Carlos Ivorra Castillo

ÁLGEBRA

Mathematics, rightly viewed, posseses not only truth, but supreme beauty —a beauty cold and austere, like that of sculpture.

Bertrand Russell

Índice General

Introducción

El propósito de este libro es introducir a un lector con conocimientos mínimos de matemáticas en el estudio de los números naturales

$$0, \quad 1, \quad 2, \quad 3, \quad 4, \quad 5, \quad \ldots$$

Quizá esta afirmación sorprenda al lector por dos posibles motivos: bien porque crea que los números naturales son algo tan simple que difícilmente se puede escribir un libro sobre ellos, bien porque crea que un libro así no debería llamarse 'Álgebra'. El primer caso es fácil de rectificar. Consideremos por ejemplo la ecuación

$$x^2 + xy - 3y^2 = 15.$$

¿Sabría decidir el lector si existen números naturales (x, y) que satisfagan esta condición? Tenemos aquí un problema de planteamiento elemental cuya solución no es nada fácil. Si existiera un par así podríamos tener suerte y encontrarlo por tanteo, pero si no lo hay necesitaremos algún tipo de razonamiento que lo justifique, pues el no encontrar soluciones no significa que no las haya. Si el problema fuera $x^2 + xy + 3y^2 = 15$ el asunto sería muy diferente, pues podríamos hacer $4(x^2 + xy + 3y^2) = (2x + y)^2 + 11y^2$ y de aquí sacaríamos una cota a las posibles soluciones, con lo que un número finito de comprobaciones bastaría para decidir si las hay. Aun así habríamos necesitado un pequeño truco que requeriría un mínimo de perspicacia.

De nada sirve despejar la y en función de x, o viceversa, pues entonces nos encontraremos con el problema de determinar si una expresión con una raíz cuadrada puede o no ser un número natural, y no podremos ir mucho más lejos.

Sin duda el lector que creía dominar los números naturales reconocerá ya la precariedad de ese dominio. Sin embargo esta situación suele causar rechazo al matemático acostumbrado a otra clase de problemas más ... ¿abstractos? La reacción natural es: ¿pero qué importa si existen o no soluciones naturales? Una pregunta interesante podría ser si existen funciones reales continuas no derivables en ningún punto, por ejemplo, porque una solución negativa consolidaría nuestro conocimiento de la continuidad y la derivabilidad, mientras que una solución positiva sería (y de hecho es) algo verdaderamente curioso e intrigante. Sin embargo, tanto si alguien encuentra una solución a esa ecuación como si prueba que no las hay, lo cierto es que nos quedamos igual, obtenemos un dato irrelevante.

Esta objeción entronca con la posible sorpresa de que un libro que promete abordar estas banalidades tenga la osadía de titularse 'Álgebra'. El reproche estaría justificado si lo único que fuéramos a ver en este libro fuera una colección de recetas o, aún peor, de trucos para resolver ecuaciones como la de antes. También en tal caso sería razonable opinar que el contenido del libro sería irrelevante, al menos según los gustos matemáticos al uso. Sin embargo, el interés de un problema puede no estar en la pregunta sino en la respuesta. Parafraseamos a Gauss al decir que la aridez de esta clase de problemas oculta una disciplina que merece el título de Reina de las Matemáticas. ¿Por qué un matemático que destacó tan prodigiosamente en análisis, geometría diferencial, física y estadística, entre otras partes de la matemática, anteponía la teoría de números a todas ellas? Sencillamente porque al abordar problemas como el que hemos propuesto se encontró con una teoría mucho más rica, sutil y abstracta que cualquier otra de su época.

Ciertamente, la teoría de números antes de Gauss era esencialmente una colección de trucos, verdaderos monumentos al ingenio humano, eso sí, pero despreciables al gusto del matemático moderno, pero estamos hablando de la teoría de números del siglo XVIII. Para los matemáticos del siglo XIX la situación era radicalmente distinta, y es esta visión moderna la que queremos transmitir al lector de este libro. Básicamente se puede describir como sigue:

Los números naturales son unos objetos extremadamente caprichosos, pero no caóticos. Es como si un pianista decide caprichosamente qué pieza va a tocar. A priori no podemos predecir lo que hará, pero una vez conocemos su decisión podemos anticipar cada uno de sus movimientos a partir de la partitura. Un pianista caótico sería por ejemplo un intérprete de Jazz que improvisara en todo momento. Así, el comportamiento de los números puede ser controlado en función de ciertos parámetros, caprichosos hasta donde hoy se sabe, y la forma de controlarlos no es la fuerza bruta (la manipulación de ecuaciones al estilo del siglo XVIII), que ofrece resultados muy limitados, sino la psicología más fina, la búsqueda de leyes generales que sólo pueden ser expresadas en términos de objetos abstractos, impensables en una primera aproximación, pero que los matemáticos han podido descubrir poco a poco a lo largo de casi dos siglos.

Pensemos por ejemplo en la introducción de los números enteros:

$$\ldots \quad -5, \quad -4, \quad -3, \quad -2, \quad -1, \quad 0, \quad 1, \quad 2, \quad 3, \quad 4, \quad 5, \quad \ldots$$

Se trata del ejemplo más elemental de cómo un artificio algebraico como es poner un signo delante de los números resulta ser de inestimable ayuda en su manejo. Tanto es así que en realidad, aunque la motivación primera en el estudio de los números proviene de los números naturales, es más justo decir que en este libro se estudian los números enteros.

Pero si queremos resolver el problema que hemos planteado necesitamos ir mucho más lejos. El paso siguiente en esta dirección es factorizar la ecuación

$$x^2 + xy - 3y^2 = \left(x + y\,\tfrac{1+\sqrt{13}}{2}\right)\left(x + y\,\tfrac{1-\sqrt{13}}{2}\right).$$

Esto puede parecer un sucio 'truco', pero en realidad es un paso obvio si se disfruta del punto de vista adecuado. Así nos encontramos con que la ecuación

está relacionada con el número $D = 13$, invisible hasta ahora, y que se conoce como discriminante de la ecuación. Por ejemplo, si antes decíamos que el problema con -3 es más difícil que el mismo problema pero con un $+3$, un algebrista verá el discriminante $D = 13$ frente al discriminante $D = -11$, y el algebrista sabe que una de las características generales de este tipo de ecuaciones es que las de discriminante negativo siempre son más fáciles. Vemos así que el verdadero problema no era el signo del -3, sino el del discriminante.

Además nos ha aparecido el número irracional $\frac{1+\sqrt{13}}{2}$, y en este punto el algebrista deja de pensar en números para fijarse en algo mucho más abstracto, como es el conjunto

$$\mathbb{Z}\Big[\tfrac{1+\sqrt{13}}{2}\Big] = \Big\{x + y\,\tfrac{1+\sqrt{13}}{2} \mid x, y \in \mathbb{Z}\Big\}.$$

Él sabe que este conjunto tiene una estructura importante conocida como dominio íntegro, que lo hace muy similar al propio conjunto $\mathbb{Z}$ de los números enteros. Sobre este conjunto está definida una aplicación llamada norma

$$\mathrm{N} : \mathbb{Z}\Big[\tfrac{1+\sqrt{13}}{2}\Big] \longrightarrow \mathbb{Z},$$

dada por $\mathrm{N}\left(x + y\,\frac{1+\sqrt{13}}{2}\right) = \left(x + y\,\frac{1+\sqrt{13}}{2}\right)\left(x + y\,\frac{1-\sqrt{13}}{2}\right) = x^2 + xy - 3y^2$ y cuyo comportamiento es extremadamente regular.

El resultado es que la penetración del algebrista convierte un arduo problema que todo el mundo entiende en un sencillo problema que sólo los algebristas entienden: ¿Existe un entero cuadrático en $\mathbb{Q}(\frac{1+\sqrt{13}}{2})$ cuya norma sea 15?

Decimos 'sencillo' pensando, por supuesto, en el punto de vista del algebrista que cuenta con el equipaje de una sólida y elegante teoría. Para él la solución se obtiene analizando unos objetos todavía más abstractos y alejados de la simple ecuación dada: los ideales del anillo anterior. No importa si el lector no sabe en este punto de qué estamos hablando (eso es lo que puede aprender en este libro, precisamente), lo que importa es que esos ideales siguen un comportamiento extremadamente simple, de modo que una comprobación elemental le permite concluir la inexistencia de soluciones enteras. Citamos aquí la comprobación sin ánimo de que el lector la entienda, sólo para que admire su sencillez formal:

Tenemos que $15 = 3 \cdot 5$ y si existiera un ideal de norma 15 éste tendría un factor primo de norma 5, pero eso significaría que el discriminante 13 sería un resto cuadrático módulo 5, pero $(13/5) = (3/5) = (5/3) = (2/3) = -1$, contradicción.

Todo esto puede ser razonado sin esfuerzo incluso mentalmente. El lector encontrará los detalles en el capítulo XIV.

En realidad este problema era muy fácil. Si el término independiente de la ecuación no hubiera sido 15, sino otro número, como 17, entonces la solución habría sido positiva, y para justificarlo el algebrista habría tenido que contar con dos datos más, todavía más abstractos:

1) El número de clases de $\mathbb{Q}\Big(\frac{1+\sqrt{13}}{2}\Big)$ es $h = 1$,

2) El cuerpo $\mathbb{Q}\Big(\frac{1+\sqrt{13}}{2}\Big)$ contiene unidades de norma negativa.

Una vez más, no esperamos que el lector entienda nada de esto. La segunda propiedad es fácil de comprobar con un mínimo tanteo, mientras que la primera es un hecho nada trivial y que ejemplifica lo que antes llamábamos comportamiento 'caprichoso' de los números. En efecto, cada cuerpo como $\mathbb{Q}\left(\frac{1+\sqrt{13}}{2}\right)$ tiene asociado un número natural h llamado su 'número de clases', que se puede calcular en la práctica mediante un algoritmo.

¿Por qué el número de clases para $\sqrt{13}$ es $h = 1$ mientras que, por ejemplo, para $\sqrt{15}$ es $h = 2$? Esto forma parte del comportamiento caprichoso de los números del que hablábamos antes, pero lo cierto es que, una vez determinado el número de clases, el algebrista sabe cuál es el 'carácter' que este capricho imprime a los problemas asociados a este número, y sabe a qué atenerse.

No creemos necesario aburrir al lector con más afirmaciones que probablemente no entienda. Éstas habrán bastado para que comprenda la situación. Los problemas numéricos como el que hemos presentado abren la puerta, a la vez que dan sentido y motivación, a una teoría cuyo 'sabor' ha podido captar hace un momento, una teoría profunda, rica en conceptos y en ideas y que nos permite llegar a elegantes principios generales más simples formalmente cuanto más elevados y complejos conceptualmente.

Se trata de una situación similar a la de la mecánica celeste: el movimiento de los planetas puede ser descrito eficientemente por las leyes ptolemaicas, meramente descriptivas y aproximadas, o por las leyes de Kepler, rigurosas pero técnicas, o por la ley de la gravitación universal de Newton, la más simple formalmente, o por las ecuaciones de la relatividad general de Einstein, las más sofisticadas de todas, pero las que proporcionan una mejor comprensión del fenómeno.

El estudio de los números enteros se conoce en general como Teoría de Números, y la teoría que hay detrás es tan vasta que no encaja en ninguna rama particular de las matemáticas, sino que en ella intervienen el álgebra, la topología, el análisis e incluso la geometría. Por ello, y a pesar de que fraccionarla no deja de ser artificial, se habla de una Teoría de Números Elemental (que no usa más que la aritmética básica), una Teoría Algebraica de Números, una Teoría Analítica de números y una Geometría de los Números. (No obstante las fronteras no pueden establecerse con precisión, y por eso se ha terminado hablando de una Teoría Algebraica de Números Analítica).

El ejemplo que hemos dado corresponde a la Teoría Algebraica de Números (al igual que el contenido de este libro). Quizá después de todo podría tener razón el lector que considerara que 'Álgebra' no es el título adecuado de este libro, sino que sería mejor haberlo llamado 'Teoría Algebraica de Números'. Sin embargo hemos decidido darle el título que tiene porque al fin y al cabo abordamos a un nivel aceptable como introducción el equivalente a un primer curso de álgebra: álgebra lineal (módulos, espacios vectoriales, matrices, determinantes), teoría de anillos, teoría de cuerpos y teoría de grupos finitos (con especial hincapié en los grupos abelianos y resolubles y los grupos de permutaciones), y no creemos que la palabra 'Álgebra' deba significar otra cosa más especializada. Más bien los libros que se ocupan de resultados algebraicos tan

abstractos que ya no tienen nada que ver con los números (cuyo valor e interés nadie pone en duda) deberían llamarse 'Álgebra abstracta' (como de hecho algunos lo hacen), y un libro de Teoría Algebraica de Números sería algo más especializado y sistemático. Preferimos pensar, pues, que éste es un libro de álgebra con ilustraciones de teoría de números, encaminado a dotar al lector de una base algebraica suficiente para un estudio posterior de la Teoría Algebraica de Números propiamente dicha.

El criterio general en la redacción ha sido usar los números como hilo conductor e ir introduciendo progresivamente los conceptos algebraicos necesarios a un nivel lo suficientemente general como para que el lector termine con un conocimiento sólido del álgebra elemental, pero nunca hasta el punto (esperamos) de que las ideas resulten oscurecidas por los conceptos formales.

La restricción principal ha sido que a todos los efectos no existen los números reales. No hemos demostrado ningún resultado que requiera el uso de números reales ni se da ninguna interpretación geométrica o aplicación a la geometría de los conceptos algebraicos. La razón de esta restricción es que, en primer lugar, los números reales no han resultado necesarios en ningún momento y, en segundo lugar, que consideramos que la introducción más razonable de los números reales es una introducción geométrica y no algebraica ni analítica, por lo que no es éste el libro adecuado para presentarlos.

La única laguna importante que este criterio ha ocasionado es que no hemos hablado de equivalencia y semejanza de matrices, vectores propios, polinomios característicos, etc., pues estos conceptos tienen una interpretación geométrica importante y sería absurdo introducirlos sin ella. También puede echarse en falta la teoría de Sylow, de la que no hemos hablado porque su indiscutible utilidad en el estudio de los números sólo se pone de manifiesto en estados más avanzados de la teoría, y por lo tanto hubiera sido forzado mostrar alguna aplicación más allá de la propia teoría de grupos. Hemos incluido tres apéndices con algunos resultados cuyo interés no puede comprenderse plenamente sin conocer el desarrollo posterior de la teoría, pero que de todos modos pueden ser ilustrativos porque son una prolongación natural de la teoría elemental.

El orden de exposición pretende combinar la naturalidad, en el sentido de que cada concepto aparezca en el momento en que resulta necesario, con el mínimo orden preciso para una correcta asimilación por parte del lector. Esto hace que algunos resultados puedan estar en capítulos donde en principio no se esperaría encontrarlos. Piénsese que éste no es un libro de consulta, sino un libro para ser leído desde el principio hasta el final, un libro donde no se pretende que esté 'todo' sino sólo lo necesario para que no haya paja que saltar.

Esperamos sinceramente que el lector disfrute, si no con la forma de este libro, de la que somos responsables, sí con su contenido, que ha cautivado a tantos matemáticos.

Preliminares conjuntistas

Citamos aquí brevemente los resultados con los que el lector debería conocer para entender este libro. De todos modos, salvo en muy contadas ocasiones todos los requisitos pueden suplirse con un poco de sentido común (o intuición, como suele decirse). Por ello el único requisito real es estar familiarizado con el lenguaje y el razonamiento matemático.

Suponemos que el lector conoce el lenguaje de la teoría de conjuntos elemental: conjuntos, subconjuntos, unión, intersección, producto cartesiano, aplicaciones, etc. Sólo hay un punto a destacar a este respecto, y es que en este libro adoptaremos siempre el convenio de que en una composición de aplicaciones actúa primero la aplicación de la izquierda, esto es, $(f \circ g)(x) = g\big(f(x)\big)$. Consideramos que, a la larga, este convenio resulta mucho más natural que el contrario.

Necesitaremos también algo de teoría de cardinales, aunque normalmente todos los cardinales que nos aparecerán serán finitos. Si el lector decide ignorar toda alusión a cardinales infinitos se perderá una mínima parte del contenido de este libro. Baste, pues, saber que el cardinal de un conjunto es, al menos si el conjunto es finito, lo que usualmente se entiende por su 'número de elementos', y que dos conjuntos tienen el mismo cardinal si y sólo si se puede establecer una aplicación biyectiva entre ellos. El cardinal de un conjunto X es menor o igual que el cardinal de un conjunto Y si y sólo si existe una aplicación inyectiva de X en Y.

Un hecho elemental de uso muy frecuente es que si un conjunto finito X tiene el mismo cardinal que un subconjunto Y, entonces $X = Y$. Más en general: una aplicación entre dos conjuntos finitos del mismo cardinal es biyectiva si y sólo si es inyectiva si y sólo si es suprayectiva.

Respecto a la aritmética cardinal usaremos a menudo que si un conjunto está dividido en subconjuntos disjuntos, entonces su cardinal es la suma de los cardinales de sus partes, y si todas ellas tienen el mismo cardinal, entonces el cardinal del conjunto total es el de una de sus partes multiplicado por el número de partes. Así mismo, el cardinal de un producto cartesiano es el producto de los cardinales de los factores.

El único punto donde necesitaremos algún resultado adicional es en la prueba de la equicardinalidad de bases (capítulo VII). Allí usaremos que si X es un conjunto infinito, entonces el número de subconjuntos finitos de X coincide con

el cardinal de X, y que si X está dividido en conjuntos finitos, entonces el cardinal de X coincide con el número de partes.

Mención especial requiere el Lema de Zorn, que nos aparecerá en pocas pero importantes ocasiones. Recordemos las definiciones que intervienen en su enunciado:

Un conjunto X está *parcialmente ordenado* por una relación $\leq$ si se cumple:

1. $x \leq x$ para todo $x \in X$.
2. Si $x \leq y$ e $y \leq x$, entonces $x = y$, para todo $x, y, \in X$.
3. Si $x \leq y$ e $y \leq z$, entonces $x \leq z$, para todo $x, y, z \in X$.

El ejemplo típico de orden parcial en una familia de conjuntos es el dado por la inclusión, es decir, $x \leq y$ si y sólo si $x \subset y$.

Un subconjunto Y de un conjunto parcialmente ordenado X es una *cadena* si cualquier par de elementos $x, y \in Y$ cumple $x \leq y$ o $y \leq x$.

Un elemento x de un conjunto parcialmente ordenado X es una *cota superior* de un conjunto $Y \subset X$ si para todo $y \in Y$ se cumple $y \leq x$.

Si X es un conjunto parcialmente ordenado, un *maximal* de X es un elemento $x \in X$ tal que no existe ningún $y \in X$ que cumpla $x \leq y$, $x \neq y$.

Lema de Zorn *Si X es un conjunto parcialmente ordenado no vacío en el que toda cadena tiene una cota superior, entonces X tiene un elemento maximal.*

En la práctica, el lema de Zorn se aplica a conjuntos ordenados por la inclusión, y la forma típica de probar que una cadena tiene cota superior es probar que la unión de todos sus elementos es también un elemento del conjunto, lo cual es siempre un ejercicio sencillo. El resultado es entonces la existencia de un miembro de la familia que no está contenido en ningún otro. Todas las verificaciones concretas de las hipótesis del lema de Zorn en este libro se dejan como un sencillo ejercicio para el lector.

Hay un teorema que puede probarse mediante el lema de Zorn pero que hemos preferido probar de otro modo para evitar tecnicismos conjuntistas demasiado prolijos. Se trata de la existencia de clausura algebraica (capítulo VIII) En su lugar usaremos un resultado equivalente al lema de Zorn, y es el principio de buena ordenación de Zermelo:

Principio de buena ordenación *Todo conjunto admite un buen orden, esto es, una relación de orden total en la que todo subconjunto no vacío tiene un mínimo elemento.*

Usaremos este hecho junto con el teorema de recursión transfinita, según el cual, si X es un conjunto bien ordenado por una relación $\leq$, podemos definir una sucesión $\{A_x\}_{x \in X}$ definiendo un término arbitrario A_x en función de la sucesión de términos anteriores $\{A_y\}_{y<x}$.

Capítulo I

Los números enteros y racionales

1.1 Construcción de los números enteros

Seguramente el lector conocerá de sobra los números enteros. Los números enteros son:

$$\ldots \quad -5, \quad -4, \quad -3, \quad -2, \quad -1, \quad 0, \quad 1, \quad 2, \quad 3, \quad 4, \quad 5, \quad \ldots$$

En definitiva los números enteros no son sino los números naturales por duplicado, de modo que mientras la operación $4-7$ no puede efectuarse con números naturales, tiene en cambio la solución entera -3.

En primer lugar vamos a indicar cómo construir los números enteros en teoría de conjuntos. Aunque la formalización conjuntista no va a ser nuestra preocupación principal, consideramos ilustrativo detenernos en ello porque se trata de un buen ejemplo del uso de las relaciones de equivalencia, y el lector debería reflexionar sobre esta construcción no sólo hasta entenderla, sino hasta verla natural.

En principio podríamos definir los números enteros como los números naturales precedidos de un signo $+/-$, con el convenio de que $+0 = -0$. Esto sería lógicamente aceptable y probablemente es la definición que más se ajusta a la idea que el lector tiene de estos números, pero no es la definición más práctica ni mucho menos en la que podríamos pensar. Por ejemplo, si a partir de dicha definición queremos definir la suma de dos números enteros deberíamos escribir algo así como:

La suma de dos números enteros del mismo signo se calcula sumando sus valores absolutos con el mismo signo. La suma de dos números enteros de signos opuestos se calcula restando sus valores absolutos con el signo del sumando de mayor valor absoluto.

El lector lo habrá entendido perfectamente, pero desde un punto de vista lógico es una ley enrevesada y si quisiéramos usarla para probar algo tan simple

como que $(n+m)+r = n+(m+r)$ nos obligaría a distinguir casos y más casos.

La idea para obtener una definición práctica parte del hecho de que un número entero puede ser determinado algebraicamente como la resta de dos números naturales. Por ejemplo, el par $(8,3)$ determina el número $8-3=+5$, mientras que el par $(3,8)$ determina al número $3-8=-5$.

No podemos establecer que el número entero $+5$ será para nosotros el par de números naturales $(8,3)$, porque, por ejemplo, el par $(7,2)$ es otro objeto distinto que tendría el mismo derecho a ser identificado con el entero $+5$.

Entonces nos preguntamos cuándo dos pares de números (a,b) y (c,d) dan lugar al mismo número entero al restar sus componentes. Obviamente se cumple $a-b=c-d$ si y sólo si $a+d=b+c$. Ahora observamos que los pares de números naturales y la relación $a+d=b+c$ no involucran en absoluto números enteros, luego podemos usarlos para definir los números enteros sin que nuestra definición resulte circular.

Definición 1.1 Suponemos conocido el conjunto de los números naturales, al que aquí llamaremos $\mathbb{N}$. Definimos en $\mathbb{N}\times\mathbb{N}$ la relación R dada por

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

Es fácil probar que se trata de una relación de equivalencia. Llamaremos $[a,b]$ a la clase de equivalencia del par (a,b), es decir, $[a,b]$ es el conjunto formado por todos los pares relacionados con (a,b).

En los términos anteriores los elementos de $[a,b]$ son todos los pares que dan lugar al mismo número entero que (a,b) al restar sus componentes, con lo que existe exactamente una clase de equivalencia por cada número entero. Por ejemplo, el número $+5$ se corresponde con la clase cuyos elementos son $(5,0),(6,1),(7,2),\ldots$ La diferencia lógica es que los números enteros no los tenemos definidos y las clases de equivalencia respecto a la relación R sí.

Llamaremos conjunto de los *números enteros* al cociente $\mathbb{Z}=(\mathbb{N}\times\mathbb{N})/\mathrm{R}$. La letra Z es por el alemán Zahl (número). Si n es un número natural llamaremos $+n=[n,0]$ y $-n=[0,n]$.

Ahora es fácil probar que todo número entero $[a,b]$ es de la forma $[a-b,0]$ o bien $[0,b-a]$, según si a es mayor o menor que b, es decir, todo número entero es de la forma $+n$ o bien $-n$ para un número natural n. Además todos éstos son distintos salvo en el caso $+0=-0=[0,0]$.

Llamaremos números positivos a los del conjunto $\mathbb{Z}^+=\{+n \mid n\in\mathbb{N}, n\neq 0\}$. Los números negativos serán los del conjunto $\mathbb{Z}^-=\{-n \mid n\in\mathbb{N}, n\neq 0\}$. De este modo el conjunto $\mathbb{Z}$ se expresa como unión disjunta $\mathbb{Z}=\mathbb{Z}^+\cup\mathbb{Z}^-\cup\{0\}$.

Para ordenar los números enteros observamos que ha de ser $a-b\leq c-d$ si y sólo si $a+d\leq b+c$ (para todos los números naturales a,b,c,d), luego podemos definir $[a,b]\leq[c,d]$ si y sólo si $a+d\leq b+c$.

Esta definición exige comprobar que es compatible con la relación R, es decir, que si $[a,b]=[a',b']$ y $[c,d]=[c',d']$ entonces $a+d\leq b+c$ si y sólo si $a'+d'\leq b'+c'$.

La comprobación es sencilla, como también lo es probar que esta relación define un orden total con el cual $\mathbb{Z}$ queda ordenado según lo hemos representado en la página 1. En lo sucesivo identificaremos los números naturales con los números enteros no negativos. En particular suprimiremos el signo +, de modo que 2 y +2 serán una misma cosa. Por tanto podemos escribir $\mathbb{N} \subset \mathbb{Z}$.

La suma y el producto de números enteros se definen como sigue:

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

(El lector debe convencerse de que éstas son las definiciones lógicas. Por ejemplo, en el caso de la suma ha de considerar que $(a-b)+(c-d)=(a+c)-(b+d)$.)

Es fácil ver que estas operaciones son compatibles con la identificación que hemos hecho entre números naturales y enteros, es decir, que $7+5=12$ visto tanto como suma de números naturales como de enteros (más concretamente: $(+m)+(+n)=+(m+n)$).

Ahora es fácil demostrar las propiedades básicas de la suma de enteros. Veamos como muestra la asociatividad que antes habíamos puesto como ejemplo: $([a,b]+[c,d])+[e,f]=[a+c+e,b+d+f]=[a,b]+([c,d]+[e,f])$.

1.2 Anillos

Nuestro estudio de los números enteros nos va a llevar más adelante a trabajar con 'números' más generales (o más abstractos, si se quiere). Por ello, en lugar de enunciar directamente las propiedades básicas de las operaciones con enteros conviene hacerlo en un contexto más general, de manera que el mismo lenguaje que introduzcamos ahora nos permita después sentir cierta familiaridad con los objetos que nos encontraremos.

Definición 1.2 Una *ley de composición interna* en un conjunto A es una aplicación $*: A \times A \longrightarrow A$. Escribiremos $a*b$ en lugar de $*(a,b)$.

Diremos que una ley de composición interna $*$ es *asociativa* si cumple que $(a*b)*c=a*(b*c)$ para todos los elementos a, b y c del conjunto A.

En tal caso las expresiones de la forma $a*b*c*d*e$, y en general $a_1*\cdots*a_n$ están bien definidas, en el sentido de que no dependen del orden en que se efectúen las operaciones (respetando la posición de los factores) y por lo tanto no se necesitan paréntesis.

Una ley de composición interna $*$ es *conmutativa* si cumple $a*b=b*a$ para todos los elementos a y b del conjunto A. Si $*$ es a la vez asociativa y conmutativa las expresiones $a_1*\cdots*a_n$ no dependen tampoco de la posición de cada factor, es decir, podemos desordenarlas sin alterar el resultado.

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.

El elemento aludido en la condición 3 ha de ser único, pues si 0 y $0'$ cumplen lo mismo entonces $0 = 0 + 0' = 0'$. Lo llamaremos *elemento neutro* o *nulo* del anillo A.

Igualmente, para cada a de A el elemento $-a$ aludido en 4 es único, pues si b cumpliera lo mismo entonces $b = 0 + b = -a + a + b = -a + 0 = -a$. Lo llamaremos *elemento simétrico* u *opuesto* de a.

En lo sucesivo usaremos siempre los signos + y $\cdot$ para nombrar las operaciones de un anillo cualquiera, aunque en cada caso se tratará de una operación distinta. A la operación + la llamaremos 'suma' y a la operación $\cdot$ la llamaremos 'producto'. Igualmente, 'A es un anillo' significará que lo es con ciertas operaciones que se sobrentienden.

Escribiremos $a - b$ en lugar de $a + (-b)$. La notación $\sum_{i=1}^{n} a_i$ se usará para representar sumas finitas mientras que $\prod_{i=1}^{n} a_i$ indicará un producto finito.

Un anillo A es *conmutativo* si $ab = ba$ para todos los elementos a y b de A.

Un anillo A es *unitario* si existe un elemento 1 en A tal que $a \cdot 1 = 1 \cdot a = a$ para todo elemento a de A. Dicho elemento 1 ha de ser único, pues si 1 y $1'$ cumplen lo mismo entonces $1 = 1.1' = 1'$. Al elemento 1 lo llamaremos *identidad* de A.

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 sus operaciones.

Teorema 1.3 *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. $0a = a0 = 0$.

5. $(-a)b = a(-b) = -(ab)$.

6. $(-a)(-b) = ab$.

7. $-(a + b) = -a - b$.

DEMOSTRACIÓN:

1. $a + b = a + c \Rightarrow -a + a + b = -a + a + c \Rightarrow 0 + b = 0 + c \Rightarrow b = c.$
2. $a + a = a \Rightarrow a + a = a + 0 \Rightarrow a = 0.$
3. $-a + a = 0 = -a + (-(-a)) \Rightarrow a = -(-a).$
4. $0a = (0 + 0)a = 0a + 0a \Rightarrow 0a = 0.$
5. $(-a)b + ab = (-a + a)b = 0b = 0 \Rightarrow (-a)b = -(ab).$
6. $(-a)(-b) = -(a(-b)) = -(-(ab)) = ab.$
7. $(-a - b) + (a + b) = a - a + b - b = 0 \Rightarrow (-a - b) = -(a + b).$ ∎

Observemos que si en un anillo unitario A se cumple $1 = 0$ entonces cualquier $a \in A$ cumple $a = a \cdot 1 = a \cdot 0 = 0$, luego $A = 0$. Los anillos que más nos van a interesar son los anillos conmutativos y unitarios distintos de este caso trivial. A tales anillos los llamaremos dominios es decir: Un *dominio* es un anillo conmutativo y unitario en el que $1 \neq 0$. Es fácil ver que $\mathbb{Z}$ es un dominio.

Notemos que en cualquier anillo $a0 = 0a = 0$, pero no es cierto en general que si $ab = 0$ uno de los factores haya de ser nulo. Por supuesto en $\mathbb{Z}$ sí ocurre así. Vamos a dar una definición que recoja este hecho.

Definición 1.4 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.

Una propiedad muy importante de los dominios íntegros en 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 $b - c = 0$.

Ejercicio: Dotar a $\mathbb{Z} \times \mathbb{Z}$ de una estructura de dominio que no sea íntegro.

Para acabar con las propiedades básicas del anillo $\mathbb{Z}$ vamos a probar que cualquier par de números no nulos se puede dividir euclídeamente, es decir, se puede obtener un cociente y un resto. Nos basamos en que los números naturales cumplen esto mismo.

Teorema 1.5 *Sean D y d números enteros con d no nulo. Entonces existen unos únicos enteros c y r tales que $D = dc + r$ y $0 \leq r < |d|$, donde $|d|$ es igual a d si d es positivo y a $-d$ si es negativo.*

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 tenemos $D = dc + r$, 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 $D = d(c+1) + (-d-r)$.

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 1.6 Un *dominio euclídeo* es un dominio íntegro A tal que existe una función $\phi : A \setminus \{0\} \longrightarrow \mathbb{N}$ que cumpla lo siguiente:

1. Si a, b son elementos de A no nulos $\phi(a) \leq \phi(ab)$.
2. 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 $0 \leq \phi(r) < \phi(d)$.

La función ϕ se llama *norma euclídea.*

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 $|r| < |d|$.

Un elemento a de un dominio A es una *unidad* si existe un elemento b en A tal que $ab = 1$. Dicho elemento b está unívocamente determinado por a, ya que si $ab = 1 = ac$ entonces $b = b1 = bac = 1c = c$. A este único elemento lo llamaremos *inverso* de a y lo representaremos por a^{-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 $b = 1b = a^{-1}ab = a^{-1}0 = 0$.

Las unidades de $\mathbb{Z}$ son exactamente 1 y -1.

Un *anillo de división* es un anillo unitario con $1 \neq 0$ en el que todo elemento no nulo es una unidad.

Un *cuerpo* es un anillo de división conmutativo. En particular todo cuerpo es un dominio íntegro.

Observemos también que todo cuerpo K es un dominio euclídeo tomando como norma la aplicación constante 1, pues la división euclídea puede realizarse siempre con resto 0, es decir, $D = d(D/d) + 0$.

Vamos a definir operaciones entre números enteros y los elementos de un anillo.

Sea A un anillo, a un elemento de A y n un número entero. Definimos el elemento na como

$$na = \begin{cases} \overbrace{a + \cdots + a}^{n \text{ veces}} & \text{si } n > 0 \\ 0 & \text{si } n = 0 \\ \overbrace{(-a) + \cdots + (-a)}^{-n \text{ veces}} & \text{si } n < 0 \end{cases}$$

Si $n > 0$ definimos también $a^n = \overbrace{a \cdots a}^{n \text{ veces}}$. Si A es unitario $a^0 = 1$, y si a es una unidad y $n < 0$, entonces $a^n = \overbrace{a^{-1} \cdots \ a^{-1}}^{-n \text{ veces}}$.

Es pura rutina comprobar los hechos siguientes.

Teorema 1.7 *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 $\mathbb{Z}$.

1.3 Cuerpos de cocientes. Números racionales

A continuación vamos a dar un método para obtener un cuerpo a partir de un dominio íntegro. A partir de $\mathbb{Z}$ obtendremos el cuerpo de los números racionales, pero el método es general y lo aplicaremos a más casos.

Sea K un cuerpo y a, b dos elementos de K con b no nulo. Llamaremos $\frac{a}{b} = ab^{-1}$. Es fácil comprobar las relaciones siguientes:

$$\frac{a}{b} = \frac{c}{d} \Leftrightarrow ad = bc, \quad \frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \quad \frac{a}{b}\frac{c}{d} = \frac{ac}{bd}.$$

Con estos hechos *in mente* definimos:

Definición 1.8 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 $(c,d) \Leftrightarrow ad = bc$. Es fácil probar que R es una relación de equivalencia en $A \times A^*$. Llamaremos $\frac{a}{b}$ a la clase de equivalencia del par (a,b). De este modo se cumple $\frac{a}{b} = \frac{c}{d} \Leftrightarrow ad = bc$. Llamaremos *cuerpo de cocientes* de A al conjunto cociente $K = (A \times A^*)/\mathrm{R}$. Es fácil comprobar que ciertamente K es un cuerpo con las operaciones dadas por $\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$, $\frac{a}{b}\frac{c}{d} = \frac{ac}{bd}$. Concretamente $0 = \frac{0}{1}$, $1 = \frac{1}{1}$,$-\frac{a}{b} = \frac{-a}{b} = \frac{a}{-b}$, y si $\frac{a}{b} \neq 0$, entonces $\left(\frac{a}{b}\right)^{-1} = \frac{b}{a}$.

Para relacionar un anillo con su cuerpo de cocientes conviene introducir algunos conceptos. 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 $a + a = b + b = a$, $a + b = b + a = b$, $aa = ab = ba = a$, $bb = b$.

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 1.9 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 $f(ma) = mf(a)$.

Una precaución es que 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 $f(1) = 1$.

Cuando $f(1) = 1$ se cumple $f(a^n) = f(a)^n$ para todo elemento a de A y todo entero n. En cualquier caso esto vale para exponentes positivos.

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 tanto podemos considerarlos el mismo anillo.

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 1.10 *Sea A un dominio íntegro y K su cuerpo de cocientes.*

a) La aplicación $\phi : A \longrightarrow K$ dada por $\phi(a) = a/1$ es un monomorfismo de anillos.

b) Si K' es un cuerpo y $\psi : A \longrightarrow K'$ es un monomorfismo de anillos, existe un único monomorfismo de cuerpos $\chi : K \longrightarrow K'$ tal que para todo a de A se cumple $\chi(\phi(a)) = \psi(a)$.

Demostración: a) es inmediato. Para probar b) basta definir $\chi(a/b) = \psi(a)\psi(b)^{-1}$. Se prueba que la definición no depende de la representación de a/b como fracción y que es un monomorfismo. ∎

Lo que afirma la parte a) 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 b) afirma es 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 aunque, si no tenemos a A metido en ningún cuerpo, siempre podemos realizar la construcción del teorema 1.10

Definición 1.11 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 $\mathbb{Z}$, $b \neq 0$. Como $\frac{a}{b} = \frac{-a}{-b}$, podemos exigir que b sea positivo.

El cuerpo $\mathbb{Q}$ está totalmente ordenado por la relación $\frac{a}{b} \leq \frac{c}{d} \Leftrightarrow ad \leq bc$ (si b, $d > 0$). Es fácil ver que este orden extiende al de $\mathbb{Z}$. Llamaremos $\mathbb{Q}^+$ al conjunto de los números racionales positivos (mayores que 0) y $\mathbb{Q}^-$ al de los números negativos.

El *valor absoluto* de un número racional r es

$$|r| = \begin{cases} r & \text{si } r \geq 0, \\ -r & \text{si } r < 0. \end{cases}$$

El *signo* de r es

$$\operatorname{sig} r = \begin{cases} 1 & \text{si } r > 0, \\ 0 & \text{si } r = 0, \\ -1 & \text{si } r < 0. \end{cases}$$

El lector puede entretenerse demostrando el teorema siguiente:

Teorema 1.12 *Sean r, s, t y u números racionales.*

1. *Si $r \le s$ y $t \le u$ entonces $r + t \le s + u$.*
2. *Si $r \le s$ entonces $-s \le -r$ y si son no nulos $1/s \le 1/r$.*
3. *Si $0 \le r$ y $s \le t$, entonces $rs \le rt$.*
4. *Existe un número natural n tal que $r < n$.*
5. *Si $r < s$ existe un número racional t tal que $r < t < s$.*
6. *$|r| = |s|$ si y sólo si $r = s$ o $r = -s$.*
7. *$|r| \le a$ si y sólo si $-a \le r \le a$.*
8. *$|rs| = |r||s|$.*
9. *$|a + b| \le |a| + |b|$.*
10. *$\big||a| - |b|\big| \le |a - b|$.*

(Veamos por ejemplo la prueba de 10: por 9) $|a| = |a - b + b| \le |a - b| + |b|$, luego $|a| - |b| \le |a - b|$, y similarmente $|b| - |a| \le |a - b|$, luego tenemos que $-|a - b| \le |a| - |b| \le |a - b|$, y por 7) concluimos 10).

Los cuerpos de cocientes nos permiten salirnos temporalmente de un anillo en nuestros cálculos aunque después volvamos a él. Veamos un ejemplo.

Definición 1.13 Definimos inductivamente el *factorial* de un número natural mediante las condiciones siguientes:

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

Por ejemplo

$$0! = 1,\ 1! = 1,\ 2! = 2,\ 3! = 6,\ 4! = 24,\ 5! = 120,\ 6! = 720,\ \text{etc.}$$

Sean n, n_1, ... , n_k números naturales tales que $n = n_1 + \ldots + n_k$. Definimos el *número combinatorio*

$$\binom{n}{n_1 \cdots n_k} = \frac{n!}{n_1! \cdots n_k!}$$

Si $0 \le m \le n$ 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 1.14 *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

$$\binom{n}{n_1\ \ldots\ n_k\ n_{k+1}} = \binom{n-n_{k+1}}{n_1\ \ldots\ n_k}\binom{n}{n_{k+1}}.$$

■

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

La utilidad principal de estos números será para nosotros el hecho siguiente:

Teorema 1.15 (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.

$$\begin{aligned}
(a+b)^{n+1} = (a+b)^n(a+b) &= \sum_{m=0}^{n} \binom{n}{m} a^m b^{n-m}(a+b) \\
&= \sum_{m=0}^{n} \binom{n}{m} a^{m+1} b^{n-m} + \sum_{m=0}^{n} \binom{n}{m} a^m b^{n-m+1} \\
&= \sum_{m=1}^{n+1} \binom{n}{m-1} a^m b^{n+1-m} + \sum_{m=0}^{n} \binom{n}{m} a^m b^{n+1-m} \\
&= \binom{n}{0} a^0 b^{n+1} + \binom{n}{n} a^{n+1} b^0 \\
&\quad + \sum_{m=1}^{n} \binom{n}{m-1} a^m b^{n+1-m} + \sum_{m=1}^{n} \binom{n}{m} a^m b^{n+1-m} \\
&= \binom{n}{0} a^0 b^{n+1} + \binom{n+1}{n+1} a^{n+1} b^0 \\
&\quad + \sum_{m=1}^{n} \left(\binom{n}{m-1} + \binom{n}{m} \right) a^m b^{n+1-m} \\
&= \binom{n}{0} a^0 b^{n+1} + \binom{n+1}{n+1} a^{n+1} b^0 + \sum_{m=1}^{n} \binom{n+1}{m} a^m b^{n+1-m} \\
&= \sum_{m=0}^{n+1} \binom{n+1}{m} a^m b^{n+1-m}.
\end{aligned}$$

■

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 1.16 *Sea A un anillo conmutativo y unitario,n un número natural y $a_1, \ldots, a_k$ elementos de A. Entonces se cumple:*

$$(a_1 + \cdots + a_k)^n = \sum_{n_1,\ldots,n_k} \binom{n}{n_1 \cdots n_k} a_1^{n_1} \cdots a_k^{n_k},$$

donde la suma se extiende sobre todos los números naturales $n_1, \ldots, n_k$ tales que $n_1 + \cdots + n_k = n$.

1.4 Cuaterniones racionales

Para terminar esbozaremos un ejemplo de un anillo de división D que no es un cuerpo. La idea es que los elementos de D han de ser de la forma $a + bi + cj + dk$, donde a, b, c y d son números racionales. Los elementos i, j, k se multiplican como sigue:

$$i^2 = j^2 = k^2 = -1, \quad ij = k, \quad ji = -k, \quad jk = i, \quad kj = -i, \quad ki = j, \quad ik = -j.$$

O sea, cuando multiplicamos en el orden $i \to j \to k \to i$, el producto es el elemento restante, pero si multiplicamos en el orden inverso obtenemos el opuesto.

Según esto, dos elementos cualesquiera se han de multiplicar así:

$$(a+bi+cj+dk)(a'+b'i+c'j+d'k) = (aa'-bb'-cc'-dd')+(ab'+ba'+cd'-dc')i$$
$$+(ac'+ca'+db'-bd')j+(ad'+da'+bc'-cb')k.$$

Como un elemento de D viene determinado por cuatro números racionales, formalmente podemos definir $D = \mathbb{Q}^4$ y las operaciones vendrán dadas por:

$$(a,b,c,d) + (a',b',c',d') = (a+a', b+b', c+c', d+d')$$

$$(a,b,c,d)(a',b',c',d')$$
$$= (aa'-bb'-cc'-dd', ab'+ba'+cd'-dc', ac'+ca'+db'-bd', ad'+da'+bc'-cb').$$

Así es fácil, aunque tedioso, probar que D es un anillo unitario. La identidad es, por supuesto, $(1,0,0,0)$.

Llamando $1 = (1,0,0,0)$, $i = (0,1,0,0)$, $j = (0,0,1,0)$ y $k = (0,0,0,1)$, es fácil probar que

$$(a,b,c,d) = (a,0,0,0)1 + (b,0,0,0)i + (c,0,0,0)j + (d,0,0,0)k.$$

Por otra parte, la aplicación que a cada número racional a le asigna $(a,0,0,0)$ es un monomorfismo de anillos, por lo que si identificamos a con $(a,0,0,0)$, obtenemos que cada elemento de D se expresa de forma única como queríamos, es decir, $(a,b,c,d) = a + bi + cj + dk$.

Los elementos de D se llaman *cuaterniones racionales.* Para probar que D es realmente un anillo de división conviene llamar *conjugado* de $q = a+bi+cj+dk$ al cuaternión $\bar{q} = a - bi - cj - dk$. Es fácil probar que $q\bar{q} = a^2 + b^2 + c^2 + d^2$. A este número lo llamaremos *norma* de q. La norma de un cuaternión q, que representaremos por $\mathrm{N}(q)$, es un número racional positivo y además claramente $\mathrm{N}(q) = 0 \Leftrightarrow q = 0$.

Si q es un cuaternión no nulo, tenemos que existe $q^{-1} = \frac{\bar{q}}{\mathrm{N}(q)}$, luego ciertamente D es un anillo de división. Como $ij \neq ji$, no es un cuerpo.

Ejercicio: Comprobar que $\overline{pq} = \bar{q}\bar{p}$, y de aquí a su vez que $\mathrm{N}(pq) = \mathrm{N}(p)\,\mathrm{N}(q)$.

Ejercicio: Escribir explícitamente la igualdad $\mathrm{N}(pq) = \mathrm{N}(p)\,\mathrm{N}(q)$ e interpretarla como una propiedad de los números naturales.

Ejercicio: ¿Qué condición ha de cumplir un cuerpo K para que podamos construir un anillo de división de cuaterniones sobre K?

Capítulo II

Anillos de polinomios

Si 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 $(xy+x)(x^2-2y) = x^2(xy+x)-2y(xy+x) = x^3y + x^3 - 2xy^2 - 2xy$.

Al trabajar con números enteros 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. Lo que vamos a hacer es dar una construcción general que permite añadir a cada anillo A un conjunto 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. Un polinomio no es nada más que esto, pero la construcción formal resulta un tanto técnica.

2.1 Construcción de los anillos de polinomios

Definición 2.1 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, 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:

$$(f+g)(u) = f(u) + g(u), \qquad (fg)(u) = \sum_{v+w=u} f(v)g(w).$$

Notar que el sumatorio que define el producto es finito.

Teorema 2.2 *Sea A un anillo y S un conjunto. Entonces $A[S]$ es un anillo. Si A es conmutativo o unitario, $A[S]$ también lo es.*

Demostración: Es fácil ver que si f, g, $h \in A[S]$, entonces $(f+g)+h = f+(g+h)$ 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 $(f+g)h = fh + gh$.

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.

Si A es unitario, 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 $1f = f$. ∎

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 2.3 *Sea A un anillo 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 si A es unitario $\phi(1) = 1$.*

Definición 2.4 En lo sucesivo, si A es un anillo, 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 $A[S]$. Supongamos que A es unitario. 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 2.5 *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 2.6 *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 2.7 *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 2.8 En la expresión de 2.6, 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, de la forma $a\, x_1^{k_1} \ldots x_n^{k_n}$) es un *monomio*. Por tanto un 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 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 $a_m \neq 0$.

Si m es el mayor natural tal que el coeficiente de x^m en un polinomio p es no nulo, entonces a dicho coeficiente se le llama *coeficiente director* del polinomio p y el número m se llama *grado* de p y lo representaremos por $\operatorname{grad} p$.

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:

$$\sum_{i=0}^{m} a_i x^i + \sum_{i=0}^{m} b_i x^i = \sum_{i=0}^{m} (a_i + b_i) x^i,$$

$$\left(\sum_{i=0}^{m} a_i x^i\right)\left(\sum_{i=0}^{n} b_i x^i\right) = \sum_{i=0}^{m+n}\left(\sum_{i+j=k} a_i b_j\right)x^k.$$

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.

2.2 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 2.9 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

$$\phi p(v) = \sum_{i=1}^{m} \phi(a_i)\, v(x_1)^{k_{i1}} \ldots\ v(x_n)^{k_{in}} \in B.$$

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 pondremos simplemente $p(b_1, \ldots, b_n)$.

Teorema 2.10 *Sean A y B dos anillos conmutativos y unitarios, $\phi : A \longrightarrow B$ un homomorfismo tal que $\phi(1) = 1$, 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 $p = \sum_{i=1}^{m} a_i\, x_1^{k_{i1}} \ldots\ x_n^{k_{in}}$ y $q = \sum_{i=1}^{m} b_i\, x_1^{k_{i1}} \ldots\ x_n^{k_{in}}$. Observar 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.

$$\Phi(p+q) = \Phi\left(\sum_{i=1}^{m} (a_i + b_i)\, x_1^{k_{i1}} \ldots\ x_n^{k_{in}}\right) = \sum_{i=1}^{m} \phi(a_i + b_i) v(x_1)^{k_{i1}} \ldots\ v(x_n)^{k_{in}}$$

$$= \sum_{i=1}^{m} \big(\phi(a_i) + \phi(b_i)\big) v(x_1)^{k_{i1}} \dots\ v(x_n)^{k_{in}} = \sum_{i=1}^{m} \phi(a_i) v(x_1)^{k_{i1}} \dots\ v(x_n)^{k_{in}}$$

$$+ \sum_{i=1}^{m} \phi(b_i) v(x_1)^{k_{i1}} \dots\ v(x_n)^{k_{in}} = \Phi(p) + \Phi(q).$$

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 2.11 *Sean A y B anillos conmutativos y unitarios y $\phi : A \longrightarrow B$ un homomorfismo tal que $\phi(1) = 1$. Sea 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 2.12 *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 2.10 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 2.10 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 $A[S] \subset A[T]$.

2.3 Propiedades algebraicas

Las principales propiedades algebraicas de los anillos de polinomios se deducen a partir 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 2.13 *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 $pq = \sum_{k=0}^{m+n} \left(\sum_{i+j=k} a_i b_j \right) x^k$ 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 2.14 *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 2.15 *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 2.13 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. ∎

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 2.16 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 empezamos a introducir en el capítulo anterior: 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 del 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.

Debemos ocuparnos ahora de la posibilidad de dividir polinomios. Esta es una característica importantísima de los anillos con una indeterminada.

Teorema 2.17 *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 el grado de r 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 $db_m^{-1}a_nx^{n-m} = \sum_{i=0}^{m} b_ib_m^{-1}a_nx^{i+n-m}$. El monomio de mayor grado es $b_m(b_m)^{-1}a_nx^{m+n-m} = a_nx^n$, luego se trata de un polinomio de grado n con coeficiente director a_n.

Consecuentemente el polinomio $D - d(b_m)^{-1}a_nx^{n-m}$ tiene grado menor que n, luego por hipótesis de inducción existen polinomios c' y r de manera que $D - db_m^{-1}a_nx^{n-m} = dc' + r$ con $\operatorname{grad} r < \operatorname{grad} d$.

Sea $c = b_m^{-1}a_nx^{n-m} + c'$. 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 2.13. 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 2.18 *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. Si bien la teoría que estamos desarrollando ha surgido para resolver una serie de problemas anteriores a ella misma, estamos ante un ejemplo (a un nivel muy simple) de cómo cada teoría plantea de forma natural nuevos problemas a la vez que resuelve otros, problemas que nunca se hubieran podido formular fuera del contexto creado por ella.

Capítulo III

Ideales

En los capítulos anteriores hemos introducido los que van a ser por ahora nuestros objetos de estudio principales: los números enteros y racionales y sus anillos de polinomios. Ahora vamos a introducir un concepto que ha resultado ser fundamental en el estudio de éstos y otros anillos relacionados. Se trata del concepto de ideal. Por razones que luego podremos entrever, el concepto de ideal surgió con cierto retraso en el estudio de los números. Nosotros lo introducimos desde un principio porque, dada su importancia, conviene familiarizarse con él cuanto antes. Sin embargo, para evitar un grado de abstracción que todavía no podemos justificar, aquí nos limitaremos a ver las mínimas ideas que nos puedan ser útiles de momento.

3.1 Ideales en un dominio

Definición 3.1 Un ideal en un dominio 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 $ab \in I$.

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.

Una observación trivial es que si un ideal I de dominio A contiene una unidad u, entonces $I = A$. En efecto, por definición de ideal se cumple que $1 = u^{-1}u \in I$ y si $a \in A$, entonces $a = a1 \in I$, es decir, 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, será una unidad, y el ideal será el cuerpo completo.

Definición 3.2 Es inmediato que la intersección de una familia 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 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 3.3 *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 $(x) = \{ax \mid a \in A\}$.

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 $\sum_{i=1}^n a_ix = (\sum_{i=1}^n a_i)\, x$ están en $\{ax \mid a \in A\}$, luego $(X) \subset \{ax \mid a \in A\}$. La otra inclusión es obvia. ∎

Entre los ideales de un anillo se puede definir una suma y un producto como sigue:

Definición 3.4 Sea A un anillo y sean $S_1, \ldots, S_n$ subconjuntos de A. Llamaremos

$$\begin{aligned} S_1 + \cdots + S_n &= \{s_1 + \cdots + s_n \mid s_i \in S_i \text{ para } \ i = 1, \ldots, n\} \\ S_1 \cdots S_n &= \Big\{ \sum_{i=1}^{m} s_{i1} \cdots s_{in} \mid m \in \mathbb{N} \text{ y } s_{ij} \in S_j \text{ para } j = 1, \ldots, n \Big\} \end{aligned}$$

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 Dominios de ideales principales

Definición 3.5 Un ideal de un dominio A es *principal* si está generado por un solo elemento, es decir, si es de la forma $(a) = aA = \{ab \mid b \in A\}$.

Un *dominio de ideales principales* (DIP) es un dominio íntegro en el que todo ideal es principal.

Teorema 3.6 *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 $b = ac + r$, para $r = 0$ o bien $\phi(r) < \phi(a)$. Como $a \in I$, por la definición de ideal $ac \in I$, y como I es un subanillo, también $b - ac \in I$, 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 $n \in Z$. También son DIP los anillos $K[x]$, donde K es un cuerpo.

Como los cuerpos no tienen más ideales que los impropios, y éstos son principales, $(0 = (0), A = (1))$, resulta que los cuerpos son trivialmente DIPs. (Alternativamente, sabemos que los cuerpos son dominios euclídeos.)

El hecho de que los anillos más importantes sean DIPs es la explicación de que el concepto de ideal tardara en surgir en teoría de números. Cualquier afirmación sobre ideales en un DIP puede reformularse como una afirmación sobre los elementos del anillo, pues cada ideal está determinado por su generador. No obstante hay anillos que no son DIP, y al estudiarlos conviene saber cuántas cosas son ciertas para ideales en general aunque no sean principales. De hecho, en ciertos casos de interés, resultados que en DIPs pueden formularse con elementos y con ideales, son falsos en otros anillos en términos de elementos, pero siguen siendo ciertos en términos de ideales.

Vamos a ver unos ejemplos de dominios íntegros que no son DIPs.

Teorema 3.7 *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. ∎

Esto nos da muchos ejemplos de dominios íntegros que no son DIP (ni por tanto euclídeos). A saber, $\mathbb{Z}[x]$ y más en general $A[S]$ cuando el cardinal de S es mayor que 1 (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 $\mathbb{Q}[x, y]$.

3.3 Anillos noetherianos

Para acabar el capítulo vamos a definir una clase de anillos más general que la de los DIPs y que jugará un papel relevante en el próximo capítulo.

Definición 3.8 Un dominio íntegro A es un anillo *noetheriano* si todo ideal de A es finitamente generado.

Evidentemente, todo DIP es un anillo noetheriano.

Teorema 3.9 *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 de ideales de A tiene un maximal 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 de ideales de A no tuviera maximal, sería posible extraer 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 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). ∎

Capítulo IV

Divisibilidad en dominios íntegros

El concepto de divisibilidad es uno de los más importantes en el estudio de los números. A partir de él se plantean los más interesantes y variados problemas cuyo estudio ha ocupado a los matemáticos durante milenios. Aquí desarrollaremos la teoría básica al respecto. En capítulos posteriores profundizaremos más en ella.

4.1 Conceptos básicos

Definición 4.1 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 (y lo representaremos $a \mid b$) si existe un elemento c de A tal que $b = ac$.

Por ejemplo en $\mathbb{Z}$ es fácil ver que 3 divide a 15, pero no a 16.

Es obvio que si $a \mid b$ y $b \mid c$ entonces $a \mid c$.

Si u es una unidad, cualquier elemento a de A se expresa como $a = u(u^{-1}a)$, luego las unidades dividen a todo elemento de A. Por otra parte si u es una unidad y $a \mid u$, entonces existe un b en A tal que $u = ab$, luego $1 = abu^{-1}$, es decir, a es una unidad. En otras palabras, los divisores de las unidades son las unidades.

Por el contrario 0 no divide a nadie salvo a sí mismo.

Diremos que dos elementos a y b de A son *asociados* si $a \mid b$ y $b \mid a$. Por ejemplo en $\mathbb{Z}$ se cumple que 3 y -3 son asociados. Ser asociado es una relación de equivalencia. Si dos elementos son asociados tienen los mismos múltiplos y divisores.

La asociación está estrechamente relacionada con la existencia de unidades. En efecto, si a y b son asociados no nulos, entonces $a = ub$ y $b = va$, para ciertos u y v del anillo A. Por lo tanto $a = uva$, de donde $uv = 1$, o sea, u y v son unidades. Así pues, si dos elementos son asociados, uno se obtiene del otro

multiplicándolo por una unidad. El recíproco es cierto, como es fácil observar. Además, cuando un mismo elemento no nulo se multiplica por unidades distintas obtenemos elementos distintos, luego un elemento no nulo de A tiene tantos asociados como unidades hay en A. Como $\mathbb{Z}$ tiene dos unidades, los asociados en $\mathbb{Z}$ forman parejas, salvo el cero, que es su único asociado.

Tenemos, pues, que todo elemento de A tiene por divisores a las unidades de A y a sus propios asociados (entre los que está 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.

Estas consideraciones nos llevan al concepto de elemento irreducible: 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. También es obvio que un elemento es irreducible si y sólo si lo es cualquiera de sus asociados.

Por ejemplo, es fácil ver que los únicos divisores de 5 en $\mathbb{Z}$ son 1, -1, 5 y -5, lo que implica que 5 es irreducible en $\mathbb{Z}$. En cambio 15 no es irreducible, pues factoriza como $15 = 3 \cdot 5$.

Si un número entero (no nulo ni unidad) no es irreducible, entonces factoriza como producto de dos enteros estrictamente menores en módulo. Si éstos no son irreducibles factorizarán a su vez en factores menores, y este proceso tiene que acabar antes o después, por lo que todo número entero se puede expresar como producto de irreducibles. Más aún, puede probarse que esta descomposición es esencialmente única. Para formular esto con precisión y en términos aplicables a otros casos (como por ejemplo a polinomios), conviene introducir el concepto siguiente:

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).

No es difícil probar por métodos elementales que $\mathbb{Z}$ es un DFU. Por ejemplo la factorización única de 140 es $140 = 2 \cdot 2 \cdot 5 \cdot 7 = (-5) \cdot 2 \cdot 7 \cdot (-2) = \cdots$ Sin embargo vamos a probar más en general que todo DIP es un DFU. Esto lo veremos en la sección siguiente. Acabaremos ésta con algunas consideraciones adicionales sobre DFUs que nos ayudarán a familiarizarnos con ellos.

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 $e_p(a)$. En una descomposición de a aparecerán $e_p(a)$ 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, $n_i = e_{p_i}(a)$ 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 $e_p(a)$ 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 $e_p(ab) = e_p(a) + e_p(b)$.

Podemos definir $e_p(a) = 0$ para todo irreducible p cuando a es una unidad y así la relación anterior es válida también si a o b es una unidad.

Notar también que un irreducible p divide a un elemento a si y sólo si $e_p(a) \neq 0$. En efecto, si $e_p(a) \neq 0$ 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 $e_p(a) = e_p(p) + e_p(b) = 1 + e_p(b) \neq 0$.

Si $a \mid b$, ha de cumplirse que $e_p(a) \leq e_p(b)$ 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 $e_p(b) - e_p(a)$ (y una unidad adecuada). Dos elementos a y b son asociados si y sólo si $e_p(a) = e_p(b)$ 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 $0 \neq e_p(a) + e_p(b)$, 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, 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 4.2 *Sea A un dominio íntegro*

1. *Todo primo de A es irreducible.*
2. *Si A es DFU, entonces 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. ∎

4.2 Ideales y divisibilidad

Aunque todavía no estamos en condiciones de comprender enteramente por qué, lo cierto es que los ideales proporcionan el lenguaje idóneo para expresar los hechos más relevantes de la divisibilidad en un anillo. En primer lugar hemos de notar que 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. Notar 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 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. 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 4.3 Un ideal P de un anillo 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 al exigencia de que los primos no sean unidades. Hay, no obstante, una discrepancia debida principalmente a motivos históricos, y es que, mientras 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 4.4 *Si A es un dominio íntegro, un ideal P de 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 4.5 Un ideal M de un anillo 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).

La existencia de ideales maximales en cualquier anillo conmutativo y unitario A está garantizada por el lema de Zorn. Más aún, todo ideal distinto de A está contenido en un ideal maximal (la familia de ideales distintos de A que contienen a un ideal está ordenada por la inclusión, y todo subconjunto totalmente ordenado tiene una cota superior, pues la unión de una cadena de ideales es claramente un ideal, además es un ideal distinto de A porque no puede contener a la identidad). No vamos a necesitar esto de momento.

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 4.6 *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 4.7 *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 4.8 *En un DIP los ideales maximales coinciden con los ideales primos 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 4.6 ∎

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 4.9 *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 $(a_o) \subsetneq (a_1) \subsetneq (a_2) \subsetneq \cdots$ en contradicción con el teorema 3.9. 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.

4.3 Divisibilidad en $\mathbb{Z}$

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.

El problema más elemental que surge a raíz de todo esto es encontrar un método para obtener las factorizaciones en primos de números cualesquiera. En particular sería conveniente hallar un método para reconocer los números primos. El método más simple para hallar todos los primos hasta un número dado es la llamada *criba de Eratóstenes,* que consiste en escribir una lista con

los primeros n naturales, tachar el 1, que no es primo por definición, después tachar todos los múltiplos de 2 (salvo el propio 2), dejar el menor número que queda (el 3) y tachar sus múltiplos, dejar el menor número que queda (el 5) y tachar sus múltiplos, etc. Los números que sobrevivan serán los primos menores que n. He aquí la lista de los primos menores que 100, que hacen un total de 25.

$$2,\quad 3,\quad 5,\quad 7,\quad 11,\quad 13,\quad 17,\quad 19,\quad 23,\quad 29,\quad 31,\quad 37,\quad 41,$$

$$43,\quad 47,\quad 53,\quad 59,\quad 61,\quad 67,\quad 71,\quad 73,\quad 79,\quad 83,\quad 89,\quad 97.$$

Para descomponer un número en factores primos podemos ir probando a dividirlo por los primos menores que él hasta hallar uno que lo divida e ir repitiendo la operación con los cocientes que vayamos obteniendo. Notar que si queremos factorizar un número n y m cumple que $n \leq m^2$, entonces, si n no es primo, el menor primo que divide a n ha de ser menor que m. Por ejemplo, el menor primo que divida a un número menor que 100 ha de ser menor que 10, es decir, si un número menor que 100 no es divisible entre 2, 3, 5 o 7, entonces es primo.

En cualquier caso, siempre es posible distinguir los números primos de los compuestos y hallar la factorización de cualquier número en un número finito de pasos. Más adelante encontraremos técnicas para abordar este problema con más elegancia.

Una cuestión importante es si el número de primos es finito o infinito. La respuesta es que es infinito. 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 4.10 *(Euclides): En $\mathbb{Z}$ hay infinitos números primos.*

Demostración: Dado un número n consideremos $n!$ Se cumple que todo número 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 4.11 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 $2^2 \cdot 3 \cdot 7^5$, $2 \cdot 5 \cdot 7$, $3^4 \cdot 5^2 \cdot 7$, el 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 un DIP el máximo común divisor de un conjunto finito de números cumple una propiedad muy importante:

Teorema 4.12 *(Relación de Bezout): Sea A un DIP y $a_1, \ldots, a_n$ elementos de A. Sea d un* mcd *de $a_1, \ldots, a_n$. Entonces $(d) = (a_1) + \cdots + (a_n)$, luego existen ciertos elementos $r_1, \ldots, r_n$ en 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 $rm + sn = 1$.

4.4 Divisibilidad en anillos de polinomios

El estudio de la divisibilidad no es tan sencillo si pasamos a los anillos de polinomios $A[x]$. Tomemos unos cuantos polinomios y multipliquémoslos:

$$(x^2+3x-5)(x-1)(2x^3+3) = 2x^6+4x^5-16x^4+13x^3+6x^2-24x+15.$$

¿Cómo encontrar los factores a partir del producto? De hecho ni siquiera sabemos si los factores que hemos tomado son irreducibles o no (es costumbre hablar de números primos pero de polinomios irreducibles, aunque en principio son términos equivalentes). El problema es que, a diferencia del caso numérico, hay infinitos posibles divisores para un polinomio dado. No hay ningún criterio general para determinar si un polinomio dado es o no irreducible, aunque algo se puede decir sobre el tema. De todos modos, antes de entrar en ello tenemos planteado otro problema más importante. Sabemos que $A[x]$ es un DFU cuando A es un cuerpo, pero en otro caso $A[x]$ no es DIP. ¿Sigue siendo $A[x]$ un DFU a pesar de ello? ¿Es $\mathbb{Z}[x]$ DFU?, ¿Es $\mathbb{Q}[x,y]$ DFU? Vamos a probar que sí, para lo cual 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 4.13 Sea D un DFU y sea $c : D[x] \longrightarrow D$ una aplicación que asigne a cada polinomio $f \in D[x]$ un mcd de sus coeficientes no nulos (y $c(0) = 0$). A $c(f)$ se le llama *contenido* del polinomio f.

Usaremos la notación $a \approx b$ para indicar que a y b son asociados. Es claro que si f es un polinomio no nulo y a es un mcd de sus coeficientes no nulos, entonces $a \approx c(f)$.

Diremos que un polinomio f es *primitivo* si $c(f)$ es una unidad, 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 (en $\mathbb{Z}[x]$ podemos elegir los contenidos naturales), mientras que el polinomio x^2-4 es primitivo. En general es inmediato que si $f \in D[x]$ y $f \neq 0$, entonces $f(x) = c(f)g(x)$ donde $g(x) \in D[x]$ 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 $D[x]$ para probar la unicidad de los factores irreducibles no constantes. Necesitamos dos resultados sobre $c(f)$.

Teorema 4.14 *Sea D un DFU.*

1. *Si $a \in D$ y $f \in D[x]$, entonces $c(af) \approx a \cdot c(f)$.*
2. *Si f, $g \in D[x]$, entonces $c(fg) \approx c(f)c(g)$.*

Demostración: 1) Es inmediato que $a \cdot c(f)$ es un mcd de los coeficientes de af.

2) Sea $f = c(f)f_1$ y $g = c(g)g_1$ con f_1 y g_1 primitivos. Entonces $c(fg) = c\,(c(f)f_1 \cdot c(g)g_1) \approx c(f)c(g)c(f_1g_1)$, luego basta probar que $c(f_1g_1)$ es una unidad.

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 $f_1 \cdot g_1 = \sum_{k=0}^{m+n} \left(\sum_{i+j=k} a_i b_j\right) x^k$. Si $c(f_1g_1)$ no fuera una unidad en D, existiría un irreducible p tal que $p \mid c(f_1g_1)$.

Entonces $p \mid \sum_{i+j=k} a_i b_j$ para cada k entre 0 y $n+m$.

Como $c(f_1)$ es una unidad, p no divide a $c(f_1)$, 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 índice t tal que $p \mid b_j$ para $j < t$ y $p \nmid b_t$ ($b_t \neq 0$).

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

$$c(f) \approx d_1 \cdots d_r.$$

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 $c(f)$ en D.

Para demostrar la unicidad del resto de la descomposición necesitamos dos resultados:

Teorema 4.15 *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 2.15 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 $c(f)$ y $c(g)$ son unidades en D tenemos que $s \approx s \cdot c(f) \approx c(sf) = c(rg) \approx r \cdot c(g) \approx r$, luego $r = sv$ para cierta unidad v de D. En consecuencia $sf = rg = svg$, luego $f = vg$, lo que prueba que f y g son asociados en $D[x]$.

El recíproco es obvio. ∎

Teorema 4.16 *(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 $g = \sum_{i=0}^{n} \frac{a_i}{b_i} x^i$ y $h = \sum_{i=0}^{m} \frac{c_i}{d_i} x^i$ 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 $g_1 = ag_2$, siendo $a = c(g_1)$ y $g_2 \in D[x]$ un polinomio primitivo.

Claramente $g = \frac{1}{b} g_1 = \frac{a}{b} g_2$, luego $\operatorname{grad} g = \operatorname{grad} g_2$. Del mismo modo se llega a que $h = \frac{c}{d} h_2$, donde c, $d \in D$ y $h_2 \in D[x]$ es primitivo, $\operatorname{grad} h = \operatorname{grad} h_2$.

Ahora $f = gh = \frac{ac}{bd} g_2 h_2$, con lo que f y $g_2 h_2$ 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 $f = u g_2 h_2$, 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, lo que significa que $g \in K$, pero como está en $D[x]$, en realidad g está en D. Ahora $1 \approx c(f) \approx gc(h)$, luego g es una unidad en D, luego en $D[x]$. Esto prueba que f es irreducible en $D[x]$. ∎

Finalmente podemos probar:

Teorema 4.17 *(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. ∎

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 4.18 *(Criterio de irreducibilidad de Eisenstein) Sea D un DFU, 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 $i = 0, \ldots n-1$, $p \nmid a_n$ 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 lo será. 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 $c(g)c(h) \approx c(f)$, 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 4.19 *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.

En general no se conoce ningún método efectivo para determinar si un polinomio dado es o no irreducible, ni mucho menos para encontrar los factores irreducibles de un polinomio dado. 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 4.20 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 4.21 *(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 2.17, que nos da dos polinomios $q(x)$ y $r(x)$ tales 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 4.22 *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 4.23 *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, $f(x) = (x - c_1)f'(x)$.

Como c_2 es raíz de f tenemos que $0 = f(c_2) = (c_2 - c_1)f'(c_2)$ y como las raíces son distintas $c_2 - c_1 \neq 0$. Como A es un dominio íntegro ha de ser $f'(c_2) = 0$, luego por el teorema anterior $f'(x) = (x - c_2)f''(x)$, de donde $f(x) = (x - c_1)(x - c_2)f''(x)$.

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 el polinomio $x^2 + 1$ tiene al menos tres raíces en el anillo D de los cuaterniones racionales.

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:

Teorema 4.24 *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 $c = \frac{a}{b}$, con $a, b \in D$. Si $c \notin D$ entonces $b \nmid a$, luego existe un primo $p \in D$ tal que $e_p(a) < e_p(b)$. 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:

$$a^n = -d_{n-1}ba^{n-1} - \cdots - d_1b^{n-1}a - d_0b^n.$$

Ahora bien, el exponente de p en el miembro izquierdo es exactamente $ne_p(a)$, mientras que en el miembro derecho es estrictamente mayor que $ne_p(a)$, con lo que tenemos una contradicción. ∎

Ejercicio: Probar que un número entero tiene una raíz n–sima en $\mathbb{Q}$ si y sólo si la tiene en $\mathbb{Z}$.

Ejercicio: Probar que las raíces enteras de un polinomio de $\mathbb{Z}[x]$ dividen a su término independiente.

Terminaremos con una aplicación más sofisticada del criterio de Eisenstein. Para ello necesitaremos un resultado sencillo que, según veremos en el capítulo siguiente, tiene una consecuencia muy importante en la teoría de anillos.

Teorema 4.25 *Sea p un número natural 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, $p \mid \binom{p}{m+1}$. ∎

Consideremos el polinomio $x^n - 1 \in \mathbb{Z}[x]$. Evidentemente no es irreducible, porque tiene la raíz $x = 1$. Eso significa que es divisible entre $x - 1$. No es difícil llegar a que

$$x^n - 1 = (x-1)(x^{n-1} + x^{n-2} + \cdots + x + 1).$$

Vamos a probar que si p es un número natural primo, entonces el polinomio $p(x) = x^{p-1} + x^{p-2} + \cdots + x + 1$ es irreducible en $\mathbb{Z}[x]$ (o equivalentemente en $\mathbb{Q}[x]$).

Por el teorema 4.19 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 anterior p divide a todos los coeficientes de $p(x+1)$ salvo al correspondiente a x^p, 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.

Capítulo V

Congruencias y anillos cociente

Las congruencias son una herramienta muy potente en el estudio de los números enteros. La solución de muchos problemas de muy diversa índole, así como algunas propiedades generales sobre los anillos, encuentran su expresión más natural en términos de congruencias. En este capítulo veremos muchos ejemplos de ello. Los anillos cociente proporcionan una visión alternativa más conceptual de la noción de congruencia.

5.1 Definiciones básicas

Definición 5.1 Consideremos un dominio A y un ideal I. Diremos que dos elementos a y b de A son *congruentes* módulo I, abreviado $a \equiv b \,(\text{mód } I)$, 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.

Consideremos el caso concreto de $\mathbb{Z}$. En general diremos que dos elementos de un anillo son congruentes módulo un tercero si lo son módulo el ideal que éste genera, por lo que en DIPs como $\mathbb{Z}$ no necesitamos mencionar ideales. En definitiva dos enteros a y b son congruentes módulo n si $n \mid a - b$. Dado un entero a, podemos dividirlo entre n de modo que $a = nc + r$, con $0 \leq r < n$. Resulta que $a - r = nc$, luego $a \equiv r \,(\text{mód } n)$. 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. Por ejemplo $\mathbb{Z}/5\mathbb{Z}$ está formado por las clases siguientes:

$$\begin{array}{cccccccccccc}
\{ & \ldots & -20, & -15, & -10, & -5, & 0, & 5, & 10, & 15, & 20, & \ldots & \} \\
\{ & \ldots & -19, & -14, & -9, & -4, & 1, & 6, & 11, & 16, & 21, & \ldots & \} \\
\{ & \ldots & -18, & -13, & -8, & -3, & 2, & 7, & 12, & 17, & 22, & \ldots & \} \\
\{ & \ldots & -17, & -12, & -7, & -2, & 3, & 8, & 13, & 18, & 23, & \ldots & \} \\
\{ & \ldots & -16, & -11, & -6, & -1, & 4, & 9, & 14, & 19, & 24, & \ldots & \}
\end{array}$$

La primera clase la forman los múltiplos de 5, la segunda los números que al ser divididos entre 5 dan resto 1, etc. Si llamamos $[a]$ a la clase de equivalencia de a tenemos por ejemplo que $[4] = [-1]$, $\mathbb{Z}/5\mathbb{Z} = \big\{[0],\ [1],\ [2],\ [3],\ [4]\big\}$.

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 5.2 *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 dominio 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

$$([a] + [b]) + [c] = [a + b] + [c] = [a + b + c] = [a] + [b + c] = [a] + ([b] + [c]).$$

∎

Ejercicio: Construir las tablas de la suma y el producto en $\mathbb{Z}/5\mathbb{Z}$.

De esta forma nos hemos encontrado con una colección de anillos finitos. Estos anillos resultan utilísimos para determinar si un número dado divide o no a otro. En efecto, se cumple que $m \mid n$ si y sólo si $[n] = 0$ módulo m, y esto es fácil de calcular. Por ejemplo, tomando clases módulo 7:

$$\begin{aligned}
[234] &= [2][10][10] + [3][10] + [4] = [2][3][3] + [3][3] + [4] \\
&= [6][3] + [2] + [4] = [-1][3] + [6] = [3]
\end{aligned}$$

Esto prueba que el resto de 234 entre 7 es 3, y por tanto no es múltiplo de 7. Hemos hecho muchas operaciones, pero todas muy sencillas. Con números pequeños no siempre es más rápido, pero con números grandes es muy práctico.

Así podemos obtener criterios sencillos de divisibilidad. Normalmente escribimos los números en base 10, es decir,

$$1.247 = 1 \cdot 1.000 + 2 \cdot 100 + 4 \cdot 10 + 7 = 1 \cdot 10^3 + 2 \cdot 10^2 + 4 \cdot 10 + 7.$$

Si en general tenemos un número de la forma $a = \sum_{i=0}^{n} c_i 10^i$, al tomar clases módulo 2 se cumple

$$[a] = \sum_{i=0}^{n} [c_i][10]^i = \sum_{i=0}^{n} [c_i][0]^i = [c_0][0]^0 = [c_0],$$

luego el número a será múltiplo de 2 si y sólo si $[a] = 0$, si y sólo si $[c_0] = 0$, es decir, si y sólo si su última cifra es múltiplo de 2. En resumen:

Un número natural n es múltiplo de 2 si y sólo si acaba en 0, 2, 4, 6 u 8.

Con lo que a simple vista sabemos que 6.278 es par mientras que 29.953 es impar.

Tomemos ahora clases módulo 3. La clave es que $[10] = [1]$:

$$[a] = \sum_{i=0}^{n} [c_i][10]^i = \sum_{i=0}^{n} [c_i][1]^i = \sum_{i=0}^{n} [c_i] = [\sum_{i=0}^{n} c_i].$$

Podemos enunciar el resultado así:

Un número natural es múltiplo de 3 si y sólo si la suma de sus cifras lo es.

Por ejemplo, para saber si 3.725 es múltiplo de 3 sumamos $3+7+2+5 = 17$ y repetimos la operación: $1 + 7 = 8$. Como 8 no es múltiplo de 3, concluimos que 3.725 tampoco lo es. De hecho al dividirlo entre 3 su resto es el mismo que el de 8, o sea, 2.

Si consideramos clases módulo 4 resulta que $[10] = 2$, luego $[100] = [2][2] = 0$, y todas las potencias de 10 superiores a 10 dan resto cero. La regla que se obtiene es:

Un número natural n es múltiplo de 4 si y sólo si el número formado por sus dos últimas cifras es múltiplo de 4.

Por ejemplo, el año 2.494 no será bisiesto, ya que, módulo 4,

$$[94] = [9][2] + [4] = [1][2] + [0] = [2].$$

Como $[10] = 0$ módulo 5, ocurre lo mismo que con el 2: Un número es múltiplo de 5 si y sólo si su última cifra lo es o, equivalentemente,

Un número natural es múltiplo de 5 si y sólo si acaba en 0 o en 5.

No hay un criterio fácil para el 6 o el 7. En todo caso, para saber si un número es múltiplo de 6 basta ver si lo es de 2 y de 3. Como $[10] = [1]$ módulo 9, la

regla del 3 vale para el 9. El lector puede deducir una regla para el 10 y otra para el 11 (para el cual se cumple $[10] = -[1]$).

Si el lector se ha entretenido en construir tablas de anillos $\mathbb{Z}/n\mathbb{Z}$ se habrá encontrado ciertamente con algo interesante: los primeros ejemplos de anillos con divisores de cero. De hecho nos ha aparecido un caso al trabajar con los múltiplos de 4. En el anillo $\mathbb{Z}/4\mathbb{Z}$ se cumple que $[2][2] = [0]$, luego no es un dominio íntegro.

El lector experimentador puede obtener empíricamente los primeros valores de n para los cuales $\mathbb{Z}/n\mathbb{Z}$ sí es dominio íntegro: 2, 3, 5, 7, 11, ... A éstos hay que añadir el 0 porque el correspondiente anillo cociente es isomorfo a $\mathbb{Z}$, y no vale el 1 porque en la definición de dominio exigimos que $0 \neq 1$, mientras que $\mathbb{Z}/1\mathbb{Z}$ sólo tiene un elemento.

Si analizamos la razón por la que un anillo $\mathbb{Z}/n\mathbb{Z}$ no es íntegro vemos que en cada caso la explicación es siempre la misma: en $\mathbb{Z}/4\mathbb{Z}$ la clase $[2]$ es un divisor de 0 porque $2 \cdot 2 = 4$. En $\mathbb{Z}/6\mathbb{Z}$ las clases $[2]$ y $[3]$ son divisores de 0 porque $2 \cdot 3 = 6$, etc. En definitiva, que las factorizaciones de n dan lugar a divisores de cero. En resumen, un anillo $\mathbb{Z}/n\mathbb{Z}$ no puede ser íntegro a menos que n sea primo. El lector puede tratar de probar directamente que $\mathbb{Z}/p\mathbb{Z}$ es íntegro si p es primo. Nosotros vamos a dar una prueba general. 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 5.3 *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. El lector que tenga a mano sus tablas podrá conjeturar una respuesta a la siguiente pregunta: ¿De entre los $\mathbb{Z}/p\mathbb{Z}$, cuáles son cuerpos?

La respuesta es que todos lo son. De nuevo el lector es invitado a justificar este hecho. Ha de probar que en un anillo $\mathbb{Z}/p\mathbb{Z}$ con p primo todo elemento es una unidad. Por ejemplo, en $\mathbb{Z}/5\mathbb{Z}$ vemos que $[2]^{-1} = [3]$, pues $[2][3] = [6] = [1]$. El lector emprendedor debería recordar la relación de Bezout (4.12). Una vez más aquí vamos a ver una prueba general.

Teorema 5.4 *Sea A un anillo conmutativo y unitario con $1 \neq 0$ 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. 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 5.5 *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 $\{u_1, \ldots, u_n\} = \{uu_1, \ldots, uu_n\}$ 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, según veremos, tiene muchas consecuencias interesantes:

Teorema 5.6 *Sea A un dominio con un número finito de unidades n 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, $\mathbb{Z}/p\mathbb{Z}$ 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:

Teorema 5.7 *(Teorema de Fermat) Para todo primo p y todo número entero a, se cumple que $a^p \equiv a$ (mód p).*

5.2 Números perfectos

Veamos un ejemplo menos elemental del uso de las congruencias en el estudio de los números enteros. Los números se han relacionado desde siempre con el esoterismo y la adivinación. Así, el número 7 se ha considerado un número de

buena suerte, mientras el 13 era considerado nefasto. Un concepto que tiene su origen en esta clase de "teorías" es el de número perfecto. Un número es *perfecto* si es la suma de sus divisores distintos de él mismo (siempre considerando números naturales). Por ejemplo, los divisores de 6 son 1, 2 y 3 y se cumple que $6 = 1 + 2 + 3$. El número 6 es perfecto. Otro ejemplo de número perfecto es el $28 = 1 + 2 + 4 + 7 + 14$.

Definamos la función $\sigma(n)$ igual a la suma de los divisores de n. En estos términos un número n es perfecto si y sólo si $\sigma(n) = 2n$ (pues también sumamos el propio n).

Una función f de números naturales es *multiplicativa* si cuando $(m, n) = 1$ cumple $f(mn) = f(m)f(n)$. Las funciones multiplicativas aparecen frecuentemente en teoría de números, y de hecho acabamos de encontrarnos con una. Para probarlo hemos de tener en cuenta que si $(m, n) = 1$, entonces cada divisor de mn se expresa de forma única como el producto de un divisor de m y un divisor de n. En efecto, si $a \mid mn$, descomponemos el número a en un producto, el primero de cuyos factores contenga a los factores primos de a que dividen a m y el segundo a los que dividen a n. Por lo tanto

$$\sigma(mn) = \sum_{d|mn} d = \sum_{u|m}\sum_{v|n} uv = \sum_{u|m} u\left(\sum_{v|n} v\right) = \Big(\sum_{u|m} u\Big)\sigma(n) = \sigma(m)\sigma(n).$$

Por otra parte, si p es un primo es fácil calcular $\sigma(p^n) = 1 + p + \cdots + p^n$. Teniendo en cuenta la identidad $(1+x+x^2+\cdots+x^n)(x-1) = x^{n+1}-1$, resulta que

$$\sigma(p^n) = \frac{p^{n+1} - 1}{p - 1}$$

Así podemos calcular fácilmente la función σ. Por ejemplo,

$$\sigma(100) = \sigma(2^2)\sigma(5^2) = \frac{2^3 - 1}{2 - 1} \cdot \frac{5^3 - 1}{5 - 1} = 7 \cdot 31 = 217.$$

Euclides recoge en sus *Elementos* el hecho de que si el número $1+2+\cdots+2^n = 2^{n+1} - 1$ es primo, entonces el número $2^n(2^{n+1} - 1)$ es perfecto.

En efecto, como para los primos p se cumple que $\sigma(p) = p+1$, en nuestro caso tenemos que $\sigma(2^{n+1} - 1) = 2^{n+1}$, y como σ es multiplicativa $\sigma\big(2^n(2^{n+1} - 1)\big) = \sigma(2^n)\sigma(2^{n+1} - 1) = (2^{n+1} - 1) \cdot 2^{n+1} = 2 \cdot \big(2^n(2^{n+1} - 1)\big)$.

Descartes afirmó (y la primera prueba conocida se debe a Euler) que un número par es perfecto si y sólo si es de la forma indicada por Euclides.

Para verlo tomemos un número par, que lo podremos expresar en la forma $2^n m$, donde $n > 0$ y m es un número impar. Si es perfecto $\sigma(2^n m) = 2^{n+1}m$. Como σ es multiplicativa

$$(2^{n+1} - 1)\sigma(m) = 2^{n+1}m. \tag{5.1}$$

Como $2^{n+1} - 1$ es impar, el exponente de 2 en $\sigma(m)$ ha de ser $n + 1$, es decir, $\sigma(m) = 2^{n+1}a$ para cierto número natural a.

Sustituyendo en (5.1) obtenemos $(2^{n+1}-1)2^{n+1}a = 2^{n+1}m$, luego

$$(2^{n+1}-1)a = m. \tag{5.2}$$

Por lo tanto $\sigma(m) = 2^{n+1}a = m + a$. Ahora bien, si $a > 1$ resulta que 1, a y m son divisores distintos de m, luego $\sigma(m)$ debe ser al menos $1 + a + m$, contradicción. Concluimos que $a = 1$ y que $\sigma(m) = m+1$, lo que sólo es posible si m no tiene más divisores que 1 y m, es decir, si m es primo.

Sustituyendo $a = 1$ en (5.2) queda que $m = 2^{n+1} - 1$, luego el número original era $2^n(2^{n+1} - 1)$ con $2^{n+1} - 1$ primo, o sea, de la forma establecida por Euclides. Con un leve cambio de índice podemos enunciar así lo que hemos probado:

Un número par es perfecto si y sólo si es de la forma $2^{n-1}(2^n - 1)$ *y* $2^n - 1$ *es primo.*

Respecto a los números perfectos impares no se conoce ninguno, pero nadie ha demostrado que no existan. Es un problema abierto.

Ejercicio: Estudiar la función $d(n)$ = número de divisores de n.

Hasta aquí no hemos usado congruencias. Estas intervienen a la hora de determinar para qué valores de n se cumple que el número $2^n - 1$ es primo. El lector empírico ya debería haber empezado a recopilar datos. He aquí lo que puede obtenerse sin esforzarse uno mucho:

n	**2**	**3**	4	**5**	6	**7**	8	9	10	¿11?
$2^n - 1$	**3**	**7**	15	**31**	63	**127**	255	511	1023	¿2047?

La tabla muestra los primeros valores de $2^n - 1$. Es fácil ver que para $n = 2$, 3, 5 y 7 obtenemos primos (el 127 es primo porque no es divisible por primos menores que 12). Tampoco cuesta ver que los números restantes no lo son. Todos son múltiplos de 3 salvo el 511, que es divisible entre 7. El número correspondiente a 11 está entre interrogantes porque ya no es fácil decidir si es primo. Habría que tratar de dividirlo entre todos los primos menores que 45 y hay un total de 14. Dejemos eso para luego y centrémonos en los casos claros. Vemos que $2^n - 1$ no siempre es primo. ¿Conjetura algo el lector sobre en qué casos lo es? No hay que forzar mucho la imaginación para sospechar que $2^n - 1$ es primo exactamente cuando n lo es.

Tratemos de probar que $2^n - 1$ no es primo si n es compuesto. Para buscar un posible divisor a $2^n - 1$ observemos los divisores que hemos encontrado en nuestros ejemplos:

n	**2**	**3**	2^2	**5**	$2 \cdot 3$	**7**	2^3	3^2	$2 \cdot 5$
$2^n - 1$	**3**	**7**	$3 \cdot 5$	**31**	$3^2 \cdot 7$	**127**	$3 \cdot 5 \cdot 17$	$7 \cdot 73$	$3 \cdot 11 \cdot 31$

Si al lector no le basta esto para conjeturar qué divisores podemos encontrar en el número $2^n - 1$ cuando n no es primo, he aquí una pista más. La factorización siguiente no ha sido hallada por tanteo: $2^{14} - 1 = 16 \cdot 383 = 3 \cdot 43 \cdot 127$.

Es fácil encontrar el factor 3, pero ¿cómo se encuentran los factores 43 y 127 en el cociente 5.461?

La mayor pista está en el 127. El 127 es el primo correspondiente a $n = 7$ y 7 es uno de los factores de 14. El otro factor es 2 y su primo correspondiente el 3, que también divide a $2^{14} - 1$. También es claro el caso $n = 10$, cuyos factores son 2 y 5 y en $2^{10} - 1$ aparecen los primos correspondientes, 3 y 31. Vemos que también aparecen otros primos, como el 11 en $2^{10} - 1$ o el 5 y el 17 en $2^8 - 1$, pero eso no importa. La conjetura es que si $d \mid n$ entonces $2^d - 1 \mid 2^n - 1$, y la prueba es trivial si usamos congruencias: tomando clases módulo $2^d - 1$ se cumple que $[2]^d = [1]$, luego si $n = dm$

$$[2^n - 1] = \big([2]^d\big)^m - [1] = [1]^m - [1] = [0],$$

lo que prueba que en efecto $2^d - 1 \mid 2^n - 1$.

Con esto ya tenemos que para que $2^n - 1$ sea primo el número n ha de ser primo. Los números de la forma $2^p - 1$ con p primo se llaman *números de Mersenne*. La cuestión es si todos los números de Mersenne son primos. Antes hemos probado que los primos de Mersenne están en correspondencia con los números pares perfectos. He aquí los números perfectos que hemos encontrado:

p	2	3	5	7
$2^p - 1$	3	7	31	127
$2^{p-1}(2^p - 1)$	6	28	496	8.128

Según nuestras conjeturas el número $2^{11} - 1$ debería ser primo. El lector animoso puede tratar de dividirlo entre los 14 primos menores que 45. Aquí vamos a estudiar el problema pensando más para trabajar menos. Es evidente que $2^{11} - 1 = 2.047$ no es múltiplo de 2, 3 o 5. Vamos a ver si puede ser múltiplo de 7. No sólo vamos a probar que no lo es, sino que vamos a encontrar razones por las que no puede serlo, con lo que podremos descartar muchos más primos aparte del 7. El número $2^{11} - 1$ será múltiplo de 7 si y sólo si al tomar clases módulo 7 se cumple que $[2]^{11} = [1]$. Las potencias de 2 módulo 7 pueden calcularse recurrentemente $[2]^{n+1} = [2]^n[2]$, lo que nos permite reducir las potencias al tiempo que las calculamos: $[2]^2 = [4]$, $[2]^3 = [4][2] = [8] = [1]$, $[2]^4 = [1][2] = [2]$, etc. Así vamos obteniendo lo siguiente:

n	0	1	2	3	4	5	6	...
$[2]^n$	[1]	[2]	[4]	[1]	[2]	[4]	[1]	...

Vemos que el [1] se alcanza cíclicamente en los múltiplos de 3, luego no es necesario llegar hasta el 11: como 11 no es múltiplo de 3, $2^{11} - 1$ no puede ser múltiplo de 7.

Por lo tanto la razón por la que $2^{11} - 1$ no es múltiplo de 7 es que 11 no es múltiplo de 3. ¿De dónde ha salido ese 3? Es obvio que al calcular potencias de un elemento dado no nulo de $\mathbb{Z}/p\mathbb{Z}$, como es [2], tarde o temprano obtendremos un valor repetido, pues sólo hay un número finito de valores posibles, o sea, dado a en $\mathbb{Z}/p\mathbb{Z}$, existen números naturales $0 < m < n$ tales que $a^m = a^n = a^m \cdot a^{n-m}$,

luego $a^{n-m} = 1$. En resumen, ha de haber un natural no nulo d tal que $a^d = 1$. Si tomamos el d menor posible es obvio que el 1 se alcanzará exactamente en los múltiplos de d, que es lo que nos hemos encontrado.

En resumen, dado un primo p existe un número d tal que $[2]^n - 1$ es divisible entre p si y sólo si $d \mid n$. El lector puede calcular los valores de d para distintos primos. No es inmediato, pero tampoco excesivamente laborioso. Como siempre, el lector perezoso tiene a continuación una tabla:

p	3	5	7	11	13	17	19
d	2	4	3	10	12	8	18

¿Encuentra el lector alguna relación entre p y d? Fermat encontró una al estudiar el problema: $d \mid p - 1$. Esto es esencialmente el teorema de Fermat que hemos probado en la sección anterior. En efecto, hemos visto que $2^{p-1} \equiv 1 \pmod{p}$, y como 2^n vale 1 en los múltiplos de d, ha de ser $d \mid p - 1$.

Tenemos que para que $2^{11} - 1$ sea divisible entre p es necesario que el d correspondiente a p divida a 11, pero como d no puede ser 1 (esto es evidente), ha de ser $d = 11$, y por el teorema de Fermat $11 \mid p - 1$. En resumen, para que un primo p pueda dividir a $2^{11} - 1$ es necesario que sea de la forma $p = 11m + 1$.

Buscamos ahora los números de la forma $11m + 1$ menores que 45. Podemos ahorrarnos los valores impares de m porque dan números pares, que no pueden ser primos. Resulta que sólo hay un valor posible: $11 \cdot 2 + 1 = 23$. Acabamos de probar que si $2^{11} - 1$ no es primo, entonces es divisible entre 23.

Ahora calculamos módulo 23:

$$\begin{aligned} [2]^5 &= [32] = [9], \\ [2]^{10} &= [9]^2 = [9][3][3] = [27][3] = [4][3] = [12], \\ [2]^{11} &= [24] = [1]. \end{aligned}$$

Luego resulta que, después de todo, $2^{11} - 1$ no es primo. Concretamente factoriza como $23 \cdot 89$.

En realidad hoy en día no se sabe si hay un número finito o infinito de primos de Mersenne. De todos modos los primos de Mersenne no terminan en el $2^7 - 1$. Vamos a probar con el siguiente candidato: $2^{13} - 1 = 8.191$. Así veremos la potencia de la técnica que hemos desarrollado.

En primer lugar $2^{13} - 1 < 2^{14} = \left(2^7\right)^2 = 128^2$, luego nos basta buscar posibles divisores primos menores que 128.

Supongamos que $p \mid 2^{13} - 1$. Entonces módulo p se cumple que $[2]^{13} = [1]$. Si d es el menor natural no nulo tal que $[2]^d = [1]$ se cumple que $d \mid 13$, y por el teorema de Fermat $13 \mid p - 1$. Así pues $p = 13n + 1$. Más aún, como n ha de ser par necesariamente, $p = 26n + 1$.

Los valores de $26n + 1$ menores que 128 son $27, \mathbf{53}, \mathbf{79}, 105$. Sólo hay dos primos, lo que significa que si $2^{13} - 1$ no es primo entonces es divisible entre 53 o 79. Es fácil ver que no es el caso. Tomamos clases módulo 53:

$$[2]^6 = [64] = [11], \quad [2]^{12} = [11]^2 = [121] = [15], \quad [2]^{13} = [30] \neq [1].$$

Y ahora módulo 79:

$$\begin{array}{lcll} [2]^6 & = & [64] & = & [-15], \\ [2]^{12} & = & [-15]^2 & = & [-15][-5][3] = [75][3] = [-4][3] = [-12], \\ [2]^{13} & = & [-24] & \neq & [1]. \end{array}$$

Con ayuda de ordenadores se ha podido ampliar la lista de primos de Mersenne. Los 15 primeros corresponden a los valores de p dados en la tabla 5.1.

Tabla 5.1: Primos p tales que $2^p - 1$ es un primo de Mersenne

p	2, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, 127, 521, 607, 1.279, ...

Ejercicio: Probar que $n - 1 \mid n^k - 1$, con lo que si $k > 1$ el número $n^k - 1$ no puede ser primo salvo si $n = 2$.

5.3 Unidades

Ahora obtendremos algunos resultados útiles sobre las unidades de un anillo. Muchos de ellos, especialmente 5.13 son inmediatos a partir de los resultados de la teoría de grupos finitos, pero preferimos dar aquí pruebas directas que no nos desvíen de la teoría de anillos. Empezamos estudiando en particular el anillo $\mathbb{Z}/n\mathbb{Z}$. ¿Cuántas unidades tiene? Para responder a esta pregunta conviene responder antes a otra más pretenciosa: ¿cuáles son?

El lector debería construirse tablas y conjeturar algo. Hemos visto que los divisores de cero en $\mathbb{Z}/n\mathbb{Z}$ se deben a las factorizaciones de n, es decir, si $d \mid n$, entonces $[d]$ es un divisor de cero, luego no es una unidad. Por otra parte un múltiplo no nulo de un divisor de cero es también un divisor de cero, luego basta con que d y n tengan un divisor común para que $[d]$ sea un divisor de cero. Por ejemplo, en $\mathbb{Z}/8\mathbb{Z}$ la clase $[6]$ no es unidad, pues como $(6, 8) = 2$, resulta que $[6][4] = [0]$, donde el 4 sale de que $8 = 2 \cdot 4$.

Esto significa que para que $[m]$ sea una unidad de $\mathbb{Z}/n\mathbb{Z}$ hace falta que $(m, n) = 1$. La condición es suficiente, pues si $(m, n) = 1$, por la relación de Bezout existen ciertos enteros r y s tales que $rm + sn = 1$, de donde $[1] = [r][m] + [s][0] = [r][m]$.

Resumiendo:

Teorema 5.8 *El conjunto de las unidades de $\mathbb{Z}/n\mathbb{Z}$ es*

$$U_n = \big\{[m] \mid (m, n) = 1\big\}.$$

Definición 5.9 Llamaremos $\phi(n)$ al número de números naturales menores que n y primos con n, o sea, al número de unidades de $\mathbb{Z}/n\mathbb{Z}$. La función ϕ se llama *función de Euler.*

Una versión más general del teorema de Fermat (consecuencia directa del teorema 5.6) es que si a y n son números enteros primos entre sí, entonces $a^{\phi(n)} \equiv 1$ (mód n). Es fácil calcular los primeros valores de ϕ:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$\phi(n)$	0	1	2	2	4	2	6	4	6	4	10	4	12	6	8	8

¿Encuentra el lector alguna regularidad? La hay. La función de Euler es multiplicativa. Lo probaremos a partir de un hecho general de interés. Notemos que si A y B son anillos el producto $A \times B$ es también un anillo con las operaciones definidas componente a componente, es decir,

$$\begin{aligned}(a,b)+(a',b') &= (a+a',b+b'),\\ (a,b)(a',b') &= (aa',bb').\end{aligned}$$

Igualmente se puede definir el producto de una familia finita de anillos. En general esta construcción no es muy útil porque los anillos producto nunca son dominios íntegros, pero ahora nos va a servir para comprender la estructura de los anillos $\mathbb{Z}/n\mathbb{Z}$.

Teorema 5.10 *Sean $m_1, \ldots, m_n$ números naturales primos entre sí dos a dos. Sea $m = m_1 \cdots m_n$. Entonces la aplicación*

$$f : \mathbb{Z}/m\mathbb{Z} \longrightarrow (\mathbb{Z}/m_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/m_n\mathbb{Z})$$

dada por $f\big([a]\big) = \big([a], \ldots, [a]\big)$ es un isomorfismo de anillos.

Demostración: Es inmediato comprobar que está bien definida. Más aún, es inyectiva, pues si $\big([a], \ldots, [a]\big) = \big([b], \ldots, [b]\big)$ entonces cada $m_i \mid b-a$, y como son primos entre sí es claro que $m \mid b-a$, es decir, $[a] = [b]$ (módulo m).

Como los dos anillos tienen m elementos podemos concluir que f es un isomorfismo. ∎

Es fácil ver que las unidades de un producto $A \times B$ son los pares (u, v) donde u es una unidad en A y v una unidad en B. Por lo tanto, si m y n son enteros primos entre sí el isomorfismo entre $\mathbb{Z}/mn\mathbb{Z}$ y $\mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ hace corresponder los elementos de U_{mn} con los de $U_m \times U_n$, y esto prueba que $\phi(mn) = \phi(m)\phi(n)$.

Esta propiedad reduce el cálculo de la función de Euler a las potencias de primos, pero es fácil ver que $\phi(p^n) = (p-1)p^{n-1}$ (pues los números menores que p^n y que no son primos con p^n son los p^{n-1} múltiplos de p). Así por ejemplo, para calcular $\phi(45)$ basta hacer $\phi(45) = \phi(3^2)\phi(5) = 2 \cdot 3 \cdot 4 = 24$.

El teorema 5.10 tiene un enunciado clásico, conocido por los chinos desde hace más de 1.500 años.

Teorema 5.11 *(Teorema chino del resto) Sean $m_1, \ldots, m_n$ números naturales primos entre sí dos a dos y sean $c_1, \ldots, c_n$ enteros cualesquiera. Entonces las congruencias $x_i \equiv c_i$ (mód m_i), para $i = 1, \ldots, n$ tienen una solución común única módulo $m = m_1 \cdots m_n$.*

Ejercicio: Probar que si $a \equiv b \pmod{d}$ y $d = (m, n)$, entonces las congruencias

$$x \equiv a \pmod{m} \qquad \text{y} \qquad x \equiv b \pmod{n}$$

tienen una solución común única módulo el m.c.m. de m y n.

Todavía hay algo importante que podemos decir de las unidades de una dominio íntegro en general cuando éstas son un número finito.

Definición 5.12 Sea A un dominio con un número finito de unidades. Sea u una unidad de A. Llamaremos *orden* de u al menor natural no nulo tal que $u^n = 1$. Lo representaremos $\mathrm{o}_A(u)$. El teorema 5.6 garantiza la existencia de $\mathrm{o}_A(u)$.

Si n y m son números enteros primos entre sí, llamaremos $\mathrm{o}_n(m)$ al orden de la clase $[m]$ en $\mathbb{Z}/n\mathbb{Z}$.

Por ejemplo, el número d que considerábamos al estudiar los primos de Mersenne, no es sino $d = \mathrm{o}_p(2)$. Como vimos entonces, si s es un número entero y $s = \mathrm{o}_A(u)c + r$ con $0 \le r < \mathrm{o}_A(u)$, entonces $u^s = \left(u^{\mathrm{o}_A(u)}\right)^c u^r = u^r \neq 1$ salvo que $r = 0$, luego $u^s = 1$ si y sólo si $\mathrm{o}_A(u) \mid s$.

Más en general, dos potencias coinciden $u^i = u^j$ si y sólo si $u^{j-i} = 1$, si y sólo si $\mathrm{o}_A(u) \mid j - i$, si y sólo si $i \equiv j \pmod{\mathrm{o}_A(u)}$.

Esto significa que si u es una unidad de un dominio con un número finito de unidades, entonces, dada una unidad u de A, las potencias $u^0, u, u^2, u^3, \ldots$ se repiten cíclicamente, de modo que 1 aparece exactamente en las potencias de exponente divisible entre $\mathrm{o}_A(u)$.

El teorema 5.6 afirma que $\mathrm{o}_A(u)$ es un divisor del número total de unidades.

He aquí unos cuantos ejemplos de órdenes de unidades en distintos anillos $\mathbb{Z}/n\mathbb{Z}$:

Para $n = 5$, $\phi(5) = 4$

m	1	2	3	4
$\mathrm{o}_5(m)$	1	4	4	2

Para $n = 7$, $\phi(7) = 6$

m	1	2	3	4	5	6
$\mathrm{o}_7(m)$	1	3	6	3	6	2

Para $n = 8$, $\phi(8) = 4$

m	1	3	5	7
$\mathrm{o}_8(m)$	1	2	2	2

Para $n = 9$, $\phi(9) = 6$

m	1	2	4	5	7	8
$\mathrm{o}_9(m)$	1	6	3	6	3	2

Para $n = 11$, $\phi(11) = 10$

m	1	2	3	4	5	6	7	8	9	10
$\mathrm{o}_{11}(m)$	1	10	5	5	5	10	10	10	5	2

Observamos que cuando n es primo existen elementos de todos los órdenes posibles, es decir, de todos los órdenes que dividen al número de unidades. En general no es cierto, como muestra el ejemplo $n = 8$. Vamos a probar este hecho importante.

Teorema 5.13 *Sea A un dominio íntegro con un número finito n de unidades. Entonces para cada divisor m de n existe una unidad en A de orden m.*

DEMOSTRACIÓN: Probemos que si u es una unidad de A, $n = \mathrm{o}_A(u)$, y $m \mid n$, entonces $\mathrm{o}_A(u^m) = n/m$.

En efecto, $(u^m)^{n/m} = u^n = 1$, luego $\mathrm{o}_A(u^m) \mid n/m$. Por otro lado, si $(u^m)^r = 1$, entonces $n \mid mr$, luego $n/m \mid r$, y así $\mathrm{o}_A(u^m) = n/m$.

Si u, v son unidades de A, $\mathrm{o}_A(u) = r$, $\mathrm{o}_A(v) = s$ y $(r, s) = 1$, entonces $\mathrm{o}_A(uv) = rs$.

En efecto, $(uv)^{rs} = (u^r)^s (v^s)^r = 1 \cdot 1 = 1$, luego $\mathrm{o}_A(uv) \mid rs$. Si $(uv)^m = 1$, entonces $u^m v^m = 1$, luego $u^m = v^{-m}$ y $\mathrm{o}_A(u^m) = \mathrm{o}_A(v^{-m})$. Como $(u^m)^r = 1$, tenemos $\mathrm{o}_A(u^m) \mid r$ e igualmente $\mathrm{o}_A(v^{-m}) \mid s$, luego $\mathrm{o}_A(u^m) = \mathrm{o}_A(v^{-m}) = 1$, o sea, $u^m = v^{-m} = 1$ y por lo tanto $r \mid m$ y $s \mid m$. Como $(r, s) = 1$, podemos concluir que $rs \mid m$. Así pues, rs es el menor natural no nulo que hace 1 a uv al calcular la exponencial, es decir, $\mathrm{o}_A(uv) = rs$.

Supongamos ahora que u y v son unidades de A, $\mathrm{o}_A(u) = r$, $\mathrm{o}_A(v) = s$, pero no necesariamente $(r, s) = 1$. Descompongamos r y s en potencias de primos y formemos como sigue dos números r' y s': el número r' es el producto de todas las potencias de primos que dividen a r con exponente mayor o igual que a s, mientras que s' está formado por el producto de las potencias de primos que dividen a s con exponente mayor estrictamente que a r. De este modo $r' \mid r$, $s' \mid s$, $r's' = \mathrm{mcm}(r, s)$ y $(r', s') = 1$.

Así $\mathrm{o}_A(u^{r/r'}) = r'$ y $\mathrm{o}_A(v^{s/s'}) = s'$, luego $\mathrm{o}_A(u^{r/r'} v^{s/s'}) = r's' = \mathrm{mcm}(r, s)$.

Es decir, si existe una unidad de orden r y otra de orden s, existe una tercera de orden $\mathrm{mcm}(r, s)$. Aplicando esto varias veces obtenemos una unidad u tal que su orden m es múltiplo de los órdenes de todas las unidades de A. Sabemos que $m = \mathrm{o}_A(u) \mid n$, pero por otro lado, toda unidad de A es raíz del polinomio $x^m - 1$, y no puede haber más de m raíces, con lo que $n \leq m$. Así pues, $\mathrm{o}_A(u) = n$.

Para cada divisor m de n, el elemento $u^{n/m}$ tiene orden m. ∎

Definición 5.14 Una unidad en un dominio cuyo orden sea igual al número de unidades se llama una *raíz primitiva de la unidad.*

En general, *una raíz de la unidad* de un dominio A es un elemento u de A que cumple $u^n = 1$ para cierto entero n no nulo (en particular es una unidad

porque $u \cdot u^{n-1} = 1$). El teorema 5.6 afirma que si hay un número finito de unidades, todas ellas son raíces de la unidad. El teorema anterior prueba que si además el dominio es íntegro, entonces existen raíces primitivas de la unidad. En el caso concreto de los anillos $\mathbb{Z}/p\mathbb{Z}$ las raíces primitivas de la unidad (o mejor los enteros cuyas clases son raíces primitivas de la unidad) se llaman *raíces primitivas* módulo p. El nombre de raíces primitivas se debe a que todas las demás unidades se obtienen de ellas por exponenciación.

Por ejemplo, una raíz primitiva módulo 7 es 3 y sus potencias módulo 7 son

$$[3]^0 = [1],\ [3]^1 = [3],\ [3]^2 = [2],\ [3]^3 = [6],\ [3]^4 = [4],\ [3]^5 = [5],\ [3]^6 = [1],$$

éstos son todos los elementos de $\mathbb{Z}/7\mathbb{Z}$ (salvo [0], que no es una unidad).

No hay criterios para obtener raíces primitivas módulo p, pero sabemos que existen y podemos encontrarlas. Se pueden dar algunas condiciones que facilitan la búsqueda, pero en la práctica es preferible consultar tablas de raíces primitivas o usar ordenadores que las calculen. He aquí la lista de las mínimas raíces primitivas para los menores primos:

Tabla 5.2: Raíces primitivas (mód p)

p	r	p	r	p	r	p	r	p	r
2	1	13	2	31	3	53	2	73	5
3	2	17	3	37	2	59	2	79	3
5	2	19	2	41	6	61	2	83	2
7	3	23	5	43	3	67	2	89	3
11	2	29	2	47	5	71	7	97	5

5.4 Homomorfismos y anillos cociente

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 5.15 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 5.16 *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, $b = a$. ∎

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 $\mathrm{N}(p) = I$.

El próximo teorema nos dice que todo epimorfismo es esencialmente de este tipo.

Teorema 5.17 *(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.*

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 $\bar{f}\big([a]\big) = \bar{f}([b])$, entonces $f(a) = f(b)$, $f(a - b) = 0$, $a - b \in \mathrm{N}(f)$, luego $[a] = [b]$. ∎

Con ayuda de estos conceptos vamos a profundizar un poco en la estructura de los anillos.

Definición 5.18 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.

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 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 0 o 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.

Si A es un anillo unitario, la aplicación $f : \mathbb{Z} \longrightarrow A$ dada por $f(n) = n1$ es claramente un homomorfismo. Su núcleo es precisamente $(\operatorname{car} A)$. Si $\operatorname{car} A = 0$ lo que tenemos es que $f[\mathbb{Z}]$ es un subanillo de A isomorfo a $\mathbb{Z}$. Podemos identificar ambos anillos, es decir, identificar el número entero m con el elemento $m1$ de A.

Si $\operatorname{car} A = n > 0$ el teorema de isomorfía nos da que $\mathbb{Z}/n\mathbb{Z}$ es isomorfo a $f[\mathbb{Z}]$, luego podemos identificar cada $m1$ con la clase $[m]$ de $\mathbb{Z}/n\mathbb{Z}$.

En la práctica identificaremos siempre m con $m1$, es decir, 3 será lo mismo que el elemento $1 + 1 + 1$ de A, con la única precaución de que si por ejemplo $\operatorname{car} A = 5$, entonces $7 = 12$ cuando los consideramos como elementos de A.

Si K es un cuerpo de característica 0, K no sólo contiene a $\mathbb{Z}$, sino que también contiene un subcuerpo isomorfo a su cuerpo de cocientes, es decir, a $\mathbb{Q}$. También en este caso identificaremos a $\mathbb{Q}$ con un subcuerpo de K. Por ejemplo, $5/3$ será lo mismo que el elemento de K

$$\frac{1+1+1+1+1}{1+1+1}.$$

De este modo, los anillos $\mathbb{Z}$ y $\mathbb{Z}/n\mathbb{Z}$ son los más pequeños en sus respectivas características, pues están contenidos en cualquier anillo unitario de la misma característica.

Si nos centramos en dominios íntegros tenemos que los menores son $\mathbb{Z}$ y los cuerpos $\mathbb{Z}/p\mathbb{Z}$ y si consideramos cuerpos, resulta que todo cuerpo K contiene un subcuerpo isomorfo a $\mathbb{Z}/p\mathbb{Z}$ o a $\mathbb{Q}$. A dicho cuerpo se le llama *cuerpo primo* de K, y como está construido a partir de la identidad mediante sumas y cocientes, es claro que está contenido en cualquier subcuerpo de K.

Veamos un par de teoremas que muestran por qué la característica de un anillo es un dato a tener en cuenta.

Teorema 5.19 *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 4.25:

$$(a+b)^p = \sum_{k=0}^{p} \binom{p}{k} a^k b^{p-k} = a^p + b^p + \sum_{k=1}^{p-1} \binom{p}{k} a^k b^{p-k} = a^p + b^p.$$

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, $(a \pm b)^{p^n} = a^{p^n} \pm b^{p^n}$. Otro resultado en el que la característica es relevante es el siguiente:

Teorema 5.20 *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

$$\eta = \frac{-b \pm \sqrt{b^2-4ac}}{2a}. \tag{5.3}$$

Si existe $\sqrt{b^2-4ac} \in K$, es fácil ver que (5.3) son raíces de $p(x)$. ∎

5.5 Cocientes de anillos de polinomios

Terminamos el capítulo mostrando cómo los anillos cociente nos dan un método para construir anillos que contengan elementos con características predeterminadas. Antes de ver el caso general estudiaremos un ejemplo concreto. Vamos a construir un anillo en el que un elemento tenga cuadrado -5.

Partimos del polinomio x^2+5, que es irreducible en $\mathbb{Q}[x]$ (porque es de grado 2 y no tiene raíces en $\mathbb{Z}$). Por lo tanto el ideal (x^2+5) es maximal y el anillo cociente $\mathbb{Q}[x]/(x^2+5)$ es un cuerpo. Vamos a llamar $\sqrt{-5}=[x]$ y $\mathbb{Q}\big[\sqrt{-5}\,\big]=\mathbb{Q}[x]/(x^2+5)$.

Como $[x^2+5]=[0]$, se cumple que $[x]^2=-[5]$, o sea, $(\sqrt{-5})^2=-[5]$.

Si $p(x)\in\mathbb{Q}[x]$, dividimos $p(x)=(x^2+5)c(x)+a+bx$. Entonces

$$[p(x)]=[x^2+5][c(x)]+[a]+[b][x]=[0]+[a]+[b]\sqrt{-5}.$$

Es decir, hemos probado que $\mathbb{Q}\big[\sqrt{-5}\,\big]=\big\{[a]+[b]\sqrt{-5}\mid a,b\in\mathbb{Q}\big\}$.

Notemos que si $[a]+[b]\sqrt{-5}=[c]+[d]\sqrt{-5}$, entonces $[a-c+(b-d)x]=[0]$, luego $a-c+(b-d)x\in(x^2+5)$ y así $x^2+5\mid a-c+(b-d)x$. Pero la única forma en que un polinomio de grado 2 puede dividir a otro de grado ≤ 1 es que éste sea nulo, o sea, $a-c+(b-d)x=0$, y consecuentemente $a=c$, $b=d$.

No hay confusión si escribimos $a+b\sqrt{-5}$ en lugar de $[a]+[b]\sqrt{-5}$.

Llamemos $\mathbb{Z}\big[\sqrt{-5}\,\big]=\big\{m+n\sqrt{-5}\mid m,n\in\mathbb{Z}\big\}$. Es inmediato comprobar que $\mathbb{Z}\big[\sqrt{-5}\,\big]$ es un subanillo de $\mathbb{Q}\big[\sqrt{-5}\,\big]$, obviamente un dominio íntegro.

Llamamos norma de un elemento de $\mathbb{Z}\big[\sqrt{-5}\,\big]$ al número natural

$$\mathrm{N}(m+n\sqrt{-5})=(m+n\sqrt{-5})(m-n\sqrt{-5})=m^2+5n^2.$$

Una simple comprobación nos da que $\mathrm{N}(uv)=\mathrm{N}(u)\,\mathrm{N}(v)$, para todo par de elementos u y v de $\mathbb{Z}\big[\sqrt{-5}\,\big]$. Por ejemplo, $\mathrm{N}(4+\sqrt{-5})=21$. De aquí se deduce que $4+\sqrt{-5}$ es irreducible en $\mathbb{Z}\big[\sqrt{-5}\,\big]$.

En efecto: no puede ser una unidad, pues si u es unidad $1=\mathrm{N}(1)=\mathrm{N}(uu^{-1})=\mathrm{N}(u)\,\mathrm{N}(u^{-1})$, luego $\mathrm{N}(u)=1$, de donde $u=\pm1$.

Además si $4+\sqrt{-5}=uv$, entonces $\mathrm{N}(u)\,\mathrm{N}(v)=21$, luego $\mathrm{N}(u)=1$, 3, 7 o 21. Pero $\mathrm{N}(u)=m^2+5n^2=3$ es imposible, pues si $n\neq0$ tenemos $m^2+5n^2\geq5$ y si $n=0$ queda $m^2=3$, y tampoco puede ser, o sea, $\mathrm{N}(u)\neq3$ para cualquier u de $\mathbb{Z}\big[\sqrt{-5}\,\big]$. Tampoco puede ser $\mathrm{N}(u)=7$ porque entonces $\mathrm{N}(v)=3$, luego ha de ser $\mathrm{N}(u)=1$ o $\mathrm{N}(u)=21$, en cuyo caso $\mathrm{N}(v)=1$, es decir, $u=\pm1$ o bien $v=\pm1$.

Del mismo modo se prueba que $4-\sqrt{-5}$ es irreducible (tiene la misma norma) y de forma similar lo son 3 y 7. Ahora bien:

$$3\cdot7=21=(4+\sqrt{-5})(4-\sqrt{-5}),$$

Así pues, un elemento de $\mathbb{Z}\big[\sqrt{-5}\,\big]$, el 21, admite dos descomposiciones distintas en irreducibles (notar que como las únicas unidades son ±1, el único asociado de 3 es -3). Esto prueba que $\mathbb{Z}\big[\sqrt{-5}\,\big]$ no es DFU. Los cuatro elementos considerados son ejemplos de irreducibles que no son primos.

Posiblemente el lector considere artificial la prueba anterior. Ello se debe a que hemos usado técnicas que estudiaremos con detalle más adelante, de modo

que en su momento nos resultarán naturales. El lector puede repetir todos los pasos de la construcción anterior con el polinomio x^2+1. Llamando $i=[x]$ obtenemos un cuerpo $\mathbb{Q}[i]=\mathbb{Q}[x]/(x^2+1)$ con la propiedad de que $i^2=-1$. Así mismo se cumple $\mathbb{Q}[i]=\{a+bi \mid a,b\in\mathbb{Q}\}$, de modo que $a+bi=c+di$ si y sólo si $a=c$ y $b=d$.

En realidad ya hemos estudiado una copia isomorfa de este cuerpo. Podemos ver a $\mathbb{Q}[i]$ como un subcuerpo del anillo de los cuaterniones racionales estudiado al final del capítulo I.

Definimos la norma en $\mathbb{Q}[i]$ como $\mathrm{N}(a+bi)=(a+bi)(a-bi)=a^2+b^2$, y también es fácil ver que $\mathrm{N}(uv)=\mathrm{N}(u)\,\mathrm{N}(v)$ (de hecho es un caso particular de la propiedad correspondiente en los cuaterniones).

Si nos restringimos al anillo $\mathbb{Z}[i]=\{a+bi \mid a,b\in\mathbb{Z}\}$, la norma toma valores naturales. Este anillo se conoce como anillo de los *enteros de Gauss,* y su comportamiento es completamente distinto al de $\mathbb{Z}\big[\sqrt{-5}\,\big]$. En efecto, vamos a ver que es un dominio euclídeo, y por lo tanto un DFU. La norma euclídea será la norma que acabamos de definir. Obviamente cumple la propiedad 1 de la definición 1.6.

Sean α y β dos enteros de Gauss, $\beta\neq 0$. Entonces $\alpha/\beta=a+bi\in\mathbb{Q}[i]$. Sean m y n los enteros más próximos a a y b respectivamente. Entonces $|a-m|\leq 1/2$, $|b-n|\leq 1/2$. Llamemos $\gamma=m+ni\in\mathbb{Z}[i]$ y $r=\alpha-\beta\gamma$. Así $\alpha=\beta\gamma+r$, con

$$\mathrm{N}(r)=\mathrm{N}(\alpha-\beta\gamma)=\mathrm{N}\Big(\frac{\alpha}{\beta}-\gamma\Big)\,\mathrm{N}(\beta)=((a-m)^2+(b-n)^2)\,\mathrm{N}(\beta)\leq\frac{\mathrm{N}(\beta)}{2}<\mathrm{N}(\beta).$$

Vemos así cómo la norma permite estudiar eficientemente las propiedades de los enteros de Gauss, por ejemplo, como en el caso anterior las unidades son los elementos de norma 1, lo que en este caso nos lleva a que hay cuatro unidades: ±1 y $\pm i$.

Otro ejemplo, como $\mathrm{N}(1+i)=2$, el mismo argumento del ejemplo anterior nos lleva a que $1+i$ es irreducible, luego primo. De hecho $2=(1+i)(1-i)$ es la descomposición en primos de 2. Notemos que los factores son asociados, pues $1-i=-i(1+i)$.

Finalmente damos un teorema general que recoge estos ejemplos.

Teorema 5.21 *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. Entonces el ideal $\big(p(x)\big)$ es maximal en $k[x]$, luego el anillo cociente $K=k[x]/\big(p(x)\big)$ es un cuerpo. La aplicación $\phi:k\longrightarrow K$ dada por $\phi(a)=[a]$ es un monomorfismo de cuerpos, pues si $[a]=0$, entonces $p(x)\mid a$, luego $a=0$ (o sea, $\mathrm{N}(f)=0$). Por lo tanto podemos considerar que k está contenido en K. Llamamos $\alpha=[x]\in K$. De este modo, si $p(x)=\sum_{i=0}^n a_ix^i$, 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 el caso en que $p(x)$ es irreducible el cuerpo K que hemos construido tiene una estructura fácil de describir. En efecto, si $\alpha=[x]$ es la raíz de $p(x)$, entonces

todo elemento de K es de la forma $[q(x)] = q(\alpha)$ (por el mismo argumento con el que hemos probado que $p(\alpha) = 0$). Así pues,

$$K = k[\alpha] = \big\{q(\alpha) \mid q(x) \in k[x]\big\}.$$

Con más exactitud, si el grado de $p(x)$ es n, todo polinomio $q(x)$ puede expresarse como $q(x) = p(x)c(x) + r(x)$, con grad $r(x) < n$, y $q(\alpha) = r(\alpha)$, es decir, los elementos de K son de la forma

$$a_{n-1}\alpha^{n-1} + \cdots + a_1\alpha + a_0, \qquad a_i \in k.$$

Más aún, esta expresión es única, pues restando dos expresiones de este tipo obtenemos una expresión del tipo $t(\alpha) = 0$, donde grad $t(x) \leq n - 1$, luego $[t(x)] = 0$, luego $p(x) \mid t(x)$, luego $t(x) = 0$, luego las dos expresiones eran la misma.

Ejercicio: Definir y estudiar el anillo $\mathbb{Z}\big[\sqrt{-2}\big]$. Probar que es un dominio euclídeo.

Capítulo VI

Algunas aplicaciones

En este capítulo estudiaremos algunos problemas clásicos de la teoría de números que ilustren la relación entre la teoría general de anillos que hemos estudiado y los problemas concretos sobre números enteros, a la vez que motiven las técnicas y conceptos algebraicos más avanzados que introduciremos en los temas siguientes. Comenzamos con un problema que se remonta a la matemática griega.

6.1 Ternas pitagóricas

Diofanto trató en su Aritmética el problema de encontrar ternas de números naturales no nulos x, y, z tales que $x^2 + y^2 = z^2$. Estas ternas se llaman *ternas pitagóricas,* pues según el teorema de Pitágoras permiten construir triángulos rectángulos con lados enteros. Los egipcios las usaban para construir ángulos rectos en arquitectura. Entre los ejemplos más conocidos están $3^2 + 4^2 = 5^2$, $5^2 + 12^2 = 13^2$, $7^2 + 24^2 = 25^2$. ¿Cómo encontrarlas todas?

En primer lugar notamos que si (x, y, z) es una terna pitagórica, también lo es (mx, my, mz) para cualquier número m y, recíprocamente, dada una terna pitagórica (x, y, z), podemos dividir sus componentes por su m.c.d. para obtener otra que cumpla además $(x, y, z) = 1$. Una terna cuyos elementos no tengan divisores comunes se llama *primitiva.* Si encontramos un método para hallar todas las ternas primitivas, las restantes se obtienen multiplicándolas por números arbitrarios, luego el problema está resuelto. Las ternas anteriores son todas primitivas.

Ante todo notemos que un divisor primo de dos de las componentes de una terna pitagórica, divide a la tercera. Por ejemplo, si $p \mid x$ y $p \mid z$, entonces $p \mid z^2 - x^2$, con lo que $p \mid y^2$ y por lo tanto $p \mid y$. Esto significa que, en realidad, las componentes de una terna pitagórica primitiva son primas entre sí dos a dos. En particular no puede haber más de una componente par. Un número es par o impar si y sólo si lo es su cuadrado, y la suma y la diferencia de números impares es par. Como consecuencia si dos de las componentes son impares, la restante

Tabla 6.1: Ternas pitagóricas

p	q	x	y	z
2	1	4	3	5
3	2	12	5	13
4	1	8	15	17
4	3	24	7	25
5	2	20	21	29
5	4	40	9	41
6	1	12	35	37
6	5	60	11	61
7	2	28	45	53
7	4	56	33	65
7	6	84	13	85

ha de ser par, es decir, en una terna primitiva hay siempre dos componentes impares y una par.

Ahora veamos que z ha de ser impar. En otro caso lo son x e y, es decir, $x = 2m+1$, $y = 2n+1$, luego $x^2 = 4m^2+4m+1$, $y^2 = 4n^2+4n+1$. Al tomar clases módulo 4 resulta que $[z]^2 = [x]^2 + [y]^2 = [1] + [1] = [2]$. Sin embargo ninguna clase módulo 4 tiene a [2] por cuadrado: $[0]^2 = [0]$, $[1]^2 = [1]$, $[2]^2 = [0]$, $[3]^2 = [1]$.

Como la situación de x e y es simétrica, podemos suponer que x es par e y impar. Según lo visto z es también impar. Consecuentemente $z+y$, $z-y$ son ambos pares. Digamos que $x = 2u$, $z+y = 2v$, $z-y = 2w$.

Ahora $x^2 = z^2 - y^2 = (z+y)(z-y)$, luego $u^2 = vw$, $v > 0$, $w > 0$.

Por otro lado $(v, w) = 1$, ya que si un primo p divide a ambos, entonces

$$\begin{aligned} p \;\mid\; (v+w) &= \frac{1}{2}(z+y) + \frac{1}{2}(z-y) = \frac{1}{2}\,2z = z, \\ p \;\mid\; (v-w) &= \frac{1}{2}(z+y) - \frac{1}{2}(z-y) = y, \end{aligned}$$

y como $(y, z) = 1$, esto es contradictorio.

Por la factorización única, es claro que si $vw = u^2$ con $(v, w) = 1$, $v > 0$, $w > 0$, entonces tanto v como w han de ser cuadrados (cada uno ha de contener cada primo un número par de veces porque así le ocurre a u). Pongamos $v = p^2$ y $w = q^2$. Obviamente $(p, q) = 1$.

Así tenemos que $z = v + w = p^2 + q^2$, $y = v - w = p^2 - q^2$. En particular $q < p$.

Como z e y son impares, p y q deben tener paridad opuesta. Sustituyendo en las fórmulas anteriores queda

$$x^2 = z^2 - y^2 = p^4 + 2p^2q^2 + q^4 - p^4 + 2p^2q^2 - q^4 = 4p^2q^2 = (2pq)^2,$$

luego $x = 2pq$.

En consecuencia la terna original queda de la forma

$$(x, y, z) = (2pq, p^2 - q^2, p^2 + q^2),$$

donde p, q son números naturales primos entre sí, $q < p$ y de paridad opuesta.

Recíprocamente, es fácil comprobar que cualquier terna en estas condiciones es una terna pitagórica primitiva. Por lo tanto ya sabemos enumerarlas todas. La tabla 6.1 contiene las correspondientes a los valores de $p \leq 7$.

En una tablilla cuneiforme aproximadamente del año 1.500 a.C. se ha encontrado una enumeración de ternas pitagóricas, entre las cuales se encontraba $(4.961, 6.480, 8.161)$. Se obtiene con $p = 81$ y $q = 40$.

Ejercicio: Comprobar que toda terna pitagórica contiene un múltiplo de 3, un múltiplo de 4 y un múltiplo de 5.

6.2 Sumas de dos cuadrados

Una pregunta relacionada con el problema anterior es ¿qué números naturales pueden expresarse como suma de dos cuadrados? Antes de teorizar sobre ellos echemos una ojeada a las sumas de los primeros diez cuadrados. Los primos están en negrita.

0	1	4	9	16	25	36	49	64	81	100
	2	**5**	10	**17**	26	**37**	50	65	82	**101**
		8	**13**	20	**29**	40	**53**	68	85	104
			18	25	34	45	58	**73**	90	**109**
				32	**41**	52	65	80	**97**	116
					50	**61**	74	**89**	106	125
						72	85	100	117	136
							98	**113**	130	**149**
								128	145	164
									162	**181**
										200

Antes de interpretar la tabla pensemos un poco qué debemos buscar en ella. Notemos que si $z = x^2 + y^2$ entonces para todo natural n se cumple también $n^2 z = (nx)^2 + (ny)^2$, es decir, que si multiplicamos una suma de cuadrados por un cuadrado, obtenemos otra suma de cuadrados. Se dice que un número n es *libre de cuadrados* si todos los primos que lo dividen tienen multiplicidad 1. Todo número n puede expresarse de forma única como producto de un cuadrado perfecto y de un número libre de cuadrados. Basta agrupar por un lado todos los pares de primos que lo dividen y por otro los primos que queden sin pareja. Lo que decíamos hace un momento es que si la parte libre de cuadrados de un número natural n es suma de dos cuadrados, entonces n también lo es. Lo primero que podemos observar en la tabla es que se cumple el recíproco, es decir, la parte libre de cuadrados de todos los números de la tabla está también en la tabla. Por ejemplo: la parte libre de cuadrados de 117 es 13. Esto nos lleva a

nuestra primera conjetura: Un número es suma de dos cuadrados si y sólo si lo es su parte libre de cuadrados.

Ahora quedémonos tan sólo con los números de nuestra lista que son libres de cuadrados, es decir:

$$0,\ 1,\ \mathbf{2},\ \mathbf{5},\ 10,\ \mathbf{13},\ \mathbf{17},\ 26,\ \mathbf{29},\ 34,\ \mathbf{37},\ \mathbf{41},\ \mathbf{53},\ 58,$$

$$\mathbf{61},\ 65,\ \mathbf{73},\ 74,\ 82,\ 85,\ \mathbf{89},\ \mathbf{97},\ \mathbf{101},\ 106,\ \mathbf{109},\ \mathbf{113}.$$

Esta lista contiene todos los números libres de cuadrados menores que 121 (el primer número que falta en la tabla) que pueden expresarse como suma de dos cuadrados.

Vemos que no están todos los primos, pero los primos que aparecen en los compuestos están también en la lista: $10 = 2 \cdot 5$, $26 = 2 \cdot 13$, $34 = 2 \cdot 17$, $58 = 2 \cdot 29$, $65 = 5 \cdot 13$, $74 = 2 \cdot 37$, $82 = 2 \cdot 41$, $85 = 5 \cdot 17$, $106 = 2 \cdot 53$.

La conjetura es, pues, que un número libre de cuadrados es suma de dos cuadrados si y sólo si lo son los primos que lo componen. Respecto a los primos que son suma de dos cuadrados nos quedan los siguientes

$$2,\ 5,\ 13,\ 17,\ 29,\ 37,\ 41,\ 53,\ 61,\ 73,\ 89,\ 97,\ 101,\ 109,\ 113.$$

Si el lector no conjetura nada a simple vista, no cuesta mucho encontrar una condición que han de cumplir. Si un primo impar es de la forma $p = x^2 + y^2$, entonces x e y han de tener paridades opuestas, digamos $x = 2u$, $y = 2v + 1$. Como consecuencia resulta que $p = 4u^2 + 4v^2 + 4v + 1$, es decir, $p = 4m + 1$. Si el lector calcula los primeros primos de la forma $4n + 1$ obtendrá precisamente la lista anterior (sin el 2, claro).

Sospechamos entonces que un primo impar es suma de dos cuadrados si y sólo si es congruente con 1 módulo 4. La conjetura general sería entonces que un número es suma de dos cuadrados si y sólo si los primos impares que lo dividen con multiplicidad impar son congruentes con 1 módulo 4. Vamos a probar que la conjetura es exacta.

Consideremos en primer lugar la fórmula

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2.$$

Esta fórmula no sale del aire, sino que es simplemente la forma explícita de expresar que la norma de un producto de elementos del cuerpo $\mathbb{Q}[i]$ es igual al producto de las normas. En este contexto nos garantiza que un producto de números expresables como suma de dos cuadrados es también expresable como suma de dos cuadrados. Ya hemos comentado que el recíproco no es totalmente cierto, pero algo podemos probar.

Veamos que si $a = pb$ donde p es primo y tanto a como p son suma de dos cuadrados, entonces b también lo es.

Para ello observemos que si $p = u^2 + v^2$ y $a = r^2 + s^2$, entonces

$$\begin{aligned}(us - rv)(us + rv) &= u^2s^2 - r^2v^2 = u^2s^2 + u^2r^2 - u^2r^2 - r^2v^2\\ &= u^2(s^2 + r^2) - r^2(u^2 + v^2) = u^2a - r^2p,\end{aligned}$$

luego $p \mid (us - rv)(us + rv)$ y por tanto $p \mid (us - rv)$ o bien $p \mid (us + rv)$.

Si $p \mid (us + rv)$, entonces, como $p^2 \mid ap$ y

$$ap = (r^2 + s^2)(u^2 + v^2) = (ru - sv)^2 + (rv + us)^2,$$

resulta que $p^2 \mid (ru - sv)^2$, luego $b = \frac{a}{p} = \frac{(ru-sv)^2}{p^2} + \frac{(rv+us)^2}{p^2}$, suma de cuadrados.

Si $p \mid (us - rv)$ razonamos igual con la fórmula

$$ap = (s^2 + r^2)(u^2 + v^2) = (su - rv)^2 + (sv + ur)^2.$$

Esto implica que si un número a, expresable como suma de dos cuadrados, es divisible por un número b que no lo es, entonces el cociente tiene un factor primo no expresable como suma de dos cuadrados, pues tendríamos $a = bc$ y si todos los factores primos de c fueran expresables como suma de dos cuadrados, una aplicación repetida del resultado anterior nos daría que b es también expresable como suma de dos cuadrados.

Ahora viene el resultado fundamental: si p es un primo tal que $(a, b) = 1$ y $p \mid (a^2 + b^2)$, entonces p es suma de dos cuadrados. Notar que si r es libre de cuadrados y $r = a^2 + b^2$, entonces $(a, b) = 1$, pues si $d \mid a$ y $d \mid b$, entonces $d^2 \mid r$. Por lo tanto un número libre de cuadrados es suma de dos cuadrados si y sólo si sus factores primos lo son.

En efecto, sea $a = pm \pm c$, $b = pn \pm d$, donde $|c|$, $|d| \leq p/2$ (si el resto de la división resulta mayor que $p/2$ sumamos 1 al cociente y tomamos resto negativo).

Entonces $a^2 + b^2 = m^2p^2 \pm 2mpc + c^2 + n^2p^2 \pm 2npd + d^2 = Ap + (c^2 + d^2)$. En consecuencia $p \mid (c^2 + d^2)$, o sea, $c^2 + d^2 = py$, para cierto y. Como $(c, d) < p$, p no lo divide, luego $(c, d)^2 \mid y$.

Ahora dividimos la ecuación $c^2 + d^2 = py$ hasta obtener $e^2 + f^2 = pz$, donde $(e, f) = 1$ y $pz \leq c^2 + d^2 \leq (p/2)^2 + (p/2)^2 = p^2/2$, luego $z \leq p/2$.

Si p no fuera suma de dos cuadrados, por el resultado anterior z tiene un factor primo q que tampoco es expresable como suma de dos cuadrados. En particular $q \leq z < p$.

Hemos probado que si existen números p, a, b, tales que p es primo, $(a, b) = 1$ y p divide a $a^2 + b^2$, entonces existen números q, e, f, en las mismas condiciones y con $q < p$. Pero si existieran tales ternas de números debería haber una con p mínimo, y según lo visto es imposible.

Esto prueba la mayor parte de nuestra conjetura: Sea $n = u^2v$ donde v es libre de cuadrados y supongamos que n es suma de dos cuadrados, $n = a^2 + b^2 = (a, b)^2(c^2 + d^2)$ con $(c, d) = 1$. Entonces $(a, b) \mid u$, luego $v \mid (c^2 + d^2)$. Por el resultado que acabamos de probar todo primo que divide a v (luego a $c^2 + d^2$) es suma de dos cuadrados, luego v es suma de dos cuadrados.

En resumen, tenemos probado que un número es suma de dos cuadrados si y sólo si lo es su parte libre de cuadrados, si y sólo si los primos que dividen a su parte libre de cuadrados son suma de dos cuadrados. Sólo falta probar que los únicos primos impares expresables como suma de dos cuadrados son exactamente los congruentes con 1 módulo 4. La necesidad ya está probada. Veamos la suficiencia. Necesitamos un hecho curioso:

Consideremos las cuartas potencias de los primeros números naturales:

$$1,\ \ 16,\ \ 81,\ \ 256,\ \ 625,\ \ 1.296,\ \ 2.401,\ \ 4.096,\ \ 6.561,\ \ 10.000.$$

Ahora calculemos las diferencias entre cada número obtenido y su anterior:

$$15,\ \ 65,\ \ 175,\ \ 369,\ \ 671,\ \ 1.105,\ \ 1.695,\ \ 2.465,\ \ 3.439.$$

Otra vez:

$$50,\ \ 110,\ \ 194,\ \ 302,\ \ 434,\ \ 590,\ \ 770,\ \ 974.$$

Y otra vez:

$$60,\ \ 84,\ \ 108,\ \ 132,\ \ 156,\ \ 180,\ \ 204.$$

Y a la cuarta vez obtenemos

$$24,\ \ 24,\ \ 24,\ \ 24,\ \ 24,\ \ 24.$$

Si el lector parte de otro exponente n distinto de cuatro llegará a un resultado similar. Unos cuantos ensayos le llevarán a convencerse de que la sucesión obtenida se vuelve constante a partir del n–simo paso y que la constante que aparece es concretamente el número $n!$ ¿Sabría dar una prueba?

Veamos, para ello, que si $p(x)$ es un polinomio con coeficientes enteros y de grado n no nulo, entonces $p(x+1)-p(x)$ es un polinomio de grado $n-1$ cuyo coeficiente director es n veces el coeficiente director de $p(x)$.

En efecto, sea $p(x)=\sum_{i=0}^{n} a_i x^i$. Así

$$p(x+1)-p(x)=\sum_{i=0}^{n} a_i(x+1)^i-\sum_{i=0}^{n} a_i x^i.$$

Cada polinomio $(x+1)^i$ tiene grado i, luego el único monomio de grado n que aparece en $p(x+1)$ es el de $a_n(x+1)^n$, o sea, $a_n x^n$, que se anula con el monomio correspondiente de $p(x)$, luego el grado de $p(x+1)-p(x)$ es a lo sumo $n-1$. Ahora calculemos el monomio de grado $n-1$.

En $\sum_{i=0}^{n} a_i(x+1)^i$ tenemos dos sumandos con grado $\geq n-1$, a saber, $a_n(x+1)^n$ y $a_{n-1}(x+1)^{n-1}$. El monomio de grado $n-1$ en cada uno de ellos es $na_n x^{n-1}$ y $a_{n-1}x^{n-1}$, respectivamente, pero el último se cancela con el correspondiente monomio de $p(x)$. Por tanto el monomio de grado $n-1$ en $p(x+1)-p(x)$ es exactamente $na_n x^{n-1}$, con lo que ciertamente se trata de un polinomio de grado $n-1$ con coeficiente director na_n.

En consecuencia, si partimos del polinomio x^n, las diferencias sucesivas son los valores que toma el polinomio $(x+1)^n-x^n$, que es un polinomio de grado $n-1$ con coeficiente director n, las siguientes diferencias vienen dadas por un polinomio de grado $n-2$ y coeficiente director $n(n-1)$, luego las diferencias n–simas vienen dadas por un polinomio de grado 0, o sea, constante y con coeficiente director igual a $n!$, o sea, todas son iguales a $n!$

Con ayuda de este hecho vamos a probar la conjetura que teníamos pendiente, según la cual si un número primo es congruente con 1 módulo 4, entonces es suma de dos cuadrados.

En efecto, sea $p = 4n + 1$. Por el teorema de Fermat sabemos que todo número a entre 1 y $p-1$ cumple $[a]^{p-1} = 1$ módulo p, es decir, $[a]^{4n} = 1$, luego $[a+1]^{4n} - [a]^{4n} = 0$, es decir, $p \mid (a+1)^{4n} - a^{4n}$ para $1 \leq a \leq 4n-1$.

Por otra parte $(a+1)^{4n} - a^{4n} = ((a+1)^{2n} + a^{2n})((a+1)^{2n} - a^{2n})$, luego o bien $p \mid (a+1)^{2n} + a^{2n}$ o bien $p \mid (a+1)^{2n} - a^{2n}$.

Si $p \mid (a+1)^{2n} + a^{2n}$ para algún número a, entonces, como $(a, a+1) = 1$, sabemos que p es suma de dos cuadrados. Veamos que no es posible que p no divida a ninguno de estos números, es decir, que no es posible que se cumpla $p \mid (a+1)^{2n} - a^{2n}$ para todo número a entre 1 y $4n-1$. En tal caso p divide a las $4n-2$ diferencias de las potencias $2n$–simas de los $4n-2$ primeros números enteros, luego también a las $4n-3$ diferencias de sus diferencias, etc. y así debería dividir a las $2n$ diferencias de orden $2n$, que valen $(2n)!$, pero como p es primo, resulta que p divide a un $m \leq 2n$, lo cual es imposible ya que $p = 4n+1$.

Lo que hemos visto es un ejemplo típico de una demostración de teoría de números de finales del siglo XVIII. Ahora vamos a probar los mismos resultados con las técnicas de principios del siglo XIX (siempre —por supuesto— con notación moderna).

La norma en el anillo $\mathbb{Z}[i]$ viene dada por $\mathrm{N}(a+bi) = a^2 + b^2$ (ver el capítulo anterior), luego un número n es suma de dos cuadrados si y sólo si existe un $u \in \mathbb{Z}[i]$ tal que $\mathrm{N}(u) = n$.

Sabemos que $\mathbb{Z}[i]$ es un dominio de factorización única. Notemos que

$$a + bi \mid (a+bi)(a-bi) = \mathrm{N}(a+bi),$$

es decir, todo entero de Gauss divide a su norma.

Si π es un entero de Gauss primo, entonces $\pi \mid \mathrm{N}(\pi)$ y $\mathrm{N}(\pi)$ es un número natural que se descompone en producto de primos en $\mathbb{Z}$, luego π ha de dividir a uno de esos primos. Es decir, existe un primo p tal que $\pi \mid p$, luego $\mathrm{N}(\pi) \mid p^2$, luego $\mathrm{N}(\pi) = p$ o $\mathrm{N}(\pi) = p^2$.

Si $\mathrm{N}(\pi) = p^2 = \mathrm{N}(p)$ entonces p y π son asociados (pues π divide a p y el cociente tiene norma 1, luego es una unidad), luego $\pi = \pm p, \pm pi$, y en particular concluimos que p es primo en $\mathbb{Z}[i]$. A su vez esto implica que no hay primos de norma p, ya que tal primo ρ debería dividir a p, pero al ser p primo ρ sería asociado de p, luego su norma sería p^2 y no p.

En resumen, un primo de Gauss π tiene norma p (y entonces p es suma de dos cuadrados) o bien tiene norma p^2 (y entonces no hay primos de norma p, luego p no es suma de dos cuadrados).

De aquí se siguen la mayoría de los hechos que hemos probado antes: un número n es suma de dos cuadrados si y sólo si existe un entero de Gauss u tal que $\mathrm{N}(u) = n$. Si descomponemos $u = \pi_1 \cdots \pi_r$ en producto de primos queda $n = \mathrm{N}(\pi_1) \cdots \mathrm{N}(\pi_r)$.

Ahora es obvio que si un primo p divide a n con exponente impar, uno de los factores $N(\pi_i)$ ha de ser igual a p (si sólo hubiera del tipo p^2 el exponente en n

sería par), luego p es suma de dos cuadrados. Recíprocamente, si los primos que dividen a n con exponente impar son suma de dos cuadrados, podemos formar un entero de Gauss u de norma n (multiplicando primos de Gauss de norma p si el exponente de p en n es impar y primos naturales p si el exponente es par).

En resumen: un número es suma de dos cuadrados si y sólo si los primos que lo dividen con exponente impar son sumas de dos cuadrados. Teniendo en cuenta que la parte libre de cuadrados de un número es el producto de los primos que lo dividen con exponente impar, esto implica todo lo que habíamos probado excepto una cosa: que los primos que son suma de dos cuadrados son 2 y los congruentes con 1 módulo 4. En términos de enteros de Gauss esto significa que hay enteros de Gauss de norma p si y sólo si $p = 2$ o bien $p \equiv 1$ (mód 4).

La prueba que hemos visto de que los primos impares $p = x^2 + y^2$ son congruentes con 1 módulo 4 es trivial, por lo que no tiene sentido buscar una alternativa. Por el contrario, vamos a dar una prueba "más moderna" del recíproco.

Supongamos que $p \equiv 1$ (mód 4). Hemos de probar que p es suma de dos cuadrados o, equivalentemente, que p no es primo en $\mathbb{Z}[i]$ (pues esto ya implica que p factoriza en dos primos de norma p, luego es suma de dos cuadrados). A su vez, para ello basta probar que existe un $n \in \mathbb{Z}$ tal que $-1 \equiv n^2$ (mód p). En efecto, en tal caso $p \mid n^2 + 1 = (n + i)(n - i)$, pero claramente $p \nmid n \pm i$ en $\mathbb{Z}[i]$, luego p no puede ser primo.

En otros términos, hemos de probar que -1 tiene raíz cuadrada en $\mathbb{Z}/p\mathbb{Z}$. Para ello consideramos el polinomio $x^{(p-1)/2} - 1$, entre cuyas raíces están todos los cuadrados no nulos de $\mathbb{Z}/p\mathbb{Z}$ (por el teorema de Fermat 5.7). Por otra parte, en $\mathbb{Z}/p\mathbb{Z}$ hay $(p - 1)/2$ cuadrados no nulos ($x^2 = y^2$ si y sólo si $y = \pm x$, y $x \neq -x$). Como un polinomio tiene a lo sumo tantas raíces en un cuerpo como indica su grado, concluimos que las raíces de $x^{(p-1)/2} - 1$ son exactamente los elementos con raíz cuadrada. Ciertamente, si $p \equiv 1$ (mód 4) tenemos que -1 es una de dichas raíces.

Con lo visto podemos comprender la utilidad de los conceptos algebraicos tales como el anillo de enteros de Gauss. La prueba que hemos obtenido con ellos es mucho más sencilla, no requiere apenas cálculos, es por tanto más fácil de recordar y explica realmente por qué la propiedad de ser suma de dos cuadrados depende sólo de la parte libre de cuadrados. En muchos casos, las técnicas algebraicas permiten llegar a resultados que serían imposibles mediante meros cálculos. Más adelante veremos más ejemplos.

6.3 Sumas de cuatro cuadrados

No estamos en condiciones de estudiar qué numeros pueden expresarse como suma de tres cuadrados (aunque el lector podría conjeturar fácilmente cuáles son), pero vamos a demostrar un importante resultado debido a Lagrange:

Teorema 6.1 *Todo número natural es suma de cuatro cuadrados.*

DEMOSTRACIÓN: El problema se reduce a estudiar los números primos en cuanto pensamos en el anillo de los cuaterniones racionales y en el hecho de que la norma del producto es el producto de las normas. Al escribirlo explícitamente obtenemos la fórmula:

$$\begin{aligned}&(x^2+y^2+z^2+w^2)(x'^2+y'^2+z'^2+w'^2)\\ &=(xx'-yy'-zz'-ww')^2+(xy'+yx'+zw'-wz')^2\\ &+(xz'+zx'+wy'-yw')^2+(xw'+wx'+yz'-zy')^2.\end{aligned}$$

Esto nos dice que el producto de números expresables como suma de cuatro cuadrados es también expresable como suma de cuatro cuadrados, luego basta probar que todo número primo es expresable como suma de cuatro cuadrados.

Por razones técnicas vamos a necesitar la variante que resulta de sustituir x' por $-x'$ en la fórmula anterior:

$$\begin{aligned}&(x^2+y^2+z^2+w^2)(x'^2+y'^2+z'^2+w'^2)\\ &=(-xx'-yy'-zz'-ww')^2+(xy'-yx'+zw'-wz')^2\\ &+(xz'-zx'+wy'-yw')^2+(xw'-wx'+yz'-zy')^2.\end{aligned}\tag{6.1}$$

Como $2=1^2+1^2+0^2+0^2$, basta probar que todo primo impar p es suma de cuatro cuadrados.

Los números x^2 con $0\le x\le\frac{1}{2}(p-1)$ son incongruentes módulo p, e igualmente ocurre con $-1-y^2$ con $0\le y\le\frac{1}{2}(p-1)$.

Como en total son $p+1$, existen x, y en estas condiciones tales que

$$x^2\equiv -1-y^2 \text{ (mód } p).\tag{6.2}$$

Entonces $x^2<(\frac{1}{2}p)^2$, $y^2<(\frac{1}{2}p)^2$, luego $x^2+y^2<2(\frac{1}{2}p)^2$,

$$x^2+y^2+1<1+2\Big(\frac{1}{2}p\Big)^2<p^2,\tag{6.3}$$

Por (6.2) $x^2+y^2+1\equiv 0$ (mód p), luego $x^2+y^2+1=mp<p^2$ (por (6.3)), de donde $0<m<p$.

Sea r el menor natural no nulo tal que existen números enteros x, y, z, w que cumplan $rp=x^2+y^2+z^2+w^2$. Como m cumple esto, será $r\le m<p$.

Necesariamente r es impar, pues si fuera par, 0, 2 o 4 de los x, y, z, w serían pares y, reordenándolos, podríamos exigir que $x+y$, $x-y$, $z+w$ y $z-w$ fueran pares.

Entonces $\frac{1}{2}rp=\big(\frac{1}{2}(x+y)\big)^2+\big(\frac{1}{2}(x-y)\big)^2+\big(\frac{1}{2}(z+w)\big)^2+\big(\frac{1}{2}(z-w)\big)^2$, en contradicción con la minimalidad de r.

Nuestro objetivo es probar que $r=1$. Supongamos que $r>1$.

Sean x', y', z', w' los restos módulo r de x, y, z, w entre $-r/2$ y $r/2$ (es posible ya que r es impar).

Claramente $n=x'^2+y'^2+z'^2+w'^2\equiv x^2+y^2+z^2+w^2=rp\equiv 0$ (mód r), pero $n>0$, pues en otro caso $x'=y'=z'=w'=0$, r dividiría x, y, z, w, luego

$r^2 \mid x^2+y^2+z^2+w^2 = rp$, de donde $r \mid p$ y en consecuencia $r = 1$, contra lo supuesto. También es claro que $n < 4(\frac{1}{2}r)^2 = r^2$.

Sea $0 < k < r$ tal que $n = kr$. Por la identidad (6.1), $krpr = z_1^2+z_2^2+z_3^2+z_4^2$ para ciertos naturales z_1, z_2, z_3, z_4 y, teniendo en cuenta cómo se obtienen a partir de x, y, z, w, x', y', z', w', es claro que los cuatro son múltiplos de r (por ejemplo $z_1 = -xx' - yy' - zz' - ww' \equiv x^2 - y^2 - z^2 - w^2 = -rp \equiv 0$ (mód r)).

Así pues $z_i = rt_i$ y por tanto $r^2kp = r^2t_1^2 + r^2t_2^2 + r^2t_3^2 + r^2t_4^2$, con lo que $kp = t_1^2 + t_2^2 + t_3^2 + t_4^2$, en contra de la minimalidad de r. ∎

6.4 Números de la forma $x^2 + 3y^2$

Fermat planteó, junto al estudio de los números expresables como suma de dos cuadrados, el problema más general de determinar qué números pueden expresarse en la forma $x^2 + 2y^2$, $x^2 + 3y^2$, etc. Ahora estudiaremos el caso $x^2 + 3y^2$ para destacar las analogías y las diferencias con el caso ya visto de las sumas de dos cuadrados. Comencemos examinando una tabla:

0	1	4	9	16	25	36	49	64	81	100
3	4	**7**	12	**19**	28	39	52	**67**	84	**103**
12	**13**	16	21	28	**37**	48	**61**	76	93	112
27	28	**31**	36	**43**	52	63	76	91	108	**127**
48	49	52	57	64	**73**	84	**97**	112	129	148
75	76	**79**	84	91	100	111	124	**139**	156	175

Como en el caso de las sumas de dos cuadrados, se observa que un número está en la tabla si y sólo si lo está su parte libre de cuadrados y que un número libre de cuadrados está en la tabla si y sólo si lo están los primos que lo componen.

Del mismo modo que 2 era un caso excepcional en el caso de la suma de dos cuadrados, aquí lo es el tres, pues si observamos los primos distintos de 3 que aparecen en la tabla:

$$7,\ 13,\ 19,\ 31,\ 37,\ 43,\ 61,\ 67,\ 73,\ 79,\ 97,$$

notamos que son los de la forma $6n + 1$.

La conjetura es, por tanto, que un número es de la forma $x^2 + 3y^2$ si y sólo si los primos que lo dividen con exponente impar diferentes de 3 son de la forma $6n + 1$.

Veamos en primer lugar la necesidad de la condición para los primos.

Supongamos que un primo $p \neq 3$ es de la forma $p = a^2 + 3b^2$. El número a no puede ser múltiplo de 3 o si no $3 \mid p$. Consecuentemente, módulo 6 el número a es igual a 1, 2, 4 o 5, luego a^2 ha de ser 1 o 4.

Por otro lado los cuadrados módulo 6 son 0, 1, 3 o 4, luego $3b^2$ ha de ser congruente con 0 o 3.

Ahora, la suma de un número congruente con 1 o 4 más un número congruente con 0 o 3 da un número congruente con 1 o 4 módulo 6, pero si $p = 6n+4$ entonces $2 \mid p$, contradicción. Así pues $p = 6n + 1$.

El hecho siguiente es que por la fórmula

$$(a^2 + 3b^2)(c^2 + 3d^2) = (ac - 3bd)^2 + 3(ad + bc)^2, \tag{6.4}$$

el producto de números expresables en la forma $x^2 + 3y^2$ es también de esta forma.

Probamos ahora que si 2 divide a un número de la forma $a^2 + 3b^2$ entonces de hecho $4 \mid a^2 + 3b^2$ y el cociente también es de esta forma.

En efecto, a y b deben ser ambos pares o ambos impares. Si ambos son pares, entonces $4 \mid a^2$ y $4 \mid b^2$, luego $4 \mid a^2 + 3b^2$ y $(a^2 + 3b^2)/4 = (a/2)^2 + 3(b/2)^2$.

Si ambos son impares $a = 4m \pm 1$ y $b = 4n \pm 1$. Escogiendo el signo de modo que los unos se cancelen resulta que $4 \mid a + b$ o bien $4 \mid a - b$.

Si $4 \mid a + b$ entonces, por (6.4),

$$4(a^2 + 3b^2) = (1^2 + 3 \cdot 1^2)(a^2 + 3b^2) = (a - 3b)^2 + 3(a + b)^2,$$

y $a - 3b = a + b - 4b$, luego $4 \mid a - 3b$ y así $16 \mid 4(a^2 + 3b^2)$, es decir, $4 \mid (a^2 + 3b^2)$ y $(a^2 + 3b^2)/4 = ((a - 3b)/4)^2 + 3((a + b)/4)^2$.

Si $4 \mid (a - b)$ se razona igual con la igualdad

$$4(a^2 + 3b^2) = (1^2 + 3(-1)^2)(a^2 + 3b^2) = (a + 3b)^2 + 3(a - b)^2.$$

Ahora probamos que si un número $a^2 + 3b^2$ es divisible por un primo de la forma $p = u^2 + 3v^2$, entonces el cociente también tiene esta forma.

Para ello observamos que

$$\begin{aligned}(ub - av)(ub + av) &= u^2b^2 - a^2v^2 = u^2b^2 + 3v^2b^2 - 3v^2b^2 - a^2v^2 \\ &= b^2(u^2 + 3v^2) - v^2(a^2 + 3b^2).\end{aligned}$$

Por lo tanto $p \mid (ub - av)(ub + av)$ y, como es primo, $p \mid (ub - av)$ o $p \mid (ub + av)$. Llamando $s = \pm 1$ según el caso tenemos que $p \mid (ub + sav)$ y (de nuevo por (6.4))

$$(u^2 + 3v^2)(a^2 + 3b^2) = (u^2 + 3(sv)^2)(a^2 + 3b^2) = (ua - s3vb)^2 + 3(ub + sav)^2,$$

luego $p \mid (ua - s3vb)^2$ y, por ser primo, $p^2 \mid (ua - s3vb)^2$, con lo que

$$\frac{a^2 + 3b^2}{p} = \left(\frac{ua - s3vb}{p}\right)^2 + 3\left(\frac{ub + sav}{p}\right)^2.$$

Como consecuencia, si un número $a^2 + 3b^2$ es divisible entre un número impar x que no es de esta forma, entonces el cociente tiene un factor impar que tampoco es de esta forma. En efecto, sea $a^2 + 3b^2 = xy$. Si $2 \mid y$, entonces hemos probado que $4 \mid y$ y además $x(y/4) = c^2 + 3d^2$. Podemos repetir hasta que $y/4^n$ sea impar, con lo que y es de la forma $y = p_1 \cdots p_m$, donde cada $p_i = 4$ o bien es un primo impar. Si no hubiera primos impares o todos fueran de la forma $u^2 + 3v^2$, entonces los dos resultados anteriores nos darían que x también es de la forma $r^2 + 3s^2$.

Si $(a,b)=1$ y un primo impar $p \mid a^2+3b^2$, entonces p es también de la forma $p=c^2+3d^2$.

En efecto, sean $a=mp\pm c$ y $b=np\pm d$, donde $|c|<p/2$ y $|d|<p/2$ (la desigualdad estricta es posible sólo si p es impar).

Módulo p se cumple que $a^2+3b^2=c^2+3d^2$, luego $p \mid c^2+3d^2$.

Sea $c^2+3d^2=py$. Notemos que $(c,d) \mid y$, pues en otro caso $p \mid (c,d)$, lo cual es imposible. Por consiguiente al dividir entre (c,d) queda $e^2+3f^2=pz$, con $(e,f)=1$, y si p no fuera de la forma pedida, entonces z tendría un factor primo impar q que tampoco sería de esta forma. Ahora, $pq \mid pz \mid e^2+3f^2 \mid c^2+3d^2$, luego $pq<(p/2)^2+3(p/2)^2=p^2$, con lo que $q<p$.

En resumen, si p no cumple lo pedido, hay un primo menor que tampoco lo cumple. Esto es imposible, pues ha de haber un mínimo primo en las condiciones de p.

Supongamos ahora que $n=a^2+3b^2$. Entonces, si n es par, $4 \mid n$ y $n/4$ es también de esta forma. Repitiendo, podemos expresar $n=4^i m$, donde m es impar. Esto significa que la parte libre de cuadrados de n es impar. Sea $n=u^2v$ con v libre de cuadrados e impar. Así $n=a^2+3b^2=(a,b)^2(c^2+3d^2)$ con $(c,d)=1$. Entonces $(a,b) \mid u$, luego $v \mid c^2+3d^2$. Por el resultado que acabamos de probar todo primo (impar) que divide a v (luego a c^2+3d^2) es suma de la forma indicada, luego v también.

Esto prueba que un número es de la forma a^2+3b^2 si y sólo si lo es su parte libre de cuadrados, y que un número libre de cuadrados es de esta forma si y sólo si lo son los primos que lo componen. Falta probar que los primos de esta forma son el 3 y los congruentes con 1 módulo 6.

Notar que un primo es de la forma $p=6n+1$ si y sólo si $p=3n+1$, pues $3n$ ha de ser par. Como en el caso de las sumas de dos cuadrados, los números a^{3n} son congruentes con 1 módulo p para $a=1,\dots,3n$ (por el teorema de Fermat). Por tanto p divide a las $3n$ diferencias $(a+1)^{3n}-a^{3n}$, para $a=1,\dots,3n-1$. Ahora bien,

$$\begin{aligned}(a+1)^{3n}-a^{3n} &= \big((a+1)^n-a^n\big)\big((a+1)^{2n}+(a+1)^n a^n+a^{2n}\big)\\ &= ((a+1)^n-a^n)(A^2+A(2B)+(2B)^2)\\ &= =((a+1)^n-a^n)(A+B)^2+3B^2,\end{aligned}$$

(donde hemos usado que o bien a, o bien $a+1$ es par).

Además $(a,a+1)=1$ implica que $(A,B)=1$. Si $p \mid (A+B)^2+3B^2$, entonces p es de la forma pedida. En otro caso $p \mid \big((a+1)^n-a^n\big)$ para todo $a=1,\dots,3n-1$, luego p divide a las $2n$ diferencias n–simas, que valen $n!$, lo que es imposible puesto que $p=3n+1$.

Los argumentos que hemos empleado recuerdan en muchos aspectos a los que usamos al tratar las sumas de dos cuadrados, pero sería ingenuo pensar que se trata de dos casos particulares de un mismo argumento general. En realidad el caso que acabamos de estudiar presenta dificultades adicionales intrínsecas, sin un equivalente en las sumas de dos cuadrados. Esto no puede verse en la maraña de cálculos que hemos empleado, y sólo se comprende adecuadamente

cuando se plantea en el contexto algebraico adecuado. En esta línea podríamos pensar en la factorización

$$x^2 + 3y^2 = (x + \sqrt{-3})(x - \sqrt{-3})$$

y tratar de simplificar la prueba utilizando el anillo

$$\mathbb{Z}\big[\sqrt{-3}\,\big] = \big\{a + b\sqrt{-3} \mid a, b \in \mathbb{Z}\big\},$$

del mismo modo que en el caso de las sumas de dos cuadrados hemos considerado el anillo de enteros de Gauss. Ciertamente este anillo está muy relacionado con el problema, pero nos encontramos con el obstáculo de que $\mathbb{Z}\big[\sqrt{-3}\,\big]$ no es un DFU. En efecto, como de costumbre, consideramos la norma definida por $\mathrm{N}(a+b\sqrt{-3}) = (a+b\sqrt{-3})(a-b\sqrt{-3}) = a^2+3b^2$. Como en los otros ejemplos se comprueba que es multiplicativa, así como que las unidades son los elementos de norma 1, lo que nos da únicamente ± 1. La factorización

$$(1 + \sqrt{-3})(1 - \sqrt{-3}) = 4 = 2 \cdot 2 \tag{6.5}$$

es un ejemplo de factorización no única, pues los cuatro factores tienen norma 4 y es fácil ver que no hay elementos de norma 2, por lo que los cuatro son irreducibles y no asociados.

Éste es el motivo que hace más complicado el problema. La razón por la que, pese a ello, el problema no es excesivamente más complicado es que puede probarse que la ecuación (6.5) es esencialmente el único caso de factorización no única en $\mathbb{Z}\big[\sqrt{-3}\,\big]$. En la sección siguiente estudiaremos esta factorización en relación con un problema similar al de la clasificación de las ternas pitagóricas.

Ejercicio: Obtener una conjetura sobre qué números son de la forma $a^2 + 2b^2$.

Ejercicio: Estudiar los números de la forma $a^2 + 5b^2$ y comprobar que no cumplen propiedades similares a las de los ejemplos anteriores.

6.5 La ecuación $x^2 + 3y^2 = z^3$

Aquí estudiaremos la factorización en el anillo $\mathbb{Z}\big[\sqrt{-3}\,\big]$ para probar el siguiente resultado técnico, que necesitaremos en la próxima sección.

Las soluciones enteras de la ecuación $x^2 + 3y^2 = z^3$ con $(x, y) = 1$ son exactamente las de la forma

$$x = a^3 - 9ab^2, \qquad y = 3a^2b - 3b^3 \qquad z = a^2 + 3b^2,$$

donde a y b son enteros primos entre sí.

Sea $\sigma : \mathbb{Q}\big[\sqrt{-3}\,\big] \longrightarrow \mathbb{Q}\big[\sqrt{-3}\,\big]$ la aplicación $\sigma(a + b\sqrt{-3}) = a - b\sqrt{-3}$. Es inmediato comprobar que se trata de un isomorfismo de anillos.

Sean a y b números cualesquiera y $x + y\sqrt{-3} = (a + b\sqrt{-3})^3$. Aplicando σ obtenemos que $x - y\sqrt{-3} = (a - b\sqrt{-3})^3$ y

$$\begin{aligned} x^2 + 3y^2 &= (x + y\sqrt{-3})(x - y\sqrt{-3}) = (a + b\sqrt{-3})^3(a - b\sqrt{-3})^3 \\ &= \big((a + b\sqrt{-3})(a - b\sqrt{-3})\big)^3 = \big(a^2 + 3b^2\big)^3. \end{aligned}$$

Así, $x^2+3y^2=z^3$, donde $z=a^2+3b^2$. Desarrollando $(a+b\sqrt{-3})^3$ por el teorema del binomio de Newton, obtenemos

$$a^3+3a^2b\sqrt{-3}+3ab^2\sqrt{-3}^{\,2}+b^3\sqrt{-3}^{\,3}=a^3-9ab^2+(3a^2b-3b^3)\sqrt{-3},$$

luego $x=a^3-9ab^2$, $y=3a^2b-3b^3$.

Esto prueba que las ternas $(a^3-9ab^2, 3a^2b-3b^3, a^2+3b^2)$ cumplen ciertamente la ecuación $x^2+3y^2=z^3$. Además si $(x,y)=1$ también $(a,b)=1$, o de lo contrario un entero no unitario dividiría a $a+b\sqrt{-3}$, luego a su cubo $x+y\sqrt{-3}$, luego a x y a y. Notar que una prueba directa que no use el anillo $\mathbb{Z}\big[\sqrt{-3}\,\big]$ es necesariamente mucho más complicada.

Además los cálculos que acabamos de hacer nos muestran que si tenemos $x^2+3y^2=z^3$ nos basta probar que $x+y\sqrt{-3}=(a+b\sqrt{-3})^3$ para ciertos enteros a y b, pues entonces x e y tendrán la forma deseada. Si el anillo $\mathbb{Z}\big[\sqrt{-3}\,\big]$ fuera un DFU esto se reduciría esencialmente a justificar que en la igualdad

$$z^3=(x+y\sqrt{-3})(x-y\sqrt{-3})$$

los dos factores de la derecha son primos entre sí, pues entonces sería fácil ver que ambos habrían de ser cubos. Lo que haremos será probar que los elementos $x+y\sqrt{-3}$ con $(x,y)=1$ admiten algo muy parecido a una factorización única.

1) Si $(a,b)=1$ y a^2+3b^2 es par, entonces $(1\pm\sqrt{-3}) \mid (a+b\sqrt{-3})$ (siempre en $\mathbb{Z}\big[\sqrt{-3}\,\big]$), donde el signo se escoge adecuadamente.

En general $p \mid q$ en $\mathbb{Z}\big[\sqrt{-3}\,\big]$ si y sólo si existe un $r\in\mathbb{Z}\big[\sqrt{-3}\,\big]$ tal que $pr=q$, es decir, si y sólo si $p^{-1}q\in\mathbb{Z}[\sqrt{-3}]$, donde el inverso se calcula, en principio, en $\mathbb{Q}\big[\sqrt{-3}\,\big]$.

Como $\mathrm{N}(u)=u\sigma(u)$, el inverso de u en $\mathbb{Q}\big[\sqrt{-3}\,\big]$ es $u^{-1}=\sigma(u)/\,\mathrm{N}(u)$.

Concretamente los inversos de $1+\sqrt{-3}$ y $1-\sqrt{-3}$ son, respectivamente, $\frac{1}{4}(1-\sqrt{-3})$ y $\frac{1}{4}(1+\sqrt{-3})$, y multiplicados por $a+b\sqrt{-3}$ dan

$$\frac{1}{4}(a+3b)+\frac{1}{4}(-a+b)\sqrt{-3} \quad \text{y} \quad \frac{1}{4}(a-3b)+\frac{1}{4}(a+b)\sqrt{-3}.$$

Así pues, $1\pm\sqrt{-3}$ dividirá a $a+b\sqrt{-3}$ si uno de estos dos elementos está en $\mathbb{Z}\big[\sqrt{-3}\,\big]$, es decir, si $4 \mid (a+3b)$ y $4 \mid (-a+b)$ o bien $4 \mid (a-3b)$ y $4 \mid (a+b)$.

Ahora bien, a y b han de tener la misma paridad, y como son primos entre sí, han de ser ambos impares. Por tanto serán de la forma $4n\pm1$, luego $a+b$ o $a-b$ ha de ser divisible entre 4.

Si $4 \mid a+b$, también $4 \mid a-3b$, pues, módulo 4, se cumple $[a-3b]=[a]+[-3][b]=[a]+[b]=[a+b]=0$.

Si $4 \mid a-b$, entonces también $4 \mid a+3b$, pues, módulo 4, $[a+3b]=[a]+[3][b]=[a]+[-1][b]=[a-b]=0$. Además $[-a+b]=[-1][a-b]=0$.

Por tanto se cumple lo pedido.

2) Si $(a,b)=1$ y a^2+3b^2 es divisible por el primo impar p, entonces p es de la forma $p=u^2+3v^2$ y $(u\pm v\sqrt{-3}) \mid (a+b\sqrt{-3})$, donde el signo hay que escogerlo adecuadamente.

La primera parte está probada en la sección anterior. También allí se vio que si un primo impar $p = u^2 + 3v^2$ divide a $a^2 + 3b^2$, entonces $p \mid (ub + sav)$, siendo $s = \pm 1$ y

$$a^2 + 3b^2 = \left(\frac{ua - s3vb}{p}\right)^2 + 3\left(\frac{ub + sav}{p}\right)^2.$$

Llamando $x = (ua - s3vb)/p$ e $y = (ub + sav)/p$, resulta que

$$x + y\sqrt{-3} = (u + sv\sqrt{-3})(a + b\sqrt{-3})/p.$$

Multiplicando por $u - sv\sqrt{-3}$ obtenemos

$$(u - sv\sqrt{-3})(x + y\sqrt{-3}) = p(a + b\sqrt{-3})/p = a + b\sqrt{-3}.$$

3) Si $(a, b) = 1$, entonces

$$a + b\sqrt{-3} = \pm(u_1 \pm v_1\sqrt{-3})(u_2 \pm v_2\sqrt{-3}) \cdots (u_n \pm v_n\sqrt{-3}), \tag{6.6}$$

donde la norma de cada factor es 4 o un primo impar.

Esto es consecuencia de los dos resultados anteriores, sin más que tener en cuenta que en una factorización $a + b\sqrt{-3} = (u + v\sqrt{-3})(x + y\sqrt{-3})$ se ha de cumplir $(x, y) = 1$ o de lo contrario $(a, b) \neq 1$.

4) Los factores de (6.6) están determinados salvo signo por el hecho de que $a^2+3b^2 = (u_1^2+3v_1^2)\cdots(u_n^2+3v_n^2)$ es una descomposición de a^2+3b^2 en producto de primos impares y cuatros. Además en la descomposición nunca aparecen a la vez dos factores $u + v\sqrt{-3}$ y $u - v\sqrt{-3}$.

Lo que hay que probar es que si $p = u^2 + 3v^2$ es un primo impar o un 4, entonces u y v están determinados salvo el signo. Para $p = 4$ es obvio. Si p es un primo impar y $p = u^2 + 3v^2 = r^2 + 3s^2$, entonces por 2) $u + v\sqrt{-3} = (r + s\sqrt{-3})(x + y\sqrt{-3})$, luego tomando normas, $x^2 + 3y^2 = 1$, con lo que $x + y\sqrt{-3} = \pm 1$ y así $u + v\sqrt{-3} = \pm(r + s\sqrt{-3})$, es decir, $u = \pm r$ y $v = \pm s$.

Por otra parte, si aparecieran simultáneamente dos factores $u + v\sqrt{-3}$ y $u - v\sqrt{-3}$, entonces $u^2 + 3v^2 \mid a + b\sqrt{-3}$, luego $u^2 + 3v^2 \mid (a, b)$, contradicción.

Ahora estamos en condiciones de demostrar el resultado que tenemos pendiente:

> *Si $(x, y) = 1$ y x^2+3y^2 es un cubo, entonces $x+y\sqrt{-3} = (a+b\sqrt{-3})^3$ para ciertos enteros a y b.*

En efecto: cada primo que divide a $x^2 + 3y^2$ lo hace con exponente múltiplo de 3, y como 2 ha de dividirlo con exponente par, de hecho 2 lo divide con exponente múltiplo de 6. Así $x^2 + 3y^2 = p_1^3 \cdots p_n^3$, donde cada p_i es un 4 o un primo impar.

En la factorización de $x + y\sqrt{-3}$ tipo (6.6), cada p_i da lugar al mismo factor $u_i \pm v_i\sqrt{-3}$, pues no pueden aparecer con signos distintos, es decir, la factorización es de la forma

$$x + y\sqrt{-3} = \pm(u_1 \pm v_1\sqrt{-3})^3(u_2 \pm v_2\sqrt{-3})^3 \cdots (u_n \pm v_n\sqrt{-3})^3,$$

luego $x + y\sqrt{-3}$ es un cubo. ∎

6.6 El Último Teorema de Fermat

Pierre de Fermat fue un eminente matemático que realizó grandes descubrimientos en teoría de números, pero, de acuerdo con las costumbres de su época, jamás publicó nada, y su trabajo es conocido por la correspondencia que mantuvo con otros matemáticos así como por la publicación póstuma de sus obras, llevada a cabo por su hijo Samuel. Entre las notas así publicadas se encontraban ciertas anotaciones en los márgenes de una edición de la Aritmética de Diofanto, una de las cuales, situada junto al punto en el que Diofanto encuentra las ternas pitagóricas, afirma que mientras —ciertamente— existen números enteros que satisfacen la ecuación $x^2 + y^2 = z^2$, el resultado es falso para exponentes mayores que dos, es decir, la ecuación $x^n + y^n = z^n$ no tiene soluciones positivas para $n > 2$. Fermat afirma tener una maravillosa demostración de este hecho, pero que no cabe en el margen del libro. Indudablemente Fermat debió de cometer un error, pues sólo recientemente se ha encontrado una compleja prueba basada en potentes técnicas algebraicas de las que Fermat distaba mucho de disponer.

La afirmación:

> *La ecuación $x^n + y^n = z^n$ no tiene soluciones enteras positivas para exponentes $n > 2$.*

se conoce con el nombre de *Último Teorema de Fermat,* aunque, según lo dicho, es prácticamente seguro que Fermat nunca lo demostró ni tampoco fue su última conjetura, ni mucho menos.

Aquí vamos a resolver algunos casos particulares del Último Teorema de Fermat y daremos algunas indicaciones del camino que permite obtener más resultados sobre él.El caso más simple, dado el estudio sobre las ternas pitagóricas que hemos llevado a cabo, es el caso $n = 4$.

Teorema 6.2 *La ecuación, $x^4 + y^4 = z^2$ no tiene soluciones enteras positivas. En particular el Último Teorema de Fermat es cierto para $n = 4$.*

Demostración: Si existen soluciones positivas de la ecuación $x^4 + y^4 = z^2$, entonces (x^2, y^2, z) es una terna pitagórica. Notar que si dividimos x, y, z por su m.c.d. obtenemos números primos entre sí que siguen cumpliendo la ecuación (notemos que si un primo p divide a x y a y, entonces p^2 divide a z), luego podemos suponer que $(x, y, z) = 1$, y claramente esto implica que en realidad son primos entre sí dos a dos y que la terna (x^2, y^2, z) es primitiva.

Según los resultados de la sección primera, $x^2 = 2pq$, $y^2 = p^2 - q^2$, $z = p^2 + q^2$, donde p y q son números enteros primos entre sí, de distinta paridad y $p > q > 0$ (intercambiamos x con y si es necesario para que x^2 sea el par).

Ahora, $p^2 = y^2 + q^2$, luego (q, y, p) es otra terna pitagórica, lo que obliga a que p sea impar, luego q ha de ser par, y así $q = 2ab$, $y = a^2 - b^2$, $p = a^2 + b^2$, para ciertos enteros a y b primos entre sí, de paridad opuesta, $a > b > 0$ (notar que se trata de una terna primitiva porque $(p, q) = 1$).

Por lo tanto $x^2 = 4ab(a^2 + b^2)$ y en consecuencia $ab(a^2 + b^2) = (x/2)^2$. Por otra parte $(a, b) = 1$ implica fácilmente que $(ab, a^2 + b^2) = 1$.

Ahora usamos una vez más un argumento muy simple pero importante: si el producto de dos números naturales primos entre sí es un cuadrado, entonces ambos son cuadrados, pues cada uno de ellos debe tener cada factor primo con exponente par.

Concluimos que ab y $a^2 + b^2$ son cuadrados y, por el mismo argumento, también lo son a y b. Digamos $a = u^2$, $b = v^2$, $a^2 + b^2 = w^2$.

Entonces $u^4 + v^4 = a^2 + b^2 = w^2 = p < p^2 + q^2 = z < z^2$.

En resumen, si existe una terna de números positivos (x, y, z) de manera que $x^4 + y^4 = z^2$, existe otra (u, v, w) que cumple lo mismo pero con $w^2 < z^2$. Si existieran tales ternas debería haber una con z mínimo, lo cual es falso según lo visto, por lo que la ecuación no tiene solución. ∎

Es importante notar que el teorema anterior no sólo prueba el Último Teorema de Fermat para $n = 4$, sino en general para $n = 4k$. En efecto, si existieran números positivos (x, y, z) tales que $x^{4k} + y^{4k} = z^{4k}$, entonces (x^k, y^k, z^k) sería una solución a la ecuación $x^4 + y^4 = z^4$, lo cual es imposible. En particular el Último Teorema de Fermat es cierto para las potencias de dos.

De aquí se sigue ahora que si el Último teorema de Fermat es cierto para exponentes primos impares, entonces es cierto para todo exponente. En efecto, si existen soluciones positivas a una ecuación $x^n + y^n = z^n$, entonces n no puede ser potencia de 2, luego existe un primo impar p tal que $p \mid n$, o sea, $n = pk$, para cierto entero k, luego (x^k, y^k, z^k) es una solución positiva a la ecuación $x^p + y^p = z^p$.

Observemos que si p es impar el Último Teorema de Fermat equivale a la no existencia de soluciones enteras no triviales (o sea, con $xyz \neq 0$) de la ecuación

$$x^p + y^p + z^p = 0,$$

lo que muestra que en realidad el papel de las tres variables es simétrico. Esto simplifica algunos argumentos.

El caso $p = 3$ fue demostrado por Euler, y en la prueba aparecen nuevas ideas de interés. Vamos a verlo.

Teorema 6.3 *No existen enteros no nulos x, y, z tales que $x^3 + y^3 = z^3$.*

Demostración: Vamos a seguir la prueba del teorema 6.2. Para empezar suponemos que existen números (x, y, z) que cumplen $x^3 + y^3 = z^3$. Dividiéndolos entre su m.c.d. podemos suponer que son primos entre sí y, al cumplir la ecuación, han de ser primos entre sí dos a dos. Es obvio que a lo sumo uno de los tres números puede ser par, pero si x, y son impares entonces z es par, luego exactamente uno de ellos es par.

Por simetría podemos suponer que x e y son impares. Entonces $x + y$, $x - y$ son pares, digamos $x + y = 2p$, $x - y = 2q$. Así $x = p + q$, $y = p - q$.

Ahora consideramos la factorización siguiente:

$$x^3 + y^3 = (x + y)(x^2 - xy + y^2)$$

[No es difícil llegar a ella: basta observar que el polinomio $x^3 + 1$ tiene una raíz igual a -1, luego es divisible entre $x + 1$, y la división da lugar a la factorización

$x^3+1=(x+1)(x^2-x+1)$. Ahora se sustituye x por x/y y se multiplica por y^3.]

Sustituyendo obtenemos

$$x^3+y^3=2p\big((p+q)^2-(p+q)(p-q)+(p-q)^2\big)=2p(p^2+3q^2).$$

Además podemos afirmar que p y q son primos entre sí (un factor común lo sería de x e y) y tienen paridades opuestas (porque $x=p+q$ es impar). Cambiando el signo de x, y, z si es necesario podemos suponer que $x+y>0$, luego $p>0$ e, intercambiando x con y si es necesario, también $q>0$ (no puede ser que $x=y$, pues q sería 0, y como $(x,y)=1$ habría de ser $x=y=1$, y entonces $z^3=2$, lo cual es imposible).

En resumen, si existe una solución (x,y,z) con x e y impares, entonces existen números naturales no nulos p y q de paridad opuesta, primos entre sí tales que el número $2p(p^2+3q^2)$ es un cubo.

El análogo en la prueba del teorema 6.2 era la factorización $x^2=4ab(a^2+b^2)$, que nos daba que $ab(a^2+b^2)$ debía ser un cuadrado. Igualmente nosotros hemos de justificar que los números $2p$ y p^2+3q^2 son primos entre sí, con lo que cada uno de ellos será un cubo.

En realidad esto no tiene por qué ser cierto, pero poco falta. Notemos primero que, como p y q tienen paridad opuesta, p^2+3q^2 es impar, de donde se sigue claramente que $(2p,p^2+3q^2)=(p,p^2+3q^2)=(p,3q^2)$ y como $(p,q)=1$ el único factor común de p y $3q^2$ es 3. En otras palabras, si 3 no divide a p, entonces $(2p,p^2+3q^2)=1$. Supongamos que es así.

Entonces, según lo dicho, $2p$ y p^2+3q^2 son cubos. En este punto usamos el resultado de la sección anterior (el análogo en el caso que nos ocupa a la clasificación de las ternas pitagóricas). Tenemos que $p=a(a-3b)(a+3b)$, $q=3b(a-b)(a+b)$. Claramente a y b son primos entre sí y tienen paridades opuestas (o si no p y q serían pares).

Por otra parte $2p=2a(a-3b)(a+3b)$ es un cubo. Veamos de nuevo que los factores $2a$, $a-3b$ y $a+3b$ son primos entre sí dos a dos, con lo que los tres serán cubos.

Como a y b tienen paridades opuestas, $a-3b$ y $a+3b$ son impares, luego un factor común de $2a$ y $a\pm3b$ es un factor de a y $a\pm3b$, luego un factor común de a y $3b$. Igualmente un factor común de $a+3b$ y $a-3b$ lo es de a y $3b$, luego basta probar que $(a,3b)=1$. Puesto que $(a,b)=1$, lo contrario obligaría a que $3\mid a$, pero entonces $3\mid p$ y estamos suponiendo lo contrario.

Así pues, $2a=u^3$, $a-3b=v^3$, $a+3b=w^3$, luego $v^3+w^3=2a=u^3$. Nuestro objetivo es encontrar una solución de la ecuación de Fermat con z^3 par y menor (en valor absoluto) que el valor del que hemos partido. Así podremos concluir que no pueden existir tales soluciones ya que no puede haber una mínima. Hemos de reordenar la terna (u,v,w) para dejar en tercer lugar la componente par. Como $u^3v^3w^3=2a(a-3b)(a+3b)=2p\mid z^3$, lo cierto es que la componente par, sea cual sea, es menor en módulo que z^3.

Falta llegar a la misma conclusión si $3\mid p$. Supongamos que $p=3s$ y que $3\nmid q$. Entonces nuestro cubo es $2p(p^2+3q^2)=3^2\cdot2s(3s^2+q^2)$ y los números

$3^2 \cdot 2s$ y $3s^2 + q^2$ son primos entre sí, pues $(s, q) = 1$ obliga a que los únicos divisores comunes posibles sean 2 y 3, pero $3s^2 + q^2$ es impar (luego 2 no sirve) y $3 \nmid q$, (luego tampoco sirve).

Consecuentemente $3^2 \cdot 2s = u^3$ y $3s^2 + q^2 = v^3$. Por el resultado de la sección anterior, $q = a(a - 3b)(a + 3b)$, $s = 3b(a - b)(a + b)$.

Por otro lado $3^2 \cdot 2s = 3^3 \cdot 2b(a - b)(a + b)$ es un cubo, luego $2b(a - b)(a + b)$ también lo es. El resto de la prueba es prácticamente igual al caso anterior. ∎

6.7 Enteros ciclotómicos

Hay dos aspectos de la prueba anterior que conviene destacar. Uno es el uso de la descomposición $x^3 + y^3 = (x + y)(x^2 - xy + y^2)$ en el comienzo de la demostración. El otro es el uso del anillo $\mathbb{Z}\left[\sqrt{-3}\right]$ como un medio de obtener resultados sobre números enteros pasando por números "imaginarios".

El caso $p = 5$ del último teorema de Fermat fue demostrado independientemente por Dirichlet y Legendre mediante técnicas similares, considerando factorizaciones más complejas y ayudándose del anillo $\mathbb{Z}\left[\sqrt{-5}\right]$. Sin embargo la prueba resulta mucho más complicada que la que acabamos de ver y los casos superiores se vuelven prácticamente intratables debido a que la complejidad aumenta desmesuradamente. Dirichlet intentó probar el caso $p = 7$, pero sólo consiguió una prueba para exponente 14.

Fue Kummer quien, basándose en ideas de Lamé, obtuvo una prueba del teorema de Fermat para una amplia clase de primos. No estamos en condiciones de abordar la teoría de Kummer, pero podemos indicar en qué se basa. Esencialmente se trata de usar números "imaginarios" para simplificar la factorización de $x^p + y^p$. Concretamente, tenemos la factorización

$$x^p - 1 = (x - 1)(x^{p-1} + \cdots + x + 1),$$

y vimos en el capítulo anterior que el factor $p(x) = x^{p-1} + \cdots + x + 1$ es irreducible. El teorema 5.21 nos da que existe un cuerpo $\mathbb{Q}[\omega]$ en el que $p(x)$ tiene una raíz ω. Concretamente

$$\mathbb{Q}[\omega] = \{a_{p-2}\omega^{p-2} + \cdots + a_1\omega + a_o \mid a_{p-2}, \ldots, a_0 \in \mathbb{Q}\},$$

y además la expresión de cada elemento es única. Por razones que ahora no podemos explicar a este cuerpo se le llama *cuerpo ciclotómico p–ésimo*.

Como $p(\omega) = 0$ y $p(x) \mid x^p - 1$, resulta que $\omega^p - 1 = 0$, o sea, $\omega^p = 1$ y, obviamente, $\omega \neq 1$.

No puede ocurrir que $\omega^n = 1$ para $0 < n < p$, pues entonces, tomando el mínimo n que cumple esto, $p = nc + r$ con $0 < r < p$ (p es primo), y entonces $\omega^p = (\omega^n)^c \cdot \omega^r = \omega^r \neq 1$, contradicción. Esto significa que las potencias $1 = \omega^0$, ω, ω^2, $\ldots$, ω^{p-1}, son distintas dos a dos, pues si $\omega^i = \omega^j$, entonces $\omega^{i-j} = 1$.

Además todas cumplen $(\omega^i)^p = 1$, luego son las p raíces del polinomio $x^p - 1$, o sea, $x^p - 1 = (x - 1)(x - \omega) \cdots (x - \omega^{p-1})$.

Sustituyendo $x = -x/y$ y multiplicando por $-y^p$ obtenemos la factorización

$$x^p + y^p = (x + y)(x + \omega y) \cdots (x + \omega^{p-1} y).$$

Esta factorización en polinomios de grado 1 es la más simple posible, y es la clave para obtener pruebas del teorema de Fermat para numerosos valores de p. Para ello es necesario estudiar el cuerpo $\mathbb{Q}[\omega]$ así como el anillo de los llamados *enteros ciclotómicos:*

$$\mathbb{Z}[\omega] = \{a_{p-2}\omega^{p-2} + \cdots + a_1\omega + a_0 \mid a_{p-2}, \ldots, a_0 \in \mathbb{Z}\}.$$

Ello supone desarrollar una compleja teoría que nos queda muy lejana. Sin embargo vamos a hacer alguna observación adicional sobre estos anillos.

Como la relación $\omega^p = 1$ es más sencilla de manejar que la relación

$$\omega^{p-1} + \cdots + \omega + 1 = 0, \tag{6.7}$$

resulta conveniente trabajar con los elementos de $\mathbb{Q}[\omega]$ en la forma

$$a_{p-1}\omega^{p-1} + \cdots + a_1\omega + a_0 \tag{6.8}$$

(de modo que, al operar, simplemente reducimos las potencias ω^p que puedan aparecer).

El único inconveniente es que la expresión ya no es única. Si tenemos

$$a_{p-1}\omega^{p-1} + \cdots + a_1\omega + a_0 = 0,$$

usando la relación (6.7) tenemos que

$$a_{p-1}\omega^{p-1} + \cdots + a_1\omega + a_0 - a_{p-1}(\omega^{p-1} + \cdots + \omega + 1) = 0,$$

o sea,

$$(a_{p-2} - a_{p-1})\omega^{p-2} + \cdots + (a_1 - a_{p-1})\omega + (a_0 - a_{p-1}) = 0,$$

y por la unicidad

$$a_{p-1} = a_{p-2} = \cdots = a_1 = a_0.$$

De aquí se sigue en general que

$$a_{p-1}\omega^{p-1} + \cdots + a_1\omega + a_0 = b_{p-1}\omega^{p-1} + \cdots + b_1\omega + b_0$$

si y sólo si (restando los dos miembros)

$$a_{p-1} - b_{p-1} = a_{p-2} - b_{p-2} = \cdots = a_1 - b_1 = a_0 - b_0,$$

o en otras palabras, si existe un número racional c tal que $a_i = b_i + c$, para todo $i = 0, \ldots, p-1$. Equivalentemente, los coeficientes de una expresión (6.8) están unívocamente determinados salvo suma de un número racional (o de un número entero si el elemento está en $\mathbb{Z}[\omega]$).

En los próximos capítulos iremos aplicando a estos anillos los resultados que vayamos obteniendo.

Para terminar, notemos que para $p = 3$ el elemento ω es raíz del polinomio $x^2 + x + 1$, y por lo tanto

$$\omega = \frac{-1 + \sqrt{-3}}{2}.$$

De aquí se sigue fácilmente que $\mathbb{Q}[\omega] = \mathbb{Q}[\sqrt{-3}]$, aunque $\mathbb{Z}[\sqrt{-3}] \subsetneq \mathbb{Z}[\omega]$. Esta inclusión es en el fondo la razón por la que se cumplen los resultados que hemos visto en la sección 6.3, pues sucede que el anillo

$$\mathbb{Z}[\omega] = \mathbb{Z}\left[\frac{1 + \sqrt{-3}}{2}\right]$$

sí tiene factorización única (es un dominio euclídeo), y lo que hemos visto en la sección 6.3 es un reflejo de dicha factorización en el subanillo $\mathbb{Z}[\sqrt{-3}]$. No estamos en condiciones de probar teóricamente la relación entre la factorización de uno y otro anillo, pero lo dicho basta para comprender que los resultados que hemos probado mediante cálculos prolijos pueden ser vistos como consecuencias claras de fenómenos abstractos conceptualmente mucho más simples y mucho más comprensibles.

Ejercicio: Justificar que la igualdad $2 \cdot 2 = (1 + \sqrt{-3})(1 - \sqrt{-3})$ no contradice la factorización única en $\mathbb{Z}[\omega]$.

Capítulo VII

Módulos y espacios vectoriales

En este capítulo introduciremos una nueva estructura algebraica más sencilla que la de anillo. La idea es que la estructura de los anillos es en general muy complicada, pero, prescindiendo en parte de la operación de producto, obtenemos una estructura más simple que puede ser analizada con facilidad y nos proporciona información importante.

7.1 Módulos

Definición 7.1 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 $r \in M$.

Observamos que la suma en un módulo ha de cumplir las mismas propiedades que la suma en un anillo, por lo que las propiedades elementales de la suma de anillos 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 $(M, +, \cdot)$.

Si D es un anillo de división, los D–módulos se llaman *espacios vectoriales.*

Tenemos disponibles muchos ejemplos de módulos:

En primer lugar, si A es un anillo unitario, entonces A es un A–módulo con su suma y su producto.

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 $A \times B$.

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).

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).

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 $\phi(m) = m1$).

Enunciemos a continuación las propiedades elementales de los módulos. Todas se demuestran igual que para anillos. Observar que el producto de números enteros por elementos de un módulo está definido exactamente igual que para anillos.

Teorema 7.2 *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 7.3 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.

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. No siempre es posible considerar a un subconjunto como submódulo (por ejemplo si no contiene al 0), las condiciones que se han de cumplir las da el teorema siguiente.

Teorema 7.4 *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. ∎

Observar que las condiciones 2) y 3) del teorema anterior pueden resumirse en una sola:

Teorema 7.5 *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* $ar + bs \in N$.

El teorema 7.4 muestra claramente que cuando consideramos a un anillo unitario A como A–módulo, entonces los submódulos coinciden con los ideales izquierdos.

Definición 7.6 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.*

También 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 $\langle X\rangle \subset \langle Y\rangle$. Notar que $\langle\varnothing\rangle = 0$.

Cuando el conjunto X sea finito, $X = \{x_1, \ldots, x_n\}$, 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.

Teniendo en cuenta que al considerar a un anillo conmutativo y unitario A como A–módulo los submódulos coinciden con los ideales, es inmediato que el submódulo generado por un subconjunto X coincide con el ideal generado por X, es decir, $(X) = \langle X\rangle$.

Es fácil reconocer los elementos del submódulo generado por un subconjunto:

Teorema 7.7 *Sea A un anillo unitario, M un A–módulo y $X \subset M$. Entonces*

$$\langle X\rangle = \left\{ \sum_{i=1}^{n} a_i r_i \;\middle|\; n \in \mathbb{N}, a_i \in A, r_i \in X \right\}.$$

La prueba es sencilla: 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 $\langle X\rangle$.

Veamos algunos ejemplos concretos. El cuerpo $\mathbb{Q}\big[\sqrt{-3}\,\big]$ considerado en el capítulo anterior es un $\mathbb{Q}$–espacio vectorial. Como cuerpo no tiene ideales propios, pero como $\mathbb{Q}$–espacio vectorial tiene infinitos subespacios. Por ejemplo $\langle 1\rangle = \{a1 \mid a \in \mathbb{Q}\} = \mathbb{Q}$, o $\big\langle\sqrt{-3}\,\big\rangle = \{a\sqrt{-3} \mid a \in \mathbb{Q}\}$, $\big\langle 5 + 2\sqrt{-3}\,\big\rangle = \{a5 + 2a\sqrt{-3} \mid a \in \mathbb{Q}\}$, etc.

El anillo $\mathbb{Z}\big[\sqrt{-3}\,\big]$ es un $\mathbb{Z}$–módulo. Cambiando $\mathbb{Q}$ por $\mathbb{Z}$ en los ejemplos anteriores tenemos ejemplos de submódulos de $\mathbb{Z}\big[\sqrt{-3}\,\big]$.

Los módulos cociente se definen exactamente igual que los anillos cociente, aunque para nosotros tendrán un interés secundario.

Definición 7.8 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

$$[r] = r + N = \{r + s \mid s \in N\}.$$

Teorema 7.9 *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.

Definimos los homomorfismos de módulos de forma análoga a los de anillos. Su interpretación es la misma.

Definición 7.10 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.
Una *aplicación lineal* es un homomorfismo de espacios vectoriales.

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, llamaremos *núcleo* de f al submódulo de M dado por $\mathrm{N}(f) = \{r \in M \mid f(r) = 0\}$, la *imagen* de f es el submódulo de N dado por $\operatorname{Im} f = f[M] = \{f(r) \mid r \in M\}$.

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 $\mathrm{N}(f) = N$.

Teorema 7.11 *Un homomorfismo de módulos es inyectivo si y sólo si su núcleo es trivial.* (cf. 5.16).

Teorema 7.12 *(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.* (cf. 5.17).

7.2 Suma de módulos

El interés que tendrán los módulos para nosotros es que, mientras la estructura de un anillo puede ser muy complicada, la estructura de módulo de ese mismo anillo puede ser fácil de describir. Un primer análisis de la estructura de un módulo consiste en descomponerlo en suma de módulos más simples en el sentido en que a continuación indicaremos.

Definición 7.13 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

$$\sum_{i\in I} N_i = \left\langle \bigcup_{i\in I} N_i \right\rangle$$

Es decir, la suma de una familia de submódulos es el menor submódulo que los contiene a todos. Por el teorema 7.7, un elemento de este submódulo es una suma de elementos de algunos de los módulos N_i multiplicados por elementos de 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, $N_1 + \cdots + N_n = \{r_1 + \cdots + r_n \mid r_i \in N_i\}$.

Descomponer un módulo en suma de submódulos nos da una información importante sobre su estructura. Por ejemplo, todo elemento de $\mathbb{Q}\big[\sqrt{-3}\,\big]$ es de la forma $a + b\sqrt{-3}$, con a, $b \in \mathbb{Q}$, luego $\mathbb{Q}\big[\sqrt{-3}\,\big] = \big\langle 1\big\rangle + \big\langle \sqrt{-3}\,\big\rangle$. Igualmente $\mathbb{Z}\big[\sqrt{-3}\,\big] = \big\langle 1\big\rangle + \big\langle \sqrt{-3}\,\big\rangle$.

En realidad se cumple más: no sólo cada elemento de $\mathbb{Q}\big[\sqrt{-3}\,\big]$ es suma de un elemento de $\big\langle 1\big\rangle$ y otro de $\big\langle \sqrt{-3}\,\big\rangle$, sino que además la descomposición es única. La razón es que los subespacios considerados tienen intersección trivial. Veamos esto con detalle:

Definición 7.14 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 $\bigoplus_{i\in I} N_i$.

Por ejemplo, tenemos que $\mathbb{Q}\big[\sqrt{-3}\,\big] = \big\langle 1\big\rangle \oplus \big\langle \sqrt{-3}\,\big\rangle$.

Ejercicio: Probar que $\mathbb{Q}[x] = M \oplus N$, con $M = \big\langle x^{2n} \mid n \in \mathbb{N}\big\rangle$, $N = \big\langle x^{2n+1} \mid n \in \mathbb{N}\big\rangle$.

Como decíamos, las sumas directas están relacionadas con la unicidad de las descomposiciones en sumas:

Teorema 7.15 *Sea A un anillo unitario, M un A–módulo y N_1, ... , 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 se da la igualdad* $m_1 + \cdots + m_n = 0$ *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) $\Rightarrow$ 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) $\Rightarrow$ 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. ∎

En definitiva, si un módulo M se expresa como suma directa de una familia de n submódulos, entonces cada elemento de M determina y está determinado por un elemento de cada uno de los submódulos. Para reflejar adecuadamente este hecho conviene introducir el concepto de producto de módulos.

Definición 7.16 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

$$(m_1, \ldots, m_n) + (m'_1, \ldots, m'_n) = (m_1 + m'_1, \ldots, m_n + m'_n),$$

$$r \cdot (m_1, \ldots, m_n) = (rm_1, \ldots, rm_n).$$

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 tienen nulas todas sus componentes 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 distintos de $M_1 \times \cdots \times M_n$ (isomorfos, pero distintos).

También es inmediato que si $(m_1, \ldots, m_n) \in M_1 \times \cdots \times M_n$ entonces

$$(m_1, \ldots, m_n) = \iota_1(m_1) + \cdots + \iota_n(m_n),$$

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 7.15.

Estos resultados son ciertos en el caso de tener infinitos módulos, pero con una matización:

Si tenemos una familia de A–módulos $\{M_i\}_{i\in I}$, entonces 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 $\iota_i : M_i \longrightarrow \prod_{i\in I} M_i$ 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

$$\sum_{i\in I} M_i = \Big\{ f \in \prod_{i\in I} M_i \;\Big|\; \{i \in I \mid f(i) \neq 0\} \text{ es finito} \Big\}.$$

Pero se cumple igualmente que la suma $\sum_{i\in I} M_i$ es directa.

Al módulo así construido, o sea, a

$$\Big\{ f \in \prod_{i\in I} M_i \;\Big|\; \{i \in I \mid f(i) \neq 0\} \text{ es finito} \Big\}$$

lo llamaremos *suma directa externa* de los módulos $\{M_i\}_{i\in I}$, y la representaremos

$$\bigoplus_{i\in I} M_i.$$

De este modo, dada una familia arbitraria de A–módulos, podemos construir un módulo que se exprese como suma directa de submódulos isomorfos a los módulos dados.

Finalmente notemos que si $M = M_1 \oplus \cdots \oplus M_n$, entonces las aplicaciones $\pi_i : M \longrightarrow M_i$ dadas por $\pi_i(m_1 + \cdots + m_n) = m_i$ son epimorfismos de módulos.

7.3 Módulos libres.

Notar que, como en la descomposición $\mathbb{Q}\big[\sqrt{3}\,\big] = \langle 1\rangle \oplus \langle \sqrt{3}\,\rangle$ los sumandos son monógenos, en lugar de decir que todo elemento de $\mathbb{Q}\big[\sqrt{3}\,\big]$ se expresa de forma única como un elemento de $\langle 1\rangle$ más un elemento de $\langle \sqrt{3}\,\rangle$, podemos precisar y decir que se expresa de forma única como $a \cdot 1 + b \cdot \sqrt{3}$, con a y b en el anillo $\mathbb{Q}$, es decir, que cada elemento de $\mathbb{Q}\big[\sqrt{3}\,\big]$ determina y está determinado por dos números racionales. Vamos a estudiar esta situación en general.

Definición 7.17 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 $x_1, \ldots, x_n \in X$ 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 $a_1 = \cdots = a_n = 0$.

En caso contrario se dice que X es un conjunto *ligado* o que sus elementos son *linealmente dependientes.*

Es claro que un conjunto que contenga a 0 es ligado, pues $1 \cdot 0 = 0$.

Los elementos de M de la forma $a_1x_1 + \cdots + a_nx_n$ se llaman *combinaciones lineales* de los elementos $x_1, \ldots, x_n$, luego un conjunto es linealmente independiente si el 0 se expresa de forma única como combinación lineal de sus elementos.

Notemos que si $X = \{x_1, \ldots, x_n\}$ es un conjunto libre, entonces

$$\langle X\rangle = \langle x_1\rangle \oplus \cdots \oplus \langle x_n\rangle\,,$$

pues por el teorema 7.7 todo elemento de $\langle X\rangle$ es de la forma $a_1x_1 + \cdots + a_nx_n$, es decir, $\langle X\rangle = \langle x_1\rangle + \cdots + \langle x_n\rangle$, y por la definición de conjunto libre se cumple la condición 2) del teorema 7.15, luego la suma es directa.

En realidad la prueba vale igual si el conjunto X es infinito, en cuyo caso $\langle X\rangle = \bigoplus_{x\in X} \langle x\rangle$.

De nuevo por el teorema 7.15 resulta que si X es libre, cada elemento de $\langle X\rangle$ se expresa de forma única como combinación lineal de los elementos de X, es decir, si un conjunto $X = \{x_1, \ldots, x_n\}$ es libre, el módulo $\langle X\rangle$ tiene tantos elementos como n–tuplas posibles de elementos de A. Para cada una de estas n–tuplas $(a_1, \ldots, a_n) \in A^n$, el elemento $a_1x_1 + \cdots + a_nx_n$ es un elemento distinto de $\langle X\rangle$. Más aún, 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 que a cada elemento de $\langle X\rangle$ le asigna su n–tupla de coeficientes $(a_1, \ldots, a_n)$ es un isomorfismo de módulos, o sea, $\langle X\rangle \cong A^n$.

El recíproco es cierto, pero antes de probarlo conviene introducir un nuevo concepto:

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.

Ya hemos probada la mitad del teorema siguiente:

Teorema 7.18 *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 producto A^n.*

DEMOSTRACIÓN: Sólo falta probar que el módulo A^n tiene una base con n elementos, pero es fácil ver que, por ejemplo, para $n = 3$, una base de A^3 está formada por las ternas $(1, 0, 0)$, $(0, 1, 0)$ y $(0, 0, 1)$. ∎

Es fácil ver que el resultado vale igualmente para bases infinitas: Un A–módulo M tiene una base si y sólo si es isomorfo a una suma directa con todos los sumandos iguales al anillo A. En tal caso M tendrá una base con tantos elementos como sumandos.

Por ejemplo, una base de $\mathbb{Q}[\sqrt{3}]$ está formada por los elementos 1 y $\sqrt{3}$. Una base de $\mathbb{Q}[x]$ está formada por todas las potencias de x: 1, x, x^2, x^3, x^4, ...

Definición 7.19 Sea A un anillo unitario y M un A–módulo libre con una base $X = \{x_1, \ldots, x_n\}$. Entonces cada elemento m de M se expresa de forma única como combinación lineal $m = a_1x_1 + \cdots + a_nx_n$. Los elementos $(a_1, \ldots, a_n)$ se llaman *coordenadas* de m en la base X (esto presupone una ordenación en la base). Hemos visto que la aplicación que a cada elemento le asigna su n–tupla de coordenadas es un isomorfismo de módulos.

El interés de las bases es que nos dan una representación clara de los elementos de un módulo. En definitiva, los elementos de un módulo libre son identificables con n-tuplas de elementos del anillo.

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 él es finito.

Hay un caso importante en el que podemos garantizar la existencia de bases:

Teorema 7.20 *Sea D un anillo de división y V un D–espacio vectorial. Entonces V es libre, más aún, todo subconjunto libre de V está contenido en una base de V.*

DEMOSTRACIÓN: La familia de los subconjuntos libres de un módulo (incluyendo al conjunto vacío) está inductivamente ordenada respecto a la inclusión, luego por el lema de Zorn todo subconjunto libre de un módulo está contenido en uno maximal para la inclusión.

Si X es un subconjunto libre de un D–espacio vectorial V, sea B un subconjunto libre maximal para la inclusión que contenga a X. Basta ver que B es un generador de V.

En otro caso existiría un elemento v en V que no podría expresarse como combinación lineal de elementos de B. Basta probar que $B \cup \{v\}$ es un subconjunto libre de V, pues esto contradice la maximalidad de B.

En efecto, si $0 = dv + d_1v_1 + \cdots + d_nv_n$ para ciertos elementos $d, d_1, \ldots, d_n$ en el anillo D, no todos nulos, y ciertos elementos $v_1, \ldots, v_n$ de B, entonces no puede ser $d = 0$, o de lo contrario $0 = d_1v_1 + \cdots + d_nv_n$ implicaría que B no es libre luego, al ser D un anillo de división, existe d^{-1} y podemos despejar $v = -d^{-1}d_1v_1 - \cdots - d^{-1}d_nv_n$, con lo que v sí es combinación lineal de los elementos de B, contradicción en cualquier caso. ∎

Notar que todo subconjunto de todo conjunto libre es libre, y que todo conjunto que contenga a un conjunto generador es también generador. Hemos probado que en un espacio vectorial todo conjunto libre está contenido en una base. Ahora vamos a ver que todo generador contiene a una base.

Teorema 7.21 *Sea D un anillo de división y V un D–espacio vectorial. Entonces todo generador de V contiene una base.*

Demostración: Sea X un generador de V y sea B un conjunto libre contenido en X y maximal respecto a la inclusión. Se cumple que $X \subset \langle B\rangle$, pues si existiera un elemento $x \in X \setminus \langle B\rangle$, entonces se comprueba igual que en 7.20 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. ∎

A continuación vamos a probar una propiedad notable de los módulos libres sobre anillos conmutativos y unitarios, y es que en ellos todas las bases tienen el mismo número de elementos. La prueba no es sencilla. En primer lugar nos ocupamos del caso infinito.

Teorema 7.22 *Sea A un anillo unitario y M un A–módulo libre con una base infinita. Entonces todas las bases de M tienen el mismo cardinal.*

Demostración: Supongamos que X es una base infinita de M y sea Y otra base. Cada elemento x de X (necesariamente no nulo) se expresa como combinación lineal de elementos de Y, luego existe un subconjunto finito F_x de Y de modo que x se expresa como combinación lineal de los elementos de F_x con todos los coeficientes no nulos. Al ser Y una base, F_x es único.

Consideremos la aplicación $f : X \longrightarrow \mathcal{P}Y$ dada por $f(x) = F_x$. Por definición, si $x \in X$ se cumple que $x \in \langle f(x)\rangle$. En consecuencia, si F es un subconjunto finito de Y, se cumple que $f^{-1}\big[\{F\}\big] \subset \langle F\rangle$.

Cada elemento de F es combinación lineal de un número finito de elementos de X, luego podemos encontrar un subconjunto finito G de X tal que $F \subset \langle G\rangle$, luego tenemos que $f^{-1}\big[\{F\}\big] \subset \langle F\rangle \subset \langle G\rangle$.

Más exactamente, $f^{-1}\big[\{F\}\big] \subset G$, pues si existiera un $x \in f^{-1}\big[\{F\}\big] \setminus G$, entonces $x \in \langle G\rangle$, es decir, x sería combinación lineal de otros elementos de X (los de G), lo que nos daría que X es linealmente dependiente.

Esto significa que $f^{-1}\big[\{F\}\big]$ es finito, para todo subconjunto finito de Y.

Tenemos una aplicación f de X en el conjunto de los subconjuntos finitos de Y tal que cada subconjunto finito de Y tiene a lo sumo una cantidad finita de antiimágenes.

Claramente $X = \bigcup_F f^{-1}\big[\{F\}\big]$, donde F recorre los subconjuntos finitos de Y, pero como los conjuntos $f^{-1}\big[\{F\}\big]$ son finitos, esto implica que Y ha de ser infinito y además

$$|X| \leq \sum_F \Big|f^{-1}\big[\{F\}\big]\Big| \leq \sum_F \aleph_0 = |Y|.$$

Intercambiando los papeles de X e Y obtenemos que $|X| = |Y|$. ∎

Nos queda considerar el caso de los módulos cuyas bases son finitas. Primero lo probamos para espacios vectoriales.

Teorema 7.23 *Sea V un espacio vectorial sobre un anillo de división D. Entonces todas las bases de V tienen el mismo cardinal.*

Demostración: Sean X e Y bases de V. Por el teorema anterior podemos suponer que son finitas. Digamos que $X = \{x_1, \ldots, x_n\}$, $Y = \{y_1, \ldots, y_m\}$. 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 $x_1 \in \langle y_1, x_2, \ldots x_n\rangle$.

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 $y_1, \ldots, y_n$, o sea, $y_{n+1} = a_1y_1 + \cdots + a_ny_n$, luego $a_1y_1 + \cdots + a_ny_n - y_{n+1} = 0$, con lo que Y sería linealmente dependiente. ∎

Ejercicio: Deducir del teorema anterior que si M es un módulo sobre un dominio íntegro entones todas las bases de M tienen el mismo cardinal.

A continuación probamos el caso más general de equicardinalidad de bases que vamos a considerar.

Teorema 7.24 *Si A es un anillo conmutativo y unitario y M es un A–módulo libre, entonces todas las bases de M tienen el mismo cardinal.*

DEMOSTRACIÓN: Sea I un ideal maximal de A. 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, y todas sus bases tienen el mismo cardinal. Basta probar, pues, que toda A–base de M tiene el mismo cardinal que una A/I–base de M/IM.

Sea X una base de M. Por simplificar la notación supondremos que es finita, digamos $X = \{x_1, \ldots, x_n\}$. Veamos que $X^* = \{[x_1], \ldots, [x_n]\}$ es una base de M/IM.

Si $[u] \in M/IM$, entonces $u \in M$, luego $u = \sum_{j=1}^{n} a_j x_j$ para ciertos $a_j \in A$. Por lo tanto $[u] = \sum_{j=1}^{n} [a_j][x_j]$, lo que prueba que X^* genera M/IM.

Veamos que $[x_1], \ldots, [x_n]$ son linealmente independientes en M/IM (y en particular que son distintos).

Supongamos que $\sum_{j=1}^{n} [a_j][x_j] = [0]$ para ciertos elementos a_j de A (que podemos suponer no nulos). Entonces $\sum_{j=1}^{n} a_j x_j \in IM$, 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 $b_k \in I$.

Así pues $[a_1] = \cdots = [a_n] = [0]$, como queríamos probar. Con esto tenemos que X^* es base de M/IM, y en particular hemos visto que si $x \neq x'$, entonces $[x] \neq [x']$, luego se cumple también que $|X^*| = |X|$. ∎

Definición 7.25 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. Si A es un anillo de división (y por lo tanto M es un espacio vectorial), al rango de M se le llama *dimensión* de M ($\dim M$).

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 7.26 *Sea V un espacio vectorial sobre un anillo de división D y sea W un subespacio de V. Entonces:*

1. $\dim V = \dim W + \dim(V/W)$. *En particular* $\dim W \leq \dim V$.

2. *Si* $\dim W = \dim V$ *y ambas son finitas, entonces* $W = V$.

Demostración: 1) Sea X una base de W. Por el teorema 7.20 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 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 7.27 *Sean V y W dos subespacios 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*

$$\dim(V + W) + \dim(V \cap W) = \dim V + \dim W.$$

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 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 7.26 se cumple lo siguiente:

Teorema 7.28 *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, o sea, es 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. Tan sólo podemos salvar lo que afirma el teorema siguiente:

Teorema 7.29 *Todo submódulo de un módulo libre sobre un dominio de ideales principales es libre de rango menor o igual.*

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. 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.

Consideremos 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 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

$$a_1 m_{b_1} + \cdots + a_{n-1} m_{b_{n-1}} \in M \cap L_{b_n}.$$

Por lo tanto,

$$0 = f_{b_n}(0) = f_{b_n}(a_1 m_{b_1} + \cdots + a_{n-1} m_{b_{n-1}}) + f_{b_n}(a_n b_n) = a_n d_n,$$

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. ∎

En el caso de que los módulos tengan rango finito podemos hacer una precisión que nos será útil en varias ocasiones:

Teorema 7.30 *Sea A un dominio euclídeo, M un A–módulo libre de rango m y N un submódulo no nulo de M. Entonces M tiene una base $b_1, \ldots, b_m$ tal que existen $a_1, \ldots, a_n \in A$ (con $n \leq m$) tales que $a_1 b_1, \cdots, a_n b_n$ es una base de N.*

Demostración: Para cada base $y_1, \ldots, y_m$ de M y cada base $z_1, \ldots, z_n$ de N podemos considerar las coordenadas de cada z_i en la base $y_1, \ldots, y_m$, que obviamente no pueden ser todas nulas. Escojemos dos bases tales que exista un z_i que tenga una coordenada no nula $a_1 \in A$ cuya norma euclídea sea la mínima posible (es decir, tal que cualquier coordenada no nula de cualquier elemento de cualquier base de N respecto de cualquier base de M tenga norma euclídea mayor o igual que la de a_1).

Reordenando las bases podemos suponer que $z_1 = a_1 y_1 + b_2 y_2 + \cdots + b_m y_m$. Dividamos $b_i = a_1 c_i + r_i$, donde cada r_i es nulo o bien tiene norma euclídea menor que la de a_1. Así

$$z_1 = a_1(y_1 + c_2 y_2 + \cdots + c_m y_m) + r_2 y_2 + \cdots + r_m y_m,$$

Llamamos $x_1 = y_1 + c_2 y_2 + \cdots + c_m y_m$, y es fácil ver que $x_1, y_2, \ldots, y_m$ es también una base de M, respecto de la cual, las coordenadas de z_1 son $(a_1, r_2, \ldots, r_m)$. Por la minimalidad de a_1 concluimos que $r_2 = \cdots = r_m = 0$. Por consiguiente, $z_1 = a_1 x_1$. Llamemos $M' = \langle y_2, \ldots, y_m \rangle$. Así, podemos expresar $z_i = b_i x_1 + z_i'$, con $b_i \in A$ y $z_i' \in M'$. Dividamos $b_i = a_1 c_i + r_i$, donde r_i es nulo o tiene norma menor que a_i. Para cada $i > 1$ llamamos $w_i = z_i - c_i z_1 = r_i x_1 + z_i'$. Es claro que $z_1, w_2, \ldots, w_n$ es una base de N, respecto de la cual las coordenadas de los w_i en x_1 son los r_i. Por la elección de a_1 concluimos que $r_i = 0$. Equivalentemente, tenemos que $N' = \langle w_2, \ldots, w_n \rangle \subset M'$.

Ahora razonamos por inducción sobre m (el rango de M). Si $m = 1$ entonces $N' = M' = 0$ y hemos encontrado las bases x_1 de M y $a_1 x_1$ de N, tal y como exige el teorema. Si el teorema es cierto para módulos de rango menor que M, entonces podemos aplicar la hipótesis de inducción a M' y N', con lo que encontramos una base $x_2, \ldots, x_m$ de M' tal que $a_2 x_2, \ldots, a_n x_n$ es una base de N', para ciertos $a_i \in A$. Es claro entonces que $x_1, \ldots, x_m$ es una base de M y que $a_1 x_1, \ldots, a_n x_n$ es una base de N. ∎

Los demás resultados sobre espacios vectoriales son falsos para módulos hasta 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 7.26). 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. $\mathbb{Z}/2\mathbb{Z}$).

La existencia de bases en un módulo es importante 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 7.31 *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. ∎

Concluimos con un resultado que generaliza en parte al teorema 7.30.

Teorema 7.32 *Sea A un dominio euclídeo y M un A–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 A–módulo libre de rango n y sea $\{y_1, \ldots, y_n\}$ una base de L. Entonces por el teorema anterior 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 $N = f\left[f^{-1}[N]\right]$, se comprueba fácilmente que $f^{-1}[N]$ es un submódulo de L, luego por 7.30 es 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. ∎

Capítulo VIII

Extensiones de cuerpos

Con la teoría de módulos y espacios vectoriales como herramienta, estamos en condiciones de profundizar notablemente en el estudio de los anillos que nos han ido apareciendo en capítulos anteriores, tales como los cuerpos cuadráticos $\mathbb{Q}\big[\sqrt{-3}\,\big]$ o los cuerpos ciclotómicos $\mathbb{Q}[\omega]$. De momento nos ocuparemos sólo de los cuerpos y dejaremos para más adelante el estudio de los anillos como $\mathbb{Z}\big[\sqrt{-3}\,\big]$ o $\mathbb{Z}[\omega]$, pues en este capítulo usaremos fuertemente las propiedades de los espacios vectoriales.

8.1 Extensiones algebraicas

Definición 8.1 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. Principalmente nos interesarán las extensiones que tienen por cuerpo base al cuerpo de los números racionales, pero conviene estudiar las extensiones en general.

La primera observación importante es que si K/k es una extensión, 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.

Por ejemplo, $\big|\mathbb{Q}\big[\sqrt{-3}\,\big] : \mathbb{Q}\big| = 2$, pues una base de $\mathbb{Q}\big[\sqrt{-3}\,\big]$ como $\mathbb{Q}$–espacio vectorial está formada por los números 1 y $\sqrt{-3}$. Los resultados vistos en el capítulo VI sobre los cuerpos ciclotómicos $\mathbb{Q}[\omega]$ para un primo p, nos dan que $\big|\mathbb{Q}[\omega] : \mathbb{Q}\big| = p-1$, pues una base de $\mathbb{Q}[\omega]$ como $\mathbb{Q}$–espacio vectorial es la formada por $1, \omega, \ldots, \omega^{p-2}$.

Nos van a interesar especialmente las extensiones finitas, aunque tendremos ocasión de trabajar con algunas infinitas. El resultado siguiente es fundamental en todo lo que sigue:

Teorema 8.2 *(Teorema de transitividad de grados) Consideremos tres cuerpos $k \subset K \subset L$. 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 $z = \sum_{j=1}^{n} b_j y_j$. Ahora 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 introducimos el concepto más importante de este capítulo.

Definición 8.3 Sea K/k una extensión y $a \in K$. Se dice que el elemento 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.*

Por ejemplo, todo elemento a del cuerpo base es algebraico, pues es raíz del polinomio $x - a \in k[x]$. El elemento $\sqrt{-3}$ es raíz de $x^2 + 3 \in \mathbb{Q}[x]$, luego es algebraico sobre $\mathbb{Q}$. Por el contrario, en la extensión $\mathbb{Q}(x)/\mathbb{Q}$, el elemento x es obviamente trascendente.

Puede parecer fuerte la exigencia de que todos los elementos de una extensión sean algebraicos, pero en realidad se cumplirá en todos los casos que vamos a manejar. De ello es responsable en gran parte el teorema siguiente:

Teorema 8.4 *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 $b_0 + b_1 x + \cdots + b_n x^n \in k[x]$. ∎

Por ejemplo, consideremos el elemento $1 + \sqrt{-3} \in \mathbb{Q}\big[\sqrt{-3}\,\big]$. Como la extensión $\mathbb{Q}\big[\sqrt{-3}\,\big]/\mathbb{Q}$ tiene grado 2, siguiendo la prueba del teorema anterior es suficiente considerar las potencias 1, $1+\sqrt{-3}$, $(1+\sqrt{-3}\,)^2 = -2 + 2\sqrt{-3}$. Hay que buscar números racionales a, b, c que cumplan

$$a + b(1+\sqrt{-3}) + c(-2 + 2\sqrt{-3}\,) = a + b - 2c + (b+2c)\sqrt{-3} = 0,$$

lo que equivale a que $a + b - 2c = 0 = b + 2c$. Por ejemplo sirven $c = 1$, $b = -2$ y $a = 4$. Así pues, $1 + \sqrt{-3}$ es raíz del polinomio $4 - 2x + x^2 \in \mathbb{Q}[x]$.

Introducimos ahora en general una notación que venimos empleando desde hace tiempo en muchos casos particulares. Se trata de una forma muy cómoda de describir extensiones de un cuerpo a partir de elementos generadores.

Definición 8.5 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

$$A[S] = \{p(a_1, \ldots, a_n) \mid n \in \mathbb{N}, p(x_1, \ldots, x_n) \in A[x_1, \ldots, x_n], a_1, \ldots, a_n \in S\},$$

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 $p(a_1, \ldots, a_n)$.

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

$$k(S) = \left\{ \frac{p(b_1, \ldots b_n)}{q(b_1, \ldots b_n)} \mid n \in \mathbb{N}, p, q \in k[x_1, \ldots, x_n], b_i \in S, q(b_1, \ldots, b_n) \neq 0 \right\}.$$

El cuerpo $k(S)$ se llama *adjunción* a k de S.
Cuando el conjunto $S = \{b_1, \ldots, b_n\}$ es finito escribiremos también

$$A[b_1, \ldots, b_n] \quad \text{y} \quad k(b_1, \ldots, b_n).$$

Notar que $A[S \cup T] = A[S][T]$ y $k(S \cup T) = k(S)(T)$. En particular

$$A[b_1, \ldots, b_n] = A[b_1] \ldots [b_n] \quad \text{y} \quad k(b_1, \ldots, b_n) = k(b_1) \ldots (b_n).$$

Observar 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. También son casos particulares los nombres que hemos dado a los anillos $\mathbb{Z}\big[\sqrt{-3}\,\big]$, etc.

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*. El elemento a (que no es único) se llama *elemento primitivo* de la extensión.

Es fácil ver que toda extensión finita es finitamente generada, pues si S es una k–base de K es claro que $K = k[S] = k(S)$.

Ejercicio: Probar que $\mathbb{Q}(\sqrt{2}, \sqrt{3}\,) = \mathbb{Q}(\sqrt{2} + \sqrt{3}\,)$

Las extensiones algebraicas tienen un comportamiento y una estructura muy simples. La práctica totalidad de sus propiedades es consecuencia del teorema siguiente. Es de destacar que en la prueba hacemos uso de los principales resultados que conocemos sobre anillos de polinomios y divisibilidad.

Teorema 8.6 *Sea K/k una extensión 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, pero en el capítulo IV vimos que en un DIP todo ideal primo 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 8.7 Sea K/k una extensión 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.

El teorema siguiente precisa un poco más los resultados que hemos obtenido.

Teorema 8.8 *Sea K/k una extensión 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 8.6 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. ∎

Así pues, al adjuntar a un cuerpo un elemento algebraico obtenemos una extensión finita, y en particular algebraica. Esto sigue siendo cierto si adjuntamos un número finito de elementos algebraicos. Si adjuntamos infinitos podemos perder la finitud, pero nunca el carácter algebraico de la extensión resultante. Veámoslo.

Teorema 8.9 *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 8.6 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. ∎

Así pues 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 8.10 *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, es decir, 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 $\operatorname{pol}\operatorname{mín}(a, K)$. Así, $\operatorname{pol}\operatorname{mín}(a, K) \in k(b_1, \ldots, b_n)[x]$, 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. Lo dejamos como ejercicio.

Ejercicio: Probar que si K/k es una extensión de cuerpos, entonces el conjunto de los elementos de K que son algebraicos sobre k es un subcuerpo de K.

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.

8.2 Homomorfismos entre extensiones

En esta sección probaremos que si tenemos una extensión de cuerpos K/k y adjuntamos a k un conjunto $S \subset K$ de elementos algebraicos, la extensión $k(S)$ está completamente determinada por los polinomios mínimos de los elementos de S, y que en particular no depende de K. Para expresar esto con precisión necesitamos el concepto de homomorfismo de extensiones. En realidad, 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 8.11 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. (En general, un automorfismo de anillos, módulos, etc. es un isomorfismo de un anillo, módulo, etc. en sí mismo).

Notar también 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.

Si K/k es una extensión, llamaremos $G(K/k)$ al conjunto de todos los k–automorfismos de K.

Si tenemos un cuerpo k y un polinomio irreducible $p(x) \in k[x]$, el teorema 5.21 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 8.6 hemos obtenido que toda extensión de la forma $k(a)$, donde a es raíz de $p(x)$, es isomorfa a la construida en 5.21. Como consecuencia dos cualesquiera de estas extensiones son isomorfas entre sí. Lo probamos en un contexto más general.

Teorema 8.12 *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, tal y como ya comentábamos, 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 8.13 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 8.14 *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.

Pensemos ahora en una extensión algebraica simple $k(a)/k$. Si σ y τ son dos k–automorfismos de $k(a)$ tales que $\sigma(a) = \tau(a)$, entonces $\sigma = \tau$, pues todo elemento de $k(a)$ es de la forma $p(a)$ para cierto polinomio $p(x) \in k[x]$, y $\sigma\big(p(a)\big) = p\big(\sigma(a)\big) = p\big(\tau(a)\big) = \tau\big(p(a)\big)$. En otras palabras, un k–automorfismo σ está determinado por el valor $\sigma(a)$ que toma en un elemento primitivo a. Ahora bien, sabemos 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 es una raíz en $k(a)$ del polinomio mínimo de a, 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 $\mathbb{Q}(i)/\mathbb{Q}$ el elemento primitivo i tiene dos conjugados, él mismo y $-i$. En consecuencia $\mathbb{Q}(i)$ tiene dos automorfismos, la identidad y el determinado por $\sigma(i) = -i$, es decir, $\sigma(a + bi) = a - bi$.

Recordemos que tanto la prueba de que un anillo como $\mathbb{Z}[i]$ es un dominio euclídeo, como la prueba de que $\mathbb{Z}\big[\sqrt{-3}\,\big]$ no es un DFU dependen fuertemente de las propiedades de las normas respectivas. 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

$$\mathrm{N}(a) = \sigma_1(a) \cdots \sigma_n(a).$$

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.

De momento podemos comprobar, al menos, que este esquema general es válido en los ejemplos que hemos manejado. Por ejemplo, la norma en $\mathbb{Q}(i)$ viene dada por $\mathrm{N}(a + bi) = a^2 + b^2 = (a + bi)(a - bi) = (a + bi)\sigma(a + bi)$, luego es ciertamente el producto de los dos automorfismos de la extensión $\mathbb{Q}(i)/\mathbb{Q}$. El lector puede igualmente contrastar el caso de $\mathbb{Q}\big(\sqrt{-3}\,\big)$, etc.

Ejercicio: Comprobar que la extensión ciclotómica $\mathbb{Q}(\omega)/\mathbb{Q}$, donde $\omega^p = 1$, tiene exactamente $p - 1$ automorfismos. Calcular explícitamente la norma de un elemento $a_0 + a_1\omega + \cdots + a_{p-2}\omega^{p-2}$ (definida como el producto de sus imágenes por todos los

automorfismos) en los casos $p = 3$ y $p = 5$. Comprobar que en ambos casos la norma es un número racional y que la norma de un elemento del anillo $\mathbb{Z}[\omega]$ es un entero.

Sin embargo 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 los razonamientos particulares que aquí vamos a emplear.

Consideramos el polinomio x^3-2, que por el criterio de Eisenstein es irreducible en $\mathbb{Q}[x]$. Podemos aplicar el teorema 5.21 para construir un cuerpo $\mathbb{Q}\big(\sqrt[3]{2}\big)$ donde tiene una raíz. En este cuerpo podemos factorizar

$$x^3 - 2 = (x - \sqrt[3]{2}\,)(x^2 + \sqrt[3]{2}x + \sqrt[3]{2}^{\,2}).$$

Vamos a ver que el segundo factor es irreducible en $\mathbb{Q}\big(\sqrt[3]{2}\big)[x]$, o lo que es lo mismo, que no tiene raíces en $\mathbb{Q}\big(\sqrt[3]{2}\big)$. Equivalentemente, hemos de ver que $\sqrt[3]{2}^{\,2} - 4\sqrt[3]{2}^{\,2} = -3\sqrt[3]{2}^{\,2}$ no tiene raíz cuadrada en $\mathbb{Q}\big(\sqrt[3]{2}\big)$, es decir, que no existe ningún elemento $\eta \in \mathbb{Q}\big(\sqrt[3]{2}\big)$ tal que $\eta^2 = -3\sqrt[3]{2}^{\,2}$.

Como la extensión $\mathbb{Q}\big(\sqrt[3]{2}\big)/\mathbb{Q}$ tiene grado 3, todo elemento de $\mathbb{Q}\big(\sqrt[3]{2}\big)$ es de la forma $\eta = a + b\sqrt[3]{2} + c\sqrt[3]{2}^{\,2}$ para ciertos números racionales a, b, c. Entonces

$$\begin{aligned}\eta^2 &= a^2 + b^2\sqrt[3]{2}^{\,2} + 2c^2\sqrt[3]{2} + 2ab\sqrt[3]{2} + 2ac\sqrt[3]{2}^{\,2} + 4bc\\ &= (a^2 + 4bc) + (2c^2 + 2ab)\sqrt[3]{2} + (b^2 + 2ac)\sqrt[3]{2}^{\,2}.\end{aligned}$$

Por la unicidad de la expresión, $\eta^2 = -3\sqrt[3]{2}^{\,2}$ equivale a que

$$\begin{aligned}a^2 + 4bc &= 0,\\ 2c^2 + 2ab &= 0,\\ b^2 + 2ac &= -3,\end{aligned}$$

para ciertos números racionales a, b, c.

Notar que $b \neq 0$, pues en otro caso se deduce $a = b = c = 0$, y no se cumplen las ecuaciones. Igualmente $c \neq 0$. De la segunda ecuación se obtiene $a = -\frac{c^2}{b}$, y sustituyendo en la primera obtenemos que $\frac{c^4}{b^2} + 4bc = 0$, de donde $\frac{c^3}{b^3} = -4$.

Tenemos, pues, que -4 tiene una raíz cúbica en $\mathbb{Q}$, lo cual es falso, pues entonces la tendría en $\mathbb{Z}$ (ver el ejercicio tras el teorema 4.24).

Así queda probado que $x^2 + \sqrt[3]{2}x + \sqrt[3]{2}^{\,2}$ es irreducible en $\mathbb{Q}\big(\sqrt[3]{2}\big)[x]$. El elemento primitivo no tiene $\mathbb{Q}$–conjugados en $\mathbb{Q}\big(\sqrt[3]{2}\big)$, luego el único $\mathbb{Q}$–automorfismo de $\mathbb{Q}\big(\sqrt[3]{2}\big)$ es la identidad y no podemos definir una norma en este cuerpo como producto de automorfismos.

Sin embargo podemos volver a aplicar el teorema 5.21 y considerar una extensión de $\mathbb{Q}\big(\sqrt[3]{2}\big)$ en la que el polinomio $x^2 + \sqrt[3]{2}x + \sqrt[3]{2}^{\,2}$ tenga una raíz α. Consideremos $\mathbb{Q}\big(\sqrt[3]{2}\big)(\alpha) = \mathbb{Q}\big(\sqrt[3]{2}, \alpha\big)$.

El polinomio $x^3 - 2$ tiene dos raíces en $\mathbb{Q}\big(\sqrt[3]{2}, \alpha\big)$, luego tiene tres. Por razones de simetría conviene abandonar la notación $\sqrt[3]{2}$ y llamar α, 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 (ver la prueba de la transitividad de grados) por los elementos

$$1,\quad \alpha,\quad \alpha^2,\quad \beta,\quad \alpha\beta,\quad \alpha^2\beta. \tag{8.1}$$

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 (8.1) como combinación lineal de los elementos de dicha base. Notar 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. No estamos en condiciones de apreciar el valor que tiene la norma de una extensión, pero lo cierto es que representa un papel fundamental no sólo en la práctica, sino también en los resultados teóricos más profundos.

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.

8.3 Clausuras algebraicas

Acabamos de ver que para definir una norma en una extensión puede hacernos falta considerar una extensión mayor, donde cierto polinomio tenga más raíces. Esto se consigue usando una o más veces el teorema 5.21. Aquí vamos a estudiar las extensiones que se obtienen de ese modo, para entender exactamente cuándo hace falta pasar a un cuerpo mayor, qué extensiones hemos de buscar y cuáles son las ventajas que presentan frente a las extensiones de partida. Todo resulta conceptualmente más simple si probamos primero que todo cuerpo tiene una máxima extensión algebraica, donde todos los polinomios tienen todas sus raíces, de manera que todos los cuerpos que nos interesarán podrán ser considerados como subcuerpos de dicha extensión máxima.

La idea de que un polinomio tenga 'todas' sus raíces en un cuerpo hay que entenderla como sigue:

Definición 8.15 Sea K un cuerpo y $p(x) \in K[x]$. Diremos que el polinomio $p(x)$ se *escinde* en $K[x]$ si existen elementos $a_0, a_1, \ldots, a_n \in K$ (no necesariamente distintos) tales que $p(x) = a_0(x - a_1)\cdots(x - a_n)$.

Notar que a_0 es el coeficiente director de $p(x)$.

Si tenemos un cuerpo k y un polinomio $p(x) \in k[x]$, aplicando el teorema 5.21 a un factor irreducible de $p(x)$ podemos encontrar una extensión donde $p(x)$ factorice como $p(x) = (x - a_1)p_1(x)$. Aplicándolo de nuevo a un factor irreducible de $p_1(x)$ encontramos una extensión mayor donde $p(x) = (x-a_1)(x-a_2)p_2(x)$, y tras un número finito de pasos llegamos a una extensión de k donde $p(x)$ se escinde.

En estos términos, lo que queremos probar es que todo cuerpo tiene una extensión algebraica en la que todos los polinomios se escinden, y por lo tanto nunca necesitaremos buscar una mayor para que un polinomio tenga más raíces. La prueba consiste en construir una cadena de extensiones de modo que en cada una de ellas se escinda un polinomio más, hasta recorrer todos los polinomios de $k[x]$. Como hemos de tratar con infinitos polinomios, necesitamos un poco de teoría de conjuntos.

Teorema 8.16 *Sea k un cuerpo. Existe una extensión algebraica K/k tal que todo polinomio no constante de $k[x]$ se escinde en $K[x]$.*

Demostración: Sea P el conjunto de todos los polinomios no constantes de $k[x]$. Consideramos un buen orden $\preceq$ en P, es decir, un orden total en el que todo subconjunto no vacío tiene mínimo. Esto equivale a poner en una lista a todos los polinomios:

$$p_0 \prec p_1 \prec \ \ldots \ \prec p_\omega \prec p_{\omega+1} \prec p_{\omega+2} \prec \ \ldots \ \prec p_{\omega+\omega} \prec p_{\omega+\omega+1} \prec \ \ldots$$

Vamos a construir una cadena $\{K_p\}_{p\in P}$ de extensiones algebraicas de k tales que si $q \preceq p$ entonces K_p es una extensión de K_q y para cada $p \in P$ el polinomio

$p(x)$ se escinde en K_p. Podemos hacerlo por recursión, es decir, basta definir K_p supuestos definidos $\{K_q\}_{q\prec p}$. Para ello consideramos

$$K_p^* = k \cup \bigcup_{q\prec p} K_q,$$

que claramente es un cuerpo y, de hecho, una extensión algebraica de k.

Definimos K_p como una extensión algebraica de K_p^* tal que $p(x)$ se escinda en $K_p^*[x]$ (existe por la observación previa al teorema). Esto termina la definición. Ahora tomamos

$$K = \bigcup_{p\in P} K_p.$$

De nuevo es claro que K es un cuerpo, es una extensión algebraica de k y cada polinomio no constante $p(x) \in k[x]$ se escinde en $K_p[x]$, luego en $K[x]$. ∎

Veamos algunas caracterizaciones del cuerpo que hemos construido.

Teorema 8.17 *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 $f(x) = (x - a_1)f_1(x) = f(x) = (x - a_1)(x - a_2)f_2(x)$, 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) ⇒ 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, luego podemos considerar el polinomio $p(x) = \text{pol mín}(a, k)$, 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 8.18 Diremos que un cuerpo K es *algebraicamente cerrado* si cumple cualquiera de las condiciones del teorema anterior.

Un cuerpo K es una *clausura algebraica* de otro cuerpo k si K/k es una extensión algebraica y K es algebraicamente cerrado.

El teorema 8.16 afirma que todo cuerpo tiene una clausura algebraica. Ahora probaremos que dos cualesquiera son isomorfas, con lo que podemos decir que todo cuerpo tiene esencialmente una única clausura algebraica. Lo probamos en un contexto más general que nos será útil después.

Teorema 8.19 *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 $p(x) = \text{pol mín}(u, A)$ 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 8.12 nos da un isomorfismo $\tau' : A(u) \longrightarrow L$ 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 8.20 *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'. ∎

Definición 8.21 Llamaremos $\mathbb{A}$ a una clausura algebraica de $\mathbb{Q}$. No importa cuál sea, pues dos cualesquiera son $\mathbb{Q}$–isomorfas, luego son el mismo cuerpo a todos los efectos. Cuando hablemos de 'números algebraicos', sin más precisión, se ha de entender que nos referimos a los elementos de $\mathbb{A}$, los números algebraicos sobre $\mathbb{Q}$.

Así tenemos una extensión algebraica $\mathbb{A}/\mathbb{Q}$, de modo que ya no necesitamos construir extensiones cada vez que queramos que un cierto polinomio tenga raíces, sino que todos los polinomios de $\mathbb{Q}[x]$ tienen raíces en $\mathbb{A}$.

Esto nos permite un cambio conceptual importante: Si tenemos un cuerpo k, hasta ahora no tenía sentido hablar de las raíces de un polinomio $p(x) \in k[x]$ mientras no justificáramos su existencia en k o no construyéramos una extensión de k donde existieran. Ahora, en cambio, podemos hablar de las raíces $p(x)$ con independencia de que estén o no en k. Si no están todas en k sabemos de todos modos que las que faltan están en su clausura algebraica (en una clausura algebraica prefijada), y podemos adjuntarlas a k si nos interesa.

En particular, los objetos $\sqrt{-3}, \sqrt[3]{2}, i, \omega$, etc. que hemos venido fabricando según los necesitábamos ya no serán para nosotros construcciones específicas aisladas, sino elementos del cuerpo $\mathbb{A}$.

Con un ligero abuso de lenguaje, si $p(x)$ es un polinomio de grado n con coeficientes en un cuerpo k y K es una clausura algebraica de k, podemos decir que $p(x)$ tiene n raíces en K. Esto no es exacto, porque algunas de estas raíces pueden ser iguales. Con rigor, si K es un cuerpo algebraicamente cerrado, cada polinomio $p(x) \in K[x]$ de grado n se escinde en la forma

$$p(x) = a_0(x - a_1) \cdots (x - a_n),$$

y si agrupamos los factores iguales podemos escribir también

$$p(x) = a_0(x - a_1)^{r_1} \cdots (x - a_m)^{r_m},$$

donde ahora $a_1, \ldots, a_m$ son distintos dos a dos y $r_1 + \cdots + r_m = n$.

Esta descomposición es única, pues no es sino la descomposición en irreducibles de $p(x)$ en $K[x]$, que es un DFU. Llamaremos *orden de multiplicidad* de la raíz a_i en el polinomio $p(x)$ al exponente r_i.

De este modo, la suma de los órdenes de multiplicidad de las raíces de un polinomio en un cuerpo algebraicamente cerrado es igual al grado del polinomio. Para completar la definición, si un elemento $a \in K$ no es raíz de un polinomio $p(x) \in K[x]$, diremos que su orden de multiplicidad en $p(x)$ es 0.

Una raíz de un polinomio es *simple* si su orden de multiplicidad es 1. En otro caso es *múltiple.* A su vez una raíz múltiple puede ser *doble, triple,* etc.

Como por el criterio de Eisenstein, el polinomio $x^n - 2$ es irreducible en $\mathbb{Q}[x]$ para todo número natural n no nulo, al tomar una raíz $u \in \mathbb{A}$, obtenemos una extensión $\mathbb{Q}(u) \subset \mathbb{A}$ de modo que $|\mathbb{Q}(u) : \mathbb{Q}| = n$. Por la transitividad de grados, la extensión $\mathbb{A}/\mathbb{Q}$ es infinita.

Ejercicio: Sea K/k una extensión no necesariamente algebraica donde K es algebraicamente cerrado. Probar que el conjunto de los elementos de K algebraicos sobre k es una clausura algebraica de k.

Volvamos ahora al problema de definir normas. El problema con que nos encontrábamos en al intentar definir la norma del cuerpo $\mathbb{Q}\big(\sqrt[3]{2}\,\big)$ (como extensión de $\mathbb{Q}$) era que este cuerpo no tiene automorfismos. La razón es que los conjugados del elemento primitivo $\sqrt[3]{2}$ están fuera del cuerpo. Por ello tuvimos que extenderlo y considerar los tres monomorfismos $\sigma : \mathbb{Q}\big(\sqrt[3]{2}\,\big) \longrightarrow \mathbb{Q}(\alpha, \beta)$.

Desde un punto de vista teórico el segundo cuerpo es irrelevante, en el sentido de que dichos monomorfismos son en realidad todos los monomorfismos $\sigma : \mathbb{Q}\big(\sqrt[3]{2}\,\big) \longrightarrow \mathbb{A}$. En efecto, cualquiera de estos monomorfismos ha de enviar $\sqrt[3]{2}$ a uno de sus tres conjugados α, β, γ, luego ha de coincidir con uno de los anteriores.

Ahora ya podemos definir la norma de una extensión finita K/k. Se calcula aplicando todos los k–monomorfismos de K en una clausura algebraica y multiplicándolos. La norma así definida es obviamente multiplicativa, pero de momento seguimos sin poder probar que las normas de los elementos de K están en k, lo cual es fundamental.

8.4 Extensiones normales

El concepto de clausura algebraica tiene interés teórico, porque nos permite hablar de las raíces de un polinomio como objetos existentes de antemano sin tener que construirlas en cada caso particular. Sin embargo suele suceder que la clausura algebraica de un cuerpo sea una extensión infinita, mientras que la mayoría de los resultados sobre extensiones de cuerpos valen sólo para extensiones finitas. Por otro lado nunca nos va a interesar trabajar con infinitos polinomios a un tiempo. Por ello conviene estudiar la mínima extensión donde un polinomio dado tiene todas sus raíces. Veremos que estas extensiones tienen un comportamiento muy parecido al de las clausuras algebraicas, pero conservan la finitud.

Definición 8.22 Sea k un cuerpo y $p(x) \in k[x]$. Se dice que un cuerpo 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 $p(x)$.

Si fijamos una clausura algebraica K de un cuerpo k y nos restringimos a considerar las extensiones algebraicas de k contenidas en K, entonces cada polinomio $p(x) \in k[x]$ tiene un único cuerpo de escisión, a saber, la adjunción a k de sus raíces en K. Si no nos restringimos a una clausura algebraica fija, entonces podemos probar que el cuerpo de escisión es único salvo k–isomorfismo:

Teorema 8.23 *Dos cuerpos de escisión cualesquiera de un mismo polinomio sobre un mismo cuerpo k son k–isomorfos.*

DEMOSTRACIÓN: Sean K y K' cuerpos de escisión de un polinomio $p(x)$ sobre un cuerpo k. Consideremos dos clausuras algebraicas L y L' de K y K' respectivamente. Por el teorema 8.20 la identidad en k se extiende a un k–isomorfismo $\sigma : L \longrightarrow L'$. Basta probar que $\sigma[K] = K'$, pero $\sigma[K]$ es un cuerpo de escisión de $p(x)$ sobre k, luego ha de ser la adjunción a k de las raíces de $p(x)$ en L', o sea, $\sigma[K] = K'$. ∎

Por ejemplo, $\mathbb{Q}(\alpha, \beta)$ es el cuerpo de escisión sobre $\mathbb{Q}$, o sobre $\mathbb{Q}(\alpha)$, del polinomio $x^3 - 2$. Así mismo, $\mathbb{Q}\big(\sqrt{2}\big)$ es el cuerpo de escisión sobre $\mathbb{Q}$ de $x^2 - 2$ y el cuerpo ciclotómico p–ésimo $\mathbb{Q}(\omega)$ es el cuerpo de escisión, sobre $\mathbb{Q}$, del polinomio $x^{p-1} + \cdots + x + 1$, o también de $x^p - 1$.

Sucede que las extensiones K/k donde K es el cuerpo de escisión sobre k de un cierto polinomio poseen muchas propiedades generales que no dependen del polinomio en cuestión. Por ello conviene introducir el concepto siguiente:

Definición 8.24 Diremos que una extensión K/k es *finita normal*[1] si K es el cuerpo de escisión sobre k de un polinomio $p(x) \in k[x]$.

Ejercicio: Probar que toda extensión de grado 2 es normal.

Los teoremas siguientes muestran que las extensiones finitas normales tienen un comportamiento similar al de las clausuras algebraicas. Todas las propiedades de estas extensiones se siguen esencialmente del próximo teorema técnico. Pese a su sencillez, contiene una de las ideas básicas en las que descansa la teoría que estamos desarrollando.

Teorema 8.25 *Sea $k \subset F \subset K \subset L$ una cadena de extensiones algebraicas, donde K/k es finita normal. Sea $\sigma : F \longrightarrow L$ un k–monomorfismo. Entonces se cumple que $\sigma[F] \subset K$ y σ se extiende a un k–automorfismo de K.*

DEMOSTRACIÓN: Sea L' una clausura algebraica de L. Obviamente también lo es de F y de $\sigma[F]$. Por el teorema 8.20 tenemos que σ se extiende a un k–automorfismo $\sigma^* : L' \longrightarrow L'$. Basta probar que $\sigma^*[K] = K$, pues entonces a *fortiori* $\sigma[F] \subset K$ y la restricción de σ^* a K será el k–automorfismo buscado.

Sea $p(x) \in k[x]$ tal que K sea el cuerpo de escisión sobre k de $p(x)$. Sean $a_1, \ldots, a_n$ las raíces de $p(x)$ en K. Entonces tenemos que $K = k(a_1, \ldots, a_n)$, luego $\sigma^*[K] = k(\sigma^*(a_1), \ldots \sigma^*(a_n))$. Ahora bien, cada $\sigma^*(a_i)$ es una raíz de $p(x)$ (no olvidemos que los k–monomorfismos envían raíces a raíces). Por lo tanto la restricción de σ^* al conjunto $\{a_1, \ldots, a_n\}$ es una inyección de dicho conjunto en sí mismo, y como es finito es de hecho una biyección. En otros términos, $\{\sigma^*(a_1), \ldots, \sigma^*(a_n)\} = \{a_1, \ldots, a_n\}$. Por lo tanto $\sigma^*[K] = k(a_1, \ldots, a_n) = K$. ∎

Así pues, un k–monomorfismo que parta de un subcuerpo de una extensión normal a un cuerpo mayor, en realidad no se sale de la extensión normal, y

[1] Puede definirse una extensión normal no necesariamente finita como el cuerpo de escisión de un conjunto de polinomios, no necesariamente finito, pero aquí sólo vamos a considerar extensiones normales finitas.

además se extiende a un k–automorfismo de la extensión normal. Informalmente, es imposible 'salirse' de una extensión normal K/k mediante un k–monomorfismo. Quien asimile esta idea verá naturales los resultados siguientes.

En primer lugar, siempre podemos pasar de un elemento a un k–conjugado mediante un k–monomorfismo. Luego una extensión normal K/k contiene a los k–conjugados de todos sus elementos. Con más detalle tenemos el teorema siguiente.

Teorema 8.26 *Sea K/k una extensión finita normal y $p(x) \in k[x]$ un polinomio irreducible con una raíz en K. Entonces $p(x)$ se escinde en $K[x]$.*

DEMOSTRACIÓN: Sea $a \in K$ una raíz de $p(x)$. Sea L una clausura algebraica de K. Sea b otra raíz de $p(x)$ en L. Entonces a y b son k–conjugados (su polinomio mínimo es $p(x)$ dividido entre su coeficiente director). Por el teorema 8.14 existe un k–isomorfismo $\sigma : k(a) \longrightarrow k(b)$. Por el teorema anterior (tomando $F = k(a)$) resulta que $k(b) \subset K$, es decir, K contiene todas las raíces de $p(x)$ en L. Como $p(x)$ se escinde en $L[x]$, de hecho se escinde en $K[x]$. ∎

Así pues, mientras en una clausura algebraica se escinden todos los polinomios, en una extensión normal se escinden al menos todos los irreducibles que tienen una mínima relación con K (una raíz). Por ejemplo, la extensión $\mathbb{Q}\big(\sqrt[3]{2}\big)/\mathbb{Q}$ no es normal, pues $x^3 - 2$ tiene una raíz en $\mathbb{Q}\big(\sqrt[3]{2}\big)$, pero no se escinde.

Es fácil probar el recíproco del teorema anterior:

Teorema 8.27 *Una extensión finita K/k es normal si y sólo si todo polinomio irreducible $p(x) \in k[x]$ con una raíz en K se escinde en $K[x]$.*

DEMOSTRACIÓN: La extensión K/k es finita, luego finitamente generada: $K = k(a_1, \ldots, a_n)$ y, por hipótesis, los polinomios $p_i(x) = \operatorname{pol}\operatorname{mín}(a_i, k)$ se escinden en $K[x]$, luego $f(x) = p_1(x) \cdots p_n(x)$ también se escinde. Entonces K es el cuerpo de escisión sobre k de $f(x)$. ∎

Observar 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). En general la extensión K/k no tiene por qué ser normal, como lo muestra $\mathbb{Q} \subset \mathbb{Q}\big(\sqrt[3]{2}\big) \subset \mathbb{Q}(\alpha, \beta)$.

El hecho de que los monomorfismos que parten de una extensión normal se extiendan a automorfismos hace que las extensiones normales tengan muchos automorfismos, y que éstos sean suficientes para controlar la extensión. El núcleo de la teoría de extensiones algebraicas consiste en obtener información de una extensión normal K/k a partir del conjunto $G(K/k)$ de todos los k–automorfismos de K. Un ejemplo en dicha línea es la siguiente mejora del teorema 8.14:

Teorema 8.28 *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 8.14 existe un k–isomorfismo $\sigma : k(a) \longrightarrow k(b)$, que por 8.25 se extiende a un k–automorfismo de K. El recíproco es obvio. ∎

Ejercicio: Dada una extensión K/k, se dice que K es el cuerpo de escisión sobre k de un conjunto $S \subset k[x]$ si todos los polinomios de S se escinden en $K[x]$ y K es la adjunción a k de sus raíces. Probar que K/k es finita normal si y sólo si K es el cuerpo de escisión de un conjunto finito de polinomios.

Ejercicio: Una extensión K/k (no necesariamente finita) es *normal* si K es el cuerpo de escisión de un conjunto de polinomios de $k[x]$. Generalizar todos los teoremas anteriores al caso de extensiones normales cualesquiera.

Para definir una norma en la extensión $\mathbb{Q}\big(\sqrt[3]{2}\,\big)/\mathbb{Q}$ hemos tenido que pasar a la extensión $\mathbb{Q}(\alpha,\beta)/\mathbb{Q}$. La razón es que la primera no es normal. La segunda es un sustituto finito de la clausura algebraica de $\mathbb{Q}$. Vamos a generalizar esta idea.

Partamos de una extensión finita K/k. En particular será finitamente generada, es decir, $K = k(u_1,\dots,u_n)$. Fijemos una clausura algebraica de K. Sea $p(x)$ el producto de todos los polinomios $\operatorname{pol\,mín}(u_i,k)$ y sea L el cuerpo de escisión de $p(x)$ sobre k (en la clausura algebraica fijada).

Entonces la extensión L/k es normal, y como entre las raíces de $p(x)$ se encuentran $u_1,\dots,u_n$, concluimos que $k \subset K \subset L$.

Más aún, si L' es cualquier otra extensión normal de k que contenga a K (siempre en la clausura algebraica fijada), entonces los polinomios $\operatorname{pol\,mín}(u_i,k)$ tienen todos una raíz en L' (precisamente u_i), luego se escinden en $L'[x]$, luego $p(x)$ se escinde en $L'[x]$, con lo que $L \subset L'$.

Es decir, L es la menor extensión normal de k que contiene a K. En particular esto prueba que la definición de L no depende de la elección de los generadores $u_1,\dots,u_n$. En resumen hemos probado:

Teorema 8.29 *Sea K/k una extensión finita. Cada clausura algebraica de K contiene una mínima extensión finita normal L/k tal que $k \subset K \subset L$. Ésta se obtiene como el cuerpo de escisión sobre k del producto de los polinomios mínimos en k de cualquier generador finito de K/k. Se la llama* clausura normal *de K/k.*

Si no fijamos una clausura algebraica de K, es fácil ver que dos clausuras normales de una extensión K/k son K–isomorfas.

Así pues, $\mathbb{Q}(\alpha,\beta)$ es la clausura normal de la extensión $\mathbb{Q}\big(\sqrt[3]{2}\,\big)/\mathbb{Q}$.

Observar que si L es la clausura normal de una extensión K/k en una clausura algebraica C, los k–monomorfismos $\sigma : K \longrightarrow C$ coinciden con los k–monomorfismos $\sigma : K \longrightarrow L$ (por el teorema 8.25).

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 algebraica, y ahora hemos visto que esto equivale a considerar los k–monomorfismos de K en la clausura normal de K/k, que es una extensión finita normal L/k.

Observar también que si K/k es normal, entonces K es la clausura normal de la extensión, luego los k–monomorfismos de K en la clausura normal son precisamente los k–automorfismos de K.

8.5 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.

Recordemos que si k es un cuerpo y $f(x) \in k[x]$, el orden de multiplicidad de una raíz $a \in k$ es el número de veces que el factor $x - a$ aparece en la descomposición de $f(x)$ en irreducibles en $K[x]$, siendo K la clausura algebraica de k, o simplemente un cuerpo de escisión de $f(x)$.

Si el orden de multiplicidad de una raíz a es n, entonces $f(x) = (x-a)^n g(x)$, donde $g(x) \in K[x]$ y $g(a) \neq 0$. Una raíz de multiplicidad 1 es una raíz simple.

Aunque no es importante, debemos notar que si $f(x) \in k[x]$ y a es una raíz de $f(x)$ en k, el orden de multiplicidad de a en $f(x)$ es el único número natural para el que se da la descomposición $f(x) = (x - a)^n g(x)$, donde $g(x) \in k[x]$ y $g(a) \neq 0$, aunque $f(x)$ no se escinda en $k[x]$. La razón es que pasando a la clausura algebraica K de k, el polinomio $g(x)$ se descompone en factores de grado 1 en $K[x]$ entre los que no puede figurar $x - a$, ya que $g(a) \neq 0$, luego la descomposición de $f(x)$ en $K[x]$ tiene n factores $x - a$ más los demás factores de $g(x)$, con lo que n es el orden de multiplicidad de a.

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 8.30 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 $f'(x) = 15x^2 - 4x$.

Las propiedades siguientes sobre derivadas se demuestran mediante meros cálculos a partir de la definición.

Teorema 8.31 *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. *$(f/g)' = (f'g - fg')/g^2$.*

La relación entre las derivadas y la multiplicidad de las raíces viene dada por el teorema siguiente:

Teorema 8.32 *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 $g(c) \neq 0$ ($n \geq 1$ 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 8.33 *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 $na_n \neq 0$, luego $f'(x) \neq 0$.

2) Si $f'(x) = 0$, entonces (con la misma notación del apartado anterior) cada coeficiente $ia_i = 0$, 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 8.34 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 $\operatorname{pol\,mín}(a,k)$.

Notar que si $p(x) = \operatorname{pol\,mín}(a,k)$ entonces a es separable si y sólo si $p'(a) \neq 0$ (por 8.32), si y sólo si $p'(x) \neq 0$, pues si $p'(a) = 0$ entonces $p(x) \mid p'(x)$, y por grados $p'(x) = 0$. Esta última condición depende sólo de p, luego si a es separable todos sus conjugados lo son también, y $\operatorname{pol\,mín}(a,k)$ 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 casi todos los cuerpos son perfectos, luego en la práctica todas las extensiones que manejaremos serán separables.

Teorema 8.35 *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 $p(x) = \operatorname{pol\,mín}(a,k)$ 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 $p(x) = \operatorname{pol\,mín}(a,k)$. Si a no fuera separable, entonces $p'(x) = 0$, luego por 8.33 $p(x) = f(x^p)$ 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

$$p(x) = \sum_{i=0}^{n} b_i^p x^{pi} = \sum_{i=0}^{n} \left(b_i x^i\right)^p = \left(\sum_{i=0}^{n} b_i x^i\right)^p,$$

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 $a \in k$. Sea K la clausura algebraica de k, sea b una raíz de $x^p - a$ en K. 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 $\operatorname{pol\,mín}(b,k) \mid (x-b)^p$, luego ha de ser $\operatorname{pol\,mín}(b,k) = (x-b)^n$, para cierto $n \leq p$, pero como k es perfecto b es raíz simple de $\operatorname{pol\,mín}(b,k)$, es decir, $n = 1$, luego $x - b = \operatorname{pol\,mín}(b,k) \in k[x]$. 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 $k^p \subset k$ y ambos tienen 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 8.36 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 8.37 *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 8.28). 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 $p(x) = \operatorname{pol}\operatorname{mín}(a, k)$ 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 8.38 Sea K/k una extensión finita. Llamaremos k–monomorfismos de K a los k–monomorfismos de K en una clausura algebraica de K.

Según hemos visto al final de la sección anterior, los k–monomorfismos de K coinciden con los k–monomorfismos de K en una clausura normal de K sobre k, y 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 (es inmediato que éste no depende de la elección de la clausura algebraica).

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 $\operatorname{pol}\operatorname{mín}(a,k)$ 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. En símbolos:

$$\mathrm{N}(k(a)/k) = |k(a) : k|.$$

Vamos a generalizar este hecho a extensiones separables cualesquiera. Primero probamos un resultado técnico:

Teorema 8.39 *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 8.25 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.

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 $\operatorname{pol}\operatorname{mín}\big(a_2, k(a_1)\big) \mid \operatorname{pol}\operatorname{mín}(a_2, k)$). 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 8.37 nos permite concluir que la extensión L/k es separable. De aquí se sigue en primer lugar:

Teorema 8.40 *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 $a_1, \ldots, a_n$ los conjugados de todos los elementos de S. Sabemos que todos ellos son separables sobre k. Entonces $L = k(a_1, \ldots, a_n)$ 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 8.41 *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 8.39 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 8.42 *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 8.10).

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 $a \in L$ (por 8.37). 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.

8.6 El teorema del elemento primitivo

Ha llegado el momento de recoger el fruto de la teoría que hemos desarrollado, aunque en realidad pospondremos los resultados más profundos que en este momento no sabríamos aprovechar. En esta sección probaremos un teorema muy útil y nada trivial que ilustra muy bien las posibilidades que da el control de los monomorfismos de una extensión.

Teorema 8.43 *(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 5.13 en K hay una raíz primitiva de la unidad, es decir, un elemento a 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 $\pm\sqrt{3}$. Un automorfismo de K está determinado por las imágenes de $\sqrt{2}$ y $\sqrt{3}$. Como sólo hay cuatro posibilidades y ha de haber cuatro automorfismos, las cuatro posibilidades se dan. Así pues, los cuatro automorfismos de K están determinados por la tabla siguiente:

Automorfismo	Imagen de $\sqrt{2}$	Imagen de $\sqrt{3}$
1	$\sqrt{2}$	$\sqrt{3}$
σ	$-\sqrt{2}$	$\sqrt{3}$
τ	$\sqrt{2}$	$-\sqrt{3}$
ρ	$-\sqrt{2}$	$-\sqrt{3}$

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. Como sólo hay seis posibilidades distintas, cada una de ellas se corresponde con un automorfismo, según indica la tabla siguiente:

Automorfismo	Imagen de α	Imagen de β
1	α	β
σ_2	α	γ
σ_3	β	α
σ_4	β	γ
σ_5	γ	α
σ_6	γ	β

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.

8.7 Normas y trazas

Finalmente podemos definir la norma de una extensión de cuerpos. Notemos primero que el teorema 8.37 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 8.44 Sea K/k una extensión separable de grado n. Sean $\sigma_1,\dots,\sigma_n$ los k-monomorfismos de K (en la clausura normal L de K/k). Definimos la *norma* y la *traza* de un elemento $\alpha\in K$ 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.

De este modo 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

$$\mathrm{N}(\alpha) = \alpha^{|K:k|}, \quad \mathrm{Tr}(\alpha) = |K:k|\,\alpha.$$

Teorema 8.45 *(Transitividad de normas) Sea $k \subset K \subset L$ una cadena de extensiones finitas separables. Entonces para todo $\alpha \in L$ se cumple*

$$\mathrm{N}_k^L(\alpha) = \mathrm{N}_k^K\big(\mathrm{N}_K^L(\alpha)\big), \quad \mathrm{Tr}_k^L(\alpha) = \mathrm{Tr}_k^K\big(\mathrm{Tr}_K^L(\alpha)\big)$$

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 $\tau_i\big(\mathrm{N}_K^L(\alpha)\big) = \rho_i\big(\sigma_1(\alpha)\big)\cdots\rho_i\big(\sigma_n(\alpha)\big)$, y $\mathrm{N}_k^K\big(\mathrm{N}_K^L(\alpha)\big)$ 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 $\mathrm{N}_k^K\big(\mathrm{N}_K^L(\alpha)\big)$, que es el producto de todos los $\rho_i\big(\sigma_j(\alpha)\big)$, 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

$$\begin{aligned}
&\Big(a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}\Big)\Big(a + b\sqrt{2} - c\sqrt{3} - d\sqrt{6}\Big)\\
&= \Big(a + b\sqrt{2}\Big)^2 - \Big(c\sqrt{3} + d\sqrt{6}\Big)^2\\
&= a^2 + 2b^2 + 2ab\sqrt{2} - 3c^2 - 6d^2 - 6cd\sqrt{2},
\end{aligned}$$

y la norma de este número en la extensión $\mathbb{Q}\big(\sqrt{2}\,\big)/\mathbb{Q}$ es

$$\begin{aligned}\mathrm{N}(\alpha) &= \left(a^2+2b^2-3c^2-6d^2+(2ab-6cd)\sqrt{2}\,\right)\\ &\cdot\ \left(a^2+2b^2-3c^2-6d^2-(2ab-6cd)\sqrt{2}\,\right)\\ &= \left(a^2+2b^2-3c^2-6d^2\right)^2-2\,(2ab-6cd)^2\,.\end{aligned}$$

■

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. 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 $i=1,\dots,p-1$.

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

$$\begin{aligned}\mathrm{Tr}\left(\sum_{i=0}^{p-1}a_i\omega^i\right) &= \sum_{i=0}^{p-1}a_i\,\mathrm{Tr}(\omega^i)=a_0\,\mathrm{Tr}(1)-\sum_{i=1}^{p-1}a_i\\ &= (p-1)a_0-\sum_{i=1}^{p-1}a_i=pa_0-\sum_{i=0}^{p-1}a_i.\end{aligned}$$

■

Capítulo IX

Grupos

A lo largo de los temas anteriores nos hemos encontrado con una extensa gama de operaciones entre objetos diversos: las operaciones de un anillo, las de un módulo, el producto y la exponenciación por enteros en un anillo, la composición de aplicaciones, etc. Sucede que la mayoría de estas operaciones presentan propiedades comunes pese a su naturaleza tan diferente, lo cual se puede poner de manifiesto introduciendo una nueva estructura algebraica, la más simple de todas, que nos proporcione una visión unitaria de las características compartidas por todas las operaciones que verifiquen unas propiedades mínimas. Ésta es la estructura de grupo. Vamos a introducirla y a tratar de convencernos de la conveniencia de pensar, siempre que sea posible, en términos de teoría de grupos.

9.1 Definición y propiedades básicas

Definición 9.1 Un *grupo* es un par $(G, \cdot)$, donde G es un conjunto y $\cdot$ es una ley de composición interna en G que cumple las propiedades siguientes:

a) $(ab)c = a(bc)$ para todos los elementos $a, b, c \in G$.

b) Existe un elemento $1 \in G$ tal que $a \cdot 1 = 1 \cdot a = a$ para todo $a \in G$.

c) Para cada elemento $a \in G$ existe otro $a^{-1} \in G$ tal que $aa^{-1} = a^{-1}a = 1$.

Como siempre, el elemento 1 cuya existencia afirma b) es único y se llama *elemento neutro* de G. También el elemento a^{-1} es único para cada $a \in G$ y se llama *elemento inverso* de a.

Un grupo $(G, \cdot)$ es *abeliano* si cumple la propiedad adicional de que $ab = ba$ para todo par de elementos $a, b \in G$.

Siguiendo nuestra costumbre, omitiremos toda mención expresa a la ley interna de un grupo cuando ello no lleve a confusión. Así, cuando hablemos de un grupo G se entenderá que lo es con cierta ley interna que representaremos con el signo $\cdot$.

En realidad con los grupos usaremos dos notaciones distintas según convenga. La *notación multiplicativa* es la notación según la cual la operación interna de

un grupo se representa mediante el signo $\cdot$, el elemento neutro se representa por 1 y el inverso de un elemento a se representa por a^{-1}. Con grupos abelianos usaremos también la llamada *notación aditiva,* según la cual la operación de un grupo se representa por el signo $+$, el elemento neutro se representa por 0 y el elemento inverso de un elemento a se representa por $-a$.

Hasta el momento nos hemos encontrado con muchos ejemplos de grupos. El objetivo de este capítulo es dar una base teórica para poder trabajar simultáneamente con todos ellos y comprender su estructura. Recopilemos ahora algunos de esos ejemplos:

1. Si $(A, +, \cdot)$ es un anillo, entonces $(A, +)$ es un grupo abeliano, al que llamaremos el *grupo aditivo* de A. El conjunto de las unidades de A es un grupo con el producto de A (abeliano también si A es conmutativo).

2. En particular, si K es un cuerpo, $K \setminus \{0\}$ es un grupo con el producto de K, lo llamaremos *grupo multiplicativo* de K. Lo representaremos por K^*.

3. Otros casos particulares de interés son los grupos $\mathbb{Z}/n\mathbb{Z}$ con la suma usual y los grupos U_n de las unidades de $\mathbb{Z}/n\mathbb{Z}$ con el producto.

4. Si A es un anillo, el conjunto $\operatorname{Aut} A$ de todos los automorfismos de A es un grupo, donde la operación es la composición de aplicaciones.

5. Si K/k es una extensión de cuerpos, entonces el conjunto $G(K/k)$ de todos los k-automorfismos de K es un grupo, llamado *grupo de Galois* de la extensión. Una extensión se dice *abeliana* si es de Galois y su grupo de Galois es abeliano.

6. Todo módulo es un grupo abeliano con la suma. Más aún, como veremos enseguida, los grupos abelianos están en correspondencia biunívoca con los $\mathbb{Z}$-módulos.

Nuestro interés por los grupos proviene en gran medida de los grupos de Galois, por lo que nos van a interesar especialmente los grupos finitos. Es costumbre llamar *orden* de un grupo G al número de sus elementos, y se representa por $|G|$.

Si G es un grupo y n un número entero, podemos definir el elemento g^n exactamente igual como hicimos en el capítulo I cuando g era una unidad en un anillo. Se cumplen las mismas propiedades. Si usamos la notación aditiva, en lugar de g^n escribiremos ng, y se cumplen las mismas propiedades que para la operación correspondiente en anillos y módulos.

De este modo, si G es un grupo abeliano, la operación externa entre números enteros y elementos de G que acabamos de definir convierte a G en un $\mathbb{Z}$-módulo cuya suma es la operación interna de G. De este modo, tal y como avanzábamos, todo $\mathbb{Z}$-módulo es un grupo abeliano y cada grupo abeliano se puede convertir en un $\mathbb{Z}$-módulo de forma natural.

Ahora vamos a definir para los grupos los conceptos análogos a los que ya tenemos definidos para anillos y módulos: subgrupos, generadores, homomorfismos, cocientes, etc.

Comenzamos por el concepto de homomorfismo de grupos:

Definición 9.2 Una aplicación $f : G \longrightarrow H$ entre dos grupos G y H es un *homomorfismo de grupos* si cumple $f(uv) = f(u)f(v)$ para todos los elementos u, v de G.

La aplicación f es un *monomorfismo, epimorfismo o isomorfismo* si además es inyectiva, suprayectiva o biyectiva, respectivamente. Un isomorfismo de un grupo G en sí mismo es un *automorfismo* de G. Llamaremos $\operatorname{Aut} G$ al conjunto de los automorfismos de un grupo G.

En general la composición de homomorfismos de grupos vuelve a ser un homomorfismo de grupos. La composición de isomorfismos es un isomorfismo, la aplicación inversa de un isomorfismo es un isomorfismo y por lo tanto, si G es un grupo, el conjunto $\operatorname{Aut} G$ resulta ser un grupo con la composición de aplicaciones como operación interna.

Diremos que dos grupos G y H son *isomorfos* cuando existe un isomorfismo entre ellos, y lo representaremos $G \cong H$. En tal caso G y H tienen las mismas propiedades, son a todos los efectos el mismo grupo.

Notar que en un grupo G, el elemento neutro 1 es el único elemento $g \in G$ que cumple $gg = g$. Como consecuencia si $f : G \longrightarrow H$ es un homomorfismo de grupos, se cumple $f(1) = 1$. Además, como $f\big(g\big)f\big(g^{-1}\big) = f(1) = 1$, también $f\big(g^{-1}\big) = f(g)^{-1}$.

Definición 9.3 Un grupo H es un *subgrupo* de un grupo G si $H \subset G$ y el producto de dos elementos de H es el mismo en H que en G. Lo representaremos $H \leq G$.

Notar que si $H \leq G$, entonces el elemento neutro de H es el mismo que el de G, pues es el único elemento $g \in G$ que cumple $gg = g$. Igualmente el inverso de un elemento $h \in H$ es su inverso en G, pues es el único elemento $g \in G$ tal que $gh = 1$.

Equivalentemente, diremos que un subconjunto H de un grupo G es un subgrupo de G si es un grupo con la operación de G. Esto supone que al operar dos elementos de H hemos de obtener un elemento de H, que el elemento neutro de G está en H y que el inverso de todo elemento de H está en H. Estas condiciones son también suficientes, aunque en realidad pueden resumirse en una sola:

Teorema 9.4 *Sea G un grupo y H un subconjunto no vacío de G. Entonces H es un subgrupo de G si y sólo si para todos los elementos g, $h \in H$ se cumple $gh^{-1} \in H$.*

Demostración: Obviamente un subgrupo ha de cumplir esta condición. Si un subconjunto H de G cumple esto, por ser no vacío existe un cierto $h \in H$, luego $1 = hh^{-1} \in H$, para todo $h \in H$ se cumple $h^{-1} = 1 \cdot h^{-1} \in H$ y para todos los g, $h \in H$ se cumple que g, $h^{-1} \in H$, luego $gh = g(h^{-1})^{-1} \in H$. Esto prueba que $H \leq G$. ∎

Todo grupo G tiene al menos dos subgrupos, el propio G y el subgrupo $1 = \{1\}$, llamado *subgrupo trivial.* Los subgrupos 1 y G se llaman *subgrupos impropios.*

Es fácil probar que si $f : G \longrightarrow H$ es un homomorfismo de grupos, entonces el *núcleo* de f, dado por $\mathrm{N}(f) = \{g \in G \mid f(g) = 1\}$ y la *imagen* de f, dada por $\operatorname{Im} f = f[G]$, son subgrupos de G y H respectivamente. Al igual que vimos en el capítulo V para anillos y en el capítulo VII para módulos, un homomorfismo de grupos es inyectivo si y sólo si su núcleo es el subgrupo trivial.

A la hora de definir los grupos cociente nos encontramos con una dificultad, y es que en grupos no abelianos podemos definir la relación de congruencia de dos formas distintas. Por ello la definición de los grupos cociente requiere algunos conceptos adicionales que veremos después. De momento nos limitaremos a definir las congruencias y probaremos algunos resultados elementales en torno a ellas, necesarios en el estudio de los subgrupos en general.

Definición 9.5 Sea G un grupo y $H \leq G$. Diremos que dos elementos u y v de G son *congruentes por la izquierda* módulo H (y lo representaremos $u \equiv_i v$ (mód H)) si cumplen $u^{-1}v \in H$.

Diremos que u y v son *congruentes por la derecha* módulo H si cumplen $uv^{-1} \in H$. Lo representaremos $u \equiv_d v$ (mód H).

Es muy fácil comprobar que ambas relaciones son de equivalencia y que la clase de equivalencia de un elemento $a \in G$ para la congruencia por la izquierda es el conjunto $aH = \{ah \mid h \in H\}$, mientras que la clase de equivalencia de a para la congruencia por la derecha es $Ha = \{ha \mid h \in H\}$.

Llamaremos $(G/H)_i$ y $(G/H)_d$ a los conjuntos cociente para las relaciones de congruencia módulo H por la izquierda y por la derecha, respectivamente.

En general se cumple que todas las clases de equivalencia tienen tantos elementos como H, pues la aplicación que a cada $h \in H$ le asigna el elemento $ah \in aH$ biyecta H con aH (e igualmente por la derecha).

Así pues, todas las clases de equivalencia de un grupo G respecto a un subgrupo H tienen cardinal igual a $|H|$, luego, si G es finito, los dos conjuntos cociente $(G/H)_i$ y $(G/H)_d$ tienen cardinal igual a $|G|/|H|$. A este cardinal lo llamaremos *índice* del subgrupo H en el grupo G, y lo representaremos por $|G : H|$. Tenemos demostrado el teorema siguiente:

Teorema 9.6 *(Teorema de Lagrange) Sea G un grupo finito y $H \leq G$. Entonces $|G| = |G : H| \cdot |H|$. En particular el orden de todo subgrupo de G es un divisor del orden de G.*

Este resultado, pese a su simplicidad, es de importancia capital en el estudio de los grupos finitos. El lector puede demostrar que en realidad se cumple también para grupos infinitos, aunque aquí no lo necesitaremos.

9.2 Grupos de permutaciones

Antes de seguir introduciendo nuevos conceptos sobre grupos, vamos a estudiar con detalle una familia concreta de grupos finitos que nos permitirá poner ejemplos claros de cuanto veremos después.

Definición 9.7 Si A es un conjunto cualquiera, llamaremos Σ_A al conjunto de todas las aplicaciones biyectivas de $f : A \longrightarrow A$. Es inmediato que Σ_A es un grupo con la operación dada por la composición de aplicaciones. Se le llama *grupo simétrico* del conjunto A. A los elementos de Σ_A se les llama también *permutaciones* de A.

Es inmediato que si dos conjuntos A y B tienen el mismo cardinal, entonces los grupos simétricos Σ_A y Σ_B son isomorfos. Basta tomar una aplicación biyectiva $f : A \longrightarrow B$ y asignar a cada permutación $g \in \Sigma_A$ la permutación $f^{-1} \circ g \circ f$.

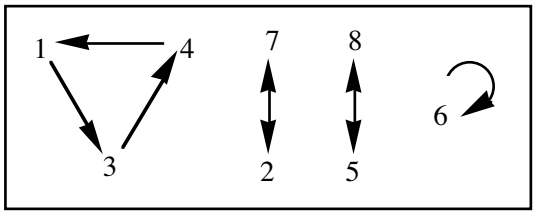
En otras palabras, da igual hablar de las permutaciones del conjunto $\{a, b, c\}$ que del conjunto $\{1, 2, 3\}$. El isomorfismo entre ambos grupos se obtiene cambiando el nombre a los elementos. Por ejemplo, la permutación que envía el 1 al 2, el 2 al 3 y el 3 al 1 se corresponde con la que envía a a b, b a c y c a a.

Por lo tanto, a efectos prácticos, y puesto que sólo nos va a interesar el caso en el que el conjunto A es finito, podemos limitarnos a estudiar los grupos simétricos sobre los conjuntos $\{1, \ldots, n\}$, donde n es un número natural. Al grupo simétrico sobre este conjunto lo representaremos por Σ_n, el grupo de las permutaciones de n elementos.

Una forma elemental de representar una permutación de n elementos f es mediante la n-tupla $\big(f(1), \ldots, f(n)\big)$. Así por ejemplo, $(3, 2, 5, 1, 4)$ es uno de los elementos del grupo Σ_5, concretamente es la permutación que envía el 1 al 3, el 2 al 2, el 3 al 5, etc.

De este modo, a cada permutación de Σ_n le corresponde una n-tupla con sus componentes distintas dos a dos, y cada una de estas n-tuplas representa a una permutación distinta. Ahora bien, una simple inducción demuestra que hay exactamente $n!$ formas distintas de disponer n objetos en una n-tupla sin repeticiones, luego podemos concluir que $|\Sigma_n| = n!$.

Consideremos la siguiente permutación de Σ_8: $(3, 7, 4, 1, 8, 6, 2, 5)$. Para ver más claramente su comportamiento podemos representarla mediante un diagrama de flechas así:



Observamos que al actuar la permutación el 1 pasa al 3, si vuelve a actuar, el 3 pasa al 4 y de aquí vuelve al 1. En definitiva, mediante sucesivas aplicaciones de la permutación, el 1 puede ir a parar al 1, al 3 o al 4, pero nunca al 2 o al 6.

En general, diremos que dos elementos a y b de $A = \{1, \ldots, n\}$ están relacionados por una permutación $\sigma \in \Sigma_n$ si existe un número entero m tal que $\sigma^m(a) = b$. Es inmediato que esta relación es de equivalencia, con lo que σ divide al conjunto A en clases de equivalencia llamadas *órbitas,* que en nuestro ejemplo son

$$\{1,\ 3,\ 4\}, \quad \{2,\ 7\}, \quad \{5,\ 8\} \quad \text{y} \quad \{6\}.$$

Sea σ una permutación de Σ_n. Sea a un número entre 1 y n. Los elementos a, $\sigma(a)$, $\sigma^2(a)$, $\sigma^3(a)$, ... están todos en la órbita de a. Como es finita, existen dos números naturales distintos tales que $\sigma^i(a) = \sigma^j(a)$. Si por ejemplo $i < j$, entonces $\sigma^{j-i}(a) = a$, o sea, existe un número natural r no nulo tal que $\sigma^r(a) = a$. Sea r el menor posible. Entonces los elementos a, $\sigma(a)$, $\sigma^2(a)$, ... , $\sigma^{r-1}(a)$ son todos distintos, pues si dos de ellos coincidieran, la diferencia de sus exponentes nos daría un número natural $s < r$ para el que $\sigma^s(a) = a$. Al seguir aplicando exponentes mayores vuelven a aparecernos los mismos elementos, $\sigma^r(a) = a$, $\sigma^{r+1}(a) = \sigma(a)$, etc. Por otra parte, como $\sigma^r(a) = a$, también $\sigma^{-1}(a) = \sigma^{r-1}(a)$, $\sigma^{-2}(a) = \sigma^{r-2}(a)$, etc, luego los exponentes negativos no introducen nuevos elementos. Esto quiere decir que en realidad toda la órbita de a, o sea, el conjunto $\{\sigma^u(a) \mid u \in \mathbb{Z}\}$, es exactamente $\big\{a, \sigma(a), \sigma^2(a), \ldots, \sigma^{r-1}(a)\big\}$.

Hemos demostrado que la órbita de un elemento a respecto a una permutación σ es de la forma $\big\{a, \sigma(a), \sigma^2(a), \ldots, \sigma^{m-1}(a)\big\}$, y $\sigma^m(a) = a$. Gráficamente, esto significa que los esquemas de flechas como el anterior dan lugar a 'círculos' de flechas

$$a \to \sigma(a) \to \sigma^2(a) \to \ldots \to \sigma^{m-1}(a) \to a$$

Las órbitas con un solo elemento se llaman *triviales.* En el ejemplo anterior hay una única órbita trivial, que es $\{6\}$.

Una permutación es un *ciclo* si forma una única órbita no trivial. Un ejemplo de ciclo en Σ_8 es la permutación $(3, 2, 4, 1, 5, 6, 7, 8)$. Si el lector representa su esquema de flechas observará que se forman las órbitas $\{1, 3, 4\}$, $\{2\}$, $\{5\}$, $\{6\}$, $\{7\}$ y $\{8\}$.

Cuando hablemos de la órbita de un ciclo, se entenderá que nos referimos a su órbita no trivial. Llamaremos *longitud* de un ciclo al cardinal de su órbita. El ciclo de nuestro ejemplo tiene longitud 3.

Resulta cómodo usar otra notación para los ciclos. Al ciclo cuya órbita es $\big\{a, \sigma(a), \sigma^2(a), \ldots, \sigma^{m-1}(a)\big\}$ lo representaremos mediante la m-tupla

$$\big(a, \sigma(a), \sigma^2(a), \ldots, \sigma^{m-1}(a)\big).$$

Por ejemplo, el ciclo que hemos mostrado antes no es sino

$$(1, 3, 4) = (3, 4, 1) = (4, 1, 3).$$

La permutación $(3,7,4,1,8,6,2,5)$ que hemos analizado no es un ciclo, pero puede expresarse como producto de ciclos: $(1,3,4)(2,7)(5,8)$.

Esto no es casual. Diremos que dos ciclos son *disjuntos* si sus órbitas no tienen elementos en común. Toda permutación distinta de la identidad se expresa como producto de ciclos disjuntos. Veamos con un ejemplo que existe un método que a partir de cualquier permutación distinta de la identidad nos permite encontrar una expresión como producto de ciclos disjuntos.

Partamos de la permutación

$$\sigma = (3,2,10,1,11,6,9,7,8,4,5) \in \Sigma_{11}$$

Esta permutación envía el 1 al 3, el 3 al 10, el 10 al 4 y el 4 al 1 de nuevo. Consideremos el ciclo $(1,3,10,4)$. Este ciclo se comporta igual que σ sobre los números 1, 3, 4 y 10. El menor número que no está aquí es el 2, pero $\sigma(2) = 2$, luego en realidad el ciclo $(1,3,10,4)$ coincide con σ también sobre el 2. El siguiente número es el 5. Ahora σ envía el 5 al 11 y el 11 al 5, luego el ciclo $(5,11)$ coincide con σ sobre el 5 y el 11. Más aún, el producto $(1,3,10,4)(5,11)$ coincide con σ sobre los números 1, 2, 3, 4, 5, 10 y 11 y deja invariantes a los restantes. El siguiente número es el 6, pero también $\sigma(6) = 6$, luego no hay nada que hacer con él. Tomamos el 7. $\sigma(7) = 9$, $\sigma(9) = 8$ y $\sigma(8) = 7$, luego el ciclo $(7,9,8)$ coincide con σ sobre 7, 8, 9, y el producto de ciclos disjuntos $(1,3,10,4)(5,11)(7,9,8)$ es igual a σ.

Notar que este proceso siempre da ciclos disjuntos porque la órbita de cada ciclo es una de las órbitas de σ y dos cualesquiera de ellas son disjuntas. La razón por la que al final obtenemos σ es que en un producto de ciclos disjuntos, la imagen de un número sólo depende del ciclo en cuya órbita se encuentre, ya que los anteriores lo fijan y los posteriores fijan también a su imagen (que está en la misma órbita).

Así pues, para saber la imagen del 4 por $(1,3,10,4)(5,11)(7,9,8)$ basta fijarse en el ciclo en el que aparece el 4, que es $(1,3,10,4)$ y observar que su imagen por él es 1. Como el 2 no aparece en ningún ciclo, su imagen es 2.

Observar que para calcular la imagen de un elemento por un producto de ciclos disjuntos no importa el orden en el que aparecen los ciclos. Esto significa que los ciclos disjuntos conmutan entre sí (su producto no depende del orden en que se multipliquen).

Otro hecho importante es que toda permutación distinta de la identidad se expresa de forma única como producto de ciclos disjuntos. Como es más fácil entenderlo que explicarlo, dejamos que el lector se convenza por sí mismo.

Ahora estamos en condiciones de representarnos claramente los grupos de permutaciones. Obviamente $\Sigma_1 = 1$, $\Sigma_2 = \big\{1,(1,2)\big\}$. Los elementos no triviales de Σ_3 pueden ser ciclos de longitud 2 o ciclos de longitud 3. En total tenemos las posibilidades siguientes:

$$\Sigma_3 = \big\{1,(1,2,3),(3,2,1),(1,2),(1,3),(2,3)\big\}.$$

Ejercicio: Considerar $G = \Sigma_3$ y $H = \big\{1,(1,2)\big\}$. Calcular los cocientes $(G/H)_i$ y $(G/H)_d$ y comprobar que son distintos.

En Σ_4 tenemos los siguientes tipos de permutaciones:

$$1, \quad (\circ,\circ,\circ,\circ), \quad (\circ,\circ,\circ), \quad (\circ,\circ), \quad (\circ,\circ)(\circ,\circ)$$

Para construir un ciclo de longitud 4 podemos comenzar con cualquiera de los 4 números. Una vez fijado el primero, tenemos 3 opciones para el siguiente (pues no se pueden repetir), 2 más para el siguiente y una sola para el último. En total hay $4 \cdot 3 \cdot 2 \cdot 1$ formas de construir un ciclo de longitud 4. Sin embargo cada ciclo puede representarse de 4 formas diferentes:

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

luego en realidad hay $3 \cdot 2 \cdot 1 = 6$ ciclos de longitud 4 en Σ_4.

Del mismo modo hay $4 \cdot 3 \cdot 2/3 = 8$ ciclos de longitud 3 y $4 \cdot 3/2 = 6$ ciclos de longitud 2. Como los elementos de tipo $(\circ,\circ)(\circ,\circ)$ pueden representarse de 8 formas distintas:

$$(1,2)(3,4) = (2,1)(3,4) = (4,3)(2,1), \ldots$$

hay un total de $4 \cdot 3 \cdot 2 \cdot 1/8 = 3$ permutaciones de este tipo. Añadiendo la identidad tenemos $1 + 6 + 8 + 6 + 3 = 24 = 4!$ elementos en Σ_4.

Ejercicio: Calcular explícitamente las 24 permutaciones de Σ_4.

Ejercicio: Determinar los tipos de permutaciones de Σ_5 así como cuántas permutaciones hay de cada tipo.

Es fácil operar con permutaciones expresadas como productos de ciclos. Por ejemplo, en Σ_4 podemos considerar el producto de $(1,2)(3,4)$ por $(1,3,2,4)$. Para calcularlo observamos que, cuando $(1,2)(3,4)(1,3,2,4)$ actúa sobre el 1, el primer ciclo lo envía al 2, el segundo deja invariante al 2 y el tercero lo envía al 4, luego la permutación total empieza así: $(1,4,\ldots)$. Ahora el primer ciclo deja fijo al 4, el segundo lo envía al 3 y el tercero envía el 3 al 2, luego tenemos $(1,4,2,\ldots)$. Igualmente vemos que el 2 va a parar al 3 y por tanto ha de ser $(1,4,2,3,\ldots)$, pero como la longitud de un ciclo no puede ser mayor que 4, seguro que el 3 va a parar al 1 y así $(1,2)(3,4)(1,3,2,4) = (1,4,2,3)$.

Con un poco de práctica el lector operará rápidamente con permutaciones.

Ejercicio: Probar que Σ_n no es abeliano, para $n \geq 3$.

Por otra parte, el inverso de un ciclo es muy fácil de hallar. Teniendo en cuenta que se trata de la aplicación inversa es inmediato que $(a,b,c,d)^{-1} = (d,c,b,a)$, es decir, el inverso de un ciclo se obtiene escribiéndolo del revés. Así mientras (a,b,c,d) lleva a a b, el ciclo (d,c,b,a) lleva b a a, etc.

El inverso de una permutación cualquiera se calcula fácilmente a partir del siguiente hecho general sobre grupos: $(gh)^{-1} = h^{-1}g^{-1}$ o, más en general todavía: $(g_1 \cdots g_n)^{-1} = g_n^{-1} \cdots g_1^{-1}$.

Por ejemplo, $\big((1,5,2)(3,6,4)(7,8)\big)^{-1} = (8,7)(4,6,3)(2,5,1)$.

Los grupos de permutaciones nos interesan, entre otros motivos, porque nos dan una representación fácil de manejar de los grupos de automorfismos de las

extensiones de Galois. En efecto, sea K/k una extensión finita de Galois, sea $p(x) \in k[x]$ un polinomio del que K sea cuerpo de escisión y sea A el conjunto de las raíces de $p(x)$ en K. Entonces $K = k(A)$ y la aplicación $f : G(K/k) \longrightarrow \Sigma_A$ dada por $f(\sigma) = \sigma|_A$ es un monomorfismo de grupos. En efecto, sabemos que cualquier k-automorfismo σ envía raíces de $p(x)$ a raíces de $p(x)$, luego $\sigma|_A$ es una aplicación inyectiva (luego biyectiva, puesto que A es finito) de A en A. Por lo tanto f está bien definida y es obviamente un homomorfismo. Del hecho de que $K = k(A)$ se sigue que si dos k-automorfismos coinciden sobre A entonces son iguales, luego f es inyectiva.

Así pues, en general, el grupo de Galois $G(K/k)$ de una extensión finita de Galois es isomorfo a un subgrupo del grupo de las permutaciones de las raíces de cualquier polinomio del cual la extensión sea cuerpo de escisión.

Ejemplo En el capítulo anterior estudiamos el cuerpo de escisión sobre $\mathbb{Q}$ del polinomio $x^3 - 2$, que es de la forma $\mathbb{Q}(\alpha, \beta)$, donde α, β, γ son las tres raíces del polinomio. Vimos que la extensión tiene seis automorfismos, que se corresponden con las seis permutaciones de las raíces, es decir, son

$$1, \quad (\alpha, \beta, \gamma), \quad (\gamma, \beta, \alpha), \quad (\alpha, \beta), \quad (\alpha, \gamma), \quad (\beta, \gamma)$$

En general la representación de un grupo de Galois como grupo de permutaciones no tiene por qué ser suprayectiva, es decir, puede haber permutaciones de raíces que no se correspondan con ningún automorfismo.

Ejemplo Consideremos el polinomio $x^4 + x^3 + x^2 + x + 1$. Su cuerpo de escisión sobre $\mathbb{Q}$ es el cuerpo ciclotómico $\mathbb{Q}(\omega)$ de grado 4. El grupo de Galois $G(\mathbb{Q}(\omega)/\mathbb{Q})$ puede identificarse con un subgrupo del grupo de las permutaciones de las cuatro raíces $\omega, \omega^2, \omega^3, \omega^4$. Como cada automorfismo está determinado por su acción sobre ω, concluimos que sólo hay 4 automorfismos. Concretamente son

$$1, \quad (\omega, \omega^2, \omega^3, \omega^4), \quad (\omega, \omega^3, \omega^4, \omega^2), \quad (\omega, \omega^4)(\omega^2, \omega^3).$$

Por ejemplo, el automorfismo que cumple $\sigma(\omega) = \omega^3$ ha de cumplir también

$$\begin{aligned} \sigma(\omega^3) &= \left(\omega^3\right)^3 = \omega^9 = \omega^4, \\ \sigma(\omega^4) &= \left(\omega^3\right)^4 = \omega^{12} = \omega^2, \\ \sigma(\omega^2) &= \left(\omega^3\right)^2 = \omega^6 = \omega. \end{aligned}$$

Por lo tanto es $(\omega, \omega^3, \omega^4, \omega^2)$.

Así pues, no existe ningún automorfismo de esta extensión que permute las raíces en la forma $(\omega, \omega^2, \omega^3)$, por ejemplo.

Ejercicio: Representar el grupo de Galois $G(\mathbb{Q}\left(\sqrt{2}, \sqrt{3}\right)/\mathbb{Q})$ como grupo de permutaciones de $\left\{\pm\sqrt{2}, \pm\sqrt{3}\right\}$.

9.3 Generadores, grupos cíclicos

Notemos que la intersección de una familia de subgrupos de un grupo es de nuevo un subgrupo, luego podemos dar una definición análoga a la vista para submódulos e ideales:

Definición 9.8 Sea G un grupo y $X \subset G$. Llamaremos *subgrupo generado* por X a la intersección de todos los subgrupos de G que contienen a X. Lo representaremos mediante $\langle X\rangle$.

Si $G = \langle X\rangle$ diremos que el conjunto X es un *generador* de G. Un grupo que admite un generador con un solo elemento es un grupo *cíclico*. Una extensión de cuerpos es *cíclica* si es de Galois y su grupo de Galois es cíclico.

Es fácil encontrar una expresión para los elementos de un grupo en función de un conjunto de generadores.

Teorema 9.9 *Sea G un grupo, X un subconjunto de G, $X^{-1} = \{x^{-1} \mid x \in X\}$ y g un elemento de G. Entonces:*

1. $\langle X\rangle = \{x_1 \cdots x_n \mid x_1, \ldots, x_n \in X \cup X^{-1}\}$.
2. *Si G es finito* $\langle X\rangle = \{x_1 \cdots x_n \mid x_1, \ldots, x_n \in X\}$.
3. $\langle g\rangle = \{g^n \mid n \in \mathbb{Z}\}$.
4. *Si* $\big|\langle g\rangle\big| = m$, *entonces* $\langle g\rangle = \{1, g, \ldots, g^{m-1}\}$ *y* $g^n = 1$ *si y sólo si* $m \mid n$.

Demostración: 1) Llamemos $H = \{x_1 \cdots x_n \mid x_1, \ldots, x_n \in X \cup X^{-1}\}$. Tenemos que $\langle X\rangle$ es un grupo que contiene a los elementos de X, luego también a sus inversos y a los productos que pueden formarse entre ellos. Por lo tanto $H \subset \langle X\rangle$. Es fácil ver que H es un subgrupo de G y obviamente contiene a X, luego $\langle X\rangle \subset H$.

2) Basta ver que si G es finito y $x \in G$, entonces $x^{-1} = x^m$ para cierto número natural m. Así un elemento de X^{-1} puede ser sustituido por varios factores (iguales) de X.

Esto es cierto porque las potencias x, x^2, x^3, ... no pueden ser todas distintas (ya que sólo pueden tomar un número finito de valores), por lo que existen dos números naturales $m < n$ tales que $x^m = x^n$.

Entonces $x^{n-m} = 1$ con $n - m > 0$, luego $x^{-1} = x^{n-m-1}$, con $n - m - 1 \geq 0$.

3) Es consecuencia inmediata de 1).

4) En la prueba de 2) hemos visto que existe un número natural $n \neq 0$ tal que $g^n = 1$. Tomemos el menor n que cumple esto. Entonces los elementos 1, g, g^2, ... g^{n-1} son todos distintos, pues si dos de ellos coincidieran, $g^i = g^j$ con $i < j$, entonces tendríamos $g^{j-i} = 1$ con $0 < j - i < n$, en contra de la elección de n. A partir de n las potencias de g se repiten, pues $g^n = 1$, $g^{n+1} = g$, $g^{n+2} = g^2$, etc. En particular $g^r = 1$ si y sólo si $n \mid r$.

Así pues, $\langle g\rangle = \{1, g, g^2, \ldots, g^{n-1}\}$, por lo que n es el orden de $\langle g\rangle$, o sea, $n = m$. ∎

El teorema anterior nos proporciona mucha información sobre los grupos cíclicos. Un hecho obvio es que todo grupo cíclico es abeliano, pues dos elementos de un grupo cíclico son de la forma g^i y g^j, para ciertos números enteros i y j, luego su producto en cualquier orden es $g^{i+j} = g^{j+i}$. Ahora conviene probar lo siguiente:

Teorema 9.10 *Todo subgrupo de un grupo cíclico es cíclico.*

DEMOSTRACIÓN: Sea $G = \langle g\rangle$ un grupo cíclico. Entonces la aplicación $f : \mathbb{Z} \longrightarrow G$ dada por $f(n) = g^n$ es un epimorfismo de grupos. Si $H \leq G$, es fácil ver que $f^{-1}[H] \leq \mathbb{Z}$ (esto es cierto para cualquier homomorfismo), pero los subgrupos de $\mathbb{Z}$ coinciden con los ideales, es decir, existe un $m \in \mathbb{Z}$ tal que $f^{-1}[H] = m\mathbb{Z}$, luego $H = \{g^{mn} \mid n \in \mathbb{Z}\} = \langle g^m\rangle$. ∎

En vista de esto resulta que las afirmaciones sobre los subgrupos de un grupo cíclico pueden reformularse en términos de sus elementos (y los subgrupos que generan). En esta línea conviene dar la definición siguiente:

Definición 9.11 Sea G un grupo y $g \in G$. Se llama *orden* de g a $\mathrm{o}(g) = \big|\langle g\rangle\big|$.

Si G es un grupo finito, entonces $\mathrm{o}(g)$ es un número natural no nulo y $\mathrm{o}(g) \;\big|\; |G|$. Un grupo de orden n es cíclico si y sólo si tiene un elemento de orden n.

Por el teorema 9.9, $g^m = 1$ si y sólo si $\mathrm{o}(g) \mid m$.

Ejercicio: Probar que todo grupo de orden primo es cíclico.

Ejercicio: Calcular todos los subgrupos de Σ_3.

Notar que el concepto de orden de una unidad en el sentido de 5.12 es un caso particular de esta definición, pues el conjunto de las unidades de un dominio es un grupo. La prueba del teorema 5.13 es válida en realidad en el siguiente contexto general:

Teorema 9.12 *Todo subgrupo finito del grupo multiplicativo de un dominio íntegro es cíclico.*

El resto del teorema 5.13, es decir, la existencia de elementos de todos los órdenes posibles, es válida para grupos cíclicos en general:

Teorema 9.13 *Sea $G = \langle g\rangle$ un grupo cíclico de orden n.*

1. *Si m es un número entero, $\langle g^m\rangle = \langle g^d\rangle$, donde $d = (m, n)$.*
2. *Si $m \mid n$, entonces $\mathrm{o}(g^m) = n/m$.*
3. *En general $\mathrm{o}(g^m) = n/(m, n)$.*

4. *Para cada divisor m de n, G tiene un único subgrupo de orden m.*

5. *Si $m \mid n$, entonces G tiene exactamente $\phi(m)$ elementos de orden m, donde ϕ es la función de Euler (5.9).*

DEMOSTRACIÓN: 1) Sea $m = rd$. Entonces $g^m = \left(g^d\right)^r \in \langle g^d \rangle$, luego $\langle g^m \rangle \subset \langle g^d \rangle$.

Por la relación de Bezout, existen enteros a y b tales que $d = am + bn$.

Así, $g^d = (g^m)^a \cdot (g^n)^d = (g^m)^a \cdot 1 = (g^m)^a \in \langle g^m \rangle$, luego $\langle g^d \rangle \subset \langle g^m \rangle$.

2) La prueba está dada en 5.13.

3) es consecuencia de 1) y 2).

4) De 2) se deduce que si $m \mid n$, entonces G tiene un subgrupo de orden m. Si G tiene dos subgrupos de orden m, digamos $\langle g^r \rangle$ y $\langle g^{r'} \rangle$, entonces $\mathrm{o}(g^r) = \mathrm{o}(g^{r'}) = m$, luego por 3) $(r, n) = (r', n) = d$, y por 1) $\langle g^r \rangle = \langle g^d \rangle = \langle g^{r'} \rangle$.

5) Sea H el único subgrupo de G de orden m. Entonces H contiene a todos los elementos de G de orden m. Aplicando a H el apartado 3) tenemos que H tiene tantos elementos de orden m como números $r \leq m$ cumplen $(r, m) = 1$, es decir, hay $\phi(m)$. ∎

Veamos algunos resultados sobre órdenes y generadores en grupos simétricos.

Teorema 9.14 *Sea n un número natural no nulo.*

1. *El orden de un ciclo de Σ_n coincide con su longitud.*

2. *El orden de un producto de ciclos disjuntos es el mínimo común múltiplo de las longitudes de dichos ciclos.*

3. *El grupo Σ_n está generado por los ciclos de longitud 2. Más aún, está generado por los $n-1$ ciclos $(1,2)$, $(2,3)$, ... , $(n-1,n)$. Los ciclos de longitud 2 se llaman trasposiciones.*

DEMOSTRACIÓN: 1) Sea un ciclo $\sigma = (a_1, \ldots, a_m)$. Es fácil ver entonces que para $1 \leq i < m$ se cumple que $\sigma^i(a_1) = a_{i+1}$, luego $\sigma^i \neq 1$. Sin embargo $\sigma^m(a_1) = a_1$ y en realidad el mismo argumento vale para cualquier a_i, luego $\sigma^m = 1$, y en consecuencia, $\mathrm{o}(\sigma) = m$.

2) Sea $\sigma = \sigma_1 \cdots \sigma_r$, donde las permutaciones σ_i son ciclos disjuntos. Sea m_i el orden de σ_i y sea m el mínimo común múltiplo de los m_i. Como los ciclos disjuntos conmutan, se cumple que $\sigma^m = \sigma_1^m \cdots \sigma_r^m = 1$, luego $\mathrm{o}(\sigma) \mid m$.

Supongamos ahora que $\sigma^s = 1$, luego como antes, $\sigma_1^s \cdots \sigma_r^s = 1$. Vamos a probar que $\sigma_1^s = 1$ y así $m_1 = \mathrm{o}(\sigma_1) \mid s$. Como las permutaciones conmutan, en realidad el argumento valdrá para todas ellas y así cada $m_i \mid s$, y por tanto $m \mid s$, lo que probará que $\mathrm{o}(\sigma) = m$.

Sea p un número en la órbita de σ_1. Como los ciclos son disjuntos, cada σ_j con $j \neq 1$ deja fijo a p, luego lo mismo hacen sus potencias y $p = \sigma^s(p) = (\sigma_1^s \cdots \sigma_r^s)(p) = \sigma_1^s(p)$.

Por otro lado, si p no está en la órbita de σ_1, entonces $\sigma_1(p) = p$ y también se cumple $\sigma^s(p) = p$. Esto prueba que $\sigma^s = 1$.

3) Se trata de probar que toda permutación puede expresarse como producto de trasposiciones. Como toda permutación es producto de ciclos, basta probar que todo ciclo se expresa como producto trasposiciones. Pero eso es fácil:

$$(a_1, \ldots, a_m) = (a_1, a_2)(a_1, a_3) \cdots (a_1, a_m).$$

■

La última afirmación del teorema anterior puede interpretarse como que podemos cambiar el orden de n elementos hasta dejarlos en cualquier disposición deseada mediante sucesivos intercambios de dos elementos cada vez.

9.4 Conjugación y subgrupos normales

Hemos visto que en general las relaciones de congruencia por la izquierda y por la derecha respecto a un subgrupo no tienen por qué coincidir, y sucede que sólo cuando coinciden puede definirse consistentemente una estructura de grupo en el conjunto cociente. Estudiemos, pues, en qué casos coinciden. Para ello nos será de utilidad el concepto de automorfismo interno de un grupo.

Definición 9.15 Sea G un grupo y g, $h \in G$. Definimos el *conjugado* de h por g como el elemento $h^g = g^{-1}hg \in G$. Definimos la función $\alpha_g : G \longrightarrow G$ dada por $\alpha_g(h) = h^g$. Es fácil comprobar que α_g es un automorfismo de G.

Más aún, la aplicación $\alpha : G \longrightarrow \operatorname{Aut} G$ dada por $\alpha(g) = \alpha_g$ es un homomorfismo de grupos. A la imagen de este homomorfismo, es decir, al subgrupo $\operatorname{Int} G = \{\alpha_g \mid g \in G\}$ se le llama grupo de los *automorfismos internos* de G.

Es evidente que $h^g = h$ si y sólo si $gh = hg$, luego $\operatorname{Int} G = 1$ si y sólo si G es abeliano.

Si $A \subset G$, escribiremos $A^g = \alpha_g[A] = \{a^g \mid a \in A\}$.

Teorema 9.16 *Sea G un grupo y H un subgrupo de G. Las siguientes condiciones son equivalentes:*

1. *La relación de congruencia módulo H por la izquierda coincide con la relación de congruencia módulo H por la derecha.*
2. *$gH = Hg$ para todo elemento $g \in G$.*
3. *$(G/H)_i = (G/H)_d$ (y entonces escribiremos simplemente G/H).*
4. *$H^g = H$ para todo elemento $g \in G$.*
5. *$H^g \subset H$ para todo elemento $g \in G$.*

DEMOSTRACIÓN: Teniendo en cuenta que gH y Hg son las clases de equivalencia de g por la izquierda y por la derecha módulo H, es claro que 1), 2) y 3) son equivalentes.

También es fácil probar que $gH = Hg$ equivale a que $H = g^{-1}Hg = H^g$. Sólo falta probar que 5) implica 4), pero esto es consecuencia de que $H^{g^{-1}} \subset H$ implica que $H \subset H^g$, luego de hecho $H = H^g$. ∎

Definición 9.17 Notar que en general, si $H \leq G$ y $g \in G$, el conjunto H^g es la imagen de H por un automorfismo de G, luego $H^g \leq G$. Los subgrupos de la forma H^g se llaman *conjugados* de H.

Dos elementos g, $h \in G$ son *conjugados* si existe un $x \in G$ tal que $g = h^x$. Es fácil probar que la conjugación es una relación de equivalencia en G. A la clase de equivalencia de un elemento $g \in G$ para la relación de conjugación se la llama *clase de conjugación* de g y se representa por $\mathrm{cl}(g)$.

Diremos que N es un *subgrupo normal* de un grupo G si cumple las condiciones del teorema anterior. Lo representaremos mediante $N \trianglelefteq G$.

Es inmediato que todo subgrupo de un grupo abeliano es normal. También es fácil ver que todo subgrupo de índice 2 es normal. La razón es que un subgrupo H con índice 2 en un grupo G da lugar a dos clases de congruencia, una es H (tanto por la izquierda como por la derecha) y la otra es $G \setminus H$, luego se cumple la condición 3) de 9.16.

Veamos las propiedades de la conjugación en grupos de permutaciones.

Teorema 9.18 *Sea n un número natural no nulo.*

1. *Si $(a_1, \ldots, a_m)$ es un ciclo en Σ_n y $\sigma \in \Sigma_n$, entonces*
$$(a_1, \ldots, a_m)^\sigma = \big(\sigma(a_1), \ldots, \sigma(a_m)\big).$$
2. *Dos permutaciones de Σ_n cuyas descomposiciones en producto de ciclos disjuntos sean $\sigma_1 \cdots \sigma_r$ y $\tau_1 \cdots \tau_s$ son conjugadas si y sólo si $r = s$ y (reordenando adecuadamente) la longitud de cada σ_i coincide con la de τ_i.*

DEMOSTRACIÓN: 1) En primer lugar,
$$\begin{aligned}\big((a_1, \ldots, a_m)^\sigma\big)\big(\sigma(a_1)\big) &= \sigma\Big((a_1, \ldots, a_m)\Big(\sigma^{-1}\big(\sigma(a_1)\big)\Big)\Big)\\ &= \sigma\big((a_1, \ldots, a_m)(a_1)\big) = \sigma(a_2)\\ &= \big(\sigma(a_1), \ldots, \sigma(a_m)\big)\big(\sigma(a_1)\big).\end{aligned}$$
Lo mismo vale para cualquier otro a_i. Si a es distinto de $\sigma(a_1), \ldots, \sigma(a_m)$, entonces $\sigma^{-1}(a)$ es distinto de $a_1, \ldots, a_m$, luego
$$(a_1, \ldots, a_m)\big(\sigma^{-1}(a)\big) = \sigma^{-1}(a)$$
y
$$\sigma\Big((a_1, \ldots, a_m)\big(\sigma^{-1}(a)\big)\Big) = \sigma\big(\sigma^{-1}(a)\big) = a = \big(\sigma(a_1), \ldots, \sigma(a_m)\big)(a).$$

Así pues $(a_1, \ldots, a_m)^\sigma$ y $\big(\sigma(a_1), \ldots, \sigma(a_m)\big)$ actúan igual sobre todos los elementos, luego son iguales.

2) Veámoslo con un ejemplo: Tomamos una permutación de Σ_7:

$$(1,3,5)(2,4)(6,7)$$

Si la conjugamos por otra cualquiera, digamos por $g = (1,2,6)(3,5)$, usando 1) obtenemos la siguiente permutación (sólo hay que reemplazar cada número por su imagen por g):

$$(2,5,3)(6,4)(1,7)$$

No es casual que obtengamos ciclos disjuntos, pues cada ciclo de órbita A se transforma en otro de órbita $g[A]$, luego si dos ciclos tienen órbitas disjuntas A y B, las órbitas de sus conjugados, $g[A]$ y $g[B]$ también son disjuntas. Esto prueba que dos permutaciones conjugadas tienen que tener descomposiciones en ciclos disjuntos del mismo tipo en el sentido de 2).

Tomemos ahora dos permutaciones del mismo tipo:

$$(1,3,5)(2,4)(6,7)$$

$$(6,1,7)(3,5)(2,4)$$

Basta considerar la permutación $g = (1,6,2,3)(4,5,7)$, es decir, la que envía a cada número de la permutación de arriba al que ocupa su lugar en la permutación de abajo, para que al conjugar la primera por g se obtenga la segunda. ∎

Algunos ejemplos:

El grupo Σ_3 tiene 6 elementos divididos en 3 clases de conjugación, a saber, la del elemento neutro $\{1\}$, la de las trasposiciones $\big\{(1,2),(1,3),(2,3)\big\}$ (de orden 2) y la de los ciclos de longitud 3, $\big\{(1,2,3),(3,2,1)\big\}$ (de orden 3).

Notar que en general todos los elementos de una misma clase de conjugación tienen el mismo orden, ya que existe un automorfismo que envía uno a otro, y los automorfismos conservan el orden.

Por definición, un subgrupo es normal si y sólo si es unión de clases de conjugación, luego los únicos subgrupos normales de Σ_3 son 1, $\big\langle(1,2,3)\big\rangle$ y Σ_3 (el segundo es la unión de $\{1\}$ y la clase de los ciclos de longitud 3).

Otro hecho general de interés es que si un grupo tiene un único subgrupo de un cierto orden, éste ha de ser normal, pues no puede tener más conjugados que él mismo. Esto nos da otra prueba de que $\big\langle(1,2,3)\big\rangle$ es normal en Σ_3. Un tercer argumento es que tiene índice 2.

Ejercicio: Determinar el número de clases de conjugación de Σ_4, así como el número de elementos de cada una de ellas.

Los órdenes posibles para los subgrupos de Σ_4 son 1, 2, 3, 4, 6, 8, 12 y 24. Tratemos de hallar al menos los subgrupos normales.

Ningún subgrupo normal propio puede contener a las trasposiciones, ya que generan todo el grupo. Tampoco puede contener a los ciclos de longitud 4, ya

que son 6, más el neutro 7, luego el grupo debería tener orden 8 o 12, pero no hay ninguna otra clase de conjugación de cardinal 1 o 5 que podamos añadir para completar esos 7 elementos.

Es fácil ver que

$$V_4 = \{1\} \cup \mathrm{cl}\big((1,2)(3,4)\big) = \big\{1, (1,2)(3,4), (1,3)(2,4), (1,4)(2,3)\big\}$$

es un subgrupo normal de Σ_4 en el que todos los elementos no triviales tienen orden 2 (o sea, que no es cíclico). A este grupo se le llama el *grupo de Klein.*

Para que un subgrupo normal propio contenga a los ciclos de longitud 3 (que son 8 más el neutro 9) ha de tener orden 12, luego faltan los 3 elementos de $\mathrm{cl}\big((1,2)(3,4)\big)$. El lector puede probar directamente que

$$A_4 = \{1\} \cup \mathrm{cl}\big((1,2)(3,4)\big) \cup \mathrm{cl}\big((1,2,3)\big)$$

es un subgrupo normal de orden 12 llamado *grupo alternado* de grado 4.

Ejercicio: Probar que $A_4 = \big\langle (1,2,3), (1,2)(3,4) \big\rangle$.

En resumen, los únicos subgrupos normales de Σ_4 son $1 < V_4 < A_4 < \Sigma_4$. Es fácil ver que V_4 es abeliano, mientras que A_4 no lo es.

Todos los subgrupos de Σ_4 de orden 2 o 3 son cíclicos, pues tienen orden primo. Hay tantos subgrupos de orden 2 como elementos de orden 2, o sea, 9. Cada subgrupo de orden 3 contiene dos elementos de orden 3 que lo generan, luego hay 4 subgrupos de orden 3. Todo subgrupo de orden 12 es normal por tener índice 2, luego A_4 es el único.

Ejercicio: Si $H \trianglelefteq K \trianglelefteq G$, ¿es necesariamente $H \trianglelefteq G$?

9.5 Producto de grupos

Introducimos ahora un producto de grupos similar a la suma de módulos, aunque en el caso que nos ocupa hay que prestar atención a ciertas particularidades, especialmente si los grupos no son abelianos.

Definición 9.19 Sea G un grupo y A y B dos subconjuntos de G. Llamaremos $AB = \{ab \mid a \in A, b \in B\}$. En notación aditiva hablaremos de suma de subconjuntos y la representaremos $A + B$.

A diferencia de lo que ocurre con módulos, el producto de dos subgrupos no es necesariamente un subgrupo. El teorema siguiente aclara la situación.

Teorema 9.20 *Sea G un grupo y H y K dos subgrupos de G.*

1. *$HK \leq G$ si y sólo si $HK = KH$.*
2. *Si $H \trianglelefteq G$ o $K \trianglelefteq G$, entonces $HK \leq G$.*
3. *Si $H \trianglelefteq G$ y $K \trianglelefteq G$, entonces $HK \trianglelefteq G$.*

DEMOSTRACIÓN: 1) Si $HK \leq G$ y $x \in HK$, entonces $x^{-1} \in HK$, luego es de la forma $x^{-1} = hk$, con $h \in H$ y $k \in K$. Por lo tanto $x = k^{-1}h^{-1} \in KH$. Igualmente se prueba la otra inclusión, luego $HK = KH$.

Si $HK = KH$, sean $x, y \in HK$. Entonces $x = hk$ e $y = h'k'$ con $h, h' \in H$ y $k, k' \in K$. Por lo tanto $xy^{-1} = hkk'^{-1}h'^{-1}$. El elemento $kk'^{-1}h'^{-1} \in KH = HK$, luego $kk'^{-1}h'^{-1} = h''k''$ para ciertos $h'' \in H$ y $k'' \in K$. Consecuentemente $xy^{-1} = hh''k'' \in HK$. Por el teorema 9.4 tenemos que $HK \leq G$.

2) Si $H \trianglelefteq G$ y $hk \in HK$, entonces $hk = kk^{-1}hk = kh^k \in KH^k = KH$ e igualmente se prueba la otra inclusión. Por lo tanto $HK = KH$ y por 1) $HK \leq G$.

3) Si $g \in G$, como la conjugación por g es un automorfismo de G, se cumple que $(HK)^g = H^gK^g = HK$. Por lo tanto $HK \trianglelefteq G$. ∎

El lector debe tener clara la diferencia entre que $HK = KH$ y que $hk = kh$ para todo $h \in H$ y todo $k \in K$. En el primer caso se dice que H y K conmutan. En el segundo se dice que conmutan elemento a elemento. La segunda propiedad implica obviamente la primera, pero el recíproco no es cierto. Es fácil encontrar ejemplos en el grupo Σ_3. A menudo es útil este sencillo resultado:

Teorema 9.21 *Si dos subgrupos normales de un grupo G tienen intersección trivial, entonces conmutan elemento a elemento.*

DEMOSTRACIÓN: Sean N y $M \trianglelefteq G$ tales que $N \cap M = 1$. Sean $n \in N$ y $m \in M$. Entonces $n^{-1}m^{-1}nm = \left(m^{-1}\right)^n m = n^{-1}n^m \in N \cap M = 1$, luego $nm = mn$. ∎

Definición 9.22 Diremos que un grupo G es *producto directo* de los subgrupos $N_1, \ldots, N_r$ si todos son normales en G, $G = N_1 \cdots N_r$ y la intersección de cada N_i con el producto de los factores restantes es trivial. En tal caso se escribe $G = N_1 \times \cdots \times N_r$.

Para grupos abelianos el concepto de producto directo coincide con el concepto de suma directa de subgrupos vistos como $\mathbb{Z}$-módulos. Simplemente hemos cambiado la notación aditiva por la multiplicativa (Notar que $N_1 \cdots N_r = \langle N_1, \ldots, N_r \rangle$). Teniendo en cuenta el teorema anterior, la prueba del siguiente resultado es análoga a la del teorema 7.15.

Teorema 9.23 *Sea G un grupo y $N_1, \ldots, N_r$ una familia de subgrupos normales tales que $G = N_1 \times \cdots \times N_r$. Entonces:*

1. *Si se da la igualdad $n_1 \cdots n_r = 1$ con cada $n_i \in N_i$, entonces cada $n_i = 1$.*
2. *Cada elemento $g \in G$ se expresa de forma única como producto $g = n_1 \cdots n_r$ con cada $n_i \in N_i$.*

Al igual que ocurre con los módulos, si tenemos una familia de grupos podemos construir un grupo que se exprese como producto directo de una familia de subgrupos isomorfos a los dados.

Definición 9.24 Sea una familia de grupos $G_1, \ldots, G_r$. Llamaremos *producto directo externo* de $G_1, \ldots, G_r$ al grupo

$$G_1 \times \cdots \times G_r = \big\{(g_1, \ldots, g_r) \;\big|\; g_j \in G_j \text{ para cada } j = 1, \ldots, r\big\}$$

con la operación dada por $(g_1, \ldots, g_r)(g'_1, \ldots, g'_r) = (g_1 g'_1, \ldots, g_r g'_r)$.

Como en el caso de los módulos, la aplicación $\iota_i : G_i \longrightarrow G_1 \times \cdots \times G_r$ que a cada elemento $g \in G_i$ le asigna la r-tupla cuya componente i-ésima es g y las restantes son 1 es un monomorfismo de grupos, por lo que podemos identificar cada G_i con su imagen, es decir, con el subgrupo N_i de $G_1 \times \cdots \times G_r$ formado por las r-tuplas que tienen todas sus componentes iguales a 1 salvo quizá la i-ésima. Es fácil ver que cada N_i es isomorfo a G_i y que $G_1 \times \cdots \times G_r$ es producto directo (en el sentido de 9.22) de los subgrupos N_i.

9.6 Grupos cociente

Pasamos por fin al estudio de los grupos cociente. Resulta que sólo pueden definirse cuando la congruencia izquierda coincide con la derecha, es decir, cuando el subgrupo es normal.

Teorema 9.25 *Sea G un grupo y $N \trianglelefteq G$. Entonces el conjunto cociente G/N es un grupo con la operación dada por $(gN)(hN) = ghN$. El elemento neutro de G/N es $1N = N$. Si $g \in G$, se cumple que $(gN)^{-1} = g^{-1}N$. Si el grupo G es abeliano o cíclico, entonces G/N también lo es.*

Demostración: Hay que probar que la operación en G/N está bien definida, es decir, que si $gN = g'N$ y $hN = h'N$, entonces $ghN = g'h'N$.

Para ello observamos que $(gh)^{-1}g'h' = h^{-1}g^{-1}g'h' = h^{-1}h'h'^{-1}g^{-1}g'h'$.

Ahora, $h^{-1}h' \in N$ y $g^{-1}g' \in N$ porque por hipótesis son congruentes módulo N, y como N es normal, $h'^{-1}g^{-1}g'h' = (g^{-1}g')^{h'} \in N$, luego $(gh)^{-1}g'h' \in N$ y así $ghN = g'h'N$.

El resto del teorema es obvio. En todo caso, nótese que si $G = \langle g \rangle$, entonces es claro que $G/N = \langle gN \rangle$. ∎

Los teoremas sobre anillos y módulos cociente se prueban con pocos cambios para grupos. El más importante es el teorema de isomorfía:

Teorema 9.26 *(Teorema de Isomorfía) Sea $f : G \longrightarrow H$ un homomorfismo de grupos. Entonces $\mathrm{N}(f) \trianglelefteq G$, y la aplicación $\bar{f} : G/\mathrm{N}(f) \longrightarrow \mathrm{Im}\, f$ dada por $\bar{f}\big(g\,\mathrm{N}(f)\big) = f(g)$ es un isomorfismo de grupos.*

Recíprocamente, todo subgrupo normal N de un grupo G es el núcleo de un epimorfismo de grupos, a saber, de la *proyección canónica* $p : G \longrightarrow G/N$ definida mediante $p(g) = gN$.

Como un ejemplo de uso del teorema anterior, notemos que si $G = \langle g \rangle$ es un grupo cíclico de orden n, entonces la aplicación $f : \mathbb{Z} \longrightarrow G$ dada por $f(n) = g^n$ es un epimorfismo de grupos y, según 9.9, tenemos que $\mathrm{N}(f) = n\mathbb{Z}$. Por el

teorema de isomorfía resulta que G es isomorfo a $\mathbb{Z}/n\mathbb{Z}$, de donde concluimos que dos grupos cíclicos finitos son isomorfos si y sólo si tienen el mismo orden. Por otra parte, si G es infinito la aplicación f es un isomorfismo, luego todo grupo cíclico infinito es isomorfo a $\mathbb{Z}$.

Teorema 9.27 *Sea G un grupo y $N \trianglelefteq G$.*

1. *Si $N \leq H \leq G$, entonces $H/N \leq G/N$.*
2. *Si $K \leq G/N$, entonces existe un subgrupo H de G tal que $N \leq H \leq G$ y $K = H/N$.*
3. *Si $N \leq H \leq G$, entonces $H/N \trianglelefteq G/N$ si y sólo si $H \trianglelefteq G$.*
4. *Las correspondencias descritas en los apartados 1) y 2) determinan una biyección entre los subgrupos de G/N y los subgrupos de G que contienen a N. Los subgrupos normales de G se corresponden con los subgrupos normales de G/N.*

Demostración: 1) Es inmediato. Los elementos de H/N son las clases hN con $h \in H$ y los de G/N son las clases gN con $g \in G$. La operación es la misma.

2) Consideremos el epimorfismo canónico $p : G \longrightarrow G/N$ dado por $p(g) = gN$. Definimos $H = p^{-1}[K] \leq G$. Como $N = p^{-1}[1N]$, claramente $N \leq H$. Así $H/N = \{hN \mid h \in H\} = p[H] = p\big[p^{-1}[K]\big] = K$.

3) Si $H/N \trianglelefteq G/N$, entonces al conjugar un elemento $h \in H$ por un elemento $g \in G$ se cumple $h^g N = (hN)^{gN} \in H/N$, luego $h^g \in H$ y así $H \trianglelefteq G$. El recíproco es análogo.

4) Si $H/N = H'/N$, entonces $H = p^{-1}[H/N] = p^{-1}[H'/N] = H'$, luego la correspondencia es inyectiva, y por 2) es suprayectiva. ∎

Veamos un par de resultados adicionales muy útiles cuando se manejan grupos cociente.

Teorema 9.28 *(Segundo teorema de isomorfía) Sea G un grupo, $H \leq G$ y $K \trianglelefteq G$. Entonces $HK/K \cong H/(H \cap K)$.*

Demostración: Consideremos la aplicación $f : H \longrightarrow HK/K$ dada por $f(h) = hK$. Es claro que se trata de un homomorfismo de grupos. Además es un epimorfismo, pues un elemento de HK/K es de la forma hkK con $h \in H$ y $k \in K$, pero $hkK = hK = f(h)$.

Un elemento $h \in H$ está en $\mathrm{N}(f)$ si y sólo si $h \in H$ y $hK \in K$, si y sólo si $h \in H \cap K$, luego $\mathrm{N}(f) = H \cap K$ y por el teorema de isomorfía concluimos $HK/K \cong H/(H \cap K)$. ∎

Teorema 9.29 *(Tercer teorema de isomorfía) Consideremos un grupo G y dos subgrupos $K \trianglelefteq G$ y $K \leq H \trianglelefteq G$. Entonces $(G/K)/(H/K) \cong G/H$.*

DEMOSTRACIÓN: Un elemento de G/K es de la forma gK con $g \in G$, luego un elemento cualquiera de $(G/K)/(H/K)$ es de la forma $(gK)(H/K)$. Además $(gK)(H/K) = (1K)(H/K)$ si y sólo si $gK \in H/K$, es decir, si y sólo si $g \in H$.

Esto significa que la aplicación $f : G \longrightarrow (G/K)/(H/K)$ definida mediante $f(g) = (gK)(H/K)$ es un epimorfismo de núcleo H, luego por el teorema de isomorfía, $(G/K)/(H/K) \cong G/H$. ∎

Ejercicio: Enunciar y demostrar teoremas de isomorfía análogos para módulos.

El segundo teorema de isomorfía implica que si H y K son subgrupos de un grupo finito G y K es normal, entonces $|HK| = |H|\,|K|/|H \cap K|$. Vamos a probar que esto sigue siendo cierto aunque ninguno de los subgrupos sea normal y HK no sea un subgrupo.

Teorema 9.30 *Sea G un grupo finito y H, K dos subgrupos de G. Entonces*

$$|HK| = \frac{|H|\,|K|}{|H \cap K|}.$$

DEMOSTRACIÓN: Consideremos la aplicación $f : H \times K \longrightarrow HK$ dada por $f(h,k) = hk$. Obviamente es suprayectiva. Si $f(h,k) = f(h',k')$, entonces $hk = h'k'$, luego $u = (h')^{-1}h = k'k^{-1} \in H \cap K$.

Hemos probado que si $f(h,k) = f(h',k')$, entonces $(h',k') = (hu, u^{-1}k)$ para cierto $u \in H \cap K$. El recíproco es trivialmente cierto. Esto significa que para cada $hk \in HK$ se cumple que $f^{-1}[hk] = \big\{(hu, u^{-1}k) \;\big|\; u \in H \cap K\big\}$, luego $\big|f^{-1}[hk]\big| = |H \cap K|$.

En consecuencia, el número de elementos de $H \times K$ es igual al número de conjuntos de la forma $f^{-1}[hk]$ (que es $|HK|$) multiplicado por el número de elementos de cada uno de estos conjuntos (que es $|H \cap K|$), es decir, hemos probado que $|H|\,|K| = |HK|\,|H \cap K|$. ∎

9.7 Grupos alternados

Terminamos el capítulo con un concepto importante sobre grupos de permutaciones.

Definición 9.31 Sea $n \geq 2$, sea $P_n = \big\{\{i,j\} \;\big|\; 1 \leq i < j \leq n\big\}$. Para cada permutación $\sigma \in \Sigma_n$ y cada $b = \{i,j\}$ con $1 \leq i < j \leq n$, sea

$$\epsilon(\sigma, b) = \begin{cases} 1 & \text{si } \sigma(i) < \sigma(j) \\ -1 & \text{si } \sigma(j) < \sigma(i) \end{cases}$$

Llamaremos *signatura* de σ a

$$\operatorname{sig}\sigma = \prod_{b \in P_n} \epsilon(\sigma, b) \in \{1, -1\}.$$

Las permutaciones de signatura 1 se llaman *permutaciones pares.* Las de signatura -1 se llaman *impares.*

Enseguida daremos una interpretación sencilla de este concepto. Primero conviene probar lo siguiente:

Teorema 9.32 *Sea $n \geq 2$. Entonces la aplicación* $\operatorname{sig} : \Sigma_n \longrightarrow \{-1, 1\}$ *es un homomorfismo de grupos.*

DEMOSTRACIÓN: Sean σ, $\tau \in \Sigma_n$ y sea $b \in P_n$. Es fácil comprobar que $\epsilon(\sigma\tau, b) = \epsilon(\sigma, b)\epsilon\big(\tau, \sigma[b]\big)$. Como la aplicación $P_n \longrightarrow P_n$ dada por $b \mapsto \sigma[b]$ es biyectiva, se cumple que

$$\operatorname{sig}(\sigma\tau) = \prod_{b\in P_n} \epsilon(\sigma\tau, b) = \prod_{b\in P_n} \epsilon(\sigma, b)\epsilon\big(\tau, \sigma[b]\big)$$

$$= \prod_{b\in P_n} \epsilon(\sigma, b) \prod_{b\in P_n} \epsilon\big(\tau, \sigma[b]\big) = \prod_{b\in P_n} \epsilon(\sigma, b) \prod_{b\in P_n} \epsilon(\tau, b) = (\operatorname{sig}\sigma)(\operatorname{sig}\tau).$$

■

Consideremos la trasposición $(1, 2) \in \Sigma_n$. Es claro que $\epsilon\big((1, 2), b\big) = -1$ si y sólo si $b = \{1, 2\}$, luego $\operatorname{sig}(1, 2) = -1$. Más aún, todas las trasposiciones son conjugadas (por el teorema 9.18), y obviamente

$$\operatorname{sig}(\sigma^\tau) = (\operatorname{sig}\sigma)^{\operatorname{sig}\tau} = \operatorname{sig}\sigma,$$

pues $\{+1, -1\}$ es un grupo abeliano. Esto implica que todas las trasposiciones tienen signatura -1. Si unimos esto al teorema anterior y al teorema 9.14, tenemos probado el teorema siguiente:

Teorema 9.33 *Una permutación es par o impar si y sólo si se descompone en un número par o impar de trasposiciones, respectivamente.*

Es fácil reconocer la signatura de una permutación descompuesta en ciclos. Basta recordar que, según la prueba de 9.14 3), un ciclo de longitud m se descompone en $m-1$ trasposiciones, luego un ciclo es par si y sólo si su longitud es impar.

Definición 9.34 Llamaremos *grupo alternado* de grado n al grupo A_n formado por las permutaciones pares de Σ_n, es decir, al núcleo del homomorfismo sig.

Por el teorema de isomorfía, $\Sigma_n/A_n \cong \{1, -1\}$, luego $|\Sigma_n : A_n| = 2$, es decir, $|A_n| = n!/2$.

Capítulo X

Matrices y determinantes

Recordemos que nuestra intención es estudiar anillos como $\mathbb{Z}\big[\sqrt{-2}\,\big]$ y hasta ahora sólo tenemos una teoría razonable sobre cuerpos como $\mathbb{Q}\big(\sqrt{-2}\,\big)$. La razón es que la teoría de cuerpos se apoya en la teoría de espacios vectoriales, mientras que el análogo para anillos es la teoría de módulos, que no es tan potente o, al menos, requiere razonamientos más delicados para conseguir resultados que en el caso de espacios vectoriales son mucho más simples. En este capítulo introduciremos dos poderosas herramientas de la teoría de módulos con las que finalmente estaremos en condiciones de abordar los anillos numéricos.

10.1 Matrices

Definición 10.1 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

$$B^j = \begin{pmatrix} b_{1j} \\ \vdots \\ b_{m,j} \end{pmatrix}$$

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):

$A(BC) = (AB)C$.
$A(B+C) = AB + AC$.
$(A+B)C = AC + BC$.
$AI_n = I_m A = A$.
Si A es conmutativo $(AB)^t = B^t A^t$.

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. 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 10.2 *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

$$f(u) = \sum_{i=1}^m x_i f(u_i) = \sum_{i=1}^m x_i \sum_{j=1}^n a_{ij} v_j = \sum_{j=1}^n \left(\sum_{i=1}^m x_i a_{ij} \right) v_j,$$

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 10.3 Si M y N son A-módulos libres de rangos m y n, llamaremos $\mathrm{Hom}_A(M, N)$ al conjunto de todos los homomorfismos entre 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 7.31 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 $\mathrm{M}_B^{B'}(f) = C$.

Ejercicio: Calcular la matriz asociada a los automorfismos del cuerpo ciclotómico $\mathbb{Q}(\omega)$, donde $\omega^5 = 1$, respecto a la base ordenada $(1, \omega, \omega^2, \omega^3)$.

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 $(af)(m) = a\big(f(m)\big)$ (notar 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 10.2, $\mathrm{M}_B^{B'}(f+g) = \mathrm{M}_B^{B'}(f) + \mathrm{M}_B^{B'}(g)$.

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 10.4 *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 10.2, $\mathrm{M}_B^{B''}(f \circ g) = \mathrm{M}_B^{B'}(f)\mathrm{M}_{B'}^{B''}(g)$. ∎

El espacio $\mathrm{Hom}_A(M, M)$ es un anillo unitario con la suma y la composición de aplicaciones. Acabamos de probar que la aplicación $\mathrm{M}_B^{B'}$ es un isomorfismo. Notar que la matriz identidad se corresponde con la aplicación identidad.

El lector debe tener presente que todas las propiedades sobre los conjuntos de matrices $\mathrm{Mat}_{m\times n}(A)$ se traducen a propiedades de los espacios $\mathrm{Hom}_A(M, N)$ a través de los isomorfismos que hemos definido.

Definición 10.5 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 $C^{-1} \in Mat_n(A)$ 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.

Teorema 10.6 *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. ∎

Definición 10.7 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'.

Ejercicio: Sabemos que el cuerpo ciclotómico $\mathbb{Q}(\omega)$ coincide, para $p = 3$, con el cuerpo cuadrático $\mathbb{Q}\big(\sqrt{-3}\big)$. Concretamente, $\omega = \big(-1 + \sqrt{-3}\big)/2$. Calcular la matriz de cambio de base asociada a $\big(1, \sqrt{-3}\big)$ y $(1, \omega)$.

Terminamos las propiedades generales sobre matrices con las observaciones siguientes:

Teorema 10.8 *Se cumple*

1. *Si A es un dominio íntegro y B, $C \in \mathrm{Mat}_n(A)$ cumplen que $BC = I_n$, entonces B y C son regulares y $C = B^{-1}$.*

2. *Si A es un dominio íntegro y $B \in \mathrm{Mat}_{m\times n}(A)$, $C \in \mathrm{Mat}_{n\times m}(A)$ cumplen que $BC = I_m$, $CB = I_n$, entonces $n = m$, B y C son regulares y $C = B^{-1}$.*

Demostración: Sea K el cuerpo de cocientes de A. Entonces $\mathrm{Mat}_n(A)$ puede considerarse como un subanillo de $\mathrm{Mat}_n(K)$. Fijemos una base del espacio vectorial K^n y consideremos las aplicaciones lineales f, $g : K^n \longrightarrow K^n$ cuyas matrices en la base considerada sean B y C respectivamente. Entonces la matriz de $f \circ g$ es I_n, lo que significa que $f \circ g$ es la aplicación identidad. De aquí se sigue que f es un monomorfismo, luego $\dim \mathrm{Im}\, f = \dim K^n = n$. Por 7.26 tenemos que $\mathrm{Im}\, f = K^n$, luego f es un isomorfismo y por el teorema anterior B es regular. Multiplicando por B^{-1} en $BC = I_n$ obtenemos que $C = B^{-1}$.

La prueba de 2 es análoga. ∎

10.2 Determinantes

Pasamos ahora al estudio de los determinantes de matrices cuadradas. En lugar de definir directamente el concepto de determinante, que puede resultar artificial, vamos a establecer las propiedades que deseamos que cumplan los determinantes y concluiremos que la única definición posible es la que vamos a adoptar.

Definición 10.9 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 10.10 *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 $\operatorname{car} A$ es impar 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 10.11 *Sea A un dominio y n un número natural no nulo. Para cada $i = 1, \ldots, n$ sea $e_i = (\delta_{i1}, \ldots, \delta_{in})$, es decir, la n-tupla que tiene un 1 en la posición i-ésima y 0 en las restantes. 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

$$f(v_1, \ldots, v_n) = a \sum_{\sigma \in \Sigma_n} (\operatorname{sig} \sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}. \tag{10.1}$$

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 (10.1). 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 $a_{ij} = \delta_{ij}$ y el único sumando no nulo en (10.1) 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 10.12 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 $a \in A$.

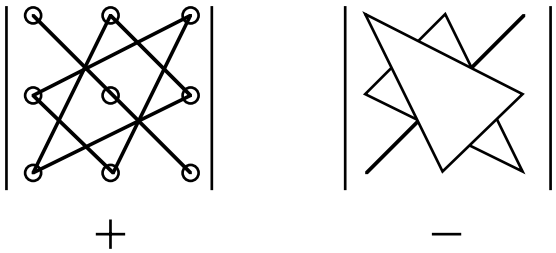
Para $n = 2$ tenemos la fórmula:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

Para $n = 3$ hay 6 sumandos, tres con signo positivo y tres con signo negativo. El lector puede comprobar que el desarrollo es el siguiente:

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

El esquema siguiente (conocido como regla de Sarrus) permite recordar fácilmente la fórmula:



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 10.13 *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

$$|B^t| = \sum_{\sigma\in\Sigma_n} (\operatorname{sig}\sigma) b_{1\sigma(1)}\cdots b_{n\sigma(n)} = |B|.$$

■

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 10.14 *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 $B_1, \ldots, B_n$. Notar 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 10.11, 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 para $a = |C|$ (fórmula (10.1)), 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 10.15 *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 10.13).

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 10.16 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 10.17 *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 10.17 es un caso particular de un resultado más general:

Teorema 10.18 *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 10.19 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

$$\tilde{b}_{ij} = (-1)^{i+j}B_{ji}.$$

Notar 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 10.20 *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 10.18, 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 10.21 *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, $BC = CB = In$, 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.

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:

$$\left.\begin{array}{rl} a_{11}x_1 + \cdots + a_{1n}x_n & = b_1 \\ \cdots\cdots\cdots\cdots\cdots & \\ a_{n1}x_n + \cdots + a_{nn}x_n & = b_n \end{array}\right\}$$

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 10.18, 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 10.22 *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 7.31 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 7.31 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 10.23 *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 7.30). Sea $R = (\mathbb{Z}/a_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/a_n\mathbb{Z}) \times \mathbb{Z}^{m-n}$.

Sea $f : M \longrightarrow R$ el homomorfismo que a cada z_i lo envía a la n-tupla que tiene un 1 en la posición i-ésima y 0 en las restantes. Obviamente f es un epimorfismo y un elemento $c_1 z_1 + \cdots + c_m z_m$ está en $\mathrm{N}(f)$ si y sólo si $\big([c_1], \ldots, [c_n], c_{n+1}, \ldots, c_m\big) = 0$, lo que equivale a que $a_i \mid c_i$ para $i = 1, \ldots, n$ y $c_i = 0$ para $i = n + 1, \ldots, m$.

Consecuentemente, $\mathrm{N}(f) = \langle a_1 z_1, \ldots, a_n z_n \rangle = N$, y por el teorema de isomorfía $M/N \cong (\mathbb{Z}/a_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/a_n\mathbb{Z}) \times \mathbb{Z}^{m-n}$.

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 10.21 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 10.4 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,

$$|\det B| = |\det P|\,|\det C|\,|\det Q| = |\det C| = |M/N|.$$

■

10.3 Formas bilineales

En las secciones anteriores hemos visto cómo las matrices describen homomorfismos entre módulos libres y cómo los determinantes aportan información sobre las matrices. Aquí emplearemos la matrices (y por lo tanto los determinantes) para describir otros objetos algebraicos de interés: las formas bilineales. Aunque podríamos trabajar más en general, nos centramos en el caso que realmente nos va a interesar en la práctica.

Definición 10.24 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 $F(v_1, v_2) = F(v_2, v_1)$.

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 10.25 *Sea $F : V \times V \longrightarrow K$ una forma bilineal en un K-espacio vectorial V. Sea $B = (v_1, \ldots, v_m)$ 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 $\Phi_B(v) = (x_1, \ldots, x_m)$ y $\Phi_B(v') = (y_1, \ldots, y_n)$. Entonces

$$\begin{aligned} F(v, v') &= F\left(\sum_{i=1}^{m} x_i v_i, \sum_{j=1}^{n} y_j v_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i\, F(v_i, v_j)\, y_j \\ &= \Phi_B(v)\, \mathrm{M}_B(F)\, \Phi_B(v')^t. \end{aligned}$$

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.

Ejercicio: Encontrar la relación entre las matrices de una misma forma bilineal en dos bases distintas.

Las formas bilineales están muy relacionadas con el concepto de espacio dual, que introducimos seguidamente.

Definición 10.26 Si K es un cuerpo y V es un K-espacio vectorial, llamaremos *espacio dual* de V a $V^* = Hom(V, K)$.

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 $v_i^*(v_j) = \delta_{ij}$. La base $B^* = (v_1^*, \ldots, v_n^*)$ 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 10.27 *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 $\iota(v_i)$, 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 10.28 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 10.6 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.

Al igual que las demás construcciones algebraicas abstractas, las formas bilineales aparecen en contextos muy diversos en el estudio de los números. Dado el nivel introductorio de este libro, a nosotros sólo nos aparecerán en el ejemplo siguiente:

Teorema 10.29 *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 $B = (v_1, \ldots, v_n)$ de K. Sean $\sigma_1, \ldots, \sigma_n$ los k-monomorfismos de K. Entonces

$$\begin{aligned} \mathrm{M}_B &= (\mathrm{Tr}(v_i v_j)) = \left(\sum_{k=1}^{n} \sigma_k(v_i v_j)\right) = \left(\sum_{k=1}^{n} \sigma_k(v_i)\sigma_k(v_j)\right) \\ &= \big(\sigma_k(v_i)\big)_{ik}\big(\sigma_k(v_j)\big)_{kj} = \big(\sigma_i(v_j)\big)\big(\sigma_i(v_j)\big)^t. \end{aligned}$$

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 $\mathbb{Q}\big(\sqrt{3}\big)$. Calcular la base dual respecto a la traza de la base $\big(1, \sqrt{3}\big)$.

Conviene extraer algunas ideas de la demostración del teorema anterior:

Definición 10.30 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:

$$\Delta[B] = \Delta[v_1, \ldots, v_n] = \big|\mathrm{Tr}(v_i v_j)\big| = \big|\sigma_i(v_j)\big|^2,$$

donde $\sigma_1, \ldots, \sigma_n$ son los k-monomorfismos de K.

Notar 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 10.31 *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

$$M_B = \mathrm{M}_B^C \, M_C \, \left(\mathrm{M}_B^C\right)^t. \tag{10.2}$$

Ahora bien, para todo $v, w \in K$,

$$\Phi_B(v)\, \mathrm{M}_B^C \, M_C \, \left(\mathrm{M}_B^C\right)^t \, \Phi_B(w)^t = \Phi_C(v)\, M_C \, \Phi_C(w)^t = Tr(vw),$$

luego por la unicidad de la matriz de una forma bilineal, se cumple (10.2). ∎

Capítulo XI

Enteros algebraicos

Con la teoría de anillos, la teoría de cuerpos y el álgebra lineal que hemos estudiado, estamos en condiciones de iniciar el estudio de los números desde un punto de vista moderno. La teoría algebraica de números dejó de ser una colección de resultados dispersos cuando, a finales del siglo XIX, Dedekind introdujo los enteros algebraicos, gracias a los cuales los profundos resultados alcanzados años atrás por Gauss, Kummer, Dirichlet, Eisenstein, y muchos más admitían un tratamiento unificado. Puede decirse que la teoría algebraica de números es precisamente el estudio de los anillos de enteros algebraicos.

Si pensamos en los números algebraicos como una generalización de los números racionales, los enteros algebraicos son entonces el paralelo de los números enteros.

11.1 Definición y propiedades básicas

Aunque la teoría de enteros algebraicos se puede desarrollar a un nivel más general, como una teoría de extensiones de anillos similar a la de extensiones de cuerpos, nosotros haremos fuerte uso de que $\mathbb{Z}$ es un dominio euclídeo, por lo que habremos de limitarnos a trabajar con $\mathbb{Z}$ como anillo base y, en consecuencia, con $\mathbb{Q}$ como cuerpo base. Por ello conviene introducir antes que nada el concepto de cuerpo numérico.

Definición 11.1 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.

Consideremos un cuerpo K de característica 0, es decir, un cuerpo que 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 11.2 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 $\mathbb{A}$. 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 11.3 *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 11.4 $\mathcal{O}_{\mathbb{Q}} = \mathbb{Z}$.

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 $q \in Z$. ∎

El lector debe notar el paralelismo entre el teorema siguiente y resultados similares sobre cuerpos.

Teorema 11.5 *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

$$\mathbb{Z}[c] = \langle c^m \mid m = 1, \ldots, n-1 \rangle . \tag{11.1}$$

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 (11.1), 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 11.6 *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 7.30). 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 11.7 *Para cada $c \in \mathbb{A}$ existe un entero racional no nulo m de manera que $mc \in \mathbb{E}$.*

DEMOSTRACIÓN: Sea $\operatorname{pol}\operatorname{mín}(c, \mathbb{Q}) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$. Sea m el producto de los denominadores de todos los coeficientes no nulos de $p(x)$.

Entonces $m^n(c^n + a_{n-1}c^{n-1} + \cdots + a_1c + a_0) = 0$, luego

$$(mc)^n + a_{n-1}m(mc)^{n-1} + \cdots + a_1m^{n-1}(mc) + a_0 = 0.$$

Por lo tanto, $x^n + a_{n-1}mx^{n-1} + \cdots + a_1m^{n-1}x + a_0$ es un polinomio mónico con coeficientes enteros del cual es raíz mc. ∎

Teorema 11.8 *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 11.7 es que los elementos primitivos siempre se pueden tomar enteros:

Teorema 11.9 *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 que definimos al final del capítulo anterior. He aquí un par de hechos adicionales sobre ellos que vamos a necesitar:

Teorema 11.10 *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 σ_1, σ_n y tomamos uno cualquiera de ellos ρ. 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 $\{1,\dots,n\}$ 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. ∎

Teorema 11.11 *Sea K un cuerpo numérico de grado n. Entonces $\mathcal{O}_K$ es un $\mathbb{Z}$-módulo libre de rango n.*

Demostración: Por 11.9 sabemos que $K = \mathbb{Q}(c)$, donde $c \in \mathcal{O}_K$. Entonces $1, c, \dots, c^{n-1}$ es una base de K formada por enteros. Podemos tomar una base de K $B = \{b_1,\dots,b_n\}$ formada por enteros tal que el número natural $\big|\Delta[b_1,\dots,b_n]\big|$ sea mínimo. Vamos a probar que entonces $\{b_1,\dots,b_n\}$ 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 $\{b_1,\dots,b_n\}$. Como en cualquier caso $\{b_1,\dots,b_n\}$ es una base de K, se cumplirá que

$$d = a_1b_1 + \cdots + a_nb_n, \tag{11.2}$$

para ciertos números racionales $a_1, \dots, a_n$ 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 (11.2) obtenemos que

$$rb_1 + a_2b_2 + \cdots + a_nb_n = d - ab_1 \in \mathcal{O}_K.$$

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 $\big|\mathrm{M}_C^B\big| = r$ y en consecuencia

$$\big|\Delta[C]\big| = r^2\big|\Delta[B]\big| < \big|\Delta[B]\big|,$$

en contra de la elección de B. Por lo tanto B es una base de $\mathcal{O}_K$ como $\mathbb{Z}$-módulo. ∎

Definición 11.12 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.

Un ejemplo más concreto: Es obvio que $1, \sqrt{5}$ forman una base de $\mathbb{Q}\big(\sqrt{5}\,\big)$ y sus miembros son sin duda enteros algebraicos, pero para que fueran una base entera haría falta que los enteros algebraicos de $\mathbb{Q}\big(\sqrt{5}\,\big)$ fueran exactamente los elementos de

$$\big\langle 1, \sqrt{5}\big\rangle = \{m + n\sqrt{5} \mid m, n \in \mathbb{Z}\},$$

y esto no es evidente en absoluto ¡como que es falso!

Ejercicio: Probar que $\frac{1+\sqrt{5}}{2}$ es un entero algebraico.

En general, 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 el teorema 10.22 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 11.13 *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.*

11.2 Ejemplos de anillos de enteros algebraicos

Enteros cuadráticos Como primer ejemplo veamos el caso de los *cuerpos 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 $K = \mathbb{Q}\big(\sqrt{b^2 - 4c}\big)$. Podemos expresar $b^2 - 4c = m^2 d$, donde d es libre de cuadrados, y así $K = \mathbb{Q}\big(m\sqrt{d}\big) = \mathbb{Q}\big(\sqrt{d}\big)$.

Tenemos, pues, que todo cuerpo cuadrático es de la forma $K = \mathbb{Q}\big(\sqrt{d}\big)$, donde d es un entero libre de cuadrados (obviamente $d \neq 1$). Como

$$\mathrm{pol\,mín}\big(\sqrt{d}, \mathbb{Q}\big) = x^2 - d = \big(x + \sqrt{d}\big)\big(x - \sqrt{d}\big),$$

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 $\mathbb{Q}(\sqrt{d}\,)$, 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:

$$\Delta\left[1, \sqrt{d}\,\right] = \begin{vmatrix} 1 & 1 \\ \sqrt{d} & -\sqrt{d} \end{vmatrix}^2 = \left(-2\sqrt{d}\,\right)^2 = 4d.$$

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 11.10 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

$$\Delta[1, \alpha] = \begin{vmatrix} 1 & 1 \\ \frac{1+\sqrt{d}}{2} & \frac{1-\sqrt{d}}{2} \end{vmatrix}^2 = \left(-\sqrt{d}\,\right)^2 = d.$$

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 11.14 *Sea d un entero libre de cuadrados y $K = \mathbb{Q}(\sqrt{d}\,)$.*

1. *Si $d \not\equiv 1 \pmod 4$ entonces $\mathcal{O}_K = \mathbb{Z}\left[\sqrt{d}\,\right]$ 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 $\mathbb{Z}\left[\sqrt{-1}\,\right]$ se conoce con el nombre de anillo de los enteros de Gauss. Éste fue el primer anillo de enteros algebraicos estudiado en profundidad. Es costumbre llamar $i = \sqrt{-1}$, tal y como hemos venido haciendo hasta ahora.

Ejercicio: Probar que si d y d' son enteros distintos libres de cuadrados entonces los cuerpos $\mathbb{Q}(\sqrt{d}\,)$ y $\mathbb{Q}(\sqrt{d'}\,)$ no son isomorfos.

En general no es fácil calcular anillos de enteros, al menos no con la poca base teórica que tenemos. Disponemos de la suficiente álgebra para desarrollar

la teoría general sobre anillos de enteros algebraicos, pero enfrentarse con casos concretos (por ejemplo para determinar bases enteras) es mucho más difícil que obtener resultados generales. Puede decirse que entre los anillos de enteros, si $\mathbb{Z}$ es la tierra firme, los cuerpos cuadráticos son la orilla del mar, pero queda un océano entero por explorar sobre el que podemos teorizar en la pizarra, pero obtener imágenes concretas de su fauna requiere expediciones demasiado bien equipadas para nuestras posibilidades. No obstante, curiosear en la orilla es un buen entrenamiento, y por ello los cuerpos cuadráticos van a ser nuestro modelo básico.

No obstante a lo dicho, en esta sección haremos dos excursiones por las profundidades. Con una base mejor los argumentos que vamos a emplear podrían ser sustituidos por otros más conceptuales y más simples.

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 VIII 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.

Con todo esto ya podemos demostrar el teorema siguiente:

Teorema 11.15 *Sea p un número primo impar y $K = \mathbb{Q}(\omega)$, donde ω es una raíz p-ésima primitiva de la unidad. Entonces $\mathcal{O}_K = \mathbb{Z}[\omega]$.*

Demostración: Sea $\alpha = \sum_{i=0}^{p-2} a_i\omega^i$ un entero ciclotómico de orden p. Hemos de probar que todos los coeficientes son enteros racionales. En principio sabemos que la traza es un entero. Más aún, para cada $0 \le k \le p-2$ tenemos que $\mathrm{Tr}(\alpha\omega^{-k}) \in \mathbb{Z}$. Así tenemos la misma información sobre todos los coeficientes:

$$\begin{aligned}\mathrm{Tr}(\alpha\omega^{-k}) &= pa_k - \sum_{i=0}^{p-2} a_i \in \mathbb{Z}, \qquad \text{para } k \neq p-1\\ \mathrm{Tr}(\alpha\omega) &= -\sum_{i=0}^{p-2} a_i \in \mathbb{Z}.\end{aligned}$$

Por lo tanto $pa_k \in \mathbb{Z}$ para todo $k = 0, \dots, p-1$. Llamemos $b_k = pa_k$. Hemos de probar que $p \mid b_k$ para todo k, con lo que los a_k serán también enteros. Consideremos $\pi = 1 - \omega$. Sustituyendo $\omega = 1 - \pi$ y desarrollando obtenemos

$$p\alpha = \sum_{i=1}^{p-2} b_i\omega^i = \sum_{i=1}^{p-2} c_i\pi^i,$$

donde

$$c_i = \sum_{j=i}^{p-2} (-1)^i \binom{j}{i} b_j \in \mathbb{Z},$$

para $i = 0, \ldots, p-2$.

Como $\pi = 1 - \omega$, por simetría se cumple también

$$b_i = \sum_{j=i}^{p-2} (-1)^i \binom{j}{i} c_j,$$

para $i = 0, \ldots, p-2$.

Por lo tanto basta probar que $p \mid c_j$ para todo j, pues entonces estas fórmulas implican que p también divide a los b_i.

Lo probaremos por inducción. Suponemos que $p \mid c_i$ para cada $i \leq k-1$ y vamos a probar que $p \mid c_k$, donde $0 \leq k \leq p-2$.

La razón por la que hemos hecho el cambio de variable es que ω es una unidad de $\mathcal{O}_K$, mientras que π cumple $\mathrm{N}(\pi) = p$ (pronto veremos que esto implica que π es primo en $\mathcal{O}_K$). Tenemos que

$$p = \mathrm{N}(1-\omega) = \prod_{i=1}^{p-1} (1-\omega^i) = (1-\omega)^{p-1} \prod_{i=1}^{p-1} (1 + \omega + \cdots + \omega^{i-1}) = \pi^{p-1}\delta,$$

para cierto $\delta \in \mathcal{O}_K$.

Esto implica que $p \equiv 0 \pmod{\pi^{k+1}}$, es decir, módulo el ideal generado por π^{k+1} en $\mathcal{O}_K$.

Por otro lado,

$$0 \equiv p\alpha = \sum_{i=0}^{p-2} c_i \pi^i \equiv c_k \pi^k \pmod{\pi^{k+1}},$$

pues los términos anteriores a $c_k\pi^k$ son múltiplos de p por hipótesis de inducción y los posteriores son múltiplos de π^{k+1} directamente.

Esto equivale a que $c_k\pi^k = \eta\pi^{k+1}$ para un cierto $\eta \in \mathcal{O}_K$, luego $c_k = \eta\pi$.

Finalmente tomamos normas: $c_k^{p-1} = \mathrm{N}(c_k) = \mathrm{N}(\eta)\,\mathrm{N}(\pi) = p\,\mathrm{N}(\eta)$, luego en efecto $p \mid c_k$. ∎

Para calcular el discriminante de las extensiones ciclotómicas nos basaremos en el siguiente resultado general.

Teorema 11.16 *Sea $K = \mathbb{Q}(a)$ un cuerpo numérico de grado n y llamemos $p(x) = \operatorname{pol\,mín}(a, \mathbb{Q})$. 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 al final del capítulo anterior,

$$\Delta[1, a, \dots, a^{n-1}] = \prod_{1\le i<j\le n} \big(\sigma_j(a) - \sigma_i(a)\big)^2, \tag{11.3}$$

donde $\sigma_1(a)$, ... , $\sigma_n(a)$ 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

$$p'\big(\sigma_j(a)\big) = \prod_{\substack{i=1\\ i\ne j}}^n \big(\sigma_j(a) - \sigma_i(a)\big)$$

para $j = 1, \dots, n$.

Multiplicando todas estas ecuaciones obtenemos

$$\mathrm{N}\big(p'(a)\big) = \prod_{j=1}^n \sigma_j\big(p'(a)\big) = \prod_{j=1}^n p'\big(\sigma_j(a)\big) = \prod_{\substack{i,j=1\\ i\ne j}}^n \big(\sigma_j(a) - \sigma_i(a)\big).$$

Agrupamos los pares $\big(\sigma_j(a) - \sigma_i(a)\big)\big(\sigma_(a) - \sigma_j(a)\big) = -\big(\sigma_j(a) - \sigma_i(a)\big)^2$. El número de factores (-1) que aparecen es $n(n-1)/2$, luego teniendo en cuenta (11.3) queda

$$\mathrm{N}\big(p'(a)\big) = (-1)^{n(n-1)/2}\Delta[1, a, \dots, a^{n-1}],$$

y de aquí se sigue el teorema. ∎

Como caso particular obtenemos:

Teorema 11.17 *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 11.16 sobre los cuerpos cuadráticos.

El anillo de enteros de $\mathbb{Q}(\sqrt[3]{2})$. Vamos a probar que si $K = \mathbb{Q}(\sqrt[3]{2})$, entonces $\mathcal{O}_K = \mathbb{Z}[\sqrt[3]{2}]$.

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 $\operatorname{pol\,mín}(\eta, \mathbb{Q})$ tiene grado 3 y sus raíces son las imágenes de η por los tres monomorfismos de K, o sea,

$$\operatorname{pol\,mín}(\eta, \mathbb{Q}) = (x - \frac{u}{d} - \frac{v}{d}\alpha - \frac{w}{d}\alpha^2)(x - \frac{u}{d} - \frac{v}{d}\beta - \frac{w}{d}\beta^2)(x - \frac{u}{d} - \frac{v}{d}\gamma - \frac{w}{d}\gamma^2).$$

Operando (con bastante paciencia) se llega a

$$\operatorname{pol\,mín}(\eta, \mathbb{Q}) = x^3 - \frac{3u}{d}x^2 + \frac{3u^2 - 6vw}{d^2}x - \frac{u^3 + 2v^3 + 4w^3 - 6uvw}{d^3}.$$

Por lo tanto η será entero si y sólo si

$$d \mid 3u \tag{11.4}$$
$$d^2 \mid 3u^2 - 6vw \tag{11.5}$$
$$d^3 \mid u^3 + 2v^3 + 4w^3 - 6uvw \tag{11.6}$$

Si existe un primo p tal que $p \mid d$ y $p \mid u$, entonces por (11.5) $p^2 \mid 6vw$, lo que implica que $p \mid v$ o $p \mid w$.

Si $p \mid v$ por (11.6) $p^3 \mid 4w^3 - 6uvw$, 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 (11.4) implica que $d = 1$ o $d = 3$. Basta probar que d no puede valer 3 y entonces η estará en $\mathbb{Z}[\sqrt[3]{2}]$.

Si $d = 3$, tenemos que $(3, u) = 1$, y de (11.5) se sigue que $(3, v) = (3, w) = 1$.

Tomando clases módulo 3 en (11.6) 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 (11.6) debería ser 0.

Con la segunda posibilidad llegamos a $[-9]$, luego concluimos que $d = 3$ es imposible. ∎

Ahora el teorema 11.16 nos permite calcular el discriminante de la extensión: La derivada del polinomio mínimo de α es $3x^2$, luego

$$\Delta_K = (-1)^{3(3-1)/2}\,\mathrm{N}(3\alpha^2) = -\,\mathrm{N}(3)\,\mathrm{N}(\alpha)^2 = -108.$$

11.3 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. No obstante, en el próximo capítulo veremos que la divisibilidad en estos anillos sigue unas leyes generales de las cuales la factorización única de $\mathbb{Z}$ es un caso particular. Una de las piezas clave en nuestro estudio será el hecho de que en los anillos $\mathcal{O}_K$ hay definida una norma $\mathrm{N} : \mathcal{O}_K \longrightarrow \mathbb{Z}$ que conserva los productos. Conviene recoger en un primer teorema sus propiedades principales:

Teorema 11.18 *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 $N(a) = \pm 1$ 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?

Tenemos caracterizadas las unidades de los cuerpos numéricos como los enteros de norma ± 1. Obtener descripciones explícitas de los grupos de unidades es, junto con la determinación de los anillos de enteros, una de las cuestiones más difíciles a la hora de estudiar ejemplos concretos de cuerpos numéricos, y resulta imprescindible al abordar muchos problemas concretos. Así, por ejemplo, los resultados de Kummer sobre el Último Teorema de Fermat dependen fuertemente de un análisis de las unidades ciclotómicas. En este libro estudiaremos únicamente las unidades de los cuerpos cuadráticos. Un teorema de Dirichlet afirma que el grupo de unidades de un cuerpo numérico es infinito salvo en el caso de los cuerpos cuadráticos de discriminante negativo. Para estos cuerpos es fácil mostrar explícitamente las unidades. Lo hacemos en el teorema siguiente. El caso de los cuerpos cuadráticos de discriminante positivo es mucho más complicado, y no estaremos en condiciones de abordarlo hasta el capítulo XIII.

Teorema 11.19 *Sea d un número entero negativo y libre de cuadrados. Entonces el grupo U de las unidades de $\mathbb{Q}\big(\sqrt{d}\,\big)$ es el siguiente:*

Para $d = -1$, $U = \big\{\pm 1, \pm\sqrt{-1}\big\}$.

Para $d = -3$, $U = \{\pm 1, \pm\omega, \pm\omega^2\}$, donde $\omega = \frac{-1+\sqrt{-3}}{2}$.

Para $d < -3$, $U = \{\pm 1\}$.

Demostración: La norma en $\mathbb{Q}\big(\sqrt{d}\,\big)$ viene dada por $\mathrm{N}(a+b\sqrt{d}) = a^2 - db^2$, luego si $d < 0$ tenemos que la norma es siempre positiva. Un entero $a + b\sqrt{d}$ es una unidad si y sólo si se cumple la ecuación $a^2 - db^2 = 1$.

Para $d = -1$ ésta se reduce a $a^2 + b^2 = 1$, y además los enteros de $\mathbb{Q}\big(\sqrt{-1}\,\big)$ son los elementos de $\mathbb{Z}\big[\sqrt{-1}\,\big]$, luego a y b han de ser enteros racionales. Es claro que las únicas soluciones enteras de $a^2 + b^2 = 1$ son $(1,0)$, $(0,1)$, $(-1,0)$ y $(0,-1)$, de donde U es el grupo indicado.

Para $d = -3$ observamos que $\mathbb{Q}\big(\sqrt{-3}\,\big)$ es el cuerpo ciclotómico de orden 3. En efecto, como ya hemos observado en otras ocasiones, una raíz del tercer polinomio ciclotómico, x^2+x+1, es el número ω del enunciado, luego $\mathbb{Q}\big(\sqrt{-3}\,\big) = \mathbb{Q}(\omega)$.

La norma de un elemento del anillo de enteros $\mathbb{Z}[\omega]$ es

$$\begin{aligned}\mathrm{N}(a+b\omega) &= (a+b\omega)(a+b\omega^2) = a^2 + ab\omega^2 + ab\omega + b^2\omega^3 = a^2 - ab + b^2\\ &= \frac{1}{4}\big((2a)^2 - 2(2a)b + b^2 + 3b^2\big) = \frac{1}{4}\big((2a-b)^2 + 3b^2\big).\end{aligned}$$

El elemento $a + b\omega$ será una unidad si y sólo si $\frac{1}{4}\big((2a-b)^2 + 3b^2\big) = \pm 1$, o sea, si y sólo si $(2a-b)^2 + 3b^2 = 4$.

Si $b = 0$ ha de ser $2a = \pm 2$, luego $a = \pm 1$.

Si $b = \pm 1$ ha de ser $2a - b = \pm 1$ y es fácil ver que entonces (a, b) ha de tomar los valores $(0, 1)$, $(0, -1)$, $(1, 1)$, $(-1, -1)$.

En total quedan las soluciones ± 1, $\pm\omega$ y $\pm 1 \pm \omega$, o de otra forma, ± 1, $\pm\omega$ y $\pm\omega^2$.

Así pues, las unidades de $\mathbb{Z}[\omega]$ son $\{\pm 1, \pm\omega, \pm\omega^2\}$.

Si $d < -3$ hemos de tener presente que a y b pueden ser enteros o semienteros, es decir, $a = A/2$, $b = B/2$, con A, $B \in \mathbb{Z}$. Como el caso semientero incluye al caso entero, basta probar que las únicas soluciones semienteras de $a^2 - db^2 = 1$ son $(1, 0)$ y $(-1, 0)$.

En efecto, tenemos $A^2 - dB^2 = 4$, pero si $B \neq 0$, entonces $A^2 - dB^2 > 4$, pues en realidad $d \leq -5$, luego ha de ser $B = 0$ y $A^2 = 4$, o sea, $a = \pm 1$, $b = 0$.

El caso restante $d = -2$ se razona igualmente. ∎

Ocupémonos ahora de los ideales de los anillos de enteros. Los resultados básicos son los siguientes:

Teorema 11.20 *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 7.30), 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 $\big\{[v_1], \ldots, [v_r]\big\}$ 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 $n \in P$. 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 5.3) y el teorema 5.5 nos da que $\mathcal{O}/P$ es un cuerpo, luego el ideal P es maximal (por 5.4). ∎

El apartado 2) del teorema anterior puede precisarse más. El teorema siguiente muestra que la conexión de la norma con la divisibilidad es más fuerte de lo que podría pensarse en un principio.

Teorema 11.21 *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 10.23 tenemos que

$$|\mathcal{O}/a\mathcal{O}| = |\det C| = |\,\mathrm{N}(a)|.$$

■

Ejercicio: Usar el cociente $\mathbb{Z}[i]/(3)$ para construir las tablas de la suma y el producto de un cuerpo de 9 elementos.

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 4.24 es una condición necesaria para que puedan tener factorización única. En primer lugar la demostramos para el anillo de todos los enteros algebraicos:

Teorema 11.22 *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 11.5 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 $1 \leq i \leq r$, $0 \leq k \leq n-1$.

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 11.5 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 11.23 *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. Es el caso, por ejemplo, de $\mathbb{Z}[\sqrt{5}]$.

11.4 Factorización única en cuerpos cuadráticos

En la sección anterior hemos podido comprobar que hay una gran similitud entre los anillos de enteros algebraicos y el anillo $\mathbb{Z}$, sin embargo esta similitud no llega hasta garantizar la factorización única. En efecto, sabemos que los anillos de enteros comparten propiedades algebraicas con los DIP's y los DFU's, como que son noetherianos, los ideales maximales coinciden con los primos, etc., pero no es cierto que todos ellos sean DIP's o DFU's (como ya hemos comprobado en algunos ejemplos). El hecho de que sean noetherianos garantiza que todo elemento de un anillo de enteros (no nulo ni unidad) se descompone en producto de irreducibles. También sabemos que la factorización en irreducibles será única (salvo orden o asociación) si y sólo si los elementos irreducibles coinciden con los primos. Sin embargo, esto no siempre es cierto. Aquí vamos a estudiar la situación en los cuerpos cuadráticos. Antes de entrar en detalles, conviene notar que los cuerpos cuadráticos se dividen en dos familias muy diferentes:

Definición 11.24 Un cuerpo cuadrático es *real* o *imaginario* según si su discriminante es positivo o negativo.

Lo que hace diferentes a los cuerpos reales de los imaginarios es en esencia que la norma de los cuerpos imaginarios es siempre positiva, mientras que en los cuerpos reales hay elementos de norma positiva y negativa. Más exactamente, puesto que $\mathrm{N}\big(a + b\sqrt{d}\,\big) = a^2 - db^2$, cuando $d < 0$ hay sólo un número finito de valores de a, b que pueden hacer que la norma tome un valor dado, mientras que si $d > 0$ no tenemos ninguna cota. Por ello los cuerpos imaginarios se comportan mucho mejor que los cuerpos reales. Por ello el argumento con el que hemos determinado las unidades de los cuerpos cuadráticos imaginarios no sirve para los cuerpos reales. En general los cuerpos cuadráticos imaginarios son los cuerpos numéricos más sencillos.

Tabla 11.1: Factorizaciones no únicas en cuerpos cuadráticos imaginarios

d					
-5	6	$=$	$2\cdot 3$	$=$	$\left(1+\sqrt{-5}\right)\left(1-\sqrt{-5}\right)$
-6	6	$=$	$2\cdot 3$	$=$	$\sqrt{-6}\left(-\sqrt{-6}\right)$
-10	14	$=$	$2\cdot 7$	$=$	$\left(2+\sqrt{-10}\right)\left(2-\sqrt{-10}\right)$
-13	14	$=$	$2\cdot 7$	$=$	$\left(1+\sqrt{-13}\right)\left(1-\sqrt{-13}\right)$
-14	15	$=$	$3\cdot 5$	$=$	$\left(1+\sqrt{-14}\right)\left(1-\sqrt{-14}\right)$
-15	4	$=$	$2\cdot 2$	$=$	$\left(\frac{1+\sqrt{-15}}{2}\right)\left(\frac{1-\sqrt{-15}}{2}\right)$
-17	18	$=$	$2\cdot 3\cdot 3$	$=$	$\left(1+\sqrt{-17}\right)\left(1-\sqrt{-17}\right)$
-21	22	$=$	$2\cdot 11$	$=$	$\left(1+\sqrt{-21}\right)\left(1-\sqrt{-21}\right)$
-22	26	$=$	$2\cdot 13$	$=$	$\left(1+\sqrt{-22}\right)\left(1-\sqrt{-22}\right)$
-23	6	$=$	$2\cdot 3$	$=$	$\left(\frac{1+\sqrt{-23}}{2}\right)\left(\frac{1-\sqrt{-23}}{2}\right)$
-26	27	$=$	$3\cdot 3\cdot 3$	$=$	$\left(1+\sqrt{-26}\right)\left(1-\sqrt{-26}\right)$
-29	30	$=$	$2\cdot 3\cdot 5$	$=$	$\left(1+\sqrt{-29}\right)\left(1-\sqrt{-29}\right)$
-30	34	$=$	$2\cdot 17$	$=$	$\left(2+\sqrt{-30}\right)\left(2-\sqrt{-30}\right)$

Comencemos, pues, estudiando la unicidad de las factorizaciones en cuerpos imaginarios. A la hora de encontrar factorizaciones no únicas es preciso asegurarse de que los factores considerados son irreducibles, lo cual no es obvio.

Por ejemplo, podría pensarse que las factorizaciones en $\mathbb{Q}(i)$

$$10 = 2\cdot 5 = (3+i)(3-i)$$

prueban que $\mathbb{Q}(i)$ no es un dominio de factorización única, pero no es así. La razón es que ninguno de los factores es irreducible.

Se cumple que $2=(1+i)(1-i)$ (no es difícil descubrirlo sin más que pensar en $a^2+b^2=2$). Como $\mathrm{N}(1+i)=2$ es primo, $1+i$ sí es irreducible, al igual que $1-i$.

Del mismo modo se concluye que $5=(1+2i)(1-2i)$ es una descomposición en irreducibles. Por lo tanto la factorización de 10 en irreducibles es

$$10 = (1+i)(1-i)(1+2i)(1-2i).$$

La otra factorización se obtiene agrupando los factores en otro orden:

$$\begin{aligned}(1+i)(1-2i) &= 3-i\\ (1-i)(1+2i) &= 3+i\end{aligned}$$

Hay otra razón por la que una aparente factorización doble puede no serlo. Por ejemplo, tenemos que $5 = (1+2i)(1-2i) = (2+i)(2-i)$. Esto se explica porque los factores son asociados. En efecto:

$$\begin{aligned} 1+2i &= (-i)(2-i) \\ 1-2i &= (-i)(2+i) \end{aligned}$$

Con estas salvedades podemos buscar auténticos ejemplos de factorizaciones dobles. Aunque este ejemplo no servía, nos da una idea provechosa, la de buscar números irreducibles cuya norma no sea un primo.

La tabla 11.1 nos muestra que las factorizaciones no únicas son algo bastante frecuente. Vamos a comprobar el primer ejemplo. En $\mathbb{Q}\big(\sqrt{-5}\,\big)$ tenemos que $\mathrm{N}(2) = 4$, $\mathrm{N}(3) = 9$, $\mathrm{N}\big(1+\sqrt{-5}\,\big) = \mathrm{N}\big(1-\sqrt{-5}\,\big) = 6$.

Como $-5 \not\equiv 1 \pmod 4$, el anillo de enteros de $\mathbb{Q}\big(\sqrt{-5}\,\big)$ es $\mathbb{Z}\big[\sqrt{-5}\,\big]$.

Si $2 = xy$ fuera una descomposición no trivial de 2, es decir, donde x e y son no unidades, entonces $\mathrm{N}(x) = \mathrm{N}(y) = 2$.

Si $x = a+b\sqrt{-5}$, entonces queda $a^2+5b^2 = 2$. Si $b \neq 0$ entonces $a^2+5b^2 \geq 5$, luego ha de ser $b = 0$, pero entonces $a^2 = 2$, lo que es imposible para $a \in \mathbb{Z}$. Así pues 2 es irreducible.

Igualmente se prueba que no existen elementos de norma 3, lo que nos da la irreducibilidad de los cuatro factores. Por último, como

$$\mathrm{N}\big(1+\sqrt{-5}\,\big) \neq \mathrm{N}(2), \mathrm{N}(3),$$

podemos asegurar que $1+\sqrt{-5}$ no es asociado ni de 2 ni de 3, luego las factorizaciones son distintas en sentido estricto.

Las comprobaciones restantes son un interesante ejercicio para el lector. Es de destacar que incluso hay factorizaciones con distinto número de factores en cada miembro.

En los cuerpos cuadráticos reales el trabajo se complica debido a que las normas pueden ser negativas, lo que impide descartar tan fácilmente casos como $\mathrm{N}(x) = 2$ en $\mathbb{Q}\big(\sqrt{-5}\,\big)$. Pese a ello, la tabla 11.2 contiene algunos ejemplos que podemos comprobar.

Tabla 11.2: Factorizaciones no únicas en cuerpos cuadráticos reales

d					
10	6	$=$	$2\cdot 3$	$=$	$\big(4+\sqrt{10}\,\big)\big(4-\sqrt{10}\,\big)$
15	10	$=$	$2\cdot 5$	$=$	$\big(5+\sqrt{15}\,\big)\big(5-\sqrt{15}\,\big)$
26	10	$=$	$2\cdot 5$	$=$	$\big(6+\sqrt{26}\,\big)\big(6-\sqrt{26}\,\big)$
30	6	$=$	$2\cdot 3$	$=$	$\big(6+\sqrt{30}\,\big)\big(6-\sqrt{30}\,\big)$

Por ejemplo, en $\mathbb{Q}(\sqrt{10})$ tenemos $\mathrm{N}(2) = 4$, $\mathrm{N}(3) = 9$, $\mathrm{N}(4+\sqrt{10}) = \mathrm{N}(4-\sqrt{10}) = 6$. Basta probar que no hay enteros de norma igual a ± 2 ni a ± 3.

En primer lugar, el anillo de enteros es $\mathbb{Z}[\sqrt{10}]$. Si existiera un entero $x = a + b\sqrt{10}$ tal que $\mathrm{N}(a + b\sqrt{10}) = a^2 - 10b^2 = \pm 2$ o ± 3, entonces tendría que ser $a^2 \equiv \pm 2$ o ± 3 (mód 10), o lo que es lo mismo, $a^2 \equiv 2, 3, 7, 8$ (mód 10).

Sin embargo, los cuadrados módulo 10 son: $0, 1, 4, 9, 6, 5, 6, 9, 4, 1$, luego la congruencia es imposible.

Los otros casos se prueban de modo similar.

Los ejemplos anteriores no deben llevar al lector a pensar que la factorización única no se da nunca o casi nunca en cuerpos numéricos. Vamos a ver ahora algunos ejemplos de factorización única. En general es difícil probar que un cuerpo numérico es un DFU, salvo en un caso: cuando se trata de un dominio euclídeo. Como siempre, los casos más simples son los cuerpos cuadráticos imaginarios.

Teorema 11.25 *Sea d un número negativo libre de cuadrados. El anillo de los enteros de $\mathbb{Q}(\sqrt{d})$ es un dominio euclídeo si y sólo si d es uno de los cinco números siguientes:*

$$-1, \quad -2, \quad -3, \quad -7, \quad -11$$

En estos casos, la norma euclídea es $\phi(a) = \mathrm{N}(a)$.

Demostración: En primer lugar veamos que los anillos indicados son en efecto dominios euclídeos.

Es inmediato que la función N cumple los requisitos para ser una norma euclídea. Sólo hay que probar que la división euclídea es realizable.

Veamos que para ello es suficiente probar que para todo $\alpha \in \mathbb{Q}(\sqrt{d})$ existe un entero γ tal que $\mathrm{N}(\alpha - \gamma) < 1$.

En efecto, si se cumple esta condición, sean Δ y δ dos enteros con $\delta \neq 0$. Consideremos $\alpha = \Delta/\delta$. Por hipótesis existe un entero γ tal que $\mathrm{N}(\Delta/\delta - \gamma) < 1$, es decir, $\mathrm{N}\big((\Delta - \delta\gamma)/\delta\big) < 1$, luego $\mathrm{N}(\Delta - \delta\gamma) < \mathrm{N}(\delta)$. Sea $\epsilon = \Delta - \delta\gamma$. De este modo hemos obtenido que $\Delta = \delta\gamma + \epsilon$, con $\mathrm{N}(\epsilon) < \mathrm{N}(\delta)$.

Sea, pues, $\alpha = r + s\sqrt{d} \in \mathbb{Q}(\sqrt{d})$. Para $d \not\equiv 1$ (mód 4) tenemos que encontrar un elemento $\gamma = x + y\sqrt{d}$ tal que x, $y \in \mathbb{Z}$ y $(r-x)^2 - d(s-y)^2 < 1$.

En los casos $d = -1$ y $d = -2$ basta tomar como x e y los enteros racionales más cercanos a los números racionales r y s, pues entonces $|r - x|$, $|s - y| \leq 1/2$, con lo que $(r-x)^2 - d(s-y)^2 \leq 1/4 + 2(1/4) = 3/4 < 1$.

Los restantes valores de d son congruentes con 1 módulo 4, luego buscamos un elemento de la forma $x + y\frac{1+\sqrt{d}}{2}$, donde x, y son enteros racionales y de modo que $\big(r - x - (1/2)y\big)^2 - d\big(s - (1/2)y\big)^2 < 1$.

Si tomamos como y el entero racional más cercano a $2s$, tenemos entonces que $|2s - y| \leq 1/2$, luego $\big(s - (1/2)y\big)^2 \leq 1/16$ y $-d\big(s - (1/2)y\big)^2 \leq 11/16$. Nos falta elegir x de modo que $\big(r - x - (1/2)y\big)^2 < 5/16$, para lo cual es suficiente

que $\big|r - x - (1/2)y\big| \leq 1/2$, es decir, tomamos como x el entero más cercano a $r - (1/2)y$.

Respecto a los valores de d no contemplados en el enunciado, son -5, -6 y los enteros libres de cuadrados menores que -11. Hemos visto que $\mathbb{Q}\big(\sqrt{-5}\,\big)$ y $\mathbb{Q}\big(\sqrt{-6}\,\big)$ no son dominios de factorización única, luego no pueden ser euclídeos. Supongamos ahora que $d < -11$ es libre de cuadrados, luego de hecho $d \leq -13$. Sea $\mathcal{O}$ el anillo de enteros. Si $\mathcal{O}$ fuera euclídeo podríamos tomar un $\delta \in \mathcal{O}$ de norma euclídea mínima entre los enteros no nulos ni unitarios.

Entonces todo $\Delta \in \mathcal{O}$ se expresa como $\Delta = \delta c + r$, donde $r = 0, 1, -1$, por la elección de δ. Esto significa que $\mathcal{O}/(\delta) = \big\{[0], [1], [-1]\big\}$, luego por el teorema 11.21 ha de ser $\mathrm{N}(\delta) \leq 3$.

Sea $\delta = (a/2) + (b/2)\sqrt{d}$, donde a y b son enteros. Tenemos que $a^2 - db^2 \leq 12$, y como $d \leq -13$, necesariamente $b = 0$ y $|a| \leq 3$, pero entonces $\delta = a/2$ es entero y no puede ser más que $\delta = 0, 1, -1$, en contra de cómo ha sido elegido. ∎

Es importante notar que la prueba del teorema anterior nos da un criterio práctico para calcular divisiones euclídeas en los cinco cuerpos a los que se aplica. Este criterio es especialmente simple en los casos $d = -1$ y $d = -2$, donde dados dos enteros Δ y δ, el cociente γ se obtiene como el entero cuyas coordenadas están más próximas a las de Δ/δ, y el resto es simplemente la diferencia $\Delta - \delta\gamma$.

Por otra parte, los cinco cuerpos anteriores no son los únicos en los que la factorización es única. No obstante no hay muchos más. Puede probarse que los únicos cuerpos cuadráticos con discriminante negativo que son dominios de factorización única son los correspondientes a los nueve valores siguientes de d:

Tabla 11.3: Cuerpos cuadráticos imaginarios con factorización única

$$d = -1, \quad -2, \quad -3, \quad -7, \quad -11, \quad -19, \quad -43, \quad -67 \quad -163.$$

En el capítulo XIII demostraremos que estos anillos son DFU's. Probar que son los únicos excede nuestras posibilidades.

El caso de los cuerpos cuadráticos reales es todavía más complicado. No se sabe si hay infinitos de ellos que sean dominios de factorización única. Los valores de d menores que 100 para los que la factorización es única son los siguientes:

Tabla 11.4: Primeros cuerpos cuadráticos reales con factorización única

$$d = 2,\ 3,\ 5,\ 6,\ 7,\ 11,\ 13,\ 14,\ 17,\ 19,\ 21,\ 22,\ 23,\ 29,\ 31,\ 33,\ 37,\ 38,\ 41,$$
$$43,\ 46,\ 47,\ 53,\ 57,\ 59,\ 61,\ 62,\ 67,\ 69,\ 71,\ 73,\ 77,\ 83,\ 86,\ 89,\ 93,\ 94,\ 97.$$

Tampoco se sabe si entre ellos hay un número finito de dominios euclídeos. Al menos se sabe que sólo un número finito de ellos son euclídeos con norma $\phi(x) = \big|\mathrm{N}(x)\big|$, pero se desconoce si puede haber cuerpos cuadráticos que sean dominios euclídeos con otra norma.

Los cuerpos cuadráticos euclídeos con la norma $\left|\mathrm{N}\right|$ son exactamente los correspondientes a los siguientes valores de d:

Tabla 11.5: Cuerpos cuadráticos reales euclídeos

$$2,\ 3,\ 5,\ 6,\ 7,\ 11,\ 13,\ 17,\ 19,\ 21,\ 29,\ 33,\ 37,\ 38,\ 41,\ 57,\ 73$$

Aunque este resultado también escapa a las posibilidades de este libro, vamos a probar una porción.

Teorema 11.26 *Los cuerpos cuadráticos correspondientes a*

$$d \ = \ 2,\ 3,\ 5,\ 6,\ 7,\ 13,\ 17,\ 21,\ 29$$

son dominios euclídeos con el valor absoluto de la norma.

Demostración: El planteamiento es el mismo que en el teorema 11.25. Para agrupar los casos en los que d es congruente con 1 módulo 4 y los casos en los que no lo es, definimos

$$\begin{array}{lll} \lambda = 0, & E = d & \text{si } d \not\equiv 1 \ (\text{mód } 4) \\ \lambda = 1/2, & E = (1/4)d & \text{si } d \equiv 1 \ (\text{mód } 4). \end{array}$$

La condición que hay que probar es entonces que para todo par de números racionales r y s existen enteros x e y tales que

$$\left|(r - x - \lambda y)^2 - E(s - y)^2\right| < 1. \tag{11.7}$$

Por reducción al absurdo vamos a suponer que existen r y s para los que (11.7) no se cumple cualesquiera que sean x e y.

En primer lugar probamos que podemos suponer que $0 \le r, s \le 1/2$.

Si $d \not\equiv 1$ (mód 4), podemos sustituir en (11.7) r, x, s, y por

$$\epsilon_1 r + u, \quad \epsilon_1 x + u, \quad \epsilon_2 s + v, \quad \epsilon_2 y + v,$$

donde ϵ_1, $\epsilon_2 = \pm 1$ y u, $v \in \mathbb{Z}$, sin que la expresión varíe, y con esto podemos reducir r y s al intervalo deseado.

Si $d \equiv 1$ (mód 4) la expresión (11.7) resulta inalterada si reemplazamos r, x, s, y por una de estas alternativas:

$$\begin{array}{llll} \epsilon_1 r + u & \epsilon_1 x + u & \epsilon_1 s & \epsilon_1 y \qquad (11.8) \\ r & x - v & s + 2v & y + 2v \qquad (11.9) \\ r & x + y & -s & -y \qquad (11.10) \\ 1/2 - r & -x & 1 - s & 1 - y \qquad (11.11) \end{array}$$

Mediante (11.8) podemos hacer $0 \le r \le 1/2$. mediante (11.9) hacemos $-1 \le s \le 1$. Si es necesario, (11.10) nos da $0 \le s \le 1$ y si s no es menor o igual que $1/2$, entonces (11.11) hace que lo sea.

Dados, pues, r y s entre 0 y 1/2 que no cumplan (11.7) para ningún par de enteros x e y, para cada uno de estos pares se ha de cumplir una de las dos afirmaciones siguientes:

$$\begin{array}{lrcl} P(x,y): & (r-x-\lambda y)^2 & \geq & 1+E(s-y)^2 \\ Q(x,y): & E(s-y)^2 & \geq & 1+(r-x-\lambda y)^2 \end{array}$$

Vamos a analizar tres casos particulares:

$$\begin{array}{llcllll} P(0,0) & r^2 & \geq & 1+Es^2 & Q(0,0) & Es^2 & \geq & 1+r^2 \\ P(1,0) & (1-r)^2 & \geq & 1+Es^2 & Q(1,0) & Es^2 & \geq & 1+(1-r)^2 \\ P(-1,0) & (1+r)^2 & \geq & 1+Es^2 & Q(-1,0) & Es^2 & \geq & 1+(1+r)^2 \end{array}$$

Como $0 \leq r \leq 1/2$ y $E > 0$, tenemos que $P(0,0)$ y $P(1,0)$ no pueden darse, luego se cumplen $Q(0,0)$ y $Q(1,0)$.

Si se cumple $P(-1,0)$, entonces, usando $Q(1,0)$ tenemos

$$(1+r)^2 \geq 1+Es^2 \geq 2+(1-r)^2, \tag{11.12}$$

lo que operando lleva a que $4r \geq 2$, luego $r = 1/2$. Sustituyendo en (11.12) queda $Es^2 = 5/4$.

Ahora bien, dando a E todos los valores permitidos por el enunciado del teorema, en ningún caso de obtiene un valor para s que tenga raíz cuadrada racional.

Por lo tanto se cumple $Q(-1,0)$, es decir, $Es^2 \geq 1+(1+r)^2 \geq 2$ y así $E \geq 8$. Pero sucede que $E < 8$ en todos los casos que estamos contemplando. ∎

11.5 Aplicaciones de la factorización única

Terminamos el capítulo con algunas aplicaciones a los números enteros. El siguiente resultado fue planteado por Fermat a sus contemporáneos.

Teorema *Las únicas soluciones enteras de* $y^2+2 = x^3$ *son* $(x,y) = (3, \pm 5)$.

Demostración: En primer lugar, y ha de ser impar, pues si fuera par también lo sería x, y llegaríamos a que $2 \equiv 0 \pmod 4$. Si x e y cumplen la ecuación, entonces

$$\big(y+\sqrt{-2}\,\big)\big(y-\sqrt{-2}\,\big) = x^3.$$

Un divisor común α de $y+\sqrt{-2}$ y de $y-\sqrt{-2}$ en $\mathbb{Z}\big[\sqrt{-2}\,\big]$ cumpliría que $\mathrm{N}(\alpha) \mid \mathrm{N}(y+\sqrt{-2}) = x^3$ y, como también dividiría a la diferencia, $2\sqrt{-2}$, también $\mathrm{N}(\alpha) \mid 8$, luego $\mathrm{N}(\alpha) = 1$ (es impar y potencia de 2).

Así pues, $y+\sqrt{-2}$, $y-\sqrt{-2}$ son primos entre sí. Teniendo en cuenta la factorización única de $\mathbb{Z}\big[\sqrt{-2}\,\big]$ así como que las unidades en este anillo son $\{\pm 1\}$ y ambas son cubos, resulta que si dos números primos entre sí son un cubo, entonces cada factor lo es, es decir, $y+\sqrt{-2} = \big(a+b\sqrt{-2}\,\big)^3$ para ciertos enteros a y b.

Igualando los coeficientes de obtenemos que $1 = b(3a^2 - 2b^2)$, lo que sólo es posible si $b = 1$ y $a = \pm 1$, de donde $y = \pm 5$ y por lo tanto $x = 3$. ∎

Ejercicio: Si el producto de dos enteros de Gauss primos entre sí es una potencia cuarta ¿lo es necesariamente cada factor? ¿Es cierto con cubos?

La variante que veremos a continuación, también propuesta por Fermat, resulta un poco más complicada.

Teorema Las únicas soluciones enteras de la ecuación $y^2 + 4 = x^3$ son
$$(x, y) = (5, \pm 11),\ (2, \pm 2).$$

Demostración: Supongamos primero que y es impar. Tenemos que

$$(2 + iy)(2 - iy) = x^3.$$

Un divisor común α de $2 + iy$, $2 - iy$ lo es también de su suma, 4, luego $\mathrm{N}(\alpha) \mid 16$ y $\mathrm{N}(\alpha) \mid x^3$, impar, lo que implica que $\alpha = 1$ y los dos factores $2 + iy$, $2 - iy$ son primos entre sí. Por consiguiente ambos son cubos, es decir, existen enteros racionales a y b tales que $2+iy = (a+ib)^3$. Conjugando, $2-iy = (a-ib)^3$ y sumando las dos ecuaciones y simplificando se llega a que $4 = 2a(a^2 - 3b^2)$, luego tenemos que $a(a^2 - 3b^2) = 2$.

Claramente entonces $a = \pm 1, \pm 2$, pero las únicas posibilidades que permiten la existencia de b son $a = -1$, $b = \pm 1$ y $a = 2$, $b = \pm 1$. Entonces

$$x^3 = \big((a + ib)(a - ib)\big)^3 = (a^2 + b^2)^3$$

y los valores de x que se obtienen son $x = 2, 5$. De $y^2 + 4 = 8, 125$, obtenemos que $y = \pm 2, \pm 11$.

Ahora queda el caso en que y es par. Digamos $y = 2Y$. Entonces x ha de ser par también, $x = 2X$ y se cumple $Y^2 + 1 = 2X^3$. De aquí se sigue que Y es impar. Factorizamos

$$(1 + iY)(1 - iY) = 2X^3. \tag{11.13}$$

El máximo común divisor de $1 + iY$ y $1 - iY$ divide a la suma $2 = -i(1 + i)^2$. Como $1 + i$ es primo (tiene norma 2), dicho mcd ha de ser 1, $1 + i$, o bien 2.

Claramente 2 no divide a $1 + iY$, sin embargo, $1 + i$ sí divide tanto a $1 + iY$ como a $1 - iY$ (se sigue de que Y es impar).

Al dividir los dos miembros de (11.13) entre $(1 + i)^2$, en el primer miembro queda un producto de dos factores primos entre sí, y en el segundo miembro queda $-iX^3 = (iX)^3$. Por lo tanto, cada factor del primer miembro ha de ser un cubo. En particular $(1+iY)/(1+i)$ es un cubo, es decir, $1+iY = (1+i)(a+ib)^3$. Conjugando y sumando como antes se llega a $1 = (a + b)(a^2 - 4ab + b^2)$ y de aquí salen las soluciones $a = \pm 1$, $b = 0$ o bien $a = 0$, $b = \pm 1$, que implican $y = \pm 2$, y entonces $x = 2$. ∎

Ahora veremos una interesante prueba debida a Nagell de una conjetura de Ramanujan.

Teorema *Las únicas soluciones de la ecuación $x^2+7=2^n$ son las siguientes:*

$$(x,n) \quad = \quad (\pm 1,3), \quad (\pm 3,4), \quad (\pm 5,5), \quad (\pm 11,7), \quad (\pm 181,15).$$

DEMOSTRACIÓN: Claramente, x tiene que ser impar, y podemos suponer que es positivo. Supongamos primero que $n=2m$. Entonces podemos factorizar

$$(2^m+x)(2^m-x)=7,$$

de donde $2^m+x=7$ y $2^m-x=1$. Sumando, $2^{m+1}=8$, luego $m=2$, $n=4$, $x=3$.

Sea ahora n impar. El caso $n=3$ lleva a la solución $(1,3)$. Supongamos que $n>3$.

Trabajamos en el cuerpo $\mathbb{Q}\big(\sqrt{-7}\,\big)$, cuyos enteros forman un dominio de factorización única. La descomposición en factores primos de 2 es la siguiente:

$$2=\left(\frac{1+\sqrt{-7}}{2}\right)\left(\frac{1-\sqrt{-7}}{2}\right)$$

Como x es impar, es fácil ver que x^2+7 es múltiplo de 4. Tomando $m=n-2$ podemos reescribir la ecuación del siguiente modo:

$$\frac{x^2+7}{4}=2^m$$

y podemos factorizarla así en $\mathbb{Q}\big(\sqrt{-7}\,\big)$:

$$\left(\frac{x+\sqrt{-7}}{2}\right)\left(\frac{x-\sqrt{-7}}{2}\right)=\left(\frac{1+\sqrt{-7}}{2}\right)^m\left(\frac{1-\sqrt{-7}}{2}\right)^m.$$

Ninguno de los primos de la derecha divide a la vez a los dos factores de la izquierda, pues entonces dividiría a su diferencia, $\sqrt{-7}$, pero las normas correspondientes no se dividen.

Por la unicidad de la factorización (y teniendo en cuenta que las unidades son ± 1), podemos concluir que

$$\frac{x\pm\sqrt{-7}}{2}=\pm\left(\frac{1\pm\sqrt{-7}}{2}\right)^m,$$

donde esta expresión representa a dos ecuaciones de las que desconocemos los signos adecuados, pero en cualquier caso la ecuación para $\frac{x+\sqrt{-7}}{2}$ debe tener a la derecha un primo y la de $\frac{x-\sqrt{-7}}{2}$ debe tener el otro, mientras que la unidad ± 1 debe ser la misma en ambas ecuaciones. Al restarlas queda

$$\pm\sqrt{-7}=\left(\frac{1+\sqrt{-7}}{2}\right)^m-\left(\frac{1-\sqrt{-7}}{2}\right)^m.$$

Ahora probamos que el signo ha de ser negativo. Supongamos que es positivo. Llamemos $a=\frac{1+\sqrt{-7}}{2}$ y $b=\frac{1-\sqrt{-7}}{2}$. Estamos suponiendo que $a^m-b^m=a-b$.

Como $a+b=1$ y $ab=2$, tenemos lo siguiente:

$$a^2=(1-b)^2=1-2b+b^2=1-ab^2+b^2\equiv 1 \pmod{b^2}$$

Por lo tanto $a^m=a(a^2)^{(m-1)/2}\equiv a$ (mód b^2), y así

$$a-b=a^m-b^m\equiv a-0 \text{ (mód } b^2),$$

es decir, $b^2 \mid b$, lo cual es imposible.

Así pues, $-2^m\sqrt{-7}=\big(1+\sqrt{-7}\big)^m-\big(1-\sqrt{-7}\big)^m$. Al desarrollar por el teorema del binomio de Newton se cancelan los términos pares y queda

$$-2^m\sqrt{-7}=2\binom{m}{1}\sqrt{-7}+2\binom{m}{3}\big(\sqrt{-7}\big)^3+\cdots+2\binom{m}{m}\big(\sqrt{-7}\big)^m.$$

Sacamos factor común 2 y concluimos que $-2^{m-1}\equiv m$ (mód 7).

Como $2^6\equiv 1$ (mód 7), es claro que si un número m cumple esta congruencia, también lo cumplen todos los congruentes con m módulo 42.

Si examinamos todos los números entre 0 y 41, vemos que los únicos que cumplen la congruencia son $m=3$, 5 y 13. El teorema estará probado si demostramos que dos soluciones de la ecuación original (en el caso que estamos estudiando) no pueden ser congruentes módulo 42, pues entonces las únicas soluciones posibles serán $n=5$, 7, 15, con las que se corresponden $x=5$, 11, 181.

Supongamos que m y m' son soluciones de la ecuación (en realidad queremos decir que $m+2$ y $m'+2$ lo son). Supongamos que $m<m'$ y $m\equiv m'$ (mód 42). Sea 7^l la mayor potencia de 7 que divide a $m'-m$. Entonces

$$a^{m'}=a^m a^{m'-m}=a^m\left(\frac{1}{2}\right)^{m'-m}\big(1+\sqrt{-7}\big)^{m'-m}. \tag{11.14}$$

Como $\phi(7^{l+1})=6\cdot 7^l$, el teorema 5.6 implica que $2^{6\cdot 7^l}\equiv 1$ (mód 7^{l+1}), y como $6\cdot 7^l \mid m'-m$, tenemos

$$2^{m'-m}\equiv 1 \text{ (mód } 7^{l+1}). \tag{11.15}$$

Una sencilla inducción (en la que se usa el teorema 4.25) prueba que

$$\big(1+\sqrt{-7}\big)^{7^l}=1+7^l\sqrt{-7}\,(1+7\alpha),$$

para cierto entero α. Al elevar después a $(m'-m)/7^l$ y desarrollar, obtenemos que

$$(1+\sqrt{-7})^{m'-m}\equiv 1+(m'-m)\sqrt{-7} \pmod{7^{l+1}}. \tag{11.16}$$

Ahora pasamos la potencia de 2 al primer miembro de (11.14), tomamos congruencias módulo 7^{l+1} en el anillo de enteros de $\mathbb{Q}\big(\sqrt{-7}\big)$ y sustituimos (11.15) y (11.16), de lo que resulta

$$a^{m'}\equiv a^m+a^m(m'-m)\sqrt{-7} \pmod{7^{l+1}}. \tag{11.17}$$

De nuevo por inducción se prueba que $a^m \equiv \frac{1+m\sqrt{-7}}{2^m}$ (mód 7), luego podemos escribir $a^m = \frac{1+m\sqrt{-7}}{2^m} + 7\alpha$ para un cierto entero α. Como $7^l \mid m'-m$, al sustituir en (11.17) queda

$$a^{m'} \equiv a^m + \frac{m'-m}{2^m}\sqrt{-7} \pmod{7^{l+1}}. \tag{11.18}$$

Conjugando:

$$b^{m'} \equiv b^m - \frac{m'-m}{2^m}\sqrt{-7} \pmod{7^{l+1}}. \tag{11.19}$$

Pero al ser m y m' soluciones de la ecuación, tenemos

$$a^m - b^m = -\sqrt{-7} = a^{m'} - b^{m'},$$

luego al restar (11.18) menos (11.19) llegamos a que

$$(m'-m)\sqrt{-7} \equiv 0 \pmod{7^{l+1}}.$$

Por lo tanto, la multiplicidad del primo $\sqrt{-7}$ en $m'-m$ es al menos $2l+1$, pero ésta ha de ser el doble de la multiplicidad de 7 (visto como primo de $\mathbb{Z}$) en $m'-m$, luego la multiplicidad de $\sqrt{-7}$ es al menos $2l+2$ y así $7^{l+1} \mid m'-m$, en contra de la elección de l. ∎

Hemos afirmado sin prueba que $\mathbb{Q}\big(\sqrt{-163}\,\big)$ es un DFU (a pesar de que hemos demostrado que no es un dominio euclídeo). Se trata de un hecho muy profundo que no estaremos preparados para probar hasta el capítulo XIII. De momento vamos a aceptarlo para probar un curioso teorema de Euler:

Teorema *El polinomio $p(x) = x^2 + x + 41$ toma valores primos sobre todos los números naturales entre 0 y 39.*

Demostración: Es claro que si $m < n$, entonces $p(m) < p(n)$, luego los cuarenta primos que se obtienen son distintos. También es fácil ver que $p(-n-1) = p(n)$, luego los 40 primeros números negativos dan los mismos primos.

Como hemos dicho, la prueba usa la factorización única en $\mathbb{Q}\big(\sqrt{-163}\,\big)$. Un entero de este cuerpo es de la forma $x + y\frac{1+\sqrt{-163}}{2}$, donde x e y son enteros racionales. Si hacemos $y = 1$ obtenemos un entero $\frac{2x+1+\sqrt{-163}}{2}$ con la propiedad de que no tiene divisores enteros racionales no unitarios, y su norma es, precisamente, $x^2 + x + 41$.

Supongamos que $m = x^2 + x + 41$ es compuesto para un cierto entero x tal que $0 \le x \le 39$. Puesto que $m < 40^2 + 40 + 41 = 41^2$, Existirá un primo $p \mid m$ tal que $p \le 39$.

Tenemos p divide al producto de $\frac{2x+1+\sqrt{-163}}{2}$ por su conjugado, pero no puede dividir a ninguno de los dos factores, luego p no es primo en $\mathbb{Q}\big(\sqrt{-163}\,\big)$.

Como tiene norma p^2, ha de descomponerse en producto de dos primos de norma p. Digamos que

$$p = \left(s + t\sqrt{-163}\,\right)\left(s - t\sqrt{-163}\,\right) = s^2 + 163t^2,$$

donde s y t son ambos enteros o ambos semienteros. Como p es un primo racional, necesariamente $t \neq 0$. Esto nos da la contradicción

$$163 \leq (2s)^2 + 163(2t)^2 = 4p \leq 4 \cdot 39 = 156.$$

■

Observar que $p(40) = 40^2 + 40 + 41 = 40 \cdot 41 + 41 = 41^2$ ya no es primo.

Los cuarenta primos alcanzados son:

41, 43, 47, 53, 61, 71, 83, 97, 113, 131, 151, 173, 197, 223, 251,

281, 313, 347, 383, 421, 461, 503, 547, 593, 641, 691, 743, 797, 853,

911, 971, 1.033, 1.097, 1.163, 1.231, 1.301, 1.373, 1.447, 1.523, 1.601.

Capítulo XII

Factorización ideal

En este capítulo nos adentraremos en la teoría de la divisibilidad, cuyo estudio comenzamos en el capítulo IV. Nuestro objetivo es demostrar que, a pesar de los ejemplos que hemos visto en el capítulo anterior, los anillos de enteros algebraicos sí tienen factorización única siempre que admitamos la existencia de divisores 'ideales' que no se corresponden con elementos del anillo. Esta factorización única se comporta de modo extremadamente regular y, aunque no es tan potente como la factorización única 'real', en muchos casos es suficiente para comprender el comportamiento de estos anillos y, en parte, el comportamiento de los números enteros. El origen de las ideas que vamos a exponer se remontan al estudio de Kummer sobre los enteros ciclotómicos. Kummer quería averiguar si los anillos de enteros ciclotómicos tienen factorización única, y para ello se dedico a buscar factorizaciones de primos racionales por si encontraba algún contraejemplo. En este proceso llegó a desarrollar una teoría que le permitía determinar a priori el número de divisores primos que debía tener un primo racional p, sus multiplicidades, así como criterios explícitos que le permitían determinar a qué enteros ciclotómicos debía dividir cada factor, todo esto antes de encontrar dichos factores primos. Cuando llegó por fin a encontrar un contraejemplo a la factorización única su teoría estaba tan perfeccionada y era tan coherente que Kummer pudo demostrar que era consistente hablar de divisores 'ideales' que se comportaran según las reglas que había obtenido, aunque no se correspondieran con ningún divisor real en el anillo (del mismo modo que los antiguos algebristas trabajaban con números 'imaginarios' aunque no se correspondieran con ningún número real). A finales del siglo XIX Dedekind formalizó la teoría de Kummer identificando sus divisores ideales con los ideales en el sentido usual de la teoría de anillos y probando que los resultados de Kummer son válidos en una clase muy general de anillos que ahora introducimos.

12.1 Dominios de Dedekind

Recordemos que si $\mathfrak{a}$ y $\mathfrak{b}$ son ideales de un anillo D, su producto es

$$\mathfrak{ab} = \left\{ \sum_{i=1}^{n} p_i q_i \;\middle|\; n \in \mathbb{N} \text{ y } p_i \in \mathfrak{a}, q_i \in \mathfrak{b} \text{ para } i = 1, \ldots, n \right\}. \tag{12.1}$$

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 12.1 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 12.2 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 (12.1) 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 12.3 *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 comentado 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 12.4 *Si D es un dominio de Dedekind, entonces los ideales fraccionales de D forman un grupo. Además los ideales primos 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

$$\big(\mathfrak{p} + (d)\big)/\mathfrak{p} = \big(\mathfrak{p}_1/\mathfrak{p}\big) \cdots \big(\mathfrak{p}_r/\mathfrak{p}\big) \quad \text{y} \quad \big(p + (d^2)\big)/\mathfrak{p} = \big(\mathfrak{q}_1/\mathfrak{p}\big) \cdots \big(\mathfrak{q}_s/\mathfrak{p}\big).$$

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 12.5 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 $\mathfrak{b} \mid \mathfrak{a}$. Notar 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.

Observar 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 (notar 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 $c = \mathfrak{ab}$ en lugar de $\mathfrak{a} \mid (c)$ o $(c) = \mathfrak{ab}$. 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. Notar 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 mayor 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 al 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 12.6 *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. ∎

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. Así, por ejemplo, hasta ahora no hemos

sabido probar que un anillo de enteros algebraicos tiene factorización única salvo en el caso en que de hecho es un dominio euclídeo, pero esta propiedad es mucho más fuerte y deja de cubrir muchos casos de interés. El concepto de dominio de Dedekind, en cambio, admite una caracterización algebraica muy fácil de verificar en la práctica. Veámosla.

Teorema 12.7 *(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: Todo dominio de Dedekind es noetheriano, pues una cadena de ideales estrictamente creciente significaría una cadena decreciente de divisores, lo cual es imposible. La propiedad 2) está probada en el teorema 12.4. La prueba del teorema 4.24 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 3.9 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.

Notar 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

$$a_i s = \sum_{j=1}^{m} b_{ij} a_j \qquad \text{para } i = 1, \ldots, m.$$

Esto puede expresarse matricialmente mediante la ecuación $s(a_j)^t = B(a_j)^t$, donde llamamos $B = (b_{ij})$. Equivalentemente, $(B - sI_m)(a_j)^t = 0$. Por consiguiente la matriz $B - sI_m$ no puede ser regular, pues entonces multiplicando por su inversa concluiríamos que $(a_j) = 0$, lo cual es imposible. Por lo tanto $|B - sI_m| = 0$ y el polinomio $p(x) = |B - xI_m| \in D[x]$ es mónico, no nulo y tiene por raíz a s. 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 $r = s$. ∎

Observar que la prueba del teorema anterior nos ha dado una expresión explícita para el inverso de un ideal en un dominio de Dedekind:

$$\mathfrak{a}^{-1} = \{x \in K \mid x\mathfrak{a} \subset D\}.$$

12.2 Factorización ideal en anillos de enteros

Los teoremas 11.20 y 11.23, junto con el teorema de Dedekind, implican que todo anillo de enteros algebraicos de un cuerpo numérico es un dominio de Dedekind. El teorema 11.20 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 cumplen esta condición. El interés de la finitud de los cocientes reside en que es la clave para definir la norma de un ideal, que jugará el mismo papel que la norma de los elementos reales en el estudio de la divisibilidad. En efecto:

Definición 12.8 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 11.21 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 12.9 *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* $\mathrm{N}(\mathfrak{ab}) = \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{b})$.

DEMOSTRACIÓN: Por la unicidad de la factorización en primos e inducción sobre el número de factores, basta probar que $\mathrm{N}(\mathfrak{ap}) = \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{p})$ cuando $\mathfrak{p}$ es un ideal primo (el caso en que uno de los factores es 1 es obvio).

Consideremos los grupos abelianos finitos $\mathfrak{a}/\mathfrak{ap} \le \mathcal{O}/\mathfrak{ap}$. 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 $|\mathcal{O}/\mathfrak{p}| = |\mathfrak{a}/\mathfrak{ap}|$. ∎

Con esto estamos en condiciones de ver ejemplos cómodamente. Vimos en el capítulo anterior que en $\mathbb{Q}\big(\sqrt{-5}\big)$ teníamos las siguientes descomposiciones en irreducibles:

$$6 = 2 \cdot 3 = \big(1 + \sqrt{-5}\big)\big(1 - \sqrt{-5}\big). \tag{12.2}$$

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 $3 = \mathfrak{qr}$. Por otra parte, los factores de la derecha en (12.2) 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

$$6 = 2 \cdot 3 = (\mathfrak{p}_1\mathfrak{p}_2)(\mathfrak{qr}) = (\mathfrak{p}_1\mathfrak{q})(\mathfrak{p}_2\mathfrak{r}) = \big(1 + \sqrt{-5}\big)\big(1 - \sqrt{-5}\big).$$

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 $\mathfrak{q} \ne \mathfrak{r}$, pues en otro caso $1 = 3 - \big(1 + \sqrt{-5} + 1 - \sqrt{-5}\big) \in \mathfrak{q}$, o sea, $\mathfrak{q} = 1$.

Si llamamos $\mathfrak{p} = \mathfrak{p}_1 = \mathfrak{p}_2$, la factorización de 6 es, en definitiva, $6 = \mathfrak{p}^2\mathfrak{qr}$. Los factores son 'ideales' porque no están en el anillo $\mathbb{Z}\big[\sqrt{-5}\big]$, pero se comportan como si lo estuviesen.

Ejercicio: Obtener las factorizaciones ideales correspondientes a las factorizaciones no únicas mostradas en el capítulo anterior en cuerpos cuadráticos reales e imaginarios.

El teorema siguiente recoge los hechos básicos sobre las normas de ideales análogos a los ya conocidos para normas de enteros reales.

Teorema 12.10 *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 12.9.

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 $\mathcal{O}/\mathfrak{a}$ es en particular un grupo finito (con la suma) y el orden de cualquier elemento es divisible entre $\mathrm{N}(\mathfrak{a})$. 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 12.11 *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).
3) se sigue de 1) porque cada ideal es un divisor de su norma. ∎

Veamos ahora que en la mayoría de los casos es muy fácil encontrar las factorizaciones ideales de los primos racionales (y con ello ir encontrando todos los ideales primos de un anillo de enteros). El teorema siguiente da cuenta de ello.

Teorema 12.12 *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 $g(x) = \operatorname{pol}\operatorname{mín}(\alpha,\mathbb{Q})$ 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 $i = 1,\dots,r$ son primos distintos y la descomposición de p en primos es $p = \mathfrak{p}_1^{e_1}\cdots\mathfrak{p}_r^{e_r}$.*

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 $\bar{g}_i = \operatorname{pol}\operatorname{mín}(\alpha_i, \mathbb{Z}/p\mathbb{Z})$.

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 $\bar{q}(a_i) - \bar{r}(a_i) = 0$.

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

$$\mathfrak{p}_1^{e_1}\cdots\mathfrak{p}_r^{e_r} \subset \big(p, g_1(\alpha)^{e_1}\cdots g_r(\alpha)^{e_r}\big) = \big(p, g(\alpha)\big) = (p,0) = (p).$$

Notar 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 $\operatorname{N}(\mathfrak{p}_1^{e_1}\cdots\mathfrak{p}_r^{e_r}) = \mathfrak{p}^{e_1\operatorname{grad} g_1+\cdots+e_r\operatorname{grad} g_r} = p^n$, 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

$$3 = \big(3, 1+\sqrt{-5}\,\big)\big(3, -1+\sqrt{-5}\,\big) = \big(3, 1+\sqrt{-5}\,\big)\big(3, 1-\sqrt{-5}\,\big).$$

∎

En el capítulo siguiente estudiaremos con detalle las factorizaciones en cuerpos cuadráticos. Veamos ahora un resultado general para cuerpos ciclotómicos.

Teorema 12.13 *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 $f = \mathrm{o}_p(q)$ (el orden de q módulo p), y $r = (p-1)/f$.

Demostración: 1) En la prueba del teorema 11.15 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) Aplicamos el teorema anterior. Sea $g(x)$ el polinomio mínimo de ω. Éste divide a x^p-1, lo cual sigue siendo cierto módulo q. Como la derivada de x^p-1 es px^{p-1}, que es no nulo módulo q, y su única raíz es 0, concluimos que las raíces de x^p-1 en una extensión de $\mathbb{Z}/q\mathbb{Z}$ han de ser simples, luego lo mismo vale para las de $\bar{g}(x)$. En particular, los factores irreducibles de $\bar{g}(x)$ en $\mathbb{Z}/q\mathbb{Z}$ son todos distintos dos a dos, luego la factorización de q es de la forma $q = \mathfrak{p}_1 \cdots \mathfrak{p}_r$, donde los factores son distintos. Sólo falta probar que $f = \mathrm{o}_p(q)$, pues comparando las normas sale que $r = (p-1)/f$. Sea $e = \mathrm{o}_p(q)$.

Sabemos que $L = \mathbb{Z}[\omega]/\mathfrak{p}_i$ es un cuerpo con q^f elementos. En él, $[\omega]$ es una raíz p-ésima de la unidad distinta de 1 ($\mathfrak{p}_i$ no puede dividir a $\pi = \omega - 1$), luego tiene orden p, pero por otro lado el grupo multiplicativo de L tiene orden q^f-1, con lo que $p \mid q^f - 1$ y, en consecuencia, $e = \mathrm{o}_p(q) \mid f$.

Un elemento de L es de la forma $h\big([\omega]\big)$, con $h \in (\mathbb{Z}/q\mathbb{Z})[x]$. Como los elementos de $\mathbb{Z}/q\mathbb{Z}$ cumplen $a^q = a$, resulta que $h\big([w]\big)^{q^e} = h\left([\omega]^{q^e}\right) = h\left([\omega^{q^e}]\right) =$

$h\big([\omega]\big)$, puesto que $q^e \equiv 1$ (mód p). Esto significa que todos los elementos de L son raíces de $x^{q^e} - x$ y así, el número de elementos de L ha de cumplir $q^f \leq q^e$, luego $e = f$. ∎

Ejemplo Vamos a considerar el caso $p = 23$ y $q = 47$ en el teorema anterior. Como $q \equiv 1$ (mód p), tenemos que $f = \mathrm{o}_p(q) = 1$, 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 $\mathbb{Q}\big(\sqrt{-47}\,\big)$ no hay enteros de norma 47.

Ahora bien, un entero de $\mathbb{Q}\big(\sqrt{-47}\,\big)$ 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 1.844 las descomposiciones en factores primos de todos los primos racionales ≤ 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.

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 12.12 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 12.12).

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 $\mathrm{o}(u)=3$, 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

12.3 Dominios de Dedekind y dominios de factorización única

Terminamos el capítulo profundizando un poco más en la relación entre la divisibilidad real e ideal, es decir, entre los dominios de factorización única y los dominios de Dedekind. Un hecho sencillo que conocemos es el siguiente:

Un dominio de Dedekind es un dominio de factorización única si y sólo si todos sus elementos irreducibles son primos.

De hecho esto es cierto para todo anillo noetheriano (por el teorema 4.9). Ahora vamos a probar un resultado dual:

Teorema 12.14 *Un dominio de factorización única es un dominio de Dedekind si y sólo si todos sus ideales primos son maximales.*

DEMOSTRACIÓN: Una implicación es obvia. Sea D un DFU donde los ideales maximales coinciden con los primos y veamos que es un dominio de Dedekind. Dividimos la prueba en varios pasos.

(i) *Si π es primo en D, entonces las unidades módulo π^r son las clases de los elementos primos con π.*

Lo probamos por inducción sobre r. Para $r = 1$ es obvio porque $D/(\pi)$ es un cuerpo.

Si vale para r y $(\alpha, \pi) = 1$, por hipótesis de inducción existe $\xi_o \in D$ de manera que $\alpha\xi_0 \equiv 1 \pmod{\pi^r}$, 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 $\alpha\xi \equiv 1 \pmod{\pi^{r+1}}$.

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) *Si $\beta \in D$ es no nulo ni unitario entonces las unidades módulo β son las clases de los elementos primos con β.*

Factoricemos $\beta = \epsilon\pi_1^{r_1} \cdots \pi_n^{r_n}$, donde los π_i son primos distintos y ϵ es una unidad. Consideramos el homomorfismo $D \longrightarrow \big(D/(\pi_1^{r_1})\big) \times \cdots \times \big(D/(\pi_n^{r_n})\big)$ definido por $\alpha \mapsto \big([\alpha], \ldots, [\alpha]\big)$. Claramente su núcleo es (β). Veamos que es suprayectivo, con lo que tendremos $D/(\beta) \cong \big(D/(\pi_1^{r_1})\big) \times \cdots \times \big(D/(\pi_n^{r_n})\big)$.

Dados $\alpha_1, \ldots, \alpha_n \in D$, sea $\beta_i = \beta/\pi_i^{r_i}$. Entonces $(\beta_i, \pi_i) = 1$, luego por (i) tenemos que β_i es una unidad módulo $\pi_i^{r_i}$. En consecuencia existe $\gamma_i \in D$ tal que $\beta_i\gamma_i \equiv \alpha_i \pmod{\pi_i^{r_i}}$. Sea $\alpha = \beta_1\gamma_1 + \cdots + \beta_n\gamma_n$. Claramente $\alpha \equiv \alpha_i \pmod{\pi_i^{r_i}}$, luego es una antiimagen de la n-tupla $\big([\alpha_1], \ldots, [\alpha_n]\big)$.

Ahora es claro que α es una unidad módulo β si y sólo si lo es módulo $\pi_i^{r_i}$ para todo i, lo que por (i) equivale a que $(\alpha, \beta) = 1$.

(iii) *Si $\delta = \mathrm{mcd}(\alpha, \beta)$, entonces $(\delta) = (\alpha, \beta)$.*

Sea $\alpha = \alpha'\delta$ y $\beta = \beta'\delta$. Entonces $(\alpha', \beta') = 1$, luego por (ii) existe ξ tal que $\alpha'\xi \equiv 1 \pmod{\beta'}$.

Así $\beta' \mid \alpha'\xi - 1$, luego multiplicando por δ queda $\beta \mid \alpha\xi - \delta$, es decir, $\delta = \alpha\xi + \beta\xi'$ para cierto ξ', y por lo tanto $(\delta) \subset (\alpha, \beta)$. La otra inclusión es obvia.

(iv) *D es DIP.*

Sea A un ideal propio de D. Sea $\alpha \in A$ no nulo. Factoricemos $\alpha = \pi_1 \cdots \pi_n$ como producto de primos. Si π es cualquier primo y $A \subset (\pi)$ entonces $\pi \mid \alpha$ luego $\pi = \pi_i$ para algún i.

Definimos $e(\pi) = \min_\alpha e_\pi(\alpha)$, donde α recorre los elementos no nulos de A y $e_p(\alpha)$ es el exponente de π en α. Acabamos de probar que si $e(\pi) > 0$ entonces $\pi = \pi_i$ para un i, es decir, que $e(\pi) = 0$ 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 $\beta_i \in A$ de modo que $e_{\pi_i}(\beta_i) = e(\pi_i)$. Sea $\delta = \pi_1^{e(\pi_1)} \cdots \pi_n^{e(\pi_n)}$. Obviamente se cumple $A \subset (\delta)$. Además $\delta = \mathrm{mcd}(\alpha, \beta_1, \ldots, \beta_n)$, pues ciertamente δ divide a α, $\beta_1, \ldots, \beta_n$ (porque están en A) y si γ es un divisor común de todos ellos entonces 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 (iii) tenemos que $(\delta) = (\alpha, \beta_1, \ldots, \beta_n) \subset A$, luego $A = (\delta)$. ∎

Conviene observar que en la prueba del teorema anterior hemos visto que (ii) → (iii) → (iv), por lo que si un DFU cumple (ii) entonces es un DIP. Así mismo, la prueba de (ii) contiene una demostración del teorema chino del resto (cf. 5.10) para DFU's que sean dominios de Dedekind, es decir, para DIP's.

Ejercicio: Probar que si un DFU cumple el teorema chino del resto entonces es un DIP.

Ejercicio: Probar el teorema chino del resto en un dominio de Dedekind arbitrario, es decir, probar el teorema 5.10 para ideales $\mathfrak{m}_1, \ldots, \mathfrak{m}_n$ primos entre sí en un dominio de Dedekind D y después dar un enunciado equivalente en términos de congruencias.

Ejercicio: Probar que si X es infinito entonces $\mathbb{Q}[X]$ es un DFU no noetheriano.

Ejercicio: Probar que $\mathbb{Z}[x]$ es un DFU noetheriano pero no es un dominio de Dedekind.

Volviendo al caso de los anillos de enteros algebraicos, puesto que todos ellos son dominios de Dedekind, probar que tienen factorización única equivale a probar que son DIP's. Sin embargo, no es fácil, no ya probar que todos los ideales de un anillo sean principales, sino tan sólo decidir si un ideal dado lo es. La factorización ideal proporciona técnicas para abordar este problema. En el capítulo siguiente nos ocuparemos de ello en el caso particular de los cuerpos cuadráticos.

Capítulo XIII

Factorización en cuerpos cuadráticos

Los cuerpos cuadráticos son los más simples de todos los cuerpos numéricos y, entre ellos, los más simples son los imaginarios. Sin embargo, su estudio proporciona resultados importantes sobre los números enteros. En otros capítulos hemos visto algunas de sus aplicaciones. Aquí vamos a estudiar más a fondo su estructura, demostrando algunos resultados que no tienen análogos en otros cuerpos y también otros válidos para cuerpos numéricos arbitrarios, pero cuyas demostraciones generales no están a nuestro alcance. Entre otras cosas, daremos un algoritmo para determinar si un cuerpo cuadrático tiene o no factorización única, así como para decidir si un ideal dado es o no principal, y obtener en su caso un generador.

13.1 Los primos cuadráticos

Comenzamos determinando los primos cuadráticos. Para facilitar el estudio conviene introducir unas definiciones.

Definición 13.1 Sea $\mathbb{Q}(\sqrt{d}\,)$ un cuerpo cuadrático, $\mathcal{O}$ su anillo de enteros y p un primo racional. Entonces $\mathrm{N}(p) = p^2$, luego p puede descomponerse a lo sumo en dos factores primos de $\mathcal{O}$ de norma p. Esto da lugar a tres modos posibles de factorización. Si p sigue siendo primo en $\mathcal{O}$, diremos que p *se conserva.* Si p se descompone en producto de dos ideales primos distintos, $p = \mathfrak{pq}$, diremos que p *se escinde.* Si p es el cuadrado de un primo, $p = \mathfrak{p}^2$, diremos que p *se ramifica.*

Ahora vamos a aplicar el teorema 12.12 a los cuerpos cuadráticos. En primer lugar, si $\mathbb{Q}(\sqrt{d}\,)$ es un cuerpo cuadrático, entonces su anillo de enteros es de la forma $\mathbb{Z}[\alpha]$, tal y como exigen las hipótesis. El entero α es $\sqrt{d}$ si $d \equiv 1 \pmod 4$ o bien $\alpha = \frac{1+\sqrt{d}}{2}$ si $d \not\equiv 1 \pmod 4$. En el primer caso el polinomio mínimo de α es $g(x) = x^2 - d$ y en el segundo es $g(x) = x^2 - x + \frac{1-d}{4}$.

Según 12.12, la factorización de un primo racional p depende de la factorización de $g(x)$ módulo p. Más concretamente, tenemos que p se conserva primo, se ramifica o se escinde según si $g(x)$ tiene 0, 1 o 2 raíces distintas módulo p respectivamente.

Consideramos primero el caso $p \neq 2$, pues bajo esta hipótesis podemos considerar la fórmula usual para las soluciones de la ecuación general de grado 2, en virtud de la cual la factorización de $g(x)$ depende únicamente de su discriminante, que resulta ser $\Delta = 4d$ si $d \not\equiv 1 \pmod 4$ y $\Delta = d$ si $d \equiv 1 \pmod 4$ (o sea, el discriminante de $g(x)$ es precisamente el discriminante de K).

Ahora conviene introducir un concepto:

Si p es un primo impar y d un entero primo con p, diremos que d es un *resto cuadrático* módulo p si existe un $u \in \mathbb{Z}$ tal que $d \equiv u^2 \pmod p$. En caso contrario (siempre suponiendo que d es primo con p) diremos que d es un *resto no cuadrático* módulo p.

En estos términos, tenemos que $g(x)$ tiene 0, 1 o 2 raíces módulo p si y sólo si Δ es un resto no cuadrático, es cero o es un resto cuadrático módulo p, respectivamente. Ahora bien, como $\Delta = d$ o $\Delta = 4d$ y $p \neq 2$, en realidad este criterio sigue siendo cierto si cambiamos Δ por d.

Así pues, si $p \mid d$ (o, equivalentemente, si $p \mid \Delta$) tenemos que

$$g(x) = x^2 - d \equiv x^2 \pmod p \quad \text{o} \quad g(x) = x^2 - x + \frac{1-d}{4} \equiv (x - 1/2)^2 \pmod p,$$

donde el $1/2$ ha de entenderse como el inverso de 2 módulo p. Por consiguiente $p = \mathfrak{p}^2$, donde

$$\mathfrak{p} = (p, \alpha) \qquad \text{o} \qquad \mathfrak{p} = (p, \alpha - 1/2).$$

Supongamos ahora que $p \nmid d$, en cuyo caso tenemos dos posibilidades. Si d es un resto no cuadrático módulo p entonces p se conserva primo, mientras que si $d \equiv u^2 \pmod p$ entonces

$$g(x) = x^2 - d \equiv (x+u)(x-u) \pmod p$$

o bien

$$g(x) = x^2 - x + \frac{1-d}{4} \equiv \left(x - \frac{1+u}{2}\right)\left(x - \frac{1-u}{2}\right),$$

de donde a su vez $p = \mathfrak{p}_1\mathfrak{p}_2$, con

$$\mathfrak{p}_i = (p, \alpha \pm u) \qquad \text{o} \qquad \mathfrak{p}_i = \left(p, \alpha - \frac{1 \pm u}{2}\right).$$

Nos falta estudiar el caso $p = 2$. Si $2 \mid d$ entonces $d \equiv 2 \pmod 4$ (pues d no puede ser múltiplo de 4). Entonces

$$g(x) = x^2 - d \equiv x^2 \pmod 2,$$

luego $2 = \mathfrak{p}^2$, con $\mathfrak{p} = (2, \alpha)$.

Si $2 \nmid d$ tenemos dos posibilidades: si $d \equiv 3 \pmod 4$ entonces

$$g(x) = x^2 - d \equiv x^2 - 1 \equiv (x-1)^2 \pmod 2,$$

luego $2 = \mathfrak{p}^2$ con $\mathfrak{p} = (2, \alpha - 1)$.

Si, por el contrario, $d \equiv 1 \pmod 4$, digamos $d = 4k + 1$, entonces

$$g(x) = x^2 - x - k.$$

A su vez hemos de distinguir dos casos: Si k es impar (o equivalentemente, si $d \equiv 5 \pmod 8$) entonces $g(x)$ no tiene raíces módulo 2 y p se conserva primo. Si k es par (o sea, $d \equiv 1 \pmod 8$), entonces

$$g(x) \equiv x^2 - x = x(x-1) \pmod 2,$$

luego $2 = \mathfrak{p}_1\mathfrak{p}_2$, donde $\mathfrak{p}_1 = (2, \alpha)$, $\mathfrak{p}_2 = (2, \alpha - 1)$.

El teorema siguiente resume todo lo que hemos obtenido, pero antes introducimos una notación útil: Si p es un primo y $u \in \mathbb{Z}$, definimos

$$[p, u] = (p, \alpha - u).$$

Es claro que el ideal $[p, u]$ depende sólo del resto de u módulo p, por lo que podemos considerar $u \in \mathbb{Z}/p\mathbb{Z}$.

Teorema 13.2 *Sea d un número entero libre de cuadrados y p un primo racional. Consideremos el anillo de los enteros algebraicos del cuerpo $\mathbb{Q}(\sqrt{d}\,)$.*

1. *Si p es impar y $p \mid d$, entonces p se ramifica:*

$$p = \begin{cases} [p,0]^2 & \text{si } d \not\equiv 1 \pmod p, \\ [p, \frac{1}{2}]^2 & \text{si } d \equiv 1 \pmod p. \end{cases}$$

2. *Si p es impar y $d \equiv u^2 \pmod p$, entonces p se escinde:*

$$p = \begin{cases} [p,u][p,-u] & \text{si } d \not\equiv 1 \pmod p, \\ [p, \frac{1+u}{2}][p, \frac{1-u}{2}] & \text{si } d \equiv 1 \pmod p. \end{cases}$$

3. *Si p es impar y d es un resto no cuadrático módulo p, entonces p se conserva primo.*

4. *Si $p = 2$ tenemos las posibilidades siguientes:*

$$\begin{aligned} &d \equiv 1 \pmod 8 && 2 = [2,0][2,1] && \textit{(se escinde),} \\ &d \equiv 2, 6 \pmod 8 && 2 = [2,0]^2 && \textit{(se ramifica),} \\ &d \equiv 3, 7 \pmod 8 && 2 = [2,1]^2 && \textit{(se ramifica),} \\ &d \equiv 5 \pmod 8 && 2 = 2 && \textit{(se conserva).} \end{aligned}$$

Observaciones:

1. No puede ocurrir $d \equiv 0, 4 \pmod 8$ porque d es libre de cuadrados.
2. Del teorema se sigue que un primo p se ramifica en K si y sólo si $p \mid \Delta$.
3. Si $\mathfrak{p} = [p, u]$ es un ideal en las condiciones del teorema, entonces $\mathrm{N}(\mathfrak{p}) = p$, luego $\mathbb{Z}[\alpha]/\mathfrak{p} = \mathbb{Z}/p\mathbb{Z}$, y $\alpha \equiv u \pmod{\mathfrak{p}}$. Esto nos permite determinar fácilmente si dos enteros cuadráticos son congruentes módulo $\mathfrak{p}$, y en particular si un entero cuadrático es divisible entre $\mathfrak{p}$.
4. El cuerpo K tiene un único automorfismo no trivial, al que llamaremos *conjugación* y representaremos por $x \mapsto \bar{x}$. Es claro que si $\mathfrak{a}$ es un ideal de K lo mismo le sucede a su imagen por la conjugación, que representaremos por $\bar{\mathfrak{a}}$. Del teorema se sigue que cuando p se escinde, entonces lo hace en la forma $p = \mathfrak{p}\bar{\mathfrak{p}}$, para cierto ideal primo $\mathfrak{p}$. Por ejemplo, en el caso $p = 2$, $d \equiv 1 \pmod 8$ tenemos que

$$2 = (2, \alpha)(2, \alpha - 1)$$

y, ciertamente, $\bar{\alpha} = -(\alpha - 1)$, luego $(2, \alpha - 1) = (2, \bar{\alpha}) = \overline{(2, \alpha)}$.

De la última observación se sigue fácilmente un resultado más general:

Teorema 13.3 *Sea K un cuerpo cuadrático y $\mathfrak{a}$ un ideal de su anillo de enteros. Entonces* $\mathrm{N}(\mathfrak{a}) = \mathfrak{a}\overline{\mathfrak{a}}$.

Demostración: Por la factorización única basta probarlo para ideales primos. Si $\mathfrak{p}$ es un ideal primo sea p el único primo racional al que divide. Si p se conserva primo, entonces $\mathfrak{p} = p$ y $\mathrm{N}(\mathfrak{p}) = p^2 = p\bar{p}$.

Si p se ramifica, entonces $p = \mathfrak{p}^2$, de donde $\overline{\mathfrak{p}} = \mathfrak{p}$ y $\mathrm{N}(\mathfrak{p}) = p$. Por lo tanto se cumple también $\mathrm{N}(\mathfrak{p}) = p = \mathfrak{p}^2 = \mathfrak{p}\overline{\mathfrak{p}}$.

Si p se escinde, hemos visto que lo hace en la forma $p = \mathfrak{p}\overline{\mathfrak{p}}$, y concluimos igualmente. ∎

13.2 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, 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 13.4 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 $h = 1$. Aquí demostraremos la finitud del número de clases en el caso particular de los cuerpos cuadráticos y daremos un procedimiento para calcularlo.

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 caso concreto de los cuerpos cuadráticos, si $\mathfrak{b}$ es un ideal no nulo tenemos que $\mathfrak{b}\overline{\mathfrak{b}} = \big(\mathrm{N}(\mathfrak{b})\big) \approx (1) = \mathfrak{b}\mathfrak{b}^{-1}$, luego $\overline{\mathfrak{b}}$ es similar a $\mathfrak{b}^{-1}$. Además $\mathfrak{a} \approx \mathfrak{b}$ si y sólo si $\mathfrak{a}\overline{\mathfrak{b}}$ es principal.

Nuestro objetivo es encontrar un conjunto finito de representantes de las clases de $\mathcal{H}$ y luego dar un algoritmo que nos permita saber cuándo dos clases son la misma, con lo que tendremos completamente determinado el grupo de clases.

El primer paso es observemos que todo ideal $\mathfrak{a}$ es similar a otro no divisible entre enteros racionales no unitarios. En efecto: si m es un entero maximal que divida a $\mathfrak{a}$, entonces $\mathfrak{a} = m\mathfrak{b}$ y obviamente $\mathfrak{b}$ no es divisible entre enteros racionales no unitarios. También es obvio que $\mathfrak{b}$ es similar a $\mathfrak{a}$. Ahora probamos que $\mathfrak{b}$ tiene un comportamiento especialmente satisfactorio.

Teorema 13.5 *Sea d un entero libre de cuadrados y $\mathcal{O}$ el anillo de los enteros de $\mathbb{Q}\big(\sqrt{d}\,\big)$. Sea $\mathfrak{a}$ un ideal de $\mathcal{O}$ que no sea divisible entre enteros racionales no unitarios. Sea $a = \mathrm{N}(\mathfrak{a})$. Entonces todo entero cuadrático es congruente módulo a con un entero racional y dos enteros racionales son congruentes módulo $\mathfrak{a}$ si y sólo si lo son módulo a. En particular se cumple que $\mathcal{O}/\mathfrak{a} \cong \mathbb{Z}/a\mathbb{Z}$.*

Demostración: Por hipótesis $\mathfrak{a}$ no es divisible entre primos que se conserven, luego su descomposición en primos es de la forma $\mathfrak{a} = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_r^{e_r}$, donde $\mathrm{N}(\mathfrak{p}_i) = p_i$. Si el primo p_i se ramifica entonces $e_i = 1$ (o de lo contrario $p_i \mid \mathfrak{a}$). Si p_i se escinde entonces $\overline{\mathfrak{p}}_i \nmid \mathfrak{a}$, o de lo contrario $p_i \mid \mathfrak{a}$. Por lo tanto los primos p_i son distintos dos a dos.

Veamos que si m es un entero racional, entonces $\mathfrak{a} \mid m$ si y sólo si $a \mid m$. En efecto, puesto que $\mathfrak{a} \mid a$, una implicación es evidente. Supongamos que $\mathfrak{a} \mid m$. Tenemos que $a = \mathrm{N}(\mathfrak{a}) = p_1^{e_1} \cdots p_r^{e_r}$.

Si p_i se ramifica, $e_i = 1$, $\mathfrak{p}_i \mid m$ y tomando normas $p_i \mid m^2$, luego $p_i^{e_i} \mid m$.

Si p_i se escinde entonces $\mathfrak{p}_i^{e_i} \mid m$ y conjugando $\overline{\mathfrak{p}}_i^{e_i} \mid m$. Como ambos ideales son primos entre sí concluimos que $p_i^{e_i} = \mathfrak{p}_i^{e_i}\overline{\mathfrak{p}}_i^{e_i} \mid m$.

Por lo tanto $a \mid m$.

Obviamente entonces, $m \equiv n \pmod{\mathfrak{a}}$ es equivalente a $m \equiv n \pmod{a}$. A su vez esto implica que la característica de $\mathcal{O}/\mathfrak{a}$ es a, luego $\mathbb{Z}/a\mathbb{Z} \subset \mathcal{O}/\mathfrak{a}$. Concretamente $\mathbb{Z}/a\mathbb{Z}$ está formado por las clases con representante entero racional.

Puesto que $a = \mathrm{N}(\mathfrak{a}) = |\mathcal{O}/\mathfrak{a}|$, ha de darse la igualdad y el teorema queda probado. ∎

El teorema siguiente contiene esencialmente la finitud del grupo de clases:

Teorema 13.6 *Sea d un entero libre de cuadrados y sea $\mathcal{O}$ el anillo de enteros de $\mathbb{Q}(\sqrt{d})$. Sea Δ el discriminante de K. Entonces todo ideal no nulo $\mathfrak{a}$ de $\mathcal{O}$ es similar a otro ideal $\mathfrak{b}$ tal que*

$$\mathrm{N}(\mathfrak{b})^2 \le \frac{|\Delta|}{3}.$$

Demostración: Podemos suponer que $\mathfrak{a}$ no es divisible entre enteros racionales no unitarios. Consideremos $a = \mathrm{N}(\mathfrak{a})$. Sea

$$\alpha = \begin{cases} \sqrt{d} & \text{si } d \not\equiv 1 \pmod 4 \\ \frac{1+\sqrt{d}}{2} & \text{si } d \equiv 1 \pmod 4 \end{cases} \tag{13.1}$$

El teorema anterior nos da que $\alpha \equiv s \pmod{\mathfrak{a}}$ para cierto entero racional s. Llamemos

$$r = \begin{cases} s & \text{si } d \not\equiv 1 \pmod 4 \\ \frac{2s-1}{2} & \text{si } d \equiv 1 \pmod 4 \end{cases}$$

En ambos casos tenemos que

$$r - \frac{\sqrt{\Delta}}{2} = s - \alpha \in \mathfrak{a}.$$

(observemos que en el segundo caso ninguno de los dos sumandos es entero, pero sí la suma.)

El teorema anterior nos da también que si a r le sumamos un múltiplo de a se sigue cumpliendo la relación anterior, luego podemos exigir que $|r| \le a/2$.

Tenemos que $\mathfrak{a} \mid r - \frac{\sqrt{\Delta}}{2}$, luego $r - \frac{\sqrt{\Delta}}{2} = \mathfrak{a}\mathfrak{b}$ para cierto ideal $\mathfrak{b}$. Además

$$\mathrm{N}(\mathfrak{a}\mathfrak{b}) = \left|\left(r - \frac{\sqrt{\Delta}}{2}\right)\left(r + \frac{\sqrt{\Delta}}{2}\right)\right| = \left|r^2 - \frac{\Delta}{4}\right| \le r^2 + \frac{|\Delta|}{4} \le \frac{1}{4}a^2 + \frac{|\Delta|}{4}.$$

Si $\mathrm{N}(\mathfrak{a}) \le \mathrm{N}(\mathfrak{b})$, entonces $\mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{a}) \le \mathrm{N}(\mathfrak{a})\,\mathrm{N}(\mathfrak{b})$, o sea, $a^2 \le (a^2 + |\Delta|)/4$, lo que equivale a que $a^2 \le |\Delta|/3$.

Por otro lado, por el teorema 13.3, $\mathrm{N}(\mathfrak{b}) = \mathfrak{b}\overline{\mathfrak{b}}$, y como $\mathfrak{a}\mathfrak{b}$ y $\mathfrak{b}\overline{\mathfrak{b}}$ son ambos principales, resulta que $\mathfrak{a} \approx \overline{\mathfrak{b}}$.

En resumen, $\mathfrak{a} \approx \overline{\mathfrak{b}}$ y $\mathrm{N}(\overline{\mathfrak{b}}) < \mathrm{N}(\mathfrak{a})$ salvo si $\mathrm{N}(\mathfrak{a})^2 \le |\Delta|/3$.

Además $\overline{\mathfrak{b}}$ no es divisible entre enteros racionales no unitarios, pues si $m \mid \overline{\mathfrak{b}}$ entonces $m \mid \mathfrak{b} \mid r - \frac{\sqrt{\Delta}}{2}$, pero esto es imposible: entonces m dividiría a $s - \alpha$, y en consecuencia $s - \alpha = m(u + v\alpha)$, luego $-1 = mv$, contradicción.

Esto significa que podemos obtener una sucesión de ideales similares con normas estrictamente decrecientes mientras los cuadrados de las normas superen a $|\Delta|/3$. Como las normas son números naturales, tras un número finito de pasos hemos de llegar a un ideal $\mathfrak{b}$ similar al de partida y tal que $\mathrm{N}(\mathfrak{b})^2 \le |\Delta|/3$. ∎

Teorema 13.7 *Sea d un entero libre de cuadrados y $\mathcal{O}$ el anillo de enteros de $\mathbb{Q}(\sqrt{d})$. Entonces el grupo de clases de $\mathcal{O}$ es un grupo abeliano finito.*

Demostración: Todo ideal fraccional es congruente con un ideal, y todo ideal es similar a un ideal de norma menor o igual que el menor natural cuyo cuadrado supera a $|\Delta|/3$. Por el teorema 12.11 sólo hay un número finito de ideales con una norma dada, luego tenemos que todo ideal fraccional es similar a un ideal de entre los elementos de un conjunto finito. Por lo tanto sólo puede haber un número finito de clases de similitud, y el grupo cociente es finito. ∎

Sólo con la ayuda de este teorema podemos probar que $h = 1$ en los casos

$$d = -1, \quad -2, \quad -3, \quad -7, \quad -11, \quad -19, \quad 2, \quad 3, \quad 5, \quad 13,$$

y que $h = 2$ en los casos $d = -5,\ -6,\ -10,\ -13,\ -15$.

Veamos algunos ejemplos: si $d = -1$ tenemos $\Delta = -4$, luego todo ideal es similar a uno de norma a tal que $a^2 \le 4/3$, luego $a = 1$. Así pues todo ideal es similar a 1 y por lo tanto $h = 1$. Esto ya lo sabíamos, porque $\mathbb{Z}[i]$ es euclídeo.

Si $d = -5$ entonces $\Delta = -20$ y la cota es $a^2 \le 20/3$, luego $a \le 2$. En $\mathbb{Q}(\sqrt{-5})$ el 2 se ramifica, luego sólo hay dos ideales de norma menor o igual que 2, a saber, 1 y $[2, 1]$. Por lo tanto $h \le 2$. En el capítulo XI vimos que $\mathbb{Q}(\sqrt{-5})$ no tiene factorización única, luego $h \ne 1$. Por lo tanto $h = 2$.

Si $d = -10$ tenemos $a^2 \le 40/3$, luego $a \le 3$. De nuevo 2 se ramifica y 3 se conserva primo, luego no hay ideales de norma 3. Tenemos sólo dos ideales y así $h \le 2$. Como la factorización no es única $h = 2$.

Más interesante es el caso $d = -19$. Tenemos que $a^2 \le 19/3$, luego $a \le 2$, es decir, todo ideal es similar a uno de norma menor o igual que 2, pero 2 se conserva primo en $\mathbb{Q}(\sqrt{-19})$, luego no hay ideales de norma 2. Así pues todo ideal es similar a 1, o sea $\mathbb{Q}(\sqrt{-19})$ tiene factorización única y por otro lado sabemos que no es euclídeo. Es el primer ejemplo que tenemos de DIP no euclídeo.

13.3 Cálculo del número de clases

Ahora veremos que refinando los argumentos de la sección anterior podemos determinar el número de clases de cualquier cuerpo cuadrático. Trataremos por separado los cuerpos imaginarios y los reales. Como siempre, los primeros son más sencillos.

13.3.1 Cuerpos cuadráticos imaginarios

Sea $\mathfrak{a}_i$ un ideal no divisible entre enteros distintos de ± 1. Sea $a_i = \mathrm{N}(\mathfrak{a}_i)$. En la prueba del teorema 13.6 partimos de un número entero o semientero r_i tal que

$$r_i - \frac{\sqrt{\Delta}}{2} \in \mathfrak{a}_i, \qquad |r_i| \le \frac{a_i}{2}.$$

Desde aquí encontramos otro ideal $\mathfrak{a}_{i+1} \approx \mathfrak{a}_i$ (no divisible entre enteros no unitarios) determinado por

$$\overline{\mathfrak{a}}_{i+1}\mathfrak{a}_i = r_i - \frac{\sqrt{\Delta}}{2}. \tag{13.2}$$

Llamemos $a_{i+1} = \mathrm{N}(\mathfrak{a}_{i+1})$. Como $\overline{\mathfrak{a}}_{i+1} \mid r_i - \frac{\sqrt{\Delta}}{2}$, conjugando $\mathfrak{a}_{i+1} \mid r_i + \frac{\sqrt{\Delta}}{2}$, es decir, $-r_i - \frac{\sqrt{\Delta}}{2} \in \mathfrak{a}_{i+1}$. Por lo tanto, si llamamos r_{i+1} a la reducción de $-r_i$ módulo a_{i+1} de modo que $|r_{i+1}| \le (1/2)a_{i+1}$, resulta que $(\mathfrak{a}_{i+1}, r_{i+1})$ vuelve a estar en las hipótesis de 13.6.

Tomando normas en la igualdad (13.2) obtenemos que $a_{i+1}a_i = r_i^2 - \Delta/4$, por lo que en total tenemos:

$$a_{i+1} = \frac{r_i^2 - \Delta/4}{a_i}, \qquad r_{i+1} \equiv -r_i \ (\text{mód } a_{i+1}), \quad |r_{i+1}| \le \frac{a_{i+1}}{2}. \tag{13.3}$$

Estas fórmulas nos permiten calcular los sucesivos a_i y r_i sin necesidad de calcular lo ideales $\mathfrak{a}_i$.

Si alguno de los a_i toma el valor 1, entonces el ideal correspondiente $\mathfrak{a}_i$ será (1), luego el ideal de partida $\mathfrak{a}_0$ será principal.

Ahora vamos a probar que el recíproco es cierto, es decir, que si el ideal de partida es principal, se alcanza el valor $a_i = 1$ para algún i. Más aún, probaremos que si la sucesión de los a_i deja de ser estrictamente decreciente sin alcanzar el valor 1, entonces nunca toma el valor 1, con lo que en un número finito de pasos sabremos siempre si el ideal de partida es o no principal.

En efecto, supongamos que $\mathfrak{a}_0$ es principal. Sea i el menor natural tal que $a_{i+1} \ge a_i$. Según la prueba del teorema 13.6, esto implica que $a_i^2 \le |\Delta|/3$.

Sea $\mathfrak{a}_i = \big(u + v\sqrt{d}\,\big)$, donde u, v son enteros o semienteros. Tenemos, pues, que

$$(u^2 - v^2 d)^2 \le \frac{|\Delta|}{3}.$$

En el caso $d \not\equiv 1 \pmod 4$ tenemos que u y v son enteros y

$$(u^2 - v^2 d)^2 \leq \frac{4|d|}{3}.$$

En el caso $d \equiv 1 \pmod 4$ multiplicamos y dividimos entre 16 y queda

$$\frac{1}{16}\big((2u)^2 - (2v)^2 d\big)^2 \leq \frac{|d|}{3}.$$

Si $v \neq 0$ tenemos en el primer caso $d^2 \leq 4|d|/3$, luego $|d| \leq 4/3$, $d = -1$ y entonces $a_i^2 \leq 4/3$, luego $a_i = 1$ y por lo tanto $\mathfrak{a}_i = 1$.

En el segundo caso queda $d^2 \leq 16|d|/3$, luego $|d| \leq 16/3$ y, puesto que $d \equiv 1 \pmod 4$, la única posibilidad es $d = -3$, luego $a_i^2 \leq 1$ y así $\mathfrak{a}_i = 1$, como antes.

Por otra parte, si $v = 0$ nos queda $\mathfrak{a}_i = (u)$, luego u ha de ser entero y, como $\mathfrak{a}_i$ no es divisible entre enteros no unitarios, $\mathfrak{a}_i = 1$.

Hemos probado que si $\mathfrak{a}_0$ es principal entonces la sucesión de los a_i decrece hasta llegar a 1, luego si deja de decrecer antes de llegar a 1 es que el ideal de partida no es principal.

Si nos fijamos todavía más en el proceso podemos obtener un generador del ideal de partida cuando es principal: Supongamos que tenemos un generador de $\mathfrak{a}_{i+1} = (u)$. De (13.2) deducimos que $(u)\left(r_i - \frac{\sqrt{\Delta}}{2}\right) = \mathfrak{a}_{i+1}\overline{\mathfrak{a}}_{i+1}\mathfrak{a}_i = a_{i+1}\mathfrak{a}_i$, y por lo tanto

$$\mathfrak{a}_i = \left(u\frac{r_i - \sqrt{\Delta}/2}{a_{i+1}}\right).$$

Como un generador del último ideal es 1, resulta que un generador del ideal de partida $\mathfrak{a}_0$ viene dado por

$$u = \frac{\left(r_0 - \sqrt{\Delta}/2\right)\cdots\left(r_{i-1} - \sqrt{\Delta}/2\right)}{a_1 \cdots a_i}, \tag{13.4}$$

donde i es el menor índice que cumple $a_i = 1$.

Ejemplos El primer caso que no podíamos abordar directamente con el teorema 13.6 es $d = -14$. El teorema nos da que todo ideal de $\mathbb{Z}\big[\sqrt{-14}\,\big]$ es similar a uno de norma menor o igual que 4.

Los primos menores que 4 son $2 = [2,0]^2$ y $3 = [3,1][3,-1]$. Todo ideal de norma menor que cuatro es producto de éstos, luego tenemos las siguientes posibilidades:

$$(1), \quad [2,0], \quad [3,1], \quad [3,-1].$$

Faltaría $[2,0]^2$, pero es principal, luego similar a (1). Esto nos da que el número de clases es $h \leq 4$.

En primer lugar, el algoritmo que hemos obtenido nos da que ninguno de los tres ideales distintos de (1) es principal. En efecto, para $[2,0]$ tenemos que $a_0 = 2$ y $r_0 = 0$. Entonces $a_1 = (0^2 + 14)/2 = 7$, y como es mayor que a_0,

resulta que $[2,0]$ no es principal. Para $[3,1]$ y $[3,-1]$ obtenemos $a_1 = 5$, luego tampoco son principales. El ideal $[2,0]^2 = (2)$ es principal. Veamos qué ocurre con $[3,1]^2$.

En este caso $a_0 = 9$ y hemos de hallar r_0. Sabemos que $[3,1] \mid 1 - \sqrt{-14}$, por lo que $[3,1]^2 \mid \left(1 - \sqrt{-14}\right)^2 = -13 - 2\sqrt{-14}$ y el inverso de 2 módulo 9 es -4, luego tenemos $[3,1]^2 \mid 52 - \sqrt{-14}$. Al reducir 52 módulo 9 llegamos a que $\sqrt{-14} \equiv -2 \pmod{[3,1]^2}$.

Así pues, $r_0 = -2$. El algoritmo nos da: $(9,-2), (2,0), (7,*)$, luego $[3,1]^2$ no es principal. Además hemos obtenido que $[3,1]^2$ es similar a un ideal de norma 2, y el único posible es $[2,0]$, luego $[3,1]^2 \approx [2,0]$.

Como el orden de $[2,0]$ en el grupo de las clases es 2, concluimos que el orden de $[3,1]$ ha de ser 4, luego el grupo tiene orden $h = 4$. Además tenemos su estructura, se trata de un grupo cíclico generado por $[3,1]$, y sus elementos son

$$(1), \quad [3,1], \quad [3,1]^2 \approx [2,0], \quad [3,1]^3 \approx [3,-1].$$

■

Veamos ahora un ejemplo más sofisticado. Consideremos el anillo $\mathbb{Z}\left[\sqrt{-74}\,\right]$.

Por el teorema 13.6 todo ideal es similar a uno de norma ≤ 9. Los primos menores que 9 son:

$$2 = [2,0]^2, \quad 3 = [3,1][3,-1], \quad 5 = [5,1][5,-1], \quad 7.$$

Como 7 se conserva tiene norma 14, luego lo eliminamos. Los ideales con norma menor o igual que 9 son:

$$(1), \quad [2,0], \quad [3,1], \quad [3,-1], \quad [5,1], \quad [5,-1], \quad 2,$$

$$[2,0][3,1], \quad [2,0][3,-1], \quad [3,1]^2, \quad [3,-1]^2, \quad 3.$$

Si eliminamos los principales no triviales nos quedan 10. Por lo tanto $h \leq 10$. El ideal $[2,0]$ no es principal, pues el algoritmo nos da $(2,0)$, $(37,*)$. Por lo tanto tiene orden 2 en el grupo de las clases.

Para el ideal $[3,1]$ tenemos $(3,1)$, $(25,*)$, luego tampoco es principal.

Consideramos $[3,1]^2$, para el que $a_0 = 9$. Para calcular r_0 partimos de que $[3,1] \mid 1 - \sqrt{-74}$, luego $[3,1]^2 \mid \left(1 - \sqrt{-74}\right)^2 = -73 - 2\sqrt{-74}$, y en consecuencia $2\sqrt{-74} \equiv 1 \pmod{[3,1]^2}$. El inverso de 2 módulo 9 es -4, luego tenemos que $\sqrt{-74} \equiv -4 \pmod{[3,1]^2}$, o sea, $r_0 = -4$. El algoritmo nos da: $(9,-4), (10,*)$, luego $[3,1]^2$ no es principal.

Calculamos $[3,1]^3$. En este caso $a_0 = 27$. Además

$$\left(1 - \sqrt{-74}\right)^3 \equiv -5 + 17\sqrt{-74} \pmod{27}.$$

El inverso de 17 módulo 27 es 17^{17} $(17 = \phi(27) - 1)$, que reducido es 8, luego r_0 se obtiene reduciendo a 40 (mód 27), es decir, $r_0 = 13$. Así: $(27,13), (9,-4), (10,*)$. Esto nos dice que $[3,1]^3$ es similar a un ideal no principal de norma 9. Los únicos candidatos son $[3,1]^2$ y $[3,-1]^2$, pero $[3,1]^2$ es

imposible: si $[3,1]^3 \approx [3,1]^2$, entonces $[3,1] \approx (1)$. Así pues $[3,1]^3 \approx [3,-1]^2$ y el orden de $[3,1]$ es al menos 4, pero $[3,1]^4 \approx [3,-1]^2[3,1] = (3)[3,-1] \approx [3,-1]$, que tampoco es principal (o lo sería $[3,-1]^2$).

Finalmente, $[3,1]^5 \approx [3,-1][3,1] = (3) \approx (1)$, lo que nos permite concluir que el orden de $[3,1]$ en el grupo de las clases es exactamente 5. Por el teorema de Lagrange, $5 \mid h \leq 10$, luego las únicas posibilidades para h son 5 y 10. Como $[2,0]$ tiene orden 2, ha de ser $h = 10$, luego los 10 ideales de nuestra lista son no similares dos a dos. Como $[3,1]$ tiene orden 5 y $[2,0]$ tiene orden 2, la teoría de grupos nos da que $[3,1][2,0]$ es un generador del grupo. ∎

Ejercicio: Identificar en la lista las diez potencias del ideal $[3,1][2,0]$.

Una aplicación distinta es encontrar generadores de ideales principales, Por ejemplo, el primo 541 se escinde en $\mathbb{Z}[i]$, pues una raíz cuadrada de -1 módulo 541 es 52, (puede hallarse a partir de que 2 es una raíz primitiva de la unidad módulo 541, luego $[-1] = [2^{540/2}]$ y una raíz cuadrada es $[2^{540/4}] = [52]$). Así tenemos el primo $[541, 52]$, que ha de ser un ideal principal. Calculamos $(541,52), (5,-2), (1,*)$, luego, de acuerdo con (13.4), un generador es

$$\frac{(52-i)(-2-i)}{5 \cdot 1} = -21 - 10i.$$

Por lo tanto $541 = (-21-10i)(-21+10i) = (21+10i)(21-10i)$. De este modo hemos obtenido la representación de $541 = 21^2 + 10^2$ como suma de dos cuadrados. ∎

En el capítulo anterior usamos la factorización única de $\mathbb{Q}\big(\sqrt{-163}\,\big)$. Ahora es fácil justificarla: hay que considerar ideales de norma ≤ 7 y los primos 2, 3, 5 y 7 se conservan, luego todos los ideales a considerar son principales.

Ejercicio: Probar que para los valores $d = -43$ y -67 el número de clases es $h = 1$.

Veamos ahora un ejemplo con $d \equiv 1 \pmod 4$. Tomemos $d = -47$. Hay que considerar ideales con normas ≤ 3. Tenemos $2 = [2,0][2,1]$ y $3 = [3,0][3,1]$. Los ideales posibles son (1), $[2,0]$, $[2,1]$, $[3,0]$, $[3,1]$.

Para $[2,0]$ tenemos que $\frac{1+\sqrt{-47}}{2} \equiv 0 \pmod{[2,0]}$, luego cambiando el signo $r_0 = -1/2$. El algoritmo nos da $(2,-1/2), (6,*)$, y $[2,0]$ no es principal.

Consideramos ahora $[2,0]^2$. Se cumple que $\left(\frac{1+\sqrt{-47}}{2}\right)^2 \equiv 0 \pmod{[2,0]^2}$, es decir, $\frac{23-\sqrt{-47}}{2} \equiv 0 \pmod{[2,0]^2}$, luego, reduciendo $23/2$ módulo 4, queda $r_0 = -1/2$. El algoritmo da $(4,-1/2), (3,1/2), (4,*)$, luego no es principal y es similar a un ideal de norma 3 con $r = 1/2$. Veamos cuál de los dos es.

Para $[3,0]$ tenemos que $\frac{1-\sqrt{-47}}{2} \equiv 0 \pmod{[3,0]}$ y $r = 1/2$.

Para $[3,1]$ tenemos que $\frac{-1-\sqrt{-47}}{2} \equiv 0 \pmod{[3,1]}$ y $r = -1/2$.

Por lo tanto $[2,0]^2 \approx [3,-1]$ y de aquí $[2,0]^3 \approx [3,-1][2,0]$.

Tabla 13.1: Número de clases de cuerpos cuadráticos imaginarios

(Los números en negrita son los que cumplen $d \equiv 1 \pmod 4$.)

d	h	d	h	d	h	d	h
-1	1	-26	6	-53	6	-78	4
-2	1	-29	6	$\mathbf{-55}$	**4**	$\mathbf{-79}$	**5**
$\mathbf{-3}$	**1**	-30	4	-57	4	-82	4
-5	2	$\mathbf{-31}$	**3**	-58	2	$\mathbf{-83}$	**3**
-6	2	-33	4	$\mathbf{-59}$	**3**	-85	4
$\mathbf{-7}$	**1**	-34	4	-61	6	-86	10
-10	2	$\mathbf{-35}$	**2**	-62	8	$\mathbf{-87}$	**6**
$\mathbf{-11}$	**1**	-37	2	-65	8	-89	12
-13	2	-38	6	-66	8	$\mathbf{-91}$	**2**
-14	4	$\mathbf{-39}$	**4**	$\mathbf{-67}$	**1**	-93	4
$\mathbf{-15}$	**2**	-41	8	-69	8	-94	8
-17	4	-42	4	-70	4	$\mathbf{-95}$	**8**
$\mathbf{-19}$	**1**	$\mathbf{-43}$	**1**	$\mathbf{-71}$	**7**	-97	4
-21	4	-46	4	-73	4		
-22	2	$\mathbf{-47}$	**4**	-74	10		
$\mathbf{-23}$	**3**	$\mathbf{-51}$	**2**	-77	8		

Tenemos que $\frac{1+\sqrt{-47}}{2} \equiv 0 \pmod{[3,0]}$ y $\frac{1+\sqrt{-47}}{2} \equiv 0 \pmod{[2,0]}$, luego también $\frac{1+\sqrt{-47}}{2} \equiv 0 \pmod{[3,0][2,0]}$. Así pues, $r_0 = -1/2$, con lo que la sucesión del algoritmo es $(6,-1/2),(2,1/2)$, de donde $[2,0]^3 \approx [2,1]$ y

$$[2,0]^4 \approx [2,1][2,0] \approx (1).$$

En conclusión, la clase del ideal $[2,0]$ tiene orden 4 en el grupo de las clases, con lo que tenemos que $4 \mid h \leq 5$, luego $h = 4$ y los representantes de las clases son

$$(1),\ [2,0],\ [3,0],\ [2,1].$$

El ideal sobrante, $[3,1]$, es similar a $[3,0]$. En efecto, de $[3,0] \approx [2,0]^2$ obtenemos conjugando que $[3,1] \approx [2,1]^2 \approx [2,0]^6 \approx [2,0]^2 \approx [3,0]$. ∎

La tabla 13.1 recoge los números de clases h de todos los cuerpos cuadráticos con $-100 \leq d \leq -1$. El lector puede comprobar los valores que desee. Obsérvese que el comportamiento de h es extremadamente irregular.

13.3.2 Cuerpos cuadráticos reales

El caso real presenta algunas dificultades adicionales. Partimos de un ideal $\mathfrak{a}$ no divisible entre enteros no unitarios. Aplicando el algoritmo que hemos visto para el caso $d < 0$ llegamos a un ideal $\mathfrak{a}_{i+1}$ en las mismas condiciones con la propiedad de que $a_{i+1}^2 = \mathrm{N}(\mathfrak{a}_{i+1})^2 < \Delta/3$.

A partir de este momento cambiaremos el criterio para elegir r módulo a. Tomaremos r_{i+1} de manera que $r_{i+1} \equiv -r_i \pmod{a_{i+1}}$ y r_{i+1} sea el máximo que cumple $r_{i+1}^2 < \Delta/4$. En definitiva, continuaremos la sucesión de ideales de acuerdo con las fórmulas:

$$\overline{\mathfrak{a}}_{i+1}\mathfrak{a}_i = r_i - \frac{\sqrt{\Delta}}{2}. \tag{13.5}$$

$$a_{i+1} = \frac{\Delta/4 - r_i^2}{a_i}, \qquad r_{i+1} \equiv -r_i \pmod{a_{i+1}}, \quad r_{i+1}^2 < \Delta/4 \quad \text{(máximo)}. \tag{13.6}$$

(Notemos que la primera ecuación de (13.6) procede de tomar normas en (13.5), y en el segundo miembro hay que poner un valor absoluto o, equivalentemente, hay que garantizar que sea positivo. Ésa es la razón del cambio de orden respecto a (13.3).)

Hemos de probar que siempre es posible encontrar un r_{i+1} en las condiciones de (13.6).

La primera vez es posible porque tenemos $a_{i+1}^2 < \Delta/3$ (obtenido por el criterio anterior) y existe un entero o semientero r que cumple $r \equiv -r_i \pmod{a_{i+1}}$ y $|r| \le a_{i+1}/2$ (el que eligiríamos con el criterio anterior). Este r cumple también $r^2 \le a_{i+1}^2/4 \le \Delta/12 < \Delta/4$. Sólo hay que cambiarlo por el máximo posible y ya tenemos r_{i+1}.

Sin embargo, a partir de aquí a_{i+1} ya no cumple necesariamente $a_{i+1}^2 < \Delta/3$, luego no tenemos asegurado que podamos encontrar r_{i+1} en las condiciones requeridas. Vamos a probarlo.

Por la maximalidad de r_i tenemos que $(r_i + a_i)^2 > \Delta/4$ (la igualdad no puede darse porque d es libre de cuadrados). De aquí

$$r_i^2 - \Delta/4 + 2r_ia_i + a_i^2 > 0,\ a_i\left(\frac{r_i^2 - \Delta/4}{a_i} + 2r_i + a_i\right) > 0,\ -a_{i+1} + 2r_i + a_i > 0$$

y multiplicando por $-a_{i+1}$ queda

$$a_{i+1}^2 - 2r_ia_{i+1} - a_ia_{i+1} < 0, \quad a_{i+1}^2 - 2r_ia_{i+1} + r_i^2 - \Delta/4 < 0,$$

luego $(a_{i+1} - r_i)^2 < \Delta/4$.

Así pues, cambiando $a_{i+1} - r_i$ por el máximo posible módulo a_{i+1}, encontramos un r_{i+1} que cumple (13.6).

Más aún, hemos visto que $a_{i+1} - r_i \le r_{i+1}$, luego $-r_i \le r_{i+1} - a_{i+1} \le r_{i+1}$. Puesto que tanto $r_{i+1}^2 < \Delta/4$ como $r_i^2 < \Delta/4$, esto implica $(r_{i+1} - a_{i+1})^2 < \Delta/4$, mientras que por definición de r_{i+1} tenemos $(r_{i+1} + a_{i+1})^2 > \Delta/4$, es decir, $(r_{i+1} - a_{i+1})^2 < (r_{i+1} + a_{i+1})^2$.

Desarrollando los cuadrados queda $-r_{i+1} < r_{i+1}$, luego $r_{i+1} > 0$, es decir, a partir del momento en que usamos el nuevo criterio, los r_i serán siempre positivos.

Ahora notamos que $a_{i+1} \le a_ia_{i+1} = \Delta/4 - r_i^2 < \Delta/4$, luego los ideales $\mathfrak{a}_i$ recorren un conjunto finito de ideales posibles. Necesariamente la sucesión ha de entrar en un ciclo.

Para garantizar que este algoritmo nos permite decidir siempre si el ideal de partida es principal basta probar que si es así entonces el ideal 1 aparece en el ciclo. Podemos suponer que $\mathfrak{a}_0$ es el comienzo del ciclo, de modo que para un k (mínimo) se cumple $\mathfrak{a}_0 = \mathfrak{a}_k$.

Supongamos que $\mathfrak{a}_0 = \big(x_0 + y_0\sqrt{d}\,\big)$, donde los coeficientes son enteros o semienteros. Por construcción, $\overline{\mathfrak{a}}_0\mathfrak{a}_1 = (r_0 + \sqrt{\Delta}/2)$, luego al multiplicar queda

$$\big(x_0 + y_0\sqrt{d}\,\big)\big(r_0 + \sqrt{\Delta}/2\big) = a_0\mathfrak{a}_1,$$

y de este modo $\mathfrak{a}_1 = (x_1 + y_1\sqrt{d})$, donde

$$x_1 + y_1\sqrt{d} = \big(x_0 + y_0\sqrt{d}\,\big)\frac{r_0 + \sqrt{\Delta}/2}{a_0}.$$

Repitiendo k veces llegamos a que $\mathfrak{a}_k = \big(x_k + y_k\sqrt{d}\,\big)$, donde

$$x_k + y_k\sqrt{d} = \big(x_0 + y_0\sqrt{d}\,\big)\frac{(r_0 + \sqrt{\Delta}/2)\cdots(r_{k-1} + \sqrt{\Delta}/2)}{a_0\cdots a_{k-1}},$$

y como $\big(x_0 + y_0\sqrt{d}\,\big) = \big(x_k + y_k\sqrt{d}\,\big)$, se ha de cumplir que el número

$$\epsilon = \frac{(r_0 + \sqrt{\Delta}/2)\cdots(r_{k-1} + \sqrt{\Delta}/2)}{a_0\cdots a_{k-1}} \tag{13.7}$$

es una unidad.

En general se cumple que

$$\mathfrak{a}_{j+nk} = \big(x_{j+nk} + y_{j+nk}\sqrt{d}\,\big), \quad \text{donde} \quad x_{j+nk} + y_{j+nk}\sqrt{d} = \big(x_j + y_j\sqrt{d}\,\big)\epsilon^n.$$

Definamos $x_j + y_j\sqrt{d} = \epsilon^{-n}\big(x_{j+nk} + y_{j+nk}\sqrt{d}\,\big)$ para $j < 0$, donde n es un entero suficientemente grande para que $j + nk$ sea positivo. Sea igualmente $\mathfrak{a}_j = \big(x_j + y_j\sqrt{d}\,\big)$, de modo que $\mathfrak{a}_{j+nk} = \mathfrak{a}_j$ para todo n y todo j (observemos que los ideales $\mathfrak{a}_j$ con $j < 0$ no son necesariamente los ideales obtenidos mediante el algoritmo antes de llegar al ciclo que comienza con $\mathfrak{a}_0$).

Sea $\epsilon = u + v\sqrt{d}$. Teniendo en cuenta la expresión de ϵ y que todos los r_j son positivos, es claro que $u, v > 0$. En particular $\epsilon \neq \pm 1$. Sea $\epsilon^n = u_n + v_n\sqrt{d}$. Los enteros (o semienteros) u_n y v_n satisfacen las relaciones:

$$u_{n+1} = uv_n + dvv_n \tag{13.8}$$

$$v_{n+1} = vu_n + uv_n \tag{13.9}$$

Estas ecuaciones muestran que u_n, $v_n > 0$, así como que ambos se hacen arbitrariamente grandes cuando n crece. (El caso menos obvio se da si u y v son semienteros. Entonces $d > 4$ y (13.8) implica $u_{n+1} > 2v_n$, a su vez (13.9) implica $v_{n+1} \geq u_n/2$ y entonces $v_{n+2} > u_{n+1}/2 > v_n$.)

Tenemos que $x_{nk} + y_{nk}\sqrt{d} = \big(x_0 + y_0\sqrt{d}\,\big)\epsilon^n$, luego

$$x_{nk} = u_nx_0 + dv_ny_0, \qquad y_{nk} = v_nx_0 + u_ny_0.$$

Supongamos que x_0 e y_0 son no nulos y tienen signos opuestos. Para saber el signo de x_{nk} restamos los cuadrados de los dos sumandos (y tenemos en cuenta que $\mathrm{N}(\epsilon^n) = \pm 1$).

$$u_n^2 x_0^2 - d^2 v_n^2 y_0^2 = (\pm 1 + dv_n^2)x_0^2 - d^2 v_n^2 y_0^2 = \pm x_0^2 + dv_n^2(x_0^2 - dy_0^2).$$

Para valores de n lo suficientemente grandes como para que $dv_n^2(x_0^2 - dy_0^2)$ supere en módulo a $\pm x_0^2$ (lo cual siempre acaba ocurriendo porque v_n se hace cada vez mayor) el signo de esta expresión es el de $x_0^2 - dy_0^2$, luego el signo de x_{nk} será el de x_0 o el de y_0 según si $x_0^2 - dy_0^2$ es mayor o menor que 0.

El cálculo análogo para y_{nk} es el siguiente:

$$v_n^2 x_0^2 - u_n^2 y_0^2 = v_n^2 x_0^2 - (\pm 1 + dv_n^2)y_0^2 = \pm y_0^2 + v_n^2(x_0^2 - dy_0^2).$$

De aquí deducimos que para valores grandes de n el signo de y_{nk} es también el de x_0 o el de y_0 según si $x_0^2 - dy_0^2$ es mayor o menor que 0. En resumen, que si x_0 e y_0 son no nulos y tienen signos opuestos, para valores grandes de n se cumple que x_{nk} e y_{nk} tienen el mismo signo.

Supongamos ahora que x_0 e y_0 son no nulos y tienen el mismo signo. Lo que hemos probado antes es que si el signo es distinto, entonces los coeficientes de $\big(x_0 + y_0\sqrt{d}\,\big)\epsilon^n$ tienen el mismo signo si n es suficientemente grande, luego en nuestro caso los coeficientes de $\big(x_0 - y_0\sqrt{d}\,\big)\epsilon^n$ tienen el mismo signo cuando n es grande.

Por definición, $x^{-nk} + y^{-nk}\sqrt{d} = \big(x_0 + y_0\sqrt{d}\,\big)\epsilon^{-n}$. Conjugamos teniendo en cuenta que $\mathrm{N}(\epsilon) = \epsilon\overline{\epsilon} = \pm 1$ (luego $\epsilon^{-1} = \pm\overline{\epsilon}$) y obtenemos $x_{-nk} - y_{-nk}\sqrt{d} = \pm\big(x_0 - y_0\sqrt{d}\,\big)\epsilon^n$. Así pues, resulta que x_{-nk} y $-y_{-nk}$ tienen el mismo signo cuando n es grande, con lo que x_{-nk} e y_{-nk} tienen signos opuestos.

Con todo esto hemos probado que o bien hay un índice j para el que $x_j y_j = 0$ o bien $x_j y_j$ toma signos distintos en distintos j's. En este último caso ha de haber un entero j tal que $x_j y_j < 0$ y $x_{j+1}y_{j+1} > 0$. Vamos a demostrar que esto es imposible.

Tenemos que $x_{j+1} + y_{j+1}\sqrt{d} = \big(x_j + y_j\sqrt{d}\,\big)\big(r_j + \sqrt{d}\,\big)/a_j$, y por otra parte $a_j = \Big|\big(x_j + y_j\sqrt{d}\,\big)\big(x_j - y_j\sqrt{d}\,\big)\Big|$. Por lo tanto

$$\big(x_{j+1} + y_{j+1}\sqrt{d}\,\big)\big(x_j - y_j\sqrt{d}\,\big) = \pm\big(r_j + \sqrt{d}\,\big).$$

Igualando los coeficientes obtenemos las ecuaciones siguientes:

$$\begin{aligned} x_{j+1}x_j - dy_{j+1}y_j &= \pm r_j, && (13.10)\\ x_j y_{j+1} - x_{j+1}y_j &= \pm 1. && (13.11) \end{aligned}$$

Si suponemos $x_j y_j < 0$ y $x_{j+1}y_{j+1} > 0$ se convierten en

$$\begin{aligned} |x_{j+1}x_j| + d|y_{j+1}y_j| &= r_j,\\ |x_j y_{j+1}| + |x_{j+1}y_j| &= 1. \end{aligned}$$

Obviamente la segunda ecuación es imposible si las variables son enteros no nulos. Si son semienteros multiplicamos las ecuaciones por 4 y queda

$$\begin{aligned} |2x_{j+1}|\,|2x_j| + d|2y_{j+1}|\,|2y_j| &= 4r_j,\\ |2x_j|\,|2y_{j+1}| + |2x_{j+1}|\,|2y_j| &= 4. \end{aligned}$$

Si, por ejemplo x_j es semientero (luego y_j también), entonces $|2x_j|$, $|2y_j|$ son congruentes con ± 1 módulo 4. Multiplicamos la primera ecuación por $|2x_j|$, tomamos congruencias módulo 4, usamos la segunda ecuación (también como congruencia) y nos queda que $\pm 4r_j \equiv |2x_j|^2|2x_{j+1}| - |2y_j|^2|2x_{j+1}| \equiv 0$ (mód 4), luego r_j es entero, pero esto es una contradicción, pues ha de ser semientero.

En consecuencia existe un entero j para el que $x_j y_j = 0$, pero no puede ser $x_j = 0$ ya que (13.10) nos da entonces $dy_{j+1}y_j = \pm r_j$, lo que contradice la condición $|r_j|^2 < \Delta/4$ (notemos que en cualquier caso y_j es entero).

Por lo tanto $y_j = 0$, y como $\mathfrak{a}_j = (x_j)$ no es divisible entre enteros no unitarios, ha de ser $x_j = \pm 1$ y, cualquiera que sea j, el ideal $\mathfrak{a}_j = 1$ es igual a uno del ciclo $0, \ldots, k$, como queríamos probar.

Notemos que la fórmula (13.4) para calcular el generador de un ideal principal sigue siendo válida en este caso.

Ejemplo Vamos a calcular el número de clases para $d = 79$. Todo ideal es similar a uno de norma menor o igual que 10. Los primos se comportan como sigue:

$$2 = [2,1]^2, \quad 3 = [3,1][3,-1], \quad 5 = [5,2][5,-2], \quad 7 = [7,3][7,-3].$$

Los ideales posibles son

$$(1), \quad [2,1], \quad [3,1], \quad [3,-1], \quad [5,2], \quad [5,-2], \quad [7,3], \quad [7,-3],$$

$$[2,1][3,1], \quad [2,1][3,-1], \quad [2,1][5,2], \quad [2,1][5,-2], \quad [3,1]^2, \quad [3,-1]^2.$$

Partiendo de $[2,1]$, tenemos que $a_0 = 2$ y r_0 es el mayor número congruente con 1 módulo 2 cuyo cuadrado no supera a 79, o sea, $r_0 = 7$. Aplicamos el algoritmo: $(2,7), (15,8), (1,*)$, luego resulta que $[2,1]$ es principal (notemos que la sucesión de las normas ya no es decreciente).

Ahora consideramos el ideal $[3,1]$. Para éste $a_0 = 3$ y r_0 es el mayor número congruente con 1 módulo 3 cuyo cuadrado no supera a 79, o sea, $r_0 = 7$. Ahora tenemos $(3,7), (10,3), (7,4), (9,5), (6,7), (5,8), (3,7)$.

Como entramos en un ciclo, el ideal $[3,1]$ no es principal. Además vemos que $[3,1]$ es similar a ideales de norma 10, 7, 9, 6 y 5, que tienen que ser algunos de los que figuran en nuestra lista. No importa cuáles sean de hecho, pues como son pares de conjugados, el que no es similar a $[3,1]$ lo es a $[3,-1]$, luego es redundante en cualquier caso. Con esto nuestros candidatos se reducen a $(1), [3,1], [3,-1]$.

Sabemos que $[3,1]$ es similar a un ideal de norma 9, que ha de ser $[3,1]^2$ o bien $[3,-1]^2$, pero si fuera $[3,1]^2 \approx [3,1]$, entonces $[3,1] \approx (1)$, luego podemos

concluir que $[3,1] \approx [3,-1]^2$, y por consiguiente $[3,-1]^2$ no es principal (ni $[3,-1]$, por ser el conjugado de $[3,1]$). Esto nos dice que la clase del ideal $[3,-1]$ tiene orden 3 en el grupo de clases, luego los tres ideales (1), $[3,1]$ y $[3,-1]$ no son similares entre sí, y el número de clases es $h = 3$. ∎

No damos más ejemplos porque las diferencias con el caso $d < 0$ son mínimas en la práctica. Veamos ahora que el algoritmo que hemos dado nos permite también determinar las unidades de los anillos de enteros cuadráticos reales.

Apliquemos el algoritmo al ideal $\mathfrak{a}_0 = (1)$ (en ningún momento hemos excluido este caso). Como es principal, al cabo de k pasos volveremos a obtener el (1), pero, siguiendo la notación anterior, el generador al que llegaremos $x_k + y_k\sqrt{d}$ no será 1 sino la unidad ϵ dada por (13.7).

Sabemos que $\epsilon = u + v\sqrt{d} \neq \pm 1$ y que $u, v > 0$. De aquí se sigue que las potencias de ϵ son todas distintas. Vamos a probar que las unidades de $\mathbb{Q}\big(\sqrt{d}\big)$ son exactamente $\pm\epsilon^n$, donde n recorre los números enteros.

En efecto, sea $x_0 + y_0\sqrt{d}$ una unidad cualquiera. Al partir del ideal $\mathfrak{a}_0 = \big(x_0 + y_0\sqrt{d}\big)$ se ha de alcanzar $\mathfrak{a}_j = (1)$ para algún j. Como $a_0 = 1$, el valor de r_0 es simplemente el mayor entero cuyo cuadrado no supera a d (la congruencia módulo 1 no es una restricción), luego el algoritmo a partir de (a_0, r_0) continúa exactamente igual a como empieza cuando partíamos de $\mathfrak{a}_0 = (1)$, luego el ciclo de a_j's y r_j's que se produce al partir de $\mathfrak{a}_0 = (1)$ es el mismo que el producido con $\mathfrak{a}_0 = \big(x_0 + y_0\sqrt{d}\big)$, luego el valor de ϵ que da la fórmula (13.7) es el mismo en ambos casos.

Además hemos probado no sólo que $\mathfrak{a}_j = (1)$, sino que para un cierto entero n se tiene además $x_{j+nk} + y_{j+nk}\sqrt{d} = \pm 1$. Por otra parte $x_{j+nk} + y_{j+nk}\sqrt{d} = \epsilon^n\big(x_0 + y_0\sqrt{d}\big)$, luego resulta que $x_0 + y_0\sqrt{d} = \pm\epsilon^{-n}$, como queríamos demostrar.

En particular hemos demostrado el teorema siguiente:

Teorema 13.8 *Sea d un número natural libre de cuadrados $d \neq 1$. Entonces existe una unidad ϵ de $\mathbb{Q}\big(\sqrt{d}\big)$ tal que los elementos $\pm\epsilon^n$ para $n \in \mathbb{Z}$ son distintos dos a dos y constituyen todas las unidades de $\mathbb{Q}\big(\sqrt{d}\big)$. Dicha unidad se llama* unidad fundamental.

Ejemplos Calculemos la unidad fundamental de $\mathbb{Q}\big(\sqrt{2}\big)$. Hemos de partir del ideal (1), luego $a_0 = 1$ y r_0 es el mayor entero cuyo cuadrado no supera a 2, o sea, $r_0 = 1$.

El algoritmo es $(1,1), (1,*)$, luego $\epsilon = \big(1 + \sqrt{2}\big)/1$ es la unidad fundamental. Las demás unidades son las de la forma $\pm\big(1 + \sqrt{2}\big)^n$.

Para $\mathbb{Q}\big(\sqrt{3}\big)$ queda $(1,1), (2,1), (1,*)$, luego

$$\epsilon = \frac{\big(1 + \sqrt{3}\big)\big(1 + \sqrt{3}\big)}{2} = 2 + \sqrt{3}.$$

Para $\mathbb{Q}\big(\sqrt{6}\big)$ resulta $(1,2), (2,2), (1,*)$, luego

$$\epsilon = \frac{\big(2 + \sqrt{6}\big)\big(2 + \sqrt{6}\big)}{2} = 5 + 2\sqrt{6}.$$

Tabla 13.2: Número de clases de cuerpos cuadráticos reales

Se indica también la unidad fundamental ϵ y su norma. Los números en negrita son los que cumplen $d \equiv 1$ (mód 4). El número α es el definido en (13.1).

d	h	ϵ	$\mathrm{N}(\epsilon)$	d	h	ϵ	$\mathrm{N}(\epsilon)$
2	1	$1+\alpha$	-1	**53**	**1**	$3+\alpha$	-1
3	1	$2+\alpha$	$+1$	55	2	$89+12\alpha$	$+1$
5	**1**	α	-1	**57**	**1**	$131+40\alpha$	$+1$
6	1	$5+2\alpha$	$+1$	58	2	$99+13\alpha$	-1
7	1	$8+3\alpha$	$+1$	59	1	$530+69\alpha$	$+1$
10	2	$3+\alpha$	-1	**61**	**1**	$17+5\alpha$	-1
11	1	$10+3\alpha$	$+1$	62	1	$63+8\alpha$	$+1$
13	**1**	$1+\alpha$	-1	**65**	**2**	$7+2\alpha$	-1
14	1	$15+4\alpha$	$+1$	66	2	$65+8\alpha$	$+1$
15	2	$4+\alpha$	$+1$	67	1	$48.842+5.967\alpha$	$+1$
17	**1**	$3+2\alpha$	-1	**69**	**1**	$11+3\alpha$	$+1$
19	1	$170+39\alpha$	$+1$	70	2	$251+30\alpha$	$+1$
21	**1**	$2+\alpha$	$+1$	71	1	$3.480+413\alpha$	$+1$
22	1	$197+42\alpha$	$+1$	**73**	**1**	$943+250\alpha$	-1
23	1	$24+5\alpha$	$+1$	74	2	$43+5\alpha$	-1
26	2	$5+\alpha$	-1	**77**	**1**	$4+\alpha$	$+1$
29	**1**	$2+\alpha$	-1	78	2	$53+6\alpha$	$+1$
30	2	$11+2\alpha$	$+1$	79	3	$80+9\alpha$	$+1$
31	1	$1.520+233\alpha$	$+1$	82	4	$9+\alpha$	-1
33	**1**	$19+8\alpha$	$+1$	83	1	$82+9\alpha$	$+1$
34	2	$35+6\alpha$	$+1$	**85**	**2**	$4+\alpha$	-1
35	2	$6+\alpha$	$+1$	86	1	$10.405+1.122\alpha$	$+1$
37	**1**	$5+2\alpha$	-1	87	2	$28+3\alpha$	$+1$
38	1	$37+6\alpha$	$+1$	**89**	**1**	$447+106\alpha$	-1
39	2	$25+4\alpha$	$+1$	91	2	$1.574+165\alpha$	$+1$
41	**1**	$27+10\alpha$	-1	**93**	**1**	$13+3\alpha$	$+1$
42	2	$13+2\alpha$	$+1$	94	1	$2.143.295+221.064\alpha$	$+1$
43	1	$3.482+531\alpha$	$+1$	95	2	$39+4\alpha$	$+1$
46	1	$24.335+3.588\alpha$	$+1$	**97**	**1**	$5.035+1.138\alpha$	-1
47	1	$48+7\alpha$	$+1$	**101**	**1**	$9+2\alpha$	-1
51	2	$50+7\alpha$	$+1$				

Para $\mathbb{Q}\big(\sqrt{5}\big)$ es $r_0 = 1/2$ y así $(1, 1/2), (1, *)$. Por lo tanto,

$$\epsilon = \frac{1+\sqrt{5}}{2}.$$

Para $\mathbb{Q}(\sqrt{97})$ es $r_0 = 9/2$, luego el algoritmo nos da

$$(1, 9/2), (4, 7/2), (3, 5/2), (6, 7/2), (2, 9/2),$$

$$(2, 7/2), (6, 5/2), (3, 7/2), (4, 9/2), (1, *).$$

La unidad fundamental es $5.604 + 569\sqrt{97}$.

Una aplicación de estos resultados es la solución de la llamada *ecuación de Pell*. Se trata de la ecuación $dy^2 + 1 = x^2$. Fermat planteó a los matemáticos ingleses de su tiempo encontrar un método para hallar las soluciones enteras de dicha ecuación. El nombre de 'ecuación de Pell' se debe a un error de Euler, que atribuyó a Pell su solución, cuando al parecer Pell no hizo ninguna contribución a este tema.

Desde nuestro punto de vista la ecuación equivale a $\mathrm{N}(x+y\sqrt{d}) = 1$. Cuando d es positivo, libre de cuadrados y no es congruente con 1 módulo 4 sabemos que tiene infinitas soluciones, a saber, las unidades de $\mathbb{Z}[\sqrt{d}]$ de norma 1.

Si $d < 0$ es fácil ver que las únicas soluciones son $x = \pm 1$, $y = 0$, luego el caso no tiene interés. También es fácil ver que d no debe ser un cuadrado perfecto, pues si $d = a^2$ una solución de la ecuación de Pell da lugar a la expresión $(ay)^2 + 1 = x^2$, y es fácil ver que dos cuadrados perfectos no pueden diferir en una unidad (salvo $y = 0$, $x = \pm 1$).

Sin embargo, no hay más limitaciones. La ecuación de Pell tiene infinitas soluciones enteras siempre que d sea un número positivo no cuadrado perfecto. No vamos a demostrar esto aquí, pero lo cierto es que las soluciones de la ecuación de Pell $dy^2 + 1 = x^2$ son exactamente los coeficientes de las potencias de la unidad ϵ que se obtiene según el algoritmo visto en este capítulo para el caso $d > 0$, $d \not\equiv 1$ (mód 4) (es decir, tomando $\Delta = 4d$, aun en el caso de que d no cumpla estas hipótesis) si $\mathrm{N}(\epsilon) = 1$, y son los coeficientes de las potencias de ϵ^2 si por el contrario $\mathrm{N}(\epsilon) = -1$. De hecho este algoritmo (conocido como 'método inglés') fue ideado por Wallis en respuesta al reto de Fermat de resolver la ecuación de Pell y de él derivan los demás algoritmos que hemos visto. Una justificación de todo esto desde nuestro punto de vista supondría estudiar los subanillos de los anillos de enteros algebraicos, en lo cual no nos vamos a detener.

Por ejemplo, la unidad fundamental para $d = 97$ es $5.604 + 569\sqrt{97}$, pero tiene norma -1, y su cuadrado es $62.809.633 + 6.377.352\sqrt{97}$, por lo que la menor solución de la ecuación de Pell para $d = 97$ es

$$97(6.377.352)^2 + 1 = 62.809.633^2.$$

No cabe duda de que Fermat conocía la solución del problema que planteó, pues mostró como ejemplos las soluciones para los casos $d = 109$ y $d = 149$, que son nada menos que $y = 15.140.424.455.100$ e $y = 2.113.761.020$, respectivamente. En el siglo XII, el matemático hindú Bháscara Achárya conocía la solución mínima para $d = 61$, que es la más grande para $d \leq 100$. Se trata de $y = 226.153.980$.

Tabla 13.3: Soluciones mínimas de la ecuación de Pell

d	y	d	y	d	y	d	y
1	$*$	26	10	51	7	76	6.630
2	2	27	5	52	90	77	40
3	1	28	24	53	9.100	78	6
4	$*$	29	1.820	54	66	79	9
5	4	30	2	55	12	80	1
6	2	31	273	56	2	81	$*$
7	3	32	3	57	20	82	18
8	1	33	4	58	2.574	83	9
9	$*$	34	6	59	69	84	6
10	6	35	1	60	4	85	30.996
11	3	36	$*$	61	226.153.980	86	1.122
12	2	37	12	62	8	87	3
13	180	38	6	63	1	88	21
14	4	39	4	64	$*$	89	53.000
15	1	41	3	65	16	90	2
16	$*$	41	320	66	8	91	165
17	8	42	2	67	5.967	92	120
18	4	43	531	68	4	93	1.260
19	39	44	30	69	936	94	221.064
20	2	45	24	70	30	95	4
21	12	46	3.588	71	413	96	5
22	42	47	7	72	2	97	6.377.352
23	5	48	1	73	267.000	98	10
24	1	49	$*$	74	430	99	1
25	$*$	50	14	75	3	100	$*$

Ejemplo Vamos a resolver la ecuación de Pell $18y^2 + 1 = x^2$ (notemos que 18 no es libre de cuadrados).

Tenemos: $(1, 4), (2, 4), (1, *)$. Por lo tanto

$$\epsilon = \frac{\left(4 + \sqrt{18}\right)\left(4 + \sqrt{18}\right)}{2} = 17 + 4\sqrt{18}.$$

Como $\mathrm{N}(\epsilon) = 1$, la menor solución es $y = 4$.

La tabla 13.3 contiene las menores soluciones de la ecuación de Pell para los primeros valores de d. Es evidente que los ejemplos mostrados por Fermat no estaban elegidos al azar. El lector emprendedor puede abordar el caso $d = 421$. La solución mínima tiene 33 dígitos.

Capítulo XIV

La ley de reciprocidad cuadrática

14.1 Introducción

En el capítulo anterior hemos obtenido resultados muy profundos sobre la aritmética de los cuerpos cuadráticos, el más importante de los cuales es sin duda el algoritmo que nos permite calcular los números de clases y las unidades fundamentales. Para hacernos una idea de su importancia vamos a considerar un cuerpo cuadrático K que cumpla dos condiciones:

1. Su número de clases de $h = 1$, es decir, $\mathcal{O}_K$ es un DFU.
2. Si K es real entonces tiene unidades de norma negativa.

Un ejemplo es $K = \mathbb{Q}(\sqrt{13})$. Su unidad fundamental es $1 + \frac{1+\sqrt{13}}{2}$ y tiene norma -1 (notar que la existencia de unidades de norma negativa equivale a que la unidad fundamental tenga norma negativa).

En general tenemos que $\mathrm{N}\left(x + y\frac{1+\sqrt{13}}{2}\right) = x^2 + xy - 3y^2$. Consideremos el problema siguiente:

¿Existen números enteros x e y tales que $x^2 + xy - 3y^2 = 15$?

Para alguien que desconozca la teoría que hemos desarrollado se trata de un problema muy difícil, pues el signo negativo delante del 3 hace que la mínima solución, si existe, pueda estar formada por números muy grandes, imposibles de obtener por tanteo.

Sin embargo para nosotros es trivial: la pregunta es si existe un entero de K de norma 15. Dicho entero factorizaría en el producto de un primo de norma 3 por uno de norma 5 (pues los primos cuadráticos han de tener norma potencia de primo), luego en particular ha de haber un primo de norma 5, lo que significa que 5 ha de escindirse en K (no se ramifica porque no divide al discriminante $\Delta = 13$). Ahora bien, sabemos que 5 se ramifica si y sólo si 13 es un resto

cuadrático módulo 5, pero los restos cuadráticos módulo 5 son 1 y 4, luego $[13] = [3]$ no es uno de ellos. Así pues, la ecuación no tiene soluciones enteras.

Si el lector se pregunta dónde hemos usado las dos hipótesis sobre K la respuesta es que en ningún sitio: el argumento anterior es válido para cualquier cuerpo cuadrático. Las hipótesis las necesitamos si queremos resolver casos en los que sí haya solución:

En general, existen enteros cuadráticos de norma un primo p si y sólo si p se escinde o se ramifica en K. Aquí sí hacemos uso de las dos hipótesis: sin ellas sólo podemos decir que existe un ideal $\mathfrak{p}$ de norma p si y sólo si p se escinde o se ramifica en K. Si $h = 1$ se cumplirá además que $\mathfrak{p} = (\pi)$, para cierto primo π, luego podemos concluir que existe un entero cuadrático π tal que $|\,\mathrm{N}(\pi)| = p$ si y sólo si π se escinde o se ramifica. Si K es imaginario todas las normas son positivas, si K es real y hay unidades de norma -1, multiplicando π por una de ellas si es preciso podemos exigir también que $\mathrm{N}(\pi) = p$, con lo que tenemos que hay un entero cuadrático π de norma p si y sólo si p se escinde o se ramifica en K.

Por lo tanto podemos construir enteros cuadráticos de cualquier norma m con la única restricción de que los primos que dividan a m con exponente impar se escindan o ramifiquen en K (pues siempre existen enteros cuadráticos de norma p^2, a saber el propio p). En nuestro caso concreto tenemos:

> *La ecuación $x^2 + xy - 3y^2 = m$ tiene soluciones enteras para un m dado si y sólo si los primos que dividen a m con exponente impar son de la forma $x^2 + xy - 3y^2$, y un primo p es de esta forma si y sólo si $p = 13$ o 13 es un resto cuadrático módulo p.*

Ésta es una solución completamente satisfactoria del problema, pues nos permite saber con un número finito de operaciones si cualquier número m es o no de la forma indicada. Las operaciones consisten en determinar si 13 es o no un resto cuadrático módulo un número finito de primos.

Fijémonos en la condición '13 es un resto cuadrático módulo p'. En el caso general de un cuerpo $K = \mathbb{Q}(\sqrt{d}\,)$ y p un primo impar sería 'd es un resto cuadrático módulo p', y como el discriminante es $\Delta = d$ o $\Delta = 4d$, y ciertamente d es un resto cuadrático módulo p si y sólo si lo es $4d$, podemos expresar la condición como

$$\Delta \text{ es un resto cuadrático módulo } p.$$

Ésta es la condición para que un primo impar que no divida a Δ se escinda en K, y es lo que hay que comprobar para un número finito de primos p a la hora de decidir si un número m es la norma de un entero cuadrático.

En principio, si nos dan cien primos p y determinamos si cumplen o no esta condición, nuestros cálculos no nos ayudarán en nada para decidir si un nuevo primo cumple la condición o no, es decir, el saber si Δ es o no un resto cuadrático módulo p no nos dice nada sobre si lo es módulo otro primo q.

Pues bien, alrededor de 1.740 Euler estaba investigando un caso particular de este problema (concretamente cuándo un primo p divide a un número de la forma $x^2 + ny^2$) y descubrió algo sorprendente. Es obvio que la condición

$$\Delta \text{ es un resto cuadrático módulo } p$$

depende sólo del resto de Δ módulo p. Lo que Euler descubrió (formulado en nuestros términos) es que dicha condición *depende sólo del resto de p módulo Δ*, es decir, si dos primos impares p, q cumplen $p \equiv q \pmod{\Delta}$ entonces Δ es un resto cuadrático módulo p si y sólo si lo es módulo q (o equivalentemente, p se escinde en K si y sólo si q se escinde en K). Esto es mucho más que un mero juego de palabras: en efecto, que la condición dependa sólo del resto de Δ módulo p no nos dice nada, porque Δ es fijo y al cambiar p cambiamos el resto módulo p, pero que la condición dependa sólo del resto de p módulo Δ es algo completamente distinto pues, como Δ es una constante de nuestro problema, hay sólo un número finito de restos módulo Δ a tener en cuenta. Concretamente, un primo p que no divida a Δ ha de pertenecer a una de las $\phi\big(|\Delta|\big)$ clases del grupo $U_{|\Delta|}$ de las unidades módulo $|\Delta|$. Lo que nos dice el descubrimiento de Euler es que si determinamos el comportamiento de $\phi\big(|\Delta|\big)$ primos, uno en cada clase, ya no necesitamos comprobar si Δ es o no un resto cuadrático módulo ningún otro primo, sino que, dado otro primo p, bastará calcular su resto módulo Δ (o sea, ver a qué clase corresponde) y ya sabremos que el comportamiento de p será el mismo que el del primo en dicha clase que ya habíamos estudiado.

En nuestro ejemplo concreto tenemos que considerar el grupo

$$U_{13} = \big\{\ [1],\ [2],\ [3],\ [4],\ [5],\ [6],\ [7],\ [8],\ [9],\ [10],\ [11],\ [12]\ \big\}.$$

Tomamos un primo en cada clase, por ejemplo:

$$\mathbf{53},\ 2,\ \mathbf{3},\ \mathbf{17},\ 5,\ 19,\ 7,\ 47,\ \mathbf{61},\ \mathbf{23},\ 37,\ \mathbf{51}.$$

Comprobamos si 13 es o no un resto cuadrático módulo cada uno de estos doce primos (es resto cuadrático módulo los que están en negrita). Si el lector se molesta en hacer los cálculos se hará una idea de lo tedioso que resulta comprobar que 13 es un resto no cuadrático módulo 37.

Admitiendo el descubrimiento de Euler, la conclusión a la que llegamos es que un primo p se escinde en K si y sólo si $p \equiv 1, 3, 4, 9, 10, 12 \pmod{13}$, p se conserva en K si y sólo si $p \equiv 2, 5, 6, 7, 8, 11 \pmod{13}$ y p se ramifica si y sólo si $p = 13$.

(En realidad nuestros razonamientos no incluyen el caso $p = 2$, pero veremos que la conclusión vale igualmente).

Ésta es, pues, una solución muchísimo más elegante desde un punto de vista teórico y muchísimo más satisfactoria desde un punto de vista práctico, pues es más fácil calcular el resto de p módulo 13 que determinar si 13 es o no un resto cuadrático módulo p. Pero los descubrimientos de Euler no acaban aquí. Notemos que de las 12 clases resulta que 6 corresponden a primos que se escinden y las otras 6 a primos que se ramifican. Esto no es casual. Más aún, si nos fijamos veremos que las clases $\big\{[1], [3], [4], [9], [10], [12]\big\}$, correspondientes a primos que se escinden, forman un subgrupo de U_{13}. También llama la atención que $1 + 12 = 3 + 10 = 4 + 9 = 13$. En general, si llamamos clases de escisión a aquellas cuyos primos se escinden, se cumple:

1. Las clases de escisión forman un subgrupo de índice 2 en $U_{|\Delta|}$.

2. Si K es real la clase $[n]$ es de escisión si y sólo si lo es $[-n]$, si K es imaginario $[n]$ es una clase de escisión si y sólo si $[-n]$ no lo es.

Sabiendo esto, el camino por el que hemos llegado a la solución del ejemplo se allana mucho más. En efecto: comprobamos sin dificultad que 13 es un resto cuadrático módulo 3 (pues $[13] = [1]$), por lo tanto $[3]$ es una clase de escisión. Por la propiedad 2) también lo es $[-3] = [10]$. Por 1) también lo es $[3]^2 = [9]$ y por 2) $[-9] = [4]$. Por último $[1]$ es clase de escisión por 1) y $[12]$ lo es por 2). Como ya tenemos 6, sabemos que son todas.

Así pues, con esta información la solución del problema no requiere más 'cálculo' que el carácter cuadrático de 13 módulo 3 (el lector que se haya molestado en calcular el carácter cuadrático de los doce primos de antes apreciará sin duda la diferencia).

Euler descubrió empíricamente estos hechos, aunque no fue capaz de demostrar sino una mínima parte de ellos. No obstante encontró una propiedad sencilla de enunciar sobre restos cuadráticos a partir de la cual demostró todos estos resultados. Hoy se conoce como Ley de Reciprocidad Cuadrática.

Legendre demostró la ley de reciprocidad a partir de un principio aparentemente más sencillo que no pudo demostrar y que nosotros hemos usado tácitamente: que toda clase de U_m contiene algún primo. En efecto, en la clase $[1]$ hemos encontrado el primo 53, en $[4]$ está el 17, etc., pero ¿qué garantía teníamos de que en todas las clases íbamos a encontrar algún primo? Si una clase no contuviera primos no tendría sentido decir si es o no una clase de escisión. Notar que si $[n]$ es una clase de U_m sus elementos son los números de la forma $mx + n$, por lo que el postulado de Legendre se puede enunciar así:

> *Si m y n son números naturales primos entre sí entonces la progresión aritmética $mx + n$ contiene un número primo.*

Cuando contaba con poco más de veinte años de edad, Gauss redescubrió la ley de reciprocidad cuadrática y la demostró. También se tropezó con el postulado de Legendre y no consiguió demostrarlo, pero pudo eludirlo a la hora de probar la ley de reciprocidad.

Años más tarde, Dirichlet demostró lo que hoy se conoce como teorema de Dirichlet sobre primos en progresiones aritméticas, según el cual si $(m, n) = 1$ la progresión aritmética $mx + n$ contiene *infinitos* primos. La prueba original de Dirichlet combinaba técnicas analíticas con la aritmética de los cuerpos ciclotómicos (de cualquier orden, no necesariamente primo) y escapa a nuestras posibilidades actuales. Hoy se conoce una prueba que no requiere teoría algebraica de números pero sí una cierta dosis de teoría de funciones analíticas, algo completamente ajeno a este libro.

Así pues, nosotros no demostraremos el teorema de Dirichlet, pero probaremos la ley de reciprocidad sin él. La prueba original de Gauss era muy complicada, al parecer capaz de desesperar a sus alumnos más aventajados.

Después encontró otra prueba basada esencialmente en la aritmética de los cuerpos cuadráticos también muy profunda pero conceptualmente más clara. Años más tarde encontró varias pruebas más de carácter más elemental. Aquí veremos una prueba más moderna completamente elemental. Después tendremos que ingeniárnoslas para evitar el teorema de Dirichlet a la hora de deducir los teoremas de Euler sobre cuerpos cuadráticos.

14.2 El símbolo de Legendre

Los resultados sobre restos cuadráticos se enuncian más cómodamente utilizando una notación debida a Legendre.

Definición 14.1 Sea $p > 0$ un primo impar y n un entero. Definimos el *símbolo de Legendre* (n/p) como

$$\left(\frac{n}{p}\right) = \begin{cases} 1 & \text{si } n \text{ es un resto cuadrático módulo } p \\ -1 & \text{si } n \text{ es un resto no cuadrático módulo } p \\ 0 & \text{si } p \mid n \end{cases}$$

Obviamente, si $n \equiv n' \pmod p$ entonces $(n/p) = (n'/p)$.

Sea p un primo impar y consideremos la aplicación $U_p \longrightarrow U_p$ dada por $x \mapsto x^2$. Obviamente es un homomorfismo de grupos cuya imagen la forman las clases de restos cuadráticos módulo p. Su núcleo son las clases de $\mathbb{Z}/p\mathbb{Z}$ que cumplen $x^2 - 1 = 0$. Como $\mathbb{Z}/p\mathbb{Z}$ es un cuerpo este polinomio tiene sólo dos raíces, a saber, $\pm[1]$. Ahora el teorema de isomorfía nos da que el número de clases de restos cuadráticos es exactamente la mitad del total de clases, o sea, $(p-1)/2$.

El paso siguiente en el estudio de los restos cuadráticos es el teorema siguiente:

Teorema 14.2 *(Criterio de Euler) Sea p un primo impar y $n \in \mathbb{Z}$. Entonces*

$$\left(\frac{n}{p}\right) \equiv n^{(p-1)/2} \pmod p.$$

DEMOSTRACIÓN: Si $p \mid n$ el teorema es evidente. Supongamos lo contrario.

Tomando clases módulo p, vemos que $\left[n^{(p-1)/2}\right]^2 = [n]^{p-1} = [1]$, pues $[n]$ es un elemento del grupo U_p, de orden $p-1$, y acabamos de comentar que las únicas clases que cumplen esto son $\pm[1]$, luego $\left[n^{(p-1)/2}\right] = \pm[1]$.

Por otro lado, si n es un resto cuadrático se cumple $[n] = [u]^2$, luego

$$\left[n^{(p-1)/2}\right] = [u]^{p-1} = [1].$$

Basta probar que esto sólo ocurre cuando n es un resto cuadrático, pero esto ocurre cuando $[n]$ es una raíz del polinomio $x^{(p-1)/2} - 1$, que a lo sumo tiene $(p-1)/2$ raíces, y como de hecho las clases de los restos cuadráticos suman

ya este número, concluimos que no hay más. Por lo tanto, si n es un resto no cuadrático $\left[n^{(p-1)/2}\right]$ ha de ser $-[1]$. ∎

Esto nos da un método para calcular símbolos de Legendre ligeramente más práctico que calcular todos los cuadrados módulo p: se calcula $n^{(p-1)/2}$ y se reduce módulo p. Tiene que dar -1, 0 o 1 y ése es el valor de (n/p). Una consecuencia teórica de interés es que el símbolo de Legendre es multiplicativo:

Teorema 14.3 *Sea p un primo impar y a, b enteros cualesquiera. Entonces:*

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$$

Demostración: El criterio de Euler nos da que ambos miembros son congruentes, pero como valen -1, 0 o 1, de hecho son iguales. ∎

Del criterio de Euler también se deduce un caso particular de la ley de reciprocidad cuadrática:

Teorema 14.4 *Si p es un primo impar,*

$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2},$$

luego -1 es un resto cuadrático módulo p si y sólo si $p \equiv 1$ (mód 4).

En efecto, el criterio de Euler nos da que ambos miembros son congruentes módulo p, pero de hecho ambos son iguales a ± 1, luego son iguales entre sí.

¿Por qué módulo 4? Porque -4 es el discriminante de $\mathbb{Q}\big(\sqrt{-1}\,\big)$. Tenemos que -1 es un resto cuadrático módulo p si y sólo si $\Delta = -4$ es un resto cuadrático módulo p y, según uno de los teoremas de Euler que hemos citado en la introducción, esto ha de depender sólo del resto de p módulo $|\Delta| = 4$, que es lo que ahora acabamos de probar.

Nuestra prueba de la ley de reciprocidad se basará en el siguiente teorema de Gauss:

Teorema 14.5 *(Gauss): Sea p un primo impar y a un entero primo con p. Sean los subconjuntos de U_p dados por*

$$N = \{\ [-1], \ldots, [-\frac{p-1}{2}]\ \} \qquad y \qquad P = \{\ [1], \ldots, [\frac{p-1}{2}]\ \}.$$

Sea $r = \big|[a]P \cap N\big|$, donde $[a]P = \big\{[a][u] \;\big|\; [u] \in P\big\}$. Entonces

$$\left(\frac{a}{p}\right) = (-1)^r.$$

DEMOSTRACIÓN: Como $[a]$ es una unidad, la aplicación que a $[u]$ le asigna $[a][u]$ es biyectiva, luego $\big|[a]P\big| = |P| = (p-1)/2$.

Dados dos elementos de P, digamos $[u]$ y $[v]$ con $0 < u < v \leq (p-1)/2$, no puede suceder que $[a][u] = -[a][v]$, pues entonces $[a][u] + [a][v] = 0$, o sea, $p \mid a(u+v)$, luego $p|u+v<p$, lo cual es imposible.

Esto significa que $[a]P = \big\{\pm[1], \ldots, \pm[(p-1)/2]\big\}$, y hay exactamente r signos negativos. Por lo tanto, si llamamos $[z]$ al producto de todos los elementos de P, tenemos que el producto de todos los elementos de $[a]P$ es por un lado $[a]^{(p-1)/2}[z]$ y por otro es $(-1)^r[z]$. Como $[z]$ es una unidad podemos simplificarla y queda $[a]^{(p-1)/2} = [(-1)^r]$.

El teorema se sigue ahora del criterio de Euler. ∎

Este criterio no nos será especialmente interesante en cuanto dispongamos de la ley de reciprocidad cuadrática, pero como ejemplo para comprenderlo adecuadamente vamos a usarlo para determinar si 3 es un resto cuadrático módulo 19.

Calculamos $[3]P$, es decir, multiplicamos $[3]$ por los números del 1 al 9:

$$[3]P = \big\{\ [3],\ [6],\ [9],\ [12],\ [15],\ [18],\ [2],\ [5],\ [8]\ \big\}.$$

Restamos 19 a los números mayores que 9 y queda:

$$\begin{aligned}[3]P &= \big\{\ [3],\ [6],\ [9],\ [-7],\ [-4],\ [-1],\ [2],\ [5],\ [8]\ \big\}\\ &= \big\{\ [-1],\ [2],\ [3],\ [-4],\ [5],\ [6],\ [-7],\ [8],\ [9]\ \big\}.\end{aligned}$$

Como hay 3 signos negativos, $(3/19) = (-1)^3 = -1$, luego 3 es un resto no cuadrático módulo 19. Nótese que este cálculo es más simple que aplicar el criterio de Euler directamente.

El -1 es un caso aparte en la ley de reciprocidad cuadrática porque es una unidad. También es un caso aparte el 2 (y muy en el fondo el motivo es que 2 es el grado de los cuerpos cuadráticos). El teorema anterior nos permite resolverlo:

Teorema 14.6 *Sea p un primo impar. Entonces*

$$\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8},$$

luego 2 es un resto cuadrático módulo p si y sólo si $p \equiv \pm 1$ (mód 8).

Aplicamos el teorema de Gauss. Consideramos $[2]P = \big\{\ [2],\ [4], \ldots, [p-1]\ \big\}$. Entonces $r = \frac{p-1}{2} - s$, donde s es el mayor natural tal que $2s \leq (p-1)/2$.

CASO 1: $(p-1)/2$ es par y, consecuentemente $2s = \frac{p-1}{2}$. Calculando obtenemos $r = (p-1)/4$ y el valor del símbolo de Legendre es $(-1)^{(p-1)/4}$, pero como $(p+1)/2$ es impar, este valor no se altera si lo elevamos a $(p+1)/2$, y como $\frac{p-1}{4}\,\frac{p+1}{2} = \frac{p^2-1}{8}$, se cumple el teorema.

CASO 2: $(p-1)/2$ es impar, y $2s = \frac{p-1}{2} - 1$, luego $r = (p+1)/4$, y como en el caso anterior podemos elevar el resultado que nos da el teorema de Gauss al exponente impar $(p-1)/2$, de donde se sigue también el enunciado. ∎

A la hora de determinar si un número n es o no un resto cuadrático módulo un primo dado p, lo mejor es usar el teorema 14.3, que nos reduce el problema a calcular los símbolos de Legendre para los primos divisores de n y para -1 si n es negativo (de hecho, los primos que tengan exponente par pueden ser ignorados, pues el cuadrado de un símbolo de Legendre siempre es 1).

Por lo tanto el problema se reduce a calcular símbolos de Legendre (q/p), donde q es primo o $q = -1$. Los casos $q = -1$ y $q = 2$ ya están resueltos. La ley de reciprocidad cuadrática cubre los casos restantes:

Teorema 14.7 *(Ley de reciprocidad cuadrática) Sean p y q primos impares distintos.*

1. *Si p o q es congruente con 1 módulo 4, entonces*

$$\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right),$$

2. *Si ninguno de los dos primos es congruente con 1 módulo 4, entonces*

$$\left(\frac{p}{q}\right) = -\left(\frac{q}{p}\right),$$

Equivalentemente, para cualquier par de primos impares p y q se cumple

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4}.$$

DEMOSTRACIÓN: Es claro que si p o q es congruente con 1 módulo 4 entonces $(p-1)(q-1)/4$ es un número par, y en otro caso es impar (los factores $(p-1)/2$ y $(q-1)/2$ son impares). Por lo tanto la última afirmación del enunciado es equivalente a las dos primeras.

Por el teorema de Gauss, $(p/q) = (-1)^r$, donde r es el número de enteros x entre 1 y $(q-1)/2$ tales que px es congruente módulo q con un entero u tal que $-(q-1)/2 \le u < 0$, o sea, $px = qy + u$ para cierto entero y, completamente determinado por x.

Por lo tanto, r es también el número de pares de números enteros (x, y) tales que

$$1/2 < x < q/2 \qquad \text{y} \qquad -q/2 < px - qy < 0.$$

Entonces $qy < px + q/2 < pq/2 + q/2$, luego $y < p/2 + 1/2$ y, al ser entero, $y < p/2$.

Notar también que, como $px > 0$ y $u < 0$, de $px = qy + u$ se sigue que $y > 0$.

En resumen, r es el número de pares de números enteros (x, y) que cumplen

$$1/2 < x < q/2, \qquad 1/2 < y < p/2 \tag{14.1}$$

$$-q/2 < px - qy < 0. \tag{14.2}$$

Análogamente, $(q/p) = (-1)^s$, donde s es el número de pares de enteros que cumplen estas mismas condiciones cambiando p por q y viceversa. Este número no se altera si cambiamos también x por y.

Si además multiplicamos las desigualdades por -1 queda:

$$1/2 < x < q/2, \qquad 1/2 < y < p/2 \tag{14.3}$$

$$0 < px - qy < p/2. \tag{14.4}$$

Lo que queremos probar es que $\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{r+s} = (-1)^{(p-1)(q-1)/4}$, lo que equivale a que $(p-1)(q-1)/4 - (r+s)$ sea un número par.

Ahora bien, $(p-1)(q-1)/4$ es el número de pares de enteros (x, y) que cumplen (14.1), luego $(p-1)(q-1)/4 - (r+s)$ resulta ser el número de pares de enteros (x, y) que cumplen (14.1) pero que no cumplen ni (14.2) ni (14.4).

Notemos que siendo $p \neq q$ es imposible $px - qy = 0$, ya que en tal caso $p \mid y < p/2$. Por lo tanto, la negación de (14.2) y (14.4) equivale a la negación de $-q/2 < px - qy < p/2$. Lo que debemos probar es que el número de pares de enteros (x, y) que cumplen

$$1/2 < x < q/2, \qquad 1/2 < y < p/2, \qquad \text{y no} \qquad -q/2 < px - qy < p/2$$

es un número par. La figura 14.1 ilustra la prueba en el caso $p = 11$ y $q = 17$.

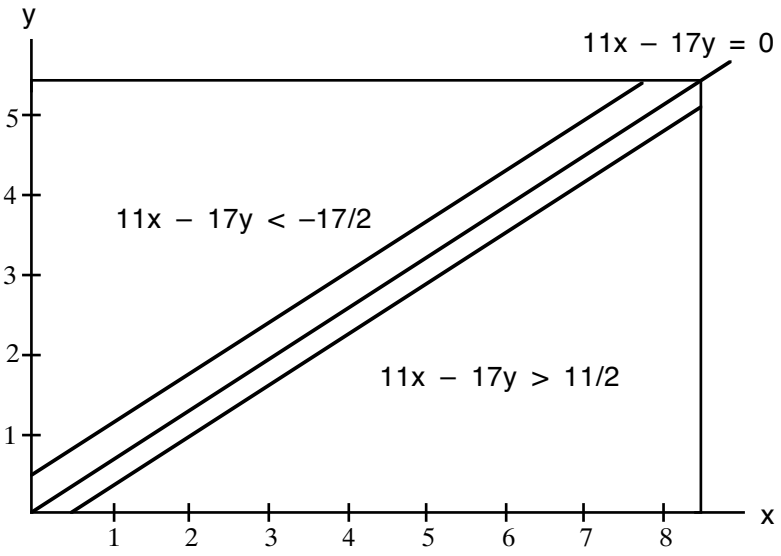


Figura 14.1: La ley de reciprocidad cuadrática

Definimos los conjuntos

$$A = \big\{(x,y) \,\big|\, 1/2 < x < q/2,\ \ 1/2 < y < p/2,\ \ px - qy < -q/2\big\}$$
$$B = \big\{(x,y) \,\big|\, 1/2 < x < q/2,\ \ 1/2 < y < p/2,\ \ px - qy > \ \ p/2\big\}.$$

Claramente se trata de conjuntos disjuntos y hemos de probar que el número de elementos de $A \cup B$ es par. Para ello basta comprobar que A y B tienen el mismo número de elementos. La aplicación $\phi : \mathbb{Z} \times \mathbb{Z} \longrightarrow Z \times \mathbb{Z}$ dada por $\phi(x, y) = \big((q+1)/2-x, (p+1)/2-y\big)$ es claramente biyectiva y es fácil comprobar que cumple $f[A] = B$, luego el teorema está probado. ∎

Antes de extraer consecuencias teóricas de la ley de reciprocidad, vamos a ver cómo nos permite calcular cualquier símbolo de Legendre mediante operaciones elementales. El método consiste, dado (n/p), en reducir n módulo p, factorizarlo en primos y descomponer según 14.3 en el producto de los símbolos de Legendre de la forma (q/p), donde q recorre los primos que dividen a n con exponente impar y el -1 si n es negativo. Después reducimos cada q módulo p y volvemos a factorizar. Cuando lleguemos a un factor (q/p), donde q ya es menor que p, invertimos el símbolo mediante la ley de reciprocidad y seguimos así hasta reducir los símbolos a casos calculables directamente o, si queremos, hasta llegar a 1 o a -1. En la práctica es mucho más simple de lo que parece. Veamos algunos ejemplos:

$$\left(\frac{2.002}{97}\right) = \left(\frac{62}{97}\right) = \left(\frac{2}{97}\right)\left(\frac{31}{97}\right) = \left(\frac{31}{97}\right) = \left(\frac{97}{31}\right) = \left(\frac{4}{31}\right) = 1,$$

donde usamos que $97 \equiv 1$ (mód 8), luego $(2/97) = 1$, y $97 \equiv 1$ (mód 4), luego $(31/97) = (97/31)$. Finalmente, 4 es un cuadrado perfecto, luego $(4/31) = 1$.

$$\left(\frac{15}{71}\right) = \left(\frac{3}{71}\right)\left(\frac{5}{71}\right) = -\left(\frac{71}{3}\right)\left(\frac{71}{5}\right) = -\left(\frac{2}{3}\right)\left(\frac{1}{5}\right) = 1.$$

14.3 El símbolo de Jacobi

Recordemos que nuestro objetivo es deducir los teoremas de Euler a partir de la ley de reciprocidad cuadrática. Sin embargo no podemos hacerlo directamente porque ello nos llevaría a tomar primos en las clases de los grupos U_m, y no sabemos justificar su existencia.

De este modo, si bien en la práctica siempre es más conveniente reducir todos los cálculos a números primos mediante factorizaciones, en teoría hemos de arreglárnoslas para trabajar con números cualesquiera. Vamos a demostrar una ley de reciprocidad en la que los números que intercambiemos no sean primos necesariamente. El primer paso es generalizar el símbolo de Legendre:

Definición 14.8 Sean m y n enteros no nulos primos entre sí. Supongamos que n es impar positivo y factoriza en la forma $n = p_1 \cdots p_r$, donde los primos son positivos y no necesariamente distintos. Definimos el *símbolo de Jacobi* de m y n como

$$\left(\frac{m}{n}\right) = \left(\frac{m}{p_1}\right) \cdots \left(\frac{m}{p_r}\right),$$

(donde los símbolos de la derecha son símbolos de Legendre).

Para $n = 1$ definimos $(m/1) = 1$.

Claramente se tiene

$$\left(\frac{mm'}{n}\right) = \left(\frac{m}{n}\right)\left(\frac{m'}{n}\right), \qquad \left(\frac{m}{nn'}\right) = \left(\frac{m}{n}\right)\left(\frac{m}{n'}\right),$$

así como que si $m \equiv m'$ (mód n), entonces $(m/n) = (m'/n)$. Además el símbolo de Jacobi coincide con el de Legendre cuando n es primo.

Es importante dejar claro que no es cierto que $(m/n) = 1$ si y sólo si m es un resto cuadrático módulo n. El interés del símbolo de Jacobi es simplemente que generaliza formalmente el símbolo de Legendre a números compuestos, pero no sirve directamente para obtener resultados sobre restos. He aquí la ley de reciprocidad generalizada:

Teorema 14.9 *Sean m, $n > 0$ enteros impares con $(m, n) = 1$.*

1. $\left(\dfrac{-1}{n}\right) = (-1)^{(n-1)/2}$.
2. $\left(\dfrac{2}{n}\right) = (-1)^{(n^2-1)/8}$.
3. $\left(\dfrac{m}{n}\right)\left(\dfrac{n}{m}\right) = (-1)^{(m-1)(n-1)/4}$.

DEMOSTRACIÓN: 1) se deduce inmediatamente a partir de la definición y de la propiedad análoga de los símbolos de Legendre. Hay que tener en cuenta que si n y n' son impares,

$$4 \mid (n-1)(n'-1) = nn' - 1 - (n-1) - (n'-1),$$

luego el número $(nn'-1)/2 - (n-1)/2 - (n'-1)/2$ es par, es decir,

$$(-1)^{(nn'-1)/2} = (-1)^{(n-1)/2}(-1)^{(n'-1)/2}.$$

2) se demuestra igual que 1), usando esta vez que si n y n' son impares

$$16 \mid (n^2-1)(n'^2-1) = \big((nn')^2 - 1\big) - (n^2-1) - (n'^2-1),$$

luego $\big((nn')^2-1\big)/8 - (n^2-1)/8 - (n'^2-1)/8$ es par.

3) Si $m = 1$ es obvio. Supongamos que m es primo. Si $m \equiv 1$ (mód 4), entonces desarrollamos (m/n) según los factores primos de n, invertimos por la ley de reciprocidad cuadrática, volvemos a agrupar y queda $(m/n) = (n/m)$.

Supongamos que $m \equiv -1$ (mód 4). Si $n \equiv 1$ (mód 4), entonces el número de factores primos de n congruentes con -1 módulo 4 ha de ser par, luego al descomponer, invertir y agrupar aparece un número par de signos negativos, luego de nuevo $(m/n) = (n/m)$.

Si $n \equiv -1$ (mód 4) el número de signos negativos que aparecen es impar, luego resulta que $(m/n) = -(n/m)$.

Si m no es primo razonamos igual descomponiendo m y usando la parte ya probada en lugar de la ley de reciprocidad. ∎

Estos resultados nos permiten calcular símbolos de Jacobi igual que los de Legendre.

El teorema siguiente contiene el núcleo de los teoremas de Euler: Los símbolos de Jacobi dependen sólo del resto del numerador módulo el denominador. Ahora usaremos la ley de reciprocidad para probar que (con ciertas restricciones) sólo dependen del resto del denominador módulo el numerador.

Teorema 14.10 *Sean m y n enteros primos entre sí con n impar y positivo. Si n' es un entero impar y positivo tal que $n \equiv n' \pmod{4m}$, entonces*

$$\left(\frac{m}{n}\right) = \left(\frac{m}{n'}\right).$$

Si se cumple $m \equiv 1 \pmod 4$ es suficiente exigir $n \equiv n' \pmod m$.

Demostración: Sea $m = \epsilon\, 2^j m'$, donde m' es impar y $\epsilon = \pm 1$. Entonces

$$\begin{aligned}\left(\frac{m}{n}\right) &= \left(\frac{\epsilon}{n}\right)\left(\frac{2}{n}\right)^j\left(\frac{m'}{n}\right) = \left(\frac{\epsilon}{n}\right)\left(\frac{2}{n}\right)^j(-1)^{(m'-1)(n-1)/4}\left(\frac{n}{m'}\right)\\ \left(\frac{m}{n'}\right) &= \left(\frac{\epsilon}{n'}\right)\left(\frac{2}{n'}\right)^j\left(\frac{m'}{n'}\right) = \left(\frac{\epsilon}{n'}\right)\left(\frac{2}{n'}\right)^j(-1)^{(m'-1)(n'-1)/4}\left(\frac{n'}{m'}\right).\end{aligned}$$

Como $n \equiv n' \pmod m$, se cumple $(n/m') = (n'/m')$.

Como $n \equiv n' \pmod 4$, también $(-1)^{(m'-1)(n-1)/4} = (-1)^{(m'-1)(n'-1)/4}$.

Si j es par, también $(2/n)^j = (2/n')^j$. Si j es impar, entonces $(2/n)^j = (2/n)$ y $(2/n')^j = (2/n')$. Además m es par, luego de $n \equiv n' \pmod{4m}$ se deduce $n \equiv n' \pmod 8$, y por el teorema 14.9 $(2/n) = (2/n')$ (vale 1 si y sólo si n (o n') es congruente con ± 1 módulo 8).

Finalmente, también por el teorema 14.9, $(\epsilon/n) = (\epsilon/n')$ (es igual a 1 si y sólo si n (o n') es congruente con 1 módulo 4). En consecuencia $(m/n) = (m/n')$.

Si $m \equiv 1 \pmod 4$ y $n \equiv n' \pmod m$ entonces $j = 0$.

Si $\epsilon = 1$ entonces directamente

$$\left(\frac{m}{n}\right) = \left(\frac{n}{m}\right) = \left(\frac{n'}{m}\right) = \left(\frac{m}{n'}\right).$$

Si $\epsilon = -1$ entonces $m' \equiv -1 \pmod 4$, luego

$$(-1)^{(m'-1)(n-1)/4} = (-1)^{(n-1)/2} = \left(\frac{\epsilon}{n}\right),$$

y llegamos a la misma conclusión. ∎

14.4 Los teoremas de Euler

Definición 14.11 Sea $d \neq 1$ un entero libre de cuadrados y sea Δ el discriminante del cuerpo $\mathbb{Q}\big(\sqrt{d}\,\big)$. Definimos la aplicación *signatura*

$$\mathrm{sig} : U_{|\Delta|} \longrightarrow \{1, -1\},$$

mediante $\mathrm{sig}\big([n]\big) = (d/n)$ (donde $n > 0$ es impar)

Por los teoremas 11.14 y 14.10, sig([n]) no depende del impar $n > 0$ escogido en la clase. Diremos que una clase $A \in U_{|\Delta|}$ es una clase de escisión para d si $\mathrm{sig}(A) = +1$.

Probamos finalmente los teoremas de Euler:

Teorema 14.12 *Sea $d \neq 1$ un entero libre de cuadrados y Δ el discriminante del cuerpo $\mathbb{Q}\big(\sqrt{d}\,\big)$.*

1. *La aplicación* $\mathrm{sig} : U_{|\Delta|} \longrightarrow \{1, -1\}$ *es un epimorfismo de grupos, por lo que el conjunto de las clases de escisión forma un subgrupo de índice* 2 *en* $U_{|\Delta|}$*, y su orden es* $\phi\big(|\Delta|\big)/2$.

2. *Si $d > 0$ entonces $[n]$ es una clase de escisión si y sólo si $[-n]$ es una clase de escisión.*

3. *Si $d < 0$ entonces $[n]$ es una clase de escisión si y sólo si $[-n]$ no es una clase de escisión.*

4. *Si p es un primo que no divide a Δ, entonces p se escinde en $\mathbb{Q}\big(\sqrt{d}\,\big)$ si y sólo si $[p]$ es una clase de escisión. En otro caso se conserva primo.*

Demostración: Es inmediato que la signatura es un homomorfismo de grupos. Los apartados 2) y 3) se reducen a probar que $\mathrm{sig}\big([-1]\big) =$ signo de d. En efecto:

Sea $d = \epsilon\, 2^j u$, donde $u > 0$ es impar y $\epsilon = \pm 1$. Entonces

$$\mathrm{sig}\big([-1]\big) = \left(\frac{d}{4|d|-1}\right) = \left(\frac{\epsilon}{4|d|-1}\right)\left(\frac{2}{4|d|-1}\right)^{j}\left(\frac{u}{4|d|-1}\right).$$

Aplicamos a cada factor el caso correspondiente del teorema 14.9:

$$\left(\frac{\epsilon}{4|d|-1}\right) = \epsilon,$$

O bien $j = 0$, o bien $8 \mid 4|d|$. En culquier caso $\left(\dfrac{2}{4|d|-1}\right)^{j} = 1$.

Finalmente,

$$\left(\frac{u}{4|d|-1}\right) = (-1)^{(u-1)/2}\left(\frac{4|d|-1}{u}\right) = (-1)^{(u-1)/2}\left(\frac{-1}{u}\right) = 1.$$

Por lo tanto $\mathrm{sig}\big([-1]\,\big) = \epsilon$.

Para probar que la signatura es un epimorfismo es suficiente encontrar un entero n tal que $\mathrm{sig}\big([n]\big) = -1$. Si $d < 0$ es consecuencia del apartado 3).

Supongamos $d > 0$. El caso $d = 2$ se sigue del teorema 14.9. Basta tomar un entero $n \not\equiv \pm 1$ (mód 8).

Si $d \neq 2$ sea p un primo impar que divida a d. Sea $d = pm$. Podemos tomar un entero u tal que $(u/p) = -1$. Por el teorema chino del resto existe un entero n tal que $n \equiv u$ (mód p) y $n \equiv 1$ (mód $4m$) (pues al ser d libre de cuadrados se cumple que $(p, m) = 1$). Notar que $(d, n) = 1$. Entonces

$$\mathrm{sig}\big([n]\big) = \left(\frac{d}{n}\right) = \left(\frac{p}{n}\right)\left(\frac{m}{n}\right) = \left(\frac{n}{p}\right)\left(\frac{m}{n}\right) = \left(\frac{u}{p}\right)\left(\frac{m}{1}\right) = -1.$$

El resto del apartado 1) es el teorema de isomorfía.

4) Por el teorema 13.2 un primo impar p que no divide a Δ se escinde si y sólo si d es un resto cuadrático módulo p, o sea, si y sólo si $\mathrm{sig}\big([p]\big) = (d/p) = 1$. En caso contrario p se conserva.

Falta considerar el caso $p = 2$. Si $2 \nmid \Delta$ entonces $\Delta = d \equiv 1$ (mód 4).

Para calcular $\mathrm{sig}\big([2]\big)$ hemos de expresar $[2] = [2 + kd]$, donde k se escoge de modo que $2 + kd$ sea impar y positivo.

Podemos exigir $k \equiv -1$ (mód 4), y así $2 + kd \equiv 1$ (mód 4). El teorema 14.9 nos da entonces que $\left(\frac{-1}{2+kd}\right) = 1$, luego

$$\mathrm{sig}\big([2]\big) = \left(\frac{d}{2+kd}\right) = \left(\frac{|d|}{2+kd}\right) = \left(\frac{2+kd}{|d|}\right) = \left(\frac{2}{|d|}\right).$$

Según el teorema 14.9 concluimos que

$$\mathrm{sig}\big([2]\big) = 1 \quad \text{ si y sólo si } \quad |d| \equiv \pm 1 \text{ (mód 8)},$$

y como $d \equiv 1$ (mód 4) esto equivale a que $d \equiv 1$ (mód 8), es decir, a que 2 se escinda en K (según el teorema 13.2). ∎

Entre los teoremas de Euler figuraban algunas propiedades más, como que los cuadrados de $U_{|\Delta|}$ son clases de escisión, pero esto es consecuencia de que la signatura es un homomorfismo. Igualmente, el producto de dos clases que no sean de escisión es una clase de escisión, etc.

Ejemplo: Consideremos $K = \mathbb{Q}\big(\sqrt{3}\,\big)$. Sabemos que su anillo de enteros $\mathbb{Z}\big[\sqrt{3}\,\big]$ es un dominio de factorización única (es euclídeo), pero en el capítulo anterior vimos que su unidad fundamental es $\epsilon = 2 + \sqrt{3}$, y $\mathrm{N}(\epsilon) = 1$. Como todas las demás unidades son potencia de ésta, resulta que $\mathbb{Z}\big[\sqrt{3}\,\big]$ no tiene unidades de norma negativa, con lo que no se cumple una de las hipótesis que poníamos al principio del capítulo.

Tratemos de estudiar, a pesar de ello, qué números son de la forma $x^2 - 3y^2$, es decir, qué números son la norma de un entero de $\mathbb{Z}\big[\sqrt{3}\,\big]$.

El discriminante de K es $\Delta = 12$ y en el grupo $U_{12} = \big\{[1],[5],[7],[11]\big\}$ son clases de escisión [1] y [11] (directamente por el teorema 14.12).

Así pues se ramifican 2 y 3, se escinden los primos $p \equiv \pm 1$ (mód 12) y se conservan los primos $p \equiv \pm 5$ (mód 12).

Por lo tanto, una condición necesaria para que un entero m sea de la forma $m = x^2 - 3y^2$ es que si un primo p divide a m con exponente impar entonces $p = 2, 3$, o $p \equiv \pm 1$ (mód 12), pues en caso contrario es imposible constuir primos cuadráticos de norma p.

El recíproco no es cierto. Tomemos $p = 11$. Entones p se escinde, pero eso sólo significa que existe un ideal primo de norma 11, sin embargo no hay enteros de norma 11. Lo único que tenemos es $\mathrm{N}\big(1 \pm 2\sqrt{3}\,\big) = -11$ (si un primo tuviera norma 11 sería asociado de uno de estos dos, y el cociente sería una unidad de norma -1).

Si prescindimos de los signos tenemos una equivalencia: para todo entero m se cumple que $\pm m = x^2 - 3y^2$ si y sólo si todo primo p que divida a m con exponente impar cumple $p = 2$, 3, o $p \equiv \pm 1$ (mód 12). Para cada m sólo un signo es posible (los detalles quedan a cargo del lector)

No es difícil determinar cuál es el signo adecuado: Si m cumple las condiciones, su signo será el mismo que el de su parte libre de cuadrados m'. Si además $(m', 3) = 1$ entonces $\pm m' = x^2 - 3y^2$ implica $\pm m' \equiv 1$ (mód 3), luego el signo adecudo es el que verifica también $m' \equiv \pm 1$ (mód 3). Si $3 \mid m'$, como $-3 = \mathrm{N}\big(1 \pm \sqrt{3}\,\big)$, el signo para m es el opuesto que para $m'/3$.

Capítulo XV

La teoría de Galois

En los capítulos anteriores hemos conocido una parte de la teoría de cuerpos cuadráticos, no en toda su profundidad, pero sí en grado suficiente como para comprender el grado de sofisticación a que puede llegar la teoría algebraica de números. Los cuerpos cuadráticos son los más simples de todos los cuerpos numéricos, por lo que no ha de extrañarnos que el estudio de los cuerpos numéricos en general requiera técnicas más potentes que las que hasta ahora hemos estudiado. En realidad muchos de los conceptos que hemos manejado hasta ahora serían excesivos si los quisiéramos tan sólo para estudiar simplemente los cuerpos cuadráticos. Por ejemplo, hemos usado la teoría de extensiones de cuerpos para dar una definición general de norma, pero es muy fácil definir directamente las normas de los cuerpos cuadráticos sin tanto artificio. El interés de la teoría de extensiones de cuerpos, de la teoría de enteros algebraicos, etc. es que nos capacitan para afrontar el estudio de los cuerpos numéricos arbitrarios. En este capítulo presentamos otra de las piedras angulares de la teoría algebraica de números: el teorema de Galois. Sobre cuerpos cuadráticos resulta trivial, pero sobre cuerpos mayores es una herramienta indispensable. Después de desarrollar la teoría general mostraremos diversos contextos en los que se aplica. No obstante, el lector debe tener presente que no estamos en condiciones de estudiar más a fondo la aritmética de los cuerpos numéricos en general, que es lo que realmente da sentido a todo lo que vamos a ver, por lo que este capítulo ha de entenderse como una base algebraica para estudios posteriores. De todos modos, en el capítulo XVII veremos un ejemplo más importante y representativo de la utilidad de la teoría de Galois.

15.1 La correspondencia de Galois

El teorema de Galois establece una biyección entre los cuerpos intermedios de una extensión finita de Galois y los subgrupos de su grupo de Galois $G(K/k)$, lo que permite reducir muchos problemas de cuerpos a problemas sobre grupos finitos.

Recordemos que si K/k es una extensión finita de Galois el grupo $G(K/k)$

tiene orden igual al grado de la extensión. La correspondencia de Galois asigna a cada subgrupo el cuerpo que definimos a continuación:

Definición 15.1 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 8.37 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 15.2 *(Teorema de independencia de Dedekind) Sea K un cuerpo y $\sigma_1, \ldots, \sigma_n$ 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

$$\sum_{i=1}^n c_i\sigma_i(a) = 0 \qquad \text{para todo } a \in K. \tag{15.1}$$

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 (15.1) 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 15.3 *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

$$\sum_{i=1}^{n+1} a_i\sigma_j(b_i) = 0 \qquad \text{para } j = 1, \ldots, n. \tag{15.2}$$

Tomando el valor de j correspondiente al automorfismo identidad, se cumple que $\sum_{i=1}^{n+1} a_ib_i = 0$, 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 $b_1a_1 = 0$ 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 (15.2) tenemos que $\sum_{i=1}^{n+1} \sigma_h(a_i)\sigma_h\big(\sigma_j(b_i)\big) = 0$ 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 (15.2) 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 15.4 *(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 (esto ya lo probamos en el capítulo VIII).

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 8.25). 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 8.28 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 8.25 garantiza que $\sigma[L] = L$, luego $r(\sigma) \in G(L/k)$. La aplicación r es suprayectiva también por el teorema 8.25, 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 $F\big(\langle(\alpha, \beta, \gamma)\rangle\big) = \mathbb{Q}(\omega) = \mathbb{Q}\big(\sqrt{-3}\,\big)$. ∎

Definición 15.5 Si k es un cuerpo, $p(x) \in k[x]$ es un polinomio no constante, y K es el cuerpo de escisión de $p(x)$ sobre k, se llama *grupo de Galois* de $p(x)$ al grupo $G(K/k)$.

Es claro que el grupo de Galois de un polinomio no depende (salvo isomorfismo) del cuerpo de escisión considerado, pues dos cualesquiera son k-isomorfos,

y un k-isomorfismo entre las extensiones induce de forma natural un isomorfismo entre los grupos de Galois.

En el ejemplo anterior hemos visto que el grupo de Galois (sobre $\mathbb{Q}$) del polinomio x^3-2 es (isomorfo a) Σ_3. Es obvio que el grupo de Galois de todo polinomio irreducible de grado 2 es un grupo cíclico de orden 2. Existen técnicas para determinar el grupo de Galois de un polinomio en algunos casos concretos, pero no vamos a entrar en ello. En las secciones siguientes veremos algunos ejemplos más. Ahora veamos un par de resultados sencillos pero muy útiles con los que acabaremos de perfilar la teoría de Galois:

Teorema 15.6 *Sean K/k y L/k extensiones finitas de un cuerpo k (contenidas en un mismo cuerpo) y además supongamos que K/k es 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 $f(x)\in K[x]$ en KL, luego KL es el cuerpo de escisión sobre K 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 15.7 *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 esción sobre k de un polinomio $p(x)$ y L es el cuerpo de escición 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 $|KL:k|=|KL:L|\,|L:k|=|K:k|\,|L:k|$. ∎

15.2 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 15.8 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 estudiar sólo el 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 9.12 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 9.13 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 $\operatorname{grad} c_n(x) = m = \phi(n)$.

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.

Observar 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 15.9 *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,

$$c_n(x) = \frac{x^n - 1}{\prod\limits_{\substack{d|n\\ d\neq n}} c_d(x)}$$

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 15.1 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:

Tabla 15.1: 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
$c_{10}(x)$	$=$	$x^4-x^3+x^2-x+1$

$$\begin{aligned} c_{105}(x) &= x^{48}+x^{47}+x^{46}-x^{43}-x^{42}-2x^{41}-x^{40}-x^{39}+x^{36}+x^{35} \\ &+ x^{34}+x^{33}+x^{32}+x^{31}-x^{28}-x^{26}-x^{24}-x^{22}-x^{20}+x^{17}+x^{16} \\ &+ x^{15}+x^{14}+x^{13}+x^{12}-x^9-x^8-2x^7-x^6-x^5+x^2+x+1. \end{aligned}$$

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 15.10 *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, $c_n(x)\in\mathbb{Z}[x]$. ∎

Notar que las divisiones necesarias para calcular $c_n(x)$ según el teorema 15.9 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.

Ya conocemos los polinomios ciclotómicos de orden primo p. En capítulos anteriores hemos visto que $c_p(x)=x^{p-1}+\cdots+x+1$ y que es irreducible en $\mathbb{Q}[x]$.

Vamos a ver que en realidad esto vale para todos los polinomios ciclotómicos sobre $\mathbb{Q}$.

Teorema 15.11 *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 de $c_n(x)$ en $\mathbb{Z}[x]$ (no constante). 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 $p \leq n$ 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 9.13 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 9.13 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 15.12 *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 y las de orden primo son cíclicas (pues los grupos U_p son cíclicos). 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 15.13 *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 15.12, 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 9.30). 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 $H_aH_b \leq H_d$, 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 $\mathbb{Q}(\omega_m) = \mathbb{Q}(\omega_{2m})$.

Con los resultados de este libro no estamos en condiciones de profundizar en los cuerpos ciclotómicos arbitrarios. En su lugar vamos a usar la teoría de Galois para comprender mejor los cuerpos ciclotómicos de orden primo.

Si p es un número primo el grupo de las raíces p-ésimas primitivas de la unidad sobre $\mathbb{Q}$ tiene orden p, luego todas las raíces distintas de 1 son primitivas (no hay más órdenes posibles que 1 y p). Si ω es una raíz p-ésima primitiva, el grado de la extensión $\mathbb{Q}(\omega)/\mathbb{Q}$ es $p-1$ y el grupo $G(\mathbb{Q}(\omega)/\mathbb{Q})$ es isomorfo a U_p, que cíclico.

De acuerdo con la prueba de 15.12, cada clase $[m] \in U_p$ se corresponde con el automorfismo que envía ω a ω^m. De este modo, si m es una raíz primitiva de la unidad módulo p, es decir, si $[m]$ es un generador de U_p, entonces el automorfismo σ que cumple $\sigma(\omega) = \omega^m$ es un generador de $G(\mathbb{Q}(\omega)/\mathbb{Q})$.

Como consecuencia del teorema de Galois, la extensión $\mathbb{Q}(\omega)/\mathbb{Q}$ tiene tantos cuerpos intermedios como divisores tiene $p-1$. Vamos a describir estos cuerpos. En primer lugar notemos que si un automorfismo fija a un elemento, también lo fijan sus potencias, luego el cuerpo $F\big(\langle\sigma\rangle\big)$ fijado por un grupo cíclico de automorfismos coincide con el conjunto de los elementos fijados por σ, por lo que escribiremos simplemente $F(\sigma)$.

Sea σ un generador de $G(\mathbb{Q}(\omega)/\mathbb{Q})$. Para cada divisor m de $p-1$, el automorfismo σ^m tiene orden $(p-1)/m$, luego $|\mathbb{Q}(\omega) : F(\sigma^m)| = o(\sigma^m) = (p-1)/m$ y por lo tanto $|F(\sigma^m) : \mathbb{Q}| = m$. Llamemos $d = (p-1)/m$.

Antes de seguir nuestro análisis en general conviene poner un ejemplo para imaginarnos la situación. Tomemos $p = 13$, $m = 3$, d= 4. Vamos a obtener el cuerpo intermedio de grado 3 sobre $\mathbb{Q}$. En primer lugar necesitamos un generador del grupo de Galois. Como 2 es una raíz primitiva módulo 13, nos sirve el automorfismo σ que cumple $\sigma(\omega) = \omega^2$. Nuestro cuerpo es el fijado por σ^3, que está determinado por $\sigma^3(\omega) = \omega^8$.

Un elemento cualquiera de $\mathbb{Q}(\omega)$ es de la forma:

$$a + b\omega + c\omega^2 + d\omega^3 + e\omega^4 + f\omega^5 + g\omega^6 + h\omega^7 + i\omega^8 + j\omega^9 + k\omega^{10} + l\omega^{11} + m\omega^{12},$$

donde es importante tener presente que la expresión no es única, sino que podemos sumar una misma cantidad a todos lo coeficientes sin alterar el elemento (ver el capítulo VIII). Ahora bien, si dos elementos de esta forma tienen igual un coeficiente, serán iguales si y sólo si tienen los mismos coeficientes.

Al aplicar σ^3 a nuestro elemento obtenemos lo siguiente:

$$a + b\omega^8 + c\omega^3 + d\omega^{11} + e\omega^6 + f\omega + g\omega^9 + h\omega^4 + i\omega^{12} + j\omega^7 + k\omega^2 + l\omega^{10} + m\omega^5.$$

El elemento estará en $F(\sigma^3)$ si y sólo si esta expresión coincide con la primera. Como el término independiente es a en ambos casos, todos los coeficientes deben coincidir. Esto ocurre si y sólo si $b = f = m = i$, $c = k = l = d$, $e = h = j = g$, es decir, si y sólo si nuestro elemento es de la forma

$$a + u(\omega + \omega^5 + \omega^8 + \omega^{12}) + v(\omega^2 + \omega^3 + \omega^{10} + \omega^{11}) + w(\omega^4 + \omega^6 + \omega^7 + \omega^8).$$

Llamemos

$$\eta_0 = \omega + \omega^5 + \omega^8 + \omega^{12}, \quad \eta_1 = \omega^2 + \omega^3 + \omega^{10} + \omega^{11}, \quad \eta_2 = \omega^4 + \omega^6 + \omega^7 + \omega^8.$$

Lo que hemos obtenido es que $F(\sigma^3)$ es el espacio vectorial generado por los elementos $1, \eta_0, \eta_1, \eta_2$. Claramente se da la relación $1 + \eta_0 + \eta_1 + \eta_2 = 0$, luego estos elementos no son linealmente independientes. De hecho sabemos que la dimensión de $F(\sigma^3)$ es 3, luego tres cualesquiera de ellos forman una base. Por ejemplo 1, η_0, η_1. Todo elemento de $F(\sigma^3)$ se expresa de forma única como combinación lineal de 1, η_0, η_1 (o de η_0, η_1, η_2). Si lo expresamos como combinación lineal de los primeros, la expresión es única salvo que se puede sumar una cantidad arbitraria a todos los coeficientes sin alterar el elemento.

Ahora vamos con el caso general. Podemos considerar a σ^m como una permutación de las raíces $\omega, \ldots, \omega^{p-1}$. Los valores $\omega, \sigma(\omega), \ldots, \sigma^{p-2}(\omega)$ son todos distintos (en caso contrario sería fácil probar que $\sigma^i(\omega) = \omega$ para $1 \le i < p-1$, luego σ no tendría orden $p-1$. Esto significa que, visto como permutación, σ es un ciclo de longitud $p-1$, concretamente, $\big(\omega, \sigma(\omega), \ldots, \sigma^{p-2}(\omega)\big)$.

Por lo tanto, los valores $\omega, \sigma^m(\omega), \sigma^{2m}(\omega), \ldots, \sigma^{(d-1)m}(\omega)$ son también distintos (son parte de los anteriores), mientras que el siguiente es $\sigma^{dm}(\omega) = \omega$. Esto vale para cualquier raíz primitiva ω y significa que si aplicamos sucesivamente σ^m sobre una raíz primitiva ω, pasamos por d raíces distintas, es decir, σ^m considerada como una permutación de las raíces $\omega, \ldots, \omega^{p-1}$ es un producto de m ciclos disjuntos de longitud d. Llamemos $\eta_0, \eta_1, \ldots, \eta_{m-1}$ a las sumas de los elementos de cada órbita.

Cuando σ^m actúa sobre un elemento de $\mathbb{Q}(\omega)$ expresado como combinación lineal de $1, \omega, \ldots, \omega^{p-1}$, da lugar a otro elemento en el que el coeficiente de una raíz ω pasa a estar junto a la raíz $\sigma^m(\omega)$ y, teniendo en cuenta que el término independiente no varía, para que el elemento sea fijado es necesario que el coeficiente de cada raíz ω coincida con el de $\sigma^m(\omega)$, y a su vez éste ha de coincidir con el de $\sigma^{2m}(\omega)$, etc., lo que significa que todos los coeficientes de una misma órbita deben coincidir, luego pueden sacarse como factores comunes y el elemento puede expresarse como combinación lineal de $1, \eta_0, \eta_1, \ldots, \eta_{m-1}$. (Esto se ve claramente en el ejemplo anterior). Por otro lado los elementos $1, \eta_0, \eta_1, \ldots, \eta_{m-1}$ son fijados por σ^m, luego hemos probado que $F(\sigma^m)$ es el subespacio generado por $1, \eta_0, \eta_1, \ldots, \eta_{m-1}$. No vale la pena repetir en general las condiciones de la unicidad de la expresión.

Los elementos $\eta_0, \eta_1, \ldots, \eta_{m-1}$ se llaman *períodos* de longitud d, pues cada uno es la suma de d raíces primitivas. Cumplen la relación $1+\eta_0+\ldots+\eta_{m-1} = 0$.

Más aún, si la raíz ω está en η_0 y ω^i es otra raíz primitiva, existe un automorfismo σ^j tal que $\sigma^j(\omega) = \omega^i$. Si $\sigma^{km}(\omega)$ es otra raíz de η_0, entonces $\sigma^j\big(\sigma^{km}(\omega)\big) = \sigma^{km}\big(\sigma^j(\omega)\big) = \sigma^{km}(\omega^i)$ está en el mismo período que ω^i, es decir, σ^j envía todas las raíces de η_0 a las raíces del período de ω^i, luego $\sigma^j(\eta_0)$ es el período de ω^i. Con esto hemos probado que los períodos son conjugados.

También es fácil ver que la imagen de un período por un automorfismo es siempre un período.

Por lo tanto, si llamamos concretamente η_0 al período de ω, la sucesión $\eta_0, \sigma(\eta_0), \sigma^2(\eta_0), \ldots, \sigma^{p-1}(\eta_0)$ recorre todos los períodos. Más exactamente: la

sucesión $\eta_0, \sigma(\eta_0), \sigma^2(\eta_0), \ldots, \sigma^{m-1}(\eta_0)$ está formada por todos los períodos. La razón es que si no están todos, alguno se repite, pero en cuanto un período se repite entramos en un ciclo en el que ya no pueden aparecer nuevos períodos.

Por lo tanto numeraremos los períodos con el orden dado por $\eta_i = \sigma^i(\eta_0)$, donde η_0 es el período de ω.

Esta definición es válida para todo entero i, de modo que $\eta_m = \sigma^m(\eta_0) = \eta_0$ (pues σ^m permuta los sumandos de cada período), $\eta_{m+1} = \eta_1$, etc. En general $\sigma^i(\eta_j) = \eta_{i+j}$, para enteros cualesquiera i, j.

Desde esta perspectiva general es fácil hallar los períodos concretos que hemos obtenido en nuestro ejemplo $p = 13$, $m = 3$, $d = 4$:

Recordemos que $\sigma(\omega) = \omega^2$ y $\sigma^3(\omega) = \omega^8$. Para hallar η_0 partimos de ω y le aplicamos repetidas veces σ^3: $\sigma^3(\omega) = \omega^8$, $\sigma^3(\omega^8) = \omega^{12}$, $\sigma^3(\omega^{12}) = \omega^5$, $\sigma^3(\omega^5) = \omega$. Así pues, $\eta_0 = \omega + \omega^5 + \omega^8 + \omega^{12}$. Para obtener η_1 simplemente hacemos

$$\eta_1 = \sigma(\eta_0) = \sigma(\omega + \omega^5 + \omega^8 + \omega^{12}) = \omega^2 + \omega^{10} + \omega^3 + \omega^{11} = \omega^2 + \omega^3 + \omega^{10} + \omega^{11}.$$

Igualmente $\eta_2 = \sigma(\eta_1) = \omega^4 + \omega^6 + \omega^7 + \omega^8$.

Es importante que los períodos no dependen de la elección del generador σ, aunque en general su orden sí depende de esta elección. Un período tiene por sumandos a las raíces a las que se puede llegar desde una dada mediante los automorfismos del subgrupo de orden m, y esto no depende de σ.

Ejercicio: Determinar todos los subcuerpos del cuerpo ciclotómico decimotercero. Indicar las inclusiones entre ellos y, para cada uno, calcular los períodos que lo generan, la tabla del producto de períodos y el subgrupo asociado en el grupo de Galois.

El conocimiento de esta estructura de las extensiones ciclotómicas es importante a la hora de trabajar con ellas. Como ejemplo vamos a ver que simplifican considerablemente el cálculo de normas.

Supongamos que queremos calcular la norma de

$$\alpha = \omega^5 - \omega^4 - 3\omega^2 - 3\omega - 2 \qquad \text{para } p = 7.$$

Una raíz primitiva módulo 7 es 3, luego un generador del grupo de automorfismos es $\sigma(\omega) = \omega^3$, y así

$$\mathrm{N}(\alpha) = \alpha\sigma(\alpha)\sigma^2(\alpha)\sigma^3(\alpha)\sigma^4(\alpha)\sigma^5(\alpha).$$

El truco está en agrupar los factores en la forma:

$$\mathrm{N}(\alpha) = \big(\alpha\sigma^2(\alpha)\sigma^4(\alpha)\big)\big(\sigma(\alpha)\sigma^3(\alpha)\sigma^5(\alpha)\big),$$

con lo que el primer factor es invariante por σ^2 y el segundo es la imagen por σ del primero. Calculamos los términos del primer factor:

$$\begin{aligned} \alpha &= \omega^5 - \omega^4 - 3\omega^2 - 3\omega - 2 \\ \sigma^2(\alpha) &= -3\omega^4 + \omega^3 - 3\omega^2 - \omega - 2 \\ \sigma^4(\alpha) &= \omega^6 - 3\omega^4 - \omega^2 - 3\omega - 2 \end{aligned}$$

Multiplicamos $\alpha\sigma^2(\alpha)$:

$$\begin{array}{rrrrrrr}
0 & 1 & -1 & 0 & -3 & -3 & -2\\
0 & 0 & -3 & 1 & -3 & -1 & -2\\
\hline
0 & -2 & 2 & 0 & 6 & 6 & 4\\
-1 & 1 & 0 & 3 & 3 & 2 & 0\\
3 & 0 & 9 & 9 & 6 & 0 & -3\\
0 & -3 & -3 & -2 & 0 & 1 & -1\\
9 & 9 & 6 & 0 & -3 & 3 & 0\\
\hline
11 & 5 & 14 & 10 & 12 & 12 & 0
\end{array}$$

El producto se ordena como en una multiplicación usual de polinomios, con la única diferencia de que los términos en $\omega^7 = 1$ no se sitúan a la izquierda de los términos en ω^6, sino bajo los términos independientes, e igualmente con exponentes superiores.

El resultado es $11\omega^6+5\omega^5+14\omega^4+10\omega^3+12\omega^2+12\omega$, que puede simplificarse restando 12 a todos los coeficientes, con lo que queda $-\omega^6-7\omega^5+2\omega^4-2\omega^3-12$.

Ahora lo multiplicamos por $\sigma^4(\alpha)$:

$$\begin{array}{rrrrrrr}
1 & 0 & -3 & 0 & -1 & -3 & -2\\
-1 & -7 & 2 & -2 & 0 & 0 & -12\\
\hline
-12 & 0 & 36 & 0 & 12 & 36 & 24\\
0 & 2 & 6 & 4 & -2 & 0 & 6\\
-2 & -6 & -4 & 2 & 0 & -6 & 0\\
21 & 14 & -7 & 0 & 21 & 0 & 7\\
2 & -1 & 0 & 3 & 0 & 1 & 3\\
\hline
9 & 9 & 31 & 9 & 31 & 31 & 40
\end{array}$$

El primer factor es, pues, $40 + 31\eta_0 + 9\eta_1 = 31 + 22\eta_0$.

El segundo se obtiene aplicando σ, luego es $40 + 31\eta_1 + 9\eta_0 = 9 - 22\eta_0$.

Por lo tanto $\mathrm{N}(\alpha) = (31 + 22\eta_0)(9 - 22\eta_0) = 279 - 484\eta_0 - 484\eta_0^2$.

Ahora es fácil calcular $\eta_0^2 = \eta_0 + 2\eta_1 = \eta_0 + 2(-1 - \eta_0) = -2 - \eta_0$. Con esto podemos concluir $\mathrm{N}(\alpha) = 279 - 484\eta_0 + 484\eta_0 + 968 = 1.247$.

No ha sido un cálculo rápido, pero el camino directo supone hacer seis multiplicaciones en lugar de dos. En general el cálculo de normas se simplifica agrupando los factores en dos grupos. Si a su vez el número de factores en un grupo no es primo, se pueden formar grupos dentro de los grupos y el cálculo se reduce más.

15.3 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, pues todo cuerpo de característica 0 contiene a los números racionales, luego es infinito. 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 tan sólo de cardinales que sean potencias de primo.

Sea ahora k un cuerpo finito de cardinal p^n. Entonces por 9.12 el grupo multiplicativo de k es cíclico de orden $p^n - 1$. Esto implica que si u es un elemento no nulo de k, se cumple $u^{p^n-1} = 1$, luego $u^{p^n} = u$, pero esto es válido también si $u = 0$. Así pues, todo elemento de k es raíz del polinomio $x^{p^n} - x$. Vamos a probar que este hecho permite construir, a la vez que caracteriza, los cuerpos de p^n elementos.

Definición 15.14 Llamaremos $\mathbb{A}_p$ a la clausura algebraica del cuerpo $\mathbb{Z}/p\mathbb{Z}$. Llamemos *cuerpo de Galois* de p^n elementos al conjunto $\mathrm{CG}(p^n)$ de las raíces en $\mathbb{A}_p$ del polinomio $x^{p^n} - x$.

El teorema siguiente prueba, entre otras cosas, que $\mathrm{CG}(p^n)$ es realmente un cuerpo de p^n elementos.

Teorema 15.15 *Sea p un número primo y $n > 0$ un número natural. Entonces*

1. *$\mathrm{CG}(p^n)$ es el cuerpo de escición sobre $\mathbb{Z}/p\mathbb{Z}$ del polinomio $x^{p^n} - x$.*
2. *$\mathrm{CG}(p^n)$ es, salvo isomorfismo, el único cuerpo de p^n elementos y es el único subcuerpo de $\mathbb{A}_p$ con p^n elementos.*
3. *$\mathbb{A}_p$ es la unión de todos los cuerpos $\mathrm{CG}(p^n)$. En particular es infinito.*

Demostración: 1) Sólo hay que probar que $\mathrm{CG}(p^n)$ es realmente un cuerpo, pero esto es inmediato, teniendo en cuenta que $(u+v)^p = u^p + v^p$, para todos los $u, v \in \mathbb{A}_p$.

2) Como la derivada formal de $x^{p^n} - x$ es $p^n x^{p^n-1} - 1 = -1$, que no tiene raíces, resulta que las raíces de $x^{p^n} - x$ son todas distintas, luego $\mathrm{CG}(p^n)$ tiene p^n elementos. Según hemos probado, un subcuerpo de $\mathbb{A}_p$ con p^n elementos ha de estar constituido por las raíces de $x^{p^n} - x$, luego ha de ser exactamente $\mathrm{CG}(p^n)$. Todo cuerpo de p^n elementos tiene característica p, es una extensión finita de $\mathbb{Z}/p\mathbb{Z}$, luego es isomorfo a un subcuerpo de $\mathbb{A}_p$, o sea, a $\mathrm{CG}(p^n)$.

3) Si $u \in \mathbb{A}_p$, entonces $(\mathbb{Z}/p\mathbb{Z})(u)$ es un cuerpo finito, luego es $\mathrm{CG}(p^n)$ para cierto número natural n. En particular $u \in \mathrm{CG}(p^n)$. ∎

En particular $\mathrm{CG}(p) = \mathbb{Z}/p\mathbb{Z}$. Al trabajar con cuerpos es más usual la notación $\mathrm{CG}(p)$ que $\mathbb{Z}/p\mathbb{Z}$. Investiguemos ahora los grupos de Galois.

Definición 15.16 Si K es un cuerpo finito se llama *automorfismo de Frobenius* de K a la aplicación $\sigma : K \longrightarrow K$ dada por $\sigma(u) = u^p$.

Es claro que σ es un automorfismo de K (ver la prueba de 8.35).

Teorema 15.17 *Sea $K = \mathrm{CG}(p^n)$ y $k = \mathrm{CG}(p)$. Sea σ el automorfismo de Frobenius de K. Entonces K/k es finita de Galois y $G(K/k) = \langle\sigma\rangle$.*

Demostración: La extensión es de Galois porque los cuerpos son perfectos y K es el cuerpo de escisión sobre K del polinomio $x^{p^n} - x$.

Sabemos que el grupo multiplicativo K^* es cíclico. Sea $K^* = \langle u\rangle$. Entonces $u^{p^n} = 1$, pero $u^{p^m} \neq 1$ 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 es un generador. ∎

Así pues, el grupo $G\big(\mathrm{CG}(p^n)/\mathrm{CG}(p^n)\big)$ es cíclico de orden n, luego tiene exactamente un subgrupo para cada divisor de m. Si $m \mid n$ y H es el subgrupo de orden n/m, o sea, $H = \langle\sigma^m\rangle$, entonces $|\mathrm{CG}(p^n) : F(H)| = n/m$, luego $|F(H) : \mathrm{CG}(p)| = m$. En consecuencia $F(H) = \mathrm{CG}(p^m)$. Con esto hemos probado:

Teorema 15.18 *Sean m y n números naturales no nulos y p un primo. Entonces*

$$\mathrm{CG}(p^m) \subset \mathrm{CG}(p^n) \quad \textit{si y sólo si} \quad m \mid n.$$

Ejercicio: Demostrar que $G\big(\mathrm{CG}(p^n)/\mathrm{CG}(p^m)\big) = \langle\sigma^m\rangle$.

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 $o_m(p)$. Nótese que la observación anterior es un caso particular de este hecho. (Ver la prueba de 12.13.)

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 enmedio. 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 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 capítulos anteriores 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 15.19 *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. ∎

15.4 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 juegan un papel relevante en la teoría de números trascendentes (por ejemplo en la prueba de la trascendencia de π). Aquí usaremos la teoría de Galois para demostrar un resultado en torno a ellos.

Definición 15.20 Sea A un dominio y $\sigma \in \Sigma_n$. Por 2.10 se cumple que la aplicación : $\overline{\sigma} : A[x_1, \ldots, x_n] \longrightarrow A[x_1, \ldots, x_n]$ definida mediante

$$\overline{\sigma}\big(p(x_1, \ldots, x_n)\big) = p(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$$

es un homomorfismo de anillos. De hecho es claro que se trata de un isomorfismo que claramente 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 $\sigma \in \Sigma_n$.

Notar 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 (u,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 grado 3 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 15.21 *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 15.22 *Sea A un dominio, $e_0,\dots,e_n$ los polinomios simétricos elementales en $A[x_1,\dots,x_n]$ y $\bar{e}_0,\dots,\bar{e}_k$ los polinomios simétricos elementales en $A[x_1,\dots,x_k]$, donde $1\le k<n$. Entonces $e_i\in A[\bar{e}_0,\dots,\bar{e}_k,x_{k+1},\dots,x_n]$ para $i=0,\dots,n$.*

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 15.23 *Sea A un dominio y $e_0,\dots,e_n$ los polinomios simétricos elementales en $A[x_1,\dots,x_n]$. Entonces*

$$(x-x_1)\cdots(x-x_n)=\sum_{k=0}^{n}(-1)^{n-k}e_{n-k}(x_1,\dots,x_n)x^k.$$

Demostración: Por inducción sobre n. Para $n=1$ es obvio. Supongámoslo para $n-1$.

Sean $\bar{e}_0,\dots,\bar{e}_{n-1}$ los polinomios simétricos elementales en $A[x_1,\dots,x_{n-1}]$. Entonces

$$\begin{aligned}
&(x-x_1)\cdots(x-x_n)=\sum_{k=0}^{n-1}(-1)^{n-1-k}\ \bar{e}_{n-1-k}x^k(x-x_n)\\
=\ &\sum_{k=0}^{n-1}(-1)^{n-1-k}\ \bar{e}_{n-1-k}x^{k+1}-\sum_{k=0}^{n-1}(-1)^{n-1-k}x_n\bar{e}_{n-1-k}x^k\\
=\ &(-1)^n x_n\bar{e}_{n-1}+x^n+\sum_{k=0}^{n-2}(-1)^{n-1-k}\ \bar{e}_{n-1-k}x^{k+1}\\
&-\sum_{k=1}^{n-1}(-1)^{n-1-k}x_n\bar{e}_{n-1-k}x^k\\
=\ &(-1)^n x_n\bar{e}_{n-1}+x^n+(-1)^{n-k}(\bar{e}_{n-k}+x_n\bar{e}_{n-1-k})x^k\\
(\text{por } 15.21)\ =\ &(-1)^n x_n e_{n-1}+x^n+\sum_{k=1}^{n-1}(-1)^{n-k}e_{n-k}x^k\\
=\ &\sum_{k=0}^{n}(-1)^{n-k}e_{n-k}x^k.
\end{aligned}$$

∎

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 15.24 *Sea A un dominio y sean $e_1,\dots,e_n$ los polinomios simétricos elementales en $A[x_1,\dots,x_n]$. Entonces el anillo de los polinomios simétricos de $A[x_1,\dots,x_n]$ es $A[e_1,\dots,e_n]$.*

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 15.21 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 $e_n\bar{h} = \sigma(h) = \sigma(h) = \sigma(e_n)\sigma(\bar{h}) = e_n\sigma(\bar{h})$, 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 $p(\alpha_1, \ldots, \alpha_n) \in A$. Enunciar casos particulares de este resultado que puedan probarse sin los teoremas anteriores (con teoría de Galois, de enteros algebraicos, etc.)

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 15.25 *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 15.23 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 VIII, 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 $E = k(e_1,\dots,e_n)$. ∎

Capítulo XVI

Módulos finitamente generados

Llegados a este punto está claro que el estudio de los números algebraicos es una combinación de muchas ramas del álgebra: teoría de cuerpos, teoría de anillos, teoría de grupos, álgebra lineal. En estados más avanzados aparecen también el álgebra homológica, así como disciplinas no algebraicas, como la teoría de funciones holomorfas, topología, análisis de Fourier, etc. Ciñéndonos a la parte que conocemos, podemos decir que uno de los teoremas fundamentales que aporta la teoría de anillos es el hecho de que los cuerpos numéricos tienen factorización única en ideales, la teoría de cuerpos aporta el teorema de Galois y ahora vamos a ver un resultado no menos importante en la vertiente del álgebra lineal (si quisiéramos buscar un hecho clave de igual peso por parte de la teoría de grupos deberíamos pensar probablemente en los teoremas de Sylow, que no veremos en este libro). Se trata de un teorema que nos da la estructura de los módulos finitamente generados sobre un dominio euclídeo (en realidad el teorema es válido para DIPs, pero la prueba es algo más técnica). Como ejemplo mostraremos la estructura de los grupos de unidades de los grupos U_n de las unidades módulo n.

16.1 Los teoremas de estructura

El punto de partida es el teorema siguiente:

Teorema 16.1 *Sea A un dominio euclídeo y M un A-módulo finitamente generado. Entonces A es suma directa de un número finito de submódulos monógenos.*

Demostración: Sea $\{a_1, \ldots, a_n\}$ un sistema generador de M. Sea N un A-módulo libre de rango n y sea $\{x_1, \ldots, x_n\}$ una base de N. Por el teorema 7.31 existe un homomorfismo de módulos $f : N \longrightarrow M$ tal que $f(x_i) = a_i$ para $i = 1, \ldots, n$. Obviamente f es un epimorfismo.

Sea $R = \mathrm{N}(f)$. Por el teorema 7.30, R es un módulo libre y podemos encontrar una base $\{y_1, \ldots, y_n\}$ de N y unos elementos $a_1, \ldots, a_m$ de A tales que $\{a_1y_1, \ldots, a_my_m\}$ es una base de R para cierto $m \leq n$.

Basta probar que N/R es suma directa de un número finito de submódulos monógenos, ya que es un módulo isomorfo a M. Concretamente, vamos a probar que $N/R = \langle [y_1] \rangle \oplus \cdots \oplus \langle [y_n] \rangle$.

En efecto, si $b_1, \ldots, b_n \in A$ cumplen $b_1[y_1] + \cdots + b_n[y_n] = 0$, entonces se cumple que $b_1y_1 + \cdots + b_ny_n \in R$, luego $b_1y_1 + \cdots + b_ny_n = c_1a_1y_1 + \cdots + c_ma_my_m$ para ciertos elementos $c_1, \ldots, c_m$ de A.

Como $\{y_1, \ldots, y_n\}$ es base de N las coordenadas son únicas, luego ha de ser $b_i = c_ia_i$ para $i = 1, \ldots, m$ y $b_i = 0$ para $i = m+1, \ldots, n$. En cualquier caso se cumple $b_i[y_i] = 0$ (pues $[a_iy_i] = 0$ para $i = 1, \ldots, m$). Por el teorema 7.15 la suma es directa. ∎

En el capítulo VII vimos que el concepto de suma directa de módulos coincide con el de producto cuando el número de factores es finito. En teoría de grupos finitos se prefiere el término producto directo al de suma directa, con lo que un caso particular del teorema anterior puede enunciarse como sigue:

Teorema 16.2 *Todo grupo abeliano finitamente generado es isomorfo a un producto directo de grupos cíclicos.*

Así por ejemplo, un grupo abeliano de orden 4 puede ser un grupo cíclico o bien el producto directo de dos grupos cíclicos de orden 2. Un ejemplo de cada caso son los grupos $\mathbb{Z}/4\mathbb{Z}$ y $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2\mathbb{Z})$. El primero lo hemos manejado en muchas ocasiones. Si llamamos $1 = \big([0],[0]\big)$, $a = \big([1],[0]\big)$, $b = \big([0],[1]\big)$ y $c = \big([1],[1]\big)$, la tabla del segundo es

	1	a	b	c
1	1	a	b	c
1	a	1	c	b
b	b	c	1	a
c	c	b	a	1

Observamos que no son grupos isomorfos puesto que el producto de dos grupos cíclicos de orden 2 no tiene elementos de orden 4.

Es costumbre usar la notación C_n para referirse a un grupo cíclico cualquiera de orden n (sabemos que dos cualesquiera son isomorfos entre sí). De este modo, un grupo abeliano de orden 4 ha de ser de tipo C_4 o $C_2 \times C_2$. Se distinguen porque en los grupos de tipo C_4 hay elementos de orden 4 y en los de tipo $C_2 \times C_2$ no los hay. Por ejemplo, el grupo de Klein V_4 que definimos en el capítulo IX es de tipo $C_2 \times C_2$.

Quizá el lector piense que un grupo abeliano de orden 6 puede ser de dos tipos diferentes, a saber, C_6 o $C_2 \times C_3$. Esto es falso y la razón nos la da el teorema siguiente. Para enunciarlo en términos de módulos vamos a introducir un concepto que generaliza al de orden de un elemento en un grupo abeliano.

Definición 16.3 Sea A un DIP, M un A-módulo y $m \in M$. Llamaremos *anulador* de m al conjunto $\mathrm{An}(m) = \{a \in A \mid am = 0\}$.

Es claro que se trata de un ideal de A. Como A es DIP existirá un $a \in A$ tal que $\mathrm{An}(m) = (a)$. El elemento a se llama *período* de m, y está unívocamente determinado salvo unidades. Lo representaremos por $\mathrm{o}(m)$.

En el caso de que $A = \mathbb{Z}$ y M es un grupo abeliano, entonces el período de un elemento no es sino su orden, cuando éste es finito, y 0 si el elemento es de orden infinito.

Teorema 16.4 *Sea A un dominio euclídeo y sea M un A-módulo monógeno, $M = \langle m\rangle$. Sea $\mathrm{o}(m) = a_1a_2$, donde $(a_1, a_2) = 1$. Entonces $M = \langle m_1\rangle \oplus \langle m_2\rangle$, para ciertos elementos m_1, m_2 de M tales que $\mathrm{o}(m_1) = a_1$ y $\mathrm{o}(m_2) = a_2$.*

Demostración: Sean $m_1 = a_2m$ y $m_2 = a_1m$. Obviamente tenemos que $\langle m_1\rangle + \langle m_2\rangle \leq \langle m\rangle$. Por la relación de Bezout existen elementos $u, v \in A$ tales que $ua_1 + va_2 = 1$. Por lo tanto $m = ua_1m + va_2m = vm_1 + um_2 \in \langle m_1\rangle + \langle m_2\rangle$, luego $\langle m_1\rangle + \langle m_2\rangle = \langle m\rangle$.

Para probar que la suma es directa nos falta ver que $\langle m_1\rangle \cap \langle m_2\rangle = 0$. Supongamos que $x \in \langle m_1\rangle \cap \langle m_2\rangle$. Entonces $x = ra_2m = sa_1m$ y de aquí resulta que $x = (ua_1 + va_2)x = ua_1ra_2m + va_2sa_1m = ur(am) + vs(am) = 0$.

Así pues, $M = \langle m_1\rangle \oplus \langle m_2\rangle$.

Como $a_1m_1 = a_1a_2m = 0$, tenemos que $(a_1) \subset \mathrm{An}(m_1)$.

Si $b \in \mathrm{An}(m_1)$, entonces $ba_2m = 0$, luego $ba_2 \in \mathrm{An}(m)$, luego $a_1a_2 \mid ba_2$, y por lo tanto $a_1 \mid b$, es decir, $b \in (a_1)$. Esto prueba que $\mathrm{An}(m_1) = (a_1)$ y por tanto que $\mathrm{o}(m_1) = a_1$. Igualmente se obtiene que $\mathrm{o}(m_2) = a_2$. ∎

La traducción a grupos abelianos finitos es la siguiente:

Teorema 16.5 *Sean m y n números naturales tales que $(m, n) = 1$. Entonces*

$$C_{mn} \cong C_m \times C_n.$$

Por lo tanto sólo hay un tipo de grupos abelianos de orden 6, a saber, $C_6 \cong C_2 \times C_3$, o dicho de otro modo, todo grupo abeliano de orden 6 es cíclico.

Teorema 16.6 *Sea A un dominio euclídeo y M un A-módulo. Supongamos que existen elementos m_1 y m_2 en M tales que $M = \langle m_1\rangle \oplus \langle m_2\rangle$, $\mathrm{o}(m_1) = a_1$, $\mathrm{o}(m_2) = a_2$ y $(a_1, a_2) = 1$. Entonces $M = \langle m_1 + m_2\rangle$ y $\mathrm{o}(m_1 + m_2) = a_1a_2$.*

Demostración: Por el teorema de Bezout existen elementos $u, v \in A$ tales que $ua_1 + va_2 = 1$. Entonces $m_1 = ua_1m_1 + va_2m_1 = va_2m_1 = va_2m_1 + va_2m_2 = va_2(m_1 + m_2)$, de donde $m_1 \in \langle m_1 + m_2\rangle$. Igualmente $m_2 \in \langle m_1 + m_2\rangle$, y así $M = \langle m_1 + m_2\rangle$.

Claramente $a_1a_2(m_1 + m_2) = 0$, luego $(a_1a_2) \subset \mathrm{An}(m_1 + m_2)$.

Si $b \in \mathrm{An}(m_1 + m_2)$, entonces $bm_1 + bm_2 = 0$, y como la suma es directa se cumple $bm_1 = bm_2 = 0$, luego $a_1 \mid b$ y $a_2 \mid b$. Como son primos entre sí, $a_1a_2 \mid b$ y $b \in (a_1a_2)$. Esto prueba que $\mathrm{An}(m_1 + m_2) = (a_1a_2)$ y así $\mathrm{o}(m_1 + m_2) = a_1a_2$.

∎

Observemos que si en un grupo abeliano tenemos un elemento u de orden m y otro v de orden n con $(m,n)=1$, acabamos de ver que uv tiene orden mn.

Los teoremas anteriores no son aplicables a elementos de orden 0 (o sea, de orden infinito en el caso de grupos cíclicos). Vamos a probar que estos elementos generan un módulo libre, que quedará completamente descrito por su rango. En esta dirección definimos los conceptos siguientes:

Definición 16.7 Sea A un dominio y M un A-módulo. Un elemento m de M es de *torsión* si existe un $a \in A$ no nulo tal que $am = 0$. En el caso de que A sea un DIP, esto equivale a que el período de m no sea 0.

Llamaremos *submódulo de torsión* de M al conjunto M_t de todos los elementos de torsión. Es inmediato comprobar que M_t es ciertamente un submódulo de M. Se dice que el módulo M es *libre de torsión* si $M_t = 0$.

Los elementos de torsión son los elementos de orden no nulo y que podemos controlar con los teoremas anteriores. Para los módulos que no son de torsión tenemos lo siguiente:

Teorema 16.8 *Sea A un dominio euclídeo y M un A-módulo finitamente generado. Entonces existe un submódulo M' de M de manera que M' es libre y $M = M' \oplus M_t$. El rango de M' no depende de la descomposición y se llama* rango *de M. En particular M es libre si y sólo si es libre de torsión.*

Demostración: Por el teorema 16.1, sabemos que M es suma directa de submódulos monógenos, digamos que $M = \langle y_1\rangle \oplus \cdots \oplus \langle y_n\rangle$. Ordenando los generadores podemos suponer que los r primeros son elementos de torsión y los restantes no. Sean $N = \langle y_1\rangle \oplus \cdots \oplus \langle y_r\rangle$ y $M' = \langle y_{r+1}\rangle \oplus \cdots \oplus \langle y_n\rangle$. Entonces $M = N \oplus M'$.

Es claro que $\{y_{r+1}, \ldots, y_n\}$ es una base de M', pues si

$$a_{r+1}y_{r+1} + \cdots + a_n y_n = 0,$$

como la suma es directa, $a_{r+1}y_{r+1} = \cdots = a_n y_n = 0$, y como los elementos $y_{r+1}, \ldots, y_n$ no son de torsión, $a_{r+1} = \cdots = a_n = 0$. Por lo tanto M' es un módulo libre.

Como $y_1, \ldots, y_r$ son elementos de torsión, tenemos que $N \subset M_t$. Recíprocamente, un elemento $x \in M_t$ se expresará de la forma $x = a_1y_1 + \cdots + a_ny_n$ para ciertos $a_1, \ldots, a_n$ de A. Si $b \in A$, $b \neq 0$ y $bx = 0$, entonces

$$ba_1y_1 + \cdots + ba_ny_n = 0$$

y, como la suma es directa, ha de ser $ba_1y_1 = \cdots = ba_ny_n = 0$. Como $y_{r+1}, \ldots, y_n$ no son de torsión, $ba_{r+1} = \cdots = ba_n = 0$ y como A es un dominio íntegro y $b \neq 0$, se cumple que $a_{r+1} = \cdots = a_n = 0$, luego

$$x = a_1y_1 + \cdots + a_ry_r \in N.$$

Así pues, $N = M_t$.

Observemos que cualquiera que sea la descomposición $M = M' \oplus M_t$, se cumple que $M' \cong M/M_t$, pues la aplicación $f : M \longrightarrow M'$ que a cada $m = m_1 + m_2$ con $m_1 \in M'$ y $m_2 \in M_t$ le asigna $f(m) = m_1$ es un epimorfismo de núcleo M_t. Por lo tanto M/M_t es libre y el rango de cualquier M' es el mismo que el de M/M_t. ∎

Finalmente vamos a clasificar la parte de torsión de un módulo. Si pensamos en grupos finitos la cuestión es saber cuántos tipos distintos de grupos abelianos hay de un orden dado. Por ejemplo de orden 12 sabemos que a lo sumo hay dos, a saber, $C_2 \times C_2 \times C_3$ y C_{12}. Otras posibilidades como $C_4 \times C_3$ las hemos descartado porque $C_4 \times C_3 \cong C_{12}$. Sólo nos falta saber si en efecto los grupos $C_2 \times C_2 \times C_3$ y C_{12} son no isomorfos. El teorema siguiente responde a esta pregunta.

Teorema 16.9 *Sea A un dominio euclídeo y M un A-módulo finitamente generado.*

1. *Existen elementos $x_1, \ldots, x_n \in M$ tales que $M_t = \langle x_1\rangle \oplus \cdots \oplus \langle x_n\rangle$ y para cada $i = 1, \ldots, n$, $\mathrm{o}(x_i) = p_i^{e_i}$, donde p_i es un primo de A y e_i es un número natural no nulo.*

2. *Existen elementos $y - 1, \ldots, y_m \in M$ tales que $M_t = \langle y_1\rangle \oplus \cdots \oplus \langle y_m\rangle$, y si llamamos $b_i = \mathrm{o}(y_i)$, entonces para cada $i = 1, \ldots, m$, se cumple que b_i no es cero ni unidad y si $i < m$, entonces $b_i \mid b_{i+1}$.*

3. *Los números n y m, los $p_i^{e_i}$ para $i = 1, \ldots, n$ y los b_i para $i = 1, \ldots, m$ están determinados por M salvo unidades, 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 b_i se llaman* factores invariantes *de M.*

Demostración: Sabemos que M_t se descompone en suma directa de submódulos monógenos, cuyos generadores son elementos de torsión y podemos tomarlos no nulos, luego sus períodos serán elementos de A no nulos ni unidades. Descomponiéndolos en potencias de primos y aplicando el teorema 16.4 conseguimos una descomposición de tipo 1).

Si ahora agrupamos los sumandos correspondientes a los divisores elementales de todas las bases posibles y los máximos exponentes obtenemos el último factor invariante, repitiendo con los sumandos restantes vamos obteniendo los factores invariantes anteriores.

Por ejemplo, si nos dan el grupo abeliano $C_6 \times C_{18} \times C_{10} \times C_{25} \times C_{15}$, la descomposición en divisores elementales es

$$C_2 \times C_3 \times C_2 \times C_9 \times C_2 \times C_5 \times C_{25} \times C_3 \times C_5.$$

Ahora agrupamos las mayores potencias de primos, es decir, $2 \times 9 \times 25$ y obtenemos un factor C_{450}. Los restantes son $C_3 \times C_2 \times C_2 \times C_5 \times C_3 \times C_5$. Ahora agrupamos $2 \times 3 \times 5$ y obtenemos un factor C_{30}. Quedan $C_2 \times C_3 \times C_5$, que nos dan otro factor C_{30}. En total la descomposición tipo 2) es $C_{30} \times C_{30} \times C_{450}$. Es

claro que siempre es posible obtener una descomposición en factores invariantes a partir de una descomposición en factores elementales. Falta probar que sólo hay un resultado posible. Daremos 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 A. Definimos

$$M_i = \langle y_i\rangle\,, \quad M_i(p) = \{r \in M_i \mid pr = 0\} \quad \text{y} \quad M(p) = \{r \in M_t \mid pr = 0\}.$$

Es obvio que todos ellos son submódulos de M.

Vamos a probar que $M(p) = M_1(p) \oplus \cdots \oplus M_m(p)$. Como cada $M_i(p) \subset M_i$ y la suma de los M_i es directa, es fácil ver que la suma de los $M_i(p)$ también es directa, luego tenemos $M_1(p) \oplus \cdots \oplus M_m(p) \subset M(p)$. Si $r \in M(p)$, entonces $r = u_1y_1 + \cdots + u_my_m$ para ciertos $u_1, \ldots, u_m$ de A. Como $pr = 0$, tenemos que $pu_1y_1 + \cdots + pu_my_m = 0$ y, como la suma es directa, $pu_1y_1 = \cdots = pu_my_m = 0$, luego cada $u_iy_i \in M_i(p)$. Así pues $M(p) = M_1(p) \oplus \cdots \oplus M_m(p)$.

b) El ideal (p) es maximal, luego $A/(p)$ es un cuerpo y $M(p)$ es un espacio vectorial sobre $A/(p)$ con la operación externa dada por $[u]r = ur$.

En efecto, lo único que hay que comprobar es que la operación está bien definida, es decir que si $[u] = [v]$ entonces $ur = vr$, pero si $[u] = [v]$, entonces $p \mid (u - v)$, es decir, $u - v = qp$ para cierto $q \in A$ y $(u - v)r = qpr = 0$, luego $ur = vr$. A partir de aquí se prueban sin problemas las condiciones de espacio vectorial.

La expresión $M(p) = M_1(p) \oplus \cdots \oplus M_m(p)$ vale también como $A/(p)$-espacios vectoriales, pues los $M_i(p)$ son claramente subespacios vectoriales y 0 se sigue expresando de forma única como suma de elementos de cada sumando.

c) $\dim M_i(p) = \begin{cases} 0 & \text{si } p \nmid \mathrm{o}(y_i) \\ 1 & \text{si } p \mid \mathrm{o}(y_i) \end{cases}$

Si $p \nmid \mathrm{o}(y_i)$, entonces un $r \in M_i(p)$ es de la forma $r = uy_i$ para un $u \in A$ y $pr = 0$, luego $puy_i = 0$ y $\mathrm{o}(y_i) \mid pu$, luego $\mathrm{o}(y_i) \mid u$, y así $r = uy_i = 0$, o sea, $M_i(p) = 0$.

Si $\mathrm{o}(y_i) = pv$, para un $v \in A$, entonces razonando igual que antes concluimos que $\mathrm{o}(y_i) \mid pu$, 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 fácil ver que $vy_i \in M_i(p)$ y que es no nulo).

d) Por lo tanto la dimensión de $M(p)$ es igual al número de factores invariantes divisibles entre p.

e) 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). Ahora bien, la definición de $M(p)$ no depende de la descomposición de tipo 2) escogida, luego su dimensión como $A/(p)$-espacio vectorial es igual al número de factores invariantes de la primera descomposición divisibles entre p (o sea, igual a m) y por otra parte es el número de factores invariantes de la segunda

descomposición divisibles entre p (o sea, $\leq m'$). 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, es decir, que 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.

f) En una descomposición de tipo 2), el último factor invariante b_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, es decir, $b_m r = 0$ para todo $r \in M_t$.

El conjunto $\{u \in A \mid ur = 0 \text{ para todo } r \in M_t\}$ es claramente un ideal de A al cual pertenece b_m. Recíprocamente, si u está en este ideal, entonces $uy_m = 0$, luego $b_m \mid u$. Por lo tanto $\{u \in A \mid ur = 0 \text{ para todo } r \in M_t\} = (b_m)$.

De aquí se desprende que dos descomposiciones tipo 2) deben tener igual (salvo unidades) el último factor invariante.

g) 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 $b_1, \ldots, b_m$ y la segunda con factores invariantes $c_1, \ldots, c_m$. Ya sabemos que $b_m = c_m$.

Si b_m se descompone en un solo primo, entonces b_m es primo, y como los restantes factores invariantes son divisores suyos, son todos salvo unidades ese mismo primo, es decir, todos los b_i y los c_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 b_m se descompone en $n+1$ primos. En e) hemos probado que b_1 y c_1 son divisibles entre los mismos primos. Sea p un primo que divida a ambos (luego divide a todos los b_i y a todos los c_i).

Sea $pM = \{pr \mid r \in M_t\}$. Es claro que pM es un submódulo de M y además

$$pM = \langle py_1 \rangle \oplus \cdots \oplus \langle py_m \rangle = \langle pz_1 \rangle \oplus \cdots \oplus \langle pz_m \rangle\,.$$

También es obvio que $\mathrm{o}(py_i) = b_i/p$ y $\mathrm{o}(pz_i) = c_i/p$.

Puede ocurrir que los primeros b_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, donde el último factor invariante es b_m/p, luego podemos aplicar la hipótesis de inducción y concluir que el número de sumandos nulos el igual para las dos descomposiciones, y que las restantes tienen los mismos factores invariantes (salvo unidades), es decir, el número de b_i's iguales a p es el mismo que el de c_i's, y para los restantes, $b_i/p = c_i/p$ (salvo unidades). Esto implica la igualdad (salvo unidades) de los b_i's y los c_i's.

h) La unicidad de los divisores elementales se deduce de la de los factores invariantes, pues es fácil ver que a partir de dos descomposiciones distintas de tipo 1) se pueden conseguir dos descomposiciones distintas de tipo 2). ∎

El teorema siguiente resume la situación de la forma más clara:

Teorema 16.10 *Sea A un dominio euclídeo y M, N dos A-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 $\mathrm{o}(x_i) = \mathrm{o}(y_i)$ para $i = 1, \ldots, m$.

Es claro que la aplicación $f_i : A \longrightarrow \langle x_i\rangle$ dada por $f_i(u) = ux_i$ es un epimorfismo de A-módulos cuyo núcleo es $\big(\mathrm{o}(x_i)\big)$, luego $\langle x_i\rangle \cong A/\big(\mathrm{o}(x_i)\big)$ y lo mismo vale para los sumandos $\langle y_i\rangle$, luego $\langle x_i\rangle \cong \langle y_i\rangle$ para $i = 1, \ldots, m$.

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, $M \cong N$. ∎

Ahora ya podemos concluir que $C_2 \times C_2 \times C_3$ y C_{12} son no isomorfos, pues los divisores elementales del primero son $(2, 2, 3)$, y los del segundo son $(4, 3)$. Alternativamente, los factores invariantes del primero son $(2, 6)$ y los del segundo son (12).

Un hecho importante que hemos obtenido en esta prueba es que en un grupo abeliano finito G existe un elemento cuyo orden es múltiplo del orden de todos los demás elementos de G, concretamente, un elementos cuyo orden sea el último factor invariante. En general, al mínimo común múltiplo de los órdenes de los elementos de un grupo finito G se le llama *exponente* de G (y es siempre un divisor de $|G|$). Lo que hemos probado es que en un grupo abeliano finito, el exponente es el último factor invariante y por lo tanto es el orden de un elemento de G, cosa que en general no es cierta. Por ejemplo el grupo Σ_3 tiene elementos de orden 2 y 3, luego su exponente es 6, pero no tiene elementos de orden 6.

Esta observación sobre el exponente nos proporciona una prueba alternativa del teorema 9.12: Si G es un subgrupo finito del grupo multiplicativo de un dominio íntegro D, sea n el exponente de G. Entonces todo elemento de G es raíz del polinomio $x^n - 1$, luego G tiene a lo sumo n elementos, y como hay un elemento de orden n, concluimos que G es cíclico. ∎

Ejercicio: Probar que un grupo abeliano finito posee subgrupos de todos los órdenes que dividen al orden del grupo.

Ejercicio: Probar que si G es un grupo abeliano finito y p es un primo que divide al orden de G, entonces G tiene un elemento de orden p.

16.2 La estructura de los grupos de unidades

Seguidamente vamos a obtener la estructura de unos grupos abelianos importantes, los grupos U_n de las unidades de los anillos $\mathbb{Z}/n\mathbb{Z}$, que según sabemos son los grupos de Galois de los cuerpos ciclotómicos. Es importante notar que aquí no vamos a necesitar los teoremas de estructura que acabamos de probar. No necesitamos teoremas generales para obtener la estructura de grupos particulares. Lo que estos teoremas garantizan es que los resultados que vamos a obtener sobre grupos de unidades tienen análogos en cualquier grupo abeliano finito. En el apéndice A se encuentra una aplicación interesante de los teoremas de estructura a la teoría de cuerpos.

El hecho más elemental es que si p es un primo el grupo U_p consta de todas las clases no nulas de $\mathbb{Z}/p\mathbb{Z}$ y es cíclico. Lo probamos esencialmente en 5.13, aunque más en general tenemos el resultado de 9.12, según el cual todo subgrupo finito del grupo multiplicativo de un dominio íntegro es cíclico.

Los teoremas de estructura nos dan una prueba más elegante de este mismo hecho: si G es un subgrupo finito del grupo multiplicativo de un cuerpo K y su exponente es m, entonces el polinomio x^m-1 tiene tantas raíces como elementos tiene G, luego $|G| \leq m$, y la otra desigualdad es obvia, pues G es abeliano y tiene un elemento g de orden m. Por lo tanto $|G| = m$ y $G = \langle g \rangle$.

Recordemos que las raíces primitivas de la unidad de $\mathbb{Z}/p\mathbb{Z}$, es decir, los generadores de U_p, se llaman raíces primitivas módulo p. Consideremos ahora los grupos U_{p^e}, donde p es primo.

Teorema 16.11 *Sea p un primo impar y $e \geq 2$. Entonces el grupo U_{p^e} es cíclico y un generador es $[r(p+1)]$, donde r es una cierta raíz primitiva módulo p (elegida adecuadamente).*

Demostración: Puesto que el orden de U_{p^e} es $\phi(p^e) = p^{e-1}(p-1)$, basta probar que $[r]$ tiene orden $p-1$ y $[p+1]$ tiene orden p^{e-1}.

Sea s una raíz primitiva módulo p cualquiera. Entonces definimos

$$r = s^{p^{e-1}} \equiv s \ (\text{mód } p),$$

con lo que r también es una raíz primitiva módulo p. Además

$$[r]^{p-1} = [s]^{p^{e-1}(p-1)} = [1],$$

pues el orden de cualquier elemento divide al orden del grupo.

Si $[r]^m = [1]$, entonces $p^e \mid r^m - 1$, luego $p \mid r^m - 1$, luego $p-1 \mid m$ (pues r tiene orden $p-1$ módulo p). Esto prueba que el orden de $[r]$ es $p-1$.

Veamos por inducción sobre $e \geq 2$ que $(1+p)^{p^{e-2}} \equiv 1 + kp^{e-1} \pmod{p^e}$, para un cierto k (dependiente de e) tal que $k \not\equiv 0 \pmod{p}$.

Para $e = 2$ se cumple con $k = 1$.

Supuesto cierto para e, tenemos $(1+p)^{p^{e-2}} = 1 + kp^{e-1} + ap^e = 1 + tp^{e-1}$, donde $t = k + ap \not\equiv 0 \pmod{p}$. Entonces

$$(1+p)^{p^{e-1}} = (1+tp^{e-1})^p = 1 + ptp^{e-1} + \cdots + t^p p^{p(e-1)}.$$

Como $p \geq 2$ y $e \geq 2$, esta expresión es de la forma $1 + tp^e + bp^{e+1}$, luego $(1+p)^{p^{e-1}} \equiv 1 + tp^e \pmod{p^{e+1}}$, con $t \not\equiv 0 \pmod{p}$.

Por lo tanto al tomar clases módulo p^e resulta $[1+p]^{p^{e-1}} = [1]$, luego el orden de la clase $[1+p]$ ha de ser un divisor de p^{e-1}. Si el orden fuera menor que p^{e-1}, entonces tendríamos que $[1+p]^{p^{e-2}} = [1]$, o sea, $[kp^{e-1}] = [0]$, luego $p \mid k$, contradicción. Por lo tanto el orden de $[1+p]$ es exactamente p^{e-1}. ∎

El teorema anterior no es válido para $p = 2$. Obviamente $U_2 = 1$ y $U_4 \cong C_2$. Para potencias mayores la situación es la siguiente:

Teorema 16.12 *Si $e \geq 3$, entonces $U_{2^e} \cong C_2 \times C_{2^{e-2}}$. Un generador del primer factor es $[-1]$ y un generador del segundo es $[5]$, es decir, todo elemento de U_{2^e} se expresa de forma única como $\pm[5]^j$, donde $j = 0, \ldots, 2^{e-2} - 1$.*

El orden de U_{2^e} es 2^{e-1}. Claramente $[-1]$ tiene orden 2. Una simple inducción muestra que para $e \geq 3$

$$5^{2^{e-3}} = (1+2^2)^{2^{e-3}} \equiv 1 + 2^{e-1} \pmod{2^e},$$

de donde $[5]^{2^{e-3}} \neq [1]$, mientras que

$$5^{2^{e-2}} \equiv (1+2^{e-1})^2 \equiv 1 \pmod{2^e},$$

luego el orden de $[5]$ es 2^{e-2}.

Por otra parte $[-1]$ no es una potencia de $[5]$, ya que si se cumpliera $2^e \mid 5^r + 1$ para algún r, entonces $4 \mid 5^r + 1$, pero módulo 4 es $[5]^r = [1]^r = [1] \neq [-1]$.

De este modo, los subgrupos $\langle[-1]\rangle$ y $\langle[5]\rangle$ tienen intersección trivial, luego su suma es directa. Así pues, $\langle[-1]\rangle \times \langle[5]\rangle$ tiene 2^{e-1} elementos y por consiguiente es todo el grupo U_{2^e}. ∎

Para completar la clasificación de los grupos de unidades recordamos un hecho que ya usamos en el capítulo V para probar que la función de Euler es multiplicativa (y que es consecuencia inmediata del teorema 5.10):

Teorema 16.13 *Si m y n son primos entre sí, entonces $U_{mn} \cong U_m \times U_n$.*

Con esto conocemos la estructura de todos los grupos de unidades. Por ejemplo

$$U_{300} \cong U_4 \times U_3 \times U_{25} \cong C_2 \times C_2 \times C_{20}.$$

Como aplicación vamos a usar esta información para determinar los restos cuadráticos módulo enteros arbitrarios, no necesariamente primos.

Teorema 16.14 *Sea p un primo impar y $e \geq 2$. Entonces un entero m es un resto cuadrático módulo p^e si y sólo si es un resto cuadrático módulo p.*

DEMOSTRACIÓN: Una implicación es obvia. Supongamos que m es un resto cuadrático módulo p. Por definición m es una unidad módulo p^e, luego por el teorema 16.11 se cumple que $m \equiv r^i(1+p)^i \pmod{p^e}$, donde r es una cierta raíz primitiva módulo p.

Tomando clases módulo p queda $[m] = [r]^i$. Como $[m]$ es un resto cuadrático módulo p existe un entero s tal que $[r]^i = [s]^2$. Como r es una raíz primitiva, existe un entero j tal que $[s] = [r]^j$, luego $[r]^i = [r]^{2j}$, de donde resulta que $i \equiv 2j \pmod{p-1}$, por lo que i es par, $i = 2k$.

Así $m \equiv r^{2k}(1+p)^{2k} \equiv (r^k(1+p)^k)^2 \pmod{p^e}$, es decir, m es un resto cuadrático módulo p^e. ∎

La situación para $p = 2$ es, como siempre, distinta.

Teorema 16.15 *Un entero impar m es un resto cuadrático módulo 4 si y sólo si $m \equiv 1 \pmod 4$, y es un resto cuadrático módulo 2^e con $e \geq 3$ si y sólo si $m \equiv 1 \pmod 8$.*

DEMOSTRACIÓN: El caso $e = 2$ es claro. Si $e \geq 3$, entonces por el teorema 16.12, $m \equiv \pm 5^i \pmod{2^e}$.

Si m es un resto cuadrático, entonces $m \equiv r^2 \pmod{2^e}$ para cierto entero r, que también será de la forma $r \equiv \pm 5^j \pmod{2^e}$, luego $r^2 \equiv 5^{2j} \pmod{2^e}$, y por lo tanto $\pm 5^i \equiv 5^{2j} \pmod{2^e}$.

La unicidad de la expresión exige que el signo sea positivo y que $i = 2j$ (si tomamos los exponentes reducidos módulo 2^{e-2}).

Por lo tanto $m \equiv 5^{2j} \pmod{2^e} \equiv 25^j \pmod 8 \equiv 1 \pmod 8$.

Recíprocamente, si $m \equiv 1 \pmod 8$, entonces $\pm 5^i \equiv 1 \pmod 8$, y como se cumple que $5^2 \equiv 1 \pmod 8$, las potencias de 5 sólo son congruentes con 1 y con 5 módulo 8, luego el signo ha de ser positivo e i ha de ser par, $i = 2j$. Entonces $m \equiv (5^j)^2 \pmod{2^e}$, luego es un resto cuadrático. ∎

El resultado siguiente es consecuencia inmediata del isomorfismo definido en el teorema 5.10 y de los teoremas anteriores.

Teorema 16.16 *Sea $m = p_1^{e_1} \cdots p_r^{e_r}$, donde $p_1, \ldots, p_r$ son primos distintos. Entonces un entero n es un resto cuadrático módulo m si y sólo si es un resto cuadrático módulo $p_i^{e_i}$ para $i = 1, \ldots, r$. Si m es impar esto equivale a que n sea un resto cuadrático módulo p_i para $i = 1, \ldots, r$.*

Ahora queda claro lo que dijimos en el capítulo XIV sobre que el hecho de que un símbolo de Jacobi (m/n) valga 1 no implica que m sea un resto cuadrático módulo n, pues $(2/15) = (2/3)(2/5) = (-1)(-1) = 1$, pero precisamente porque $(2/3) = (2/5) = -1$, tenemos que 2 no es un resto cuadrático módulo 15.

Capítulo XVII

Resolución de ecuaciones por radicales

En el capítulo V (teorema 5.20) vimos que las soluciones de una ecuación de segundo grado $ax^2 + bx + c = 0$ con coeficientes a, b y c en un cuerpo de característica distinta de 2 pueden obtenerse mediante la fórmula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a},$$

donde $\sqrt{b^2 - 4ac}$ representa a una raíz cuadrada del elemento $b^2 - 4ac$, es decir, un elemento cuyo cuadrado es $b^2 - 4ac$.

Esta fórmula es conocida desde la antigüedad, mientras que una fórmula similar que permita resolver ecuaciones polinómicas de tercer grado no fue hallada hasta el siglo XVI, por el matemático Cardano (si bien Tartaglia afirmaba que fue él quien la encontró y se la comunicó a Cardano bajo palabra de no revelarla). La fórmula de Cardano es demasiado complicada para que resulte de utilidad práctica, por lo que no vamos a demostrarla. No obstante es instructivo conocer su aspecto:

Dada una ecuación cúbica

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

con $a \neq 0$, una de sus raíces tiene la forma $x = u + v$, donde

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \quad \text{y} \quad v = -\frac{p}{3u} \quad \text{(en el supuesto de que } u \neq 0.\text{)}$$

a su vez

$$p = \frac{3ac - b^2}{3a^2} \qquad \text{y} \qquad q = \frac{2b^3 - 9abc + 27a^2d}{27a^3}.$$

Las dos soluciones restantes se obtienen cambiando la elección de la raíz cúbica que define a u. En caso de que $u = 0$, los valores de v se obtienen mediante otra expresión del mismo estilo.

Vemos, pues, que las soluciones de la ecuación $ax^3+bx^2+cx+d=0$ pueden obtenerse a partir de los coeficientes a, b, c y d mediante fórmulas consistentes en sumas, restas, productos, cocientes y extracción de raíces.

Posteriormente el matemático Ferrari obtuvo la solución (todavía más compleja) de las ecuaciones polinómicas de grado 4, o sea, obtuvo una expresión similar de las raíces de una ecuación polinómica de grado 4 en función de sus coeficientes.

Con esto quedaba planteado el problema de resolver la ecuación general de grado n, es decir, la ecuación

$$a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = 0.$$

Se entiende que resolver la ecuación de grado n significa encontrar una fórmula que exprese sus raíces en términos de sumas, restas, productos, cocientes y extracción de raíces a partir de sus coeficientes $a_0, \ldots, a_n$. Sin embargo en este capítulo probaremos que el teorema de Ferrari es el mejor resultado posible en este terreno, que no existe tal fórmula para polinomios de grado ≥ 5.

17.1 Extensiones radicales

El primer paso en el problema que hemos planteado es dar una definición algebraica precisa del concepto de 'expresabilidad mediante raíces'.

Definición 17.1 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.

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).

En estos términos, las raíces de la ecuación son $\omega^i\alpha$ y $\omega^i\beta$, para $i = 1,\dots,5$, 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, notar que ciertamente toda extensión radical es finita.

Teorema 17.2 *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 $K = k(a_1,\dots,a_n)$ según la definición 17.1. Llamemos $p_i(x) = \operatorname{pol\,mín}(a_i,k)$. 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 $K_1,\dots,K_r$ 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 $a_{i1},\dots,a_{in}$ de manera que $K_i = k(a_{i1},\dots,a_{in})$ y se cumpla la definición de extensión radical. Es fácil ver que tomando $N = k(a_{11},\dots,a_{1n},\dots,a_{r1},\dots,a_{rn})$ 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 $K = k(a_1,\dots,a_n)$ es una extensión radical, podemos considerar la cadena de cuerpos intermedios $K_i = k(a_1,\dots,a_i)$, 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

$$K_i = K_{i-1}(a_i), \qquad \text{con } a_i^{r_i} \in K_{i-1}. \tag{17.1}$$

Si además la extensión es de Galois podemos considerar los grupos asociados $G_i = G(K/K_i)$, que forman una cadena

$$1 = G_n \le G_{n-1} \le \cdots \le G_1 \le G(K/k). \tag{17.2}$$

Ahora la clave es que vamos a probar que, salvo una restricción técnica, la condición (17.1) equivale a que

$$G_i \trianglelefteq G_{i-1} \qquad \text{y} \qquad G_{i-1}/G_i \text{ es cíclico.} \tag{17.3}$$

De este modo, tendremos que una extensión finita de Galois K/k es radical si y sólo si el grupo $G(K/k)$ tiene una sucesión de subgrupos (17.2) que cumplan (17.3). Esta caracterización algebraica es fácil de comprobar si se conoce el grupo de Galois y, como veremos, nos permitirá resolver el problema que nos hemos planteado.

Teorema 17.3 *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) $\to$ 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 15.2 existe un $w \in K$ tal que

$$v = \sum_{i=0}^{d-1} \eta^i \sigma^i(w) \neq 0.$$

Entonces $\eta\sigma(v) = \eta \sum_{i=0}^{d-1} \eta^i \sigma^{i+1}(w) = \sum_{i=0}^{d-1} \eta^{i+1}\sigma^{i+1}(w) = v$. Llamando $u = v^{-1}$ tenemos que $\sigma(u) = \eta u$.

En general, $\sigma^i(u) = \eta^i u$, luego los elementos $\eta^i u$ para $i = 0, \ldots, d-1$ 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 $i = 1, \ldots, d-1$, 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) $\to$ 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) $\to$ 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 $i = 1, \ldots, n$. 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 elemento 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.

Ahora vemos la restricción técnica de la que hablábamos antes y que nos impide pasar directamente de (17.1) a (17.3) y viceversa. Para sortear este obstáculo necesitaremos algo de teoría de grupos.

17.2 Grupos resolubles

Las consideraciones de la sección anterior llevan de forma natural a la definición siguiente:

Definición 17.4 Un grupo finito G es *resoluble* si existe una sucesión de subgrupos de G

$$1 = G_0 \trianglelefteq G_1 \trianglelefteq \cdots \trianglelefteq G_n = G \tag{17.4}$$

tal que cada cociente G_{i+1}/G_i sea un grupo abeliano.

En general, una sucesión de subgrupos (17.4) se llama una *serie* de G. Los subgrupos G_i se llaman *términos* de la serie y los cocientes G_{i+1}/G_i se llaman *factores* de la serie. Una serie es *abeliana, cíclica,* etc. si todos sus factores son grupos abelianos, cíclicos, etc. En estos términos, un grupo es resoluble si tiene una serie abeliana.

El nombre de 'resoluble' se debe a que, según veremos, un polinomio es resoluble por radicales si y sólo si su grupo de Galois es resoluble. Según lo dicho en la sección anterior, quizá el lector piense que deberíamos haber definido un grupo resoluble como un grupo con una serie cíclica. Sucede que es equivalente —después lo probaremos—, pero la definición que hemos dado resulta ser más manejable.

Ejercicio: Admitiendo que un grupo es resoluble si y sólo si tiene una serie cíclica, probar que si K/k es una extensión de Galois de grado n y k contiene una raíz n-sima primitiva de la unidad, entonces K/k es radical si y sólo si $G(K/k)$ es resoluble.

Es inmediato que todo grupo abeliano G es resoluble, pues $1 \trianglelefteq G$ es ya una serie abeliana de G.

Sin embargo hay grupos resolubles que no son abelianos. Por ejemplo una serie abeliana de Σ_3 viene dada por $1 \trianglelefteq A3 \trianglelefteq \Sigma_3$. Los factores son $A_3/1 \cong C_3$ y $\Sigma_3/A_3 \cong C_2$.

Una serie abeliana de Σ_4 es $1 \trianglelefteq V_4 \trianglelefteq A_4 \trianglelefteq S_4$. Ahora $V_4/1 \cong C_2 \times C2$, $A_4/V_4 \cong C_3$ y $\Sigma_4/A_4 \cong C_2$ (téngase en cuenta que un grupo de orden primo ha de ser cíclico).

Más adelante veremos ejemplos de grupos no resolubles. Informalmente podemos decir que la mayoría de los grupos finitos son resolubles. El menor grupo no resoluble tiene 60 elementos y es el único grupo no resoluble de orden menor que 120.

Una de las razones por las que la mayoría de los grupos son resolubles es que la resolubilidad se conserva por la mayoría de las operaciones que pueden realizarse con grupos. El teorema siguiente da cuenta de ello. Al mismo tiempo, las propiedades que vamos a ver son las que nos permitirán esquivar las raíces de la unidad en la caracterización de las extensiones radicales.

Teorema 17.5 *Se cumple:*

1. *Si G es un grupo resoluble y $H \leq G$, entonces H es resoluble.*
2. *Si G es un grupo resoluble y $N \trianglelefteq G$, entonces G/N es resoluble.*
3. *Si G es un grupo y N es un subgrupo normal de G tal que N y G/N son resolubles, entonces G es resoluble.*
4. *Si H y K son subgrupos resolubles de un grupo G y $H \trianglelefteq G$, entonces HK es resoluble. En particular el producto directo de grupos resolubles es resoluble.*

Demostración: 1) Sea $1 = G_0 \trianglelefteq G_1 \trianglelefteq \cdots \trianglelefteq G_n = G$ una serie abeliana de G. Entonces

$$1 = G_0 \cap H \trianglelefteq G_1 \cap H \trianglelefteq \cdots \trianglelefteq G_n \cap H = H$$

es una serie abeliana de H, pues, por el segundo teorema de isomorfía,

$$\begin{aligned}(G_{i+1} \cap H)\Big/(G_i \cap H) &= (G_{i+1} \cap H)\Big/\big(G_i \cap (G_{i+1} \cap H)\big)\\ &\cong G_i(G_{i+1} \cap H)\Big/G_i \leq G_{i+1}/G_i,\end{aligned}$$

y como el último grupo es abeliano, el primero también lo es. Por lo tanto H es resoluble.

2) Si $1 = G_0 \trianglelefteq G_1 \trianglelefteq \cdots \trianglelefteq G_n = G$ es una serie abeliana de G entonces

$$1 = G_0N/N \trianglelefteq G_1N/N \trianglelefteq \cdots \trianglelefteq G_nN/N = G/N$$

es una serie abeliana de G/N, pues por los teoremas de isomorfía

$$\begin{aligned}\big(G_{i+1}N/N\big)\Big/(G_iN/N) &\cong G_{i+1}N\Big/G_iN = G_{i+1}(G_iN)\Big/G_iN\\ &\cong G_{i+1}\Big/(G_{i+1} \cap G_iN)\\ &\cong (G_{i+1}/G_i)\Big/\big((G_{i+1} \cap G_iN)/G_i\big),\end{aligned}$$

y el último grupo es un cociente del grupo abeliano G_{i+1}/G_i. Por lo tanto G/N es resoluble.

3) Sean

$$1 = N_o \trianglelefteq N_1 \trianglelefteq \cdots \trianglelefteq N_n = N$$

y

$$1 = H_o/N \trianglelefteq H_1/N \trianglelefteq \cdots \trianglelefteq H_m/N = G/N$$

series abelianas de N y G/N respectivamente. Entonces una serie abeliana de G es claramente

$$1 = N_0 \trianglelefteq N_1 \trianglelefteq \cdots \trianglelefteq N_n = H_0 \trianglelefteq H_1 \trianglelefteq \cdots \trianglelefteq H_m = G.$$

4) Por el teorema de isomorfía, $HK/H \cong K/(H \cap K)$, que es resoluble por ser cociente de K. Como H también es resoluble, por el apartado anterior HK también lo es. ∎

Ahora ya podemos probar que, como anticipábamos, los grupos resolubles tienen series cíclicas. La idea es tomar series lo más largas posibles.

Diremos que una serie de un grupo es *estricta* si cada término es un subgrupo estricto del siguiente, es decir, si ningún factor es trivial.

Dado un grupo no trivial G, la serie estricta $1 \lhd G$ puede extenderse hasta llegar a una serie estricta $1 = G_0 \lhd G_1 \lhd \cdots \lhd G_n = G$ en la que ya no sea posible insertar más términos. No es posible insertar términos indefinidamente porque el producto de los órdenes de los factores de una serie es igual al orden de G, luego el número de factores nunca puede superar el orden de G.

El hecho de que entre dos términos $G_i \lhd G_{i+1}$ no sea posible insertar ningún subgrupo, o sea, que no haya subgrupos H tales que $G_i \lhd H \lhd G_{i+1}$, equivale, según el teorema 9.27, a que ningún factor G_{i+1}/G_i posea subgrupos normales $1 \lhd H/G_i \lhd G_{i+1}/G_i$.

Un grupo no trivial G que no posea subgrupos normales propios es un grupo *simple.* Acabamos de probar que todo grupo no trivial posee una serie estricta cuyos factores son simples.

Por el tercer teorema de isomorfía cualquier factor de cualquier serie de un grupo resoluble es resoluble, luego un grupo resoluble tiene una serie estricta formada por grupos simples y resolubles a la vez. Ahora bien:

Teorema 17.6 *Un grupo finito es simple y resoluble si y sólo si es cíclico de orden primo.*

Demostración: Un grupo cíclico de orden primo no tiene subgrupos propios, luego es simple, y es abeliano, luego es resoluble.

Si un grupo G es simple y resoluble su única serie es $1 \lhd G$, luego ha de ser una serie abeliana, pero entonces G es abeliano. Como todos los subgrupos de un grupo abeliano son normales sólo puede ser simple si no tiene subgrupos.

Por definición de grupo simple G no es trivial, luego existe un $g \in G$ no trivial. Entonces $\langle g\rangle \neq 1$, y como G no tiene subgrupos propios, $G = \langle g\rangle$, o sea, G es cíclico. Pero un grupo cíclico posee subgrupos para todos los divisores del orden del grupo, luego el orden de G ha de ser un número primo. ∎

Consecuentemente hemos probado:

Teorema 17.7 *Si G es un grupo resoluble, entonces G tiene una serie cuyos factores son grupos cíclicos de orden primo.*

Con esto ya podemos caracterizar las extensiones radicales, pero todavía necesitaremos un hecho más para demostrar que la ecuación general de grado n no es resoluble por radicales cuando $n \geq 5$. La causa reside fundamentalmente en el teorema que probamos a continuación:

Teorema 17.8 *Si $n \geq 5$ entonces el grupo alternado A_n es un grupo simple y no abeliano. En particular A_n y Σ_n no son resolubles.*

Demostración: Dividimos la prueba en varios pasos:

1) A_n está generado por los ciclos de longitud 3.

Consideremos un producto de dos trasposiciones. Ha de ser de la forma $(a,b)(c,d)$ o bien de la forma $(a,b)(a,c)$.

En el primer caso $(a,b)(c,d) = (a,b,c)(c,a,d)$.

En el segundo $(a,b)(a,c) = (a,b,c)$.

Todo elemento de A_n se puede expresar como producto de un número par de trasposiciones, y cada par de trasposiciones se puede sustituir por un producto de ciclos de longitud 3, luego toda permutación de A_n se expresa como producto de ciclos de longitud 3.

2) Si un subgrupo $N \trianglelefteq A_n$ contiene un ciclo de longitud 3, entonces $N = A_n$.

En efecto, sea $(a,b,c) \in N$. Si σ es cualquier otro ciclo de longitud 3, por el teorema 9.18 existe una permutación $\tau \in \Sigma_n$ tal que $(a,b,c)^\tau = \sigma$.

Si $\tau \in A_n$ entonces $\sigma = (a,b,c)^\tau \in N^\tau = N$.

Si τ es impar, $(a,b,c) = (c,b,a)^{-1} = (c,b,a)^{(a,c)}$, y $\sigma = (c,b,a)^{(a,c)\tau}$, donde ahora la permutación $(a,c)\tau$ es par y, como $(c,b,a) = (a,b,c)^{-1} \in N$, concluimos igualmente que $\sigma \in N$.

Por lo tanto N contiene a todos los ciclos de longitud 3, que generan A_n, luego $N = A_n$.

3) A_5 es simple.

Si $N \trianglelefteq A_5$ y $N \neq 1$, un elemento de N ha de ser de la forma (a,b,c), (a,b,c,d,e) o bien $(a,b)(c,d)$.

Si N contiene un elemento de la forma (a,b,c) ya hemos visto que $N = A_5$.

Si N contiene un elemento de la forma (a,b,c,d,e), entonces

$$(a,b,c,d,e)(a,b,c,d,e)^{(a,b)(d,e)} = (a,b,c,d,e)(b,a,c,e,d) = (b,e,c) \in N,$$

luego también $N = A_5$.

Si N contiene un elemento de la forma $(a,b)(c,d)$, sea e distinto de a,b,c,d. Entonces

$$(a,b)(c,d)\big((a,b)(c,d)\big)^{(a,b,e)} = (a,b)(c,d)(b,e)(c,d) = (a,e,b) \in N,$$

luego también $N = A_5$. Por lo tanto A_5 es simple.

4) Finalmente probamos que A_{n+1} es simple suponiendo que lo es A_n.

Sea $N \trianglelefteq A_{n+1}$ y $N \neq 1$. Notemos que una permutación como $(1,2,3)$ puede considerarse como permutación de A_3 o de A_4, etc. Con precisión, podemos identificar A_n con el conjunto de las permutaciones de A_{n+1} que dejan fijo a $n+1$. Así $A_n \leq A_{n+1}$.

Tenemos que $N \cap A_n \trianglelefteq A_n$, y por hipótesis de inducción es simple, luego $N \cap A_n = 1$ o bien $N \cap A_n = A_n$. Si se da el segundo caso, entonces N contiene todos los ciclos de longitud 3 de A_n, luego $N = A_{n+1}$.

Veamos que no puede ocurrir $N \cap A_n = 1$. En tal caso sea $\sigma \in N$, $\sigma \neq 1$.

Sea $j = \sigma(n+1) \neq n+1$. La permutación σ no puede ser una trasposición porque sería impar. Tampoco puede ser un ciclo de longitud 3 porque entonces $N = A_{n+1}$. En consecuencia podemos encontrar índices $k \neq l$ distintos de j y de $n+1$ tales que $\sigma(k) = l$ (en la expresión de σ como producto de ciclos disjuntos aparecerán al menos 4 índices).

Sea $\tau = \sigma^{(n+1,j)(k,l,u,v)}$, donde u y v son elementos distintos de los anteriores (estamos en el caso en que $n \geq 6$). Como hemos conjugado por una permutación par, tenemos que $\sigma\tau \in N$.

Como $\sigma(n+1) = j$, se cumple que $\tau(j) = n+1$ y como $\sigma(k) = l$, también $\tau(l) = u$. Por lo tanto $(\sigma\tau)(n+1) = n+1$, mientras que $(\sigma\tau)(k) = u$, luego $\sigma\tau \neq 1$, pero $\sigma\tau \in N \cap A_n$, contradicción.

El hecho de que A_n es no abeliano para $n \geq 5$ es consecuencia inmediata de que A_n contiene a A_4, que es un grupo no abeliano:

$$(1,2,3)(1,2) = (2,3), (1,2)(1,2,3) = (1,3).$$

∎

Puede probarse que A_5 es el menor grupo simple no abeliano y también el menor grupo no resoluble. También hemos visto que Σ_n es resoluble para $n \leq 4$.

La teoría de grupos resolubles va mucho más allá del alcance de este libro. Aunque no nos va a hacer falta más adelante demostraremos un último resultado sobre estos grupos porque involucra un concepto básico de la teoría general de grupos que es conveniente conocer. La serie descrita en el teorema 17.7 tiene la mayor longitud posible. Tiene tantos factores como primos aparecen en el orden de G. Su interés reside en que sus factores son los más sencillos posibles, pero a veces interesa todo lo contrario, es decir, trabajar con una serie abeliana de longitud mínima. Para construirla partiremos de G y en cada paso tomaremos el menor subgrupo que podamos sin que el cociente deje de ser abeliano.

Definición 17.9 Sea G un grupo. Llamaremos *subgrupo derivado* de G al subgrupo G' generado por los elementos de la forma $[x,y] = x^{-1}y^{-1}xy$, para todos los x, $y \in G$.

El elemento $[x,y]$ se llama *conmutador* de x, y. Su nombre se debe a que obviamente $xy = yx[x,y]$. En particular $[x,y] = 1$ si y sólo si $xy = yx$, es decir, si y sólo si x e y conmutan.

El teorema siguiente afirma que G' es el menor subgrupo normal de G cuyo cociente es abeliano.

Teorema 17.10 *Sea G un grupo. Entonces*

1. *$G' \trianglelefteq G$ y G/G' es un grupo abeliano.*
2. *Si $N \trianglelefteq G$, entonces G/N es abeliano si y sólo si $G' \leq N$.*

Demostración: 1) Es inmediato que si x, y, $g \in G$ entonces $[x,y]^g = [x^g, y^g]$. Los elementos de G' son productos de conmutadores, luego sus conjugados también, es decir, $G'^g \leq G'$, lo que prueba que G' es normal.

Claramente el cociente G/G' es abeliano, pues si xG', yG' son dos de sus elementos, tenemos que $x^{-1}y^{-1}xy \in G'$, luego $(xG')(yG') = (yG')(xG')$.

2) Si $G' \leq N$, entonces $G/N \cong (G/G')/(N/G')$, que es un grupo abeliano por ser un cociente de un grupo abeliano.

Si G/N es abeliano entonces para todo x, $y \in G$ se cumple $(xN)(yN) = (yN)(xN)$, luego $[x,y] = x^{-1}y^{-1}xy \in N$ y, como los conmutadores generan G', concluimos que $G' \leq N$. ∎

En particular un grupo G es abeliano si y sólo si $G' = 1$. Un grupo simple no abeliano ha de cumplir $G' = G$, pues ha de ser $G' \neq 1$ y no hay más subgrupos normales. Otra propiedad evidente es que si $H \leq G$, entonces $H' \leq G'$.

Definimos el *derivado* n-simo $G^{n)}$ de un grupo G mediante

$$G^{0)} = G, \qquad G^{n+1)} = \left(G^{n)}\right)'.$$

Claramente se cumple

$$\cdots \quad G^{n+1)} \quad \trianglelefteq \quad G^{n)} \quad \trianglelefteq \quad \cdots \quad \trianglelefteq \quad G^{1)} \quad \trianglelefteq \quad G^{0)} \quad = \quad G.$$

Si existe un número n tal que $G^{n)} = 1$ entonces la serie derivada

$$1 = G^{n)} \quad \trianglelefteq \quad \cdots \quad \trianglelefteq \quad G^{1)} \quad \trianglelefteq \quad G^{0)} \quad = \quad G.$$

es abeliana, luego G es resoluble. Recíprocamente, si G es resoluble y

$$1 = H_n \quad \trianglelefteq \quad \cdots \quad \trianglelefteq \quad H_1 \quad \trianglelefteq \quad H_0 = G$$

es una serie abeliana de G, aplicando repetidamente el teorema anterior obtenemos que $G^{i)} \leq H_i$ para $i = 1, \ldots, n$. En efecto, si vale para i, como H_i/H_{i+1} es abeliano tenemos que $G^{i+1)} = \left(G^{i)}\right)' \leq H_i' \leq H_{i+1}$.

Así concluimos que $G^{n)} = 1$ y que la longitud de la serie derivada es menor o igual que la longitud de la serie dada. La serie derivada es, pues, la menor serie abeliana de un grupo resoluble.

17.3 Caracterización de las extensiones radicales

Por fin estamos en condiciones de probar el teorema (debido a Galois) que caracteriza las extensiones radicales en términos de su grupo de automorfismos. Recordemos que el único problema era que hasta ahora teníamos que suponer que el cuerpo base contenía una raíz de la unidad adecuada. Aunque es posible dar resultados más finos, para no tener problemas con la existencia de raíces primitivas, trabajaremos con cuerpos de característica cero.

Teorema 17.11 *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 17.2 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 17.1.

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 15.6 la extensión L_n/L_0 es finita de Galois y $G(L_n/L_0) \cong G\Big(K/\big(K \cap k(\omega)\big)\Big)$.

Llamemos $b_i = a_i^{r_i} \in K_{i-1}$. Por el teorema 17.3 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

$$H_i/H_{i-1} = G(L_n/L_i)\Big/G(L_n/L_{i-1}) \cong G(L_i/L_{i-1})$$

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 abeliana. Además

$$G\big(\big(K \cap k(\omega)\big)\Big/k\big) \cong G\big(k(\omega)/k\big)\Big/G\Big(K/\big(K \cap k(\omega)\big)\Big)$$

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 $G\Big(K/\big(K\cap k(a)\big)\Big)$ 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 17.12 *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

$$G = G\big(K(\omega)/k(\omega)\big) \cong G\Big(\big(K/\big(K\cap k(\omega)\big)\big)\Big) \leq G(K/k)$$

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

$$r_i = \big|F_i : F_{i-1}\big| \;\Big|\; \big|K(\omega) : k(\omega)\big| = \big|G\big(K(\omega)/k(\omega)\big)\big|\,\big|G(K/k)\big| = n.$$

Podemos aplicar el teorema 17.3, 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 17.13 *(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.*

Si $p(x)$ es un polinomio de grado menor o igual que 4, su grupo de Galois es isomorfo a un subgrupo del grupo de las permutaciones de sus raíces, es decir, de Σ_n para $n \leq 4$, luego es un grupo resoluble. Por lo tanto todas las ecuaciones de grado menor o igual que 4 son resolubles por radicales (y la forma concreta de resolverlas nos la dan los teoremas de Cardano y Ferrari).

Es obvio que existen ecuaciones de grado superior a 4 resolubles por radicales. Hemos visto un ejemplo tras la definición 17.1. Sin embargo sucede que no todas lo son. No es fácil encontrar ejemplos concretos con los métodos de los que disponemos. Aunque no vamos a demostrarlo, lo cierto es que la ecuación $x^5 - 4x + 2 = 0$ no es resoluble por radicales.

17.4 La ecuación general de grado n

Como suele suceder en teoría de números, es más fácil obtener resultados generales que estudiar casos particulares, y así, más fácil que mostrar ejemplos concretos de ecuaciones no resolubles por radicales es probar que la ecuación general de grado n no es resoluble por radicales cuando $n \geq 5$, es decir, que no existen fórmulas similares a las de Cardano y Ferrari que nos expresen las raíces de una ecuación polinómica arbitraria a partir de sus coeficientes mediante una expresión radical. Para plantear el problema con precisión necesitamos una definición algebraica de 'ecuación general'.

Definición 17.14 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 17.15 *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 sean $u_1, \ldots, u_n$ 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 15.23 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 $k(a_0, \ldots, a_{n-1})$ de $p_n(x)$ y $k(x_1, \ldots, x_n)$ es el cuerpo de escisión sobre $k(e_1, \ldots, e_n)$ de $(x - x_1) \cdots (x - x_n)$, es claro que los respectivos grupos de Galois deben ser isomorfos, es decir,

$$G\big(F/k(a_0, \ldots, a_{n-1})\big) \cong G\big(k(x_1, \ldots, x_n)/k(e_1, \ldots, e_n)\big).$$

El primero es el grupo de Galois de $p_n(x)$ y el segundo es isomorfo a Σ_n por el teorema 15.25. ∎

Como consecuencia inmediata tenemos:

Teorema 17.16 *(Abel) La ecuación general de grado no es resoluble por radicales para $n \geq 5$.*

Apéndice A

El teorema de la base normal

En este apéndice 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 1 *Si K/k es una extensión finita de Galois, el cuerpo k es infinito y $G(K/k) = \{\sigma_1, \ldots, \sigma_n\}$, 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 $\sigma_1(u), \ldots, \sigma_n(u)$ 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 2 *Si K/k es una extensión finita de Galois y el cuerpo k es infinito entonces K/k tiene una base normal.*

Demostración: Sea $G(K/k) = \{\sigma_1, \ldots, \sigma_n\}$ y sea v según el teorema anterior. Basta probar que los conjugados $\sigma_1(v), \ldots, \sigma_n(v)$ forman una k-base de K. De hecho basta ver que son linealmente independientes. Supongamos que existen elementos $a_1, \ldots, a_n \in k$ tales que

$$a_1\sigma_1(v) + \cdots + a_n\sigma_n(v) = 0.$$

Aplicamos σ_j^{-1} para $j = 1, \ldots, n$ y obtenemos un sistema de ecuaciones de la forma

$$\sum_{i=1}^{n} a_i(\sigma_i\sigma_j^{-1})(v) = 0, \qquad j = 1, \ldots, n.$$

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 el capítulo XV, estas extensiones son cíclicas, luego el caso restante está incluido en el teorema próximo, con el que concluye la prueba.

Teorema 3 *Toda extensión cíclica tiene una base normal.*

Demostración: Sea $G(K/k) = \langle\sigma\rangle$. Vamos a dotar a K de estructura de módulo sobre el anillo $k[x]$. Para ello definimos

$$(a_m x^m + \cdots + a_1 x + a_0)\alpha = a_m\sigma^m(\alpha) + \cdots + a_1\sigma(\alpha) + a_0\alpha.$$

Es fácil ver que con este producto (y su suma) K es ciertamente un $k[x]$-módulo.

Más aún, es un módulo finitamente generado, pues una k-base de K es también un generador de K como $k[x]$-módulo (notar que el producto por elementos de k como subconjunto de $k[x]$ coincide con el producto como k-espacio vectorial).

Como $k[x]$ es un dominio de ideales principales podemos usar los teoremas de estructura vistos en el capítulo XVI.

Si $n = |K : k|$ tenemos que $\sigma^n = 1$, es decir, $\sigma^n(\alpha) - \alpha = 0$, o equivalentemente, $(x^n - 1)\alpha = 0$ para todo $\alpha \in K$. Esto significa que todos los elementos de K son de torsión, luego el teorema 16.9 nos da la descomposición

$$K = \langle y_1 \rangle \oplus \cdots \oplus \langle y_m \rangle ,$$

donde, si llamamos $p_i(x) = \mathrm{o}(y_i)$, los p_i son polinomios no constantes y cumplen $p_i \mid p_{i+1}$. Es obvio entonces que el polinomio p_m anula a todos los elementos de K. Su grado no puede ser menor que n, pues en tal caso, si $p_m(x) = a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, 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 15.2 (es fácil ver que de hecho $p_m(x) = x^n - 1$).

Ahora podemos probar que los conjugados de y_m forman una base normal, es decir, son linealmente independientes. Si se cumpliera que

$$a_{n-1}\sigma^{n-1}(y_m) + \cdots + a_1\sigma(y_m) + a_0 y_m = 0,$$

para ciertos elementos $a_i \in k$, entonces tendríamos

$$(a_{n-1}x^{n-1} + \cdots + a_1 x + a_0)y_m = 0,$$

es decir, $q(x)y_m = 0$ para un cierto polinomio $q(x)$ de grado menor que n. Entonces $\mathrm{o}(y_m) \mid q(x)$, y comparando los grados ha de ser $q(x) = 0$, o sea, $a_{n-1} = \cdots = a_0 = 0$. ∎

El lector con conocimientos adicionales de álgebra lineal se habrá dado cuenta de que la prueba se reduce a demostrar que el polinomio mínimo de σ es $x^n - 1$, con lo que el submódulo generado por y_m tiene dimensión n y es, por lo tanto, todo K, y por otra parte es sabido que una base de dicho módulo es la formada por $y_m, \sigma(y_m), \ldots, \sigma^{n-1}(y_m)$.

Apéndice B

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 este apéndice. 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 todo este apéndice supondremos que todos los cuerpos son de característica prima p. En primer lugar veremos cómo factorizan los polinomios en el caso general.

Definición 1 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 tal que $f(x) = g(x^{p^n})$, 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 2 *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 8.33 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 8.32, 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 3 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 4 *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 5 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 8.40 se sigue fácilmente que K_s es un cuerpo, y el teorema 4 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 6 *Sea K/k una extensión algebraica. Entonces la extensión K/K_s es puramente inseparable y K/K_p es separable.*

DEMOSTRACIÓN: Sea $a \in K$. Sea $p(x) = \text{pol mín}(a, k)$. 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 2 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 4 tenemos que K/K_s es puramente inseparable.

Por otra parte, sea L la clausura normal de K sobre k. Entonces K_p es el cuerpo fijado de la extensión L/k. En efecto, un elemento de L está en el cuerpo fijado de L/k si y sólo si él es su único k-conjugado, si y sólo si es la única raíz de su polinomio mínimo sobre k, si y sólo si está en K_p. Obviamente K_p es también el cuerpo fijado de L/K_p y por el teorema 8.37 L/K_p es separable, y K/K_p también. ∎

Así pues, tenemos las cadenas

$$k \subset K_s \subset K \qquad \text{y} \qquad k \subset K_p \subset K.$$

La primera tiene el primer tramo separable y el segundo puramente inseparable, y la segunda tiene el primer tramo puramente inseparable y el segundo separable. Veamos ahora el comportamiento de los monomorfismo con relación a estas descomposiciones. Nos basamos en el teorema siguiente:

Teorema 7 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 8.19 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 8 *Sea K/k una extensión algebraica. Entonces*

$$K = K_s K_p \qquad y \qquad k = K_s \cap Kp.$$

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 $K/K_s K_p$ es separable y puramente inseparable, luego $K = K_s K_p$.

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 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.

Terminamos con algunas observaciones sencillas: toda extensión puramente inseparable es normal, y si K/k es una extensión normal entonces

$$G(K/k) = G(K/K_p) \cong G(K_s/k), \qquad G(K_p/k) = 1.$$

Apéndice C

La resultante

Dedicamos este apéndice a estudiar los conceptos de resultante y discriminante, que son útiles al profundizar en el estudio de los polinomios.

Definición 1 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 $A = \mathbb{Z}[u_0, \ldots, u_n, v_0, \ldots, v_m]$ 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* $R \in \mathbb{Z}[u_0, \ldots, u_n, v_0, \ldots, v_m]$ 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 $g(x)$, concluimos 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 n en $s_1, \ldots, s_n$ y cada u_j (para $j \geq 1$) es un polinomio de grado 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 mn en cada una de las variables s_i, t_i. Como lo mismo es cierto para el producto que aparece en la igualdad anterior, 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 derecho, 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

$$f = a_0(x - a_1) \cdots (x - a_n), \quad g = b_0(x - b_1) \cdots (x - b_m).$$

El homomorfismo $\mathbb{Z}[u_0, v_0, s_1, \ldots, s_n, t_1, \ldots, t_m] \longrightarrow k$ dado por

$$u_0 \mapsto a_0,\ v_0 \mapsto b_0,\ s_i \mapsto a_i,\ t_i \mapsto b_i$$

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

$$R(f, g) = a_0^m b_0^n \prod_{i,j} (a_i - b_j) = a_0^m \prod_{i=1}^{n} g(a_i).$$

Ahora es inmediato el teorema siguiente:

Teorema 2 *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

$$f(x) = a_0 \prod_{i=1}^{n} (x - a_i),$$

entonces

$$f'(x) = a_0 \sum_{i=1}^{n} \prod_{j\neq i} (x - a_j).$$

y, según hemos visto,

$$\begin{aligned} R(f,f') &= a_0^{2n-1} \prod_{i=1}^{n} f'(a_i) = a_0^{2n-1} \prod_{i\neq j} (a_i - a_j) \\ &= (-1)^{n(n-1)/2} a_0^{2n-1} \prod_{i<j} (a_i - a_j)^2 \end{aligned}$$

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 3 Si $f(x)$ es un polinomio de grado n con coeficientes en un cuerpo k, definimos su *discriminante* como

$$\Delta(f) = (-1)^{n(n-1)/2} R(f,f')/a_0 = a_0^{2n-2} \prod_{i<j} (a_i - a_j)^2,$$

donde $a_1, \ldots, a_n$ son las raíces de $f(x)$ en una clausura algebraica 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 4 *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

$$\Delta(f) = -\frac{1}{a} \begin{vmatrix} a & b & c \\ 2a & b & 0 \\ 0 & 2a & b \end{vmatrix} = b^2 - 4ac.$$

■

Bibliografía

[1] Baker, A. *Breve introducción a la teoría de números.* Alianza Ed., Madrid, 1986.

[2] Bastida, J.R. *Field extensions and Galois theory.* Addison-Wesley P.C., California, 1984.

[3] Borevich, Z.I., Shafarevich, I.R. *Number Theory.* Academic Press, New York, 1967.

[4] Edwards, H. M. *Fermat's last theorem.* Springer-Verlag, New York, 1977.

[5] Hungerford, T.W. *Algebra.* Springer-Verlag, New York, 1974.

[6] Lang, S., *Algebra* Addison-Wesley P.C., California, 1965.

[7] Stewart, I. y Tall, D. *Algebraic number theory.* Chapman and Hall, Londres, 1979.

Índice de Tablas

Índice de Materias