Universidad Nacional

Creación de algunas alternativas para el mejoramiento de la educación matemática

Myriam Acevedo
Contrato 208-93

Informe final

Santa Fe de Bogotá

1997

0.1 Componente experimental

La componente experimental del proyecto presente en toda la investigación consideró dos momentos: consulta a docentes acerca de experiencias e inquietudes relativas a su propio conocimiento matemático así como a la transposición de éste al aula de clase; implementación de algunas unidades del texto en cursos regulares con docentes en formación y en cursillos con docentes en ejercicio. Nos referiremos brevemente a la encuesta para discutir sus resultados mas sobresalientes.

En ella se pretendía explorar tanto la profundidad del conocimiento del docente como la transposición didáctica (las formas en que se pone este conocimiento al servicio de una mejor docencia). Las respuestas a las preguntas abiertas mostraron diferentes niveles de dominio. Una pregunta sencilla como la siguiente:

1. Considere los datos siguientes: El producto de dos números enteros es -24 y una de las diferencias entre ellos es -11. Una pregunta natural que surge, mirando estos datos es: ¿Cuáles son los dos números? ¿Qué otras preguntas y que tengan solución única, se pueden contestar con base en estos datos?

suscitó respuestas variadas, unas totalmente adecuadas (como cuál es el cociente de los dos números), otras parcialmente adecuadas (producen preguntas con solución única y otras que no tienen solución única, otras que muestran falta de comprensión del enunciado y finalmente unos que resuelven el problema propuesto ($-3, 8$ y $3, -8$) sin responder la pregunta planteada que explora qué mas puede construir el profesor con base en la situación planteada.

Por otra parte, se plantean situaciones hipotéticas de aula explorando los nexos que el docente establece entre conocimientos del álgebra superior y problemas típicos del álgebra de la secundaria. Un ejemplo se da a continuación.

1. Un estudiante, confrontado en una previa de selección múltiple con varias funciones polinómicas, marcó las siguientes de entre las alternativas (A, B, C, D) que se le presentaron:

(i) $p(x) = x^3 + x + 1$	X (B) no tiene ceros reales.
(ii) $q(x) = x^2 - 7x + 13$	X (C) tiene ceros reales porque el producto de éstos es positivo.
(iii) $t(x) = x^3 - 3x^2 - 27x - 7$	X (A) no puede factorizarse en los enteros porque 7 es un número primo.
(iv) $s(x) = x^4 - 2x^3 - 5x + 2$	X (D) No se puede factorizar en los enteros porque no tiene ceros enteros.

(a) En cada caso decir si el estudiante está en lo correcto o no y dar razones para cada una de sus decisiones.

(b) En los casos en que Ud. opina que el estudiante está equivocado, intente explicar lo que estaba pensando el estudiante que podría haber causado el error.

Especialmente interesante resultaron las respuestas a las partes (i) y (iv) por los argumentos que usaron los docentes para justificar sus respuestas. En el primer caso varios (un 28%) docentes citaron los teoremas apropiados para concluir que el estudiante estaba equivocado. Un porcentaje similar aplicó la división sintética con 1 y -1, para afirmar que el estudiante tenía la razón. El 22% de los docentes encuestados no contestó mientras que otro 22% afirman que el estudiante está en lo correcto sin explicar porqué.

En cuanto a la parte (iv), tan sólo el 3% identificó el error del estudiante. El 97% contestó que el estudiante tenía la razón argumentando de nuevo con una incorrecta utilización de la división sintética. Esto es especialmente sorprendente dada la facilidad con que se puede factorizar la expresión, pues $x^4 - 2x^3 - 5x + 2 = \left(x^2 + x + 2\right)\left(x^2 - 3x + 1\right)$.

Aquí se aprecia la diagramación final del Capítulo 1 del texto cuyo título es Recorrido del álgebra: de la teoría de ecuaciones al álgebra abstracta.

También se incluyen las versiones finales del resto del texto, capítulos 2 - 10, introducción y bibliografía.

Maiteracta

INTRODUCCION

Este texto, que está pensado básicamente para formación de docentes de matemáticas de nivel medio, es el resultado principal del proyecto de investigación "Creación de alternativas para el mejoramiento de la enseñza del Algebra" que bajo el auspicio de Colciencias y de la Universidad Nacional de Colombia desarrollamos entre 1993 y 1996.

En el, hemos intentado plasmar nuestras concepciones acerca de lo que debería ser la formación de los docentes de la educación básica y media en esta área. En primer lugar, en estas páginas se explicitan significativamente nexos entre la "teoría formal" del álgebra moderna y las "nociones elementales" de la aritmética y el álgebra de la secundaria. Un cuidadoso examen histórico permite mostrar, por ejemplo, cómo los temas y métodos del álgebra escolar evolucionaron hasta transformarse en la llamada álgebra moderna. En segundo lugar, la estructura y planteamientos (que incluyen problemas de diversos niveles, puntos de discusión y puntos de investigación) en el texto están orientados a desarrollar y enriquecer el pensamiento matemático de los docentes en formación. El desarrollo histórico del conocimiento algebraico revela la actividad y el pensamiento matemáticos en estado de evolución; se enfrenta al lector por ejemplo, a analizar cómo el desarrollo de algoritmos para solucionar ecuaciones abrió caminos hacia la construcción de significado y hacia la generalidad. En tercer lugar, aquí se abordan a fondo los nexos entre el álgebra por una parte, y la teoría de números y el cálculo por otra. En el contexto de métodos de solución de ecuaciones se exploran también los nexos entre el álgebra y la geometría, y un estudio de la solución de ecuaciones por aproximación desemboca como es natural en la generación de métodos numéricos. Es importante anotar, además, que al interior del texto se plantean algunas alternativas metodológicas en la presentación de los diferentes temas. Finalmente, característica fundamental del texto es que se invita a un proceso continuo de *construir, profundizar y ampliar.*

El libro consta de diez capítulos que van desde la teoría de ecuaciones al álgebra moderna y su eje central es el problema de la solubilidad de ecuaciones polinómicas que estuvo, y sigue estando, significativamente ligada al planteamiento y solución de problemas fundamentales de la matemática.

En el Capítulo 1 se trabaja la solución de ecuaciones lineales en una o dos variables, tema aparentemente muy elemental pero que, tratado con profundidad, ilustra la potencia de los sistemas numéricos que condicionan la posibilidad de modelar y resolver este tipo de ecuaciones. Se ilustra allí por primera vez en el texto (con la matemática egipcia) cómo los métodos utilizados para resolver ecuaciones en diferentes etapas de la historia, así como el contexto en el cual se plantean, revelan en todas las etapas tanto conocimiento, como ignorancia u otro tipo de limitación, por lo demás muy similar a la situación en el aula de clase.

El Capítulo 2 está dedicado a estudiar las ecuaciones cuadráticas que surgen de manera natural en dos contextos, tanto a partir de problemas aritméticos como de problemas geométricos. Se muestra cómo la interrelación entre estos dos contextos marcó todo el desarrollo, no sólo de la teoría de ecuaciones cuadráticas, sino de la matemática misma. Se incluye una discusión de la solución de ecuaciones cuadráticas en enteros que permite explorar varios de los temas tradicionalmente asociados con la teoría de números.

En el Capítulo 3 se aborda la solución de ecuaciones cúbicas y cuárticas fundamentalmente desde los enfoques geométrico y algebraico, pero se inicia una exploración de los métodos de aproximación que permitirá posteriormente enfocar la solución desde un punto de vista análitico.

Dado que en los capítulos anteriores se ha trabajado a fondo la solución de ecuaciones polinómicas de grado menor que cinco, en el capítulo 4 se desarrolla una teoría elemental de ecuaciones lo mas general posible, partiendo de un cambio fundamental de orientación dado por Viète que impulsa el álgebra hacia nuevos niveles de abstracción con la introducción de coeficientes literales y que abre la posibilidad de representaciones generales.

En el Capítulo 5 se analiza el papel de los métodos numéricos en la solución de ecuaciones polinómicas. Se examina la etapa de aritmetización de la matemática y se discute el significado a nivel teórico tanto de métodos numéricos de aproximación como de solución por computador, además de plantear sus ventajas y limitaciones.

En los Capítulos 6, 7 y 8, con el ánimo de enriquecer la experiencia matemática del futuro profesor y hacer mas vívida su comprensión de ideas formales y abstractas, se analizan elementos y argumentos de la teoría de números involucrados en la construcción de la teoría de grupos y la teoría de anillos; se hace énfasis por ejemplo en resultados de estas dos teorías que son generalizaciones de las propiedades familiares de los números enteros, así como en propiedades de estas estructuras abstractas que se sustentan en propiedades de los números enteros y que aprovechan la riqueza de su estructura. Se contrastan además argumentos usados para demostrar propiedades en la aritmética de enteros con argumentos usados en grupos y anillos generales.

En el Capítulo 6 se discuten todos los aspectos fundamentales de la teoría de grupos, tomando como referente inicial los sistemas numéricos y enfatizando el análisis de grupos familiares (de aritméticas modulares, de transformaciones geométricas, de funciones, etc).

Partiendo de modelos familiares de anillos. a saber, los números enteros, los polinomios y los enteros modulares, se construye en el capítulo 7 una teoría general de anillos haciendo énfasis en determinar bajo qué condiciones un anillo tiene estructura similar a la de nuestros modelos (dominio de integridad, dominio euclidiano, dominio de factorización única,...) y en construir estructuras cada vez mas completas.

El capítulo 8 está dedicado a trabajar un anillo que, a pesar de constituir para

nosotros un modelo, no se analizó de manera general en el capítulo anterior, a saber, el anillo de los polinomios. El problema central a considerar es la factorización de polinomios con coeficientes en un campo y en particular la factorización de polinomios con coeficientes en los racionales. Se analizan criterios disponibles para decidir acerca de la irreducibilidad. Se culmina con una caracterización de los ideales de estos anillos que permite construir nuevos campos a partir de adjunciones formales, logrando una primera aproximación a los campos de descomposición.

En el capítulo 9, dedicado al estudio los números complejos y el Teorema Fundamental del Algebra se explora a fondo el nexo entre la teoría de ecuaciones polinómicas y el planteamiento y solución de problemas fundamentales del análisis. Allí se observa cómo la paulatina construcción de significado para los números complejos, que surgen de la solución algebraica de ecuaciones polinómicas, se constituyó en punto de apoyo para la caracterización de las raíces y así mismo permitió garantizar la existencia de raíces para tales ecuaciones, lo cual no sólo da solidez a la teoría de ecuaciones sino proporciona solución a problemas petenecientes al cálculo, como la integración por fracciones parciales o la solución de ecuaciones diferenciales.

El último capítulo, que presenta algunos elementos de la teoría de cuerpos en torno a la consideración de la pregunta ¿Es soluble por radicales una ecuación polinómica de grado mayor o igual que cinco?, da un balance del objetivo milenario del álgebra de dar solución a las ecuaciones polinómicas, mostrando que la noción de factorización resulta ser la mas generalizable. En el transcurso de esta evaluación, se enlazan los dos tópicos discutidos a lo largo del libro, la teoría de ecuaciones y las estructuras algebraicas. Se inicia con una discusión de las ecuaciones polinómicas de grado mayor que cuatro, que en general no pueden ser abordadas algebraicamente como las de grados inferiores; se genera formalmente en este estudio la noción de extensiones algebraicas que desemboca en la construcción de una teoría general de factorización de polinomios, con el objeto de garantizar la existencia de campos de descomposición para un polinomio dado. Se concluye construyendo el puente fundamental entre la teoría de cuerpos y la teoría de grupos que se resume en el Teorema Fundamental de Galois.

La estructura del texto procura que el estudiante logre una sólida construcción personal del conocimiento por medio de los puntos de discusión. No se debe avanzar en la lectura y el estudio sin dedicar pensamiento serio a éstos. El profesor puede encontrar en ellos, además, puntos centrales alrededor de los cuales impulsar la discusión en clase para así contar con una efectiva participación de los alumnos. En los problemas allí planteados también se profundizan los elementos que enlazan temas tomados de diferentes áreas de la matemática.

Los puntos de investigación, como su nombre lo indica, invitan al lector a seguir indagando por su cuenta, especialmente en cuanto a las ramificaciones que algunas de las ideas presentadas puedan tener.

No faltan los ejercicios y otros problemas de aplicación inmediata o de práctica, que se encuentran comunmente en los textos, pero los problemas que se encuentran al final de cada capítulo tienen otra finalidad. Explicitan los nexos que hay entre

los temas estudiados y la matemática escolar, facilitando así la transposición que permite que el futuro docente logre un dominio de la matemática superior que en efecto enriquecerá e informará su entendimiento de la matemática elemental.

Capítulo 1

Ecuaciones lineales

En esta sección introductoria queremos considerar la solución de ecuaciones lineales en una o dos variables. A primera vista parece un tema supremamente elemental, pero nuestro estudio muestra que no es así y que proporciona una oportunidad para desarrollar el pensamiento matemático y analizar la potencia de los sistemas numéricos y de ciertas ideas matemáticas.

La misma forma en que sea apropiado modelar una ecuación dada esta ligada con el sistema numérico y la aritmética con los cuales se está trabajando y los procedimientos y métodos que se pueden usar para resolverla. Esta coherencia del pensamiento es indispensable para lograr una comprensión amplia y profunda del tema, al tiempo que juega un papel clave en la solución de problemas relacionados con las ecuaciones lineales. En las secciones que siguen tendremos presentes todas estas facetas del tema.

1.1 Un nuevo examen de temas conocidos

Sabemos que desde las primeras experiencias con la adición en la escuela primaria se comienza por proponer problemas como los siguientes.

-En la siguiente adición el cuadrado está tapando un número. ¿Cuál es el número que está tapado?

$$4 + \square = 7. \tag{1.1}$$

Para responder la pregunta, se espera que el alumno utilice la experiencia repetida con ciertas sumas o que aplique un método como el "seguir contando a partir de 4 hasta 7" ...5...6...7 para llegar a la respuesta, 3. También podría usar el ensayo y el error, colocando un valor en el cuadrado, sumando este valor con 4, observando el resultado y haciendo la respectiva corrección en el valor colocado en el cuadrado. Otra alternativa sería la de restar 4 de ambos miembros de la ecuación, para obtener $\square = 3$.

El modelo de la balanza

Si se busca lograr una comprensión amplia y profunda de esta situación por parte del alumno en el nivel básico, un primer paso en el tratamiento de estas ecuaciones debe ser el de construir un modelo adecuado, y resulta que a nuestro alcance tenemos uno sorprendentemente sencillo. Se trata del modelo de una balanza de brazos que se encuentra en equilibrio. Por ejemplo, para modelar la ecuación 1.1 tendríamos una situación en la cual en uno de los brazos de la balanza se encuentra un objeto desconocido y cuatro pesas iguales mientras que en el otro lado hay siete de las mismas pesas y la balanza se encuentra en equilibrio (Figura 1.1).

Figura 1.1

Para averiguar entonces cuánto pesa el objeto desconocido podríamos proceder de muchas formas diferentes.

A pesar de ser muy sencillo, este modelo puede explorarse de muchas maneras, cada una de las cuales promueve el pensamiento matemático autónomo. Veamos algunas posibilidades. Después de resolver los siguientes problemas, discutir el kilometraje matemático que se puede extraer de la solución de cada uno de ellos.

Problemas

1. Un ladrillo está en equilibrio en una balanza con $\frac{3}{4}$ de ladrillo y una pesa de $\frac{3}{4}$ de libra. ¿Cuánto pesa el ladrillo?

2. ¿Cómo se pueden dividir $180g$ de cebada en dos porciones, una que contiene $40g$ y la otra $140g$, usando una balanza sólo tres veces, si se cuenta con dos pesas, una de $1g$ y una de $4g$?

3. Una botella y un vaso están en equilibrio en una balanza con una jarra. La botella sola está en equilibrio con un vaso y un plato, mientras que tres platos están en equilibrio con dos jarras. ¿Cuántos vasos estarán en equilibrio con una botella?

4. En un cierto supermercado, se tiene una balanza defectuosa, pues los dos platos no están a la misma distancia del fulcro (Figura 1.2).

Figura 1.2

El tendero sugiere que es posible medir dos libras de azucar de la siguiente manera. Se coloca una pesa de una libra en el platillo de la izquierda y se añade azucar en el de la derecha hasta que estén en equilibrio. Luego, se coloca una pesa de una libra en el platillo de la derecha y se añade azucar al platillo de la izquierda hasta que los dos platillos estén en equilibrio. Las dos cantidades de azucar así medidas sumarán dos libras. ¿Tiene razón el tendero? ¿Puede usted pensar en alguna otra forma de medir dos libras de azucar con esta balanza defectuosa?

5. El modelo de la balanza puede usarse también para representar sistemas de ecuaciones lineales. ¿Puede usted dar un ejemplo interesante?

Este mismo modelo permite tratar desigualdades lineales. Considerar y resolver los siguientes dos ejemplos.

Ejercicio
Cuatro niños Alberto, Beatriz, Carola y David juegan en un balancín. Una

Puntos de discusión

1. *¿Cuál sería el procedimiento con la balanza que corresponde a restar 4 de ambos miembros de la ecuación?*

2. *¿Cuáles serían la disposición original y el procedimiento con la balanza que corresponden a "seguir contando"?*

3. *¿La balanza es útil para proceder por ensayo y error? ¿Cómo?*

profesora los observa jugando en tres momentos diferentes y esto es lo que ve (Figura 1.3).

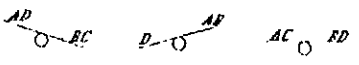
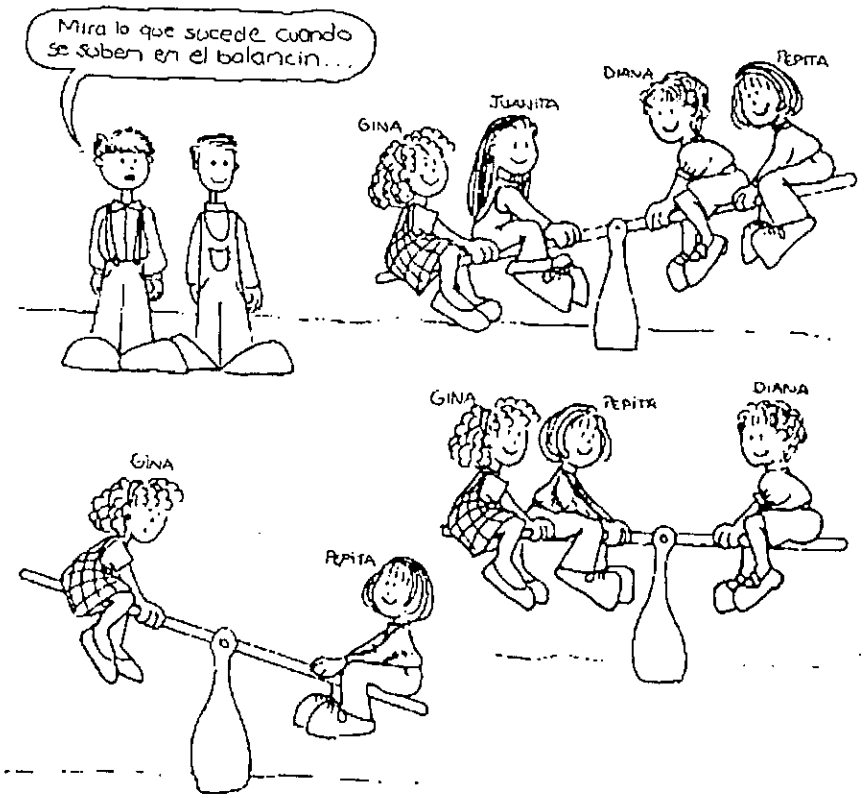


Fig 1.3

¿Quién es el mas pesado de los niños?

Ejercicio
Un día que Pepe y José observaban jugar a sus amigas se dieron cuenta que las cuatro pesaban diferente.



Con los diagramas anteriores, puedes decir cuál de las cuatro niñas es más pesada.

El ensayo y error

No cabe duda que uno de los medios mas potentes que tiene a su disposición un estudiante en la solución de ecuaciones lineales es el de ensayo y error; éste se constituye en un ejercicio fuerte del pensamiento matemático. Consideremos nuevamente la ecuación en cuestión.

$$4 + \square = 7.$$

Si yo adivino 2, coloco el 2 en el cuadrado y sumo; obtengo $4+2=6$. Ahora no sólo descarto el valor de 2 sino que, además, pienso: 2 es demasiado pequeño, ya que la suma que obtuve, 6, es menor de la que busco, 7. Así que cuando vuelvo a ensayar, pondré un número mayor en el cuadrado. Si ensayo el 3, ya está.

Pero, si en lugar de seguir ensayando, pienso un poco mas, puedo decir que la suma que obtuve, 6, es 1 menos que la suma que busco, 7, luego el número que coloqué en el cuadrado es 1 menos que el que busco. Coloqué el 2; luego, busco 1 más, o sea el 3.

El ejemplo es supremamente sencillo, pero la historia consagra la importancia del ensayo y error, en términos muy similares a éstos, en un método conocido desde los tiempos de los antiguos egipcios y llamado el *método de posición falsa*. Tendremos oportunidad de mirarlo un poco mas adelante.

Nuevos modelos y nuevos sistemas numéricos

Por otra parte, sabemos que un gran paso se toma cuando en algún momento se pide hallar el número escondido en una adición como ésta

$$7 + \square = 5, \tag{1.2}$$

que no se presta para modelar en la balanza. No sólo hay que buscar un nuevo modelo, sino que proporciona la oportunidad de introducir un nuevo sistema numérico. De hecho, es común motivar la introducción de nuevos sistemas numéricos a partir de la necesidad de resolver ciertas ecuaciones. Así las cosas, la ecuación anterior puede servir para motivar la introducción del modelo de desplazamientos sobre una recta (numérica) y, en consecuencia, de los números negativos. Por ejemplo, el modelo representa el número positivo 8 en términos de un desplazamiento de longitud 8 en un sentido sobre una recta numérica y el número negativo -2 por un desplazamiento de longitud 2 en el sentido contrario sobre la recta. La suma de dos números enteros está representada por la resultante de los dos desplazamientos.

Puntos de discusión

1. ¿Cómo se modelaría la ecuación (1.2) usando desplazamientos sobre la recta numérica?

2. ¿Se podría modificar el modelo de la balanza de modo muy particular para representar esta ecuación?

Otras ecuaciones lineales, como $(5 \times \square) + 3 = 12$ exigen la introducción de los números racionales y así sucesivamente.

Mas adelante en este libro estudiaremos algunas estructuras algebraicas y en particular pondremos mucho énfasis en el estudio de grupos y anillos. Por el momento, queremos comentar que es posible enfocar la solución de ecuaciones lineales no sólo en términos del sistema numérico particular en el cual se está trabajando, sino en términos de ciertas estructuras algebraicas. Por ejemplo, un grupo es una estructura algebraica en la cual se pueden resolver ecuaciones lineales que involucran una sola operación ($a + x = b$ o $c \cdot x = d$), mientras que un campo es una estructura algebraica en la cual se pueden resolver ecuaciones lineales generales del tipo $ax + b = c$.

Ecuaciones lineales con dos variables

Por otra parte, sucede algo similar con ecuaciones lineales en dos variables. Veamos un primer ejemplo.

-En la siguiente adición el cuadrado y el triángulo están tapando dos números (enteros positivos). ¿Cuáles son los números que están tapando?

$$\square + \triangle = 9.$$

La idea en este problema es que no hay solución única y que, al responder, el alumno debe aportar todas las parejas de números que "hacen 9": $1+8, 2+7, 3+6, ..., 7+2, 8+1$. Una sencilla variación puede hacer que el anterior problema resulte muy interesante. Basta buscar todas las soluciones, digamos en números enteros, de la ecuación

$$(5 \times \square) + (3 \times \triangle) = 9,$$

para comenzar a ver todo el proceso de pensamiento que se requiere en general para tratar ecuaciones de este tipo. Aquí delimitar el sistema numérico en el cual se está trabajando y las operaciones que en el se pueden efectuar incide también de manera fundamental en los modelos y los procedimientos de solución que sean apropiados.

La construcción paulatina de significado

Resulta desde luego mas interesante y formativo en niveles superiores estudiar la solución de ecuaciones desde el punto de vista de su desarrollo histórico, analizando cada método usado en términos de las concepciones numéricas, tanto de los sistemas numéricos mismos como de las operaciones entre números, que se habían construido en cada cultura matemática. Este enfoque tiene en cuenta que no se construye de entrada una concepción matemática completa, que los conceptos construidos son apenas parcialmente adecuados y son susceptibles a perfeccionarse, sometidos a un largo proceso de maduración precisamente en el proceso histórico. Además estas formas de aproximarse a la construcción del conocimiento matemático se dan en los alumnos que encontramos en nuestras clases. Cada uno ha logrado una construcción parcial de los conceptos numéricos y de las operaciones, y según sus concepciones, estará en condiciones de proceder frente a los retos que nosotros le planteamos. Esto implica que no hay una única forma adecuada de resolver una ecuación dada. En este primer capítulo, entonces, queremos explorar varias alternativas y aproximaciones, tanto históricas como pedagógicas, para revelar las riquezas del pensamiento matemático que yacen detrás de un tema tan aparentemente sencillo y cerrado como es la solución de ecuaciones lineales.

1.2 Métodos de solución de ecuaciones lineales en la historia

Los métodos utilizados para resolver ecuaciones en diferentes etapas de la historia, así como el contexto en el cual se plantean, revelan en todas las etapas tanto conocimiento como ignorancia u otro tipo de limitación. Quizás una de las lecciones más importantes que se puede derivar de un estudio histórico de la solución y la teoría de ecuaciones es la *genialidad de la ignorancia*. Para el profesor de matemáticas esto puede servir de guía

para comprender que la enseñanza debe valorar las muestras de ingenio del alumno frente a un problema nuevo para cuya solución no conoce previamente ningún método y es por ello que constantemente debe proponer problemas que no se encierren en el esquema de métodos conocidos, pues es precisamente de esta forma que el estudiante podrá extender los límites de su conocimiento y desarrollar su pensamiento matemático.

1.2.1 Los métodos egipcios de solución de una ecuación lineal

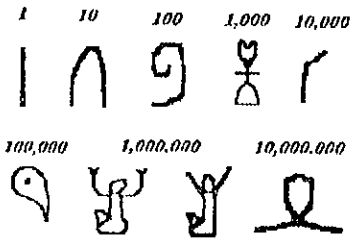


Figura 1.4

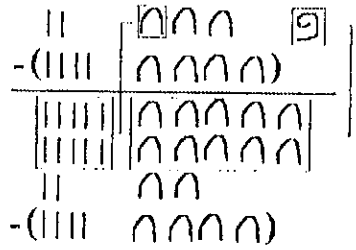


Figura 1.5

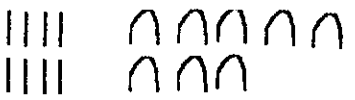


Figura 1.6

Al analizar los métodos usados por los egipcios para resolver ecuaciones, salta a la vista la dificultad de hacer "álgebra" cuando el sistema numérico en el cual se trabaja está severamente limitado. Si lo comparamos con un sistema conocido diríamos que el sistema numérico egipcio tiene semejanzas con el romano: es un sistema decimal con un símbolo especial para la unidad y para cada potencia de 10 (Figura 1.4).

No se aprecia en él ninguna noción de valor posicional. Así las cosas, la adición y sustracción egipcias se reducen a reunión de símbolos y, si fuera el caso, 'trueque' de un símbolo de valor mayor por 10 del valor inmediatamente anterior, una decena por 10 unidades, una centena por 10 decenas, y así sucesivamente. La sustracción es precisamente esto; la supresión de símbolos con el 'trueque' en caso de que fuera necesario.

Cambiando diez | por un ∩ y diez ∩ por un ϑ, tenemos (Figura 1.5).

Luego se puede restar, suprimiendo símbolos correspondientes al sustraendo, para obtener la diferencia (Figura 1.6).

La multiplicación egipcia se efectúa por medio de duplicaciones sucesivas y la división por medio de la multiplicación del divisor hasta obtener el dividendo. Por ejemplo, para hacer el producto 19×71 se lleva a cabo sucesivas duplicaciones de 71 hasta la última potencia de 2 menor que 19. Luego se escogen las potencias de 2 cuya suma sea 19 (se expresa 19 como suma de potencias de 2) y se adicionan los respectivos productos de 71. Como $19 = 16 + 2 + 1$,

$$19 \times 71 = (16+2+1) \times 71 = 16 \times 71 + 2 \times 71 + 1 \times 71 = 1136 + 142 + 71 = 1349.$$

√	1	71
√	2	142
	4	284
	8	568
√	16	1136
Totales	19	1349

Para efectuar la división, se lleva a cabo un proceso similar. Por ejemplo, para hacer $35 \div 8$ se efectúan sucesivas duplicaciones del divisor parando para no exceder el dividendo. Si es necesario, como lo es en este caso, se hacen sucesivas divisiones por 2. Se suman aquellos múltiplos y submúltiplos que suman 35. La suma de las respectivas potencias de 2 corresponde al cociente.

	1	8	
	2	16	
	4	32	✓
	$\frac{1}{2}$	4	
	$\frac{1}{4}$	2	✓
	$\frac{1}{8}$	1	✓
Totales	$4+\frac{1}{4}+\frac{1}{8}$	35	

Ejercicios

1. Efectuar 15×39 y 20×47 usando el algoritmo egipcio de la multiplicación.
2. Efectuar $117 \div 3$ y $1001 \div 77$ usando el algoritmo egipcio de la división.
3. Dada una división, ¿puede usted predecir si para efectuarla será necesario recurrir o no a submúltiplos del divisor?

Punto de discusión

El algoritmo de la multiplicación usado por los egipcios debe su validez al hecho de que cualquier número puede expresarse (de manera única) como suma de potencias de 2. ¿Puede usted explicar esta afirmación?

Los últimos dos algoritmos dejan en claro que los egipcios entendían la división como la operación inversa de la multiplicación.

Sin embargo, como ilustra con claridad el último ejemplo, dividir para los egipcios no es multiplicar por el recíproco. De hecho, los egipcios no tenían una conceptualización ágil y adecuada de los fraccionarios. Por algún motivo, para ellos las fracciones se debían limitar a aquellas de numerador 1, con la excepción de la fracción $\frac{2}{3}$. Por ejemplo, en lugar de escribir la fracción $\frac{3}{17}$ un egipcio debió descomponerla en una suma de fracciones con numerador 1 y denominadores distintos. El primer paso, escribir $\frac{3}{17} = \frac{1}{17} + \frac{2}{17}$ es obvio. Luego, se debe descomponer $\frac{2}{17}$ de la manera exigida. En uno de los papiros egipcios existentes, el papiro de Rhind, se incluye una gran tabla de descomposiciones de las fracciones con numerador 2 y denominador impar desde $\frac{2}{5}$ hasta $\frac{2}{101}$. La descomposición dada para $\frac{2}{17}$ es $\frac{1}{12} + \frac{1}{51} + \frac{1}{68}$.

Punto de discusión

Explicar exactamente en qué sentido la división es para los egipcios la operación inversa de la multiplicación.

Punto de discusión

¿Cuál es el recíproco del fraccionario p/q? Explicar por qué el tratamiento egipcio de los fraccionarios impide en general concebir la división como multiplicación por el recíproco.

Ahora bien, en un sistema numérico como éste resulta bien difícil imaginar métodos adecuados para resolver ecuaciones y, de hecho, las ecuaciones más sencillas causan algunas dificultades grandes. Por ejemplo, del Papiro Rhind (c. 2300 a.C.) tenemos este problema: "Una cantidad y su $\frac{1}{7}$ sumadas se hacen 19. ¿Cuál es la cantidad?" Obviamente en nuestra notación esto es

$$x + \frac{x}{7} = 19 \quad o \quad \frac{8x}{7} = 19.$$

Pero recuerden que los egipcios no permitían la fracción $\frac{8}{7}$. Describiendo su método de solución dice el papiro

> "Cuantas veces se debe multiplicar a 8 para obtener 19, tantas veces se debe multiplicar a 7 para obtener el número correcto."

Esto describe el método egipcio de posición falsa que depende del razonamiento proporcional y que se puede explicar como sigue.

En el ejemplo $x + \frac{x}{7} = 19$, si se supone que $x = 7$, se calcula que $x + \frac{x}{7}$ es igual a 8. Se debe multiplicar 8 por $\frac{19}{8} = 2 + \frac{1}{4} + \frac{1}{8}$ para obtener 19; el valor correcto de x se obtiene, por lo tanto, multiplicando 7 por $2 + \frac{1}{4} + \frac{1}{8}$ que da por resultado

$$x = \left(2 + \frac{1}{4} + \frac{1}{8}\right) 7 = 16 + \frac{1}{2} + \frac{1}{8}.$$

En palabras se puede describir este procedimiento como sigue. Se da un valor, apropiadamente escogido, para la cantidad desconocida y se sustituye éste en la ecuación obteniendo un cierto resultado. La solución a la ecuación tiene la misma relación con el valor escogido que el número dado (constante en la ecuación) tiene con el resultado que se calcula.

Punto de discusión
Comparar y contrastar el método de posición falsa de los egipcios con el procedimiento por ensayo y error tal como fue presentado en la sección anterior.

Ahora bien, si analizamos nuestro método contemporáneo de solución a la ecuación anterior, escribiríamos

$$x + \frac{x}{7} = 19.$$

Luego sumaríamos $x + \frac{x}{7}$. (Como hemos comentado, este resultado $\frac{8x}{7}$ violaría el tratamiento egipcio de las fracciones.) Luego multiplicaríamos $19 \times \frac{7}{8}$ para obtener nuestra respuesta, es decir, dividiríamos multiplicando por el recíproco de $\frac{8}{7}$. Pero, aunque este último paso no involucrara una fracción con numerador diferente de 1, la división como fue practicada por los egipcios no permitiría este planteamiento, pues aunque conciben la división como la inversa de la multiplicación, no la conciben como multiplicación por el recíproco.

Puntos de discusión

1. Uno de los modelos mas importantes que usan los niños para la división es la repartición. La repartición se utiliza mas que todo en números naturales donde, para tratar el caso general, hay que hablar de un cociente y un residuo. Dar un ejemplo. ¿Es la división así modelada la inversa de la multiplicación?

2. El hecho de no tener un modelo totalmente adecuado para la división impide que un alumno pueda resolver ecuaciones lineales generales?

Volviendo al método de los egipcios, queremos dejar en claro que, para el profesor, éste no es solamente un cuento curioso del pasado. El hecho es que en el salón de clase todos los días pequeños seres humanos construyen modelos de significado para las operaciones aritméticas que enfrentan y generan métodos para resolver problemas acordes con el entendimiento que han logrado. Todos sabemos que el primer modelo de la multiplicación que se genera es la adición repetida y que hay que superar este modelo cuando, en la misma escuela primaria, se llega a sistemas numéricos distintos de los naturales (en particular, al sistema de los fraccionarios). En ese momento se pierden para siempre algunos niños para la matemática porque no logran construir nuevos modelos capaces de dar significado a las operaciones que están efectuando.

De igual manera, a veces el maestro no comprende que el modelo de la división como inversa de la multiplicación es en sí bastante sofisticado y se encuentra en conflicto con el modelo de repartición (división con residuo) que es tal vez la más básica para el niño.

De todas maneras, los egipcios nos enseñan que, aunque se logre pasar de esta noción intuitiva de la división como repartición a la de la división

como inversa de la multiplicación esto no es garantía de haber construido un significado de la división que es adecuado para estudiar las fracciones o el álgebra. El camino es más largo y tortuoso.

Ejercicios

Punto de discusión
Especificar el conflicto que existe entre el modelo de la división como repartición (con posible residuo) y el modelo de la división como inversa de la multiplicación.

Resolver estos problemas tomados del Papiro de Rhind por el método de posición falsa.

1. (Problemas 3 a 6) Dividir 6, 7, 8 y 9 panes entre 10 hombres.
2. (Problemas 25 a 27) Una cantidad y su $\frac{1}{2}$ juntas son 16. ¿Cuál es la cantidad? Una cantidad y su $\frac{1}{4}$ juntas son 15. ¿Cuál es la cantidad? Una cantidad y su $\frac{1}{5}$ juntas son 21, ¿Cuál es la cantidad?
3. (Problema 40) Dividir 100 panes entre 50 hombres de manera que la suma de las tres partes mayores sea 7 veces la suma de las dos menores.

1.3 Ecuaciones lineales con solución en números enteros

En la escuela secundaria se aprende que una ecuación lineal de la forma $ax + by = c$ representa una línea recta en el plano cartesiano. Hay infinitos puntos (x, y) sobre la recta. Para determinar uno cualquiera de ellos, basta dar un valor a una de las variables y despejar el valor correspondiente de la otra. Pero el procedimiento no es tan directo cuando se trata de restringir valores a los números enteros.

Consideremos por ejemplo la ecuación $2x + 7y = 12$. Dando un valor cualquiera a y, digamos $y = 13$, se puede despejar y obtener el correspondiente valor de x. En este caso, tenemos $x = \frac{12-7(13)}{2} = -\frac{79}{2}$. Pero, como es evidente, no se asegura con este procedimiento que el valor de x sea entero. Como hemos querido enfatizar, el sistema numérico en el cual se está trabajando (o la concepción de tal sistema que hemos podido desarrollar) influye fuertemente en los métodos y procedimientos que emplearemos. Resulta que resolver ecuaciones lineales en enteros implica desarrollar una serie de temas de la teoría de números.

Ahora bien, una de las relaciones que mas común y marcadamente se resalta entre el álgebra (moderna o abstracta) y otras ramas de la matemática es la que existe entre álgebra y teoría de números. Sin embargo, no sólo es posible, sino urgente, replantear el estudio de los elementos y argumentos compartidos por estas dos áreas con el ánimo de enriquecer la experiencia matemática del profesor y hacer mas vívida su comprensión de los aspectos formales y abstractos que inciden en o colindan con temas elementales y que permiten un acercamiento, desde la matemática elemental, a la comprensión de la naturaleza de la matemática contemporánea.

Estudiaremos en esta sección algunos tópicos de la teoría de números relacionados con la solución de ecuaciones y que contribuyeron a la formación del modo de pensar del álgebra abstracta. En particular, seguiremos la trayectoria histórica de algunos problemas de representación de enteros,

deteniéndonos primero en el estudio de las ecuaciones diofánticas lineales. En el siguiente capítulo trataremos la solución de ecuaciones diofánticas cuadráticas; en ambos casos nuestros métodos preferidos de solución usan fracciones continuas.

Además, posteriormente consideraremos el método de descenso infinito de Fermat y su uso para la solución de problemas relacionados con los anteriores. Nuestro interés en estos temas se arraiga en dos aspectos importantes. Primero, queremos mostrar como este método, original de Fermat, resalta una de las propiedades básicas del conjunto de los números naturales que es fundamental en toda la argumentación posterior del álgebra abstracta. En segundo lugar, queremos subrayar la naturaleza abstracta de los problemas de representación propios de la teoría de números, característica básica del álgebra moderna, es decir, mostrar que se trata de enunciados sobre la posibilidad de lograr una cierta representación sin que se basen las demostraciones en la construcción de la misma.

Para abordar esta importante discusión, es necesario primero construir (o recordar) algunos de los elementos y argumentos básicos de la teoría de números.

1.3.1 Divisibilidad

En la teoría de números, el instrumento de análisis que tradicionalmente se usa para explorar propiedades e interrelaciones numéricas es el estudio de la divisibilidad.

Si consideramos una de las clasificaciones básicas de los números, en pares e impares, vemos que se está haciendo referencia al hecho de que un número sea o no divisible por 2. La clasificación de los números en primos o compuestos es también basada en propiedades de divisibilidad; los números primos son aquellos que tienen exactamente dos divisores, 1 y el número mismo, los compuestos los que tienen más de dos divisores.

Desde los primeros días de la matemática griega, estas definiciones hicieron su aparición, apreciadas como verdades profundas acerca de los números. No han perdido su vigencia como conceptos claves alrededor de los cuales se analizan las muy variadas situaciones en que tienen inherencia los números enteros. En el álgebra en particular podemos mencionar el orden de un grupo, el grado de una ecuación, la dimensión de un espacio, etc.

Definiciones básicas

Todos nosotros estamos familiarizados con las nociones básicas de divisibilidad; sin embargo, a manera de repaso presentamos la siguiente lista de definiciones.

Definición 1.1 *Sean a, b números enteros. Decimos que a* divide *a b si existe un número entero m tal que $am = b$. Se dice también que a es* factor *o* divisor *de b, que b es* divisible *por a o que b es* múltiplo *de a. Escribimos $a|b$.*

Definición 1.2 *Un número es* par *si 2 lo divide. Se sigue que los números pares pueden escribirse en la forma $2k$ donde k es un número entero.*

Un número se dice impar, si 2 no lo divide.

Ya que los únicos residuos posibles al dividir un número por 2 son 0 o 1, se sigue que los números impares pueden escribirse en la forma $2k+1$ donde k es un número entero.

Definición 1.3 *Un número entero positivo se dice primo si tiene exactamente dos divisores. 2, 19 y 37 son números primos.*

Definición 1.4 *Un número entero positivo se dice* compuesto *si tiene más de dos divisores.*

6 es compuesto ya que el conjunto de sus divisores es $\{1,2,3,6\}$. Otros ejemplos de números compuestos son 91 cuyo conjunto de divisores es $\{1,7,13,91\}$ y 1001 con divisores $\{1,7,11,13,77,91,143,1001\}$.

Nota. *El número 1 no es ni primo ni compuesto.*

Definición 1.5 *Sean a,b dos números enteros. Se dice que un número entero c es divisor común de a y b si $c|a$ y $c|b$.*

Definición 1.6 *Un número entero d se dice máximo común divisor de a y b si*

(i) d es divisor común de a y b. $[d|a \wedge d|b]$.

(ii) Si c es cualquier divisor común de a y b, entonces c divide a d. $[c|a \wedge c|b \rightarrow c|d$.$]$

Escribimos $d = mcd(a,b)$. Donde no hay lugar a confusión escribiremos $d = (a,b)$.

Definición 1.7 *Sean a,b números enteros. Decimos que a y b son primos relativos si su máximo común divisor es 1. Escribimos $(a,b) = 1$.*

1.3.2 Propiedades elementales de la divisibilidad.

Para adelantar nuestra discusión acerca de las relaciones entre la teoría de números y el álgebra, requerimos de algunos resultados básicos sobre la divisibilidad que no demostraremos ya que sus demostraciones son inmediatas. En los enunciados que siguen a, b, c son números enteros.

Teorema 1.1 *Si $a|b$ y $a|c$, entonces $a|(b \pm c)$.*

Teorema 1.2 *Si $a|b$, entonces $a|bc$.*

Ejercicio

Demostrar el siguiente teorema.

Teorema 1.3 *Si $a|b$ y $a|c$, entonces $a|(bx+cy)$, para x,y enteros. La expresión $ax+by$ se llama una combinación lineal de a y b.*

1.3.3 Algoritmo de Euclides.

Es evidente que, dados dos números enteros, no necesariamente se da el caso que uno sea divisor del otro. El algoritmo de la división y el algoritmo de Euclides hacen referencia a casos mas generales.

Teorema 1.4 *(Algoritmo de la división). Sean a y b dos números enteros positivos. Entonces existen q, r enteros no negativos, $0 \leq r < b$, tales que*

$$a = qb + r. \tag{1.3}$$

En la expresión anterior q se llama el *cociente* y r el *residuo* al dividir a por b. (Es claro que, si en (1.3) $r = 0$ entonces b es un divisor de a.)

El Algoritmo de Euclides, que enunciamos a continuación, utiliza el algoritmo de la división para producir unos resultados mas interesantes desde el punto de vista de la teoría de números.

Teorema 1.5 *(Algoritmo de Euclides). Sean a, b números enteros positivos. Entonces, podemos aplicar el algoritmo de la división repetidamente como sigue.*

Existen $q_1, q_2, \ldots, q_{n+1}, r_1, r_2, \ldots, r_n$ tales que

$$\begin{array}{lcll} a & = & bq_1 + r_1, & 0 \leq r_1 < b; \\ b & = & q_1 r_1 + r_2, & 0 \leq r_2 < r_1; \\ r_1 & = & q_3 r_2 + r_3, & 0 \leq r_3 < r_2; \\ \ldots & & \ldots & \ldots \\ r_{n-2} & = & q_n r_{n-1} + r_n, & 0 \leq r_n < r_{n-1}; \\ r_{n-1} & = & q_{n+1} r_n & \end{array} \tag{1.4}$$

Nótese que el residuo en el último renglón es 0. Para demostrar que siempre se llega a esta situación, notamos que la sucesión de residuos $r_1, r_2, \ldots r_n$ es una sucesión decreciente de números enteros no negativos, ninguno mayor que r_1. Esto significa que en algún momento habrá un residuo de 0, tal como indica precisamente el último renglón del algoritmo.

Ejemplos. Si $a = 289$ y $b = 102$, se tiene

$$\begin{aligned} 289 &= 102 \times 2 + 85 \\ 102 &= 1 \times 85 + 17 \\ 85 &= 17 \times 5. \end{aligned}$$

Si $a = 55$ y $b = 34$ se tiene

$$\begin{aligned} 55 &= 34 \times 1 + 21 \\ 34 &= 21 \times 1 + 13 \\ 21 &= 13 \times 1 + 8 \\ 13 &= 8 \times 1 + 5 \\ 8 &= 5 \times 1 + 3 \end{aligned}$$

$$\begin{aligned} 5 &= 3 \times 1 + 2 \\ 3 &= 2 \times 1 + 1 \\ 2 &= 2 \times 1 \end{aligned}$$

Ahora bien, queremos demostrar que, cuando se aplica el Algoritmo de Euclides a dos números a y b, r_n, el último residuo diferente de 0, es el máximo común divisor de a y b ($mcd(a,b)$). Veamos.

$$\begin{aligned} a &= bq_1 + r_1, && (1) \\ b &= q_1 r_1 + r_2, && (2) \\ r_1 &= q_3 r_2 + r_3, && (3) \\ &\dots \quad \dots \\ r_{n-2} &= q_n r_{n-1} + r_n, && (n) \\ r_{n-1} &= q_{n+1} r_n && (n+1) \end{aligned}$$

Primero demostraremos que r_n es común divisor de a y b. De $(n+1)$ se sigue que $r_n | r_{n-1}$. Pero entonces en (n), se tiene que $r_n | r_n$ y $r_n | r_{n-1}$, de donde $r_n | r_{n-2}$. Ahora bien, el miembro derecho de $(n-1)$ contiene una combinación lineal de r_{n-1} y r_{n-2}, mientras que el miembro derecho es r_{n-3}. Se sigue del Teorema 1.3.3 que $r_n | r_{n-3}$.

Continuando este proceso, podemos demostrar que r_n es divisor de los demás residuos, en particular $r_n | r_2$ y $r_n | r_1$, entonces de (2) se tiene que $r_n | b$. Y aplicando el mismo razonamiento a (1) se sigue que $r_n | a$. Es decir, r_n es divisor común de a y b.

Ahora bien, para demostrar que r_n es el máximo común divisor de a y b, debemos considerar cualquier divisor común c de a y b. Ya que $c | a$ y $c | b$, de (1) se sigue que $c | r_1$. De $c | b$ y $c | r_1$ se sigue de (2) que $c | r_2$. Procediendo de esta manera, podemos demostrar que c es factor de todos los residuos y, en particular, que es divisor de r_n.

Según la definición de máximo común divisor, hemos demostrado que $r_n = mcd(a,b)$.

1.3.4 Representación del máximo común divisor de dos números.

Podemos usar el Algoritmo de Euclides para demostrar el Teorema de la representación del máximo común divisor de dos números.

Teorema 1.6 *Sean $a, b \in \mathbb{N}$, $d = mcd(a,b)$. Existen $s, t \in \mathbb{Z}$ tales que*

$$as + bt = d.$$

Diremos que d puede *representarse como combinación (forma) lineal* en a y b.

Empezaremos por mirar un ejemplo; nuestra demostración seguirá los mismos pasos. Consideremos los números 289 y 102; aplicando el algoritmo de Euclides, concluimos que $17 = mcd(289, 102)$. Tenemos

$$\begin{aligned} 289 &= 102 \times 2 + 85 \qquad (1)\\ 102 &= 85 \times 1 + 17 \qquad (2)\\ 85 &= 17 \times 5 \end{aligned}$$

Ahora bien, para hallar s y t tales que $17 = 289s + 102t$, de (2) se sigue que

$$17 = 102 - 85 \times 1$$

y de (1) que

$$85 = 289 - 102 \times 2. \qquad (3)$$

Luego, sustituyendo en (3), obtenemos

$$17 = 102 - 85 = 102 - (289 - 102 \times 2) = 102 \times 3 - 289.$$

En los términos de nuestro teorema, aquí se tiene $17 = d$, $289 = a$, $102 = b$, $s = -1$, $t = 3$.

Como un segundo ejemplo, consideremos los números 173 y 29. Aplicando el algoritmo de Euclides, tenemos

$$\begin{aligned} 173 &= 29 \times 5 + 28 \qquad (4)\\ 29 &= 28 \times 1 + 1 \qquad (5)\\ 28 &= 1 \times 28. \end{aligned}$$

Puntos de discusión

Lo anterior puede modelarse sobre la recta numérica en términos de saltos iguales.

1. Por ejemplo, ¿cómo pueden representarse los múltiplos de 5 en términos de saltos iguales a partir de 0?

2. ¿Cómo puede representar los múltiplos comunes de 5 y 7 con el mismo modelo? ¿El mínimo común múltiplo de 5 y 7?

3. Si se comienza en el punto correspondiente a 0 y se permite hacer saltos de a 8 hacia adelante y saltos de a 11 hacia atrás, ¿cuál es el punto más próximo a 0 al cual se puede llegar? Esto corresponde a un modelo del máximo común divisor de 8 y 11. ¿Por qué?

4. ¿Puede usted replantear los principales temas y resultados sobre divisibilidad que hemos desarrollado hasta el momento en términos de este modelo?

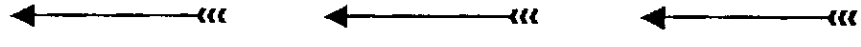
Se sigue que $1 = mcd(173, 29)$. Ahora, queremos hallar s, t tales que $1 = 173s + 29t$. De (4) y (5), tenemos

$$\begin{aligned} 28 &= 173 - 29 \times 5\\ 1 &= 29 - 28 \times 1. \end{aligned}$$

Entonces,

$$1 = 29 - 28 = 29 - (173 - 29 \times 5) = 173(-1) + 29(6). \qquad (1.5)$$

Demostración. Nuestra demostración del Teorema de la representación procede refiriéndonos al Algoritmo de Euclides para despejar r_{n-1} en $(n-1)$, luego r_{n-2} en $(n-2)$. De esta manera se logra expresar r_{n-1} como combinación lineal de r_{n-2} y r_{n-3}. Luego, se depeja r_{n-3} en $(n-3)$ y así sucesivamente hasta expresar r_{n-1} en la forma $as + bt$.



Regresamos ahora a considerar dos propiedades fundamentales de los números.

Teorema 1.7 *Si $a|bc$ y $(a, b) = 1$, entonces $a|c$.*

Demostración. Como $(a, b) = 1$, por el Teorema de la representación existen enteros s, t tales que $sa + tb = 1$. Multiplicando la anterior ecuación por c produce

$$sac + tbc = c.$$

Ahora bien, $a|bc$, de donde, $a|tbc$. Además, $a|sac$. Se sigue que $a|(sac+tbc)$, es decir, $a|c$.

Es corolario inmediato el llamado Lema de Euclides.

Corolario 1.1 *(Lema de Euclides). Sea p primo. Si $p|ab$ y $p \nmid a$, entonces $p|b$.*

Ahora, estamos en condiciones de demostrar el siguiente teorema.

Teorema 1.8 *(Teorema fundamental de la aritmética.) Dado un número entero positivo N. N puede expresarse de manera única (salvo orden) como producto de potencias de números primos. $N = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_n^{\alpha_n}$.*

Los números $p_1, p_2, \ldots p_n$ se llaman los *factores primos* y el producto $p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_n^{\alpha_n}$ la *factorización prima* de N.

Punto de discusión

Usando el lema de Euclides, esbozar una demostración del teorema fundamental de la aritmética.

Esta propiedad de los números enteros es supremamente potente; permite obtener solución a problemas que parecen ser indeterminados. Antes de proseguir nuestro estudio, consideremos un ejemplo sencillo de la potencia de la unicidad de la factorización prima de un número.

Problema

El producto de las edades de dos personas adultas es 770. ¿Cuál es la suma de sus edades?

Solución. Para resolver el problema, ya que en el se habla de un producto, expresemos 770 en su factorización prima.

$$770 = 11 \times 7 \times 5 \times 2$$

Ahora bien, ninguno de estos factores por sí solo representa la edad de una persona adulta. Así las cosas, las dos edades que buscamos deben expresarse como productos de estos factores. Además, ni 2×5 ni 2×7 corresponden a la edad de un adulto. Se sigue que una de las edades que buscamos es $2 \times 11 = 22$ y la otra $5 \times 7 = 35$ cuya suma es $22 + 35 = 57$.

1.3.5 Contraste de enfoques

El enfoque que hemos dado al Teorema de representación lineal no sólo nos permite afirmar la posibilidad de expresar el máximo común divisor d de dos números enteros a y b en la forma $d = as + bt$, donde s y t son enteros, sino que nos permite hallar s y t (resolver la ecuación $ax + by = d$ en enteros). Contrastemos ahora con un tratamiento contemporáneo del mismo teorema. Con esta comparación queremos comenzar a entender cómo el álgebra vista como una teoría de ecuaciones puede transformarse en una álgebra abstracta que comporta el carácter puramente formal de la matemática moderna.

Teorema 1.9 *Dados dos números enteros a, b, no ambos iguales a 0, entonces (a, b) el máximo común divisor de a y b existe; además podemos encontrar enteros s y t tales que $(a, b) = sa + tb$.*

Demostracion. Consideremos el conjunto S de todos los enteros de la forma $ma + nb$ donde m y n recorren los enteros. Hay enteros no nulos en S ya que a y b no son ambos iguales a 0. Si $w = ma + nb$ está en S, entonces $-w = (-m)a + (-n)s$ también pertenece a S, lo cual nos permite concluir que hay enteros positivos en S. Por el principio de buena ordenación, hay un menor entero positivo d que pertenece a S. Es claro que $d = sa + tb$ para algunos enteros s, t. Vamos a demostrar que d es el máximo común divisor de a y b.

Primero, observamos que si $c|a$ y $c|b$, entonces $c|sa + tb = d$, es decir, cualquier divisor común de a y b es un divisor de d. Nos falta mostrar que $d|a$ y $d|b$. Ahora, sea $x = ma + nb$ cualquier elemento de S. Por el algoritmo de Euclides, $x = qd + r$ donde $0 \leq r < d$. Reescribiendo esta última ecuación, obtenemos $ma + nb = q(sa + tb) + r$, de donde, $r = (m - qs)a + (n - qt)b$, o sea, $r \in S$. Ya que $0 \leq r$ y $r < d$, por la forma en que elegimos d, debe tenerse que $r = 0$. De allí se tiene que $x = qd$, es decir, que d es divisor de cualquier elemento de S. Ahora bien, $a \in S$ ya que $a = 1 \cdot a + 0 \cdot b$ y $b \in S$ ya que $b = 0 \cdot a + 1 \cdot b$, se sigue que $d|a$ y $d|b$ como queríamos.

De acuerdo con la definición de máximo común divisor, hemos demostrado que $mcd(a, b)$ existe y que puede expresarse como $(a, b) = sa + tb$ donde s, t son enteros.

Ahora que hemos visto dos enfoques y dos demostraciones del teorema de representación del máximo común divisor, contrastemos los dos. Nuestra primera demostración utiliza el algoritmo de Euclides y muestra la existencia del máximo común divisor mostrando una manera de encontrarlo (último residuo diferente de 0 cuando se aplica el algoritmo a los números a y b). Además muestra que el máximo común divisor puede representarse como combinación lineal de a y b en efecto mostrando cómo lograr dicha representación (despejando d sucesivamente en las ecuaciones que se obtienen al aplicar el algoritmo de Euclides).

En cambio, el enfoque contemporáneo no nos dice cómo encontrar el máximo común divisor ni cómo expresarlo como combinación de a y b. Este tipo de demostración se llama una demostración pura de existencia, mientras que el enfoque tradicional se llama una demostración constructiva.

De manera similar, en el terreno de la solución de ecuaciones, el álgebra tradicional se ocupa de encontrar las soluciones a una ecuación o un sistema de ecuaciones, mientras que el álgebra abstracta se ocupa, entre otras cosas, de enunciar condiciones, bajo las cuales, soluciones a una cierta ecuación existen sin determinar cuales son.

Como tendremos oportunidad de apreciar mas adelante, en su estudio de la representación lineal y cuadrática, tanto el matemático Alejandrino Diofanto como su discípulo moderno Pierre de Fermat comenzaron a estudiar el problema de existencia sin ocuparse de manera fundamental con el problema de construcción, de manera que modos de pensar naturales a la teoría de números serían luego adoptados por el álgebra.

1.4 Ecuaciones diofánticas lineales

Podemos usar el procedimiento conocido como el algoritmo de Euclides para resolver ecuaciones diofánticas lineales, es decir, solucionar en números enteros ecuaciones lineales en dos variables con coeficientes enteros. Este problema tiene su origen en los trabajos algebraicos de Diofanto de Alejandría quizás el último matemático de gran importancia en la escuela griega de Alejandría, escuela que había producido a Euclides, Herón, Ptolomeo y muchos otros sobresalientes matemáticos durante mas de seis siglos. En nuestra discusión de la historia de la solución de ecuaciones polinómicas mas adelante estudiaremos el álgebra geométrica de los Pitagóricos recopilada magistralmente por Euclides en el Libro 2 de los *Elementos*. Allí veremos que la imposibilidad de expresar numéricamente magnitudes irracionales había llevado a representarlas por segmentos de recta e identificarlas con las longitudes de esos segmentos. Naturalmente, en el estudio de la solución de ecuaciones, quedaba un espacio para desarrollar un tratamiento estrictamente algebraico-numérico siempre que se circunscribiera el conjunto numérico a los números racionales. Tal fue el planteamiento de Diofanto. A lo largo de los siglos de estudio de la obra diofántica, se efectuó una delimitación mas estricta. Hoy día cuando se habla de la solución de ecuaciones diofánticas se hace referencia a la solución de ecuaciones (polinómicas con coeficientes enteros) en los números enteros.

Tal restricción al dominio de los enteros constituye un ejercicio exigente para quienes estamos acostumbrados a la solución de ecuaciones en los números reales. Aun en el caso de ecuaciones lineales se encuentran limitaciones e impedimentos desconocidos. Comencemos por considerar tres de tales ecuaciones.

$$2x + 5y = 121 \quad (1) \qquad 4x + 16y = 79 \quad (2) \qquad 6x + 21y = 102 \quad (3)$$

Notemos, en primera instancia que la ecuación (2) no tiene solución en los enteros. Nuestro análisis se basa en consideraciones de divisibilidad. Sean cuales fueran los números enteros x y y, tanto $4x$ como $16y$ son pares (de hecho, son divisibles por 4), de donde $4x + 16y$ es par. Ya que 79 no es par, es imposible encontrar enteros x y y tales que $4x + 16y = 79$.

Es claro que, para que tenga solución en los enteros la ecuación $ax + by = c$, donde a, b, c son enteros, es necesario que cualquier divisor común de a y b también sea un divisor de c. Pues si $m|a$ y $m|b$, entonces $m|ax + by$, pero $ax + by = c$, entonces $m|c$. En particular, si $d = mcd(a, b)$, para que la ecuación anterior tenga solución en los enteros, es necesario que $d|c$.

Ahora bien, en el caso de la ecuación (3), vemos que $3|6$ y $3|21$, y en este caso vemos también que $3|102$. Es mas, $3 = mcd(6, 21)$.

Resulta que la condición anterior es también suficiente para garantizar que la ecuación tenga solución en los enteros, pues, por el teorema de representación sabemos que existen enteros s, t tales que $6s + 21t = 3$. De allí, como $102 = 3 \times 34$ se sigue que

$$6(34s) + 21(34t) = 102,$$

y hemos obtenido una solución a la ecuación.

En cuanto a la ecuación (1), $2x + 5y = 121$, notamos que $mcd(2, 5) = 1$. Por nuestra discusión anterior la existencia de soluciones en los enteros está

garantizada, pues en la ecuación $ax + by = c$, si $mcd(a, b) = 1$, es claro que $1|c$ y, por consiguiente, se cumple la condición suficiente y necesaria para que la ecuación tenga solución en los enteros.

Ahora bien, consideremos la ecuación diofántica

$$3x + 7y = 141. \tag{1.6}$$

Notemos de inmediato que la ecuación tiene solución ya que $mcd(3, 7) = 1$. Resolvamos en enteros la ecuación $3s + 7t = 1$. Podríamos usar el algoritmo de Euclides para resolverla, pero en este caso basta adivinar. Si $s = -2$, $t = 1$, tenemos

$$3(-2) + 7(1) = 1. \tag{1.7}$$

Ahora, multipliquemos ambos miembros de (1.7) por 141 para obtener

$$3(-282) + 7(141) = 141. \tag{1.8}$$

Esto es, hemos encontrado una solución a la ecuación (1.6). Ahora, preguntémonos: ¿cuántas soluciones tiene esa ecuación? Recordando nuestra geometría analítica, es claro que la gráfica de la ecuación es una línea recta con pendiente $-\frac{3}{7}$ e intercepto $\frac{141}{7}$. Del hecho de que la pendiente sea racional, podemos deducir que si encontramos una solución a la ecuación en números enteros, entonces tiene infinitas soluciones en los enteros. El diagrama muestra cómo son éstas soluciones (Figura 1.7).

Traduciendo a términos algebraicos la pendiente, diremos que si (a, b) es un punto de la recta, entonces, el punto $(a + 7, b - 3)$ también pertenece a ella. Desde que encontremos una solución (a, b) en enteros, es decir, desde que se sepa que la recta pasa por un punto con coordenadas enteras (punto reticular), entonces se sigue que pasa por infinitos puntos reticulares, o lo que es equivalente, que tiene infinitas soluciones en enteros.

Para hallar estas soluciones, resolvamos el sistema de ecuaciones (1.6) y (1.8)

$$\begin{aligned} 3x + 7y &= 141 \\ 3(-282) + 7(141) &= 141 \end{aligned}$$

Restando, (1.6) − (1.8), obtenemos

$$3(x + 282) + 7(y - 141) = 0.$$

Equivalentemente, $3(x + 282) = 7(141 - y)$ o

$$\frac{3}{7} = \frac{141 - y}{x + 282}.$$

Ya que, para todo $t \in \mathbb{Z}$, $t \neq 0$, $\frac{3}{7} = \frac{3t}{7t}$, obtenemos

$$\begin{aligned} 141 - y &= 3t & x + 282 &= 7t \\ y &= 141 - 3t & x &= 7t - 282. \end{aligned}$$

Dando valores enteros al parámetro t, se obtienen todas las (infinitas) soluciones en enteros a la ecuación (1.6).

Notemos que el paso crucial en el anterior proceso es encontrar una solución en enteros a la ecuación $3x + 7y = 1$. Este paso no siempre resulta tan fácil de sortear. Para el caso desarrollado, simplemente adivinamos una solución. Pero no es inmediato adivinar una solución en enteros, por ejemplo, a la ecuación

$$98x + 199y = 1.$$

Como vimos, la teoría de números se ha encargado de estudiar este problema y uno de los métodos que se emplea para resolverlo se basa en el algoritmo de Euclides. Existe, sin embargo, una interesante variedad de métodos que se han inventado para hacer frente a este problema, entre los cuales mencionamos las fracciones continuas, la sucesión de Farey y la aritmética modular.

En este momento, para poder resolver en todo caso la ecuación $ax + by = 1$, donde $mcd(a, b) = 1$, desarrollaremos sólo uno de los anteriores la parte pertinente de la teoría de las fracciones continuas.

1.5 Fracciones continuas

La irrupción en la escena matemática de las fracciones continuas se dio cuando Rafael Bombelli (s.XVI) las usó para hallar aproximaciones de las raíces irracionales de ecuaciones polinómicas, tema que examinaremos en mayor detalle en el capítulo siguiente. Euler y Lagrange (s.XVIII) desarrollaron la teoría de fracciones continuas de manera sistemática y haremos uso de esta teoría para concretar nuestro método de solución de las ecuaciones diofánticas lineales. Veamos una versión contemporánea de esa teoría.

1.5.1 Notación básica y ejemplos

Comencemos por expresar un número racional como fracción continua. El primer paso es expresar el número como la suma de su parte entera y su parte fraccionaria, donde es evidente que ésta está entre 0 y 1.

Tenemos

$$\frac{p}{q} = a_1 + \frac{r_1}{q}, \qquad 0 \leq r_1 < q.$$

Obviamente, si $r_1 = 0$, el proceso termina y se tiene $\frac{p}{q} = a_1$ (se trata de un número entero). Si $r_1 \neq 0$, escribimos

$$\frac{p}{q} = a_1 + \frac{1}{\frac{q}{r_1}}, \qquad 0 < r_1 < q.$$

Ahora bien, se sigue que $\frac{q}{r_1} > 1$ y podemos iterar el proceso de separarlo en su parte entera y su parte fraccionaria como sigue

$$\frac{q}{r_1} = a_2 + \frac{r_2}{r_1}, \qquad 0 \leq r_2 < r_1.$$

Ahora bien, si $r_2 = 0$, el proceso termina, $\frac{q}{r_1} = a_2$ y

$$\frac{p}{q} = a_1 + \frac{1}{a_2},$$

que denotaremos $(a_1; a_2)$ y llamaremos la expansión de $\frac{p}{q}$ en fracción continua. Si $r_2 \neq 0$, el proceso continua de manera similar. Ahora bien, tal como en el caso del Algoritmo de Euclides, tenemos una sucesión decreciente de números enteros positivos $q > r_1 > r_2 > \ldots$, de donde, es claro que, en algún momento el residuo será 0 y el proceso terminará.

Por ejemplo, la expansión de $\frac{19}{7}$ en fracción continua es

$$\frac{19}{7} = 2 + \frac{5}{7} = 2 + \frac{1}{\frac{7}{5}} = 2 + \frac{1}{1 + \frac{2}{5}} = 2 + \frac{1}{1 + \frac{1}{\frac{5}{2}}} = 2 + \frac{1}{1 + \frac{1}{2 + \frac{1}{2}}}.$$

Ejercicios

1. Hallar la expansión en fracción continua de (a) $\frac{1}{13}$; (b) $\frac{34}{21}$; (c) $\frac{17}{13}$.
2. Hallar el número cuya expansión en fracción continua es $3 + \frac{1}{2 + \frac{1}{3 + \frac{1}{4}}}$.

Punto de discusión

¿Existe alguna relación evidente entre la expansión en fracción continua de 13/7 y la de 11/7? ¿Hay alguna información sobre un número racional que por simple inspección se puede obtener a partir de su expansión en fracción continua?

Ahora bien, consideremos la expansión en fracción continua de un número irracional x. Nuevamente, separamos la parte entera y la fraccionaria de x, obteniendo

$$x = a_1 + \frac{1}{x_1}, \qquad 0 < \frac{1}{x_2} < 1,$$

donde el número $x_2 = \frac{1}{x - a_1} > 1$ es irracional. Una vez mas, podemos expresar x_2 en la forma

$$x_2 = a_2 + \frac{1}{x_3}, \qquad 0 < \frac{1}{x_3} < 1, \qquad a_2 \geq 1.$$

Nuevamente, el número $x_3 = \frac{1}{x_2 - a_2}$ es mayor que 1 e irracional. Se sigue que este proceso puede continuar indefinidamente, dando lugar a la sucesión de ecuaciones

$$x = a_1 + \frac{1}{x_2}, \qquad x_2 > 1,$$

$$x_2 = a_2 + \frac{1}{x_3}, \qquad x_3 > 1, \quad a_2 \geq 1,$$

$$\ldots \qquad \ldots \qquad \ldots$$

$$x_n = a_n + \frac{1}{x_{n+1}}, \qquad x_{n+1} > 1, \quad a_n \geq 1,$$

en la cual los a_i son todos enteros y los x_i son todos irracionales. Al contrario de la expansión en fracción continua de un número racional, la expansión de un irracional no puede terminar ya que esto sólo podría darse si uno de los a_n fuera igual a uno de los x_n, que es claramente imposible.

$$x = a_1 + \frac{1}{x_2} = a_1 + \frac{1}{a_2 + \frac{1}{x_3}} = a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{x_4}}} = \ldots$$

Este proceso lleva a la expresión

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n + \cfrac{1}{x_{n+1}}}}}.$$

existencia de una raíz en determinado intervalo y, por ende, una sustentación del método de aproximación sólo se lograría con la geometría analítica de Descartes y el cálculo de Newton y Leibniz (s. XVII).

3.2 Ecuación de tercer grado

En esta sección estudiaremos la solución al problema de la ecuación cúbica dada por los algebristas italianos y algunos de sus sucesores.

3.2.1 El *Ars Magna* de Girolamo Cardano

En 1545 Girolamo Cardano publicó su obra *Ars Magna* en la cual expone los métodos de solución de la ecuación cúbica descubiertos por Nicola Tartaglia y extendidos por el mismo Cardano a nuevos casos. El *Ars Magna* es, desde todo punto de vista, un libro extraordinario que demuestra hasta qué punto la mentalidad del matemático italiano del Renacimiento, en la gran tradición de Leonardo de Pisa ha roto con las ataduras mentales de la matemática griega (y también hasta qué punto permanecen cuestiones heredadas de sus antecesrores matemáticos que no han sido resueltos).

Consideraciones sobre las raíces de ciertas ecuaciones

El libro comienza con una discusión de lo que Cardano llama soluciones dobles. Considera las raíces positivas (denominadas por Cardano verdaderas) y negativas (falsas) que surgen de ecuaciones donde intervienen potencias pares de las incógnitas como las siguientes.

$$x^2 = 9; x^4 = 81; x^4 + 3x^2 = 28; x^4 + 12 = 7x^2; x^4 + 12 = 6x^2.$$

Analizando cada caso, Cardano nos dice. La primera tiene una raíz verdadera y una falsa (que escribiremos 3 y -3), la segunda también tiene una verdadera y una falsa, a saber, 3 y -3. Si se suman el cuadrado del cuadrado con el cuadrado, como en la tercera, tendremos el mismo caso, es decir una raíz verdadera y una falsa, en este caso 2 y -2.

En cambio, en la cuarta y quinta ecuación se suman el cuadrado de un cuadrado con un número. Cardano distingue entre ellos, pues en la cuarta, se tienen cuatro soluciones, 2 y -2, $\sqrt{3}$ y $-\sqrt{3}$. En cambio en la quinta, en palabras de Cardano, si no hay una solución verdadera tampoco hay una falsa.

Puntos de discusión

1. Nótese que Cardano considera tres casos para la ecuación de grado cuatro, a saber, cuatro raíces, dos raíces y ninguna raíz. ¿Qué tipo de raíz acepta Cardano que no sería aceptada por los matemáticos de la tradición griega-árabe?

2. Notamos que aquí tenemos lo que se denomina ecuación bicuadrática, pues en ellas sólo aparecen potencias de x^2. ¿En qué sentido pueden éstas reducirse al caso de la cuadrática y en qué sentido no? ¿Son éstos todos los casos posibles de la ecuación bicuadrática?

En cuanto a las potencias impares, los ejemplos que Cardano considera son

$$\begin{aligned} 2x &= 16; 2x^3 = 16; x^3 + 6x = 20; x^3 + 16 = 12x; x^3 = 1 \\ x^3 + 9 &= 12x; x^3 + 21 = 2x; x^3 = 2x + 21 \end{aligned}$$

Su conclusión relativa a las primeras tres ecuaciones es que cuando una potencia impar o aún mil de ellas se comparan con un número siempre hay una solución verdadera.

Puntos de discusión

1. ¿A qué tipo de solución está haciendo referencia Cardano aquí?

2. ¿Es cierto que toda ecuación polinómica de grado impar tiene una raíz real positiva?

En cuanto a la cuarta ecuación de esta lista, Cardano dice que hay que determinar primero si $\frac{2}{3}$ del coeficiente del término en x por la raíz cuadrada de un tercio del mismo coeficiente (aquí $\frac{2}{3}(12) \cdot \sqrt{\frac{12}{3}}$) es mayor, menor o igual al término constante. En el caso de que sean iguales, Cardano nos dice que la ecuación tiene una raíz verdadera y una falsa. Por ejemplo, respecto de la ecuación $x^3 + 16 = 12x$, según Cardano, las soluciones son 4 y -2.

Puntos de discusión

1. Considerar los últimos cinco ejemplos dados por Cardano en (**). Determinar cuántas soluciones positivas y cuántas negativas tiene cada ecuación.

2. ¿Cardano tiene en cuenta las soluciones que nosotros llamamos complejas?

3. ¿Cardano considera que pueden existir raíces múltiples?

4. De acuerdo con esta discusión ¿es posible pensar en una propiedad general como, por ejemplo, que una ecuación polinómica de grado 3 tenga exactamente 3 raíces?

Cardano alarga su discusión mirando casos en los cuales se mezclan potencias pares e impares y extensiones de la aplicación de la fórmula cuadrática a la bicuadrática.

Punto de discusión

El francés J.E. Montucla, historiador de la matemática, dice de Cardano que

> "Cardano fue el primero en percibir la multiplicidad de raíces que puede tener la incógnita en una ecuación y la distinción entre positivos y negativos. ... Sin embargo, debemos notar, para no darle demasiado crédito a Cardano que su descubrimiento no se desarrolló plenamente; no sólo no dijo nada acerca del uso de estas raíces negativas, las cuales probablemente creó inútiles, sino que tuvo un error en cuanto a las ecuaciones que tienen varias raíces iguales con el mismo signo.... No obstante, podemos excusar este error en una época en la cual se usaba el álgebra únicamente para la solución de problemas numéricos."

Discutir esta cita de Montucla a la luz de los anteriores ejemplos dados por Cardano.

Transformación de ecuaciones

Un segundo tema discutido por Cardano es la transformción de ecuaciones. Cardano muestra aquí, por ejemplo, que ecuaciones polinómicas con coeficientes racionales pueden transformarse a ecuaciones en los enteros. Para ello, presenta la ecuación $x + \frac{1}{3}x^2 + \frac{1}{9}x^3 = 19$ que transforma (multiplicando por 9) en $x^3 + 3x^2 + 9x = 171$.

Trata otros ejemplos y métodos mas interesantes que, en efecto, muestran que ciertas transformaciones pueden introducir coeficientes irracionales. Algunos ejemplos se escriben a continuación.

1. Dada la solución a la ecuación $x^2 = 6x + 16$, se puede encontrar la solución a la ecuación $x^2 + 6x = 16$. Basta restar el coeficiente de x (o sea, 6) de la solución a la primera ecuación.

2. Dada la ecuación $x^3 + 8 = 18x$ ésta puede convertirse en la ecuación $x^3 + 8 = 9x^2$.

3. La ecuación $x^4 + 8 = 12x$ puede transformarse en la ecuación $x^4 + 512 = 12x^3$.

Punto de investigación

Examinar cada una de las anteriores transformaciones. Determinar bajo cuáles condiciones son correctas y bajo cuáles condiciones transformaciones como las registradas en 2 y 3 producirán coeficientes irracionales.

Solución de las ecuaciones cúbicas

En seguida de la discusión de transformaciones Cardano expone su solución a la ecuación cúbica. Todos los ejemplos que presenta son numéricos con coeficientes enteros positivos y todas sus justificaciones son geométricas. Varios de los ejemplos que considera se dan a continuación.

1. Cubo y primera potencia igual a número. Resuelve $x^3 + 6x = 20, x^3 + 3x = 10; x^3 + 6x = 2$.

Cardano atribuye la solución de ecuaciones de esta forma a Scippio del Ferro de Bologna quien lo pasó a Antonio María Fior de Venecia. De acuerdo con Cardano, Fior participó en una competencia de solución de ecuaciones cúbicas con Niccolò Tartaglia de Brescia en el transcurso del cual Tartaglia descubrió la solución de del Ferro. Finalmente, Tartaglia reveló el método a Cardano.

La regla de solución que aduce Cardano es: Elevar un tercio del coeficiente del término en x al cubo; sumar el cuadrado de un medio del término constante. Sacar raíz cuadrada de la suma. Escribir este valor dos veces. A uno sumar un medio del número que ya ha elevado al cuadrado; al otro restar un medio del mismo número.

Punto de discusión

Aplicar el método de Cardano a la primera ecuación que él considera. ¿Qué observa usted?

2. Cubo igual a primera potencia mas número. Resuelve $x^3 = 6x + 40, x^3 = 6x + 6$.

Punto de discusión

¿Se puede usar el mismo método anterior? ¿Cuáles modificaciones hay que hacerle? ¿Hay solución en todo caso? ¿Bajo cuáles condiciones puede llevar el procedimiento a tomar la raíz cuadrada de un número negativo?

3. Cubo y número igual primera potencia. Resuelve $x^3 + 3 = 8x, x^3 + 60 = 46x$.

Punto de discusión

Cardano resuelve $x^3 + 3 = 8x$ resolviendo $y^3 = 8y + 3$ y luego sumando sus dos raíces. Justificar este procedimiento.

4. Cubo igual a cuadrado mas número. Resuelve $x^3 = 6x^2 + 20$.

1. 5. Cubo y cuadrado igual número. Resuelve $x^3+6x^2 = 100, x^3+6x^2 = 16, x^3 + 6x^2 = 7$.

   6. Cubo y número igual cuadrado. Resuelve $x^3 + 64 = 18x^2, x^3 + 24 = 8x^2$.

   7. Cubo, cuadrado y primera potencia igual número. Resuelve $x^3 + 6x^2+12x = 22, x^3+3x^2+9x = 171, x^3+6x^2+x = 14, x^3+12x^2+27x = 400$.

   8. Cubo y primera potencia igual cuadrado y número. Resuelve $x^3 + 12x = 6x^2 + 25$; $x^3 + 48x = 12x^2 + 48$; $x^3 + 12x = 6x^2 + 8$.

   9. Cubo y cuadrado igual primera potencia y número. Resuelve $x^3 + 6x^2 = 20x + 56$; $x^3 + 6x^2 = 20x + 112$; $x^3 + 6x^2 = 20x + 21$.

   10. Cubo igual cuadrado, primera potencia y número. Resuelve $x^3 = 6x^2 + 72x + 729$.

   11. Cubo y número igual cuadrado y primera potencia. Resuelve $x^3 + 64 = 6x^2 + 24x$.

   12. Cubo, primera potencia y número igual al cuadrado. Resuelve $x^3 + 4x + 8 = 6x^2$; $x^3 + 4x + 16 = 6x^2$; $x^3 + 4x + 1 = 6x^2$.

   13. Cubo, cuadrado y número igual primera potencia. Resuelve $x^3 + 6x^2 + 12 = 31x$.

Punto de discusión

¿Por qué Cardano tuvo que dividir sus soluciones en casos? ¿Por qué no pudo presentar una única fórmula como en el caso de la ecuación cuadrática?

Ejercicios

En la siguiente sección se expone una versión unificada de estos casos. Cuando la haya estudiado, volver a resolver cada una de estas ecuaciones que Cardano planteó.

3.2.2 La fórmula general de solución de la ecuación de tercer grado

Consideremos ahora la ecuación general de tercer grado en una variable

$$y^3 + ay^2 + by + c = 0. \tag{3.4}$$

Para ello basta observar primero (como allí se anotaba) que dicha ecuación puede ser reducida a una de la forma

$$x^3 + px + q = 0; \tag{3.5}$$

en la cual el coeficiente de x^2 es nulo. Dicho objetivo se logra realizando una sustitución adecuada, a saber

$$y = x - \frac{a}{3}.$$

La ecuación (1) tomaría entonces la forma

$$\left(x - \frac{a}{3}\right)^3 + a\left(x - \frac{a}{3}\right)^2 + b\left(x - \frac{a}{3}\right) + c = 0.$$

Desarrollando y reduciendo términos obtendremos

$$x^3 + \left[-\frac{a^2}{3} + b\right]x + \left[\frac{2}{27}a^3 - \frac{ab}{3} + c\right] = 0.$$

Miremos ahora como Cardano (Tartaglia) determinó la solución de la cúbica reducida (2).

Consideró la ecuación

$$x^3 + px = q;$$

con $p > 0$ y $q > 0$, ecuación que denominaremos la *cúbica reducida.* Dado que el Renacimiento era un período de especial aprecio por la matemática griega y que no se había desarrollado aún medios algebraicos de demostración, no es sorprendente que su solución a esté problema se basará en argumentos geométricos.

Usó para ello la identidad algebraica

$$(a - b)^3 + 3ab(a - b) = a^3 - b^3.$$

Punto de discusión

Mostrar cómo la identidad $(a - b)^3 + 3ab(a - b) = a^3 - b^3$ se puede derivar de un modelo geométrico.

Al comparar ésta con la ecuación que nos ocupa, nos llevará a concluir que si a y b se escogen tales que

$$3ab = p$$

y

$$a^3 - b^3 = q,$$

la identidad algebraica se transformaría en

$$(a-b)^3 + p(a-b) = q,$$

con lo cual $x = a - b$ será una solución de la cúbica reducida.

El problema consiste ahora en resolver el par de ecuaciones simultáneas

$$\begin{aligned} a^3 - b^3 &= q \\ ab &= \frac{p}{3}. \end{aligned}$$

Elevando la primera de estas ecuaciones al cuadrado y la segunda al cubo tenemos

$$a^6 - 2a^3b^3 + b^6 = q^2$$

$$4a^3b^3 = 4\frac{p^3}{27}.$$

Sumando estas dos ecuaciones

$$\left(a^3 + b^3\right)^2 = a^6 + 2a^3b^3 + b^6 = q^2 + 4\frac{p^3}{27},$$

de donde,

$$a^3 + b^3 = \sqrt{q^2 + 4\frac{p^3}{27}}.$$

De esta última ecuación y de la ecuación $a^3 - b^3 = q$ podemos determinar a^3 y b^3.

$$a^3 = \frac{1}{2}\left(q + \sqrt{q^2 + 4\frac{p^3}{27}}\right)$$

$$b^3 = \frac{1}{2}\left(-q + \sqrt{q^2 + 4\frac{p^3}{27}}\right)$$

Análogamente para resolver la ecuación

$$x^3 = px + q,$$

con $p > 0$ y $q > 0$,

el correspondiente argumento geométrico requerirá de la identidad

$$(a+b)^3 = a^3 + b^3 + 3ab(a+b).$$

¡Ojo! Es natural que las consideraciones se restrinjan en estos análisis a números positivos; ¡desde luego, sin ningún problema el método puede ser generalizado y presentarse con la versión conocida en la actualidad!

Consideremos entonces el caso general. Dada la ecuación (1) $y^3 + ay^2 + by + c = 0$, ésta puede ser reducida a una de la forma (2) $x^3 + px + q = 0$ haciendo la sustitución $y = x - \frac{a}{3}$, como lo anotamos anteriomente.

Considerando la ecuación reducida $x^3 + px + q = 0$, introducimos dos nuevas variables u y v, tales que $x = u + v$, la ecuación se transforma en una eucación cúbica en dos variables,

$$(u + v)^3 + p(u + v) + q = 0$$

o

$$u^3 + v^3 + q + (3uv + p)(u + v) = 0.$$

Cualquiera que sea la suma de los números $u + v$ siempre es posible exigir que su producto uv tnega un valor fijo previamente establecido. Si $u + v = A$ y exigimos que $uv = B$, entonces puesto que $v = A - u$, obtenemos

$$u(A - u) = B.$$

Por tanto, es suficiente que u sea solución de la ecuación cuadrática

$$u^2 - Au + B = 0,$$

y ya sabemos que las raíces de esta ecuación pueden ser determinadas sin dificultad por la fórmula cuadrática.

En el caso que nos ocupa, $u + v$ es la raíz x que se busca de la ecuación cúbica; se pide entonces que

$$uv = -\frac{p}{3};$$

esto es, $3uv + p = 0$. Escogiendo u y v con estas características obtenemos

$$u^3 + v^3 + q = 03uv + p = 0. \tag{3.6}$$

En consecuencia, si encontramos números u y v que satisfacen este sistema de ecuaciones, el número $x = u + v$ será la raíz de la ecuación.

A partir del sistema (3) es fácil construir una ecuación cuadrática cuyas raíces sean u^3 y v^3.

$$u^3 + v^3 = -q$$

$$u^3v^3 = -\frac{p^3}{27}$$

Por relación entre coeficientes y raíces de la ecuación cuadrática, tenemos que u^3 y v^3 son raíces de la cuadrática

$$z^2 + qz - \frac{p^3}{27} = 0.$$

Resolviendo ésta por la fórmula usual obtenemos,

$$u^3 = -\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}} \tag{3.7}$$

$$v^3 = -\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}, \tag{3.8}$$

de donde,

$$x = -\sqrt[3]{\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

Nota. Obsérvese que cada una de las cantidades del lado derecho de (2.2) y (2.3) tiene tres raíces cúbicas. Podríamos entonces pensar que x tiene nueve valores, pero el siguiente argumento muestra que éste no es el caso. Dado que $uv = -\frac{p}{3}$ las raíces cúbicas son tomadas en pares para que el producto de cada par sea racional.

Por tanto, si u y v denotan los valores de algún par de raíces cúbicas que satisfacen esta condición, los otros pares admisibles serían ωu, $\omega^2 v$ y $\omega^2 u$, ωv, donde ω y ω^2 son las raíces complejas cúbicas de la unidad, es decir, raíces de la ecuación $z^3 - 1 = 0$.

Puntos de discusión

1. Resolver la ecuación $z^3 - 1 = 0$ observando que $z^3 - 1 = (z-1)(z^2 + z + 1) = 0$.

2. Sea ω una de las raíces de $z^2 + z + 1 = 0$. Calcular ω^2 y ω^3 directamente a partir de ω. ¿Qué se observa?

3. Al aplicar este procedimiento, Cardano consideraría un solo par de valores para u y v. ¿Por qué?

Por tanto las raíces de la cúbica reducida serían: $u+v$; $\omega u+\omega^2 v$; $\omega^2 u+\omega v$.

Ejemplo. Analicemos la solución de la cúbica reducida

$$x^3 - 15x = 126.$$

Haciendo $x = u + v$ en la ecuación tenemos

$$u^3 + v^3 + (3uv - 15)x = 126.$$

Basados en las condiciones que se analizaron anteriormente para u y v,

$$\begin{aligned} 3uv - 15 &= 0 \\ u^3 + v^3 &= 126 \\ u^3 v^3 &= 125. \end{aligned}$$

u^3 y v^3 son las raíces de la ecuación

$$z^2 - 126z + 125 = 0.$$

Por consiguiente, $u^3 = 125$ y $v^3 = 1$; $u = 5$ y $v = 1$.

Por tanto las tres raíces de la cúbica serían

$$r_1 = u + v = 6$$

$$r_2 = \omega u + \omega^2 v = \frac{-1+\sqrt{-3}}{2}.5 + \frac{-1-\sqrt{-3}}{2} = -3 + 2\sqrt{-3}$$

$$r_3 = \omega^2 u + \omega v = -73 - 2\sqrt{-3},$$

donde ω es la *raíz tercera primitiva de la unidad* que es obtenida al solucionar en la ecuación $z^3 = 1$.

Analicemos ahora más detenidamente las raíces de la cúbica

$$x^3 + px + q = 0.$$

(i) Si $\frac{q^2}{4} + \frac{p^3}{27}$ es positivo, entonce u^3 y v^3 son ambas reales, y las raíces serían:

$$u + v; -\frac{u+v}{2} + \frac{u-v}{2}.\sqrt{-3}; -\frac{u+v}{2} - \frac{u-v}{2}.\sqrt{-3}.$$

(ii) Si $\frac{q^2}{4} + \frac{p^3}{27}$ es cero, entonces $u^3 = v^3$; en este caso $u = v$, y las raíces serían:

$$2u; u(\omega + \omega^2); u(\omega + \omega^2),$$

o sea, $2u$; $-u$; $-u$.

Punto de discusión

¡Estamos afirmando aquí que $\omega + \omega^2 = -1$! ¿Por qué?

(iii) Si $\frac{q^2}{4} + \frac{p^3}{27}$ es negativo, u^3 y v^3 serían complejos; sus raíces cúbicas serían también complejas de la forma $m+in$ y $m-in$; las raíces de la reducida serían en este caso: $2m$; $-m - n\sqrt{3}$; $-m + n\sqrt{3}$, todas reales.

Ejercicios

Resuelva los siguientes ejercicios usando el método anteriormente descrito.

Ejercicio 1 $x^3 + 11x = 6x^2 + 6$

Ejercicio 2 $x^3 + 6x^2 + x = 14$

Ejercicio 3 $x^3 + 6x^2 = 20x + 56$

Ejercicio 4 $x^3 + 64 = 6x^2 + 24$

Punto de investigación.

1. El método de Viète. Para resolver la cúbica reducida $x^3 + ax = b$, Vieta usó la sustitución $x = \frac{a}{3y} - y$. Determinar la solución usando este método y aplicarlo para resolver las ecuaciones

(a) $x^3 + 81x = 702$

(b) $x^3 + 6x^2 + 18x + 15 = 0$

2. Haciendo la sustitución $x = y + \frac{5}{y}$, encuentre una raíz de la cúbica $x^3 = 15 + 126$.

3.3 Solución de la ecuación de cuarto grado

Tiempo después de resolver la ecuación cúbica, Luigi Ferrari (1522-1565), encontró una solución para la ecuación de grado cuatro. El desafío para resolver una ecuación de grado cuatro fue lanzado en 1540 por el matemático italiano Zuanne de Tonini da Coi quien propuso a Cardano el siguiente problema.

"Dividir 10 en tres partes en proporción continua de forma que el producto de las dos primeras sea 6."

Luigi Ferrari, secretario de Cardano, llegó a la solución de este problema a través de razonamientos geométricos que ilustraremos a continuación.

En el lenguaje actual el problema se traduce en determinar números positivos x, y, z tales que

$$x + y + z = 10$$
$$\frac{y}{x} = \frac{x}{z}$$
$$x \cdot y = 6.$$

Combinando estas ecuaciones, el problema se transforma en resolver la ecuación

$$\frac{6}{x} + x + \frac{x^3}{6} = 10,$$

que es equivalente a la cuártica

$$x^4 + 6x^2 + 36 = 60x.$$

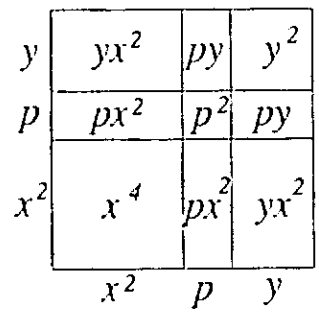


Figure 3.3:

El primer paso consistió en establecer la identidad (Figura 3.3).

$$\left(x^2+p+y\right)^2 = x^4+p^2+y^2+2x^2p+2x^2y+2py$$

A continuación Ferrari propuso completar en el lado izquierdo de la ecuación un trinomio cuadrado perfecto (adicionando el término adecuado)

$$x^4+12x^2+36 = 60x+6x^2,$$

obteniendo

$$\left(x^2+6\right)^2 = 6x^2+60. \tag{3.9}$$

A continuación adiciona a ambos lados de la ecuación los términos necesarios para que, al introducir una nueva variable y, el miembro de la izquierda continúe siendo un cuadrado perfecto. Tenemos

$$\left(x^2+6+y\right)^2 = 6x^2+60x+y^2+12y+2x^2y = (2y+6)\,x^2+60x+\left(y^2+12y\right) \tag{3.10}$$

El problema consiste ahora en seleccionar y de manera que el trinomio de la derecha sea un cuadrado perfecto; para ello es claro que el discriminante de la ecuación cuadrática asociada en x^2 debe ser 0..

$$D = (60)^2-4(2y+6)(y^2+12y) = 0.$$

Esta condición da origen a una cúbica en y

$$y^3+15y^2+36y = 450$$

que puede ser resuelta por el método descrito en la sección anterior (hallando la reducida). Sustituyendo el valor de y así obtenido en (3.1) y extrayendo raíz cuadrada resultará una cuadrática que al ser resuelta nos permite encontrar la solución de la ecuación cuártica original.

Es claro que, de manera análoga a la solución de la cúbica que requería de la solución de una cuadrática auxiliar, la solución de la cuártica dada por Ferrari está basada en la solución de una cúbica auxiliar, conocida como la *resolvente cúbica.*

Describamos ahora el método de solución en el caso general. Consideremos sin pérdida de generalidad la cuártica simplificada (es decir, con coeficiente de x^4 igual a uno)

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

que puede ser escrita como

$$x^4 + ax^3 = -bx^2 - cx - d.$$

Si sumamos en ambos miembros de esta ecuación $\frac{a^2x^2}{4}$, obtenemos en uno de los lados un cuadrado perfecto

$$x^4 + ax^3 + \frac{a^2x^2}{4} = -bx^2 - cx - d + \frac{a^2x^2}{4}$$

$$\left(x^2 + \frac{ax}{2}\right)^2 = \left(\frac{a^2}{4} - b\right)x^2 - cx - d.$$

Si ahora adicionamos a ambos lados de esta expresión el término $\left(x^2 + \frac{ax}{2}\right)y + \frac{y^2}{4}$ siendo y una nueva variable, tendremos

$$\left(x^2 + \frac{ax}{2} + \frac{y}{2}\right)^2 = \left(\frac{a^2}{4} - b + y\right)x^2 + \left(\frac{ay}{2} - c\right)x + \left(\frac{y^2}{4} - d\right).$$

Nos interesa analizar las condiciones bajo las cuales el trinomio cuadrado en x cuyos coeficientes dependen de y sea un cuadrado perfecto. Recordemos que para que un trinomio cuadrado de la forma $Ax^2 + Bx + C$ sea el cuadrado de un binomio de la forma $\alpha x + \beta$, es suficiente que $B^2 - 4AC = 0$.

Trasladando esta idea a nuestro problema, y debe ser tal que

$$\left(\frac{ay}{2} - c\right)^2 - 4\left(\frac{a^2}{4} - b + y\right)\cdot\left(\frac{y^2}{4} - d\right) = 0.$$

Desarrollando, obtenemos la expresión

$$y^3 - by^2 + (ac - 4d)y - [d(a^2 - 4b) + c^2] = 0,$$

que es la resolvente cúbica.

Resolviendo esta cúbica determinaremos los valores de α y β en términos de su solución y_0, para los cuales

$$\left(x^2 + \frac{ax}{2} + \frac{y_0}{2}\right)^2 = (\alpha x + \beta)^2,$$

de donde,

$$x^2 + \frac{ax}{2} + \frac{y_0}{2} = \alpha x + \beta,$$

o,

$$x^2 + \frac{ax}{2} + \frac{y_0}{2} = -\alpha - \beta.$$

De estas últimas expresiones encontramos las raíces de la cuártica.

Ejemplo. Analicemos ahora las raíces de una ecuación cuártica particular aplicando el *Método de Ferrari*. Consideremos la ecuación

$$x^4 + 8x^3 + 9x^2 - 8x - 10 = 0,$$

que puede ser escrita

$$x^4 + 8x^3 = -9x^2 + 8x + 10.$$

Si adicionamos en ambos lados de la ecuación $16x^2$, obtenemos en uno de los miembros un cuadrado perfecto

$$\left(x^2 + 4x\right)^2 = 7x^2 + 8x + 10.$$

Si ahora adicionamos en ambos lados de la expresión $(x^2 + 4x)y + \frac{y^2}{4}$ siendo y una nueva variable, tendremos

$$\left(x^2 + 4x + \frac{y}{2}\right)^2 = (7 + y)x^2 + (4y + 8)x + \left(\frac{y^2}{4} + 10\right).$$

Nos interesa determinar las condiciones bajo las cuales el trinomio cuadrado en x (miembro derecho) cuyos coeficientes dependen de y sea un cuadrado perfecto. y debe ser tal que

$$(4y + 8)^2 - 4(7 + y)\left(\frac{y^2}{4} + 10\right) = 0.$$

Transformando esta nueva ecuación, tenemos

$$y^3 - 9y^2 - 24y + 216 = 0.$$

Aplicando a esta última ecuación el *Método de Cardano*, haciendo $y = t+3$, obtenemos la cúbica reducida

$$t^3 - 51t + 90 = 0,$$

que ya sabemos resolver.

Nos interesa estudiar otros dos métodos que se desarrollaron para resolver las ecuaciones cuárticas, los métodos de Vieta y de Descartes. Para llegar a sus planteamientos debe darse un paso fundamental en el álgebra que abre el camino hacia consideraciones de factorización.

3.4 Una idea genial y una observación clave de Francois Viète

Se debe a Francois Viète (1540-1603) nacido en Fontenay-le-Comte y educado en Poitiers (y conocido en castellano con el nombre de Vieta) la institucionalización de un simbolismo algebraico que, aunque deficiente en cuanto a notación de operaciones y exponentes, permitió el tratamiento de ecuaciones (y otras expresiones) algebraicas generales. La idea clave de Viète es la de usar ciertas letras para representar incógnitas o variables y otras letras para representar coeficientes o constantes. Los historiadores reportan distintas convenciones, unos diciendo que para Viète las consonantes representaban cantidades conocidas y las vocales incógnitas, otros que las letras del principio del alfabeto representaban variables y las del final constantes. Sin embargo, el punto importante es precisamente la posibilidad de representar expresiones generales, hablar de la ecuación general de segundo grado o la ecuación general de tercer grado y mostrar con ello propiedades compartidas por todas, mientras que con la notación anterior a Viète sólo se podía mostrar un caso numérico particular que ilustraba, sin generalidad total, dichas propiedades (y con alguna frecuencia mantenía ocultas varias propiedades importantes).

Ahora bien, todo matemático o profesor de matemáticas sabe que la buena notación es clave para adelantar investigaciones matemáticas, dominar nuevas áreas del conocimiento matemático o resolver problemas. Los frutos de esta innovación de Viète no se hicieron esperar. En particular, el mismo Viète observó las relaciones que existen entre las raíces y los coeficientes de una ecuación polinómica. En su escrito *De equationen emendatione*, Viète dice (traducido a nuestra notación para exponentes y operaciones):

> "Si $A^3+(-B-D-G)A^2+(BD+BG+DG)A = BDG$, entonces A es igual a cualquiera de las cantidades B, D o G."

Punto de discusión

Explicitar las relaciones entre raíces y coeficientes de la ecuación cúbica reconcidas por Viète de acuerdo con sus palabras.

Que esta realización era fruto de la notación es claro. Si bien es cierto que la dependencia sobre el método de completar el cuadrado usado por al-Khwarizmi puede dar alguna idea de la relación entre raíces y coeficientes (¿por qué?), se pone rápidamente de manifiesto que una relación tan directa está escondida tanto por la variedad en los coeficientes mismos como por los métodos geométricos (intersección de cónicas) usadas en su solución por matemáticos como Omar Khayyam. Por ejemplo, ¿qué se puede decir acerca de las interrelaciones entre estos 'casos' de la cúbica: $x^3 + 3x^2 + 3x + 1$ y $x^3 - 7x + 6$? Parecen esencialmente distintos. Y en cuanto a su solución por intersección de cúbicas, sabemos que Omar Khayyam tuvo que emplear distintas secciones cónicas de acuerdo a los términos que intervienen en la ecuación. Así su tratamiento, en lugar de mostrar relaciones generales entre raíces y coeficientes de la cúbica, parece implicar que no existe tal generalidad.

Punto de discusión

Precisar en sus propias palabras en qué sentido este descubrimiento de Viète sólo puede hacerse cuando se dejan los ejemplos numéricos atrás y se representan los coeficientes de una ecuación por letras.

Nuevamente, los grandes algebristas italianos del siglo XVI como Cardano, aunque cuentan entre sus conquistas métodos estrictamente algebraicos de solución de las cúbicas, también se ven obligados a dividir éstos en casos, empleando métodos ligeramente distintos para cada caso y presentando sus resultados por intermedio de ejemplos numéricos generalizables. Quizás sí vale la pena comentar que la reducción de un caso a otro, efectuada por ellos por medio de una manipulación algebraica, debió indicar que las diferencias observadas no eran tan esenciales como aparecen en el tratamiento geométrico, abriendo así un espacio a la mente unificadora de Viète.

Ahora bien, los resultados directamente atribuibles a Viète no son tan generales como los que hoy día manejamos, ya que sólo reconocía como válidas las soluciones reales positivas de la ecuación. Así las cosas, Viète muestra que los coeficientes de una ecuación polinómica cuyas raíces son reales positivas son funciones de esas raíces.

Punto de discusión

Nuestro estudio de la obra de Cardano nos indica que él no usó coeficientes literales sino numéricos. Tampoco consideró coeficientes negativos en sus ecuaciones pero sí reconocía las raíces negativas. Por otra parte rechazó las soluciones complejas. Comparar y contrastar las concepciones de Vieta y Cardano frente a los sistemas numéricos y la notación algebraica.

Anotamos aquí lo que parece ser una anomalía histórica. La primera publicación que habla de la solución de ecuaciones por factorización (uno de los métodos más comunmente empleados al nivel secundario) apareció en la obra de Harriot *Artis Analyticae Praxis* (1631). Dada su fácil y bonita interpretación geométrica en el caso de la cuadrática y la cúbica (véase la sección sobre Tabit ben Qurra), nos preguntamos cómo podría haber sucedido que no fuera descubierto y utilizado con anterioridad. Pero es también claro que el método de factorización depende de las relaciones observadas entre raíces y coeficientes, y fueron precisamente los medios deficientes para expresar generalidad los que bloquearon su descubrimiento.

Punto de discusión

Precisar en qué sentido el método de factorización depende de las relaciones de Vieta.

Cuando posteriormente a la época de Viète se llegue a la aceptación completa de las raíces negativas y complejas, será posible concluir con toda generalidad que los coeficientes de una ecuación polinómica son funciones simétricas de las raíces. Este hecho, combinado con otros desarrollos en el concepto de función principalmente en relación al cálculo, permitiría a los matemáticos lograr una nueva perspectiva sobre las ecuaciones polinómicas. Con ello, aparte de las ventajas mencionadas, la observación central de Viète se convertiría más adelante precisamente en la avenida de acceso a las propiedades esenciales de las ecuaciones polinomiales. En la obra de Vandermonde, Lagrange, Cauchy y finalmente Galois, es la permutación o sustitución de las raíces y la invarianza de las funciones simétricas de las raíces la clave para una teoría general de solución de ecuaciones por fórmula algebraica y la materia prima para los primeros estudios de grupos.

Pero, hay una importante observación que es necesario hacer aquí. Si bien Viète impulsa el álgebra hacia nuevos niveles de generalidad que anticipan su conversión en el álgebra abstracta que conocemos, la mentalidad matemática de Viète no deja de ser clásica, pues su rechazo de raíces negativas y complejas, conjuntamente con el cuidado que ejerció para que sus ecuaciones fueran homogéneas nos muestra que Viète siguió trabajando dentro del marco teórico de la geometría y la posibilidad de interpretar y representar, tanto las ecuaciones mismas como sus soluciones, geométricamente.

3.5 Otros métodos de solución de la ecuación cuártica

Método de Descartes

Analizaremos ahora el método de solución planteado por Descartes a una cuártica del siguiente tipo

$$x^4 + ax^2 + bx + c = 0.$$

Basándose en la noción de factorización, Descartes asume que el polinomio de grado 4 asociado a esta ecuación puede ser factorizado como el producto de dos polinomios (especiales) de grado dos, a saber,

$$x^4 + ax^2 + bx + c = (x^2 + dx + e)(x^2 - dx + f).$$

Igualando coeficientes (aplicación del teorema de identidad), tenemos $e + f - d^2 = a$; $d(f - e) = b$; $c = ef$. Combinando estas relaciones obtenemos la siguiente ecuación

$$d^6 + 2ad^4 + (a^2 - 4c)d^2 - b^2 = 0,$$

que es una cúbica en d^2.

Como esta cúbica siempre tiene una raíz real, d^2 es conocido y pueden determinarse los valores de e y f. Finalmente la solución de la ecuación original se encontrará al resolver $x^2 + dx + e = 0$ y $x^2 - dx + f = 0$.

Miremos un ejemplo.

Ejemplo. Resolver la ecuación cuártica $x^4 + 3x^2 - 6x - 2 = 0$. Para ello, supongamos que el polinomio puede ser factorizado

$$x^4 - 3x^2 - 6x - 2 = (x^2 + dx + e)(x^2 - dx + f).$$

Igualando coeficientes, tenemos $e + f - d^2 = -3$; $d(f - e) = -6$; $ef = -2$. Combinando estas relaciones obtenemos la ecuación .

$$d^6 - 6d^4 + 7d^2 - 36 = 0,$$

que es una cúbica en d^2.

Haciendo $y = d^2$, nuestro problema consiste ahora en resolver

$$y^3 - 6y^2 + 7y - 36 = 0,$$

que a su vez con la sustitución $y = t + 2$ se transforma en

$$t^3 - 5t - 38 = 0.$$

Usando las fórmulas de Cardano con $p = -5$; $q = -38$, obtenemos

$$t = \sqrt[3]{19 + \sqrt{\frac{19^2}{4} + \frac{-5^3}{27}}} + \sqrt[3]{19 - \sqrt{\frac{19^2}{4} + \frac{-5^3}{27}}}.$$

Con el valor de t seleccionado aquí, determinamos y, d, e y f, y con estos valores podremos identificar los factores cuadráticos y por ende la solución de la ecuación original.

Ejercicios

Usando el método de Ferrari o de Descartes según el caso, resuelva las siguientes ecuaciones.

Ejercicio 5 $x^4 + 3 = 12x$

Ejercicio 6 $x^4 + 6x^2 + 8x + 21 = 0$, *si sabemos que 4 es una solución de la resolvente cúbica.*

Ejercicio 7 $x^4 + 9x + 4 = 4x^2$ *si se sabe que* $\frac{13}{2}$ *es solución de la resolvente cúbica.*

Ejercicio 8 $x^4 + 6x^3 = 6x^2 + 30x + 11$

El método de Vieta

Vieta planteó un método para resolver una bicuadrática distinto al planteado por Descartes. Veamos. Dada la ecuación $x^4 + py^2 + qy + r = 0$, sugiere sumar a ambos lados de la ecuación $x^2z^2 + (\frac{1}{4})z^4$; donde z es una nueva variable.

Punto de discusión

Obtenga la resolvente cúbica aplicando esta idea. ¿Por qué funciona esta sustitución?

Ejercicio 9 *Use este método para resolver*

$$y^4 - y^3 + y^2 - y = 10.$$

3.6 Caracterización de las raíces de la cúbica y la cuártica usando el discriminante

En el inicio de la discusión de la solución de ecuaciones polinómicas presentamos la muy conocida caracterización de las raíces de la ecuación cuadrática a través del análisis de su *discriminante*, pero no hemos tocado esta idea en

el caso de la cúbica y la cuártica. En primer lugar es pertinente aclarar que podemos hablar del discriminante para una ecuación polinómica arbitraria de grado n, $n > 1$ y es realmente interesante hacerlo pues nos proporciona una herramienta potente para analizar el carácter de las raíces.

Definición 1 *Sea* $p(x) = a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_0 = 0$, *con* $a_i \in R$ *una ecuación polinómica de grado* n *con* $n > 1$ *y sean* $r_1, r_2, \ldots, r_n$ *sus raíces. Entonces el* discriminante *de* $p(x)$ *es*

$$D = (r_1 - r_2)^2(r_1 - r_3)^2 \cdots (r_1 - r_n)^2(r_2 - r_3)^2 \cdots (r_{n-1} - r_n)^2 = \prod_{i,j=1}^{n} (x_i - x_j)^2,$$

con $i < j$.

¡Nótese que $D = 0$ si y sólo si $p(x)$ tiene una raíz múltiple! Si $n = 2$, considerando la polinómica simplificada $p(x) = x^2 + ax + b$ con r_1 y r_2 sus raíces,

$$D = (r_1 - r_2)^2 = (r_1 + r_2)^2 - 4r_1r_2 = (-a)^2 - 4b = a^2 - 4b.$$

Si a y b son reales (como lo estamos considerando aquí) el signo de D determina cuando las raíces de la polinómica $p(x) = 0$ son reales o complejas. Algo similar ocurrirá en la cúbica. Veamos.

Teorema 3.6.1 *Si* $x^3 + ax^2 + bx + c$ *tiene coeficientes reales, entonces*

(1) Si $D > 0$ todas las raíces son reales y distintas.
(2) Si $D = 0$ todas las raíces son reales y existe una raiz multiple.
(3) Si $D < 0$ existe una raiz real y dos complejas conjugadas.

Demostración. Como los coeficientes son reales, las raíces complejas ocurren en pares conjugados (¿En qué se basa esta afirmación?) Por lo tanto al menos una raíz es real. Si las otras raíces son reales y no hay dos raíces iguales,

$$D = (r_1 - r_2)^2(r_1 - r_3)^2(r_2 - r_3)^2 > 0.$$

Si las otras son reales y una raíz es múltiple entonces $D = 0$. Si las otras raíces son complejas conjugadas, a saber, $\alpha + \beta i$, $\alpha - \beta i$. α y β reales y $\beta \neq 0$, entonces

$$D = [r_1 - (\alpha + \beta i)]^2 [r_1 - (\alpha - \beta i)]^2 [\alpha + \beta i - (\alpha - \beta i)]^2 = \left[(r_1 - \alpha)^2 + \beta^2\right]^2 \cdot (-4\beta^2) < 0.$$

Por tanto, si $D > 0$ las tres raíces son reales. ¡(1) y (3) se demuestran de manera similar!

Ejercicio

Demostrar las partes (1) y (3) del teorema anterior.

Nota. El hecho que el signo del discriminante determine la naturaleza de las raíces de la cúbica es muy útil, dado que el discriminante puede ser identificado sin conocer explicitamente las raíces, usando por ejemplo las relaciones entre coeficientes y raíces de la ecuación que estudiaremos en la siguiente sección.

Punto de discusión

Volvamos a las consideraciones de Cardano sobre el número de raíces que puede tener una ecuación (Sección 2.1.1). ¿El criterio que usa Cardano para discriminar entre casos es equivalente al discriminante? Explicar.

Teorema 3.6.2 *Si*

$$g(y) = y^3 + ay^2 + by + c$$

y

$$g(x) = x^3 + px + q$$

es la cúbica reducida, entonces $f(y)$ y $g(x)$ tienen el mismo discriminante.

Demostración. Si s_1, s_2, s_3 son las raíces de $g(y)$, afirmamos entonces que $r_1 = s_1 + \frac{a}{3}$, $r_2 = s_2 + \frac{a}{3}$, $r_3 = s_3 + \frac{a}{3}$ son las raíces de $g(x)$. ¿Por qué? De la afirmación anterior podemos deducir que

$$r_i - r_j = \left(s_i + \frac{a}{3}\right) - \left(s_j + \frac{a}{3}\right) = s_i - s_j$$

para $i < j$. Por tanto, los discriminantes de las ecuaciones $f(x) = 0$ y $g(x) = 0$ son iguales.

Teorema 3.6.3 *El discriminante de la ecuación $x^3 + px + q = 0$ es $-4p^3 - 27q^2$.*

Demostración. Las raíces de esta ecuación son

$$r_1 = \sqrt[3]{A} + \sqrt[3]{B};\ r_2 = \omega\sqrt[3]{A} + \omega^2\sqrt[3]{B};\ \text{y}\ r_3 = \omega^2\sqrt[3]{A} + \omega\sqrt[3]{B},$$

donde $\omega = \frac{1}{2}(-1 \pm i\sqrt{3})$. A y B son los valores $\frac{1}{2}\left[-q \pm \sqrt{q^2 + 4\frac{p^3}{27}}\right]$ y $\sqrt[3]{A}\sqrt[3]{B} = -\frac{p}{3}$. Es posible deducir ahora que

$$r_1 - r_2 = (1 - \omega)\left[\sqrt[3]{A} - \omega^2\sqrt[3]{B}\right]$$

$$r_2 - r_3 = \omega(1 - \omega)\left[\sqrt[3]{A} - \sqrt[3]{B}\right]$$

$$r_3 - r_1 = \omega^2(1-\omega)\left[\sqrt[3]{A} - \omega\sqrt[3]{B}\right].$$

Punto de Investigación

¡Analice usted como se determinan estas expresiones!

Usando las relaciones anteriores podemos concluir que

$$(r_1 - r_2)(r_2 - r_3)(r_3 - r_1) = -3(1+2\omega)(A-B).$$

Pero, como $1 + 2\omega = \pm i\sqrt{3}$, y $A - B = \pm\sqrt{q^2 + 4\frac{p^3}{27}}$, podemos deducir que $D = -27q^2 + 4p^3$.

Si consideramos por ejemplo la cúbica $2x^3 - 7x + 1 = 0$, reescribiéndola tendríamos

$$x^3 - (\frac{7}{2})x + \frac{1}{2} = 0.$$

Aquí $p = -\frac{7}{2}$ y $q = \frac{1}{2}$.

Por tanto $D = -503.5 < 0$. Se deduce entonces que la ecuación tiene una raíz real y dos complejas conjugadas.

3.6.1 Problemas propuestos

1. Si a es un número real arbitrario caracterizar las raíces de la cúbica $x^3 - 3x + a = 0$.

2. ¿Cuál es en términos de a, b y c el discriminante de $y^3 + ay^2 + by + c = 0$? Use este resultado para determinar la naturaleza de las raíces de la ecuación polinómica $y^3 - 3y^2 - 3y - 2 = 0$

3. Recordemos que en el proceso de solución de la ecuación cuártica (usando el método de Ferrari) llegamos a una ecuación cúbica. (conocida como la resolvente cúbica). Es posible también ahora relacionar como en el caso de la cúbica y de la cúbica reducida los *discriminantes*! Explore usted qué relación existe entre el discriminante de la cuártica y el discriminante de su resolvente cúbica. ¡Establezca para ello relaciones entre las raíces de la cuártica y las raíces de su resolvente cúbica!

3.7 Relaciones entre los coeficientes y las raíces de una ecuación cúbica o cuártica

En el análisis de la ecuación cuadrática ya tuvimos ocasión de recordar y usar las conocidas relaciones entre coeficientes y raíces. Esto es, si consideramos

la ecuación cuadrática simplificada $x^2 + a_1x + a_2 = 0$ cuyas raíces son r_1 y r_2, concluimos que $r_1 + r_2 = -a_1$; y $r_1r_2 = a_2$. Cardano fue el primero en observar que estas relaciones también se cumplen para la ecuación cuadrática aunquando las raíces sean complejas. Tal como sugieren las ideas de Cardano y Viète, es posible realizar un análisis similar para la cúbica y la cuártica sin necesidad de exigir que las raíces sean todas reales y positivas, pero para ello es importante que usted tenga presentes dos resultados de la teoría de polinomios, el Teorema del Factor y el Teorema de Identidad.

Si sabemos que las raíces de la cúbica

$$x^3 + a_1x^2 + a_2x + a_3 = 0$$

son r_1, r_2 y r_3, por Teorema del Factor se sigue que

$$x^3 + a_1x^2 + a_2x + a_3 = (x - r_1)(x - r_2)(x - r_3),$$

de donde,

$$x^3 - a_1x^2 + a_2x + a_3 = x^3 - (r_1 + r_2 + r_3)x^2 + (r_1r_2 + r_1r_3 + r_2r_3)x - r_1r_2r_3.$$

Por Teorema de Identidad tenemos

$$\begin{aligned} r_1 + r_2 + r_3 &= -a_1 \\ r_1r_2 + r_1r_3 + r_2r_3 &= a_2 \\ r_1r_2r_3 &= -a_3. \end{aligned}$$

3.7.1 Aplicaciones

Mirémos algunas aplicaciones sencillas de estas relaciones que pueden ser trabajadas a nivel del bachillerato, y que permiten caracterizar el comportamiento de las raíces de la ecuación, hacer interrelaciones interesantes tanto al interior del álgebra misma, como son solución de sistemas, manejo de identidades, etc., así como con otras áreas de la matemática, como la aritmética (argumentos de paridad, progresiones, divisibilidad...) y solucionar completamente la ecuación, conocida una condición sobre las raíces sin necesidad de emplear mecánicamente una fórmula.

1. Resolver la ecuación

$$4x^3 + 16x^2 - 9x - 36 = 0,$$

sabiendo que la suma de dos de sus raíces es 0.

Solución. Escribiendo la ecuación en la forma

$$x^3 + 4x^2 - \frac{9}{4}x - 9 = 0$$

y llamando r_1, r_2, r_3 sus raíces, podemos aplicar las relaciones anteriores para obtener

$$\begin{aligned} r_1 + r_2 + r_3 &= -4 \\ r_1r_2 + r_1r_3 + r_2r_3 &= -\frac{9}{4} \\ r_1r_2r_3 &= 9. \end{aligned}$$

Pero además sabemos que la suma de dos de sus raíces es 0. Supongamos pues que $r_1 + r_2 = 0$. Sustituyendo, las relaciones se transforman en: $r_3 = -4$; $r_1^2 = \frac{9}{4}$, de donde, las raíces de esta cúbica son: $\pm\frac{3}{2}$ y -4.

2. Resolver la ecuación

$$3x^3 - 26x^2 + 52x - 24 = 0,$$

si sabemos que sus raíces están en progresión geométrica.

Solución. Reescribiendo la ecuación, tenemos

$$x^3 - \frac{26}{3}x^2 + \frac{52}{3}x - 8 = 0.$$

De nuevo si r_1, r_2, r_3 son sus raíces, tenemos

$$\begin{aligned} r_1 + r_2 + r_3 &= \frac{26}{3} \\ r_1r_2 + r_1r_3 + r_2r_3 &= \frac{52}{3} \\ r_1r_2r_3 &= 8. \end{aligned}$$

3. Si a, b, c son las raíces de la ecuación

$$x^3 - px^2 + qx - r = 0,$$

determinar el valor de $\frac{1}{a^2} + \frac{1}{b^2} + \frac{1}{c^2}$, esto es, determinar

$$\frac{b^2c^2 + a^2c^2 + a^2b^2}{a^2b^2c^2}.$$

Solución. Basta usar las relaciones entre raíces y coeficientes. Como $abc = r$, $a^2b^2c^2 = r^2$. De otra parte,

$$b^2c^2 + a^2c^2 + a^2b^2 = [ab + ac + bc]^2 - 2\,[abc(a + b + c)] = q^2 - 2[r(p)].$$

De lo anterior concluimos que

$$\frac{1}{a^2} + \frac{1}{b^2} + \frac{1}{c^2} = \frac{q^2 - 2pr}{r^2}.$$

4. Encontrar una ecuación polinómica cuyas raíces sean $a+b$, $a-b$, $-a+b$ y $-a-b$.

Solución. Asumimos que la ecuación polinómica tiene la forma

$$x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0.$$

Punto de discusión

¿Cuáles son las relaciones entre raíces y coeficientes en el caso de la ecuación cuártica?

Utilizando las relaciones anteriores

$$\begin{aligned} -a_3 &= (a+b) + (a-b) + (-a+b) + (-a-b) \\ a_2 &= -2a^2 - 2b^2 = a_2 \\ a_1 &= 0 = -a_3 \\ a_0 &= a_2^4a^2b^2 + b^4. \end{aligned}$$

La ecuación nos quedaría entonces

$$x^4 + (-2a^2 - 2b^2)x^2 + (a^4 - 2a^2b^2 + b^4) = 0.$$

5. Resolver la ecuación polinómica

$$x^4 - 2x^3 - 21x^2 + 22x + 40 = 0,$$

si sabemos que sus raíces están en progresión aritmética.

Solución. Sean $r_1, r_1+a, r_1+2a, r_1+3a$ las raíces de la ecuación. Entonces

$$a_3 = -(4r_1 + 6a) = -2 \tag{3.11}$$

$$a_2 = 6r_1^2 + 18ar_1 + 11a^2 = -21. \tag{3.12}$$

Combinándolas, obtenemos la ecuación cuadrática

$$r_1^2 - r_1 - 20 = 0,$$

cuyas raíces son 5 y -4. Podemos determinar ahora el valor de a y concluir que las otras dos raíces son 2 y -1.

6. Encontrar la suma de los cuadrados y la suma de los cubos de las raíces de la ecuación

$$x^4 + qx^2 + rx + s = 0.$$

Solución. Sean r_1, r_2, r_3y r_4 las raíces de la ecuación.

$$\begin{aligned} a_3 = -(r_1 + r_2 + r_3 + r_4) &= 0 \\ a_2 = r_1r_2 + r_1r_3 + r_1r_4 + r_2r_3 + r_2r_4 + r_3r_4 &= q \qquad (3.13) \\ a_1 = r_1r_2r_3 + r_1r_2r_4 + r_1r_3r_4 + r_2r_3r_4 &= -r \qquad (3.14) \\ a_0 = r_1r_2r_3r_4 &= s, \qquad (3.15) \end{aligned}$$

de donde,

$$\begin{aligned} 0 &= [(r_1 + r_2) + (r_3 + r_4)]^2 \\ &= r_1^2 + r_2^2 + r_3^2 + r_4^2 + 2r_1r_2 + 2r_1r_3 + 2r_1r_4 + 2r_2r_3 + 2r_2r_4 + 2r_3r_4 \qquad (3.16) \\ &= r_1^2 + r_2^2 + r_3^2 + r_4^2 + 2a_2. \qquad (3.17) \end{aligned}$$

Por lo tanto,

$$-2q = r_1^2 + r_2^2 + r_3^2 + r_4^2.$$

De manera análoga podemos concluir que

$$r_1^3 + r_2^3 + r_3^3 + r_4^3 = 6r.$$

3.8 El método de descenso infinito.

A continuación onsideraremos el método de descenso infinito de Fermat y su uso para la solución de problemas relacionados con la representación de enteros. Nuestro interés en este tema se arraiga en dos aspectos importantes. Primero, queremos mostrar como este método, original de Fermat, resalta una de las propiedades básicas del conjunto de los números naturales que es fundamental en toda la argumentación posterior del álgebra abstracta. En segundo lugar, queremos subrayar la naturaleza abstracta de los problemas de representación propios de la teoría de números, característica básica del álgebra moderna, es decir, mostrar que se trata de enunciados sobre la posibilidad de lograr una cierta representación sin que se basen las demostraciones en la construcción de la misma.

Fermat reclamó la distinción de haber descubierto el teorema siguiente:

Teorema 3.8.1 *El área de un triángulo rectángulo cuyos lados son números racionales no puede ser un cuadrado perfecto.*

La anotación que transcribimos a continuación y en la cual Fermat hizo tal reclamo apareció escrita en el ejemplar del libro de Bachet, una recopilación de la obra de Diofanto, en el margen al lado del Problema 20 del libro 6 de la Aritmética .

> "Esta proposición, que es mi propio descubrimiento, he podido demostrar después de mucho tiempo, no sin mucho trabajo y pensamiento. Doy la demostración aquí ya que este método posibilitará hacer unos desarrollos extraordinarios en la teoría de números.
>
> Si el área de un triángulo rectángulo (de lados racionales) fuera un cuadrado, entonces existirían dos bicuadrados (potencias cuartas) cuya diferencia sería un cuadrado. En consecuencia habría dos números cuadrados cuya suma y diferencia serían ambos cuadrados...."

Despíes de algunos argumentos adicionales, Fermat escribe:

> "Luego, si existen dos cuadrados cuya suma y diferencia son ambos cuadrados, también existirían otros dos cuadrados enteros que tienen la misma propiedad, pero cuya suma es menor. Por el mismo razonamiento, encontraremos una suma menor que esta última, y podemos proseguir ad infinitum hallando números enteros cuadrados perfectos cada vez menores con la misma propiedad. Sin embargo, esto es imposible ya que no puede existir una serie infinita de números enteros menores que cualquier entero dado que escogemos. ———————————————
>
> Este margen es demasiado pequeño para permitirme dar la demostración completa y en todo detalle."

De esta manera, charlada y resumida, expuso Fermat su famoso método de descenso infinito. El método de descenso infinito es una particularización de la demostración por contradicción. Es decir, en cuanto a su estructura lógica es apenas una versión del método de contradicción, pero en cuanto a la teoría de números es un método propio que depende de una de las propiedades de los nńeros enteros positivos, de la misma manera que la inducción es un método propio relacionado con estos números.

Para poder entender el método de descenso infinito y usarlo en el futuro, debemos, en primer lugar, reconstruir el argumento charlado por Fermat en términos mas formales y completos.

Las longitudes de los lados de un triángulo rectángulo racional pueden expresarse en la forma

$$x^2 + y^2, \quad x^2 - y^2, \quad 2xy, \quad x, y \in \mathbf{Z}.$$

Dado que un factor común en los lados aparecería como un cuadrado perfecto en el área, podemos suponer que $x^2 - y^2$ y, por lo tanto, $x + y$ y $x - y$ son números impares. Además, esto nos permite suponer que x, y son primos relativos y, por ende, que x,y,$x+y$, y $x-y$ son todos primos relativos tomados dos a dos.

Si el área del triángulo fuera un cuadrado, tendríamos que $xy(x-y)(x+y)$ sería un cuadrado. Ya que los factores son primos relativos, cada uno de estos factores por separado debe ser un cuadrado perfecto. Entonces, tenemos

$$x = u^2; \quad y = v^2; \quad u^2 + v^2 = p^2; \quad u^2 - v^2 = q^2.$$

De las últimas dos de estas ecuaciones, obtenemos que

$$2v^2 = p^2 - q^2 = (p + q)(p - q). \tag{3.18}$$

Ahora bien, p y q son ambos pares porque p^2 y q^2 han de ser ambos pares según nuestras condiciones iniciales. Sin embargo, no pueden tener otro factor en común diferente de 2 porque u^2, v^2 son primos relativos.

Se sigue de (1) que las únicas posibilidades para $p + q$ y $p - q$ son

$$p + q = 2m^2n^2, \qquad p - q = n^2 2m^2,$$

donde n es un número par.

De aquí se obtiene

$$u^2 = \frac{p^2 + q^2}{2} = (m^2)^2 + \left(\frac{n}{2}\right)^2,$$

de donde, los números m^2, $n^2/2$ son los lados de un nuevo triángulo rectángulo de área cuadrada, a saber, $\frac{m^2n^2}{4}$.

Observamos que los lados de este nuevo triángulo son menores que los del triángulo original ya que el cuadrado de su hipotenusa, u^2 o x, es un factor de uno de los catetos del triángulo original.

Ahora bien, de esta manera, estamos en las condiciones iniciales del problema y podemos aplicar el mismo razonamiento para obtener un tercer triángulo con lados menores que el segundo y que cumple las mismas

condiciones. Pero, dado que no es posible tener una sucesión decreciente infinita de números enteros positivos, se sigue que, a partir de suponer la existencia de un tal triángulo, llegamos a una contradicción. Luego, el área de un triángulo rectángulo con lados racionales no puede ser un cuadrado perfecto.

Punto de discusión

En sus investigaciones Fermat enunció el siguiente teorema, conocido con el nombre del "ultimo teorema de Fermat":

La ecuación $x^n + y^n = z^n$ no tiene solución en enteros para $ngeq3$.

Fermat mencionó que tenía una ingeniosa demostración para el teorema pero que 'no cabía en este margen' (del libro mencionado de Bachet donde anotaba sus resultados). No resultó fácil para otros matemáticos producir dicha demostración. En consecuencia, el teorema ha sido objeto de estudio durante los últimos tres siglos esperando demostración. Al fin en 1993, un joven matemático inglés, Andrew Wiles, hizo pública una ingeniosa demostración que recopilaba las investigaciones recientes de muchos importantes matemáticos. Sin embargo, se descubrió que la demostración de Wiles es incompleta; un gran número de investigadores trabajan en su compleción.

Por supuesto, los matemáticos suponen que la demostración que Fermat creó tener no pudo ser completa. No obstante, una demostración parcial del caso $n = 4$ sigue directamente de la demostración del teorema sobre triángulos rectángulos que acabamos de estudiar.

Pues, en ella se pone

$$u^2 + v^2 = p^2; \quad u^2 - v^2 = q^2,$$

de donde,

$$(u^2 + v^2)(u^2 - v^2) = u^4 - v^4 = (pq)^2.$$

Ahora bien, acabamos de demostrar que conlleva contradicción suponer que existen dos potencias cuartas cuya diferencia es un cuadrado. De allí se sigue que no es posible resolver en enteros la ecuación

$$x^4 + y^4 = z^4.$$

¿Por qué ? Discutir.

Problemas propuestos

1. Usar el método de descenso infinito de Fermat para demostrar que el sistema de ecuaciones

$$x^2 + y^2 = z^2$$
$$y^2 + z^2 = t^2$$

no tiene solución en los enteros.

2. Si no lo ha hecho aún, terminar la demostración de que no existen enteros todos diferentes de cero tales que

$$x^4 - y^4 = z^4.$$

3.8.1 El método de descenso infinito y un problema de representación cuadrática

No es difícil darse cuenta que en todos los casos anteriores, el método de descenso infinito ha sido empleado para demostrar resultados 'negativos', es decir, para demostrar que una cierta ecuación no tiene solución en los enteros. De hecho, Fermat trabajó mucho para llegar a usar el método de descenso infinito para demostrar resultados positivos. El mismo describe sus experiences en las siguientes palabras.

> "Por mucho tiempo no fui capaz de aplicar mi método a proposiciones afirmativas porque el truco y el giro que se requieren para encararlos es mucho mas difícil que aquél que se usa para proposiciones negativas. Así que, cuando tuve que demostrar que todo primo que excede en 1 un múltiplo de 4 está compuesto por dos cuadrados , me encontré en una tormenta. Pero al fin una meditación repetida muchas veces me dio las luces que me hacían falta, y ahora proposiciones afirmativas se someten a mi método, con la ayuda de nuevos principios que necesariamente se le deben adjuntar. El curso de mi modo de razonar en proposiciones afirmativas es como sigue: si un número primo de la forma $4n + 1$ arbitrariamente escogido no es la suma de dos cuadrados, [demuestro que] habrá otro de la misma naturaleza, menor que el escogido, y por tanto un tercer aun menor, y así sucesivamente. Haciendo de esta manera un descenso infinito finalmente llegamos al número 5, el menor de todos los números de este tipo $[4n + 1]$. Por la demostración mencionada y el argumento que ella contiene, se sigue que 5 no es la suma de dos cuadrados. Pero sí lo es. Por consiguiente, debemos inferir por reductio ad absurdum , que todos los números de la forma $4n+1$ pueden expresarse como la suma de dos cuadrados."

3.8.2 Demostración por contraejemplo mínimo

Es bien conocido que el álgebra abstracta, y en particular la teoría de grupos, debe muchos de sus ejemplos mas importantes a la teoría de números y que, argumentos basados en teoremas básicos de la teoría de números aparecen constantemente en la demostración de los teoremas de la teoría de grupos finitos. Al tratar de concretar estas relaciones y estudiarlas a fondo en un contexto histórico, nos preguntamos por el método de descenso infinito puesto que aparentemente se ha arrinconado en la teoría de números sin dejar rastro alguno en posteriores desarrollos del álgebra abstracta.

Como el título de esta sección sugiere, se quiere mostrar aquí que en efecto el método de Fermat está en el fondo del modo de pensar del álgebra abstracta donde aparece como el argumento del contraejemplo mínimo.

Examinemos ahora una demostración proveniente de la teoría de grupos y referente al tema desarrollado en la Sección 6.6 de este texto. El lector debe regresar al presente aparte cuando esté estudiando dicha sección para comprender los pormenores de esta exposición. Por el momento basta observar los paralelos entre el razonamiento de la demostración y el método de descenso infinito de Fermat.

El teorema que nos interesa tiene el siguiente enunciado.

Sea G un grupo abeliano finito. Si p es un primo y $p \mid o(G)$, entonces G contiene un elemento de orden p.

Se describe el procedimiento de la demostración en estas palabras, similares a la descripción que Fermat nos dio de su método.

Usamos una técnica de demostración que es una aplicación del principio de la buena ordenación y que es particularmente aplicable a grupos finitos. El argumento procede como sigue: Si el teorema en cuestión no se cumple, entonces el conjunto de grupos para los cuales es falso es no vacío; y de allí el conjunto de los órdenes de estos grupos es un conjunto no vacío de enteros positivos. Sea n el menor de tales enteros y sea G uno de los grupos de orden n para los cuales el teorema es falso. G es, por lo tanto, un contraejemplo mínimo. La demostración procede mostrando que tal contraejemplo mínimo no existe.

Demostración. Sea G un contraejemplo mínimo. Es decir, $p \mid o(G)$ para algún primo p, pero G no contiene ningún elemento de orden p. Además, si G' es cualquier otro grupo con estas propiedades, se tiene que $o(G) \leq o(G')$.

Sea $g \in G, g \neq e$. Si $o(g) = pm$, entonces $o(g^m) = p$, que constituye una contradicción de la suposición de que G no contiene ningún elemento de orden p. Entonces se puede suponer que el orden de todo elemento de G es primo relativo con p.

Entonces para cualquier $g \neq e \in G$, $\langle g \rangle$ es subgrupo propio de G, ya que,

de lo contrario, G sería cíclico y $o(G) = o(g)$. Pero $p \mid o(G)$ y p no divide a $o(g)$. Se sigue que G no es cíclico.

Dado que G es abeliano, $\langle g \rangle \triangleleft G$. Consideremos ahora el grupo cociente $G/\langle g \rangle$. $\langle e \rangle \subset \langle g \rangle$ y $o(g) > 1$,de donde, $o\left(G/\langle g \rangle\right) < o(G)$. Y como $o\left(G/\langle g \rangle\right) o(g) = o(G)$, se sigue que $p \mid o\left(G/\langle g \rangle\right)$. Ya que G es un contraejemplo mínimo, tenemos que $G/\langle g \rangle$ contiene un elemento que notaremos $h\langle g \rangle$ de orden p. Pero el orden de una clase lateral $(h\langle g \rangle)$ divide el orden de su representante h en G. Se sigue que $p \mid o(h)$, lo cual implica que G tiene un elemento de orden p, una contradicción. La contradicción muestra que no hay ningún contraejemplo, o sea, que el teorema se cumple.

3.9 Problemas del capítulo

1. Si 2 es una solución de la ecuación $x^3 + hx + 10 = 0$ hallar h.

2. Sea $f(x)$ un polinomio cúbico. Demostrar que r es raíz de $f(x) = 0$ si y sólo si $f(x) = (x - r)q(x)$ para algún polinomio cuadrático $q(x)$.

3. Determinar, por simple inspección, una raíz de la ecuación polinómica $x^3 - 4t + 3 = 0$ y usar esta información para obtener todas las raíces.

4. Resolver la ecuación $(x - 2)(x - 3)(x - 4)(x - 5) = 360$.

5. Si

$$\begin{aligned} x + y + z &= 4 \\ xy + yz + zx &= 6 \\ xyz &= 3 \end{aligned}$$

hallar los valores de $\dfrac{1}{x} + \dfrac{1}{y} + \dfrac{1}{z}$ y $x^2 + y^2 + x^2$ sin determinar x, y, y z.

6. Si $x + y = 1$ y $x^3 + y^3 = 19$, hallar $x^2 + y^2$.

7. Sean p, q y r las raíces de la ecuación $x^3 + ax^2 + bx + c = 0$. Expresar $p^2q^2 + q^2r^2 + r^2p^2$ en términos de a, b y c. Resolver el sistema de ecuaciones

$$\begin{aligned} p + q + r &= \frac{7}{2} \\ pq + qr + rp &= -\frac{5}{2} \\ pqr &= -2. \end{aligned}$$

8. Sean p y q dos de las raíces de $x^3 - x + 1 = 0$. Demostrar que pq es raíz de $x^3 + x^2 - 1 = 0$.

9. Si $a \neq b$, $a^3 - b^3 = 19x^3$ y $a - b = x$, hallar a en términos de x.

10. Si p, q y r son raíces de $x^3 - x^2 + x - 2 = 0$, hallar el valor (numérico) de $p^3 + q^3 + r^3$.

11. Hallar la ecuación cúbica cuyas raíces son los cubos de las raíces de la ecuación
$$x^3 + ax^2 + bx + c = 0.$$

12. Consideremos el conjunto de todas las ecuaciones de la forma
$$x^3 + a_2x^2 + a_1x + a_0 = 0,$$
donde los coeficientes a_2, a_1, a_0 son reales y $\mid a_i \mid \leq 2$ para $i = 0, 1, 2$. Sea r el mayor número real positivo que satisface alguna de estas ecuaciones. ¿En cuál de los intervalos $[1, \frac{3}{2}), [\frac{3}{2}, 2), [2, \frac{5}{2}), [\frac{5}{2}, 3), [3, \frac{7}{2})$ se encuentra r?

13. Hallar una relación general entre a y b tal que la ecuación $x^3 + ax + b = 0$ puede escribirse en la forma $x^4 = (x^2 + cx + d)^2$ y usarla para resolver $8x^3 - 36x + 27 = 0$.

14. Supongamos que $x^3 + px + q = 0$ tiene una raíz no real $a + bi$ y a, b, p, q son todos reales con $q \neq 0$. Demostrar que $aq > 0$.

15. Si a, b, c, d son números reales, demostrar que cada uno de los sistemas de tres ecuaciones ((3.19) y (3.20)) es equivalente al otro.
$$a^2 + b^2 = 2; \qquad c^2 + d^2 = 2; \qquad ac = bd; \quad (3.19)$$
$$a^2 + c^2 = 2 \qquad b^2 + d^2 = 2; \qquad ab = cd. \quad (3.20)$$

16. Hallar una expresión sencilla correspondiente a una raíz de $x^3 - 3x^2 - x - \sqrt{2} = 0$.

17. Si
$$u = 1 + \frac{x^3}{3!} + \frac{x^6}{6!} + \ldots,$$
$$v = \frac{x}{1!} + \frac{x^4}{4!} + \frac{x^7}{7!} \ldots,$$
$$w = \frac{x^2}{2!} + \frac{x^5}{5!} + \frac{x^8}{8!} \ldots,$$
demostrar que $u^3 + v^3 + w^3 - 3uvw = 1$.

18. Si $ax^3 + bx + c$, donde $a, c \neq 0$ tiene un factor de la forma $x^2 + px + 1$ demostrar que $a^2 - c^2 = ab$.

19. Hallar todos los valores de m para los cuales la ecuación

$$x^4 - (3m + 2)x^2 + m^2 = 0$$

tiene cuatro raíces reales en progresión aritmética.

20. Resolver en enteros el sistema

$$\begin{aligned} a^3 - b^3 - c^3 &= 3abc \\ a^2 &= 2(b + c). \end{aligned}$$

21. Resolver en enteros

$$x_1^4 + x_2^4 + \cdots + x_{14}^4 = 1599.$$

22. Demostrar que la única solución en enteros de la ecuación

$$x^2 + y^2 + z^2 = 2xyz$$

es $x = y = z = 0$.

23. Hallar enteros x, y, z, w tales que $x^2 + y^2 + z^2 + w^2 = 2xyzw$.

24. Resolver en enteros (tanto positivos como negativos)

$$\text{(a) } \frac{1}{x} + \frac{1}{y} = \frac{1}{14}; \qquad \text{(b) } \frac{1}{x} + \frac{1}{y} = \frac{1}{z}.$$

25. Sean a y b enteros positivos tales que $ab = 1$ divide a a^2+b^2. Demostrar que $\frac{a^2+b^2}{ab+1}$ es el cuadrado de un entero.

26. Demostrar que la ecuación $x^2 + y^2 = z^5 + z$ tiene infinitas soluciones que son primos relativos.

Chapter 4

Teoría de ecuaciones

4.1 La generación de nuevos métodos y la búsqueda de resultados más generales

Ya que hemos tratado la solución de ecuaciones polinómicas de grado menor que cinco, es el momento de desarrollar una teoría de ecuaciones lo mas general posible. La historia nos ha indicado que se produce un cambio fundamental de orientación justo en el momento en que Viète introduce la costumbre de representar no sólo las cantidades variables o incógnitas sino también las cantidades constantes o dadas por medio de letras. No cabe duda que el paso tomado por Viète marca un hito en la teoría de ecuaciones por muchas razones de central importancia. Veamos algunas de ellas.

4.1.1 Consideraciones acerca de las relaciones de Viète

Un paso hacia la generalidad. La geometría, desde la obra de Euclides escrita alrededor de 300 a.C., había logrado la generalidad en sus demostraciones. Uno de los mecanismos usados por Euclides fue la construcción de un sistema lógico-deductivo precursor de los sistemas modernos. Los ingredientes principales son definiciones, axiomas, postulados y demostraciones. También en el interior de las demostraciones se logró la generalidad dando por sentado que, aunque apareciera una figura y se hiciera referencia a ella, el argumento presentado era aplicable a cualesquiera dos puntos, cualquier triángulo, cualquier círculo, etc.

No se tenía un mecanismo similar en la teoría de ecuaciones. No se contó con ningún sistema axiomático relacionado con el álgebra hasta el siglo XIX. Pero el trabajo de Viète representó un primer paso hacia la generalidad análoga a la generalizabilidad de las figuras geométricas. Antes de Viète, se

presentaba un método de solución mostrando como éste podía usarse para resolver una ecuación específica dada. A veces los coeficientes de la ecuación indicaba que eran arbitrarios en el sentido en qué sus valores específicos no jugaban ningún papel que hiciera funcionar el método. Si volvemos a nuestros ejemplos chinos y en particular al sistema de ecuaciones lineales cuya solución se expuso, observamos que los coeficientes son 1, 2 y 3 en varias combinaciones. Esto es un indicio de su arbitrariedad. A veces el método iba acompañado por comentarios que indicaban a cuáles otras ecuaciones era aplicable.

Sin la notación introducida por Viète no hay plena generalidad ni es posible pensar en una axiomatización del álgebra.

El problema de la homogeneidad. Recordemos que el problema de la homogeneidad consiste en la observación de que, en cuanto a su interpretación geométrica, términos como x, x^2, x^3 representan entes de distinta dimensión. Siguiendo el modelo geométrico del álgebra generado por los griegos es inconcebible escribir una expresión como $x^3 + 4x^2 + 2x + 7$, pues aquí estaríamos sumando un volumen con una área, una longitud, etc. La representación de las raíces de una ecuación por constantes literales y el desarrollo de la misma a partir de sus factores, rescata la noción de homogeneidad, pues

$$(x-a)(x-b)(x-c) = (x^2-(a+b)x+ab)(x-c) = x^3-(a+b+c)x^2+(ab+ac+bc)x-abc$$

muestra que cada uno de los términos de la ecuación representa un volumen y, por lo tanto, la suma tiene sentido geométrico. Esto es importante precisamente porque poco después de que se conocieran las ideas de Viète, Descartes se encargaría de contradecirlas. Las razones que tuvo para hacerlo son interesantes. Primero, como hemos comentado, aunque su notación produjo un medio de lograr generalidad, el mismo Viète sólo introdujo su estudio de la relación entre las raíces y los coeficientes en el caso de una ecuación de tercer grado que tiene todas sus tres raíces reales y positivas. No contempló la posibilidad de aceptar raíces negativas o imaginarias como soluciones genuinas a las ecuaciones y su prejuicio geométrico muestra claramente por qué. Pero Cardano ya había resuelto el problema de encontrar dos números cuya suma sea 10 y cuyo producto sea 40 y esto le había conducido a una solución en complejos, pues la correspondiente ecuación sería $x^2 - 10x + 40 = 0$, solución es $\left\{x = 5 + i\sqrt{15}\right\}, \left\{x = 5 - i\sqrt{15}\right\}$. Es decir, ya se tenía evidencia de que se cumplieran relaciones similares entre raíces y coeficientes aunque aquellas no sean reales y positivas.

Para Descartes, las relaciones de Viète, ligadas íntimamente a nociones geométricas, contenían prohibiciones artificiales. El meollo del problema, tanto la exigencia de homogeneidad, como las restricciones sobre las raíces, como las restricciones sobre la 'dimensión' es la interpretación geométrica.

En su desarrollo de la geometría analítica, Descartes representa geométricamente la división de un segmento por otro, destruyendo así toda justificación de la necesidad de trabajar con expresiones homogéneas.

Punto de investigación

Investigar sobre la posibilidad de representar geométricamente la suma, la diferencia, el producto y el cociente de dos números que representan longitudes de segmentos.

Así mismo la posición de Descartes lo libera de las ataduras mentales que supeditan el álgebra a la geometría y le permiten enunciar proposiciones realmente generales como lo es el Teorema Fundamental del Algebra que afirma que una ecuación polinómica de grado n tiene exactamente n raíces. Por otra parte, de esta manera se abre el camino para que se considere la posibilidad de estudiar geometrías en espacios de n dimensiones, estudio que por primera vez se realiza desde la perspectiva del álgebra en la obra de Hermann Grassman hacia 1840.

El método de factorización. Las relaciones de Viète dejan en claro la posibilidad de explorar el método de factorización para la solución de ecuaciones polinómicas. Hemos trazado la historia de la generación de fórmulas para la solucion de ecuaciones polinómicas de grado 2, 3 y 4. Notamos que, aunque desde los babilonios se tenía en claro que el problema de encontrar dos números cuya suma y producto son conocidos conlleva la solución de una ecuación cuadrática, no hay evidencia del uso del método de factorización en factores lineales antes de la obra de Harriot (1631). Es claro que la generalidad de este método sólo es concebible sobre la base de las relaciones de Viète.

4.1.2 Pierre de Fermat, René Descartes y sus contemporáneos

Un importante paso en la nueva exploración del significado del álgebra y un proceso indispensable en la eventual transformación de la misma, es evidentemente la creación de la geometría analítica que produce un vuelco en la apreciación de la forma en que se relacionan la geometría y el álgebra.

Obra casi simultánea de Fermat y Descartes, por su notación más adecuada y su publicidad más agresiva, hoy día usamos el enfoque cartesiano y casi ni nos acordamos de la contribución de Fermat. Sin embargo, lo que nos interesa comentar aquí es que el método de la geometría analítica o de coordenadas deja completamente clara la misma unidad notada por Viète cuando introdujo coeficientes literales, pero ahora a nivel metodológico. Descartes hace particular énfasis en el hecho de que su enfoque proporciona un método que resuelve todos los problemas formulables, mientras que los geómetras

sintéticos debían buscar diferentes enfoques para diferentes problemas y a veces contentarse con la imposibilidad de solución de una cuestión dada. Descartes comenta en el *Discurso del método* que

> "He dado estos (métodos) muy sencillos para mostrar que es posible construir todos los problemas de la geometría ordinaria haciendo no más de lo poco cubierto por las cuatro figuras que he explicado. Esta es una cosa que creo los matemáticos antiguos no observaron. pues de otro modo no hubieran puesto tanto esfuerzo en escribir tantos libros en los cuales la misma secuencia de las proposiciones muestra que no poseían un método seguro para hallar todo, sino que recogían aquellas proposiciones que habían encontrado por accidente."

Se pone sobre la mesa histórica, entonces, la apreciación de que el tratamiento netamente algebraico permite una generalidad metodológica en la solución de ecuaciones y otros problemas. que la geometría es incapaz de alcanzar.

Por otra parte, en este contexto regresamos ~~por primera vez~~ a la observación tan genial de Leonardo de Pisa, pues combinado con el desarrollo del concepto de función y la representación de las funciones por intermedio de curvas, característicos del cálculo basado en la geometría analítica de Descartes, este enfoque permite estudiar las funciones polinómicas e identificar el problema de la solución de una ecuación polinómica con el de hallar los ceros de la función. Esta relación fue intuida por Leonardo cuando concluyó que la ecuación que estaba estudiando tenía una raíz entre 1 y 2. Pero con la geometría analítica y el cálculo. se instituye un análisis del crecimiento, decrecimiento, máximos y mínimos de las funciones (incluidas las polinómicas aunque no exclusivamente referido a ellas) que enriquece su comprensión y conlleva nuevos enfoques para aproximar las raíces de las respectivas ecuaciones independientemente de poseer o no una fórmula de solución.

Por otra parte, uno de los tabúes del clásico marco teórico que fundamenta los números y el álgebra en la geometría, que es superada por la geometría analítica, es la del grado de una ecuación (aunque ésta fue parcialmente superada ya por los italianos que resolvieron las ecuaciones de cuarto grado sin preocuparse por la interpretación geométrica de un producto con cuatro factores.). Otro es evidentemente el estatus privilegiado de que gozaban la recta y el círculo (la regla y el compás). En la geometría analítica no hay curvas privilegiadas.

Es evidente. de las bien conocidas reglas de Descartes para acotar el número de raíces reales positivas o negativas de una ecuación polinómica, que para Descartes la discusión de Cardano sobre lo ficticio de los números negativos ha sido resuelto a favor de una aceptación de ellos, aunque los sigue

denominando falsos. Se cuenta que Descartes reconoció la posibilidad, dada una ecuación polinómica, de transformarla en una ecuación cuyas raíces han sido aumentadas en cualquier cantidad (correspondiente a la traslación de la gráfica). Esto permite relacionar números negativos con positivos de una manera que resulta muy aprovechable y que se encontraría pocos años mas tarde en las ideas del inglés John Wallis.

En esta misma línea de pensamiento, el contemporáneo de Descartes, Albert Girard, siguiendo unas observaciones de Cardano concernientes a las ecuaciones cúbicas y cuárticas, afirma que una ecuación polinómica de grado n tiene n raíces incluyendo las absurdas (complejas) y las repetidas. Esta es tal vez la primera vez que se enuncia lo que hoy día se conoce como el Teorema Fundamental del Algebra. Dado el poco entendimiento que se tenía entonces de los números negativos y complejos, vemos que se trata de introducir elementos ideales en la teoría que permiten que ésta goce de mayor simetría, generalidad y simplicidad.

Un comentario que viene al caso es que con pensadores como Viète y Descartes se efectúa un cambio en el sistema de clasificación de ecuaciones que refleja el cambio de perspectiva que se estaba dando. Las ecuaciones ahora se clasifican según su grado, cuando con anterioridad (hasta la obra de Pacioli) eran clasificadas según el número de términos que contenían (monomios, binomios, etc.). Descartes es el primero en hablar claramente de lo que él denomina la "dimensión" de una ecuación en su *Géométrie* de 1637. Debemos tener presente que el mismo lenguaje que usamos refleja los énfasis que estamos haciendo o la estructura que estamos empleando. En particular, debemos cuestionar cualquier residuo del lenguaje anterior que perdura en la matemática escolar y determinar si se justifica su preservación; así mismo averiguar si las mezclas de lenguaje tienen un efecto adverso sobre la comprensión y dominio del tema que nuestros alumnos logran construir. Por último, el estudio de las funciones polinómicas en sí como entes u objetos matemáticos llevará eventualmente, en el álgebra abstracta, al estudio del anillo de los polinomios. Veremos, entonces, el desenlace de estas contribuciones históricas.

4.2 La ecuación polinómica general

Teniendo en cuenta estos acontecimientos, no es sorprendente, por lo tanto, que las relaciones de Viète se generalizan poco después de ser enunciadas a ecuaciones polinómicas de cualquier grado y sin restricciones sobre las raíces, sobre la base del Teorema Fundamental del Algebra.

Sean $r_1, r_2, \ldots, r_n$ las raíces de la ecuación polinómica

$$a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0 = 0.$$

Entonces

$$-(r_1 + r_2 + \cdots + r_n) = \frac{a_{n-1}}{a_n}$$

$$r_1 r_2 + r_1 r_3 + \cdots + r_{n-1} r_n = \sum r_i r_j = \frac{a_{n-2}}{a_n}$$

$$r_1 r_2 r_3 + \cdots + r_{n-2} r_{n-1} r_n = \sum r_i r_j r_k = \frac{a_{n-3}}{a_n}$$

$$\cdots$$

$$r_1 r_2 \cdots r_n = (-1)^n \frac{a_0}{a_n}.$$

En palabras, el coeficiente de x^k está dado por la suma de los productos de las raíces, tomados de a $n-k$. Es decir, el miembro izquierdo de cada una de estas expresiones es una *función simétrica* de las raíces de la ecuación.

Definición 1 *Una función $f(x_1, x_2, \ldots, x_n)$ se llama simétrica si permanece invariante bajo cualquier permutación de $x_1, x_2, \ldots, x_n$.*

Luego es claro que, si $a_n = 1$, los coeficientes son funciones tanto simétricas como homogéneas de las raíces. Cuando $a_n = 1$ se dice que la ecuación es *mónica*; es claro que toda ecuación polinómica puede transformarse en una ecuación mónica de modo que, en lo que sigue, restringiremos nuestra atención a las ecuaciones mónicas.

Sin cuidar todos los detalles, es posible ofrecer una demostración por inducción de estas relaciones como sigue. Es claro que es cierto para los primeros casos, $n = 2, 3$. Supongamos que es cierto para $n = t$ y consideramos la ecuación

$$(x - r_{t+1})(x^t + a_{t-1}x^{t-1} + \cdots + a_1 x + a_0) =$$

$$x^{t+1} + (a_{t-1} + r_{t+1})x^t + (a_{t-2} + r_{t+1}(a_{t-1}))x^{t-1} + \cdots + r_{t+1}a_0$$

Lo anterior produce precisamente las relaciones requeridas para el caso $n = t + 1$, completando nuestra demostración.

Ahora bien, este resultado, que en principio puede ser utilizado para resolver ecuaciones de cualquier grado por factorización, no tiene, como podría pensarse, una aplicación directa y general. Aparentemente las relaciones de Viète nos proporcionan n ecuaciones en n variables $r_1, r_2, \ldots, r_n$. Parece que así se determinarían las raíces de manera única. Pero el proceso es circular como se aprecia en el siguiente ejemplo correspondiente al caso de la ecuación cúbica. Consideremos la ecuación

$$x^3 + a_2x^2 + a_1x + a_0 = 0.$$

Sean r_1, r_2, r_3 sus raíces. Las relaciones de Viète nos dice que

$$r_1 + r_2 + r_3 = -a_2 \tag{4.1}$$

$$r_1r_2 + r_1r_3 + r_2r_3 = a_1 \tag{4.2}$$

$$r_1r_2r_3 = -a_0. \tag{4.3}$$

Intentemos despejar r_1. Tenemos de (4.3) que $r_1 = \dfrac{-a_0}{r_2r_3}$. De (4.2) se sigue que

$$r_1 = \frac{-a_0}{a_1 - (r_1r_2 + r_1r_3)} = \frac{-a_0}{a_1 - r_1(r_2 + r_3)}.$$

Usando (4.1), obtenemos $r_1 = \dfrac{-a_0}{a_1 - r_1(-a_2 - r_1)}$ que es equivalente a

$$r_1^3 + a_2r_1^2 + a_1r_1 + a_0 = 0.$$

Es claro que no hemos hecho mas que caminar en un círculo, pues tenemos la ecuación original con la raíz r_1 tomando el lugar de la variable x. Esto significa que las relaciones de Viète no son suficientes para resolver ecuaciones polinómicas. Sin embargo, estas relaciones pueden usarse conjuntamente con otros datos proporcionados. Esto lo estudiaremos en la siguiente sección.

4.2.1 Problemas cuya solución está basada en las relaciones de Viète

Si se conoce algún dato adicional sobre las raíces, además de la relación entre raíces y coeficientes, a veces sí es factible usar las relaciones de Viète para resolver una ecuación polinómica. Esto ha dado lugar a toda suerte de problemas exóticos que tienen interés histórico para nosotros, pues muestran

la exploración del alcance que las relaciones de Viète puedan tener para la solución de ecuaciones. Veamos algunos ejemplos.

1. Resolver la ecuación $4x^3 - 24x^2 + 23x + 18 = 0$ si se sabe que sus raíces están en progresión aritmética (Hall and Knight, 1888). Sean $a - b, a, a + b$ las tres raíces. Entonces,

$$(a - b) + a + (a + b) = 3a = \frac{24}{4} = 6 \Longrightarrow a = 2.$$

Por otra parte,

$$(a-b)a+(a-b)(a+b)+a(a+b) = 3a^2-b^2 = \frac{23}{4} \Longrightarrow b^2 = \frac{48-23}{4} = \frac{25}{4} \Longrightarrow b = \pm\frac{5}{2}.$$

Es fácil ver que estos valores satisfacen también la tercera relación de Viète,

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

de donde se puede concluir que las tres raíces son $-\frac{1}{2}, 2, \frac{9}{2}$.

2. El anterior problema puede invertirse como sigue. Consideremos la ecuación $x^3 + px^2 + qx + r = 0$. Encontrar condiciones necesarias y suficientes sobre p, q y r para que las raíces de la ecuación estén en progresión aritmética.

Para resolverlo, sean (como en el problema #1), $a - b, a, a + b$ las tres raíces. Entonces, la suma es igual a $3a$. Se sigue que $3a = -p$ o $a = -p/3$. Es decir, una de las raíces es $-p/3$. (¿Por qué?) Sustituyendo este valor en la ecuación original produce

$$\left(-\frac{p}{3}\right)^3 + p\left(\frac{p^2}{9}\right) + q\left(-\frac{p}{3}\right) + r = 0 \Longrightarrow 2p^3 + 27r = 9pq,$$

que es la condición buscada.

3. Resolver la ecuación $24x^3 - 14x^2 - 63x + 45 = 0$ si se sabe que una de las raíces es el duplo de otra. Sean $a, 2a, b$ las raíces de la ecuación. Entonces,

$$3a + b = \frac{7}{12}; 2a^2 + 3ab = -\frac{21}{8}; 2a^2b = -\frac{15}{8}.$$

Se puede despejar b en la primera ecuación y sustituir este valor en la segunda, obteniendo la ecuación cuadrática en $a, 8a^2 - 2a - 3 = 0$. Esta ecuación produce dos valores para a, a cada uno de los cuales le corresponde un valor de b. Sólo un par de valores satisface también la tercera ecuación. ¿Cuáles son estos valores? ¿Cuáles son las raíces de la ecuación dada?

4. Hallar una condición necesaria sobre a, b y c para que las raíces de la ecuación $x^3 + ax^2 + bx + c = 0$ estén en progresión geométrica. De allí idear una ecuación cúbica (con coeficientes enteros) cuyas raíces estén en progresión geométrica y resolver dicha ecuación. ¿Es la condición que Ud. generó también suficiente? ¿Por qué?

4.2.2 ¿Los problemas para qué?: un primer problema para escudriñar.

En cualquier libro tradicional de álgebra en la sección de teoría de ecuaciones encontramos problemas como éste.

Dado $f(x) = a_n x^n + a_{n-a} x^{n-1} + ... + a_0 = 0$ con raíces $r_1, r_2, ..., r_n$. Hallar $F(x) = 0$ tal que cada una de sus raíces es menor que una raíz de $f(x) = 0$ en una cantidad dada h.

Para resolverlo escribimos

$$F(x) = A_n y^n + A_{n-1} y^{n-1} + ... + A_0 = A_n (x-h)^n + A_{n-1} (x-h)^{n-1} + ... + A_0.$$

Ahora, la expresión a la derecha es igual a 0 para $x = r_i$ $(i = 1, 2, ..., n)$, de donde,

$$A_n (x-h)^n + A_{n-1} (x-h)^{n-1} + ... + A_0 = a_n x^n + a_{n-a} x^{n-1} + ... + a_0.$$

La forma mas fácil de determinar los A_j es por división. Si se dividen ambos miembros de la ecuación por $(x-h)$, el lado izquierdo nos proporciona un cociente de

$$A_n (x-h)^{n-1} + ... + A_1$$

y un residuo de A_0. Por lo tanto, si se divide el miembro derecho por $(x-h)$ se obtendrá el mismo cociente y el mismo residuo. En seguida se divide

$$A_n (x-h)^{n-1} + ... + A_1$$

por $(x-h)$ y obtenemos un cociente de

$$A_n (x-h)^{n-2} + ... + A_2$$

y un residuo de A_1. Continuamos dividiendo sucesivamente los cocientes hasta obtener todos los A_j. El proceso completo es fácil si usamos únicamente los coeficientes. Por ejemplo, consideremos la ecuación cúbica

$$x^3 - 14x^2 + 63x - 90 = 0$$

y reduzcamos cada raíz en 2. Escribimos el proceso completo como sigue.

$$\begin{array}{ccccc|c}
1 & -14 & 63 & -90 & & -2 \\
 & 2 & 24 & -78 & & \\
\hline
1 & -12 & 39 & -12 & & \\
 & -2 & 20 & & & \\
\hline
1 & -10 & 19 & & & \\
 & -2 & & & & \\
\hline
1 & -8 & & & &
\end{array}$$

Punto de discusión

Comparar este resultado con el de efectuar la división $x^3 - 14x^2 + 63x - 90 \div (x - 2)$.

La ecuación reducida es, por lo tanto,

$$x^3 - 8x^2 + 19x - 12 = 0.$$

Una raíz es evidentemente igual a 1, y las otras se encuentran fácilmente dividiendo por $(x - 1)$ y factorizando el cociente cuadrático. Tenemos

$$x^3 - 8x^2 + 19x - 12 = (x - 1)(x - 3)(x - 4) = 0.$$

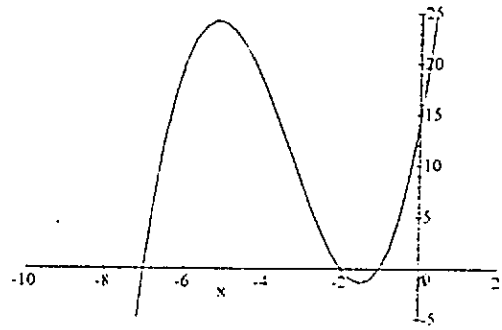
Luego, las raíces son $\{x = 1\}, \{x = 3\}, \{x = 4\}$y se sigue que las raíces de la ecuación original se obtienen sumando 2 a cada una de éstas. Son $x = 2, x = 5, x = 6$.

No cabe duda que la manipulación que hemos efectuado es tan clara como lo es abstrusa. Tenemos derecho de preguntarnos por el por qué. Si bien podemos lograr una transformación como ésta, nuestros motivos para hacerlo se encuentran totalmente obfuscados. Aquí la historia nos esclarece el panorama de manera contundente. Descartes usaba transformaciones de este tipo en sus consideraciones acerca de las soluciones negativas de ecuaciones polinómicas. Recordemos que había cierta polémica y mucho desconcierto en torno a los números negativos e imaginarios, sobre todo porque no gozaban de una interpretación geométrica clara, es decir, no pueden interpretarse como magnitudes de segmentos. Esto a su vez conlleva un cuestionamiento de su realidad, cuestionamiento que se revela en la terminología entonces común, pues recordemos que las raíces negativas se llamaban falsas o ficticias.

Una discusión similar se desarrolló en torno a los números complejos o imaginarios, nombre que sigue usándose y que señala los problemas filosóficos suscitados por ellos.

Ahora bien, Descartes considera una ecuación cuyas raíces son negativas y por medio de una transformación apropiada la asocia con una ecuación cuyas raíces son positivas. Es mas; si estudiamos las gráficas de las correspondientes funciones polinómicas, tema originado por Descartes con la geometría analítica, vemos que una es la imagen de la otra por una sencilla traslación.

Por ejemplo, consideremos la ecuación cúbica $x^3 + 10x^2 + 23x + 14 = 0$. Sus raíces son $x = -1, x = -2$ y $x = -7$ y su gráfica aparece en la figura siguiente.

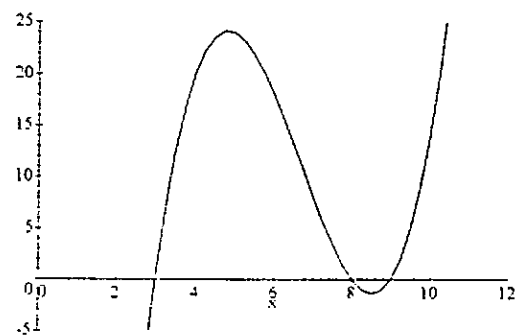


Gráfica parcial de $x^3 + 10x^2 + 23x + 14$

Ahora transformemos la ecuación en una cuyas raíces son cada una 10 mas que una raíz de la ecuación anterior.

$$\begin{array}{rrrrr} 1 & 10 & 23 & 14 & -10 \\ & -10 & 0 & -230 & \\ 1 & 0 & 23 & -216 & \\ & -10 & 100 & & \\ 1 & -10 & 123 & & \\ & -10 & & & \\ 1 & -20 & & & \end{array}$$

Obtenemos la ecuación $x^3 - 20x^2 + 123x - 216 = 0$ que se factoriza como $x^3 - 20x^2 + 123x - 216 = (x - 3)(x - 8)(x - 9) = 0$, y cuyas raíces claramente son cada una mayor en 10 que una de las raíces de la ecuación original. Ahora bien, veamos la gráfica de esta última función polinómica.



Gráfica de $f(x) = x^3 - 20x^2 + 123x - 216$

Aunque es parcial, es claro que la gráfica es la misma que la asociada con la ecuación original simplemente trasladada 10 unidades hacia la derecha.

Ahora bien, no sólo nos indica que no hay nada inherente que diferencia entre raíces negativas y positivas, sino que también aquí se indica un motivo por un cambio de énfasis. Poco después del trabajo de Descartes, lo invariante en la gráfica después de la traslación se volverá de mayor interés para la matemática que la misma solución de las ecuaciones. Posiblemente es aquí donde el concepto que llegará a denominarse función se volverá mas importante para el matemático que la misma teoría de ecuaciones. Sin embargo, Descartes estaba interesado en el análisis de las soluciones de las ecuaciones polinómicas y su lección es clara: no hay diferencia fundamental entre las soluciones positivas y las negativas.

Un segundo punto que debemos considerar relativo a nuestro problema original versa sobre el método de Cardano para la solución de la ecuación cúbica. Recordemos que éste se basa en la reducción de la cúbica general

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

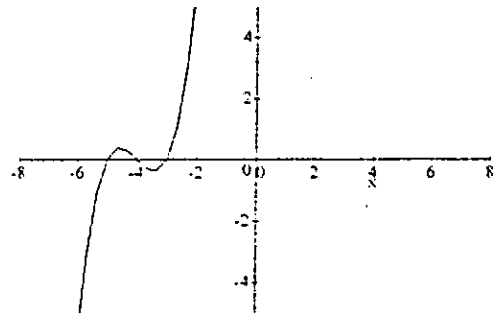
a una cúbica de la forma

$$y^3 + py + q = 0$$

por medio de una transformación similar a la usada por Descartes. En efecto, el método descansa sobre la posibilidad de efectuar una transformación para obtener una ecuación relacionada tal que la suma de sus raíces sea 0. Por ejemplo, consideremos la ecuación $x^3 + 12x^2 + 47x + 60 = 0$ cuyas raíces son $\{x = -5\}, \{x = -4\}, \{x = -3\}$. La transformación exigida para aplicar el método de Cardano consta de poner $y = x + \frac{12}{3} = x + 4$ o $x = y - 4$. Obtenemos

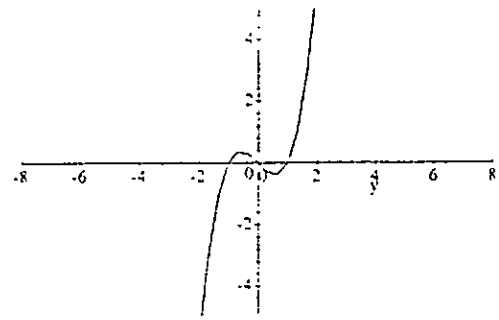
$$\begin{array}{ccccc} 1 & 12 & 47 & 60 & -4 \\ & -4 & -32 & -60 & \\ 1 & 8 & 15 & 0 & \\ & -4 & -16 & & \\ 1 & 4 & -1 & & \\ & -4 & & & \\ 1 & 0 & & & \end{array}$$

es decir, la ecuación $y^3 - y = 0$. De hecho, ésta es una ecuación supremamente sencilla de resolver y cuyas raíces son $y = -1, 0$ y 1. Es evidente que la suma de estas raíces es 0 como queríamos. Se sigue que las raíces de la ecuación original son $x = -5, -4$ o -3 y las gráficas de las funciones polinómicas correspondientes son



Gráfica parcial de $f(x) = x^3 + 12x^2 + 47x + 60$

y



Gráfica parcial de $g(y) = y^3 - y$

Así las cosas, hemos podido identificar al menos dos motivaciones fundamentales que subyacen tras un problema que superficialmente parece exigir manipulación algebraica sin sentido, rumbo, finalidad ni propósito. Por una parte, el poder llegar a una comprensión de los nexos profundos entre soluciones negativas y positivas contribuirá a la construcción de un sistema numérico que comprehende ambas y que se denominará el sistema de los números reales. Por otra parte, tal manipulación tiene una raíz teórica como procedimiento que permite generar un método de solución para la ecuación cúbica.

Anticipamos que estas consideraciones sobre la naturaleza de los números negativos producirán también un importante resultado teórico. Pues, Descartes junto con su compatriota Girard, basándose entre otras consideraciones en la identificación de raíces negativas y positivas, fue uno de los primeros matemáticos en aceptar como cierto el Teorema Fundamental del Algebra

(una ecuación polnómica de grado n tiene exactamente n raíces (contando multiplicidades)).

4.2.3 Un nuevo problema de interés.

Es también corriente encontrar en los libros tradicionales de álgebra preguntas como ésta. Dada la ecuación

$$f(x) = (x - r_1)(x - r_2) \cdots (x - r_n) = 0,$$

hallar la ecuación cuyas raíces son las de $f(x) = 0$ multiplicada cada una por una constante k, es decir, buscamos la ecuación

$$F(x) = (x - kr_1)(x - kr_2) \cdots (x - kr_n) = 0.$$

Ahora, conociendo las relaciones entre coeficientes y raíces, esta ecuación será igual a la anterior con el coeficiente de x^{n-1} multiplicado por k, el coeficiente de x^{n-2} multiplicado por k^2, y así sucesivamente hasta el término constante que será igual a el de la ecuación original multiplicado por k^n.

Punto de discusión

Justificar completamente la anterior afirmación.

Una vez mas nos preguntamos por el por qué de un problema como éste. En lo que sigue ofreceremos dos razones importantes. La primera de ellas es de tipo práctico. nos permite resolver cierta clase de ecuaciones mas fácilmente. Supongamos que queremos resolver la ecuación cuadrática

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

Multipliquemos las raíces por a. Nos resulta

$$ax^2 + abx + a^2c = 0.$$

Ahora, dividiendo por a, obtenemos una ecuación cuadrática mónica, a saber,

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

que puede ser mas fácil de factorizar. Por ejemplo, dada la ecuación

$$6x^2 - 17x + 12 = 0,$$

si multiplicamos las raíces por 6 y simplificamos, obtenemos

$$x^2 - 17x + 72 = (x - 9)(x - 8) = 0.$$

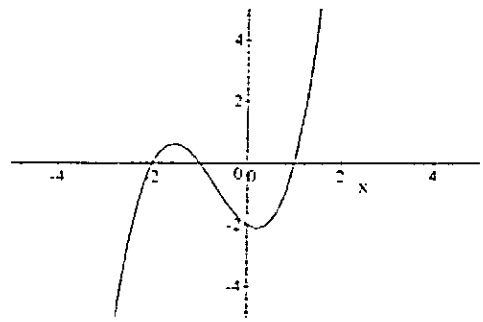
Luego, las raíces de la ecuación original son $x = \frac{9}{6}, x = \frac{8}{6}$ o $x = \frac{3}{2}, x = \frac{4}{3}$.

Primer encuentro con la regla de signos de Descartes.

Descartes produjo una regla, asociada con los cambios en los signos de los coeficientes en una ecuación polinómica, que proporciona una cota para el número de raíces reales positivas de la ecuación. La regla puede enunciarse como sigue.

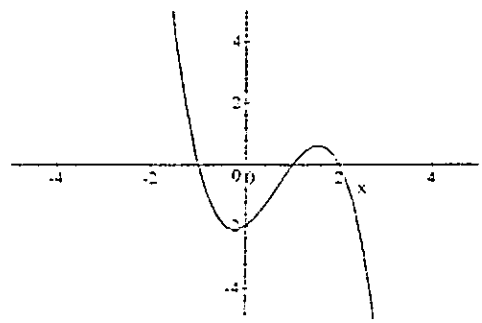
Una ecuación polinómica $f(x) = 0$ con coeficientes reales no puede tener mas raíces reales positivas que cambios de signos hay en $f(x)$ y no puede tener mas raíces reales negativas que cambios de signo hay en $f(-x)$. Mas adelante ofreceremos un enunciado distinto y mas potente de la regla de Descartes.

Por ahora, para analizar este criterio y tratar de relacionarlo con el problema original de esta sección, veamos las gráficas de las ecuaciones $f(x) = x^3 + 2x^2 - x - 2 = 0$ y $f(-x) = -x^3 + 2x^2 + x - 2 = 0$. Tenemos



Gráfica de $f(x) = x^3 + 2x^2 - x - 2$

y



Gráfica de $f(-x) = -x^3 + 2x^2 + x - 2$

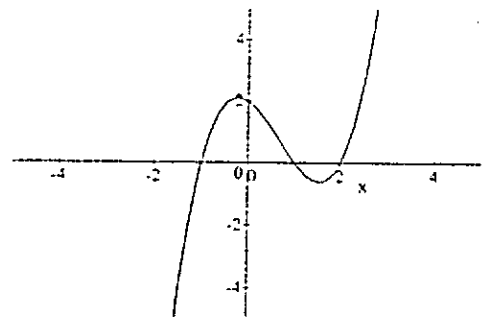
Punto de discusión

Explicar el enunciado de la regla de signos de Descartes en relación con estas dos gráficas.

Punto de investigación

Sea $g(x)$ una función polinómica. Determinar la transformación geométrica que lleva la gráfica de $g(x)$ en la de $g(-x)$.

Ahora comparemos ésta última con la ecuación cuyas raíces son las de $f(x) = 0$ multiplicada cada una por -1. Esta sería $g(x) = x^3 - 2x^2 - x + 2 = 0$. Tenemos



Gráfica de $g(x) = x^3 - 2x^2 - x + 2$

Punto de discusión

¿Puede usted identificar las relaciones entre las gráficas de las funciones $f(x), f(-x)$ y $g(x)$? ¿Entre sus raíces? Quizás es apropiado usar nociones de simetría en sus consideraciones.

Demostraremos la validez de la regla de Descartes mas adelante, por el momento consideremos un método ágil para hallar algunas de las raíces de una ecuación polinómica.

4.3 Del álgebra al cálculo

4.3.1 El Teorema del Binomio y sus extensiones

Al menos desde el tiempo de Euclides se conoce una relación que llamaremos el teorema del binomio para exponente $n = 2$. Siguiendo a Euclides, podemos interpretar esta relación geométricamente: $(a + b)^2$ representa el área de un cuadrado de lado $a + b$.

Es inmediato observar de la figura que

$$(a + b)^2 = aa + ab + ba + bb$$

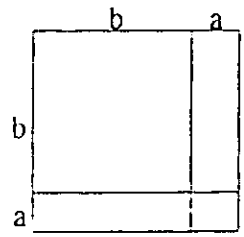


Figure 4.1:

$$\begin{aligned} &= a^2 + ab + ba + b^2 \\ &= a^2 + 2ab + b^2. \end{aligned}$$

En nuestra interpretación, todos los términos de esta expansión representan áreas y como tales se dan como producto de dos factores, "largo × ancho".

Puntos de discusión

1. Consideremos $(a+b)^3$. ¿Cuál sería su interpretación geométrica?
2. ¿Qué representa cada uno de los sumandos en su expansión?
3. ¿Puede Ud. hacer un diagrama que demuestra que $(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$?
4. Desarrollar la expresión $(a+b)^3 = (a+b)^2(a+b)$ desde el punto de vista algebraico. Expresarla como suma de ocho términos y como suma de cuatro términos.

Ahora bien, para considerar el caso general $(a+b)^n$, $n > 3$, es necesario dejar atrás la interpretación geométrica. ¿Por qué? En nuestras consideraciones generales, podemos comenzar por la cuestión de homogeneidad y ofrecer una demostración por inducción del hecho de que cada uno de los sumandos que aparece en la expansión de $(a+b)^n$ es un producto de n factores entre $a$es y $b$es.

Ya hemos considerado los primeros casos. Nuestra hipótesis de inducción es que cada uno de los sumandos que aparecen en la expansión de $(a+b)^{n-1}$ es un producto de $n-1$ factores entre $a$es y $b$es.

Ahora, por definición

$$\begin{aligned} (a+b)^n &= (a+b)(a+b)^{n-1} \\ &= a(a+b)^{n-1} + b(a+b)^{n-1}. \end{aligned}$$

Aplicando la propiedad distributiva y la hipótesis de inducción, es claro

que cada uno de los sumandos en la expansión de $(a+b)^n$ es un producto de n factores entre $a$es y $b$es.

Observemos una vez más la primera expresión que obtuvimos para $(a+b)^2$,

$$(a+b)^2 = aa + ab + ba + bb.$$

Para relacionar el teorema con nociones de combinatoria, observemos que en cada sumando tenemos dos factores y hay dos posibilidades para cada factor, pues o bien puede ser a o bien puede ser b; además estos factores han sido escogidos de todas las formas posibles (de entre dos, a y b).

Ahora bien, en la expansión de $(a+b)^n$ ya sabemos que cada sumando contiene n factores y cada factor o bien puede ser a o bien b. De allí se sigue que el número total de sumandos es 2^n. ¿Por qué?

En esta expansión, preguntemos ¿cuántos términos hay de la forma a^n? Esto corresponde al número de maneras de escoger, de n factores, n de los cuales son iguales a a. El término respectivo de la expansión es $\binom{n}{n}a^n$. Ahora, ¿cuántos términos habrá de la forma $a^{n-1}b$? Nuevamente, nuestro modelo nos indica que es igual al número de maneras de escoger, de n factores, $n-1$ iguales a a. El factor restante en cada caso es igual a b. Esto es $\binom{n}{n-1}a^{n-1}b$.

Procediendo de esta manera, obtenemos que

$$(a+b)^n = \binom{n}{n}a^n + \binom{n}{n-1}a^{n-1}b + \binom{n}{n-2}a^{n-2}b^2 + \ldots + \binom{n}{2}a^2b^{n-2} + \binom{n}{1}ab^{n-1} + \binom{n}{0}b^n,$$

fórmula conocida como el Teorema del Binomio. Nótese que n aquí es un número natural.

El teorema era conocido ya por Omar Khayyam (1048?-1122) quien en una de sus obras (*El álgebra*) menciona que puede desarrollar dicha expansión para $n = 4, 5, 6$ y más(*) y que la ha demostrado en una obra ya realizada. (Sin embargo, de ésta última aparentemente no quedan copias.)

Para que sea útil, necesitamos una forma de calcular las expresiones de combinatorias. Si queremos contar el número de formas de escoger de n 'cosas', k de ellas, $\binom{n}{k}$, podemos pensar como sigue. Si hay n cosas, tenemos n posibilidades para la primera cosa escogida, nos quedan $n-1$ posibilidades para la segunda, $n-2$ para la tercera, ..., $n-(k-1) = n-k+1$ para la k-ésima. O sea, $n(n-1)(n-2)\cdots(n-k+1)$. Sin embargo, este análisis toma en cuenta el orden en que se encogen las 'cosas', mientras que el orden no nos interesa (recuerde los casos que estudiamos para $n = 2, 3$). Cada conjunto de k cosas puede ordenarse de $k!$ maneras diferentes, de donde, en la expresión anterior cada combinación de k 'cosas' fue contada $k!$ veces. Luego, el valor de

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} = \frac{n!}{(n-k)!k!}.$$

4.3.2 El triángulo de Pascal

En la escuela secundaria con alguna frecuencia se presenta (sin demostración) el Teorema del Binomio por medio de este conocido triángulo.

$$
\begin{array}{ccccccc}
1 & & & & & & \\
1 & 1 & & & & & \\
1 & 2 & 1 & & & & \\
1 & 3 & 3 & 1 & & & \\
1 & 4 & 6 & 4 & 1 & & \\
1 & 5 & 10 & 10 & 5 & 1 & \\
\vdots & & & & & &
\end{array}
$$

Este enfoque o presentación del teorema tiene una larga historia. El matemático chino, Chu Shi-kié (1303) produjo este mismo arreglo triangular de coeficientes correspondientes a la expansión binomial. Hacia 1544 el matemático alemán Michael Stifel (1486?-1567) introdujo la expresión 'coeficiente binomial' y mostró como calcular $(1+a)^{n+1}$ a partir de $(1+a)^n$. Es claro que este arreglo de números en el cual cada número es la suma de los dos inmediatamente superiores era conocido por Tartaglia (quien dice en 1566 que es invención suya). Stifel y (el matemático e ingeniero belga Simón) Stevin. En 1654 Blaise Pascal (1623-62) investigó el arreglo usando el enfoque de la combinatoria y obtuvo muchas propiedades nuevas que fueron dadas a conocer en una publicación pósthumo *Traité du triangle arithmétique*. Estas son especialmente interesantes en cuanto los aplicó a la teoría de probabilidad. Subsecuentemente el arreglo recibió el nombre que todavía se usa de Triángulo de Pascal.

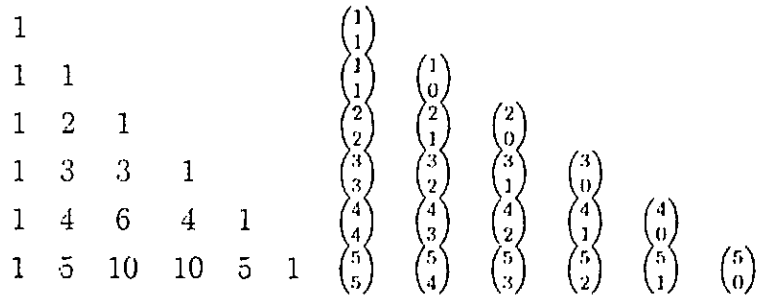
En el año 1665, Isaac Newton (1642-1727) mostró cómo se puede calcular $(1+a)^n$ directamente sin referencia a $(1+a)^{n-1}$. Esto involucra el conteo de combinaciones precisamente como lo presentamos anteriormente.

Veamos cómo podemos reconciliar estas distintas formas de calcular los coeficientes binomiales considerando sus leyes de formación. La característica más sobresaliente del triángulo de Pascal es su ley de formación: que cada nuevo término es la suma de dos términos. el inmediatamente encima de él y el que está a la izquierda de éste. en el triángulo.

Pues, pensemos ahora cómo podemos calcular una combinatoria, digamos el número de combinaciones de 5 cosas tomadas de 2. Es claro que las combinaciones se pueden formar de dos maneras. o bien se escogen 2 de 4 y no se escoge la quinta, o bien se escoge 1 de 4 y luego se escoge la quinta. En símbolos

$$\binom{5}{2} = \binom{4}{1} + \binom{4}{2}.$$

O sea, la ley de formación de estas combinatorias es la misma que la de los números en el triángulo de Pascal, dando lugar a la pareja de triángulos (parciales) de la siguiente figura:



En su escrito sobre el triángulo, Pascal demostró una gran cantidad de propiedades interesantes que pueden constituirse en problemas no rutinarios y retadores para el estudiante de la secundaria. A continuación proponemos algunos referidos a la disposición original del triángulo ideada por Pascal y que se muestra (en parte) a continuación. En el los coeficientes binomiales se leen diagonalmente.

1	1	1	1	1	1	1	1	1	1
1	2	3	4	5	6	7	8	9	
1	3	6	10	15	21	28	36		
1	4	10	20	35	56	84			
1	5	15	35	70	126				
1	6	21	56	126					
1	7	28	84						
1	8	36							
1	9								
1									

Puntos de discusión

1. Cada número en el triángulo es la suma de los números en la fila anterior a partir de la primera columna hasta la columna donde está ubicado el número.

2. Cada número en el triángulo es igual a la suma de todos los números en la columna anterior, desde la primera fila hasta la fila donde está ubicado el número.

3. Cada números del triángulo disminuido en 1 es igual a la suma de todos los números en el rectángulo encerrado por su fila y su columna (sin incluir a los números que están en éstas).

4. Los elementos de la fila j son los mismos que los de la columna j en el mismo orden.

5. La suma de los números en cualquier diagonal es igual al duplo de la suma de los números en la anterior diagonal.

6. La suma de los números en la n-ésima diagonal es igual a 2^{n-1}.

7. Si dos números están en la misma diagonal, la razón entre el número superior y el número inferior es igual a la razón entre la cantidad de elementos desde el número superior (inclusive) hasta el extremo superior de su columna y la cantidad de elementos desde el número inferior (inclusive) hasta el extremo inferior de su columna.

8. Enunciar y demostrar al menos dos propiedades adicionales de los elementos que se encuentran en el triángulo de Pascal.

Punto de investigación

No cabe duda que un campo donde el trabajo con el teorema del binomio tiene gran influencia es en el cálculo mismo. Hagamos memoria sobre cómo se suelá calcular la derivada de la función $f(x) = x^n$. (La manera contemporánea es muy similar a la forma como el mismo Newton la calculó.) Tenemos

$$f'(x) = lim_{\Delta x \to 0} \frac{f(x + \Delta x)}{\Delta x}$$

Esto equivale a calcular el límite siguiente

$$lim_{\Delta x \to 0} \frac{(x + \Delta x)^n}{\Delta x}.$$

Para proceder se requiere aplicar el teorema del binomio para desarrollar el numerador de esta expresión y 'simplificar' luego dividiendo por Δx.

Investigar otros usos del Teorema del Binomio en el cálculo.

Volveremos a usar el Teorema cuando tratamos las raíces complejas de un polinomio.

4.3.3 Las relaciones de Newton

Ahora, volvamos a considerar funciones simétricas de las raíces de un polinomio. Entre las funciones mas sencillas se encuentran la suma de las potencias de las raíces, es decir, sumas de la forma $s_k = r_1^k + r_2^k + ... + r_n^k$, donde $r_1, r_2, ..., r_n$ son las raíces de un polinomio de grado n. Para nuestras consideraciones, necesitamos saber cómo encontrar la derivada de una función polinómica $f(x)$, y estamos suponiendo que el lector está familiarizado con

dicho procedimiento. Sin embargo, notamos que podemos definir la derivada estricatmente a partir de nociones algebraicas. Comenzamos por la función $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0$ y expandámos $f(x+h)$ en potencias de h. Obtenemos

$$f(x+h) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0 + h[na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \ldots + a_1]$$

$$+\frac{h}{2!}[n(n-1)a_n x^{n-2} + (n-1)(n-2)a_1 x^{n-3} + \ldots + 2a_{n-2}] + \ldots.$$

Definimos la primera derivada de f como el coeficiente de h, la segunda derivada como el coeficiente de $\frac{h}{2!}$, y así sucesivamente. Es claro que la segunda derivada es, a su vez, la primera derivada de la primera derivada, etc.

Punto de discusión

¿Cómo se relaciona la definición anterior de las sucesivas derivadas de $f(x)$ con las definiciones usuales?

Así las cosas, hemos deducido una fórmula muy importante para las llamadas series de Taylor, a saber

$$f(x+h) = f(x) + \frac{h}{1}f'(x) + \frac{h}{2!}f''(x) + \ldots + \frac{h}{n!}f^{(n)}(x).$$

Esta fórmula es simétrica en x y h, de modo que es posible intercambiar los papeles de x y h para obtener

$$f(x+h) = f(h) + \frac{x}{1}f'(h) + \frac{x}{2!}f''(h) + \ldots + \frac{x}{n!}f^{(n)}(h).$$

Ahora intentemos obtener expresiones en términos de los coeficientes de $f(x)$ para las distintas potencias de las raíces $r_1, r_2, .., r_n$. Para ello definimos

$$s_1 = r_1 + r_2 + \ldots + r_n$$

$$s_2 = r_1^2 + r_2^2 + \ldots + r_n^2$$

$$s_3 = r_1^3 + r_2^3 + \ldots + r_n^3$$

.....

De hecho, ya sabemos que $a_n s_1 + a_{n-1} = 0$. Sin embargo, para hallar $s_2, s_3, ...$, usaremos las derivadas. Escribamos $f(x) = (x-r_1)(x-r_2)\cdots(x-r_n)$ y calculemos $f'(x)$ a partir de $f(x+h)$. Tenemos

$$f(x+h) = (x - r_1 + h)(x - r_2 + h) \cdots (x - r_n + h).$$

Y es claro que el coeficiente de h en esta expresión será

$$f'(x) = [(x-r_2)(x-r_3)\cdots(x-r_n)]+[(x-r_1)(x-r_3)\cdots(x-r_n)]+...+[(x-r_1)(x-r_2)\cdots(x-r_{n-1})]$$

Esta última expresión puede escribirse como

$$f'(x) = \frac{f(x)}{x - r_1} + \frac{f(x)}{x - r_2} + ... + \frac{f(x)}{x - r_n}.$$

Realizando las divisiones indicadas y sumando los coeficientes de cada potencia de x, tenemos

$$f'(x) = a_n x^{n-1} + (a_n r_1 + a_{n-1})x^{n-2} + (a_n r_1^2 + a_{n-1} r_1 + a_2)x^{n-3} + ...$$

$$a_n x^{n-1} + (a_n r_2 + a_{n-1})x^{n-2} + (a_n r_2^2 + a_{n-1} r_2 + a_{n-2})x^{n-3} + ...$$

$$.... + ...$$

$$a_n x^{n-1} + (a_n r_n + a_{n-1})x^{n-2} + (a_n r_n^2 + a_{n-1} r_n + a_{n-2})x^{n-3} + ...$$

lo cual es igual a

$$n a_n x^{n-1} + (a_n s_1 + n a_{n-1})x^{n-2} + (a_n s_2 + a_{n-1} s_1 + n a_2)x^{n-3} + ...$$

Ahora bien, usando la definición para la derivada, tenemos que

$$f'(x) = n a_n x^{n-1} + (n-1)a_{n-1} x^{n-2} + (n-2)a_{n-2} x^{n-3} + ...$$

e igualando términos, encontramos finalmente que

$$a_n s_1 + a_{n-1} = 0 \tag{4.4}$$

$$a_n s_2 + a_{n-1} s_1 + 2a_{n-2} = 0 \tag{4.5}$$

$$a_n s_3 + a_{n-1} s_2 + a_{n-2} s_1 + 3a_{n-3} = 0 \tag{4.6}$$

y, en general, que

$$a_n s_i + a_{n-1} s_{i-1} + a_{n-2} s_{i-2} + ... + i a_{n-i} = 0.$$

Este desarrollo nos permite escribir las fórmulas de Newton para $s_1, s_2, ...$

4.3.4 Uso de las fórmulas de Newton en la solución de problemas

Vamos a resolver un problema, primero directamente, usando manipulación algebraica conocida y, segundo, usando las fórmulas anteriores.

- Hallar la suma de los cuadrados y de los cubos de las raíces de la ecuación $x^3 - px^2 + qx - r = o$.

Para ello, sean a, b y c las raíces de la ecuación dada. Entonces, a partir de la familiar identidad algebraica

$$(a+b+c)^2 = a^2 + b^2 + c^2 + 2(ab + bc + ca)$$

y teniendo en cuenta que

$$a+b+c = p \qquad y \qquad ab + bc + ca = q,$$

encontramos que la suma de los cuadrados de las raíces se puede expresar como

$$a^2 + b^2 + c^2 = p^2 - 2q.$$

Ahora, aplicando directamente nuestras fórmulas, tenemos que la suma de los cuadrados de las raíces,

$$s_2 = -\frac{a_{n-1}}{a_n} s_1 - \frac{2a_{n-2}}{a_n}; \qquad o \qquad s_2 = -\frac{a_{n-1}}{a_n}\frac{-a_{n-1}}{a_n} - \frac{2a_{n-2}}{a_n},$$

que en este caso nos da $s_2 = p^2 - 2q$.

Ejercicios

Ejercicio 1 *Hallar la suma de los cubos de las raíces de* $x^3 - px^2 + qx - r = o$ *de dos maneras diferentes.*

Ejercicio 2 *Hallar la suma de los cuadrados y los cubos de las raíces de la ecuación*

$$x^4 + qx^2 + rx + s = 0.$$

Ejercicio 3 *Hallar la suma de las potencias cuartas de las raíces de la ecuación* $x^3 + qx + r = 0$.

4.4 Del cálculo al álgebra

4.4.1 Caracterización y distribución de las raíces de un polinomio

Un nuevo acercamiento al Teorema del Factor: el método de Horner

Como hemos comentado, con el reconocimiento de las relaciones de Viète, se puede pensar en la solución de una ecuación polinómica no sólo por fórmula sino también por factorización. Hemos mostrado que la solución por factorización no puede aplicarse algorítmicamente. Sin embargo, el llamado método de Horner permite un planteamiento interesante en este respecto. Pues, aunque originalmente fue diseñado para evaluar mas eficientemente un polinomio para cierto valor de la variable, el método de Horner también puede usarse en la identificación de factores.

Veamos primero cómo este método permite evaluar un polinomio mas eficientemente. Pensemos en la siguiente situación. En el salón de clase se coloca un reloj que emite una señal sonora cada segundo. Cada vez que suena el reloj el alumno puede oprimir una tecla de su calculadora. Hay que evaluar el polinomio

$$x^4 + 20x^3 - 6x^2 - 7x + 11$$

en $x = 4$. ¿Quién puede terminar primero?

[Esta pregunta se presta para iniciar una competencia animadora entre sus alumnos. ¿Cree Ud. que los alumnos nunca inventarán por su cuenta un método que no se les ha enseñado? ¿No pueden ni siquiera usar correctamente los métodos que los profesores se les ha explicado? Tal vez el obstáculo está precisamente allí en enseñar sin involucrar al alumno como agente activo, o en enseñar sin motivar al alumno o despertar su interés, en resumen, en no orientar correctamente la posibilidad de enfrentar nuevas situaciones sin una guía rígida del profesor. Ahora pensemos. La eficiencia en el cálculo es una característica importante de los programas de computador. El computador ha vuelto la espalda a ciertos métodos teóricos reemplazándolos por métodos numéricos; éste es un problema que ejemplifica un tema de actualidad. ¿Cómo lo podemos enfocar? Tal vez, si nuestros alumnos saben programar podemos pedirles que desarrollen el programa más eficiente que puedan. Si no tienen acceso a un computador, nos tocará pedirles que ahorren esfuerzo a sí mismos. Si no llegan a inventar un método como el de Horner, tenemos que estar listos con una nueva aproximación a la solución.]

Para aproximarnos al método respondamos los siguientes

1. *Puntos de discusión*

2. ¿Cuál es el modo más eficiente de calcular $97x+81x$? ¿No es cierto que es más rápido y eficiente factorizar la x, reduciendo en uno el número de operaciones que hay que efectuar?

3. ¿Cuál es el modo más eficiente de calcular $97x^2 + 81x$?

4. ¿Cuál es el modo más eficiente de calcular $97x^3 + 81x^2 - 47x$?

5. ¿Puede usted decirnos cuál es el método de Horner de evaluación de un polinomio?

Evaluemos el polinomio $f(x) = x^4 + 20x^3 - 6x^2 - 7x + 11$ en $x = 4$ usando el método de Horner. ¿Puede usted relacionar su conjetura sobre el funcionamiento del método de Horner con la siguiente tabla?

1	20	−6	−7	11	4
	4	96	360	1412	
1	24	90	353	1423	

Puntos de discusión

1. Usar la tabla anterior para expresar $f(x)$ en la forma $q(x)(x-4)+r$, para algún polinomio $q(x)$ y constante r. ¿Qué representa la constante r?

2. ¿Cómo se puede usar el método de Horner para identificar un factor de un polinomiodado?

Ahora repitamos el proceso hasta expresar $f(x)$ como un polinomio en la 'variable' $x-4$.

1	20	−6	−7	11	4
	4	96	360	1412	
1	24	90	353	1423	
	4	108	792		
1	28	202	1161		
	4	128			
1	32	330			
	4				
1	36				

lo que nos da $f(x) = (x-4)^4+36(x-4)^3+330(x-4)^2+1161(x-4)+1423$.
Puntos de discusión

1. Explicar completamente la afirmación anterior.

2. Usar este procedimiento para expresar el polinomio $f(x) = x^4 - 25x^3 + 159x^2 - 92x + 407$ como un polinomio en la 'variable' $x - 11$.

3. Usar el procedimiento anterior para hallar $q(x)$ en la expresión

$$f(x) = x^4 - 25x^3 + 159x^2 - 92x + 407 = (x - 11)q(x).$$

Puntos de investigación

1. Hallar el método más eficiente para evaluar c^n en una calculadora.

2. Para valores pequeños del entero n, determinemos la expansión de x^n en términos de $(x - 1)$. Busquemos regularidades que dependen de n y k con respecto de los coeficientes de $(x - 1)^k$. Hallar una fórmula general para estos coeficientes. Volver a escribir la ecuación que se obtiene haciendo la sustitución $x' = x + 1$.

4.4.2 Raíces múltiples

Ya hemos visto que, para lograr enunciar el Teorema Fundamental del Algebra es necesario llegar a aceptar como genuinas tanto las raíces reales negativas como las raíces complejas. Otro ingrediente indispensable es tomar en cuenta las multiplicidades de las raíces. Que un mismo valor de la variable debe contarse varias veces como raíz no es inmediatamente claro y su importancia se reconoce, de hecho, o bien a partir de la factorización del polinomio o bien a partir de su gráfica. Estas técnicas a su vez aparecen sólo en la primera mitad del siglo XVII. Con la introducción del cálculo, se logra encaminar una discusión amplia que arroja los principales resultados al respecto. En la sección anterior vimos (Teorema del Factor) que a es una raíz de la ecuación polinómica $f(x) = 0$ si y sólo si $x - a$ es factor de $f(x)$, es decir, si $f(x) = (x - a)q(x)$.

Definición 2 *Decimos que a es una raíz simple o una raíz de multiplicidad 1, si $(x - a)^2$ no es factor de $f(x)$.*

Definición 3 *Mas generalmente, se dice que a es una raíz de multiplicidad k de $f(x)$ si $(x - a)^k$ es factor de $f(x)$ y $(x - a)^{k+1}$ no lo es.*

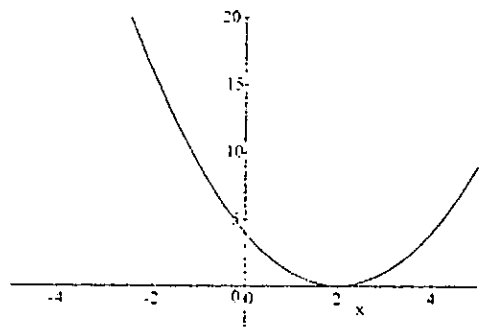
En este último caso, si separamos $f(x)$ en factores lineales, es claro que tendremos k factores $(x - a)$ en la expresión. Por ello, se cuenta a un total de k veces cuando se están contando las raíces del polinomio y se dice que a es una raíz múltiple.

La derivada del polinomio nos proporciona un criterio para determinar la multiplicidad de una raíz, pues si a es raíz de $f(x)$ y a no es raíz de $f'(x)$ entonces a es raíz simple. Además, si a es raíz de $f(x)$ de multiplicidad k, entonces a es raíz de $f'(x)$ de multiplicidad $k-1$, pues si tenemos $f(x) = (x-a)^k q(x)$, donde $(x-a)$ no divide a $q(x)$, entonces,

$$f'(x) = k(x-a)^{k-1}q(x)+(x-a)^k q'(x) = (x-a)^{k-1}[kq(x)+(x-a)q'(x)] = (x-a)^{k-1}Q(x).$$

Ahora, $(x-a)$ no es factor de $Q(x)$ puesto que $Q(a) = kq(a) \neq 0$.

Es interesante mirar las gráficas de las funciones polinomicas con raíces múltiples. Un primer ejemplo sencillo sería una función de la forma $f(x) = (x-2)^2$. Nuestra experiencia con la geometría analítica y las secciones cónicas nos dice que se trata de una parábola con vértice en el punto $(2,0)$



Gráfica de $f(x) = (x-2)^2$

Notamos que la gráfica es tangente al eje x en el punto $x = 2$.

Puntos de discusión

1. Calcular $f'(x)$ y hacer la gráfica. ¿Para qué valor de x se tiene que $p'(x) = 0$?
2. Hacer la gráfica de la función polinómica $l(x) = (x-3)^2(x+5)$ y la de su derivada.
3. ¿Será cierto que, siempre que una función polinómica tiene una raíz a de multiplicidad 2, sucede que la gráfica de la función no cruza el eje x sino que es tangente al eje en el punto $(a,0)$?
4. Estudiar la gráfica de la función polinómica $g(x) = (x-1)^3$ y la de su derivada. ¿Qué observa?

Ahora estamos en condiciones de resolver los siguientes problemas.

Ejercicios

Ejercicio 4 *Resolver en números reales el sistema de ecuaciones*

$$x + y + z = 3$$

$$x^2 + y^2 + z^2 = 3$$

$$x^5 + y^5 + z^5 = 3.$$

Sean x, y y z las raíces de la ecuación cúbica $t^3 + a_2t^2 + a_1t + a_0 = 0$. Usaremos las relaciones de Newton para resolver el problema. Sabemos que $a_3 = 1, s_1 = 3, s_2 = 3$ y $s_5 = 3$. Si calculamos a_2, a_1 y a_0 y resolvamos la correspondiente ecuación cúbica, encontraremos los valores de x, y y z que satisfacen el sistema de ecuaciones dado.

De $a_n s_1 + a_{n-1} = 0$ con $n = 3$ y $a_3 = 1$, tenemos $s_1 + a_2 = 3 + a_2 = 0$, de donde, $a_2 = -3$.

De $a_n s_2 + a_{n-1}s_1 + 2a_{n-2} = 0$ con $n = 3$ y $a_3 = 1$, obtenemos

$$s_2 + a_2 s_1 + 2a_1 = 3 + (-3)(3) + 2a_1 = 0,$$

de donde, $a_1 = 3$.

Sustituyendo estos valores en $a_n s_3 + a_{n-1}s_2 + a_{n-2}s_3 + 3a_{n-3} = 0$, y teniendo en cuenta que $n = 3$ y $a_n = 1$, se obtiene

$$s_3 = -3a_{n-3} = -3a_0.$$

Una sustitución mas en $a_n s_4 + a_{n-1}s_3 + a_{n-2}s_2 + a_{n-3}s_1 + 4a_{n-4} = 0$, donde es claro que $a_{n-4} = 0$, nos da

$$s_4 = -12a_0 - 9.$$

Finalmente, obtenemos la expresión para s_5 sin olvidar que tanto a_{n-4} como a_{n-5} son iguales a 0, como sigue.

$$a_n s_5 + a_{n-1}s_4 + a_{n-2}s_3 + a_{n-3}s_2 = a_3 s_5 + a_2 s_4 + a_1 s_3 + a_0 s_2 =$$

$$3 + 3(-12a_0 - 9) + (3)(-3a_0) + 3a_0 = 0,$$

lo cual nos permite concluir que $a_0 = -1$.

Se sigue que la ecuación a resolver es $t^3 - 3t^2 + 3t - 1 = 0 = (t-1)^3$ y 1 es raíz de multiplicidad 3 de esta ecuación. En conclusión, $x = y = z = 1$ es la única solución del sistema de ecuaciones dado. ■

Ejercicio 5 *Usar ideas similares a las del problema anterior para resolver el sistema de ecuaciones*

$$x + y = 8$$

$$x^2 + y^2 = 34.$$

4.4.3 El Teorema de Rolle

En la presente sección estudiaremos una serie de resultados que caractarizan las raíces de una ecuación polinómica. Estaremos interesados en analizar (acotar) el número de raíces reales de una ecuación determinada y, entre ellas, queremos contar las que son racionales, las que son positivas y las que son negativas. Para ello es conveniente echar mano a un resultado que pertenece al cálculo y que afirma

Teorema 4.4.1 *(Teorema de Rolle) Si los números reales a y b son ceros de un polinomio $f(x)$ con coeficientes reales, entonces existe un número real c, $a \leq c \leq b$, que es cero de $f'(x)$, la derivada del polinomio.*

Punto de discusión

¿Hay alguna relación entre este resultado y las observaciones que se hicieron en los anteriores puntos de discusión? Si lo hay, ¿cuál es?

Del teorema de Rolle se sigue de inmediato que, si todas las raíces del polinomio (con coeficientes reales) $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ son reales, entonces todas las raíces de su derivada $f'(x)$ también son reales. Es mas, entre dos raíces adyacentes de $f(x)$ hay exactamente una raíz de $f'(x)$ y esta raíz es simple. Es mas. Si $x_1, x_2, ..., x_k$ son las raíces reales de $f(x)$ con multiplicidades $m_1, m_2, ..., m_k$, respectivamente, Es claro que $m_1 + m_2 + \cdots + m_k = n$. Por otra parte, la derivada tendrá por raíces $x_1, x_2, ..., x_k$ con multiplicidades $m_1 - 1, m_2 - 1, ..., m_k - 1$, respectivamente. Además, por el Teorema de Rolle, tendrá al menos una raíz real en cada uno de los $k - 1$ intervalos $(x_1, x_2), (x_2, x_3), ..., (x_{k-1}, x_k)$. En total son $m_1 - 1 + m_2 - 1 + \cdots + m_k - 1 + k - 1 = n - 1$ raíces reales. Pero, $f'(x)$ es un polinomio de grado $n - 1$. Se sigue que éstas son todas sus raíces.

Una segunda consecuencia del Teorema de Rolle es que, si todas las raíces de $f(x)$ son reales y t de estas raíces son positivas, entonces $f'(x)$ tiene o bien t o bien $t - 1$ raíces reales positivas.

Punto de discusión

Sean $x_1, x_2, ..., x_j$ las raíces positivas de $f(x)$, demostrar esta última afirmación tomando en consideración el resultado anterior. Por otra parte, sea x_0 el mayor de las raíces de $f(x)$ que no son positivas. ¿Qué puede suceder entre x_0 y x_1?

4.4.4 Regla de signos de Descartes

Ahora estamos en condiciones de exponer y demostrar uno de los primeros instrumentos importantes que tuvieron los matemáticos para estudiar la distribución de las raíces de un polinomio, conocido como la regla de signos de

Descartes. Primero demostraremos una parte del resultado de Descartes y luego la complementaremos.

Teorema 4.4.2 *(Regla de Descartes) Si los coeficientes de un polinomio* $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ *son reales y si todas sus raíces también son reales, entonces el número de raíces positivas, teniendo en cuenta multiplicidades, es igual al número de cambios de signos en la sucesión de los coeficientes del polinomio. Si* $f(x)$ *tiene raíces complejas, entonces el número de raíces reales es igual al número de cambios de signo en la sucesión de coeficientes menos un número par.*

Punto de discusión
Considerar las funciones polinómicas

$$f_1(x) = x^2 - 7x + 10, f_2(x) = x^2 + 4x - 20, f_3(x) = x^3 - 3x^2 - 6x + 8.$$

Comprobar la validez de la regla de Descartes en cada caso.

Demostración. Demostraremos la primera parte de la regla de Descartes por inducción suponiendo, sin pérdida de generalidad, que a_n es positivo. Primero, observamos que, si todas las raíces de la ecuación son reales y t de ellas son positivas, entonces, el signo del último coeficiente, diferente de 0, de $f(x)$ es $(-1)^t$. Sea a_{n-t} ese coeficiente. Tenemos

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_{n-k}x^{n-k} =$$

$$a_n x^{n-k}(x - x_1)(x - x_2)\cdots(x - x_t)(x - x_{t+1})\cdots(x - x_{n-k}),$$

donde $x_1, x_2, ..., x_t$ son las raíces positivas de $f(x)$, y $x_{t+1}, ..., x_{n-k}$ son las raíces negativas y donde se ha tenido en cuenta la multiplicidad de cada raíz. Se sigue que

$$a_{n-k} = a_0(-1)^t x_1 \cdots x_t(-x_{t+1})\cdots(-x_{n-k})$$

y que cada uno de estos factores (con la posible excepción de $(-1)^t$ es positivo. Entonces, a_{n-k} tiene el mismo signo que $(-1)^t$.

Ahora, nuestro resultado puede demostrarse por inducción. Para polinomios de grado uno, el resultado es trivial. Pues, si $f(x) = a_1 x + a_0$ su única raíz es $\frac{-a_0}{a_1}$ y ésta es positiva solamente si a_1 y a_0 tienen signos opuestos, es decir, solamente si hay un cambio de signo.

Nuestra hipótesis de inducción es que el teorema se tiene para todo polinomio de grado $n-1$ con raíces reales y, con base en ella demostraremos que

también se tiene para cualquier polinomio $f(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_0$ de grado n.

Ahora bien, si $a_0 = 0$, consideremos el polinomio $f_1(x) = a_n x^{n-1} + a_{n-1}x^{n-2} + ... + a_1$. Las raíces positivas de $f(x)$ y $f_1(x)$ son las mismas y también el número de cambios de signos en la sucesión de sus coeficientes es el mismo en cada caso. Como, por hipótesis de inducción, la ley de signos de Descartes es válida para $f_1(x)$, se sigue que también es válida para $f(x)$.

Por otra parte, si $a_n \neq 0$, consideremos la derivada de $f(x)$,

$$f'(x) = na_n x^{n-1} + (n-a)a_{n-1}x^{n-2} + \cdots + a_1.$$

Es claro que el número de cambios de signo en el polinomio $f(x)$ y el número homólogo para $f'(x)$ coinciden si los signos de a_0 y el último coeficiente de la derivada coinciden y es mayor en 1, si estos signos son opuestos.

De acuerdo con nuestras consideraciones de entrada, en el primer caso el número de raíces positivas de $f(x)$ y de $f'(x)$ tiene la misma paridad y en el segundo tienen paridad opuesta. Pero, de lo que hemos deducido del Teorema de Rolle, el número de raíces positivas de un polinomio, si todas sus raíces son reales, puede ser igual al número de raíces positivas de su derivada o puede ser mayor en 1 que este número. Tomando esto en consideración, vemos que en el primer caso, $f(x)$ y $f'(x)$ tienen el mismo número de raíces positivas y que, en el segundo caso, $f(x)$ tiene una mas. Ahora, por nuestra hipótesis de inducción, la regla de signos de Descartes se cumple para $f'(x)$, de donde, se sigue que, en ambos casos, el número de raíces reales positivas de $f(x)$ es igual al número de cambios de signo en la sucesión de sus coeficientes. ■

Punto de discusión

. El teorema de Budan nos permite calcular el número de raíces positivas de un polinomio en un intervalo determinado (a, b). Usando la noción de Descartes de trasladar el polinomio (haciendo un cambio de variable) y la regla de signos de Descartes, determinar cómo esto puede lograrse.

4.4.5 Raíces complejas de un polinomio

Para terminar de demostrar la regla de Descartes, debemos estudiar la posibilidad de que no todas las raíces de una ecuación polinómica sean reales. Queremos mostrar que un polinomio con coeficientes reales siempre tiene un número par de raíces complejas, es decir, que las raíces complejas siempre se presentan en parejas.

Para ello haremos unas consideraciones acerca del conjugado de un número complejo.

Sea $z = a + bi$ un número complejo donde a, b son reales e $i^2 = -1$. El conjugado de z, que se escribe $\overline{z}$ es el número complejo $a - bi$.

De inmmediato podemos considerar algunas propiedades del conjugado.

1. Si $\overline{z} = z$, entonces $b = 0$, es decir. z es real. Para ver esto, observamos que si $a + bi = a - bi$, entonces $2bi = 0$, pero $i \neq 0$, de donde, $b = 0$.

2. $\overline{\overline{z}} = z$. Basta observar que $\overline{\overline{a + bi}} = \overline{a - bi} = a - (-b)i = a + bi = z$.

3. $\overline{z + w} = \overline{z} + \overline{w}$. ¡Demostrar!

3. $\overline{cz} = c\overline{z}$, donde $c \in \backslash BbbR$. Pues, $\overline{cz} = \overline{c(a + bi)} = \overline{ca + cbi} = ca - cbi = c(a - bi) = c\overline{z}$.

4. $\overline{z^n} = \overline{z}^n$. La demostración de esta última propiedad puede hacerse citando el Teorema del Binomio, ya que

$$z^n = (a+bi)^n = a^n + \binom{n}{1}a^{n-1}bi + \binom{n}{2}a^{n-2}(bi)^2 + \ldots + \binom{n}{n-1}a(bi)^{n-1} + \binom{n}{n}(bi)^n,$$

de donde,

$$\overline{z^n} = \overline{(a+bi)^n} = \overline{a^n + \binom{n}{1}a^{n-1}bi + \binom{n}{2}a^{n-2}(bi)^2 + \ldots + \binom{n}{n-1}a(bi)^{n-1} + \binom{n}{n}(bi)^n}$$

$$= \overline{a^n} + \overline{\binom{n}{1}a^{n-1}bi} + \overline{\binom{n}{2}a^{n-2}(bi)^2} + \ldots + \overline{\binom{n}{n-1}a(bi)^{n-1}} + \overline{\binom{n}{n}(bi)^n}$$

$$= a^n + \binom{n}{1}a^{n-1}\overline{bi} + \binom{n}{2}a^{n-2}\overline{(bi)^2} + \ldots + \binom{n}{n-1}a\overline{(bi)^{n-1}} + \binom{n}{n}\overline{(bi)^n}.$$

Ahora bien, recordando que $i^2 = -1, i^3 = -i$, e $i^4 = 1$, observamos que las potencias pares de i son reales mientras que las potencias impares son imaginarias, lo cual nos da que $\overline{(bi)^k} = (bi)^k$, k par y $\overline{(bi)^k} = -(bi)^k$, k impar. Se sigue que

$$\overline{(a+bi)^n} = a^n - \binom{n}{1}a^{n-1}bi + \binom{n}{2}a^{n-2}(bi)^2 - + \ldots + (-1)^{n-1}\binom{n}{n-1}a(bi)^{n-1} + (-1)^n\binom{n}{n}(bi)^n$$

$$= (a - bi)^n = \overline{z}^n.$$

Luego, sea $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0$ un polinomio con coeficientes reales. El número complejo z es raíz de

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0 \Longleftrightarrow a_n z^n + a_{n-1}z^{n-1} + \ldots + a_0 = 0$$

$$\iff \overline{a_n z^n + a_{n-1} z^{n-1} + \ldots + a_0} = \overline{0} \iff a_n \overline{z^n} + a_{n-1} \overline{z^{n-1}} + \ldots + a_0 = 0$$

$$\iff a_n \overline{z}^n + a_{n-1} \overline{z}^{n-1} + \ldots + a_0 = 0.$$

De esta discusión es claro que, siempre que un número complejo z es raíz de una ecuación polinómica, su conjugado $\overline{z} \neq z$ también lo es, o sea, las raíces complejas se presentan en parejas. De allí se cumple el segundo enunciado de la regla de signos de Descartes.

Existen otras dos demostraciones alternas del hecho de que ocurren en parejas las raíces complejas de una ecuación polinómica con coeficientes reales cuyos métodos nos interesan aquí. Uno de ellos usa el algoritmo de la división para demostrar que si $(x - (a + bi)) \mid f(x)$ entonces

$$(x - (a + bi))(x - (a - bi)) = [(x - a)^2 + b^2] \mid f(x).$$

El segundo se basa en la forma trigonométrica de un número complejo $z = r(\cos\theta + i \sin\theta)$ y utiliza el Teorema de de Moivre para demostrar que. $\overline{z^n} = \overline{z}^n$.

Puntos de investigación

1. ¿Cómo se formula el algoritmo de la división para polinomios? Usarla para demostrar el resultado anterior.

2. ¿Cuál es el Teorema de de Moivre? Usarlo para demostrar que $\overline{z^n} = \overline{z}^n$.

Puntos de discusión

1. De esta discusión se sigue que un polinomio con coeficientes reales y de grado impar tiene al menos una raíz real (como lo había observado Cardano). ¿Puede Ud. decir por qué?

2. ¿Cómo se puede determinar el número máximo de raíces reales negativas que tiene un polinomio cualquiera $f(x)$?

4.4.6 El Teorema de Sturm

Por toda su utilidad. la regla de signos de Descartes nos permite acotar el número de raíces reales positivas y reales negativas de un polinomio, pero no nos proporciona una respuesta exacta a las preguntas de si un polinomio con coeficientes reales tiene al menos una raíz real, de cuál es el número

de raíces reales que tiene en total y de cuántas raíces reales tiene en un intervalo determinado (a, b). Muchos matemáticos dedicaron mucho esfuerzo durante muchos años para dar solución a este problema hasta que, en 1835, el matemático francés Sturm encontró un procedimiento que permite determinar las respuestas a todas las tres. La idea es la siguiente. Sea $f(x)$ un polinomio con coeficientes reales y sea $f_1(x)$ la derivada ($f'(x) = f_1(x)$). Dividamos $f(x)$ por $f_1(x)$ y sea $f_2(x)$ el residuo que se obtiene, tomándolo con signo opuesto. En seguida, dividamos $f_1(x)$ por $f_2(x)$ y sea $f_3(x)$ el residuo tomado con signo opuesto, y así sucesivamente. Se puede demostrar que el último polinomio no nulo que se obtiene aplicando este procedimiento será un polinomio constante, digamos c. El teorema de Sturm dice que si $a < b$ son dos números reales que no son raíces de $f(x)$, y si se sustituye $x = a$ y $x = b$ en la sucesión de polinomios $f(x), f_1(x), f_2(x), ..., f_{n-1}(x), c$ se obtienen dos sucesiones

$$f(a), f_1(a), ..., f_{n-1}(a), c$$

$$f(b), f_1(b), ..., f_{n-1}(b), c$$

tales que el número de cambios de signos en la sucesión #1 es mayor o igual que el número de cambios de signo en la sucesión #2 y que la diferencia entre estos dos números es exactamente igual al número de raíces reales de $f(x)$ en el intervalo (a, b).

El Teorema de Sturm obviamente permite calcular el número de raíces reales de un polinomio con coeficientes reales en cualquier segmento de la recta real, lo que constituye una herramienta valiosísima.

4.5 Raíces enteras y racionales

4.5.1 Raíces enteras de una ecuación polinómica con coeficientes enteros

Ya que usaremos la aritmética modular fuertemente en la presente sección, comencemos por hacer una lista de las propiedades que utilizaremos.

(a) Si $a \equiv b(\mathrm{mod} m)$ y $c \equiv d(\mathrm{mod} m)$, entonces $a \pm c \equiv b \pm d(\mathrm{mod} m)$.

(b) Si $a \equiv b(\mathrm{mod} m)$ y $c \equiv d(\mathrm{mod} m)$, entonces $ac \equiv bd(\mathrm{mod} m)$. En particular, si $a \equiv b(\mathrm{mod} m)$, entonces $ac \equiv bc(\mathrm{mod} m)$ y $a^n \equiv b^n(\mathrm{mod} m)$.

(c) Si $ac \equiv bc(\mathrm{mod} m)$ y $m.c.d.(c, m) = 1$, entonces $a \equiv b(\mathrm{mod} m)$.

Con estas herramientas, podemos tratar el problema de encontrar ceros enteros de algunos polinomios. Por ejemplo, pensemos en primer lugar en una manera rápida de mostrar que el polinomio $x^2 - 131x + 267$ no puede tener un cero entero. Un tal cero debe ser o bien par o bien impar. Si es

par, los sumandos x^2 y $131x$ serán pares, dando un valor impar, y por ende diferente de 0, al polinomio. Igual cosa sucede si x es impar.

El argumento anterior utiliza el hecho de que 0 es par. Pero, 0 es múltiplo de cualquier número entero y, por consiguiente, de cualquier número primo. Se sigue que, si r es un cero entero de $q(x)$, entonces, $q(r) \equiv 0(\mathrm{mod}m)$ para cada entero m. Por una parte, las soluciones a esta congruencia son candidatos para ser ceros enteros del polinomio q, y por otra parte, si se puede encontrar un entero m para el cual la congruencia no tiene solución no podrá haber ceros enteros del polinomio.

En el transcurso de la solución de la siguiente serie de problemas, obtendremos métodos para resolver congruencias de la forma $q(x) \equiv 0(\mathrm{mod}m)$.

Consideremos la congruencia

$$x^2 - 9x - 36 \equiv 0(\mathrm{mod}m). \tag{4.7}$$

Punto de discusión

1. Mostrar primero que las soluciones de esta congruencia para $m = 8$ son las mismas que las soluciones de la congruencia

$$x^2 - x - 4 \equiv 0(\mathrm{mod}8).$$

2. Verificando que las únicas soluciones de la congruencia mod8 son $x \equiv 4$ y $x \equiv 5$, escribir todas las soluciones enteras de la congruencia entre 0 y 39, inclusive.

Ahora resolvamos la congruencia

$$x^2 - 9x - 36 \equiv 0(\mathrm{mod}5).$$

3. Escribir todas las soluciones enteras de la congruencia entre 0 y 39, inclusive.

4. Mostrar ahora que toda solución de la congruencia $x^2 - 9x - 36 \equiv 0(\mathrm{mod}40)$ es solución de una de las dos congruencias anteriores.

5. Usar este hecho para escribir toda solución de la congruencia módulo 40 entre 0 y 39, inclusive.

6. Ahora, usar los resultados anteriores para adivinar cuáles son los ceros del polinomio $x^2 - 9x - 36$.

7. Sea $q(x)$ un polinomio con coeficientes enteros. Demostrar que, si $a \equiv b(\mathrm{mod}m)$, entonces $q(a) \equiv q(b)(\mathrm{mod}m)$, para cualquier entero positivo m.

Ahora, sean $q(x)$ un polinomio con coeficientes enteros y m un entero positivo que es producto de potencias de primos p^k.

8. Mostrar que toda solución de la congruencia

$$q(x) \equiv 0(\mathrm{mod}m)$$

es también solución a la congruencia

$$q(x) \equiv 0(\mathrm{mod} p^k).$$

4.5.2 Raíces racionales de una ecuación polinómica con coeficientes enteros

Una de las actividades mas comunes del álgebra escolar es la de encontrar las raíces racionales de polinomios con coeficientes enteros. Los resultados teóricos requeridos son de fácil demostración y los consideraremos a continuación. Necesitaremos un resultado preliminar que desarrollaremos en el conjunto de puntos de discusión que se encuentra a continuación.

Puntos de discusión

Si tenemos el polinomio lineal $a_1x + a_0$, con a_1, a_0 enteros, sabemos que tiene una raíz racional, a saber $-\frac{a_0}{a_1}$. Y es claro que, si $-\frac{a_0}{a_1} = \frac{p}{q}$, donde p y q son primos relativos, se sigue que $a_0 = kp$ y $a_1 = kq$.

1. ¿Qué sucede si reemplazamos $x = \frac{p}{q}$ en la ecuación $a_nx^n + a_{n-1}x^{n-1} + ... + a_1x + a_0 = 0$, expresamos todos los sumandos con denominador común q^n, y luego multiplicamos ambos miembros de la ecuación por q^n?

2. ¿Por qué podemos concluir que $p \mid a_0$?

3. ¿Por qué podemos concluir que $q \mid a_n$?

En resumen tenemos el siguiente teorema.

Teorema 4.5.1 *Sea $f(x) = a_nx^n + a_{n-1}x^{n-1} + ... + a_1x + a_0$ un polinomio con coeficientes enteros. Si $f(\frac{p}{q}) = 0$, donde p y q son enteros relativamente primos, entonces $p \mid a_0$ y $q \mid a_n$.*

El procedimiento para encontrar las raíces racionales de un polinomio dado consta de elaborar una lista de todas las posibilidades de acuerdo con el Teorema 4.6.1, y comprobar en cada caso si se trata de una raíz o no usando, por ejemplo, el método de Horner. El procedimiento podría volverse bastante largo y tedioso, pero con la ayuda de la regla de signos de Descartes y de acotamiento de las raíces, se puede reducir la lista completa y hacer el procedimiento mas manejable. Pero existen dos teoremas especiales que nos permiten avanzar en el procedimiento con mayor rapidez. Veamos.

Teorema 4.5.2 *Sea* $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$, *donde* n *es mayor a igual a 2. Si* a_0, a_n *y* $f(1)$ *son todos impares, entonces* $f(x)$ *no tiene raíces reales.*

Notemos que $f(0) = a_0$ es impar. Ahora bien, sean p y q dos enteros, ambos pares o ambos impares. Entonces la diferencia $f(p) - f(q)$ es par dado que el valor

$$f(p) - f(q) = a_n(p^n - q^n) + a_{n-1}(p^{n-1} - q^{n-1}) + \ldots + a_2(p^2 - q^2) + a_1(p - q)$$

es divisible por el número par $p - q$. En particular, si p es par, la diferencia $f(p) - f(0)$ es par. Pero $f(0)$ es impar. Se sigue que $f(p)$ también debe ser impar lo que implica que $f(p) \neq 0$. Análogamente para p impar, la diferencia $f(p) - f(1)$ es par. Como por hipótesis $f(1)$ es impar, se sigue por el mismo razonamiento anterior que $f(p) \neq 0$. En consecuencia, $f(x)$ no tiene raíces enteras para ningún valor entero de x, par o impar.

Nótese que este resultado, donde intervienen nociones de divisibilidad entre enteros, proporciona un excelente criterio para analizar la existencia o no de raíces enteras de un polinomio.

Teorema 4.5.3 *Sea* $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$ *y sea* a *un entero tal que* $f(a) \neq 0$. *Si* p *y* q *son enteros relativamente primos tales que* $f(\frac{p}{q}) = 0$, *entonces* $(p - aq) \mid f(a)$.

Nuestra demostración comienza por notar que $f(x) - f(a)$ es un polinomio con coeficientes enteros y que a es una de sus raíces. Luego, por el Teorema del Factor, se tiene que $f(x) - f(a) = (x - a)g(x)$. De allí se sigue que

$$\left(\frac{p}{q} - a\right) g\left(\frac{p}{q}\right) = f\left(\frac{p}{q}\right) - f(a) = -f(a).$$

Ahora, multiplicando ambos miembros de la ecuación anterior por q^n nos da

$$q^{n-1}(p - aq)g(\frac{p}{q}) = -q^n f(a).$$

Ahora bien, si el polinomio $f(x)$ tiene grado n, entonces el polinomio $g(x)$ tiene grado $n - 1$ y es claro, del anterior conjunto de puntos de discusión, que $q^{n-1}g(\frac{p}{q})$ es un entero. Se sigue que $p - aq$ divide al miembro izquierdo de la ecuación, de donde, también divide al miembro derecho. Además, ya que p y q son primos relativos, también es cierto que $p - aq$ y q son primos relativos, de donde; $(p - aq) \mid f(a)$, como queríamos.

En el siguiente conjunto de puntos de discusión, se ilustra cómo se pueden usar estos dos resultados para acortar la búsqueda de raíces racionales de una ecuación polinómica.

Puntos de discusión

1. Consideremos el polinomio $f(x) = 36x^6 - 96x^5 + 49x^4 + 47x^3 - 33x^2 - 7x + 4$. Aplicando la regla de los signos de Descartes, ¿qué se puede decir acerca del número de raíces reales positivas que $f(x)$ puede tener? ¿Qué se puede decir acerca del número de raíces reales negativas que $f(x)$ puede tener?

2. ¿A qué es igual $f(-1)$?

3. ¿Cuáles son las posibilidades para las raíces de $f(x)$ de acuerdo con el teorema que se desarrolló en el anterior conjunto de puntos de discusión?

4. ¿Cuáles son las posibilidades teniendo en cuenta el último teorema que demostramos?

4.6 Mas contribuciones del álgebra al cálculo

4.6.1 Generalizaciones intuitivas que conllevan series y productos infinitos

Para finalizar este capítulo queremos explorar otras interrelaciones entre el álgebra y el cálculo desde el punto de vista de los matemáticos involucrados en la creación de este último, es decir, queremos comprender la inspiración algebraica detrás de varias de las herramientas fundamentales del cálculo.

Nadie duda que la buena notación algebraica asociada con la geometría analítica es de fundamental importancia para el desarrollo del cálculo. Pero para comprender cuán profundo fue el cambio de perspectiva generado por el enfoque algebraico (analítico) en la geometría, es necesario apreciar los nuevos problemas planteados por esta transformación.

Es conocido que a través de la geometría analítica se supera el inventario clásico de curvas definidas individualmente como lugares geométricos (círculo, secciones cónicas, algunos espirales) y se logra estudiar toda curva correspondiente a cualquier expresión algebraica. En este proceso se abre una avenida hacia el concepto de función.

En La Geométrie Descartes proyecta con claridad la importancia que ejercerá el enfoque algebraico y el concepto de función en la geometría cuando dice

> "Yo podría dar aquí varias otras formas de trazar y concebir una serie de líneas curvas, cada curva mas compleja que la anterior, pero creo que la mejor manera de agrupar todas las tales curvas y luego clasificarlas en orden, es por medio de reconocer el hecho de que todos los puntos de estas curvas que pueden llamarse "geométricas", es decir, que admiten una medición precisa y exacta, deben tener una relación definida con todos los puntos de una línea recta, y que esta relación debe expresarse por medio de una sola ecuación."

Ya que la geometría analítica de un solo golpe extiende sin límites las curvas geométricas conocidas, se hace urgente la búsqueda de métodos generales para resolver problemas tradicionalmente geométricos, como son los de encontrar tangentes (derivada) y normales a estas curvas o el de su "cuadratura" (integral o área debajo de la curva), pues los métodos conocidos tendían a tratar cada curva por separado e involucraban propiedades específicas a la curva para hacer sus análisis.

No es una exageración decir que las interrelaciones entre expresión algebraica y representación geométrica están implícitas en el significado de los resultados del cálculo.

Pero el álgebra permitió el desarrollo de muchas de las ideas del cálculo en formas no relacionadas directamente con la transformación de curva a función y de ello nos ocuparemos en lo que sigue.

Otras formas de explotación de las relaciones de Vieta

Hacia mediados del siglo XVII y prácticamente sin precedentes históricos, comenzaron a aparecer una gran cantidad de series y productos infinitos, que expresaban relaciones del todo insospechadas y que se lograron con base en la extensión a casos infinitos de los resultados concernientes a las relaciones que existen entre los coeficientes y las raíces de una ecuación polinómica. Vamos a considerar dos de éstos, uno estudiado por John Wallis y otro por James Gregory (1638-1675), contemporáneos y corresponsales de Newton; en algunos de sus apartes nuestros tratamientos se benefician de posteriores contribuciones de Euler (quien repensó y replanteó el trabajo de Wallis y Gregory).

Consideremos primero la ecuación

$$sen\, x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + - \ldots$$

Esta expresión fue lograda por Gregory y de hecho Newton se refirió a ella como un 'polinomio infinito'. (Hoy día se dice que la función *sen* x está expresada como una serie de potencias.)

Ahora bien, es claro que *sen* $x = 0$ para $x = 0, \pm\pi, \pm 2\pi, \pm 3\pi$, etc. Por otra parte, si las raíces de un polinomio son $r_1, r_2, r_3, \ldots$ se puede escribir la ecuación como el producto

$$(x - r_1)(x - r_2)(x - r_2) \cdots = 0.$$

En este punto haremos una suposición fuerte (y por el momento injustificada), tal como hicieron los matemáticos de aquél entonces, a saber, que esta última relación sirve no sólo para ecuaciones con un número finito de raíces, sino también en el caso de infinitas raíces. Por otra parte, supondremos que dos ecuaciones que tienen las mismas raíces son iguales. Así las cosas,

$$sen x = x\left(1 - \frac{x^2}{\pi^2}\right)\left(1 - \frac{x^2}{4\pi^2}\right)\left(1 - \frac{x^2}{9\pi^2}\right)\cdots,$$

pues la función del miembro derecho tiene los mismos ceros que la función *sen* x. Pero, al expandir este producto infinito vemos que el coeficiente de x^3 es el negativo de

$$\left(\frac{1}{\pi^2} + \frac{1}{4\pi^2} + \frac{1}{9\pi^2} + \ldots\right).$$

Además, en la serie infinita que escribimos al principio, el coeficiente de x^3 es $-\frac{1}{3!}$. Igualando coeficientes nos da

$$\frac{1}{\pi^2}\left(\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \ldots\right) = \frac{1}{3!}$$

de donde se obtiene

$$\frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \ldots = \frac{\pi^2}{6}.$$

Si examinamos los procedimientos anteriores, vemos que usamos una serie infinita (la de *sen* x), un producto infinito (la función con los mismos ceros que *sen* x) y finalmente obtuvimos el valor de otra serie infinita (la de los recíprocos de los cuadrados). Este último es del todo intrigante, ya que involucra el 'enigmático' número π.

Ahora bien, Wallis obtuvo $\frac{\pi}{2}$ como producto infinito, haciendo unas consideraciones similares. Escribimos la ecuación

$$sen(\pi x) = \pi x\left(1 - \frac{x^2}{1^2}\right)\left(1 - \frac{x^2}{2^2}\right)\left(1 - \frac{x^2}{3^2}\right)\cdots,$$

en virtud del hecho (que, recordamos, es una suposición hasta el momento no justificada) de que las dos funciones tienen los mismos ceros, a saber $x = 1, 2, 3, \ldots$. Ahora bien, poniendo $x = \frac{1}{2}$, se tiene

$$sen\left(\frac{\pi}{2}\right) = \frac{\pi}{2}\left(1 - \frac{1}{1^2 2^2}\right)\left(1 - \frac{1}{2^2 2^2}\right)\left(1 - \frac{1}{2^2 3^2}\right)\cdots.$$

De allí se sigue que

$$1 = \frac{\pi}{2}\left(\frac{2^2 - 1}{1^2 2^2}\right)\left(\frac{4^2 - 1}{2^2 2^2}\right)\left(\frac{6^2 - 1}{2^2 3^2}\right)\cdots.$$

Después de despejar y factorizar, se obtiene

$$\frac{\pi}{2} = \frac{2 \times 2}{1 \times 3} \cdot \frac{4 \times 4}{3 \times 5} \cdot \frac{6 \times 6}{5 \times 7} \cdots.$$

No cabe duda que el resultado de Wallis es completamente inesperado y, si causa sorpresa e incredulidad hoy, es difícil imaginar la maravilla de sus contemporáneos.

De estos dos ejemplos es claro que resultados algebraicos conocidos en el caso finito se desplazaron (indiscriminadamente) a casos infinitos y sirvieron para evaluar tanto sumas como productos infinitos. Aquí hay algunas dificultades de fondo, en particular, se supone sin demostración que las sumas y productos infinitos, en especial los que se usan como equivalentes a la función $sen\, x$, convergen. Mas adelante veremos que los matemáticos que trabajaron con ellos conocían la posibilidad de producir series y productos divergentes, pero no tenían adecuados criterios para establecer convergencia y destacar series y productos convergentes.

También notamos, para volver luego al mismo punto, que en las expresiones obtenidas, números irracionales son representados por sumas y productos infinitos de números racionales.

Una nota de importancia en este momento es el uso que Wallis dio a estas expresiones, empleando una analogía intuitiva. Para ubicar el trabajo de Wallis, debemos primero estudiar algunos de los resultados del matemático francés Giles Personne de Roberval (1602-75). Roberval utiliza la siguiente concepción de una superficie:

> "Una superficie está dividida en un número infinito de pequeñas superficies que son iguales, o tienen diferencias iguales o mantienen alguna progresión regular tal como de cuadrados a cuadrados, de cubos a cubos, y así sucesivamente. Y dado que las superficies están limitadas por líneas, en lugar de comparar las superficies, se puede comparar las líneas. La infinidad de líneas representa la infinidad de pequeñas superficies que componen la superficie total...."

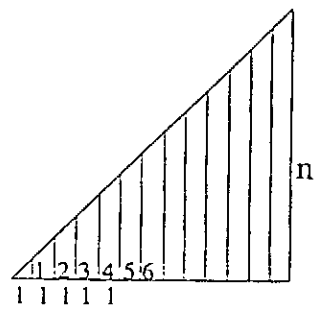


Figure 4.2:

Ahora, esta concepción algo 'cruda' de Roberval reaparece en el trabajo de Wallis en la siguiente forma. Wallis usa la suma límite de las potencias de los números naturales así. Observa primero que

$$\frac{0+1}{1+1}=\frac{1}{2},\frac{0+1+2}{2+2+2}=\frac{1}{2},\ldots,\frac{0+1+2+\ldots+n}{n+n+\ldots+n}=\frac{n(n+1)}{2n(n+1)}=\frac{1}{2}.$$

Observa luego que un triángulo puede pensarse como si estuviera compuesto por un número infinito de líneas cuyas longitudes están en progresión aritmética, la mayor de las cuales es su base. Se sigue que

$$\frac{\text{área del triángulo}}{\text{área del rectángulo}}=\frac{1}{2}.$$

Puntos de discusión

1. A partir de la ecuación de la recta $y=mx$, mostrar cómo se puede construir la interpretación de Wallis, a saber, que "un triángulo está compuesto por infinitas líneas cuyas longitudes están en progresión aritmética". ¿Habrá que modificar la interpretación en el caso de la ecuación $y=mx+b$?

2. ¿Cuáles son algunas de las suposiciones fuertes que encontramos en este tratamiento? ¿Qué es lo que se ha extendido (sin justificación) del caso finito al caso infinito ?

3. ¿Es siempre posible extender propiedades que se tienenen todo caso finito al caso infinito? ¿Qué hay de la proposición "el todo es mayor que la parte"?

Por otra parte, Wallis parte de la consideración de la suma de los cuadrados de los números naturales

$$\frac{0+1}{1+1}=\frac{1}{2}=\frac{1}{3}+\frac{1}{6},\frac{0+1+4}{4+4+4}=\frac{5}{12}=\frac{1}{3}+\frac{1}{12},\ldots,\frac{0+1+4+\ldots n^2}{n^2+n^2+\ldots+n^2}=\frac{1}{3}+\frac{1}{6n}$$

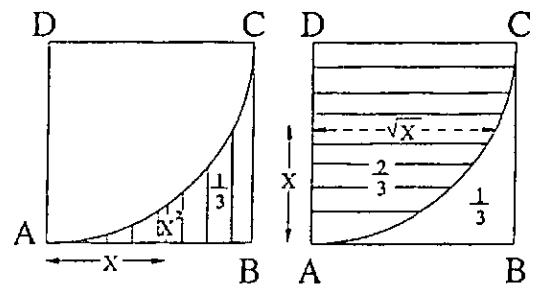


Figure 4.3:

para tomar el límite de esta razón cuando n tiende a infinito, a saber, $\frac{1}{3}$ y usarlo para determinar el área de la parábola, el volumen del cono, etc.

Punto de investigación

Investigar cómo se puede derivar la fórmula para el volumen del cono a partir de esta segunda sucesión estudiada por Wallis.

4.6.2 Generalización a suma de las k-ésimas potencias

Wallis extendió su método a sumas de una potencia cualquiera p de los números naturales obteniendo lo que es esencialmente este resultado.

$$lim_{n\to\infty}\frac{0^p+1^p+2^p+\ldots+n^p}{n^p+n^p+n^p+\ldots+n^p}=\frac{1}{p+1}.$$

Esta identidad sería la base para unas primeras consideraciones sobre la posibilidad de extender la noción de 'exponente' a números que no sean enteros positivos, como veremos a continuación.

Para calcular el área debajo de la parábola, cuya ecuación es $f(x) = x^2$, recurrimos, por supuesto, a la suma de los cuadrados que desarrollamos anteriormente. Obtuvimos

$$lim_{n\to\infty}\frac{0^2+1^2+2^2+\ldots+n^2}{n^2+n^2+n^2+\ldots+n^2}=\frac{1}{3}=\frac{\text{área } ABC}{\text{área } ABCD}.$$

Interpretando en términos gráficos, esto nos da derecho a suponer que el área de la parte restante del rectángulo es $\frac{2}{3}$.

Ahora bien, éste último resultado corresponde a

$$lim_{n\to\infty}\frac{\sqrt{0}+\sqrt{1}+\sqrt{2}+\ldots+\sqrt{n}}{\sqrt{n}+\sqrt{n}+\sqrt{n}+\ldots+\sqrt{n}} = \frac{\text{área rectángulo } ABCD - \text{área } ABC}{\text{área rectángulo } ABCD}$$
$$= \frac{2}{3}=\frac{1}{\frac{3}{2}} = \frac{1}{1+\frac{1}{2}}.$$

De aquí conjeturó Wallis que $\sqrt{x}$ puede pensarse como $x^{\frac{1}{2}}$. Consideraciones similares lo condujeron a concluir que $\sqrt[3]{n} = n^{\frac{1}{3}}, 1 = n^0$, etc. Finalmente, generalizó sus resultados afirmando que

$$lim_{n\to\infty}\frac{0^p+1^p+2^p+\ldots+n^p}{n^p+n^p+n^p+\ldots n^p} = \frac{1}{p+1};$$

para cualquier p positivo, negativo o fraccionario. El trabajo pionero de Wallis introdujo el tema de exponentes negativos y fraccionarios que manejaría Newton con maestría en su generalización del Teorema del Binomio.

4.6.3 El trabajo de Newton con series infinitas

La inspiración de Newton para el trabajo que adelantó con series infinitas se encuentra en una analogía inesperada con las fracciones decimales que transcribimos a continuación.

> "Ya que las operaciones de calcular con números y con variables son realmente similares, de hecho no parece haber diferencia alguna entre ellas excepto en cuanto a los símbolos que denotan cantidades, definidas en un caso e indefinidas en el otro, me sorprende que no se le ha ocurrido a ninguno (con excepción de N. Mercator con su cuadratura de la parábola) adaptar la doctrina recientemente establecida para números decimales de modo similar a variables, especialmente porque ello abre luego el camino a consecuencias mucho mas impactantes. Pues, ya que esta doctrina en especies tiene la misma relación con el álgebra que la doctrina de números decimales tiene a la aritmética común, las operaciones de adición, sustracción, multiplicación, división y extracción de raíces de ésta pueden aprenderse fácilmente de las de aquella si la persona tiene destreza en ambas, tanto la aritmética como el álgebra, y aprecia la correspondencia entre números decimales y términos algebraicos continuados al infinito: a saber, que a cada puesto en una sucesión decimal que decrece continuamente hacia la derecha le corresponde un único término en un

> arreglo de variables ordenado según la sucesión de dimensiones de numeradores o denominadores continuados en progresión uniforme al infinito. Y tal como la ventaja de los decimales consiste en que, cuando todas las fracciones y raíces han sido reducidas a decimales asumen en algún modo la naturaleza de enteros; así también la ventaja de las sucesiones infinitas de variables es que clases de términos más complicados (tales como fracciones cuyos denominadores son cantidades complejas, raíces de cantidades complejas y las raíces de ecuaciones 'afectadas') pueden reducirse a clases de términos sencillos: es decir, a series infinitas de fracciones con numeradores y denominadores simples y despojados de la complejidad casi insuperable que aquejaba los otros."

Aquí hay una visión grandiosa. Las series infinitas son al álgebra lo que las fracciones decimales son a la aritmética. Es especialmente notable la analogía newtoniana entre las expansiones decimales infinitas para números irracionales y las expansiones en series infinitas para funciones no polinómicas (racionales y radicales). Tal como la representación decimal de los irracionales permite operar con ellos como si fueran enteros, la representación en series infinitas de esas funciones, permitía operar con ellas como si fueran polinómicas.

Ahora intentemos explorar la visión Newtoniana que penetra todo el espíritu de la nueva teoría que hoy llamamos 'cálculo'.

Newton empleó dos medios diferentes para llegar a las series infinitas. Una de ellos consiste en aplicar las operaciones aritméticas comunes a las funciones algebraicas. Las operaciones de especial interés en este contexto son la división y la extracción de raíces. El propósito de efectuarlas es el de expresar la función en cuestión como una serie infinita de modo que se pueda sumarla o restarla fácilmente con cualquier otra función expresada de modo similar. Además, expresada la función como serie infinita, es fácil derivarla o integrarla, usando para ello únicamente la derivada o integral de funciones de la forma $f(x) = x^k$ y conocidas propiedades de la derivada e integral, a saber,

$$(f+g)' = f'+g' \quad (cf)' = c(f') \quad \int (f+g) = \int f + \int g \quad \int cf = c\int f, etc.$$

Ahora bien, Newton está especialmente interesado en funciones racionales y muestra su método con el ejemplo siguiente donde busca expresar la función racional $f(x) = \frac{a^2}{b+x}$ como una suma infinita.

$$
\begin{array}{lll}
a^2+ & \dots & \\
a^2+ & \frac{a^2x}{b} & \\
 & \underline{\frac{-a^2x}{b}} & \\
 & \frac{-a^2x}{b}- & \frac{a^2x^2}{b^2} \\
 & & \underline{\frac{a^2x^2}{b^2}} \\
 & & \frac{a^2x^2}{b^2}+ \quad \frac{a^2x^3}{b^3} \\
 & & \qquad \underline{\frac{-a^2x^3}{b^3}} \\
 & \vdots &
\end{array}
\qquad
\begin{array}{l}
\underline{|b+x} \\
\frac{a^2}{b} - \frac{a^2x}{b^2} + \frac{a^2x^2}{b^3} - \frac{a^2x^3}{b^4} + - \dots
\end{array}
$$

Puntos de discusión

1. ¿Usando esta serie infinita podemos hallar f' o $\int f$? ¿Cómo?

2. ¿Qué ventajas nos ofrece poder expresar la función como serie infinita?

3. ¿Qué tiene que ver esto con las palabras de Newton que transcribimos con anterioridad?

4. En particular, ¿cuál es la analogía entre las potencias $x, x^2, x^3, \dots$ en estas series de potencias y las potencias $10^{-1}, 10^{-2}, 10^{-3}, \dots = (10^{-1}), (10^{-1})^2, (10^{-1})^3, \dots$ de las fracciones decimales?

5. ¿Qué se sabe acerca de la convergencia de la series? ¿Para cuáles valores de x converge la suma de la progresión geométrica $1, x, x^2, x^3, \dots, x^n, \dots$?

6. ¿Por qué esta forma de expresar la función racional acorta y simplifica el trabajo matemático, por ejemplo, el de calcular su derivada o su integral?

Igualmente a Newton le interesó expresar funciones que contienen radicales en forma de series infinitas y para ello empleó el conocido algoritmo de extracción de raíces, de manera muy similar al procedimiento anterior con la división.

En la actualidad, este algoritmo no se enseña ya, dada la capacidad de las calculadoras del bolsillo mas ordinarias para calcular la raíz cuadrada de un número. Para comprender el ejemplo de Newton, sin embargo, hay que comprender el algoritmo que usaba.

Expresemos cada número M de dos o más cifras como $M = a \cdot 10 + b$. Entonces,

$$M^2 = a^2 \cdot 100 + 2ab \cdot 10 + b^2 = 100a^2 + (2 \cdot 10a + b)b.$$

El algoritmo de extracción de raíces se basa en la identidad $(a+b)^2 = a^2 + 2ab + b^2$, como mostramos a continuación por medio de un ejemplo.

Si vamos a extraer la raíz cuadrada de un número, digamos 2209, el algoritmo nos dirige a separar los dígitos del número en grupos de dos de derecha a izquierda. ¿Qué tiene que ver esta instrucción con la identidad anterior? Estamos considerando $\sqrt{2209} = \sqrt{22'09}$. Ahora, debemos escribir el mayor número cuyo cuadrado es ≤ 22, esto es 4, encima del 2, elevar al cuadrado, así

$$\begin{array}{c} 4 \\ \sqrt{2209} \\ 16 \end{array}$$

Luego se debe restar (es decir, de M^2 restamos $100a^2$), bajar los dos dígitos siguientes, multiplicar el 4 por 20 (¿por qué?), escribir 80 y buscar un número tal que cuando se suma este número a 80 y se multiplica por el número el resultado sea menor que el residuo 609 y lo mas próximo posible. En este caso el número buscado será 7; hacer la suma $80 + 7$ y multiplicar ésta por 7, así

$$\begin{array}{cccc} 4 & 7 & & \\ \sqrt{22} & \overline{09} & & \\ 16 & & & \\ 6 & 09 & (80{+}7) & \\ 6 & 09 & & \end{array}$$

En seguida volvemos a restar obteniendo un residuo de 0. ¿Qué hemos restado a M^2 en esta ocasión? Tenemos, pues,

$$2209 = 47^2 = (40{+}x)^2 = 40^2{+}(2{\cdot}40{+}x)x = 40^2{+}(80{+}7){\cdot}7 = 40^2{+}2{\cdot}40{\cdot}7{+}7^2.$$

Es obvio que este proceso sólo termina si el número cuya raíz se está extrayendo es el cuadrado de un número racional. Se puede aproximar una raíz cuadrada irracional al grado de exactitud deseado continuando el proceso anterior haciendo grupos de a dos dígitos a la derecha del punto decimal, etc.

Un ejemplo que trabaja Newton es $\sqrt{a^2 + x^2}$ como se aprecia en el siguiente desarrollo.

$$\begin{array}{llllllll}
a^2 & +x^2 & (a & +\frac{x^2}{2a} & -\frac{x^4}{8a^3} & +\frac{x^6}{16a^5} & -\frac{5x^8}{128a^7} & +\frac{7x^{10}}{256a^9} \quad -\frac{21x^{12}}{1024a^{11}} \cdots \\
a^2 & & & & & & & \\
 & +x^2 & & & & & & \\
 & x^2 & +\frac{x^4}{4a^2} & & & & & \\
 & & \frac{x^4}{4a^2} & & & & & \\
 & & \frac{x^4}{4a^2} & -\frac{x^6}{8a^4} & +\frac{x^8}{64a^6} & & & \\
 & & & \frac{x^6}{8a^4} & -\frac{x^8}{64a^6} & & & \\
 & & & \frac{x^6}{8a^4} & +\frac{x^8}{16a^6} & -\frac{x^{10}}{64a^8} & +\frac{x^{12}}{256a^{10}} & \\
 & & & & -\frac{5x^8}{64a^6} & +\frac{x^{10}}{64a^8} & -\frac{x^{12}}{256a^{10}} & \\
 & & & & -\frac{5x^8}{64a^6} & -\frac{5x^{10}}{128a^8} & +\frac{5x^{12}}{512a^{10}} & \cdots \\
 & & & & & \frac{7x^{10}}{128a^8} & -\frac{7x^{12}}{512a^{10}} & \cdots \\
 & & & & & \frac{7x^{10}}{128a^8} & -\frac{7x^{12}}{512a^{10}} & \cdots \\
 & & & & & & -\frac{21x^{12}}{512a^{10}} & \cdots \\
 & & & & & & & \\
\cdots & & & & & & &
\end{array}$$

Puntos de discusión

1. ¿Cómo se puede construir una función radical tal que, cuando se le aplica el algoritmo de extracción de raíces, dé por resultado un polinomio (finito)?

2. ¿Cómo se puede construir una función radical tal que, cuando se le aplique el algoritmo de extracción de raíces, dé por resultado una serie infinita de potencias?

3. Sumar $f(x) = \frac{a^2}{b+x}$ y $g(x) = \sqrt{a^2+x^2}$ de modo que la suma sea fácilmente derivable e integrable.

4.6.4 Generalización del Teorema del Binomio

Nuestro propósito principal en esta sección es estudiar la generalización, debida a Isaac Newton, del Teorema del Binomio a casos en los cuales el exponente no es un número natural, sino que puede ser fraccionario o negativo. Para ello es importante conocer más de cerca la forma en que Newton estaba pensando las cosas.

Newton llegó a enunciar y utilizar el Teorema del Binomio para exponentes fraccionarios y negativos, los cuales producen series infinitas, como una generalización natural del caso de exponentes enteros positivos; nunca ofreció una demostración de estos descubrimientos, sino se limitó a verificar algunos casos particulares. Por ejemplo, comprobó que al elevar al cuadrado la serie que obtuvo para $(1+x)^{1/2}$ se obtiene el resultado $1+x$. Newton y su colaborador James Gregory estuvieron convencidos de la necesidad de producir una demostración, pero del trabajo de Newton al respecto sólo quedan algunas cartas. Por ejemplo, en 1676 dirigió dos cartas a Henry Oldenburg, secretario de la Royal Society en las cuales enuncia el resultado general de la expansión de $(P+PQ)^{m/n}$ que dice haber investigado antes de 1669. Para Newton el resultado para exponentes fraccionarios era especialmente importante para extraer raíces, pues si Q es menor que 1, al factorizar P de la expresión anterior, la suma de las potencias de Q converge.

Para Newton éste era uno de los descubrimientos más bellos que él había podido hacer. Se dio cuenta de la posibilidad de la extensión y vio de inmediato su utilidad para el cálculo de raíces cuadradas y en otras situaciones. La generalización del Teorema del Binomio se constituyó en un segundo camino que llevó a Newton al trabajo con series infinitas. Veamos por qué.

Si se considera el caso $(a+b)^{\frac{1}{2}}$ y si simplemente se sustituye el valor $\frac{1}{2}$ por n en la fórmula binomial, tenemos

$$(a+b)^{\frac{1}{2}} = a^{\frac{1}{2}} + \binom{1/2}{1} a^{\frac{1}{2}-1} + \binom{1/2}{2} a^{\frac{1}{2}-2}b^2 + \binom{1/2}{3} a^{\frac{1}{2}-3}b^3 + \ldots$$

$$= a^{\frac{1}{2}} + \frac{\frac{1}{2}}{1} a^{\frac{1}{2}-1} b + \frac{\frac{1}{2}(\frac{1}{2}-1)}{2!} a^{\frac{1}{2}-2} b^2 + \frac{\frac{1}{2}(\frac{1}{2}-1)(\frac{1}{2}-2)}{3!} a^{\frac{1}{2}-3} b^3 + \ldots$$

Nótese que para un exponente natural n las diferencias en los numeradores de los coeficientes llegan a producir un factor de 0 cuando se llega al término con numerador $n(n-1)(n-2)\cdots(n-n)$ y en todos los términos de allí en adelante. Es decir, la expansión es finita. En cambio, para un exponente racional no entero esto no ocurre, dando lugar a una suma infinita.

Newton utilizó esta fórmula para aproximar raíces. Por ejemplo, para hallar un valor aproximado de $\sqrt{10}$, lo escribimos como

$$\begin{aligned}
(9+1)^{1/2} &= 9^{1/2} + \frac{\frac{1}{2}}{1} 9^{-1/2} \cdot 1 + \frac{\frac{1}{2}(-\frac{1}{2})}{2!} 9^{-3/2} \cdot 1^2 + \ldots \\
&= 3 + \frac{1}{2} \cdot \frac{1}{3} - \frac{1}{8} \cdot \frac{1}{27} + - \ldots \\
&= 3 + 0.16\overline{6} - 0.004\overline{6} + - \ldots \\
&\approx 3.162
\end{aligned}$$

Cuando el exponente es un entero negativo, también se obtiene una expansión infinita. Pues, aplicando directamente la fórmula binomial, tenemos, por ejemplo,

$$(1+x)^{-n} = 1^{-n} + \frac{-n}{1} x + \frac{(-n)(-n-1)}{2!} x^2 + \frac{(-n)(-n-1)(-n-2)}{3!} x^3 + \ldots.$$

Puntos de discusión

1. ¿Por qué es infinita esta expresión?
2. ¿Cuál es el coeficiente de x^r?

Ahora bien, fue en este contexto que Newton primero advertió que podrían presentarse dificultades importantes en el manejo de series infinitas y, en particular, notó el problema de convergencia. Para apreciar el problema, consideremos estos dos casos.

La serie infinita para $(1-x)^{-1} = \frac{1}{1-x}$ es $1 + x + x^2 + x^3 + x^4 + \ldots$. Si sustituimos el valor $x = 2$, obtenemos

$$-1 = 1 + 2 + 4 + 8 + 16 + \ldots.$$

En efecto, Newton se dio cuenta de que la expansión es válida únicamente para $|x| < 1$.

En la expansión que consideramos para $\sqrt{10}$, si en lugar de $(9+1)^{\frac{1}{2}}$ hubieramos tomado $(1+9)^{\frac{1}{2}}$ obtendríamos

$$(1+9)^{\frac{1}{2}} = 1^{\frac{1}{2}} + \frac{\frac{1}{2}}{1}1^{\frac{1}{2}-1}9 + \frac{\frac{1}{2}(-\frac{1}{2})}{2!}1^{\frac{1}{2}-2}9^2 + \ldots = 1 + \frac{9}{2} - \frac{81}{8} + - \ldots$$

Puntos de discusión

1. ¿Qué podemos decir de esta suma? ¿Nos dará una aproximación de $\sqrt{10}$?

2. Volviendo a la carta de Newton a Oldenburg, ¿qué condición impone a la expresión $(P+PQ)^{\frac{m}{n}}$? ¿Por qué es esta condición suficiente para garantizar convergencia?

4.6.5 Algunas consideraciones finales

En conclusión, es evidente que hay un rico intercambio entre temas y resultados generalmente considerados algebraicos y el desarrollo del cálculo. La analogía entre representación decimal infinita (irracionales) con la posibilidad de manipulación de estos números como si fueran enteros y la representación en series infinitas (funciones racionales y radicales) con la posibilidad de manipular estas funciones como si fueran polinómicas, es desconocida pero muy enriquecedora.

Es también intrigante enfatizar que, mientras que sumas y productos finitos de números racionales siempre producen racionales (clausurativiad), en cambio, sumas y productos infinitos de racionales pueden producir irracionales. Esto es implícito en la representación decimal de los irracionales, pero con poca frecuencia se hace explícito.

Si bien es común, de hecho indispensable, usar la manipulación algebraica en el cálculo, esperamos con esta sección haber contribuido a mostrar que también ideas fundamentales de la aritmética y el álgebra están plasmadas en las bases conceptuales históricas del cálculo.

4.7 Problemas del capítulo

1. Hallar una ecuación cúbica cuyas raíces son los cuadrados de las raíces de la ecuación $x^3 - x^2 + 3x - 10 = 0$.

2. Sean u, v, w los ceros del polinomio cúbico $4x^3 - 7x^2 - 3x + 2$. Construir un polinomio cúbico cuyos ceros son $u-(1/vw), v-(1/wu), w-(1/uv)$.

3. Sean m, n, p, q los ceros del polinomio cuártico $x^4 - 3x^3 + 2x^2 + 4x - 1$. Sin hallar m, n, p, q explícitamente, construir un polinomio de grado seis cuyos ceros son mn, mp, mq, np, nq, pq.

4. Hallar el polinomio mónico cuyos ceros son los recíprocos de los ceros del polinomio $x^3 - 2x^2 + 6x + 5$. Hallar, además, el polinomio de grado 3 con coeficientes enteros cuyos cero son esos mismos recíprocos y tal que el máximo común divisor de sus coeficientes sea 1.

5. Si la suma de los ceros de un polinomio (teniendo en cuenta multiplicidades) es 0, demostrar que la suma de los ceros de su derivada también es 0.

   (a) Demostrar que , si todos los ceros de un polinomio $f(x) = \sum^{n} a_r x^r$ son reales, entonces $a_{n-1}^2 \geq 2a_{n-2}a_n$ y $a_1^2 \geq 2a_0a_2$. Demostrar, además, que la inversa de esta afirmación no se cumple.

   (b) Usar la parte (a) para verificar que no todos los ceros del polinomio $x^6 + 2x^5 + 3x^4 - 4x^3 + 5x^2 + 6x + 7$ no son todos reales.

   (c) Aplicando el Teorema de Rolle, mostrar que si todos los ceros del polinomio son reales entonces $(n-1)a_{n-1}^2 \geq 2na_n a_{n-2}$. Dar un ejemplo para mostrar que la inversa de esta proposición no se tiene para $n \geq 3$.

6. Sea $f(x) = x^3 + ax^2 + bx + c$, donde a, b, c son reales. Sean u, v, w sus ceros. Verificar que $uv = w^2 + aw + b$ y que

$$(u - v)^2 = -[3w^2 + 2aw - (a^2 - 4b)] = (a^2 - 3b - f'(w).$$

7. Resolver la ecuación $x^4 - x^3 - 7x^2 + 23x - 20 = 0$ dado que el producto de dos de sus raíces es -5.

8. Considerar la ecuación polinómica

$$x^4 + px^3 + qx^2 + rx + s = 0.$$

   (a) Demostrar que el producto de dos de sus raíces es igual al producto de las otras dos si y sólo si $r^2 = p^2 s$.

   (b) Demostrar que la suma de dos de sus raíces es igual a la suma de las otras dos si y sólo si $p^3 + 8r = 4pq$.

   (c) Si se supone que $p^3 + 8r = 4pq$, es necesariamente cierto que la suma de dos de las raíces es igual a la suma de las otras dos?

9. Si las raíces de $x^3 + ax^2 + bx + c = 0$ están en progresión aritmética, demostrar que $2a^3 - 9ab + 27c = 0$.

10. Dados el producto p de los senos de los ángulos de un triángulo y el producto q de sus cosenos, demostrar que las tangentes de los ángulos son las raíces de la ecuación

$$qx^3 - px^2 + (1+q)x - p = 0.$$

11. Si a, b, c son las raíces de la ecuación

$$x^3 - x^2 - x - 1 = 0$$

(a) Demostrar que a, b, c son distintos.

(b) Demostrar que

$$\frac{b^n - c^n}{b-c} + \frac{c^n - a^n}{c-a} + \frac{a^n - b^n}{a-b}$$

es entero para $n = 1, 2, \ldots$.

12. Si el producto de dos de las cuatro raíces de la ecuación

$$x^4 - 18x^3 + kx^2 + 200x - 1984 = 0$$

es -32, hallar k.

13. Si x, y, z son reales y satisfacen $x + y + z = 5$ y $yz + zx + xy = 3$, demostrar que $-1 \leq z \leq \frac{13}{3}$.

14. Hallar todos los valores enteros de a tales que todos los ceros de $x^4 - 14x^3 + 61x^2 - 84x + a$ son enteros.

15. Determinar todos los polinomios de grado n que tengan todos sus coeficientes iguales a $+1$ o -1 y tales que tengan solamente ceros reales.

Chapter 5

Métodos numéricos en la solución de una ecuación polinómica

5.1 Solución de una ecuación polinómica por factorización

Como hemos comentado, con el reconocimiento de las relaciones de Viète, se puede pensar en la solución de una ecuación polinómica no sólo por fórmula sino también por factorización. Pero hemos demostrado además que la solución por factorización no puede aplicarse algorítmicamente. En el caso de las ecuaciones cuadráticas, el procedimiento mas sencillo de solución por factorización es el ensayo y error. Dada la ecucación $x^2 - 20x + 91 = 0$ buscamos dos números cuyo producto sea 91 y cuya suma sea 20. Ahora bien, las únicas factorizaciones de 91 son $91 = 1 \times 91$ y $91 = 7 \times 13$. Sólo este último cumple las dos condiciones impuestas por las relaciones de Viète. Obtenemos, por lo tanto, la factorización

$$x^2 - 20x + 91 = (x - 7)(x - 13) = 0.$$

La solución termina observando que, si un producto es igual a 0, uno de sus factores debe ser igual a 0, así que o bien $x - 7 = 0 \Longrightarrow x = 7$ o bien $x - 13 = 0 \Longrightarrow x = 13$. Toda aplicación de la factorización es del mismo estilo que este ejemplo, complicándose el procedimiento por ensayo y error cuando el término constante tiene muchas formas de expresarse como producto de dos factores.

Puntos de discusión

1. Analizar la solución de la ecuación $x^2 - x - 12 = 0$ por factorización a la luz de estas últimas consideraciones.

2. Analizar la solución de la ecuación $3x^2 + 15x - 84 = 0$ por factorización.

En el caso de ecuaciones de grado mayor que dos, el procedimiento por factorización puede resultar aun mas engorroso, aunque hemos estudiado algunos criterios adicionales que permiten acortar el proceso en el Capítulo IV). Allí vimos que el llamado *método de Horner* nos sirve para evaluar un polinómio dado $f(x)$ en determinado valor de la variable, digamos $x = b$ de manera económica. De hecho, el método de Horner permite expresar $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ como un polinomio en la 'variable' $(x - b)$ así

$$f(x) = A_n(x - b)^n + A_{n-1}(x - b)^{n-1} + \cdots + A_1(x - b) + A_0.$$

De aquí es evidente que

$$f(b) = A_0,$$

o sea, que el método evalúa el polinomio $f(x)$ en $x = b$. Además, si resulta que $f(b) = A_0 = 0$, se deduce de allí que b es raíz de la ecuación polinómica $f(x) = 0$ y que $(x - b)$ es factor de $f(x)$.

Recordemos el procedimiento.

El método de Horner y la solución por factorización

Consideremos el polinomio $f(x) = x^4 - 25x^3 + 159x^2 - 92x + 407$. Aplicando criterios de divisibilidad al término constante, vemos que 11 es uno de sus factores. En efecto, $407 = 11 \times 37$. Evaluemos el polinomio en $x = 11$. Tenemos

$$\begin{array}{rrrrrr} 1 & -25 & 159 & -92 & 407 & 11 \\ & 11 & -154 & 55 & -407 & \\ 1 & -14 & 5 & -37 & 0 & \end{array}$$

En efecto, $f(11) = 0$, es decir, 11 es raíz de la ecuación polinómica $f(x) = 0$ y

$$f(x) = (x - 11)(x^3 - 14x^2 + 5x - 37).$$

[Si seguimos el mismo procedimiento con el cociente $x^3 - 14x^2 + 5x - 37$ y así sucesivamente llegaremos a la expresión de f como polinomio en $(x - 11)$.

Veamos

1	−25	159	−92	407	11
	11	−154	55	−407	
1	−14	5	−37	0	
	11	−33	−308		
1	−3	−28	−345		
	11	88			
1	8	60			
	11				
1	19				

De este modo obtenemos $f(x) = (x-11)^4 + 19(x-11)^3 + 60(x-11)^2 - 345(x-11)$.]

Para terminar de resolver la ecuación $f(x) = 0$, debemos ahora buscar raíces de $x^3-14x^2+5x-37 = 0$ pero resulta que ésta ya no produce resultados ante la búsqueda de factores racionales (basta probar $\pm 1, \pm 37$ usando el método de Horner) y será necesario recurrir a otros medios. Veremos en la siguiente sección cómo puede usarse el método de Horner para aproximar una de las raíces de esta última ecuación cúbica, pues se sabe que tiene otra raíz real. (¿Por qué?)

Ejercicios

Usando el método de Horner, o de otro modo, resolver por factorización las siguientes ecuaciones polinómicas

1. $x^4 + 2x^3 - 2x - 4 = 0$
2. $x^4 - 7x^3 + 14x^2 - 8x = 0$
3. $-4x^2 - x^3 + 16x + 64 = 0$
4. $x^3 + 6x^2 + 11x + 6 = 0$

Puntos de discusión

5. ¿Es el método de Horner siempre aplicable para hallar un factor lineal y de allí resolver una ecuación cúbica? ¿Una ecuación cuártica?
6. ¿Cómo se puede resolver la ecuación $x^4 + 64 = 0$ por factorización?

Ejercicios

1. Factorizar sobre los enteros

   (a) $4x^2 + 1$

(b) $x^4 - 20x^2 + 4$

(c) $x^{10} + x^5 + 1$

2. Demostrar que las ecuaciones

$$x^4 - x^3 + x^2 + 2x - 6 = 0$$
$$x^4 + x^3 + 3x^2 + 4x + 6 = 0$$

tienen un par de raíces complejas en común.

3. Hallar un polinomio $f(x)$ tal que $(f(x))^5 - x$ es divisible por $(x-1)(x-2)(x-3)$.

4. Hallar valores de a, b tales que $(ax+b)(x^5+1) - (5x+1)$ sea divisible por x^2+1.

5. Un polinomio mónico con coeficientes enteros tiene la propiedad de que uno de sus ceros es igual al producto de los otros dos. Demostrar que debe ser factorizable (reducible) sobre los enteros.

6. ¿Para cuál entero a es $x^2 - x + a$ factor de $x^{13} + x + 90$?

7. Factorizar en enteros $(x^4 - 1)^4 - x - 1$.

5.2 Aproximación de las raíces reales de una ecuación polinómica

5.2.1 Los métodos aproximativos tienen una larga historia

La solución de una ecuación polinómica por aproximación tiene una historia muy larga. Hemos sugerido que el mismo uso de la base sesenta por parte de los babilonios pudo haberse motivado por la posibilidad de lograr expresar un mayor número de fracciones usando fracciones sexagesimales finitas. Lo anterior evidentemente prepara el terreno para comprender la necesidad de acortar expresiones infinitas por aproximación, introduciendo de manera natural la expresión aproximada. Y, en efecto, vimos que alrededor de unos 2000 años a.C. los babilonios ya habían producido una aproximación del valor de $\sqrt{2}$ expresado como fracción sexagesimal. Volvamos a ella.

La aproximación del valor de $\sqrt{2}$ dada por los babilonios se escribe en notación sexagesimal así

$$1; 24, 51, 10 = 1 \times 60^0 + 24 \times 60^{-1} + 51 \times 60^{-2} + 10 \times 60^{-3}.$$

Elevando al cuadrado, se obtiene

$$(1; 24, 51, 10)^2 = 1; 59, 59, 59, 38, 1, 40.$$

Punto de discusión

Convertir ésta en fracción decimal. ¿Qué tan precisa es la aproximación?

Los griegos siguieron esta actividad, produciendo por ejemplo métodos generales para aproximar raíces cuadradas.

Punto de discusión

Arquímedes utilizó el siguiente procedimiento de aproximación de la raíz cuadrada de un número A. Sea a la primera aproximación a $\sqrt{A}$, es decir, la parte entera. Para hallar x el próximo término de la aproximación usando la identidad $A = (x+a)^2 = a^2 + 2ax + x^2$, observamos que $2ax + x^2 = (2a + x)x$ debe ser menor que, pero próximo a, $A - a^2$. Entonces, dividimos $A - a^2$ por $2a$, teniendo en cuenta que no sólo $2ax$ sino $(2a + x)x$ ha de ser menor que $A - a^2$. De esta manera se encuentra, por ensayo y error, el mayor valor de x que sirve. Se encuentra un tercer término de la aproximación de la misma manera y así sucesivamente. Por ejemplo, Teón de Alejandría explica el método usado como sigue. Se quiere hallar $\sqrt{4500}$. Para ello se observa que la parte entera es $a = 67$ y ya que $67^2 = 4489$, se sigue que $A - a^2 = 4500 - 4489 = 11$. Ahora bien, suponiendo que la raíz cuadrada está expresada en fracciones sexagesimales, tenemos

$$\sqrt{4500} = 67 + \frac{x}{60} + \frac{y}{60^2},$$

donde todavía han de hallarse x, y. Ahora bien, x ha de ser tal que $\frac{2 \cdot 67x}{60}$ sea algo menos que 11, o sea, x debe ser menor que $\frac{11 \cdot 60}{2 \cdot 67} = \frac{330}{67} > 4$. En efecto, 4 sirve dando lugar a la aproximación $(67 + \frac{4}{60})$, y así sucesivamente.

Ejercicio

Utilizar este procedimiento para aproximar $\sqrt{3768}$, expresando su resultado como fracción sexagesimal.

Nota. Aunque se encuentra alejado de nuestra discusión es importante hacer notar que Arquímedes (278-212 a.C.) produjo una excelente aproximación para el valor del número irracional que nosotros denotamos π. El método que utilizó Arquímedes es conocido con el nombre del *método de exhaución*. Omitiendo los detalles, el planteamiento consiste en inscribir un polígono regular de 96 lados en un círculo y circunscribir un tal polígono al mismo círculo, calculando en ambos casos los perímetros de los polígonos y

argumentando que la longitud de la circunferencia es mayor que el perímetro del polígono inscrito y menor que el del circunscrito. Así Arquímedes logra la desigualdad

$$3\frac{10}{71} < \pi < 3\frac{10}{70}.$$

La aproximación es bastante buena dando $3.140845... < \pi < 3.142857...$

No se encuentra entre los manuscritos griegos una explicación para la aproximación de la raíz cúbica de un número, pero sin divulgar su método Herón da la aproximación $\sqrt[3]{100} \approx 4\frac{9}{14}$.

Recordemos que el primer encuentro que tuvimos con la aproximación de la raíz de un polinomio cúbico fue en la solución que dio Leonardo de Pisa a la ecuación $x^3 + 2x^2 + 10x = 20$. Nuevamente, Leonardo no divulga su método. La gran diferencia que se introduce en el trabajo de Leonardo es el reconocmiento de que la raíz no es construible con regla y compás, es decir, no es susceptible de una representación geométrica tradicional. El procedimiento de Leonardo se basa en la manipulación de la ecuación. Primero se factoriza 10 para obtener $10(x + \frac{1}{10}x^3 + \frac{1}{5}x^2) = 20$, o sea, $x + \frac{1}{10}x^3 + \frac{1}{5}x^2 = 2$. Esta última ecuación implica que $x < 2$. Pero poniendo $x = 1$, se obtiene de la primera que $1 + 2 + 10 = 13 < 20$, de donde, $x > 1$. Leonardo procede a afirmar que la raíz se encuentra entre 1 y 2. Aquí el punto importante es la aplicación implícita del Teorema del Valor Medio, lo cual supone a su vez la continuidad de la 'función' $x + \frac{1}{10}x^3 + \frac{1}{5}x^2$ (por supuesto sin contar con una adecuada noción de función, concepto que se desarrolla posteriormente, ni conocer el mencionado teorema cuyo enunciado versa sobre funciones continuas).

Auncuando se posea un método satisfactorio para obtener la solución de una ecuación polinómica, como es el caso de las ecuaciones cuadráticas, cúbicas y cuárticas, puede resultar que la expresión exacta de la raíz no sea suficiente para los propósitos del momento. Ya hemos visto, por ejemplo, que Rafael Bombelli propuso un método de aproximación de la raíz positiva de la ecuación $x^2 - 13 = 0$, método que en efecto corresponde a una expresión por fracciones continuas. Las sucesivas convergentes de la fracción proporcionan aproximaciones racionales cada vez mas acertadas de $\sqrt{13}$. Ahora bien, cuando estudiamos este procedimiento nuestro interés se centró en mostrar una creciente comprensión aritmética del sistema de los números reales, puesto que las fracciones continuas permiten generar un criterio que distingue entre números racionales e irracionales, por una parte, y por otra entre números irracionales que son producto de la extracción de una raíz cuadrada (todos construibles con regla y compás) y los demás números irracionales. Pero, de hecho, a veces el propósito de una tal aproximación es

simplemente asignar un valor numérico con el grado deseado de precisión a la solución.

5.2.2 Método de Horner

En secciones anteriores estudiamos algunas de las características del método que permiten evaluar un polinomio para cierto valor de la variable con gran eficiencia. En este aparte queremos capitalizar esta característica para generar un método numérico que permita resolver ecuaciones polinómicas por aproximación. Antes de analizarlo observaremos algunos algoritmos de Horner más elementales que nos permiten manipular ágilmente las funciones polinomiales (usando calculadora).

Consideremos la función polinomial

$$f(x) = x^4 - 3x^3 + 7x^2 - 15x + 1.$$

Nos interesa evaluar este polinomio en 2 y en 3, efectuando el menor número posible de operaciones. Si escribimos

$$f(x) = ((x-3)x + 7)x - 15)\,x + 1,$$

el proceso y el resultado de la evaluación se puede leer en la siguiente tabla

1	-3	7	-15	1
	2	-2	10	-10
1	-1	5	-5	-9

¡Dése cuenta que $f(2) = -9$!

Análogamente, para evaluarla en 3 tendríamos

1	-3	7	-15	1
	3	0	21	18
1	0	7	6	19

Tenemos que el polinomio

$$x^4 - 3x^3 + 7x^2 - 15x + 1 = (x^3 - x^2 + 5x - 5)(x - 2) + (-9)$$

$$x^4 - 3x^3 + 7x^2 - 15x + 1 = (x^3 + 7x + 6)(x - 3) + 19$$

Si analizamos las tablas anteriores observamos que en ellas aparecen los coeficientes de los polinomios $x^3 - x^2 + 5x - 5$ y $x^3 + 7x + 6$, respectivamente así como los residuos correspondientes $f(2)$ y $f(3)$. Estamos en condiciones

de afirmar que hay una raíz de este polinomio entre 2 y 3. ¿Por qué? Vamos a usar el algoritmo de Horner para aproximarnos a él. Observemos las siguientes tablas

1	-3	7	-15	1
	2	-2	10	-10
1	-1	5	-5	-9
	2	2	14	
1	1	7	9	
	2	6		
1	3	13		
	2			
1	5			

Lo cual nos permite escribir $f(x) = (x-2)^4 + 5(x-2)^3 + 13(x-2)^2 + 9(x-2) + (-9)$. Si hacemos $u = x - 2$, tenemos

$$f(x) = q(u) = u^4 + 5u^3 + 13u^2 + 9u + (-9).$$

Volvamos a aplicar el algoritmo de Horner a este polinomio evaluándolo en 0.6 y 0.5.

1	5	13	9	-9
	0.6	3.36	9.816	11.896
1	5.6	16.36	18.816	2.896

1	5	13	9	-9
	0.5	2.75	7.875	8.4375
1	5.5	15.75	16.875	-0.5625
	0.5	3.00	9.375	
1	6.0	18.75	26.25	
	0.5	3.25		
1	6.5	22		
	0.5			
1	7			

A este tercer polinomio $h(w) = w^4 + 7w^3 + 22w^2 + 26.25w - 0.5625$ lo evaluamos en 0.02 y en 0.03.

1	7	22	26.25	-0.5625
	0.02	0.1404	0.4428	0.5339
1	7.02	22.1404	26.6928	-0.0286

1	7	22	26.25	-0.5625
	0.03	0.2109	0.666327	0.8074
1	7.03	22.2109	26.916327	0.2449

Ahora bien, si estudiamos las tablas anteriores, de la primera pareja podemos afirmar que $q(u)$ tiene una raíz entre 0.5 y 0.6; y de la segunda que $h(w)$ tiene una raíz entre 0.02 y 0.03, de donde $f(x)$ tiene una raíz entre 2.52 y 2.53.

Analicemos ahora más sistemáticamente el método de Horner. Si nosotros ya hemos determinado que una raíz de la ecuación polinómica $f(x) = 0$ se encuentra localizada entre los enteros consecutivos a y $a+1$, para localizar la raíz entre décimas sucesivas transformamos la ecuación $f(x) = 0$ en una ecuación $g(x) = 0$ cuyas raíces son las de $f(x)$ disminuidas en a. Esto puede hacerse efectuando divisiones sintéticas sucesivas. Como $f(x)$ tiene una raíz r entre a y $a+1$, $g(x)$ tendrá una raíz $r-a$ entre 0 y 1. A continuación esta raíz puede ser ubicada entre décimas sucesivas. Supongamos que hemos localizado la raíz $r-a$ de $g(x)$ entre las décimas sucesivas $\frac{b}{10}$ y $\frac{b+1}{10}$. Podemos proceder entonces con $g(x)$ como lo hicimos con $f(x)$, obteniendo un polinomio $h(x)$ cuyas raíces son las de $g(x)$ disminuidas en $\frac{b}{10}$. El polinomio $h(x)$ tendrá entonces una raíz entre 0 y 0.1. Cuando localicemos esta raíz entre centésimas sucesivas, habremos pues localizado la raíz de $f(x)$ entre centésimas sucesivas.

Ejercicios

Para cada uno de los siguientes polinomios, hallar valores enteros consecutivos de x para los cuales los valores del polinomio difieren en signo y, de allí, usando el método de Horner, dar un cero de cada uno aproximado a dos cifras decimales.

1. $x^4 - x^3 - x^2 - x - 1$
2. $2x^3 - 9x^2 + 12x + 7$
3. $2x^6 - 7x^5 + x^4 + x^3 - 12x^2 - 5x + 1$

5.2.3 Métodos que utilizan conceptos del cálculo

Los resultados de Leonardo de Pisa indican con claridad la incidencia de algunos conceptos del cálculo en el planteamiento de un método aproximativo de solución de una ecuación polinómica. Antes de analizar los métodos de aproximación de las raíces de una ecuación polinomial es conveniente aclarar

que, en efecto, gran parte de dichos métodos requieren argumentos netamente analíticos acerca de las funciones polinomiales en contraposición a los métodos algebraicos o de corte geométrico que exploramos anteriormente; pero esto no significa que no sean de fundamental importancia no sólo como una enriquecedora oportunidad de trabajo con métodos numéricos (uso inteligente de la calculadora y el computador, y construcción de algoritmos) sino porque a través de ellos podemos establecer un nexo significativo entre álgebra y cálculo y aprovechar de paso la herramienta visual (gráfica en el plano cartesiano), como una alternativa de acercamiento intuitivo al problema de la solución de una ecuación polinomial.

Si consideramos ahora la función polinómica de grado n

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 + a_0,$$

donde a_i es real para todo i.

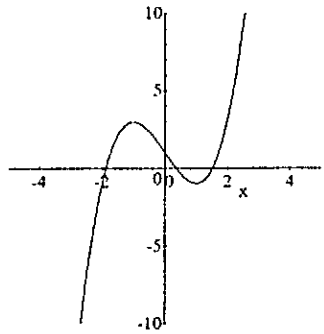
Recordemos aquí algunas características importantes que tiene la gráfica de esta función.

1. Para todo x real, $f(x)$ y $f'(x)$ son funciones continuas; la gráfica de f es una curva suave.

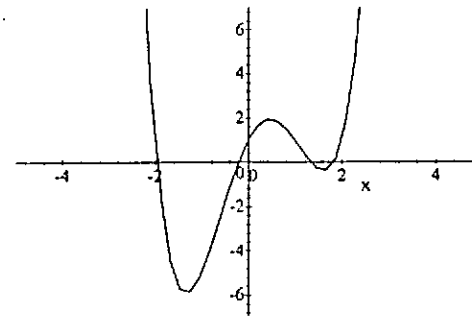
2. Para valores grandes de x, $|a_n x^n| > |a_{n-1} x^{n-1} + a_{n-2} x^{n-2} + \cdots + a_0|$. De lo anterior se sigue que si n es par y $a_n > 0$, la gráfica de f crece hacia infinito por ambas ramas (y si $a_n < 0$, decrece); pero si n es impar y $a_n > 0$, a la derecha crece y a la izquierda decrece (lo recíproco si $a_n < 0$).

3. Los puntos de intersección de la gráfica de f con el eje x, es decir, los puntos tales que $y = f(x) = 0$, corresponden a las raíces reales de la ecuación $f(x) = 0$. (Existen a lo más n de éstos.) Como sabemos, en el máximo y en el mínimo de la gráfica de $y = f(x)$, la derivada $f'(x) = 0$. Por consiguiente el número de máximos y mínimos es a lo más $n - 1$. De otra parte si $f''(x) > 0$, la primera derivada crece, es decir, la gráfica es cóncava hacia arriba; si $f''(x) < 0$, la gráfica es concava hacia abajo. Finalmente, como algunas de las raíces de $f'(x) = 0$ pueden ser complejas, el número de máximos y mínimos de

la gráfica puede ser menor que $n-1$.



Gráfica de $f(x) = x^3 - 3x + 1$



Gráfica de $f(x) = x^4 - x^3 - 4x^2 + 4x + 1$

$f(x) = x^3 - 3x + 1$ $\qquad$ $f(x) = x^4 - x^3 - 4x^2 + 4x + 1.$

Una vez construida la gráfica de un polinomio es fácil encontrar una aproximación a sus raíces.

5.2.4 Método del "exceso" y el "defecto"

Si analizamos la función polinomial f en el intervalo $[a, b]$, y determinamos que $f(a)$ y $f(b)$ tienen signos contrarios, podemos afirmar que entre a y b la polinomial $f(x)$ tiene al menos una raíz (aplicando el Teorema del Valor Intermedio). Considerando ahora un valor de r_1 suficientemente pequeño y tal que $a < a + r_1 < b$, si determinamos que $f(a)$ y $f(a + r_1)$ tienen signos contrarios podemos afirmar ahora que entre a y $a + r_1$, $f(x)$ tiene una raíz. Si continuamos con este proceso encontraremos una aproximación cada vez mejor de una raíz del polinomio.

El proceso de aproximación se reduce así a una aplicación reiterada del Teorema del Valor Intermedio. Es importante anotar que si $f(a)$ y $f(b)$ tienen signos contrarios, existe una raíz de la ecuación entre a y b, pero esto no significa que tenga una sola raíz en este intervalo. Además, si $f(a)$ y $f(b)$ tienen el mismo signo tampoco se puede decir que $f(x)$ no tiene raíces reales entre a y b.

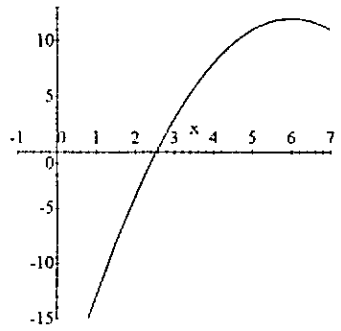
Puntos de discusión

1. Una ecuación polinomial de grado impar, tiene por lo menos una raíz *real*. (¿Por qué?) ¿Puede usted caracterizar más especificamente esta raíz usando el Teorema del Valor Intermedio?

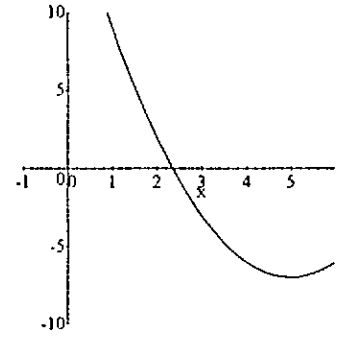
2. ¿Qué puede usted decir acerca de las raíces de una ecuación polinomial de grado par cuyo término independiente sea positivo? Para contestar explore ejemplos.

5.2.5 El método de tangentes y el método de cuerdas

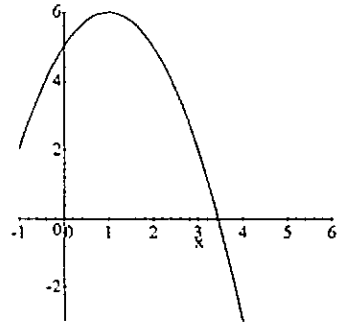
El método de tangentes, llamado el *método de Newton* y el método de cuerdas o de *interpolación lineal*, llamado también de *posición falsa*, son usados separada o simultáneamente para obtener un estimativo del error de una aproximación de una raíz de una ecuación polinómica. Supongamos que entre a y b el polinomio $f(x)$ tiene solamente una raíz, sabemos que $f(a)$ y $f(b)$ tienen signos opuestos y asumimos que $f''(x)$, no cambia de signo en este intervalo. Luego, la gráfica de la función en este intervalo tiene una de las cuatro formas siguientes.



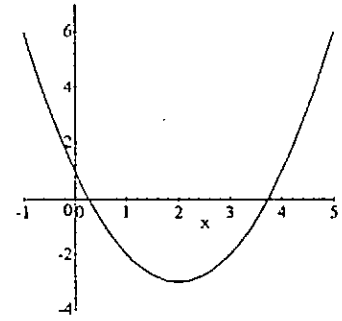
Caso I.



Caso II.



Caso III



Caso IV

En los casos I y II la tangente a la gráfica en el punto de abscisa a intercepta al eje x en el punto de abscisa α_1, que está entre la raíz buscada y

a. Si nosotros calculamos la abscisa α_1 y consideramos luego la tangente a la gráfica en el punto de abscisa α_1, encontraremos un punto α_2 que esta entre α_1 y la raíz deseada; continuando determinamos α_3 de manera análoga y así sucesivamente. De esta manera, identificaremos una mejor aproximación por *defecto* de la raíz buscada. En los casos III y IV es necesario trabajar con la tangente a la gráfica en el punto de abscisa b para obtener $\beta_1, \beta_2, \beta_3, \ldots$, es decir cada vez mejores y mejores aproximaciones por *exceso*.

Dado que la ecuación de la recta tangente a la curva $y = f(x)$ en el punto de abscisa a es

$$y - f(a) = f'(a)(x - a),$$

la abscisa α_1 del punto de intersección de esta tangente con el eje x se obtiene de la igualdad

$$0 - f(a) = f'(a)(\alpha_1) - a),$$

de donde, $\alpha_1 = a - \frac{f(a)}{f'(a)}$.

Análogamente, $\alpha_2 = \alpha_1 - \frac{f(\alpha_1)}{f'(\alpha_1)}, \alpha_3 = \alpha_2 - \frac{f(\alpha_2)}{f'(\alpha_2)}, \ldots$.

De manera similar,

$$\beta_1 = b - \frac{f(b)}{f'(b)}, \quad \beta_2 = \beta_1 - \frac{f(\beta_1)}{f'(\beta_1)}, \quad \beta_3 = \beta_2 - \frac{f(\beta_2)}{f'(\beta_2)}, \ldots.$$

Ejemplos

Consideremos la función polinomial

$$f(x) = x^4 - x^3 - x^2 - x - 1.$$

Usemos el método de Newton para aproximar una raíz de la ecuación polinomial asociada. Consideremos el intervalo cerrado $[1, 2]$. Tenemos que $f(1) = -3$ y $f(2) = 1$. Luego podemos garantizar que entre 1 y 3 hay una raíz. En este intervalo además $f''(x) > 0$; es por ello que el comportamiento de la curva corresponde al Caso IV. Trazamos la tangente a la curva en $(2, f(2))$ y así determinamos $\beta_1 = 1.93$. A continuación trazamos la tangente en $(1.93, f(1.93))$; obtenemos $\beta_2 = 1.9278$; luego $\beta_3 = 1.9275$, y así sucesivamente.

5.2.6 Método de interpolación lineal

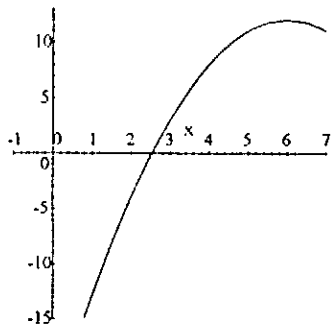
El método de interpolación lineal se basa en considerar la ecuación de una cuerda que pasa a través de los puntos dados $(a, f(a))$,$(b, f(b))$. Esta ecuación tiene la forma

$$\frac{x - a}{b - a} = \frac{y - f(a)}{f(b) - f(a)}.$$

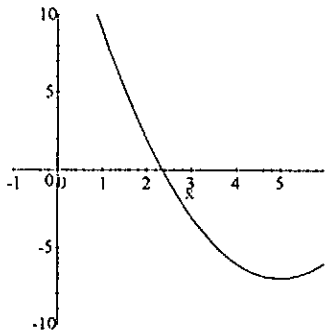
La abscisa γ_1 del punto de intersección de esta cuerda con el eje x se obtiene sustituyendo en la ecuación el punto de coordenadas $(\gamma_1, 0)$, de donde,

$$\gamma_1 = \frac{af(b) - bf(a)}{f(b) - f(a)}.$$

Tomando este valor como el nuevo b en los casos I y II.

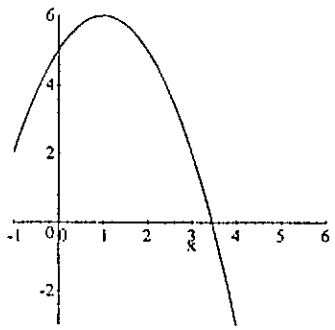


Caso I.

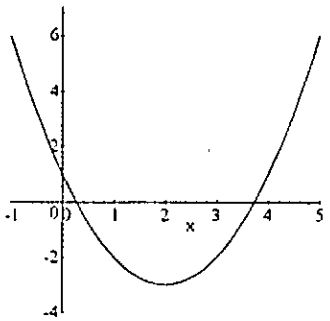


Caso II.

y como el nuevo a en los casos III y IV,



Caso III



Caso IV

se obtienen los siguientes resultados. Para I y II se tiene que

$$\gamma_2 = \frac{af(\gamma_1) - \gamma_1 f(a)}{f(\gamma_1) - f(a)}$$

y así sucesivamente.

En los casos III y IV, tomando γ_1 como el nuevo a, encontramos

$$\gamma_2 = \frac{\gamma_1 f(b) - bf(\gamma_1)}{f(b) - f(\gamma_1)}, \qquad \gamma_3 = \frac{\gamma_2 f(b) - bf(\gamma_2)}{f(b) - f(\gamma_2)},$$

y así sucesivamente.

Ejemplo

Usando el método de interpolación para la polinomial $f\ (x) = x^4 - x^3 - x^2 - x - 1$que analizamos con el método de Newton, tenemos $f(1) = -3$ y $f(2) = 1$ que corresponde al Caso IV. Entonces en este caso

$$\gamma_1 = \frac{af(b) - bf(a)}{f(b) - f(a)} = 1.75.$$

De manera similar determinamos $\gamma_2 = 1.91048$, $\gamma_3 = 1.92608, \ldots$.

Nota. La combinación de estos dos métodos nos permite realizar un buen estimativo del error, que desde luego no será mayor que la diferencia entre estas aproximaciones; por tanto la raíz buscada estará entre ellas.

5.2.7 Método de Lobacevskii

Uno de los métodos más usados para cálculo aproximado de raíces, especialmente cuando de raíces complejas se trata, es el método propuesto por N.I. Lobacevskii en su libro *Algebra* publicado en 1834. La idea básica de este método viene de Bernoulli. Observemos primero que si nos dan un polinomio cuyas raíces son $r_1, r_2, \ldots, r_n$ es fácil construir un polinomio también de grado n cuyas raíces sean $r_1^2, r_2^2, \ldots, r_n^2$, es decir, los cuadrados de las raíces del polinomio dado. Veamos. Si $r_1, r_2, \ldots, r_n$. son las raíces de la polinomial

$$x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_0,$$

ésta se puede ser factorizada en $\mathbf{C}$ como

$$(x - r_1)(x - r_2)(x - r_3)\cdots(x - r_n).$$

Y la polinomial

$$x^n - a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots \pm a_0,$$

cuyas raíces son las opuestas de las raíces de la polinomial dada, puede ser escrita en la forma

$$(x + r_1)(x + r_2)(x + r_3)\cdots + (x + r_n).$$

El producto de estas dos polinomiales es entonces

$$(x^2 - r_1^2)(x^2 - r_2^2)\cdots(x^2 - r_n^2),$$

el cual contiene solamente potencias pares de x. Si hacemos en el la sustitución $x^2 = y$, obtenemos una polinomial de grado n en y

$$y^n + b_{n-1}y^{n-1} + b_{n-2}y^{n-2} + \cdots + b_0,$$

que puede ser factorizada como

$$(y - r_1^2)(y - r_2^2)(y - r_3^2)\cdots(y - r_n^2)$$

y cuyas raíces son $r_1^2, r_2^2, r_3^2, \cdots, r_n^2$. Si multiplicamos directamente las polinomiales

$$x^n + a_{n-1}x^{n-1} + \cdots + a_0, \quad x^n - a_{n-1}x^{n-1} + a_{n-2}x^{n-2} - \cdots \pm a_0,$$

podremos obtener los coeficientes b_k. Después de obtener los coeficientes $b_{n-1}, b_{n-2}, b_{n-3}, \ldots, b_0$ de polinomial cuyas raíces son $r_1^2, r_2^2, \ldots, r_n^2$, estaremos en condiciones de construir los coeficientes $c_{n-1}, c_{n-2}, c_{n-3}, \ldots, c_0$ de la polinomial cuyas raíces son los cuadrados de las raíces de

$$y^n + b_1y^{n-1} + \cdots + b_n,$$

es decir $r_1^4, r_2^4, \cdots, r_n^4$. Análogamente podemos obtener los coeficientes $d_{n-1}, d_{n-2}, \ldots, d_0$ de la polinomial cuyas raíces son $r_1^8, r_2^8, \ldots, r_n^8$, y así sucesivamente.

Para llegar a la idea fundamental del método consideremos por simplicidad el caso en que todas las raíces de la ecuación son reales y distintas en valor absoluto. Asumimos

$$|r_1| > |r_2| > |r_3| > \cdots > |r_n|.$$

Sea N suficientemente grande y consideremos la polinomial

$$x^N + A_{N-1}x^{N-1} + A_{N-2}x^{N-2} + \cdots + A_0$$

que tiene como raíces las N-ésimas potencias de las raíces $r_1, r_2, \ldots, r_n$ del polinomio dado. Tenemos entonces

$$\begin{aligned} -A_{N-1} &= r_1^N + r_2^N + \cdots r_n^N \\ A_{N-2} &= r_1^N r_2^N + r_1^N r_3^N + \cdots + x_{n-1}^N x_n^N \\ &\vdots \\ \pm A_0 &= r_1^N r_2^N \cdots r_n^N. \end{aligned} \tag{5.1}$$

Teniendo en cuenta que la sucesión de números $|r_1^N|, |r_2^N|, \cdots, |r_n^N|$, para N suficientemente grande, es tal que $|r_1^N| > |r_2^N| > \cdots > |r_n^N|$, en las expresiones para los A_i bastará quedarnos con el primer sumando, puesto que los restantes

son despreciables comparados con el primero. De esta forma obtenemos las fórmulas aproximadas

$$r_1^N \approx -A_{N-1}, \quad r_1^N r_2^N \approx A_{N-2}, \quad r_1^N r_2^N r_3^N \approx -A_{N-3} \cdots r_1^N r_2^N \cdots r_n^N \approx \pm A_0,$$

de donde,

$$r_1 \approx \sqrt[N]{-A_{N-1}}, r_2 \approx \sqrt[N]{-\frac{A_{N-2}}{A_{N-1}}}, r_3 \approx \sqrt[N]{-\frac{A_{N-3}}{A_{N-2}}}, \cdots, r_n \approx \sqrt[N]{-\frac{A_0}{A_1}}.$$

Ejercicios

Usar el método de Lobacevskii para obtener estimativos del valor absoluto de los ceros de los siguientes polinomios

1. $x^2 - x + 1$
2. $x^5 - x^4 - x^3 + 4x^2 - x - 1$
3. $x^{11} + x^8 - 3x^5 + x^4 + x^3 - 2x^2 + x - 2$
4. $3x^2 - x + 4$

5.2.8 Métodos numéricos en el siglo XX

El gran cambio introducido en los procedimientos contemporáneos de aproximación ha sido la posibilidad de efectuar todos los cálculos usando una calculadora o plasmar el método en un algoritmo o programa de computador que lo efectúe con gran rapidez y precisión. Es por todos sabido que aun la calculadora mas común tiene la extracción de raíz cuadrada incorporada, mientras que las calculadoras graficadoras tienen posibilidad de evaluar y producir la gráfica de funciones polinómicas y de algunas trascendentes. Las calculadoras usan algoritmos de evaluación de las funciones que son, en efecto, métodos numéricos. En ellos una de las herramientas mas poderosas será la recursión. En lo que sigue analizaremos algunos algoritmos o programas de computador que efectúan procedimientos aproximativos para evaluar una función polinómica, la raíz cuadrada de un número y la raíz de un polinómio.

Evaluación de la raíz cuadrada Supongamos que queremos encontrar la raíz cuadrada de $a > 1$, es decir, queremos encontrar $z > 1$ tal que

$$z * z = a. \tag{5.2}$$

La técnica que usaremos reemplaza z por variables x, y tales que

$$x > y \quad \text{y} \quad x * y = a \tag{5.3}$$

y luego hace que los valores de x, y tiendan hacia z manteniendo (*.9) invariante. Podemos inicializar en todo caso con $x = a, y = 1$. Entonces se quiere mantener (*.9) válido al tiempo que la diferencia $x - y$ decrece. Nos sirve la sustitución $(x, y) \longleftarrow (x', y')$ tal que

$$x' = \frac{x+y}{2} \quad \text{y} \quad y' = \frac{2xy}{x+y}$$

donde x' es la media aritmética y y' la media armónica de los dos números. Es bien sabido que la media aritmética es mayor que la media armónica $(x' > y')$ y se tiene además

$$a = x * y = \frac{x+y}{2} * \frac{2xy}{x+y} = x' * y'.$$

Ahora bien

$$0 < x' - y' = \frac{x+y}{2} - \frac{2xy}{x+y} = \frac{(x-y)^2}{2(x+y)} = \frac{x-y}{2} \cdot \frac{x-y}{x+y} < \frac{x-y}{2},$$

lo cual implica que un sólo paso reduce la diferencia al menos por la mitad. Resulta que el procedimiento es muy rápido y permite casi duplicar el número de dígitos correctos en la raíz en cada paso pues si

$$0 < x - y < 2^{-n}$$

entonces

$$0 < x' - y' < \frac{2^{-n}}{2(x+y)} < \frac{2^{-n}}{4y}.$$

Nuestra discusión permite generar el programa en *Pascal*

```
program raiz;{initially x>y}
var x,y : real;
function r(x,y : real) : real;
begin
            if x=y then r:=x
            else r:=r((x+y)/2,2*x*y/(x+y))
end;
begin
write ('x,y='); readln(x,y); writeln (r(x,y)); readln
end
```

Punto de discusión

Usando una calculadora seguir los pasos del anterior algoritmo para hallar $\sqrt{3}$, es decir, para $x = 3, y = 1$.

Desarrollemos ahora un programa iterativo para hallar la raíz cuadrada. Este tiene la ventaja sobre el programa anterior de que puede usarse para cualquier $a > 0$. En el se reemplaza la sustitución simultánea $(x, y) \longleftarrow (x', y')$ por las dos sustituciones sucesivas

$$x \longleftarrow \frac{x+y}{2}; \quad y \longleftarrow \frac{a}{x}.$$

```
program raiz1;{ x>y}
   var x,y,eps : real;
   function r(x,y : real) : real;
   begin
                writeln(x:20,y:20);
                if abs(x-y)/x<eps then r:=(x+y)/2
                else r:=r((x+y)/2,2*x*y/(x+y))
   end;
   begin
   write ('x,y,eps='); read(eps,x,y); x:= r(x,y); readln
   end
```

Si ponemos $x = 2, y = 1, eps = 1E - 07$ el programa anterior genera el siguiente 'output'.

2.0000000000	1.0000000000
1.5000000000	1.3333333333
1.4166666667	1.4117647059
1.4142156863	1.4142114385
1.4141135624	1.4142135624

Ahora bien, éste es el conocido método de extracción de la raíz cuadrada.

Puntos de discusión

1. Comparar este último algoritmo con el método empleado por Arquímedes para aproximar la raíz cuadrada y con la aproximación por fracciones continuas de Bombelli. ¿Hay alguna diferencia en los tres métodos?

2. Este es un ejemplo particular del método de Newton para aproximar los ceros de una función. ¿Puede usted decir por qué?

En el siguiente conjunto de puntos de investigación, exploramos la solución de la ecuación cuadrática por iteración.

Puntos de investigación

1. Para resolver la ecuación $x^2 - 4x - 1 = 0$ por iteración, despejamos $x^2 = 4x + 1$ o $x = 4 + \frac{1}{x}$ lo cual nos lleva a usar la recurrencia

$$x_1 = 4, \quad x_{n+1} = 4 + \frac{1}{x_n}.$$

Mostrar que x_n converge a una raíz de la ecuación y que, en cada paso, el error se reduce en un factor de 18.

2. Para resolver $x^2 = 10$ por iteración, observamos que al depejar x obtenemos $x = \frac{10}{x}$ lo cual no lleva a ninguna parte ya que es periódico de período 2. Así las cosas, ponemos $x = z - 3$, reemplazamos en la ecuación original y despejamos z para obtener $z = 6 + \frac{1}{z}$, lo que permite plantear una recurrencia similar a la del punto anterior. ¿Por qué se toma $x = z - 3$? Plantear la recurrencia y estudiar la convergencia.

3. En el caso de la cuadrática $x^2 - 2x + 1 = 0$, si la transformamos en $x = 2 - \frac{1}{x}$ y ponemos

$$x_1 = 2, \quad x_{n+1} = 2 - \frac{1}{x_n},$$

mostrar, por inducción que $x_n = 1 + \frac{1}{n}$ con una lenta convergencia hacia 1. ¿Cuántas raíces distintas tiene la ecuación original?

4. Comprobar que un intento por usar un procedimiento similar a los anteriores para las ecuaciones

$$x^2 - 2x + 2 = 0 \quad x^2 - 3x + 3 = 0$$

lleva a sucesiones periódicas de períodos 4 y 6, respectivamente. Hallar los valores que se repiten en cada una de estas sucesiones (usar la convención $\frac{\text{diferente de } 0}{\infty} = 0$ y $\frac{\text{diferente de } 0}{0} = \infty$). Analizar y discutir el problema de la convergencia para ecuaciones con raíces complejas.

5. Considerar la ecuación $x^2 - px + q = 0$ y la sucesión asociada

$$x_1 = p, \quad x_{n+1} = p - \frac{q}{x_n}. \tag{5.4}$$

   (a) ¿Bajo qué condiciones converge la sucesión (*)?

   (b) Si converge, ¿hacia cuál de las raíces de la ecuación original converge?

   (c) ¿Con qué rata de convergencia lo hace?

Programas de computador y software potente Es importante tener en cuenta exactamente qué pueden hacer programas como *Mathematica* o *Maple* en la solución de ecuaciones polinómicas y qué no pueden hacer. Toda la discusión que sigue se basa en cómo responde un programa de *Maple* para estudiantes cuando se da el comando de resolver una determinada ecuación. El programa ofrece dos opciones: solución exacta y solución numérica. Primero resumiremos algunos resultados específicos y luego reflexionaremos sobre la potencialidad del programa.

Ejemplos

1. Ecuación: $x^2-15x+7=0$. Comando - solución exacta: $\left\{x=\frac{15}{2}+\frac{1}{2}\sqrt{197}\right\}$, $\left\{x=\frac{15}{2}-\frac{1}{2}\sqrt{197}\right\}$. Comando - solución numérica: $\{x=.48217\}$, $\{x=14.518\}$

2. Ecuación: $x^2+4x+11=0$. Solución exacta: $\left\{x=-2+i\sqrt{7}\right\}$, $\left\{x=-2-i\sqrt{7}\right\}$. Solución numérica: No responde.

3. Ecuación: $x^3+7x^2-x+31=0$. Solución exacta: $\left\{x=-\sqrt[3]{\left(\frac{793}{27}+\frac{13}{9}\sqrt{321}\right)}-\frac{52}{9\sqrt[3]{\left(\frac{793}{27}+\frac{13}{9}\sqrt{321}\right)}}\right.$

$$\left\{x=\frac{1}{2}\sqrt[3]{\left(\frac{793}{27}+\frac{13}{9}\sqrt{321}\right)}+\frac{26}{9\sqrt[3]{\left(\frac{793}{27}+\frac{13}{9}\sqrt{321}\right)}}-\frac{7}{3}+\frac{1}{2}i\sqrt{3}\left(-\sqrt[3]{\left(\frac{793}{27}+\frac{13}{9}\sqrt{321}\right)}+\frac{52}{9\sqrt[3]{\left(\frac{793}{27}+\frac{13}{9}\sqrt{321}\right)}}\right.\right.$$

$$\left\{x=\frac{1}{2}\sqrt[3]{\left(\frac{793}{27}+\frac{13}{9}\sqrt{321}\right)}+\frac{26}{9\sqrt[3]{\left(\frac{793}{27}+\frac{13}{9}\sqrt{321}\right)}}-\frac{7}{3}-\frac{1}{2}i\sqrt{3}\left(-\sqrt[3]{\left(\frac{793}{27}+\frac{13}{9}\sqrt{321}\right)}+\frac{52}{9\sqrt[3]{\left(\frac{793}{27}+\frac{13}{9}\sqrt{321}\right)}}\right.\right.$$

1. Solución numérica: $\{x=-7.659\}$.

4. Ecuación: x^4-2x^3-5x+2. Solución exacta: $\left\{x=-\frac{1}{2}+\frac{1}{2}i\sqrt{7}\right\}$, $\left\{x=-\frac{1}{2}-\frac{1}{2}i\sqrt{7}\right\}$, $\left\{x=\frac{3}{2}+\frac{1}{2}\sqrt{5}\right\}$, $\left\{x=\frac{3}{2}-\frac{1}{2}\sqrt{5}\right\}$.
Solución numérica: $\{x=.38197\}$, $\{x=2.618\}$.

5. Ecuación: $x^4+2x^3+18x^2+5x+2$. Solución exacta: $\{x=\rho\}$ where ρ is a root of $Z^4+2Z^3+18Z^2+5Z+2$. Solución numérica: No responde.

6. Ecuación: $x^4-\frac{5017}{11753}x^3+\frac{11228622}{108962063}x^2-\frac{13048747}{108962063}x-\frac{12749}{64897}=0$. Solución exacta: $\left\{x=-\frac{253}{511}\right\}$, $\left\{x=\frac{19}{23}\right\}$, $\left\{x=\frac{7}{146}+\frac{1}{18542}i\sqrt{164344731}\right\}$, $\left\{x=\frac{7}{146}-\frac{1}{18542}i\sqrt{164344731}\right\}$. Solución numérica: $\{x=-.49511\}$, $\{x=.82609\}$.

7. Ecuación: x^5-19x^2+23. Solución exacta: $\{x=\rho\}$ where ρ is a root of Z^5-19Z^2+23. Solución numérica: $\{x=-1.0667\}$, $\{x=1.1467\}$, $\{x=2.4805\}$.

8. Ecuación: $x^6 + 13x^4 - 19x^3 + 5x^2 - 4x + 23 = 0$. Solución exacta: $\{x = \rho\}$donde ρ es raíz de $Z^6 + 13Z^4 - 19Z^3 + 5Z^2 - 4Z + 23 = 0$. Solución numérica: No responde.

Nótese de las ecuaciones (1) y (2) que el programa *Maple* aplica perfectamente la fórmula de solución para la ecuación cuadrática y que conoce métodos, probablemente similares a los que hemos mirado, para asignar un valor aproximado a las raíces reales. Igualmente, no responde cuando se pide un valor numérico para las raíces complejas. Nótese, además, de la ecuación (3) que el programa conoce la fórmula de Cardano para la solución de la ecuación cúbica. Las ecuaciones (4), (5) y (6) nos confirman que, dada una ecuación de grado cuatro que es factorizable sobre enteros o racionales, el programa es capaz de factorizarla completamente en factores lineales o cuadráticos y dar las respectivas soluciones. Por otra parte, es claro que el programa no domina la solución general de la cuártica y, por ende, no puede encontrar las raíces complejas de una cuártica que es irreducible.

Puntos de discusión

1. Estudiar las ecuaciones (7) y (8) y analizar lo que nos comunica sobre la potencia del programa *Maple* para resolver ecuaciones polinomiales de grados cinco o seis.

2. A partir de los ejemplos anteriores, ¿cuáles otras conclusiones puede hacer Ud. con respecto de la capacidad del programa de resolver ecuaciones polinomiales?

5.3 Estatus teórico de los métodos numéricos

No sólo es cierto que, dada la gran facilidad de encontrar soluciones aproximadas a las ecuaciones polinómicas usando calculadora, programas de computador o paquetes como Mathematica o Maple, los métodos numéricos se han vuelto accesibles a todo el mundo, sino de mayor importancia aun, el significado de tales procedimientos a nivel teórico indica un último y fundamental cambio de énfasis en la solución de ecuaciones.

Si bien es cierto que hemos trazado los enfoques desde el geométrico pasando por el algebraico y llegando al analítico, hay razones de peso para tomar los métodos numéricos como los mas generales. Pero no hay que perder de vista que éstos sólo son aplicables cuando la ecuación tiene raíces reales y para aproximar éstas.

Como tendremos oportunidad de estudiar mas adelante, el siglo XIX vio el nacimiento de una teoría general del álgebra, o una álgebra abstracta,

cuyos intereses se centran mas en las estructuras algebraicas (grupo, anillo, campo,...) que en la teoría de ecuaciones. Entre otras razones, esto se debe a una teoría conocida como la Teoría de Galois que demuestra, basándose en resultados pertinentes a las estructuras algebraicas, que en general no existen métodos algebraicos para resolver ecuaciones de grado 5 o mayor. Otra manera de decir lo mismo es que no existe ninguna fórmula de solución para dichas ecuaciones que, tal como en el caso de las ecuaciones cuadráticas, cúbicas y cuárticas, puede expresarse en términos de operaciones aritméticas (adición, sustracción, multiplicación, división, extracción de raíces) con los coeficientes de la ecuación.

Si bien Leonardo de Pisa mostró lo inadecuados que son los métodos geométricos (pregonados por los matemáticos griegos y árabes) para tratar el problema de solución de una ecuación polinómica, Galois demostró también las deficiencias de los métodos algebraicos (pregonados por los algebristas italianos). *El problema de resolver una ecuación polinómica no goza en general de solución.* Esto lleva al álgebra a centrarse en otros puntos, en particular, en las estructuras algebraicas que le dan nueva vida y creciente importancia dentro de la matemática vista como un todo. Por ende, después de culminar esta discusión de la teoría de ecuaciones, proseguiremos nuestro estudio centrado en estas nuevas direcciones de desarrollo algebraico.

Por otra parte, coincide esta metamorfosis del álgebra con la aritmetización de la matemática, proceso que se lleva a cabo a lo largo de la segunda mitad del siglo XIX. Parte de este proceso es la aritmetización del cálculo que sustituye aproximaciones geométricas e intuitivas a conceptos básicos del análisis, como son límite y continuidad, por ejemplo, por definiciones aritméticas rigurosas como son las definiciones $\delta - \epsilon$ de Weierstrass. (La aritmetización de la matemática a su vez induce a Georg Cantor a llevar el proceso otro paso, fundamentando la aritmética en la teoría de conjuntos, y a otros matemáticos, como Gottlob Frege y Bertrand Russell, a dar otro paso hacia atrás fundamentando la teoría de conjuntos en la lógica.)

Punto de investigación

Investigar acerca de este proceso de fundamentación de la matemática, qué lo motivó, las líneas de fundamentación que se desarrollaron, así como sus alcances y fracasos.

Así las cosas, debemos mantener presente que el presente siglo no sólo ha dado herramientas muy eficaces que posibiliten la solución numérica de ecuaciones algebraicas y mucho otros problemas centrales de la matemática, sino que la calculadora y el computador reflejan directamente la importancia de nueva fundamentación de la matemática sobre la base de la aritmética.

5.4 Problemas del capítulo

1. Si $x^4+4x^3+6px^2+4qx+r$ es divisible por x^3+3x^2+9x+3 (es decir, la división deja residuo 0), hallar $(p+q)r$.

2. Si se divide y^2+my+2 por $y-1$ el cociente es $f(y)$ y el residuo es r. Si se divide y^2+my+2 por $y+1$ el cociente es $g(y)$ y el residuo es r. Hallar m.

3. Si $3x^3-9x^2+kx-12$ es divisible por $x-3$, hallar k.

4. El método de bisección para la aproximación de raíces. Sea $f(x)$ cualquier polinomio con coeficientes reales. Supóngase además que $f(a)$ es positivo y $f(b)$ negativo. La gráfica de $f(x)$ es una curva continua que une el punto $(a, f(a))$ que está por encima del eje x con el punto $(b, f(b))$ que está por debajo. Por ende, cruzará el eje en alguno punto intermedio entre a y b.

   (a) Demostrar que, o bien $\frac{a+b}{2}$ es un cero de $f(x)$ o bien $f(\frac{a+b}{2})$ difiere en signo de $f(a)$ o de $f(b)$.

   (b) Sea n un entero positivo. Dar un algoritmo que produce un intervalo de longitud menor o igual a $(b-1)^{-2n}$ que contiene un cero de $f(x)$

5. Usar el método de Horner y el de Newton para aproximar un cero del polinomio $2x^6-7x^5+x^4+x^3-12x^2-5x+1$ entre $x=3$ y $x=4$.

   (a) Sea $c>0$. Demostrar que el método de Newton aplicado al polinomio x^2-c da, a partir de una aproximación positiva a_1 de $\sqrt{c}$, una sucesión $\{a_n\}$ con

   $$a_{n+1}=\frac{1}{2}\left(a_n+\frac{c}{a_n}\right), n\geq 1.$$

   (b) Hallar una expresión que relaciona la diferencia a_{n+1}^2-c con la diferencia a_n^2-c, y mostrar que, cuando n crece, a_n se aproxima cada vez mas a $\sqrt{c}$. Mostrar, además, que $\{a_n\}$ es decreciente para $n\geq 2$.

6. Determinar todos los ceros reales del polinomio $3x^4-2x^3-x^2-3x+1$, aproximados a tres cifras decimales, usando el método de Horner, el de Newton y el de interpolación lineal.

7. Considerar la ecuación polinómica $x^3-4x-18=0$.

(a) Mostrar que hay una raíz positiva única r entre 3 y 4.

(b) La ecuación puede volver a escribirse en la forma $x = g(x)$ donde $2g(x) = x^3 - 18$. Comenzando por un valor de u_1 (digamos 3) y definiendo una sucesión $u_n = g(u_{n-1})$ para $n \geq 2$, determinar si u_n se aproxima a una solución de la ecuación dada.

8. Demostrar que el polinomio $6x^8 + 5x^6 + 12x^4 + 2x^2 + 1$ no tiene ceros reales.

9. Demostrar que un polinomio no trivial cuyos coeficientes no nulos son todos reales positivos no puede tener ceros reales no negativos.

10. Supongamos que $f(x)$ es un polinomio de grado n con coeficientes reales y tal que el coeficiente de x^n es positivo. Si $f(k) < 0$ para algún número real k, demostrar que $f(x)$ tiene una raíz real que es mayor que k.

11. Demostrar que el polinomio lineal con coeficientes reales $ax + b (ab \neq 0)$ tiene un cero positivo si y sólo si los coeficientes a, b tienen signos opuestos.

12. Sea $f(x)$ un polinomio con coeficientes reales. Demostrar que r es raíz de $f(x) = 0$ si y sólo si $-r$ es raíz de $f(-x) = 0$.

13. Localizar intervalos que contienen ceros reales de los siguientes polinomios y hallar cuánta información pueda acerca de estos ceros.

(a) $x^5 - 3x^4 - x^2 - 4x + 14$

(b) $24x^5 + 143x^4 - 136x^3 + 281x^2 + 36x - 140$

(c) $2x^4 + 5x^3 + x^2 + 5x + 2$

(d) $16x^6 + 3x^4 - 3x^3 - 142x^2 - 9x - 21$

(e) $16x^7 - 5x^5 - 97x^4 - 95x^3 - 79x^2 + 36$

Capítulo 5

Teoría de grupos

Nuestro propósito en este capítulo es discutir aspectos fundamentales de la teoría de grupos y analizar como elementos y argumentos muy familiares de la aritmética que se trabajan en los niveles elementales se presentan continuamente al interior de esta teoría.

Las ideas preliminares de la *teoría de grupos* (y en general las del *álgebra abstracta*) pueden ser construidas a partir de conceptos elementales, por ejemplo, los énfasis iniciales en el estudio de los números y sus propiedades, sus operaciones (adición y multiplicación de naturales, enteros, reales...). Las relaciones y los argumentos y elementos presentes en este análisis pueden ser aprovechados para derivar propiedades de sistemas algebraicos (objeto central del álgebra moderna); es importante destacar que en contraste con la aritmética y el álgebra elemental, el álgebra moderna (en general) y la teoría de grupos (en particular) estudian operaciones sobre objetos que no necesariamente son números pero que satisfacen ciertas propiedades (análogas a las consideradas elementales en la matemática escolar).

5.1 Ley de composición interna

Precisamente las llamadas operaciones elementales, que se trabajan ampliamente en los niveles básicos: la adición y multiplicación en naturales, enteros, racionales, reales constituyen ejemplos y modelos para abordar el primer objeto de estudio fundamental en el álgebra moderna: las *leyes de composición interna u operaciones binarias*, que se definen a continuación.

Definición 5.1 *Una ley de composición interna u operación binaria sobre un conjunto E es una función de $E \times E$ en E. Es decir es una relación que a un par de elementos de E le hace corresponder uno y sólo un elemento de E.*

Los símbolos mas usados para notar la composición son $+$ y $\cdot$, si $a, b \in E$, el compuesto de a y b se notará usualmente $a + b$, $a \cdot b$ o simplemente ab.

Analizaremos algunos ejemplos en los que es posible determinar si la ley definida es o no de composición interna. Usaremos algunos de ellos en nuestra discusión posterior, desde luego el lector puede complementar con múltiples ejemplos adicionales.

Ejemplos

1. Suma o producto, en los naturales, en los enteros, en los racionales, en los reales..etc.

2.Consideremos ahora el conjunto de los enteros impares $I = \{x \in \mathbb{Z} | x = 2k+1 | k \in \mathbb{Z}\}$, y en I la suma usual de enteros. Obsérvese que si $x = 2k+1$ y $y = 2m+1$ con k y m enteros. $x + y = 2(k+m+1)$, ¡esto significa que la suma no es una ley interna sobre este conjunto! ¿Qué ocurre con el producto?

3. Para cada entero positivo m consideremos el conjunto

$$\mathbb{Z}_m = \{0, 1, 2 \cdots m-1\}.$$

Definimos dos importantes leyes de composición que retomaremos continuamente en nuestro análisis, notadas (si no da lugar a confusión) $+$ y $\cdot$ y llamadas adición módulo m y multiplicación módulo m, de la siguiente forma.

Si $x, y \in \mathbb{Z}_m$, $x + y$ es el residuo de la división de $x + y$ (suma usual en enteros) por m y $x \cdot y$ es el residuo de la división de $x \cdot y$ (producto usual entre enteros) por m.

Las tablas de la adición y multiplicación módulo 3 se ilustran a continuación.

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\cdot$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Punto de discusión

Construir las tablas para adición y multiplicación módulos 4 y 6 y contrastar con las anteriores.

4. Sea ahora $B = \{\frac{p}{q} | p, q \in \mathbb{Z}, q = 1, 2, 4\}$; considere en B la adición usual de racionales. Obsérvese que los divisores positivos de 4 son precisamente 1, 2 y 4.

Punto de discusión

¿Es la adición una ley de composición interna en el conjunto B ? ¿Puede usted construir un resultado general si q varía en el conjunto de divisores positivos de un entero arbitrario d ?

5. En el conjunto $C = \{1, -1, i, -i\}$ consideremos el producto usual de complejos (i es la unidad imaginaria, $i^2 = -1$). Nótese que

$$C = \{x \in \mathbb{C} | x^4 = 1\},$$

esto es, el conjunto C está constituido precisamente por las raíces cuartas de la unidad.

De manera mas general podemos considerar en $U = \{x \in \mathbb{C} | x^n = 1, n \in \mathbb{N}, n \geq 2\}$, el producto usual de complejos. Nótese que $U = \left\{ e^{\frac{2k\pi}{n}} | k = 0, 1, 2, \ldots n-1 \right\}$, es decir, sus elementos son las raíces n-ésimas de la unidad.

6. Sea § el conjunto de las simetrías del cuadrado; § contiene ocho miembros, cuatro rotaciones r_0, r_1, r_2 y r_3 en el plano del cuadrado y en el sentido

contrario de las manecillas del reloj, de 0°, 90°, 180° y 270°; dos rotaciones h y v en el espacio alrededor de los ejes de simetría horizontal y vertical del cuadrado y dos rotaciones d_1 y d_2 en el espacio alrededor de las diagonales del cuadrado. (Las cuatro últimas se consideran usualmente como reflexiones respecto a los cuatro ejes de simetría.) Consideremos en § la composición usual de funciones, esto es, si x y y son elementos de §, $x \circ y$ es el resultado de aplicar primero la simetría x y a continuación la simetría y y constituye otro ejemplo de ley de composición interna. Observemos el compuesto de algunos elementos de §, marcamos para ello los vértices del cuadrado como se ilustra en la Figura 5.1.

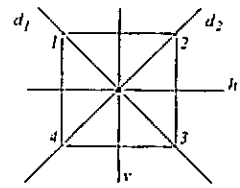


Fig 5.1

Aplicando r_3 a continuación de v esto es efectuando $r_3 \circ v$, tendríamos Figura 5.2.

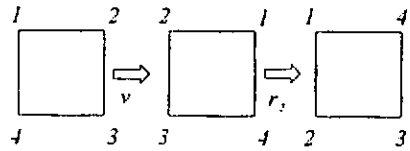


Fig 5.2

y

Por tanto $r_3 \circ v = d_2$. De manera similar, aplicando r_3 y a continuación v tendremos (Figura 5.3),

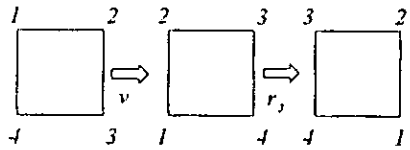


Fig 5.3

de donde, $v \circ r_3 = d_1$.

Punto de discusión

¡Completar la tabla de esta ley de composición y observar en ella regularidades! Construir ahora la tabla de composición para las simetrías de un triángulo equilátero; compararla con las simetrías de un rectángulo (no cuadrado).

7. En $Q = \{1, i, j, k, -1, -i, -j, -k\}$ se define una ley interna por medio de las siguientes condiciones.

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

8. Si consideramos ahora $x, y \in \mathbb{R}$ y la ley definida por

$$x * y = x + y - 3,$$

por propiedades de la suma entre reales $*$ es claramente una ley interna sobre $\mathbb{R}$. Exploremos con ella otras ideas. Sean $x, y, z \in \mathbb{R}$

$$\begin{aligned}(x * y) * z &= (x + y - 3) * z \\ &= (x + y - 3) + z - 3 \\ &= x + (y + z - 3) - 3 \\ &= x + (y * z) - 3 = x * (y * z)\end{aligned}$$

De otra parte, observamos que

$$\begin{aligned}y * x &= y + x - 3 \\ &= x + y - 3 \\ &= x * y\end{aligned}$$

Punto de discusión

¡Considerar sobre $\mathbb{R}^*$ la ley

$$x \triangle y = \frac{x \cdot y}{x + y}$$

y explorar las propiedades que acabamos de observar en $*$!

9. La tabla que aparece a continuación define una ley interna $\diamond$ sobre el subconjunto de enteros $\{0, 1, 2\}$.

$\diamond$	0	1	2
0	0	1	2
1	1	0	0
2	2	0	0

Obsérvese que en este caso

$$\begin{aligned}(1 \diamond 1) \diamond 2 &= 0 \diamond 2 \\ &= 2\end{aligned}$$

y

$$\begin{aligned}1 \diamond (1 \diamond 2) &= 1 \diamond 0 \\ &= 1.\end{aligned}$$

Si analizamos ahora detenidamente cada una de las leyes de composición exploradas hasta aquí, podemos notar en algunas de ellas regularidades que se expresan en las definiciones que se presentan a continuación.

Definición 5.2 *Sea $\circ$ una ley de composición interna (o una operación binaria) sobre un conjunto E. Esta ley es asociativa si $\forall a, b, c \in E$.*

$$(a \circ b) \circ c = a \circ (b \circ c).$$

Es claro que, por ejemplo, la adición y la multiplicación definidas en el conjunto de los números enteros, racionales, reales o complejos son asociativas, así como las leyes definidas en 3, 4, 5, 6, 7 y 8. La observación que hicimos respecto a la ley definida en 9 nos muestra que ésta no es asociativa.

Punto de discusión

¡Construir otros ejemplos de leyes internas asociativas y no asociativas, entre ellas estudiar el producto vectorial (producto cruz) de vectores en el espacio!

Definición 5.3 *La ley de composición $\circ$ sobre E es conmutativa si* $\forall a, b \in E$

$$a \circ b = b \circ a.$$

Nótese que la composición definida sobre el conjunto de simetrías del cuadrado es no conmutativa, pero las leyes definidas en 8 sí lo son.

Punto de discusión

¡Analizar esta idea en todas las leyes presentadas hasta el momento y construir nuevos ejemplos de leyes conmutativas y no conmutativas; explorar entre ellas el producto usual de matrices definido sobre el conjunto de matrices 2×2 con elementos reales o complejos y de nuevo el producto vectorial!

Es importante anotar en este momento que si E es un conjunto finito (con n elementos) una ley de composición $\circ$ sobre E puede ser completamente descrita por una tabla. Esta tabla es además fácil de construir cuando $\circ$ es conmutativa, pues $\circ$ es conmutativa si y sólo si su tabla es simétrica con respecto a la diagonal principal.

Punto de discusión

Analizar con este criterio las tablas construidas en los ejemplos anteriores.

Continuando la exploración de propiedades, nos podemos preguntar por ejemplo si para la ley definida en 8 sobre los números reales, a saber, $x * y = x + y - 3$, es o no posible identificar un elemento $e \in \mathbb{R}$ tal que.

$$x * e = x = e * x,$$

es decir, un elemento $e \in \mathbb{R}$ tal que

$$x + e - 3 = x.$$

Aplicando propiedades de la suma de reales a esta última igualdad deducimos que $e = 3$. Nótese además que $3 * x = 3 + x - 3 = x$.

Si nos planteamos esta misma pregunta para la ley Δ definida sobre $\mathbb{R}^*$ veremos que no existe tal elemento. ¡Comprobarlo! Lo anterior motiva entonces una nueva definición.

Definición 5.4 *Dada una ley de composición $\circ$ sobre E, un elemento $e \in E$ es neutro (o elemento identidad) si* $\forall x \in E$

$$x \circ e = e \circ x = x.$$

y es llamado un inverso de x para $\circ$.

De la tabla del ejemplo 9 se puede extraer una información interesante acerca de los inversos. Nótese que

$$1 \diamond 1 = 0$$

y

$$1 \diamond 2 = 0,$$

donde 0 es el neutro para esta operación, es decir, que 1 y 2 son inversos de 1. Obsérvese además que $\diamond$ no es una ley asociativa, pues

$$(1 \diamond 1) \diamond 2 = 0 \diamond 2 = 2$$

y

$$1 \diamond (1 \diamond 2) = 1 \diamond 0 = 1.$$

Esto nos hace pensar, como efectivamente se expresa a continuación, que la unicidad del inverso depende de la asociatividad de la ley.

Teorema 5.1 *Si $\circ$ es asociativa sobre E, un elemento $x \in E$ admite a lo mas un inverso para $\circ$.*

Para demostrarlo, suponemos que para $x \in E$, existen y y z en E tales que,

$$x \circ y = e = y \circ x$$

$$x \circ z = e = z \circ x.$$

Entonces

$$\begin{aligned} y &= y \circ e \\ &= y \circ (x \circ z) \\ &= (y \circ x) \circ z \\ &= e \circ z \\ &= z. \end{aligned}$$

Si x es inversible para una ley asociativa $\circ$, podemos referirnos al inverso de x. Usualmente en notación aditiva lo notamos $-x$ y en notación multiplicativa lo notamos x^{-1}.

Nota. En $\mathbb{Q}$, $\mathbb{R}$ o $\mathbb{C}$ todo elemento no nulo es inversible para la multiplicación, pero en $\mathbb{Z}$ los únicos elementos inversibles para la multiplicación son 1 y -1. Todo elemento de $\mathbb{Z}_6$ es inversible para la adición, pero solamente 1 y 5 son inversibles para la multiplicación (hecho que retomaremos mas adelante). Toda simetría del cuadrado es inversible para la composición, por ejemplo, $r_1^{-1} = r_3$ y $d_1^{-1} = d_1$.

Puntos de discusión

1. Analizar el problema de los inversos para todas las leyes presentadas anteriormente.

2. Construir otras leyes (sobre conjuntos de simetrías, conjuntos de matrices, conjuntos de permutaciones, vectores geométricos, etc) y explorar inversos.

Las proposiciones que se discuten a continuación nos permitirán trabajar con los inversos (cuando estos existen) de manera mas ágil.

Teorema 5.2 *Si y es el inverso de x para una ley de composición $\circ$ sobre E entonces x es un inverso de y para $\circ$.*

El inverso de un elemento inversible, es el mismo, inversible. En particular, si x es inversible para una ley de composición asociativa $\circ$, entonces $(x^{-1})^{-1} = x$.

Demostración. Las dos primeras afirmaciones se deducen de las ecuaciones

$$x \circ y = e = y \circ x.$$

En particular, entonces, si x es inversible para una ley de composición asociativa $\circ$ entonces x^{-1} es inversible y el inverso de x^{-1} es único. Como

$$x \cdot x^{-1} = e$$

y

$$x^{-1} \cdot (x^{-1})^{-1} = (x^{-1})^{-1} \cdot x^{-1} = e,$$

concluimos por unicidad del inverso que $(x^{-1})^{-1} = x$.

Teorema 5.3 *Si x y y son elementos inversibles para una ley de composición asociativa $\circ$ sobre E, entonces $x \circ y$ es inversible para $\circ$ y*

$$(x \circ y)^{-1} = y^{-1} \circ x^{-1}.$$

Demostración. Analicemos

$$\begin{aligned} (x \circ y) \circ (y^{-1} \circ x^{-1} &= (x \circ (y \circ y^{-1})) \circ x^{-1} \\ &= (x \circ e) \circ x^{-1} \\ &= x \circ x^{-1} \\ &= e. \end{aligned}$$

Y similarmente

$$(y^{-1} \circ x^{-1}) \circ (x \circ y) = e.$$

Concluimos por unicidad del inverso que

$$(x \circ y)^{-1} = y^{-1} \circ x^{-1}.$$

Notas.

1. Si $\circ$ es conmutativa, por ejemplo el producto en $\mathbb{Q}$, $\mathbb{R}$ o $\mathbb{C}$, se tiene que $(x \circ y)^{-1} = x^{-1} \circ y^{-1}$.
2. La conclusión de la proposición anterior no es necesariamente válida para operaciones no asociativas, si sobre el conjunto $T = \{0, 1, 2\}$ definimos la ley que se ilustra en la tabla

$\circ$	0	1	2
0	0	1	2
1	1	1	0
2	2	0	1

0 es el elemento neutro para $\circ$ y cada elemento admite inverso único, pero

$$(1 \circ 1)^{-1} = 1^{-1} = 2$$

y

$$1^{-1} \circ 1^{-1} = 2 \circ 2 = 1.$$

Teorema 5.4 *Si $\circ$ es una ley de composición asociativa sobre E y x, y, z son elementos de E, se tiene*

i. Si x y y conmutan con z entonces $x \circ y$ conmuta con z.

ii. Si x conmuta con y y si y es inversible entonces x conmuta con y^{-1}.

iii. Si x conmuta con y y si ambos x y y son inversibles entonces x^{-1} conmuta con y^{-1}.

Demostración. (i) Si $x \circ z = z \circ x$ y $y \circ z = z \circ y$, entonces

$$(x \circ y) \circ z = x(\circ(y \circ z) = x \circ (z \circ y) = (x \circ z) \circ y = z \circ (x \circ y).$$

(ii) Si $x \circ y = y \circ x$ y y es inversible, entonces

$$y^{-1} \circ x = y^{-1} \circ (x \circ (y \circ y^{-1})) = y^{-1} \circ ((y \circ x) \circ y^{-1}) = (y^{-1} \circ y) \circ (x \circ y^{-1}) = x \circ y^{-1}.$$

(iii) Finalmente si x y y son inversibles y si $x \circ y = y \circ x$, tenemos,

$$x^{-1} \circ y^{-1} = (y \circ x)^{-1} = (x \circ y)^{-1} = y^{-1} \circ x^{-1}.$$

Puntos de discusión.

1. Sean x y y reales no negativos. Se define $x \sim y = |x - y|$. ¿Es $\sim$ una ley asociativa? ¿Cuál es el neutro para $\sim$? ¿Existen inversos?

2. En el conjunto de vectores del espacio, se definen el producto cruz y el producto punto (producto escalar). ¿Son leyes de composición interna? Identificar para ellas posibles elementos identidad.

3. Dados $a, b, c, d \in \mathbb{Z}$, se define en $\mathbb{Z} \times \mathbb{Z}$,

   $$(a, b) * (c, d) = (ac, bc + d).$$

   Determinar si $*$ es una ley asociativa y si es conmutativa, encontrar neutro e inversos.

Si analizamos ahora la suma en el conjunto de los números naturales, es claramente una ley de composición interna asociativa. Con esta simple caracterización los naturales constituyen un ejemplo de una estructura primaria, la estructura de *semigrupo*.

Definición 5.6 *Se dice que un conjunto § es un semigrupo si esta dotado de una ley de composición interna asociativa.*

Un semigrupo es conmutativo o abeliano si la composición es además conmutativa.

Puntos de discusión

1. Estudiar las matrices de tamaño 2×2 con elementos reales y sobre este conjunto el producto usual de matrices. ¿Tiene este conjunto estructura de semigrupo abeliano?

2. ¿Tienen los reales positivos con la operación $\sim$ definida en un punto de discusión anterior estructura de semigrupo?

3. ¿Son los enteros con el producto usual un semigrupo conmutativo?.

5.2 Grupos

Es claro entonces que, si consideramos la adición en el conjunto de los enteros, es allí una ley interna asociativa y conmutativa; esto es, los enteros constituyen un semigrupo abeliano con esta operación. Pero aun mas, existe un entero que actúa como neutro y todo entero tiene inverso (opuesto aditivo) para esta operación. De manera similar si estudiamos la multiplicación en los racionales, diríamos que con ésta tienen estructura de semigrupo conmutativo. Pero, de nuevo en este conjunto existe un elemento que actúa como neutro, el uno, y todo racional diferente de cero es inversible. Lo anterior nos sugiere desde luego que estos conjuntos con estas operaciones están dotados de una estructura mas completa que la de semigrupo, ésta es precisamente la estructura de *grupo*. Un grupo se puede definir entonces como sigue. Un conjunto G es un grupo si $(G, \circ)$ es un semigrupo tal que existe un elemento neutro para $\circ$ y todo elemento de G es inversible para $\circ$. O mas precisamente

Definición 5.7 *Sea G un conjunto dotado de una ley de composición interna notada $(\cdot)$. Decimos que G es un grupo si*

i. Dados $a, b, c \in G$ $a \cdot (b \cdot c) = (a \cdot b) \cdot c$;

ii. Existe e en G tal que $a \cdot e = e \cdot a = a$, para todo a en G;

iii. Dado a en G existe b en G tal que $a \cdot b = b \cdot a = e$.

e es llamado el neutro del grupo G y el elemento b del punto (iii), que se acostumbra a notar a^{-1}, es el inverso de a en G.

En los ejemplos anteriores $(\mathbb{Z}, +)$, $(\mathbb{R}^*, \cdot)$,

$(\mathbb{Z}_m, +)$, $(\mathbb{Z}_3^*, \cdot)$, $(U, \cdot)$, $(\S, \circ)$ y $(\mathbb{Q}, \cdot)$ son grupos. Verificarlo y explorar qué ocurre con los otros ejemplos presentados en la sección anterior.

Notas.

1. Del numeral (ii) de la definición de grupo se deduce que si G es un grupo, G es no vacío.

2. De las observaciones que hicimos anteriormente sobre leyes de composición interna asociativas podemos concluir que si G es un grupo

   (a) En G sólo hay un elemento neutro;

   (b) Todo elemento $a \in G$ tiene un único inverso;

   (c) Para todo $a \in G$, $(a^{-1})^{-1} = a$ el neutro y el inverso son únicos.

Hablamos en la sección anterior de leyes internas conmutativas y nos referimos también a *semigrupos conmutativos o abelianos*, naturalmente podemos hablar de grupos, conmutativos, o no, según que la ley interna sea o no conmutativa. Mas específicamente tenemos

Definición 5.8 *Si $(G, \cdot)$ es un grupo y si para cualesquiera $a, b \in G$ se cumple que $a \cdot b = b \cdot a$, decimos que G es un grupo abeliano o conmutativo.*

Claramente los grupos que comunmente trabajamos en la matemática elemental, los enteros con la adición, los racionales distintos de cero con la multiplicación, los complejos con la suma, etc. son abelianos.

Puntos de discusión

1. ¿ Son todos los grupos de los ejemplos que usted analizó en el punto de discusión anterior abelianos?
2. Explorar variados ejemplos de grupos no abelianos y analizar el cardinal de los conjuntos. ¿ Se observa alguna regularidad?

5.2.1 Otros ejemplos de grupos

1. Sea A un conjunto no vacío y consideremos

$$P(A) = \{B_i | B_i \subset A\},$$

el conjunto de partes de A. Definimos allí la diferencia simétrica ∇

$$B_i \nabla B_j = (B_i - B_j) \cup (B_j - B_i).$$

∇ es una ley de composición interna asociativa. El conjunto vacío, $\emptyset \in P(A)$ es tal que, $B_i \nabla \emptyset = \emptyset \nabla B_i = B_i$, para todo $B_i \in P(A)$.

Además, dado $B_i \in P(A)$, $B_i \nabla B_i = \emptyset$, es decir B_i es su propio inverso, o sea, todo elemento de $P(A)$ es inversible.

Podemos afirmar entonces que $(P(A), \nabla)$ es un grupo; aun mas, si $B_i, B_j \in P(A)$, $B_i \nabla B_j = B_j \nabla B_i$, esto es $(P(A), \nabla)$ es un grupo conmutativo o abeliano.

2. Consideremos $S = \{1, 2\}$ y construimos un conjunto S_2 cuyos elementos son todas las biyecciones (aplicaciones uno a uno y sobre) de S en S. En este conjunto están precisamente todas las posibles permutaciones de los elementos de S que son. La función identidad I, tal que $I(1) = 1$, $I(2) = 2$ y la función ϕ_1, tal que $\phi_1(1) = 2$ y $\phi_1(2) = 1$. Para mayor comodidad podemos notarlas como sigue.

$$I = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \phi_1 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Sobre S_2 consideramos, $\circ$, composición usual de funciones y observamos que $I \circ I = I$, $I \circ \phi_1 = \phi_1 \circ I = \phi_1$ y $\phi_1 \circ \phi_1 = I$. Concluimos, pues, que $(S_2, \circ)$ es un grupo abeliano.

3. Si ahora tomamos $S = \{1, 2, 3\}$ y S_3 el conjunto de todas las biyecciones sobre S (permutaciones de 3 elementos), y la composición de funciones sobre S_3 tenemos.

$$I = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \phi_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \phi_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix},$$
$$\phi_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \phi_4 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \phi_5 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}.$$

Obsérvese que $I = \phi_1 \circ \phi_1 = \phi_2 \circ \phi_2 = \phi_3 \circ \phi_3 = \phi_3 \circ \phi_4 = \phi_4 \circ \phi_3, \phi_4 \circ \phi_3 = \phi_1$ y $\phi_3 \circ \phi_4 = \phi_2$. Si usted considera todos los compuestos entre elementos de S_3 que nos restan por analizar podrá comprobar que $(S_3, \circ)$ es un *grupo no conmutativo*.

Punto de discusión

Si usted considera $S = \{1, 2, 3, \cdots n\}$, ¿cuántas biyecciones puede construir? ¿Es $(S_n, \circ)$ un grupo para cualquier $n \in \mathbb{N}$?

4. Sean ahora $f(x) = ax + b$ y $g(x) = cx + d$, $c, d \neq 0$, esto es, las ecuaciones de dos rectas que no son paralelas a ninguno de los ejes coordenados. Dichas rectas quedan completamente determinadas por los pares de números reales (a, b) y (c, d). Nótese que si consideramos la compuesta de estas dos funciones

$$(f \circ g)(x) = f(g(x) = f(cx + d) = a(cx + d) + b = acx + ad + b,$$

se obtiene de nuevo una función lineal y $a \cdot c \neq 0$ pues $a \neq 0$ y $c \neq 0$.

Esta composición de funciones lineales induce una operación $*$ en el conjunto $G = \mathbb{R}^* \times \mathbb{R}$ definida de la siguiente manera

$$(a, b) * (c, d) = (ac, ad + b),$$

que es una ley de composición interna sobre G. Es mas $(G, *)$ es un grupo.

Comprobemos que $*$ es asociativa.

$$\begin{aligned} ((a, b) * (c, d)) * (e, f) &= (ac, ad + b) * (e, f) \\ &= ((ac)e, (ac)f + (ad + b)) \\ &= (a(ce), a(cf) + ad + b) \\ &= (a(ce), a(cf + d) + b) \\ &= (a, b) * (ce, cf + d) \\ &= (a, b) * ((c, d) * (e, f)) \end{aligned}$$

El elemento neutro para $*$ es la pareja $(1, 0)$ pues $(a, b) * (1, 0) = (1, 0) * (a, b) = (a, b)$.

Encontremos ahora el inverso de un par (a, b) en G. Para ello debemos identificar un par (x, y) en G tal que

$$(a, b) * (x, y) = (1, 0),$$

esto es

$$(ax, ay + b) = (1, 0).$$

Luego, $ax = 1$ y $ay + b = 0$. Como $a \neq 0$, $x = \frac{1}{a} = a^{-1}$ y $y = -b \cdot a^{-1}$, esto es, el par $(x, y) = (a^{-1}, -b \cdot a^{-1})$. Claramente

$$(a, b) * (a^{-1}, -b \cdot a^{-1}) = (aa^{-1}, a(-b \cdot a^{-1}) + b) = (1, 0).$$

Obsérvese que $x = \frac{1}{a}y - \frac{-b}{a}$ es la función inversa de $y = ax + b$.

Finalmente usted puede comprobar que $(G, *)$ es no abeliano, $(a, b)*(c, d) \neq (c, d) * (a, b)$; este último hecho es de esperarse, pues la composición de funciones (que no es conmutativa) indujo la operación $*$.

Nota. Es importante resaltar aquí cómo es posible construir interesantes ejemplos de una estructura abstracta con objetos completamente familiares en la matemática básica, las funciones, y en este último caso en particular con las funciones lineales.

Puntos de discusión

1. ¿Es el conjunto de los números irracionales un grupo bajo la multiplicación?
2. Las funciones x, $\frac{1}{x}$, $-x$ y otra función $g(x)$, forman un grupo con la composición Identificar la función $g(x)$ y construir la respectiva tabla.
3. Sea $G = \{f_0, f_1, f_2, f_3, f_4, f_5\}$ donde $f_0, f_1, f_2, f_3, f_4, f_5$ son funciones definidas en los reales de la siguiente manera.
   $f_0(x) = x$, $f_1(x) = \frac{1}{1-x}$, $f_2(x) = \frac{x-1}{x}$, $f_3(x) = \frac{x+1}{2-x}$, $f_4(x) = \frac{2-x}{1-2x}$, $f_5(x) = \frac{2x+1}{x+1}$.
   Considerar en G la composición de funciones y determinar si $(G, \circ)$ es un grupo.
4. Si sobre el conjunto $A = \{a\sqrt{2} + b\sqrt{3} | a, b \in \mathbb{Q}\}$, toma usted el producto usual, ¿es $(A,)$ un grupo?
5. Sean r una rotación de 60° en el plano, r^2 una rotación en el plano de 120°,y así sucesivamente. Sobre el conjunto $\{r, r^2, r^3, r^4, r^5, r^6\}$, considerar la composición de las rotaciones. ¿Es este conjunto un grupo con esta operación binaria?

Para enriquecer nuestro análisis de los grupos es pertinente determinar si los resultados familiares para nosotros en los conjuntos numéricos, por ejemplo, aquellos que se derivan de los axiomas de los reales como son la generalización de propiedades, la posibilidad de transformar igualdades, la existencia de solución para ecuaciones lineales, etc. tienen o no significado al interior de cualquier grupo (de un grupo abstracto). Los teoremas que se incluyen a continuación nos abren esta posibilidad.

Teorema 5.5 *Si $(G, .)$ es un grupo, en G son válidas las leyes cancelativas, es decir, $\forall a, b, c \in G$, si $a \cdot b = a \cdot c \Longrightarrow b = c$ y si $b \cdot a = c \cdot a \Longrightarrow b = c$.*

Demostración. Analizaremos tan solo la primera, pues el argumento para la segunda es completamente análogo.

Dado que $a \in G$ y G es un grupo, existe $a^{-1} \in G$, tal que

$$a \cdot a^{-1} = a^{-1} \cdot a = e,$$

donde e es el neutro de G. Como por hipótesis $a \cdot b = a \cdot c$, multiplicando los dos miembros de esta igualdad (a izquierda) por a^{-1}, tenemos

$$a^{-1} \cdot (a \cdot b) = a^{-1} \cdot (a \cdot c)$$

$$(a^{-1} \cdot a) \cdot b = (a^{-1} \cdot a) \cdot c,$$

de donde, $e \cdot b = e \cdot c$ y esto implica que $b = c$.

Teorema 5.6 *Sea $(G,\cdot)$ un grupo si $a,b,c,d \in G$ entonces $(a\cdot b)\cdot(c\cdot d) = a\cdot((b\cdot c)\cdot d)$ Es decir, es posible generalizar la asociatividad de $\cdot$ en G).*

Demostración. Si hacemos $x = c\cdot d$, y empleamos el hecho de que $\cdot$ es una ley asociativa, tenemos que

$$\begin{aligned}(a\cdot b)\cdot(c\cdot d) &= (a\cdot b)\cdot x\\ &= a\cdot(b\cdot x)\\ &= a\cdot(b\cdot(c\cdot d))\\ &= a\cdot((b\cdot c)\cdot d)).\end{aligned}$$

Nota. Acerca de una ley interna asociativa $\cdot$, con elemento neutro e inversos, anotamos ya que para cualesquiera a,b, $(a\cdot b)^{-1} = b^{-1}\cdot a^{-1}$ y que tan solo cuando dicha ley es además conmutativa, $(a\cdot b)^{-1} = a^{-1}\cdot b^{-1}$. Tales afirmaciones resultan válidas desde luego si $(G,\cdot)$ es un grupo, esto es,

Teorema 5.7 *Sea $(G,\cdot)$ un grupo, $a,b\in G$. Entonces $(a\cdot b)^{-1} = b^{-1}\cdot a^{-1}$.*

Punto de discusión

Demostrar el teorema anterior.

En general $(a\cdot b)^{-1}$ difiere de $a^{-1}\cdot b^{-1}$, excepto cuando $(G,\cdot)$ es un grupo abeliano.

En un ejemplo anterior tomabamos, $G = \mathbb{R}^*\times\mathbb{R}$ y $(a,b)*(c,d) = (ac, ad+b)$. Analicemos

$$((2,3)*(7,1))^{-1} = (14, 2+3)^{-1} = (14,5)^{-1} = \left(\frac{1}{14}, \frac{-5}{14}\right).$$

Pero $(2,3)^{-1} = (\frac{1}{2}, \frac{-3}{2})$ y $(7,1)^{-1} = (\frac{1}{7}, \frac{-1}{7})$, de donde,

$$(2,3)^{-1} * (7,1)^{-1} = \left(\frac{1}{14}, \frac{-22}{14}\right) \neq ((2,3)*(7,1))^{-1},$$

en tanto que $(7,1)^{-1} * (2,3)^{-1} = ((2,3)*(7,1))^{-1}$.

De la posibilidad de cancelar en un grupo G se deduce una interesante propiedad de los grupos.

Teorema 5.8 *En un grupo $(G,\cdot)$ existe un único elemento tal que $x\cdot x = x$, y este elemento es e, el neutro de G.*

Demostración. Como e es el neutro $e\cdot e = e$. Si existe en G otro elemento x tal que $x\cdot x = x$, de nuevo por ser e neutro, tendríamos que $x\cdot e = x$, de donde, $x\cdot x = x\cdot e$. Aplicando cancelativa concluimos que $x = e$.

Nota. Nótese que si en un conjunto tenemos una ley asociativa, conmutativa, que posee elemento neutro, no podemos garantizar la validez del resultado anterior. Por ejemplo en $\mathbb{Z}_6$ el producto módulo 6 cumple las anteriores propiedades, pero $1\cdot 1 = 1$ y $3\cdot 3 = 3$, esto es, la ecuación $x\cdot x = x$ tiene mas de una solución.

Si en $\mathbb{Z}_6$ con multiplicación módulo 6, se considera una ecuación del tipo $a\cdot x = b$, por ejemplo $4x = 2$, son soluciones de esta ecuación en $\mathbb{Z}_6$, $x = 2$ y $x = 5$. Es decir aquí tampoco la solución es única. Este hecho no se presenta en un grupo como lo garantiza el siguiente teorema.

Teorema 5.9 *Si $(G,\cdot)$ es un grupo y $a,b \in G$, entonces las ecuaciones $a\cdot x = b$ y $y\cdot a = b$ tienen solución y esta solución es única.*

Demostración.

(i) Veamos que $x = a^{-1}\cdot b$ es solución de la primera ecuación.

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

Se procede de manera análoga para mostrar que $b\cdot a^{-1}$ es solución de la segunda.

(ii) Nos restaría demostrar que esta solución es única. Para ello se supone que existen x_1 y x_2, soluciones de la primera ecuación, esto es que $a\cdot x_1 = b$ y que $a\cdot x_2 = b$. Se concluye entonces que $a\cdot x_1 = a\cdot x_2$ y aplicando la ley cancelativa se tiene que $x_1 = x_2$, esto es, la solución es única.

Puntos de discusión

1. Considerar en $\mathbb{R}^+$ la siguiente ley de composición interna

$$x * y = x\log_{10} y$$

¿Es $*$ una ley asociativa? ¿Existe módulo? ¿Hay inversos? ¿Es $(\mathbb{R}^+,*)$ un grupo.

2. En el conjunto $A = \{c,a,b\}$ se define una ley de composición interna por la siguiente tabla

$\circ$	c	a	b
c	c	a	b
a	b	c	a
b	a	b	c

¿Es $\circ$ una ley asociativa? ¿Existe un elemento en A que actúe como módulo? En caso afirmativo, ¿son los elementos de A inversibles para $\circ$? ¿Es $(A,\circ)$ un grupo?

Dada sobre $\mathbb{R}$ la ley ∇ definida por

$$x\nabla y = x + y + x^2 y.$$

(a) ∇ no es una ley asociativa. Comprobarlo.

(b) Demostrar que ∇ admite un elemento neutro y todo número real tiene un inverso único a derecha para ∇, pero existen elementos que no tienen inverso a izquierda para ∇.

Los puntos de discusión anteriores nos motivan a presentar condiciones bajo las cuales un conjunto dotado de una ley interna asociativa es un grupo. Veamos esta idea en el siguiente teorema.

Teorema 5.10 *Sea $(G,\cdot)$ un conjunto con una operación interna asociativa. Entonces $(G,\cdot)$ es un grupo si y sólo si*

i. Existe $e \in G$ tal que $e\cdot x = x$, para todo $x \in G$. (Existe neutro a izquierda.)

ii. Para cada $a \in G$, existe $y \in G$ tal que $y \cdot a = e$. (Existen inversos a izquierda.)

Demostración. Si $(G, .)$ es un grupo (i) y (ii) se cumplen.

Supongamos ahora que (i) y (ii) se cumplen y sea $x \in G$. Por (ii) existe $x' \in G$ tal que

$$x' \cdot x = e.$$

A su vez para x', también por (ii) existe x'' tal que

$$x'' \cdot x' = e.$$

Entonces

$$x \cdot x' = e \cdot (x \cdot x') = (x'' \cdot x') \cdot (x \cdot x') = x'' \cdot (x' \cdot x) \cdot x' = (x'' \cdot e) \cdot x' = x'' \cdot x' = e,$$

lo cual significa que x' es también inverso a derecha de x.

Veamos ahora que e es neutro a derecha. Sean $x \in G$ y x' inverso a izquierda de x. Entonces

$$x \cdot e = x \cdot (x' \cdot x) = (x \cdot x') \cdot x = e \cdot x = x.$$

Se usó aquí el hecho, ya demostrado, que si x' es inverso a izquierda es también inverso a derecha.

Concluimos entonces que $(G, \cdot)$ es un grupo.

Nota. Nótese que en la demostración se aplica reiteradamente el hecho de que $\cdot$ es asociativa.

Por otra parte, un resultado análogo al del Teorema 5.2.6 se tiene, si para una ley asociativa sobre G existen neutro e inverso a derecha.

Puntos de discusión

1. Considerar de nuevo $\mathbb{Z}_6$ con la multiplicación y comprobar que en este conjunto la ecuación $ax = b$ no siempre tiene solución.
2. Analizar las leyes ∇, $\circ$ y $*$ del punto de investigación anterior y determinar si en estos casos las ecuaciones $a \cdot x = b$, y, $y \cdot a = b$ tienen siempre solución. ($\cdot$ representa en cada caso la operación mencionada.)

El poder garantizar la solubilidad de ecuaciones lineales, problema muy familiar e importante en la matemática básica resulta fundamental también en la caracterización de los grupos. El siguiente teorema expresa este hecho.

Teorema 5.11 *Sean G un conjunto no vacío, $\cdot$ una operación binaria asociativa. $(G, \cdot)$ es un grupo si y sólo si dados $a, b \in G$ las ecuaciones $a \cdot x = b$ y $y \cdot a = b$ admiten siempre solución en G.*

Demostración. Si $(G, .)$ es un grupo, ya demostramos el Teorema 5.2.5 en el que se garantiza que estas ecuaciones tienen solución y esta solución es única. Nos resta entonces suponer que estas dos ecuaciones tienen solución y demostrar que $(G, \cdot)$ es grupo.

Como $G \neq \emptyset$, existe por lo menos un elemento $a \in G$. Se tiene entonces por hipótesis que la ecuación $a \cdot x = a$ tiene una solución en G. Llamemos e esta

solución, esto es, $a \cdot e = a$. El objetivo es demostrar que e es precisamente un neutro a derecha para $\cdot$ en G. Para ello sea $x \in G$. Por hipótesis existe $y \in G$ tal que $y \cdot a = x$. Por lo tanto,

$$x \cdot e = (y \cdot a) \cdot e = y \cdot (a \cdot e) = y \cdot a = x.$$

Como x es arbitrario, con esto se demuestra que e es neutro a derecha para $(\cdot)$ en G.

De otra parte, dado $x \in G$ la ecuación $x \cdot x' = e$, admite solución $x' \in G$, es decir, x' es inverso a derecha de x. Nos encontramos ahora en las condiciones del Teorema 5.2.6, existencia de neutro a derecha e inversos a derecha. Concluimos entonces que $(G, \cdot)$ es un grupo.

Punto de discusión

Considerar en $\mathbb{Z}$ la ley definida por

$$a \circ b = a^b.$$

Mostrar que en $\mathbb{Z}$ existe un elemento que actúa como neutro a derecha para esta operación. ¿Posee todo elemento de $\mathbb{Z}$ un inverso a derecha? ¿Es $(\mathbb{Z}, \circ)$ un grupo?

Con los elementos anteriores podemos ya introducirnos en un punto que nos interesa, los nexos entre la aritmética "elemental" y la teoría de grupos. Es claro que ya se han vislumbrado algunos, el partir de la familiaridad con las operaciones de adición y multiplicación en los naturales, enteros, racionales, reales; el analizar sus propiedades motiva de hecho la presentación de ideas mas abstractas, como son las de ley de composición interna, noción de semigrupo, y de grupo. Pero lo importante, como lo anotamos anteriormente es observar cuáles de las características especiales de estos semigrupos o grupos que nos son tan familiares se tienen en un grupo abstracto y cómo la aritmética misma nos da elementos para caracterizarlo.

La primera idea, que por su misma naturaleza usa explícitamente elementos de la aritmética, es la noción de orden de un grupo.

Definición 5.9 *Sea $(G, \cdot)$ un grupo. Se llama orden de G y se nota $o(G)$ al cardinal de G.*

Definición 5.10 *Si $o(G)$ es finito el grupo es finito. En caso contrario se dirá que el grupo es infinito.*

Nótese que los grupos mas familiares para nosotros son de orden infinito. En el análisis de los grupos finitos el orden es fundamental para caracterizarlos. Un estudio por ejemplo de las tablas de los grupos de órdenes menores o iguales a cinco nos permite afirmar por qué tales grupos resultan siempre abelianos, mostrándonos de paso las posibles presentaciones de estos grupos. Analizaremos a continuación esta idea mas ampliamente.

5.2.2 Tablas de grupos finitos

Para el caso de conjuntos finitos, con 1,2,3 o 4 elementos es posible hacer un análisis exhaustivo de las distintas tablas determinadas por las diferentes leyes de composición que dan a estos conjuntos una estructura de grupo.

En primer lugar de la definición de grupo se deduce que un grupo es no vacío, existe por lo menos un elemento, el neutro, de aquí que el caso mas simple sería

1. $G = \{e\}$ y como e actúa como neutro, la única posible ley de composición interna estaría dada por $e \cdot e = e$. Lo anterior implicaría que e es su propio inverso, esto es $e^{-1} = e$

2. Si ahora el conjunto está formado por dos elementos distintos $\{e, a\}$, uno de ellos debe actuar como neutro, si tomamos e como en neutro tenemos

$$e \cdot e = e \qquad e \cdot a = a \cdot e = a.$$

De nuevo e es su propio inverso. ¿Cuál es el inverso a' de a? No tendríamos sino dos posibilidades: $a' = e$ o $a' = a$.

Si $a' = e$, $a \cdot a' = a \cdot e = e$, pero tenemos que $a \cdot e = a$, concluiríamos que $a = e$ y esto contradice la hipótesis inicial de que G tiene dos elementos distintos. Entonces $a' = a$ es decir $a \cdot a = e$ y a es también su propio inverso. La única tabla posible es

$\cdot$	e	a
e	e	a
a	a	e

La asociatividad puede ser verificada facilmente; además, es claro que en éste caso $(G, \cdot)$ resulta un grupo abeliano.

Ilustremos con algunos ejemplos de grupos que pueden ser representados por esta tabla.

(a) $G = \{1, -1\}$, con el producto usual de reales. G es el grupo de las raíces cuadradas de la unidad, esto es, $G = \{x \in \mathbb{R} | x^2 = 1\}$ y la respectiva tabla es

$\cdot$	1	-1
1	1	-1
-1	-1	1

(b) El grupo $(S_2, \circ)$ de todas las permutaciones del conjunto $\{1, 2\}$ puede ser también representado por esta tabla. Recordemos $I \circ I = I$, $I \circ \phi_1 = \phi_1 \circ I = \phi_1$ y $\phi_1 \circ \phi_1 = I$.

(c) Considerar $(\mathbb{Z}_3^*, \cdot)$ y determinar si puede ser representado por esta tabla.

3. Consideremos ahora un conjunto constituido por tres elementos distintos $\{e, a, b\}$y exploremos posibles tablas.

Denotemos nuevamente el neutro por e. Tenemos $e \cdot e = e$, $a \cdot e = e \cdot a = a$ y $b \cdot e = e \cdot b = b$. Debemos definir ahora $a \cdot a$

$\cdot$	e	a	b
e	e	a	b
a	a	?	
b	b		

Se darían tres opciones $a \cdot a = e$, $a \cdot a = a$ o $a \cdot a = b$. La segunda nos llevaría a que $a = e$ y esto es una contradicción. Analicemos entonces las

otras opciones. Si $a \cdot a = e$, esto es $a^{-1} = a$, nos preguntamos qué ocurriría con $a \cdot b$. De nuevo se tendrían tres posibilidades, $a \cdot b = e$, $a \cdot b = a$, o $a \cdot b = b$; pero cualquiera de éstas nos llevaría a una contradición, en su orden $b = a$, $b = e$ y $a = e$.

De lo anterior se concluye que no es posible tomar $a \cdot a = e$, queda entonces una sola posiblidad para $a \cdot a$, $a \cdot a = b$, $(a^2 = b)$, realizando un análisis similar con las otras posiciones se concluye que $a \cdot b = e$, $b \cdot a = e$ y $b \cdot b = a$.

$\cdot$	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

Esto implica que e es su propio inverso, a es el inverso de b (y visceversa); una presentación usual para este grupo es $G = \{e, a, a^2\}$.

Veamos algunos grupos que pueden ser representados por esta tabla.

(a) $G = \{x \in \mathbb{C} | x^3 = 1\}$, las raíces cúbicas de la unidad, con el producto usual de complejos.

$$G = \{e^{\frac{2\pi i}{3}}, e^{\frac{4\pi i}{3}}, e^{2\pi i}\},$$

donde $e^{2\pi i} = \cos 2\pi + i \sin 2\pi = 1$ (aquí e es el número real base de los logaritmos naturales). Si $a = e^{\frac{2\pi i}{3}}$, $a^2 = e^{\frac{4\pi i}{3}}$ y $a^3 = e^{2\pi i} = 1$. Es decir $G = \{1, a, a^2\}$.

(b) $G = \mathbb{Z}_3$ con la suma módulo 3 y elemento identidad 0 cuya tabla es

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

El inverso (opuesto aditivo) de 1 es 2, $1 + 2 = 2 + 1 = 0$; $2 + 2 = 1$ y $2 + 2 + 2 = 0$.

Punto de discusión

Explorar otros grupos con tres elementos y caracterizarlos. ¿Pueden todos ser representados por la misma tabla? ¿Qué concluye?

4. Para un conjunto $G = \{e, a, b, c\}$, con cuatro elementos distintos es posible definir, a través de un análisis similar a los anteriores, dos leyes de composición interna diferentes que proveen a G de estructura de grupo abeliano. En las tablas siguientes se ilustran tales operaciones.

$\cdot$	e	a	b	c
e	e	a	b	c
a	a	b	c	e
b	b	c	e	a
c	c	e	a	b

Tabla 1.

$*$	e	a	b	e
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

Tabla 2.

Veamos ahora algunos de los grupos representados por la tabla 1.

(a) $(\mathbb{Z}_4, +)$, con la suma módulo 4.

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

(b) Las raíces cuartas de la unidad con el producto usual de complejos.

$$U = \{x \in \mathbb{C} | x^4 = 1\} = \{1, -1, i, -i\}$$

La segunda tabla corresponde al conocido *cuarto grupo de Klein* que tiene variedad de representaciones en contexto diversos, a saber, como un grupo de permutaciones de cuatro objetos, como un grupo de transformaciones (funciones), como el grupo de simetrías del tetraedro, como un grupo en la aritmética binaria, etc.

Consideremos aquí como ejemplo un conjunto F de funciones de $\mathbb{R}^*$ en $\mathbb{R}^*$,

$$F = \{i_d, f_1, f_2, f_3\},$$

donde: $i_d(x) = x$, $f_1(x) = -x$, $f_2(x) = \frac{1}{x}$ y $f_3(x) = -\frac{1}{x}$. F es un grupo con la composición de funciones; obsérvese que

$$(f_1 \circ f_1)(x) = f_1(f_1(x) = f_1(-x) = x$$

$$(f_2 \circ f_2)(x) = f_2(f_2(x) = f_2(\frac{1}{x}) = x$$

$$(f_3 \circ f_3)(x) = f_3(f_3(x) = f_3(-\frac{1}{x}) = x.$$

Esto es cada elemento es su propio inverso. Además

$$f_1 \circ f_2 \equiv f_2 \circ f_1 \equiv f_3$$

$$f_1 \circ f_3 \equiv f_3 \circ f_1 \equiv f_2$$

$$f_2 \circ f_3 \equiv f_3 \circ f_2 \equiv f_1$$

Puntos de discusión

1. Considerar las matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

Mostrar que forman un grupo con el producto usual y construir la tabla de dicho grupo. ¿Tiene la misma representación que el cuarto grupo de Klein?

1. Estudiar los otros grupos que se menciona pueden ser representados por la tabla 2.

2. Buscar otros ejemplos de grupos con cuatro elementos. ¿ Pueden ser siempre representados por una de estas tablas? ¿Qué concluye?

Parece natural en este momento tratar de presentar unas ideas mas generales que permitan orientar la construcción de tablas de un grupo finito $G = \{a_1, a_2, a_3, \cdots, a_n\}$. Se debe seleccionar en primer lugar un elemento, por ejemplo a_1, como elemento neutro. En dicha tabla el elemento $a_i \cdot a_j$ es el situado en la intersección de la fila i con la columna j, y $a_j \cdot a_i$ es el situado en la intersección de la fila j con la columna i. Una consecuencia del Teorema 5.2.5 (en un grupo las ecuaciones $a \cdot x = b$ y $y \cdot a = b$ tienen siempre solución y esa solución es única.), es que en la tabla de un grupo finito cada elemento debe aparecer una y sólo una vez en cada fila y en cada columna, pues las soluciones de las ecuaciones $a_i \cdot x = a_j$ y $y \cdot a_j = a_k$ son únicas.

Recíprocamente, es posible afirmar que si en una tabla para una operación asociativa $\cdot$ sobre un conjunto finito $G = \{a_1, a_2, \cdots, a_n\}$ existe un elemento que actúa como neutro y cada elemento de G aparece una y sólo una vez en cada fila y columna, entonces $(G, \cdot)$ es un grupo.

Nótese que observando la tabla se puede decir si G es o no abeliano, si la tabla es simétrica con respecto a la diagonal, el grupo es abeliano, en caso contrario no lo es.

Nota. Analicemos la siguiente tabla.

$*$	a	b	c	d	e
a	a	b	c	d	e
b	b	c	e	a	d
c	c	a	d	e	b
d	d	e	a	b	c
e	e	d	b	c	a

En esta tabla cada fila o columna contiene todos los elementos del conjunto sin repetición, pero esto no es suficiente para afirmar que el conjunto con esta ley es un grupo. En este caso $*$ no es asociativa, pues basta mostrar un contraejemplo.

$$(b * c) * d = e * d = c$$

pero

$$b * (c * d) = b * e = d.$$

¿Hay elemento neutro? ¿Es todo elemento inversible?

Para afirmar entonces que un conjunto dotado de una ley interna es un grupo no basta garantizar que cada elemento aparece una y sólo una vez en cada fila o columna, se requiere además comprobar que la ley es asociativa aunque el proceso sea tedioso.

Punto de discusión

Considerar un conjunto con cinco elementos distintos. Construir todas las posibles tablas y encontrar grupos que puedan ser representados por estas tablas.

Un análisis informal de algunas tablas de grupos finitos nos permitirá aproximarnos a un importante teorema de representación que formalizaremos mas adelante. Para un grupo con dos elementos la única tabla posible es

$\cdot$	e	a
e	e	a
a	a	e

Si analizamos las filas de esta tabla observamos que corresponden a permutaciones de los elementos $\{e, a\}$. La primera corresponde a la permutación identidad

$$I = \begin{pmatrix} e & a \\ e & a \end{pmatrix}$$

y la segunda a

$$x = \begin{pmatrix} e & a \\ a & e \end{pmatrix}.$$

El conjunto formado por estas dos permutaciones $\{i, x\}$ es un grupo con la composición y su tabla

$\circ$	i	x
i	i	x
x	x	i

es idéntica a la original.

Si consideramos ahora $(\mathbb{Z}_3, +)$

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

y hacemos

$$I = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 1 & 2 \end{pmatrix}, \; x = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \end{pmatrix}, \; y = \begin{pmatrix} 0 & 1 & 2 \\ 2 & 0 & 1 \end{pmatrix}$$

$\{i, x, y\}$ es un grupo con la composición (subgrupo de S_3, el grupo de permutaciones de tres elementos y su tabla es

$\circ$	i	x	y
i	i	x	y
x	x	y	i
y	y	i	x

que es idéntica a la tabla de $(\mathbb{Z}_3, +)$.

Similarmente si usted considera, por ejemplo, $\{0, 1, 2, 3, 4\}$ con la suma módulo 5 y elabora su respectiva tabla, puede observar que es idéntica a la tabla del grupo,

$\circ$	i	x	y	z	w
i	i	x	y	z	w
x	x	y	z	w	i
y	y	z	w	i	x
z	z	w	i	x	y
w	w	i	x	y	z

donde i, x, y, z, w son las permutaciones de 5 elementos

$$i = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 & 4 \end{pmatrix}, x = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 & 0 \end{pmatrix}, y = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 0 & 1 \end{pmatrix},$$

$$z = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 3 & 4 & 0 & 1 & 2 \end{pmatrix}, w = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 4 & 0 & 1 & 2 & 3 \end{pmatrix}.$$

Este es de nuevo un subgrupo de un grupo de permutaciones, el grupo de permutaciones de 5 elementos que tiene 5! elementos.

Las anteriores son tan solo ilustraciones del *Teorema de Cayley*, que analizaremos posteriormente y que garantiza que siempre es posible encontrar un grupo de n permutaciones que tienen una tabla de grupo idéntica a algún grupo dado de orden n.

En algunos de los ejemplos que hemos discutido en la presente sección usamos ya en grupos multiplicativos una notación para las potencias de elementos. Ahora vamos a darle significado a esta notación.

Definición 5.11 *Sean $(G, \cdot)$ un grupo y $a \in G$, las potencias enteras de a se pueden definir inductivamente así (en notación multiplicativa)*

i. $a^0 = e$, donde e es el neutro de G.
ii. $a^n = a \cdot a \cdots a$, n veces donde $n \in \mathbb{N}$.
iii. $a^{-n} = (a^{-1})^n = a^{-1} \cdot a^{-1} \cdots a^{-1}$, donde $n \in \mathbb{N}$ y a^{-1} es el inverso de a.

En notación aditiva, a^n significa $na = a + a + \cdots + a$ y a^{-n}, $(-n)a = n(-a) = (-a) + (-a) + \cdots + (-a)$, donde $-a$ aparece como sumando n veces y $n \in \mathbb{N}$.

Con esta presentación es posible demostrar para un grupo propiedades completamente análogas a las que se tienen para la potenciación en los reales.

Teorema 5.12 *Sean $(G, \cdot)$ un grupo, $a \in G$, y $m, n \in \mathbb{Z}$ entonces*

i. $a^n \cdot a^m = a^{n+m} = a^m \cdot a^n$
ii. $(a^n)^m = anm = (a^m)^n$
iii. $a^{-n} = (a^{-1})^n = (a^n)^{-1}$.
iv. $e^n = e$.

Punto de discusión

¡Demostrar el teorema anterior!

Analicemos cómo usar estas propiedades en un grupo particular. Sea $G = \{(x, y, z) | x, y, z \in \{0, 1, 2\}\}$ y definimos

$$(x_1, y_1, z_1) * (x_2, y_2, z_2) = (x_1 + x_2 + y_2 z_1, y_1 + y_2, z_1 + z_2)$$

donde $+$ es adición módulo 3.

Por ejemplo, $(2, 2, 0) * (1, 1, 2) = (2 + 1 + 4, 2 + 1, 0 + 2) = (1, 0, 2)$. $(G, *)$ es un grupo no conmutativo. ¡Comprobarlo!

Para $(x, y, z) \in G$ determinemos $((x, y, z))^3$

$$\begin{aligned}((x,y,z))^3 &= ((x,y,z) * (x,y,z)) * (x,y,z) \\ &= (2x + yz, 2y, 2z) * (x,y,z) \\ &= (3x + yz + 2yz, 3y, 3z) \\ &= (3x + 3yz, 3y, 3z) \\ &= (0,0,0).\end{aligned}$$

Es decir, al elevar cualquier elemento de G al cubo obtenemos la tripla $(0, 0, 0)$ que es el neutro del grupo.

Si consideramos por ejemplo las triplas $a = (1, 0, 0)$, $b = (0, 1, 0)$, $c = (0, 0, 1)$. Tenemos

$a^2 = (2, 0, 0)$, $b^2 = (0, 2, 0)$, $c^2 = (0, 0, 2)$,

$a \cdot b = b \cdot a = (1, 1, 0)$, $c \cdot b = (1, 1, 1)$, $b \cdot c = (0, 1, 1)$

$a \cdot c = c \cdot a = (1, 0, 1)$.

De estas últimas observaciones es posible concluir que todo elemento de G puede ser expresado en términos de a, b, c; es decir, en este caso usando potencias de algunos elementos del grupo es posible determinarlo completamente.

Es muy usual que en nuestro trabajo en grupos finitos se presenten situaciones similares a la anterior, esto es que a partir de potencias de algunos elementos del grupo, se pueda caracterizar completamente éste y lo mas importante se pueda explorar una posible representación.

Puntos de discusión

1. Un grupo tiene elementos $\{e, a, b, c\}$ que satisfacen las siguientes relaciones:

   $$a^3 = b^3 = c^3 = e, a \cdot b = b \cdot a, a \cdot c = c \cdot a y c \cdot b = b \cdot c.$$

   Caracterizar completamente este grupo.

2. Demostrar que no existe un grupo en el cual $r^5 = e = a^2$ y $ar^3 = ra$, donde a, r, e son elementos distintos.

En uno de los puntos de discusión, seguramente ya usted construyó todas las tablas posibles para un conjunto con 5 elementos y observó, así como en las tablas de grupos de ordenes menores a 5, que todos resultan ser conmutativos. Sinembargo, hemos hecho mención de grupos no abelianos $((S_3, \circ)$ por ejemplo). ¿Cómo son sus ordenes? El siguiente teorema nos aclara este punto.

Teorema 5.13 *Si $(G, \cdot)$ es un grupo no abeliano, entonces $o(G) \geq 6$.*

Demostración. Si $(G, \cdot)$ es no abeliano entonces existen $a, b \in G$, $a \neq b$ tales que $a \cdot b \neq b \cdot a$. De esto se deduce que $e, a, b, a \cdot b, b \cdot a$ son cinco elementos diferentes de G. Un sexto es a^2 si $a^2 \neq e$. Ahora bien, si $a^2 = a$, concluimos que $a = e$.

3. Considerar el grupo $(\mathbb{R}^*, .)$ y analizar si los siguientes subconjutos son o no subgrupos de $\mathbb{R}$

   (a) Todos los enteros y sus recíprocos (inversos múltiplicativos).

   (b) Todos los números irracionales y el uno.

   (c) El conjunto $\{\cdots, \frac{1}{16}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4, 8, 16, \cdots\}$

4. ¿Puede usted encontrar subgrupos finitos de ¿$(\mathbb{R}, +)$?, ¿$(\mathbb{R}^*, .)$?

5. Es el conjunto de los enteros Gaussianos, $\{x + iy | x, y \in \mathbb{Z}\}$, ¿un subgrupo de $(\mathbb{C}, +)$?, ¿un subgrupo de $(\mathbb{C}, \cdot)$?

6. Si $a, b \in \mathbb{C}$ las matrices de la forma

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$$

   Constituyen un subgrupo de $(M_{2\times 2}(\mathbb{C}), \cdot)$? Investigar qué ocurre con las matrices 2×2 con determinante ± 1 respecto al grupo anterior.

7. Dar ejemplos de subgrupos abelianos de grupos no abelianos.

8. Si $(V_3, +)$ el grupo de los vectores del espacio con la adición, buscar allí varios subgrupos infinitos. ¿Puede encontrar algún subgrupo finito?

9. Para los pares ordenados de reales, se define $\bigoplus$

$$(a, b) \bigoplus (c, d) = (ac, bc + d)$$

   $(\mathbb{R}^* \times \mathbb{R}^*, \bigoplus)$ es grupo. Identificar todos sus subgrupos.

Existen desde luego caracterizaciones (que usted puede usar en las investigaciones anteriores) que permiten agilizar el estudio de los subgrupos, y están expresadas en los teoremas siguientes.

Teorema 5.14 *Si $H \subset G$, $H \neq \emptyset$ y $(G, \cdot)$ grupo. H es subgrupo de G si y sólo si $\forall a, b \in H$ $a \cdot b^{-1} \in H$.*

Demostración.

(i) Supongamos que H es un subgrupo de G. Por definición entonces en $(H, \cdot)$, si $a, b \in H$, el inverso de b, $b^{-1} \in H$ y por ser $\cdot$ ley interna en H, $a \cdot b^{-1} \in H$.

(ii) Si suponemos ahora que $\forall a, b \in H$, $a \cdot b^{-1} \in H$, debemos demostrar que H es un subgrupo de G. En primer lugar debemos ver que e, el neutro del grupo está en H. Si $H \neq \emptyset$ existe $a \in H$ pero entonces por hipótesis $a \cdot a^{-1} \in H$, es decir $e \in H$.

Similarmente, si $a \in H$, como $e \in H$, $e \cdot a^{-1} = a^{-1} \in H$, esto es si un elemento está en H su inverso también.

Finalmente, si $a, b \in H$, entonces $a, b^{-1} \in H$ pero por hipótesis $a \cdot (b^{-1})^{-1} = a \cdot b \in H$.

Como $(G, \cdot)$ es grupo, la operación $\cdot$ es asociativa y por lo tanto también lo es para los elementos de H

Si $a^2 = b$, $a \cdot b = a \cdot a^2 = a^3$ y $b \cdot a = a^2 \cdot a = a^3$. De donde $a \cdot b = b \cdot a$ (contradicción).

Si $a^2 = a \cdot b$ o si $a^2 = b \cdot a$, llegaríamos a que $a = b$, de nuevo una contradicción. Concluimos que en ese caso a^2 sería el sexto elemento. Si $a^2 = c$, por un análisis similar concluiríamos que $a \cdot b \cdot a$ es el sexto elemento. Es decir $o(G) \geq 6$

5.3 Subgrupos

Una pregunta interesante se refiere a si la estructura de estos grupos se conserva o no en subconjuntos arbitrarios de ellos y cómo usar estructuras conocidas para construir nuevas estructuras. Si analizamos, por ejemplo, el conjunto de los números naturales con la suma, vemos que es un subconjunto del grupo aditivo de los enteros y la suma restringida a ellos es una ley de composición interna. Sin embargo, los naturales no tienen estructura de grupo aditivo. Pero si consideramos los enteros pares con la suma (como veremos mas adelante) sí conservan la estructura. Desde luego es de esperarse que en un grupo tan familiar para nosotros, estemos en capacidad de caracterizar los subconjuntos que conservan la estructura y que para ello nos sirva precisamente la aritmética de los enteros. Los subconjuntos especiales de los grupos a los que hemos hecho referencia en este comentario son los *subgrupos*.

Definición 5.12 *Sea $(G, \cdot)$ un grupo y $H \subset G$, $H \neq \emptyset$, H es subgrupo de G si la operación de G restringida a H hace de H un grupo.*

Ejemplos

1. En el grupo $(S, \circ)$ de las simetrías del cuadrado (ejemplo 6, sección 5.2) el subconjunto $R = \{r_0, r_1, r_2, r_3\}$, donde r_0, r_1, r_2, r_3 y r_4 son respectivamente las rotaciones de 0°, 90°, 180° y 270°, $(R, \circ)$ es un grupo, esto es, R es un subgrupo de $\{S\}$. ¡Comprobarlo!

2. Los subconjuntos $\{e, a\}$, $\{e, b\}$ y $\{e, c\}$, son subgrupos del cuarto grupo de Klein cuya tabla se analizó en la sección anterior.

3. En el grupo $G = \{(x, y, z) | x, y, z \in \{1, 2, 3\}\}$ con una ley $*$ definida,

$$(x_1, y_1, z_1) * (x_2, y_2, z_2) = (x_1 + x_2 + y_2 z_1, y_1 + y_2, z_1 + z_2)$$

donde $+$ es adición módulo 3, el subconjunto

$$H = \{(0,0,0), (1,0,0), (2,0,0), (0,1,0), (0,2,0), (1,1,0)\}$$

no es un subgrupo de G pues si consideramos por ejemplo $(1,0,0) * (0,2,0) = (1,2,0) \notin H$, es decir $*$ no es una ley de composición interna en H.

Puntos de discusión

1. En el grupo S de las simetrías del cuadrado, ¿es el conjunto $\{h, v, d_1, d_2\}$ (reflexiones respecto a los ejes de simetría) un subgrupo de S? Identificar todos los subgrupos de S.

2. Sea $E = \{1, 2, 3\}$, $(P(E), \nabla)$ es un grupo. ¿Es $\{\emptyset, \{1, 2\}, \{1, 3\}\}$ un subgrupo de $P(E)$? ¿Tiene $P(E)$ subgrupos de ordenes 2,4,5 o 6?

Teorema 5.15 *Sean $(G,\cdot)$ es grupo, $H \subset G$, H finito. H es un subgrupo de G si y sólo si $\forall a, b \in H$, $a \cdot b \in H$.*

Demostración.

(i) Si H es subgrupo de G (es un grupo con la operación $\cdot$) se tiene entonces que $\cdot$ es ley interna en H, es decir $\forall a, b \in H$, $a \cdot b \in H$.

(ii) Si suponemos ahora que $\forall a, b \in H$, $a \cdot b \in H$, nos bastará demostrar que $\forall a, b \in H$, $a \cdot b^{-1} \in H$ y aplicar el Teorema 5.3.1.

Como por hipótesis $a \cdot b \in H$, $\forall a, b \in H$, se tiene que si $a \in H$, $a^2, a^3, a^4, \ldots, a^n \in H$. Pero H es un subconjunto finito de G. Deben existir entonces enteros positivos n y m con $n > m$ tales que $a^n = a^m$ (las potencias de a deben repetirse). De aquí podemos concluir que $a^{n-m} = e$ con $n - m > 0$, es decir, $e = a^{n-m} \in H$.

Si multiplicamos ahora por el inverso multiplicativo de a, a^{-1} tenemos $a^{-1} = a^{n-m-1}$, pero como $n - m > 0$ $n - m - 1 \leq 0$, nuestro objetivo es demostrar que $a^{-1} \in H$. En efecto,

Si $n - m - 1 = 0$, $a^{-1} = a^0 = e \in H$.

So $n - m - 1 > 0$, $a^{-1} = a^{n-m-1} \in H$. Considerando ahora $a, b \in H$, se tiene entonces que $a, b^{-1} \in H$ y por hipótesis $a \cdot b^{-1} \in H$, aplicando teorema 5.3.1, podemos concluir que H es subgrupo de G.

Nota: En grupos infinitos, subconjuntos infinitos cerrados para la operación, no necesariamente son subgrupos. Basta analizar por ejemplo el conjunto de los números naturales con la suma o el de los enteros con el producto.

Puntos de discusión

Sea $(G,\cdot)$ un grupo.

1. Sea $C = \{x \in G | x \cdot a = a \cdot x \forall a \in G\}$ es un subgrupo de G. ¡Demostrarlo! (este subgrupo se conoce con el nombre de centro de G).
   (a) Identificar el centro del grupo S de las simetrías del cuadrado.
   (b) ¿Cuál el centro de S_3?
   (c) ¿Cuál el centro de un grupo abeliano?
2. Sea $C(a) = \{y \in G | y \cdot a = a \cdot y\}$ $a \in G$ fijo. Demostrar que éste es también un subgrupo de G (llamado centralizador o normalizador de a).
   (a) Usar el centralizador para encontrar posibles subgrupos de S_4 (grupo de permutaciones de cuatro elementos).
   (b) Si un elemento está en el centro de un grupo G, ¿cuál es su centralizador?

Caracterización de subgrupos de $(Z,+)$

Consideremos ahora el grupo aditivo de los números enteros $(\mathbb{Z},+)$ y el subconjunto de éste, $H = 2\mathbb{Z} = \{2m | m \in \mathbb{Z}\}$. Veamos que H es subgrupo de $\mathbb{Z}$.

Dados $x, y \in H$, arbitrarios, para aplicar el teorema 5.3.1 bastará con demostrar que $x - y \in H$.

Como $x \in H$, $x = 2m$ para algún $m \in \mathbb{Z}$ y como $y \in H$, $y = 2n$ para algún $n \in \mathbb{Z}$, pero entonces $x - y = 2m - 2n = 2(m - n) \in H$, es decir H es un subgrupo de $\mathbb{Z}$.

Similarmente razonaríamos si consideramos $3\mathbb{Z}$, $4\mathbb{Z}$, $7\mathbb{Z}$, ..., $n\mathbb{Z}$, esto es, podemos demostrar que todos los subconjuntos de esta forma son subgrupos de $\mathbb{Z}$. Surge entonces una pregunta natural, ¿ tiene $(\mathbb{Z}, +)$ subgrupos que no sean de esta forma?

El siguiente teorema responde esta pregunta, caracterizando completamente los subgrupos de $\mathbb{Z}$, y en su demostración se usan fundamentalmente elementos de la teoría de números. Se aprecian pues allí (tal vez por primera vez en nuestra discusión) los fuertes nexos a los que hacíamos referencia al iniciar el capítulo.

Teorema 5.16 *Sea $(\mathbb{Z}, +)$ el grupo aditivo de los enteros. Entonces H es un subgrupo de $\mathbb{Z}$ si y sólo si existe $n \in \mathbb{N} \cup \{0\}$ tal que $H = n\mathbb{Z}$ donde $n\mathbb{Z} = \{nx | x \in \mathbb{Z}\}$.*

Demostración. Sea H un subgrupo de $\mathbb{Z}$.

Si $H = \{0\}$ entonces $H = \{0x = 0 | x \in \mathbb{Z}\} = 0\mathbb{Z}$, es decir es valida la afirmación.

Si $H \neq \{0\}$, existe $y \in H$, $y \neq 0$ pero entonces como H es subgrupo, el opuesto de y, $-y \in H$. Tiene sentido considerar entonces el conjunto

$$S = \{x > 0 | x \in H\}.$$

$S \neq \emptyset$, pues $y \in S$ y $-y \in S$. Pero por $S \subseteq \mathbb{N}$ podemos aplicar entonces el *principio de buena ordenación:* Todo subconjunto no vacío de naturales posee un elemento mínimo. Existe pues, $n = min(S)$, n es claramente mayor que 0 y $n \in H$. Nuestro objetivo, es mostrar que $H = n\mathbb{Z}$.

Veamos primero que, z es múltiplo de n, $\forall z \in H$. Como $z, n \in \mathbb{Z}$ aplicamos a este par de enteros el algoritmo de la división. Existen $q, r \in \mathbb{Z}$ tales que $z = nq + r$ con $r = 0$ o $0 \leq r < n$.

Si $r \neq 0$, $r = z - nq$, como $z \in H$ y $nq \in H$, por ser H subgrupo de $\mathbb{Z}$, $z - nq \in H$, es decir, $r \in H$ y $r > 0$ de donde $r \in S$. Pero, esto contradice la escogencia de n, pues r sería un elemento de S menor que el mínimo. De allí se sigue que $r = 0$ y todo elemento de H es múltiplo de n, esto significa que $H \subset n\mathbb{Z}$.

Por otra parte, $n\mathbb{Z} = \{nx | x \in \mathbb{Z}\} \subset H$ puesto que $n \in H$, y cualquier múltiplo entero de n estará en H por ser H subgrupo de $\mathbb{Z}$. Concluimos entonces que $H = n\mathbb{Z}$.

Nos resta demostrar que si $H = n\mathbb{Z}$ con $n \in \mathbb{N} \cup \{0\}$, H es un subgrupo de $\mathbb{Z}$. Para ello notamos que

(i) $0 = n0 \in H$.

(ii) Si $a, b \in n\mathbb{Z}$ $a = nq$ y $b = ns$ para algunos q y s en $\mathbb{Z}$, entonces $a - b = nq - ns = n(q - s) \in n\mathbb{Z}$, de donde, $n\mathbb{Z}$ es un subgrupo de $\mathbb{Z}$.

Apreciemos en qué medida esta caracterización de los subgrupos de $\mathbb{Z}$ nos

permite explorar representaciones para $\mathbb{Z}$ y generalizar ideas propias de la aritmética

Si seleccionamos un par de enteros, por ejemplo 2 y 3, su máximo común divisor (en este caso 1, dado que son primos relativos) puede ser expresado como combinación lineal de ellos, esto es, existen q y r enteros tales que

$$2q + 3r = 1.$$

Si multiplicamos la relación anterior por n, donde n es un entero arbitrario, obtenemos

$$n = 2(nq) + 3(nr) = 2t + 3s, \text{ con } t, s \in \mathbb{Z}.$$

Concluimos que cualquier $n \in \mathbb{Z}$ puede ser expresado como combinación lineal de 2 y 3, esto es $\mathbb{Z} \subset 2\mathbb{Z} + 3\mathbb{Z}$. De otra parte, es fácil demostrar que $2\mathbb{Z} + 3\mathbb{Z}$ es un subgrupo de $\mathbb{Z}$, esto es $\mathbb{Z} = 2\mathbb{Z} + 3\mathbb{Z}$ puede ser expresado en este caso como suma de dos de sus subgrupos.

Nota. Definimos $H + K = \{h + k | h \in H, k \in K\}$.

Punto de discusión

¿ Es posible expresar $\mathbb{Z}$ como suma de sus subgrupos $m\mathbb{Z}$ y $n\mathbb{Z}$ si m y n son enteros arbitrarios? Analizar condiciones.

Una de las relaciones fundamentales que se definen en el conjunto de los números enteros, y que se trabajan en otro de los capítulos de este libro, es la relación de divisibilidad, nos interesa ahora observar cómo las construcciones que se derivan de ella pueden ser ampliadas a grupos abstractos. Recordemos como se define la relación de congruencia modular en $\mathbb{Z}$

Definición 5.13 *Dado $n \in \mathbb{Z}$ fijo, $a, b \in \mathbb{Z}$ decimos que a es congruente con b módulo n, y notamos $a \equiv b(mod\ n)$, si y sólo si $n|a - b$.*

La relación $\equiv$ es reflexiva, simétrica y transitiva, como se demuestra fácilmente y es, por ello, una relación de equivalencia. Por ser $\equiv$ una relación de equivalencia podemos construir las clases de equivalencia de los elementos de $\mathbb{Z}$ según $\equiv$, de la siguiente manera.

Si $a \in \mathbb{Z}$ y n es un entero fijo, $[a] = \{x \in \mathbb{Z} | a \equiv x(\text{mod}n)\}$. Analicemos estas ideas para un entero n particular, digamos para $n = 3$. Dados $a, b \in \mathbb{Z}$, $a \equiv b(\text{mod}3)$ si y sólo si $3|a - b$; pero esto significa precisamente que $a - b = 3m$ para algún $m \in \mathbb{Z}$, de donde $a - b \in 3\mathbb{Z}$. Además, $3\mathbb{Z}$ es subgrupo de $\mathbb{Z}$ como se demostró en el Teorema 5.3.3.

Si $a \in \mathbb{Z}$, $[a] = \{x \in \mathbb{Z} | a \equiv x(3)\} = \{x \in \mathbb{Z} | a = x + 3m, m \in \mathbb{Z}\}$, aplicando el algoritmo de la división a x y a 3, concluimos que para $[a]$, se dan tan sólo tres posibilidades:

$[a] = [0] = \{0, 3, 6, 9, \ldots\}$; $[a] = [1] = \{1, 4, 7, 10, \ldots\}$; o, $[a] = [2] = \{2, 5, 8, 11, \ldots\}$. Obsérvese que una partición de $\mathbb{Z}$ es entonces $\{[0], [1], [2]\}$ pues $[0] \cup [1] \cup [2] = \mathbb{Z}$ y $[i] \cap [j] = \emptyset$ para todo $i, j = 0, 1, 2$.

Para un entero $n > 1$ arbitrario la relación $\equiv$ puede entonces definirse como sigue.

$$a \equiv b(n) \Longleftrightarrow (\exists q \in \mathbb{Z})(a - b = nq) \Longleftrightarrow a - b \in n\mathbb{Z}.$$

Lo interesante de expresarla de esta manera es que plantea un camino para generalizarla a cualquier subgrupo de un grupo abstracto tal como se muestra a continuación.

Definición 5.14 *Sean $(G, \cdot)$ un grupo y H un subgrupo de G. Se define la relación $\equiv$ en G como sigue. Para $a, b \in G$ $a \equiv b(mod H) \iff a \cdot b^{-1} \in H$ (en notación aditiva $a - b \in H$).*

Demostremos que $\equiv$ es una relación de equivalencia.

(i) Es reflexiva, pues dado $a \in G$, $a \cdot a^{-1} = e \in H$, esto es, $a \equiv a(\mathrm{mod} H)$.

(ii) Es simétrica, pues si $a \equiv b(H)$, $a \cdot b^{-1} \in H$. Como H es subgrupo de G $(a \cdot b^{-1})^{-1} \in H$, es decir $b \cdot a^{-1} \in H$, pero esto significa que $b \equiv a(\mathrm{mod} H)$.

(iii) Es transitiva, dado que si $a \equiv b(\mathrm{mod} H)$ y $b \equiv c(\mathrm{mod} H)$, entonces $a \cdot b^{-1} \in H$ y $b \cdot c^{-1} \in H$. Por ser H subgrupo, $(a \cdot b^{-1}) \cdot (b \cdot c^{-1}) \in H$. Asociando concluimos que $a \cdot c^{-1} \in H$, es decir, $a \equiv c(\mathrm{mod} H)$ por definición de $\equiv$.

La partición inducida por la anterior relación de equivalencia nos permitirá caracterizar completamente los órdenes de los posibles subgrupos de un grupo finito dado. Pero para llegar a esta idea debemos analizar primero las clases de equivalencia inducidas en G según esta relación.

Definición 5.15 *Sean $(G, \cdot)$ un grupo, H un subgrupo de G y $a \in G$. Al conjunto $[a] = \{x \in G | x \equiv a(mod H)\}$ lo llamaremos clase de equivalencia de a según $\equiv$ (o según H).*

Si retomamos el ejemplo de la congruencia módulo 3 en los enteros y la expresamos con la idea de la congruencia módulo un subgrupo, en este caso el subgrupo $H = 3\mathbb{Z}$, las clases en el grupo $\mathbb{Z}$ según este subgrupo serían:

$[0] = \{x \in \mathbb{Z} | x \equiv 0(3\mathbb{Z})\} = \{x \in \mathbb{Z} | x \in 3\mathbb{Z}\} = 3\mathbb{Z}$ que corresponde a los *múltiplos de 3.*

$[1] = \{x \in \mathbb{Z} | x \equiv 1(3\mathbb{Z})\} = \{x \in \mathbb{Z} | x - 1 \in 3\mathbb{Z}\} = \{x \in \mathbb{Z} | x = 3q + 1\}, q \in \mathbb{Z}\}$. De manera similar,

$[2] = \{x \in \mathbb{Z} | x = 3q + 2$, para algún $q \in \mathbb{Z}\}$.

Nuestro objetivo es analizar en el caso general el cardinal de cada una de las clases de equivalencia para un grupo y un subgrupo dados, para ello requerimos dar primero a estas clases una presentación alterna, por medio de la idea de clases laterales.

Definición 5.16 *Sean $(G, \cdot)$ un grupo, $a \in G$ y H un subgrupo de G. El conjunto $Ha = \{h \cdot a | h \in H\}$ se denomina clase a derecha de a según el subgrupo H.*

Veamos un ejemplo. Si tomamos $H = 3\mathbb{Z}$, sin olvidar que para este caso la notación es aditiva, tenemos

Si $a = 0$, $H + 0 = 3\mathbb{Z} + 0 = 3\mathbb{Z}$.

Si $a = 1$, $H + 1 = 3\mathbb{Z} + 1 = \{h + 1 | h \in 3\mathbb{Z}\} = \{3m + 1 | m \in \mathbb{Z}\}$.

Si $a = 2$, $H + 2 = 3\mathbb{Z} + 2 = \{h + 2 | h \in 3\mathbb{Z}\} = \{3n + 2 | n \in \mathbb{Z}\}$.

Obsérvese que $3\mathbb{Z} = [0]$, $3\mathbb{Z} + 1 = [1]$ y $3\mathbb{Z} + 2 = [2]$. En general es posible afirmar

Teorema 5.17 *Si $(G, \cdot)$ es un grupo y H es un subgrupo de G, $\forall a \in G$, $[a] = Ha$.*

Demostración. Sea $x \in [a]$, esto significa que $x \equiv a(H)$, es decir $x \cdot a^{-1} \in H$. Luego existe $h \in H$ tal que $x \cdot a^{-1} = h$. Si multiplicamos por a, $x \cdot a^{-1} \cdot a = h \cdot a$, de donde, $x = h \cdot a$, es decir $x \in Ha$. Concluimos que $[a] \subseteq Ha$.

Similarmente si tomamos $y \in Ha$, $y = h \cdot a$ para algún $h \in H$. Entonces, $y \cdot a^{-1} = h$, es decir, $y \cdot a^{-1} \in H$ y esto significa que $y \equiv a(H)$ o, equivalentemente que $y \in [a]$. Por ende, $Ha \subseteq [a]$.

Analicemos en algunos ejemplos el comportamiento de estas clases.

1. Sea $G = \{i, x, y, z\}$ donde

$$i = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}, x = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}, y = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}, z = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}.$$

$(G, \circ)$ es un grupo. Consideremos el subgrupo de G, $H = \{i, x\}$ y construyamos las clases de G según H

$$\begin{aligned} [i] &= Hi = \{i, x\} = H \\ [x] &= Hx = \{i, x\} = H \\ [y] &= Hy = \{y, z\} \\ [z] &= Hz = \{z, y\}. \end{aligned}$$

Obsérvese que todas las clases tienen dos elementos y que $\{[i], [x], [y], [z]\}$ es una partición de G.

2. Sea $G = S_3$, el grupo de permutaciones de tres elementos, que analizamos en la sección anterior, $G = \{i, \phi_1, \phi_2, \phi_3, \phi_4, \phi_5\}$ donde

$$\begin{aligned} I &= \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \phi_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \phi_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \\ \phi_3 &= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \phi_4 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \phi_5 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \end{aligned}$$

Consideremos el subgrupo $H = \{i, \phi_4, \phi_5\}$ y construimos todas las clases laterales. Se observa que, como las dos clases $[i] = Hi = H$ y $[\phi_1] = H\phi_1 = \{\phi_1, \phi_2, \phi_5\}$, todas tienen el mismo número de elementos y precisamente el número de elementos del subgrupo H. Además el conjunto formado por todas las clases de equivalencia según H es una partición de G.

Si usted explora un grupo arbitrario y construye todas las clases según un subgrupo encontrará este mismo hecho, en general

Teorema 5.18 *Si $(G, \cdot)$ es un grupo, H un subgrupo de G y $a, b \in H$, entonces $[a]$ y $[b]$ tienen el mismo cardinal. Es decir, existe una correspondencia biunívoca entre los elementos de $[a]$ y los elementos de $[b]$.*

Demostración. Sabemos por el teorema 5.3.4 que, $[a] = Ha$ y $[b] = Hb$. Si definimos

$$f : Ha \longrightarrow Hb \quad ha \longrightarrow f(ha) = hb,$$

f así definida es biyectiva. (¡Demostrar!) Se concluye entonces que Ha y Hb tienen el mismo cardinal, es decir, $[a]$ y $[b]$ tienen el mismo cardinal.

Notas.

1. Dado que $[e] = He = \{he | h \in H\} = H$, entonces, como lo observabamos ya en los ejemplos, se tiene que toda clase de equivalencia en G según H tiene $o(H)$ elementos.

2. La relación $a \equiv b(mod H)$ determina una partición de G esto es

$G = \cup[a]$, $a \in G$ y $\forall a, b \in G$ $a \neq b$, $[a] \cap [b] = \emptyset$.

Como consecuencia de los resultados anteriores es posible demostrar el primer teorema fundamental de la teoría de grupos.

Teorema 5.19 *(Teorema de Lagrange). Sean $(G, \cdot)$ un grupo finito y H un subgrupo de G. Entonces $o(H)|o(G)$.*

Demostración. Como la relación $\equiv$ determina una partición de G, cada $a \in G$ pertenece a una y sólo una clase de equivalencia según H, la clase $[a]$. Dado que G es finito, el número de estas clases es finito, digamos m. Como cada clase posee el mismo número de elementos, y , $[e]$ donde e es el neutro del grupo coincide con el subgrupo H, cada clase tiene exactamente $o(H)$ elementos. El orden del grupo es, por lo tanto, $o(G) = m \cdot o(H)$, es decir, $o(H)|o(G)$.

Una consecuencia inmediata de este teorema que usted puede analizar es que si G es un grupo de orden primo sus únicos subgrupos son G mismo y $\{e\}$.

Punto de discusión

Si $o(G) = n$ y $q|n$ el teorema anterior no garantiza la existencia de un subgrupo de orden q. Ilustrar porqué. Estudiar los ejemplos presentados inicialmente.

De otra parte, si $(G, \cdot)$ es un grupo finito, se conoce su orden y el orden de un subgrupo H de éste, es posible determinar el número de clases de equivalencia en el grupo según el subgrupo. Este número es llamado el *índice* de H en G, se denota por $i_G(H)$ o, cuando no hay lugar a confusión, $i(H)$, y se tiene entonces como corolario del teorema de Lagrange que $i(H) = \frac{o(G)}{o(H)}$. En el ejemplo 1 anterior, el $i_G(H)$ es 3, en el segundo ejemplo usted podrá concluir que es índice es 2.

Punto de discusión

1. Identificar todas las clases del subgrupo de las rotaciones en el grupo de las simetrías del cuadrado y determinar el índice de éste subgrupo.

2. En el grupo infinito $(\mathbb{Z},+)$ identificar todas las clases de los subgrupos $2\mathbb{Z}$, $5\mathbb{Z}$, $n\mathbb{Z}$, donde $n \in \mathbb{Z}$, $n > 1$. Usted puede observar que todo subgrupo distinto de los triviales es de índice finito. ¿Contradice este hecho las afirmaciones anteriores?

3. Un subgrupo del grupo $(V_3, +)$ de los vectores en el espacio, es $(V_2, +)$, vectores en el plano. Analizar por qué e identificar clases en V_3 según este subgrupo. ¿Es de índice finito?

4. Un subgrupo finito de $(\mathbb{R}, \times)$ es $(\{1, -1\}, \times)$. Si $a \in \mathbb{R}$, caracterizar $[a]$. ¿Qué concluye?

5. Considerar el conjunto de funciones de la forma $f(x) = \frac{(ax+b)}{(cx+d)}$ con $a, b, c, d \in \mathbb{R}$. Estas forman un grupo con la composición (¡Comprobarlo!). Un subgrupo de éste es el conjunto de las funciones para las cuales $c = 0$ y $d = 1$. Encontrar la clase según este subgrupo de la función $\frac{1}{x}$.

A continuación se presenta la tabla del grupo de orden seis, $(\mathbb{Z}_7^*, \cdot)$.

$\cdot$	1	2	3	4	5	6
1	1	1	3	4	5	6
2	2	4	6	1	3	5
3	3	6	2	5	1	4
4	4	1	5	2	6	3
5	5	3	1	6	4	2
6	6	5	4	3	2	1

De esta tabla podemos concluir que,

$H = \{1, 2, 2^2\}$ y $K = \{1, 4, 4^2\}$ son subgrupos de $\mathbb{Z}_7^*$ de orden 3. ¿Existen otros subgrupos de orden 3? Nótese que $2^2 = 4$, $2^3 = 1$, $2^4 = 2$, $2^5 = 4$ y que $4^2 = 2$, $4^3 = 1$, $4^4 = 4$, etc.

$T = \{1, 6\}$ es subgrupo de orden 2. (¿Hay mas subgrupos de orden 2?) Y,

$$\mathbb{Z}_7^* = \{1, 3, 3^2, 3^3, 3^4, 3^5\} = \{1, 5, 5^2, 5^3, 5^4, 5^5\}.$$

Respecto al índice de estos subgrupos, observamos que, $i(H) = \frac{o(\mathbb{Z}_7^*)}{o(H)} = \frac{6}{3} = 2 = i(K)$. Esto significa que hay 2 clases distintas de $\mathbb{Z}_7^*$ según H, y dos clases según K. Como $i(T) = \frac{6}{2} = 3$, hay tres clases distintas de $\mathbb{Z}_7$ según T. ¡Identificar completamente estas clases!

Obsérvese además que para cada elemento $a \in \mathbb{Z}_7^*$ existe $k_i \in \mathbb{N}$ tal que $a^{k_i} = 1$. En efecto,

$a = 1$, $k_1 = 1$; $a = 2$, $k_2 = 3$; $a = 3$, $k_3 = 6$; $a = 4$, $k_4 = 3$; $a = 5$, $k_5 = 6$ y $a = 6$, $k_6 = 2$.

Nótese que para $a = 3$ y $a = 5$, $k_3 = k_5 = 6 = o(\mathbb{Z}_7^*)$, mientras que en los otros casos $k_i | 6$.

5.3.1 Grupos cíclicos

El análisis que hicimos con el grupo $(\mathbb{Z}_7^*, .)$ motiva la presentación del siguiente teorema.

Teorema 5.20 *Sean $(G, \cdot)$ un grupo y $a \in G$ un elemento para el que existen $n, m \in N$, $n > m$ tales que $a^n = a^m$. Entonces existe $m_0 \in \mathbb{N}$ tal que*

i. $a^{m_0} = e$ y si $0 \leq i < j < m_0$, $a^i \neq a^j$;
ii. Si $a^t = e$, $t \in \mathbb{Z}$ entonces $m_0 | t$;
iii. $\{a^k | k \in \mathbb{Z}\} = \{e, a, a^2, \cdots, a^{m_0 - 1}\}$.

Demostración.

(i) Si existen $n, m \in \mathbb{N}$, $n > m$ tales que $a^n = a^m$, $a^{n-m} = e$, $n - m > 0$. El conjunto $T = \{t \in \mathbb{Z}^+ | a^t = e\}$, es entonces no vacío y podemos aplicar de nuevo aquí el *principio de buena ordenación*. Sea $m_0 = \min(T)$, $a^{m_0} = e$ y $m_0 \geq 1$.

Si $0 \leq i < j < m_0$ y suponemos que $a^j = a^i$, entonces $a^{j-i} = e$. Como $j - i > 0$, $j - i \in T$, pero esto no es posible ya que $j - i < m_0$ y $m_0 = \min(T)$. Se concluye entonces que $a^j \neq a^i$.

(ii) Sean $t \in \mathbb{Z}$ y $a^t = e$. Supongamos que m_0 no divide a t. Aplicando entonces a t y a m_0 el algoritmo de la división, tenemos que existen $q, r \in \mathbb{Z}$ tales que

$$t = m_0 q + r, \text{donde } 0 < r < m_0.$$

Pero entonces $t - m_0 q = r > 0$ y $a^t = a^{m_0 q + r} = a^{m_0 q} \cdot a^r = a^r$. Como $a^t = e$ concluimos que $a^r = e$. Esta última igualdad contradice la escogencia de $m_0 = \min(T)$, pues estamos suponiendo $0 < r < m_0$. Se tiene entonces que $r = 0$, esto es $t = m_0 q$, o lo que es lo mismo $m_0 | t$.

(iii) Consideramos ahora $k \in \mathbb{Z}$ y aplicamos el algoritmo de la división a m_0 y a k

$$k = m_0 q + r, \text{con } 0 \leq r < m_0.$$

Se tiene que

$$a^k = a^{m_0 q + r} = a^{m_0 q} \cdot a^r = a^r, \text{con } 0 \leq r < m_0.$$

Concluimos que $\{a^k | k \in \mathbb{Z}\} \subseteq \{e, a, a^2, \ldots, a^{m_0 - 1}\}$, desde luego se tiene también que $\{e, a, a^2, \ldots, a^{m_0 - 1}\} \subseteq \{a^k | k \in \mathbb{Z}\}$,esto es $\{a^k | k \in \mathbb{Z}\} = \{e, a, a^2, \ldots, a^{m_0 - 1}\}$.

Veamos ahora que $\{a^k | k \in \mathbb{Z}\}$ es el menor subgrupo de $(G, \cdot)$ que contiene a a.

Teorema 5.21 *Sean $(G, \cdot)$ un grupo y $a \in G$. Entonces $H = \{a^m | m \in \mathbb{Z}\}$ es un subgrupo de G y si K es un subgrupo de G y $a \in K$ entonces $H \subseteq K$.*

Demostración. Sean $x, y \in H$, $x = a^{m_1}$ y $y = a^{m_2}$, $m_1, m_2 \in \mathbb{Z}$. $y^{-1} = a^{-m_2} \in H$ pues $-m_2 \in \mathbb{Z}$, y $x \cdot y^{-1} = a^{m_1} \cdot a^{-m_2} = a^{m_1 - m_2} \in H$, dado que $m_1 - m_2 \in \mathbb{Z}$.

Sean ahora K un subgrupo de G y $a \in K$. Como K es subgrupo, $a^m \in K$ para todo $m \in \mathbb{Z}$. Pero entonces $\{a^m | k \in \mathbb{Z}\} \subseteq K$, es decir, $H \subseteq K$.

Definición 5.17 *El subgrupo H del teorema anterior se llama subgrupo cíclico generado por a y suele notarse*

$$\langle a \rangle = \{a^m | m \in \mathbb{Z}\}.$$

Definición 5.18 *Un grupo G se dice cíclico si $G = \langle a \rangle$ para algún a en G.*

En el ejemplo preliminar a esta sección, $(\mathbb{Z}_7^*, \cdot)$ es un grupo cíclico finito de orden 6 , $\mathbb{Z}_7^* = \langle 3 \rangle = \{1, 3, 3^2, 3^3, 3^4, 3^5\} = \langle 5 \rangle$ y los subgrupos H, K y, T son también cíclicos, $H = \langle 2 \rangle$, $K = \langle 4 \rangle$, de orden 3, y $T = \langle 6 \rangle$ de orden 2.

El grupo, $(\mathbb{Z}, +)$ es también cíclico pero de orden infinito, $\mathbb{Z} = \langle 1 \rangle$.

Teorema 5.22 *Sea $G = \langle a \rangle$ un grupo cíclico finito de orden n. Entonces $G = \{e, a, a^2, \ldots, a^{n-1}\}$.*

Demostración. Como $o(G) = n < \infty$, entonces no todas las potencias de a son distintas, es decir, existen i, j enteros positivos, $i < j$, tales que $a^i = a^j$. Se sigue que $a^{j-i} = e$, con $j - i > 0$. Por el Teorema 5.3.7, existe $m_0 \in \mathbb{N}$ tal que $m_0 = \min\{t \in \mathbb{N} | a^t = e\}$, con la propiedad de que $G = \{e, a, a^2, \cdots, a^{m_0-1}\}$. Pero estas potencias son distintas y como $o(G) = n$, entonces $m_0 = n = o(G)$, de donde $G = \{e, a, a^2, \cdots, a^{n-1}\}$.

Puntos de discusión

1. Considerar el grupo de las raíces séptimas de la unidad,

$$G = \{x \in \mathbb{C} | x^7 = 1\},$$

   con el producto usual de complejos. $(G, \cdot)$ es un grupo cíclico. Encontrar un elemento $a \in G$, que genere este grupo. ¿Existe solamente un elemento en el grupo que cumple esta condición?

2. ¿ Es el grupo $(S_3, \circ)$ cíclico?

3. Considerar el grupo de simetrías del triángulo equilatero y explorar sus subgrupos. ¿Tiene este grupo un subgrupo cíclico?

4. Sea $f(x) = \frac{1}{2} - x$.

   (a) Identificar $k \in \mathbb{N}$ tal que $f^k(x) = x$ para todo $x \in \mathbb{R}$ donde $f^k = f \circ f \circ f \circ \cdots \circ f$, k veces.

   (b) Usar el hecho anterior para construir un grupo cíclico del cual $f(x)$ sea un generador y caracterizar subgrupos. ¿Son todos cíclicos?

5. Si todos los subgrupos no triviales de un grupo finito son cíclicos, ¿es el grupo cíclico?

De nuevo, como en gran parte de los resultados que hemos discutido en esta sección requerimos para caracterizar completamente los grupos y subgrupos cíclicos finitos, propiedades fundamentales de los números enteros.

5.3.2 Caracterización de los subgrupos de un grupo cíclico.

Teorema 5.23 *Todo subgrupo H de un grupo cíclico G es cíclico.*

Demostración. Sean $G = \langle a \rangle = \{a^n | n \in \mathbb{Z}\}$ y H un subgrupo de G.

(i) Si $H = \{e\}$, H es cíclico pues $e^n = e$, $\forall n \in \mathbb{Z}$.

(ii) Si $H \neq \{e\}$, sea $g \in H$, $g \neq e$. Existe entonces $m \in \mathbb{Z}$ tal que $g = a^m$ con $m \neq 0$. Por ser H subgrupo, $g^{-m} \in H$ y el conjunto $\{\alpha \in \mathbb{N} | a^\alpha \in H\}$ es no vacío. De nuevo por el *principio de buena ordenación* existe $m_0 = \min\{\alpha \in \mathbb{N} | a^\alpha \in H\}$.Veamos que $H = \langle a^{m_0} \rangle$. Tenemos

$$\langle a^{m_0} \rangle = \{a^{m_0 k} | k \in \mathbb{Z}\}.$$

Como H es subgrupo de G y $a^{m_0} \in H$, entonces $a^{m_0 k} \in H$, $\forall k \in \mathbb{Z}$, es decir, $\langle a^{m_0} \rangle \subseteq H$.

Tomemos ahora $a^z \in H$. Como $m_0 > 0$, aplicamos el algoritmo de la división a z y a m_0. Existen $q, r \in \mathbb{Z}$|

$$z = m_0 q + r, \text{con } 0 \leq r < m_0.$$

Como $a^{m_0 q} \in H$ entonces $a^{-m_0 q} \in H$, de donde, $a^z \cdot a^{-m_0 q} \in H$ y, de aquí, $a^{z-m_0 q} = a^r \in H$. Pero, por la escogencia de m_0, r debe ser 0, de donde, $z = m_0 q$ y $a^z = a^{m_0 q} \in \langle a^{m_0} \rangle$, es decir, $H \subseteq \langle a^{m_0} \rangle$. Concluimos entonces que $H = \langle a^{m_0} \rangle$, esto es, H es cíclico.

1. (a) Si $G = \langle a \rangle$ y H es subgrupo de G, $H \neq e$, entonces $H = \langle a^{m_0} \rangle$, donde $m_0 = \min\{m \in \mathbb{N} | a^m \in H\}$.
   Si G es un grupo cíclico infinito, todos los subgrupos diferentes de $\{e\}$, son también infinitos.

Punto de discusión

Demostrar el corolario y analizar sus implicaciones.

Consideremos ahora el grupo de los cuaterniones, al que hicimos referencia ya anteriormente, $Q = \{1, i, j, k, -1, -i, -j, -k\}$ con una ley notada $\cdot$ que satisface

$$i^2 = j^2 = k^2 = -1, i \cdot j = k, j \cdot k = i, k \cdot i = j, j \cdot i = -k, i \cdot k = -j \text{ y } k \cdot j = -i.$$

1 es el elemento neutro. A continuación aparece la tabla correspondiente.

$\cdot$	1	i	j	k	-1	$-i$	$-j$	$-k$
1	1	i	j	K	-1	$-i$	$-j$	$-k$
i	i	-1	k	$-j$	$-i$	1	$-k$	j
j	j	$-k$	-1	i	$-j$	k	1	$-i$
k	k	j	$-i$	-1	$-k$	$-j$	i	1
-1	-1	$-i$	$-j$	$-k$	1	i	j	k
$-i$	$-i$	1	$-k$	j	i	-1	k	$-j$
$-j$	$-j$	k	1	$-i$	j	$-k$	-1	j
$-k$	$-k$	$-j$	i	1	k	j	$-i$	1

Q es un grupo no abeliano de orden 8; los subgrupos de Q son

$$H_0 = \{1\}, H_1 = \{1, -1\} = \langle -1 \rangle, H_2 = \{1, i, -i, -1\} = \{1, i, i^2, i^3\} = \langle i \rangle,$$
$$H_3 = \{1, j, -j, -1\} = \{1, j, j^2, j^3\} = \langle j \rangle \text{ y } H_4 = \{1, k, -k, -1\} = \{1, k, k^2, k^3\} = \langle k \rangle.$$

Nótese que los subgrupos H_0, H_1, H_2, H_3, y H_4 son cíclicos. ¿Es Q cíclico? Obsérvese además que $(-1)^2 = 1$, $i^4 = j^4 = k^4 = 1$, y que para cada uno de estos elementos no existe una potencia menor de éste que nos dé el neutro del grupo.

i. **Definición 5.19** *Sean $(G, \cdot)$ un grupo y $a \in G$. El orden de a, notado $o(a)$ es el menor entero positivo m tal que $a^m = e$, si tal entero existe. Si no existe, decimos que $o(a) = +\infty$.*

En otras palabras, $o(a) = \min\{m \in \mathbb{N} | a^m = e\}$, si este conjunto es no vacío. En el ejemplo anterior $o(-1) = 2$, $o(i) = o(j) = o(k) = 4$.

Del teorema anterior se deduce,

Teorema 5.24 $o(a) = o(\langle a \rangle)$.

Corolario 5.1 *Si $(G, \cdot)$ es un grupo finito y $a \in G$ entonces $o(a)|o(G)$.*

Demostración. Basta considerar el subgrupo $\langle a \rangle = \{a^k | k \in \mathbb{N}\}$. Por el Teorema de Lagrange, $o(\langle a \rangle)|o(G)$, pero entonces $o(a)|o(G)$.

Corolario 5.2 *Si $(G, \cdot)$ es finito y $a \in G$, entonces $a^{o(G)} = e$.*

Demostración. Por el corolario anterior, existe $t \in \mathbb{Z}$ tal que $o(G) = t \cdot o(a)$. Pero entonces $a^{o(G)} = a^{t \cdot o(a)} = (a^{o(a)})^t = e^t = e$.

El Corolario 5.2 tiene interesantes aplicaciones en la teoría de números, que abordaremos una vez aclaremos unas ideas acerca de los enteros modulares.

Anteriormente anotamos que, si en el conjunto $\mathbb{Z}_m = \{1, 2, 3, \ldots, m-1\}$ consideramos la adición módulo m, $\mathbb{Z}_m$ es un grupo conmutativo. Esto no es cierto en general, si sobre los elementos no nulos de $\mathbb{Z}_m$, definimos la multiplicación módulo m. Para ver esto, analicemos $\mathbb{Z}_6^* = \{1, 2, 3, 4, 5, 6\}$. No todo elemento de este conjunto es inversible para el producto; por ejemplo, no existe en $\mathbb{Z}_6^*$ un elemento que multiplicado (módulo 6) por 2 dé como resultado 1, el elemento neutro de la multiplicación. Pero si seleccionamos un subconjunto especial de $\mathbb{Z}_6^*$, a saber, $S = \{1, 5\}$ y restringimos la multiplicación módulo 6 a este conjunto, $(S, \cdot)$ es un grupo de orden 2.

Veamos ahora qué pasa con el conjunto $\mathbb{Z}_5^* = \{1, 2, 3, 4\}$; éste sí es un grupo con el producto módulo 5 y tiene orden 4.

Los ejemplos anteriores nos sugieren que existen condiciones bajo las cuales uno de estos subconjuntos es un grupo; éstas se expresan en el siguiente teorema.

Teorema 5.25 *Sean $m \in \mathbb{N}$, $m > 1$ en $\mathbb{Z}_m$, $S \subset \mathbb{Z}_m$ el conjunto de todos las clases de equivalencia 1 y $m-1$ tales que sus representantes son primos relativos con m. Entonces S forma un grupo bajo la multiplicación módulo m y su orden es $\phi(m)$ (donde ϕ representa la función ϕ de Euler).*

Demostración.

(i) Veamos que S es cerrado para el producto. Sean $1 \leq a_j \leq m$, $j = 1, 2$. Nuestro objetivo es demostrar que si $a_1, a_2 \in S$ entonces $a_1 \cdot a_2 \in S$.

Si $mcd(a_1, m) = 1$ y $mcd(a_2, m) = 1$, se sigue que existen $\alpha_1, \alpha_2, \beta_1, \beta_2 \in \mathbb{Z}$ tales que

$$\alpha_1 m + \beta_1 a_1 = 1 \quad \text{y} \quad \alpha_2 m + \beta_2 a_2 = 1.$$

Luego,

$$(\alpha_1 m + \beta_1 a_1)(\alpha_2 m + \beta_2 a_2) = 1,$$

de donde,

$$(\beta_1 \beta_2)(a_1 a_2) + m(\alpha_1 \alpha_2 m + \alpha_1 \beta_2 a_2 + \alpha_2 \beta_1 a_1) = 1.$$

Entonces existen $\beta = \beta_1 \beta_2$ y $\alpha = (\alpha_1 \alpha_2 m + \alpha_1 \beta_2 a_2 + \alpha_2 \beta_1 a_1)$ tales que

$$\beta a_1 a_2 + \alpha m = 1.$$

Es decir, $mcd(a_1 a_2, m) = 1$. •

De otra parte, si $a_3 \in \mathbb{Z}$ y $1 \leq a_3 \leq m$ es tal que $a_1 \cdot a_2 = a_3$, veamos que, $mcd(a_3, m) = 1$.

Si $d = mcd(a_3, m)$, $d|a_3$ y $d|m$. Como $a_1 \cdot a_2 = a_3$, se sigue que $d|a_1a_2 - a_3$. Además, como $d|a_3$, concluimos que $d|a_1a_2$. Pero, como $d|m$ y $mcd(a_1a_2, m) = 1$, se sigue que $d|1$, esto es, $d = 1$, de donde a_3 es primo relativo con 1.

(ii) $1 \in S$, pues 1 es primo relativo con m.

(iii) El producto módulo m es asociativo en S, pues es asociativo en $\mathbb{Z}_m$.

(iv) Nos resta ver que todo elemento de S es inversible. Para ello sea $0 < x < m$ tal que $mcd(x, m) = 1$. Como x y m son primos relativos, existen α y β enteros tales que

$$\alpha x + \beta m = 1, \alpha x - 1 = -\beta m,$$

esto es, $m|\alpha x - 1$, lo que es lo mismo, $\alpha x - 1 \equiv 0(\mathrm{mod} m)$, $\alpha x \equiv 1(\mathrm{mod} m)$, o equivalentemente, $\alpha \cdot x = 1$ (multiplicación módulo m). α es entonces el inverso multiplicativo de x.

Por todo lo anterior concluimos que $(S, \cdot)$ es un grupo cuyo orden es claramente $\phi(m)$, ya que esta función de Euler se define precisamente como el número de enteros menores que y primos relativos con m.

En $\mathbb{Z}_6$, $S = \{1, 5\}$, S tiene $\phi(6) = 2$ elementos. Además, el inverso de 5 es 5, pues $1 = mcd(5, 6) = 5 \cdot 5 + -4 \cdot 6$.

Puntos de discusión

1. En $\mathbb{Z}_7$, $S = \{1, 2, 3, 4, 5, 6\}$ tiene $\phi(7) = 6$ elementos. Explorar los inversos multiplicativos de los elementos de S en este caso. Encontrar un elemento $m \in S$ tal que $\langle m \rangle = S$
2. Considerar el grupo $\mathbb{Z}_p^* = \{1, 2, 3, \ldots, p-1\}$, con la multiplicación módulo p, p primo. Identificar todos los $x \in \mathbb{Z}_p^*$ tales que $x^2 = 1$.
3. ¿La tabla del grupo $\mathbb{Z}_5^*$ es idéntica a la del cuarto grupo de Klein?

Corolarios del Teorema 5.25 son el Teorema de Euler y el Teorema de Fermat a los cuales se hace referencia en otro capítulo de este libro (solución de congruencias). Importantes teoremas de la teoría de números se demuestran usando exclusivamente herramientas de la teoría de grupos.

Corolario 5.3 *(Teorema de Euler). Sean $a, m \in \mathbb{Z}$ con $m > 0$. Entonces $a^{\phi(m)} \equiv 1(mod m)$ si $mcd(a, m) = 1$.*

Demostración. Por análisis anterior, el orden del grupo S es $\phi(m)$ y, por ello, si $0 < a < m$, $a^{\phi(m)} \equiv 1(\mathrm{mod} m)$.

Si $a > m$ y $mcd(a, m) = 1$, aplicando el algoritmo de la división a a y a m tenemos $a = qm + r$, $0 \leq r < m$, el resto de dividir a por m es r. Como $mcd(a, m) = 1$, entonces $mcd(m, r) = 1$, de donde, $a^{\phi(m)} \equiv r^{\phi(m)} \equiv 1(\mathrm{mod} m)$.

Corolario 5.4 *(Teorema de Fermat). Si p es primo entonces $a^p \equiv a(mod p)$ para todo $a \in \mathbb{Z}$.*

Demostración.

(i) Si p es primo y $mcd(a,p) = 1$, $\phi(p) = p-1$ y por teorema de Euler $a^{\phi(p)} = a^{p-1} \equiv 1(p)$.

(ii) Si a y p no son primos relativos, $a = pq$ para algún $q \in \mathbb{Z}$ entonces $p|a$ y $p|a^p$, es decir, $p|a^p - a$ y esto significa que $a^p \equiv a(mod p)$.

El otro corolario que se deduce del Teorema 5.25 caracteriza completamente los grupos de orden primo.

Corolario 5.5 *Si $(G,\cdot)$ es un grupo tal que $o(G) = p$, p primo entonces*

i. *G es cíclico.*

ii. *Los únicos subgrupos de G son G y $\{e\}$.*

Demostración. La afirmación (ii) es consecuencia del Teorema de Lagrange pues los únicos divisores positivos de p son p y 1, y por tanto si H es un subgrupo de G, $o(H) = 1$, o, $o(H) = p$.

Veamos que G es cíclico. Como $o(G) = p$, primo, $p > 1$, $G \neq \{e\}$. Existe entonces $a \in G$, $a \neq e$. Si consideramos el subgrupo cíclico generado por a, este subgrupo tiene orden p, de donde $G = \langle a \rangle$, o sea, G es cíclico.

Consideremos ahora $U = \{z \in \mathbb{C} | z^3 = 1\}$, el conjunto de las raíces cúbicas de la unidad. Como hemos visto, $(U,\cdot)$ es un grupo con el producto usual de complejos y $U = \{1, e^{\frac{2\pi i}{3}}, e^{\frac{4\pi i}{3}}\}$. Los únicos subgrupos de U son $\{1\}$ y U. Si hacemos $w = e^{\frac{2\pi i}{3}}$, observamos que $w^2 = e^{\frac{4\pi i}{3}}$ y $w^3 = 1$, es decir $U = \langle w \rangle$ es cíclico de orden 3.

Análogamente si $U = \{z \in \mathbb{C} | z^p = 1\}$, p primo, U, el grupo de las raíces p-ésimas de la unidad, es un grupo con el producto usual de complejos, cuyos únicos subgrupos son los triviales. ¡Mostrar que U es cíclico de orden p e identificar un generador!

El recíproco del Corolario 5.5 también se cumple. Veamos

Corolario 5.6 *Sea $(G,\cdot)$ un grupo, $G \neq \{e\}$, tal que sus únicos subgrupos son G y $\{e\}$. Entonces G es cíclico y $o(G)$ es primo.*

Demostración. Si consideramos $a \in G$, $a \neq e$, nuevamente $\langle a \rangle$ es un subgrupo distinto de $\{e\}$. Concluimos que $G = \langle a \rangle$ es cíclico. Antes de demostrar que $o(G)$ es primo, demostremos que $o(G) < \infty$.

Supongamos que G es infinito. Como $G = \langle a \rangle$, esto significa que $\forall n \in \mathbb{N}$ $a^n \neq e$. En particular, $a^2 \neq e$ y $a \neq a^{2m}$, $\forall m \in \mathbb{Z}$. Pero entonces $a \notin \langle a^2 \rangle$, de donde $\langle a^2 \rangle$ es un subgrupo no trivial tal que $\{e\} \subset \langle a^2 \rangle \subset \langle a \rangle = G$. Esto contradice la hipótesis de que los únicos subgrupos son los triviales.

Supongamos ahora que $o(G) = n = rs$, con $0 < r < n$ y $0 < s < n$; entonces $\langle a^r \rangle$ es un subgrupo de G que posee s elementos, $a^r, a^{2r}, a^{3r}, \ldots, a^{sr}$, pero esto contradice de nuevo la hipótesis sobre los subgrupos de G.

5.4 Generación de Grupos

Dados subgrupos de un grupo, ¿es posible usarlos para construir nuevos subgrupos? Si en $(\mathbb{Z},+)$, tomamos por ejemplo los subgrupos $H = 3\mathbb{Z}$ y $K = 5\mathbb{Z}$, ¿cómo construir con ellos un nuevo subgrupo de $\mathbb{Z}$?

Una sugerencia natural es considerar operaciones elementales de conjuntos, unión e intersección. Veamos.

$$H \cup K = \{x \in \mathbb{Z} | x \in 3\mathbb{Z}, \text{o}, x \in 5\mathbb{Z}\};$$

$$H \cup K = \{x \in \mathbb{Z} | x = 3m, \text{o}, x = 5n, m, n \in \mathbb{Z}\}.$$

Basta considerar $3 \in H$ y $5 \in K$ y su suma $3 + 5 = 8 \notin H \cup K$. Se sigue que $H \cup K$ no es un subgrupo de $(\mathbb{Z}, +)$. En este caso podría plantearse la pregunta, ¿es posible construir un subgrupo propio de $(\mathbb{Z}, +)$ que contenga a $H \cup K$?

Pero si ahora consideramos la intersección,

$$H \cap K = \{x \in \mathbb{Z} | x = 3m, \text{y}, x = 3n, m, n \in \mathbb{Z}\},$$

tenemos que x es múltiplo común de 3 y 5. El mínimo común múltiplo de 3 y 5, 15, es tal que, $H \cap K = 15\mathbb{Z}$ ¡Demostrar! Tenemos, entonces, que $H \cap K$ es un subgrupo de $(\mathbb{Z}, +)$.

Consideremos ahora $(Q, .)$, el grupo de los cuaterniones,

$$Q = \{\pm 1, \pm i, \pm j, \pm k\}$$

y los subgrupos de Q, $H_0 = \{1\}$, $H_1 = \{1, -1\}$, $H_2 = \{1, i, -i, -1\}$, $H_3 = \{1, j, -1, -j\}$ y $H_4 = \{1, k, -1, -k\}$. En este caso $H_i \cap H_j$ es siempre un subgrupo de Q. Además, $H_1 \cup H_2 = H_2$ es subgrupo de Q, pero $H_2 \cup H_3 = \{\pm 1, \pm i, \pm j\}$ no lo es.

Seleccionemos ahora un subconjunto arbitrario de Q, digamos, $T = \{-1\}$. Dicho conjunto está contenido en los subgrupos H_1, H_2, H_3, H_4 y Q; además, el subgrupo H_1 está contenido en todos ellos, es el menor subgrupo de $(Q, \cdot)$ que contiene a T. Un análisis similar con el subconjunto $\{1, j\}$ nos conduce concluir que tan solo los subgrupos H_3 y Q lo contienen y que la intersección entre ellos, H_3, es el menor subgrupo que lo contiene.

Los ejemplos anteriores motivan una idea importante en cualquier estructura algebraica, en particular en este caso en los grupos, en relación a la cuestión de construir nuevos subgrupos a partir de subgrupos conocidos y aún mas general como construirlos a partir de un subconjunto arbitrario del grupo.

Para formalizar esta idea requerimos una definición y un teorema que ya seguramente se ha intuido a partir de los ejemplos.

Definición 5.20 *Sea $(G, \cdot)$ un grupo. Un subgrupo H de G, $H \neq G$ se dice maximal, si entre el y el grupo total no es posible encontrar un subgrupo de G, es decir, si K es un subgrupo de G tal que $H \subseteq K \subseteq G$, entonces $K = H$ o $K = G$ (no existen subgrupos intermedios).*

Nótese que esto ocurre con el subgrupo $3\mathbb{Z}$ de $(\mathbb{Z}, +)$. Si suponemos que K es un subgrupo de $(\mathbb{Z}, +)$ tal que $3\mathbb{Z} \subseteq K \subseteq \mathbb{Z}$, por ser subgrupo de $\mathbb{Z}$, $K = m\mathbb{Z}$, para algún $m \in \mathbb{Z}$. Pero entonces dado $3s \in 3\mathbb{Z}$, $3s = mn$, para algún $n \in \mathbb{Z}$ y esto nos lleva a concluir que o bien $K = 3\mathbb{Z}$ o $K = \mathbb{Z}$. Esto es $3\mathbb{Z}$ es subgrupo maximal.

Punto de discusión

Mostrar que en $(\mathbb{Z}, +)$ un subgrupo $H = n\mathbb{Z}$ es maximal si y sólo si n es primo.

Teorema 5.26 *Dado $(G,\cdot)$ un grupo. Si H y K son subgrupos de G entonces $H \cap K$ es subgrupo de G.*

Demostración. Basta considerar $x, y \in H \cap K$ y demostrar que $x \cdot y^{-1} \in H \cap K$. Pero si $x, y \in H \cap K$, $x, y \in H$ y $x, y \in K$. Como H es subgrupo de G, $x \cdot y^{-1} \in H$, análogamente $x \cdot y^{-1} \in K$, de donde $x \cdot y^{-1} \in H \cap K$, o sea, $H \cap K$ es subgrupo de G.

El resultado anterior puede ser generalizado a una familia de subgrupos de $(G,\cdot)$, a saber, Si $(H_i)_{i\in I}$ es una familia de subgrupos de G entonces $\bigcap_i H_i$ es un subgrupo de G. (El argumento de la demostración es completamente similar a la demostración del teorema anterior.)

El teorema que aparece a continuación nos garantiza la posibilidad de construir subgrupos de un grupo a partir de un subconjunto de éste, sinembargo no nos da una presentación manejable en la práctica de tal modo que se mejorará la presentación en resultados posteriores.

Teorema 5.27 *Sean $(G,\cdot)$ un grupo y $A \subseteq G$. Existe un subgrupo H de G que contiene a A y es el menor subgrupo que lo contiene. (Es decir, si K es un subgrupo de G tal que $A \subseteq K$ entonces $H \subseteq K$).*

Demostración. Consideramos $(H_i)_{i\in I}$, la familia de todos los subgrupos de G que contienen a A. Podemos garantizar que esta familia es no vacía pues por lo menos G está en ella.

Consideramos $H = \bigcap_i H_i$. H es un subgrupo de G y como $A \subset H_i$ para todo i, $A \subset H$. H es claramente el menor subgrupo de G que contiene a A.

Definición 5.21 *El subgrupo H del teorema anterior se denomina el subgrupo generado por A y se nota $\langle A\rangle$.*

Punto de discusión

Explorar con ejemplos y demostrar estas propiedades del subgrupo generado que usaremos posteriormente.

i. $A = \langle A\rangle$ si y sólo si A es subgrupo de G.
ii. Dados $A, B \subseteq G$. Si $A \subseteq B$ entonces $\langle A\rangle \subseteq \langle B\rangle$.
iii. Dados $A, B \subseteq G$, $\langle A\rangle \cup \langle B\rangle \subseteq \langle A \cup B\rangle$, y $\langle A \cap B\rangle \subseteq \langle A\rangle \cap \langle B\rangle$.
iv. Si $A = \{a\}$, $\langle A\rangle = \langle a\rangle$.

Respecto a (iv), nótese que como $\langle A\rangle = \langle\{a\}\rangle$ es el menor subgrupo que contiene a $\{a\}$ éste es precisamente el cíclico generado por a.

Retomemos el grupo de los cuaterniones $(Q,\cdot)$ y en el los subconjuntos $A = \{j\}$ y $A' = \{-j\}$ y formamos

$$F = \{a_1 a_2 a_3 \cdots a_n | a_i \in A \cup A', n \in N\}$$

$$F = \{j \cdot j, j \cdot -j, -j \cdot -j, \ldots\} = \{j^2, -j^2, j^3, -j^3, \ldots\} = \{-1, 1, j, -j\}.$$

Observamos que F es subgrupo de Q, contiene a A y coincide con $\langle A\rangle$. En este ejemplo se muestra otra forma de construir el subgrupo generado. Veamos la idea con mayor generalidad.

Consideremos $A \subseteq G$, $A \neq \emptyset$. Como G es un grupo, todo elemento de A tiene inverso; sea entonces $A^{-1} = \{a^{-1} | a \in A\}$ y construimos

$$F = \{a_1 a_2 a_3 \cdots a_n | a_i \in A \cup A^{-1}, n \in \mathbb{N}\}.$$

F está pues formado por productos finitos de elementos de $A \cup A^{-1}$. Se afirma que F es precisamente el subgrupo generado por A, esto es el menor subgrupo que contiene a A.

Demostremos primero que F es subgrupo de G. Para ello consideramos, $x, y \in F$

$$x = a_1 a_2 \cdots a_n, a_i \in A \cup A^{-1}$$

y

$$y = b_1 b_2 \cdots b_m, b_j \in A \cup A^{-1}.$$

$y \in G$ y es en consecuencia inversible. $y^{-1} = b_m^{-1} b_{m-1}^{-1} \cdots b_2^{-1} b_1^{-1}$, producto finito de elementos de $A \cup A^{-1}$. Por tanto $x \cdot y^{-1} = a_1 \cdots a_n \cdot b_m^{-1} \cdots b_1^{-1}$, que es de nuevo producto finito de elementos de $A \cup A^{-1}$, es decir $x \cdot y^{-1} \in F$.

Es claro además que $A \subseteq F$, pues si $a \in A$, $a = a$, producto finito de elementos de $A \cup A^{-1}$, esto es $a \in F$.

Como es el menor subgrupo que contiene a A y F es un subgrupo que contiene a A, entonces $\langle A \rangle \subseteq F$; de otra parte $A \cup A^{-1} \subseteq \langle A \rangle$ y $\langle A \rangle$ es subgrupo, $F \subseteq \langle A \rangle$, entonces $F = \langle A \rangle$.

Notas.

i. Si $A \neq \emptyset$, $a \subseteq G$, $\langle A \rangle = \{a_1^{m_1} \cdot a_2^{m_2} \cdots a_n^{m_n} | n \in \mathbb{N}, a_i \in A, m_i = \pm 1\}$.

ii. Si $(G, \cdot)$ es abeliano y $A \subseteq G$, $A \neq \emptyset$,

$$\langle A \rangle = \{a_1^{m_1} \cdot a_2^{m_2} \cdots a_n^{m_n} | n \in \mathbb{N}, a_i \in A, m_i \in \mathbb{Z}\}.$$

iii. Si la operación de G es aditiva y $(G, +)$ es abeliano, $\langle A \rangle = \{m_1 a_1 + m_2 a_2 + \cdots + m_n a_n \mid n \in \mathbb{N}, m_i \in \mathbb{Z}, a_i \in A\}$.

iv. Si exploramos de nuevo el grupo no conmutativo de los cuaterniones, $Q = \{\pm 1, \pm i, \pm j, \pm k\}$, este grupo es generado por $A = \{i, j, k\}$, es decir $\langle \{i, j, k\} \rangle = Q$.

v. En el grupo no abeliano $(S_3, \circ)$, si consideramos

$$\phi_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \phi_5 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

Podemos observar que,

$$S_3 = \{e = \phi_3^2, \phi_3, \phi_5, \phi_3 \circ \phi_5, \phi_5 \circ \phi_3, \phi_3 \circ \phi_5 \circ \phi_3 = (\phi_5)^2\}.$$

Este grupo es generado por el conjunto $\{\phi_3, \phi_5\}$, $S_3 = \langle \{\phi_3, \phi_5\} \rangle$.

5.5 Hacia un recíproco del Teorema de Lagrange en grupos cíclicos

Para aproximarnos a la construcción de un recíproco del Teorema de Lagrange requerimos completar nuestra discusión acerca del orden de un elemento de un grupo que, recordamos, habíamos definido como $\min\{k \in \mathbb{N} | a^k = e\}$.

Consideremos los subgrupos $H = \{i, h\}$ y $K = \{i, v\}$ y construimos,

$$HK = \{hk | h \in H, k \in K\}$$

$$HK = \{i, h, v, hv\} = \{i, h, v, r_1^2\}.$$

HK es en este caso un subgrupo de S y si construimos, KH podemos notar que en este ejemplo particular $KH = HK$.

Puntos de discusión

1. Investigar qué ocurre si se consideran los subgrupos, $H_1 = \{i, r_1, r_1^2 h, h\}$ y $K_1 = \{i, h\}$ y se construyen H_1K_1 y K_1H_1. ¿Es H_1K_1 un subgrupo de S?

2. Identificar todos los subgrupos de S, analizar productos de estos subgrupos, y determinar el número de elementos del conjunto producto en cada caso. ¿Coincide éste con el producto de los órdenes de los subgrupos?.

3. Sean A y B subconjuntos de un grupo G, cada uno con m elementos. Demostrar que AB contiene k elementos, donde $m \leq k \leq m^2$. Estudiar condiciones bajo las cuales $k = m$ y bajo las cuales $k = m^2$.

De las observaciones anteriores se percibe que para el objetivo planteado para esta sección de *contar* subgrupos de un determinado orden, es importante el análisis del producto de subgrupos, requerimos entonces caracterizarlo.

Teorema 5.33 *Dados H, K subgrupos de un grupo $(G, \cdot)$. HK es subgrupo de G si y sólo si $HK = KH$.*

Demostración. Supongamos que HK es subgrupo de G y demostremos que $HK = KH$.

Sea $x = hk \in HK$, $h \in H$ y $k \in K$. Como HK es subgrupo de G, $x^{-1} \in HK$, esto es $x^{-1} = h_1 \cdot k_1$, $h_1 \in H$, $k_1 \in K$, pero entonces $x = k_1^{-1} \cdot h_1^{-1}$, $k_1^{-1} \in K$ y $(h_1)^{-1} \in H$, de donde $x \in KH$. De manera similar si $y \in KH$, concluimos que $y \in HK$.

Si asumimos ahora que $HK = KH$, nuestro objetivo es demostrar que HK es subgrupo de G. Tomemos $x, y \in HK$, $x = h_1 \cdot k_1$ y $y = h_2 \cdot k_2$. Entonces

$$x \cdot y^{-1} = h_1 \cdot k_1 \cdot (k_2)^{-1} (h_2)^{-1} = h_1 \cdot k_3 \cdot (h_2)^{-1} = h_1 \cdot (h_3 \cdot k_4) = (h_1 \cdot h_3) \cdot k_4 = h_5 \cdot k_4,$$

donde por ser $HK = KH$, hemos escrito $k_3 \cdot h_2^{-1} = h_3 \cdot k_4$. Concluimos entonces que $x \cdot y^{-1} \in HK$, de donde HK es un subgrupo de G.

Corolario 5.8 *Si $(G, \cdot)$ es un grupo abeliano y H, K son subgrupos de G, entonces HK es subgrupo de G.*

En el grupo óctico usted pudo observar que para los subgrupos H y K seleccionados, el producto HK tiene 4 elementos, H_1K_1 tiene también 4 elementos en este ejemplo, y usted habrá realizado un análisis mas general con los otros subgrupos y al trabajar el Punto de discusión 2. Es importante ahora encontrar una manera de contar los elementos de productos de subgrupos finitos de un grupo cualquiera.

Teorema 5.34 *Si H y K son subgrupos finitos de un grupo G, entonces*

$$\#(HK) = \frac{o(H)o(K)}{o(H \cap K)}.$$

Demostración. La observación de los ejemplos nos permite intuir que

$$\#(HK) < \#(H \times K) = \#(H)\#(K),$$

ya que pueden existir $h_1, h_2 \in H$, $h_1 \neq h_2$ y $k_1, k_2 \in K$, $k_1 \neq k_2$, tales que $h_1 \cdot k_1 = h_2 \cdot k_2$; esto es, pueden repetirse algunos productos. Para excluir estas repeticiones requerimos definir en $H \times K$ una relación de equivalencia de la siguiente manera

$$(h_1, k_1) \sim (h_2, k_2) \Longleftrightarrow h_1 \cdot k_1 = h_2 \cdot k_2.$$

$\sim$ es una relación de equivalencia. ¡Comprobarlo! Entonces $\#(HK) =$ número de clases diferentes de $H \times K$ según $\sim$.

Debemos ahora determinar el número de elementos de cada clase de equivalencia. La idea es demostrar que cada clase tiene $o(H \cap K)$ elementos.

Sea $(a, b) \in H \times K$, $a \in H$, $b \in K$ y analicemos $[(a, b)]$. Si consideramos $x \in H \cap K$ y observamos $a \cdot x \cdot x^{-1} b = a \cdot b$, podemos afirmar que $a \cdot x \in H$, $x^{-1} \cdot b \in K$ y $(a \cdot x, x^{-1} \cdot b) \sim (a, b)$, es decir $(a \cdot x, x^{-1} \cdot b) \in [(a, b)]$. Entonces

$$\{(a \cdot x, x^{-1} \cdot b) | x \in H \cap K\} \subseteq [(a, b)].$$

Si ahora tomamos $(h, k) \in [(a, b)]$, $(h, k) \sim (a, b)$, $h \cdot k = a \cdot b$, $a^{-1} \cdot h = b \cdot k^{-1}$, pero $a^{-1} \cdot h \in H$ y $b \cdot k^{-1} \in K$, es decir $a^{-1} \cdot h = b \cdot k^{-1} \in H \cap K$.

Sea $x = a^{-1} \cdot h$, $x^{-1} = h^{-1} \cdot a$, nótese que

$$a \cdot (a^{-1} \cdot h)(h^{-1} \cdot a \cdot b) = a \cdot b \Longrightarrow (a \cdot x, x^{-1} \cdot b) \in [(a, b)].$$

De la forma en que tomamos x, se sigue que $h = a \cdot x$ y $b \cdot k^{-1} = x$, de donde, $k^{-1} = b^{-1} \cdot x$. Entonces $k = x^{-1} \cdot b$, esto es, los elementos de $[(a, b)]$ son de la forma $(a \cdot x, x^{-1} \cdot b)$ con $x \in H \cap K$, es decir, $(h, k) \in \{(a \cdot x, x^{-1} \cdot b) | x \in H \cap K\}$. Concluimos que

$$[(a, b)] = \{(a \cdot x, x^{-1} \cdot b) | x \in H \cap K\}.$$

Basta definir ahora una aplicación

$$f : H \cap K \longrightarrow [(a, b)], x \longrightarrow (a \cdot x, x^{-1} \cdot b).$$

Dicha aplicación resulta ser biyectiva. ¡Demostrarlo! Podemos afirmar entonces que

$$o(H \cap K) = \#[(a, b)].$$

De otra parte, como $\#(H \times K) = o(H)o(K)$ y cada clase de equivalencia según $\sim$ tiene $o(H \cap K)$ elementos, se sigue que

$$\#(HK) = N^0 \text{de clases diferentes} = \frac{\#(H \times K)}{o(H \cap K)} = \frac{o(H)o(K)}{o(H \cap K}.$$

Puntos de discusión

Sea $(G,\cdot)$ un grupo, $a,b \in G$, se tiene

i. $o(a) = o(a^{-1})$.

ii. $o(xax^{-1}) = o(a)$, $\forall x \in G$.

iii. $o(ab) = o(ba)$.

1. Explorar las propiedades anteriores en los grupos, $(S_3, \circ)$ y en $(M^*_{2\times 2}, \cdot)$, el grupo de las matrices no singulares de 2×2, con el producto.

2. Demostrar (i), (ii) y (iii).

Una aplicación importante del Lema 5.2 es la posibilidad de caraterizar los generadores de un grupo cíclico y a partir de allí abordar el reciproco del Teorema de Lagrange en dichos grupos, que desde luego, como usted lo habrá explorado en los ejemplos, no es válido en general.

Teorema 5.31 *Sean $(G,\cdot)$ un grupo cíclico, $o(G) = n$ finito y $G = \langle a\rangle = \{e, a, a^2, a^3, \cdots, a^{n-1}\}$. Sea $1 \leq k \leq n$. Tenemos*

i. Si $(n,k) = 1$, entonces a^k genera a G, es decir, $G = \langle a^k\rangle$.

ii. Recíprocamente, si $b = a^k$ genera a G entonces $mcd(k,n) = 1$.

iii. El número de generadores de G es $\phi(n)$

Demostración.

(i) Sea $1 \leq k \leq n$ tal que $mcd(k,n) = 1$. $o(a^k) = \frac{o(a)}{mcd(k,n)} = n$, luego $\langle a^k\rangle = G = \langle a\rangle$, es decir, a^k genera a G.

(ii) Sea $1 \leq k \leq n$ tal que $\langle a^k\rangle = G$. Como $G = \langle a^k\rangle = \langle a\rangle$ y como $o(a^k) = o(a) = n$, entonces dado que $n = o(a^k) = \frac{o(a)}{mcd(n,k)}$, concluimos que $mcd(n,k) = 1$.

Hemos mostrado entonces que $G = \langle a^k\rangle$ si y sólo si $mcd(k,n) = 1$. De esto se deduce que hay tantos generadores como enteros primos relativos con n, es decir, hay $\phi(n)$ generadores.

Teorema 5.32 *(Recíproco del Teorema de Lagrange). Si G es un grupo cíclico finito de orden n y $k|n$, existe H subgrupo de G tal que $o(H) = k$.*

Demostración. Como $k|n$, existe r entero positivo, tal que $n = kr$, así que el subgrupo

$$H = \{a^r, a^{2r}, a^{3r}, ..., a^{(k-1)r}, e = a^{kr}\}$$

es un subgrupo de G de orden k.

Puntos de discusión

1. Considerar ejemplos de grupos no cíclicos de ordenes 6 y 8 y analizar sus subgrupos. ¿Existen en el primer caso subgrupos de órdenes 2 y 3? ¿Existen en el segundo caso subgrupos de ordenes 2 y 4? ¿Percibe usted alguna regularidad?

2. Si el orden de un grupo es pq donde p y q son primos distintos, ¿Es posible garantizar la existencia de subgrupos de ordenes p y q respectivamente? Es decir, ¿podemos construir otra aproximación al recíproco del Teorema de Lagrange?

En la búsqueda de un recíproco se dan pasos muy fuertes al construir la teoría de Sylow, teoría que no incluimos en este libro; allí se involucran continuamente elementos de la teoría de números (factorización, teorema fundamental de la aritmética, etc.) que permiten estructurar una clasificación de los grupos finitos.

5.6 Subgrupos Normales

En este momento tenemos por el Teorema de Lagrange que si $(G,\cdot)$ es un grupo finito y H es un subgrupo de G, $o(H)|o(G)$; aún mas hemos demostrado un recíproco de este teorema para grupos cíclicos que nos garantiza la existencia de subgrupos de un determinado orden. Nos interesa ahora *contar* cuántos subgrupos de un orden determinado pueden existir en un grupo finito. Para ello, requerimos definir una relación de equivalencia que nos permita clasificar y así facilite tal conteo. Comenzaremos por analizar algunas ideas preliminares.

Multiplicación de Subgrupos. Si $A=\{a_1,a_2,\cdots,a_n\}$ y $B=\{b_1,b_2,\cdots,b_m\}$ son dos subconjuntos de un grupo G, definimos

$$AB=\{a_i\cdot b_j | a_i\in A, b_j\in B\}.$$

El número de tales productos es claramente mn, pero algunos de éstos pueden ser repetidos. Nótese que en general $AB\neq BA$, a no ser que el grupo sea abeliano.

Si consideramos ahora H un subgrupo, $H=\{e,h_1,h_2,\cdots,h_{r-1}\}$, $HH=H$, $h_i.h_j\in H$, por ser H subgrupo, y en este caso tan solo hay r productos distintos.

Consideremos el grupo óctico (grupo de simetrías del cuadrado).

constituido por cuatro rotaciones de 0°, 90°, 180° y 270°, en sentido contrario de las manecillas que corresponden a las permutaciones

$$\begin{aligned} i &= \begin{pmatrix}1&2&3&4\\1&2&3&4\end{pmatrix}, r_1=\begin{pmatrix}1&2&3&4\\2&3&4&1\end{pmatrix},\\ r_2 &= r_1^2=\begin{pmatrix}1&2&3&4\\3&4&1&2\end{pmatrix}, r_3=r_1^3=\begin{pmatrix}1&2&3&4\\4&1&2&3\end{pmatrix}, \end{aligned}$$

y por cuatro reflexiones con respecto a las rectas de simetría h, v, d_1 (que pasa por los vértices 1,3) y d_2 (que pasa por los vértices 2,4) y que corresponden a las permutaciones

$$h=\begin{pmatrix}1&2&3&4\\4&3&2&1\end{pmatrix}, v=\begin{pmatrix}1&2&3&4\\2&1&4&3\end{pmatrix}, d_1=\begin{pmatrix}1&2&3&4\\1&4&3&2\end{pmatrix}, d_2=\begin{pmatrix}1&2&3\\3&2&1\end{pmatrix}$$

Lema 5.1 *Sean $(G,\cdot)$ un grupo, $a \in G$ tal que $o(a) = n$. Entonces $a^k = e$ si y sólo si $k \equiv 0(mod n)$.*

Demostración. Supongamos que $a^k = e$ y apliquemos el algoritmo de la división a k y a n. Existen entonces enteros q y r tales que $k = nq + r$, con $0 \leq r < n$. Ahora,

$$e = a^k = a^{nq+r} = (a^n)^q \cdot a^r = e \cdot a^r = a^r.$$

Tendríamos un entero positivo r estrictamente menor que n tal que $a^r = e$. Por ser n el mínimo, $r = 0$. De donde $k = nq$, esto es, $n|k$ o lo que es lo mismo $k \equiv 0(\mathrm{mod} n)$. El análisis de la otra afirmación es similar.

Lema 5.2 *$a^k = e$ es equivalente a que $k \equiv 0(mod n)$ si y sólo si $n = o(a)$.*

Demostración. Supongamos que $a^k = e$ es equivalente a $k \equiv 0(\mathrm{mod} n)$. Como $n \equiv 0(\mathrm{mod} n)$ concluimos que $a^n = e$.

Además si $k \in \mathbb{Z}$, es tal que $0 < k < n$ entonces $k \not\equiv 0(\mathrm{mod} n)$, por hipótesis $a^k \neq e$. Concluimos que $o(a) = n$.

Recíprocamente si suponemos que $o(a) = n$, $a^k = e$ si y sólo sí $k \equiv 0(\mathrm{mod} n)$.

La siguiente pregunta surge de manera natural. Si $o(a) = n$, ¿cuál es el orden de a^i para i entero positivo? Analicemos esta pregunta en tres de los ejemplos que discutimos anteriormente.

En el grupo de los cuaterniones $o(i) = o(j) = o(k) = 4$. ¿Cuál es el orden de i^2? Como $i^2 = -1$, $i^{2^2} = 1$, entonces $o(i^2) = 2$. ¿Y el de i^3? $i^3 = -i$, $i^{3^2} = -1$, $i^{3^3} = i$ y $i^{3^4} = 1$, esto es, $o(i^3) = 4$. Similarmente para las potencias de j y de k.

En el grupo de las raíces p-ésimas de la unidad ¿Cuál es el orden de w^2 si $w = e^{\frac{2\pi i}{p}}$? Afirmamos que $o(w^2) = p$, y aún mas que $o(w^k) = p$ para cualquier $1 < k < p$. ¡Explique usted porqué!

Pero, si analizamos $U = \{z \in \mathbb{C} | z^6 = 1\}$, raíces sextas de la unidad y tomamos $w = e^{\frac{\pi i}{3}}$ en U, $o(w^2) = 3$, $o(w^3) = 2$, $o(w^4) = 3$ y $o(w^5) = 6$. La observación de estos ejemplos seguramente nos sugiere ya alguna regularidad que se expresa en el teorema que se enuncia a continuación.

Teorema 5.28 *Dado $a \in G$ tal que $o(a) = n$, entonces $o(a^i) = \frac{n}{d}$, donde $d = mcd(n, i)$.*

Demostración. Si $k \in \mathbb{Z}$ es tal que $(a^i)^k = e$, esto es $a^{ik} = e$, entonces como $o(a) = n$, concluimos que $n|ik$, es decir $ik \equiv 0(\mathrm{mod} n)$, entonces $k \equiv 0(\mathrm{mod} \frac{n}{d})$, donde $d = mcd(n, i)$. Es decir $(a^i)^k = e$ si y sólo si $k \equiv 0(\mathrm{mod} \frac{n}{d})$, y por el Lema 5.2 $o(a^i) = \frac{n}{d}$.

Nos interesa explorar ahora bajo qué condiciones es posible garantizar la existencia de elementos de un orden determinado en un grupo, pues hemos observado ya en los ejemplos que en general no es posible dar tal garantía.

Teorema 5.29 *Si $o(a) = nm$ y $mcd(n, m) = 1$, entonces existen b, c, únicos, en G tales que $o(b) = n$, $o(c) = m$ y $a = b \cdot c = c \cdot b$.*

Demostración. Como $mcd(n, m) = 1$, 1 puede ser expresado como combinación lineal de n y m, es decir existen $r, s \in \mathbb{Z}$ tales que

$$rn + sm = 1.$$

Pero entonces, $a = a^1 = a^{rn+sm} = a^{rn} \cdot a^{sm} = a^{sm} \cdot a^{rn}$.

Tomando $c = a^{rn}$ y $b = a^{sm}$, tenemos que $o(c) = o(a^{rn}) = \frac{nm}{mcd(rn,nm)} = m$, pues $mcd(rn, nm) = n$. (¿Por qué?) Análogamente, $o(b) = o(a^{sm}) = \frac{nm}{mcd(sm,nm)} = n$.

La demostración de la unicidad de b y c recurre de nuevo a la representación del máximo común divisor, tantas veces citada. Si suponemos que existen $x, y \in G$ tales que $o(x) = n$, $o(y) = m$ y $a = x \cdot y = y \cdot x$, veamos que $y = a^{rn}$ y $x = a^{sm}$.

Como $rn + sm = 1$, tenemos que $a^{rn} = (x \cdot y)^{rn} = x^{rn} \cdot y^{rn} = y^{rn}$, porque $x \cdot y = y \cdot x$ y $o(x) = n$. De allí se sigue que $a^{rn} = y^{rn} = y^{1-sm} = y \cdot y^{-sm} = y \cdot (y^m)^{-s} = y$, pues el orden de y es m. En conclusión, $a^{rn} = y$ y esto muestra la unicidad de y.

De otra parte, como $a = x \cdot y = x \cdot a^{rn}$, $a = x \cdot a^{rn}$, entonces $x = a^{1-rn} = a^{sm}$ y esto muestra la unicidad de x.

Corolario 5.7 *Dados un grupo $(G, \cdot)$ y $a \in G$ tal que $o(a) = p_1^{r_1} p_2^{r_2} \cdots p_n^{r_n}$ donde los p_i son primos distintos y los r_i son enteros positivos, existen $b_1, b_2, \ldots, b_n$ únicos en G tales que $a = b_1 b_2 \cdots b_n$, $b_i \cdot b_j = b_j \cdot b_i$, i, j en $1, 2, \cdots n$ y además $o(b_i) = p_i^{r_i}$.*

Puntos de discusión

1. Construir un argumento inductivo para demostrar el Corolario 5.7.
2. En $(\mathbb{Z}_7^*, \cdot)$, $o(5) = 6$, $5 = 2 \cdot 6 = 6 \cdot 2$ y $o(2) = 3, o(6) = 2$. Nótese que también $5 = 4 \cdot 3 = 3 \cdot 4$, pero allí no es válida la afirmación acerca de los órdenes. ¿Hay alguna contradicción en este hecho?

Si en el grupo abeliano $(\mathbb{Z}^*, \cdot)$ consideramos los subgrupos $K = \langle 4 \rangle$ y $H = \langle 6 \rangle$, podemos observar que la intersección de estos, se reduce al neutro, $K \cap H = \{1\}$ y que el orden de sus respectivos generadores es, $o(4) = 3$ y $o(6) = 2$ y además $o(4 \cdot 6) = o(3) = 6 = o(4) \cdot o(6)$. Una pregunta que se podría hacer es la siguiente. Si en general como en este caso, el orden de $a \cdot b$ se puede expresar en términos del orden de a y del orden de b, ya que desde el punto de vista operativo en la determinación de los órdenes de los elementos de un grupo finito resultaría de gran ayuda. Es desde luego natural que existan condiciones para ello y se expresan en el siguiente teorema.

Teorema 5.30 *Sean $(G, \cdot)$ un grupo, $a, b \in G$ tales que $\langle a \rangle \cap \langle b \rangle = \{e\}$, $ab = ba$ y $o(a) = m$, $o(b) = n$, finitos. Entonces $o(ab) = mcm(m, n)$, el mínimo común múltiplo entre m y n.*

Demostración. Dado que $ab = ba$, si $(ab)^k = e$ entonces $a^k \cdot b^k = e$, de donde, $a^k = e$ y $b^k = e$. ¿Por qué razón? Concluimos entonces que $k \equiv 0 (\mathrm{mod} m)$ y $k \equiv 0 (\mathrm{mod} n)$, y de aquí, aplicando propiedades de las congruencias, $k \equiv 0 (\mathrm{mod}\ mcm(m, n))$. Luego el orden de ab es $mcm(m, n)$.

Nótese que si n y m son primos relativos, el orden de ab es precisamente mn, pues del hecho de que los órdenes sean primos relativos podemos deducir fácilmente que la intersección de los subgrupos $\langle a \rangle$ y $\langle b \rangle$ se reduce al neutro. Este es el caso en nuestro ejemplo de $\mathbb{Z}_7^*$.

Corolario 5.9 *Dados $(G,\cdot)$ finito y H y K subgrupos de G tales que $o(H) > \sqrt{o(G)}$ y $o(K) > \sqrt{o(G)}$, entonces $H \cap K \neq \{e\}$.*

Demostración. Basta analizar igualdad anterior

$$\#(HK) = \frac{o(H)o(K)}{o(H \cap K)} > \frac{o(G)}{o(H \cap K)}.$$

Como $o(G) \leq \#(HK)$, si $o(H \cap K) = 1$, tendríamos que $o(G) > o(G)$, que es una contradicción, luego $H \cap K \neq \{e\}$.

Corolario 5.10 *Sean $(G,\cdot)$ un grupo, $o(G) = pq$, con p, q primos, $p > q$. Entonces G posee a lo mas un subgrupo de orden p.*

Demostración. Supongamos que existen H, K subgrupos de $(G,\cdot)$ tales que $o(H) = o(K) = p$. Nuestro objetivo es demostrar que $H = K$.

Como $p > \sqrt{pq}$ nos encontramos en la hipótesis del Corolario 5.9, así que $H \cap K \neq \{e\}$. Y como $H \cap K \subseteq H$ y $H \cap K \subseteq K$ y H y K son subgrupos de orden p, (p primo), $H \cap K = H = K$.

Corolario 5.11 *Sean H y K subgrupos de un grupo finito G entonces*

$$i_{\langle H \cup K \rangle}(H) \geq i_K(H \cap K).$$

Demostración. Como $HK \subset \langle H \cup K \rangle$, entonces $\#(HK) \geq o(\langle H \cup K \rangle)$. Pero por teorema anterior

$$\frac{o(H)o(K)}{o(H \cap K)} = \#(HK) \leq o(\langle H \cup K \rangle),$$

y por consiguiente

$$\frac{o(K)}{o(H \cap K)} \leq \frac{o(\langle H \cup K \rangle)}{o(H)},$$

o lo que es lo mismo,

$$i_K(H \cap K) \leq i_{\langle H \cup K \rangle}(H).$$

5.6.1 La transformación de un subgrupo dado: subgrupos conjugados

Dado N un subgrupo dado de un grupo G, podemos considerar la transformación de N por un elemento a fijo de G (que no esté en N). Notaremos esta transformación como aNa^{-1} y la definiremos como el conjunto $\{aea^{-1}, an_1a^{-1}, an_2a^{-1}, \cdots\}$, donde $e, n_1, n_2, \cdots$ son los elementos de N. Demostraremos que este nuevo conjunto es también un subgrupo de G. Para ello, en el caso de grupos finitos, tan solo es necesario establecer la clausurativa, a saber, si an_ra^{-1} y an_sa^{-1} son dos elementos de aNa^{-1}, entonces

$$(an_ra^{-1})(an_sa^{-1}) = an_r(a^{-1}a)(n_sa^{-1}) = an_rn_sa^{-1}.$$

Pero n_r, n_s están en el subgrupo N, de donde, $n_rn_s = n_t \in N$ y se tiene que el producto an_ta^{-1} es un elemento de aNa^{-1}. Concluimos entonces

que aNa^{-1} es un subgrupo de G. Este subgrupo tiene el mismo orden de N, pues no hay elementos repetidos; en efecto, si suponemos que $an_ra^{-1} = an_sa^{-1}$, aplicando cancelativas concluimos que $n_r = n_s$. En conclusión, los subgrupos N y aNa^{-1} tienen el mismo orden. El subgrupo aNa^{-1} es llamado el subgrupo conjugado de N por el elemento a.

Surge naturalmente la pregunta: ¿son ellos el mismo subgrupo de G? La respuesta a esta pregunta es en general ¡*no*! Para sustentar esta respuesta, analicemos el grupo $(A_4, \circ)$, el subgrupo de permutaciones pares del grupo de permutaciones de cuatro elementos $(S_4, \circ)$. Sus elementos son

$$e = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}, a = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}, b = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix},$$

$$c = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}, p = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 2 & 3 \end{pmatrix}, p^2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \end{pmatrix},$$

$$q = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix}, q^2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix}, r = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix},$$

$$r^2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}, s = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}, s^2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}.$$

Si consideramos aquí el subgrupo $N = \{e, p, p^2\}$ y el elemento $a \in A_4$, tenemos que $aNa^{-1} = \{1, q, q^2\}$, que no es un subgrupo de A_4. Por otra parte, $pNp^{-1} = p^2N(p^2)^{-1} = N$.

Tomemos ahora el subgrupo de S_4, $H = \{i, a, b, c\}$. Consideremos

$$pHp^{-1} = pHp^2 = \{i, pap^2, pbp^2, pcp^2\} = \{i, c, a, b\} = H.$$

De la misma forma,

$$r^2H(r^2)^{-1} = r^2Hr = \{i, r^2ar, r^2br, r^2cr\} = \{i, b, c, a\} = H.$$

Obsérvese además que si $H = \{i, a, b, c\}$, $pH = \{p, r, s, q\}$ y $Hp = \{p, s, q, r\}$.

Puntos de discusión

1. Seleccionar otros subgrupos de A_4, construir los subgrupos conjugados y determinar si coinciden o no con el subgrupo inicial.
2. En el grupo óctico considerar $H = \{i, h\}$ y $K = \{i, v\}$ y construir subgrupos conjugados. ¿Qué concluye?
3. ¿Tiene el grupo óctico un subgrupo de orden 4? En caso afirmativo, construir subgrupos conjugados para este subgrupo.

La discusión anterior motiva una nueva definición.

Definición 5.22 *Si N es un subgrupo de un grupo G, se dice que N es un subgrupo normal de G si para todo $a \in G$, $aNa^{-1} \subseteq N$. Se nota $N \triangleleft G$.*

Lema 5.3 *Si $N \triangleleft G$, entonces para todo $a \in G$, $aNa^{-1} = N$.*

Demostración: Si $aNa^{-1} = N$, para todo $a \in G$, N es desde luego normal en G.

Supongamos que $N \triangleleft G$. Por definición para todo $a \in G$, $aNa^{-1} \subset N$. Luego $a^{-1}Na \subset N$, para todo $a \in G$, y de aquí $N \subset aNa^{-1}$ para todo $a \in G$. Se sigue que $aNa^{-1} = N$.

Nota. Nótese que el lema no afirma que dado $n \in N$ y $a \in G$, $ana^{-1} = n$ sino que dado $a \in G$ existen n_1 y n_2 en N tales que $an_1a^{-1} = n_2$.

Para dar una presentación alterna de los subgrupos normales, que seguramente se ha vislumbrado ya en los ejemplos y en las notas anteriores, requerimos de nuevo definir una relación de equivalencia.

Definición 5.23 *Dados $a, b \in G$, y H un subgrupo de G, se define la relación R.*

$$a \ R \ b, \text{ si y sólo si } b^{-1}a \in H$$

Puntos de discusión

1. Demostrar que R es una relación de equivalencia.
2. Dado $(G, \cdot)$ un grupo y $a \in G$, notamos $C(a)$ la clase de equivalencia de a según la relación R y $aH = \{ah | h \in H\}$ la clase a izquierda de a según H (como lo anotamos en una sección anterior, simétricamente la clase a derecha). Demostrar que $C(a) = aH$.

Teorema 5.35 *Sean $(G, \cdot)$ un grupo, N subgrupo de G. N es un subgrupo normal de G si y sólo si toda clase a izquierda es clase a derecha según N.*

Demostración. Si $N \triangleleft G$ y $a \in G$ por Lema 5.6.1, $aNa^{-1} = N$, es decir, $aN = Na$ para todo $a \in G$ que nos dice que toda clase lateral izquierda es clase lateral derecha.

Si ahora suponemos que dado $a \in G$ existe $b \in G$, tal que $Na = bN$, se sigue que $a \in bN$ y de allí que la clase izquierda de a coincidirá con la clase izquierda de b, esto es $aN = bN$, de donde, $Na = aN$ para todo $a \in G$. Por lo tanto, $aNa^{-1} = N$ para todo $a \in G$ y N es subgrupo normal de G.

Teorema 5.36 *Sean $(G, \cdot)$ un grupo y N un subgrupo de G, N es subgrupo normal de G si y sólo sí el producto de clases a derecha según N es clase a derecha según N.*

Demostración. N subgrupo normal implica que, para todo $a \in G$, $Na = aN$. Ahora, si consideramos Nb otra clase a derecha, se tiene

$$(Na)(Nb) = N(aN)b = N(Na)b = NNab = Nab,$$

pues por ser N subgrupo de G, $NN = N$. Se concluye entonces que $(Na)(Nb) = Nab$.

Si suponemos ahora que $(Na)(Nb) = Nc$, debemos demostrar que N es subgrupo normal de G. Para ello, veamos primero que $ab = (ea)(eb) \in (Na)(Nb)$, luego $ab \in Nc$, esto es $Nab = Nc$, $(Na)(Nb) = Nab$.

Considerando ahora a un elemento arbitrario de G y $n \in N$, se tiene

$$ana^{-1} = (ea)(na^{-1}) \in (Na)(Na^{-1}) = Naa^{-1} = N.$$

Luego para todo $n \in N$, $ana^{-1} \in N$, es decir $aNa^{-1} \subset N$ para todo $a \in G$ y se sigue que N es subgrupo normal de G.

Nota. Si N es un subgrupo normal de G, y consideramos el conjunto $\{Na | a \in G\}$, de las clases a derecha (según N) de elementos de G, que notaremos $\frac{G}{N}$, el Lema 5.6.2 nos sugiere considerar en este conjunto la siguiente ley de composición interna.

$$\begin{array}{rcl} (\frac{G}{N}) \times (\frac{G}{N}) & \longrightarrow & \frac{G}{N} \\ (Na, Nb) & \longrightarrow & (Na)(Nb) = Nab. \end{array}$$

Esta operación estáa bien definida, puesto que, si $Na = Na'$ y $Nb = Nb'$, entonces $Nab = (Na)(Nb) = (Na')(Nb') = Na'b'$.

Punto de discusión

Verificar que, la anterior operación es asociativa, que N actúa como elemento neutro y que el inverso de Na es Na^{-1}. En resumen, verificar que si N es normal en G, entonces $\frac{G}{N}$ es un grupo con esta operación. ¿Es $\frac{G}{N}$ abeliano?

Definición 5.24 *Si $N \triangleleft G$, $(\frac{G}{N}, \cdot)$ se llama el grupo cociente de G por N.*

Notas.

1. (a) De la definición de $i_G(N)$ resulta que si $(G, \cdot)$ es un grupo finito y $N \triangleleft G$, entonces

$$o\left(\frac{G}{N}\right) = i_G(N) = \frac{o(G)}{o(N)}.$$

2. Si consideramos el conjunto $\{aN | a \in G\}$ de las clases a izquierda de a según N y lo notamos GN, podemos afirmar que existe una correspondencia biunivoca entre $\frac{G}{N}$ y GN. ¡Demostrarlo! Concluimos entonces que

$$i_G(N) = \#\left(\frac{G}{N}\right) = \#(G\mathbb{N}).$$

Lema 5.4 *Si N es un subgrupo de G tal que $i_G(N) = 2$, entonces N es normal en G.*

Demostración. Si $a \in G$ y $a \notin N$, las clases a derecha de N son N y Na, y las clases a izquierda son N y aN. Luego estos dos pares conforman una partición de G, es decir,

$$N \cup aN = G \text{ y } N \cap aN = \emptyset$$

y

$$N \cup Na = G, N \cap Na = \emptyset.$$

Resulta entonces que si $a \notin N$, $Na = aN$, para $a \in N$, se tiene desde luego que $Na = aN = N$.

Concluimos entonces que para todo $a \in G$, $Na = aN$, esto es $N \lhd G$.

Exploremos ahora con un poco mas de detenimiento cómo descubrir si un subgrupo dado es normal o no.

1. Por consideración de las clases de conjugación, pues si n es cualquier elemento de un subgrupo normal N, entonces N debe contener todos los miembros de la clase de n, es decir, todos los elementos de la forma ana^{-1}, para cualquier $a \in G$. Con este argumento usted puede determinar, por ejemplo, que A_4 es un subgrupo normal de S_4.

2. En un grupo finito, cuya tabla ha sido construida, se puede listar clases a derecha y a izquierda según el subgrupo. Si coinciden, el subgrupo es normal. En la sección de grupos, en el aparte titulado Otros ejemplos de grupos, usted completó seguramente la tabla del grupo $(S_3, \circ)$ y posteriormente identificó todos los subgrupos de este grupo. Consideremos, con la notación usada allí,

$$H = \{i, \phi_1\}.$$

Basta considerar en este caso $\phi_3 \in S_3$ y construir $H\phi_3$ y $\phi_3 H$. Se observa que $H\phi_3 = \{\phi_3, \phi_4\}$, mientras que, $\phi_3 H = \{\phi_3, \phi_5\}$, o sea, las clases derecha e izquierda no coinciden indicando que el subgrupo H no es normal en G.

Si en este mismo grupo seleccionamos el subgrupo

$$K = \{i, \phi_4, \phi_5\}.$$

Observamos que $K\phi_1 = \{\phi_1, \phi_2, \phi_3\}$ y $\phi_1 K = \{\phi_1, \phi_3, \phi_2\}$, $K\phi_2 = \{\phi_2, \phi_3, \phi_1\}$ y $\phi_2 K = \{\phi_2, \phi_1, \phi_3\}$, y $K\phi_3 = \{\phi_3, \phi_1, \phi_2\} = \phi_3 K$. Por otra parte, desde luego, las clases derechas o izquierdas de todos los elementos de K coinciden con K. Concluimos entonces que K es un subgrupo normal de S_3, además $\frac{G}{K} = \{K, K\phi_1\}$ es un grupo de orden 2, dicho orden es precisamente el $i_{S_3}(K)$.

3. Podemos usar las relaciones que definen el grupo, esto es, dado un grupo abstracto G, determinar si un subgrupo N es normal o no dependerá de si aNa^{-1} coincide o no con el subgrupo N para todo a en el grupo.

Para ilustrar el procedimiento que tenemos en mente, consideremos el grupo abstracto definido por $p^3 = r^4 = e$; $rpr^{-1} = p^{-1}$. Demostraremos, sin construir la tabla, que $\langle p \rangle$ es normal y el $\langle r \rangle$ no lo es.

Sea $N = \{e, p, p^2\}$. Es claro que $pNp^{-1} = N$ y $p^2 N p^{-2} = N$. Ahora se tiene que

$$rNr^{-1} = \{e, rpr^{-1}, rp^2r^{-1}\} = \{e, p^{-1}, p\},$$

pues $rpr^{-1} = p^{-1}$ y el inverso de rpr^{-1} es rp^2r^{-1}. Concluimos que $rNr^{-1} = N$.

Además,

$$r^2 N r^{-2} = r(rNr^{-1})r^{-1} = rNr^{-1} = N,$$

y similarmente $r^3 N r^{-3} = N$. Como también se tiene $pNp^{-1} = p^2 N p^{-2} = N$, entonces si a es cualquier elemento en G, digamos $p^2r^3pr^2$, entonces

$$aNa^{-1} = p^2r^3pr^2N(p^2r^3pr^2)^{-1}$$

$$
\begin{aligned}
&= p^2r^3pr^2Nr^{-2}p^{-1}r^{-3}p^{-2}\\
&= p^2r^3p(r^2Nr^{-2})p^{-1}r^{-3}p^{-2}\\
&= p^2r^3pNp^{-1}r^{-3}p^{-2}\\
&= p^2r^3(pNp^{-1})r^{-3}p^{-2}\\
&= p^2(r^3Nr^{-3})p^{-2}\\
&= p^2Np^{-2}\\
&= N.
\end{aligned}
$$

Por tanto, N es invariante bajo conjugación por cualquier elemento de G y esto significa que N es un subgrupo normal de G.

Considerando ahora $H = \{e, r, r^2, r^3\}$, $pHp^{-1} = \{e, prp^{-1}, pr^2p^{-1}, ppr^3p^{-1}\}$. Pero por relación inicial, $rpr^{-1} = p^{-1}$ y esto equivale a decir que $r = prp$, de donde, $prp^{-1} = p(prp)p^{-1} = p^2r$. Ahora, éste último no puede ser ninguno de los elementos de H pues el suponerlo nos lleva a contradecir la relación $p^3 = r^4 = e$. (¡Verificar!) Se concluye entonces que pHp^{-1} no coincide con el subgrupo H. H puede ser transformado por p en diferentes subgrupos, y por tanto $H = \langle r \rangle$ no es normal en G.

4. Cuando se tiene una realización de un grupo, la tarea de analizar los subgrupos normales se hace mas fácil, ilustremos esta idea con un ejemplo en el cual el grupo es infinito.

Las matrices de la forma

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

a, b, c, d números reales, $ad - bc = 1$, forman un subgrupo de $(M_{2\times 2}, \cdot)$. Investigaremos si este subgrupo es normal o no.

Sea $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ cualquier matriz del subgrupo H, esto es tal que $ad - bc = 1$, y sea

$M = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$ cualquier otra matriz 2×2 no singular (inversible). Consideremos la matriz MAM^{-1}. Por un teorema muy conocido del álgebra lineal, el determinante de esta matriz es el producto de los determinantes de las tres matrices

$$\det(MAM^{-1}) = \det(M) \cdot \det(A) \cdot \det(M^{-1}) = \det(A),$$

pues $\det(M) \cdot \det(M^{-1}) = 1$. Concluimos entonces que $\det(MAM^{-1}) = \det(A) = 1$, y esto significa que MAM^{-1} está en el subgrupo H. Pero A es cualquier elemento de H y M es cualquier elemento del grupo de matrices 2×2 con elementos reales. Por tanto, el subgrupo H es transformado en sí mismo por cualquier elemento del grupo que significa que H es un subgrupo normal.

5. Otro tipo de análisis sería pertinente si estudiamos las transformaciones geométricas en dos dimensiones y nos preguntamos si el grupo de las traslaciones en el plano es un subgrupo normal del grupo de las isometrías directas (rotaciones y traslaciones). Si R es cualquier rotación y T es cualquier traslación, es fácil analizar geométricamente que RTR^{-1} es también una traslación, dado que restaura la orientación de la figura y esto nos permite

concluir que el subgrupo de las translaciones es normal en el grupo de las isometrías directas.

Puntos de discusión

1. Sea $(G, \cdot)$ un grupo. Demostrar que el centro de G, $\{x \in G | gx = xg, \forall g \in G\}$ es normal en G.

2. Sea H subgrupo de G y $N(H) = \{g \in G | gHg^{-1} = H.\}$ Demostrar que H es subgrupo normal en $N(H)$.

3. Se define un grupo G por las relaciones $p^3 = a^2 = (pa)^3 = e$. Demostrar sin construir la tabla que $\{e, a\}$ no es un subgrupo normal, pero que $\{e, a, pap^{-1}, p^{-1}ap\}$ sí es normal en G.

4. Identificar todos los subgrupos normales de S_4.

5. En un grupo de orden 10, ¿es posible garantizar que un subgrupo de orden 5 es normal? ¿Y uno de orden 2? Investigar si es posible afirmar que un subgrupo H de orden m es normal en un grupo G de orden mr, si m no tiene factores primos menores que r.

6. Si a es un elemento de orden 2 en un grupo finito G, demostrar que a está en el centro de G.

7. Determinar todos los subgrupos normales del grupo óctico.

8. Encontrar un contraejemplo para demostrar que, si dos elementos están en la misma clase, entonces no necesariamente uno de ellos está en el centralizador del otro.

9. En un grupo G, elementos de la forma $x^{-1} \cdot y^{-1} \cdot x \cdot y$ son llamados conmutadores. Formar conmutadores con pares de elementos seleccionados en los grupos S_3, S_4, el grupo óctico, el grupo de los cuaterniones, etc. Después de esta exploración demostrar que el conjunto de todos los conmutadores (cuando x, y recorren todos los elementos del grupo) genera un subgrupo normal de G.

10. Es claro que si $(G, \cdot)$ es un grupo abeliano, todo subgrupo es normal, explorar si existe un grupo no abeliano en el que todos los subgrupos no triviales (propios) sean normales.

5.7 Aplicaciones que preservan estructura. Homomorfismos de Grupos

Nuestro objetivo en esta sección es establecer una correspondencia entre el grupo cociente $\frac{G}{H}$ y el grupo G, que conserve la estructura de estos grupos, a pesar de que sus objetos y operaciones sean esencialmente distintos. Exploremos antes de abordar este problema algunas aplicaciones especiales definidas entre grupos que hemos estudiado en secciones anteriores.

1. Sea f una aplicación definida del grupo $(\mathbb{Z}, +)$ en el grupo $(\{\pm 1, \pm i\}, \cdot)$ por $f(n) = i^n$. Entonces todos los enteros divisibles por 4 tienen como imagen a 1; el conjunto $\{\ldots, -3, 1, 5, 9, \ldots\}$ tiene como imagen a i; el conjunto $\{\ldots, -6, -2, 2, 6, \ldots\}$ es enviado en -1 por f y los restantes enteros

$\{\ldots,-5,-1,3,7,\ldots\}$ son enviados en $-i$. Esta aplicación desde luego no es uno a uno, pero

$$f(m+n)=i^{m+n}=i^m\cdot i^n=f(m)\cdot f(n)$$

Obsérvese que si ahora consideramos f del grupo $(\mathbb{Z}_4,+)$ en el grupo $(\{\pm 1,\pm i\},\cdot)$, f es ahora uno a uno y sobre.

Recuerde usted que las clases de residuales módulo 4, son precisamente

$$[0]=\{\cdots,-4,0,4,8,\ldots\}$$
$$[1]=\{\ldots,-7,-3,1,5,9,\ldots\}$$
$$[2]=\{\ldots,-6,-2,2,6,10,\ldots\}$$
$$[3]=\{\ldots,-5,-1,3,7,11,\ldots\}.$$

2. Consideremos el grupo $(M_{2\times 2},\cdot)$ de matrices no singulares con el producto y la aplicación que envía a cada matriz A en su determinante $|A|$. Es bien conocido que si $A,B\in M_{2\times 2}$, $|A\cdot B|=|A||B|$, los productos son preservados por esta aplicación.

3. Consideremos $(S_3,\circ)$ grupo formado por las permutaciones

$$I=\begin{pmatrix}1&2&3\\1&2&2\end{pmatrix},\phi_1=\begin{pmatrix}1&2&3\\1&3&2\end{pmatrix},\phi_2=\begin{pmatrix}1&2&3\\3&2&1\end{pmatrix},$$

$$\phi_3=\begin{pmatrix}1&2&3\\2&1&3\end{pmatrix},\phi_4=\begin{pmatrix}1&2&3\\3&1&2\end{pmatrix},\phi_5=\begin{pmatrix}1&2&3\\2&3&1\end{pmatrix}$$

y el grupo $(G,\cdot)$ definido por la tabla,

$\cdot$	e	a
e	e	a
a	a	e

Y definimos f de la siguiente manera,

$$f(i)=e,f(\phi_4)=e,f(\phi_5)=e$$

y

$$f(\phi_1)=f(\phi_2)=f(\phi_3)=a.$$

Esto significa que no se hace distinción entre las permutaciones (1,2,3), (3,1,2) y (2,3,1), las permutaciones pares; todas son enviadas en el elemento identidad e. Ni tampoco se diferencian entre las permutaciones (1,3,2), (3,2,1) y (2,1,3), las permutaciones impares, que son todas enviadas en a. Nótese que, de nuevo, $f(\phi_i\circ\phi_j)=f(\phi_i)\cdot f(\phi_j)$, o sea, f conserva operaciones y está definida de un grupo de orden 6, S_3, en el grupo de orden 2, G.

Si uno piensa en un triángulo equilátero e interpreta las permutaciones anteriores como permutaciones de sus vértices, las pares corresponden a movimientos del triángulo que "no lo voltean" (rotaciones de 120° en el plano de éste) y las impares a simetrías en las cuales la cara opuesta del triángulo queda expuesta (reflexiones con respecto a las alturas).

4. Observemos ahora las tablas de los grupos $\{0,1,2,3\}$ con la suma módulo 4 y $\{1,2,3,4\}$ con el producto módulo 5.

$+(\text{mod}4)$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot(\text{mod}5)$	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1

Son dos grupos de orden 4, cuyas tablas aparentemente difieren, observando mas detenidamente en la primera hay un elemento de orden 2 (el 2) y dos de orden 4 $(1,3)$; en la segunda hay uno de orden 2 (el 4) y dos de orden 4 $(2,3)$. En escencia, ambos son grupos cíclicos de orden 4, el primero es el $\langle 3\rangle$ y el segundo el $\langle 2\rangle$. Si definimos una aplicación

$$g : (\mathbb{Z}_4, +) \longrightarrow (\mathbb{Z}_5^*, \cdot)$$

por $g(3) = 2$, g es desde luego biyectiva y además para cualesquiera $a, b \in \mathbb{Z}_4$, $g(a+b) = g(a) \cdot g(b)$.

Si usted reordena las tablas de estos dos grupos, podrá observar que los dos pueden ser representados por la tabla.

$\cdot$	e	a	b	c
e	e	a	b	c
a	a	b	c	e
b	b	c	e	a
c	c	e	a	b

Es importante observar aquí que si entre grupos de orden 4 cuyas tablas son respectivamente

$*$	e'	x	y	z
e'	e'	x	y	z
x	x	e'	z	y
y	y	z	e'	x
z	z	y	x	e'

$\cdot$	e	a	b	c
e	e	a	b	c
a	a	b	c	e
b	b	c	e	a
c	c	e	a	b

definimos h tal que

$$h(e') = e, h(x) = a, h(y) = b \text{ y } h(z) = c,$$

h es biyectiva. Pero, por ejemplo, $h(x * y) = h(z) = c$, mientras que $h(x) = a$ y $h(y) = b$ y $h(x) \cdot h(y) = a \cdot b = e$. Es decir, h no conserva las operaciones.

5. Si entre los grupos $(\mathbb{R}^+, \cdot)$ y $(\mathbb{R}, +)$, definimos la aplicación f tal que para todo $x \in \mathbb{R}^+$, $f(x) = \log x$, se tiene que $f(x \cdot y) = \log(x \cdot y) = \log x + \log y = f(x) + f(y)$. De nuevo en este caso f conserva las operaciones.

Como lo destacamos en los ejemplos anteriores, las aplicaciones definidas entre los respectivos grupos (excepto h) conservan las operaciones; una aplicación de este tipo se conoce como un *homomorfismo.*

Definición 5.25 *Dados $(G, \cdot)$ y $(G', *)$ dos grupos. Una aplicación $f : G \to G'$ se dice un homomorfismo si $f(a \cdot b) = f(a) * f(b)$, para todo $a, b \in G$.*

Si además f es biyectiva, f se dice un *isomorfismo*, correspondencia que preserva la estructura, y los grupos G y G' se dicen *isomorfos*. Notaremos $G \approx G'$.

Si $G = G'$ y f es isomorfismo de G en si mismo, f se dice un automorfismo.

Puntos de discusión

1. Demostrar que la relación $\approx$ entre grupos, es un relación de equivalencia.

2. Sea $(G = \{e, a, b, c\}, \cdot)$, es un grupo definido por las relaciones $a^2 = b^2 = c^2 = e$; $a = bc$, $b = ca$ y $c = ab$, esto es G es el cuarto grupo de Klein. Consideramos $f : G \to G$, tal que

   $$f(e) = e, f(a) = e, f(b) = b y f(c) = a.$$

   Demostrar que f es un automorfismo. Construir todos los posibles automorfismos de G en G y demostrar que el conjunto formado por todos estos automorfismos es un grupo bajo la composición usual de funciones.

3. Sea $G = \{z \in \mathbb{C} | z^6 = 1\} = \{1, w, w^2, w^3, w^4, w^5\} = \langle w \rangle$, con el producto usual de complejos, el grupo de las raíces sextas de la unidad, donde $w = \cos\frac{\pi}{3} + i \sin\frac{\pi}{3}$. Se define $h : G \to G$, por $h(w) = w^5$. ¿Es h un automorfismo? ¿Cuál es la imagen por h de w^2?, de w^3?, de w^4?, de w^5? ¿Puede usted definir un automorfismo de G en G distinto de h? (Sugerencia: Tenga en cuenta los órdenes de los elementos de G.)

4. A continuación se encuentra la tabla de un grupo no abeliano de orden 6. Los elementos p y q son de orden 3 y los elementos a, b, y c son de orden 2.

$\cdot$	e	p	q	a	b	c
e	e	p	q	a	b	c
p	p	q	e	b	c	a
q	q	e	p	c	a	b
a	a	c	b	e	q	p
b	b	a	c	p	e	q
c	c	b	a	q	p	e

Consideremos las aplicaciones f y g definidas respectivamente por

$$f : G \to G$$

$$e \to e, p \to q, q \to p, a \to c, b \to b, c \to a$$

y

$$g : G \to G$$

$$e \to e, p \to p, q \to q, a \to b, b \to c, c \to a$$

son automorfismos. Demostrarlo. Construir todos los posibles automorfismos. ¿Cuántos hay? ¿Observa usted alguna relación entre el número de automorfismos y el orden del grupo o de los subgrupos de este?

1. Sean $(G, \cdot)$ un grupo no abeliano, x un elemento fijo de G. Si definimos $h : G \to G$ por $f(y) = xyx^{-1}$, para todo $y \in G$, h es un automorfismo (conocido como automorfismo interior). Demostrar este hecho y explorar todos los automorfismos interiores del grupo óctico.

2. Construir un isomorfismo entre los grupos $(\mathbb{Z}, +)$ y $(2\mathbb{Z}, +)$.

3. ¿Por qué no es posible construir un isomorfismo entre $(\mathbb{Z}, +)$ y $(\mathbb{Z}_5, +)$?

4. ¿Porqué no es posible construir un isomorfismo entre los números complejos con la multiplicación y los pares ordenados de reales (x, y) con la adición?

5. El grupo de simetrías del rectángulo y el grupo de simetrías del rombo son isomorfos. Demostrarlo usando consideraciones geométricas.

6. Considerar el grupo $\{0, 1, 2, 3, 4\}$ con la adición módulo 5. Encontrar un grupo multiplicativo al cual éste sea isomorfo.

7. Sea $(G, \cdot)$ un grupo. Encontrar condiciones sobre G para que la aplicación que envía x en x^{-1} sea un automorfismo.

8. Demostrar que $z \to \overline{z}$ es un automorfismo de $(\mathbb{C}, +)$. ¿Es también un automorfismo de $(\mathbb{C}, \cdot)$?

9. Si $(G, \cdot)$ es un grupo y notamos $A(G)$ el conjunto de todos los automorfismos de G, es decir, $A(G) = \{f : G \to G | f \text{ es automorfismo}\}$, entonces $(A(G), \circ)$ es un grupo donde $\circ$ representa composición de funciones. Demostrarlo.

10. Si un grupo es abeliano, ¿es su grupo de automorfismos siempre abeliano? Explorar ejemplos.

11. ¿Puede un grupo de automorfismos de un grupo infinito ser finito?

Teorema 5.37 *Sea $(G, \cdot)$ un grupo, $N \lhd G$ y $\frac{G}{N} = \{Na | a \in G\}$. Entonces $f : G \to \frac{G}{N}$ definida por $x \to f(x) = Nx$, es un homomorfismo sobre $\frac{G}{N}$. f es llamado el homomorfismo canónico o natural.*

Demostración. Veamos primero que f es un homomorfismo. Sean $x, y \in G$, $f(x \cdot y) = N(x \cdot y)$, pero como N es normal en G, $N(x \cdot y) = (Nx) \cdot (Ny)$, entonces,

$$f(x \cdot y) = N(x \cdot y) = (Nx) \cdot (Ny) = f(x)f(y).$$

f es sobre, porque si escogemos un elemento arbitario, $w \in \frac{G}{N}$, existe $x \in G$ tal que $w = Nx$, es decir, $w = f(x)$, para algún $x \in G$.

En el Ejemplo 3, en los inicios de esta sección, definimos un homomorfismo f entre el grupo $(S_3, \circ)$ y el grupo cíclico de orden 2, en el que tres elementos del grupo S_3, a saber i, ϕ_4 y ϕ_5,tenían como imagen a e, elemento identidad del cíclico. Es fácil verificar que el subconjunto $\{i, \phi_4, \phi_5\}$ es un *subgrupo normal* de S_3.

Si en el ejemplo 5 del mismo aparte se quieren identificar los reales positivos cuya imagen es el modulo de la adición en $\mathbb{R}$, los x tales que $f(x) = \log x = 0$, se llegaría a la conclusión de que sólo sirve $x = 1$, el módulo de la multiplicación en $\mathbb{R}^+$. Y de nuevo el conjunto formado por $\{1\}$ es un subgrupo normal de $(\mathbb{R}^+, \cdot)$.

Planteando un estudio similar en cada uno de los ejemplos y en los puntos de discusión, se obtendrían conclusiones análogas. Realmente lo que se está explorando en estos casos, es el llamado nucleo (o kernel) de los homomorfismos, noción que resultará de gran importancia para solucionar el problema de la representación.

Definición 5.26 *Sean $(G,\cdot)$ y $(G',*)$ grupos y $f: G \to G'$ un homomorfismo. Se llama núcleo (Kernel) de f al conjunto* $\ker(f) = \{a \in G | f(a) = e'\}$ *donde e' es el neutro de G', es decir,* $\ker(f) = f^{-1}(e')$.

Teorema 5.38 *Sea $f: (G,\cdot) \to (G',*)$ un homomorfismo, entonces*

i. $f(e) = e'$, donde e es el neutro de G y e' el neutro de G'.
ii. Si $a \in G$, $f(a^{-1}) = (f(a))^{-1}$.
iii. El nucleo de f, $\ker(f)$ es un subgrupo normal de G.

Demostración.

(i) $e = e \cdot e$, por ser e neutro de G. Y por ser f homomorfismo,

$$f(e) = f(e \cdot e) = f(e) * f(e).$$

Pero, por otro lado como e' es neutro de G', $f(e) = f(e) * e'$; concluimos por unicidad del neutro que $f(e) = e'$.

(ii) Si $a \in G$, existe $a^{-1} \in G$ tal que $e = a \cdot a^{-1}$, pero entonces,

$$e' = f(e) = f(a \cdot a^{-1}) = f(a) * f(a^{-1}).$$

De otra parte, $e' = f(a) * (f(a))^{-1}$. Ahora por unicidad del inverso concluimos que $f(a^{-1}) = (f(a))^{-1}$.

(iii) En primer lugar, $\ker(f) \neq \emptyset$, pues por (i) $e \in \ker(f)$. Para ver que $\ker r(f)$ es un subgrupo basta mostrar que si $a, b \in \ker(f)$, entonces $a \cdot b^{-1} \in \ker(f)$.

Como por (ii), $f(b^{-1}) = (f(b))^{-1}$, si $b \in Ker(f)$, $f(b) = e'$ y

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

esto es, $b^{-1} \in \ker(f)$. Ahora, si $a, b \in \ker(f)$, se tiene

$$f(a \cdot b^{-1}) = (f(a)) * (f(b^{-1}) = e' * e' = e',$$

de donde, se concluye que $a \cdot b^{-1} \in \ker(f)$.

Veamos ahora que $\ker(f) \lhd G$. Sean $g \in G$ y $a \in \ker(f)$. Entonces

$$f(g \cdot a \cdot g^{-1}) = f(g) * f(a) * f(g^{-1}) = f(g) * e' * (f(g))^{-1} = f(g) * (f(g))^{-1} = e'.$$

Se tiene entonces que, para cualquier $g \in G$, $a \in \ker(f)$, $g \cdot a \cdot g^{-1} \in \ker(f)$, de donde $\ker(f) \lhd G$.

Punto de discusión

Sea $f: (G,\cdot) \to (G',*)$ un homomorfismo. Consideremos $f(G) = \{f(a) | a \in G\}$, el conjunto imagen. Explorar las características de este conjunto en los homomorfismos trabajados en esta sección y demostrar que $f(G)$ es un subgrupo de G'.

Si usted analiza los homomorfismos estudiados a lo largo de esta sección, y elige entre ellos aquellos que son isomorfismos, observará que en todos ellos el núcleo consta de un único elemento, el neutro del grupo. Pero existen otros homomorfismos, que no son isomorfismos, para los cuales el núcleo es también trivial. La siguiente proposición caracteriza completamente dichos homomorfismos.

Teorema 5.39 *Sea $f : (G, \cdot) \to (G', *)$ un homomorfismo. Entonces f es uno a uno si y sólo si* $\ker(f) = \{e\}$.

Demostración. (i) Sean f uno a uno y $a \in \ker(f)$. Se tiene entonces que $f(a) = e'$ donde e' es el neutro de G'. Pero si e es el neutro de G, por el Teorema 5.3.8, $f(e) = e'$y, ya que por hipótesis f es uno a uno, concluimos que $a = e$, es decir $\ker(f) = \{e\}$.

(ii) Si suponemos ahora que $\ker(f) = \{e\}$, debemos demostrar que f es uno a uno. Sean $a, b \in G$ tales que $f(a) = f(b)$, entonces $(f(a)) * (f(b))^{-1} = e'$. Pero por ser f homomorfismo,

$$e' = (f(a)) * (f(b))^{-1} = f(a \cdot b^{-1}),$$

que implica que $a \cdot b^{-1} \in \ker(f)$. Además, por hipótesis $\ker(f) = \{e\}$, luego, $a \cdot b^{-1} = e$, $a = b$ y se puede concluir que f es uno a uno.

Nuestro objetivo ahora es analizar mas a fondo la idea de que "un homomorfismo conserva la estructura". Queremos ir mas allá de la conservación de operaciones y explorar qué pasa con la imagen directa e inversa de un subgrupo, así como la imagen directa e inversa de subgrupos normales, por un homomorfismo.

Por ejemplo, exploremos el cuarto grupo de Klein al que se hacía referencia en el Punto de discusión 3 de esta sección, y sobre él, el homomorfismo f tal que

$$f(e) = e, f(a) = e, f(b) = a \text{ y } f(c) = b (a^2 = b^2 = c^2 = e \text{ y } c = ab = ba).$$

Si seleccionamos un subgrupo de este grupo, digamos $H = \{e, c\}$, y analizamos

$$f(H) = \{f(x) | x \in H\} = \{e, b\},$$

la imagen resulta ser de nuevo un subgrupo del grupo de Klein. Desde luego como el grupo en este caso es abeliano, todos los subgrupos resultan ser normales.

Consideramos los grupos $(\mathbb{Z}, +)$ y $(\{\pm 1, \pm i\}, \cdot)$, y el homorfismo $g : n \to i^n$, así como el subgrupo por $K = \{1, -1\}$ y analizamos $g^{-1}(K) = \{b \in K | b = g(a)$, para algún $a \in G\}$, observamos que $g^{-1}(K) = 2\mathbb{Z}$, un subgrupo de $(\mathbb{Z}, +)$.

Usted puede explorar diversos subgrupos y homomorfismos para aproximarse aún mas a esta idea, que se expresa completamente en el siguiente teorema.

Teorema 5.40 *Sea $f : (G, \cdot) \to (G', *)$ un homomorfismo sobre. Entonces*

i. Si H' es un subgrupo de G', entonces $f^{-1}(H')$ es un subgrupo de G.

ii. Si $H \triangleleft G$, entonces $f(H) \triangleleft G'$.

iii. Si $H' \triangleleft G'$, entonces $f^{-1}(H') \triangleleft G$.

iv. Si H es un subgrupo de G y $\ker(f) \subset H$, entonces $H = f^{-1}(f(H))$.

v. Sean $\mathcal{H} = \{\mathcal{H}, \mathcal{H} \text{ subgrupo de } G \mid \ker(f) \subset H\}$ y $\mathcal{K} = \{K, K \text{ subgrupo de } G'\}$. La aplicación $\Psi : \mathcal{H} \to \mathcal{H}'$, definida por $\Psi(H) = f(H)$ es una biyección. Además si $H \triangleleft G$, $f(H) \triangleleft G'$.

Demostración.

(i) Sea H' subgrupo de G' veamos que $f^{-1}(H')$ es subgrupo de G.

$f^{-1}(H')$ es no vacío, puesto que por lo menos $e \in f^{-1}(H')$, ya que $e' \in H'$ y $f(e) = e'$. Sean $x, y \in f^{-1}(H')$, entonces $f(x) \in H'$ y $f(y) \in H'$. Como H' es subgrupo,

$$f(x) * (f(y))^{-1} \in H',$$

es decir $f(x \cdot y^{-1}) \in H'$, o equivalentemente, $x \cdot y^{-1} \in f^{-1}(H')$. $f^{-1}(H')$ es entonces un subgrupo de G.

(ii) Sea $H \triangleleft G$. Debemos demostrar que $f(H) \triangleleft G'$. Tomamos para ello $h \in H$, $g' \in G'$; como f es sobre, existe $g \in G$ tal que $f(g) = g'$. Consideremos entonces,

$$g' * f(h) * (g)'^{-1} = f(g) * f(h) * (f(g))^{-1} = f(g \cdot h \cdot g^{-1}).$$

Dado que $H \triangleleft G$, $g \cdot h \cdot g^{-1} \in H$. Entonces $g' * f(h) * (g)'^{-1} = f(g \cdot h \cdot g^{-1}) \in f(H)$, y esto significa que $f(H) \triangleleft G'$.

(iii) Sea $H' \triangleleft G'$ y demostremos que $f^{-1}(H') \triangleleft G$. Para ello, sean $g \in G$, $h \in f^{-1}(H')$. Entonces $f(h) \in H'$. Apliquémosle f a $g \cdot h \cdot h^{-1}$,

$$f(g \cdot h \cdot g^{-1}) = (f(g)) * f(h) * (f(g))^{-1} \in H',$$

por ser H' normal. Entonces $f(g \cdot h \cdot g^{-1}) \in H'$, es decir, $g \cdot h \cdot g^{-1} \in f^{-1}(H')$, lo cual significa que $f^{-1}(H) \triangleleft G$.

(iv) Dado H subgrupo de G, tal que $\ker(f) \subset H$, debemos ver que $H = f^{-1}(f(H))$. Siempre se tiene que $H \subseteq f^{-1}(f(H))$, pero queda por demostrar la otra contenencia.

Sea $x \in f^{-1}(f(H))$. Entonces $f(x) \in f(H)$; existe pues $h \in H$ tal que $f(x) = f(h)$, y por consiguiente, $f(x \cdot h^{-1}) = e'$ lo cual equivale a decir que $x \cdot h^{-1} \in \ker(f)$. Ahora bien, $\ker(f) \subset H$ implica que $x \cdot h^{-1} \in H$. Existe entonces $h_1 \in H$ tal que $x \cdot h^{-1} = h_1$, $x = h_1 \cdot h \in H$ que nos lleva a concluir que $f^{-1}(f(H) \subseteq H$.

(v) Veamos que Ψ es una biyección. Demostraremos que Ψ es uno a uno. ¡Demostrar que Ψ es sobre; usar para ello que el homomorfismo f es sobre!

Supongamos que $\Psi(H) = \Psi(H_1)$, entonces $f(H) = f(H_1)$ y, como H, H_1 son subgrupos de G que contiene al $\ker(f)$, (iv) implica que $H = f^{-1}(f(H)$ y $H_1 = f^{-1}(f(H_1)$. Como $f(H) = f(H_1)$, se concluye que $H = H_1$.

En resumen, para $H \in H$, $H \triangleleft G$ si y sólo si $f(H) \triangleleft G'$ si y sólo si $\Psi(H) \triangleleft G'$.

Si usted analiza de nuevo la tabla del grupo óctico, observará que el subgrupo $H = \{i, r_1, r_2, r_3\}$, constituido por las rotaciones, es un subgrupo normal generado por r_1. El grupo $\frac{G}{H}$ está formado en este caso por dos clases, una de ellas la clase de H. Su tabla es la de un grupo cíclico de orden 2 y el homorfismo canónico de G en $\frac{G}{H}$ tiene como núcleo al subgrupo H.

Si en el grupo $(S_3, \circ)$ considera usted el subgrupo de las permutaciones pares, que es normal, y construye el grupo cociente obtendrá que este grupo resulta isomorfo a un grupo cíclico de orden 2. Exploraremos estas ideas en los Teoremas Fundamentales de Isomorfía que presentamos a continuación.

Teorema 5.41 *Sean $(G, \cdot)$ un grupo, $N \triangleleft G$, y $\frac{G}{N}$ el grupo cociente. Entonces todo subgrupo de $\frac{G}{N}$ es de la forma $\frac{H}{N}$, donde H es subgrupo de G y $N \subset H$. Además $\frac{N}{H} \triangleleft \frac{G}{H}$ si y sólo si $H \triangleleft G$.*

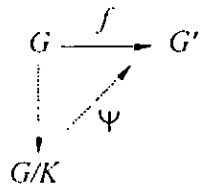
Demostración. El homomorfismo canónico $f : G \to \frac{G}{N}$, tal que $f(g) = Ng$ es sobre. Entonces por el Teorema 5.7.4, parte (v), dado un subgrupo H' de $\frac{G}{N}$, existe un único subgrupo H de G, $H \subset \ker(f) = N$, tal que $f(H) = H'$. Pero por definición

$$f(H) = \{f(x) | x \in H\} = \{Nx | x \in H\} = \frac{H}{N}.$$

Es también claro por el mismo teorema que $H \triangleleft G$ si y sólo si $f(H) \triangleleft \frac{G}{N}$, esto es, si y sólo si $\frac{H}{N} \triangleleft \frac{G}{N}$.

Teorema 5.42 *Sean $(G, \cdot)$ y $(G', *)$ dos grupos, $f : G \to G'$ un homomorfismo sobre, y $K = \ker(f)$. Entonces $\frac{G}{K} \approx G'$.*

Demostración. Para demostrar que los grupos $\frac{G}{K}$ y G' son isomorfos, es preciso definir un isomorfismo. La observación del diagrama



nos induce a definir $\Psi : \frac{G}{K} \to G'$, por $\Psi(Kg) = f(g)$. Veamos que Ψ satisface las condiciones requeridas.

En primer lugar, Ψ está bien definida, pues si $Kg = Kh$, entonces $Kgh^{-1} = K$, y esto significa que $g \cdot h^{-1} \in K = \ker(f)$, de donde, $f(g \cdot h^{-1}) = e'$, $f(g) = f(h)$ y esto implica que $\Psi(Kg) = \Psi(Kh)$.

Si suponemos ahora que $\Psi(Kg) = \Psi(Kh)$, concluiremos similarmente que $Kg = Kh$, o sea que Ψ es uno a uno.

Además, Ψ es sobre. Pues dado $g' \in G'$, por ser f sobre, existe $g \in G$, tal que $f(g) = g'$. Pero entonces existe $Kg \in \frac{G}{K}$ tal que $\Psi(Kg) = f(g) = g'$.

Ψ es además un homomorfismo, pues si $Kg, Kh \in \frac{G}{K}$,

$$\Psi((Kg)(Kh)) = \Psi(Kgh) = f(gh) = f(g) * f(h) = \Psi(Kg) * \Psi(Kh).$$

Puntos de discusión

1. Sean $(G, \cdot)$ y $(G', *)$ dos grupos y $f : G \to G'$ un homomorfismo sobre. Si $N' \triangleleft G'$ y $N = f^{-1}(N')$, demostrar que $\frac{G}{N} \approx \frac{G'}{N'}$.
2. Construir el grupo cociente de $(\mathbb{Z}, +)$ por el subgrupo $(5\mathbb{Z}, +)$. ¿Es este grupo isomorfo a $(\mathbb{Z}_5, +)$?

3. Considerar el grupo cociente de $(\mathbb{C}, +)$ por el subgrupo de los enteros gaussianos, los complejos de la forma $a+bi$ donde $a, b \in \mathbb{Z}$. Identificar un grupo isomorfo a este cociente.

4. Explorar la siguiente afirmación. Si H es un subgrupo normal de índice 3 en un grupo G, el grupo cociente es isomorfo a C_3 (grupo cíclico de orden 3).

5. Demostrar que si $m, n \in \mathbb{Z}$, $C_m \triangleleft C_{mn}$ y $\frac{C_{mn}}{C_m} \approx C_n$.

Teorema 5.43 *Sean $(G, \cdot)$ un grupo, H subgrupo de G, $K \triangleleft G$. Entonces $\frac{H}{H \cap K} \approx \frac{HK}{K}$.*

Demostración. Consideramos $f : H \to \frac{HK}{K}$, definida por $f(h) = hK$. Por ser K normal, $f(h) = hK = Kh$. f es homomorfismo pues, si $h_1, h_2 \in H$,

$$f(h_1 \cdot h_2) = Kh_1h_2 = (Kh_1)(Hh_2) = f(h_1)f(h_2).$$

Ahora si $Kg \in \frac{HK}{K}$, con $g \in HK$, $g = hk$, $h \in H$, $k \in K$. Pero entonces $Kg = gK = (hk)K = hK$, es decir, existe $h \in H$ tal que $Kg = f(h)$.

Si $h \in \ker(f)$, $f(h) = Kh = K$, entonces $h \in K$, de donde ,$h \in H \cap K$. Concluimos que $\ker(f) \subseteq H \cap K$. Además, es claro que si $x \in H \cap K$ entonces $f(x) = K$, esto es, $x \in \ker(f)$. Por consiguiente, $\ker(f) = H \cap K$.

Aplicando ahora teorema anterior, se concluye que $\frac{H}{H \cap K} \approx \frac{HK}{K}$.

Puntos de discusión

1. En un punto de discusión anterior definimos los automorfismos interiores de un grupo $(G, \cdot)$ de la siguiente manera. Dado $a \in G$ fijo, $f_a : G \to G$ se define por $f_a(g) = a \cdot g \cdot a^{-1}$, para todo $g \in G$. Si $I(G)$ es el conjunto formado por todos los automorfismos interiores del grupo G, demostrar que es subgrupo normal del grupo de todos los automorfismos de G.

2. Sean $(G, \cdot)$ es un grupo y $Z(G) = \{x \in G | g \cdot x = x \cdot g, \forall g \in G\}$ el centro de G. Demostrar que $Z(G) \triangleleft G$.

Teorema 5.44 *Sea $(G, \cdot)$ un grupo y $Z(G)$ su centro entonces $\frac{G}{Z(G)} \approx I(G)$.*

Demostración. Basta considerar la aplicación $\phi : G \to I$ definida por $\phi(a) = f_a$, donde $f_a(g) = a \cdot g \cdot a^{-1}$, para $a \in G$. ϕ es homomorfismo, pues $\phi(a \cdot b) = f_{a \cdot b}$ y

$$f_{a \cdot b}(g) = (a \cdot b) \cdot g \cdot (a \cdot b)^{-1} = a \cdot (b \cdot g \cdot b^{-1}) \cdot a^{-1} = f_a(b \cdot g \cdot b^{-1}) = f_a(f_b(g) = (f_a \circ f_b)(g).$$

Concluimos entonces que $\phi(a \cdot b) = f_a \circ f_b = \phi(a) \circ \phi(b)$, o sea, ϕ es homomorfismo. ϕ es además sobre (¡demostrar!).

Si consideramos ahora $a \in \ker(\phi)$, $\phi(a) = i_d$, donde i_d denota el automorfismo identidad, tenemos que $f_a(g) = i_d(g) = g$, de donde, para todo $g \in G$, $a \cdot g \cdot a^{-1} = g$, lo cual implica que $a \cdot g = g \cdot a$. Se sigue entonces que $\ker(\phi) \subseteq Z(G)$. De otra parte, es claro que $Z(G) \subseteq \ker(\phi)$ lo que nos permite concluir que $\ker(\phi) = Z(G)$, y por el Teorema 5.4.3 concluimos que $\frac{G}{Z(G)} \approx I(G)$.

Analicemos ahora algunas aplicaciones de la teoría discutida en esta sección.

1. Planteamos en uno de los puntos de discusión que en un grupo $(G, \cdot)$ arbitrario la aplicación $f(a) = a^{-1}$, no es en general un automorfismo; desde luego si G es abeliano, f sí es automorfismo. Es importante resaltar aquí que si por hipótesis la aplicación definida sobre un grupo es un automorfismo, de allí podemos concluir que el grupo es abeliano. La mencionada aplicación nos proporciona pues una manera de caracterizar los grupos abelianos.

2. Sean $(G, \cdot)$ un grupo finito, $o(G) = n$ y $m \in \mathbb{N}$, tal que m, n son primos relativos. Entonces la aplicación $h : G \to G$, definida por $h(a) = a^m$ es una biyección. Además si $(G, \cdot)$ es abeliano, h es automorfismo.

Veamos. Como $(m, n) = 1$, existen $r, s \in \mathbb{Z}$ tales que $rn + sm = 1$. Luego si $g \in G$,

$$g = g^{rn+sm} = g^{rn} \cdot g^{sm} = e \cdot g^{sm} = (g^s)^m.$$

Si hacemos $a = g^s$, tenemos que $g = a^m = h(a)$, es decir, dado $g \in G$, existe $a \in G$ tal que $g = h(a)$. h es , por lo tanto, una aplicación sobre y, como G es finito, h es también uno a uno, de donde, h es una biyección.

Si el grupo es abeliano, h es homomorfismo,

$$h(a \cdot b) \quad = \quad (a \cdot b)^m = (a^m) \cdot (b^m) = h(a) \cdot h(b).$$

Puntos de discusión

1. Construir el grupo $I(G)$ de los automorfismos interiores para el grupo óctico, el grupo de los cuaterniones, el grupo $(S_3, \circ)$ y el grupo $(A_4, \circ)$ de las permutaciones pares de S_4. Identificar en cada caso el centro del grupo y verificar que $\frac{G}{Z(G)} \approx I(G)$.

2. Si C_9 y C_6 son grupos cíclicos de órdenes 9 y 6 respectivamente. Demostrar que $A(C_9) \approx C_6$, donde $A(G)$ es el conjunto de los automorfismos de G.

3. Sea $(G, \cdot)$ un grupo finito y $\gamma \in A(G)$ tal que $\gamma(x) = x$ si y sólo si $x = e$. Entonces dado $a \in G$ existe $x \in G$ tal que $a = x^{-1} \cdot \gamma(x)$.

4. Investigar la siguiente afirmación. Si $(G, \cdot)$ es un grupo finito de orden mayor que 2, existe un automorfismo no trivial sobre G. Explorar casos.

5.8 Permutaciones

En la última sección de este capítulo retomaremos elementos muy importantes que hemos mencionado varias veces a lo largo de el, los relativos a las permutaciones.

Definición 5.27 *Una permutación de un conjunto S es una biyección de S sobre S.*

Nota. Dado S un conjunto, notaremos $B(S)$ el conjunto de todas las biyecciones de S en S. En un punto de discusión anterior usted ya demostró que $(B(S), \circ)$ es un grupo.

Si S posee n elementos, el grupo $B(S)$ se nota S_n y es llamado grupo simétrico; es el grupo de todas las permutaciones de los n elementos y tiene orden $n!$. También en uno de los puntos de discusión usted exploró ya estas ideas. Se han mencionado además algunos subgrupos de los grupos simétricos entre ellos A_4 el subgrupo de permutaciones pares de S_4. Sobre este aspecto tendremos mas que decir mas adelante.

En la sección en que analizamos las tablas de grupos finitos hicimos una aproximación informal al siguiente teorema que nos permite representar los elementos de cualquier grupo por permutaciones, es decir, realizar un grupo G como algo mas concreto.

Teorema 5.45 *Teorema de Representación de Cayley. Todo grupo G es isomorfo a un subgrupo de $(B(S), \circ)$, para algún conjunto S.*

Demostración. Sea $(G, \cdot)$ un grupo. Si consideramos $a \in G$ fijo, la aplicación $\phi_a : G \to G$ definida por $\phi_a(g) = g \cdot a$ es una biyección de G en G. Veamos.

(i) Si $\phi_a(g) = \phi_a(g')$, entonces $g \cdot a = g' \cdot a$, que implica que $g = g'$, es decir ϕ_a es uno a uno.

(ii) Dado $g' \in G$, $g' = (g' \cdot a^{-1}) \cdot a = \phi_a(g' \cdot a^{-1})$, esto es, existe $x = g' \cdot a^{-1} \in G$ tal que $g' = \phi_a(x)$ y esto significa que ϕ_a es sobre.

Si b es otro elemento de G, analicemos como se comporta la compuesta de ϕ_a y ϕ_b.

$$(\phi_b \circ \phi_a)(g) = \phi_b(\phi_a(g)) = \phi_b(g \cdot a) = (g \cdot a) \cdot b = g \cdot (a \cdot b) = \phi_{a \cdot b}(g).$$

De allí se sigue que $\phi_{a \cdot b} = \phi_b \circ \phi_a$. Este comportamiento nos lleva a considerar $\Phi : G \longrightarrow B(G)$ definida por $\Phi(a) = \phi_a$.Veamos que Φ es un homorfismo uno a uno. En primer lugar,

$$\Phi(a \cdot b) = \phi_{a \cdot b} = \phi_b \circ \phi_a = \Phi(b) \circ \Phi(a),$$

de donde Φ es un homomorfismo. Ahora si $\Phi(a) = \Phi(b)$, $\phi_a = \phi_b$ puesto que para todo

$$g \in G, \phi_a(g) = \phi_b(g), g \cdot a = g \cdot b$$

Por cancelativa se concluye que $a = b$, esto es Φ es uno a uno.

Si consideramos $\Phi : G \to \Phi(G)$, Φ es un isomorfismo y como $\Phi(G)$ es un subgrupo de $B(G)$ y G es isomorfo a el, el teorema queda demostrado.

Punto de discusión

¿ Es posible hallar un conjunto S, $S \neq G$, tal que se consiga una representación de G por permutaciones que posea menos elementos que la que acabamos de construir? Explorar de nuevo los ejemplos que analizamos en la sección de tablas de grupos finitos, analizar otros ejemplos e investigar un resultado general al respecto.

En el resto de la sección nos dedicaremos a analizar permutaciones sobre conjuntos finitos.

Consideremos la permutación

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 5 & 6 & 2 & 10 & 1 & 8 & 3 & 7 & 4 & 9 \end{pmatrix}.$$

Contamos cuántos pares están fuera de su orden natural. Por ejemplo, en el orden natural 6 va después de 3, pero en esta permutación el orden de 6 y 3 está cambiado. Se dice que hay una inversión. Deseamos enumerar tales inversiones. El mejor camino es tomar cada dígito y contar a partir de él el número de inversiones, como se indica con este ejemplo.

5 está seguido por 2,1,3 y 4, por tanto contribuye con 4 inversiones.

6 está seguido por 2,1,3,4, contribuye con 4 inversiones.

2 está seguido por 1, contribuye con 1 inversión.

10 está seguido por 1,8,3,7,4,9, por tanto contribuye con 6 inversiones.

1 no contribuye con inversiones.

8 está seguido por 3,4,7, contribuye con 3.

3 no contribuye con inversiones.

7 está seguido por 4, contribuye con una.

9 no contribuye.

El número total de inversiones es en este caso 19, un número impar; se dice que ésta es una permutación impar.

Puntos de discusión

1. Usar el método descrito en el ejemplo anterior para determinar la paridad de las siguientes permutaciones

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 2 & 4 & 1 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 1 & 6 & 5 & 4 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 5 & 6 & 8 & 7 & 1 & 2 & 4 & 3 \end{pmatrix}.$$

1. ¿Cómo se afecta la paridad de la permutación del ejemplo inicial si se intercambian los siguientes pares: $(5\ 9), (21), (68)$?, es decir, si intercambia las imágenes.

Observe que si allí intercambiamos tan sólo los números 2 y 7, aparece una nueva inversión y la permutación resultante es par.

Nuestro objetivo es ahora simplificar la notación de las permutaciones, para ello consideremos un ejemplo

$$P = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 1 & 4 & 5 \end{pmatrix}.$$

En esta permutación se tiene que $P(1) = 2$, escribimos entonces simplemente (12), 1 es enviado en 2. Pero además en este caso $P(2) = 3$, 2 es enviado en 3. Cerramos aquí un primer ciclo (123). Como 4 y 5 son dejados fijos por P, escribimos los llamados 1-ciclos (4) y (5); la permutación P puede ser escrita entonces como producto de estos ciclos

$$P = (123)(4)(5).$$

Respecto al ejemplo inicial

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 5 & 6 & 2 & 10 & 1 & 8 & 3 & 7 & 4 & 9 \end{pmatrix},$$

partiendo de los números en su orden natural, podemos obtener una permutación por sucesivos intercambios de pares de números, o *transposiciones*. Por ejemplo, efectuando la transposición (15) el orden natural se transforma en (5 2 3 4 1 6 7 8 9 10). Requerimos ahora que 6 ocupe la segunda posición, hacemos la transposición (2 6), obteniendo

$$(5\,6\,3\,4\,1\,2\,7\,8\,9\,10).$$

Ahora aplicando (2 3), se obtiene

$$5\quad 6\quad 2\quad 10\quad 1\quad 8\quad 3\quad 7\quad 4\quad 9.$$

En seguida (1 4) nos da

$$5\quad 6\quad 2\quad 10\quad 1\quad 3\quad 7\quad 8\quad 9\quad 4.$$

A continuación aplicamos (3 8) para obtener

$$5\quad 6\quad 2\quad 10\quad 1\quad 8\quad 7\quad 3\quad 9\quad 4.$$

Y ahora (3 7) produce

$$5\quad 6\quad 2\quad 10\quad 1\quad 8\quad 3\quad 7\quad 9\quad 4.$$

Finalmente aplicamos, (4 9)

$$5\quad 6\quad 2\quad 10\quad 1\quad 8\quad 3\quad 7\quad 4\quad 9.$$

Lo anterior nos muestra que la permutación dada puede ser obtenida ("resuelta") por un número finito de transposiciones. Y nos induce a pensar que el método puede ser extendido a una permutación de cualquier número de objetos. Precisemos un poco más las ideas anteriores.

Definición 5.28 *Un m-ciclo* $(s_1, s_2, s_3, \cdots, s_m)$ *es una permutación de* m *elementos distintos de un conjunto* S, *que envía a* s_i *en* s_{i+1}, *para* $i = 1, 2, 3, \cdots, m-1$ *y a* s_m *en* s_1.

Representaremos la composición de dos ciclos escribiéndolos en posiciones adyacentes, con la aplicación de la derecha actuando primero. Un 2-ciclo es llamado una *transposición*.

Teorema 5.46 *Todo m-ciclo,* $m > 1$ *puede ser escrito como un producto de transposiciones. Toda permutación de un número finito* $n > 1$ *de elementos puede ser expresada como producto de transposiciones.*

Demostración. Veamos que el producto de transposiciones

$$(s_1 s_m)(s_1 s_{m-1})(s_1 s_{m-2}) \cdots (s_1 s_3)(s_1 s_2)$$

es igual al ciclo $(s_1 s_2 s_3 \cdots s_{m-1} s_m)$. En el producto la transposición $(s_1 s_2)$ envía a s_1 en s_2. Como s_2 no aparece en las otras transposiciones, éstas lo dejan fijo. Entonces la transposición $(s_1 s_3)$ envía a s_1 en s_3 que es dejado fijo por el resto de las transposiciones. En general si $1 \leq i < m$, entonces s_i aparece primero en la transposición $(s_1 s_i)$, la cual envía a s_i en s_1 y después la siguiente transposición $(s_1 s_{i+1})$ envía a s_1 en s_{i+1}. Las siguientes transposiciones dejan a s_{i+1} sin cambio y por tanto el efecto final es que $s_i \to s_{i+1}$. Ahora, s_m aparece solamente en la última transposición y ésta lo envía en s_1.

Una permutación P de un número finito n elementos puede ser factorizada en n o menos ciclos. Comenzando el primer ciclo con el primer elemento s_1, seguido por su imagen $P(s_1)$. A continuación cada elemento con su imagen bajo P hasta que una de las imágenes sea s_1. La mínima longitud del ciclo es 1, si s_1 es dejado fijo por P y la maxima longitud es n, pues s_1 debe aparecer como imagen de algún elemento. El número de elementos no usados en el primer ciclo es menor que el número original n. Por este camino los n elementos pueden agotarse en n o menos ciclos, y cada ciclo puede ser reemplazado por un producto de transposiciones como se demostró en la parte inicial.

Nótese que cada permutación puede ser escrita de diversas maneras como un producto de transposiciones. Por ejemplo si consideramos la permutación identidad de 5 elementos

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix},$$

ésta puede ser escrita como (12)(12) o (25)(25) o (12)(34)(12)(34) o (12)(23)(23)(12) o (45)(45)(45)(45)(45)(45) etc. Nótese, además, que cada una de estas factorizaciones tiene un número par de transposiciones.

El siguiente teorema establece que la paridad o imparidad del número de transposiciones para las diversas representaciones de cualquier permutación es invariante.

Teorema 5.47 *Si una permutación P puede ser escrita como el producto de t transposiciones, entonces toda factorización de P en transposiciones tiene $t + 2i$ transposiciones donde i es un entero.*

En otras palabras, toda factorización tiene un número par de transposiciones o toda factorización tiene un número impar de transposiciones.

Demostración. Sea $P = p_1p_2p_3\cdots p_t = q_1q_2q_3\cdots q_s$ donde las p_i y las q_j son transposiciones. Consideremos el número

$$\alpha = \prod_{i<j}(j - i),$$

$i, j = 1, 2, 3, \cdots, n$. Por ejemplo, si $n = 3$, $\alpha = (2 - 1)(3 - 1)((3 - 2)$. Sea Q en S_n y aplicamos Q a α

$$Q(\alpha) = \prod_{i<j}(Q(j) - Q(i)).$$

Este producto contiene también las diferencias entre enteros positivos, menores o iguales que n; pero en este caso ya estas diferencias no son todas positivas como en α. Tenemos entonces

Si $Q = (lk)$, Q es una transposición $(l > k)$

$$Q(\alpha) = \prod_{i<j}(Q(j) - Q(i)) = -\alpha.$$

Veamos ahora que un número impar de factores cambia el signo. Suponemos $k < l$

Los factores que no tienen a k o l son dejados fijos por Q. Un factor positivo de la forma $k - t$, con $1 \leq t < k$, es llevado por Q en $l - t$, pues $Q(k) = l$ y $Q(t) = t$, similarmente $l - t$ es llevado en $k - t > 0$.

De manera análoga, un factor positivo de la forma $w - l$, con $l < w \leq n$ es llevado en $w - k > 0$ y recíprocamente $w - k$ es llevado en $w - l$.

Para $k < z < l$, los factores $z - k$ y $l - z$ son llevados por Q en $z - l < 0$ y $k - z < 0$ y recíprocamente. Luego z contribuye con dos cambios de signo a los factores del producto α.

Nos falta por considerar el factor $l - k$, éste es llevado por Q en $k - l < 0$; esto muestra que el número de cambios de signo es impar y por lo tanto el producto α cambia de signo al actuar Q sobre α.

La acción de la transposición $Q = (lk)$ resulta ser cambiar el signo a α, $Q(\alpha) = -\alpha$.

(ii) Sea ahora $P \in S_n$, $P = p_1 p_2 \cdots p_t = q_1 q_2 \cdots q_s$, donde las p_i y las q_j son transposiciones. Entonces

$$P(\alpha) = p_t(p_{t-1}(\cdots(p_2(p_1(\alpha))))) = (-1)^s \alpha = (-1)^t \alpha.$$

Ya que para cada transposición p_i se verifica que $p_i(\alpha) = -\alpha$ y similarmente para cada q_j se tiene que $q_j(\alpha) = -\alpha$. De aquí resulta que $(-1)^t = (-1)^s$, y de esto que t es par si y sólo si s lo es, o lo que es lo mismo, $s = t + 2i$ donde i es un entero.

La invarianza de la paridad permite precisar entonces la definición de permutación par e impar.

Definición 5.29 *Una permutación de n elementos es llamada una permutación par si puede ser factorizada en un número par de transposiciones y es llamada impar si puede ser factorizada en un número impar de transposiciones.*

Las permutaciones de S_7

$$f = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 4 & 2 & 1 & 6 & 5 & 7 \end{pmatrix} \text{ y } g = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 3 & 6 & 5 & 1 & 7 & 2 \end{pmatrix}$$

se pueden expresar respectivamente como $f = (13)(12)(14)(56)$ y $g = (14)(15)(23)(26)(27)$, por lo que f es par y g es impar.

Puntos de discusión

1. Usar ejemplos para explorar las siguientes ideas. Un producto arbitrario de permutaciones pares es par. El producto de una permutación par por una impar es impar en cualquier orden. El producto de dos permutaciones impares es par. ¿Le permiten estas ideas hacer alguna afirmación sobre los subgrupos de S_n?

2. Demostrar que una permutación y su inversa tienen la misma paridad. Explorar en primera instancia algunos ejemplos.

3. En S_5 si $f = (123)(45)$ y $g = (243)(15)$, encontrar $h \in S_5$ tal que $hfh^{-1} = g$.

4. Expresar la permutación

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 5 & 6 & 2 & 10 & 1 & 8 & 3 & 7 & 4 & 9 \end{pmatrix}$$

como producto de transposiciones. ¿Es una permutación par?

5. Clasificar en par o impar las permutaciones de S_3. ¿Cuántas son pares? ¿Cuántas impares? Explorar en S_4 las mismas preguntas. ¿Puede usted vislumbrar alguna regularidad?

En el punto de discusión 1 anterior se exploró a través de ejemplos el hecho de que un producto arbitrario de permutaciones pares es una permutación par, o sea, que las permutaciones pares son cerradas bajo sucesivas aplicaciones. Esto nos induce a pensar que esta propiedad es válida en general y que entonces el conjunto de todas las permutaciones pares forman un grupo. Este es precisamente el llamado grupo alternante A_n al que nos hemos referido anteriormente, pero cuyo carácter queda establecido por el siguiente teorema.

Teorema 5.48 *$A_n = \{f \in S_n | f$ es par $\}$ es un subgrupo de S_n.*

Demostración. A_n es no vacío. Pues si consideramos (lk) cualquier transposición en S_n, $(lk)(lk) = i$, es decir, la permutación identidad es par.

Para concluir que A_n es subgrupo de S_n basta entonces demostrar que es cerrado para el producto (composición). Para ello, dada $f \in S_n$, podemos asignar a f un número real de la siguiente manera

$$s(f) = \prod_{i<j} \left(\frac{f(j) - f(i)}{j - i} \right).$$

Se tiene entonces que si f, g son elementos de S_n, $s(g \circ f) = s(f)s(g) = s(g)s(f)$. ¡Demostrar esta afirmación!

Además si $f = i$, i idéntica en S_n

$$s(i) = \prod_{i<j} \left(\frac{f(j) - f(i)}{j - i} \right) = 1.$$

Sea ahora $f = (lk)$ una transposición. Tenemos

$$s(f) = \prod_{i<j} \left(\frac{f(j) - f(i)}{j - i} \right) = -1.$$

Si consideramos $f = f_m \circ f_{m-1} \cdots f_2 \circ f_1$, donde cada f_i es una transposición, se tiene que $s(f) = s(f_1) \cdot s(f_2) \cdots s(f_m)$. Luego $s(f) = (-1)^m$.

Si f es una permutación par, es el producto de un número par de transposiciones, y entonces m es par y $s(f) = 1$. Si dadas $f, g \in A_n$, suponemos que $f \circ g$ es impar, su signo sería entonces -1; pero por otro lado se tendría que $s(f \circ g) = s(f) \cdot (g) = 1$. De esta contradicción concluimos que si $f, g \in A_n$, $f \circ g \in A_n$, esto es, A_n es subgrupo de S_n.

Puntos de discusión

1. Construir las tablas de los subgrupos A_3 y A_4 e identificar en cada caso un grupo de transformaciones geométricas que sea isomorfo a estos subgrupos.

2. Identificar para cada uno de los subgrupos del punto anterior un conjunto de generadores.

3. Demostrar que si un grupo de permutaciones tiene orden impar, entonces toda permutación del grupo es par.

En la demostración del teorema anterior observamos que si $f \in S_n$ es un permutación arbitraria $s(f) \in \{-1, 1\}$. Esto nos lleva a definir una aplicación entre el grupo $(S_n, \circ)$ y el grupo multiplicativo $(\{-1, 1\}, \cdot)$, que nos permite caracterizar completamente el subgrupo A_n, como se expresa en la siguiente proposición

Teorema 5.49 *La aplicación $s : (S_n, \circ) \longrightarrow (\{-1, 1\}, \cdot)$ es un homomorfismo sobre, cuyo núcleo es A_n que tiene orden $\frac{n!}{2}$.*

Demostración. En la demostración anterior vimos que si $f, g \in S_n$, $s(f \circ g) = s(f) \cdot (g)$, o sea, s es un homomorfismo. Además de los comentarios previos es claro también que s es sobre. Analicemos pues el núcleo de este homorfismo.

Sea $f \in \ker(s)$. Entonces $s(f) = 1$. Si ahora expresamos f como producto de transposiciones,

$$f = f_m \circ \cdots \circ f_2 \circ f_1, s(f) = (-1)^m,$$

Pero entonces $(-1)^m = 1$, de donde, m es par, es decir $f \in A_n$ y $\ker(s) \subseteq A_n$. Si ahora consideramos $f \in A_n$, $s(f) = 1$ y por tanto $f \in \ker(s)$. Concluimos que $\ker(s) = A_n$.

En la sección de homomorfismos demostramos que el núcleo de un homomorfismo es un subgrupo normal del grupo dominio, en este caso tenemos que $\ker(s) \lhd S_n$, entonces A_n es subgrupo normal de S_n. Pero además, por el primer teorema fundamental de isomorfismos,

$$\frac{S_n}{\ker(s)} \approx \{-1, 1\}, \frac{S_n}{A_n} \approx \{-1, 1\},$$

luego

$$o\left(\frac{S_n}{A_n}\right) = \frac{o(S_n)}{o(A_n)} = 2.$$

Entones $o(A_n) = \frac{n!}{2}$.

Analicemos ahora algunas permutaciones. En S_5,

$$p = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 2 & 5 & 4 \end{pmatrix}$$

Nótese que $p = (23)(45)$ y $p^2 = i$, es decir $o(p) = 2$. Por otra parte, si

$$q = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 1 & 3 & 5 \end{pmatrix},$$

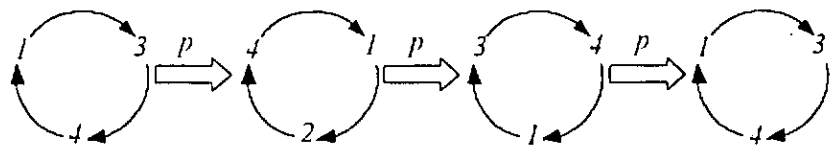
q no es de orden 2 pues

$$q^2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 2 & 4 & 1 & 5 \end{pmatrix}.$$

Pero $q^3 = i$, $o(q) = 3$. Obsérvese que lo único que nos interesa de la permutación q es su efecto sobre el conjunto $\{1, 3, 4\}$, pues deja 2 y 5 invariantes, y por ende basta considerar el ciclo (1 3 4) y ver cómo se transforma al aplicar sucesivamente q.

$$134 \to 413 \to 341 \to 134$$

El orden del ciclo (1 3 4) es entonces 3. En el siguiente diagrama se puede observar el efecto de cada permutación y en el se aprecia claramente la motivación de referirnos a permutaciones de esta forma como ciclos.



Si ahora tomamos p el ciclo (4 3 1), se tiene

$$p = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 1 & 3 & 5 \end{pmatrix},$$

y q la transposición (2 5), esto es,

$$q = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 5 & 3 & 4 & 2 \end{pmatrix},$$

la permutación compuesta

$$q \circ p = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 1 & 3 & 2 \end{pmatrix}$$

es el producto de los ciclos (2 5) de orden 2 y (4 3 1) de orden 3.

Analicemos el orden de $q \circ p$, basta para ello listar sus potencias, i, (qop), $(qop)^2$, $(qop)^3$, $(qop)^4$, $(qop)^4$, $(qop)^5$ y $(qop)^6$

$$\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 1 & 3 & 2 \\ 3 & 2 & 4 & 1 & 5 \\ 1 & 5 & 3 & 4 & 2 \\ 4 & 2 & 1 & 3 & 5 \\ 3 & 5 & 4 & 1 & 2 \\ 1 & 2 & 3 & 4 & 5 \end{array}$$

De la anterior tabla es claro que el orden de $q \circ p$ es 6 que es el producto de los órdenes de los ciclos (2 5) y (4 3 1)

Punto de discusión

Determinar "algebraicamente" el orden de la anterior permutación, teniendo en cuenta que $pq = qp$ y que $q^2 = p^3 = i$.

Formalizaremos en los dos teoremas siguientes las ideas que usted seguramente ya ha vislumbrado en los ejemplos.

Teorema 5.50 *Sea $f = (a_1, a_2, \cdots, a_m)$ un m-ciclo en S_n. Entonces el orden de f es m. $(o(f)) = m$ =longitud de f.*

Demostración. Por definición del m-ciclo,

$$f(a_1) = a_2, f(a_2) = a_3, f(a_3) = a_4 \ldots f(a_{m-1}) = a_m y f(a_m) = a_1,$$

de donde se concluye que

$$f(a_1) = a_2, f^2(a_1) = a_3, f^3(a_1) = a_4, ..., f^{m-1}(a_1) = a_m \text{ y } f^m(a_1) = a_1.$$

Para cada $i = 1, 2, \ldots, m$

$$f^m(a_i) = f^m(f^{i-1}(a_1)) = f^{i-1}(f^m(a_1)) = f^{i-1}(a_1) = a_i.$$

Además, si $k \notin \{a_1, a_2, \cdots, a_m\}$, $f(k) = k$ y por tanto $f^m(k) = k$. Entonces $f^m = i$ y m es el mínimo de los enteros positivos t tal que $f^t = i$. Siendo el menor entero positivo tal que $f^m(a_1) = a_1$, concluimos que m es el orden de f.

Teorema 5.51 *Sea $f = f_m \circ f_{m-1} \circ \cdots \circ f_2 \circ f_1 \in S_n$, donde los f_i son ciclos disyuntos de longitudes $l_m, \ldots, l_2, l_1$ respectivamente. Entonces $o(f) = mcm(l_1, l_2, \ldots, l_m)$.*

Demostración. Basta demostrar que ciclos disyuntos conmutan y aplicar directamente el Teorema 5.3.0. Sean $g, h \in S_n$ dos ciclos disyuntos $g = (a_1 a_2 \cdots a_m)$ y $h = (b_1 b_2 \cdots b_k)$.

Consideremos $S = \{1, 2, 3, \ldots, n\}$ y sean

$$A = \{x \in S | g(x) \neq x\} = \{a_1, a_2, \ldots, a_m\}$$

$$B = \{x \in S | h(x) \neq x\} = \{b_1, b_2, \ldots, b_k\}$$

y $A_1 = S - A$ y $B_1 = S - B$. Como $A \cap B = \emptyset$, dado un $x \in S$ se tiene tan solo una de las tres posibilidades siguientes

i. $x \in A$, (ii)$x \in B$ o (iii)$x \in A_1 \cap B_1$.

(i) Si $x \in A$, $g(x) \in A$ entonces $g(x) \notin B$. Luego $(h \circ g)(x) = h(g(x)) = g(x)$ y $(g \circ h)(x) = g(h(x)) = g(x)$, pues $x \notin B$ es dejado invariante por h. Concluimos que en este caso $g \circ h = h \circ g$.

(ii) Si $x \in B$ el argumento es similar.

(iii) Sea $x \in A_1 \cap A_2$, entonces $x \notin A$ y $x \notin B$. Luego $g(x) = x$ y $h(x) = x$, y entonces $g(h(x)) = g(x) = x$ y $h(g(x) = h(x) = x$. Concluimos también aquí que $g \circ h = h \circ g$.

Usando un argumento inductivo usted puede demostrar que el producto de m ciclos disyuntos también conmuta y aplicar un corolario que generaliza el mencionado teorema.

Puntos de discusión

1. Usar ejemplos para verificar la siguiente afirmación. $p = (a_1, a_2, \cdots, a_m)$, un m-ciclo, es permutación par si y sólo si $m - 1$ es par. Demostrar. ¿Es posible construir un enunciado similar para permutaciones impares?

2. Listar todas las permutaciones de 4 elementos que tienen orden 2.

3. Si $x = (12345)$ hallar x^2, x^3, x^4, etc. ¿Cuál es el orden de x?

4. Si los períodos de tres ciclos disyuntos son respectivamente 4, 6, 3, ¿cuál es el período del producto?

5. Considerar en S_4 las permutaciones $p = (12)$ y $q = (14)$. Determinar $p \circ q$ y $q \circ p$ e identificar el orden de cada uno de estos productos. Explorar otros ejemplos de ciclos no disyuntos e intentar construir alguna generalización.

Nos interesa ahora identificar conjuntos de generadores para los grupos de permutaciones. Seguramente en secciones anteriores usted ya exploró generadores para los subgrupos S_3 y S_4, por ejemplo, las permutaciones

$$a = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix}, b = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix} \text{ y } c = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix}$$

constituyen un conjunto de generadores para S_4. Esto es $S_4 = \langle\{(12), (23), (34)\}\rangle$. Pero también usted puede comprobar que se puede generar el mismo grupo con los ciclos disyuntos (12), (13) y (14) o tan solo con los ciclos (1234) y (12) de órdenes 4 y 2 respectivamente.

Los siguientes teoremas nos identifican conjuntos de generadores para el caso general.

Teorema 5.52 *El grupo $(S_n, \circ)$ es generado por $(12), (13), \cdots, (1n-1)(1n)$.*

Demostración. Si $n = 2$ Los elementos de S_2 son precisamente la identidad i y (12) y desde luego $((12))^2 = i$ es decir $S_2 = \langle(12)\rangle$.

Sea $f \in S_n$, con $n > 2$, $n \in \{1, 2, \cdots, n\}$, $f(n) = m \in \{1, 2, \cdots, n\}$. Si construimos

$$g = (1n) \circ (1m) \circ f,$$

g fija a n, $g(n) = n$, y puede ser considerada como un elemento de S_{n-1}.

Si utilizamos un argumento inductivo y suponemos la proposición válida para $n - 1$, la permutación g puede entonces ser expresada como producto de las transposiciones $(12), (13), \cdots, (1, n-1)$. Entonces como $g = (1n) \circ (1m) \circ f$,

$$f = (1m) \circ (1n) \circ g,$$

de donde f se expresa como producto de las transposiciones $(12), (13), \cdots, (1n-1), (1n)$.

Antes de enunciar el siguiente teorema, requerimos introducir una regla debida a Jordan, que nos facilita la determinación de $f \circ g \circ f^{-1}$ dados $f, g \in S_n$. "Para calcular $f \circ g \circ f^{-1}$ se sustituye cada elemento l en g por $f(l)$"

Si por ejemplo en S_7, $f = (12345) \circ (67)$ y $g = (15) \circ (74) \circ (23)$

$$f \circ g \circ f^{-1} = (21) \circ (65) \circ (34).$$

¡Verificar esta última expresión y demostrar que la regla de Jordan es válida en general!

Teorema 5.53 *El grupo* $(S_n, \circ)$ *es generado por* (12) *y* $(123\cdots n)$.

Demostración. Ya demostramos que si $f \in S_n$, f es producto de transposiciones. Además si (jk) es una transposición

$$(jk) = (1j) \circ (1k) \circ (1j).$$

Por lo tanto, es suficiente demostrar que las transposiciones $(12), (13), \cdots, (1n)$ están en el subgrupo generado por (12) y $(12\cdots n)$. Para ello, usaremos la regla de Jordan.

$$(12) \circ (123\cdots n) \circ ((12))^{-1} = (21345\cdots n).$$

Luego

$$(21345\cdots n)^{-1} \circ (12) \circ (21345\cdots n) = (31) = (13).$$

Entonces (13) y $(21345\cdots n)$ están en el subgrupo generado por (12) y $(12\cdots n)$.

Considerando ahora

$$(13) \circ (213\cdots n) \circ (13)^{-1} = (23145\cdots n),$$

podemos concluir que (14) y $(23145\cdots n)$ están en el generado. Continuando el proceso llegamos a que $(12), (13), \ldots, (1n)$ están en el subgrupo generado por (12) y $(12\cdots n)$. Como ya se demostró que el grupo S_n es generado por las transposiciones $(12), \ldots, (1n)$, es también generado por (12) y $(12\cdots n)$.

Puntos de discusión

1. En S_3 y en S_4 construir los subgrupos generados por todos los 3-ciclos. ¿Qué observa? Investigar una caracterización de estos subgrupos para cualquier $n \geq 3$.

2. Demostrar que $(S_n, \circ)$ puede ser generado por los siguientes conjuntos de ciclos

   (a) $\{(12), (23), \ldots, (n-1n)\}$.

   (b) $\{(123\cdots n-1), (n-1n)\}$.

   (c) ¿ Cúal es el mínimo número de generadors de S_n?

3. En S_8 considerar las permutaciones $p = (1234)(5678)$ y $q = (1537)(2846)$. Demostrar que $\langle\{p, q\}\rangle$ es isomorfo al grupo de los cuaterniones.

4. En S_5 si $a = (12)$, $b = (23)$ y $c = (45)$, encontrar $\langle\{a, b, c\}\rangle$.

5. Sean $f = (ijk)$ un 3-ciclo y g una permutación que reemplaza a i, j, k por l, m y n, respectivamente. Demostrar que $g \circ f \circ g^{-1}$ es el 3-ciclo (lmn).

2.9 Problemas del capítulo

1. Sea $S = \{x \mid x \in \mathbb{Q}, x \neq 0, x \neq 1\}$. Consideremos el siguiente conjunto de funciones definidas sobre S.

$$\left\{ f_1(x) = x, f_2(x) = \frac{1}{1-x}, f_3(x) = \frac{x-1}{x} \right\}$$

Demostrar que $G = \{f_1(x), f_2(x), f_3(x)\}$ forma un grupo con la composición de funciones $(\circ)$.

2. Sea $\mathbb{Q}' = \mathbb{Q} \cup \{\infty\}$. Para cada cuádrupla ordenada de enteros (a, b, c, d) definimos

$$\text{Si } c \neq 0, \qquad f_{a,b,c,d} = \left\{ \begin{array}{ll} \frac{ax+b}{cx+d} & \text{si } x \neq -\frac{d}{c}, \infty \\ \frac{a}{c} & \text{si } x = \infty \\ \infty & \text{si } x = -\frac{d}{c} \end{array} \right\}.$$

$$\text{Si } c = 0, \qquad f_{a,b,c,d} = \left\{ \begin{array}{ll} \frac{ax+b}{d} & \text{si } x \neq \infty \\ \infty & \text{si } x = \infty \end{array} \right\}.$$

Demostrar que estas funciones forman un grupo con la composición $(\circ)$.

3. Para cada número natural n, hallar un conjunto de generadores T del grupo $\mathbb{Z}(+)$ con n elementos tal que si $S \subset T, S \neq T$, S genera un subgrupo propio de $\mathbb{Z}$.

4. Demostrar que el conjunto H de vectores de la forma $\binom{m}{n}$ donde $m, n \in \mathbb{Z}$ forma un subgrupo del grupo $(V, +)$ donde V consta de los vectores $\binom{x}{y}, x, y \in \mathbb{R}$.

5. Demostrar que H es generado por los vectores $\binom{a}{b}$ y $\binom{c}{d}$ si y sólo si $\mid ad - bc \mid = 1$.

6. Usar propiedades del grupo de funciones $G = \{g_1(x) = x, g_2(x) = 1 - x\}$para resolver la siguiente ecuación funcional, es decir, hallar todas las funciones $f(x)$ que satisfacen la ecuación

$$xf(x) = 2f(1-x).$$

Sugerencia: Reemplazar x por $1-x$ y resolver el sistema de ecuaciones simultáneas.

7. Usar propiedades de un cierto grupo de funciones para resolver la siguiente ecuación funcional, es decir, hallar todas las funciones $f(x)$ que satisfacen

$$xf(x) + 2f\left(\frac{x-1}{x+1}\right) = 1.$$

8. Resolver la ecuación funcional

$$xf(x) + 2f\left(-\frac{1}{x}\right) = 1$$

para todo número real $x \neq 0$.

Capítulo 6

Teoría de Anillos

Cuando se habla de los números enteros usualmente se hacen afirmaciones como: tienen estructura de *anillo*, son un dominio de integridad, tienen factorización única, etc. Recordemos el significado de estas afirmaciones y exploremos estas ideas desde una perpectiva mas general.

6.1 Anillos

Definición 6.1 *Por un anillo A se entiende un conjunto A dotado de dos leyes de composición interna, u operaciones binarias, que se notan usualmente $+$ y $\cdot$, llamadas adición y multiplicación tales que*

i. $(A, +)$ es un grupo abeliano.

ii. En A la operación $\cdot$ es asociativa.

iii. Para $a, b, c \in A$, $(a + b) \cdot c = ac + bc$ y $a \cdot (b + c) = ab + ac$.

Definición 6.2 *Si para todo $a, b \in A$, $a \cdot b = b \cdot a$, A se dice anillo conmutativo.*

A continuación consideramos algunos anillos, muy familiares para nosotros desde la básica que se constituyen en modelos para la construcción de una teoría general.

Ejemplos

$(\mathbb{Z},+,\cdot)$ los números enteros con sus operaciones usuales de adición y multiplicación, similarmente los números racionales, los números reales y los complejos.

$(\mathbb{Z}_n,+,\cdot)$ con la adición y multiplicación módulo n a la que hacíamos referencia en el capítulo de grupos.

El conjunto de los polinomios con coeficientes enteros (racionales, reales,...) con la adición y multiplicación usuales.

Todos ellos son además anillos conmutativos.. Muchas de las caracterizaciones que discutiremos estarán marcadamente referidas a estos anillos, pues son aquellos que se trabajan con mayor familiaridad en la matemática escolar.

Otros Ejemplos de Anillos

1. Sea $(G,+)$ un grupo abeliano, y $A(G)$ el conjunto de los automorfismos de G. Entonces $(A(G),+,\circ)$, es un anillo, en general no conmutativo. Es claro ya que la composición de automorfismos es un automorfismo ($\circ$ es operación binaria), que es asociativa y que si $f,g,h \in A(G)$, $f\circ(g+h) = (f\circ g)+(f\circ h)$, esto es, $\circ$ es distributiva con respecto a $+$.

2. Si E es un conjunto y Δ es la diferencia simétrica definida sobre $P(E)$, $(P(E),\Delta,\cap)$ es un anillo conmutativo. Usando propiedades de las operaciones Δ y $\cap$ usted puede verificar que satisface las condiciones requeridas; sin embargo, es interesante reconstruir los argumentos para demostrar estas propiedades, especialmente el relativo al hecho de que la intersección es distributiva con respecto a la diferencia simétrica, a saber, dados $A,B,C \in P(E)$, $A\cap(B\Delta C) = (A\cap B)\Delta(A\cap C))$.

3. Sea
$$M_{2\times 2}(\mathbb{R}) = \{A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} | a,b,c,d \in \mathbb{R}\}$$
se tiene que $(M_{2\times 2}(\mathbb{R}),+,\cdot)$ es un anillo no conmutativo, con $+$ y $\cdot$ adición y multiplicación usual de matrices.

4. Sea A el conjunto de las funciones continuas a valor real definidas sobre el intervalo $[0,1]$. Entonces $(A,+,\circ)$ es un anillo conmutativo, siendo $+$ y $\circ$ la adición y composición usual de funciones.

5. Si $A_1 = \{a+b\sqrt{2} | a,b \in \mathbb{Z}\}$, $(A_1,+,\cdot)$, adición y multiplicación usual de reales, es un anillo conmutativo. Resulta también un anillo, como lo veremos posteriormente, si usted considera $a,b \in \mathbb{Q}$, con la diferencia que éste último tiene una estructura mas completa.

6. Sea $A_2 = \{a+bi | a,b \in \mathbb{Z}\}$, i unidad imaginaria, $(A_2,+,\cdot)$ es un anillo conmutativo, conocido como el anillo de los enteros gaussianos. Similar al ejemplo anterior, si usted considera $a,b \in \mathbb{Q}$ obtiene de nuevo un anillo con una estructura mas completa que la de los enteros gaussianos.

7. Si ahora A es un conjunto que tiene solamente un elemento y $+$ y $\cdot$ denotan las únicas posibles composiciones sobre A, entonces $(A,+,\cdot)$ es un anillo conmutativo. Cualquier anillo que tenga solamente un elemento es llamado anillo nulo (cero), pues su único elemento debe ser el elemento neutro para la adición. En consecuencia cualquier anillo que tenga por lo menos dos elementos es llamado anillo no nulo (no cero).

8. Sea $(A,+)$ un grupo abeliano. Si $\cdot$ es la ley de composición sobre A definida por
$$x\cdot y = 0$$
para todo $x,y \in A$, donde 0 es elemento neutro para $+$, entonces $(A,+,\cdot)$ es un anillo conmutativo. Esta idea resulta interesante pues cualquier grupo abeliano puede ser transformado en un anillo simplemente por definir el producto de cualesquiera dos elementos como el cero. Un anillo cuya multiplicación satisface $x\cdot y = 0$ es llamado un anillo trivial, el anillo nulo es un ejemplo de un anillo trivial.

Notas.

Si $(A, +, \cdot)$ es un anillo, entonces para todo $x, y \in A$:

i. Si 0 es el neutro de $+$, $x \cdot 0 = x \cdot (0+0) = x \cdot 0 + x \cdot 0$ y esto implica por unicidad del neutro que $x \cdot 0 = 0$.

ii. Dados $x, y \in A$, $x \cdot y \in A$, por ser $(A, +)$ un grupo existe el opuesto aditivo (único) de $x \cdot y$, $-(x \cdot y)$, tal que

$$(x \cdot y) + (-(x \cdot y)) = 0$$

Pero,

$$x \cdot y + x \cdot (-y) = x \cdot (y + (-y)) = x \cdot 0 = 0.$$

Por unicidad se tiene entonces que $x \cdot (-y) = -(x \cdot y)$ y por un razonamiento similar se llega también a $(-x) \cdot (y) = -(x \cdot y)$ y de allí a $(-x) \cdot (-y) = x \cdot y$. Todas estas propiedades son muy familiares en nuestro trabajo con los sistemas numéricos.

iii. Si $(A, +, \cdot)$ es un anillo no nulo que posee un elemento identidad para $\cdot$, entonces $1 \neq 0$, pues si $x \in A$, $x \neq 0$, entonces

$$1 \cdot x = x \neq 0 = 0 \cdot x.$$

En el análisis de anillos que tengan un elemento identidad para la multiplicación es conveniente excluir la posibilidad $1 = 0$, es decir el anillo nulo. Se precisa la idea de anillo con elemento identidad en la siguiente definición.

Definición 6.3 *Un anillo con identidad es un anillo no nulo que posee un elemento identidad para la multiplicación.*

El entero 1 es la identidad multiplicativa para el anillo de los enteros modulo m, el automorfismo identidad lo es para el anillo de automorfismos de un grupo abeliano y E es la identidad para el anillo $P(E)$.

Puntos de discusión

1. Explorar si todos los anillos presentados en los ejemplos poseen o no elemento identidad. Identificar éstos.

2. Investigar, tres ejemplos de anillos no conmutativos que no tengan elemento identidad.

3. Sea $(A, +, \cdot)$ un anillo, demostrar que para cualesquiera $x, y, z, w \in A$,

$$(x + y) \cdot (z + w) = x \cdot z + y \cdot z + x \cdot w + y \cdot w.$$

4. Un seudo-anillo es una estructura algebraica $(A, +, \cdot)$ tal que $(A, +)$ es un grupo abeliano, $(A, \cdot)$ es un semigrupo, y

$$(x + y) \cdot (z + w) = x \cdot z + y \cdot z + +x \cdot w + y \cdot w$$

para cualesquiera $x, y, z, w \in A$

(a) Si $\cdot$ es la composición definida en $\mathbb{Z}_3$, por

$$x.y = 1$$

para cualesquiera $x, y \in \mathbb{Z}_3$, entonces $(\mathbb{Z}_3, +, \cdot)$ (donde $+$ es adición módulo 3) es un seudo-anillo pero no un anillo.

(b) Sea $(A,+,\cdot)$ un seudo-anillo y sea $z = 0.0$, demostrar que $a \cdot 0 = z$, para todo $a \in A$.

5. Sea $(A,+)$ un grupo y $\cdot$ una ley de composición asociativa sobre A y distributiva sobre $+$, que posee un elemento identidad 1. Demostrar que $(A,+,\cdot)$ es un anillo.

6. Considerar $(A,+,\cdot)$ un anillo. Si se define sobre A una ley $\circ$ tal que $x \circ y = y \cdot x$, demostrar que $(A,+,\circ)$ es también un anillo. Este anillo es conocido como el anillo recíproco. Explorar esta idea con el anillo $(M_{2\times 2}(\mathbb{R}),+,\cdot)$.

7. Sea p un número primo. $A = \{\frac{m}{n} | m,n \in \mathbb{Z}, n \neq 0$ y p no divide a $n\}$. A es un anillo con la adición y multiplicación usual de racionales. ¿Posee A elemento identidad?

8. Sea $(A,+,\cdot)$ un anillo con identidad 1, se define en A un nuevo par de operaciones así: $a \bigoplus b = a + b + 1$ y $a \bigotimes b = a.b + a + b$. Demostrar que $(A, \bigoplus, \bigotimes)$ es anillo conmutativo con identidad.

Una pregunta que puede surgir aquí de manera natural es la siguiente. Si $(A,+,\cdot)$ es un anillo con identidad, ¿son todos los elementos de A inversibles para $\cdot$?, es decir, para cualquier $x \in A$ existe $y \in A$ tal que $x \cdot y = y \cdot x = 1$. La respuesta inmediata surge de considerar el anillo de los enteros, un anillo con identidad 1, donde es claro que los elementos inversibles son tan solo ± 1. Pero si se considera, por ejemplo, el anillo $(\mathbb{Z}_3,+,\cdot)$ (adición y multiplicación módulo 3), en el todos los elementos no nulos son inversibles, a saber, $1 \cdot 1 = 1$, $2 \cdot 2 = 1$. Por otra parte, no ocurre lo mismo en el anillo $(\mathbb{Z}_6,+,\cdot)$.

Puntos de discusión

1. Identificar los elementos inversibles de cada uno de los anillos de los ejemplos anteriores que posean elemento identidad. En los grupos nos referimos a la existencia de neutro a derecha (o izquierda), inverso a derecha (o izquierda) ¿se da en algunos de estos anillos esta situación?

2. Si $(A,+,\cdot)$ es un anillo con identidad, entonces $T = \{x \in A, \exists y \in A | x \cdot y = y \cdot x = 1\}$ es un grupo con $\cdot$.

Notas. Dado $(A,+,\cdot)$ un anillo,

1. Como $\cdot$ es distributivo respecto a $+$, es decir $\forall x,y,z \in A$ $x.(y+z) = x.y + x.z$ y $(x+y).z = x.z + y.z$, es posible establecer recursivamente la generalización de esta propiedad, esto es si $x, y_1, y_2, \cdots y_m \in A$, entonces

$$x.(y_1 + y_2 + \cdots + y_m) = x.y_1 + x.y_2 + \cdots + x.y_m$$

Y similarmente si $x_1, x_2, \cdots, x_n \in A$, entonces

$$(x_1 + x_2 + \cdots + x_n) \cdot (y_1 + y_2 + \cdots + y_m) = x_1.y_1 + \cdots x_n.y_m$$

En particular $\forall x,y \in A$ y $\forall n \in \mathbb{Z}$,

$$(nx).y = n(x.y) = x.(ny)$$

2. Si $a \in A$ y $m \in \mathbb{Z}^+$, a^m se define de la manera usual. Además, como lo anotamos para los grupos, para todo $m, n \in \mathbb{Z}^+$

$$a^m \cdot a^n = a^{m+n}$$

y

$$(a^m)^n = a^{mn}.$$

Pero en general, así como sucede en los grupos, si $a, b \in A$, $(a \cdot b)^m \neq a^m \cdot b^m$; la igualdad se tiene tan sólo cuando el anillo es conmutativo. De esto se desprende que el conocido Teorema del Binomio, que aplicamos sin problema en el "algebra elemental" y al cual nos hemos referido en otro aparte de este libro, sea válido solamente en el caso en que el anillo A es conmutativo. Se enuncia este resultado en el siguiente teorema. Observe usted cómo un resultado tan familiar se generaliza de manera natural en el contexto de una estructura abstracta, un anillo en el cual los objetos son arbitrarios, y cómo los argumentos de la demostración resultan completamente similares a los usados cuando los objetos son números.

Teorema 6.1 *Sea $(A, +, \cdot)$ un anillo conmutativo. Si $a, b \in A$, entonces para todo entero $n \geq 2$*

$$(a+b)^n = a^n + \sum_{m=1}^{n-1} \binom{n}{m} a^{n-m} \cdot b^m + b^n.$$

Demostración.

Sea S el conjunto de todos los números naturales $n \geq 2$, para los cuales es válido el teorema. Es claro que $2 \in S$, pues

$$[(a+b)^2 = (a+b) \cdot (a+b) = a \cdot (a+b) + b \cdot (a+b) = a^2 + 2a \cdot b + b^2].$$

Por otra parte si $n \in S$,

$$[(a+b)^{n+1} = (a+b)^n \cdot (a+b) = (a+b)^n \cdot a + (a+b)^n \cdot b = a \cdot (a+b)^n + (a+b)^n \cdot b].$$

Formulamos ahora la hipótesis de inducción; suponemos válido el teorema para un n fijo en S, debemos demostrar que es válido para $n+1$.

$$\begin{aligned} a \cdot (a+b)^n &= a \cdot (a^n + \sum_{m=1}^{n-1} \binom{n}{m} a^{n-m} \cdot b^m + b^n) \\ &= a^{n+1} + (\sum_{m=1}^{n-1} \binom{n}{m} a^{n-m+1} \cdot b^m + a \cdot b^n) \\ &= a^{n+1} + (\sum_{m=1}^{n} \binom{n}{m} a^{n-m+1} b^m) \end{aligned}$$

y

$$(a+b)^n \cdot b = (a^n + \sum_{m=1}^{n-1} \binom{n}{m} a^{n-m} \cdot b^m + b^n) \cdot b$$

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

Ahora bien, dado que $\binom{n}{m} + \binom{n}{m-1} = \binom{n+1}{m}$ se tiene

$$(a+b)^{n+1} = a^{n+1} + \sum_{m=1}^{n} \binom{n+1}{m} a^{n+1-m} \cdot b^m + b^{n+1},$$

y el teorema es válido para $n+1$.

Nótese que tan sólo cuando existe un elemento identidad para la multiplicación, la expresión inicial puede transformarse en

$$(a+b)^n = \sum_{m=0}^{n} \binom{n}{m} a^{n-m} b^m,$$

que es la forma que usamos en el algebra elemental. Desde luego allí no nos preocupamos por la existencia de elemento identidad, pues siempre existe.

Analicemos ahora algunos anillos especiales. Si en el anillo de las matrices 2×2 con elementos reales consideramos las matrices

$$A = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}, B = \begin{pmatrix} 2 & 1 \\ 2 & 5 \end{pmatrix}, C = \begin{pmatrix} 7 & 1 \\ 7 & 5 \end{pmatrix}.$$

Nótese que a pesar de que sean diferentes B y C se tiene $A \cdot B = A \cdot C$. Aún mas, si D es cualquier matriz no nula de la forma

$$\begin{pmatrix} a & b \\ a & b \end{pmatrix}$$

$A \cdot D = D \cdot A = 0_{2\times 2}$, matriz cero de tamaño 2×2.

De manera similar si se considera el anillo $(\mathbb{Z}_6, +, \cdot)$, allí $2 \cdot 3 = 4 \cdot 3 = 0$ por una parte, y $2 \cdot 2 = 2 \cdot 5 = 4$ por otra. Esto nos lleva a pensar que no todos los anillos se comportan como nuestros modelos familiares. En los números enteros, en contraste con lo anterior, si $a, b, c \in \mathbb{Z}$, son tales que $a \neq 0$ y $a \cdot b = a \cdot c$ entonces $b = c$, es decir, es válida la propiedad cancelativa. Igual cosa sucede en los enteros, $a \cdot b = 0$ si y sólo si $a = 0$ o $b = 0$. No existen números enteros, racionales, o reales que multiplicados den cero, si ambos son distintos de cero. Estas últimas propiedades son pues características de anillos especiales, así como la conmutatividad o la existencia de elemento identidad. Formalizaremos estas ideas en las siguientes definiciones.

Definición 6.4 *Si $(A, +, \cdot)$ es un anillo, un elemento $a \in A$ se dice un divisor de cero de A si existe $b \in A$, $b \neq 0$ tal que $a \cdot b = 0$ o $b \cdot a = 0$. Un divisor propio de cero es un elemento no nulo de A que es divisor de cero.*

En el anillo de las matrices 2×2, la matriz D es un divisor propio de cero, en el anillo modular $\mathbb{Z}_6$, 2 y 3 son divisores propios de cero.

Si $(A, +, \cdot)$ es un anillo conmutativo con identidad, que no tiene divisores propios de cero, A se dice un *dominio de integridad* (o dominio entero).

Son ejemplos de dominios de integridad, los enteros, los polinomios con coeficientes enteros (o racionales, o reales), el anillo $(\mathbb{Z}_7, +, \cdot)$. En los comentarios previos a las definiciones 4 y 5 se presentaban dos anillos que no son desde luego dominios de integridad y si se analiza el anillo $P(E)$ se podrá concluir que si E tiene mas de dos elementos, este anillo tampoco es un dominio de integridad.

Punto de discusión

Estudiar los anillos presentados en los ejemplos iniciales y determinar cuales de ellos son dominios de integridad.

De nuevo refiriéndonos a los ejemplos que preceden a las definiciones 4 y 5, usted seguramente observó en ellos que existen nexos entre la posibilidad de cancelar y el hecho de ser un divisor de cero. Explicitamos esto en el siguiente teorema.

Teorema 6.2 *Un elemento a de un anillo no nulo A es un divisor de cero si y sólo si a no es cancelable para la multiplicación.*

Demostración.

Si $a \in A$ es un divisor de cero, existe $b \in A$, no nulo tal que $a \cdot b = 0$ o $b \cdot a = 0$. Pero además se tiene que $a \cdot 0 = 0 \cdot a = 0$, es decir $a \cdot b = a \cdot 0$ (o $b \cdot a = 0 \cdot a$) y $b \neq 0$, no es válida pues la cancelativa para a.

Si suponemos ahora que existen $x, y \in A$ tales que $x \neq y$ y $a \cdot x = a \cdot y$, entonces

$$a \cdot (x - y) = a \cdot x - a \cdot y = 0$$

con $(x - y) \neq 0$ esto implica que a es divisor de 0.

Corolario 6.1 *Si $(A, +, \cdot)$ es un anillo conmutativo con identidad las siguientes afirmaciones son equivalentes*

i. A es un dominio de integridad.

ii. $\forall a, b, c \in A$, si $a \neq 0$ y $a \cdot b = a \cdot c$ entonces $b = c$.

Lo anterior significa que un anillo es un dominio de integridad si y sólo si es un anillo conmutativo con identidad en el cual es válida la cancelativa para la multiplicación.

Corolario 6.2 *Si $(A, +, \cdot)$ es un dominio de integridad, entonces un elemento $x \in A$ satisface*

$$x^2 = x$$

si y sólo si $x = 0$ o $x = 1$.

Demostración.

Si $x^2 = x$ y $x \neq 0$, se tiene que $x^2 = x \cdot x = x = x \cdot 1$. Ahora, por ser un dominio de integridad vale la cancelativa y concluimos que $x = 1$. Es claro además que 0 y 1 son soluciones pues $0 \cdot 0 = 0$ y $1 \cdot 1 = 1$.

Consideremos ahora el conjunto

$$A = \{B \in M_{2\times 2}(\mathbb{R}) \mid \det(B) \neq 0\}$$

donde $\det(B)$ es el determinante de B. Usted puede comprobar que $(A, +, \cdot)$ con adición y multiplicación usual de matrices, es un anillo no conmutativo con elemento identidad. Si se explora mas a fondo se observará que no hay allí divisores de cero. Además, se da algo que nos interesa resaltar y que no es válido en el anillo total de las matrices 2×2, toda matriz distinta de la idénticamente nula es inversible para la multiplicación. Es éste un ejemplo de un anillo de tipo especial que definimos a continuación.

Definición 6.5 *Un anillo no nulo con elemento identidad es llamado anillo de división si todo elemento no nulo es inversible para la multiplicación. Un anillo conmutativo con división es llamado un campo. Es decir $(A, +, \cdot)$ es un campo si $(A, +)$ es un grupo abeliano y $(A^*, \cdot)$ es un grupo abeliano.*

Los enteros no son desde luego un anillo de división, pues sus únicos elementos inversibles son ± 1. Mientras que el anillo $(\mathbb{Z}_7, +, \cdot)$ es mas que un anillo de división, es un campo con un número finito de elementos. Son también campos y muy familiares para nosotros $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ y $(\mathbb{C}, +, \cdot)$.

Puntos de discusión

1. Un elemento a de un anillo con identidad que tiene inverso no puede ser un divisor de 0.

2. ¿Constituyen los polinomios con coeficientes racionales un anillo de división?

3. ¿ Es cualquier campo un dominio de integridad? Explorar ejemplos. Construir un argumento general.

4. Ya usted observó en el análisis del anillo de los enteros que no todo dominio de integridad es un campo. Investigar condiciones necesarias y /o suficientes para que un dominio de integridad sea un campo.

5. Si A es un anillo con más de un elemento y con la propiedad de que para cada elemento no nulo a existe un único elemento $b \in A$ tal que $a \cdot b \cdot a = a$. Demostrar que
   i. A no tiene divisores de 0.
   ii. $b \cdot a \cdot b = b$.
   iii. A tiene elemento identidad.
   iv. A es un anillo de división.

6. Sea $(A, +, \cdot)$ un anillo. Un elemento $b \in A$ se dice nilpotente si para algún $n \in \mathbb{N}^*$, $b^n = 0$. (El mínimo de los enteros positivos m tales que $b^m = 0$, es llamado el índice de nilpotencia).
   i. Explorar elementos nilpotentes en el anillo $(\mathbb{Z}_8, +, \cdot)$, en el anillo de los polinomios con coeficientes en $\mathbb{Z}_4$, en el anillo de las matrices 2×2 con elementos reales, en el anillo de los enteros gaussianos.
   ii. Demostrar que un elemento nilpotente no nulo es un divisor propio de cero.
   iii. Demostrar que si A es un anillo con identidad y $b \in A$ es un elemento nilpotente entonces $1 - b$ es inversible. Explorar ejemplos inicialmente.
   iv. Demostrar que si a, b son elementos nipotentes de un anillo conmutativo entonces $a + b$ es nilpotente.

7. Verificar que el anillo de las funciones definidas sobre $[0,1]$ a valor real posee divisores de 0.

Usted habrá observado en los ejemplos que algunos de los anillos modulares tienen divisores de cero, otros no, algunos son dominios de integridad otros no, algunos son campos otros no. Es importante caracterizar completamente el comportamiento de estos anillos pues constituyen una fuente interesante de ejemplos y contraejemplos y porque involucran significativamente elementos y argumentos de la aritmética elemental. El siguiente teorema nos proporciona esta caracterización.

Teorema 6.3 *El anillo de los enteros módulo n, $(\mathbb{Z}_n,+,\cdot)$ es un campo si y sólo si n es un número primo.*

Demostración. Sea $\mathbb{Z}_n = \{0,1,2,\ldots,n-1\}$, con las operaciones de adición y multiplicación módulo n. Si suponemos que $\mathbb{Z}_n$ es un campo y que $n = rs$ con $r,s \in \mathbb{Z}$, $1 < r < n$, $1 < s < n$, es decir que n no es primo, entonces $n = 0 = r \cdot s$, con $r \neq 0$, $s \neq 0$. ($n = 0$ porque el resto de dividir n por n es 0.) Pero en ese caso $(\mathbb{Z}_n^*,\cdot)$ no sería un grupo, pues r y s no serían elementos inversibles. Concluimos entonces que n es un número primo.

Supongamos ahora que $n = p$ es un primo y veamos que $\mathbb{Z}_p$ es un campo. Dado $r \in \mathbb{Z}$ tal que $1 < r < p-1$, r y p son primos relativos y sabemos que existen s y t tales que $rs + tp = 1$. Luego, tenemos que $r \cdot s + t \cdot p = 1$ ($\cdot$, multiplicación módulo p). Como $p = 0$, tenemos que $r \cdot s = 1$. Así que r posee inverso multiplicativo en $\mathbb{Z}_p^*$. Es claro además que $s \neq 0$, es decir r posee inverso multiplicativo en $\mathbb{Z}_p^*$. Concluimos que $(\mathbb{Z}_p,+,\cdot)$ es un campo si p es primo.

Teorema 6.4 *Todo dominio de integridad finito es un campo.*

Demostración. Sea D un dominio de integridad finito, debemos mostrar que D^* es un grupo con la multiplicación y para ver que $(D^*,\cdot)$ es grupo. Basta ver que existe $1 \in D^*$ tal que $1 \cdot a = a \cdot 1 = a$ para todo $a \in D^*$ y que para todo $a \in D^*$, existe $b \in D^*$, tal que $a \cdot b = b \cdot a = 1$.

Consideremos $D = \{a_1, a_2, \cdots, a_n\}$ y $a \in D^*$ fijo. Definimos la aplicación

$$g : D \longrightarrow D$$

por $g(x) = a \cdot x$. g es uno a uno, pues si $g(x) = g(y)$, se tiene que $a \cdot x = a \cdot y$ y como estamos en un dominio de integridad, vale la cancelativa dando $x = y$. Ahora, g es uno a uno y, por ser D finito, g es sobre, de donde, g es una biyección.

Como $a \in D^*$, existe un único $e \in D^*$ tal que $a = a \cdot e$. Veamos que e es elemento identidad para la multiplicación en D^*.

Dado $y \in D$, existe $x \in D$, tal que $y = a \cdot x$, pero entonces

$$y \cdot e = (a \cdot x) \cdot e = (a \cdot e) \cdot x = a \cdot x = y,$$

e actúa como elemento identidad así que notamos $e = 1$.

Dado $1 \in D^*$, existe un único $b \in D^*$, tal que $1 = a \cdot b$, esto es b es inverso de a, $b \neq 0$, pues si $b = 0$, llegaríamos a que $1 = 0$ y esto implicaría $D = \{0\}$.

i. Un subconjunto no vacío A de $P(E)$ es un subanillo si y sólo si para cualesquiera $A, B \in A$, $A \cup B$ y $A - B$ están en A.

ii. Todo subconjunto propio no vacío de E es un divisor propio de cero en el anillo $P(E)$.

5. Sea $(A, +, \cdot)$ un anillo.

i. Sea $a \in A$ fijo, y $S = \{x \in A | x \cdot a = a \cdot x\}$. S es un subanillo de A.

ii. Sea $S \subset A$. El centralizador de S en A, $C_A(S)$ o simplemente $C(S)$ se define como $\{x \in A | s \cdot x = x \cdot s, \forall s \in S\}$. $C(S)$ es un subanillo de A. Si $S = A$, $C(A)$ es el llamado centro de A y es desde luego un subanillo conmutativo de A.

iii. Si S es el subanillo de matrices del punto 2, ¿cúal es el centralizador de S?

iv. En el anillo de las funciones lineales en $\mathbb{R}$, con la adición usual y $\cdot$ la composición de funciones, consideramos $S = \{f(x) = 2x - 1\}$. Identificar $C(S)$.

6. Si $(A, +, \cdot)$ es un anillo en el cual $x^2 - x \in A$, para todo $x \in A$, entonces A es conmutativo. Sugerencia. Demostrar primero que $x \cdot y + y \cdot x \in C(A)$ y a continuación demostrar que $x^2 \in C(A)$.

6.3 Hacia la construcción de nuevas estructuras

6.3.1 Ideales

En los anillos existen subconjuntos estables que son de mayor importancia que los subanillos, pues permiten construir nuevas estructuras, que resultan ser mas completas que la del anillo original. Estos subconjuntos son los llamados *ideales*.

Definición 6.9 *Un subconjunto I de un anillo A se dice un ideal (bilátero) de A si*

i. I es subgrupo aditivo de $(A, +)$.

ii. Dados $a \in I$, $r \in A$, $ar, ra \in I$, o sea $aA \subset I$, $Aa \subset I$, donde, $Aa = \{xa | x \in A\}$ y $aA = \{ax | x \in A\}$.

Un ideal propio de A es un ideal de A que es subconjunto propio.

Notas.

1. Un ideal de un anillo es necesariamente un subanillo, pero un subanillo no necesariamente es un ideal. Por ejemplo $\mathbb{Z}$ es un subanillo pero no un ideal del anillo $\mathbb{R}$.

2. Si I es subgrupo aditivo de $(A, +)$ y se tiene que dado $a \in I$, $aI \subset I$, I se dice ideal a derecha de A. Similarmente si es subgrupo aditivo y dado $a \in I$, $Ia \subset I$, I se dice ideal a izquierda de A. Desde luego todo ideal bilátero es ideal a izquierda y a derecha.

Puntos de discusión

1. En el anillo de las matrices 2×2 con elementos reales buscar ejemplos de ideales a derecha pero no a izquierda, a izquierda pero no a derecha y biláteros.

2. En el anillo de los enteros, $n\mathbb{Z}$, n entero, considerar los subanillos $n\mathbb{Z}=\{nx|x \in \mathbb{Z}\}$. Demostrar que para n arbitrario $n\mathbb{Z}$ es ideal. Explorar primero casos particulares.

3. Para m, n enteros distintos, considerar los ideales de $\mathbb{Z}$, $m\mathbb{Z}$ y $n\mathbb{Z}$. Demostrar que $m\mathbb{Z}+n\mathbb{Z} = \{mx + ny|x, y \in \mathbb{Z}\}$es un ideal de $\mathbb{Z}$? Si $d = mcd(m, n)$, considerar el ideal $d\mathbb{Z}$; explorar ejemplos para determinar la relación que existe entre este último ideal y el ideal $m\mathbb{Z} + n\mathbb{Z}$.

4. Sean $m = 3$ y $n = 5$. Verificar que $m\mathbb{Z} \cap n\mathbb{Z}$ es un ideal. Si tomamos $M = mcm(m, n)$, ¿existe alguna relación entre el ideal $M\mathbb{Z}$ y $m\mathbb{Z} \cap n\mathbb{Z}$? Considerar otros ejemplos y establecer regularidades.

5. Sea A el anillo de las funciones continuas definidas sobre el intervalo $[0, 1]$ a valor real y $J = \{f|f\left(\frac{1}{2}\right) = 0\}$. Verificar que J es un ideal.

6. Sean A un anillo y $a \in A$ fijo. Demostrar que el conjunto $I = \{x \cdot a|x \in A\}$ es un ideal a izquierda de A. En general si $a_1, a_2, \cdots, a_n \in A$, ¿es el conjunto $\{x_1 \cdot a_1 + x_2 \cdot a_2 + \cdots + x_n \cdot a_n|x_i \in A\}$ un ideal a izquierda de A?

7. Sean I, K ideales a derecha de un anillo A. ¿ Es IK ideal a derecha de A? ¿Es $I + K$ ideal a derecha de A?

8. Si $(F, +, \cdot)$ es un campo, caracterizar completamente los ideales de F.

9. Sea A el anillo de las funciones diferenciables definidas sobre el intervalo $(-1, 1)$, y $J_1 = \{f \in A|f'(0) = 0\}$. Demostrar que J_1 es un ideal de A e intentar generalizar la idea para funciones infinitamente diferenciables.

Nota. Sabemos ya que H es un subgrupo de $(\mathbb{Z}, +)$ si y sólo si existe un número natural n tal $H = n\mathbb{Z}$. Además, en uno de los puntos de discusión anterior usted mostró que $n\mathbb{Z}$ es un ideal de $\mathbb{Z}$ para todo n entero. Es claro entonces que los ideales de $\mathbb{Z}$ son de la forma $n\mathbb{Z}$ para algún número natural n.

Los siguientes teoremas generalizan ideas que seguramente usted ya exploró en casos particulares y proporcionan herramientas para construir nuevos ideales de un anillo.

Nota. Dado un anillo A la familia de todos los ideales del anillo es no vacía, pues A y $\{0\}$ son ideales, conocidos como ideales triviales.

Teorema 6.9 *Sean A un anillo y S_i una familia no vacía de ideales de A. Entonces $S = \bigcap_{i \in I} S_i$ es ideal de A.*

Demostración. Como $0 \in S_i$, para todo $i \in I$, la intersección es no vacía. Además dados x, y en la intersección, $x, y \in S_i$ para todo i y como cada S_i es subgrupo aditivo, $x - y \in S_i$ para todo i, de donde, $x - y$ está en la intersección. En resumen, la intersección es un subgrupo de $(A, +)$.

Ahora si $x \in S$ y $y \in A$, $x \in S_i$ para todo i, por ser cada S_i un ideal, $x \cdot a, a \cdot x \in S_i$ para todo i. Concluimos entonces que $x \cdot a, a \cdot x \in S$, y esto significa que S es un ideal de A.

Teorema 6.10 *Dados$(A, +, \cdot)$ un anillo, $S \subset A$. Existe el menor ideal de A que contiene a S y es $\bigcap_{i \in I} T_i$, donde $(T_i)_{i \in I} = \mathcal{T}$, es la familia de los ideales de A que contienen a S.*

Demostración.

La familia $\mathcal{T}$ de los ideales de A que contienen a S es no vacía pues A está en la familia. La intersección de todos los ideales, por teorema anterior, es un ideal y como $S \subseteq T_i$ para todo i, entonces $S \subseteq \bigcap_{i \in I} T_i$. Este es entonces el mas pequeño ideal que contiene a S.

Definición 6.10 - *Sea $(A, +, \cdot)$ un anillo. Si S es un subconjunto de A, el ideal generado por S, que se nota $\langle S \rangle$, es el menor ideal que contiene a S, esto es $\bigcap_{i \in I} T_i$ tal que T_i es ideal y $S \subseteq T_i$.*

Nota. Si $S = \{a_1, a_2, \cdots, a_n\}$, notaremos $\langle S \rangle = \langle \{a_1, a_2, \cdots, a_n\} \rangle = \langle a_1, a_2, \cdots, a_n \rangle$.

Punto de discusión

En los enteros, construir los ideales generados por los siguientes subconjuntos $\{2\}$, $\{3\}$,... $\{n\}$, siendo n cualquier entero positivo. ¿ Qué concluye?

En el punto de discusión anterior usted seguramente intentó dar una presentación mucho mas operativa al ideal generado por un elemento (como lo hicimos con los subgrupos generados), formalicemos esta idea.

Dado un anillo $(A, +, \cdot)$ y $a \in A$, ¿como será el ideal (bilátero) generado por a, $\langle a \rangle$?

Como $\langle a \rangle$ debe ser un subgrupo aditivo de $(A, +)$ y $aA \subset \langle a \rangle$, y, $Aa \subset \langle a \rangle$, tenemos que $x \cdot a \cdot y$, $x' \cdot a$, $a \cdot y'$ y na deben ser elementos $\langle a \rangle$, donde $x, x', y, y' \in A$ y n es entero. Por ser $\langle a \rangle$ subgrupo aditivo, las sumas finitas de elementos de esta forma deben ser elementos del generado. Luego si consideramos

$$I = \{na + x \cdot a + a \cdot y + \sum_{i=1}^{m} x_i \cdot a \cdot y_i | x, y, x_i, y_i \in A, n \in \mathbb{Z}\},$$

¡demostrar las siguientes afirmaciones acerca de I

i. Es un ideal.

ii. $a \in I$, por tanto $\langle a \rangle \subset I$ y $I \subset \langle a \rangle$!

Concluimos entonces que $I = \langle a \rangle$. Si el anillo A posee identidad, la expresión que define I se reduce, pues $na + x \cdot a + a \cdot y = (n1 + x) \cdot a + a \cdot y = x' \cdot a + a \cdot y = x' \cdot a \cdot 1 + 1 \cdot a \cdot s$. Entonces

$$\langle a \rangle = \{\sum_{i=1}^{m} x_i \cdot a \cdot y_i | x_i, y_i \in A, m \geq 1 \in \mathbb{Z}\}.$$

Definición 6.11 *Dado $(A, +, \cdot)$ un anillo, y I un ideal de A.*

i. I se dice *finitamente generado* si existe $S = \{a_1, a_2, \cdots, a_n\} \subset A$ tal que $I = \langle a_1, \cdots, a_n \rangle$.

ii. Un ideal I del anillo A se dice *principal* si existe $b \in A$ tal que $I = \langle b \rangle$.

Definición 6.12 . *Un anillo $(A, +, \cdot)$ se dice de ideales principales, si todo ideal de A es principal. Si además $(A, +, \cdot)$ es un dominio de integridad y todo ideal de éste dominio es principal, diremos que A es dominio de ideales principales.*

El anillo $(\mathbb{Z}, +, \cdot)$ de los números enteros es un dominio de ideales principales; ya hemos observado que si I es ideal de $\mathbb{Z}$, existe $n \in \mathbb{Z}$, $n \geq 0$ tal que $I = n\mathbb{Z} = \langle n \rangle$. Usted puede demostrar además que n es el mínimo de los enteros no negativos que pertenece a I.

Puntos de discusión

1. Sean $(A, +, \cdot)$ un anillo, S, T subconjuntos de A. ¿ Es $\langle S \cap T \rangle = \langle S \rangle \cap \langle T \rangle$? ¿Es $\langle S \rangle \cup \langle T \rangle = \langle S \cup T \rangle$?

2. ¿Es la intersección de ideales a derecha (o a izquierda) también un ideal a derecha (o a izquierda)? Explorar ejemplos.

3. Buscar otros ejemplos de anillos de ideales principales. Sugerencia: Explorar anillos de polinomios con coeficientes enteros, racionales, reales.

Nuestro interés, como lo dijimos al inicio de esta sección, es construir nuevas estructuras a partir de ya conocidas, en particular construir anillos cociente (idea análoga a la de grupos cociente) usando los *ideales*. Pero, como esto nos conduce a un problema de representación, requerimos retomar aquí una idea que trabajamos ampliamente en grupos, la idea de *homomorfismo*.

6.3.2 Homomorfismos de Anillos

Definición 6.13 *Sean $(A, +, \cdot)$ y $(A', \bigoplus, \bigotimes)$ anillos, una aplicación $f : A \longrightarrow A'$, se dice un homomorfismo del anillo A en el anillo A', si para cualesquiera $x, y \in A$,*

i. $f(x + y) = f(x) \bigoplus f(y)$;

ii. $f(x \cdot y) = f(x) \bigotimes f(y)$.

Si existe un homomorfismo del anillo A en el anillo A', decimos que los anillos son *homomorfos*.

Definición 6.14 *Una aplicación $f : A \longrightarrow A'$ se dice un isomorfismo de A sobre A' si f es un homomorfismo uno a uno de A sobre A'. Si existe un tal isomorfismo diremos que A es isomorfo a A' y notaremos $A \approx A'$.*

Nota.

1. La relación "ser isomorfos" es de equivalencia en el conjunto de todo los anillos. El argumento para demostrarlo es completamente similar al que usamos para grupos.

2. Por comodidad en la notación en adelante usaremos $+$ y $\cdot$ para las operaciones de los anillos A y A', pero es importante que usted tenga en cuenta que corresponden a operaciones generales.

Consideremos los anillos $(\mathbb{Z}, +, \cdot)$, de los enteros y el anillo $(M_{2\times 2}(\mathbb{Z}), +, \cdot)$ de las matrices 2×2 con elementos enteros. Definimos una aplicación f de la siguiente manera $f : \mathbb{Z} \to M_{2\times 2}(\mathbb{Z})$, tal que

$$f(n) = \begin{pmatrix} n & 0 \\ 2n & 0 \end{pmatrix}.$$

Nótese que

$$f(n \cdot m) = \begin{pmatrix} nm & 0 \\ 2(nm) & 0 \end{pmatrix} = \begin{pmatrix} n & 0 \\ 2n & 0 \end{pmatrix} \cdot \begin{pmatrix} m & 0 \\ 2m & 0 \end{pmatrix}.$$

Es decir, $f(m \cdot n) = f(m) \cdot f(n)$, similarmente se puede comprobar que $f(m+n) = f(m) + f(n)$ con lo cual se ha demostrado que es los anillos son homomórfos. Si suponemos ahora que $f(m) = f(n)$, concluimos usando propiedades de las matrices que $m = n$, esto es, f es uno a uno. Naturalmente f no es sobre, pero sí podemos afirmar que el anillo $\mathbb{Z}$ es isomórfo a la imagen, esto es, al conjunto de todas las matrices de la forma

$$\begin{pmatrix} a & 0 \\ 2a & 0 \end{pmatrix}$$

que es un ideal del anillo de matrices.

Nótese además que

$$f(0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, f(-n) = \begin{pmatrix} -n & 0 \\ -2n & 0 \end{pmatrix} = -\begin{pmatrix} n & 0 \\ 2n & 0 \end{pmatrix},$$

esto es, $f(-n) = -f(n)$. Finalmente observemos que, como el anillo de los enteros es un anillo con identidad 1, podemos determinar su imagen por f

$$f(1) = \begin{pmatrix} 1 & 0 \\ 2 & 0 \end{pmatrix}.$$

Se puede comprobar que esta matriz actúa como identidad en el ideal de las matrices de la forma

$$\begin{pmatrix} a & 0 \\ 2a & 0 \end{pmatrix}$$

o sea, es la identidad de $f(\mathbb{Z})$. Estas observaciones nos llevan a enunciar un teorema cuya demostración omitimos pues sus argumentos son completamente similares a los utilizados para homomorfismos de grupos.

Teorema 6.11 *Dados $(A, +, \cdot)$ y $(A', +, \cdot)$ dos anillos y f un homomorfismo de A en A' se tiene que*

i. $f(0) = 0'$, donde 0 y $0'$ son los módulos de la adición en A y en A'.

ii. Para todo $a \in A$, $f(-a) = -f(a)$.

iii. $f(A)$ es un subanillo de A'.

iv. Si A es un anillo con identidad 1 entonces $f(1)$ actua como identidad de $f(A)$.

v. Si dado $a \in A$, existe a^{-1}, entonces $f(a^{-1}) = (f(a))^{-1}$.

Por el núcleo (kernel) de un homomorfismo de anillos $f : A \longrightarrow A'$ entenderemos el núcleo de f visto simplemente como un homomorfismo de grupos aditivos, es decir, el conjunto $\{x \in A | f(x) = 0'\}$ donde $0'$ es el neutro aditivo del anillo A'.

Puntos de discusión

1. Por el comentario anterior ya es claro que si $f : A \longrightarrow A'$ es un homomorfismo de anillos $\ker(f)$ es subgrupo de $(A', +)$. Pero realmente es algo mas que un subgrupo aditivo, es un *ideal*. Demostrar esta afirmación.

2. Sea $(L, +, \circ)$ el anillo de las funciones lineales definidas de reales en reales. La aplicación que a cada función f le asocia su valor en 1, $f(1)$ es un homomorfismo de $\mathbb{R}$ en $\mathbb{R}$. Demostrarlo e identificar el núcleo de este homomorfismo.

3. Construir dos homomorfismos del anillo $(\mathbb{Z}, +, \cdot)$ de los enteros en el anillo de las matrices $(M_{2\times 2}(\mathbb{Z}), +, \cdot)$, uno que sea uno a uno y el otro no. Determinar los núcleos de estos homomorfismos.

4. Sea $f : A \longrightarrow A'$ un homomorfismo de anillos. Demostrar que f es uno a uno sí y sólo sí $f^{-1}(\{0'\}) = \{0\}$, donde $0, 0'$ son los respectivos módulos de la adición.

5. Explorar ejemplos para verificar que, si $f : A \longrightarrow A'$ y $g : A' \longrightarrow A''$ son homomorfismos de anillos entonces $g \circ f : A' \longrightarrow A''$ es también, un homomorfismo de anillos.

Vamos a construir ahora la idea de anillo cociente, similar a la de grupo cociente; el papel que jugaban allí los subgrupos normales lo jugarán aquí los ideales.

Consideremos el anillo $(\mathbb{Z}, +, \cdot)$ y un ideal de éste, digamos $2\mathbb{Z} = (2)$. Retomando la idea que desarrollamos en grupos, dados $m, n \in \mathbb{Z}$, definimos en $\mathbb{Z}$ la relación congruencia módulo el subgrupo (ideal) $2\mathbb{Z}$ de la siguiente manera, m es congruente con n módulo el subgrupo (ideal) $2\mathbb{Z}$, si $m - n \in 2\mathbb{Z}$. Esto significa que $m - n$ es un múltiplo de 2. La relación así definida resulta ser de equivalencia y, en consecuencia podemos construir las clases de equivalencia de los enteros según esta relación, a saber,

$$[0] = 0 + 2\mathbb{Z} = \{x \in \mathbb{Z} | x \in 2\mathbb{Z}\} \text{ y } [1] = 1 + 2\mathbb{Z} = \{x | x = 2m + 1 | m \in \mathbb{Z}\}.$$

Formamos un conjunto con estas clases que notamos $\frac{\mathbb{Z}}{2\mathbb{Z}}$. En el caso de los grupos nos interesaba definir entre estas clases una operación, inducida por la del grupo $(+)$; como ahora tenemos un anillo y en el dos operaciones $(+, \cdot)$ tendremos que hablar de las dos operaciones inducidas que van a dotar al cociente de estructura de anillo.

Generalicemos ahora estas ideas.

Definición 6.15 *Sean $(A,+,\cdot)$ un anillo, I un ideal bilátero de A. Si $x,y \in A$ se dice que x es congruente con y módulo I, $x \equiv y(mod I)$ si y sólo si $x-y \in I$.*

Notas.

1. Esta noción de congruencia generaliza en forma muy natural (como lo anotamos en grupos) la congruencia entre números enteros, pues $x \equiv y(\mathrm{mod} n)$ significa $n|x-y$ y esto equivale, desde luego, a la propiedad de que $x-y$ esté en el ideal generado por n, $\langle n \rangle$.

2. La relación $\equiv$ es de equivalencia, hecho que se demuestra con un argumento completamente análogo al de grupos.

3. Dados $x,y,x',y',z \in A$ se tiene

   i. Si $x \equiv y(\mathrm{mod}\, I)$, y, $z \in A$, $x \cdot z \equiv y \cdot z(\mathrm{mod} I)$ y también $z \cdot x \equiv z \cdot y(\mathrm{mod} I)$.

   ii. Si $x \equiv y(\mathrm{mod} I)$ y $x' \equiv y(\mathrm{mod} I)$ entonces $x+x' \equiv y+y'(\mathrm{mod} I)$ y $x \cdot x' \equiv y \cdot y'(\mathrm{mod} I)$.

Para ilustrar demostraremos (ii) que nos servirá de base para afirmar que las operaciones inducidas en el cociente resultan bien definidas.

La hipótesis significa que existen $a,a' \in I$ tales que

$$x = y+a, \qquad \text{y,} \qquad x' = y'+a'.$$

Entonces

$$x+x' = y+a+y'+a',$$

de donde se deduce que $(x+x')-(y+y') \in I$, es decir, $x+x' \equiv y+y'(\mathrm{mod} I)$.

De otra parte

$$x \cdot x' = (y+a)\cdot(y'+a') = y \cdot y' + a \cdot y' + y \cdot a' + a \cdot a'.$$

Dado que I es un ideal bilátero, cada uno de los productos $a \cdot y', y \cdot a', a \cdot a'$ están en I, desde luego sus sumas también, es decir $x \cdot x' - y \cdot y' \in I$, de donde $x \cdot x' \equiv y \cdot y'(\mathrm{mod} I)$.

Con estos elementos podemos ya construir el *anillo cociente* $\frac{A}{I}$. Sea $x \in A$, la clase de equivalencia de x, $[x]$, es como antes, el conjunto de todos los elementos de A que son congruentes con x módulo I, es decir

$$[x] = \{y \in A | x-y \in I\} \ = \{y \in A | y \in x+I.\}$$

Entonces $[x]$ es precisamente $x+I$. Consideramos el conjunto formado todas las clases de elementos de A, notado $\frac{A}{I}$. Podemos dotar este conjunto de estructura aditiva con la adición inducida $(x+I)+(y+I) = (x+y)+I$, y referirnos al grupo cociente aditivo $(\frac{A}{I},+)$ que resulta aquí abeliano. Nos resta solamente definir la multiplicación entre clases para completar la estructura. Dados $x,y \in A$ se define

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

Por observaciones anteriores sobre la relación $\equiv$, la multiplicación está bien definida. Afirmamos que $\frac{A}{I}$ es un anillo, conocido como el anillo cociente. !Revisar que se cumplen todos los axiomas que definen un anillo!.

Teorema 6.12 *Sean $(A,+,\cdot)$ un anillo, I un ideal de A. Entonces $f : A \to \frac{A}{I}$ es un homomorfismo sobreyectivo, llamado canónico, cuyo nucleo* $\ker(f) = I$.

Demostración.

(i) Demostremos que f está bien definida. Si tomamos $x = y$, entonces $0 = x - y$, pero por ser I un ideal $0 \in I$, de donde,

$$x - y \in I, x \in y + I, x + I = y + I,$$

es decir $f(x) = f(y)$.

(ii) Veamos que es un homomorfismo sobreyectivo.

$$f(x + y) = (x + y) + I = (x + I) + (y + I) = f(x) + f(y),$$

por ser I un grupo aditivo $I + I = I$.

$$f(x \cdot y) = (x \cdot y) + I = (x + I) \cdot (y + I) = f(x) \cdot f(y),$$

por ser I un ideal y además $I \cdot I = I$. Por la misma construcción de $\frac{A}{I}$, es claro que si tomamos cualquier elemento en este anillo existe un elemento en A cuya imagen es la clase escogida, de donde se concluye que f es sobre.

(iii) Analicemos $\ker(f)$. Sea $x \in \ker(f)$, entonces $f(x) = I$, ya que I es el elemento identidad ($[0]$) de la adición en el anillo cociente. Se tiene entonces que $x + I = I$, que significa que $x \in I$, es decir, $\ker(f) \subset I$. Si ahora $x \in I$, $x + I = I$, esto es, $f(x) = I$, de donde, $x \in \ker(f)$, o sea, $I \subset \ker(f)$. Se concluye que $\ker(f) = I$.

Teorema 6.13 *(Primer Teorema de Isomorfía). Sean $f : A \longrightarrow A'$ un homomorfismo sobreyectivo del anillo A sobre el anillo A',* $\ker(f) = K$ *su núcleo. Entonces* $A' \approx \frac{A}{K}$.

Demostración.

Consideremos $\phi : \frac{A}{K} \longrightarrow A'$, $\phi(x + k) = f(x)$. Veamos que ϕ está bien definida y que es un homomorfismo biyectivo.

(i) Si $x + K = y + K$, entonces $x - y \in K$, de donde $f(x - y) = 0'$ ($0'$, módulo aditivo de A'). Por ser f homomorfismo se tiene entonces que $f(x) - f(y) = 0'$, y de aquí $f(x) = f(y)$, esto es, $\phi(x + K) = \phi(y + K)$. Luego, ϕ está bien definida.

(ii) Veamos que ϕ es un homomorfismo. Aplicamos para ello que f es un homomorfismo.

$$\begin{aligned}
\phi((x + K) + (y + K)) &= \phi((x + y) + K) = f(x + y) = f(x) + f(y) \\
&= \phi(x + K) + \phi(y + K). \\
\phi((x + K)(y + K)) &= \phi(x \cdot y + K) = f(x \cdot y) = f(x) \cdot f(y) \\
&= \phi(x + k) \cdot \phi(y + K).
\end{aligned}$$

(iii) Si suponemos ahora que $\phi(x + K) = \phi(y + K)$, esto significa que $f(x) = f(y)$, $f(x - y) = 0'$, de donde, $x - y \in K$ o $x \in y + K$. Entonces $x + K = y + K$, esto es, ϕ es uno a uno.

(iv) Que ϕ es sobre resulta directamente del hecho de que f sea sobre. Pues si $y' \in A'$, existe $x \in A$ tal que $f(x) = y'$, por lo tanto $x + K$ es tal que $\phi(x + K) = f(x) = y'$.

Nota.

Si $(A, +, \cdot)$ es un anillo con elemento identidad e para $\cdot$, es posible demostrar por inducción que la aplicación $f : \mathbb{Z} \longrightarrow A$ definida por $f(n) = ne$ es un homomorfismo de anillos. Por ser f un homomorfismo se debe tener que $f(1) = e$; se sigue de nuevo por inducción que la aplicación f está unívocamente determinada. Si A es un anillo no trivial $\ker(f)$ no es todo $\mathbb{Z}$ y por tanto es un ideal propio de la forma $m\mathbb{Z}$. Se sigue entonces del teorema anterior que $\frac{\mathbb{Z}}{m\mathbb{Z}}$ es isomorfo a $f(\mathbb{Z})$. Sin importar como sean los objetos de anillo, se puede representar una "parte de él" por un anillo cociente que está completamente caracterizado.

Recordemos que dado un homomorfismo sobreyectivo entre dos grupos, es posible establecer una correspondencia biunívoca entre subgrupos normales de éstos. El teorema siguiente nos garantiza que es posible establecer una tal correspondencia entre ideales de dos anillos homomorfos volviendo a la idea de que "los homomorfismos conservan la estructura".

Puntos de discusión

1. En el ejemplo previo al teorema 6.11 definíamos un homomorfismo f entre el anillo de los enteros y el de las matrices 2×2 con elementos enteros. Considerar el ideal de $\mathbb{Z}$, $5\mathbb{Z} = \langle 5 \rangle$. Determinar la imagen por f de este ideal, $f(\langle 5 \rangle)$. ¿Es un ideal del anillo de matrices? Si considera usted el ideal I' del anillo de matrices generado por $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. ¿Tiene sentido preguntarse por la preimagen de este ideal, $f^{-1}(I')$?

2. Si $f : A \longrightarrow A'$ un homomorfimo de anillos. Si I es un ideal de A, ¿es $f(I)$ siempre un ideal de A? Si I' es un ideal de A', explorar ejemplos para verificar que, si f es sobre, $f^{-1}(I')$ es ideal de A. Demostrar esta última afirmación.

Teorema 6.14 *Si $f : A \longrightarrow A'$ es un homomorfismo sobreyectiva del anillo A sobre el anillo A'. Existe una correspondencia biunívoca entre el conjunto de ideales del anillo A que contienen a $K = \ker(f)$ y los ideales del anillo A'. Tal correspondencia está dada por asociar a un ideal I de A que contiene a K el ideal $f(I)$.*

Demostración.

Sea I' ideal de A', por ser f sobre, $f^{-1}(I')$ es ideal de A. Se tiene además que dado $x \in K$, $f(x) = 0'$. Pero por ser I' ideal, $0' = f(x) \in I'$ y entonces $x \in f^{-1}(I')$. Luego, el ideal $f^{-1}(I')$ contiene a K.

Consideramos $\mathcal{I} = \{ \text{ I ideal de } A | K \subset I \}$ y $\mathcal{I}' = \{ I' | \text{ ideal de } A' \}$, definimos $F : \mathcal{I} \longrightarrow \mathcal{I}'$ tal que dado $I \in \mathcal{I}$, $F(I) = f(I)$, donde $f(I)$ es ideal de A'.

Veamos que F es uno a uno. Supongamos que $F(I) = F(J)$, siendo $I, J \in \mathcal{I}$, se tiene que $f(I) = f(J)$. Como I, J contienen al núcleo de f, se tiene que $f^{-1}(f(I)) = I$ y $f^{-1}(f(J)) = J$. Concluimos entonces que $I = J$, esto es F es uno a uno.

Demostremos ahora que F es sobre. Sea $I' \in \mathcal{I}'$, si tomamos $I = f^{-1}(I')$, tenemos que I es un ideal que contiene a K y tal que $F(I) = f(I) = I'$; es decir, dado $I' \in \mathcal{I}'$ existe $I \in \mathcal{I}$ tal que $K \subset I$ y $F(I) = I'$. Entonces F es sobre. F es entonces una biyección.

Puntos de discusión

1. Con las hipótesis del teorema 6.14, considerar $I \in I$ e I' su ideal correspondiente, es decir $F(I) = f(I) = I'$. Definir un homomorfismo biyectivo de $\frac{A}{I}$, en $\frac{A'}{I'}$, para concluir que $\frac{A}{I} \approx \frac{A'}{I'}$.

2. Construir un homomorfismo sobreyectivo, del anillo de los enteros en el anillo de las matrices 2×2 con elementos enteros, para verificar la anterior afirmación.

3. Si $(\mathbb{C}, +, \cdot)$ es el campo (cuerpo conmutativo) de los números complejos. Demostrar que $f : \mathbb{C} \longrightarrow \mathbb{C}$ definida por $f(z) = \bar{z}$ es un homomorfismo biyectivo.

4. Construir todos los posibles homomorfismos del anillo $(\mathbb{Z}, +, \cdot)$ en sí mismo. ¿Qué concluye?

5. Sea X un conjunto no vacío. Considerar el anillo $(P(X), \Delta, \cap)$. Sea B un subconjunto de X, fijo. Explorar si la aplicación $f : P(X) \longrightarrow P(X)$ definidio por $f(x) = x \cap B$. es un homomorfismo. Si lo es identificar K núcleo de f y caracterizar $\frac{P(X)}{K}$. (Si lo considera necesario seleccione un conjunto con dos o tres elementos para analizar este problema).

6.3.3 Cuerpo de Cocientes

Nuestra discusión en esta sección requiere que retomemos conocimientos acerca de los números racionales, con los que estamos familiarizados desde la aritmética elemental, pues con éstos como módelo podremos construir conceptos mucho mas abstractos.

Recordemos unos pasos en la construcción de los números racionales a pártir del conjunto de los números enteros. En primer lugar definimos los números racionales a partir del cociente de enteros y a continuación determinamos condiciones para que dos racionales sean iguales, por ejemplo, $\frac{2}{5} = \frac{4}{20}$. El punto importante es que una fracción está determinada por un par de enteros, en este caso especial por $(2, 5)$ pero también por $(4, 20)$. Si nosotros miramos todos los pares que tienen el mismo cociente como equivalentes, hemos encontrado la manera de definir la fracción como una cierta clase de equivalencia de pares de enteros. A continuación se dan "reglas" para la adición y la multiplicación. Veremos que éstas serán exactamente los mismos pasos que usaremos en la generalización.

El anillo de los enteros es un dominio de integridad, o sea, un anillo no trivial, conmutativo, con elemento identidad y sin divisores de cero. Para la discusión general consideremos entonces $(D, +, \cdot)$ un dominio de integridad.

Definición 6.16 *Sean (a, b) y (c, d) pares de elementos en D, con $b \neq 0$ y $d \neq 0$. Decimos que estos pares son equivalentes $(a, b) \equiv (c, d)$ si y sólo si $a \cdot d = b \cdot c$.*

Veamos que $\equiv$ es en efecto una relación de equivalencia.

(i) Es reflexiva, pues para cualesquiera $a, b \in D$, $a \cdot b = b \cdot a$ ya que el anillo es conmutativo, de donde $(a, b) \equiv (a, b)$.

(ii) Si $(a, b) \equiv (c, d)$ entonces $a \cdot d = b \cdot c$, lo cual implica que $c \cdot b = d \cdot a$. Concluimos que $(c, d) \equiv (a, b)$. La relación es simétrica.

(iii) Sean $a, b, c, d, e, f \in D$, con $b, d, f \neq 0$. Si $(a, b) \equiv (c, d)$ y $(c, d) \equiv (e, f)$, se tiene que

$$a \cdot d = b \cdot c \text{ y } c \cdot f = d \cdot e,$$

si multiplicamos la primera igualdad por f y la segunda por b, tenemos,

$$a \cdot d \cdot f = b \cdot c \cdot f \text{ y } b \cdot c \cdot f = b \cdot d \cdot e,$$

entonces $a \cdot d \cdot f = b \cdot d \cdot e$, de donde $d \cdot (a \cdot f - b \cdot e) = 0$. Como D no tiene divisores de cero y $d \neq 0$ se concluye que $a \cdot f - b \cdot e = 0$, y de aquí $a \cdot f = b \cdot e$, o lo que es lo mismo, $(a, b) \equiv (e, f)$. La relación es pues transitiva.

Notamos la clase de equivalencia del par (a, b), por $\frac{a}{b}$ y pasamos a definir la adición y la multiplicación de tales clases.

Si $\frac{a}{b}$ y $\frac{c}{d}$ son clases definimos la adición por

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd},$$

y la multiplicación por

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{a \cdot c}{b \cdot d}.$$

Requerimos ahora demostrar que el resultado de estas operaciones es independiente de la escogencia de los pares (a, b) y (c, d) como representantes de las clases. Para ello suponemos que

$$\frac{a}{b} = \frac{a'}{b'} \text{ y } \frac{c}{d} = \frac{c'}{d'},$$

Debemos probar que

$$\frac{ad + bc}{bd} = \frac{a'd' + b'c'}{b'd'}.$$

Y esto es verdad si y sólo sí

$$b'd'(ad + bc) = bd(a'd' + b'c')$$

o lo que es lo mismo

$$b'd'ad + b'd'bc = bda'd' + bdb'c'.$$

Pero por hipótesis $ab' = a'b$ y $cd' = c'd$. Usando esto podemos concluir que la igualdad anterior se tiene, esto es, la adición está bien definida. ¡Usar un argumento similar para demostrar que la multiplicación está bien definida!

Punto de discusión

Demostrar que el conjunto de los cocientes $\frac{a}{b}$, con $b \neq 0$, con las operaciones de adición y multiplicación antes definidas es un anillo conmutativo con elemento identidad.

Pero resulta aun mas interesante en el paralelo que estamos estableciendo con los racionales, que este conjunto de cocientes no es simplemente un

anillo, sino que es un campo. Si notamos con K dicho anillo, debemos ver que todo elemento no nulo en K tiene inverso multiplicativo. Ahora bien, todo elemento no nulo puede ser escrito en la forma $\frac{a}{b}$, con $b \neq 0$ y $a \neq 0$. Su inverso, como es fácil de comprobar con la definición de multiplicación, resulta ser $\frac{b}{a}$.

Finalmente si consideramos la aplicación $f : D \longrightarrow K$, definida por $f(a) = \frac{a}{1}$, f resulta ser un homomorfismo uno a uno de anillos, llamado una inmersión. Decimos entonces que D puede ser sumergido (inmerso) en K. K se llama el campo de cocientes de D.

Veamos ahora que K es el menor campo que contiene a D, es decir, el menor campo en el cual D está inmerso. Si F es un campo en el que D está inmerso. El conjunto de elementos de la forma $a \cdot b^{-1}$ con $a, b \in D$ y $b \neq 0$ forman un subcampo de F ¡Demostrar!

Si consideramos la aplicación $\phi : K \longrightarrow F$ definida por $\phi(\frac{a}{b}) = a \cdot b^{-1}$, ϕ es un isomorfismo, pues se tiene que

$$\phi\left(\frac{a}{b} + \frac{c}{d}\right) = \phi\left(\frac{ad+bc}{bd}\right) = (ad+bc)\cdot(bd)^{-1} = a\cdot b^{-1} + c\cdot d^{-1} = \phi\left(\frac{a}{b}\right) + \phi\left(\frac{c}{d}\right).$$

Además

$$\phi\left(\frac{a}{b}\cdot\frac{c}{d}\right) = \phi\left(\frac{ac}{bd}\right) = (ac)\cdot(bd)^{-1} = (ab^{-1})\cdot(cd^{-1}) = \left(\phi\left(\frac{a}{b}\right)\right)\left(\phi\left(\frac{c}{d}\right)\right).$$

ϕ es uno a uno, pues si $\phi\left(\frac{a}{b}\right) = \phi\left(\frac{c}{d}\right)$, entonces $a \cdot b^{-1} = c \cdot d^{-1}$. Multiplicando esta igualdad por bd se concluye que $ad = bc$ y esto significa que los pares (a, b) y (c, d) pertenecen a la misma clase de equivalencia,o sea, $\frac{a}{b} = \frac{c}{d}$.

ϕ es además sobre. ¡Explicar porque! Concluimos que K es isomorfo a F, es decir el llamado campo de cocientes es único, salvo isomorfismos.

Nota.

Cuando $D = \mathbb{Z}$, entonces K es por definición el campo de los números racionales. Cuando D es el anillo de los polinomios (con coeficientes enteros), su campo de cocientes es es llamado el campo de las funciones racionales.

Puntos de discusión

1. Sea $A = \{m + n\sqrt{2} | m, n \in \mathbb{Z}\}$. Demostrar que

   (a) A es un dominio de integridad.

   (b) El campo de cocientes de A es $\{r + s\sqrt{2} | r, s \in \mathbb{Q}\}$.

2. Sean $C = \{m + in | m, n \in \mathbb{Z}\}$ y $D = \{m + in\sqrt{2} | m, n \in \mathbb{Z}\}$.

   (a) Demostrar que C y D son dominios de integridad (subdominios de $\mathbb{C}$).

   (b) Identificar subcampos de los complejos que sean campos de cocientes de estos dominios.

   (c) ¿ Son C y D dominios de integridad isomorfos?

3. Algunos textos de Algebra Moderna definen un dominio de integridad como un anillo no trivial, conmutativo, sin divisores de cero. (Esto es, no se exige que el anillo tenga elemento identidad). Investigar cómo construir el campo de cocientes del dominio en este caso.

4. Sean D un dominio de integridad, $a, b \in D$ tales que para m, n enteros primos relativos se tiene que $an = bn$ y $am = bm$. Demostrar que $a = b$.

5. Sea A un anillo con identidad, $f : A \longrightarrow D$ un homomorfismo del anillo en el dominio de integridad D, tal que $\ker(f) \neq A$. Demostrar que $f(1)$ es la identidad de D.

6.3.4 Otros ideales especiales

Nos interesa aquí continuar explorando las formas de construir nuevas estructuras, usando para ello otros tipos especiales de ideales.

Definición 6.17 *Sean A un anillo e I un ideal de A. I se dice un ideal maximal de A si*

i. $I \subset A$, $I \neq A$;

ii. Si J es ideal de A tal que $I \subset J \subset A$ entonces $J = I$ o $J = A$. Es decir I es ideal maximal de A, si I es ideal propio de A y ningún ideal propio de A lo contiene.

Teorema 6.15 . *En $\mathbb{Z}$, $n\mathbb{Z}$ es ideal maximal de $\mathbb{Z}$ si y sólo si n es un número primo.*

Demostración.

Si $n\mathbb{Z}$ es ideal maximal de $\mathbb{Z}$, entonces $n \geq 2$ y si suponemos que $n = a_1 a_2$ con $1 < a_i < n$, veamos que esta suposición origina una contradicción. $n\mathbb{Z} \subset a_1\mathbb{Z}$, pues $n = a_1 a_2 \in a_1\mathbb{Z}$. Pero $1 \notin n\mathbb{Z}$ dado que este ideal es maximal y si 1 estuviera allí coincidiría con $\mathbb{Z}$. Además, $a_1\mathbb{Z} \neq \mathbb{Z}$ pues $a_1 \notin n\mathbb{Z}$, ya que en tal caso $a_1\mathbb{Z}$ coincidiría con $\mathbb{Z}$ y el módulo 1 pertenecería a $a_1\mathbb{Z}$. Esto es absurdo pues $a_1 > 1$ y se tendría que $1 = a_1 \cdot q$ para algún q entero, es decir, a_1 sería un elemento de $\mathbb{Z}$ inversible y los únicos inversibles de $\mathbb{Z}$ son ± 1.

Recíprocamente, supongamos que $n = p$ primo en $\mathbb{Z}$. $p\mathbb{Z}$ es ideal propio de$\mathbb{Z}$, pues $1 \notin p\mathbb{Z}$ por ser $1 < p$, esto es, el ideal $p\mathbb{Z}$ no coincide con $\mathbb{Z}$. De otra parte, si existe un ideal $m\mathbb{Z}$ que contiene pero es distinto de $p\mathbb{Z}$, p y m serían primos relativos. Existen pues, enteros s y t tales que $sp + tm = 1$. Pero $sp + tm \in m\mathbb{Z}$ es decir $1 \in m\mathbb{Z}$ y de aquí $m\mathbb{Z}$=$\mathbb{Z}$; esto es el ideal $p\mathbb{Z}$ es maximal.

Una proposición de la teoría de números, que usamos continuamente en la construcción de argumentos de divisibilidad y que dice que si p es un primo y $p|c \cdot d$ entonces $p|c$ o $p|d$, motiva la siguiente definición.

Definición 6.18 *Si P es ideal de A, P se dice ideal primo de A si dados $a, b \in A$ tales que $ab \in P$ entonces $a \in P$ o $b \in P$. Es decir, si $ab \in P$ y $a \notin P$ entonces $b \in P$.*

Teorema 6.16 *Sea A un anillo conmutativo con identidad 1. Si I es ideal máximal de A entonces I es ideal primo de A.*

Demostración

Sean $a, b \in A$ tales que $ab \in I$, y supongamos que $a \notin I$. Veamos que el ideal generado por I y b es I, esto es $b \in I$.

Si $a \notin I$, el ideal generado por I y a, coincide con A. ¡Explicar por qué! Pero entonces 1 está en el ideal generado por I y por a y existen $s \in A$ y $t \in I$, tales que $1 = t + sa$, de donde $b = bt + sab$, pero $ab \in I$ y $bt \in I$. Se concluye por ser I un ideal, que $b \in I$.

Notas.

1. Nótese que $\langle 4 \rangle = \{4x | x \in \mathbb{Z}\}$ es ideal maximal de los pares, pero no es primo. Pues $2 \cdot 2 = 4 \in \langle 4 \rangle$, pero naturalmente $2 \notin \langle 4 \rangle$. La existencia de identidad requerida en el teorema, no se tiene en este caso.
2. En el anillo de los números enteros, un ideal es primo si y sólo si es maximal; esto es, los ideales primos allí, son precisamente los de la forma $p\mathbb{Z}$ donde p es un primo, hecho que resulta coherente con la idea de primo que manejamos precisamente en los números enteros.

Teorema 6.17 *Sea A un anillo conmutativo con identidad 1 y sea I un ideal de A. I es maximal de A si y sólo si*

$$\frac{A}{I}$$

es un campo (cuerpo conmutativo).

Demostración.

Supongamos que I es ideal maximal de A y veamos que el anillo cociente es un campo. Sabemos ya que $\frac{A}{I}$ es un anillo. Debemos demostrar además que si $a + I \neq I$, existe $b \in A$ tal que $b + I \neq I$ tal que $(a + I)(b + I) = 1 + I$. Sea entonces $a + I \neq I$, esto significa que $a \notin I$. Luego el ideal generado por I y a, $\langle I, a \rangle = \langle 1 \rangle = A$. Existe entonces $b \in A$ tal que $s + ab = 1$, con $s \in I$, luego $ab = 1 - s$ $(s \in I)$. Por tanto,

$$(a + I)(b + I) = ab + I = 1 + I.$$

$b \notin I$, pues si $b \in I$, tendríamos que $ab + s = 1$, la unidad estaría en I y esto contradice que I sea un ideal maximal. Por lo tanto, $a + I$ es inversible en $\frac{A}{I}$ si $a + I \neq I$.

Recíprocamente, si suponemos que $\frac{A}{I}$ es un campo, demostaremos que I es un ideal maximal de A. Sea J ideal de A, tal que $I \subset J \subset A$ y veamos que $J = A$. Si $J \neq A$, existe $a \notin J$. Entonces $a \notin I$ y $a + I \neq I$. Como $\frac{A}{I}$ es campo, existe $b \in A$ tal que

$$(a + I)(b + I) = 1 + I,$$

es decir $ab + I = 1 + I$, luego $ab + s = 1$ para algún $s \in I$, esto es $ab - 1 \in I$, de donde, $ab - 1 \in J$. Como J es ideal de A y $a \in J$ entonces $ab \in J$, y por lo tanto $1 \in J$, esto es $J = A$. Concluimos que I es ideal maximal.

Notas.

1. En $(\mathbb{Z}, +, .)$, $I = m\mathbb{Z}$ es maximal si y sólo si $m = p$, p primo. De lo anterior se tiene que $\frac{\mathbb{Z}}{(m)}$ es campo si y sólo si $m = p$, p primo. Si p es primo el campo $\frac{\mathbb{Z}}{(p)}$ es isomorfo a $\mathbb{Z}_p$. Basta enviar a $a + \langle p \rangle \to [a]$ donde $[a] = \{b \in \mathbb{Z} | b \equiv a (\mathrm{mod} p)\}$.

2. Si el anillo no tiene elemento identidad puede ocurrir que, a pesar de ser I ideal maximal, $\frac{A}{I}$ no sea un campo. Basta considerar nuevamente el anillo de los pares y allí el ideal maximal $\langle 4 \rangle$. En $\frac{A}{\langle 4 \rangle}$, $2 + \langle 4 \rangle$ es un divisor de cero, pues

$$(2 + \langle 4 \rangle)(2 + \langle 4 \rangle) = 4 + \langle 4 \rangle = \langle 4 \rangle.$$

En la nota 1 anterior aparece una condición necesaria y suficiente para que un ideal de $\mathbb{Z}$ sea primo; el siguiente teorema presenta una condición necesaria y suficiente para que un ideal de un anillo mas general sea primo.

Teorema 6.18 *Sean A un anillo conmutativo (no necesariamente con identidad) P un ideal propio de A. P es ideal primo de A si y sólo si $\frac{A}{P}$ es un anillo conmutativo sin divisores de cero.*

Demostración.

Sea P ideal propio de A y supongamos que P es primo. Veamos que $\frac{A}{P}$ no tiene divisores de cero. Sean $a + P$ y $b + P$ distintos de P, $(a + P)(b + P) = ab + P$. Si suponemos que $ab + P = P$, tendríamos que $ab \in P$. Pero, como P es primo, entonces $a \in P$ o $b \in P$. En este caso $a + P = P$ o $b + P = P$, una contradicción. Es claro además que $\frac{A}{P}$ es conmutativo.

Nota.

En los textos en que no se exige existencia de identidad en la definición de dominio, se concluiría entonces aquí que $\frac{A}{P}$ es un dominio de integridad

Si ahora suponemos que $\frac{A}{P}$ no tiene divisores de cero y si $ab \in P$,

$$(a + P)(b + P) = ab + P = P.$$

Pero entonces $a + P = P$ o $b + P = P$, o sea, $a \in P$ o $b \in P$ lo cual implica que P es un ideal primo.

Puntos de discusión

1. En un anillo arbitrario A, si usted considera el ideal $\{0\}$, ¿es un ideal primo?

2. Sobre el producto cartesiano $\mathbb{Z} \times \mathbb{Z}$ se definen dos operaciones

$$(a, b) + (c, d) = (a + c, b + d) y (a, b) \cdot (c, d) = (ac, bd).$$

   Demostrar que $(\mathbb{Z} \times \mathbb{Z}, +, \cdot)$ es un anillo conmutativo con identidad. Encontrar en este anillo un ideal que sea primo pero no maximal.

3. Sean K un campo y S un conjunto. Sea x_0 un elemento de S. Consideramos A, el anillo de las aplicaciones de S en K y J el conjunto de las aplicaciones $f : S \longrightarrow K$, tales que $f(x_0) = 0$. Demostrar que J es un ideal maximal y que $\frac{A}{J}$ es isomorfo a K.

En el anillo de los números enteros, como se deduce del siguiente teorema, ser ideal primo es equivalente a ser maximal. Veamos qué tipo de anillos tienen este mismo comportamiento.

Teorema 6.19 *Sea D un dominio de ideales principales, $I = \langle a \rangle$ ideal no trivial de D, esto es, $I \neq \{0\}$, $I \neq D$. I es ideal primo si y sólo si I es ideal maximal.*

Demostración.

Sean $I = \langle a \rangle$, ideal primo de D, $a \in \langle a \rangle$. Sea J es otro ideal $J = \langle b \rangle$, tal que $\langle a \rangle \subset \langle b \rangle \subset D$, se tiene entonces que $a = br$ para algún $r \in D$. Pero entonces $br \in I$ y como I es primo, se tiene que $b \in I$ o $r \in I$. Si $b \in \langle a \rangle$, se tendría $\langle b \rangle \subset \langle a \rangle$, que contradice que $\langle a \rangle \subset \langle b \rangle$ y $\langle a \rangle \neq \langle b \rangle$. Pero entonces $r \in \langle a \rangle$, de donde $r = as$ para algún $s \in D$. Por tanto $a = br = bas$, o sea, $a = abs$, y por tanto $a(1 - bs) = 0$. Como $a \neq 0$ y D dominio de integridad, concluimos que $bs = 1$. Entonces $1 \in \langle b \rangle$ y esto implica que $\langle b \rangle = D$. I es pues maximal.

De otra parte, en anillos conmutativos con identidad todo ideal maximal es primo, es decir se tiene el recíproco. En el anillo $(\mathbb{Z}, +, \cdot)$, un ideal es primo si y sólo si es maximal.

6.4 Divisibilidad en anillos

En este aparte nos dedicaremos a explorar un punto de especial importancia, la generalización a otros anillos de los conceptos de divisibilidad en números enteros que son presentados desde los niveles básicos. La idea es analizar en qué tipo de anillos podemos definir estas generalizaciones y hasta qué punto se asemejan con resultados ya familiares para nosotros.

Tomaremos para esta discusión A un anillo conmutativo con elemento identidad ($\mathbb{Z}$ desde luego es un ejemplo de tal tipo de anillo).

Definición 6.19 *Sean $(A, +, \cdot)$ un anillo conmutaivo con elemento identidad, $a, b \in A$, $a \neq 0$. Se dice que a divide a b y se nota $a|b$ si existe $c \in A$ tal que $b = ac$.*

Por ejemplo, en el anillo $\mathbb{Z}_7$ que, en efecto, es un anillo conmutativo con identidad, se tiene que $3|5$ pues $5 = 3 \cdot 4$, pero además $3|4$ pues $3 \cdot 6 = 4$.

Nota.

Nótese que si $b = 0$ cualquier $a \neq 0$ de A divide a b, pues $ab = 0$.

Para estos anillos es posible enunciar un primer teorema completamente similar al que se enuncia para los números enteros y los argumentos usados para demostrarlo son completamente análogos.

Teorema 6.20 . *Sean A un anillo conmutativo con unidad, $a, b, c \in A$. Entonces*

i. $a|0$, $1|a$, $a|a$.

ii. $a|1$ si y sólo si a es inversible en A.

iii. Si $a|b$ entonces $ac|bc$.

iv. Si $a|b$ y $b|c$ entonces $a|c$.

v. Si $c|a$ y $c|b$ entonces $c|ax + by$, $\forall x, y \in A$.

Demostración.

Demostraremos tan solo (ii). ¡Demstrar los otros numerales!.

$a|1$, si y sólo si existe $c \in A$ tal que $1 = ac$, y esto es significa precisamente que a es inversible.

Nota.

En los enteros si $a|1$ se concluye que $a = \pm 1$, pues estos son los únicos elementos inversibles de $\mathbb{Z}$.

En nuestro ejemplo de $\mathbb{Z}_7$ nótese que $\forall a \neq 0$, $a|1$. ¡Explicar por qué! Además, como $3|5$, $6|10$, esto es $6|3$ y en efecto $3 = 6 \cdot 4$. Obsérvese que en $\mathbb{Z}_6$, $2 = 5 \cdot 4$, 5 es un elemento inversible de este anillo y $5|2$, pero 2 y 4 no son inversibles.

Definición 6.20 *Sean $a, b \in A$, se dice que a y b son asociados, si existe $c \in A$, c inversible, tal que $a = bc$.*

Notas.

En $\mathbb{Z}_6$, 2 y 4 son asociados. En $\mathbb{Z}$, los asociados de n son n y $-n$, porque como ya lo comentamos antes los únicos elementos inversibles son ± 1.

Puntos de discusión

1. Demostrar que la relación "ser asociados" es una relación de equivalencia.

2. Sea A un anillo conmutativo con unidad, $a, b \in A$, $a \neq 0$. Demostrar que $a|b$ si y sólo si $\langle b \rangle \subset \langle a \rangle$. Verificar esta afirmación en uno de los anillos modulares.

3. Sea $D = \{a + ib | a, b \in \mathbb{Z}\}$, el dominio de los enteros gaussianos.

   (a) Encontrar todos los elementos inversibles de este anillo.

   (b) Si $a + ib \in D$, ¿cúales son sus asociados?

4. Si D es un dominio de integridad y para $a, b \in D$ se tiene que $a|b$ y $b|a$, ¿qué se puede concluir acerca de a y b?

Cuando hablamos de divisores en $\mathbb{Z}$, la idea natural que surge es la de divisor común, es mas, en este caso también nos podemos referir a la noción de *máximo común divisor*.

Definición 6.21 *Sean $(a_1, a_2, \cdots, a_n)$ no nulos en un anillo A, $d \in A$ se dice un máximo común divisor de $a_1, a_2, \cdots, a_n$ si*

1. (a) $d|a_i$, para $i = 1, 2, \cdots n$.

   (b) Si $c|a_i$ para $i = 1, 2, \cdots, n$, entonces $c|d$.

Punto de discusión

En $\mathbb{Z}_7$ hallar un máximo común divisor para $4, 5, 6$. ¿Es este *mcd* único? Si no lo es, explorar qué relación existe entre estos máximos.

Para avanzar en nuestra analogía con los números enteros, nos interesa garantizar, por ejemplo, que el máximo común divisor de dos elementos del anillo puede expresarse como combinación lineal de ellos, relación que hemos usado continuamente. Para ello, requerimos mas que un anillo conmutativo con identidad. Veamos.

Teorema 6.21 *Sea $(A, +, \cdot)$ un anillo y sean $a_1, a_2, \cdots, a_n$ elementos no nulos en A. Existe $mcd(a_1, a_2, \cdots, a_n) = d$, tal que d se puede expresar en la forma*

$$d = r_1 a_1 + r_2 a_2 + \cdots + r_n a_n$$

con $r_i \in A$ si y sólo si $\langle a_1, a_2, \cdots, a_n \rangle$ es principal.

Nota. $\langle a_1, a_2, \cdots, a_n \rangle = \{s_1 a_1 + s_2 a_2 + \cdots + s_n a_n | s_1, s_2, \cdots, s_n \in A\}$, el ideal generado por $a_1, a_2, \cdots, a_n$.

Demostración.

Supongamos que $d = mcd(a_1, a_2, \cdots, a_n)$ existe y puede expresarse en la forma $d = r_1 a_1 + r_2 a_2 + \cdots + r_n a_n$, con $r_i \in A$. Entonces $d \in \langle a_1, a_2, \cdots, a_n \rangle$, luego $\langle d \rangle \subset \langle a_1, a_2, \cdots, a_n \rangle$. De otra parte, como $d = mcd(a_1, a_2, \cdots, a_n)$, para cada a_i existe c_i tal que $a_i = d \cdot c_i$, luego todo elemento de la forma $s_1 a_1 + s_2 a_2 + \cdots + s_n a_n$ que está en el ideal generado por $a_1, a_2, \cdots, a_n$ puede ser escrito como

$$s_1 c_1 d + s_2 c_2 d + \cdots + s_n a_n d = (s_1 c_1 + s_2 c_2 + \cdots, + s_n a_n) \cdot d,$$

y este elemento está en el ideal generado por d. Es decir $\langle a_1, a_2, \cdots, a_n \rangle \subset \langle d \rangle$. En conclusión, $\langle d \rangle = \langle a_1, a_2, \cdots, a_n \rangle$.

Recíprocamente, supongamos que el ideal $\langle a_1, a_2, \cdots, a_n \rangle$ es principal, es decir, existe $d \in A$ tal que $\langle d \rangle = \langle a_1, a_2, \cdots, a_n \rangle$. Entonces, como cada $a_i \in \langle d \rangle$, existe c_i tal que $a_i = c_i \cdot d$, esto es, $d | a_i$ para cada $i = 1, 2, \cdots, n$. Ahora bien, si $c \in A$ es tal que $c | a_i$ para todo i, $a_i = c \cdot d_i$. Pero $d = r_1 a_1 + \cdots + r_n a_n$ por estar en el ideal generado. Entonces

$$d = r_1 d_1 c + r_2 d_2 c + \cdots + r_n d_n c = (r_1 d_1 + \cdots + r_n d_n) \cdot c,$$

es decir, $c | d$ y se sigue que d es máximo común divisor y $d = r_1 a_1 + \cdots + r_n a_n$.

Corolario 6.5 *Si A es anillo de ideales principales y $a_1, a_2, \cdots, a_n \in A$, existe $mcd(a_1, a_2, \cdots, a_n)$ y puede ser escrito como combinación lineal de $a_1, a_2, \cdots, a_n$.*

Aplicando el teorema anterior, dado que $\mathbb{Z}$ es un dominio de ideales principales, se puede garantizar en $\mathbb{Z}$ la existencia del máximo común divisor para cualquier conjunto finito de enteros y su expresión como combinación lineal de éstos. ¡Contrastar los argumentos presentados a continuación con los que se construyeron en el Capítulo 1 para garantizar la existencia del máximo común en el anillo de los enteros!.

Definición 6.22 *Sean A un anillo, $a_1, a_2, \cdots, a_n \in A$. $a_1, a_2, \cdots, a_n$ se dicen primos relativos si $mcd(a_1, a_2, \cdots, a_n) = 1$.*

Nota.

Si en un anillo A, $\langle a_1, a_2, \cdots, a_n \rangle = \langle 1 \rangle = A$, $a_1, a_2, \cdots, a_n$ deben tener un divisor común, el cual es inversible en A.

Punto de discusión

¿Puede usted identificar en el anillo $\mathbb{Z}_7$ un par de elementos que sean primos relativos?

Se presenta ahora una caracterización de los primos relativos completamente análoga a la construida en los enteros, que se cita en algunos textos como identidad de Bezout.

Teorema 6.22 *Sea A es un anillo de ideales principales.*

$$a_1, a_2, \cdots, a_n \in A$$

son primos relativos si y sólo si existen $r_1, r_2, \cdots, r_n \in A$ tales que $1 = r_1a_1 + r_2a_2 + \cdots + r_na_n$.

Demostración.

Si suponemos que $a_1, a_2, \cdots, a_n$ son primos relativos, por definición $1 = mcd(a_1, a_2, \cdots, a_n)$. Aplicando el teorema anterior, existen $r_i \in A$ tal que

$$1 = r_1a_1 + r_2a_2 + \cdots + r_na_n.$$

Si ahora suponemos que existen $r_1, r_2, \cdots, r_n$ en A tales que $1 = r_1a_1 + r_2a_2 + \cdots + r_na_n$, esto significa que 1 está en el ideal generado pór $a_1, a_2, \cdots, a_n$. Pero este ideal coincide entonces con el anillo A. Además para $i = 1, 2, \cdots, n$, $1|a_i$ y si $c|a_i$ para $i = 1, 2, \cdots, n$,

$$c|r_1a_1 + r_2a_2 + \cdots + r_na_n$$

esto es $c|1$. Esto significa que $mcd(a_1, a_2, \cdots, a_n) = 1$, es decir, son primos relativos.

Nótese la importancia de este último teorema. Un anillo de ideales principales puede ser generado por un conjunto finito de elementos primos relativos. Es claro que ésta es la idea presente en los números enteros; con los enteros 5 y 7, por ejemplo, se puede generar el anillo de los enteros; cualquier entero puede ser expresado como combinación lineal de los enteros 5 y 7, o la ecuación diofántica $5x + 7y = k$ tiene solución en $\mathbb{Z}$ para cualquier k entero. Este no es el caso de los enteros 6 y 8 que no son primos relativos. Nuestro objetivo es ahora analizar si en un anillo de ideales principales es posible factorizar "completamente" e identificar elementos primos como en los enteros. Ya tenemos herramientas para una primera aproximación.

Teorema 6.23 . *Sea A un anillo de ideales principales. Si $a, b, c \in A$ y $c|ab$ y $mcd(a, c) = 1$ entonces $c|b$.*

Demostración.

Como $mcd(a, c) = 1$, existen $r, s \in A$ tales que $ra + sc = 1$, de donde, multiplicando por b, $b = rab + scb$. Como $c|ab$, $ab = xc$ para algún $x \in A$ y entonces

$$b = rab + scb = rxc + scb = c(rx + sb)$$

de donde $c|b$.

6.4.1 Elementos primos e irreducibles

En anillos conmutativos con elemento identidad generales se plantean dos ideas que en el anillo de los enteros o en anillo de los polinomios se identifican. Veamos.

Definición 6.23 *Sean A un anillo y $p \in A$. p se dice primo si y sólo si p no es inversible (en $\mathbb{Z}$, $p \neq 1, -1$) y si $p|ab$ entonces $p|a$ o $p|b$.*

Definición 6.24 *$q \in A$ se dice irreducible (o no factorizable) si y sólo si q no es inversible y $q = bc$, con $b, c \in A$ implica b es inversible o c es inversible.*

Notas.

1. q es irreducible equivale a decir que q, no puede ser factorizado de una manera no trivial, es decir, sus únicos factores son q, sus asociados y los elementos inversibles del anillo. En $\mathbb{Z}$, p es primo o irreducible si sus unicos divisores son $\pm p$ y ± 1.

2. 2 es un irreducible (primo) de $\mathbb{Z}$ pero 2 no es un irreducible de $\mathbb{Q}$. Un campo no contiene elementos irreducibles dado que todo elemento no nulo del campo es inversible (es una unidad).

Puntos de discusión

1. En el anillo $(\mathbb{Z}_8, +, \cdot)$, identificar elementos primos e irreducibles.

2. En el anillo $(\mathbb{Z}, +, \cdot)$ de los enteros gaussianos identificar algunos elementos primos. ¿Son irreducibles?

3. Considere usted el anillo de los polinomios con coeficientes enteros, $\mathbb{Z}[x]$. Demostrar que el máximo común divisor de los polinomios 1 y x es 1, pero no existen polinomios $g, h \in \mathbb{Z}$ tales que $1 = 2g + xh$.

4. Sea a un elemento no nulo de un anillo conmutativo con identidad. a es irreducible si y sólo si $\langle a \rangle$ es ideal maximal.

5. Sea A anillo conmutativo con identidad. Si p es primo en A y $p|a_1 a_2 \cdots a_n$ entonces $p|a_i$ para algún i.

6. Verificar, explorando ejemplos, que en un dominio de integridad todo elemento primo es irreducible. Demostrarlo.

Teorema 6.24 *Sean D un dominio de ideales principales, $p \in D$, $p \neq 0$. p es irreducible si y sólo si p es primo.*

Demostración.

Como D es un dominio de integridad, en el Punto de discusión 6 anterior usted demostró que si p es primo es irreducible.

Veamos que si D es dominio de ideales principales y p es irreducible entonces p es primo. Debemos ver que si $p|a \cdot b$, con $a, b \in D$ entonces $p|a$ o $p|b$.

Supongamos que $p|a \cdot b$. Se tiene entonces que existe $c \in D$ tal que $a \cdot b = p \cdot c$. Como D es un dominio de ideales principales, $\langle p, a \rangle = \langle d \rangle$, para algún

$d \in D$, de donde, $p = s \cdot d$ para algún $s \in D$. Como p es irreducible entonces s es inversible o d es inversible. Si d es inversible entonces $\langle p, a \rangle = \langle d \rangle = \langle 1 \rangle = D$. Luego existen $\alpha, \beta \in D$, tales que $1 = \alpha p + \beta a$, de donde

$$b = b\alpha p + b\beta a = \alpha bp + \beta pc = (\alpha b + \beta c) \cdot p.$$

Concluimos en este caso que $p|b$. Si ahora s es inversible en D, entonces $d = p \cdot s^{-1} \in \langle p \rangle$, de donde $\langle d \rangle \subset \langle p \rangle$, luego $\langle a, p \rangle = \langle d \rangle \subset \langle p \rangle$, concluimos que $a \in \langle p \rangle$, esto es $p|a$.

Retomemos aquí algunos elementos de la teoría acerca de números primos y factorización en el anillo de los enteros para contrastar con las ideas que estamos desarrollando.

Un entero $p > 1$ es llamado un número primo cuando sus únicos divisores son los triviales, ± 1, $\pm p$. Un número $m > 1$ que no es primo es llamado compuesto.

Se enuncian a continuación una serie de lemas.

Lema 6.1 *Un primo p es primo relativo con un número n o lo divide.*

Lema 6.2 *Un producto es divisible por un primo p solamente cuando divide a uno de los factores.*

Lema 6.3 *Un producto $q_1 q_2 \cdots, q_r$ de factores primos q_i es divisible por un primo p solamente cuando p es igual a uno de los q_i.*

Lema 6.4 *Todo número n es divisible por algún primo.*

Y se culmina con el

Teorema 6.25 *(Teorema de la Factorización Unica.) Todo número compuesto puede ser factorizado de manera única en factores primos.*

La idea del teorema de factorización única y los lemas que se usan para demostrarlo, aparecen ya en los *Elementos* de Euclides en los libros VII y IX. Incluimos aquí la demostración de este teorema para que posteriormente en la generalización de éste a un anillo, se observen elementos y argumentos comunes.

Demostración.

El primer paso es demostrar que todo número compuesto n es el producto de factores primos. Por uno de los lemas citados anteriormente, existe p_1 primo tal que $p_1|n$, es decir, $n = p_1 n_1$. Si n_1 es compuesto existe p_2 tal que $n_1 = p_2 n_2$. Este proceso puede continuarse con los enteros $n_1, n_2, n_3, \ldots$ hasta encontrar un n_k que sea primo, y entonces $n = p_1 p_2 \cdots p_k$.

Una vez establecida la existencia de una factorización prima el segundo paso consiste en demostrar la unicidad de esta factorización, y para ello basta suponer la existencia de dos factorizaciones primas.

$$n = p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_l$$

Dado que p_i divide al producto de los q-es, se sigue por uno de los lemas que p_i es igual a algún q_j y recíprocamente cada q es igual a algún p. Esto

$d \in D$, de donde, $p = s \cdot d$ para algún $s \in D$. Como p es irreducible entonces s es inversible o d es inversible. Si d es inversible entonces $\langle p, a \rangle = \langle d \rangle = \langle 1 \rangle = D$. Luego existen $\alpha, \beta \in D$, tales que $1 = \alpha p + \beta a$, de donde

$$b = b\alpha p + b\beta a = \alpha bp + \beta pc = (\alpha b + \beta c) \cdot p.$$

Concluimos en este caso que $p|b$. Si ahora s es inversible en D, entonces $d = p \cdot s^{-1} \in \langle p \rangle$, de donde $\langle d \rangle \subset \langle p \rangle$, luego $\langle a, p \rangle = \langle d \rangle \subset \langle p \rangle$, concluimos que $a \in \langle p \rangle$, esto es $p|a$.

Retomemos aquí algunos elementos de la teoría acerca de números primos y factorización en el anillo de los enteros para contrastar con las ideas que estamos desarrollando.

Un entero $p > 1$ es llamado un número primo cuando sus únicos divisores son los triviales, ± 1, $\pm p$. Un número $m > 1$ que no es primo es llamado compuesto.

Se enuncian a continuación una serie de lemas.

Lema 6.1 *Un primo p es primo relativo con un número n o lo divide.*

Lema 6.2 *Un producto es divisible por un primo p solamente cuando divide a uno de los factores.*

Lema 6.3 *Un producto $q_1 q_2 \cdots, q_r$ de factores primos q_i es divisible por un primo p solamente cuando p es igual a uno de los q_i.*

Lema 6.4 *Todo número n es divisible por algún primo.*

Y se culmina con el

Teorema 6.25 *(Teorema de la Factorización Unica.) Todo número compuesto puede ser factorizado de manera única en factores primos.*

La idea del teorema de factorización única y los lemas que se usan para demostrarlo, aparecen ya en los *Elementos* de Euclides en los libros VII y IX. Incluimos aquí la demostración de este teorema para que posteriormente en la generalización de éste a un anillo, se observen elementos y argumentos comunes.

Demostración.

El primer paso es demostrar que todo número compuesto n es el producto de factores primos. Por uno de los lemas citados anteriormente, existe p_1 primo tal que $p_1|n$, es decir, $n = p_1 n_1$. Si n_1 es compuesto existe p_2 tal que $n_1 = p_2 n_2$. Este proceso puede continuarse con los enteros $n_1, n_2, n_3, \ldots$ hasta encontrar un n_k que sea primo, y entonces $n = p_1 p_2 \cdots p_k$.

Una vez establecida la existencia de una factorización prima el segundo paso consiste en demostrar la unicidad de esta factorización, y para ello basta suponer la existencia de dos factorizaciones primas.

$$n = p_1 p_2 \cdots p_k = q_1 q_2 \cdots q_l$$

Dado que p_i divide al producto de los q-es, se sigue por uno de los lemas que p_i es igual a algún q_j y recíprocamente cada q es igual a algún p. Esto

demuestra que ambos lados contienen el mismo número de primos. La única diferencia podría ser que un primo p ocurriera un número mayor de veces en uno de los lados. Pero entonces si cancelamos p un número suficiente de veces en uno de los lados, tendría en uno de los lados p y en el otro no. Esto constituye una contradicción.

Nuestro interés es ahora garantizar como en el anillo de los enteros, la factorización en un anillo mas general. Para ello nos apoyaremos en el Punto de discusión 4 del que usted puede deducir y demostrar el siguiente teorema que relaciona los ideales maximales (primos) en dominios de ideales principales con los elementos irreducibles (primos).

Teorema 6.26 *Sea D un dominio de ideales principales, el ideal $\langle p \rangle$ es maximal (primo) de D si y sólo si p es irreducible (primo) en D.*

Aplicaremos directamente en nuestra discusión un corolario de éste teorema.

Corolario 6.6 *Sea $a \neq 0$, no inversible en un dominio de ideales principales D (en el anillo de los enteros $a \neq 0$, $a \neq \pm 1$). Entonces existe $p \in A$, p primo, tal que $p|a$.*

Nótese la similitud con el Lema 6.3.4 enunciado previo al Teorema de factorización en $\mathbb{Z}$.

Demostración. Como a no es inversible $\langle a \rangle \neq A$, entonces existe I ideal maximal de D tal que $\langle a \rangle \subseteq I$. Pero por teorema anterior, si I es maximal $I = \langle p \rangle$ donde p es primo de D, no hay distinción entre irreducible y primo en este anillo. Luego $\langle a \rangle \subseteq \langle p \rangle$,es decir $p|a$.

El corolario anterior motiva la aproximación definitiva a la factorización.

Definición 6.25 *Un dominio de integridad D se dice un dominio de factorización única si*

i. Para todo $a \in D$, $a \neq 0$ y a no inversible, a puede factorizarse como producto de un número finito de elementos irreducibles, y

ii. Si $a = p_1 p_2 \cdots p_n = q_1 q_2 \cdots q_m$, con p_i y q_j irreducibles, entonces $n = m$ y existe una permutación ϕ de los indices tal que p_i y $q_{\phi(i)}$ son asociados, para $i = 1, 2, \cdots, n$.

Naturalmente $(\mathbb{Z}, +, \cdot)$ es un dominio de factorización única.

Puntos de discusión

1. ¿Es $\mathbb{Z}_5$ un dominio de factorización única?
2. Considerar el anillo $(\mathbb{Z}[x], +, \cdot)$, de los polinomios con coeficientes enteros. ¿Es un dominio de integridad? ¿Es un dominio de factorización única?
3. ¿Es $\mathbb{Q}$ un dominio de ideales principales? ¿ Es $\mathbb{Q}$ un dominio de factorización única?
4. $\mathbb{Z}[\sqrt{2}] = \{a + b\sqrt{2} | a, b \in \mathbb{Z}\}$ es un dominio de integridad. Investigar si es un dominio de factorización única. Sugerencia: Caracterizar elementos irreducibles.

Teorema 6.27 *Sea D un dominio de ideales principales. Entonces todo elemento de D que no es cero, ni inversible posee una factorización como producto de un número finito de primos.*

Demostración. Sea $a \neq 0$ no inversible en D. Por corolario anterior, existe p_1 primo, $p_1 \in D$, $p_1|a$ y, por ende, existe $a_1 \in D$ tal que, $a = p_1 a_1$. Luego $\langle a \rangle \subseteq \langle a_1 \rangle$. Pero si $\langle a \rangle = \langle a_1 \rangle$, $a_1 = ra$ para $r \in A$ y se concluye que $a = p_1 a_1 = p_1 ra$, de donde $1 = p_1 r$. Esto contradice que p_1 es primo pues p_1 sería inversible. Luego $\langle a \rangle \subset \langle a_1 \rangle$ y, repitiendo el procedimiento para a_1, se obtiene una sucesión de ideales principales

$$\langle a \rangle \subset \langle a_1 \rangle \subset \langle a_2 \rangle \subset \cdots \subset \langle a_n \rangle \subset \cdots$$

donde $a_{n-1} = p_n a_n$ para algún primo $p_n \in D$. Este proceso continúa hasta que a_n sea inversible en D. Es posible demostrar que este proceso debe parar, esto es, que existe n tal que a_n es inversible y $\langle a \rangle \subset \langle a_1 \rangle \subset \cdots \subset \langle a_n \rangle \subset D$.

Se deduce entonces que a es expresable como producto de un número finito de primos, $a = p_1 p_2 \cdots p_{n-1} p'_n$ donde $p'_n = p_n \cdot a_n$ es asociado de un primo y es, por tanto, también un p'_n primo.

Puntos de discusión

1. Contrastar los argumentos de esta demostración con los del Teorema de Factorización Unica en en $\mathbb{Z}$.
2. Demostrar que si D es un dominio de ideales principales e I un ideal propio de D, entonces I es producto de un número finito de ideales (maximales) primos. Explorar inicialmente esta idea con ideales del anillo $\mathbb{Z}$.

El siguiente teorema sobre los números primos apareció por primera vez en los *Elementos* de Euclides (Proposición 20, Libro IX).

Teorema 6.28 *Existe una infinitud de primos.*

La demostración de Euclides es como sigue. Sean $a, b, c \cdots k$ una familia de números primos. Tome su producto $P = ab \cdots k$ y sume 1. Entonces $P + 1$ es un primo o no es un primo. Si es primo entonces agregaríamos un primo a los anteriores. Si no lo es, debe ser divisible por algún primo p. Pero p no puede coincidir con ninguno de los primos dados $a, b, \cdots k$, porque entonces dividiría a P y a $P + 1$ y de allí dividiría a su diferencia, esto es dividiría a 1, y esto es imposible. Por tanto, siempre es posible encontrar un nuevo primo dado un conjunto (finito) de primos.

Ilustremos antes de continuar con nuestra discusión general, la construcción de algunos primos por el método de Euclides

$$\begin{aligned} 2 \cdot 3 + 1 &= 7 \\ 2 \cdot 3 \cdot 5 + 1 &= 31 \\ 2 \cdot 3 \cdot 5 \cdot 7 + 1 &= 211 \end{aligned}$$

y así sucesivamente.

Enunciamos ahora el resultado análogo a éste teorema de Euclides en un dominio de ideales principales D.

Teorema 6.29 *Existe un número infinito de primos en D.*

Demostración. Supongamos que existe un número finito $p_1, p_2, \ldots, p_n$, de primos en D. Definimos en D, $a = (p_1 \cdot p_2 \cdots p_n) + 1$. a no es divisible por ninguno de los p_i; pues si $p_i | a$, $p_i(a - p_1 p_2 \cdots p_n) = 1$, que significa que p_i es inversible y esto contradice que p_i es primo. Como $a > 1$, por Corolario 6.6 a debe tener un factor primo, luego a es divisible por un primo que no está entre los $p_1, p_2, \cdots, p_n$. Concluimos entonces que el número de primos no puede ser finito.

Nótese la similitud entre las dos demostraciones; la diferencia central está en la referencia a los inversibles, que el caso de los enteros positivos se reduciría a la imposibilidad de que p_i sea 1 (único elemento inversible en los enteros positivos).

Nota.

La demostración sobre unicidad de la factorización en el caso general, que no se incluyó en la discusión anterior, tiene lineamientos completamente similares a los presentados en el caso de los enteros.

Punto de discusión

Demostrar que todo dominio de ideales principales es un dominio de factorización única.

A lo largo nuestra discusión de la teoría de grupos en el capítulo anterior y de la teoría de anillos en él presente, hemos aplicado en múltiples ocasiones el llamado *algoritmo de la división* a pares de números enteros. Es claro que este algoritmo constituye una valiosa herramienta en la caracterización de la estructura del anillo $(\mathbb{Z}, +, \cdot)$,así como en la construcción de argumentos que involucran enteros positivos, como son el orden de un grupo o de un elemento. Es por eso que la pregunta a tratar ahora es la siguiente. ¿Es válido el algoritmo de la división en anillos mas generales?

Para responder esta pregunta recordemos la versión actual del algoritmo de la división en $\mathbb{Z}$ con el objeto de motivar por medio de ella la introducción de las ideas generales.

"Si $a, b \in \mathbb{Z}$, $b \neq 0$, existen $q, r \in \mathbb{Z}$ tales que $a = qb + r$, con $0 \leq r < |b|$."

Definición 6.26 *Un dominio de integridad D, es decir, un anillo conmutativo, con identidad y sin divisores de cero, se dice un dominio euclidiano si existe una función*

$$d : A - \{0\} \to \mathbb{Z}^+$$

que cumple

i. $d(a)$ es entero no negativo para todo $a \in D - \{0\}$.

ii. Si $a, b \in D$, $a \neq 0$, $b \neq 0$, $d(ab) \geq d(a)$.

iii. Para $a, b \in D$, $b \neq 0$, existen $q, r \in D$, $a = qb + r$, con $r = 0$, o, $d(r) < d(b)$ si $r \neq 0$.

Notas.

1. Para el anillo de los números enteros, $d(a) = |a|$; este anillo es desde luego un dominio Euclidiano.

2. En $K[x]$, anillo de polinomios con coeficientes en un campo K, si $p(x) \in K[x]$, $d(p) =$ grado de $p(x)$. En este anillo son válidas además (ii) y (iii), como el lector puede verificar; se constituye, por tanto, en otro ejemplo de un dominio Euclidiano.

Restaría observar ahora si el "resto" y el "cociente" que aparecen en la definición de dominio euclidiano son únicos, como se garantiza en el algoritmo de la división. Para ello analicemos primero algunas propiedades de la función d en un dominio euclidiano.

Teorema 6.30 *Sea D un dominio euclidiano, con función d, entonces*

i. Si $a \neq 0$, $a \in D$, $d(a) \geq d(1)$.

ii. Si $a \neq 0$ en D, a es inversible si y sólo si $d(a) = d(1)$.

Demostración. (i) Como $a = a \cdot 1$, luego $d(a) = d(a \cdot 1) \geq d(1)$. Para ejemplificar, en $\mathbb{Z}$, si $a \neq 0$, $|a| \geq |1|$, en el anillo de los polinomios todo polinomio no constante tiene grado mayor o igual que 1.

(ii) Sea $a \neq 0$, a inversible. Existe b tal que $ab = 1$, luego

$$d(a) \leq d(ab) = d(1) \leq d(a),$$

y entonces $d(a) = d(1)$. Nuevamente, veamos lo que sucede en un anillo modelo. En $\mathbb{Z}$ los únicos elementos inversibles son ± 1 y $|1| = |-1|$.

Recíprocamente sea $a \neq 0$, tal que $d(a) = d(1)$. Existen $q, r \in D$ tales que $1 = aq + r$, donde $r = 0$ o $d(r) < d(a)$, aplicando la propiedad (iii) del dominio euclidiano. Pero $d(a) = d(1)$, entonces $d(r) = d(r \cdot 1) < d(1)$, lo que es una contradición. Luego $r = 0$ y de aquí $1 = aq$, es deci,r a es inversible en D. En el anillo de los polinomios con coeficientes en un campo los únicos elementos inversibles son los polinomios constantes no nulos y su grado, que es 0, coincide con el grado del polinomio constante 1, que desde luego es también 0.

Estamos ahora preparados para dar condiciones para la unicidad.

Teorema 6.31 *El resto y el cociente en la condición (iii) de la definición de dominio euclidiano son únicos si y sólo si $d(a + b) \leq \max(d(a), d(b))$.*

Demostración. Si existen a, b no nulos en D tales que $d(a+b) > \max(d(a), d(b))$, tenemos que

$$b = 0(a + b) + b = 1(a + b) - a,$$

con $d(-a) = d(a) < d(a + b)$ y $d(b) < d(a + b)$. En este caso entonces se observa la falta de unicidad del resto y el cociente.

Recíprocamente si la desigualdad se tiene y $a \in D$, $a = qb + r$, con $r = 0$ o $d(r) < d(b)$ y suponemos que el resto y el cociente no son únicos. Sea también $a = q'b + r'$, con $r' = 0$ o $d(r') < d(b)$. Si $r \neq r'$ y $q \neq q'$,

$$d(b) \leq d((q - q')b) = d(r - r') < \max(d(r), d(-r')) < d(b).$$

Pero esto es posible sólo si $r - r' = 0$ o $q - q' = 0$. Si una de estas dfierencias es igual a 0, la otra también lo es. Se concluye entonces la unicidad del cociente y el resto.

Puntos de discusión

1. Contrastar el argumento de la demostración anterior con el que se usa para demostrar la unicidad del cociente y el residuo, en $\mathbb{Z}$.

2. Explorar si, en un dominio euclidiano y por aplicación reiterada del algoritmo de la división, es posible construir un algoritmo similar al de Euclides que permita encontrar el máximo común divisor de dos o mas elementos.

Teorema 6.32 *Si D es un dominio euclidiano entonces D es un dominio de ideales principales.*

Demostración.

Sea D un dominio euclidiano, con función d. Si I es ideal de D y $I = \{0\} = \langle 0\rangle$ o si $I = D = \langle 1\rangle$, se sigue que es principal. Supongamos entonces que I es un ideal de D no trivial y sea

$$S = \{d(a) | a \in I, a \neq 0\}.$$

S es un subconjunto no vacío de los naturales, de manera que, por el *principio de buena ordenación*, existe un elemento mínimo en S, es decir, existe $b \in S$ tal que $d(b)$ es mínimo en S. Veamos que $I = \langle b\rangle$. Si $a \in I$ por ser éste un dominio euclidiano, existen $q, r \in D$ tales que $a = bq + r$, con $r = 0$ o $d(r) < d(b)$. Ahora, $r = a - bq \in I$, dado que $a \in I$ y $bq \in I$. Si $r \neq 0$, se tendría que $d(r) < d(b)$ y esto contradice la minimalidad de $d(b)$ en S. Concluimos entonces que $a = qb \in \langle b\rangle$, esto es $I \subseteq \langle b\rangle$. Como además $b \in I$, se tiene que $\langle b\rangle \subseteq I$, esto es, $I = \langle b\rangle$. D es entonces un dominio de ideales principales.

Corolario 6.7 *Todo dominio euclidiano es un dominio de factorización única.*

Demostración.

Por el Teorema 6.32 si D es dominio euclidiano entonces es dominio de ideales principales. Pero ya sabemos además que todo dominio de ideales principales es dominio de factorización única. Luego, D es dominio de factorización única.

Para concluir esta sección analicemos el dominio de integridad $\mathbb{Z}(\sqrt{n}) = \{a + b\sqrt{n} | a, b \in \mathbb{Z}\}$, donde n es un entero libre de cuadrados, es decir, n no es divisible por m^2, $m \in \mathbb{Z}$, $m > 1$. Nuestro objetivo es concluir que $\mathbb{Z}(\sqrt{n})$ es dominio euclidiano para valores especiales de n.

Requerimos para llegar a este resultado definir en este dominio el concepto de norma análogo al de valor absoluto en $\mathbb{Z}$.

Si $\alpha = a + b\sqrt{n} \in \mathbb{Z}(\sqrt{n})$, $N(\alpha) = \alpha\overline{\alpha} = (a + b\sqrt{n})(a - b\sqrt{n}) = a^2 - b^2 n$ se llama la norma de α.

Puntos de discusión

1. Si $\alpha, \beta \in \mathbb{Z}$ demostrar que
   i. $N(\alpha) = 0 \iff \alpha = 0$.
   ii. $N(\alpha\beta) = N(\alpha)N(\beta)$.
   iii. $N(1) = 1$.

2. Verificar que la norma permite caracterizar elementos inversibles e irreducibles de $\mathbb{Z}(\sqrt{n})$.

3. Demostrar que si $\alpha \in \mathbb{Z}$ entonces

   i. $N(\alpha) = \pm 1 \iff \alpha$ es inversible en $\mathbb{Z}$.

   ii. Si $N(\alpha) = \pm p$, p primo, entonces α es irreducible en $\mathbb{Z}$. Explorar estas ideas en $\mathbb{Z}(\sqrt{2})$

Nota.

Consideremos por ejemplo $D = \{a + b\sqrt{10} | a, b \in \mathbb{Z}\}$. Demostremos que $2, 3, 4 + \sqrt{10}, 4 - \sqrt{10}$ son irreducibles pero no primos en D.

En primer lugar $N(a + b\sqrt{10}) = a^2 - 10b^2$. Si suponemos entonces que $2 = a \cdot b$,

$$N(2) = 2^2 = 4 = N(a)N(b).$$

Entonces $N(a) = \pm 1, \pm 2$ o $N(b) = \pm 1, \pm 2$.

Pero la ecuación de Pell $x^2 - 10y^2 = c$, tiene soluciones enteras de valor absoluto menor o igual que 10 sólo para $c = 0, \pm 1$. Concluimos entonces que $N(a) = \pm 1$ o $N(b) = \pm 1$ y entonces a o b son inversibles, es decir, 2 es irreducible.

Similarmente para 3 y $4 \pm \sqrt{10}$, como, $N(3) = 9$ y $N(4 \pm \sqrt{10}) = 6$, se concluye que 3 y $4 \pm \sqrt{10}$ son irreducibles de D.

De otra parte, 2 no divide a $4 \pm \sqrt{10}$, pues $N(2) = 4$ no divide a $N(4 \pm \sqrt{10}) = 6$ en $\mathbb{Z}$, pero $2|6 = 2 \cdot 3 = (4 + \sqrt{10})(4 - \sqrt{10})$, esto es 2 no es primo.

Argumentos similares se tienen para 3 y $4 \pm \sqrt{10}$ pues $3|(4 + \sqrt{10})(4 - \sqrt{10})$ y $6 = 2 \cdot 3|(4 \pm \sqrt{10})$.

Teorema 6.33 $\mathbb{Z}(\sqrt{n})$, *para* $n = -1, -2, 2, 3$ *es dominio euclidiano.*

Demostración.

Consideremos

$$d : \mathbb{Z}(\sqrt{n}) - \{0\} \longrightarrow \mathbb{Z}^+(\sqrt{n})$$

por $d(\alpha) = |N(\alpha)|$ y veamos que d es la función correspondiente al dominio euclidiano cuando $n = -1, -2, 2, 3$. En primer lugar $d(\alpha) = 0$ si y sólo si $\alpha = 0$, de donde $d(\alpha) \geq 1$ para todo $\alpha \neq 0$.

Si $\alpha, \beta \neq 0$ en $\mathbb{Z}(\sqrt{n})$,

$$d(\alpha\beta) = |N(\alpha\beta)| = |(N(\alpha)N(\beta)| = d(\alpha)d(\beta) \geq d(\alpha).$$

Es decir, se cumple (ii) de la definición de dominio euclidiano.

Como $\beta \neq 0$, $\alpha\beta^{-1} \in \mathbb{Q}(\sqrt{n})$, campo de cocientes de $\mathbb{Z}(\sqrt{n})$, de donde, $\alpha\beta^{-1} = a + b\sqrt{n}$, con $a, b \in \mathbb{Q}$.

Sean $x, y \in \mathbb{Z}(\sqrt{n})$ tales que $|x - a| \geq \frac{1}{2}$, $|y - b| \geq \frac{1}{2}$. Si $z = x + y\sqrt{n}$, entonces $z \in \mathbb{Z}(\sqrt{n})$ y, como en $\mathbb{Q}(\sqrt{n})$ está también definida la norma, se tiene que

$$|N(\alpha\beta^{-1}) - z| = |N((a - x) - (b - y)\sqrt{n})| = |(a - x)^2 - n(b - y)^2|.$$

De la manera como se escogen x, y se deduce que

$$\begin{aligned} -\frac{n}{4} &\geq (a-x)^2 - n(b-y)^2 \geq \frac{1}{4}, \text{ si } n > 0, \\ 0 &\geq (a-x)^2 - n(b-y)^2 \geq \frac{1}{4} + (-n)\frac{1}{4}, \text{ si } n < 0 \end{aligned}$$

En términos de la función d esto equivale a

$$d(\alpha\beta^{-1} - z) = |(a-x)^2 - n(b-y)^2| < 1,$$

para $n = -1, -2, 2, 3$. Tomando $w = \beta(\alpha\beta^{-1} - z)$, se tiene que $\alpha = z\beta + w$. Como α y $z\beta$ están en $\mathbb{Z}(\sqrt{n})$, $w \in \mathbb{Z}(\sqrt{n})$, para $n = -1, -2, 2, 3$

$$d(w) = d(\beta(\alpha\beta^{-1} - z) = (d(\beta))(d(\alpha\beta^{-1} - z)) < d(\beta).$$

Se verifica pues la definición de dominio euclidiano para $n = -1, -2, 2, 3$.

Corolario 6.8 *$\mathbb{Z}(i)$es dominio de factorización única.*

Nota.

En $\mathbb{Z}(i\sqrt{5})$, $9 = 3 \cdot 3 = (2+\sqrt{-5})(2-\sqrt{-5})$, donde $3, (2+\sqrt{-5})$ y $(2-\sqrt{-5})$ son irreducibles, esto es $\mathbb{Z}(i\sqrt{5})$ no es un dominio de factorización única.

2.5 Problemas del capítulo

1. (a) Sea A un anillo conmutativo con al menos uno, pero tan solo, un número finito de divisores de cero. Demostrar que A es finito.

   (b) Explorar que ocurre con la afirmación anterior si el anillo A es no conmutativo y el número de divisores de cero a derecha (o a izquierda) es finito.

2. Los enteros 2, -3 y 5 tienen la propiedad de que la diferencia de cualesquiera dos de ellos es un entero múltiplo del otro. $(2-(-3) = 1 \times 2, (-3) - 5 = (-4_{\times} 2, 5 - 2 = (-1) \times (-3).)$ Sean a, b, c tres enteros que tienen esta propiedad; (a) Demostrar que a, b, c no pueden ser todos positivos. (b) Suponer ahora que a, b, c no tienen factores comunes (excepto 1 y -1). ¿Es verdad que uno de los enteros tiene que ser 1, 2, -1 o -2.

3. Sea $(A, +, \cdot)$ un anillo en el cual siempre que la ecuación $ax = b$ con $a, b \in A$ y $a \neq 0$ tenga solución, esta es única. Demostrar que A es un dominio de integridad.

4. Sea $p \in \mathbb{Z}$, p primo. Demostrar que $ab^p - ba^p$ es divisible por $6p$.

5. Sea A el anillo con identidad de todas las matrices diagonales de la forma

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$$

   con $a, b \in \mathbb{Z}$ y B el subanillo de las matrices de la forma

$$\begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}$$

   Demostrar que en el subanillo B existen dos elementos que actuan como identidad. ¿En un anillo sin divisores de cero es posible encontrar un subanillo con esta característica?.

6. Sean m, n enteros positivos. ¿Para qué valores de m y n, $m^4 + 4n^4$ es un número primo?.

7. Sea $(A, +, \cdot)$ un anillo tal que las ecuaciones $ax = b$ y $ya = b$ son solubles siempre que $a \neq 0$. Demostrar que A es un campo.

8. (a) Si A es un anillo con elemento identidad de característica $n > 1$, demostrar que A contiene un subanillo isomorfo a $\mathbb{Z}_n$. Si A tiene característica 0, entonces A contiene un subanillo isomorfo a $\mathbb{Z}$.

   (b) Todo campo contiene un subcampo isomorfo a $\mathbb{Z}_p$ para algún primo p o un subcampo isomorfo a $\mathbb{Q}$.

9. (a) Sea $\phi : (\mathbb{Z}, +) \longrightarrow (2\mathbb{Z}, +)$ definida por $\phi(x) = 2x$. Demostrar ϕ es un isomorfismo. ¿Es ϕ un homomorfismo del anillo $(\mathbb{Z}, +, \cdot)$ en el anillo $(2\mathbb{Z}, +, \cdot)$?.

   (b) Demostrar que los anillos $2\mathbb{Z}$ y $3\mathbb{Z}$ no son isomorfos.

10. Demostrar que 1 y $p-1$ son los únicos elementos del campo $\mathbb{Z}_p$ que son sus propios inversos multiplicativos.

Capítulo 8

Anillos de polinomios

8.1 Una teoría de polinomios

Como lo comentamos en el capítulo 7 algunos anillos, muy familiares para nosotros desde la básica, se constituyen en modelos para la construcción de una teoría general; éste es el caso del anillo de polinomios. En la básica secundaria tienen lugar las primeras experiencias de trabajo con los polinomios; se estudia su aritmética, se aprende a sumar, restar, multiplicar y dividir polinomios (usando un algoritmo similar al algoritmo de Euclides para números naturales). Se analiza, como lo hemos hecho en secciones anteriores, la solución de ecuaciones polinómicas y se caracterizan sus raíces. Esto conlleva desde luego un examen de la factorización e involucra un análisis de las funciones polinómicas y sus gráficas. Todo lo anterior aunque no se perciba inicialmente, forma parte de una teoría mas general, la llamada teoría de polinomios; el estudio de la estructura de los polinomios es fundamental en la teoría de anillos, y se constituye en una interesante fuente de ejemplos y contraejemplos. En lo que sigue desarrollaremos una tal teoría.

Definición 8.1 *Sea A un anillo. Las expresiones de la forma*

$$a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0 = \sum_{i=0}^{n} a_i x^i$$

donde $n \geq 0$ es un entero y $a_i \in A$ para $i = 1, 2, ..., n$ forman un anillo $A[x]$ llamado el anillo de polinomios sobre A con las operaciones.

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

y

$$\sum_{i=0}^{n} a_i x^i \cdot \sum_{j=0}^{m} b_j x^j = \sum_{k=0}^{m+n} c_k x^k, \text{ donde } c_k = \sum_{i+j=k} a_i b_j.$$

Nótese que para sumar dos polinomios de grados distintos, $m > n$,

$$\sum_{i=0}^{n} a_i x^i + \sum_{j=0}^{m} b_j x^j$$

es conveniente pensar en el primero como $a_m x^m + \ldots + a_{n+1} x^{n+1} + a_n x^n \ldots + a_1 x + a_0$ donde $a_{n+1} = \cdots = a_m = 0$. Por este motivo, al escribir la suma de dos polinomios sobre A, siempre podemos suponer, sin pérdida de generalidad, que son del mismo grado.

Cada elemento de $A[x]$ se llama un *polinomio* y los elementos $a_i \in A$ se llaman los *coeficientes* del polinomio. Dos polinomios en $A[x]$ son iguales si y sólo si sus coeficientes respectivos son iguales en A. El grado $\deg P$ de un polinomio no nulo P es el mayor i para el cual $a_i \neq 0$. Si el anillo A tiene unidad 1 y si $a_n = 1$, se dice que el polinomio es mónico. Los polinomios de grado 0 corresponden a los elementos de A y se llaman polinomios constantes, mientras que los polinomios de grado 1 son lineales, de grado 2 cuadráticos, de grado 3 cúbicos, de grado 4 cuárticos y de grado 5 quínticos. No se asigna un grado al polinomio 0.

Teorema 8.1 *Sea A un anillo. Entonces $A[x]$ es un anillo. Si A es un anillo con elemento identidad (módulo), entonces $A[x]$ es un anillo con identidad. Si A es anillo conmutativo, entonces $A[x]$ es conmutativo.*

Punto de discusión

Demostrar el teorema anterior.

Una vez definida la multiplicación entre polinomios, que es desde luego una operación binaria sobre el conjunto de los polinomios, surge naturalmente una pregunta (familiar ya en los números enteros). ¿Es siempre posible efectuar un proceso inverso?, es decir, dado un polinomio expresarlo como producto de dos o mas polinomios (factorizarlo). Este no es en general un problema sencillo (dependerá desde luego de la estructura del anillo donde estamos formulando tal cuestión); nosotros hemos visto ya algunos procedimientos viables pero no aplicables en todo caso. A partir de la experiencia con la multiplicación se llegan a identificar algunos casos especiales que, con alguna frecuencia son llamados productos notables. Retómelos usted para resolver los siguientes ejercicios.

Ejercicios

1. Si $a \neq b$, $a^3 - b^3 = 19x^3$ y $a - b = x$, expresar a en términos de x.
2. Si se factoriza completamente el polinomio $x^9 - x$ como producto de polinomios con coeficientes enteros, ¿cuántos factores tendrá?
3. Si $(r + \frac{1}{r})^3 = 3$, ¿a qué es igual $r^3 + \frac{1}{r^3}$?
4. Para factorizar el polinomio $x^4 + 4$ en los enteros basta aplicar la factorización de una diferencia de cuadrados es decir: $t^2 - a^2 = (t - a)(t + a)$. Efectuar tal factorización.
5. Demostrar que la diferencia de los cuadrados de dos números impares consecutivos es siempre divible por 8.
6. Escribir los siguientes polinomios como diferencia de cuadrados y luego factorizar sobre $\mathbb{Z}$.
   (a) $4x^2 - 20x - 11$
   (b) $5x^2 - 6x + 1$
   (c) $x^4 - 47x^2 + 1$

En cualquier anillo de polinomios sobre un anillo A, es inmediato observar que las potencias de x (llamada una *indeterminada*) en la expresión polinómica marcan la posición de cada coeficiente de manera similar a cómo, en la aritmética decimal, una potencia de diez determina la posición de cada dígito. La analogía anterior puede extenderse y, de hecho, fue extendida por Isaac Newton, hecho que originó cambios en la manera de pensar de los matemáticos frente al álgebra.

A pesar de que la indeterminada x no tome valores en el anillo A (ni en ningún otro conjunto), es posible considerar una suma s de la forma

$$s = \sum_{i=0}^{n} a_i r^i \tag{1}$$

siendo r es un elemento fijo de A. Por las propiedades del anillo, es claro que s pertenece a A. Resulta de interés en nuestra teoría asociar expresiones polinómicas sobre A con expresiones como (1), es decir, establecer una correspondencia

$$\sum_{i=0}^{n} a_i x^i \longrightarrow \sum_{i=0}^{n} a_i r^i$$

y esperaríamos que una tal correspondencia resultara ser un homomorfismo de anillos, esto es, conservara la estructura. Sin embargo, esto no necesariamente es cierto si A no es un anillo conmutativo.

Punto de discusión

Sean A un anillo con elemento identidad 1 y en $A[x]$ los polinomios $(0+1x)$ y $(t+0x)$. Considerar su producto y determinar la imagen de éste bajo la correspondencia definida anteriormente, así como la imagen de cada uno de los polinomios y el producto de estas imágenes. ¿Qué se observa? ¿Son necesariamente iguales? ¿Constituye la correspondencia un homomorfismo entre los grupos $(A,+)$ y $(A[x],+)$?

La discusión anterior nos sugiere que requerimos anillos de coeficientes con estructuras cada vez mas completas. Considerando polinomios con coeficientes en un anillo conmutativo con elemento identidad, o en una estructura mas completa aun, un campo por ejemplo, la correspondencia mencionada es claramente un homomorfismo de anillos. Este hecho, así como el trabajo familiar con teoría de polinomios donde los coeficientes son tomados en un dominio de integridad, $\mathbb{Z}$, por ejemplo o en los campos, $\mathbb{Q}$, $\mathbb{R}$ o $\mathbb{C}$ nos motiva a estudiar mas a fondo anillos de polinomios especiales. En lo que sigue consideraremos el anillo $F[x]$ de polinomios sobre un campo F.

Puntos de discusión

1. Si $p(x)$ y $g(x)$ son dos elementos distintos de cero de $F[x]$. Demostrar que $leg(f(x)g(x)) = deg(f(x)) \cdot deg(g(x))$.

2. Investiguar qué ocurre con la afirmación anterior cuando se considera los coeficientes de los polinomios en un dominio de integridad y qué si los considera en un anillo arbitrario. Explorar ejemplos.

3. Si F es un campo demostrar que $F[x]$ es un dominio de integridad. (¿Es válido también cuando el anillo de coeficientes es un dominio de integridad?)

4. Si consideramos la función $deg(f(x))$ para todo $f(x) \neq 0$ en $F[x]$, ¿disponemos ya de todos los elementos para afirmar que $F[x]$ es anillo euclidiano?

5. ¿Será posible encontrar inversos multiplicativos para los polinomios en $F[x]$? En particular, ¿existe un inverso en $F[x]$ del polinomio $p(x) = x$?

Teorema 8.2 *Sean F un campo y $F[x]$ el anillo de polinomios sobre F. Entonces*

$$\phi : \sum_{i=0}^{n} a_i x^i \longrightarrow \sum_{i=0}^{n} a_i t^i,$$

donde t es un elemento fijo del campo F, es un homomorfismo de anillos de $F[x]$ en F.

Demostración
Sean $p(x) = \sum_{i=0}^{n} a_i x$, $q(x) = \sum_{i=0}^{n} b_i x^i$. Veamos qué sucede con $\phi(p+q)$ y $\phi(p \cdot q)$. Tenemos $\phi(p+q) = \phi\left(\sum_{i=0}^{n} a_i x^i + \sum_{i=0}^{n} b_i x^i\right) = \phi\left(\sum_{i=0}^{n}(a_i + b_i)x^i\right) = \sum_{i=0}^{n}(a_i+b_i)t^i = \sum_{i=0}^{n} a_i t^i + \sum_{i=0}^{n} b_i t^i = \phi(p)+\phi(q)$. Aplicando reiteradamente la asociatividad, conmutatividad y distributividad de la suma y multiplicación en el campo F concluimos que ϕ es hasta el momento un homomorfismo de grupos entre $(F[x], +)$ y $(F, +)$.

Ejercicios

Para concluir que ϕ es un homomorfismo de anillos, demostrar que $\phi(p \cdot q) = \phi(p) \cdot \phi(q)$.

Definición 8.2 *Sea F un campo, $a_0, a_1, ..., a_n$ elementos de F. Una función polinómica $f(x)$ es una aplicación de F en F definida por $f(t) = \sum_{i=0}^{n} a_i t^i$. Cualquier elemento de F con imagen 0 bajo una tal aplicación se llama un cero de $f(x)$.*

8.2 Factorización de polinomios sobre un campo

Teorema 8.3 *(Algoritmo de la división para $F[x]$.) Dados dos polinomios $f(x)$ y $g(x)$ en $F[x]$, $g(x) \neq 0$. Entonces existen un polinomio cociente $q(x)$ y un polinomio residuo $r(x)$ tales que $f(x) = q(x)g(x) + r(x)$ donde $r(x)$ es el polinomio 0 o $\deg r(x) < \deg g(x)$.*

Demostración
En primer lugar si el grado de $f(x)$ es menor que el grado de $g(x)$, nos bastará tomar $q(x) = 0$ y $r(x) = f(x)$ y tenemos que $f(x) = 0 \cdot g(x) + f(x)$ donde $deg(f(x)) < deg(g(x))$.

Suponemos ahora que $f(x) = a_0 + a_1 x + \cdots + a_m x^m$ y $g(x) = b_0 + b_1 x + \cdots + b_n x^n$ con $a_m \neq 0, b_n \neq 0$ y $m \leq n$. Consideramos el polinomio

$$f_1(x) = f(x) - \left(\frac{a_m}{b_n} x^{m-n} g(x)\right)$$

Dicho polinomio tiene grado menor o igual que $m-1$, esto es menor que el grado de $f(x)$. Podemos argumentar entonces inductivamente sobre el grado de $f(x)$ y suponer válido el teorema para polinomios de grado menor que m. En particular es válido para $f_1(x)$. Existen entonces $q_1(x)$ y $r(x)$ en $F[x]$, con $r(x)=0$ o $deg(r(x)) < deg(g(x))$, tal que $f_1(x) = q_1(x)g(x) + r(x)$. Pero entonces tenemos que

$$f(x) - \left(\frac{a_m}{b_n}x^{m-n}g(x)\right) = q_1(x)g(x) + r(x).$$

Transponiendo términos se tiene que

$$f(x) = \left[(\frac{a_m}{b_n})x^{m-n} + q_1(x)\right]g(x) + r(x).$$

Si hacemos $q(x) = (\frac{a_m}{b_n})x^{m-n} + q_1(x)$, entonces $f(x) = q(x)g(x) + r(x)$ con $q(x)$ y $r(x)$ en $F[x]$, $r(x) = 0$ o $deg(r(x)) < deg(g(x))$.

Nota

1. Nótese que la demostración es en realidad el proceso de división larga que todos usamos para dividir un polinomio entre otro.

2. Obsérvese que la división procede hasta que se presente un residuo de grado menor que el grado del polinomio divisor. Demostremos ahora que el residuo y el cociente de la división son únicos. Para ello supongamos que $f(x) = q_1(x)g(x) + r_1(x)$ y $f(x) = q_2(x)g(x) + r_2(x)$. Se sigue que $q_1(x)g(x) + r_1(x) = q_2(x)g(x) + r_2(x)$, de donde, se concluye que $g(x)$ divide a $r_2(x) - r_1(x)$, pero $\deg r_1(x) < \deg g(x)$ o $r_1(x) = 0$ y $\deg r_2(x) < \deg g(x)$ o $r_2(x) = 0$. La única posibilidad es que $r_2(x) - r_1(x) = 0$. Se tiene entonces que el residuo (y por ende el cociente) es único.

Teorema 8.4 *(Teorema del factor). Sea $p(x)$ un polinomio en $F[x]$, donde F es un campo. Entonces, $x-d$, con $d \in F$, es un factor de $p(x)$ si y sólo si d es un cero de la función polinómica $f(x)$.*

Demostración
Apliquemos el teorema anterior tomando $g(x) = x - d$. Entonces el residuo $r(x)$ es tal que $\deg r(x) < \deg g(x) = 1$. Se sigue que $r(x) = 0$ o $r(x) = r$, donde r un elemento del campo F y

$$p(x) = q(x)(x-d) + r.$$

Ahora, el homomorfismo del Teorema 8.2 asegura que

$$p(d) = q(d)(d-d) + r.$$

De allí se sigue que, si d es un cero de $f(x)$, es decir, si $p(d) = 0$, entonces $r = 0$. Inversamente, si $x - d$ es un factor de $p(x)$, de modo que $r = 0$, entonces

$$p(d) = q(d)(d-d) + 0 = q(d) \cdot 0 + 0 = 0,$$

y esto implica que d es un cero de $p(x)$.

Definición 8.3 *Una ecuación polinómica es una ecuación de la forma* $f(x) = 0$, *donde* $f(x)$ *es una función polinómica. Las raíces de la ecuación son los ceros de función* $f(x)$, *es decir, los elementos de* F *cuya imagen bajo* $f(x)$ *es 0.*

Teorema 8.5 *(Teorema del residuo). Sea* $f(x)$ *una función polinómica con coeficientes de un campo* F. *Entonces el valor de la función* $f(d)$, *con* $d \in F$, *es igual al residuo cuando se divide* $f(x)$, *por* $x - d$.

Demostración
A partir del algoritmo de la división, se puede escribir

$$f(x) = q(x)(x - d) + r(x),$$

donde $r = 0$ o $\deg r(x) < 1$. Se sigue que $r(x)$, es una constante, digamos r, que pertenece al campo F. Aplicando el homomorfismo, tenemos

$$f(d) = q(d)(d - d) + r = r,$$

como queríamos.

Puntos de discusión

1. Sean D un dominio de integridad, $c \in D$, y $p(x)$ un polinomio sobre D. Demostrar que $x - c$ divide a $p(x)$ si y sólo si $p(c) = 0$.

2. El algoritmo de la división se formuló para polinomios con coeficientes en un campo. Ilustrar por qué dicho algoritmo no es siempre válido para polinomios sobre un dominio de integridad. ¡Analizar polinomios en $\mathbb{Z}[x]$! Formular y demostrar una versión modificada del algoritmo en dominios de integridad.

3. Para determinar el resto de dividir un polinomio $f(x) \in F[x]$ por $x - c$ basta determinar $f(c)$. Investigar expresiones para el resto cuando el polinomio $f(x)$ se divide por $(x - a)(x - b)$, $(x - c)^n$, $(x - a_1)(x - a_2)(x - a_3) \cdots (x - a_m)$ y $(x - a)^r(x - b)^s$.

Definición 8.4 *Un polinomio no constante* $f(x) \in F[x]$ *es irreducible sobre* F *o es un polinomio irreducible en* $F[x]$, *si* $f(x)$ *no puede ser expresado como un producto* $g(x)h(x)$ *de dos polinomios* $g(x), h(x) \in F[x]$ *ambos de grado menor que el grado de* $f(x)$.

Notas

1. Nótese que la irreducibilidad de un polinomio depende del campo sobre el que lo estemos considerando. Un polinomio $f(x)$ puede ser irreducible sobre un campo F, pero puede no ser irreducible sobre un campo E que contiene a F. Un ejemplo muy familiar para nosotros es el polinomio $x^2 - 2$, que es irreducible en $\mathbb{Q}$, pero si lo consideramos en $\mathbb{R}[x]$, allí se factoriza en $(x - \sqrt{2})(x + \sqrt{2})$.

2. En campos finitos con un número reducido de elementos, identificar los ceros de polinomios cuadráticos o cúbicos puede resultar efectivo para determinar si son o no irreducibles. Si consideramos por ejemplo

el polinomio $f(x) = x^3+3x+2$ en $\mathbb{Z}_5[x]$ y suponemos que se factoriza como producto de polinomios de grado menor, debe existir al menos un factor lineal $x-a$ de $f(x)$, para algún $a \in \mathbb{Z}_5$. Pero entonces $f(a) = 0$. Sinembargo $f(0) = 2$, $f(1) = 1$, $f(2) = 1$, $f(3) = 3$ y $f(4) = 3$, lo cuál prueba que $f(x)$ no tiene ceros en $\mathbb{Z}_5$; es pues irreducible sobre $\mathbb{Z}_5$ En las notas anteriores hemos trabajado con polinomios de grados dos y tres para los que resulta sencillo decidir sobre su irreducibilidad, siempre que podamos caracterizar sus ceros; esta idea se generaliza en el siguiente teorema

Teorema 8.6 *Sea $f(x) \in F[x]$, y sea $f(x)$ de grado dos o tres. Entonces $f(x)$ es reducible sobre F si y sólo si tiene un cero en F.*

Demostración
Si $f(x)$ es reducible, $f(x) = g(x)h(x)$, donde el grado de $g(x)$ y el grado de $h(x)$ son ambos menores que el grado de $f(x)$, entonces dado que $f(x)$ es cuadrático o cúbico, $g(x)$ o $h(x)$, o ambos tienen grado uno. Si $g(x)$ tiene grado 1, entonces salvo por un posible factor en F, $g(x)$ es de la forma $x-a$, de donde $g(a) = 0$. Esto implica que $f(a) = 0$ y por tanto $f(x)$ tiene un cero en F. Recíprocamente por teorema del factor, si $f(a) = 0$ para algún $a \in F$, $x - a$ es un factor de $f(x)$, es decir $f(x)$ es reducible.

8.2.1 Polinomios sobre el campo racional -La factorización de polinomios en Z[x].

Puntos de Discusión. Para iniciar recordemos algunas ideas familiares acerca de los polinomios con coeficientes enteros.

1. Sea $c \in \mathbb{Z}$.
   (a) Si $c > 0$ el polinomio $x^2 + c$ es irreducible (no puede ser factorizado como el producto de dos o mas polinomios de grado positivo) sobre $\mathbb{Z}$, $\mathbb{Q}$ y $\mathbb{R}$, pero es reducible sobre $\mathbb{C}$.
   (b) Si $c = -d^2$ para algún $d \in \mathbb{Z}$, demostrar que $x^2 + c$ es reducible sobre $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ y $\mathbb{C}$.
   (c) Si $c < 0$, pero no es el opuesto de un cuadrado perfecto, $x^2 + c$ es irreducible sobre $\mathbb{Z}$ y $\mathbb{Q}$ pero reducible sobre $\mathbb{R}$ y $\mathbb{C}$.
2. Discutir la irreducibilidad del polinomio $ax^2 + bx + c$, sobre $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ y $\mathbb{C}$, cuando $a, b, c \in \mathbb{Z}$.

Definición 8.5 *Se dice que el polinomio $p(x) = a_0 + a_1x + \cdots + a_nx^n$ donde $a_0, a_1, \cdots, a_n$ son enteros, es primitivo si el máximo común divisor de $a_0, a_1, \cdots, a_n$ es 1.*

Punto de discusión
Si $f(x)$ y $g(x)$ son polinomios primitivos, demostrar que $f(x)g(x)$ es un polinomio primitivo.

Definición 8.6 *El contenido del polinomio $p(x) = a_0 + a_1 + \cdots a_n$, donde todos los a_i son enteros es el máximo común divisor de los enteros $a_0, a_1, \cdots, a_n$.*

Nota. Nótese que todo polinomio $p(x)$ con coeficientes enteros puede escribirse como $p(x) = mq(x)$ donde m es el contenido de $p(x)$ y $q(x)$ es un polinomio primitivo.

Teorema 8.7 *(Lema de Gauss).*

Si el polinomio primitivo $p(x)$ puede factorizarse como el producto de dos polinomios con coeficientes racionales, puede factorizarse como el producto de dos polinomios con coeficientes enteros.

Demostración
Suponemos que $p(x) = r(x)s(x)$ donde $r(x)$ y $s(x)$ tienen coeficientes racionales. Reduciendo a común denominador y sacando factores comunes podemos escribir $p(x) = (\frac{a}{b})u(x)v(x)$, donde a, b son enteros, y $u(x)$ y $v(x)$ tienen coeficientes enteros y son primitivos. Entonces $bp(x) = au(x)v(x)$. El contenido del primer miembro es b pues $p(x)$ es primitivo, y el del segundo es a, pues el producto de dos primitivos $u(x)$ y $v(x)$ es primitivo, como el lector demostró en el anterior punto de discusión. Por lo tanto $a = b$, $\frac{a}{b} = 1$ y $p(x) = u(x)v(x)$, donde $u(x)$ y $v(x)$ tienen coeficientes enteros.

Definición 8.7 *Un polinomio $p(x) = a_0 + a_1x + \cdots + a_nx^n$ se dice entero mónico si todos sus coeficientes son enteros y $a_n = 1$.*

Puntos de discusión

1. Demostrar que si un polinomio entero mónico se factoriza como el producto de dos polinomios de grado positivo y coeficientes racionales, se factoriza como el producto de dos polinomios enteros mónicos.

2. Demostrar el siguiente teorema. Si $f(x) \in \mathbb{Z}[x]$, entonces $f(x)$ se factoriza en un producto de dos polinomios de grado positivo en $\mathbb{Q}[x]$ si y sólo si se factoriza como el producto de dos polinomios de grado positivo en $\mathbb{Z}[x]$.

Un corolario al teorema que usted demostró en el punto de discusión 2 anterior es el siguiente.

Corolario 8.1 *Si $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ es un polinomio entero mónico con $a_0 \neq 0$ y si $f(x)$ tiene un cero $m \in \mathbb{Z}$, $m|a_0$.*

Demostración Si $f(x)$ tiene un cero a en $\mathbb{Q}$, entonces $f(x)$ tiene un factor lineal $x - a$ en $\mathbb{Q}[x]$ pero por teorema anterior $f(x)$ tiene una factorización en factores lineales en $\mathbb{Z}[x]$; entonces para algún $m \in \mathbb{Z}$ se tiene

$$f(x) = (x - m)\left(x^{n-1} + \cdots + \frac{a_0}{m}\right),$$

de donde $\frac{a_0}{m} \in \mathbb{Z}$, es decir $m|a_0$.

Notas

1. Una demostración de la irreducibilidad de $x^2 - 2$ sobre $\mathbb{Q}$ podría ser la siguiente: $x^2 - 2$ se factoriza no trivialmente en $\mathbb{Q}[x]$ si y solamente si tiene un cero en $\mathbb{Q}$. Por corolario anterior tiene un cero en $\mathbb{Q}$ si y sólo si tiene un cero en $\mathbb{Z}$, y además los únicos ceros posibles son los divisores de 2, ± 1 y ± 2, y ninguno de estos números es cero de este polinomio.

Demostrar que los polinomios $x^3+26x^2-52x+13$ y x^5-4x^4+2x+2 son irreducibles en $\mathbb{Q}[x]$.

Puntos de discusión

1. Aplicando el criterio de Eisenstein, ¿puede usted determinar si todos los polinomios que se listan a acontinuación son irreducibles sobre Q?

   $x^2-12, 8x^3+6x^2-9x+24, 4x^{10}-9x^3+24x-18, 2x^{10}-25x^3+10x^2-30.$

2. ¿Es el polinomio x^n-p irreducible sobre los racionales para cualquier primo p?

3. Demostrar que el polinomio $1+x+\cdots+x^{p-1}$ es irreducible sobre los racionales para cualquier primo p.

4. Sean m y n enteros primos relativos, si

   $$\left(x-\frac{m}{n}\right)|(a_0+a_1x+a_sx^s)$$

   donde los a_i son enteros, demostrar que $m|a_0$ y $n|a_s$.

Puntos de investigación

1. Escribir una lista completa de los polinomios de grados 0, 1 2, 3 y 4 en el anillo $Z_2[x]$.

2. Encontrar todos los polinomios irreducibles de grados dos y tres en $Z_2[x]$ y en $Z_3[x]$.

3. ¿El polinomio x^4+4 puede ser factorizado en factores lineales en $Z_5[x]$?.

4. ¿Es el polinomio x^3+2x+3 un polinomio irreducible de $Z_5[x]$? Si no lo es determinar su factorización.

5. Demostrar que el polinomio x^7-x toma el valor 0 para todo $x \in Z_7$.

6. Sea p un primo. ¿Cúantos polinomios diferentes de grado n hay sobre Z_p (para un n fijo)?.

8.3 Caracterización de ideales de $F[x]$

En uno de los puntos de discusión del capítulo de anillos se afirmaba que el ideal de $F[x]$ formado por todos los polinomios que tienen término constante nulo, es un Ideal Principal, pues es precisamente el ideal $< x >$. (Los ideales Principales de un anillo $(A,+,\cdot)$ conmutativo con identidad, son de la forma $\{ra|r \in A\}$). La pregunta es entonces, ¿qué otros ideales del anillo $F[x]$ (anillo conmutativo con identidad) resultan ser también principales? esta pregunta es interesante en Teoría General de Anillos, pues como lo comentamos anteriomente, nos va a permitir construir nuevas estructuras, pero, es aun mas interesante, en anillos de polinomios con coeficientes en un campo pues nos va a permitir dar una solución formal al problema fundamental abordado en este texto: la existencia de raíces y la factorización. El siguiente teorema nos proporciona elementos importantes en este sentido.

2. Demostremos ahora que el polinomio $f(x) = x^4 - 2x^2 + 8x + 1$ es irreducible sobre $\mathbb{Q}$. Si suponemos que $f(x)$ tiene un factor lineal en $\mathbb{Q}[x]$, entonces tiene un cero en $\mathbb{Z}$, y por corolario anterior este cero debe ser un divisor en $\mathbb{Z}$ de 1, esto es ± 1. Pero $f(1) = 8$ y $f(-1) = -8$, por tanto tal factorización es imposible. Supongamos que $f(x)$ se factoriza como producto de dos factores cuadráticos en $\mathbb{Q}[x]$. Tiene entonces una factorización

$$(x^2 + ax + b)(x^2 + cx + d)$$

en $\mathbb{Z}[x]$. Si igualamos coeficientes de las potencias de x tenemos $bd = 1$, $ad + bc = 8$, $ac + b + d = -2$ y $a + c = 0$ para enteros a, b, c, d. Analizando el sistema de ecuaciones concluimos que no existen enteros que sean solución de éste, de donde, se deduce que la factorización en dos factores cuadráticos es también imposible. Se concluye entonces que $f(x)$ es irreducible sobre $\mathbb{Q}$.

El problema de determinar si un polinomio puede (o no) factorizarse sobre un campo determinado resulta en general dispendioso, inclusive en el caso de polinomios sobre un campo tan familiar como el de los racionales; esto se debe a que existen pocos criterios que permitan decidir sobre tal factorización. Uno de tales criterios se presenta a continuación

Teorema 8.8 *(Criterio de Eisenstein). Sea $q(x) = a_0 + a_1 x + \cdots + a_n x^n$ un polinomio con coeficientes enteros. Si existe un número primo p tal que $p|a_0$, $p|a_1$, ... $p|a_{n-1}$, p no divide a a_n y p^2 no divide a a_0; $q(x)$ es irreducible sobre los racionales.*

Demostración
Puesto que al sacar el máximo común divisor de los coeficientes de $p(x)$, no se modifica la hipótesis de que p no divide a a_n, podemos suponer sin perdida de generalidad que $q(x)$ es primitivo. Supongamos que $q(x)$ se factoriza como producto de dos polinomios con coeficientes racionales. Del lema de Gauss, $q(x)$ se puede factorizar como producto de polinomios con coeficientes enteros, esto es

$$q(x) = (b_0 + b_1 x + \cdots + b_s x^s)(c_0 + c_1 x + \cdots + c_t x^t)$$

donde b_i y c_j son enteros, $s, t > 0$. Comparando coeficientes de ambos miembros tenemos que $a_0 = b_0 c_0$; como $p|a_0$, se sigue que $p|b_0$ o $p|c_0$. Como p^2 no divide a a_0, p no los puede dividir a ambos. Supongamos que $p|a_0$ y que p no divide a c_0, pero no todos los b_i son divisibles por p, pues en ese caso todos los coeficientes de $q(x)$ serían divisibles por p y esto contradice la hipótesis de que p no divide a a_n. Sea entonces b_l el primero de los b no divisible por p con $l \leq s < n$. $a_l = b_l c_0 + b_{l-1} c_1 + \cdots + b_0 c_l$, lo que nos lleva a concluir por suposiciones anteriores que $p|b_l c_0$, pero esto es imposible pues p no divide a c_0 y p no divide a c_l. Esta contradicción demuestra que el polinomio $q(x)$ no puede ser factorizado en los racionales, es decir es irreducible.

Por ejemplo si tomamos $p = 3$, y aplicamos el criterio de Eisenstein al polinomio $25x^5 - 9x^4 + 3x^2 - 12$ podemos afirmar que es irreducible sobre $\mathbb{Q}$.

Ejercicio

Teorema 8.9 *Si F es un campo, todo ideal en $F[x]$ es principal.*

Demostración
Sea I un ideal de $F[x]$. Si $I = 0$, entonces $I =< 0 >$. Supongamos entonces que $I \neq 0$ y que $g(x)$ es un elemento no nulo de I de grado mínimo. Si el grado de $g(x)$ es cero, entonces $g(x) \in F$ y es una unidad (elemento inversible del anillo $F[x]$), el ideal I contiene una unidad y esto implica que $I =< 1 >= F[x]$, I es entonces ideal principal. Si el grado de $g(x) \geq 1$, sea $f(x)$ cualquier elemento de I. Por algoritmo de la división, existen $q(x)$ y $r(x)$ en $F[x]$, tales que $f(x) = g(x)q(x)+r(x)$, donde $deg(r(x)) < deg\, g(x)$. Como $f(x)$ y $g(x)$ pertenecen a I, que es un ideal $f(x)-g(x)q(x) = r(x) \in I$. Como $g(x)$ es un elemento no nulo de I de grado mínimal, se tiene que $r(x) = 0$. Por tanto $f(x) = g(x)q(x)$, y $I =< g(x) >$.

Teorema 8.10 *Un ideal $< p(x) > \neq 0$ de $F[x]$ es maximal si y sólo si $p(x)$ es irreducible sobre F.*

Demostración
Supongamos que $< p(x) > \neq 0$ es un ideal maximal de $F[x]$. Entonces $< p(x) > \neq F[x]$, dado que $p(x) \notin F$. Supongamos que $p(x)$ se factoriza no trivialmente en $F[x]$, $p(x) = f(x)g(x)$. Dado que $< p(x) >$ es maximal y por tanto también ideal primo, $(f(x)g(x)) \in < p(x) >$ implica que $f(x) \in < p(x) >$, o, $g(x) \in < p(x) >$. Es decir, $f(x)$ o $g(x)$ tienen a $p(x)$ como un factor. Pero esto es imposible pues tanto el grado de $f(x)$ como el grado de $g(x)$ son menores que el grado de $p(x)$. Esto demuestra que $p(x)$ es irreducible sobre F. Recíprocamente si $p(x)$ es irreducible sobre F, supongamos que I es un ideal tal que $< p(x) > \subseteq I \subseteq F[x]$. Como I es un ideal principal $I =< g(x) >$, para algún $g(x) \in I$. Entonces $p(x) \in I$ implica que $p(x) = g(x)q(x)$, para algún $q(x) \in F[x]$. Pero $p(x)$ es irreducible, lo cual implica que $g(x)$ o $q(x)$ tienen grado cero. Si $g(x)$ tiene grado cero, es una constante no nula en F, entonces $g(x)$ es una unidad en $F[x]$, y por tanto $< g(x) >= I = F[x]$. Si $q(x)$ es de grado cero entonces $q(x) = c$, donde $c \in F$ y $g(x) = \frac{1}{c}p(x)$, de donde, $I =< p(x) >$ y por consiguiente I es un ideal maximal.

Nota - Recordemos si A es un anillo conmutativo con elemento identidad, I es ideal máximo de A si y sólo si A/I es un campo. Usando el teorema anterior podemos afirmar que $p(x)$ es irreducible sobre F si y solo si $F[x]/ < p(x) >$ es un campo.

Mostramos en una nota anterior, que el polinomio x^3+3x+2 es irreducible en $Z_5[x]$. $\frac{Z_5[x]}{<x^3+3x+2>}$ es entonces un campo. Similarmente $\frac{Q[x]}{<x^2-2>}$ es un campo. Sinembargo esta nota no nos permite visualizar dichos campos, ni ilustra acerca de su importancia, por ello analizaremos unos ejemplos y enunciaremos un teorema que motiva nuestro trabajo con extensiones algebraicas en el siguiente capítulo.

Definición 8.8 *Sea F un campo, y $p(x)$ un polinomio con coeficientes en F irreducible en $F[x]$,*

$$\frac{F[x]}{< p(x) >} = \{r(x) + p(x)q(x) | q(x) \in F[x]\},$$

clases de equivalencia de los polinomios de $F[x]$, módulo el ideal $< p(x) >$.

Ejemplo. Sea $F = \mathbb{Q}$ el campo de los números racionales. Consideremos el polinomio irreducible sobre $\mathbb{Q}$, $p(x) = x^2 - 2$ e $I =< x^2 - 2 >$. Cualquier

elemento en $\frac{\mathbb{Q}[x]}{<x^2-2>}$ es una clase lateral de la forma $f(x)+I$ con $f(x) \in \mathbb{Q}[x]$. Pero, dado un polinomio cualquiera $f(x) \in \mathbb{Q}[x]$, por algoritmo de la división, $f(x) = q(x)(x^2-2)+r(x)$, donde $r(x) = 0$ o $deg(r(x)) < deg(x^2-2) = 2$. Luego $r(x) = a_0 + a_1 x$ con $a_0, a_1 \in \mathbb{Q}$. Por consiguiente $f(x)+I = a_0 + a_1 x + q(x)(x^2-2) + < x^2-2 > = a_0 + a_1 x + < x^2-2 >$, pues $q(x)(x^2-2) \in < x^2-2 >$. Por adición y multiplicación de clases

$$f(x)+I = (a_0 + < x^2-2 >) + a_1(x + < x^2-2 >) = (a_0+I) + a_1(x+I).$$

Si notamos $\overline{x} = x+I$, todo elemento de $\frac{\mathbb{Q}[x]}{<x^2-2>}$, es de la forma $a_0 + a_1\overline{x}$, con $a_0, a_1 \in \mathbb{Q}$. Además $(\overline{x})^2 - 2 = (x+I)^2 - 2 = (x^2-2)+I = I = 0$ (pues I es el neutro del campo $\frac{\mathbb{Q}[x]}{I}$). Concluimos que $(\overline{x})^2 = 2$, $\overline{x}$ es raíz de la ecuación x^2-2 en el campo cociente.

Sea ahora $a_0 + a_1\overline{x}$ un elemento no nulo del campo, a_0 y a_1 no son ambos nulos. ¿Cúal es su inverso?. Debe ser un elemento $b_0 + b_1\overline{x}$ en el campo tal que

$$(a_0 + a_1\overline{x})(b_0 + b_1\overline{x}) = 1.$$

Desarrollando el producto obtenemos

$$a_0b_0 + (a_0b_1 + a_1b_0)\overline{x} + (a_1b_1)(\overline{x})^2 = 1.$$

Usando que $(\overline{x})^2 = 2$ e igualando, se tiene el sistema

$$a_0b_0 + 2a_1b_1 = 1$$

$a_1b_0 + a_0b_1 = 0$ Al resolverlo encontramos que $b_0 = \frac{-a_0}{2a_1{}^2 - a_0{}^2}$ y $b_1 = \frac{a_1}{2a_1{}^2 - a_0{}^2}$.

Puntos de Discusión

1. ¿Es $\frac{\mathbb{Q}[x]}{<x^2-6x+6>}$ un campo?

2. Sea $p(x) = x^2 - \frac{1}{3}x - 1 \in \mathbb{Q}[x]$. Demostrar que x^3+1, $\frac{-x^2}{3} + x + 1$ y $\frac{10}{9}x + \frac{2}{3}$ están en la misma clase de equivalencia según el ideal $I = < p(x) >$. Encontrar un polinomio que esté en la misma clase de $x^4 + x^2 + 1$ según I.

3. Sabemos que en $\mathbb{Q}[x]$ el polinomio x no es inversible. Encontrar el inverso de la clase de equivalencia de este polinomio en $\frac{\mathbb{Q}[x]}{<x^2+1>}$.

4. Sea F un campo y $f(x), g(x) \in F[x]$. Demostrar que $f(x)$ divide a $g(x)$ si y sólo si $g(x) \in < f(x) >$.

5. Sea $F = Z_{11}$. Demostrar que $\frac{F[x]}{<x^2+1>}$ es un campo que tiene 121 elementos.

Punto de investigación

Sea $F = \mathbb{R}$, campo de los números reales. Investigar por qué el campo $\frac{F[x]}{<x^2+1>}$ es isomorfo al campo de los números complejos.

Formalizaremos las ideas discutidas en los puntos anteriores con el siguiente teorema

Teorema 8.11 *(Teorema de Kronecker). Sea F un campo, y sea $p(x)$ un irreducible en $F[x]$. Entonces $\frac{F[x]}{<p(x)>}$ es un campo que contiene un subcampo isomorfo a F y el polinomio $p(x)$ tiene un cero en $\frac{F[x]}{<p(x)>}$.*

Demostración Los demostración de que $\frac{F[x]}{<p(x)>}$ es un campo, se deja al lector. ¡Importante analizar lo relativo a los inversos! Para demostrar que el campo $\frac{F[x]}{<p(x)>}$ tiene un subcampo isomorfo a F basta considerar la aplicación $\phi : F \longrightarrow \frac{F[x]}{<p(x)>}$, tal que $\phi(c) = c + < p(x) >= \overline{c}$ para cualquier $c \in F$. ϕ es un homomorfismo uno a uno, F es entonces isomorfo a su imagen que es un subcampo de $\frac{F[x]}{<p(x)>}$.

Veamos ahora que el polinomio $p(x)$ tiene una raíz en $\frac{F[x]}{<p(x)>}$. Sea $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ y consideremos

$$p(x)+ < p(x) >= a_0+a_1(x+ < p(x))+a_2(x+ < p(x))^2+\cdots+a_n(x+ < p(x) >)^n$$

Hemos aplicado aquí propiedades de la adición y multiplicación de clases. Si notamos ahora $\overline{x} = x+ < p(x) >$ se tiene que

$$p(x)+ < p(x) >= a_0 + a_1\overline{x} + a_2(\overline{x})^2 + \cdots + a_n(\overline{x})^n$$

Pero $p(x) \in < p(x) >$ y por ende $p(x)+ < p(x) >=< p(x) >$. Además como usted lo notaría en la primera parte $< p(x) >$ es el módulo del campo cociente, o sea, $< p(x) >= 0$. Concluimos entonces que

$$a_0 + a_1\overline{x} + a_2(\overline{x})^2 + \cdots + a_n(\overline{x})^n = 0$$

de donde $\overline{x}$ es raíz de $p(x)$ en $\frac{F[x]}{<p(x)>}$.

8.3.1 *Factorización Unica en $F[x]$*

Para concluir esta sección discutiremos otros aspectos importantes de la factorización de polinomios con coeficientes en un campo arbitrario F. La discusión es completamente análoga a la que se dió en nuestro otro anillo modelo, el de los números enteros. Las ideas son además muy familiares desde la básica en el trabajo con polinomios en $\mathbb{Q}[x]$ o en $\mathbb{R}[x]$, pero, como anotamos en el capítulo de anillos nuestro punto es analizar hasta donde es posible generalizarlas.

Definición 8.9 *Sea F un campo. Si dado $f(x) \in F[x]$, $f(x) = g(x)h(x)$ con $g(x), h(x) \in F[x]$, se dice que $g(x)$ divide $f(x)$ y $h(x)$ divide a $f(x)$, o en símbolos, $g(x) \mid f(x)$ y $h(x) \mid f(x)$.*

Definición 8.10 *Sean $p(x)$ y $q(x)$ polinomios en $F[x]$, no ambos nulos. Un polinomio $d(x)$ se dice el máximo común divisor de $p(x)$, y $q(x)$ si es un divisor común de $p(x)$ y $q(x)$ ($d(x) \mid p(x)$ y $d(x) \mid q(x)$) que es divisible por todos sus divisores comunes. (Es decir, si $e(x) \mid p(x)$ y $e(x) \mid q(x)$, entonces $e(x) \mid d(x)$.) Se escribe $d(x) = m.c.d(p(x), q(x))$ o simplemente $d(x) = (p(x), q(x))$.*

Uno de los resultados que continuamente usamos del anillo de los enteros es el hecho de que el máximo común divisor entre dos enteros puede ser expresado como combinación lineal de éstos y tanto el máximo común divisor como la combinación pueden encontrarse aplicando el Algoritmo de

Euclides, que consiste simplemente en la aplicación reiterada del Algoritmo de la División. Como ya demostramos que este último algoritmo se tiene en $F[x]$, nos resultará muy natural el siguiente teorema,

Teorema 8.12 *(Algoritmo de Euclides). Sean $p(x)$ y $q(x)$ polinomios en $F[x], q(x) \neq 0$. Existen polinomios $s(x)$ y $t(x)$ en $F[x]$ tales que $d(x) = (p(x), q(x)) = s(x)p(x) + t(x)q(x)$.*

Demostración
Por algoritmo de la división existen $q_1(x)$ y $r_1(x)$ en $F[x]$ tales que

$$p(x) = q_1(x)q(x) + r_1(x),$$

con $r_1(x) = 0$ o $deg(r_1(x)) < deg(q(x))$. Si $r_1(x) \neq 0$, aplicando el algoritmo de la división a $q(x)$ y a $r_1(x)$

$$q(x) = q_2(x)r_1(x) + r_2(x)$$

con $r_2(x) = 0$ o $deg(r_2(x)) < deg(r_1(x))$. Si $r_2(x) \neq 0$, aplicando el algoritmo de la división a $r_1(x)$ y a $r_2(x)$

$$r_1(x) = q_3(x)r_2(x) + r_3(x)$$

con $r_3(x) = 0$ o $deg(r_3(x)) < deg(r_2(x))$; y así sucesivamente. En el paso en el cual $r_i(x) = 0$, el proceso termina. Demostraremos que $r_{i-1}(x)$ para $i > 1$, o $q(x)$ en el caso en que $i = 1$ es el máximo común divisor. Supongamos que $r_i(x) = 0$. Se tiene entonces

$$r_{i-2}(x) = q_i(x)r_{i-1}(x) + 0$$

Se sigue que $r_{i-1}(x)|r_{i-2}(x)$. De la división anterior

$$r_{i-3}(x) = q_{i-1}(x)r_{i-2}(x) + r_{i-1}(x),$$

de donde, se concluye que $r_{i-1}(x)|r_{i-2}(x)$. Trabajando hacia atrás la sucesión de divisiones, podemos probar similarmente que $r_{i-1}(x)$ divide a $r_2(x)$ y a $r_1(x)$, entonces

$$r_{i-1}(x)|q(x) = q_2(x)r_1(x) + r_2(x),$$

de donde, se concluye que

$$r_{i-1}|p(x) = q_1(x)q(x) + r_1(x).$$

En el caso de que $i = 1$, $r_1(x) = 0$ y $p(x) = q_1(x)q(x)$ y $q(x) = (p(x), q(x))$. Hemos probado que $r_{i-1}(x)$ es un divisor común de $p(x)$ y $q(x)$ si $r_i(x) = 0$. Supongamos ahora que no existiera residuo cero pero si un residuo de grado cero, esto es, un elemento no nulo del campo. Tal elemento es automáticamente un divisor de $p(x)$ y de $q(x)$ polinomios de $F[x]$. Si ahora reescribimos $r_{j-1}(x) = q_{j+1}(x)r_j(x) + r_{j+1}(x)$ en cualquier paso j, como $r_{j+1}(x) = r_{j-1}(x) - q_j(x)r_j(x)$, tendríamos una manera de expresar cada residuo en términos de los primeros residuos de la sucesión. De esta forma sustituyendo cada residuo, podremos expresar $r_{i-1}(x)$ cuando $r_i(x) = 0$ como una combinación lineal de la forma

$$s(x)p(x) + t(x)q(x).$$

Como ya demostramos que $r_i(x)$ es divisor común de $p(x)$ y de $q(x)$, la combinación anterior lo es, pero además cualquier divisor común de $p(x)$ y $q(x)$ divide a $s(x)p(x) + t(x)q(x)$, es decir, tenemos precisamente que

$$(p(x), q(x)) = s(x)p(x) + t(x)q(x).$$

Punto de investigación
- Revisar la Teoría General de factorización en anillos y proponer una demostración alterna del Algoritmo de Euclides en el anillo $F[x]$.

Nota -$F[x]$ es claramente entonces un dominio Euclidiano.

Ilustraremos la aplicación de algoritmo en el siguiente ejemplo.

Sea $p(x) = 6x^5 + 27x^4 - 23x^3 - 78x^2 - 57x - 14$ y $q(x) = 3x^4 + 12x^3 - 19x^2 - 28x - 12$, las divisiones sucesivas nos dan

$$p(x) = q_1(x)q(x) + r_1(x)$$

$$p(x) = (2x+1)q(x) + (3x^3 - 3x^2 - 5x - 2)$$

$$q(x) = q_2(x)r_1(x) + r_2(x)$$

$$q(x) = (x+5)(3x^3 - 3x^2 - 5x - 2) + (x^2 - x - 2)$$

$$r_1(x) = q_3(x)r_2(x) + r_3(x)$$

$$r_1(x) = (3x)(x^2 - x - 2) + (x - 2)$$

y

$$r_2(x) = q_4(x)r_3(x) + r_4(x)$$

$$x^2 - x - 2 = (x+1)(x-2) + 0.$$

Como $r_4(x) = 0$ el máximo común divisor es $r_3(x) = x - 2$. En ese caso despejando de atrás hacia adelante en el conjunto de igualdades, obtenemos que

$$x - 2 = s(x)p(x) + t(x)q(x),$$

con $s(x) = 3x^2 + 15x + 1$ y $t(x) = -6x^3 - 33x^2 - 20x - 1$.

Punto de discusión
Sean $p(x), q(x) \in F[x]$. Demostrar que si $p(x)$ y $q(x)$ no tienen factores comunes de grado positivo, entonces existen en $F[x]$ polinomios $s(x)$ y $t(x)$ tales que

$$1 = s(x)p(x) + t(x)q(x).$$

Teorema 8.13 *Sea $p(x)$ un polinomio irreducible en $F[x]$. Si $p(x)$ divide a $q(x)h(x)$, para $q(x), h(x) \in F[x]$, entonces $p(x)|q(x)$ o $p(x)|h(x)$.*

Punto de discusión
- Usando el punto de discusión anterior demostrar el teorema(8.13).

Presentaremos una demostración alterna.

Demostración
Como $p(x)|g(x)h(x)$ entonces $g(x)h(x) \in < p(x) >$ el cual es ideal maximal de $F[x]$. Pero entonces $< p(x) >$ es ideal primo y como $g(x)h(x) \in < p(x) >$, concluimos que $g(x) \in < p(x) >$ y esto significa que $p(x)|g(x)$ o $h(x) \in < p(x) >$ y esto significa que $p(x)|h(x)$.

Punto de discusión
Usar inducción para demostrar que si $p(x)$ es irreducible en $F[x]$ y $p(x)|r_1(x)\cdots r_n(x)$ para $r_i(x) \in F[x]$ entonces $p(x)|r_i(x)$ para algún i.

Teorema 8.14 *Sea F un campo, entonces todo polinomio no constante $f(x) \in F[x]$ puede ser factorizado en $F[x]$ en un producto de polinomios*

irreducibles, los polinomios irreducibles son únicos excepto por el orden y por factores unidades (constantes no nulas en F).

Demostración.
Sea $f(x) \in F[x]$ un polinomio no constante. Si $f(x)$ no es irreducible entonces $f(x) = g(x)h(x)$ con el grado de $f(x)$ y el de $g(x)$ menores ambos que el grado de $f(x)$. Si $g(h)$ y $h(x)$ son ambos irreducibles, entonces el proceso termina. Si no al menos uno de estos se factoriza en polinomios de grado menor. Continuando este proceso, encontramos la factorización

$$f(x) = p_1(x)p_2(x)\cdots p_r(x)$$

donde los $p_i(x)$ son irreducibles. Para demostrar la unicidad basta suponer dos factorizaciones en irreducibles

$$f(x) = p_1(x)p_2(x)\cdots p_r(x) = q_1(x)q_2(x)\cdots q_s(x).$$

Entonces $p_1(x)|q_j(x)$, para algún j. Asumimos sin perdida de generalidad que $j = 1$. Dado que $q_1(x)$ es irreducible,

$$q_1(x) = k_1 p_1(x),$$

donde $k_1 \neq 0$, pero $k_1 \in F$ (es un unidad). Si sustituimos $q_1(x)$ y cancelamos, tenemos que

$$p_2(x)\cdots p_r(x) = k_1 q_2(x)\cdots q_s(x).$$

Por un razonamiento similar, tenemos que $q_2(x) = k_2 p_2(x)$, por tanto

$$p_3(x)\cdots p_r(x) = k_1 k_2 q_3(x)\cdots q_s(x).$$

Si continuaramos de esta manera llegaríamos a

$$1 = k_1 k_2 \cdots k_r q_{r+1}(x)\cdots q_s(x).$$

Es claro que esto sólo es posible si $s = r$, entonces

$$1 = k_1 k_2 \cdots k_r.$$

Por tanto, los factores irreducibles $p_i(x)$ y $q_j(x)$ son los mismos, excepto posiblemente por orden y factores constantes.

Nota. - El anillo $F[x]$ es un *dominio de factorización única.*

Puntos de Discusión

1. Sea $f(x)$ un polinomio de grado n en $F[x]$. Demostrar que la ecuación polinómica $f(x) = 0$ no puede tener mas de n raíces distintas.
2. Si A es un anillo y $f(x) \in A[x]$, ¿Que condiciones mínimas debe tener A para que la afirmación de (1) se cumpla?.

Puntos de investigación

1. Sea D un dominio de integridad. ¿Es siempre $D[x]$ un dominio de integridad?. Explorar inicialmente el anillo $\mathbb{Z}[x]$.
2. $\mathbb{Z}$ es un dominio euclidiano. ¿Es $\mathbb{Z}[x]$ un dominio euclidiano?.
3. ¿Es $D[x]$ un dominio de factorización única cuando el anillo D lo es?. Investigar ejemplos.
4. Sea A un anillo conmutativo con elemento identidad. Encontrar condiciones sobre los coeficientes para que el elemento $a_0 + a_1 x + \cdots + a_n x^n \in A[x]$ sea inversible.

8.4 Problemas del capítulo

1. Sea p un primo. ¿ Cuántos diferentes de grado n sobre $\mathbb{Z}_p$ hay?. Explorar una formula para el número de polinomios mónicos en $\mathbb{Z}_p[x]$ de grados 2,3,4; ¿cúa les no pueden ser expresados como un producto de polinomios de grado menor?.

2. Sea $p(t)$ un polinomio cuadrático mónico con coeficientes en $\mathbb{Q}$. Demostrar que para cualquier entero n, existe un entero k tal que:

$$p(n)p(n+1) = p(k)$$

3. Considerar un polinomio $f(x)$ con coeficientes reales que tiene la propiedad de que $f(g(x)) = g(f(x)$ para todo polnomio $g(x))$ con coeficientes reales. Determinar la naturaleza de $f(x)$.

4. Encontrar condiciones necesarias y suficientes para que el polinomio $ax^4 + bx^3 + cx^2 + dx + e$, $(a, b, c, d, e \in \mathbb{R}, a \neq 0)$ sea de la forma $p(q(x))$ donde p y q son ambos cuadráticos.

5. (a) Considerar el polinomio cuadrático $6t^2+2t-20$. Determinar dos enteros u y v para los cuales $u + v = 2$ y $uv = -120$. Verificar que $6t + u$ y $vt - 20$ son múltiplos del mismo polinomio lineal. Escribir el polinomio inicial en la forma

$$6t^2 + ut + vt - 20$$

y factorizarlo.

   (b) Para el polinomio cuadrático $at^2 + bt + c$ demostrar como, la determinación de enteros u y v con $u + v = b$ y $uv = ac$ puede dar una factorización del cuadrático en factores lineales sobre $\mathbb{Z}$.

6. Demostrar que si un polinomio de grado n es reducible entonces al menos uno de sus factores tiene grado que no excede $\frac{n}{2}$.

7. Encontrar polinomios $p(x)$ tales que $p(x)$ sea divisible por $x^2 + 1$ y $p(x) + 1$ sea divisible por $x^3 + x^2 + 1$.

8. Demostrar que la fración $\frac{n^3+2n}{n^4+3n^2+1}$ es irreducible para todo natural n.

9. Encontrar todos los enteros positivos m tales que x^2+2 sea un divisor de $x^5 - 10x + 12$ en $\mathbb{Z}_m[x]$.

10. (a) ¿Pueden dos polinomios distintos (con coeficientes racionales por ejemplo) tener los mismos ceros?. ¿Pueden dos polinomios mónicos distintos tener los mismos ceros?. ¿Pueden dos polinomios mónicos distintos tomar por ejemplo el valor -7, exactamente en el mismo conjunto de valores de la variable?.

    (b) ¿Pueden existir dos números distintos a y b y dos polinomios mónicos distintos $p(x)$ y $q(x)$, tales que $p(x)$ y $q(x)$ tomen el valor a exactamente en el mismo conjunto de valores de la variable y también p y q tomen el valor b en el mismo conjunto de valores de la variable?.

Capítulo 9

Números complejos y el Teorema Fundamental del Algebra

La solución de ecuaciones polinómicas (como ya lo hemos comentado en capítulos anteriores) estuvo, y sigue estando, significativamente ligada al planteamiento y solución de problemas fundamentales de la matemática. En este capítulo discutiremos este nexo desde dos perspectivas, una es, cómo la paulatina construcción de significado para los números complejos, que aparecen precisamente en la solución algebraica de ecuaciones polinómicas se constituyó en punto de apoyo para la caracterización completa de las raíces; y a su vez el poder garantizar la existencia de raíces para tales ecuaciones permitió no sólo dar solidez a la teoría de ecuaciones, sino resolver importantes problemas del cálculo. Y la otra, cómo, para demostrar la existencia de las mencionadas raíces (Teorema Fundamental del Algebra) se requieren resultados esencialmente analíticos, esto es argumentos del cálculo.

Partiremos de una revisión del proceso histórico de construcción de significado de los números complejos, del que usted puede derivar alternativas para su transposición en el aula que contrasten con la árida y esquemática presentación tradicional. Dicha revisión culmina en esta primera sección con una aproximación intuitiva a las ideas del análisis que sirvieron como base para fundamentar definitivamente la teoría de ecuaciones.

9.1 Números Complejos

9.1.1 Antecedentes históricos

La primera referencia a la raíz cuadrada de un número negativo (número complejo) se encuentra en los trabajos de Estereometría de Herón de Alejandría (segunda mitad del primer siglo D.C), donde aparece $\sqrt{81-144}$, que es sustituida (sin explicación) por $\sqrt{144-81}$. Realmente el problema original involucraría, una (para la época) "solución imposible" $3\sqrt{-7}$ que no es planteada por Herón; sin embargo existe la duda si éste es un error

de los copistas o del propio Herón.

Punto de Discusión.
En la práctica el "supuesto error" de Herón es corriente en los estudiantes, para ellos no tiene significado alguno $\sqrt{-4}$, pero $\sqrt{4}$ sí, optan simplemente por sustituir. ¿Cómo abordaría usted este problema?.

La siguiente mención a la "imposibilidad"de determinar la raíz cuadrada de un número negativo se encuentra en la Aritmética de Diofanto (275 D.C). El pretendía calcular los lados de un triángulo rectángulo de perímetro 7 y área 12 y encontró que para ello era necesario resolver la ecuación

$$336x^2 + 24 = 172x.$$

¡Explorar porqué!

Afirmaba Diofanto que esta ecuación no podía ser resuelta a menos que el cuadrado de la mitad del coeficiente de x menos $24 \cdot 336$ fuese un cuadrado, para nosotros, a menos que este número fuese un real positivo (recuérde que Diofanto estaba trabajando en los racionales).

En el 850 D.C, el hindu Mahavira es el primero en establecer claramente la dificultad en la solución de este tipo de problemas, comentando en su discusión acerca de los números negativos: "Dada la naturaleza de las cosas, un negativo no es un cuadrado; es por esto que no tiene raíz cuadrada".

Tarda en aparecer en la historia una nueva referencia a este problema; es en Italia después de la invención de la imprenta, cuando en su *Summa*, Luca Pacioli (1494) establece que la ecuación cuadrática $x^2 + c = bx$ no puede ser resuelta a menos que $\frac{1}{4}b^2 \geq c$, es decir, reconocía la imposibilidad de encontrar el valor de $\sqrt{-a}$ cuando a es positivo. Presenta pues, un criterio general para que la ecuación cuadrática tenga solución real, que involucra lo que hoy conocemos como el *Discriminante de la Ecuación* al que hicimos referencia en el capítulo 2.

Pero, es Cardano, quien da un paso importante en la construcción de significado de los números complejos. En el capítulo 37 de su *Ars Magna* en 1545 para solucionar el siguiente problema: "Dividir 10 en dos partes cuyo producto sea 40" requiere resolver la ecuación $x(10 - x) = 40$, y obtiene las raíces $5 + \sqrt{-15}$ y $5 - \sqrt{-15}$, habla de la solución por la 'menos raíz' y dice: "Dejando de lado la tortura mental involucrada", multiplico $5+\sqrt{-15}$ por $5 - \sqrt{-15}$; el producto es $25 - (-15)$, esto es, 40. Estos números que desde luego no tienen un claro significado, satisfacen pues las condiciones del problema planteado. Argumenta luego que, en aras del progreso de la aritmética, es importante entonces involucrar y refinar el trabajo con este tipo de números; los usa también posteriormente en su solución de las ecuaciones cúbicas. Es decir, opera por primera vez con estos números y les concede importancia; no tiene una significación clara de ellos pero los manipula y los utiliza, situación esta última, de ocurrencia corriente en las aulas respecto a los complejos, pero que debe constituirse tan solo en un avance en el proceso de construcción de significado.

En 1572, Bombelli en su álgebra al discutir la solución de la ecuación cúbica, se refiere a cantidades tales como $+\sqrt{-a}$ y $-\sqrt{-a}$ y formula lo que en nuestra práctica actual corresponde a las cuatro operaciones entre números complejos. Desafortunadamente las ideas de Bombelli permanecieron guardadadas por mucho tiempo, pues se catalogaron de sofisticadas y poco útiles.

Es Albert Girard en 1629 quien cataloga los números complejos como una solución 'formal', para poder establecer un resultado general sobre el número de raíces de una ecuación polinómica de grado n. En *L'Invention nouvelle en l'algebre*, el dice: "Uno podría decir: ¿Por qué usar estas soluciones imposibles (raíces complejas)? Yo respondería: Por tres cosas, por la certidumbre de reglas generales, por su utilidad y porque no hay otras soluciones." Sin embargo la visión avanzada de Girard no tuvo tampoco gran influencia en esa época.

En 1673 Wallis da un paso realmente fundamental en el proceso de construcción de significado de los números complejos; es el primero que presenta una idea de la representación gráfica de estos números. En su *Algebra* mostró cómo representar geométricamente las raíces complejas de una ecuación cuadrática con coeficientes reales. Wallis dice: " Los números complejos no son mas absurdos que los negativos y, dado que estos pueden ser representados sobre una recta, es posible representar los números complejos en un plano ". Consideró para esto, un eje sobre el cual ubicó los números reales con respecto a un origen, la distancia del origen al punto representaba la parte real de la raíz (distancia medida en la dirección positiva o negativa, según esta parte fuera positiva o negativa); desde un punto en el eje real levantaba una perpendicular al eje cuya longitud representaba el número que multiplicado por $\sqrt{-1}$ daba la parte imaginaria de la raíz; la línea era dibujada en una dirección o en la opuesta de acuerdo a que el número fuera positivo o negativo. Wallis discute entonces una construcción geométrica para las raíces de $ax^2 + bx + c = 0$ cuando las raíces son reales y cuando son complejas.

Punto de Discusión.
Wallis, pretende dar significación a los números complejos estableciendo una analogía con los números negativos y su posibilidad de representarlos. Analice usted hasta qué punto esta analogía puede resultar provechosa en el aula. Discuta las limitaciones de dicha analogía.

En lo que concierne a Descartes, realmente él deshecha este tipo de raíces y les asigna el término "imaginario". Dice en *La Géométrie:* "Ni las raíces verdaderas, ni las falsas (negativas) son siempre reales, algunas veces ellas son imaginarias". Argumentaba que las raíces negativas pueden al menos hacerse 'reales' transformando la ecuación correspondiente en otra cuyas raíces son positivas (efectuando una traslación, como ya se comentó en el capítulo 4), pero, esto no puede hacerse para las raíces complejas. Es por esto que estas no son reales sino imaginarias; ellas no son números, concluía. Descartes establece pues una clara distinción entre raíces reales y raíces imaginarias.

Punto de Discusión. ¿Por qué el argumento de Descartes respecto a las raíces negativas no funciona con las raíces complejas?.

Newton descartó las raíces complejas probablemente por su falta de significado físico, es probable que lo hiciera mas bien, por la falta de significado físico. En la *Universal Arithmetic* decía: "...Las raíces de una ecuación pueden ser imposibles (complejas), esto puede ocurrir tanto en casos en que el problema es imposible, como en casos en que el problema es posible". Es decir, si el problema no tiene solución física o geométrica real deberá tener raíces complejas.

Leibniz trabajó formalmente con los números complejos, por ejemplo, fac-

torizó la polinomial $x^4 + a^4$ como

$$\left(x + a\sqrt{-\sqrt{-1}}\right)\left(x - a\sqrt{-\sqrt{-1}}\right)\left(x + a\sqrt{\sqrt{-1}}\right)\left(x - a\sqrt{\sqrt{-1}}\right),$$

pero no entendió su naturaleza como se hace patente en comentarios acerca de las raíces imaginarias, posiblemente porque no tenía una idea acerca de su representación gráfica.

Los números complejos permanecen prácticamente escondidos para los matemáticos hasta el siglo XVIII; son casi ignorados desde que los introduce Cardano hasta cerca del 1700, cuando son usados para integrar por el método de las fracciones parciales. Esto origina una larga controversia acerca de los números complejos referida especialmente a las exponenciales y los logaritmos de éstos. Es Euler quien a pesar de no presentar una definición explícita de tales números, los manipula, define y usa correctamente la exponencial y el logaritmo complejos. Euler intenta aproximarse a una noción de número complejo, en su *Complete Introducción to Algebra* publicada en Alemania en 1770 dice: "Porque todo número concebible es mayor, igual, o menor que 0, es claro que las raíces cuadradas de números negativos no pueden ser incluidas entre estos números posibles (reales). En consecuencia, debemos decir que son números imposibles. Esta circunstancia conduce al concepto de tales números, que por su naturaleza son imposibles, y ordinariamente son llamados imaginarios pues existen solamente en la imaginación". Podría decirse además que es Euler quien introduce la variable compleja. En el texto antes citado afirma: "Dada una cantidad variable ésta también incluye valores imaginarios'; demuestra en él el teorema de factorización de un polinomio con coeficientes enteros y anota que estos factores simples son reales o imaginarios de acuerdo a la naturaleza de las raíces. Euler comete algunos errores con los números complejos; por ejemplo, en su álgebra escribe $\sqrt{-1}\sqrt{-4} = \sqrt{4} = 2$ porque $\sqrt{a}\sqrt{b} = \sqrt{ab}$. ¡Nótese de nuevo la similitud con argumentos usados por los estudiantes, para justificar un "error" que por la falta de significado no puede ser considerado como tal!.

Punto de Discusión.
¿Cúal o cúales de las siguientes afirmaciones dadas por un estudiante considera usted incorrectas?

1. $$(\sqrt{-4})(\sqrt{-16}) = \sqrt{(-4)(-16)}$$

2. $$\sqrt{(-4)(-16)} = \sqrt{64}$$

3. $$\sqrt{64} = 8$$

Construir estrategias para discutir estas ideas en el aula.

A pesar de los avances en la construcción de significado de etapas anteriores en los inicios del siglo XVIII los matemáticos pensaban aún que las raíces distintas de un número complejo introducían diferentes tipos de órdenes al número y que esto podría dar lugar a raíces ideales, de naturaleza no especificada, lo que impediría su completa determinación. Es D'Alembert quien en su trabajo *Réflexions sur le cause générale des vents* (1747), aclara

este punto. Afirma que toda expresión construida a partir de un número complejo por medio de operaciones algebraicas (en las que incluye raíces y potencias), es un número complejo de la forma $A + B\sqrt{-1}$. La única dificultad que él tiene es demostrar esta afirmación en el caso de $(a+bi)^{c+di}$. Se da pues, un paso adelante en la construcción de significado para la estructura misma de los complejos así como para la caracterización de las raíces de una ecuación polinomial.

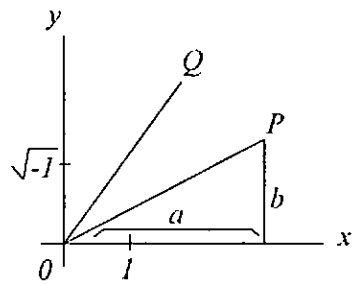
En el transcurso del siglo XVIII los números complejos son usados cada vez de manera más efectiva, los matemáticos van adquiriendo mayor confianza con ellos; en 1799 cuando Gauss da su primera demostración del teorema fundamental de álgebra (que discutiremos posteriormente en este capítulo) y dado que ésta depende del reconocimiento de tales números, Gauss solidifica su posición.

Hacia 1830 los conceptos de número complejo y función de variable compleja están en vía de clarificación se da entonces en ese momento un paso vital para que la teoría de complejos y de funciones en variable compleja resulte mas intuitiva y razonable, se consolida la representación geométrica de estos números y de las operaciones algebraicas con ellos. Varios matemáticos Côtes, De Moivre, Euler y Vandermonde, usando la representación de los números complejos como puntos del plano, concluyen que las soluciones de la ecuación $x^n - 1 = 0$,

$$cos\frac{2k\pi}{n} + isen\frac{2k\pi}{n}$$

son los vértices de un polígono regular. Algunos historiadores afirman que la correspondencia entre números complejos y parejas de puntos del plano era conocida ya en 1800, sin embargo la identificación decisiva no se hace hasta que a las operaciones algebraicas con estos números se les da significado geométrico.

En 1797 el matemático noruego Caspar Wessel escribe un trabajo titulado "Sobre la representación analítica de la dirección"; en él representa geométricamente segmentos de recta dirigidos (vectores) y realiza operaciones con ellos. En él introduce un eje imaginario "y" en el que asocia $\sqrt{-1}$ con una unidad y el eje real x con una unidad real 1.



En la representación geométrica de Wessel el vector OP es el segmento de línea OP dibujado del origen O en el plano de unidades $+1$ y $\sqrt{-1}$, el vector es representado por el número complejo $a + b\sqrt{-1}$. Similarmente el vector OQ es el segmento OQ y está representado por otro número $c + d\sqrt{-1}$. Wessel define las operaciones con vectores, definiendo las operaciones con complejos en términos geométricos. Su definición de las cuatro operaciones es practicamente la que estudiamos hoy. La suma de $a + bi$ y $c + di$ es la diagonal del paralelogramo con lados adyacentes OP y OQ. El producto de $a + bi$ y $c + di$ es un nuevo vector OR, tal que OR es a OQ como OP es a la unidad real, y el ángulo formado por OR y el eje x es la suma de los

ángulos formados por OP y OQ. Claramente Wessel razona en términos de vectores mas que en términos de asociar números complejos con puntos en el plano.

Una representación geométrica diferente de los números complejos fue dada por el suizo Jean-Robert Argand (1768-1822); el observó que los números negativos eran una extensión de números positivos que resultaba de combinar magnitud con dirección. Se preguntó entonces ¿Puede uno extender el sistema de los números reales adicionando un nuevo concepto? Consideró entonces la secuencia $1, x, -1$. Podemos encontrar una operación que transforme 1 en x y x en -1?. Si rotamos OP (Figura 2) 90° en el sentido contrario de las manecillas del reloj alrededor de O y entonces repetimos la rotación, vamos de P a Q por operaciones sucesivas. Entonces Argand notó que esto es lo que ocurre precisamente si multiplicamos 1 por $\sqrt{-1}$ y luego multiplicamos este producto por $\sqrt{-1}$; esto es, nosotros obtenemos -1. Podemos pensar entonces a $\sqrt{-1}$ como una rotación de 90° en el sentido contrario de las manecillas, es decir $-\sqrt{-1}$ es una rotación de 90° en el sentido de las manecillas.

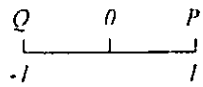


Figura 2

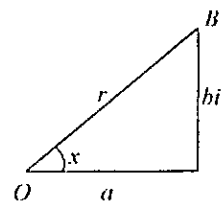


Figura 2a

Para utilizar este significado operacional de los números complejos, Argand decidió que un segmento de línea típico OB (Fig 2a) que parta del origen puede ser representado por

$$r(\cos\alpha + i\,sen\,\alpha),$$

donde r es la longitud. Planteó además que el número complejo $a + bi$ puede ser representado por OB que es combinación de a y bi. Demostró también que los números complejos pueden ser adicionados y multiplicados geométricamente y aplicó estos resultados para demostrar teoremas de trigonometría, geometría y álgebra.

Gauss (como lo analizaremos posteriormente) uso de manera significativa estos argumentos para demostrar el teorema fundamental del álgebra. En su trabajo presupuso una correspondencia uno a uno entre puntos del plano cartesiano y números complejos, separando además las funciones de variable compleja en parte real y parte imaginaria; utilizó en su analisis la representación geométrica. Al respecto comentaba: "En la representación geométrica uno encuentra un significado intuitivo de los números complejos completamente establecido y mas aun, uno no necesita admitir estas cantidades en el dominio de la aritmética". Introduce Gauss el término 'número complejo' y usa por primera vez i para $\sqrt{-1}$.

En 1853 W.R. Hamilton en su trabajo sobre los cuaterniones presenta un concepto de número complejo aun mas intuitivo, observando que la base esencial del concepto está en las operaciones y sus propiedades. El número

complejo se define entonces formalmente como un par ordenado de reales.

9.2 La Estructura de los Números Complejos

9.2.1 Caracterización

Como lo comentamos en la sección anterior el estudio de la solución algebraica de ecuaciones polinómicas de segundo grado inicialmente generó un largo debate y un interesante proceso de construcción de significado que desembocó en la caracterización (conceptualización y representación) de los números complejos, sus operaciones y propiedades. Es por ello que tradicionalmente en la educación básica este tópico se aborda tan solo desde esta perpectiva, mas específicamente, a partir de la imposibilidad de encontrar soluciones reales para la ecuación $x^2+1=0$. Se habla de la llamada unidad imaginaria i, $i=\sqrt{-1}$, "i" satisface la ecuación, esto es $i^2=-1$, i tiene entonces el caracter de "solución formal". La solución de ecuaciones menos simples como $2x^2+3=0$, $x^2+x+1=0$ o $x^2-x+4=0$ da luego lugar a expresiones de la forma $a+bi$ donde a y b son reales, e i es la unidad imaginaria.

Puntos de Discusión

1. ¿Cúal es el menor valor de k para el cual la ecuación

$$2x(kx-4)-x^2+6=0$$

no tiene raíces reales?.

2. Si se sabe que una de las raíces de la ecuación

$$2x^2+rx+s=0$$

es $3+2i$. ¿Cúal es el valor de s?

El conjunto de los números complejos se define entonces como

$$\mathbb{C}=\{a+bi : a,b\in\mathbb{R}\},$$

y es dotado de una estructura, caracterizando completamente sus objetos y definiendo operaciones entre ellos.

Dados dos complejos z y w, $z=a+bi$ y $w=c+di$,

$$z=w \Longleftrightarrow a=b \wedge c=d.$$

La adición y multiplicación entre números complejos se define simplemente asumiéndolos como expresiones polinómicas y teniendo en cuenta que $i^2=-1$.

$$z+w=(a+bi)+(c+di)=(a+c)+i(b+d)$$
$$z\cdot w=(a+bi)(c+di)=(ac-bd)+i(ad+bc)$$

Punto de Discusión

1. ¿Cúal o cúales de las afirmaciones que aparecen a continuación, relativas a la ecuación
$$ix^2 - x + 2i = 0,$$
no es verdadera? Explicar por qué.

   (a) La suma de sus raíces es 2.

   (b) El discriminante es 9.

   (c) Sus raíces son imaginarias.

   (d) Las raíces pueden ser encontradas a partir de la fórmula cuadrática.

   (e) Las raíces pueden encontrarse por factorización.

Nota. Nótese que en la esquemática presentación anterior no es fácil imaginar desde un punto de vista intuitivo "¿Qué es i?".

A partir del siglo XIX (como lo comentamos anteriormente) los matemáticos tomaron plena conciencia de que la base esencial del concepto de número está en la caracterización de las operaciones que se definen sobre el conjunto numérico y en las propiedades de éstas, es decir en la *ESTRUCTURA*.

Precisamente en estas ideas se basó W.R. Hamilton para dar una presentación mas intuitiva de los números complejos como sigue;

Un número complejo "z" se define como un par ordenado de reales

$$z = (a, b) \text{ tales que } a, b \in \mathbb{R}.$$

Para que tal presentación resulte consistente debemos extender a estos pares de reales (números complejos) la noción de igualdad y definir el significado de las cuatro operaciones fundamentales (adición, sustración, multiplicación y división) entre estos pares en la siguiente forma:

Definición 9.1 *IGUALDAD: Dos complejos* (a, b) *y* (c, d) *se dice que son iguales si y sólo si* $a = c$ *y* $b = d$.

Definición 9.2 *ADICION:* $(a, b) + (c, d) = (a + c, b + d)$.

Definición 9.3 *MULTIPLICACION:* $(a, b) \cdot (c, d) = (ac - bd, ad + bc)$

Puntos de Discusión.

1. Definir sustracción y división entre pares ordenados de reales a partir de definiciones 1.1, 1.2 y 1.3.

2. Demostrar que el conjunto de los pares ordenados de reales, con la adición y multiplicación definidas, tiene estructura de cuerpo.

Notas.

1. Todo número complejo puede ser escrito entonces en la forma

$$(a, b) = (a, 0) + (0, b),$$

Además

$$(0, b) = (b, 0), \cdot (0, 1)$$

Y por tanto

$$(a, b) = (a, 0) + (b, 0) \cdot (0, 1).$$

El número complejo puede ser expresado entonces mediante complejos de la forma $(a, 0)$ y un complejo especial $(0, 1)$, para el que se tiene que

$$(0, 1) \cdot (0, 1) = (1, 0).$$

2. Obsérvese que si se consideran las cuatro operaciones definidas sobre complejos de la forma $(a, 0)$, el resultado es siempre un par con segunda componente 0; resulta natural por tanto identificar complejos de la forma $(a, 0)$ con el real a. En particular el producto $(0, 1) \cdot (0, 1) = (1, 0)$, puede ser identificado con el real 1 y la llamada unidad imaginaria i, solución de la ecuación $x^2 = -1$, puede ser representada por el par $(0, 1)$.

Puntos de Discusión

1. Encontrar valores reales de x y y que satisfacen la ecuación

$$(1 + i)(x + 2y) - (3 - 2i)(x - y) = 8 + 3y.$$

2. Analizar la siguiente afirmación. Los números a y b satisfacen la ecuación

$$\frac{a + b}{a} = \frac{b}{a + b}$$

solamente si uno es real y el otro es imaginario puro.

3. Determinar raíces reales de la ecuación

$$(1 + i)x^3 + (1 + 2i)x^2 - (1 + i)x - 1 - 2i = 0.$$

4. Hallar todas las ecuaciones de la forma $x^2 + px + q = 0$ con p y q racionales, tales que sus raíces r_1 y r_2 no son racionales cuando

   (a) $r_1^2 = r_2$

   (b) $r_1^3 = r_2$

Cada número complejo se puede representar ahora geométricamente (representación adoptada por K.F Gauss, C. Wessel y J.R Argand a comienzos del siglo XIX) por un punto del plano; las partes real e imaginaria son las coordenadas de este punto en un sistema cartesiano ortogonal, cuyos ejes se llaman eje real y eje imaginario. El plano utilizado para representar el conjunto $\mathbb{C}$ de los complejos se llama PLANO COMPLEJO.

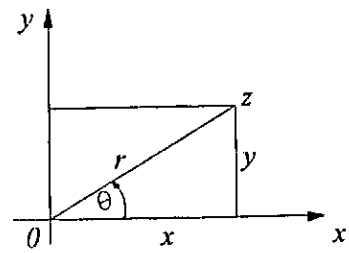


Figura 3

Si z es un número complejo no nulo, $z = x + iy \neq 0$, el punto P que lo representa (Fig 3) queda unívocamente determinado por su distancia al origen, O, que es r y el ángulo θ que forma el eje real con el vector $\overrightarrow{OP}$. El ángulo θ es llamado el *ARGUMENTO* del complejo z. r es la magnitud (módulo) del complejo z notada $|z|$. De la figura

$$r = |z| = \sqrt{x^2 + y^2} = \sqrt{z.\overline{z}},$$

donde $\overline{z} = x - iy$, el complejo conjugado de z. Y $\tan\theta = \frac{y}{x}$.

Puntos de Discusión

1. ¿Cúal es la parte real del complejo $\frac{1-z}{1+z}$, si $z = cos\,\theta + isen\,\theta$?
2. Encontrar en cada caso un complejo z tal que:
   (a) $z = (\overline{z})^2$.
   (b) $|z| = 1$ y $Re(z^2) = 0$. (Re parte real).
3. Si z es un número complejo tal que $|z| \leq 2$. ¿Cuál es el máximo de $|1 + z + z^2 + z^3|$?. Y para qué valor de z se alcanza este máximo?.

Los números r y θ son las *COORDENADAS POLARES DEL COMPLEJO* z. Pero como existen infinitos argumentos posibles para el complejo z, todos los de la forma $\theta + 2k\pi$ con k entero, basta considerar el llamado *ARGUMENTO PRINCIPAL,*

$$-\pi < \theta \leq \pi.$$

Se tiene entonces que $x = r \cdot cos\,\theta$; $y = r \cdot sen\,\theta$ Reemplazando, z queda expresado en la forma *POLAR O TRIGONOMETRICA*

$$z = r \cdot (cos\,\theta + isen\,\theta) = r \cdot cis\theta$$

Donde $cis\,\theta = cos\,\theta + isen\,\theta$. Esta expresión es de gran utilidad para operar. Si ahora z y x son dos complejos con expresión polar:

Veamos

$$z = a + bi = r_1(cos\,\theta_1 + isen\,\theta_1)$$
$$w = c + di = r_2(cos\,\theta_2 + isen\,\theta_2)$$

entonces

$$zw = (ac - bd) + i(ad + bc) = r_1 r_2(cos(\theta_1 + \theta_2) + isen(\theta_1 + \theta_2),$$

y

$$\frac{z}{w} = \frac{r_1}{r_2}[cos(\theta_1 - \theta_2) + isen(\theta_1 - \theta_2)].$$

Punto de discusión
Además si $n \in \mathbb{N}$ y $z = r(cos\,\theta + isen\,\theta)$, entonces demostrar que

$$z^n = r^n(cos\,n\theta + isen\,n\theta).$$

Analicemos qué ocurre con la raíz n-ésima de un número complejo
Si $z = r_1(cos\,\theta_1 + isen\,\theta_2)$, nos interesa encontrar w en $\mathbb{C}$ tal que $w^n = z$.
Si $w = r(cos\,\theta + isen\,\theta)$

$$w^n = r^n(cosn\theta + isennn\theta) = z = r_1(cos\,\theta_1 + isen\,\theta_1),$$

de donde podemos concluir
$r = r_1^{\frac{1}{n}}$ y $n\theta = \theta_1 + 2k\pi$.

Por tanto $\theta = \frac{\theta_1 + 2k\pi}{n}$ con $k = 0, 1, 2, \cdots, n-1$. Esto es,

$$z^{\frac{1}{n}} = r_1^{\frac{1}{n}}\left(cos\frac{\theta_1 + 2k\pi}{n} + isen\frac{\theta_1 + 2k\pi}{n}\right)$$

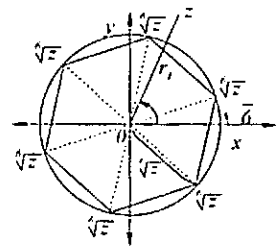


Figura 4

Puntos de discusión

1. ¿Es $(1+i)^{20} - (1-i)^{20}$ un número real?.
2. Encontrar todos los valores enteros de n para los cuales $(n+i)^4$ es un entero.
3. Demostrar que si $x + \frac{1}{x} = 2cos\alpha$, entonces

$$x^n + \frac{1}{x^n} = 2cosn\alpha$$

4. Explicar porqué el número complejo $\frac{\sqrt{2}}{2}(21 - 20i)$ puede ser expresado por la suma

$$cos45° + icos135° + \cdots + i^n cos(45 + 90n)° + \cdots + i^{40}cos3645°$$

5. Si $S = i^n + i^{-n}$, donde n es un entero arbitrario. ¿Cúal es el número de posibles valores distintos de S?.
6. Si n es un entero múltiplo de 4, determinar S donde

$$S = 1 + 2i + 3i^2 + \cdots + (n+1)i^n.$$

Geométricamente la expresión para las raíces n-ésimas puede ser descrita de la siguiente manera. Los puntos del plano complejo que corresponden a los valores de la raíz n-ésima de z son los vértices de un polígono regular de n lados inscrito en un círculo de centro en el origen y radio $r^{\frac{1}{n}}$ y tal que uno de los vértices de este n-ágono regular tiene argumento $\frac{\theta_1}{n}$.

Puntos de Discusión

1. Encontrar todas las soluciones de la ecuación

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

2. Factorizar $x^8 + x^4 + 1$ en factores irreducibles sobre los complejos.

3. Factorizar $x^4 - 2x^3 + 6x^2 + 22x + 13$ sobre los números complejos dado que $2 + 3i$ es un cero.

4. Sean a, b y n enteros positivos, demostrar que existen emteros x y y tales que

$$(a^2 + b^2)^n = x^2 + y^2$$

5. Sea n un entero mayor o igual que 3, y sea α, β y γ complejos tales que $(\alpha)^n = (\beta)^n = (\gamma)^n = 1$, $\alpha + \beta + \gamma = 0$. Demostrar que n es un múltiplo de 3.

6. Dado que $13 = 2^2 + 3^2$ y $74 = 5^2 + 7^2$, expresar 13×74 como una suma de dos cuadrados. (Sug: Considere $z = 2 + 3i$ y $w = 5 + 7i$ y use que $(|z|)^2(|w|)^2 = (|zw|)^2$).

7. Sea $n = 2m$ donde m es un entero impar mayor que 1. Sea $\theta = e^{2\frac{2\pi}{n}}$. Expresar $(1 - \theta)^{-1}$ como un polinomio en θ con coeficientes enteros.

8. Demostrar que si $z = e^{i\theta}$ entonces $z - z^{-1} = 2isenn\theta$.

9. Dado un punto P sobre la circunferencia de un círculo unitario y los vértices $A_1, A_2, \cdots, A_n$ de un polígono regular inscrito de n lados. Demostrar que

$$(PA_1)^2 + (PA_2)^2 + \cdots + (PA_n)^2$$

es una constante.

10. Demostrar que si los puntos en un plano complejo correspondientes a dos complejos distintos z_1 y z_2 son dós vértices de un triángulo equilatero, entonces el tercer vértice corresponde a $-wz_1 - w^2z_2$ donde w es raiz cúbica primitiva de la unidad.

9.2.2 Significado Geométrico de las Operaciones con Números Complejos

La representación geométrica de los números complejos abre el camino, como se mencionó en las notas históricas y en la discusión de la estructura, a las aplicaciones de los números complejos a la geometría. Una construcción geométrica resultará como una imagen de cierta operación realizada con números complejos. Ampliaremos aquí la discusión al respecto.

Nos centraremos en el examen de las construcciones que corresponden a la adición, sustracción, multiplicación y división de números complejos. Estando representados los números complejos z y w por los puntos Z y W, si construimos los vectores OZ y OW con punto inicial el origen de un sistema de coordenadas para visualizar la adición de los complejos bastará aplicar la llamada ley del paralelogramo para sumar estos vectores, el vector resultante $O\Xi$ será su suma y el punto Ξ representará el número complejo $z+w$ (Figura 5).

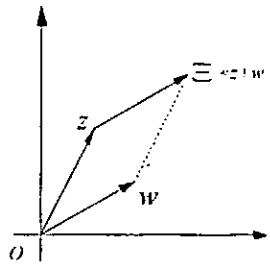


Figura 5

De hecho si

$$z = z_1 + z_2 i \text{ y } W = w_1 + w_2 i$$

las proyecciones de OZ y OW sobre los ejes real e imaginario, serán respectivamente: $z_1; w_1; z_2; w_2$, por lo tanto las proyecciones de $O\Xi$ son

$$z_1 + w_1 \text{ y } z_2 + w_2$$

y en consecuencia

$$\Xi = z_1 + w_1 + i(z_2 + w_2) = z + w.$$

Nota -Nótese que la figura $OZ\Xi W$ es en general un paralelogramo, siempre que los puntos O, Z y W no esten alineados.

Puntos de discusión

- Usar la representación anterior para demostrar que módulo de la suma de dos números complejos es menor o igual que la suma de los módulos. La construcción geométrica para la suma de dos números complejos conduce inmediatamente a la correspondiente construcción para la diferencia $W-Z$. Esta diferencia está representada por el cuarto vértice del paralelogramo, tres vértices consecutivos del cual son: O, Z, W. es claro que el vector que representa la diferencia $W-Z$ tiene la misma magnitud, dirección y sentido que el vector ZW (es equipolente con él) (Figura 6).

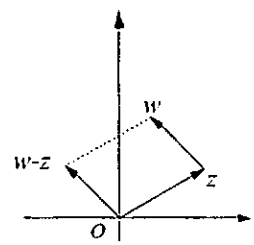


Figura 6

La regla para la multiplicación de números complejos en forma trigonómetrica nos proporciona una construcción simple para el producto de dos números complejos Z y W. Antes de abordar esta construcción es necesario explicar

lo que significa *SENTIDO* al referirnos al triángulo ABC cuyos vértices se toman en el orden indicado en la figura 7.

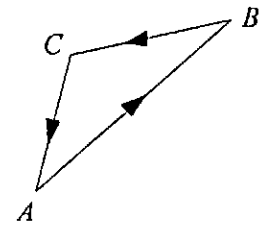


Figura 7

Si vamos de A a B, de B a C y volviendo nuevamente de C a A, el interior del triángulo puede quedar ya a la izquierda, ya a la derecha; en el primer caso decimos que tiene sentido positivo; en el último, que tiene sentido negativo. Así el triángulo ABC representado en la figura tiene sentido positivo; pero el mismo triángulo si sus vértices se toman en el orden ACB, tendrá sentido negativo. Uniendo el punto Z con O y con el punto que en el eje real corresponde al real 1, se forma el triángulo $O1Z$. Tomando ahora OW como lado correspondiente a $O1$, construimos otro triángulo $OW\Xi$ semejante a $O1Z$, es decir, con el mismo sentido y ángulos iguales en los vértices correspondientes (Figura 8).

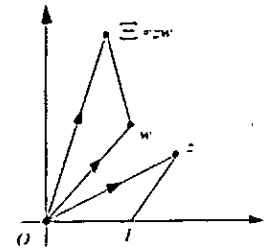


Figura 8

Si ϕ_1 y ϕ_2 son los ángulos entre el eje real y los vectores OZ y OW, por construcción el ángulo entre el eje real y $O\Xi$ es $\phi_1 + \phi_2$. El argumento de Ξ es $\phi_1 + \phi_2$. Además si designamos con ρ, r_1, r_2 las distancias de O a Ξ, Z y W respectivamente, se desprende de la semejanza de los triángulos $O1Z$ y $OW\Xi$ que

$$\frac{\phi}{r_1} = \frac{r_2}{1};$$

en consecuencia $\phi = r_1 r_2$ es el módulo de Ξ. Por lo tanto, Ξ representa el producto ZW.

Punto de discusión

- Efectúe usted una construcción similar para representar $\frac{Z}{W}$.

Sean ahora los números complejos Z, W, Ξ, representados por tres puntos alineados. En este caso como se observa en la figura

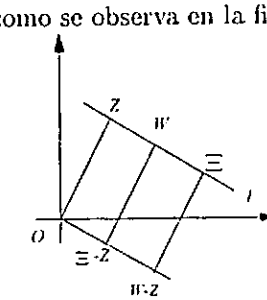


Figura 9

los puntos O, $\Xi - Z$, $W - Z$ están también alineados, y por lo tanto los argumentos de $\Xi - Z$ y $W - Z$ o son iguales o difieren en π. Por consiguiente

$$\Xi - Z = m(W - Z),$$

o sea

$$\Xi = (1 - m)Z + mW,$$

donde m es un número real. Es claro que la recíproca es también cierta, esto es, si m es real, los puntos Ξ, Z, W están alineados. Veamos el significado del número m. Si designamos por r la distancia entre Z y W y por ρ la medida del segmento $Z\Xi$, m es precisamente la razón $\frac{r}{\rho}$.
El vector correspondiente a $i(W - Z)$ es perpendicular a la recta l que une a Z y a W; por tanto usted puede ver que los números complejos

$$\Xi = a + im(W - Z),$$

donde m es un real, representan puntos de la recta trazada por un punto arbitrario a y perpendicular a l.

Puntos de discusión

1. Usar ideas de la subsección anterior para construir un triángulo XYZ, dados los puntos medios P, Q, R de los lados XY, YZ, ZX.

2. Los números complejos $z = a + (b - a)t$ donde a y b son complejos dados y t es un real representan puntos de una recta. Demostrar que las rectas

$$z = a + (b - a)t$$
$$z = c + (d - c)t'$$

   (a) son paralelas si $Im(\frac{b-a}{d-c}) = 0$ (Im parte imaginaria).

   (b) son perpendiculares si $Re(\frac{b-a}{d-c} = 0)$

   (c) ¿Cómo se determina el punto de intersección entre las dos rectas si no son paralelas?

3. Si los números complejos Z, W, Ξ representan los vértices de un triángulo, demostrar que las medianas se cortan en un punto y hallar el número complejo correspondiente a ese punto.

4. Los números complejos Z, W, Ξ representan los vértices de un triángulo $Z, W\Xi$ de sentido positivo. Si a, b, c son las longitudes de los lados ZW, $W\Xi$, $Z\Xi$ y A, B, C los ángulos opuestos, demostrar que

$$\frac{Z - W}{\Xi - W} = \frac{a}{b}(cosC + isenC)$$

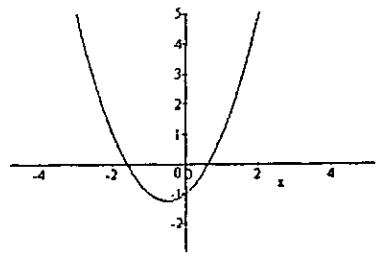
   Y

$$\frac{\Xi - Z}{\Xi - W} = \frac{c}{b}(cosA - isenA)$$

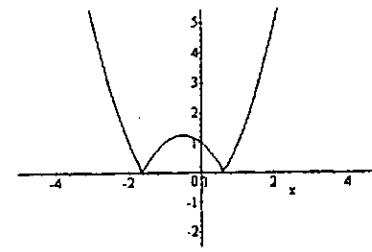
Las ideas relativas a la representación y a la correspondencia entre parejas ordenadas de reales y puntos del plano que hemos comentado en la primera parte pueden ser aprovechadas para analizar más a fondo la estructura del plano complejo; simplemente tocaremos dos aspectos que nos interesan para nuestro análisis del teorema fundamental, la visualización de las raíces de una polinomial y una aproximación intuitiva a la solución de ecuaciones en variable compleja.

9.2.3 Visualizando los Ceros Complejos de un Polinomio

Una inquietud muy corriente en el trabajo con la solución de ecuaciones tiene que ver con el problema de la visualización de las raíces. No existe dificultad alguna si consideramos por ejemplo, el polinomio $x^2 + x - 1$. Geométricamente los ceros de éste son los puntos de intersección de su gráfica con el eje de las x. (Fig 10a). Pero, ¿qué ocurre cuando nos preguntamos por la representación de los ceros de $x^2 + x + 1$, que no posee ceros reales?



$x^2 + x - 1$
Figura 10a



$x^2 + x - 1$
Figura 10b

Diríamos entonces, los ceros de esta polinomial 'desaparecen' de la gráfica, para valores reales de x, pero, esto no sucede si consideramos la gráfica de $p(z) = z^2 + z + 1$, cuando z es un complejo.

Desafortunadamente la gráfica de una función compleja no es fácil de manejar. Pero, como estamos interesados solamente en los ceros, no tenemos necesidad de considerar la gráfica completa. Consideraremos los ceros de una polinomial (y los de cualquier función) como los puntos en la base de una superficie llamada superficie modular. El método es una generalización del hecho de que los ceros de una polinomial $P(X)$ son los puntos mínimos de la gráfica de $|P(X)|$ los cuales estarían sobre el eje x (Fig 10b).

Recordemos que si w es un número complejo $w = u + iv$, con u y v reales, entonces su módulo o magnitud se define como $|w| = \sqrt{u^2 + v^2}$, la cual es no negativa. Si pensamos en la polinomial $p(z) = a_2z^2 + a_1z + a_0$ como una función compleja, de variable compleja $z = x + iy$; el módulo de $p(z)$ es $|p(z)|$. La superficie modular de la polinomial $w = p(z)$ es la superficie definida por las tres variables, x, y y $|w|$. Los ceros de una función a valor complejo definida sobre el plano complejo son los puntos de la superficie modular que tocan el plano complejo, dado que el complejo $w = 0$ si y sólo si su módulo es cero.

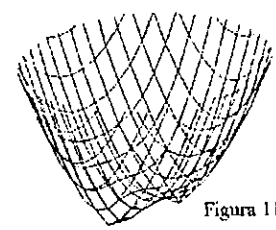


Figura 11

En la figura 11 aparece la superficie modular de la cuadrática $z^2 + z - 1$. Comparando la figura 10b con la figura 11 se muestra que la representación de $|w|$ para z real es la traza de la superficie modular sobre el plano $y = imag(z) = 0$. Podemos imaginar entonces que la superficie modular para el

nuevo polinomio z^2+z+1 sería similar. Dado que este polinomio tiene ceros $\frac{(-1\pm i\sqrt{3})}{2}$, la correspondiente superficie tendrá exactamente dos puntos que tocan el plano complejo, uno por cada cero. La superficie no negativa resultante se muestra en la figura 12.

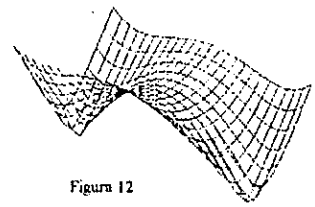


Figura 12

Una cuestión interesante respecto a las dos polinomiales anteriores se referiría a observar como cambian los ceros cuando la polinomial z^2+z-1 cambia a z^2+z+1. Dado que los ceros de una polinomial son funciones continuas de los coeficientes y que el coeficiente a_0 de la polinomial z^2+z+a_0 cambia continuamente de -1 a 1 sobre la recta real, la posición de los puntos bajos sobre la correspondiente superficie modular se mueve a lo largo de una curva continua sobre el plano complejo comenzando y terminando en las posiciones indicadas en las figuras 2 y 3. Como los ceros de z^2+z+a_0 están dados por $\frac{(-1\pm\sqrt{1-4a_0})}{2}$, estos ceros serán reales cuando $a_0 < \frac{1}{4}$. Cuando la superficie modular toma la apariencia de la Figura 13 se presenta un cero doble en $z=-\frac{1}{2}$ de la polinomial $z^2+z+\frac{1}{4}$; los puntos bajos de la superficie modular se agrupan alrededor del cero doble.

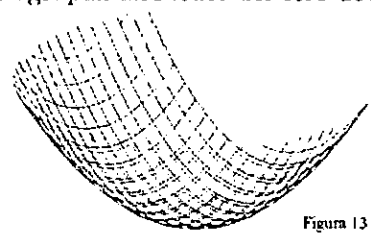


Figura 13

Cuando a_0 crece de $\frac{1}{4}$ a 1, el punto doble se rompe en dos puntos simétricamente localizados con respecto al eje x; no son reales, son complejos conjugados. Para a_0 en $\left(\frac{1}{4},1\right]$, los ceros de z^2+z+a_0 tienen parte real $x=-\frac{1}{2}$ y por tanto están en sobre la línea $x=-\frac{1}{2}$. La figura 14 muestra el recorrido de los ceros de z^2+z+a_0 cuando a_0 se incrementa de -1 a 1.

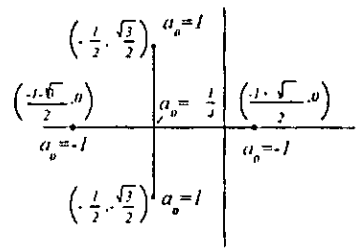


Figura 14

9.2.4 Solución de Ecuaciones en Variable Compleja - Una Aproximación Intuitiva

Ilustraremos a través de los ejemplos una idea central en el análisis complejo: Una ecuación polinomial en variable compleja z es equivalente a un par de ecuaciones en variables reales x y y. Este hecho nos permite visualizar las curvas asociadas a estas ecuaciones en el plano. La solución

$S \neq \emptyset$ es subanillo de $\mathbb{Z}$ si y sólo si existe un único entero $n \geq 0$ tal que $S = n\mathbb{Z}$ con $n\mathbb{Z}$=$\{nx | x \in \mathbb{Z}\}$.

Demostración. (i) Si $S = \{0\}$, $\{0\} = 0\mathbb{Z}$.

(ii) Sea $S \neq \{0\}$. Luego existe $a \neq 0$, $a \in S$. Como S es subanillo, $-a \in S$. Podemos afirmar entonces que $a > 0$ o $-a > 0$ está en S. Construimos el conjunto $T = \{x \in S | x > 0\}$. T es un conjunto no vacío; en virtud del principio de buena ordenación existe $n \in S$, $n = \min(T)$. Dado entonces $m \in S$ podemos aplicar el algoritmo de la división a m y a n. Existen entonces $q, r \in \mathbb{Z}$ únicos tales que $m = nq + r$, con $1 \leq r \leq n - 1$. Pero

$$nq = \underbrace{n + n + n + \cdots + n}_{q \text{ veces}}$$

está en S, de donde, $r = m - nq$ está en S. Por la definición de n la única posibilidad es que $r = 0$ y entonces $m \in n\mathbb{Z}$ lo que implica $S \subseteq n\mathbb{Z}$. De otra parte, como $n \in S$, $nx \in S$,es decir $n\mathbb{Z} \subseteq \mathbb{S}$. En conclusión, $S = n\mathbb{Z}$.

Puntos de discusión

1. Determinar cuáles de los siguientes conjuntos son subanillos del anillo de los números reales.
   i. $S = \{n \in \mathbb{Z} | n = 0 \text{ o } |n| \geq 17\}$.
   ii. $S = \{n + m\sqrt{3} | n, m \in \mathbb{Z}\}$.
   iii. $S = \{\frac{p}{q} | p, q \in \mathbb{Z}, q \neq 0, \text{ y, } q \text{ no es múltiplo de } 5\}$.

2. Considerar el anillo de las matrices 2×2 con elementos reales.
   i. Demostrar que el conjunto $S = \left\{ \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} | a, b \in \mathbb{R} \right\}$es un subanillo.
   ii. Verificar que la matriz
   $$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$
   actúa como identidad a izquierda en S y que todo elemento no idénticamente nulo de S es inversible a izquierda pero no a derecha.
   iii. Identificar por lo menos otros tres subanillos de este anillo. Usted ya ha observado que el anillo de las matrices con elementos reales no es en general un dominio de integridad, ¿Existe algún subanillo de este anillo que sea un dominio?

3. Sea $\mathbb{R}[x]$ el conjunto de todos los polinomios en variable x con coeficientes reales.
   i. Demostrar que con las operaciones usuales de adición y multiplicación de polinomios, $\mathbb{R}[x]$ es un anillo conmutativo con elemento identidad. ¿Es $\mathbb{R}[x]$ un anillo de división?
   ii. Identificar por lo menos tres subanillos de $\mathbb{R}[x]$.

4. Sea E un conjunto. Usted ya demostró en otro de los puntos de discusión que $(P(E), \Delta, \cap)$ es un anillo conmutativo. Explorar ejemplos para verificar y luego demostrar las siguientes afirmaciones.

6.2 Subanillos

En el análisis de los grupos, hacíamos referencia a los subconjuntos "especiales" que conservan la estructura, los subgrupos. Aquí también nos interesa esta caracterización; nos referiremos a los *subanillos*.

Definición 6.7 *Sea $(A,+,\cdot)$ un anillo. Por un subanillo de A entendemos cualquier subconjunto S de A tal que las operaciones de A restringidas a S hacen de S un anillo.*

Por ejemplo, los números enteros constituyen un subanillo de los números racionales y estos últimos forman un subanillo de los reales. $4\mathbb{Z}$ es un subanillo de $\mathbb{Z}$. Las funciones definidas sobre $\mathbb{R}$, a valor real y diferenciables, forman un subanillo del anillo de las funciones continuas. ¡Verificar estas afirmaciones!

Nota. Una pregunta interesante que surge de inmediato es ésta. ¿Heredan los subanillos de un anillo, la estructura mas completa de éste? Si el anillo es conmutativo, es claro que el subanillo también lo es. Pero si posee por ejemplo elemento identidad o es un dominio de integridad, esto puede o no ser cierto para el subanillo. Basta que usted analice subanillos de $\mathbb{Z}$ como, $2\mathbb{Z}$, $3\mathbb{Z}$, $4\mathbb{Z}$.

Definición 6.8 *Si S, un subconjunto de un dominio de integridad A, es tal que, las operaciones de A restringidas a S hacen de S un dominio de integridad se dice que S es un subdominio.*

Así como en los grupos, existen en los anillos criterios para determinar si un subconjunto de un anillo es un subanillo o si un subconjunto de un dominio es un subdominio y para caracterizar completamente los subanillos de nuestro anillo modelo de los números enteros. Se expresan en los siguientes teoremas.

Teorema 6.7 *Un subconjunto S de un anillo $(A,+,\cdot)$ es un subanillo si y solamente si $\forall x,y \in S$, $x+y$, $x\cdot y$ y $-y$ están en S. Un subconjunto S de un dominio de integridad A es un subdominio si y sólo si el elemento identidad de A para la multiplicación está en S.*

Demostración.

La demostración de la primera parte es completamente análoga a la que desarrollamos para subgrupos. ¡Reconstruirla!

Si ahora suponemos que S es un subdominio, se tiene que es un dominio y tiene elemento identidad que debe ser la identidad de A, pues si A no tiene divisores de cero la única solución no nula de la ecuación $x^2 = 1$ es $x = 1$.

Si se tiene que $1 \in S$ y suponemos que existe en S un divisor propio de cero, éste sería divisor propio de cero en A y esto contradice que A sea dominio de integridad. Se concluye entonces que S es dominio de integridad.

En la caracterización de los subanillos de $(\mathbb{Z},+,\cdot)$, se usan de nuevo elementos que provienen de la teoría de números y los argumentos se asemejan, como es natural, a los utilizados para caracterizar los subgrupos.

Teorema 6.8 *Dado $(\mathbb{Z},+,\cdot)$ el anillo de los números enteros. $S \subset \mathbb{Z}$,*

6.1.1 Característica de un anillo

En los grupos la noción de orden constituyó punto importante en nuestro análisis de los nexos con la teoría de números; en el caso de los anillos la noción de característica tiene elementos similares.

Definición 6.6 *Sea A un anillo y consideremos el conjunto*

$$S = \{n \in \mathbb{Z} | na = 0, \forall a \in A\}.$$

Se define característica del anillo A, notada $char(A)$ al entero dado por

i. El mínimo de S si $S \neq \emptyset$.

ii. 0 si $S = \emptyset$.

Los anillos $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$ tienen característica cero, el anillo $\mathbb{Z}_n$ tiene característica n. Si consideramos un conjunto S no vacío y el anillo $(P(S), \Delta, \cap)$, $\forall B \subset S$, $B\Delta B = \emptyset$, donde $\emptyset$ es el módulo para Δ, concluimos entonces que $Char(P(S)) = 2$.

En los anillos con elemento identidad la característica puede ser identificada de manera mas sencilla usando criterios de minimalidad en el conjunto de los naturales.

Teorema 6.5 *Sea A un anillo con elemento identidad 1. A tiene característica $m > 0$ si y sólo si $m = \min(T)$ donde $T = \{n \in \mathbb{N} | n \cdot 1 = 0, 1$ y $0 \in A\}$ es no vacío.*

Demostración. Sea $S = \{n \in \mathbb{N} | n \cdot a = 0, \forall a \in A\}$ y supongamos que la característica de A es $m > 0$. Por definición se tiene que

$$m = \min\{n \in \mathbb{N} | na = 0, \forall a \in A\},$$

y como S es no vacío, $T = \{n \in \mathbb{N} | n \cdot 1 = 0\}$ es no vacío. Por el principio de buena ordenación, existe $m_0 = \min(T)$. Veamos que $m_0 = m$. Como la característica de A, $m > 0$, es tal que $m \cdot 1 = 0$ entonces $m \in T$. Luego $m_0 \cdot a = m_0(1 \cdot a) = (m_0 \cdot 1) \cdot a = 0 \cdot a = 0$, para todo $a \in A$. Se sigue que $m = m_0 = \min(T)$. La demostración en el otro sentido es similar; usa tan sólo argumentos de minimalidad.

Nota. Obsérvese que en el anillo $(\mathbb{Z}_3, +, \cdot)$, el orden de 1 como elemento del grupo aditivo $(\mathbb{Z}_3, +)$ es 3, similarmente $o(2) = 3$ y $char(\mathbb{Z}_3^*) = 3$.

Corolario 6.3 *Sea D un dominio de integridad,*

i. Si $Char(D) = n > 0$, entonces si $a \in D^*$, el orden de a como elemento del grupo adtivo es n.

ii. Si $Char(D) = 0$ y $a \in D^*$ entonces a como elemento del grupo aditivo $(D, +)$ tiene orden infinito.

Demostración.

(i) Supongamos que $Char(D) = n > 0$. Dado $a \in D^*$, $n \cdot a = 0$ el neutro de D. Entonces el conjunto $S = \{t \in \mathbb{N} | t \cdot 1 = 0\}$ es un subconjunto no vacío de los naturales y por el principio de buena ordenación tiene un elemento

mínimo que es precisamente $o(a)$. $o(a) = \min(S) = m$, m entero positivo. Pero como $n \in S$ entonces $o(a) \geq n = Char(D)$.

De otra parte como $n = \min\{s \in \mathbb{N} a \cdot 1 = 0\}$, dado que $o(a) \cdot a = 0$,

$$0 = o(a) \cdot a = o(a) \cdot (1 \cdot a) = (o(a) \cdot 1) \cdot a.$$

Ya que $a \neq 0$ y D es dominio de integridad, se tiene que $o(a) \cdot 1 = 0$ y por tanto $Char(D) \geq o(a)$, por definición. Luego $n \geq m$. Concluimos que $o(a) = n = m$.

(ii) Sea $Char(D) = 0$ y $a \in D^*$. El conjunto $\{t \in \mathbb{N} | n \cdot 1 = 0\}$ es vacío. Si $n \in \mathbb{N}$ y $a \in D^*$, $n \cdot a = (n \cdot 1) \cdot a \neq 0$, por ser $(n \cdot 1) \neq 0$ y $a \neq 0$ en un dominio de integridad. Entonces el conjunto $\{s \in \mathbb{N} | s \cdot a = 0\}$ es vacío si $a \in D^*$, de donde $o(a) = \infty$ y D es un conjunto infinito.

Punto de discusión

Si D es un dominio de integridad. Demostrar que $Char(D) = n > 0$ si y sólo si $sn \cdot a = 0$ para algún $a \in D^*$ y para algún $m \in \mathbb{N}$.

Teorema 6.6 *Sea D un dominio de integridad. Entonces la característica de D es cero, o es un número primo.*

Demostración. Supongamos que $Char(D) = n > 0$, si $n = st$, con $1 < s < n$ y $1 < t < n$. Dado que $0 = n \cdot 1 = (st) \cdot 1 = (s \cdot 1)(t \cdot 1)$, por ser D un dominio de integridad concluiríamos que $s \cdot 1 = 0$ o $t \cdot 1 = 0$, pero esto contradice la elección de n como el mínimo de los enteros positivos tal que $n \cdot 1 = 0$. Se sigue que n es primo.

Corolario 6.4 *Si D es un dominio de integridad finito entonces $Char(D) = p$, p primo.*

Nota. Explorar los puntos de discusión siguientes una vez analice la sección de homomorfismos.

Puntos de discusión

1. Sea A un anillo con identidad 1. Demostrar que la función

   $$f : \mathbb{Z} \longrightarrow A,$$

   definida por $f(n) = n \cdot 1$ es un homorfismo. ¿Es sobre? ¿Es uno a uno?

2. Algunos textos de Algebra Moderna entre ellos el de Seth Warner, definen la característica de un anillo con identidad como "El número natural p tal que $\langle p \rangle$ es el núcleo del homorfismo f de $\mathbb{Z}$ en A definido por $f : n \longrightarrow n \cdot 1$".

   Contrastar la definición anterior con la presentada en nuestra discusión y usarla para demostrar que si D es un dominio de integridad con característica un primo p entonces

   i. La función $g : x \longrightarrow nx$, donde $n \in \mathbb{Z}$ es un homorfismo no trivial del anillo D en sí mismo si y sólo si $n \equiv 1 (mod p)$.

   ii. La función $h : x \longrightarrow x^p$ es un homomorfismo uno a uno de D en D.

   iii. Para todo $m \in \mathbb{N}$, la función $h_m : x \longrightarrow x^{p^m}$ es un homomorfismo uno a uno de D en D.

de una ecuación polinomial en variable compleja se traduce, entonces, en el problema de resolver un sistema de ecuaciones polinomiales en variable real, esto es, identificar la intersección de dos curvas en el plano (la demostración del teorema fundamental dependerá precisamente de garantizar que ciertas curvas se intersectan).

Ejemplos

1. Consideremos la ecuación lineal

$$2z + 5 = 0;$$

con z en los complejos. Como $z \in \mathbb{C}$, $z = x + iy$. De donde

$$2(x + iy) + 5 = 0$$

$$(2x + 5) + i(2y) = 0 = 0 + 0i$$

Por tanto $2x + 5 = 0$ y $2y = 0$, sistema de ecuaciones simultáneas que tiene como única solución $x = -\frac{5}{2}$; $y = 0$. Gráficamente la intersección de la recta $x = -\frac{5}{2}$ con el eje x (Figura 15).

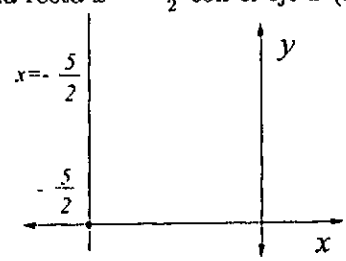


Figura 15

2. ¿Qué ocurre si los coeficientes de la lineal son también complejos? Por ejemplo

$$(1 + i)z + (2 - 3i) = 0$$

De manera similar, haciendo $z = x + iy$, obtenemos el sistema

$$\begin{aligned} x - y + 2 &= 0 \\ x + y - 3 &= 0, \end{aligned}$$

cuya solución gráfica se observa en la figura 16; corresponde al punto $(\frac{1}{2}, \frac{5}{2})$.

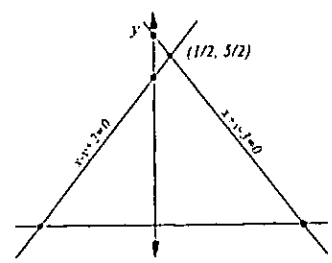


Figura 16

3. Si r y s son complejos, $r \neq 0$, ¿qué se puede decir de la solución de la ecuación lineal

$$rz + s = 0?$$

Tenemos ahora $r = a + bi$, $s = c + di$ y $z = x + yi$. Sustituyendo en la ecuación obtenemos

$$(ax - by + c) + i(ay + bx + d),$$

que nos lleva al sistema de ecuaciones simultáneas

$$\begin{aligned} ax - by + c &= 0 \\ bx + ay + d &= 0 \end{aligned}$$

que tiene solución única si su determinante es distinto de cero. Claramente esto ocurre en este caso pues $\Delta = a^2 + b^2 \neq 0$, ya que $r \neq 0$. Hay entonces un único punto de intersección de las rectas que puede ser determinado gráficamente y nos permite identificar la solución de la ecuación lineal en variable compleja z con coeficientes en los complejos.

4. Realizaremos ahora un análisis similar con ecuaciones cuadráticas en variable compleja z. Analicemos, por ejemplo, la solución de

$$z^2 + z + 1 = 0$$

De nuevo si hacemos $z = x + iy$, tenemos

$$\begin{aligned} (x + iy)^2 + (x + iy) + 1 &= 0 \\ (x^2 + x - y^2 + 1) + i(2xy + y) &= 0, \end{aligned}$$

de donde,

$$\begin{aligned} x^2 + x - y^2 + 1 &= 0 \\ y(2x + 1) &= 0. \end{aligned}$$

9.2.5 Sistema de ecuaciones simultáneas en variables y .

La primera ecuación puede ser reescrita como

(a)
$$(x + \frac{1}{2})^2 - y^2 = -\frac{3}{4}$$

Gráficamente el problema consiste en encontrar la intersección de esta hipérbola con

(b)
$$y(2x + 1) = 0,$$

que representa un par de rectas perpendiculares que se intersectan en $(-\frac{1}{2}, 0)$.

Combinando las dos ecuaciones (es claro que $y \neq 0$), tenemos $x = -\frac{1}{2}$; sustituyendo en (1) obtenemos

$$y^2 = \frac{3}{4}; y = \pm\sqrt{\frac{3}{4}}.$$

Las dos curvas se intersectan en los puntos $(-\frac{1}{2}, \sqrt{\frac{3}{4}})$ y $(-\frac{1}{2}, -\sqrt{\frac{3}{4}})$. En la figura 17 se observan estos interceptos. Las soluciones de la cuadrática inicial son entonces

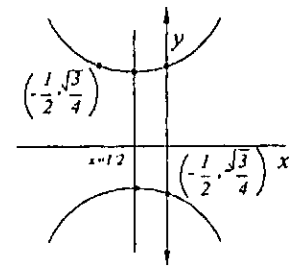


Figura 17

$z = -\frac{1}{2} + \sqrt{\frac{3}{4}}i$ y $z = -\frac{1}{2} - \sqrt{\frac{3}{4}}i$, dos raíces complejas conjugadas.

Para la cuadrática $z^2 + 3z + 1 = 0$, obtenemos el sistema

$$\begin{aligned} (x + \frac{3}{2})^2 - y^2 &= \frac{5}{4} \\ y(2x + 3) &= 0, \end{aligned}$$

cuya solución se observa en la Fig(18).

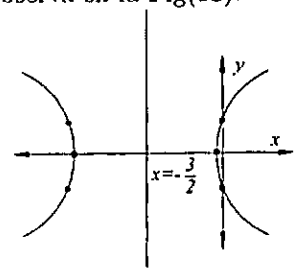


Figura 18

5. Similarmente para caracterizar completamente la solución de la cuadrática general

$$az^2 + bz + c = 0.$$

Se sustituye z por $x + iy$ aparece entonces el sistema

$$\begin{aligned} \left(x + \frac{b}{2a}\right)^2 - y^2 &= \left(\frac{1}{2a}\right)^2 (b^2 - 4ac) \\ y(2ax + b) &= 0, \end{aligned}$$

cuya solución se puede discutir a partir del análisis del término $b^2 - 4ac$.

Nota

Observense (19), (20) y (21); aparecen allí las curvas que resultan al análizar las ecuaciones polinómicas $z^2 + z + 1 = 0$, $z^2 + 3z + 1 = 0$ y $z^2 + 2z + 1 = 0$. En los tres casos se ha construido una circunferencia de centro en el origen, con la condición de que en su interior contenga todas las soluciones de los sistemas de ecuaciones simultáneas correspondientes.

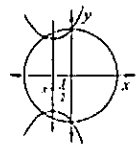


Figura 19

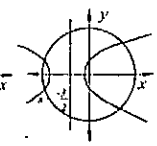


Figura 20

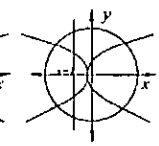


Figura 21

Estudiemos para concluir este aparte. Ecuaciones polinomiales cúbicas en variable compleja z. Miremos por ejemplo

$$z^3 - 3z + 2 = 0$$

De nuevo al hacer $z = x + iy$ obtenemos el sistema

$$\begin{aligned} x^3 - 3xy^2 - 3x + 2 &= 0 \\ y(3x^2 - y^2 - 3) &= 0. \end{aligned}$$

La gráfica de la segunda ecuación corresponde al eje x junto con la hipérbola de centro en $(0,0)$ y asíntotas $y = \sqrt{3}x$ y $y = -\sqrt{3}x$.

Nótese que la primera ecuación puede ser escrita

$$y^2 = \frac{1}{3}\left(x^2 - 3 + \frac{2}{x}\right).$$

Esto significa que el lugar geométrico de esta ecuación es asintótico al par de líneas rectas cuya ecuación está dada por $3y^2 = x^2$; además la curva definida por ella es asintótica al eje y. ¡Justifique!

De la segunda ecuación $3x^2 - y^2 - 3 = 0$; esto es $3x^2 - 3 = y^2$. Sustituyendo en la primera

$$\begin{aligned} 3x^2 - 3 &= \frac{1}{3}\left(x^2 - 3 + \frac{2}{x}\right) \\ 8x^2 - \frac{2}{x} - 6 &= 0 \\ 8x^3 - 6x - 2 &= 0. \end{aligned}$$

Puntos de discusión

1. La ecuación cúbica $z^3 - 3z + 2 = 0$ tiene entonces dos raíces reales, una de multiplicidad dos. Explicar ¿por qué¿. Para ello observe la figura 22.

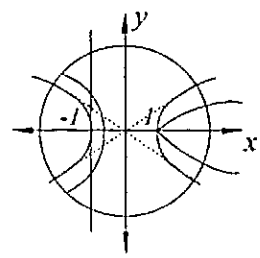


Figura 22

2. Usar un procedimiento similar para analizar las ecuaciones

$$z^3 + pz + q = 0$$

"D'Alembert dedicó mucho de su tiempo y esfuerzo a demostrar el teorema conjeturado por Girard y conocido hoy como el Teorema Fundamental del Algebra: Toda ecuación polinomial $f(x) = 0$, que tenga coeficientes complejos y grado $n \geq 1$, tiene al menos una raíz compleja [...]. El enunciado dado por Gauss al referirse al Teorema Fundamental es esencialmente la proposición conocida en Francia como el Teorema de D'Alembert; Gauss prueba que todas las demostraciones anteriores incluidas algunas de Euler y Lagrange son inadecuadas." [Boyer Carl B. *A history of mathematics,* 1968].

De estas citas se desprende la concepción de una historia marcada por tres momentos importantes:

- Las primeras formulaciones del teorema sin demostración, aparecen en el siglo XVII (Girard).

- Las primeras tentativas de demostración, en el siglo XVIII (D'Alembert, Euler).

- La primera demostración rigurosa, en el siglo XIX (Gauss).

Es necesario diferenciar en el análisis de este desarrollo dos resultados (que en ocasiones se confunden) en el marco de la teoría de ecuaciones algebraicas: el Teorema Fundamental (conocido como Teorema de D'Alembert o de D'Alembert-Gauss) y el llamado Teorema de Factorización Lineal (conocido como Teorema de Kronecker). El Teorema fundamental del álgebra (TFA) puede ser expresado en una de las siguientes formas equivalentes.

(a) Todo polinomio de grado $n \geq 1$ con coeficientes complejos tiene al menos una raíz compleja.

(b) Todo polinomio de grado $n \geq 1$ con coeficientes complejos se descompone en un producto de n factores lineales con coeficientes complejos y admite n raíces complejas (distintas o repetidas)

(a') y (b') las mismas conclusiones que en (a) y (b) a partir de un polinomio con coeficientes reales.

(c) Todo polinomio de grado $n > 1$ con coeficientes reales puede ser descompuesto en un producto de factores con coeficientes reales de primero o de segundo grado.

Es importante aclarar aquí que el Teorema de Factorización Lineal (TFL) afirma: "Si $p(x)$ es un polinomio de grado $n \geq 1$ con coeficientes en un cuerpo conmutativo K, existe un cuerpo L (cuerpo de descomposición) que contiene los coeficientes de $p(x)$ y en el que $p(x)$ se descompone en producto de n factores lineales (esto es tiene n raíces distintas o múltiples, todas en L)". Es decir, para el caso de los polinomios con coeficientes reales o complejos (que nos concierne) los enunciados se confunden; sin embargo en el análisis histórico es importante establecer que el teorema de factorización lineal garantiza para un polinomio cualquiera la existencia de factorización y de raíces, independientemente de la forma de estas últimas, mientras que el TFA corresponde a un resultado preciso sobre la naturaleza de las raíces (o de la factorización), en una estructura particular, el cuerpo de los complejos o el de los reales. Esta distinción de los teoremas en el plano histórico permite comprender la historia del TFA y más generalmente la historia de la teoría de ecuaciones algebraicas.

con p y q reales no nulos y

$$z^4 + 2z^2 + z + 1 = 0.$$

Un punto importante a resaltar aquí y que seguramente se ha apreciado más claramente en el análisis de los ejemplos es el siguiente. Si

$$f(z) = z^n + c_1 z^{n-1} + \cdots + c_{n-1} + c_n = 0$$

es una polinomial en variable compleja z con coeficientes $c_1, c_2, \cdots, c_n$ reales o complejos, sustituimos en $f(z)$ los valores $z = x + iy$, $c_1 = a_1 + b_1 i$, $c_2 = a_2 + b_2 i$, $\cdots$, $c_n = a_n + b_n i$, y realizamos todos los cálculos encontramos que

$$f(z) = X + iY,$$

donde $X = P(x, y)$, $Y = Q(x, y)$ son polinomios de grado n en x y y con coeficientes reales expresados en términos de a_i y b_i. Lo anterior nos permite afirmar que, si las variables x y y cambian continuamente, la polinomial $f(z)$ también cambia continuamente.

9.3 Teorema Fundamental del Algebra

La complejidad del problema de la solubilidad por radicales de una ecuación polinomial motivó a los matemáticos a trabajar la teoría de ecuaciones en tres direcciones completamente distintas.

- Garantizar la existencia de por lo menos una raíz.

- Caracterizar las raíces (sin identificarlas explicitamente) usando relaciones entre raíces y coeficientes.

- Encontrar métodos para determinar aproximaciones de las raíces de una polinomial.

En lo relativo a las dos últimas direcciones presentamos ya una discusión en capítulos anteriores y se apreciaba allí como la determinación de aproximaciones se basa en argumentos del cálculo (método de aproximación de Newton, método de Horner, Teorema de Taylor), en esta sección nos centraremos en el primer punto, específicamente, en el llamado "*Teorema Fundamental del Algebra*" no sólo porque se "menciona" y se usa en el trabajo con polinomios en los distintos niveles, sin analizar su significado, ni explorar realmente sus implicaciones en la fundamentación de la teoría de ecuaciones, sino porque, desde sus orígenes estuvo ligado a problemas del cálculo y requiere para su demostración de argumentos analíticos (como lo comentamos en el inicio de la sección anterior).

9.3.1 Sobre la historia del teorema fundamental del algebra

"La disertación de Gauss contiene la primera demostración rigurosa del llamado "Teorema Fundamental de Algebra" (TFA), el cual establece que toda ecuación algebraica con coeficientes reales tiene al menos una raíz y por tanto tiene n raíces. El teorema se remonta a Albert Girard [...]; d'Alembert trató de dar una demostración en 1746." [Struik Dirk J. *A concise history of mathematics*, 1967].

9.3.2 Teoría de ecuaciones en el siglo XVII

En 1629, Albert Girard enuncia en su *Invention nouvelle en l'algébre* un teorema en el que afirma que una ecuación polinomial de grado n tiene exactamente n raíces, si uno cuenta las raíces imposibles, esto es las complejas, y si uno considera las raíces repetidas. Girard fundamenta su teoría como lo enuncia en la segunda parte de este teorema, sobre el sistema de relaciones entre los coeficientes del polinomio y las raíces de la ecuación, y usa de manera sistemática esta relación para calcular las raíces. El enunciado general de Girard sobre el número de raíces y el grado de una ecuación algebraica que supone la aceptación de las no reales representa un avance comparado con afirmaciones anteriores, que se referían al grado como una cota superior de este número. Es posible afirmar entonces, que el teorema de Girard constituye una conjetura del TFA; aunque él no afirma que todas las raíces sean complejas y no presenta, un enunciado general sobre la naturaleza de las raíces no reales. Para algunos historiadores el enunciado de Girard puede mas bien ser interpretado como una primera forma del Teorema de Factorización Lineal que sirve de fundamento a la naciente teoria de ecuaciones algebraicas.

En 1637, en el libro III de *La Géométrie,* René Descartes dice.

Una ecuación puede tener tantas raíces distintas como el número de dimensiones (grado) de la variable.

El dice "puede tener" porque considera las raíces negativas como raíces falsas; incluye posteriormente raíces negativas e imaginarias con el propósito de contar, y concluye que hay tantas como el grado. Inicialmente afirma que el número de raíces es a lo mas igual al grado, justifica esta afirmación construyendo ecuaciones de tercero y cuarto grado por multiplicación de ciertos factores lineales. Su argumento mas general consiste en decir que, puesto que uno forma una ecuación de grado n haciendo el producto de n ecuaciones simples, recíprocamente toda ecuación de grado n puede ser descompuesta en n factores lineales. Las raíces imaginarias necesarias para asegurar la generalidad de esta factorización aparecen como adjunciones formales, sin precisar su naturaleza. Se está ocupando Descartes realmente del TFL, y no del TFA. La mayoría de los matemáticos del siglo XVII presentan planteamientos similares a los de Descartes, Wallis (1685), Prestel (1689), Ozaman (1702), Newton (1707), Reyneaw (1708). Nuevamente sus enunciados constituyen fundamentos de la teoría de ecuaciones pero en la dirección del Teorema de Factorización Lineal. Es interesante anotar que en la mayoría de tratados de la época no se intenta demostrar la descomposición de un polinomio en factores lineales; se asume ésta como un principio legítimo, base de la teoría de ecuaciones.

9.3.3 Leibniz, la integración de diferenciales racionales y el problema del Teorema Fundamental del Algebra.

En 1702 se produce un evento de especial importancia para la discusión acerca de la teoría de ecuaciones, pero en un dominio nuevo de la matemática de reciente creación: el cálculo infinitesimal. Leibniz y Bernoulli publican sendas memorias sobre cálculo integral; las dos memorias tienen un punto en común, usan el método de descomposición de fracciones racionales en

suma de elementos simples. La base de esta descomposición reposa en ambos casos sobre la expresión del denominador de la fracción racional en un producto de factores lineales. Después de reducir por división el grado del numerador (grado inferior al del denominador), Bernoulli asume una factorización lineal para el denominador, descompone en fracciones simples y dice que la determinación de los numeradores de estas fracciones reposa sobre el método de los coeficientes indeterminados. Leibniz escribe un enunciado similar pero hace anotaciones en las que manifiesta que los conocimientos algebraicos no permiten calcular todas las raíces exactamente; se apoya pues en la existencia *a priori* de factorización lineal (los dos suponen la validez del TFL). Leibniz en su memoria de 1702 supone que las raíces del denominador de la fracción son todas distintas, pero al año siguiente considera el caso de raíces múltiples, mientras que el análisis de Bernoulli corresponde solamente a raíces distintas.

Punto de Discusión
El problema de integrar fracciones racionales depende pues directamente de la teoría de ecuaciones algebraicas y mas precisamente del TFA en su forma "Todo polinomio de grado $n \geq 1$ con coeficientes reales puede ser descompuesto en un producto de factores reales de primero y segundo grado". Sobre esta cuestión Leibniz hace explícita una respuesta negativa; él no piensa que uno pueda obtener una tal factorización real para todo polinomio y justifica su afirmación con un famoso "contraejemplo". Considera las descomposiciones sucesivas del polinomio $x^4 + a^4$

$$(xx - aa\sqrt{-1})(xx + aa\sqrt{-1})$$

$$(x + a\sqrt{\sqrt{-1}})(x - a\sqrt{\sqrt{-1}})(x + a\sqrt{-\sqrt{-1}})(x - a\sqrt{-\sqrt{-1}})$$

y afirma, a partir de esta descomposición, que si se combinan los cuatro factores, no es posible que el producto sea un trinomio real. Discutir el por qué es incorrecta la argumentación de Leibniz.

Las memorias de Leibniz y de Bernoulli constituyen una etapa importante en la historia del cálculo integral, pero el aspecto de interés en nuestro análisis presentado por Leibniz es que enuncia de manera precisa el TFA y refuta su validez; esto le concede un papel especial en la historiografía de la teoría de ecuaciones algebraicas. Esto se da de una parte porque por primera vez se tiene conocimiento del TFA de una forma general y precisa al preguntar "¿Todo polinomio con coeficientes reales puede ser descompuesto en factores reales de primero y de segundo grado?" La otra parte, la respuesta negativa, aporta una nueva interrogación la dualidad entre el TFL que Leibniz admite como un principio indiscutible y el TFA que él refuta. Sobre la primera reposa la descomposición de las fracciones racionales en elementos simples con denominador lineal, la segunda contiene la posibilidad de una descomposición real en denominadores lineales y cuadráticos. La refutación del TFA que hace Leibniz es claramente una concepción de que los imaginarios se clasifican en diferentes clases.

Otro personaje que jugó un papel importante para el problema que nos ocupa (en la primera parte del siglo XVIII) fue el matemático inglés Roger Cotes quien, en su obra *Harmonia Mensurarum* publica un método para resolver integrales de la forma

$$\int \frac{dx}{x^4 + a^4}, \int \frac{dx}{x^8 + a^8}.$$

En el hace una descomposición en factores reales lineales o cuadráticos de polinomios de la forma $a^n \pm x^n$ (estableciendo relaciones geométricas en un círculo de radio a dividido en n partes iguales); una vez descompuesto el denominador de la correspondiente fracción en factores reales irreducibles de primero y segundo grado, expresa la integral con ayuda de funciones algebraicas, logarítmicas y trigonométricas.

Puntos de Investigación.

1. ¿A qué tipo de relaciones geométricas se hace referencia en la descomposición de polinomios de la forma $a^n \pm x^n$?
2. ¿Por qué una vez descompuesto el denominador de la fracción racional en factores irreducibles de primero y segundo grado, es siempre posible determinar la integral de la función racional?

En 1719 Jean Bernoulli hace la descomposición de $x^4 + a^4$ en dos trinomios reales $x^2 + ax\sqrt{2} + a^2$ y $x^2 - ax\sqrt{2} + a^2$ y de allí calcula la integral $\int \frac{dx}{x^4+a^4}$. En 1730 Abraham de Moivre no sólo expone en su obra los resultados de Cotes sino que descompone en factores cuadráticos irreducibles polinomios del tipo $A + Bx^m + Cx^{2m}$, $m \in \mathbb{N}$; pero de Moivre no afirma que tal descomposición sea siempre posible, es mas, en su trabajo hace un análisis de la descomposición del polinomio

$$z^4 + pz^3 + qz^2 + pz + 1$$

en un producto de dos polinomios cuadráticos reales. Considera que al abordar la desco mposición en la forma $(z^2 + az + 1)(z^2 + bz + 1)$, los coeficientes a y b obtenidos por el método de los coeficientes indeterminados podrían ser imaginarios; es esto lo que lo conduce a concluir que la factorización no es posible.

Puntos de Discusión
- Analizar la descomposición del polinomio $z^4+pz^3+qz^2+pz+1$ en factores cuadráticos.

9.3.4 El papel de Euler

Hacia 1740 cuando no estaba aun resuelto completamente el problema de la integral de fracciones racionales, surge en el cálculo otro problema que será igualmente soluble con el TFA y que se constituirá en objeto central de estudio en esta década: la integración de ecuaciones diferenciales lineales homogéneas de orden n con coeficientes constantes. El problema es expuesto de manera general y resuelto por Euler en su memoria "De integratione aequationum differentialium altiorum graduum". En la búsqueda de soluciones de la forma $y = e^{px}$ (p contante) para la ecuación diferencial

$$0 = Ay + B\frac{dy}{dx} + C\frac{d^2y}{dx^2} + \cdots + N\frac{d^ny}{dx^n},$$

él encuentra la ecuación algebraica

$$0 = A + Bp + Cp^2 + \cdots + Np^n.$$

Supone sistemáticamente la descomposición del polinomio en factores lineales o cuadráticos con coeficientes reales; discute casos en que la ecuación

tiene raíces simples o múltiples, y en que tiene raíces reales o imaginarias; para cada uno muestra n integrales particulares independientes de la ecuación original, las expresa en términos finitos reales y deduce la integral general como una combinación lineal. Este trabajo constituye una etapa importante en la historia de la teoría de ecuaciones diferenciales; sin embargo nuevamente este resultado general del cálculo integral depende de la veracidad del TFA. Euler utiliza sistemáticamente en sus memorias el hecho de que las raíces complejas pueden ser asociadas por pares, con suma y producto reales. En una carta que envía a Bernoulli en 1739 anuncia su descubrimiento sobre las ecuaciones diferenciales y manifiesta su convicción en cuanto a la generalidad de la descomposición de todo polinomio en factores reales de primero y segundo grado, citando un ejemplo.

"La expresión $1 - ap + bp^2 - cp^3 + dp^4 - ep^5 + etc = 0$ si fuese posible de factorizar en factores reales simples de la forma $1 - \alpha p$, se factoriza y se resuelve; si éste no es el caso se factoriza en factores de la forma $1 - \alpha p + \beta pp$, cuya resolución es siempre posible; no aparecen expresiones de grado superior, esto es, siempre se factoriza en expresiones simples $1 - \alpha p$ o en cuadráticas $1 - \alpha p + \beta pp$, siempre reales."

Este constituye para los historiadores el primer enunciado general y positivo del TFA; en un largo intercambio de cartas entre Euler y Jean Bernoulli, Euler enuncia reiteradamente el teorema, pero reconoce que no tiene una demostración general para el. En 1742 le comunica sinembargo que ha desarrollado una demostración rigurosa para polinomios de grado menor que cuatro, insiste nuevamente en la importancia excepcional del resultado para resolver los dos problemas del cálculo antes mencionados, y lo exhorta a contribuir a la investigación aportando una respuesta positiva o negativa. La gran memoria de Euler sobre el TFA *Recherches sur les racines imaginaires des équations* aparece en Francia en 1751. En la primera parte (como lo comentamos en la sección anterior) discute directamente la descomposición de polinomios en factores reales. Su siguiente paso consiste en mostrar que toda raíz de una ecuación algebraica es de la forma $M + N\sqrt{-1}$ (M y N reales), para deducir la factorización real. Demuestra Euler la estabilidad de los símbolos $M + N\sqrt{-1}$ bajo las operaciones algebraicas fundamentales y de manera análoga deduce que todas las raíces de una ecuación son de esta forma.

Punto de Discusión
Discutir el significado que tiene para la caracterización de la estructura de los números complejos la estabilidad de éstos respecto a las operaciones algebraicas fundamentales.

Es importante comentar que la demostración de Euler inaugura la tradición de las llamadas demostraciones 'algebraicas'; la parte analítica mínima está aislada de la parte algebraica. Esta última consiste en reducir el problema de la factorización real a "garantizar" (mostrar la existencia) de una solución real de una ecuación, de un tipo tratado en la parte analítica. En sus argumentos Euler usa tres resultados: el Teorema de Factorización Lineal (sin demostración puesto que lo considera como una verdad evidente); el número de raíces de una ecuación igual al grado de la ecuación, que deduce del anterior acudiendo a la unicidad de la descomposición; y las relaciones entre raíces y coeficientes, deducidas también del primero por un cálculo como si se trataran de números ordinarios. Todo este conjunto de propiedades generales constituye la base algebraica indiscutible del TFA y

existencia. Inician comparando el enunciado de Euler y D'Alembert para argumentar que en el de Euler hay una suposición *a priori* de la existencia de las raíces pero que esto no ocurre en el caso de D'Alembert. En lo relativo a las demostraciones, la demostración de Euler es algebraica, supone la validez del TFL, esto es la existencia *a priori* de raíces, mientras que D'Alembert usa un método analítico o mas bien analítico-geométrico para tratar de demostrar directamente la existencia de por lo menos una raíz. Para ellos la demostración de D'Alembert contiene lagunas lógicas mas no errores de principio. El aporte de D'Alembert no se reduce a haber hecho pública la primera tentativa seria de demostrar el TFA; mas bien reside en el estatus que le da al teorema en la teoría de ecuaciones algebraicas, pues a nivel de estructura reorganiza y unifica la teoría sobre la base del teorema; se constituye éste en fundamento riguroso de la teoría. De otra parte, el uso de argumentos analíticos en la demostración, en contraste con los netamente algebraicos de la época, abre la puerta que permite salir del círculo vicioso en que se encontraba la fundamentación de la teoría de ecuaciones algebraicas.

Nota. Nótese que para el caso de ecuaciones cuadráticas, cúbicas y cuárticas la existencia no constituye un problema, todas ellas son solubles por radicales (existen fórmulas que permiten calcular sus raíces), pero, para grados superiores no es posible garantizar existencia por el camino de exhibir una solución.

9.4 La teoría moderna: el papel de Gauss.

Carl Friedrich Gauss consideró el teorema tan importante que dio de él cuatro demostraciones. (La primera publicada en 1799, la segunda y tercera en 1816 y la cuarta en 1849.)

La primera fue presentada por Gauss en su tesis doctoral en 1799; en ella no calcula una raíz, sino demuestra su existencia. La idea consiste en que las raíces complejas $a+bi$ de $P(x+iy)=0$ corresponden a puntos del plano y si$P(x+iy)=u(x,y)+iv(x,y)$, entonces (a,b) debe ser intersección de las curvas $u=0$ y $v=0$. Por un estudio cualitativo de las curvas él prueba que existe un arco de una que conecta los puntos de dos regiones distintas separadas por la otra. Entonces la curva $u=0$ debe cortar a la curva $v=0$. El argumento es supremamente original; sin embargo depende de las gráficas de estas curvas, y puede resultar a veces complicado demostrar que ellas deberían cruzarse. En este mismo trabajo Gauss demostró que una polinomial de grado n puede ser expresada como producto de factores lineales y cuadráticas con coeficientes reales.

Como se dijo, Gauss da tres demostraciones mas de este teorema. En la segunda demostración, deja de lado los argumentos geométricos y prueba que el producto de las diferencias de cada dos raíces (conocido hoy como el *Discriminante de Sylvester*) puede ser expresado como una combinación lineal de polinomiales y sus derivadas. Por tanto, una condición necesaria y suficiente que la polinomial y su derivada tengan raíces comunes es que el discriminante se anule. Sin embargo, esta segunda demostración asume que una polinomial no puede cambiar de signo en dos valores distintos de x sin que se anule entre ellos. La demostración de este hecho estaba fuera del alcance de rigor de esa época. La tercera demostración usa lo que nosotros conocemos hoy como el Teorema Integral de Cauchy. La cuarta es una

por ende de la teoría de ecuaciones. En la demostración de Euler la descomposición de todo polinomio en factores lineales o cuadráticos implica que toda raíz es de la forma $M + N\sqrt{-1}$, pero para llegar a este resultado él supone *a priori* la existencia de n raíces, y las calcula como si ellas aparecieran justamente de las operaciones algebraicas habituales sobre los números, entonces su naturaleza es realmente desconocida. "Esta es una contradicción que constituye una verdadera laguna lógica" (comentario de Gauss). No da Euler una demostración general del teorema, realmente, estudia diversos casos particulares (considera polinomios de grado superior) y se podría afirmar que "generaliza" apresuradamente tratando de convencer al lector de la validez del TFA.

Punto de Discusión
Enunciar e ilustrar (use sus propias palabras) la diferencia existente entre el Teorema de Factorización Lineal y el Teorema Fundamental del Algebra.

9.3.5 El aporte de D'alembert.

La demostración del Teorema Fundamental del Algebra dada por D'Alembert figura en una memoria titulada *Recherches sur le calcul intégral* que fue publicada por la Academia de Berlín en 1748. Se sitúa este trabajo en el marco del estudio de la integración de fracciones racionales. En el inicio del estudio D'Alembert sistematiza los conocimientos de la época y muestra la estabilidad de los símbolos $a + b\sqrt{-1}$ $(a, b \in \mathbb{Z})$, para operaciones algebraicas y analíticas elementales. Establece que una función (de variable compleja) en variables del tipo $x + y\sqrt{-1}$ puede ser escrita en la forma $p + q\sqrt{-1}$ donde p y q son funciones de variable real.

Nota. ¡Este es precisamente el argumento que usamos en la sección 2 para trabajar ecuaciones en variable compleja!

Deduce también que una raíz imaginaria de una ecuación es ella misma una cantidad de la forma $p + q\sqrt{-1}$. Al respecto dice.

"Una ecuación que tiene raíces imaginarias puede ser divida en trinomios donde los coeficientes son reales".

La demostración que da de este hecho es errónea pues supone que la raíz de una ecuación algebraica se expresa en términos finitos con la ayuda de funciones elementales explícitas; ello confirma el problema que existe en la época respecto a la ligazón de dos resultados generales, uno sobre la forma de las 'cantidades imaginarias' (raíces imaginarias de ecuaciones algebraicas) y el otro la posición de que la demostración del TFA se basa en el resultado que toda raíz de una ecuación algebraica puede ser expresada por radicales. Es interesante observar como evoluciona el pensamiento de D'Alembert al respecto; en 1745 presenta una demostración directa usando resultados sobre las raíces, que sustituye, en 1746 por una suposición sobre la expresión de éstas en función de los coeficientes. La crítica de Gauss y de otros contemporáneos a la demostración de D'Alembert es que trabaja sobre la forma de las raíces imaginarias de la ecuación, pero no sobre el problema mismo del TFA, la existencia de tales raíces; en concepto de Gauss, D'Alembert asume la existencia. Sin embargo para algunos historiadores el TFA la demostración de D'Alembert sí tiene el estatus de un teorema de

variación de la primera; sin embargo en ella Gauss usa números complejos mas libremente (y dice que ahora son de conocimiento común). Gauss en sus primeras tres demostraciones y mas tarde Cauchy, Jacobi y Abel, asumen que los coeficientes representan números reales, pero en la cuarta Gauss toma los coeficientes como complejos dando a ella mayor generalidad.

Nota. Un análisis detenido de las mencionadas demostraciones se encuentra en un apéndice.

Con su trabajo en el TFA Gauss inaugura una nueva manera de acercarse a la cuestión de la existencia en matemáticas. Para los griegos el criterio de existencia era la constructibilidad, en trabajos mas formales de siglos posteriores, la existencia era establecida por exhibir la cantidad en cuestión. Por ejemplo, la existencia de la solución de una ecuación cuadrática es establecida por exhibir cantidades que satisfacen la ecuación; pero en el caso de ecuaciones de grado superior a cuatro este método no es adecuado. El desarrollo de la demostración de Gauss no nos ayuda a 'calcular' el objeto cuya existencia se está estableciendo.

La teoría de ecuaciones después de Gauss se proyecta a niveles de mayor generalidad. En 1887 Kronecker, inspirado en la demostración algebraica de Gauss del TFA y con la ayuda de la noción de congruencia de polinomios, muestra que todo polinomio con coeficientes racionales puede ser descompuesto en un producto de factores lineales. Steinitz en 1910 enuncia un teorema que corresponde a la existencia de un cuerpo de descomposición, una versión mas general para un polinomio con coeficientes en cualquier cuerpo conmutativo; obtiene así la versión moderna del TFL. Trabajos posteriores de E. Artin y O. Schreier establecen en la teoría de cuerpos reales una generalización del TFA, a saber si K es un cuerpo real cerrado, entonces el cuerpo $K(i)$ es algebraicamente cerrado. En la demostración de este teorema se hace referencia de nuevo al argumento algebraico de Gauss en su demostración del TFA.

9.5 Anexo

9.5.1 Gauss y sus demostraciones del Teorema Fundamental

9.5.2 Primera demostración

Enuncia Gauss inicialmente el TFA de la siguiente forma.

Toda polinomial X con coeficientes reales o complejos puede ser factorizada en factores lineales en el campo de los números complejos.

Y afirma que es suficiente demostrar el teorema para polinomios con coeficientes reales, pues si X tiene coeficientes complejos el producto $X\overline{X}$ es real y su factorización implica la factorización de X y $\overline{X}$. En realidad, Gauss demuestra la siguiente forma del teorema fundamental.

Toda polinomial X con coeficientes reales puede ser factorizada en factores

lineales y cuadráticos.

Empieza Gauss con una polinomial real

$$X = x^m + Ax^{m-1} + Bx^{m-2} + \cdots + Lx + M$$

en la cual x es una indeterminada. El desea probar la existencia de un factor lineal o cuadrático de X. Un factor lineal implica la existencia de una raíz real $\pm r$, donde r es positivo o cero. Un factor cuadrático implica la existencia de dos raíces complejas $r(cos\phi \pm isen\phi)$; de donde el factor cuadrático puede ser escrito como

$$x^2 - 2xrcos\phi + r^2$$

con $r > 0$. Sustituyendo una de las raíces en la ecuación $X = 0$ y separando parte real y parte imaginaria, uno puede obtener un par de ecuaciones reales para r y ϕ

$$(1) r^m cos\, m\phi + Ar^{m-1} cos(m-1)\phi + \cdots + Lrcos\phi + M = 0;$$

$$(2) r^m sen\, m\phi + Ar^{m-1} se(m-1)\phi + \cdots + Lrsen\phi = 0.$$

Gauss nota que Euler obtenía este par de ecuaciones usando números complejos. Gauss no considera complejos; él deriva (1) y (2) de manera directa de la suposición de que X tiene un factor lineal $x \pm r$ o un factor cuadrático. Interpreta (1) y (2) como ecuaciones de curvas algebraicas en coordenadas polares y procede a demostrar que estas curvas se intersectan al menos en un punto. Una vez demostrado esto, se sigue que X tiene un factor lineal o cuadrático y, continuando el proceso, se obtiene una factorización de X en factores lineales o cuadráticos. Notemos la ecuación (1) $U = 0$ y la (2) $T = 0$. Para ilustrar la demostración dibujemos las curvas $U = 0$ y $T = 0$ para el caso de la ecuación cuadrática $x^2 + 3x + 1 = 0$, cuya solución estudiamos ya anteriormente (Figura 23).

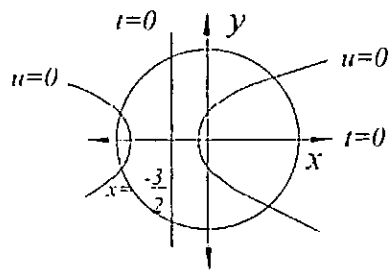


Figura 23

En coordenadas rectángulares x y y tendremos dos curvas de orden m y el eje $y = 0$ es siempre parte de la segunda curva $T = 0$

Gauss estudiá a continuación la intersección de estas dos curvas con un círculo de radio r y demúestra que

"Para un radio R suficientemente grande existen exactamente $2m$ intersecciones del círculo con $T = 0$ y $2m$ intersecciones del círculo con $U = 0$ y todo punto de intersección de la segunda clase está entre dos puntos de intersección de la primera clase."

Nota Si usted retoma las ecuaciones que se trabajaron en la sección 2 de este capitulo podrá interpretar claramente el lema anterior.

Gauss presentó una completa demostración de este lema, anotando a continuación que

"Los $4m$ puntos cambian muy poco si R se hace un poco más grande o pequeño. (En términos modernos los $4m$ puntos cambian continuamente en función de R.) No demuestra la continuidad (dice es fácil ver..). Pasa luego a estudiar las ramas de las curvas $U = 0$ y $T = 0$ interiores al círculo y comenta que

"Existe un punto de intersección de una rama de la primera curva con una rama de la segunda curva"

y para demostrar esta afirmación usa un argumento geométrico. Nota el punto de intersección del círculo con el eje x negativo por 0, el siguiente punto por 1 y así sucesivamente (como se ha hecho en la figura (24)). Los números impares denotan puntos sobre $U = 0$ y los pares puntos sobre $T = 0$. Ahora él dice si una rama de una curva algebraica entra a un cierto dominio debe tener el mismo dominio siempre; si este punto se acepta los "puntos pares" se conectan con al menos otro punto par de una rama de la curva $T = 0$ y todo punto impar con otro impar de una rama de la curva $U = 0$. Por este camino resulta complicado demostrar la existencia de un punto de intersección (es más bien una aproximación intuitiva); para ello utilizó el siguiente argumento.

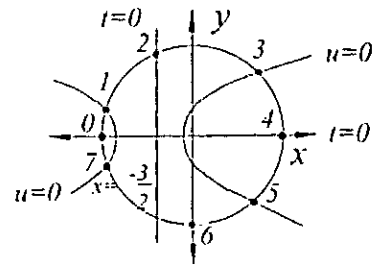


Figura 24

Supongamos que no existen puntos de intersección, el punto 0 es conectado con el punto $2m$ por el eje x. El punto 1 no puede ser conectado con ningún punto del otro lado del eje x sin intersectar el eje x. Por tanto si el punto 1 se conecta con el punto impar n, debe ocurrir que $n < 2m$. De otra parte, si 2 se conecta con n', $n' < n$, pero la diferencia $n' - 2$ es par, pues 2 y n' son pares. Continuando por este camino finalmente se encuentra un punto h conectado con $h + 2$. Pero ahora la rama que entra al círculo en el punto $h + 1$ debe necesariamente intersectar la rama que conecta a h con $h + 2$ y esto contradice la hipótesis. Por tanto existe un punto de intersección.

Comenta B.L der Waerden en su libro *A History of Algebra* respecto a esta primera demostración de Gauss que las suposiciones acerca de las ramas de curvas algebraicas parecen plausibles a nuestra intuición geométrica pero no son probadas estrictamente por Gauss. Sin embargo es de anotar que en 1920 Alexander Ostrowski publica un artículo en el que demuestra todas las suposiciones de Gauss con argumentos irrefutables.

9.5.3 Segunda demostración

La segunda demostración es puramente algebraica. Las únicas suposiciones acerca del campo de los números reales son las siguientes.

- Toda ecuación polinomial con coeficientes reales de grado impar tiene una raíz real.

- Toda ecuación cuadrática con coeficientes complejos tiene dos raíces complejas.

Gauss empieza con el polinomio real de grado m

$$Y = x^m - L'x^{m-1} + L''x^{m-2} + \cdots + \cdots .(1')$$

¿Qué ocurre si suponemos que el polinomio Y puede ser factorizado en algún campo de extensión en factores lineales, esto es

$$Y = (x-a)(x-b)(x-c)\cdots ?$$

Se podría entonces formar una combinación lineal con una nueva indeterminada t de la forma

$$(a+b)t - ab.(3)$$

Si las raíces $a, b, c, \cdots$ se permutan, la función lineal (3) asume $m' = \frac{1}{2}m(m+1)$ valores. Por tanto es una raíz de una ecuación de grado m'. Las raíces de esta ecuación auxiliar son funciones lineales de t de la forma (3). Si una raíz de la auxiliar es conocida $a+b$ y ab son conocidas; esto implica que a y b pueden ser expresadas por medio de una raíz cuadrada.

Punto de discusión
Discutir la validez del argumento anterior.

Si m es de la forma $m = 2^n k$, $.0k$ impar. El grado de la auxiliar sería $m' = 2^{n-1}k'$ con k' par. Por tanto si conocemos una raíz compleja de la ecuación auxiliar se pueden calcular dos raíces a y b de la original hallando raíces cuadradas de un número complejo.

Continuando por el mismo camino se llega a una ecuación de grado impar; los coeficientes de esta ecuación son funciones simétricas de las raíces a y b, con coeficientes reales y por tanto son números reales conocidos. Dado que el grado de la ecuación es impar, tiene al menos una raíz real. Sustituyendo en secuencia de ecuaciones auxiliares uno puede calcular al menos una raíz compleja de la ecuación original.

El trabajo de la demostración radica en probar que la ecuación $Y = 0$ tiene m raíces $a, b, c, \cdots$, en algún campo de extensión de los números reales. Esto puede ser demostrado por el método de "adjunción simbólica" de Kronecker.

Gauss no sigue este camino sino construye su ecuación auxiliar sin asumir la existencia de raíces; la ecuación auxiliar de grado m' es construida de la siguiente manera.

En primer lugar reemplaza el polinomio (1') por una polinomial y, cuyas raíces son las indeterminadas $a, b, c, \cdots$

(2')
$$y = (x-a)(x-b)(x-c)\cdots$$

Luego forma una ecuación auxiliar ξ en una nueva variable u, definiendo ξ como el producto de m' expresiones de la forma

(3')
$$u - (a+b)t + ab$$

obtenidas por permutación de las raíces. Esta polinomial ξ es simétrica en las indeterminadas $a, b, c, \cdots$, por tanto puede ser expresada de manera

única como una polinomial en u y t y los coeficientes de y, los cuales son funciones simétricas elementales de $a, b, c, \cdots$. Por tanto estos coeficientes pueden ser reemplazados por los coeficientes $L', L'', \ldots$ de la polinomial (1), obteniendo de esta forma una ecuación polinomial auxiliar Z.

A continuación Gauss demuestra el teorema siguiente Si el discriminante de Y no es cero, el discriminante de Z no puede ser cero. Después sustituye un valor de t tal que el discriminante de Z sea diferente de cero y demuestra que si una raíz de Z es conocida es posible calcular un par de raíces de la polinomial original Y. Comenta B.L. van der Waerden que obviamente Gauss deriva su método para encontrar una raíz de Y de la suposición de que Y es producto de factores primos, y después rehace la prueba para hacerla independiente de esta suposición.

9.5.4 Tercera demostración.

Gauss empieza considerando la polinomial

$$X = z^m + Az^{m-1} + Bz^{m-2} + \cdots + Lz + M(4)$$

con coeficientes reales. Llama

$$r^m cos\, m\phi + Ar^{m-1} cos(m-1)\phi + \cdots + Lr cos\phi + M = t$$

$$r^m sen\, m\phi + Ar^{m-1} sen(m-1)\phi + \cdots + Lr sen\phi = u.$$

Las expresiones t y u son las mismas U y T de la primera demostración. Son la parte real e imaginaria obtenidas al sustituir la variable compleja z

$$z = r(cos\phi + i sen\phi),$$

en (4). Las derivadas de t y u con respecto a ϕ son llamadas respectivamente $-u'$ y t'. Tenemos entonces

$$t' = mr^m cos\, m\phi + (m-1)Ar^{m-1} cos(m-1)\phi + \cdots + Lr cos\phi;$$

$$u' = mr^m sen\, m\phi + (m-1)Ar^{m-1} sen(m-1)\phi + \cdots + Lr sen\phi.$$

A continuación Gauss demostró que

$$tt' + uu'$$

es positivo para un valor R de r suficientemente grande sin importar el valor de ϕ. Lo anterior se demuestra observando que, para valores grandes de r, los términos dominantes de t y u son

$$r^m cos\, m\phi$$

y

$$r^m sen\, m\phi$$

y los términos dominantes de t' y u' son

$$mr^m cos\, m\phi$$

y

$$mr^m sen\, m\phi.$$

Por tanto el término dominante de $tt' + uu'$ es

$$mr^{2m}(cos\, m\phi^2 + sen\, m\phi^2) = mr^{2m}$$

que es siempre positivo. Las segundas derivadas de t y u con respecto a ϕ son llamadas $-u''$ y t''.

$$t'' = m^2 r^m cos\, m\phi + \cdots + Lrcos\phi$$

$$u'' = m^2 r^m sen\, m\phi + \cdots + Lrsen\phi$$

El objetivo de Gauss era probar que existía un punto del plano en el cual t y u son ambas nulas. Ya que la existencia de tal punto implica la existencia de una raíz compleja de la polinomial X. Si la raíz es real, X tiene un factor lineal y el proceso continúa. Si la raíz no es real X tiene un factor cuadrático y de manera similar el proceso continúa.

Supongamos que no existe un punto del plano tal que $t = u = 0$, entonces $t^2 + u^2$ es siempre diferente de cero, y la función

$$y = \frac{(t^2 + u^2)(tt'' + uu'') + (tu' - ut')^2 - (tt' + uu')^2}{r(t^2 + u^2)^2}$$

es siempre finita. Nótese que el factor r en el denominador se cancela puesto que t', u', t'', u'' son divisibles por r. Gauss considera a continuación la doble integral

$$\omega = \int_0^R \int_0^3 60y\, dr\, d\phi.$$

El resultado es el mismo si integramos primero con respecto a ϕ y luego con respecto a r. ¿Por qué razón? La integral indefinida con respecto a ϕ es

$$\int y d\phi = \frac{tu' - ut'}{r(t^2 + u^2)}.$$

Es claro que si diferenciamos la función del lado derecho obtenemos y; además la función del lado derecho tiene el mismo valor para $\phi = 0$ y para $\phi = 360^0$, por tanto el valor de la integral entre 0 y 360 es cero. Esto implica que

$$\omega = 0.$$

De otra parte si uno integra primero con respecto a r, obtiene la integral indefinida

$$\int y dr = \frac{tt' + uu'}{t^2 + u^2}.$$

Para $r = 0$ esta expresión es nula y para $r = R$ es positiva. Esto significa que la integral es positiva y por tanto ω es positiva esto contradice nuestra conclusión en (12). Por tanto la hipótesis de que t y u nunca son ambos cero conduce a una contradicción.

La pregunta es como Gauss encontró esta demostración, No se sabe claramente pero se puede adivinar como; comenta *B.L.van der Waerden*:

Pues si consideramos $X = t + iu$ como una función de variable compleja

$$z = r(cos\phi + isen\phi),$$

la función

$$X(z) = X(r, \phi) = t + iu$$

geométricamente hablando define una función del z-plano en el X-plano. En el X-plano podemos introducir coordenadas polares

$$X = s(cos\beta + isen\beta)$$

y tenemos

$$tan\beta = \frac{u}{t}.$$

Diferenciando con respecto a ϕ y r obtenemos

$$\beta\phi = cos^2\beta\frac{tt' + uu'}{t^2} = \frac{tt' + uu'}{t^2 + u^2} = U$$

y

$$\beta r = cos^2\beta\frac{tu' - ut'}{rt^2} = \frac{tu' - ut'}{r(t^2 + u^2)} = V.$$

Diferenciando una vez más obtenemos

$$Ur = V\phi = y.$$

Gauss hace uso de la ecuación anterior para concluir

$$\int y\, dr = U$$

y

$$\int y\, d\phi = V.$$

Es posible que Gauss encontrara su función en y derivando U con respecto a r y V con respecto a ϕ. Conociendo además que Ur y $V\phi$ coinciden, U y V son derivadas de una misma función β con respecto a ϕ y a r. A pesar de que el ángulo β no está unívocamente definido, en una vecindad de un punto (r, ϕ) el ángulo β es función diferenciable de r y ϕ y por tanto la diferencial total está bien definida. Sin embargo, Gauss encontró una manera de evitar el uso de los multivalores de ϕ trabajando con una integral doble e intercambiando el orden de integración; llegando finalmente por este camino a obtener una contradicción.

1.6 Problemas del capítulo

1. La ecuación $|z-1|+|z+1|=3$ siendo z una variable compleja, representa una curva plana. Describir la curva y expresar su ecuación en coordenadas rectangulares.

2. (a) Demostrar que las raíces de la ecuación $z^4+2z+1=0$ son las raíces cuartas primitivas de la unidad.

   (b) Demostrar que la ecuación $z\bar{z}-2|z|+1=0$ tiene un número infinito de raíces.

3. Demostrar que si c es una constante real positiva, entonces la ecuación $|\frac{z+1}{z-1}|=c$ representa una circunferencia si $c\neq 1$ y una linea recta si $c=1$.

4. La ecuación $6x^4-x^3+10x^2-x+6=0$ tiene cuatro raíces distintas de igual módulo determinarlas.

5. Sea $(\mathbb{C}^*,\cdot)$, el grupo multiplicativo de los números complejos. Sea n un entero positivo. Demostrar que la función $f(z)=z^n$ es un homomorfismo de $(\mathbb{C}^*,\cdot)$ en $\mathbb{C}^*,\cdot)$. ¿Es f sobre?. Identificar el nucleo de f.

6. Si $P(z)$ y $Q(z)$ son polinomios con coeficientes complejos que tienen el mismo conjuto de ceros, posiblemente con diferentes multiplicidades y lo mismo ocurre con $P(z)+1$ y $Q(z)+1$. Demostrar que $P(z)\equiv Q(z)$.

7. ¿Cúal es el máximo valor de $|z^3-z+2|$, si z es un número complejo tal que $|z|=1$?.

8. Encontrar todas las raíces de la ecuación $z^n=(z+1)^n$ y demostrar que los puntos que corresponden a estas raíces son colineales.

9. Dada la ecuación cuadrática

$$z^2+(p+ip')z+q+iq'=0$$

demostrar que

   (a) Si la ecuación tiene una raíz real entonces

$$(q')^2-pp'q'+q(p')^2=0$$

   (b) Si la ecuación tiene dos raíces reales entonces

$$p^2-(p')^2-4q \text{ y } pp'=2q'$$

Capítulo 10

Algunos elementos de la Teoría de Cuerpos

En el presente capítulo nos proponemos estudiar elementos que enlazan los dos temas que nos han ocupado: la solución de ecuaciones y las estructuras algebraicas. Para ello, comenzaremos por una discusión de la solución de ecuaciones de grado mayor que cuatro y sus diferencias con las ecuaciones de grado menor. Esto nos llevará directamente a un estudio de las extensiones algebraicas y la noción de invarianza bajo permutaciones. En seguida consideraremos formalmente el problema de la factorización de polinomios, llegando a enunciar los resultados generales asociados con el concepto de campo de descomposición. Lo anterior nos lleva al Teorema Fundamental de la teoría de Galois en el cual se relacionan los campos de descomposición y ciertos grupos asociados con las permutaciones de las raíces de un polinomio conocido como el grupo de Galois. Finalmente, nos concentraremos en analizar las aplicaciones de los resultados de Galois pues nos proporcionan soluciones a problemas clásicos de la geometría y el álgebra, como son los problemas de construcciones con regla y compás (trisección de un ángulo, duplicación del cubo, construcción de polígonos regulares) y la solución de una ecuación polinomial por "radicales", precisamente estableciendo nexos entre ellos y dando una perspectiva mas amplia.

Una cosa que nos debe sorprender sobremanera es que, después de tantos siglos de estudio y dedicación a ella, la solución de ecuaciones polinómicas continúa eludiéndonos. Contamos apenas con algunos métodos que funcionan en unos casos, para ecuaciones de cierto grado o para aquellas que resultan factorizables. Hemos visto además que existen procedimientos de aproximación que sirven en el caso de ecuaciones que tienen raíces reales; programas sofisticados de computador que aplican automáticamente las conocidas fórmulas de solución para ecuaciones de grado menor que cinco, pero es claro que ni los métodos numéricos ni otros métodos computacionales resuelven cualquier ecuación polinómica.

Sin embargo existe una teoría que caracteriza toda ecuación soluble por fórmula y explica por qué otras ecuaciones no lo son. Esta es la teoría de Galois desarrollada por el joven matemático francés Evariste Galois hacia 1835. Los resultados obtenidos por varios de sus antecesores (y especialmente por el matemático noruego Niels Abel) indicaron a Galois que el problema que buscaban resolver los matemáticos, a saber, encontrar una

fórmula de solución para la ecuación quíntica en términos de operaciones aritméticas y radicación con los coeficientes de la ecuación) era imposible. Galois se dedicó determinar condiciones generales de solubilidad por fórmula, tarea que llevó a un estudio de la estructura de los campos de factorización (descomposición) de los polinomios e inició una nueva etapa en el álgebra. Después de Galois el álgebra superior, o sea el álgebra en las fronteras de investigación, se caracteriza como abstracta. No se centra en problemas concretos de buscar solución a una ecuación o un conjunto particular de ecuaciones, sino en problemas abstractos de estructuras y sus caracterizaciones.

10.1 Teoría de Galois

10.1.1 Antecedentes históricos

El estudio de las fórmulas de solución de ecuaciones cuadráticas, cúbicas y cuárticas y caracterización de las resolventes.

En esta primera sección analizaremos el estudio que hicieron los algebristas del siglo XVIII de las fórmulas conocidas para la solución de ecuaciones cuadráticas, cúbicas y cuárticas en un intento por dominar el caso de las quínticas, así como algunos de los problemas que encontraron en su estudio y que los llevaron a centrar su atención en lo que hoy llamamos campos de extensión y grupos de permutaciones de las raíces.

Resolvente de la ecuación cuadrática

Comenzaremos nuestras consideraciones con una nueva forma de escribir la solución a la ecuación cuadrática, con la intención de expresar de manera análoga la solución a la ecuación cúbica, es decir, para interrelacionar los resultados obtenidos en estos dos casos particulares. Sea $x^2+bx+c=0$ la ecuación cuadrática en consideración. Si denotamos por x, y la soluciones, entonces

$$\begin{aligned} x &= \frac{1}{2}\left[(x+y)+(x-y)\right] = \frac{1}{2}\left[(x+y)+\sqrt{(x-y)^2}\right] \qquad (10.1)\\ y &= \frac{1}{2}\left[(x+y)-(x-y)\right] = \frac{1}{2}\left[(x+y)-\sqrt{(x-y)^2}\right] \end{aligned}$$

Es claro que $x+y$ y $(x-y)^2$ son polinomios simétricos en x, y y que son expresables directamente en términos de los coeficientes, pues $x+y=-b$ y $(x-y)^2=(x+y)^2-4xy=b^2-4c$. Es claro que (*.*) corresponde a la familiar fórmula de solución para la cuadrática.

Resolvente de la ecuación cúbica

Ahora, al buscar una analogía para la ecuación cúbica $x^3+bx^2+cx+d=0$, observamos que si denotamos las raíces de la cúbica por x, y, z tenemos

$$\begin{aligned} x &= \frac{1}{3}\left[(x+y+z)+(x+\alpha y+\alpha^2 z)+(x+\alpha^2 y+\alpha z)\right] \qquad (10.2)\\ &= \frac{1}{3}\left[(x+y+z)+\sqrt[3]{(x+\alpha y+\alpha^2 z)^3}+\sqrt[3]{(x+\alpha^2 y+\alpha z)^3}\right], \end{aligned}$$

donde α es una raíz cúbica primitiva de la unidad. Recordemos que ser raíz cúbica primitiva de la únidad significa que $\alpha^3 = 1$ y que α es raíz de la ecuación cuadrática $x^2 + x + 1 = 0$ ya que $x^3 - 1 = (x-1)(x^2+x+1) = 0$. Se tienen expresiones similares para y y z de acuerdo con la escogencia de las raíces cúbicas en la expresión anterior, a saber

$$\begin{aligned} y &= \frac{1}{3}\left[(x+y+z) + \alpha^2(x+\alpha y+\alpha^2 z) + \alpha(x+\alpha^2 y+\alpha z)\right] \\ z &= \frac{1}{3}\left[(x+y+z) + \alpha(x+\alpha y+\alpha^2 z) + \alpha^2(x+\alpha^2 y+\alpha z)\right]. \end{aligned}$$

Ahora bien, es claro que $x+y+z = -b$. Además, queremos mostrar que las cantidades subradicales

$$\begin{aligned} u &= (x+\alpha y+\alpha^2 z)^3, \\ v &= (x+\alpha^2 y+\alpha z)^3, \end{aligned}$$

también pueden expresarse en términos de los coeficientes de la ecuación. Para ello notamos que la característica principal de los coeficientes de una ecuación polinómica es que representan funciones simétricas elementales de las raíces (relaciones de Viète) y, por ende, cualquier función simétrica de las raíces puede expresarse en función de ellas. Nos basta, entonces, mostrar que uv y $u+v$ son simétricas, es decir, que son invariantes bajo permutaciones de las raíces de la ecuación. Tenemos

$$\begin{aligned} u+v &= (x+\alpha y+\alpha^2 z)^3 + (x+\alpha^2 y+\alpha z)^3 \\ &= 2(x^3+y^3+z^3) + 12xyz - 3(x^2y+xy^2+x^2z+xz^2+y^2z+yz^2) \\ &= 2(x+y+z)^3 - 9(x^2y+xy^2+x^2z+xz^2+y^2z+yz^2) \\ uv &= (x+\alpha y+\alpha^2 z)^3(x+\alpha^2 y+\alpha z)^3 \\ &= \left[(x^2+\alpha^3y^2+\alpha^3z^2) + xy(\alpha+\alpha^2) + xz(\alpha+\alpha^2) + yz(\alpha^4+\alpha^2)\right]^3 \\ &= \left[(x^2+y^2+z^2) + (1+\alpha+\alpha^2)(xy+xz+yz) - (xy+xz+yz)\right]^3 \\ &= \left[(x^2+y^2+z^2) - (xy+xz+yz)\right]^3 \end{aligned}$$

Su invarianza bajo permutaciones de x, y, z es clara. Una vez se expresen u, v en función de los coeficientes, la expresión de x, y, z en función de los mismos sigue. Ahora bien, como nuestra expresión para x involucra la extracción de una raíz cúbica, resulta que hay tres posibles valores para u y tres para v, dando un total de nueve posibilidades. Tres de ellos corresponden a x, y y z mientras que los seis restantes no son soluciones. Se puede determinar cuáles de las nueve son en efecto soluciones sustituyendo en la ecuación y comprobando.

El estudio de las resolventes que hemos expuesto se debe al matemático francés Vandermonde que lo presentó ante la Academia Francesa en 1770. Unos pocos meses mas tarde fue publicada una extensa exposición del mismo problema por Lagrange cuya obra eclipsó casi por completo la de Vandermonde. En su estudio de la ecuación cúbica, el planteamiento de Lagrange considera la cantidad $t = x + \alpha y + \alpha^2 z$ donde x, y y z son las soluciones de la ecuación cúbica y α es una raíz cúbica de la unidad ($\alpha \neq 1$). Lagrange observa que t tiene seis valores posibles dependiendo del orden en que se toman las raíces, es decir, dependiendo de las seis permutaciones de x, y, z. Lagrange observa que estos seis valores son soluciones de la ecuación

$$f(X) = (X-t_1)(X-t_2)(X-t_3)(X-t_4)(X-t_5)(X-t_6) = 0, \quad (10.3)$$

y que los coeficientes de la ecuación, por ser funciones simétricas de los t_i son también funciones simétricas de x, y y z, y por ende, pueden expresarse en términos de los coeficientes de la ecuación cúbica original. Ahora bien, Lagrange llama (*.*) una ecuación resolvente porque, aunque es de grado seis, es en efecto una ecuación cuadrática en X^3. Para ver esto, basta observar que se pueden ordenar los seis valores de t de la siguiente manera. Sea $t_1 = x + \alpha y + \alpha^2 z$,entonces tenemos

$$t_1 = x + \alpha y + \alpha^2 z \quad t_2 = \alpha t_1 = \alpha x + \alpha^2 y + z \quad t_3 = \alpha^2 t_1 = \alpha^2 x + y + \alpha z$$
$$t_4 = x + \alpha z + \alpha^2 y \quad t_5 = \alpha t_4 = \alpha x + \alpha^2 z + y \quad t_6 = \alpha^2 t_4 = \alpha^2 x + z + \alpha y$$

Ahora bien, es fácil ver que $(X - t_1)(X - t_2)(X - t_3) = X^3 - t_1^3$ y, procediendo de la misma manera con los otros tres factores, que

$$f(X) = (X^3 - t_1^3)(X^3 - t_4^3).$$

Se sigue que es soluble aplicando la fórmula cuadrática y luego sacando raíz cúbica. Por otra parte, referente al trabajo de Vandermonde que expusimos anteriormente, se tiene que $u = t_1^3, v = t_4^3$ y que los coeficientes de $f(X) = 0$ son precisamente $u + v, uv$. Una vez resuelta la ecuación y conocidos los seis valores de t sólo resta identificar entre ellos t_1, t_4 pues las soluciones a la cúbica son

$$x = \frac{1}{3}\left[(x + y + z) + t_1 + t_4\right], y = \frac{1}{3}\left[(x + y + z) + \alpha^2 t_1 + \alpha t_4\right], z = \frac{1}{3}\left[(x + y + z) + \alpha t_1 +\right.$$

Para ello, observa Lagrange que si t es cualquiera de las seis soluciones y $w = (x + \alpha y + \alpha^2 z)(x + \alpha^2 y + \alpha z)$, w es simétrico en x, y, z y, por tanto, es conocido. Las tres soluciones a la cúbica son

$$x = \frac{1}{3}\left[(x + y + z) + t + (w/t)\right], y = \frac{1}{3}\left[(x + y + z) + \alpha t + (w/\alpha t)\right], z = \frac{1}{3}\left[(x + y + z) + \alpha^2\right.$$

Resolvente de la cuártica

Tanto Vandermonde como Lagrange estudiaron también la solución de la cuártica, llegando a conclusiones similares a los anteriores aunque su procedimiento fue ligeramente diferente. Tomando i como raíz primitiva cuártica de la unidad, denotando las raíces de la ecuación por x, y, z, r y comenzando con la fórmula análoga para x, a saber

$$x = \frac{1}{4}\left[(x + y + z + r) + \sqrt[4]{(x + iy - z - ir)^4} + \sqrt[4]{(x - y + z - r)^4} + \sqrt[4]{(x - iy - z + ir)^4}\right.$$

tanto Lagrange como Vandermonde se dieron cuenta que basta evaluar solamente una de estas expresiones subradicales para posibilitar la solución de la ecuación de grado cuatro. La cantidad en cuestión es $t = x - y + z - r$. Las veinticuatro permutación de las raíces en este caso dan lugar a sólo seis valores de t que son

$$\pm(x - y + z - r), \pm(x + y - z - r), \pm(x - y - z + r),$$

cada uno de los cuales aparece cuatro veces. Si denotamos estos seis valores por $\pm t_1, \pm t_2, \pm t_3$ se sigue que la ecuación resolvente asociada con ellos es

$$f(X) = (X - t_1)^4(X + t_1)^4(X - t_2)^4(X + t_2)^4(X - t_3)^4(X + t_3)^4 = 0 = g(X)^4$$
$$= \left[(X^2 - t_1^2)(X^2 - t_2^2)(X^2 - t_3^2)\right]^4.$$

10.1.2 Raíces de la unidad

En esta sección nos ocuparemos con otro paso histórico en el desarrollo de la teoría de Galois, cual es la solución *algebraica* de ecuaciones de la forma $x^n - 1 = 0$, y cuyos aspectos geométricos serán tenidos en cuenta mas adelante. El caso $n = 1$ corresponde a la ecuación $x - 1 = 0$ o $x = 1$y $n = 2$ corresponde a la ecuación $x^2 - 1 = 0$ cuyas raíces son $x = \pm 1$. Estos primeros casos nos dan pie para definir raíz primitiva de la unidad.

Definición 10.1 *El número complejo ξ se llama raíz $m-$ésima primitiva de la unidad si ξ es raíz de la ecuación $x^m - 1 = 0$ y ξ no es raíz de la ecuación $x^k - 1 = 0$ para ningún $k < m$.*

Se sigue que 1 es raíz primitiva para $m = 1$ pero no lo es para $m = 2$,mientras que -1 es una raíz segunda primitiva de la unidad. En general,

$$x^m - 1 = (x - 1)(x^{m-1} + x^{m-2} + \cdots + x + 1) \tag{10.4}$$

mostrando que las raíces primitivas son raíces del segundo factor.

De allí se sigue que la ecuación $x^3 - 1 = 0$ puede factorizarse como $x^3 - 1 = (x - 1)(x^2 + x + 1) = 0$. Las raíces del segundo factor son, claramente, $\alpha = \frac{-1+\sqrt{3}i}{2}$ y $\overline{\alpha} = \frac{-1-\sqrt{3}i}{2}$ que son las raíces cúbicas primitivas de las unidad.

Ejercicio

Comprobar que $\alpha^2 = \overline{\alpha}$ y que $\alpha^3 = \alpha \cdot \alpha^2 = 1$.

Ahora bien, a pesar de ser susceptible a la anterior factorización (*.*), si $m = 2k$ la ecuación $x^m - 1 = x^{2k} - 1 = \left(x^k\right)^2 - 1 = (x^k + 1)(x^k - 1)$, de manera que cualquier raíz $m-$ésima primitiva de la unidad será raíz de la ecuación $x^k + 1 = 0$. En particular, las raíces cuartas primitivas de la unidad satisfacen $x^2 + 1 = 0$ y son $\alpha = i, \overline{\alpha} = -i$.

Ejercicio

Comprobar que todas las raíces cuartas de la unidad son potencias de i, en particular, que son i, i^2, i^3, i^4.

Para resolver algebraicamente la ecuación $x^5 - 1 = (x-1)(x^4+x^3+x^2+x+1)$ notamos que dividiendo $x^4 + x^3 + x^2 + x + 1 = 0$ por x^2 se obtiene

$$x^2 + x + 1 + \frac{1}{x} + \frac{1}{x^2} = \left(x + \frac{1}{x}\right)^2 + \left(x + \frac{1}{x}\right) + 1 - 2$$

que es una cuadrática en $x + \frac{1}{x}$. La solución que se obtiene es $\beta = x + \frac{1}{x} = \frac{-1\pm\sqrt{1+4}}{2} = \frac{-1\pm\sqrt{5}}{2}$. Ahora, multiplicando esta ecuación por x nos lleva a la cuadrática

$$x^2 - \beta x + 1 = 0$$

cuyas soluciones son

$$x = \frac{\beta \pm \sqrt{\beta^2 - 4}}{2} = \frac{\frac{-1\pm\sqrt{5}}{2} \pm \sqrt{-\frac{10\pm 2\sqrt{5}}{4}}}{2}$$

que produce las cuatro raíces primitivas $\alpha = \frac{\sqrt{5}-1\pm\sqrt{2\sqrt{5}-10}}{4}, \frac{-\sqrt{5}-1\pm\sqrt{2\sqrt{5}-10}}{4}$.

Puntos de discusión

Por consiguiente, t_1^2, t_2^2, t_3^2 son las raíces de una ecuación cúbica conocida ya que los coeficientes de $g(X)$ son funciones simétricas en t_1^2, t_2^2, t_3^2 y, por ende, son simétricas en x, y, z, r. Ya que esta ecuación cúbica puede resolverse, podemos obtener t_1^2, t_2^2, t_3^2 y, sacando raíz cuadrada, $\pm t_1, \pm t_2, \pm t_3$. Además, como se tiene que

$$\begin{aligned} x &= \frac{1}{4}\left[(x+y+z+r)+t_1+t_2+t_3\right], \\ y &= \frac{1}{4}\left[(x+y+z+r)-t_1+t_2-t_3\right], \\ z &= \frac{1}{4}\left[(x+y+z+r)+t_1-t_2-t_3\right], \\ r &= \frac{1}{4}\left[(x+y+z+r)-t_1-t_2+t_3\right], \end{aligned}$$

se puede resolver la ecuación.

Resolvente de la ecuación de grado cinco

Cuando se quieren extender estas ideas al estudio de la ecuación quíntica, se presenta una situación nueva que no es susceptible a ser dominada por las mismas tácticas. Volviendo a los casos de la cúbica y cuártica, podemos observar que las resolventes son, en efecto, de grado superior al grado de la ecuación original. Sin embargo, circunstancias particulares relacionadas con los grupos de permutaciones de las raíces, permiten que se "rebaje" el grado de la resolvente. En el caso de la cúbica, en lugar de una ecuación de grado $3! = 6$ se obtiene una cuadrática en X^3 debido, precisamente, a que las permutaciones cíclicas producen tres valores que son las raíces cúbicas de (usando la misma notación anterior) t_1^3, mientras que las trasposiciones producen tres que representan las raíces cúbicas de t_4^3. En el caso de la cuártica sucede algo similar; la resolvente es de grado 24 pero se reduce a una ecuación de grado 6 en X^4 que se resuelve con facilidad porque es, en efecto, una ecuación cúbica en X^2. Esto se debe al número de raíces primitivas de la unidad que existen en ambos casos. Pero en el caso de la ecuación de grado cinco, hay cuatro raíces primitivas y la resolvente, que es de grado $5! = 120$, se reduce a una ecuación de grado 24 en X^5 debido a la siguiente relación. Sea, análogamente a los casos anteriores,

$$t = x_1 + \alpha x_2 + \alpha^2 x_3 + \alpha^3 x_4 + \alpha^4 x_5; \alpha^5 = 1; \alpha \neq 1.$$

Entonces, $\alpha t = x_5 + \alpha x_1 + \alpha^2 x_2 + \alpha^3 x_3 + \alpha^4 x_4, \alpha^2 t = x_4 + \alpha x_5 + \alpha^2 x_1 + \alpha^3 x_2 + \alpha^4 x_3$, $\alpha^3 t = x_3 + \alpha x_4 + \alpha^2 x_5 + \alpha^3 x_1 + \alpha^4 x_2 y \alpha^4 t = x_2 + \alpha x_3 + \alpha^2 x_4 + \alpha^3 x_5 + \alpha^4 x_1$ son tales que cumplen

$$f(X) = (X-t)(X-\alpha t)(X-\alpha^2 t)(X-\alpha^3 t)(X-\alpha^4 t) = X^5 - t^5.$$

Nótese que estos valores corresponden a la aplicación de las sucesivas potencias de la permutación cíclica $(x_1 x_2 x_3 x_4 x_5)$ a t. Ahora, la resolvente completa consta de 24 factores de este mismo tipo, mostrando que el procedimiento, que en los casos anteriores permitió rebajar el grado de la ecuación a resolver, en este caso resulta en una ecuación de grado mayor que la original.

1. Usar un procedimiento similar al caso $n = 5$ para mostrar cómo resolver la ecuación $x^7 - 1 = 0$.

2. Demostrar que si α es una raíz j-ésima primitiva y β una raíz $k-$ésima primitiva de la unidad y $(j, k) = 1$, entonces $\alpha\beta$ es una raíz jk-ésima de la unidad.

3. Hallar una raíz décima y una raíz vigésima primitiva de la unidad.

4. Sea α una raíz undécima primitiva de la unidad. Sustituyendo en la ecuación $x^{10} + x^9 + \cdots + x^2 + x + 1 = 0$ demostrar que las raíces undécimas primitivas de la unidad son $\alpha, \alpha^2, \alpha^3, \ldots, \alpha^{10}$.

5. ¿Cuántas raíces 2^n-ésimas de la unidad son primitivas?

Vandermonde también resolvió la ecuación $x^{11} - 1 = 0$.

Puntos de investigación

1. Estudiar el tratamiento de Vandermonde de esta ecuación.

2. Investigar acerca del método (algebraico) de Gauss para la solución de la ecuación $x^p - 1 = 0$ con $p > 11$.

Consideremos ahora el método geométrico que desarrolló Gauss.

Gauss y la ecuación ciclotómica

Observaremos aquí que precisamente con elementos de la teoría de números se resuelven problemas geométricos una vez transformados en problemas algebraicos.

Describiremos aquí las investigaciones de Gauss sobre los polígonos regulares, pues en ellas se aprecian, como ya lo comentamos, nexos importantes con la teoría de números. Inicia Gauss su discusión afirmando "Un polígono regular con n lados tiene sus vértices equidistantes sobre un círculo". El radio del círculo no es importante, así que se puede asumir que tiene radio $r = 1$. Dado que a cada uno de los lados del polígono le corresponde un ángulo central de

$$\frac{360°}{n} = \frac{2\pi}{n}.$$

El problema se traduce en dividir el ángulo de 360° en n partes iguales. Cualquier ángulo puede ser bisectado, por ello, si se puede construir un polígono de n lados, uno puede construir sucesivamente uno con $2n, 4n, \cdots$, en general con $2^k n$ lados.

Por otra parte, de un polígono con $2n$ lados uno puede construir uno con n. En consecuencia, uno puede limitar las consideraciones a polígonos regulares con un número impar de lados. Del hecho de que los polígonos regulares con $3, 4$ y 5 lados puedan ser construidos, se sigue que todos los polígonos con: $2^k, 3 \cdot 2^k, 5 \cdot 2^k$ lados son construibles.

Es evidente que si tenemos un polígono con n lados y $a|n$, esto es $n = ab$, el polígono con a lados puede ser derivado tomando cada b vértices. Mas interesante resulta el hecho de que los resultados básicos sobre ecuaciones

indeterminadas nos proporcionan, bajo ciertas condiciones, un camino para construir polígonos con un gran número de lados.

Por ejemplo de polígonos con a y b lados, donde a y b son primos relativos se puede obtener un polígono con ab lados.

Para demostrar éste hecho, observamos que como a y b son primos relativos, existen enteros x y y tales que $ax - by = 1$, esto es 1 es una combinación lineal de a y b. (Nótese como toma esta relación. ¿Pierde generalidad en las consideraciones?)

Dividiendo esta relación por ab tenemos

$$\frac{1}{a \cdot b} = \frac{x}{b} - \frac{y}{a}$$

o

$$\frac{360°}{a \cdot b} = x\frac{360°}{b} - y\frac{360°}{a}.$$

Esto demuestra que el ángulo central de un polígono con ab lados es la diferencia entre dos múltiplos del ángulo central de los polígonos con a y b lados. Por ejemplo, como los polígonos con 3 y 5 lados son construibles, el polígono con 15 lados es construible. Se concluye de esto también que es suficiente mostrar la constructibilidad de polígonos para los cuales el número de lados es una potencia de un primo impar.

La construcción de un polígono de n lados o equivalentemente un ángulo de $\frac{2\pi}{n}$, puede ser realizada usando funciones trigonométricas del ángulo, por ejemplo $\cos\frac{2\pi}{n}$ o $\sin\frac{2\pi}{n}$.

Por medio de la Ley de los Cosenos uno puede encontrar la expresión

Un complejo puede ser expresado en forma polar como $a + bi = r(\cos\phi + \sin\phi)$, $i = \sqrt{-1}$, donde a y b son las coordenadas en el plano complejo, r es el radio vector o valor absoluto y ϕ es el ángulo o amplitud que el radio vector forma con el eje real.

Si este complejo se multiplica por otro $a_1 + ib_1 = r_1(\cos\phi_1 + i\sin\phi_1)$, tenemos

$$(a + ib)(a_1 + ib_1) = rr_1[\cos(\phi + \phi_1) + i\sin(\phi + \phi_1)].$$

De la anterior se deriva el conocido Teorema de Moivre

$$(a + ib)^n = r^n(\cos n\phi + i\sin n\phi).$$

Gauss asumió que un círculo con radio 1 puede ser dibujado en el plano complejo. En éste círculo él inscribió un polígono regular con n lados tal

que uno de los vértices esté sobre el eje real positivo en el punto $x = 1$ (ver figura 1). Entonces el siguiente vértice corresponderá al número complejo

$$\xi = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}$$

y los siguientes

$$\xi_2 = \cos\frac{4\pi}{n} + i\sin\frac{4\pi}{n}, \cdots \xi_{n-1} = \cos\frac{2\pi\cdot(n-1)}{n} + \sin\frac{2\pi\cdot(n-1)}{n}.$$

El teorema de Moivre demuestra que estos números son potencias de ξ ,

$$\xi_0 = \xi^0 = 1, \xi_1 = \xi, \xi_2 = \xi^2, \cdots, \xi_{n-1} = \xi^{n-1}.$$

De la fórmula de de Moivre se concluye

$$\xi^n = \left(\cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}\right)^n = \cos 2\pi + i\sin 2\pi = 1,$$

de donde,

$$\xi^n = 1.$$

Este resultado establece un punto de singular importancia en esta discusión, a saber, ξ y todas sus potencias son raíces de la ecuación algebraica

$$x^n - 1 = 0.$$

Por esta razón, uno se refiere a los ξ^j como las n-ésimas raíces de la unidad y a la última ecuación como *ecuación de división del círculo* o *ecuación ciclotómica*.

Recuerde que cuando estudiamos grupos, anotamos que el conjunto formado por las raíces n-ésimas de la unidad, con la multiplicación de complejos es un grupo, aún mas, es un grupo cíclico de orden n.

Las dos funciones trigonométricas de las que depende la construcción del polígono regular, aparecen como componentes de las n-ésimas raíces de la unidad y, cuando ellas pueden ser expresadas como operaciones con raíces cuadradas, las ξ también. De allí se sigue que si el polígono regular con n lados puede ser construido con regla y compás, la correspondiente ecuación ciclotómica puede ser resuelta por raíces cuadradas.

La ecuación $x^n - 1 = 0$, no es la ecuación de grado mínimo que la n-ésima raíz de la unidad ξ satisface, dado que ésta puede ser factorizada en los racionales, $x - 1$ es desde luego un factor. En general las raíces de la ecuación ciclotómica se clasifican en dos grupos. Algunas que no son raíces de la unidad para exponentes menores que n son llamadas *raíces primitivas*, las otras que satisfacen ecuaciones de este tipo con exponentes menores son llamadas no primitivas.

Es sencillo decidir cuando una raíz de la unidad

$$\xi_k = \xi^k = \cos\frac{2\pi k}{n} + i\sin\frac{2\pi k}{n}$$

es primitiva. Esto es, si satisface una ecuación

$$x^t = 1,\ t < n.$$

Por el teorema de de Moivre

$$(\xi_k)^t = \cos\frac{2\pi kt}{n} + i\sin\frac{2\pi kt}{n} = 1.$$

Esto solamente es posible si la amplitud $2\pi\frac{kt}{n}$, es un múltiplo entero de 2π. Cuando k es primo relativo con n, el menor entero t que satisface esta condición es $t = n$, mientras que un entero menor t puede ser encontrado cuando k y n tienen un factor común. (Este análisis permite caracterizar el orden de los elementos de este grupo, a la vez que identificar los posibles generadores a que hacíamos referencia en un capítulo anterior.) Tenemos entonces que

Una raíz n-ésima de la unidad ξ_k, definida anteriomente, es primitiva sólo cuando k es primo relativo con n.

Para $n = 6$ son primitivas entonces ξ y ξ_5, pero para $n = 7$ son primitivas $\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6$, el grupo correspondiente puede ser generado por cualquiera de ellas.

Puntos de investigación

1. Determinar las raíces cúbicas de la unidad por un análisis geométrico.
2. ¿Cuántas raíces cuartas primitivas de la unidad hay? ¿Cuántas sextas primitivas? ¿Cuántas onceavas primitivas?

Por el análisis anterior nosotros podemos afirmar que el número de raíces primitivas de la unidad es $\phi(n)$ donde ϕ corresponde a la función de Euler.

El siguiente paso en el álgebra de las raíces de la unidad es mostrar que las $\phi(n)$ raíces primitivas de la unidad satisfacen una ecuación con coeficientes racionales de grado $\phi(n)$ y que esta ecuación no puede ser factorizada en los racionales (es minimal). No incluiremos esta demostración aquí, dado que nuestro interés en el momento es mostrar los antecedentes que tuvo a la mano Galois cuando generó su teoría.

Retornemos a las resolventes

El método de resolventes para la solución de ecuaciones polinómicas no es general en el sentido que no proporciona un método de solución para ecuaciones polinómicas de grado mayor que 4, aunque sí puede usarse para la solución general de la ecuación $x^n - 1 = 0$, dadas las características especiales de ésta. El tratamiento de resolventes implicó, sin embargo, que se pusiera atención especial a las permutaciones de las raíces de una ecuación, tema que resultaría básica para el álgebra abstracta (en sus inicios el estudio de grupos de permutaciones). Por otra parte, un examen de los resolventes condujo a Galois a importantes descubrimientos que formarían la base de su teoría y que serán nuestro próximo tema de estudio.

10.1.3 Evariste Galois

Al estudiar la obra de Galois se pone en claro la deuda que él tenía con sus antecesores, principalmente Lagrange, Gauss, Cauchy y Abel. Pero al mismo tiempo se observa el punto clave de transición en la cual el álgebra

pasa de ser el estudio de ecuaciones y su solución a ser principalmente el estudio de estructuras abstractas.

Al iniciar su trabajo, Galois se interesó por determinar las características de una ecuación que, o bien permiten que se resuelva por radicales, o bien no permiten dicha solución. Pero en el transcurso de la solución de este problema, Galois encontró resultados cuya trascendencia llega mucho las allá de la solución de ecuaciones polinómicas por radicales. En efecto, se considera a Galois como fundador de la teoría de grupos o del álgebra abstracta en general.

Nuestro estudio de la teoría de Galois cubrirá tres aspectos: (1) una descripción del trabajo original de Galois; (2) una exposición breve se la teoría de Galois como comunmente se estudia hoy día; y (3) el estudio de algunas de las implicaciones y aplicaciones de la teoría.

Comentarios preliminares

La investigación de Galois se apoya en varios resultados históricamente anteriores. La influencia mas notoria es la de Lagrange, pues Lagrange había tomado el primer paso cuando en su análisis de las fórmulas de solución por radicales de ecuaciones polinómicas descubrió que dichas fórmulas dependían a su vez del comportamiento de ciertas funciones convenientemente escogidas bajo (el grupo de) las permutaciones de las raíces. Por cierto, Lagrange no usó el concepto abstracto de grupo, pero sí desarrolló algunas propiedades aisladas de grupos a estudiar estas permutaciones. De los resultados de Lagrange utilizados por Galois se destacan estos dos.

1. Si ϕ y ψ son dos funciones de las raíces de una ecuación polinómica tales que ϕ toma r valores bajo las permutaciones que dejan a ψ invariante, entonces ϕ satisface un polinomio de grado r con coeficientes que son funciones racionales de ψ y los coeficientes de la ecuación original.

2. El teorema de Lagrange que, recordamos, dice que el orden (y el índice) de un subgrupo de un grupo finito divide al orden del grupo.

Punto de discusión

Detallar la inherencia de las permutaciones de las raíces de la ecuación en el comportamiento de las resolventes estudiadas por Lagrange.

De Gauss, Galois toma un resultado a cerca de la solución de las llamadas ecuaciones binomiales (relacionadas con las ciclotómicas), a saber, que ecuaciones del tipo $x^p - A = 0$, con p primo, son solubles por radicales.

Punto de discusión

Detallar la solución de la ecuación $x^p - A = 0$, con p primo, y el papel que juega en ella las raíces p-ésimas de la unidad.

De Cauchy y Abel Galois utiliza varios resultados, entre ellos el teorema, debido a Cauchy y utilizado por Abel, que el número de valores que una función no simétrica de n (raíces) no puede ser menor que el mayor primo p menor o igual a n al no ser que sea 2. Este es un precursor del Teorema de Cauchy, pero dicho teorema posdata el trabajo de Galois.

Además de estos teoremas particulares que fueron usados por Galois, es importante destacar las nociones que se tenían en ese tiempo de campo, campo de extensión y grupo.

Concepto de campo

Para los fines de los algebristas de la época de Galois, un campo es un conjunto en el cual se pueden efectuar las cuatro operaciones básicas (adición, sustracción, multplicación y división, exceptuando por supuesto la división por 0). Es por este motivo que se destacan las funciones racionales puesto que éstas son expresiones que contienen dichas operaciones.

Concepto de extensión de un campo Cuando Lagrange dice, en el teorema citado arriba, que ϕ es raíz de una ecuación polinómica cuyos coeficientes con funciones racionales de ψ y los coeficientes originales, está diciendo que ϕ satisface una ecuación en el campo de extensión $F(\psi)$ donde F es el campo al cual pertenecen los coeficientes de la ecuación.

Dado un campo F se piensa en el campo que se obtiene de F adjuntando algunos elementos 'nuevos' $\alpha_1, \alpha_2, \ldots, \alpha_n$ como el conjunto de todas las expresiones raciionales en $\alpha_1, \alpha_2, \ldots, \alpha_n$ sobre F.

$$F(\alpha_1, \alpha_2, \ldots, \alpha_n) = \left\{ \frac{p(\alpha_1, \alpha_2, \ldots, \alpha_n)}{q(\alpha_1, \alpha_2, \ldots, \alpha_n)} \mid p \text{ y } q \text{ son polinomios con coeficientes en } F, q(\alpha_1, \alpha \right.$$

$F(\alpha_1, \alpha_2, \ldots, \alpha_n)$ se llama una extensión del campo F. Para Galois los campos extensión son aquellos que se obtienen adjuntando un número finito de elementos y, se puede observar en los comentarios sobre el trabajo de Galois que estos elementos eran siempre algebraicos de tal forma que la extensión era de dimensión finita sobre F. Finalmente, en la obra de Galois no se observa el reconocimiento que el campo posee estructura de espacio vectorial y, por lo tanto, no se llega a reducir o simplificar los elementos del campo extensión. Es decir, siempre dichos elementos se consideran como expresiones racionales y no como combinaciones lineales de los elementos de la base del campo de extensión sobre el campo original.

Punto de discusión

¿Cómo se pueden expresar elementos de la forma $\frac{p(\sqrt{3})}{q(\sqrt{3})}$, donde p y q son polinomios con coeficientes racionales como elementos de la forma $a + b\sqrt{3}$ donde a, b son racionales, es decir, a elementos de un espacio vectorial con base $\{1, \sqrt{3}\}$ sobre los racionales?

El concepto de grupo

El concepto de grupo con que trabaja Galois está limitado al de un grupo finito de permutaciones. El único criterio que se usa para determinar si un conjunto dado de permutaciones es un grupo es que el conjunto debe estar cerrado respecto de la composición. Las demás propiedades que hoy día se incluyen en la definición de grupo abstracto se conocían pero no formaban parte de la definición que se manejaba. Lo anterior se debe a que, dentro del contexto finito donde se estudiaba la estructura de grupo, el único prerrequisito que se tiene para que un subconjunto sea subgrupo es precisamente la propiedad clausurativa.

10.1.4 Un bosquejo de la obra de Galois

La obra original de Galois es difícil; Cauchy y Fourier tuvieron dificultades para entenderla. Por lo tanto, vamos a numerar la serie de pasos demostrados por Galois e ilustrarlos con un ejemplo. Nuestro desarrollo se basa fuertemente en el tratamiento de la misma obra que da Morria Kline en su libro *El pensamiento matemático desde los tiempos antiguos hasta los modernos.*

Paso 1. En primer lugar, dada una ecuación (general o particular) se muestra cómo se puede hallar el grupo G de dicha ecuación en el campo F de sus coeficientes, es decir, el grupo de permutacioines de las raíces que deja invariante toda relación entre las raíces con coeficientes en dicho campo. Galois muestra cómo se puede encontrar G sin conocer las raíces.

En el ejemplo que trataremos, $x^4 + px^2 + q = 0$, no nos encontramos en el caso general porque conocemos las raíces, a saber

$$x_1 = \sqrt{\frac{-p + \sqrt{p^2 - 4q}}{2}}, x_2 = -\sqrt{\frac{-p + \sqrt{p^2 - 4q}}{2}}, x_3 = \sqrt{\frac{-p - \sqrt{p^2 - 4q}}{2}},$$

$$x_4 = -\sqrt{\frac{-p - \sqrt{p^2 - 4q}}{2}}.$$

Llamemos F al campo de los coeficientes de la ecuación. El procedimiento de Galois requiere que se encuentre el grupo de las permutaciones de las raíces x_1, x_2, x_3, x_4 que deja invariante a toda relación sobre F entre dichas raíces, es decir, toda relación que tenga 'coeficientes' en F. Fuera de las relaciones simétricas entre las raíces, podemos destacar las relaciones sobre F

$$x_1 + x_2 = 0; \qquad x_3 + x_4 = 0. \tag{10.5}$$

(Aparentemente éstas son las únicas relaciones entre las raíces de la ecuación, con coeficientes en F, con excepción de las simétricas, pero es difícil demostrar este hecho.)

Ahora bien, por la definición de funciones simétricas, toda permutación que pertenece al grupo S_4 deja invariante toda función simétrica de x_1, x_2, x_3, x_4. El subgrupo de S_4 que también deja invariante a 10.5 es

$$\begin{aligned}
G &= \{e_0 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}, e_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix}, \\
e_2 &= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix}, e_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}, \\
e_4 &= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}, e_5 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}, \\
e_6 &= \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{pmatrix}, e_7 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{pmatrix}\}
\end{aligned}$$

Paso 2. Se procede a buscar el mayor subgrupo propio H de G. Si hay más de uno, se escoge cualquiera. Después de hallar H, usando propiedades del grupo G, por medio de una serie de procedimientos con operaciones racionales (que Galois da teóricamente, es decir, no es un método práctico pero sí se da una forma constructiva de proceder), se halla una función

ϕ de las raíces, con coeficientes en F, tal que ϕ sea invariante bajo toda permutación que pertenece a H pero no sea invariante bajo las demás permutaciones que pertenecen a G. Hay varias de dichas funciones, pero una es suficiente para proceder. El método de Galois consiste en construir una ecuación con coeficientes en F tal que una de sus raíces sea la función ϕ. El grado de dicha ecuación es igual al índice de H en G (de acuerdo con Lagrange (1)), y la ecuación se llama una 'resolvente parcial'. La solución de la ecuación resolvente nos da ϕ y se adjunta ϕ a F para formar un nuevo campo $F_1 = F(\phi)$.

El grupo de la ecuación original con respecto del campo F_1 es precisamente H.

En nuestro ejemplo, ϕ es raíz de la ecuación

$$t^2 - (p^2 - 4q) = 0$$

cuyo grado es 2, que corresponde al índice de H en G. Esto nos da $\phi = \pm\sqrt{p^2 - 4q}$, pero $-\sqrt{p^2 - 4q} = x_1^2 - x_3^2$. Es decir, ϕ expresado como relación de las raíces sobre F es $x_1^2 - x_3^2$. Se comprueba fácilmente que $H = \{e_0, e_1, e_2, e_3\}$ deja ϕ invariante, mientras que e_4, e_5, e_6, e_7 'mueven' ϕ. Para este efecto es necesario tener en cuenta que

$$x_1 + x_2 = 0 \longrightarrow x_1 = -x_2 \longrightarrow x_1^2 = x_2^2$$

y similarmente $x_3^2 = x_4^2$. El campo que tenemos es $F_1 = F(\phi) = F(\sqrt{p^2 - 4q})$ y H es el grupo de la ecuación $x^4 + px + q = 0$ sobre F_1.

Paso 3. Como próximo paso, se busca el mayor subgrupo K de H empleando en la búsqueda conocimiento de la teoría de grupos. Ahora, se puede obtener una función ϕ_1 de las raíces de la ecuación original con coeficientes en F_1 que sea invariante bajo las permutaciones de K pero no bajo las demás permutaciones de H. Para hallar ϕ_1 (sin concer las raíces) se construye una ecuación sobre F_1 que tenga a ϕ_1 como una de sus raíces y cuyo grado sea igual al índice de K en H. Dicha ecuación se llama la segunda resolvente parcial y al solucionarla se halla ϕ_1. Se adjunta ϕ_1 a F_1 para obtener un campo $F_2, F_2 = F_1(\phi_1) = F(\phi, \phi_1)$. Con respecto de F_2 el grupo de la ecuación original $x^4 + px^2 + q = 0$ es K.

En el ejemplo que se está tratando, la ecuación resolvente construida es $t^2 - 2(-p - \sqrt{p^2 - 4q}) = 0$ que tiene por raíz a

$$\phi_1 = x_3 - x_4 = 2\sqrt{\frac{-p - \phi}{2}} = 2\sqrt{\frac{-p - \sqrt{p^2 - 4q}}{2}},$$

de donde,

$$F_2 = F_1(\phi_1) = F_1\left(\sqrt{\frac{-p - \sqrt{p^2 - 4q}}{2}}\right) = F\left(\sqrt{p^2 - 4q}, \sqrt{\frac{-p - \sqrt{p^2 - 4q}}{2}}\right).$$

Ahora, $\phi_1 = x_3 - x_4$, función de las raíces que es invariante únicamente bajo las permutaciones e_0, e_1, es decir, el subgrupo K es $\{e_0, e_1\}$ cuyo índice en H es 2, el grado de la ecuación resolvente que proporciona ϕ_1. Ademas, K es el grupo de la ecuación original sobre el campo F_2.

Paso 4. Se repite el proceso tantas veces sea necesario hasta que se obtenga un campo F_l tal que el grupo de la ecuación original sobre F_l sea $E = (e_0)$.

En el ejemplo que se está desarrollando se requiere apenas un paso más dado que el único subgrupo propio de $K = \{e_0, e_1\}$ es $E = (e_0)$. La ecuación resolvente es, en este caso,

$$t^2 - 2\left(-p + \sqrt{p^2 - 4q}\right) = 0$$

y

$$\phi_2 = x_1 - x_2 = 2\sqrt{\frac{-p + \sqrt{p^2 - 4q}}{2}}.$$

Ahora, dado que $F_3 = F_2(\phi_2)$ contiene a F_2, la única permutación que deja invariante a toda relación entre las raíces sobre F_2 es e_0. Tenemos el siguiente esquema

$$\begin{array}{ccc} G & \longleftrightarrow & F \\ \cup & & \cap \\ H & \longleftrightarrow & F_1 \\ \cup & & \cap \\ K & \longleftrightarrow & F_2 \\ \cup & & \cap \\ E & \longleftrightarrow & F3 \end{array}$$

Paso 5. Luego, Galois demuestra que si el grupo de una ecuación respecto de un campo es E, entonces el campo contiene todas las raíces de dicha ecuación. Además, Galois proporciona un método para hallar las raíces por medio de operaciones racionales en F. Es decir, la obra de Galois contiene un método constructivo para seguir todos los pasos anteriores y hallar el campo donde la ecuación tiene todas sus raíces, llamado campo de descomposición de la ecuación polinómica dada, dado que la ecuación puede ser factorizada en factores lineales sobre dicho campo.

Ahora bien, antes de culminar nuestra consideración de la obra de Galois falta relacionar el procedimiento que hemos descrito con el problem de la solución de la ecuación por radicales.

En primera instancia Galois aisla el concepto de subgrupo normal. Recordamos que el subgrupo H del grupo G se dice normal si $\forall g \in G, gH = Hg$. El resultado principal de Galois es el siguiente. Si la ecuación resolvente que sirve para 'reducir' el grupo de una ecuación, digamos de un grupo G a un subgrupo H, es una ecuación binomial del tipo $x^p - A = 0$ de grado primo p, entonces H es subgrupo normal de G de índice primo p; e inversamente, si H es subgrupo normal de G de índice primo p, entonces la ecuación resolvente correspondiente es una ecuación binomial de grado p (o puede reducirse a una ecuación de este tipo). Si todas las resolventes sucesivas son ecuaciones binomiales, entonces por los resultados de Gauss, se puede resolver la ecuación original por radicales, pues se sabe que se puede pasar del campo original (de los coeficientes) al campo final donde se encuentran todas las raíces, por medio de sucesivas extensiones radicales, es decir, cada vez se adjunta un nuevo elemento de forma radical. Inversamente, si una ecuación es soluble por radicales, entonces el conjunto de ecuaciones radicales debe existir, y éstas han de ser ecuaciones binomiales.

Por otra parte, debido a la equivalencia mencionada, la ecuación es soluble por radicales si el grupo correspondiente contiene una sucesión de subgrupos tal que cada uno es normal en el anterior y de índice primo. Pero esto

describe precisamente un grupo soluble. Luego, la ecuación es soluble por radicales si el grupo respectivo es soluble, y viceversa.

Ahora bien, el proceso de análisis al cual Galois sujeta el tema de campos de extensión y los correspondientes grupos de permutaciones (referentes a la adjunción de sucesivas raíces de una ecuación polinómica) trasciende al problema particular de la solución de ecuaciones por radicales. Es decir, el proceso es generalizable a las raíces de ecuaciones que no son necesariamente solubles por radicales y proporciona una forma adecuada de estudiarlas. Sin embargo, dichas generalizaciones no fueron puestas en claro sino varios años después de la muerte de Galois, cuando el álgebra abstracta y, en particular, la teoría de grupos había avanzado y se había generalizado.

También es importante señalar la relación que existe entre este análisis y las fórmulas de solución por radicales que hemos estudiado. Según la teoría de Galois, estas fórmulas deben depender de una sucesión de ecuaciones binomiales. Para ver esto es conveniente notar que cualquier ecuación cuadrática puede expresarse en forma binomial sencillamente completando el cuadrado.

$$ax^2 + bx + c = 0 \longleftrightarrow \left(x + \frac{b}{2a}\right)^2 - \left(\frac{b^2}{4a} - \frac{c}{a}\right) = 0$$

y, poniendo $y = x + \frac{b}{2a}$ vemos que ésta tiene la formal $y^2 - A = 0$, de grado primo 2.

Ahora, refiriéndonos al análisis que hizo Lagrange de la fórmula de solución por radicales de la ecuación general de grado tres, observamos que allí se encuentran dos ecuaciones resolventes cuyas formas son

$$y^2 - A = 0, \qquad z^3 - B = 0.$$

Punto de discusión

Volviendo sobre nuestra discusión del trabajo de Lagrange identificar específicamente estas dos ecuaciones.

La ecuación general de grado cuatro puede resolverse con cuatro ecuaciones resolventes binomiales. Recordando el método de Ferrari, estas ecuaciones son de grados $2, 3, 2$ y 2 en ese orden.

En el caso particular de la ecuación que hemos venido estudiando, $x^4 + px^2 + q = 0$, las ecuaciones binomiales sucesivas son

$\left(x^2 + \frac{p}{2}\right)^2 = \frac{p^2}{4} - q$	primera ecuación resolvente binomial
$\left(x^2 + \frac{p}{2} = \pm\sqrt{\frac{p^2}{4} - q} = \pm\sqrt{C}\right)$	
$x^2 = \sqrt{C} - \frac{p}{2}$	segunda ecuación resolvente binomial
$x^2 = -\sqrt{C} - \frac{p}{2}$	tercera ecuación binomial resolvente

La serie de descomposición de los índices de los subgrupos formados es $2, 2, 2$.

Estudiando este análisis de Galois, aunque sólo sea un bosquejo, podemos darnos cuenta de inmediato que existen diferencias esenciales entre éste y el tratamiento contemporáneo de la teoría de Galois. La mas importante entre estas diferencias es la ausencia total (en el estudio original) de nociones de espacio vectorial o sea, el tratamiento de los campos también como espacios vectoriales.

10.1.5 Extensiones algebraicas simples

Nuestro propósito en esta sección es comenzar la búsqueda de maneras de enlazar la solución por resolventes tal como la venimos estudiando con la idea de la factorización o descomposición de los polinomios en factores lineales.

Para comenzar esta búsqueda consideremos la ecuación cuadrática $x^2 + a_1x + a_0 = 0$ y el resolvente que obtuvimos en la sección anterior. Teníamos

$$\begin{aligned} r_1 &= \frac{1}{2}\left[(r_1 + r_2) + (r_1 - r_2)\right] = \frac{1}{2}\left[(r_1 + r_2) + \sqrt{(r_1 - r_2)^2}\right] \\ r_2 &= \frac{1}{2}\left[(r_1 + r_2) - (r_1 - r_2)\right] = \frac{1}{2}\left[(r_1 + r_2) - \sqrt{(r_1 - r_2)^2}\right]. \end{aligned}$$

Ahora bien, la idea es que la expresión subradical puede darse en términos de los coeficientes de la ecuación puesto que

$$\sqrt{a_1^2 - 4a_0} = \sqrt{(r_1 + r_2)^2 - 4r_1r_2} = \sqrt{(r_1 - r_2)^2} = \pm(r_1 - r_2).$$

Como los coeficientes son conocidos, conocemos también la resolvente. Por ejemplo, para la ecuación $f(x) = x^2 - 5x + 3 = 0$ tenemos

$$\begin{aligned} r_1 &= \frac{1}{2}\left[-5 + \sqrt{13}\right] \\ r_2 &= \frac{1}{2}\left[-5 - \sqrt{13}\right]. \end{aligned}$$

Nótese que esto corresponde simplemente a la aplicación de la fórmula cuadrática. Considerando que $f(x) \in \mathbb{Q}[x]$, o sea que los coeficientes de la ecuación dada son racionales, si adjuntamos $\sqrt{13}$ a $\mathbb{Q}$, queremos ver que tendremos un nuevo campo que denotaremos $\mathbb{Q}(\sqrt{13})$ en el cual se encuentran las raíces de la ecuación dada y tal que la ecuación puede factorizarse en factores lineales en ese campo. Veamos.

Al decir que adjuntamos $\sqrt{13}$ a $\mathbb{Q}$, queremos decir que consideraremos el conjunto

$$\mathbb{Q}(\sqrt{13}) = \{a + b\sqrt{13} \mid a, b \in \mathbb{Q}\}.$$

El conjunto es cerrado para la adición y la multiplicación pues

$$(a + b\sqrt{13}) + (c + d\sqrt{13}) = (a + b) + (c + d)\sqrt{13}$$

y

$$(a + b\sqrt{13}) \cdot (c + d\sqrt{13}) = (ac + 13bd) + (ad + bd)\sqrt{13}.$$

Es evidente que el elemento idéntico para la adición es $0 + 0\sqrt{13}$ y para la multiplicación es $1 + 0\sqrt{13}$. La única propiedad de los campos que no es inmediata es la existencia de inversos multiplicativos para los elementos no nulos, pero si $a + b\sqrt{13} \neq 0$, tenemos que

$$\frac{1}{a + b\sqrt{13}} \cdot \frac{a - b\sqrt{13}}{a - b\sqrt{13}} = \frac{a - b\sqrt{13}}{a^2 - 13b^2} = \frac{a}{a^2 - 13b^2} + \frac{-b}{a^2 - 13b^2}\sqrt{13} \in \mathbb{Q}(\sqrt{13}).$$

Por otra parte, nuestra afirmación sobre la factorización resulta inmediata, pues

$$f(x) = x^2 - 5x + 3 = \left(x - \left(\frac{-5 + \sqrt{13}}{2}\right)\right)\left(x - \left(\frac{-5 - \sqrt{13}}{2}\right)\right),$$

donde $\frac{-5+\sqrt{13}}{2}, \frac{-5-\sqrt{13}}{2}$ pertenecen a $\mathbb{Q}(\sqrt{13})$. Nótese que esta extensión corresponde a la adjunción a $\mathbb{Q}$ de un cero del binomio $x^2 - 13$.

Puntos de discusión

1. Comparar este procedimiento con la adjunción formal descrita en el Teorema de Kronecker (Capítulo 7).

2. Determinar el campo de extensión sobre $\mathbb{Q}$ que contiene las raíces de $x^2 + 1 = 0$.

Como hemos notado, la meta del álgebra clásica es la solución de ecuaciones, en particular, la posibilidad de expresar las raíces de la ecuación general

$$x^n + a_{n-1}x^{n-1} + ... + a_1x + a_0 = 0, \tag{10.6}$$

en términos de los coeficientes $a_0, a_1, ..., a_{n-1}$ usando un número finito de operaciones $+, -, \times, \div$ y radicales $\sqrt{}, \sqrt[n]{}$.

Ahora bien, el conjunto de elementos que se obtienen a partir de $\mathbb{Q}$ y $a_0, a_1, ..., a_{n-1}$ es el campo $\mathbb{Q}(a_0, a_1, ..., a_{n-1})$. Si se denotan las raíces de (10.6) por $r_1, r_2, ..., r_n$, de modo que

$$(x - r_1)(x - r_2) \cdots (x - r_n) = x^n + a_{n-1}x^{n-1} + ... + a_1x + a_0,$$

se tiene que $a_0, a_1, ..., a_{n-1}$ son funciones polinómicas de $r_1, r_2, ..., r_n$, llamadas las funciones simétricas elementales.

$$a_0 = (-1)^n r_1 r_2 \cdots r_n, ..., a_{n-1} = -(r_1 + r_2 + ... + r_n).$$

El objetivo de la solución por radicales de una ecuación polinómica es el de extender $\mathbb{Q}(a_0, a_1, ..., a_{n-1})$ por medio de la adjunción de radicales hasta que se construya un campo que contenga todas las raíces $r_1, r_2, ..., r_n$.

Por ejemplo, las raíces r_1, r_2 de la cuadrática $x^2 + a_1x + a_0 = 0$ se encuentran en la extensión de $\mathbb{Q}(a_0, a_1) = \mathbb{Q}(r_1r_2, r_1 + r_2)$ adjuntando

$$\sqrt{a_1^2 - 4a_0} = \sqrt{(r_1 + r_2)^2 - 4r_1r_2} = \sqrt{(r_1 - r_2)^2} = \pm(r_1 - r_2).$$

En este caso, $\mathbb{Q}(r_1, r_2)$ es el mismo la extensión radical de $\mathbb{Q}(a_0, a_1, \sqrt{a_1^2 - 4a_0})$, pero en otros casos, la extensión radical de $\mathbb{Q}(a_0, a_1, ..., a_{n-1})$ que contiene $r_1, r_2, ..., r_n$ es mayor que $\mathbb{Q}(r_1, r_2, ..., r_n)$. Por ejemplo, en el caso de la ecuación cúbica, que ya hemos estudiado, la solución arroja una extensión radical de $Q(a_0, a_1, a_2)$ que incluye las raíces cúbicas de la unidad además de r_1, r_2, r_3.

Puntos de discusión

1. Distinguir entre los pormenores de la solución de la ecuación cuadrática y la cúbica que dan lugar a las diferencias que notamos en esta última afirmación.

2. Comparar la situación que aquí se describe con las resolventes de Lagrange y Vandermonde.

10.1.6 Extensiones algebraicas

En general, el adjuntar un elemento a un campo K significa formar la clausura de $K \cup \{u\}$ bajo las operaciones de $+, -, \times, \div$ (por un elemento diferente de 0, por supuesto), es decir, tomar la intersección de todos los campos que contienen a $K \cup \{u\}$. La adjunción se dice radical si alguna potencia entera u^m de u pertenece a K. Es decir, si existe $t \in K$ tal que $u^m = t$, en cuyo caso $u = \sqrt[m]{t}$. El resultado de adjunciones sucesivas $K(u_1)(u_2)...(u_n)$ se denota por $K(u_1, u_2, ..., u_n)$ y si cada una de las adjunciones es radical, decimos que $K(u_1, u_2, ..., u_n)$ es una extensión radical de K.

Puntos de discusión

1. Sean $K = \mathbb{Q}(\sqrt{6})$, $L = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. Demostrar que K es subcampo de L y hallar $[L : K]$.
2. Hallar $[\mathbb{Q}(\sqrt[3]{2}) : Q(\sqrt{2})]$.
3. Hallar $[\mathbb{Q}(\sqrt{1+\sqrt{5}}) : \mathbb{Q}(\sqrt{5})]$.
4. Hallar $[\mathbb{Q}(\sqrt{1+\sqrt[3]{2}}) : \mathbb{Q}(\sqrt[3]{2})]$.

Antes de seguir nuestro camino principal, consideremos una serie de teoremas que tratan de manera general la adjunción sucesiva de elementos a un campo base.

Teoremas sobre las extensiones de dimensión finita

Sean K un campo, L un campo que contiene a K. Entonces L se puede pensar como un espacio vectorial sobre K. La dimensión de este espacio vectorial se llama la dimensión (o el grado) de L sobre K y se denota $[L : K]$. Decimos que L es de dimensión finita o infinita sobre K según $[L : K]$ sea finita o infinita.

Teorema 10.1 *Sean K, L, M campos con $K \subset L \subset M$. Entonces $[M : K]$ es finita si y sólo si tanto $[M : L]$ como $[L : K]$ son finitas, y en este caso, $[M : K] = [M : L][L : K]$.*

Demostración. Supongamos primero que $[M : K]$ es finita. Como L es subespacio vectorial de M, considerado como espacio vectorial sobre K, $[L : K]$ también es finita. Ahora, cualquier base que genera a M sobre K también genera a M sobre L, dado que $K \subset L$; entonces $[M : L]$ es también finita.

Ahora, supongamos que $[M : L] = n$ y $[L : K] = m$ con n, m finitos. Queremos mostrar que $[M : K] = mn$, y es por lo tanto finita.

Sean $u_1, u_2, \ldots, u_n$ una base para M sobre L y $v_1, v_2, \ldots, v_m$ una base para L sobre K. Afirmamos que los mn elementos $u_i v_j (i = 1, \ldots, n; j = 1, \ldots, m)$ forman una base de M sobre K. Hay que demostrar que estos elementos generan a M sobre K y que son linealmente independientes en M sobre K.

Primero, sea $z \in M$, entonces

$$z = \sum_i b_i u_i$$

donde los b_i pertenecen a L. Además, cada b_i puede expresarse como

$$b_i = \sum_j c_{ij} v_j$$

donde los $c_{ij} \in K$, puesto que los v_j generan L sobre K. Se sigue que

$$z = \sum_{i,j} c_{ij} u_i v_j \quad i = 1, \ldots, n; j = 1, \ldots, m.$$

Por lo tanto, los $u_i v_j$ generan a M sobre K.

Para demostrar que son linealmente independientes sobre K, supongamos que

$$\sum_{i,j} c_{ij} u_i v_j = 0, \text{ con } c_{ij} \in K.$$

Hay que demostrar que todos los c_{ij} son iguales a cero. Ahora, como $b_i = \sum_j c_{ij} v_j$ con $b_i \in L$, tenemos

$$\sum_i b_i u_i = 0.$$

Pero los u_i son linealmente independientes sobre L, entonces cada $b_i = 0$. Es decir,

$$\sum_j c_{ij} v_j = 0.$$

Pero los v_j son linealmente independientes sobre K, por lo tanto, $c_{ij} = 0$, para cada c_{ij}, y los $u_i v_j$ se han mostrado linealmente independientes en M sobre K.

Ejemplo Consideremos el elemento $u = \sqrt{2} + \sqrt{3}$, es fácil ver satisface una ecuación de cuadrática sobre el campo $\mathbb{Q}(\sqrt{2})/$. Basta despejar $u - \sqrt{2}) = \sqrt{3}$, elevar al cuadrado para pbtener

$$(u - \sqrt{2})^2 = 3$$

que simplifica a la ecuación

$$u^2 - 2\sqrt{2}u - 1 = 0.$$

Por otra parte, el elemento $v = \sqrt{2}$ satisface la ecuación cuadrática $x^2 - 2 = 0$ sobre $\mathbb{Q}$. Si $K = \mathbb{Q}$, $M = \mathbb{Q}(\sqrt{2})$ y $L = M(\sqrt{2} + \sqrt{3})$, tenemos $L \supseteq M \supseteq K$.

Puntos de discusión

1. Investigar la dimensión de L sobre K.
2. Hallar un polinomio de $\mathbb{Q}[x]$ uno de cuyos ceros sea $\sqrt{2} + \sqrt{3}$.

Subcampo generado por un subconjunto

Sean M cualquier campo y S un subconjunto de $M, S \neq \emptyset$. Está claro que hay un único subcampo mínimo de M que contiene a S, a saber, la intersección de todos los subcampos de M que contienen a S. Esta construcción nos interesa particularmente cuando S consta de un subcampo K de M mas

un elemento adicional u de M. Hemos denotado el subcampo así generado por $K(u)$. Se pueden presentar dos casos.

Caso 1. No existe ningún polinomio f con coeficientes en K (diferente del polinomio que es idénticamente 0) tal que $f(u) = 0$. En este caso, se dice que u es *trascendente* sobre K. Además, es evidente que el campo $K(u)$ es el campo de todas las funciones racionales en u (cocientes de polinomios en u) con coeficientes en K, donde u se comporta exactamente como un indeterminado sobre K.

Nota. Tal como lo habíamos anotado, éste es el concepto de campo de extensión que se halla en la obra original de Galois.

Caso 2. Sí existe un polinomio f con coeficientes en K tal que $f(u) = 0$. En este caso se dice que u es *algebraico* sobre K. Una descripción del campo $K(u)$, para este caso, está contenida en el teorema siguiente.

Teorema 10.2 *Sean K un campo cualquiera, u un elemento de un campo mayor. Supongamos que u es algebraico sobre K. Sea f un polinomio mónico con coeficientes en K de menor grado tal que $f(u) = 0$, y sea ese grado mínimo n. Entonces*

1. *f es único;*
2. *f es irreducible sobre K;*
3. *$1, u, u^2, \ldots, u^{n-1}$ forman una base del espacio vectorial $K(u)$ sobre K.*
4. *$[K(u) : K] = n$;*
5. *Un polinomio g con coeficientes en K satisface que $g(u) = 0$ si y sólo si g es múltiplo de f.*

Antes de proceder con la demostración anotamos que el enunciado del teorema está afirmando que si u es algebraico sobre K, entonces satisface un polinomio mónico de grado mínimo sobre K. Está claro que si $f(u) = 0$ con f no mónico de grado n, basta dividir por el coeficiente a_n del término en x^n para obtener un polinomio mónico que u satisface. Además, la condición de que exista un f de grado mínimo sigue del principio de la buena ordenación.

Demostracion.

1. Si f_0 es otro polinomio mónico de grado n tal que $f_0(u) = 0$, entonces, poniendo $f_1 = f - f_0$ obtenemos que $f_1(u) = 0$, donde f_1 tiene grado menor que n. Si $f_1 \not\equiv 0$, esto contradice la escogencia de f de menor grado. (Es claro que una multiplicación por un elemento de K hará que f_1 sea mónico.) Entonces, $f_1 \equiv 0$ y $f = f_0$, es decir, f es único.

2. Si $f = f_0 f_1$ donde f_0, f_1 son polinomios de grado menor que n con coeficientes en K, entonces $f(u) = f_0(u) f_1(u) = 0$, de donde, f_0 o f_1 tiene a u por raíz y, de nuevo, se tiene una contradicción de la escogencia mínima de f.

3. La existencia de una relación lineal de $1, u, \ldots, u^{n-1}$ igual a cero con coeficientes en K implicaría la existencia de g tal que $g(u) = 0$ donde g es un polinomio de grado menor que n. Por lo tanto, $1, u, \ldots, u^{n-1}$ son

linealmente independientes sobre K. Debemos mostrar, además, que generan a $K(u)$. Para tal efecto, llamemos T al subespacio vectorial de $K(u)$ generado por $1, u, \ldots, u^{n-1}$. Si demostramos que T es un campo, tendremos que $T = K(u)$ puesto que, al permitir la multiplicación de elementos de T, se tendrá que $u^k \in T, \forall k \in \mathbb{N}$. Además, $K \subset T$, de donde se sigue que $T = K(u)$ por definición de $K(u)$.
Primero mostraremos que $u^k \in T, \forall k \in \mathbb{N}$. Ya sabemos que es cierto hasta $k = n - 1$. Ahora, supongamos que sea cierto para un entero positivo $m - 1$. Entonces,

$$u^{m-1} = \alpha_0 + \alpha_1 u + \cdots + \alpha_{n-1} u^{n-1}, \qquad \alpha_i \in K.$$

Multiplicando ambos miembros de esta ecuación por u, obtenemos

$$u^m = \alpha_0 u + \alpha_1 u^2 + \cdots + \alpha_{n-1} u^n.$$

Pero u^n es una combinación lineal de $1, u, \ldots, u^{n-1}$ en virtud de la ecuación $f(u) = 0$. Entonces, $u^m \in T$ y el resultado sigue por inducción.
El resultado anterior nos dice que T es un anillo, y en realidad sabemos que T es un dominio de integridad por estar contenido en $K(u)$. Falta demostrar que cualquier elemento $z \neq 0$ en T tiene un inverso multiplicativo en T. Ahora bien, $z = h(u)$ donde h es un polinomio no trivial de grado menor que n. Como f es irreducible y h es de grado menor que el grado de f, se sigue que el máximo común divisor de f y h es 1, pues f es mónico. Entonces, existen polinomios r, s tales que

$$rf + sh = 1,$$

de donde, $(rf + sh)(u) = 1(u) = 1$. Luego, $s(u)h(u) = 1$ (¿por qué?) y $s(u)$ es el inverso de $z = h(u)$.

4. Esto sigue directamente de (3).

5. Si g no fuera múltiplo de f, donde $g(u) = 0$, como f es irreducible tendríamos $\text{mcd}(f, g) = 1$. Entonces, existirían polinomios r, s tales que $rf + sg = 1$ y $(rf + sh)(u) = 1(u) = 1$, que implica que $s(u)g(u) = 1$. Pero $g(u) = 0$, que es una contradicción. Luego, g es múltiplo de f. Claramente, si g es múltiplo de f entonces $g(u) = 0$.

Diremos que un campo L que contiene a un campo K es *algebraico* sobre K si todo elemento de L es algebraico sobre K; de otro modo diremos que L es trascendente sobre K. Si L es una extensión de dimensión finita sobre K, es evidente que L es algebraico sobre K, pues si $[L : K] = n$ y si $u \in L$, entonces $1, u, \ldots, u^n$ son linealmente dependientes sobre K y esto nos proporciona un polinomio sobre K que u satisface.

Es posible que una extensión de dimensión infinita sobre K sea algebraica sobre K. En particular, si $K = \mathbb{Q}$ y A es el conjunto de todos los números complejos que son algebraicos sobre $\mathbb{Q}$, es posible demostrar que A es un campo algebraico sobre $\mathbb{Q}$ pero cuya dimensión sobre $\mathbb{Q}$ no es finita.

Se dice que el campo $K(u)$ es el campo que se obtiene *adjuntando* u a K, donde hasta el momento requerimos que u pertenezca a un campo que contiene a K. La dimensión de $K(u)$ sobre K se llama el *grado* de u sobre K y el polinomio cuya existencia se demuestra en el Teorema 10.2 se llama el *polinomio irreducible* de u sobre K.

Teorema 10.3 *Sean L y M subcampos de un campo N y supongamos que tanto L como M contienen un campo K. Si escribimos*

$$[L:K]=m,[M:K]=n,[L\cup M:K]=t$$

entonces

1. *t es finito si y sólo si tanto m como n son finitos;*
2. *En este caso, t es múltiplo de n y de m, y además, $t\leq mn$.*
3. *Si $(m,n)=1$, entonces $t=mn$.*

Demostración. Si t es finito, entonces también lo son m y n porque L y M son subcampos de $L\cup M$. Supongamos, entonces, que tanto m como n son finitos. Demostraremos que t es finito y a lo sumo mn por inducción sobre n. El caso $n=1$ es trivial, pues entonces $M=K, L\cup M=L$, y

$$t=[L\cup M:K]=[L:K]=m=m\cdot 1=m\cdot n.$$

Entonces sea $n>1$ y supongamos que el teorema se tiene para todo $k<n$.

Como $n>1$ existe $u\in M$ tal que $u\notin K$. Sean r el grado de u sobre K y s el grado de u sobre L. Tenemos que $s\leq r$ puesto que el polinomio irreducible de u sobre K es múltiplo del polinomio irreducible de u sobre L (Teorema *.*). Ahora bien, por el Teorema *.1 tenemos

$$[L(u):K]=[L(u):L][L:K]=sm$$

y

$$ms=[L(u):K]=[L(u):K(u)][K(u):K],$$

de donde, $[L(u):K(u)]=ms/r$.

Además,

$$[M:K(u)][K(u):K]=[M:K]\Longrightarrow[M:K(u)]=\frac{n}{r}<n.$$

Aplicando la hipótesis de inducción a los campos $L(u)$ y M sobre $K(u)$ deducimos que

$$[L(u)\cup M:K(u)]\leq\frac{mns}{r^2}\leq\frac{mn}{r}.$$

Pero el campo $L(u)\cup M$ es el mismo que $L\cup M$ puesto que $u\in M$. Por consiguiente,

$$t=[L\cup M:K]=[L\cup M:K(u)][K(u):K]\leq\frac{mn}{r}\cdot r=mn,$$

que nos da $t\leq mn$. Ahora,

$$[L\cup M:K]=[L\cup M:L][L:K]\Longrightarrow t=xm\Longrightarrow m\mid t$$

y

$$[L\cup M:K]=[L\cup M:M][M:K]\Longrightarrow t=yn\Longrightarrow n\mid t.$$

Combinando $m\mid t, n\mid t, t\leq mn$, si $(m,n)=1$, se tiene que $t=mn$, y se ha completado la demostración del teorema.

Si $L=K(u), M=K(v)$, $L\cup M$ tiene una forma mas explícita, pues en este caso $L\cup M=K(u,v)$ el cual puede pensarse como el resultado de adjuntar

v a $K(u)$, el resultado de adjuntar u a $K(v)$, o el resultado de adjuntar u y v a K.

En general podemos hablar de la adjunción de un conjunto S de (un número arbitrario de elementos nuevos) $u, v, w \ldots$ a un campo dado K si se ha construido una extensión del campo K que contiene a S. Nos interesa pensar en la menor extensión tal y su existencia está garantizada definiéndola como la intersección de todos los campos que contienen a K y a S. De esta manera se pretende construir un campo que contenga todos los ceros de un polinomio dado y, por ende, en el cual el polinomio puede descomponerse en factores lineales. Tal campo se llama *campo de descomposición* del polinomio.

Puntos de discusión

1. Sean $K = \mathbb{Q}$, y sean u_1, u_2 dos ceros distintos del polinomio x^3+6x+2. Sean $K_1 = K(u_1)$ y $K_2 = K_1(u_2)$. Determinar $[K_1 : K]$,$[K_2 : K_1]$ y $[K_2 : K]$.
2. Sea v un cero de $x^2 + 3$. Demostrar que v puede escogerse en K_2 de tal modo que $K' = K(v)$ es subcampo de K_2. Hallar $[K_2 : K']$ y $[K' : K]$.

Formalizaremos estas ideas en el siguiente aparte.

10.1.7 Campos de descomposición

Hasta el momento no hemos mostrado ningún método constructivo para obtener una extensión finita de un campo dado. Para tal efecto, estudiaremos nuevamente polinomios y los campos donde se encuentran sus raíces. Comenzaremos por recordar el Teorema de Kronecker.

Teorema 10.4 *Sea f un polinomio irreducible de grado $n \geq 1$ con coeficientes en un campo K. Entonces existe un campo que contiene a K y a una raíz de f cuya dimensión sobre K es n.*

Demostración. Sea $K[x]$ el anillo de polinomios en x sobre K y sea $A = \langle f(x) \rangle$ el ideal de $K[x]$ generado por $f(x)$. Aplicando varios teoremas de la teoría de anillos euclidianos, A es un ideal maximal de $K[x]$ y $L = K[x]/A$ es un campo.

Ahora bien, L contiene a $L_0 = \{k + A \mid k \in K\}$, el cual es isomorfo a K, pues si consideramos la aplicación ψ de $K[x]$ en $K[x]/A$ definida por

$$\psi(g(x)) = g(x) + A,$$

es claro que, considerando la restricción de ψ a K, es decir, a los polinomios constantes en $K[x]$, que son de la forma $g(x) = k, k \in K$, tenemos

$$\psi(k) = k + A.$$

Queremos mostrar que $L_0 = K$, o sea, ψ es homomorfismo, 1-1 y sobre. Ahora bien, ψ es trivialmente sobre, pues $\forall l_0 \in L_0, l_0 = k_0 + A$ con $k_0 \in K$, de donde, $l_0 = \psi(k_0)$.

Por otra parte, ψ es 1-1 puesto que si $l_1 = k_1 + A$ y $l_2 = k_2 + A$ y si $l_1 = l_2$ entonces $k_1 + A = k_2 + A$ y $k_2 - k_1 \in A$. Ahora, si $k_1 \neq k_2$ tenemos que

$k_1 - k_2$ es un polinomio de grado 0, pero los elementos de A son múltiplos de $f(x)$ cuyo grado es n. Es decir, son de grado $\geq n$. Luego, $k_1 - k_2$ debe ser el polinomio 0, es decir, $k_1 = k_2$. Se sigue que ψ es 1-1.

Además, ψ es (trivialmente) un homomorfismo, dado que

$$\begin{aligned} \psi(k_1 + k_2) &= (k_1 + k_2) + A = (k_1 + A) + (k_2 + A) = \psi(k_1) + \psi(k_2). \\ \psi(k_1 k_2) &= (k_1 k_2) + A = (k_1 + A)(k_2 + A) = \psi(k_1)\psi(k_2). \end{aligned}$$

De esta manera, como $L_0 \approx K, L$ puede considerarse como una extensión de K.

Ahora, falta mostrar que $f(x)$ tiene una raíz en el campo $L = K[x]/A$. Para ello, recordemos que si $g(x) \in K[x]$, entonces por propiedades de un anillo euclidiano, $\exists q(x), r(x) \in K[x]$ tales que

$$g(x) = q(x)f(x) = r(x), \text{ donde } r(x) \equiv 0, \text{ o, } deg(r(x)) < deg(f(x)).$$

Entonces, para cualquier $g(x) \in K[x], g(x) + A \in L = K[x]/A$ y

$$g(x)+A = q(x)f(x)+r(x)+A = r(x)+A \text{ donde } r(x) \equiv 0, \text{ o, } deg(r(x)) < deg(f(x)),$$

dado que $q(x)f(x) \in A$. Es decir, $L = \{r(x) + A \mid r(x) \in K[x], \text{ y } r(x) \equiv 0, \text{ o, } deg(r(x)) < n\}$. Ahora bien, es claro que poniendo $t = x + A$, tendremos que t es raíz de f en L porque

$$f(t) = f(x + A) = f(x) + A = A, \text{ el 'cero' del campo } L,$$

donde recordamos que los ideales 'absorben' productos. Podemos demostrar, además, que $[L : L_0] = n$ (donde recordamos que $L_0 \approx K$). Para tal efecto, consideremos las clases laterales

$$1 + A, x + A, \ldots, x^{n-1} + A.$$

Queremos mostrar que éstas constituyen una base de L sobre K. Es claro que si $l \in L, l = r(x) + A$ donde $deg(r(x)) < n$. Luego, si $r(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1}, a_j \in K$,

$$l = r(x) + A = a_0(1 + A) + a_1(x + A) + \cdots + a_{n-1}(x^{n-1} + A), a_j \in K.$$

Luego , las clases laterales bajo consideración generan L sobre K. Además son linealmente independientes en L sobre K puesto que si

$$\begin{aligned} k_0(1 + A) + k_1(x + A) + \cdots \quad &+ \quad k_{n-1}(x^{n-1} + A) = A \\ &\Longrightarrow \quad k_0 + k_1 x + \cdots + k_{n-1}x^{n-1} + A = A \\ &\Longrightarrow \quad k_0 + k_1 x + \cdots + k_{n-1}x^{n-1} \in A. \end{aligned}$$

Pero, nuevamente, los elementos de A son múltiplos de un polinomio $f(x)$ de grado n. Entonces, $k_0 + k_1 x + \cdots + k_{n-1}x^{n-1} \equiv 0$ y $k_0 = k_1 = \cdots = k_{n-1} = 0$. Se sigue que los elementos $1 + A, x + A, \ldots, x^{n-1} + A$ son linealmente independientes sobre K.

Corolario 10.1 *Sea $f(x) \in K[x]$. Entonces existe un campo L que contiene a K y a una raíz de f. Además, $[L : K] \leq deg(f(x))$.*

Demostración. Basta considerar los factores irreducibles de f (sobre K). Según el teorema anterior, existe un campo que contiene a K y a una raíz de alguno de los factores irreducibles de f. Pero ésta es también raíz de f.

Teorema 10.5 *Sea $f(x) \in K[x]$ un polinomio de grado $n \geq 1$. Entonces existe una extensión L de K en la cual $f(x)$ tiene n raíces (teniendo en cuenta raíces múltiples). Además, $[L : K] \leq n!$.*

Demostración. Por el corolario anterior existe una extensión L_1 de K donde $f(x)$ tiene una raíz u tal que $[L_1 : K] \leq n$. En el campo L_1 podemos escribir $f(x)(x-u)g(x)$ y sabemos que $deg(g(x)) = n-1$.

Podemos aplicar inducción sobre el grado del polinomio $g(x)$. Entonces, existe una extensión L de L_1 donde $g(x)$ tiene $n-1$ raíces y tal que $[L : L_1] \leq (n-1)!$ Ahora bien, una raíz de $f(x)$ es u o es raíz de $g(x)$. Entonces en $L, f(x)$ tiene n raíces y

$$[L : K] = [L : L_1][L_1 : K] \leq n(n-1)! = n!. \tag{10.7}$$

Definición 10.2 *Sea $f(x) \in K[x]$. Decimos que M es campo de descomposición de f sobre K si f se factoriza completamente en M, es decir, si f se factoriza en factores lineales en M. Además, $M = K(u_1, \ldots, u_r)$ donde los u_j son todas las raíces de f.*

El teorema 10.5 nos dice que todo polinomio tiene un campo de descomposición.

Puntos de discusión

1. El polinomio $f(x) = x^2 - 3$ es irreducible sobre $\mathbb{Q}$. Comparar la estrcutura del campo generado por $1 + (x^2 - 3)$ y $x + (x^2 - 3)$ y el campo generado sobre $\mathbb{Q}$ por $1, \sqrt{3}$.

2. Sean $K = \mathbb{Q}$, u un cero del polinomio $f(x) = x^2 - 3x + 1$. Entonces $K(u)$ ya es un campo de descomposición. Demostrar que los otros dos ceros de $f(x)$ son u^-2 y $-u^2 - u + 2$.

3. Demostrar que el polinomio $x^3 + 6x + 2$ es irreducible sobre $\mathbb{Q}$ y que si u_1, u_2 son dos ceros distintos del polinomio, entonces el menor campo de descomposición es $\mathbb{Q}(u_1)1, u_2)$.

4. Contestar las mismas preguntas del punto 3 para los polinomios

   (a) $x^3 - 3x^2 + 3x - 17$.

   (b) $x^3 - 2$.

   (c) $x^3 - 3x + 1$. (Demostrar en este caso que $\mathbb{Q}(u_1, u_2) = \mathbb{Q}(u_1)$.)

5. Determinar cuáles de los siguientes polinomios son irreducibles sobre $\mathbb{Q}$ y determinar el menor campo de descomposición para los reducibles.

   (a) $x^3 - 5x^2 + 3x + 1$

   (b) $x^3 - 2x + 1$

   (c) $x^3 - x^2 - 8x + 12$

   (d) $x^3 - 5x^2 - 2x + 7$

10.1.8 El grupo de Galois

Sea K un subcampo de un campo L (equivalentemente, L una extensión de K). Sea

$$G = \{\text{automorfismos } \alpha \text{ de } L \mid \alpha(x) = x, \forall x \in K\}.$$

G es subgrupo del grupo de todos los automorfismos de L puesto que si $\alpha, \beta \in G$, entonces $\forall x \in K$

$$\beta^{-1}\alpha(x) = \beta^{-1}(\alpha(x)) = \beta^{-1}(x) = x,$$

de donde, $\beta^{-1}\alpha \in G$. G se llama el *grupo Galois* de L sobre K y se denota por $Gal(L/K)$. Para estudiar el grupo Galois requerimos, en primer lugar, de dos teoremas técnicos que enunciaremos sin demostración.

Teorema 10.6 *Sean G un grupo y K un campo. Supongamos que $\chi_1, \chi_2, \ldots, \chi_n$ son homomorfismos distintos de G en K^*, el grupo multiplicativo de elementos distintos de cero en K. Entonces $\chi_1, \chi_2, \ldots, \chi_n$ son linealmente independientes sobre K.*

Teorema 10.7 *Sean K un campo, $\rho_1, \rho_2, \ldots, \rho_n$ homomorfismos distintos, uno-a-uno de K en un campo L. Sea*

$$K_0 = \{x \in K \mid \rho_1(x) = \rho_2(x) = \cdots = \rho_n(x)\}.$$

Entonces

1. K_0 es subcampo de K;
2. $[K : K_0] \geq n$).

Este resultado es particularmente interesante en el caso en el cual $L = K$ como veremos en el siguiente desarrollo.

Automorfismos de un campo

Sea G un grupo de automorfismos de un campo K (no necesariamente todos). Definimos

$$K^G = \{x \in K \mid g(x) = x, \forall g \in G\}.$$

K^G es subcampo de K puesto que $\forall x, y \in K^G$,

(a) $g(x - y) = g(x) - g(y) = x - y, \forall g \in G$. Entonces, $(x - y) \in K^G$

(b) Si $y \neq 0, g(xy^{-1}) = g(x)(g(y))^{-1} = xy^{-1}, \forall g \in G$. Entonces $xy^{-1} \in K^G$.

Recordando la definición de grupo Galois, tenemos que

$$G \subseteq Gal(K/K^G).$$

Nota. Los elementos de G son automorfismos de K que fijan el subcampo K^G, pero pueden no ser todos, por ello, escribimos la contenencia y no la igualdad estricta.

Además, si L es una extensión de K, tenemos que $L^{Gal(L/K)} \supseteq K$.

Teorema 10.8 *Sea G un grupo finito de automorfismos de un campo K. Entonces, $[K : K^G] = \mid G \mid$, el orden de G.*

Demostración. Recordamos que $K^G = \{x \in K \mid g(x) = x, \forall g \in G\}$. Si denotamos por i el elemento idéntico de G (automorfismo idéntico), podemos escribir

$$K^G = \{x \in K \mid g(x) = i(x), \forall g \in G\}.$$

Aplicando el Teorema 10.7, obtenemos $[K : K^G] \geq n$, donde $n = \mid G \mid$. Nos queda mostrar que $[K : K^G]$ no puede ser mayor que n, y esto lo haremos por contradicción. Supongamos que $[K : K^G] > n$. Entonces existen $a_1, a_2, \ldots, a_{n+1}$ en K que son linealmente independientes sobre K^G. Luego, el sistema de ecuaciones

$$\sum_{i=1}^{n+1} g(a_i)x_i = 0, \qquad g \in G$$

tiene mas variables $(n+1)$ que ecuaciones (n, uno para cada elemento de G), y entonces, tiene una solución no trivial en K. Entre las soluciones no triviales escogemos la que tiene el menor número posible de x_i diferentes de 0. Acordándonos que $a_1, a_2, \ldots, a_{n+1}$ es una base, se puede reordenar si es necesario de tal manera que podemos suponer que en cualquier solución no trivial por lo menos r de los términos son diferentes de 0, es decir,

$$x_1, \ldots, x_r \neq 0; x_{r+1} = \cdots = x_{n+1} = 0.$$

Ahora, multiplicando esta solución por x_1^{-1} obtenemos nuevamente una solución, de tal modo que podemos suponer además que $x_1 = 1$. Tenemos

$$\sum_{i=1}^{r} g(a_i)x_i = 0, \qquad \forall g \in G. \tag{10.8}$$

Ahora, sea h cualquier automorfismo en G.

$$h\left(\sum_{i=1}^{r} g(a_i)x_i\right) = h(0) = 0,$$

y por propiedades de los automorfismos

$$\sum_{i=1}^{r}((hg)(a_i))h(x_i) = 0, \forall g \in G.$$

Cuando g recorre todos los valores en G, hg también recorre estos mismos valores, o sea, esta ultima expresión equivale a

$$\sum_{i=1}^{r}(g(a_i))h(x_i) = 0, \forall g \in G. \tag{10.9}$$

De 10.10 y 10.11 se tiene

$$\sum_{i=1}^{r}(g(a_i))(h(x_i) - x_i) = 0, \forall g \in G.$$

Como $x_1 = 1$ entonces $h(x_1) - x_1 = 0$ y tenemos que $h(x_i) - x_i, i = 1, \ldots, r$ es una solución de la ecuación original con menos que r términos diferentes

de 0. Esto contradice la elección de r, a no ser que se trate de una solución trivial. Es decir, $h(x_i) = x_i, \forall h \in G$. Esto es, $x_i \in K^G, \forall i = 1, \ldots, n+1$.

Ahora bien, en el sistema original de ecuaciones pongamos $g = i$. Obtenemos la ecuación

$$\sum_{i=1}^{n+1} x_i a_i = 0, \quad \text{con } x_i \in K^G,$$

la cual tiene solución no trivial. Pero esto contradice la independencia lineal de $a_1, a_2, \ldots, a_{n+1}$ en K sobre K^G. Como no pueden existir $n+1$ elementos de K que sean linealmente independientes sobre $K^G, [K : K^G]$ no puede ser mayor que n. De allí se sigue que $[K : K^G] = n = \mid G \mid$.

10.1.9 Automorfismos de las extensiones radicales

Retomando el hilo principal de la discusión, es claro, a partir de nuestras definiciones, que cualquier extensión radical E de $\mathbb{Q}(a_0, a_1, ..., a_{n-1})$ que contiene a $r_1, r_2, ..., r_n$ es también una extensión radical de $\mathbb{Q}(r_1, r_2, ..., r_n)$. Ahora la propiedad mas sobresaliente de $\mathbb{Q}(r_1, r_2, ..., r_n)$ es que es simétrica con respecto a $r_1, r_2, ..., r_n$ en el sentido en que cualquier permutación σ de $r_1, r_2, ..., r_n$ puede extenderse a una biyección σ de $\mathbb{Q}(r_1, r_2, ..., r_n)$ definida por

$$\sigma f(r_1, r_2, ..., r_n) = f(\sigma r_1, \sigma r_2, ..., \sigma r_n)$$

para cada función racional f de $r_1, r_2, ..., r_n$. Además, esta biyección σ claramente satisface que

$$\begin{aligned} \sigma(f+g) &= \sigma f + \sigma g, \\ \sigma(fg) &= \sigma f \cdot \sigma g, \end{aligned}$$

de donde, σ es un automorfismo de $\mathbb{Q}(r_1, r_2, ..., r_n)$.

Una extensión radical E de $\mathbb{Q}(r_1, r_2, ..., r_n)$ no es necesariamente simétrica en este sentido. Por ejemplo, la extensión $\mathbb{Q}(r_1, r_2, ..., r_n, \sqrt{r_1})$ contiene una raíz cuadrada de r_1 pero no de r_2. Se sigue que no hay ningún automorfismo que intercambia r_1 con r_2. Pero es posible reponer la simetría si se adjuntan $\sqrt{r_2}, \sqrt{r_3}, ..., \sqrt{r_n}$ además de $\sqrt{r_1}$. Esto lo consignamos en el siguiente teorema.

Teorema 10.9 *Para cada extensión radical E de $\mathbb{Q}(r_1, r_2, ..., r_n)$ existe una extensión radical $\overline{E} \supseteq E$ con automorfismos σ que extienden todas las permutaciones de $r_1, r_2, ..., r_n$.*

Demostración. Para cada elemento que se adjunta en la extensión, representado por la expresión radical $e(r_1, r_2, ..., r_n)$, y para cada permutación σ de $r_1, r_2, ..., r_n$, adjuntemos el elemento $e(\sigma r_1, \sigma r_2, ..., \sigma r_n)$. Ya que sólo hay un número finito de permutaciones σ, el campo que resulta $\overline{E} \supseteq E$ es también una extensión radical de $\mathbb{Q}(r_1, r_2, ..., r_n)$. Así se obtiene una biyección, que también denotaremos por σ, de $\overline{E}$ que envía cada $f(r_1, r_2, ..., r_n) \in \overline{E}$, es decir, cada función racional de $r_1, r_2, ..., r_n$ y los radicales adjuntos, en $f(\sigma r_1, \sigma r_2, ..., \sigma r_n)$ y esta biyección es claramente un automorfismo de $\overline{E}$ que extiende la permutación σ.

La razón por la cual queremos un automorfismo σ que extienda cada permutación de $r_1, r_2, ..., r_n$ es que los coeficientes $a_0, a_1, ..., a_{n-1}$ son fijos bajo cada una de tales permutaciones, de modo que cada elemento del campo

$\mathbb{Q}(a_o, a_1, ..., a_{n-1})$ también es fijado por cada permutación. El siguiente corolario nos dice algo respecto del grupo de Galois de las extensiones radicales.

Corolario 10.2 *Si E es una extensión radical de $\mathbb{Q}(a_o, a_1, ..., a_{n-1})$ que contiene a $r_1, r_2, ..., r_n$, entonces existe una extensión radical adicional $\overline{E} \supseteq E$ tal que $Gal(\overline{E}/\mathbb{Q}(a_o, a_1, ..., a_{n-1}))$ incluye aquellos automorfismos σ que extienden todas las permutaciones de $r_1, r_2, ..., r_n$.*

Demostración. El corolario sigue de inmediato del Teorema 10.9 y del hecho de que una extensión radical de $\mathbb{Q}(a_o, a_1, ..., a_{n-1})$ que contiene a $r_1, r_2, ..., r_n$ es también extensión radical de $\mathbb{Q}(r_1, r_2, ..., r_n)$.

Puntos de discusión

1. Sea $E = \mathbb{Q}(r_1, r_2, \ldots, r_n)$ donde los r_i son todas las ra'ices n-ésimas de la unidad. Demostrar que $E = \mathbb{Q}(r)$ donde r es una raíz n-ésima primitiva de la unidad.

2. Sea E_0 un campo que contiene todas las raíces n-ésimas de la unidad. Sea $t \in E_0$ y sea E un campo de descomposición de $x^n - t$ sobre E_0. Si ω es cualquier cero de $x^n - t$ en E, demostrar que $E = E_0(\omega)$ y que $Gal(E/E_0)$ es cíclico.

3. Sean $t \neq 0$ un elemento de un campo E y D un campo de descomposición de $x^n - t$. Demostrar que D contiene las raíces n-ésimas de la unidad.

4. En referencia al punto 2, sean $E_0 = \mathbb{Q}$, $n = 2$ y $t = 4$. Demostrar que el campo de descomposición E de $x^n - t$ sobre E_0 es $\mathbb{Q}$ y, por ende, que $Gal(E/E_0)$ contiene únicamente el automorfismo idéntico.

5. En referencia al punto 2, sean $E_0 = \mathbb{Q}(r)$ donde r es una raíz cúbica primitiva de la unidad, $t = 7$. Factorizar $x^3 - 7$ en factores lineales en su campo de descomposición E sobre E_0. Demostrar que $Gal(E/E_0)$ es cíclico de orden 3.

6. Demostrar que el grupo de Galois de $x^3 - 7$ sobre $\mathbb{Q}$ es isomorfo a S_3, el grupo de las permutaciones de tres elementos.

7. Sean K, K_2, K_2 y K' como el los puntos de discusión al final de la sección 1.1.6. Sea u_3 el tercer cero del polinomio $x^3 + 6x + 2$. Considerar los grupos $Gal(K_2/K)$,$Gal(K_2/K')$, $Gal(K'/K)$ Demostrar que $Gal(K_2/K)$ es isomorfo a S_3. Determinar los campos que son las imágenes del campo K_1 bajo los automorfismos de $Gal(K'/K)$.

8. Con respecto del punto anterior, determinar

   (a) Los subgrupos de $Gal(K_2/K)$

   (b) Los campos intermedios entre K_2 y K.

   (c) La correspondencia entre campos intermedios y subgrupos.

Con el teorema anterior tenemos las bases necesarias para demostrar el teorema fundamental de la teoría de Galois, pero necesitaremos una serie de lemas con resultados parciales para poder enunciarlo con toda generalidad.

10.1.10 Preparativos para el Teorema Fundamental de la teoría de Galois

Sean L una extensión de un campo K y $G = Gal(L/K)$. Sean $\mathcal{L}$ la colección de subcampos de L que contienen a K y $\mathcal{G}$ la colección de subgrupos de G. Queremos averiguar qué relaciones existen entre $\mathcal{L}$ y $\mathcal{G}$.

Primero, sea $M \in \mathcal{L}$, es decir, un subcampo de L que contiene a K. Definimos

$$M' = \{g \in G \mid g(x) = x, \forall x \in M\}.$$

Como $K \subset M$, es claro que M' es subgrupo de G pues si $g_1, g_2 \in M'$, entonces $\forall x \in M$,

$$g_1(x) = x \wedge g_2(x) = x \Longrightarrow g_2^{-1}g_1(x) = g^{-1}(g_1(x)) = g_2^{-1}(x) = x,$$

de donde, $g_2^{-1}g_1 \in M'$ y M' es subgrupo de G ($M' \in \mathcal{G}$).

Segundo, sea $H \in \mathcal{G}$, un subgrupo de G. Definimos

$$H' = \{x \in L \mid h(x) = x, \forall h \in H\}.$$

Es claro pues que H' es un subcampo de L que contiene a K.

A todo subcampo de L que contiene a K podemos asociar un subgrupo de G y a todo subgrupo de G podemos asociar un subcampo de L que contiene a K. ¿Qué mas podemos afirmar sobre esta asociación?

Lema 10.1 *Si $M_1, M_2 \in \mathcal{L}$ y $M_1 \supseteq M_2$, entonces $M_1' \subseteq M_2'$.*

Demostración. Si $g \in M_1'$, entonces $g(x) = x, \forall x \in M_1$. Pero $M_2 \subseteq M_1$. Entonces $g(x) = x, \forall x \in M_2$. Luego, $g \in M_2'$, es decir, $M_1' \subseteq M_2'$.

Lema 10.2 *Si $H_1, H_2 \in \mathcal{G}$ y $H_1 \supseteq H_2$, entonces $H_1' \subseteq H_2'$.*

Demostración. Si $x \in H_1'$ entonces $h(x) = x, \forall h \in H_1$. Pero $H_2 \subseteq H_1$. En particular, entonces, $h(x) = x, \forall h \in H_2$. Se sigue que $x \in H_2'$ y, por ende, que $H_1' \subseteq H_2'$.

Lema 10.3 *Si $M \in \mathcal{L}$, entonces $M'' \supseteq M$ y si $H \in \mathcal{G}$, entonces $H'' \supseteq H$.*

Demostración. $M'' = \{x \in L \mid g(x) = x, \forall g \in M'\}$. Pero M' fija los elementos de M. Entonces, $M \subseteq M''$. De manera similar se demuestra que $H'' \supseteq H$.

Los resultados anteriores nos permiten hacer la siguiente definición.

Definición 10.3 *Una extensión L se dice* normal *sobre K si $K'' = K$, o equivalentemente, si $G' = K$.*

Esta definición dice que el grupo de automorfismos de L que fijan a K ($Gal(L/K)$) no fijan a 'mas que' K. Es decir, si $u \in L$ y $u \notin K$, existe $\alpha \in G$ tal que $\alpha(u) \neq u$ (hay un automorfismo que no fija a u).

Lema 10.4 $\forall M \in \mathcal{L}$, $M''' = M$ y $\forall H \in \mathcal{G}$, $H''' = H$.

Demostración. Por el Lema 10.3, $M'' \supseteq M$. Entonces $M''' \subseteq M'$. Pero $H'' \supseteq H$, luego $M' \subseteq M'''$. De manera similar se demuestra que $H''' = H$.

Definición 10.4 *M'' se llama la* clausura *de M y se dice que M es cerrado si $M'' = M$. Además, H' se llama la clausura de H y se dice que H es cerrado si $H'' = H$.*

De lo anterior es claro que M', H' son cerrados.

Teorema 10.10 *Sean $\mathcal{G}_1, \mathcal{L}_1$ los subconjuntos de $\mathcal{G}$, $\mathcal{L}$, respectivamente, formados por los elementos cerrados. Entonces existen una aplicación uno-a-uno Φ de $\mathcal{L}_1$ sobre $\mathcal{G}_1$ y una aplicación uno-a-uno Ψ de $\mathcal{G}_1$ sobre $\mathcal{L}_1$ definidas por $\Phi : M \longrightarrow M' \quad y \quad \Psi : H \longrightarrow H'$.*

Demostración. Si $M_1 \in \mathcal{L}_1$ entonces $M_1' = M_1'''$, de modo que $M_1 \in' \mathcal{G}_1$. Además, para $M_{1,2} \in \mathcal{L}_1$, si $M_1' = M_2'$, entonces $M_1'' \subseteq M_2''$ y $M_2'' \subseteq M_1''$, por el Lema 10.2 Por lo tanto, $M_1'' = M_2''$. Pero M_1 y M_2 son cerrados. Se sigue que $M_1 = M_2$. Luego, Φ es uno-a-uno.

Por otra parte, Φ es sobre, pues si $H \in \mathcal{G}_1, H = (H')'$. Entonces existe $H' \in \mathcal{L}_1$ tal que $\Phi(H') = H$.

Para demostrar la segunda parte del teorema basta poner $\Psi = \Phi^{-1}$.

Teorema 10.11 *Supongamos que $M_1, M_2 \in L$ y $M_1 \supseteq M_2$ con $[M_1 : M_2] = n$, finito. Entonces*

$$i_{M_2'}(M_1') \leq n.$$

Nota. *Otra forma de escribir la conclusión es: $[M_2' : M_1'] \leq n$.*

Demostración.
Supongamos que es falso, o sea, que $[M_2' : M_1'] > n$. Entonces, sean $\phi_0 = e, \phi_1, \ldots, \phi_n$ elementos de distintas clases laterales derechas de M_1' en M_2'. Sean $\psi_0, \psi_1, \ldots, \psi_n$ las restricciones de los ϕ_i a M_1. Entonces $\psi_0, \psi_1, \ldots, \psi_n$ son distintos homomorfismos uno-a-uno de M_1 en L. Pue si para $i \neq j, \Psi_i = \Psi_j$, entonces

$$\begin{aligned} \Phi_i(x) &= \Phi_j(x), \forall x \in M_1 \\ \Phi_j^{-1}(\Phi_i(x)) &= x, \forall x \in M_1, \end{aligned}$$

de donde, $\Phi_j^{-1}\Phi_i \in M_1'$, lo cual contradice que $\phi_0, \phi_1, \ldots, \phi_n$ se encuentran en clases laterales distintas de M_1' en M_2'. De allí se tiene que los $\psi_0, \psi_1, \ldots, \psi_n$ son distintos.

Ahora bien, si $N = \{x \in M_1 \mid x = \psi_0(x) = \psi_1(x) = \cdots = \psi_n(x)\}$, entonces $[M_1 : N] \geq n + 1$, por el Teorema *.5. Pero $M_2 \subseteq N$ puesto que todo elementos de M_2' fija a M_2 y los Ψ_i son restricciones de elementos de M_2' a M_1. además, $[M_1 : M_2] = n$. Hemos llegado a una contradicción ya que

$$[M_1 : M_2] = [M_1 : N][N : M_2], \text{ con } [M_1 : M_2] = n \text{ y } [M_1 : N] \geq n + 1,$$

quedando demostrado el teorema

En la presente serie de teoremas hemos intentado mostrar, hasta donde sea posible, una dualidad entre $\mathcal{L}$ y $\mathcal{G}$. Por lo tanto, resulta natural enunciar el teorema siguiente cuya demostración será omitida.

Teorema 10.12 *Sean $H_1, H_2 \in \mathcal{G}$, con $H_1 \supseteq H_2$ y $[H_1 : H_2] = n$, finito. Entonces $[H_2' : H_1'] \leq n$.*

El próximo paso en nuestro desarrollo es demostrar que, en el caso de los elementos cerrados, se tiene la igualdad en los dos teoremas anteriores.

Teorema 10.13 *Si $M_2 \in \mathcal{L}$ y $M_2 = M_2''$ (M_2 es cerrado) y si $[M_1 : M_2] = n$, entonces $M_1 = M_1''$ y $[M_2' : M_1'] = n$.*

Demostración. Por el Teorema 10.11, $[M_2' : M_1'] = n_0 \leq n$. Por el teorema 10.12, $[M_1'' : M_2''] \leq n_0$. Pero $M_2'' = M_2$. Luego, $[M_1'' : M_2] \leq n_0$. Además, hemos mostrado en el caso general que $M_1'' \supseteq M_1$. Entonces

$$n = [M_1 : M_2] \leq [M_1'' : M_2] \leq n_0 \leq n.$$

De allí es claro que $n = n_0$ y, por lo tanto, que $M_1'' = M_1$.

Teorema 10.14 *Si $H_2 \in \mathcal{G}$, con $[H_1 : H_2] = n$ y $H_2'' = H_2$, entonces $H_1'' = H_1$ y $[H_2' : H_1'] = n$.*

Demostración. Es similar a la demostración anterior.

Ejercicio

Demostrar el Teorema 10.14.

Corolario 10.3 *Cualquier subgrupo finito de G es cerrado.*

Demostración. Si i es el automorfismo idéntico de G, el subgrupo (i) es cerrado dado que $(i)' = L$ y $(i)'' = L' = (i)$. Entonces, si H es subgrupo de G y H es finito, $[H : (i)] = \mid H \mid$ es finito. Por el Teorema 10.14, $H'' = H$, o sea, H es cerrado.

Corolario 10.4 *Si L es extensión normal de K, entonces un campo M tal que $[M : K]$ es finito es necesariamente cerrado.*

Demostración. Si L es extensión normal, tenemos por definición que $K'' = K$. Además, $[M : K]$ es finito. Por el Teorema 10.13, $M'' = M$, es decir, M es cerrado.

En el transcurso de las demostraciones anteriores hemos mostrado que hay una correspondencia 1-1 entre los subcampos cerrados de L que contienen a K y los subgrupos cerrados de $G = Gal(L/K)$. Por medio de los dos corolarios anteriores, vemos que, al imponer las condiciones de que la extensión L sea normal y finita, podemos concluir que todos los subcampos y subgrupos en cuestión son cerrados. Esta conclusión sale claramente del corolario siguiente.

Corolario 10.5 *Sean L una extensión normal y finita de $K, G = Gal(L/K)$. Si $[L : K] = n$, entonces $\mid G \mid = n$.*

Demostración. Si i es el automorfismo idéntico de G, es inmediato que $(i)' = L$. Además, L es extensión normal. Luego, $K'' = K$ y $G' = K$.

$$n = [L : K] = [K' : L'] = [G : (i)] = \mid G \mid .$$

Estamos en condiciones de enunciar el **Teorema Fundamental de la Teoría de Galois.**

Teorema 10.15 *Sean L una extensión normal y finita de un campo K y $G = Gal(L, K)$. Entonces existe una correspondencia 1-1 entre el conjunto de los subgrupos de G y el conjunto de los campos M tales que $K \subseteq M \subseteq L$. Donde, para $H \in \mathcal{G}$, el elemento correspondiente de $\mathcal{L}$ es $\{x \in$*

$L \mid h(x) = x, \forall h \in H\}$ *y para* $M \in \mathcal{L}$ *el elemento correspondiente de* $\mathcal{G}$ *es* $\{g \in G \mid g(x) = x, \forall x \in M\}$. *Si* H *corresponde a* M, *se tiene que* $[M : K] = [G : H]$. *En particular,* $[L : K] = \mid G \mid$.

Nota. Es posible, además, mostrar que $M \in \mathcal{L}$ es extensión normal de K si y solamente si el elemento correspondiente de $\mathcal{G}$, M', es subgrupo normal de G.

El Teorema Fundamental y los campos de descomposición

Nuestro próximo paso será el de interrelacionar las extensiones normales y finitas de los cuales habla el Teorema Fundamental con los campos de descomposición. Como esta es la parte mejor conocida de la teoría de campos, haremos un resumen de los teoremas principales dejando las demostraciones de la mayoría de ellos al lector.

Teorema 10.16 *Sean* L *una extensión finita de un campo* K, φ *un automorfismo de* L *tal que* $\varphi(k) = k, \forall k \in K$. *Entonces, para cualquier* $u \in L, \varphi(u)$ *es un conjugado de* u.

Demostración. Como $u \in L$, una extensión finita de K, u es algebraico sobre K. Sea $f(x) \in K[x]$ el polinomio irreducible de u sobre K. Entonces, si $f(x) = a_0 + a_1 x + \cdots + a_{n-1}x^{n-1} + x^n$,

$$f(u) = a_0 + a_1 u + \cdots + a_{n-1}u^{n-1} + u^n = 0$$

y

$$\varphi(f(u)) = \varphi\left(a_0 + a_1 u + \cdots + a_{n-1}u^{n-1} + u^n\right) = \varphi(0) = 0.$$

Aplicando propiedades de los automorfismos tenemos

$$\varphi(a_0) + \varphi(a_1)\varphi(u) + \cdots + \varphi(a_{n-1})(\varphi(u))^{n-1} + (\varphi(u))^n = 0.$$

Pero φ fija a los a_i ya que éstos pertenecen a K. Por lo tanto,

$$a_0 + a_1\varphi(u) + \cdots + a_{n-1}(\varphi(u))^{n-1} + (\varphi(u))^n = 0$$

que muestra que $\varphi(u)$ es también raíz de $f(x)$ y por lo tanto conjugado de u.

El teorema anterior nos dice que si M es un campo de extensión de K que contiene algunas raíces de un polinomio $f(x) \in K[x]$, irreducible sobre K, entonces los elementos del grupo Galois de M sobre K permutan dichas raíces. Estamos en condiciones de demostrar el siguiente teorema que nos proporciona un eslabon que une las extensiones normales con los campos de descomposición.

Teorema 10.17 *Sean* M *una extensión normal de* K *y* $f(x)$ *un polinomio irreducible sobre* K *tal que* f *tenga una raíz en* M. *Entonces* f *factoriza sobre* M *en factores lineales distintos.*

Demostración. Sean $u = u_1, u_2, \ldots, u_r$ las distintas imágenes de u bajo automorfismos de M sobre K. Cada u_i es raíz de f y entonces $r \leq n$ donde $deg(f(x)) = n$.

Sea $g(x) = (x - u_1)(x - u_2)\cdots(x - u_r)$. Los coeficientes de g pertenecen a M, pero cualquier automorfismo de M sobre K permuta los u_i. Entonces,

los coeficientes de g, siendo funciones simétricas de los u_i son invariantes bajo todo automorfismo de M/K y, como M es normal sobre K, dichos coeficientes deben estar en K. Es decir, $g(x) \in K[x]$. Ahora bien, f es el polinomio irreducible de u sobre K y g es otro polinomio sobre K con raíz u. Por lo tanto, $f \mid g$, de donde, $n \leq r$. Pero ya teníamos $r \leq n$. Luego, $g = f$. Por consiguiente, f factoriza sobre M como producto de factores lineales distintos.

Puntos de discusión El polinomio $f(x) = x^3 + 3x^2 + 4$ tiene un cero real $u_1 = a+b-1$ donde $a = (-3+\sqrt{8})^{\frac{1}{3}}$ y $b = (-3-\sqrt{8})^{\frac{1}{3}}$, así como dos ceros complejos $u-2, u_3$.

1. Demostrar que el polinomio es irreducible en $\mathbb{Q}[x]$. Es decir, hay un polinomio irreducible con un cero en $E = \mathbb{Q}(u_1)$ que no factoroiza en factores lineales en E. Demostrar que E no es una extensión normal de Q.

2. Demostrar que $D = \mathbb{Q}(u_1, u_2) = \mathbb{Q}(u_1, u_2, u_3)$ sí es campo de descomposición de $f(x)$ sobre $\mathbb{Q}$. Demostrar que D es extensión normal de $\mathbb{Q}$.

Teorema 10.18 *Sean G el grupo Galois de M sobre K, L un campo intermedio ($K \subseteq L \subseteq M$) tal que L es extensión normal y finita de K. Entonces $L' \lhd G$ y G/L' es isomorfo al $Gal(L/K)$.*

Demostración. Sean $u \in L$ y $\varphi \in G = Gal(M/K)$. Queremos mostrar que L es *estable* respecto de G, es decir, que $\varphi(u) \in L$. Ahora, como L es extensión finita, u es algebraico sobre K. Sea f su polinomio mínimo irreducible sobre K. Entonces, por el teorema *.16, f factoriza en factores lineales distintos en L. Como $\varphi(u)$ es raíz de f, tenemos que $\varphi(u) \in L$. En resumen, $\forall u \in L, \forall \varphi \in G, \varphi(u) \in L$.

Veamos ahora que $L' \lhd G$, es decir, que $\forall \alpha \in G, \forall \varphi \in L', \alpha^{-1}\varphi\alpha \in L'$. Para ello, sea x un elemento cualquiera de L. Entonces

$$\alpha^{-1}\varphi\alpha(x) = \alpha^{-1}\varphi(\alpha(x)).$$

Pero L es estable, luego $\alpha(x) \in L$ y, por otra parte, $\varphi \in L'$, de donde, φ fija todo elemento de L. Es decir, $\varphi(\alpha(x)) = \alpha(x)$. Entonces,

$$\alpha^{-1}\varphi\alpha(x) = \alpha^{-1}\varphi(\alpha(x)) = \alpha^{-1}(\alpha(x)) = x.$$

Lo anterior significa que $\alpha^{-1}\varphi\alpha$ fija todo elemento de L, es decir, $\alpha^{-1}\varphi\alpha \in L'$, y de allí se tiene que $L' \lhd G$.

Ahora bien, si $\alpha \in Gal(M/K)$, como L es estable es claro que $\alpha \mid_L$ es un automorfismo de L sobre K. Dicha restricción produce un homomorfismo ρ de G en $Gal(L/K)$, a saber

$$\rho : \alpha \longrightarrow \alpha \mid_L .$$

Aplicando teoremas de la teoría de grupos, obtenemos

$$G/\ker\rho \approx im\rho.$$

Ahora bien, $\ker\rho = \{\alpha \in G \mid \alpha \mid_L$ es el automorfismo idéntico$\}$. Claramente, $\ker\rho = L'$. Además, $im\rho \approx Gal(L/K)$, o sea, $G/L \approx' Gal(L/K)$.

Ya hemos demostrado que una extensión normal y finita M de K es campo de descomposición de un polinomio cuyas raíces son distintas. (Esto sigue directamente del Teorema 10.17 puesto que si $[M : K]$ es finito y $u \in M, u \notin K$, u satisface un polinomio irreducible f sobre K. Luego, f tiene una raíz en M y por el Teorema 10.17 se tiene el resultado.) Ahora queremos establecer, hasta donde sea posible, la inversa de esta proposición. Primero necesitamos la siguiente definición.

Definición 10.5 *Sea $f(x) \in K[x]$ un polinomio irreducible sobre K. Decimos que f es* separable *sobre K si f factoriza en factores lineales distintos en algún campo de descomposición.*

Se puede mostrar que, equivalentemente, f es separable si f' (la derivada de f) $\neq 0$. Luego, podemos enunciar el siguiente teorema cuya demostración omitiremos

Teorema 10.19 *Sea M una extensión finita de K. M es normal sobre K si y solamente si M es campo de descomposición sobre K de un polinomio separable (cuyos factores irreducibles son separables).*

El Teorema *.18 nos proporciona precisamente los campos que satisfacen los prerrequisitos del Teorema Fundamental, éstos son campos de descomposición de polinomios irreducibles separables. Entre otras cosas, esto nos muestra que, aunque la investigación original de Galois se limitó a los campos de extensión que se obtienen en relación con la solución de ecuaciones polinómicas, esto *no resta generalidad* a los resultados de Galois.

10.1.11 La estructura de las extensiones radicales

Lo que hemos determinado hasta este momento es que una solución por radicales de la ecuación general de grado n involucra una extensión radical de $\mathbb{Q}(r_1, r_2, ..., r_n)$ que contiene a $a_o, a_1, ..., a_{n-1}$ y, por ende, una extensión radical $\overline{E}$ con la simetría que describimos en el Corolario 10.2. Este enfoque abre el camino para demostrar la no existencia de una tal solución si se averigua lo suficiente acerca de $Gal(\overline{E}/\mathbb{Q}(a_o, a_1, ..., a_{n-1}))$ para demostrar, por ejemplo, que tal simetría no se da, al menos para el caso $n \geq 5$. En esta sección demostraremos que el grupo Galois $Gal(K(u_1, u_2, ..., u_n)/K)$ de cualquier extensión radical posee una estructura especial, conocida con el nombre de *solubilidad* que se hereda de la estructura de $K(u_1, u_2, ..., u_n)$. Luego, en la siguiente sección demostraremos que esta estructura de hecho es incompatible con la simetría descrita en este corolario al Teorema 10.9. Para simplificar el proceso de descripción de la estructura se harán ciertas suposiciones acerca de la adjunción de radicales u_i que como demostraremos pueden hacerse sin pérdida de generalidad.

En primera instancia puede suponerse que cada radical u_i que se adjunta es una raíz p-ésima para algún primo p. Es decir, en lugar de adjuntar directamente algo como $\sqrt[6]{u}$, podemos adjuntar primero $v = \sqrt{u}$ y luego $\sqrt[3]{v}$. En segundo lugar, si u_i es una p-ésima raíz, podemos suponer que $K(u_1, u_2, ..., u_i)$ no contiene ninguna raíz p-ésima de la unidad que no esté en $K(u_1, u_2, ..., u_{i-1})$ a no ser que el mismo u_i es una raíz p-ésima de la unidad. Si éste no es el caso, simplemente se adjunta una raíz p-ésima de la unidad $\zeta \neq 1$ a $K(u_1, u_2, ..., u_{i-1})$ para obtener $K(u_1, u_2, ..., u_{i-1}, \zeta)$ antes de adjuntar u_i. De este modo $K(u_1, u_2, ..., u_{i-1}, \zeta)$ contiene todas las raíces

p-ésimas de la unidad, $1, \zeta, \zeta^2, ..., \zeta^{p-1}$. Con estas dos modificaciones, el campo extensión final $K(u_1, u_2, ..., u_k)$ es el mismo y sigue siendo el mismo aunque las raíces adjuntadas ζ se incluyan en la lista $u_1, u_2, ..., u_k$.

Se sigue que cualquier extensión radical $K(u_1, u_2, ..., u_r)$ es la union de una torre ascendente de campos

$$K = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_t = K(u_1, u_2, ..., u_t),$$

donde cada $K_i = K_{i-1}(u_i)$, con u_i una raíz p-ésima de un elemento en K_{i-1} a no ser que u_i es el mismo una raíz p-ésima de la unidad.

A esta torre de campos le corresponde una torre descendiente de grupos

$$Gal(K_t/K_0) = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_t = Gal(K_t/K_t) = \{\mathbf{1}\},$$

donde $G_i = Gal(K_t/K_i) = Gal(K_t/K_{i-1}(u_i))$ y $\mathbf{1}$ denota el automorfismo idéntico. Las contenencias siguen de inmediato de la definición de $Gal(L/K)$, para campos cualesquiera $L \supseteq K$, como el grupo de automorfismos que fijan cada elemento de K. A la medida en que K se acerque a L (en el sentido de contener mas y mas elementos de L), $Gal(L/K)$ debe decrecer a $\{\mathbf{1}\}$. El punto importante aquí es que el paso de G_{i-1} a su subgrupo G_i es lo suficientemente pequeño como para describirse en términos pertenecientes a la teoría de grupos, a saber, G_i es un subgrupo normal de G_{i-1} y G_{i-1}/G_i es abeliano, como demostraremos a continuación. Sin embargo, primero queremos simplificar la notación un poco mas. Sean $L = K_t, M = K_{i-1}, u = u_i, p = p_i$.

Teorema 10.20 *Sean $L \supseteq M(u) \supseteq M$ campos con $u^p \in M$ para algún primo p, y tales que si $M(u)$ no contiene ninguna raíz p-ésima de la unidad que no está en M, a no ser que u mismo es una raíz p-ésima de la unidad, entonces $Gal(L/M(u))$ es un subgrupo normal de $Gal(L/M)$ y $Gal(L/M)/Gal(L/M(u))$ es abeliano.*

Demostración. En virtud del teorema de homomorfismos de grupos, es suficiente hallar un homorfismo de $Gal(L/M)$ cuyo núcleo es $Gal(L/M(u))$ y cuya imagen es un grupo abeliano. La aplicación obvia con núcleo $Gal(L/M(u))$ es la restricción a $M(u)$ ya que, por definición,

$$\sigma \in Gal(L/M(u)) \Leftrightarrow \sigma \mid_{M(u)}$$

es la aplicación idéntica.

La propiedad de los homomorfismos

$$\sigma'\sigma \mid_{M(u)} = \sigma' \mid_{M(u)} \sigma$$

para todo $\sigma', \sigma \in Gal(L/M)$,es automática siempre que $\sigma \mid_{M(u)} (b) \in B(\alpha)$ para cada $x \in M(u)$, es decir, siempre que $M(u)$ sea cerrado bajo cada $\sigma \in Gal(L/M)$.

Ya que σ fija M, se sigue que $\sigma \mid_{M(u)}$está completamente determinada por el valor de $\sigma(u)$. Si u es una raíz p-ésima de la unidad ζ, entonces

$$(\sigma(u))^p = \sigma(u^p) = \sigma(\zeta^p) = \sigma(1) = 1,$$

y de allí se tiene que $\sigma(u) = \zeta^i = u^i \in M(u)$, ya que cada raíz p-ésima de la unidad es alguno de los ζ^i. Si u no es una raíz de la unidad, entonces

$$(\sigma(u))^p = \sigma(u^p) = u^p$$

porque $u^p \in M$, y de allí $\sigma(u) = \zeta^i u$ para alguna raíz p-ésima de la unidad ζ, y $\zeta \in M$ por hipótesis. Nuevamente, $\sigma(u) \in M(u)$. Se sigue que $M(u)$ es cerrado, como queríamos.

Lo anterior también implica que $|_{M(u)}$envía a $Gal(L/M)$ en $Gal(M(u)/M)$ de tal modo que queda por comprobar que $Gal(M(u)/M)$ es abeliano. Si u es una raíz de la unidad, se sigue que cada $\sigma\ |_{M(u)} \in Gal(M(u))/M)$ es de la forma σ_i donde $\sigma_i(u) = u^i$. Por lo tanto,

$$\sigma_i\sigma_j(u) = \sigma_i(u^j) = u^{ji} = u^{ij} = \sigma_j\sigma_i(u).$$

De manera similar, si u no es una raíz de la unidad, entonces cada $\sigma\ |_{M(u)} \in Gal(M(u))/M)$ es de la forma σ_i donde $\sigma_i(u) = \zeta^i u$, de donde se sigue que

$$\sigma_i\sigma_j(u) = \sigma_i(\zeta^j u) = \zeta^{j+1}u = \zeta^{i+j}u = \sigma_j\sigma_i(u),$$

ya que $\zeta \in M$ y, por tanto, ζ permanece fijo. En cualquiera de los dos casos, $Gal(M(u)/M)$ es abeliano.

Recordemos que la propiedad de $Gal(K(u_1, u_2, ..., u_t)\ /K)$ que sigue del teorema anterior es que contiene subgrupos

$$Gal(K(u_1, u_2, ..., u_t)/K) = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_t = \{1\}$$

con cada G_i subgrupo normal de G_{i-1} y G_{i-1}/G_i abeliano se denomina la *solubilidad* de $Gal(K(u_1, u_2, ..., u_t)\ /K)$. Se sigue de lo demostrado que si $K(u_1, u_2, ..., u_t)$ es una extensión radical, entonces $Gal(K(u_1, u_2, ..., u_t)\ /K)$ es un grupo *soluble*. Esta caracterización permite analizar la solubilidad por radicales de las ecuaciones polinómicas, como se hará en la siguiente sección.

10.2 Aplicaciones

10.2.1 La no existencia de solución por radicales para la ecuación general de grado mayor o igual a 5

De acuerdo con lo que hemos dicho, la no existencia de una solución por radicales a una ecuación polinómica de grado mayor o igual a 5, es equivalente a mostrar que una extensión radical de $\mathbb{Q}(a_o, a_1, ..., a_{n-1})$ no contiene a $r_1, r_2, ..., r_n$ o, equivalentemente, a $\mathbb{Q}(r_1, r_2, ..., r_n)$. Ya hemos reducido el problema a el de demostrar que la simetría de la extensión hipotética $\overline{E}$ que contiene a $r_1, r_2, ..., r_n$, enunciada en el corolario del Teorema 10.9, es incompatible con la solubilidad de $Gal(\overline{E}/\mathbb{Q}(a_o, a_1, ..., a_{n-1}))$, enunciada en el Teorema 10.18.

La demostración examina el efecto sobre $r_1, r_2, ..., r_n$ de los automorfismos hipotéticos de $\overline{E}$ y, por ende, realmente concierne el *grupo simétrico* S_n de permutaciones de $r_1, r_2, ..., r_n$. De hecho, en lo que sigue se adapta una demostración estandar de que $S_n, n \geq 5$ no es un grupo soluble.

Teorema 10.21 *Una extensión radical de* $\mathbb{Q}(a_o, a_1, ..., a_{n-1})$ *no contiene a* $\mathbb{Q}(r_1, r_2, ..., r_n)$ *para* $n \geq 5$.

Demostración. Procederemos por contradicción. Supongamos que E es una extensión radical de $\mathbb{Q}(a_o, a_1, ..., a_{n-1})$ que contiene a $\mathbb{Q}(r_1, r_2, ..., r_n)$.

Se sigue que E es también una extensión radical de $\mathbb{Q}(r_1, r_2, ..., r_n)$ y, por el Corolario 1.2, existe una extensión radical $\overline{E} \supseteq E$ tal que

$$G_0 = Gal(\overline{E}/\mathbb{Q}(a_o, a_1, ..., a_{n-1}))$$

incluye automorfismos σ que extiende todas las permutaciones de $r_1, r_2, ..., r_n$.

Por el Teorema 10.18, G_0 tiene una descomposición

$$G_o \supseteq G_1 \supseteq \cdots \supseteq G_k = \{1\},$$

donde cada G_i es un subgrupo normal de G_{i-1} y G_{i-1}/G_i es abeliano. Esto contradice la existencia de los automorfismos σ, ya que, como G_{i-1}/G_i es abeliano, G_i es el núcleo del homorfismo de G_{i-1} sobre un grupo abeliano y, por lo tanto,

$$\sigma, \tau \in G_{i-1} \Longrightarrow \sigma^{-1}\tau^{-1}\sigma\tau \in G_i.$$

Este último hecho puede usarse para demostrar, por inducción sobre i, que, si $n \geq 5$, entonces cada G_i contiene automorfismos σ que extienden todos los 3-ciclos (r_a, r_b, r_c). Esto es cierto para G_0 por hipótesis. Y, cuando $n \geq 5$, la propiedad se transmite de G_{i-1} a G_i porque

$$(r_a, r_b, r_c) = (r_d, r_a, r_c)^{-1}(r_c, r_e, r_b)^{-1}(r_d, r_a, r_c)(r_c, r_e, r_b)$$

donde a, b, c, d y e son distintos. Luego si hay al menos cinco raíces, existe σ en cada G_i que extienden 3-ciclos arbitrarios (r_a, r_b, r_c), lo cual significa, en particular, que $G_k \neq \{1\}$. Esta contradicción demuestra que $\mathbb{Q}(r_1, r_2, ..., r_n)$ no está contenido en ninguna extensión radical de $\mathbb{Q}(a_o, a_1, ..., a_{n-1})$ para $n \geq 5$.

Punto de discusión

Respecto de los Puntos de discusión al final de la Sección 1.1.9, construir una serie de descomposición de subgrupos nornales (condiciones para un grupo soluble) para $Gal(K_2/K)$ y determinar la correspondiente sucesión de campos intermedios.

10.2.2 El problema de las construcciones: los problemas clásicos

Probablemente en los inicios del siglo V a.C. los matemáticos griegos se dedicaron a investigar algunos problemas de construcción que pasaron a la historia de las matemáticas. Tres de ellos adquirieron particular fama.

(i) La cuadratura del círculo. Dado un círculo el problema pide construir un cuadrado cuya área sea igual al área del círculo.

(ii) Trisección de un ángulo. Este problema pide un método para dividir un ángulo arbitrario en tres partes iguales.

(iii) La duplicación del cubo. Se pide construir un cubo que tenga el doble del volumen que un cubo dado.

Los geómetras griegos enfocaron su interés en variados problemas de este tipo, especialmente en la construcción de polígonos regulares. A pesar del esfuerzo de estos habilidosos geómetras muchos de los problemas quedaron sin resolver. En el siglo XIX en la última sección de sus *Disquisitiones arithmeticae*, Gauss retomó el problema de la construcción de los polígonos regulares, precisamente para introducir tópicos geométricos en su trabajo sobre

teoría de números. Después de presentar reglas para las construcciones por medio de regla y compás, pasó al análisis de los problemas de construcción estableciendo una relación entre estos problemas y la teoría de ecuaciones algebraicas.

Para determinar qué cantidades pueden ser construidas, el asumió que uno puede representar las operaciones geométricas en un sistema coordenado en el plano y examinar las operaciones algebraicas involucradas en cada paso. Cuando dos puntos están dados por medio de sus coordenadas, los coeficientes de la ecuación de la línea recta que pasa a través de ellos pueden ser calculados racionalmente (cocientes, operaciones con racionales), recíprocamente cuando los coeficientes de dos pares de líneas son conocidos, las coordenadas de su punto intersección pueden ser determinadas por operaciones racionales de éstos. El cálculo de los puntos de intersección de un círculo y una línea recta o de un círculo con otro círculo produce una ecuación de segundo grado. Las coordenadas pueden ser obtenidas entonces como sumas de expresiones racionales en los coeficientes conocidos y raíces cuadradas de tales expresiones. La distancia entre los dos puntos es también expresable como una raíz cuadrada.

Como toda otra construcción puede ser obtenida por composición de estas operaciones simples, se concluye que tales magnitudes pueden ser construidas si se establece una manera de calcular algebraicamente por aplicación repetida de las cuatro operaciones aritméticas y por extracción de raíces cuadradas. En resumen, el construir con regla y compás las cantidades geométricas corresponde algebraicamente a que éstas pueden ser deducidas por uso repetido de las cuatro operaciones racionales y extracción de raíz cuadrada.

El problema era ahora solucionar una ecuación por radicales; dicho problema no pudo ser atacado antes de los trabajos de los dos eminentes matemáticos, N.H Abel (1802-1829) y E. Galois (1811- 1832). Como comentamos anteriormente, ellos observaron que a pesar de que cualquier expresión construible satisface una ecuación con coeficientes racionales, para muchas de estas ecuaciones existe un polinomio de grado mínimal que no puede ser factorizado con coeficientes racionales. De la Teoría de Galois se sigue que para que esta ecuación sea soluble por radicales debe tener propiedades especiales; una de ellas es que tenga como grado una potencia de 2. (El problema de la solubilidad caracterizado en términos del grado de la ecuación.)

Regresando con estos argumentos a los problemas clásicos, observamos que si el cubo dado tiene lado a y el cubo del doble del volumen lado b tendríamos, $b^3 = 2a^3$, o, $b = \sqrt[3]{2}a$. El problema es entonces construir $x = \sqrt[3]{2}$, la cual es raíz de la ecuación

$$x^3 - 2 = 0.$$

Esta ecuación no puede ser factorizada en los racionales y, dado que el grado no es potencia de 2, se concluye que el cubo no puede ser duplicado por medio de construcciones con regla y compás.

Acerca de la trisección del ángulo, haremos las siguientes consideraciones. En primer lugar un ángulo α puede ser construido si y sólo si $\cos\alpha$ ($\sin\alpha$) es construible con regla y compás. El problema se traduce entonces en discutir la posibilidad de construir $x = \cos\frac{\alpha}{3}$ cuando $a = \cos\alpha$ es conocido.

Usando una fórmula de la trigonometría elemental, a saber

$$\cos 3\theta = 4\cos\theta - 3\cos\theta,$$

tenemos

$$\cos\alpha = 4(\cos\frac{\alpha}{3})^3 - 3\cos\frac{\alpha}{3},$$

y esto puede ser reescrito como la ecuación cúbica

$$4x^3 - 3x - a = 0.$$

Esta ecuación no puede ser factorizada en los racionales y, dado que es cúbica, se concluye entonces que no se puede trisectar con regla y compás un ángulo arbitrario. De nuevo un problema geométrico es resuelto argumentando sobre la solubilidad de una ecuación algebraica.

El problema de la cuadratura del círculo, equivalente a encontrar una construcción para π con regla y compás, es de corte distinto, pues lo que ocurre aquí es que no existe una ecuación algebraica asociada con este problema. π no es construible con regla y compás, aún mas, π es transcendente, hecho cuya demostración fue dada por F. Lindemann en 1873. Quedaban así resueltos los famosos problemas clásicos.

Punto de investigación

¿Es posible dividir con regla y compás un ángulo de 60° en cinco partes iguales?

10.2.3 Construcción de polígonos regulares

Volvamos ahora al estudio que adelantó Gauss de la ecuación ciclotómica. En esta sección vamos a considerar la ecuación minimal para las raíces n-ésimas primitivas de la unidad y a aplicar el criterio que limita el grado de una ecuación soluble por radicales, análisis que produce un resultado interesante.

Para que la ecuación de las raíces n-ésimas de la unidad sea soluble por medio de raíces cuadradas, es necesario que $\phi(n)$ (la función ϕ de Euler) sea una potencia de 2.

Este hecho origina una poderosa restricción sobre el número n. Para analizar sus implicaciones, sea

$$n = 2^{\alpha_0} p_1^{\alpha_1} \cdots p_r^{\alpha_r}$$

la factorización prima de n. En ese caso $\phi(n)$ es

$$\phi(n) = 2^{\alpha_0 - 1} p_1^{\alpha_1 - 1}(p_1 - 1) \cdots p_r^{\alpha_r - 1}(p_r - 1)$$

y será una potencia de 2 sólo cuando cada uno de sus factores es potencia de 2. Se concluye entonces que todos los exponentes $\alpha_1, \alpha_2, \cdots \alpha_r$ deben se iguales a 1. Y en segundo lugar, se tiene que los números $p_i - 1$ deben ser potencias de 2. Por tanto, un primo impar que divide a n es de la forma

$$p_i = 2^{k_i} + 1.$$

Pero éstos serían precisamente los primos de Fermat. En la caracterización de los primos de Fermat se observa que números de esta forma son primos

solamente cuando el exponente k_i es a su vez una potencia de 2. Quedan pues definidos por la expresión

$$\begin{aligned} F_t &= 2^{2^t}+1. \\ F_0 &= 3, F_1 = 5, F_2 = 17, F_3 = 257, \ldots \end{aligned}$$

Basados en estas observaciones, podemos enunciar el resultado fundamental presentado por Gauss en las *Disquisitiones.*

Un polígono regular con n lados puede ser construido con regla y compás solamente cuando el número n es de la forma $n = 2^\alpha p_1 p_2 \cdots p_r$ donde los factores primos son primos de Fermat.

La discusión anterior estuvo dirigida a demostrar que esta condición es necesaria. Gauss demostró recíprocamente que también es suficiente, demostrando que un polígono con p lados puede ser construido cuando p es un primo de Fermat. En ese caso encontró que la ecuación para las p-ésimas raíces primitivas de la unidad puede ser resuelta por una serie de ecuaciones de segundo grado. Para terminar, analicemos un par de ejemplos que ilustran la discusión presentada y nos permiten aclarar los nexos que nos interesaba destacar:

Ejemplos

1. Cuando p es primo el número de p-ésimas raíces primitivas de la unidad es $\phi(p) = p-1$ y claramente la única raíz no primitiva es $x = 1$. Dado que todas satisfacen la ecuación

$$x^p - 1 = 0,$$

las raíces primitivas son las raíces de

$$\frac{x^p-1}{x-1} = x^{p-1} + x^{p-2} + \cdots + x^2 + x + 1 = 0.$$

Las raíces, como se desarrolló en la Sección 1.1.2, son todas las potencias de

$$\xi = \cos\frac{2\pi}{p} + i\sin\frac{2\pi}{p}$$

Nótese además que las dos raíces

$$\xi^k = \cos\frac{2\pi}{p}k + i\sin\frac{2\pi}{p}k$$

$$\xi^{p-k} = \xi^{-k} = \cos\frac{2\pi}{p}k - i\sin\frac{2\pi}{p}k$$

son conjugadas, y su suma es

$$\xi^k + \xi^{-k} = 2\cos\frac{2\pi}{p}k,$$

que es un número real. En particular

$$\eta = \xi + \xi^{-1} = 2\cos\frac{2\pi}{p} > 0, \tag{10.10}$$

y este número puede servir para construir el polígono.

Los dos períodos son reales. Veamos para ρ

$$\rho = 2\cos\frac{2\pi}{17} + 2\cos\frac{4\pi}{17} + 2\cos\frac{8\pi}{17} + 2\cos\frac{16\pi}{17},$$

y se puede comprobar además que ρ es positivo y ρ_1 es negativo.

Estos períodos son realmente raíces de la ecuación de segundo grado

$$r(x) = (x-\rho)(x-\rho_1) = x^2 - (\rho+\rho_1)x + \rho\rho_1.$$

Esta ecuación tiene coeficientes racionales puesto que, claramente, $\rho+\rho_1 = -1$. Para el producto, un cálculo y análisis de las potencias permite concluir que $\rho\rho_1 = -4$.

La ecuación entonces tiene la forma

$$x^2 + x - 4$$

y sus raíces son

$$\rho = \frac{\sqrt{17}-1}{2}, \qquad \rho_1 = \frac{-\sqrt{17}-1}{2}.$$

En un segundo paso Gauss estudia los que llama segundos períodos

$$\sigma = \xi + \xi^{-1} + \xi^4 + \xi^{-4}$$

$$\sigma_1 = \xi^2 + \xi^{-2} + \xi^8 + \xi^{-8}$$

$$\sigma_2 = \xi^3 + \xi^{-3} + \xi^5 + \xi^{-5}$$

$$\sigma_3 = \xi^6 + \xi^{-6} + \xi^7 + \xi^{-7}.$$

De nuevo todos estos períodos son reales, por ejemplo,

$$\sigma = \cos\frac{2\pi}{17} + \cos\frac{8\pi}{17} > 0.$$

Similarmente, se muestra que σ_2 es positivo mientras que σ_1 y σ_3 son negativos.

Los cuatro períodos en pares, σ, σ_1 y σ_2, σ_3 son raíces de una ecuación de segundo grado cuyos coeficientes pueden ser expresados en términos de los primeros períodos. Tenemos

$$\sigma + \sigma_1 = \rho, \qquad \sigma_2 + \sigma_3 = \rho_1$$

y por multiplicación

$$\sigma\sigma_1 = -1, \qquad \sigma_2\sigma_3 = -1.$$

Entonces las ecuaciones cuadráticas son

$$x^2 - \rho x - 1 = 0, \qquad x^2 - \rho_1 x - 1 = 0.$$

Para la solución de estas ecuaciones, teniendo en cuenta anotaciones sobre signo de estos períodos, encontramos

$$\sigma = \frac{\rho + \sqrt{\rho^2+4}}{2}, \qquad \sigma_1 = \frac{\rho - \sqrt{\rho^2+4}}{2},$$

$$\sigma_2 = \frac{\rho_1 + \sqrt{\rho_1{}^2+4}}{2}, \qquad \sigma_3 = \frac{\rho_1 - \sqrt{\rho_1{}^2+4}}{2}.$$

2. Para el menor primo de Fermat $p = 3$, tenemos

$$x^2 + x + 1 = 0.$$

Substituyendo $x = \xi$ en esta ecuación, y después de dividir por ξ obtenemos

$$\eta + 1 = \xi + \xi^{-1} + 1 = 0.$$

Por (1) tenemos

$$\cos \frac{2\pi}{3} = -\frac{1}{2},$$

y para el lado del polígono, se encuentra entonces

$$S_3 = \sqrt{3}.$$

Esto es, el polígono regular de tres lados es construible con regla y compás.

3. Para el siguiente primo de Fermat $p = 5, \xi$ satisface la ecuación

$$\xi^4 + \xi^3 + \xi^2 + \xi + 1 = 0,$$

y dividiendo por ξ^2 se tiene

$$\xi^2 + \xi^{-2} + \xi + \xi^{-1} + 1 = 0.$$

Haciendo $\eta = \xi + \xi^{-1}$ y elevando al cuadrado obtenemos $\eta^2 - 2 = \xi^2 + \xi^{-2}$. Al sustituir estos valores en la ecuación, vemos que η es raíz de la cuadrática

$$\eta^2 + \eta - 1 = 0.$$

Una solución de esta ecuación es

$$\eta = \frac{\sqrt{5} - 1}{2}.$$

Seleccionamos $\eta > 0$. Obtenemos $\cos \frac{2\pi}{5} = \frac{\sqrt{5} - 1}{4}$ y el lado del pentágono es entonces $s_5 = \frac{1}{2}\sqrt{10 - 2\sqrt{5}}$.

4. Para el primo de Fermat $p = 17$, Gauss utiliza los residuos cuadráticos módulo 17. Para $p = 17$ cuando la raíz $x = \xi$ se substituye y la ecuación se divide por ξ^8, se sigue que

$$\xi + \xi^{-1} + \xi^2 + \xi^{-2} + \cdots + \xi^8 + \xi^{-8} = -1$$

En éste punto Gauss introduce dos cantidades que llama *primeros períodos*

$$\rho = \xi + \xi^{-1} + \xi^2 + \xi^{-2} + \xi^4 + \xi^{-4} + \xi^8 + \xi^{-8}$$

$$\rho_1 = \xi^3 + \xi^{-3} + \xi^5 + \xi^{-5} + \xi^6 + \xi^{-6} + \xi^7 + \xi^{-7}.$$

Los exponentes en la primera expresión son los residuos cuadráticos módulo 17 y los de la segunda los no residuos cuadráticos módulo 17 (soluciones de $x^2 \equiv a (\mathrm{mod} 17)$).

Bibliografía

Aleksandrov, A.D, A.N.Kolmogorov & M.A.Lavrentlev,eds. *Mathematics its Content, Meanings and Methods.*. Translated by S.H. Goul and T. Bartha. Cambridge, Mass. The MIT Press, 1986.

Arcavi Abraham y Maxim Bruckheimer. *'Reading Bombelli's x-purgated Algebra".* The College Mathematics Journal. Vol 22, No. 3 Mayo,1991. pp212-219.

Barbeau, E.J. *Polynomials.* Problem Books in Mathematics. Springer-Verlag, 1989.

Baron, Margaret. *The Origins of the Infinitesimal Calculus.* Dover. New York. 1969.

Behnke, H et al. *Fundamentals of Mathematics*, Vol I, MIT Press. Cambridge. 1990.

Birkhoff, G. MacLane, S. *A Brief Survey of Modern Algebra.* The Macmillan Company. New York. 1953.

Bottazini, Umberto. *The Higher Calculus: A History of Real and Complex Analysis from Euler to Weierstrass.*

Boyer, Carl.B. *The History of the Calculus and its Conceptual Development.* New York. Dover. 1959.

Budden, Frank. *The Fascination of Groups.* Cambridge, Cambridge. University Press. 1985.

Burnside, W. *Theory of Groups of Finite Order.* Dover. New York. 1955.

Burnside, W. *Theory of Groups of Finite Order.* Second Edition, Dover, Publications Inc. New York. 1955.

Burton, David. *The History of Mathematics: An Introduction.* Wm. Brown Publishers. Dubuque, Iowa. 1985.

Burton, L. Mason, J. Stacey, K. *Pensar Matemáticamente.* Labor. Madrid. 1982.

Burton, L. *Thinking Things Through.* Oxford. Blackwell. 1984.

Canadian Mathematics Competition: *Problems, Problems, Problems.* Vol 5. Faculty of Mathematics. University of Waterloo. Waterloo. Canada. 1992.

Caycedo, José Francisco. *Teoría de Grupos.* Departamento de Matemáticas Universidad Nacional de Colombia, Santafé de Bogotá, 1987.

Se pueden substituir entonces los valores de ρ y ρ_1, para obtener expresiones explícitas para los segundos períodos. Por ejemplo,

$$\sigma = \frac{1}{4}\left(\sqrt{17} - 1 + \sqrt{34 - 2\sqrt{17}}\right), \quad \sigma_2 = \frac{1}{4}\left(-\sqrt{17} - 1 + \sqrt{34 - 2\sqrt{17}}\right).$$

Finalmente se necesitan dos períodos de tercer orden

$$\eta = \xi + \xi^{-1} = 2\cos\frac{2\pi}{17}, \qquad \eta_1 = \xi^4 + \xi^{-4} = 2\cos\frac{8\pi}{17}.$$

De nuevo, estas dos cantidades satisfacen una ecuación de segundo grado cuyos coeficientes pueden ser expresados por medio de los segundos períodos

$$\eta + \eta_1 = \sigma, \qquad \eta\eta_1 = \sigma_2,$$

y la ecuación es entonces

$$x^2 - \sigma x + \sigma_2 = 0.$$

Las expresiones para η y η_1 muestran que $\eta > \eta_1$ y, por tanto, las soluciones de la ecuación cuadrática están dadas por

$$\eta = 2\cos\frac{2\pi}{17} = \frac{\sigma + \sqrt{\sigma^2 - 4\sigma_2}}{2}.$$

En esta expresión se substituyen los valores de σ y σ_2, y aparece la fórmula en términos de raíces cuadradas, que puede ser usada para calcular S_{17} el lado del polígono regular de 17 lados.

Punto de investigación

Usando el método de Gauss, demostrar que las raíces séptimas primitivas de la unidad pueden ser obtenidas por sucesivas soluciones de polinomios de grado 2 y 3.

10.3 Conclusiones

No podemos dejar esta exploración extendida de la solución de ecuaciones polinomiales sin recalcar las implicaciones de los resultados que hemos obtenido.

No existe una fórmula de solución para las ecuaciones polinomiales con coeficientes racionales en general, es decir, si la ecuación es irreducible y de grado mayor o igual que 5, en general no es soluble por radicales. Los métodos numéricos pueden aproximar las raíces reales, pero no tienen relevancia para las raíces complejas. La solución por factorización es viable en el caso en que la respectiva función polinómica no sea irreducible sobre los racionales (o en que sus factores irreducibles sean de grado menor o igual a 4). Sin embargo, la teoría de Galois nos dice que existe un campo de descomposición para cada polinomio con coeficientes racionales, es decir, la factorización existe, pero no hay en general ninguna forma práctica para efectuarla.

Todo esto presentaría un panorama muy desolador si no fuera por las nuevas direcciones que logra tomar el álgebra al concentrarse en el estudio de las estructuras algebraicas y en cuestiones de la existencia de cierto tipo de soluciones o de cierto número de soluciones, sin preocuparse por dar procedimientos efectivos para calcular éstas.

Courant, Richard. *What is Mathematics?.* Oxford University Press. New York. 1978.

Descartes, René. *Geometry.* Traducido del francés por D.E. Smith y Marcia Lathan. Dover. New York. 1954.

Dickson, L.E. *Introduction to the Theory of Algebraic Equations.* 1903.

Dunham, William. *Euler and the Fundamental Theorem of Algebra.* College Mathematics Journal. Vol 22. No 4.

Edwards, Harold. *Galois Theory.* Springer-Verlag. 1984.

Engel, Arthur. *Exploring Mathematics with Your Computer.* Washington. The Mathematical Association of America. 1993.

Eves, Howard. *An Introduction to the Foundations and Fundamental Concepts of Mathematics.* Hold, Rinehart and Winston. New York. 1965.

Fennema, E. and M. Loef. *Teachers' Knowledge and its Impact.* En D.A. Grouws(Ed) Handbook of Research on Mathematics Teaching and Learning. The Macmillan Company. New York. 1992.

Grouws, Douglas (Editor). *Handbook of Research on Mathematics Teaching and Learning.* The MacMillan Company. New York. 1992.

Gilain, Christian. *'Sur l'historie du théorème fondamental de l'algébre: théorie des équations et calcul intégral.'* Memoire presenté par U. Bottazini en Archive for History of Exact Sciences. Vol 42. No 2. Springer-Verlag. 1991.

Gleason, A.M. R.E. Greenwod and L.M. Kelly. *The William Lowell Putnam Mathematics Competition: Problems and Solutions 1938-1964.* The Mathematical Association of America. 1980.

Greitzer, Samuel. *Arbelos.* Mathematical Association of America. Enero 1983, pp 10-17.

Greitzer, Samuel. *Un Modelo de Entrenamiento.* Universidad Antonio Nariño. Olimpiadas Colombianas de Matamáticas. Bogotá. 1988.

Hadlock, Charles. *Field Theory and its Classical Problems.* Mathematical Association of America. Washington. 1978.

Hall, H.S. & S.R Knight. *Higher Algebra: a Sequel to Elementary Algebra for Schools.* London. MacMillan @ Co. Ltd. 1957.

Hall, Marshall. *Theory of Groups.* The Macmillan Company. New York. 1959.

Heath, Sir Thomas. *A History of Greek Mathematics.* Vol 2. Dover. New York. 1980.

Herstein, I.N. Algebra Moderna. Editorial F. Trillas. México. 1970.

Houston, W. Robert (Ed) *Handbook of Research on Teacher Education.* The MacMillan Company. New York. 1990.

Jacobson, N. *Lectures in Abstracts Algebra.* Vol 1. Van Nostrand Company. Inc. 1957.

Kaplansky, Irving. *Fields and Rings.* Chicago. The University of Chicago Press. 1970.

Kline, Jacob. *Greek Mathematical thought and the originin of Algebra. Dover.* New York. (Translated by Eva Brann). 1992.

Kline, Morris. *Mathematical Thought from Ancient to Modern Times.* New York. Oxford University Press. 1972.

Lang, Serge. *Algebra.* Adisson Wesley Publishing Company Inc. 1965.

Long, Cliff and Thomas Hern. *Graphing the Complex Zeros of Polynomials Using Modulus Surfaces.* The College Mathematics Journal. Vol 20. No 2. 1989.

Maxfield, John E. y Mayaret W. Maxfield. *Abstract Algebre and Solution by Radicals.* Dover. New York, 1971.

McCoy, Neal.H. *The Theory of Rings.* The MacMillan Company. New York. 1968.

Meserve, Bruce. *Fundamental Concepts of Algebra.* Dover. New York. 1982.

Newton, Isaac. *The Mathematical Papers of Isaac Newton.* Vol III, 1670-1673. D.T. Whiteside, Editor. Cambridge University Press. 1969.

Olds, C.D. *Continued Fractions.* The Mathematical Association of America. Washington. 1963.

Ore, Oystein. *Number Theory and its History.* Dover. New York. 1968.

Pascal, Blaise. *On the Arithmetic Triangle* en Smith, D.E. *A Source Books in Mathematics.*. Dover. New York. 1959.

Polya, G. *Mathematical Discovery: On Understanding, Teaching and Learning Problem Solving. Vol I.* John Wiley & Sons. New York. 1962.

Sawyer, W.W. *A Concrete Approach to Abstract Algebra.* Dover. New York. 1959.

Scharlau, Winfried & Opolka Hans. *From Fermat to Minkowski.* Springer-Verlag. 1985.

Smith, D.E. *History of Mathematics. Vol II.* New York. Dover. 1959.

Uspenski, J.V. *Teoría de Ecuaciones.* Limusa. México. 1958.

van der Waerden, B.L. *Geometry and Algebra in Ancient Civilizations.* Springer-Verlag. 1983.

van der Waerden, B.L. *A History of Algebra: From al-Khwarizmi to Emmy Noether.* Springer-Verlag. 1985.

Warner, Seth. *Modern Algebra.* Dover. New York. 1990.

Zassenhauss, Hans. *On the Fundamental Theorem of Algebra.* American Mathematical Monthly. Mayo, 1967.