Carlos Ivorra Castillo

ÁLGEBRA

Esencialmente, el álgebra y el dinero determinan clases; la primera a nivel intelectual, el segundo a nivel práctico.

Simone Weil

Índice General

Preámbulo

Este libro forma parte de la "tercera edición" de mis libros de *Álgebra* [Al], *Geometría* [G] y *Análisis Matemático* [An], de los que he separado la *Teoría de grupos* en un nuevo volumen [TG]. La diferencia esencial es que he suprimido muchos ejemplos y aplicaciones que ahora se encuentran en la serie de libros "introductorios" *Introducción a la teoría algebraica de números* [ITAl], *Introducción a la geometría euclídea* [IGE], *Introducción a la teoría analítica de números* [ITAn] e *Introducción al cálculo diferencial* [IC] y los he sustituido por contenidos nuevos.

Como en la edición anterior, los contenidos están distribuidos entre los cuatro libros de modo que pueden leerse simultáneamente siguiendo el orden que muestra el esquema de la página siguiente.

El primer capítulo de [Al] es una introducción a la teoría de conjuntos, cuyos aspectos más técnicos (los relacionados con el axioma de elección y la teoría de cardinales infinitos) se han relegado a dos apéndices. La teoría descrita es la teoría de Zermelo, que resulta más que suficiente para formalizar los contenidos de los cuatro libros. El único inconveniente es que "se queda corta" para desarrollar plenamente la teoría de cardinales infinitos, pero hemos preferido reducirla a lo imprescindible, aun al precio de no poder enunciar con total precisión algunos resultados sobre rango y dimensión de módulos y espacios vectoriales de dimensión infinita que, aunque resulta natural presentarlos al tratar estos conceptos, no son realmente necesarios en ningún momento.

El libro de Álgebra consta ahora (tras el capítulo [Al I] de fundamentos) de un primer bloque de cinco temas con los contenidos básicos del "álgebra abstracta" (incluyendo el álgebra lineal) y un segundo bloque de aplicaciones y resultados más avanzados.

El libro de Geometría empieza con dos capítulos que exponen la geometría euclídea a partir de unos axiomas al estilo de Hilbert, seguidos de un segundo bloque de cuatro capítulos dedicados a la geometría analítica y sus aplicaciones y un tercer bloque en el que analizamos más a fondo las geometrías afín y euclídea estudiadas en el bloque precedente e introducimos la geometría proyectiva y las geometrías no euclídeas.

El libro de Análisis consta de un primer bloque con los preliminares topológicos, un segundo bloque con los hechos básicos del cálculo diferencial e integral y un tercer bloque con temas más avanzados.

Álgebra	Geometría	Análisis	Grupos
Al I Conjuntos			
Al II Anillos	**G I** G. absoluta		
Al III Aritmética	**G II** G. euclídea	**An I** Números reales	**TG I** Elementos
Al IV Módulos	**G III** G. analítica	**An II** Topología I	**TG II** Permutaciones I
Al V Cuerpos I	**G IV** Cuaternios	**An III** Topología II	**TG III** Estructura
Al VI Álgebra lineal	**G V** Biyecciones afines	**An IV** Tª de la medida I	**TG IV** Resolubles
Al VII Ecuaciones	**G VI** Regla y compás	**An V** Calc. diferencial	**TG V** Nilpotentes
Al VIII Enteros algebraicos	**G VII** G. afín	**An VI** Tª de la medida II	**TG VI** Caracteres
Al IX Cuerpos II	**G VIII** G. proyectiva	**An VII** Variedades	**TG VII** Permutaciones II
Al Ap A Ax. de elección	**G IX** Cónicas	**An VIII** Cálc. vectorial	**TG VIII** Clásicos I
Al Ap B Ctos. infinitos	**G X** G. parabólica	**An IX** An. harmónico	**TG IX** Clásicos II
Al Ap C Cuadrados latinos	**G XI** G. hiperbólica	**An X** Holomorfas	**TG Ap A** Banach-Tarski
	G XII G. elíptica	**An Ap A** Sólidos rígidos	
	G Ap A G. Inversiva	**An Ap B** Gravitación	
	G Ap B Hamming		

Finalmente, el libro de Teoría de grupos aparece dividido en el esquema anterior en un primer bloque con la teoría básica y un segundo bloque con contenidos más avanzados.

Salvo los dos apéndices sobre teoría de conjuntos de [Al], los demás apéndices contienen aplicaciones que, por motivos diversos, era preferible exponer unificadamente en lugar de dejarlas dispersas por el texto.

Remitimos a las introducciones de cada uno de los libros para una panorámica general más detallada de sus contenidos.

Introducción

En [ITAl] hemos visto cómo algunos problemas que, en principio, sólo hacen referencia a los números naturales o, a lo sumo, a los enteros, requieren para ser resueltos, o pueden resolverse más fácilmente, si usamos "conceptos abstractos" tales como polinomios, números algebraicos, matrices, determinantes, etc. Sin embargo, allí presentamos sólo la teoría más rudimentaria imprescindible para abordar con tales conceptos los problemas considerados. El propósito de este libro es presentar la teoría correspondiente en un nivel adecuado de generalidad y abstracción para entender mejor la matemática abstracta subyacente y para aumentar drásticamente las posibilidades de aplicación de dicha teoría.

Resolución de ecuaciones por radicales Por ejemplo, es bien sabido que una ecuación de segundo grado

$$x^2 + bx + c = 0$$

(con coeficientes complejos, por simplificar) puede resolverse con la fórmula

$$x = \frac{-b \pm \sqrt{b^2 - 4c}}{2}.$$

En 1535 Gerolamo Cardano publicó su libro *Ars magna,* en el que incluía una fórmula (debida en realidad a Tartagila) para resolver cualquier ecuación de tercer grado:

$$x^3 + ax^2 + bx + c = 0.$$

La fórmula es bastante sofisticada. Requiere calcular:

$$p = \frac{3b - a^2}{3}, \qquad q = \frac{2a^3 - 9ab + 27c}{27}, \qquad D = \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3,$$

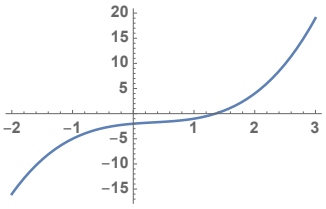
y entonces las soluciones son

$$x = \sqrt[3]{-q/2 + \sqrt{D}} - \frac{p}{3\sqrt[3]{-q/2 + \sqrt{D}}} - \frac{a}{3},$$

donde obtenemos tres soluciones considerando las tres raíces cúbicas del radicando (si cambiamos $\sqrt{D}$ por $-\sqrt{D}$ volvemos a obtener las mismas tres soluciones).

Por ejemplo, para resolver la ecuación

$$x^3 - x^2 + x - 2 = 0,$$



(la figura muestra la gráfica del polinomio) tenemos

$$a = -1, \quad b = 1, \quad c = -2,$$
$$p = \frac{2}{3}, \quad q = -\frac{47}{27}, \quad D = \frac{83}{108},$$

de donde

$$x = \sqrt[3]{\frac{\sqrt{\frac{83}{3}}}{6} + \frac{47}{54}} - \frac{2}{9\sqrt[3]{\frac{\sqrt{\frac{83}{3}}}{6} + \frac{47}{54}}} + \frac{1}{3}.$$

Si consideramos la raíz cúbica real, obtenemos $x = 1.35321\ldots$ Las otras dos raíces son imaginarias.

La *Ars magna* contenía también una fórmula para resolver ecuaciones de cuarto grado descubierta por Ludovico Ferrari, un discípulo de Cardano. Dada una ecuación

$$x^4 + ax^3 + bx^2 + cx + d = 0,$$

primero calculamos

$$p = \frac{8b - 3a^2}{8}, \quad q = \frac{8c - 4ab + a^3}{8}, \quad r = \frac{256d - 64ac + 16a^2b - 3a^4}{256},$$

luego hallamos una raíz de la ecuación cúbica:

$$P^3 - \frac{p}{2}P^2 - rP + \frac{4pr - q^2}{8} = 0,$$

con ella calculamos

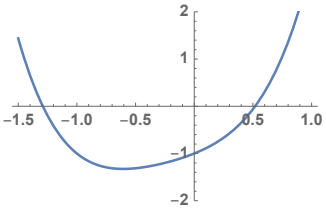
$$R = \sqrt{r - P^2}, \qquad Q = -\frac{q}{2R},$$

y entonces las soluciones de la ecuación dada son

$$x = \frac{Q \pm \sqrt{Q^2 - 4(P - R)}}{2} - \frac{a}{4}, \qquad x = \frac{-Q \pm \sqrt{Q^2 - 4(P + R)}}{2} - \frac{a}{4}.$$

Por ejemplo, para resolver la ecuación

$$x^4 + x^3 + x^2 + x - 1 = 0,$$



calculamos

$$p = q = \frac{5}{8}, \quad r = -\frac{307}{256},$$

lo que nos lleva la cúbica:

$$P^3 - \frac{5}{16}P^2 + \frac{307}{256}P - \frac{1\,735}{4\,096} = 0.$$

Para resolverla calculamos

$$p' = \frac{7}{6}, \quad q' = -\frac{65}{256}, \quad D = \frac{563}{6\,912},$$

lo que nos lleva a la solución

$$P = \sqrt[3]{\frac{\sqrt{\frac{563}{3}}}{48} + \frac{65}{432}} - \frac{7}{18\sqrt[3]{\frac{\sqrt{\frac{563}{3}}}{48} + \frac{65}{432}}} + \frac{5}{48} \approx 0.349454,$$

de donde $Q \approx -0.271859$, $R \approx 1.14949$, lo que nos da las soluciones reales

$$x \approx -1.29065, \qquad x \approx 0.51879.$$

Las otras dos soluciones son imaginarias.

Cardano observó que estas fórmulas requieren a veces calcular raíces cuadradas de números negativos, incluso cuando al final el resultado es un número real, aunque no entendía realmente lo que estaba haciendo. Fue Rafael Bombelli quien en su *Álgebra* de 1572 consideró explícitamente los números complejos, y los algebristas no tardaron en plantearse el problema de encontrar fórmulas similares a las de Cardano y Ferrari para las ecuaciones de grado 5, es decir, fórmulas que permitieran expresar las soluciones de una ecuación dada en términos de sus coeficientes mediante sumas, productos, cocientes y extracción de raíces (lo que se abrevia diciendo que la ecuación es "resoluble por radicales").

Paralelamente, en 1629 Albert Girard conjeturó algo más débil, a saber, que toda ecuación polinómica de grado n debía tener n raíces complejas (contando sus multiplicidades). Muchos matemáticos trataron de demostrar esta conjetura (aunque Leibniz y Nicolás Bernoulli creyeron que era falsa). El primer intento de demostrarlo lo hizo D'Alembert en 1746, pero su prueba tenía una laguna. También Euler (1749), Lagrange (1772), Laplace (1795) y Gauss (1799), entre otros, publicaron pruebas incompletas. Muchos de ellos suponían implícitamente que el polinomio tenía solución, y lo que demostraban era que podía expresarse en la forma $a + bi$. La primera prueba correcta de lo que hoy se conoce como el teorema fundamental del álgebra la publicó un matemático aficionado, Jean-Robert Argrand, en 1814. Sin embargo, esto demostraba que los polinomios siempre tienen raíces complejas, pero no que fueran resolubles por radicales.

En 1799 Paolo Ruffini publicó una demostración incompleta, que fue corregida por Abel en 1813, según la cual no existe ninguna fórmula general análoga a las de Cardano y Ferrari (o a la de la ecuación de segundo grado) aplicable a cualquier ecuación de grado 5 que exprese sus raíces en términos de sus coeficientes mediante sumas, restas, productos, cocientes o raíces, pero el argumento no excluía que las raíces de cada ecuación particular pudieran expresarse en términos de sus coeficientes mediante una fórmula distinta en cada caso. En 1832, Évariste Galois encontró un criterio para determinar si una ecuación dada es resoluble por radicales, y probó que existen ecuaciones de cualquier grado

mayor o igual que 5 no resolubles por radicales. Su demostración permaneció inédita hasta 1846, cuando la divulgó Joseph Liouville. (Galois escribió precipitadamente sus resultados a sus 20 años porque sabía que iba a morir en un duelo, y así fue.)

El teorema fundamental del álgebra lo demostramos ya en [ITAn 3.33], y la justificación de las fórmulas de Cardano y Ferrari es elemental, por lo que podríamos perfectamente haberla incluido en [ITAl], pero la existencia de ecuaciones polinómicas no resolubles por radicales no podríamos haberla demostrado con las técnicas expuestas en [ITAl] o [ITAn] o incluso [IC]. Con los resultados que expondremos en este libro sí dispondremos de la teoría necesaria.

Construcciones con regla y compás Otro ejemplo de resultado que no pudimos probar en [ITAl] por falta de la teoría necesaria es la caracterización de Wantzel de los números complejos constructibles con regla y compás. En [IGE 6.16] demostramos que todos los puntos constructibles con regla y compás (identificados con números complejos) son algebraicos, lo que implica inmediatamente la imposibilidad de cuadrar el círculo o rectificar la circunferencia con regla y compás. Sin embargo, con las técnicas disponibles en [IGE] o en [ITAl], no estábamos en condiciones de demostrar que, además, los números constructibles tienen que ser raíces de polinomios de grado potencia de 2, lo que —según vimos— implica que el problema de la duplicación del cubo es irresoluble con regla y compás, al igual que el de la trisección del ángulo. Sucede que la misma teoría de Galois que permite estudiar la resolubilidad por radicales de una ecuación polinómica dada permite también caracterizar los números complejos constructibles con regla y compás, y demostrar, por ejemplo, el teorema siguiente, que fue enunciado por Gauss en sus *Disquisitiones arithmeticae* de 1801, si bien la primera demostración la publicó Wantzel en 1837:

> *El polígono regular de n lados es constructible con regla y compás si y sólo si* $n = 2^k p_1 \cdots p_m$*, donde los* p_i *son primos de Fermat distintos entre sí.*

En particular, los polígonos regulares de 7 o 9 lados no son constructibles con regla y compás.

Por su naturaleza geométrica, esta aplicación (y muchas otras) las expondremos en el capítulo VI de [G].

La teoría de conjuntos En [ITAl] demostramos todos los resultados que utilizamos a partir de hechos evidentes cuya validez no puede ser cuestionada salvo por alguien que se tome la filosofía demasiado en serio. Sin embargo, precisamente para dejar la exposición de las matemáticas claramente a salvo de preguntas demasiado filosóficas, resulta útil constatar que todos esos hechos evidentes pueden demostrarse a partir de unos pocos axiomas, los axiomas de la teoría de conjuntos, de modo que una "definición operativa" de "matemáticas" sería considerar que las matemáticas estudian las consecuencias de los axiomas

de la teoría de conjuntos.[1] Por otro lado, la teoría de conjuntos no sólo nos proporciona unos axiomas capaces de marcar una frontera definida entre lo que es una demostración matemática rigurosa y algo no concluyente, sino que dota a la matemática de un lenguaje muy cómodo y potente para expresar (casi) cualquier teoría matemática. Dedicamos el capítulo I a presentar dicho lenguaje y dichos axiomas.

Por conveniencia hemos aislado en el apéndice A el análisis del último de los axiomas (el axioma de elección) pues sus consecuencias más importantes requieren argumentos más técnicos y sofisticados que los resultados típicos de este libro y en el apéndice B probaremos también algunos resultados técnicos sobre conjuntos infinitos que tal vez el lector prefiera considerar únicamente en el momento en que vaya a necesitarlos realmente.

Los axiomas de la teoría de conjuntos permiten reducir todos los conceptos matemáticos a dos conceptos básicos o primitivos: el concepto de conjunto y el concepto de pertenencia entre conjuntos. Esto hace que cualquier otro concepto, por elemental que sea, deba definirse a partir de éstos dos, lo que obliga a dedicar cierto tiempo a dar definiciones artificiosas de conceptos inmediatos y demostrar laboriosamente cosas "evidentes". Un caso típico es la definición 1.5 de los números naturales y la demostración de sus propiedades. Por ejemplo, la artificiosa demostración del teorema 1.7 no debe verse realmente como una demostración de que todo número natural distinto de 0 es el siguiente de otro número natural, aunque eso es ni más ni menos lo que afirma el enunciado, sino como parte de la justificación de que los números naturales definidos en el seno de la teoría de conjuntos se comportan como cabe esperar que tienen que comportarse los números naturales. Si pudiéramos demostrar que hay un número natural distinto de 0 que no es el siguiente de otro número natural, la conclusión no sería que estábamos equivocados al creer que era así, sino que la definición conjuntista dada de los números naturales sería inadmisible. Otros ejemplos son las demostraciones de las propiedades elementales de los conjuntos finitos que veremos en la sección 1.6 o las propiedades de las sumas y productos finitos, que demostraremos en el apéndice al capítulo II.

Pero lo más importante de este primer capítulo no son los axiomas ni las demostraciones artificiosas, sino el propio lenguaje conjuntista que en [ITAl] usamos muy restringidamente. El lector debe familiarizarse con los conceptos de "producto cartesiano", "relación", "relación de equivalencia", "relación de orden",

[1] Puede objetarse que algunas matemáticas requieren axiomas adicionales, como la famosa *hipótesis del continuo,* que no es demostrable ni refutable a partir de los axiomas de los que hablamos, pero a eso puede contraobjetarse que un teorema que se apoye en la hipótesis del continuo es también una demostración matemática en el sentido que hemos indicado, a saber, que una consecuencia de los axiomas de la teoría de conjuntos es que la hipótesis del continuo implica el teorema en cuestión. Una objeción más seria a nuestra "definición operativa" es que hay otros sistemas axiomáticos incompatibles con el sistema típico que nosotros vamos a considerar que también dan lugar a matemáticas interesantes que estamos excluyendo con nuestra "definición operativa", pero precisamente por eso la llamamos "definición operativa", porque nos sirve como guía para definir las matemáticas que vamos a describir aquí (y para las expuestas en el 99.99% de los libros y artículos de matemáticas publicados a lo largo de la historia) aunque no sea adecuada para tratar con otras teorías muy alejadas de nuestro propósito.

"aplicación inyectiva, suprayectiva, biyectiva", "dominio", etc., que impregnan cualquier exposición moderna de cualquier rama de la matemática. Hay que incluir en este bloque las "leyes de composición interna" y todos los conceptos relacionados (conmutatividad, asociatividad, etc.) que presentamos en la sección 2.1, ya en el capítulo II, mientras que la sección 2.2 contiene otro ejemplo de "teoría artificiosa" que enmarca los números enteros en el contexto de la teoría de conjuntos, pero que el lector puede ver como un buen ejercicio de uso de buena parte de los conceptos conjuntistas introducidos hasta el momento.

Anillos Podríamos considerar que es en la sección 2.3 donde empieza el contenido algebraico propiamente dicho de este libro. A partir de aquí, el capítulo II se corresponde bastante fielmente con los resultados sobre anillos probados en [ITAl], salvo que consideramos anillos de polinomios con cualquier número de indeterminadas (finito o infinito) y que demostramos algunos resultados fundamentales no triviales sobre divisibilidad de polinomios.

El tratamiento de la aritmética en dominios íntegros que presentamos en el capítulo III se caracteriza por introducir desde el primer momento el concepto de ideal, mientras que en [ITAl] esperamos hasta el capítulo XIII para introducirlos motivadamente. Por lo demás, salvo por el énfasis en el concepto de ideal, su contenido también se ajusta bastante a la exposición de la divisibilidad en dominios íntegros vista en [ITAl].

Módulos y espacios vectoriales En [ITAl] pudimos ver cómo el concepto general de "anillo" (con posibles condiciones adicionales que definen clases importantes de anillos, como los dominios íntegros, los dominios de factorización única, etc.) permite aplicar principios generales válidos en principio para la aritmética de los números enteros a muchos otros casos de interés, de modo que objetos aparentemente muy distintos de los números enteros pueden ser tratados hasta cierto punto como si fueran números enteros. Del mismo modo, en el capítulo IV introduciremos la teoría básica sobre módulos y espacios vectoriales, de los que en [ITAl] vimos sólo casos muy particulares, que también nos permite aplicar resultados generales a contextos muy diversos, mucho más generales que aquellos en los que la teoría de anillos es aplicable.

Si A es un anillo unitario, un A-módulo M es un conjunto en el que tenemos definidas dos operaciones, que podemos hablar de la suma $r+s$ de dos elementos de M y del producto ar de un elemento de A por un elemento de M, y de modo que se cumplan una serie de propiedades que enumeramos a continuación simplemente para que el lector se forme una primera idea superficial de lo que es la estructura de módulo. En esencia, tenemos un módulo allí donde podemos plantear expresiones de la forma

$$a_1r_1 + \cdots + a_nr_n,$$

donde los factores de la izquierda están en un anillo unitario y los de la derecha en un conjunto M en el que tenemos definida una suma, pero no el producto de dos elementos de M cualesquiera. Cuando el anillo A es un cuerpo es costumbre llamar A-espacios vectoriales a los A-módulos.

Axiomas de módulo

1. $(r+s)+t = r+(s+t)$ para todos los r, s, $t \in M$.
2. $r+s = s+r$ para todos los r, $s \in M$.
3. Existe un elemento $0 \in M$ tal que $r+0 = r$ para todo $r \in M$.
4. Para todo $r \in M$ existe un elemento $-r \in M$ tal que $r+(-r) = 0$.
5. $a(r+s) = ar+as$ para todo $a \in A$ y todos los r, $s \in M$.
6. $(a+b)r = ar+br$ para todos los a, $b \in A$ y todo $r \in M$.
7. $a(br) = (ab)r$ para todos los a, $b \in A$ y todo $r \in M$.
8. $1r = r$ para todo $r \in M$.

Por ejemplo, todo elemento v de $\mathbb{R}^3$ se puede expresar de forma única como

$$v = a_1e_1 + a_2e_2 + a_3e_3,$$

donde a_1, a_2, a_3 son números reales y $e_1 = (1,0,0)$, $e_2 = (0,1,0)$, $e_3 = (0,0,1)$, lo que se expresa diciendo que $\mathbb{R}^3$ es un $\mathbb{R}$-espacio vectorial "generado" por e_1, e_2, e_3.

Igualmente, todo entero ciclotómico de orden 5 (véase la sección [ITAl 17.2]) se puede expresar en la forma

$$\alpha = a_0 + a_1\omega + a_2\omega^2 + a_3\omega^3,$$

donde a_0, a_1, a_2, a_3 son números enteros y $\omega \neq 1$ es una raíz quinta de la unidad, lo que se expresa diciendo que el anillo de los enteros ciclotómicos es un $\mathbb{Z}$-módulo "generado" por $1, \omega, \omega^2, \omega^3$.

Vemos así que la analogía entre ambos contextos no es superficial, sino que se debe a que se trata de dos casos particulares del concepto de "módulo", lo que nos permite aplicar a ambos todos los conceptos y resultados de la teoría general sobre módulos, como el concepto de "sistema generador". Más precisamente, las expresiones de la forma $a_1r_1 + \cdots + a_nr_n$ se llaman "combinaciones linales" de $r_1, \ldots, r_n$, por lo que un sistema generador de un módulo M es un conjunto de elemento de M tales que cualquier otro se puede expresar como combinación lineal de generadores. En los dos ejemplos que hemos puesto se cumple que los coeficientes a_i están unívocamente determinados, de modo que cada elemento del módulo correspondiente se puede expresar *de forma única* como combinación lineal de los generadores. Cuando un generador tiene esta propiedad de unicidad se dice que es una "base" del módulo y los coeficientes unívocamente determinados se llaman "coordenadas" del elemento que determinan.

Por ejemplo, uno de los resultados no triviales de la teoría de módulos afirma que dos bases de un mismo módulo sobre un anillo conmutativo y unitario tienen necesariamente el mismo número de elementos, el cual recibe el nombre de

"rango" del módulo (aunque en los espacios vectoriales se llama "dimensión"). En particular podemos decir que $\mathbb{R}^3$ es un espacio vectorial de dimensión 3, porque cada uno de sus elementos está determinado por tres coordenadas reales, y que el anillo de los enteros ciclotómicos de orden 5 es un $\mathbb{Z}$-módulo de rango 4, porque cada entero ciclotómico está determinado por 4 coordenadas enteras.

Así, la teoría de módulos nos proporciona una serie de resultados generales para trabajar con objetos determinables por coordenadas en un anillo como en los dos ejemplos anteriores. Esto se aplica en particular a la geometría analítica descrita en los respectivos apéndices A de [IGE] e [IC], de modo que, cuando decimos que las rectas tienen una dimensión, que los planos tienen dos dimensiones y que el espacio tiene tres dimensiones, estamos calculando la dimensión de ciertos espacios vectoriales asociados a las rectas, a los planos y al espacio. Esta conexión la desarrollaremos con detalle en el capítulo III de [G].

Al tratar con módulos y espacios vectoriales aparece de forma natural el concepto de "matriz". En la sección 4.4 generalizamos algunos de los resultados sobre matrices que en la sección [ITAl 11.2] presentamos para matrices 2×2 para el caso de matrices de dimensiones arbitrarias, y en la sección 6.1 generalizamos el concepto de determinante que en [ITAl 11.6] introdujimos para matrices 2×2 y en el apéndice A [IC] definimos para matrices 3×3:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad-bc, \qquad \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei+bfg+cdh-ceg-afh-bdi.$$

Por ejemplo, en [ITAl 11.7] vimos que una matriz 2×2 tiene inversa si y sólo si su determinante es una unidad, pero adaptar la prueba para comprobar que también es válida para matrices 3×3 resulta muy farragoso, y no aporta ninguna pista sobre cómo habría que definir un determinante 4×4 para que la propiedad siguiera siendo cierta en este caso. En la sección 6.1 daremos una definición general y muy conceptual de "determinante de una matriz" que nos permitirá probar fácilmente las propiedades básicas de los determinantes de matrices de cualquier tamaño, lo que nos dotará de una herramienta algebraica muy potente.

En general, en el capítulo VI presentamos algunos aspectos más avanzados de la teoría de módulos para los que es preferible haber visto antes algunos resultados del capítulo V y del capítulo IV de [G]. Un ejemplo de aplicación de los conceptos básicos del álgebra lineal se encuentra en el apéndice C, donde estudiaremos "el problema de los 36 oficiales", que Euler no pudo resolver:

> *El emperador se disponía a visitar una ciudad en la que estaban acuartelados seis regimientos, y el comandante de la guarnición seleccionó seis oficiales de distinta graduación en cada uno de ellos y quiso disponerlos en formación* 6×6 *para que el emperador pasara revista, y de modo que, cualquiera que fuera la fila o la columna que éste decidiera recorrer, encontrara en ella un oficial de cada regimiento y uno de cada una de las graduaciones. ¿Cómo había que disponer para ello a los* 36 *oficiales?*

(La introducción y las dos primeras secciones del apéndice se pueden leer sin necesidad de ningún conocimiento algebraico previo.)

En la sección 4.5 probaremos un profundo teorema general que nos dará la estructura de los módulos finitamente generados sobre un dominio de ideales principales, que tiene aplicaciones a contextos muy diversos. Por una parte, nos permitirá clasificar todos los grupos abelianos finitamente generados (en particular, nos permitirá calcular fácilmente de forma explícita todos los grupos abelianos finitos que hay de un orden dado). En la sección 6.4 lo usaremos para estudiar la estructura de los endomorfismos de un espacio vectorial y en el capítulo V de [G] nos permitirá clasificar todas las isometrías de un espacio euclídeo de cualquier dimensión (todas las aplicaciones que conservan las distancias entre puntos).

Cuerpos La teoría general sobre espacios vectoriales presentada en el capítulo IV es un ingrediente fundamental para el estudio que hacemos en el capítulo V de las extensiones de cuerpos, es decir, para estudiar la relación entre un cuerpo y un subcuerpo prefijado. La conexión básica consiste en que si tenemos un cuerpo y un subcuerpo, $k \subset K$, entonces K es un k-espacio vectorial. Cuando su dimensión es finita se dice que K/k es una extensión finita de cuerpos, la dimensión se llama "grado" de la extensión, y se representa por $|K : k|$.

Por ejemplo, $|\mathbb{C} : \mathbb{R}| = 2$, pues todo número complejo se expresa de forma única como $a+bi$, donde a, b son números reales, lo que significa que $1, i$ forman una base de $\mathbb{C}$ como $\mathbb{R}$-espacio vectorial. Similarmente, si llamamos $\mathbb{Q}(\omega)$ al cuerpo de los números ciclotómicos de orden 5, formado por todos los números complejos de la forma

$$a_0 + a_1\omega + a_2\omega^2 + a_3\omega^3,$$

donde los coeficientes son números racionales y $\omega \neq 1$ es una raíz quinta de la unidad, tenemos que $|\mathbb{Q}(\omega) : \mathbb{Q}| = 4$.

Un hecho fundamental es la transitividad de grados: si tenemos tres cuerpos $k \subset L \subset K$, se cumple que

$$|K : k| = |K : L||L : k|.$$

Así, por ejemplo, en [G 6.4] Demostraremos que un número complejo z es constructible con regla y compás si y sólo si existe una cadena de cuerpos

$$\mathbb{Q} = F_0 \subset F_1 \subset \cdots \subset F_n \subset \mathbb{C}$$

tal que $z \in F_n$ y cada extensión F_i/F_{i-1} tiene grado 2. En particular, esto implica que la extensión $F_n/\mathbb{Q}$ tiene grado 2^n y de ahí se puede deducir que z tiene que ser raíz de un polinomio de grado potencia de 2, que es lo que nos faltó probar en [IGE] para concluir que los problemas de la trisección del ángulo y la duplicación del cubo no son resolubles con regla y compás.

Sin embargo, que z esté en un subcuerpo de $\mathbb{C}$ que tenga grado potencia de 2 sobre $\mathbb{Q}$ es una condición necesaria, pero, en principio, no suficiente, para que

z sea constructible con regla y compás, por lo que no basta para probar, por ejemplo, la caracterización de los polígonos regulares constructibles con regla y compás que hemos enunciado más arriba. La condición necesaria y suficiente es la existencia de la cadena indicada de cuerpos intermedios.

Algo similar sucede con la resolubilidad por radicales de una ecuación polinómica, que puede caracterizarse con relativa facilidad en términos de la existencia de una cadena de subcuerpos de $\mathbb{C}$ que empieza en $\mathbb{Q}$, cuyo último término contiene a las raíces del polinomio, y cuyos eslabones satisfacen ciertas condiciones. De este modo, tanto el problema de si un determinado número complejo es constructible con regla y compás como el problema de si un polinomio es resoluble con radicales se pueden reducir a determinar si un cierto subcuerpo del cuerpo $\mathbb{C}$ de los números complejos puede expresarse como el último término de una cadena de subcuerpos en ciertas condiciones.

Ahora bien, en principio, no es fácil encontrar subcuerpos de un cuerpo dado. Por ejemplo, aunque tengamos una descripción muy simple de los números ciclotómicos de orden 5, con la que no es difícil identificar ciertos subcuerpos, no tenemos ningún criterio para encontrarlos todos. De hecho, ni siquiera es evidente que una extensión finita de cuerpos tenga sólo un número finito de subcuerpos, como de hecho sucede. Es en este punto donde interviene la teoría de Galois. El teorema fundamental de la teoría de Galois establece que, dada una extensión finita de cuerpos $k \subset K$, bajo ciertas hipótesis que, en el caso de subcuerpos de $\mathbb{C}$, siempre pueden garantizarse cambiando K por un cuerpo mayor, los subcuerpos intermedios $k \subset L \subset K$ están en correspondencia biunívoca con los subgrupos del "grupo de Galois" $G(K/k)$ formado por todos los automorfismos de K que dejan fijos a los elementos de k, que es un grupo finito.

Por ejemplo, si K es el cuerpo de los números ciclotómicos de orden p, el grupo $G(K/\mathbb{Q})$ está formado por las conjugaciones definidas en [ITAl 17.5], y allí probamos que se corresponden con los elementos del grupo U_p de las unidades del anillo de restos $\mathbb{Z}_p$, que tiene orden $p-1$. La teoría de Galois permite calcular explícitamente los cuerpos intermedios a partir de los subgrupos de U_p, que son fáciles de calcular.

En general, la teoría de Galois permite reducir problemas sobre cuerpos a problemas mucho más simples sobre grupos. Por ejemplo, en el caso de la constructibilidad de los polígonos regulares, permite probar que una raíz de la unidad compleja (un vértice de un polígono regular de radio unitario) es constructible con regla y compás si y sólo si es raíz de un polinomio de grado potencia de 2, lo cual se traduce fácilmente en la condición en términos de los primos de Fermat que hemos enunciado más arriba. En el caso de la resolubilidad por radicales, la teoría de Galois reduce el problema de si un polinomio es resoluble por radicales a que un determinado grupo de Galois G tenga una propiedad llamada precisamente por ello "resolubilidad", que no es sino la existencia de una cadena de subgrupos con ciertas propiedades. Esto lo veremos en el capítulo VII, en los que daremos también los detalles sobre las fórmulas de Cardano y Ferrari para la resolución de ecuaciones cúbicas y bicuadráticas.

Enteros algebraicos En el capítulo XIII de [ITAl] demostramos que los anillos de enteros algebraicos de los cuerpos cuadráticos tienen factorización única ideal. Esto es cierto para los anillos de enteros algebraicos de todos los subcuerpos de $\mathbb{C}$ de grado finito sobre $\mathbb{Q}$, pero las técnicas con las que contábamos en [ITAl] no nos permitían ir más allá de los cuerpos cuadráticos. En el capítulo VIII demostraremos que si K es un "cuerpo numérico" de grado n (un subcuerpo de $\mathbb{C}$ de grado n sobre $\mathbb{Q}$) su anillo de enteros algebraicos es un $\mathbb{Z}$-módulo de rango n, lo que nos permitirá expresar los enteros algebraicos de K como combinaciones lineales de una "base entera" prefijada. La teoría de cuerpos y la teoría de módulos desarrollada en los capítulos precedentes nos permitirá definir normas y estudiar la divisibilidad en anillos de enteros algebraicos sobre cuerpos numéricos arbitrarios. En 8.32 caracterizaremos la factorización ideal en un dominio íntegro arbitrario en términos de tres propiedades algebraicas sencillas y probaremos que se cumplen en los anillos de enteros algebraicos de cuerpos numéricos. Definiremos el "grupo de clases", que en [ITAl] sólo pudimos definir para cuerpos cuadráticos, y demostraremos que es finito en el caso de los anillos de enteros ciclotómicos de orden primo. Con ello podremos demostrar el teorema de Kummer que da una condición suficiente para que se cumpla el Último Teorema de Fermat para un exponente p.

La condición suficiente que daremos es la primera que obtuvo Kummer, pero no es fácil de comprobar en la práctica y el propio Kummer la simplificó posteriormente. Sin embargo, el argumento requiere técnicas de teoría de números que van más allá del contenido de este libro. Dichas técnicas también permiten mejorar sustancialmente muchos resultados del capítulo VIII. Por ejemplo, puede probarse que el grupo de clases de cualquier cuerpo numérico es finito. No obstante, aunque los resultados probados en este capítulo no sean todo lo generales que podrían ser, lo hemos incluido porque ilustran la potencia del álgebra presentada en este libro frente a los rudimentos algebraicos con los que contábamos en [ITAl].

Finalmente, el capítulo IX contiene varios resultados de naturaleza diversa sobre cuerpos que no han sido necesarios en los capítulos precedentes pero que tienen interés en teorías más avanzadas, como la teoría algebraica de números, la cohomología de grupos o la geometría algebraica. (La sección 9.1 sobre cuerpos finitos la usaremos en [G] y en [TG].) Probablemente el lector sabrá apreciar el interés por sí mismos de muchos de ellos, como el estudio de los cuerpos finitos o de las extensiones trascendentes, pero tal vez prefiera pasar por alto este capítulo o sus secciones más técnicas (especialmente las últimas) mientras no le sean necesarias.

Capítulo I

El lenguaje de la teoría de conjuntos

Es habitual distinguir distintas "ramas" o "especialidades" dentro de la matemática según sus objetos de estudio respectivos, como el álgebra, la geometría, la topología, el análisis matemático, etc. Ninguna de ellas puede considerarse aislada y claramente delimitada de las restantes, sino que todas tienen múltiples conexiones entre sí. En principio, la teoría de conjuntos es una más de estas "ramas", la que tiene por objeto el estudio de los conjuntos en general. Sin embargo, el lenguaje que los matemáticos han desarrollado para el estudio de los conjuntos en general ha resultado ser el más adecuado para expresar con claridad y precisión los conceptos y resultados de todas las demás ramas de la matemática, y los hechos más elementales de la teoría de conjuntos han resultado ser el punto de partida idóneo para todas ellas. Es por ello que dedicamos este primer capítulo a introducir el lenguaje, los conceptos y los resultados más elementales de la teoría de conjuntos, a modo de base sobre la cual presentaremos en los capítulos siguientes los resultados algebraicos que constituyen el auténtico objeto de este volumen.

1.1 Conjuntos

Toda la matemática puede construirse a partir de dos conceptos fundamentales: el concepto de conjunto y el de pertenencia. Cuando decimos que son conceptos fundamentales (o primitivos) queremos decir que no es posible dar definiciones operativas de estos conceptos, es decir, definiciones a partir de las cuales podamos deducir lógicamente sus propiedades, al contrario de lo que sucede con cualquier concepto matemático que no sea uno de estos dos. Si intentamos dar una definición de "conjunto", lo máximo que podemos decir es que un conjunto es una colección de objetos, y esto no es operativo porque "conjunto" y "colección" son sinónimos, con lo que realmente no hemos definido nada. Aun así, con esta "definición" obtenemos una idea informal de lo que pretendemos

que sea un "conjunto" en el sentido técnico específico que tiene esta palabra en matemáticas, idea que podemos terminar de perfilar con algunos ejemplos:

Ejemplos Si escribimos $A = \{a, b, c\}$, entenderemos que A es un conjunto que tiene tres[1] elementos a, b, c. Para expresar que a es uno de los elementos del conjunto A escribiremos[2] $a \in A$, y leeremos "el elemento a pertenece al conjunto A". Por el contrario, si d es un elemento distinto de a, b y c, escribiremos $d \notin A$ para indicar que el elemento d no pertenece al conjunto A. Con esto hemos introducido el segundo (y último) concepto fundamental de la teoría de conjuntos, la relación de pertenencia.

Hay que precisar que, aunque hayamos "definido" un conjunto como una colección de elementos, no hay que entender de ahí que todo conjunto deba tener al menos un elemento. Al contrario, vamos a admitir la existencia de un conjunto vacío, que representaremos por $\varnothing$, determinado por la propiedad de no tener elementos. Podemos pensar que un conjunto es como una bolsa, que es susceptible de contener elementos, pero que nada impide que esté vacía.

Es muy importante tener presente que el concepto matemático de conjunto permite que unos conjuntos sean a su vez elementos de otros conjuntos. Por ejemplo, el conjunto $B = \{a, \{b, c\}\}$ no debe ser confundido con el conjunto A anterior. Si suponemos que a, b, c son distintos entre sí, entonces A es un conjunto con tres elementos, mientras que B es un conjunto con dos elementos, uno de ellos es a, que es común a A y a B, y el otro es el conjunto $\{b, c\}$, de modo que

$$\{b, c\} \in B, \qquad \{b, c\} \notin A.$$

Esto podría llevarnos a concluir que en el lenguaje de la teoría de conjuntos hay dos clases de objetos: los *elementos puros* como a, b, c, que no son conjuntos, sino objetos susceptibles de pertenecer a conjuntos, y los conjuntos como A y B, formados por tales elementos o por otros conjuntos. El lector puede pensar que esto es así si lo prefiere, pero la realidad es que los únicos conjuntos que vamos a necesitar en todo momento son los *conjuntos puros,* es decir, los conjuntos cuyos elementos son todos conjuntos, formados a su vez por conjuntos, todos ellos formados a su vez por conjuntos, y así sucesivamente. No hay ningún inconveniente técnico en suponer que existen elementos puros, pero lo cierto es que en ningún momento nos van a hacer falta para nada, por lo que es más práctico restringir el significado técnico de la palabra "conjunto" en matemáticas para entender que significa "conjunto puro".

Tanto si el lector prefiere pensar que existen elementos puros —y, por lo tanto, conjuntos "mixtos" que los incluyen en su composición— como si prefiere adoptar el convenio usual de trabajar únicamente con conjuntos puros, el hecho es que todos los conceptos matemáticos que vamos a definir en este libro (o en

[1] En realidad, dicha notación no presupone que, por ejemplo, tenga que ser $a \neq b$. Si se diera este caso, A tendría dos elementos, o uno solo si $a = b = c$.

[2] El signo $\in$ es una deformación de la letra griega ϵ y fue introducido como abreviatura de la palabra griega $\epsilon\sigma\tau\acute{\iota}$ (está), de modo que $a \in A$ es una abreviatura de "a está en A".

cualquier otro libro que no entre en cuestiones muy específicas de la teoría de conjuntos) serán conjuntos puros.

Esto no significa que una expresión como $A = \{a, b, c\}$ sea inaceptable por contener "elementos puros", sino que meramente estamos adoptando el convenio de que si consideramos un conjunto de este tipo, sus elementos a, b y c serán a su vez otros conjuntos, cuya composición puede que sea irrelevante para las consideraciones que queramos hacer, por lo que tal vez nunca lleguemos a precisarla, pero eso no significa que a no tenga a su vez sus elementos, digamos $a = \{x, y\}$, donde x e y serán nuevos conjuntos cuya composición podríamos investigar a su vez, etc. ∎

Los ejemplos precedentes pretenden precisar el sentido que los matemáticos dan a las palabras "conjunto" y "pertenencia", pero fundamentar una teoría matemática en una definición circular de conjunto y en unos ejemplos aclaratorios dista mucho del canon de rigor que se exige en la matemática formal. La forma válida de precisar operativamente el significado de las palabras "conjunto" y "pertenencia" no es a través de unas definiciones necesariamente insatisfactorias ni a través de unos ejemplos, sino a través de unos axiomas que constituyan las afirmaciones básicas que aceptaremos que cumplen los conjuntos, de las cuales se pueden deducir todas las demás. Dedicaremos la sección siguiente a presentar los axiomas básicos de la teoría de conjuntos.

1.2 Los axiomas de la teoría de conjuntos

La idea de "conjunto" como colección de objetos está contenida en el siguiente axioma fundamental:

Axioma de extensionalidad *Dos conjuntos son iguales si y sólo si tienen los mismos elementos.*

En términos prácticos, si tenemos dos conjuntos A y B y podemos demostrar que, bajo el supuesto de que un x cumple $x \in A$, necesariamente $x \in B$ y viceversa, entonces este axioma nos permite concluir que $A = B$.

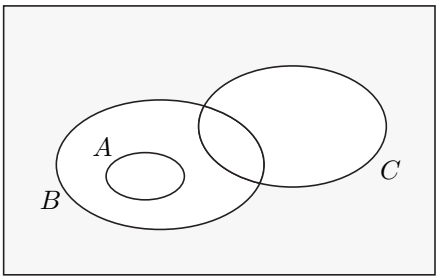
Más conceptualmente, lo que dice el axioma de extensionalidad es que si dos conjuntos A y B se diferencian en algo, necesariamente tiene que ser en sus elementos. Los conjuntos no tienen ningún otro rasgo distintivo aparte de sus elementos. Esta concepción de los conjuntos puede no ser la más adecuada en otros contextos. Por ejemplo, uno podría considerar que un equipo de fútbol es algo distinto de un equipo de béisbol, incluso si ambos están formados por los mismos jugadores. Lo que dice el axioma de extensionalidad es que en el sentido matemático de "conjunto" esto no es admisible. No hay "conjuntos de fútbol o de béisbol", ni "conjuntos rojos o verdes". Las únicas propiedades de los conjuntos son las determinadas por quiénes son sus elementos.

A menudo es conveniente descomponer el contenido del axioma de extensionalidad en dos partes a través del concepto de inclusión:

Definición 1.1 Diremos que un conjunto A es un *subconjunto* de un conjunto B o que A *está contenido, o incluido, en* B, y lo representaremos por $A \subset B$, si todo elemento de A es también un elemento de B.

En estos términos, el axioma de extensionalidad afirma que una igualdad de conjuntos equivale a una doble inclusión: $A = B$ si y sólo si $A \subset B$ y $B \subset A$.

A veces es útil visualizar los conjuntos mediante los llamados *diagramas de Venn.* La figura muestra la representación de tres conjuntos A, B, C, donde hay que entender que los elementos de A están representados por los puntos encerrados por la curva que tiene la A junto a ella. Vemos entonces que $A \subset B$, pero no $A = B$, pues hay elementos de B que no están dentro de A.



Para indicar que se da una inclusión $A \subset B$, pero que ésta es *estricta,* es decir, que todo elemento de A pertenece también a B, pero existen elementos de B que no están en A (tal y como les sucede a los conjuntos de la figura), usaremos la notación $A \subsetneq B$.

Esto equivale a que $A \subset B$ y $A \neq B$, o también a que $A \subset B$ y $B \not\subset A$. La última notación es la que emplearemos para indicar que B no es un subconjunto de A. Así, por ejemplo, la figura muestra que $B \not\subset C$ y $C \not\subset B$.

Notemos que otras propiedades obvias de la inclusión de conjuntos son las siguientes: $A \subset A$ (todo conjunto es un subconjunto de sí mismo) y que si $A \subset B$ y $B \subset C$ entonces $A \subset C$ (si todo elemento de A está en B y todo elemento de B está en C es evidente que todo elemento de A está en C).

En este punto debemos señalar que la noción de conjunto que estamos manejando es tan general y abstracta que resulta peligroso hacer afirmaciones generales a la ligera sobre lo que podemos esperar que cumplan los conjuntos. Por ejemplo, en contextos no matemáticos, podemos hablar del conjunto de todos los planetas del sistema solar, del conjunto de los dedos de mi mano, etc., y podemos sentirnos tentados de extrapolar de ahí un principio general:

Axioma de comprensión *Dada cualquier propiedad*[3] P*, existe un conjunto cuyos elementos son exactamente los conjuntos que cumplen la propiedad* P*.*

[3] Quizá el lector eche en falta un concepto preciso de "propiedad". Ciertamente, es necesario especificar qué entendemos por "propiedad" para que un axioma como éste pueda tenerse por preciso y riguroso, pero una determinación del concepto de "propiedad" requiere tecnicismos relacionados con la lógica matemática, y no vamos a entrar en ello aquí. Baste saber que es posible dar una definición totalmente satisfactoria del concepto de "propiedad" y que, en la práctica, la idea subyacente es que una propiedad es cualquier propiedad —valga la circularidad que, como decimos, es evitable— definible con precisión en términos de los conceptos primitivos de conjunto y pertenencia o a partir de cualesquiera otros previamente definidos a partir de ellos, como el de "inclusión".

Notemos que, para cada propiedad P, sólo puede haber un conjunto cuyos elementos sean precisamente los conjuntos que cumplen la propiedad P, pues si hubiera dos, tendrían exactamente los mismos elementos (los conjuntos con la propiedad P), luego el axioma de extensionalidad implicaría que ambos son en realidad el mismo conjunto. Por lo tanto, si aceptamos el axioma de comprensión, para cada propiedad P podemos representar mediante $\{x \mid Px\}$ (léase "el conjunto de todos los x que cumplen Px") el (único) conjunto de todos los conjuntos que tienen la propiedad P.

Por desgracia, no podemos admitir el axioma de comprensión, ya que resulta ser contradictorio. En efecto, basta considerar la propiedad $Px \equiv x \notin x$, que nos daría el conjunto

$$R = \{x \mid x \notin x\},$$

es decir, el conjunto de todos los conjuntos que no son elementos de sí mismos. Este "conjunto" resulta ser contradictorio, ya que llegamos a un absurdo tanto si suponemos que $R \in R$ como si suponemos que $R \notin R$. En el primer caso, por ser un elemento de R debe cumplir la propiedad P que lo define, es decir, debe cumplir $R \notin R$, que es justo lo contrario de lo que habíamos supuesto. Si, por el contrario, suponemos que $R \notin R$, entonces R tiene la propiedad P que define a R, luego debería cumplir $R \in R$, y de nuevo tenemos una contradicción.

Esto se conoce como *paradoja de Russell,* y nos permite concluir que el axioma de comprensión afirma la existencia de conjuntos que no pueden existir, y por ello debemos descartarlo.

A los matemáticos les llevó un tiempo encontrar la manera de hablar rigurosamente de conjuntos sin poder postular que las propiedades definen conjuntos, y la solución más simple la obtuvo Ernst Zermelo, que se dio cuenta de que no hay inconveniente en aceptar que las propiedades definen *subconjuntos.* Concretamente, Zermelo propuso el axioma siguiente:

Axioma de especificación *Dado un conjunto A y una propiedad P, existe un conjunto cuyos elementos son los elementos de A que cumplen P.*

Nuevamente, sólo puede existir un conjunto con los elementos de A que cumplen P, pues si hubiera dos, ambos tendrían los mismos elementos (los elementos de A que cumplen P), y por el axioma de extensionalidad serían el mismo. Por lo tanto, aceptando el axioma de especificación, dado cualquier conjunto A y cualquier propiedad P, podemos representar mediante

$$\{x \in A \mid Px\}$$

al (único) subconjunto de A formado por los conjuntos que cumplen P.

Notemos que no llegamos a nada absurdo aunque apliquemos este axioma a la propiedad $Px \equiv x \notin x$. No hay ningún inconveniente en considerar el conjunto

$$R = \{x \in A \mid x \notin x\} \subset A$$

cuyos elementos son los elementos de A que no se pertenecen a sí mismos.

Lo que podemos probar ahora es que $R \notin A$, pues si $R \in A$, tendríamos la misma contradicción de antes tanto si $R \in R$ como si $R \notin R$. Y una vez determinado que $R \notin A$, podemos asegurar, más concretamente, que $R \notin R$, ya que si fuera $R \in R \subset A$ entonces, en particular, tendríamos que $R \in A$, y ya hemos visto que eso es imposible.[4]

Aceptando el axioma de especificación nos encontramos con que $x \notin x$ no es la única propiedad que da lugar a contradicciones si intentamos aplicar el axioma de comprensión. Lo mismo le sucede a la propiedad $Px \equiv x = x$. En efecto, si pudiéramos aplicar el axioma de comprensión a esta propiedad, tendríamos el conjunto

$$V = \{x \mid x = x\},$$

es decir, el conjunto de todos los conjuntos que son iguales a sí mismos, y como eso sucede siempre, V sería simplemente el conjunto de todos los conjuntos. Pero si existiera tal conjunto, entonces podríamos definir, por el axioma de especificación, el conjunto

$$R = \{x \in V \mid x \notin x\},$$

y de nuevo tenemos una contradicción, la condición $x \in V$ se cumple siempre y podemos suprimirla, lo que nos da que $R = \{x \mid x \notin x\}$ es el conjunto paradójico de Russell.

En suma, el axioma de especificación implica que no existe ningún conjunto universal V que contenga a todos los conjuntos. Notemos que no hay ninguna contradicción en ello. El problema con el axioma de comprensión es que afirma que existen conjuntos que no pueden existir. En cambio, el axioma de especificación no afirma que exista el conjunto de todos los conjuntos, y no hay ninguna contradicción en que hayamos concluido que, de hecho, no existe.

Éste es un buen punto para reflexionar sobre la "zona gris" de la figura de la página 4. Podemos considerar que el rectángulo representa la totalidad de los conjuntos, pero hay que entender que dicha totalidad es un abismo inabarcable. Los conjuntos nos permiten parcelar cualquier zona de tamaño moderado dentro de esa totalidad, pero fuera de cada conjunto queda siempre un abismo inconmensurable e "imparcelable".

Notemos que el axioma de especificación sólo nos permite definir subconjuntos de un conjunto dado, y no es posible construir todos los conjuntos que necesitan los matemáticos yendo siempre "hacia abajo", sino que necesitamos otros axiomas que nos permitan, a partir de un conjunto, construir otros mayores, que no sean subconjuntos suyos. Hay tres axiomas que cumplen esta función:

[4] Tal vez el lector se pregunte si es posible que un conjunto se pertenezca a sí mismo. La respuesta es que la pregunta es irrelevante: en ningún momento vamos a considerar conjuntos que se pertenezcan a sí mismos, de modo que da igual suponer que no existen o suponer que existen y no tenerlos en cuenta para nada. Si optamos por lo más simple, que es suponer que no existen, entonces $R = \{x \in A \mid x \notin x\}$ es simplemente $R = A$, y la conclusión a la que hemos llegado es a que $A \notin A$.

Axioma del par *Dados dos conjuntos A y B, existe otro conjunto C cuyos elementos son exactamente A y B.*

Axioma de la unión *Dado un conjunto A, existe otro conjunto B cuyos elementos son exactamente los que pertenecen a alguno de los elementos de A.*

Axioma del conjunto de partes *Dado un conjunto A, existe otro conjunto cuyos elementos son exactamente todos los subconjuntos de A.*

Observemos que, en los tres casos, el conjunto cuya existencia se afirma es único, por el axioma de extensionalidad, ya que dos conjuntos en las condiciones indicadas tendrían los mismos elementos. Por lo tanto podemos darles nombre:

El conjunto cuyos únicos elementos son dos conjuntos dados A y B lo representaremos por $\{A, B\}$ y lo llamaremos *par desordenado* de A y B.

El conjunto cuyos elementos son todos los elementos de alguno de los elementos de un conjunto dado A lo representaremos por $\bigcup A$ y lo llamaremos *gran unión* de A.

El conjunto cuyos elementos son todos los subconjuntos de un conjunto dado A lo representaremos por $\mathcal{P}A$ y lo llamaremos conjunto de las *partes* de A.

Por ejemplo, si un conjunto X consta de los tres conjuntos A, B y C representados en la página 4, entonces $\bigcup X$ se corresponde con la región de color blanco de la figura, es decir, la región que incluye tanto los puntos de A, como los de B como los de C.

Es importante que en el axioma del par no exigimos que los dos conjuntos A y B sean distintos. En caso de que sean el mismo, escribiremos $\{A\} = \{A, A\}$, que es el conjunto cuyo único elemento es A.

Con estos axiomas ya podemos introducir algunas construcciones conjuntistas básicas:

- Dados dos conjuntos A y B, definimos su *unión* como el conjunto

$$A \cup B = \bigcup\{A, B\},$$

  cuyos elementos son los elementos que están ya sea en A o en B. Más precisamente, $x \in A \cup B$ es equivalente a que se cumpla $x \in A$ o $x \in B$.

- Dados dos conjuntos A y B, definimos su *intersección* como el conjunto

$$A \cap B = \{x \in A \mid x \in B\},$$

  que resulta de aplicar el axioma de especificación al conjunto A y la propiedad $Px \equiv x \in B$. Así, los elementos de $A \cap B$ son los que están a la vez en A y en B. Por lo tanto, $x \in A \cap B$ es equivalente a que se cumpla $x \in A$ y $x \in B$.

- Dados dos conjuntos A y B, definimos su *complemento* como el conjunto

$$A \setminus B = \{x \in A \mid x \notin B\},$$

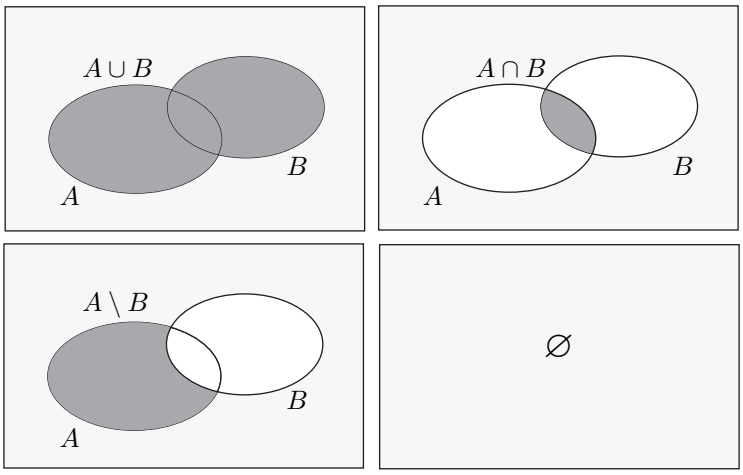
de modo que $x \in A \setminus B$ es equivalente a $x \in A$ y $x \notin B$.

- Dado cualquier conjunto A, definimos el *conjunto vacío* como

$$\varnothing = \{x \in A \mid x \neq x\},$$

que obviamente es un conjunto sin elementos y no depende del conjunto A a partir del cual lo calculamos, pues dos conjuntos sin elementos tienen los mismos elementos (ninguno), y por el axioma de extensionalidad son el mismo conjunto.

La figura siguiente muestra la representación (en gris oscuro) de la unión, la intersección y el complemento de dos conjuntos A y B, así como la "no representación" del conjunto vacío.



Es importante tener presente que, para todo conjunto X, se cumple la relación $\varnothing \subset X$, es decir, el conjunto vacío es un subconjunto de cualquier conjunto. Esto es "pura lógica": en general, para que suceda $A \not\subset B$ es necesario que exista un elemento de A que no sea elemento de B, luego, para que pudiera suceder $\varnothing \not\subset X$, tendría que existir un elemento $x \in \varnothing$ tal que $x \notin X$, pero como $\varnothing$ no tiene elementos, esto es imposible.

Se dice que dos conjuntos A y B son *disjuntos* si $A \cap B = \varnothing$, es decir, si no tienen elementos en común. Por ejemplo, en la figura de la página 4 se tiene que A y C son disjuntos, mientras que B y C no lo son.

Es posible demostrar muchas propiedades elementales sobre uniones, intersecciones y complementos. Veamos una muestra:

Teorema 1.2 (Leyes de De Morgan) *Si A, B son dos subconjuntos de un conjunto X, entonces*

$$X \setminus (A \cup B) = (X \setminus A) \cap (X \setminus B), \quad X \setminus (A \cap B) = (X \setminus A) \cup (X \setminus B).$$

DEMOSTRACIÓN: Veamos la primera igualdad (la segunda se demuestra análogamente). Para probar una igualdad de conjuntos demostramos la doble inclusión. Para probar $\subset$ tomamos un $x \in X \setminus (A \cup B)$. Esto significa que $x \in X$ y $x \notin A \cup B$. Como la unión contiene los elementos que están ya sea en A como en B, para que $x \notin A \cup B$ tiene que suceder que $x \notin A$ y $x \notin B$. Entonces, como $x \in X$ y $x \notin A$, resulta que $x \in X \setminus A$, e igualmente $x \in X \setminus B$. Por definición de intersección $x \in (X \setminus A) \cap (X \setminus B)$.

Ahora probamos la inclusión opuesta: tomamos $x \in (X \setminus A) \cap (X \setminus B)$. Esto significa que $x \in X \setminus A$ y $x \in X \setminus B$. De la primera parte deducimos que $x \in X$ y $x \notin A$, y de la segunda además que $x \notin B$. Como $x \notin A$ y $x \notin B$, podemos concluir que $x \notin A \cup B$, luego $x \in X \setminus (A \cup B)$. ∎

Hemos definido la gran unión $\bigcup X$ como la unión de todos los conjuntos que forman parte de X. Igualmente podemos definir una gran intersección:

Teorema 1.3 *Si $X \neq \varnothing$, existe un único conjunto Y cuyos elementos son los que pertenecen a todos los elementos de X.*

DEMOSTRACIÓN: Como X no es el conjunto vacío, existe un $A \in X$. Ahora consideramos la propiedad $Px \equiv x$ pertenece a todos los elementos de X, que es una propiedad a la que podemos aplicar el axioma de especificación, para formar el conjunto

$$Y = \{x \in A \mid Px\}.$$

El conjunto Y cumple lo pedido, pues si x pertenece a todos los elementos de X, entonces cumple Px y, en particular, $x \in A$, luego $x \in Y$. Recíprocamente, si $x \in Y$ entonces cumple Px, luego x pertenece a todos los elementos de X. El conjunto Y es único por el axioma de extensionalidad, ya que dos conjuntos que cumplieran lo requerido tendrían los mismos elementos (los que pertenecen a todos los elementos de X), luego serían el mismo conjunto. ∎

Al conjunto dado por el teorema anterior lo llamaremos *gran intersección* de X y lo representaremos por $\bigcap X$. Es importante tener en cuenta que la gran intersección $\bigcap \varnothing$ no está definida. El lector familiarizado con las sutilezas de la lógica puede plantearse quién debería ser dicha intersección en caso de existir.

El sexto y último axioma que vamos a considerar en esta sección (y el penúltimo que vamos a necesitar) es el siguiente:

Axioma de infinitud *Existe un conjunto X con la propiedad de que $\varnothing \in X$ y, siempre que $x \in X$, se cumple también que $\{x\} \in X$.*

Vamos a analizar el contenido de este axioma. Para ello llamaremos conjuntos *inductivos* a los conjuntos X que cumplen lo que afirma el axioma de

infinitud. A diferencia de lo que sucedía con los demás axiomas sobre existencia de conjuntos, no podemos concluir por el axioma de extensionalidad que exista un único conjunto inductivo, pero podemos demostrar lo siguiente:

Teorema 1.4 *Existe un único conjunto inductivo X con la propiedad de que está contenido en cualquier otro conjunto inductivo.*

DEMOSTRACIÓN: Sea X_0 un conjunto inductivo cualquiera. Sabemos que existe por el axioma de infinitud. Sea $\mathcal{I} = \{A \in \mathcal{P}X_0 \mid A \text{ es inductivo}\}$, que es un conjunto bien definido por el axioma de especificación. Sus elementos son todos los subconjuntos inductivos de X_0. En particular $\mathcal{I} \neq \varnothing$, pues $X_0 \in \mathcal{I}$.

Esto nos permite formar la gran intersección $X = \bigcap \mathcal{I}$. Vamos a probar que el conjunto X cumple lo requerido. En primer lugar probamos que X es inductivo. Para ello, en primer lugar debemos ver que $\varnothing \in X$. Por definición de gran intersección, esto equivale a que $\varnothing$ pertenece a todos los elementos de $\mathcal{I}$, y eso es cierto, porque los elementos de $\mathcal{I}$ son conjuntos inductivos, y por definición todos tienen a $\varnothing$ por elemento.

En segundo lugar tomamos $x \in X$ y tenemos que probar que $\{x\} \in X$. Para ello tenemos que probar que $\{x\}$ pertenece a todos los elementos de $\mathcal{I}$. Tomemos uno cualquiera, digamos $A \in \mathcal{I}$. Como $x \in X$, sabemos que x pertenece a todos los elementos de $\mathcal{I}$, y en particular $x \in A$. Como $A \in \mathcal{I}$, resulta que A es inductivo, luego por definición de conjunto inductivo, si $x \in A$, también tiene que cumplirse que $\{x\} \in A$. Con esto hemos probado que $\{x\}$ está en cualquier elemento de $\mathcal{I}$ prefijado, luego $\{x\} \in X$.

Ahora consideramos cualquier conjunto inductivo Y y tenemos que probar que $X \subset Y$. Para ello empezamos demostrando que $X_0 \cap Y$ es un conjunto inductivo. En efecto, como X_0 e Y son ambos inductivos, tenemos que $\varnothing \in X_0$ y $\varnothing \in Y$, luego $\varnothing \in X_0 \cap Y$. Por otra parte, si $x \in X_0 \cap Y$, tenemos que $x \in X_0$ y $x \in Y$, y al ser ambos inductivos $\{x\} \in X_0$ y $\{x\} \in Y$, luego $\{x\} \in X_0 \cap Y$.

Así pues, hemos probado que $X_0 \cap Y$ es inductivo y claramente $X_0 \cap Y \subset X_0$, luego $X_0 \cap Y \in \mathcal{P}X_0$. Esto equivale a que $X_0 \cap Y \in \mathcal{I}$, pues es un subconjunto inductivo de X_0. Por último, si $x \in X$, entonces x está en todos los elementos de $\mathcal{I}$, luego $x \in X_0 \cap Y$ luego $x \in Y$. Esto prueba que $X \subset Y$. ∎

Definición 1.5 Llamaremos *conjunto de los números naturales* al menor conjunto inductivo (en el sentido del teorema anterior) y lo representaremos por $\mathbb{N}$.

Observemos que $\varnothing \in \mathbb{N}$, por definición de conjunto inductivo. Cuando pensemos en el conjunto vacío como número natural lo representaremos por 0 y lo llamaremos *cero,* si bien técnicamente 0 y $\varnothing$ son dos nombres alternativos para el mismo conjunto.

Si $n \in \mathbb{N}$, por definición de conjunto inductivo tenemos que $\{n\} \in \mathbb{N}$. Cuando pensemos en $\{n\}$ como número natural escribiremos Sn en lugar de $\{n\}$, y diremos que Sn es *el siguiente* de n.

Al siguiente de 0 lo llamaremos *uno* y lo representaremos por $1 = S0 = \{\varnothing\}$, al siguiente de 1 lo llamaremos *dos* y lo representaremos por

$$2 = S1 = \{1\} = \{\{\varnothing\}\},$$

igualmente definimos el *tres,* el *cuatro,* el *cinco,* el *seis,* el *siete,* el *ocho* y el *nueve* como:

$$\begin{array}{ll} 3 = S2 = \{\{\{\varnothing\}\}\} & 7 = S6 = \{\{\{\{\{\{\{\varnothing\}\}\}\}\}\}\} \\ 4 = S3 = \{\{\{\{\varnothing\}\}\}\} & 8 = S7 = \{\{\{\{\{\{\{\{\varnothing\}\}\}\}\}\}\}\} \\ 5 = S4 = \{\{\{\{\{\varnothing\}\}\}\}\} & 9 = S8 = \{\{\{\{\{\{\{\{\{\varnothing\}\}\}\}\}\}\}\}\}. \\ 6 = S5 = \{\{\{\{\{\{\varnothing\}\}\}\}\}\} & \end{array}$$

Es claro que 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 son números naturales. En general:

Teorema 1.6 (Axiomas de Peano) *El conjunto* $\mathbb{N}$ *de los números naturales cumple las propiedades siguientes:*

1. $0 \in \mathbb{N}$ *(el cero es un número natural).*
2. *Si* $n \in \mathbb{N}$*, entonces* $Sn \in \mathbb{N}$ *(el siguiente de un número natural es un número natural).*
3. 0 *no es el siguiente de ningún número natural.*
4. *Si* $m, n \in \mathbb{N}$ *cumplen* $Sm = Sn$*, entonces* $m = n$ *(si dos números naturales tienen el mismo siguiente, es que son el mismo).*
5. *Si* $A \subset \mathbb{N}$ *tiene la propiedad de que* $0 \in A$ *y siempre que* $n \in A$ *también* $Sn \in A$*, entonces* $A = \mathbb{N}$ *(principio de inducción).*

Demostración: 1) El cero es un número natural porque $\mathbb{N}$ es un conjunto inductivo y $0 = \varnothing \in \mathbb{N}$ por definición de conjunto inductivo.

2) Si $n \in \mathbb{N}$, entonces $Sn = \{n\} \in \mathbb{N}$, también por definición de conjunto inductivo.

3) No puede suceder que $0 = Sn$, porque entonces $\varnothing = \{n\}$, cuando el conjunto de la derecha no es vacío, ya que contiene a n.

4) Si $Sm = Sn$, esto es lo mismo que $\{m\} = \{n\}$, y entonces $m \in \{m\} = \{n\}$, luego $m = n$.

5) La hipótesis es que A es un conjunto inductivo, y entonces $\mathbb{N} \subset A$ porque $\mathbb{N}$ es el menor conjunto inductivo (está contenido en cualquier otro, según el teorema 1.4). Como por hipótesis $A \subset \mathbb{N}$, tenemos que $A = \mathbb{N}$. ∎

Observaciones Ahora podemos entender plenamente el axioma de infinitud: hemos elegido una operación conjuntista, concretamente $x \mapsto \{x\}$, que cuando la vamos aplicando sucesivamente al conjunto $\varnothing$ va produciendo conjuntos distintos de los precedentes: $\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\{\{\varnothing\}\}\}, \ldots$ y hemos postulado que existe un conjunto que contiene a todos los conjuntos que se obtienen de este modo. Obviamente dicho conjunto no es único, pues puede contener elementos "extra" no deseados, pero siempre podemos quedarnos con el menor de todos ellos, $\mathbb{N}$, que es el menor conjunto que puede obtenerse a partir de $0 = \varnothing$ aplicando la operación "siguiente", sin ningún añadido. A los conjuntos que vamos obteniendo al aplicar la operación "siguiente" los llamamos números naturales $0, 1, 2, 3, \ldots$

Abusando un poco del lenguaje, podemos expresar que $\mathbb{N}$ es el menor conjunto inductivo mediante $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$, donde los puntos suspensivos indican que los elementos de $\mathbb{N}$ son únicamente los que resultan de calcular siguientes y más siguientes, pero la versión formal de este hecho es el teorema anterior.

A partir de aquí, el hecho de que hayamos definido concretamente el número tres como $3 = \{\{\{\varnothing\}\}\}$ se vuelve completamente irrelevante. Lo único que importa de los números naturales es que cumplen los axiomas de Peano. Qué conjuntos hemos seleccionado concretamente para que esto suceda no tiene importancia. Ninguno de los resultados que vamos a demostrar sobre los números naturales dependerán de cuáles sean concretamente los elementos del 3 o del 7.

Aquí tenemos un primer ejemplo de un concepto matemático (el de los números naturales) que nadie concibe como conjuntos, sino más bien como "elementos puros", pero esto no contradice que técnicamente podamos definir los números naturales de modo que sean conjuntos (y de hecho conjuntos puros).

Podríamos haber elegido otra operación "siguiente". Por ejemplo, si hubiéramos elegido $x \mapsto x \cup \{x\}$ entonces los números naturales hubieran sido

$$0 = \varnothing, \quad 1 = \{0\}, \quad 2 = \{0, 1\}, \quad 3 = \{0, 1, 2\}, \quad 4 = \{0, 1, 2, 3\}, \quad \ldots$$

Esta alternativa es mucho más conveniente para quien pretenda adentrarse en la teoría de conjuntos, pero para nuestros fines es irrelevante una u otra opción y la que hemos elegido simplifica la demostración de los axiomas de Peano. Más adelante volveremos sobre la comparación de conjuntos de números naturales obtenidos a partir de elecciones diferentes de la función "siguiente" o incluso del 0 (que tampoco es obligatorio que sea el conjunto vacío). ∎

Conviene destacar el significado del quinto axioma de Peano, el principio de inducción. Para ello consideramos cualquier propiedad P y lo aplicamos al conjunto $A = \{n \in \mathbb{N} \mid Pn\}$. Su traducción a este caso es la siguiente:

Principio de inducción *Si probamos que* 0 *tiene la propiedad* P *y, bajo la hipótesis de que* $n \in \mathbb{N}$ *cumple la propiedad* P *(hipótesis de inducción), podemos demostrar que* Sn *también cumple la propiedad* P*, entonces podemos asegurar que todo número natural cumple la propiedad* P*.*

Veamos un ejemplo sencillo de demostración por inducción:

Teorema 1.7 *Todo número natural distinto de* 0 *es el siguiente de otro número natural.*

Demostración: Tomamos como propiedad P la propiedad $Pn \equiv$ "si $n \neq 0$ entonces n es el siguiente de otro número natural". Y vamos a probar por inducción que la cumplen todos los números naturales. Se cumple $P0$, pues no se da el caso de que $0 \neq 0$. Si suponemos por hipótesis de inducción que se cumple Pn, entonces se cumple PSn, simplemente porque Sn es el siguiente de un número natural. Por lo tanto todos los números tienen la propiedad P. ∎

Para extraer más consecuencias de los axiomas de Peano necesitamos algunos conceptos conjuntistas adicionales.

1.3 Funciones

El concepto de función es una de las piezas fundamentales del vocabulario conjuntista. Para introducirlo necesitamos otros conceptos previos.

Tenemos definidos los pares desordenados $\{a, b\}$, y es claro que de una igualdad $\{a, b\} = \{c, d\}$ no podemos concluir necesariamente que $a = c$ y $b = d$, pues $\{a, b\} = \{b, a\}$, y perfectamente podría suceder que $a = d \neq b = c$.

Esto hace conveniente definir el *par ordenado* con *primera componente* a y *segunda componente* b como $(a, b) = \{\{a\}, \{a, b\}\}$. Entonces:

Teorema 1.8 *Si a, b, c, d son conjuntos cualesquiera, entonces*

$$(a, b) = (c, d) \text{ si y sólo si } a = c \text{ y } b = d.$$

DEMOSTRACIÓN: Es inmediato que si $a = c$ y $b = d$ entonces $(a, b) = (c, d)$. Vamos a probar la implicación opuesta. Para ello suponemos primero que $a = b$. Entonces, por definición de par ordenado, $(a, b) = \{\{a\}, \{a, a\}\} = \{\{a\}\}$ y $(c, d) = \{\{c\}, \{c, d\}\} = \{\{a\}\}$. Esto obliga a que $\{c\} = \{c, d\} = \{a\}$, de donde se sigue claramente que $c = d = a = b$.

Llegamos a la misma conclusión si $c = d$, por lo que podemos suponer que $a \neq b$ y $c \neq d$. Tenemos que $\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}$. Entonces, como $\{a\} \in \{\{c\}, \{c, d\}\}$, tiene que ser $\{a\} = \{c\}$ o bien $\{a\} = \{c, d\}$, pero el segundo caso es imposible, porque entonces $c = a = d$. Por lo tanto, $\{a\} = \{c\}$ y esto implica que $a = c$.

Igualmente, $\{a, b\} \in \{\{c\}, \{c, d\}\}$, y no puede ser $\{a, b\} = \{c\}$, porque entonces $a = c = b$, luego tiene que ser $\{a, b\} = \{c, d\}$, luego $b \in \{c, d\}$, luego $b = c$ o bien $b = d$, pero no puede ser $b = c = a$, luego $b = d$, como había que probar. ∎

Observemos ahora que si $a \in A$ y $b \in B$, entonces $a, b \in A \cup B$, luego $\{a\}, \{a, b\} \subset A \cup B$, luego $\{a\}, \{a, b\} \in \mathcal{P}(A \cup B)$, luego

$$(a, b) = \{\{a\}, \{a, b\}\} \subset \mathcal{P}(A \cup B),$$

luego $(a, b) \in \mathcal{PP}(A \cup B)$. Esto nos permite probar:

Teorema 1.9 *Dados dos conjuntos A y B, existe un único conjunto cuyos elementos son todos los pares ordenados (a, b) con $a \in A$ y $b \in B$.*

DEMOSTRACIÓN: Basta aplicar el axioma de especificación a la propiedad $Px \equiv x$ es un par ordenado de la forma $x = (a, b)$, con $a \in A$ y $b \in B$ y definir

$$X = \{x \in \mathcal{PP}(A \cup B) \mid Px\}.$$

Así, todo elemento de X es un par ordenado en las condiciones requeridas, y si $x = (a, b)$ es cualquiera de dichos pares, justo antes del enunciado de este teorema hemos probado que $x \in \mathcal{PP}(A \cup B)$ y además Px, luego $x \in X$. La unicidad es por el axioma de extensionalidad, ya que dos conjuntos que cumplieran lo requerido tendrían los mismos elementos (los pares ordenados con primera componente en A y segunda componente en B), luego serían el mismo conjunto. ∎

Llamaremos *producto cartesiano* de los conjuntos A y B al conjunto dado por el teorema anterior, y que representaremos por $A \times B$. Así, si $a \in A$ y $b \in B$, tenemos que $(a, b) \in A \times B$, y todos los elementos de $A \times B$ son de esta forma.

A partir de este punto podemos olvidarnos de que los pares ordenados (a, b) son conjuntos. Nunca más necesitaremos pensar en ellos como conjuntos, sino que lo único que importa es que determinan el orden de sus componentes y que pueden "recolectarse" en los productos cartesianos.

Definición 1.10 Diremos que un conjunto f es una *aplicación* o *función* de un conjunto A en un conjunto B (y lo representaremos por $f : A \longrightarrow B$) si $f \subset A \times B$ y para cada $a \in A$ existe un único $b \in B$ tal que $(a, b) \in f$. Dicho b recibe el nombre de *imagen* de a por f y se representa por $f(a)$. También se dice que a es una *antiimagen* de b por f. El conjunto A sobre el que está definida f se llama *dominio* de f y lo representaremos por $\mathcal{D}f$.

Diremos que $f : A \longrightarrow B$ es *inyectiva* si cuando dos elementos x, $y \in A$ cumplen $f(x) = f(y)$, entonces $x = y$ (elementos distintos tienen imágenes distintas).

Diremos que $f : A \longrightarrow B$ es *suprayectiva* si para todo $b \in B$ existe un $a \in A$ tal que $f(a) = b$ (todo elemento de B tiene al menos una antiimagen en A).

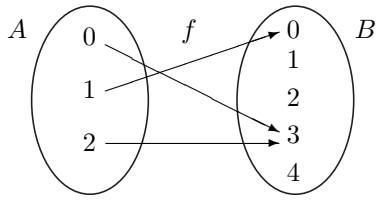
Diremos que $f : A \longrightarrow B$ es *biyectiva* si es inyectiva y suprayectiva.

Notemos que la suprayectividad y, por consiguiente, la biyectividad de una aplicación no depende únicamente de f, sino también del conjunto B en el que consideramos que toma sus imágenes. Una misma función puede ser o no suprayectiva según el conjunto B considerado.

Ejemplos Sean[5] $A = \{0, 1, 2\}$ y $B = \{0, 1, 2, 3, 4\}$ y sea

$$f = \{(0, 3), (1, 0), (2, 3)\}.$$

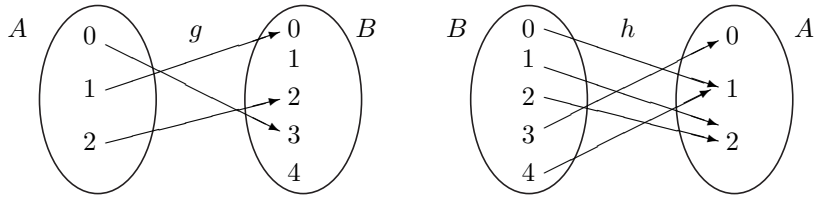
Entonces f es una aplicación $f : A \longrightarrow B$. Ciertamente es un subconjunto de $A \times B$ (sus elementos son pares con primera componente 0, 1 o 2 y segunda componente 0, 1, 2, 3 o 4) y cada elemento de A tiene una única imagen por f en B, a saber, $f(0) = 3$, $f(1) = 0$, $f(2) = 3$. Podemos representar f mediante un diagrama de flechas, así:



[5]Por definición $\{0, 1, 2\} = \{0\} \cup \{1\} \cup \{2\}$ es el conjunto cuyos elementos son 0, 1 y 2, y análogamente con $\{0, 1, 2, 3, 4\}$.

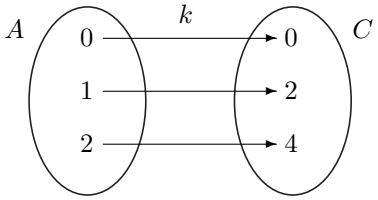
Vemos que f no es inyectiva, porque hay dos elementos distintos de A (el 0 y el 2) que tienen la misma imagen en B: se cumple $f(0) = f(2)$, pero $0 \neq 2$. También vemos f no es suprayectiva, porque hay varios elementos en B que no tienen antiimagen en A (el 1, el 2 y el 4).

La figura siguiente muestra una aplicación inyectiva $g : A \longrightarrow B$ y otra suprayectiva de $h : B \longrightarrow A$:



Vemos que, en efecto, la aplicación g no asigna una misma imagen a dos elementos distintos, por lo que es inyectiva, pero no suprayectiva, ya que hay elementos sin antiimagen. En cambio, la de la derecha no es inyectiva, pues $h(0) = h(4)$ y también $h(1) = h(2)$, pero sí que es suprayectiva, ya que todo elemento de B tiene al menos una antiimagen.

La figura siguiente muestra una aplicación biyectiva entre dos conjuntos A y C:



Notemos que éste es el aspecto "típico" de una aplicación biyectiva entre dos conjuntos: cada elemento de A tiene una única imagen en C y cada elemento de C tiene una única antiimagen en A (una porque la aplicación es suprayectiva y única porque es inyectiva). De esta suerte, los elementos de A quedan emparejados con los elementos de C sin que falte ni sobre ninguno. ∎

En la práctica, para definir una aplicación $f : A \longrightarrow B$ basta especificar (una vez dados A y B) cuál es la imagen $f(a) \in B$ de cada elemento $a \in A$.

Por ejemplo, podemos definir $S : \mathbb{N} \longrightarrow \mathbb{N}$ sin más que especificar que si $n \in \mathbb{N}$, entonces $S(n) = \{n\}$. Toda definición hecha de este modo puede traducirse a una aplicación del axioma de especificación. Concretamente, en este caso sería

$$S = \{x \in \mathbb{N} \times \mathbb{N} \mid \text{existe un } n \in \mathbb{N} \text{ tal que } x = (n, S(n))\}.$$

Igualmente, para probar que dos aplicaciones $f : A \longrightarrow B$ y $g : A \longrightarrow C$ son iguales (como conjuntos) no usaremos normalmente una doble inclusión que

nos obligue a "recordar" que f y g son conjuntos de pares ordenados. Eso es equivalente a probar (teniendo en cuenta que estamos suponiendo de antemano que f y g tienen el mismo dominio A) que para todo $a \in A$ se cumple que $f(a) = g(a)$. En otras palabras:

> *Dos aplicaciones son iguales si y sólo si tienen el mismo dominio y asignan la misma imagen a cada elemento de dicho dominio.*

En efecto, si se cumple esto podemos probar que $f = g$ por doble inclusión: si $x \in f$, entonces x es un par ordenado $x = (a, f(a))$, pero como $f(a) = g(a)$, de hecho $x = (a, g(a)) \in g$. Esto prueba que $f \subset g$ e igualmente se prueba que $g \subset f$.

Introducimos ahora algunos conceptos adicionales sobre una aplicación cualquiera $f : A \longrightarrow B$ (los ejemplos que las acompañan hacen referencia a las funciones g y h correspondientes a los gráficos de la página precedente):

1. Si $X \subset A$, definimos $f[X] = \{b \in B \mid \text{existe } x \in X \text{ tal que } f(x) = b\}$. Así, $f[X]$ es el conjunto de las imágenes de los elementos de X.

   Por ejemplo, $h[\{1, 2, 3\}] = \{0, 2\}$.

2. En particular $f[A]$ es el conjunto de todas las imágenes de los elementos de A y se denomina *rango* o *imagen* de f. Lo representaremos por $\mathcal{R}f$ o por $\operatorname{Im} f$. Notemos que f es suprayectiva si y sólo si $f[A] = B$.

   Por ejemplo, $\mathcal{R}g = \{0, 2, 3\}$.

3. Si $Y \subset B$, definimos $f^{-1}[Y] = \{a \in A \mid f(a) \in Y\}$. Se trata del conjunto de todas las antiimágenes de los elementos de Y. (Es habitual escribir $f^{-1}[b] = f^{-1}[\{b\}]$ para hacer referencia al conjunto de las antiimágenes de $b \in B$.)

4. Si $X \subset A$ definimos $f|_X : X \longrightarrow B$ como la aplicación que coincide con f sobre los elementos de X, es decir, su dominio es X y, para cada $x \in X$, se define $f|_X(x) = f(x)$.

5. Si f es inyectiva y $C = f[A]$, definimos $f^{-1} : C \longrightarrow A$ como la única aplicación que a cada elemento de C le asigna su única antiimagen por f, es decir, si $c \in C$, entonces $f^{-1}(c)$ es el único elemento $a \in A$ tal que $f(a) = c$. Así $f(a) = b$ es equivalente a $f^{-1}(b) = a$, pero para que esto tenga sentido es necesario que f sea inyectiva, pues si b tuviera varias antiimágenes, no estaría bien definido $f^{-1}(b)$.

6. En particular, si $f : A \longrightarrow B$ biyectiva, también $f^{-1} : B \longrightarrow A$ biyectiva.

Si $A \subset B$, la aplicación $i : A \longrightarrow B$ dada por $i(a) = a$ se llama *inclusión* de A en B. Si, más concretamente, $A = B$, la inclusión $i : A \longrightarrow A$ se llama *aplicación identidad* en A.

Si $f : A \longrightarrow B$, $g : C \longrightarrow D$ y $f[A] \subset C$, podemos definir la *composición* $f \circ g : A \longrightarrow D$ como la aplicación dada por $(f \circ g)(a) = g(f(a))$, es decir, la aplicación que se calcula pasando de $a \in A$ a $f(a) \in f[A] \subset C$ y luego de $f(a)$ a $g(f(a)) \in D$.

Por ejemplo, si g y h son las aplicaciones de las figuras de la página anterior, se cumple que $g \circ h$ es la identidad en A, pues $(g \circ h)(0) = h(g(0)) = h(3) = 0$, e igualmente se comprueba que $(g \circ h)(1) = 1$ y $(g \circ h)(2) = 2$.

Los conceptos que acabamos de introducir satisfacen numerosas relaciones de interés que conviene conocer. Veamos algunas de ellas:

1. Si $f : A \longrightarrow B$, $g : B \longrightarrow C$ y $h : C \longrightarrow D$, entonces $(f \circ g) \circ h = f \circ (g \circ h)$.

   En efecto, los dos miembros de la igualdad son aplicaciones $A \longrightarrow D$ y, para cada $a \in A$, se cumple que

   $$((f \circ g) \circ h)(a) = h((f \circ g)(a)) = h(g(f(a))) = (g \circ h)(f(a)) = (f \circ (g \circ h))(a).$$

   Por lo tanto, en esta situación podemos escribir $f \circ g \circ h$ sin necesidad de poner paréntesis. Se trata de la aplicación que actúa primero como f, luego aplica g y luego h.

2. Si $f : A \longrightarrow B$ es biyectiva, entonces $f \circ f^{-1} = i_A$ es la identidad en A y $f^{-1} \circ f = i_B$ es la identidad en B.

   En efecto, si $a \in A$ y $f(a) = b$, entonces $f^{-1}(b) = a$, luego

   $$(f \circ f^{-1})(a) = f^{-1}(f(a)) = f^{-1}(b) = a = i_A(a),$$

   e igualmente se comprueba que $(f^{-1} \circ f)(b) = b = i_B(b)$, para todo $b \in B$.

3. Si $f : A \longrightarrow B$ e i_A, i_B son las identidades en A y en B, respectivamente, entonces $i_A \circ f = f = f \circ i_B$.

4. Si $f : A \longrightarrow B$ y X, $Y \subset B$, entonces

   $$f^{-1}[X \cup Y] = f^{-1}[X] \cup f^{-1}[Y], \quad f^{-1}[X \cap Y] = f^{-1}[X] \cap f^{-1}[Y],$$

   $$f^{-1}[X \setminus Y] = f^{-1}[X] \setminus f^{-1}[Y].$$

   Comprobamos la primera como ejemplo: si $a \in f^{-1}[X \cup Y]$, entonces $f(a) \in X \cup Y$, luego $f(a) \in X$ o bien $f(a) \in Y$, luego $a \in f^{-1}[X]$ o bien $a \in f^{-1}[Y]$, luego $a \in f^{-1}[X] \cup f^{-1}[Y]$. Esto nos da la inclusión $\subset$, la inclusión $\supset$ se demuestra análogamente.

5. Si $f : A \longrightarrow B$ y $g : B \longrightarrow A$ y $f \circ g = i_A$ es la identidad en A, entonces f es inyectiva y g es suprayectiva.

   En efecto, para ver que f es inyectiva tomamos x, $y \in A$ y suponemos que $f(x) = f(y)$. Entonces $g(f(x)) = g(f(y))$, luego $(f \circ g)(x) = (f \circ g)(y)$, luego $x = y$.

   Para ver que g es suprayectiva tomamos $a \in A$ y observamos que se cumple $g(f(a)) = a$, luego $f(a) \in B$ es una antiimagen de a por g.

Ejercicio: Sea $f : A \longrightarrow B$ y sean $X, Y \subset A$. Estudiar si se cumplen las igualdades

$$f[X \cup Y] = f[X] \cup f[Y], \qquad f[X \cap Y] = f[X] \cap f[Y], \qquad f[X \setminus Y] = f[X] \setminus f[Y].$$

Por último observamos que si $f : A \longrightarrow B$, por definición $f \subset A \times B$, luego $f \in \mathcal{P}(A \times B)$. Por lo tanto, podemos usar el axioma de especificación y definir

$$B^A = \{f \in \mathcal{P}(A \times B) \mid f : A \longrightarrow B\},$$

que es el conjunto de todas las aplicaciones de A en B, ya que la restricción $f \in \mathcal{P}(A \times B)$ no elimina a ninguna de ellas.

1.4 Los números naturales

Para definir los números naturales hemos elegido arbitrariamente un conjunto como 0 (concretamente $\varnothing$) y una operación conjuntista adecuada como "siguiente" (concretamente $x \mapsto \{x\}$). Veamos ahora que hay muchas otras elecciones posibles. En lugar de la versión del axioma de infinitud que hemos adoptado, podríamos haber tomado esta variante más abstracta:

Axioma de infinitud (variante) *Existe un conjunto X con una aplicación $S : X \longrightarrow X$ inyectiva y no suprayectiva.*

Observemos que la versión anterior del axioma implica ésta, ya que la aplicación sucesor $S : \mathbb{N} \longrightarrow \mathbb{N}$ es inyectiva y no suprayectiva (eso es precisamente lo que afirman los axiomas de Peano 3 y 4). Ahora vamos a ver que esta versión abstracta es suficiente para construir un conjunto de números naturales.

Sea $S : X \longrightarrow X$ una aplicación inyectiva y no suprayectiva, tal y como postula el axioma. Entonces podemos elegir un elemento $0 \in X$ que no tiene antiimagen. Diremos que $A \subset X$ es *inductivo* si $0 \in A$ y cuando $n \in A$ entonces también $S(n) \in A$. Sea $\mathcal{I} = \{A \in \mathcal{P}X \mid A \text{ es inductivo}\}$. Observemos que $\mathcal{I} \neq \varnothing$, ya que $X \in \mathcal{I}$. Sea $N = \bigcap \mathcal{I}$. Exactamente igual que en la demostración del teorema 1.4 podemos probar que N es inductivo. En particular, si $n \in N$, se cumple que $S(n) \in N$, por lo que podemos restringir $S|_N : N \longrightarrow N$. Vamos a llamar S a esta restricción. Entonces:

Teorema 1.11 (Axiomas de Peano) *Existe un conjunto N, una aplicación $S : N \longrightarrow N$ y un elemento 0 de modo que se cumplen las propiedades siguientes:*

1. *$0 \in N$.*
2. *Si $n \in N$, entonces $S(n) \in N$.*
3. *No existe ningún $n \in N$ tal que $S(n) = 0$.*
4. *Si $m, n \in N$ y $S(m) = S(n)$, entonces $m = n$.*
5. *Si $A \subset N$ tiene la propiedad de que $0 \in A$ y siempre que $n \in A$ también $S(n) \in A$, entonces $A = N$.*

DEMOSTRACIÓN: Las propiedades 1) y 2) se cumplen porque N es inductivo, la propiedad 3) porque hemos elegido 0 sin antiimagen por S, la propiedad 4) porque S es inyectiva y la 5) porque lo que afirma es que $A \subset N \subset X$ es inductivo, luego $A \in \mathcal{I}$, luego $N \subset A$, ya que los elementos de N están en todos los elementos de $\mathcal{I}$. ∎

Podemos definir igualmente $1 = S(0)$, $2 = S(1)$, etc. Así tenemos la misma situación que antes, salvo que ahora no hemos precisado cuáles son concretamente los elementos de N ni quién es concretamente el cero ni en qué consiste concretamente el paso al siguiente número natural. Conviene recoger esto en una definición:

Definición 1.12 Un *sistema de Peano* es una terna[6] $(N, S, 0)$ que cumple las cinco propiedades del teorema anterior.

Vemos entonces que el axioma de infinitud que hemos dado en esta sección es equivalente a la existencia de una terna de Peano (pues hemos probado que existe una a partir del axioma y, recíprocamente, la función S de una terna de Peano es una aplicación inyectiva y no suprayectiva). El axioma de infinitud dado en la sección 1.2 implica la existencia de un sistema de Peano en el que, concretamente, $0 = \varnothing$ y $S(n) = \{n\}$, pero esto es anecdótico.

Para poner de manifiesto que lo único que importa realmente de los números naturales es que forman un sistema de Peano ahora vamos a trabajar con uno cualquiera de ellos. Empezamos demostrando la propiedad más importante de los números naturales junto con el principio de inducción:

Teorema 1.13 (Principio de recursión) *Sea $(N, S, 0)$ un sistema de Peano, sea $g : A \longrightarrow A$ una aplicación arbitraria y sea $a \in A$. Entonces existe una única aplicación $f : N \longrightarrow A$ tal que $f(0) = a$ y para todo $n \in N$ se cumple que $f(S(n)) = g(f(n))$.*

La última propiedad se expresa a menudo diciendo que el diagrama siguiente es conmutativo:

$$\begin{array}{ccc} A & \xrightarrow{g} & A \\ {\scriptstyle f}\uparrow & & \uparrow{\scriptstyle f} \\ N & \xrightarrow[S]{} & N \end{array}$$

Decir que un diagrama de este tipo es conmutativo quiere decir que si vamos de un punto a otro por dos caminos diferentes el resultado es el mismo, en este caso $f \circ g = S \circ f$, que es justo lo que afirma el teorema.

En la práctica, el principio de recursión afirma que para definir una aplicación $f : N \longrightarrow A$ no es necesario definir explícitamente $f(0)$, $f(1)$, $f(2)$, etc., sino que basta definir $f(0)$ como un cierto $a \in A$ y explicar cómo se calcula $f(S(n))$ supuesto que ya hayamos calculado $f(n)$, es decir, dar una función

[6] En general podemos definir una terna como $(a, b, c) = ((a, b), c)$.

g que determine $f(S(n))$ a partir de $f(n)$. Esto determina completamente la función f, que vendrá dada por:

$$f(0) = a, \quad f(1) = f(S(0)) = g(a), \quad f(2) = f(S(1)) = g(g(a)),$$

$$f(3) = g(g(g(a))), \quad f(4) = g(g(g(g(a)))), \quad \ldots$$

Cuando definimos una función de este modo se dice que la estamos definiendo *por recurrencia*, o que la definición es *recursiva.*

Antes de ver ejemplos y aplicaciones vamos a demostrar el teorema:

DEMOSTRACIÓN: Diremos que $h : X \longrightarrow A$ es una *aproximación* si:

1. $X \subset N$, $0 \in X$ y siempre que $n \in X$ y $n \neq 0$ existe un $m \in X$ tal que $n = S(m)$
2. $h(0) = a$ y siempre que $n \in X$ y $S(n) \in X$, entonces $h(S(n)) = g(h(n))$.

Así, las aproximaciones son funciones que cumplen lo que requiere el teorema salvo que no tienen por qué estar definidas en todo N.

Veamos en primer lugar que si $h : X \longrightarrow A$ y $h' : X' \longrightarrow A$ son aproximaciones y $n \in X \cap X'$, entonces $h(n) = h'(n)$.

Lo probamos por inducción sobre n. Concretamente, consideramos la propiedad Pn dada por: si $n \in X \cap X'$, entonces $h(n) = h'(n)$, y vamos a comprobar que la cumplen todos los elementos de N.

Se cumple para 0, porque por definición de aproximación $0 \in X \cap X'$ y $h(0) = a = h'(0)$.

Supongamos, como hipótesis de inducción, que si $n \in X \cap X'$, entonces $h(n) = h'(n)$ y vamos a probar que lo mismo vale para $S(n)$. Para ello suponemos que $S(n) \in X \cap X'$. Entonces, por definición de aproximación, $n \in X \cap X'$ (aquí usamos el cuarto axioma de Peano, que dice que n es el único anterior de $S(n)$). Entonces, por hipótesis de inducción $h(n) = h'(n)$, y por definición de aproximación $h(S(n)) = g(h(n)) = g(h'(n)) = h'(S(n))$. Esto termina la prueba.

Ahora probamos que para todo $n \in N$ existe una aproximación h con n en su dominio. Lo probamos nuevamente por inducción sobre n. Es inmediato que la aplicación $h : \{0\} \longrightarrow A$ dada por $h(0) = a$ es una aproximación, y tiene a 0 en su dominio, luego el resultado es cierto para 0.

Supongamos, por hipótesis de inducción que $h : X \longrightarrow A$ es una aproximación tal que $n \in X$. Si $S(n) \in X$, entonces h cumple lo requerido para $S(n)$. En caso contrario es fácil ver que $h' : X \cup \{S(n)\} \longrightarrow A$ definida como h sobre los elementos de X y como $h'(S(n)) = g(h(n))$ es una aproximación con $S(n)$ en su dominio.

Ahora ya podemos definir $f : N \longrightarrow A$ como sigue: para cada $n \in N$, sabemos que existe una aproximación $h : X \longrightarrow A$ con n en su dominio y que, si tomamos dos cualesquiera, el valor de $h(n)$ va a ser el mismo para ambas. Por lo tanto podemos definir $f(n) = h(n)$.

Esto hace que, en particular, $f(0) = a$, porque todas las aproximaciones asignan a 0 la imagen a. Por otra parte, si $n \in N$ y $h : X \longrightarrow A$ es una aproximación tal que $S(n) \in X$, entonces por definición de aproximación $n \in X$, luego por definición de f tenemos que $f(n) = h(n)$ y $f(S(n)) = h(S(n))$, y por definición de aproximación

$$f(S(n)) = h(S(n)) = g(h(n)) = g(f(n)).$$

Por lo tanto, f cumple lo pedido, y sólo falta probar que sólo hay una f que cumpla esta propiedad. Para ello tenemos que probar que si $f : N \longrightarrow A$ y $f' : N \longrightarrow A$ cumplen las condiciones del enunciado, entonces $f = f'$, para lo cual a su vez basta probar que si $n \in N$ entonces $f(n) = f'(n)$. Esto lo probamos por inducción. Para $n = 0$ se cumple, pues $f(0) = a = f'(0)$. Si suponemos como hipótesis de inducción que $f(n) = f'(n)$, entonces

$$f(S(n)) = g(f(n)) = g(f'(n)) = f'(S(n)).$$

esto termina la prueba de la unicidad de f. ∎

Como primera aplicación demostramos la equivalencia entre todos los sistemas de Peano:

Teorema 1.14 *Si $(N, S, 0)$ y $(N', S', 0')$ son dos sistemas de Peano, entonces existe $f : N \longrightarrow N'$ biyectiva tal que $f(0) = 0'$ y que hace conmutativo el diagrama siguiente:*

$$\begin{array}{ccc} N & \xrightarrow{f} & N' \\ {\scriptstyle S}\uparrow & & \uparrow{\scriptstyle S'} \\ N & \xrightarrow[f]{} & N' \end{array}$$

Observemos que esto significa que $f(1) = f(S(0)) = S'(f(0)) = S'(0') = 1'$, $f(2) = f(S(1)) = S'(f(1)) = S'(1') = 2'$ y, en general, que f transforma el 0 de un sistema en el $0'$ del otro, el 1 de un sistema en el $1'$ del otro, y así sucesivamente. En definitiva, f es como un "diccionario" que traduce "cero" por "zero", "uno" por "one", "dos" por "two" y así sucesivamente, poniendo en evidencia que los dos sistemas de Peano no son más que dos ristras de nombres alternativos para los números naturales, de modo que es lo mismo partir de "dos", calcular el siguiente "tres" y luego traducir "three" que partir de "dos", traducir "two" y luego calcular el siguiente "three".

Demostración: Aplicamos el teorema de recursión a $0' \in N'$ y a la función $S' : N' \longrightarrow N'$. La conclusión es que existe una función $f : N \longrightarrow N'$ tal que $f(0) = 0'$ y para todo $n \in N$ se cumple que $f(S(n)) = S'(f(n))$, que es la condición de conmutatividad para el diagrama. Sólo falta probar que f es biyectiva.

Para ello invertimos los papeles y definimos $f' : N' \longrightarrow N$ mediante el teorema de recursión aplicado al segundo sistema de Peano, de modo que $f'(0') = 0$ y para todo $n' \in N'$ se cumple que $f'(S'(n')) = S(f'(n'))$.

Ahora demostramos por inducción que $f'(f(n)) = n$ para todo $n \in N$. Para $n = 0$ tenemos que $f'(f(0)) = f'(0') = 0$. Si vale para n, entonces $f'(f(S(n))) = f'(S'(f(n))) = S(f'(f(n))) = S(n)$.

Esto prueba que $f \circ f'$ es la identidad en N, y la propiedad 5 de la página 17 nos da que f es inyectiva. Pero invirtiendo los papeles obtenemos también que $f' \circ f$ es la identidad, luego f es suprayectiva, y concluimos que es biyectiva. ∎

A partir de aquí fijamos un sistema de Peano cualquiera $(\mathbb{N}, S, 0)$, que puede ser el que hemos construido concretamente en la sección 1.2 u otro cualquiera, y llamaremos *números naturales* a los elementos de $\mathbb{N}$.

Vamos a aplicar el teorema de recursión tomando como a un número natural $m \in \mathbb{N}$ y como g la aplicación siguiente $S : \mathbb{N} \longrightarrow \mathbb{N}$. El teorema nos da que existe una única $f_m : \mathbb{N} \longrightarrow \mathbb{N}$ tal que $f_m(0) = m$ y $f_m(S(n)) = S(f_m(n))$. Sin embargo, en lugar de usar la notación habitual para aplicaciones, es decir, $f_m(n)$, vamos a usar una notación específica, y escribiremos $m + n = f_m(n)$. Con esta notación, las propiedades que determinan por recurrencia la función f_m se escriben así:

$$m + 0 = m, \qquad m + S(n) = S(m + n).$$

En particular, observamos que $m+1 = m+S(0) = S(m+0) = S(m)$, por lo que a partir de ahora ya no volveremos a escribir $S(n)$ para referirnos al siguiente de un número natural, sino que usaremos la notación $m + 1$. Puesto que ésta será la notación que emplearemos en lo sucesivo, conviene reescribir en estos términos los resultados que hemos enunciado hasta ahora con la notación S:

Principio de inducción *Si probamos que* 0 *tiene la propiedad* P *y bajo la hipótesis de que* $n \in \mathbb{N}$ *cumple la propiedad* P *(hipótesis de inducción) podemos demostrar que* $n+1$ *también cumple la propiedad* P*, entonces podemos asegurar que todo número natural cumple la propiedad* P*.*

Principio de recursión *Si* $g : A \longrightarrow A$ *es una aplicación arbitraria y* $a \in A$*, existe una única aplicación* $f : \mathbb{N} \longrightarrow A$ *tal que* $f(0) = a$ *y para todo* $n \in \mathbb{N}$ *se cumple que* $f(n + 1) = g(f(n))$

Suma de números naturales La *suma* de dos números naturales m y n está unívocamente determinada por las propiedades siguientes:

$$m + 0 = m, \qquad m + (n + 1) = (m + n) + 1.$$

Esta definición se corresponde con la suma que el lector conoce sin duda desde sus primeros años. Por ejemplo:

$$2 + 3 = 2 + (2 + 1) = (2 + 2) + 1 = (2 + (1 + 1)) + 1 = ((2 + 1) + 1) + 1$$
$$= (3 + 1) + 1 = 4 + 1 = 5.$$

Producto de números naturales El *producto* de dos números naturales está unívocamente determinado por las propiedades

$$m \cdot 0 = 0, \qquad m(n+1) = m \cdot n + m.$$

Observemos que la definición de producto se basa en el teorema de recursión tomando $a = 0$ y como $g : \mathbb{N} \longrightarrow \mathbb{N}$ la aplicación dada por $g(n) = n + m$. Esto nos define una función f_m, pero en lugar de $f_m(n)$ escribimos $m \cdot n$, o simplemente mn. Como en el caso de la suma, este producto es el producto "de toda la vida". No obstante, antes de hacer cuentas con él es conveniente demostrar las propiedades básicas de estas operaciones:

1. $(m + n) + r = m + (n + r)$.

   Por inducción[7] sobre r. Para $r = 0$ se reduce a $m + n = m + n$. Si vale para r, entonces

   $$(m+n)+(r+1) = ((m+n)+r)+1 = (m+(n+r))+1$$
   $$= m+((n+r)+1) = m+(n+(r+1)),$$

   donde hemos aplicado la definición de suma, la hipótesis de inducción, y dos veces más la definición de suma.

   Esta propiedad nos permite escribir expresiones $m + n + r$ sin necesidad de intercalar paréntesis. Así, por ejemplo,

   $$3+3 = 3+2+1 = 3+1+1+1 = 4+1+1 = 5+1 = 6.$$

2. $m + n = n + m$.

   Para probar esto demostramos antes algunos casos particulares. En primer lugar demostramos que $0 + n = n$ por inducción sobre n (notemos que $n + 0 = n$ se cumple por la definición de suma). Para $n = 0$ es $0 + 0 = 0$, Si vale para n, entonces

   $$0 + (n + 1) = (0 + n) + 1 = n + 1.$$

   En segundo lugar demostramos que $1 + n = n + 1$, también por inducción sobre n. Para $n = 0$ es $1 + 0 = 1 = 0 + 1$, la primera igualdad por la definición de suma y la segunda porque ya hemos visto que sumar 1 equivale a pasar al siguiente. Si vale para n, entonces

   $$1 + (n + 1) = (1 + n) + 1 = (n + 1) + 1.$$

   Ahora probamos el caso general por inducción sobre n. Para $n = 0$ es $m + 0 = m = 0 + m$. Si vale para n, entonces

   $$m+(n+1) = (m+n)+1 = (n+m)+1 = n+(m+1)$$
   $$= n+(1+m) = (n+1)+m,$$

   donde hemos usado las propiedades precedentes.

[7] "Por inducción sobre r" significa que vamos a aplicar el principio de inducción a la propiedad $Pr \equiv (m + n) + r = m + (n + r)$, considerando m y n fijos.

3. $(m+n)r = mr + nr$.

   Por inducción sobre r. Para $r = 0$ es $(m+n)0 = 0 = 0+0 = m\cdot 0 + n\cdot 0$. Si vale para r entonces

   $$(m+n)(r+1) = (m+n)r + m + n = mr + nr + m + n$$
   $$= mr + m + nr + r = m(r+1) + n(r+1).$$

4. $m(n+r) = mn + mr$.

   Por inducción sobre r. Para $r = 0$ es $mn = mn$. Si vale para r, entonces

   $$m(n+(r+1)) = m((n+r)+1) = m(n+r) + m$$
   $$= mn + mr + m = mn + m(r+1).$$

5. $(mn)r = m(nr)$.

   Por inducción sobre r. Para $r = 0$ es $(mn)0 = 0 = m(0) = m(n0)$. Si vale para r, entonces

   $$(mn)(r+1) = (mn)r + mn = m(nr) + mn = m(nr+n) = m(n(r+1)),$$

   donde hemos usado la propiedad anterior.

   Como en el caso de la suma, esta propiedad hace innecesarios los paréntesis entre los términos de una multiplicación.

6. $n \cdot 1 = n$.

   $n \cdot 1 = n(0+1) = n\cdot 0 + n = 0 + n = n$.

7. $mn = nm$.

   Demostramos antes algunos casos particulares. En primer lugar vemos que $0 \cdot n = 0$. Para $n = 0$ es $0 \cdot 0 = 0$. Si vale para n, entonces

   $$0 \cdot (n+1) = 0 \cdot n + 0 = 0 + 0 = 0.$$

   En segundo lugar $1 \cdot n = n$. Para $n = 0$ es $1 \cdot 0 = 0$. Si vale para n, entonces $1(n+1) = 1 \cdot n + 1 \cdot 1 = n + 1$. Ahora probamos el caso general, por inducción sobre n. Para $n = 0$ es $m0 = 0 = 0m$. Si vale para n, entonces

   $$m(n+1) = mn + m = nm + m = nm + 1 \cdot m = (n+1)m,$$

   donde hemos usado la propiedad 3.

8. Si $m + r = n + r$, entonces $m = n$.

   Por inducción sobre r. Para $r = 0$ tenemos $m+0 = n+0$, luego ciertamente $m = n$. Si vale para r y tenemos $m + (r+1) = n + (r+1)$, esto es lo mismo que $(m+r)+1 = (n+r)+1$, o también $S(m+r) = S(n+r)$, luego por el cuarto axioma de Peano $m + r = n + r$, luego $m = n$ por hipótesis de inducción.

9. Si $m + n = 0$, entonces $m = n = 0$.

   En efecto, si fuera $n \neq 0$, entonces $n = r + 1$ para cierto r (teorema 1.7), y entonces $m + n = (m + r) + 1 \neq 0$, pues el 0 no es el siguiente de ningún número natural.

10. Si $mn = 0$, entonces $m = 0$ o bien $n = 0$.

    En caso contrario, $m = m' + 1$, $n = n' + 1$, luego, al igual que antes,

$$mn = (m' + 1)(n' + 1) = (m' + 1)n' + m' + 1 \neq 0.$$

Ahora ya es fácil operar con números naturales. Por ejemplo,

$$3 \cdot 2 = 3(1 + 1) = 3 + 3 = 6.$$

El último concepto básico de la aritmética de los números naturales es la relación de orden:

Ordenación de los números naturales Diremos que un número natural m es menor o igual que otro n, y lo representaremos por $m \leq n$, si existe un $r \in \mathbb{N}$ tal que $m + r = n$. Observemos que en tal caso dicho r es único por la propiedad 8 precedente, por lo que podemos llamarlo *resta* de n y m, y lo representaremos por $n - m$.

Notemos que $m \leq m$, porque $m + 0 = m$. Escribiremos $m < n$ para indicar que $m \leq n$ y $m \neq n$.

Por ejemplo, como $2 + 3 = 5$, se cumple que $2 < 5$ y que $5 - 2 = 3$.

Veamos ahora las propiedades correspondientes:

1. Para todo natural n se cumple que $0 \leq n$.

   En efecto, $0 + n = n$.

2. Si m y n son números naturales, entonces $m \leq n$ o bien $n \leq m$.

   Por inducción sobre n, si $n = 0$ se cumple $n = 0 \leq m$. Si vale para n, tenemos que $m \leq n$ o bien $n \leq m$. Si se da el primer caso, existe un r tal que $m + r = n$, luego $m + r + 1 = n + 1$, luego $m \leq n + 1$, como había que probar.

   Supongamos ahora que $n \leq m$ y sea r tal que $n + r = m$. Si $r = 0$ entonces $n = m$, luego $n + 1 = m + 1$, luego $m \leq n + 1$, como había que probar. Si $r \neq 0$, existe un r' tal que $r = r' + 1$, luego $n + r' + 1 = m$, luego $n + 1 \leq m$.

3. Si $m \leq n$ y $n \leq m$, entonces $m = n$.

   Tenemos que $m + r = n$ y $n + r' = m$, luego $m + r + r' = m = m + 0$, luego $r + r' = 0$, luego $r = r' = 0$, luego $m = n$.

4. Si $m \leq n$ y $n \leq r$, entonces $m \leq r$.

   Tenemos que $m + u = n$ y $n + v = r$, luego $m + u + v = r$, luego $m \leq r$.

5. Se cumple $m \leq n$ si y sólo si $m + r \leq n + r$.

   En efecto, $m \leq n$ equivale a que exista un u tal que $m + u = n$, lo cual equivale a que $m + r + u = n + r$, lo cual equivale a que $m + r \leq n + r$.

6. Si $r \neq 0$ y $mr = nr$, entonces $m = n$.

   En efecto, no perdemos generalidad si suponemos $m \leq n$, de modo que $n = m + u$, luego $nr = mr + ur$, luego $mr + 0 = mr + ur$, luego $ur = 0$, luego $u = 0$, luego $m = n$.

7. Si $r \neq 0$, entonces $m \leq n$ si y sólo si $mr \leq nr$.

   Si $m \leq n$, existe un u tal que $m + u = n$, luego $mr + ur = nr$, luego $mr \leq nr$. Recíprocamente, si $mr \leq nr$, entonces o bien $m \leq n$, como queremos probar, o bien $n \leq m$, en cuyo caso $nr \leq mr$ por la parte ya probada, luego $nr = mr$ por 3, luego $n = m$ por 6, luego $m \leq n$.

Notemos que, trivialmente, $n < S(n)$, luego la ordenación que acabamos de introducir se corresponde con la que resulta al ir generando los números naturales a partir del 0 mediante la operación "siguiente", es decir:

$$0 < 1 < 2 < 3 < 4 < \cdots$$

La relación de orden en $\mathbb{N}$ permite dar un significado preciso a algunas expresiones con "puntos suspensivos". Por ejemplo, para cada $n \in \mathbb{N}$ podemos definir

$$I_n = \{1, \ldots, n\}, \qquad I_n^* = \{0, \ldots, n-1\}$$

y esto hay que entenderlo como que $I_n = \{m \in \mathbb{N} \mid 1 \leq m \leq n\}$ (con lo que en particular $I_0 = \varnothing$) e igualmente $I_n^* = \{m \in \mathbb{N} \mid m < n\}$, que son definiciones válidas por especificación.

Exponenciación de números naturales La *exponenciación* de dos números naturales está unívocamente determinada por las propiedades siguientes:

$$m^0 = 1, \qquad m^{n+1} = m^n \cdot m.$$

Dejamos como ejercicio demostrar sus propiedades básicas:

1. Si $n \neq 0$, entonces $0^n = 0$ (pero $0^0 = 1$, por definición).
2. $1^n = 1$.
3. $m^{n+r} = m^n \cdot m^r$.
4. $(m^n)^r = m^{nr}$.
5. $(mn)^r = m^r \cdot n^r$.
6. Si $m < n$ y $1 \leq r$, entonces $m^r < n^r$.
7. Si $m \leq n$, entonces $m^r \leq n^r$.

1.5 Relaciones de orden

Acabamos de definir una ordenación de los números naturales (que no es sino la ordenación usual que el lector conocía sin duda de antemano). Sin embargo, el conjunto de los números naturales no es el único conjunto que vamos a manejar en el que es posible considerar una relación de orden, y por ello es conveniente desarrollar una teoría general —o, más bien, un lenguaje general— sobre conjuntos ordenados para estar en condiciones de aplicar los mismos hechos generales a distintos casos particulares sin necesidad de redefinir los conceptos y repetir las demostraciones en cada caso. Para ello necesitamos en primer lugar una teoría general sobre relaciones.

Definición 1.15 Una *relación* R en un conjunto A es un subconjunto del producto cartesiano $R \subset A \times A$.

En la práctica, en lugar de escribir $(a, b) \in R$, escribiremos $a\,R\,b$ y diremos que a *está relacionado* con b (respecto de la relación R).

Observaciones Como en el caso de las funciones, en la práctica podemos olvidar que las relaciones son técnicamente conjuntos de pares ordenados. Así, para definir una relación R en un conjunto A basta especificar que $a\,R\,b$ si y sólo si a y b cumplen una determinada propiedad Pab. Esto equivale a definir R como el conjunto

$$R = \{(a, b) \in A \times A \mid Pab\}.$$

Por ejemplo, si R es una relación en un conjunto A, podemos definir su *relación inversa* como la relación R^{-1} dada por

$$a\,R^{-1}\,b \quad \text{si y sólo si} \quad b\,R\,a.$$

Técnicamente, R^{-1} se define por el axioma de especificación, como el conjunto

$$R^{-1} = \{x \in A \times A \mid \text{existen } a, b \in A \text{ tales que } x = (a, b) \text{ y } (b, a) \in R\},$$

pero la primera forma de definir ésta o cualquier otra relación es más práctica.

Otro ejemplo: en la sección anterior hemos definido la ordenación de los números naturales como una propiedad $m \leq n$ que pueden tener o no dos números naturales dados. Ahora podemos considerar el conjunto

$$\{(m, n) \in \mathbb{N} \times \mathbb{N} \mid m \leq n\}$$

y, si llamamos $\leq$ a este conjunto, tenemos que $m \leq n$ en el sentido de la sección anterior si y sólo si se cumple esto mismo en el sentido que acabamos de introducir, es decir, en el sentido de que el par (m, n) pertenece al conjunto $\leq$.

■

Veamos algunos conceptos del lenguaje de las relaciones.

Definición 1.16 Una relación R en un conjunto A es:

- *Reflexiva* si para todo $a \in A$ se cumple que $a\,R\,a$.
- *Irreflexiva* si para ningún $a \in A$ se cumple que $a\,R\,a$.
- *Simétrica* si cuando a, $b \in A$ cumplen $a\,R\,b$ también cumplen $b\,R\,a$.
- *Antisimétrica* si cuando a, $b \in A$ cumplen $a\,R\,b$ y $b\,R\,a$ entonces $a = b$.
- *Asimétrica* si no existen a, $b \in A$ que cumplan $a\,R\,b$ y $b\,R\,a$.
- *Transitiva* si cuando a, b, $c \in A$ cumplen $a\,R\,b$ y $b\,R\,c$ entonces $a\,R\,c$.
- *Conexa* si cuando a, $b \in A$, o bien se cumple $a\,R\,b$ o bien $b\,R\,a$.

Una relación $\leq$ en un conjunto A es una *relación de orden parcial* si es reflexiva, antisimétrica y transitiva. Si además es conexa, se dice que es una *relación de orden total.*

Una relación $<$ en un conjunto A es una *relación de orden parcial estricto* si es irreflexiva, asimétrica y transitiva. Si además cuando a, $b \in A$ se cumple $a < b$ o $b < a$ o $a = b$ se dice que es una *relación de orden total estricto.*

Un *conjunto parcialmente (resp. totalmente) ordenado* es un par $(A, \leq)$, donde $\leq$ es una relación de orden parcial (resp. total) en A.

Observaciones En la sección anterior hemos demostrado que la relación $\leq$ definida allí sobre el conjunto $\mathbb{N}$ de los números naturales es una relación de orden total, y es fácil ver que la relación $<$ definida como $m < n$ si $m \leq n$ y $m \neq n$ es una relación de orden total estricto.

En realidad esto es un caso particular de un hecho general: siempre que en un conjunto A tenemos definida una relación de orden (total o parcial) no estricto $\leq$, la relación dada por $a < b$ si y sólo si $a \leq b$ y $a \neq b$ es una relación de orden estricto (total o parcial según lo sea la relación dada) y, recíprocamente, si $<$ es cualquier relación de orden estricto en un conjunto A, la relación dada por $a \leq b$ si y sólo si $a < b$ o $a = b$ es una relación de orden no estricto, de modo que una y otra se determinan mutuamente. Por lo tanto, es indiferente hablar de relaciones de orden estricto o no estricto en un conjunto: siempre que tenemos definida una, podemos dar por hecho que tenemos definida de este modo una del otro tipo.

Al hablar de conjuntos ordenados es frecuente omitir la relación de orden, y decir, por ejemplo que A es un conjunto ordenado, de modo que hay que entender que $(A, \leq)$ es un conjunto ordenado con cierta relación de orden que no se menciona, ya sea porque es una relación arbitraria denotada por $\leq$, ya porque sea una relación específica que se deduce del contexto. Por ejemplo, siempre que hablemos de $\mathbb{N}$ como conjunto ordenado se entenderá que hablamos del par $(\mathbb{N}, \leq)$, donde $\leq$ es la relación definida en la sección precedente.

El hecho de emplear el signo $\leq$ para referirnos tanto a una relación de orden arbitraria como a relaciones de orden concretas definidas en ciertos conjuntos (de momento sólo tenemos definida la de $\mathbb{N}$) no da lugar a confusión porque rara vez se consideran varias relaciones de orden definidas sobre un mismo conjunto, por lo que el significado de $\leq$ se deduce en cada contexto de los elementos sobre los que actúa. Sólo en caso de considerar varias relaciones de orden sobre un mismo conjunto será necesario distinguir unas de otras con signos como $\leq'$, $\leq^*$, $\trianglelefteq$, etc.

En general, si X es un conjunto parcialmente ordenado, además de la relación $\leq$ sobrentendida, tenemos la relación estricta asociada $<$, y las relaciones inversas de ambas, que representaremos siempre por $\geq$ y $>$ respectivamente.

Así, por ejemplo, $5 \geq 3$ es lo mismo que $3 \leq 5$. ∎

Ampliamos ahora nuestro vocabulario sobre un conjunto parcialmente ordenado A:

1. Si $X \subset A$, consideraremos a X como conjunto ordenado con la restricción[8] a X de la relación de orden de A.

2. Si $X \subset A$ y $m \in X$, diremos que m es el *mínimo* de X si para todo $x \in X$ se cumple que $m \leq x$. Notemos que si X tiene mínimo, éste es único, pues si m y m' son dos mínimos, entonces ambos están en X y deben cumplir $m \leq m'$ y $m' \leq m$, luego $m = m'$. Por ello podemos representarlo por $\min X$ (aunque recordemos que, en general, puede no existir).

3. Si $X \subset A$ y $M \in X$, diremos que M es el *máximo* de X si para todo $x \in X$ se cumple que $x \leq M$. Al igual que sucede con los mínimos, si X tiene máximo éste es único, y podemos representarlo por $\max X$.

4. Si $X \subset A$ y $c \in A$, diremos que c es una *cota inferior* de X si para todo $x \in X$ se cumple que $c \leq x$. La diferencia con el mínimo es que no exigimos que $c \in X$, y esto hace que un mismo conjunto pueda tener muchas cotas inferiores distintas (o ninguna).

5. Si $X \subset A$ y $C \in A$, diremos que C es una *cota superior* de X si para todo $x \in X$ se cumple $x \leq C$.

Por ejemplo, si $X = \{3, 5, 6\}$, tenemos que, respecto al orden usual de $\mathbb{N}$, los números 0, 1, 2, 3 son cotas inferiores de X y $3 = \min X$, mientras que cualquier número $n \geq 6$ es una cota superior, y $6 = \max X$.

También es claro que $0 = \min \mathbb{N}$, mientras que $\mathbb{N}$ no tiene máximo elemento, ya que si $M \in \mathbb{N}$ fuera el máximo, debería cumplir $M + 1 \leq M$, lo cual es falso.

Observemos que si un conjunto tiene máximo entonces éste es el mínimo del conjunto de sus cotas superiores, e igualmente el mínimo (si existe) es el máximo del conjunto de sus cotas inferiores.

La ordenación de los números naturales tiene una propiedad notable:

[8] Técnicamente, la restricción es $\leq_X = \leq_A \cap (X \times X)$, de modo que, si $x, y \in X$, entonces $x \leq_X y$ es equivalente a $x \leq_A y$, pero $\leq_X$ sólo está definida para elementos de X.

Teorema 1.17 (Principio de buena ordenación) *Todo subconjunto no vacío de $\mathbb{N}$ tiene un mínimo elemento.*

DEMOSTRACIÓN: Sea $X \subset \mathbb{N}$ un conjunto no vacío, de modo que existe un cierto $m \in X$. Consideremos la propiedad $Pc \equiv c$ es una cota inferior de X. Claramente se cumple $P0$, y $m+1$ no tiene la propiedad P, pues no es cierto que $m+1 \leq m$. Si fuera cierto que cuando c cumple la propiedad P también la cumple $c+1$, el principio de inducción nos daría que todo número natural cumple la propiedad P, y hemos visto que eso es falso. Por lo tanto, tiene que existir un $c \in \mathbb{N}$ que cumpla Pc, pero de modo que $c+1$ no cumpla P. En otras palabras, c es una cota inferior de X, pero $c+1$ no lo es. Lo segundo significa que existe un $n \in X$ tal que $n < c+1$, y lo primero implica que $c \leq n < c+1$. De aquí se sigue[9] que $c = n$, luego $c \in X$ es el mínimo de X. ∎

Por ejemplo, es fácil probar que $n+1 = \text{mín}\{m \in \mathbb{N} \mid m > n\}$, es decir, que el siguiente de un número natural n es el mínimo número natural mayor que n.

En general, el principio de buena ordenación hace que si tenemos que existe un número natural que cumple una propiedad P, podamos considerar "el mínimo número natural que cumple la propiedad P", es decir, el mínimo del conjunto no vacío $\{n \in \mathbb{N} \mid Pn\}$, con lo que tenemos un número que, además de cumplir P, cumple que ningún natural menor cumple P. Veamos un ejemplo de este tipo de argumentación:

Teorema 1.18 (División euclídea) *Dados dos números naturales D y d, con $d \neq 0$, existen unos únicos c y r tales que $D = dc + r$ y $r < d$.*

DEMOSTRACIÓN: Como $d \neq 0$, tenemos que $1 \leq d$, luego $D \leq dD$. Así pues, D es un número natural x que cumple $D \leq dx$, luego podemos tomar el mínimo número natural x que cumple $D \leq dx$.

Si $D = dx$, entonces basta tomar $c = x$ y $r = 0$. Si, por el contrario, $D < dx$ entonces necesariamente $x \neq 0$, luego podemos considerar $c = x - 1 < x$. Por la minimalidad de x tiene que ser $dc < D < d(c+1) = dc + d$. Sea $r = D - dc$, de modo que $dc + r < dc + d$, luego $r < d$.

Con esto tenemos probado que existen c y r que cumplen lo pedido. Ahora tenemos que ver que son únicos. Si c_1, c_2, r_1, r_2 cumplieran lo requerido y $c_1 \neq c_2$, podemos suponer que $c_1 < c_2$. Entonces

$$D = dc_1 + r_1 < dc_1 + d = d(c_1 + 1) \leq dc_2 \leq dc_2 + r_2 = D,$$

contradicción. Por lo tanto, $c_1 = c_2$, y entonces $D = dc_1 + r_1 = dc_1 + r_2$, luego $r_1 = r_2$. ∎

En las condiciones del teorema anterior, se dice que c y r son, respectivamente, el *cociente* y el *resto* de la *división euclídea* del *dividendo* D entre el *divisor* d.

[9] Tenemos que $c+r = n$ y $n+s = c+1$, con $s \neq 0$, luego $s = s'+1$, luego $c+r+s'+1 = c+1$, luego $r + s' = 0$, luego $r = 0$, luego $c = n$.

El proceso de tomar el menor número natural que cumple una propiedad puede usarse en una forma fuerte de razonamiento por reducción al absurdo conocida como *"demostración por contraejemplo mínimo"*. Esto significa que si queremos probar que todo número natural cumple una propiedad P, podemos suponer, por reducción al absurdo que existe un número natural que no cumple P, y entonces tomar el mínimo número natural n que no cumple P. Así, para llegar a un absurdo no sólo contamos con que n no cumple P, sino también con que todos los números menores que n cumplen P.

A su vez, la demostración por contraejemplo minimal es equivalente a una versión fuerte del principio de inducción. En principio, en una demostración por inducción suponemos que el anterior de un número cumple una propiedad P y demostramos que dicho número cumple P. Ahora vamos a ver que, si queremos, podemos suponer como hipótesis de inducción que, no sólo el inmediato anterior del número cumple P, sino que todos los números menores la cumplen:

Teorema 1.19 (Principio de inducción fuerte) *Sea $A \subset \mathbb{N}$ tal que, para todo número natural n, si todo $m < n$ cumple que $m \in A$, también $n \in A$. Entonces $A = \mathbb{N}$.*

Demostración: Si no es cierto que $A = \mathbb{N}$, como $A \subset \mathbb{N}$, tiene que existir un $n \in \mathbb{N}$ que no esté en A. Podemos tomar entonces un contraejemplo mínimo, es decir, el mínimo número natural n que no está en A. Dicho número tiene la propiedad de que todos los menores que él sí que están en A, luego la hipótesis del teorema nos dice que n tiene que estar en A, y eso es una contradicción. ∎

Si aplicamos esto a un conjunto de la forma $A = \{n \in \mathbb{N} \mid Pn\}$, vemos que si suponemos como hipótesis de inducción que todos los números menores que uno dado n tienen la propiedad P y llegamos a que n también tiene la propiedad P, podemos concluir que todo número natural n tiene la propiedad P.

Observemos que con esta formulación no hace falta exigir una demostración aparte de $P0$, porque es trivialmente cierto que "todo número natural menor que 0 cumple la propiedad P", y estamos suponiendo que bajo este caso trivial de la hipótesis de inducción podemos demostrar $P0$. Esto no quita para que, en la práctica, al suponer que todo número natural menor que n cumple P, pueda ser conveniente tratar por separado los casos $n = 0$ (donde la hipótesis de inducción no aporta nada) y $n \neq 0$.

Existe un principio de recursión fuerte análogo al principio de inducción fuerte. Según el principio de recursión, para definir una función $f : \mathbb{N} \longrightarrow A$ por recurrencia basta especificar $f(0)$ y cómo puede calcularse $f(n+1)$ a partir de $f(n)$, es decir, basta imponer una relación de la forma $f(n+1) = g(f(n))$ que exprese $f(n+1)$ como una función dada de $f(n)$. Ahora vamos a probar que, si queremos, podemos definir $f(n)$ como función de todos los valores previos $f(0), \ldots, f(n-1)$ o, en otras palabras, que para definir $f(n)$ podemos suponer que f ya está definida sobre los números anteriores a n.

Enunciar esto con precisión requiere algunas definiciones previas. Recordemos que hemos definido $I_n^* = \{0, \ldots, n-1\}$ (en particular $I_0^* = \varnothing$). Vamos a llamar A^* al conjunto[10] de todas las aplicaciones $h : I_n^* \longrightarrow A$, para algún $n \in \mathbb{N}$. Si $f : \mathbb{N} \longrightarrow A$, para cada $n \in \mathbb{N}$ definimos $f_n = f|_{I_n^*} \in A^*$, así f_n es la restricción de f a los números naturales menores que n. En particular $f_0 = \varnothing$.

Principio de recursión fuerte Si A es un conjunto y $g : A^* \longrightarrow A$, existe una única aplicación $f : \mathbb{N} \longrightarrow A$ tal que para todo $n \in \mathbb{N}$ se cumple que $f(n) = g(f_n)$.

Notemos que esto expresa lo que hemos indicado: que podemos definir $f(n)$ como una cierta función de f_n, es decir, definir f suponiendo que f ya está definida sobre los números menores que n.

Demostración: Diremos que h es una *aproximación* si existe un $n \in \mathbb{N}$ tal que $h : I_n^* \longrightarrow A$ y para todo $m < n$ se cumple que $h(m) = g(h|_{I_m^*})$. Veamos que si $m \leq n$ y $h : I_m^* \longrightarrow A$, $h' : I_n^* \longrightarrow A$ son dos aproximaciones, entonces $h'|_{I_m^*} = h$, es decir, que ambas coinciden sobre la parte común de su dominio.

En efecto, si no es así, existe un $i \in I_m^*$ tal que $h(i) \neq h'(i)$. Podemos tomar el mínimo número natural en el que esto sucede, y entonces $h|_{I_i^*} = h'|_{I_i^*}$, luego $h(i) = g(h|_{I_i^*}) = g(h'|_{I_i^*}) = h'(i)$, por definición de aproximación, contradicción.

Ahora veamos que para cada $m \in \mathbb{N}$ existe una aproximación $h : I_m^* \longrightarrow A$. Lo probamos por inducción sobre m. Para $m = 0$ es $I_0^* = \varnothing$ y basta tomar[11] $h = \varnothing$. Si vale para m y tenemos una aproximación $h : I_m^* \longrightarrow A$, definimos $h' : I_{m+1}^* \longrightarrow A$ como la aplicación que sobre I_m^* coincide con h y además $h'(m) = g(h)$. Es claro que h' es una aproximación, pues si $n \in I_m^*$ se cumple que $h'(n) = h(n) = g(h|_{I_n^*}) = g(h'|_{I_n^*})$, y para $n = m$ tenemos también que $h'(m) = g(h) = g(h'|_{I_m^*})$.

Ahora basta definir $f : \mathbb{N} \longrightarrow A$ como $f(n) = h(n)$, donde h es cualquier aproximación que tenga a n en su dominio. Hemos probado que no importa cuál elijamos, pues todas coinciden en la parte común de sus dominios. En particular $f_n = h|_{I_n^*}$, luego se cumple que $f(n) = h(n) = g(h|_{I_n^*}) = g(f_n)$, como queríamos probar.

La unicidad de f se debe a que si existiera otra $f' : \mathbb{N} \longrightarrow A$ que cumpliera lo mismo y $f' \neq f$, entonces existiría un $n \in \mathbb{N}$ tal que $f(n) \neq f'(n)$, y podríamos tomar el menor número natural que cumple esto, en cuyo caso $f_n = f'_n$, luego $f(n) = g(f_n) = g(f'_n) = f'(n)$, contradicción. ∎

En la práctica nunca hay necesidad de explicitar cuál es exactamente la función g que estamos considerando al aplicar el principio de recursión fuerte.

[10] Notemos que A^* es un conjunto, pues si $h : I_n^* \longrightarrow A$, entonces $h \subset I_n^* \times A \subset \mathbb{N} \times A$, luego $h \in \mathcal{P}(\mathbb{N} \times A)$, luego podemos definir por especificación

$$A^* = \{h \in \mathcal{P}(\mathbb{N} \times A) \mid \text{existe un } n \in \mathbb{N} \text{ tal que } h : I_n^* \longrightarrow A\}.$$

[11] Notemos que, de acuerdo con la definición de aplicación, se cumple trivialmente que $\varnothing : \varnothing \longrightarrow A$.

La idea es que podemos definir $f(n)$ con cualquier criterio que involucre la restricción f_n. Cualquier criterio explícito que asigne un elemento de A a una función f_n que se supone definida podrá plasmarse en una función g adecuada. En particular, puesto que en la definición de $f(n+1)$ podemos hacer referencia a f_{n+1}, también podemos hacer referencia al dominio de f_{n+1}, que es I^*_{n+1}, o bien al máximo del dominio de f_{n+1}, que es n. Esto nos da un principio de recursión intermedio entre el que acabamos de demostrar y el que teníamos ya probado:

> *Para definir una función $f : \mathbb{N} \longrightarrow A$ podemos definir $f(0)$ y luego definir $f(n+1)$ en función de $f(n)$ y de n.*

Veamos un ejemplo de esta situación:

Definición 1.20 Definimos la *función factorial* $\mathbb{N} \longrightarrow \mathbb{N}$, en la que la imagen de un número natural n se representa por $n!$, como la dada por las condiciones

$$0! = 1, \qquad (n+1)! = n! \cdot (n+1).$$

De este modo

$$0! = 1, \quad 1! = 1, \quad 2! = 1 \cdot 2, \quad 3! = 1 \cdot 2 \cdot 3, \quad 4! = 1 \cdot 2 \cdot 3 \cdot 4, \quad \ldots$$

y en general "resumiremos" la definición recurrente de $n!$ mediante la expresión con puntos suspensivos: $n! = 1 \cdot 2 \cdots n$, con el convenio adicional de que $0! = 1$.

Técnicamente, esta definición no se ajusta a las condiciones de la versión original del principio de recursión, porque no se define $(n+1)!$ como función únicamente de $n!$, sino de $n!$ y de n, pero, por lo que acabamos de explicar, la existencia y unicidad de la función factorial sí que está justificada por la versión fuerte del principio de recursión.

1.6 Conjuntos finitos

Ya hemos observado que $\mathbb{N}$ es un conjunto infinito, pero no hemos introducido ninguna definición formal que nos permita distinguir entre conjuntos finitos e infinitos. Ahora estamos en condiciones de hacerlo. Para ello conviene introducir un concepto general:

Definición 1.21 Diremos que dos conjuntos X e Y son *equipotentes* y lo representaremos por $X \sim Y$, si existe una aplicación $f : X \longrightarrow Y$ biyectiva.

Por ejemplo, los conjuntos $A = \{0, 1, 2\}$ y $C = \{0, 2, 4\}$ son equipotentes, pues en la página 15 se muestra una biyección $k : A \longrightarrow C$. Recordemos que las biyecciones "emparejan" los elementos de un conjunto con los de otro, de modo que cuando entre dos conjuntos se puede establecer una biyección esto se traduce en lo que normalmente expresamos diciendo que "tienen el mismo número de elementos" (en el ejemplo, A y C tienen ambos tres elementos). Ésa es la idea que vamos a explotar aquí. Antes demostramos unas propiedades elementales:

Teorema 1.22 *Se cumple:*[12]

1. $X \sim X$.

2. *Si* $X \sim Y$, *entonces* $Y \sim X$.

3. *Si* $X \sim Y$ *e* $Y \sim Z$, *entonces* $X \sim Z$.

DEMOSTRACIÓN: 1) La identidad $i_X : X \longrightarrow X$ es una aplicación biyectiva que prueba que todo conjunto es equipotente a sí mismo.

2) Si $X \sim Y$, existe $f : X \longrightarrow Y$ biyectiva, y es claro entonces que la aplicación inversa $f^{-1} : Y \longrightarrow X$ es biyectiva, luego $Y \sim X$.

3) Por hipótesis existen $f : X \longrightarrow Y$ y $g : Y \longrightarrow Z$ biyectivas, y es fácil comprobar entonces que $f \circ g : X \longrightarrow Z$ es biyectiva. ∎

La existencia de una aplicación inyectiva $f : A \longrightarrow B$ se traduce en que los objetos de A pueden emparejarse uno a uno con los de B, aunque pueden sobrar objetos en B (que no sobren equivale a que f sea suprayectiva y, por consiguiente, biyectiva). Por lo tanto, el teorema siguiente es intuitivamente obvio, aunque una prueba formal requiere distinguir algunos casos:

Teorema 1.23 *Para todo par de números naturales* m *y* n, *existe una aplicación inyectiva* $f : I_m \longrightarrow I_n$ *si y sólo si* $m \leq n$.

DEMOSTRACIÓN: Recordemos que $I_n = \{1, \ldots, n\}$. Si $m \leq n$, es claro que $I_m \subset I_n$, por lo que la inclusión $i : I_m \longrightarrow I_n$ es una aplicación inyectiva.

Falta probar que si $m > n$ no puede existir $f : I_m \longrightarrow I_n$ inyectiva. De no ser así, podemos tomar el menor número natural m para el que existe una aplicación $f : I_m \longrightarrow I_n$ inyectiva para cierto $n < m$.

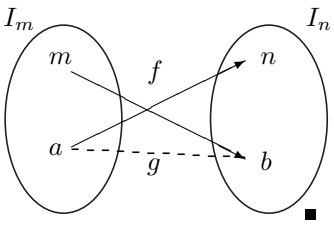
Notemos que $m > n \geq 0$, luego $1 \in I_m$ y $f(1) \in I_n$, luego $n \neq 0$. Digamos que $m = m' + 1$ y que $n = n' + 1$.

Si n no tiene antiimagen por f, entonces $f|_{I_{m'}} : I_{m'} \longrightarrow I_{n'}$ inyectiva, y $n' < m'$, con lo que $m' < m$ cumple la misma propiedad que m, cuando se suponía que m era el mínimo que la cumplía, contradicción.

Supongamos pues que existe un $a \in I_m$ tal que $f(a) = n$. Si $a = m$, sigue siendo cierto que $f|_{I_{m'}} : I_{m'} \longrightarrow I_{n'}$ es inyectiva, pues al quitar m del dominio de la aplicación, ya no necesitamos a n en la imagen, y tenemos la misma contradicción.

[12] Quizá el lector piense que la equipotencia puede considerarse una relación en el sentido de la sección anterior, es decir, que podamos considerar que la relación $\sim$ es el conjunto $E = \{(X, Y) \mid X \sim Y\}$. Sin embargo, no existe tal conjunto E. La razón es que, si existiera, todo conjunto X cumpliría que $(X, X) = \{\{X\}, \{X, X\}\} \in E$, luego $\{X\} \in \bigcup E$, luego $X \in \bigcup\bigcup E$, luego podríamos definir $V = \{X \in \bigcup\bigcup E \mid X = X\}$, y sería el conjunto de todos los conjuntos, que ya sabemos que lleva a una contradicción. Ésta es una de las pocas ocasiones en las que "rozaremos el abismo".

Supongamos, por último, que $a < m$. Entonces $f(m) = b \in I_n$, con $b \neq n$, pues si fuera $b = n$ sería $f(m) = f(a)$, luego $m = a$. Por lo tanto $b \in I_{n'}$. Esto nos permite definir $g : I_{m'} \longrightarrow I_{n'}$ como la aplicación que coincide con f salvo que $g(a) = b$, que sigue siendo inyectiva y nos lleva a la misma contradicción. ∎



En particular, si $I_m \sim I_n$, tenemos aplicaciones biyectivas (luego inyectivas) de uno en otro y viceversa, luego el teorema anterior nos da que $m = n$. Esto justifica la definición siguiente:

Definición 1.24 Un conjunto X es *finito* si existe un $n \in \mathbb{N}$ tal que $X \sim I_n$. En tal caso n es único y recibe el nombre de *cardinal* de X. Lo representaremos por $|X|$. Los conjuntos que no son finitos se llaman *infinitos.*

El cálculo del cardinal de un conjunto es lo que normalmente se llama "contar". Por ejemplo, un conjunto $A = \{a, b, c\}$, donde a, b y c son distintos dos a dos cumple $|A| = 3$, porque podemos definir $f : I_3 \longrightarrow A$ biyectiva mediante $f(1) = a$, $f(2) = b$, $f(3) = c$.

De este modo, por la propia definición vemos que $|I_n| = n$. En particular $|\varnothing| = 0$ y, de hecho, $\varnothing$ es el único conjunto de cardinal 0.

El resultado fundamental sobre cardinales es el siguiente:

Teorema 1.25 *Dos conjuntos finitos son equipotentes si y sólo si tienen el mismo cardinal. Todo conjunto equipotente a un conjunto finito es finito.*

Demostración: Si X es un conjunto finito y $X \sim Y$, entonces existe un n tal que $I_n \sim X \sim Y$, luego $I_n \sim Y$, luego Y es finito y $|Y| = |X| = n$. Si $|X| = |Y| = n$, entonces $X \sim I_n \sim Y$, luego $X \sim Y$. ∎

Otro hecho básico es el siguiente:

Teorema 1.26 *Si X es un conjunto finito e $Y \subset X$, entonces Y es finito, $|Y| \leq |X|$ y se da la igualdad $|Y| = |X|$ si y sólo si $Y = X$.*

Demostración: Sea X un conjunto finito y supongamos que existe un $x \in X$. Sea $n = |X|$ y sea $f : I_n \longrightarrow X$ biyectiva. Notemos que $n \neq 0$, pues estamos suponiendo que $X \neq \varnothing$. Por lo tanto $n = n' + 1$.

Podemos suponer que $f(n) = x$, pues si la antiimagen de x es un $m < n$, podemos considerar la aplicación $g : I_n \longrightarrow X$ que coincide con f salvo que $g(n) = f(m)$ y $g(m) = f(n)$. Es fácil ver que g sigue siendo biyectiva, pero así se restringe a una biyección $I_{n'} \longrightarrow X \setminus \{x\}$. Por lo tanto, $X \setminus \{x\}$ es finito y $|X \setminus \{x\}| = |X| - 1$.

Pasemos ya a probar el enunciado, por inducción sobre $|X|$. Si $|X| = 0$ entonces $X = \varnothing$, luego necesariamente $Y = \varnothing$ y $|X| = |Y| = 0$.

Si vale para conjuntos de cardinal n y $|X| = n+1$, distinguimos dos casos: si $Y = X$ entonces trivialmente Y es finito y tiene el mismo cardinal que X. En caso contrario existe un $x \in X \setminus Y$, con lo que $Y \subset X \setminus \{x\}$, que según hemos probado es un conjunto finito de cardinal n. Por la hipótesis de inducción Y es finito y $|Y| \leq |X \setminus \{x\}| = n < n+1 = |X|$. ∎

Por ejemplo, ahora podemos probar:

Teorema 1.27 $\mathbb{N}$ *es infinito.*

Demostración: Si $\mathbb{N}$ fuera finito, tendría un cardinal $n \in \mathbb{N}$, pero eso es imposible, porque $I_{n+1} \subset \mathbb{N}$ y, según el teorema anterior, se cumpliría que $n+1 = |I_{n+1}| \leq |\mathbb{N}| = n$, contradicción. ∎

Para probar que un conjunto X tiene cardinal menor o igual que otro conjunto Y no hace falta que $X \subset Y$, sino que basta con que exista una aplicación inyectiva de X en Y. Más en general:

Teorema 1.28 *Sea $X \neq \varnothing$ un conjunto arbitrario e Y un conjunto finito. Las afirmaciones siguientes son equivalentes:*[13]

1. *X es finito y $|X| \leq |Y|$.*
2. *Existe una aplicación $f : X \longrightarrow Y$ inyectiva.*
3. *Existe una aplicación $g : Y \longrightarrow X$ suprayectiva.*

Demostración: $1 \Rightarrow 2$. Sea $|X| = n$, $|Y| = m$. Basta componer una biyección $X \longrightarrow I_n$ con la inclusión $I_n \longrightarrow I_m$ con una biyección $I_m \longrightarrow Y$. El resultado es una aplicación $f : X \longrightarrow Y$ inyectiva.

$2 \Rightarrow 1$. Tenemos que $f[X] \subset Y$, luego $f[X]$ es finito y $|f[X]| \leq |Y|$. Además $f : X \longrightarrow f[X]$ biyectiva, luego X es finito y $|X| = |f[X]| \leq |Y|$.

$2 \Rightarrow 3$. Fijemos $x_0 \in X$ y definimos $g : Y \longrightarrow X$ mediante

$$g(y) = \begin{cases} f^{-1}(y) & \text{si } y \in f[X], \\ x_0 & \text{si } y \in Y \setminus f[X]. \end{cases}$$

Es inmediato comprobar que g es suprayectiva, pues cada $x \in X$ tiene a $f(x)$ por antiimagen.

$3 \Rightarrow 2$. Sea $|Y| = n$ y fijemos una aplicación $h : I_n \longrightarrow Y$ biyectiva. Entonces $g' = h \circ g : I_n \longrightarrow X$ es suprayectiva. Sea $f' : X \longrightarrow I_n$ la aplicación dada por $f'(x) = \min g'^{-1}[x]$. Es fácil ver que es inyectiva, al igual que lo es $f = f' \circ h : X \longrightarrow Y$. ∎

Otro resultado notable sobre aplicaciones entre conjuntos finitos es el siguiente:

[13] El hecho de que la negación de 1 implique la negación de 2 se conoce como "principio del palomar", pues se ilustra con este ejemplo: si en un palomar hay más palomos que nidos, tiene que haber al menos un nido con más de un palomo (porque la aplicación que a cada palomo le asigna su nido no puede ser inyectiva).

Teorema 1.29 *Sea $f : X \longrightarrow Y$ una aplicación entre dos conjuntos finitos del mismo cardinal. Entonces f es inyectiva si y sólo si f es suprayectiva si y sólo si f es biyectiva.*

DEMOSTRACIÓN: Si f es inyectiva, entonces $f : X \longrightarrow f[X]$ biyectiva, luego $|f[X]| = |X| = |Y|$, luego $f[X] = Y$ por 1.26, luego f es suprayectiva.

Si f es suprayectiva pero no inyectiva, existen x, $x' \in X$ distintos tales que $f(x) = f(x')$, pero entonces $f|_{X\setminus\{x\}} : X \setminus \{x\} \longrightarrow Y$ sigue siendo suprayectiva, luego $|Y| \leq |X \setminus \{x\}| < |X| = |Y|$, contradicción. ∎

Hemos definido la suma de números naturales mediante una relación recurrente que la caracteriza. Ahora podemos demostrar que la suma se corresponde con la idea que todos tenemos de ella: $m+n$ es el número de cosas que tenemos cuando juntamos m cosas a otras n cosas.

Teorema 1.30 *Si X e Y son conjuntos finitos disjuntos, entonces $X \cup Y$ es finito, y $|X \cup Y| = |X| + |Y|$.*

DEMOSTRACIÓN: Sean $f : I_m \longrightarrow X$, $g : I_n \longrightarrow Y$ biyectivas. Entonces podemos definir $h : I_{m+n} \longrightarrow X \cup Y$ mediante

$$h(u) = \begin{cases} f(u) & \text{si } 1 \leq u \leq m, \\ g(u-m) & \text{si } m+1 \leq u \leq m+n. \end{cases}$$

Notemos que si $m+1 \leq u \leq m+n$, entonces $1 \leq u-m \leq n$, luego $u-m \in I_n$. Es fácil ver que h es biyectiva, luego $X \cup Y$ es finito y $|X \cup Y| = m+n = |X|+|Y|$. ∎

Ahora probamos un resultado más general:

Teorema 1.31 *Si X e Y son conjuntos finitos, entonces*

$$|X \cup Y| = |X| + |Y| - |X \cap Y|.$$

DEMOSTRACIÓN: Basta tener en cuenta que $X \cup Y = X \cup (Y \setminus (X \cap Y))$, donde X e $Y \setminus (X \cap Y)$ son disjuntos. Por el teorema anterior

$$|X \cup Y| = |X| + |Y \setminus (X \cap Y)|.$$

Por otra parte $Y = (X \cap Y) \cup (Y \setminus (X \cap Y))$ y los dos conjuntos son disjuntos, luego nuevamente por el teorema anterior

$$|Y| = |X \cap Y| + |Y \setminus (X \cap Y)|.$$

Por lo tanto $|Y \setminus (X \cap Y)| = |Y| - |X \cap Y|$, luego sustituyendo en la primera ecuación obtenemos la fórmula del enunciado. ∎

Ahora interpretamos el producto:

Teorema 1.32 *Si X e Y son conjuntos finitos, entonces $X \times Y$ es finito y $|X \times Y| = |X||Y|$.*

DEMOSTRACIÓN: Lo probamos por inducción sobre $|Y|$. Si $|Y| = 0$ entonces $Y = \varnothing$, y es claro que $X \times Y = \varnothing$, luego es finito y $|X \times Y| = 0 = |X||Y|$.

Si vale cuando $|Y| = n$ y suponemos que $|Y| = n + 1$, tomamos $y \in Y$, de modo que $|Y \setminus \{y\}| = n$, y por hipótesis de inducción $X \times (Y \setminus \{y\})$ es finito y $|X \times (Y \setminus \{y\})| = |X| \cdot n$. Ahora basta observar que

$$X \times Y = (X \times (Y \setminus \{y\}) \cup (X \times \{y\}),$$

que la unión es disjunta y que $X \times \{y\} \sim X$ (a través de la biyección $(x, y) \mapsto x$), luego $X \times \{y\}$ es finito y $|X \times \{y\}| = |X|$. Por el teorema 1.30 concluimos que $X \times Y$ es finito y que

$$|X \times Y| = |X| \cdot n + |X| = |X|(n + 1) = |X||Y|.$$

∎

Vamos a contar algunos conjuntos más. Observemos que si $f : A \longrightarrow B$ entonces $f \subset A \times B$, luego $f \in \mathcal{P}(A \times B)$. Por lo tanto, si definimos, por especificación,

$$B^A = \{f \in \mathcal{P}(A \times B) \mid f : A \longrightarrow B\},$$

entonces B^A es el conjunto de todas las aplicaciones de A en B. El que se represente con esa notación se debe al teorema siguiente:

Teorema 1.33 *Si A y B son conjuntos finitos, entonces B^A es un conjunto finito y $|B^A| = |B|^{|A|}$.*

DEMOSTRACIÓN: Por inducción sobre $|A|$. Si $|A| = 0$, entonces $A = \varnothing$, y "técnicamente" existe una aplicación de A en B, porque $\varnothing : \varnothing \longrightarrow B$ es cierto, si nos ajustamos a la definición de aplicación. Por lo tanto $B^A = \{\varnothing\}$ y es un conjunto finito de cardinal $|B^A| = 1 = |B|^0 = |B|^{|A|}$.

Supongamos que el resultado es cierto cuando $|A| = n$ y supongamos que $|A| = n + 1$. Sea $a \in A$ y sea $A' = A \setminus \{a\}$, que es un conjunto finito de cardinal n. Por hipótesis de inducción $B^{A'}$ es finito y $|B^{A'}| = |B|^n$.

Definimos $f : B^A \longrightarrow B^{A'} \times B$ mediante $f(g) = (g|_{A'}, g(a))$. Es fácil ver que f es biyectiva. Por lo tanto $B^A \sim B^{A'} \times B$, y el teorema anterior nos da que B^A es finito y $|B^A| = |B^{A'}||B| = |B|^n|B| = |B|^{n+1} = |B|^{|A|}$. ∎

Teorema 1.34 *Si X es un conjunto finito, $\mathcal{P}X$ es finito y $|\mathcal{P}X| = 2^{|X|}$.*

DEMOSTRACIÓN: Basta observar que $f : \{0, 1\}^X \longrightarrow \mathcal{P}X$ definida por $f(g) = g^{-1}[1]$ es biyectiva, y luego aplicar el teorema anterior. ∎

Terminamos con otra propiedad importante de los conjuntos finitos:

Teorema 1.35 *Si A es un conjunto finito no vacío y $\leq$ es cualquier relación de orden total sobre A, entonces A tiene máximo y mínimo elemento.*

DEMOSTRACIÓN: Sea $|A| = n > 0$ y sea $f : A \longrightarrow I_n$ biyectiva. Supongamos que A no tiene máximo elemento (para el caso del mínimo se razona análogamente). Eso significa que, para todo $x \in A$, existe un $y \in A$ tal que $x < y$. De entre todos los y que cumplen esto, podemos elegir aquel para el que $f(y)$ es mínimo. Esto nos da un criterio para definir una aplicación $g : A \longrightarrow A$ con la propiedad de que $x < g(x)$, para todo $x \in A$. Podemos aplicar el principio de recursión, que nos da una aplicación $h : \mathbb{N} \longrightarrow A$ tal que $h(0)$ es cualquier elemento prefijado de A y $h(n+1) = g(h(n))$.

Observemos que si $m < n$, entonces $h(m) < h(n)$. En efecto, lo razonamos por inducción sobre n. Si $n = 0$ no se cumple $m < 0$, luego la implicación se cumple trivialmente. Si vale para n y suponemos que $m < n + 1$, entonces $m \leq n$. Si $m = n$ entonces $h(m) < g(h(m)) = h(m+1) = h(n+1)$, como había que probar, y si $m < n$ entonces por hipótesis de inducción

$$h(m) < h(n) < g(h(n)) = h(n+1)$$

y concluimos igualmente. Pero esto significa que h es inyectiva, con lo que podemos afirmar que $\mathbb{N}$ es finito, y así tenemos una contradicción. ∎

En particular, si $A \subset \mathbb{N}$ no es vacío y consideramos en A el orden usual de los números naturales, sabemos que A tiene mínimo elemento, y ahora podemos afirmar que tiene máximo si y sólo si es finito, pues si es finito tiene máximo por el teorema anterior, y si tiene máximo m, entonces $A \subset I_m \cup \{0\}$, luego es finito.

1.7 Productos cartesianos

Hemos definido los conceptos de par ordenado y de producto cartesiano de dos conjuntos, pero conviene dar una definición alternativa que se generalice fácilmente a familias de más de dos conjuntos. Para ello conviene introducir antes algunos convenios de notación muy habituales:

A veces, en lugar de expresar mediante $X : I \longrightarrow A$ que X es una aplicación de un conjunto I en otro conjunto A, escribiremos $\{X_i\}_{i\in I}$, y escribiremos X_i en lugar de $X(i)$.

Cuando empleamos esta notación tenemos que distinguir entre $\{X_i\}_{i\in I}$, que es una aplicación, y $\{X_i \mid i \in I\} = X[I]$, que es el conjunto formado por todos los conjuntos X_i, es decir, la imagen de la aplicación.

En el caso particular en que $I = I_n$ es frecuente usar notaciones alternativas, como $\{X_i\}_{i=1}^n$ o $(X_i)_{i=1}^n$ o incluso, con puntos suspensivos: $(X_1, \ldots, X_n)$. Los objetos de esta forma se llaman —según el contexto— *n-tuplas* o *sucesiones finitas* de conjuntos[14]. En definitiva, lo que estamos diciendo es que el objeto representado por $(X_1, \ldots, X_n)$ o por $\{X_i\}_{i=1}^n$ no es más que una aplicación con dominio I_n y rango en un conjunto que no especificamos, que a cada número i le asigna una imagen $X(i) = X_i$.

[14]para $n = 2$ son pares, para $n = 3$ son ternas, para $n = 4$ son cuádruplas, para $n = 5$ quíntuplas y, en general, n-tuplas.

Por ejemplo, la quíntupla $x = (4, 5, 5, 2, 0)$ es una aplicación $x : I_5 \longrightarrow \mathbb{N}$ tal que $x_1 = 4$, $x_2 = 5$, $x_3 = 5$, $x_4 = 2$, $x_5 = 0$. Técnicamente es el conjunto

$$x = \{(1, 4), (2, 5), (3, 5), (4, 2), (5, 0)\},$$

pero no ganamos nada representándolo así.

Cuando el conjunto de índices es $I = \mathbb{N}$ o incluso $I = \{n \in \mathbb{N} \mid n \geq n_0\}$, se habla de *sucesiones (infinitas)* de conjuntos y es frecuente emplear la notación alternativa $\{X_n\}_{n=0}^{\infty}$ o $\{X_n\}_{n=n_0}^{\infty}$.

Volviendo al caso general, a menudo se dice que $\{X_i\}_{i\in I}$ es una *familia*[15] *(indexada) de conjuntos.* Recordemos que $X : I \longrightarrow B$, para cierto conjunto B. Entonces definimos el *producto cartesiano* de $\{X_i\}_{i\in I}$ como el conjunto

$$\prod_{i\in I} X_i = \{x \in (\bigcup B)^I \mid \text{para cada } i \in I \text{ se cumple } x(i) \in X_i\}.$$

Con la notación que estamos introduciendo, los elementos $x \in \prod_{i\in I} X_i$ son de la forma $x = \{x_i\}_{i\in I} = (x_i)_{i\in I}$, es decir, familias de conjuntos (que son como n-tuplas generalizadas, no necesariamente finitas) tales que $x_i \in X_i$.

En el caso particular en que todos los conjuntos X_i coinciden con un mismo conjunto X, entonces el producto cartesiano coincide con el conjunto que ya hemos definido como X^I, es decir, el conjunto de todas las aplicaciones de I en X. Sin embargo, con la notación que estamos introduciendo, cada $x \in X^I$ lo representaremos de la forma $x = (x_i)_{i\in I}$.

En el caso de una familia $\{X_i\}_{i=1}^n$ (es decir, cuando $I = I_n$), representaremos también el producto cartesiano con las notaciones alternativas

$$\prod_{i=1}^{n} X_i = X_1 \times \cdots \times X_n,$$

y sus elementos son las n-tuplas $(x_1, \ldots, x_n)$ tales que $x_i \in X_i$.

En el caso de los productos con $I = I_n$ y con todos los factores iguales a un mismo conjunto X, escribiremos

$$X^n = \prod_{i=1}^{n} X = \overbrace{X \times \cdots \times X}^{n \text{ veces}},$$

cuyos elementos son n-tuplas $(x_1, \ldots, x_n)$ con cada $x_i \in X$.

[15] También es frecuente llamar familias (no indexadas) de conjuntos a cualquier conjunto cuyos elementos sean conjuntos. Según los convenios que estamos adoptando todos los conjuntos están formados por conjuntos, por lo que la diferencia entre un conjunto y una familia de conjuntos es puramente psicológica: al decir "familia de conjuntos" estamos indicando que nos interesa ver a sus elementos como conjuntos. Por ejemplo, nadie llama a $\mathbb{N}$ familia de conjuntos, porque, aunque los números naturales son conjuntos, eso es algo que en la práctica "se olvida", mientras que $\mathcal{P}\mathbb{N}$ sí se considera una familia de conjuntos, porque sus elementos, no sólo son conjuntos, sino que además se conciben como tales.

De este modo, ahora tenemos dos interpretaciones posibles de $(3,5) \in \mathbb{N}\times\mathbb{N}$. Podemos considerar que $\mathbb{N}\times\mathbb{N}$ es el producto cartesiano definido originalmente, en cuyo caso $(3,5) = \{\{3\},\{3,5\}\}$, o bien que es el que acabamos de definir, en cuyo caso $(3,5) = \{(1,3),(2,5)\}$. Sin embargo, resulta totalmente irrelevante si nos referimos a uno o a otro, pues la naturaleza conjuntista del par $(3,5)$ no tiene importancia. Lo único que importa es que $(3,5)$ es un conjunto que determina unívocamente una primera componente 3 y una segunda componente 5. Podemos adoptar el convenio de que, a partir de ahora, todos los productos cartesianos que consideremos lo son en el nuevo sentido que acabamos de introducir, pero eso no tiene ninguna traducción sensible en la práctica.

Notemos que A^1 es casi lo mismo que A, pero no exactamente. Está formado por todas las aplicaciones de $\{1\}$ en A, luego $A^1 = \{\{(1,a)\} \mid a \in A\}$. Así la aplicación $A \longrightarrow A^1$ dada por $a \mapsto (1,a)$ es biyectiva y hace que todo lo que digamos para A tenga una traducción obvia a A^1 y viceversa.

En general, cada producto cartesiano $\prod_{i\in I} X_i$ tiene asociadas las *proyecciones,* que son las aplicaciones $p_i : \prod_{i\in I} X_i \longrightarrow X_i$ dadas por $p_i(x) = x_i$. Claramente son suprayectivas.

Terminamos esta sección con algunos resultados de combinatoria[16]:

Si $\{X_i\}_{i\in I}$ es una familia de conjuntos, podemos definir

$$\bigcup_{i\in I} X_i = \bigcup\{X_i \mid i \in I\}, \qquad \bigcap_{i\in I} X_i = \bigcap\{X_i \mid i \in I\}$$

(la segunda definición requiere que $I \neq \varnothing$). Cuando $I = I_n$ usaremos también las notaciones alternativas

$$\bigcup_{i=1}^{n} X_i = X_1 \cup \cdots \cup X_n, \quad \bigcap_{i=1}^{n} X_i = X_1 \cap \cdots \cap X_n.$$

Teorema 1.36 *Toda unión finita de conjuntos finitos es finita.*

DEMOSTRACIÓN: Por inducción sobre el número de conjuntos que unimos. Si tenemos una familia $\{X_i\}_{i=1}^{1}$, entonces la unión es X_1 que es finito por hipótesis. Si vale para uniones de n conjuntos, entonces

$$\bigcup_{i=1}^{n+1} X_i = \bigcup_{i=1}^{n} X_i \cup X_{n+1},$$

que es una unión de dos conjuntos finitos, el primero por hipótesis de inducción y el segundo por hipótesis. Por lo tanto, es finita. ∎

Similarmente se prueba que todo producto cartesiano de un número finito de conjuntos finitos es finito.

Definición 1.37 *Si A es un conjunto, llamaremos $\Sigma_A \subset A^A$ al conjunto de todas las aplicaciones $f : A \longrightarrow A$ biyectivas.*

[16]Se llama combinatoria al cálculo de cardinales de conjuntos finitos.

Ya hemos probado que si A es un conjunto finito con $|A| = n$, entonces $|A^A| = n^n$. Ahora vamos a calcular el cardinal de Σ_A.

Teorema 1.38 *Si A es un conjunto finito y $|A| = n$, entonces $|\Sigma_A| = n!$*

Demostración: Si $f : A \longrightarrow B$ es biyectiva, podemos definir una aplicación $F : \Sigma_A \longrightarrow \Sigma_B$ biyectiva mediante $F(x) = f^{-1} \circ x \circ f$, por lo que $|\Sigma_A| = |\Sigma_B|$. Esto hace que no perdamos generalidad si suponemos $A = I_n$. Escribiremos Σ_n en lugar de Σ_{I_n}.

Notemos que Σ_n es el conjunto de todas las n-tuplas $(m_1, \ldots, m_n)$ cuyas componentes son distintas dos a dos.

Ahora observamos que podemos definir $H : \Sigma_{n+1} \longrightarrow I_{n+1} \times \Sigma_n$ como la aplicación que a cada $n+1$-tupla $(m_1, \ldots, m_{n+1})$ con $m_i = n + 1$ le asigna el par formado por i como primera componente y la n-tupla que resulta de suprimir m_i en $(m_1, \ldots, m_{n+1})$. Una comprobación rutinaria muestra que H es biyectiva, luego $|\Sigma_{n+1}| = |I_{n+1} \times \Sigma_n| = (n + 1)|\Sigma_n|$.

Ahora basta razonar que $|\Sigma_n| = n!$ por inducción. Si $n = 0$ entonces, por las sutilezas de la lógica, es $\Sigma_0 = \{\varnothing\}$, luego $|\Sigma_0| = 1 = 0!$ Si se cumple para n, entonces $|\Sigma_{n+1}| = |\Sigma_n|(n + 1) = n!(n + 1) = (n + 1)!$ ∎

1.8 Relaciones de equivalencia

Terminamos este capítulo introduciendo otro concepto conjuntista básico, que hasta ahora no hemos necesitado, pero que aparecerá constantemente en los temas siguientes.

Definición 1.39 Una *relación de equivalencia* en un conjunto A es una relación R en A que es reflexiva, simétrica y transitiva.

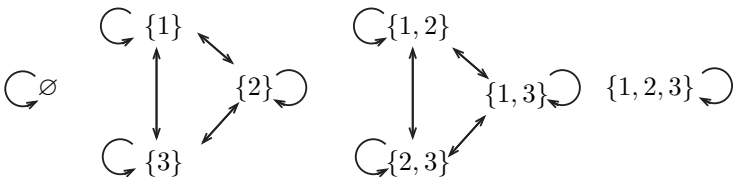
Recordemos que, explícitamente esto significa que todo $a \in A$ cumple $a\,R\,a$, si $a\,R\,b$ entonces también $b\,R\,a$ y si $a\,R\,b$ y $b\,R\,c$ entonces $a\,R\,c$.

Ejemplo Consideremos $X = \{1, 2, 3\}$, $A = \mathcal{P}X$ y R la relación en A dada por

$$U\,R\,V \quad \text{si y sólo si} \quad |U| = |V|.$$

Es inmediato que se trata de una relación de equivalencia, pues todo conjunto tiene el mismo cardinal que él mismo, si U tiene el mismo cardinal que V entonces V tiene el mismo cardinal que U, y si U tiene el mismo cardinal que V y V tiene el mismo cardinal que W entonces U tiene el mismo cardinal que W.

El diagrama siguiente nos permite formarnos una imagen de la relación:



Cada flecha $a \to b$ indica que $a\,R\,b$. El hecho de que la relación sea reflexiva se traduce en que cada elemento tiene una flecha en forma de "rizo", el hecho de que sea simétrica se traduce en que todas las flechas son dobles $\leftrightarrow$, y el hecho de que sea transitiva se traduce en que los ocho elementos de A se agrupan en bloques totalmente conectados. Esos bloques son las clases de equivalencia que vamos a definir a continuación. ∎

Definición 1.40 Sea R una relación de equivalencia en un conjunto A. Para cada $a \in A$ la *clase de equivalencia* de a es el conjunto

$$[a]_R = \{x \in A \mid x\,R\,a\}$$

de todos los elementos de A relacionados con a. El conjunto de todas las clases de equivalencia[17] de A respecto de R se llama *conjunto cociente* y se representa por A/R. La aplicación $p : A \longrightarrow A/R$ dada por $p(a) = [a]_R$ se llama *proyección canónica* y claramente es suprayectiva.

Así, en el ejemplo precedente vemos que hay cuatro clases de equivalencia, por lo que el cociente A/R tiene cuatro elementos, uno correspondiente a cada cardinal posible de un subconjunto de X. También podemos constatar en él los hechos generales que recoge el teorema siguiente:

Teorema 1.41 *Sea R una relación de equivalencia en un conjunto A.*

1. *Si $a \in A$, entonces $a \in [a]_R$.*
2. *Si $a, b \in A$, entonces $a\,R\,b$ si y sólo si $[a]_R = [b]_R$.*
3. *Si $a, b \in A$, entonces no $a\,R\,b$ si y sólo si $[a]_R \cap [b]_R = \varnothing$.*

Por lo tanto, A/R es una partición de A, es decir, una familia de subconjuntos de A no vacíos y disjuntos dos a dos cuya unión es A.

Demostración: 1) es inmediato, porque $a\,R\,a$.

2) Si $a\,R\,b$, probamos la igualdad de sus clases por doble inclusión: si se cumple $x \in [a]_R$, entonces $x\,R\,a$, luego $x\,R\,b$ por transitividad, luego $x \in [b]_R$. La otra inclusión se prueba igual, teniendo en cuenta que $b\,R\,a$ por simetría. Recíprocamente, si $[a]_R = [b]_R$, entonces $a \in [a]_R = [b]_R$, luego $a\,R\,b$.

3) Si $[a]_R \cap [b]_R \neq \varnothing$, existe $x \in [a]_R \cap [b]_R$, luego $x\,R\,a$ y $x\,R\,b$, luego por simetría y transitividad $a\,R\,b$. Recíprocamente, si $a\,R\,b$, entonces $a \in [a]_R \cap [b]_R$, luego $[a]_R \cap [b]_R \neq \varnothing$. ∎

La notación $[a]_R$ que hemos introducido para las clases de equivalencia es genérica, pero a menudo introduciremos notación específica para cada relación que consideremos.

[17] Notemos que dicho conjunto se puede definir por especificación, pues se trata de

$$A/R = \{X \in \mathcal{P}A \mid \text{ existe un } a \in A \text{ tal que } X = [a]_R\}.$$

La idea subyacente al concepto de clases de equivalencia y conjunto cociente es que formalizan un proceso de abstracción: si en un conjunto de bolas de colores consideramos la relación de equivalencia "ser del mismo color", cada clase de equivalencia contendrá a las bolas de un mismo color, y el conjunto cociente tendrá un elemento por cada color que puedan tener las bolas, igual que en el ejemplo anterior el cociente tenía un elemento por cada cardinal posible que pueden tener los conjuntos considerados.

Capítulo II

Anillos

En el capítulo anterior, a la vez que hemos introducido el lenguaje conjuntista básico, hemos demostrado muchos hechos que en [ITAl] dimos por evidentes, como las propiedades de los números naturales, o las de los conjuntos finitos, etc. Ahora vamos a introducir —conectándolos con la base conjuntista expuesta en el capítulo anterior— algunos de los conceptos básicos que estudiamos en el capítulo I de [ITAl], entre ellos los números enteros y racionales y los polinomios. Los números reales los introduciremos en [G 2.6] y [An 1.44].

Más en general, vamos a estudiar algunas "estructuras algebraicas", es decir, conjuntos dotados de ciertas funciones y/o relaciones que satisfacen ciertos axiomas. En el capítulo anterior ya nos hemos encontrado con una de ellas, la de conjunto ordenado, que hemos definido como un par $(A, \leq)$ formado por un conjunto A y una relación de orden en A, y en [ITAl] presentamos someramente algunas más, como las estructuras de "anillo", "cuerpo", etc., que aquí analizaremos con más detalle.

2.1 Leyes de composición interna

En esta primera sección daremos una definición general de "operación" en un conjunto, que generalice a la suma y el producto de números naturales, que son las únicas operaciones que hemos introducido hasta ahora.

Definición 2.1 Una *ley de composición interna* o una *operación interna* en un conjunto A es una aplicación $* : A \times A \longrightarrow A$. En la práctica, en lugar de escribir $*((a, b))$ para representar la imagen del par (a, b) por la aplicación $*$, usaremos la notación $a * b$.

En definitiva, una ley de composición interna en un conjunto A es cualquier criterio que nos permita operar dos cualesquiera de sus elementos a y b para obtener un nuevo elemento $a * b$ de A.

Al final de la sección 1.4 hemos introducido la suma y el producto de números naturales. Técnicamente, hemos definido la suma $m + n$ como $f_m(n)$, donde

$f_m : \mathbb{N} \longrightarrow \mathbb{N}$ es una función definida por recurrencia en términos de m. Ahora podemos definir $+$ como la aplicación que a cada par $(m, n) \in \mathbb{N} \times \mathbb{N}$ le asigna esta suma $m+n$, y así $+$ pasa a tener un "significado autónomo", concretamente el de una ley de composición interna en $\mathbb{N}$. Lo mismo se aplica al producto.

Veamos algunas definiciones generales. Fijemos un conjunto A con una ley de composición interna $*$.

1. Se dice que $*$ es *asociativa* si cumple

$$(a * b) * c = a * (b * c),$$

   para todos los elementos a, b, $c \in A$.

2. Se dice que $*$ es *conmutativa* si cumple $a*b = b*a$ para todos los elementos a, $b \in A$.

3. Un *elemento neutro* para $*$ es un elemento $e \in A$ tal que $a * e = e * a = a$, para todo $a \in A$.

   Observemos que una operación $*$ no puede tener más de un elemento neutro, pues si tiene dos, digamos e y e', de la definición se sigue que $e = e * e' = e'$.

4. Si A tiene elemento neutro e, un *elemento opuesto* para un $a \in A$ es un $b \in A$ tal que $a * b = b * a = e$.

De todas las propiedades que puede tener una ley de composición interna, la asociativa va a ser para nosotros "irrenunciable", es decir, no vamos a estudiar ninguna operación que no la posea. Esto nos lleva a introducir una de las estructuras algebraicas más elementales:

Definición 2.2 Un *semigrupo* es un par $(A, *)$, donde A es un conjunto y $*$ es una ley de composición interna en A que cumple la propiedad asociativa. Si además cumple la propiedad conmutativa diremos que $(A, *)$ es un *semigrupo conmutativo.*

Así, estudiar los semigrupos es estudiar las propiedades disponibles siempre que contemos con una ley asociativa, sea ésta cual sea.

Por ejemplo, una propiedad elemental es que si $(A, *)$ es un semigrupo con elemento neutro e y $a \in A$ tiene opuesto, entonces dicho opuesto es único. En efecto, si tiene dos opuestos b y b', se cumple que

$$b = b * e = b * (a * b') = (b * a) * b' = e * b' = b'.$$

Vamos a introducir dos convenios habituales de notación para las leyes de composición interna:

Notación multiplicativa Usaremos $\cdot$ para representar cualquier ley de composición interna asociativa en un conjunto dado A. En tal caso, si la operación tiene elemento neutro, lo representaremos por 1, y si un $a \in A$ tiene elemento opuesto, lo representaremos por a^{-1} y lo llamaremos también *elemento inverso*.

Notación aditiva Usaremos $+$ para representar cualquier ley de composición interna asociativa y conmutativa en un conjunto dado A. En tal caso, si la operación tiene elemento neutro, lo representaremos por 0, y si un $a \in A$ tiene elemento opuesto, lo representaremos por $-a$ y lo llamaremos también *elemento simétrico*. También adoptaremos el convenio de escribir $a-b$ en lugar de $a+(-b)$.

Hay que entender que la notación multiplicativa la aplicaremos tanto para leyes conmutativas como no conmutativas, mientras que la notación aditiva sólo la usaremos para leyes conmutativas.

Otro convenio habitual es el de omitir las leyes de composición (u otras relaciones o funciones asociadas) al nombrar una estructura algebraica, tal y como ya hemos establecido para el caso de los conjuntos ordenados. Así, hablaremos de un semigrupo A en lugar de escribir $(A, \cdot)$ o $(A, +)$, dejando al contexto si usamos la notación aditiva o multiplicativa. En otras palabras, cuando decimos que A es un semigrupo, hay que entender que en realidad nos estamos refiriendo a un par $(A, +)$ o $(A, \cdot)$.

Veamos otro ejemplo en el que difieren las notaciones multiplicativa y aditiva: para cada $a \in A$ en un semigrupo con elemento neutro y cada $n \in \mathbb{N}$ podemos definir a^n (si la notación es multiplicativa) o na (si es aditiva) mediante las relaciones recurrentes:

$$a^0 = 1, \quad a^{n+1} = a^n \cdot a, \qquad\qquad 0 \cdot a = 0, \quad (n+1)a = na + a.$$

Notemos que no hemos definido dos operaciones, sino una sola, escrita con dos notaciones distintas. Dejamos como ejercicio demostrar por inducción sus propiedades básicas —que expresamos con notación multiplicativa— y traducirlas a notación aditiva.

$$a^{m+n} = a^m a^n, \qquad (a^m)^n = a^{mn}, \qquad (ab)^n = a^n b^n.$$

La tercera sólo es válida si $ab = ba$ (de donde, por inducción, se prueba como paso previo que $b^n a = ab^n$).

También conviene observar que al particularizar estas definiciones a los semigrupos $(\mathbb{N}, +)$ y $(\mathbb{N}, \cdot)$ obtenemos las definiciones usuales de producto y exponenciación de números naturales.

Una potencia a^n es el resultado de multiplicar a por sí mismo n veces. Ahora vamos a definir el producto de n elementos de un semigrupo, no necesariamente iguales entre sí.

Definición 2.3 Sea $(A,\cdot)$ un semigrupo con elemento neutro y sea $\{a_i\}_{i=0}^n$ una sucesión de elementos de A. Definimos por recurrencia,[1]

$$\prod_{i<0} a_i = 1, \qquad \prod_{i<j+1} a_i = (\prod_{i<j} a_i)\cdot a_j,$$

para $j \leq n$, de modo que en particular tenemos definido el producto finito de los elementos dados:

$$a_0\cdots a_n = \prod_{i=0}^{n} a_i = \prod_{i<n+1} a_i.$$

Ésta es la notación que usaremos cuando empleemos la notación multiplicativa en el semigrupo. Si usamos notación aditiva usaremos el signo de sumatorio $\sum$, es decir, la definición anterior se traduce a

$$\sum_{i<0} a_i = 0, \qquad \sum_{i<j+1} a_i = \sum_{i<j} a_i + a_j,$$

y en particular $a_0 + \cdots + a_n = \sum_{i=0}^{n} a_i = \sum_{i<n+1} a_i$.

Hemos definido productos y sumas finitas con índices que empiezan en 0, pero podemos partir de cualquier número natural si definimos:

$$\prod_{i=m}^{n} a_i = \prod_{i=0}^{n-m} a_{m+i}, \qquad \text{para } m \leq n,$$

y análogamente con notación aditiva.

Las sumas y productos finitos definidos de este modo satisfacen una serie de propiedades "obvias", como que se pueden partir por cualquier punto $0 \leq i \leq n$:

$$a_0\cdots a_n = (a_0\cdots a_i)(a_{i+1}\cdots a_n).$$

Desde un punto de vista técnico, estas "obviedades" se traducen en una serie de tediosos argumentos rutinarios por inducción que el lector puede consultar en el apéndice al final de este capítulo si está interesado en ellos.

Veamos una aplicación de las sumas finitas:

Teorema 2.4 *Si $k > 1$ es un número natural, para cada $m \in \mathbb{N}$ no nulo existe una única sucesión $\{c_i\}_{i=0}^n$ de números menores que k tal que $c_n \neq 0$ y*

$$m = \sum_{i=0}^{n} c_i k^i.$$

[1] Técnicamente estamos aplicando el teorema de recursión de este modo: definimos una función $f : \mathbb{N} \longrightarrow A$ mediante $f(i) = a_i$ si $i \leq n$ y $f(i) = 1$ en otro caso. A su vez, definimos $g : \mathbb{N} \longrightarrow A$ mediante $g(0) = 1$ y $g(j+1) = g(j)\cdot f(j)$, de modo que $g(j+1)$ se define como función de $g(j)$ y de j. Finalmente, $\prod_{i=0}^{n} a_i = g(n+1)$.

DEMOSTRACIÓN: Observemos en primer lugar que $m < k^m$. En efecto, para $m = 0, 1$ es inmediato, y si vale para $m \geq 1$, entonces

$$m + 1 \leq m + m = 2m < k \cdot k^m = k^{m+1}.$$

Por lo tanto, hay un mínimo $n^* \leq m$ tal que $m < k^{n^*}$. No puede ser $n^* = 0$, pues entonces sería $m = 0$, luego $n^* = n + 1$ y se cumple $k^n \leq m < k^{n+1}$. Observemos que la unicidad de n^* implica la de n, es decir, hemos probado que para todo natural $m > 0$ existe un único natural n tal que $k^n \leq m < k^{n+1}$. Llamaremos $o(m)$ a este único n.

Si dividimos $m = k^n c + m'$, con $m' < k^n$, tiene que ser $c < k$, ya que si fuera $c \geq k$, tendríamos $m \geq k^n k = k^{n+1}$. Además $c > 0$, o de lo contrario $m = m' < k^n$. Así pues, todo número natural $m > 0$ puede expresarse en la forma

$$m = ck^n + m', \qquad 0 < c < k, \qquad m' < k^n \leq m.$$

Observemos que esta expresión es única, pues necesariamente $n = o(m)$ (ya que $m < (k-1)k^n + k^n = kk^n = k^{n+1}$) y entonces c y m' están unívocamente determinados como cociente y resto de la división euclídea. Vamos a probar que todo natural m no nulo admite una descomposición como la que indica el enunciado con $n = o(m)$.

Razonamos por inducción sobre m. Supuesto cierto para todo $m' < m$, consideramos la expresión $m = ck^n + m'$. Si $m' = 0$ entonces $m = ck^n$ ya tiene la forma deseada. En caso contrario, por hipótesis de inducción

$$m' = \sum_{i=0}^{n'} c_i k^i,$$

donde $n' = o(m') < n$ (porque $m' < k^n$) y, definiendo $c_i = 0$ para $n' < i < n$ y $c_n = c$, tenemos que

$$m = \sum_{i=0}^{n'} c_i k^i + \sum_{i=n'+1}^{n-1} c_i k^i + c_n k^n = \sum_{i=0}^{n} c_i k^i.$$

Para probar la unicidad observamos que

$$k^n \leq \sum_{i=0}^{n} c_i k^i < k^{n+1}.$$

En efecto, la primera desigualdad se sigue de que $c_n \neq 0$ separando el último sumando, y la segunda se prueba por inducción sobre n. Si vale para n, entonces

$$\sum_{i=0}^{n+1} c_i k^i = \sum_{i=0}^{n} c_i k^i + c_{n+1} k^{n+1} < k^{n+1} + (k-1)k^{n+1} = k^{n+2}.$$

Por lo tanto, razonando por inducción sobre m, si tenemos dos descomposiciones

$$m = \sum_{i=0}^{n} c_i k^i = \sum_{i=0}^{n'} c'_i k^i$$

en las condiciones del enunciado, necesariamente $n = n' = o(m)$, con lo que, por la unicidad de la descomposición

$$c_n k^n + \sum_{i=0}^{n-1} c_i k^i = c'_n k^n + \sum_{i=0}^{n-1} c'_i k^i$$

(notemos que hemos probado que los segundos sumandos son $< k^n$), tenemos que $c_n = c'_n$ y

$$\sum_{i=0}^{n-1} c_i k^i = \sum_{i=0}^{n-1} c'_i k^i.$$

Llamamos $\bar{n}$ y $\bar{n}'$ a los máximos naturales tales que $c_{\bar{n}} \neq 0$ y $c'_{\bar{n}'} \neq 0$, respectivamente, de modo que

$$\sum_{i=0}^{\bar{n}} c_i k^i = \sum_{i=0}^{\bar{n}'} c'_i k^i.$$

Por hipótesis de inducción $\bar{n} = \bar{n}'$ y $c_i = c'_i$ para todo $i < \bar{n}$, lo que implica que las dos sucesiones $\{c_i\}_{i\leq n}$ y $\{c'_i\}_{i\leq n}$ son iguales. ∎

Definición 2.5 La sucesión $\{c_i\}_{i\leq n}$ dada por el teorema anterior se llama *representación en base k* del número m. Es habitual usar la notación

$$c_n \cdots c_{0(k)} = \sum_{i=0}^{n} c_i k^i.$$

Por ejemplo, $232_{(5)} = 2\cdot 5^2 + 3\cdot 5 + 2$. Observemos que, en general, $k = 1\cdot k + 0$, por lo que la representación de k en base k es $k = 10_{(k)}$.

Sin más razón de peso que el hecho de que es el número de dedos que tenemos en las manos, cuando no se especifica la base se entiende que es $k = 9 + 1$, de modo que la notación usual para este valor de k es $k = 10$. Así pues, convenimos en que

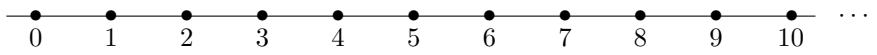
$$c_n \cdots c_0 = \sum_{i=0}^{n} c_i 10^i.$$

De este modo, todo número natural tiene un nombre canónico en términos de las diez cifras $0, \ldots, 9$, aunque la elección del número 10 es puramente arbitraria. En principio, la menor base admisible es $k = 2$, de modo que, por ejemplo, $10 = 2^3 + 2 = 1010_{(2)}$. Por lo tanto, todo número natural admite un nombre canónico en términos únicamente de las cifras 0 y 1.

Con este resultado tenemos ya justificados todos los resultados de la aritmética elemental (la que conocen los niños), salvo que no vamos a describir aquí los algoritmos sobradamente conocidos para sumar, restar, multiplicar y dividir números naturales a través de sus desarrollos decimales.

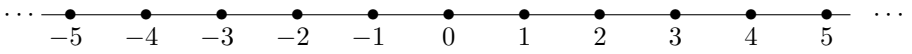
2.2 Los números enteros

Los números naturales y su aritmética tienen una interpretación geométrica intuitiva. Podemos visualizarlos como puntos equidistantes en una recta:



De este modo la relación de orden $m < n$ se interpreta como "m está a la izquierda de n", sumar $2 + 3$ se interpreta como situarse en el punto asignado al 2 y desplazarse 3 lugares hacia la derecha, mientras que restar $5 - 3$ consiste en situarse sobre el 5 y desplazarse tres lugares hacia la izquierda. El hecho de que no se pueda restar $5 - 7$ se interpreta como que no podemos desplazarnos 7 lugares a la izquierda del 5 sin salirnos de los puntos asignados a los números naturales.

Ahora vamos a introducir los números enteros, que extienden a los números naturales de modo que la figura anterior se completa hasta:



Sumar dos números enteros se interpreta como partir del punto asignado al primero y desplazarse tantos lugares como indica el segundo hacia la derecha si éste es positivo o hacia la izquierda si es negativo. Este "paso" de los números naturales a los enteros tiene interés por muchas razones. Señalamos algunas:

- Desde un punto de vista puramente algebraico, obtenemos una estructura más rica. Concretamente, todo número entero tiene un opuesto para la suma, cosa que no le sucede a ningún número natural, salvo el 0.

- Esto tiene a su vez consecuencias aritméticas. Por ejemplo, si tratamos únicamente con números naturales y tenemos una igualdad $a + b = c + d$, no podemos pasar a $a - c = d - b$ a menos que tengamos la garantía de que $a \geq c$ y $d \geq b$, lo cual no tiene por qué suceder *a priori* y puede obligarnos a descomponer el razonamiento en varios casos. En cambio, si trabajamos con números enteros un paso así siempre es válido.

- Por otra parte, aunque es todavía un ejemplo muy elemental, es un primer paso hacia la conceptualización abstracta de la aritmética. Para alguien que sólo conozca los números naturales, la suma y la resta son dos operaciones distintas, consistentes en "moverse" en sentidos opuestos, en añadir en un caso y en sustraer en otro. Sin embargo, para quien concibe a los números naturales como una parte de los números enteros, una resta $m - n$ es lo mismo que la suma $m + (-n)$. Los números enteros permiten concebir como una misma cosa lo que parecían ser dos cosas bastante diferentes.

- Por último, de cara a las aplicaciones geométricas, se trata de un paso más en el camino que nos llevará a asignar un número a cada punto de una recta, lo cual a su vez permite traducir a términos algebraicos los problemas geométricos.

En la sección 1.2 de [ITAl] introdujimos los números enteros como números naturales precedidos de un signo positivo o negativo. Esto captura, ciertamente, la esencia de lo que es un número entero, pero aquí vamos a dar una definición que, siendo formalmente más artificiosa, simplifica drásticamente la comprobación de las propiedades básicas de los números enteros, que se reducen a comprobaciones casi mecánicas.

La idea es que cada número entero puede expresarse como resta de dos números naturales. Por ejemplo, $5 = 7 - 2$ y $-5 = 2 - 7$. En general, cada par $(m, n) \in \mathbb{N}\times\mathbb{N}$ determina el número entero $m-n$ (que aún no tenemos definido), pero podemos pensar en definir el número entero $m-n$ precisamente como el par (m, n). Entonces el conjunto de los números enteros sería técnicamente $\mathbb{N} \times \mathbb{N}$, aunque al final "nos olvidáramos" de ello.

Sin embargo, esto no es viable literalmente, porque pares como $(2, 7)$ y $(10, 15)$ deben "ser" el mismo número entero -5, y dos cosas distintas no pueden ser la misma cosa. Este problema ontológico se resuelve mediante un procedimiento que nos aparecerá de forma recurrente en muchos contextos: cuando queremos definir un objeto como "abstracción" de una propiedad común de varios de ellos —en este caso queremos abstraer de cada par (m, n) la "resta" que determina— definimos la relación de equivalencia que relaciona a dos objetos si y sólo si "deberían" representar al mismo objeto y formamos el conjunto cociente.

En nuestro caso, si nos preguntamos cuándo dos pares (a, b) y (c, d) de números naturales "deberían" dar lugar a la misma resta $a - b = c - d$, la respuesta es cuando $a + d = b + c$. Lo importante es que esta última condición es una igualdad entre números naturales que, al contrario que $a - b = c - d$, no involucra ningún concepto que no tengamos ya definido. En conclusión:

Definición 2.6 Definimos en $\mathbb{N} \times \mathbb{N}$ la relación de equivalencia dada por

$$(a, b)\, R\, (c, d) \qquad \text{si y sólo si} \qquad a + d = b + c.$$

Llamaremos *conjunto de los números enteros* al cociente $\mathbb{Z} = (\mathbb{N} \times \mathbb{N})/R$. La Z es por el alemán *Zahl* (número).

La comprobación de que la relación R es de equivalencia no ofrece ninguna dificultad. Por ejemplo, la transitividad se debe a que si $(a, b)\, R\, (c, d)$ y $(c, d)\, R\, (e, f)$, entonces

$$a + d = b + c, \qquad c + f = d + e,$$

luego sumando miembro a miembro $a+d+c+f = b+c+d+e$ y, simplificando, $a + f = b + e$, luego $(a, b)\, R\, (e, f)$.

De momento representaremos por $[a, b]$ la clase de equivalencia de par (a, b), si bien enseguida encontraremos una notación más conveniente. En principio, como caso particular de 1.41, sabemos que $[a, b] = [c, d]$ si y sólo si $a+d = b+c$.

Si $n \in \mathbb{N}$, definimos

$$+n = [n, 0] \in \mathbb{Z}, \qquad -n = [0, n] \in \mathbb{Z}$$

Llamaremos números enteros *positivos* y *negativos,* respectivamente, a los elementos de los conjuntos

$$\mathbb{Z}^+ = \{+n \mid n \in \mathbb{N}, n > 0\}, \qquad \mathbb{Z}^- = \{-n \mid n \in \mathbb{N}, n > 0\},$$

y aparte definimos $0 = [0, 0] \in \mathbb{Z}$.

Se cumple entonces que $\mathbb{Z} = \mathbb{Z}^- \cup \{0\} \cup \mathbb{Z}^+$ y los tres conjuntos son disjuntos dos a dos. En efecto:

- Todo número entero es de la forma $[m, n]$, con $m,\ n \in \mathbb{N}$. Si $m < n$ entonces $[m, n] = [0, n - m] \in \mathbb{Z}^-$, si $m = n$ entonces $[m, n] = [0, 0]$ y si $m > n$ entonces $[m, n] = [m - n, 0] \in \mathbb{Z}^+$.
- Si $[m, n] \in \mathbb{Z}^+ \cap \mathbb{Z}^-$ entonces $[m, n] = [r, 0] = [0, s]$, para ciertos $r,\ s > 0$, pero entonces $r + s = 0$, lo cual sólo es posible si $r = s = 0$, contradicción. Igualmente, si $[m, 0] = [0, 0]$ es que $m = 0$, luego ningún número de $\mathbb{Z}^+$ es igual a 0, e igualmente con $\mathbb{Z}^-$.

En otras palabras, todo número entero es positivo, negativo o cero, y no puede ser dos cosas a la vez. Por consiguiente, $\mathbb{Z}$ consta exactamente de los elementos

$$0, +1, -1, +2, -2, +3, -3, \ldots$$

y es importante que en esta enumeración no aparecen términos repetidos, es decir, que sólo puede suceder que $+n = +m$ o que $-n = -m$ si $m = n$ (por ejemplo, $+n = +m$ equivale a $[n, 0] = [m, 0]$, y a su vez a $n = m$).

Ordenación de los números enteros Definimos en $\mathbb{Z}$ la relación dada por[2]

$$[a, b] \leq [c, d] \text{ si y sólo si } a + d \leq b + c.$$

No es inmediato que esto sea una definición válida. En general, cuando definimos una relación o una función sobre las clases de equivalencia de un conjunto cociente, es necesario comprobar que la definición no depende de la elección de los representantes de las clases. Concretamente, si $[a, b] = [a', b']$ y $[c, d] = [c', d']$, sería incoherente que usando la primera representación obtuviéramos que $[a, b] \leq [c, d]$, mientras que usando la segunda obtuviéramos $[a', b'] \not\leq [c', d']$.

Pero esto no sucede: tenemos que $a + b' = b + a'$ y $c + d' = d + c'$. Si además $a + d \leq b + c$, entonces, sumando todo, $a + d + b + a' + c + d' \leq b + c + a + b' + d + c'$, y simplificando, $a' + d' \leq b' + c'$.

Es fácil comprobar que se trata de una relación de orden total:

- $[a, b] \leq [a, b]$, pues $a + b \leq a + b$.

[2]La idea detrás de esta definición es que pretendemos que las clases $[a, b]$ y $[c, d]$ representen a las restas $a - b$ y $c - d$, respectivamente, y $a - b \leq c - d$ equivale a $a + d \leq b + d$, de modo que esta última relación no involucra las restas $a - b$ y $c - d$ que no tenemos definidas.

- Si $[a,b] \leq [c,d]$ y $[c,d] \leq [a,b]$, entonces $a+d \leq b+c$ y $c+b \leq d+a$, luego $a+d=b+c$, luego $[a,b]=[c,d]$.
- Si $[a,b] \leq [c,d]$ y $[c,d] \leq [e,f]$, entonces $a+d \leq b+c$ y $c+f \leq d+e$. Sumando ambas desigualdades y simplificando, queda $a+f \leq b+e$, luego $[a,b] \leq [e,f]$.
- Si $m \leq n$, entonces $+m \leq +n$ y $-n \leq -m$, pues lo primero equivale a que $[m,0] \leq [n,0]$, es decir, a que $m \leq n$, y lo segundo a que $[0,n] \leq [0,m]$, es decir, también a que $m \leq n$.
- Para todo $m, n \in \mathbb{N}$ no nulos se cumple que $-m \leq 0 \leq +n$, pues lo primero equivale a que $[0,m] \leq [0,0]$, y a su vez a que $0 \leq m$, e igualmente con la segunda desigualdad.

Teniendo en cuenta que todo número entero es 0 o de la forma $\pm n$, con $n > 0$, los dos últimos puntos implican que la relación de orden es total y que con esta ordenación

$$\cdots < -5 < -4 < -3 < -2 < -1 < 0 < +1 < +2 < +3 < +4 < +5 < \cdots$$

En particular, el conjunto $\mathbb{Z}^+$ de los enteros positivos es el formado por los enteros mayores que 0, mientras que los enteros negativos $\mathbb{Z}^-$ son los enteros menores que 0.

Suma y producto de números enteros Definimos ahora una suma y un producto[3] en $\mathbb{Z}$:

$$[a,b]+[c,d]=[a+c,b+d], \qquad [a,b][c,d]=[ac+bd,ad+bc].$$

De nuevo hay que justificar que estas operaciones están bien definidas en el sentido de que no dependen de la elección de los representantes, es decir, que si $[a,b]=[a',b']$ y $[c,d]=[c',d']$, entonces $[a+c,b+d]=[a'+c',b'+d']$ y también $[ac+bd,ad+bc]=[a'c'+b'd',a'd'+b'c']$.

Dejamos como ejercicio el caso de la suma y comprobamos el producto. Tenemos que $a+b'=b+a'$ y que $c+d'=d+c'$, y queremos probar que

$$ac+bd+a'd'+b'c'=ad+bc+a'c'+b'd'.$$

Esto equivale a

$$\begin{aligned} ac+bd+a'd'+b'(c'+d) &= (a+b')d+bc+a'c'+b'd', \\ ac+bd+a'd'+b'c+b'd' &= a'd+bd+bc+a'c'+b'd', \\ (a+b')c+a'd' &= a'(c'+d)+bc, \\ a'c+bc+a'd' &= a'c+a'd'+bc. \end{aligned}$$

Por lo tanto, se da la igualdad requerida.

[3] Si pensamos que $[a,b]$ y $[c,d]$ pretenden representar las restas (que aún no tenemos definidas) $a-b$ y $c-d$, el producto se define teniendo en cuenta que $(a-b)(c-d)=ac+bd-(bc+ad)$, y la definición adecuada de la suma se deduce análogamente.

Veamos las propiedades de estas operaciones (en todas ellas se entiende que m, n, r, etc. son números enteros cualesquiera):

1. $(m+n)+r = m+(n+r)$.

   En efecto, $([a,b]+[c,d])+[e,f] = [a+c+e, b+d+f] = [a,b]+([c,d]+[e,f])$, donde hemos usado la asociatividad de la suma de números naturales.

2. $m+n = n+m$.

   $[a,b]+[c,d] = [a+c, b+d] = [c+a, d+c] = [c,d]+[a,b]$.

3. $m+0 = m$.

   $[a,b]+[0,0] = [a,b]$

4. Todo número entero tiene un opuesto para la suma, pues

   $$[a,b]+[b,a] = [a+b, a+b] = 0.$$

   Más precisamente, si $n \in \mathbb{N}$ tenemos que $+n+(-n) = 0$, luego el opuesto de $+n$ es $-n$ y viceversa.

5. $(mn)r = m(nr)$.

   Hay que comprobar que $([a,b][c,d])[e,f] = [a,b]([c,d][e,f])$. Basta aplicar rutinariamente la definición del producto a ambos miembros hasta comprobar que la expresión resultante es la misma (usando las propiedades de la suma y el producto de los números naturales).

6. $mn = nm$.

   La comprobación es análoga a la de la propiedad anterior.

7. $m(+1) = m$.

   La definición de producto da que $[a,b][1,0] = [a,b]$.

8. $m(n+r) = mn+mr$.

   La comprobación es análoga a la de 5.

9. Si $mn = 0$, entonces $m = 0$ o $n = 0$.

10. Si $m \leq n$, entonces $m+r \leq n+r$.

    Suponemos que $[a,b] \leq [c,d]$, luego $a+d \leq b+c$, y queremos probar que $[a+b]+[e,f] \leq [c+d]+[e,f]$, es decir, que $[a+e, b+f] \leq [c+e, d+f]$ o, equivalentemente, que $a+e+d+f \leq b+f+c+e$, lo cual es claramente cierto.

11. Si $m \geq 0$ y $n \geq 0$, entonces $mn \geq 0$.

    Sabemos que los enteros no negativos son los de la forma $+m$, $+n$, con m y n naturales, y entonces $(+m)(+n) = +(mn) \geq 0$. En efecto, por la propia definición, $[m,0][n,0] = [mn,0]$.

Podríamos demostrar muchas propiedades más, pero éstas son las únicas que necesitamos demostrar antes de "olvidarnos" de que hemos construido los números enteros como clases de equivalencia de pares de números naturales. En efecto, para prescindir completamente de ese tecnicismo sólo nos falta una observación:

Consideremos la aplicación $i : \mathbb{N} \longrightarrow \{0\} \cup \mathbb{Z}^+$ dada por $i(n) = +n$ (notemos que $+0 = 0$). Ya hemos probado que es biyectiva, y que $m < n$ equivale a $i(m) < i(n)$. Ahora podemos observar además que

$$i(m+n) = i(m) + i(n), \qquad i(mn) = i(m)i(n).$$

Más explícitamente, que $+(m+n) = +m + (+n)$ y que $+(mn) = (+m)(+n)$.

Esto significa que la aplicación i es un "diccionario" que hace corresponder cualquier afirmación aritmética sobre los números naturales $0, 1, 2, 3, \ldots$ con la afirmación correspondiente sobre los enteros no negativos $0, +1, +2, +3, \ldots$ de modo que $3 < 5$ es equivalente a $+3 < +5$, y $2 + 5 = 7$ es equivalente a que $+2+(+5) = +7$, y $2 \cdot 3 = 6$ es equivalente a que $(+2)(+3) = +6$. Así, todo lo que sabemos sobre números naturales sigue siendo cierto si los sustituimos por los enteros no negativos[4] y, por consiguiente, a partir de ahora podemos "olvidarnos" de los conjuntos que hasta ahora tomábamos como números naturales y pasar a considerar que los números naturales son los enteros no negativos. Esto nos permite afirmar que

$$\mathbb{N} = \{0\} \cup \mathbb{Z}^+ \subset \mathbb{Z},$$

de modo que la suma y el producto de números naturales son las restricciones a $\mathbb{N} \times \mathbb{N}$ de las operaciones correspondientes en $\mathbb{Z}$. En particular, ya no es necesario emplear la notación $+1, +2, +3$, etc., que usábamos para distinguir a los números enteros positivos de los números naturales correspondientes, sino que podemos eliminar el signo $+$ y llamar $1, 2, 3$, etc. a los enteros positivos, que a partir de ahora *son* también los números naturales.

Además, a partir de ahora ya nunca necesitaremos recordar que los números enteros son clases de equivalencia de pares de los "antiguos" números naturales, sino que nos bastará saber que $\mathbb{Z}$ está formado por los números naturales y sus opuestos (junto con que se cumplen las propiedades que hemos demostrado).

No obstante, recordamos por última vez la construcción de $\mathbb{Z}$ para explicitar la idea subyacente que la ha dirigido en todo momento, es decir, que la clase $[m, n]$ pretendía representar la resta no definida $m - n$. Ahora podemos afirmar (sin hacer referencia a restas no definidas) que

$$[m, n] = [m, 0] + [0, n] = +m + (-n) = m - n,$$

donde m y n son números naturales "antiguos" en la primera expresión y números naturales "nuevos" —enteros no negativos— en la última.

[4]Más precisamente, es fácil ver que $\{0\} \cup \mathbb{Z}^+$ es un sistema de Peano tomando como operación siguiente la función $S(n) = n+(+1)$, por lo que es un conjunto de números naturales tan válido como cualquier otro.

2.3 Conceptos básicos sobre anillos

Finalmente introducimos el concepto central de este capítulo:

Definición 2.7 Un *anillo* es una terna $(A, +, \cdot)$ en la que A es un conjunto y $+$, $\cdot$ son dos leyes internas en A, de modo que se cumplan las propiedades siguientes:

1. $(a + b) + c = a + (b + c)$ para todos los a, b, c de A.
2. $a + b = b + a$ para todos los a, b de A.
3. Existe un elemento 0 en A tal que $a + 0 = a$ para todo a de A.
4. Para todo a de A existe un $-a$ en A tal que $a + (-a) = 0$.
5. $(ab)c = a(bc)$ para todos los a, b, c de A.
6. $a(b + c) = ab + ac$
   $(a + b)c = ac + bc$ para todos los a, b, c de A.

En otros términos, un anillo es una combinación de dos semigrupos $(A, +)$ y $(A, \cdot)$, el primero tiene que ser conmutativo y con elemento neutro 0 (que, según sabemos, es único) y además todo elemento a tiene que tener un opuesto $-a$, que también es único. Al segundo semigrupo no le pedimos en principio que sea conmutativo ni que tenga elemento neutro, pero sí que satisfaga la propiedad 6, que se conoce como *propiedad distributiva* del producto respecto de la suma.

Como es habitual, usaremos los signos $+$ y $\cdot$ para referirnos a las operaciones de un anillo arbitrario, a las que llamaremos "suma" y "producto", de modo que estos signos tendrán una interpretación distinta según el contexto. Del mismo modo, cuando digamos que "A es un anillo", querremos decir que hablamos de una terna $(A, +, \cdot)$, para ciertas operaciones no indicadas explícitamente.

Si el producto de un anillo cumple la propiedad conmutativa diremos que el anillo es *conmutativo.* Si tiene elemento neutro (que será único, lo representaremos por 1 y lo llamaremos la *identidad* de A) diremos que el anillo es *unitario.*

Las propiedades que hemos demostrado en la sección precedente demuestran que $(\mathbb{Z}, +, \cdot)$ es un anillo, al que en lo sucesivo, según lo indicado, nos referiremos simplemente como $\mathbb{Z}$.

En [ITAl] estudiamos ejemplos muy variados de anillos que ilustran el interés de trabajar con este grado de abstracción, como los anillos de restos $\mathbb{Z}_n$, o los anillos de enteros cuadráticos, como los enteros de Gauss o de Eisentein, o los anillos de enteros ciclotómicos, etc.

El teorema siguiente contiene unas cuantas propiedades sencillas de los anillos. Todas ellas se cumplen en particular en el caso de $\mathbb{Z}$, pero aún más importante es saber que podremos usarlas al trabajar con cualquier conjunto del que sepamos que tiene estructura de anillo, por muy abstracta que pueda ser la naturaleza de sus elementos y de sus operaciones.

Teorema 2.8 *Sea A un anillo y a, b, c elementos de A.*

1. *Si $a+b=a+c$ entonces $b=c$.*
2. *Si $a+a=a$ entonces $a=0$.*
3. *$-(-a)=a$.*
4. *$0\cdot a=a\cdot 0=0$.*
5. *$(-a)b=a(-b)=-(ab)$.*
6. *$(-a)(-b)=ab$.*
7. *$-(a+b)=-a-b$.*

Demostración:

1. Si $a+b=a+c$, entonces $-a+a+b=-a+a+c$, luego $0+b=0+c$, luego $b=c$.
2. Si $a+a=a$, entonces $a+a=a+0$, luego $a=0$.
3. $-a+a=0=-a+(-(-a))$, luego $a=-(-a)$.
4. $0\cdot a=(0+0)a=0\cdot a+0\cdot a$, luego $0\cdot a=0$.
5. $(-a)b+ab=(-a+a)b=0b=0$, luego $(-a)b=-(ab)$.
6. $(-a)(-b)=-(a(-b))=-(-(ab))=ab$.
7. $(-a-b)+(a+b)=a-a+b-b=0$, luego $(-a-b)=-(a+b)$. ∎

De la construcción de los números enteros hemos obtenido que los hay de dos clases: positivos, como $+5$, y negativos, como -5. Ya hemos justificado que podíamos suprimir el $+$ del $+5$ al identificarlo con el número natural 5, y ahora vemos que el $-$ del -5 puede entenderse como la notación general que expresa que -5 es el simétrico de 5, del mismo modo que tenemos que $5=-(-5)$.

Si aplicamos en particular las propiedades 5 y 6 a números naturales a y b obtenemos las conocidas "reglas de los signos" para $\mathbb{Z}$:

$$\text{negativo}\times\text{positivo}=\text{negativo},\ \text{negativo}\times\text{negativo}=\text{positivo}$$

a las que hay que añadir, por supuesto, que positivo $\times$ positivo $=$ positivo, que no expresa sino que el producto de números naturales es un número natural.

Con estas propiedades también quedan justificadas las reglas de los signos para la suma. Por ejemplo:

$$5-9=-(-5+9)=-(9-5)=-4.$$

En general, ahora es fácil justificar que para restar dos números naturales se resta el mayor menos el menor y se pone el signo del mayor.

Las propiedades que hemos demostrado sobre $\mathbb{Z}$ en la sección anterior van bastante más allá de lo que exige la definición de anillo conmutativo y unitario, lo cual significa que no todas las propiedades de $\mathbb{Z}$ son extrapolables a cualquier anillo. A continuación vamos a ir introduciendo algunas definiciones generales correspondientes a propiedades que se cumplen en $\mathbb{Z}$.

Empezamos por un hecho básico, y es que en $\mathbb{Z}$ se cumple que $1 \neq 0$, mientras que la definición de anillo unitario no excluye la posibilidad de que $1 = 0$. Ahora bien, este caso es muy particular, pues si un anillo A cumple esta propiedad, entonces todo $a \in A$ cumple $a = a \cdot 1 = a \cdot 0 = 0$. En suma, la igualdad $1 = 0$ sólo puede darse si $A = \{0\}$.

Los anillos que más nos van a interesar son los que llamaremos *dominios,* que son los anillos conmutativos y unitarios en los que $1 \neq 0$.

Claramente, $\mathbb{Z}$ es un dominio.

Notemos, por otra parte, que en cualquier anillo se cumple $a \cdot 0 = 0 \cdot a = 0$, pero no es cierto en general que si $ab = 0$ uno de los factores tenga que ser nulo, propiedad que sí se cumple en $\mathbb{Z}$, según hemos visto.

Definición 2.9 Un elemento a de un dominio A es un *divisor de cero* si es no nulo y existe un b en A no nulo tal que $ab = 0$. Un *dominio íntegro* es un dominio sin divisores de cero.

Tenemos, pues, que $\mathbb{Z}$ es un dominio íntegro.

Una propiedad muy importante de los dominios íntegros es que en ellos podemos simplificar elementos no nulos de las igualdades, es decir, si en un dominio íntegro tenemos que $ab = ac$ y $a \neq 0$, entonces $b = c$, pues $a(b-c) = 0$, luego $b - c = 0$.

El anillo de restos $\mathbb{Z}_6$ introducido en el capítulo III de [ITAl] y que aquí introduciremos en el capítulo siguiente es un ejemplo de dominio con divisores de 0.

Ejercicio: Dotar a $\mathbb{Z} \times \mathbb{Z}$ de una estructura de dominio que no sea íntegro.

Definición 2.10 Un *anillo ordenado* es una cuádrupla $(A, +, \cdot, \leq)$ de manera que $(A, +, \cdot)$ es un dominio, $(A, \leq)$ es un conjunto totalmente ordenado y se cumplen las dos propiedades de compatibilidad siguientes:

1. Si $a \leq b$, entonces $a + c \leq b + c$, para todos los a, b, $c \in A$.
2. Si $a \geq 0$ y $b \geq 0$ entonces $ab \geq 0$.

Los elementos no nulos de cualquier anillo ordenado pueden dividirse en *positivos* y *negativos* según que sean mayores o menores que 0. Representaremos por A^+ y A^- al conjunto de los elementos positivos y negativos de A, respectivamente.

En la sección anterior hemos demostrado que $\mathbb{Z}$ es un anillo ordenado. Veamos las propiedades que se deducen de este hecho (y que, por lo tanto, son válidas en cualquier anillo ordenado):

Teorema 2.11 *Sea A un anillo ordenado. En las propiedades siguientes se entiende que a, b, etc. son elementos de A:*

1. *Si $a \leq b$ y $c \leq d$, entonces $a + c \leq b + d$.*
2. *Si $a < b$ y $c \leq d$, entonces $a + c < b + d$.*
3. *$a \leq b$ si y sólo si $b - a \geq 0$.*
4. *Si $a \leq b$, entonces $-b \leq -a$.*
5. *$a \geq 0$ si y sólo si $-a \leq 0$.*
6. *Si $a \leq b$ y $c \geq 0$, entonces $ac \leq bc$.*
7. *Si $a \leq b$ y $c \leq 0$, entonces $bc \leq ac$.*
8. *$a^2 \geq 0$.*
9. *$1 > 0$*

DEMOSTRACIÓN: 1) Claramente $a + c \leq b + c \leq b + d$.

2) Si $a < b$ no puede ser $a + c = b + c$, pues esto implica $a = b$, luego $a + c < b + c \leq b + d$.

3) $a \leq b$ si y sólo si $0 = a - a \leq b - a$.

4) Por la propiedad anterior, ambas afirmaciones equivalen a $b - a \geq 0$.

5) Es un caso particular de la propiedad precedente.

6) Tenemos que $b-a \geq 0$, luego $(b-a)c \geq 0$, luego $bc-ac \geq 0$,luego $ac \leq bc$.

7) Tenemos que $-c \geq 0$ y por la propiedad anterior $-ac \leq -bc$, luego $bc \leq ac$.

8) Si $a \geq 0$ entonces $a^2 \geq 0$ por la propiedad 2) de compatibilidad. Si $a \leq 0$, entonces $-a \geq 0$, luego $a^2 = (-a)^2 \geq 0$.

9) Basta tener en cuenta que $1 = 1^2$. ∎

En particular, la propiedad 1 implica que si $b \geq 0$ entonces $a \leq a+b$ y esto se generaliza fácilmente a sumas finitas: si $a_1, \ldots, a_n$ son elementos no negativos de A, entonces $a_i \leq a_1 + \cdots + a_n$. En particular, si una suma de elementos no negativos es 0, necesariamente todos los sumandos son 0. Equivalentemente, una suma finita de términos positivos es positiva.

Si A es un dominio íntegro y $a \in A$ es no nulo, entonces $a^2 \neq 0$, luego la propiedad 8 puede refinarse a $a^2 > 0$. Teniendo en cuenta la observación precedente, si $a_1, \ldots, a_n$ son elementos de A cualesquiera y $a_1^2 + \cdots + a_n^2 = 0$, necesariamente $a_1 = \cdots = a_n = 0$.

Definición 2.12 Si A es un anillo ordenado, definimos las aplicaciones *valor absoluto* $|\ |: A \longrightarrow A$ y *signo* $\operatorname{sig}: A \longrightarrow \{-1, 0, 1\}$ mediante

$$|a| = \begin{cases} a & \text{si } a \geq 0, \\ -a & \text{si } a < 0, \end{cases} \qquad \operatorname{sig}(a) = \begin{cases} 1 & \text{si } a > 0, \\ 0 & \text{si } a = 0, \\ -1 & \text{si } a < 0. \end{cases}$$

Observemos que la imagen del valor absoluto está, más concretamente, en el conjunto de los elementos no negativos de A. En particular, para $\mathbb{Z}$ tenemos que $|\ |: \mathbb{Z} \longrightarrow \mathbb{N}$.

En general, cada elemento de A puede reconstruirse a partir de su signo y su valor absoluto mediante la relación $a = \operatorname{sig}(a)|a|$.

Teorema 2.13 *Sea A un anillo ordenado. Entonces*

1. $|a| = 0$ *si y sólo si* $a = 0$.
2. $|\pm 1| = 1$.
3. $|a| = |-a|$.
4. $|a| \leq b$ *si y sólo si* $-b \leq a \leq b$.
5. $|a + b| \leq |a| + |b|$.
6. $||a| - |b|| \leq |a - b|$.
7. $|ab| = |a||b|$.

Demostración: Las tres primeras propiedades son inmediatas por la definición de valor absoluto.

4) Si $|a| \leq b$ y $a \geq 0$, entonces $a \leq b$ y $-b \leq -a \leq 0 \leq a \leq b$. Si $a \leq 0$ entonces tenemos $-a \leq b$, luego $-b \leq a \leq 0 \leq -a \leq b$.

Recíprocamente, si $-b \leq a \leq b$, entonces multiplicando por -1 queda que $-b \leq -a \leq b$, luego, sea $a \geq 0$ o $a \leq 0$, se cumple que $|a| \leq b$.

5) Como $|a| \leq |a|$, por el apartado anterior $-|a| \leq a \leq |a|$, e igualmente $-|b| \leq b \leq |b|$. Sumando las desigualdades:

$$-(|a| + |b|) \leq a + b \leq |a| + |b|,$$

luego $|a + b| \leq |a| + |b|$.

6) $|a| = |a - b + b| \leq |a - b| + |b|$, luego $|a| - |b| \leq |a - b|$. Igualmente $|b| - |a| \leq |b - a| = |a - b|$, luego

$$-|a - b| \leq |a| - |b| \leq |a - b|,$$

luego $||a| - |b|| \leq |a - b|$.

7) Es claro que $|a||b| = \pm ab$, y es positivo, luego es $|ab|$. ∎

Vamos a definir operaciones entre números enteros y los elementos de un anillo.

Definición 2.14 Sea A un anillo, a un elemento de A y n un número entero. Definimos el elemento na como[5]

$$na = \begin{cases} \overbrace{a + \cdots + a}^{n \text{ veces}} & \text{si } n > 0 \\ 0 & \text{si } n = 0 \\ \overbrace{(-a) + \cdots + (-a)}^{-n \text{ veces}} & \text{si } n < 0 \end{cases}$$

Si $n > 0$ definimos también $a^n = \overbrace{a \cdots a}^{n \text{ veces}}$. Si A es unitario $a^0 = 1$, y si a tiene inverso y $n < 0$, entonces $a^n = \overbrace{a^{-1} \cdots \ a^{-1}}^{-n \text{ veces}}$.

Es pura rutina comprobar los hechos siguientes.

Teorema 2.15 *Sea A un anillo unitario y a, b elementos de A (que supondremos inversibles cuando proceda). Sean m y n números enteros. Se cumple:*

1. $m(a + b) = ma + mb$.
2. $(m + n)a = ma + na$.
3. $(-m)a = -(ma) = m(-a)$.
4. $m(na) = (mn)a$.
5. *Si $ab = ba$ entonces* $(ab)^m = a^m b^m$.
6. $a^{m+n} = a^m a^n$.
7. $(a^m)^n = a^{mn}$.
8. $a^{-m} = (a^{-1})^m = (a^m)^{-1}$.

Además si $A = \mathbb{Z}$, ma es lo mismo en el sentido de la definición anterior que en el sentido del producto usual en $\mathbb{Z}$.

Para terminar de exponer las propiedades básicas de $\mathbb{Z}$ probamos que es posible dividir euclídeamente números enteros:

Teorema 2.16 *Sean D y d números enteros con d no nulo. Entonces existen unos únicos enteros c y r tales que $D = dc + r$ y $0 \leq r < |d|$.*

Demostración: Consideremos los números naturales $|D|$ y $|d|$. Sabemos que existen naturales c y r tales que $|D| = |d|c + r$, con $0 \leq r < |d|$.

Si $r = 0$ entonces cambiando el signo de c si es preciso tenemos $D = dc + 0$. Supongamos $r > 0$.

Si $D \geq 0$ y $d > 0$ entonces $D = dc + r$, como queríamos.

Si $D \geq 0$ y $d < 0$ entonces sirve $D = d(-c) + r$.

Si $D < 0$ y $d > 0$ entonces $D = d(-c - 1) + (d - r)$.

Si $D < 0$ y $d < 0$ entonces $D = d(c + 1) + (-d - r)$.

[5]En general, estas definiciones "con puntos suspensivos" se traducen fácilmente en definiciones recurrentes.

Si tuviéramos dos expresiones distintas $D = dc + r = dc' + r'$, entonces sea $\bar{c} = c$ si $d > 0$ y $\bar{c} = -c$ si $d < 0$. Igualmente definimos $\bar{c}'$. Así $dc = |d|\bar{c}$, $dc' = |d|\bar{c}'$. Supongamos que $\bar{c} < \bar{c}'$. Entonces

$$D = dc + r = |d|\bar{c} + r < |d|\bar{c} + |d| = |d|(\bar{c} + 1) \leq |d|\bar{c}' = dc' \leq dc' + r' = D,$$

y esto es una contradicción. Por lo tanto ha de ser $c = c'$ y de aquí que $dc + r = dc + r'$, luego $r = r'$. ∎

Esta propiedad de los números enteros confiere propiedades muy importantes al anillo $\mathbb{Z}$ y es poseída también por otros anillos de interés. Por ello conviene tratarla en general:

Definición 2.17 Un *dominio euclídeo* es un dominio íntegro A tal que existe una función $\phi : A \setminus \{0\} \longrightarrow \mathbb{N}$ tal que si D y d son elementos de A con $d \neq 0$ entonces existen c y r en A de manera que $D = dc + r$ con $r = 0$ o bien $\phi(r) < \phi(d)$. La función ϕ se llama *norma euclídea.*

En [ITAl 2.18] incluimos una propiedad adicional en la definición de dominio euclídeo. Vamos a probar que la definición que hemos dado no es más general:

Teorema 2.18 *Todo dominio euclídeo A admite una norma $\phi : A \setminus \{0\} \longrightarrow \mathbb{N}$ tal que si a, b son elementos de A no nulos, entonces $\phi(a) \leq \phi(ab)$.*

Demostración: Sea $\bar{\phi} : A \setminus \{0\} \longrightarrow \mathbb{N}$ una norma en A según la definición que hemos dado y definamos $\phi : A \setminus \{0\} \longrightarrow \mathbb{N}$ mediante

$$\phi(a) = \min_{b \neq 0} \bar{\phi}(ab) = \min\{\bar{\phi}(ab) \mid b \in A \setminus \{0\}\}.$$

Así ϕ también cumple la definición de dominio euclídeo, pues, dados D y d en A con $d \neq 0$, podemos expresar $\phi(d) = \bar{\phi}(de)$, para cierto $e \in A$ no nulo, y podemos dividir $D = dec' + r$, donde $r = 0$ o bien $\bar{\phi}(r) < \bar{\phi}(de) = \phi(d)$, con lo que $c = ec'$ y r cumplen lo requerido respecto de la norma ϕ.

Además, dados $a, b \in A$ no nulos, se cumple que $\phi(ab) = \bar{\phi}(abc)$, para cierto $c \in A$ no nulo, y entonces $\phi(a) \leq \bar{\phi}(abc) = \phi(ab)$, luego ϕ cumple la propiedad adicional. ∎

En lo sucesivo, siempre que consideremos un dominio euclídeo, consideraremos en él una norma euclídea que cumpla la propiedad dada por el teorema anterior.

Es obvio que $\mathbb{Z}$ es un dominio euclídeo con la norma ϕ dada por $\phi(a) = |a|$. Ahora bien, observemos que el cociente y el resto no son únicos. Por ejemplo, para dividir 8 entre 3 podemos hacer $8 = 3 \cdot 2 + 2$ o bien $8 = 3 \cdot 3 - 1$. En ambos casos $|r| < |d|$.

Definición 2.19 Un elemento a de un dominio A es una *unidad*[6] si existe un elemento b en A tal que $ab = 1$, es decir, si tiene inverso para el producto.

En tal caso sabemos que el inverso de a es único y, de acuerdo con la notación multiplicativa, lo representaremos por a^{-1}.

[6] Es importante recordar que la palabra "unidad" en un dominio no se refiere necesariamente al 1, sino a cualquier elemento con inverso. La palabra para referirse al 1 es "identidad".

Si a es una unidad, entonces a^{-1} es obviamente una unidad, y $(a^{-1})^{-1} = a$. Si a y b son unidades, entonces ab es una unidad, y $(ab)^{-1} = a^{-1}b^{-1}$.

Obviamente 1 es una unidad y $1^{-1} = 1$. En cambio 0 no puede ser una unidad. Una unidad no puede ser divisor de cero, pues si a es una unidad y $ab = 0$, entonces $b = 1b = a^{-1}ab = a^{-1}0 = 0$.

En todo anillo, -1 es también una unidad, pues $(-1)(-1) = 1$. Las unidades de $\mathbb{Z}$ son exactamente 1 y -1.

En efecto, si m, $n \in \mathbb{Z}$ cumplen $mn = 1$, entonces $|m||n| = 1$ y, como $|n| \neq 0$ (y es un número natural) se cumple $1 \leq |n|$, luego $|m| \leq |m||n| = 1$, luego $|m| = 1$, luego $m = \pm 1$.

Un *cuerpo* es un dominio en el que todo elemento no nulo es una unidad. En particular todo cuerpo es un dominio íntegro.

De momento no tenemos ningún ejemplo de cuerpo. Dedicaremos la sección siguiente a extender $\mathbb{Z}$ hasta el cuerpo de los números racionales, pero antes terminamos ésta con algunos hechos sobre cuerpos que nos servirán de guía para dicho fin.

Sea K un cuerpo y a, b dos elementos de K con $b \neq 0$. La *fracción* determinada por a y b es $a/b = ab^{-1}$. a y b se llaman, respectivamente, *numerador* y *denominador* de la fracción a/b.

Veamos las propiedades básicas de las fracciones. La mayoría de ellas son consecuencia inmediata de la definición. En todas ellas se presupone la hipótesis de que los elementos que aparecen como denominadores son no nulos.

1. $\dfrac{a}{1} = a, \qquad \dfrac{1}{b} = b^{-1}, \qquad \left(\dfrac{a}{b}\right)^{-1} = \dfrac{b}{a}.$
2. $\dfrac{a}{b} = \dfrac{c}{d}$ si y sólo si $ad = bc$

   La igualdad de fracciones es $ab^{-1} = cd^{-1}$, y esto equivale a la igualdad que resulta de multiplicar por $bd \neq 0$, que es $ad = bc$.
3. Si $c \neq 0$, entonces $\dfrac{a}{b} = \dfrac{ac}{bc}$.
4. $c\,\dfrac{a}{b} = \dfrac{ca}{b}$.
5. $\dfrac{a}{b} + \dfrac{c}{d} = \dfrac{ad + bc}{bd}$.

   $$\frac{a}{b} + \frac{c}{d} = \frac{ad}{bd} + \frac{bc}{bd} = ad\,\frac{1}{bd} + bc\,\frac{1}{bd} = (ad + bc)\frac{1}{bd} = \frac{ad + bc}{bd}.$$
6. $\dfrac{a}{b}\dfrac{c}{d} = \dfrac{ac}{bd}$.
7. $\dfrac{a/b}{c/d} = \dfrac{ad}{bc}$.

Terminamos con una última observación sobre los cuerpos, y es que trivialmente son dominios euclídeos tomando como norma la aplicación constante 1, pues la división euclídea puede realizarse siempre con resto 0, es decir, descomponiendo $D = d(D/d) + 0$.

2.4 Cuerpos de cocientes. Números racionales

Hemos visto que $\mathbb{Z}$ dista mucho de ser un cuerpo, pues sus únicas unidades son ± 1. Por lo tanto, una fracción como $4/5$ no tiene sentido en $\mathbb{Z}$. Ahora vamos a construir el cuerpo $\mathbb{Q}$ de los números racionales, que contiene a $\mathbb{Z}$ y justifica el uso de fracciones de números enteros.

Como hicimos en la sección 1.4 de [ITAl], vamos a construir un cuerpo K a partir de cualquier dominio íntegro A. La idea es que K esté formado por todas las fracciones de elementos de A, pero en [ITAl] hicimos algo que, técnicamente, no podemos hacer sobre la base conjuntista sobre la que estamos trabajando, y es que allí "definimos" la igualdad de fracciones, mientras que ahora tenemos que definir las fracciones como ciertos conjuntos y no podemos "definir" cuándo dos conjuntos son iguales, ya que lo serán si y sólo si tienen los mismos elementos, y lo que tenemos que hacer es definir las fracciones como los conjuntos adecuados para que sean iguales exactamente cuando queremos que esto suceda.

Si llamamos $A^* = A \setminus \{0\}$, cada fracción está determinada por un par $(a, b) \in A \times A^*$, pero no podemos definir las fracciones como los elementos de este producto cartesiano porque distintos pares pueden determinar la misma fracción.

La situación es análoga a la que nos encontramos en su momento a la hora de construir $\mathbb{Z}$ a partir de $\mathbb{N}$: debemos definir una relación de equivalencia en el conjunto de pares de modo que dos de ellos estén relacionados si y sólo si deben definir la misma fracción, y tomar como fracciones las clases de equivalencia. Los resultados sobre fracciones que hemos visto en la sección precedente nos indican cuáles son las definiciones adecuadas, tanto de la relación de equivalencia como de las operaciones en el cociente:

Definición 2.20 Sea A un dominio íntegro y $A^* = A \setminus \{0\}$. Sea R la relación en el conjunto $A \times A^*$ dada por $(a, b)\, R\, (c, d)$ si y sólo si $ad = bc$.

Es fácil ver que es una relación de equivalencia. Veamos por ejemplo la transitividad: si $(a, b)\, R\, (c, d)$ y $(c, d)\, R\, (e, f)$, tenemos que $ad = bc$ y que $cf = de$, luego $adcf = bcde$. Como $dc \neq 0$ y A es un dominio íntegro, se puede simplificar, y resulta $af = be$, luego $(a, b)\, R\, (e, f)$.

Representaremos por $\dfrac{a}{b}$ la clase de equivalencia del par (a, b) y llamaremos $K = (A \times A^*)/R$ al conjunto cociente. Por las propiedades básicas de las clases de equivalencia sabemos que

$$\frac{a}{b} = \frac{c}{d} \quad \text{si y sólo si} \quad ad = bc.$$

Definimos en K la suma y el producto dadas por

$$\frac{a}{b}+\frac{c}{d}=\frac{ad+bc}{bd}, \qquad \frac{a}{b}\frac{c}{d}=\frac{ac}{bd}.$$

Notemos que, como estamos exigiendo que A sea un dominio íntegro, el hecho de que b, $d\neq 0$ implica que $bd\neq 0$. Si no fuera así estas definiciones no serían correctas. No obstante, todavía falta algo por comprobar para dar por válidas estas definiciones, y es que, como siempre que definimos relaciones o funciones sobre clases de equivalencia, hay que comprobar que la definición no depende de la elección de los representantes. Por ejemplo, para el caso de la suma, si $a/b=a'/b'$ y $c/d=c'/d'$, tiene que cumplirse que

$$\frac{ad+bc}{bd}=\frac{a'd'+b'c'}{b'd'}.$$

Esto es equivalente a que $(ad+bc)b'd'=(a'd'+b'c')bd$. En efecto, teniendo en cuenta que $ab'=ba'$ y que $cd'=dc'$, vemos que

$$(ad+bc)b'd'=ab'dd'+bb'cd'=ba'dd'+bb'dc'=(a'd'+b'c')bd.$$

Ahora es pura rutina comprobar que K cumple todos los axiomas de cuerpo con estas definiciones. El cuerpo K así construido se llama *cuerpo de cocientes* de A. Los elementos neutros de K son $0=0/1$ y $1=1/1$, el simétrico de una fracción es $-(a/b)=(-a)/b$, y el inverso es $(a/b)^{-1}=b/a$. Aquí hay que observar que si $a/b\neq 0/1$, entonces $a\neq 0$.

Con esto no hemos terminado la construcción del cuerpo de cocientes, pues nos falta llegar a que podemos ver a A como un subconjunto de K. Esta afirmación requiere varias precisiones, y para plantear exactamente lo que queremos demostrar conviene introducir algunos conceptos abstractos:

Homomorfismos de anillos Es claro que lo que interesa de un anillo no es en absoluto la naturaleza conjuntista de sus elementos, sino el modo en que los relacionan las leyes internas. Por ejemplo, si $A=\{a,b\}$ es cualquier conjunto con dos elementos, es fácil convertirlo en un anillo (cuerpo, de hecho) con las leyes dadas por

$$a+a=b+b=a, \quad a+b=b+a=b,$$
$$aa=ab=ba=a, \quad bb=b.$$

Con estas operaciones, $a=0$ y $b=1$ (es decir, a es el neutro de la suma y b el del producto) y en estos términos las operaciones son

$$0+0=1+1=0, \quad 0+1=1+0=1,$$
$$0\cdot 0=0\cdot 1=1\cdot 0=0, \quad 1\cdot 1=1.$$

Si hacemos lo mismo con otro conjunto $A'=\{a',b'\}$ obtenemos un anillo distinto conjuntistamente, pero el mismo anillo algebraicamente. La forma de plasmar esta relación es el concepto de homomorfismo de anillos que definimos a continuación.

Definición 2.21 Sean A y B dos anillos. Una aplicación $f : A \longrightarrow B$ es un *homomorfismo de anillos* si cumple $f(a+b) = f(a) + f(b)$, $f(ab) = f(a)f(b)$ para todos los elementos a y b de A.

Una consecuencia inmediata es que $f(0) = f(0+0) = f(0) + f(0)$, luego $f(0) = 0$, y que $f(a) + f(-a) = f(a-a) = f(0) = 0$, luego $f(-a) = -f(a)$. También es claro que si m es un número entero $f(ma) = mf(a)$.

Sin embargo, hay que tener presente que, aunque A y B sean unitarios, no tiene por qué ocurrir $f(1) = 1$. Por ejemplo la aplicación que vale constantemente 0 es un homomorfismo (el único) que cumple $f(1) = 0$.

Suponiendo $f(1) \neq 0$, una condición suficiente para que $f(1) = 1$ es que B sea un dominio íntegro, pues entonces $f(1)f(1) = f(1 \cdot 1) = f(1) = f(1)1$, luego $f(1) = 1$.

Otra condición que garantiza que $f(1) = 1$ es que $f : A \longrightarrow B$ sea suprayectiva, pues entonces tiene que existir un $a \in A$ tal que $f(a) = 1$ y así

$$f(1) = f(1)1 = f(1)f(a) = f(1 \cdot a) = f(a) = 1.$$

Cuando $f(1) = 1$ y $a \in A$ es inversible, se cumple que $f(a^{-1})$ es inversible y $f(a^{-1}) = f(a)^{-1}$.

Más en general, se cumple que $f(a^n) = f(a)^n$ para todo $n \in \mathbb{Z}$. Si a no es inversible, esto se cumple de todos modos para todo $n \in \mathbb{N}$.

Es inmediato comprobar que la composición de homomorfismos es un homomorfismo.

Un *isomorfismo* de anillos es un homomorfismo biyectivo. Notemos que si $f : A \longrightarrow B$ es un isomorfismo, entonces $f^{-1} : B \longrightarrow A$ también es un isomorfismo, pues $f\big(f^{-1}(a) + f^{-1}(b)\big) = f\big(f^{-1}(a)\big) + f\big(f^{-1}(b)\big) = a + b$, luego $f^{-1}(a+b) = f^{-1}(a) + f^{-1}(b)$, e igualmente ocurre con el producto.

Dos anillos A y B son *isomorfos* (abreviadamente, $A \cong B$) si existe un isomorfismo $f : A \longrightarrow B$. Cuando dos anillos son isomorfos son algebraicamente indistinguibles, es decir, uno es conmutativo si y sólo si lo es el otro, etc. Por lo tanto podemos considerarlos como "el mismo" anillo (aunque conjuntistamente sean distintos).

Un anillo A es un *subanillo* de un anillo B si $A \subset B$ y las operaciones de A son las mismas que las de B. Por ejemplo, $\{2n \mid n \in \mathbb{Z}\}$ es un subanillo de $\mathbb{Z}$ (no unitario, por cierto). En general, si $f : A \longrightarrow B$ es un homomorfismo, es fácil ver que $f[A]$ es un subanillo de B.

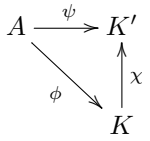
Ejercicio: Considerando $\mathbb{Z} \times \mathbb{Z}$ y $\mathbb{Z} \times \{0\}$, probar que la identidad de un subanillo puede ser distinta de la del anillo. Probar que esto es imposible si los anillos son dominios íntegros.

Un *monomorfismo* de anillos es un homomorfismo inyectivo. Si $f : A \longrightarrow B$ es un monomorfismo es claro que $f : A \longrightarrow f[A]$ es un isomorfismo, o sea, A es isomorfo a un subanillo de B, luego podemos identificar A con su imagen y considerar que A es un subanillo de B.

Éste es el caso de un dominio íntegro y su cuerpo de cocientes:

Teorema 2.22 *Sea A un dominio íntegro y K su cuerpo de cocientes.*

1. *La aplicación $\phi : A \longrightarrow K$ dada por $\phi(a) = a/1$ es un monomorfismo de anillos.*
2. *Si K' es un cuerpo y $\psi : A \longrightarrow K'$ es un monomorfismo de anillos, existe un único monomorfismo de cuerpos $\chi : K \longrightarrow K'$ que hace conmutativo el diagrama siguiente:*



Demostración: 1) es inmediato. En 2), puesto que $a/b = \phi(a)\phi(b)^{-1}$, el monomorfismo χ tiene que venir dado necesariamente por

$$\chi(a/b) = \chi(\phi(a))\chi(\phi(b))^{-1} = \psi(a)\psi(b)^{-1}.$$

Vamos a ver que esto define realmente un monomorfismo, que necesariamente será único. En general, siempre que definamos una función sobre fracciones, tenemos que comprobar que la definición no depende de la representación que usamos, pues ésta no es única. En este caso se trata de probar que si $a/b = a'/b'$ entonces $\psi(a)\psi(b)^{-1} = \psi(a')\psi(b')^{-1}$, ahora bien, sabemos que $ab' = ba'$ y, como ψ es un homomorfismo, $\psi(a)\psi(b') = \psi(b)\psi(a')$, de donde se sigue la relación requerida. Es fácil probar que χ es un monomorfismo y trivialmente hace conmutativo el diagrama, pues $\chi(\phi(a)) = \chi(a/1) = \psi(a)$. ∎

Lo que afirma la parte 1) del teorema anterior es que podemos considerar a A como un subanillo de su cuerpo de cocientes sin más que identificar cada elemento a con $a/1$, es decir, considerando que dividir entre 1 es no hacer nada.

La parte 2) afirma en particular que si un cuerpo K' contiene a A, entonces también contiene una copia isomorfa de K, a saber, el conjunto $\{ab^{-1} \mid a, b \in A\}$. En otras palabras, si ya tenemos a A contenido en un cuerpo K' no necesitamos salirnos de K' para construir el cuerpo de cocientes de A. Basta tomar todas las fracciones posibles con elementos de A y el resultado será un cuerpo isomorfo al cuerpo de cocientes de A.

Definición 2.23 Llamaremos cuerpo de los *números racionales* $\mathbb{Q}$ al cuerpo de cocientes de $\mathbb{Z}$. Los elementos de $\mathbb{Q}$ son las fracciones a/b con a, b en $\mathbb{Z}$, $b \neq 0$.

En lo sucesivo, los números racionales de la forma $n/1$, donde $n \in \mathbb{Z}$, los representaremos simplemente por n. Así, a partir de ahora nos "olvidaremos" de que habíamos definido los números enteros como clases de equivalencia de pares de números naturales y pasaremos a considerar que, por ejemplo, el número entero 5 es el número racional $5/1$. Hemos probado que con esto estamos

cambiando el $\mathbb{Z}$ "antiguo" por otro anillo isomorfo, por lo que todos los resultados que conocíamos sobre el "antiguo" $\mathbb{Z}$ siguen siendo válidos para "el nuevo", con la diferencia de que ahora podemos afirmar que $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q}$.

En general, si D es un dominio íntegro ordenado, toda fracción de K puede expresarse en la forma a/b con $b > 0$. En efecto, si $b < 0$ la podemos transformar en $a/b = (-a)/(-b)$, y ahora $-b > 0$.

Teorema 2.24 *Si A es un dominio íntegro ordenado, entonces su cuerpo de cocientes K es un cuerpo ordenado con la relación dada por*

$$\frac{a}{b} \leq \frac{c}{d} \quad \textit{si y sólo si} \quad ad \leq bc,$$

donde las fracciones se toman con denominadores positivos.

Demostración: Hay que comprobar que la relación está bien definida, es decir, que si $a/b = a'/b'$ y $c/d = c'/d'$ (con denominadores positivos) y $ad \leq bc$, entonces también $a'd' \leq b'c'$. En efecto, sabemos que $ab' = a'b$ y $cd' = c'd$. Como b', $d' > 0$, tenemos que $ad \leq bc$, luego $ab'd'd \leq bb'd'c$, luego $a'bdd' \leq bb'dc'$, luego $a'd' \leq b'c'$.

Es claro que se trata de una relación de orden total. Veamos como ejemplo la prueba de la transitividad: si $a/b \leq c/d \leq e/f$ (con denominadores positivos) tenemos que $ad \leq bc$ y $cf \leq de$. Por lo tanto $adf \leq bcf \leq bde$, luego $af \leq be$, luego $a/b \leq e/f$.

En cuanto a las propiedades de compatibilidad, veamos por ejemplo la de la suma: Suponemos que $a/b \leq c/d$ y queremos probar que $a/b + e/f \leq c/d + e/f$. Equivalentemente, suponemos que $ad \leq bc$ y queremos probar que

$$\frac{af + be}{bf} \leq \frac{cf + de}{df},$$

que a su vez equivale a $afdf + bedf \leq cfbf + debf$, o también a $adf^2 \leq cbf^2$, pero esto es cierto porque $ad \leq cb$ y $f^2 \geq 0$. ∎

Observemos que, según la relación de orden dada por el teorema anterior, $a/1 \leq b/1$ si y sólo si $a \leq b$, es decir, que al identificar a A con un subanillo de K, la relación de orden en A no es sino la restricción de la de K.

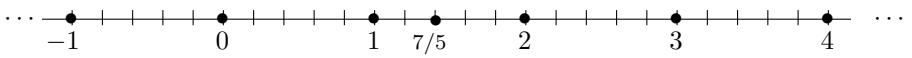
En particular esto se aplica a $\mathbb{Q}$, que resulta ser un cuerpo ordenado, por lo que tenemos definidos los conceptos de valor absoluto y signo de un número racional, y cumplen las propiedades que hemos demostrado en general.

Otra característica general válida para todo cuerpo ordenado K es que si u, $v \in K$, con $u < v$, existe un $w \in K$ tal que $u < w < v$. Esto se expresa diciendo que los cuerpos ordenados son *densos* en sí mismos.

En efecto, podemos definir $2 = 1 + 1 \in K$, que claramente cumple $2 > 0$, y entonces basta tomar $w = \dfrac{u + v}{2}$. Es inmediato comprobar que $u < w < v$.

Notemos que $\mathbb{Z}$ no cumple esto, sino que entre dos números enteros consecutivos no hay ningún otro número entero. Para que el argumento anterior funcione hace falta que 2 tenga inverso, cosa que no sucede en $\mathbb{Z}$.

Los números racionales tienen también su representación geométrica como puntos de una recta. Por ejemplo, para representar $7/5 = 1 + 2/5$ dividimos en cinco partes cada intervalo entre dos números enteros consecutivos y tomamos, desde 0, 7 de dichos intervalos.



Notemos que si quisiéramos señalar en la recta todos los puntos correspondientes a números racionales, por muy finos que trazáramos los puntos, en la práctica "llenaríamos" toda la recta. Por ejemplo, para representar todas las fracciones con denominador 1 000 000 tendríamos que poner un millón de puntos entre cada dos números enteros consecutivos. Sin embargo, el hecho de que los puntos con números racionales asignados "estén por todas partes", no resuelve el problema teórico de si quedan todavía puntos de la recta sin asignar o no. Volveremos sobre esto más adelante.

Números combinatorios Los cuerpos de cocientes nos permiten salirnos temporalmente de un anillo en nuestros cálculos aunque después volvamos a él. Veamos un ejemplo.

Sean $n, n_1, \ldots, n_k$ números naturales tales que $n = n_1 + \cdots + n_k$. Definimos el *número combinatorio*

$$\binom{n}{n_1 \cdots n_k} = \frac{n!}{n_1! \cdots n_k!}$$

Si $0 \leq m \leq n$ abreviaremos

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

Por ejemplo, $\binom{5}{3} = 10$. Vamos a demostrar las propiedades principales de los números combinatorios.

Teorema 2.25 *Sean $m \leq n$ números naturales.*

1. $\binom{n}{m} = \binom{n}{n-m}$.
2. $\binom{n}{0} = \binom{n}{n} = 1, \quad \binom{n}{1} = n$.
3. *Si $m < n$,* $\binom{n}{m} + \binom{n}{m+1} = \binom{n+1}{m+1}$.
4. *Los números combinatorios son números naturales.*

DEMOSTRACIÓN: 3) Hay que probar que

$$\frac{n!}{m!\,(n-m)!}+\frac{n!}{(m+1)!\,(n-m-1)!}=\frac{(n+1)!}{(m+1)!\,(n-m)!}.$$

Ahora bien,

$$\left(\frac{n!}{m!\;(n-m)!}+\frac{n!}{(m+1)!\;(n-m-1)!}\right)(m+1)!\;(n-m)!$$

$$=\frac{n!\,(m+1)\,m!\,(n-m)!}{m!\,(n-m)!}+\frac{n!\,(m+1)!\,(n-m)(n-m-1)!}{(m+1)!\,(n-m-1)!}$$

$$=n!\,(m+1)+n!\,(n-m)=m\,n!+n!+n\,n!-m\,n!=(n+1)\,n!=(n+1)!$$

4) Una simple inducción nos da que $\binom{n}{m}$ es un número natural, pues cada número combinatorio con $n+1$ es suma de dos con n, por el apartado 3).

Para el caso general basta usar que

$$\binom{n}{n_1\ \dots\ n_k\ n_{k+1}}=\binom{n-n_{k+1}}{n_1\ \dots\ n_k}\binom{n}{n_{k+1}}.$$

■

La forma más fácil de calcular los números combinatorios es disponerlos en forma de triángulo, de modo que cada uno es la suma de los dos que hay sobre él. El triángulo así construido se suele llamar *triángulo de Tartaglia.*

					1					
				1		1				
			1		2		1			
		1		3		3		1		
	1		4		6		4		1	
1		5		10		10		5		1

Triángulo de Tartaglia

Los números combinatorios tienen una interpretación combinatoria:

Teorema 2.26 *Si A es un conjunto de cardinal n y $m \leq n$, entonces A tiene $\binom{n}{m}$ subconjuntos de cardinal m.*

DEMOSTRACIÓN: Por inducción sobre n. Si $n=0$ entonces $A=\varnothing$ y tiene un único subconjunto de cardinal 0, y ciertamente $\binom{0}{0}=1$.

Supongamos que es cierto para conjuntos de cardinal n y supongamos que $|A|=n+1$. Es inmediato que A tiene un único subconjunto de cardinal 0, y $\binom{n+1}{0}=1$. Consideremos ahora los subconjuntos de cardinal $m+1$. Fijado un $a\in A$, el conjunto X de todos los subconjuntos de A de cardinal $m+1$ puede dividirse en dos subconjuntos disjuntos, el subconjunto Y formado por los subconjuntos que contienen a a y el subconjunto Z formado por los que no lo contienen. Claramente, hay tantos subconjuntos de A con $m+1$ elementos que contienen a a como subconjuntos tiene $A\setminus\{a\}$ con m elementos, luego,

por hipótesis de inducción, $|Y| = \binom{n}{m}$. Por otra parte, también es claro que $|Z| = \binom{n}{m+1}$. Por consiguiente:

$$|X| = |Y| + |Z| = \binom{n}{m} + \binom{n}{m+1} = \binom{n+1}{m+1}.$$

■

La utilidad principal de estos números será para nosotros el hecho siguiente:

Teorema 2.27 (Binomio de Newton) *Sea A dominio, n un número natural y a, b dos elementos de A. Entonces*

$$(a+b)^n = \sum_{m=0}^{n} \binom{n}{m} a^m b^{n-m}.$$

DEMOSTRACIÓN: Por inducción sobre n. Para $n = 0$ es inmediato.

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

$$= \sum_{m=0}^{n} \binom{n}{m} a^{m+1} b^{n-m} + \sum_{m=0}^{n} \binom{n}{m} a^m b^{n-m+1}$$

$$= \sum_{m=1}^{n+1} \binom{n}{m-1} a^m b^{n+1-m} + \sum_{m=0}^{n} \binom{n}{m} a^m b^{n+1-m}$$

$$= \binom{n}{0} a^0 b^{n+1} + \binom{n}{n} a^{n+1} b^0$$

$$+ \sum_{m=1}^{n} \binom{n}{m-1} a^m b^{n+1-m} + \sum_{m=1}^{n} \binom{n}{m} a^m b^{n+1-m}$$

$$= \binom{n+1}{0} a^0 b^{n+1} + \binom{n+1}{n+1} a^{n+1} b^0$$

$$+ \sum_{m=1}^{n} \left(\binom{n}{m-1} + \binom{n}{m} \right) a^m b^{n+1-m}$$

$$= \binom{n+1}{0} a^0 b^{n+1} + \binom{n+1}{n+1} a^{n+1} b^0 + \sum_{m=1}^{n} \binom{n+1}{m} a^m b^{n+1-m}$$

$$= \sum_{m=0}^{n+1} \binom{n+1}{m} a^m b^{n+1-m}.$$

■

Una consecuencia inmediata es que $\sum_{m=0}^{n} \binom{n}{m} = (1+1)^n = 2^n$.

De forma similar se demuestra en general:

Teorema 2.28 *Sea A un anillo conmutativo y unitario, n un número natural y $a_1, \dots, a_k$ elementos de A. Entonces se cumple:*

$$(a_1 + \cdots + a_k)^n = \sum_{n_1,\dots,n_k} \binom{n}{n_1 \cdots n_k} a_1^{n_1} \cdots a_k^{n_k},$$

donde la suma se extiende sobre todos los números naturales $n_1, \dots, n_k$ tales que $n_1 + \cdots + n_k = n$.

2.5 Anillos de polinomios

En la sección 1.3 de [ITAl] vimos cómo es posible añadir una "indeterminada" a cualquier dominio A para obtener el anillo $A[x]$ de los polinomios con coeficientes en A, así como que el proceso se puede repetir para construir anillos de polinomios con cualquier número de indeterminadas. Recordemos que la idea básica es que si, por ejemplo, x e y son números enteros, $xy + x$ y $x^2 - 2y$ son otros números enteros. Su suma es $x^2 + xy + x - 2y$ y su producto

$$(xy + x)(x^2 - 2y) = x^2(xy + x) - 2y(xy + x) = x^3y + x^3 - 2xy^2 - 2xy.$$

Pero al trabajar con números enteros o, más en general, con elementos de cualquier dominio, surgen fácilmente relaciones de este estilo y a menudo resulta muy útil poder tratarlas como objetos y no como meros términos que relacionan números concretos. Ahora vamos a presentar una construcción general que permite añadir a cada anillo unitario A un conjunto (tal vez infinito) de elementos indeterminados, como aquí son x e y, de modo que obtengamos un nuevo anillo con elementos como $x^3y + x^3 - 2xy^2 - 2xy$. A estos objetos los llamaremos polinomios.

Puesto que vamos a incorporar "de una vez" un conjunto arbitrario de indeterminadas a un dominio dado, la construcción resulta ser algo más técnica que la que vimos en [ITAl], que ya era un tanto técnica.

Definición 2.29 Sea S un conjunto. Llamemos M el conjunto de las aplicaciones $u : S \longrightarrow \mathbb{N}$ tales que el conjunto $\{i \in S \mid u(i) \neq 0\}$ es finito.

Por ejemplo, si $S = \{x, y, z\}$ y una función $u \in M$ cumple $u(x) = 3$, $u(y) = 1$, $u(z) = 7$, nuestra intención es que u represente al monomio puro x^3yz^7.

Si u, v son funciones de M llamaremos $u + v$ a la función dada por

$$(u + v)(i) = u(i) + v(i).$$

Claramente $u + v$ está en M.

Notemos que la suma $u + v$ representa al producto de los monomios representados por u y por v. Si $m \in \mathbb{N}$ y $u \in M$ llamaremos mu a la función dada por $(mu)(i) = m\big(u(i)\big)$. También es claro que mu está en M. Es claro que mu representa a la potencia m–sima del monomio representado por u. Llamaremos 0 a la función de M que toma constantemente el valor 0.

Si $x \in S$ llamaremos $\epsilon_x \in M$ a la función que toma el valor 1 en x y vale 0 en cualquier otro punto. Claramente, ϵ_x representa al monomio x.

Notemos que si $u \in M$ y $x_1, \ldots, x_n$ son los puntos donde u no se anula, entonces u puede expresarse como $u = u(x_1)\epsilon_{x_1} + \cdots + u(x_n)\epsilon_{x_n}$. Si pensamos en el primer ejemplo, esto se interpreta como que el monomio u es el producto del monomio x elevado a 3, por el monomio y, por el monomio z elevado a 7.

Un polinomio arbitrario, como $x^3y + x^3 - 2xy^2 - 2xy$, es una suma de monomios no necesariamente puros, sino multiplicados por coeficientes en un anillo dado. Esto nos lleva a la definición siguiente:

Si A es un anillo unitario, llamaremos conjunto de los *polinomios* con *indeterminadas* en S sobre A al conjunto $A[S]$ formado por las funciones $f : M \longrightarrow A$ tales que el conjunto $\{u \in M \mid f(u) \neq 0\}$ es finito.

Así, si $f \in A[S]$ y $u \in M$, el elemento $f(u)$ se interpreta como el coeficiente del monomio u en f. Con estas ideas el lector puede convencerse de que la definición lógica de las operaciones en $A[S]$ es la siguiente:

$$(f+g)(u) = f(u) + g(u), \qquad (fg)(u) = \sum_{v+w=u} f(v)g(w).$$

Notemos que el sumatorio que define el producto es finito.

Teorema 2.30 *Sea A un anillo unitario y S un conjunto. Entonces $A[S]$ es un anillo unitario. Si A es conmutativo, $A[S]$ también lo es.*

Demostración: Es fácil ver que para todos los polinomios f, g, $h \in A[S]$, se cumple $(f+g)+h = f+(g+h)$ y $f+g = g+f$.

La aplicación $0 : M \longrightarrow A$ que toma constantemente el valor 0 es el elemento neutro de $A[S]$ y si $f \in A[S]$, la función dada por $(-f)(u) = -f(u)$ es el simétrico de f. Si f, g, $h \in A[S]$ y $u \in M$ se cumple

$$\begin{aligned}\big((fg)h\big)(u) &= \sum_{v+w=u}\sum_{s+t=v} f(s)g(t)h(w) = \sum_{w+s+t=u} f(s)g(t)h(w)\\ &= \sum_{s+v=u}\sum_{t+w=v} f(s)g(t)h(w) = \big(f(gh)\big)(u),\end{aligned}$$

luego $(fg)h = f(gh)$.

$$\begin{aligned}\big(f(g+h)\big)(u) &= \sum_{v+w=u} f(v)\big(g(w)+h(w)\big)\\ &= \sum_{v+w=u} f(v)g(w) + \sum_{v+w=u} f(v)h(w) = (fg)(u) + (fh)(u),\end{aligned}$$

luego $f(g+h) = fg + fh$, e igualmente $(f+g)h = fh + gh$.

Sea 1 la aplicación que vale 1 sobre $0 \in M$ y vale 0 en otro caso. Entonces $(f1)(u) = f(u)$, luego $f1 = f$. Igualmente $1f = f$.

Si A es conmutativo

$$(fg)(u) = \sum_{v+w=u} f(v)g(w) = \sum_{v+w=u} g(w)f(v) = (gf)(u),$$

luego $fg = gf$, es decir, $A[S]$ es conmutativo. ∎

Los teoremas siguientes prueban que los polinomios son lo que esperamos que sean. El primer paso es sumergir A en $A[S]$. El teorema siguiente es una comprobación rutinaria.

Teorema 2.31 *Sea A un anillo unitario y S un conjunto. Para cada $a \in A$ sea f_a el polinomio que cumple $f_a(0) = a$ y que toma el valor 0 en cualquier otro caso. Sea $\phi : A \longrightarrow A[S]$ la aplicación dada por $\phi(a) = f_a$. Entonces ϕ es un monomorfismo de anillos y $\phi(1) = 1$.*

Definición 2.32 En lo sucesivo, si A es un anillo unitario, S un conjunto y $a \in A$, escribiremos a en lugar de $\phi(a)$ y A en lugar de $\phi[A]$. De este modo A es un subanillo de $A[S]$. Para cada $x \in S$ llamaremos $\bar{x}$ al polinomio que cumple $\bar{x}(\epsilon_x) = 1$ y que toma el valor 0 en cualquier otro caso. La aplicación que a cada x le asigna $\bar{x}$ es biyectiva, luego podemos identificar x con $\bar{x}$ y así considerar que $S \subset A[S]$. A los elementos de S los llamaremos *indeterminadas*.

El teorema siguiente recoge el comportamiento de los polinomios construidos a partir de las indeterminadas mediante productos. Inmediatamente después probaremos que todo polinomio puede construirse a partir de las indeterminadas mediante sumas y productos.

Teorema 2.33 *Sea A un anillo unitario y S un conjunto.*

1. *Si $k \in \mathbb{N}$, $a \in A$ y $x \in S$, entonces el polinomio ax^k toma el valor a sobre $k\epsilon_x$ y 0 en otro caso.*
2. *Si $k_1, \ldots, k_n \in \mathbb{N}$, $a \in A$ y $x_1, \ldots, x_n$ son indeterminadas distintas, entonces el polinomio $ax_1^{k_1} \cdots x_n^{k_n}$ toma el valor a sobre $k_1\epsilon_{x_1} + \cdots + k_n\epsilon_{x_n}$ y 0 en otro caso.*
3. *Si $x, y \in S$, entonces $xy = yx$.*
4. *Si $a \in A$ y $x \in S$, entonces $ax = xa$.*

Demostración:

1. Por inducción sobre k. Para $k = 0$ es inmediato. Supuesto cierto para k, entonces $(ax^{k+1})(u) = \big((ax^k)x\big)(u) = (ax^k)(v)x(w) = 0$ salvo si $v = k\epsilon_x$ y $w = \epsilon_x$, es decir, salvo si $u = (k+1)\epsilon_x$, en cuyo caso da a.
2. Por inducción sobre n. Para $n = 1$ es el caso anterior. Supuesto cierto para n tenemos que $(ax_1^{k_1} \cdots x_{n+1}^{k_{n+1}})(u) = (ax_1^{k1} \cdots x_n^{k_n})(v)(x_{n+1}^{k_{n+1}})(w) = 0$ salvo que $v = k_1\epsilon_{x_1} + \cdots + k_n\epsilon_{x_n}$ y $w = k_{n+1}\epsilon_{x_{n+1}}$, es decir, salvo si $u = k_1\epsilon_{x_1} + \cdots + k_{n+1}\epsilon_{x_{n+1}}$, en cuyo caso vale a.

3. es inmediato por 2, pues ambos polinomios son la misma función.

4. Basta notar que el caso 1 se prueba igual con a por la derecha. ∎

Como consecuencia inmediata tenemos:

Teorema 2.34 *Sea A un anillo unitario y S un conjunto. El polinomio*

$$\sum_{i=1}^{m} a_i\, x_1^{k_{i1}} \cdots x_n^{k_{in}},$$

donde $a_1, \ldots, a_m \in A$, $x_1, \ldots, x_n$ son indeterminadas distintas y las n–tuplas de naturales $(k_{i1}, \ldots, k_{in})$ son todas distintas, vale a_i sobre $k_{i1}\epsilon_{x_1} + \cdots + k_{in}\epsilon_{x_n}$ y vale 0 en cualquier otro caso.

Como los polinomios de esta forma cubren todas las aplicaciones posibles de M en A (con un número finito de imágenes no nulas) hemos demostrado:

Teorema 2.35 *Sea A un anillo unitario y S un conjunto. Todo polinomio no nulo de $A[S]$ se expresa en la forma descrita en el teorema anterior para ciertas indeterminadas, ciertos elementos de A y ciertas n–tuplas de naturales. La expresión es única (salvo el orden) si exigimos que todos los a_i sean no nulos y que cada indeterminada tenga exponente no nulo en al menos un sumando.*

Definición 2.36 En la expresión de 2.34, los elementos a_i se llaman *coeficientes* del polinomio. Concretamente a_i es el coeficiente del término en $x_1^{k_{i1}} \cdots x_n^{k_{in}}$. Se entiende que si un término no aparece en la expresión, su coeficiente es 0 (siempre puede añadirse multiplicado por 0).

Un polinomio con un único coeficiente no nulo (o sea, un polinomio de la forma $a\, x_1^{k_1} \ldots x_n^{k_n}$) es un *monomio* de *grado* $k_1 + \cdots + k_n$ y *coeficiente* a. Por lo tanto, todo polinomio se expresa siempre como suma de monomios. A veces se les llama binomios, trinomios, etc. según el número de monomios que los compongan. El *grado* de un polinomio no nulo p es el mayor de los grados de sus monomios con coeficiente no nulo y lo representaremos por $\operatorname{grad} p$.

El coeficiente del término del monomio cuyos exponentes son todos nulos se llama *término independiente,* es decir, el término independiente de f es $f(0)$. Un polinomio cuyo único coeficiente no nulo sea a lo sumo el término independiente es un polinomio *constante.* Los polinomios constantes son exactamente los elementos de A, según la identificación que hemos realizado.

Tenemos definidos anillos de polinomios con cualquier cantidad de indeterminadas, posiblemente infinitas. Cuando $S = \{x_1, \ldots, x_n\}$ es finito, en lugar de $A[S]$ se escribe también $A[x_1, \ldots, x_n]$.

Por ejemplo, un elemento de $\mathbb{Z}[x, y, z]$ es $3x^5y^2z^2 + 8x^2z - 6z^2 + 5$. El término independiente es 5, el coeficiente del monomio en x^2z es 8 (en el cual la indeterminada y tiene exponente 0), el coeficiente del monomio en x^5 es 0.

Cuando sólo hay una indeterminada la expresión de un polinomio es más sencilla. Cada polinomio no nulo de $A[x]$ es de la forma $\sum_{i=0}^{m} a_i x^i$, y la expresión es única si exigimos que $a_m \neq 0$.

El coeficiente del monomio de mayor grado de p se llama *coeficiente director* del polinomio p. Un polinomio de $A[x]$ es *mónico* si su coeficiente director es 1.

La suma y el producto de polinomios con una indeterminada es más simple:

$$\sum_{i=0}^{m} a_i x^i + \sum_{i=0}^{m} b_i x^i = \sum_{i=0}^{m} (a_i + b_i) x^i,$$

$$\left(\sum_{i=0}^{m} a_i x^i\right)\left(\sum_{i=0}^{n} b_i x^i\right) = \sum_{k=0}^{m+n}\left(\sum_{i+j=k} a_i b_j\right) x^k.$$

Por ejemplo, un elemento de $\mathbb{Z}[x]$ es $2x^5 + 5x^2 - 11x + 6$. Se trata de un polinomio de grado 5 con coeficiente director igual a 2.

En la práctica escribiremos $p = p(x_1, \ldots, x_n)$ para indicar que las indeterminadas $x_1, \ldots, x_n$ son las únicas (a lo sumo) que aparecen en el polinomio p con exponentes no nulos.

Evaluación de polinomios La evaluación de polinomios es un concepto muy sencillo: si $p(x) = 2x^2 - 4x$, pretendemos que $p(3)$ sea $2 \cdot 3^2 - 4 \cdot 3 = 6$. No obstante, vamos a definir las evaluaciones en un contexto más general que nos será útil después.

Definición 2.37 Sean A y B dos anillos conmutativos y unitarios, $\phi : A \longrightarrow B$ un homomorfismo, S un conjunto y $v : S \longrightarrow B$ cualquier aplicación. Para cada polinomio $p = \sum_{i=1}^{m} a_i\, x_1^{k_{i1}} \ldots x_n^{k_{in}} \in A[S]$ definimos la *evaluación*

$$\phi p(v) = \sum_{i=1}^{m} \phi(a_i)\, v(x_1)^{k_{i1}} \ldots v(x_n)^{k_{in}} \in B.$$

La conmutatividad de B y la unicidad de la expresión hacen que $\phi p(v)$ esté bien definido, pues dos expresiones de p difieren sólo en el orden de las indeterminadas y en la presencia de monomios con coeficiente 0, o de indeterminadas con exponente 0, pero en cualquier caso se obtiene el mismo elemento de B.

Tenemos, por tanto, una aplicación $\Phi : A[S] \longrightarrow B$ dada por $\Phi(p) = \phi p(v)$.

En definitiva $\Phi(p)$ se calcula reemplazando los coeficientes de p por su imagen por ϕ y las indeterminadas por sus imágenes por v.

En la práctica, si $p = p(x_1, \ldots, x_n)$ escribiremos $\phi p(b_1, \ldots, b_n)$ para indicar el polinomio que resulta de evaluar cada indeterminada x_i con el elemento b_i. Notar que aunque S pueda ser infinito, $\phi p(v)$ sólo depende de la forma en que v actúa sobre las indeterminadas que aparecen en p, que son siempre un número finito. Cuando ϕ sea simplemente la identidad en A no lo escribiremos, y expresaremos la evaluación mediante $p(b_1, \ldots, b_n)$.

Teorema 2.38 *Sean A y B dos anillos conmutativos y unitarios, $\phi : A \longrightarrow B$ un homomorfismo, S un conjunto y $v : S \longrightarrow B$ cualquier aplicación. Entonces la evaluación $\Phi : A[S] \longrightarrow B$ es el único homomorfismo que coincide con ϕ sobre A y con v sobre S.*

DEMOSTRACIÓN: Sean p, $q \in A[S]$, digamos

$$p = \sum_{i=1}^{m} a_i\, x_1^{k_{i1}} \dots\ x_n^{k_{in}}, \qquad q = \sum_{i=1}^{m} b_i\, x_1^{k_{i1}} \dots\ x_n^{k_{in}}.$$

Observemos que no hay problema en suponer que los exponentes de los monomios son los mismos, pues podemos añadir monomios con coeficiente 0 hasta igualar ambas expresiones.

$$\Phi(p+q) = \Phi\left(\sum_{i=1}^{m}(a_i+b_i)\, x_1^{k_{i1}} \dots\ x_n^{k_{in}}\right) = \sum_{i=1}^{m}\phi(a_i+b_i)v(x_1)^{k_{i1}} \dots\ v(x_n)^{k_{in}}$$

$$= \sum_{i=1}^{m}\big(\phi(a_i)+\phi(b_i)\big)v(x_1)^{k_{i1}} \dots\ v(x_n)^{k_{in}} = \sum_{i=1}^{m}\phi(a_i)v(x_1)^{k_{i1}} \dots\ v(x_n)^{k_{in}}$$

$$+\sum_{i=1}^{m}\phi(b_i)v(x_1)^{k_{i1}} \dots\ v(x_n)^{k_{in}} = \Phi(p)+\Phi(q).$$

Para probar que Φ conserva productos usaremos el hecho ya probado de que conserva las sumas.

$$\begin{aligned}
\Phi(pq) &= \Phi\left(\sum_{i,j=1}^{m} a_i b_j\, x_1^{k_{i1}+k_{j1}} \dots\ x_n^{k_{in}+k_{jn}}\right)\\
&= \sum_{i,j=1}^{m} \Phi(a_i b_j\, x_1^{k_{i1}+k_{j1}} \dots\ x_n^{k_{in}+k_{jn}})\\
&= \sum_{i,j=1}^{m} \phi(a_i b_j)\, v(x_1)^{k_{i1}+k_{j1}} \dots\ v(x_n)^{k_{in}+k_{jn}}\\
&= \sum_{i,j=1}^{m} \phi(a_i)\phi(b_j)\, v(x_1)^{k_{i1}+k_{j1}} \dots\ v(x_n)^{k_{in}+k_{jn}}\\
&= \left(\sum_{i=1}^{m}\phi(a_i)\, v(x_1)^{k_{i1}} \dots v(x_n)^{k_{in}}\right)\left(\sum_{j=1}^{m}\phi(b_j)\, v(x_1)^{k_{j1}} \dots v(x_n)^{k_{jn}}\right)\\
&= \Phi(p)\Phi(q).
\end{aligned}$$

La unicidad es evidente. ∎

De este teorema se deducen varios casos particulares de interés.

Teorema 2.39 *Sean A y B anillos conmutativos y unitarios, $\phi : A \longrightarrow B$ un homomorfismo de anillos y S un conjunto. Entonces existe un único homomorfismo $\bar{\phi} : A[S] \longrightarrow B[S]$ que coincide con ϕ en A y deja invariantes a las indeterminadas. Además es inyectivo, suprayectivo o biyectivo si ϕ lo es.*

DEMOSTRACIÓN: El homomorfismo no es sino el construido en el teorema anterior tomando como v la identidad en S. Concretamente

$$\bar{\phi}\left(\sum_{i=1}^{m} a_i\, x_1^{k_{i1}} \dots\ x_n^{k_{in}}\right) = \sum_{i=1}^{m} \phi(a_i)\, x_1^{k_{i1}} \dots\ x_n^{k_{in}}.$$

Todo lo pedido es obvio. ∎

Esto significa en particular que si A es un subanillo de B podemos considerar $A[S]$ como un subanillo de $B[S]$. Así por ejemplo, $\mathbb{Z}[S] \subset \mathbb{Q}[S]$.

Teorema 2.40 *Sea A un anillo conmutativo y unitario. Sea S un conjunto y supongamos que $S = X \cup Y$ con X e Y disjuntos. Sea B el conjunto de los polinomios de $A[S]$ tales que todos sus monomios con coeficientes no nulos tengan tan sólo indeterminadas de X con exponentes no nulos. Entonces B es un subanillo de $A[S]$ isomorfo a $A[X]$ y $A[S]$ es isomorfo a $A[X][Y]$.*

DEMOSTRACIÓN: Sea $\phi : A[X] \longrightarrow A[S]$ el homomorfismo construido en 2.38 con la identidad en A y la identidad en X. Es claro que B es la imagen de ϕ y que ϕ es un monomorfismo.

Ahora sea $\psi : A[X][Y] \longrightarrow A[S]$ el homomorfismo construido en 2.38 a partir de ϕ y de la identidad en Y. Es inmediato probar que se trata de un isomorfismo de anillos. ∎

Por ejemplo, el polinomio $3x^5y^2z^2 + 8x^2z - 6z^2 + 5$ de $\mathbb{Z}[x, y, z]$ puede ser identificado con $(3x^5y^2 - 6)z^2 + (8x^2)z + 5 \in \mathbb{Z}[x, y][z]$, donde ahora $3x^5y^2 - 6$ es el coeficiente de z^2.

Si lo queremos en $\mathbb{Z}[z][x, y]$ será: $3z^2(x^5y^2) + (8z)x^2 + (-6z^2 + 5)$, donde ahora $-6z^2 + 5$ es el término independiente.

Por otra parte si $S \subset T$ podemos considerar $A[S] \subset A[T]$.

Propiedades algebraicas Las principales propiedades algebraicas de los anillos de polinomios se deducen de consideraciones sobre los grados. Es obvio que el grado de la suma de dos polinomios f y g de $A[x]$ es menor o igual que el máximo de los grados de f y g. Será igual a dicho máximo si sus grados son distintos, pero si coinciden se pueden cancelar los coeficientes directores y el grado de la suma disminuye:

$$(3x^5 - 2x^2 + 5x + 2) + (-3x^5 + x^3 - x^2 + 1) = x^3 - 3x^2 + 5x + 3.$$

El grado del producto es a lo sumo la suma de los grados. Normalmente se da la igualdad. Las únicas excepciones se dan si uno de los factores es nulo, o si alguno de los coeficientes directores es un divisor de cero.

Teorema 2.41 *Sea A un anillo unitario y p, q dos polinomios no nulos de $A[x]$ tales que al menos el coeficiente director de uno de ellos no sea un divisor de cero. Entonces $pq \neq 0$, $\operatorname{grad}(pq) = \operatorname{grad}(p) + \operatorname{grad}(q)$ y el coeficiente director del producto es el producto de los coeficientes directores.*

DEMOSTRACIÓN: Sean $p = \sum_{i=0}^{m} a_i x^i$, $q = \sum_{i=0}^{n} b_i x^i$, con $a_m \neq 0 \neq b_n$. Entonces $pq = \sum_{k=0}^{m+n} \big(\sum_{i+j=k} a_i b_j \big) x^k$ y el coeficiente de x^{m+n} es exactamente $a_m b_n \neq 0$, puesto que uno de ellos no es divisor de cero. Por lo tanto $a_m b_n$ es el coeficiente director de pq y el grado es $m + n$. ∎

Teorema 2.42 *Sea A un dominio íntegro y S un conjunto cualquiera. Entonces $A[S]$ es un dominio íntegro.*

DEMOSTRACIÓN: El teorema anterior nos da que si A es un dominio íntegro entonces $A[x]$ también lo es. Aplicándolo un número finito de veces obtenemos que si A es un dominio íntegro y S es finito, entonces $A[S]$ también lo es. Si S es arbitrario y f, g son dos polinomios no nulos de $A[S]$, entonces los monomios con coeficientes no nulos de f y g contienen un número finito de indeterminadas con exponente no nulo, luego f y g están en un subanillo $A[X]$ con X finito, luego $A[X]$ es un dominio íntegro, luego $fg \neq 0$. Por tanto $A[S]$ es un dominio íntegro. ∎

Teorema 2.43 *Sea A un dominio íntegro y S un conjunto. Entonces las unidades de $A[S]$ son las mismas que las de A.*

DEMOSTRACIÓN: Veámoslo primero para $A[x]$. Si $p \in A[x]$ es una unidad, entonces existe otro polinomio no nulo q tal que $pq = 1$. Por 2.41 tenemos que $\operatorname{grad} p + \operatorname{grad} q = \operatorname{grad} 1 = 0$, luego ha de ser $\operatorname{grad} p = \operatorname{grad} q = 0$, es decir, p y q están en A, luego p es una unidad en A.

De aquí se sigue el resultado para $A[S]$ con S finito y, por el mismo argumento que en el teorema anterior, vale para todo S. ∎

Fracciones algebraicas En particular vemos que $A[S]$ no es un cuerpo aunque A lo sea. Como sí es un dominio íntegro, podemos definir su cuerpo de fracciones.

Definición 2.44 Sea A un dominio íntegro y S un conjunto. Llamaremos cuerpo de las *fracciones algebraicas* o *funciones racionales* sobre A con indeterminadas en S al cuerpo de cocientes de $A[S]$. Lo representaremos por $A(S)$.

Así, por ejemplo, un elemento de $\mathbb{Z}(x, y)$ es $\frac{x^4 - x^3 y}{x^3 - 4xy^2 + 4}$.

Ejercicio: Probar que $\mathbb{Z}(S) = \mathbb{Q}(S)$.

Quizá éste es un buen momento para empezar a entender la utilidad del lenguaje algebraico que hemos empezado a introducir en este capítulo: el hecho de que $\mathbb{Z}[x]$ sea un anillo (y más concretamente un dominio íntegro) nos permite tratar formalmente a sus elementos con las mismas reglas básicas que a los números enteros. El hecho de que conozcamos la construcción general del cuerpo de cocientes de un dominio íntegro justifica que hablemos de fracciones de polinomios exactamente igual que de fracciones de enteros, y estos ejemplos son sólo una mínima parte de los que nos vamos a encontrar.

División de polinomios Debemos ocuparnos ahora de la posibilidad de dividir polinomios. La división euclídea es una característica importantísima de los polinomios con una única indeterminada.

Teorema 2.45 *Sea A un anillo unitario, D y d dos polinomios no nulos de $A[x]$ tales que el coeficiente director de d sea una unidad en A. Entonces existen unos únicos polinomios c y r en $A[x]$ tales que $D = dc+r$ con $r = 0$ o de grado menor estrictamente que el grado de d (también podemos exigir que $D = cd + r$, pero si A no es conmutativo los polinomios que cumplan esto no tienen por qué ser los mismos).*

DEMOSTRACIÓN: Si $\operatorname{grad} D < \operatorname{grad} d$ basta tomar $c = 0$ y $r = D$. Supongamos que $\operatorname{grad} d \leq \operatorname{grad} D$.

Sea $D = \sum_{i=0}^{n} a_i x^i$, $d = \sum_{i=0}^{m} b_i x^i$, con $a_n \neq 0 \neq b_m$ y $m \leq n$. Además estamos suponiendo que b_m es una unidad de A. Veamos el teorema por inducción sobre n.

Si $n = 0$, entonces también $m = 0$, es decir, $D = a_0$, $d = b_0$, luego basta tomar $c = (b_0)^{-1}a_0$ y $r = 0$. Supongámoslo cierto para polinomios de grado menor que n.

Consideremos $db_m^{-1}a_n x^{n-m} = \sum_{i=0}^{m} b_i b_m^{-1} a_n x^{i+n-m}$. El monomio de mayor grado es $b_m(b_m)^{-1}a_n x^{m+n-m} = a_n x^n$, luego se trata de un polinomio de grado n con coeficiente director a_n.

Consecuentemente el polinomio $D - d(b_m)^{-1}a_n x^{n-m}$ tiene grado menor que n, luego por hipótesis de inducción existen polinomios c' y r de manera que $D - db_m^{-1}a_n x^{n-m} = dc' + r$ con $\operatorname{grad} r < \operatorname{grad} d$. Sea $c = b_m^{-1}a_n x^{n-m} + c'$. Así $D = dc + r$ como se pedía.

Veamos ahora la unicidad. Supongamos que $D = dc+r = dc'+r'$. Entonces $d(c-c') = r'-r$. Si $c-c' \neq 0$, como el coeficiente director de d es una unidad, por el teorema 2.41. resulta que $\operatorname{grad}(r'-r) = \operatorname{grad}(d(c-c')) = \operatorname{grad} d + \operatorname{grad}(c-c')$, pero $\operatorname{grad}(r'-r) < \operatorname{grad} d \leq \operatorname{grad} d + \operatorname{grad}(c-c')$, contradicción. Concluimos entonces que $c = c'$, luego también $r = r'$. ∎

El lector que sepa dividir números naturales puede adaptar su método para dividir también polinomios. No hay ninguna diferencia esencial.

Es importante que para poder dividir polinomios el divisor debe tener coeficiente director unitario. En particular podemos dividir siempre entre polinomios mónicos. Cuando A es un cuerpo todos los coeficientes son unidades, luego se pueden dividir polinomios cualesquiera. Como en este caso el grado del producto es la suma de los grados, tenemos todas las condiciones exigidas en la definición de dominio euclídeo, es decir:

Teorema 2.46 *Si K es un cuerpo, entonces el anillo de polinomios $K[x]$ es un dominio euclídeo.*

Sin embargo esto es falso si K no es un cuerpo. Por ejemplo $\mathbb{Z}[x]$ no es un dominio euclídeo. Tampoco es cierto en anillos de polinomios con más de una indeterminada. Por ejemplo $\mathbb{Q}[x,y]$ no es un dominio euclídeo. Estos hechos los probaremos en el capítulo siguiente. Es interesante notar que en estos momentos no tenemos idea de cómo puede probarse la no existencia de una norma euclídea.

2.6 Apéndice: Sumas y productos finitos

Presentamos en este apéndice las propiedades básicas que justifican las operaciones con sumas y productos finitos de elementos de un semigrupo A con elemento neutro. En realidad podríamos prescindir del elemento neutro, pero así se simplifica la exposición. Recordemos la definición 2.3 de producto finito: dada una sucesión $\{a_i\}_{i=1}^n$ de elementos de A, definimos

$$\prod_{i<0} a_i = 1, \qquad \prod_{i<j+1} a_i = (\prod_{i<j} a_i)\cdot a_j,$$

para $j \leq n$, de modo que en particular tenemos definido el producto finito de los elementos dados:

$$a_0 \cdots a_n = \prod_{i=0}^{n} a_i = \prod_{i<n+1} a_i.$$

A su vez, definiendo

$$\prod_{i=m}^{n} a_i = \prod_{i=0}^{n-m} a_{m+i}, \qquad \text{para } m \leq n$$

se cumple la relación recurrente

$$\prod_{i=u}^{u} a_i = a_u, \qquad \prod_{i=u}^{v+1} a_i = \left(\prod_{i=u}^{v} a_i\right)\cdot a_{v+1},$$

para $u \leq v < n$. Esto puede "resumirse" con la notación $\prod_{i=u}^{v} a_i = a_u \cdots a_v$.

El teorema siguiente afirma que un producto puede partirse en bloques:

$$a_1 \cdots a_n = (a_1 \cdots a_{n_1})(a_{n_1+1} \cdots a_{n_2}) \cdots (a_{n_{m-1}+1} \cdots a_n).$$

Teorema 2.47 (Propiedad asociativa generalizada) *Sea A un semigrupo con elemento neutro, sea $\{a_i\}_{i=1}^n$ una sucesión de elementos de A, y sea $\{n_j\}_{j=0}^m$ una sucesión de números naturales $0 = n_0 < n_1 < \cdots < n_m = n$. Entonces*

$$\prod_{i=1}^{n} a_i = \prod_{j=1}^{m} \prod_{i=n_{j-1}+1}^{n_j} a_i.$$

Demostración: Observemos que el primer producto del miembro derecho es el definido a partir de la sucesión $\left\{ \prod_{i=n_{j-1}+1}^{n_j} a_i \right\}_{j=1}^m$.

En primer lugar demostramos que, para todo si $1 \leq u \leq n - n_k$, se cumple la igualdad

$$\prod_{i=1}^{n_k+u} a_i = \prod_{i=1}^{n_k} a_i \prod_{i=n_k+1}^{n_k+u} a_i.$$

Razonamos por inducción sobre u. Para $u = 0$ no hay nada que probar y para $u = 1$ es claramente cierta. Si vale para $u \geq 1$, entonces

$$\prod_{i=1}^{n_k+u+1} a_i = \left(\prod_{i=1}^{n_k+u} a_i\right) a_{n_k+u+1} = \prod_{i=1}^{n_k} a_i \left(\prod_{i=n_k+1}^{n_k+u} a_i\right) a_{n_k+u+1}$$

$$= \prod_{i=1}^{n_k} a_i \prod_{i=n_k+1}^{n_k+u+1} a_i.$$

En particular, si aplicamos esto a $u = n_{k+1} - n_k$ obtenemos que

$$\prod_{i=1}^{n_{k+1}} a_i = \prod_{i=1}^{n_k} a_i \prod_{i=n_k+1}^{n_{k+1}} a_i.$$

Ahora probamos por inducción sobre k que si $1 \leq k \leq m$, entonces

$$\prod_{i=1}^{n_k} a_i = \prod_{j=1}^{k} \prod_{i=n_{j-1}+1}^{n_j} a_i.$$

Para $k = 0$ la igualdad se reduce a $1 = 1$. Supuesto cierto para k, si $k + 1 \leq m$ tenemos que

$$\prod_{i=1}^{n_{k+1}} a_i = \prod_{i=1}^{n_k} a_i \prod_{i=n_k+1}^{n_{k+1}} a_i = \left(\prod_{j=1}^{k} \prod_{i=n_{j-1}+1}^{n_j} a_i\right) \prod_{i=n_k+1}^{n_{k+1}} a_i = \prod_{j=1}^{k+1} \prod_{i=n_{j-1}+1}^{n_j} a_i.$$

Aplicando esto a $k = m$ obtenemos la igualdad del enunciado. ∎

A partir de este punto vamos a suponer que el semigrupo es conmutativo y, pasamos a emplear notación aditiva por ilustrar también el uso de esta notación, si bien podríamos mantener también la notación multiplicativa.

Teorema 2.48 (Propiedad conmutativa generalizada) *Sea A un semigrupo conmutativo y con elemento neutro. Sea $\{a_i\}_{i=1}^n$ una sucesión de elementos de A y sea $\sigma : I_n \longrightarrow I_n$ biyectiva. Entonces*

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} a_{\sigma(i)}.$$

Demostración: Lo probamos por inducción sobre n. Si $n = 0$ por definición ambos miembros son 0. Si vale para n, consideramos una sucesión $\{a_i\}_{i=1}^{n+1}$ en A y una biyección $\sigma : I_{n+1} \longrightarrow I_{n+1}$. Pongamos que $\sigma(k) = n + 1$. Por la propiedad asociativa generalizada,

$$\sum_{i=1}^{n+1} a_{\sigma(i)} = \sum_{i=1}^{k-1} a_{\sigma(i)} + a_{n+1} + \sum_{i=k+1}^{n+1} a_{\sigma(i)} = \sum_{i=1}^{k-1} a_{\sigma(i)} + \sum_{i=k+1}^{n+1} a_{\sigma(i)} + a_{n+1}$$

$$= \sum_{i=1}^{k-1} a_{\sigma(i)} + \sum_{i=k}^{n} a_{\sigma(i+1)} + a_{n+1}.$$

La última igualdad se debe a que, por definición,

$$\sum_{i=k+1}^{n+1} a_{\sigma(i)} = \sum_{i=0}^{n-k} a_{\sigma(k+i+1)} = \sum_{i=k}^{n} a_{\sigma(i+1)}.$$

Sea $\tau : I_n \longrightarrow I_n$ dada por

$$\tau(i) = \begin{cases} \sigma(i) & \text{si } i < k, \\ \sigma(i+1) & \text{si } i \geq k. \end{cases}$$

Entonces:

$$\sum_{i=1}^{n+1} a_{\sigma(i)} = \sum_{i=1}^{k-1} a_{\tau(i)} + \sum_{i=k}^{n} a_{\tau(i)} + a_{n+1} = \sum_{i=1}^{n} a_{\tau(i)} + a_{n+1}$$

$$= \sum_{i=1}^{n} a_i + a_{n+1} = \sum_{i=1}^{n+1} a_i.$$

donde hemos usado la propiedad asociativa generalizada y la hipótesis de inducción. ∎

Este teorema afirma que una suma finita no depende del orden de los sumandos. Para hacer esto más explícito vamos a considerar una sucesión $\{a_i\}_{i\in I}$, donde I es un conjunto finito cualquiera, no necesariamente de la forma I_n. Sea $n = |I|$ y sean $f : I_n \longrightarrow I$, $g : I_n \longrightarrow I$ dos biyecciones. Entonces $\sigma = g \circ f^{-1} : I_n \longrightarrow I_n$ es también una biyección y cumple que

$$a_{f(\sigma(i))} = a_{g(f^{-1}(f(i)))} = a_{g(i)}.$$

Por el teorema anterior, $\sum_{i=1}^{n} a_{f(i)} = \sum_{i=1}^{n} a_{f(\sigma(i))} = \sum_{i=1}^{n} a_{g(i)}$. Esto justifica la definición siguiente:

Definición 2.49 Sea A un semigrupo cuya operación sea conmutativa y tenga elemento neutro. Sea I un conjunto finito no vacío, sean $\{a_i\}_{i\in I}$ elementos de A. Definimos

$$\sum_{i\in I} a_i = \sum_{j=1}^{n} a_{f(j)},$$

donde $n = |I|$ y $f : I_n \longrightarrow I$ es cualquier biyección.

Hemos demostrado que la suma así definida es la misma sea cual sea la biyección f empleada para calcularla. Convenimos además en que $\sum_{i\in\varnothing} a_i = 0$.

El teorema siguiente sobre cambio del conjunto de índices es inmediato:

Teorema 2.50 *Sea A un semigrupo conmutativo con elemento neutro, consideremos una biyección $h : I \longrightarrow J$ entre conjuntos finitos y sean $\{a_j\}_{j\in J}$ elementos de A. Entonces*

$$\sum_{i\in I} a_{h(i)} = \sum_{j\in J} a_j.$$

DEMOSTRACIÓN: Si los conjuntos son vacíos la igualdad se reduce a $0 = 0$. En caso contrario tomamos $n = |I|$ y $f : I_n \longrightarrow I$ biyectiva y observamos que si aplicamos la definición de suma finita con f para el miembro izquierdo y $f \circ h : I_n \longrightarrow J$ para el miembro derecho, obtenemos en ambos casos la misma suma finita. ∎

Ahora podemos expresar la propiedad asociativa generalizada bajo hipótesis más generales que en la versión precedente:

Teorema 2.51 (Propiedad asociativa generalizada) *Sea A un semigrupo conmutativo con elemento neutro, sea $\{I_j\}_{j=1}^m$ una familia de conjuntos finitos disjuntos dos a dos, sea $I = \bigcup_{j=1}^m I_j$ y sean $\{a_i\}_{i\in I}$ elementos de A. Entonces,*

$$\sum_{i\in I} a_i = \sum_{j=1}^m \sum_{i\in I_j} a_i.$$

DEMOSTRACIÓN: Sea $k_j = |I_j|$ y sea $f_j : I_{k_j} \longrightarrow I_j$ biyectiva. Definimos por recurrencia:

$$n_0 = 0, \quad n_j = n_{j-1} + k_j, \qquad \text{para } 1 \le j \le m.$$

Sea $f'_j : \{i \in \mathbb{N} \mid n_{j-1} + 1 \le i \le n_j\} \longrightarrow I_j$ dada por $f'_j(i) = f_j(i - n_{j-1})$, claramente biyectiva. Como los conjuntos I_j son disjuntos, las funciones f'_j determinan una función $f : I_{n_m} \longrightarrow I$ biyectiva. Entonces

$$\sum_{i\in I} a_i = \sum_{i=1}^{n_m} a_{f(i)} = \sum_{j=1}^m \sum_{i=n_{j-1}+1}^{n_j} a_{f'_j(i)} = \sum_{j=1}^m \sum_{i=1}^{k_j} a_{f_j(i)} = \sum_{j=1}^m \sum_{i\in I_j} a_i. \quad ∎$$

Todas las propiedades elementales de las sumas y productos finitos pueden demostrarse ya fácilmente, bien a partir de las propiedades que hemos demostrado, bien mediante inducciones sencillas.

Por ejemplo, consideremos la propiedad de intercambio de sumatorios:

$$\sum_{i=1}^n \sum_{j=1}^m a_{ij} = \sum_{j=1}^m \sum_{i=1}^n a_{ij}.$$

Para probarla basta considerar $I_n \times I_m = \bigcup_{i=1}^n \{i\} \times I_m$ y aplicar el teorema anterior:

$$\sum_{(i,j)\in I_n\times I_m} a_{ij} = \sum_{i=1}^n \sum_{(i,j)\in\{i\}\times I_m} a_{ij} = \sum_{i=1}^n \sum_{j\in I_m} a_{ij} = \sum_{i=1}^n \sum_{j=1}^m a_{ij},$$

e igualmente

$$\sum_{(j,i)\in I_m\times I_n} a_{ij} = \sum_{j=1}^m \sum_{i=1}^n a_{ij},$$

y los dos miembros izquierdos son iguales por el teorema 2.50, considerando la biyección $f : I_n \times I_m \longrightarrow I_m \times I_n$ dada por $f(i, j) = (j, i)$.

Capítulo III

Aritmética en dominios íntegros

En el capítulo anterior hemos presentado la aritmética elemental de los números enteros, es decir, hemos definido la suma y el producto y hemos visto que podemos dividir números enteros, sea para obtener números racionales, sea mediante la división euclídea, para obtener un cociente y un resto enteros. Ahora vamos a ascender un nivel introduciendo los conceptos de la aritmética clásica: múltiplos, divisores, primos, etc. Todo esto puede hacerse mediante técnicas elementales que no involucren más que los conceptos que ya conocemos sobre números enteros (incluso podríamos trabajar exclusivamente con números naturales, como en el capítulo II de [ITAl]), pero aquí vamos a adoptar un punto de vista eminentemente algebraico, y formularemos los resultados en términos que los hagan válidos para una clase muy amplia de anillos, de modo que todos los resultados sobre divisibilidad, primos, etc. que demostremos para $\mathbb{Z}$ serán aplicables a otros contextos en los que se cumplan unos requisitos mínimos.

3.1 Ideales

En el análisis algebraico de la aritmética es fundamental el concepto de ideal. Este concepto surgió relativamente tarde en el estudio de los números debido en gran parte a que para el caso de los números enteros se vuelve trivial y, por lo tanto prescindible, e incluso "invisible", en el sentido de que a nadie se le ocurriría usar ideales para formular una serie de hechos que se enuncian con total naturalidad sin hacer referencia a ellos. Sin embargo, cuando las propiedades aritméticas de $\mathbb{Z}$ se expresan en términos de ideales, entonces son directamente generalizables a otros contextos, mientras que las formulaciones clásicas no lo permiten, de modo que quien piensa en términos de ideales puede ver que dos contextos diferentes son esencialmente análogos cuando a quien sólo piense en términos de números o, más en general, de elementos de un anillo, puede no ver relación entre ellos.

Definición 3.1 Un *ideal* en un dominio[1] A es un conjunto $I \subset A$ que cumpla las propiedades siguientes:

1. $0 \in I$,
2. si $a, b \in I$, entonces $a + b \in I$,
3. si $a \in A$ y $b \in I$ entonces $ab \in I$.

Notemos que la propiedad 3. nos da en particular que si $b \in I$ entonces $-b = (-1)b \in I$, y combinando esto con 2. resulta que si $a, b \in I$, entonces también $a - b = a + (-b) \in I$. En particular, todo ideal es un subanillo.

Si $a \in A$, se llama *ideal principal* generado por a al ideal

$$aA = Aa = \{ab \mid b \in A\}.$$

El conjunto aA es lo que habitualmente se llama el conjunto de los *múltiplos* de a en A. Es inmediato comprobar que es realmente un ideal de A. Lo que expresan los axiomas de la definición de ideal es que el 0 siempre es múltiplo de a, que la suma de dos múltiplos de a es también múltiplo de a y que todo múltiplo de un múltiplo de a es también un múltiplo de a.

Del mismo modo que podemos ver el concepto de "dominio" como una selección de propiedades básicas que satisfacen los números enteros y que abstraemos para aplicarlas a otros objetos que pueden ser muy distintos de los números enteros, el concepto de "ideal" surge como una selección de las propiedades básicas que cumplen los conjuntos de múltiplos, con la idea de que pueden ser aplicables a otros subconjuntos de un anillo que no sean realmente conjuntos de múltiplos.

Todo anillo tiene al menos dos ideales, a saber, $\{0\}$ y el propio A. Se les llama ideales *impropios.* El ideal $\{0\}$ es el ideal *trivial* y se representa simplemente por 0. Notemos que ambos son principales, pues $0 = 0A$ y $A = 1A$.

Un poco más en general, un hecho simple, pero importante, sobre ideales es que si un ideal I de un dominio A contiene una unidad, entonces $I = A$.

En efecto, si $u \in I$ es una unidad, por definición de ideal se cumple que $1 = u^{-1}u \in I$ y si $a \in A$, entonces $a = a \cdot 1 \in I$. Así pues, todo elemento de A está en I.

Por lo tanto, los únicos ideales de un cuerpo son los impropios, pues si un ideal de un cuerpo posee un elemento no nulo, éste será una unidad, y el ideal será el cuerpo completo.

Así pues, el concepto de ideal se vuelve trivial cuando se aplica a los cuerpos, y lo mismo sucede en una amplia clase de anillos, en un sentido menos radical que recoge la definición siguiente:

[1]La definición es válida para anillos unitarios arbitrarios sin más que añadir que también $ba \in I$ en la propiedad 3.

Un *dominio de ideales principales* (DIP) es un dominio íntegro en el que todo ideal es principal.

La observación precedente muestra que todo cuerpo es trivialmente DIP, pero se cumple algo mucho más general:

Teorema 3.2 *Todo dominio euclídeo es un dominio de ideales principales.*

DEMOSTRACIÓN: Sea A un dominio euclídeo y sea $\phi : A \setminus \{0\} \longrightarrow \mathbb{N}$ su norma euclídea. Sea $I \neq 0$ un ideal de A (si $I = 0$ ya es principal). Sea $a \in I$ tal que $\phi(a)$ sea el mínimo del conjunto $\{\phi(b) \mid b \in I, b \neq 0\}$.

Si $b \in I$, entonces $b = ac + r$, con $r = 0$ o bien $\phi(r) < \phi(a)$. Como $a \in I$, por la definición de ideal $ac \in I$, y a su vez $b - ac \in I$, es decir, $r \in I$. Como $\phi(a)$ es mínimo, no puede ser $\phi(r) < \phi(a)$, luego $r = 0$, es decir, $b = ac \in aA$.

Hemos probado que $I \subset aA$. Como $a \in I$, la otra inclusión es consecuencia de la definición de ideal. Por tanto $I = aA$ es un ideal principal. ∎

En particular tenemos que $\mathbb{Z}$ es un DIP, es decir, los únicos ideales de $\mathbb{Z}$ son los de la forma $n\mathbb{Z}$, para $n \in \mathbb{Z}$. También son DIP los anillos de polinomios $K[x]$, donde K es un cuerpo.

Esto es lo que hace que el concepto de ideal sea superfluo al estudiar la aritmética de $\mathbb{Z}$ o de anillos como $K[x]$. Cada ideal I de un DIP es de la forma $I = aA$, para cierto $a \in A$, y todo lo que se diga en términos de I puede decirse igualmente en términos de a, sin necesidad de mencionar ideales para nada. Sin embargo, vamos a ver las propiedades aritméticas expresadas en términos de ideales son más generales que las expresadas en términos de elementos "reales" del anillo, pues siguen siendo válidas en contextos en los que no existe la correspondencia entre (elementos) ideales y "elementos reales" que se da en los DIP.

Terminamos esta sección introduciendo algunos conceptos convenientes para tratar con ideales en anillos que no sean DIP.

Definición 3.3 Es inmediato que la intersección de una familia (no vacía) de ideales de un anillo A sigue siendo un ideal de A. Por lo tanto si $X \subset A$, existe un mínimo ideal de A que contiene a X, a saber, la intersección de la familia de todos los ideales de A que contienen a X (existe al menos uno, el propio A). Lo llamaremos *ideal generado* por X y lo representaremos por (X). También se dice que el conjunto X es un *generador* del ideal (X).

Así, para todo subconjunto X de A tenemos que (X) es un ideal de A, $X \subset (X)$ y si I es un ideal de A tal que $X \subset I$, entonces $(X) \subset I$. Otro hecho obvio es que si $X \subset Y \subset A$, entonces $X \subset (Y)$, luego $(X) \subset (Y)$.

Cuando el conjunto X es finito, $X = \{x_1, \ldots, x_n\}$, el ideal generado por X se representa por $(x_1, \ldots, x_n)$. Entonces se dice que el ideal está *finitamente generado.*

El teorema siguiente nos da la forma de los elementos de un ideal a partir de sus generadores.

Teorema 3.4 *Sea A un dominio y $X \subset A$. Entonces*

$$(X) = \{a_1x_1 + \cdots + a_nx_n \mid n \in \mathbb{N}, a_i \in A, x_i \in X\}$$

En particular si $x \in A$, entonces $(x) = Ax = \{ax \mid a \in A\}$.

Demostración: Se comprueba sin dificultad que el conjunto de la derecha es un ideal de A y claramente contiene a X, luego

$$(X) \subset \{a_1x_1 + \cdots + a_nx_n \mid n \in \mathbb{N}, a_i \in A, x_i \in X\}.$$

Por otra parte (X) ha de contener a los elementos de la forma ax, con x en X, y por ser un subanillo a las sumas de estos elementos, luego se da la igualdad.

Si X tiene un sólo elemento x, las sumas $\sum_{i=1}^{n} a_ix = \big(\sum_{i=1}^{n} a_i\big)x$ están en el conjunto $\{ax \mid a \in A\}$, luego $(X) \subset \{ax \mid a \in A\}$. La otra inclusión es obvia. ∎

Así pues, los ideales principales son los ideales generados por un único elemento del anillo.[2]

El teorema siguiente nos proporciona ejemplos de dominios que no son DIP:

Teorema 3.5 *Sea A un dominio íntegro. Entonces $A[x]$ es DIP si y sólo si A es un cuerpo.*

Demostración: Si A es un cuerpo sabemos que $A[x]$ es un dominio euclídeo, luego es un DIP. Recíprocamente, si $A[x]$ es DIP, sea $a \in A$ un elemento no nulo y veamos que es una unidad en A. Para ello consideramos el ideal (x, a) de $A[x]$. Como ha de ser un ideal principal existe un polinomio $p \in A[x]$ tal que $(x, a) = (p)$, luego $a = pq$ para cierto $q \in A[x]$, pero entonces $\operatorname{grad} p + \operatorname{grad} q = \operatorname{grad} a = 0$, luego $\operatorname{grad} p = 0$ y por tanto $p \in A$. Por otra parte también $x = pr$, para cierto $r \in A[x]$, pero entonces el coeficiente director de x, que es 1, es el producto de p por el coeficiente director de r, luego p es una unidad y $(p) = A[x]$.

Entonces $1 \in (p) = (x, a)$, luego $1 = ux + va$, para ciertos polinomios $u, v \in A[x]$. Sin embargo el término independiente de ux es 0 y el de va es ba, donde b es el término independiente de v. Resulta, pues, que $1 = ba$, con lo que a es una unidad en A. ∎

En particular $\mathbb{Z}[x]$ no es DIP, ni en particular euclídeo. Lo mismo vale para $A[S]$ cuando S tiene más de un elemento (pues $A[S] = A\big[S \setminus \{x\}\big][x]$ y $A[S \setminus \{x\}]$ no es un cuerpo).

Ejercicio: Probar que $(x, 2)$ no es un ideal principal de $\mathbb{Z}[x]$, y que (x, y) no es un ideal principal de $\mathbb{Q}[x, y]$.

[2] Pero esto no contradice que un mismo ideal principal pueda tener varios generadores. Por ejemplo, en $\mathbb{Z}$ es claro que $(3) = (-3)$. El ideal admite dos generadores, pero *sólo es necesario uno* para generarlo.

Esto responde a un problema que habíamos dejado planteado al final del capítulo anterior: como $\mathbb{Z}[x]$ no es DIP, podemos asegurar que no es un dominio euclídeo. Así pues, no es posible dividir en general polinomios con coeficientes enteros en los términos del teorema 2.45, de modo que la restricción allí impuesta de que el coeficiente director del divisor sea una unidad no puede eliminarse. Tampoco es un dominio euclídeo $\mathbb{Q}[x, y]$, por la misma razón.

Sin embargo, estos anillos, como $\mathbb{Z}[x]$ o $\mathbb{Q}[x, y]$, aunque no son DIP, cumplen una propiedad más débil de interés:

Definición 3.6 Un dominio íntegro A es un anillo *noetheriano* si todo ideal de A es finitamente generado.

Evidentemente, todo DIP es un anillo noetheriano.

Teorema 3.7 *Sea A un dominio íntegro. Son equivalentes:*

1. *A es un anillo noetheriano.*
2. *Para toda cadena ascendente de ideales de A*

$$I_0 \subset I_1 \subset I_2 \subset I_3 \subset \ \ldots$$

   existe un número natural n tal que $I_n = I_m$ para todo $m \geq n$.

3. *Toda familia no vacía de ideales de A tiene un maximal[3] para la inclusión.*

DEMOSTRACIÓN: Si $I_0 \subset I_1 \subset I_2 \subset I_3 \subset \ \ldots$ es una cadena ascendente de ideales de A, es fácil ver que la unión $\bigcup_{i=0}^{\infty} I_i$ es también un ideal de A. Si A es noetheriano ha de tener un generador finito X. Cada elemento de X está en uno de los ideales I_i, y como X es finito y los ideales forman una cadena, existirá un natural n tal que $X \subset I_n$, pero entonces $\bigcup_{i=0}^{\infty} I_i = (X) \subset I_n$, lo que implica que $I_i = I_n$ para todo $i \geq n$. Por tanto 1) implica 2).

Si una familia no vacía de ideales de A no tuviera maximal, se podría extraer[4] una cadena ascendente de ideales que contradijera 2), luego 2) implica 3).

Si A tuviera un ideal I que no admitiera un generador finito, entonces, dado cualquier elemento a_0 de I, se cumple que $(a_0) \neq I$, luego existe un elemento $a_1 \in I \setminus (a_0)$, luego $(a_0) \subset (a_0, a_1) \neq I$, y de esta forma podemos conseguir[5] una cadena de ideales $(a_0) \subset (a_0, a_1) \subset (a_0, a_1, a_2) \subset \ldots$ sin que ninguno de ellos sea maximal. Por lo tanto 3) implica 1). ∎

[3]En general, en un conjunto parcialmente ordenado X, un elemento maximal m es un elemento tal que no existe otro $x \in X$ tal que $m < x$. Aquí estamos considerando a la inclusión como un orden parcial en la familia de los ideales de A, y si $\mathcal{F}$ es una familia de ideales de A, un ideal $I \in \mathcal{F}$ es maximal para la inclusión ni no existe otro $J \in \mathcal{F}$ tal que $I \subsetneq J$.

[4]Éste es el primer uso que hacemos del axioma de elección, pero en la versión débil del Principio de elecciones dependientes (A.3). En efecto, lo aplicamos tomando como A una familia de ideales $\mathcal{F}$ y como R la relación dada por $I\,R\,J$ si y sólo si $I \subsetneq J$. La hipótesis de que $\mathcal{F}$ no tiene maximal equivale a la hipótesis que requiere el Principio de elecciones dependientes, que a su vez nos da la cadena ascendente requerida.

[5]Técnicamente aquí se aplica también el principio de elecciones dependientes. Ahora el conjunto A es la familia de las funciones $s : I_n^* \longrightarrow I$ tales que $s(i+1) \in I \setminus (s(0), \ldots, s(i))$ para cada $i < n$, y la relación R es la inclusión estricta.

Hemos visto que, aunque $\mathbb{Z}$ o $\mathbb{Q}$ sean DIP, eso no implica que $\mathbb{Z}[x]$ o $\mathbb{Q}[x,y]$ lo sean. En cambio, eso no pasa con la propiedad de Noether:

Teorema 3.8 (Teorema de Hilbert) *Si A es un anillo noetheriano entonces $A[x_1,\dots,x_n]$ también lo es.*

Demostración: Basta probar que si A es noetheriano también lo es $A[x]$, pues esto implica que lo es $A[x_1]$, luego también $A[x_1][x_2] = A[x_1,x_2]$, y así llegamos hasta $A[x_1,\dots,x_n]$.

Sea I un ideal de $A[x]$. Sea I_i el conjunto de los coeficientes directores de los polinomios de I de grado i (más el 0).

Es claro que I_i es un ideal de A, así como que $I_0 \subset I_1 \subset I_2 \subset I_3 \subset \cdots$ (para ver que un elemento de I_i está en I_{i+1} basta multiplicar por x el polinomio que justifica que está en I_i). Como A es noetheriano, los ideales I_i son iguales a partir de un I_r.

Sea $I_i = (b_{i1},\dots,b_{in})$ para $i = 0,\dots,r$ (no es restricción suponer que el número de generadores es el mismo para todos los ideales, pues siempre podemos añadir generadores redundantes). Podemos suponer también que los $b_{ij} \neq 0$.

Sea p_{ij} un polinomio en I de grado i cuyo coeficiente de grado i sea b_{ij}. Vamos a probar que $I = (p_{ij} \mid i = 0,\dots,r,\ j = 1,\dots,n)$. Claramente este ideal está contenido en I.

Sea $f \in I$ un polinomio de grado d. Veremos que está en el ideal generado por los p_{ij} por inducción sobre d. El coeficiente director de f está en I_d. Si $d > r$ notamos que los coeficientes directores de $x^{d-r}p_{r1},\dots,x^{d-r}p_{rn}$ son los números $b_{r1},\dots,b_{rn}$, que generan $I_d = I_r$. Por consiguiente, existen elementos $c_1,\dots,c_n$ en A tales que el coeficiente director de f coincide con el de

$$c_1x^{d-r}p_{r1} + \cdots + c_nx^{d-r}p_{rn}.$$

Por consiguiente, el polinomio $f - c_1x^{d-r}p_{r1} - \cdots - c_nx^{d-r}p_{rn}$ tiene grado menor que d (ya que los términos de mayor grado se cancelan) y está en I, luego por hipótesis de inducción f está en el ideal generado por los p_{ij}.

Si $d \leq r$ obtenemos un polinomio $f - c_1p_{d1} - \cdots - c_np_{dn}$ de grado menor que d y contenido en I, con lo que se concluye igualmente. ∎

Terminamos esta sección observando que entre los ideales de un anillo se puede definir una suma y un producto como sigue:

Definición 3.9 Sea A un anillo y sean $S_1,\dots,S_n$ subconjuntos de A. Llamaremos

$$\begin{aligned} S_1 + \cdots + S_n &= \{s_1 + \cdots + s_n \mid s_i \in S_i \text{ para } \ i = 1,\dots,n\},\\ S_1 \cdots S_n &= \Big\{\sum_{i=1}^{m} s_{i1}\cdots s_{in} \mid m \in \mathbb{N} \text{ y } s_{ij} \in S_j \text{ para } j = 1,\dots,n\Big\}. \end{aligned}$$

Es pura rutina comprobar que la suma y el producto de ideales de A vuelve a ser un ideal de A. Además son operaciones asociativas, conmutativas y distributivas, es decir, $P(Q+R) = PQ + PR$. De la definición de ideal se sigue que $PQ \subset P \cap Q$.

3.2 Divisibilidad en dominios íntegros

Es bien conocido que todo número natural (mayor que 1) se descompone de forma única salvo el orden en producto de números primos. Por ejemplo, la descomposición de 360 es

$$360 = 2^3 \cdot 3^2 \cdot 5.$$

Demostraremos este hecho en la sección siguiente, mientras que en ésta introduciremos todos los conceptos necesarios para enunciar el teorema con precisión y con un cierto grado de generalidad. Empezamos con los conceptos de múltiplo y divisor, que ya hemos introducido de pasada al discutir la definición de ideal:

Definición 3.10 Sea A un dominio íntegro y a, b dos elementos de A. Diremos que a *divide* a b, o que a es un *divisor* de b, o que b es un *múltiplo* de a, o que b es *divisible* entre a (y lo representaremos $a \mid b$) si existe un elemento c de A tal que $b = ac$.

Por ejemplo en $\mathbb{Z}$ tenemos que 3 divide a 15, pero no a 16.

Existen algunos casos triviales de divisibilidad que conviene tener presentes:

- 0 no divide a ningún elemento de A salvo a sí mismo.
- Todo elemento de a divide[6] a 0.
- Si $u \in A$ es una unidad, entonces divide a todo $a \in A$, pues $a = u(u^{-1}a)$.
- Si $u \in A$ es una unidad, sus únicos divisores son las demás unidades de A, pues si $a \mid u$, entonces existe un b en A tal que $u = ab$, luego $1 = abu^{-1}$, luego a es una unidad.

Es obvio que $a \mid a$ y que si $a \mid b$ y $b \mid c$ entonces $a \mid c$. Esto significa que la divisibilidad es una relación reflexiva y transitiva. De hecho está cerca de determinar una relación de orden parcial, pero en realidad no es antisimétrica. Por el contrario, el fallo de la antisimetría da lugar a un concepto no trivial:

Diremos que dos elementos a y b de A son *asociados* si $a \mid b$ y $b \mid a$.

Por ejemplo, en $\mathbb{Z}$ tenemos que 3 y -3 son asociados. En general:

Dos elementos no nulos a, b de un dominio íntegro A son asociados si y sólo si existe una unidad $u \in A$ tal que $a = ub$.

En efecto, si a y b son asociados, entonces $a = ub$ y $b = va$, para ciertos u y v del anillo A. Por lo tanto $a = uva$, de donde $uv = 1$, luego u y v son unidades. El recíproco es trivial, teniendo en cuenta que $a = ub$ equivale a $b = u^{-1}a$.

[6] Lingüísticamente, existe una contradicción aparente entre el hecho de que A, por ser un dominio íntegro, no tiene divisores de 0 y que todos los elementos de A dividen a 0. Lo que sucede es que todos los elementos de A dividen a 0 "trivialmente", es decir, con cociente 0, mientras que un "divisor de 0" es, por definición, un elemento que divide a 0 "no trivialmente", con cociente no nulo.

Es claro que la asociación en un dominio íntegro A es una relación de equivalencia. Si A tiene un número finito de unidades, $u_1, \ldots, u_n$, entonces los asociados de un $a \in A$ no nulos serán $au_1, \ldots, au_n$ (entre los que está ya el propio a, ya que una de las unidades será la identidad). Así pues, las clases de equivalencia para la relación de asociación son la formada sólo por el 0 y las restantes, que tendrán n elementos cada una. Como en $\mathbb{Z}$ hay dos unidades, ± 1, los asociados no nulos en $\mathbb{Z}$ se agrupan por parejas: ± 1, ± 2, ± 3, etc.

El concepto de asociación es importante por el motivo siguiente: si queremos enunciar un teorema de descomposición única en factores primos válido para $\mathbb{Z}$, nos vemos obligados a relajar la condición de unicidad, ya que, por ejemplo,

$$360 = 2 \cdot 2 \cdot 2 \cdot 3 \cdot 3 \cdot 5 = (-2) \cdot 2 \cdot 2 \cdot 3 \cdot (-3) \cdot 5.$$

Aquí tenemos dos descomposiciones en factores primos que no se diferencian únicamente en el orden de los factores. Si quisiéramos enunciar el teorema para números enteros podríamos decir que la descomposición es única "salvo signos", pero si queremos un enunciado válido para anillos más generales la fórmula correcta es "salvo unidades" o "salvo asociados".

En efecto, la propia definición de "asociado" implica que dos elementos asociados tienen los mismos múltiplos y divisores, por lo que si un primo aparece en una descomposición en factores, ésta se puede modificar para que aparezca en su lugar cualquiera de sus asociados, tal y como hemos hecho con la factorización del 360, en la que hemos cambiado un 2 por un -2 a cambio de cambiar también un 3 por un -3.

Hay que tener la misma precaución al definir el concepto de "primo", pues cuando trabajamos con números naturales podemos decir que 5 es primo porque sus únicos divisores son 1 y 5, pero si trabajamos con números enteros resulta que 5 tiene cuatro divisores: ± 1 y ± 5.

En general, en cualquier dominio íntegro A, tenemos que todo elemento $a \in A$ tiene por divisores a las unidades de A y a sus propios asociados (entre los que figura él mismo). A estos divisores los llamaremos *divisores impropios* de a. Cualquier otro divisor es un *divisor propio.*

Por ejemplo, los divisores impropios de 4 en $\mathbb{Z}$ son 1, -1, 4 y -4. Sus divisores propios son 2 y -2. En este sentido podemos decir que 5 sólo tiene (en $\mathbb{Z}$) los divisores impropios.

Con estas precauciones, ya podemos dar una definición razonable de elemento irreducible en un anillo:

Un elemento a de un dominio íntegro A es *irreducible* en A si es no nulo, no es una unidad y no admite ninguna descomposición $a = bc$ con b y c elementos de A, salvo que uno de ellos sea una unidad (y, por lo tanto, el otro es un asociado de a).

Equivalentemente, un elemento (no nulo ni unidad) es irreducible si sus únicos divisores son los impropios. Es obvio que un elemento es irreducible si y sólo si lo es cualquiera de sus asociados.

En este sentido podemos decir que 5 (al igual que -5) es irreducible en $\mathbb{Z}$. Es importante que decimos "irreducible" y no primo, la definición de "primo" la daremos un poco más adelante y, aunque ambas serán equivalentes en $\mathbb{Z}$, en otros anillos no lo serán, por lo que el lector debe acostumbrarse a que "no tener divisores propios" es la definición de elemento irreducible y no la de elemento primo. Observemos que dicha definición excluye también al cero y a las unidades, de modo que 1 y -1 no son irreducibles en $\mathbb{Z}$, por definición.

Los elementos no nulos ni unitarios que no son irreducibles se dicen *reducibles*. Así pues, 1 y -1 no son ni reducibles ni irreducibles, sino que en todo dominio íntegro tenemos el 0, las unidades, los elementos irreducibles y los reducibles.

Ahora ya podemos introducir la noción fundamental que perseguíamos:

Un dominio íntegro A es un *dominio de factorización única* (DFU) si todo elemento a de A no nulo y que no sea una unidad se descompone como producto de elementos irreducibles $a = c_1 \cdots c_n$ y la descomposición es única salvo ordenación o cambio por asociados (es decir, si $a = c_1 \cdots c_n = d_1 \cdots d_m$ son dos descomposiciones de a en elementos irreducibles, entonces $m = n$ y, ordenando los factores adecuadamente, cada c_i es asociado a d_i).

En la sección siguiente demostraremos que $\mathbb{Z}$ es un DFU en este sentido. El resto de esta sección lo dedicaremos introducir algunos conceptos más relacionados con las descomposiciones en irreducibles.

Si A es un DFU y a es un elemento no nulo ni unitario, para cada elemento irreducible p de A llamaremos *exponente* de p en a al número de veces que p o sus asociados aparecen en cualquier descomposición de a en factores irreducibles (puede ser igual a 0). Lo denotaremos por $v_p(a)$.

Por ejemplo, teniendo en cuenta la descomposición de 360 en factores irreducibles que hemos mostrado antes, se cumple que

$$v_2(360) = 3, \qquad v_3(360) = 2, \qquad v_5(360) = 1, \qquad v_p(360) = 0 \quad \text{para } p \geq 7.$$

En general, en una descomposición de un $a \in A$ aparecerán $v_p(a)$ factores asociados a p, es decir, factores de la forma up donde u es una unidad. Si multiplicamos todas las unidades que así aparecen, resulta que a admite una descomposición en la forma

$$a = u \cdot p_1^{n_1} \cdots p_n^{n_m},$$

donde los p_i son irreducibles distintos, $n_i = v_{p_i}(a)$ y u es una unidad.

La presencia de u es necesaria, pues, por ejemplo, la única forma de factorizar en $\mathbb{Z}$ el -25 de este modo es $-25 = (-1)5^2$. Lo importante es que cada p aparece siempre con exponente $v_p(a)$ en virtud de la unicidad de la factorización.

Además el exponente de un irreducible en un elemento a es por definición el mismo que el de sus asociados, y el exponente de un irreducible en un elemento a es el mismo que en los asociados de a (pues una factorización de un asociado de a se obtiene multiplicando una factorización de a por una unidad, sin cambiar los irreducibles).

La factorización en irreducibles de un producto puede obtenerse como el producto de las factorizaciones de los factores, de donde se sigue la relación

$$v_p(ab) = v_p(a) + v_p(b).$$

Podemos definir $v_p(a) = 0$ para todo irreducible p cuando a es una unidad, y así la relación anterior es válida también si a o b es una unidad.

Notemos que un irreducible p divide a un elemento a si y sólo si $v_p(a) \neq 0$.

En efecto, si $v_p(a) \neq 0$ eso significa que p aparece en una factorización de a, luego $p \mid a$. Por otra parte si $p \mid a$, entonces $a = pb$ para cierto elemento b, luego $v_p(a) = v_p(p) + v_p(b) = 1 + v_p(b) \neq 0$.

Si $a \mid b$, ha de cumplirse que $v_p(a) \leq v_p(b)$ para todo irreducible p de A. La condición es también suficiente, pues si se cumple esto, entonces b se obtiene como producto de a por el producto de todos los irreducibles p que dividen a b elevados al exponente $v_p(b) - v_p(a)$ (y una unidad adecuada).

Dos elementos a y b son asociados si y sólo si $v_p(a) = v_p(b)$ para todo irreducible p de A.

Como consecuencia de estos hechos tenemos que en un DFU, si p es irreducible y $p \mid ab$, entonces $p \mid a$ o $p \mid b$.

En efecto, estamos suponiendo que $0 \neq v_p(a) + v_p(b)$, luego una de los dos exponentes ha de ser no nulo.

Este hecho resulta ser muy importante en la teoría de la divisibilidad, hasta el punto de que conviene introducir un nuevo concepto para comprenderlo adecuadamente:

Si A es un dominio íntegro, diremos que un elemento p de A es *primo* si es no nulo, no es una unidad y cuando $p \mid ab$ entonces $p \mid a$ o $p \mid b$, para todos los elementos a y b de A.

Ya hemos probado la mitad del siguiente teorema fundamental:

Teorema 3.11 *Sea A un dominio íntegro.*

1. *Todo primo de A es irreducible.*
2. *Si A es DFU, un elemento de A es primo si y sólo si es irreducible.*

Demostración: Efectivamente, si p es primo y se descompone como $p = ab$, entonces $p \mid a$ o $p \mid b$, pero como $a \mid p$ y $b \mid p$, lo que tenemos es que p es asociado con a o con b, lo que implica que el otro es una unidad. La segunda afirmación ya está probada. ∎

Así pues, una vez sabemos que un dominio íntegro es DFU, ya podemos hablar de descomposiciones en factores primos en lugar de descomposiciones en factores irreducibles, pues ambos conceptos coinciden. Pero hay que recordar que en ausencia de factorización única "primo" es un concepto más fuerte que "irreducible". En la sección 9.3 de [ITAl] tenemos ejemplos de esta situación.

3.3 Ideales y divisibilidad

Como se desprende, por ejemplo, del capítulo XIII de [ITAl], los ideales proporcionan el lenguaje idóneo para expresar los hechos más relevantes de la divisibilidad en un anillo. Como sabemos, si a es un elemento de un dominio íntegro A, entonces el ideal $(a) = Aa$ es precisamente el conjunto de todos los múltiplos de a.

Es claro que $a \mid b$ equivale a $(b) \subset (a)$, de donde se sigue que a y b son asociados si y sólo si $(a) = (b)$, es decir, si y sólo si generan el mismo ideal.

Hemos de pensar que dos elementos asociados son una misma cosa a efectos de divisibilidad (ambos tienen los mismos múltiplos y divisores). Ahora vemos que a cada familia de elementos asociados de un dominio íntegro le corresponde un único ideal principal. En particular el 0 se corresponde con el ideal $0 = (0)$ y las unidades de A se corresponden todas ellas con el ideal $A = (1)$.

El lector que quiera comprender adecuadamente la teoría de la divisibilidad debe esforzarse por llegar a entender que los ideales principales representan mejor que los elementos mismos del anillo los posibles divisores de un elemento dado. Quizá en esta dirección le ayude conocer un débil esbozo informal del modo en que el concepto de ideal era concebido cuando apareció en la teoría:

Consideremos las dos afirmaciones siguientes relativas a $\mathbb{Z}$. Por una parte $2 \mid 6$ y por otra $-2 \mid 6$. A efectos de divisibilidad ambas son equivalentes, puesto que 2 y -2 son asociados. Podemos resumirlas en una sola si consideramos que es el ideal $(2) = (-2)$ el que divide a 6, y escribimos en consecuencia $(2) \mid 6$. Podemos pensar que los divisores de los elementos de un dominio íntegro no son otros elementos del anillo, sino sus ideales. Así, podemos definir $(a) \mid b$ como $a \mid b$, lo cual no depende del generador elegido para el ideal, pues dos cualesquiera son asociados. Notemos que esto equivale a que $b \in (a)$, luego si I es un ideal principal tenemos (por definición) que $I \mid b \Leftrightarrow b \in I$.

Lo que hace de esto una idea brillante es que en realidad no tenemos por qué exigir a I que sea principal, con lo que cualquier ideal I puede dividir a un elemento en este sentido.

En un DIP cada 'divisor ideal' se corresponde con una familia de 'divisores reales' asociados (sus generadores), pero —como se ve en el capítulo XIII de [ITAl]— hay anillos no DIP en los que se puede hablar coherentemente de divisores ideales en este sentido sin que estén asociados a divisores reales, es decir, sin que sean principales. Tales 'divisores ideales' resultan esenciales para formular una teoría de divisibilidad razonable (y útil) en dichos anillos. De hecho, los ideales en el sentido moderno fueron introducidos por Dedekind a finales del siglo XIX para formalizar esta idea de divisor ideal que no se corresponde con ningún divisor real.

Más en general, podemos extender la relación de divisibilidad de modo que los ideales puedan dividirse entre sí. Podemos pensar que un ideal I divide a un ideal J si $J \subset I$ (comparar con $a \mid b \Leftrightarrow (b) \subset (a)$). De momento no entraremos en la teoría de divisores ideales, sino que nos limitaremos a desarrollar la teoría de divisibilidad en dominios íntegros mostrando su conexión con los ideales del

anillo. El lector debe tener presente que esta conexión se volverá esencial en capítulos posteriores, por lo que debe acostumbrarse a pensar e interpretar las cosas en términos de ideales en la medida de lo posible.

Como primer ejemplo del paso a términos de ideales, veamos el equivalente del concepto de elemento primo:

Definición 3.12 Un ideal P de un dominio A es *primo* si $P \neq A$ y para todo par de ideales I, J de A tales que $IJ \subset P$, se cumple que $I \subset P$ o $J \subset P$.

Si tenemos *in mente* la equivalencia $I \mid J \Leftrightarrow J \subset I$ vemos que la definición de ideal primo es paralela a la de elemento primo. La condición $P \neq A$ se corresponde con la exigencia de que los primos no sean unidades. Hay, no obstante, una discrepancia debida principalmente a motivos históricos, y es que, si bien hemos exigido que el elemento 0 no sea considerado primo, sí admitimos que el ideal 0 sea considerado primo. He aquí una caracterización práctica del concepto de ideal primo.

Teorema 3.13 *Un ideal P de un dominio A es primo si y sólo si $P \neq A$ y para todo par de elementos a, b de A, si $ab \in P$ entonces $a \in P$ o $b \in P$.*

Demostración: Si P es primo y $ab \in P$, entonces $(a)(b) \subset (ab) \subset P$, de donde resulta que $(a) \subset P$ o $(b) \subset P$, o sea, $a \in P$ o $b \in P$.

Recíprocamente, si $IJ \subset P$, pero I no está contenido en P, entonces existe un $a \in I \setminus P$. Ahora, si $b \in J$ tenemos que $ab \in IJ \subset P$, luego $a \in P$ o $b \in P$, y ha de ser $b \in P$, es decir, $J \subset P$. ∎

Ahora es inmediato que en un dominio íntegro A se cumple que un elemento no nulo a es primo si y sólo si el ideal (a) es un ideal primo. No obstante recordamos que el ideal trivial (0) es primo, aunque el elemento 0 no lo es por definición. Si un elemento es irreducible cuando no tiene divisores propios, el concepto análogo para ideales es el siguiente:

Definición 3.14 Un ideal M de un dominio A es un ideal *maximal* si $M \neq A$ y si I es un ideal de A tal que $M \subset I \subset A$, entonces $M = I$ o $I = A$.

Como en el caso de ideales primos, estamos admitiendo la posibilidad de que el ideal 0 sea maximal (si bien no tiene por qué serlo necesariamente).

Al contrario de lo que ocurre con el concepto de 'primo', no es cierto que un elemento a de un dominio íntegro A sea irreducible si y sólo si el ideal (a) es maximal. La situación es un poco más delicada. Concretamente a es irreducible si y sólo si (a) es maximal entre los ideales principales, es decir, si $(a) \neq A$ y cuando $(a) \subset (b) \subset A$, entonces $(a) = (b)$ o $(b) = A$.

En efecto, si a es irreducible y $(a) \subset (b) \subset A$, entonces $b \mid a$, luego o bien b es una unidad (y entonces $(b) = A$) o bien b es asociado de a (con lo que $(b) = (a)$). El recíproco es igual. Por lo tanto tenemos:

Teorema 3.15 *Sea A un dominio íntegro y $a \neq 0$ un elemento de A.*

1. *a es primo si y sólo si (a) es primo.*
2. *a es irreducible si y sólo si (a) es maximal entre los ideales principales de A.*
3. *Si A es DIP, entonces a es irreducible si y sólo si (a) es maximal.*

La tercera afirmación es inmediata, pues en un DIP los ideales maximales coinciden con los ideales maximales entre los ideales principales.

Hemos visto que todo elemento primo de un anillo es irreducible. Entre ideales podemos demostrar justo la implicación contraria:

Teorema 3.16 *En un dominio, todo ideal maximal es primo.*

DEMOSTRACIÓN: Si M es un ideal maximal en A y $ab \in M$, pero $a, b \notin M$, tendríamos que $M \subsetneq M + (a) \subset A$, luego la maximalidad de M implica que $M + (a) = A$. Por lo tanto $1 = m + xa$ para cierto $m \in M$ y cierto $x \in A$. Así pues $b = mb + xab \in M$, con lo que tenemos una contradicción. ∎

Ahora estamos en condiciones de probar dos hechos clave.

Teorema 3.17 *En un DIP los ideales maximales coinciden con los ideales primos no triviales y los elementos irreducibles coinciden con los elementos primos.*

DEMOSTRACIÓN: Si A es un DIP y (a) es un ideal primo no trivial, supongamos que (b) es un ideal tal que $(a) \subset (b) \subset A$. Entonces $a = bc$ para cierto $c \in A$. Como (a) es primo se ha de cumplir o bien $b \in (a)$ (en cuyo caso $(a) = (b)$) o bien $c \in (a)$, en cuyo caso $c = da$ para cierto $d \in A$, y así $a = bc = bda$, luego (dado que $a \neq 0$), $bd = 1$, o sea, b es una unidad y por lo tanto $(b) = A$.

La segunda afirmación se sigue de la primera y de 3.15 ∎

Con esto podemos probar el resultado principal de esta sección. Diremos que un dominio íntegro A tiene la *propiedad de factorización* si todo elemento de A no nulo ni unidad se descompone en producto de irreducibles.

Teorema 3.18 *Todo anillo noetheriano A tiene la propiedad de factorización. Si además todo elemento irreducible de A es primo, entonces A es DFU. En particular todo DIP es DFU.*

DEMOSTRACIÓN: Sea A un anillo noetheriano. Llamemos S al conjunto de los elementos de A no nulos ni unidades pero que no admitan una descomposición en irreducibles. Hemos de probar que S es vacío.

Si existe un elemento a en S, entonces a no es unidad, luego $(a) \neq A$. Si a fuera irreducible entonces él mismo sería una descomposición en irreducibles, luego no lo es. Podemos factorizar $a = bc$ donde ni b ni c es una unidad (ni 0). Si ninguno estuviera en S entonces se descompondrían en producto de irreducibles, y a también. Por tanto al menos uno de los dos está en S. Digamos que $b \in S$. Como $b \mid a$ se cumple que $(a) \subset (b)$. La inclusión es estricta, pues si $(a) = (b)$

entonces a y b serían asociados, es decir, $a = bu$ para cierta unidad u, pero entonces $bu = bc$, luego $c = u$ sería una unidad, cuando no lo es.

En definitiva hemos probado que para cada $a \in S$ existe un $b \in S$ tal que $(a) \subsetneq (b)$. Repitiendo este proceso obtendríamos una sucesión creciente de ideales $(a_0) \subsetneq (a_1) \subsetneq (a_2) \subsetneq \cdots$ en contradicción con el teorema 3.7. Por lo tanto S ha de ser vacío y así todo elemento no nulo ni unitario de A admite una descomposición en irreducibles.

Supongamos que los irreducibles coinciden con los primos y que tenemos dos descomposiciones en irreducibles de un mismo elemento $a = c_1 \cdots c_n = d_1 \cdots d_m$. Podemos suponer que $m \leq n$.

Como d_m es primo, ha de dividir a uno de los factores de $c_1 \cdots c_n$ y como éstos son irreducibles, de hecho ha de ser asociado a uno de ellos. Pongamos que d_m es asociado a c_n. Entonces $c_n = u_m d_m$ para cierta unidad u_m.

Simplificando d_m obtenemos que $c_1 \cdots c_{n-1} u_m = d_1 \cdots d_{m-1}$. Repitiendo el proceso con d_{m-1} (y teniendo en cuenta que un irreducible no puede dividir a una unidad), llegamos tras m pasos a que $c_1 \cdots c_{n-m}\, u_1 \cdots u_m = 1$, lo que obliga a que $n = m$, pues ningún irreducible puede dividir a 1. Además hemos obtenido que cada c_i es asociado a d_i, luego la descomposición es única. ∎

Con esto tenemos probada la factorización única de $\mathbb{Z}$ y de los anillos $K[x]$ donde K es un cuerpo. Para el caso de $\mathbb{Z}$ es posible dar argumentos directos más elementales basados en el buen orden de $\mathbb{N}$. Por ejemplo, para encontrar un factor irreducible de un número entero basta tomar el menor natural que lo divide. Lo mismo ocurre con $K[x]$ considerando el grado de los polinomios.

Para terminar probamos un resultado general sobre existencia de ideales maximales. Es el primer teorema que probamos que depende del axioma de elección (véase el apéndice A), aunque éste no es necesario si el anillo es noetheriano, pues el teorema 3.7 nos da en la prueba siguiente el ideal maximal que, en general, nos da el lema de Zorn:

Teorema 3.19 (AE) *Todo anillo unitario A en el que $1 \neq 0$ tiene un ideal maximal. Más aún, todo ideal $I \neq A$ está contenido en un ideal maximal de A.*

Demostración: Dado un anillo unitario A, basta aplicar el lema de Zorn a la familia de ideales de A distintos de A y que contienen a I, considerada como conjunto parcialmente ordenado por la inclusión. Si $\mathcal{C}$ es una familia de tales ideales totalmente ordenada por la inclusión, es fácil ver que $\bigcup \mathcal{C}$ es un ideal de A que contiene a I, y es distinto de A porque no puede contener a 1 (es en este punto donde se requiere que A sea unitario). El lema de Zorn nos da entonces un ideal maximal de A que contiene a I. ∎

3.4 Divisibilidad en DFUs

En $\mathbb{Z}$ podemos afinar la unicidad de la descomposición en factores primos exigiendo que éstos sean positivos, es decir, números naturales. Así, la descomposición en primos del número 60 es $60 = 2 \cdot 2 \cdot 3 \cdot 5$, y no consideraremos

otras como $2 \cdot 2 \cdot (-3) \cdot (-5)$. Si no se indica lo contrario, cuando hablemos de primos en $\mathbb{Z}$ nos referiremos a naturales primos. Supondremos que el lector está familiarizado con las técnicas para descomponer en factores primos enteros pequeños.

Otro hecho básico sobre la aritmética de los números enteros es que hay infinitos números primos. Para probarlo observemos que en general un primo p no puede dividir al mismo tiempo a un número n y a $n+1$, pues entonces dividiría a su diferencia, que es 1. De hecho, en $\mathbb{Z}$, si p divide a n, el próximo número al que divide es $n+p$. Sabiendo esto demostramos:

Teorema 3.20 (Euclides) *En $\mathbb{Z}$ hay infinitos números primos.*

DEMOSTRACIÓN: Dado un número $n > 0$, consideremos $n!$ Se cumple que todo número natural menor o igual que n divide a $n!$, luego ningún número menor o igual que n divide a $n!+1$. En consecuencia un divisor primo de $n!+1$ ha de ser mayor que n. Por lo tanto por encima de cada número n hay siempre un número primo. Esto implica que hay infinitos primos. ∎

Definición 3.21 Sea A un dominio íntegro y X un subconjunto de A. Diremos que un elemento d de A es un *máximo común divisor* (mcd) de los elementos de X si d divide a los elementos de X y cualquier elemento de A que cumpla lo mismo es un divisor de d.

Diremos que un elemento m de A es un *mínimo común múltiplo* (mcm) de los elementos de X si es múltiplo de todos los elementos de X y todo elemento de A que cumpla lo mismo es un múltiplo de m.

Es obvio que m es un mcd o un mcm de X si y sólo si lo es cualquiera de sus asociados, es decir, estos conceptos son únicos salvo unidades. Por supuesto el mcd o el mcm de un conjunto dado no tiene por qué existir. No obstante, cualquier subconjunto finito de un DFU tiene mcd y mcm. El lector puede entretenerse probando que las siguientes "recetas" nos dan un mcd y un mcm de cualquier subconjunto finito X de un DFU:

> *Un* mcd *de X está formado por el producto de los primos que dividen a todos los elementos de X elevados al mínimo exponente con el que aparecen en alguno de los elementos de X.*
>
> *Un* mcm *de X está formado por el producto de todos los primos que dividen a algún elemento de X elevados al mayor exponente con el que aparecen en los elementos de X.*

Por ejemplo, dados los números

$$2^2 \cdot 3 \cdot 7^5, \qquad 2 \cdot 5 \cdot 7, \qquad 3^4 \cdot 5^2 \cdot 7,$$

su mcd es 7 y el mcm es $2^2 \cdot 3^4 \cdot 5^2 \cdot 7^5$.

Escribiremos $\text{mcd}(a_1, \ldots, a_n)$ y $\text{mcm}(a_1, \ldots, a_n)$ para representar el mcd y el mcm de los elementos $a_1, \ldots, a_n$. A veces el mcd lo representaremos simplemente por $(a_1, \ldots, a_n)$.

En $\mathbb{Z}$ el mcd es único si lo exigimos positivo. Si no se indica lo contrario siempre lo supondremos así.

Hay que prestar un poco de atención al cero: por definición todo elemento de un anillo divide a 0, de donde se sigue fácilmente que el mcd de un conjunto de elementos que contenga a 0 es el mismo que el del conjunto que resulte de eliminarlo. Por otra parte si un conjunto de elementos contiene una unidad, su mcd es 1.

Los elementos de un conjunto son *primos entre sí* si su mcd es 1, es decir, si no tienen divisores primos comunes. No hay que confundir esto con que sean primos entre sí dos a dos, que es más fuerte. Si dividimos los elementos de un conjunto por su mcd obtenemos un conjunto de elementos primos entre sí, pues si d es el mcd y p es un primo que dividiera al conjunto resultante, entonces dp dividiría al conjunto original, luego $dp \mid d$ y p sería una unidad.

En particular, si A es un DFU, K es su cuerpo de cocientes y a, $b \in A$ son no nulos, podemos expresarlos como $a = da'$ y $b = db'$, donde $(a', b') = 1$ y entonces $a/b = a'/b'$, por lo que toda fracción de K se puede elegir *irreducible,* en el sentido de que el numerador y el denominador sean primos entre sí.

En el caso concreto de $\mathbb{Z}$ y $\mathbb{Q}$ concluimos que cada número racional admite una única representación como fracción irreducible con denominador positivo.

En $\mathbb{Z}$, nada nos impide considerar un ideal como $(6, 9)$, con dos generadores, pero sabemos que $\mathbb{Z}$ es DIP, luego dicho ideal tiene que ser principal, luego tiene que poder expresarse en términos de un único generador. Concretamente, sucede que $(6, 9) = (3)$, y el teorema siguiente lo prueba en un contexto general:

Teorema 3.22 (Relación de Bezout) *Sea A un DIP, sean $a_1, \ldots, a_n$ elementos de A y sea d un* mcd *de $a_1, \ldots, a_n$. Entonces $(d) = (a_1) + \cdots + (a_n)$, luego existen ciertos $r_1, \ldots, r_n \in A$ de manera que $d = r_1 a_1 + \cdots + r_n a_n$.*

Demostración: Sea $(d) = (a_1) + \cdots + (a_n)$ (por definición). Vamos a ver que d es un mcd de $a_1, \ldots, a_n$.

Como cada a_i está en (d), ciertamente $d \mid a_i$. Si s divide a todos los a_i, entonces $(a_i) \subset (s)$, luego $(d) = (a_1) + \cdots + (a_n) \subset (s)$, luego $s \mid d$.

Observemos que si d' es cualquier otro mcd de los elementos dados, entonces $(d') = (d)$, luego la relación de Bezout es válida para cualquiera de ellos. ∎

Este resultado se aplica especialmente a pares de elementos primos entre sí: si m y n son primos entre sí, existen r y s tales que $rm + sn = 1$.

En [ITAl 2.20] demostramos la relación de Bezout para dos elementos en un dominio euclídeo. Acabamos de ver que en realidad la propiedad relevante no es que el dominio sea eucídeo, sino que sea un DIP.

Terminamos esta sección recordando el teorema [ITAl 3.15]:

Teorema 3.23 *Sea p un número primo y $0 < m < p$. Entonces $p \mid \binom{p}{m}$.*

DEMOSTRACIÓN: Si $m = 1$, entonces $\binom{p}{m} = p$, luego efectivamente $p \mid \binom{p}{m}$. Si $p \mid \binom{p}{m}$ y $m + 1 < p$, entonces

$$\binom{p}{m+1} = \frac{p!}{(m+1)!\,(p-m-1)!} = \frac{p-m}{m+1}\,\frac{p!}{m!\,(p-m)!} = \frac{p-m}{m+1}\binom{p}{m}.$$

Así pues, $p \mid (p-m)\binom{p}{m}$, y como $(m+1, p) = 1$, la divisibilidad se conserva al dividir entre $m+1$, es decir, $p \mid \binom{p}{m+1}$. ∎

3.5 Divisibilidad en anillos de polinomios

El estudio de la divisibilidad no es tan sencillo si pasamos a los anillos de polinomios $A[x]$. Sabemos que $A[x]$ es un DFU cuando A es un cuerpo, pero en otro caso $A[x]$ no es DIP, pero ¿es igualmente un DFU cuando A es DFU?, por ejemplo, ¿es $\mathbb{Z}[x]$ un DFU?, ¿es $\mathbb{Q}[x, y]$ un DFU? Vamos a probar que la respuesta es afirmativa, pero para ello necesitamos un trabajo previo.

Si D es un DFU y K es su cuerpo de cocientes, vamos a probar que $D[x]$ es un DFU apoyándonos en que $K[x]$ lo es. La situación típica con la que tenemos que encontrarnos es la siguiente: El polinomio $6x^2 - 24$ factoriza en $\mathbb{Z}[x]$ como

$$6x^2 - 24 = 6(x^2 - 4) = 2 \cdot 3 \cdot (x - 2) \cdot (x + 2).$$

Vemos que tiene 4 divisores primos. Sin embargo, en $\mathbb{Q}[x]$ sólo tiene dos, pues los primeros factores pasan a ser unidades. Conviene dar la definición siguiente:

Definición 3.24 Sea D un DFU y sea $c : D[x] \longrightarrow \mathcal{P}D$ la aplicación que a cada polinomio $f \in D[x]$ le asigna el ideal generado por un mcd de sus coeficientes no nulos (no importa cuál, pues dos cualesquiera son asociados, luego generan el mismo ideal). Convenimos además en que $c(0) = 0$. A $c(f)$ se le llama *contenido* del polinomio f.

Diremos que un polinomio f es *primitivo* si $c(f) = (1)$, o sea, si sus coeficientes son primos entre sí. En particular, todo polinomio mónico es primitivo.

Por ejemplo, el contenido del polinomio $6x^2 - 24 \in \mathbb{Z}[x]$ es (6), mientras que el polinomio $x^2 - 4$ es primitivo. En general es inmediato que si $f \in D[x]$ y $f \neq 0$, entonces $f(x) = c\,g(x)$ donde $c \in D$ es un generador de $c(f)$ y $g(x) \in D[x]$ es un polinomio primitivo. Así, para probar que todo polinomio $f(x) \in D[x]$ se descompone en irreducibles basta probar que podemos factorizar por una parte polinomios constantes y por otra polinomios primitivos.

La factorización de las constantes es obvia, puesto que estamos suponiendo que D es un DFU. Notemos que todo $a \in D$ es irreducible en D si y sólo si lo es en $D[x]$. (Una descomposición de a en factores no unitarios de $D[x]$ tendría que constar de polinomios de grado 0, luego serían factores no unitarios de D, y el recíproco es obvio.)

Para probar que todo polinomio primitivo $p(x) \in D[x]$ se descompone en irreducibles observamos que los polinomios primitivos no son divisibles entre constantes no unitarias, ya que una constante que divida a $p(x)$ divide también a su contenido.

Así, si $p(x)$ (no unitario) no pudiera descomponerse en irreducibles en $D[x]$, en particular no sería irreducible, luego se descompondría en dos factores, digamos $p(x) = p_1(x)q_1(x)$, donde ninguno de los dos es una unidad, luego ambos tienen grado menor que el grado de $p(x)$. Al menos uno de los dos no podría descomponerse en irreducibles (digamos que $p_1(x)$), luego $p_1(x)$ no es irreducible y factoriza como $p_1(x) = p_2(x)q_2(x)$, donde ambos factores son no constantes, pues dividen a $p(x)$, luego el grado de $p_2(x)$ es menor que el de $p_1(x)$. De este modo obtenemos una sucesión de polinomios $p(x), p_1(x), p_2(x), \ldots$ cuyos grados son estrictamente decrecientes, lo cual es absurdo.

Con esto tenemos demostrado que todo polinomio de $D[x]$ no nulo ni unitario se descompone en producto de irreducibles. La parte delicada es demostrar la unicidad de la descomposición. La idea es usar la factorización única en D para probar la unicidad de los factores irreducibles constantes y la unicidad en $K[x]$ para probar la unicidad de los factores irreducibles no constantes. Necesitamos dos resultados sobre $c(f)$.

Teorema 3.25 *Sea D un DFU.*

1. *Si $a \in D$ y $f \in D[x]$, entonces $c(af) = (a)c(f)$.*
2. *Si f, $g \in D[x]$, entonces $c(fg) = c(f)c(g)$.*

Demostración: 1) Si $c(f) = (c)$, es inmediato que ac es un mcd de los coeficientes de af, luego $c(af) = (ac) = (a)(c)$.

2) Sea $f = c_f f_1$ y $g = c_g g_1$ con $c(f) = (c_f)$, $c(g) = (c_g)$ y f_1, g_1 primitivos. Entonces, por la propiedad precedente,

$$c(fg) = c\,(c_f f_1 \cdot c_g g_1) = c(f)c(g)c(f_1 g_1),$$

luego basta probar que $f_1 g_1$ es primitivo. Sean $f_1 = \sum_{i=0}^{n} a_i x^i$, $g_1 = \sum_{i=0}^{m} b_i x^i$, con $a_n \neq 0 \neq b_m$. Entonces $f_1 \cdot g_1 = \sum_{k=0}^{m+n} \big(\sum_{i+j=k} a_i b_j \big) x^k$.

Si $f_1 g_1$ no fuera primitivo, existiría un irreducible $p \in D$ que dividiría a todos sus coeficientes, es decir $p \mid \sum_{i+j=k} a_i b_j$, para cada k entre 0 y $n+m$.

Como f_1 es primitivo, p no divide a todos los a_i, luego existe un mínimo índice s tal que $p \mid a_i$ para $i < s$ y $p \nmid a_s$ (en particular $a_s \neq 0$).

Igualmente existe un mínimo t tal que $p \mid b_j$ para $j < t$ y $p \nmid b_t$ (luego $b_t \neq 0$).

Ahora, tomando $k = s+t$, resulta que p divide a $\sum_{i+j=k} a_i b_j$ y también divide a todos los sumandos salvo quizá a $a_s b_t$, de donde divide a la diferencia, o sea, a $a_s b_t$. Como p es primo divide a uno de los factores, lo que nos da una contradicción. ∎

Así, si un polinomio $f(x) \in D[x]$ no nulo ni unitario admite una descomposición en irreducibles de $D[x]$ de la forma $d_1 \cdots d_r \cdot p_1(x) \cdots p_s(x)$, donde los d_i son los factores constantes, entonces cada $p_i(x)$ es primitivo, pues en caso contrario sería divisible entre un polinomio constante irreducible. Por consiguiente

$$c(f) = (d_1 \cdots d_r).$$

Ahora, la factorización única en D garantiza que si f admite dos descomposiciones en irreducibles, los factores de grado 0 son los mismos salvo orden y asociación, pues ambos constituyen descomposiciones en irreducibles de un generador de $c(f)$ en D.

Para demostrar la unicidad del resto de la descomposición necesitamos dos resultados:

Teorema 3.26 *Sea D un DFU, sea K su cuerpo de cocientes y sean f, g polinomios primitivos en $D[x]$. Entonces f y g son asociados en $D[x]$ si y sólo si lo son en $K[x]$.*

Demostración: Si f y g son asociados en $K[x]$ entonces $f = gu$ para cierta unidad u de $K[x]$. Por el teorema 2.43, sabemos que u es un polinomio constante, es decir, está en K. Por lo tanto $u = r/s$ para ciertos r, $s \in D$ no nulos. Entonces $sf = rg$. Como $c(f) = c(g) = (1)$, tenemos que

$$(s) = (s) \cdot c(f) = c(sf) = c(rg) = (r) \cdot c(g) = (r),$$

luego $r = sv$ para cierta unidad $v \in D$. En consecuencia $u = r/s = v \in D$, luego f y g son asociados en $D[x]$. El recíproco es obvio. ∎

El teorema siguiente tiene gran interés en sí mismo:

Teorema 3.27 (Criterio de irreducibilidad de Gauss) *Sea D un DFU, sea K su cuerpo de cocientes y $f \in D[x]$ un polinomio primitivo no constante. Entonces f es irreducible en $D[x]$ si y sólo si lo es en $K[x]$.*

Demostración: Supongamos que f es irreducible en $D[x]$ pero $f = gh$ donde g, $h \in K[x]$ no son unidades. Entonces $\operatorname{grad} g \geq 1$, $\operatorname{grad} h \geq 1$. Sean

$$g = \sum_{i=0}^{n} \frac{a_i}{b_i} x^i, \qquad h = \sum_{i=0}^{m} \frac{c_i}{d_i} x^i,$$

para ciertos a_i, b_i, c_i, $d_i \in D$ con $b_i \neq 0 \neq d_i$. Llamemos $b = b_0 \cdots b_n$ y, para cada i, sea $b_i^* = b/b_i \in D$.

Consideremos el polinomio $g_1 = \sum_{i=0}^{n} a_i b_i^* x^i \in D[x]$. Podemos descomponer $g_1 = a g_2$, siendo $c(g_1) = (a)$ y $g_2 \in D[x]$ un polinomio primitivo. Claramente $g = \dfrac{1}{b} g_1 = \dfrac{a}{b} g_2$, luego $\operatorname{grad} g = \operatorname{grad} g_2$.

Del mismo modo se llega a que $h = \dfrac{c}{d}h_2$, donde c, $d \in D$ y $h_2 \in D[x]$ es primitivo y $\operatorname{grad} h = \operatorname{grad} h_2$. Ahora

$$f = gh = \frac{ac}{bd}\, g_2h_2,$$

con lo que f y g_2h_2 son dos polinomios primitivos en $D[x]$ asociados en $K[x]$. Por el teorema anterior son asociados en $D[x]$, luego existe una unidad $u \in D$ tal que $f = ug_2h_2$, y así f es reducible en $D[x]$, contradicción.

Si f es irreducible en $K[x]$ y $f = gh$ con g, $h \in D[x]$, entonces g o h es una unidad en $K[x]$, por ejemplo g, pero en la prueba del teorema anterior hemos visto que, en esta situación, de hecho g es una unidad en D, y esto prueba que f es irreducible en $D[x]$. ∎

Finalmente podemos probar un resultado anunciado tras [ITAl 2.31]:

Teorema 3.28 (Gauss) *Si D es un DFU y S un conjunto, entonces $D[S]$ es un DFU.*

Demostración: Veamos primeramente que $D[x]$ es DFU. Sea $f \in D[x]$ no nulo ni unidad. Hemos visto que f admite una descomposición en polinomios irreducibles en $D[x]$

$$f = d_1 \cdots d_r p_1(x) \cdots p_s(x),$$

donde los d_i son los factores de grado 0 y los $p_i(x)$ son necesariamente primitivos. También sabemos que los d_i son únicos salvo orden y asociación.

Por otra parte, si K es el cuerpo de cocientes de D, tenemos que los $p_i(x)$ son irreducibles en $K[x]$ y $d_1 \cdots d_r$ es una unidad en $K[x]$. Así, si f admite dos factorizaciones en $D[x]$, los correspondientes $p_i(x)$ han de ser los mismos salvo orden y asociación en $K[x]$, pero también sabemos que la asociación en $K[x]$ coincide con la asociación en $D[x]$ para polinomios primitivos de $D[x]$. Esto prueba la unicidad de la descomposición y nos da que $D[x]$ es un DFU.

Aplicando este resultado un número finito de veces obtenemos que $D[S]$ es un DFU cuando el conjunto S es finito. El caso general se reduce al finito del modo usual (dos factorizaciones en irreducibles en $D[S]$ serían también dos factorizaciones en irreducibles en un $D[S_0]$, con S_0 finito). ∎

Tenemos así demostrados dos resultados que, sin ser excesivamente complejos, no son triviales en absoluto y representan un papel relevante en la teoría que estamos desarrollando. Con las técnicas que hemos necesitado para llegar a ellos podemos probar fácilmente un criterio útil de irreducibilidad de polinomios:

Teorema 3.29 (Criterio de irreducibilidad de Eisenstein) *Sea D un DFU, sea K su cuerpo de cocientes, $f = \sum_{i=0}^{n} a_i x^i \in D[x]$ un polinomio no constante con $a_n \neq 0$ y $p \in D$ un primo. Supongamos que $p \mid a_i$ para todo $i = 0, \ldots, n-1$, así como que $p \nmid a_n$ y $p^2 \nmid a_0$. Entonces f es irreducible en $K[x]$ y, si es primitivo, en $D[x]$.*

DEMOSTRACIÓN: Sea $f = c_f f_1$ donde $f_1 \in D[x]$ es primitivo. Basta probar que f_1 es irreducible en $K[x]$, pues c_f es una unidad en K, luego f también será irreducible. El resto del teorema es consecuencia del criterio de Gauss.

También por el criterio de Gauss, basta probar que f_1 es irreducible en $D[x]$. Notemos que p no divide a c_f (porque no divide a a_n).

Cada coeficiente de f es el producto de c_f por el correspondiente coeficiente de f_1, luego f_1 sigue cumpliendo las hipótesis del teorema. Por no cambiar de notación podemos suponer que $f = f_1$ (pero ahora f es primitivo).

Supongamos que $f = gh$, donde $g = \sum_{i=0}^{r} b_i x^i$ y $h = \sum_{j=0}^{s} c_j x^j$. Ninguno de los dos puede ser constante o de lo contrario f no sería primitivo. Como $c(g)c(h) = c(f) = (1)$, tanto g como h son primitivos.

Tenemos que $p \mid a_0 = b_0 \cdot c_0$, luego p divide a uno de los factores. Pongamos por caso que $p \mid b_0$. Como $p^2 \nmid a_0$ no puede ser que p divida también a c_0.

Como g es primitivo, p no puede dividir a todos los b_i. Tomemos el menor natural k tal que $p \mid b_i$ para $i < k$ y $p \nmid b_k$. Así $1 \leq k \leq r < n$.

El coeficiente $a_k = \sum_{i+j=k} b_i c_j$ es divisible entre p y por otra parte p divide a todos los sumandos salvo quizá a $b_k c_0$, luego divide a la diferencia, que es justo $b_k c_0$. Sin embargo $p \nmid b_k$ y $p \nmid c_0$, contradicción. ∎

Ejercicio: Probar que en $\mathbb{Q}[x]$ hay polinomios irreducibles de grado arbitrariamente grande.

Aunque los polinomios a los que podemos aplicar el criterio de Eisenstein han de cumplir unas propiedades muy particulares, en realidad este criterio es útil en más ocasiones de las que en principio se podría pensar. Ello se debe al resultado siguiente:

Teorema 3.30 *Sea A un dominio. Sea a una unidad de A y b cualquier elemento de A. La aplicación $f : A[x] \longrightarrow A[x]$ dada por $f\big(p(x)\big) = p(ax + b)$ es un isomorfismo de anillos, luego en particular un polinomio $p(x)$ es irreducible en $A[x]$ si y sólo si $p(ax + b)$ lo es.*

DEMOSTRACIÓN: Claramente f es un homomorfismo porque no es sino la evaluación en $ax + b$. Es biyectivo porque tiene por inverso a la evaluación en $a^{-1}x - a^{-1}b$. ∎

Por ejemplo, para probar que el polinomio $p(x) = 8x^3 - 6x - 1$ es irreducible en $\mathbb{Z}[x]$, basta ver que lo es en $\mathbb{Q}[x]$, pero por el teorema anterior basta ver que lo es el polinomio $p(\frac{1}{2}\,x + \frac{1}{2}) = x^3 + 3x - 3$, que es irreducible en $\mathbb{Q}[x]$ por el criterio de Eisenstein.

La búsqueda de factores irreducibles de grado 1 de un polinomio equivale a la búsqueda de sus raíces, tal y como explicamos a continuación.

Definición 3.31 Sea A un dominio y $f(x)$ un polinomio en $A[x]$. Podemos considerar a f como una función $f : A \longrightarrow A$. En efecto, para cada elemento $a \in A$ tenemos definida la evaluación $f(a) \in A$. Diremos que a es una *raíz* del polinomio f si $f(a) = 0$.

La relación básica entre la divisibilidad y la existencia de raíces se sigue del siguiente teorema elemental:

Teorema 3.32 (Teorema del resto) *Sea A un anillo conmutativo y unitario, $f(x) \in A[x]$ y $c \in A$. Entonces existe un único polinomio $q(x) \in A[x]$ tal que $f(x) = q(x)(x-c) + f(c)$.*

DEMOSTRACIÓN: Si $f = 0$ basta tomar $q = 0$. En otro caso el resultado se sigue del teorema 2.45, que nos da dos polinomios $q(x)$ y $r(x)$ de manera que $f(x) = q(x)(x-c) + r(x)$ con el grado de r menor que el de $x - c$, o sea, r es de grado cero, luego constante. Sustituyendo x por c resulta que $f(c) = q(c)(c-c) + r$, o sea, $r = f(c)$. ∎

Como consecuencia:

Teorema 3.33 *Sea A un dominio íntegro. Sea $f(x) \in A[x]$ y $c \in A$. Entonces c es una raíz de $f(x)$ si y sólo si $(x-c) \mid f(x)$.*

DEMOSTRACIÓN: Si c es una raíz de $f(x)$ entonces $f(c) = 0$, luego el teorema anterior se reduce a que existe un polinomio $q(x)$ tal que $f(x) = q(x)(x-c)$, luego $(x-c) \mid f(x)$.

Si $(x-c) \mid f(x)$ entonces $f(x) = q(x)(x-c)$ para cierto polinomio $q(x)$. Por lo tanto $f(c) = q(c)(c-c) = 0$, es decir, c es una raíz de $f(x)$. ∎

De aquí que un polinomio irreducible de grado mayor que 1 no puede tener raíces. Por otro lado un polinomio de grado 1, $ax + b$ siempre tiene una raíz en un cuerpo, a saber, $-b/a$. Así obtenemos lo siguiente:

Si K es un cuerpo, un polinomio de grado 2 o 3 es irreducible en $K[x]$ si y sólo si no tiene raíces en K.

En efecto, si se pudiera descomponer en producto de irreducibles, al menos uno de sus factores irreducibles tendría grado 1, luego tendría una raíz en K. Para polinomios de grado mayor que 3 la no existencia de raíces ya no implica la irreducibilidad. Por ejemplo, el polinomio $(x^2+1)^2$ es reducible en $\mathbb{Q}[x]$ y no tiene raíces en $\mathbb{Q}$.

El teorema anterior implica un hecho muy importante sobre las raíces de un polinomio en un dominio íntegro:

Teorema 3.34 *Sea A un dominio íntegro y sea $f(x) \in A[x]$ un polinomio de grado n. Entonces $f(x)$ tiene a lo sumo n raíces en A.*

DEMOSTRACIÓN: Sean $c_1, \ldots, c_m$ raíces distintas de $f(x)$ en A. Por el teorema anterior tenemos que $(x-c_1) \mid f(x)$, es decir, $f(x) = (x-c_1)f_1(x)$.

Como c_2 es raíz de f tenemos que $0 = f(c_2) = (c_2 - c_1)f_1(c_2)$ y como las raíces son distintas $c_2 - c_1 \neq 0$. Como A es un dominio íntegro ha de ser $f_1(c_2) = 0$, luego por el teorema anterior $f_1(x) = (x-c_2)f_2(x)$, de donde $f(x) = (x-c_1)(x-c_2)f_2(x)$.

Repitiendo esto m veces tenemos que $(x-c_1)\cdots(x-c_m) \mid f(x)$, por lo que el grado de $(x-c_1)\cdots(x-c_m)$, que es m, ha de ser menor o igual que el grado de $f(x)$, que es n. ∎

Sin embargo el número de raíces de un polinomio no tiene por qué igualar a su grado. Por ejemplo, x^2+1 no tiene raíces en $\mathbb{Q}[x]$, pues el cuadrado de un número racional no puede ser negativo. Por otro lado $(x-1)^2$ tiene grado 2 y una única raíz.

Ejercicio: Probar que las raíces enteras de un polinomio de $\mathbb{Z}[x]$ dividen a su término independiente.

Hay un análogo al criterio de Gauss, pero referente a la existencia de raíces en lugar de la irreducibilidad. La prueba es mucho más sencilla [ITAl 2.29]:

Teorema 3.35 *Sea D un DFU y K su cuerpo de cocientes. Sea $p(x)$ un polinomio mónico no constante con coeficientes en D. Si c es una raíz de $p(x)$ en K, entonces $c \in D$.*

Demostración: Sea $c = a/b$, con $a, b \in D$. Si $c \notin D$ entonces $b \nmid a$, luego existe un primo $p \in D$ tal que $v_p(a) < v_p(b)$. Sea $p(x) = \sum_{i=0}^{n} d_i x^i$, donde $d_n = 1$. Entonces

$$\frac{a^n}{b^n} + d_{n-1}\frac{a^{n-1}}{b^{n-1}} + \cdots + d_1\frac{a}{b} + d_0 = 0.$$

Multiplicando por b^n queda:

$$a^n = -d_{n-1}ba^{n-1} - \cdots - d_1 b^{n-1}a - d_0 b^n.$$

Ahora bien, el exponente de p en el miembro izquierdo es exactamente $nv_p(a)$, mientras que en el miembro derecho es estrictamente mayor que $nv_p(a)$, con lo que tenemos una contradicción. ∎

Ejemplo Si A es un anillo y $n \in \mathbb{N}$ es un número natural $n \geq 2$, una *raíz n-sima* de a es otro elemento $b \in A$ tal que $b^n = a$. Si aplicamos el teorema anterior al polinomio $x^n - a$ vemos que si A es un DFU, entonces $a \in A$ tiene raíz n-sima en el cuerpo de cocientes K si y sólo si la tiene en A.

Por ejemplo, es inmediato que 2 no tiene raíz cuadrada en $\mathbb{Z}$, por lo que podemos afirmar que tampoco la tiene en $\mathbb{Q}$. Esto es significativo porque hay razones geométricas por las que debería haber un punto en cada recta asociado a una raíz cuadrada de 2, y, en general, de cualquier número positivo, lo que podemos concluir que la representación geométrica de los números racionales no abarca todos los puntos de una recta. ∎

Polinomios ciclotómicos Veamos una aplicación más sofisticada del criterio de Eisenstein, a saber, una prueba alternativa del teorema [ITAl 3.34]. Vamos a ver que si p es primo, el polinomio ciclotómico

$$c_p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$$

es irreducible en $\mathbb{Z}[x]$ (o, equivalentemente, en $\mathbb{Z}[x]$, pues es primitivo). Para ello observamos que

$$x^p - 1 = (x-1)(x^{p-1} + x^{p-2} + \cdots + x + 1).$$

Por el teorema 3.30 basta probar que $p(x+1)$ es irreducible en $\mathbb{Z}[x]$. Aplicamos la evaluación en $x+1$ a la igualdad anterior y obtenemos

$$(x+1)^p - 1 = (x+1-1)p(x+1) = xp(x+1).$$

Por tanto

$$xp(x+1) = \sum_{k=0}^{p} \binom{p}{k} x^k - 1 = \sum_{k=1}^{p} \binom{p}{k} x^k,$$

y en consecuencia $p(x+1) = \sum_{k=1}^{p} \binom{p}{k} x^{k-1}$.

Por el teorema 3.23 tenemos que p divide a todos los coeficientes de $p(x+1)$ salvo al correspondiente a x^{p-1}, que es 1. Además p^2 no divide al término independiente, que es p. Por el criterio de Eisenstein, $p(x+1)$ es irreducible, luego $p(x)$ también. ∎

Ejemplo Veamos otro ejemplo no trivial de irreducibilidad de una familia de polinomios. Vamos a probar que, para todo $n \geq 2$ el polinomio $x^n - x - 1$ es irreducible en $\mathbb{Q}(x)$. Por el criterio de irreducibilidad de Gauss, basta probar que es irreducible en $\mathbb{Z}[x]$.

En general, dado un polinomio $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, definimos $\tilde{f}(x) = x^n f(1/x)$. Explícitamente:

$$\tilde{f}(x) = a_n + a_{n-1}x + \cdots + a_1 x + a_0 x^n.$$

Las propiedades siguientes se demuestran sin dificultad:

1. Si $f(0) \neq 0$, entonces $\operatorname{grad} \tilde{f} = \operatorname{grad} f$ y $\tilde{\tilde{f}} = f$.
2. Si $f = gh$, entonces $\tilde{f} = \tilde{g}\tilde{h}$.
3. Si $c \neq 0$, entonces $\widetilde{cf} = c\tilde{f}$.
4. El coeficiente de x^n en $f\tilde{f}$ es $a_0^2 + a_1^2 + \cdots + a_n^2$.

En el caso en que $f(x) = x^n - x - 1$, tenemos que $\tilde{f}(x) = -x^n - x^{n-1} + 1$. Vamos a probar que, para $n \geq 2$, los polinomios f y $\tilde{f}$ no tienen raíces comunes en $\mathbb{C}$. Si α fuera una raíz común, cumpliría

$$\alpha^n = \alpha + 1, \qquad \alpha^n = -\alpha^{n-1} + 1,$$

luego $\alpha^{n-1} = -\alpha$, luego $\alpha^n = -\alpha^2$, luego $\alpha^2 = -\alpha - 1$, luego $\alpha^3 = 1$. Esto implica que toda potencia de α es igual a $1, \alpha$ o α^2. Pero $\alpha^n = 1$ nos da que $1 = \alpha + 1$ y $\alpha = 0$, pero $f(0) = 1 \neq 0$; si fuera $\alpha^n = \alpha$ tendríamos $\alpha = \alpha + 1$, que es imposible, y si fuera $\alpha^n = \alpha^2$ queda $\alpha^2 = -\alpha^2$ y de nuevo $\alpha = 0$, que también es imposible.

Veamos ahora que si $f(x) \in \mathbb{Z}[x]$ cumple que $f(0) \neq 0$, $f(x)$ y $\tilde{f}(x)$ no tienen raíces comunes y $f(x) = g(x)h(x)$, para ciertos polinomios no constantes $g(x), h(x) \in \mathbb{Z}[x]$, existe un polinomio $k(x) \in \mathbb{Z}[x]$ del mismo grado que f tal que $f\tilde{f} = k\tilde{k}$, $k \neq \pm f$, $k \neq \pm\tilde{f}$. Además, si f es mónico y $f(0) = \pm 1$, podemos tomar $k(x)$ mónico con $k(0) = \pm 1$.

En efecto, como $f(0) \neq 0$, también $g(0) \neq 0 \neq h(0)$, luego $\operatorname{grad}\tilde{g} = \operatorname{grad} g$, $\operatorname{grad}\tilde{h} = \operatorname{grad} h$. Tomamos $k(x) = g(x)\tilde{h}(x)$, con lo que ciertamente se cumple $\operatorname{grad} k = \operatorname{grad} g + \operatorname{grad} h = \operatorname{grad} f$.

Si fuera $k = \pm f$, es decir, $g\tilde{h} = \pm gh$, sería $\tilde{h} = \pm h$, con lo que una raíz de h en $\mathbb{C}$ sería a la vez raíz de $f(x)$ y de $\tilde{f}(x) = \tilde{g}(x)\tilde{h}(x)$, en contra de lo supuesto. Igualmente se razona que $k \neq \pm\tilde{f}$.

Si además suponemos que $f(x)$ es mónico y $f(0) = \pm 1$, entonces $\tilde{f}$ tiene también su coeficiente director y su término independiente iguales a ± 1, luego $(f\tilde{f})(0) = \pm 1$, luego $(k\tilde{k})(0) = \pm 1$, pero, como son polinomios con coeficientes enteros, esto implica que el coeficiente director de k y su término independiente son ± 1, luego cambiando k por $-k$ si es preciso, podemos suponer que k es mónico y $k(0) = \pm 1$.

Veamos finalmente que $f(x) = x^n - x - 1$ es irreducible en $\mathbb{Z}[x]$ cuando $n \geq 2$. El caso $n = 2$ se puede comprobar directamente, así que podemos suponer que $n > 2$. Si f fuera reducible, podríamos expresarlo como $f(x) = g(x)h(x)$, donde $g(x), h(x) \in \mathbb{Z}[x]$ no son constantes. Se cumplen todas las condiciones del resultado precedente, luego existe un polinomio mónico $k(x)$ de grado n con $k(0) = \pm 1$ y tal que $f\tilde{f} = k\tilde{k}$, $\pm f \neq k \neq \pm\tilde{f}$.

Pongamos que $k(x) = b_n x^n + \cdots + b_1 x + b_0$, donde $b_n = 1$, $b_0 = \pm 1$. Comparando los términos de grado n en $f\tilde{f} = k\tilde{k}$ obtenemos

$$3 = b_0^2 + b_1^2 + \cdots b_n^2,$$

de donde $b_1^2 + \cdots + b_{n-1}^2 = 1$. Como los coeficientes son enteros, tienen que ser todos nulos menos un $b_i = \pm 1$. Así, $k(x) = x^n + b_i x^i + b_0$. Calculamos:

$$f\tilde{f} = (x^n - x - 1)(-x^n - x^{n-1} + 1) = -x^{2n} - x^{2n-1} + x^{n+1} + 3x^n + x^{n-1} - x - 1.$$

Notemos que $2n - 1 > n + 1$ porque $n > 2$. Por otro lado:

$$k\tilde{k} = (x^n + b_i x^i + b_0)(b_0 x^n + b_i x^{n-i} + 1) =$$

$$= b_0 x^{2n} + b_i x^{2n-i} + b_0 b_i x^{n+i} + 3x^n + b_0 b_i x^{n-1} + b_i x^i + b_0.$$

Al comparar los coeficientes directores vemos que $b_0 = -1$. Igualamos los términos de grado mayor que n:

$$-x^{2n} - x^{2n-1} + x^{n+1} = -x^{2n} + b_i x^{2n-i} - b_i x^{n+i}.$$

Como los tres monomios de la izquierda son distintos, tiene que ser $2n - i \neq n + i$. Si es $2n - i > n + i$, entonces tiene que ser $i = 1$, $b_i = -1$ y queda $k(x) = x^n - x - 1 = f$, lo cual es imposible.

Si es $n+1 > 2n-i$ entonces $i = n-1$, $b_i = 1$ y queda $k(x) = x^n + x^{n-1} - 1 = -\tilde{f}$, y de nuevo tenemos una contradicción, lo que prueba que $f(x)$ es irreducible en $\mathbb{Z}[x]$, luego también en $\mathbb{Q}(x)$. ∎

Descomposición en fracciones simples Terminamos con una ilustración del uso de la factorización única en los anillos de polinomios. Vamos a probar que toda fracción algebraica se descompone en forma única en suma de fracciones simples, lo que justifica la validez general del método de integración de funciones racionales visto en la sección C2 de [IC]:

Teorema 3.36 *Si k es un cuerpo, toda función racional $z \in k(x)$ se descompone de forma única como*

$$z = f + \sum_{i=1}^{n} \sum_{j=1}^{r_i} \frac{f_{ij}}{p_i^j}$$

donde $f, f_{ij}, p_i \in k[x]$, los polinomios p_i son mónicos, irreducibles, distintos dos a dos, y f_{ij} es nulo o tiene grado menor que p_i (pero $f_{ir_i} \neq 0$).

Observemos que en [IC] nos interesaba el caso en que k es el cuerpo de los números reales, en el cual los polinomios mónicos irreducibles son necesariamente de la forma $x - a$ o bien $x^2 + bx + c$, por lo que las fracciones simples son necesariamente de la forma

$$\frac{A}{(x-a)^n}, \quad \text{o} \quad \frac{Mx+N}{(x^2+bx+c)^n}.$$

Demostración: Sea $z = u/v$, con $u, v \in k[x]$. Podemos suponer que u y v son primos entre sí y que v es mónico. Si $v = 1$ tenemos la descomposición con $f = u$. En caso contrario consideremos un factor primo p_1 de v, de modo que $v = p_1^{r_1}\bar{v}$ y p_1 no divide a $\bar{v}$. Podemos suponer que tanto p_1 como $\bar{v}$ son mónicos. Vamos a probar que

$$\frac{u}{v} = \frac{\bar{u}}{\bar{v}} + \sum_{j=1}^{r_1} \frac{f_{1j}}{p_1^j},$$

donde $\bar{u}$ y $\bar{v}$ son primos entre sí y los polinomios f_{1j} tienen grado menor que el grado de p_1. Si obtenemos esta descomposición, podemos aplicar el mismo resultado a $\bar{u}/\bar{v}$ y así sucesivamente hasta que el denominador acabe siendo 1 (lo cual ocurrirá necesariamente tras un número finito de pasos, porque a cada paso le quitamos un factor primo). Con esto habremos llegado a una descomposición de z como la que aparece en el enunciado.

Como p_1 y $\bar{v}$ son primos entre sí, por la relación de Bezout existen polinomios c, d tales que $c\bar{v} + dp_1 = 1$. Esta igualdad implica en particular que p_1 no divide

a c, y tampoco divide a u porque u y v son primos entre sí. Por lo tanto, en la división $uc = gp_1 + f_{1,r_1}$, el resto f_{1,r_1} no puede ser nulo, y tiene grado menor que el de p_1. Así pues,

$$\frac{u}{\bar{v}} = uc + \frac{du}{\bar{v}}\,p_1 = \frac{g\bar{v}+du}{\bar{v}}\,p_1 + f_{1,r_1} = \frac{u_1}{\bar{v}}\,p_1 + f_{1,r_1},$$

donde u_1 y $\bar{v}$ son primos entre sí. Aplicamos el mismo razonamiento a $u_1/\bar{v}$ para obtener la descomposición

$$\frac{u}{\bar{v}} = \Big(\frac{u_2}{\bar{v}}\,p_1 + f_{1,r_1-1}\Big)\,p_1 + f_{1,r_1} = \frac{u_2}{\bar{v}}\,p_1^2 + f_{1,r_1-1}p_1 + f_{1,r_1},$$

donde u_2 y $\bar{v}$ son primos entre sí (aunque ahora f_{1,r_1-1} sí puede ser 0, lo que sucederá si p_1 divide a u_1). Repetimos el proceso hasta obtener:

$$\frac{u}{\bar{v}} = \frac{\bar{u}}{\bar{v}}\,p_1^{r_1} + \sum_{j=1}^{r_1} f_{1j}p_1^{r_1-j},$$

donde cada f_{1j} es nulo o tiene grado menor que el de p_1 y $\bar{u} = u_r$ es primo con $\bar{v}$. Dividiendo entre $p_1^{r_1}$ obtenemos

$$z = \frac{u}{v} = \frac{\bar{u}}{\bar{v}} + \sum_{j=1}^{r_1} \frac{f_{1j}}{p_1^j},$$

que es la descomposición que buscábamos.

Ahora vamos a probar la unicidad. Supongamos que una misma fracción $z \in k(x)$ admite dos descomposiciones en factores simples. Pongamos que una de las descomposiciones tiene m fracciones (sin contar el polinomio inicial) y la otra n, con $m \le n$. Razonamos por inducción sobre n. Si $n = 0$ tenemos simplemente dos polinomios, que trivialmente han de ser el mismo, pues ambos han de coincidir con z.

Consideremos una descomposición cualquiera:

$$z = f + \sum_{i=1}^{n}\sum_{j=1}^{r_i} \frac{f_{ij}}{p_i^j}$$

Al sumar todas las fracciones correspondientes a p_i, el denominador común es $p_i^{r_i}$, y el numerador es una suma en la que todos los sumandos son múltiplos de p_i menos uno ($f_{i,r_i} \neq 0$), luego tenemos que

$$z = f + \sum_{i=1}^{n} \frac{h_i}{p_i^{r_i}},$$

donde p_i no divide a h_i. Al efectuar la suma (incluyendo a f), con denominador común $p_1^{r_1}\cdots p_n^{r_n}$, el numerador es una suma de $n+1$ términos, todos los cuales son múltiplos de p_i menos uno, a saber, $p_1^{r_1}\cdots h_i \cdots p_n^{r_n}$, luego el numerador no es múltiplo de p_i. Así pues, concluimos que, si $z = u/v$ con u y v primos entre sí y v mónico, necesariamente $v = p_1^{r_1}\cdots p_n^{r_n}$ y que u no es divisible entre ninguno de los p_i.

Con esto hemos probado que, en cualquier descomposición de z en fracciones simples, los polinomios p_i son necesariamente los factores irreducibles de v, y los r_i son los exponentes con que aparecen en v.

Consideremos de nuevo las dos descomposiciones que estamos suponiendo que tiene z. Pongamos que una contiene el sumando $f_{1r_1}/p_1^{r_1}$ y la otra $g_{1r_1}/p_1^{r_1}$. Según hemos visto,

$$u = f_{1,r_1}p_2^{r_2}\cdots p_n^{r_n} + \text{ múltiplos de } p_1 = g_{1,r_1}p_2^{r_2}\cdots p_n^{r_n} + \text{ múltiplos de } p_1,$$

luego p_1 divide a $f_{1,r_1} - g_{1,r_1}$, lo cual sólo es posible si $f_{1,r_1} = g_{1,r_1}$, ya que ambos tienen grado menor que p_1.

Esto significa que podemos cancelar el sumando $f_{1,r_1}/p_1^{r_1}$ de ambas descomposiciones, y así llegamos a otra fracción que admite dos descomposiciones en factores simples de longitudes $m-1$ y $n-1$. Por hipótesis de inducción, ambas tienen que ser iguales, luego las dos descomposiciones de partida también coinciden. ∎

3.6 Congruencias y anillos cociente

Estudiamos ahora las congruencias definidas en la sección 3.1 de [ITAl], aunque las introducimos en el contexto general de la definición [ITAl 13.3]:

Definición 3.37 Consideremos un dominio A y un ideal I. Diremos que dos elementos a y b de A son *congruentes* módulo I, abreviado $a \equiv b \pmod{I}$, si $a - b \in I$.

Teniendo en cuenta que $0 \in I$, que el opuesto de un elemento de I está en I y que la suma de elementos de I está en I, se sigue fácilmente que la congruencia módulo I es una relación de equivalencia en A.

Diremos que dos elementos de A son congruentes módulo un tercero si lo son módulo el ideal que éste genera. Explícitamente:

$$a \equiv b \pmod{m} \quad \text{si y sólo si} \quad m \mid a - b.$$

En el caso de un DIP, como es $\mathbb{Z}$, podemos hablar de congruencias en estos términos sin perder generalidad y sin mencionar para nada ideales, pero es bueno saber que, en el fondo, lo que hay es un ideal.

En el caso concreto de $\mathbb{Z}$ podemos dividir $a = nc + r$, con $0 \le r < n$, y resulta que $a \equiv r \pmod{n}$. Así, todo número entero es congruente con un número natural menor que n, exactamente con su resto al dividirlo entre n. Por otra parte dos números naturales distintos menores que n no pueden ser congruentes entre sí, ya que su diferencia en el orden adecuado es un número natural no nulo menor que n, luego no puede ser un múltiplo de n. Vamos a expresar adecuadamente lo que hemos obtenido:

Si I es un ideal de un anillo A, llamaremos A/I al conjunto cociente originado por la relación de congruencia módulo I.

Lo que acabamos de probar es que el conjunto $\mathbb{Z}/n\mathbb{Z}$ tiene exactamente n elementos. Si llamamos $[a]$ a la clase de equivalencia de a tenemos, por ejemplo, que $\mathbb{Z}/5\mathbb{Z} = \{[0],\ [1],\ [2],\ [3],\ [4]\}$. La primera clase la forman los múltiplos de 5, la segunda los números que al ser divididos entre 5 dan resto 1, etc.

En general, en un anillo A, los elementos congruentes con un cierto a módulo un ideal I son los b que cumplen $b - a \in I$, es decir, los elementos de la forma $a + i$ para un cierto $i \in I$. Por lo tanto $[a] = a + I = \{a + i \mid i \in I\}$.

El interés de todo esto radica en el hecho siguiente:

Teorema 3.38 *Sea A un dominio e I un ideal de A. Si a, a', b, b' son elementos de A que cumplen $[a] = [a']$ y $[b] = [b']$, entonces también $[a+b] = [a'+b']$ y $[ab] = [a'b']$. El conjunto A/I se convierte en un anillo conmutativo y unitario con las operaciones definidas mediante $[a] + [b] = [a + b]$, $[a][b] = [ab]$. Al anillo A/I se le llama* anillo cociente *de A módulo el ideal I.*

Demostración: Tenemos que $a - a' \in I$ y que $b - b' \in I$, de donde

$$(a + b) - (a' + b') = (a - a') + (b - b') \in I,$$

$$ab - a'b' = ab - a'b + a'b - a'b' = (a - a')b + a'(b - b') \in I,$$

por las propiedades de los ideales. Esto nos garantiza que las operaciones definidas mediante $[a] + [b] = [a + b]$ y $[a][b] = [ab]$ no dependen de la elección de a y b en cada una de las clases. El resto del teorema es inmediato. Por ejemplo la asociatividad de la suma se cumple porque

$$([a] + [b]) + [c] = [a + b] + [c] = [a + b + c] = [a] + [b + c] = [a] + ([b] + [c]).$$

■

Remitimos a la sección 3.2 de [ITAl] para algunas aplicaciones elementales de las congruencias. El teorema siguiente es [ITAl 13.13] y generaliza a [ITAl 3.9]. Notemos que en un anillo cociente A/I se cumple $0 = [0] = 0 + I = I$, luego $[a] = 0$ si y sólo si $a \in I$.

Teorema 3.39 *Sea A un dominio y P un ideal de A. Entonces A/P es un dominio íntegro si y sólo si P es un ideal primo.*

Demostración: Si P es primo entonces $P \neq A$, luego $1 \notin P$, luego en A/P se cumple que $[1] \neq [0]$. Así pues, A/P es un dominio. Si $[a]$, $[b]$ son dos elementos de A/P tales que $[a][b] = 0$, entonces $[ab] = [0]$, es decir, $ab \in P$. Como P es primo, $a \in P$ o $b \in P$, luego se cumple que $[a] = 0$ o $[b] = 0$, es decir, A/P es un dominio íntegro.

Recíprocamente, si A/P es un dominio íntegro, ha de ser $[1] \neq [0]$, luego $1 \notin P$ y en consecuencia $P \neq A$. Si a, b son elementos de A tales que $ab \in P$, entonces $[a][b] = [ab] = 0$, y como A/P es íntegro $[a] = 0$ o $[b] = 0$, es decir, $a \in P$ o $b \in P$, luego P es primo. ■

Queda así justificado que los anillos $\mathbb{Z}/p\mathbb{Z}$ son dominios íntegros cuando p es un primo.

Teorema 3.40 *Sea A un dominio y sea M un ideal de A. Entonces A/M es un cuerpo si y sólo si M es un ideal maximal.*

Demostración: Si M es maximal entonces es primo, luego A/M es un dominio íntegro. Veamos que si $[a] \neq [0]$ entonces $[a]$ es una unidad en A/M. Como $a \notin M$ tenemos que $M + (a) \neq M$, luego por la maximalidad de M ha de ser $M + (a) = A$. En particular $1 \in M + (a)$, luego 1 se puede expresar de la forma $1 = m + ba$, para cierto $m \in M$ y cierto $b \in A$. Tomando clases resulta que $[1] = [m] + [b][a] = [b][a]$, luego $[a]$ es en efecto una unidad.

Si A/M es cuerpo entonces M es un ideal primo, y en particular $M \neq A$. Consideremos un ideal N de A tal que $M \subsetneq N \subset A$. Sea a un elemento de N que no esté en M. Así $[a] \neq 0$, luego existe un $b \in A$ tal que $[a][b] = [1]$, es decir, $[ab - 1] = [0]$, luego $ab - 1 \in M \subset N$ y como $a \in N$, también $ab \in N$, luego $1 \in N$ y por tanto $N = A$. Esto prueba que M es un ideal maximal. ∎

Como en $\mathbb{Z}$ los ideales maximales coinciden con los primos, resulta que los anillos $\mathbb{Z}/p\mathbb{Z}$ son cuerpos. La situación general es la siguiente [ITAl 3.8]:

Teorema 3.41 *El conjunto U_n de las unidades del anillo $\mathbb{Z}/n\mathbb{Z}$ está formado por las clases $[m]$ tales que $(m, n) = 1$.*

Demostración: Si $(m, n) = 1$, por la relación de Bezout, existen enteros u, v tales que $um + vn = 1$, con lo que $[u][m] = 1$, lo que prueba que $[m] \in U_n$. Recíprocamente, si $(m, n) = d > 1$, podemos expresar $m = dm'$, $n = dn'$, y entonces $[m][n'] = [m'][dn'] = [m'][n] = 0$, con $[n'] \neq 0$, pues $0 < n' < n$, y esto significa que $[m]$ es un divisor de 0 en $\mathbb{Z}/n\mathbb{Z}$, luego no puede ser una unidad. ∎

Hay otra razón por la cual los anillos $\mathbb{Z}/n\mathbb{Z}$ que son dominios íntegros son también cuerpos, y es que son finitos:

Teorema 3.42 *Todo dominio íntegro finito es un cuerpo.*

Demostración: Sea A un dominio íntegro finito. Hay que probar que todo elemento no nulo u de A es una unidad. Si $A = \{0, u_1, \ldots, u_n\}$, entonces los elementos $uu_1, \ldots, uu_n$ son todos distintos y no nulos, luego se da la igualdad de conjuntos $\{u_1, \ldots, u_n\} = \{uu_1, \ldots, uu_n\}$ y por lo tanto: $u_1 \cdots u_n = uu_1 \cdots uu_n$, es decir, $u_1 \cdots u_n = u^n \cdot (u_1 \cdots u_n)$. Como $u_1 \cdots u_n$ no es 0, podemos cancelarlo, lo que nos da $u^n = 1$, o sea $u \cdot u^{n-1} = 1$, luego u es una unidad. ∎

Hay una variante de este argumento que tiene muchas consecuencias interesantes:

Teorema 3.43 *Sea A un dominio con un número finito n de unidades y sea u una unidad de A. Entonces $u^n = 1$.*

DEMOSTRACIÓN: Sean $u_1, \ldots, u_n$ las unidades de A. Entonces los elementos $uu_1, \ldots, uu_n$ son todos distintos, pues si $uu_i = uu_j$, como u es una unidad, se cumple $u_i = u_j$. Además el producto de unidades es una unidad, luego en realidad $\{u_1, \ldots, u_n\} = \{uu_1, \ldots, uu_n\}$ y en particular los productos coinciden: $u_1 \cdots u_n = uu_1 \cdots uu_n$, es decir, $u_1 \cdots u_n = u^n \cdot (u_1 \cdots u_n)$, y como $u_1 \cdots u_n$ es una unidad, podemos cancelarlo, lo que nos da $u^n = 1$. ∎

En particular, el cuerpo $\mathbb{Z}/p\mathbb{Z}$ tiene $p-1$ unidades, luego si $[a] \in \mathbb{Z}/p\mathbb{Z}$ cumple $[a] \neq [0]$, entonces $[a]^{p-1} = [1]$, luego $[a]^p = [a]$ (y esto último vale incluso si $[a] = [0]$). Ésta es una de las formas en las que se suele enunciar el teorema de Fermat (compárese con [ITAl 3.24]):

Teorema 3.44 (Teorema de Fermat) *Para todo primo p y todo número entero a, se cumple que $a^p \equiv a$* (mód p).

Homomorfismos y cocientes Ya hemos visto que la existencia de un monomorfismo de anillos $f : A \longrightarrow B$ se interpreta como que A puede ser considerado como un subanillo de B. Ahora veremos que si f es un epimorfismo entonces B puede ser considerado un cociente de A.

Definición 3.45 Si $f : A \longrightarrow B$ es un homomorfismo de anillos, llamaremos *núcleo* de f al conjunto $\mathrm{N}(f) = \{a \in A \mid f(a) = 0\}$.

No ofrece ninguna dificultad probar que $\mathrm{N}(f)$ es un ideal de A. Una propiedad útil es la siguiente:

Teorema 3.46 *Si $f : A \longrightarrow B$ es un homomorfismo de anillos, se cumple que f es monomorfismo si y sólo si $\mathrm{N}(f) = 0$.*

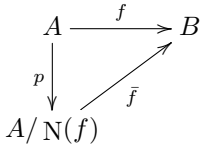
DEMOSTRACIÓN: Una implicación es evidente, y para ver la otra supongamos que $\mathrm{N}(f) = 0$ y que $f(a) = f(b)$. Entonces $f(b-a) = 0$, y en consecuencia $b - a \in \mathrm{N}(f) = 0$, es decir, $b = a$. ∎

Por ejemplo, si $f : K \longrightarrow A$ es un homomorfismo de un cuerpo K en un anillo A, o bien f es constante igual a 0 o bien es un monomorfismo, porque el núcleo de f tiene que ser un ideal de K, y sólo hay dos opciones: que sea K, en cuyo caso f es nulo, o bien que sea 0, en cuyo caso f es un monomorfismo.

Notemos ahora que si A es un anillo e I es un ideal de A, entonces la aplicación $p : A \longrightarrow A/I$ dada por $p(a) = [a]$ es un epimorfismo, llamado *epimorfismo canónico*. Claramente $\mathrm{N}(p) = I$.

Sucede que todo epimorfismo es esencialmente de este tipo:

Teorema 3.47 (Teorema de isomorfía) *Sea $f : A \longrightarrow B$ un homomorfismo de anillos. Entonces la aplicación $\bar{f} : A/\mathrm{N}(f) \longrightarrow f[A]$ dada por $\bar{f}\big([a]\big) = f(a)$ es un isomorfismo de anillos que hace conmutativo el diagrama siguiente:*



DEMOSTRACIÓN: La aplicación está bien definida, pues si $[a] = [b]$, entonces $a - b \in \mathrm{N}(f)$, con lo que $f(a - b) = 0$, o sea, $f(a) = f(b)$.

Es inmediato comprobar que se trata de un homomorfismo y es inyectivo porque si $\bar{f}([a]) = \bar{f}([b])$, entonces $f(a) = f(b)$, $f(a - b) = 0$, $a - b \in \mathrm{N}(f)$, luego $[a] = [b]$. La conmutatividad del diagrama es inmediata. ∎

Notemos que, en las condiciones del teorema anterior, si f es un epimorfismo, entonces $\bar{f}$ es un isomorfismo entre B y $A/\mathrm{N}(f)$ de modo que si "sustituimos" cada elemento de B por su equivalente en el cociente, entonces f se corresponde con la proyección canónica.

La característica de un anillo Recordamos ahora la noción de característica de un anillo introducida en [ITAl 3.12]. Observemos que si A es un dominio, entonces tenemos definidos los elementos

$$1, \quad 2 \cdot 1 = 1 + 1, \quad 3 \cdot 1 = 1 + 1 + 1, \quad \ldots \quad n \cdot 1, \quad \ldots$$

Aquí hay que entender que en la expresión $n \cdot 1$ el número n es un número natural, mientras que el 1 es la identidad de A. El resultado no es un número natural, sino un elemento de A. Por ejemplo, si hacemos el cálculo en $\mathbb{Z}/6\mathbb{Z}$ obtenemos que

$$6 \cdot 1 = 1 + 1 + 1 + 1 + 1 + 1 = 0.$$

Definición 3.48 Llamaremos *característica* de un dominio A ($\operatorname{car} A$) al mínimo número natural no nulo n tal que $n1 = 0$, o bien $\operatorname{car} A = 0$ si no existe tal n.

En otras palabras, la característica de un dominio es el mínimo número de veces que hay que sumar 1 consigo mismo para obtener 0, o bien 0 si esto es imposible.

Claramente $\mathbb{Z}$ y $\mathbb{Q}$ son anillos de característica 0, mientras que $\mathbb{Z}/n\mathbb{Z}$ tiene característica n.

Notemos que si $A \subset B$ son dominios íntegros, entonces la identidad en A es la misma que la identidad en B (pues la identidad en A cumple $1 = 1 \cdot 1$, y el único elemento no nulo que cumple esto en B es la identidad), luego $\operatorname{car} A = \operatorname{car} B$.

Otro hecho notable es que la característica de un dominio íntegro ha de ser o bien 0 o bien un número primo, pues si $\operatorname{car} A = mn$, donde m y n no son 1, entonces $m1 \neq 0 \neq n1$, pero $(m1)(n1) = (mn)1 = 0$, luego A no es íntegro.

Fijemos ahora un anillo unitario A y consideremos la aplicación $f : \mathbb{Z} \longrightarrow A$ dada por $f(n) = n1$, que claramente es un homomorfismo.

Si A tiene característica 0, entonces $f(n) \neq 0$ para todo $n > 0$, luego también para todo $n \neq 0$. Por lo tanto, el núcleo de f es el ideal 0 y f es un monomorfismo o, equivalentemente, un isomorfismo entre $\mathbb{Z}$ y $f[\mathbb{Z}]$, de modo que $f[\mathbb{Z}]$ es un anillo indistinguible de $\mathbb{Z}$ a efectos algebraicos.

Por consiguiente, en la práctica podemos identificar cada $n \in \mathbb{Z}$ con $n \cdot 1 \in A$ y considerar así que $\mathbb{Z} \subset A$. En otras palabras, si A es un anillo de característica 0

y escribimos $3 \in A$, nos referimos al elemento $3 = 1 + 1 + 1 \in A$, e igualmente $-3 = -1 - 1 - 1 \in A$. El hecho de que f sea un monomorfismo se traduce en que en A se cumple $2 + 3 = 5$, $3 \cdot 4 = 12$ o $5 \neq 18$ porque todo esto es cierto en $\mathbb{Z}$.

Supongamos ahora que A tiene característica $m > 0$ y vamos a probar que el núcleo de f es $m\mathbb{Z}$. En principio sabemos que $f(m) = 0$, luego dicho núcleo es un ideal no nulo de $\mathbb{Z}$, que será de la forma $r\mathbb{Z}$, para cierto $r > 0$ tal que $r \mid m$. En particular $0 < r \leq m$, pero, como $r \cdot 1 = 0$ y m es el menor natural no nulo que cumple esto, tiene que ser $r = m$, luego el núcleo es ciertamente $m\mathbb{Z}$.

El teorema de isomorfía nos da entonces un monomorfismo $\bar{f} : \mathbb{Z}/m\mathbb{Z} \longrightarrow A$, dado por $\bar{f}([n]) = n \cdot 1$. Nuevamente, esto nos legitima a identificar a $\mathbb{Z}/m\mathbb{Z}$ con un subanillo de A, el formado también por las sumas de unos, con la diferencia de que ahora dicho anillo de sumas de unos no es isomorfo a $\mathbb{Z}$, sino a $\mathbb{Z}/m\mathbb{Z}$. En resumen:

- Si A es un dominio de característica 0, entonces tiene como subanillo a $\mathbb{Z}$ (identificado con el anillo formado por las sumas de unos y menos unos).
- Si K es un cuerpo de característica 0, entonces, por el teorema 2.22 tenemos que K no sólo contiene a $\mathbb{Z}$, sino también a su cuerpo de cocientes $\mathbb{Q}$.
- Si A es un dominio de característica $n > 0$, entonces tiene como subanillo a $\mathbb{Z}/n\mathbb{Z}$ (identificado con el anillo formado por las sumas de unos de A).
- Si A es un dominio íntegro (en particular un cuerpo) de característica $p > 0$ entonces p es primo y contiene como subcuerpo a $\mathbb{Z}/p\mathbb{Z}$.

Más precisamente, si K es un cuerpo de característica 0, podemos considerar que, por ejemplo, $5/3 \in K$ sin más que entender que nos referimos al elemento.

$$\frac{1+1+1+1+1}{1+1+1}.$$

Así pues, los anillos $\mathbb{Z}$ y $\mathbb{Z}/n\mathbb{Z}$ son los menores en sus respectivas características, pues están contenidos en cualquier dominio de la misma característica.

Más precisamente, si K es un cuerpo y $k \subset K$ es un subcuerpo, el cuerpo formado a partir del 1 mediante sumas, restas y cocientes, que es isomorfo a $\mathbb{Q}$ o a un cuerpo de restos $\mathbb{Z}/p\mathbb{Z}$, está contenido en k (porque $1 \in k$ y al sumar, restar o dividir unos no nos salimos de k). Así pues, se trata del menor subcuerpo de K, en cuanto a que está contenido en todos sus subcuerpos. Dicho cuerpo mínimo se llama el *cuerpo primo* de K.

Veamos un par de teoremas que muestran por qué la característica de un anillo es un dato a tener en cuenta. El primero es [ITAl 3.16]:

Teorema 3.49 *Sea A un dominio de característica prima p. Entonces para todos los elementos a y b de A se cumple $(a \pm b)^p = a^p \pm b^p$.*

DEMOSTRACIÓN: Usamos el teorema 3.23:

$$(a+b)^p = \sum_{k=0}^{p} \binom{p}{k} a^k b^{p-k} = a^p + b^p + \sum_{k=1}^{p-1} \binom{p}{k} a^k b^{p-k} = a^p + b^p.$$

Si p es impar $(a-b)^p = a^p + (-b)^p = a^p - b^p$. Si $p = 2$ entonces $b+b = 2b = 0$, luego $b = -b$ y el resultado vale. ∎

Obviamente de aquí se sigue que, más en general, $(a \pm b)^{p^n} = a^{p^n} \pm b^{p^n}$.

Teorema 3.50 *Sea K un cuerpo con* $\operatorname{car} K \neq 2$ *y sea $p(x) = ax^2+bx+c \in K[x]$ con $a \neq 0$. El polinomio $p(x)$ tiene una raíz en K si y sólo si existe un $\alpha \in K$ de manera que $\alpha^2 = b^2 - 4ac$. En tal caso, si llamamos $\alpha = \sqrt{b^2-4ac}$, las raíces de $p(x)$ son*

$$\frac{-b \pm \sqrt{b^2-4ac}}{2a}.$$

DEMOSTRACIÓN: Supongamos que $\eta \in K$ cumple $a\eta^2 + b\eta + c = 0$. Multiplicando por $4a$ tenemos que $(2a\eta)^2 + 2(2a\eta b) + 4ac = 0$, de donde $(2a\eta + b)^2 = b^2 - 4ac$ y por lo tanto $2a\eta + b = \pm\sqrt{b^2-4ac}$. Como $\operatorname{car} K \neq 2$, $2a \neq 0$ y así

$$\eta = \frac{-b \pm \sqrt{b^2-4ac}}{2a}.$$

Si existe $\sqrt{b^2-4ac} \in K$, es fácil ver que los elementos η definidos por la fórmula anterior son raíces de $p(x)$. ∎

Veamos algunas observaciones más:

- Todo anillo de característica 0 es infinito (porque contiene a $\mathbb{Z}$).
- Sin embargo, un anillo de característica no nula puede ser infinito. Por ejemplo, el anillo de polinomios $(\mathbb{Z}/n\mathbb{Z})[x]$.
- Todo anillo ordenado tiene característica 0, pues una simple inducción demuestra que $n \cdot 1 > 0$ para todo $n > 0$.

  En particular vemos que en un anillo finito no es posible definir ninguna relación de orden que cumpla los axiomas de anillo ordenado.

Podría pensarse que, por el contrario, todos los anillos de característica 0 pueden ordenarse, pero no es así. Por ejemplo, si un cuerpo K, de característica 0, contiene un elemento $i \in K$ tal que $i^2 = -1$, entonces no puede ser dotado de una relación de orden que lo convierta en un cuerpo ordenado, porque tendría que ser $i^2 = -1 < 0$ y, por otra parte, los cuadrados tienen que ser positivos en un cuerpo ordenado.

Cocientes de anillos de polinomios Veamos una última aplicación de los anillos cociente:

Teorema 3.51 *Sea k un cuerpo y $p(x) \in k[x]$ no constante. Entonces existe un cuerpo K que contiene a k en donde $p(x)$ tiene una raíz.*

DEMOSTRACIÓN: Tomando un factor irreducible, podemos suponer que $p(x)$ es irreducible (pues toda raíz de un factor de $p(x)$ es también raíz de $p(x)$). Entonces el ideal $\big(p(x)\big)$ es maximal en $k[x]$ (teorema 3.15), luego el anillo cociente $K = k[x]/\big(p(x)\big)$ es un cuerpo.

La proyección canónica $p : k[x] \longrightarrow K$ se restringe a un homomorfismo $j : k \longrightarrow K$ dado por $j(a) = [a]$. Como es no nulo (pues $j(1) = 1$), tiene que ser un monomorfismo de cuerpos, a través del cual podemos identificar a k con su imagen, es decir, a cada $a \in k$ con la clase $[a] \in K$.

Por otra parte, llamamos $\alpha = [x] \in K$. La clave del argumento es que, al haber tomado el cociente respecto del ideal $(p(x))$, en K se cumple que $[p(x)] = 0$. Más explícitamente, si $p(x) = \sum_{i=0}^{n} a_i x^i$, entonces

$$0 = [p(x)] = \sum_{i=0}^{n} [a_i][x]^i = \sum_{i=0}^{n} a_i \alpha^i = p(\alpha),$$

donde hemos usado la identificación $[a_i] = a_i$. Así pues, α es una raíz de $p(x)$ en K. ∎

Definición 3.52 Si k es un cuerpo y $p(x) \in k[x]$ es un polinomio irreducible, llamaremos $k[\alpha]$ al cuerpo construido en la demostración del teorema anterior, es decir, $k[\alpha] = k[x]/(p(x))$ y $\alpha = [x]$, de modo que α es una raíz de $p(x)$ en $k[\alpha]$.

La estructura de este cuerpo es fácil de describir (compárese con [ITAl 6.1]): En principio, sus elementos son clases de equivalencia de la forma $[q(x)]$, donde $q(x) \in k[x]$. Si $q(x) = \sum_{i=0}^{m} b_i x^i$, entonces

$$[q(x)] = \sum_{i=0}^{m} [b_i][x]^i = \sum_{i=0}^{m} b_i \alpha^i = q(\alpha),$$

de modo que $k[\alpha] = \{q(\alpha) \mid q(x) \in k[x]\}$. Esto significa que todo elemento de $k[\alpha]$ se obtiene sumando y multiplicando potencias de α y elementos de k. Pero podemos precisar más. Si $p(x)$ tiene grado n, dado cualquier $q(x) \in k[x]$ podemos dividir $q(x) = p(x)c(x) + r(x)$, donde $r(x) = \sum_{i=0}^{n-1} c_i x^i$, y entonces, como $p(\alpha) = 0$, queda que $q(\alpha) = r(\alpha)$, luego

$$k[\alpha] = \{c_{n-1}\alpha^{n-1} + \cdots + c_1\alpha + c_0 \mid c_0, \ldots, c_{n-1} \in k\}.$$

Así pues, todo elemento de $k[\alpha]$ se expresa como suma de las n potencias consecutivas $1, \alpha, \ldots, \alpha^{n-1}$ multiplicadas por elementos de k. Más aún, cada elemento de $k[\alpha]$ admite una única expresión de esta forma, es decir, si

$$c_{n-1}\alpha^{n-1} + \cdots + c_1\alpha + c_0 = c'_{n-1}\alpha^{n-1} + \cdots + c'_1\alpha + c'_0,$$

entonces $c_i = c'_i$ para todo i. En efecto, tenemos que

$$(c_{n-1} - c'_{n-1})\alpha^{n-1} + \cdots + (c_1 - c'_1)\alpha + c_0 - c'_0 = 0.$$

Si llamamos $t(x) = \sum_{i=0}^{n-1}(c_i - c'_i)x^i$, entonces $t(\alpha) = 0$, lo cual, haciendo $\alpha = [x]$ y eliminando la identificación de k con su imagen en $k[\alpha]$, equivale a que $[t(x)] = 0$, es decir, a que $p(x) \mid t(x)$, pero la única forma en que un polinomio de grado n puede dividir a otro de grado a lo sumo $n-1$ es que el segundo sea 0. Por lo tanto $t(x) = 0$, lo cual significa que $c_i = c'_i$ para todo i, como queríamos probar.

En los capítulos IV, V y VI de [ITAl] estudiamos con detalle varios cuerpos de esta forma.

El teorema chino del resto Vamos a mostrar distintos grados de abstracción del teorema chino del resto, que en su versión más particular [ITAl 3.7] afirma que un sistema de congruencias lineales

$$x \equiv a_1 \ (\text{mód } m_1), \quad \ldots, \quad x \equiv a_n \ (\text{mód } m_n)$$

tiene solución siempre que los módulos $m_1, \ldots, m_n$ son primos entre sí dos a dos. Más aún, las soluciones x del sistema forman una misma clase de congruencia módulo $m = m_1 \cdots m_n$.

En términos más modernos, este mismo enunciado puede expresarse así:

Teorema 3.53 (Teorema chino del resto) *Si $m_1, \ldots, m_n$ son números enteros primos entre sí dos a dos y $m = m_1 \cdots m_n$, se tiene el isomorfismo de anillos*

$$\mathbb{Z}/m\mathbb{Z} \cong (\mathbb{Z}/m_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/m_n\mathbb{Z})$$

dado por $[a] \mapsto ([a], \ldots, [a])$.

Demostración: Es claro que la aplicación $f(a) = ([a], \ldots, [a])$ define un homomorfismo de anillos de $\mathbb{Z}$ en el producto de los anillos cociente (considerado como anillo con las operaciones definidas coordenada a coordenada), y su núcleo es $m\mathbb{Z}$, pues obviamente $f(m) = 0$ y, si $f(a) = 0$, entonces cada m_i divide a a y, como son primos entre sí, esto implica que $m \mid a$, luego $a \in m\mathbb{Z}$. Por tanto, el teorema de isomorfía nos da un monomorfismo de $\mathbb{Z}/m\mathbb{Z}$ en el producto de anillos, y ahora basta observar que ambos anillos tienen el mismo número de elementos, luego una aplicación inyectiva entre ellos tiene que ser biyectiva. ∎

La demostración anterior es muy simple, porque prueba la inyectividad, que es trivial, y automáticamente obtiene la suprayectividad, que es la parte no trivial. Sin embargo, así no tenemos ninguna información sobre cómo encontrar un entero a que resuelva el sistema de congruencias para unos a_i dados. Vamos a dar una prueba alternativa que sí nos indica cómo obtener antiimágenes explícitamente y que además sirve en un contexto más general (más general incluso que [ITAl 4.9]):

Teorema 3.54 (Teorema chino del resto) *Si D es un DIP, $m_1, \ldots, m_n$ son elementos de D primos entre sí dos a dos y $m = m_1 \cdots m_n$, se tiene el isomorfismo de anillos*

$$D/(m) \cong D/(m_1) \times \cdots \times D/(m_n)$$

dado por $[a] \mapsto ([a], \ldots, [a])$.

Demostración: El mismo argumento del teorema anterior prueba que la aplicación indicada es un monomorfismo de anillos. Falta probar que es suprayectivo. Llamemos $m_i' = m/m_i \in D$. Es claro que m_i y m_i' son primos entre sí, luego por la relación de Bezout existen $r_i, s_i \in D$ tales que $r_i m_i + s_i m_i' = 1$. Si llamamos $e_i = s_i m_i'$, tenemos que

$$e_i \equiv 1 \ (\text{mód } m_i), \qquad e_i \equiv 0 \ (\text{mód } m_j), \quad \text{para } j \neq i.$$

Ahora, dados $u_1, \ldots, u_n \in D$, es claro que la clase de $a = u_1 e_1 + \cdots + u_n e_n$ módulo m es una antiimagen de $([u_1], \ldots, [u_n])$. ∎

La hipótesis de que el anillo considerado es DIP la hemos usado al aplicar la relación de Bezout. Vamos a probar que dicha hipótesis es necesaria:

Teorema 3.55 *Si D es un dominio de factorización única, las afirmaciones siguientes son equivalentes:*

1. *D es un dominio de ideales principales.*
2. *Los ideales primos no triviales de D son maximales.*
3. *Si $a_1, \ldots, a_n$ son elementos de D primos entre sí dos a dos y $m = a_1 \cdots a_n$, se tiene el isomorfismo de anillos*

   $$D/(m) \cong D/(a_1) \times \cdots \times D/(a_n)$$

   dado por $[a] \mapsto ([a], \ldots, [a])$.
4. *Si $a_1, \ldots, a_n$ son elementos de D y d es su máximo común divisor, existen $r_1, \ldots, r_n \in D$ tales que $d = r_1 a_1 + \cdots + r_n a_n$.*

Demostración: 1) ⇒ 2) es el teorema 3.17.

2) ⇒ 3) Es claro que el homomorfismo descrito en el enunciado es un monomorfismo. El problema es ver que es suprayectivo. Dividimos la prueba en varios pasos:

i) *Si π es primo en D, entonces las unidades de $D/(\pi^r)$ son las clases de los elementos primos con π.*

Lo probamos por inducción sobre r. Para $r = 1$ es trivial, pues el ideal (π) es primo, luego por hipótesis es maximal, luego $D/(\pi)$ es un cuerpo.

Si vale para r y $(\alpha, \pi) = 1$, por hipótesis de inducción existe $\xi_0 \in D$ de manera que $\alpha\xi_0 \equiv 1 \ (\text{mód } \pi^r)$, es decir, $\alpha\xi_0 - 1 = \pi^r \beta$. Sea $\lambda \in D$ y $\xi = \xi_0 + \lambda\pi^r$. Veamos que eligiendo λ adecuadamente se cumple que $\alpha\xi \equiv 1 \ (\text{mód } \pi^{r+1})$.

Tenemos que $\alpha\xi - 1 = \alpha\xi_0 - 1 + \alpha\lambda\pi^r = \pi^r(\beta + \alpha\lambda)$, luego $\alpha\xi \equiv 1 \pmod{\pi^{r+1}}$ si y sólo si $\alpha\lambda \equiv -\beta \pmod{\pi}$. Ahora bien, por el caso $r = 1$ resulta que α es una unidad módulo π, luego existe un λ que cumple la congruencia.

El recíproco es obvio: si $\alpha\xi \equiv 1 \pmod{\pi^{r+1}}$ entonces $\alpha\xi + \beta\pi^{r+1} = 1$, con lo que claramente $\pi \nmid \alpha$.

ii) *Se cumple 3) bajo la hipótesis de que $a_i = \pi_i^{r_i}$, donde los π_i son primos no asociados dos a dos.*

Llamemos $m_i = m/a_i$, de modo que m_i es primo con a_i, luego, por i), sabemos que m_i es una unidad en $D/(a_i)$, es decir, que existe un $s_i \in D$ tal que $e_i = s_i m_i \equiv 1 \pmod{a_i}$, y obviamente $e_i \equiv 0 \pmod{a_j}$, para $j \neq i$. Entonces, dados $u_1, \ldots, u_n \in D$, es claro que $u = u_1 e_1 + \cdots + u_n e_n$ es una antiimagen de $([u_1], \ldots, [u_n])$ por el homomorfismo del enunciado.

Por consiguiente, $[u]$ es una unidad de $D/(m)$ si y sólo si $([u], \ldots, [u])$ es una unidad en el producto, lo que equivale a que $[u]$ sea una unidad en cada cociente $D/(\pi_i^{r_i})$ y hemos probado que esto equivale a que u sea primo con cada π_i, lo que equivale a que u sea primo con m.

iii) *Si $m \in D$ no es nulo ni unitario, las unidades de $D/(m)$ son las clases de los elementos primos con m.*

Podemos descomponer $m = \epsilon\pi_1^{r_1} \cdots \pi_n^{r_n}$, donde ϵ es una unidad y los π_i son primos no asociados dos a dos. Como $(m) = (m\epsilon^{-1})$ y los primos con m son los mismos que los primos con $m\epsilon^{-1}$, podemos suponer que $\epsilon = 1$. Llamamos $a_i = \pi_i^{r_i}$ y por ii) tenemos el isomorfismo del enunciado.

Por lo tanto, si $u \in D$, la clase $[u]$ es una unidad en $D/(m)$ si y sólo si $([u], \ldots, [u])$ es una unidad en el producto, lo cual equivale a que cada clase $[u]$ sea una unidad en $D/(a_i)$, y por i) esto equivale a que u sea primo con cada π_i, lo que a su vez equivale a que u sea primo con m.

Por último, usando iii) en lugar de i), la prueba de ii) vale en general, sin suponer que cada a_i es una potencia de primo (notemos que no se pierde generalidad si suponemos que cada a_i no es nulo ni unitario), lo que nos da 3).

3) $\Rightarrow$ 4) Razonamos por inducción sobre n. Consideramos primero el primer caso no trivial, que es $n = 2$. Sea $a_1 = db_1$, $a_2 = db_2$, de modo que b_1 y b_2 son primos entre sí. Por 3) existe un $u \in D$ tal que $u \equiv 1 \pmod{b_1}$, $u \equiv 0 \pmod{b_2}$. Explícitamente, $u = 1 + b_1c_1 = b_2c_2$, luego

$$d = d(-b_1c_1 + b_2c_2) = -c_1a_1 + c_2a_2,$$

como había que probar. Si vale para n, sea $d^* = \mathrm{mcd}(a_1, \ldots, a_n)$, de modo que $d = \mathrm{mcd}(d^*, a_{n+1})$. Por hipótesis de inducción y por el caso $n = 2$ ya probado, tenemos que

$$d^* = r_1a_1 + \cdots + r_na_n, \qquad d = rd^* + sa_{n+1},$$

de donde $d = rr_1a_1 + \cdots rr_na_n + sa_{n+1}$.

4) $\Rightarrow$ 1) Notemos que 4) equivale a que $(a_1,\dots,a_n)=(d)$. Sea I un ideal propio de D y tomemos $\alpha\in I$ no nulo. Factoricemos $\alpha=\epsilon\pi_1^{e_1}\cdots\pi_n^{e_n}$ como producto de primos, donde ϵ es una unidad y los π_i son no asociados dos a dos.

Definimos $e(\pi)=\min\limits_{\delta} e_\pi(\delta)$, donde δ recorre los elementos no nulos de I y $e_\pi(\delta)$ es el exponente de π en δ. Así, si $e(\pi)>0$ entonces $e_\pi(\alpha)\ge e(\pi)>0$, luego $\pi=\pi_i$ para algún i. Equivalentemente, $e(\pi)=0$ para todos los primos excepto quizá algunos de los π_i.

Por la propia definición de $e(\pi)$ tenemos que para cada i existe un $\beta_i\in I$ de modo que $e_{\pi_i}(\beta_i)=e(\pi_i)$. Sea $\delta=\pi_1^{e(\pi_1)}\cdots\pi_n^{e(\pi_n)}$. Se cumple que $I\subset(\delta)$, pues si $\beta\in I$, entonces $e(\pi_i)\le e_{\pi_i}(\beta)$, para todo i, luego $\delta\mid\beta$.

Además $\delta=\mathrm{mcd}(\alpha,\beta_1,\dots,\beta_n)$, pues ciertamente δ divide a α, $\beta_1,\dots,\beta_n$ (porque están en I) y, si γ es un divisor común de todos ellos, los divisores primos de γ serán algunos de los π_i (porque $\gamma\mid\alpha$) y $e_{\pi_i}(\gamma)\le e_{\pi_i}(\beta_i)=e(\pi_i)$, luego $\gamma\mid\delta$.

Por 4) tenemos que $(\delta)=(\alpha,\beta_1,\dots,\beta_n)\subset I$, luego $I=(\delta)$. ∎

Así pues, el teorema chino del resto, al igual que la relación de Bezout o la maximalidad de los ideales primos son características que determinan a los dominios de ideales principales entre los dominios de factorización única. Aparentemente, pues, no es posible demostrar el teorema chino del resto en dominios de factorización única que no sean DIPs, pero sucede que si reformulamos el teorema en términos de ideales, entonces es válido en cualquier dominio íntegro:

Teorema 3.56 (Teorema chino del resto) *Sea D un dominio íntegro y sean $\mathfrak{a}_1,\dots,\mathfrak{a}_n$ ideales de D tales que $\mathfrak{a}_i+\mathfrak{a}_j=D$ para $i\ne j$. Si $\alpha_1,\dots,\alpha_n\in D$, existe un $x\in D$ tal que $x\equiv\alpha_i\pmod{\mathfrak{a}_i}$ para $i=1,\dots,n$.*

Demostración: Para $n=2$ tenemos que $a_1+a_2=1$, para ciertos elementos $a_i\in\mathfrak{a}_i$, y basta tomar $x=\alpha_2a_1+\alpha_1a_2$.

En el caso general, para cada $i\ge2$ elegimos $a_i\in\mathfrak{a}_1$, $b_i\in\mathfrak{a}_i$ tales que $a_i+b_i=1$. Entonces

$$1=\prod_{i=2}^n(a_i+b_i)\in\mathfrak{a}_1+\prod_{i=2}^n\mathfrak{a}_i=D.$$

Por el caso ya probado existe un elemento $y_1\in D$ tal que $y_1\equiv1\pmod{\mathfrak{a}_1}$, $y_1\equiv0\pmod{\mathfrak{a}_i}$ para $i\ge2$. Similarmente podemos definir elementos $y_i\in D$ que cumplan

$$y_i\equiv1\pmod{\mathfrak{a}_i},\qquad y_i\equiv0\pmod{\mathfrak{a}_j}\quad\text{para } j\ne i.$$

Basta tomar $x=\alpha_1y_1+\cdots+\alpha_ny_n$. ∎

Capítulo IV

Módulos y espacios vectoriales

En el apéndice A de [IG] hemos visto que la geometría analítica nos permite identificar los puntos del plano con los elementos de $\mathbb{R}^2$, y en el apéndice A de [IC] hemos visto que, análogamente, es posible identificar los puntos del espacio con los elementos de $\mathbb{R}^3$. En general, podemos ver a $\mathbb{R}^n$ como un "espacio de n-dimensiones". Al interpretar geométricamente $\mathbb{R}^2$, $\mathbb{R}^3$ o, en general, $\mathbb{R}^n$, resulta natural considerar dos operaciones: la suma de vectores, dada por

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

y el producto por escalares, dado por

$$\alpha(x_1, \ldots, x_n) = (\alpha x_1, \ldots, \alpha x_n).$$

Ambas tienen una interpretación geométrica natural que estudiaremos con detalle en el capítulo III de [G], pero antes vamos a analizar aquí la estructura algebraica que adquiere $\mathbb{R}^n$ con estas operaciones, que resulta ser una de las estructuras algebraicas más versátiles, que aparece en los contextos matemáticos más diversos. Se trata de la estructura de "espacio vectorial", que a su vez es un caso particular de la estructura de "módulo". Estas estructuras aparecen, por ejemplo, en muchos contextos en los que unos objetos matemáticos pueden determinarse mediante "coordenadas". Es el caso de los puntos de una recta, de un plano o del espacio, pero nos encontramos situaciones análogas en la teoría algebraica de números. Por ejemplo, el teorema [ITAl 8.12] dice que los elementos de un cuerpo de la forma $\mathbb{Q}(\alpha)$, donde α es un número algebraico de grado n, están determinados por n coordenadas racionales. En [ITAl 4.3] definimos los enteros de Gauss, cada uno de los cuales está determinado por dos coordenadas enteras, y en la sección 17.2 de [ITAl] vimos que todo entero ciclotómico de orden 5 está determinado por cuatro coordenadas enteras, etc.

Vamos a ver que la teoría de módulos permite tratar conceptualmente de forma unificada todos estos objetos matemáticos y muchos más.

4.1 Módulos

Definición 4.1 Sea A un anillo unitario. Un *A–módulo izquierdo* es una terna $(M, +, \cdot)$ tal que M es un conjunto, $+ : M \times M \longrightarrow M$ es una operación interna en M y $\cdot$ es lo que se llama una *operación externa* en M con dominio de operadores en A, lo que significa simplemente que $\cdot : A \times M \longrightarrow M$. Además se han de cumplir las propiedades siguientes:

1. $(r + s) + t = r + (s + t)$ para todos los r, s, $t \in M$.
2. $r + s = s + r$ para todos los r, $s \in M$.
3. Existe un elemento $0 \in M$ tal que $r + 0 = r$ para todo $r \in M$.
4. Para todo $r \in M$ existe un elemento $-r \in M$ tal que $r + (-r) = 0$.
5. $a(r + s) = ar + as$ para todo $a \in A$ y todos los r, $s \in M$.
6. $(a + b)r = ar + br$ para todos los a, $b \in A$ y todo $r \in M$.
7. $a(br) = (ab)r$ para todos los a, $b \in A$ y todo $r \in M$.
8. $1r = r$ para todo $r \in M$.

Observemos que la suma en un módulo ha de cumplir las mismas propiedades que la suma en un anillo (que afirman que $(M, +)$ es un grupo abeliano, por lo que las propiedades elementales de la suma en anillos (o en grupos abelianos) valen para módulos. Por ejemplo, el elemento 0 que aparece en la propiedad 3 es único, así como los elementos simétricos que aparecen en 4.

Un A–módulo derecho se define igualmente cambiando la operación externa por otra de la forma $\cdot : M \times A \longrightarrow M$. La única diferencia significativa es que la propiedad 7 se convierte en $(rb)a = r(ba)$, que escrito por la izquierda sería $a(br) = (ba)r$ (en lugar de $a(br) = (ab)r$, que es la propiedad de los módulos izquierdos). Mientras no se indique lo contrario sólo consideraremos módulos izquierdos, aunque todo vale para módulos derechos. Seguiremos el mismo convenio que con los anillos, según el cual las operaciones de un módulo se representarán siempre con los mismos signos, aunque sean distintas en cada caso. También escribiremos M en lugar de $(M, +, \cdot)$.

Veremos que los módulos sobre cuerpos tienen un comportamiento mucho más simple y regular que sobre anillos arbitrarios, aunque en realidad conviene notar que la propiedad conmutativa del producto en los cuerpos no es relevante para ello.

Por eso definimos un *anillo de división* como un anillo unitario en el que $1 \neq 0$ y donde todo elemento no nulo tiene inverso para el producto. En otras palabras, es un anillo que cumple todos los axiomas de la definición de cuerpo salvo quizá la conmutatividad del producto.

Si D es un anillo de división, los D–módulos se llaman *espacios vectoriales,* y en tal caso es costumbre llamar *vectores* a los elementos del espacio y *escalares* a los elementos de D.

Tenemos disponibles muchos ejemplos de módulos:

1. Un $\mathbb{Z}$-módulo es esencialmente lo mismo que un grupo abeliano, pues todo grupo abeliano $(A, +)$ se convierte en $\mathbb{Z}$-módulo con el producto usual de un entero por un elemento del grupo y, recíprocamente, en todo $\mathbb{Z}$-módulo $(A, +, \cdot)$ el producto es el definido de forma usual a partir de la estructura de grupo de $(A, +)$.

2. Si A es un anillo unitario, entonces A es un A–módulo con su suma y su producto.

3. Más en general, si B es un anillo unitario y A es un subanillo que contenga a la identidad, entonces B es un A–módulo con la suma de B y el producto restringido a $A \times B$.

   Por ejemplo, podemos ver a un anillo de polinomios $A[x_1, \ldots, x_n]$ como A-módulo, o a $\mathbb{C}$ como $\mathbb{R}$-espacio vectorial, etc.

4. Más en general aún, si $\phi : A \longrightarrow B$ es un homomorfismo de anillos unitarios tal que $\phi(1) = 1$, entonces B es un A–módulo con la suma en B y el producto dado por $ab = \phi(a)b$ (el ejemplo anterior sería un caso particular de éste tomando como homomorfismo la inclusión).

5. Un caso particular de este ejemplo es que si A es un anillo unitario e I es un ideal de A, entonces el anillo cociente A/I es un A–módulo con su suma y el producto dado por $a[b] = [ab]$ (basta tomar como ϕ el epimorfismo canónico).

6. Otro caso particular es que si A es un anillo unitario, entonces A es un $\mathbb{Z}$–módulo con el producto usual de un entero por un elemento de A (tomando $\phi(m) = m1$).

7. Si M es cualquier A-módulo, también lo es M^n con las operaciones que hemos descrito en la introducción a este capítulo para el caso de $\mathbb{R}^n$.

8. Si A es cualquier conjunto en el que hay definida una suma que cumpla las cuatro primeras propiedades de la definición de módulo (que son las mismas que aparecen en la definición de anillo[1]), entonces A se convierte en un $\mathbb{Z}$-módulo sin más que definir el producto $\cdot : \mathbb{Z} \times A \longrightarrow A$ exactamente igual que en un anillo (definición 2.14).

Por otra parte, la definición 2.14 puede extenderse para definir el producto de números enteros por elementos de cualquier anillo. Enunciemos a continuación las propiedades elementales de los módulos. Todas se demuestran igual que para anillos.

[1]Un conjunto dotado de tal suma es lo que se conoce como un grupo abeliano, y lo que vemos es que un grupo abeliano es esencialmente lo mismo que un $\mathbb{Z}$-módulo, porque todo $\mathbb{Z}$-módulo es un grupo abeliano con su suma y todo grupo abeliano se convierte en $\mathbb{Z}$-módulo con el producto natural.

Teorema 4.2 *Sea A un anillo unitario y M un A–módulo.*

1. *Si $r+s=r+t$ entonces $s=t$ para todos los r, s, $t\in M$,*
2. *$r+r=r$ si y sólo si $r=0$, para todo $r\in M$,*
3. *$-(-r)=r$ para todo $r\in M$,*
4. *$-(r+s)=-r-s$, para todos los r, $s\in M$,*
5. *$a0=0r=0$, para todo $a\in A$ y todo $r\in M$,*
6. *$n(ar)=(na)r=a(nr)$, para todo $n\in\mathbb{Z}$, $a\in A$ y $r\in M$,*
7. *Si A es un anillo de división, $a\in A$, $r\in M$ y $ar=0$, entonces $a=0$ o $r=0$.*

(Por ejemplo, la propiedad 7 se cumple porque si $a\neq 0$, entonces existe a^{-1} y por tanto $a^{-1}ar=a^{-1}0=0$, $1r=0$, $r=0$).

Definición 4.3 Sea A un anillo unitario y M un A–módulo. Diremos que un módulo N es un *submódulo* de M si $N\subset M$ y las operaciones de N son las mismas que las de M. Cuando M es un espacio vectorial, sus submódulos se llaman *subespacios vectoriales.*

Evidentemente, si un subconjunto de un módulo dado puede ser dotado de estructura de submódulo, la forma de hacerlo es única (pues las operaciones en N han de ser las restricciones de las de M). Por tanto es indistinto hablar de submódulos de M que de subconjuntos que pueden ser estructurados como submódulos. Pero no siempre es posible considerar a un subconjunto como submódulo (por ejemplo si no contiene al 0). Las condiciones que se han de cumplir para ello las da el teorema siguiente:

Teorema 4.4 *Sea A un anillo unitario y M un A–módulo. Un subconjunto N de M puede ser dotado de estructura de submódulo si y sólo si cumple las condiciones siguientes:*

1. *$N\neq\varnothing$,*
2. *Si r, $s\in N$ entonces $r+s\in N$,*
3. *Si $a\in A$ y $r\in N$, entonces $ar\in N$.*

Demostración: Obviamente si N es un submódulo ha de cumplir estas condiciones. Si N cumple estas condiciones entonces por 2) y 3) la suma y el producto están definidos en N. Por 1) existe un $r\in N$, por 3) $-r=(-1)r\in N$, por 2) $0=r-r\in N$. Por tanto N tiene neutro y de nuevo por 3) el simétrico de cada elemento de N está en N. El resto de las propiedades exigidas por la definición se cumplen por cumplirse en M. ∎

Las condiciones 2) y 3) del teorema anterior pueden resumirse en una sola:

Teorema 4.5 *Sea A un anillo unitario y M un A–módulo. Un subconjunto N de M puede ser dotado de estructura de submódulo si y sólo si $N \neq \varnothing$ y para todos los a, $b \in A$ y todos los r, $s \in N$ se cumple que $ar + bs \in N$.*

Definición 4.6 De los teoremas anteriores se desprende que si A es un anillo unitario y M es un A–módulo, entonces M y $0 = \{0\}$ son submódulos de M, y se llaman *submódulos impropios.* Cualquier otro submódulo de M se llama *submódulo propio.* El submódulo 0 se llama también *submódulo trivial.*

Es obvio que la intersección de una familia de submódulos de M es un submódulo de M. Si X es un subconjunto de M, llamaremos *submódulo generado* por X a la intersección de todos los submódulos de M que contienen a X. Lo representaremos $\langle X\rangle$.

Es inmediato a partir de la definición que si N es un submódulo de M y $X \subset N$, entonces $\langle X\rangle \subset N$. Igualmente si $X \subset Y \subset M$, entonces se cumple $\langle X\rangle \subset \langle Y\rangle$. Notemos que $\langle\varnothing\rangle = 0$. Cuando el conjunto X sea finito, digamos $X = \{x_1, \ldots, x_n\}$, escribiremos también $\langle X\rangle = \langle x_1, \ldots, x_n\rangle$.

Si $M = \langle X\rangle$ diremos que el conjunto X es un *sistema generador* de M. Diremos que M es *finitamente generado* si tiene un sistema generador finito. El modulo M es *monógeno* si admite un generador con un solo elemento.

Al considerar a un anillo conmutativo y unitario A como A–módulo los submódulos coinciden con los ideales, de donde se sigue que el submódulo generado por un subconjunto X coincide con el ideal generado por X, es decir, $(X) = \langle X\rangle$.

Similarmente, en un $\mathbb{Z}$-módulo los submódulos coinciden con los subgrupos abelianos, y el submódulo generado por un subconjunto es el mismo que el subgrupo generado.

Es fácil reconocer los elementos del submódulo generado por un subconjunto:

Teorema 4.7 *Sea A un anillo unitario, M un A–módulo y $X \subset M$. Entonces*

$$\langle X\rangle = \{a_1r_1 + \cdots + a_nr_r \mid n \in \mathbb{N}, a_i \in A, r_i \in X\}\,.$$

La prueba es sencilla (es idéntica a la del teorema 3.4): un submódulo que contenga a X ha de contener necesariamente al conjunto de la derecha, pero es fácil ver que este subconjunto es de hecho un submódulo, luego contiene a $\langle X\rangle$.

Los elementos de la forma $a_1r_1 + \cdots + a_nr_n$, con los $a_i \in A$, se llaman *combinaciones lineales* de $r_1, \ldots, r_n$. En estos términos, el teorema anterior afirma que el submódulo generado por un conjunto X es el conjunto de todas las combinaciones lineales de los elementos de X.

Ejemplos Si A es un anillo unitario, en A^n podemos considerar la *base canónica,* formada por las n-tuplas e_i que tienen todas sus componentes nulas salvo la i-ésima, que es igual a 1. Por ejemplo, la base canónica de A^3 está formada por las ternas $(1,0,0), (0,1,0), (0,0,1)$, y es claro que, en general, $A^n = \langle e_1, \ldots, e_n\rangle$, pues claramente

$$(a_1, \ldots, a_n) = a_1e_1 + \cdots + a_ne_n.$$

Aunque esto lo analizaremos en el capítulo III de [G], sabemos por el apéndice A de [IC] que si $v \in \mathbb{R}^3$ no es nulo, el subespacio $\langle v\rangle = \{\alpha v \mid \alpha \in \mathbb{R}\}$ está formado por los puntos de la recta que pasa por el punto 0 con vector director v, mientras que si $v, w \in \mathbb{R}^3$ son vectores no nulos con direcciones distintas, el subespacio $\langle v, w\rangle = \{\alpha v + \beta w \mid \alpha, \beta \in \mathbb{R}\}$ contiene los puntos del plano que contiene a las rectas $\langle v\rangle$ y $\langle w\rangle$.

Podemos considerar el cuerpo $\mathbb{C}$ de los números complejos como un $\mathbb{R}$-espacio vectorial, y entonces el hecho de que todo número complejo sea de la forma $a + bi$, donde a, b son números reales equivale a que $\mathbb{C} = \langle 1, i\rangle$, mientras que los subespacios $\langle 1\rangle$ y $\langle i\rangle$ son los ejes real e imaginario, respectivamente. Por otro lado, también podemos considerar a $\mathbb{C}$ como $\mathbb{C}$-espacio vectorial, en cuyo caso, trivialmente $\mathbb{C} = \langle 1\rangle$.

En situaciones como ésta en las que puede haber ambigüedad es costumbre explicitar el anillo que estamos tomando como domino de operadores para distinguir, por ejemplo, el eje real $\langle 1\rangle_{\mathbb{R}}$ de $\langle 1\rangle_{\mathbb{C}} = \mathbb{C}$.

Si comparamos 3.4 con 4.7 vemos que si A es un anillo conmutativo y unitario y $X \subset A$, entonces $(X) = \langle X\rangle_A$, es decir, que el ideal generado por X no es sino el A-módulo generado por X.

En [ITAl 12.1] definimos módulos en un cuerpo cuadrático $\mathbb{Q}(\sqrt{d})$, que no son sino los $\mathbb{Z}$-submódulos de $\mathbb{Q}(\sqrt{d})$ que admiten un generador finito. En los capítulos XII y XIII de [ITAl] trabajamos a menudo con estos módulos, por ejemplo, el teorema [ITAl 13.8] determina cuáles de ellos son ideales de un orden cuadrático, y en el ejemplo posterior vimos que, en el anillo $\mathbb{Z}\big[\sqrt{-5}\,\big]$, tenemos la igualdad

$$\mathfrak{p} = (2, 1+\sqrt{-5}) = \big\langle 2, 1+\sqrt{-5}\big\rangle_{\mathbb{Z}[\sqrt{-5}]} = \big\langle 2, 1+\sqrt{-5}\big\rangle_{\mathbb{Z}},$$

donde la única desigualdad no trivial es la última, que expresa que las combinaciones lineales con coeficientes en $\mathbb{Z}\big[\sqrt{-5}\,\big]$ de los dos generadores se reducen en realidad a sus combinaciones lineales con coeficientes en $\mathbb{Z}$. Esto no es cierto en general, y supone una simplificación drástica en la descripción del ideal.

El hecho de que todo entero ciclotómico $\mathbb{Z}[\omega]$ de orden 5 se exprese en la forma

$$a_3\omega^3 + a_2\omega^2 + a_1\omega + a_0,$$

donde los a_i son números enteros, significa que el anillo $\mathbb{Z}[\omega]$ de los enteros ciclotómicos de orden 5 es $\mathbb{Z}[\omega] = \big\langle 1, \omega, \omega^2, \omega^3\big\rangle_{\mathbb{Z}}$, mientras que $\big\langle 1, \omega, \omega^2, \omega^3\big\rangle_{\mathbb{Q}}$ es el conjunto de todas las combinaciones lineales de los generadores con coeficientes en $\mathbb{Q}$, es decir, el cuerpo ciclotómico $\mathbb{Q}[\omega]$.

Un ejemplo de generador infinito lo obtenemos considerando un anillo de polinomios $A[x]$, donde A es un anillo unitario. Claramente $A[x] = \langle X\rangle$, donde $X = \{x^n \mid n \in \mathbb{N}\}$ es el conjunto de las potencias de x. El submódulo $\big\langle 1, x, x^2\big\rangle$ está formado por los polinomios de grado ≤ 2. ∎

Los módulos cociente se definen exactamente igual que los anillos cociente:

Definición 4.8 Sea A un anillo unitario, M un A–módulo y N un submódulo de M. Definimos en M la relación de *congruencia* módulo N mediante

$$r \equiv s \text{ (mód } N) \text{ si y sólo si } r - s \in N.$$

Es fácil probar que se trata de una relación de equivalencia en M. Llamaremos M/N al conjunto cociente. La clase de equivalencia de un elemento $r \in M$ es

$$[r] = r + N = \{r + s \mid s \in N\}.$$

La prueba del teorema siguiente es completamente análoga a la de 3.38:

Teorema 4.9 *Sea A un anillo unitario, M un A–módulo y N un submódulo de M. El conjunto M/N es un A–módulo con las operaciones dadas por*

$$[r] + [s] = [r + s] \quad y \quad a[r] = [ar].$$

Se le llama módulo cociente.

Es fácil ver que si I es un ideal de un anillo A, entonces el A-módulo cociente A/I es el módulo que resulta de considerar al anillo cociente A/I como A-módulo a través del epimorfismo canónico $A \longrightarrow A/I$.

Definimos los homomorfismos de módulos de forma análoga a los de anillos. Su interpretación es la misma:

Definición 4.10 Sea A un anillo unitario y M, N dos A–módulos. Una aplicación $f : M \longrightarrow N$ es un *homomorfismo de módulos* si cumple:

$f(r + s) = f(r) + f(s)$, para todos los $r, s \in M$,

$f(ar) = af(r)$, para todo $a \in A$ y todo $r \in M$.

Obviamente esto equivale a que $f(ar + bs) = af(r) + bf(s)$, para $a, b \in A$, $r, s \in M$.

- Un *monomorfismo* de módulos es un homomorfismo inyectivo.
- Un *epimorfismo* de módulos es un homomorfismo suprayectivo.
- Un *isomorfismo* de módulos es un homomorfismo biyectivo.
- Un *endomorfismo* de módulos es un homomorfismo de un módulo en sí mismo.
- Un *automorfismo* de módulos es un isomorfismo de un módulo en sí mismo.
- Los homomorfismos de espacios vectoriales se llaman *aplicaciones lineales.*

La composición de homomorfismos es un homomorfismo, la inversa de un isomorfismo es un isomorfismo. Dos módulos M y N son *isomorfos* ($M \cong N$) si existe un isomorfismo entre ellos.

Si $f : M \longrightarrow N$ es un homomorfismo de módulos, es fácil ver que si N' es un submódulo de N entonces $f^{-1}[N']$ es un submódulo de M y si M' es un submódulo de M, entonces $f[M']$ es un submódulo de N.

En particular, se define el *núcleo* de f como $\mathrm{N}(f) = \{r \in M \mid f(r) = 0\}$, es decir, $\mathrm{N}(f) = f^{-1}[0]$, que es un submódulo de M. La *imagen* de f es $\operatorname{Im} f = f[M] = \{f(r) \mid r \in M\}$, que es un submódulo de N.

Si A es un anillo unitario, M es un A–módulo y N es un submódulo de M, la aplicación $f : M \longrightarrow M/N$ dada por $f(r) = [r]$ es un epimorfismo de módulos llamado *epimorfismo canónico.* Se cumple que $\mathrm{N}(f) = N$.

Dado un módulo M y submódulos $N \subset N' \subset M$ es claro que N'/N es un submódulo de M/N. Es fácil probar que todo submódulo de M/N es de esta forma.

Los teoremas siguientes son análogos a 3.46 y 3.47, respectivamente:

Teorema 4.11 *Un homomorfismo de módulos es inyectivo si y sólo si su núcleo es trivial.*

Teorema 4.12 (Teorema de isomorfía) *Consideremos un anillo unitario A y sea $f : M \longrightarrow N$ un homomorfismo de A–módulos. Entonces la aplicación $\bar{f} : M/\mathrm{N}(f) \longrightarrow \operatorname{Im} f$ definida por $\bar{f}([r]) = f(r)$ es un isomorfismo de módulos.*

Ejemplo Consideremos el conjunto c formado por todas las sucesiones convergentes de números reales, que adquiere estructura de $\mathbb{R}$-espacio vectorial con las operaciones dadas por

$$\{x_n\}_{n=0}^{\infty} + \{y_n\}_{n=0}^{\infty} = \{x_n + y_n\}_{n=0}^{\infty}, \qquad \alpha\{x_n\}_{n=0}^{\infty} = \{\alpha x_n\}_{n=0}^{\infty}.$$

La aplicación $L : c \longrightarrow \mathbb{R}$ que a cada sucesión le hace corresponder su límite es claramente un epimorfismo de espacios vectoriales, cuyo núcleo es el subespacio c_0 formado por las sucesiones convergentes a 0. El teorema de isomorfía nos da que $c/c_0 \cong \mathbb{R}$.

Por otra parte, podemos considerar la aplicación $i : \mathbb{R} \longrightarrow c$ que a cada número real α le hace corresponder la sucesión constante $i(\alpha) = \{\alpha\}_{n=0}^{\infty}$. También es claro que i es un monomorfismo de espacios vectoriales, que nos permite identificar a $\mathbb{R}$ con un subespacio vectorial de c (el subespacio de las sucesiones constantes).

Consideramos ahora la aplicación $\pi : c \longrightarrow c_0$ dada por $\pi(s) = s - i(L(s))$, es decir, la aplicación que a cada sucesión convergente le asigna la sucesión que resulta de restarle la sucesión constantemente igual a su límite, que obviamente converge a 0. Se cumple que su núcleo es $\mathbb{R}$ (el subespacio de las sucesiones constantes), pues si $\pi(s) = 0$, entonces $s = i(L(s)) \in \mathbb{R}$, y el recíproco es inmediato. Así pues, el teorema de isomorfía nos da también que $c/\mathbb{R} \cong c_0$.

■

4.2 Suma de módulos

Una de las razones por las que es útil la estructura de módulo es porque permite obtener descripciones sencillas de los elementos de un módulo que resultan fáciles de manejar. Un primer análisis de la estructura de un módulo consiste en descomponerlo en suma de módulos más simples en el sentido en que exponemos a continuación. Antes veamos un par de ejemplos sencillos:

Ejemplo Continuando con el ejemplo precedente, podemos afirmar que

$$c = c_0 + \mathbb{R},$$

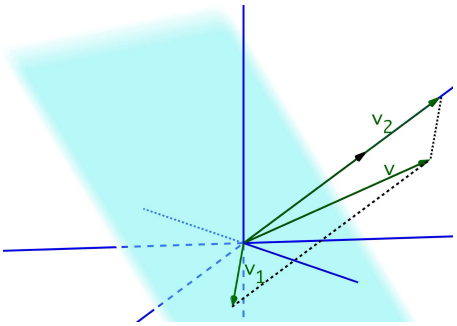
donde esto hay que entenderlo como que todo elemento de c es suma de un elemento de c_0 más un elemento de $\mathbb{R}$, es decir, que toda sucesión convergente es la suma de una sucesión convergente a 0 más una sucesión constante. ∎

Ejemplo Sea K un cuerpo (por simplicidad, de característica 0, por ejemplo $\mathbb{R}$, o cualquier otro). Consideramos el espacio vectorial $V = K^3$. Es fácil ver que

$$V_1 = \{(x, y, z) \in V \mid x + y + z = 0\}$$

es un subespacio vectorial de V (la suma de ternas cuyas componentes suman 0 es una terna de suma 0 y el producto de un escalar por una terna de suma 0 tiene suma 0). Por otro lado, $V_2 = \langle(1, 1, 1)\rangle$ es por definición un subespacio vectorial.

Ahora es fácil ver que $V = V_1 + V_2$, donde, como en el ejemplo anterior, esto significa que todo vector de V se puede expresar como suma de un vector de V_1 y otro de V_2. En efecto, si $v = (x, y, z) \in V$ y $s = x + y + z$, sólo tenemos que considerar $v_2 = (s/3)(1, 1, 1)$ y $v_1 = v - v_2$, con lo que $v_1 \in V_1$, $v_2 \in V_2$ y $v = v_1 + v_2$. Esto tiene una interpretación geométrica sencilla:



Si $K = \mathbb{R}$, entonces V_1 es un plano y V_2 una recta perpendicular, y lo que estamos afirmando es que todo vector del espacio se puede descomponer en suma de un vector del plano y otro de la recta. ∎

Estas descomposiciones de un espacio vectorial en suma de dos subespacios son casos particulares de la definición siguiente:

Definición 4.13 Sea A un anillo conmutativo y unitario y M un A–módulo. Llamaremos *suma* de una familia de submódulos $\{N_i\}_{i\in I}$ de M al submódulo

$$\sum_{i\in I} N_i = \langle \bigcup_{i\in I} N_i \rangle$$

Así, la suma de una familia de submódulos es el menor submódulo que los contiene a todos. Por el teorema 4.7, un elemento de este submódulo es una combinación lineal de elementos de algunos de los módulos N_i con coeficientes en A, pero al multiplicar un elemento de N_i por un elemento de A obtenemos otro elemento de N_i, y la suma de dos elementos de un mismo N_i está también en N_i. Por lo tanto un elemento de $\sum_{i\in I} N_i$ es de la forma $r_1 + \cdots + r_n$, donde cada sumando está en un submódulo distinto. En particular si la familia es finita,

$$N_1 + \cdots + N_n = \{r_1 + \cdots + r_n \mid r_i \in N_i\}.$$

Ahora bien, en los dos ejemplos que hemos puesto, no sólo es cierto que cada sucesión de c sea suma de dos sucesiones de c_0 y $\mathbb{R}$, o que cada vector de V se expresa como suma de un vector de V_1 y otro de V_2, sino que en ambos casos la expresión es única. Esto se puede probar directamente sin dificultad, pero es aún más inmediato si contamos con la teoría adecuada:

Definición 4.14 Sea A un anillo unitario y M un A–módulo. Se dice que una familia de submódulos $\{N_i\}_{i\in I}$ es *independiente* si para cada índice i se cumple

$$N_i \cap \sum_{j\neq i} N_j = 0.$$

Si $\{N_i\}_{i\in I}$ es una familia de submódulos independientes, se dice que su suma es *directa* y en lugar de $\sum_{i\in I} N_i$ se escribe $\bigoplus_{i\in I} N_i$.

Teorema 4.15 *Sea A un anillo unitario, M un A–módulo y $N_1, \ldots, N_n$ una familia de submódulos tales que $M = N_1 + \cdots + N_n$. Equivalen:*

1. *$M = N_1 \oplus \cdots \oplus N_n$.*
2. *Si $m_1 + \cdots + m_n = 0$ con cada $m_i \in N_i$, entonces cada $m_i = 0$.*
3. *Cada elemento $m \in M$ se expresa de forma única como suma*

   $$m = m_1 + \cdots + m_n$$

   con cada $m_i \in N_i$.

Demostración: Por simplificar, vamos a probarlo para el caso $n = 3$. El caso general es análogo. De hecho el resultado es cierto incluso con infinitos sumandos.

1) $\Rightarrow$ 2). Si $m_1 + m_2 + m_3 = 0$ con cada $m_i \in N_i$, entonces

$$m_1 = -m_2 - m_3 \in N_1 \cap (N_2 + N_3) = 0,$$

y análogamente se concluye que $m_2 = m_3 = 0$.

2) ⇒ 3). Como $M = N_1 + N_2 + N_3$, todo elemento de M se descompone en una suma de la forma $m_1 + m_2 + m_3$, con cada $m_i \in N_i$. Si un elemento admite dos descomposiciones

$$m_1 + m_2 + m_3 = m'_1 + m'_2 + m'_3$$

entonces $(m_1 - m'_1) + (m_2 - m'_2) + (m_3 - m'_3) = 0$, luego por 2) podemos concluir que $(m_1 - m'_1) = (m_2 - m'_2) = (m_3 - m'_3) = 0$, luego ambas descomposiciones son la misma.

3) ⇒ 1). Si un elemento $m_1 \in N_1 \cap (N_2 + N_3)$, entonces $m_1 = m_2 + m_3$, $m_i \in N_i$, o sea, $m_1 + 0 + 0 = 0 + m_2 + m_3$, luego por la unicidad, $m_1 = 0$, es decir, $N_1 \cap (N_2 + N_3)$ es el submódulo trivial. Igualmente ocurre si permutamos los índices. ∎

Nota Hemos enunciado el teorema anterior para sumas finitas porque es el caso que realmente vamos a necesitar, y el enunciado general para sumas posiblemente infinitas oculta parcialmente la idea básica. En general, si $M = \sum_{i \in I} N_i$, que la suma sea directa equivale a que siempre que una suma $m_{i_1} + \cdots + m_{i_n} = 0$ con $m_{i_j} \in N_{i_j}$ y los índices $i_j \in I$ son distintos dos a dos, entonces cada $m_{i_j} = 0$ (ésta es la versión general de 2), lo cual equivale a su vez a que todo $m \in M$ *no nulo* se expresa de forma única como suma $m = m_{i_1} + \cdots + m_{i_n}$ con $m_{i_j} \in N_{i_j}$ *no nulo* y los índices $i_j \in I$ distintos dos a dos (y ésta es la versión general de 3). La prueba del caso general se obtiene modificando ligeramente la que hemos visto. ∎

Así pues, la condición para que una suma de submódulos sea directa (con la consecuente unicidad de los sumandos) la condición necesaria y suficiente es que la intersección de uno de los sumandos con la suma de los demás sea trivial. En particular, si sólo hay dos sumandos, la condición es simplemente que su intersección sea trivial, y esto sucede en los dos ejemplos que hemos puesto:

Ejemplos Por una parte, $c = c_0 \oplus \mathbb{R}$, pues $c_0 \cap \mathbb{R} = 0$, ya que una sucesión que sea a la vez constante y convergente a 0 tiene que ser necesariamente la sucesión nula.

Por otra parte, $V = V_1 \oplus V_2$, pues $V_1 \cap V_2 = 0$, ya que si un vector tiene componentes que suman 0 y además es de la forma $v = \alpha(1,1,1) = (\alpha, \alpha, \alpha)$, necesariamente $\alpha = 0$, y entonces $v = 0$. ∎

En general, si un módulo M se expresa como suma directa de una familia de n submódulos, entonces la estructura de módulo de M está completamente determinada por las estructuras de los submódulos, a través del concepto siguiente de producto de módulos:

Definición 4.16 Sea A un anillo unitario y $M_1, \ldots, M_n$ una familia de A–módulos. Entonces el producto cartesiano

$$M_1 \times \cdots \times M_n = \big\{(m_1, \ldots, m_n) \mid m_j \in M_j \text{ para cada } j = 1, \ldots, n\big\}$$

es un A–módulo con las operaciones dadas por

$$(m_1,\dots,m_n)+(m'_1,\dots,m'_n)=(m_1+m'_1,\dots,m_n+m'_n),$$
$$r\cdot(m_1,\dots,m_n)=(rm_1,\dots,rm_n).$$

Esto vale en particular cuando todos los factores son un mismo módulo M, con lo que tenemos la estructura de A–módulo sobre el producto M^n que ya habíamos puesto más arriba como ejemplo.

Es obvio que la aplicación $\iota_i : M_i \longrightarrow M_1\times\cdots\times M_n$ que a cada elemento $m\in M_i$ le asigna la n–tupla cuya componente i–ésima es m y las restantes son 0 es un monomorfismo de módulos, por lo que podemos identificar a cada M_i con su imagen, es decir, con el submódulo de $M_1\times\cdots\times M_n$ formado por las n–tuplas que tiene todas sus componentes nulas salvo la i–ésima. La única precaución es que en principio puede ocurrir que $M_i=M_j$, mientras que sus imágenes respectivas por ι_i y ι_j serán submódulos de $M_1\times\cdots\times M_n$ distintos entre sí (isomorfos, pero distintos).

También es inmediato que si $(m_1,\dots,m_n)\in M_1\times\cdots\times M_n$ entonces

$$(m_1,\dots,m_n)=\iota_1(m_1)+\cdots+\iota_n(m_n),$$

luego $M_1\times\cdots\times M_n=M_1+\cdots+M_n$. Además $M_2+\cdots+M_n$ está formado por las n–tuplas con la primera componente nula, luego $M_1\cap(M_2+\cdots+M_n)=0$. Lo mismo vale con otros índices, con lo que $M_1\times\cdots\times M_n=M_1\oplus\cdots\oplus M_n$.

En resumen, dada una familia de A–módulos, hemos construido un A–módulo que es suma directa de una familia de submódulos isomorfos a los dados. Recíprocamente, si un módulo M es suma directa de una familia de submódulos $M=M_1\oplus\cdots\oplus M_n$, entonces M es isomorfo al producto cartesiano $M_1\times\cdots\times M_n$. El isomorfismo es la aplicación que a cada elemento de M le asigna la n–tupla formada por los elementos en los que se descompone según el teorema 4.15.

Estos resultados son ciertos en el caso de tener infinitos módulos, pero con una precaución:

Si tenemos una familia de A–módulos $\{M_i\}_{i\in I}$, el producto cartesiano $\prod_{i\in I} M_i$ es un A–módulo con las operaciones definidas por

$$(x_i)_{i\in I}+(y_i)_{i\in I}=(x_i+y_i)_{i\in I},\qquad r\cdot(x_i)_{i\in I}=(r\cdot x_i)_{i\in I}.$$

Igualmente, las aplicaciones $\iota_i : M_i\longrightarrow\prod_{i\in I} M_i$ dadas por

$$\iota_i(m)(j)=\begin{cases} m & \text{si } j=i,\\ 0 & \text{si } j\neq i,\end{cases}$$

son monomorfismos de módulos, pero al identificar cada módulo M_i con su imagen ya no es cierto que $\prod_{i\in I} M_i=\sum_{i\in I} M_i$, sino que

$$\sum_{i\in I} M_i=\Big\{f\in\prod_{i\in I} M_i \;\Big|\; \{i\in I\mid f(i)\neq 0\}\text{ es finito}\Big\},$$

pues el miembro derecho es claramente un submódulo del producto cartesiano que contiene a todos los submódulos M_i, luego también a su suma. Esto nos da la inclusión $\subset$, y la contraria se sigue inmediatamente de que todo elemento no nulo del miembro derecho es suma de un número finito de elementos de los submódulos M_i:

$$f = \iota_{i_1}(f(i_1)) + \cdots + \iota_{i_n}(f(i_n)),$$

donde $\{i_1, \ldots, i_n\} = \{i \in I \mid f(i) \neq 0\}$.

Pero la suma de los submódulos $\iota_i[M_i]$ sigue siendo directa (por el mismo argumento que en el caso finito), por lo que definimos la *suma directa externa* de los módulos $\{M_i\}_{i\in I}$ como

$$\bigoplus_{i\in I} M_i = \Big\{ f \in \prod_{i\in I} M_i \;\Big|\; \{i \in I \mid f(i) \neq 0\} \text{ es finito} \Big\}.$$

De este modo, dada una familia arbitraria de A–módulos, podemos construir un módulo que se expresa como suma directa de submódulos isomorfos a los módulos dados. Pero es importante recordar que esta suma directa externa no coincide con el producto cartesiano cuando el número de factores es infinito. En tal caso tenemos dos módulos distintos:

$$\bigoplus_{i\in I} M_i \subset \prod_{i\in I} M_i.$$

El segundo consta de "tuplas infinitas" arbitrarias, mientras que el primero contiene las que sólo tienen una cantidad finita de componentes no nulas.

Alrededor de las sumas directas y productos hay definidos diversos homomorfismos naturales. Todas las propiedades concernientes a ellos se demuestran fácilmente y los dejamos como ejercicios para el lector:

- Las proyecciones: $\pi_i : \prod_{i\in I} M_i \longrightarrow M_i$ dadas por $\pi_i(f) = f(i)$ son claramente epimorfismos de módulos, que a su vez se restringen a epimorfismos $\pi_i : \bigoplus_{i\in I} M_i \longrightarrow M_i$.

  Cuando el número de factores es finito, $M = M_1 \oplus \cdots \oplus M_n$, las proyecciones vienen dadas, más explícitamente, por $\pi_i(m_1 + \cdots + m_n) = m_i$.

- Si tenemos homomorfismos de módulos $f_i : M_i \longrightarrow N_i$, podemos definir un homomorfismo $\prod_{i\in I} f_i : \prod_{i\in I} M_i \longrightarrow \prod_{i\in I} N_i$ mediante

  $$(\prod_{i\in I} f_i)((m_i)_{i\in I}) = (f_i(m_i))_{i\in I}.$$

  Éste se restringe a su vez a un homomorfismo $\bigoplus_{i\in I} f_i : \bigoplus_{i\in I} M_i \longrightarrow \bigoplus_{i\in I} N_i$.

  En el caso finito la definición es

  $$(f_1 \oplus \cdots \oplus f_n)(m_1, \ldots, m_n) = (f_1(m_1), \ldots, f_n(m_n)).$$

- Dados homomorfismos de módulos $f_i : M_i \longrightarrow N$, podemos definir otro homomorfismo $\sum_{i\in I} f_i : \bigoplus_{i\in I} M_i \longrightarrow N$ mediante

$$(\sum_{i\in I} f_i)(x) = \sum_{i\in I} f_i(x_i),$$

donde es fundamental que la suma de la derecha tiene un número finito de sumandos no nulos, pues en otro caso no estaría definida.

En el caso finito queda

$$(f_1 + \cdots + f_n)(m_1, \ldots, m_n) = f_1(m_1) + \cdots + f_m(m_n).$$

- Dados homomorfismos de módulos $f_i : M \longrightarrow N_i$, podemos definir otro homomorfismo $\prod_{i\in I} f_i : M \longrightarrow \prod_{i\in I} N_i$ mediante

$$(\prod_{i\in I} f_i)(m)(i) = f_i(m).$$

En el caso finito: $(f_1 \times \cdots \times f_n)(m) = (f_1(m), \ldots, f_n(m))$.

Es posible enunciar y demostrar muchas propiedades elementales de estos homomorfismos. Por ejemplo, si $f_i : M_i \longrightarrow N_i$ son isomorfismos, entonces $\prod_{i\in I} f_i : \prod_{i\in I} M_i \longrightarrow \prod_{i\in I} N_i$ y $\bigoplus_{i\in I} f_i : \bigoplus_{i\in I} M_i \longrightarrow \bigoplus_{i\in I} N_i$ son también isomorfismos.

4.3 Módulos libres

Si X es un sistema generador de un módulo M, eso significa que cada elemento de M se expresa como combinación lineal de elementos de X. En general la expresión no tiene por qué ser única, pero lo cierto es que en muchos casos lo es. Por ejemplo, tenemos que $\mathbb{C} = \langle 1, i\rangle_{\mathbb{R}}$ y no sólo es cierto que todo número complejo es de la forma $z = x + yi$, sino que z determina unívocamente su parte real x y su parte imaginaria y. Similarmente, si e_1, e_2, e_3 es la base canónica de $\mathbb{R}^3$, entonces $\mathbb{R}^3 = \langle e_1, e_2, e_3\rangle$ y todo vector de $\mathbb{R}^3$ se expresa de forma única como $v = xe_1 + ye_2 + ze_3$.

En cambio, si consideramos el anillo de enteros ciclotómicos de orden 5, resulta natural considerar el generador $\mathbb{Z}[\omega] = \langle 1, \omega, \omega^2, \omega^3, \omega^4\rangle_{\mathbb{Z}}$, pero las combinaciones lineales de los elementos de este generador no son únicas, sino que, por ejemplo,

$$3\omega^4 + \omega^3 + 5\omega^2 + 2\omega + 7 = 4\omega^4 + 2\omega^3 + 6\omega^2 + 3\omega + 8.$$

Sabemos que para tener la unicidad basta eliminar la última potencia, y considerar $\mathbb{Z}[\omega] = \langle 1, \omega, \omega^2, \omega^3\rangle_{\mathbb{Z}}$. Tenemos así dos sistemas generadores de $\mathbb{Z}[\omega]$, uno que puede parecer más natural, pero que no garantiza la unicidad de las combinaciones lineales, y otro que sí que la garantiza. Todos estos hechos son casos particulares de una teoría algebraica general.

Definición 4.17 Sea A un anillo unitario, M un A–módulo y X un subconjunto de M. Diremos que X es un conjunto *libre,* o que sus elementos son *linealmente independientes,* si para todos los $x_1, \ldots, x_n \in X$ distintos dos a dos[2] y todos los $a_1, \ldots, a_n \in A$, la igualdad $a_1x_1 + \cdots + a_nx_n = 0$ sólo se da en el caso trivial en que $a_1 = \cdots = a_n = 0$. En caso contrario se dice que X es un conjunto *ligado* o que sus elementos son *linealmente dependientes.*

En otras palabras, un conjunto X es libre si la única forma de expresar el 0 como combinación lineal de elementos de X es tomando todos los coeficientes nulos.

Es claro que un conjunto que contenga a 0 es ligado, pues $1 \cdot 0 = 0$ es una combinación lineal no trivial nula.

Un subconjunto X de un A–módulo M es una *base* de M si es un generador libre. Un módulo es *libre* si tiene una base.

Notemos que las definiciones implican trivialmente que $\varnothing$ es una base del módulo nulo.

Por ejemplo, es fácil ver que una base de A^n es $\{e_1, \ldots, e_n\}$, donde $e_i \in A^n$ es la n-tupla dada por

$$e_{ij} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j, \end{cases}$$

que es la que ya habíamos definido como *base canónica* de A^n.

Teorema 4.18 *Si A es un anillo unitario, M es un A-módulo y $x_1, \ldots, x_n$ es un sistema generador de M, entonces $x_1, \ldots, x_n$ es una base de M si y sólo si todo elemento $m \in M$ se expresa de forma única como combinación lineal $m = a_1x_1 + \cdots + a_nx_n$.*

Demostración: Si se da la unicidad, al aplicarla a $m = 0$ obtenemos que 0 se expresa de forma única como combinación lineal de $x_1, \ldots, x_n$, lo que implica que es un sistema libre. Recíprocamente, si el sistema es una base, por ser un sistema generador, todo elemento $m \in M$ es combinación lineal de sus elementos, y si tenemos dos expresiones, digamos

$$m = a_1x_1 + \cdots + a_nx_n = b_1x_1 + \cdots + b_nx_n,$$

entonces

$$(a_1 - b_1)x_1 + \cdots + (a_n - b_n)x_n = 0,$$

luego, por la independencia lineal, $a_1 - b_1 = \cdots = a_n - b_n = 0$, es decir, ambas combinaciones lineales tienen los mismos coeficientes. ∎

[2]Notemos que los elementos de X con los que formamos la combinación lineal tienen que ser distintos dos a dos. De lo contrario, todo conjunto no vacío sería ligado. Podemos definir también que una n-tupla $(x_1, \ldots, x_n)$ es linealmente independiente si cuando se cumple $a_1x_1 + \cdots + a_nx_n = 0$ necesariamente $a_1 = \cdots = a_n = 0$, y esto implica que los x_i son distintos dos a dos, pues si $x_i = x_j$ con $i \neq j$, bastaría tomar $a_i = 1$, $a_j = -1$ y los demás coeficientes nulos para tener una combinación lineal no trivial nula.

Definición 4.19 Si A es un anillo unitario, M es un A-módulo libre de base $X = \{x_1, \ldots, x_n\}$ y $m \in M$, los coeficientes que cumplen $m = a_1x_1 + \cdots + a_nx_n$ se llaman *coordenadas* de m en la base X (esto presupone que hemos fijado una ordenación de la base).

Teniendo en cuenta que

$$(a_1x_1 + \cdots + a_nx_n) + (b_1x_1 + \cdots + b_nx_n) = (a_1 + b_1)x_1 + \cdots + (a_n + b_n)x_n,$$

$$a(a_1x_1 + \cdots + a_nx_n) = aa_1x_1 + \cdots + aa_nx_n,$$

es obvio que la aplicación $M \longrightarrow A^n$ que a cada elemento de M le asigna su n–tupla de coordenadas es un isomorfismo de módulos, o sea, $M \cong A^n$.

Teniendo en cuenta que A^n siempre es un A-módulo libre, tenemos probado el teorema siguiente:

Teorema 4.20 *Sea A un anillo unitario y M un A–módulo. Se cumple que M tiene una base con n elementos si y sólo si M es isomorfo al A–módulo A^n.*

Observemos que si $X = \{x_1, \ldots, x_n\}$ es una base de un A–módulo M, entonces

$$M = \langle x_1 \rangle \oplus \cdots \oplus \langle x_n \rangle,$$

pues por el teorema 4.7 todo elemento de M es de la forma $a_1x_1 + \cdots + a_nx_n$, es decir, $M = \langle x_1 \rangle + \cdots + \langle x_n \rangle$, y por la definición de conjunto libre se cumple la condición 2) del teorema 4.15, luego la suma es directa.

Nota Es fácil ver que todo lo anterior vale igualmente para bases infinitas:

Si X es una base de un A-módulo M, entonces cada elemento $m \in M$ se expresa de forma única como suma

$$m = \sum_{x \in X} a_x x,$$

donde todos los a_x son nulos salvo a lo sumo una cantidad finita de ellos (entendiendo entonces que la suma en principio infinita es, por definición, la suma finita formada por los sumandos no nulos). Alternativamente, todo elemento $m \in M$ no nulo se expresa de forma única como combinación lineal de un número finito de elementos de X con coeficientes no nulos.

A su vez, un A–módulo M tiene una base X si y sólo si es isomorfo a la suma directa $M \cong \bigoplus_{x \in X} A$. La *base canónica* de la suma directa es la formada por los e_x dados por

$$e_x(y) = \begin{cases} 1 & \text{si } y = x, \\ 0 & \text{si } y \neq x. \end{cases}$$

Ejemplos En estos términos podemos decir que $\{1, i\}$ es una base de $\mathbb{C}$ como $\mathbb{R}$-espacio vectorial, así como que $\{1, \omega, \omega^2, \omega^3\}$ es una base de $\mathbb{Z}[\omega]$ como $\mathbb{Z}$-módulo, mientras que $\{1, \omega, \omega^2, \omega^3, \omega^4\}$ es un sistema generador, pero no una base, ya que tenemos una combinación lineal nula:

$$1 + \omega + \omega^2 + \omega^3 + \omega^4 = 0$$

con coeficientes no nulos.

Es fácil ver que una base de un anillo de polinomios $A[x]$ como A-módulo es la formada por todas las potencias de x:

$$1, \quad x, \quad x^2, \quad x^3, \quad x^4, \quad \dots$$

Es importante señalar que no todos los módulos son libres, aunque sí lo serán casi todos los que nos van a interesar. Por ejemplo, el anillo $\mathbb{Z}/n\mathbb{Z}$ es un $\mathbb{Z}$–módulo no libre, ya que si fuera libre debería ser isomorfo a una suma directa de varias veces $\mathbb{Z}$, lo cual es imposible, ya que tales sumas son infinitas y $\mathbb{Z}/n\mathbb{Z}$ es finito. ∎

Las bases son útiles a la hora de determinar los homomorfismos de un módulo en otro. Es inmediato que si dos homomorfismos de módulos coinciden sobre los elementos de un sistema generador, entonces son el mismo homomorfismo. Sobre las bases se puede decir más:

Teorema 4.21 *Sea A un anillo unitario, M y N dos A–módulos y X una base de M. Entonces cada aplicación $f : X \longrightarrow N$ se extiende a un único homomorfismo $f^* : M \longrightarrow N$.*

Demostración: Cada elemento no nulo de M se expresa de forma única como combinación lineal $a_1x_1 + \cdots + a_nx_n$ de elementos de X con coeficientes en A no nulos. La unicidad nos permite definir sin ambigüedad

$$f^*(a_1x_1 + \cdots + a_nx_n) = a_1f(x_1) + \cdots + a_nf(x_n)$$

y es fácil ver que la aplicación así definida (con la condición adicional $f^*(0) = 0$) es un homomorfismo. ∎

Los resultados que hemos demostrado hasta aquí son poco más que consecuencias inmediatas de las definiciones. Ahora vamos a demostrar resultados no triviales, y empezamos estudiando con más detalle los espacios vectoriales, es decir, los módulos sobre un anillo de división D, en los que se cumplen hechos que, o bien no valen para módulos arbitrarios, o bien las demostraciones son más complicadas. El teorema siguiente es un ejemplo del primer caso:

Teorema 4.22 *Si V es un D-espacio vectorial, entonces un conjunto $X \subset V$ es ligado si y sólo si uno de sus elementos es combinación lineal de los restantes.*

Demostración: Si $x = \alpha_1x_1 + \cdots + \alpha_nx_n$, con $x_i \in X \setminus \{x\}$, entonces $-x+\alpha_1x_1+\cdots+\alpha_nx_n = 0$ es una combinación lineal nula no trivial de elementos de X, luego X es ligado. Recíprocamente, si X es ligado, existen elementos $x_1, \dots, x_n \in X$ distintos dos a dos y escalares $\alpha_1, \dots, \alpha_n \in D$ no todos nulos tales que $\alpha_1x_1 + \cdots + \alpha_nx_n = 0$. Renumerando podemos suponer que $\alpha_1 \neq 0$, y entonces (y aquí usamos que D es un anillo de división)

$$x_1 = -\alpha_1^{-1}\alpha_2x_2 - \cdots - \alpha_1^{-1}\alpha_nx_n,$$

luego x_1 es combinación lineal de $x_2, \dots, x_n$. ∎

Nota Esto no es cierto para módulos arbitrarios. Por ejemplo, en $\mathbb{Z}$ como $\mathbb{Z}$-módulo tenemos que $X = \{2, 3\}$ es ligado, ya que $3 \cdot 2 + (-2) \cdot 3 = 0$, pero no es cierto que $2 = n3$ ni $3 = n2$ para ningún entero n, luego ninguno de los dos es combinación lineal del otro. ∎

Una consecuencia importante resulta de unir el teorema anterior a un sencillo hecho general: si x es combinación lineal de los elementos de X, entonces $\langle X\rangle = \langle X \cup \{x\}\rangle$. Esto es inmediato, porque $X \cup \{x\} \subset \langle X\rangle$, y esto implica $\langle X \cup \{x\}\rangle \subset \langle X\rangle$. La otra inclusión es cierta en general.

Teorema 4.23 *Si V es un D-espacio vectorial finitamente generado (es decir, si tiene un generador finito), entonces tiene una base. Más aún, todo sistema generador contiene una base.*

Demostración: Consideremos un generador finito de V, de modo que $V = \langle v_1, \ldots, v_n\rangle$. Si no es libre, por el teorema anterior uno de los generadores es combinación lineal de los demás, y por la observación precedente puede ser eliminado, con lo que pasamos a tener un sistema generador de V con $n - 1$ elementos. Tras un número finito de pasos tenemos que llegar a un generador libre (que puede ser $\varnothing$, si $V = 0$).

Con esto hemos probado que todo generador finito contiene una base. Si X es un sistema generador arbitrario de V, basta probar que contiene un generador finito. Ahora bien, si $v_1, \ldots, v_n$ es un generador finito de V, entonces cada v_i puede expresarse como combinación lineal de un cierto subconjunto finito $X_i \subset X$. El conjunto $X^* = \bigcup_{i=1}^{n} X_i \subset X$ es finito y es un sistema generador de V, pues $V = \langle v_1, \ldots, v_n\rangle \subset \langle X^*\rangle \subset V$. ∎

Y ahora podemos probar uno de los resultados fundamentales:

Teorema 4.24 *Si V es un D-espacio vectorial finitamente generado, entonces todas las bases de V tienen el mismo cardinal.*

Demostración: En la prueba del teorema anterior hemos visto que todo generador contiene una base finita, luego todas las bases de V tienen que ser finitas (una base no puede estar estrictamente contenida en otra base, ya que entonces la mayor no sería libre). Sean, pues $X = \{x_1, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_m\}$ dos bases de V. Podemos tomar $n \leq m$.

El elemento y_1 se expresa como combinación lineal de los elementos de X

$$y_1 = d_1x_1 + \cdots + d_nx_n,$$

y como es no nulo, alguno de los coeficientes será no nulo. Reordenando la base podemos suponer que $d_1 \neq 0$. Entonces podemos despejar

$$x_1 = d_1^{-1}y_1 - d_1^{-1}d_2x_2 - \cdots - d_1^{-1}d_nx_n,$$

luego resulta que $x_1 \in \langle y_1, x_2, \ldots x_n\rangle$. Obviamente $X \subset \langle y_1, x_2, \ldots x_n\rangle$ y así $\langle y_1, x_2, \ldots, x_n\rangle = V$.

Consecuentemente $y_2 = e_1y_1 + \cdots + e_nx_n$ y alguno de los coeficientes distintos de e_1 ha de ser no nulo (o si no $y_2 = e_1y_1$, luego $e_1y_1 - y_2 = 0$, con lo que Y sería ligado).

Reordenando la base podemos suponer que $e_2 \neq 0$ y repitiendo el argumento anterior concluimos que $\langle y_1, y_2, x_3, \ldots, x_n\rangle = V$.

De este modo llegamos finalmente a que $\langle y_1, \ldots, y_n\rangle = V$. De aquí se sigue que $m = n$, pues en otro caso existiría y_{n+1} y sería combinación lineal de $y_1, \ldots, y_n$, con lo que Y sería ligado. ∎

Antes hemos visto que si en un sistema generador un vector es combinación lineal del resto, podemos eliminarlo y seguimos teniendo un sistema generador. Otro hecho elemental es que si X es un sistema libre y v es un vector que no es combinación lineal de elementos de X, entonces $X \cup \{v\}$ sigue siendo libre.

En efecto, si fuera ligado, existiría una combinación lineal no trivial

$$\alpha_1x_1 + \cdots + \alpha_rx_r + \beta v = 0,$$

donde necesariamente $\beta \neq 0$, o de lo contrario X sería ligado, y esto nos permitiría despejar v (usando que β tiene inverso, de modo que este argumento requiere que D sea un anillo de división) para concluir que v sí que es combinación lineal de elementos de X, en contra de lo supuesto.

De aquí deducimos otra consecuencia importante sobre los espacios vectoriales:

Teorema 4.25 *Si V es un D-espacio vectorial finitamente generado, entonces todo conjunto libre X de vectores de V está contenido en una base.*

DEMOSTRACIÓN: Sea $v_1, \ldots, v_n$ una base de V. Si X no es generador, alguno de los v_i no es combinación lineal de los elementos de X, pues si todos lo fueran tendríamos que $V = \langle v_1, \ldots, v_n\rangle \subset \langle X\rangle \subset V$. Pongamos que es v_1. Entonces, por la observación precedente, $X \cup \{v_1\}$ es libre.

Si $X \cup \{v_1\}$ no es generador, entonces, como antes, algún v_i no es combinación lineal de sus elementos (y obviamente no será v_1, luego podemos suponer que es v_2) y por lo tanto $X \cup \{v_1, v_2\}$ es libre. Tras un número finito de pasos tenemos que llegar a que $X \cup \{v_1, \ldots, v_r\}$ es un generador libre (pues en el peor de los casos lo será cuando añadamos todos los vectores de la base), y así tenemos una base que contiene a X. ∎

Definición 4.26 Si V es un espacio vectorial sobre un anillo de división D, se llama *dimensión* de V, y la representaremos por $\dim_D V$, al número de elementos de cualquier base de V.

El teorema 4.24 implica que todo espacio vectorial finitamente generado tiene asociada una dimensión, que es un número natural (el espacio vectorial nulo tiene dimensión 0). Por ello, en lugar de hablar de espacios vectoriales finitamente generados, es más frecuente referirse a ellos como *espacios vectoriales de dimensión finita.*

En la sección 4.6 recogemos los resultados que prueban que el concepto de dimensión está bien definido también para espacios no finitamente generados (que se llaman, consecuentemente, *espacios vectoriales de dimensión infinita*), pero nunca vamos a necesitar esos resultados.

Que un espacio vectorial V tenga dimensión n significa que, fijada una base, cada elemento de V está determinado por n elementos de D. Por ejemplo, $\dim \mathbb{Q}[\omega] = 4$ porque cada elemento de $\mathbb{Q}[\omega]$ está determinado por cuatro coordenadas racionales.

Nos gustaría afirmar lo mismo para $\mathbb{Z}[\omega]$, pero primero tenemos que justificar que también en el caso de un $\mathbb{Z}$-módulo libre todas las bases tienen el mismo número de elementos. El teorema siguiente nos lo garantiza:

Teorema 4.27 (AE) *Si A es un anillo conmutativo y unitario y M es un A–módulo libre finitamente generado, entonces todas las bases de M tienen el mismo número de elementos.*[3]

Demostración: Sea I un ideal maximal de A (teorema 3.19). Es claro que

$$IM = \left\{ \sum_{i=1}^{n} a_i m_i \;\middle|\; n \in \mathbb{N}, a_i \in I, m_i \in M \right\}$$

es un submódulo de M. El A–módulo cociente M/IM se convierte en un A/I–módulo con el producto dado por $[a][m] = [am]$.

Esta definición es correcta, pues si $[a] = [a']$ y $[m] = [m']$, entonces

$$am - a'm' = am - am' + am' - a'm' = a(m - m') + (a - a')m' \in IM,$$

pues $m - m' \in IM$ y $a - a' \in I$.

Como I es un ideal maximal, el anillo cociente A/I es en realidad un cuerpo, luego M/IM es un espacio vectorial. Ahora basta probar que que toda A–base de M tiene el mismo cardinal que una A/I–base de M/IM, pues entonces M/IM será un espacio vectorial de dimensión finita y todas las bases de M tendrán como cardinal la dimensión de M/IM.

En efecto, si $X = \{x_1, \ldots, x_n\}$ es una A–base de M, vamos a demostrar que $X^* = \{[x_1], \ldots, [x_n]\}$ es una A/I–base de M/IM y que las n clases son distintas dos a dos, con lo que también tiene n elementos.

Todo elemento de M/IM es de la forma $[u]$, con $u \in M$, luego $u = \sum_{j=1}^{n} a_j x_j$, y entonces $[u] = \sum_{j=1}^{n} [a_j][x_j]$, lo que prueba que X^* genera M/IM.

Supongamos ahora que $\sum_{j=1}^{n} [a_j][x_j] = [0]$ para ciertos elementos a_j de A y vamos a probar que $[a_1] = \cdots [a_n] = 0$, lo cual demuestra a la vez que los $[x_j]$ son distintos dos a dos y que son linealmente independientes.

[3]Notemos que el axioma de elección sólo aparece en la prueba al aplicar el teorema 3.19, luego no es necesario si suponemos que el anillo A es noetheriano.

Tenemos que $\sum_{j=1}^{n} a_j x_j \in IM$, luego $\sum_{j=1}^{n} a_j x_j = \sum_{k=1}^{m} b_k m_k$, para ciertos elementos $m_k \in M$ y ciertos $b_k \in I$.

Como X es una base de M, cada m_k se expresa como $m_k = \sum_{j=1}^{n} c_{jk} x_j$.

Por lo tanto

$$\sum_{j=1}^{n} a_j x_j = \sum_{k=1}^{m} b_k m_k = \sum_{k=1}^{m} b_k \sum_{j=1}^{n} c_{jk} x_j = \sum_{j=1}^{n} \left(\sum_{k=1}^{m} b_k c_{jk} \right) x_j.$$

Pero como X es base, $a_j = \sum_{k=1}^{m} b_k c_{jk} \in I$, porque cada $b_k \in I$, luego $a_j \in I$, luego $[a_j] = 0$. Con esto tenemos que X^* es base de M/IM y tiene el mismo número de elementos que X. ∎

Definición 4.28 Si M es un módulo libre sobre un anillo conmutativo y unitario A, llamaremos *rango* de M ($\operatorname{rang} M$) al número de elementos de cualquier base de M.

Por supuesto, el que se use la palabra "rango" al hablar de módulos libres arbitrarios y de "dimensión" al hablar de espacios vectoriales no es más que una costumbre arraigada, pero en realidad el concepto es el mismo.

Como en el caso de la dimensión, hemos probado que el rango está definido para todo A-módulo libre finitamente generado —con A conmutativo y unitario—, por lo que es habitual referirse a ellos como A-módulos libres de rango finito. No obstante, la prueba del teorema anterior se adapta trivialmente al caso de módulos no necesariamente finitamente generados usando la generalización 4.56 de 4.24 a espacios vectoriales no necesariamente finitamente generados.

Ahora ya podemos afirmar que $\operatorname{rang} \mathbb{Z}[\omega] = 4$, y esto significa simplemente que cada entero ciclotómico está completamente determinado por 4 números enteros.

Observemos que si M y N son A-módulos libres, entonces

$$\operatorname{rang}(M \oplus N) = \operatorname{rang} M + \operatorname{rang} N.$$

En efecto, si los rangos son finitos tenemos que $M \cong A^m$, $N \cong A^n$, y entonces $M \oplus N \cong A^m \oplus A^n \cong A^{m+n}$, donde el último isomorfismo es el dado por

$$((a_1, \ldots, a_m), (b_1, \ldots, b_n)) \mapsto (a_1, \ldots, a_m, b_1, \ldots, b_n).$$

El argumento se generaliza fácilmente al caso en que alguno de los rangos es infinito, usando que si $I \cap J = \varnothing$, entonces

$$\bigoplus_{i \in I} A \oplus \bigoplus_{j \in J} A \cong \bigoplus_{i \in I \cup J} A.$$

La dimensión en espacios vectoriales se comporta mucho mejor que el rango en módulos libres en general. Veamos primero algunos resultados positivos sobre espacios vectoriales y después comentaremos la situación general.

Teorema 4.29 *Sea V un espacio vectorial de dimensión finita sobre un anillo de división D y sea W un subespacio de V. Entonces:*[4]

1. $\dim V = \dim W + \dim(V/W)$. *En particular* $\dim W \leq \dim V$.

2. *Si* $\dim W = \dim V$ *entonces* $W = V$.

Demostración: 1) Sea X una base de W. Por el teorema 4.25 sabemos que X se extiende a una base Y de V. Veamos que si $y_1, \ldots, y_n \in Y \setminus X$, entonces las clases $[y_1], \ldots, [y_n]$ son linealmente independientes (y en particular distintas) en V/W.

En efecto, si $a_1[y_1] + \cdots + a_n[y_n] = 0$ entonces $a_1y_1 + \cdots + a_ny_n \in W$, luego se expresa como combinación lineal de elementos de X, es decir,

$$a_1y_1 + \cdots + a_ny_n = b_1x_1 + \cdots + b_mx_m,$$

y esto nos da dos expresiones distintas de un mismo elemento de V como combinación lineal de elementos de la base Y, lo cual es imposible salvo que todos los coeficientes sean nulos.

Por otra parte, todo elemento de V/W es de la forma $[v]$, donde $v \in V$. El elemento v se expresa como combinación lineal de elementos de Y, digamos

$$v = a_1y_1 + \cdots + a_ny_n + b_1x_1 + \cdots + b_mx_m,$$

donde $y_1, \ldots, y_n \in Y \setminus X$ y $x_1, \ldots, x_m \in X$. Por lo tanto

$$[v] = a_1[y_1] + \cdots + a_n[y_n] + b_1[x_1] + \cdots + b_m[x_m] = a_1[y_1] + \cdots + a_n[y_n] + 0$$

y esto prueba que $\big\{[y] \mid y \in Y \setminus X\big\}$ es un generador, luego una base de V/W, que según lo visto tiene el mismo cardinal que $Y \setminus X$. Así pues

$$\dim V = |Y| = |X| + |Y - X| = \dim W + \dim(V/W).$$

2) Si $\dim W = \dim V = n$ finito, entonces una base de W (con n elementos) se ha de extender hasta una base de V, también con n elementos, luego toda base de W lo es también de V. Esto implica que $W = V$. ∎

Un razonamiento similar permite probar el resultado siguiente:

Teorema 4.30 *Sean V y W dos subespacios de dimensión finita de un espacio vectorial sobre un anillo de división D. Entonces existen conjuntos X e Y tales que X es base de V, Y es base de W, $X \cap Y$ es base de $V \cap W$ y $X \cup Y$ es base de $V + W$. En particular*

$$\dim(V + W) + \dim(V \cap W) = \dim V + \dim W.$$

[4] El lector familiarizado con la teoría de cardinales infinitos se dará cuenta de que la prueba de 1) vale igualmente aunque las dimensiones sean infinitas. No puede decirse lo mismo de 2.

La prueba consiste esencialmente en partir de una base de $V \cap W$ y extenderla hasta una base X de V y hasta una base Y de W. En realidad, la prueba vale también si las dimensiones no son finitas.

En particular la dimensión de una suma directa de subespacios es igual a la suma de las dimensiones de los subespacios.

Como consecuencia inmediata del teorema de isomorfía y del teorema 4.29 se cumple lo siguiente:

Teorema 4.31 *Sea $f : V \longrightarrow W$ una aplicación lineal entre espacios vectoriales sobre un anillo de división D. Entonces* $\dim V = \dim \mathrm{N}(f) + \dim \mathrm{Im}\, f$.

Otra propiedad sencilla en torno a las dimensiones es que si V es un espacio vectorial de dimensión finita n, entonces todo sistema libre con n elementos es una base, al igual que todo sistema generador con n elementos. En efecto, todo sistema libre con n elementos se extiende hasta una base, que ha de tener n elementos, luego tiene que ser ya una base. Igualmente, todo sistema generador con n elementos contiene una base con n elementos, luego él mismo es una base.

Casi todos estos resultados son falsos sobre módulos libres cualesquiera, incluso en los casos más simples. Por ejemplo, $2\mathbb{Z}$ es un submódulo de $\mathbb{Z}$ con el mismo rango finito, pero $2\mathbb{Z} \neq \mathbb{Z}$ (al contrario que 4.29). Por otro lado $\{2\}$ es un subconjunto libre de $\mathbb{Z}$ que no puede extenderse hasta una base de $\mathbb{Z}$ y el conjunto $\{2, 3\}$ es un generador de $\mathbb{Z}$ que no contiene una base. Por otra parte, no todo cociente de un módulo libre es libre (p.ej. $\mathbb{Z}/2\mathbb{Z}$).

4.4 Matrices

Una base finita en un A-módulo libre M no sólo permite determinar los elementos de M en términos de una n-tupla de coordenadas en A^n, sino que también permite reducir a un número finito de coordenadas los homomorfismos entre M y cualquier otro A-módulo libre que tenga también una base finita. Para mostrarlo necesitamos el concepto de matriz:

Definición 4.32 Sea A un anillo unitario y m, n números naturales no nulos. Una *matriz* $m \times n$ sobre A es una aplicación $B : \{1, \ldots, m\} \times \{1, \ldots, n\} \longrightarrow A$. Escribiremos b_{ij} en lugar de $B(i, j)$ y también $B = (b_{ij})$. En la práctica escribiremos los elementos de una matriz $m \times n$ dispuestos en m filas y n columnas así:

$$B = \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix}$$

Llamaremos $\mathrm{Mat}_{m\times n}(A)$ al conjunto de todas las matrices $m \times n$ sobre A. Las matrices $n \times n$ se llaman *matrices cuadradas.* Escribiremos $\mathrm{Mat}_n(A)$ en lugar de $\mathrm{Mat}_{n\times n}(A)$.

Evidentemente dos matrices $B = (b_{ij})$ y $C = (c_{ij})$ son iguales si y sólo si tienen las mismas dimensiones $m \times n$ y $b_{ij} = c_{ij}$ para todo par de índices i, j.

Podemos identificar los elementos de A^n con las matrices $1 \times n$, es decir, con las matrices con una sola fila y n columnas. A estas matrices se las llama *matrices fila.* Cuando A es un anillo de división se las llama también *vectores fila.*

Por analogía, las matrices $m \times 1$, es decir, las matrices que constan de una sola columna, se llaman *matrices columna* o *vectores columna* cuando A es un anillo de división.

En las matrices fila y columna suprimiremos el índice fijo, es decir, las representaremos así:

$$(a_1, \ldots, a_n), \qquad \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Llamaremos *matriz traspuesta* de una matriz $B \in \mathrm{Mat}_{m\times n}(A)$ a la matriz $B^t \in \mathrm{Mat}_{n\times m}(A)$ que resulta de intercambiar las filas de B por sus columnas, es decir, la componente (i, j) de B^t es la componente (j, i) de B.

De este modo, la traspuesta de una matriz fila es una matriz columna y viceversa. Claramente $B^{tt} = B$.

Una matriz cuadrada B es *simétrica* si $B = B^t$, es decir, si $b_{ij} = b_{ji}$ para todo par de índices i, j.

La *fila* i-ésima de una matriz B es la matriz fila $B_i = (b_{i1}, \ldots, b_{in})$. La *columna* j-ésima de la matriz B es la matriz columna

$$B^j = \begin{pmatrix} b_{1j} \\ \vdots \\ b_{mj} \end{pmatrix}$$

luego, en este sentido, una matriz $m \times n$ tiene m filas y n columnas.

Llamaremos *matriz nula* de orden $m \times n$ a la matriz $m \times n$ que tiene todas sus componentes iguales a 0.

Llamaremos *diagonal principal* de una matriz cuadrada $B \in \mathrm{Mat}_n(A)$ a la n-tupla $(b_{11}, \ldots, b_{nn})$.

Una matriz cuadrada es una *matriz diagonal* si tiene nulas todas sus componentes que no están en la diagonal principal.

Una matriz diagonal es una *matriz escalar* si tiene todas sus componentes de la diagonal principal iguales entre sí.

La *matriz identidad* de orden n es la matriz escalar $n \times n$ cuyas componentes de la diagonal principal son iguales a 1. La representaremos por I_n. Si definimos la *delta de Kronecker* mediante

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

entonces $I_n = (\delta_{ij})$.

Ahora definimos unas operaciones con matrices:

Si $B = (b_{ij})$ y $C = (c_{ij})$ son matrices $m \times n$, llamaremos $B + C$ a la matriz $m \times n$ dada por $B + C = (b_{ij} + c_{ij})$.

Si $B = (b_{ij})$ es una matriz $m \times n$ y $a \in A$, llamaremos aB a la matriz $m \times n$ dada por $aB = (ab_{ij})$.

Con estas operaciones $\mathrm{Mat}_{m\times n}(A)$ se convierte en un A-módulo libre de rango mn. Una base la forman las mn matrices que tienen un 1 en cada una de las posiciones posibles y las restantes componentes nulas.

La estructura de A-módulo en los espacios de matrices fila no es sino la estructura usual en los espacios A^n.

Finalmente definimos el siguiente producto de matrices:

Si $B \in \mathrm{Mat}_{m\times n}(A)$ y $C \in \mathrm{Mat}_{n\times r}(A)$, la matriz $BC \in \mathrm{Mat}_{m\times r}(A)$ es la que tiene en la posición (i, j) el elemento $\sum_{k=1}^{n} b_{ik}c_{kj}$.

Es pura rutina comprobar las propiedades siguientes (que se cumplen cuando las dimensiones de las matrices son las apropiadas):

1. $B(CD) = (BC)D$.
2. $B(C + D) = BC + BD$.
3. $(B + C)D = BD + CD$.
4. $BI_n = I_mB = B$.
5. Si A es conmutativo $(BC)^t = C^tB^t$.

En general, el producto de matrices no es una operación interna en el conjunto $\mathrm{Mat}_{m\times n}(A)$, pero sí lo es en los espacios de matrices cuadradas. Los espacios $\mathrm{Mat}_n(A)$ son anillos unitarios con la suma y el producto de matrices. Salvo en casos triviales no son conmutativos. La aplicación que a cada elemento $a \in A$ le asigna la matriz escalar aI_n es un monomorfismo de anillos, con lo que podemos identificar los elementos de A con las matrices escalares, y así A es un subanillo de $\mathrm{Mat}_n(A)$. El producto de una matriz por el elemento a coincide con el producto por la matriz aI_n.

Vamos a dar una interpretación de todo esto en términos de módulos.

Sea M un A-módulo libre de rango finito n. Una *base ordenada* de M es una n-tupla $B = (u_1, \ldots, u_n)$ tal que $u_1, \ldots, u_n$ forman una base de M.

Llamaremos *sistema de coordenadas* asociado a la base ordenada B a la aplicación $\Phi_B : M \longrightarrow A^n$ que a cada elemento $m \in M$ le asigna la n-tupla $(a_1, \ldots, a_n)$ tal que $m = a_1u_1 + \cdots + a_nu_n$. A $\Phi_B(m)$ se le llama n-tupla de *coordenadas* de m respecto a la base B.

Sea $f : M \longrightarrow N$ un homomorfismo entre módulos libres de rangos m y n respectivamente. Sean $B = (u_1, \ldots, u_m)$ y $B' = (v_1, \ldots, v_n)$ bases ordenadas de M y N. Entonces, para cada u_i existen unos únicos elementos $a_{ij} \in A$ tales que $f(u_i) = \sum_{j=1}^{n} a_{ij} v_j$.

Llamaremos *matriz asociada* a f en las bases B y B' a $\mathrm{M}_B^{B'}(f) = (a_{ij})$, es decir, a la matriz que tiene por filas a las coordenadas en la base B' de las imágenes de los miembros de la base B.

Teorema 4.33 *Sea $f : M \longrightarrow N$ un homomorfismo entre A-módulos libres de rangos m y n respectivamente. Sean B y B' bases ordenadas de M y N. Entonces $\mathrm{M}_B^{B'}(f)$ es la única matriz que cumple:*

$$\Phi_{B'}\big(f(u)\big) = \Phi_B(u)\mathrm{M}_B^{B'}(f),$$

para todo $u \in M$.

Demostración: Sean $B = (u_1, \ldots, u_m)$ y $B' = (v_1, \ldots, v_n)$, sea $\mathrm{M}_B^{B'}(f) = (a_{ij})$ y sea $\Phi_B(u) = (x_1, \ldots, x_m)$. Entonces $u = \sum_{i=1}^{m} x_i u_i$ y

$$f(u) = \sum_{i=1}^{m} x_i f(u_i) = \sum_{i=1}^{m} x_i \textstyle\sum_{j=1}^{n} a_{ij} v_j = \displaystyle\sum_{j=1}^{n} \left(\sum_{i=1}^{m} x_i a_{ij} \right) v_j,$$

luego

$$\Phi_{B'}\big(f(u)\big) = \left(\sum_{i=1}^{m} x_i a_{ij} \right) = \Phi_B(u)\mathrm{M}_B^{B'}(f).$$

Si una matriz C cumple $\Phi_{B'}\big(f(u)\big) = \Phi_B(u)C$, entonces tomando $u = u_i$ la n-tupla $\Phi_B(u)$ es la que tiene un 1 en el lugar i-ésimo y 0 en los restantes. El producto $\Phi_B(u)C$ no es sino la fila i-ésima de C, luego dicha fila i-ésima está formada por las coordenadas $\Phi_{B'}\big(f(u_i)\big)$, al igual que la fila i-ésima de $\mathrm{M}_B^{B'}(f)$. Por lo tanto $C = \mathrm{M}_B^{B'}(f)$. ∎

Definición 4.34 Si M y N son A-módulos, llamaremos $\mathrm{Hom}_A(M, N)$ al conjunto de todos los homomorfismos entre M y N. Si ambos son libres de rangos m y n, fijadas dos bases ordenadas B y B' de M y N respectivamente, tenemos definida una aplicación

$$\mathrm{M}_B^{B'} : \mathrm{Hom}_A(M, N) \longrightarrow \mathrm{Mat}_{m\times n}(A),$$

que claramente es biyectiva.

En efecto, si $\mathrm{M}_B^{B'}(f) = \mathrm{M}_B^{B'}(g)$, entonces por el teorema anterior para todo elemento u de M, se cumple $\Phi_{B'}\big(f(u)\big) = \Phi_{B'}\big(g(u)\big)$, luego ha de ser $f(u) = g(u)$ y por lo tanto $f = g$, la aplicación es inyectiva. Por otra parte, dada una matriz $C \in \mathrm{Mat}_{m\times n}(A)$, por el teorema 4.21 existe $f \in \mathrm{Hom}_A(M, N)$ que envía a cada componente de la base B al elemento de N que en la base B' tiene por n-tupla de coordenadas a la correspondiente fila de C, con lo que $\mathrm{M}_B^{B'}(f) = C$.

Ejemplo Consideremos la simetría $f : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ respecto del plano de ecuación $x + y + z = 0$. Hemos visto que $\mathbb{R}^3 = V_1 \oplus V_2$, donde V_1 es el subespacio formado por los vectores cuyas coordenadas suman 0 (el plano considerado) y $V_2 = \langle(1, 1, 1)\rangle$. Entonces, la simetría f es la aplicación que a cada $v = v_1 + v_2$, con $v_i \in V_i$, le asigna el vector $f(v) = v_1 - v_2$. Equivalentemente, si llamamos $\pi_i : \mathbb{R}^3 \longrightarrow V_i$ a las dos proyecciones, se cumple que $f(v) = \pi_1(v) - \pi_2(v)$, y esta expresión muestra que f es una aplicación lineal. Vamos a calcular su matriz respecto de la base canónica de $\mathbb{R}^3$.

Para ello descomponemos los vectores de la base canónica:

$$\begin{aligned} e_1 &= (1,0,0) = (2/3,-1/3,-1/3) + (1/3,1/3,1/3) \\ e_2 &= (0,1,0) = (-1/3,2/3,-1/3) + (1/3,1/3,1/3) \\ e_3 &= (0,0,1) = (-1/3,-1/3,2/3) + (1/3,1/3,1/3) \end{aligned}$$

y calculamos sus imágenes:

$$\begin{aligned} f(e_1) &= (2/3,-1/3,-1/3) - (1/3,1/3,1/3) = (1/3,-2/3,-2/3) \\ f(e_2) &= (-1/3,2/3,-1/3) - (1/3,1/3,1/3) = (-2/3,1/3,-2/3) \\ f(e_3) &= (-1/3,-1/3,2/3) - (1/3,1/3,1/3) = (-2/3,-2/3,1/3) \end{aligned}$$

con lo que la matriz buscada es

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

Teniendo en cuenta que las coordenadas de un vector respecto de la base canónica son sus componentes, esto nos dice que

$$f(x,y,z) = (x,y,z)\begin{pmatrix} 1/3 & -2/3 & -2/3 \\ -2/3 & 1/3 & -2/3 \\ -2/3 & -2/3 & 1/3 \end{pmatrix} =$$

$$\left(\frac{x}{3} - \frac{2y}{3} - \frac{2z}{3}, -\frac{2x}{3} + \frac{y}{3} - \frac{2z}{3}, -\frac{2x}{3} - \frac{2y}{3} + \frac{z}{3}\right).$$

■

Si A es un anillo conmutativo, el conjunto $\mathrm{Hom}_A(M, N)$ puede ser dotado de estructura de A-módulo de forma natural:

Definimos $f + g$ como el homomorfismo que sobre cada $m \in M$ actúa mediante $(f + g)(m) = f(m) + g(m)$, y si $a \in A$, entonces af es el homomorfismo determinado por $(af)(m) = a\big(f(m)\big)$ (notemos que si A no es conmutativo af no tiene por qué ser un homomorfismo).

Es fácil comprobar que la aplicación $\mathrm{M}_B^{B'}$ es un isomorfismo de módulos, es decir, que $\mathrm{M}_B^{B'}(f + g) = \mathrm{M}_B^{B'}(f) + \mathrm{M}_B^{B'}(g)$ y que $\mathrm{M}_B^{B'}(af) = a\mathrm{M}_B^{B'}(f)$.

Por ejemplo, si $u \in M$, entonces

$$\begin{aligned}\Phi_B(u)\big(\mathrm{M}_B^{B'}(f)+\mathrm{M}_B^{B'}(g)\big) &= \Phi_B(u)\mathrm{M}_B^{B'}(f)+\Phi_B(u)\mathrm{M}_B^{B'}(g)\\ &= \Phi_{B'}\big(f(u)\big)+\Phi_{B'}\big(g(u)\big)\\ &= \Phi_{B'}\big(f(u)+g(u)\big)\\ &= \Phi_{B'}\big((f+g)(u)\big),\end{aligned}$$

luego por la unicidad de 4.33, $\mathrm{M}_B^{B'}(f+g)=\mathrm{M}_B^{B'}(f)+\mathrm{M}_B^{B'}(g)$.

Esto explica las definiciones que hemos dado de suma de matrices y producto de una matriz por un elemento de A: la suma de dos matrices es la operación que nos da la matriz asociada al homomorfismo suma de los homomorfismos asociados a los sumandos, y similarmente con el producto por elementos de A. Respecto al producto de matrices, su interpretación es la siguiente:

Teorema 4.35 *Sean $f : M \longrightarrow N$ y $g : N \longrightarrow R$ homomorfismos de A-módulos libres de rango finito. Sean B, B' y B'' bases ordenadas de M, N y R respectivamente. Entonces* $\mathrm{M}_B^{B''}(f\circ g)=\mathrm{M}_B^{B'}(f)\mathrm{M}_{B'}^{B''}(g)$.

Demostración: Si $u \in M$, entonces

$$\begin{aligned}\Phi_B(u)\mathrm{M}_B^{B'}(f)\mathrm{M}_{B'}^{B''}(g) &= \Phi_{B'}\big(f(u)\big)\mathrm{M}_{B'}^{B''}(g)=\Phi_{B''}\Big(g\big(f(u)\big)\Big)\\ &= \Phi_{B''}\big((f\circ g)(u)\big),\end{aligned}$$

luego por la unicidad de 4.33, $\mathrm{M}_B^{B''}(f\circ g)=\mathrm{M}_B^{B'}(f)\mathrm{M}_{B'}^{B''}(g)$. ∎

Definición 4.36 Si M es un A-módulo, representaremos por $\mathrm{End}_A(M)=\mathrm{Hom}_A(M,M)$ al conjunto de todos los endomorfismos de M. Es fácil ver que $\mathrm{End}_A(M)$ es un anillo con la suma de homomorfismos que ya tenemos definida y con la composición de aplicaciones como producto.

Si M es libre de rango n y B es una base ordenada, el teorema anterior prueba que la aplicación $M_B^B : \mathrm{End}_A(M) \longrightarrow \mathrm{Mat}_n(A)$ es un isomorfismo de anillos. Notemos que la matriz identidad se corresponde con la aplicación identidad. Se dice que $M_B(f)=M_B^B(f)$ es la *matriz* de f en la base B.

El grupo $\mathrm{Aut}_A(M)$ de las unidades de $\mathrm{End}_A(M)$ está formado por los automorfismos de M. Si M es libre de rango finito, la aplicación M_B^B hace corresponder los automorfismos de M con las matrices regulares en el sentido siguiente:

Definición 4.37 Una matriz $C \in \mathrm{Mat}_n(A)$ es *regular* si es una unidad del anillo $\mathrm{Mat}_n(A)$, es decir, si existe una matriz $C^{-1} \in \mathrm{Mat}_n(A)$ tal que $CC^{-1}=C^{-1}C=I_n$. En tal caso la matriz C^{-1} es única y se llama *matriz inversa* de C. Una matriz cuadrada que no es regular es una *matriz singular*.

Una propiedad elemental es que si A es conmutativo y B es una matriz regular, entonces la matriz traspuesta B^t también es regular y $(B^t)^{-1} = (B^{-1})^t$. En efecto, basta observar que $(B^{-1})^t B^t = (BB^{-1})^t = I_n^t = I_n$, e igualmente en orden inverso.

El conjunto $\mathrm{LG}(n, A)$ de matrices regulares con coeficientes en A es claramente un grupo con el producto de matrices, y recibe el nombre de *grupo lineal general* de A. Según hemos observado, si M es un A-módulo libre y B es una base de M, la aplicación $\mathrm{Aut}_A(M) \longrightarrow \mathrm{LG}(n, A)$ dada por $f \mapsto \mathrm{M}_B(f)$ es un isomorfismo de grupos. Un poco más en general:

Teorema 4.38 *Si $f : M \longrightarrow N$ es un homomorfismo entre módulos libres del mismo rango finito y B, B' son bases ordenadas de M y N respectivamente, entonces f es un isomorfismo si y sólo si $\mathrm{M}_B^{B'}(f)$ es regular y, en tal caso, $\mathrm{M}_{B'}^{B}(f^{-1}) = \mathrm{M}_B^{B'}(f)^{-1}$.*

Demostración: Sea $g \in \mathrm{Hom}_A(N, M)$ tal que $g = f^{-1}$ si suponemos que f es isomorfismo o tal que $\mathrm{M}_{B'}^{B}(g) = \mathrm{M}_B^{B'}(f)^{-1}$ si suponemos que $\mathrm{M}_B^{B'}(f)$ es regular.

En cualquier caso se cumple que $\mathrm{M}_B^{B'}(f)\mathrm{M}_{B'}^{B}(g) = \mathrm{M}_B^B(f \circ g) = I_n = \mathrm{M}_B^B(I)$ y $\mathrm{M}_{B'}^{B}(g)\mathrm{M}_B^{B'}(f) = \mathrm{M}_{B'}^{B'}(g \circ f) = I_n = \mathrm{M}_{B'}^{B'}(I)$, de donde se siguen las dos implicaciones. ∎

Conviene observar lo siguiente:

Teorema 4.39 *El centro del grupo $\mathrm{LG}(n, A)$ está formado por las matrices de la forma uI_n, donde u es una unidad de A.*

Demostración: El enunciado equivale a probar que si M es un A-módulo libre de rango n, el centro del grupo $\mathrm{Aut}_A(M)$ está formado por los automorfismos de la forma $f_u(m) = um$, donde u es una unidad de A, pues éstos son los automorfismos con matriz de la forma indicada uI_n.

Es inmediato que las matrices de la forma uI_n conmutan con todas las matrices, luego están en el centro de $\mathrm{LG}(n, A)$, y por consiguiente los automorfismos f_u están en el centro de $\mathrm{Aut}_A(M)$. Hay que probar el recíproco.

Sea $e_1, \ldots, e_n$ una base de M y, fijado $i > 1$, consideremos el automorfismo de M dado por

$$g_i(e_1) = e_1 + e_i, \qquad g_i(e_j) = e_j, \quad \text{para } j > 1.$$

Si $f \in \mathrm{Aut}_A(M)$ está en el centro del grupo, tiene que cumplir que $f \circ g_i = g_i \circ f$. Pongamos que $f(e_1) = a_1e_1 + \cdots + a_ne_n$. Entonces

$$f(g_i(e_1)) = f(e_1 + e_i) = f(e_1) + f(e_i),$$

mientras que

$$g_i(f(e_1)) = g_i(a_1e_1 + \cdots + a_ne_n) = a_1e_1 + \cdots + a_ne_n + a_1e_i = f(e_1) + a_1e_i.$$

Por lo tanto, tiene que ser $f(e_i) = a_1e_i$, para todo $i > 1$. Pero, partiendo de que $f(e_n) = a_1e_n$, podemos probar que $f(e_i) = a_1e_i$ para todo $i < n$, luego en particular se cumple que $f(e_1) = a_1e_1$, y así $f(e_i) = a_1e_i$ para todo i.

Notemos que el argumento anterior supone que $n > 1$, pero si $n = 1$ la conclusión es trivial. Ahora bien, f^{-1} también tiene que estar en el centro de $\mathrm{Aut}(M)$ (porque el centro de un grupo es un subgrupo), luego tiene que ser $f^{-1}(e_i) = ve_i$ para todo i, luego $e_i = f(f^{-1}(e_i)) = a_1 ve_i$, luego $a_1 v = 1$, luego a_1 es una unidad de A, luego, llamando $u = a_1$, tenemos que $f = f_u$. ∎

En particular, si K es un cuerpo, el centro del grupo $\mathrm{LG}(n, K)$ está formado por las matrices diagonales aI_n, donde $a \in K$ es no nulo.

Las matrices también proporcionan una forma sencilla de obtener las coordenadas de un elemento de un módulo libre en una base conocidas sus coordenadas en otra base:

Definición 4.40 Si $B = (u_1, \ldots, u_n)$ y $B' = (v_1, \ldots v_n)$ son dos bases ordenadas de un mismo A-módulo M, se llama *matriz de cambio de base* a la matriz $\mathrm{M}_B^{B'} = \mathrm{M}_B^{B'}(I)$, donde I es la identidad en M.

Claramente $\mathrm{M}_B^{B'}$ es regular y $(\mathrm{M}_B^{B'})^{-1} = \mathrm{M}_{B'}^{B}$. La fila i-ésima de $\mathrm{M}_B^{B'}$ es $\Phi_{B'}(u_i)$ y para todo $m \in M$ se cumple la relación

$$\Phi_{B'}(m) = \Phi_B(m)\mathrm{M}_B^{B'},$$

es decir, el producto por $\mathrm{M}_B^{B'}$ transforma las coordenadas de m en B en las coordenadas de m en B'.

Ejemplo Sea K un cuerpo (de característica 0) y consideremos el sistema de vectores

$$B = \{(1, -1, 0), (1, 0, -1), (1, 1, 1)\} \subset K^3.$$

Vamos a probar que son un sistema generador de K^3. Para ello vemos cómo expresar los vectores de la base canónica como combinación lineal de éstos:

$$(1, 0, 0) = x(1, -1, 0) + y(1, 0, -1) + z(1, 1, 1),$$

lo que equivale al sistema de ecuaciones

$$\begin{aligned} x + y + z &= 1 \\ -x + z &= 0 \\ -y + z &= 0 \end{aligned}$$

del que se desprende que $x = y = z$ y, sustituyendo en la primera ecuación, $x = 1/3$, luego la solución es

$$e_1 = (1, 0, 0) = \frac{1}{3}(1, -1, 0) + \frac{1}{3}(1, 0, -1) + \frac{1}{3}(1, 1, 1).$$

Similarmente obtenemos que

$$e_2 = (0, 1, 0) = -\frac{2}{3}(1, -1, 0) + \frac{1}{3}(1, 0, -1) + \frac{1}{3}(1, 1, 1).$$

$$e_3 = (0, 0, 1) = \frac{1}{3}(1, -1, 0) - \frac{2}{3}(1, 0, -1) + \frac{1}{3}(1, 1, 1).$$

Esto implica que

$$e_1, e_2, e_3 \in \langle B \rangle\,,$$

luego $K^3 = \langle e_1, e_2, e_3\rangle \subset \langle B\rangle \subset K^3$, luego $K^3 = \langle B\rangle$. Por lo tanto, B es un sistema generador de K^3, y por el teorema 4.23 debe contener una base de K^3, pero $\dim K^3 = 3$, luego dicha base tiene que ser el propio B. La matriz

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

tiene por filas las coordenadas de los vectores de la base canónica en la base B, luego es la matriz $M_{B'}^B$, donde B' es la base canónica, es decir, es la matriz que nos da las coordenadas de un vector en la base B a partir de sus coordenadas en la base canónica.

Por ejemplo, el vector $v = (1, 1, 0)$ tiene coordenadas

$$\Phi_B(v) = (1,1,0)\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ -2/3 & 1/3 & 1/3 \\ 1/3 & -2/3 & 1/3 \end{pmatrix} = (-1/3, 2/3, 2/3).$$

Esto significa que

$$(1,1,0) = -\frac{1}{3}(1,-1,0) + \frac{2}{3}(1,0,-1) + \frac{2}{3}(1,1,1).$$

■

4.5 Módulos finitamente generados sobre DIPs

Podemos resumir los resultados que hemos demostrado sobre espacios vectoriales de dimensión finita diciendo que un espacio vectorial de dimensión finita sobre un anillo de división D está completamente determinado por un número natural, su dimensión d, en el sentido de que todos los D–espacios vectoriales de dimensión d son isomorfos a D^d y, desde un punto de vista algebraico, como espacios vectoriales, podemos decir que todos son "el mismo espacio", aunque difieran como conjuntos. Podemos expresar esto diciendo que los espacios vectoriales de dimensión finita quedan completamente "clasificados" por su dimensión, en el sentido de que, conocida la dimensión (y el anillo D), conocido el espacio.

En esta sección llegaremos a resultados similares para módulos finitamente generados sobre un dominio de ideales principales D. Sabemos que lo anterior es cierto igualmente para los D–módulos libres de rango finito, pero también sabemos que puede haber D-módulos finitamente generados que no sean libres.

Vamos a probar que también es posible asignar a cada D-módulo finitamente generado un número finito de "datos" que lo determinarán salvo isomorfismo, aunque ahora los "datos" no se reducirán a un mero número natural, como en el caso de los espacios vectoriales. El resultado será una clasificación completa de

los D-módulos finitamente generados en el mismo sentido que en el caso de los espacios vectoriales: cada módulo tendrá asignados unos "datos" de modo que, conocidos los datos, conocido el módulo y, aun sin conocer los datos que determinan a un módulo dado, podremos enumerar en la práctica todas las posibilidades que pueden darse, de modo que podamos analizar cada una separadamente si es necesario sabiendo que con ello cubrimos todos los casos posibles.

El teorema de clasificación que vamos a probar aquí es un resultado mucho más abstracto que cualquier otro que hayamos probado hasta el momento, y tal vez el lector no esté en condiciones de apreciar el valor de trabajar con tanta generalidad. Sin embargo, veremos que el nivel de abstracción de esta sección se traduce en que los resultados que vamos a obtener serán aplicables en contextos muy diversos, como por ejemplo la clasificación de los grupos abelianos finitamente generados o la clasificación las isometrías de un espacio euclídeo.

Nuestro punto de partida será el teorema siguiente:

Teorema 4.41 (AE) *Todo submódulo de un módulo libre sobre un dominio de ideales principales es libre de rango menor o igual que el del módulo.*[5]

Demostración: Sea L un módulo libre sobre un dominio de ideales principales D. Sea $B \subset L$ una base y consideremos en ella un buen orden $\leq$. Para cada $b \in B$ definimos

$$L_b = \langle c \in B \mid c < b \rangle\,, \qquad \bar{L}_b = \langle c \in B \mid c \leq b \rangle\,.$$

Cada $a \in \bar{L}_b$ se expresa de forma única como $a = u + db$, con $u \in L_b$ y $d \in D$. La aplicación $f_b : \bar{L}_b \longrightarrow D$ dada por $a \mapsto d$ es claramente un homomorfismo de módulos.

Tomemos ahora un submódulo $M \subset L$ y vamos a considerar los homomorfismos f_b restringidos a $f_b : M \cap \bar{L}_b \longrightarrow D$. Así, el núcleo de f_b es claramente $M \cap L_b$. La imagen de f_b será un ideal de D. Como D es un dominio de ideales principales, estará generada por un cierto $d_b \in D$. Sea $B' = \{b \in B \mid d_b \neq 0\}$ y, para cada $b \in B'$ elegimos un $m_b \in M \cap \bar{L}_b$ tal que $f_b(m_b) = d_b$.

El teorema quedará probado si demostramos que $C = \{m_b \mid b \in B'\}$ es una base de M (pues, ciertamente, su cardinal es menor o igual[6] que el de B).

En primer lugar demostramos que C es linealmente independiente. Supongamos, para ello, que tenemos una combinación lineal nula $a_1 m_{b_1} + \cdots + a_n m_{b_n} = 0$, donde $a_i \in D$ y $b_1 < \cdots < b_n$ son elementos de B. En esta situación, para cada $i < n$, tenemos que $m_{b_i} \in M \cap \bar{L}_{b_i} \subset M \cap L_{b_n}$, luego

$$a_1 m_{b_1} + \cdots + a_{n-1} m_{b_{n-1}} \in M \cap L_{b_n}\,.$$

[5]El teorema usa AE únicamente al suponer que el módulo dado tiene una base que admite un buen orden, pero esto se cumple trivialmente para módulos de rango numerable (en particular, finito). En tal caso no es necesario AE.

[6]En el caso en que la base B sea infinita, el lector debe saber que el cardinal de un subconjunto es menor o igual que el cardinal del conjunto. Nosotros hemos visto que esto es así cuando B es finito o incluso infinito numerable, y nunca vamos a necesitar otros casos. De hecho, en esta sección sólo necesitaremos el caso finito.

Por lo tanto,

$$0 = f_{b_n}(0) = f_{b_n}(a_1 m_{b_1} + \cdots + a_{n-1} m_{b_{n-1}}) + f_{b_n}(a_n m_{b_n}) = a_n d_{b_n},$$

y concluimos que $a_n = 0$. Aplicando ahora $f_{b_{n-1}}$ se obtiene que $a_{n-1} = 0$, e igualmente con todos los coeficientes.

Veamos ahora que C es un sistema generador de M. Por reducción al absurdo, supongamos que existe un $m \in M$ que no puede expresarse como combinación lineal de elementos de C. Entonces $m \in M \cap \bar{L}_b$ para cierto $b \in B$, y podemos tomar el mínimo b tal que existe un m en estas condiciones.

Si $b \notin B'$, entonces la imagen de f_b es nula, luego $f_b(m) = 0$, lo que significa que $m \in M \cap L_b$, pero entonces existirá un $b' < b$ tal que $m \in M \cap \bar{L}_{b'}$, en contradicción con la minimalidad de b. Concluimos, pues, que $b \in B'$, y entonces $f_b(m) = dd_b$, para cierto $d \in D$.

Llamemos $m' = m - dm_b \in M \cap \bar{L}_b$. Claramente $f_b(m') = dd_b - dd_b = 0$. Consecuentemente, $m' \in M \cap L_b$, luego, existe un $b' < b$ tal que $m' \in M \cap \bar{L}_{b'}$ y, por la minimalidad de b, tenemos que m' es combinación lineal de elementos de C, pero entonces m también lo es, y llegamos a una contradicción. ∎

Un poco más en general, todo submódulo de un módulo finitamente generado es finitamente generado:

Teorema 4.42 *Sea D un dominio de ideales principales y M un D–módulo con un generador finito de n elementos. Entonces todo submódulo de M admite un generador finito con a lo sumo n elementos.*

Demostración: Sea $\{x_1, \ldots, x_n\}$ un generador de M, sea L un D–módulo libre de rango n y sea $\{y_1, \ldots, y_n\}$ una base de L. Entonces por el teorema 4.21 existe un homomorfismo $f : L \longrightarrow M$ tal que $f(y_i) = x_i$ para cada $i = 1, \ldots, n$.

Como $\mathrm{Im} f$ es un submódulo que contiene a un generador de M, necesariamente ha de ser $\mathrm{Im} f = M$, luego f es suprayectiva.

Ahora, si N es un submódulo de M, se cumple que $N = f\left[f^{-1}[N]\right]$, y por 4.41 tenemos que $f^{-1}[N]$ es un submódulo de L libre y de rango menor o igual que n. La imagen de una base de $f^{-1}[N]$ es claramente un sistema generador de N. ∎

Ya hemos señalado que no todo $\mathbb{Z}$-módulo es libre (por ejemplo, $\mathbb{Z}/n\mathbb{Z}$). Sin embargo, ahora podemos dar una condición sencilla que determina cuándo un módulo finitamente generado sobre un DIP es libre. Para ello introducimos algunos conceptos:

Definición 4.43 Si A es un dominio y M es un A–módulo, un elemento $m \in M$ es *de torsión* si existe un $a \in A$ no nulo tal que $am = 0$. Llamaremos M_t al conjunto de todos los elementos de torsión de M. Es inmediato comprobar que M_t es un submódulo de M, que recibe el nombre de *submódulo de torsión* de M. Si $M = M_t$ se dice que M es un *módulo de torsión.* Si $M_t = 0$ se dice que M es *libre de torsión.*

Por ejemplo, $\mathbb{Z}/n\mathbb{Z}$ es un $\mathbb{Z}$–módulo de torsión, pues $nx = 0$ para todo $x \in \mathbb{Z}/n\mathbb{Z}$.

En general, si D es un dominio íntegro y M es un D–módulo libre, entonces es libre de torsión, pues si $m \in M$ es un elemento de torsión, podemos expresarlo como combinación lineal de los elementos de una base:

$$m = d_1 b_1 + \cdots + d_n b_n,$$

y entonces, si existe un $d \in D$ no nulo tal que $dm = 0$, tenemos que

$$0 = dm = dd_1 b_1 + \cdots + dd_n b_n,$$

luego, por la independencia lineal de la base, $dd_1 = \cdots dd_n = 0$, y como D es íntegro, esto implica que $d_1 = \cdots = d_n = 0$, luego $m = 0$.

En el contexto en el que estamos trabajando, se cumple el recíproco:

Teorema 4.44 *Si D es un dominio de ideales principales, entonces un D-módulo finitamente generado es libre si y sólo si es libre de torsión.*

Demostración: Sea M un D–módulo finitamente generado (que podemos suponer no nulo). Acabamos de ver que si es libre también es libre de torsión. Ahora supongamos que es libre de torsión. Esto equivale a que cualquier $y \in M$ no nulo es libre. Sea G un sistema generador finito de M y sea y_1 uno cualquiera de sus elementos no nulos. En general, podemos ir eligiendo elementos $y_1, \ldots, y_n \in G$ tales que $\{y_1, \ldots, y_n\}$ sea un sistema libre hasta agotar los elementos de G o bien hasta que cualquier $y \in G \setminus \{y_1, \ldots, y_n\}$ haga que $\{y_1, \ldots, y_n, y\}$ sea ligado. Pongamos que $G = \{y_1, \ldots, y_n, y_{n+1}, \ldots y_k\}$, donde tal vez $k = n$ y sea $L = \langle y_1, \ldots, y_n \rangle$, que es un D–módulo libre finitamente generado.

Si $k > n$, para cada $n < i \leq k$, tenemos una combinación lineal nula

$$d_1 y_1 + \cdots + d_n y_n + d_i y_i = 0,$$

donde no todos los coeficientes son nulos, pero, concretamente, tiene que ser $d_i \neq 0$, pues de lo contrario $\{y_1, \ldots, y_n\}$ sería ligado. Entonces $d_i y_i \in L$. Sea $d = d_{n+1} \cdots d_k \neq 0$, de modo que $dy_i \in L$ para todo $i = 1, \ldots, k$. El hecho de que G es un generador de M implica claramente que $dm \in L$ para todo $m \in M$. Consideremos ahora la aplicación $f : M \longrightarrow L$ dada por $f(m) = dm$. Claramente es un homomorfismo de módulos, y es un monomorfismo (tiene núcleo trivial) porque M es libre de torsión. Por lo tanto M es isomorfo a su imagen, que es un submódulo de L, luego es libre por el teorema 4.41. ∎

Ejercicio: Probar que $\mathbb{Q}$ es un $\mathbb{Z}$–módulo libre de torsión, pero no libre.

Podemos precisar más la situación:

Teorema 4.45 *Sea D un dominio de ideales principales y sea M un D–módulo finitamente generado. Entonces M/M_t es un D–módulo libre y existe un submódulo libre L de M tal que $M = M_t \oplus L$. En particular $L \cong M/M_t$, luego su rango está determinado por M.*

DEMOSTRACIÓN: Observemos en primer lugar que M/M_t es libre de torsión. En efecto, si $[x] \in M/M_t$ es un elemento de torsión, esto significa que existe un $d \in D$ no nulo tal que $[dx] = 0$, es decir, $dx \in M_t$, luego existe un $d' \in D$ no nulo tal que $d'dx = 0$, luego $x \in M_t$, luego $[x] = 0$. También es claro que M/M_t es finitamente generado. Por el teorema anterior tenemos que M/M_t es libre.

Sea $[m_1], \ldots, [m_n]$ una base de M/M_t y sea $L = \langle m_1, \ldots, m_n \rangle$. Se cumple que $m_1, \ldots, m_n$ son linealmente independientes, pues si

$$d_1 m_1 + \cdots + d_n m_n = 0,$$

entonces $d_1[m_1] + \cdots + d_n[m_n] = 0$, luego $d_1 = \cdots = d_n = 0$, por la independencia lineal de las clases $[m_i]$.

Por otra parte, $M = M_t + L$, ya que si $m \in M$, entonces existen $d_i \in D$ tales que $[m] = d_1[m_1] + \cdots + d_n[m_n]$, luego $x = d_1 m_1 + \cdots d_n m_n \in L$ cumple $[m] = [x]$, luego $m - x = t \in M_t$, luego $m = t + x \in M_t + L$.

Finalmente, $M = M_t \oplus L$, pues esto equivale a que $M_t \cap L = 0$ y, en efecto, si $m \in M_t \cap L$, entonces $m = d_1 m_1 + \cdots + d_n m_n$ para ciertos $d_i \in D$, pero entonces $d_1[m_1] + \cdots + d_n[m_n] = [m] = 0$, luego $d_1 = \cdots = d_n = 0$, luego $m = 0$.

Notemos que la proyección $M = M_t \oplus L \longrightarrow L$ es un epimorfismo de módulos con núcleo M_t, lo que nos da el isomorfismo $L \cong M/M_t$. ∎

Definición 4.46 Si M es un módulo finitamente generado sobre un DIP D, llamaremos *rango* de M al rango del módulo libre M/M_t.

Con esto tenemos ya uno de los "datos" que nos permitirán clasificar los módulos finitamente generados sobre un DIP. Dos módulos tienen el mismo rango si y sólo si tienen partes libres isomorfas. Ahora vamos a encontrar otros "datos" que determinen todas las posibilidades para la parte de torsión.

Si M es un D-módulo arbitrario y $m \in M$, llamaremos *orden* de m al conjunto

$$o(m) = \{d \in D \mid dm = 0\}.$$

Claramente es un ideal de D, que será no nulo si y sólo si m es un elemento de torsión. En tal caso, si $o(m) = (a)$, entonces tenemos que $a \neq 0$ y un $d \in D$ anula a m (es decir, cumple $dm = 0$) si y sólo si $a \mid d$. Diremos que a es un *periodo* de m.

Notemos que los generadores de un ideal principal están determinados salvo unidades, por lo que cada $m \in M$ determina sus periodos salvo unidades. En los casos en que tengamos un criterio para seleccionar un generador de cada ideal no distinguiremos entre orden y periodos. Por ejemplo, en el caso de un $\mathbb{Z}$-módulo M podemos considerar que el orden de un elemento de torsión m es su único periodo positivo, y es el menor natural a tal que $am = 0$. Notemos que se trata del orden de m cuando consideramos a M como grupo abeliano.

Si $a \in D$, definimos $M[a] = \{m \in M \mid am = 0\}$, que claramente es un submódulo de M.

Definimos el *anulador* de M como $\mathrm{An}(M) = \{d \in D \mid M[d] = M\}$, es decir, el conjunto de los elementos de D que anulan a todos los elementos de M. Es claro que también es un ideal de D, y si M es un módulo de torsión finitamente generado entonces es no nulo, pues basta tomar un generador finito de M y multiplicar un periodo de cada uno de sus miembros para obtener un elemento no nulo de $\mathrm{An}(M)$.

Si $\mathrm{An}(M) = (e)$ se dice que e es un *exponente* de M. Así, cada D–módulo de torsión tiene un exponente no nulo determinado salvo unidades. En el caso de $\mathbb{Z}$ podemos considerar que el exponente de un $\mathbb{Z}$-módulo de torsión es el único exponente positivo.

Por ejemplo, un exponente de D/eD es e.

Es claro que el orden de todo elemento de un módulo de torsión finitamente generado tiene que dividir a su exponente (que es lo que habíamos observado en el caso de $\mathbb{Z}/6\mathbb{Z}$).

Ejemplos Los $\mathbb{Z}$-módulos de torsión más sencillos que podemos considerar son los módulos $\mathbb{Z}/n\mathbb{Z}$, pero no son los únicos, pues, por ejemplo, $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ es un $\mathbb{Z}$-módulo de torsión con cuatro elementos y no es el mismo que (no es isomorfo a) $\mathbb{Z}/4\mathbb{Z}$. Para comprobar que no son isomorfos basta calcular los órdenes de sus elementos:

m	0	1	2	3
o	1	4	2	4

m	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
o	1	2	2	2

Si los módulos fueran isomorfos, tendrían que tener los mismos elementos de los mismos órdenes. ∎

Similarmente, podríamos pensar que $\mathbb{Z}/6\mathbb{Z}$ y $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$ son dos ejemplos de $\mathbb{Z}$-módulos con seis elementos no isomorfos entre sí, pero no es cierto, pues el teorema chino del resto 3.53 afirma que son isomorfos. Vamos a demostrar que los únicos D-módulos de torsión finitamente generados son, salvo isomorfismo, las sumas directas de módulos de tipo D/aD, y que son todas distintas entre sí salvo las que resultan isomorfas en virtud del teorema anterior.

Por ejemplo, los únicos $\mathbb{Z}$-módulos con 12 elementos son

$$\mathbb{Z}/4\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}, \qquad \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z}.$$

No incluimos $\mathbb{Z}/12\mathbb{Z}$ porque es isomorfo al primero, ni $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/6\mathbb{Z}$ porque es isomorfo al segundo. Observemos que el primero tiene exponente 12 y el segundo 6.

En general, el teorema chino del resto implica que, a la hora de cubrir todos los posibles $\mathbb{Z}$-módulos de torsión, sólo hemos de considerar sumandos de cardinal potencia de primo, pues si n no es potencia de primo, entonces $\mathbb{Z}/n\mathbb{Z}$ se descompone en sumandos de tipo $\mathbb{Z}/(p^r)$.

Ejemplo Dado el $\mathbb{Z}$-módulo de torsión

$$M = \mathbb{Z}/6\mathbb{Z} \oplus \mathbb{Z}/18\mathbb{Z} \oplus \mathbb{Z}/10\mathbb{Z} \oplus \mathbb{Z}/25\mathbb{Z} \oplus \mathbb{Z}/15\mathbb{Z},$$

aplicando el teorema anterior vemos que es isomorfo a

$$M \equiv \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z} \oplus \mathbb{Z}/3\mathbb{Z} \oplus \mathbb{Z}/9\mathbb{Z} \oplus \mathbb{Z}/5\mathbb{Z} \oplus \mathbb{Z}/5\mathbb{Z} \oplus \mathbb{Z}/25\mathbb{Z},$$

donde simplemente hemos descompuesto el orden de cada sumando en potencias de primo. A los números $2, 2, 2, 3, 3, 9, 5, 5, 25$ los llamaremos *divisores elementales* de M, y vamos a demostrar que están determinados y determinan a M salvo isomorfismo, en el sentido de que, por una parte, no es posible obtener otra descomposición de M que dé lugar a un juego distinto de divisores elementales y, por otra parte, dos $\mathbb{Z}$-módulos de torsión con juegos diferentes de divisores elementales no pueden ser isomorfos entre sí. ∎

Observemos que, en general, un D-módulo M es isomorfo a un $D/(a)$ si y sólo si $M = \langle x \rangle$ con $o(x) = (a)$, pues en tal caso la aplicación $D \longrightarrow M$ dada por $d \mapsto dx$ es un epimorfismo de módulos de núcleo (a), luego ciertamente $M \cong D/(a)$, y el recíproco es inmediato, ya que $D/(a) = \langle [1] \rangle$ y $o([1]) = (a)$.

El teorema siguiente muestra que a todo D-módulo de torsión finitamente generado le podemos asignar unos divisores elementales. Luego probaremos que son únicos:

Teorema 4.47 *Sea D un dominio de ideales principales y M un D–módulo de torsión no nulo finitamente generado. Entonces $M = \langle m_1 \rangle \oplus \cdots \oplus \langle m_r \rangle$, donde cada m_i tiene periodo potencia de primo.*

Demostración: Observemos en primer lugar que si $a, b \in D$ son primos entre sí, entonces $M[ab] = M[a] \oplus M[b]$.

En efecto, por la relación de Bezout 3.22 podemos expresar $ua + vb = 1$, para ciertos $u, v \in D$. Así, dado $m \in M[ab]$, tenemos que $m = uam + vbm$, con $uam \in M[b]$ y $vbm \in M[a]$, luego $M[ab] = M[a] + M[b]$. Además la suma es directa, porque si $m \in M[a] \cap M[b]$, entonces $m = uam + ubm = 0$.

Ahora tomamos un exponente e de M, lo descomponemos en potencias de primos no asociados dos a dos, $e = p_1^{e_1} \cdots p_r^{e_r}$ (notemos que no hace falta añadir un factor unitario porque podemos eliminarlo y seguimos teniendo un periodo) y aplicamos lo anterior:

$$M = M[e] = M[p_1^{e_1}] \oplus \cdots \oplus M[p_r^{e_r}].$$

Ahora basta probar que cada $M[p_i^{e_i}]$ se descompone en suma directa de submódulos monógenos con generadores de periodo potencia de primo. Alternativamente, podemos suponer que $M = M[p^e]$ tiene exponente p^e, donde $p \in D$ es primo. Entonces los periodos de todos los elementos de M son también (salvo unidades) potencias de p.

Sea $k = D/(p)$, que es un cuerpo (por 3.17), y observemos que $M[p]$ tiene estructura de k-espacio vectorial (de dimensión finita) con el producto dado por $[d]m = dm$ (el punto relevante es que el producto está bien definido).

Vamos a probar, por inducción sobre $\dim_k M[p]$, que

$$M = \langle m_1\rangle \oplus \cdots \oplus \langle m_r\rangle,$$

de modo que cada m_i tiene periodo p^{e_i} con $e_1 \geq e_2 \geq \cdots \geq e_r$.

Para ello tomamos $m_1 \in M$ tal que su periodo sea p^{e_1} con el mayor exponente e_1 posible. Si $M = \langle m_1\rangle$ tenemos ya la conclusión. En caso contrario consideramos el cociente $\bar{M} = M/\langle m_1\rangle$, que será no nulo. Como todos los elementos de M tienen periodo divisor de p^{e_1}, tenemos que $\bar{M} = \bar{M}[p^{e_1}]$. Ahora probamos un hecho general:

> *En toda clase $[y] \in \bar{M}$ podemos elegir un representante $[y] = [y']$ de modo que $[y]$ e y' tienen el mismo periodo.*

En efecto: Notemos que, en general, el periodo de una clase divide al de cualquiera de sus representantes. Si el periodo de $[y]$ es p^s, entonces $p^s y \in \langle m_1\rangle$, luego $p^s y = p^t c m_1$, con $(p, c) = 1$, para cierto $t \leq e_1$. Si $t = e_1$, entonces $p^s y = 0$, luego el periodo de y es p^s.

Si $t < e_1$, entonces $p^t c m_1$ tiene periodo p^{e_1-t}, luego y tiene periodo p^{s+e_1-t}, luego $s + e_1 - t \leq e_1$, porque p^{e_1} anula a y, luego $s \leq t$ y concluimos que $y' = y - p^{t-s} c m_1$ cumple $[y'] = [y]$ y que y' tiene periodo p^s.

A continuación veamos que $\dim_k \bar{M}[p] < \dim_k M[p]$.

Para ello tomamos una k-base $[x_1], \ldots, [x_s]$ de $\bar{M}[p]$. Por el resultado que acabamos de obtener, podemos suponer que cada x_i tiene periodo p, es decir, que $x_1, \ldots, x_s \in M[p]$ y se cumple que $p^{e_1-1}m_1, x_1, \ldots, x_s \in M[p]$ son linealmente independientes sobre k, pues si

$$[d]p^{e_1-1}m_1 + [d_1]x_1 + \cdots + [d_s]x_s = 0,$$

entonces $[d_1][x_1] + \cdots + [d_s][x_s] = 0$, luego $[d_1] = \cdots = [d_s] = 0$, luego $[d]p^{e_1-1}m_1 = 0$, luego $p^{e_1} \mid dp^{e_1-1}$, luego $p \mid d$, luego $[d] = 0$.

Así pues, podemos aplicar la hipótesis de inducción, según la cual

$$\bar{M} = \langle [m_2]\rangle \oplus \cdots \oplus \langle [m_r]\rangle,$$

donde cada $[m_i]$ tiene periodo p^{e_i} con $e_2 \geq \cdots \geq e_r$. Según hemos visto, podemos exigir que cada m_i tenga también periodo p^{e_i} y, como e_1 era el mayor posible, se cumple que $e_1 \geq e_2$. Ahora basta ver que

$$M = \langle m_1\rangle \oplus \cdots \oplus \langle m_r\rangle.$$

En efecto, es claro que M se descompone en suma de estos submódulos. Falta ver que la suma es directa, pero es que si

$$d_1 m_1 + \cdots + d_r m_r = 0,$$

entonces en $\bar{M}$ tenemos que $[d_2m_2] + \cdots + [d_rm_r] = 0$, luego, al ser la suma directa, $[d_2m_2] = \cdots = [d_rm_r] = 0$, luego $p^{e_i} \mid d_i$ (para $i \geq 2$), luego $d_im_i = 0$ (porque el periodo de m_i es el mismo que el de $[m_i]$), luego $d_1m_1 = 0$. ∎

Según la última observación previa al teorema, ahora tenemos probado que todo D–módulo de torsión finitamente generado es isomorfo a una suma directa de módulos de tipo D/aD con a potencia de primo.

Ahora falta probar la unicidad, es decir, que si un mismo módulo admite dos descomposiciones en las condiciones del teorema anterior, aunque los generadores m_i puedan ser distintos, el número de sumandos y los órdenes de éstos serán los mismos, lo cual nos permitirá hablar de los divisores elementales del módulo, que no dependerán de la elección de la descomposición elegida para calcularlos.

Hay otro tipo de descomposición de un módulo de torsión que a menudo es más conveniente. La ilustramos primero con un ejemplo:

Ejemplo Consideremos de nuevo el $\mathbb{Z}$-módulo de torsión de divisores elementales $2, 2, 2, 3, 3, 9, 5, 5, 25$, es decir,

$$M = \langle m_1\rangle \oplus \langle m_2\rangle \oplus \langle m_3\rangle \oplus \langle m_4\rangle \oplus \langle m_5\rangle \oplus \langle m_6\rangle \oplus \langle m_7\rangle \oplus \langle m_8\rangle \oplus \langle m_9\rangle\,,$$

donde los órdenes de los m_i son los indicados $(2, 2, 2, 3, \ldots)$. Tomemos los generadores cuyos órdenes son las mayores potencias de cada primo, en este caso $o(m_3) = 2$, $o(m_6) = 9$, $o(m_9) = 25$, y entonces el teorema 3.53 nos da que

$$\langle m_3\rangle \oplus \langle m_6\rangle \oplus \langle m_9\rangle = \langle m_3'\rangle\,,$$

con $o(m_3') = 2 \cdot 9 \cdot 25 = 450$.

De los sumandos restantes, tomamos los que corresponden a divisores elementales con las mayores potencias de primos, en este caso $o(m_2) = 2$, $o(m_5) = 3$, $o(m_8) = 5$, y observamos que

$$\langle m_2\rangle \oplus \langle m_5\rangle \oplus \langle m_8\rangle = \langle m_2'\rangle\,,$$

con $o(m_2') = 30$. Ahora nos quedan $o(m_1) = 2$, $o(m_4) = 3$, $o(m_7) = 5$, con lo que podemos agrupar $\langle m_1\rangle \oplus \langle m_4\rangle \oplus \langle m_7\rangle = \langle m_1'\rangle$, con $o(m_1') = 30$. En total

$$M = \langle m_1'\rangle \oplus \langle m_2'\rangle \oplus \langle m_3'\rangle\,,$$

donde los tres generadores tienen órdenes $30, 30, 450$. Observemos que hemos empezado a numerar los generadores por el m_3' porque el número de generadores que obtenemos mediante esta agrupación es igual al máximo número de divisores elementales con la misma base.

Estos órdenes, caracterizados por que cada uno divide al siguiente, se llaman *factores invariantes* de M, y vamos a ver que sucede lo mismo que con los divisores elementales, es decir, que, por una parte, no es posible obtener dos descomposiciones distintas de M que lleven a dos juegos distintos de factores invariantes y, por otra parte, que dos módulos de torsión con factores invariantes distintos no pueden ser isomorfos. ∎

Teorema 4.48 *Sea D un dominio de ideales principales y M un D-módulo finitamente generado de torsión.*

1. *Existen elementos $x_1, \dots, x_n \in M$ tales que $M_t = \langle x_1\rangle \oplus \cdots \oplus \langle x_n\rangle$ y para cada $i = 1, \dots, n$, $o(x_i) = (p_i^{e_i})$, donde p_i es un primo de D y e_i es un número natural no nulo.*

2. *Existen elementos $y_1, \dots, y_m \in M$ tales que $M_t = \langle y_1\rangle \oplus \cdots \oplus \langle y_m\rangle$, y si $o(y_i) = (f_i)$, entonces para cada $i = 1, \dots, m$, se cumple que f_i no es cero ni unidad y si $i < m$, entonces $f_i \mid f_{i+1}$.*

3. *Los números n y m, los ideales $(p_i^{e_i})$ para $i = 1, \dots, n$ y los ideales (f_i) para $i = 1, \dots, m$ están determinados por M, es decir, cualquier descomposición de M en la forma indicada en 1) o en 2) da lugar a los mismos n, m, etc. Los elementos $p_i^{e_i}$ se llaman* divisores elementales *de M, los elementos f_i se llaman* factores invariantes *de M.*

Demostración: Tenemos probado 1), mientras que 2) es una consecuencia inmediata: Dada una descomposición de tipo 1), multiplicamos todos los primos que aparecen como base de divisores elementales elevados al mayor exponente posible, con lo que obtenemos el último factor invariante f_m (donde m es el mayor número de divisores elementales con la misma base) y el teorema 3.54 nos da que las sumas de los generadores correspondientes a los divisores elementales que hemos tomado es un y_m de periodo f_m. Luego formamos el factor invariante f_{m-1} repitiendo el proceso con los divisores elementales que quedan, y seguimos así hasta que se acaben los divisores elementales.

En cuanto a 3), basta probar la unicidad de los factores invariantes, pues es claro que si existieran dos descomposiciones distintas en divisores elementales, a partir de ellas podríamos obtener dos descomposiciones distintas en factores invariantes por el procedimiento que acabamos de describir. Descomponemos la prueba en varios pasos:

a) Sea $M_t = \langle y_1\rangle \oplus \cdots \oplus \langle y_m\rangle$ una descomposición tipo 2) y sea p un primo de D. Llamando $M_i = \langle y_i\rangle$, es claro que $M[p] = M_1[p] \oplus \cdots \oplus M_m[p]$, y que esto es cierto también si consideramos a todos los módulos como $D/(p)$-espacios vectoriales.

b) Se cumple que $\dim M_i[p] = \begin{cases} 0 & \text{si } p \nmid f_i, \\ 1 & \text{si } p \mid f_i. \end{cases}$

En efecto, si $p \nmid f_i$, entonces un $r \in M_i[p]$ es de la forma $r = uy_i$ para un $u \in D$ y $pr = 0$, luego $puy_i = 0$ y $f_i \mid pu$, luego $f_i \mid u$, y así $r = uy_i = 0$, o sea, $M_i[p] = 0$. Si $f_i = pv$, para un $v \in D$, entonces razonando igual que antes concluimos que $f_i \mid pu$, luego $v \mid u$, digamos $u = tv$, con lo que $r = tvy_i$, y por lo tanto $M_i[p] = \langle vy_i\rangle$ (es claro que $vy_i \in M_i[p]$ y que es no nulo).

c) Por lo tanto la dimensión de $M[p]$ es igual al número de factores invariantes divisibles entre p.

d) Si tenemos dos descomposiciones tipo 2), una en m sumandos y otra en m' sumandos, tomamos un primo que divida al primer factor invariante de la

primera descomposición (y que por lo tanto divide a los m factores invariantes). La definición de $M[p]$ no depende de la descomposición de tipo 2) escogida y su dimensión como $D/(p)$-espacio vectorial es, por una parte, igual al número de factores invariantes de la primera descomposición divisibles entre p (o sea, $= m$) y, por otra parte, es el número de factores invariantes de la segunda descomposición divisibles entre p (es decir, $\leq m'$). Esto prueba que $m \leq m'$, e igualmente se prueba $m' \leq m$, luego tenemos que dos descomposiciones tipo 2) han de tener el mismo número de sumandos. Así pues, el número de factores invariantes es invariante.

Más aún, hemos probado que todo primo que divide al primer factor invariante de una descomposición divide al primer factor invariante de cualquier otra descomposición.

e) En una descomposición de tipo 2), el último factor invariante f_m es múltiplo de todos los anteriores, luego anula a todos los generadores de M_t, y por lo tanto a todos los elementos de M_t, luego es un exponente de M_t (en principio, hemos probado que es un múltiplo de cualquier exponente de M_t, pero cualquier exponente anula a y_m, luego es múltiplo de f_m y concluimos que es asociado a f_m). Puesto que M_t determina sus exponentes, concluimos que dos descomposiciones tipo 2) deben tener igual (salvo unidades) el último factor invariante.

f) Vamos a probar la unicidad de los factores invariantes por inducción sobre el número de factores primos en que se descompone el último factor invariante de M (que ya sabemos que es invariante). Sean dos descomposiciones tipo 2:

$$M_t = \langle y_1\rangle \oplus \cdots \oplus \langle y_m\rangle \quad \text{y} \quad M_t = \langle z_1\rangle \oplus \cdots \oplus \langle z_m\rangle\,,$$

la primera con factores invariantes $f_1, \ldots, f_m$ y la segunda con factores invariantes $g_1, \ldots, g_m$. Ya sabemos que $(f_m) = (g_m)$.

Si f_m se descompone en un solo primo, entonces f_m es primo, y como los restantes factores invariantes son divisores suyos, son todos salvo unidades ese mismo primo, es decir, todos los f_i y los g_i son iguales, luego tenemos la unicidad.

Supongamos que la unicidad se cumple para módulos cuyo último factor invariante se descomponga en n factores primos y que f_m se descompone en $n+1$ primos. En d) hemos probado que f_1 y g_1 son divisibles entre los mismos primos. Sea p un primo que divida a ambos (luego divide a todos los f_i y a todos los g_i).

Sea $pM_t = \{pr \mid r \in M_t\}$. Es claro que pM_t es un submódulo de M_t y además

$$pM_t = \langle py_1\rangle \oplus \cdots \oplus \langle py_m\rangle = \langle pz_1\rangle \oplus \cdots \oplus \langle pz_m\rangle\,.$$

También es obvio que $o(py_i) = (f_i/p)$ y $o(pz_i) = (g_i/p)$.

Puede ocurrir que los primeros f_i sean iguales a p, con lo que los primeros sumandos de estas descomposiciones serían nulos. Si en ambas descomposiciones eliminamos los primeros sumandos si son nulos, obtenemos dos descomposiciones tipo 2) del módulo pM_t, donde el último factor invariante es f_m/p, luego podemos aplicar la hipótesis de inducción y concluir que el número de sumandos nulos es igual para las dos descomposiciones, y que las restantes tienen los

mismos factores invariantes (salvo unidades), es decir, el número de f_i's iguales a p es el mismo que el de g_i's, y para los restantes, $(f_i/p) = (g_i/p)$. Esto implica la igualdad de los (f_i)'s y los (g_i)'s. ∎

En resumen, tenemos la clasificación siguiente de los módulos finitamente generados sobre un dominio de ideales principales:

Teorema 4.49 *Sea D un DIP y M, N dos D-módulos finitamente generados. Entonces $M \cong N$ si y sólo si M y N tienen el mismo rango y los mismos factores invariantes (o el mismo rango y los mismos divisores elementales).*

Demostración: Una implicación es obvia. Supongamos que M y N tienen el mismo rango y los mismos factores invariantes o divisores elementales. Entonces existen descomposiciones

$$M = M' \oplus \langle x_1 \rangle \oplus \cdots \oplus \langle x_m \rangle \qquad N = N' \oplus \langle y_1 \rangle \oplus \cdots \oplus \langle y_m \rangle\,,$$

donde M' y N' son módulos libres del mismo rango y $o(x_i) = o(y_i)$ para cada $i = 1, \ldots, m$.

Pero ya hemos observado que $\langle x_i \rangle \cong D/o(x_i) = D/o(y_i) \cong \langle y_i \rangle$, y por otra parte $M' \cong N'$ porque son dos módulos libres del mismo rango. A partir de un isomorfismo entre cada sumando directo podemos construir un isomorfismo entre las dos sumas, es decir, $M \cong N$. ∎

Es inmediato (y ya lo hemos usado en la demostración de 4.48) que el último factor invariante de un módulo de torsión finitamente generado sobre un DIP es precisamente el exponente del módulo. Esto es un hecho no trivial que resulta útil en algunas ocasiones: el exponente de un módulo de torsión finitamente generado (que en principio es el mínimo común múltiplo de los órdenes de sus elementos) es siempre el orden de uno de sus elementos. Para el caso de grupos abelianos esto es [ITAl 3.28], y de aquí podemos deducir [ITAl 3.29], que ahora enunciamos en una versión un poco más general:

Teorema 4.50 *Todo subgrupo finito del grupo de unidades de un dominio íntegro es cíclico.*

Demostración: Sea D un dominio íntegro y sea K su cuerpo de cocientes. Entonces $U(D) \leq K^*$, luego basta probar que todo subgrupo finito del grupo multiplicativo de un cuerpo es cíclico. Sea, pues, $G \leq K^*$ finito. Entonces G es un grupo abeliano finito, luego podemos considerar su último factor invariante m, que, según acabamos de señalar, es el orden de un elemento de G. Consideramos ahora el polinomio $x^m - 1 \in K[x]$. El hecho de que m sea el exponente de G se traduce en que todos los elementos de G son raíces de este polinomio, pero el número de raíces de un polinomio en un cuerpo no puede exceder a su grado, luego $|G| \leq m$. Como G tiene un elemento de orden m, tiene que ser $|G| = m$, y G es cíclico. ∎

Observemos que, por 4.20, un $\mathbb{Z}$-módulo libre de rango finito r no es más que un producto de $\mathbb{Z}$-módulos isomorfos a $\mathbb{Z}$, es decir, un producto directo de un número finito de grupos cíclicos infinitos, por lo que los teoremas 4.45 y 4.48 nos dan en particular el resultado siguiente:

Teorema 4.51 *Todo grupo abeliano finitamente generado (en particular todo grupo abeliano finito) es producto directo de un número finito de grupos cíclicos.*

Más aún, hemos probado que el número de factores infinitos está unívocamente determinado por el grupo (es su rango) y los factores finitos pueden elegirse de órdenes potencia de primo (y son los divisores elementales del grupo) o bien de modo que cada orden divida al siguiente (y entonces dichos órdenes son sus factores invariantes), y el rango, junto con los factores invariantes o con los divisores elementales, determinan el grupo salvo isomorfismo.

Otra consecuencia inmediata es la siguiente:

Teorema 4.52 *Un grupo abeliano finito tiene subgrupos de todos los órdenes que dividen al orden del grupo.*

DEMOSTRACIÓN: Basta tener en cuenta que todo grupo cíclico finito tiene subgrupos de todos los órdenes que dividen al orden del grupo [TG 1.16]. Así, si $G = G_1 \times \cdots \times G_n$ es un producto de grupos cíclicos, tomando subgrupos $H_i \leq G_i$ de órdenes adecuados y formando productos $H = H_1 \times \cdots \times H_n$ podemos obtener subgrupos de cualquier orden que divida al orden de G. ∎

Terminamos esta sección con un hecho que necesitaremos más adelante:

Teorema 4.53 *Sea D un DIP, sea L un D-módulo libre de rango finito*[7] *y sea M un submódulo de L no nulo y finitamente generado. Entonces existe una base B de L, existen elementos $e_1, \ldots, e_m \in B$ y $f_1, \ldots, f_m \in D$ tales que $f_1e_1, \ldots, f_me_m$ forman una base de M y para cada $i < m$ se cumple que $f_i \mid f_{i+1}$. Además, los ideales (f_i) están unívocamente determinados por L y M.*

DEMOSTRACIÓN: Vamos a probar la existencia por inducción sobre el rango de M (que es también un módulo libre por el teorema 4.41). Observemos que si $B = \{x_1, \ldots, x_n\}$ es una base de L, las aplicaciones $\pi_B^i : L \longrightarrow D$ que a cada $x \in L$ le asignan respectivamente su coordenada i-ésima en la base B son homomorfismos de módulos y, como $M \neq 0$, los módulos $\pi_B^i[M]$ no pueden ser todos nulos.

En general, para cada homomorfismo $h : L \longrightarrow D$, tenemos que $h[M]$ es un ideal de D y, como D es noetheriano, toda familia no vacía de ideales tiene un elemento maximal, es decir, podemos encontrar un homomorfismo h_1 tal que el ideal $h_1[M] = (f_1)$ no esté contenido estrictamente en ningún otro ideal de esta forma. En particular $f_1 \neq 0$, pues los $\pi_B^i[M]$ no son todos nulos.

Sea $y_1 \in M$ tal que $h_1(y_1) = f_1$. Entonces, si $h : L \longrightarrow D$ es cualquier homomorfismo, se tiene que cumplir que $h(y_1) \in (f_1)$, pues en caso contrario $(f_1) \subsetneq (h(y_1)) + (f_1) = (d)$, para cierto $d \in D$, y podemos tomar $u, v \in D$ tales que $uh(y_1) + vh_1(y_1) = d$, luego $h' = uh + vh_1 : L \longrightarrow D$ es un homomorfismo tal que $h'(y_1) = d$, luego $(f_1) \subsetneq (d) \subset h'[M]$.

[7] Si el rango de L no es finito podemos tomar un subconjunto finito de una base de L que permita expresar como combinaciones lineales a todos los generadores de M, y podemos sustituir L por el submódulo generado por dicho conjunto finito, que será un D-módulo libre de rango finito, por lo que el teorema vale igualmente, pero en este paso se usa AE.

En particular, esto implica que si B es cualquier base de L, las coordenadas de y_1 en la base B tienen que ser divisibles entre f_1, luego y_1 es divisible entre f_1. Pongamos que $y_1 = f_1 e_1$, para cierto $e_1 \in L$.

Veamos ahora que $L = \langle e_1\rangle \oplus \mathrm{N}(h_1)$. En efecto, tenemos que $h_1(e_1) = 1$, luego todo $x \in L$ se descompone como $x = h_1(x)e_1 + (x - h_1(x)e_1)$ y el segundo sumando está en el núcleo $\mathrm{N}(h_1)$. Además la suma es directa, porque si tomamos $de_1 \in \langle e_1\rangle \cap \mathrm{N}(h_1)$ al aplicar h_1 sale $d = 0$, luego $de_1 = 0$.

Si $M = \langle f_1 e_1\rangle$ ya se cumplen las condiciones del enunciado. En caso contrario consideramos $L_1 = \mathrm{N}(h_1)$ y $M_1 = M \cap L_1$, y es claro entonces que $M = \langle y_1\rangle \oplus M_1$, pues todo $x \in M$ es de la forma $x = h_1(x)e_1 + a$, con $a \in L_1$ y $h_1(x) = df_1$, con $d \in D$, luego $x = dy_1 + a$, con $a = x - dy_1 \in M_1$.

Por lo tanto, $M_1 \neq 0$ es libre de rango una unidad menos que M. Por hipótesis de inducción existe una base B_1 de L_1, existen $e_2, \ldots, e_m \in B_1$ y existen $f_2, \ldots, f_m \in D$ tales que $f_2e_2, \ldots, f_me_m$ son una base de M_1 y $f_i \mid f_{i+1}$. Es claro entonces que $B = \{e_1\} \cup B_1$ es una base de L y que $f_1e_1, \ldots, f_me_m$ es una base de M.

Además, si $h = \pi_B^1 + \pi_B^2 : L \longrightarrow D$, tenemos que $h(f_1e_1) = f_1$, $h(f_2e_2) = f_2$, luego $h_1[M] = (f_1) \subset (f_1) + (f_2) \subset h[M]$. Como la inclusión no puede ser estricta, tiene que ser $f_1 \mid f_2$.

Veamos ahora la unicidad de los (f_i). Notemos que, a diferencia de lo que sucede en el teorema 4.48, los f_i pueden ser unidades de D (y entonces podemos tomarlos iguales a 1). Pongamos que $f_1 = \cdots = f_r = 1$, que $f_{r+1}, \ldots, f_m$ no son unidades y que $B = \{e_1, \ldots, e_n\}$, con $m \leq n$. Entonces

$$L = \langle e_1, \ldots, e_n\rangle\,, \quad M = \langle f_1e_1, \ldots, f_me_m\rangle\,,$$

de donde se sigue que

$$L/M \cong (\langle e_{r+1}\rangle \,/\, \langle f_{r+1}e_{r+1}\rangle) \oplus \cdots \oplus (\langle e_m\rangle \,/\, \langle f_me_m\rangle) \oplus \langle e_{m+1}\rangle \oplus \cdots \oplus \langle e_n\rangle\,.$$

El último bloque de sumandos forma un D–módulo libre de rango $n - m$, mientras que los primeros forman el módulo de torsión del cociente, y como es claro que $o(\bar{e}_i) = (f_i)$, resulta que los factores invariantes de L/M son precisamente $f_{r+1}, \ldots, f_m$. Así pues, tenemos que m es el rango de M, luego está unívocamente determinado, los f_i no unitarios son los factores invariantes de L/M, luego los ideales que generan están unívocamente determinados, y el número de f_i's unitarios es exactamente m menos el número de f_i's no unitarios, luego todos los (f_i) están unívocamente determinados. ∎

4.6 Apéndice: Espacios vectoriales de dimensión infinita

Recogemos aquí las pruebas de existencia y equicardinalidad de bases para espacios vectoriales no finitamente generados. Todas ellas usan de forma esencial el axioma de elección. En primer lugar probamos la existencia de bases:

Teorema 4.54 (AE) *Si V es un D–espacio vectorial, entonces V es libre. Más aún, todo subconjunto libre de V está contenido en una base.*

DEMOSTRACIÓN: La familia de los subconjuntos libres de V que contienen a uno dado (y siempre podemos tomar como conjunto libre dado el conjunto vacío) está parcialmente ordenada por la inclusión y cumple claramente las hipótesis del lema de Zorn. Por lo tanto todo subconjunto libre de un espacio vectorial está contenido en uno maximal para la inclusión.

En realidad, la prueba vale hasta aquí para módulos cualesquiera, pero ahora observamos que todo subconjunto libre X de un D-espacio vectorial maximal para la inclusión es generador y, por lo tanto, base. En efecto, si no es generador, existe un $v \in V$ que no es combinación lineal de elementos de X, y por la observación previa al teorema 4.25 concluimos que $X \cup \{v\}$ es libre y contradice la maximalidad de X. ∎

También es cierto en general que todo generador contiene una base:

Teorema 4.55 (AE) *Si V es un D–espacio vectorial, todo generador de V contiene una base.*

DEMOSTRACIÓN: Sea X un generador de V. El lema de Zorn nos garantiza la existencia de un conjunto libre $B \subset X$ maximal respecto a la inclusión (es decir, tal que no está estrictamente contenido en ningún otro subconjunto libre de X). Se cumple que $X \subset \langle B\rangle$, pues si existiera un elemento $x \in X \setminus \langle B\rangle$, entonces tenemos que $B \cup \{x\}$ es libre y está contenido en X, en contradicción con la maximalidad de B.

Por lo tanto $V = \langle X\rangle \subset \langle B\rangle$, es decir, $V = \langle B\rangle$, y así el conjunto B es una base de V contenida en X. ∎

Ahora podemos demostrar que todas las bases de un espacio vectorial tienen el mismo cardinal:

Teorema 4.56 (AE) *Si X e Y son dos bases de un mismo D-espacio vectorial, entonces existe una biyección $X \longrightarrow Y$.*

DEMOSTRACIÓN: Sea V un espacio vectorial con bases X e Y. Como ya hemos probado el teorema 4.24, podemos suponer que V no es finitamente generado, con lo que las dos bases son infinitas.

Cada $x \in X$ puede expresarse como combinación lineal de un conjunto finito $Y_x \subset Y$. Si llamamos $Y_0 = \bigcup_{x\in X} Y_x$, tenemos que todo elemento de X es combinación lineal de elementos de Y_0, luego $X \subset \langle Y_0\rangle$, luego $V = \langle X\rangle \subset \langle Y_0\rangle$. Esto implica que $Y = Y_0$, pues un elemento de $Y \setminus Y_0$ podría expresarse como combinación lineal de elementos de Y_0, y entonces Y no sería libre.

Fijando una enumeración de cada conjunto Y_x, podemos definir una aplicación inyectiva $Y \longrightarrow X \times \mathbb{N}$ (a cada $y \in Y$ le asignamos un par (x, i), donde y es el i-ésimo elemento de Y_x). El teorema B.11 nos da entonces una aplicación inyectiva $Y \longrightarrow X$. Similarmente obtenemos una aplicación inyectiva $X \longrightarrow Y$, luego el teorema de Cantor-Bernstein implica que existe la biyección del enunciado. ∎

Admitiendo, tal y como indicamos en el apéndice B, que es posible definir el cardinal de un conjunto arbitrario de modo que dos conjuntos tienen el mismo cardinal si y sólo si son biyectables, ahora es claro que la definición de la dimensión 4.26 es válida para todo espacio vectorial, si bien nosotros sólo la usaremos en el caso de espacios de dimensión finita.

Igualmente, el lector familiarizado con la teoría de cardinales infinitos concluirá ahora que la prueba del teorema 4.27 es válida para módulos no necesariamente finitamente generados, por lo que la definición del rango 4.28 libre vale para todo módulo libre sobre un anillo conmutativo y unitario.

Capítulo V

Extensiones de cuerpos

En [ITAl] estudiamos a fondo los cuerpos cuadráticos, de la forma $\mathbb{Q}\big[\sqrt{d}\,\big]$, y vimos que algunos problemas aritméticos llevan a considerar otros cuerpos de números algebraicos más complejos, como los cuerpos ciclotómicos $\mathbb{Q}[\omega]$. En este capítulo desarrollaremos una teoría general sobre cuerpos de números algebraicos, para lo cual es esencial el aparato algebraico que proporcionan la teoría de espacios vectoriales y la teoría de grupos, sin los cuales difícilmente podíamos ir más allá de estudiar los ejemplos sencillos que consideramos en [ITAl].

5.1 Extensiones algebraicas

He aquí las definiciones básicas de la teoría de extensiones de cuerpos:

Definición 5.1 Diremos que K/k es una *extensión de cuerpos* (o simplemente una extensión) si K es un cuerpo y k es un subcuerpo de K. El cuerpo k se llama *cuerpo base* de la extensión.

Si $a \in K$, se dice que a es *algebraico* sobre k si existe un polinomio $p(x) \in k[x]$ no nulo tal que $p(a) = 0$. En caso contrario se dice que es *trascendente* sobre k.

La extensión K/k es *algebraica* si todos los elementos de K son algebraicos sobre k. En caso contrario se dice que es *trascendente.*

Ejemplos

- Todo elemento a del cuerpo base es algebraico, pues es raíz del polinomio $x - a \in k[x]$.
- Los números reales $\sqrt{2}$, $\sqrt[3]{5}$ son algebraicos sobre $\mathbb{Q}$, pues son raíces de los polinomios $x^2 - 2$, $x^3 - 5 \in \mathbb{Q}[x]$.
- No es trivial en absoluto, pero en [ITAn 10.7] vimos que el número e es trascendente (sobre $\mathbb{Q}$), y en [ITAn 10.9] vimos que π también lo es.
- Un ejemplo sencillo de elemento trascendente sobre $\mathbb{Q}$ es la indeterminada x en $\mathbb{Q}(x)$ (el cuerpo de cocientes del anillo de polinomios $\mathbb{Q}[x]$).

Introducimos ahora la notación adecuada para describir las extensiones de cuerpos:

Sea B un dominio íntegro y A un subanillo unitario. Sea S un subconjunto de B. Llamaremos $A[S]$ a la intersección de todos los subanillos de B que contienen a A y a S.

Es fácil probar que

$$A[S] = \{p(a_1, \dots, a_n) \mid n \in \mathbb{N}, p(x_1, \dots, x_n) \in A[x_1, \dots, x_n], a_1, \dots, a_n \in S\},$$

pues este conjunto es ciertamente un subanillo de B que contiene a A y a S, luego contiene a $A[S]$, y por otra parte, como $A[S]$ es un anillo, ha de contener a todos los elementos de la forma $p(a_1, \dots, a_n)$.

Sea ahora K un cuerpo y k un subcuerpo de K. Si S es un subconjunto de K, llamaremos $k(S)$ a la intersección de todos los subcuerpos de K que contienen a k y a S. Se prueba igualmente que

$$k(S) = \left\{ \frac{p(b_1, \dots b_n)}{q(b_1, \dots b_n)} \mid n \in \mathbb{N}, p, q \in k[x_1, \dots, x_n], b_i \in S, q(b_1, \dots, b_n) \neq 0 \right\}.$$

El cuerpo $k(S)$ se llama *adjunción* a k de S. Cuando $S = \{b_1, \dots, b_n\}$ es un conjunto finito escribiremos también

$$A[b_1, \dots, b_n] \quad \text{y} \quad k(b_1, \dots, b_n).$$

Notemos que $A[S \cup T] = A[S][T]$ y $k(S \cup T) = k(S)(T)$. En particular

$$A[b_1, \dots, b_n] = A[b_1] \dots [b_n] \quad \text{y} \quad k(b_1, \dots, b_n) = k(b_1) \dots (b_n).$$

Observemos que la notación $A[S]$ y $k(S)$ para los anillos de polinomios y los cuerpos de fracciones algebraicas que venimos utilizando es un caso particular de la que acabamos de introducir.

Una extensión K/k es *finitamente generada* si $K = k(S)$, para un conjunto finito $S \subset K$. Si $K = k(a)$ la extensión es *simple*. Todo a que cumpla esto (que no es único) se llama *elemento primitivo* de la extensión.

El teorema siguiente nos describe las extensiones que obtenemos al adjuntar a un cuerpo un elemento algebraico:

Teorema 5.2 *Sea K/k una extensión de cuerpos y $a \in K$ un elemento algebraico sobre k. Entonces:*

1. *Existe un único polinomio mónico irreducible $p(x) \in k[x]$ tal que $p(a) = 0$.*
2. *Un polinomio $g(x) \in k[x]$ cumple $g(a) = 0$ si y sólo si $p(x) \mid g(x)$.*
3. *$k(a) = k[a] = \big\{r(a) \mid r(x) \in k[x]$ y $\operatorname{grad} r(x) < \operatorname{grad} p(x)\big\}$.*

DEMOSTRACIÓN: Consideremos el epimorfismo $\phi : k[x] \longrightarrow k[a]$ dado por $\phi(g(x)) = g(a)$. El hecho de que a sea algebraico significa que $\mathrm{N}(\phi)$ es un ideal no nulo y, como $k[x]$ es DIP, existe un polinomio no nulo $p(x) \in k[x]$ tal que $\mathrm{N}(\phi) = \big(p(x)\big)$. Como las constantes son unidades, podemos exigir que $p(x)$ sea mónico (al dividir por el coeficiente director obtenemos un asociado que genera el mismo ideal).

Por el teorema de isomorfía, $k[x]/\big(p(x)\big) \cong k[a]$, que es un dominio íntegro, luego el ideal $(p(x))$ es primo, y por 3.17 es maximal, luego $p(x)$ es irreducible y $k[x]/\big(p(x)\big)$ es un cuerpo. Por lo tanto $k[a]$ resulta ser un cuerpo, de donde se sigue que $k[a] = k(a)$.

El polinomio p es único, pues si $q(x)$ también cumple 1), entonces $q(a) = 0$, es decir, $q(x) \in \mathrm{N}(\phi) = \big(p(x)\big)$, luego $p(x) \mid q(x)$, pero si $q(x)$ es irreducible han de ser asociados, es decir, difieren en una constante, y al ser ambos mónicos deben coincidir.

El apartado 2) es consecuencia de que $\mathrm{N}(\phi) = \big(p(x)\big)$. Respecto a 3), un elemento de $k[a]$ es de la forma $q(a)$ con $q(x) \in k[x]$. Existen polinomios $c(x)$ y $r(x)$ tales que $q(x) = p(x)c(x) + r(x)$ y $\operatorname{grad} r(x) < \operatorname{grad} p(x)$. Entonces $q(a) = p(a)c(a) + r(a) = 0 \cdot c(a) + r(a) = r(a)$, luego tiene la forma pedida. ∎

Definición 5.3 Sea K/k una extensión de cuerpos y $a \in K$ un elemento algebraico sobre k. Llamaremos *polinomio mínimo* de a sobre k al polinomio $\operatorname{pol\,mín}(a, k) \in k[x]$ que cumple el teorema anterior.

Así pues, $\operatorname{pol\,mín}(a, k)$ es el menor polinomio no nulo de $k[x]$ que tiene a a por raíz, en el sentido de que divide a cualquier otro que cumpla lo mismo.

Ejemplo El teorema 3.51 nos asegura que todo cuerpo tiene una extensión en la que un polinomio dado tiene una raíz. Por ejemplo, si partimos de $\mathbb{Q}$ y del polinomio $x^2 + 5$, sabemos que existe un cuerpo K en el que existe un elemento $\sqrt{-5}$ que es raíz de dicho polinomio y, como éste es irreducible en $\mathbb{Q}[x]$, el teorema anterior nos da que $\operatorname{pol\,mín}(\sqrt{-5}, \mathbb{Q}) = x^2 + 5$. Más aún, por el apartado 3), sabemos que

$$\mathbb{Q}\big(\sqrt{-5}\,\big) = \mathbb{Q}\big[\sqrt{-5}\,\big] = \{a + b\sqrt{-5} \mid a, b \in \mathbb{Q}\}.$$

∎

A continuación observamos otro hecho fundamental:

Definición 5.4 Si K/k es una extensión de cuerpos, entonces K es un k–espacio vectorial con las operaciones obvias. Llamaremos *grado* de la extensión a la dimensión de K como k–espacio vectorial. Lo representaremos por $|K : k|$. Una extensión es *finita* o *infinita* según lo sea su grado.

Es claro que toda extensión finita es finitamente generada, pues si S es una k–base de K es claro que $K = k[S] = k(S)$.

Ahora podemos precisar el teorema anterior:

Teorema 5.5 *Sea K/k una extensión de cuerpos y $a \in K$ un elemento algebraico sobre k. Sea $p(x) = \operatorname{pol\,mín}(a, k)$. Entonces:*

1. *La extensión $k(a)/k$ es finita y $|k(a) : k| = \operatorname{grad} p(x)$.*
2. *Una base de $k(a)$ sobre k es $\{1, a, \ldots, a^{n-1}\}$, donde $n = \operatorname{grad} p(x)$.*

DEMOSTRACIÓN: El teorema 5.2 3) afirma que $\{1, a, \ldots, a^{n-1}\}$ es un generador de $k(a)$ como k–espacio vectorial. Por otra parte ha de ser libre, pues una combinación lineal de sus elementos no es sino un polinomio $q(a)$ con coeficientes en k y grado menor o igual que $n-1$, luego $q(a) = 0$ implica que $p(x) \mid q(x)$, luego por grados $q(x) = 0$ y los coeficientes de la combinación lineal son nulos. ∎

Ejemplos Si $d \in \mathbb{Z}$ es libre de cuadrados, el polinomio $x^2 + d$ es irreducible en $\mathbb{Q}[x]$ porque no tiene raíces (si no tiene raíces en $\mathbb{Z}$, tampoco las tiene en $\mathbb{Q}$, por 3.35), luego si $\sqrt{d}$ es una raíz en una extensión de $\mathbb{Q}$, el cuerpo $\mathbb{Q}\big[\sqrt{d}\,\big]$ está formado por los elementos de la forma $a + b\sqrt{d}$, con $a, b \in \mathbb{Q}$. Obtenemos así los cuerpos cuadráticos que introdujimos en [ITAl 9.1].

En el capítulo III, como aplicación del criterio de Eisenstein, hemos probado que si $p \geq 3$ es un número primo, el polinomio ciclotómico

$$c_p(x) = x^{p-1} + x^{p-1} + \cdots + x + 1$$

es irredudible en $\mathbb{Q}[x]$, luego si llamamos ω a una raíz de $c_p(x)$ en una extensión de $\mathbb{Q}$, el cuerpo $\mathbb{Q}[\omega]$ tiene grado $p-1$ sobre $\mathbb{Q}$ y está formado por elementos de la forma

$$a_{p-2}\omega^{p-1} + \cdots + a_1\omega + a_0, \qquad a_0, a_1, \ldots, a_{p-1} \in \mathbb{Q}.$$

Éstos son los cuerpos ciclotómicos de orden primo introducidos en [ITAl 17.1]. ∎

Ahora probamos que al adjuntar a un cuerpo un elemento algebraico, todos los elementos de la extensión resultante son algebraicos. Más en general:

Teorema 5.6 *Toda extensión finita es algebraica.*

DEMOSTRACIÓN: Sea K/k una extensión finita de grado n y sea $a \in K$. Consideremos las potencias

$$1, \quad a, \quad a^2, \quad \ldots \quad , \quad a^n.$$

Si hay dos iguales, digamos $a^i = a^j$ con $i \neq j$, entonces a es raíz del polinomio no nulo $x^i - x^j \in k[x]$.

Si son distintas, son $n+1$ elementos distintos de un espacio vectorial de dimensión n, luego no pueden ser linealmente independientes. Existen coeficientes $b_0, \ldots, b_n$ en k no todos nulos de modo que $b_0 + b_1 a + \cdots + b_n a^n = 0$, con lo que a es raíz del polinomio no nulo $b_0 + b_1 x + \cdots + b_n x^n \in k[x]$. ∎

Ejemplo Consideremos el elemento $1+\sqrt{-5} \in \mathbb{Q}\big[\sqrt{-5}\,\big]$. Como la extensión $\mathbb{Q}\big[\sqrt{-5}\,\big]/\mathbb{Q}$ tiene grado 2, siguiendo la prueba del teorema anterior es suficiente considerar las potencias 1, $1+\sqrt{-5}$, $(1+\sqrt{-5}\,)^2 = -4+2\sqrt{-5}$. Hay que buscar números racionales a, b, c que cumplan

$$a + b(1+\sqrt{-5}) + c(-4+2\sqrt{-5}\,) = a + b - 4c + (b+2c)\sqrt{-5} = 0,$$

lo que equivale a que $a+b-4c = 0 = b+2c$. Por ejemplo sirven $c=1$, $b=-2$ y $a=6$. Así pues, $1+\sqrt{-5}$ es raíz del polinomio $6-2x+x^2 \in \mathbb{Q}[x]$. ∎

La potencia del teorema anterior se amplia al combinarlo con el siguiente:

Teorema 5.7 (Teorema de transitividad de grados) *Sea $k \subset K \subset L$ una cadena de extensiones de cuerpos. Entonces $|L:k| = |L:K| \cdot |K:k|$.*

Lo probaremos para extensiones finitas, aunque el caso general se prueba sin ningún cambio importante.

Demostración: Sea $\{x_1, \ldots, x_m\}$ una base de K como k–espacio vectorial. Sea $\{y_1, \ldots, y_n\}$ una base de L como K–espacio vectorial. Basta probar que el conjunto

$$\{x_i y_j \mid i = 1, \ldots, m, j = 1, \ldots, n\}$$

es una base de L como k–espacio vectorial con exactamente mn elementos.

Supongamos que $\sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij} x_i y_j = 0$ para ciertos coeficientes a_{ij} en k. Entonces, para cada j, se cumple que $\sum_{i=1}^{m} a_{ij} x_i \in K$ y como $\{y_1, \ldots, y_n\}$ es una base de L sobre K, podemos concluir que $\sum_{i=1}^{m} a_{ij} x_i = 0$ para cada j. Ahora, al ser $\{x_1, \ldots, x_m\}$ una base de K sobre k, podemos concluir que todos los coeficientes a_{ij} son nulos.

Esto prueba que los $x_i y_j$ son distintos para índices distintos y que forman un conjunto linealmente independiente. En particular tiene mn elementos.

Ahora sea z cualquier elemento de L. Como $\{y_1, \ldots, y_n\}$ es una base de L sobre K, existen elementos $b_1, \ldots, b_n$ en K tales que $z = \sum_{j=1}^{n} b_j y_j$. A su vez, cada elemento b_j se expresa como combinación lineal $b_j = \sum_{i=1}^{m} a_{ij} x_i$ para ciertos coeficientes a_{ij} en k.

Por tanto $z = \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij} x_i y_j$ es combinación lineal de los elementos $x_i y_j$, que son, pues un generador de L. ∎

En particular una extensión finita de una extensión finita es una extensión finita del cuerpo menor.

Ahora podemos estudiar el efecto de adjuntar a un cuerpo cualquier cantidad de elementos algebraicos:

Teorema 5.8 *Sea K/k una extensión y S un conjunto de elementos de K algebraicos sobre k. Entonces $k(S) = k[S]$ y $k(S)/k$ es una extensión algebraica. Si el conjunto S es finito, entonces $k(S)/k$ es finita.*

DEMOSTRACIÓN: Supongamos primero que $S = \{a_1, \ldots, a_n\}$ es finito. Entonces $k(a_1) = k[a_1]$, ahora bien, como $k[x] \subset k(a_1)[x]$, todo elemento algebraico sobre k lo es sobre $k(a_1)$, luego aplicando de nuevo el teorema 5.2 tenemos que $k(a_1)(a_2) = k(a_1)[a_2]$, luego $k(a_1, a_2) = k[a_1][a_2] = k[a_1, a_2]$. Además $k(a_1, a_2)/k(a_1)$ es finita y por el teorema de transitividad de grados $k(a_1, a_2)/k$ también es finita. Repitiendo n veces llegamos a que la extensión $k(a_1, \ldots, a_n)/k$ es finita y $k[a_1, \ldots, a_n] = k(a_1, \ldots, a_n)$.

Sea ahora S un conjunto cualquiera. Si $a \in k(S)$, entonces

$$a = \frac{p(b_1, \ldots b_n)}{q(b_1, \ldots b_n)},$$

para ciertos polinomios p, $q \in k[x_1, \ldots, x_n]$ y ciertos $b_1, \ldots, b_n \in S$.

Por lo tanto $a \in k(b_1, \ldots, b_n) = k[b_1, \ldots, b_n] \subset k[S]$. Además, según lo ya probado, la extensión $k(b_1, \ldots, b_n)/k$ es algebraica, luego a es algebraico sobre k. En consecuencia $k(S) = k[S]$ es una extensión algebraica de k. ∎

Por consiguiente, una extensión algebraica es finita si y sólo si es finitamente generada. Una propiedad que acaba de redondear el comportamiento de las extensiones algebraicas es la siguiente:

Teorema 5.9 *Consideremos cuerpos $k \subset K \subset L$. Entonces la extensión L/k es algebraica si y sólo si lo son L/K y K/k.*

DEMOSTRACIÓN: Si L/k es algebraica, es obvio que K/k lo es. Por otra parte, todo elemento de L es raíz de un polinomio no nulo con coeficientes en k, luego en K, y esto quiere decir que L/K también es algebraica.

Supongamos que L/K y K/k son algebraicas. Tomemos un $a \in L$. Entonces a es algebraico sobre K. Sean $b_1, \ldots, b_n \in K$ los coeficientes de $\operatorname{pol\,mín}(a, K)$. Así, $\operatorname{pol\,mín}(a, K) \in k(b_1, \ldots, b_n)[x]$, luego a es algebraico sobre $k(b_1, \ldots, b_n)$, luego la extensión $k(b_1, \ldots, b_n)(a)/k(b_1, \ldots, b_n)$ es finita. Como $b_1, \ldots, b_n$ son algebraicos sobre k, la extensión $k(b_1, \ldots, b_n)/k$ también es finita, luego por transitividad de grados, $k(b_1, \ldots, b_n)(a)/k$ es finita, luego algebraica, luego a es algebraico sobre k. ∎

Una consecuencia sencilla, pero importante, de estas propiedades es que las operaciones con elementos algebraicos dan elementos algebraicos.

Teorema 5.10 *Si K/k es una extensión de cuerpos, el conjunto de los elementos de K que son algebraicos sobre k es un subcuerpo de K.*

DEMOSTRACIÓN: Si $\alpha, \beta \in K$ son algebraicos sobre K, entonces la extensión $k(\alpha, \beta)/k$ es finita, luego es algebraica, luego $\alpha \pm \beta$, $\alpha\beta$, α/β (el cociente si $\beta \neq 0$) son algebraicos sobre k, pues están en $k(\alpha, \beta)$. ∎

Ejercicio: Probar que si K/k es una extensión y $p(x) \in K[x]$ tiene coeficientes algebraicos sobre k, entonces toda raíz de $p(x)$ en K es algebraica sobre k.

Estudiamos ahora los homomorfismos entre extensiones. Como los cuerpos no tienen ideales propios, un homomorfismo de cuerpos no nulo es de hecho un monomorfismo, por lo que definiremos tan sólo monomorfismos e isomorfismos entre extensiones.

Definición 5.11 Sean K/k y L/l dos extensiones de cuerpos. Un *isomorfismo* entre ellas es un isomorfismo de cuerpos $\phi : K \longrightarrow L$ tal que $\phi[k] = l$.

Si K/k y L/k son extensiones de un mismo cuerpo k, entonces un *k–monomorfismo (k–isomorfismo)* entre ellas es un monomorfismo (isomorfismo) $\phi : K \longrightarrow L$ que deja invariantes a los elementos de k. En particular los k–monomorfismos de extensiones son monomorfismos de k–espacios vectoriales.

Si K/k es una extensión, un *k–automorfismo* de K es un k–isomorfismo de K/k en K/k, es decir, un isomorfismo de K en K que deja invariantes a los elementos de k.

Notemos que si $\phi : K \longrightarrow K$ es un isomorfismo de cuerpos, entonces el conjunto $\{a \in K \mid \phi(a) = a\}$ es un subcuerpo de K, luego contiene al cuerpo primo. Esto quiere decir, por ejemplo, que los $\mathbb{Q}$–automorfismos de una extensión $K/\mathbb{Q}$ son todos los automorfismos de K, es decir, la condición de fijar a los elementos de $\mathbb{Q}$ no es una restricción en realidad.

Llamaremos *grupo de Galois* de una extensión de cuerpos K/k el grupo $G(K/k)$ de todos los k–automorfismos de K.

Veremos que, a través de los grupos de Galois, la teoría de grupos se convierte en una poderosa herramienta para el estudio de las extensiones de cuerpos.

Si tenemos un cuerpo k y un polinomio irreducible $p(x) \in k[x]$, el teorema 3.51 nos da una extensión de k donde $p(x)$ tiene una raíz. Si recordamos la prueba veremos que la extensión es concretamente $K = k[x]/\big(p(x)\big)$ y la raíz es $a = [x]$ (identificando a k con las clases de polinomios constantes). Esta extensión cumple además que $K = k(a)$. Por otra parte, en la prueba del teorema 5.2 hemos obtenido que toda extensión de la forma $k(a)$, donde a es raíz de $p(x)$, es isomorfa a la construida en 3.51. Como consecuencia, dos cualesquiera de estas extensiones son isomorfas entre sí. Lo probamos en un contexto más general.

Teorema 5.12 *Sean K/k y L/l dos extensiones y $\sigma : k \longrightarrow l$ un isomorfismo. Sea $a \in K$ un elemento algebraico sobre k. Sea $p(x) = \operatorname{pol\,mín}(a, k)$. Consideremos la extensión de σ a los anillos de polinomios $\sigma : k[x] \longrightarrow l[x]$. Sea b una raíz en L de $\sigma p(x)$. Entonces σ se extiende a un isomorfismo $\sigma^* : k(a) \longrightarrow l(b)$ tal que $\sigma^*(a) = b$.*

Demostración: La aplicación $\phi : k[x] \longrightarrow k(a)$ dada por $\phi\big(g(x)\big) = g(a)$ es un epimorfismo cuyo núcleo es precisamente el ideal $\big(p(x)\big)$, luego por el teorema de isomorfía $k[x]/\big(p(x)\big) \cong k(a)$, y la imagen de $[x]$ por el isomorfismo es a.

Es obvio que $\sigma p(x) = \operatorname{pol\,mín}(b, l)$, luego por el mismo argumento tenemos también que $l[x]/\big(\sigma p(x)\big) \cong l(b)$, y la imagen de $[x]$ por el isomorfismo es b.

Por otra parte el isomorfismo $\sigma : k[x] \longrightarrow l[x]$ cumple $\sigma(x) = x$ e induce un isomorfismo $k[x]/\big(p(x)\big) \cong l[x]/\big(\sigma p(x)\big)$ que lleva $[x]$ a $[x]$.

La composición de todos estos isomorfismos nos da el isomorfismo buscado. ∎

En particular, un cuerpo de la forma $k(a)$ (con a algebraico) está totalmente determinado por k y el polinomio mínimo de a. Para enunciar esto de la forma más adecuada conviene introducir un concepto:

Definición 5.13 Sean K/k y L/k dos extensiones de un mismo cuerpo k y sean $a \in K$ y $b \in L$ dos elementos algebraicos sobre k. Diremos que son *k–conjugados* si su polinomio mínimo sobre k es el mismo.

Teorema 5.14 *Sean K/k y L/k dos extensiones del mismo cuerpo k, sean $a \in K$ y $b \in L$ algebraicos sobre k. Entonces a y b son k–conjugados si y sólo si existe un k–isomorfismo $\sigma : k(a) \longrightarrow k(b)$ tal que $\sigma(a) = b$.*

Demostración: Si a y b son k–conjugados el resultado se sigue del teorema anterior. Si existe σ en dichas condiciones y $p(x) = \operatorname{pol\,mín}(a, k)$, entonces $p(a) = 0$, luego también $p(b) = p\big(\sigma(a)\big) = \sigma\big(p(a)\big) = 0$, con lo que $p(x) = \operatorname{pol\,mín}(b, k)$ ∎

En la prueba anterior hemos usado dos hechos elementales, pero de uso muy frecuente: el primero es que un polinomio mónico irreducible es el polinomio mínimo de cualquiera de sus raíces. El segundo es que la imagen por un k–monomorfismo de una raíz de un polinomio de $k[x]$ es necesariamente otra raíz de dicho polinomio.

Ejemplo Si tomamos cualquier extensión de $\mathbb{Q}$ en la que el polinomio $x^2 + 5$ tenga una raíz, podemos formar el cuerpo $\mathbb{Q}\big[\sqrt{5}\,\big]$ que resulta de adjuntar a $\mathbb{Q}$ dicha raíz. Por ejemplo, podemos considerar concretamente el cuerpo construido en la prueba del teorema 3.51, y entonces $\mathbb{Q}\big[\sqrt{5}\,\big] = \mathbb{Q}[x]/(x^2 - 5)$, con $\sqrt{-5} = [x]$.

Ahora bien, ya conocemos un cuerpo donde $x^2 - 5$ tiene una raíz, a saber, el cuerpo $\mathbb{R}$ de los números reales, por lo que también podemos considerar el cuerpo $\mathbb{Q}\big[\sqrt{5}\,\big] \subset \mathbb{R}$. El teorema anterior nos dice que los dos cuerpos $\mathbb{Q}\big[\sqrt{5}\,\big]$ son $\mathbb{Q}$-isomorfos, por lo que es irrelevante que pensemos en uno o en otro. Da igual pensar que $3 + 7\sqrt{5}$ es un número real o una clase de equivalencia de polinomios. Los elementos de la forma $a + b\sqrt{5}$, sean números reales o clases de equivalencia de polinomios, constituyen cuerpos $\mathbb{Q}$-isomorfos, y todas las propiedades algebraicas que valen para uno, valen para el otro. ∎

Definición 5.15 Llamamos *cuerpo de los números complejos* a cualquier cuerpo de la forma $\mathbb{C} = \mathbb{R}[i]$, donde i es una raíz del polinomio $x^2 + 1$.

Por la misma discusión del ejemplo precedente, es irrelevante qué cuerpo en concreto consideramos, pues dos cualesquiera serán $\mathbb{R}$-isomorfos.

El hecho de que $i^2 = -1$ determina completamente la aritmética de $\mathbb{C}$. Explícitamente:

$$(x+yi)+(x'+y'i) = (x+x')+(y+y')i,\ (x+yi)(x'+y'i) = (xx'-yy')+(xy'+yx')i.$$

Por ejemplo, en el capítulo I de [ITAn] definimos $\mathbb{C} = \mathbb{R}^2$, y convertimos las fórmulas precedentes en la definición de la suma y el producto en $\mathbb{C}$, es decir:

$$(x,y) + (x',y') = (x+x', y+y'), \qquad (x,y)(x',y') = (xx'-yy', xy'+yx').$$

Esta definición obliga a comprobar que estas operaciones realmente definen un cuerpo que contiene un subcuerpo isomorfo a $\mathbb{R}$ (el formado por los pares $(x,0)$) y un elemento $i = (0,1)$ que cumple $i^2 = -1$.

En cambio, en el capítulo VIII de [ITAl] vimos que una construcción alternativa de $\mathbb{C}$ es definirlo como el cuerpo que proporciona la demostración del teorema 3.51, es decir:

$$\mathbb{C} = \mathbb{R}[x]/(x^2+1),$$

de modo que $i = [x]$.

Ahora sabemos que es irrelevante qué definición de $\mathbb{C}$ adoptemos, al igual que es irrelevante construir $\mathbb{R}$ mediante secciones de Dedekind o mediante sucesiones de Cauchy. En cualquier caso, los números complejos serán los objetos de la forma $z = x + yi$, con $x, y \in \mathbb{R}$, donde la expresión es única. ∎

Veamos un primer resultado sobre grupos de Galois:

Teorema 5.16 *Si K/k es una extensión finita, entonces el grupo de Galois $G(K/k)$ es finito.*

Demostración: Sea $K = k[S]$, donde $S = \{\alpha_1, \ldots, \alpha_n\}$. Entonces, cada $\alpha \in K$ es de la forma $\alpha = p(\alpha_1, \ldots, \alpha_n)$, con $p(x_1, \ldots, x_n) \in k[x_1, \ldots, x_n]$. Si $\sigma \in G(K/k)$, entonces $\sigma(\alpha) = p(\sigma(\alpha_1), \ldots, \sigma(\alpha_n))$. Por lo tanto, si dos k-automorfismos coinciden sobre S, son iguales.

Ahora bien, si $p_i(x) \in k[x]$ es el polinomio mínimo de α_i, tenemos que $p_i(\alpha_i) = 0$, luego $p_i(\sigma(\alpha_i)) = \sigma(p(\alpha_i)) = 0$, luego $\sigma(\alpha_i)$ es una de las raíces de $p_i(x)$ en K. Como un polinomio tiene un número finito de raíces en un cuerpo, resulta que $\sigma(\alpha_i)$ sólo puede tomar un número finito de valores. Si llamamos T al conjunto de las imágenes de elementos de S por elementos de $G(K/k)$, tenemos que es un conjunto finito, y la aplicación $G(K/k) \longrightarrow T^S$ dada por $\sigma \mapsto \sigma|_S$ es inyectiva, luego $G(K/k)$ es finito. ∎

Particularicemos el argumento del teorema anterior al caso de una extensión algebraica simple $k(a)/k$. Tenemos que un k–automorfismo σ está determinado por el valor $\sigma(a)$ que toma en el elemento primitivo a, así como que $\sigma(a)$ ha de ser una raíz de $\operatorname{pol}\operatorname{mín}(a,k)$, luego la extensión $k(a)/k$ tiene a lo sumo tantos k–automorfismos como raíces tiene (en $k(a)$) el polinomio mínimo de a.

Recíprocamente, si $b \in k(a)$ es una raíz de $\operatorname{pol\,mín}(a,k)$, entonces a y b son k–conjugados, luego existe un k–isomorfismo $\sigma : k(a) \longrightarrow k(b)$, pero $k(b) \subset k(a)$ y ambos tienen el mismo grado sobre k (el grado del polinomio mínimo de a). Así pues, $k(a) = k(b)$ y σ es un k–automorfismo de $k(a)$.

En resumen, una extensión algebraica simple $k(a)/k$ tiene exactamente tantos k–automorfismos como raíces tiene en $k(a)$ el polinomio mínimo de a. En particular el número de k–automorfismos no puede superar al grado de la extensión.

Por ejemplo, en la extensión $\mathbb{C}/\mathbb{R}$, el elemento primitivo i tiene dos conjugados, él mismo y $-i$, por lo que $\mathbb{C}$ tiene exactamente dos $\mathbb{R}$-automorfismos, la identidad y la *conjugación compleja* dada por $i \mapsto -i$, es decir, $x + yi \mapsto x - yi$. En general se representa por $\bar{z} = x - yi$ al conjugado del número complejo $z = x + yi$.

Similarmente, el cuerpo $\mathbb{Q}\big[\sqrt{-5}\,\big]$ tiene dos automorfismos, la identidad y el determinado por $\sigma(a + b\sqrt{-5}) = a - b\sqrt{-5}$.

Recordemos que en el estudio de la aritmética de los cuerpos de números algebraicos que mostramos en [ITAl] era fundamental el concepto de norma. Uno de los resultados que proporciona la teoría de extensiones algebraicas es la posibilidad de definir normas similares en cualquier extensión finita. Si una extensión K/k tiene grado n y $\sigma_1, \dots \sigma_n$ son k–automorfismos de K, la norma de un elemento $a \in K$ puede definirse como

$$\mathrm{N}(a) = \sigma_1(a) \cdots \sigma_n(a)$$

(véase por ejemplo la definición [ITAl 17.8] para el caso de los cuerpos ciclotómicos de orden primo). Es claro que una norma así definida conserva productos, pero todavía no sabemos probar otro hecho fundamental, y es que, para que sirva de algo, $\mathrm{N}(a)$ ha de pertenecer al cuerpo base k. En [ITAl 17.9] lo probamos para el caso de los cuerpos ciclotómicos de orden primo, y para los cuerpos cuadráticos es inmediato, pero necesitamos profundizar más en la teoría de extensiones algebraicas para obtener un resultado general.

Para empezar, no todas las extensiones algebraicas simples tienen tantos isomorfismos como su grado. Se trata de una patología relativamente frecuente que debemos comprender. A continuación analizamos con detalle un ejemplo concreto. Nuestro objetivo a medio plazo será obtener un marco general que sustituya a los razonamientos particulares que aquí vamos a emplear:

Ejemplo Consideremos el polinomio $x^3 - 2$, que no tiene raíces en $\mathbb{Q}$ porque no las tiene en $\mathbb{Z}$ (criterio de Gauss). Por lo tanto es irreducible en $\mathbb{Q}[x]$. Sin embargo, tiene una única raíz $\sqrt[3]{2} \in \mathbb{R}$, que nos permite formar la extensión $\mathbb{Q}[\sqrt[3]{2}]/\mathbb{Q}$, de grado 3. Tenemos la factorización

$$x^3 - 2 = (x - \sqrt[3]{2})(x^2 + \sqrt[3]{2}\,x + \sqrt[3]{4}),$$

donde el segundo factor no tiene raíces en $\mathbb{R}$ (porque la raíz cúbica de un número es única), luego tampoco en $\mathbb{Q}[\sqrt[3]{2}]$. Así pues, el elemento primitivo $\sqrt[3]{2}$ no tiene conjugados en $\mathbb{Q}[\sqrt[3]{2}]$ (aparte de sí mismo), luego $G(\mathbb{Q}[\sqrt[3]{2}\,]/\mathbb{Q}) = 1$.

Sin embargo, podemos aplicar el teorema 3.51 al cuerpo $\mathbb{Q}[\sqrt[3]{2}]$ para obtener un cuerpo en el que el polinomio $x^2 + \sqrt[3]{2}\,x + \sqrt[3]{4}$ tiene una raíz β, lo que nos permite formar el cuerpo $\mathbb{Q}[\sqrt[3]{2}, \beta]$, que tiene grado 2 sobre $\mathbb{Q}[\sqrt[3]{2}]$, luego grado 6 sobre $\mathbb{Q}$.

El polinomio $x^3 - 2$ tiene dos raíces en $\mathbb{Q}\big(\sqrt[3]{2}, \beta\big)$, luego tiene tres. Por razones de simetría conviene abandonar la notación $\sqrt[3]{2}$ y llamar $\alpha = \sqrt[3]{2}$, β, γ a las tres raíces. Así

$$x^3 - 2 = (x - \alpha)(x - \beta)(x - \gamma).$$

De este modo, $x^3 - 2$ es el polinomio mínimo en $\mathbb{Q}$ de α, y $x^2 + \alpha x + \alpha^2$ es el polinomio mínimo en $\mathbb{Q}(\alpha)$ de β y γ. Por lo tanto

$$|\mathbb{Q}(\alpha, \beta) : \mathbb{Q}| = |\mathbb{Q}(\alpha, \beta) : \mathbb{Q}(\alpha)||\mathbb{Q}(\alpha) : \mathbb{Q}| = 2 \cdot 3 = 6.$$

Una $\mathbb{Q}$–base de $\mathbb{Q}(\alpha)$ la forman los elementos $1, \alpha, \alpha^2$. Una $\mathbb{Q}(\alpha)$–base de $\mathbb{Q}(\alpha, \beta)$ la forman $1, \beta$, luego una $\mathbb{Q}$–base de $\mathbb{Q}(\alpha, \beta)$ está formada (según la prueba de la transitividad de grados) por los elementos

$$1, \quad \alpha, \quad \alpha^2, \quad \beta, \quad \alpha\beta, \quad \alpha^2\beta. \tag{5.1}$$

Nos falta la expresión de γ en esta base, pero teniendo en cuenta que

$$x^2 + \alpha x + \alpha^2 = (x - \beta)(x - \gamma),$$

resulta que $\alpha = -\beta - \gamma$, luego $\gamma = -\beta - \alpha$.

Ejercicio: Expresar el producto de cada par de elementos de la base (5.1) como combinación lineal de los elementos de dicha base. Notemos que $\beta^2 + \alpha\beta + \alpha^2 = 0$, luego $\beta^2 = -\alpha\beta - \alpha^2$.

Ahora tenemos tres monomorfismos $\sigma_\alpha, \sigma_\beta, \sigma_\gamma : \mathbb{Q}\big(\sqrt[3]{2}\,\big) \longrightarrow \mathbb{Q}(\alpha, \beta)$ que asignan a $\sqrt[3]{2}$ los valores α, β y γ. Con ellos ya podemos definir una norma. Para cada $u \in \mathbb{Q}\big(\sqrt[3]{2}\,\big)$ sea

$$\mathrm{N}(u) = \sigma_\alpha(u)\sigma_\beta(u)\sigma_\gamma(u)$$

y así tenemos una norma multiplicativa como en $\mathbb{Q}(i)$. Vamos a ver que, efectivamente, $\mathrm{N}(u) \in \mathbb{Q}$. El elemento u será de la forma $u = a + b\sqrt[3]{2} + c\sqrt[3]{2}^{\,2}$, para ciertos a, b, $c \in \mathbb{Q}$. Sus conjugados son

$$a + b\alpha + c\alpha^2, \quad a + b\beta + c\beta^2, \quad a + b\gamma + c\gamma^2.$$

Por tanto la norma será

$$(a + b\alpha + c\alpha^2)(a + b\beta + c\beta^2)(a + b\gamma + c\gamma^2).$$

Tras un cálculo no muy complejo (teniendo en cuenta las relaciones que hemos obtenido entre α, β, γ) se obtiene

$$\mathrm{N}(a + b\sqrt[3]{2} + c\sqrt[3]{2}^{\,2}) = a^3 + 2b^3 + 4c^3 - 6abc \in \mathbb{Q}.$$

Más aún, la norma en el anillo $\mathbb{Z}\big[\sqrt[3]{2}\,\big]$ toma valores enteros.

Una muestra de la potencia de la teoría es que no hubiera sido nada fácil probar directamente que esta expresión es multiplicativa, ni mucho menos haber llegado hasta ella sin el auxilio de la teoría de extensiones. ∎

En general vemos que si queremos definir normas en una extensión simple K/k donde no hay suficientes automorfismos, lo que hay que hacer es considerar una extensión mayor L/K que contenga los conjugados del elemento primitivo y definir la norma como el producto de todos los k–monomorfismos $\sigma : K \longrightarrow L$. En las secciones siguientes nos ocuparemos de justificar que este procedimiento siempre funciona.

5.2 Extensiones normales

En esta sección generalizaremos la técnica empleada en el último ejemplo de la sección precedente en virtud de la cual es posible extender un cuerpo hasta otro que contenga todas las raíces de un polinomio dado. Precisamos esta idea mediante la definición siguiente:

Definición 5.17 Sea k un cuerpo y $p(x) \in k[x]$. Sea K una extensión de k. Se dice que $p(x)$ se *escinde* en $K[x]$ si existen elementos $a_0, a_1, \ldots, a_n \in K$ (no necesariamente distintos) tales que $p(x) = a_0(x - a_1)\cdots(x - a_n)$. Notemos que a_0 es necesariamente el coeficiente director de p.

Se dice que K es un *cuerpo de escisión* sobre k del polinomio $p(x)$ si $k \subset K$, el polinomio $p(x)$ se escinde en $K[x]$ y $K = k(a_1, \ldots, a_n)$, donde $a_1, \ldots, a_n$ son las raíces en K de $p(x)$.

Ejemplo Sea $p \geq 3$ un número primo y $\mathbb{Q}[\omega]$ el cuerpo ciclotómico de orden p, es decir, la adjunción a $\mathbb{Q}$ de una raíz del polinomio $c_p(x) = x^{p-1} + \cdots + x + 1$.

Entonces $\omega^p = 1$ y $\omega \neq 1$, lo que implica que todas las potencias ω^i, para $0 \leq 1 < p$ son distintas entre sí (por ejemplo, porque ω es un elemento del grupo multiplicativo $\mathbb{Q}(\omega)^*$ y su orden tiene que dividir a p, luego tiene que ser p). Ahora bien, teniendo en cuenta que $x^p - 1 = (x-1)c_p(x)$, es claro que todas las potencias ω^i para $1 \leq i < p$ son raíces de $c_p(x)$, y no puede haber más, luego

$$c_p(x) = (x - \omega)(x - \omega^2)\cdots(x - \omega^{p-1}),$$

por lo que $\mathbb{Q}[\omega]$ es un cuerpo de escisión sobre $\mathbb{Q}$ del polinomio $c_p(x)$. ∎

Ahora probamos que los cuerpos de escisión son esencialmente únicos:

Teorema 5.18 *Si k es un cuerpo y $p(x) \in k[x]$ es un polinomio no nulo, entonces existe una extensión K/k tal que K es un cuerpo de escisión de p sobre k, y si K' cumple lo mismo, entonces K y K' son k-isomorfos.*

Demostración: Vamos a probar una versión de la unicidad un poco más general: si $f : k \longrightarrow k'$ es un isomorfismo de cuerpos (que se extiende a un isomorfismo $f : k[x] \longrightarrow k'[x]$), $p' = f(p)$ y K' es un cuerpo de escisión de $p'(x)$ sobre k', entonces existe un isomorfismo $g : K \longrightarrow K'$ que extiende a f. La unicidad del enunciado se obtiene tomando como f la identidad en k.

Razonamos por inducción sobre el grado de $p(x)$. Si tiene grado 0 entonces el único cuerpo de escisión es k y se cumple trivialmente el teorema. Supongamos que es cierto para polinomios de grado n y que $p(x)$ tiene grado $n+1$. Por el teorema 3.51 existe una extensión de k donde $p(x)$ tiene una raíz a. Sea $k_1 = k(a)$ y sea $p(x) = (x-a)p_1(x)$, con $p_1(x) \in k_1[x]$. Por hipótesis de inducción p_1 tiene un cuerpo de escisión K sobre k_1.

Esto significa que $p_1(x) = a_0(x-a_1)\cdots(x-a_n)$, con los $a_i \in K$ y además $K = k_1(a_1,\ldots,a_n)$, pero entonces también $p(x) = a_0(x-a)(x-a_1)\cdots(x-a_n)$ y $K = k(a,a_1,\ldots,a_n)$, luego K es un cuerpo de escisión de p sobre k.

Supongamos ahora que tenemos un isomorfismo $f : k \longrightarrow k'$, que $p' = f(p)$ y que K' es un cuerpo de escisión de p' sobre k'. Sea a' una raíz de p' en K' y sea $k'_1 = k'(a')$. El teorema 5.12 nos da que f se extiende a un isomorfismo $f_1 : k_1 \longrightarrow k'_1$ tal que $f(a) = a'$, que a su vez se extiende a un isomorfismo $k_1[x] \longrightarrow k'_1[x]$. Sea $p'_1 = f(p_1)$, de modo que $p'(x) = (x-a')p'_1(x)$. Es claro entonces que K' es un cuerpo de escisión sobre k'_1 de $p'_1(x)$, luego por hipótesis de inducción f_1 se extiende a un isomorfismo $g : K \longrightarrow K'$, que es también una extensión de f. ∎

Así, en el ejemplo final de la sección anterior hemos construido el cuerpo de escisión sobre $\mathbb{Q}$ del polinomio x^3-2, que es una extensión de grado 6 sobre $\mathbb{Q}$.

Ejercicio: Probar que el cuerpo de escisión de un polinomio de grado n sobre un cuerpo k tiene grado sobre k menor o igual que n!

Vemos así que si un polinomio $p(x) \in k[x]$ tiene en un cuerpo un número de raíces menor que su grado (contando cada raíz a tantas veces como el factor lineal $x-a$ divida a $p(x)$) podemos pasar a una extensión K donde tenga exactamente tantas raíces como indica su grado, y entonces podemos decir que $p(x)$ "tiene todas sus raíces en K", en el sentido de que no pueden aparecer más raíces en ninguna extensión de K, ya que el número máximo de raíces es el grado del polinomio. En este sentido, podemos decir que el cuerpo de escisión de un polinomio es el cuerpo donde éste "tiene todas sus raíces". Ahora vamos a probar que los cuerpos de escisión garantizan la presencia de "todas las raíces" de muchos más polinomios. Para ello introducimos el concepto siguiente:

Definición 5.19 Una extensión algebraica K/k es *normal* si cuando un polinomio irreducible $p(x) \in k[x]$ tiene una raíz en K, entonces se escinde en $K[x]$.

Vamos a probar que el cuerpo de escisión de un polinomio $p(x)$ es siempre una extensión normal, con lo cual no sólo contiene a todas las raíces de $p(x)$, sino también las de todos los polinomios con alguna raíz en la extensión, en particular las de todos los polinomios mínimos de sus elementos. Nos apoyamos en el teorema siguiente:

Teorema 5.20 *Sea $k \subset F \subset K \subset L$ una cadena de extensiones algebraicas tal que K sea el cuerpo de escisión de un polinomio de $k[x]$. Sea $\sigma : F \longrightarrow L$ un k-monomorfismo. Entonces $\sigma[F] \subset K$ y σ se extiende a un k-automorfismo de K.*

DEMOSTRACIÓN: Sea $p(x) \in k[x]$ tal que $p(x) = a_0(x - a_1)\cdots(x - a_n)$ y $K = k[a_1, \ldots, a_n]$. Sea $F' = \sigma[F] \subset L$ y sea $K' = F'(a_1, \ldots, a_n)$. Entonces K es un cuerpo de escisión de $p(x)$ sobre F y K' es un cuerpo de escisión de $p(x)$ sobre F'. En la prueba del teorema 5.18 hemos visto que σ se extiende a un isomorfismo $\bar{\sigma} : K \longrightarrow K'$.

Tomemos un $\alpha \in F$. Entonces $\alpha = h(a_1, \ldots, a_n)$, para cierto polinomio $h(x_1, \ldots, x_n) \in k[x_1, \ldots, x_n]$, y resulta que

$$\sigma(\alpha) = \bar{\sigma}(h(a_1, \ldots, a_n)) = h(\bar{\sigma}(a_1), \ldots, \bar{\sigma}(a_n)),$$

pero $\bar{\sigma}$ envía raíces de $p(x)$ en raíces de $p(x)$, luego los elementos $\bar{\sigma}(a_1), \ldots, \bar{\sigma}(a_n)$ son $a_1, \ldots, a_n$, tal vez en otro orden, y así $\sigma(\alpha) \in k(a_1, \ldots, a_n) = K$. Con esto hemos probado que $F' = \sigma[F] \subset K$, pero entonces

$$K = k(a_1, \ldots, a_n) \subset K' = F'(a_1, \ldots, a_n) \subset K,$$

luego $K' = K$, luego $\bar{\sigma} \in G(K/k)$ es una extensión de σ. ∎

Teorema 5.21 *Una extensión finita K/k es normal si y sólo si K es el cuerpo de escisión de un polinomio de $k[x]$.*

DEMOSTRACIÓN: Supongamos que K es el cuerpo de escisión sobre k de un polinomio y sea $p(x) \in k[x]$ un polinomio irreducible con una raíz $a \in K$. Tenemos que probar que $p(x)$ se escinde en $K[x]$.

Sea L un cuerpo de escisión de $p(x)$ sobre K, de modo que $L = K(a_1, \ldots, a_n)$ y $p(x) = a_0(x - a_1)\cdots(x - a_n)$. Como a y a_i son raíces de $p(x)$ en L, existe un k-monomorfismo $\sigma : k(a) \longrightarrow k(a_i)$, y por el teorema anterior $k(a_i) \subset K$, es decir, $a_i \in K$ para todo i, luego $a_1, \ldots, a_n \in K$, luego $p(x)$ se escinde en $K[x]$.

Recíprocamente, si K/k es normal, como es finita, existen $a_1, \ldots, a_n \in K$ tales que $K = k[a_1, \ldots, a_n]$. Sea $p_i = \operatorname{pol\,mín}(a_i, k)$ y sea $p = p_1 \cdots p_n$. Como K/k es normal, cada p_i se escinde en $K[x]$, luego lo mismo le sucede a p, y es claro que al adjuntar a k todas las raíces de p, en particular estamos adjuntando los a_i, luego obtenemos K. Esto prueba que K es un cuerpo de escisión de $p(x)$ sobre k. ∎

Ejercicio: Probar que toda extensión de grado 2 es normal.

Por conveniencia reformulamos 5.20 en términos de extensiones normales:

Teorema 5.22 *Sea $k \subset F \subset K \subset L$ una cadena de extensiones algebraicas tal que K/k sea finita y normal. Sea $\sigma : F \longrightarrow L$ un k-monomorfismo. Entonces $\sigma[F] \subset K$ y σ se extiende a un k-automorfismo de K.*

En particular:

Teorema 5.23 *Sea K/k una extensión finita normal y $a, b \in K$. Entonces a y b son k–conjugados si y sólo si existe un $\sigma \in G(K/k)$ tal que $\sigma(a) = b$.*

DEMOSTRACIÓN: Si a y b son conjugados, por 5.14 existe un k–isomorfismo $\sigma : k(a) \longrightarrow k(b)$, que por el teorema anterior se extiende a un k–automorfismo de K. El recíproco es obvio. ∎

Así pues, en una extensión normal la conjugación está "controlada" por el grupo de Galois. El núcleo de la llamada "teoría de Galois" consiste en obtener información de una extensión normal a partir de su grupo de Galois. En realidad veremos que la teoría requiere una propiedad adicional en la extensión que estudiaremos en la sección siguiente.

Observemos que si $k \subset K \subset L$ y L/k es finita normal, entonces L/K también es normal (pues L es el cuerpo de escisión sobre K del mismo polinomio que sobre k), pero en general la extensión K/k no tiene por qué ser normal, como lo muestra $\mathbb{Q} \subset \mathbb{Q}\big(\sqrt[3]{2}\,\big) \subset \mathbb{Q}(\alpha, \beta)$.

Definición 5.24 Dada una extensión finita K/k, una *clausura normal* de K sobre k es una extensión L/K tal que L/k es normal y no existe ninguna extensión intermedia $k \subset K \subset L' \subset L$ tal que L' sea normal sobre k.

Es claro que toda extensión finita tiene una clausura normal, pues podemos expresar $K = k[a_1, \ldots, a_n]$ y si $p_i = \operatorname{pol\,mín}(a_i, k)$ y $p = p_1 \cdots p_n$, basta tomar como L un cuerpo de escisión de p sobre K. Entonces L también es un cuerpo de escisión de p sobre K y cualquier extensión intermedia L' normal sobre K tiene que contener todas las raíces de p, y entonces tiene que ser L.

Además, dos clausuras normales de una misma extensión K/k son K-isomorfas, pues ambas tienen que ser cuerpos de escisión de p sobre K.

Nuestra intención es definir la norma de un elemento de una extensión finita K/k como el producto de todas sus imágenes por los k–monomorfismos de K en una clausura normal de K/k, tal y como hemos hecho con para definir la norma de $\mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$. Sin embargo, para probar que esto es viable todavía necesitamos estudiar una propiedad adicional de las extensiones algebraicas.

5.3 Extensiones separables

En esta sección vamos a obtener toda la información que necesitamos sobre los monomorfismos de una extensión de cuerpos. La idea básica es que si la normalidad de una extensión nos da propiedades cualitativas importantes acerca de sus automorfismos (que luego resultan aplicables a una extensión finita cualquiera tomando su clausura normal), aquí vamos a introducir otra propiedad, la separabilidad, que nos dará propiedades cuantitativas, es decir, acerca del número de automorfismos. La separabilidad concierne a la multiplicidad de las raíces de los polinomios irreducibles.

En general, si $p(x) \in k[x]$ es un polinomio no constante y K es una extensión de k tal que $p(x)$ se escinde en $K[x]$, tenemos que

$$p(x) = a_0(x - a_1) \cdots (x - a_n),$$

y ahora tenemos que analizar la posibilidad de que algunas raíces aparezcan repetidas en esta descomposición. Agrupando las repeticiones podemos escribir

$$p(x) = a_0(x - a_1)^{r_1} \cdots (x - a_m)^{r_m},$$

donde $r_1 + \cdots + r_m = n$, y los a_i son distintos dos a dos.

Los exponentes r_i están unívocamente determinados por p y K, pues son los exponentes de la descomposición de p en factores primos en $K[x]$, que es un DFU. Diremos que r_i es el *orden de multiplicidad* de la raíz a_i en $p(x)$. Si un elemento $a \in K$ no es raíz de $p(x)$, diremos también que su orden de multiplicidad en $p(x)$ es 0.

Una raíz se dice *simple, doble, triple,* etc. según si su orden de multiplicidad es $1, 2, 3$, etc.

Observemos que si $p(x)$ tiene una raíz a en un cuerpo K, podemos factorizar $p(x) = (x - a)^n g(x)$, con $g(x) \in K[x]$ tal que $g(a) \neq 0$, y entonces n es el orden de multiplicidad de a en $p(x)$ aunque $p(x)$ no se escinda en $K[x]$. La razón es que si pasamos a un cuerpo de escisión L de $g(x)$ sobre K, ninguna de las raíces de $g(x)$ en L será a, puesto que $g(a) \neq 0$, por lo que p se escindirá en $L[x]$ y la multiplicidad de a en la factorización seguirá siendo n.

Otra observación menor es que podemos definir el orden de multiplicidad para polinomios con coeficientes en un dominio íntegro, pues todo dominio íntegro está contenido en su cuerpo de fracciones.

Para estudiar la multiplicidad de las raíces de polinomios es indispensable el concepto de derivada formal:

Definición 5.25 Sea D un dominio íntegro y $f(x) = \sum_{i=0}^{n} a_i x^i \in D[x]$, llamaremos *derivada formal* del polinomio $f(x)$ al polinomio

$$f'(x) = \sum_{i=1}^{n} i a_i x^{i-1} \in D[x].$$

Por ejemplo, si $f(x) = 5x^3 - 2x^2 + 4 \in \mathbb{Q}[x]$, entonces $f'(x) = 15x^2 - 4x$.

Claramente, si $D = \mathbb{R}$ o $D = \mathbb{C}$ la derivada formal de un polinomio coincide con su derivada en el sentido analítico. El teorema siguiente muestra que las reglas usuales del cálculo de derivadas son válidas para las derivadas formales:

Teorema 5.26 *Sea D un dominio íntegro y f, $g \in D[x]$.*

1. *Si $f \in D$ entonces $f' = 0$.*
2. *Si $c \in D$, entonces $(cf)' = cf'$.*
3. $(f + g)' = f' + g'$.
4. $(fg)' = f'g + fg'$.
5. $(f/g)' = (f'g - fg')/g^2$.

DEMOSTRACIÓN: Las tres primeras propiedades son inmediatas. Para probar la cuarta tomamos $f(x) = \sum_{i=0}^{m} a_i x^i$, $g(x) = \sum_{i=0}^{n} b_i x^i$, con lo que, usando las propiedades segunda y tercera,

$$(fg)' = \big(\sum_{ij} a_i b_j x^{i+j}\big)' = \sum_{ij} a_i b_j (x^{i+j})' = \sum_{ij} (i+j) a_i b_j (x^{i+j-1})$$

$$= \sum_{ij} i a_i x^{i-1} b_j x^j + \sum_{ij} a_i x^i j b_j x^{j-1}$$

$$= \big(\sum_i i a_i x^{i-1}\big)\big(\sum_j b_j x^j\big) + \big(\sum_i a_i x^i\big)\big(\sum_j j b_j x^{j-1}\big) = f'g + fg'.$$

En la quinta propiedad se entiende que g divide a f, y basta calcular $(g(f/g))'$ con la regla cuarta y despejar la derivada del cociente. ∎

La relación entre las derivadas y la multiplicidad de las raíces viene dada por el teorema siguiente:

Teorema 5.27 *Sean $D \subset E$ dominios íntegros, $f \in D[x]$ y $c \in E$ tal que $f(c) = 0$. Entonces c es una raíz simple de f si y sólo si $f'(c) \neq 0$.*

DEMOSTRACIÓN: Sea $f(x) = (x-c)^n g(x)$, donde $g(c) \neq 0$ (y $n \geq 1$ es el orden de multiplicidad de c). Entonces $f'(x) = n(x-c)^{n-1} g(x) + (x-c)^n g'(x)$.

Si c es raíz simple de f, entonces $n = 1$, luego $f'(x) = g(x) + (x-c)g'(x)$, y por lo tanto $f'(c) = g(c) + 0 = g(c) \neq 0$. Si c es raíz múltiple, entonces $n > 1$, luego

$$f'(c) = n(c-c)^{n-1} g(c) + (c-c)^n g'(c) = 0.$$ ∎

Este resultado nos lleva a investigar los casos en que la derivada de un polinomio puede anularse. Un caso trivial es el de los polinomios constantes.

Teorema 5.28 *Sea D un dominio íntegro y $f(x) \in D[x]$ un polinomio no constante.*

1. *Si $\operatorname{car} D = 0$ entonces $f'(x) \neq 0$.*
2. *Si $\operatorname{car} D = p$, entonces $f'(x) = 0$ si y sólo si existe $g(x) \in D[x]$ tal que $f(x) = g(x^p)$.*

DEMOSTRACIÓN: 1) Sea $f(x) = \sum_{i=0}^{n} a_i x^i$, con $n > 0$ y $a_n \neq 0$. Entonces $f'(x) = \sum_{i=1}^{n} i a_i x^{i-1}$, donde el coeficiente director es $n a_n \neq 0$, luego $f'(x) \neq 0$.

2) Si $f'(x) = 0$, entonces (con la misma notación del apartado anterior) cada coeficiente $i a_i = 0$, para $i = 1, \ldots, n$. Si $a_i \neq 0$ es necesario que $i = 0$ (en D), es decir, que $p \mid i$. En otras palabras, que los monomios de $f(x)$ con coeficientes no nulos tienen exponente múltiplo de p, es decir,

$$f(x) = \sum_{i=0}^{r} a_i x^{pi} = \sum_{i=0}^{r} a_i (x^p)^i = g(x^p),$$

donde $g(x) = \sum_{i=0}^{r} a_i x^i$. El recíproco es evidente. ∎

Ahora ya sabemos lo necesario para estudiar la separabilidad.

Definición 5.29 Sea K/k una extensión y $a \in K$ un elemento algebraico sobre k. Diremos que a es *separable* sobre k si a es raíz simple de $\operatorname{pol}\operatorname{mín}(a,k)$.

Notemos que si $p(x) = \operatorname{pol}\operatorname{mín}(a,k)$ entonces a es separable si y sólo si $p'(a) \neq 0$ (por 5.27), si y sólo si $p'(x) \neq 0$, pues si $p'(a) = 0$ entonces $p(x) \mid p'(x)$ y, teniendo en cuenta los grados, $p'(x) = 0$. Esta última condición depende sólo de p, luego si a es separable todos sus conjugados lo son también, y $\operatorname{pol}\operatorname{mín}(a,k)$ tiene todas sus raíces simples.

Una extensión K/k es *separable* si todos los elementos de K son separables sobre k (en particular una extensión separable es algebraica). Un cuerpo k es *perfecto* si todas sus extensiones algebraicas son separables.

El interés de esta definición reside en que prácticamente todos los cuerpos que nos van a interesar son perfectos, luego en la práctica todas las extensiones que manejaremos serán separables.

Teorema 5.30 *Se cumple:*

1. *Todo cuerpo de característica* 0 *es perfecto.*
2. *Un cuerpo k de característica p es perfecto si y sólo si*
$$k = k^p = \{a^p \mid a \in k\}.$$
3. *Todo cuerpo finito es perfecto.*

Demostración: 1) Si $\operatorname{car} k = 0$ y K/k es una extensión algebraica, sea $a \in K$. Entonces el polinomio $p(x) = \operatorname{pol}\operatorname{mín}(a,k)$ no es constante y $p'(x) \neq 0$, luego a es separable.

2) Si $\operatorname{car} k = p$ y $k = k^p$, sea K/k una extensión algebraica y $a \in K$, sea $p(x) = \operatorname{pol}\operatorname{mín}(a,k)$. Si a no fuera separable, entonces $p'(x) = 0$, luego por 5.28 $p(x) = f(x^p)$ para cierto polinomio $f(x) \in k[x]$.

Sea $f(x) = \sum_{i=0}^{n} a_i x^i$. Como $a_0, \ldots, a_r \in k = k^p$, existen $b_0, \ldots, b_r \in k$ de manera que $a_i = b_i^p$. Por lo tanto

$$p(x) = \sum_{i=0}^{n} b_i^p x^{pi} = \sum_{i=0}^{n} \big(b_i x^i\big)^p = \big(\sum_{i=0}^{n} b_i x^i\big)^p,$$

donde el polinomio $\sum_{i=0}^{n} b_i x^i \in k[x]$. Por lo tanto $p(x)$ no es irreducible en $k[x]$, contradicción.

Ahora supongamos que k es perfecto y $\operatorname{car} k = p$. Veamos que $k = k^p$.

Sea $a \in k$. Por 3.51 existe una extensión K de k donde $x^p - a$ tiene una raíz b. Entonces $b^p - a = 0$, o sea, $a = b^p$.

El polinomio $x^p - a = x^p - b^p = (x-b)^p$, y por otro lado tenemos que $\operatorname{pol}\operatorname{mín}(b,k) \mid (x-b)^p$, luego ha de ser $\operatorname{pol}\operatorname{mín}(b,k) = (x-b)^n$, para cierto $n \leq p$, pero como k es perfecto b es raíz simple de $\operatorname{pol}\operatorname{mín}(b,k)$, es decir, $n = 1$, luego $x - b = \operatorname{pol}\operatorname{mín}(b,k) \in k[x]$. Por consiguiente $b \in k$ y $a = b^p \in k^p$.

3) Si k es un cuerpo finito de característica p, entonces la aplicación $\phi : k \longrightarrow k^p$ dada por $\phi(a) = a^p$ es suprayectiva, pero también inyectiva, puesto que si $a^p = b^p$, entonces $a^p - b^p = (a-b)^p = 0$, luego $a - b = 0$, o sea, $a = b$.

Por lo tanto $k^p \subset k$ tienen ambos el mismo cardinal, lo que obliga a que $k = k^p$. ∎

Como una primera muestra del interés de la separabilidad, vamos a ver el efecto que tiene combinarla con la normalidad. Para ello introducimos algunos conceptos.

Definición 5.31 Sea K/k una extensión. Definimos su *cuerpo fijado* como

$$F = \big\{a \in K \mid \sigma(a) = a \text{ para todo } \sigma \in G(K/k)\big\}.$$

Es claro que efectivamente F es un cuerpo y $k \subset F \subset K$.

Una extensión *de Galois* es una extensión normal y separable. El teorema siguiente nos da un importante criterio para determinar cuándo un elemento de una extensión finita de Galois pertenece de hecho al cuerpo base:

Teorema 5.32 *Una extensión finita K/k es de Galois si y sólo si su cuerpo fijado es k.*

Demostración: Sea K/k una extensión finita de Galois. Sea a un elemento de su cuerpo fijado y sea $p(x)$ su polinomio mínimo sobre k.

Por la normalidad, $p(x)$ tiene todas sus raíces en K. Si b es cualquiera de ellas, ha de existir un $\sigma \in G(K/k)$ tal que $\sigma(a) = b$ (teorema 5.23). Pero a está en el cuerpo fijado, luego $b = a$, o sea, a es la única raíz de $p(x)$, que ha de ser, pues, $p(x) = (x-a)^n$. Pero por otra parte a es separable sobre k, luego ha de ser una raíz simple de $p(x)$. Así pues $p(x) = x - a \in k[x]$ y por lo tanto $a \in k$.

Supongamos ahora que k es el cuerpo fijado de la extensión. Veamos que K/k es normal. Sea $p(x) \in k[x]$ un polinomio irreducible (que podemos suponer mónico) con una raíz $a \in K$. Hemos de ver que $p(x)$ se escinde en $K[x]$.

Sean $a_1, \ldots, a_n$ todas las raíces de $p(x)$ en K (sin repeticiones). Sea

$$g(x) = (x - a_1) \cdots (x - a_n) \in K[x].$$

Para cada $\sigma \in G(K/k)$ se cumple que $\sigma g(x) = \big(x - \sigma(a_1)\big) \cdots \big(x - \sigma(a_n)\big)$, pero es obvio que $\sigma(a_1), \ldots, \sigma(a_n)$ son los mismos $a_1, \ldots, a_n$ cambiados de orden, luego en realidad $\sigma g(x) = g(x)$. Esto significa que todos los coeficientes de $g(x)$ son fijados por σ, luego están en el cuerpo fijado de K/k, que por hipótesis es k, o sea, $g(x) \in k[x]$.

Como $p(x)$ es el polinomio mínimo de a y a es una de las raíces de $g(x)$, tenemos que $p(x) \mid g(x)$, pero por otra parte todas las raíces de $g(x)$ lo son de $p(x)$ y además son simples, luego $g(x) \mid p(x)$. Como ambos son mónicos $p(x) = g(x)$ se escinde en $K[x]$ y además con raíces simples.

Con esto hemos probado también que K/k es separable, pues dado $a \in K$, si tomamos $p(x) = \operatorname{pol\,mín}(a, k)$ hemos probado que las raíces de $p(x)$ son simples. ∎

Comenzamos ahora la labor de 'contar' los automorfismos de una extensión finita. En realidad si la extensión no es normal puede no haber más automorfismo que la identidad (como le ocurre a $\mathbb{Q}(\sqrt[3]{2}\,)/\mathbb{Q}$), luego si queremos resultados generales no hemos de contar automorfismos, sino monomorfismos.

Definición 5.33 Sea K/k una extensión finita. Llamaremos *k–monomorfismos* de K a los k–monomorfismos de K en una clausura normal de K sobre k.

En particular, si K/k es normal entonces los k–monomorfismos de K son de hecho los k–automorfismos de K.

Llamaremos $\mathrm{N}(K/k)$ al número de k–monomorfismos de K (que no depende de la elección de la clausura normal de K/k, pues dos cualesquiera son K-isomorfas, luego todo k-monomorfismo en una se corresponde con un k-monomorfismo en otra).

Por ejemplo, si $k(a)/k$ es una extensión simple de grado n sabemos que el número de k–monomorfismos de K es igual al número de k–conjugados de a, y si a es separable sobre k entonces $\operatorname{pol\,mín}(a, k)$ tiene todas sus raíces simples, luego a tiene exactamente n k–conjugados, es decir, el número de k–monomorfismos de $k(a)$ es igual al grado de la extensión. Equivalentemente:

$$\mathrm{N}(k(a)/k) = |k(a) : k|.$$

Vamos a generalizar este hecho a extensiones separables cualesquiera. Primero probamos un resultado técnico:

Teorema 5.34 *Consideremos una cadena de extensiones $k \subset K \subset L$ con L/k finita normal. Entonces* $\mathrm{N}(L/k) = \mathrm{N}(L/K)\,\mathrm{N}(K/k)$.

Demostración: Por 5.22 tenemos que cada k-monomorfismo de K se extiende a un k-automorfismo de L. Basta probar que de hecho se extiende exactamente a $\mathrm{N}(L/K)$ de ellos. Notemos que $\mathrm{N}(L/K) = |G(L/K)|$ porque la extensión L/K es normal.

Sean σ y τ dos extensiones a L de un mismo k–monomorfismo de K. Como L/k es normal σ y τ son k–automorfismos de L, luego $\tau\sigma^{-1}$ es un k–automorfismo de L que de hecho fija a K, es decir, $\rho = \tau\sigma^{-1} \in G(L/K)$ y $\tau = \sigma\rho$. Así pues, si σ es una extensión cualquiera de un k–monomorfismo de K, las restantes son de la forma $\sigma\rho$ con $\rho \in G(L/K)$.

Por otra parte, si $\sigma\rho = \sigma\rho'$, componiendo con σ^{-1} concluimos que $\rho = \rho'$, luego en efecto, hay tantas extensiones como elementos de $G(L/K)$ ∎

Sea ahora una extensión finita normal $L = k(a_1, \ldots, a_n)$, donde los elementos a_i son separables sobre k. Podemos aplicar el teorema anterior con $K = k(a_1)$ y concluir que $\mathrm{N}(L/k) = \mathrm{N}\big(L/k(a_1)\big)\ |k(a_1) : k|$. Ahora consideramos la cadena $k(a_1) \subset k(a_1, a_2) \subset L$. Es claro que está también en las hipótesis del teorema anterior (a_2 es separable sobre $k(a_1)$ porque $\operatorname{pol\,mín}\big(a_2, k(a_1)\big) \mid \operatorname{pol\,mín}(a_2, k)$).

Por lo tanto concluimos que

$$\begin{aligned} \mathrm{N}(L/k) &= \mathrm{N}\big(L/k(a_1,a_2)\big)\ |k(a_1,a_2):k(a_1)|\ |k(a_1):k| \\ &= \mathrm{N}\big(L/k(a_1,a_2)\big)\ |k(a_1,a_2):k|. \end{aligned}$$

Repitiendo el proceso llegamos a $\mathrm{N}(L/k) = |L:k|$.

Esto es esencialmente el resultado al que queremos llegar (y, como vemos, es una mera generalización del caso trivial de extensiones simples). Notemos ahora que si F es el cuerpo fijado de la extensión L/k entonces $G(L/k) = G(L/F)$, luego $|L:k| = |G(L/k)| = |G(L/F)| = |L:F|$, con lo que $k = F$ y el teorema 5.32 nos permite concluir que la extensión L/k es separable. De aquí se sigue en primer lugar:

Teorema 5.35 *Si $K = k(S)$, donde S es un conjunto de elementos separables sobre k, entonces la extensión K/k es separable.*

Demostración: Supongamos primero que S es finito. Sean $a_1, \dots, a_n$ los conjugados de todos los elementos de S. Sabemos que todos ellos son separables sobre k. Entonces $L = k(a_1, \dots, a_n)$ es la clausura normal de K/k, y en estas condiciones hemos probado que L/k es separable, luego K/k también. El caso infinito se reduce trivialmente al caso finito. ∎

Notemos que acabamos de probar que la clausura normal de una extensión finita separable es una extensión finita de Galois. Finalmente estamos en condiciones de enunciar el teorema principal de esta sección:

Teorema 5.36 *Si K/k es una extensión finita separable de grado n entonces el número de k–monomorfismos de K es exactamente igual a n. En particular si K/k es finita de Galois $|G(K/k)| = |K:k|$.*

Demostración: Lo tenemos probado para extensiones normales. Si K/k no es normal tomamos la clausura normal L/k. Entonces las extensiones L/k y L/K son normales, luego el teorema 5.34 y el caso normal nos dan $|L:k| = \mathrm{N}(K/k)\ |L:K|$, luego también $\mathrm{N}(K/k) = |K:k|$. ∎

El teorema siguiente termina de perfilar el comportamiento de las extensiones separables, totalmente análogo al de las algebraicas en general:

Teorema 5.37 *Sea $k \subset K \subset L$ una cadena de extensiones. Entonces L/k es separable si y sólo si L/K y K/k lo son.*

Demostración: Una implicación es sencilla. La otra se reduce fácilmente al caso en que las extensiones son finitas (ver por ejemplo 5.9).

Supongamos, pues, que L/K y K/k son finitas separables. Si $a \in L$ y b es un k–conjugado de a, entonces existe un k–monomorfismo de K tal que $\sigma(a) = b$.

Si $p(x) = \mathrm{pol\,mín}(a, K)$ entonces $\mathrm{pol\,mín}(b, K) = \sigma p(x)$, y es claro que si a es una raíz simple de $p(x)$ entonces b es raíz simple de $\sigma p(x)$, es decir, los k–conjugados de elementos de L son separables sobre K. De aquí se sigue que la clausura normal de L sobre k es separable sobre K, luego podemos suponer que L/k es normal.

La clausura normal de K/k está contenida en L, es separable y obviamente L es separable sobre ella (por la implicación opuesta a la que estamos probando). Por lo tanto podemos suponer también que K/k es normal.

Veamos que el cuerpo fijado de L/k es k. Si a está en dicho cuerpo fijado, en particular a es fijado por $G(L/K)$, luego $a \in K$ (por 5.32). Todo automorfismo de K/k se extiende a un automorfismo de L/k que fija a a, luego a está en el cuerpo fijado de K/k, que es k. ∎

Ejercicio: Probar que si K/k es una extensión de cuerpos, el conjunto K_s de los elementos de K separables sobre k es un subcuerpo de K.

Terminamos esta sección con un teorema no trivial según el cual en muchos casos los esfuerzos por generalizar el caso de extensiones simples a extensiones finitas arbitrarias es innecesario:

Teorema 5.38 (Teorema del elemento primitivo) *Toda extensión finita separable es simple.*

Demostración: Sea K/k una extensión finita separable. Distingamos dos casos según que el cuerpo base k sea finito o infinito. Si k es finito y K es una extensión finita de k, entonces K también es un cuerpo finito (cada elemento de K está determinado por sus coordenadas en una k-base, y sólo hay un número finito de coordenadas posibles). Por el teorema 4.50 el grupo multiplicativo K^* es cíclico, es decir, existe un elemento $a \in K$ tal que sus potencias recorren todos los elementos de K salvo el cero. Obviamente, $K = k(a)$.

Supongamos ahora que k es infinito. Toda extensión finita es finitamente generada, es decir, es de la forma $k(a_1, \ldots, a_n)/k$.

Razonando por inducción es suficiente probar que si a, b son elementos separables sobre un cuerpo k entonces existe un $c \in k(a, b)$ tal que $k(a, b) = k(c)$.

Sea A el conjunto de todos los pares (a', b'), donde a' es un k–conjugado de a y b' es un k–conjugado de b. Es claro que si (a_1, b_1), (a_2, b_2) son dos pares distintos en A, existe a lo sumo un $u \in k$ tal que $a_1 + ub_1 = a_2 + ub_2$. Así pues, como A es finito y k es infinito existe un elemento $v \in k$ distinto de cero y para el que $a_1 + vb_1 \neq a_2 + vb_2$, para todo par de pares distintos (a_1, b_1), $(a_2, b_2) \in A$.

Sea $c = a + vb$. Entonces, si σ, τ son k–monomorfismos distintos de $k(a, b)$ los pares $\big(\sigma(a), \sigma(b)\big)$ y $\big(\tau(a), \tau(b)\big)$ son pares distintos de A, luego

$$\sigma(c) = \sigma(a) + v\sigma(b) \neq \tau(a) + v\tau(b) = \tau(c).$$

Esto significa que c tiene tantos conjugados como k–monomorfismos tiene la extensión $k(a, b)/k$. Por los resultados que hemos probado, el grado de $\operatorname{pol\,mín}(c, k)$ coincide con $|k(a, b) : k|$ o también $|k(c) : k| = |k(a, b) : k|$. Puesto que $k(c) \subset k(a, b)$, de hecho $k(a, b) = k(c)$. ∎

Así pues, para reducir dos generadores a, b de una extensión separable a uno solo, hemos de tomar un elemento de la forma $c = a + vb$, con v en el cuerpo base. La finalidad de v es simplemente romper la simetría de la expresión para que el número de automorfismos sea el máximo posible. Unos ejemplos aclararán esta idea.

Ejemplo Consideremos el cuerpo $K = \mathbb{Q}\big(\sqrt{2}, \sqrt{3}\,\big)$. La extensión $K/\mathbb{Q}$ es normal, pues K es el cuerpo de escisión sobre $\mathbb{Q}$ del polinomio $(x^2-2)(x^2-3)$. Trivialmente es separable, pues los cuerpos son de característica 0. Así pues, $K/\mathbb{Q}$ es una extensión finita de Galois. Veamos que tiene grado 4. Para ello consideramos la cadena

$$\mathbb{Q} \subset \mathbb{Q}\big(\sqrt{2}\,\big) \subset \mathbb{Q}\big(\sqrt{2}, \sqrt{3}\,\big).$$

Es claro que el grado del primer tramo es 2, luego basta probar que el segundo tramo también tiene grado 2. A su vez esto equivale a que $\sqrt{3} \notin \mathbb{Q}\big(\sqrt{2}\,\big)$. Ahora bien, es fácil ver que la ecuación

$$\big(a + b\sqrt{2}\,\big)^2 = a^2 + 2b^2 + 2ab\sqrt{2} = 3$$

no tiene solución para $a, b \in \mathbb{Q}$.

Los conjugados de $\sqrt{2}$ son $\pm\sqrt{2}$ y los conjugados de $\sqrt{3}$ son $\pm\sqrt{3}$. Si llamamos $S = \{\pm\sqrt{2}, \pm\sqrt{3}\}$, tenemos que cada elemento de $G(K/\mathbb{Q})$ está determinado por su restricción a S, es decir, que la aplicación $G(K/\mathbb{Q}) \longrightarrow \Sigma_S$ dada por $\sigma \mapsto \sigma|_S$ es un monomorfismo de grupos. Como, además, la imagen de $\sqrt{2}$ sólo puede ser $\pm\sqrt{2}$ y la de $\sqrt{3}$ sólo puede ser $\pm\sqrt{3}$, vemos que sólo hay cuatro permutaciones posibles en la imagen, que expresadas como productos de ciclos disjuntos son:

$$1, \quad (\sqrt{2}, -\sqrt{2}), \quad (\sqrt{3}, -\sqrt{3}), \quad (\sqrt{2}, -\sqrt{2})(\sqrt{3}, -\sqrt{3}).$$

Como la imagen tiene que tener cuatro permutaciones, tienen que ser éstas. Es claro entonces que $G(K/\mathbb{Q}) \cong C_2 \times C_2$.

Una $\mathbb{Q}$–base de K es 1, $\sqrt{2}$, $\sqrt{3}$, $\sqrt{2}\cdot\sqrt{3} = \sqrt{6}$. Un elemento primitivo es $\sqrt{2} + \sqrt{3}$, pues al aplicar los cuatro automorfismos obtenemos los conjugados

$$\sqrt{2} + \sqrt{3}, \quad -\sqrt{2} + \sqrt{3}, \quad \sqrt{2} - \sqrt{3}, \quad -\sqrt{2} - \sqrt{3}.$$

Los cuatro son, efectivamente, distintos porque sus coordenadas en la base dada son distintas. Por lo tanto $\mathbb{Q}\big(\sqrt{2} + \sqrt{3}\,\big)$ tiene grado 4 sobre $\mathbb{Q}$ y está contenido en K, luego es K.

Ejercicio: Calcular $\operatorname{pol\,mín}(\sqrt{2} + \sqrt{3}, \mathbb{Q})$.

Ejemplo Consideremos de nuevo $K = \mathbb{Q}(\alpha, \beta)$, donde α y β son dos de las raíces del polinomio $x^3 - 2$. Según hemos visto en la sección anterior, la extensión $K/\mathbb{Q}$ es de Galois y tiene grado 6. La imagen de α y β por cada uno de los seis automorfismos de K ha de ser α, β o la tercera raíz γ. Si llamamos $S = \{\alpha, \beta, \gamma\}$, la restricción $\sigma \mapsto \sigma|_S$ es un monomorfismo $G(K/\mathbb{Q}) \longrightarrow \Sigma_3$, y como ambos grupos tienen orden 6, tiene que ser un isomorfismo. Así pues $G(K/\mathbb{Q}) \cong \Sigma_3$ y los seis automorfismos permutan α, β, γ de todas las formas posibles.

En este caso el elemento $\alpha + \beta$ no es un elemento primitivo de $K/\mathbb{Q}$, pues al aplicarle los seis automorfismos obtenemos sólo tres conjugados: $\alpha + \beta$, $\alpha + \gamma$, $\beta + \gamma$. Por lo tanto $\operatorname{pol\,mín}(\alpha + \beta, \mathbb{Q})$ tiene grado 3 y $\mathbb{Q}(\alpha + \beta)$ es un cuerpo

intermedio de grado 3 sobre $\mathbb{Q}$. Un elemento primitivo es, por ejemplo $\alpha - \beta$. En efecto, sus conjugados son

$$\begin{array}{ll} \alpha-\beta, & \beta-\gamma = \quad \alpha+2\beta,\\ \alpha-\gamma = 2\alpha+\beta, & \gamma-\alpha = -2\alpha - \ \beta,\\ \beta-\alpha, & \gamma-\beta = -\ \alpha-2\beta, \end{array}$$

y son todos distintos por la independencia lineal de α y β. Puede comprobarse que $\operatorname{pol\,mín}(\alpha-\beta) = x^6 + 108$. Notemos que la presencia del -1 se traduce en una pérdida de la simetría de la expresión $\alpha+\beta$, que hace que las las seis imágenes por los automorfismos sean distintas. La prueba del teorema del elemento primitivo justifica que siempre podemos conseguir el máximo número posible de conjugados mediante esta técnica. ∎

5.4 Normas y trazas

Finalmente podemos definir en general la norma de una extensión de cuerpos. Notemos primero que el teorema 5.32 vale en parte para extensiones separables (no necesariamente normales): Si K/k es separable, un elemento $u \in K$ está en k si y sólo si $\sigma(u) = u$ para todo k-monomorfismo σ de K. En efecto, si consideramos la clausura normal L/k, tenemos que $u \in k$ si y sólo si $\sigma(u) = u$ para todo $\sigma \in G(L/k)$, pero las restricciones a K de los k-automorfismos de L son los k-monomorfismos de K.

Definición 5.39 Sea K/k una extensión separable de grado n. Sean $\sigma_1, \ldots, \sigma_n$ los k-monomorfismos de K (en la clausura normal L de K/k). Definimos la *norma* y la *traza* de un $\alpha \in K$ como

$$\mathrm{N}(\alpha) = \sigma_1(\alpha)\cdots\sigma_n(\alpha) \in L, \qquad \mathrm{Tr}(\alpha) = \sigma_1(\alpha) + \cdots + \sigma_n(\alpha) \in L.$$

Si $\sigma \in G(L/k)$, entonces $\sigma_i \circ \sigma$ es un k-monomorfismo de K en L, luego $\sigma_i \circ \sigma = \sigma_j$ para algún j. Más aún, si $i \neq j$, entonces $\sigma_i \circ \sigma \neq \sigma_j \circ \sigma$, pues actúan de forma distinta sobre un elemento primitivo a de K.

Por lo tanto, la composición con σ permuta los monomorfismos y, en consecuencia, $\sigma\big(\mathrm{N}(\alpha)\big) = \mathrm{N}(\alpha)$, $\sigma\big(\mathrm{Tr}(\alpha)\big) = \mathrm{Tr}(\alpha)$ para todo α y todo σ, es decir, $\mathrm{N}(\alpha)$, $\mathrm{Tr}(\alpha) \in k$.

Tenemos así dos aplicaciones $\mathrm{N}, \mathrm{Tr} : K \longrightarrow k$. Es obvio que

$$\mathrm{N}(uv) = \mathrm{N}(u)\,\mathrm{N}(v), \qquad \mathrm{Tr}(u+v) = \mathrm{Tr}(u) + \mathrm{Tr}(v).$$

De hecho la traza es una aplicación lineal de k-espacios vectoriales. Una propiedad elemental es que si $\alpha \in k$, entonces

$$\mathrm{N}(\alpha) = \alpha^{|K:k|}, \qquad \mathrm{Tr}(\alpha) = |K:k|\,\alpha.$$

Teorema 5.40 (Transitividad de la norma) *Sea $k \subset K \subset L$ una cadena de extensiones finitas separables. Entonces para todo $\alpha \in L$ se cumple*

$$\mathrm{N}_k^L(\alpha) = \mathrm{N}_k^K\big(\mathrm{N}_K^L(\alpha)\big), \qquad \mathrm{Tr}_k^L(\alpha) = \mathrm{Tr}_k^K\big(\mathrm{Tr}_K^L(\alpha)\big).$$

DEMOSTRACIÓN: Sean $\sigma_1, \ldots, \sigma_n$ los K-monomorfismos de L. Así

$$\mathrm{N}_K^L(\alpha) = \sigma_1(\alpha) \cdots \sigma_n(\alpha).$$

Sean $\tau_1, \ldots, \tau_m$ los k-monomorfismos de K. Escojamos para cada uno de ellos τ_i una extensión a un k-automorfismo de la clausura normal de L sobre k, digamos ρ_i. Entonces $\tau_i\left(\mathrm{N}_K^L(\alpha)\right) = \rho_i\left(\sigma_1(\alpha)\right) \cdots \rho_i\left(\sigma_n(\alpha)\right)$, y $\mathrm{N}_k^K\left(\mathrm{N}_K^L(\alpha)\right)$ es el producto de estos términos para $i = 1, \ldots, m$.

Los monomorfismos $\sigma_j \circ \rho_i$, definidos sobre L, son distintos dos a dos, pues si $\sigma_j \circ \rho_i = \sigma_u \circ \rho_v$, restringiendo a K tenemos $\rho_i = \rho_v$, luego $i = v$, y componiendo con el automorfismo inverso queda $\sigma_j = \sigma_u$, luego $j = u$.

Como en total son mn monomorfismos, de hecho son todos los k-monomorfismos de L, luego $\mathrm{N}_k^K\left(\mathrm{N}_K^L(\alpha)\right)$, que es el producto de todos los $\rho_i\left(\sigma_j(\alpha)\right)$, es igual a $\mathrm{N}_k^L(\alpha)$.

El mismo razonamiento vale para las trazas. ∎

Ejemplo Vamos a calcular la norma de la extensión $\mathbb{Q}\big(\sqrt{2}, \sqrt{3}\,\big)/\mathbb{Q}$.

Un elemento arbitrario de $\mathbb{Q}\big(\sqrt{2}, \sqrt{3}\,\big)$ es de la forma

$$\alpha = a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}, \quad \text{con } a, b, c, d \in \mathbb{Q}.$$

Su norma en la extensión $\mathbb{Q}\big(\sqrt{2}, \sqrt{3}\,\big)/\mathbb{Q}\big(\sqrt{2}\,\big)$ es

$$\begin{aligned} &\left(a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}\right)\left(a + b\sqrt{2} - c\sqrt{3} - d\sqrt{6}\right) \\ &= \left(a + b\sqrt{2}\right)^2 - \left(c\sqrt{3} + d\sqrt{6}\right)^2 \\ &= a^2 + 2b^2 + 2ab\sqrt{2} - 3c^2 - 6d^2 - 6cd\sqrt{2}, \end{aligned}$$

y la norma de este número en la extensión $\mathbb{Q}\big(\sqrt{2}\,\big)/\mathbb{Q}$ es

$$\begin{aligned} \mathrm{N}(\alpha) &= \left(a^2 + 2b^2 - 3c^2 - 6d^2 + (2ab - 6cd)\sqrt{2}\right) \\ &\cdot \left(a^2 + 2b^2 - 3c^2 - 6d^2 - (2ab - 6cd)\sqrt{2}\right) \\ &= \left(a^2 + 2b^2 - 3c^2 - 6d^2\right)^2 - 2\left(2ab - 6cd\right)^2. \end{aligned}$$

∎

Ejercicio: Calcular la traza de las extensiones $\mathbb{Q}(\sqrt{2}, \sqrt{3}\,)/\mathbb{Q}$ y $\mathbb{Q}(\sqrt[3]{2}\,)/\mathbb{Q}$.

Ejemplo Consideramos ahora la extensión ciclotómica p-ésima $\mathbb{Q}(\omega)/\mathbb{Q}$ para un primo impar p. Es normal y su grado es $p-1$, luego tiene $p-1$ automorfismos, determinados por la imagen que toman sobre el elemento primitivo ω, que ha de ser uno de sus conjugados ω^i para $i = 1, \ldots, p-1$. Evaluando en 0 el polinomio

$$x^{p-1} + \cdots + x + 1 = (x - \omega)(x - \omega^2) \cdots (x - \omega^{p-1}),$$

obtenemos $\mathrm{N}(\omega) = 1$. Como la norma conserva productos se cumple $\mathrm{N}(\omega^i) = 1$ para todo i.

Evaluando en 1 el mismo polinomio queda

$$\mathrm{N}(1-\omega) = (1-\omega)(1-\omega^2)\cdots(1-\omega^{p-1}) = 1^{p-1} + \cdots + 1 + 1 = p.$$

Si $p \nmid i$, entonces $\mathrm{Tr}(\omega^i)$ es la suma de los $p-1$ conjugados de ω^i, es decir,

$$\mathrm{Tr}(\omega^i) = \omega + \omega^2 + \cdots + \omega^{p-1} = -1.$$

Si $a \in \mathbb{Q}$ entonces $\mathrm{Tr}(a) = a + a + \cdots + a = (p-1)a$. En resumen,

$$\mathrm{Tr}(\omega^i) = \begin{cases} -1 & \text{si } p \nmid i \\ p-1 & \text{si } p \mid i \end{cases}$$

Una ventaja de la traza frente a la norma es que es mucho más fácil de calcular. En efecto, si $\sum_{i=0}^{p-1} a_i\omega^i$ es un elemento cualquiera de $\mathbb{Q}(\omega)$, entonces

$$\begin{aligned} \mathrm{Tr}\left(\sum_{i=0}^{p-1} a_i\omega^i\right) &= \sum_{i=0}^{p-1} a_i \,\mathrm{Tr}(\omega^i) = a_0\,\mathrm{Tr}(1) - \sum_{i=1}^{p-1} a_i \\ &= (p-1)a_0 - \sum_{i=1}^{p-1} a_i = pa_0 - \sum_{i=0}^{p-1} a_i. \end{aligned}$$

■

5.5 La teoría de Galois

Presentamos ahora el teorema fundamental que permite obtener consecuencias relevantes sobre extensiones de Galois a través del estudio de sus grupos de automorfismos. Básicamente afirma que los cuerpos intermedios de una extensión finita de Galois se corresponden de forma natural con los subgrupos del grupo de Galois. Más concretamente, la correspondencia de Galois asigna a cada subgrupo el cuerpo que definimos a continuación:

Definición 5.41 Sea K/k una extensión de cuerpos y H un subgrupo del grupo de Galois $G(K/k)$. Llamaremos *cuerpo fijado* por H al cuerpo

$$F(H) = \big\{a \in K \mid \sigma(a) = a \text{ para todo } \sigma \in H\big\}.$$

Es muy fácil probar que ciertamente $F(H)$ es un cuerpo y $k \subset F(H) \subset K$.

El teorema 5.32 afirma que una extensión algebraica K de un cuerpo k es de Galois si y sólo si $F\big(G(K/k)\big) = k$. El teorema siguiente contiene la parte técnica del teorema de Galois y resulta útil por sí mismo en algunas ocasiones:

Teorema 5.42 (Teorema de independencia de Dedekind) *Consideremos un cuerpo K y sean $\sigma_1, \ldots, \sigma_n$ automorfismos distintos de K. Si $c_1, \ldots, c_n$ son elementos de K tales que $\sum_{i=1}^{n} c_i\sigma_i(a) = 0$ para todo $a \in K$, entonces $c_1 = \cdots = c_n = 0$.*

DEMOSTRACIÓN: Por inducción sobre n. Si $n = 1$, entonces $c_1\sigma_1(a) = 0$ para todo $a \in K$, luego en particular $c_1\sigma_1(1) = 0$, es decir, $c_1 = 0$. Supongamos que $n > 1$ y que

$$\sum_{i=1}^{n} c_i\sigma_i(a) = 0 \qquad \text{para todo } a \in K. \tag{5.2}$$

Si algún c_i es nulo entonces todos lo son por hipótesis de inducción. Supongamos que son todos no nulos. Como $\sigma_1 \neq \sigma_n$, existe un elemento $b \in K$ tal que $\sigma_1(b) \neq \sigma_n(b)$. Obviamente $b \neq 0$. Por hipótesis,

$$\sum_{i=1}^{n} c_i\sigma_i(ba) = 0 \quad \text{para todo } a \in K.$$

Multiplicando por $\sigma_n(b^{-1})$ y restando de (5.2) queda:

$$\sum_{i=1}^{n} c_i\big(1 - \sigma_n(b^{-1})\sigma_i(b)\big)\sigma_i(a) = 0 \quad \text{para todo } a \in K.$$

Como el último sumando es nulo, podemos aplicar la hipótesis de inducción y concluir que $c_i\big(1 - \sigma_n(b^{-1})\sigma_i(b)\big) = 0$ para $i = 1, \ldots, n-1$. En particular $\big(1 - \sigma_n(b^{-1})\sigma_1(b)\big) = 0$ y $\sigma_1(b) = \sigma_n(b)$, contradicción. ∎

Con la ayuda de este resultado podemos probar el teorema siguiente, que esencialmente contiene la biyectividad de la correspondencia de Galois:

Teorema 5.43 *Sea K/k una extensión y H un subgrupo finito de $G(K/k)$. Entonces*

$$|K : F(H)| = |H|.$$

DEMOSTRACIÓN: Supongamos que $|K : F(H)| = r < |H| = n$. Sea $b_1, \ldots, b_r$ una base de K como $F(H)$-espacio vectorial. Sea $H = \{\sigma_1, \ldots, \sigma_n\}$. La aplicación $f : K^n \longrightarrow K^r$ dada por

$$f(x_1, \ldots, x_n) = \big(\sum_{i=1}^{n} x_i\sigma_i(b_1), \ldots, \sum_{i=1}^{n} x_i\sigma_i(b_r)\big)$$

es claramente lineal y, como $n = \dim \mathrm{N}(f) + \dim \mathrm{Im}\, f \leq \dim \mathrm{N}(f) + r$, concluimos que $\dim \mathrm{N}(f) > 0$, luego existe una n-tupla $(c_1, \ldots, c_n)$ de elementos de K no todos nulos de modo que $\sum_{i=1}^{n} c_i\sigma_i(b_j) = 0$ para $j = 1, \ldots, r$.

Si $a \in K$, entonces $a = a_1b_1 + \cdots + a_rb_r$ para ciertos $a_1, \ldots, a_r \in F(H)$. Como son fijados por los automorfismos de H se cumple que

$$\sum_{i=1}^{n} c_i\sigma_i(a_jb_j) = a_j \sum_{i=1}^{n} c_i\sigma_i(b_j) = 0,$$

y sumando para todos los j obtenemos que $\sum_{i=1}^{n} c_i\sigma_i(a) = 0$ para todo $a \in K$, pero esto contradice al teorema anterior.

Supongamos ahora que $|K : F(H)| > |H| = n$. Sean $b_1, \ldots, b_{n+1}$ elementos de K linealmente independientes sobre $F(H)$.

Como antes podemos concluir que existen elementos $(a_1, \ldots, a_{n+1})$ de K no todos nulos de modo que

$$\sum_{i=1}^{n+1} a_i \sigma_j(b_i) = 0 \qquad \text{para } j = 1, \ldots, n. \tag{5.3}$$

Tomando el valor de j correspondiente al automorfismo identidad, se cumple que $\sum_{i=1}^{n+1} a_i b_i = 0$, lo que significa que alguno de los $a_i \notin F(H)$, pues los b_i son independientes.

Reordenando podemos suponer que $a_i \neq 0$ para $i = 1, \ldots, r$ y que los restantes son nulos. También podemos escoger los a_i con r mínimo. Ha de ser $r > 1$, porque en otro caso tendríamos $b_1 a_1 = 0$ y entonces $a_1 = 0$, contradicción.

Podemos multiplicar todos los a_i por a_r^{-1} y así suponer que $a_r = 1$. Como algún $a_i \notin F(H)$ y éste no es ciertamente a_r, podemos tomar $a_1 \notin F(H)$. Entonces existe un índice h tal que $\sigma_h(a_1) \neq a_1$.

Aplicando σ_h a (5.3) tenemos que se cumple $\sum_{i=1}^{n+1} \sigma_h(a_i)\sigma_h\big(\sigma_j(b_i)\big) = 0$ para todo $j = 1, \ldots, n$. Como $\sigma_j\sigma_h$ recorre todo H cuando $j = 1, \ldots, n$, podemos escribir $\sum_{i=1}^{n+1} \sigma_h(a_i)\sigma_j(b_i) = 0$ y restando de (5.3) llegamos a que

$$\sum_{i=1}^{n+1} \big(a_i - \sigma_h(a_i)\big)\sigma_j(b_i) = 0,$$

para $j = 1, \ldots, n$. Pero ahora los coeficientes son nulos para $i = r, \ldots, n+1$, aunque no lo es el primero. O sea, hemos encontrado unos valores que cumplen lo mismo que $(a_1, \ldots, a_{n+1})$ pero con más ceros, en contra de la minimalidad de r. ∎

Con esto llegamos al teorema de Galois. En el apartado 7) usamos la notación KL para el cuerpo $K(L) = L(K)$, o sea, el mínimo cuerpo que contiene a K y L (donde K y L son dos cuerpos contenidos en un cuerpo común).

Teorema 5.44 (Teorema Fundamental de la Teoría de Galois) *Sea K/k una extensión finita de Galois.*

1. *Existe una biyección entre los cuerpos intermedios $k \subset L \subset K$ y los subgrupos de $G(K/k)$. Esta biyección asigna a cada cuerpo L el grupo $G(K/L)$ y su inversa asigna a cada grupo H el cuerpo $F(H)$.*
2. *Si $k \subset L \subset L' \subset K$, entonces $G(K/L') \leq G(K/L) \leq G(K/k)$.*
3. *Si $H \leq H' \leq G(K/k)$, entonces $k \subset F(H') \subset F(H) \subset K$.*
4. *Si $k \subset L \subset K$ entonces K/L es una extensión normal (luego de Galois).*
5. *Si $k \subset L \subset K$, la extensión L/k es normal (luego de Galois) si y sólo si $G(K/L)$ es un subgrupo normal de $G(K/k)$.*

6. *Si $k \subset L \subset K$ y L/k es de Galois, la aplicación $r : G(K/k) \longrightarrow G(L/k)$ dada por $r(\sigma) = \sigma|_L$ es un epimorfismo de grupos cuyo núcleo es $G(K/L)$, luego $G(L/k) \cong G(K/k)/G(K/L)$*

7. *Si $H_1, H_2 \leq G(K/k)$ entonces*

$$F\big(\langle H_1, H_2\rangle\big) = F(H_1) \cap F(H_2) \quad y \quad F(H_1 \cap H_2) = F(H_1)F(H_2).$$

DEMOSTRACIÓN: 4) Si K/k es normal, K es el cuerpo de escisión sobre k de un cierto polinomio $p(x)$, pero, obviamente, K también es el cuerpo de escisión sobre L de $p(x)$, luego K/L es normal.

1) La aplicación que a cada L le asigna $G(K/L)$ es inyectiva, pues si tenemos $G(K/L) = G(K/L')$, entonces $F\big(G(K/L)\big) = F\big(G(K/L')\big)$, y como las extensiones son de Galois, $L = L'$.

Si $H \leq G(K/k)$ y $L = F(H)$, es obvio que $H \leq G(K/L)$ y, por el teorema anterior, $|H| = |K : L| = \big|G(K/L)\big|$ porque la extensión es de Galois, luego $H = G(K/L)$. Por lo tanto L es una antiimagen de H, luego la aplicación es biyectiva y su inversa es la que indica el enunciado.

2) y 3) son inmediatos.

5) Supongamos que L/k es normal. Sea $\sigma \in G(K/L)$ y $\tau \in G(K/k)$. Para cada $a \in L$ se cumple que $\tau^{-1}(a) \in L$ (por el teorema 5.22). Consecuentemente, $\sigma\big(\tau^{-1}(a)\big) = \tau^{-1}(a)$ y $\sigma^\tau(a) = \tau\Big(\sigma\big(\tau^{-1}(a)\big)\Big) = \tau\big(\tau^{-1}(a)\big) = a$, luego tenemos que $\sigma^\tau \in G(K/L)$ y en consecuencia $G(K/L) \trianglelefteq G(K/k)$.

Si $G(K/L) \trianglelefteq G(K/k)$, tomemos un polinomio irreducible $p(x) \in k[x]$ con una raíz a en L. Como K/k es normal $p(x)$ se escinde en K. Basta probar que todas las raíces de $p(x)$ en K están en L, y así $p(x)$ se escindirá en L. Sea b otra raíz de $p(x)$ en K. Por el teorema 5.23 existe un automorfismo $\tau \in G(K/k)$ tal que $\tau(b) = a$. Hemos de probar que $b \in F\big(G(K/L)\big) = L$.

Sea $\sigma \in G(K/L)$. Entonces $\sigma^\tau \in G(K/L)$, luego $\tau\big(\sigma(\tau^{-1}(a)\big) = a$, o sea, $\tau\big(\sigma(b)\big) = a$ o, lo que es lo mismo, $\sigma(b) = b$.

6) La aplicación r está bien definida porque si $\sigma \in G(K/k)$, entonces el teorema 5.22 garantiza que $\sigma[L] = L$, luego $r(\sigma) \in G(L/k)$. La aplicación r es suprayectiva también por el teorema 5.22, y obviamente es un epimorfismo de grupos. El resto es inmediato.

7) Es una consecuencia inmediata de 1), 2) y 3). El grupo $\langle H_1, H_2\rangle$ es el menor subgrupo de $G(K/k)$ que contiene a H_1 y a H_2, luego su cuerpo fijado ha de ser el mayor cuerpo intermedio contenido en $F(H_1)$ y en $F(H_2)$, o sea, ha de ser $F(H_1) \cap F(H_2)$. Análogamente se tiene la otra igualdad. ∎

Ejemplo El estudio de los grupos de Galois nos permite conocer todos los cuerpos intermedios de una extensión finita de Galois. Como ilustración vamos a aplicarlo al cuerpo de escisión sobre $\mathbb{Q}$ del polinomio $x^3 - 2$. Sabemos que es de la forma $K = \mathbb{Q}(\alpha, \beta)$, donde α, β y γ son las raíces del polinomio, así como que el grado de la extensión es 6, que el grupo de Galois G es isomorfo a Σ_3 y que por lo tanto permuta las tres raíces de todos los modos posibles.

Como Σ_3 tiene cuatro subgrupos propios, el teorema de Galois nos dice que la extensión $\mathbb{Q}(\alpha,\beta)/\mathbb{Q}$ tiene exactamente cuatro cuerpos intermedios. Vamos a calcularlos.

Los subgrupos de Σ_3 son los tres subgrupos de orden 2 generados por las trasposiciones (β,γ), (α,γ) y (α,β) y el subgrupo de orden 3 generado por el ciclo (α,β,γ).

El cuerpo $F\big(\langle(\beta,\gamma)\rangle\big)$ cumple que $\Big|\mathbb{Q}(\alpha,\beta) : F\big(\langle(\beta,\gamma)\rangle\big)\Big| = \big|\langle(\beta,\gamma)\rangle\big| = 2$, luego $\Big|F\big(\langle(\beta,\gamma)\rangle\big) : \mathbb{Q}\Big| = 3 = \big|\mathbb{Q}(\alpha) : \mathbb{Q}\big|$. Como obviamente $\alpha \in F\big(\langle(\beta,\gamma)\rangle\big)$, tenemos la inclusión $\mathbb{Q}(\alpha) \subset F\big(\langle(\beta,\gamma)\rangle\big)$ y, al coincidir los grados, ha de ser $\mathbb{Q}(\alpha) = F\big(\langle(\beta,\gamma)\rangle\big)$.

Igualmente, $\mathbb{Q}(\beta) = F\big(\langle(\alpha,\gamma)\rangle\big)$ y $\mathbb{Q}(\gamma) = F\big(\langle(\alpha,\beta)\rangle\big)$.

Nos falta calcular $F\big(\langle(\alpha,\beta,\gamma)\rangle\big)$, que ha de tener grado 2 sobre $\mathbb{Q}$. Para ello observamos que como $\alpha \neq \beta$, se cumple $\alpha/\beta \neq 1$, pero $(\alpha/\beta)^3 = 2/2 = 1$, luego $\omega = \alpha/\beta$ es una raíz del polinomio $x^3 - 1$ distinta de 1. Por consiguiente $\mathbb{Q}(\omega)$, el cuerpo ciclotómico de orden 3, está contenido en K y tiene grado 2 sobre $\mathbb{Q}$, luego ha de ser $F\big(\langle(\alpha,\beta,\gamma)\rangle\big) = \mathbb{Q}(\omega) = \mathbb{Q}\big(\sqrt{-3}\,\big)$. ∎

En la sección siguiente mostraremos un primer ejemplo de cómo la teoría de grupos permite obtener resultados no triviales sobre cuerpos, pero antes veamos un par de resultados sencillos, pero muy útiles, con los que acabaremos de perfilar la teoría de Galois:

Teorema 5.45 *Sean K/k y L/k extensiones de un cuerpo k (contenidas en un mismo cuerpo) y además supongamos que K/k es de finita de Galois. Entonces la extensión KL/L es también de Galois y $G(KL/L) \cong G(K/K\cap L)$.*

Demostración: La extensión K/k es normal, luego K es el cuerpo de escisión sobre k de un cierto polinomio $f(x) \in k[x]$. Sean $a_1,\dots,a_n$ las raíces de $f(x)$ en K. Entonces $K = k(a_1,\dots,a_n)$, luego es obvio que $KL = L(a_1,\dots,a_n)$.

Además $a_1,\dots,a_n$ son las raíces de $f(x) \in L[x]$ en KL, luego KL es el cuerpo de escisión sobre L de $f(x)$, lo que implica que la extensión KL/L es normal. Como los a_i son separables sobre k, lo son sobre L y así KL/L es de Galois.

Si $\sigma \in G(KL/L)$, como K/k es normal, se cumple que $\sigma|_K : K \longrightarrow K$, y claramente $\sigma|_K \in G(K/K\cap L)$.

La aplicación $\phi : G(KL/L) \longrightarrow G(K/K\cap L)$ dada por $\phi(\sigma) = \sigma|_K$ es ciertamente un homomorfismo de grupos y de hecho es un monomorfismo, pues si $\phi(\sigma) = I|_K$, entonces, como $\sigma|_L = I|_L$, resulta que $\sigma = I|_{KL}$.

Consideremos el cuerpo $E = F(\operatorname{Im}\phi)$. Entonces $K\cap L \subset E \subset K$. Si $a \in E$ se cumple $\sigma|_K(a) = a$ para todo $\sigma \in G(KL/L)$, luego $a \in F\big(G(KL/L)\big) = L$.

Por lo tanto $E \subset K\cap L$, es decir, $E = K\cap L$ y así $F(\operatorname{Im}\phi) = F\big(G(K/K\cap L)\big)$, con lo que $\operatorname{Im}\phi = G(K/K\cap L)$ y ϕ es un isomorfismo. ∎

Teorema 5.46 *Sean K/k y L/k dos extensiones finitas de Galois de un mismo cuerpo k (ambas contenidas en un cuerpo mayor) y tales que $K\cap L = k$. Entonces KL/k es finita de Galois y $G(KL/k) \cong G(K/k)\times G(L/k)$.*

DEMOSTRACIÓN: Si K es el cuerpo de escisión sobre k de un polinomio $p(x)$ y L es el cuerpo de escisión de $q(x)$, es claro que KL es el cuerpo de escisión de pq, luego KL/k es una extensión finita de Galois (la separabilidad es clara). Por el teorema de Galois podemos definir $\phi : G(KL/k) \longrightarrow G(K/k) \times G(L/k)$ mediante $\phi(\sigma) = (\sigma|_K, \sigma|_L)$.

Obviamente ϕ es un homomorfismo de grupos. De hecho es un monomorfismo porque su núcleo es claramente trivial. Para probar que ϕ es biyectiva basta ver que ambos grupos tienen el mismo orden, pero por el teorema anterior $\big|G(K/k)\big| = \big|G(KL/L)\big|$, y así $|KL : k| = |KL : L|\,|L : k| = |K : k|\,|L : k|$. ∎

5.6 Cuerpos algebraicamente cerrados

Sabemos que si k es un cuerpo y $p(x) \in k[x]$, entonces existe una extensión de k en la que $p(x)$ tiene todas sus raíces (el cuerpo de escisión del polinomio), pero cabe preguntarse si no es posible encontrar un cuerpo k que contenga ya todas las raíces de todos los polinomios de $k[x]$. El teorema siguiente precisa esta idea:

Teorema 5.47 *Sea K un cuerpo. Las condiciones siguientes son equivalentes:*

1. *K no tiene extensiones algebraicas distintas de sí mismo.*
2. *Los polinomios irreducibles en $K[x]$ son los polinomios de grado 1.*
3. *Todo polinomio no constante de $K[x]$ tiene una raíz en K.*
4. *Todo polinomio de $K[x]$ se escinde en $K[x]$.*
5. *K contiene un subcuerpo k tal que la extensión K/k es algebraica y todo polinomio no constante de $k[x]$ se escinde en $K[x]$.*

DEMOSTRACIÓN: 1) ⇒ 2) Sabemos que todo polinomio de grado 1 es irreducible, y los de grado 0 son unidades o el 0. Si existiera un polinomio irreducible de grado mayor que 1, entonces dicho polinomio no tendría raíces en K, luego existiría una extensión de K en el que tendría una raíz a, y $K(a)$ sería una extensión algebraica propia de K.

2) ⇒ 3) Si un polinomio no es constante entonces no es nulo ni unitario, luego tiene un factor irreducible, que será de la forma $ax + b \in K[x]$ con $a \neq 0$, luego el polinomio tendrá por raíz $-b/a$.

3) ⇒ 4) Si $f(x) \in K[x]$ es constante entonces $f(x) = a_0 \in K$, luego se escinde en $K[x]$. Si no es constante tiene una raíz $a_1 \in K$, luego $f(x) = (x - a_1)f_1(x)$, para cierto polinomio $f_1(x) \in K[x]$. Si $f_1(x)$ tampoco es constante tiene una raíz $a_2 \in K$, luego $f(x) = (x - a_1)f_1(x) = (x - a_1)(x - a_2)f_2(x)$, para cierto polinomio $f_2(x) \in K[x]$. Como el grado de los polinomios que vamos obteniendo es cada vez una unidad menor, al cabo de un número finito n de pasos llegaremos a un polinomio de grado 0, es decir, a una constante a_0 y tendremos $f(x) = a_0(x - a_1)\cdots(x - a_n)$.

4) $\Rightarrow$ 5) Basta tomar $k = K$.

5) $\Rightarrow$ 1) Si L/K es una extensión algebraica y $a \in L$ entonces a es algebraico sobre k, por 5.9, luego podemos considerar el polinomio $p(x) = \operatorname{pol}\operatorname{mín}(a, k)$, que por hipótesis se escinde en $K[x]$. Existen $a_0, a_1, \ldots, a_n \in K$ tales que

$$p(x) = a_0(x - a_1)\cdots(x - a_n).$$

Como $p(a) = 0$, necesariamente $a = a_i$ para algún i, luego $a \in K$ y en consecuencia $L = K$. ∎

Definición 5.48 Diremos que un cuerpo K es *algebraicamente cerrado* si cumple cualquiera de las condiciones del teorema anterior.

La propiedad 5. está encaminada a encontrar cuerpos algebraicamente cerrados, pues lo que afirma es que si tenemos un cuerpo k que no es algebraicamente cerrado, para que una extensión algebraica K de k lo sea basta con comprobar que los polinomios de $k[x]$ se escinden[1] en $K[x]$, sin necesidad de considerar los "nuevos" polinomios (los de $K[x]$) que hemos obtenido al extender el cuerpo.

Por ejemplo, es obvio que $\mathbb{Q}$ no es algebraicamente cerrado y, a pesar de que muchos polinomios irreducibles en $\mathbb{Q}[x]$ se escinden en $\mathbb{R}[x]$, lo cierto es que $\mathbb{R}$ no es "lo suficientemente grande" como para ser algebraicamente cerrado, pues, por ejemplo, el polinomio $x^2 + 1$ no tiene raíces en $\mathbb{R}$.

El teorema fundamental del álgebra afirma que $\mathbb{C}$ sí que es un cuerpo algebraicamente cerrado. Hemos dado una prueba "topológica" en [ITAn 3.33] y otra "analítica" en [IC 7.24]. Ahora vamos a dar otra algebraica (aunque requiere dos hechos elementales que se comprueban analíticamente). Conviene enunciar el resultado en un contexto general, que constituye la primera aplicación profunda de la teoría de Galois que vamos a presentar. Necesitamos un resultado previo:

Teorema 5.49 *Sea R un cuerpo ordenado en el que todo elemento positivo tiene raíz cuadrada y sea $C = R(i)$ la adjunción a C de una raíz del polinomio $x^2 + 1$. Entonces todo elemento de C tiene raíz cuadrada. En particular todo número complejo tiene raíz cuadrada.*

Demostración: En principio, todo $a \in R$ tiene raíz cuadrada en C, pues si $a \geq 0$ entonces la tiene en R y si $a < 0$, entonces $-a > 0$, luego existe $\sqrt{-a} \in R$, luego $\sqrt{-a}\, i \in C$ es una raíz cuadrada de a en C.

Ahora consideramos un elemento arbitrario de C, que será de la forma $a+bi$, con $a, b \in R$, y definimos

$$c = \sqrt{\frac{1}{2}(\sqrt{a^2+b^2}+a)}, \quad d = \sqrt{\frac{1}{2}(\sqrt{a^2+b^2}-a)} \in C.$$

Esto es correcto, pues $a^2 + b^2 \geq 0$, luego por hipótesis existe la raíz cuadrada $\sqrt{a^2+b^2} \in R$, luego las raíces cuadradas exteriores existen en C.

[1] El teorema 9.20 afirma que, de hecho, basta con que tengan una raíz en K.

Se comprueba trivialmente que $c^2 - d^2 = a$ y que $2cd = \pm b$. Cambiando c por $-c$ si es preciso se sigue cumpliendo la primera ecuación y la segunda pasa a ser $2cd = b$, de donde se sigue que $(c + di)^2 = a + bi$. ∎

Teorema 5.50 *Sea R un cuerpo ordenado que cumpla las dos propiedades siguientes:*

1. *Todo $a \in R$ positivo tiene una raíz cuadrada en R.*
2. *Todo polinomio de grado impar en $R[x]$ tiene al menos una raíz en R.*

Sea $K = R(i)$, donde i es una raíz del polinomio $x^2 + 1$. Entonces el cuerpo K es algebraicamente cerrado.

DEMOSTRACIÓN: Notemos que $x^2 + 1$ no puede tener raíces en un cuerpo ordenado, por lo que $|K : R| = 2$. El teorema anterior nos da que todo elemento de K tiene raíz cuadrada. La fórmula de las ecuaciones de segundo grado nos da a su vez que todo polinomio de grado 2 con coeficientes en K tiene sus raíces en K, lo cual equivale a que K no tiene extensiones algebraicas de grado 2.

Supongamos ahora que K no es algebraicamente cerrado. Entonces existe un polinomio no constante sin raíces en K. Sea L la adjunción a K de una raíz de dicho polinomio, con lo que tenemos una extensión finita L/K de grado mayor que 1. La extensión L/R también es finita. Cambiando L por la clausura normal de L sobre R podemos suponer que L/R es una extensión finita de Galois (notemos que R, al ser un cuerpo ordenado, tiene característica 0, luego la extensión es separable). Como $R \subset C \subset L$, tenemos que $|L : R|$ es par.

Consideramos el grupo de Galois $G = G(L/R)$, que tiene orden par, digamos $|G| = 2^n m$, con m impar y $n \geq 1$, y vamos a usar un hecho nada trivial:

El grupo G tiene un subgrupo P de orden 2^n.

Este subgrupo P es lo que se llama un 2-subgrupo de Sylow de G, y probaremos su existencia en [TG 3.27]. Sea F el cuerpo fijado por P, de modo que $R \subset F \subset L$. La extensión F/R es finita separable, luego tiene un elemento primitivo, digamos $F = R(a)$. Sea $p = \text{pol mín}(a, R)$. Entonces p es irreducible en $R[x]$, pero es un polinomio de grado impar, pues $|F : R| = |G : P| = m$ es impar. Por hipótesis p tiene que tener una raíz en R, lo cual sólo es posible si p tiene grado 1, luego $F = R$ y, por consiguiente, $P = G$.

Así pues, hemos concluido que la extensión L/R tiene grado potencia de 2, luego lo mismo le sucede a L/K. Concretamente, $|L : K| = |G(L/K)| = 2^{n-1}$. Ahora usamos otro hecho no trivial:

El grupo G tiene un subgrupo N de índice 2.

Esto lo demostraremos en [TG 3.6]. Si F es el cuerpo fijado de N, tenemos que $|F : K| = 2$, pero hemos visto que K no puede tener extensiones de grado 2, con lo que tenemos una contradicción. ∎

Observemos que el cuerpo $\mathbb{R}$ de los números reales cumple las dos hipótesis del teorema anterior. La existencia de raíces cuadradas está probada en [An 1.15], mientras que la segunda hipótesis es [ITAn 3.25], y daremos otra prueba en [An 3.26]. Por lo tanto, tenemos así otra demostración de que el cuerpo $\mathbb{C}$ de los números complejos es algebraicamente cerrado. El teorema siguiente nos da otro ejemplo:

Teorema 5.51 *Si K/k es una extensión de cuerpos con K algebraicamente cerrado y A es el conjunto de los elementos de K algebraicos sobre k, entonces A es un cuerpo algebraicamente cerrado.*

DEMOSTRACIÓN: Es claro que $A = k(A)$, luego es un cuerpo, y si tomamos un polinomio no constante $p(x) \in k[x]$, sabemos que se escinde en $K[x]$, pero sus raíces serán algebraicas sobre k, luego estarán en A, luego $p(x)$ se escinde en $A[x]$, luego A es algebraicamente cerrado por el apartado 5) de 5.47. ∎

Definición 5.52 Llamaremos $\mathbb{A}$ al conjunto de los números complejos algebraicos sobre $\mathbb{Q}$ que, por el teorema anterior, es un cuerpo algebraicamente cerrado. Sus elementos se llaman simplemente *números algebraicos*.

Llamamos $\mathbb{R}_a$ al conjunto de los números reales algebraicos (sobre $\mathbb{Q}$). Es claro entonces que $a + bi \in \mathbb{A}$ si y sólo si a, $b \in \mathbb{R}_a$.

En efecto, si $a, b \in \mathbb{R}_a$, entonces, $a, b, i \in \mathbb{A}$, luego $a + bi \in \mathbb{A}$. Recíprocamente, si $a + bi \in \mathbb{A}$, entonces el conjugado $a - bi$ también es algebraico (con el mismo polinomio mínimo sobre $\mathbb{Q}$), luego $a + bi + a - bi = 2a \in \mathbb{A}$, luego $a \in \mathbb{A}$, luego $bi = a + bi - a \in \mathbb{R}_a$, luego $b \in \mathbb{R}_a$.

Es claro entonces que $\mathbb{A} = \mathbb{R}_a(i)$. De hecho, es fácil ver que $\mathbb{R}_a$ cumple las hipótesis del teorema 5.50, por lo que la clausura algebraica de $\mathbb{A}$ se deduce también de este teorema.

Observemos que $\mathbb{R}_a \neq \mathbb{R}$ porque $\mathbb{R}_a$ es numerable [An 1.55] y $\mathbb{R}$ no lo es [An 1.57]. Claramente entonces $\mathbb{A}$ también es numerable, mientras que $|\mathbb{C}| = \mathfrak{c}$.

Ejercicio: Probar que la extensión $\mathbb{A}/\mathbb{Q}$ es algebraica, pero infinita.

Así pues, si queremos trabajar con extensiones algebraicas de $\mathbb{Q}$, sólo necesitamos considerar subcuerpos de $\mathbb{C}$, o incluso de $\mathbb{A}$, pero hay muchos casos en los que es necesario estudiar extensiones algebraicas de otros cuerpos. Por ejemplo, de cuerpos finitos. Vamos a probar que siempre podemos considerar un cuerpo análogo a $\mathbb{A}$ para cualquier cuerpo dado:

Definición 5.53 Diremos que un cuerpo K es una *clausura algebraica* de un cuerpo k si K es una extensión algebraica de k y es algebraicamente cerrado.

Por ejemplo, $\mathbb{C}$ es una clausura algebraica de $\mathbb{R}$, y es la única salvo $\mathbb{R}$-isomorfismo, es decir, que si K es cualquier clausura algebraica de $\mathbb{R}$, entonces K es un cuerpo $\mathbb{R}$-isomorfo a $\mathbb{C}$.

En efecto, podemos considerar una raíz $\alpha \in K$ del polinomio x^2+1, de modo que existe un $\mathbb{R}$-isomorfismo $\sigma : \mathbb{C} \longrightarrow \mathbb{R}(\alpha) \subset K$, pero entonces $\mathbb{R}(\alpha)$ es un cuerpo algebraicamente cerrado y la extensión $K/\mathbb{R}(\alpha)$ es algebraica, luego tiene que ser $K = \mathbb{R}(\alpha)$, y así $\sigma : \mathbb{C} \longrightarrow K$ es un $\mathbb{R}$-isomorfismo.

Similarmente, $\mathbb{A}$ es una clausura algebraica de $\mathbb{Q}$, y también es única salvo $\mathbb{Q}$-isomorfismo. Vamos a deducir esto de un resultado general:

Teorema 5.54 (AE)[2] *Sea K/k una extensión algebraica y $\sigma : k \longrightarrow L$ un monomorfismo de cuerpos, con L es algebraicamente cerrado. Entonces σ se extiende a un monomorfismo $\sigma^* : K \longrightarrow L$.*

Demostración: Sea M el conjunto de todos los pares (A, τ) tales que $k \subset A \subset K$ y $\tau : A \longrightarrow L$ es un monomorfismo que extiende a σ. El conjunto M está inductivamente ordenado por la relación

$$(A, \tau) \leq (A', \tau') \text{ si y sólo si } A \subset A' \text{ y } \tau'|_A = \tau.$$

Sea (A, τ) un elemento maximal. Es suficiente probar que $A = K$. En otro caso sea $u \in K \setminus A$. Sea $p(x) = \operatorname{pol}\operatorname{mín}(u, A)$ y sea $\tau p(x)$ el polinomio correspondiente en $\tau[A][x] \subset L[x]$ que, al ser L algebraicamente cerrado, tiene una raíz $v \in L$.

El teorema 5.12 nos da un monomorfismo $\tau' : A(u) \longrightarrow L$ que extiende a τ, en contra de la maximalidad de (A, τ). Por tanto $A = K$. ∎

El caso particular que nos interesa de momento es:

Teorema 5.55 (AE)[2] *Si K es una clausura algebraica de k, K' es una clausura algebraica de k' y $\sigma : k \longrightarrow k'$ es un isomorfismo, entonces σ se extiende a un isomorfismo $\sigma^* : K \longrightarrow K'$. En particular dos clausuras algebraicas de un cuerpo k son k–isomorfas.*

Demostración: Por el teorema anterior, σ se extiende a un monomorfismo $\sigma^* : K \longrightarrow K'$.

Como K es algebraicamente cerrado, $\sigma^*[K]$ también lo es, y la extensión $K'/\sigma^*[K]$ es algebraica, luego ha de ser $\sigma^*[K] = K'$, es decir, σ^* es un isomorfismo.

Si K y K' son dos clausuras algebraicas de un mismo cuerpo k, entonces la identidad en k se extiende a un k–isomorfismo de K en K'. ∎

Notemos que si consideramos extensiones algebraicas de $\mathbb{Q}$, no perdemos generalidad si suponemos que son subcuerpos de $\mathbb{C}$ y, más concretamente de $\mathbb{A}$, pues ahora sabemos que toda extensión algebraica de $\mathbb{Q}$ es $\mathbb{Q}$-isomorfa a un subcuerpo de $\mathbb{A}$.

Aunque hemos construido explícitamente una clausura algebraica para $\mathbb{R}$ y para $\mathbb{Q}$, lo cierto es que todo cuerpo tiene una clausura algebraica:

[2]No se necesita AE si la extensión K/k es finita, pues entonces puede expresarse como $K = k(a_1, \ldots, a_n)$ y σ puede ir extendiéndose sucesivamente a cada $k(a_1, \ldots a_i)$ usando el teorema 5.12. Si $|K : k| = \aleph_0$ expresamos $K = k(a_1, a_2, \ldots)$ y sólo necesitamos el axioma ED junto con 5.12.

Teorema 5.56 (AE)[3] *Todo cuerpo tiene una clausura algebraica.*

DEMOSTRACIÓN: Sea k un cuerpo y consideremos el conjunto $C = k[x] \times \mathbb{N}$. Para cada $\alpha \in k$, llamemos $\bar{\alpha} = (x - \alpha, 0) \in C$. Sea $\bar{k} = \{\bar{\alpha} \mid \alpha \in k\} \subset C$. Notemos que la aplicación $k \longrightarrow \bar{k}$ dada por $\alpha \mapsto \bar{\alpha}$ es biyectiva, por lo que todo elemento de $\bar{k}$ se expresa de forma única como $\bar{\alpha}$, para un cierto $\alpha \in k$. Dotamos a $\bar{k}$ de estructura de cuerpo mediante $\bar{\alpha} + \bar{\beta} = \overline{\alpha + \beta}$, $\bar{\alpha}\bar{\beta} = \overline{\alpha\beta}$. Es claro que $\alpha \mapsto \bar{\alpha}$ es ahora un isomorfismo de cuerpos, que podemos extender a un isomorfismo de anillos $k[x] \longrightarrow \bar{k}[x]$. Si $p(x) \in k[x]$, representaremos su imagen por $\bar{p}(x) \in \bar{k}[x]$.

Así hemos construido otro cuerpo $\bar{k}$ que algebraicamente es el mismo que k, pero que desde un punto de vista conjuntista está contenido en C. Basta probar que $\bar{k}$ tiene una clausura algebraica.

Sea $\mathcal{M}$ el conjunto[4] de todos los cuerpos K tales que $K/\bar{k}$ es una extensión algebraica y, como conjunto, $K \subset C$ y además, si $\xi = (p(x), n) \in K$, entonces $\bar{p}(\xi) = 0$.

Notemos que $\bar{k} \in \mathcal{M}$ pues si $\xi \in \bar{k}$, entonces $\xi = \bar{\alpha} = (x - \alpha, 0)$, para cierto $\alpha \in k$, de modo que $\overline{x - \alpha} = x - \bar{\alpha} = x - \xi$, y ciertamente ξ es raíz de $x - \xi$.

Definimos en $\mathcal{M}$ la relación de orden dada por $K \leq K'$ si y sólo si K es un subcuerpo de K'. Vamos a comprobar que $\mathcal{M}$ cumple las hipótesis del lema de Zorn. Si $\mathcal{C} \subset \mathcal{M}$ es una cadena, entonces $L = \bigcup \mathcal{C} \subset C$ y si $\xi, \xi' \in L$, existe un $K \in \mathcal{C}$ tal que $\xi, \xi' \in K$ (porque $\mathcal{C}$ es una cadena) y las operaciones $\xi + \xi'$, $\xi\xi'$ son independientes de la elección de K. Por lo tanto, esto define una suma y un producto en L de modo que cada $K \in \mathcal{C}$ es un subcuerpo de L. Esto implica claramente que la extensión $L/\bar{k}$ es algebraica, y si $\xi = (p(x), n) \in L$, existe un $K \in \mathcal{C}$ tal que $\xi \in K$, luego por definición de $\mathcal{M}$ tenemos que $\bar{p}(\xi) = 0$, luego $L \in \mathcal{M}$ es una cota superior de $\mathcal{C}$.

Por el lema de Zorn existe un elemento maximal $K \in \mathcal{M}$. Basta probar que K es algebraicamente cerrado. En caso contrario, K admite una extensión algebraica K'. Para cada polinomio mónico irreducible $p(x) \in k[x]$ tal que $\bar{p}(x)$ tenga al menos una raíz en $K' \setminus K$, sean $\alpha_1, \dots \alpha_m$ sus raíces en $K' \setminus K$. Como $\bar{p}(x)$ tiene a lo sumo un número finito de raíces en K, sólo puede haber un número finito de números naturales n tales que $(p(x), n) \in K$, luego podemos tomar un n mayor que todos ellos y definir $\bar{\alpha}_i = (p(x), n + i) \in C \setminus K$.

Así tenemos una aplicación inyectiva $K' \longrightarrow C$ que sobre K es la identidad y sobre $K' \setminus K$ es $\alpha_i \mapsto \bar{\alpha}_i$. Si llamamos $\bar{K}' \subset C$ a la imagen de esta aplicación, es claro que podemos dotar a $\bar{K}'$ de estructura de cuerpo K-isomorfo a K' y, como $\bar{p}(\alpha_i) = 0$, tenemos que $\bar{K}' \in \mathcal{M}$ y contradice la maximalidad de K. ∎

[3] Se puede evitar el axioma de elección para cuerpos numerables.

[4] Más precisamente, los elementos de $\mathcal{M}$ son ternas $(K, +, \cdot)$, de manera que $\bar{k} \subset K \subset C$ y $+, \cdot : K \times K \longrightarrow K$. En particular $+, \cdot \subset C \times C \times C$, luego

$$(K, +, \cdot) \in \mathcal{P}C \times \mathcal{P}(C \times C \times C) \times \mathcal{P}(C \times C \times C),$$

luego $\mathcal{M} \subset \mathcal{P}(\mathcal{P}C \times \mathcal{P}(C \times C \times C) \times \mathcal{P}(C \times C \times C)) = \overline{\mathcal{M}}$. La función de C en la demostración es garantizar que $\mathcal{M}$ pueda definirse por especificación a partir del conjunto $\overline{\mathcal{M}}$.

Extensiones infinitas normales Los resultados de esta sección nos permiten generalizar algunos resultados que hemos demostrado para extensiones finitas normales. Para eliminar la hipótesis de finitud debemos generalizar el concepto de cuerpo de escisión para considerar un conjunto arbitrario de polinomios:

Definición 5.57 Si K/k es una extensión y $P \subset k[x]$, se dice que K es el *cuerpo de escisión* sobre k de los polinomios de P si todos los polinomios de P se escinden en K y K es la adjunción a k del conjunto de todas sus raíces.

En primer lugar generalizamos 5.18:

Teorema 5.58 (AE) *Si k es un cuerpo y $P \subset k[x]$, existe una extensión K/k tal que K es un cuerpo de escisión de P sobre k, y si K' cumple lo mismo, entonces K y K' son k-isomorfos.*

Demostración: Basta tomar una clausura algebraica $\bar{K}$ de k y tomar $K = k(S)$, donde S es el conjunto de las raíces en $\bar{K}$ de los polinomios de P. Claramente es un cuerpo de escisión de P sobre k. Si K'/k es otro cuerpo de escisión, tomamos una clausura algebraica $\bar{K}'$, de modo que existe un k-isomorfismo $\sigma : \bar{K} \longrightarrow \bar{K}'$, el cual transforma las raíces de los polinomios de P en $\bar{K}$ en las raíces de estos mismos polinomios en $\bar{K}'$, luego transforma K en K', luego se restringe a un k-isomorfismo $\sigma : K \longrightarrow K'$. ∎

Ahora generalizamos 5.21:

Teorema 5.59 (AE) *Una extensión K/k es normal si y sólo si K es el cuerpo de escisión de un conjunto de polinomios $P \subset k[x]$.*

Demostración: Si K/k es normal, entonces K es la adjunción a k de las raíces del conjunto P de todos los polinomios que tienen al menos una raíz en K (luego se escinden en K, por la normalidad). Recíprocamente, si $K = k(S)$, donde S es el conjunto de las raíces de un conjunto P de polinomios que se escinden en K, en particular K/k es una extensión algebraica. Sea ahora $f(x) \in k[x]$ un polinomio no nulo con una raíz $\alpha \in K = k(S)$. Entonces

$$\alpha = \frac{p(b_1, \dots b_n)}{q(b_1, \dots b_n)},$$

donde los b_i son raíces de polinomios de P, luego podemos tomar $p_1, \dots, p_r \in P$ de modo que los b_i son raíces de los p_j. Sea $p = p_1 \cdots p_r$. Como cada p_i se escinde en $K[x]$, lo mismo le sucede a p. Sea S_0 el conjunto de todas las raíces de p, de modo que $K_0 = k(S_0) \subset K$ es el cuerpo de escisión de p, luego la extensión K_0/k es normal y $\alpha \in K_0$. Como f tiene una raíz en K_0, se escinde en $K_0[x]$, luego también en $K[x]$, luego la extensión es normal. ∎

Ahora generalizamos 5.22:

Teorema 5.60 (AE) *Sea $k \subset F \subset K \subset L$ una cadena de extensiones algebraicas tal que K/k sea normal. Sea $\sigma : F \longrightarrow L$ un k-monomorfismo. Entonces $\sigma[F] \subset K$ y σ se extiende a un k-automorfismo de K.*

DEMOSTRACIÓN: Sea $\bar{L}$ una clausura algebraica de L, con lo que también es una clausura algebraica de F y de $\sigma[F]$. Por el teorema 5.55 tenemos que σ se extiende a un k-automorfismo $\sigma : \bar{L} \longrightarrow \bar{L}$, pero $\sigma[K] = K$, porque K es la adjunción a k de las raíces en $\bar{L}$ de un conjunto de polinomios de $k[x]$ que se escinden en $\bar{L}[x]$, luego $\sigma[K]$ es la adjunción a k de esas mismas raíces. En particular $\sigma[F] \subset \sigma[K] = K$ y la restricción de σ a K es un k-automorfismo. ∎

En particular tenemos la generalización de 5.23:

Teorema 5.61 *Sea K/k una extensión normal y $a, b \in K$. Entonces a y b son k–conjugados si y sólo si existe un $\sigma \in G(K/k)$ tal que $\sigma(a) = b$.*

Y por último observamos que la prueba de 5.32 vale sin cambio alguno para extensiones infinitas usando el teorema anterior en lugar de su versión para extensiones finitas:

Teorema 5.62 *Una extensión algebraica K/k es de Galois si y sólo si su cuerpo fijado es k.*

5.7 Cuerpos formalmente reales

En geometría se trabaja a menudo con cuerpos ordenados, y un cuerpo ordenado nunca puede ser algebraicamente cerrado. Vamos a estudiar más detalladamente qué cuerpos admiten una relación de orden y a continuación introduciremos un concepto análogo al de clausura algebraica para cuerpos ordenados.

Definición 5.63 Un cuerpo R es *(formalmente) real* si existe una relación de orden en R con la que cumple los axiomas de cuerpo ordenado.

La diferencia entre "cuerpo real" y "cuerpo ordenado" es que el segundo concepto presupone una relación de orden dada en el cuerpo, mientras que un cuerpo real puede admitir distintas relaciones de orden que cumplan los axiomas de cuerpo ordenado.

Ejemplo Consideremos el cuerpo $R = \mathbb{Q}\big[\sqrt{2}\,\big] \subset \mathbb{R}$. Obviamente es un cuerpo ordenado en el que $\sqrt{2} > 0$, pero la conjugación $\sigma(a + b\sqrt{2}) = a - b\sqrt{2}$ es un $\mathbb{Q}$-automorfismo de R que nos permite definir otra relación de orden en R con la que R es también un cuerpo ordenado, a saber, la dada por

$$\alpha \leq^* \beta \quad \text{si y sólo si} \quad \sigma(\alpha) \leq \sigma(\beta).$$

Respecto a este orden, $\sqrt{2} <^* 0$. ∎

Que un cuerpo R sea formalmente real equivale a que exista un subconjunto con ciertas propiedades:

Sea R un cuerpo. Un conjunto $P \subset R$ es un *cono positivo* si cumple las propiedades siguientes:

1. Si a, $b \in P$, entonces $a + b \in P$ y $ab \in P$.
2. $R = P \cup (-P)$ y $P \cap (-P) = \{0\}$ (donde $-P = \{-a \mid a \in P\}$).

Es trivial comprobar que si R es un cuerpo ordenado, entonces el conjunto $P = \{a \in R \mid a \geq 0\}$ es un cono positivo y, recíprocamente, si P es un cono positivo en R, entonces la relación definida por $a \leq b$ si y sólo si $b - a \in P$ es una relación de orden total en R con la que R se convierte en cuerpo ordenado y respecto a la cual $P = \{a \in R \mid a \geq 0\}$.

Por lo tanto, un cuerpo es real si y sólo si admite un cono positivo. Esta condición se puede debilitar bastante. Empecemos por relajarla un poco:

Diremos que $P \subset R$ es un *precono positivo* si cumple las propiedades siguientes:

1. Si a, $b \in P$, entonces $a + b$, $ab \in P$.
2. Si $a \in R$, entonces $a^2 \in P$.
3. $-1 \notin P$.

Es claro que todo cono positivo es un precono positivo (basta tener en cuenta que es el conjunto de elementos ≥ 0 del orden que define en R). Recíprocamente:

Teorema 5.64 (AE) *Si R es un cuerpo, todo precono positivo en R está contenido en un cono positivo.*[5]

Demostración: Sea P_0 un precono positivo en R y consideremos la familia de todos los preconos positivos en R que contienen a P_0, considerada como conjunto parcialmente ordenado por la inclusión. Es claro que el lema de Zorn implica que existe un precono positivo P que contiene a P_0 y es maximal respecto de la inclusión. Vamos a probar que es un cono positivo. Obviamente cumple la propiedad 1) de la definición.

Veamos ahora que si $x \in R$, entonces, o bien $xP \cap (1 + P) = \varnothing$, o bien $-xP \cap (1 + P) = \varnothing$.

Supongamos lo contrario. Entonces existen elementos $a, b, c, d \in P$ tales que $xa = 1 + b$, $-xc = 1 + d$. Multiplicando ambas ecuaciones resulta que $-acx^2 = 1 + b + d + bd$, luego $-1 = acx^2 + b + d + bd \in P$, contradicción.

[5]Si el cuerpo R del teorema anterior es numerable, entonces no es necesario el axioma de elección. En efecto, podemos partir de una enumeración $\{a_n\}_{n=1}^{\infty}$ de R y definir una sucesión de preconos positivos $\{P_n\}_{n=0}^{\infty}$ que empiece con el precono dado P_0 y de modo que P_{n+1} sea el conjunto de todos los elementos de R de la forma

$$\sum_{i=1}^{s} a_{i1} \cdots a_{im_i}$$

con los $a_{ij} \in P_n \cup \{a_n\}$ si -1 no está en dicho conjunto, o $P_{n+1} = P_n$ en caso contrario. Es claro entonces que $P = \bigcup_n P_n$ es un precono positivo maximal.

Salvo que indiquemos otra cosa, esta observación se aplica a todos los resultados posteriores de esta sección, pues el axioma de elección sólo interviene a través de la aplicación de este teorema.

Supongamos, pues que $xP \cap (1+P) = \varnothing$. Sea $P' = P - xP$. Como $0 = 0^2 \in P$, todo $p \in P$ se puede poner como $p = p - x0 \in P'$, luego $P \subset P'$. Por otra parte, P' es un precono positivo: tomamos $p - xp'$, $q - xq' \in P'$, con $p, p', q, q' \in P$:

- $(p - xp') + (q - xq') = (p + q) - x(q + q') \in P'$.
- $(p - xp')(q - xq') = (pq + x^2 p'q') - x(pq' + qp') \in P'$
- Como $P \subset P'$, tenemos que P' contiene los cuadrados.
- Si $-1 \in P'$, entonces existen $a, b \in P$ tales que $-1 = a - bx$, luego $a + 1 = bx \in xP \cap (1 + P)$, contradicción.

La maximalidad de P implica entonces que $P = P'$ y, como $0, 1 \in P$ (porque son cuadrados) $-x = 0 - 1x \in P' = P$.

Similarmente, si suponemos que $-xP \cap (1+P) = \varnothing$, el mismo razonamiento aplicado a $-x$ nos da que $x \in P$. Por consiguiente, $R = P \cup (-P)$.

Supongamos por último que $a \in P \cap (-P)$ con $a \neq 0$. Entonces $a, -a \in P$ y, por la parte ya probada, $1/a \in P \cup (-P)$, luego $1/a \in P$ o $-1/a \in P$, luego resulta que $-1 = \mp a(\pm 1/a) \in P$, contradicción. ∎

Así pues, un cuerpo es real si y sólo si tiene un precono positivo. Ahora observamos que en todo cuerpo existe un conjunto que casi es un precono positivo:

Definición 5.65 Si K es un cuerpo, llamaremos S_K al conjunto de todas las sumas de cuadrados en K.

Como los preconos positivos contienen a todos los cuadrados y son cerrados para sumas, es obvio que S_K está contenido en todo precono positivo de K. Más aún, es claro que S_K cumple todas las condiciones de la definición de precono positivo salvo quizá la última, es decir, salvo que $-1 \notin S_K$.

Más aún, conviene observar que $S_K \setminus \{0\}$ es un subgrupo de K^*.

Es claro que $1 \in S_K \setminus \{0\}$ y que $S_K \setminus \{0\}$ es cerrado para productos, luego la única comprobación necesaria es que también es cerrado para inversos. Ahora bien, si $a \in S_K \setminus \{0\}$, entonces $1/a = a(1/a^2) \in S_K \setminus \{0\}$.

Así llegamos a la caracterización más simple posible de los cuerpos reales:

Teorema 5.66 (AE) *Sea R un cuerpo. Las afirmaciones siguientes son equivalentes:*

1. *R es real.*
2. *-1 no es suma de cuadrados en R.*
3. *No todo elemento de R es suma de cuadrados y $\operatorname{car} R \neq 2$.*
4. *Una suma de cuadrados en R es nula si y sólo si todos los sumandos son nulos.*

DEMOSTRACIÓN: 1) ⇒ 3) es trivial, pues si R es real toda suma de cuadrados es ≥ 0 y $\operatorname{car} R = 0$.

3) ⇒ 2) Si $-1 \in S_R$ y $\operatorname{car} R \neq 2$, entonces para todo $a \in R$, se cumple que

$$a = \frac{(a+1)^2 - (a-1)^2}{4} = \left(\frac{a+1}{2}\right)^2 + (-1)\left(\frac{a-1}{2}\right)^2 \in S_R,$$

luego $R = S_R$.

2) ⇒ 1) La condición $-1 \notin S_R$ es la única que falta para asegurar que S_R es un precono positivo, luego implica que R es real, por el teorema 5.64.

2) ⇒ 4) Supongamos que $a_1^2 + \cdots + a_n^2 = 0$ y no todos los sumandos son nulos. No perdemos generalidad si suponemos que ninguno lo es. Necesariamente entonces, $n \geq 2$, luego $-a_n^2 \in S_R \setminus \{0\}$, luego $-1 \in S_R \setminus \{0\}$ (porque hemos visto que $S_R \setminus \{0\}$ es un grupo), contradicción.

4) ⇒ 2) Si $-1 \in S_R$, entonces $1^2 + (-1) = 0$ es una suma de cuadrados nula cuyos términos no son nulos. ∎

Así pues, una condición necesaria y suficiente para que un cuerpo admita un orden compatible con su estructura algebraica es simplemente que -1 no pueda expresarse como suma de cuadrados.

Teorema 5.67 (AE) *Si R es un cuerpo real y $a \in R$ cumple que $R(\sqrt{a})$ no es real, entonces $-a$ es suma de cuadrados en R.*

DEMOSTRACIÓN: Observemos que a no puede ser un cuadrado en R, pues entonces $R(\sqrt{a}) = R$ sería real. Por el teorema anterior, podemos expresar

$$-1 = \sum_{j=1}^{n} (u_i + v_i\sqrt{a})^2 = \sum_{j=1}^{n} u_i^2 + a\sum_{j=1}^{n} v_i^2 + 2\sum_{j=1}^{n} u_i v_i \sqrt{a}.$$

Como $1, \sqrt{a}$ es una base de $R(\sqrt{a})$ sobre R, igualando las coordenadas vemos que

$$-1 = \sum_{j=1}^{n} u_i^2 + a\sum_{j=1}^{n} v_i^2,$$

de donde se sigue que $-a$ es cociente de sumas de cuadrados, luego es suma de cuadrados. ∎

Como consecuencia:

Teorema 5.68 (AE) *Si R es un cuerpo real y $a \in R$, entonces a es suma de cuadrados si y sólo si es positivo en toda ordenación de R.*

DEMOSTRACIÓN: Una implicación es obvia y, si a no es suma de cuadrados, el teorema anterior nos da que $R(\sqrt{-a})$ es real y en cualquier ordenación de este cuerpo sucede que $-a = (\sqrt{-a})^2 > 0$, luego $a < 0$. Un orden en $R(\sqrt{-a})$ se restringe a un orden en R en el que $a < 0$. ∎

Ahora necesitamos un resultado técnico:

Teorema 5.69 *Sea K/k una extensión finita de grado impar de cuerpos de característica 0 y consideremos $a_1, \ldots, a_n \in k$ no nulos tales que la ecuación $a_1x_1^2 + \cdots + a_nx_n^2 = 0$ tenga una solución no trivial en K (es decir, con alguna $x_i \neq 0$). Entonces existe también una solución no trivial en k.*

DEMOSTRACIÓN: Por el teorema del elemento primitivo $K = k(\alpha)$, para cierto $\alpha \in K$. Sea $g = \operatorname{pol}\operatorname{mín}(\alpha, k)$, cuyo grado será de la forma $2m+1$. Ahora usamos que $K \cong k[x]/(g)$, como se ve en la prueba del teorema 5.2. Podemos trabajar con $k[x]/(g)$, de modo que los x_i que estamos suponiendo que existen en K son clases $[f_i]$, para ciertos $f_i \in k[x]$, de modo que

$$a_1f_1(x)^2 + \cdots + a_nf_n(x)^2 = h(x)g(x).$$

Podemos exigir además que $\operatorname{grad} f_i \leq 2m$ y que su m.c.d. es 1. El polinomio del miembro izquierdo de la igualdad anterior tiene grado par $\leq 4m$, luego si $m \geq 1$ el grado de h tiene que ser impar y $\leq 2m-1$. Entonces, algún factor irreducible de h tiene que tener grado impar. Digamos que $h = h_1h_2$, donde h_1 es irreducible de grado impar. Sea β una raíz de h_1 en una extensión de k y sea $K' = k(\beta)$, que es una extensión de k de grado impar con $|K' : k| \leq 2m-1 < 2m+1 = |K : k|$.

Como h_1 no puede dividir a todos los polinomios f_i, tenemos que las clases $[f_i] \in k[x]/(h_1) \cong K'$ no son todas nulas, y determinan una solución no trivial de $a_1x_1^2 + \cdots + a_nx_n^2$ en K'. Como el grado va disminuyendo, tras un número finito de pasos tenemos que llegar a un cuerpo K en el que $m = 0$, es decir, tal que $K = k$. ∎

Teorema 5.70 (AE) *Sea R un cuerpo ordenado y K/R una extensión finita. Supongamos que se da uno de los dos casos siguientes:*

1. *Existe un $a \in R$ positivo tal que $K = R(\sqrt{a})$.*
2. *$|K : R|$ es impar.*

Entonces K es real y el orden de R se extiende a un orden en K.

DEMOSTRACIÓN: Basta probar que el conjunto P_0 formado por los elementos de la forma $\sum_{j=1}^{n} c_j\alpha_j^2$, con $c_j \in R$, $\alpha_j \in K$, $c_j > 0$ es un precono positivo en K, pues entonces estará contenido en un cono positivo que contendrá a todos los elementos positivos de R, luego determinará un orden en K que extiende al de R.

Es claro que P_0 cumple todas las condiciones de la definición de precono positivo salvo quizá que $-1 \notin P_0$, luego sólo se trata de probar que es imposible que

$$-1 = \sum_{j=1}^{n} c_j\alpha_j^2,$$

con $c_j > 0$ en R y $\alpha_j \in K$. Vamos a dar una prueba separada para cada caso. En el primero tenemos que

$$-1 = \sum_{j=1}^{n} c_j(a_j + b_j\sqrt{a})^2 = \sum_{j=1}^{n} c_j(a_j^2 + ab_j^2) + \sum_{j=1}^{n} 2a_jb_jc_j\sqrt{a},$$

luego de hecho $-1 = \sum_{j=1}^{n} c_j(a_j^2 + ab_j^2)$, pero esto es imposible, porque la suma es obviamente positiva.

En el segundo caso basta aplicar el teorema anterior: tenemos que

$$x_0^2 + \sum_{j=1}^{n} c_j x_j^2 = 0$$

tiene solución en K con $x_0 = 1$, luego también tiene solución no trivial en R, lo cual es imposible, al igual que antes. ∎

Definición 5.71 Un cuerpo R es *realmente cerrado* si es formalmente real y no tiene extensiones algebraicas formalmente reales (distintas de él mismo).

Teorema 5.72 (AE) *Si R es un cuerpo realmente cerrado, entonces admite una única relación de orden (que cumpla los axiomas de cuerpo ordenado), concretamente, la que tiene por cono positivo a R^2 (el conjunto de los cuadrados en R).*

Demostración: Si R es realmente cerrado, admite una relación de orden. Sea C su cono positivo. Obviamente $R^2 \subset C$, pero si $a \in C$, entonces el teorema 5.70 nos da que el cuerpo $R(\sqrt{a})$ es real, y obviamente es una extensión algebraica de R, luego tiene que ser $R = R(\sqrt{a})$, luego $a \in R^2$. Por lo tanto, R^2 es un cono positivo y es, de hecho, el único posible. ∎

Así pues, a partir de aquí consideraremos a todo cuerpo realmente cerrado como cuerpo ordenado con la única ordenación posible.

Teorema 5.73 (AE) *Si R es un cuerpo, las afirmaciones siguientes son equivalentes:*

1. *R es realmente cerrado.*
2. *Todo polinomio en $R[x]$ de grado impar tiene una raíz en R y R admite un orden respecto al que todo elemento positivo tiene raíz cuadrada en R.*
3. *El cuerpo $R[i]$ (donde $i^2 = -1$) es algebraicamente cerrado y contiene estrictamente a R.*

Demostración: 1) ⇒ 2) Ya hemos visto que R admite una ordenación en la que los elementos positivos son los cuadrados. Si $p(x) \in R[x]$ es un polinomio de grado impar, existe una extensión de R en la que tiene una raíz α. Consideramos entonces la extensión $R(\alpha)$, que tiene grado impar sobre R, luego el teorema 5.70 nos da que el cuerpo $R(\alpha)$ es real, luego $R = R(\alpha)$, con lo que $p(x)$ tiene una raíz en R.

2) ⇒ 3) Como R admite un orden, no puede ser que -1 tenga raíz cuadrada en R. El resto es el teorema 5.50.

3) ⇒ 1) Vamos a probar que toda suma de cuadrados en R es un cuadrado. Esto implicará que -1 no es suma de cuadrados en R, por lo que 5.66 nos dará que R es real.

Dados $a, b \in R$, tenemos que $a + bi = (c + di)^2 = c^2 - d^2 + 2cdi$, luego $a = c^2 - d^2$ y $b = 2cd$, luego

$$a^2 + b^2 = (c^2 - d^2)^2 + 4c^2d^2 = (c^2 + d^2)^2.$$

Si K/R es una extensión algebraica real, entonces adjuntando a K una raíz α de $x^2 + 1$ obtenemos otra $K(\alpha)$ de modo que $R(\alpha) \cong R(i)$, luego $R(\alpha)$ es algebraicamente cerrado, luego $K \subset K(\alpha) = R(\alpha)$, pero $R(\alpha)$ no es real, luego $K = R$. Esto prueba que R no tiene extensiones algebraicas reales, luego es realmente cerrado. ∎

Como consecuencia:

Teorema 5.74 *Si R es un cuerpo realmente cerrado, entonces todo $a \in R$ tiene raíz n-sima para n impar y, si $a > 0$, para todo n.*

Demostración: Sea $n = 2^u m$, con m impar. El polinomio $x^m - a$ tiene que tener una raíz $b \in R$. Si $a > 0$, es claro que también $b > 0$, y entonces podemos extraer sucesivamente u raíces cuadradas, lo que nos da una raíz n-sima de a. ∎

Es fácil ver que si n es impar entonces cada $a \in R$ tiene una única raíz n-sima, mientras que si es par y $a > 0$ tiene dos, de las cuales sólo una es positiva. Por lo tanto, podemos usar la notación usual $\sqrt[n]{a}$ para referirnos a la única raíz n-sima de a si n es impar o a la única raíz positiva si n es par y $a \geq 0$.

Definición 5.75 Si R es un cuerpo formalmente real, una *clausura real* de R es una extensión algebraica K/R tal que K sea realmente cerrado con un orden que extienda al de R.

Por ejemplo, el conjunto $\mathbb{R}_a$ de los números reales algebraicos es una clausura real de $\mathbb{Q}$. Más aún:

Teorema 5.76 *Toda clausura real de $\mathbb{Q}$ es isomorfa a $\mathbb{R}_a$.*

Demostración: Sea K una clausura real de $\mathbb{Q}$. Recordemos de [An 1.2] que un cuerpo ordenado K es arquimediano si $\mathbb{N}$ no está acotado en K. Vamos a probar que K es arquimediano. En caso contrario existe $\alpha \in K$ mayor que todos los números naturales, luego también mayor que todos los números racionales. Veamos por inducción sobre n que si $p(x) \in \mathbb{Q}[x]$ es un polinomio de grado n, entonces $\alpha^{n+1} > p(\alpha)$. Para $n = 0$ es precisamente la hipótesis de que α es mayor que todo elemento de $\mathbb{Q}$. Si vale para n y

$$p(x) = a_{n+1}x^{n+1} + \cdots + a_1x + a_0,$$

por la hipótesis de inducción aplicada al polinomio $(p(x) - a_0)/x + 1$,

$$\alpha^{n+1} > a_{n+1}\alpha^n + \cdots + a_2\alpha + a_1 + 1,$$

luego, multiplicando por $\alpha > 0$,

$$\alpha^{n+2} > a_{n+1}\alpha^{n+1} + \cdots + a_2\alpha^2 + a_1\alpha + \alpha > a_{n+1}\alpha^{n+1} + \cdots + a_2\alpha^2 + a_1\alpha + a_0.$$

Esto implica que α no puede ser raíz de ningún polinomio mónico de $\mathbb{Q}[x]$, en contra de que K es algebraico sobre $\mathbb{Q}$.

Por el teorema [An 1.43] tenemos que K es isomorfo (como cuerpo ordenado) a un subcuerpo $R \subset \mathbb{R}$. Entonces R es también una clausura real de $\mathbb{Q}$. En particular, como $R/\mathbb{Q}$ es algebraica, tiene que ser $R \subset \mathbb{R}_a$, pero como $\mathbb{R}_a$ es real y R no admite extensiones algebraicas reales, tiene que ser $R = \mathbb{R}_a$, luego K es isomorfo a $\mathbb{R}_a$. ∎

En general, todo cuerpo real tiene una clausura real:

Teorema 5.77 (AE) *Todo cuerpo ordenado R tiene una clausura real*[6] *cuya relación de orden extiende a la de R.*

Demostración: Sea C una clausura algebraica de R. Podemos aplicar el lema de Zorn al conjunto de todos los cuerpos ordenados $R \subset K \subset C$ tales que el orden de K extiende al de R. Obtenemos así un elemento maximal K, es decir, un cuerpo ordenado que contiene a R como subcuerpo ordenado y que no admite extensiones algebraicas dentro de C que cumplan lo mismo. Basta probar que K es realmente cerrado.

El mismo argumento del teorema 5.68 prueba que todo elemento positivo en K tiene raíz cuadrada, por lo que el orden en K es único (los elementos positivos son los cuadrados).

Es claro que si K admite una extensión real algebraica, admite una que cumple $K \subset K' \subset C$. Pero entonces el orden de K' se restringe a un orden en K, pero como el orden de K es único, de hecho el orden de K' extiende al de K, luego al de R, pero esto contradice la maximalidad de R salvo que $K' = K$. ∎

También es cierto en general que dos clausuras reales de un mismo cuerpo real R son R-isomorfas, pero la demostración tendrá que esperar hasta 6.58. Terminamos ahora con un teorema adicional sobre sumas de cuadrados:

Definición 5.78 Un número $\alpha \in \mathbb{R}_a$ es *totalmente real* si todos sus conjugados (sobre $\mathbb{Q}$) son números reales.

Por ejemplo, $\sqrt[3]{2}$ es real, pero no es totalmente real, pues sus conjugados (las otras raíces cúbicas de 2) son imaginarias.

Es claro que el conjunto $\mathbb{R}_{\text{tr}}$ de los números totalmente reales es un subcuerpo de $\mathbb{R}_a$, y de hecho la extensión $\mathbb{R}_{\text{tr}}/\mathbb{Q}$ es de Galois, pues es obvio que todo conjugado de un número totalmente real es totalmente real.

Diremos que $\alpha \in \mathbb{R}_a$ es *totalmente positivo* si todos sus conjugados son reales y positivos.

[6] Aquí usamos por segunda vez el lema de Zorn, pero una vez más es fácil ver que no se requiere AE si el cuerpo es numerable.

Teorema 5.79 *Si $K \subset \mathbb{R}_{\mathrm{tr}}$, un elemento de K es totalmente positivo si y sólo si es suma de cuadrados en K.*

DEMOSTRACIÓN: Es inmediato que si $\alpha \in K$ es suma de cuadrados, entonces es totalmente positivo, pues sus conjugados serán números reales sumas de cuadrados de números reales. Recíprocamente, si α no es suma de cuadrados, el teorema 5.68 nos da que existe un orden en K respecto al cual $\alpha < 0$. Entonces K es isomorfo (como cuerpo ordenado) a un subcuerpo de $\mathbb{R}$ (por el mismo argumento empleado en 5.76, porque K tiene que ser arquimediano), y la imagen de α por el isomorfismo es un conjugado negativo, luego α no es totalmente positivo. ∎

5.8 Extensiones ciclotómicas

En esta sección mostraremos cómo la teoría de Galois nos da un buen control sobre los cuerpos ciclotómicos. Más en general, aprovechamos la ocasión para introducir el concepto de extensión ciclotómica de un cuerpo arbitrario y de un orden arbitrario, no necesariamente primo.

Definición 5.80 Llamaremos *extensión ciclotómica* n-sima de un cuerpo k al cuerpo de escisión sobre k del polinomio $x^n - 1$.

Si $\operatorname{car} k = p$ y $n = p^u m$ con $(m, p) = 1$, entonces $x^n - 1 = (x^m - 1)^{p^u}$, lo que implica que el cuerpo de escisión de $x^n - 1$ es el mismo que el de $x^m - 1$, o en otros términos, que la extensión ciclotómica n-sima coincide con la extensión ciclotómica m-sima. Por esta razón podemos restringirnos al caso en el que $\operatorname{car} k \nmid n$ (incluyendo el caso $\operatorname{car} k = 0$).

Sea K/k una extensión ciclotómica n-sima tal que $\operatorname{car} k \nmid n$. Entonces la derivada del polinomio $x^n - 1$ es $nx^{n-1} \neq 0$, y la única raíz de este polinomio es 0, que no es raíz de $x^n - 1$. Por lo tanto las raíces de $x^n - 1$ en K son todas simples (separables) y hay n de ellas. Así pues, toda extensión ciclotómica es finita de Galois.

Las raíces del polinomio $x^n - 1$ en un cuerpo cualquiera se llaman *raíces n-simas de la unidad*. En una extensión ciclotómica n-sima (bajo las hipótesis indicadas) hay n raíces n-simas de la unidad.

Es obvio que el producto de dos raíces n-simas de la unidad vuelve a ser una raíz n-sima, así como que el inverso de una raíz n-sima es también una raíz n-sima. Esto significa que el conjunto de las raíces n-simas de la unidad en un cuerpo cualquiera forman un subgrupo finito del grupo multiplicativo del cuerpo. Por el teorema 4.50 se trata de un grupo cíclico.

Así pues, si K/k es una extensión ciclotómica n-sima tal que $\operatorname{car} k \nmid n$, el conjunto de las raíces n-simas de la unidad es un grupo cíclico de orden n. A los elementos de orden n (o sea, a los generadores) se les llama *raíces n-simas primitivas* de la unidad. Por [TG 1.16] hay exactamente $\phi(n)$ de ellas, donde ϕ es la función de Euler. Así, si ω es una raíz n-sima primitiva de la unidad, las raíces restantes son $1, \omega, \omega^2, \ldots, \omega^{n-1}$. Obviamente, $K = k(\omega)$.

Llamaremos *polinomio ciclotómico* n-simo al polinomio

$$c_n(x) = (x - \omega_1) \cdots (x - \omega_m),$$

donde $\omega_1, \ldots, \omega_m$ son las raíces n-simas primitivas de la unidad en K. Notemos que $c_n(x)$ es mónico y $\operatorname{grad} c_n(x) = m = \phi(n)$.

Si K/k es una extensión ciclotómica n-sima, $\omega \in K$ es una raíz n-sima primitiva de la unidad y P es el cuerpo primo de k, entonces $c_n(x) \in P[x]$. En efecto, la extensión $P(\omega)/P$ es también ciclotómica n-sima y, por definición, el polinomio $c_n(x)$ para la extensión $P(\omega)/P$ es el mismo que para K/k. Los automorfismos de $P(\omega)$ permutan las raíces primitivas, luego sus extensiones a $P(\omega)[x]$ dejan invariante a $c_n(x)$ (permutan sus factores), pero esto es lo mismo que decir que dejan invariantes a sus coeficientes y, por lo tanto, estos coeficientes están en $F\big(G(P(\omega)/P)\big) = P$.

Con esto hemos probado que los polinomios $c_n(x)$ pueden obtenerse siempre a partir de una extensión ciclotómica de un cuerpo primo P. Como dos cuerpos de escisión de un mismo polinomio sobre P son P-isomorfos, en realidad $c_n(x)$ no depende de la extensión K de P que tomemos. En resumen, que hay un único polinomio $c_n(x)$ para cada cuerpo primo, o dicho de otro modo, si K es cualquier cuerpo en el que exista una raíz n-sima primitiva de la unidad, es decir, una raíz n-sima de la unidad de orden n, entonces el polinomio cuyas raíces (simples) son todas las raíces n-simas primitivas de la unidad en K es $c_n(x)$, un polinomio que no depende más que de la característica de K.

Pronto probaremos que, esencialmente, el polinomio $c_n(x)$ tampoco depende de la característica.

Observemos que $c_1(x) = x - 1$ y $c_2(x) = x + 1$ (pues -1 es la única raíz cuadrada primitiva de la unidad). El teorema siguiente nos permite calcular fácilmente los polinomios ciclotómicos.

Teorema 5.81 *Sea k un cuerpo tal que $\operatorname{car} k \nmid n$. Entonces*

$$x^n - 1 = \prod_{d|n} c_d(x).$$

Demostración: Sea K/k una extensión ciclotómica n-sima. Para cada divisor d de n sea A_d el conjunto de las raíces de la unidad de orden d. De este modo $\{A_d\}_{d|n}$ es una partición del conjunto de las raíces n-simas de la unidad, es decir, una partición del conjunto de las raíces de $x^n - 1$, y los elementos de cada A_d son las raíces d-ésimas primitivas de la unidad en K, luego $c_d(x)$ tiene por raíces a los elementos de A_d. A partir de aquí el teorema es inmediato. ∎

Por lo tanto,

$$c_n(x) = \frac{x^n - 1}{\prod\limits_{\substack{d|n \\ d \neq n}} c_d(x)}$$

Tabla 5.1: Polinomios ciclotómicos

$c_1(x)$	$=$	$x-1$
$c_2(x)$	$=$	$x+1$
$c_3(x)$	$=$	x^2+x+1
$c_4(x)$	$=$	x^2+1
$c_5(x)$	$=$	$x^4+x^3+x^2+x+1$
$c_6(x)$	$=$	x^2-x+1
$c_7(x)$	$=$	$x^6+x^5+x^4+x^3+x^2+x+1$
$c_8(x)$	$=$	x^4+1
$c_9(x)$	$=$	x^6+x^3+1
$c_{10}(x)$	$=$	$x^4-x^3+x^2-x+1$

Así podemos calcular recurrentemente los polinomios ciclotómicos:

$$\begin{aligned} c_3(x) &= \frac{x^3-1}{x-1} = x^2+x+1, \\ c_6(x) &= \frac{x^6-1}{(x-1)(x+1)(x^2+x+1)} = x^2-x+1. \end{aligned}$$

La tabla 5.1 contiene los primeros polinomios ciclotómicos. El lector observará sin duda que los coeficientes no nulos de los polinomios ciclotómicos son ± 1. Así se cumple hasta llegar al polinomio ciclotómico de orden 104, en cambio, éste es el centésimo quinto polinomio ciclotómico:

$$\begin{aligned} c_{105}(x) &= x^{48}+x^{47}+x^{46}-x^{43}-x^{42}-2x^{41}-x^{40}-x^{39}+x^{36}+x^{35} \\ &+ x^{34}+x^{33}+x^{32}+x^{31}-x^{28}-x^{26}-x^{24}-x^{22}-x^{20}+x^{17}+x^{16} \\ &+ x^{15}+x^{14}+x^{13}+x^{12}-x^9-x^8-2x^7-x^6-x^5+x^2+x+1. \end{aligned}$$

Como vemos, tiene dos coeficientes iguales a -2. Puede probarse que existen polinomios ciclotómicos con coeficientes tan grandes en módulo como se quiera. Lo que sí se cumple siempre es que los coeficientes son enteros:

Teorema 5.82 *Los polinomios ciclotómicos sobre $\mathbb{Q}$ tienen los coeficientes enteros.*

DEMOSTRACIÓN: Por inducción. Para $n=1$ tenemos $c_1(x)=x-1$.
Supongamos que $c_m(x) \in \mathbb{Z}[x]$ para todo $m<n$. Sea

$$q(x) = \prod_{\substack{d|n \\ d\neq n}} c_d(x).$$

Por hipótesis de inducción $q(x) \in \mathbb{Z}[x]$ y es mónico. Por el teorema anterior $x^n-1 = c_n(x)q(x)$.

En $\mathbb{Z}[x]$ existen polinomios $p(x)$ y $r(x)$ tales que $x^n - 1 = q(x)p(x) + r(x)$, donde el grado de $r(x)$ es menor que el de $q(x)$, pero esto también es cierto en $\mathbb{Q}[x]$ y, por la unicidad de la división euclídea, ha de ser $p(x) = c_n(x)$ y $r(x) = 0$, o sea, $c_n(x) \in \mathbb{Z}[x]$. ∎

Notemos que las divisiones necesarias para calcular $c_n(x)$ según el teorema 5.81 se pueden hacer como si los polinomios fueran de $\mathbb{Z}[x]$ aunque en realidad sean de $(\mathbb{Z}/p\mathbb{Z})\,[x]$, lo que significa que los polinomios ciclotómicos de característica p (cuando existen) son los mismos que los de característica 0, pero considerando a sus coeficientes en $\mathbb{Z}/p\mathbb{Z}$. En definitiva, $c_n(x)$ es esencialmente único.

En el capítulo III hemos visto que $c_p(x) = x^{p-1} + \cdots + x + 1$ es irreducible en $\mathbb{Q}[x]$. Vamos a ver que en realidad esto vale para todos los polinomios ciclotómicos sobre $\mathbb{Q}$.

Teorema 5.83 *El polinomio $c_n(x)$ es irreducible en $\mathbb{Q}[x]$.*

Demostración: Por el criterio de Gauss es suficiente probar que es irreducible en $\mathbb{Z}[x]$. Sea $f(x)$ un factor mónico irreducible no constante de $c_n(x)$ en $\mathbb{Z}[x]$. Hemos de probar que $f(x) = c_n(x)$. Sea ω una raíz n-sima primitiva de la unidad tal que $f(\omega) = 0$. Sea p un primo tal que $p \le n$ y $(p, n) = 1$. Veamos que ω^p es raíz de $f(x)$.

Sea $c_n(x) = f(x)g(x)$, con $f(x)$, $g(x) \in \mathbb{Z}[x]$. Por el teorema [TG 1.5], el orden de ω^p es también n, o sea, ω^p es otra raíz n-sima primitiva de la unidad y, en consecuencia, es raíz de $c_n(x)$. Si no fuera raíz de $f(x)$ lo sería de $g(x)$, o sea, $g(\omega^p) = 0$. Entonces ω es raíz del polinomio $g(x^p)$ y, como $f(x) = \operatorname{pol\,mín}(\omega, \mathbb{Q})$, se ha de cumplir $f(x) \mid g(x^p)$.

Sea $g(x^p) = f(x)h(x)$. Dividiendo euclídeamente en $\mathbb{Z}[x]$ y por la unicidad de la división en $\mathbb{Q}[x]$ podemos concluir que $h(x) \in \mathbb{Z}[x]$.

Ahora tomamos clases módulo p en los coeficientes de los polinomios, con lo que $\big[g(x^p)\big] = \big[f(x)\big]\,\big[h(x)\big]$. Pero todo $u \in \mathbb{Z}/p\mathbb{Z}$ cumple $u^p = u$ (pues si $u \neq 0$ pertenece al grupo multiplicativo, de orden $p-1$, luego $u^{p-1} = 1$), y esto nos permite extraer el exponente: $\big[g(x^p)\big] = \big[g(x)\big]^p$. Todo factor irreducible de $\big[f(x)\big]$ divide a $\big[g(x)\big]^p$, luego a $\big[g(x)\big]$.

De aquí llegamos a una contradicción, pues $x^n - 1 = c_n(x)s(x)$, para un cierto polinomio en $\mathbb{Q}[x]$. De nuevo por la unicidad de la división euclídea, de hecho $s(x) \in \mathbb{Z}[x]$, luego luego $x^n - [1] = \big[f(x)\big]\,\big[g(x)\big]\,\big[s(x)\big]$, ahora bien, sabemos que el polinomio $x^n - [1]$ debe escindirse en factores distintos, pero por otro lado los polinomios $\big[f(x)\big]$ y $\big[g(x)\big]$ tienen factores comunes.

Con esto hemos probado que ω^p es raíz de $f(x)$ para todo primo $p < n$ con $(p, n) = 1$.

Si $(m, n) = 1$ y $m < n$, entonces los primos en los que se descompone m son menores que n y primos con n, luego aplicando repetidas veces lo anterior llegamos a que ω^m es también raíz de $f(x)$, pero por el teorema [TG 1.5] toda raíz primitiva es de la forma ω^m con $(n, m) = 1$, luego $f(x)$ tiene todas las raíces de $c_n(x)$ y, en consecuencia, $f(x) = c_n(x)$. ∎

Tras estos resultados generales, pasamos a estudiar los grupos de Galois de las extensiones ciclotómicas. En el teorema siguiente incluimos algunos hechos básicos que ya hemos usado y probado.

Teorema 5.84 *Sea K/k una extensión ciclotómica n-sima, tal que* $\operatorname{car} k \nmid n$. *Entonces:*

1. *La extensión K/k es finita de Galois.*
2. *Si $\omega \in K$ es una raíz n-sima primitiva de la unidad, $K = k(\omega)$.*
3. $\operatorname{pol\,mín}(\omega, k) \mid c_n(x)$.
4. *$G(K/k)$ es isomorfo a un subgrupo del grupo U_n de las unidades de $\mathbb{Z}/n\mathbb{Z}$.*
5. *El polinomio $c_n(x)$ es irreducible en $k[x]$ si y sólo si $G(K/k) \cong U_n$, y en tal caso el grado de la extensión es $\phi(n)$.*

Demostración: 1), 2) y 3) ya están probados. Para probar 4) fijamos una raíz n-sima primitiva de la unidad ω. Para cada $\sigma \in G(K/k)$ es claro que $\sigma(\omega)$ ha de ser otra raíz n-sima primitiva, que será de la forma ω^m, con $(m, n) = 1$. Como el orden de ω es n, el número m sólo está determinado módulo n, o lo que es lo mismo, la clase $[m] \in U_n$ está unívocamente determinada por σ.

Sea $\phi : G(K/k) \longrightarrow U_n$ que a cada automorfismo σ le asigna la clase $[m]$ del modo descrito. Así, si $\phi(\sigma) = [m]$, entonces $\sigma(\omega) = \omega^m$. Es fácil ver que ϕ es un homomorfismo de grupos. Además si $\phi(\sigma) = [1]$ entonces $\sigma(\omega) = \omega$, luego $\sigma = 1$. Por lo tanto ϕ es inyectivo y $G(K/k)$ es isomorfo a un subgrupo de U_n.

5) El polinomio $c_n(x)$ es irreducible en $k[x]$ si y sólo si es el polinomio mínimo de las raíces primitivas (por 3), si y sólo si $|K : k| = \phi(n)$ (por 2), si y sólo si $\big|G(K/k)\big| = \big|U_n\big|$, si y sólo si $G(K/k) \cong U_n$ (por 4). ∎

En particular todas las extensiones ciclotómicas son abelianas[7] y las de orden primo son cíclicas (pues los grupos U_p son cíclicos, por 4.50).

Ejemplo Vamos a estudiar el cuerpo $K = \mathbb{Q}(\omega)$, donde ω es una raíz séptima primitiva de la unidad. Esto significa que ω cumple la relación

$$\omega^6 + \omega^5 + \omega^4 + \omega^3 + \omega^2 + \omega + 1 = 0.$$

Una $\mathbb{Q}$-base de K está formada por $1, \omega, \omega^2, \omega^3, \omega^4, \omega^5$, aunque también podemos considerar la base $\omega, \omega^2, \omega^3, \omega^4, \omega^5, \omega^6$ (es la imagen de la anterior por el automorfismo de espacios vectoriales dado por $\alpha \mapsto \omega\alpha$).

Sabemos que $K/\mathbb{Q}$ es una extensión finita de Galois con grupo de Galois $G = G(K/\mathbb{Q}) \cong U_7$. Un generador de U_7 es la clase $[3]$, por lo que $G = \langle\sigma\rangle$, donde el automorfismo σ está determinado por que $\sigma(\omega) = \omega^3$.

En general, cuando tenemos una extensión simple de Galois con un elemento primitivo ω, si llamamos Ω al conjunto de los conjugados de ω, que en este caso

[7] En general, se dice que una extensión de Galois es abeliana, cíclica, etc. si lo es su grupo de Galois.

son las potencias $\Omega = \{\omega, \omega^2, \omega^3, \omega^4, \omega^5, \omega^6\}$, el grupo de Galois G actúa sobre Ω, de modo que la restricción $G \longrightarrow \Sigma_\Omega$, es decir, la aplicación $g \mapsto g|_\Omega$, es un monomorfismo de grupos que nos permite ver a G como grupo de permutaciones. En este caso, el generador σ se corresponde con el ciclo

$$\sigma = (\omega, \omega^3, \omega^2, \omega^6, \omega^4, \omega^5).$$

El grupo G tiene dos subgrupos propios de órdenes 2 y 3, que son $\langle\sigma^3\rangle$ y $\langle\sigma^2\rangle$. Llamemos K_3 y K_2 a sus cuerpos fijados respectivos, que tiene grados 3 y 2 sobre $\mathbb{Q}$, respectivamente.

Vista como permutación, tenemos que $\sigma^2 = (\omega, \omega^2, \omega^4)(\omega^3, \omega^6, \omega^5)$, lo que se traduce en que

$$\eta_1 = \omega + \omega^2 + \omega^4, \quad \eta_2 = \omega^3 + \omega^5 + \omega^6$$

están en el cuerpo K_2 fijado por σ^2. El hecho de que las potencias de ω sean linealmente independientes se traduce en que η_1 y η_2 también lo son, luego son una base de K_2. Además, $\sigma(\eta_1) = \eta_2$, luego son conjugados, es decir, tienen el mismo polinomio mínimo. Éste será

$$(x - \eta_1)(x - \eta_2) = x^2 - (\eta_1 + \eta_2) + \eta_1\eta_2.$$

Claramente $\eta_1 + \eta_2 = -1$, mientras que

$$\eta_1\eta_2 = (\omega^4 + \omega^2 + \omega)(\omega^6 + \omega^5 + \omega^3) = \omega^3 + \omega^2 + 1 + \omega + 1 + \omega^5 + 1 + \omega^6 + \omega^4 = 2,$$

luego $\operatorname{pol\,mín}(\eta_1, \mathbb{Q}) = x^2 + x + 2$ y así

$$\eta_i = \frac{-1 \pm \sqrt{-7}}{2},$$

luego $K_2 = \mathbb{Q}(\eta_i) = \mathbb{Q}\big[\sqrt{-7}\,\big]$.

Similarmente, vemos que $\sigma^3 = (\omega, \omega^6)(\omega^2, \omega^5)(\omega^3, \omega^4)$, por lo que una base de K_3 la forman

$$\rho_1 = \omega + \omega^6, \quad \rho_2 = \omega^2 + \omega^5, \quad \rho_3 = \omega^3 + \omega^4.$$

Además $\sigma(\rho_1) = \rho_2$ y $\sigma(\rho_2) = \rho_3$, por lo que los tres son conjugados. Para calcular su polinomio mínimo se comprueba sin dificultad que

$$\rho_1^2 = \rho_2 + 2, \qquad \rho_1^3 = \rho_3 + 3\rho_1,$$

luego, teniendo en cuenta que $\rho_1 + \rho_2 + \rho_3 + 1 = 0$, resulta que

$$\rho_1^3 + \rho_1^2 - 2\rho_1 - 1 = 0,$$

luego el polinomio mínimo es $x^3 + x^2 - 2x - 1$. Como K_3 no tiene subcuerpos intermedios, tiene que ser $K_3 = \mathbb{Q}(\rho_i)$ para cualquier i. Si tomamos $\rho = \rho_1$, resulta que $K_3 = \mathbb{Q}(\rho)$, de modo que cada elemento de K_3 se expresa de forma

única como $a\rho^2+b\rho+c$, con $a,b,c\in\mathbb{Q}$, y el producto está determinado por la relación $\rho^3=-\rho^2+2\rho+1$. Según hemos visto, los conjugados de ρ son

$$\rho_2=\rho^2-2,\qquad \rho_3=\rho^3-3\rho=-\rho^2-\rho+1.$$

■

Ejemplo Ahora vamos a estudiar el cuerpo $K=\mathbb{Q}(\omega)$, donde ω es una raíz novena primitiva de la unidad. Según la tabla de polinomios ciclotómicos, esto significa que $\omega^6+\omega^3+1=0$. Ahora

$$G=G(K/\mathbb{Q})\cong U_9=\{\bar1,\bar2,\bar4,\bar5,\bar7,\bar8\}=\langle\bar2\rangle\,,$$

luego $G=\langle\sigma\rangle$, donde $\sigma(\omega)=\omega^2$. Como permutación de los conjugados de ω es

$$\sigma=(\omega,\omega^2,\omega^4,\omega^8,\omega^7,\omega^5).$$

De nuevo G es un grupo cíclico de orden 6 con dos subgrupos propios, de órdenes 2 y 3, que se corresponden con dos subcuerpos intermedios K_3 y K_2 de grados 3 y 2, respectivamente, sobre $\mathbb{Q}$. Ahora bien, es inmediato que ω^3 es una raíz cúbica primitiva de la unidad, cuyo polinomio mínimo es x^2+x+1, luego $K_2=\mathbb{Q}(\omega^3)=\mathbb{Q}\big[\sqrt{-3}\,\big]$.

Para calcular K_3, que es el cuerpo fijado por $\sigma^3=(\omega,\omega^8)(\omega^2,\omega^7)(\omega^4,\omega^5)$, por lo que una base de K_3 está formada por

$$\begin{aligned}\rho_1&=\omega+\omega^8=\omega+\omega^2(-\omega^4-1)=\omega-\omega^2-\omega^5,\\ \rho_2&=\omega^2+\omega^7=\omega^2+\omega(-\omega^3-1)=-\omega+\omega^2-\omega^4,\\ \rho_3&=\omega^4+\omega^5.\end{aligned}$$

Vemos así que $\rho_1+\rho_2+\rho_3=0$, y un cálculo laborioso muestra que

$$\rho_1\rho_2\rho_3=-1,\qquad \rho_1\rho_2+\rho_1\rho_3+\rho_2\rho_3=-3,$$

de donde se sigue que el polinomio mínimo de los ρ_i es x^3-3x+1. Así pues, llamando $\rho=\rho_1$, tenemos que $K_3=\mathbb{Q}(\rho)$, donde $\rho^3=3\rho-1$. Cada elemento de K_3 se expresa de forma única como $a\rho^2+b\rho+c$, donde $a,b,c\in\mathbb{Q}$. Por ejemplo, $\rho^2=\omega^2+\omega^7+2=\rho_2+2$, por lo que

$$\rho_2=\rho^2-2,\qquad \rho_3=-\rho_1-\rho_2=-\rho^2-\rho+2.$$

■

El teorema siguiente es un ejemplo sencillo de cómo la teoría de Galois transforma un problema de cuerpos en otro más manejable sobre grupos finitos.

Teorema 5.85 *Para cada número natural n sea ω_n una raíz n-sima primitiva de la unidad. Sean a y b dos números naturales, $d=\mathrm{mcd}(a,b)$ y $m=\mathrm{mcm}(a,b)$. Entonces*

$$\mathbb{Q}(\omega_a)\mathbb{Q}(\omega_b)=\mathbb{Q}(\omega_m)\qquad y\qquad \mathbb{Q}(\omega_a)\cap\mathbb{Q}(\omega_b)=\mathbb{Q}(\omega_d).$$

DEMOSTRACIÓN: Es claro que $\mathbb{Q}(\omega_d)$, $\mathbb{Q}(\omega_a)$ y $\mathbb{Q}(\omega_b)$ están contenidos en $\mathbb{Q}(\omega_m)$, pues ω_d, ω_a y ω_b son potencias de ω_m. Para $i = d$, a, b, m, llamemos $H_i = G\big(\mathbb{Q}(\omega_m)/\mathbb{Q}(\omega_i)\big)$. Si identificamos $G\big(\mathbb{Q}(\omega_m)/\mathbb{Q}\big) \cong U_m$, de acuerdo con el teorema 5.84, entonces

$$H_i = \big\{[x] \in U_m \;\big|\; x \equiv 1 \text{ (mód } i)\big\},$$

pues $[x] \in H_i$ si y sólo si $\omega_i^x = \omega_i$ si y sólo si $x \equiv 1$ (mód i).

Hemos de probar que $H_a \cap H_b = H_m = 1$ y que $H_aH_b = H_d$. Ahora bien, es obvio que $x - 1$ es múltiplo de a y b si y sólo si es múltiplo de m, lo que nos da la primera igualdad.

A su vez esto implica que $|H_aH_b| = |H_a|\,|H_b|$ (por el teorema [TG 1.37]). Teniendo en cuenta que $|H_i| = \big|\mathbb{Q}(\omega_m) : \mathbb{Q}(\omega_i)\big| = \phi(m)/\phi(i)$, concluimos que

$$|H_aH_b| = \frac{\phi(m)}{\phi(a)}\frac{\phi(m)}{\phi(b)} = \frac{\phi(m)}{\phi(d)} = |H_d|.$$

Como $H_aH_b \le H_d$, la igualdad de órdenes implica $H_aH_b = H_d$. ∎

Ejercicio: Probar que $\omega_a\omega_b$ es una raíz m-sima primitiva de la unidad. Deducir directamente la primera igualdad del teorema anterior.

Ejercicio: Probar que si m es impar entonces $\mathbb{Q}(\omega_m) = \mathbb{Q}(\omega_{2m})$.

Cuando el cuerpo base es $\mathbb{Z}/p\mathbb{Z}$ la situación es la siguiente:

Teorema 5.86 *Sea p un número primo y $n = p^a m \ge 3$, donde $p \nmid m$ un número natural. Entonces la descomposición en factores irreducibles de imagen de $c_n(x)$ en $(\mathbb{Z}/p\mathbb{Z})[x]$ es de la forma $(p_1(x)\cdots p_r(x))^e$, donde $e = \phi(p^a)$ y cada factor $p_i(x)$ tiene grado $o_m(p)$ (el orden de p en el grupo de unidades U_m).*

DEMOSTRACIÓN: Supongamos en primer lugar que $p \nmid m$, sea $p(x)$ un factor irreducible de $c_n(x)$ de grado d y sea K el cuerpo que resulta de adjuntar a $k = \mathbb{Z}/p\mathbb{Z}$ una raíz ω de $p(x)$. Entonces $|K : k| = d$, luego $|K| = p^d$, luego el grupo de unidades K^* tiene $p^d - 1$ elementos, y contiene un elemento ω de orden m, luego $m \mid p^d - 1$, es decir, $o_m(p) \mid d$.

Sea $d' = o_m(p)$. Como el grupo $G(K/k)$ es abeliano, por 4.52 tiene subgrupos de todos los órdenes que dividen a su orden, luego existe una extensión intermedia[8] $k \subset L \subset K$ tal que $|L : k| = d'$, con lo que $|L| = p^{d'}$. Según el teorema 4.50 el grupo de unidades L^* es cíclico de orden $p^{d'} - 1$ y $m \mid p^{d'} - 1$, luego L^* tiene un elemento ω' de orden m, es decir, una raíz m-sima primitiva de la unidad, luego L^* contiene a todas las raíces m-simas de la unidad, en particular a todas las raíces primitivas, luego toda extensión ciclotómica de k está contenida en L y tiene grado divisor de d'. En particular $d \mid d'$ y tenemos que $d = d' = o_m(p)$.

[8]Más fácilmente, en 9.2 probaremos que existe un cuerpo de $p^{d'}$ elementos, y esto es lo único que requiere la prueba.

Así pues, todos los factores irreducibles de $c_n(x)$ en $(\mathbb{Z}/p\mathbb{Z})[x]$ tienen grado $o_m(p)$, y tienen multiplicidad 1 porque $c_n(x) \mid x^n - 1$ y $x^n - 1$ tiene sus raíces simples. Esto prueba el teorema cuando $p \nmid n$.

El caso general se reduce al caso anterior si probamos que $c_n(x) = c_m(x)^{\phi(p^a)}$. Lo razonamos por inducción sobre n. Para $n = 1$ es trivial. Supuesto cierto para todo $n' < n$, tenemos que

$$x^n - 1 = \prod_{d|m}\prod_{p^i|p^a} c_{dp^i}(x) = \prod_{\substack{d|m\\ d\neq m}}\prod_{p^i|p^a} c_d(x)^{\phi(p^i)} \prod_{p^i|p^{a-1}} c_m(x)^{\phi(p^i)}\ c_n(x)$$

$$= \prod_{\substack{d|m\\ d\neq m}} c_d(x)^{\sum_{i=0}^{a}\phi(p^i)} c_m(x)^{\sum_{i=0}^{a-1}\phi(p^i)} c_n(x) = \prod_{\substack{d|m\\ d\neq m}} c_d(x)^{p^a} c_m(x)^{p^{a-1}} c_n(x),$$

donde hemos calculado

$$\sum_{i=0}^{a}\phi(p^i) = 1 + \sum_{i=1}^{a}(p-1)p^{i-1} = 1 + (p-1)\frac{p^a-1}{p-1} = p^a.$$

Pero, por otra parte,

$$x^n - 1 = (x^m-1)^{p^a} = \Big(\prod_{d|m} c_d(x)\Big)^{p^a}$$

y comparando ambas expresiones obtenemos que $c_m(x)^{p^{a-1}}c_n(x) = c_m(x)^{p^a}$, luego $c_n(x) = c_m(x)^{p^a - p^{a-1}} = c_m(x)^{\phi(p^a)}$. ∎

Por ejemplo, la tabla siguiente muestra la descomposición en factores primos de $c_9(x)$ en distintos cuerpos $\mathbb{Z}/p\mathbb{Z}$, y en ella podemos comprobar que los distintos casos se dan según lo que establece el teorema anterior:

p	p (mód 9)	$o_9(p)$	$c_9(x)$
2	2	6	x^6+x^3+1
3	3	–	$(x+1)^6$
5	5	6	x^6+x^3+1
7	7	3	$(x^3+3)(x^3+5)$
11	2	6	x^6+x^3+1
13	4	3	$(x^3+4)(x^3+10)$
17	8	2	$(x^2+3x+1)(x^2+4x+1)(x^2+10x+1)$
19	1	1	$(x+2)(x+3)(x+10)(x+13)(x+14)(x+15)$

Capítulo VI

Álgebra lineal

Presentamos aquí una serie de conceptos, técnicas y resultados relacionados con módulos y espacios vectoriales que constituyen el núcleo de lo que se conoce como álgebra lineal.

6.1 Determinantes

En [ITAl 11.6] definimos el concepto de determinante de una matriz 2×2, y vimos que una matriz tiene inversa si y sólo si su determinante es una unidad (en el caso de matrices sobre un cuerpo, si y sólo si su determinante no es nulo). Esto se traducía en que al definir en [ITAl 11.16] la equivalencia de formas cuadráticas con coeficientes enteros era fundamental exigir que los cambios de variable considerados tuvieran determinante ± 1. Los determinantes nos han aparecido en los contextos más diversos, como al definir la norma de un módulo completo en un cuerpo cuadrático en [ITAl 12.12], o la definir la orientación de bases en [ITAl 12.18] o en [IGE A.8]. En [ITAn A.16] nos aparecieron determinantes en un contexto muy distinto, al estudiar la relación entre la medida de Jordan de un conjunto y la de su imagen por una aplicación lineal. Finalmente, en [IC A.7] introdujimos los determinantes de matrices 3×3 y los interpretamos (en el caso de matrices con coeficientes reales) como el volumen del paralelepípedo determinado por las tres filas de la matriz.

En realidad no es evidente que los determinantes 3×3 definidos en [IC] tengan ninguna relación con los determinantes 2×2 definidos previamente, pero ahora vamos a ver que ambas definiciones son casos particulares de una definición general mucho más natural. En efecto, para obtener de forma natural la definición de determinante empezaremos estableciendo las propiedades que deseamos que cumplan los determinantes y concluiremos que la única definición posible es la que vamos a adoptar.

Definición 6.1 Sea A un dominio y n un número natural no nulo. Entonces A^n es un A-módulo libre de rango n. Una aplicación $f : (A^n)^n \longrightarrow A$ es una *forma multilineal* si para todos los elementos $v_1, \ldots, v_n, v' \in A^n$, todos los $a, a' \in A$ y

todo índice $1 \leq i \leq n$ se cumple

$$f(v_1, \ldots, av_i + a'v', \ldots, v_n) = af(v_1, \ldots, v_i, \ldots, v_n) + a'f(v_1, \ldots, v', \ldots, v_n).$$

Una forma multilineal f es *antisimétrica* si cuando la n-tupla $x \in (A^n)^n$ resulta de intercambiar el orden de dos componentes de la n-tupla $y \in (A^n)^n$, entonces $f(x) = -f(y)$.

Una forma multilineal f es *alternada* si toma el valor 0 sobre todas las n-tuplas que tienen dos componentes iguales.

Antes de discutir estos conceptos conviene destacar algunas consecuencias sencillas de la definición:

Teorema 6.2 *Sea A un anillo conmutativo y unitario y $f : (A^n)^n \longrightarrow A$ una forma multilineal.*

1. *La forma f es antisimétrica si y sólo si para toda permutación $\sigma \in \Sigma_n$ y todos los $v_1, \ldots, v_n \in A^n$ se cumple*

$$f(v_{\sigma(1)}, \ldots, v_{\sigma(n)}) = (\operatorname{sig}\sigma) f(v_1, \ldots, v_n)$$

2. *Si f es alternada, entonces es antisimétrica.*

Demostración: 1) Si f cumple esta propiedad es antisimétrica, pues al intercambiar dos elementos estamos aplicando una trasposición y las trasposiciones tienen signatura -1. Si f es antisimétrica, el valor de $f(v_{\sigma(1)}, \ldots, v_{\sigma(n)})$ puede obtenerse aplicando sucesivas trasposiciones sobre $f(v_1, \ldots, v_n)$, que cambiarán el signo de f tantas veces como trasposiciones compongan a σ. Por lo tanto el resultado final será $(\operatorname{sig}\sigma) f(v_1, \ldots, v_n)$.

2) Supongamos que f es alternada y sean $1 \leq i < j \leq n$, $v_1, \ldots, v_n \in A^n$. Entonces

$$\begin{aligned}
0 &= f(v_1, \ldots, \overset{(i)}{v_i + v_j}, \ldots, \overset{(j)}{v_i + v_j}, \ldots v_n) \\
&= f(v_1, \ldots, \overset{(i)}{v_i}, \ldots, \overset{(j)}{v_i}, \ldots v_n) + f(v_1, \ldots, \overset{(i)}{v_i}, \ldots, \overset{(j)}{v_j}, \ldots v_n) \\
&+ f(v_1, \ldots, \overset{(i)}{v_j}, \ldots, \overset{(j)}{v_i}, \ldots v_n) + f(v_1, \ldots, \overset{(i)}{v_j}, \ldots, \overset{(j)}{v_j}, \ldots v_n) \\
&= 0 + f(v_1, \ldots, \overset{(i)}{v_i}, \ldots, \overset{(i)}{v_j}, \ldots v_n) + f(v_1, \ldots, \overset{(i)}{v_j}, \ldots, \overset{(j)}{v_i}, \ldots v_n) + 0,
\end{aligned}$$

luego $f(v_1, \ldots, v_j, \ldots, v_i, \ldots v_n) = -f(v_1, \ldots, v_i, \ldots, v_j, \ldots v_n)$. ∎

Ejercicio: Probar que si $\operatorname{car} A \neq 2$ entonces una forma multilineal sobre $(A^n)^n$ es antisimétrica si y sólo si es alternada.

De este modo vemos que los conceptos de forma antisimétrica y forma alternada son casi equivalentes. El segundo nos evita algunos problemas que surgen cuando puede ocurrir $x = -x$ sin que x sea 0, pero en tal caso la teoría que vamos a desarrollar es de poca utilidad. Ahora probamos que siempre existe una forma multilineal alternada en $(A^n)^n$, y que es esencialmente única, lo que dará pie a la definición de la función determinante.

Teorema 6.3 *Sea A un dominio y n un número natural no nulo. Sea $e_1, \ldots, e_n$ la base canónica de A^n y sea $a \in A$. Existe una única forma multilineal alternada $f : (A^n)^n \longrightarrow A$ tal que $f(e_1, \ldots, e_n) = a$.*

DEMOSTRACIÓN: Supongamos que existe f y veamos que es única. De este modo obtendremos la forma que ha de tener y podremos construirla.

Sea $(v_1, \ldots, v_n) \in (A^n)^n$. Para cada $i = 1, \ldots, n$ sea

$$v_i = (a_{i1}, \ldots, a_{in}) = \sum_{j=1}^{n} a_{ij} e_j.$$

Por la multilinealidad,

$$\begin{aligned} f(v_1, \ldots, v_n) &= f\left(\sum_{j_1=1}^{n} a_{1j_1} e_{j_1}, \ldots, \sum_{j_n=1}^{n} a_{nj_n} e_{j_n} \right) \\ &= \sum_{j_1=1}^{n} \cdots \sum_{j_n=1}^{n} a_{1j_1} \cdots a_{nj_n} f(e_{j_1}, \ldots, e_{j_n}). \end{aligned}$$

Como f es alternada, todas las asignaciones $k \mapsto j_k$ que no sean biyecciones harán que $f(e_{j_1}, \ldots, e_{j_n}) = 0$, luego podemos eliminarlas de las sumas y así

$$f(v_1, \ldots, v_n) = \sum_{\sigma \in \Sigma_n} a_{1\sigma(1)} \cdots a_{n\sigma(n)} f(e_{\sigma(1)}, \ldots, e_{\sigma(n)}).$$

Como f es alternada tenemos que $f(e_{\sigma(1)}, \ldots, e_{\sigma(n)}) = (\operatorname{sig} \sigma) f(e_1, \ldots, e_n)$, luego

$$f(v_1, \ldots, v_n) = a \sum_{\sigma \in \Sigma_n} (\operatorname{sig} \sigma)\, a_{1\sigma(1)} \cdots a_{n\sigma(n)}. \tag{6.1}$$

Dado que esta expresión no depende de f, concluimos que si f existe es única. Además esto nos lleva a definir $f : (A^n)^n \longrightarrow A$ por la fórmula (6.1). Si probamos que la función así definida es una forma multilineal alternada y además $f(e_1, \ldots, e_n) = a$, el teorema quedará demostrado.

Tomemos $v_1, \ldots, v_n, v' \in A^n$, $b, b' \in A$ y $1 \leq i \leq n$. Sea $v_i = (a_{i1}, \ldots, a_{in})$, $v' = (a_1, \ldots, a_n)$. Claramente $bv_i + b'v' = (ba_{i1} + b'a_1, \ldots, ba_{in} + b'a_n)$, luego

$$f(v_1, \ldots, bv_i + b'v', \ldots, v_n) = a \sum_{\sigma \in \Sigma_n} (\operatorname{sig} \sigma) a_{1\sigma(1)} \cdots (ba_{i\sigma(i)} + b'a_{\sigma(i)}) \cdots a_{n\sigma(n)}$$

$$= ba \sum_{\sigma \in \Sigma_n} (\operatorname{sig} \sigma) a_{1\sigma(1)} \cdots a_{i\sigma(i)} \cdots a_{n\sigma(n)} + b'a \sum_{\sigma \in \Sigma_n} (\operatorname{sig} \sigma) a_{1\sigma(1)} \cdots a_{\sigma(i)} \cdots a_{n\sigma(n)}$$

$$= bf(v_1, \ldots, v_i, \ldots, v_n) + b'f(v_1, \ldots, v', \ldots, v_n).$$

Esto prueba que f es multilineal. Para probar que es alternada supongamos que $v_i = v_j$ con $i < j$. Entonces $a_{ik} = a_{jk}$ para $k = 1, \ldots, n$.

Sea A_n el grupo alternado, formado por las permutaciones de signatura positiva, y sea B_n el conjunto de las permutaciones impares.

Es inmediato que la aplicación $g : A_n \longrightarrow B_n$ dada por $g(\sigma) = (i, j)\sigma$ es biyectiva.

Si $\sigma \in A_n$ y $\tau = g(\sigma)$, entonces $a_{i\sigma(i)} = a_{i\tau(j)} = a_{j\tau(j)}$ e igualmente se cumple $a_{j\sigma(j)} = a_{i\tau(i)}$. De aquí resulta que $a_{1\sigma(1)} \cdots a_{n\sigma(n)} = a_{1\tau(1)} \cdots a_{n\tau(n)}$, y como $\operatorname{sig}\sigma = -\operatorname{sig} g(\sigma)$, el sumando correspondiente a σ se cancela con el correspondiente a $g(\sigma)$ y, en total, $f(v_1, \ldots, v_n) = 0$.

Por último, si $(v_1, \ldots, v_n) = (e_1, \ldots, e_n)$, entonces $a_{ij} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j, \end{cases}$ y el único sumando no nulo en (6.1) es el correspondiente a $\sigma = 1$, con lo que queda $f(e_1, \ldots, e_n) = a$. ∎

Ahora ya podemos definir la aplicación determinante. Conviene observar que si A es un dominio y n un número natural no nulo, podemos identificar $(A^n)^n$ con $\mathrm{Mat}_n(A)$. Concretamente, cada $(v_1, \ldots, v_n) \in (A^n)^n$ puede identificarse con la matriz que tiene por filas a $v_1, \ldots, v_n$. De hecho esta correspondencia es un isomorfismo de A-módulos.

Por ello es indistinto considerar que el dominio de una forma multilineal es $(A^n)^n$ o $\mathrm{Mat}_n(A)$. El teorema anterior puede reformularse como que existe una única forma multilineal alternada f sobre $\mathrm{Mat}_n(A)$ tal que $f(I_n)$ sea un valor dado $a \in A$.

Definición 6.4 Si A es un dominio y n es un número natural no nulo, llamaremos *función determinante* $\det : \mathrm{Mat}_n(A) \longrightarrow A$ a la única forma multilineal alternada que cumple $\det(I_n) = 1$.

Dada una matriz cuadrada B, escribiremos indistintamente $\det(B)$ o $|B|$ para representar al determinante de B.

Según la construcción del teorema anterior, si $B = (b_{ij})$, entonces

$$|B| = \sum_{\sigma \in \Sigma_n} (\operatorname{sig}\sigma)\, b_{1\sigma(1)} \cdots b_{n\sigma(n)}.$$

Por ejemplo, si $n = 1$ es claro que $|a| = a$ para todo $a \in A$.

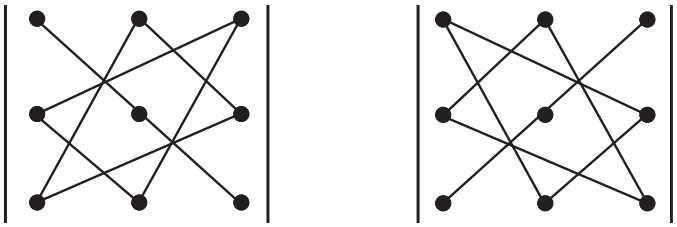
Para $n = 2$ tenemos la fórmula:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

Para $n = 3$ hay 6 sumandos, tres con signo positivo y tres con signo negativo. El lector puede comprobar que el desarrollo es el éste:

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - bdi - afh.$$

El esquema siguiente (conocido como *regla de Sarrus*) permite recordar fácilmente la fórmula:



Con esto vemos que el concepto de determinante que acabamos de introducir generaliza a los casos particulares que ya conocíamos. La fórmula de los determinantes de orden 4 contiene 24 sumandos, por lo que no resulta práctica. Más adelante veremos formas razonables de calcular determinantes de cualquier orden.

Teorema 6.5 *El determinante de una matriz cuadrada coincide con el de su traspuesta.*

DEMOSTRACIÓN: Sea $B = (b_{ij})$ una matriz $n \times n$ con coeficientes en un dominio A. Entonces

$$|B^t| = \sum_{\sigma \in \Sigma_n} (\operatorname{sig} \sigma)\, b_{\sigma(1)1} \cdots b_{\sigma(n)n}.$$

Reordenando los factores de cada sumando queda

$$|B^t| = \sum_{\sigma \in \Sigma_n} (\operatorname{sig} \sigma)\, b_{1\sigma^{-1}(1)} \cdots b_{n\sigma^{-1}(n)} = \sum_{\sigma \in \Sigma_n} (\operatorname{sig} \sigma^{-1})\, b_{1\sigma^{-1}(1)} \cdots b_{n\sigma^{-1}(n)},$$

y como la correspondencia $\sigma \mapsto \sigma^{-1}$ es biyectiva queda

$$|B^t| = \sum_{\sigma \in \Sigma_n} (\operatorname{sig} \sigma)\, b_{1\sigma(1)} \cdots b_{n\sigma(n)} = |B|. \qquad \blacksquare$$

El interés de este teorema reside en que, gracias a él, todas las propiedades que cumplen los determinantes respecto a las filas de una matriz se cumplen también respecto a las columnas.

Una de las propiedades más importantes de los determinantes es la siguiente:

Teorema 6.6 *Consideremos un dominio A, un número natural no nulo n y dos matrices B, $C \in \mathrm{Mat}_n(A)$. Entonces $|BC| = |B|\,|C|$.*

DEMOSTRACIÓN: Sea $f : \mathrm{Mat}_n(A) \longrightarrow A$ dada por $f(B) = |BC|$. Vamos a probar que f es una forma multilineal alternada.

Por comodidad usaremos la notación $f(B_1, \ldots, B_n)$ para indicar la imagen de la matriz B que tiene filas $B_1, \ldots, B_n$. Notemos que la fila i-ésima de BC es $(B_iC^1, \ldots, B_iC^n)$, donde $C^1, \ldots, C^n$ son las columnas de la matriz C.

Así, $f(B_1, \ldots, B_n) = \det(Z_1, \ldots, Z_n)$, donde $Z_i = (B_iC^1, \ldots, B_iC^n)$. A partir de aquí se sigue inmediatamente la multilinealidad de f. Además, si $B_i = B_j$, entonces $Z_i = Z_j$, luego $f(B_1, \ldots, B_n) = 0$.

Con la notación del teorema 6.3, tenemos además que

$$f(e_1, \dots, e_n) = f(I_n) = |I_n C| = |C|,$$

luego f ha de ser la aplicación construida en la prueba de dicho teorema cuando $a = |C|$ (fórmula (6.1)), que en términos de matrices y determinantes es simplemente $f(B) = |B|a$. Así pues: $|BC| = f(B) = |B||C|$. ∎

Ahora vamos a dar algunas propiedades elementales que permiten manipular determinantes.

Teorema 6.7 *Sea A un dominio y B, $C \in \mathrm{Mat}_n(A)$. Entonces*

1. *Si C resulta de intercambiar dos filas o columnas de la matriz B, entonces $|C| = -|B|$.*
2. *Si C resulta de multiplicar una fila o columna de B por un cierto $a \in A$, entonces $|C| = a|B|$.*
3. *Si C resulta de sumar a la fila (o columna) i-ésima de B la fila (o columna) j-ésima de B con $i \neq j$, multiplicada por un $a \in A$, entonces $|C| = |B|$.*

Demostración: 1) y 2) son consecuencias inmediatas de la definición de determinante (las variantes con columnas se cumplen por el teorema 6.5).

3) Se cumple porque $|C|$ se descompone por multilinealidad en dos sumandos, uno es $|B|$ y otro el determinante de la matriz que resulta de repetir en el lugar i-ésimo la columna j-ésima (multiplicado por a), y éste es nulo. ∎

Estos resultados nos permiten calcular determinantes de cualquier orden mediante manipulaciones adecuadas. Basta notar que si una matriz cuadrada B tiene nulos todos los coeficientes bajo la diagonal principal, es decir, si $b_{ij} = 0$ cuando $i > j$, entonces $|B|$ es el producto de los coeficientes de la diagonal principal (pues la única permutación que no da lugar a un sumando nulo en la definición de determinante es la identidad).

Por otro lado conviene observar que si A es un dominio íntegro y K es su cuerpo de cocientes, una matriz en $\mathrm{Mat}_n(A)$ está también en $\mathrm{Mat}_n(K)$ y su determinante es el mismo en cualquier caso. Por ello a la hora de calcular determinantes podemos trabajar siempre en los cuerpos de cocientes, es decir, podemos hacer divisiones cuando convenga.

Calculemos por ejemplo:

$$\begin{vmatrix} 2 & -3 & 2 & 3 \\ 4 & 3 & 5 & 0 \\ 3 & 2 & 0 & -3 \\ 5 & 2 & 4 & 7 \end{vmatrix} = \begin{vmatrix} 2 & -3 & 2 & 3 \\ 0 & 9 & 1 & -6 \\ 0 & \frac{13}{2} & -3 & -\frac{15}{2} \\ 0 & \frac{19}{2} & -1 & -\frac{1}{2} \end{vmatrix} =$$

El segundo determinante resulta de sumar a la segunda fila la primera multiplicada por -2, a la tercera fila la primera multiplicada por $-3/2$ y a la cuarta la primera multiplicada por $-5/2$. De este modo conseguimos ceros bajo el término a_{11}.

Por el mismo proceso hacemos ceros en todas las posiciones bajo la diagonal principal: sumamos a la tercera fila la segunda multiplicada por $-13/18$ y a la cuarta la segunda multiplicada por $-19/18$. Después sumamos a la cuarta fila la tercera multiplicada por $-37/67$ y obtenemos una matriz triangular, es decir, con ceros bajo la diagonal:

$$= \begin{vmatrix} 2 & -3 & 2 & 3 \\ 0 & 9 & 1 & -6 \\ 0 & 0 & -\frac{67}{18} & -\frac{19}{6} \\ 0 & 0 & -\frac{37}{18} & \frac{36}{6} \end{vmatrix} = \begin{vmatrix} 2 & -3 & 2 & 3 \\ 0 & 9 & 1 & -6 \\ 0 & 0 & -\frac{67}{18} & -\frac{19}{6} \\ 0 & 0 & 0 & \frac{508}{67} \end{vmatrix} =$$

Ahora el determinante se reduce al producto de los elementos de la diagonal, o sea:

$$= 2 \cdot 9 \cdot (-67/18) \cdot (508/67) = -508.$$

De este modo se puede calcular cualquier determinante, pero vamos a probar que el trabajo puede reducirse considerablemente.

Definición 6.8 Sea A un dominio y $B \in \mathrm{Mat}_n(A)$. Llamaremos *menor complementario* de b_{ij} al determinante de la matriz que resulta de eliminar la fija i-ésima y la columna j-ésima de B. Lo representaremos por B_{ij}.

Teorema 6.9 *Sea A un dominio y sea $B \in \mathrm{Mat}_n(A)$ tal que en su fila i-ésima el único elemento no nulo sea b_{ij}. Entonces $|B| = (-1)^{i+j} b_{ij} B_{ij}$.*

Demostración: Supongamos en primer lugar que $i = j = n$, o sea, $b_{nj} = 0$ si $j = 1, \ldots, n-1$. Así

$$\begin{aligned} |B| &= \sum_{\sigma \in \Sigma_n} (\operatorname{sig}\sigma) b_{1\sigma(1)} \cdots b_{n\sigma(n)} = \sum_{\substack{\sigma \in \Sigma_n \\ \sigma(n)=n}} (\operatorname{sig}\sigma) b_{1\sigma(1)} \cdots b_{(n-1)\sigma(n-1)} b_{nn} \\ &= b_{nn} \sum_{\sigma \in \Sigma_{n-1}} (\operatorname{sig}\sigma) b_{1\sigma(1)} \cdots b_{(n-1)\sigma(n-1)} = b_{nn} B_{nn} = (-1)^{n+n} b_{nn} B_{nn} \end{aligned}$$

Si i y j son cualesquiera, sea B' la matriz que resulta de llevar la fila i-ésima de B a la posición n-sima. Para hacer este cambio hay que permutar la fila i-ésima con las $n-i$ filas que le siguen, luego el signo del determinante cambia $n-i$ veces: $|B| = (-1)^{n-i}|B'|$.

Sea ahora B'' la matriz que resulta de llevar la columna j-ésima de B' a la posición n-sima. De nuevo $|B'| = (-1)^{n-j}|B''|$ y así $|B| = (-1)^{2n-i-j}|B''| = (-1)^{i+j}|B''|$.

La fila n-sima de B'' tiene únicamente la componente n-sima no nula, y además es igual a b_{ij}.

Por lo ya probado $|B| = (-1)^{i+j} b_{ij} B''_{nn}$, pero es obvio que $B''_{nn} = B_{ij}$, luego $|B| = (-1)^{i+j} b_{ij} B_{ij}$. ∎

Teniendo esto en cuenta, en nuestro ejemplo era suficiente con hacer ceros en la primera columna:

$$\begin{vmatrix} 2 & -3 & 2 & 3 \\ 4 & 3 & 5 & 0 \\ 3 & 2 & 0 & -3 \\ 5 & 2 & 4 & 7 \end{vmatrix} = \begin{vmatrix} 2 & -3 & 2 & 3 \\ 0 & 9 & 1 & -6 \\ 0 & \frac{13}{2} & -3 & -\frac{15}{2} \\ 0 & \frac{19}{2} & -1 & -\frac{1}{2} \end{vmatrix} = 2 \begin{vmatrix} 9 & 1 & -6 \\ \frac{13}{2} & -3 & -\frac{15}{2} \\ \frac{19}{2} & -1 & -\frac{1}{2} \end{vmatrix},$$

y el determinante que resulta se puede calcular fácilmente. Por supuesto el teorema anterior vale para columnas igual que para filas.

Ejemplo Como aplicación vamos a calcular los llamados *determinantes de Vandermonde*. Concretamente probaremos que

$$\begin{vmatrix} 1 & \cdots & 1 \\ a_1 & \cdots & a_n \\ \vdots & & \vdots \\ a_1^{n-1} & \cdots & a_n^{n-1} \end{vmatrix} = \prod_{i<j} (a_j - a_i),$$

donde $a_1, \ldots, a_n$ son elementos de un dominio A. En particular, si A es un dominio íntegro, un determinante de Vandermonde es no nulo si y sólo si los elementos de su segunda fila son distintos dos a dos.

Lo probaremos por inducción sobre n. Para $n = 1$ o incluso $n = 2$ es inmediato. Supuesto para $n - 1$ restamos a cada fila la anterior multiplicada por a_1, con lo que obtenemos

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ 0 & a_2 - a_1 & \cdots & a_n - a_1 \\ \vdots & \vdots & & \vdots \\ 0 & a_2^{n-1} - a_1 a_2^{n-2} & \cdots & a_n^{n-1} - a_1 a_n^{n-2} \end{vmatrix}.$$

Ahora aplicamos el teorema anterior y obtenemos

$$\begin{vmatrix} a_2 - a_1 & \cdots & a_n - a_1 \\ \vdots & & \vdots \\ a_2^{n-1} - a_1 a_2^{n-2} & \cdots & a_n^{n-1} - a_1 a_n^{n-2} \end{vmatrix}.$$

Por la multilinealidad sobre las columnas este determinante es igual a

$$(a_2 - a_1) \cdots (a_n - a_1) \begin{vmatrix} 1 & \cdots & 1 \\ a_2 & \cdots & a_n \\ \vdots & & \vdots \\ a_2^{n-2} & \cdots & a_n^{n-2} \end{vmatrix}.$$

Finalmente por la hipótesis de inducción esto es igual a $\prod_{i<j} (a_j - a_i)$. ∎

En realidad el teorema 6.9 es un caso particular de un resultado más general:

Teorema 6.10 *Sea A un dominio y $B \in \mathrm{Mat}_n(A)$. Entonces*

$$|B| = \sum_{j=1}^{n} (-1)^{i+j} b_{ij} B_{ij} = \sum_{i=1}^{n} (-1)^{i+j} b_{ij} B_{ij}.$$

DEMOSTRACIÓN: Sean $B_1, \ldots, B_n$ las filas de B. Así $B_i = \sum_{j=1}^{n} b_{ij} e_j$, donde $e_j = (\delta_{ij})$. Claramente,

$$\begin{aligned} \det(B) &= \det(B_1, \ldots, \sum_{j=1}^{n} b_{ij} e_j, \ldots, B_n) = \sum_{j=1}^{n} b_{ij} \det(B_1, \ldots, e_j, \ldots, B_n) \\ &= \sum_{j=1}^{n} b_{ij} (-1)^{i+j} B_{ij}. \end{aligned}$$

La otra igualdad se prueba análogamente. ∎

Pasemos ahora a mostrar el interés teórico de los determinantes. En primer lugar veremos que los determinantes determinan cuándo una matriz es regular.

Definición 6.11 Sea A un dominio y $B \in \mathrm{Mat}_n(A)$. Llamaremos *matriz adjunta* de B a la matriz $\tilde{B} \in \mathrm{Mat}_n(A)$ dada por

$$\tilde{b}_{ij} = (-1)^{i+j} B_{ji}.$$

Notemos que en la posición (i, j) está el menor complementario de b_{ji}, es decir, $\tilde{B}$ se forma sustituyendo en B cada elemento por su menor complementario multiplicado por el signo adecuado y después trasponiendo la matriz resultante.

Por ejemplo, si

$$B = \begin{pmatrix} 1 & 3 & -2 \\ 5 & 1 & 0 \\ -3 & 4 & 2 \end{pmatrix},$$

entonces

$$B_{11} = \begin{vmatrix} 1 & 0 \\ 4 & 2 \end{vmatrix} = 2, \quad B_{12} = \begin{vmatrix} 5 & 0 \\ -3 & 2 \end{vmatrix} = 10, \quad B_{13} = \begin{vmatrix} 5 & 1 \\ -3 & 4 \end{vmatrix} = 23,$$

$$B_{21} = \begin{vmatrix} 3 & -2 \\ 4 & 2 \end{vmatrix} = 14, \quad B_{22} = \begin{vmatrix} 1 & -2 \\ -3 & 2 \end{vmatrix} = -4, \quad B_{23} = \begin{vmatrix} 1 & 3 \\ -3 & 4 \end{vmatrix} = 13,$$

$$B_{31} = \begin{vmatrix} 3 & -2 \\ 1 & 0 \end{vmatrix} = 2, \quad B_{32} = \begin{vmatrix} 1 & -2 \\ 5 & 0 \end{vmatrix} = 10, \quad B_{33} = \begin{vmatrix} 1 & 3 \\ 5 & 1 \end{vmatrix} = 14.$$

Al reemplazar cada elemento por su menor con el signo adecuado queda

$$\begin{pmatrix} 2 & -10 & 23 \\ -14 & -4 & -13 \\ 2 & -10 & -14 \end{pmatrix}$$

luego la matriz adjunta de B es

$$\tilde{B} = \begin{pmatrix} 2 & -14 & 2 \\ -10 & -4 & -10 \\ 23 & -13 & -14 \end{pmatrix}.$$

Teorema 6.12 *Sea A un dominio y $B \in \mathrm{Mat}_n(A)$. Entonces*

$$B\tilde{B} = \tilde{B}B = |B|I_n.$$

DEMOSTRACIÓN: El término (i,j) de $B\tilde{B}$ es igual a $\sum_{k=1}^{n} b_{ik}(-1)^{k+j}B_{jk}$.

Si $i = j$ queda $\sum_{k=1}^{n} b_{ik}(-1)^{k+j}B_{ik} = |B|$ por el teorema 6.10, luego los elementos de la diagonal principal de $B\tilde{B}$ son todos iguales a $|B|$.

Si $i \neq j$ llamemos D la matriz cuyas filas son las de B salvo que en la posición j-ésima tiene repetida la fila i-ésima. Entonces $|D| = 0$ y desarrollando por la fila j-ésima queda

$$0 = |D| = \sum_{k=1}^{n} d_{jk}(-1)^{k+j}D_{jk} = \sum_{k=1}^{n} b_{ik}(-1)^{k+j}B_{jk},$$

o sea, los elementos fuera de la diagonal principal de $B\tilde{B}$ son nulos. Por lo tanto, $B\tilde{B} = |B|I_n$. Del mismo modo se prueba la otra igualdad. ∎

Esto significa que la matriz adjunta de una matriz B es casi su matriz inversa. Para obtener la inversa sólo falta que sea lícito dividir entre el determinante de B. La situación es la siguiente:

Teorema 6.13 *Sea A un dominio y $B \in \mathrm{Mat}_n(A)$. Entonces la matriz B es regular si y sólo si $|B|$ es una unidad de A, y en tal caso*

$$B^{-1} = \frac{1}{|B|}\tilde{B}.$$

DEMOSTRACIÓN: Si la matriz B es regular, entonces existe B^{-1} de manera que $BB^{-1} = I_n$, luego tomando determinantes $|B|\,|B^{-1}| = |I_n| = 1$, lo que prueba que $|B|$ es una unidad de A.

Si $|B|$ es una unidad de A, entonces sea $C = \frac{1}{|B|}\tilde{B} \in \mathrm{Mat}_n(A)$. Por el teorema anterior, $BC = CB = I_n$, luego B es regular y $B^{-1} = C$. ∎

En particular una matriz con coeficientes en un cuerpo es regular si y sólo si su determinante es distinto de cero.

Ejemplo Si $B, C \in \mathrm{Mat}_n(A)$ cumplen $BC = I_n$, podemos concluir que B y C son regulares y que $C = B^{-1}$. En efecto, basta tomar determinantes para concluir que $|B||C| = 1$, luego B y C son regulares y, multiplicando por B^{-1} en $BC = I_n$ resulta $C = B^{-1}$. ∎

Aplicación: La regla de Cramer Los resultados anteriores nos dan una expresión sencilla en términos de determinantes para las soluciones de un sistema de ecuaciones lineales:

$$\left.\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{n1}x_n + \cdots + a_{nn}x_n &= b_n \end{aligned}\right\}$$

Claramente podemos escribirlo matricialmente como

$$Ax^t = b^t,$$

donde $A = (a_{ij})$ es la matriz de los coeficientes, $b = (b_i)$ es el vector de términos independientes y $x = (x_1, \ldots, x_n)$. Si $|A| \neq 0$, el sistema tiene una única solución, dada por

$$x^t = A^{-1}b^t = \frac{1}{|A|}\tilde{A}b^t.$$

En particular,

$$x_j = \frac{1}{|A|}\sum_{i=1}^{n}(-1)^{i+j}b_iA_{ij}.$$

Ahora bien, si llamamos $C = (c_{uv})$ a la matriz que resulta de sustituir en A su columna j-ésima por el vector b, tenemos que $C_{ij} = A_{ij}$ para todo i, así como que $c_{ij} = b_i$, luego, aplicando el teorema 6.10, llegamos a que

$$x_j = \frac{1}{|A|}\sum_{i=1}^{n}(-1)^{i+j}c_{ij}C_{ij} = \frac{|C|}{|A|}.$$

En resumen:

> **Regla de Cramer:** *La j-ésima coordenada de la solución de un sistema de n ecuaciones lineales con n incógnitas (cuya matriz de coeficientes A tenga determinante no nulo) puede calcularse dividiendo entre $|A|$ el determinante de la matriz que resulta de sustituir la columna j-ésima de A por el vector de términos independientes.*

Ejemplo La solución del sistema de ecuaciones

$$\left.\begin{aligned} x - 2y + z &= 3\\ 2x + 2y - z &= 1\\ x + y + z &= 2 \end{aligned}\right\}$$

Puede calcularse con la regla de Cramer, pues

$$\begin{vmatrix} 1 & -2 & 1\\ 2 & 2 & -1\\ 1 & 1 & 1 \end{vmatrix} = 9 \neq 0,$$

y viene dada por

$$x = \frac{1}{9}\begin{vmatrix} 3 & -2 & 1\\ 1 & 2 & -1\\ 2 & 1 & 1 \end{vmatrix} \quad y = \frac{1}{9}\begin{vmatrix} 1 & 3 & 1\\ 2 & 1 & -1\\ 1 & 2 & 1 \end{vmatrix}, \quad z = \frac{1}{9}\begin{vmatrix} 1 & -2 & 3\\ 2 & 2 & 1\\ 1 & 1 & 2 \end{vmatrix},$$

lo que nos da $(x, y, z) = (4/3, -1/3, 1)$. ∎

Los determinantes también nos permiten decidir si n elementos de un módulo libre de rango n son o no linealmente independientes, y si son o no una base.

Teorema 6.14 *Sea A un dominio íntegro, sea M un A-módulo libre de rango n, sea $(v_1, \ldots, v_n)$ una base ordenada de M, sean $w_1, \ldots, w_n$ elementos de M y sea $B = (b_{ij})$ la matriz cuyas filas son las coordenadas de $w_1, \ldots, w_n$ en la base dada. Entonces:*

1. *$(w_1, \ldots, w_n)$ es una base de M si y sólo si $|B|$ es una unidad de A.*

2. *$(w_1, \ldots, w_n)$ son linealmente independientes si y sólo si $|B| \neq 0$.*

Demostración: 1) Por 4.21 existe un homomorfismo $f : M \longrightarrow M$ tal que $f(v_i) = w_i$ para cada $i = 1, \ldots, n$. La matriz de f en la base $(v_1, \ldots, v_n)$ es precisamente B.

Si $w_1, \ldots, w_n$ forman una base de M entonces también existe un homomorfismo $g : M \longrightarrow M$ tal que $f(w_i) = v_i$ para cada $i = 1, \ldots, n$. La composición $f \circ g$ es la identidad sobre la base $(v_1, \ldots, v_n)$, luego por la unicidad del teorema 4.21 se cumple que $f \circ g$ es la aplicación identidad en M. Igualmente $g \circ f$ es la identidad en M. Esto prueba que f es un isomorfismo y por lo tanto $|B|$ es una unidad.

Si $|B|$ es una unidad, la aplicación f es un isomorfismo, luego $(w_1, \ldots, w_n)$ es una base de M (pues son la imagen de una base por un isomorfismo).

2) Como la aplicación que asocia a cada elemento de M sus coordenadas en la base dada es un isomorfismo entre M y A^n, los elementos $w_1, \ldots, w_n$ son linealmente dependientes si y sólo si lo son sus coordenadas, es decir, las filas de B.

Las filas de B son linealmente independientes en A^n si y sólo si son linealmente independientes en K^n, donde K es el cuerpo de fracciones de A. En efecto, si tenemos una combinación lineal de las filas de B con coeficientes en K no todos nulos y que es igual a 0, multiplicando por el producto de los denominadores de los coeficientes no nulos, obtenemos una nueva combinación lineal que también anula a las filas de A, ahora con los coeficientes en A y no todos nulos. La otra implicación es obvia.

Como K^n es un espacio vectorial de dimensión n, las filas de B son linealmente independientes en K^n si y sólo si son una base de K^n.

Por 1), las filas de B son una base de K^n si y sólo si $|B|$ es una unidad en K, o sea, si y sólo si $|B| \neq 0$. ∎

Concluimos la sección con otra aplicación de los determinantes, esta vez al cálculo del cardinal de los módulos cociente de los $\mathbb{Z}$-módulos libres.

Teorema 6.15 *Sea M un $\mathbb{Z}$-módulo libre de rango m y sea N un submódulo de rango n. Entonces:*

1. *El módulo cociente M/N es finito si y sólo si $n = m$.*

2. *Si $n = m$, $(v_1, \ldots, v_n)$ es una base de M, $(w_1, \ldots, w_n)$ es una base de N y B es la matriz cuyas filas son las coordenadas de $w_1, \ldots, w_n$ en la base $(v_1, \ldots, v_n)$, entonces el cardinal de M/N es $|\det B|$.*

DEMOSTRACIÓN: 1) Sea $(z_1, \ldots, z_m)$ una base de M tal que $(a_1 z_1, \ldots, a_n z_n)$ sea una base de N para ciertos elementos $a_1, \ldots, a_n \in \mathbb{Z}$ (de acuerdo con el teorema 4.53). En la prueba de dicho teorema se ve que

$$M/N \cong (\mathbb{Z}/a_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/a_n\mathbb{Z}) \times \mathbb{Z}^{m-n}.$$

El cociente será finito si y sólo si $m - n = 0$, o sea, si y sólo si $m = n$.

2) Si $m = n$, en las condiciones de 1) tenemos

$$|M/N| = |(\mathbb{Z}/a_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/a_n\mathbb{Z})| = |a_1 \cdots a_n| = |\det C|,$$

donde C es la matriz que tiene a $a_1, \ldots, a_n$ en la diagonal y los restantes coeficientes nulos.

Sea $f : N \longrightarrow M$ la aplicación dada por $f(x) = x$ para todo $x \in N$. La matriz de f en las bases $(a_1 z_1, \ldots, a_n z_n)$ y $(z_1, \ldots, z_n)$ es precisamente C.

Sea P la matriz de la aplicación identidad en N respecto de las bases $(w_1, \ldots, w_n)$ y $(a_1 z_1, \ldots, a_n z_n)$ (la matriz de cambio de base). Por el teorema 6.13 tenemos que $|P| = \pm 1$.

Sea Q la matriz de la aplicación identidad en M respecto de las bases $(z_1, \ldots, z_n)$ y $(v_1, \ldots, v_n)$. Por la misma razón $|Q| = \pm 1$.

Por el teorema 4.35 la matriz PCQ es la matriz de f respecto de las bases $(w_1, \ldots, w_n)$ y $(v_1, \ldots, v_n)$. Por lo tanto las filas de PCQ son las coordenadas de $w_1, \ldots, w_n$ en la base $(v_1, \ldots, v_n)$, es decir, $B = PCQ$.

Tomando determinantes y valores absolutos,

$$|\det B| = |\det P|\,|\det C|\,|\det Q| = |\det C| = |M/N|.$$

■

6.2 Clasificación de homomorfismos de módulos

Sabemos que una aplicación lineal entre dos espacios vectoriales está completamente determinada por su matriz respecto de dos bases dadas, pero dicha matriz puede ser más sencilla o más complicada en función de las bases elegidas. Ni siquiera es evidente cómo saber si dos matrices dadas pueden ser la matriz de la misma aplicación lineal respecto a dos pares de bases. Por razones que veremos más adelante, conviene plantear estos problemas en el contexto más general de los homomorfismos entre módulos libres sobre anillos.

Supongamos que $f : M \longrightarrow N$ es un homomorfismo entre A-módulos libres de rangos m y n respectivamente y que B y B' son bases respectivas. Sea S la matriz de f en estas bases. Supongamos que f tiene otra matriz T respecto a otras bases C y C'. ¿Cuál es entonces la relación entre S y T? Es sencilla: Sea P la matriz de cambio de base de C a B (es decir, la matriz de la identidad en M respecto de las bases C y B) y sea Q la matriz de cambio de base de C' a B'. Entonces si x es la m-tupla de coordenadas en C de un elemento $m \in M$, tendremos que xP es la m-tupla de coordenadas de m en la base B,

luego xPS es la n-tupla de coordenadas de $f(m)$ en B', luego $xPSQ$ es la n-tupla de coordenadas de $f(m)$ en C'. Por la unicidad de la matriz, se ha de cumplir $T = PSQ$. Además como P y Q son matrices de isomorfismos, ambas son regulares. En vista de esto definimos:

Definición 6.16 Sea A un anillo conmutativo y unitario. Diremos que dos matrices $S, T \in \mathrm{Mat}_{m\times n}(A)$ son *equivalentes* si existen matrices regulares P, Q (de orden $m \times m$ y $n \times n$, respectivamente) tales que $T = PSQ$.

Es evidente que la equivalencia de matrices es una relación de equivalencia en el conjunto $\mathrm{Mat}_{m\times n}(A)$. Acabamos de demostrar que si S y T son matrices de un mismo homomorfismo entre A-módulos libres $f : M \longrightarrow N$, entonces S y T son equivalentes. Recíprocamente, si S es la matriz de f en ciertas bases y T es equivalente a S, es fácil ver que T es la matriz de f en otras bases adecuadas (las que convierten a P y Q en las matrices del cambio de base).

Ahora vamos a dar un criterio sencillo para determinar cuándo dos matrices dadas son o no equivalentes, al tiempo que encontraremos para cada matriz una equivalente lo más sencilla posible. Para ello usaremos el resultado siguiente:

Teorema 6.17 *Sea A un anillo conmutativo y unitario y sean S y T dos matrices $m \times n$ en A. Si T resulta de realizar sobre S una de las operaciones siguientes, entonces T es equivalente a S:*

1. *Permutar dos filas (o columnas).*
2. *Cambiar la fila (o columna) i-ésima por la suma de la fila (columna) i-ésima más la fila (columna) j-ésima ($j \neq i$) multiplicada por un $a \in A$.*
3. *Multiplicar una fila (columna) por una unidad $u \in A$.*

Demostración: Realizar una de estas operaciones equivale a multiplicar la matriz por la izquierda o por la derecha (según sea sobre filas o columnas) por una de las matrices regulares siguientes (todas tienen unos en la diagonal y ceros fuera de ella salvo cuando se indica explícitamente lo contrario):

$$E_1^{ij} = \begin{array}{c} \\ i \\ \\ j \\ \\ \end{array}\!\!\begin{array}{c} \begin{array}{ccccccc} & & i & & j & & \end{array} \\ \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \end{array}, \quad E_2^{ij}(a) = \begin{array}{c} \\ \\ \\ i \\ \\ \\ \end{array}\!\!\begin{array}{c} \begin{array}{ccccccc} & & & & & j & \end{array} \\ \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & \ddots & & a & & \\ & & & \ddots & & & \\ & & & & \ddots & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \end{array},$$

$$E_3^i(u) = \begin{array}{c} \\ \\ i \\ \\ \\ \end{array}\!\!\begin{array}{c} \begin{array}{ccccc} & & i & & \end{array} \\ \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & u & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} \end{array}.$$

■

Las matrices consideradas en la prueba del teorema anterior se llaman *matrices elementales.* Observemos que

$$(E_1^{ij})^{-1} = E_1^{ij}, \qquad (E_2^{ij}(a))^{-1} = E_2^{ij}(-a), \qquad (E_3^i(u))^{-1} = E_3^i(u^{-1}).$$

Así, la inversa de una matriz elemental es otra matriz elemental del mismo tipo.

El teorema siguiente nos resuelve el problema de determinar cuándo dos matrices sobre un dominio euclídeo son equivalentes:

Teorema 6.18 *Sea E un dominio euclídeo y $A \in \mathrm{Mat}_{m\times n}(E)$. Entonces A es equivalente a una única matriz de la forma*

$$\begin{pmatrix} d_1 & & & & & \\ & \ddots & & & & \\ & & d_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix},$$

donde cada $d_i \neq 0$ y $d_i \mid d_{i+1}$. La unicidad se entiende salvo sustitución de los d_i por asociados (o sea, salvo unidades). Los elementos d_i se llaman factores invariantes *de A.*

Demostración: La prueba que vamos a ver nos da un algoritmo para calcular los factores invariantes de cualquier matriz dada. Llamemos ϕ a la norma euclídea del anillo E.

Si $A = 0$ ya es de la forma requerida. En otro caso, sea a_{ij} un coeficiente de A no nulo con norma mínima. Intercambiando filas y columnas podemos llevarlo a la posición $(1,1)$, es decir, pasamos a una matriz equivalente donde $a_{11} \neq 0$ tiene norma mínima.

Para cada $k > 1$ dividimos $a_{1k} = a_{11}b_k + b_{1k}$, de modo que $b_{1k} = 0$ o bien $\phi(b_{1k}) < \phi(a_{11})$.

Restamos a la columna k-ésima la primera multiplicada por b_k, con lo que la primera fila se convierte en $(a_{11}, b_{12}, \ldots, b_{1n})$.

Si algún b_{1k} es no nulo llevamos a la posición $(1,1)$ el de menor norma y repetimos el proceso. Como cada vez la norma del coeficiente $(1,1)$ disminuye, el proceso no puede continuar indefinidamente, por lo que al cabo de un número finito de pasos llegaremos a una primera fila de la forma $(a_{11}, 0, \ldots, 0)$.

Repitiendo el proceso con la primera columna llegamos a una matriz equivalente de la forma:

$$\left(\begin{array}{c|ccc} a_{11} & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right)$$

Si a_{11} no divide a alguno de los restantes coeficientes a_{ij}, entonces hacemos $a_{ij} = a_{11}c + d$ con $d \neq 0$ y $\phi(d) < \phi(a_{11})$, llevamos d a la posición $(1,1)$ y

repetimos el proceso de hacer ceros. Como la norma sigue disminuyendo, tras un número finito de pasos llegaremos a una matriz como la anterior y en la que a_{11} divide a todos los coeficientes restantes.

Ahora repetimos el proceso con la matriz B, lo cual no altera los ceros de la fila y la columna primera ni el hecho de que a_{11} divide a todos los coeficientes. De este modo llegamos a una matriz como la del enunciado.

Ahora demostramos que si dos matrices como la del enunciado son equivalentes, entonces son iguales (salvo multiplicación de sus coeficientes por unidades). Para ello fijamos una cualquiera M y consideramos el homomorfismo $f : E^m \longrightarrow E^n$ que en las bases canónicas $\{e_1, \ldots, e_m\}$ y $\{f_1, \ldots, f_n\}$ tiene matriz M, es decir, $f(e_i) = d_i f_i$ para $i = 1, \ldots, r$ y $f(e_i) = 0$ en otro caso. Entonces $d_1 f_1, \ldots, d_r f_r$ forman una base de $\operatorname{Im} f$ y el teorema 4.53 nos da que r y $f_1, \ldots, f_r$ están unívocamente determinados (salvo unidades) por E^n e $\operatorname{Im} f$ (y no dependen de la elección de las bases). Por lo tanto, si otra matriz en las condiciones del enunciado es equivalente a M, sería la matriz de f en otras bases, luego tendría que ser la misma M (salvo unidades). ∎

Así pues, dos matrices son equivalentes si y sólo si tienen los mismos factores invariantes. La matriz equivalente a una matriz dada A que tiene la forma indicada en el teorema anterior se llama *forma canónica* de A, de modo que dos matrices son equivalentes si y sólo si tienen la misma forma canónica (salvo unidades).

Observemos que en la demostración del teorema anterior hemos llegado a la forma canónica realizando únicamente operaciones elementales sobre la matriz dada (es decir, multiplicaciones por matrices elementales). Por lo tanto, hemos probado lo siguiente:

Teorema 6.19 *Dos matrices $S, T \in \operatorname{Mat}_{m\times n}(E)$ sobre un dominio euclídeo E son equivalentes si y sólo si existen matrices elementales tales que*

$$S = P_1 \cdots P_r T Q_1 \cdots Q_s.$$

A partir de aquí nos centramos en el caso de las matrices sobre un cuerpo (recordemos que todo cuerpo es trivialmente un dominio euclídeo). Como todos los elementos no nulos son unidades, todos los factores invariantes de una matriz A pueden tomarse iguales a 1, luego lo único que puede variar es su número r. Este número se llama *rango* de A y lo representaremos por $\operatorname{rang} A$. Tenemos, pues, que dos matrices sobre un cuerpo son equivalentes si y sólo si tienen el mismo rango.

El rango de A tiene una interpretación muy sencilla: sea f una aplicación lineal de matriz A. Las filas de A son las coordenadas de las imágenes de los vectores de la primera base respecto a la segunda base. Estas imágenes generan el subespacio $\operatorname{Im} f$, luego contienen exactamente $\dim \operatorname{Im} f$ vectores independientes. Como la aplicación que a cada vector le asigna sus coordenadas es un isomorfismo, resulta que A tiene $\dim \operatorname{Im} f$ filas independientes. La forma canónica de A es también una matriz de f y tiene r filas independientes, luego $\dim \operatorname{Im} f = r$ y el rango de una matriz es el número de filas independientes que contiene.

Por otra parte es obvio que si dos matrices son equivalentes sus traspuestas también lo son, y la traspuesta de una forma canónica es ella misma, luego el rango de una matriz es el mismo que el de su traspuesta. Por lo tanto:

> *El rango de una matriz A con coeficientes en un cuerpo es el número de filas y el número de columnas linealmente independientes.*

Si $f : V \longrightarrow W$ es una aplicación lineal entre k-espacios vectoriales de dimensión finita, podemos definir su rango $\operatorname{rang} f$ como el rango de su matriz respecto de cualquier par de bases. Como las matrices equivalentes tienen el mismo rango, no importa la elección de las bases. De hecho, el rango de una aplicación lineal tiene una interpretación directa que no depende de ninguna elección de bases, pues es simplemente $\operatorname{rang} f = \dim_k \operatorname{Im} f$.

En efecto, basta tener en cuenta que, fijadas unas bases $\{v_1, \ldots, v_m\}$ y $\{w_1, \ldots, w_n\}$, entonces $\operatorname{Im} v = \langle f(v_1), \ldots, f(v_m)\rangle$ y el isomorfismo $W \cong k^n$ que a cada vector le asigna sus coordenadas en la base fijada en W transforma $\operatorname{Im} f$ en el subespacio vectorial generado por las filas de la matriz de f en las bases fijadas, cuya dimensión es el rango de la matriz.

Nota La prueba del teorema 6.18 se puede simplificar sustancialmente para el caso de matrices regulares sobre un cuerpo k, y la posibilidad de tal simplificación tiene consecuencias teóricas de interés. En principio sabemos que, mediante productos a izquierda y derecha por matrices elementales, toda matriz regular A se puede transformar en la matriz identidad. Ahora veremos que, multiplicando sólo por la derecha (o sólo por la izquierda) por matrices elementales sólo de tipo $E_2^{ij}(a)$, es posible transformar A en una matriz de la forma

$$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \alpha \end{pmatrix},$$

para cierto $\alpha \in k$ que, de hecho, es necesariamente $\alpha = |A|$, pues las matrices de tipo $E_2^{ij}(a)$ tienen todas determinante 1, luego al multiplicar por ellas no cambia el determinante de la matriz.

En efecto, sabemos que multiplicar una matriz (por ejemplo, por la derecha) por matrices de tipo $E_2^{ij}(a)$ equivale a sumar a una columna otra multiplicada por a. Veamos que sólo con operaciones elementales de este tipo es posible llegar a una matriz de la forma indicada.

Como A tiene determinante no nulo, no puede tener ninguna fila nula. Si $a_{12} = 0$, sumamos a la segunda columna otra columna adecuada para hacer $a_{12} \neq 0$. Seguidamente le sumamos a la primera columna la segunda multiplicada por $(1 - a_{11})/a_{12}$, y así pasamos a una matriz con $a_{11} = 1$.

Luego sumamos a la columna $j > 1$ la columna 1 multiplicada por $-a_{1j}$, con lo que la primera fila pasa a ser $(1, 0, \ldots, 0)$.

Ahora, no puede ocurrir que todos los a_{2i} sean nulos, para $i > 1$, porque entonces sería $|A| = 0$. Esto significa que podemos hacer que $a_{23} \neq 0$ sumándole si es preciso una columna distinta de la primera, lo cual no altera la primera fila. A su vez, con la tercera columna podemos hacer $a_{22} = 1$ sin alterar la primera fila, y con la segunda columna podemos hacer que la segunda fila pase a ser $(0, 1, 0, \ldots, 0)$ sin alterar la primera fila.

El proceso puede continuar de este modo hasta convertir la penúltima fila en $(0, \ldots, 0, 1, 0)$. Con la última fila ya no podemos razonar igual porque ya no tenemos filas posteriores. Podemos asegurar igualmente que $a_{nn} = \alpha \neq 0$, pues de lo contrario sería $|A| = 0$, pero ahora no tenemos una columna posterior con la que hacer $a_{nn} = 1$. No obstante, si a la columna i-ésima le sumamos la n-sima multiplicada por $-a_{ni}/a_{nn}$, convertimos la última fila en $(0, \ldots, 0, \alpha)$ sin alterar las filas anteriores. ∎

6.3 Grupos de automorfismos

Si V es un espacio vectorial y $f : V \longrightarrow V$ es un endomorfismo de V, en principio podemos considerar su matriz respecto de dos bases distintas de V, una cuando lo consideramos como espacio inicial y otra cuando lo consideramos como espacio final, pero lo habitual es que sea necesario considerar la misma base en ambos casos, y hablamos entonces de la matriz de f respecto de una base B de V, tal y como ya señalamos en la definición 4.36.

Esto es relevante porque no es cierto que si una matriz es equivalente a la matriz de un endomorfismo f en cierta base, ello implique que es la matriz de f respecto de otra base (lo será respecto de un cierto par de bases de V, pero que no serán necesariamente iguales). Esto lo estudiaremos con detalle en la sección siguiente. Ahora vamos a obtener algunos resultados notables sobre automorfismos de un espacio vectorial.

Definición 6.20 Recordemos que el grupo lineal general $\mathrm{LG}(n, k)$ es el grupo de las matrices regulares $n \times n$ sobre el cuerpo k. El teorema 6.6 implica que la aplicación determinante $\det : \mathrm{LG}(n, k) \longrightarrow k^*$ es un epimorfismo de grupos. Llamaremos *grupo lineal especial* $\mathrm{LE}(n, k)$ al núcleo de este epimorfismo, es decir, el grupo formado por las matrices $n \times n$ de determinante 1.

Si V es un k-espacio vectorial de dimensión n, llamamos $\mathrm{LG}(V)$ al grupo de los automorfismos de V. Es claro que la aplicación que a cada automorfismo le asigna su matriz en una base prefijada de V es un isomorfismo de grupos $\mathrm{LG}(V) \cong \mathrm{LG}(n, k)$.

Si S y T son las matrices de un mismo endomorfismo $h : V \longrightarrow V$ respecto de dos bases de V, entonces $T = P^{-1}SP$, por lo que $|T| = |P|^{-1}|S||P| = |S|$.

Por consiguiente, podemos definir el determinante $\det h$ de un endomorfismo h como el determinante de su matriz en cualquier base. Claramente $\det : \mathrm{LG}(V) \longrightarrow k^*$ es también un epimorfismo de grupos, a cuyo núcleo llamaremos $\mathrm{LE}(V)$.

Claramente, el isomorfismo $\mathrm{LG}(V) \cong \mathrm{LG}(n,k)$ determinado por una base cualquiera se restringe a un isomorfismo $\mathrm{LE}(V) \cong \mathrm{LE}(n,k)$.

Vamos a ver que la nota precedente nos permite calcular sistemas generadores especialmente simples de los grupos lineales $\mathrm{LG}(V)$ y $\mathrm{LE}(V)$. Para ello necesitamos un par de definiciones más:

Sea V un k-espacio vectorial de dimensión n y sea H un hiperplano de V (es decir, un subespacio vectorial de dimensión $n-1$). La *dilatación* de V de hiperplano H, dirección $w \in V \setminus H$ y razón $\alpha \neq 0$ es el automorfismo $u \in \mathrm{LG}(V)$ determinado por que fija a los puntos de H y $u(w) = \alpha w$.

Una *transvección* de V de hiperplano H es un automorfismo $u \in \mathrm{LG}(V)$ tal que existe un $h \in H$ no nulo y una aplicación lineal $f : V \longrightarrow k$ de núcleo H de modo que para todo $x \in V$ se cumpla $u(x) = x + f(x)h$.

Equivalentemente, u es una dilatación de V de razón α si en cierta base su matriz es

$$\begin{pmatrix} \alpha & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}.$$

En particular vemos que $\det u = \alpha$. Similarmente, u es una transvección de V si en cierta base tiene matriz

$$\begin{pmatrix} 1 & & & \\ \alpha & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}.$$

(Basta tomar la base de la forma $h, v, h_2, \ldots, h_{n-1}$, donde $h, h_2, \ldots, h_{n-1}$ es una base de H y $v \in V \setminus H$.) Claramente, las transvecciones tienen determinante 1.

Teorema 6.21 *Sea V un espacio vectorial de dimensión finita sobre un cuerpo k y $u \in \mathrm{LG}(V)$ un automorfismo distinto de la identidad que fije a todos los puntos de un hiperplano H. Entonces u es una dilatación o una transvección, según si su determinante es distinto o igual a 1.*

Demostración: Tomemos $v \in V \setminus H$. Entonces $V = H \oplus \langle v \rangle$. Pongamos que $u(v) = h + \alpha v$, con $h \in H$ y $\alpha \in k$. Si $\alpha \neq 1$ llamamos $w = h + (\alpha - 1)v$, y es fácil ver que $u(w) = \alpha w$, luego h es una dilatación (con determinante α). Supongamos ahora que $\alpha = 1$, de modo que $u(v) = h + v$. Como u no es la identidad, tiene que ser $h \neq 0$. Sea $f : V \longrightarrow k$ la composición de la proyección $V = H \oplus \langle v \rangle \longrightarrow \langle v \rangle$ con el isomorfismo $\beta v \mapsto \beta$. En definitiva, $f(h' + \beta v) = \beta$. Entonces, si $x = h' + \beta v$, tenemos que

$$u(x) = h' + \beta(h + v) = x + \beta h = x + f(x)h,$$

luego u es una transvección. ∎

El resultado al que queríamos llegar es el siguiente:

Teorema 6.22 *Si V es un espacio vectorial de dimensión finita sobre un cuerpo k, el grupo* $\mathrm{LG}(V)$ *está generado por las dilataciones, y el subgrupo* $\mathrm{LE}(V)$ *está generado por las transvecciones. (La primera afirmación requiere claramente que $k \neq \{0,1\}$, pues, si no, la única dilatación es la identidad.)*

Demostración: Es claro que todas las matrices elementales de tipo 2 son matrices de transvecciones, luego la nota final de la sección anterior, traducida de matrices a automorfismos, dice que todo automorfismo $u \in \mathrm{LG}(V)$ se expresa como composición de transvecciones y una dilatación, cuya razón será necesariamente $\det u$. Por lo tanto, si $u \in \mathrm{LE}(V)$, entonces $\alpha = 1$ y la dilatación es la identidad, luego tenemos u expresado como composición de transvecciones.[1]

En general, tenemos también que $\mathrm{LG}(V)$ está generado por las dilataciones y las transvecciones. Ahora basta probar que, suponiendo que $k \neq \{0,1\}$, toda transvección es producto de dos dilataciones. Para ello, fijada una transvección u, consideramos una dilatación cualquiera w respecto al mismo hiperplano con razón distinta de 1. Entonces $w' = u \circ w^{-1}$ es un automorfismo de V que fija a H y tiene determinante $\neq 1$. Por el teorema 6.21 tenemos que es una dilatación, luego la transvección $u = w' \circ w$ es composición de dos dilataciones. ∎

En ocasiones puede ser más conveniente la siguiente versión, en la que sólo interviene una transvección:

Teorema 6.23 *Sea V un k-espacio vectorial de dimensión finita y sea $e_1, \ldots, e_n$ una base de V. Entonces el grupo* $\mathrm{LG}(V)$ *está generado por las aplicaciones siguientes:*

1. *La aplicación f_{ij} que intercambia e_i con e_j y deja fijos a los demás vectores de la base.*
2. *La transvección s dada por $s(e_1) = e_1 + e_2$ y $s(e_i) = e_i$ para $i = 2, \ldots, n$.*
3. *Las dilataciones h_a dadas por $h_a(e_1) = ae_1$ y $h_a(e_i) = e_i$ para $i = 2, \ldots, n$, donde $a \in k \setminus \{0\}$.*

Demostración: Sea $f \in \mathrm{LG}(V)$. Su matriz S en la base dada es equivalente a la matriz identidad, lo que equivale a que es producto de matrices elementales. Podemos considerar las aplicaciones lineales que en la base dada tienen por matriz a cada una de las matrices en que se descompone S, y entonces f es la composición de dichas aplicaciones lineales. Por lo tanto, basta probar que cada una de ellas se expresa a su vez como composición de las aplicaciones consideradas en el enunciado.

Las matrices elementales de tipo 1 se corresponden con las aplicaciones descritas en 1. Las matrices de tipo 2 se corresponden con las transvecciones que cumplen $s^a_{ij}(e_i) = e_i + ae_j$ y dejan invariantes a los demás vectores de la base.

[1] Más precisamente, hemos probado que $\mathrm{LE}(V)$ está generado por las transvecciones cuyas matrices *en una misma base prefijada* tienen la forma que hemos indicado (con unos en la diagonal y ceros salvo en una única posición).

Podemos suponer que $a \neq 0$, pues en caso contrario se trata de la aplicación identidad. Es fácil ver que $s^a_{ij} = p \circ s^a_{12} \circ p^{-1}$, donde p es la aplicación que cumple $p(e_i) = e_1$, $p(e_j) = e_2$ y deja fijos a los demás vectores de la base. Las aplicaciones p y p^{-1} son composiciones de aplicaciones de tipo 1, luego basta estudiar s^a_{12}, pero es fácil ver que

$$s^a_{12} = h_a \circ s^1_{12} \circ h_{1/a},$$

y s^1_{12} es la aplicación del punto 2. del enunciado.

Por último, las matrices elementales del tercer tipo determinan las aplicaciones $f_{i1} \circ h_a \circ f_{i1}$. ∎

6.4 Clasificación de endomorfismos

Abordamos ahora el problema que plantea el hecho de que, al considerar un endomorfismo $f : M \longrightarrow M$ de un módulo libre M, normalmente interesa considerar su matriz respecto a una única base de M en lugar de respecto a dos bases elegidas independientemente. Concretamente, si S, T son las matrices de f respecto a ciertas bases B y C, el mismo razonamiento que cuando teníamos dos módulos nos da ahora que $T = P^{-1}SP$, para cierta matriz regular P (P es la matriz del cambio de base de B a C y P^{-1} es la matriz del cambio de base de C a B). Por ello definimos:

Definición 6.24 Sea A un anillo conmutativo y unitario. Diremos que dos matrices $S, T \in \mathrm{Mat}_n(A)$ son *semejantes* si existe una matriz $P \in \mathrm{LG}(n, A)$ tal que $T = P^{-1}SP$.

Es evidente que la semejanza de matrices es una relación de equivalencia en el conjunto $\mathrm{Mat}_n(A)$. Dos matrices de un mismo endomorfismo de A-módulos libres son semejantes, y si S es la matriz de un endomorfismo f y T es semejante a S, entonces T es la matriz de f en otra base.

Dos matrices semejantes son equivalentes, pero el recíproco no es cierto. Por ejemplo, dos matrices semejantes tienen el mismo determinante, y es fácil encontrar matrices equivalentes con determinantes distintos, luego no semejantes.

La idea fundamental en el estudio de un endomorfismo de un espacio V es encontrarle subespacios invariantes, es decir, encontrar subespacios W tales que $h[W] \subset W$. Por ejemplo, si h es un giro en $\mathbb{R}^3$ (cuyo eje pasa por 0), podemos encontrar dos subespacios invariantes: el plano de giro, donde h es un giro bidimensional, y el eje de giro, donde h es la identidad. La matriz de h será la más simple si tomamos una base que tenga dos vectores en el plano de giro y el tercero en el eje.

Para encontrar subespacios invariantes podemos valernos de la teoría de módulos gracias al planteamiento siguiente: Sea V un espacio vectorial de dimensión finita sobre un cuerpo K. Entonces el conjunto $\mathrm{End}_K(V)$ de todos los endomorfismos de V tiene estructura de anillo (no conmutativo) con la suma definida punto a punto: $(f + g)(v) = f(v) + g(v)$ y tomando como producto

la composición de aplicaciones. Si $h \in \mathrm{End}_K(V)$ definimos el homomorfismo $K[x] \longrightarrow \mathrm{End}_K(V)$ que a cada polinomio $p(x)$ le asigna $p(h)$. Notemos que la unidad de $\mathrm{End}_K(V)$ es el endomorfismo identidad, por lo que la imagen del polinomio 1 es dicha identidad.

Ahora definimos una operación $K[x] \times V \longrightarrow V$ dada por $p(x)v = p(h)(v)$. Así, por ejemplo, $xv = h(v)$, $(x^2+2)v = h\big(h(v)\big) + 2v$, etc. Es fácil ver que V, con su suma de espacio vectorial y esta operación, es un $K[x]$-módulo.

En resumen, hemos dotado a V de estructura de $K[x]$-módulo de modo que la multiplicación por elementos de K es la que ya teníamos en V como espacio vectorial y la multiplicación por x consiste en aplicar h. Esto hace que los subespacios invariantes que estamos buscando sean precisamente los submódulos de V. El teorema siguiente recoge este hecho junto con los resultados que garantizan que podemos aplicar los teoremas de estructura de módulos sobre dominios de ideales principales (notemos que $K[x]$ es un dominio euclídeo):

Teorema 6.25 *Sea V un espacio vectorial de dimensión finita n sobre un cuerpo K, sea h un endomorfismo de V. Entonces*

1. *V es un $K[x]$-módulo finitamente generado.*
2. *Sus submódulos son los subespacios vectoriales W tales que $h[W] \subset W$.*
3. *El núcleo $\mathrm{N}(h)$ es un submódulo de V.*
4. *V es un módulo de torsión.*

Demostración: 1) Notemos que el producto de un polinomio constante por un elemento de V es el producto dado de V como espacio vectorial. Por ello una combinación lineal con coeficientes en K también puede considerarse con coeficientes en $K[x]$, luego una base de V como espacio vectorial es un generador de V como módulo.

2) Si W es un submódulo de V, entonces la suma de elementos de W está en W y el producto de un escalar por un elemento de W está en W, luego W es un subespacio vectorial. Más aún, si $v \in W$, $xv = h(v) \in W$, luego $h[W] \subset W$.

Recíprocamente, si W cumple estas condiciones entonces W es estable para la suma y para el producto por escalares y por x, de donde fácilmente se sigue que W es estable para el producto por cualquier polinomio.

3) es consecuencia de 2).

4) Si $v \in V$, entonces los vectores $v, xv, \ldots, x^n v$ han de estar repetidos o ser linealmente dependientes, luego existen escalares no todos nulos tales que $t_0 v + t_1 xv + \cdots + t_n x^n v = 0$, o sea, $(t^n x^n + \cdots + t_1 x + t_0)v = 0$, luego v es un elemento de torsión. ∎

Definición 6.26 Si V es un espacio vectorial de dimensión finita, h es un endomorfismo de V y $v \in V$, llamaremos *polinomio mínimo* de v ($\mathrm{pol\,mín}\, v$) al único polinomio mónico que genera el ideal

$$o(v) = \{p \in K[x] \mid pv = 0\},$$

que es único si exigimos además que sea un polinomio mónico.

El teorema 4.48 nos garantiza ahora que V se descompone en la forma

$$V = \langle v_1 \rangle_{K[x]} \oplus \cdots \oplus \langle v_r \rangle_{K[x]},$$

donde la notación $\langle v \rangle_{K[x]}$ representa al submódulo generado por v, (mientras que $\langle v \rangle$ representará al subespacio vectorial generado por v), y de modo que los polinomios $\operatorname{pol\,mín}(v_i)$ sean, o bien potencias de primo (divisores elementales) o bien que se dividan sucesivamente (factores invariantes).

Llamaremos *divisores elementales* y *factores invariantes* de h a los de V como $K[x]$-módulo.

Nuestra intención es describir h en términos de sus factores invariantes o sus divisores elementales. En primer lugar probamos que la acción de h sobre cada submódulo monógeno $\langle v \rangle_{K[x]}$ está determinada por $\operatorname{pol\,mín} v$. Empezamos probando que $\operatorname{pol\,mín} v$ determina la dimensión de $\langle v \rangle_{K[x]}$ como espacio vectorial:

Teorema 6.27 *Sea K un cuerpo y V un K-espacio vectorial de dimensión finita n. Sea h un endomorfismo de V. Entonces* $\dim \langle v \rangle_{K[x]} = \operatorname{grad} \operatorname{pol\,mín} v$.

Demostración: Sea $p(x) = \operatorname{pol\,mín} v$ y sea m su grado. Un elemento cualquiera de $\langle v \rangle_{K[x]}$ es de la forma $q(x)v$ con $q(x) \in K[x]$.

Dividamos $q(x) = c(x)p(x) + r(x)$, con $\operatorname{grad} r(x) < m$. Entonces

$$q(x)v = c(x)(p(x)v) + r(x)v = c(x)0 + r(x)v = r(x)v, \tag{6.2}$$

que es combinación lineal de $v, xv, \ldots, x^{m-1}v$.

Estos vectores han de ser distintos y linealmente independientes, pues lo contrario significa que hay un polinomio $q(x) \neq 0$ y de grado a lo sumo $m-1$ tal que $q(x)v = 0$, pero entonces $p(x) \mid q(x)$, lo cual es imposible según los grados. Por lo tanto $v, xv, \ldots, x^{m-1}v$ es una base de $< v >_{K[x]}$ como espacio vectorial. ∎

Ahora ya estamos en condiciones de caracterizar la acción de h sobre un submódulo monógeno:

Teorema 6.28 *Sea K un cuerpo, V un K-espacio vectorial de dimensión finita y h un endomorfismo de V. Son equivalentes:*

1. *Existe un $v \in V$ tal que $V = \langle v \rangle_{K[x]}$ y $\operatorname{pol\,mín} v = \sum_{i=0}^{n} a_i x^i$ (con $a_n = 1$).*
2. *$\dim_K V = n$ y existe una base de V en la cual la matriz de h es*

$$\begin{pmatrix} 0 & 1 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-2} & -a_{n-1} \end{pmatrix}$$

DEMOSTRACIÓN: 2) equivale a que exista una base $\{v_0,\dots,v_{n-1}\}$ de V de manera que $h(v_i)=v_{i+1}$ para $i=0,\dots,n-2$ y $h(v_{n-1})=-\sum_{i=0}^{n-1} a_i v_i$.

Si se cumple 1), entonces $\{v, xv,\dots,x^{n-1}v\}$ es una base que cumple 2).

Si $\{v_0,\dots,v_{n-1}\}$ cumple 2), entonces con $v=v_0$ se cumple que $v_i=x^iv_0$, así como que $x^nv=-\sum_{i=0}^{n-1} a_ix^iv$, es decir, $p(x)=\sum_{i=0}^{n} a_ix^i$ cumple $p(x)v=0$, luego es un múltiplo del polinomio mínimo de v, pero por otro lado es obvio que $V=\langle v\rangle_{K[x]}$, luego el grado del polinomio mínimo de v ha de ser $n=\operatorname{grad}p(x)$. Así pues, $p(x)=\operatorname{pol\,mín}v$. ∎

Ahora veamos cómo describir la acción de h sobre todo el espacio V a través de sus factores invariantes o sus divisores elementales.

Definición 6.29 Si V es un espacio vectorial de dimensión finita y h es un endomorfismo de V, llamamos *polinomio mínimo* de h ($\operatorname{pol\,mín}h$) al último factor invariante.

El polinomio mínimo $p(x)$ de h es múltiplo de los polinomios mínimos de todos los generadores, luego $p(x)$ los anula a todos, y con ellos a todos los elementos de V. Así pues, $p(x)v=p(h)(v)=0$ para todo vector v, o equivalentemente, $p(h)=0$. Además, cualquier otro polinomio que cumpla $p(h)=0$ anula en particular al generador del submódulo correspondiente al último factor invariante, luego tiene que ser múltiplo de $p(x)$. En definitiva, el polinomio mínimo de h es el menor polinomio que anula a h.

Teorema 6.30 *Sea K un cuerpo y V un K-espacio vectorial de dimensión finita n. Sea h un endomorfismo de V. Las afirmaciones siguientes son equivalentes:*

1. *Los factores invariantes (o divisores elementales) de h son $p_1,\dots,p_r$.*
2. *Los polinomios $p_1,\dots,p_r$ cumplen $p_i \mid p_{i+1}$ (o que cada p_i es potencia de primo) y, en una cierta base, la matriz de h es de la forma*

$$\begin{pmatrix} M_1 & & & \\ & M_2 & & \\ & & \ddots & \\ & & & M_r \end{pmatrix}$$

donde cada M_i es la matriz asociada a p_i según el teorema 6.28.

DEMOSTRACIÓN: Consideramos una descomposición de V de tipo (6.2) de modo que los polinomios mínimos de los generadores sean los factores invariantes o los divisores elementales. Como los submódulos son subespacios invariantes, la restricción de h a cada uno de ellos es un endomorfismo y podemos aplicar el teorema 6.28 para obtener una base de cada uno de ellos de modo que la matriz de la restricción de h sea la indicada, o sea, la M_i del enunciado. La unión de las bases de los sumandos directos forma una base de V y es claro que la matriz de h en esta base es la indicada.

Recíprocamente, si la matriz de h en una base es de la forma indicada, podemos expresar V como suma directa de los subespacios generados por los vectores de la base correspondientes a cada una de las cajas M_i de la matriz. Es claro que estos subespacios son invariantes y que M_i es la matriz de la restricción de h al subespacio i-ésimo. Por el teorema 6.28, cada uno de estos subespacios está generado por un v_i cuyo polinomio mínimo es p_i. Por la unicidad concluimos que los p_i son los factores invariantes o los divisores elementales de h. ∎

Las matrices de la forma descrita en el teorema anterior son formas canónicas para la relación de semejanza. En efecto, si $A \in \mathrm{Mat}_n(K)$ y V es cualquier K-espacio vectorial de dimensión n, entonces A es la matriz de un endomorfismo h de V en una base cualquiera de V, pero en otra base h tiene una matriz de la forma del teorema anterior (con factores invariantes), luego A es semejante a una matriz de este tipo.

Por otra parte, si A y B son matrices del tipo del teorema anterior (con factores invariantes) y son semejantes, entonces son matrices de un mismo endomorfismo h de V, y por la unicidad de los factores invariantes se ha de cumplir que $A = B$.

Para divisores elementales se cumple lo mismo salvo por el hecho de que los factores invariantes están ordenados por la divisibilidad, mientras que los divisores elementales no están ordenados de ningún modo, luego las matrices que resultan de cambiar el orden de las cajas (con divisores elementales) son semejantes entre sí sin dejar de ser del tipo del teorema 6.30.

Definición 6.31 Si K es un cuerpo y $A \in \mathrm{Mat}_n(K)$, llamaremos 1^{a} *forma canónica* de A (respectivamente, 2^{a} *forma canónica*) a la única matriz del tipo indicado en el teorema 6.30 para factores invariantes (respectivamente para divisores elementales, con la consiguiente pérdida parcial de unicidad) que es semejante a la matriz A.

De este modo, dos matrices son semejantes si y sólo si sus formas canónicas son iguales. Los teoremas siguientes nos permiten calcular la forma canónica de cualquier matriz.

Teorema 6.32 *Sea K un cuerpo y $A \in \mathrm{Mat}_n(K)$ una matriz del tipo del teorema* 6.28 *para el polinomio mónico $p(x)$. Entonces los factores invariantes (los construidos en el teorema* 6.18*) de la matriz $xI_n - A \in \mathrm{Mat}_n(K[x])$ son todos iguales a* 1 *excepto el último, que es $p(x)$.*

Demostración: Sea $p(x) = \sum_{i=0}^{n} a_i x^i$, con $a_n = 1$. La matriz $xI_n - A$ es

$$\begin{pmatrix} x & -1 & & & & \\ & x & -1 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & x & -1 \\ a_0 & a_1 & a_2 & \cdots & a_{n-2} & a_{n-1}+x \end{pmatrix}$$

Sumamos a la primera columna la segunda multiplicada por x, la tercera multiplicada por x^2, etc. Luego a la segunda columna le sumamos la tercera multiplicada por x, la cuarta multiplicada por x^2, etc. y así sucesivamente. El resultado es:

$$\begin{pmatrix} 0 & -1 & & \\ & \ddots & \ddots & \\ & & & -1 \\ \sum_{i=0}^{n} a_i x^i & \sum_{i=1}^{n} a_i x^i & \cdots & a_{n-1}+x \end{pmatrix}$$

Sumando a la fila n-sima la i-ésima multiplicada por el coeficiente adecuado queda:

$$\begin{pmatrix} 0 & -1 & & \\ & \ddots & \ddots & \\ & & \ddots & -1 \\ \sum_{i=0}^{n} a_i x^i & 0 & \cdots & 0 \end{pmatrix}$$

Finalmente multiplicamos todas las filas salvo la última por -1 y reordenamos las columnas, con lo que llegamos a la forma canónica de la matriz:

$$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \sum_{i=0}^{n} a_i x^i \end{pmatrix}$$

■

Teorema 6.33 *Sea K un cuerpo, V un K-espacio vectorial de dimensión finita n y h un endomorfismo de V. Sea A la matriz de h en 1^a forma canónica. Entonces los factores invariantes de h son los factores invariantes no unitarios de la matriz $xI_n - A$.*

DEMOSTRACIÓN: Por comodidad llamaremos $[A_1, \ldots, A_r]$ a la matriz formada por las cajas $A_1, \ldots, A_r$ situadas sobre su diagonal. Sean $p_1, \ldots, p_r$ los factores invariantes de h. Entonces $xI_n - A = [M_1, \ldots, M_r]$, donde cada M_i tiene forma indicada en el teorema 6.28 para el polinomio p_i.

Por el teorema anterior, M_i es equivalente a una matriz diagonal de la forma $N_i = [1, \ldots, 1, p_i]$. Por lo tanto existen matrices P_i, $Q_i \in \mathrm{LG}_n(K[x])$ tales que $P_i M_i Q_i = N_i$.

Sean $P = [P_1, \ldots, P_r]$ y $Q = [Q_1, \ldots, Q_r]$. Es fácil ver que P y Q son regulares, así como que

$$\begin{aligned} P(xI_n - A)Q &= [P_1 M_1 Q_1, \ldots P_r M_r Q_r] = [N_1, \ldots, N_r] \\ &= [1, \ldots, 1, p_1, \ldots, 1, \ldots, 1, p_r], \end{aligned}$$

luego la forma canónica de $xI_n - A$ es $[1, \ldots, 1, p_1, \ldots, p_r]$. ■

El resultado definitivo es el siguiente:

Teorema 6.34 *Sea K un cuerpo, V un K-espacio vectorial de dimensión finita n y h un endomorfismo de V. Sea A la matriz de h en cualquier base. Entonces los factores invariantes de h son los factores invariantes no unitarios de la matriz $xI_n - A$.*

DEMOSTRACIÓN: Sea B la matriz de h en 1ª forma canónica. Como A y B son matrices de h, son semejantes, es decir, existe una matriz $P \in \mathrm{LG}_n(K)$ tal que $B = P^{-1}AP$. Entonces

$$P^{-1}(xI_n - A)P = xP^{-1}I_nP - P^{-1}AP = xI_n - B.$$

Por lo tanto, las matrices $xI_n - A$ y $xI_n - B$ son equivalentes, luego tienen los mismos factores invariantes y, por el teorema anterior, los no unitarios son los factores invariantes de h. ∎

Definición 6.35 Sea K un cuerpo y $A \in \mathrm{Mat}_n(K)$. Llamaremos *factores invariantes* de A a los factores invariantes de $xI_n - A$ (en el sentido de 6.18).

Hemos demostrado que los factores invariantes de un endomorfismo son los factores invariantes no unitarios de cualquiera de sus matrices.

También es obvio ahora que dos matrices son semejantes si y sólo si tienen los mismos factores invariantes (si tienen los mismos factores invariantes tienen la misma forma canónica).

Observemos que en principio tenemos dos definiciones de factores invariantes de una matriz de $\mathrm{Mat}_n(K)$, la dada en el teorema 6.18 y la que acabamos de dar, pero sucede que la definición según el teorema 6.18 no tiene interés para matrices sobre un cuerpo, ya que todos los factores invariantes en este sentido son iguales a 1. Lo único que interesa es su número, o sea, el rango.

Los factores invariantes de una matriz A se calculan sin dificultad mediante el algoritmo dado en el teorema 6.18. Los divisores elementales se calculan descomponiendo los factores invariantes en potencias de primos.

Llamaremos *polinomio mínimo* de una matriz A a su último factor invariante. Así, si A es la matriz de un endomorfismo h, se cumple que $\mathrm{pol\,mín}\, A = \mathrm{pol\,mín}\, h$. Dado el isomorfismo entre endomorfismos y matrices, es claro que si $p(x) = \mathrm{pol\,mín}\, A$, entonces $p(A) = 0$, y si $q(x) \in K[x]$, entonces $q(A) = 0$ si y sólo si $p(x) \mid q(x)$.

Ejemplo Vamos a calcular los factores invariantes de la matriz

$$A = \begin{pmatrix} 8 & 2 & 10 & -6 \\ 12 & 6 & 20 & -12 \\ 2 & 3 & 7 & -4 \\ 13 & 9 & 25 & -15 \end{pmatrix}$$

Para ello aplicamos el algoritmo del teorema 6.18 a $xI_4 - A$:

$$\begin{pmatrix} x-8 & -2 & -10 & 6 \\ -12 & x-6 & -20 & 12 \\ -2 & -3 & x-7 & 4 \\ -13 & -9 & -25 & x+15 \end{pmatrix}$$

En primer lugar intercambiamos las dos primeras columnas para situar en la posición $(1,1)$ un polinomio de norma mínima (o sea, de grado mínimo) (el -2). Dividiendo la fila 1 entre -2 tenemos un 1 en el lugar $(1,1)$ y restando a cada columna la primera multiplicada por su primer coeficiente llegamos hasta la matriz

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ x-6 & \frac{x^2}{2} - 7x + 12 & -5x+10 & 3x-6 \\ -3 & -\frac{3}{2}x + 10 & x+8 & -5 \\ -9 & -\frac{9}{2}x + 23 & 20 & x+12 \end{pmatrix}$$

Ahora restamos a cada fila la primera multiplicada por su primer coeficiente y hacemos ceros debajo del primer 1 (el resto de la matriz no se modifica). Como el 1 divide a todos los coeficientes restantes podemos continuar con la submatriz 3×3.

Llevamos el -5 al lugar $(2,2)$ (otra opción sería llevar el 20). Dividimos la segunda fila entre -5 y obtenemos otro 1, con el que hacemos ceros a su derecha y bajo él. El resultado es:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3x^2 - 7x + 2 & -2x^2 + 4x \\ 0 & 0 & 2x^2 - 8x + 8 & -3x^2 - 10 \end{pmatrix}$$

El cociente de $-2x^2 + 4x$ entre $3x^2 - 7x + 2$ es $-2/3$ y el resto $-(2/3)(x-2)$. Restamos a la cuarta columna la tercera multiplicada por $-2/3$ con lo que en la posición $(3,4)$ queda $-(2/3)(x-2)$. Lo pasamos a la posición $(3,3)$ y multiplicamos por $-3/2$:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & x-2 & -\frac{9}{2}x^2 + \frac{21}{2}x - 3 \\ 0 & 0 & -\frac{5}{3}x^2 + \frac{17}{3}x - \frac{11}{3} & 2x^2 - 8x + 8 \end{pmatrix}$$

Al dividir el polinomio de la posición $(3,4)$ entre $x-2$ la división nos da exacta, a saber: $-(3/2)(3x-1)(x-2)$. Restamos a la cuarta fila la primera multiplicada por $-(3/2)(3x-1)$ y nos queda (factorizando los polinomios):

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & x-2 & 0 \\ 0 & 0 & -\frac{1}{3}(5x-7)(x-2) & -\frac{15}{2}(x-2)(x-1)^2 \end{pmatrix}$$

Restamos a la cuarta fila la tercera multiplicada por $-(1/3)(5x-7)$ y queda un 0 en el lugar $(4,3)$. Multiplicando la última fila por $-2/15$ queda:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & x-2 & 0 \\ 0 & 0 & 0 & (x-2)(x-1)^2 \end{pmatrix}$$

Por lo tanto los factores invariantes de la matriz A de partida son

$$x-2 \quad \text{y} \quad (x-2)(x-1)^2.$$

Los divisores elementales son $x-2$, $x-2$ y $(x-1)^2$. El polinomio mínimo de A es $\operatorname{pol\,mín} A = (x-2)(x-1)^2 = x^3-4x^2+5x-2$. Las formas canónicas de A son:

$$\left(\begin{array}{c|ccc} 2 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 2 & -5 & 4 \end{array}\right) \qquad \left(\begin{array}{c|c|cc} 2 & 0 & 0 & 0 \\ \hline 0 & 2 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 2 \end{array}\right)$$

■

Hemos visto que si A y B son matrices matrices semejantes, o sea, si existe $P \in \mathrm{LG}_n(K)$ tal que $B = P^{-1}AP$, entonces también $xI_n - B = P^{-1}(xI_n - A)P$, luego tomando determinantes queda $|xI_n - B| = |P|^{-1}|xI_n - A||P| = |xI_n - A|$.

Definición 6.36 *Llamaremos polinomio característico de $A \in \mathrm{Mat}_n(K)$ al polinomio* $\operatorname{pol\,car} A = |xI_n - A|$. *Es fácil ver que es mónico de grado n.*

Acabamos de probar que las matrices semejantes tienen el mismo polinomio característico, por lo que podemos definir el polinomio característico de un endomorfismo como el de cualquiera de sus matrices.

El polinomio característico es mucho más fácil de calcular que los factores invariantes o los divisores elementales y en ocasiones proporciona suficiente información sobre una matriz dada.

Por ejemplo, el polinomio característico de la matriz A anterior resulta ser $\operatorname{pol\,car} A = (x-2)^2(x-1)^2$.

Este ejemplo muestra que el polinomio característico está muy relacionado con los factores invariantes y los divisores elementales. La relación exacta nos la da el teorema siguiente:

Teorema 6.37 *Sea K un cuerpo y $A \in \mathrm{Mat}_n(K)$. Entonces el polinomio característico de A es el producto de sus factores invariantes (luego también el de los divisores elementales).*

Demostración: Supongamos primero que A es de la forma indicada en el teorema 6.28, es decir, que tiene un único factor invariante no unitario igual a $\sum_{i=0}^{n} a_i x^i$.

Es fácil ver que en este caso $|xI_n - A| = \sum_{i=0}^{n} a_i x^i$ (basta desarrollar el determinante por la primera columna).

Ahora, si una matriz A está formada por dos cajas ($A = [M, N]$ con la notación del teorema 6.33) es fácil ver que $[M, N] = [M, I][I, N]$, donde I representa en cada caso a la matriz identidad de las dimensiones adecuadas. Por lo tanto $|A| = \big|[M, I]\big|\,\big|[I, N]\big| = |M|\,|N|$. De aquí se sigue en general que $\big|[M_1, \ldots, M_r]\big| = |M_1| \cdots |M_r|$.

Si A es una 1^{a} forma canónica, entonces $A = [M_1, \ldots, M_r]$, donde las matrices M_i son del tipo del teorema 6.28 para cada factor invariante de A.

Entonces $xI_n - A = [xI - M_1, \ldots, xI - M_r]$, luego

$$\operatorname{pol}\operatorname{car} A = |xI - M_1| \cdots |xI - M_r|.$$

Por el caso particular ya probado estos factores son los factores invariantes de A, luego el teorema es cierto para formas canónicas.

Finalmente, si A es cualquier matriz, su polinomio característico es el mismo que el de su 1^{a} forma canónica, que es el producto de sus factores invariantes, que son también los factores invariantes de A. ∎

Una consecuencia es el llamado *teorema de Cayley,* según el cual toda matriz es raíz de su polinomio característico (porque éste es un múltiplo de su polinomio mínimo).

Las matrices semejantes tienen el mismo polinomio característico, aunque el recíproco no es cierto, pues el polinomio característico no determina los factores invariantes o los divisores elementales.

Por ejemplo, sabiendo que $\operatorname{pol}\operatorname{car} A = (x-2)^2(x-1)^2$, los invariantes de A podrían ser:

Factores invariantes	Divisores elementales
$(x-2)^2(x-1)^2$	$(x-2)^2,\ (x-1)^2$
$x-2,\ (x-2)(x-1)^2$	$x-2,\ x-2(x-1)^2$
$x-1,\ (x-2)^2(x-1)$	$x-1,\ x-1,\ (x-2)^2$
$(x-2)(x-1),\ (x-2)(x-1)$	$x-2,\ x-2,\ x-1,\ x-1$

Los coeficientes del polinomio característico $p(x)$ de una matriz A son invariantes por semejanza. El primero y el último de estos coeficientes (aparte del director, que es igual a 1) son especialmente simples.

El término independiente es $p(0) = |0I_n - A| = |-A| = (-1)^n|A|$. De aquí se sigue algo que ya sabíamos: que las matrices semejantes tienen el mismo determinante.

Para calcular el coeficiente de x^{n-1} observamos que, según la definición de determinante, uno de los sumandos de $p(x)$ es

$$(x - a_{11}) \cdots (x - a_{nn}) = x^n - (a_{11} + \cdots + a_{nn})x^{n-1} + \cdots$$

y ningún otro sumando tiene monomios de grado $n-1$. Así pues, el coeficiente de grado $n-1$ de $p(x)$ es $-(a_{11} + \cdots + a_{nn})$.

Definición 6.38 Llamaremos *traza* de una matriz $A \in \mathrm{Mat}_n(K)$ a

$$\mathrm{Tr}(A) = a_{11} + \cdots + a_{nn}.$$

Hemos probado que las matrices semejantes tienen el mismo determinante y la misma traza. Por ello podemos definir el determinante y la traza de un endomorfismo como el de cualquiera de sus matrices.

Las raíces del polinomio característico de un endomorfismo contienen información muy importante sobre el mismo. Vamos a comprobarlo. Sea h un endomorfismo de un espacio vectorial V sobre un cuerpo K y sea A su matriz en una base. Sea $\alpha \in K$ una raíz de $\operatorname{pol}\operatorname{car} A$. Entonces $|\alpha I_n - A| = 0$, lo que significa que la matriz $\alpha I_n - A$ no es regular, luego el endomorfismo que determina, es decir, $\alpha I - h$, no es inyectivo, luego su núcleo no es trivial, luego existe un vector $v \in V$ no nulo tal que $h(v) = \alpha v$. Recíprocamente, si existe un vector v no nulo tal que $h(v) = \alpha v$, invirtiendo el razonamiento llegamos a que α es raíz de $\operatorname{pol}\operatorname{car} A$.

Definición 6.39 Sea h un endomorfismo de un espacio vectorial V sobre un cuerpo K. Si un vector $v \in V$ no nulo cumple $h(v) = \alpha v$ para cierto $\alpha \in K$, diremos que v es un *vector propio* de h y que α es un *valor propio* de h.

Acabamos de probar que los valores propios de un endomorfismo h son exactamente las raíces de su polinomio característico.

Cada vector propio tiene asociado un único valor propio. Recíprocamente, si $\alpha \in K$ es un valor propio de h, entonces α tiene asociado el espacio

$$S(h, \alpha) = \{v \in V \mid h(v) = \alpha v\},$$

que es un subespacio vectorial no trivial de V al que llamaremos *espacio fundamental* de α.

Un caso especialmente interesante es el subespacio $S(h, 1)$, que está formado por los puntos fijos de h.

Ejemplo Sea h un endomorfismo de $\mathbb{R}^4$ cuya matriz en la base canónica sea la matriz A del ejemplo anterior. Hemos calculado

$$\operatorname{pol}\operatorname{car} A = (x-2)^2(x-1)^2,$$

luego sus valores propios son 1 y 2.

El sistema de ecuaciones lineales $(u, v, w, x)A = (u, v, w, x)$ tiene solución $(-4v, v, -5v, 2v)$ para cualquier v, luego

$$S(h, 1) = \langle (-4, 1, -5, 2) \rangle .$$

Igualmente se calcula

$$S(h, 2) = \langle (-2, 1, 0, 0), (-3, 0, -6, 2) \rangle .$$

■

Como el polinomio característico es el producto de los divisores elementales, los valores propios son las raíces de los divisores elementales. Como éstos son potencias de irreducibles, es obvio que α es un valor propio de h si y sólo si uno de los divisores elementales de h tiene la forma $(x-\alpha)^e$. El teorema siguiente muestra exactamente la relación entre los valores propios y los divisores elementales.

Teorema 6.40 *Sea K un cuerpo, V un K-espacio vectorial de dimensión finita y h un endomorfismo de V. Sea $V = \langle v_1\rangle_{K[x]} \oplus \cdots \oplus \langle v_r\rangle_{K[x]}$ la descomposición de V en submódulos asociada a los divisores elementales, es decir, tal que $\operatorname{pol\,mín} v_i = p_i(x)^{e_i}$ con $p_i(x)$ irreducible. Sea $\alpha \in K$ un valor propio de h y supongamos que $p_i = x-\alpha$ para $i = 1,\dots,s$. Entonces:*

1. $S(h,\alpha) = \big\langle (x-a)^{e_1-1}v_1, \dots, (x-\alpha)^{e_s-1}v_s\big\rangle_k$.
2. $\dim S(h,\alpha) = s$.

Si $\alpha_1,\dots,\alpha_t$ son todos los valores propios de h, entonces se tiene la suma directa

$$S = S(h,\alpha_1) \oplus \cdots \oplus S(h,\alpha_t).$$

Al espacio S se le llama espacio fundamental *de h.*

Demostración: Un vector $v \neq 0$ es propio con valor propio α si y sólo si $(x-\alpha)v = 0$, luego es inmediato que los vectores $(x-\alpha)^{e_1-1}v_1, \dots, (x-\alpha)^{e_s-1}v_s$ están en $S(h,\alpha)$. Por lo tanto tenemos la inclusión

$$\big\langle (x-\alpha)^{e_1-1}v_1, \dots, (x-\alpha)^{e_s-1}v_s\big\rangle_k \leq S(h,\alpha).$$

Veamos ahora la otra inclusión. Si $v \in S(h,\alpha)$, por la descomposición de V tenemos en principio que $v = f_1(x)v_1 + \cdots + f_r(x)v_r$ para ciertos polinomios $f_i(x)$. Según el teorema 6.27 podemos exigir que $\operatorname{grad} f_i(x) < e_i$.

Como v es un vector propio, se cumple que $(x-\alpha)v = 0$, o sea,

$$(x-\alpha)f_1(x)v_1 + \cdots + (x-\alpha)f_r(x)v_r = 0.$$

Como la suma es directa cada sumando es nulo y, por definición de polinomio mínimo, resulta que $p_i(x)^{e_i} \mid (x-\alpha)f_i(x)$ para $i = 1,\dots,r$.

Si $i > s$ tenemos que $p_i(x) \neq x-\alpha$, luego $p_i(x)^{e_i} \mid f_i(x)$ y así $f_i(x)v_i = 0$, con lo que la expresión de v se reduce a

$$v = f_1(x)v_1 + \cdots + f_s(x)v_s.$$

Para $i \leq s$ tenemos que $(x-\alpha)^{e_i} \mid (x-\alpha)f_i(x)$, de donde $(x-\alpha)^{e_i-1} \mid f_i(x)$, es decir, $f_i(x) = q_i(x)(x-\alpha)^{e_i-1}$, pero tenemos también que $\operatorname{grad} f_i(x) < e_i$, luego la única posibilidad para $q_i(x)$ es que sea una constante $q_i \in K$. Con esto la expresión de v queda en

$$v = q_1(x-\alpha)^{e_1-1}v_1 + \cdots + q_s(x-\alpha)^{e_s-1}v_s,$$

lo que prueba la igualdad.

Estos generadores son independientes, pues cada uno está contenido en un sumando directo distinto. Por la misma razón los distintos espacios fundamentales tienen suma directa, pues están contenidos en sumandos directos distintos (los correspondientes a los divisores elementales con raíz el valor propio correspondiente). ∎

La matriz A de los ejemplos anteriores tiene divisores elementales $x-2$, $x-2$ y $(x-1)^2$, por lo que el teorema anterior nos da que $\dim S(h,2) = 2$ y $\dim S(h,1) = 1$ (como ya habíamos comprobado).

6.5 Formas bilineales

Al estudiar los determinantes hemos introducido el concepto de forma multilineal. Conviene estudiar más detalladamente el caso de las formas bilineales sobre espacios vectoriales de dimensión finita.

Definición 6.41 Sea V un espacio vectorial sobre un cuerpo K. Una *forma bilineal* en V es una aplicación $F : V \times V \longrightarrow K$ tal que para todo $v_1, v_2, v \in V$, y todo $a, b \in K$ se cumple

$$F(av_1 + bv_2, v) = aF(v_1, v) + bF(v_2, v),$$
$$F(v, av_1 + bv_2) = aF(v, v_1) + bF(v, v_2).$$

La aplicación determinante es un ejemplo de forma bilineal en K^2, pero es antisimétrica, y ahora nos vamos a interesar por las formas bilineales opuestas:

Una forma bilineal F es *simétrica* si para todo par de vectores $v_1, v_2 \in V$ se cumple $F(v_1, v_2) = F(v_2, v_1)$.

Por ejemplo, un producto escalar en el sentido de [G 3.19] es un ejemplo de forma bilineal simétrica (al que se le exigen propiedades adicionales).

Al igual que ocurre con las aplicaciones lineales, las formas bilineales quedan determinadas por una base y una matriz:

Si $B = (v_1, \ldots, v_m)$ es una base ordenada de V, llamaremos *matriz* de F *respecto* a dicha base a

$$\mathrm{M}_B(F) = \big(F(v_i, v_j)\big).$$

Teorema 6.42 *Sea $F : V \times V \longrightarrow K$ una forma bilineal en un K-espacio vectorial V. Sea $B = (v_1, \ldots, v_n)$ una base ordenada de V. Entonces $\mathrm{M}_B(F)$ es la única matriz de $\mathrm{Mat}_n(K)$ tal que para todo $v, v' \in V$ se cumple*

$$F(v, v') = \Phi_B(v)\, \mathrm{M}_B(F)\, \Phi_B(v')^t.$$

Demostración: Sea $\Phi_B(v) = (x_1, \ldots, x_n)$ y $\Phi_B(v') = (y_1, \ldots, y_n)$. Entonces

$$\begin{aligned} F(v, v') &= F\Big(\sum_{i=1}^n x_i v_i, \sum_{j=1}^n y_j v_j\Big) = \sum_{i=1}^n \sum_{j=1}^n x_i\, F(v_i, v_j)\, y_j \\ &= \Phi_B(v)\, \mathrm{M}_B(F)\, \Phi_B(v')^t. \end{aligned}$$

Si una matriz tiene esta propiedad, aplicándola a los vectores v_i y v_j obtenemos que su componente (i,j) ha de ser precisamente $F(v_i,v_j)$. ∎

Ejercicio: Probar que una forma bilineal es simétrica si y sólo si su matriz en una base cualquiera es simétrica.

Definición 6.43 La *forma cuadrática* asociada a una forma bilineal simétrica $F : V \times V \longrightarrow K$ es la aplicación $q : V \longrightarrow K$ dada por $q(v) = F(v,v)$.

Si el cuerpo K tiene característica distinta de 2 entonces F está determinada por q en el sentido de que si dos formas bilineales simétricas determinan la misma forma cuadrática es que son iguales. En efecto, basta calcular

$$F(v+w, v+w) = F(v,v) + F(w,w) + 2F(v,w)$$

y despejar

$$F(v,w) = \frac{q(v+w) - q(v) - q(w)}{2}.$$

En particular, fijada una base de V, tenemos una biyección entre las matrices simétricas $n \times n$ y las formas cuadráticas en V.

Las formas bilineales están muy relacionadas con el concepto de espacio dual, que introducimos seguidamente:

Definición 6.44 Si K es un cuerpo y V es un K-espacio vectorial, llamaremos *espacio dual* de V a $V^* = \mathrm{Hom}_K(V,K)$, es decir, al espacio vectorial de todas las aplicaciones lineales de V en K.

Si V tiene dimensión finita n sabemos que $V^* \cong \mathrm{Mat}_{n\times 1}(K)$, luego en particular $\dim V^* = \dim V$.

Con más exactitud, si $B = (v_1, \ldots, v_n)$ es una base de V y fijamos 1 como base de K, entonces la aplicación M_B^1 determina un isomorfismo entre V^* y $\mathrm{Mat}_{n\times 1}(K)$. Una base del segundo espacio la forman los vectores $e_i = (\delta_{ij})$. La antiimagen de e_i por el isomorfismo es una aplicación v_i^* tal que $M_B^1(v_i^*) = e_i$, lo que significa que $v_i^*(v_j) = \delta_{ij}$ (donde δ_{ij} vale 1 si $i = j$ y 0 en caso contrario).

La base $B^* = (v_1^*, \ldots, v_n^*)$ de V^* determinada por la relación $v_i^*(v_j) = \delta_{ij}$ se llama *base dual* de $(v_1, \ldots, v_n)$.

Es fácil hallar las coordenadas de un elemento de V^* respecto a esta base:

Teorema 6.45 *Sea K un cuerpo, V un K-espacio vectorial y $B = (v_1, \ldots, v_n)$ una base V. Entonces, las coordenadas en la base dual $B^* = (v_1^*, \ldots, v_n^*)$ de un elemento cualquiera $v^* \in V^*$ son $\Phi_{B^*}(v^*) = \big(v^*(v_1), \ldots, v^*(v_n)\big)$.*

Demostración: Basta observar que

$$\left(\sum_{i=1}^n v^*(v_i)v_i^*\right)(v_j) = \sum_{i=1}^n v^*(v_i)v_i^*(v_j) = \sum_{i=1}^n v^*(v_i)\delta_{ij} = v^*(v_j),$$

y como ambas aplicaciones lineales coinciden sobre la base B, han de ser iguales, o sea, $v^* = \sum_{i=1}^n v^*(v_i)v_i^*$. ∎

Ejercicio: Probar que si $f : V \longrightarrow W$ es una aplicación lineal entre espacios vectoriales, entonces la *aplicación dual* $f^* : W^* \longrightarrow V^*$ dada por $f^*(w^*) = f \circ w^*$ es también lineal. Además f^* es inyectiva (suprayectiva) si y sólo si f es suprayectiva (inyectiva). Dadas bases B y C, hallar la relación entre $\mathrm{M}_B^C(f)$ y $\mathrm{M}_{C^*}^{B^*}(f^*)$.

Si $F : V \times V \longrightarrow K$ es una forma bilineal simétrica tenemos definida una aplicación lineal

$$\begin{array}{rccccc} \iota_F : V & \longrightarrow & V^* & & \\ v & \mapsto & V & \longrightarrow & K \\ & & w & \mapsto & F(v,w) \end{array}$$

Consideremos una base $B = (v_1, \ldots, v_n)$ de V y vamos a calcular la matriz de esta aplicación respecto a las bases B y B^*. El elemento (i, j) de esta matriz será la coordenada j-ésima de $\iota_F(v_i)$, pero según el teorema anterior se trata de $\iota_F(v_i)(v_j) = F(v_i, v_j)$. Por lo tanto $\mathrm{M}_B^{B^*}(\iota_F) = \mathrm{M}_B(F)$.

Definición 6.46 Una forma bilineal simétrica $F : V \times V \longrightarrow K$ es *regular* si la aplicación ι_F es inyectiva.

Si el espacio V tiene dimensión finita esto equivale a que ι_F sea un isomorfismo, y por el teorema 4.38 esto equivale a que $\mathrm{M}_B(F)$ sea una matriz regular (para una base cualquiera B de V), o a que $|\mathrm{M}_B(F)| \neq 0$.

Así pues, toda forma bilineal simétrica regular en un espacio vectorial de dimensión finita induce un isomorfismo entre V y su espacio dual, del que hay que destacar que no depende de ninguna elección de bases. Si $B = (v_1, \ldots, v_n)$ es una base de V, la antiimagen por ι_F de su base dual es una base $B^* = (v_1^*, \ldots, v_n^*)$ de V (a la que también llamaremos *base dual* de B) caracterizada por que $F(v_i, v_j^*) = \delta_{ij}$, para todo par de índices i, j.

Nos planteamos ahora el problema análogo al que ya nos hemos planteado para los endomorfismos de un espacio vectorial: si tenemos dos bases de un espacio vectorial, ¿qué relación hay entre las matrices de una misma forma bilineal simétrica en ambas bases? ¿Cómo saber si dos matrices dadas corresponden a la misma forma bilineal en bases distintas? La primera pregunta tiene una respuesta muy simple:

Definición 6.47 Diremos que dos matrices simétricas A y B son *congruentes* si existe una matriz regular M tal que $A = MBM^t$.

Es claro que todas las matrices correspondientes a una misma forma bilineal simétrica son congruentes entre sí, y que si B es la matriz de una forma F en una cierta base, cualquier matriz A congruente con B es la matriz de F en una base adecuada. En efecto, dadas dos bases de V, si X e Y son las coordenadas de v y w respecto a una de ellas, sus coordenadas respecto de la otra serán de la forma XM e YM, donde M es la matriz de cambio de base, luego $F(v, w) = XMBM^tY^t$, donde B es la matriz de F en esta segunda base. Por la unicidad de la matriz de una forma bilineal, la matriz de F en la primera base ha de ser $A = MBM^t$. Esto prueba que las matrices de F respecto de dos

bases de V son congruentes. Igualmente se razona que si A es la matriz de F en una base, entonces B es la matriz de F en la base determinada por M a partir de la base dada.

Ahora encontraremos invariantes y formas canónicas para las matrices de las formas bilineales simétricas, de modo análogo a como hicimos con las matrices de los endomorfismos de un espacio vectorial. Consideraremos únicamente matrices y espacios vectoriales definidos sobre un cuerpo arbitrario K de característica distinta de 2.

La guía del procedimiento que vamos a seguir consiste en tratar a las formas bilineales sobre un espacio vectorial como si fueran un producto escalar. Así, diremos que dos vectores de un espacio vectorial V son *ortogonales* respecto a una forma bilineal simétrica $F : V \times V \longrightarrow K$ si cumplen $F(v, w) = 0$. Lo representaremos $v \perp w$.

Hemos de tener presente que la semejanza sólo llega hasta cierto punto. Así, para una forma arbitraria pueden ocurrir casos como que un vector no nulo sea ortogonal a sí mismo, o incluso a todos los demás vectores. Pese a ello, vamos a probar que V tiene siempre una base ortogonal, es decir, una base $v_1, \ldots, v_n$ tal que $v_i \perp v_j = 0$ cuando $i \neq j$.

Teorema 6.48 *Sea V un espacio vectorial de dimensión finita sobre un cuerpo de característica distinta de 2 y sea F una forma bilineal simétrica en V. Entonces V admite una base ortogonal respecto a F.*

Demostración: Lo probaremos por inducción sobre la dimensión de V. En dimensión 1 cualquier base es ortogonal. Supuesto cierto para espacios de dimensión $n-1$, consideremos un espacio de dimensión n. Si F es idénticamente nula cualquier base es ortogonal. En caso contrario existen vectores v, w tales que $F(v, w) \neq 0$. Entonces

$$F(v + w, v + w) = F(v, v) + 2F(v, w) + F(w, w).$$

Necesariamente, uno de los tres vectores v, w, $v + w$ no es ortogonal a sí mismo. Llamemos v_1 a este vector. Sea

$$W = \{v \in V \mid v \perp v_1\}.$$

Claramente W es un subespacio vectorial de V (el subespacio ortogonal a v_1). En general los subespacios ortogonales no tienen el buen comportamiento del caso euclídeo, pero en este caso podemos probar que $V = \langle v_1 \rangle \oplus W$. En efecto, si $w \in W \cap \langle v_1 \rangle$ entonces $w = \alpha v_1 \perp v_1$, es decir, $\alpha F(v_1, v_1) = 0$, lo que implica $\alpha = 0$. Por consiguiente $\langle v_1 \rangle \cap W = 0$.

Por otra parte, W es el núcleo de la aplicación lineal $V \longrightarrow K$ dada por $v \mapsto F(v_1, v)$, cuya imagen es todo K (porque no es nula), luego $\dim W = n-1$, y esto implica que $V = \langle v_1 \rangle \oplus W$.

Por hipótesis de inducción W tiene una base ortogonal respecto a la restricción de F, digamos $v_2, \ldots, v_n$. Es claro que $v_1, \ldots, v_n$ es una base ortogonal de V. ∎

Es obvio que la matriz de una forma bilineal en una base ortogonal es una matriz diagonal. Por lo tanto hemos probado lo siguiente:

Teorema 6.49 *Toda matriz simétrica sobre un cuerpo de característica distinta de* 2 *es congruente con una matriz diagonal.*

(Observemos que la simetría de la matriz es necesaria. Éste es el motivo por el cual nos hemos restringido a formas simétricas.)

Las matrices diagonales no son formas canónicas, pues nada impide que dos matrices diagonales sean congruentes. El problema de encontrar invariantes y formas canónicas se vuelve extremadamente delicado a partir de este punto, y depende profundamente del cuerpo considerado. Sólo hay un sencillo resultado que podemos añadir en general: Una matriz diagonal de la forma $A = [a_1^2 b_1, \ldots, a_n^2 b_n]$, con cada $a_i \neq 0$ es congruente con $B = [b_1, \ldots, b_n]$, pues $A = MBM^t$, donde $M = [a_1, \ldots, a_n]$, es decir, podemos eliminar cuadrados de la diagonal. Esto nos resuelve el problema en cuerpos donde todo elemento sea un cuadrado, como es el caso de $\mathbb{C}$ y, en general, de todo cuerpo algebraicamente cerrado. En efecto:

Teorema 6.50 *Sea K un cuerpo de característica distinta de* 2 *en el que todo elemento sea un cuadrado. Entonces toda matriz simétrica A sobre K es congruente con una única matriz de la forma $[\underbrace{1, \ldots, 1}_{r}, 0, \ldots, 0]$, donde r es el rango de A.*

En efecto, toda matriz A es congruente con una matriz diagonal, que será de la forma $[a_1^2, \ldots, a_r^2, 0, \ldots, 0]$, y podemos eliminar los cuadrados para convertirlos en unos. Es obvio que dos matrices congruentes son equivalentes, luego tienen el mismo rango, luego el número de ceros y unos que aparecen es fijo.

En particular dos matrices simétricas sobre $\mathbb{C}$ (de las mismas dimensiones) son congruentes si y sólo si tienen el mismo rango, y esto es comprobable en la práctica.

Consideramos ahora el caso de una forma bilineal simétrica sobre un cuerpo ordenado R. En este caso, podemos encontrar otro invariante además del rango:

Teorema 6.51 (Ley de inercia de Sylvester) *Todas las matrices diagonales congruentes con una matriz simétrica A sobre un cuerpo ordenado R tienen el mismo número s de coeficientes positivos, al que llamaremos* signatura *de A. El número de coeficientes negativos será necesariamente $r - s$, donde r es el rango de A.*

Demostración: Consideremos una forma bilineal F en un espacio vectorial V definida por la matriz A en una cierta base. Sean $v_1, \ldots, v_n$ y $w_1, \ldots, w_n$ dos bases de V en las que la matriz de F sea diagonal. Podemos suponer que estas matrices son $[a_1, \ldots, a_r, 0, \ldots, 0]$ y $[b_1, \ldots, b_r, 0, \ldots, 0]$, donde r es el rango de las matrices,

$$a_i > 0 \quad \text{para } i = 1, \ldots, s, \quad a_i < 0 \quad \text{para } i = s+1, \ldots, r,$$

$$b_i > 0 \quad \text{para } i = 1, \ldots, s', \quad b_i < 0 \quad \text{para } i = s'+1, \ldots, r.$$

Hemos de probar que $s = s'$.

Sean $S = \langle v_1, \ldots, v_s\rangle$, $T = \langle v_{s'+1}, \ldots, v_n\rangle$. Veamos que $S \cap T = 0$. En efecto, todo $v \in S$ es de la forma $v = \sum_{i=1}^{s} \alpha_i v_i$, luego

$$F(v,v) = \sum_{i,j=1}^{s} \alpha_i\alpha_j F(v_i,v_j) = \sum_{i=1}^{s} \alpha_i^2 F(v_i,v_i) = \sum_{i=1}^{s} \alpha_i^2 a_i \geq 0.$$

Además $F(v,v) = 0$ si y sólo si $\alpha_i = 0$ para todo i, si y sólo si $v = 0$.

Similarmente se prueba que si $v \in T$ entonces $F(v,v) \leq 0$ y $F(v,v) = 0$ si y sólo si $v = 0$. Ahora es claro que $S \cap T = 0$.

Tomando dimensiones vemos que $s + n - s' \leq n$, o sea, $s \leq s'$. Del mismo modo se prueba la desigualdad contraria. ∎

El teorema anterior nos permite definir la *signatura* de una forma bilineal simétrica sobre un cuerpo ordenado como la de cualquiera de sus matrices. Esto no nos da todavía formas canónicas, pero si el cuerpo R es euclídeo, es decir, si todo elemento positivo es un cuadrado, entonces todo elemento de R es de la forma $\pm a^2$. Por lo tanto toda matriz simétrica sobre R es congruente con una matriz diagonal de la forma $[\pm a_1^2, \ldots, \pm a_r^2, 0, \ldots, 0]$, y ésta a su vez es congruente con una de la forma $[1, \ldots, 1, -1, \ldots, -1, 0, \ldots, 0]$ y estas matrices sí que son canónicas, es decir, dos de ellas no pueden ser congruentes por el teorema anterior. Así pues:

Teorema 6.52 *Dos matrices simétricas sobre un cuerpo euclídeo son congruentes si y sólo si tienen el mismo rango y la misma signatura.*

La obtención de formas canónicas en cuerpos como $\mathbb{Q}$ requiere resultados importantes de teoría de números.

Con el teorema anterior tenemos resuelto el problema teórico que nos habíamos planteado, pero conviene que nos detengamos a dar un método para calcular en la práctica la signatura de una matriz simétrica sobre un cuerpo realmente cerrado, pues la forma de hacerlo es muy simple. Vamos a probar que todo se reduce a calcular los valores propios de la matriz. Nos basamos en el teorema siguiente:

Teorema 6.53 *Sea $h : V \longrightarrow V$ un endomorfismo en un espacio vectorial euclídeo sobre un cuerpo realmente cerrado. Si la matriz de h en una base ortonormal es simétrica, entonces V tiene una base ortonormal formada por vectores propios de h.*

Demostración: Sea R el cuerpo de escalares y sea $C = R[i]$. La hipótesis sobre R equivale a que C es algebraicamente cerrado. Sea A la matriz de h a la que alude el enunciado. En primer lugar probaremos que su polinomio característico tiene todas sus raíces en R. Sea $\lambda \in C$ una de sus raíces. Entonces $zA = \lambda z$ para un cierto $z \in C^n$ no nulo. Conjugando, $\bar{z}A = \bar{\lambda}\bar{z}$, luego $A\bar{z}^t = \bar{\lambda}\bar{z}^t$. Multiplicando por z queda $\bar{\lambda}z\bar{z}^t = zA\bar{z}^t = \lambda z\bar{z}^t$. Es claro que $z\bar{z}^t \neq 0$, luego $\lambda = \bar{\lambda}$.

Sean $\lambda_1, \ldots, \lambda_r$ las raíces distintas del polinomio característico de A, es decir, los valores propios de A. Para cada i, sea V_i el espacio de vectores propios asociado. Veamos que si $i \neq j$ entonces $V_i \perp V_j$.

Sean x e y las coordenadas de dos vectores en V_i y V_j respectivamente. Entonces $xA = \lambda_i x$, $yA = \lambda_j y$, luego $Ay^t = \lambda_j y^t$. Multiplicando obtenemos $\lambda_i xy^t = xAy^t = \lambda_j xy^t$. Puesto que $\lambda_i \neq \lambda_j$ ha de ser $xy^t = 0$, lo que prueba que los vectores correspondientes son ortogonales.

Sea $W = V_1 \perp \cdots \perp V_r$. Veamos que $h[W^\perp] \subset W^\perp$. En efecto, sea $v \in W^\perp$. Sean x las coordenadas de v e y las coordenadas de un vector de V_i. Entonces $xy^t = 0$ y $yA = \lambda_i y$, luego $xAy^t = \lambda_i xy^t = 0$. Como xA son las coordenadas de $h(v)$, concluimos que $h(v)$ es ortogonal a cada V_i, luego a W.

Esto prueba que h puede restringirse a un endomorfismo de $W^\perp$. Es fácil ver que la matriz de h en cualquier base ortonormal es simétrica (basta probar que toda matriz de la forma MAM^t lo es). Si $W^\perp \neq 0$, la matriz de $h|_{W^\perp}$ en una base ortonormal será simétrica (basta completarla hasta una base ortonormal de V y usar que la matriz de h en esta base lo es). Aplicando lo ya probado llegamos a que h tiene un vector propio en $W^\perp$, pero esto es absurdo, pues todos los vectores propios de h están en W. Así pues, $W^\perp = 0$ y $V = W \perp W^\perp = W$.

Ahora basta tomar una base ortonormal en cada V_i y unirlas todas. ∎

Equivalentemente, hemos probado que toda matriz simétrica es semejante a una matriz diagonal[2]. En términos de los invariantes de una matriz, lo que hemos probado es que las matrices simétricas sobre un cuerpo realmente cerrado tienen todos sus divisores elementales de grado 1, por lo que la dimensión del espacio fundamental de un valor propio es su multiplicidad en el polinomio característico. Ahora probamos el resultado que nos interesa:

Teorema 6.54 *Sea F una forma bilineal simétrica en un espacio vectorial euclídeo sobre un cuerpo realmente cerrado. Entonces existe una base ortonormal de V que es ortogonal para F.*

Demostración: Sea A la matriz de F en una base ortonormal de V, que será simétrica. Sea h el endomorfismo de V que en dicha base tiene matriz A. Por el teorema anterior V tiene una base $e_1, \ldots, e_n$ formada por vectores propios de h, es decir, si llamamos $x_1, \ldots, x_n$ a sus coordenadas en la base original, $x_i A = \lambda_i x_i$, luego si $i \neq j$ se cumple $F(e_i, e_j) = x_i A x_j^t = \lambda_i x_i x_j^t = 0$, luego la base es ortogonal para F. ∎

Así pues, dada una matriz simétrica A sobre un cuerpo realmente cerrado, podemos considerar un espacio euclídeo y en él una forma bilineal que tenga matriz A en una base ortonormal. Si B es la matriz de F en la base dada por el teorema anterior, tenemos que B es diagonal y $B = MAM^t$, donde M es la matriz del cambio de base entre dos bases ortonormales, luego es ortogonal,

[2]Más aún, que la matriz de semejanza M puede tomarse ortogonal, es decir, de manera que $MM^t = I$, pues esta propiedad caracteriza a las matrices que transforman una base ortonormal en otra base ortonormal.

luego $M^t = M^{-1}$. Esto significa que las matrices A y B son semejantes, luego los coeficientes de la diagonal de B son los valores propios de A. En resumen:

Teorema 6.55 *Dada una matriz simétrica A sobre un cuerpo realmente cerrado, la matriz diagonal B formada con los valores propios de A (repetidos según su multiplicidad en el polinomio característico) es congruente*[3] *con A. Por lo tanto, la signatura de A es el número de valores propios positivos.*

6.6 Aplicaciones

Las aplicaciones más relevantes de los resultados que hemos presentado en este capítulo (aparte del concepto de determinante, que aparece en los contextos más diversos) se dan en la geometría. No obstante, vamos a presentar aquí algunas consecuencias algebraicas.

La unicidad de la clausura real Veamos en primer lugar que cada cuerpo ordenado tiene, salvo isomorfismo, una única clausura real. Para ello fijamos un cuerpo ordenado K y una clausura real R, de modo que $C = R[i]$ es una clausura algebraica de K. Sea $f(x) \in K[x]$ un polinomio mónico irreducible y sean $\alpha_1, \ldots, \alpha_n$ sus raíces en C. Definimos

$$\sigma_j = \sum_{t=1}^{n} \alpha_t^j.$$

Como los automorfismos de $K(\alpha_1, \ldots, \alpha_n)/K$ permutan los α_j, es claro que fijan a cada σ_j, luego $\sigma_j \in K$. Llamamos q_f a la forma cuadrática en K^n que en la base canónica tiene por matriz $(\sigma_{i+j-2})_{i,j}$. Si esta forma tiene rango r y signatura s, llamamos $c_f = 2s - r$, que es el número de términos positivos menos el número de términos negativos en cualquier matriz diagonal congruente con $(\sigma_{i+j-2})_{i,j}$.

Teorema 6.56 *En las condiciones anteriores, c_f es el número de raíces de f que pertenecen a R.*

Demostración: Como $|C : R| = 2$, es claro que todo polinomio irreducible en $R[x]$ tiene grado 1 o 2. Por lo tanto f se descompone en $R[x]$ como producto de factores

$$f(x) = (x - \beta_1) \cdots (x - \beta_m) q_1(x) \cdots q_l(x),$$

donde los polinomios $q_j(x)$ tienen grado 2, y lo que queremos probar es que $c_f = m$. Cada $q_j(x)$ tiene a su vez dos raíces conjugadas en $C[x]$, de modo que $q_j(x) = (x - \gamma_j)(x - \bar{\gamma}_j)$, con $\gamma_j \in C$. Ahora observamos que si $x \in K^n$, entonces

$$q_f(x) = \sum_{u,v=1}^{n} \sigma_{u+v-2} x_u x_v = \sum_{j,u,v=1}^{n} \alpha_j^{u+v-2} x_u x_v$$

[3] Y la matriz de congruencia puede tomarse ortogonal.

$$= \sum_{j=1}^{n} \left(\sum_{u=1}^{n} \alpha_j^{u-1} x_u \right) \left(\sum_{v=1}^{n} \alpha_j^{v-1} x_v \right) = \sum_{j=1}^{n} \left(\sum_{u=1}^{n} \alpha_j^{u-1} x_u \right)^2$$

$$= \sum_{j=1}^{m} \left(\sum_{u=1}^{n} \beta_j^{u-1} x_u \right)^2 + \sum_{j=1}^{l} \left(\left(\sum_{u=1}^{n} \gamma_j^{u-1} x_u \right)^2 + \left(\sum_{u=1}^{n} \bar{\gamma}_j^{u-1} x_u \right)^2 \right)$$

$$= \sum_{j=1}^{m} \left(\sum_{u=1}^{n} \beta_j^{u-1} x_u \right)^2 + \sum_{j=1}^{l} 2 \left(\left(\sum_{u=1}^{n} \frac{\gamma_j^{u-1} + \bar{\gamma}_j^{u-1}}{2} x_u \right)^2 + \left(\sum_{u=1}^{n} \frac{\gamma_j^{u-1} - \bar{\gamma}_j^{u-1}}{2} x_u \right)^2 \right)$$

Ahora consideramos la matriz $M =$

$$\begin{pmatrix} \beta_1^0 & \cdots & \beta_m^0 & \frac{\gamma_1^0 + \bar{\gamma}_1^0}{2} & \frac{\gamma_1^0 - \bar{\gamma}_1^0}{2i} & \cdots & \frac{\gamma_l^0 + \bar{\gamma}_l^0}{2} & \frac{\gamma_l^0 - \bar{\gamma}_l^0}{2i} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \beta_1^{n-1} & \cdots & \beta_m^{n-1} & \frac{\gamma_1^{n-1} + \bar{\gamma}_1^{n-1}}{2} & \frac{\gamma_1^{n-1} - \bar{\gamma}_1^{n-1}}{2i} & \cdots & \frac{\gamma_1^{n-1} + \bar{\gamma}_1^{n-1}}{2} & \frac{\gamma_l^{n-1} - \bar{\gamma}_l^{n-1}}{2i} \end{pmatrix}.$$

Enseguida probamos que es regular. Admitiéndolo de momento, vemos que determina un cambio de base respecto de la base canónica de modo que, en la nueva base, las coordenadas de x son

$$y_j = \sum_{u=1}^{n} \beta_j^{u-1} x_u, \qquad j = 1, \ldots, m,$$

$$y_{m+2j-1} = \sum_{u=1}^{n} \frac{\gamma_j^{u-1} + \bar{\gamma}_j^{u-1}}{2} x_u, \qquad j = 1, \ldots, l,$$

$$y_{m+2j} = \sum_{u=1}^{n} \frac{\gamma_j^{u-1} - \bar{\gamma}_j^{u-1}}{2i} x_u, \qquad j = 1, \ldots, l,$$

por lo que la expresión coordenada de q_f en la nueva base es

$$q_f(y) = \sum_{j=1}^{m} y_j^2 + \sum_{j=1}^{l} 2(y_{m+2j-1}^2 - y_{m+2j}^2),$$

luego la matriz de q_f en la nueva base es diagonal, de la forma

$$[\overbrace{1, \ldots, 1}^{m}, \overbrace{2, \ldots, 2}^{l}, \overbrace{-2, \ldots, -2}^{l}],$$

luego $c_f = m$, como queríamos probar. Sólo queda pendiente justificar que, realmente, M es regular. Para ello consideramos la matriz

$$N = \begin{pmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & 1 & 1 & & & \\ & & & i & -i & & & \\ & & & & & \ddots & & \\ & & & & & & 1 & 1 \\ & & & & & & i & -i \end{pmatrix}$$

y observamos que

$$MN = \begin{pmatrix} \beta_1^0 & \cdots & \beta_m^0 & \gamma_1^0 & \bar{\gamma}_1^0 & \cdots & \gamma_l^0 & \bar{\gamma}_l^0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \beta_1^{n-1} & \cdots & \beta_m^{n-1} & \gamma_1^{n-1} & \bar{\gamma}_1^{n-1} & \cdots & \gamma_l^{n-1} & \bar{\gamma}_l^{n-1} \end{pmatrix},$$

luego $|MN|$ es un determinante de Vandermonde, que es no nulo porque las raíces de f son distintas entre sí. Por lo tanto, también $|M| \neq 0$. ∎

Así pues, el número de raíces reales de un polinomio f está determinado por el propio f, con independencia del cuerpo que tomemos como R. Con esto podemos probar:

Teorema 6.57 *Sea $\phi : k_1 \longrightarrow k_2$ un isomorfismo de cuerpos ordenados y sean R_1, R_2 clausuras reales de k_1 y k_2 respectivamente. Sea $k_1 \subset K \subset R_1$ tal que K/k_1 es una extensión finita. Entonces ϕ se extiende a un monomorfismo $K \longrightarrow R_2$.*

Demostración: Sea $K = k_1(\alpha)$ y sea $f_1(x) \in k_1[x]$ el polinomio mínimo de α. Sea $f_2(x) \in k_2[x]$ la imagen de f_1 por ϕ. Si $\alpha_1, \ldots, \alpha_n$ son las raíces de f_1 en $R_1[i]$, entonces ϕ se extiende a un monomorfismo $\bar{\phi} : k_1(\alpha_1, \ldots, \alpha_n) \longrightarrow R_2[i]$, de modo que $\bar{\phi}(\alpha_1), \ldots, \bar{\phi}(\alpha_n)$ son las raíces de f_2 en $R_2[i]$, y es claro que las imágenes por ϕ de los σ_j definidos antes del teorema anterior para f_1 son los correspondientes a f_2. De aquí se sigue a su vez que q_{f_1} y q_{f_2} tienen el mismo rango y la misma signatura, por lo que $c_{f_1} = c_{f_2}$.

Pero sabemos que $c_{f_1} > 0$, pues α es una raíz de f_1 en R_1, luego f_2 tiene una raíz en R_2, digamos α'. Por consiguiente ϕ se extiende a un isomorfismo $\bar{\phi} : K \longrightarrow k_2(\alpha') \subset R_2$. ∎

Por último:

Teorema 6.58 (Artin-Schreier) (AE) *Dos clausuras reales de un mismo cuerpo ordenado k son k-isomorfas.*

Demostración: Sean R_1/k y R_2/k dos clausuras reales de k. Basta aplicar el lema de Zorn al conjunto de todos los k-isomorfismos $\phi : k_1 \longrightarrow k_2$, donde $k \subset k_1 \subset R_1$, $k \subset k_2 \subset R_2$, ordenado por la inclusión. Si $\phi : k_1 \longrightarrow k_2$ es maximal, es decir, no se puede extender a cuerpos mayores, tiene que ser $k_1 = R_1$, pues en caso contrario podríamos tomar $\alpha \in R_1 \setminus k_1$ y $K = k_1(\alpha)$, y el teorema anterior nos daría una extensión. Similarmente, tiene que ser $k_2 = R_2$, o podríamos aplicar el teorema anterior a ϕ^{-1}. Por lo tanto $\phi : R_1 \longrightarrow R_2$ es un k-isomorfismo de cuerpos (y de hecho de cuerpos ordenados, porque al transformar cuadrados en cuadrados conserva el orden). ∎

El teorema de Frobenius Para nuestra segunda aplicación tenemos que introducir el concepto de álgebra:

Definición 6.59 Si k es un cuerpo, una *k-álgebra* es una cuádrupla $(A, +, \circ, \cdot)$ tal que $(A, +, \circ)$ sea un anillo unitario, $(A, +, \cdot)$ sea un k-espacio vectorial y además se cumpla la siguiente relación de compatibilidad entre ambas estructuras:

$$\alpha(a \circ b) = (\alpha a) \circ b = a \circ (\alpha b), \quad \text{para } \alpha \in k, a, b \in A.$$

En la práctica usaremos la misma notación $\cdot$ para los dos productos de un álgebra. La aplicación $k \longrightarrow A$ dada por $\alpha \mapsto \alpha \cdot 1$ es claramente un monomorfismo de anillos, por lo que podemos suponer que $k \subset A$, de modo que el producto como espacio vectorial es simplemente la restricción del producto como anillo.

Por ejemplo, el cuerpo $\mathbb{C}$ de los números complejos y el anillo de división $\mathbb{H}$ de los cuaternios son claramente $\mathbb{R}$-álgebras. La primera es conmutativa y la segunda no, pero es un álgebra de división, en el sentido de que, como anillo, es un anillo de división.

Teorema 6.60 (Frobenius) *Las únicas $\mathbb{R}$-álgebras de división de dimensión finita sobre $\mathbb{R}$ son $\mathbb{R}$, $\mathbb{C}$ y $\mathbb{H}$.*

Demostración: Sea D un álgebra de división de dimensión m sobre $\mathbb{R}$. Dado $a \in D$, la aplicación $\phi_a : D \longrightarrow D$ dada por $\phi_a(d) = ad$ es un endomorfismo de D como espacio vectorial sobre $\mathbb{R}$. Definimos

$$V = \{a \in D \mid \mathrm{Tr}(\phi_a) = 0\},$$

que claramente es un subespacio vectorial de D de codimensión 1.

Vamos a probar que $V = \{a \in D \mid a^2 \in \mathbb{R},\ a^2 \leq 0\}$.

Dado $a \in D$, sea $p(x)$ el polinomio característico de ϕ_a, que podemos factorizar como

$$p(x) = (x - t_1) \cdots (x - t_r)(x - z_1)(x - \bar{z}_1) \cdots (x - z_s)(x - \bar{z}_s),$$

donde $t_1, \ldots t_r \in \mathbb{R}$ y $z_1, \ldots, z_s \in \mathbb{C} \setminus \mathbb{R}$. Se cumple que $m = r + 2s$. Los factores irreducibles de $p(x)$ en $\mathbb{R}[x]$ son los polinomios $x - t_i$ y los productos $x^2 - 2x\mathrm{Re}\, z_i + |z_i|^2$.

El teorema de Cayley afirma que $p(\phi_a) = 0$ o, lo que es lo mismo, que $p(a)d = 0$ para todo $d \in D$. Como D es un anillo de división, esto implica que $p(a) = 0$ y, esto a su vez nos lleva a que $a = t_i$ o bien $a^2 - 2a\mathrm{Re}\, z_i + |z_i|^2 = 0$.

En el primer caso $a \in \mathbb{R}$ y $\mathrm{Tr}(\phi_a) = a$, luego $a \in V$ si y sólo si $a = 0$ si y sólo si $a^2 \leq 0$.

En el segundo caso $a \notin \mathbb{R}$ y el polinomio mínimo de ϕ_a es $x^2 - 2x\mathrm{Re}\, z_i + |z_i|^2$. Como el polinomio característico y el polinomio mínimo de un endomorfismo tienen las mismas raíces, sucede que

$$p(x) = (x^2 - 2x\mathrm{Re}\, z + |z|^2)^k,$$

para cierto $z \in \mathbb{C}$. Concluimos que $a \in V$ si y sólo si $\mathrm{Re}\, z = 0$, si y sólo si $a^2 = -|z|^2$ si y sólo si $a^2 \leq 0$.

Como V tiene codimensión 1 y $V \cap \mathbb{R} = 0$, sucede que $D = \mathbb{R} \oplus V$.

Consideramos ahora la forma bilineal $B : V \times V \longrightarrow \mathbb{R}$ dada por

$$B(a,b) = -\frac{ab+ba}{2}.$$

Esto es correcto, pues $ab + ba = (a+b)^2 - a^2 - b^2 \in \mathbb{R}$. Más aún, si $a \neq 0$, entonces $B(a,a) = -a^2 > 0$. Por lo tanto, B es un producto escalar en V.

Sea $W \subset V$ un subespacio vectorial de dimensión mínima tal que W genere D como álgebra, es decir, tal que todo elemento de D sea combinación lineal de 1 y de productos finitos de elementos de W. (Existe porque el propio V cumple esta condición.) Sea $e_1, \ldots, e_n$ una base ortonormal de W respecto del producto B. Esto significa que

$$e_i^2 = -1, \qquad e_i e_j = -e_j e_i, \quad \text{si } i \neq j.$$

Si $n = 0$ entonces $D = \mathbb{R}$. Si $n = 1$ entonces $D = \langle 1, e_1 \rangle$ con $e_1^2 = -1$, de donde se sigue que $D = \mathbb{C}$.

Si $n = 2$ entonces $D = \langle 1, e_1, e_2, e_1 e_2 \rangle$, pues las relaciones $e_1^2 = e_2^2 = -1$, $e_1 e_2 = -e_2 e_1$ reducen cualquier producto de los cuatro elementos anteriores a una combinación lineal de ellos, por lo que el miembro derecho es una subálgebra de D. Es fácil deducir de aquí que $D = \mathbb{H}$.

Sólo falta probar que $n > 2$ es imposible. En tal caso, llamemos $u = e_1 e_2 e_n$, y observamos que $u^2 = e_1 e_2 e_n e_1 e_2 e_n = 1$, luego $(u+1)(u-1) = 0$, luego $u = \pm 1$, pero entonces $e_n = \pm e_1 e_2$ y $\langle e_1, \ldots e_{n-1} \rangle$ genera D como álgebra, en contra de la minimalidad de W. ∎

Así pues, si para dotar a $\mathbb{R}^4$ de estructura de anillo era preciso renunciar a que sea un cuerpo, para cualquier otra dimensión distinta de 2 o 4 hay que renunciar también a la estructura de anillo de división. Es posible definir muchas estructuras de álgebra en cualquier $\mathbb{R}^n$, pero, salvo en los casos contemplados por el teorema, habrá elementos no nulos sin inverso.

Capítulo VII

Resolución de ecuaciones por radicales

Las soluciones de una ecuación de segundo grado $ax^2+bx+c=0$ (sobre un cuerpo de característica distinta de 2) vienen dadas por la conocida fórmula

$$x = \frac{-b \pm \sqrt{b^2-4ac}}{2a},$$

entendiendo que si el discriminante $\Delta = b^2 - 4ac$ no tiene raíces cuadradas en el cuerpo, entonces la ecuación no tiene solución (si bien su solución en una clausura algebraica tendrá esta forma). En este capítulo estudiaremos la existencia de fórmulas análogas para ecuaciones de grados superiores, junto con algunos temas relacionados. No obstante, antes de intentar resolverlas, en las primeras secciones presentaremos algunos conceptos y hechos relacionados con los polinomios de una variable.

7.1 Polinomios simétricos

Los polinomios simétricos proporcionan una relación importante entre los coeficientes de un polinomio y sus raíces. Sus propiedades (conceptualmente más simples) pueden, en ocasiones, sustituir a la teoría de Galois. También representan un papel relevante en la teoría de números trascendentes (por ejemplo en la prueba de la trascendencia de π).

Definición 7.1 Sea A un dominio y $\sigma \in \Sigma_n$. Por 2.38 se cumple que la aplicación $\overline{\sigma} : A[x_1,\dots,x_n] \longrightarrow A[x_1,\dots,x_n]$ definida mediante

$$\overline{\sigma}\big(p(x_1,\dots,x_n)\big) = p(x_{\sigma(1)},\dots,x_{\sigma(n)})$$

es un homomorfismo de anillos. De hecho es claro que se trata de un isomorfismo que se extiende a un automorfismo del cuerpo $A(x_1,\dots,x_n)$ (al que seguiremos llamando $\overline{\sigma}$).

También es fácil comprobar que la aplicación $\Sigma_n \longrightarrow \mathrm{Aut}\big(A(x_1,\dots,x_n)\big)$ dada por $\sigma \mapsto \overline{\sigma}$ es un monomorfismo de grupos.

Si no hay confusión escribiremos $\sigma(p)$ en lugar de $\overline{\sigma}(p)$ cuando p sea un polinomio o una fracción algebraica en las indeterminadas $x_1,\dots,x_n$. En definitiva, $\sigma(p)$ se obtiene a partir de p intercambiando sus variables del modo indicado por σ.

Diremos que una fracción algebraica p (en particular un polinomio) es *simétrica* si $\sigma(p) = p$ para toda permutación $\sigma \in \Sigma_n$.

Notemos que la definición depende del anillo de polinomios (o el cuerpo de fracciones algebraicas) que consideremos pues, por ejemplo, $xy + xz + yz$ es simétrico como elemento de $\mathbb{Q}[x,y,z]$, pero no como elemento de $\mathbb{Q}[v,x,y,z]$, pues la trasposición (v,x) no lo deja fijo.

Del hecho de que las aplicaciones $\overline{\sigma}$ sean isomorfismos se deduce inmediatamente que el conjunto de todas las fracciones algebraicas simétricas del cuerpo $A(x_1,\dots,x_n)$ es un subcuerpo, así como que el conjunto de todos los polinomios simétricos de $A[x_1,\dots,x_n]$ es un subanillo.

Llamaremos *polinomios simétricos elementales* de $A[x_1,\dots,x_n]$ a los polinomios $e_0,\dots,e_n$ dados por

$$e_0 = 1, \qquad e_k = \sum_{1\le i_1<\cdots<i_k\le n} x_{i_1}\cdots x_{i_k} \qquad \text{para } k = 1,\dots,n.$$

Vemos, pues, que cada e_k es un polinomio simétrico de grado k. Por ejemplo, los polinomios simétricos de 3 indeterminadas son

$$1, \qquad x+y+z, \qquad xy+xz+yz, \qquad xyz.$$

En otras palabras, el polinomio e_k es la suma de todos los monomios que pueden construirse multiplicando k variables distintas.

El teorema siguiente proporciona una relación útil entre los polinomios simétricos elementales de orden n y los de orden $n-1$. La demostración es muy sencilla y queda a cargo del lector.

Teorema 7.2 *Sea A un dominio, $n > 1$, $e_0,\dots,e_n$ los polinomios simétricos elementales en $A[x_1,\dots,x_n]$ y $\bar{e}_0,\dots,\bar{e}_{n-1}$ los polinomios simétricos elementales en $A[x_1,\dots,x_{n-1}]$. Entonces:*

1. $e_n(x_1,\dots,x_n) = x_n\bar{e}_{n-1}(x_1,\dots,x_{n-1})$.
2. $e_k(x_1,\dots,x_n) = \bar{e}_k(x_1,\dots,x_{n-1}) + x_n\bar{e}_{k-1}(x_1,\dots,x_{n-1})$, *para* $1 \le k < n$.

Esto significa que los polinomios simétricos elementales de n variables pueden obtenerse a partir de los polinomios simétricos elementales de $n-1$ variables mediante sumas y productos en las que intervenga también la variable x_n que les falta. Una simple inducción permite probar la siguiente generalización de este hecho (que no nos va a ser necesaria luego):

Teorema 7.3 *Sea A un dominio, sean $e_0, \ldots, e_n$ los polinomios simétricos elementales en $A[x_1, \ldots, x_n]$ y $\bar{e}_0, \ldots, \bar{e}_k$ los polinomios simétricos elementales en $A[x_1, \ldots, x_k]$, donde $1 \leq k < n$. Entonces $e_i \in A[\bar{e}_0, \ldots, \bar{e}_k, x_{k+1}, \ldots, x_n]$ para $i = 0, \ldots, n$.*

El interés de los polinomios simétricos elementales reside principalmente en que son los que nos dan los coeficientes de un polinomio a partir de sus raíces en un cuerpo de escisión. Veámoslo:

Teorema 7.4 (Viète) *Sea A un dominio y $e_0, \ldots, e_n$ los polinomios simétricos elementales en $A[x_1, \ldots, x_n]$. Entonces*

$$(x - x_1) \cdots (x - x_n) = \sum_{k=0}^{n} (-1)^{n-k} e_{n-k}(x_1, \ldots, x_n) x^k.$$

Demostración: Por inducción sobre n. Para $n = 1$ es obvio. Supongámoslo para $n - 1$.

Sean $\bar{e}_0, \ldots, \bar{e}_{n-1}$ los polinomios simétricos elementales en $A[x_1, \ldots, x_{n-1}]$. Entonces

$$\begin{aligned}
& (x - x_1) \cdots (x - x_n) = \sum_{k=0}^{n-1} (-1)^{n-1-k}\ \bar{e}_{n-1-k} x^k (x - x_n) \\
= & \sum_{k=0}^{n-1} (-1)^{n-1-k}\ \bar{e}_{n-1-k} x^{k+1} - \sum_{k=0}^{n-1} (-1)^{n-1-k} x_n \bar{e}_{n-1-k} x^k \\
= & (-1)^n x_n \bar{e}_{n-1} + x^n + \sum_{k=0}^{n-2} (-1)^{n-1-k}\ \bar{e}_{n-1-k} x^{k+1} \\
& - \sum_{k=1}^{n-1} (-1)^{n-1-k} x_n \bar{e}_{n-1-k} x^k \\
= & (-1)^n x_n \bar{e}_{n-1} + x^n + \sum_{k=1}^{n-1} (-1)^{n-k} (\bar{e}_{n-k} + x_n \bar{e}_{n-1-k}) x^k \\
\text{(por 7.2)} \quad = & (-1)^n e_n + x^n + \sum_{k=1}^{n-1} (-1)^{n-k} e_{n-k} x^k \\
= & \sum_{k=0}^{n} (-1)^{n-k} e_{n-k} x^k.
\end{aligned}$$

■

Por ejemplo,

$$(x - a)(x - b)(x - c) = x^3 - (a + b + c)x^2 + (ab + ac + bc)x - abc,$$

lo que en algunos casos puede evitarnos muchas operaciones. Veamos ahora otro de los hechos clave sobre polinomios simétricos:

Teorema 7.5 *Sea A un dominio y sean $e_1, \ldots, e_n$ los polinomios simétricos elementales en $A[x_1, \ldots, x_n]$. Entonces el anillo de los polinomios simétricos de $A[x_1, \ldots, x_n]$ es $A[e_1, \ldots, e_n]$.*

DEMOSTRACIÓN: Por inducción sobre n. Para $n = 1$ resulta que todo polinomio de $A[x]$ es simétrico, y como $e_1 = x$, se cumple el teorema.

Supongamos el teorema para $n-1$. Sea $p(x_1, \ldots, x_n)$ un polinomio simétrico del anillo $A[x_1, \ldots, x_n]$. Veamos que $p(x_1, \ldots, x_n) \in A[e_1, \ldots, e_n]$ por inducción sobre el grado de p. En caso de que p sea de grado 0 es evidente. Supongamos que todo polinomio simétrico de grado menor que m está en $A[e_1, \ldots, e_n]$ y que p tiene grado m.

Sea $\bar{p}(x_1, \ldots, x_{n-1}) = p(x_1, \ldots, x_{n-1}, 0) \in A[x_1, \ldots, x_{n-1}]$. Claramente $\bar{p}$ es simétrico. Por la primera hipótesis de inducción $\bar{p} \in A[\bar{e}_1, \ldots, \bar{e}_{n-1}]$, donde $\bar{e}_1, \ldots, \bar{e}_{n-1}$ son los polinomios simétricos elementales de $A[x_1, \ldots, x_{n-1}]$. Esto significa que existe un polinomio $g \in A[x_1, \ldots, x_{n-1}]$ tal que $\bar{p} = g(\bar{e}_1, \ldots, \bar{e}_{n-1})$.

Sea $h(x_1, \ldots, x_n) = p(x_1, \ldots, x_n) - g(e_1, \ldots, e_{n-1})$, simétrico. Por 7.2 se cumple que $e_i(x_1, \ldots, x_{n-1}, 0) = \bar{e}_i(x_1, \ldots, x_{n-1})$ para $i = 1, \ldots, n-1$, luego $h(x_1, \ldots, x_{n-1}, 0) = \bar{p}(x_1, \ldots, x_{n-1}) - g(\bar{e}_1, \ldots, \bar{e}_{n-1}) = 0$.

Así pues, x_n divide a $h(x_1, \ldots, x_n)$ y por simetría todas las variables lo dividen, y su producto también, o sea, e_n divide a h. Sea $h = e_n\bar{h}$.

Si $\sigma \in \Sigma_n$, tenemos que $e_n\bar{h} = h = \sigma(h) = \sigma(e_n)\sigma(\bar{h}) = e_n\sigma(\bar{h})$, luego $\sigma(\bar{h}) = \bar{h}$. Esto prueba que $\bar{h}$ es simétrico y de grado menor que m, luego por la segunda hipótesis de inducción $\bar{h} \in A[e_1, \ldots, e_n]$, con lo que también $p = g(e_1, \ldots, e_{n-1}) + e_n\bar{h} \in A[e_1, \ldots, e_n]$. ∎

Ejercicio: Probar que si A es un dominio, $\alpha_1, \ldots, \alpha_n$ son todas las raíces de un polinomio mónico de $A[x]$ en una extensión en la cual se escinda (repetidas según su multiplicidad), y $p(x_1, \ldots, x_n)$ es un polinomio simétrico, entonces $p(\alpha_1, \ldots, \alpha_n) \in A$.

No es inmediato, ni siquiera a partir del teorema anterior, que el subcuerpo de las fracciones algebraicas simétricas de un cuerpo $k(x_1, \ldots, x_n)$ sea precisamente el cuerpo $k(e_1, \ldots, e_n)$. Esto es tanto como afirmar que toda fracción algebraica simétrica se expresa como cociente de dos polinomios simétricos. Daremos una prueba basada en la teoría de Galois.

Teorema 7.6 *Sea k un cuerpo y $n \geq 1$. Entonces el cuerpo de las fracciones algebraicas simétricas de $k(x_1, \ldots, x_n)$ es $k(e_1, \ldots, e_n)$. Además la extensión $k(x_1, \ldots, x_n)/k(e_1, \ldots, e_n)$ es finita de Galois y su grupo de Galois es Σ_n.*

DEMOSTRACIÓN: Consideremos el polinomio $p(x) = (x - x_1) \cdots (x - x_n)$. El teorema 7.4 afirma que $p(x) \in k(e_1, \ldots, e_n)[x]$, lo que implica que las indeterminadas $x_1, \ldots, x_n$ son algebraicas sobre $k(e_1, \ldots, e_n)$. Más aún, son separables (porque son raíces simples de $p(x)$) y $k(x_1, \ldots, x_n)$ es el cuerpo de escisión sobre $k(e_1, \ldots, e_n)$ de $p(x)$, luego la extensión es finita de Galois.

Según vimos en el capítulo V, el grupo de Galois

$$G = G\big(k(x_1, \ldots, x_n)/k(e_1, \ldots, e_n)\big)$$

puede identificarse con un subgrupo del grupo de las permutaciones de las raíces de $p(x)$, es decir, de Σ_n, por lo que $|G| \leq n!$

Por otro lado tenemos un monomorfismo $\Sigma_n \longrightarrow \mathrm{Aut}\big(k(x_1,\dots,x_n)\big)$ y las imágenes de los elementos de Σ_n, por definición de simetría, fijan a los elementos de $k(e_1,\dots,e_n)$, es decir, de hecho tenemos un monomorfismo $\Sigma_n \longrightarrow G$. Teniendo en cuenta los órdenes, este monomorfismo es de hecho un isomorfismo, es decir, los automorfismos de la extensión son exactamente las permutaciones de las indeterminadas.

Ahora bien, si llamamos E al cuerpo de las fracciones algebraicas simétricas de $k(x_1,\dots,x_n)$, tenemos las inclusiones

$$k \subset k(e_1,\dots,e_n) \subset E \subset k(x_1,\dots,x_n),$$

y por definición de simetría es obvio que $G(k(x_1,\dots,x_n)/E) = G$, luego el teorema de Galois implica que $E = k(e_1,\dots,e_n)$. ∎

7.2 La resultante de dos polinomios

Ya sabemos que el valor $\Delta = b^2 - 4ac$ recibe el nombre de discriminante de un polinomio de segundo grado ax^2+bx+c. Aquí vamos a presentar un concepto general de discriminante de un polinomio arbitrario. De hecho estudiaremos un concepto más general aún:

Definición 7.7 Consideremos un anillo conmutativo y unitario A y dos polinomios

$$f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n, \quad g(x) = b_0x^m + b_1x^{m-1} + \cdots + b_m,$$

donde $a_0b_0 \neq 0$. Definimos la *resultante* de $f(x)$ y $g(x)$ como el determinante

$$R(f,g) = \begin{vmatrix}
a_0 & a_1 & \cdots & \cdots & a_n & & & \\
 & a_0 & a_1 & \cdots & \cdots & a_n & & \\
 & & \ddots & \ddots & & & \ddots & \\
 & & & a_0 & a_1 & \cdots & \cdots & a_n \\
b_0 & b_1 & \cdots & \cdots & b_m & & & \\
 & b_0 & b_1 & \cdots & \cdots & b_m & & \\
 & & \ddots & & & & \ddots & \\
 & & & b_0 & b_1 & \cdots & \cdots & b_m
\end{vmatrix}$$

donde los coeficientes de f aparecen en m filas y los de g en n filas (y donde se entiende que los huecos corresponden a coeficientes nulos).

En particular, si en el anillo de polinomios $A = \mathbb{Z}[u_0,\dots,u_n,v_0,\dots,v_m]$ consideramos los polinomios

$$f(x) = u_0x^n + u_1x^{n-1} + \cdots + u_n, \quad g(x) = v_0x^m + v_1x^{m-1} + \cdots + v_m,$$

tenemos la *resultante general* $R \in \mathbb{Z}[u_0,\dots,u_n,v_0,\dots,v_m]$ de grados n y m, que es un polinomio que se particuliza a la resultante de cualquier par de polinomios f y g de grados n y m en cualquier anillo A, cuando sus variables se sustituyen por los coeficientes de f y g.

De la propia definición de determinante se sigue inmediatamente que la resultante general $R(u_0,\dots,u_n,v_0,\dots,v_m)$ es un polinomio cuyos monomios tienen todos m variables u_i y n variables v_i. El monomio $u_0^m v_m^n$ (la diagonal del determinante) aparece con coeficiente 1. En particular tenemos que $R\neq 0$.

Tenemos las relaciones siguientes en $\mathbb{Z}[u_0,\dots,u_n,v_0,\dots,v_m,x]$:

$$\begin{array}{rcccccc}
a_0x^{n+m-1}+a_1x^{n+m-2}+ & \cdots & & & & & = x^{m-1}f(x)\\
a_0x^{n+m-2}+ & \cdots & \cdots & +a_nx^{m-2} & & & = x^{m-2}f(x)\\
& \ddots & & & \ddots & & \\
& a_0x^n+a_1x^{n-1} & \cdots & \cdots & a_n & = & f(x)\\
b_0x^{n+m-1}+b_1x^{n+m-2}+ & \cdots +b_mx^{n-1} & & & & & = x^{n-1}g(x)\\
b_0x^{n+m-2}+ & \cdots & \cdots & & & & = x^{n-2}g(x)\\
& & & \ddots & & & \\
& b_0x^n+b_1x^{n-1} & \cdots & \cdots & b_m & = & g(x)
\end{array}$$

Esto se interpreta como que el sistema de ecuaciones lineales que tiene por matriz la que define a R y con términos independientes los de los miembros derechos, tiene por solución $(x^{n+m-1},x^{n+m-2},\dots,x,1)$. Como $R\neq 0$, podemos resolver este sistema por la regla de Cramer y, concretamente, la última componente (igual a 1) se obtiene como cociente de dos determinantes, uno de ellos es R y el otro es el determinante que resulta de sustituir la última columna de la matriz que define a R por el vector de términos independientes. Concluimos que R es igual a este último determinante y, como los términos independientes son múltiplos de $f(x)$ o de $g(x)$, llegamos a que

$$R = Ff + Gg,$$

para ciertos polinomios F, $G\in\mathbb{Z}[u_0,\dots,u_n,v_0,\dots,v_m,x]$. Sustituyendo las indeterminadas u_i, v_i por elementos de un anillo A, obtenemos que, para todo par de polinomios $f(x)$, $g(x)\in A[x]$, existen polinomios $F(x)$, $G(x)\in A[x]$ tales que

$$R(f,g) = F(x)f(x) + G(x)g(x).$$

En particular, vemos que si $f(x)$ y $g(x)$ tienen una raíz común (en A o en cualquier extensión), entonces $R(f,g)=0$. El resultado principal que queremos probar es el recíproco.

Para ello consideramos el anillo $B=\mathbb{Z}[u_0,v_0,s_1,\dots,s_n,t_1,\dots,t_m]$, y en $B[x]$ los polinomios

$$F(x)=u_0(x-s_1)\cdots(x-s_n),\quad G(x)=v_0(x-t_1)\cdots(x-t_m).$$

Llamemos $u_i,v_i\in B$ a los coeficientes de F y G respectivamente, y vamos a calcular $R(F,G)$. Observemos que $u_i=u_0u_i'$, donde u_i' es el coeficiente i-ésimo de F/u_0, e igualmente $v_i=v_0v_i'$. Como cada monomio de R contiene m variables u_i y n variables v_i, es claro que

$$R(F,G) = u_0^m v_0^n h,$$

para cierto $h \in C = \mathbb{Z}[s_1, \ldots, s_n, t_1, \ldots, t_m]$. Si en R sustituimos un t_j por un s_i, obtenemos la resultante de dos polinomios con una raíz en común, luego se anula (y h también). Si vemos a h como polinomio en

$$\mathbb{Z}[s_1, \ldots, s_n, t_1, \ldots \hat{t_i}, \ldots t_m][t_i]$$

(donde el circunflejo significa que quitamos t_i), tenemos que s_i es raíz de h, luego h es divisible entre $s_i - t_j$, para todo i y todo j. Como estos polinomios son primos entre sí dos a dos, concluimos que

$$R(F, G) = u_0^m v_0^n \prod_{i,j} (s_i - t_j) h',$$

para cierto $h' \in C$.

Ahora bien, cada u_i (para $i \geq 1$) es un polinomio de grado $\leq n$ en $s_1, \ldots, s_n$ y cada v_j (para $j \geq 1$) es un polinomio de grado $\leq m$ en $t_1, \ldots, t_m$. Por otra parte, en cada monomio de $R(F, G)$ aparecen m variables u_i y n variables v_i, luego $R(f, g)$ tiene grado $\leq mn$ en cada una de las variables s_i, t_i. Como el grado en el producto de la derecha exactamente mn, concluimos que $h' \in \mathbb{Z}[u_0, v_0]$. Ahora bien, podemos expresar

$$R(F, G) = u_0^m h' \prod_{i=1}^{n} G(s_i),$$

de donde se sigue que el miembro derecho (sin contar h') contiene el monomio $u_0^m v_m^n = (-1)^m u_0^m t_1^n \cdots t_m^n$, al igual que el miembro izquierdo, luego ha de ser $h' = 1$. En definitiva, hemos probado que

$$R(F, G) = u_0^m v_0^n \prod_{i,j} (s_i - t_j) = u_0^m \prod_{i=1}^{n} G(s_i).$$

Consideremos ahora un cuerpo k y dos polinomios f, g no constantes que se escindan en k, es decir, que se descompongan como

$$f = a_0(x - \alpha_1) \cdots (x - \alpha_n), \quad g = b_0(x - \beta_1) \cdots (x - \beta_m).$$

El homomorfismo $\mathbb{Z}[u_0, v_0, s_1, \ldots, s_n, t_1, \ldots, t_m] \longrightarrow k$ dado por

$$u_0 \mapsto a_0, \ v_0 \mapsto b_0, \ s_i \mapsto \alpha_i, \ t_i \mapsto \beta_i$$

transforma los u_i y los v_i en los coeficientes de f y g, luego transforma $R(F, G)$ en $R(f, g)$, lo que nos da la relación

$$R(f, g) = a_0^m b_0^n \prod_{i,j} (\alpha_i - \beta_j) = a_0^m \prod_{i=1}^{n} g(\alpha_i).$$

Ahora es inmediato el teorema siguiente:

Teorema 7.8 *Sean f y g dos polinomios no constantes con coeficientes en un cuerpo k que se escindan en $k[x]$. Entonces $R(f, g) = 0$ si y sólo si f y g tienen una raíz en común.*

Un caso de particular interés se da cuando g es la derivada de f, pues una raíz común entre f y f' es una raíz múltiple de f. Si

$$f(x) = a_0 \prod_{i=1}^{n} (x - \alpha_i),$$

entonces

$$f'(x) = a_0 \sum_{i=1}^{n} \prod_{j \neq i} (x - \alpha_j).$$

y, según hemos visto,

$$\begin{aligned} R(f, f') &= a_0^{n-1} \prod_{i=1}^{n} f'(\alpha_i) = a_0^{2n-1} \prod_{i \neq j} (\alpha_i - \alpha_j) \\ &= (-1)^{n(n-1)/2} a_0^{2n-1} \prod_{i<j} (\alpha_i - \alpha_j)^2 \end{aligned}$$

Notemos que el coeficiente director de f' es na_0, luego la definición de resultante muestra que $R(f, f')/a_0$ depende polinómicamente —con coeficientes enteros— de los coeficientes de f (porque podemos sacar a_0 de la primera columna del determinante).

Definición 7.9 Si $f(x)$ es un polinomio de grado n con coeficientes en un cuerpo k, definimos su *discriminante* como

$$\Delta(f) = (-1)^{n(n-1)/2} R(f, f')/a_0 = a_0^{2n-2} \prod_{i<j} (\alpha_i - \alpha_j)^2 \in k,$$

donde $\alpha_1, \ldots, \alpha_n$ son las raíces de $f(x)$ en un cuerpo de escisión de k (repetidas según su multiplicidad).

Hemos probado que $\Delta(f)$ depende polinómicamente —con coeficientes enteros— de los coeficientes de f. El teorema siguiente es inmediato:

Teorema 7.10 *Un polinomio $f(x)$ con coeficientes en un cuerpo tiene una raíz múltiple (en una extensión algebraica) si y sólo si su discriminante es cero.*

Ejemplo Si $f(x) = ax^2 + bx + c$, entonces su discriminante es

$$\Delta(f) = -\frac{1}{a} \begin{vmatrix} a & b & c \\ 2a & b & 0 \\ 0 & 2a & b \end{vmatrix} = b^2 - 4ac.$$

■

Ejemplo Cálculos rutinarios muestran que el discriminante de un polinomio cúbico

$$f(x) = ax^3 + bx^2 + cx + d$$

es

$$\Delta(f) = b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd.$$

■

7.3 El grupo de Galois de un polinomio

Vamos a introducir algunos resultados que nos proporcionarán ejemplos para ilustrar el contenido de las secciones siguientes. El concepto básico es el siguiente:

Definición 7.11 Sea k un cuerpo y $f(x) \in k[x]$. Se llama *grupo de Galois* de $f(x)$ al grupo $G(K/k)$, donde K es cualquier cuerpo de escisión de $f(x)$ sobre k.

Sabemos que dos cuerpos de escisión de un mismo polinomio son k-isomorfos, luego el grupo de Galois de un polinomio está definido salvo isomorfismo.

Si $f(x)$ es irreducible de grado n y tiene n raíces distintas en K (esto sucede en particular si k es perfecto), llamando $A = \{\alpha_1, \ldots, \alpha_n\}$ al conjunto de todas ellas tenemos un monomorfismo de grupos $G(K/k) \longrightarrow \Sigma_A \cong \Sigma_n$ dado por la restricción a A. Por lo tanto, el grupo de Galois de $f(x)$ se identifica con un subgrupo del grupo de permutaciones de sus raíces. En particular su orden divide a $n!$

Por otra parte, como K tiene que contener a los cuerpos $k(\alpha_i)$, el orden del grupo de Galois tiene que ser múltiplo de n.

Es evidente que si $f(x)$ tiene grado 2 su grupo de Galois es Σ_2, es decir, C_2.

Observemos ahora que si Δ es el discriminante de f, por la relación entre el discriminante y las raíces, se cumple que $\sqrt{\Delta} \in K$. Más concretamente, se cumple que

$$\sqrt{\Delta} = a_0^{n-1} \prod_{i<j} (\alpha_i - \alpha_j),$$

donde la ordenación de las raíces determina si estamos considerando una de las dos raíces cuadradas de Δ o bien su opuesta. También es claro que si un automorfismo $\sigma \in G(K/k)$ se restringe a una trasposición de dos raíces, entonces $\sigma(\sqrt{\Delta}) = -\sqrt{\Delta}$, por lo que, en general, $\sigma(\sqrt{\Delta}) = \operatorname{sig}\sigma\, \sqrt{\Delta}$.

Teniendo esto en cuenta es inmediato el teorema siguiente:

Teorema 7.12 *Sea $f(x) \in k[x]$ un polinomio irreducible de grado n con raíces simples en su cuerpo de escisión K. Supongamos que la característica de k no es 2. Entonces, si Δ es el discriminante de f, se cumple que $k(\sqrt{\Delta})$ es el cuerpo fijado por $G(K/k) \cap A_n$. En particular $\sqrt{\Delta} \in k$ si y sólo si $G(K/k)$ consta únicamente de permutaciones pares de las raíces de f.*

En particular:

Teorema 7.13 *Sea k un cuerpo de característica distinta de 2, consideremos un polinomio irreducible $f(x) \in k[x]$ de grado 3 y discriminante Δ y sea K un cuerpo de escisión de f sobre k. Entonces:*

1. *Si Δ es un cuadrado en k entonces $|K : k| = 3$ (luego $G(K/k) \cong C_3$).*
2. *Si Δ no es un cuadrado en k entonces $|K : k| = 6$ y $G(K/k) \cong \Sigma_3$.*

DEMOSTRACIÓN: Es consecuencia inmediata del teorema anterior, teniendo en cuenta que $|K : k|$ es necesariamente múltiplo de 3, pues si α es una raíz de f se cumple que $|k(\alpha) : k| = 3$ y es un subgrupo de Σ_3. ∎

Ejemplo Los polinomios $x^3 - 3x + 1$ y $x^3 - 4x + 2$ en $\mathbb{Q}[x]$ parecen similares, pero sus discriminantes son, respectivamente, 81 y 148 (y ambos son irreducibles porque no tienen raíces enteras). Como el discriminante del primero es un cuadrado en $\mathbb{Q}$ y el del segundo no, sus grupos de Galois son distintos: el del primero es C_3 y el del segundo Σ_3. ∎

Observación Si k es un cuerpo ordenado, y R es una clausura real de k, entonces podemos interpretar también el signo del discriminante de un polinomio de grado 3. Sabemos que al menos una de sus raíces tiene que ser real (y las otras dos pueden ser reales o bien un par de raíces complejas conjugadas). Si tiene tres raíces reales distintas α, β, γ, entonces

$$\Delta(f) = a^4(\alpha - \beta)^2(\alpha - \gamma)^2(\beta - \gamma)^2 > 0.$$

Si $f(x)$ tiene una raíz real α dos raíces complejas conjugadas $\beta + \gamma i, \beta - \gamma i$, entonces

$$\Delta(f) = a^4(\alpha - \beta - \gamma i)^2(\alpha - \beta + \gamma i)^2(2\gamma i)^2 = -a^4|\alpha - \beta + \gamma i|^4 4\gamma^2 < 0.$$

Por último, si $f(x)$ tiene dos raíces iguales (necesariamente reales, al igual que la tercera), entonces $\Delta(f) = 0$. Por lo tanto, el signo del discriminante determina el carácter real o imaginario de las raíces de f. ∎

Consideremos el caso de un polinomio de grado 4:

$$f(x) = x^4 + ax^3 + bx^2 + cx + d \in k[x].$$

Supongamos que es irreducible y tiene cuatro raíces distintas $\alpha_1, \alpha_2, \alpha_3, \alpha_4$. Sea $K = k(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ el cuerpo de escisión y llamemos

$$\beta_1 = \alpha_1\alpha_2 + \alpha_3\alpha_4, \quad \beta_2 = \alpha_1\alpha_3 + \alpha_2\alpha_4, \quad \beta_3 = \alpha_1\alpha_4 + \alpha_2\alpha_3.$$

Notemos que se trata de tres elementos distintos de K. Más precisamente:

$$\begin{aligned} \beta_1 - \beta_2 &= (\alpha_1 - \alpha_4)(\alpha_2 - \alpha_3), \\ \beta_1 - \beta_3 &= (\alpha_1 - \alpha_3)(\alpha_2 - \alpha_4), \\ \beta_2 - \beta_3 &= (\alpha_1 - \alpha_2)(\alpha_3 - \alpha_4). \end{aligned} \tag{7.1}$$

Así, si identificamos a $G = G(K/k)$ con un subgrupo de Σ_4 a través de su acción sobre las raíces de f, tenemos que el cuerpo $L = k(\beta_1, \beta_2, \beta_3)$ queda fijo por $G \cap V$, donde $V = \{1, (12)(34), (13)(24), (14)(23)\}$.

Más aún, se comprueba que L es el cuerpo fijado por $G \cap V$. Por ejemplo, si $\sigma = (12) \in G$, entonces $\sigma(\beta_2) = \beta_3 \neq \beta_2$, y lo mismo vale para cualquier otra trasposición, si $\sigma = (123) \in G$, entonces $\sigma(\beta_1) = \beta_3 \neq \beta_1$, y lo mismo vale para cualquier otro ciclo de longitud 3, y si $\sigma = (1234) \in G$, entonces $\sigma(\beta_1) = \beta_3 \neq \beta_1$, y lo mismo vale para cualquier otro ciclo de longitud 4.

Como $V \trianglelefteq \Sigma_4$, también $V \cap G \trianglelefteq G$ y por lo tanto L/k es una extensión de Galois tal que $G(L/k) \cong G/(G \cap V)$.

Es pura rutina comprobar que

$$(x-\beta_1)(x-\beta_2)(x-\beta_3) = x^3 - bx^2 + (ac-4d)x - a^2d + 4bd - c^2. \tag{7.2}$$

Por ejemplo, por 7.4, el coeficiente de x es $A = \beta_1\beta_2 + \beta_1\beta_3 + \beta_2\beta_3$ y a su vez, por la definición de los β_i, esto es la suma de todos los productos $\alpha_i^2\alpha_j\alpha_k$, donde los tres índices son distintos entre sí. Por otra parte, a es la suma de los cuatro $-\alpha_i$ y c es la suma de los $-\alpha_i\alpha_j\alpha_k$, donde los índices no se repiten en cada término. Al multiplicar ac obtenemos todos los productos $\alpha_i^2\alpha_j\alpha_k$ más cuatro veces $\alpha_1\alpha_2\alpha_3\alpha_4 = d$, por lo que $A = ac - 4d$, como había que probar. Igualmente se comprueban los otros dos coeficientes.

El polinomio (7.2) se llama *resolvente cúbica* de $f(x)$.

De (7.1) se sigue inmediatamente que un polinomio y su resolvente cúbica tienen el mismo discriminante (incluso si éste es 0).

Teorema 7.14 *Sea $f(x) \in k[x]$ un polinomio irreducible de grado 4 con raíces simples en un cuerpo de escisión K, sea $G = G(K/k)$ y sea $k \subset L \subset K$ el cuerpo de escisión de la resolvente cúbica de $f(x)$. Sea $m = |L : k|$.*

1. *Si $m = 6$ entonces $G \cong \Sigma_4$.*
2. *Si $m = 3$ entonces $G \cong A_4$.*
3. *Si $m = 1$ entonces $G \cong V$.*
4. *Si $m = 2$ entonces $G \cong D_8$ o bien $G \cong C_4$. El primer caso se da si y sólo si $f(x)$ es irreducible en $L[x]$.*

Demostración: Sabemos que G tiene que ser un subgrupo de Σ_4 de orden múltiplo de 4, luego su orden puede ser $4, 8, 12$ o 24. Si el orden es 24 es obvio que $G = \Sigma_4$, si el orden es 12, entonces $G = A_4$, pues éste es el único subgrupo de orden 12 de Σ_4, si el orden es 8 entonces G es un 2-Sylow de Σ_4, luego es isomorfo a D_8 y si el orden es 4 entonces puede ser V o un C_4 generado por un ciclo de longitud 4.

Como no hay más posibilidades, basta probar la implicación $\Leftarrow$ de cada caso del enunciado. Sabemos que $m = |G : G \cap V|$, luego si G es Σ_4, A_4, D_8 o V, entonces $V \subset G$ y claramente $|G : V|$ es el indicado en el enunciado. Si $G = C_4$ entonces $|G \cap V| = 2$, luego también $m = 2$, de acuerdo con el apartado 4).

Sólo falta distinguir los dos subcasos de 4). Si $G \cong D_8$, entonces $V_4 \subset G$, y para cada par de raíces α_i, α_j de f existe $\sigma \in V$ tal que $\sigma(\alpha_i) = \alpha_j$, luego α_i y α_j son L-conjugadas, luego $f(x) = \operatorname{pol\,mín}(\alpha_i, L)$ es irreducible en $L[x]$. Por el contrario, si $G \cong C_4$, entonces $G \cap V$ tiene sólo una trasposición como elemento no trivial, y no todas las raíces de $f(x)$ son L-conjugadas, luego $f(x)$ no es irreducible en $L[x]$. ∎

Ejemplo Vamos a calcular el grupo de Galois sobre $\mathbb{Q}$ del polinomio

$$x^4 - x^3 - 5x^2 + 1.$$

Para ello calculamos su resolvente cúbica, que resulta ser $x^3 + 5x^2 - 4x - 21$. Su discriminante es $\Delta = 6\,809 = 11 \cdot 619$, que no es un cuadrado en $\mathbb{Z}$, luego tampoco en $\mathbb{Q}$. El teorema 7.13 nos da que el cuerpo de escisión de la resolvente cúbica tiene grado $m = 6$, luego por el teorema anterior $G \cong \Sigma_4$.

En particular, una raíz α de este polinomio cumple que $|\mathbb{Q}(\alpha) : \mathbb{Q}| = 4$, pero no es constructible con regla y compás (como ya advertíamos que podía suceder tras [G 6.7]), porque según [G 6.8] el cuerpo $\mathcal{C}$ de los números constructibles es una extensión normal de $\mathbb{Q}$, luego si α fuera constructible, todos sus conjugados lo serían también, y el cuerpo de escisión K del polinomio estaría contenido en $\mathcal{C}$, lo que contradice a [G 6.10] ya que $|K : \mathbb{Q}| = 24$, que no es potencia de 2. ∎

Teorema 7.15 *Sea k un cuerpo de característica 0, sean $d, e, f \in k$ tales que d no sea un cuadrado en k, pero que $d(e^2 - f^2 d)$ sí que lo sea. Entonces el cuerpo $K = k\Big(\sqrt{e + f\sqrt{d}}\Big)$ es una extensión cíclica de k de grado 4, y el polinomio mínimo de $\sqrt{e + f\sqrt{d}}$ es $p(x) = x^4 - 2ex^2 + (e^2 - f^2 d)$.*

Demostración: Claramente, $e^2 - f^2 d$ no puede ser un cuadrado en k, o d también lo sería. Veamos ahora que $e + f\sqrt{d}$ no es un cuadrado en $k' = k(\sqrt{d})$. En efecto, si

$$e + f\sqrt{d} = (r + s\sqrt{d})^2 = r^2 + s^2 d + 2rs\sqrt{d},$$

entonces $r^2 + s^2 d = e$, $f = 2rs$, pero entonces

$$e^2 - f^2 d = (r^2 + s^2 d)^2 - 4r^2 s^2 d = (r^2 - s^2 d)^2,$$

contradicción. Por lo tanto tenemos una cadena de extensiones cuadráticas:

$$k \subset k(\sqrt{d}) \subset k'\Big(\sqrt{e + f\sqrt{d}}\Big) = k\Big(\sqrt{e + f\sqrt{d}}\Big).$$

La última igualdad se debe a que K contiene a $e + f\sqrt{d}$ luego también a $\sqrt{d}$, luego a $k'\Big(\sqrt{e + f\sqrt{d}}\Big)$ y, por otra parte, es claro que $p(x)$ tiene a $\sqrt{e + f\sqrt{d}}$ por raíz, luego K/k tiene grado 4 y $p(x)$ es el polinomio mínimo del elemento primitivo. Sus raíces son

$$\alpha_1 = \sqrt{e + f\sqrt{d}}, \quad \alpha_2 = -\sqrt{e + f\sqrt{d}}, \quad \alpha_3 = \sqrt{e - f\sqrt{d}}, \quad \alpha_4 = -\sqrt{e - f\sqrt{d}}.$$

Notemos que $\sqrt{d}\,\alpha_1\alpha_3 = \sqrt{d(e^2 - f^2 d)} \in k$, luego $\alpha_3 \in k(\alpha_1) = K$, luego K es el cuerpo de escisión de f. Además, una raíz de la resolvente cúbica es $\beta_1 = \alpha_1\alpha_2 + \alpha_3\alpha_4 = 2f\sqrt{d} \notin k$, lo cual descarta el caso 3) del teorema 7.14, y nos permite concluir que K/k es una extensión cíclica. ∎

Ejemplo Tomando $d = e = 2$, $f = 1$ en el teorema anterior, concluimos que el polinomio $x^4 - 4x^2 + 2$ es irreducible en $\mathbb{Q}[x]$ y que su grupo de Galois es C_4. Una de sus raíces es $\sqrt{2+\sqrt{2}}$. ∎

Veamos ahora un ejemplo de grado 5, para lo cual probamos en general:

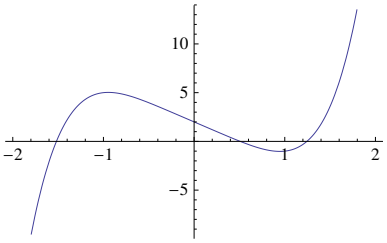
Teorema 7.16 *Si $f(x) \in \mathbb{Q}[x]$ es un polinomio irreducible de grado primo p con exactamente dos raíces imaginarias, entonces su grupo de Galois es Σ_p.*

DEMOSTRACIÓN: Sea K el cuerpo de escisión de $f(x)$ sobre $\mathbb{Q}$. Entonces $G = G(K/\mathbb{Q})$ se identifica de forma usual con un subgrupo de Σ_p y tiene orden múltiplo de p (porque f es irreducible, luego al adjuntar a $\mathbb{Q}$ una raíz obtenemos una extensión de grado p), luego contiene un elemento de orden p. Pero los elementos de orden p en Σ_p son los ciclos de longitud p. Por lo tanto, G contiene uno de estos ciclos.

Por otra parte, la conjugación compleja se restringe a un elemento de G que deja invariantes a todas las raíces reales y permuta a las dos imaginarias, luego es una trasposición. Ahora basta aplicar el teorema [TG 2.13], que nos da que $G = \Sigma_p$. ∎

El teorema 9.8 permite construir polinomios con grupo de Galois Σ_n, para cualquier natural $n \geq 2$. Un ejemplo explícito es $x^n - x - 1$, pero no es fácil de probar.[1]

Ejemplo El polinomio $x^5 - 4x + 2$ es irreducible en $\mathbb{Q}[x]$ por el criterio de Eisenstein. Vemos en la gráfica que tiene tres raíces reales. Esto puede justificarse considerando su derivada, $5x^4 - 4$, cuyas raíces son $\pm\alpha$, con $\alpha = \sqrt[4]{4/5}$, y es fácil ver que es positiva en $]-\infty, -\alpha[$ y en $]\alpha, +\infty[$, por lo que el polinomio crece en dichos intervalos y decrece en $]-\alpha, \alpha[$. Esto implica que sólo puede tener una raíz en cada intervalo, tres en total. La existencia de raíces se sigue del teorema de los valores intermedios. Por consiguiente, hay dos raíces complejas, y el teorema anterior implica que su grupo de Galois es Σ_5. ∎



Ejemplo Vamos a probar que el grupo de Galois del polinomio

$$x^8 - 72x^6 + 180x^4 - 144x^2 + 36 \in \mathbb{Q}[x]$$

es isomorfo a Q_8.

En lugar de partir de dicho polinomio, vamos a construir una extensión de $\mathbb{Q}$ con grupo de Galois isomorfo a Q_8 y veremos que el polinomio anterior es el polinomio mínimo de un elemento primitivo. Partimos de $k_1 = \mathbb{Q}\big[\sqrt{2}\,\big]$,

[1] Véase mi libro de Teoría algebraica de números, teorema [TAl 9.47].

$k_2 = \mathbb{Q}\big[\sqrt{3}\,\big]$, $k_3 = \mathbb{Q}\big[\sqrt{6}\,\big]$, de modo que las tres extensiones cuadráticas de $\mathbb{Q}$ están contenidas en $k' = \mathbb{Q}[\sqrt{2}, \sqrt{3}]$. Definimos

$$\theta = (2+\sqrt{2})(2+\sqrt{3})(3+\sqrt{6}) = 18 + 12\sqrt{2} + 10\sqrt{3} + 7\sqrt{6}.$$

Ahora observamos que

$$\theta = \begin{cases} (18+12\sqrt{2}) + (10+7\sqrt{2})\sqrt{3} & = e_1 + f_1\sqrt{d_1}, \\ (18+10\sqrt{3}) + (12+7\sqrt{3})\sqrt{2} & = e_2 + f_2\sqrt{d_2}, \\ (18+\ \ 7\sqrt{6}) + (12+5\sqrt{6})\sqrt{2} & = e_3 + f_3\sqrt{d_3}. \end{cases}$$

Observemos que los tres grupos de números e_i, f_i, d_i cumplen las hipótesis del teorema 7.15 sobre el cuerpo k_i. En primer lugar, $\sqrt{3} \notin k_1$, $\sqrt{2} \notin k_2$, $\sqrt{2} \notin k_3$. Esto puede comprobarse directamente,[2] viendo que, por ejemplo, $(a+b\sqrt{2})^2 = 2$ es imposible en k_1. En segundo lugar comprobamos que $d_i(e_i^2 - f_i^2 d_i)$ sí que es un cuadrado en k_i:

$$3((18+12\sqrt{2})^2 - (10+7\sqrt{2})^2 \cdot 3) = (3(2+\sqrt{2}))^2,$$

$$2((18+10\sqrt{3})^2 - (12+7\sqrt{3})^2 \cdot 2) = (2(3+2\sqrt{3}))^2,$$

$$2((18+7\sqrt{6})^2 - (12+5\sqrt{6})^2 \cdot 2) = (2(3+\sqrt{6}))^2.$$

Observemos además que $k_i(\sqrt{d_i}) = k'$, independientemente de i. Por lo tanto, el teorema 7.15 nos da que $K = k_i(\sqrt{\theta}) = k'(\sqrt{\theta})$ es una extensión cíclica de grado 4 de todos los k_i, y a su vez $|K:\mathbb{Q}| = 8$. El polinomio mínimo sobre k_1' de $\sqrt{\theta}$ es

$$p(x) = x^4 - 2e_1x^2 + (e_1^2 - 3f_1^2)$$

Como no está en $\mathbb{Q}[x]$, el polinomio mínimo de $\sqrt{\theta}$ en $\mathbb{Q}[x]$ tiene que ser un múltiplo de éste de mayor grado, luego tiene que ser de grado 8 y por consiguiente $K = \mathbb{Q}(\sqrt{\theta})$. Para calcular dicho polinomio basta encontrar un polinomio mónico de grado 8 que tenga a $\sqrt{\theta}$ por raíz. Para ello consideramos el automorfismo de $G(k'/\mathbb{Q})$ dado por $\sqrt{2} \mapsto -\sqrt{2}$, $\sqrt{3} \mapsto \sqrt{3}$. Representaremos por $\bar{\alpha}$ la imagen de cada $\alpha \in k'$ por dicho automorfismo. Claramente entonces

$$\bar{p}(x) = x^4 - 2\bar{e}_2x^2 + (\bar{e}_1^2 - 3\bar{f}_1^2)$$

es el polinomio mínimo sobre k_1 de $\sqrt{\bar{\theta}}$, donde

$$\bar{\theta} = (2-\sqrt{2})(2+\sqrt{3})(3-\sqrt{6}).$$

Un cálculo muestra que $\theta\bar{\theta} = 6(2+\sqrt{3})^2$, luego

$$\sqrt{\bar{\theta}} = \frac{(2+\sqrt{3})\sqrt{6}}{\theta} \in K$$

[2] Más adelante podremos decir que esto es inmediato, ya que los tres cuerpos son distintos por tener discriminantes distintos.

Como la extensión K/k_1 es de Galois y $\bar{p}(x) \in k_1[x]$ tiene una raíz en K, tiene de hecho todas sus raíces en K, luego el polinomio

$$q(x) = p(x)\bar{p}(x) = (x^4 - 2e_1x^2 + (e_1^2 - 3f_1^2))(x^4 - 2\bar{e}_1x^2 + (\bar{e}_1^2 - 3\bar{f}_1^2))$$
$$= x^8 - 72x^6 + 180x^4 - 144x^2 + 36$$

tiene todas sus raíces en K. Como una de ellas es $\sqrt{\theta}$ y tiene grado 8, es $\operatorname{pol\,mín}(\sqrt{\theta}, \mathbb{Q})$, lo que prueba que K es su cuerpo de escisión sobre $\mathbb{Q}$. Sólo falta probar que $G = G(K/\mathbb{Q}) \cong Q_8$.

Para ello usamos el teorema [TG 1.40] (más la clasificación de los grupos abelianos), según el cual G tiene que ser isomorfo a uno de los grupos

$$C_8, \qquad C_4 \times C_2, \qquad C_2 \times C_2 \times C_2 \qquad D_8 \qquad Q_8.$$

La clave es que los tres subcuerpos k_i se corresponden con tres subgrupos cíclicos de G de orden 4 (son cíclicos porque las extensiones K/k_i son cíclicas). Esto descarta a C_8, que sólo tiene un subgrupo de orden 4, a $C_4 \times C_2$, que sólo tiene dos subgrupos cíclicos de orden 4, a $C_2 \times C_2 \times C_2$, que no tiene ninguno, y a D_8, que sólo tiene uno. ∎

7.4 Ecuaciones cúbicas

Presentamos ahora la fórmula general para la resolución de una ecuación cúbica:

Teorema 7.17 (Fórmula de Cardano) *Sea K un cuerpo de característica distinta de 2 o 3 y sean a, b, $c \in K$. Entonces, las raíces de la ecuación*

$$x^3 + ax^2 + bx + c = 0 \tag{7.3}$$

en una clausura algebraica de K vienen dadas por

$$x = \sqrt[3]{-q/2 + \sqrt{D}} + \sqrt[3]{-q/2 - \sqrt{D}} - a/3,$$

donde

$$p = \frac{3b - a^2}{3}, \qquad q = \frac{2a^3 - 9ab + 27c}{27}, \qquad D = \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3,$$

la raíz cuadrada de D se escoge arbitrariamente y, fijada ésta, las raíces cúbicas u y v se escogen de modo que $p = -3uv$ (es decir, se escoge una arbitrariamente y la otra se calcula mediante esta relación).

Notemos ante todo que la relación entre las raíces cúbicas es correcta, es decir, que si

$$u = \sqrt[3]{-q/2 + \sqrt{D}}$$

es una raíz cúbica arbitraria del radicando y definimos v mediante $p = -3uv$, entonces

$$v = \sqrt[3]{-q/2 - \sqrt{D}}.$$

En efecto, elevando al cubo vemos que $p^3 = -27(-q/2 + \sqrt{D})v^3$, luego

$$v^3 = -\frac{(p/3)^3(-q/2 - \sqrt{D})}{(q/2)^2 - D} = -q/2 - \sqrt{D},$$

como queríamos probar.

Demostración: En primer lugar realizamos a la ecuación el cambio de variable

$$x = t - \frac{a}{3}, \tag{7.4}$$

es decir, consideramos el polinomio $f(t - a/3)$, que resulta ser la llamada "forma incompleta" de la ecuación cúbica:

$$t^3 + pt + q = 0, \tag{7.5}$$

donde p y q son los valores indicados en el enunciado.

No es casual que haya desaparecido el término de grado 2. Pensemos que $-a$ es la suma de las tres raíces de $f(x)$, luego $-a/3$ es su media. Las raíces de $f(t - a/3)$ los las que resultan de sumar $a/3$ a las de $f(x)$, y es claro que si a tres números les restamos su media, los tres números resultantes tienen media (luego suma) 0. Por lo tanto, el resultado del cambio de variable tenía que tener nulo el coeficiente de t^2.

Observemos también que, como las raíces de (7.5) se obtienen sumando un mismo número a las de (7.3), la fórmula que da el discriminante en términos de las raíces proporciona el mismo resultado para ambos, y aplicando a (7.5) la fórmula que hemos dado al final de la sección anterior, el discriminante resulta ser

$$\Delta = -4p^3 - 27q^2 = -2^2 \cdot 3^3 D.$$

Así pues, a efectos prácticos D contiene la misma información que el determinante. En particular sabemos $D \neq 0$ si y sólo si las tres raíces son distintas.

Ahora basta probar que las raíces de (7.5) son las dadas por la fórmula del enunciado sin el término $-a/3$ final, pues entonces la relación (7.4) nos da que las raíces de (7.3) se obtienen añadiendo dicho término. Vamos a deducir la fórmula de Cardano bajo la hipótesis adicional $p \neq 0$, y luego veremos que es válida incluso si $p = 0$. Partimos del desarrollo

$$(u + v)^3 = u^3 + v^3 + 3u^2v + 3uv^2 = u^3 + v^3 + 3uv(u + v),$$

que nos da la identidad

$$(u + v)^3 - 3uv(u + v) - u^3 - v^3 = 0. \tag{7.6}$$

Por lo tanto, si encontramos valores de u, v tales que

$$p = -3uv, \qquad q = -u^3 - v^3, \tag{7.7}$$

tendremos que una solución de (7.5) será $t = u + v$. Podemos expresar (7.7) en términos de u despejando $v = -p/3u$. (Notemos que ha de ser $u \neq 0$, ya que suponemos $p \neq 0$.) Concluimos que una condición suficiente para que t sea solución de (7.5) es que sea de la forma

$$t = u - \frac{p}{3u} \tag{7.8}$$

para un u que cumpla la ecuación

$$u^3 + q - (p/3u)^3 = 0 \tag{7.9}$$

o, equivalentemente (multiplicando por u^3):

$$u^6 + qu^3 - (p/3)^3 = 0. \tag{7.10}$$

Veamos ahora que la condición es necesaria, es decir, que toda raíz t de (7.5) es de la forma (7.8), para un cierto u que cumple (7.10). En primer lugar, dado cualquier valor de t, siempre existe un valor de $u \neq 0$ que cumple (7.8). Basta tomar una raíz de la ecuación

$$u^2 - tu - p/3 = 0. \tag{7.11}$$

Llamando $v = -p/3u$, tenemos que $t = u + v$, $p = -3uv$. Si t cumple (7.5), entonces

$$(u+v)^3 - 3uv(u+v) + q = 0$$

y, comparando con la identidad (7.6), vemos que se cumple (7.7), lo cual implica que u cumple (7.9) y, por consiguiente, (7.10).

Así pues, concluimos que resolver (7.5) es equivalente a resolver (7.10), en el sentido de que las soluciones de (7.5) son las que se obtienen a partir de las de (7.10) a través de (7.8). Ahora bien, las ecuación (7.10) es cuadrática en u^3, luego sus soluciones son

$$u^3 = \frac{-q \pm \sqrt{q^2 + 4(p/3)^3}}{2} = -\frac{q}{2} \pm \sqrt{(q/2)^2 + (p/3)^3}$$

o, más sencillamente:

$$u^3 = -q/2 \pm \sqrt{D}. \tag{7.12}$$

Combinando las dos raíces cuadradas de D con las tres raíces cúbicas, nos salen las seis raíces de (7.10), pero (7.5) sólo puede tener tres raíces. Ello se debe a que, para cada valor de t, hay dos valores de u que cumplen (7.8). Dado uno de ellos, el otro está determinado por la relación

$$uu' = -p/3, \tag{7.13}$$

que se deduce de que u y u' son las dos raíces de (7.11).

Fijemos un u_+ que cumpla (7.12) con signo positivo y una raíz cuadrada prefijada, y sea u_- una raíz de (7.12) con signo negativo (y la misma raíz cuadrada). Entonces,

$$(u_+u_-)^3 = (q/2)^2 - (q/2)^2 - (p/3)^3 = -(p/3)^3,$$

luego $u_+u_- = -(p/3)\omega$, donde ω es una cierta raíz cúbica de la unidad. Por consiguiente, $u_+(\omega^2 u_-) = -(p/3)$. Cambiando u_- por $\omega^2 u_-$ (que también cumple (7.12) con el signo negativo), concluimos que, para cada u_+ elegido arbitrariamente, existe un u_- elegido adecuadamente tal que u_+ y u_- determinan la misma raíz t de (7.5). En definitiva, vemos que las soluciones de (7.5) son de la forma (7.8), con

$$u = \sqrt[3]{-q/2 + \sqrt{D}},$$

donde en esta expresión hay que entender que la raíz cuadrada de D es fija (elegida arbitrariamente de entre las dos posibilidades) y que al variar la elección de la raíz cúbica recorremos las distintas raíces de (7.5). Finalmente observamos que, de la segunda ecuación de (7.7), se sigue que

$$v^3 = -q/2 - \sqrt{D},$$

luego podemos escribir

$$v = \sqrt[3]{-q/2 - \sqrt{D}},$$

si bien es crucial tener presente que las elecciones de las raíces cúbicas u y v no pueden hacerse de forma independiente, lo cual nos daría 9 raíces para (7.5), sino que u y v han de cumplir la relación $p = -3uv$.

Con esto hemos probado la fórmula de Cardano bajo el supuesto de que $p \neq 0$. Ahora bien, en caso contrario $D = (q/2)^2$, luego $\sqrt{D} = \pm q/2$. Al sustituir cualquiera de las dos elecciones del signo en los radicandos de las raíces cúbicas obtenemos que una de las dos se anula y la otra se reduce a $\sqrt[3]{-q}$ y, ciertamente, las tres elecciones de la raíz cúbica nos dan las tres raíces de (7.5) en este caso. Además, la relación $p = -3uv$ se cumple trivialmente. ∎

Sabemos que si $D = 0$ el polinomio tiene una raíz doble (o triple). En tal caso hay fórmulas más sencillas para calcularlas:

Teorema 7.18 *Si $D = 0$ hay dos posibilidades:*

1. *Si $p = q = 0$, entonces la ecuación tiene una raíz triple $x = -a/3$.*
2. *Si $pq \neq 0$, entonces la ecuación tiene una raíz doble y una raíz simple, dadas respectivamente por*

$$x = -\frac{3q}{2p} - \frac{a}{3}, \quad y \quad x = -\frac{4p^2}{9q} - \frac{a}{3}.$$

En particular, las tres raíces están en el mismo cuerpo que los coeficientes de la ecuación.

DEMOSTRACIÓN: Si $p = q = 0$ la fórmula de Cardano nos da inmediatamente 1). En caso contrario, la condición $D = 0$ equivale a $27q^2 = -4p^3$, y usando esto se comprueba sin dificultad que $t = -3q/2p$ es raíz tanto de (7.5) como de su derivada $3t^2 + p = 0$, por lo que es una raíz doble, y al sumarle $-a/3$ tenemos una raíz doble de (7.3).

Para calcular la raíz simple usamos que el producto de las tres raíces de (7.5) tiene que ser $-q$, lo que nos da que la tercera raíz es $t = -4p^2/9q$. ∎

Supongamos ahora que k es un cuerpo ordenado y fijemos una clausura real R. Entonces cabe distinguir dos casos según el signo de D (que es el signo opuesto al del discriminante del polinomio). Sabemos que si $D > 0$ entonces $f(x)$ tiene una raíz real y dos raíces imaginarias conjugadas, mientras que si $D < 0$ las tres raíces son reales. Veamos cómo queda la fórmula de Cardano en el primer caso:

Teorema 7.19 *Si $D > 0$, una raíz real viene dada por*

$$x = \sqrt[3]{-q/2 + \sqrt{D}} + \sqrt[3]{-q/2 - \sqrt{D}} - a/3,$$

donde las raíces cúbicas u y v son reales. Las otras dos raíces son imaginarias, y vienen dadas por

$$x = -\frac{u+v}{2} - \frac{a}{3} \pm \frac{\sqrt{3}}{2}(u-v)\,i.$$

DEMOSTRACIÓN: El radicando de u es un número real, por lo que la única raíz cúbica real es una elección posible para u. La relación $p = -3uv$, implica que la raíz cúbica v correspondiente es también la raíz cúbica real, luego $t = u+v$ es una raíz real. Las otras elecciones para u son ωu y $\omega^2 u$, donde

$$\omega = -\frac{1}{2} + \frac{\sqrt{3}}{2}\,i,$$

es una raíz cúbica de la unidad, y el valor de v correspondiente es, respectivamente, $\omega^2 v$ y ωv. Por lo tanto, una raíz imaginaria de (7.5) es

$$t = \omega u + \omega^2 v = -\frac{u+v}{2} + \frac{\sqrt{3}}{2}(u-v)\,i$$

(y la otra ha de ser su conjugada), de donde se sigue inmediatamente la expresión dada para las raíces de (7.3) en este caso. ∎

Ejemplo Vamos a resolver la ecuación

$$x^3 + 3x^2 + 6x + 2 = 0.$$

El cambio de variable $x = t - 1$ la transforma en

$$t^3 + 3t - 2 = 0.$$

Vemos entonces que $D = 2 > 0$, luego la raíz real es

$$x = \sqrt[3]{1+\sqrt{2}} + \sqrt[3]{1-\sqrt{2}} - 1.$$

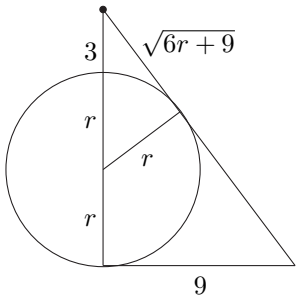
Las raíces imaginarias son

$$x = -\frac{\sqrt[3]{1+\sqrt{2}} + \sqrt[3]{1-\sqrt{2}}}{2} \pm \frac{\sqrt{3}}{2}\left(\sqrt[3]{1+\sqrt{2}} - \sqrt[3]{1-\sqrt{2}}\right) i. \qquad \blacksquare$$

Nota Observemos que si en el ejemplo anterior tomamos a $\mathbb{Q}$ como cuerpo de coeficientes de la ecuación y llamamos α a la raíz real, no es cierto que α se obtenga a partir de una raíz cúbica de un elemento de $\mathbb{Q}$ (o de varios), sino que para obtener α necesitamos extraer además la raíz cuadrada $\sqrt{2}$. $\blacksquare$

Ejemplo El problema siguiente fue planteado en el siglo XIII por el matemático chino Qin Jinshao:

> *Una ciudad está rodeada por una muralla circular con dos puertas, una al norte y otra al sur. Saliendo por la puerta norte y caminando* 3 *li hacia el norte se llega hasta un árbol. Saliendo por la puerta sur, hay que caminar* 9 *li hacia el este para ver el mismo árbol. Calcular el diámetro de la ciudad.*



Los dos triángulos que muestra la figura son semejantes, de modo que se ha de cumplir

$$\frac{2r+3}{9} = \frac{\sqrt{3(2r+3)}}{r}.$$

Elevando al cuadrado, simplificando un factor $2r+3$ y operando llegamos a la ecuación

$$r^3 + \frac{3}{2}r^2 - \frac{243}{2} = 0.$$

El cambio $r = t - 1/2$ la reduce a $t^3 - \dfrac{3}{4}t - \dfrac{485}{4} = 0$. Como $D = 29\,403/8 > 0$, sólo tiene una raíz real, que viene dada por

$$t = \sqrt[3]{\frac{485}{8} + \frac{198}{8}\sqrt{6}} + \sqrt[3]{\frac{485}{8} - \frac{198}{8}\sqrt{6}}$$

$$= \frac{1}{2}\left(\sqrt[3]{485+198\sqrt{6}} + \sqrt[3]{485-198\sqrt{6}}\right) = 5.$$

Concluimos que $r = 4.5$ y que el diámetro de la ciudad es de 9 li. $\blacksquare$

Veamos ahora la situación cuando el discriminante es negativo:

Teorema 7.20 *Si $D < 0$ la ecuación tiene tres raíces reales simples, que vienen dadas por*

$$x = 2\sqrt{-\frac{p}{3}}\,\cos\frac{\theta + 2k\pi}{3} - \frac{a}{3},$$

donde $k = 0, 1, 2$ y el ángulo $0 < \theta < \pi$ está determinado por

$$\cos\theta = \frac{-q/2}{\sqrt{-(p/3)^3}}. \tag{7.14}$$

Demostración: Para aplicar la fórmula de Cardano podemos elegir

$$\sqrt{D} = \sqrt{|D|}\,i = \sqrt{-(q/2)^2 - (p/3)^3}\,i,$$

es decir, nos quedamos con la raíz cuadrada con parte imaginaria positiva. El módulo de $-q/2 + \sqrt{D}$ es

$$\rho = \sqrt{(q/2)^2 - (q/2)^2 - (p/3)^3} = \sqrt{-(p/3)^3},$$

y su argumento es el ángulo $0 < \theta < \pi$ que cumple (7.14). Por consiguiente, los valores posibles de u son los dados por

$$u = \sqrt{-\frac{p}{3}}\left(\cos\frac{\theta + 2k\pi}{3} + i\,\mathrm{sen}\,\frac{\theta + 2k\pi}{3}\right).$$

La ecuación $p = -3uv$ nos da que

$$v = \sqrt{-\frac{p}{3}}\left(\cos\frac{\theta + 2k\pi}{3} - i\,\mathrm{sen}\,\frac{\theta + 2k\pi}{3}\right),$$

y al calcular $x = u + v - a/3$ obtenemos la expresión anunciada. En particular vemos que las raíces son reales. ∎

Ejemplo La fórmula $\cos 3\alpha = 4\cos^3\alpha - 3\cos\alpha$ nos da que $\cos 20^\circ$ satisface la ecuación

$$x^3 - \frac{3}{4}x - \frac{1}{8} = 0.$$

Uno podría pensar que resolviéndola con la fórmula de Cardano obtendremos una expresión algebraica para $\cos 20^\circ$, pero no es así. Al aplicar el teorema anterior obtenemos que $\theta = 60^\circ$, y que las soluciones de la ecuación son $\cos 20^\circ$, $\cos 140^\circ$ y $\cos 260^\circ$. Más precisamente, la fórmula de Cardano expresa las soluciones de una ecuación con $D > 0$ en términos de raíces cúbicas de números complejos, que implican trisecar ángulos, y para calcularlas tenemos que usar las funciones trigonométricas, luego al final la solución queda irremediablemente en términos de un coseno. ∎

7.5 Ecuaciones cuárticas

También es posible expresar en términos de radicales las raíces de una ecuación de cuarto grado:

Teorema 7.21 (Fórmula de Ferrari) *Sea K un cuerpo de característica distinta de 2 o 3 y sean a, b, c, $d \in K$. Entonces, las raíces de la ecuación*

$$x^4 + ax^3 + bx^2 + cx + d = 0. \tag{7.15}$$

en una clausura algebraica de K vienen dadas por

$$x = \frac{Q \pm \sqrt{Q^2 - 4(P-R)}}{2} - \frac{a}{4}, \qquad x = \frac{-Q \pm \sqrt{Q^2 - 4(P+R)}}{2} - \frac{a}{4},$$

donde, llamando

$$p = \frac{8b - 3a^2}{8}, \quad q = \frac{8c - 4ab + a^3}{8}, \quad r = \frac{256d - 64ac + 16a^2b - 3a^4}{256}, \tag{7.16}$$

P es una raíz de la ecuación

$$P^3 - \frac{p}{2}P^2 - rP + \frac{4pr - q^2}{8} = 0, \tag{7.17}$$

y Q, R se determinan mediante las ecuaciones

$$p = 2P - Q^2, \quad q = -2QR, \quad r = P^2 - R^2. \tag{7.18}$$

En la prueba veremos, más concretamente, que, si $q \neq 0$, la primera ecuación de (7.18) es redundante, de modo que, a partir de una solución P de (7.17), la tercera ecuación de (7.18) nos da un valor para R, necesariamente no nulo, y la segunda ecuación nos da un valor para Q que necesariamente cumplirá la primera ecuación. Si $q = 0$ el sistema (7.18) tiene también una solución fácil de calcular, pero enseguida veremos que en este caso hay un procedimiento más rápido para encontrar las raíces de la ecuación. En efecto, el cambio de variable

$$x = t - \frac{a}{4}$$

nos lleva a la ecuación incompleta

$$t^4 + pt^2 + qt + r = 0, \tag{7.19}$$

donde p, q, r son los dados por (7.16). Así, si $q = 0$, tenemos lo que se conoce como una ecuación bicuadrada, cuyas raíces cumplen:

$$t^2 = \frac{-p \pm \sqrt{p^2 - r}}{2},$$

luego las cuatro raíces de (7.15) son

$$x = \pm\sqrt{\frac{-p \pm \sqrt{p^2 - r}}{2}} - \frac{a}{4}.$$

Es fácil ver que si a (7.17) le hacemos el cambio de variable $P = x/2$ y dividimos entre el coeficiente director obtenemos la resolvente cúbica de (7.19).

DEMOSTRACIÓN: Ya hemos visto que basta resolver (7.19). Para ello observamos que

$$(t^2 + P)^2 - (Qt + R)^2 = t^4 + (2P - Q^2)t^2 - 2QRt + P^2 - R^2.$$

Por lo tanto, si encontramos valores P, Q, R que cumplan (7.18), las soluciones de (7.19) serán las mismas que las de la ecuación

$$(t^2 + P)^2 = (Qt + R)^2.$$

Ésta puede descomponerse en dos ecuaciones cuadráticas:

$$t^2 + P = \pm(Qt + R).$$

Equivalentemente,

$$t^2 - Qt + P - R = 0, \qquad t^2 + Qt + P + R = 0, \tag{7.20}$$

y las raíces de estas ecuaciones son las indicadas en el enunciado.

Así pues, sólo hemos de encontrar una solución de (7.18). Para ello despejamos

$$Q = -\frac{q}{2R}, \quad Q^2 = \frac{q^2}{4R^2} = \frac{q^2}{4(P^2 - r)},$$

con lo que la primera ecuación se convierte en

$$p = 2P - \frac{q^2}{4(P^2 - r)}$$

o, equivalentemente, en la cúbica

$$4(P^2 - r)(2P - p) = q^2 \tag{7.21}$$

que se simplifica hasta (7.17).

De este modo, podemos tomar una raíz cualquiera P de (7.17), luego una raíz cualquiera R de $r = P^2 - R^2$ y, por último, si $R \neq 0$, hacemos $Q = -q/2R$. Así, las dos últimas ecuaciones de (7.18) se cumplen por las elecciones de Q y R, mientras que la primera se cumple porque es equivalente a (7.17).

Si nos encontramos con $R = 0$ es porque $P^2 - r = 0$ y, a la vista de (7.21), para que esto pueda suceder ha de ser $q = 0$. Más aún, en tal caso, esta misma ecuación muestra que las raíces de (7.17) son $P = \sqrt{r}$ o bien $P = p/2$ y, tomando $R = 0$ en el primer caso o $Q = 0$ en el segundo, encontramos igualmente una solución de (7.18).

En cualquier caso, vemos que cualquier raíz P de (7.17) se puede completar (fácilmente) hasta una solución (P, Q, R) del sistema (7.18). ∎

Ejemplo (Euler) Vamos a resolver la ecuación

$$x^4 - 8x^3 + 14x^2 + 4x - 8 = 0.$$

El cambio de variable $x = t + 2$ la reduce a

$$t^4 - 10t^2 - 4t + 8 = 0.$$

La cúbica auxiliar es

$$P^3 + 5P^2 - 8P - 42 = 0.$$

Al resolverla se obtiene que una raíz es $P = -3$, y ahora hemos de resolver

$$-4 = -2QR, \quad 8 = 9 - R^2.$$

Tomamos $R = 1$, $Q = 2$ y entonces las cuatro raíces de la ecuación original son:

$$x = \frac{2 \pm \sqrt{2^2 - 4(-3-1)}}{2} - \frac{-8}{4} = 3 \pm \sqrt{5},$$

$$x = \frac{-2 \pm \sqrt{2^2 - 4(-3+1)}}{2} - \frac{-8}{4} = 1 \pm \sqrt{3}.$$

■

Ejercicio: Nicolaus Bernoulli afirmó que no era cierta la conjetura según la cual todo polinomio (con coeficientes reales) puede descomponerse en producto de polinomios de grados 1 y 2, y presentó como contraejemplo el polinomio

$$x^4 - 4x^3 + 2x^2 + 4x + 4.$$

Sin embargo, Euler demostró que dicho polinomio factoriza como

$$\left(x^2 - \left(2 + \sqrt{4 + 2\sqrt{7}}\right) x + 1 + \sqrt{7} + \sqrt{4 + 2\sqrt{7}}\right) \left(x^2 - \left(2 - \sqrt{4 + 2\sqrt{7}}\right) x + 1 + \sqrt{7} - \sqrt{4 + 2\sqrt{7}}\right).$$

Demostrar esta factorización (sin hacer uso de ella, es decir, que no vale limitarse a multiplicar y ver que el resultado es correcto).

7.6 Extensiones radicales

Las fórmulas de Cardano y Ferrari (junto con la fórmula para ecuaciones de segundo grado) muestran que toda ecuación polinómica de grado 2, 3 o 4 es "resoluble por radicales", en el sentido de que sus raíces se pueden expresar mediante sumas, productos, cocientes y raíces a partir de sus coeficientes. Vamos a caracterizar algebraicamente los números que pueden expresarse de este modo a partir de elementos de un cuerpo dado. Por simplicidad trabajaremos únicamente con cuerpos de característica 0.

Definición 7.22 Una extensión de cuerpos K/k es *radical* si existen elementos $a_1, \ldots, a_n$ en K tales que $K = k(a_1, \ldots, a_n)$ y existen naturales no nulos $r_1, \ldots, r_n$ de manera que $a_1^{r_1} \in k$ y $a_i^{r_i} \in k(a_1, \ldots, a_{i-1})$, para $i = 2, \ldots, n$.

Así, los elementos de $k(a_1)$ son polinomios en a_1 con coeficientes en k y a_1 es una raíz r_1-ésima de un elemento de k, los elementos de $k(a_1, a_2)$ son polinomios en a_2 con coeficientes en $k(a_1)$ y a_2 es una raíz r_2-ésima de un elemento de $k(a_1)$. En general todos los elementos de K se pueden obtener a partir de los de k mediante sumas, productos, cocientes y extracción de raíces.

Recíprocamente, si un elemento a admite una expresión de este tipo a partir de ciertos elementos de k, es claro que a está contenido en una extensión radical de k.

Por lo tanto, si $p(x)$ es un polinomio no constante con coeficientes en un cuerpo k, diremos que la ecuación $p(x) = 0$ es *resoluble por radicales* si existe una extensión radical K/k tal que $p(x)$ se escinde en K.

Esto equivale a que las raíces de $p(x)$ se puedan expresar mediante sumas, productos, cocientes y raíces a partir de los elementos de k. En estos términos, toda ecuación de grado 2, 3 o 4 es resoluble por radicales.

Ejemplo Consideremos la ecuación $x^{10} - 5x^5 + 5 = 0$. Sus diez raíces cumplen

$$x^5 = \frac{5 \pm \sqrt{5}}{2}, \qquad \text{luego} \qquad x = \sqrt[5]{\frac{5 \pm \sqrt{5}}{2}}.$$

Esta fórmula prueba que la ecuación es resoluble por radicales. Para expresarlo en términos de la definición que hemos dado consideramos los cuerpos

$$\mathbb{Q} \subset \mathbb{Q}\big(\sqrt{5}\,\big) \subset \mathbb{Q}\big(\sqrt{5}, \omega\big) \subset \mathbb{Q}\big(\sqrt{5}, \omega, \alpha\big) \subset \mathbb{Q}\big(\sqrt{5}, \omega, \alpha, \beta\big) = K,$$

donde ω es una raíz quinta primitiva de la unidad, y α, β son

$$\alpha = \sqrt[5]{\frac{5 + \sqrt{5}}{2}}, \qquad \beta = \sqrt[5]{\frac{5 - \sqrt{5}}{2}}$$

(las raíces quintas se eligen arbitrariamente).

Así, las raíces de la ecuación son $\omega^i\alpha$ y $\omega^i\beta$, para $i = 1, \ldots, 5$, luego todas ellas están en K, y también es claro que $K/\mathbb{Q}$ es una extensión radical. ∎

El teorema siguiente justifica que podemos limitarnos a trabajar con extensiones radicales de Galois. Ante todo, notemos que ciertamente toda extensión radical es finita.

Teorema 7.23 *Sea k un cuerpo de característica 0 y K/k una extensión radical. Sea N la clausura normal de K sobre k. Entonces la extensión N/k es radical (y de Galois).*

Demostración: Sea $K = k(a_1, \ldots, a_n)$ según la definición 7.22. Llamemos $p_i(x) = \text{pol mín}(a_i, k)$. Los polinomios $p_i(x)$ se escinden en N y N es la adjunción a k de sus raíces (pues esta extensión es normal y contiene a K, y N es la mínima extensión que cumple esto). Si v es una raíz en N de $p_i(x)$, entonces

a_i y v son conjugados, luego existe un $\sigma \in G(N/k)$ tal que $\sigma(a_i) = v$. Entonces $\sigma[K]$ es un cuerpo k-isomorfo a K que contiene a v. De aquí se sigue que existen cuerpos $K_1, \ldots, K_r$ todos ellos k-isomorfos a K y tales que $N = K_1 \cdots K_r$.

Es obvio que las extensiones K_i/k son todas radicales, luego existen elementos $a_{i1}, \ldots, a_{in}$ de manera que $K_i = k(a_{i1}, \ldots, a_{in})$ y se cumpla la definición de extensión radical. Es fácil ver que tomando $N = k(a_{11}, \ldots, a_{1n}, \ldots, a_{r1}, \ldots, a_{rn})$ se cumple la definición de extensión radical. ∎

Por lo tanto una ecuación es resoluble por radicales si y sólo si su cuerpo de escisión está contenido en una extensión radical de Galois.

Si $K = k(a_1, \ldots, a_n)$ es una extensión radical, podemos considerar la cadena de cuerpos intermedios $K_i = k(a_1, \ldots, a_i)$, de manera que

$$k = K_0 \subset K_1 \subset K_2 \subset \cdots \subset K_n = K,$$

y cada extensión intermedia es de la forma

$$K_i = K_{i-1}(a_i), \qquad \text{con } a_i^{r_i} \in K_{i-1}. \tag{7.22}$$

Si además la extensión es de Galois podemos considerar los grupos asociados $G_i = G(K/K_i)$, que forman una cadena

$$1 = G_n \leq G_{n-1} \leq \cdots \leq G_1 \leq G(K/k). \tag{7.23}$$

Ahora la clave es que vamos a probar que, salvo una restricción técnica, la condición (7.22) equivale a que

$$G_i \trianglelefteq G_{i-1} \qquad \text{y} \qquad G_{i-1}/G_i \text{ es cíclico.} \tag{7.24}$$

Teorema 7.24 *Sea K/k una extensión de cuerpos, n un número natural no nulo y $\omega \in k$ una raíz n-sima primitiva de la unidad. Las afirmaciones siguientes son equivalentes:*

1. *K/k es una extensión cíclica de grado $d \mid n$.*
2. *K es el cuerpo de escisión sobre k de un polinomio de la forma $x^d - a$ irreducible en $k[x]$, y si u es una raíz en K de dicho polinomio, entonces $K = k(u)$.*
3. *$K = k(u)$ para un cierto u tal que $u^n \in k$.*

Demostración: 1) ⇒ 2) Sea $G(K/k) = \langle\sigma\rangle$ y $\eta = \omega^{n/d} \in k$, raíz d-ésima primitiva de la unidad. Por el teorema 5.42 existe un $w \in K$ tal que

$$v = \sum_{i=0}^{d-1} \eta^i \sigma^i(w) \neq 0.$$

Entonces $\eta\sigma(v) = \eta \sum_{i=0}^{d-1} \eta^i \sigma^{i+1}(w) = \sum_{i=0}^{d-1} \eta^{i+1} \sigma^{i+1}(w) = v$. Si $u = v^{-1}$ tenemos que $\sigma(u) = \eta u$.

En general, $\sigma^i(u) = \eta^i u$, luego los elementos $\eta^i u$ para $i = 0, \dots, d-1$ son k-conjugados. Además $\sigma(u^d) = \sigma(u)^d = \eta^d u^d = u^d$, luego $a = u^d$ es fijado por el grupo de Galois y está, por consiguiente, en k. El polinomio $x^d - a \in k[x]$ tiene por raíces a todos los $\eta^i u$ para $i = 0, \dots, d-1$, luego son todas sus raíces, es decir, se escinde en $K[x]$. Además, como son k-conjugadas, $x^d - a$ es irreducible en $k[x]$. Si adjuntamos a k cualquiera de las raíces de $x^d - a$ obtenemos una extensión de grado d, pero $|K : k| = d$, luego obtenemos K.

2) $\Rightarrow$ 3) Sea u una raíz de $x^d - a$ en K. Por hipótesis $K = k(u)$. Además $u^n = (u^d)^{n/d} = a^{n/d} \in k$.

3) $\Rightarrow$ 1) Sea $b = u^n$. Es claro que el polinomio $p(x) = x^n - b \in k[x]$ tiene n raíces distintas en K, a saber: $\omega^i u$, para $i = 1, \dots, n$. Por lo tanto $p(x)$ se escinde en K y sus raíces son separables. Concluimos que K/k es una extensión finita de Galois.

Para cada $\sigma \in G(K/k)$, el elemento $\sigma(u)$ ha de ser otra raíz de $p(x)$, luego existe un entero i determinado módulo n tal que $\sigma(u) = \omega^i u$. Consideremos la aplicación $f : G(K/k) \longrightarrow \mathbb{Z}/n\mathbb{Z}$ dada por $f(\sigma) = [i]$.

Se trata de un homomorfismo de grupos, pues si $\sigma(u) = \omega^i v$ y $\tau(u) = \omega^j v$, entonces $\tau\big(\sigma(u)\big) = \omega^j \omega^i(u)$, luego $f(\sigma\tau) = f(\sigma) + f(\tau)$.

También es fácil ver que se trata de un monomorfismo, luego $G(K/k)$ es isomorfo a un subgrupo de $\mathbb{Z}/n\mathbb{Z}$, luego cíclico de orden divisor de n. ∎

Observemos que si en 2) queremos obtener el u que cumple $a = u^d \in k$ y $K = k(u)$, simplemente hemos de buscar un u que cumpla $\sigma(u) = \eta u$, donde σ es un generador del grupo de Galois y η una raíz d-ésima primitiva de la unidad en k.

Aquí es donde la teoría entronca con la teoría de grupos resolubles que hemos presentado en la sección [TG 4.1]: salvo la restricción técnica de que se requiere la presencia de una raíz de la unidad, las cadenas de cuerpos de las extensiones radicales de Galois se corresponden con series cíclicas del grupo de Galois, de modo que una extensión finita de Galois es radical si y sólo si su grupo de Galois es resoluble. En realidad esto no es inmediato por culpa de la condición sobre las raíces de la unidad, pero vamos a ver que las propiedades de los grupos resolubles permiten tenerla en cuenta sin que se restrinja el alcance de la equivalencia:

Teorema 7.25 *Sea k un cuerpo de característica 0, sea K/k una extensión radical y L un cuerpo tal que $k \subset L \subset K$ con L/k de Galois. Entonces $G(L/k)$ es resoluble.*

Demostración: Por el teorema 7.23 podemos suponer que K/k es de Galois. Por el teorema de Galois, $G(L/k) \cong G(K/k)/G(K/L)$, luego basta probar que $G(K/k)$ es resoluble.

Sean $K = k(a_1, \dots, a_n)$ y $r_1, \dots, r_n$ según la definición 7.22.

Sea $K_0 = k$ y para cada $i = 0, \dots, n-1$, sea $K_{i+1} = K_i(a_{i+1})$. De este modo tenemos

$$k = K_0 \subset K_1 \subset \cdots \subset K_n = K.$$

Sea $m = r_1 \cdots r_n$ y sea ω una raíz m-sima primitiva de la unidad en una extensión de K. Sea $L_i = K_i(\omega)$. Entonces

$$k(\omega) = L_0 \subset L_1 \subset \cdots \subset L_n = K(\omega)$$

y $L_{i+1} = L_i(a_{i+1})$.

Ahora, $L_n/L_0 = K(\omega)/k(\omega) = Kk(\omega)/k(\omega)$. Por el teorema 5.45 la extensión L_n/L_0 es finita de Galois y $G(L_n/L_0) \cong G\Big(K/\big(K \cap k(\omega)\big)\Big)$.

Por el teorema 7.24 la extensión L_i/L_{i-1} es finita de Galois y el grupo $G(L_i/L_{i-1})$ es cíclico. Sea $H_i = G(L_n/L_i)$. Entonces

$$1 = H_n \trianglelefteq H_{n-1} \trianglelefteq \cdots \trianglelefteq H_1 \trianglelefteq H_0 = G(L_n/L_0).$$

y los factores

$$H_{i-1}/H_i = G(L_n/L_{i-1})\Big/G(L_n/L_i) \cong G(L_i/L_{i-1})$$

son cíclicos. Esto prueba que el grupo $G(L_n/L_0)$ es resoluble y por consiguiente $G\big(K/(K \cap k(\omega))\big)$ también lo es.

La situación es $k \subset K \cap k(\omega) \subset k(\omega)$. La extensión $k(\omega)/k$ es ciclotómica, luego abeliana, por lo que $\big(K \cap k(\omega)\big)/k$ es una extensión finita de Galois y

$$G\big(\big(K \cap k(\omega)\big)/k\big) \cong G\big(k(\omega)/k\big)\Big/G\Big(k(\omega)/\big(K \cap k(\omega)\big)\Big)$$

es un grupo abeliano, luego resoluble.

Finalmente consideramos $k \subset K \cap k(\omega) \subset K$. Se cumple que

$$G(K/k)\Big/G\Big(K/\big(K \cap k(\omega)\big)\Big) \cong G\Big(\big(K \cap k(\omega)\big)/k\Big)$$

y tanto $G\Big(K/\big(K \cap k(\omega)\big)\Big)$ como $G\Big(\big(K \cap k(\omega)\big)/k\Big)$ son resolubles, por lo que $G(K/k)$ es resoluble. ∎

Ahora vamos a probar el recíproco. Notemos que al igual que ocurre con el teorema anterior, la prueba se complica por la necesidad de incorporar una raíz primitiva.

Teorema 7.26 *Sea k un cuerpo de característica 0 y K/k una extensión finita de Galois tal que $G(K/k)$ sea resoluble. Entonces existe una extensión radical de k que contiene a K.*

Demostración: Sea n el grado de la extensión K/k y sea ω una raíz n-sima primitiva de la unidad en una extensión de K. Como en el teorema anterior se prueba que la extensión $K(\omega)/k(\omega)$ es finita de Galois y

$$G = G\big(K(\omega)/k(\omega)\big) \cong G\Big(\big(K/\big(K \cap k(\omega)\big)\Big) \leq G(K/k)$$

es resoluble.

En consecuencia existe una serie cíclica

$$1 = G_m \trianglelefteq G_{m-1} \trianglelefteq \cdots \trianglelefteq G_1 \trianglelefteq G_0 = G.$$

Sea F_i el cuerpo fijado por G_i. Entonces $G_i = G\big(K(\omega)/F_i\big)$ y

$$k(\omega) = F_0 \subset F_1 \subset \cdots \subset F_m = K(\omega).$$

Para cada i tenemos $F_{i-1} \subset F_i \subset K(\omega)$, y como $G_{i-1} \trianglelefteq G_i$, también F_i/F_{i-1} es de Galois y $G(F_i/F_{i-1}) \cong G_i/G_{i-1}$ es cíclico. Sea

$$r_i = \big|F_i : F_{i-1}\big| \;\Big|\; \big|K(\omega) : k(\omega)\big| = \big|G\big(K(\omega)/k(\omega)\big)\big| \;\Big|\; \big|G(K/k)\big| = n.$$

Podemos aplicar el teorema 7.24, que nos da que $F_i = F_{i-1}(a_i)$, donde a_i es una raíz de un polinomio $x^{r_i} - b_i \in F_{i-1}[x]$. Así

$$k \subset k(\omega) \subset k(\omega, a_1) \subset \cdots \subset k(\omega, a_1, \ldots, a_m) = K(\omega)$$

y claramente $K(\omega)/k$ resulta ser una extensión radical que contiene a K. ∎

El teorema siguiente es consecuencia inmediata de los dos anteriores:

Teorema 7.27 (Galois) *Sea k un cuerpo de característica 0 y $p(x)$ un polinomio no constante con coeficientes en k. La ecuación $p(x) = 0$ es resoluble por radicales si y sólo si el grupo de Galois de $p(x)$ sobre k es resoluble.*

Ejemplo La ecuación $x^5 - 4x + 2 = 0$ no es resoluble por radicales, pues hemos visto que el grupo de Galois del polinomio que la define es Σ_5. ∎

Por otro lado, existen ecuaciones de grados mayores que 4 que sí que son resolubles por radicales. El caso más obvio es el de las de la forma $x^n - a = 0$, y un ejemplo menos obvio es la ecuación

$$x^8 - 72x^6 + 180x^4 - 144x^2 + 36 = 0,$$

cuyo grupo de Galois es, según hemos visto, Q_8.

Puede probarse que, para todo natural n, existen polinomios en $\mathbb{Q}[x]$ de grado n cuyo grupo de Galois es Σ_n lo que implica que no pueden existir resultados análogos a las fórmulas de Cardano y Ferrari para polinomios de grado mayor que 4. La prueba no es sencilla, pero hay un camino más simple para llegar a la misma conclusión. Necesitamos una definición:

Definición 7.28 Sea $n \geq 1$, k un cuerpo y $K = k(a_0, \ldots, a_{n-1})$ el cuerpo de las fracciones algebraicas en las indeterminadas $a_0, \ldots, a_{n-1}$. Llamaremos *polinomio general* de grado n sobre k al polinomio

$$p_n(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \in K[x].$$

La ecuación $p_n(x) = 0$ se llama *ecuación general* de grado n.

De este modo transformamos el concepto lógico de variables arbitrarias $a_0, \ldots, a_{n-1}$ en el concepto algebraico de indeterminadas de un cuerpo de fracciones algebraicas. A efectos prácticos es equivalente. Por ejemplo, la fórmula

$$x = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_0}}{2}$$

se interpreta ahora como la expresión de las raíces de $p_2(x)$ en una extensión de K. En general, una expresión para la solución de la ecuación general de grado n sobre un cuerpo k es una fórmula para resolver todas las ecuaciones particulares de grado n con coeficientes en k (de aquí el nombre de ecuación general).

Por lo tanto, si probamos que la ecuación general de grado n no es resoluble por radicales para $n \geq 5$, habremos probado que no existen teoremas similares a los de Cardano y Ferrari para grados superiores a 4. Para ello es suficiente demostrar que el grupo de Galois del polinomio general de grado n es Σ_n, pues hemos visto que el grupo Σ_n no es resoluble para $n \geq 5$.

Teorema 7.29 *Sea k un cuerpo de característica 0. Entonces el grupo de Galois del polinomio general de grado n sobre k es isomorfo a Σ_n.*

Demostración: Sean $a_0, \ldots, a_{n-1}$ los coeficientes de $p_n(x)$ y $u_1, \ldots, u_n$ las raíces de $p_n(x)$ en una extensión de $k(a_0, \ldots, a_{n-1})$. Entonces,

$$F = k(a_0, \ldots, a_{n-1}, u_1, \ldots, u_n)$$

es un cuerpo de escisión de $p_n(x)$ sobre $k(a_0, \ldots, a_{n-1})$.

Tenemos que $p_n(x) = (x-u_1)\cdots(x-u_n)$, luego por el teorema 7.4 se cumple que $a_k = (-1)^{n-k} e_{n-k}(u_1, \ldots, u_n)$ para $k = 0, \ldots, n-1$.

Consideremos la aplicación

$$\begin{array}{rccc} \phi: & k[a_0, \ldots, a_{n-1}] & \longrightarrow & k[e_1, \ldots, e_n] \\ & h(a_0, \ldots, a_{n-1}) & \mapsto & h\big((-1)^n e_n, \ldots, -e_1\big) \end{array}$$

Claramente se trata de un epimorfismo de anillos.

Además si $\phi\big(h(a_0, \ldots, a_{n-1})\big) = 0$, entonces $h\big((-1)^n e_n, \ldots, -e_1\big) = 0$ y en particular

$$h\big((-1)^n e_n(u_1, \ldots, u_n), \ldots, -e_1(u_1, \ldots, u_n)\big) = 0,$$

o sea, $h(a_0, \ldots, a_{n-1}) = 0$, lo que prueba que se trata de un isomorfismo de anillos, que se extiende a un isomorfismo entre los cuerpos $k(a_0, \ldots, a_{n-1})$ y $k(e_1, \ldots, e_n)$ y que, a su vez, se extiende a un isomorfismo entre sus respectivos anillos de polinomios en la indeterminada x.

La imagen por este isomorfismo del polinomio general $p_n(x)$ es

$$x^n - e_1 x^{n-1} + \cdots + (-1)^n e_n = (x - x_1)\cdots(x - x_n).$$

Como F es el cuerpo de escisión sobre $k(a_0,\dots,a_{n-1})$ de $p_n(x)$ y $k(x_1,\dots,x_n)$ es el cuerpo de escisión sobre $k(e_1,\dots,e_n)$ de $(x-x_1)\cdots(x-x_n)$, es claro que los respectivos grupos de Galois deben ser isomorfos, es decir,

$$G\big(F/k(a_0,\dots,a_{n-1})\big) \cong G\big(k(x_1,\dots,x_n)/k(e_1,\dots,e_n)\big).$$

El primero es el grupo de Galois de $p_n(x)$ y el segundo es isomorfo a Σ_n por el teorema 7.6. ∎

Como consecuencia inmediata tenemos:

Teorema 7.30 (Abel) *La ecuación general de grado n no es resoluble por radicales para $n \geq 5$.*

Resolución de cúbicas mediante raíces reales Hemos visto que, cuando una cúbica tiene tres raíces reales simples, la fórmula de Cardano las expresa como suma de dos raíces cúbicas imaginarias conjugadas. Vamos a probar que, en general, no es posible expresar las soluciones de la ecuación en términos de raíces reales (es decir, con radicando positivo cuando el índice es par). Concretamente, vamos a demostrar el teorema siguiente:

Teorema 7.31 *Sea K un subcuerpo del cuerpo de los números reales, y consideremos una cúbica con coeficientes en K que cumpla $D<0$ y que no tenga raíces en K. Entonces no es posible expresar (ninguna de) sus raíces en función de sus coeficientes mediante sumas, productos, cocientes y raíces que sean números reales.*

Demostración: Observemos que un número real α se puede expresar en función de elementos de K mediante sumas, productos, cocientes y raíces reales si y sólo si existe una cadena de cuerpos

$$K = K_0 \subset K_1 \subset \cdots \subset K_n \subset \mathbb{R}$$

tales que $\alpha \in K_n$ y cada cuerpo intermedio es de la forma $K_i = K_{i-1}(\alpha_i)$, para un cierto $\alpha_i \in K_i$ tal que existe un número natural $m_i \geq 2$ tal que $\alpha_i^{m_i} \in K_{i-1}$.

En general, pongamos que la ecuación es la definida por el polinomio

$$f(x) = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3),$$

donde $\alpha_i \notin K$, y supongamos que existe una cadena de cuerpos

$$K = K_0 \subset K_1 \subset \cdots \subset K_n \subset \mathbb{R}$$

en las condiciones anteriores y tal que un $\alpha_i \in K_n$. Vamos a llegar a una contradicción. Sea $L = K(\alpha_1,\alpha_2,\alpha_3)$ el cuerpo de escisión de $f(x)$ sobre K. La hipótesis $D<0$ equivale a que el discriminante de $f(x)$ cumple $\Delta > 0$ y sabemos que $\sqrt{\Delta} \in L$. Si $\sqrt{\Delta} \notin K$, como $K(\sqrt{\Delta})$ tiene grado 2 sobre K, no

contiene a los α_i, luego f es también una cúbica sobre $K(\sqrt{\Delta})$ sin raíces en este cuerpo. Además, la cadena de cuerpos

$$K = K_0(\sqrt{\Delta}) \subset K_1(\sqrt{\Delta}) \subset \cdots \subset K_n(\sqrt{\Delta}) \subset \mathbb{R}$$

cumple las mismas condiciones, luego podemos sustituir K por $K(\sqrt{\Delta})$ y suponer que $\sqrt{\Delta} \in K$. El teorema 7.13 nos da entonces que $|L : K| = 3$, luego $L = K(\alpha_1) = K(\alpha_2) = K(\alpha_3)$ y así, $\alpha_i \in K_n$ implica que $L \subset K_n$.

También podemos suponer que n es el mínimo natural tal que K_n contiene a un α_i (o, equivalentemente, a todos los α_i), con lo que $f(x)$ no tiene raíces en K_{n-1} y, cambiando K por K_{n-1}, podemos suponer que $n = 1$.

En definitiva, tenemos que $K \subset L \subset K(\alpha)$, donde $\alpha^m \in K$, para cierto $m \geq 2$. Podemos suponer que el número natural m es el mínimo posible que cumple esto. Sea p un divisor primo de m. Entonces $K(\alpha^{m/p}) \subsetneq K(\alpha)$, luego, por la minimalidad de m, tenemos que $f(x)$ no tiene raíces en este subcuerpo y, cambiando K por $K(\alpha^{m/p})$, podemos suponer que $\alpha^p \in K$.

Sea ω una raíz p-ésima primitiva de la unidad. Entonces $\alpha \notin K(\omega)$, porque la extensión $K(\omega)/K$ es abeliana (teorema 5.84), luego, si $K(\alpha) \subset K(\omega)$, tendríamos que $K(\alpha)/K$ sería de Galois, lo cual es absurdo, porque los conjugados de α son los $\omega^j\alpha$ y son imaginarios.

Por consiguiente, la extensión $K(\omega, \alpha)/K(\omega)$ no es trivial y, de hecho, tiene grado p por el teorema 7.24. Así pues, p divide al grado

$$|K(\omega, \alpha) : K| = |K(\alpha, \omega) : K(\alpha)||K(\alpha) : K|,$$

pero el primer factor tiene grado $\leq p - 1$, luego $p \mid |K(\alpha) : K|$ y así concluimos que $|K(\alpha) : K| = p$.

Como $3 = |L : K| \mid |K(\alpha) : K|$, de hecho, ha de ser $p = 3$ y $K(\alpha) = L$, pero esto es absurdo, porque L/K es una extensión de Galois y $K(\alpha)/K$ no lo es. ∎

Ejemplo Consideremos una raíz séptima de la unidad:

$$\omega = \cos(2\pi/7) + i\,\mathrm{sen}(2\pi/7)$$

y sean $\eta_1 = \omega + \omega^6$, $\eta_2 = \omega^2 + \omega^5$, $\eta_3 = \omega^3 + \omega^4$. Notemos que $\eta_1 = 2\cos(2\pi/7)$. Teniendo en cuenta que $1 + \omega + \omega^2 + \omega^3 + \omega^4 + \omega^5 + \omega^6 = 0$, es fácil ver que

$$\eta_1 + \eta_2 + \eta_3 = -1, \quad \eta_1\eta_2 + \eta_1\eta_3 + \eta_2\eta_3 = -2, \quad \eta_1\eta_2\eta_3 = 1.$$

Por consiguiente, η_1, η_2, η_3 son las raíces de la ecuación $u^3 + u^2 - 2u - 1 = 0$. Haciendo $u = 2x$ concluimos que $\cos(2\pi/7)$ es una de las raíces de la ecuación

$$x^3 + \frac{1}{2}x^2 - \frac{1}{2}x - \frac{1}{8} = 0.$$

La fórmula de Cardano nos da que

$$\cos\frac{2\pi}{7} = \sqrt[3]{\frac{7}{144}\left(\frac{1}{3} + \sqrt{3}\,i\right)} + \sqrt[3]{\frac{7}{144}\left(\frac{1}{3} - \sqrt{3}\,i\right)} - \frac{1}{6},$$

para cierta elección de las raíces cúbicas. (Las otras dos elecciones nos dan $\cos(4\pi/7)$ y $\cos(8\pi/7)$.) Ahora bien, el teorema anterior nos asegura que esta expresión es "esencialmente imaginaria", en el sentido de que es imposible obtener el resultado

$$\cos\frac{2\pi}{7} = 0.6234898019\ldots$$

con una calculadora pulsando únicamente las teclas numéricas y las de las operaciones de suma, resta, producto, división y extracción de raíces. ∎

Capítulo VIII

Anillos de enteros algebraicos

En [ITAl] tuvimos ocasión de comprobar que muchos problemas que, en principio, involucran únicamente números enteros pueden resolverse trabajando en anillos de enteros algebraicos adecuados, como $\mathbb{Z}\big[\sqrt{2}\,\big]$, etc. Vimos que estos anillos no siempre tienen factorización única, pero en el caso de los anillos de enteros de los cuerpos cuadráticos, en [ITAl 13.15] probamos que todos ellos tienen una "factorización única ideal". Ahora disponemos del aparato algebraico necesario para ir más allá de los enteros cuadráticos y obtener resultados generales aplicables a cualquier anillo de enteros algebraicos.

8.1 La forma bilineal asociada a la traza

Antes de entrar en lo que va a ser propiamente el objeto de estudio de este capítulo necesitamos introducir un nuevo concepto que combina elementos de la teoría de cuerpos y del álgebra lineal:

Teorema 8.1 *Sea K/k una extensión de cuerpos finita separable. Entonces la traza determina una forma bilineal simétrica regular dada por*

$$\begin{array}{ccc} K\times K & \longrightarrow & k \\ (\alpha,\beta) & \mapsto & \mathrm{Tr}(\alpha\beta) \end{array}$$

Demostración: Es claro que la aplicación así definida es una forma bilineal simétrica. Calculemos su matriz en una base cualquiera $B=(v_1,\dots,v_n)$ de K. Sean $\sigma_1,\dots,\sigma_n$ los k-monomorfismos de K. Entonces

$$\begin{aligned} \mathrm{M}_B &= (\mathrm{Tr}(v_iv_j)) = \left(\sum_{k=1}^n \sigma_k(v_iv_j)\right) = \left(\sum_{k=1}^n \sigma_k(v_i)\sigma_k(v_j)\right) \\ &= \big(\sigma_k(v_i)\big)_{ik}\big(\sigma_k(v_j)\big)_{kj} = \big(\sigma_i(v_j)\big)\big(\sigma_i(v_j)\big)^t. \end{aligned}$$

Por lo tanto $|\mathrm{M}_B| = \big|\sigma_i(v_j)\big|^2$. Basta comprobar que este determinante es no nulo para una base en particular. Concretamente, por el teorema del elemento

primitivo sabemos que $K = k(\alpha)$ para cierto $\alpha \in K$, y una k-base de K es $B = (1, \alpha, \ldots, \alpha^{n-1})$. Para esta base, el determinante que hemos de calcular es

$$\begin{vmatrix} 1 & \cdots & 1 \\ \sigma_1(\alpha) & \cdots & \sigma_n(\alpha) \\ \vdots & & \vdots \\ \sigma_1(\alpha)^{n-1} & \cdots & \sigma_n(\alpha)^{n-1} \end{vmatrix},$$

y este determinante es de Vandermonde, y su segunda fila la forman los n conjugados de α, que son todos distintos (pues α es separable y su polinomio mínimo tiene grado n). Por lo tanto $|\mathrm{M}_B| \neq 0$ y la forma es regular. ∎

En particular hemos probado que la traza de una extensión finita separable es siempre no nula (lo que en característica prima no es trivial).

Ejercicio: Considerar el cuerpo $\mathbb{Q}\left(\sqrt{3}\right)$. Calcular la base dual respecto a la traza de la base $\left(1, \sqrt{3}\right)$.

Conviene extraer algunas ideas de la demostración del teorema anterior:

Definición 8.2 Sea K/k una extensión finita separable y $B = (v_1, \ldots, v_n)$ una k-base de K. Llamaremos *discriminante* de B al determinante de la matriz de la forma bilineal asociada a la traza respecto a la base B. Equivalentemente:

$$\Delta[B] = \Delta[v_1, \ldots, v_n] = \left|\mathrm{Tr}(v_i v_j)\right| = \left|\sigma_i(v_j)\right|^2,$$

donde $\sigma_1, \ldots, \sigma_n$ son los k-monomorfismos de K.

Notemos que la segunda expresión implica que $\Delta[B]$ no depende del orden de los elementos de la base B (ni por supuesto del de los monomorfismos), pues una alteración de dicho orden se traduce en una permutación de las filas o columnas del determinante, lo que a lo sumo implica un cambio de signo que a su vez es absorbido por el cuadrado.

Hemos probado que $\Delta[B] \in k$ es no nulo y si α es un elemento primitivo de la extensión, entonces

$$\Delta\big[1, \alpha, \ldots \alpha^{n-1}\big] = \prod_{i<j} \big(\sigma_j(\alpha) - \sigma_i(\alpha)\big)^2.$$

Hay una última propiedad de los discriminantes que conviene observar:

Teorema 8.3 *Sea K/k una extensión de cuerpos finita separable y sean B, C dos k-bases de K. Entonces $\Delta[B] = \left|\mathrm{M}_B^C\right|^2 \Delta[C]$.*

Demostración: Basta probar que las matrices de la forma bilineal asociada a traza guardan la relación

$$M_B = \mathrm{M}_B^C \, M_C \, \left(\mathrm{M}_B^C\right)^t. \tag{8.1}$$

Ahora bien, para todo $v, w \in K$,

$$\Phi_B(v)\, \mathrm{M}_B^C \, M_C \, \left(\mathrm{M}_B^C\right)^t \, \Phi_B(w)^t = \Phi_C(v)\, M_C\, \Phi_C(w)^t = \mathrm{Tr}(vw),$$

luego por la unicidad de la matriz de una forma bilineal, se cumple (8.1). ∎

8.2 Enteros algebraicos

En la sección [ITAl 8.4] introdujimos el concepto de entero algebraico, pero ahora podemos estudiar más cómodamente los enteros algebraicos usando la teoría de Galois y la teoría de módulos en vez de los resultados sobre polinomios simétricos en los que nos apoyamos allí. En primer lugar conviene introducir el concepto de cuerpo numérico:

Definición 8.4 *Un cuerpo numérico es una extensión finita de* $\mathbb{Q}$.

En general, siempre que apliquemos a un cuerpo numérico K conceptos de la teoría de extensiones de cuerpos se entenderá que se refieren a la extensión $K/\mathbb{Q}$. Por ejemplo, el grado de un cuerpo numérico será el grado sobre $\mathbb{Q}$, un cuerpo numérico cíclico o abeliano será una extensión finita de Galois de $\mathbb{Q}$ cuyo grupo de Galois es cíclico o abeliano, etc.

Un poco más en general, consideremos de momento un cuerpo K de característica 0, que, por lo tanto, contiene al cuerpo $\mathbb{Q}$ de los números racionales. Un elemento $a \in K$ es algebraico sobre $\mathbb{Q}$ (o, simplemente, algebraico) si y sólo si es la raíz de un polinomio no nulo con coeficientes racionales, pero multiplicando dicho polinomio, en caso de que exista, por el producto de los denominadores de sus coeficientes no nulos obtenemos un polinomio con coeficientes enteros con las mismas raíces. Así pues un elemento de K es algebraico si y sólo si es la raíz de un polinomio con coeficientes enteros. El concepto de entero algebraico surge imponiendo una restricción:

Definición 8.5 Sea K un cuerpo de característica 0. Un elemento $a \in K$ es un *entero algebraico* si es la raíz de un polinomio mónico con coeficientes enteros.

Como los enteros algebraicos son en particular números algebraicos, podemos limitarnos a estudiar los enteros algebraicos del cuerpo $\mathbb{A}$ (la clausura algebraica de $\mathbb{Q}$). Llamaremos $\mathbb{E}$ al conjunto de los enteros algebraicos de $\mathbb{A}$. Si K es un cuerpo numérico llamaremos $\mathcal{O}_K$ al conjunto de los enteros algebraicos de K (La O hace referencia a 'orden', aunque aquí no introduciremos este concepto en general). Claramente tenemos que $\mathcal{O}_K = K \cap \mathbb{E}$.

La caracterización siguiente muestra entre otras cosas que no todos los números algebraicos son enteros algebraicos.

Teorema 8.6 *Un elemento algebraico a de una extensión de $\mathbb{Q}$ es un entero algebraico si y sólo si* $\operatorname{pol\,mín}(a, \mathbb{Q}) \in \mathbb{Z}[x]$.

Demostración: Una implicación es obvia. Supongamos que a es un entero algebraico y sea $p(x) \in \mathbb{Z}[x]$ un polinomio mónico tal que $p(a) = 0$. Sea $q(x)$ un factor irreducible de $p(x)$ en $\mathbb{Z}[x]$ tal que $q(a) = 0$. Existe un polinomio $r(x) \in \mathbb{Z}[x]$ tal que $p(x) = q(x)r(x)$. Como el producto de los coeficientes directores de $q(x)$ y $r(x)$ debe ser igual al coeficiente director de $p(x)$ que es 1, el coeficiente director de $q(x)$ debe ser ± 1. Podemos exigir que sea 1 y así $q(x)$ es un polinomio mónico irreducible en $\mathbb{Z}[x]$ del que a es raíz. Por el criterio de Gauss, $q(x)$ también es irreducible en $\mathbb{Q}[x]$, luego $q(x) = \operatorname{pol\,mín}(a, \mathbb{Q}) \in \mathbb{Z}[x]$. ∎

Otra consecuencia de este teorema es que los enteros algebraicos de $\mathbb{Q}$ son precisamente los números enteros:

Teorema 8.7 $\mathcal{O}_{\mathbb{Q}} = \mathbb{Z}$.

DEMOSTRACIÓN: Un número racional q es un entero algebraico si y sólo si $\operatorname{pol\,mín}(q, \mathbb{Q}) \in \mathbb{Z}[x]$, si y sólo si $x - q \in \mathbb{Z}[x]$, si y sólo si $q \in \mathbb{Z}$. ∎

El lector debe notar el paralelismo entre el teorema siguiente y resultados similares sobre cuerpos.

Teorema 8.8 *Sea K un cuerpo de característica 0. Un elemento $c \in K$ es un entero algebraico si y sólo si $\mathbb{Z}[c] = \{q(c) \mid q(x) \in \mathbb{Z}[x]\}$ es un $\mathbb{Z}$-módulo finitamente generado. En tal caso es libre de rango $|\mathbb{Q}(c) : \mathbb{Q}|$.*

DEMOSTRACIÓN: Supongamos que c es un entero algebraico. Entonces su polinomio mínimo $p(x)$ tiene coeficientes enteros y su grado es $n = |\mathbb{Q}(c) : \mathbb{Q}|$. Veamos que

$$\mathbb{Z}[c] = \langle c^m \mid m = 0, \ldots, n-1 \rangle . \tag{8.2}$$

Un elemento arbitrario de $\mathbb{Z}[c]$ es de la forma $q(c)$, donde $q(x)$ es un polinomio con coeficientes enteros. Dividimos $q(x) = p(x)u(x) + r(x)$, donde u y r tienen ambos coeficientes enteros y el grado de r es menor que n. Entonces resulta que $q(c) = r(c)$, luego pertenece al miembro derecho de (8.2), y la otra inclusión es obvia. De hecho el generador $(1, c, \ldots, c^{n-1})$ es una base, pues una combinación lineal nula es de la forma $r(c) = 0$, con $r(x) \in \mathbb{Z}[x]$ de grado menor que n, luego concluimos que $r = 0$.

Supongamos ahora que $\mathbb{Z}[c]$ es finitamente generado. Digamos que admite n generadores $v_1, \ldots, v_n$. Cada v_i es un polinomio en c con coeficientes enteros. Sea m mayor que el grado de cualquiera de dichos polinomios.

Entonces c^m se expresa como combinación lineal con coeficientes enteros de los v_i, luego en definitiva $c^m = q(c)$, con $q(x) \in \mathbb{Z}[x]$ de grado menor que m. La ecuación $c^m - q(c) = 0$ justifica que c es un entero algebraico. ∎

De aquí podemos deducir las propiedades básicas de los enteros algebraicos. En primer lugar probamos que forman un anillo.

Teorema 8.9 *El conjunto $\mathbb{E}$ es un subanillo de $\mathbb{A}$. Si K es un cuerpo numérico entonces $\mathcal{O}_K$ es un subanillo de K.*

DEMOSTRACIÓN: Sean c, $d \in \mathbb{E}$. Hay que probar que $c + d$ y cd están en $\mathbb{E}$. Sea $\{v_1, \ldots, v_n\}$ un generador de $\mathbb{Z}[c]$ y sea $\{w_1, \ldots, w_m\}$ un generador de $\mathbb{Z}[d]$. Sea M el $\mathbb{Z}$-módulo generado por los todos los productos $v_i w_j$.

Todo c^r se expresa como combinación lineal con coeficientes enteros de los v_i y todo d^s se expresa como combinación lineal con coeficientes enteros de los w_j. Al multiplicar estas expresiones obtenemos una expresión de $c^r d^s$ como combinación lineal con coeficientes enteros de los generadores de M, luego cada $c^r d^s \in M$.

En particular, $\mathbb{Z}[cd] \subset M$, luego es un $\mathbb{Z}$-módulo finitamente generado (teorema 4.42). Por el teorema anterior $cd \in \mathbb{E}$.

Al desarrollar $(c+d)^k$ obtenemos una combinación lineal con coeficientes enteros de elementos de la forma $c^r d^s$, que están en M, luego $\mathbb{Z}[c+d] \subset M$ y también se cumple que $c+d \in \mathbb{E}$.

Obviamente $\mathcal{O}_K = \mathbb{E} \cap K$ es un subanillo de K. ∎

Es costumbre referirse a los elementos del anillo $\mathcal{O}_K$ como a los *enteros* de K, es decir, reservar la palabra 'entero' para los enteros algebraicos en lugar para los enteros de $\mathbb{Z}$. Por este mismo convenio, los enteros de $\mathbb{Z} = \mathcal{O}_\mathbb{Q}$ son los *enteros racionales,* si $K = \mathbb{Q}(\omega)$ es el cuerpo ciclotómico p-ésimo, los elementos de $\mathcal{O}_K$ se llaman *enteros ciclotómicos,* etc.

La relación entre K y $\mathcal{O}_K$ es similar a la relación existente entre $\mathbb{Q}$ y $\mathbb{Z}$. Por ejemplo, igual que $\mathbb{Q}$ es el cuerpo de cocientes de $\mathbb{Z}$, se cumple que K es el cuerpo de cocientes de $\mathcal{O}_K$. Esto se deduce del resultado siguiente:

Teorema 8.10 *Para cada $c \in \mathbb{A}$ existe un entero racional no nulo m de manera que $mc \in \mathbb{E}$.*

Demostración: Sea $\operatorname{pol\,mín}(c, \mathbb{Q}) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. Sea m el producto de los denominadores de todos los coeficientes no nulos de $p(x)$.

Entonces $m^n(c^n + a_{n-1}c^{n-1} + \cdots + a_1 c + a_0) = 0$, luego

$$(mc)^n + a_{n-1}m(mc)^{n-1} + \cdots + a_1 m^{n-1}(mc) + a_0 = 0.$$

Por lo tanto, $x^n + a_{n-1}mx^{n-1} + \cdots + a_1 m^{n-1}x + a_0$ es un polinomio mónico con coeficientes enteros del cual es raíz mc. ∎

Teorema 8.11 *Si K es un cuerpo numérico, entonces K es el cuerpo de cocientes de $\mathcal{O}_K$.*

Demostración: Si $a \in K$, entonces existe un entero racional no nulo m tal que $ma \in \mathcal{O}_K$, por lo tanto $a = (ma)/m$ está en el cuerpo de cocientes de $\mathcal{O}_K$. ∎

Otra consecuencia importante del teorema 8.10 es que los elementos primitivos siempre se pueden tomar enteros:

Teorema 8.12 *Sea K un cuerpo numérico. Entonces existe un $c \in \mathcal{O}_K$ tal que $K = \mathbb{Q}(c)$.*

Demostración: Por el teorema del elemento primitivo existe un $a \in K$ tal que $K = \mathbb{Q}(a)$. Sea m un entero racional no nulo tal que $c = ma$ sea entero en K. Claramente $K = \mathbb{Q}(c)$. ∎

Un paso más de cara a reproducir para anillos de enteros los resultados que conocemos sobre cuerpos es demostrar que si K es un cuerpo numérico de grado n entonces $\mathcal{O}_K$ es un $\mathbb{Z}$-módulo libre de rango n. Para ello nos apoyaremos en los discriminantes definidos en la sección anterior. El teorema siguiente recoge un par de hechos sencillos, pero útiles, sobre ellos:

Teorema 8.13 *Sea K un cuerpo numérico de grado n y B una base de K formada por enteros. Entonces $\Delta[B] \in \mathbb{Z}$ y $\Delta[B] \equiv 0, 1 \pmod 4$.*

DEMOSTRACIÓN: Sea $B = (b_1, \ldots, b_n)$. Por hipótesis $b_1, \ldots, b_n$ son enteros, luego sus conjugados también lo son (los polinomios mínimos son los mismos), luego $\Delta[B]$ es por definición el cuadrado del determinante de una matriz de coeficientes enteros, luego es entero y además es un número racional, luego es un entero racional.

Para obtener la segunda parte consideramos los monomorfismos de K, digamos $\sigma_1, \ldots \sigma_n$, sea L la clausura normal de $K/\mathbb{Q}$ y sea ρ un automorfismo de L. Sea $A = \big(\sigma_i(b_j)\big)$.

El determinante de A es una suma de productos de la forma

$$\pm\sigma_{\tau(1)}(b_1)\cdots\sigma_{\tau(n)}(b_n),$$

donde $\tau \in \Sigma_n$. Si le aplicamos ρ obtenemos un término de la forma

$$\pm\rho\big(\sigma_{\tau(1)}(b_1)\big)\cdots\rho\big(\sigma_{\tau(n)}(b_n)\big).$$

Ahora bien, cada monomorfismo $\sigma_i\rho$ ha de ser un $\sigma_{\rho(i)}$, para cierto índice $\rho(i)$ (y ahora estamos llamando ρ a una permutación de $\{1, \ldots, n\}$ inducida por el automorfismo ρ). Por lo tanto la imagen por ρ del producto es

$$\pm\sigma_{\rho(\tau(1))}(b_1)\cdots\sigma_{\rho(\tau(n))}(b_n),$$

es decir, el sumando del determinante correspondiente a la permutación $\tau\rho$.

Si (la permutación inducida por) ρ es una permutación par entonces ρ envía sumandos con signo positivo a sumandos con signo positivo y sumandos con signo negativo a sumandos con signo negativo, mientras que si ρ es impar entonces intercambia los sumandos positivos con los negativos. En otras palabras, si llamamos respectivamente P y N a la suma de términos positivos y negativos (sin el signo) del determinante de A, tenemos que $\det A = P - N$ y o bien $\rho(P) = P$ y $\rho(N) = N$, o bien $\rho(P) = N$ y $\rho(N) = P$.

En cualquier caso $\rho(P + N) = P + N$ y $\rho(PN) = PN$, para todo automorfismo ρ, luego concluimos que $P+N$, $PN \in \mathbb{Q}$. Además son enteros algebraicos, luego están en $\mathbb{Z}$. Finalmente,

$$\Delta[B] = (P - N)^2 = (P + N)^2 - 4PN \equiv (P + N)^2 \equiv 0, 1 \pmod 4,$$

pues todo cuadrado es 0 o 1 módulo 4. ∎

Ahora podemos probar un resultado básico de estructura:

Teorema 8.14 *Sea K un cuerpo numérico de grado n. Entonces $\mathcal{O}_K$ es un $\mathbb{Z}$-módulo libre de rango n.*

DEMOSTRACIÓN: Por 8.12 sabemos que $K = \mathbb{Q}(c)$, donde $c \in \mathcal{O}_K$. Entonces $1, c, \ldots, c^{n-1}$ es una base de K formada por enteros. Podemos tomar una base de K $B = \{b_1, \ldots, b_n\}$ formada por enteros tal que el número natural $\big|\Delta[b_1, \ldots, b_n]\big|$ sea mínimo. Vamos a probar que entonces $\{b_1, \ldots, b_n\}$ es una base de $\mathcal{O}_K$ como $\mathbb{Z}$-módulo. Obviamente sus elementos son linealmente independientes sobre $\mathbb{Z}$, pues lo son sobre $\mathbb{Q}$. Basta probar que generan $\mathcal{O}_K$.

Supongamos, por el contrario, que existe un elemento $d \in \mathcal{O}_K$ que no pertenezca al submódulo generado por $\{b_1, \ldots, b_n\}$. Como en cualquier caso $\{b_1, \ldots, b_n\}$ es una base de K, se cumplirá que

$$d = a_1 b_1 + \cdots + a_n b_n, \tag{8.3}$$

para ciertos números racionales $a_1, \ldots, a_n$ no todos enteros. Podemos suponer que $a_1 \notin \mathbb{Z}$. Sea $a_1 = a + r$, donde $a \in \mathbb{Z}$ y $0 < r < 1$. Sustituyendo en (8.3) obtenemos que

$$r b_1 + a_2 b_2 + \cdots + a_n b_n = d - a b_1 \in \mathcal{O}_K.$$

Si llamamos c_1 a este elemento y $c_i = b_i$ para $i = 2, \ldots, n$ obtenemos una nueva base C de K formada por enteros tal que

$$\mathrm{M}_C^B = \begin{pmatrix} r & a_2 & a_3 & \cdots & a_n \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Claramente $\left|\mathrm{M}_C^B\right| = r$ y en consecuencia

$$\left|\Delta[C]\right| = r^2 \left|\Delta[B]\right| < \left|\Delta[B]\right|,$$

en contra de la elección de B. Por lo tanto B es una base de $\mathcal{O}_K$ como $\mathbb{Z}$-módulo. ∎

Definición 8.15 Sea K un cuerpo numérico. Una *base entera* de K es una base de $\mathcal{O}_K$ como $\mathbb{Z}$-módulo.

Como todo elemento de K es de la forma c/m, donde $c \in \mathcal{O}_K$ y $m \in \mathbb{Z}$, es inmediato que una base entera de K es un generador de K como $\mathbb{Q}$-espacio vectorial, luego es de hecho una base de K.

Así, si $\alpha_1, \ldots, \alpha_n$ es una base entera de K, tenemos que

$$\begin{aligned} K &= \{a_1\alpha_1 + \cdots + a_n\alpha_n \mid a_1, \ldots, a_n \in \mathbb{Q}\}, \\ \mathcal{O}_K &= \{a_1\alpha_1 + \cdots + a_n\alpha_n \mid a_1, \ldots, a_n \in \mathbb{Z}\}. \end{aligned}$$

En otros términos, los enteros de K son los elementos cuyas coordenadas son enteras.

Es importante tener claro que una base de un cuerpo K formada por enteros no es necesariamente una base entera. Basta pensar que si $v_1, \ldots, v_n$ es una base entera de K, entonces $2v_1, \ldots, v_n$ sigue siendo una base de K formada por enteros, pero ya no es una base entera, pues v_1 es un entero algebraico y no tiene coordenadas enteras respecto a esta segunda base.

Por ejemplo, es obvio que $1, \sqrt{5}$ forman una base de $\mathbb{Q}(\sqrt{5})$ y sus miembros son sin duda enteros algebraicos, pero para que fueran una base entera haría falta que los enteros algebraicos de $\mathbb{Q}(\sqrt{5})$ fueran exactamente los elementos de

$$\langle 1, \sqrt{5}\rangle = \{m + n\sqrt{5} \mid m, n \in \mathbb{Z}\},$$

y ya vimos en [ITAl 9.2] que esto es falso.

Determinar el anillo de enteros algebraicos de un cuerpo numérico dado es un problema, cuanto menos, laborioso. Antes de ver algunos ejemplos conviene entender un poco mejor la situación en general.

Sea B una base entera de un cuerpo numérico K y C cualquier otra base formada por enteros. Entonces cada componente de C se expresa como combinación lineal de B con coordenadas enteras, es decir, $\mathrm{M}_C^B \in \mathrm{Mat}_n(\mathbb{Z})$, luego $n = \left|\det \mathrm{M}_C^B\right|$ es un número natural no nulo y

$$\Delta[C] = n^2\Delta[B].$$

La base C será una base entera si y sólo si es una base de $\mathcal{O}_K$. Por 6.14 esto sucede si y sólo si $\det \mathrm{M}_C^B = \pm 1$, o sea, si y sólo si $n = 1$, si y sólo si $\Delta[C] = \Delta[B]$. En particular todas las bases enteras de K tienen el mismo discriminante.

Llamaremos *discriminante* de un cuerpo numérico K al discriminante de cualquier base entera de K. Lo representaremos por Δ_K.

Así pues, hemos probado que si C es cualquier base de K formada por enteros, entonces C es una base entera si y sólo si $\Delta[C] = \Delta_K$, y en general se tiene $\Delta[C] = n^2\Delta_K$, es decir, el discriminante de cualquier base formada por enteros es divisible entre el discriminante de K (y el cociente es un cuadrado). Ahora se entiende mejor por qué hemos demostrado la existencia de bases enteras tomando una con discriminante mínimo.

Una consecuencia obvia es la siguiente:

Teorema 8.16 *Sea K un cuerpo numérico y B una base de K formada por enteros y tal que $\Delta[B]$ sea libre de cuadrados. Entonces B es una base entera de K.*

Ahora ya es fácil determinar el anillo de enteros de un cuerpo cuadrático:

Enteros cuadráticos Como primer ejemplo veamos el caso de los *cuerpo cuadráticos,* es decir, los cuerpos numéricos de grado 2.

En primer lugar, si K es un cuerpo cuadrático, $K = \mathbb{Q}(\alpha)$, donde α es un entero algebraico, y por lo tanto raíz de un polinomio $x^2 + bx + c \in \mathbb{Z}[x]$. Así pues, $\alpha = \frac{-b\pm\sqrt{b^2-4c}}{2}$ y $K = \mathbb{Q}(\sqrt{b^2-4c})$. Podemos expresar $b^2 - 4c = m^2d$, donde d es libre de cuadrados, y así $K = \mathbb{Q}(m\sqrt{d}) = \mathbb{Q}(\sqrt{d})$.

Tenemos, pues, que todo cuerpo cuadrático es de la forma $K = \mathbb{Q}(\sqrt{d})$, donde d es un entero libre de cuadrados (obviamente $d \neq 1$). Como

$$\mathrm{pol\,mín}(\sqrt{d}, \mathbb{Q}) = x^2 - d = (x + \sqrt{d})(x - \sqrt{d}),$$

resulta que los elementos de K son de la forma $a + b\sqrt{d}$, donde $a, b \in \mathbb{Q}$, la extensión $K/\mathbb{Q}$ es una extensión de Galois y sus automorfismos son la identidad y el determinado por $\sigma\big(\sqrt{d}\,\big) = -\sqrt{d}$. A este automorfismo lo llamaremos simplemente *conjugación* de K, y lo representaremos por una barra horizontal, es decir,

$$\overline{a + b\sqrt{d}} = a - b\sqrt{d}.$$

En lo sucesivo, cuando hablemos de un cuerpo cuadrático $\mathbb{Q}\big(\sqrt{d}\,\big)$, sobrentenderemos que d es un entero libre de cuadrados.

A la hora de encontrar el anillo de enteros de un cuerpo numérico, el primer paso es encontrar un elemento primitivo entero, en nuestro caso tenemos $\sqrt{d}$. Esto nos da una base formada por enteros, concretamente $\{1, \sqrt{d}\}$. Calculemos su discriminante:

$$\Delta\Big[1, \sqrt{d}\,\Big] = \begin{vmatrix} 1 & 1 \\ \sqrt{d} & -\sqrt{d} \end{vmatrix}^2 = \big(-2\sqrt{d}\,\big)^2 = 4d.$$

Con esto sabemos que el discriminante de K se diferencia de $4d$ a lo sumo en un cuadrado, y como d es libre de cuadrados, sólo hay dos posibilidades, $\Delta_K = 4d$ o bien $\Delta_K = d$.

El teorema 8.13 da una condición necesaria para el segundo caso, y es que $d \equiv 1 \pmod 4$ (no puede ser $d \equiv 0 \pmod 4$ porque d es libre de cuadrados). Veamos que la condición es también suficiente. Consideremos el número

$$\alpha = \frac{1 + \sqrt{d}}{2}.$$

Es fácil calcular su polinomio mínimo, que resulta ser

$$x^2 - x + \frac{1 - d}{4}.$$

Vemos, pues, que si $d \equiv 1 \pmod 4$ entonces α es un entero algebraico y

$$\Delta[1, \alpha] = \begin{vmatrix} 1 & 1 \\ \frac{1+\sqrt{d}}{2} & \frac{1-\sqrt{d}}{2} \end{vmatrix}^2 = \big(-\sqrt{d}\,\big)^2 = d.$$

Como d es libre de cuadrados concluimos que $\{1, \alpha\}$ es en este caso una base entera de K. Resumimos en un teorema lo que hemos obtenido:

Teorema 8.17 *Sea d un entero libre de cuadrados y $K = \mathbb{Q}\big(\sqrt{d}\,\big)$.*

1. *Si $d \not\equiv 1 \pmod 4$ entonces $\mathcal{O}_K = \mathbb{Z}\big[\sqrt{d}\,\big]$ y $\Delta_K = 4d$.*
2. *Si $d \equiv 1 \pmod 4$ entonces $\mathcal{O}_K = \mathbb{Z}\left[\frac{1+\sqrt{d}}{2}\right]$ y $\Delta_K = d$.*

Recordemos que el anillo $\mathbb{Z}[i]$, donde $i = \sqrt{-1}$, se conoce con el nombre de anillo de los enteros de Gauss. Éste fue el primer anillo de enteros algebraicos estudiado en profundidad.

Ejercicio: Probar que si d y d' son enteros distintos libres de cuadrados entonces los cuerpos $\mathbb{Q}\big(\sqrt{d}\,\big)$ y $\mathbb{Q}\big(\sqrt{d'}\,\big)$ no son isomorfos.

El cálculo del anillo de enteros de un cuerpo numérico requiere a menudo técnicas específicas en las que no vamos a entrar aquí, porque el propósito de este capítulo es ilustrar las técnicas algebraicas que hemos desarrollado en los capítulos precedentes, no introducir técnicas propias de la teoría de números. No obstante, recordamos a continuación el caso de los cuerpos ciclotómicos de orden primo tratado en [ITAl] y luego un ejemplo correspondiente a un cuerpo cubico.

Enteros ciclotómicos Sea ω una raíz p-ésima primitiva de la unidad, donde p es un número primo impar, y consideremos el cuerpo $K = \mathbb{Q}(\omega)$. Recordemos que en el capítulo V obtuvimos que

$$\mathrm{Tr}\left(\sum_{i=0}^{p-1} a_i\omega^i\right) = pa_0 - \sum_{i=0}^{p-1} a_i,$$

así como que $\mathrm{N}(\omega^i) = 1$ para todo i, $\mathrm{N}(1-\omega) = p$.

Ahora debemos notar además que la norma y la traza de los enteros de un cuerpo numérico cualquiera son enteros racionales, pues por una parte son números racionales y por otra son producto (o suma) de enteros algebraicos, luego enteros. En esto se basa la prueba del teorema siguiente que dimos en [ITAl 17.12]:

Teorema 8.18 *Sea p un número primo impar y $K = \mathbb{Q}(\omega)$, donde ω es una raíz p-ésima primitiva de la unidad. Entonces $\mathcal{O}_K = \mathbb{Z}[\omega]$.*

Para calcular el discriminante de las extensiones ciclotómicas nos basaremos en el siguiente resultado general.

Teorema 8.19 *Sea $K = \mathbb{Q}(a)$ un cuerpo numérico de grado n y llamemos $p(x) = \mathrm{pol\,mín}(a, \mathbb{Q})$. Entonces*

$$\Delta[1, a, \ldots, a^{n-1}] = (-1)^{n(n-1)/2}\,\mathrm{N}\big(p'(a)\big),$$

donde $p'(x)$ es la derivada formal de $p(x)$.

Demostración: Según vimos en la sección anterior,

$$\Delta[1, a, \ldots, a^{n-1}] = \prod_{1\le i<j\le n} \big(\sigma_j(a) - \sigma_i(a)\big)^2, \tag{8.4}$$

donde $\sigma_1(a), \ldots, \sigma_n(a)$ son los conjugados de a.

Por otro lado, $p(x) = \prod_{i=1}^n \big(x - \sigma_i(a)\big)$, y se demuestra fácilmente (por inducción sobre n) que

$$p'(x) = \sum_{j=1}^{n} \prod_{\substack{i=1\\ i\ne j}}^{n} \big(x - \sigma_i(a)\big),$$

luego

$$p'\big(\sigma_j(a)\big) = \prod_{\substack{i=1\\ i\neq j}}^{n} \big(\sigma_j(a) - \sigma_i(a)\big)$$

para $j = 1, \ldots, n$. Multiplicando todas estas ecuaciones obtenemos

$$\mathrm{N}\big(p'(a)\big) = \prod_{j=1}^{n} \sigma_j\big(p'(a)\big) = \prod_{j=1}^{n} p'\big(\sigma_j(a)\big) = \prod_{\substack{i,j=1\\ i\neq j}}^{n} \big(\sigma_j(a) - \sigma_i(a)\big).$$

Agrupamos los pares $\big(\sigma_j(a)-\sigma_i(a)\big)\big(\sigma_i(a)-\sigma_j(a)\big) = -\big(\sigma_j(a)-\sigma_i(a)\big)^2$. El número de factores (-1) que aparecen es $n(n-1)/2$, luego teniendo en cuenta (8.4) queda

$$\mathrm{N}\big(p'(a)\big) = (-1)^{n(n-1)/2}\Delta[1, a, \ldots, a^{n-1}],$$

y de aquí se sigue el teorema. ∎

Como caso particular obtenemos:

Teorema 8.20 *Sea p un primo impar. El discriminante del cuerpo ciclotómico de orden p es igual a $(-1)^{(p-1)/2}p^{p-2}$.*

Demostración: Sea ω una raíz p-ésima primitiva de la unidad. Como los enteros ciclotómicos son el anillo $\mathbb{Z}[\omega]$, una base entera de $\mathbb{Q}(\omega)$ está formada por 1, ω, …, ω^{p-1}. El polinomio mínimo de ω es $p(x) = \frac{x^p-1}{x-1}$ y su derivada vale

$$p'(x) = \frac{px^{p-1}(x-1) - (x^p-1)}{(x-1)^2},$$

luego $p'(\omega) = \frac{p\omega^{p-1}}{\omega-1}$. Así pues,

$$\mathrm{N}\big(p'(\omega)\big) = \frac{p^{p-1}\cdot 1^{p-1}}{p} = p^{p-2}.$$

Como p es impar, $(-1)^{(p-1)(p-2)/2} = (-1)^{(p-1)/2}$ y, por el teorema anterior,

$$\Delta[1, \omega, \ldots, \omega^{p-1}] = (-1)^{(p-1)/2}p^{p-2}.$$ ∎

Ejercicio: Comprobar el teorema 8.19 sobre los cuerpos cuadráticos.

El anillo de enteros de $\mathbb{Q}\big(\sqrt[3]{2}\big)$. Vamos a probar que si $K = \mathbb{Q}\big(\sqrt[3]{2}\big)$, entonces $\mathcal{O}_K = \mathbb{Z}\big[\sqrt[3]{2}\big]$.

Recordemos que si llamamos $\alpha = \sqrt[3]{2}$ y β, γ a las otras raíces del polinomio $x^3 - 2$, entonces la clausura normal de K es $\mathbb{Q}(\alpha, \beta)$ y los monomorfismos de K envían α a α, β y γ respectivamente.

Un elemento de K es de la forma $\eta = \frac{u}{d} + \frac{v}{d}\sqrt[3]{2} + \frac{w}{d}\sqrt[3]{2}^{\,2}$, donde u, v, w, d son enteros racionales primos entre sí. Como la extensión tiene grado primo no

hay cuerpos intermedios, luego $\operatorname{pol}\operatorname{mín}(\eta,\mathbb{Q})$ tiene grado 3 y sus raíces son las imágenes de η por los tres monomorfismos de K, o sea,

$$\operatorname{pol}\operatorname{mín}(\eta,\mathbb{Q}) = (x - \frac{u}{d} - \frac{v}{d}\alpha - \frac{w}{d}\alpha^2)(x - \frac{u}{d} - \frac{v}{d}\beta - \frac{w}{d}\beta^2)(x - \frac{u}{d} - \frac{v}{d}\gamma - \frac{w}{d}\gamma^2).$$

Operando (con bastante paciencia) se llega a

$$\operatorname{pol}\operatorname{mín}(\eta,\mathbb{Q}) = x^3 - \frac{3u}{d}x^2 + \frac{3u^2 - 6vw}{d^2}x - \frac{u^3 + 2v^3 + 4w^3 - 6uvw}{d^3}.$$

Por lo tanto η será entero si y sólo si

$$\begin{aligned} d &\mid 3u && (8.5)\\ d^2 &\mid 3u^2 - 6vw && (8.6)\\ d^3 &\mid u^3 + 2v^3 + 4w^3 - 6uvw && (8.7) \end{aligned}$$

Si existe un primo p tal que $p \mid d$ y $p \mid u$, entonces por (8.6) $p^2 \mid 6vw$, lo que implica que $p \mid v$ o $p \mid w$.

Si $p \mid v$ por (8.7) $p^3 \mid 4w^3 - 6uvw$, y como $p^2 \mid uv$, también $p^2 \mid 4w^3$, luego $p = 2$. Pero entonces $4 \mid 2w^3 - 3uvw$ y $4 \mid uv$, luego $4 \mid 2w^3$ y $p \mid w$, contradicción.

Si $p \mid w$ entonces $p^3 \mid 2v^3 - 6uvw$, luego $p^2 \mid 2v^3$ y $p \mid v$, contradicción.

Así pues $(d, u) = 1$ y entonces (8.5) implica que $d = 1$ o $d = 3$. Basta probar que d no puede valer 3 y entonces η estará en $\mathbb{Z}\big[\sqrt[3]{2}\,\big]$.

Si $d = 3$, tenemos que $(3, u) = 1$, y de (8.6) se sigue que $(3, v) = (3, w) = 1$.

Tomando clases módulo 3 en (8.7) queda $[u] - [v] + [w] = 0$. Las únicas posibilidades son $[u] = [w] = [1]$, $[v] = [-1]$ o bien $[u] = [w] = [-1]$, $[v] = [1]$.

Si hacemos $u = 3k + 1$, $v = 3l - 1$, $w = 3r + 1$, sustituimos en

$$u^3 + 2v^3 + 4w^3 - 6uvw$$

y tomamos clases módulo 27, después de operar queda [9], cuando por (8.7) debería ser 0.

Con la segunda posibilidad llegamos a $[-9]$, luego concluimos que $d = 3$ es imposible.

Ahora el teorema 8.19 nos permite calcular el discriminante de la extensión: La derivada del polinomio mínimo de α es $3x^2$, luego

$$\Delta_K = (-1)^{3(3-1)/2}\,\mathrm{N}(3\alpha^2) = -\,\mathrm{N}(3)\,\mathrm{N}(\alpha)^2 = -108. \qquad \blacksquare$$

8.3 Divisibilidad en anillos de enteros

A la hora de estudiar la divisibilidad en anillos de enteros algebraicos es práctico hablar en términos de sus correspondientes cuerpos de cocientes, es decir, si K es un cuerpo numérico las unidades, los elementos irreducibles, primos etc. de K son por definición las unidades, elementos irreducibles, primos, etc.

de $\mathcal{O}_K$. Todos estos conceptos serían triviales aplicados literalmente a K, por ser un cuerpo. Vamos a ver que los anillos de enteros tienen propiedades similares a las de $\mathbb{Z}$, aunque no siempre son dominios de factorización única. Como vimos en [ITAl], una herramienta clave en el estudio de la divisibilidad en anillos de enteros algebraicos es la norma, que ahora tenemos definida sobre cuerpos numéricos arbitrarios y que se restringe a una aplicación $\mathrm{N} : \mathcal{O}_K \longrightarrow \mathbb{Z}$ que conserva los productos. El teorema siguiente recoge sus propiedades básicas:

Teorema 8.21 *Sea K un cuerpo numérico y $\mathrm{N} : \mathcal{O}_K \longrightarrow \mathbb{Z}$ la norma asociada.*

1. *Para todo $a, b \in \mathcal{O}_K$, se cumple $\mathrm{N}(ab) = \mathrm{N}(a)\,\mathrm{N}(b)$.*
2. *Si $a \in \mathcal{O}_K$, entonces $\mathrm{N}(a) = 0$ si y sólo si $a = 0$.*
3. *$\mathrm{N}(1) = 1$.*
4. *Si $a \in \mathcal{O}_K$, entonces $a \mid \mathrm{N}(a)$.*
5. *Si $a \in \mathcal{O}_K$, entonces a es una unidad si y sólo si $\mathrm{N}(a) = \pm 1$ y entonces $\mathrm{N}(a^{-1}) = \mathrm{N}(a)$.*
6. *Si a, $b \in \mathcal{O}_K$ son asociados, entonces $\mathrm{N}(a) = \pm\mathrm{N}(b)$.*
7. *Si $a \in \mathcal{O}_K$ y $\mathrm{N}(a)$ es un número primo, entonces a es irreducible en $\mathcal{O}_K$.*

Demostración: Las propiedades 1), 2), y 3) son consecuencia inmediata de la definición de norma.

4) Basta observar que $\mathrm{N}(a)/a$ es un producto de conjugados de a, luego es entero, y por otro lado está en K, luego $\mathrm{N}(a)/a \in \mathcal{O}_K$.

5) Si a es una unidad, $aa^{-1} = 1$, luego $\mathrm{N}(a)\,\mathrm{N}(a^{-1}) = 1$ y por lo tanto $\mathrm{N}(a) = \pm 1$.

Si $\mathrm{N}(a) = \pm 1$ entonces $a \mid \pm 1$, luego es una unidad. Como $\mathrm{N}(a)\,\mathrm{N}(a^{-1}) = 1$, se cumple $\mathrm{N}(a^{-1}) = \mathrm{N}(a)$.

6) es consecuencia de 5), pues dos asociados se diferencian en una unidad.

7) Si $a = bc$, con b, $c \in \mathcal{O}_K$, entonces $\mathrm{N}(a) = \mathrm{N}(b)\,\mathrm{N}(c)$, pero como $\mathrm{N}(a)$ es primo, $\mathrm{N}(b) = \pm 1$ o bien $\mathrm{N}(c) = \pm 1$, luego uno de los dos es una unidad. ∎

Ejercicio: Probar que 3 es primo en el anillo $\mathbb{Z}[i]$ a pesar de que su norma no es prima.

Ejercicio: Si K es un cuerpo numérico y $\alpha \in K$ tiene norma 1, ¿es necesariamente una unidad?

Ahora probamos los resultados básicos sobre los ideales de los anillos de enteros:

Teorema 8.22 *Sea K un cuerpo numérico y $\mathcal{O}$ su anillo de enteros. Entonces*

1. *$\mathcal{O}$ es un anillo noetheriano.*
2. *Si I es un ideal no nulo de $\mathcal{O}$ entonces el cociente $\mathcal{O}/I$ es finito.*
3. *Un ideal no nulo de $\mathcal{O}$ es primo si y sólo si es maximal.*

DEMOSTRACIÓN: 1) Un ideal I de $\mathcal{O}$ es un $\mathbb{Z}$-submódulo de $\mathcal{O}$. Como $\mathcal{O}$ es un $\mathbb{Z}$-módulo libre de rango finito, I también lo es (teorema 4.42), luego I está finitamente generado como $\mathbb{Z}$-módulo y *a fortiori* como ideal.

2) Sea $a \in I$, $a \neq 0$. Sea $n = N(a) \neq 0$. Como $a \mid n$, se cumple que $n \in I$.

Sea $\{v_1, \ldots, v_r\}$ una base de $\mathcal{O}$ como $\mathbb{Z}$-módulo. Entonces $\{[v_1], \ldots, [v_r]\}$ es un generador de $\mathcal{O}/I$ como $\mathbb{Z}$-módulo, es decir, todo elemento de $\mathcal{O}/I$ se puede expresar de la forma $m_1[v_1] + \cdots + m_r[v_r]$, para ciertos números enteros $m_1, \ldots, m_r$.

Por otra parte $n[v_i] = [nv_i] = 0$, para $i = 1, \ldots, r$, pues $n \in I$. Esto significa que el orden de cada $[v_i]$ es menor o igual que n y por lo tanto los números $m_1, \ldots, m_r$ pueden tomarse siempre entre 0 y $|n| - 1$. En consecuencia el anillo $\mathcal{O}/I$ es finito.

3) Si P es un ideal primo no nulo, el anillo $\mathcal{O}/P$ es un dominio íntegro (por el teorema 3.39) y 3.42 nos da que $\mathcal{O}/P$ es un cuerpo, luego el ideal P es maximal (por 3.40). ∎

El apartado 2) del teorema anterior puede precisarse más. El teorema siguiente generaliza a cuerpos numéricos arbitrarios el teorema que en [ITAl 13.9] sólo pudimos probar para cuerpos cuadráticos:

Teorema 8.23 *Sea K un cuerpo numérico y $\mathcal{O}$ su anillo de enteros. Para cada $a \in \mathcal{O}$ $a \neq 0$ se cumple que $|\mathcal{O}/a\mathcal{O}| = |\,\mathrm{N}(a)|$.*

DEMOSTRACIÓN: Sea $v_1, \ldots, v_n$ una base entera de K. Entonces todo elemento de $\mathcal{O}$ es de la forma $m_1v_1 + \cdots + m_nv_n$, con $m_1, \ldots, m_n \in \mathbb{Z}$, y todo elemento de $a\mathcal{O}$ es de la forma $m_1av_1 + \cdots + m_nav_n$, luego una base de $a\mathcal{O}$ es $av_1, \ldots, av_n$.

Sean $\sigma_1, \ldots, \sigma_n$ los monomorfismos de K. Entonces

$$\Delta[av_1, \ldots, av_n] = \big|\sigma_i(av_j)\big|^2 = \big|\sigma_i(a)\sigma_i(v_j)\big|^2.$$

Por la multilinealidad del determinante podemos sacar un factor $\sigma_i(a)$ de cada fila de la matriz, con lo que obtenemos

$$\Delta[av_1, \ldots, av_n] = \mathrm{N}(a)^2\big|\sigma_i(v_j)\big|^2 = \mathrm{N}(a)^2\Delta_K.$$

Sea C la matriz cuya fila i-ésima la forman las coordenadas de av_i en $v_1, \ldots, v_n$. Entonces sabemos que $\Delta[av_1, \ldots, av_n] = |C|^2\Delta_K$, luego ha de ser $|C| = \pm\,\mathrm{N}(a)$.

Por otro lado en virtud del teorema 6.15 tenemos que

$$|\mathcal{O}/a\mathcal{O}| = |\det C| = |\,\mathrm{N}(a)|. \qquad ∎$$

Ejercicio: Sea K un cuerpo numérico y $\alpha \in K$ un entero de norma prima. Probar que α es primo en $\mathcal{O}_K$.

Por último demostramos que los anillos de enteros satisfacen una propiedad que, según el teorema 3.35, es una condición necesaria para que puedan tener factorización única. En primer lugar la demostramos para el anillo de todos los enteros algebraicos:

Teorema 8.24 *Si $c \in \mathbb{A}$ es raíz de un polinomio mónico $p(x) \in \mathbb{E}[x]$, entonces $c \in \mathbb{E}$.*

DEMOSTRACIÓN: Sea $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$, donde cada $a_i \in \mathbb{E}$.

Sea $B = \mathbb{Z}[a_0, \ldots, a_{n-1}]$. Entonces B es un submódulo del anillo de enteros algebraicos de $\mathbb{Q}(a_0, \ldots, a_{n-1})$, luego es un $\mathbb{Z}$-módulo finitamente generado. Digamos que $B = \langle v_1, \ldots, v_r\rangle$. El mismo argumento empleado en el teorema 8.8 prueba ahora que $B[c] = \langle 1, c, \ldots, c^{n-1}\rangle_B$ (como B-módulo).

Sea N el $\mathbb{Z}$-módulo generado por los elementos $v_i \cdot c^k$, donde $1 \le i \le r$, $0 \le k \le n-1$.

Así, un elemento de $B[c]$ es una combinación lineal con coeficientes en B de los c^k y cada coeficiente es una combinación lineal con coeficientes enteros de los v_i.

Por lo tanto $\mathbb{Z}[c] \subset B[c] \subset N$ y es, en consecuencia, un $\mathbb{Z}$-módulo finitamente generado. Por el teorema 8.8 concluimos que c es un entero algebraico. ∎

Este teorema sirve para probar que determinados números algebraicos son enteros. Por ejemplo, un caso particular es que toda raíz n-sima de un entero algebraico es un entero algebraico. Como consecuencia inmediata tenemos la versión análoga para anillos de enteros algebraicos:

Teorema 8.25 *Sea K un cuerpo numérico y $\mathcal{O}$ su anillo de enteros algebraicos. Sea $p(x)$ un polinomio mónico no constante con coeficientes en $\mathcal{O}$. Si c es una raíz de $p(x)$ en K, entonces $c \in \mathcal{O}$.*

De aquí se sigue que cualquier anillo A cuyo cuerpo de cocientes sea K pero que esté estrictamente contenido en $\mathcal{O}$ no puede tener factorización única, pues un elemento de $\mathcal{O} \setminus A$ es raíz de un polinomio con coeficientes enteros (luego en A) y, sin embargo, no está en A, en contra de lo que exige el teorema 3.35. Es el caso, por ejemplo, de $\mathbb{Z}[\sqrt{5}]$.

8.4 Factorización ideal

Acabamos de ver que hay una gran similitud entre los anillos de enteros algebraicos y el anillo $\mathbb{Z}$ de los enteros ordinarios (son noetherianos, los ideales maximales coinciden con los primos, etc.), pero no es cierto que todos ellos sean DIP's o DFU's. En la sección [ITAl 13.2] vimos que todo elemento de un anillo de enteros algebraicos de un cuerpo cuadrático admite una descomposición única en "factores primos ideales", que coincide con su descomposición usual en factores primos cuando el anillo correspondiente es un DFU. Ahora estamos en condiciones de probar esto mismo para cualquier anillo de enteros algebraicos.

Recordemos que si $\mathfrak{a}$ y $\mathfrak{b}$ son ideales de un anillo D, su producto es

$$\mathfrak{ab} = \left\{ \sum_{i=1}^{n} p_i q_i \;\middle|\; n \in \mathbb{N} \text{ y } p_i \in \mathfrak{a}, q_i \in \mathfrak{b} \text{ para } i = 1, \ldots, n \right\}. \tag{8.8}$$

En otras palabras, $\mathfrak{ab}$ es el menor ideal que contiene a todos los productos ab tales que $a \in \mathfrak{a}$ y $b \in \mathfrak{b}$. Como $\mathfrak{a}$ y $\mathfrak{b}$ son ideales, estos productos están contenidos en ambos, luego se cumple que $\mathfrak{ab} \subset \mathfrak{a} \cap \mathfrak{b}$.

Definición 8.26 Un dominio íntegro D es un *dominio de Dedekind* si todo ideal propio de D (o sea, distinto de 0 y D) se descompone de forma única salvo el orden en producto de ideales primos.

Vamos a probar que la factorización ideal es formalmente análoga a la factorización real de los DFU's. Sin embargo tenemos un obstáculo justo al principio, y es que hay un hecho obvio en todo dominio íntegro cuyo análogo ideal no es evidente: los elementos no nulos son simplificables. Para probar que los ideales no nulos son simplificables demostraremos que el conjunto de los ideales de un dominio de Dedekind se puede sumergir en un grupo, con lo que para simplificar un ideal en ambos miembros de una igualdad bastará con multiplicar por su inverso en dicho grupo.

Definición 8.27 Sea D un dominio íntegro y K su cuerpo de cocientes. Un *ideal fraccional* de D es un D-submódulo no nulo $\mathfrak{a}$ de K tal que existe un $c \in D$ no nulo de manera que $c\mathfrak{a} \subset D$ (donde $c\mathfrak{a} = \{ca \mid a \in \mathfrak{a}\}$).

Si $\mathfrak{a}$ es un ideal fraccional de D, entonces $c\mathfrak{a}$ es D-submódulo de K contenido en D, luego es un D-submódulo de D, o también, $\mathfrak{b} = c\mathfrak{a}$ es un ideal no nulo de D y $\mathfrak{a} = c^{-1}\mathfrak{b}$.

El recíproco se prueba igualmente, luego, en definitiva, los ideales fraccionales de D son los conjuntos de la forma $c^{-1}\mathfrak{b}$, donde $\mathfrak{b}$ es un ideal no nulo de D y $c \in D$ es no nulo.

Tomando $c = 1$ deducimos que todos los ideales no nulos de D son ideales fraccionales. Recíprocamente, un ideal fraccional $\mathfrak{a}$ es un ideal si y sólo si $\mathfrak{a} \subset D$ (por la propia definición).

Podemos definir el producto de dos ideales fraccionales por la misma fórmula (8.8) que para ideales. Es fácil comprobar que efectivamente el producto de ideales fraccionales es un ideal fraccional, así como que cumple la propiedad asociativa.

Si $c \in K$ es no nulo, llamaremos ideal fraccional *principal* generado por c al ideal fraccional $(c) = cD$. Es fácil ver que $(c)\mathfrak{a} = c\mathfrak{a}$. En particular $(c)(d) = (cd)$.

Llamaremos $1 = (1) = D$. Es claro que $\mathfrak{a}\,1 = \mathfrak{a}$ para todo ideal fraccional $\mathfrak{a}$.

Diremos que un ideal fraccional $\mathfrak{a}$ es *inversible* si existe otro ideal fraccional $\mathfrak{b}$ tal que $\mathfrak{ab} = 1$. Es claro que si existe tal $\mathfrak{b}$ entonces es único, y lo representaremos por $\mathfrak{a}^{-1}$.

Todo ideal fraccional principal es inversible, pues $(c)^{-1} = (c^{-1})$.

Antes hemos visto que todo ideal fraccional es de la forma $c^{-1}\mathfrak{b}$, para cierto ideal $\mathfrak{b}$ y cierto entero c. En términos del producto de ideales fraccionales tenemos que todo ideal fraccional es de la forma $(c)^{-1}\mathfrak{b}$, o sea, una fracción de dos ideales. Para probar que los ideales fraccionales de un dominio de Dedekind forman un grupo necesitamos unos hechos sencillos válidos en cualquier dominio íntegro.

Teorema 8.28 *Sea D un dominio íntegro.*

1. *Todo ideal fraccional principal de D es inversible.*
2. *Un producto de ideales de D es inversible si y sólo si lo es cada factor.*
3. *Si un ideal inversible de D factoriza como producto de ideales primos, entonces la descomposición es única salvo el orden.*

DEMOSTRACIÓN: 1) Ya hemos observado que $(c)^{-1} = (c^{-1})$.

2) Es obvio que si cada factor es inversible el producto también lo es (su inverso es el producto de los inversos). Si el producto es inversible entonces el inverso de un factor es el inverso del producto multiplicado por los factores restantes.

3) Supongamos que un mismo ideal no nulo se expresa de dos formas

$$\mathfrak{p}_1 \cdots \mathfrak{p}_r = \mathfrak{q}_1 \cdots \mathfrak{q}_s$$

como producto de ideales primos (necesariamente no nulos). Podemos suponer que $r \leq s$.

Tomamos un factor (digamos $\mathfrak{p}_1$) que no contenga estrictamente a ninguno de los restantes. Por definición de ideal primo, y puesto que $\mathfrak{q}_1 \cdots \mathfrak{q}_s \subset \mathfrak{p}_1$, ha de existir un índice i de modo que $\mathfrak{q}_i \subset \mathfrak{p}_1$. Reordenando podemos suponer que $\mathfrak{q}_1 \subset \mathfrak{p}_1$. Igualmente ha de existir un índice j tal que $\mathfrak{p}_j \subset \mathfrak{q}_1 \subset \mathfrak{p}_1$. Por la elección de $\mathfrak{p}_1$ ha de ser $\mathfrak{p}_j = \mathfrak{q}_1 = \mathfrak{p}_1$. Tomando inversos podemos eliminarlos de la factorización, hasta llegar a que $D = \mathfrak{q}_{s-r} \cdots \mathfrak{q}_s \subset \mathfrak{q}_s$, lo que contradice la definición de ideal primo a no ser que $r = s$. Es claro que con esto el teorema queda demostrado. ∎

Teorema 8.29 *Si D es un dominio de Dedekind, entonces los ideales fraccionales de D forman un grupo. Además los ideales primos (no nulos) coinciden con los maximales.*

DEMOSTRACIÓN: Basta probar que todo ideal primo (no nulo) tiene un inverso y es maximal, pues entonces todo ideal no nulo será inversible por ser producto de ideales primos (inversibles) y todo ideal fraccional será inversible porque es de la forma $(c)^{-1}\mathfrak{b}$, donde $(c)^{-1}$ es ciertamente inversible y $\mathfrak{b}$ es un ideal, luego inversible también.

Vemos primero que todo ideal primo inversible es maximal. Sea $\mathfrak{p}$ un ideal primo. Hay que demostrar que si $d \in D \setminus \mathfrak{p}$ entonces $\mathfrak{p} + (d) = D$. En caso contrario existen ideales primos tales que $\mathfrak{p}+(d) = \mathfrak{p}_1 \cdots \mathfrak{p}_r$ y $\mathfrak{p}+(d^2) = \mathfrak{q}_1 \cdots \mathfrak{q}_s$. Es fácil ver que

$$\big(\mathfrak{p} + (d)\big)/\mathfrak{p} = \big(\mathfrak{p}_1/\mathfrak{p}\big) \cdots \big(\mathfrak{p}_r/\mathfrak{p}\big) \quad \text{y} \quad \big(\mathfrak{p} + (d^2)\big)/\mathfrak{p} = \big(\mathfrak{q}_1/\mathfrak{p}\big) \cdots \big(\mathfrak{q}_s/\mathfrak{p}\big).$$

El ideal $\big(\mathfrak{p} + (d)\big)/\mathfrak{p} = \big([d]\big)$ es principal y $D/\mathfrak{p}$ es un dominio íntegro, luego tiene inverso por el teorema anterior, el cual nos da también que todos los ideales primos $\mathfrak{p}_1/\mathfrak{p}$, …, $\mathfrak{p}_r/\mathfrak{p}$ tienen inverso en $D/\mathfrak{p}$.

Lo mismo ocurre con $\mathfrak{q}_1/\mathfrak{p}$, ..., $\mathfrak{q}_s/\mathfrak{p}$. Igualamos:

$$\big(\mathfrak{q}_1/\mathfrak{p}\big)\cdots\big(\mathfrak{q}_s/\mathfrak{p}\big) = \big([d^2]\big) = \big([d]\big)^2 = \big(\mathfrak{p}_1/\mathfrak{p}\big)^2\cdots\big(\mathfrak{p}_r/\mathfrak{p}\big)^2.$$

Otra aplicación del teorema anterior nos da que $s = 2r$ y que, ordenando adecuadamente, $\mathfrak{p}_i/\mathfrak{p} = \mathfrak{q}_{2i}/\mathfrak{p} = \mathfrak{q}_{2i-1}/\mathfrak{p}$. De aquí se sigue que $\mathfrak{p}_i = \mathfrak{q}_{2i} = \mathfrak{q}_{2i-1}$, y de aquí a su vez obtenemos que $\mathfrak{p} + (d^2) = \big(\mathfrak{p} + (d)\big)^2$. Consecuentemente

$$\mathfrak{p} \subset \mathfrak{p} + (d^2) = \big(\mathfrak{p} + (d)\big)^2 \subset \mathfrak{p}^2 + (d).$$

Todo elemento de $\mathfrak{p}$ es, pues, de la forma $c + ad$, con $c \in \mathfrak{p}^2$ y $a \in D$, pero como $\mathfrak{p}$ es primo y $d \notin \mathfrak{p}$, ha de ser $a \in \mathfrak{p}$, lo que prueba que $\mathfrak{p} \subset \mathfrak{p}^2 + \mathfrak{p}(d) \subset \mathfrak{p}$, es decir, $\mathfrak{p} = \mathfrak{p}^2 + \mathfrak{p}(d)$, y como $\mathfrak{p}$ tiene inverso, $1 = \mathfrak{p} + (d)$, contradicción.

Finalmente, si $\mathfrak{p}$ es cualquier ideal primo no nulo, sea $c \in \mathfrak{p}$, $c \neq 0$. Como D es un dominio de Dedekind podemos factorizar $(c) = \mathfrak{p}_1\cdots\mathfrak{p}_r \subset \mathfrak{p}$, donde los ideales primos $\mathfrak{p}_i$ son todos inversibles (por el teorema anterior, ya que (c) lo es) y en consecuencia maximales (por lo ya probado). Por definición de ideal primo, algún ideal $\mathfrak{p}_i$ está contenido en $\mathfrak{p}$, luego por maximalidad $\mathfrak{p} = \mathfrak{p}_i$ es maximal y tiene inverso. ∎

Ahora ya podemos trabajar con dominios de Dedekind como si fueran DFU's.

Definición 8.30 Sea D un dominio de Dedekind. Diremos que un ideal $\mathfrak{b}$ *divide* a un ideal $\mathfrak{a}$ si existe un ideal $\mathfrak{c}$ tal que $\mathfrak{a} = \mathfrak{b}\mathfrak{c}$. Lo representaremos $\mathfrak{b} \mid \mathfrak{a}$. Notemos que en tal caso $\mathfrak{c} = \mathfrak{a}\mathfrak{b}^{-1}$. Claramente $\mathfrak{b} \mid \mathfrak{a}$ si y sólo si $\mathfrak{a}\mathfrak{b}^{-1}$ es un ideal.

Observemos que $\mathfrak{b} \mid \mathfrak{a}$ si y sólo si $\mathfrak{a} \subset \mathfrak{b}$. En efecto, si $\mathfrak{b} \mid \mathfrak{a}$ entonces $\mathfrak{a} = \mathfrak{b}\mathfrak{c} \subset \mathfrak{b}$ y si $\mathfrak{a} \subset \mathfrak{b}$ la propia definición de producto nos da que $\mathfrak{a}\mathfrak{b}^{-1} \subset \mathfrak{b}\mathfrak{b}^{-1} = 1 = D$, luego el ideal fraccional $\mathfrak{a}\mathfrak{b}^{-1}$ es de hecho un ideal y por lo tanto $\mathfrak{b} \mid \mathfrak{a}$.

Así, un ideal $\mathfrak{p}$ es primo si y sólo si $\mathfrak{p} \neq 1$ y cuando $\mathfrak{p} \mid \mathfrak{a}\mathfrak{b}$ entonces $\mathfrak{p} \mid \mathfrak{a}$ o $\mathfrak{p} \mid \mathfrak{b}$, es decir, el concepto de ideal primo en un dominio de Dedekind es formalmente análogo al de primo real en un DFU.

Similarmente, un ideal $\mathfrak{p}$ es maximal si y solo si $\mathfrak{p} \neq 1$ y cuando $\mathfrak{a} \mid \mathfrak{p}$ entonces $\mathfrak{a} = 1$ o $\mathfrak{a} = \mathfrak{p}$, es decir, el concepto de ideal maximal en un dominio de Dedekind es formalmente análogo al de elemento irreducible en un DFU (notemos que en términos de ideales no hay ni unidades ni asociados). Hemos probado que en un dominio de Dedekind maximal equivale a primo, lo cual es análogo al hecho de que en un DFU irreducible equivale a primo.

Si $c \in D$ escribiremos $\mathfrak{a} \mid c$ o $c = \mathfrak{a}\mathfrak{b}$ en lugar de $\mathfrak{a} \mid (c)$ o $(c) = \mathfrak{a}\mathfrak{b}$. De este modo los divisores ideales pueden dividir a elementos reales. Concretamente, tenemos $\mathfrak{a} \mid c$ si y sólo si $(c) \subset \mathfrak{a}$, si y sólo si $c \in \mathfrak{a}$, es decir, un ideal, como conjunto, es el conjunto de todos sus múltiplos reales. Nótese también que $a \mid b$ si y sólo si $(a) \mid (b)$.

La factorización única ideal nos permite hablar de la multiplicidad de un ideal primo en otro ideal (o en un elemento real) exactamente en el mismo sentido que en un DFU. Toda familia finita de ideales tiene un máximo común

divisor y un mínimo común múltiplo que se pueden calcular como en un DFU, aunque en realidad hay una caracterización más simple: Teniendo en cuenta que $\mathfrak{a} \mid \mathfrak{b}$ es lo mismo que $\mathfrak{b} \subset \mathfrak{a}$, resulta que el máximo común divisor de una familia de ideales es el menor ideal que los contiene, y el mínimo común múltiplo es el mayor ideal contenido en ellos, o sea:

$$\begin{aligned} \mathrm{mcd}(\mathfrak{a}_1, \ldots, \mathfrak{a}_r) &= \mathfrak{a}_1 + \cdots + \mathfrak{a}_r, \\ \mathrm{mcm}(\mathfrak{a}_1, \ldots, \mathfrak{a}_r) &= \mathfrak{a}_1 \cap \cdots \cap \mathfrak{a}_r. \end{aligned}$$

(Esto generaliza el teorema de Bezout.)

En particular $(a, b) = (a) + (b)$ puede entenderse como el ideal generado por a y b o como el máximo común divisor de (a) y (b). Es equivalente. Podemos hablar de ideales primos entre sí, etc. con las mismas propiedades que en un DFU.

Es fácil encontrar DFU's que no sean dominios de Dedekind. Por ejemplo $\mathbb{Z}[x]$ no es un dominio de Dedekind ya que $\mathbb{Z}[x]/(x) \cong \mathbb{Z}$, luego (x) es un ideal primo no maximal. Recíprocamente veremos que todos los anillos de enteros algebraicos son dominios de Dedekind y muchos de ellos no son DFU's. Por lo tanto la divisibilidad ideal no es una generalización de la real, sino que ambas son paralelas. Las dos pueden darse simultáneamente. Esto ocurre exactamente en los DIP's:

Teorema 8.31 *Un dominio íntegro D es un* DIP *si y sólo si es un dominio de Dedekind y un dominio de factorización única.*

Demostración: Si D es DIP ya sabemos que es DFU, y todo ideal propio de D es de la forma (c), donde c no es 0 ni una unidad. Entonces c se descompone en producto de primos $c = p_1 \cdots p_n$, con lo que $(c) = (p_1) \cdots (p_n)$ también es producto de ideales primos. Recíprocamente, una descomposición de (c) en ideales primos da lugar a una factorización de c, de donde se sigue la unicidad.

Si D es a la vez un dominio de Dedekind y un DFU entonces dado un ideal primo $\mathfrak{p}$ tomamos un $c \in \mathfrak{p}$ no nulo y lo factorizamos $c = p_1 \cdots p_n$ en producto de primos. Tenemos que $\mathfrak{p} \mid c$, luego $\mathfrak{p} \mid p_i$ para algún i, luego $(p_i) \subset \mathfrak{p}$ y como los ideales primos son maximales, $\mathfrak{p} = (p_i)$ es principal, y todo ideal propio de D es principal por ser producto de ideales primos principales. ∎

Notemos que todo dominio de Dedekind es noetheriano, pues una cadena de ideales estrictamente creciente significaría una cadena decreciente de divisores, lo cual es imposible. Por consiguiente, el teorema 3.18 implica que un dominio de Dedekind es un dominio de factorización única si y sólo si todos sus elementos irreducibles son primos, mientras que 3.55 implica que un dominio de factorización única es un dominio de Dedekind si y sólo si todos sus ideales primos no triviales son maximales.

Ejercicio: Probar que si X es infinito entonces $\mathbb{Q}[X]$ es un DFU no noetheriano.

El concepto de DFU es muy útil en cuanto que proporciona un gran control sobre los anillos que tienen esta propiedad, pero está el inconveniente de que

no es fácil determinar cuándo se da el caso. En cambio, la propiedad de ser un dominio de Dedekind admite una caracterización algebraica muy fácil de verificar en la práctica. Veámosla:

Teorema 8.32 (Teorema de Dedekind) *Sea D un dominio íntegro y K su cuerpo de cocientes. Entonces D es un dominio de Dedekind si y sólo si cumple las tres propiedades siguientes:*

1. *D es noetheriano.*
2. *Los ideales primos no nulos de D son maximales.*
3. *Si $a \in K$ es raíz de un polinomio mónico con coeficientes en D, entonces $a \in D$.*

Demostración: Ya hemos observado que dominio de Dedekind es noetheriano. La propiedad 2) está probada en el teorema 8.29. La prueba del teorema 3.35 vale para probar 3) sin más cambio que considerar divisores ideales en vez de reales.

Supongamos ahora que un dominio íntegro D cumple las tres propiedades del enunciado y veamos que es un dominio de Dedekind. Dividimos la prueba en varios pasos.

(i) *Sea $\mathfrak{a} \neq 0$ un ideal de D. Entonces existen ideales primos $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ de manera que $\mathfrak{p}_1 \cdots \mathfrak{p}_r \subset \mathfrak{a}$.*

En caso contrario por el teorema 3.7 existe un ideal $\mathfrak{a}$ tal que no existen ideales primos en las condiciones pedidas y que es maximal entre los ideales para los que esto ocurre.

En particular $\mathfrak{a}$ no puede ser primo, o cumpliría (i) trivialmente. Tampoco puede ser que $\mathfrak{a} = D$. Por lo tanto existen dos ideales $\mathfrak{b}$ y $\mathfrak{c}$ tales que $\mathfrak{bc} \subset \mathfrak{a}$, pero no $\mathfrak{b} \subset \mathfrak{a}$ o $\mathfrak{c} \subset \mathfrak{a}$.

Por la maximalidad de $\mathfrak{a}$, existen ideales primos $\mathfrak{p}_1, \ldots, \mathfrak{p}_s$ y $\mathfrak{p}_{s+1}, \ldots, \mathfrak{p}_r$ tales que

$$\mathfrak{p}_1 \cdots \mathfrak{p}_s \subset \mathfrak{a} + \mathfrak{b}, \qquad \mathfrak{p}_{s+1} \cdots \mathfrak{p}_r \subset \mathfrak{a} + \mathfrak{c},$$

de donde $\mathfrak{p}_1 \cdots \mathfrak{p}_r \subset (\mathfrak{a} + \mathfrak{b})(\mathfrak{a} + \mathfrak{c}) \subset \mathfrak{aa} + \mathfrak{ab} + \mathfrak{ac} + \mathfrak{bc} \subset \mathfrak{a}$, contradicción.

(ii) *Si $\mathfrak{a}$ es un ideal no nulo de D, llamaremos $\mathfrak{a}^{-1} = \{x \in K \mid x\mathfrak{a} \subset D\}$.*

Es claro que $\mathfrak{a}^{-1}$ es un D-submódulo de K, y para cualquier $c \in \mathfrak{a}$ no nulo se cumple que $c\mathfrak{a}^{-1} \subset D$, luego $\mathfrak{a}^{-1}$ es un ideal fraccional de D.

También es inmediato que $D \subset \mathfrak{a}^{-1}$, luego $\mathfrak{a} = \mathfrak{a}D \subset \mathfrak{aa}^{-1}$.

De la definición de $\mathfrak{a}^{-1}$ se sigue que $\mathfrak{aa}^{-1} = \mathfrak{a}^{-1}\mathfrak{a} \subset D$. Esto significa que el ideal fraccional $\mathfrak{a}^{-1}\mathfrak{a}$ es de hecho un ideal de D.

Nótese también que si $\mathfrak{a} \subset \mathfrak{b}$ son dos ideales de D, entonces $D \subset \mathfrak{b}^{-1} \subset \mathfrak{a}^{-1}$.

(iii) *Si $\mathfrak{a}$ es un ideal propio, entonces $D \subsetneq \mathfrak{a}^{-1}$.*

Sea $\mathfrak{p}$ un ideal maximal de D tal que $\mathfrak{a} \subset \mathfrak{p}$. Entonces $\mathfrak{p}^{-1} \subset \mathfrak{a}^{-1}$. Basta probar que $\mathfrak{p}^{-1}$ contiene estrictamente a D. Sea $a \in \mathfrak{p}$ no nulo. Por (i), sea r el menor natural tal que existen ideales primos para los que $\mathfrak{p}_1 \cdots \mathfrak{p}_r \subset (a)$. Como $(a) \subset \mathfrak{p}$ y $\mathfrak{p}$ es primo, existe un índice i tal que $\mathfrak{p}_i \subset \mathfrak{p}$. Reordenando podemos suponer que $\mathfrak{p}_1 \subset \mathfrak{p}$. Como $\mathfrak{p}_1$ es maximal ha de ser $\mathfrak{p}_1 = \mathfrak{p}$, y por la minimalidad de r tenemos que $\mathfrak{p}_2 \cdots \mathfrak{p}_r$ no está contenido en (a). Tomamos, pues, un elemento $b \in \mathfrak{p}_2 \cdots \mathfrak{p}_r \setminus (a)$.

Claramente $b\mathfrak{p} \subset (a)$, luego $ba^{-1}\mathfrak{p} \subset a^{-1}(a) = D$ y $ba^{-1} \in \mathfrak{p}^{-1}$, pero por otra parte $b \notin (a) = aD$, luego $ba^{-1} \notin D$. Así pues, $\mathfrak{p}^{-1} \neq D$.

(iv) *Si $\mathfrak{a}$ es un ideal no nulo de D y S es un subconjunto de K tal que $\mathfrak{a}S \subset \mathfrak{a}$, entonces $S \subset D$.*

Sea $s \in S$. Como D es noetheriano tenemos que $\mathfrak{a} = (a_1, \ldots, a_m)$ para ciertos $a_1, \ldots, a_m \in D$. Por hipótesis $a_i s \in \mathfrak{a}$ para $i = 1, \ldots, m$, luego existen elementos $b_{ij} \in D$ de manera que

$$a_i s = \sum_{j=1}^{m} b_{ij} a_j \qquad \text{para } i = 1, \ldots, m.$$

Como los a_i no son todos nulos, esto quiere decir que s es un valor propio de la aplicación lineal en K^n definida por la matriz $B = (b_{ij})$, luego s es raíz del polinomio característico $p(x) = |B - xI_m| \in D[x]$, que es mónico. Por la hipótesis 3) tenemos que $s \in D$.

(v) *Si $\mathfrak{p}$ es un ideal maximal de D, entonces $\mathfrak{p}\mathfrak{p}^{-1} = D$.*

Por (ii) sabemos que $\mathfrak{p}\mathfrak{p}^{-1}$ es un ideal de D tal que $\mathfrak{p} \subset \mathfrak{p}\mathfrak{p}^{-1} \subset D$. Puesto que $\mathfrak{p}$ es maximal, ha de ser $\mathfrak{p} = \mathfrak{p}\mathfrak{p}^{-1}$ o bien $\mathfrak{p}\mathfrak{p}^{-1} = D$. Si se diera el primer caso, por (iv) tendríamos que $\mathfrak{p}^{-1} \subset D$, lo que contradice a (iii).

(vi) *Si $\mathfrak{a} \neq 0$ es un ideal, entonces $\mathfrak{a}\mathfrak{a}^{-1} = D$.*

Supongamos lo contrario. Como D es noetheriano existe un ideal $\mathfrak{a}$ maximal entre los que incumplen (vi). Obviamente $\mathfrak{a} \neq D$. Sea $\mathfrak{p}$ un ideal maximal tal que $\mathfrak{a} \subset \mathfrak{p}$.

Por (ii) $D \subset \mathfrak{p}^{-1} \subset \mathfrak{a}^{-1}$, luego $\mathfrak{a} \subset \mathfrak{a}\mathfrak{p}^{-1} \subset \mathfrak{a}\mathfrak{a}^{-1} \subset D$. En particular el ideal fraccional $\mathfrak{a}\mathfrak{p}^{-1}$ es un ideal de D. No puede ocurrir que $\mathfrak{a} = \mathfrak{a}\mathfrak{p}^{-1}$, pues entonces (iv) implicaría que $\mathfrak{p}^{-1} \subset D$ en contradicción con (iii). Así pues, $\mathfrak{a} \subsetneq \mathfrak{a}\mathfrak{p}^{-1}$, luego la maximalidad de $\mathfrak{a}$ implica que $\mathfrak{a}\mathfrak{p}^{-1}$ cumple (vi), es decir, que $\mathfrak{a}\mathfrak{p}^{-1}(\mathfrak{a}\mathfrak{p}^{-1})^{-1} = D$. Por definición de $\mathfrak{a}^{-1}$ esto significa que $\mathfrak{p}^{-1}(\mathfrak{a}\mathfrak{p}^{-1})^{-1} \subset \mathfrak{a}^{-1}$. Por lo tanto $D = \mathfrak{a}\mathfrak{p}^{-1}(\mathfrak{a}\mathfrak{p}^{-1})^{-1} \subset \mathfrak{a}\mathfrak{a}^{-1} \subset D$, de donde $\mathfrak{a}\mathfrak{a}^{-1} = D$, en contradicción con nuestra hipótesis.

(vii) *Todo ideal propio de D es producto de ideales primos.*

En caso contrario sea $\mathfrak{a}$ un ideal propio maximal entre los que no pueden expresarse como producto de ideales primos. En particular $\mathfrak{a}$ no es primo. Sea $\mathfrak{p}$ un ideal maximal tal que $\mathfrak{a} \subset \mathfrak{p}$. Como en (vi) concluimos que $\mathfrak{a} \subset \mathfrak{a}\mathfrak{p}^{-1} \subset D$ y de nuevo por (iv) y (iii), la primera inclusión es estricta.

Por la maximalidad de $\mathfrak{a}$ tenemos que $\mathfrak{a}\mathfrak{p}^{-1} = \mathfrak{p}_1 \cdots \mathfrak{p}_r$ para ciertos ideales primos $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$. Por lo tanto $\mathfrak{a} = \mathfrak{p}_1 \cdots \mathfrak{p}_r \mathfrak{p}$, en contra de la elección de $\mathfrak{a}$.

(viii) *La descomposición de un ideal en primos es única salvo el orden.*

Supongamos que un mismo ideal propio se expresa de dos formas

$$\mathfrak{p}_1 \cdots \mathfrak{p}_r = \mathfrak{q}_1 \cdots \mathfrak{q}_s$$

como producto de ideales primos (necesariamente no nulos). Podemos suponer que $r \leq s$.

Entonces, puesto que $\mathfrak{p}_1$ es primo y $\mathfrak{q}_1 \cdots \mathfrak{q}_s \subset \mathfrak{p}_1$, ha de existir un índice i tal que $\mathfrak{q}_i \subset \mathfrak{p}_1$. Reordenando podemos suponer que $\mathfrak{q}_1 \subset \mathfrak{p}_1$ y, por maximalidad, de hecho $\mathfrak{q}_1 = \mathfrak{p}_1$. Multiplicando por el inverso tenemos $\mathfrak{p}_2 \cdots \mathfrak{p}_r = \mathfrak{q}_2 \cdots \mathfrak{q}_s$. Repitiendo el proceso llegamos a que $\mathfrak{p}_i = \mathfrak{q}_i$ para $i = 1, \ldots, r$ y (si $r < s$) a que $D = \mathfrak{q}_{s-r} \cdots \mathfrak{q}_s$, pero entonces $D \subset \mathfrak{q}_s$, lo cual es imposible. Por lo tanto ha de ser $r = s$. ∎

Observemos que la prueba del teorema anterior nos ha dado una expresión explícita para el inverso de un ideal en un dominio de Dedekind:

$$\mathfrak{a}^{-1} = \{x \in K \mid x\mathfrak{a} \subset D\}.$$

Los teoremas 8.22 y 8.25, junto con el teorema de Dedekind, implican que todo anillo de enteros algebraicos de un cuerpo numérico es un dominio de Dedekind. El teorema 8.22 afirma también que estos anillos de enteros cumplen una propiedad adicional muy importante que no poseen todos los dominios de Dedekind, y es que los cocientes módulo ideales no nulos son finitos. El anillo $\mathbb{Q}[x]$ es un ejemplo de dominio de Dedekind que no cumple esta condición. El interés de la finitud de los cocientes reside en que es la clave para definir la norma de un ideal, que representará el mismo papel que la norma de los elementos reales en el estudio de la divisibilidad. En efecto:

Definición 8.33 Sea K un cuerpo numérico. Si $\mathfrak{a}$ es un ideal no nulo de $\mathcal{O}_K$, llamaremos *norma* de $\mathfrak{a}$ al cardinal del anillo cociente: $\mathrm{N}(\mathfrak{a}) = |\mathcal{O}/\mathfrak{a}|$.

Así, la norma de $\mathfrak{a}$ es un número natural no nulo y $\mathrm{N}(\mathfrak{a}) = 1$ si y sólo si $\mathfrak{a} = 1$. El teorema 8.23 implica además que si $a \in \mathcal{O}$ entonces $\mathrm{N}\big((a)\big) = |\,\mathrm{N}(a)|$, es decir, que la norma de ideales extiende (salvo signo) a la de enteros reales. El teorema siguiente acaba de garantizar que la norma de ideales se comporta satisfactoriamente:

Teorema 8.34 *Sea K un cuerpo numérico y $\mathcal{O}$ su anillo de enteros. Si $\mathfrak{a}$, $\mathfrak{b}$ son dos ideales no nulos de $\mathcal{O}$, entonces* $\mathrm{N}(\mathfrak{a}\mathfrak{b}) = \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{b})$.

Demostración: Por la unicidad de la factorización en primos e inducción sobre el número de factores, basta probar que $\mathrm{N}(\mathfrak{a}\mathfrak{p}) = \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{p})$ cuando $\mathfrak{p}$ es un ideal primo (el caso en que uno de los factores es 1 es obvio).

Consideremos los grupos abelianos finitos $\mathfrak{a}/\mathfrak{ap} \leq \mathcal{O}/\mathfrak{ap}$. El tercer teorema de isomorfía implica que $|\mathcal{O}/\mathfrak{ap}| = |\mathcal{O}/\mathfrak{a}|\,|\mathfrak{a}/\mathfrak{ap}|$, o sea, $\mathrm{N}(\mathfrak{ap}) = \mathrm{N}(\mathfrak{a})\,|\mathfrak{a}/\mathfrak{ap}|$. Basta probar que $|\mathfrak{a}/\mathfrak{ap}| = |\mathcal{O}/\mathfrak{p}|$. Notemos que por la factorización única $\mathfrak{ap}$ no puede ser igual a $\mathfrak{p}$, luego $\mathfrak{ap} \subsetneq \mathfrak{a}$, es decir, $|\mathfrak{a}/\mathfrak{ap}| > 1$.

Por el mismo motivo no pueden existir ideales $\mathfrak{b}$ de $\mathcal{O}$ tales que $\mathfrak{ap} \subsetneq \mathfrak{b} \subsetneq \mathfrak{a}$, pues entonces $\mathfrak{a} \mid \mathfrak{b} \mid \mathfrak{ap}$, luego la descomposición en factores de $\mathfrak{b}$ debe contener a la de $\mathfrak{a}$ y estar contenida en la de $\mathfrak{ap}$, luego $\mathfrak{b}$ será igual a $\mathfrak{ap}$ o a $\mathfrak{a}$ según que la multiplicidad de $\mathfrak{p}$ en $\mathfrak{b}$ sea la de $\mathfrak{ap}$ o la de $\mathfrak{a}$.

Por lo tanto, si $a \in \mathfrak{a}\setminus\mathfrak{ap}$, entonces $\mathfrak{a} = \mathfrak{ap}+(a)$ y a su vez esto implica que la aplicación $f : \mathcal{O} \longrightarrow \mathfrak{a}/\mathfrak{ap}$ dada por $f(x) = [xa]$ es un epimorfismo de $\mathcal{O}$-módulos con la propiedad de que $\mathfrak{p} \subset \mathrm{N}(f)$. Ahora, $\mathrm{N}(f)$ es un $\mathcal{O}$-submódulo de $\mathcal{O}$, o sea, un ideal. Como $\mathfrak{p}$ es maximal, ha de ser $\mathrm{N}(f) = \mathfrak{p}$ o $\mathrm{N}(f) = \mathcal{O}$, pero el segundo caso implicaría que $\mathfrak{a}/\mathfrak{ap} \cong \mathcal{O}/\mathcal{O}$, con lo que $|\mathfrak{a}/\mathfrak{ap}| = 1$, contradicción. Lo correcto es $\mathfrak{a}/\mathfrak{ap} \cong \mathcal{O}/\mathfrak{p}$, y así $|\mathcal{O}/\mathfrak{p}| = |\mathfrak{a}/\mathfrak{ap}|$. ∎

El teorema siguiente recoge los hechos básicos sobre las normas de ideales análogos a los ya conocidos para normas de enteros reales.

Teorema 8.35 *Sea K un cuerpo numérico y $\mathfrak{a}$, $\mathfrak{b}$ ideales de $\mathcal{O}_K$.*

1. *Si $\mathfrak{a} \mid \mathfrak{b}$ entonces $\mathrm{N}(\mathfrak{a}) \mid \mathrm{N}(\mathfrak{b})$.*
2. *Si $\mathrm{N}(\mathfrak{a})$ es un número primo, entonces $\mathfrak{a}$ es un ideal primo.*
3. *$\mathfrak{a} \mid \mathrm{N}(\mathfrak{a})$.*
4. *Si $\mathfrak{a}$ es un ideal primo no nulo, entonces $\mathfrak{a}$ divide a un único primo racional p y se cumple que $\mathrm{N}(\mathfrak{a}) = p^m$ para cierto natural m menor o igual que el grado de K.*

Demostración: 1) es consecuencia inmediata del teorema 8.34.

2) Un ideal de norma prima no puede descomponerse en primos (por 1), luego ha de ser primo.

3) Por definición, $\mathrm{N}(\mathfrak{a}) = |\mathcal{O}_K/\mathfrak{a}|$. El anillo $\mathcal{O}_K/\mathfrak{a}$ es en particular un grupo finito (con la suma) y el orden de cualquier elemento es divisor de $\mathrm{N}(\mathfrak{a})$. Por lo tanto $\mathrm{N}(\mathfrak{a})[1] = [0]$, lo que equivale a que $\mathrm{N}(\mathfrak{a}) \in \mathfrak{a}$.

4) Como $\mathfrak{a} \mid \mathrm{N}(\mathfrak{a})$ y $\mathfrak{a}$ es primo, $\mathfrak{a}$ debe dividir a uno de los primos racionales que dividen a $\mathrm{N}(\mathfrak{a})$. Digamos que $\mathfrak{a} \mid p$. Entonces $\mathrm{N}(\mathfrak{a}) \mid \mathrm{N}(p) = p^n$, donde n es el grado de K. Consecuentemente, $\mathrm{N}(\mathfrak{a}) = p^m$ para un cierto $m \leq n$.

Si $\mathfrak{a}$ dividiera a otro primo q, el mismo argumento nos daría que $\mathrm{N}(\mathfrak{a})$ habría de ser potencia de q, lo cual es imposible salvo si $q = p$. ∎

Este teorema contiene información relevante a la hora de estudiar los ideales propios de un anillo de enteros. El apartado 4) nos dice que todo ideal primo divide a un primo racional, por lo que factorizando los primos racionales se encuentran todos los ideales primos. La unicidad de 4) implica que los primos racionales (no asociados) son primos entre sí, de donde se sigue la existencia de infinitos ideales primos en cada anillo de enteros (al menos uno distinto para cada primo racional).

Los hechos siguientes son muy sencillos, pero a menudo resultan útiles:

Teorema 8.36 *Sea K un cuerpo numérico.*

1. *Cada ideal no nulo de $\mathcal{O}_K$ tiene sólo un número finito de divisores.*
2. *Cada elemento no nulo de $\mathcal{O}_K$ pertenece a un número finito de ideales.*
3. *Sólo un número finito de ideales pueden tener una norma dada.*

DEMOSTRACIÓN: 1) es consecuencia inmediata de la factorización única. 2) es un caso particular de 1) y 3) se sigue de 1) porque cada ideal es un divisor de su norma. ∎

Con esto estamos en condiciones de ver ejemplos cómodamente. Por ejemplo, en $\mathbb{Q}(\sqrt{-5}\,)$ tenemos[1] las siguientes descomposiciones en irreducibles:

$$6 = 2 \cdot 3 = \big(1 + \sqrt{-5}\,\big)\big(1 - \sqrt{-5}\,\big). \tag{8.9}$$

Esto prueba que 2 no es primo, y como $\mathrm{N}(2) = 4$, el ideal (2) sólo puede descomponerse en producto de dos ideales primos de norma 2, o sea, $2 = \mathfrak{p}_1\mathfrak{p}_2$. Igualmente 3 ha de ser producto de dos ideales de norma 3, digamos $3 = \mathfrak{q}\mathfrak{r}$. Por otra parte, los factores de la derecha en (8.9) tienen los dos norma 6, luego han de descomponerse en producto de un ideal de norma 2 por otro de norma 3. La unicidad de la factorización obliga a que sea $\big(1+\sqrt{-5}\,\big) = \mathfrak{p}_1\mathfrak{q}$ y $\big(1-\sqrt{-5}\,\big) = \mathfrak{p}_2\mathfrak{r}$, de modo que la factorización única de 6 es

$$6 = 2 \cdot 3 = (\mathfrak{p}_1\mathfrak{p}_2)(\mathfrak{q}\mathfrak{r}) = (\mathfrak{p}_1\mathfrak{q})(\mathfrak{p}_2\mathfrak{r}) = \big(1 + \sqrt{-5}\,\big)\big(1 - \sqrt{-5}\,\big).$$

Más aún, evidentemente $\mathfrak{p}_1$ es el máximo común divisor de 2 y $1 + \sqrt{-5}$, es decir, que $\mathfrak{p}_1 = \big(2, 1 + \sqrt{-5}\,\big)$.

Similarmente $\mathfrak{p}_2 = \big(2, 1 - \sqrt{-5}\,\big)$, $\mathfrak{q} = \big(3, 1 + \sqrt{-5}\,\big)$ y $\mathfrak{r} = \big(3, 1 - \sqrt{-5}\,\big)$.

Finalmente observamos que $\mathfrak{p}_1 = \mathfrak{p}_2$, pues $1 - \sqrt{-5} = 2 - \big(1 + \sqrt{-5}\,\big)$. Por el contrario $\mathfrak{q} \neq \mathfrak{r}$, pues en otro caso $1 = 3 - \big(1 + \sqrt{-5} + 1 - \sqrt{-5}\,\big) \in \mathfrak{q}$, o lo que es lo mismo, $\mathfrak{q} = 1$.

Si llamamos $\mathfrak{p} = \mathfrak{p}_1 = \mathfrak{p}_2$, la factorización de 6 es, en definitiva, $6 = \mathfrak{p}^2\mathfrak{q}\mathfrak{r}$. Los factores son 'ideales' porque no están en el anillo $\mathbb{Z}\big[\sqrt{-5}\,\big]$, pero se comportan como si lo estuviesen.

En realidad esta factorización del 6 en $\mathbb{Z}\big[\sqrt{-5}\,\big]$ puede obtenerse mecánicamente aplicando el teorema siguiente:

Teorema 8.37 *Sea K un cuerpo numérico y supongamos que el anillo de los enteros de K es de la forma $\mathbb{Z}[\alpha]$, para un entero algebraico α. Sea $g(x) = \operatorname{pol\,mín}(\alpha, \mathbb{Q})$ y p un primo racional. Sea $\bar{g}(x)$ la imagen de $g(x)$ por el epimorfismo de $\mathbb{Z}[x]$ sobre $(\mathbb{Z}/p\mathbb{Z})[x]$ y sea $\bar{g} = \bar{g}_1^{e_1} \cdots \bar{g}_r^{e_r}$ la descomposición de $\bar{g}$ en polinomios mónicos irreducibles en $(\mathbb{Z}/p\mathbb{Z})[x]$. Entonces los ideales $\mathfrak{p}_i = \big(p, g_i(\alpha)\big)$, para $i = 1, \ldots, r$ son primos distintos y la descomposición de p en primos es $p = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$.*

[1] Véase la tabla 10.1 de [ITAl] y la justificación de su primera línea en la misma página.

DEMOSTRACIÓN: Para cada $i = 1, \dots, r$, sea α_i una raíz de $\bar{g}_i(x)$ en una extensión de $\mathbb{Z}/p\mathbb{Z}$. Entonces $(\mathbb{Z}/p\mathbb{Z})(\alpha_i)$ es una extensión finita de $\mathbb{Z}/p\mathbb{Z}$ y $\bar{g}_i = \operatorname{pol\,mín}(\alpha_i, \mathbb{Z}/p\mathbb{Z})$.

Consideremos la aplicación $\phi_i : \mathbb{Z}[\alpha] \longrightarrow (\mathbb{Z}/p\mathbb{Z})(\alpha_i)$ dada por $\phi_i\big(q(\alpha)\big) = \bar{q}(\alpha_i)$. Está bien definida, pues si $q(\alpha) = r(\alpha)$, entonces $(q-r)(\alpha) = 0$, luego $g \mid q - r$, de donde $\bar{g} \mid \bar{q} - \bar{r}$, y también $\bar{g}_i \mid \bar{q} - \bar{r}$, luego $\bar{q}(\alpha_i) - \bar{r}(\alpha_i) = 0$.

Obviamente ϕ_i es un epimorfismo, luego $\mathbb{Z}[\alpha]/\operatorname{N}(\phi_i) \cong (\mathbb{Z}/p\mathbb{Z})(\alpha_i)$, y el segundo anillo es un cuerpo, de donde $\operatorname{N}(\phi_i)$ es un ideal primo de $\mathbb{Z}[\alpha]$.

Es claro que $\big(p, g_i(\alpha)\big) \subset \operatorname{N}(\phi_i)$ (la imagen de p es $[p] = 0$). Veamos la otra inclusión. Si $q(\alpha) \in \operatorname{N}(\phi_i)$, entonces $\bar{q}(\alpha_i) = 0$, luego $\bar{q}(x) = \bar{h}(x)\bar{g}_i(x)$. El hecho de que $\bar{q}(x) - \bar{h}(x)\bar{g}_i(x) = 0$ significa que todos los coeficientes del polinomio $q(x) - h(x)g_i(x)$ son múltiplos de p. Consecuentemente

$$q(\alpha) = \big(q(\alpha) - h(\alpha)g_i(\alpha)\big) + h(\alpha)g_i(\alpha) \in \big(p, g_i(\alpha)\big).$$

Por lo tanto, $\mathfrak{p}_i = \big(p, g_i(\alpha)\big) = \operatorname{N}(\phi_i)$ es un ideal primo que obviamente divide a p.

Aplicando que, en general, $(p,u)(p,v) \subset (p,uv)$ concluimos que

$$\mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r} \subset \big(p, g_1(\alpha)^{e_1} \cdots g_r(\alpha)^{e_r}\big) = \big(p, g(\alpha)\big) = (p,0) = (p).$$

Nótese que la primera igualdad se debe a que $g(\alpha)$ y $g_1(\alpha)^{e_1} \cdots g_r(\alpha)^{e_r}$ se diferencian en un entero múltiplo de p.

Así pues, $p \mid \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$. La igualdad la obtendremos considerando las normas.

Por definición de norma, $\operatorname{N}(\mathfrak{p}_i) = |\mathbb{Z}[\alpha]/\mathfrak{p}_i| = \big|(\mathbb{Z}/p\mathbb{Z})(\alpha_i)\big| = p^{\operatorname{grad} g_i}$, pues $(\mathbb{Z}/p\mathbb{Z})(\alpha_i)$ es un espacio vectorial de dimensión $\operatorname{grad} g_i$ sobre $\mathbb{Z}/p\mathbb{Z}$, luego es isomorfo al espacio $(\mathbb{Z}/p\mathbb{Z})^{\operatorname{grad} g_i}$.

En total $\operatorname{N}(\mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}) = p^{e_1 \operatorname{grad} g_1 + \cdots + e_r \operatorname{grad} g_r} = p^n$, donde n es el grado de K. Así pues $\operatorname{N}(\mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}) = \operatorname{N}(p)$, lo que nos da que $p = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$.

Los primos $\mathfrak{p}_i$ son distintos, pues si $\mathfrak{p}_i = \mathfrak{p}_j$, entonces $\mathfrak{p}_i \mid g_j(\alpha)$, luego los polinomios $\bar{g}_i$ y $\bar{g}_j$ tienen la raíz $[\alpha]$ en común en el cuerpo $\mathbb{Z}[\alpha]/\mathfrak{p}_i$ (este cuerpo tiene característica p, luego contiene a $\mathbb{Z}/p\mathbb{Z}$), pero eso es imposible porque ambos polinomios son irreducibles en $\mathbb{Z}/p\mathbb{Z}$, luego son primos entre sí. ∎

La hipótesis $\mathcal{O}_K = \mathbb{Z}[\alpha]$ no la cumplen todos los cuerpos numéricos, pero es bastante frecuente y sabemos que al menos la cumplen los cuerpos cuadráticos y ciclotómicos de orden primo.

Ejemplo Las factorizaciones de 2 y 3 en $\mathbb{Z}\big[\sqrt{-5}\,\big]$ que hemos obtenido más arriba pueden obtenerse también como sigue:

En primer lugar, $\operatorname{pol\,mín}(\sqrt{-5}, \mathbb{Q}) = x^2 + 5$. Su imagen en el cuerpo $(\mathbb{Z}/2\mathbb{Z})[x]$ es $x^2 + 1 = (x+1)^2$, luego 2 factoriza como $2 = \big(2, 1 + \sqrt{-5}\,\big)^2$.

La imagen en $(\mathbb{Z}/3\mathbb{Z})[x]$ es $x^2 + 2 = x^2 - 1 = (x+1)(x-1)$, lo que nos da la factorización

$$3 = \big(3, 1 + \sqrt{-5}\,\big)\big(3, -1 + \sqrt{-5}\,\big) = \big(3, 1 + \sqrt{-5}\,\big)\big(3, 1 - \sqrt{-5}\,\big). \quad ∎$$

En el capítulo siguiente estudiaremos con detalle las factorizaciones en cuerpos cuadráticos. Veamos ahora un resultado general para cuerpos ciclotómicos.

Teorema 8.38 *Sea p un primo racional impar y sea $\mathcal{O} = \mathbb{Z}[\omega]$ el anillo de los enteros ciclotómicos de orden p.*

1. *La factorización de p en $\mathcal{O}$ es $p = \mathfrak{p}^{p-1}$, donde $\mathfrak{p} = (\omega - 1)$.*
2. *Si $q \neq p$ es un primo racional, entonces la factorización de q es de la forma*
$$q = \mathfrak{p}_1 \cdots \mathfrak{p}_r,$$
donde los primos son distintos, $\mathrm{N}(\mathfrak{p}_i) = q^f$, con $f = o_p(q)$ (el orden de q módulo p), y $r = (p-1)/f$.

DEMOSTRACIÓN: 1) En la prueba del teorema 8.18 vimos que $\pi = \omega - 1$ cumple $\pi^{p-1} \mid p$, luego $\mathfrak{p}^{p-1} \mid p$ y considerando las normas tenemos la igualdad.

2) Es consecuencia inmediata del teorema anterior y del teorema 5.86. ∎

Ejemplo Vamos a considerar el caso $p = 23$ y $q = 47$ en el teorema anterior. Como $q \equiv 1$ (mód p), tenemos que $f = o_p(q) = 1$, luego 47 factoriza en 22 primos distintos de norma 47. Vamos a probar que en $\mathbb{Z}[\omega]$ no hay elementos de norma ± 47, con lo que los factores primos de 47 serán ideales no principales, y habremos probado que $\mathbb{Z}[\omega]$ no tiene factorización única.

El discriminante del cuerpo es $\Delta = -23^{21}$. Si llamamos $\sigma_1, \ldots, \sigma_{22}$ a los monomorfismos de $\mathbb{Q}(\omega)$, como $\mathbb{Q}(\omega)/\mathbb{Q}$ es normal concluimos que todos los conjugados $\sigma_i(\omega^j)$ están en $\mathbb{Q}(\omega)$, luego

$$\sqrt{\Delta} = 23^{10}\sqrt{-23} = \det\big(\sigma_i(\omega^j)\big) \in \mathbb{Q}(\omega),$$

y de aquí concluimos que $\mathbb{Q}\big(\sqrt{-23}\,\big) \subset \mathbb{Q}(\omega)$.

Si en $\mathbb{Q}(\omega)$ hubiera un entero de norma ± 47, la norma de dicho entero respecto a la extensión $\mathbb{Q}(\omega)/\mathbb{Q}\big(\sqrt{-23}\,\big)$ sería un entero cuadrático de norma ± 47 (necesariamente $+47$). Basta ver, pues, que en $\mathbb{Q}\big(\sqrt{-23}\,\big)$ no hay enteros de norma 47.

Ahora bien, un entero de $\mathbb{Q}\big(\sqrt{-23}\,\big)$ es de la forma $a + b\frac{1+\sqrt{-23}}{2}$, con a, b enteros racionales, y su norma es

$$\begin{aligned}\mathrm{N}\left(a + b\frac{1+\sqrt{-23}}{2}\right) &= \left(\frac{2a+b}{2} + b\frac{\sqrt{-23}}{2}\right)\left(\frac{2a+b}{2} - b\frac{\sqrt{-23}}{2}\right)\\ &= \frac{1}{4}\big((2a-b)^2 + 23b^2\big).\end{aligned}$$

Si hubiera un elemento de norma 47 tendríamos

$$188 = 47 \cdot 4 = (2a-b)^2 + 23b^2,$$

pero 188 no es un cuadrado perfecto, ni $188 - 23 = 165$, ni $188 - 23 \cdot 4 = 96$, luego b no puede tomar los valores $0, \pm 1, \pm 2$, y para valores mayores resulta que $(2a-b)^2 + 23b^2 > 188$. ∎

En su estudio de los enteros ciclotómicos de orden primo, Kummer publicó en 1844 las descomposiciones en factores primos de todos los primos racionales $\leq 1\,000$ para las extensiones de orden primo $p \leq 19$. Sucede que todas estas extensiones son de hecho dominios de factorización única, por lo que Kummer pudo encontrar todas las factorizaciones, no sin grandes esfuerzos. Para el siguiente caso, $p = 23$, encontró las factorizaciones de todos los primos menores que 47 y demostró que 47 no podía ser descompuesto en primos, con lo que halló el menor contraejemplo a la factorización única en enteros ciclotómicos de orden primo. Afortunadamente su teoría estaba tan avanzada que Kummer confió más en ella que en la evidencia que le mostraba que los cuerpos ciclotómicos no tenían factorización única, y así descubrió la factorización ideal de los anillos de enteros algebraicos.

El teorema 8.37 puede refinarse cuando se aplica a extensiones de Galois de $\mathbb{Q}$. Ello se debe esencialmente a que los automorfismos obligan a que las factorizaciones presenten un alto grado de simetría. En efecto, ante todo, si K es una extensión finita de Galois de $\mathbb{Q}$ y $\sigma \in G(K/\mathbb{Q})$, es claro que la imagen $\sigma[\mathfrak{a}]$ de un ideal fraccional cualquiera de K es de nuevo un ideal fraccional, que será un ideal (entero) si y sólo si lo es $\mathfrak{a}$. Así pues, podemos extender a σ a un automorfismo del grupo de los ideales fraccionales de K dado por $\sigma(\mathfrak{a}) = \sigma[\mathfrak{a}]$. Decimos 'extender' porque la acción sobre los ideales es consistente con la acción sobre elementos reales en el sentido de que $\sigma\big((\alpha)\big) = \big(\sigma(\alpha)\big)$, para todo $\alpha \in K$ no nulo.

Diremos que dos ideales fraccionales $\mathfrak{a}$ y $\mathfrak{b}$ son *conjugados* si existe un automorfismo $\sigma \in G(K/\mathbb{Q})$ tal que $\sigma(\mathfrak{a}) = \mathfrak{b}$.

Teorema 8.39 *Sea K una extensión de Galois de grado n sobre $\mathbb{Q}$ y sea p un primo racional. Entonces la factorización de p en K es de la forma*

$$p = (\mathfrak{p}_1 \cdots \mathfrak{p}_r)^e,$$

donde los ideales $\mathfrak{p}_i$ son primos distintos, forman una clase de conjugación y todos tienen la misma norma $\mathrm{N}(\mathfrak{p}_i) = p^f$, para un cierto f tal que $efr = n$.

Demostración: Es obvio que si $\mathfrak{p} \mid p$, entonces todo conjugado de $\mathfrak{p}$ cumple lo mismo. Veamos que cualquier otro divisor $\mathfrak{q}$ de p es un conjugado de $\mathfrak{p}$. Supongamos, por reducción al absurdo, que $\sigma(\mathfrak{p}) \neq \mathfrak{q}$ para todo automorfismo σ. Por el teorema chino del resto existe un $\alpha \in \mathcal{O}_K$ tal que

$$\begin{aligned} \alpha &\equiv 0 \ (\text{mód } \mathfrak{q}), \\ \alpha &\equiv 1 \ (\text{mód } \sigma(\mathfrak{p})) \ \text{ para todo } \sigma \in G(K/\mathbb{Q}). \end{aligned}$$

Pero entonces $\mathfrak{q} \mid \alpha \mid \mathrm{N}(\alpha)$, luego $p \mid \mathrm{N}(\alpha)$, luego $\mathfrak{p} \mid \mathrm{N}(\alpha)$ y por consiguiente $\mathfrak{p} \mid \sigma(\alpha)$ para algún $\sigma \in G(K/\mathbb{Q})$, de donde $\sigma^{-1}(\mathfrak{p}) \mid \alpha$, contradicción.

Es claro que el exponente de un primo $\mathfrak{p}$ en la descomposición en primos de p debe ser el mismo que el de todos sus conjugados. Como todos los divisores primos de p son conjugados, de hecho todos tienen el mismo exponente e, luego la

factorización es del tipo indicado en el enunciado. También es obvio que primos conjugados tienen la misma norma, necesariamente potencia de p. La igualdad $n = efr$ se sigue de tomar normas en ambos miembros de la factorización. ∎

En particular, el teorema anterior afirma que dos primos de un cuerpo numérico normal son conjugados si y sólo si dividen al mismo primo racional, si y sólo si tienen la misma norma. En general no tiene por qué ser así:

Ejemplo Factorización en el anillo $\mathbb{Z}\big[\sqrt[3]{2}\,\big]$.

Vamos a determinar cómo se descompone un primo racional p en el anillo de enteros del cuerpo $\mathbb{Q}\big(\sqrt[3]{2}\,\big)$. Para aplicar el teorema 8.37 hemos de considerar el polinomio $x^3 - 2$.

Si $p = 2$ tenemos que $x^3 - 2 \equiv x^3$ (mód 2), luego la factorización es de la forma $2 = \mathfrak{p}^3$, con $\mathrm{N}(\mathfrak{p}) = 2$ (De hecho $2 = \sqrt[3]{2}^{\,3}$, y para esto no hace falta el teorema 8.37).

Supongamos que p es impar y sea $G = (\mathbb{Z}/p\mathbb{Z})^*$. Consideremos el homomorfismo de grupos $\phi : G \longrightarrow G$ dado por $\phi(u) = u^3$. Su núcleo está formado por las raíces cúbicas de la unidad de $\mathbb{Z}/p\mathbb{Z}$. Puede haber una o tres de ellas. Concretamente, $\mathbb{Z}/p\mathbb{Z}$ tiene tres raíces cúbicas de la unidad si y sólo si $p \equiv 1$ (mód 3). En efecto:

Si $u \in G$ es una raíz cúbica de la unidad distinta de 1, entonces $o(u) = 3$, y por el teorema de Lagrange $3 \mid p - 1$, luego $p \equiv 1$ (mód 3).

Si $p \equiv 1$ (mód 3) y v es una raíz primitiva de la unidad módulo p, entonces $u = v^{(p-1)/3}$ es una raíz cúbica de la unidad distinta de 1.

Por lo tanto, si $p \not\equiv 1$ (mód 3) el núcleo de ϕ es trivial, luego ϕ es un monomorfismo, luego un isomorfismo. Por lo tanto [2] tiene una única antiimagen por ϕ, una única raíz cúbica módulo p, luego $x^3 - 2$ se descompone en un factor de grado 1 y otro de grado 2, y en consecuencia $p = \mathfrak{pq}$, donde $\mathrm{N}(\mathfrak{p}) = p$ y $\mathrm{N}(\mathfrak{q}) = p^2$.

Si $p \equiv 1$ (mód 3) entonces, el núcleo de ϕ tiene tres elementos, con lo que o bien $[2] \in \operatorname{Im}\phi$, y entonces 2 tiene tres raíces cúbicas módulo p, o bien $[2] \notin \operatorname{Im}\phi$, y entonces 2 no tiene raíces cúbicas módulo p.

En el primer caso la factorización es $p = \mathfrak{pqr}$, con los tres factores de norma p, y en el segundo p se conserva primo. Resumimos los resultados en la tabla siguiente:

Primo			Factorización	Norma
$p = 2$			$\mathfrak{p}^3$	p
$p \equiv 1$ (mód 3)	$x^3 \equiv 2$ (mód p)	resoluble	$\mathfrak{pqr}$	p
	$x^3 \equiv 2$ (mód p)	no resoluble	p	p^3
$p \not\equiv 1$ (mód 3)	$p \neq 2$		$\mathfrak{pq}$	p / p^2

8.5 El grupo de clases

El problema más importante que presenta la factorización ideal es el de determinar en qué casos un divisor ideal se corresponde con un divisor real (o sea, cuándo es principal) y, en particular, cuándo todos los divisores son reales, y por lo tanto el anillo es DFU. Para abordar este problema introducimos un concepto fundamental en el estudio de los enteros algebraicos.

Definición 8.40 Sea K un cuerpo numérico y $\mathcal{O}$ su anillo de enteros. Sea $\mathcal{F}$ el grupo de los ideales fraccionales de $\mathcal{O}$ y $\mathcal{P}$ el subgrupo formado por los ideales fraccionales principales, es decir, $\mathcal{P} = \big\{(a)(b)^{-1} \mid a, b \in \mathcal{O} \setminus \{0\}\big\}$. Llamaremos *grupo de clases* de K al grupo cociente $\mathcal{H} = \mathcal{F}/\mathcal{P}$.

Notemos que todo ideal fraccional es de la forma $\mathfrak{a}(b)^{-1}$, donde $\mathfrak{a}$ es un ideal, luego al tomar clases módulo $\mathcal{P}$ resulta que $\big[\mathfrak{a}(b)^{-1}\big] = [\mathfrak{a}]$, es decir, que podemos considerar a los elementos de $\mathcal{H}$ como clases de ideales, en el sentido de que siempre podemos trabajar con representantes ideales.

Por otra parte, si un ideal $\mathfrak{c}$ está en $\mathcal{P}$, o sea, si $\mathfrak{c} = (a)(b)^{-1}$, entonces $(a) = (b)\mathfrak{c}$, luego $(b) \mid (a)$, luego $b \mid a$, luego $a = bc$ para cierto entero c, y $(b)\mathfrak{c} = (a) = (b)(c)$. Por lo tanto $\mathfrak{c} = (c)$. Esto prueba que $[\mathfrak{c}] = 1$ si y sólo si $\mathfrak{c}$ es principal.

En consecuencia $\mathcal{O}$ es un DIP si y sólo si $\mathcal{H} = 1$. Puede probarse, aunque ello excede nuestras posibilidades, que el grupo $\mathcal{H}$ siempre es finito, y a su número de elementos se le llama *número de clases* de K, y se representa por h. En estos términos $\mathcal{O}$ es DIP si y sólo si $h = 1$.

En [ITAl] demostramos la finitud del grupo de clases de los cuerpos cuadráticos (véase la nota al final de la sección [ITAl 13.2]) y aquí vamos a probar la finitud del número de clases en el caso de los cuerpos ciclotómicos de orden primo.

Diremos que dos ideales $\mathfrak{a}$ y $\mathfrak{b}$ son *similares* ($\mathfrak{a} \approx \mathfrak{b}$) si son congruentes módulo $\mathcal{P}$. Concretamente, $\mathfrak{a} \approx \mathfrak{b}$ si y sólo si existen enteros algebraicos a y b tales que $\mathfrak{b} = (a)(b)^{-1}\mathfrak{a}$, o equivalentemente, $(a)\mathfrak{a} = (b)\mathfrak{b}$.

Notemos que si $\mathfrak{a}$, $\mathfrak{b}$ y $\mathfrak{c}$ son ideales no nulos tales que $\mathfrak{ac}$ y $\mathfrak{bc}$ son principales, entonces $\mathfrak{a} \approx \mathfrak{b}$, pues módulo $\mathcal{P}$ tenemos $[\mathfrak{ac}] = 1 = [\mathfrak{bc}]$, luego $[\mathfrak{a}] = [\mathfrak{b}]$.

En el teorema siguiente consideramos a $\mathbb{Q}(\omega)$ como subcuerpo del cuerpo $\mathbb{C}$ de los números complejos:

Teorema 8.41 *Si p es un primo racional y $\mathbb{Q}(\omega)$ es el cuerpo ciclotómico de orden p, entonces el grupo de clases de K es finito.*

Demostración: Un entero de K es de la forma

$$\alpha = g(\omega) = a_0 + a_1\omega + \cdots + a_{p-2}\omega^{p-2},$$

donde los coeficientes a_i son enteros racionales.

Entonces $\mathrm{N}(\alpha) = g(\omega)\cdots g(\omega^{p-1})$ o, agrupando los pares de conjugados complejos ω, ω^{p-1} y ω^2, ω^{p-2}, etc., vemos que

$$\mathrm{N}(\alpha) = |g(\omega)|^2\cdots|g(\omega^{(p-1)/2})|^2,$$

y claramente $|g(\omega^i)|^2 \le (|a_0| + \cdots + |a_{p-2}|)^2$. En particular, si todos los coeficientes a_i están acotados en la forma $|a_i| \le c$, tenemos que

$$\mathrm{N}(\alpha) \le ((p-1)^2c^2)^{(p-1)/2} = (p-1)^{p-1}c^{p-1}.$$

Fijemos ahora un ideal no nulo $\mathfrak{a}$ del anillo $\mathcal{O} = \mathbb{Z}[\omega]$, sea $\mathfrak{a}'$ otro ideal tal que $[\mathfrak{a}] = [\mathfrak{a}']^{-1}$ en el grupo de clases. Sea $n = \mathrm{N}(\mathfrak{a}')$ y sea c el menor número natural tal que $(c+1)^{p-1} > n$, de modo que $c^{p-1} \le n$. Como $|\mathcal{O}/\mathfrak{a}'| = n$ y el número de $p-1$-tuplas $(a_0,\ldots,a_{p-2})$ de números naturales tales que $0 \le a_i \le c$ es mayor que n, tiene que haber dos de ellas tales que los enteros ciclotómicos que determinan, digamos $g_1(\omega)$ y $g_2(\omega)$, sean congruentes módulo $\mathfrak{a}'$. Entonces $\alpha = g(\omega) = g_1(\omega) - g_2(\omega)$ cumple que $\alpha \in \mathfrak{a}'$ y

$$\mathrm{N}(\alpha) \le (p-1)^{p-1}c^{p-1} \le (p-1)^{p-1}\,\mathrm{N}(\mathfrak{a}').$$

La condición $\alpha \in \mathfrak{a}'$ equivale a $\mathfrak{a}' \mid (\alpha)$, de modo que existe un ideal $\mathfrak{b}$ tal que $(\alpha) = \mathfrak{a}'\mathfrak{b}$ y la desigualdad precedente implica que $\mathrm{N}(\mathfrak{b}) \le (p-1)^{p-1}$, y por otra parte tenemos que $\mathfrak{b} \approx \mathfrak{a}$.

Así pues, hemos probado que todo ideal es similar a otro de norma menor o igual que la constante $(p-1)^{p-1}$. Como sólo hay un número finito de ideales de una norma dada (teorema 8.36), concluimos que el número de clases es finito. ∎

En particular podemos hablar del número de clases h de los cuerpos ciclotómicos de orden primo. Sabemos que $h = 1$ en los casos $p = 3, 5$, pues los enteros ciclotómicos de orden 3 son los enteros de Eisenstein y en [ITAl 6.3] hemos probado que son un dominio euclídeo, mientras que el caso $p = 5$ está tratado en [ITAl 17.3].

8.6 El teorema de Kummer

Llegados a este punto estamos en condiciones de demostrar una parte significativa de la teoría de Kummer sobre el Último Teorema de Fermat. Para ello empezamos con algunas observaciones elementales sobre el anillo $\mathbb{Z}[\omega]$ de los enteros ciclotómicos de orden primo $p > 2$:

Al hacer $x = 1$ en la factorización

$$(x-\omega)\cdots(x-\omega^{p-1}) = 1 + x + \cdots x^{p-1}$$

obtenemos que

$$(1-\omega)\cdots(1-\omega^{p-1}) = p.$$

El miembro izquierdo es la norma de cualquiera de los factores, luego por el teorema 8.35 b) todos ellos son primos ciclotómicos. Más aún, si $\pi = \omega - 1$,

tomando congruencias módulo π es inmediato que divide a todos los factores, luego tenemos la descomposición en factores primos $p = \eta\pi^{p-1}$, donde η es una unidad ciclotómica.

Observemos ahora que las únicas raíces de la unidad contenidas en $\mathbb{Z}[\omega]$ son las de la forma $\pm\omega^n$.

En efecto, Si $\zeta \in \mathbb{Z}[\omega]$ es una raíz de la unidad de orden r, entonces $\zeta\omega$ tiene orden $\mathrm{mcm}(p, r) = p^i s$, donde $p \nmid s$. Como $\mathbb{Q}(\zeta\omega) \subset \mathbb{Q}(\omega)$, al tomar grados obtenemos que $\phi(p^i s) = (p-1)p^{i-1}\phi(s) \mid p-1$, luego $i = \phi(s) = 1$, luego $s = 1, 2$, luego $r \mid 2r$, luego $\zeta = \omega^n$.

Compárese la prueba del teorema siguiente con la dada en [ITAl 17.19]:

Teorema 8.42 (Lema de Kummer) *Sea ω una raíz p-ésima primitiva de la unidad y $\epsilon \in \mathbb{Z}[\omega]$ una unidad. Entonces $\epsilon = \omega^n\eta$, donde $\eta \in \mathbb{Z}[\omega]$ es una unidad real.*

Demostración: Sea $\alpha = \epsilon/\bar{\epsilon} \in \mathbb{Z}[\omega]$, donde la barra representa la conjugación compleja. Si $\sigma \in G(\mathbb{Q}(\omega)/\mathbb{Q})$, teniendo en cuenta que la conjugación compleja es también un elemento del grupo de Galois y que éste es abeliano, vemos que $\sigma(\alpha^n) = \sigma(\epsilon^n)/\overline{\sigma(\epsilon^n)}$. Por lo tanto $|\sigma(\epsilon^n)| = 1$, donde las barras representan el valor absoluto complejo. El polinomio mínimo de α^n sobre $\mathbb{Q}$ tiene coeficientes enteros y todas sus raíces tienen módulo 1. Si su grado es $k \le p-1$, la fórmula de Vieta 7.4 implica que el coeficiente de x^r en dicho polinomio está acotado por el número combinatorio $\binom{k}{r}$. Por consiguiente, los polinomios mínimos de las potencias α^n sólo pueden variar en un conjunto finito, luego las potencias de α son un número finito, luego $\alpha^m = \alpha^n$ para dos exponentes distintos, luego $\alpha^{m-n} = 1$, luego α es una raíz de la unidad. Por la observación precedente al teorema, $\alpha = \pm\omega^n$, para cierto n.

Equivalentemente, $\epsilon = \pm\omega^n\bar{\epsilon}$. Veamos que el signo negativo es imposible. Si fuera el caso, como $\omega \equiv 1 \pmod{\pi}$, vemos que $\epsilon \equiv -\bar{\epsilon} \pmod{\pi}$. Por otro lado, si expresamos $\epsilon = r(\omega)$ como polinomio en ω con coeficientes enteros y $\bar{\epsilon} = r(\omega^{-1})$, al tomar congruencias módulo π resulta que $\epsilon \equiv \bar{\epsilon} \pmod{\pi}$ (pues ambos son congruentes con la suma de los coeficientes de $r(x)$). Esto implica que $\epsilon \equiv 0 \pmod{\pi}$, es decir, que $\pi \mid \epsilon$, contradicción.

Así pues, $\epsilon = \omega^n\bar{\epsilon}$. Tomamos i tal que $n \equiv 2i \pmod{p}$, con lo que $\epsilon = \omega^{2i}\bar{\epsilon}$, luego $\eta = \epsilon\bar{\omega}^i = \bar{\epsilon}\omega^i = \bar{\eta}$. Concluimos que $\eta \in \mathbb{Z}[\omega]$ es una unidad real tal que $\epsilon = \omega^{-i}\eta$. ∎

También se llama Lema de Kummer a un resultado que Kummer demostró bajo una hipótesis algo más general que la factorización única de $\mathbb{Z}[\omega]$, pero que nosotros probaremos únicamente para el caso $p = 5$ (se trata de [ITAl 17.26]), ya que el caso general no es nada trivial.

Observemos antes que, como $\omega + \omega^4 = 2\cos(2\pi/5) = \frac{-1+\sqrt{5}}{2}$, es claro que $\mathbb{Q}(\sqrt{5}) \subset \mathbb{Q}(\omega)$, luego $\mathbb{Q}(\sqrt{5}) = \mathbb{Q}(\omega) \cap \mathbb{R}$, luego el teorema anterior implica que toda unidad de $\mathbb{Z}[\omega]$ es de la forma $\epsilon = \omega^n\eta$, donde η es una unidad en $\mathbb{Z}[\sqrt{5}]$,

pero (véase [ITAl 9.7] y el ejemplo posterior) toda unidad de $\mathbb{Z}[\sqrt{5}]$ es de la forma:

$$\eta = \pm\left(\frac{1+\sqrt{5}}{2}\right)^m = \pm(1+\omega+\omega^4)^m,$$

luego $\epsilon = \pm\omega^n(1+\omega+\omega^4)^m$.

Teorema 8.43 (Lema de Kummer) *Sea $\mathbb{Z}[\omega]$ el anillo de los enteros ciclotómicos de orden 5. Una unidad $\epsilon \in \mathbb{Z}[\omega]$ es una potencia quinta si y sólo si es congruente módulo 5 con un entero (racional).*

DEMOSTRACIÓN: Una implicación es trivial: si $\epsilon = p(\omega)$, con $p(x) \in \mathbb{Z}[x]$, entonces $\epsilon^5 \equiv p(\omega^5) = p(1)$ (mód 5). Supongamos ahora que $\epsilon \equiv a$ (mód 5), para cierto $a \in \mathbb{Z}$. Por la observación precedente al teorema tenemos que

$$\epsilon = \pm\omega^n(1+\omega+\omega^4)^m.$$

Como $\pi^2 = \omega^2 - 2\omega + 1$, tenemos que $\omega^2 \equiv 2\omega - 1$ (mód π^2), luego

$$1+\omega+\omega^4 \equiv 1+\omega+(2\omega-1)^2 = 4\omega^2-3\omega+2 \equiv 8\omega-4-3\omega+2 \equiv -2 \text{ (mód } \pi^2).$$

Por otra parte $\omega^n = (1+\pi)^n \equiv 1+n\pi$ (mód π^2), luego en total

$$a \equiv \pm(-2)^m(1+n\pi) \text{ (mód } \pi^2).$$

A su vez, $a \equiv \pm(-2)^m$ (mód π) y, como ambos miembros son enteros, de hecho $a \equiv \pm(-2)^m$ (mód 5), luego $n\pi \equiv 0$ (mód π^2), luego $\pi \mid n$, luego $5 \mid n$, luego $\omega^n = 1$. En definitiva,

$$\epsilon = (\pm(1+\omega+\omega^4))^{5k}(1+\omega+\omega^4)^j,$$

donde $0 \le j < 5$. Tenemos que probar que $j = 0$. Ahora bien

$$(1+\omega+\omega^4)^5 \equiv 1+\omega^5+\omega^{20} \equiv 3 \text{ (mód } 5),$$

luego $(1+\omega+\omega^4)^j = (-\omega^2-\omega^3)^j$ es congruente con un entero módulo 5, pero una simple comprobación muestra que esto es falso para $j = 1, 2, 3, 4$:

$$(-\omega^2-\omega^3)^2 = 1-\omega-\omega^3, \quad (-\omega^2-\omega^3)^3 = 1-2\omega^2-2\omega^3,$$

$$(-\omega^2-\omega^3)^4 = -4-3\omega^2-3\omega^3$$

y ninguno de éstos números es congruente con un entero módulo 5 (para ello sería necesario que todos los coeficientes menos el término independiente fueran múltiplos de 5). ∎

Definición 8.44 Sea p un primo impar y $\mathbb{Z}[\omega]$ el anillo de los enteros ciclotómicos de orden p. Diremos que p es *regular* si cumple las dos condiciones siguientes:

A) p no divide al número de clases h de $\mathbb{Q}(\omega)$.

B) Una unidad $\epsilon \in \mathbb{Z}[\omega]$ es una potencia p-ésima si y sólo si es congruente módulo p con un entero (racional).

Observemos que la propiedad A) se cumple trivialmente si $\mathbb{Z}[\omega]$ es un dominio de ideales principales. Kummer demostró que la propiedad A) implica de hecho la propiedad B), por lo que ésta es en realidad redundante en la definición.

Una consecuencia elemental de la propiedad A) es que si $\mathfrak{a}$ es un ideal de $\mathbb{Z}[\omega]$ tal que $\mathfrak{a}^p$ es principal, entonces $\mathfrak{a}$ es principal, pues la hipótesis es que $[\mathfrak{a}]^p = 1$ en el grupo de clases y, tomando enteros tales que $up + vh = 1$, vemos que $[\mathfrak{a}] = ([\mathfrak{a}]^p)^u([\mathfrak{a}]^h)^v = 1$.

El teorema anterior junto con [ITAl 17.3] implica que $p = 5$ es un primo regular, y es muy fácil comprobar que $p = 3$ también lo es.

La prueba del teorema siguiente es esencialmente la misma que la de [ITAl 17.27], salvo que hemos sustituido la exigencia de que el anillo de enteros ciclotómicos tenga factorización única real por la hipótesis mucho más débil de que p no divida al número de clases:

Teorema 8.45 (Kummer) *El último teorema de Fermat es cierto para exponentes regulares.*

DEMOSTRACIÓN: Sea p un primo regular y supongamos que existen enteros no nulos tales que $x^p + y^p = z^p$. En [ITAl 6.6] probamos el teorema de Fermat para exponente 3, así que podemos suponer que $p \geq 5$. También podemos suponer que x, y, z son primos entre sí dos a dos, ya que si un primo q divide a dos de ellos, la ecuación hace que divida al tercero, y entonces $(x/q, y/q, z/q)$ es también una solución. Repitiendo este proceso podemos eliminar todos los primos comunes.

En particular, esto determina los dos casos que clásicamente se han considerado al tratar el Último Teorema de Fermat. El caso I se da cuando $p \nmid xyz$ y el caso II cuando $p \mid z$. (Notemos que si p divide, por ejemplo, a x, entonces $(x', y', z') = (-z, y, -x)$ es otra solución en la que $p \mid z'$, luego si $p \mid xyz$, siempre podemos reordenar la solución para que sea $p \mid z$.)

Vamos a tratar ambos casos por separado. Empezamos considerando el caso I. Si ω es una raíz p-ésima primitiva de la unidad, tenemos la factorización

$$z^p = x^p + y^p = (x + y)(x + \omega y) \cdots (x + \omega^{p-1} y).$$

Se cumple que los factores son primos entre sí dos a dos. En efecto, si un primo $\mathfrak{q}$ dividiera a $x + \omega^i y$ y a $x + \omega^j y$, entonces dividiría también a

$$x + \omega^i y - (x - \omega^j y) = (\omega^i - \omega^j)y = \omega^i(1 - \omega^{j-i})y$$

así como a

$$x + \omega^i y - \omega^{i-j}(x + \omega^j y) = (1 - \omega^{i-j})x,$$

y al principio de la sección hemos visto que los enteros $(1 - \omega^k)$ son primos asociados a π, luego necesariamente $\mathfrak{q} = (\pi)$, luego $\pi \mid z$, luego $p \mid z$, en contradicción con la hipótesis de que estamos en el caso I.

Por la factorización única en ideales, cada ideal $(x + \omega^i y)$ es una potencia p-ésima, luego por la propiedad A) de la definición de primo regular podemos concluir que $x + \omega y = \epsilon\beta^p$, para una cierta unidad ϵ y un entero ciclotómico β.

Vamos a llegar a una contradicción tan sólo a partir de aquí, sin necesidad de usar la condición B). Para ello aplicamos la conjugación que envía ω a ω^{-1} (que no es sino la conjugación compleja). Así obtenemos que $x + \omega^{-1}y = \bar{\epsilon}\bar{\beta}^p$.

Del teorema 8.42 se sigue que $\epsilon/\bar{\epsilon} = \omega^r$, donde $0 \le r < p$. Por la parte trivial del teorema 8.43 (que claramente vale todo p y todo entero ciclotómico), sabemos que toda potencia p-ésima es congruente módulo p con un entero racional, luego $\bar{\beta}^p \equiv m \pmod p$, de donde se sigue que $\beta^p \equiv \bar{\beta}^p \pmod p$. Reuniendo todo esto vemos que

$$x + \omega^{-1}y = \bar{\epsilon}\bar{\beta}^p = \omega^{-r}\epsilon\bar{\beta}^p \equiv \omega^{-r}\epsilon\beta^p = \omega^{-r}(x + \omega y) \pmod p.$$

Equivalentemente:

$$x\omega^r + y\omega^{r-1} - y\omega - x \equiv 0 \pmod p. \tag{8.10}$$

Notemos que si p divide a un entero ciclotómico α y tenemos una expresión de α como combinación lineal entera de $p-1$ potencias de ω, como éstas son una base entera, es necesario que p divida a cada uno de los coeficientes. Usaremos esto para probar que $r = 1$ descartando cualquier otra posibilidad (notemos también que podemos suponer $p \ge 5$, ya que el caso $p = 3$ está probado).

Si $r = 0$ la congruencia (8.10) se convierte en $y\omega^{-1} - y\omega \equiv 0 \pmod p$, luego $p \mid y$, contradicción.

Si $r = 2$ queda $x\omega^2 - x \equiv 0 \pmod p$, luego $p \mid x$, contradicción.

Si $r > 2$ todas las potencias de ω que aparecen en (8.10) son distintas, y como sólo hay $4 \le p - 1$, concluimos igualmente que $p \mid x$.

Así pues, ha de ser $r = 1$, y entonces (8.10) es $(x - y)\omega + y - x \equiv 0 \pmod p$, con lo que concluimos que $x \equiv y \pmod p$.

Ahora bien, si escribimos la ecuación de Fermat como $x^p + y^p + z^p = 0$, el caso I es simétrico respecto a tres variables x, y, z luego intercambiando los papeles podemos llegar igualmente a que $x \equiv y \equiv z \pmod p$.

Pero $0 = x^p + y^p + z^p \equiv x + y + x \equiv 3x \pmod p$, y como $p > 3$, llegamos una vez más a la contradicción $p \mid x$.

Supongamos ahora que $p \mid z$ (caso II). Sustituyamos z por $p^k z$, donde ahora z es primo con p. Tenemos entonces que $x^p + y^p = p^{kp}z^p$, donde x, y, z son enteros primos con p.

En el anillo de enteros ciclotómicos, p factoriza como $p = \eta(\omega - 1)^{p-1}$, donde η es una unidad. La ecuación se convierte en

$$x^p + y^p = \epsilon(\omega - 1)^{pm}z^p, \tag{8.11}$$

donde ϵ es una unidad y $m = k(p - 1) > 0$.

Hemos de probar que esta ecuación no tiene soluciones enteras primas con $\omega - 1$. Para ello probaremos más en general que no existen enteros ciclotómicos x, y, z primos con $\omega - 1$ que satisfagan (8.11). Supongamos por reducción al absurdo que existen enteros ciclotómicos que cumplan (8.11) con el menor valor posible para m. Factorizando el miembro izquierdo de (8.11) tenemos

$$(x + y)(x + \omega y) \cdots (x + \omega^{p-1} y) = \epsilon(\omega - 1)^{pm} z^p. \tag{8.12}$$

Como en el caso I, tenemos que $\omega - 1$ divide a las diferencias

$$x + \omega^i y - (x - \omega^j y) = (\omega^i - \omega^j)y = \omega^i(1 - \omega^{j-i})y$$

(porque $1-\omega^{j-1}$ es un primo asociado a $\omega-1$). Por otra parte, como $\omega-1$ divide al miembro derecho de (8.11), divide a al menos uno de los factores del miembro derecho, pero como divide a sus diferencias, de hecho los divide a todos.

Más aún, como $\omega - 1 \nmid y$, tenemos que $(\omega - 1)^2$ no divide a ninguna de las diferencias de los factores. Equivalentemente, los números

$$\frac{x + \omega^i y}{\omega - 1}$$

son primos entre sí dos a dos.

Sabemos que en el caso II el primo $\omega - 1$ divide de hecho a todos los factores de la izquierda y que los números

$$\frac{x + \omega^i y}{\omega - 1}, \quad i = 0, \ldots, p - 1,$$

son no congruentes dos a dos módulo $\omega - 1$.

Como $\mathrm{N}(\omega - 1) = p$, estos números forman un conjunto completo de representantes de las clases de congruencia módulo $\omega - 1$. En particular existe un único i entre 0 y $p - 1$ tal que $(\omega - 1)^2 \mid x + \omega^i y$. Si llamamos y a $\omega^i y$, se sigue cumpliendo (8.11) y ahora $(\omega - 1)^2 \mid x + y$, mientras que los factores restantes $x + \omega^i y$ son divisibles entre $\omega - 1$ pero no entre $(\omega - 1)^2$.

En consecuencia el miembro izquierdo de (8.12) es divisible entre $(\omega-1)^{p+1}$, y en particular ha de ser $m > 1$.

Sea $\mathfrak{m} = (x, y)$. Como x e y son primos con $\omega - 1$, lo mismo le ocurre a $\mathfrak{m}$. Por lo tanto si $i \neq 0$ tenemos que $(x + \omega^i y) = (\omega - 1)\mathfrak{m}\mathfrak{c}_i$, mientras que $x + y$ ha de ser divisible entre los $p(m - 1) + 1$ factores $\omega - 1$ restantes que dividen el miembro derecho de (8.12), es decir,

$$(x + y) = (\omega - 1)^{p(m-1)+1}\mathfrak{m}\mathfrak{c}_0.$$

Los ideales $\mathfrak{c}_i$, para $i = 0, \ldots, p - 1$, son primos entre sí dos a dos, pues si un primo $\mathfrak{p}$ divide a dos de ellos, entonces $\mathfrak{m}\mathfrak{p}$ divide a dos números $x + \omega^i y$, $x + \omega^j y$, luego también divide a su suma y a su diferencia, es decir, a $(\omega - 1)y$, $(\omega - 1)x$, luego a $\mathfrak{m} = (x, y)$, pero esto es imposible. La ecuación dada queda ahora del modo siguiente:

$$\mathfrak{m}^p(\omega - 1)^{pm}\mathfrak{c}_0\mathfrak{c}_1 \cdots \mathfrak{c}_{p-1} = (\omega - 1)^{pm}(z)^p.$$

Puesto que los $\mathfrak{c}_i$ son primos entre sí, todos han de ser potencias p-ésimas. Digamos que $\mathfrak{c}_i = \mathfrak{b}_i^p$, con lo que

$$\begin{aligned}(x+y) &= (\omega-1)^{p(m-1)+1}\mathfrak{m}\mathfrak{b}_0^p,\\ (x+\omega^i y) &= (\omega-1)\mathfrak{m}\mathfrak{b}_i^p, \qquad i=1,\dots,p-1.\end{aligned}$$

Despejamos $\mathfrak{m}$ en la primera ecuación y lo sustituimos en la segunda:

$$(\omega-1)^{p(m-1)}\mathfrak{b}_0^p(x+\omega^i y) = (x+y)\mathfrak{b}_i^p, \quad i=1,\dots,p-1. \tag{8.13}$$

Por 8.35 podemos expresar $\mathrm{N}(\mathfrak{b}_0) = \mathfrak{b}_0\bar{\mathfrak{b}}_0$, y entonces

$$(\omega-1)^{p(m-1)}\,\mathrm{N}(\mathfrak{b}_0)^p(x+\omega^i y) = (x+y)\bar{\mathfrak{b}}_0^p\mathfrak{b}_i^p.$$

Tenemos así una ecuación de la forma $(\alpha) = (\beta)\bar{\mathfrak{b}}_0^p\mathfrak{b}_i^p$, luego $\alpha/\beta \in \bar{\mathfrak{b}}_0^p\mathfrak{b}_i^p$ es entero, luego $\bar{\mathfrak{b}}_0^p\mathfrak{b}_i^p = (\alpha/\beta)$ es principal y, por la propiedad A) de la definición de primo regular, concluimos que el ideal $\mathfrak{b}_i\bar{\mathfrak{b}}_0$ también es principal, digamos $\mathfrak{b}_i\bar{\mathfrak{b}}_0 = (\alpha_i)$. Multiplicando por $\mathfrak{b}_0$ queda $\mathrm{N}(\mathfrak{b}_0)\mathfrak{b}_i = (\alpha_i)\mathfrak{b}_0$. Notemos que tanto $\mathrm{N}(\mathfrak{b}_0)$ como (α_i) son primos con $\omega-1$. Elevamos a p y sustituimos en (8.13):

$$(\omega-1)^{p(m-1)}\,\mathrm{N}(\mathfrak{b}_0)^p(x+\omega^i y) = (x+y)(\alpha_i)^p, \quad i=1,\dots,p-1.$$

Eliminando los ideales queda

$$(\omega-1)^{p(m-1)}\,\mathrm{N}(\mathfrak{b}_0)^p(x+\omega^i y) = \epsilon_i(x+y)\alpha_i^p,$$

donde ϵ_i es una unidad, o equivalentemente

$$(\omega-1)^{p(m-1)}(x+\omega^i y) = \epsilon_i(x+y)\gamma_i^p, \tag{8.14}$$

donde $\gamma_i = \alpha_i/\,\mathrm{N}(\mathfrak{b}_0)$ (que no es necesariamente entero).

Nuestro objetivo es combinar estas ecuaciones para llegar a una ecuación similar a (8.11) pero con un valor menor para m. Una forma rápida de hacerlo es partir de la identidad

$$(x+\omega y)(1+\omega) - (x+\omega^2 y) = \omega(x+y).$$

Si la multiplicamos por $(\omega-1)^{p(m-1)}$ y usamos (8.14) para $i=1,2$ obtenemos

$$(x+y)\gamma_1^p\epsilon_1(1+\omega) - (x+y)\gamma_2^p\epsilon_2 = (x+y)\omega(\omega-1)^{p(m-1)}.$$

Como $1+\omega$ es una unidad, esta ecuación se puede poner en la forma

$$\gamma_1^p + \gamma_2^p\epsilon = \eta(\omega-1)^{p(m-1)},$$

donde ϵ y η son unidades. Multiplicando por $\mathrm{N}(\mathfrak{b}_0)^p$ queda una ecuación de tipo

$$\alpha^p + \epsilon\beta^p = \eta(\omega-1)^{p(m-1)}\gamma^p,$$

donde α, β y γ son enteros ciclotómicos primos con $\omega-1$. Esta ecuación será de tipo (8.11) si ϵ es una potencia p-ésima. Lo probaremos usando la propiedad B) de la definición de primo regular.

En efecto, basta observar que $p(m-1) \geq p$, pues hemos probado que $m > 1$, luego

$$\alpha^p + \epsilon\beta^p \equiv 0 \text{ (mód } p).$$

Despejando ϵ (lo cual es posible porque β es primo con p), vemos que es congruente con una potencia p-ésima módulo p, luego es congruente con un entero racional módulo p, luego es una potencia p-ésima (por la propiedad B). ∎

En particular tenemos probado el Último Teorema de Fermat para $p = 5$. Kummer encontró una caracterización sencilla de los primos regulares, fácil de comprobar, que permite justificar fácilmente que, por ejemplo, todos los primos menores que 100 son regulares excepto $37, 59$ y 67. Sin embargo, llegar a dicha caracterización requiere técnicas que van más allá de los contenidos de este libro.

No se sabe si el conjunto de los primos regulares es finito o infinito, pero sí se ha probado que el conjunto de los primos irregulares es infinito.

Capítulo IX

Complementos sobre cuerpos

En este último capítulo recogemos una serie de hechos adicionales sobre cuerpos para los que no podemos mostrar aquí aplicaciones relevantes que justifiquen su interés, pero que, no obstante, son de gran importancia para quien quiera profundizar en las materias que hemos introducido en este libro.

9.1 Cuerpos finitos

Los cuerpos finitos aparecen en teoría de números como cocientes de los anillos de enteros de los cuerpos numéricos sobre sus ideales primos, y resultan indispensables cuando se estudian extensiones arbitrarias de cuerpos numéricos (donde el cuerpo base no es necesariamente $\mathbb{Q}$). Fueron estudiados por primera vez por Galois, por lo que se les llama cuerpos de Galois.

Comencemos observando que los cuerpos finitos tienen necesariamente característica prima. Además, si k es un cuerpo finito de característica p, es inmediato que k ha de ser una extensión finita de su cuerpo primo $\mathbb{Z}/p\mathbb{Z}$ (no puede contener una base infinita), luego si $|k : \mathbb{Z}/p\mathbb{Z}| = n$, tenemos que k es un espacio vectorial de dimensión n sobre el cuerpo $\mathbb{Z}/p\mathbb{Z}$, luego es isomorfo al producto cartesiano de $\mathbb{Z}/p\mathbb{Z}$ por sí mismo n veces, luego en particular $|k| = p^n$.

En resumen, si k es un cuerpo finito, $\operatorname{car} k = p$, y $|k : \mathbb{Z}/p\mathbb{Z}| = n$, entonces $|k| = p^n$, luego no hay cuerpos finitos de todos los cardinales posibles, sino a lo sumo de cardinales que sean potencias de primo. Veamos que, salvo isomorfismo, existe un único cuerpo de orden p^n.

Definición 9.1 Llamaremos *cuerpo de Galois* de p^n elementos al cuerpo de escisión sobre $\mathbb{Z}/p\mathbb{Z}$ del polinomio $x^{p^n} - x$, y lo representaremos por $\mathrm{CG}(p^n)$.

Si k es un cuerpo de característica prima p, llamaremos *automorfismo de Frobenius* de k al automorfismo $\sigma : k \longrightarrow k$ dado por $\sigma(a) = a^p$.

Teorema 9.2 *Sea $K = \mathrm{CG}(p^n)$ y sea $k = \mathbb{Z}/p\mathbb{Z}$ su cuerpo primo. Entonces:*

1. *K es, salvo isomorfismo, el único cuerpo de p^n elementos.*

2. *La extensión K/k es de Galois y el grupo de Galois $G(K/k)$ es cíclico y está generado por el automorfismo de Frobenius σ.*

3. *El cuerpo $\mathrm{CG}(p^m)$ es isomorfo a un (único) subcuerpo de K si y sólo si $m \mid n$, y en tal caso es isomorfo al cuerpo fijado por σ^m.*

Demostración: 1) El polinomio $x^{p^n} - x$ tiene p^n raíces distintas, pues su derivada es -1, que no tiene raíces. Sea, pues, $K_0 \subset K$ el conjunto de las p^n raíces. Ahora bien, es inmediato que K_0 es un subanillo de K, y todo dominio íntegro finito es un cuerpo, luego K_0 es un cuerpo, luego $K = K_0$, y así K tiene p^n elementos.

Si L es un cuerpo de p^n elementos, entonces por 4.50 sabemos que L^* es un grupo cíclico de orden $p^n - 1$, luego todos sus elementos son raíces del polinomio $x^{p^n-1} - 1$, luego todos los elementos de L son raíces de $x^{p^n} - x$, luego L es un cuerpo de escisión de este polinomio, luego por la unicidad del cuerpo de escisión L es isomorfo a $\mathrm{CG}(p^n)$.

2) La extensión K/k es de Galois porque es normal porque K es un cuerpo de escisión y es separable porque los cuerpos finitos son perfectos. Sabemos que el grupo multiplicativo K^* es cíclico. Sea $K^* = \langle u \rangle$. Entonces $u^{p^n} = u$, pero $u^{p^m} \neq u$ para todo $m < n$. Esto equivale a que $\sigma^m(u) \neq 1$ para todo $m < n$, luego el orden de σ es como mínimo n, pero como el grupo de Galois tiene n elementos, el orden de σ ha de ser exactamente n, y σ es un generador.

3) El grupo $G(K/k)$, al ser cíclico, tiene un único subgrupo para cada divisor d de n. Si $n = dm$, entonces el subgrupo de orden d es el generado por σ^m y su cuerpo fijado L cumple que $|K : L| = d$, luego $|L : k| = m$, luego $|L| = p^m$. Así pues, los únicos subcuerpos de K son los cuerpos de p^m elementos, para cada $m \mid n$, y además K tiene un único subcuerpo isomorfo a cada $\mathrm{CG}(p^m)$. ∎

En particular $\mathrm{CG}(p) = \mathbb{Z}/p\mathbb{Z}$. Al trabajar con cuerpos es más usual la notación $\mathrm{CG}(p)$ que $\mathbb{Z}/p\mathbb{Z}$.

Fijada una clausura algebraica[1] $\mathbb{A}_p$ de $\mathbb{Z}/p\mathbb{Z}$, podemos identificar cada cuerpo $\mathrm{CG}(p^n)$ con el conjunto de las raíces en $\mathbb{A}_p$ del polinomio $x^{p^n} - x$, y entonces el apartado 3) del teorema anterior se reduce a que

$$\mathrm{CG}(p^m) \subset \mathrm{CG}(p^n) \leftrightarrow m \mid n.$$

Además es claro que entonces $\mathbb{A}_p = \bigcup_{n=1}^{\infty} \mathrm{CG}(p^n)$, y que $\mathrm{CG}(p^n)$ es el único subcuerpo de $\mathbb{A}_p$ de p^n elementos.

[1] No es necesario el axioma de elección para construir $\mathbb{A}_p$. Basta describir explícitamente una construcción de una cadena de cuerpos $K_0 \subset K_1 \subset K_2 \subset \cdots$ de modo que $|K_n| = p^{n!}$ y definir $\mathbb{A}_p$ como la unión de estos cuerpos. Todos los cuerpos K_n se pueden construir como subconjuntos de un conjunto numerable prefijado X y usar una enumeración de X para realizar todas las elecciones que requiere el proceso sin necesidad de recurrir al axioma de elección. La prueba de que dos clausuras algebraicas son isomorfas requiere a lo sumo ED.

Conviene observar que las extensiones entre cuerpos finitos no sólo son cíclicas, sino también ciclotómicas. En efecto, los generadores del grupo multiplicativo de $\mathrm{CG}(p^n)$ son raíces p^n-1-ésimas primitivas de la unidad. El cuerpo $\mathrm{CG}(p^n)$ es una extensión ciclotómica de orden p^n-1 de cualquiera de sus subcuerpos.

Ejercicio: Sea m un número natural no nulo y p un primo que no divida a m. Entonces el grado de la extensión ciclotómica m-sima sobre $\mathrm{CG}(p)$ es el orden de p módulo m. (Véase la prueba del teorema 8.38.)

Ejemplo Vamos a describir el cuerpo de 8 elementos. Se trata de la única extensión de grado 3 de $\mathrm{CG}(2)$. Buscamos un polinomio mónico irreducible de grado 3 en $\mathrm{CG}(2)[x]$. Hay 8 polinomios mónicos de grado 3. Descartamos los que no tienen término independiente porque no son irreducibles, con lo que quedan 4, a saber:

$$x^3+x^2+x+1, \quad x^3+x^2+1, \quad x^3+x+1 \quad \text{y} \quad x^3+1$$

El primero y el último tienen raíz 1, luego sólo nos quedan los dos de en medio. Por ejemplo, tomamos

$$p(x)=x^3+x+1.$$

Necesariamente $\mathrm{CG}(8)=\mathrm{CG}(2)[\alpha]$, donde α es una raíz de $p(x)$. Por consiguiente,

$$\mathrm{CG}(8)=\{a\alpha^2+b\alpha+c \mid a,\, b,\, c\in \mathrm{CG}(2)\}.$$

La suma de dos elementos de $\mathrm{CG}(8)$ se calcula sumando las coordenadas, mientras que el producto se calcula operando de forma habitual y reduciendo las potencias de α mediante la relación $\alpha^3=-\alpha-1$. ∎

Ejercicio: Describir el CG(9). Mostrar un isomorfismo entre este cuerpo y $\mathbb{Z}[i]/(3)$.

En este libro hemos visto que algunos problemas de la teoría de números pueden formularse en términos de si un número entero dado es o no una norma de un entero algebraico. En cuerpos finitos la situación es mucho más simple:

Teorema 9.3 *En una extensión de cuerpos finitos, la norma y la traza son suprayectivas.*

Demostración: Sea K/k una extensión de cuerpos finitos. Digamos que $k=\mathrm{CG}(m)$ y que $|K:k|=d$. Entonces $G(K/k)=\langle\sigma\rangle$, donde $\sigma(x)=x^m$. Por lo tanto la norma es

$$\mathrm{N}(x)=x^{1+m+\cdots+m^{d-1}}=x^{(m^d-1)/(m-1)}.$$

Considerando a la norma como homomorfismo de grupos $\mathrm{N}: K^*\longrightarrow k^*$, vemos que su núcleo está formado por las raíces del polinomio $x^{(m^d-1)/(m-1)}-1$, luego a lo sumo tiene $(m^d-1)/(m-1)$ elementos. Por el teorema de isomorfía la imagen tiene al menos $m-1$ elementos (observemos que $|K^*|=m^d-1$), pero éstos son todos los elementos de k^*, luego la norma es suprayectiva.

Similarmente, el núcleo de la traza $\mathrm{Tr} : K \longrightarrow k$ está formado por las raíces del polinomio

$$x^{m^{d-1}} + \cdots + x^m + x,$$

luego a lo sumo tiene m^{d-1} elementos, y la imagen de la traza tiene como mínimo m elementos, luego también es suprayectiva. ∎

Al estudiar los cuerpos finitos estamos estudiando, de hecho, todos los anillos de división finitos:

Teorema 9.4 (Teorema de Wedderburn) *Todo anillo de división finito es un cuerpo.*

Demostración: Sea K un anillo de división finito. Sea Z el centro de K, es decir, el conjunto de todos los elementos de K que conmutan con todos los demás. Es fácil ver que se trata de un cuerpo, por lo que K es un espacio vectorial sobre Z. Si éste tiene q elementos, entonces el número de elementos de K es q^n, donde n es la dimensión de K sobre Z.

Para cada $\alpha \in K$, el conjunto C_α formado por todos los elementos de K que conmutan con α es también un anillo de división que contiene a Z, luego su número de elementos es q^d, donde $d \mid n$ (por la transitividad de grados, que vale igualmente para anillos de división). Indicamos con un asterisco el conjunto que resulta de eliminar el 0. Entonces Z^* es el centro del grupo K^* y C_α^* es el centralizador de $\alpha \neq 0$. La ecuación de clases [TG 2.4] es en este caso:

$$q^n - 1 = q - 1 + \sum_d \frac{q^n - 1}{q^d - 1},$$

donde d recorre ciertos divisores de n, posiblemente repetidos, distintos de n.

Sea $c_n(x)$ el polinomio ciclotómico de orden n. Es claro que

$$x^n - 1 = c_n(x)f(x), \quad \frac{x^n - 1}{x^d - 1} = c_n(x)g(x),$$

para ciertos polinomios $f(x), g(x) \in \mathbb{Z}[x]$.

Evaluando en q vemos que $c_n(q)$ divide a todos los términos de la ecuación de clases. En particular $c_n(q) \mid q - 1$.

Por otra parte, $c_n(q)$ es el producto de factores de la forma $q - \zeta$, donde ζ recorre las raíces n-simas primitivas (complejas) de la unidad. Claramente

$$|q - \zeta| \geq \big|q - |\zeta|\big| = |q - 1| \geq 1,$$

y como $q \geq 2$ la igualdad sólo se da si $\zeta = 1$, y entonces $n = 1$. En otro caso concluimos que $|c_n(q)| > q - 1$, lo que nos da una contradicción. Así pues, $n = 1$ y $K = Z$ es un cuerpo. ∎

El grupo de descomposición de un ideal primo Para terminar veamos un ejemplo de cómo intervienen los cuerpos finitos en el estudio de los anillos de enteros algebraicos. Consideremos un cuerpo numérico K que sea una extensión de Galois de $\mathbb{Q}$ de grado n, sea $G = G(K/\mathbb{Q})$ su grupo de Galois, sea $\mathcal{O}_K$ su anillo de enteros algebraicos, sea $p \in \mathbb{Z}$ un primo y sea $\mathfrak{P}$ uno de los divisores primos de p en $\mathcal{O}_K$. Sea $K_\mathfrak{P} = \mathcal{O}_K/\mathfrak{P}$, que es un cuerpo de $N(\mathfrak{P}) = p^f$ elementos y, análogamente, llamemos $\mathbb{Q}_p = \mathbb{Z}/p\mathbb{Z}$. Observemos que la inclusión $\mathbb{Z} \longrightarrow \mathcal{O}_K$ induce un monomorfismo de cuerpos $\mathbb{Q}_p \longrightarrow K_\mathfrak{P}$ que nos permite considerar a $K_\mathfrak{P}/\mathbb{Q}_p$ como una extensión de grado f, que es de Galois, por ser una extensión de cuerpos finitos.

Si $\sigma \in G$, entonces $\sigma|_{\mathcal{O}_K}$ es un automorfismo de $\mathcal{O}_K$, luego $\sigma[\mathfrak{P}]$ es también un divisor primo de p en $\mathcal{O}_K$ (como $p \in \mathfrak{P}$, también $p = \sigma(p) \in \sigma[\mathfrak{P}]$). Es claro que

$$G_\mathfrak{P} = \{\sigma \in G \mid \sigma[\mathfrak{P}] = \mathfrak{P}\}$$

es un subgrupo de G, llamado *grupo de descomposición* de $\mathfrak{P}$. Es fácil calcular su orden. Para ello consideramos la factorización de p en $\mathcal{O}_K$, que según el teorema 8.39 es de la forma $p = (\mathfrak{P}_1 \cdots \mathfrak{P}_r)^e$, donde $\mathrm{N}(\mathfrak{P}_i) = f$ y $n = efr$. Podemos suponer que $\mathfrak{P} = \mathfrak{P}_1$. Dicho teorema nos dice además que la aplicación $G \longrightarrow \{\mathfrak{P}_1, \ldots, \mathfrak{P}_r\}$ dada por $\sigma \mapsto \sigma[\mathfrak{P}]$ es suprayectiva, de modo que, para cada índice i, existe un $\sigma_i \in G$ tal que $\mathfrak{P}_i = \sigma_i[\mathfrak{P}]$. Podemos tomar $\sigma_1 = 1$.

Ahora bien, la aplicación $(G/G_\mathfrak{P})_d \longrightarrow \{\mathfrak{P}_1, \ldots, \mathfrak{P}_r\}$ dada por $G_\mathfrak{P}\sigma \mapsto \sigma[\mathfrak{P}]$ está bien definida y es claramente biyectiva, luego $|G : G_\mathfrak{P}| = r$ y, por consiguiente, $|G_\mathfrak{P}| = ef$.

Más precisamente, ahora es claro que $(G/G_\mathfrak{P})_d = \{G_\mathfrak{P}\sigma_1, \ldots, G_\mathfrak{P}\sigma_r\}$.

Llamamos L a su cuerpo fijado, de modo que $\mathbb{Q} \subset L \subset K$ con $|L : \mathbb{Q}| = r$ y $G_\mathfrak{P} = G(K/L)$. Sea $\mathcal{O}_L$ el anillo de enteros algebraicos de L y sea $\mathfrak{p} = \mathfrak{P} \cap \mathcal{O}_L$, que claramente es uno de los divisores primos de p en $\mathcal{O}_L$. Sea $L_\mathfrak{p} = \mathcal{O}_L/\mathfrak{p}$. Como antes, las inclusiones $\mathbb{Z} \longrightarrow \mathcal{O}_L \longrightarrow \mathcal{O}_K$ inducen monomorfismos de cuerpos $\mathbb{Q}_p \longrightarrow L_\mathfrak{p} \longrightarrow K_\mathfrak{P}$ que nos permiten considerar a cada cuerpo como un subcuerpo del siguiente.

Vamos a demostrar que $L_\mathfrak{p} = \mathbb{Q}_p$, lo que significa que $\mathrm{N}(\mathfrak{p}) = p$.

En primer lugar observamos que los primos $\mathfrak{p}_i = \mathfrak{P}_i \cap \mathcal{O}_L$ son distintos de $\mathfrak{p}$, para $i \neq 1$. Para ello tomamos $\alpha \in \mathfrak{P}_i \setminus \mathfrak{P}$ y observamos que la norma $\mathrm{N}_L^K(\alpha) = \prod_{\sigma \in G_\mathfrak{P}} \sigma(\alpha) \in L$ es un entero algebraico, luego $\mathrm{N}(\alpha) \in \mathcal{O}_L$.

Como $\alpha \in \mathfrak{P}_i$, también $\mathrm{N}(\alpha) \in \mathfrak{P}_i$, luego $\mathrm{N}(\alpha) \in \mathfrak{p}_i$, pero $\mathrm{N}(\alpha) \notin \mathfrak{p}$, pues en caso contrario $\mathrm{N}(\alpha) \in \mathfrak{P}$ y, como es un ideal primo, existe un $\sigma \in G_\mathfrak{P}$ tal que $\sigma(\alpha) \in \mathfrak{P} = \sigma[\mathfrak{P}]$, luego $\alpha \in \mathfrak{P}$, contradicción. Esto prueba que $\mathfrak{p}_i \not\subset \mathfrak{p}$.

Si $i > 1$, tenemos que $\sigma_i^{-1}[\mathfrak{P}] \neq \mathfrak{P}$, luego $\sigma_i^{-1}[\mathfrak{P}]$ es un $\mathfrak{P}_j$, con $j \neq 1$, luego $\mathfrak{q}_i = \sigma_i^{-1}[\mathfrak{P}] \cap \mathcal{O}_L$ es un $\mathfrak{p}_j \neq \mathfrak{p}$. Aplicamos el teorema chino del resto 3.56 a los ideales $\mathfrak{p}$ y $\mathfrak{m} = \mathfrak{q}_2 \cdots \mathfrak{q}_r$, que son primos entre sí, por la factorización única ideal. Dado $\alpha \in \mathcal{O}_L$, existe $\alpha' \in \mathcal{O}_L$ tal que $\alpha' \equiv \alpha \pmod{\mathfrak{p}}$ y $\alpha' \equiv 1 \pmod{\mathfrak{m}}$. En particular tenemos que $\alpha' \equiv \alpha \pmod{\mathfrak{P}}$ y $\alpha' \equiv 1 \pmod{\sigma_i^{-1}[\mathfrak{P}]}$, para todo índice $i > 1$.

Así $m = \mathrm{N}_{\mathbb{Q}}^{L}(\alpha') = \sigma_1(\alpha')\cdots\sigma_r(\alpha') \in \mathbb{Z}$ y, para cada índice $i > 1$, se cumple que $\sigma_i(\alpha') \equiv 1 \pmod{\mathfrak{P}}$, luego $m \equiv \alpha' \equiv \alpha \pmod{\mathfrak{P}}$. Y como son elementos del anillo $\mathcal{O}_L$, de hecho $\alpha \equiv m \pmod{\mathfrak{p}}$. Esto significa que, en $L_{\mathfrak{p}}$, se cumple que $[\alpha] = [m] \in \mathbb{Q}_p$, y como $\alpha \in \mathcal{O}_L$ era arbitrario, hemos probado que $L_{\mathfrak{p}} = \mathbb{Q}_p$.

Ahora observamos que si $\sigma \in G_{\mathfrak{P}}$, entonces $\alpha \equiv \beta \pmod{\mathfrak{P}}$ equivale a $\sigma(\alpha) \equiv \sigma(\beta) \pmod{\mathfrak{P}}$, luego σ induce un automorfismo $\bar{\sigma} : K_{\mathfrak{P}} \longrightarrow K_{\mathfrak{P}}$, de modo que la aplicación $G_{\mathfrak{P}} \longrightarrow G(K_{\mathfrak{P}}/\mathbb{Q}_p)$ dada por $\sigma \mapsto \bar{\sigma}$ es un homomorfismo de grupos. Vamos a probar que es suprayectivo.

Para ello tomamos un elemento primitivo $K_{\mathfrak{P}} = \mathbb{Q}_p([\alpha])$, con $\alpha \in \mathcal{O}_K$. Sea $f(x) = \operatorname{pol\,mín}(\alpha, L)$. Como las raíces de f son enteros algebraicos, los coeficientes de $f(x)$ también lo son, luego $f(x) \in \mathcal{O}_L[x]$ y podemos considerar el polinomio $\bar{f}(x) \in L_{\mathfrak{p}}[x] = \mathbb{Q}_p[x]$ que resulta de tomar clases módulo $\mathfrak{p}$ en los coeficientes de f.

Como la extensión K/L es normal, $f(x) = (x-\alpha_1)\cdots(x-\alpha_l)$, para ciertos $\alpha_i \in \mathcal{O}_K$ distintos dos a dos, con $\alpha = \alpha_1$, luego $\bar{f}(x) = (x-[\alpha_1])\cdots(x-[\alpha_l])$, y así las clases $[\alpha_i]$ son todos los conjugados de $[\alpha]$.

Por lo tanto, si $\rho \in G(K_{\mathfrak{P}}/\mathbb{Q}_p)$, tenemos que $\rho([\alpha]) = [\alpha_i]$, para un cierto i, y existe un $\sigma \in G_{\mathfrak{P}}$ tal que $\sigma(\alpha) = \alpha_i$ (porque α y α_i son conjugados en K/L), luego $\bar{\sigma}([\alpha]) = [\alpha_i] = \rho([\alpha])$, luego $\rho = \bar{\sigma}$.

El teorema siguiente resume lo que hemos obtenido:

Teorema 9.5 *Sea K un cuerpo numérico normal sobre $\mathbb{Q}$, sea $\mathcal{O}_K$ su anillo de enteros algebraicos y $\mathfrak{P}$ un primo en $\mathcal{O}_K$. Sea $G_{\mathfrak{P}}$ el grupo de descomposición de $\mathfrak{P}$ y sea $K_{\mathfrak{P}} = \mathcal{O}_K/\mathfrak{P}$ el cuerpo de restos. Entonces, el homomorfismo natural $G_{\mathfrak{P}} \longrightarrow G(K_{\mathfrak{P}}/\mathbb{Q}_p)$ es suprayectivo.*

En general el epimorfismo considerado no es inyectivo, pues $G_{\mathfrak{P}}$ tiene orden ef y el grupo de Galois de los cuerpos de restos tiene orden f, luego la inyectividad equivale a $e = 1$. Vamos a dar una condición suficiente para que esto suceda:

El teorema 8.12 nos da que $K = \mathbb{Q}(\alpha)$, para cierto $\alpha \in \mathcal{O}_K$. Consideremos $f(x) = \operatorname{pol\,mín}(\alpha, \mathbb{Q}) \in \mathbb{Z}[x]$ y sea $\Delta \in \mathbb{Z}$ su discriminante. Vamos a suponer que $p \nmid \Delta$, de modo que $[\Delta] \neq 0$ en $\mathbb{Q}_p$.

Así, si $\theta_1, \ldots, \theta_n \in \mathcal{O}_K$ son las raíces de $f(x)$, tenemos que

$$f(x) = (x-\theta_1)\cdots(x-\theta_n), \text{ luego } \bar{f}(x) = (x-[\theta_1])\cdots(x-[\theta_n]),$$

$$\Delta = \prod_{i<j}(\theta_i-\theta_j)^2, \text{ luego } [\Delta] = \prod_{i<j}([\theta_i]-[\theta_j])^2 \neq 0,$$

lo que implica que las clases $[\theta_i] \in K_{\mathfrak{P}}$ son distintas dos a dos y son todas las raíces del polinomio $\bar{f}(x) \in \mathbb{Q}_p[x]$. En particular $\bar{f}(x)$ tiene raíces simples.

Ahora la conclusión es inmediata: si $\sigma \in G_{\mathfrak{P}}$ no es la identidad, existe un i tal que $\sigma(\theta_i) = \theta_j \neq \theta_i$, luego $\bar{\sigma}([\theta_i]) = [\theta_j] \neq [\theta_i]$, luego $\bar{\sigma} \neq 1$. Con esto hemos probado el teorema siguiente:

Teorema 9.6 *Sea $f(x) \in \mathbb{Z}[x]$ un polinomio mónico irreducible, sea K su cuerpo de escisión, sea p un primo que no divida al discriminante de f y sea $\mathfrak{P}$ un divisor primo de p en $\mathcal{O}_K$. Entonces el epimorfismo $G_\mathfrak{P} \longrightarrow G(K_\mathfrak{P}/\mathbb{Q}_p)$ es un isomorfismo.*

De hecho hemos probado algo más, y es que si identificamos a $G_\mathfrak{P}$ con un subgrupo del grupo de permutaciones de las raíces $\theta_1, \ldots, \theta_n$ de $f(x)$ y a $G(K_\mathfrak{P}/\mathbb{Q}_p)$ con un subgrupo del grupo de permutaciones de las raíces $[\theta_1], \ldots, [\theta_n]$ de $\bar{f}(x)$, entonces las acciones correspondientes son isomorfas, en el sentido de que $\sigma(\theta_i) = \theta_j$ si y sólo si $\bar{\sigma}([\theta_i]) = [\theta_j]$. En particular, la descomposición en ciclos disjuntos de σ es del mismo tipo que la de $\bar{\sigma}$.

Hasta aquí no hemos usado ninguna propiedad específica de los cuerpos finitos, salvo que las extensiones de cuerpos finitos son de Galois. Ahora probamos un resultado en el que es crucial que, de hecho, son cíclicas:

Teorema 9.7 (Dedekind) *Sea $f(x) \in \mathbb{Z}[x]$ un polinomio mónico irreducible y sea p un primo que no divida a su discriminante. Sea $\bar{f}(x) \in (\mathbb{Z}/p\mathbb{Z})[x]$ la reducción de $f(x)$ módulo p y sea $\bar{f}(x) = f_1(x)\cdots f_m(x)$ la descomposición de $\bar{f}(x)$ en polinomios irreducibles. Sea $e_i = \operatorname{grad} f_i(x)$. Sea K el cuerpo de escisión de $f(x)$. Entonces $G(K/\mathbb{Q})$, identificado con un subgrupo del grupo de permutaciones de las raíces de $f(x)$, contiene una permutación de tipo $e_1, \ldots, e_m$, es decir, que se descompone en m ciclos disjuntos de longitudes e_i.*

Demostración: Continuando con la notación precedente, basta probar que el automorfismo de Frobenius $\sigma \in G(K_\mathfrak{P}/\mathbb{Q}_p)$ determina una permutación de tipo $e_1, \ldots, e_m$ en el grupo de las permutaciones de $[\theta_1], \ldots, [\theta_n]$. Ahora bien, como σ genera el grupo de Galois, es claro que dos elementos de $K_\mathfrak{P}$ son conjugados si y sólo si están en la misma órbita respecto a σ, y como las raíces de cada f_i tienen que formar una clase de conjugación, resulta que σ induce en $\{[\theta_1], \ldots, [\theta_n]\}$ exactamente m órbitas de longitudes e_i, y eso es tanto como decir que σ se descompone en m ciclos disjuntos de longitudes e_i. ∎

Observemos que este teorema es puramente un teorema de teoría de extensiones de cuerpos, de modo que a partir de su enunciado nada indica que la prueba requiera considerar enteros algebraicos, factorizaciones de primos, etc. Como aplicación vamos a construir polinomios de $\mathbb{Q}[x]$ con grupo de Galois isomorfo a Σ_n:

Teorema 9.8 *Para cada natural $n \geq 2$, existe un polinomio mónico irreducible $f(x) \in \mathbb{Z}[x]$ de grado n cuyo grupo de Galois (sobre $\mathbb{Q}$) es isomorfo a Σ_n.*

Demostración: Podemos suponer que $n > 3$, pues para $n = 2, 3$ ya conocemos ejemplos de polinomios en las condiciones del teorema. Sea $f_1(x) \in \mathbb{Z}[x]$ un polinomio mónico de grado n tal que su reducción módulo 2 sea irreducible. Existe tal polinomio pues basta tomar el polinomio mínimo sobre $\mathbb{Z}/2\mathbb{Z}$ de un elemento primitivo de $\mathrm{CG}(2^n)$ y pasar a un polinomio mónico cuya reducción módulo 2 sea dicho polinomio.

Similarmente, sea $f_2(x) = xg(x)$, donde $g(x) \in \mathbb{Z}[x]$ es un polinomio mónico de grado $n-1$ cuya reducción módulo 3 sea irreducible. Por último tomemos $f_3(x) = u(x)v(x)$, donde $u(x)$ es un polinomio mónico de grado 2 cuya reducción módulo 5 sea irreducible y $v(x)$ es un polinomio mónico de grado $n-2$ que tenga también reducción irreducible módulo 5 si su grado es impar, o bien $v(x) = xv'(x)$, donde $v'(x)$ es mónico de grado $n-3$ con reducción irreducible módulo 5 si $n-2$ es par.

Por último llamamos $f(x) = -15f_1(x) + 10f_2(x) + 6f_3(x)$, de modo que

$$f(x) \equiv f_1(x) \text{ (mód 2)}, \quad f(x) \equiv f_2(x) \text{ (mód 3)}, \quad f(x) \equiv f_3(x) \text{ (mód 5)}.$$

Así $f(x)$ es mónico de grado n (pues su coeficiente director es $-15+10+6 = 1$) y es irreducible módulo 2, lo cual implica que es irreducible. Sea G su grupo de Galois. El teorema anterior implica que, visto como grupo de permutaciones de las raíces de f, el grupo G contiene un ciclo σ de longitud $n-1$ (por su reducción módulo 3) y una trasposición τ multiplicada por un ciclo ρ de longitud impar l (por su reducción módulo 5). Ahora bien, como $(\tau\rho)^l = \tau$, resulta que también $\tau \in G$.

Pongamos que $\sigma = (\theta_1, \ldots, \theta_{n-1})$, donde $\theta_1, \ldots, \theta_n$ son las raíces de f, numeradas adecuadamente en función de σ. Sea $\tau = (\theta_i, \theta_j)$. Como todas las raíces son conjugadas, existe $\alpha \in G$ tal que $\alpha(\theta_j) = \theta_n$, y cambiando τ por τ^α, podemos suponer que $\tau = (\theta_i, \theta_n)$. Pero entonces es claro que las permutaciones τ^{σ^k} son todas las trasposiciones (θ_j, θ_n), para $j = 1, \ldots, n-1$. A su vez, si $i \neq j$, tenemos que $(\theta_i, \theta_n)^{(\theta_j, \theta_n)} = (\theta_i, \theta_j)$, luego G contiene todas las trasposiciones, luego es Σ_n. ∎

Ejercicio: Construir un polinomio con grupo de Galois Σ_6 sobre $\mathbb{Q}$.

La sección siguiente contiene otra aplicación de los cuerpos finitos en la prueba de un resultado general de la teoría de cuerpos.

9.2 El teorema de la base normal

En esta sección demostraremos un teorema de cierta importancia en la teoría de Galois (especialmente en relación con la cohomología de grupos). Su enunciado es muy sencillo: si K/k es una extensión de cuerpos, una *base normal* de K sobre k es una base cuyos elementos forman una clase de conjugación, es decir, son todas las raíces de un mismo polinomio irreducible de $k[x]$. El teorema de la base normal afirma que toda extensión finita de Galois tiene una base normal.

Ejercicio: Probar que si una extensión finita tiene una base normal entonces es de Galois.

Un enunciado alternativo del teorema de la base normal es que en toda extensión finita de Galois K/k existe un $v \in K$ tal que $\{\sigma(v) \mid \sigma \in G(K/k)\}$ es una k-base de K.

Como primer paso de la demostración probamos el resultado siguiente:

Teorema 9.9 *Si K/k es una extensión finita de Galois, el cuerpo k es infinito y $G(K/k) = \{\sigma_1, \dots, \sigma_n\}$, entonces existe un $v \in K$ tal que la matriz $\big(\sigma_i\sigma_j^{-1}(v)\big)_{ij}$ tiene determinante no nulo.*

DEMOSTRACIÓN: Sea $n = |K : k|$. Sea u un elemento primitivo, es decir, $K = k(u)$. Esto implica que $\operatorname{pol\,mín}(u, k)$ tiene grado n, luego los conjugados $\sigma_1(u), \dots, \sigma_n(u)$ son distintos dos a dos. Sea $g(x) \in K[x]$ un polinomio cuyas raíces sean exactamente los conjugados de u distintos del propio u. Multiplicándolo por la constante adecuada podemos exigir que $g(u) = 1$.

Claramente entonces $g\big(\sigma_j\sigma_i^{-1}(u)\big) = \delta_{ij}$, donde

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

Al aplicar el automorfismo $\sigma_i\sigma_j^{-1}$ a la igualdad $g\big(\sigma_j\sigma_i^{-1}(u)\big) = \delta_{ij}$ obtenemos

$$\big((\sigma_i\sigma_j^{-1})(g)\big)(u) = \delta_{ij},$$

donde $(\sigma_i\sigma_j^{-1})(g)$ es el polinomio que resulta de sustituir los coeficientes de g por sus imágenes por el automorfismo.

Consideremos ahora el polinomio $p(x) = \det\big((\sigma_i\sigma_j^{-1})(g)(x)\big)$. Obviamente es no nulo, pues $p(u) = 1$.

Como tiene un número finito de raíces y el cuerpo k es infinito, existe un $a \in k$ tal que $p(a) \neq 0$. Como $(\sigma_i\sigma_j^{-1})(a) = a$, es claro que

$$\big((\sigma_i\sigma_j^{-1})(g)\big)(a) = (\sigma_i\sigma_j^{-1})\big(g(a)\big),$$

luego si llamamos $v = g(a)$ se cumple $p(a) = \det\big((\sigma_i\sigma_j^{-1})(v)\big) \neq 0$. ∎

Con esto ya podemos probar:

Teorema 9.10 *Si K/k es una extensión finita de Galois y el cuerpo k es infinito entonces K/k tiene una base normal.*

DEMOSTRACIÓN: Sea $G(K/k) = \{\sigma_1, \dots, \sigma_n\}$ y sea v según el teorema anterior. Basta probar que los conjugados $\sigma_1(v), \dots, \sigma_n(v)$ forman una k-base de K. De hecho basta ver que son linealmente independientes. Supongamos que existen elementos $a_1, \dots, a_n \in k$ tales que

$$a_1\sigma_1(v) + \cdots + a_n\sigma_n(v) = 0.$$

Aplicamos σ_j^{-1} para $j = 1, \dots, n$ y obtenemos un sistema de ecuaciones de la forma

$$\sum_{i=1}^{n} a_i(\sigma_i\sigma_j^{-1})(v) = 0, \qquad j = 1, \dots, n.$$

El hecho de que la matriz $\big(\sigma_i\sigma_j^{-1}(v)\big)_{ij}$ tenga determinante no nulo significa que sus columnas son linealmente independientes, luego $a_1 = \cdots = a_n = 0$. ∎

Nos falta demostrar que las extensiones finitas de cuerpos finitos tienen bases normales. Según hemos visto en la sección anterior, estas extensiones son cíclicas, luego el caso restante está incluido en el teorema próximo, con el que concluye la prueba.

Teorema 9.11 *Toda extensión cíclica tiene una base normal.*

DEMOSTRACIÓN: Sea $G(K/k) = \langle\sigma\rangle$. Podemos considerar a σ como un endomorfismo de K como k-espacio vectorial. Sea $p(x)$ su polinomio mínimo (en el sentido de la definición 6.29). Si $|K : k| = n$, tenemos que $\sigma^n = 1$, luego $p(x) \mid x^n - 1$, pero el grado de $p(x)$ no puede ser menor que n, pues si fuera $p_m(x) = a_{n-1}x^{n-1} + \cdots + a_1x + a_0$, tendríamos que

$$a_{n-1}\sigma^{n-1}(\alpha) + \cdots + a_1\sigma(\alpha) + a_0\alpha = 0 \qquad \text{para todo } \alpha \in K,$$

en contradicción con el teorema 5.42. Así pues, $p(x) = x^n - 1$. Esto significa que el $k[x]$-submódulo de K asociado a $p(x)$ tiene dimensión n sobre k, luego es todo K. En otras palabras, que $K = \langle\alpha\rangle_{k[x]}$, para cierto $\alpha \in K$ y, como se ve en la prueba de 6.28, tenemos que $\alpha, \sigma(\alpha), \ldots, \sigma^{n-1}(\alpha)$ es una k-base de K, luego es una base normal. ∎

9.3 Extensiones inseparables

Prácticamente todas las extensiones de cuerpos que hemos manejado en este libro han sido de característica 0 y por tanto separables. Aquí describiremos el comportamiento básico de las extensiones no separables. El lector debe tener presente que es posible adentrarse bastante en la teoría de números sin encontrarse nunca con extensiones no separables, por lo que quizá nunca necesite conocer los resultados de esta sección. No obstante, cuando se trabaja con extensiones de cuerpos infinitos de característica prima (aunque sean separables) es útil conocer el comportamiento de las extensiones no separables para justificar que las extensiones que interesan son realmente separables.

Puesto que todas las extensiones de característica 0 son separables, en toda esta sección supondremos que todos los cuerpos son de característica prima p. En primer lugar veremos cómo factorizan los polinomios en el caso general.

Definición 9.12 Sea k un cuerpo y $f(x) \in k[x]$ un polinomio no constante. Se llama *grado de inseparabilidad* de $f(x)$ a la mayor potencia p^n para la cual $f(x) = g(x^{p^n})$, para cierto $g(x) \in k[x]$.

Obviamente $g(x)$ no puede ser constante, y $\operatorname{grad} f(x) = \operatorname{grad} g(x^{p^n}) = p^n \operatorname{grad} g(x)$, luego ciertamente hay un máximo n que cumple $f(x) = g(x^{p^n})$ para cierto g. El interés de este concepto lo muestra el teorema siguiente:

Teorema 9.13 *Sea k un cuerpo y $f(x) \in k[x]$ un polinomio irreducible con grado de inseparabilidad p^n. Entonces la factorización de $f(x)$ en su cuerpo de escisión es*

$$f(x) = a_0(x - a_1)^{p^n} \cdots (x - a_r)^{p^n},$$

donde $a_1, \ldots, a_r$ son distintos dos a dos.

DEMOSTRACIÓN: Tenemos que $f(x) = g(x^{p^n})$ y el polinomio $g(x)$ no puede expresarse en la forma $g(x) = h(x^p)$, o de lo contrario $f(x) = h(x^{p^{n+1}})$, en contra de la definición del grado de inseparabilidad. El teorema 5.28 nos da entonces que $g'(x) \neq 0$.

Por otra parte $g(x)$ es irreducible, ya que si $g(x) = u(x)v(x)$ entonces también $f(x) = u(x^{p^n})v(x^{p^n})$, luego uno de los factores, digamos $u(x^{p^n})$, es constante, y $u(x)$ también.

Según el teorema 5.27, las raíces de $g(x)$ son simples, luego su factorización en una clausura algebraica de k es de la forma

$$g(x) = a_0(x - b_1)\cdots(x - b_r),$$

donde $b_1, \ldots, b_r$ son distintos dos a dos.

Sea ahora a_i una raíz de $x^{p^n} - b_i$. Entonces

$$f(x) = g(x^{p^n}) = a_0(x^{p^n} - a_1^{p^n})\cdots(x^{p^n} - a_r^{p^n}) = a_0(x - a_1)^{p^n}\cdots(x - a_r)^{p^n}.$$

■

Así pues, todas las raíces de un mismo polinomio irreducible tienen la misma multiplicidad y ésta es potencia de p. El concepto clave en el estudio de las extensiones no separables es el de la inseparabilidad pura, que en cierto sentido es el complementario de la separabilidad.

Definición 9.14 Un elemento algebraico sobre un cuerpo k es *puramente inseparable* si es la única raíz de su polinomio mínimo sobre k. Una extensión K/k es *puramente inseparable* si y sólo si todo elemento de K es puramente inseparable sobre k.

Obviamente un elemento a es a la vez separable y puramente inseparable sobre k si y sólo si $\operatorname{pol\,mín}(a, k) = x - a$, si y sólo si $a \in k$.

Más en general, según el teorema anterior, un elemento a es puramente inseparable sobre k si y sólo si su polinomio mínimo es

$$\operatorname{pol\,mín}(a, k) = (x - a)^{p^n} = x^{p^n} - a^{p^n}.$$

Entonces $a^{p^n} \in k$ y, recíprocamente, si $a^{p^n} \in k$ para cierto n entonces a es raíz del polinomio $(x - a)^{p^n} \in k[x]$, luego $\operatorname{pol\,mín}(a, k) \mid (x - a)^{p^n}$, y a es puramente inseparable sobre k. Es decir, hemos probado el teorema siguiente:

Teorema 9.15 *Un elemento a en una extensión de un cuerpo k es puramente inseparable sobre k si y sólo si existe un número natural n tal que $a^{p^n} \in k$.*

De aquí se sigue fácilmente que la adjunción a un cuerpo de elementos puramente inseparables da lugar a extensiones puramente inseparables y que una cadena de extensiones es puramente inseparable si y sólo si lo son sus términos.

Definición 9.16 Sea K/k una extensión de cuerpos. Definimos la *clausura separable* de K sobre k como el conjunto K_s de todos los elementos de K separables sobre k, y la *clausura puramente inseparable* (o *clausura perfecta*) de K sobre k como el conjunto K_p de todos los elementos de K puramente inseparables sobre k.

Del teorema 5.35 se sigue fácilmente que K_s es un cuerpo, y el teorema 9.15 implica que K_p también lo es.

Definimos el *grado de separabilidad* y el *grado de inseparabilidad* de una extensión K/k como los grados, respectivamente, $|K_s : k|$ y $|K_p : k|$.

Si la extensión K_p/k es finita su grado es potencia de p (por transitividad de grados se reduce al caso de una extensión simple y éste es evidente porque los polinomios mínimos de los elementos puramente inseparables tienen grado potencia de p).

El teorema siguiente nos da un gran control sobre las extensiones de característica prima, pues las reduce a una extensión separable y una puramente inseparable:

Teorema 9.17 *Sea K/k una extensión algebraica. Entonces la extensión K/K_s es puramente inseparable.*

Demostración: Sea $a \in K$. Sea $p(x) = \operatorname{pol}\operatorname{mín}(a, k)$. Sea p^n su grado de inseparabilidad. Entonces $f(x) = g(x^{p^n})$, con $g(x) \in k[x]$. Entonces $g(a^{p^n}) = f(a) = 0$ e igual que en la prueba del teorema 9.13 vemos que $g(x)$ es irreducible y $g'(x) \neq 0$. Por lo tanto a^{p^n} es raíz simple de $g(x)$, y en consecuencia a^{p^n} es separable sobre k, o sea, $a^{p^n} \in K_s$. Según el teorema 9.15 tenemos que K/K_s es puramente inseparable. ∎

Así pues, tenemos la cadena

$$k \subset K_s \subset K,$$

cuyo primer tramo es separable y el segundo puramente inseparable.

Ejemplo Vamos a ver que no es necesariamente cierto que la extensión K/K_p sea separable. Más concretamente, vamos a construir una extensión K/k que no es separable, pero que no contiene elementos puramente inseparables, de modo que $K_p = k$.

Partimos de un cuerpo cualquiera k_0 de característica prima p, tomamos $k = k_0(x, y)$, es decir, el cuerpo de cocientes del anillo de polinomios $k_0[x, y]$, y consideramos el polinimio $f(t) = t^{p^2} + xyt^p + x \in k[t]$. Claramente es irreducible en k por el criterio de Eisenstein.

Sea $K = k(\alpha)$, donde α es una raíz de $f(t)$, de modo que $|K : k| = p^2$.

Como $f'(t)$ es el polinomio nulo, tenemos que α no es separable sobre k. Sin embargo, α^p es raíz del polinomio $t^p + xyt + x$, cuya derivada es la constante xy, luego α^p es separable sobre k. Así pues, $K_s = k(\alpha^p)$.

Supongamos ahora que $K_p \neq k$. Como no puede ser $K_p = K$, ya que K tiene elementos separables sobre k, tiene que ser $|K_p : k| = |K : K_p| = p$.

La factorización de f en su cuerpo de escisión será de la forma

$$f(t) = (t - \alpha_1)^p \cdots (t - \alpha_p)^p.$$

Sea $g(t) = \text{pol mín}(\alpha, K_p)$. Entonces $g(t) \mid f(t)$ en $K_p[t]$. Como $K = K_p(\alpha)$, resulta que $\operatorname{grad} g = p$. Pongamos que $g(t) = \sum_{i=0}^{p} a_i t^i$, con $a_i \in K_p$.

Se cumple que todos los α_j son raíces de $g(t)$. En efecto, para cada j, existe un k-monomorfismo $\sigma_j : K \longrightarrow k(\alpha_j)$ que cumple $\sigma_j(\alpha) = \alpha_j$, pero σ_j deja invariantes a los coeficientes de g, pues, como $a_i \in K_p$, se cumple que $a_i^p \in k$, luego $\sigma_j(a_i)^p = \sigma_j(a_i^p) = a_i^p$, luego $\sigma_j(a_i) = a_i$). Por lo tanto, aplicando σ_j a $g(\alpha) = 0$, resulta que $g(\alpha_j) = 0$. Por consiguiente,

$$g(t) = (t - \alpha_1) \cdots (t - \alpha_p),$$

luego $f(t) = g(t)^p = \sum_{i=0}^{p} a_i^p t^{pi}$.

Comparando coeficientes vemos que $g(t) = t^p + a_1 t + a_0$, con $a_1^p = xy$, $a_0^p = x$. Si llamamos $u = a_0$, $v = a_1/a_0$, tenemos que $u, v \in K_p$ y $x = u^p$, $y = v^p$. Así pues, tenemos la cadena de extensiones:

$$k \subset k(u) \subset k(u, v) \subset K_p.$$

La extensión completa tiene grado p. Sin embargo, vamos a ver que las dos primeras extensiones tienen grado p, con lo que tendremos una contradicción.

El caso de la primera extensión es inmediato: el polinomio $t^p - x$ es irreducible en $k[t]$, por el criterio de Eisenstein. Como u es raíz de dicho polinomio, necesariamente $|k(u) : k| = p$.

Consideremos ahora la evaluación $k_0[x, y] \longrightarrow k_0[x^p, y]$ definida mediante $p(x, y) \mapsto p(x^p, y)$. Se trata de un isomorfismo de anillos, que induce un isomorfismo de cuerpos $\sigma : k \longrightarrow l$, donde $l = k_0(x^p, y)$ es el cuerpo de cocientes del anillo $k_0[x^p, y]$. Como u es raíz de $t^p - x \in k[t]$ y x es raíz de $t^p - x^p \in l[t]$, el teorema 5.12 nos da que σ se extiende a un isomorfismo $\sigma^* : k(u) \longrightarrow l(x) = k$ tal que $\sigma^*(u) = x$ (y $\sigma^*(y) = y$).

Pero el polinomio $t^p - y$ es irreducible en $k[t]$ y su imagen en $l[t]$ es él mismo, luego también es irreducible en este anillo, luego no tiene raíces en l, luego su antiimagen en $k(u)[t]$ (que es él mismo) no tiene raíces en $k(u)$. Esto implica que $|k(u, v) : k(u)| = p$. ∎

El teorema siguiente nos permitirá relacionar los cuerpos K_p y K_s de una extensión arbitraria:

Teorema 9.18 *Sea K/k una extensión puramente inseparable y $\sigma : k \longrightarrow C$ un monomorfismo de k en una clausura algebraica de K. Entonces σ admite una única extensión a K.*

Demostración: Por el teorema 5.54 sabemos que σ admite al menos una extensión σ^*. Ésta es única, pues si $a \in K$ entonces existe un n tal que $a^{p^n} \in k$, luego $\sigma^*(a)^{p^n} = \sigma(a^{p^n})$, con lo que $\sigma^*(a)$ es necesariamente la única raíz del polinomio $x^{p^n} - \sigma(a^{p^n})$ en C. ∎

Teorema 9.19 *Sea K/k una extensión algebraica. Entonces*

$$K = K_s K_p \qquad y \qquad k = K_s \cap K_p.$$

Si además es finita de grado n, su grado de separabilidad es n_s y su grado de inseparabilidad es n_p entonces

$$n = n_s n_p, \qquad |K : K_p| = n_s \qquad y \qquad |K : K_s| = n_p.$$

Demostración: Ya sabemos que $k = K_s \cap K_p$. La extensión K/K_sK_p es separable y puramente inseparable, luego $K = K_sK_p$.

El grado de separabilidad de K/k es el número de k-monomorfismos de K_s y por el teorema anterior cada uno de ellos se extiende a un único k-monomorfismo de K.

Por otra parte es obvio que todo k-monomorfismo de K es un K_p-monomorfismo de K (un k-monomorfismo de K envía un elemento puramente inseparable de K a un k-conjugado, o sea, a sí mismo). Así pues n_s es el número de K_p-monomorfismos de K, y como K/K_p es separable esto es el grado $|K : K_p|$. La cadena $k \subset K_p \subset K$ nos da ahora la igualdad $n = n_s n_p$ y la cadena $k \subset K_s \subset K$ nos da $|K : K_s| = n_p$. ∎

Conviene destacar que en la prueba anterior hemos visto que en general el número de k-monomorfismos de una extensión finita K/k es el grado de separabilidad de la extensión.

Ejercicio: Probar que si $k \subset K \subset K$ es una cadena de extensiones finitas, entonces el grado de separabilidad de K/k es el producto de los grados de separabilidad/inseparabilidad de las extensiones intermedias.

Algunas observaciones sencillas: toda extensión puramente inseparable es normal, y si K/k es una extensión normal entonces

$$G(K/k) = G(K/K_p) \cong G(K_s/k), \qquad G(K_p/k) = 1.$$

Terminamos con una aplicación:

Teorema 9.20 *Si K/k es una extensión algebraica tal que todo polinomio de $k[x]$ no constante tiene una raíz en K, entonces K es una clausura algebraica de K.*

Demostración: Por 5.47 basta probar que todo polinomio no constante $f(x) \in k[x]$ se escinde en $K[x]$. Sea F un cuerpo de escisión de f sobre K y sea L la adjunción a k de las raíces de f en F, de modo que L es un cuerpo de escisión de f sobre k. Basta probar que $L \subset K$ y, por el teorema anterior, para ello basta con que L_s, $L_p \subset K$.

El caso de L_p es sencillo: si $\alpha \in L_p$, entonces α es la única raíz de su polinomio mínimo sobre k, luego por hipótesis $\alpha \in K$.

Por el teorema del elemento primitivo, $L_s = k(\beta)$, para cierto $\beta \in L_s$. Por hipótesis el polinomio mínimo de β sobre k tiene una raíz $\beta' \in K$. Ahora bien, $\beta' \in L$ y es separable sobre k, luego $\beta' \in L_s$, luego $L_s = k(\beta') \subset K$. ∎

9.4 Extensiones trascendentes

En este libro hemos estudiado únicamente las extensiones algebraicas de cuerpos, pero en algunos contextos, sobre todo relacionados con la geometría algebraica, aparecen también extensiones trascendentes y es necesario conocer algunos resultados fundamentales sobre ellas, que ahora vamos a presentar. El ejemplo más simple de extensión trascendente es la determinada por un cuerpo de fracciones algebraicas $k(S)$ (es decir, el cuerpo de cocientes de un anillo de polinomios $k[S]$, donde S es un conjunto de indeterminadas) sobre su cuerpo de constantes.

Vamos a ver que toda extensión de cuerpos es esencialmente una extensión algebraica sobre una extensión de este tipo. Esto nos lleva al concepto de independencia algebraica:

Definición 9.21 Sea K/k una extensión de cuerpos. Un conjunto $S \subset K$ es *algebraicamente dependiente* sobre k si existen elementos $s_1, \ldots, s_n \in S$ distintos dos a dos y un polinomio $f \in k[x_1, \ldots, x_n]$ no nulo tal que $f(s_1, \ldots, s_n) = 0$. En caso contrario se dice que S es *algebraicamente independiente* sobre k.

Claramente, $\{s\}$ es algebraicamente independiente sobre k si y sólo si s es trascendente sobre k. Si $S \subset T$ y T es algebraicamente independiente, también lo es S. En particular, todos los elementos de un conjunto algebraicamente independiente sobre k son trascendentes sobre k.

La idea es que unos elementos son algebraicamente independientes si y sólo si no satisfacen ninguna clase de relación algebraica. Por ejemplo, π y π^2 son trascendentes sobre $\mathbb{Q}$, pero no son algebraicamente independientes, porque uno es el cuadrado del otro, y esto hace que $f(\pi, \pi^2) = 0$, donde $f(x, y) = x^2 - y$.

El ejemplo más simple de conjunto algebraicamente independiente es el del conjunto S de indeterminadas de un cuerpo de fracciones algebraicas $k(S)$. Al sustituir un número finito de indeterminadas en un polinomio de varias variables, lo que obtenemos es un polinomio con otras indeterminadas, pero con los mismos coeficientes, y el resultado será nulo si y sólo si el polinomio original era nulo.

Teorema 9.22 *Sean K/k y K'/k dos extensiones de un mismo cuerpo k, sean $S \subset K$ y $S' \subset K'$ algebraicamente independientes y sea $f : S \longrightarrow S'$ una aplicación biyectiva. Entonces f se extiende a un único k-isomorfismo de cuerpos $\bar{f} : k(S) \longrightarrow k(S')$.*

Demostración: Lo probamos primero en el caso en que $K = k(S)$ es un cuerpo de fracciones algebraicas, es decir, el cuerpo de cocientes del anillo de polinomios $k[S]$. Entonces la sustitución $\bar{f} : k[S] \longrightarrow k(S')$ dada por 2.38 es un homomorfismo de anillos que extiende a f y fija a los elementos de k. El hecho de que S' sea algebraicamente independiente significa precisamente que el núcleo de $\bar{f}$ es trivial, luego se trata de un monomorfismo. El teorema 2.22 nos da entonces que $\bar{f}$ se extiende a un k-monomorfismo de cuerpos $\bar{f} : k(S) \longrightarrow k(S')$. Como la imagen es un cuerpo que contiene a k y a S', tiene que ser $k(S')$, luego $\bar{f}$ es un k-isomorfismo. La unicidad de los teoremas que hemos citado implica claramente la unicidad de $\bar{f}$.

En el caso general, consideramos un cuerpo de fracciones algebraicas $k(S'')$ cuyo conjunto de indeterminadas tenga el mismo cardinal que S y S'. Tomamos cualquier biyección $h : S'' \longrightarrow S$ y definimos $h' = h \circ f : S'' \longrightarrow S'$. Por la parte ya probada existen k-isomorfismos $\bar{h} : k(S'') \longrightarrow k(S)$ y $\bar{h}' : k(S'') \longrightarrow k(S')$, y al componerlos obtenemos el k-isomorfismo $\bar{f} : k(S) \longrightarrow k(S')$ requerido. La unicidad de $\bar{h}$ y $\bar{h}'$ implica la de $\bar{f}$. ∎

Definición 9.23 Sea K/k una extensión de cuerpos. Una *base de trascendencia* de K sobre k es un conjunto $S \subset K$ algebraicamente independiente sobre k y maximal respecto de la inclusión.

Una extensión de cuerpos K/k es *puramente trascendente* si $K = k(S)$, donde S es un conjunto algebraicamente independiente sobre k (y por consiguiente una base de trascendencia). El teorema anterior implica que una extensión K/k es puramente trascendente si y sólo si K es k-isomorfo a un cuerpo de fracciones algebraicas sobre k.

Las propiedades básicas de las bases de trascendencia se siguen del resultado siguiente:

Teorema 9.24 *Sea K/k una extensión de cuerpos y $S \subset K$ un conjunto algebraicamente independiente sobre k. Entonces un elemento $u \in K \setminus k(S)$ es trascendente sobre $k(S)$ si y sólo si $S \cup \{u\}$ es algebraicamente independiente.*

Demostración: Supongamos que u es trascendente sobre $k(S)$ y supongamos que $f(s_1, \ldots, s_n, u) = 0$, donde f es un polinomio con coeficientes en k y $s_1, \ldots, s_n \in S$. Podemos expresar esta ecuación en la forma

$$\sum_i g_i(s_1, \ldots, s_n) u^i = 0,$$

para ciertos polinomios g_i con coeficientes en k. El hecho de que u sea trascendente sobre $k(S)$ implica que $g_i(s_1, \ldots, s_n) = 0$ para todo i, y como S es algebraicamente independiente los coeficientes de cada g_i son nulos, pero éstos son los coeficientes de f, luego $f = 0$.

Supongamos ahora que $S \cup \{u\}$ es algebraicamente independiente y que $f(u) = 0$, donde $f \in k(S)[x]$. Entonces

$$f(u) = \sum_i \frac{g_i(s_1, \ldots, s_n)}{h_i(s_1, \ldots, s_n)} u^i = 0,$$

donde g_i y h_i son polinomios con coeficientes en k. Multiplicando por el producto de los denominadores obtenemos una ecuación similar, pero en la que los coeficientes son polinomios. De la hipótesis se sigue fácilmente que todos los coeficientes han de ser nulos, de donde también $g_i = 0$ para todo i, es decir, $f(x)$ es el polinomio nulo, luego u es trascendente sobre $k(S)$. ∎

Como consecuencia inmediata tenemos:

Teorema 9.25 *Sea K/k una extensión de cuerpos y $S \subset K$ un conjunto algebraicamente independiente. Entonces S es una base de trascendencia si y sólo si la extensión $K/k(S)$ es algebraica.*

Así, por ejemplo, si una extensión K/k es algebraica, entonces $\varnothing$ es una base de trascendencia. Veamos otra consecuencia sencilla:

Teorema 9.26 (AE) *Sea K/k una extensión de cuerpos y $S \subset K$ un conjunto arbitrario tal que la extensión $K/k(S)$ sea algebraica. Entonces S contiene*[2] *una base de trascendencia de K/k.*

DEMOSTRACIÓN: Sea $T \subset S$ un conjunto algebraicamente independiente maximal respecto a la inclusión. Por el teorema 9.24 resulta que todo elemento de $S \setminus T$ es algebraico sobre $k(T)$, luego la extensión $k(S)/k(T)$ es algebraica, y $K/k(T)$ también lo es. Por consiguiente, T es una base de trascendencia. ∎

En particular, toda extensión finitamente generada (es decir, de la forma $K = k(S)$, con $S \subset K$ finito) contiene una base de trascendencia finita.

Ahora es claro que toda extensión de cuerpos se descompone en una extensión puramente trascendente seguida de una extensión algebraica.

El resultado fundamental sobre bases de trascendencia es el siguiente:

Teorema 9.27 (AE) *Sea K/k una extensión de cuerpos. Si S y T son dos bases de trascendencia de K sobre k, entonces existe una biyección $S \longrightarrow T$.*

DEMOSTRACIÓN: Supongamos en primer lugar que la extensión tiene una base de trascendencia finita $S = \{s_1, \ldots, s_n\}$ y sea T cualquier otra base de trascendencia. No puede ocurrir que todo elemento de T sea algebraico sobre $k(s_2, \ldots, s_n)$, pues entonces K sería algebraico sobre este cuerpo, y en particular lo sería s_1, lo cual es imposible. Por lo tanto existe $t_1 \in T$ trascendente sobre $k(s_2, \ldots, s_n)$. Por el teorema 9.24 tenemos que el conjunto $\{t_1, s_2, \ldots, s_n\}$ es algebraicamente independiente sobre K. Más aún, s_1 es algebraico sobre este conjunto, o de lo contrario podríamos añadirlo y tendríamos un conjunto algebraicamente independiente que contendría estrictamente a S. De aquí se sigue fácilmente que K es algebraico sobre $k(t_1, s_2 \ldots, s_n)$, lo que implica que $\{t_1, s_2, \ldots, s_n\}$ es una base de trascendencia. Repitiendo el proceso podemos llegar a una base de trascendencia $\{t_1, \ldots, t_n\}$ formada por elementos de T. Por maximalidad $T = \{t_1, \ldots, t_n\}$ luego, efectivamente, T tiene también n elementos.

Supongamos ahora que K/k tiene una base de trascendencia infinita S. Por la parte ya probada, cualquier otra base de trascendencia T es también infinita. Cada $s \in S$ es algebraico sobre $k(T)$, luego es algebraico sobre $k(T_s)$, para un

[2]Si S es finito o numerable no es necesario el axioma de elección. En la prueba, éste se emplea al tomar el conjunto maximal T, lo cual, en general, requiere el lema de Zorn, pero si S es finito o numerable, el conjunto T puede construirse recurrentemente a partir de una enumeración de S.

cierto conjunto finito $T_s \subset T$. Entonces K es algebraico sobre la adjunción a k de $\bigcup_{s\in S} T_s$, luego este conjunto es una base de trascendencia de K/k. Como está contenido en T ha de ser igual a T, es decir, $T = \bigcup_{s\in S} T_s$. A partir de aquí la prueba concluye exactamente igual que la de 4.56. ∎

Definición 9.28 Llamaremos *grado de trascendencia* de una extensión de cuerpos K/k al cardinal[3] de cualquiera de sus bases de trascendencia. Lo representaremos por $\mathrm{gt}(K/k)$.

Así, las extensiones algebraicas son las extensiones con grado de trascendencia igual a 0.

Una aplicación:

Teorema 9.29 (AE) *Si K y K' son cuerpos algebraicamente cerrados de la misma característica y con el mismo grado de trascendencia sobre su cuerpo primo, entonces son isomorfos.*

Demostración: Sean k y k' los cuerpos primos correspondientes. Como son de la misma característica, son isomorfos. La hipótesis sobre los grados de trascendencia significa que K y K' tienen bases de trascendencia S y S' sobre k y k', respectivamente, tales que existe una biyección $f : S \longrightarrow S'$. Dicha biyección se extiende a un isomorfismo $k(S) \longrightarrow k'(S')$. Además K es la clausura algebraica de $k(S)$ y análogamente con K' y $k'(S')$. Como cuerpos isomorfos tienen clausuras algebraicas isomorfas, concluimos que K y K' son isomorfos. ∎

Nota El lector familiarizado con la teoría de cardinales infinitos no tendrá dificultad en probar los hechos siguientes: sea K un cuerpo arbitrario, sea k su cuerpo primo y sea S una base de trascendencia de K sobre k. Si S es infinito, teniendo en cuenta que k es numerable (finito o infinito), entonces $|k(S)| = |S|$ y, como $K/k(S)$ es una extensión algebraica, se cumple también que $|K| = |k(S)| = |S|$. Por el contrario, si S es infinito, del mismo modo llegamos a que $|K| = \aleph_0$.

Como conclusión, si K es cualquier cuerpo no numerable, entonces su grado de trascendencia sobre su cuerpo primo es simplemente $|K|$, luego el teorema anterior implica lo siguiente:

> *Dos cuerpos algebraicamente cerrados no numerables de la misma característica son isomorfos si y sólo si tienen el mismo cardinal.*

Otra consecuencia es que el grado de trascendencia de $\mathbb{R}$ sobre $\mathbb{Q}$ (o de $\mathbb{C}$ sobre $\mathbb{Q}$) es $\mathfrak{c}$. ∎

[3]En principio tenemos que hacer las mismas observaciones que hemos hecho sobre el concepto de dimensión de un espacio vectorial: con los resultados expuestos en el apéndice B tenemos definido el concepto de grado de trascendencia para extensiones finitamente generadas o, a lo sumo, con un generador numerable. La definición general requiere el hecho de que es posible asignar un cardinal a un conjunto arbitrario.

Así como el grado de una cadena de extensiones es multiplicativo, el grado de trascendencia es aditivo:

Teorema 9.30 *Si $k \subset K \subset L$ es una cadena de extensiones de cuerpos, entonces*

$$\mathrm{gt}(L/k) = \mathrm{gt}(L/K) + \mathrm{gt}(K/k).$$

DEMOSTRACIÓN: El enunciado de este teorema, tal y como lo hemos formulado involucra el concepto general del cardinal de un conjunto arbitrario, así como la suma de cardinales, pero es posible reformularlo en términos que no requieren ninguno de estos conceptos:

> *Si S es una base de trascendencia de L sobre K y S' es una base de trascendencia de K sobre k, entonces $S \cup S'$ es una base de trascendencia de L sobre k (y $S \cap S' = \varnothing$).*

En efecto: $S \cap S' = \varnothing$ porque, más en general, $S \cap K = \varnothing$, ya que los elementos de S son trascendentes sobre K. Si llamamos $A = L \setminus K(S)$ y $A' = K \setminus k(S')$, entonces todos los elementos de A son algebraicos sobre $K(S)$ y todos los elementos de A' son algebraicos sobre $k(S')$. Además

$$L = K(S \cup A) = k(S' \cup A')(S \cup A) = k(S \cup S' \cup A \cup A') = k(S \cup S')(A')(A),$$

y los elementos de A' son algebraicos sobre $k(S')$, luego también sobre $k(S \cup S')$, y los elementos de A son algebraicos sobre $K(S) = k(S \cup S' \cup A')$, luego también sobre $k(S \cup S' \cup A')$, luego la extensión $L/k(S \cup S')$ es algebraica.

Sólo falta probar que $S \cup S'$ es algebraicamente independiente. Ahora bien, si $f(x_1, \ldots, x_n, y_1, \ldots, y_m)$ es un polinomio tal que $f(s_1, \ldots, s_n, s'_1, \ldots, s'_m) = 0$, con $s_i \in S$, $s'_i \in S'$ distintos entre sí, entonces, $s_1, \ldots, s_n$ anulan el polinomio $f(x_1, \ldots, x_n, s'_1, \ldots, s'_m) \in K[x_1, \ldots, x_n]$, luego todos sus coeficientes, que son polinomios de $k[y_1, \ldots, y_m]$ evaluados en $s'_1, \ldots, s'_m)$ son nulos, por la independencia algebraica de S, luego los coeficientes de dichos coeficientes (que son los coeficientes de f) son nulos por la independencia algebraica de S'. Por lo tanto $f = 0$ y $S \cup S$ es algebraicamente independiente. ∎

9.5 Cuerpos linealmente disjuntos

Vamos a introducir un concepto con el que obtendremos varias aplicaciones de interés.

Teorema 9.31 *Sea Ω/k una extensión de cuerpos y sean K y L dos cuerpos intermedios. Las afirmaciones siguientes son equivalentes:*

1. *Si $\{\alpha_i\}_{i\in I} \subset K$ y $\{\beta_j\}_{j\in J} \subset L$ son familias linealmente independientes sobre k, entonces $\{\alpha_i\beta_j\}_{(i,j)\in I\times J}$ es una familia (de elementos distintos dos a dos) linealmente independiente sobre k.*

2. *Todo subconjunto de K linealmente independiente sobre k es linealmente independiente sobre L.*

3. *Todo subconjunto de L linealmente independiente sobre k es linealmente independiente sobre K.*

DEMOSTRACIÓN: Por simetría basta probar la equivalencia entre 1) y 2). Supongamos 1), sea $\{\alpha_i\}_{i\in I} \subset K$ una familia linealmente independiente sobre k y sea $\{\beta_j\}_{j\in J} \subset L$ una k-base de L.

Supongamos que $I_0 \subset I$ es finito y que$\{\delta_i\}_{i\in I_0}$ son elementos de L tales que $\sum_{i\in I_0} \alpha_i\delta_i = 0$. Podemos expresar $\delta_i = \sum_{j\in J_0} a_{ij}\beta_j$, para cierto $J_0 \subset J$ finito y ciertos $a_{ij} \in k$. Entonces $\sum_{i,j} a_{ij}\alpha_i\beta_j = 0$ y los $\alpha_i\beta_j$ son linealmente independientes sobre k, luego $a_{ij} = 0$ y, por consiguiente, $\delta_i = 0$.

Supongamos ahora 2) y, en las hipótesis de 1), supongamos que existen $I_0 \subset I$, $J_0 \subset J$ finitos y $a_{ij} \in k$ tales que $\sum_{(i,j)\in I_0\times J_0} a_{ij}\alpha_i\beta_j = 0$. Entonces, por la independencia lineal de los α_i sobre L tenemos que $\sum_{j\in J_0} a_{ij}\beta_j = 0$ y de aquí a su vez $a_{ij} = 0$. Esto prueba en particular que los $\alpha_i\beta_j$ son distintos dos a dos. ∎

Definición 9.32 Sea Ω/k una extensión de cuerpos y sean K y L dos cuerpos intermedios. Diremos que K y L son *linealmente disjuntos* sobre k si cumplen cualquiera de las condiciones del teorema anterior.

Observaciones 1) En realidad es suficiente comprobar cualquiera de las condiciones del teorema anterior para familias finitas (pues una familia infinita es linealmente independiente si y sólo si lo son todos sus subconjuntos finitos).

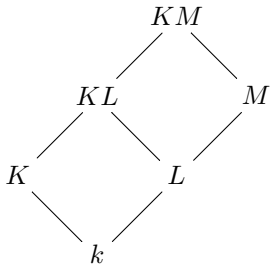
2) Notemos que si K y L son linealmente disjuntos sobre k, se cumple que $K\cap L = k$, pues si $\alpha \in K\cap L$ no está en k, tenemos que $1, \alpha \in K$ son linealmente independientes sobre k, pero no sobre L.

3) Supongamos que, en las condiciones de la definición anterior, K es el cuerpo de cocientes de un dominio D que contiene a k y sea $\{d_i\}_{i\in I} \subset D$ una k-base de D. Entonces, para que K y L sean linealmente disjuntos basta con que $\{d_i\}_{i\in I}$ sea linealmente independiente sobre L.

En efecto, si $\alpha_1, \ldots, \alpha_n$ son elementos de K linealmente independientes sobre k, podemos expresarlos como $\alpha_i = d_i'/d$, con $d_i \in D$, $d \in D$, $d \neq 0$. Es obvio que $d_1', \ldots, d_n'$ son también linealmente independientes sobre k. Sean $i_1, \ldots, i_m \in I$ tales que $d_1', \ldots, d_n' \in \langle d_{i_1}, \ldots, d_{i_m}\rangle_k$ y completemos una base $d_1', \ldots, d_m'$ de este espacio vectorial. Por hipótesis $d_{i_1}, \ldots, d_{i_m}$ son linealmente independientes sobre L y están en $\langle d_1', \ldots, d_m'\rangle_k \subset \langle d_1', \ldots, d_m'\rangle_L$. Por consiguiente $d_1', \ldots, d_m'$ tienen que ser linealmente independientes sobre L, y a su vez $\alpha_1, \ldots, \alpha_n$ también. ∎

A menudo usaremos el teorema siguiente:

Teorema 9.33 *Sea Ω/k una extensión de cuerpos y sean K, L, M cuerpos intermedios tales que $L \subset M$. Entonces K y M son linealmente disjuntos sobre k si y sólo si K y L son linealmente disjuntos sobre k y KL y M son linealmente disjuntos sobre L.*



DEMOSTRACIÓN: Si K, L y KL, M son linealmente disjuntos sobre k y $\{\alpha_i\}_{i\in I} \subset K$ es linealmente independiente sobre k, entonces también lo es sobre L y, a su vez, sobre M.

Supongamos ahora que K y M son linealmente disjuntos. Entonces toda familia en K linealmente independiente sobre k lo es sobre M, luego también sobre L, lo que prueba que K y L son linealmente disjuntos sobre k. Por otra parte, KL es el cuerpo de cocientes del dominio $K[L]$, formado por los elementos de la forma $\sum_i \alpha_i\beta_i$, con $\alpha_i \in K$, $\beta_i \in L$. Ahora bien, es obvio que una k-base $\{\alpha_i\}_{i\in I}$ de K es un generador de $K[L]$ como L-espacio vectorial y, por la parte ya probada es, de hecho, una base. Por la nota previa al teorema, basta probar que esta base es linealmente independiente sobre M, pero esto es justo lo que afirma la hipótesis. ∎

Un caso sencillo de cuerpos linealmente disjuntos es el dado por el teorema siguiente:

Teorema 9.34 *Sea Ω/k una extensión de cuerpos, sea K un cuerpo intermedio y sea $S \subset \Omega$ un conjunto de elementos algebraicamente independientes sobre K. Entonces K y $k(S)$ son linealmente disjuntos sobre k.*

DEMOSTRACIÓN: El cuerpo $k(S)$ es el cuerpo de cocientes del dominio $k[S]$, y una k-base de $k[S]$ la forman los monomios con coeficiente 1. Por la tercera observación tras la definición, basta probar que dicha base es linealmente independiente sobre K, pero esto es justo lo que significa que S es algebraicamente independiente sobre K. ∎

Con esto ya podemos presentar algunas aplicaciones. Si $k \subset L \subset K$ y la extensión K/k es finitamente generada, es obvio que K/L también lo es, pero no es tan obvio que L/k lo sea. Vamos a probarlo:

Teorema 9.35 *Si $k \subset L \subset K$ son extensiones de cuerpos y K/k es finitamente generada, entonces L/k también lo es.*

DEMOSTRACIÓN: Por 9.30 sabemos que el grado de trascendencia de L/k es finito. Más aún, si $y_1, \ldots, y_r$ es una base de trascendencia de L/k y $y_{r+1}, \ldots, y_n$ es una base de trascendencia de K/L, entonces $y_1, \ldots, y_n$ es una base de trascendencia de K/k. El teorema anterior nos da que L y $k(y_1, \ldots, y_n)$ son linealmente disjuntos sobre $k(y_1, \ldots, y_r)$.

Basta probar que la extensión algebraica $L/k(y_1, \ldots, y_r)$ tiene grado finito, pues entonces es finitamente generada, y lo mismo vale para L/k. Si hubiera familias infinitas en L linealmente independientes sobre $k(y_1, \ldots, y_r)$, también lo serían sobre $k(y_1, \ldots, y_n)$, luego el grado de $K/k(y_1, \ldots, y_n)$ sería infinito, pero esto es imposible, porque se trata de una extensión algebraica finitamente generada. ∎

Si K/k es una extensión de cuerpos, se dice que k es *algebraicamente cerrado* en k si todo elemento de K algebraico sobre k está en k.

Teorema 9.36 *Si S es algebraicamente independiente sobre k, entonces k es algebraicamente cerrado en $k(S)$.*

DEMOSTRACIÓN: Sea $\bar{k}$ una clausura algebraica de k y consideremos el cuerpo $\bar{k}(S')$ de fracciones algebraicas con coeficientes en $\bar{k}$ para un conjunto de indeterminadas S' del mismo cardinal que S. Por el teorema 9.34 tenemos que $k(S')$ es linealmente disjunto de $\bar{k}$, luego $k(S') \cap \bar{k} = k$, lo que significa que k es algebraicamente cerrado en $k(S')$ y, como $k(S)$ y $k(S')$ son k-isomorfos, lo mismo vale para $k(S)$. ∎

En otras palabras, si K/k es una extensión puramente trascendente, todos los elementos de $K \setminus k$ son trascendentes sobre k.

Teorema 9.37 *Dados cuerpos $k \subset K \subset \Omega$ con k algebraicamente cerrado en K, si $\alpha \in \Omega$ es algebraico sobre k, entonces $k(\alpha)$ y K son linealmente disjuntos sobre k y $|k(\alpha) : k| = |K(\alpha) : K|$.*

DEMOSTRACIÓN: Sea $p(X)$ el polinomio mínimo de α sobre k. Se cumple que $p(X)$ es irreducible en $K[X]$, pues un factor de $p(X)$ será de la forma $(X - \alpha_1) \cdots (X - \alpha_r)$, donde los $\alpha_i \in \Omega$ son algebraicos sobre k, luego los coeficientes del factor serían elementos de K algebraicos sobre k, luego tendrían que estar en k. Esto prueba la igualdad de los grados del enunciado. Además, las primeras potencias de α forman una k-base de $k(\alpha)$ y también una K-base de $K(\alpha)$, y el hecho de que una base siga siendo independiente basta para probar que $k(\alpha)$ y K son linealmente disjuntos (es un caso particular de tercera observación tras la definición 9.32). ∎

Hay una propiedad de interés que es ligeramente más débil que la de ser linealmente disjuntos:

Definición 9.38 Sea Ω/k una extensión de cuerpos y sean K y L dos cuerpos intermedios. Diremos que K y L son *libres* sobre k si cuando $S \subset K$ es algebraicamente independiente sobre k, también lo es sobre L.

Notemos que, en tal caso, también se cumple que si $S \subset L$ es algebraicamente independiente sobre k, también lo es sobre K.

En efecto, basta probarlo si $S \subset L$ es finito, digamos $S = \{s_1, \ldots, s_n\}$. Si S no es algebraicamente independiente sobre K, existe $F \in k[X_1, \ldots, X_n]$ no nulo tal que $F(s_1, \ldots, s_n) = 0$. Sea K' la adjunción a k de los coeficientes de F, de modo que K'/k es una extensión finitamente generada y S no es algebraicamente independiente sobre K'.

Así, si $S' = \{s'_1, \ldots, s'_r\}$ es una base de trascendencia de K' sobre k, por hipótesis es algebraicamente independiente sobre L, luego sobre $k(S)$. Por lo tanto, $\mathrm{gt}(K'/k) = r = \mathrm{gt}(K'(S)/k(S))$, luego

$$\mathrm{gt}(K'(S)/k) = \mathrm{gt}(K'(S)/k(S)) + \mathrm{gt}(k(S)/k) = r + n,$$

pero

$$\mathrm{gt}(K'(S)/k) = \mathrm{gt}(K'(S)/K') + \mathrm{gt}(K'/k) < n + r,$$

contradicción.

Teorema 9.39 *Sea Ω/k una extensión de cuerpos y sean K y L dos cuerpos intermedios. Si K y L son linealmente disjuntos sobre k, entonces son libres sobre k.*

Demostración: Sean $s_1, \ldots, s_n \in K$ algebraicamente independientes sobre k. Si fueran dependientes sobre K, tendríamos una relación de la forma $\sum_i \alpha_i M_i(s_1, \ldots, s_n)$, donde los M_i son monomios y con los $\alpha_i \in L$ no todos nulos.

Pero los elementos $M_i(s_1, \ldots, s_n)$ son linealmente independientes sobre k, o de lo contrario $s_1, \ldots, s_n$ serían algebraicamente dependientes, luego por hipótesis son linealmente independientes sobre L, lo que implica que todos los α_i son nulos, contradicción. ∎

9.6 Extensiones separables

Vamos a generalizar el concepto de extensión separable al caso de extensiones no necesariamente algebraicas. Conviene probar algunos resultados previos para tratar con los cuerpos de característica prima.

Si k es un cuerpo de característica prima p, representaremos por k^{1/p^n} a la adjunción a k de las raíces p^n-ésimas de los elementos de k en una clausura algebraica. Similarmente, k^{1/p^∞} será la adjunción a k de todas las raíces p^n-ésimas de los elementos de k, para todo n, en una clausura algebraica. El teorema 5.30 implica que k es perfecto si y sólo si $k = k^{1/p}$, si y sólo si $k = k^{1/p^\infty}$.

Teorema 9.40 *Consideremos cuerpos $k \subset K \subset \Omega$, donde Ω es algebraicamente cerrado de característica prima p. Un conjunto $X \subset K$ es linealmente independiente sobre k^{1/p^n} si y sólo si el conjunto $X^{p^n} = \{x^{p^n} \mid x \in X\}$ es linealmente independiente sobre k. Además, X es linealmente independiente sobre k^{1/p^∞} si y sólo si lo es sobre k^{1/p^n} para todo n.*

DEMOSTRACIÓN: Observemos que todo $a \in k$ es de la forma $a = u^{p^n}$, para cierto $u \in k^{1/p^n}$ y, recíprocamente, todo $u \in k^{1/p^n}$ cumple que $a = u^{p^n} \in k$.

Así, $\sum_i a_i x_i^{p^n} = 0$ (con $a_i \in k$, $x_i \in X$) es equivalente a $\sum_i u_i^{p^n} x_i^{p^n} = 0$, que a su vez equivale a $\sum_i u_i x_i = 0$ (con $u_i \in k^{1/p^n}$), de donde se sigue inmediatamente la primera parte del enunciado. Para la segunda parte basta observar que si $\sum_i u_i x_i = 0$ con $u_i \in k^{1/p^\infty}$, entonces existe un n suficientemente grande tal que todos los u_i están en k^{1/p^n}. ∎

Teorema 9.41 *Consideremos cuerpos $k \subset K \subset \Omega$ de modo que Ω es algebraicamente cerrado de característica prima p y la extensión K/k es algebraica y separable. Entonces K y k^{1/p^∞} son linealmente disjuntos sobre k.*

DEMOSTRACIÓN: Sean $\alpha_1, \ldots, \alpha_n \in K$ linealmente independientes sobre k. Tenemos que probar que son linealmente independientes sobre k^{1/p^∞}. No perdemos generalidad si suponemos que $K = k(\alpha_1, \ldots, \alpha_n)$, con lo que la extensión K/k es finita, digamos de grado r, y podemos extender el conjunto dado hasta una k-base $\alpha_1, \ldots, \alpha_r$ de K.

Si $\alpha \in K$ y $m \geq 1$, entonces $\alpha^m = \sum_i a_i \alpha_i$, con $a_i \in k$, luego $\alpha^{mp} = \sum_i a_i^p \alpha_i^p$. Como α es separable sobre k, tenemos que $k(\alpha)$ es separable y puramente inseparable sobre $k(\alpha^p)$, luego $k(\alpha) = k(\alpha^p)$, luego α se expresa como polinomio en α^p, es decir, como combinación lineal de los α^{mp}, luego también de los α_i^p. Esto prueba que $\alpha_1^p, \ldots, \alpha_r^p$ son un generador de K sobre k, y como su número coincide con la dimensión, son de hecho una base.

Partiendo ahora de esta base concluimos que $\alpha_1^{p^2}, \ldots, \alpha_r^{p^2}$ también es una base y, en general, que $\alpha_1^{p^m}, \ldots, \alpha_r^{p^m}$ son linealmente independientes sobre k, luego, por el teorema anterior, $\alpha_1, \ldots, \alpha_n$ son linealmente independientes sobre k^{1/p^n} para todo n, luego son linealmente independientes sobre k^{1/p^∞}. ∎

Si k es un cuerpo de característica prima, aunque sea perfecto, un cuerpo de fracciones algebraicas $K = k(S)$ ya no lo es, pues no cumple $K = K^{1/p}$ (puesto que las indeterminadas no tienen raíz p-ésima en K). Por lo tanto, las extensiones algebraicas de $k(S)$ no son necesariamente separables o, equivalentemente, si S es una base de trascendencia de una extensión K/k, la extensión $K/k(S)$ es algebraica, pero no necesariamente separable. No obstante, vamos a ver que en ocasiones (por ejemplo, siempre que k es perfecto y la extensión es finitamente generada) la base S puede elegirse de modo que $K/k(S)$ sea separable.

Definición 9.42 Una extensión K/k está *separablemente generada* si existe una base de trascendencia S de K sobre k tal que la extensión algebraica $K/k(S)$ es separable.

Diremos que K/k es *separable* si cuando $k \subset L \subset K$ con L/k finitamente generada, se cumple que L/k es separablemente generada.

Las propiedades básicas de las extensiones separables se deducen de las caracterizaciones que proporciona el teorema siguiente:

Teorema 9.43 *Sea K/k una extensión de cuerpos de característica prima p. Las afirmaciones siguientes son equivalentes:*

1. *K/k es separable.*
2. *K y k^{1/p^∞} son linealmente disjuntos.*
3. *K y k^{1/p^n} son linealmente disjuntos, para cierto n.*

DEMOSTRACIÓN: 1) $\Rightarrow$ 2) Sean $\alpha_1, \dots, \alpha_n \in K$ linealmente independientes sobre k. Tenemos que probar que también lo son sobre k^{1/p^∞}, para lo cual, llamando $K' = k(\alpha_1, \dots, \alpha_n)$, basta probar que K' y k^{1/p^∞} son linealmente disjuntos. Equivalentemente, cambiando K por K', podemos suponer que K/k es finitamente generada, luego separablemente generada. Sea $S \subset K$ una base de trascendencia tal que $K/k(S)$ sea separable.

Observemos que S es algebraicamente independiente sobre k^{1/p^∞}, pues si $F(s_1, \dots, s_n) = 0$, donde $F \in k^{1/p^\infty}[X_1, \dots, X_n]$ no es nulo, existe un n tal que todos los coeficientes de F elevados a p^n están en k, luego si G es el polinomio que resulta de elevar a p^n los coeficientes y las variables de F, resulta que $G \in k[X_1, \dots, X_n]$ no es nulo y $G(s_1, \dots, s_n) = 0$, contradicción.

Por 9.34 tenemos entonces que $k(S)$ es linealmente disjunto de k^{1/p^∞}. Por otro lado, como $K/k(S)$ es algebraica separable, el teorema 9.41 nos da que K es linealmente disjunto de $k(S)^{1/p^\infty}$, luego en particular de $k^{1/p^\infty}k(S)$, y el teorema 9.33 nos da entonces que K y k^{1/p^∞} son linealmente disjuntos.

2) $\Rightarrow$ 3) es trivial y 3) para un n arbitrario implica 3) para $n = 1$. Suponemos, pues, que K y $k^{1/p}$ son linealmente disjuntos. Sea $k \subset K' \subset K$ de modo que K'/k sea finitamente generada, y tenemos que probar que es separablemente generada. Como K' y $k^{1/p}$ también son linealmente disjuntos, no perdemos generalidad si suponemos que $K = K'$.

Tenemos que probar que K/k tiene una base de trascendencia S tal que $K/k(S)$ es separable, pero vamos a probar algo más fuerte, y es que si fijamos cualquier generador finito $K = k(x_1, \dots, x_n)$, podemos elegir $S \subset \{x_1, \dots, x_n\}$. Sea r el grado de trascendencia de la extensión.

Si $r = n$ tenemos que K/k es separablemente generada con $S = \varnothing$, así que supongamos que $r < n$. Por 9.26, reordenando los generadores podemos suponer que $x_1, \dots, x_r$ es una base de trascendencia de K/k.

Como x_{r+1} es algebraico sobre $k(x_1, \dots, x_r)$, existe un polinomio no nulo $F \in k[X_1, \dots, X_{r+1}]$ tal que $F(x_1, \dots, x_{r+1}) = 0$, y podemos tomarlo de grado mínimo, lo que hace que sea irreducible. Veamos que no puede ser que todas las indeterminadas aparezcan en F con exponente múltiplo de p. Si fuera el caso, podríamos expresar

$$F(X_1, \dots, X_{r+1}) = \sum_i c_i M_i(X_1, \dots, X_{r+1})^p,$$

donde cada M_i es un monomio y $c_i \in k$. Expresando $c_i = a_i^p$, con $a_i \in k^{1/p}$, tendríamos que $\sum_i a_i M_i(x_1, \dots, x_{r+1}) = 0$, con lo que los $M_i(x_1, \dots, x_{r+1}) \in K$ son linealmente dependientes sobre $k^{1/p}$. Sin embargo, son linealmente dependientes sobre k, pues una combinación lineal nula nos daría un polinomio $G \in k[X_1, \dots, X_{r+1}]$ tal que $G(x_1, \dots, x_{r+1}) = 0$ y con grado p unidades menor que el de F, luego todos los coeficientes de la combinación lineal tienen que ser nulos. Esto contradice que K sea linealmente disjunto sobre $k^{1/p}$.

Supongamos, pues, que X_1 aparece en F, pero no siempre elevado a un múltiplo de p. Entonces $F(X_1, x_2, \dots, x_{r+1})$ es un polinomio no nulo (si fuera nulo, sus coeficientes serían de la forma $G_i(x_2, \dots, x_{r+1}) = 0$, para ciertos polinomios $G_i \in k[X_2, \dots, X_{r+1}]$ no nulos que contradirían la minimalidad del grado de F). Además dicho polinomio es irreducible, anula a x_1 y su derivado es no nulo, luego x_1 es separable sobre $k(x_2, \dots, x_{r+1})$, luego sobre $k(x_2, \dots, x_n)$. Si $x_2, \dots, x_n$ es una base de trascendencia, ya tenemos la conclusión. En caso contrario, repetimos el proceso con $k(x_2, \dots, x_n)$ para encontrar un generador, digamos x_2, tal que x_2 es separable sobre $k(x_3, \dots, x_n)$. Tras un número finito de pasos llegamos a que $x_1, \dots, x_{n-r}$ son separables sobre $k(x_{n-r+1}, \dots, x_n)$ y $S = \{x_{n-r+1}, \dots, x_n\}$ es la base de trascendencia buscada. ∎

Destacamos la información adicional que hemos obtenido en la prueba precedente (notemos que es trivialmente cierto para cuerpos de característica 0):

Teorema 9.44 *Si K/k es una extensión separable finitamente generada, todo generador finito contiene una base de trascendencia S de manera que la extensión $K/k(S)$ es separable.*

Notemos que el teorema 9.43 implica también que si k es un cuerpo perfecto (en particular si es algebraicamente cerrado) toda extensión K/k es separable, pues esto es trivial para cuerpos de característica 0 y, si la característica es p, entonces $k = k^{1/p}$ y K y k son linealmente disjuntos sobre k.

Ahora es fácil probar:

Teorema 9.45 *Si $k \subset K \subset L$ son extensiones de cuerpos y K/k y L/K son separables, también lo es L/k.*

Demostración: Tenemos que K y $k^{1/p}$ son linealmente disjuntos, al igual que L y $K^{1/p}$, luego también L y $Kk^{1/p}$, luego L y $k^{1/p}$ son linealmente disjuntos por el teorema 9.33. ∎

Teorema 9.46 *Sea Ω/k una extensión de cuerpos y sean K y L dos cuerpos intermedios libres sobre k. Si la extensión K/k es separable, también lo es la extensión KL/L.*

Demostración: Consideremos un cuerpo $L \subset F \subset KL$ tal que la extensión F/L sea finitamente generada. Todo generador de F se expresa en términos de un número finito de elementos de K y de L, luego podemos tomar $k \subset F' \subset K$ de

modo que F'/k es finitamente generada y $L \subset F \subset LF'$. Como F'/k también es separable y F' y L son libres, no perdemos generalidad si suponemos que K/k es finitamente generada. Sea S una base de trascendencia de K/k tal que $K/k(S)$ sea separable.

Como los elementos de K son separables sobre $k(S)$, también lo son sobre $L(S)$, luego la extensión $KL/L(S)$ es algebraica y separable, y por hipótesis S es algebraicamente independiente sobre K, luego es una base de trascendencia y la extensión KL/L está separablemente generada, y lo mismo vale para F/L. ∎

De los dos teoremas anteriores se sigue inmediatamente:

Teorema 9.47 *Sea Ω/k una extensión de cuerpos y sean K y L dos cuerpos intermedios libres sobre k. Si las extensiones K/k y L/k son separables, también lo es la extensión KL/k.*

Terminamos demostrando un teorema fundamental para la geometría algebraica. Es una consecuencia sencilla del teorema siguiente:

Teorema 9.48 *Sea k un cuerpo algebraicamente cerrado y $F_1, \ldots, F_m$ un conjunto finito de polinomios de $k[x_1, \ldots, x_n]$. Si el sistema de ecuaciones $F_i = 0$ tiene solución en una extensión de k, entonces tiene solución en k.*

Demostración: Sea K la extensión donde el sistema tiene solución. No perdemos generalidad si suponemos que es finitamente generada. Por el teorema 9.43 sabemos que K/k es separable y, usando el teorema del elemento primitivo, tenemos que $K = k(t_1, \ldots, t_r, \alpha)$, donde $t_1, \ldots, t_r$ es una base de trascendencia de K/k. Sea $p(t_1, \ldots, t_r, x) \in k(t_1, \ldots, t_r)[x]$ el polinomio mínimo de α. Sea $F_i(\xi_1, \ldots, \xi_n) = 0$, con $\xi_j \in K$. Cada elemento ξ_j será de la forma $\xi_j = c_j(t_1, \ldots, t_r, \alpha)$, con $c_j(t_1, \ldots, t_r, x) \in k(t_1, \ldots, t_r)[x]$. Entonces

$$F_i(c_1(t_1, \ldots, t_r, x), \ldots, c_n(t_1, \ldots, t_r, x)) = p(t_1, \ldots, t_r, x)q_i(t_1, \ldots, t_r, x),$$

para un cierto polinomio q_i.

Tomemos $(\alpha_1, \ldots, \alpha_r) \in k^r$ de modo que no anule al denominador de ningún coeficiente de p, q_i, $c_1, \ldots, c_n$ ni al coeficiente director de p (es decir, que no anule a su producto, lo cual es posible porque un polinomio que se anule en todo punto ha de ser nulo). Tomemos $\beta \in k$ tal que $p(\alpha_1, \ldots, \alpha_r, \beta) = 0$ y definamos $\lambda_i = c_i(\alpha_1, \ldots, \alpha_r, \tau)$. Es claro entonces que $(\lambda_1, \ldots, \lambda_n)$ es una solución del sistema de ecuaciones. ∎

Hay una razón obvia por la que un sistema de ecuaciones polinómicas

$$F_1(x_1, \ldots, x_n) = 0, \ldots, F_m(x_1, \ldots, x_n) = 0$$

con coeficientes en un cuerpo k pueda no tener solución en k^n, y es que se cumpla $(F_1, \ldots, F_n) = 1$, es decir, que 1 pueda expresarse como combinación lineal de los polinomios F_i. En tal caso, una solución del sistema llevaría a la identidad $0 = 1$. El teorema de los ceros de Hilbert afirma que, si el cuerpo k es algebraicamente cerrado, este caso trivial es el único caso en el que un sistema de ecuaciones puede no tener solución:

Teorema 9.49 (Teorema de los ceros de Hilbert) *Sea k un cuerpo algebraicamente cerrado y $F_1, \ldots, F_m \in k[x_1, \ldots, x_n]$. Si $(F_1, \ldots, F_m) \neq 1$, entonces el sistema de ecuaciones $F_i = 0$ tiene solución en k.*

DEMOSTRACIÓN: Sea M un ideal maximal[4] que contenga a $(F_1, \ldots, F_m)$ y sea $K = k[x_1, \ldots, x_n]/M$. Es claro que las clases de las indeterminadas son una solución del sistema en K (visto como extensión de k), luego por el teorema anterior existe una solución en k. ∎

9.7 Extensiones regulares

Finalmente introducimos una clase de extensiones de cuerpos de interés en geometría algebraica cuyo estudio se basa esencialmente en las técnicas que acabamos de presentar.

Teorema 9.50 *Dados cuerpos $k \subset K \subset \Omega$, con Ω algebraicamente cerrado, las afirmaciones siguientes son equivalentes:*

1. *k es algebraicamente cerrado en K y la extensión K/k es separable.*
2. *K y $\bar{k}$ son linealmente disjuntos sobre k, donde $\bar{k} \subset \Omega$ es la clausura algebraica de k.*

DEMOSTRACIÓN: Si se cumple 1), para probar 2) podemos suponer que K/k es finitamente generada, y basta probar que K es linealmente disjunto de todas las extensiones finitas de k. Si L/k es una extensión finita separable, por el teorema del elemento primitivo es $L = k(\alpha)$, y basta aplicar el teorema 9.37. En general, sea L_s la clausura separable de k en L. Acabamos de probar que K y L_s son linealmente disjuntos sobre k, luego por el teorema 9.33 basta probar que KL_s y L son linealmente disjuntos sobre L_s.

Sea S una base de trascendencia de K/k tal que $K/k(S)$ sea separable. Como K y $\bar{k}$ son linealmente disjuntos sobre k, son libres sobre k, luego S es algebraicamente independiente sobre $\bar{k}$, luego también sobre L_s. Además, KL_s es algebraica y separable sobre $L_s(S)$, pues los elementos de K son algebraicos y separables sobre $k(S)$, luego también sobre $L_s(S)$.

Como KL_s/L_s es separable, tenemos que KL_s y L_s^{1/p^∞} son linealmente disjuntos sobre L_s (por 9.43), luego lo mismo vale para KL_s y L, ya que, como L/L_s es puramente inseparable, se cumple que $L \subset L_s^{1/p^\infty}$ (por 9.15).

Si se cumple 2), en particular K es linealmente disjunto de $k^{1/p}$, luego la extensión K/k es separable, y además $K \cap \bar{k} = k$, lo que significa que k es algebraicamente cerrado en K. ∎

Definición 9.51 Diremos que una extensión de cuerpos K/k es *regular* si cumple las condiciones del teorema anterior.

[4]Esto no requiere el axioma de elección porque el anillo $k[x_1, \ldots, x_n]$ es noetheriano.

Observemos que si k es algebraicamente cerrado, todas sus extensiones son regulares. El teorema siguiente se consecuencia del hecho análogo 9.45 para extensiones separables:

Teorema 9.52 *Si $k \subset K \subset L$ son extensiones de cuerpos y K/k y L/K son regulares, también lo es L/k.*

En cambio, el resultado análogo a 9.46 no es trivial. Para probarlo necesitaremos un refinamiento del teorema del elemento primitivo:

Teorema 9.53 *Si $K = k(\alpha, \beta)$ es una extensión finita de cuerpos y α es separable sobre k, entonces K/k es simple.*

DEMOSTRACIÓN: Sea M una extensión de K en la que se escindan los polinomios mínimos de α y β. Claramente podemos suponer que k es infinito, con lo que podemos tomar $\gamma \in k$ tal que

$$\gamma \neq \frac{\beta' - \beta}{\alpha' - \alpha},$$

para todo par β' y α' de k-conjugados de β y α respectivamente ($\alpha' \neq \alpha$). Veamos que $\sigma = \gamma\alpha + \beta$ es un elemento primitivo. Ciertamente $E = k(\sigma) \subset K$. Sea $g(X) = \operatorname{pol}\operatorname{mín}(\beta, k)$ y sea $h(X) = g(\sigma - \gamma X) \in E[X]$. Claramente

$$h(\alpha) = g(\sigma - \gamma\alpha) = g(\beta) = 0.$$

Más aún, si $\alpha' \neq \alpha$ es un k-conjugado de α, tenemos que

$$\sigma - \gamma\alpha' = \gamma\alpha + \beta - \gamma\alpha' = \gamma(\alpha - \alpha') + \beta \neq \beta'$$

para todo k-conjugado β' de β.

Por consiguiente $h(\alpha') \neq 0$. Así, si $f = \operatorname{pol}\operatorname{mín}(\alpha, k)$ tenemos que α es la única raíz común de f y h en M. Como g se escinde en $M[X]$, lo mismo le sucede a h, luego $X - \alpha$ es el máximo común divisor de f y h en $M[X]$. Ahora bien, el máximo común divisor en $E[X]$ de ambos polinomios debe dividir a $X - \alpha$ en $M[X]$, luego $\alpha \in E$ y $\beta = \sigma - \gamma\alpha \in E$. Esto nos permite concluir que $K = E = k(\sigma)$. ∎

Observemos ahora que si K/k es una extensión finitamente generada, podemos descomponerla en una torre finita $k = K_0 \subset K_1 \subset K_n = K$ de modo que $K_i = K_{i-1}(x_i)$, para cierto $x_i \in K_i$. En efecto, basta expresar $K = k(x_1, \ldots, x_n)$ y definir $K_i = k(x_1, \ldots, x_i)$.

Más aún, podemos exigir que cada x_i sea trascendente, separable o (si los cuerpos tienen característica prima p) puramente inseparable de grado p.

En efecto, si x_i es algebraico sobre K_{i-1}, podemos descomponer

$$K_{i-1} \subset (K_i)_s \subset (K_i)_s(x_i)$$

y, usando que $(K_i)_s = K_{i-1}(x_i')$ por el teorema del elemento primitivo, tenemos que x_i' es separable sobre K_{i-1} y que x_i es puramente inseparable sobre $(K_i)_s$, luego podemos exigir que cada x_i sea trascendente, separable o puramente inseparable sobre el cuerpo precedente.

A su vez, si x_i es puramente inseparable sobre K_{i-1}, el teorema 9.15 nos da que existe un n tal que $x_i^{p^n} \in K_{i-1}$, y entonces podemos intercalar las extensiones

$$K_{i-1} \subset K_{i-1}(x_i^{p^{n-1}}) \subset K_{i-1}(x_i^{p^{n-2}}) \subset \cdots \subset K_{i-1}(x_i^p) \subset K_{i-1}(x_i) = K_i,$$

de modo que cada extensión intermedia es puramente inseparable de grado p.

Teorema 9.54 *Sea Ω/k una extensión de cuerpos y sean K y L dos cuerpos intermedios linealmente disjuntos sobre k. Si la extensión K/k es regular, también lo es la extensión KL/L.*

Demostración: El teorema 9.46 nos da que KL/L es separable. Sólo hay que probar que L es algebraicamente cerrado en KL. Tomamos $\alpha \in KL$ algebraico sobre L. Entonces α se expresa en términos de un número finito de elementos de K y de L. Si llamamos $L' \subset L$ a la adjunción a k de los elementos de L, tenemos que L'/k es finitamente generada y $\alpha \in KL'$. Si probamos que $\alpha \in K'$, tendremos en particular que $\alpha \in L$. Equivalentemente, podemos suponer que la extensión L/k es finitamente generada.

Por la observación previa al enunciado, podemos descomponer L/k en una cadena finita de extensiones simples $k = L_0 \subset L_1 \subset \cdots \subset L_r = L$ de modo que $L_i = L_{i-1}(x_i)$ con x_i trascendente, separable o puramente inseparable de grado igual a la característica p de los cuerpos.

Vamos a probar que cada extensión $K(x_1, \ldots, x_r)/k(x_1, \ldots, x_r)$ es regular. Equivalentemente, podemos suponer que $L = k(x)$, con x trascendente, separable o puramente inseparable de grado p sobre k.

Si es trascendente y el resultado es falso, existe un $\alpha \in K(x)\backslash k(x)$ algebraico sobre $k(x)$. Dicho α será de la forma $\alpha = f/g$, con f, $g \in K[x]$ primos entre sí, y satisfará una ecuación irreducible de la forma

$$a_n(x)(f/g)^n + a_{n-1}(x)(f/g)^{n-1} + \cdots + a_1(x)(f/g) + a_0(x) = 0,$$

con $a_i(x) \in k[x]$. La irreducibilidad implica que $a_0(x) \neq 0$. Cambiando α por α^{-1} si es preciso, podemos suponer que $f \notin k[x]$, y a su vez, existe un factor irreducible f_1 de f en $k[x]$ tal que $f_1 \notin k[x]$. A su vez, alguna raíz α de f_1 no es algebraica sobre k, pues si todas las raíces lo fueran, teniendo en cuenta que $f_1 = (x - \alpha_1) \cdots (x - \alpha_r)$, sus coeficientes serían elementos de K algebraicos sobre k, luego estarían en k. Por consiguiente, $a_0(\alpha) \neq 0$, $g(\alpha) \neq 0$, pero al sustituir x por α en la ecuación, obtenemos $a_0(\alpha) = 0$, contradicción.

Supongamos ahora que x es algebraico sobre k. Sea $y \in K(x)$ algebraico sobre $k(x)$. Tenemos que probar que $y \in k(x)$. Vamos a ver primero que existe un z tal que $k(x, y) = k(z)$. Si x es separable sobre k, basta aplicar el teorema 9.53. La otra alternativa que estamos considerando es que x sea puramente inseparable de grado p.

Sea $S \subset K$ una base de trascendencia tal que $K/k(S)$ sea separable. Entonces $K(x)/k(S \cup \{x\})$ es una extensión algebraica separable, luego y es separable sobre $k(S \cup \{x\}) = k(x)(S)$.

Ahora bien, como K y $k(x)$ son libres sobre k, tenemos que S es algebraicamente independiente sobre $k(x)$, luego $k(x)$ es algebraicamente cerrado en $k(x)(S)$ (teorema 9.36). El polinomio mínimo $p(X)$ de y sobre $k(x)(S)$ es de la forma $(x - \alpha_1) \cdots (x - \alpha_r)$, donde los α_i son algebraicos sobre $k(x)$, luego los coeficientes de $p(X)$ son algebraicos sobre $k(x)$ y están en $k(x)(S)$, luego están en $k(x)$. En otras palabras, el polinomio mínimo de y sobre $k(x)(S)$ es el mismo que el polinomio mínimo de y sobre $k(x)$, luego y es separable sobre $k(x)$.

Ahora bien, si $|k(x, y) : k| = n$, considerando la cadena $k \subset k(x) \subset k(x, y)$, vemos que el grado de separabilidad de la extensión es n/p, mientras que la cadena $k \subset k(y) \subset k(x, y)$ nos da que, o bien $k(y) = k(x, y)$, con lo que ya tenemos que la extensión es simple, o bien $k(x, y)/k(y)$ es puramente inseparable de grado p, por lo que el grado de separabilidad es $|k(y) : k|$, luego y es separable sobre k, y de nuevo podemos aplicar el teorema 9.53.

Ahora, $K(x) = Kk(x, y) = Kk(z) = K(z)$, aplicando dos veces el teorema 9.37 obtenemos que

$$|k(z) : k| = |K(z) : K| = |K(x) : K| = |k(x) : k|,$$

pero $k(x) \subset k(z)$, luego tenemos la igualdad $k(x) = k(z) = k(x, y)$ y así concluimos que $y \in k(x)$. ∎

A partir de aquí el análogo a 9.52 ya es inmediato:

Teorema 9.55 *Sea Ω/k una extensión de cuerpos y sean K y L dos cuerpos intermedios linealmente disjuntos sobre k. Si las extensiones K/k y L/k son regulares, también lo es la extensión KL/k.*

Para terminar presentamos una versión débil de la regularidad que equivale a ésta para extensiones de cuerpos perfectos:

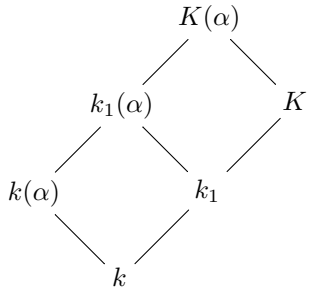
Teorema 9.56 *Consideremos cuerpos $k \subset K \subset \Omega$, con Ω algebraicamente cerrado. Sea k_s la clausura separable de k en Ω y k_1 la clausura algebraica de k en K. Las afirmaciones siguientes son equivalentes:*

1. *La extensión k_1/k es puramente inseparable.*

2. *K y k_s son linealmente disjuntos sobre k.*

Demostración: Si suponemos 1), para probar 2) basta probar que si tenemos $k \subset k' \subset k_s$ con k'/k finita, entonces k' es linealmente disjunto de K sobre k. Por el teorema del elemento primitivo es $k' = k(\alpha)$. Aplicamos el teorema 9.33 a las extensiones indicadas en el esquema de la página siguiente, teniendo en cuenta que $k_1(\alpha)$ y K son linealmente disjuntos sobre k_1 por 9.37

y $k(\alpha)$ y k_1 son linealmente disjuntos sobre k_1, mientras que $k(\alpha)$ y k_1 son linealmente disjuntos por 9.41, ya que, si k tiene característica prima p, entonces $k_1 \subset k_1^{1/p^\infty}$ por 9.15.



Esto prueba que K y $k' = k(\alpha)$ son linealmente disjuntos sobre k. ∎

Definición 9.57 Las extensiones de cuerpos que cumplen las condiciones del teorema anterior se llaman *extensiones primarias.*

Obviamente, toda extensión de un cuerpo perfecto es primaria. También es fácil ver que si tenemos extensiones $k \subset L \subset K$, entonces, si L/k y K/L son primarias, también lo es K/k, así como que si lo es K/k, también lo es L/k.

Apéndice A

El axioma de elección

En la sección 1.2 hemos presentado una teoría axiomática de conjuntos, que no es sino la *teoría de Zermelo,* determinada por seis axiomas:

Extensionalidad *Dos conjuntos son iguales si y sólo si tienen los mismos elementos.*

Especificación *Dado un conjunto A y una propiedad P, existe un conjunto cuyos elementos son los elementos de A que cumplen P.*

Par *Dados dos conjuntos A y B, existe otro conjunto C cuyos elementos son exactamente A y B.*

Unión *Dado un conjunto A, existe otro conjunto B cuyos elementos son exactamente los que pertenecen a alguno de los elementos de A.*

Partes *Dado un conjunto A, existe otro conjunto cuyos elementos son exactamente todos los subconjuntos de A.*

Infinitud *Existe un conjunto X con una aplicación $S : X \longrightarrow X$ inyectiva y no suprayectiva.*

Hemos incluido la versión abstracta del axioma de infinitud presentada en la sección 1.4, aunque en la práctica podemos optar sin ningún cambio esencial por la versión concreta dada en la sección 1.2 o por cualquier otra alternativa que implique la versión abstracta.

Estos axiomas son suficientes para demostrar la mayor parte de los resultados de este libro, pero en algunos casos es necesario un axioma más:

Axioma de elección (AE) *Para todo conjunto X, existe $F : X \longrightarrow \bigcup X$ tal que para todo conjunto $A \in X$ que no sea $\varnothing$ se cumple que $F(A) \in A$.*

Una función F en las condiciones indicadas por el axioma se llama *función de elección* para X, pues elige un elemento de cada elemento de X. Para algunos

conjuntos X es posible demostrar que existen funciones de elección (por ejemplo, en el caso de los conjuntos finitos), pero para otros puede no haber ningún criterio que determine una, ni por especificación ni empleando ninguno de los resultados sobre existencia de conjuntos deducibles de los demás axiomas, y entonces el axioma de elección resulta imprescindible.

Es fácil ver que los dos teoremas siguientes son, de hecho, equivalentes al axioma de elección:

Teorema A.1 (AE) *Si $\{X_i\}_{i\in I}$ es una familia de conjuntos no vacíos, entonces $\prod_{i\in I} X_i \neq \varnothing$.*

DEMOSTRACIÓN: Sea $A = \{X_i \mid i \in I\}$ y sea $F : A \longrightarrow \bigcup A$ una función de elección. Así, para cada $i \in I$, como $X_i \neq \varnothing$, se cumple que $F(X_i) \in X_i$. Sea ahora $x : I \longrightarrow \bigcup A$ la función dada por $x(i) = F(X_i)$. Es claro entonces que $x \in \prod_{i\in I} X_i$. ∎

Teorema A.2 (AE) *Si A es una familia de conjuntos no vacíos disjuntos dos a dos, existe un conjunto que tiene exactamente un elemento de cada elemento de A.*

DEMOSTRACIÓN: Sea $F : A \longrightarrow \bigcup A$ una función de elección en A y sea $B = F[A]$. Si $X \in A$, como es no vacío, sabemos que $F(X) \in X \cap B$. Si $u \in X \cap B$, entonces $u = F(Y)$, para cierto $Y \in A$, luego $u \in Y \cap X$, luego $Y = X$, porque los elementos de A son disjuntos dos a dos, luego $u = F(X)$, que es, por consiguiente, el único elemento de $X \cap B$. ∎

Otra consecuencia destacada es la siguiente:

Teorema A.3 (Principio de elecciones dependientes) *Si $A \neq \varnothing$ es un conjunto y R es una relación en A con la propiedad de que para todo $a \in A$ existe un $b \in A$ tal que $a\,R\,b$, entonces existe $f : \mathbb{N} \longrightarrow A$ tal que, para todo $n \in \mathbb{N}$, se cumple $f(n)\,R\,f(n+1)$.*

DEMOSTRACIÓN: Consideramos la familia $\{X_a\}_{a\in A}$ formada por los conjuntos $X_a = \{b \in A \mid a\,R\,b\}$. Por hipótesis todos los conjuntos X_a son no vacíos, luego existe $g \in \prod_{a\in A} X_a$, que es una función $g : A \longrightarrow A$ tal que, para todo $a \in A$, se cumple que $g(a) \in X_a$, es decir, $a\,R\,g(a)$. Ahora basta aplicar el principio de recursión, que nos da una función $f : \mathbb{N} \longrightarrow A$ tal que $f(0)$ es cualquier elemento prefijado de A y para todo $n \in \mathbb{N}$ se cumple que $f(n+1) = g(f(n))$, luego en particular $f(n)\,R\,f(n+1)$. ∎

El nombre de "principio de elecciones dependientes" se debe a que aquí no tenemos una familia de conjuntos $\{X_n\}_{n\in\mathbb{N}}$ dada de antemano y elegimos un $f(n) \in X_n$ para cada n, sino que cada $f(n+1)$ se elige entre los posteriores a $f(n)$ según la relación R, y así, el conjunto de los valores admisibles entre los que podemos elegir $f(n+1)$ depende de las elecciones precedentes que nos han llevado hasta $f(n)$.

El principio de elecciones dependientes nos permite hacer en particular elecciones "independientes" siempre y cuando sean sobre una familia numerable, en el sentido siguiente:

Teorema A.4 (Axioma de elección numerable) *Si $\{X_n\}_{n\in\mathbb{N}}$ es una familia de conjuntos no vacíos, entonces $\prod_{n\in\mathbb{N}} X_n \neq \varnothing$.*

DEMOSTRACIÓN: Sea $X = \bigcup_{n\in\mathbb{N}} X_n$, recordemos que $I_m^* = \{0,\dots,m-1\}$, y sea A el conjunto[1] de todas las aplicaciones $s : I_m^* \longrightarrow X$ tales que, para todo $n < m$, se cumple que $s(n) \in X_n$. Sea R la relación en A según la cual $s\,R\,t$ si y sólo si $s \subsetneq t$.

Se cumple que, $A \neq \varnothing$, pues si tomamos $x \in X_0$, la función $s : I_0^* \longrightarrow X$ dada por $s(0) = x$ cumple $s \in A$. Además, para todo $s \in A$, existe un $t \in A$ tal que $s\,R\,t$. En efecto, tenemos que $s : I_m^* \longrightarrow X$ y por hipótesis existe $y \in X_{m+1}$, luego podemos extender s a una aplicación $s' : I_{m+1}^* \longrightarrow X$ mediante $s'(m+1) = y$, y claramente $s' \in A$ cumple $s\,R\,s'$.

Podemos aplicar el principio de elecciones dependientes, que nos da una función $f : \mathbb{N} \longrightarrow A$ tal que $f(n) \subsetneq f(n+1)$, para todo $n \in \mathbb{N}$. Podemos llamar $s_n = f(n)$, que es una función definida sobre cierto I_m^*. Ahora bien, una simple inducción demuestra que $I_n^* \subset \mathcal{D}s_n$, así como que si $m \leq n$, entonces $s_m \subset s_n$. Esto nos permite definir $g : \mathbb{N} \longrightarrow X$ estableciendo que $g(n) = s_{n+1}(n)$. Es claro entonces que $g(n) \in X_n$, luego $g \in \prod_{n\in\mathbb{N}} X_n$. ∎

Notemos que el *axioma de elección finito,* es decir, el teorema anterior para familias de la forma $\{X_n\}_{n=1}^m$, se demuestra por inducción sobre m sin necesidad del axioma de elección y no es, por tanto, un axioma, sino un teorema.

Todas las aplicaciones "fuertes" del axioma de elección que vamos a necesitar (es decir, sin contar aquellas que sólo requieran el Principio de elecciones dependientes) las haremos a través de una consecuencia, bastante técnica, pero muy práctica. Para enunciarlo tenemos que introducir algunos conceptos.

Sea $(A, \leq)$ un conjunto parcialmente ordenado, es decir, A es un conjunto y $\leq$ es una relación de orden parcial en A.

Una *cadena* en A es un subconjunto $c \subset A$ que está totalmente ordenado, es decir, para todo x, $y \in c$ se cumple $x \leq y$ o bien $y \leq x$.

Un $m \in A$ es *maximal* si no existe ningún otro $a \in A$ tal que $m < a$.

Notemos que si A está totalmente ordenado, maximal es lo mismo que máximo, pero en general no es así: si A tiene máximo, es su único elemento maximal, pero puede haber distintos maximales que no sean máximos.

Se dice que $(A, \leq)$ está *bien ordenado,* o que $\leq$ es un *buen orden* sobre A si A está totalmente ordenado y todo subconjunto de A no vacío tiene un mínimo elemento.

[1] Notemos que A puede definirse por especificación a partir de $\mathcal{P}(\mathbb{N} \times X)$.

Teorema A.5 (Lema de Zorn) (AE) *Todo conjunto parcialmente ordenado en el que toda cadena tiene una cota superior, tiene un elemento maximal.*

DEMOSTRACIÓN: Sea $(A, \leq)$ un conjunto parcialmente ordenado que cumpla las hipótesis. Sea $C \subset \mathcal{P}A$ el conjunto de todas las cadenas en A. Sabemos que toda cadena $c \in C$ tiene cota superior, pero pueden darse dos casos si existe una cota superior $u \in A \setminus c$, es decir, una cota superior estricta, que es mayor que todos los elementos de c, entonces c puede prolongarse a una cadena $c \cup \{u\}$. En tal caso, llamamos $X_c \subset \mathcal{P}A$ al conjunto de todas las prolongaciones de c de la forma $c \cup \{u\}$.

La alternativa es que c no tenga cotas superiores estrictas. Entonces una cota superior u de c es un maximal de A, pues si existiera $v \in A$ tal que $u < v$, tendríamos que v es una cota superior estricta de c. Nuestro objetivo será demostrar que existe una cadena c sin cotas superiores estrictas. En cualquier caso, si c es una cadena en estas condiciones, llamamos $X_c = \{c\}$.

Con esto tenemos definida una aplicación $X : C \longrightarrow \mathcal{P}\mathcal{P}A$ o, equivalentemente, una familia $\{X_c\}_{c\in C}$ de conjuntos no vacíos. Por el teorema A.1 podemos tomar $G \in \prod_{c\in C} X_c$. Equivalentemente, $G : C \longrightarrow \mathcal{P}A$, de modo que, para toda cadena $c \in C$, se cumple que $G(c) \in X_c$.

Basta probar que existe una cadena c tal que $G(c) = c$, pues esto significa que c no tiene cotas estrictas (ya que en tal caso $c \notin X_c$, por construcción), y por lo tanto A tiene un elemento maximal.

Si no se cumple $G(c) = c$, entonces $G(c)$ es una cadena que resulta de añadir un único elemento por encima de todos los elementos de c.

Diremos que c es una *buena cadena* en $(A, \leq)$ si c está bien ordenada por $\leq$ y, para cada $a \in c$, se cumple que

$$G(\{u \in c \mid u < a\}) = \{u \in c \mid u \leq a\}.$$

Notemos que $\{u \in c \mid u < a\}$ es trivialmente una cadena, y lo que estamos pidiendo es que G le añada un elemento, y que éste sea precisamente a. Es fácil comprobar lo siguiente:

- $\varnothing$ es (trivialmente) una buena cadena.
- Si c es una buena cadena y $a \in c$, entonces los conjuntos $\{u \in c \mid u < a\}$ y $\{u \in c \mid u \leq a\}$ son también buenas cadenas.
- Si c es una buena cadena, $G(c)$ también lo es.

Ahora demostraremos este hecho:

Si c y c' son buenas cadenas y $a \in c$ cumple que $\{u \in c \mid u < a\} = \{u \in c' \mid u < a\}$, entonces $a \in c'$ o bien $c' = \{u \in c \mid u < a\}$.

En efecto, si $c' \neq \{u \in c \mid u < a\} = \{u \in c' \mid u < a\}$, como por hipótesis se da la inclusión $\supset$, tiene que haber un $a' \in c'$ que no esté en $\{u \in c' \mid u < a\}$, para lo cual a' tiene que ser una cota superior estricta de este conjunto. Como c' está bien ordenado, podemos suponer que a' es la mínima cota superior estricta (en c') de dicho conjunto. Por lo tanto:

$$\{u \in c' \mid u < a'\} = \{u \in c' \mid u < a\} = \{u \in c \mid u < a\}.$$

Como el primer y el tercer conjunto son buenas cadenas, al aplicar G obtenemos un conjunto cuyo máximo tiene que ser tanto a' como a, luego $a = a' \in c'$.

Veamos ahora que dos buenas cadenas c y c' están necesariamente contenidas una en la otra. En efecto, si c no está contenida en c', podemos tomar el mínimo $a \in c \setminus c'$, de modo que $\{u \in c \mid u < a\} \subset \{u \in c' \mid u < a\}$. Veamos que tiene que darse la igualdad.

En caso contrario, existe un $a' \in c'$ tal que $a' < a$, pero $a' \notin c$. Podemos tomar el mínimo posible. Así, $\{u \in c \mid u < a'\} = \{u \in c' \mid u < a'\}$, luego, según hemos visto, esto implica que $c = \{u \in c' \mid u < a'\}$, pero entonces $a < a'$, contradicción.

Por consiguiente, $\{u \in c \mid u < a\} = \{u \in c' \mid u < a\}$ y, aplicando de nuevo la propiedad que hemos demostrado, $c' \subset c$.

Más aún, si c y c' son dos buenas cadenas distintas, $c' \subset c$ y llamamos $a = \text{mín}(c \setminus c')$, entonces $c' = \{u \in c \mid u < a\}$, pues obviamente se cumple $a \notin c'$ y $\{u \in c \mid u < a\} = \{u \in c' \mid u < a\}$.

Ahora es fácil ver que la unión c_0 de todas las buenas cadenas es una buena cadena. En efecto, es una cadena, porque si u, $v \in c_0$, existen buenas cadenas c y c' tales que $u \in c$ y $v \in c'$, pero como $c \subset c'$ o viceversa, tenemos que u y v están en una misma cadena, luego $u \leq v$ o $v \leq u$.

Además c_0 está bien ordenada, pues si $X \subset c_0$ no es vacío, un elemento de X estará también en cierta buena cadena c, de modo que $c \cap X \neq \varnothing$. Como c está bien ordenada, existe $m = \text{mín}(X \cap c)$. Vamos a probar que m es el mínimo de X. En caso contrario existe un $x \in X$ tal que $x < m$ (luego $x \notin c$). Dicho x tiene que cumplir $x \in c'$, para cierta buena cadena c' que no puede estar contenida en c, por lo que $c \subset c'$. Sea $a = \text{mín}(c' \setminus c)$, de modo que $c = \{u \in c' \mid u < a\}$, pero entonces $x < m < a$, luego $x \in c$, contradicción.

Por último, si $a \in c_0$, entonces $a \in c$, para cierta buena cadena c, y vamos a ver que $\{u \in c_0 \mid u < a\} = \{u \in c \mid u < a\}$. Esto implicará que

$$G(\{u \in c_0 \mid u < a\}) = \{u \in c \mid u \leq a\} = \{u \in c_0 \mid u \leq a\},$$

lo que terminará la prueba de que c_0 es una buena cadena. Si $u \in c_0$ cumple $u < a$, pero $u \notin c$, entonces $u \in c'$, para cierta buena cadena c' que no puede cumplir $c' \subset c$, luego $c \subset c'$. Sea $a' = \text{mín}(c' \setminus c)$, de modo que $a' \leq u$. Entonces $c = \{v \in c' \mid v < a'\}$, luego $a < a' \leq u < a$, contradicción.

Con esto terminamos la prueba sin más que observar que, como $G(c_0)$ también es una buena cadena, tiene que ser $G(c_0) \subset c_0$ (porque c_0 es la unión de todas), y la otra inclusión se da siempre, luego $G(c_0) = c_0$, como había que probar. ∎

Veamos un ejemplo de aplicación del lema de Zorn:

Teorema A.6 (AE) *Todo conjunto puede ser bien ordenado.*

DEMOSTRACIÓN: Sea X un conjunto cualquiera (que podemos suponer no vacío) y sea A el conjunto de todos los pares $(Y, \leq)$, donde $Y \subset X$ y $\leq$ es un buen orden en Y. Definimos en A la relación de orden dada por $(Y, \leq) \preceq (Y', \leq')$ si y sólo si:

1. $Y \subset Y'$.
2. Para todo $u, v \in Y$ se cumple $u \leq v$ si y sólo si $u \leq' v$.
3. Si $u \in Y$, $v \in Y' \setminus Y$, entonces $u <' v$.

Esto significa que Y' mantiene el mismo orden sobre los elementos de Y y los elementos de $Y' \setminus Y$ son todos posteriores a los de Y, es decir, que el orden de Y' prolonga al de Y.

Es claro que $A \neq \varnothing$, pues cualquier subconjunto de X con un elemento se puede ordenar bien de forma trivial. Sea $C \subset A$ una cadena, sea $Y^* = \bigcup_{(Y,\leq)\in C} Y$ y consideremos en Y^* el orden según el cual $u \leq^* v$ si y sólo si existe un $(Y, \leq) \in C$ tal que $u, v \in Y$ y $u \leq v$.

Es fácil comprobar que se trata de una relación de orden en Y^*, y es un buen orden, pues si $B \subset Y^*$ no es vacío, tomamos $u \in B$, y entonces existe un $(Y, \leq) \in C$ tal que $u \in Y$ por lo que $B \cap Y \neq \varnothing$, luego existe $m = \text{mín}(B \cap Y)$, y se cumple que $m = \text{mín}\, B$, pues si $v \in B$, entonces, existe un $(Y', \leq') \in C$ tal que $v \in Y'$. Como C es una cadena, o bien $(Y, \leq) \preceq (Y', \leq')$ o bien $(Y', \leq') \preceq (Y, \leq)$. En el segundo caso se cumple $v \in Y \cap B$, luego $m \leq v$, luego $m \leq^* v$. En el primer caso, distinguiendo si $v \in Y$ o si $v \in Y' \setminus Y$ se llega igualmente a que $m \leq' v$ y de aquí a que $m \leq^* v$.

Por lo tanto $(Y^*, \leq^*) \in A$. Observemos ahora que si $(Y, \leq) \in C$, entonces $(Y, \leq) \preceq (Y^*, \leq^*)$. En efecto, se cumplen obviamente las propiedades 1. y 2. de la definición de $\preceq$, y en cuanto a la tercera, si $u \in Y$, $v \in Y^* \setminus Y$, existe un $(Y', \leq) \in C$ tal que $v \in Y' \setminus Y$. Como C es una cadena tiene que ser $Y \preceq Y'$, luego $u <' v$, luego $u <^* v$.

Esto prueba que $(Y^*, \leq^*)$ es una cota superior de C, luego por el lema de Zorn existe un par $(Y, \leq)$ maximal en A. Pero necesariamente entonces $Y = X$, ya que si fuera $Y \subsetneq X$, podríamos tomar $x \in X \setminus Y$ y definir una relación de orden en $Y \cup \{x\}$ que prolongara a la de Y poniendo a x como máximo elemento. Claramente se trataría de un buen orden, de modo que $(Y, \leq) \prec (Y \cup \{x\}, \leq')$ en contra de la maximalidad de $(Y, \leq)$. Así pues, $\leq$ es un buen orden en X. ∎

Apéndice B

Conjuntos infinitos

La mayor parte de los conjuntos con los que vamos a trabajar son infinitos, y en algunas ocasiones necesitaremos ciertos resultados conjuntistas que muestran que es posible hablar del "número de elementos" de un conjunto infinito, y hacer consideraciones sobre el "tamaño" de un conjunto infinito. Un desarrollo sistemático de la teoría de cardinales infinitos queda fuera del alcance de este libro, pero aquí vamos a presentar los pocos resultados que nos van a ser necesarios.

Conjuntos numerables Vamos a extender parcialmente a conjuntos infinitos el concepto de "contar" los elementos de un conjunto. Recordemos que un conjunto A es finito si existe una biyección $f : I_n \longrightarrow A$, donde n es un número natural e $I_n = \{1, \ldots, n\}$, entendiendo que $I_0 = \varnothing$. En tal caso, el número natural n es único y recibe el nombre de cardinal de A, y se representa por $|A|$. La definición siguiente generaliza estos conceptos:

Definición B.1 Un conjunto A es *numerable* si es finito o bien existe una biyección $f : \mathbb{N} \longrightarrow A$. En tal caso diremos que el *cardinal* de A es $\mathbb{N}$, aunque cuando consideramos a $\mathbb{N}$ como cardinal de los conjuntos numerables es costumbre escribir $\mathbb{N} = \aleph_0$ (áleph 0).

Así pues, si A es un conjunto numerable, su cardinal $|A|$ puede ser un número natural o bien $\aleph_0$. Los conjuntos de cardinal $\aleph_0$ son, pues, los conjuntos *infinitos numerables.* En otros términos, un conjunto A es infinito numerable si sus elementos se pueden organizar como una sucesión infinita

$$a_0, \quad a_1, \quad a_2, \quad a_3, \quad a_4, \quad \ldots$$

sin repeticiones.

Quizá el lector pueda pensar que todos los conjuntos infinitos son numerables, pero no es así:

Teorema B.2 (Cantor) *Si A es un conjunto, no existe una aplicación suprayectiva $f : A \longrightarrow \mathcal{P}A$.*

DEMOSTRACIÓN: Supongamos que $f : A \longrightarrow \mathcal{P}A$ es suprayectiva y consideremos el conjunto $X = \{x \in A \mid x \notin f(x)\} \in \mathcal{P}A$. Por hipótesis existe $a \in A$ tal que $f(a) = X$, pero entonces tenemos una contradicción:

$$a \in X \quad \text{si y sólo si} \quad a \notin f(a) \quad \text{si y sólo si} \quad a \notin X.$$

Así pues, no puede existir tal aplicación f. ∎

En particular $\mathcal{P}\mathbb{N}$ es un ejemplo de conjunto infinito no numerable. A continuación vamos a dar criterios para reconocer qué conjuntos son numerables.

Extendemos la relación de orden en $\mathbb{N}$ para incluir a $\aleph_0$ como máximo elemento, de modo que si A y B son conjuntos numerables, la desigualdad $|A| \leq |B|$ se cumple en particular si $|B| = \aleph_0$.

Evidentemente, todo conjunto biyectable con un conjunto numerable es numerable. Más aún:

Teorema B.3 *Si $f : A \longrightarrow B$ es inyectiva y B es numerable, entonces A es numerable y $|A| \leq |B|$.*

DEMOSTRACIÓN: Sabemos que el resultado es cierto si B es finito, así que podemos suponer que $|B| = \aleph_0$. Entonces, si A es finito la conclusión es trivial, luego podemos suponer que A es infinito, y tenemos que probar que es numerable. Sea $g : \mathbb{N} \longrightarrow B$ biyectiva, de modo que $h = f \circ g^{-1} : A \longrightarrow \mathbb{N}$ inyectiva. Como $h : A \longrightarrow h[A]$ es biyectiva, tenemos que $h[A]$ es infinito, y basta probar que es numerable. Equivalentemente, podemos suponer que $A \subset \mathbb{N}$, es decir, sólo hay que probar que todo subconjunto infinito de $\mathbb{N}$ es numerable.

Definimos recurrentemente una aplicación $k : \mathbb{N} \longrightarrow A$: suponiendo definidos $k(0), \ldots, k(n-1)$, tomamos

$$k(n) = \min(A \setminus \{k(0), \ldots, k(n-1)\}).$$

Notemos que la definición es correcta, porque $\{k(0), \ldots, k(n)\} \subset A$ es un conjunto finito y, como A es infinito, tenemos que $A \setminus \{k(0), \ldots, k(n-1)\} \neq \varnothing$, luego tiene un mínimo elemento, que es el que tomamos como $k(n)$.

En particular, si $m < n$, tenemos que $k(m) \neq k(n)$, pues se cumple que $k(m) \in \{k(0), \ldots, k(n-1)\}$. Esto prueba que k es inyectiva. Veamos que también es suprayectiva. Si existe un $a \in A$ que no tiene antiimagen, podemos tomar el mínimo de ellos. Consideremos $I = \{n \in \mathbb{N} \mid k(n) < a\}$. Entonces $k|_I : I \longrightarrow \{0, \ldots, a-1\}$ biyectiva, luego I es un subconjunto finito de $\mathbb{N}$, y esto implica que tiene un máximo elemento n. Pero entonces tiene que ser $k(n+1) = a$, porque ciertamente a es el mínimo del conjunto $A \setminus \{k(0), \ldots, k(n)\}$, ya que todo número $m < a$ es de la forma $k(i)$ con $i \in I$, luego $i \leq n$, luego $m \in \{k(0), \ldots, k(n)\}$, y así tenemos una contradicción. ∎

En particular, todo subconjunto de un conjunto numerable es numerable. Veamos una variante del teorema anterior:

Teorema B.4 *Si $f : A \longrightarrow B$ es suprayectiva y A es numerable, entonces B es numerable y $|B| \leq |A|$.*

DEMOSTRACIÓN: Sea $g : \mathbb{N} \longrightarrow A$ biyectiva, de modo que $g \circ f : \mathbb{N} \longrightarrow B$ es también suprayectiva. Equivalentemente, podemos suponer que $A = \mathbb{N}$. Pero entonces podemos definir $h : B \longrightarrow \mathbb{N}$ inyectiva mediante $h(a) = \min f^{-1}[a]$, y basta aplicar el teorema anterior. ∎

Otra consecuencia inmediata del teorema B.3 es que un conjunto A es numerable (finito o infinito) si y sólo si existe $f : A \longrightarrow \mathbb{N}$ inyectiva.

A continuación vamos a demostrar que $\mathbb{N} \times \mathbb{N}$ es numerable. La idea subyacente a la prueba se muestra en la tabla siguiente:

⋮						
4	10					
3	6	11				
2	3	7	12			
1	1	4	8	13		
0	0	2	5	9	14	
	0	1	2	3	4	⋯

En ella vamos disponiendo los números naturales completando diagonales: en la primera diagonal ponemos el 0, en la diagonal siguiente el 1 y el 2, en la siguiente el 3, el 4 y el 5, y así sucesivamente. Entonces a cada par ordenado de números naturales le corresponde el número natural que ponemos en su fila y su columna. Por ejemplo, $f(3,1) = 13$. Esto determina una biyección $f : \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N}$. En realidad f admite una definición algebraica muy simple:

$$f(x,y) = \frac{(x+y)(x+y+1)}{2} + x.$$

La idea para llegar a esta expresión también es sencilla: por ejemplo, la imagen $f(3,1) = 13$ está en la diagonal formada por los pares cuyas componentes suman $x + y = 4$. Para llegar a ella ha que pasar antes por las diagonales anteriores, que contienen $1+2+3+4 = 10$ números, pero como empezamos en el 0 resulta que el $f(0,4) = 10$ es ya ya el primero de dicha diagonal. Para llegar a $f(3,1)$ hemos de avanzar $x = 3$ posiciones. En general, el par $f(x,y)$ se alcanza en la posición

$$1 + 2 + \cdots + (x+y) + x = \frac{(x+y)(x+y+1)}{2} + x.$$

Vamos a dar una justificación puramente aritmética de estos hechos. Para ello observamos que la función

$$g(n) = \frac{n(n+1)}{2}$$

cumple $n \leq g(n)$, pues esto equivale a que $2n \leq n^2 + n$, o a que $n \leq n^2$, lo cual se cumple si $n = 0$ y, en caso contrario equivale a $1 \leq n$, que también se cumple.

Por lo tanto, para cada natural z existe un mínimo n' tal que $z < g(n')$. No puede ser $n' = 0$, luego existe un único $n = n' - 1$ tal que $g(n) \leq z < g(n+1)$.

Ahora bien,

$$g(n+1) - g(n) = \frac{(n+1)(n+2) - n(n+1)}{2} = \frac{(n+1)2}{2} = n+1,$$

luego existe un único número natural $0 \leq x \leq n$ tal que $z = g(n)+x$. Llamando $y = n - x$ tenemos que $z = g(x+y) + x = f(x,y)$. Esto prueba que f es suprayectiva.

La inyectividad se debe a que si $f(x,y) = f(x',y')$, entonces, llamando $n = x+y$ y $n' = x'+y'$ tenemos que $g(n) \leq z < g(n+1)$ y $g(n') \leq z < g(n'+1)$, lo cual sólo es posible si $n = n'$, luego

$$f(x,y) = g(n) + x = g(n) + x',$$

de donde $x = x'$ y, por consiguiente, de $x + y = x' + y'$ obtenemos que $y = y'$.

Más en general, ahora podemos probar:

Teorema B.5 *Si A y B son conjuntos numerables, entonces el producto cartesiano $A \times B$ es numerable.*

DEMOSTRACIÓN: Sean $g_1 : A \longrightarrow \mathbb{N}$ y $g_2 : B \longrightarrow \mathbb{N}$ inyectivas. Entonces la función $g : A \times B \longrightarrow \mathbb{N} \times \mathbb{N}$ dada por $g(a,b) = (g_1(a), g_2(b))$ es inyectiva, y su composición con la biyección $\mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N}$ que hemos construido es una aplicación $A \times B \longrightarrow \mathbb{N}$ inyectiva, luego $A \times B$ es numerable. ∎

Es claro que si uno de los dos conjuntos es infinito numerable y el otro es no vacío (finito o infinito), entonces $A \times B$ es infinito numerable. Teniendo en cuenta la relación $|A \times B| = |A||B|$ válida para conjuntos finitos, ahora podemos extenderla a conjuntos numerables si definimos

$$\aleph_0 \cdot 0 = 0, \quad \aleph_0 \cdot n = \aleph_0 \quad (\text{si } n > 0), \quad \aleph_0\aleph_0 = \aleph_0.$$

Teorema B.6 *La unión de un conjunto numerable de conjuntos numerables es numerable.*

DEMOSTRACIÓN: Sea $\{A_n\}_{n\in\mathbb{N}}$ una familia numerable de conjuntos numerables. Podemos suponer que no son todos vacíos, y fijamos un a_0 perteneciente a uno de ellos. Sea[1] $f_n : A_n \longrightarrow \mathbb{N}$ inyectiva y sea $g : \mathbb{N} \longrightarrow \mathbb{N} \times \mathbb{N}$ biyectiva. Definimos como sigue una aplicación $h : \mathbb{N} \longrightarrow \bigcup_{n\in\mathbb{N}} A_n$:

Si $g(m) = (n,i)$ y se cumple que $i \in f_n[A_n]$, entonces $h(m) = f_n^{-1}(i)$, y en caso contrario $h(m) = a_0$. Así la aplicación h es suprayectiva, pues dado $a \in \bigcup_{n\in\mathbb{N}} A_n$, existe un n tal que $a \in A_n$, y podemos tomar $i = f_n(a)$. Si $m = g^{-1}(n,i)$, es claro que $h(m) = f_n^{-1}(i) = a$. Ahora basta aplicar B.4. ∎

[1] Aquí estamos eligiendo una aplicación para cada n, con lo que estamos usando el axioma de elección. No obstante, se trata del axioma de elección numerable. En general no destacaremos los usos de las formas débiles del axioma de elección, como el axioma de elección numerable o el principio de elecciones dependientes.

En particular, la unión de dos conjuntos numerables, uno de ellos infinito, es infinita numerable. Esto puede expresarse extendiendo la suma de cardinales de modo que

$$\aleph_0 + n = \aleph_0, \qquad \aleph_0 + \aleph_0 = \aleph_0.$$

Hemos visto antes que si A es un conjunto infinito numerable, entonces $\mathcal{P}A$ no es numerable. En cambio:

Teorema B.7 *Si X es un conjunto numerable, el conjunto $\mathcal{P}^f X$ formado por todos los subconjuntos finitos de X es numerable.*

DEMOSTRACIÓN: Si X es finito, entonces $\mathcal{P}^f X$ es finito, luego la conclusión es trivial. Supongamos, pues, que X es infinito. Si $f : X \longrightarrow \mathbb{N}$ es una biyección, podemos definir $g : \mathcal{P}^f X \longrightarrow \mathcal{P}^f\mathbb{N}$ mediante $g(A) = f[A]$, y es claro que g es biyectiva, luego basta probar que $\mathcal{P}^f\mathbb{N}$ es numerable.

Sea $\mathcal{P}^n\mathbb{N}$ el conjunto de los subconjuntos de $\mathbb{N}$ de cardinal n. Una simple inducción prueba que cada $\mathcal{P}^n\mathbb{N}$ es numerable. En efecto, $\mathcal{P}^0\mathbb{N} = \{\varnothing\}$ es finito, y si $\mathcal{P}^n\mathbb{N}$ es numerable, entonces la aplicación $h : \mathcal{P}^n\mathbb{N} \times \mathbb{N} \longrightarrow \mathcal{P}^{n+1}\mathbb{N}$ dada por $h(A, n) = A \cup \{m\}$, donde m es el mínimo número natural $m \geq n$ tal que $m \notin A$ es claramente suprayectiva, por lo que $\mathcal{P}^{n+1}\mathbb{N}$ también es numerable.

Finalmente, $\mathcal{P}^f\mathbb{N} = \bigcup_{n\in\mathbb{N}} \mathcal{P}^n\mathbb{N}$ es numerable por ser unión numerable de conjuntos numerables. ∎

El teorema siguiente viene a decir que los conjuntos numerables son los conjuntos infinitos más pequeños:

Teorema B.8 *Todo conjunto infinito contiene un subconjunto infinito numerable.*

DEMOSTRACIÓN: Si X es un conjunto infinito, podemos definir una aplicación $f : \mathbb{N} \longrightarrow X$ inyectiva por recurrencia. Definidos $f(0), \ldots, f(n-1)$, tenemos que $\{f(0), \ldots, f(n-1)\} \subsetneq X$, porque X es infinito, luego podemos elegir[2] un elemento $f(n) \in X \setminus \{f(0), \ldots, f(n-1)\}$, y es claro que f así definida es inyectiva. ∎

Conjuntos no numerables Es posible asignar un cardinal $|A|$ a todo conjunto A, aunque no sea numerable, de tal modo que $|A| = |B|$ equivalga a que exista una aplicación $A \longrightarrow B$ biyectiva, y $|A| \leq |B|$ equivalga a que exista una aplicación $A \longrightarrow B$ inyectiva. No vamos a demostrar estos hechos, porque no nos harán falta en ningún momento, pero vamos a probar el resultado básico que sirve de fundamento a la teoría de los cardinales infinitos. Nos basamos en un hecho previo:

[2]Esta prueba requiere el axioma de elección, pero puede reducirse al principio de elecciones dependientes (teorema A.3) sin más que considerar el conjunto de todas las aplicaciones inyectivas $s : \{1, \ldots, n\} \longrightarrow X$ ordenado por la inclusión estricta. Una cadena $s_0 \subsetneq s_1 \subsetneq s_2 \subsetneq \cdots$ define una aplicación $\bigcup_{n\in\mathbb{N}} s_n : \mathbb{N} \longrightarrow X$ inyectiva.

Teorema B.9 *Sea X un conjunto y $F : \mathcal{P}X \longrightarrow \mathcal{P}X$ una aplicación tal que si $u \subset v \subset X$ entonces $F(u) \subset F(v)$. Entonces existe un $z \in \mathcal{P}X$ tal que $F(z) = z$.*

DEMOSTRACIÓN: Sea $A = \{u \in \mathcal{P}X \mid F(u) \subset u\}$. Se cumple que A es un conjunto no vacío (pues contiene a X). Llamemos $z = \bigcap_{u\in A} u \in \mathcal{P}X$.

Si $u \in A$, entonces $z \subset u$, luego $F(z) \subset F(u) \subset u$, con lo que $F(z) \subset z$.

Por la hipótesis, $F(F(z)) \subset F(z)$, luego $F(z) \in A$, luego $z \subset F(z)$, lo que nos da la igualdad $F(z) = z$. ∎

Teorema B.10 (Teorema de Cantor-Bernstein) *Sean X e Y conjuntos tales que existen aplicaciones inyectivas $f : X \longrightarrow Y$ y $g : Y \longrightarrow X$. Entonces existe $h : X \longrightarrow Y$ biyectiva.*

DEMOSTRACIÓN: Sea $F : \mathcal{P}X \longrightarrow \mathcal{P}X$ la aplicación dada por $F(u) = X \setminus g[Y \setminus f[u]]$. Estamos en las hipótesis del teorema anterior, pues si $u \subset v \subset X$, entonces

$$f[u] \subset f[v], \quad Y \setminus f[v] \subset Y \setminus f[u], \quad g[Y \setminus f[v]] \subset g[Y \setminus f[u]],$$

$$X \setminus g[Y \setminus f[u]] \subset X \setminus g[Y \setminus f[v]],$$

luego $F(u) \subset F(v)$.

En consecuencia existe un subconjunto $z \subset X$ tal que $F(z) = z$, es decir, $X \setminus g[Y \setminus f[z]] = z$ o, equivalentemente, $X \setminus z = g[Y \setminus f[z]]$. Por consiguiente, $f|_z : z \longrightarrow f[z]$ y $g|_{Y\setminus f[z]} : Y \setminus f[z] \longrightarrow X \setminus z$ son ambas biyectivas, luego la unión de la primera con la inversa de la segunda nos da la aplicación h buscada. ∎

En términos de cardinales, lo que afirma el teorema anterior es que si se cumple $|A| \leq |B|$ y $|B| \leq |A|$, entonces $|A| = |B|$. Existe una aritmética de los cardinales infinitos que generaliza a la que hemos expuesto para conjuntos numerables. Para nuestros fines nos bastará con el resultado siguiente, que en términos de cardinales se interpreta como que si κ es un cardinal infinito, entonces $\kappa \cdot \aleph_0 = \kappa$:

Teorema B.11 (AE) *Si X es infinito, existe una biyección $f : X \times \mathbb{N} \longrightarrow X$.*

DEMOSTRACIÓN: Consideramos el conjunto $\mathcal{F}$ de todas las aplicaciones biyectivas $f : Y \times \mathbb{N} \longrightarrow Y$, con $Y \subset X$ infinito. Observamos que es no vacío, pues por B.8 sabemos que X contiene un subconjunto numerable Y, y entonces existe $f : Y \times \mathbb{N} \longrightarrow Y$ biyectiva, luego $f \in \mathcal{F}$.

Consideramos a $\mathcal{F}$ como conjunto parcialmente ordenado por la inclusión, es decir, una función $f : Y \times \mathbb{N} \longrightarrow Y$ es anterior a otra $f' : Y' \times \mathbb{N} \longrightarrow Y'$ si $Y \subset Y'$ y $f'|_{Y\times\mathbb{N}} = f$.

Es fácil ver que si $\mathcal{C}$ es una cadena en $\mathcal{F}$, entonces $\bigcup_{f\in C} f \in \mathcal{C}$ es una cota superior de $\mathcal{C}$, luego el lema de Zorn nos asegura que existe $f : Y \times \mathbb{N} \longrightarrow Y$ maximal.

Veamos que $X \setminus Y$ tiene que ser finito. En caso contrario contiene un subconjunto numerable Z y existe $g : Z \times \mathbb{N} \longrightarrow Z$ biyectiva. Es claro entonces que $(f \cup g) : (Y \cup Z) \times \mathbb{N} \longrightarrow Y \cup Z$ es biyectiva, luego $f \cup g \in \mathcal{F}$ y contradice la maximalidad de f.

Así pues, $Z = X \setminus Y$ es finito y $f : (X \setminus Z) \times \mathbb{N} \longrightarrow X \setminus Z$ es biyectiva. Basta probar que existe una biyección $X \longrightarrow X \setminus Z$, pues entonces es fácil construir biyecciones

$$X \times \mathbb{N} \longrightarrow (X \setminus Z) \times \mathbb{N} \longrightarrow X \setminus Z \longrightarrow X.$$

Tomamos $W \subset X \setminus Z$ numerable. Entonces $W \cup Z$ es numerable, luego existe una biyección $W \cup Z \longrightarrow W$, que se extiende a una biyección

$$X = (X \setminus (W \cup Z)) \cup (W \cup Z) \longrightarrow (X \setminus (W \cup Z)) \cup W = X \setminus Z.$$

■

La sección 1.5 de [An] contiene algunos resultados adicionales sobre cardinales infinitos.

Apéndice C

Cuadrados latinos

A principios del siglo XVIII era conocido el problema consistente en tomar las dieciséis cartas A, J, Q, K de la baraja y disponerlas en un cuadrado 4 de modo que en ninguna fila o columna haya dos cartas del mismo valor o del mismo palo. El problema aparece planteado y resuelto en un libro de matemática recreativa de 1725. Hay muchas soluciones, una de las cuales es ésta:

$A\spadesuit$	$K\heartsuit$	$Q\diamondsuit$	$J\clubsuit$
$K\diamondsuit$	$A\clubsuit$	$J\spadesuit$	$Q\heartsuit$
$Q\clubsuit$	$J\diamondsuit$	$A\heartsuit$	$K\spadesuit$
$J\heartsuit$	$Q\spadesuit$	$K\clubsuit$	$A\diamondsuit$

Años más tarde, alguien —se dice que la emperatriz Catalina la Grande— planteó a Leonhard Euler una variante de este problema conocida como "el problema de los 36 oficiales":

> *El emperador se disponía a visitar una ciudad en la que estaban acuartelados seis regimientos, y el comandante de la guarnición seleccionó seis oficiales de distinta graduación en cada uno de ellos y quiso disponerlos en formación* 6×6 *para que el emperador pasara revista, y de modo que, cualquiera que fuera la fila o la columna que éste decidiera recorrer, encontrara en ella un oficial de cada regimiento y uno de cada una de las graduaciones. ¿Cómo había que disponer para ello a los* 36 *oficiales?*

Como vemos, "adornos aparte", se trata del mismo problema de la baraja, pero con seis valores y palos en vez de cuatro. Euler decidió abordar el problema como quien se pone a resolver un pasatiempo, pero, para su sorpresa, no fue capaz de encontrar una solución. Más intrigante aún era que, considerando variantes, pudo resolver el problema análogo tanto con 5 oficiales y regimientos como con 7, pero con 6 no consiguió nada. Sus investigaciones alrededor de este problema culminaron con su trabajo *Recherches sur una nouvelle espèce de quarrés magiques,* de 1779.

C.1 Cuadrados latinos y grecolatinos

Euler introdujo el concepto de *"cuadrado latino",* que no es sino una disposición de n signos en n filas y n columnas (repetidos n veces cada uno) de modo que ninguno de ellos aparezca dos veces en una misma fila o columna. El nombre de "cuadrado latino" hace referencia a que Euler usaba letras latinas $a, b, c \ldots$ como signos para sus cuadrados, pero cualquier signo vale. Por ejemplo, dos cuadrados latinos 4×4 son:

A	K	Q	J
K	A	J	Q
Q	J	A	K
J	Q	K	A

♠	♡	♢	♣
♢	♣	♠	♡
♣	♢	♡	♠
♡	♠	♣	♢

Más técnicamente podemos pensar en un cuadrado latino como una matriz (a_{ij}) con n filas y n columnas (y con valores en cualquier conjunto).

Dos cuadrados latinos (a_{ij}) y (α_{ij}) de orden n son *ortogonales* cuando en la matriz $n \times n$ que en la posición (i, j) tiene el par (a_{ij}, α_{ij}) cada uno de los pares aparece una única vez (o, equivalentemente, cuando en ella aparecen los n^2 pares posibles). Los cuadrados obtenidos de esta forma se llaman *grecolatinos,* porque Euler usaba letras latinas $(a, b, c \ldots)$ como signos para uno de los cuadrados y letras griegas $(\alpha, \beta, \gamma \ldots)$ para el otro, de modo que sus cuadrados grecolatinos contenían pares $a\alpha, a\beta, b\alpha, \ldots$ de letras griegas y latinas.

Por ejemplo, los dos cuadrados latinos anteriores son ortogonales, y el cuadrado grecolatino que forman es la solución que hemos dado al problema de las 16 cartas. En estos términos, el problema de los 36 oficiales consiste en encontrar un cuadrado grecolatino de orden 6.

Con esto ya podemos plantear una ligera reformulación del problema: se trata de encontrar un cuadrado latino de orden 6 que tenga un cuadrado latino ortogonal.

Euler dio un criterio sencillo para determinar si un cuadrado latino admite un cuadrado ortogonal, para lo cual introdujo el concepto de *transversal* de un cuadrado latino, que es una selección de n posiciones, de modo que no haya dos en la misma fila o columna y que los elementos del cuadrado en dichas posiciones sean distintos dos a dos. Por ejemplo, la figura siguiente muestra con subíndices cuatro transversales del primero de los cuadrados latinos anteriores:

A_1	K_2	Q_3	J_4
K_3	A_4	J_1	Q_2
Q_4	J_3	A_2	K_1
J_2	Q_1	K_4	A_3

Es inmediato que un cuadrado latino de orden n admite un cuadrado ortogonal si y sólo si admite n transversales disjuntas dos a dos, pues en tal caso

podemos construir el cuadrado ortogonal poniendo un mismo signo en las posiciones correspondientes a cada una de las transversales.

Por ejemplo, el segundo cuadrado latino que hemos mostrado antes (el de los palos) tiene las picas en las posiciones determinadas por la primera transversal (la marcada con los subíndices 1), tiene los corazones en las posiciones de la segunda transversal, los diamantes en las de la tercera y los tréboles en las de la cuarta.

Recíprocamente, un cuadrado ortogonal determina n transversales disjuntas dos a dos correspondientes a las posiciones que ocupa cada uno de sus signos. De hecho, el cuadrado precedente en el que hemos numerado las cuatro transversales es esencialmente el cuadrado grecolatino 4×4 que hemos mostrado en la introducción, con la única diferencia de que los palos están ahora indicados por los subíndices.

Podría parecer que esto es una obviedad que no lleva a ninguna parte, pero no es así. Por ejemplo, es obvio que la tabla de cualquier grupo finito es un cuadrado latino, pero el criterio para que admita un cuadrado ortogonal se simplifica mucho en este caso:

Teorema C.1 *Un cuadrado latino determinado por la tabla de un grupo finito admite un cuadrado latino ortogonal si y sólo si admite una transversal.*

Demostración: En vista de la observación precedente, basta ver que si un cuadrado latino determinado por la tabla de un grupo finito admite una transversal, entonces admite n transversales disjuntas.

En efecto, pongamos que $G = \{g_1, \ldots, g_n\}$ y que tenemos una transversal en su tabla formada por el elemento h_1 del la primera fila, el elemento h_2 de la segunda, etc., hasta h_n, de modo que también $G = \{h_1, \ldots, h_n\}$.

Entonces, para cada $g \in G$, podemos formar una transversal T_g de la tabla de G sin más que tomar el elemento $h_1 g$ en la primera fila, el elemento $h_2 g$ en la segunda fila, etc. Ciertamente, los elementos $h_1 g, h_2 g, \ldots, h_n g$ son distintos dos a dos, por construcción están en filas distintas y, si hubiera dos en la misma columna, por ejemplo, si $h_i g$, $h_j g$ estuvieran en la columna de g_k, eso significaría que $h_i g = g_i g_k$, $h_j g = g_j g_k$, pero entonces $h_i = g_i g_k g^{-1}$, $h_j = g_j g_k g^{-1}$, de modo que h_i, h_j estarían ambos en la columna de $g_k g^{-1}$, contradicción.

Así pues, cada T_g es ciertamente una transversal de la tabla de G, y las n transversales obtenidas de este modo son disjuntas dos a dos, ya que si $g \neq g'$, entonces $h_i g \neq h_i g'$, luego el elemento de la fila i-ésima de T_g es distinto del correspondiente a $T_{g'}$. ∎

Ejemplo Aquí tenemos las tablas de los dos grupos de orden 4:

	1	g	g^2	g^3
1	1	g	g^2	g^3
g	g	g^2	g^3	1
g^2	g^2	g^3	1	g
g^3	g^3	1	g	g^2

	1	a	b	c
1	1	a	b	c
a	a	1	c	b
b	b	c	1	a
c	c	b	a	1

El cuadrado latino correspondiente a la primera tabla (la del grupo C_4) no admite transversales. En efecto, por el teorema anterior, si admitiera una transversal, admitiría cuatro disjuntas, luego cada elemento de la tabla estaría en una transversal. Tratemos de determinar una que contenga al 1 de la primera fila. En la segunda fila tendría que contener a g^2 o a g^3 (ya que no podemos repetir ni la primera columna ni el 1). Si contuviera a g^2, en la tercera fila estaríamos forzados a tomar g, ya que no podemos repetir las dos primeras columnas ni el 1 de la tercera, y en la cuarta fila nos veríamos obligados a tomar el elemento de la tercera columna, que repite g:

	1	g	g^2	g^3
1	1			
g		g^2		
g^2				g
g^3			g	

	1	g	g^2	g^3
1	1			
g			g^3	
g^2				g
g^3		1		

Similarmente, si en la segunda fila tomamos g^3, en la cuarta fila nos vemos obligados a repetir el 1.

Concluimos que la tabla de C_4 da lugar a un cuadrado latino sin complemento ortogonal. En cambio, la tabla de $C_2 \times C_2$ sí que admite una transversal y, por consiguiente, cuatro. De hecho, si cambiamos $1, a, b, c$ por A, K, Q, J, vemos que el cuadrado latino correspondiente a dicha tabla es precisamente el que hemos usado para distribuir las cartas del problema de las 16 cartas según su valor, y ya hemos determinado cuatro transversales disjuntas en él. ∎

Vemos, pues, que algunos grupos permiten construir cuadrados grecolatinos, pero otros no. Ahora bien:

Teorema C.2 *La tabla de un grupo de orden impar es un cuadrado latino que admite siempre un cuadrado latino ortogonal.*

DEMOSTRACIÓN: Si G es un grupo de orden impar, basta probar que la diagonal de su tabla es una transversal. La diagonal está formada por los cuadrados en G, luego basta probar que dichos cuadrados son todos los elementos de G. Ello se debe a que la aplicación $g \mapsto g^2$ es inyectiva, pues si $g^2 = h^2$, entonces $(gh^{-1})^2 = 1$, pero si G tiene orden impar no puede tener elementos de orden 2, luego tiene que ser $gh^{-1} = 1$, es decir, $g = h$. ∎

Así pues:

Teorema C.3 (Euler) *Existen cuadrados grecolatinos de todos los órdenes impares.*

Ejemplo La tabla de C_3 nos da un cuadrado grecolatino:

	1	g	g^2
1	1	g	g^2
g	g	g^2	1
g^2	g^2	1	g

a	b	c
b	c	a
c	a	b

α	β	γ
γ	α	β
β	γ	α

$a\alpha$	$b\beta$	$c\gamma$
$b\gamma$	$c\alpha$	$a\beta$
$c\beta$	$a\gamma$	$b\alpha$

Ahora el lector puede resolver sistemáticamente, es decir, sin necesidad de tanteos, el "problema de los 25 oficiales" y "el problema de los 49 oficiales", pero si intentamos aplicar el método al problema de los 36 oficiales nos encontramos con que ninguno de los dos grupos de orden 6, es decir, ni C_6 ni Σ_3, determinan cuadrados grecolatinos. Sin embargo, esto no implica que no existan cuadrados grecolatinos de orden 6, ya que no es cierto que todo cuadrado grecolatino se obtenga necesariamente de un grupo. ∎

Veamos ahora que cada cuerpo finito nos proporciona familias de cuadrados latinos ortogonales dos a dos:

Teorema C.4 *Si k es un cuerpo finito de n elementos, para cada $a \in k$ no nulo, un cuadrado latino L_a de orden n viene dado por $L_a(i,j) = ai+j$ (donde usamos como índices los elementos $i, j \in k$). Ademas los $n-1$ cuadrados latinos L_a son ortogonales dos a dos.*

DEMOSTRACIÓN: Los elementos de la fila i-ésima de L_a son distintos dos a dos, pues si $L_a(i,j) = L_a(i,j')$, entonces $ai + j = ai + j'$, luego $j = j'$. Lo mismo sucede con las columnas: si $L_a(i,j) = L_a(i',j)$, entonces $ai+j = ai'+j$, luego $ai = ai'$ y, como $a \neq 0$, llegamos de nuevo a que $i = i'$. Esto prueba que cada L_a es un cuadrado latino.

Supongamos ahora que $a, a' \in k$ son no nulos y distintos entre sí y veamos que L_a es ortogonal a $L_{a'}$. Para ello basta ver que si $L_a(i,j) = L_a(i',j')$ y $L_{a'}(i,j) = L_{a'}(i',j')$, necesariamente $(i,j) = (i',j')$, es decir, que el cuadrado que se forma al combinar ambos cuadrados latinos es grecolatino. En efecto, tenemos que

$$ai + j = ai' + j', \qquad a'i + j = a'i' + j',$$

lo que equivale a que

$$a(i-i') + j - j' = 0, \qquad a'(i-i') + j - j' = 0,$$

o también a que

$$(i-i', j-j') \begin{pmatrix} a & a' \\ 1 & 1 \end{pmatrix} = (0,0),$$

pero la matriz tiene determinante no nulo, luego multiplicando por su inversa llegamos a que $(i-i', j-j') = (0,0)$, es decir, a que $i = i'$, $j = j'$. ∎

Como consecuencia:

Teorema C.5 (Euler) *Si $n > 2$ es potencia de primo, entonces existen cuadrados grecolatinos de orden n.*

En efecto, basta tener en cuenta que, según el teorema 9.2, existen cuerpos finitos de orden cualquier potencia de primo, pero no podemos aplicar el teorema al caso $n = 2$ ya que entonces sólo nos garantiza la existencia de un cuadrado latino (sin otros cuadrados ortogonales).

Ejemplo Vamos a aplicar el teorema anterior al cuerpo de 4 elementos. Un polinomio irreducible sobre el cuerpo $\{0,1\}$ de dos elementos es $x^2 + x + 1$,

luego el cuerpo k de cuatro elementos se obtiene de adjuntarle una raíz α de este polinomio, es decir, un elemento α tal que $\alpha^2 = \alpha + 1$. Los elementos de k son, pues, $0, 1, \alpha, \alpha+1$, y es fácil ver entonces que los cuadrados latinos $L_1, L_\alpha, L_{\alpha+1}$ son:

L_1	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

L_α	0	1	α	$\alpha+1$
0	0	1	α	$\alpha+1$
1	α	$\alpha+1$	0	1
α	$\alpha+1$	α	1	0
$\alpha+1$	1	0	$\alpha+1$	α

$L_{\alpha+1}$	0	1	α	$\alpha+1$
0	0	1	α	$\alpha+1$
1	$\alpha+1$	α	1	0
α	1	0	$\alpha+1$	α
$\alpha+1$	α	$\alpha+1$	0	1

Equivalentemente:

a	b	c	d
b	a	d	c
c	d	a	b
d	c	b	a

α	β	γ	δ
γ	δ	α	β
δ	γ	β	α
β	α	δ	γ

α	β	γ	δ
δ	γ	β	α
β	α	δ	γ
γ	δ	α	β

Los dos primeros son los cuadrados latinos con los que hemos resuelto el problema de las 16 cartas, pero al combinar estos tres cuadrados obtenemos otras dos soluciones alternativas. ∎

Conviene hacer dos observaciones en torno al teorema anterior. La primera es que, ciertamente, no existen cuadrados grecolatinos de orden 2, y la razón es que, como es fácil comprobar, sólo existen dos cuadrados latinos de orden 2, pero no son ortogonales. La segunda observación es en realidad una generalización de este hecho:

Teorema C.6 *Para cada número natural n, no puede haber más de $n-1$ cuadrados latinos ortogonales dos a dos.*

Demostración: Obviamente, un cuadrado latino no deja de serlo (ni deja de ser ortogonal a otro) si cambiamos sus signos por otros, por lo que, dada una familia de cuadrados latinos de orden n ortogonales dos a dos, podemos suponer que sus signos son $1, \ldots, n$ y que la primera fila de todos los cuadrados es precisamente $(1, \ldots, n)$.

Consideremos entonces los elementos situados en la posición $(2, 1)$. Ninguno de ellos puede ser 1, porque entonces el 1 aparecería repetido en la primera columna. Y no puede haber dos cuadrados latinos en la familia con el mismo elemento, digamos x, en la posición $(2, 1)$, porque en tal caso, en el cuadrado grecolatino determinado por ambos, el par (x, x) aparecería en la posición $(2, 1)$ y también en la posición $(1, x)$. Por lo tanto, la familia dada no puede tener más de $n-1$ elementos. ∎

Veamos una última técnica elemental de construcción de cuadrados grecolatinos:

Definición C.7 Si $L = (a_{ij})$ y $M = (b_{ij})$ son dos cuadrados latinos de órdenes m y n, respectivamente, definimos su *producto* como el cuadrado $L \otimes M$ de orden mn que, en la posición $((i,j),(i',j'))$ tiene el par $(a_{ii'}, b_{jj'})$.

Ejemplo Vamos a calcular el producto de los cuadrados latinos

	1	2	3
1	a	b	c
2	b	c	a
3	c	a	b

	1	2	3	4
1	a	b	c	d
2	b	a	d	c
3	c	d	a	b
4	d	c	b	a

(los correspondientes a los grupos C_3 y $C_2 \times C_2$). Se trata simplemente de:

	(1,1)	(1,2)	(1,3)	(1,4)	(2,1)	(2,2)	(2,3)	(2,4)	(3,1)	(3,2)	(3,3)	(3,4)
(1,1)	aa	ab	ac	ad	ba	bb	bc	bd	ca	cb	cc	cd
(1,2)	ab	aa	ad	ac	bb	ba	bd	bc	cb	ca	cd	cc
(1,3)	ac	ad	aa	ab	bc	bd	ba	bb	cc	cd	ca	cb
(1,4)	ad	ac	ab	aa	bd	bc	bb	ba	cd	cc	cb	ca
(2,1)	ba	bb	bc	bd	ca	cb	cc	cd	aa	ab	ac	ad
(2,2)	bb	ba	bd	bc	cb	ca	cd	cc	ab	aa	ad	ac
(2,3)	bc	bd	ba	bb	cc	cd	ca	cb	ac	ad	aa	ab
(2,4)	bd	bc	bb	ba	cd	cc	cb	ca	ad	ac	ab	aa
(3,1)	ca	cb	cc	cd	aa	ab	ac	ad	ba	bb	bc	bd
(3,2)	cb	ca	cd	cc	ab	aa	ad	ac	bb	ba	bd	bc
(3,3)	cc	cd	ca	cb	ac	ad	aa	ab	bc	bd	ba	bb
(3,4)	cd	cc	cb	ca	ad	ac	ab	aa	bd	bc	bb	ba

■

Teorema C.8 *El producto de dos cuadrados latinos L y M es de nuevo un cuadrado latino, y si ambos tienen cuadrados latinos ortogonales L' y M', entonces $L \otimes M$ es ortogonal a $L' \otimes M'$.*

DEMOSTRACIÓN: Pongamos que $L = (a_{ij})$, $M = (b_{ij})$. Fijemos una fila (i,j) de $L \otimes M$ y vamos a probar que en ella se encuentra cualquier par (x,y). Como L es un cuadrado latino, existe una única columna i' tal que $a_{ii'} = x$. Igualmente, como M es un cuadrado latino, existe una única columna j' tal que $b_{jj'} = y$, y entonces xy está en la fila (i,j) columna (i',j') del producto. Igualmente se razona que todos los pares aparecen en cada columna del producto, luego se trata de un cuadrado latino.

Para la segunda parte, pongamos que $L' = (\alpha_{ij})$, $M' = (\beta_{ij})$ y vamos a probar que todo par de pares ordenados $((x_1,y_1),(x_2,y_2))$ aparece en el producto.

Esto equivale a que existan índices $((i,j),(i',j'))$ tales que $(a_{ii'},b_{jj'})=(x_1,y_1)$, $(\alpha_{ii'},\beta_{jj'})=(x_2,y_2)$. Equivalentemente, necesitamos que

$$a_{ii'}=x_1,\qquad \alpha_{ii'}=x_2,\qquad b_{jj'}=y_1,\qquad \beta_{jj'}=y_2,$$

pero la existencia de (i,i') que cumplen las dos primeras igualdades se sigue de que L y L' son ortogonales, y la existencia de (j,j') se sigue de que lo sean M y M'. ∎

Por consiguiente:

Teorema C.9 *Si existen cuadrados grecolatinos de órdenes m y n, también existen de orden mn.*

Así pues, si hubiera cuadrados grecolatinos de orden 2, los habría de todos los órdenes primos, luego por el teorema anterior los habría de todos los órdenes. La excepción que supone el 2 nos obliga a reducir la conclusión:

Teorema C.10 (Euler) *Si $n\not\equiv 2 \pmod 4$, existen cuadrados grecolatinos de orden n.*

Demostración: Basta tener en cuenta que n se descompone en productos de potencias de primos distintos donde la potencia de 2 no será exactamente 2, por lo que existen cuadrados latinos de orden todas las potencias de primo que aparecen en la descomposición de n, y por el teorema anterior también los hay de orden n. ∎

Teniendo en cuenta este resultado, la inexistencia de cuadrados grecolatinos de orden 2 y la dificultad de encontrar uno de orden 6, Euler formuló la conjetura siguiente:

> **Conjetura de Euler** *No existen cuadrados grecolatinos de ningún orden $n\equiv 2 \pmod 4$.*

En 1901 un matemático aficionado francés llamado Gaston Tarry publicó un artículo de unas 30 páginas en las que demostraba que no existen cuadrados grecolatinos de orden 6. Para ello tuvo que generar y comparar 9 408 pares de cuadrados latinos para comprobar que no eran ortogonales. Ello le llevó dos años de trabajo dominical.

Sin embargo, en 1959 los matemáticos indios Raj Chandra Bose y Sharadchandra Shankar Shrikhande construyeron un cuadrado grecolatino de orden 22, refutando así la conjetura de Euler. Poco después, el estadounidense Ernest Tilden Parker encontró un cuadrado grecolatino de orden 10 ejecutando durante una hora un programa informático en un ordenador militar de la época. Finalmente, en el mismo año 1959, Bose, Shrikhande y Parker demostraron que la conjetura de Euler es falsa para todo $n\geq 10$ o, en otras palabras, que existen cuadrados grecolatinos de todos los órdenes excepto $n=2,6$.

Sucede que los cuadrados grecolatinos de órdenes $n\equiv 2 \pmod 4$ no pueden construirse con técnicas elementales como las que hemos visto en esta sección (a

partir de grupos, cuerpos o combinando cuadrados de orden menor), sino que todas las construcciones conocidas requieren técnicas combinatorias sofisticadas o, más precisamente, de lo que se conoce como *teoría de diseños.*

En la sección siguiente mostramos algunas caracterizaciones combinatorias de las familias de cuadrados mutuamente ortogonales y en la sección C.3 las usaremos para demostrar la no existencia de cuadrados grecolatinos de orden 6 con un argumento de D.R. Stinson (de 1982) mucho más breve y conceptual que el de Tarry.

Por otra parte, la dificultad de construir un cuadrado grecolatino de orden 10 es mucho mayor que la de justificar que existe uno exhibiéndolo explícitamente:

$a\alpha$	$b\beta$	$c\gamma$	$d\delta$	$e\epsilon$	$f\zeta$	$g\eta$	$h\theta$	$i\iota$	$j\kappa$
$b\kappa$	$a\theta$	$h\iota$	$c\alpha$	$g\zeta$	$e\beta$	$f\epsilon$	$j\delta$	$d\gamma$	$i\eta$
$c\epsilon$	$g\gamma$	$i\zeta$	$e\iota$	$j\beta$	$h\delta$	$b\theta$	$f\kappa$	$a\eta$	$d\alpha$
$d\beta$	$c\iota$	$a\kappa$	$h\zeta$	$f\eta$	$i\theta$	$j\alpha$	$e\gamma$	$b\delta$	$g\epsilon$
$e\zeta$	$j\eta$	$b\alpha$	$g\theta$	$i\delta$	$c\kappa$	$d\iota$	$a\beta$	$h\epsilon$	$f\gamma$
$f\iota$	$e\delta$	$d\theta$	$j\gamma$	$h\alpha$	$b\epsilon$	$i\beta$	$c\eta$	$g\kappa$	$a\zeta$
$g\delta$	$f\alpha$	$j\epsilon$	$i\kappa$	$a\iota$	$d\eta$	$h\gamma$	$b\zeta$	$c\beta$	$e\theta$
$h\eta$	$d\zeta$	$g\beta$	$a\epsilon$	$b\gamma$	$j\iota$	$e\kappa$	$i\alpha$	$f\theta$	$c\delta$
$i\gamma$	$h\kappa$	$e\eta$	$f\beta$	$c\theta$	$g\alpha$	$a\delta$	$d\epsilon$	$j\zeta$	$b\iota$
$j\theta$	$i\epsilon$	$f\delta$	$b\eta$	$d\kappa$	$a\gamma$	$c\zeta$	$g\iota$	$e\alpha$	$h\beta$

Cuadrados mágicos Terminamos esta sección señalando una conexión entre los cuadrados grecolatinos y los cuadrados mágicos. Un *cuadrado mágico* se define como una matriz $n \times n$ que contiene los números $1, \ldots, n^2$, de modo que todas las filas, columnas y diagonales tienen la misma suma.

Observemos que dicha suma es necesariamente

$$\frac{n(n^2+1)}{2},$$

pues al sumar todos los números del cuadrado obtenemos

$$1 + 2 + \cdots + n^2 = \frac{n^2(n^2+1)}{2},$$

y este valor tiene que ser n veces la suma de las filas o de las columnas.

Si no exigimos que las diagonales sumen lo mismo que las filas y las columnas, tenemos un cuadrado *semimágico.*

Todo cuadrado grecolatino determina un cuadrado semimágico sin más que asignar a sus signos los valores $0, 1, \ldots, n-1$ (en cualquier orden) e interpretar cada par $(x, \xi) = nx + \xi + 1$, es decir, como el número natural cuyos dígitos en base n son (x, ξ), al que le sumamos una unidad para que los elementos del cuadrado resultante recorran $1, \ldots, n^2$ en lugar de $0, \ldots, n^2 - 1$. De este modo, cada fila o columna del cuadrado resultante sumará

$$n(0+1+2+\cdots+n-1)+1+2+\cdots+n = \frac{n^2(n-1)}{2} + \frac{n(n+1)}{2} = \frac{n(n^2+1)}{2}.$$

Un cuadrado latino es *diagonal* si los signos que se encuentran en sus dos diagonales no se repiten. Es claro que si un cuadrado grecolatino se obtiene de dos cuadrados latinos diagonales ortogonales, el cuadrado semimágico que determina será mágico.

Por ejemplo, el cuadrado grecolatino

$$\begin{array}{|cccc|}\hline a\alpha & b\beta & c\gamma & d\delta \\ c\delta & d\gamma & a\beta & b\alpha \\ d\beta & c\alpha & b\delta & a\gamma \\ b\gamma & a\delta & d\alpha & c\beta \\ \hline\end{array}$$

se obtiene de dos cuadrados latinos diagonales, y haciendo $a = \alpha = 0$, $b = \beta = 1$, $c = \gamma = 2$, $d = \delta = 3$, obtenemos el cuadrado mágico

$$\begin{array}{|cccc|}\hline 1 & 6 & 11 & 16 \\ 12 & 15 & 2 & 5 \\ 14 & 9 & 8 & 3 \\ 7 & 4 & 13 & 10 \\ \hline\end{array}$$

en el que las filas, las columnas y las diagonales suman 34 (al igual que las esquinas, que las cuatro casillas centrales, que los cuatro cuadrantes, etc., como se ve claramente examinando el cuadrado grecolatino del que procede).

Consideremos ahora el cuadrado grecolatino

$$\begin{array}{|ccc|}\hline a\alpha & b\beta & c\gamma \\ b\gamma & c\alpha & a\beta \\ c\beta & a\gamma & b\alpha \\ \hline\end{array}$$

Los cuadrados latinos de los que procede no son diagonales (es fácil probar que no existen cuadrados latinos diagonales de orden 3). Si damos a a, b, c y a α, β, γ los valores $0, 1, 2$, sabemos que el cuadrado grecolatino se convertirá en un cuadrado semimágico en el que las filas y las columnas sumen 15, pero podemos forzar a que sea mágico imponiendo que las diagonales también sumen 15:

$$3a + \alpha + 1 + 3c + \alpha + 1 + 3b + \alpha + 1 = 15,$$

$$3c + \gamma + 1 + 3c + \alpha + 1 + 3c + \beta + 1 = 15.$$

Como $a + b + c = \alpha + \beta + \gamma = 0 + 1 + 2 = 3$, esto equivale a que

$$3 \cdot 3 + 3\alpha + 3 = 15, \qquad 9c + 3 + 3 = 15,$$

y así, obtendremos un cuadrado latino sin más que exigir que $c = \alpha = 1$. Tomando, por ejemplo, $a = \beta = 0$, $b = \gamma = 2$ queda el cuadrado mágico

$$\begin{array}{|ccc|}\hline 2 & 7 & 6 \\ 9 & 5 & 1 \\ 4 & 3 & 8 \\ \hline\end{array}$$

que era conocido por los chinos desde alrededor de 650 a.C.

C.2 Caracterizaciones combinatorias

Supongamos que tenemos k cuadrados latinos de orden n mutuamente ortogonales (de modo que $k \leq n-1$, por el teorema C.6). Podemos suponer que los signos de todos ellos son los números $1, \ldots, n$. Por ejemplo, antes hemos obtenido 3 cuadrados latinos mutuamente ortogonales de orden 4:

$$\begin{array}{|cccc|}\hline 1&2&3&4\\ 2&1&4&3\\ 3&4&1&2\\ 4&3&2&1\\ \hline\end{array}\qquad \begin{array}{|cccc|}\hline 1&2&3&4\\ 3&4&1&2\\ 4&3&2&1\\ 2&1&4&3\\ \hline\end{array}\qquad \begin{array}{|cccc|}\hline 1&2&3&4\\ 4&3&2&1\\ 2&1&4&3\\ 3&4&1&2\\ \hline\end{array}$$

Con ellos podemos construir una matriz $(k+2)\times n^2$ cuyas dos primeras filas contengan los n^2 pares (i,j) y las k filas siguientes contengan el elemento que el cuadrado k-ésimo de la familia tiene en la posición (i,j). En el caso de los tres cuadrados anteriores la matriz es

$$\begin{array}{c|cccccccccccccccc} F&1&1&1&1&2&2&2&2&3&3&3&3&4&4&4&4\\ C&1&2&3&4&1&2&3&4&1&2&3&4&1&2&3&4\\ 1&1&2&3&4&2&1&4&3&3&4&1&2&4&3&2&1\\ 2&1&2&3&4&3&4&1&2&4&3&2&1&2&1&4&3\\ 3&1&2&3&4&4&3&2&1&2&1&4&3&3&4&1&2 \end{array}$$

Por ejemplo, el 2 de la esquina inferior derecha indica que el tercer cuadrado latino tiene un 2 en la posición $(4,4)$.

Esta matriz tiene la propiedad de que, fijadas dos filas, cada par (i,j) aparece exactamente una vez en ellas. Para las filas F y C esto es cierto por construcción, para las filas F y r esto equivale a que en el cuadrado r cada número $1,\ldots,n$ aparece exactamente una vez en cada fila, para las filas C y r esto expresa que en el cuadrado r cada número aparece exactamente una vez en cada columna, y para dos filas r y r' esto equivale a la ortogonalidad de los cuadrados r y r'. Por esto mismo, a partir de una matriz en estas condiciones es posible construir k cuadrados latinos mutuamente ortogonales. En resumen:

Teorema C.11 *Existe una familia de k cuadrados latinos de orden n mutuamente ortogonales si y sólo si existe una matriz $(k+2)\times n^2$ de números $1,\ldots,n$ de modo que, en cada par de filas, cada par (i,j) aparece exactamente una vez.*

Dada una matriz en las condiciones del teorema anterior, llamamos X al conjunto de los n^2 pares (i,j), a los que llamaremos "puntos" y, para cada par de índices $i = 1,\ldots,k+2$, $j = 1,\ldots,n$, definimos la "recta" L_{ij} formada por todos los puntos (u,v) tales que, donde en las dos primeras filas de la matriz está el par (u,v), en la fila i se encuentra el valor j.

En la práctica llamaremos a las filas de la matriz $F, C, 1, \ldots, k$ y, consecuentemente, a las rectas las llamaremos $F_1,\ldots,F_n$, $C_1,\ldots,C_n$, L_{ij}, con $i = 1,\ldots,k$, $j = 1,\ldots n$. En nuestro ejemplo tenemos 16 puntos y las 20 rectas siguientes:

1	$F_1 = \{(1,1),(1,2),(1,3),(1,4)\}$	5	$C_1 = \{(1,1),(2,1),(3,1),(4,1)\}$
2	$F_2 = \{(2,1),(2,2),(2,3),(2,4)\}$	6	$C_2 = \{(1,2),(2,2),(3,2),(4,2)\}$
3	$F_3 = \{(3,1),(3,2),(3,3),(3,4)\}$	7	$C_3 = \{(1,3),(2,3),(3,3),(4,3)\}$
4	$F_4 = \{(4,1),(4,2),(4,3),(4,4)\}$	8	$C_4 = \{(1,4),(2,4),(3,4),(4,4)\}$
9	$L_{11} = \{(1,1),(2,2),(3,3),(4,4)\}$	13	$L_{21} = \{(1,1),(2,3),(3,4),(4,2)\}$
10	$L_{12} = \{(1,2),(2,1),(3,4),(4,3)\}$	14	$L_{22} = \{(1,2),(2,4),(3,3),(4,1)\}$
11	$L_{13} = \{(1,3),(2,4),(3,1),(4,2)\}$	15	$L_{23} = \{(1,3),(2,1),(3,2),(4,4)\}$
12	$L_{14} = \{(1,4),(2,3),(3,2),(4,1)\}$	16	$L_{24} = \{(1,4),(2,2),(3,1),(4,3)\}$
17	$L_{31} = \{(1,1),(2,4),(3,2),(4,3)\}$		
18	$L_{32} = \{(1,2),(2,3),(3,1),(4,4)\}$		
19	$L_{33} = \{(1,3),(2,2),(3,4),(4,1)\}$		
20	$L_{34} = \{(1,4),(2,1),(3,3),(4,2)\}$		

Por ejemplo, para calcular la recta L_{34} miramos la fila 3 de la tabla, buscamos las columnas en las que hay un 4 y anotamos los pares (i,j) que hay en las filas F y C de dichas columnas.

De este modo, dos rectas de la forma L_{ij}, $L_{ij'}$ (con $j \neq j'$) son "paralelas" en el sentido de que no tienen puntos en común, ya que si un punto p está en L_{ij} eso significa que en la fila i y en la única columna que tiene a p en las filas F y C se encuentra el valor j, lo que a su vez implica que $p \notin L_{ij'}$.

Por el contrario, dos rectas de la forma L_{ij}, $L_{i'j}$ son "secantes", en el sentido de que tienen un único punto en común. En efecto, que un punto esté en ambas rectas significa que en la columna que tiene a p en las filas F y C está el par (j,j') en las filas (i,i'), y, en efecto, partimos de una matriz que tiene cada par una única vez en cada par de columnas, luego hay un único punto en la intersección. Con esto tenemos probada la mitad del teorema siguiente:

Teorema C.12 *Existe una matriz en las condiciones del teorema anterior si y sólo si existe un conjunto X de n^2 puntos y una familia de $(k+2)n$ rectas (conjuntos de n puntos) divididas en $k+2$ familias de n rectas "paralelas" (con intersección vacía) de modo que dos rectas de familias distintas se cortan en un único punto.*

Falta razonar que a partir de una estructura de puntos y rectas en las condiciones indicadas es posible construir una matriz en las condiciones del teorema C.11. Para ello fijamos dos de las familias de n rectas paralelas, a las que llamamos $F_1,\dots,F_n$ y $C_1,\dots,C_n$. Como son familias de n conjuntos disjuntos de n puntos, cada punto pertenece exactamente a una recta de cada familia. Llamamos (i,j) al único punto que está en $F_i \cap C_j$. Por otra parte, numeramos las otras k familias de rectas paralelas con los números $i = 1,\dots,k$ y numeramos como $L_{i,1},\dots,L_{i,n}$ las rectas de la familia i-ésima.

Ahora basta formar una matriz con dos filas F y C en las que aparezcan todos los pares $p \in X$ y que en cada una de las otras filas $i = 1,\dots,k$ ponemos en la columna p el valor j correspondiente a la recta L_{ij} a la que pertenece p. Es

fácil ver que la matriz así construida cumple las condiciones del teorema C.11. Por ejemplo, si en dos filas i, i' apareciera repetido un par (j, j') eso significaría que los puntos p yp' en los que aparece repetido el par estarían ambos en las rectas $L_{i,j}$ y $L_{i',j'}$. ∎

Dado un sistema de puntos y rectas en las condiciones del teorema anterior, numeramos todas las rectas consecutivamente, $L_1, \ldots, L_{(k+2)n}$, con lo que podemos formar una familia de $k+2$ "grupos" disjuntos de n índices correspondientes a las familias de rectas paralelas y una familia de n^2 "bloques" formados por los conjuntos de índices de rectas a las que pertenece un mismo punto.

En nuestro ejemplo, con la numeración de las rectas que hemos indicado al definirlas, los 5 grupos son:

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

y los 16 bloques son:

B_{11}	$1,5,9,13,17$	B_{12}	$1,6,10,14,18$	B_{13}	$1,7,11,15,19$	B_{14}	$1,8,12,16,20$
B_{21}	$2,5,10,15,20$	B_{22}	$2,6,9,16,19$	B_{23}	$2,7,12,13,18$	B_{24}	$2,8,11,14,17$
B_{31}	$3,5,11,16,18$	B_{32}	$3,6,12,15,17$	B_{33}	$3,7,9,14,20$	B_{34}	$3,8,10,13,19$
B_{41}	$4,5,12,14,19$	B_{42}	$4,6,11,13,20$	B_{43}	$4,7,10,16,17$	B_{44}	$4,8,9,15,18$

Por ejemplo, el bloque B_{44} contiene los índices de las cinco rectas que pasan por el punto $(3,4)$.

Claramente, dos elementos de un mismo grupo no forman parte de un mismo bloque (pues esto equivale a que dos rectas paralelas no se cortan) y dos elementos de grupos distintos figuran en un único bloque (pues esto equivale a que dos rectas no paralelas se cortan en un único punto). Por consiguiente:

Teorema C.13 *Existe una estructura de puntos y rectas en las condiciones del teorema anterior si y sólo si existe una terna $(X, \mathcal{G}, \mathcal{B})$ donde X es un conjunto con $n(k+2)$ elementos, $\mathcal{G}$ es una partición de X en $k+2$ grupos de n elementos y $\mathcal{B}$ es una familia de n^2 bloques de $k+2$ elementos de X de modo que dos elementos de grupos distintos están contenidos en un único bloque y dos elementos de un mismo grupo no están en ningún bloque.*

Las configuraciones descritas en los teoremas anteriores son estructuras combinatorias. Por ejemplo, la estructura del último teorema es lo que llama un *diseño transversal de índice* 1 (tendría índice l si cada par de elementos de grupos distintos perteneciera exactamente a l bloques). En estos términos, lo que hemos probado es que existen k cuadrados latinos de orden n mutuamente ortogonales si y sólo si existe un diseño transversal de índice 1 con k grupos de tamaño n.

Los diseños transversales no recuerdan en nada a los cuadrados latinos, pero sucede que es en términos de ésta y otras estructuras relacionadas como se ha logrado abordar con éxito la conjetura de Euler. En la sección siguiente mostramos la utilidad de este enfoque demostrando que no existen cuadrados grecolatinos de orden 6.

C.3 El problema de los 36 oficiales

Ahora estamos en condiciones de dar una demostración elegante de que el problema de los 36 oficiales no tiene solución. Según hemos visto, esto equivale a demostrar que no existen $k = 2$ cuadrados latinos ortogonales de orden $n = 6$ y, en virtud del último teorema de la sección anterior, basta probar que no existe ninguna terna $(X, \mathcal{G}, \mathcal{B})$, donde X es un conjunto de $n(k+2) = 24$ elementos, $\mathcal{G}$ es una partición de X en $k+2 = 4$ grupos de $n = 6$ elementos y $\mathcal{B}$ es una familia de $n^2 = 36$ bloques de $k+2 = 4$ elementos de X cada uno, de modo que dos elementos de grupos distintos están contenidos en un único bloque y dos elementos de un mismo grupo no coinciden en ningún bloque.

No podemos poner ningún ejemplo de terna en estas condiciones porque, precisamente, vamos a demostrar que no existen, pero podemos ilustrar el argumento con el caso correspondiente a $k = 2$, $n = 4$, que resulta de eliminar el último grupo y sus cuatro elementos del ejemplo que hemos considerado en la sección anterior. En este ejemplo tenemos entonces un conjunto X de $n(k+2) = 16$ elementos con los $k+2 = 4$ grupos siguientes:

$$B_1 = \{1,2,3,4\},\ B_2 = \{5,6,7,8\},\ B_3 = \{9,10,11,12\},\ B_4 = \{13,14,15,16\}$$

y con los 16 bloques

B_5	$1,5,\ 9,13$	B_6	$1,6,10,14$	B_7	$1,7,11,15$	B_8	$1,8,12,16$
B_9	$2,5,10,15$	B_{10}	$2,6,\ 9,16$	B_{11}	$2,7,12,13$	B_{12}	$2,8,11,14$
B_{13}	$3,5,11,16$	B_{14}	$3,6,12,15$	B_{15}	$3,7,\ 9,14$	B_{16}	$3,8,10,13$
B_{17}	$4,5,12,14$	B_{18}	$4,6,11,13$	B_{19}	$4,7,10,16$	B_{20}	$4,8,\ 9,15$

Pero recordemos que en el caso que realmente nos interesa tenemos 4 grupos de 6 elementos y 36 bloques de 4 elementos. Los vamos a numerar todos consecutivamente $B_1, \ldots, B_N$ de modo que los grupos sean los 4 primeros conjuntos y los bloques los siguientes. En realidad es $N = 40$, pero en nuestro ejemplo es $N = 20$.

Observemos que cada $x \in X$ forma $3n$ pares con elementos de otros grupos, y cada bloque que contenga a x contiene exactamente 3 pares que contienen a x, luego cada elemento x de X tiene que estar exactamente en n bloques.

Podemos construir entonces la *matriz de incidencia* $M = (m_{ij})$, definida por

$$m_{ij} = \begin{cases} 1 & \text{si } i \in B_j, \\ 0 & \text{si } i \notin B_j. \end{cases}$$

Se trata de una matriz $4n \times N$, es decir, 24×40 en realidad, pero que en nuestro ejemplo es la matriz 16×20 que mostramos en la página siguiente. Cada fila de la matriz corresponde a un elemento $x \in X$ y contiene exactamente $n+1$ unos, uno en una de las 4 primeras columnas (que indica el grupo al cual pertenece x) y los otros n en las siguientes (que indican los n bloques a los que pertenece x).

M	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0
4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
5	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
6	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0
7	0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0
8	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
9	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1
10	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	1	0
11	0	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0	1	0	0
12	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0
13	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0
14	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0
15	0	0	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	1
16	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1	0

Podemos considerar cada fila de M como un elemento de $V = k^N$, donde $k = \{0, 1\}$ es el cuerpo de 2 elementos. V es un espacio vectorial en el que podemos considerar el producto escalar

$$(v, w) = \sum_{i=1}^{N} v_i w_i.$$

Entonces, si F_i, F_j son dos filas cualesquiera de M, se cumple que $(F_i, F_j) = 1$. En efecto, si $i = j$, entonces el producto es una suma de $n + 1$ unos, que es un número impar, luego el resultado es 1. Si $i \neq j$, distinguimos dos casos según si las dos filas tienen su primer 1 en la misma columna o no.

Si F_i, F_j tienen el primer 1 en la misma columna, eso significa que i, j están en el mismo grupo, luego el par $\{i, j\}$ no está contenido en ningún bloque, luego cuando hay un 1 en la fila i, hay un 0 en la fila j y viceversa, por lo que el producto (F_i, F_j) sólo tiene un sumando igual a 1.

Si F_i, F_j tienen el primer 1 en columnas distintas, eso significa que i, j están en grupos distintos, luego $\{i, j\}$ está en un único bloque, luego sólo hay una columna en la que F_i y F_j tienen simultáneamente un 1, luego el producto (F_i, F_j) también tiene un sumando igual a 1.

Sea $W = \langle F_1, \ldots, F_{n^2} \rangle$ el subespacio de V generado por las filas de M y sea

$$W^\perp = \{v \in V \mid (u, F_i) = 0,\ i = 1, \ldots, n^2\} \leq V.$$

El producto escalar que estamos considerando en V no es completamente análogo al producto escalar usual en $\mathbb{R}^n$ porque un vector puede ser ortogonal a sí mismo, pero la demostración del teorema de ortogonalización de Gram-Schmidt [G 3.21] es elemental y es válida en este contexto. Por lo tanto, podemos tomar una base ortogonal de W y extenderla hasta una base ortogonal de V,

con lo que los vectores añadidos estarán en $W^\perp$ (pues al ser ortogonales a una base de W, lo son a todos los vectores de W, y en particular a los generadores F_i), y esto nos permite concluir que $V = W + W^\perp$.

Como V está generado por los vectores F_i, el cociente $W/(W \cap W^\perp)$ está generado por las clases $[F_i]$, pero se cumple que

$$(F_i + F_j, F_k) = (F_i, F_k) + (F_j, F_k) = 1 + 1 = 0,$$

luego $F_i + F_j \in W^\perp$, luego $[F_i] = [F_j]$, luego $W/(W \cap W^\perp) = \langle [F_1] \rangle$. Además, como $(F_1, F_1) = 1$, tenemos que $F_1 \notin W^\perp$, por lo que $[F_1] \neq 0$, luego

$$\dim W/(W \cap W^\perp) = 1,$$

luego $\dim(W \cap W^\perp) = \dim W - 1$, luego

$$\dim W^\perp \geq \dim(W \cap W^\perp) = \dim W - 1,$$

pero $\dim W + \dim W^\perp = \dim V = N = 40$, luego $40 \geq \dim W + \dim W - 1$, de donde $\dim W \leq 41/2$, luego $\dim W \leq 20$. [En nuestro ejemplo es $N = 20$ y queda $\dim W \leq 21/2$, luego $\dim W \leq 10$ y puede comprobarse que, concretamente, la dimensión es 9.]

Así pues, las filas de M son linealmente dependientes. Como no hay más escalares que $0, 1$, una combinación lineal de filas de M es de la forma

$$\sum_{i \in Y} F_i = 0,$$

para cierto conjunto $Y \subset X$ no vacío, y que la combinación lineal sea nula equivale a que la matriz M tenga un número par de unos en cada columna $i \in Y$, lo cual a su vez equivale a que $|Y \cap B_j|$ sea par para todo $j = 1, \ldots, N$.

En nuestro ejemplo, $Y = \{1, 2, 5, 8, 10, 12, 15, 16\}$ da lugar a una combinación lineal nula, pues las filas correspondientes son:

M	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
5	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
8	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
10	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	1	0
12	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0
15	0	0	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	1
16	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1	0

y vemos que, en efecto, Y tiene exactamente 0, 2 o 4 elementos en cada uno de los 20 conjuntos B_j.

En general, como cada B_i tiene un número impar de elementos, resulta que $Y = X$ determina una combinación lineal nula. Esto se traduce en que si $Y \subset X$ determina una combinación lineal nula, lo mismo sucede con $X \setminus Y$.

Por otra parte, los conjuntos $Y = B_i \cup B_j$, para $1 \leq i < j \leq 4$ siempre dan combinaciones lineales nulas, pues, por ejemplo, en el caso de $Y = B_1 \cup B_2$, tenemos que

$$|(B_1 \cup B_2) \cap B_1| = |B_1| = n, \qquad |(B_1 \cup B_2) \cap B_2| = |B_2| = n,$$

$$|(B_1 \cap B_2) \cap B_3| = 0, \qquad |(B_1 \cap B_2) \cap B_4| = 0,$$

y, para cualquier bloque B_i, se cumple que $|(B_1 \cup B_2) \cap B_i| = 2$, ya que B_i contiene exactamente un elemento en el grupo B_1 y otro elemento en el grupo B_2 (cada bloque sólo puede tener un elemento en cada grupo y tiene tantos elementos como grupos hay, luego tiene exactamente un elemento en cada grupo).

La combinación lineal nula dada por $B_1 \cup B_2$ nos permite expresar una fila correspondiente a un $j \in B_2$ como combinación lineal de las filas de B_1 y las demás filas de B_2, luego podemos eliminar dicha fila y las filas restantes siguen generando W. Similarmente, la combinación lineal asociada a $B_1 \cup B_3$ nos permite eliminar una fila de B_3 y la asociada a $B_1 \cup B_4$ nos permite eliminar una fila de B_4. Con esto obtenemos un generador de W con $36-3 = 33$ vectores (13 en nuestro ejemplo), que todavía tienen que ser linealmente dependientes, luego todavía tiene que haber más combinaciones lineales nulas de los vectores restantes, lo que significa que existe un $Y \subset X$ que da lugar a una combinación lineal nula y que no contiene a todos los vectores de B_2, ni a todos los de B_3 ni a todos los de B_4. Esto implica que Y no es ninguno de los conjuntos (no vacíos) que ya hemos visto que dan combinaciones lineales nulas o, explícitamente, que no es de ninguna de estas formas:

- X,
- $B_i \cup B_j$, para $1 \leq i < j \leq 4$,
- $X \setminus (B_i \cup B_j)$, para $1 \leq i < j \leq 4$.

Ahora vamos a probar que si $Y \subset X$ cumple

$$\sum_{i \in Y} F_i = 0,$$

necesariamente es de uno de los tres tipos anteriores, con lo que tendremos una contradicción y habremos probado que no existen cuadrados grecolatinos de orden 6.

En nuestro ejemplo eso no es cierto, sino que ya hemos mostrado un conjunto $Y = \{1, 2, 5, 8, 10, 12, 15, 16\}$ que da lugar a una combinación lineal nula sin ser de ninguno de los tipos precedentes.

Sea, pues, $Y \subset X$ un conjunto no vacío que dé lugar a una combinación lineal nula de las filas de M y supongamos que no es de los tres tipos anteriores. Pongamos que $|Y| = m$. Como $|Y \cap B_j|$ tiene que ser par, sólo puede tomar los valores $0, 2, 4, 6$. Para $j = 0, 2, 4, 6$, sea b_j el número de índices i tales que $|Y \cap B_j|$ toma el valor j. (En nuestro ejemplo, los únicos valores posibles son $0, 2, 4$, y en el caso de $Y = \{1, 2, 5, 8, 10, 12, 15, 16\}$ tenemos $b_0 = 2, b_2 = 16, b_4 = 2$.)

Se tienen que cumplir las ecuaciones siguientes:

$$\begin{aligned} b_0 + b_2 + \ b_4 + \quad b_6 &= 40 \\ 2b_2 + 4b_4 + \ 6b_6 &= 7m \\ b_2 + 6b_4 + 15b_6 &= m(m-1)/2. \end{aligned}$$

En efecto, la primera sólo indica que cada índice j tiene que estar contado en un único b_j, luego la suma de los b_j es el conjunto total de índices. La segunda ecuación se debe a que, según hemos visto, cada $j \in X$ tiene que estar exactamente en $n+1 = 7$ bloques, por lo que al calcular $2b_2 + 4b_4 + 6b_6$ estamos contando 7 veces cada elemento de Y. La tercera ecuación se debe a que cada par de elementos de Y se encuentra exactamente en un bloque, pero si $|Y \cap B_j| = 2$, entonces B_j contiene 1 par de Y, si $|Y \cap B_j| = 4$, entonces B_j contiene 6 pares de elementos de Y, y si $|Y \cap B_j| = 6$, entonces B_j contiene 15 pares de elementos de Y, luego al sumar $b_2 + 6b_4 + 15b_6$ estamos calculando todos los pares de elementos de Y, que son $m(m-1)/2$.

Multiplicando por 2 la última ecuación y restándole la penúltima queda

$$8b_4 + 24b_6 = m^2 - m - 7m = m^2 - 8m,$$

de donde

$$b_4 + 3b_6 = \frac{m(m-8)}{8}.$$

Esto implica que $m \geq 8$ y que $m^2 \equiv m^2 - 8m \equiv 0 \pmod 8$, lo que sucede cuando $m \equiv 0, 4 \pmod 8$, luego $4 \mid m$. Reemplazando Y por $X \setminus Y$, podemos suponer además que $m \leq 12$ (si Y no es de los tres tipos indicados, $X \setminus Y$ tampoco lo es). Por lo tanto, si existe una combinación lineal nula que no sea de los tipos indicados, existe una con $m = 8, 12$ sumandos. Veamos ahora que si existe una con 12 sumandos, también hay otra con 8 sumandos.

Para ello consideramos todas las formas en las que los elementos de Y pueden repartirse entre los cuatro grupos B_1, B_2, B_3, B_4. Teniendo en cuenta que las intersecciones tienen que ser pares, las posibilidades son

$$\begin{array}{cccccc} a) & 6 & 6 & 0 & 0 \\ b) & 6 & 4 & 2 & 0 \\ c) & 6 & 2 & 2 & 2 \\ d) & 4 & 4 & 4 & 0 \\ e) & 4 & 4 & 2 & 2 \end{array}$$

En el caso a), es decir, que haya 6 elementos de Y en un grupo y 6 en otro, entonces Y es la unión de dos grupos, pero estamos suponiendo que Y no es de esa forma.

Si se diera el caso b) tendríamos una combinación lineal nula de la forma

$$(F_1 + F_2 + F_3 + F_4 + F_5 + F_6) + (F_7 + F_8 + F_9 + F_{10}) + (F_{13} + F_{14}) = 0,$$

donde las filas en cada paréntesis corresponden a un mismo grupo. Por otro lado, sabemos que

$$(F_1 + F_2 + F_3 + F_4 + F_5 + F_6) + (F_7 + F_8 + F_9 + F_{10} + F_{11} + F_{12}) = 0,$$

y al sumas ambas ecuaciones obtenemos

$$F_{11} + F_{12} + F_{13} + F_{14} = 0,$$

pero hemos visto que no puede haber combinaciones lineales nulas con 4 sumandos, luego este caso no puede darse.

Si se da el caso c) tenemos una combinación lineal nula de la forma

$$(F_1 + F_2 + F_3 + F_4 + F_5 + F_6) + (F_7 + F_8) + (F_{13} + F_{14}) + (F_{19} + F_{20}) = 0,$$

donde, como antes, cada paréntesis corresponde a elementos de Y del mismo grupo. Si sumamos la misma combinación lineal nula de antes, ahora obtenemos

$$(F_9 + F_{10} + F_{11} + F_{12}) + (F_{13} + F_{14}) + (F_{19} + F_{20}) = 0,$$

que es una combinación lineal nula con 8 sumandos.

Los casos d) y e) dan lugar igualmente a combinaciones lineales de 8 sumandos, luego basta probar que no existen tales combinaciones lineales.

Así pues, a partir de aquí suponemos que Y tiene $m = 8$ elementos, y vamos a llegar a una contradicción.

Tenemos entonces que $b_4 + 3b_6 = 0$, con lo que $b_4 = b_6 = 0$, luego $b_2 = 28$ y $b_0 = 12$. En resumen: el conjunto Y corta exactamente a 28 conjuntos B_i, y en todos los casos[1] $|Y \cap B_i| = 2$.

Como $Y \subset B_1 \cup B_2 \cup B_3 \cup B_4$ y, de los 8 elementos de Y, no puede haber más de 2 en cada B_i, necesariamente $|Y \cap B_i| = 2$ para $i = 1, 2, 3, 4$. En otras palabras, entre los 28 índices i tales que $Y \cap B_i \neq \varnothing$ se encuentran necesariamente los cuatro primeros.

Vamos a renombrar $Y = \{a, b, c, d, e, fg, h\}$, y $X \setminus Y = \{1, 2, \ldots, 16\}$, de modo que los cuatro grupos sean

$$B_1 = \{1, 2, 3, 4, a, b\}, \quad B_2 = \{5, 6, 7, 8, c, d\},$$

$$B_3 = \{9, 10, 11, 12, e, f\}, \quad B_4 = \{13, 14, 15, 16, g, h\}.$$

En nuestro ejemplo sería:

[1] Incidentalmente, esto prueba que el caso c) anterior no puede darse, porque daba lugar a una combinación lineal con cuatro sumandos en un mismo grupo, y lo mismo vale para el caso d).

M	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
a	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
b	1	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
c	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
d	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
e	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	1	0
f	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0
g	0	0	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	1
h	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1	0
1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
3	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0
4	0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0
5	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1
6	0	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0	1	0	0
7	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0
8	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0

Hasta ahora hemos trabajado con $P = (X, \{B_1, \ldots, B_{40}\})$ y ahora vamos a considerar también $Q = (X \setminus Y, \{B'_1, \ldots, B'_{40}\})$, donde $B'_i = B_i \setminus Y$.

En P tenemos que B_1, B_2, B_3, B_4 tienen seis elementos, pero B'_1, B'_2, B'_3, B'_4 pasan a tener cuatro. De los 36 bloques restantes, $B_5, \ldots, B_{40}$, que tienen cuatro elementos, hay 24 a los que les quitamos dos elementos y los B'_i correspondientes pasan a tener 2, mientras que los 12 restantes conservan todos sus elementos. Así pues, hay 16 bloques B'_i con cuatro elementos y 24 con dos elementos.

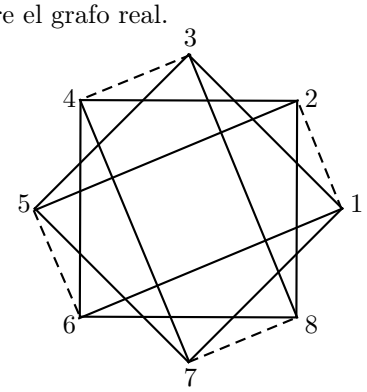
En nuestro ejemplo tenemos 16 bloques de dos elementos, 2 de cuatro y 2 vacíos:

M	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
3	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0
4	0	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0
5	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1
6	0	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0	1	0	0
7	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0
8	0	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0
	2	2	2	2	2	2	2	0	0	2	2	2	2	2	4	2	2	4	2	2

Ahora construimos el grafo[2] G que tiene por vértices los 16 elementos de $X \setminus Y$ como aristas los pares de vértices que forman uno de los bloques B'_i de dos elementos. El resultado en nuestro ejemplo (con 8 vértices en lugar de 16) es el que muestra la figura siguiente, en la que hemos marcado con trazo

[2]Un *grafo* es simplemente un par (X, A), donde X es un conjunto de *vértices* y A es un conjunto de *aristas,* es decir, de pares $\{v_1, v_2\}$ de elementos (distintos) de X.

punteado las aristas correspondientes a los bloques B'_1, B'_2, B'_3, B'_4, que no se corresponden con la situación en el caso real, porque en éste los cuatro primeros bloques tienen 4 elementos en lugar de 2. Vamos a ver que, si prescindimos de esas tres aristas, el grafo de nuestro ejemplo tiene las mismas características que vamos a demostrar sobre el grafo real.



Por ejemplo, en primer lugar vamos a probar que:

> *De cada vértice de G salen 3 aristas, y cada una se dirige a un elemento de cada uno de los grupos en los que no está el vértice.*

En efecto, si $i \in X \setminus Y$ aparece en x bloques B'_j con 2 elementos y en y bloques con 4 elementos, entonces $x + y = 7$, pues sabemos que cada elemento de X está exactamente en 7 conjuntos B_i.

Por otra parte, cada uno de los x bloques de cardinal 2 al que pertenece i contiene un único par con i, mientras que cada uno de los y bloques de cardinal 2 al que pertenece i contiene 3 pares con i y, como todo par de elementos de $X \setminus Y$ está en un único bloque B'_i, concluimos que $x + 3y$ es el número total de pares de elementos de $X \setminus Y$ que contienen a i, es decir, que tenemos las ecuaciones

$$x + y = 7, \qquad x + 3y = 15.$$

Resolviendo llegamos a que $x = 3$, $y = 4$. Por lo tanto, de i salen exactamente 3 aristas.

Los 3 bloques de Q de cardinal 2 que contienen a i proceden de 3 bloques de P de cardinal 4 a los que hemos quitado 2 elementos de Y, y estos 6 elementos de Y que hemos quitado a los 3 bloques son necesariamente los 6 elementos de Y que no están en el grupo de i. Pongamos que los bloques son:

$$\{u_1, u_2, j_1, i\}, \qquad \{u_3, u_4, j_2, i\}, \qquad \{u_5, u_6, j_3, i\},$$

donde $u_i \in Y$, $j_i \in X \setminus Y$. Pongamos que $i \in B_4$. Entonces, si j_1, j_2 es tuvieran en el mismo grupo, pongamos B_3, sería $u_1 \in B_1$, $u_2 \in B_2$, pero también $u_3 \in B_1$, $u_4 \in B_2$, lo que obligaría a que $u_5, u_6 \in B_3$, lo cual es imposible, porque un bloque no puede tener dos elementos del mismo grupo. Por eso j_1, j_2, j_3 tienen que estar en grupos distintos.

En G no hay triángulos, es decir, no hay tres vértices distintos i, j, k unidos por tres aristas.

Supongamos que las aristas forman un triángulo, por ejemplo, con vértices en los grupos $i \in B_1, j \in B_2, k \in B_3$. Entonces $\{i, j\}$ tiene que aparecer en un bloque junto con uno de los dos elementos de $B_4 \cap Y = \{g, h\}$ (pongamos g) e igualmente $\{i, k\}$ tiene que aparecer en otro bloque con otro de los dos elementos de $B_4 \cap Y$. No puede ser g, porque entonces el par $\{i, g\}$ aparecería en dos bloques. Así pues, hay un bloque que contiene a $\{i, j, g\}$ y otro que contiene a $\{i, k, h\}$, pero tendría que haber un tercer bloque que contuviera a $\{j, k\}$ y un elemento de $B_4 \cap Y = \{g, h\}$, pero entonces, o bien $\{i, g\}$ o bien $\{k, h\}$ estaría en dos bloques distintos, lo cual es imposible.

Recordemos que los grupos de P son

$$B_1 = \{1, 2, 3, 4, a, b\}, \quad B_2 = \{5, 6, 7, 8, c, d\},$$

$$B_3 = \{9, 10, 11, 12, e, f\}, \quad B_4 = \{13, 14, 15, 16, g, h\},$$

de modo que cada bloque B_i con $i > 4$ consta de un elemento de cada uno de estos grupos, que pueden ser los cuatro números o dos números y dos letras.

Veamos ahora que es imposible que los tres vértices conectados con un mismo vértice formen parte de un mismo bloque. Eligiendo la numeración de los vértices, no perdemos generalidad si suponemos que el vértice 1 está conectado con los vértices $5, 9, 13$ y vamos a ver que no puede ser que un bloque de P contenga a $\{5, 9, 13\}$. Puesto que $\{1, 5\}$ está en un bloque de Q con 2 elementos, es decir, en un bloque de P que contiene 2 elementos de Y, dicho bloque no puede ser el que contiene a $\{5, 9, 13\}$, luego el elemento del grupo 1 de este bloque no puede ser 1, ya que entonces el par $\{1, 5\}$ aparecería en dos bloques. No perdemos generalidad si llamamos 2 al elemento que falta en el bloque, de modo que éste es $\{2, 5, 9, 13\}$.

Entonces, el par $\{2, 5\}$ no es una arista de G, de modo que el vértice del segundo grupo conectado con 2 no puede ser 5, sino que tiene que estar entre $6, 7, 8$. No perdemos generalidad si llamamos concretamente 8 al vértice conectado con 2.

Como $\{1, 6\}, \{1, 7\}, \{1, 8\}$ no son aristas de G, estos pares tienen que estar contenidos en bloques de Q de cuatro elementos. El bloque que contiene a $\{1, 6\}$ no puede contener a 9, porque $\{1, 9\}$ está en un bloque con dos elementos de Y, y por el mismo motivo no puede contener al 13. No perdemos generalidad si llamamos 10 y 14 a los elementos de dicho bloque de los grupos 3 y 4, de modo que el bloque es $\{1, 6, 10, 14\}$. Similarmente, el bloque que contiene a $\{1, 7\}$ tiene que contener un elemento del grupo 3 que no sea $9, 10$ y un elemento del grupo 4 que no sea $13, 14$, luego no perdemos generalidad si numeramos estos elementos como $11, 15$, de modo que el bloque es $\{1, 7, 11, 15\}$. Esto determina el bloque que debe contener a $\{1, 8\}$. En total, tenemos los bloques

$$\{1, 6, 10, 14\}, \quad \{1, 7, 11, 15\}, \quad \{1, 8, 12, 16\}.$$

Por construcción tenemos que $\{2,8\}$ es una arista de G, luego $\{2,5\}$, $\{2,6\}$ y $\{2,7\}$ tienen que estar en bloques de P sin elementos de Y. El de $\{2,5\}$ ya lo tenemos. El de $\{2,6\}$ no puede contener ni a 9 ni a 10, porque los pares $\{2,9\}$ y $\{6,10\}$ ya están en otros bloques, luego tenemos las posibilidades

$$\{2,5,9,13\},\quad \{2,6,\frac{11}{12},\frac{16}{15}\},\quad \{2,7,\frac{12}{10},\frac{14}{16}\},$$

donde las "fracciones" indican posibilidades alternativas. Si se da la alternativa superior, vemos que 2 no está conectado en G con $11,12,14,16$, luego los vértices conectados con 2 son $8,10,15$, mientras que si se da la alternativa inferior son $8,11,14$.

De los 16 bloques de Q de 4 elementos, hay 4 correspondientes a los grupos, y otros 12 que tienen un elemento de cada grupo. De estos 12, ya tenemos determinados 6 (salvo dos alternativas), y faltan otros 6 bloques. Los bloques que hemos encontrado tienen ya a todos los pares $\{1,i\}$ menos $\{1,5\}$ y a todos los pares $\{2,i\}$ menos $\{2,8\}$ (pero los dos que faltan tienen que estar en bloques de 2 elementos), luego los 6 bloques que faltan tienen que tener todos un 3 o un 4 y tienen que contener los pares

$$\{3,5\},\quad \{3,9\},\quad \{3,13\},\quad \{4,5\},\quad \{4,9\},\quad \{4,13\},$$

pues, como $5,9,13$ están conectados con 1 en el grafo G, no pueden estar conectados ni con 3 ni con 4, luego los pares correspondientes tienen que aparecer todos en bloques de Q de 4 elementos. Así pues, los 6 bloques que nos faltan tienen que extender estos 6 pares.

Por otra parte, como 2 está conectado en G con $8,10,15$ (o bien $8,11,14$), estos tres vértices no están conectados entre sí (porque G no contiene triángulos), con lo que los 6 bloques que nos faltan tienen que contener los pares $\{8,10\},\{8,15\},\{10,15\}$ (o bien $\{8,11\},\{8,14\},\{11,14\}$).

En el primer caso tres de los bloques tienen que ser

$$\{\frac{3}{4},8,10,13\},\quad \{\frac{4}{3},8,9,15\},\quad \{?,5,10,15\},$$

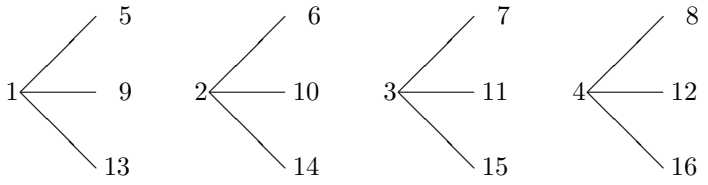
porque si un bloque contiene a $\{8,10\}$, de los seis pares con $3/4$ que hemos enumerado antes, tiene que contener a $\{3,13\}$ o a $\{4,13\}$, si contiene a $\{8,15\}$, tiene que contener a $\{3,9\}$ o a $\{4,9\}$ y si contiene a $\{10,15\}$, tiene que contener a $\{3,5\}$ o a $\{4,5\}$, pero tanto si completamos el ? con un 3 o un 4 se nos repite un par, luego no hay solución posible. Todo el razonamiento vale igualmente si cambiamos el 10 por el 11 y el 15 por el 14.

Así pues, hemos probado lo siguiente:

> $(*)$ *Ningún bloque de P contiene los tres vértices conectados en G con un vértice dado.*

Empezando de nuevo teniendo esto en cuenta, podemos suponer que los tres vértices conectados con el vértice 1 son $5,9,13$, los vértices conectados con 2

no pueden ser ningunos de estos tres (porque dicho vértice estaría conectado a la vez con 1 y con 2, del mismo grupo, lo cual es imposible), luego podemos suponer que son $6, 10, 14$, igualmente, los vértices conectados con 3 no pueden ser ningunos de los anteriores, luego podemos suponer que son $7, 11, 15$ y así los vértices conectados con 4 tienen que ser $8, 12, 16$. Gráficamente, tenemos estos fragmentos del grafo G:



Como G no contiene triángulos, los vértices $6, 10, 14$ no están conectados entre sí, luego los pares $\{6, 10\}$, $\{6, 14\}$, $\{10, 14\}$ están contenidos en bloques de Q de cuatro elementos, y tienen que ser tres bloques distintos por $(*)$. Esos tres bloques no contendrán al 2, pues los pares $\{2, 6\}$, etc. forman bloques de 2 elementos, luego uno contendrá al 1, otro al 3 y otro al 5. Renumerando los grupos podemos suponer que $\{1, 6, 10\}$ está en un bloque. Éste no puede completarse con 13 porque 1 está conectado con 13, ni con 14 por $(*)$, luego tiene que ser

$$\{1, 6, 10, \frac{15}{16}\}.$$

Intercambiando si es preciso los vértices 3 y 4, podemos suponer que el bloque es, concretamente, $\{1, 6, 10, 15\}$.

Similarmente, uno de los pares $\{7, 11\}$, $\{7, 15\}$, $\{11, 15\}$ tiene que estar en un bloque con el 1, pero no puede ser que $\{1, 7, 15\}$ o $\{1, 11, 15\}$ estén en un bloque, ya que el par $\{1, 15\}$ ya está en $\{1, 6, 10, 15\}$. Por lo tanto, $\{1, 7, 11\}$ tiene que estar en un bloque, que no puede completarse con 13 (porque 13 está contectado con 1) ni con 15 por $(*)$, ya que $7, 11, 15$ están conectados con 3, luego es

$$\{1, 7, 11, \frac{14}{16}\}.$$

El par $\{1, 8\}$ tiene que aparecer en un bloque de Q de cuatro elementos, porque 1 y 8 no están conectados, el cual no puede contener ni el 9 (porque está conectado con 1) ni el 10 ni el 11, porque aparecen en los dos bloques anteriores, ni el 13 (porque está conectado con 1), ni el 15 (porque $\{1, 15\}$ aparece en el primer bloque), luego tenemos tres bloques

$$\{1, 6, 10, 15\}, \qquad \{1, 7, 11, \frac{14}{16}\}, \qquad \{1, 8, 12, \frac{16}{14}\},$$

pero el 16 no puede aparecer en el último bloque por $(*)$, luego tiene que ser

$$\{1, 6, 10, 15\}, \qquad \{1, 7, 11, 16\}, \qquad \{1, 8, 12, 14\}.$$

Como 5 y 9 no están conectados en G (porque ambos están conectados con 1), tienen que estar en un bloque de Q de cuatro elementos, que no puede contener al 1 y, por $(*)$, tampoco al 13. Distinguimos cuatro casos:

Si dicho bloque contiene al 2, no puede contener al 14 (conectado con 2), luego tendremos

$$\{2,5,9,\frac{15}{16}\},\qquad \{2,7,12,\frac{13}{15}\},$$

pues $\{2,7\}$ tiene que estar en un bloque de Q de cuatro elementos, y no puede contener al 9 (porque $\{2,9\}$ ya está en el primer bloque), ni al 10 (conectado con 2) ni al 11 por el bloque $\{1,7,11,16\}$, ni al 14 (conectado con 2) ni al 16 (por el bloque $\{1,7,11,16\}$).

Pero 2 tiene que ir en un bloque con uno de los pares $\{8,12\},\{12,16\},\{8,16\}$, y el segundo bloque que hemos obtenido descarta las dos primeras posibilidades, luego tiene que ser

$$\{2,5,9,15\},\qquad \{2,7,12,13\},\qquad \{2,8,?,16\},$$

y el ? no puede sustituirse por ningún número, pues 9 y 12 están en los bloques anteriores, 10 está conectado con 2 y 11 no puede ser por el bloque $\{1,7,11,16\}$.

Similarmente, si el bloque de $5,9$ contiene al 3, tiene que ser

$$\{3,5,9,\frac{14}{16}\},\qquad \{3,8,10,\frac{16}{13}\},$$

donde descartamos el 9 por el primer bloque, el 11 porque está conectado con el 3, el 12 por el bloque $\{1,8,12,14\}$, el 15 porque está conectado con el 3 y el 14 por el bloque $\{1,8,12,14\}$.

Pero 3 tiene que ir en un bloque con uno de los pares $\{6,10\},\{6,14\},\{10,14\}$, y el segundo bloque que hemos obtenido descarta las dos primeras posibilidades, luego tiene que ser

$$\{3,5,9,16\},\qquad \{3,8,10,13\},\qquad \{3,?,10,14\},$$

pero el ? no tiene solución, ya que 5 y 8 están en los bloques anteriores, 7 está conectado con 3 y 6 no puede ser por el bloque $\{1,6,10,15\}$.

Falta considerar la posibilidad de que el bloque de $\{5,9\}$ contenga al 4. Entonces tenemos

$$\{4,5,9,\frac{14}{15}\},\qquad \{4,6,11,\frac{13}{14}\},$$

donde descartamos el 12 porque está conectado con el 4, el 9 por el primer bloque, el 10 por el bloque $\{1,6,10,15\}$, el 16 porque está conectado con el 4 y el 15 por el bloque $\{1,6,10,15\}$.

Ahora 4 tiene que ir en un bloque con uno de los pares $\{7,11\},\{11,15\},\{7,15\}$, y el segundo bloque que hemos obtenido descarta las dos primeras posibilidades, luego tiene que ser

$$\{4,5,9,14\},\qquad \{4,6,11,13\},\qquad \{4,7,?,15\}$$

y en el lugar del ? no pueden ir $9,11$ por los bloques anteriores, ni 12 porque está conectado con 4 ni 10 por el bloque $\{1,6,10,15\}$. ∎

Índice de Materias