

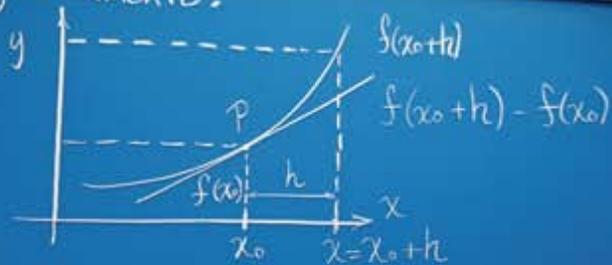
BC es igual a la suma del rectángulo BMLD

Los rectángulos AELM y BMLD suman el cuadrado de la hipotenusa AB

$$ME = \frac{AM \times ML}{2}$$



gráficamente:



Ejemplo:

Calcular la derivada de la función $f(x) = 3x^2$, en el punto $x=2$
se sabe que $x = x_0 + h$

siendo $x_0 = 2$

$$f'(2) = \lim_{h \rightarrow 0} \frac{f(2+h) - f(2)}{h}$$

$$= \lim_{h \rightarrow 0} \frac{3(2+h)^2 - 3(2)^2}{h}$$

$$= \lim_{h \rightarrow 0} \frac{3(4+4h+h^2) - 12}{h}$$

$$= \lim_{h \rightarrow 0} \frac{12+12h+3h^2 - 12}{h}$$

$$= \lim_{h \rightarrow 0} (12+12h) = 12+0 = 12$$

Regla de la cadena

$$h(x) = g[f(x)]$$

$$\Rightarrow h'(x) = g'[f(x)] \cdot f'(x)$$

se expresa

$$\frac{dh}{dx} = \frac{dg}{df} \cdot \frac{df}{dx}$$

$$\Rightarrow \frac{d}{dx} = \frac{d}{dy} \cdot \frac{dy}{dx}$$

Ejemplo:

$$f(x) = 2x$$

$$\frac{d}{dx}(f(g(x))) = f'(g(x)) \cdot g'(x)$$

$$= 2g'(x) = 2 \cdot 1 = 2$$

$$= 4(2x+1)$$

$$= 8x+4$$

$$b^2 - 4ac$$

$$5, b = -3, c = -2$$

$$(-3)^2 - 4(5)(-2)$$

$$= 9 + 40$$

$$= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$= \frac{-(-3) \pm \sqrt{(-3)^2 - 4(5)(-2)}}{2(5)}$$

$$x_1 = \frac{3 + \sqrt{49}}{10}, x_2 = \frac{3 - \sqrt{49}}{10}$$

$$x_1 = \frac{9+7}{10} = \frac{16}{10}$$

$$x_2 = \frac{9-7}{10} = \frac{2}{10}$$

Ecuación 3er grado

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

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\Delta}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\Delta}} - \frac{a}{3}$$

donde

$$p = \frac{3b - a^2}{3}, q = \frac{2a^3 - 9ab + 27c}{27}$$

$$\Delta = \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3$$

si $p = -3uv$

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\Delta}}$$

$$v = \sqrt[3]{-\frac{q}{2} - \sqrt{\Delta}}$$

$$v^3 = -\left(\frac{p}{3}\right) \left(-\frac{q}{2} - \sqrt{\Delta}\right)$$

$$v^3 = -\frac{q}{2} - \sqrt{\Delta}$$

$$= \pi \cdot \left(2L \frac{\Delta r}{2} + 2r \frac{\Delta L}{2}\right)$$

$$= \pi \cdot (L \Delta r + r \Delta L)$$

por semejanza de triángulos

$$\frac{\Delta r}{\Delta L} = \frac{r}{L}$$

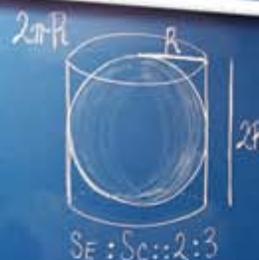
$$\therefore L \Delta r = r \Delta L$$

$$\Delta S = 2\pi r \Delta L$$



$$\sum \Delta Si = 2\pi R H$$

donde H es la altura total cubierta de la esfera



$$V_c = \frac{1}{3} \pi r^2 h$$

$$V_o = \frac{4}{3} \pi R^3$$

$$V_g = \pi r^2 h$$



$$S = 4\pi R^2$$

por teorema de Pit

$$L^2 = H^2 + r^2 = H^2 + R^2$$

Fue profesor en la Facultad de Ciencias, miembro fundador del CIMAT, investigador y director del IIMAS, profesionista autónomo en España especializado en el desarrollo de software educativo, Director técnico del proyecto MALTED de la Comisión Europea y director de Desarrollo Tecnológico en el Instituto Latinoamericano de la Comunicación Educativa. Actualmente es técnico académico en el Instituto de Matemáticas de la UNAM. Creador de la herramienta DESCARTES con la que se desarrollan contenidos educativos interactivos de matemáticas y física en España y México. Sus intereses son el desarrollo de herramientas para la creación de contenidos digitales interactivos y la publicación en web de tales contenidos para la enseñanza, la difusión y la investigación.

**JOSÉ LUIS
ABREU LEÓN**

Estudió matemáticas en la Universidad de Zúrich y obtuvo el doctorado en el Instituto de Matemáticas de la UNAM. Por su tesis doctoral recibió el premio Weizman. Desde 1998 trabaja como investigador en el Instituto de Matemáticas. Es miembro del Sistema Nacional de Investigadores y de la Academia Mexicana de Ciencias. Ha participado en la Maestría en Docencia para la Educación Media Superior (MADEMS), en la elaboración de la serie de videos Aventuras Matemáticas y también, en el proyecto IXTLI donde fungió como co-responsable. Tiene tres libros publicados, uno para nivel secundaria, otro de bachillerato y el último, de licenciatura.

MICHAEL BAROT

Estudió la licenciatura en matemáticas en la Facultad de Ciencias de la UNAM y obtuvo el doctorado en el Instituto Tecnológico de Massachusetts (MIT). A partir de entonces es profesor de la Facultad de Ciencias e investigador del Instituto de Matemáticas de la UNAM, del cual es actualmente director. Es, desde su inicio, miembro del Sistema Nacional de Investigadores. Recibió la Distinción Universidad Nacional para Jóvenes Académicos en Docencia en Ciencias Exactas en 1993. Tiene dos libros publicados, uno de divulgación y otro de texto en el Fondo de Cultura Económica, donde también participa en el comité de Ciencias.

JAVIER BRACHO

AGRADECIMIENTOS

En la elaboración de este libro participaron muchas personas a las que debemos agradecer. En las múltiples sesiones de discusión, planeación y diseño conceptual participaron Concepción Ruiz Ruiz-Funes, Emiliano Mora y Paloma Zubieta López. Concha y Paloma también contribuyeron con textos de algunos apartados y Paloma, además, corrigió el estilo y preparó los textos para su formación; también coordinó las fases finales de producción del libro y, cual hada madrina, juntó los fragmentos para hacerlo realidad. Patricia Covarrubias y Juan Andrés Burgueño contribuyeron en el apartado de Estadística. Las maestras Ma. Eugenia Otero Ulibarri y Dora Lidia Rodríguez Zúñiga del Colegio de Ciencias y Humanidades, junto con los maestros Heriberto Marín Arellano y Emilio Velarde González Baz de la Escuela Nacional Preparatoria tuvieron la amabilidad de leer nuestros manuscritos y escuchar nuestras ideas e intenciones; sus críticas y comentarios, siempre atinados y constructivos, influyeron positivamente en el resultado. Les agradecemos el tiempo y dedicación que generosamente nos brindaron.

Por la producción del DVD agradecemos a las siguientes personas: Oscar Escamilla coordinó la producción general y la de los materiales interactivos. El propio Oscar, junto con Carlos Alberto Jaimes, Abraham Pita, Julio Prado, Carlos Serrato y Erika Tovilla diseñaron y programaron los materiales interactivos en los que Mariana Villada Carbó corrigió el estilo y la redacción. Carlos Alberto Jaimes coordinó la producción de los videos y también creó la mayoría de las ilustraciones y animaciones. Roberto Elier, Mariana Villada y Paloma Zubieta escribieron los guiones. Felipe Bonilla fue el realizador de los videos y estuvo a cargo también, de la edición y la posproducción. La voz en los videos es de Pablo Flores. Las fotografías usadas en los fondos son de Yamina del Real. El diseñador gráfico de todo el contenido del disco es Alfonso Pascal. En la asesoría académica de los contenidos de los videos participó Gonzalo Zubieta Badillo.

Con seguridad olvidamos mencionar la generosa contribución de algunas personas que apoyaron la realización de este libro. A ellos también damos las gracias y rogamos disculpen la omisión. Finalmente, reconocemos el excelente trabajo de edición que orquestaron Rosanela Álvarez desde la UNAM y María Oscos desde Siglo XXI.

Este libro es una suerte de mosaico y caleidoscopio. Mosaico, pues consta de piezas independientes que se ensamblan en un todo, y caleidoscopio porque repite y repite la misma imagen reflejada en tres espejos: la imagen es la creación matemática; los espejos son la actividad humana, la naturaleza y la matemática misma. Ante todo, las matemáticas son una actividad creativa —cada pieza del mosaico da muestra de ello—, surgen de considerar problemas de diversa índole cuyas soluciones van armando, como si fueran piezas en un rompecabezas, una estructura de conocimiento de una consistencia y solidez sorprendente. Esta contundencia reside en que se crean o recrean dentro de cada mente humana que se acerca a ellas, en que se basan en la razón a tal grado, que trascienden culturas y épocas históricas y en que, siendo tan abstractas resultan estar ligadas a la realidad de maneras insólitas y fundamentales. Sin ellas, lo que llamamos ciencia no existiría y nuestra vida cotidiana actual sería impensable.

Es muy común que en una clase de matemáticas surja la pregunta ¿esto para qué sirve? En este libro, además hacemos otra: ¿por qué se creó, de dónde nace? Al responderla, identificamos tres fuentes básicas: la actividad humana, la comprensión de la naturaleza y la matemática misma, que se tratan en los temas 2, 3 y 4, respectivamente. En el tema 1, a manera de introducción, ahondamos en el porqué de esta estructura, en qué queremos decir con ella. Los otros tres temas consisten en apartados independientes y autónomos que inician desde cero. Al avanzar en la lectura de los apartados, es inevitable que ésta vaya complicándose, pues trata sobre ideas —ideas matemáticas— que requieren de tiempo, reflexión, creatividad y esfuerzo por parte del lector. Sin embargo, si alguien tropieza demasiado o desespera, puede pasar al siguiente apartado o a cualquier otro, pues tampoco están serios —salvo por el tema 1, concebido como un todo—. Podría decirse que es un libro para “picar” en el que, después de la introducción, se vale dar una probadita por aquí, otra por allá, echarse algún bocadillo completo o bien, leerlo de corrido.

No es un libro de texto, pero sí pretende ser un apoyo para el alumno y para el maestro. Este último puede encontrar nuevas ideas para tratar temas en clase, maneras distintas de enfocarlos o ejemplos interesantes para los alumnos. El estudiante podrá descubrir aspectos novedosos que fomenten su curiosidad por la materia. El público en general encontrará una visión amplia y actualizada de la actividad matemática.

El nivel en el que inicia cada apartado es de bachillerato, pero no necesariamente es el

mismo en el que acaba, a veces se vuelve técnico y otras, meramente descriptivo y cultural. Las partes más técnicas o difíciles de seguir están marcadas con el icono , que indica que se requiere de más tiempo y esfuerzo para comprenderlas e, insistimos, pueden posponerse sin menoscabo del entendimiento general. Preferimos incluirlas que omitirlas, pues las matemáticas siempre representan retos intelectuales y no pretendemos esconderlos o disfrazarlos, sino exponerlos tal cual son.

En este libro presentamos a las matemáticas sobre todo como un fenómeno cultural, y para ello escogimos algunos ejemplos de su desarrollo. Tocamos diversos temas de las matemáticas para que, cómo en la música, al escuchar diversos géneros, comprendamos mejor su variedad y riqueza. En son de broma, quisiéramos que fuera “un viaje a Acapulco”: con algo para todos los bolsillos y todos los gustos.

El libro incluye un DVD que se produjo con el mismo espíritu; contiene cápsulas históricas y biográficas además de unidades educativas interactivas. Algunas tienen relación con partes del texto pero otras son independientes. Amplían la visión que, a nuestro entender, debiera tener un bachiller de las matemáticas.

¿CÓMO SE ESCRIBEN LAS MATEMÁTICAS?

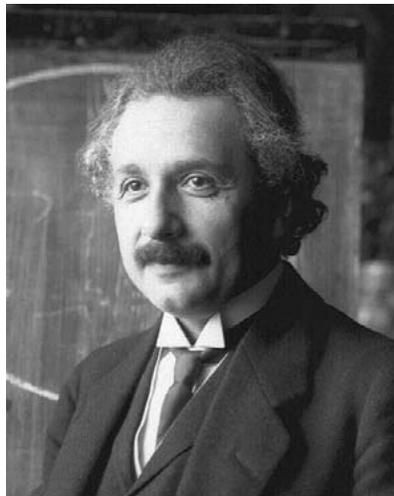
Las matemáticas son una ciencia peculiar: usan un lenguaje muy especial al que es necesario acostumbrarse. Este lenguaje, como cualquier otro, requiere ser aprendido, de lo contrario representa una barrera de entendimiento que impide acercarse, disfrutar y comprender el contenido.

En su libro *Breve historia del tiempo*, el físico británico Stephen Hawking relata su experiencia con la editorial, donde le dijeron que “por cada fórmula en tu texto, reduces tu público a la mitad”. Hawking intentó no poner una sola fórmula, pero al final, no lo pudo resistir e incluyó la fórmula más famosa de todas:

$$E = mc^2.$$

La fórmula es de Albert Einstein, físico alemán del siglo XX y es consecuencia de su Teoría general de la relatividad.

Para muchas personas esta fórmula tiene un significado casi religioso. Se sabe que significa algo, mucho, que expresa algo trascendental. Es claro que está escrita en lenguaje matemático. Se ve por el signo de igualdad —las dos rayas paralelas— que es una *ecuación*, donde se igualan los dos lados. Cada vez que aparece el símbolo de igualdad “=”, separa dos lados —el izquierdo y el derecho— y los iguala. En la ecuación de Einstein, el lado izquierdo es la simple letra E que significa la cantidad física *energía*. El lado derecho es más complicado: en él aparecen tres símbolos: m, c y un dos. La letra m significa *masa*, la letra c es la *velocidad de la luz* y el dos significa que se eleva la velocidad de la luz al cuadrado, es decir, se multiplica por sí misma. Se sobreentiende, además, que cuando dos letras que representan cantidades independientes están juntas, las cantidades se multiplican. Así que la fórmula en palabras se lee:



Einstein en 1921.

La energía es igual al producto de la masa por el cuadrado de la velocidad de la luz.

La velocidad de la luz es fija, una constante. En el vacío, la luz avanza 299 792 458 metros cada segundo, es decir, daría casi ocho vueltas a la Tierra en un segundo. Pero ¿de qué masa se está hablando, de qué energía? Lo que expresa esta fórmula es que la energía y la masa de cualquier sistema físico se pueden convertir una en la otra. Como la velocidad de la luz es tan grande, la ecuación dice que muy poca masa equivale a una energía enorme.

En este ejemplo se pueden apreciar varios aspectos:

- Una ecuación tiene dos lados, o como se dice en matemáticas, dos *miembros*.
- Cada uno de los miembros se puede constituir de diferente forma. Puede ser una simple letra o algo más complejo.
- En una ecuación se deben identificar los diferentes ingredientes: las letras usadas, que pueden ser *variables* como la masa y la energía, o *constantes*, como la velocidad de la luz. Sus significados dependen del *contexto* en el que se plantea la ecuación.
- Una ecuación debe interpretarse y su lectura no siempre es fácil. Se requiere de una buena instrucción para lograr la familiaridad con las ecuaciones. A veces, hay que leerla una y otra vez, hay que juntar sus piezas como en un rompecabezas y entonces, podrá desplegar una belleza similar a la de Einstein.

Sabemos de la dificultad que provoca el uso de las ecuaciones, fórmulas y símbolos en las matemáticas, pero no podemos prescindir de ellos. En la historia, no siempre se usó la simbología actual. Por ejemplo, en 1559, el matemático francés Jean Buteau escribía:

$$I \diamond P 6_P 9 [I \diamond P 3_P 24 \quad (1)$$

Lo que hoy se escribe como:

$$x^2 + 6x + 9 = x^2 + 3x + 24$$

Unos sesenta años antes, en 1494, el matemático italiano Luca Pacioli escribía:

Trouame.I.n^o.che.gi_to al suo quadrat^o faccia.12.

Lo que hoy se escribe como:

$$x + x^2 = 12$$

Con buena voluntad se pueden descifrar estas maneras exóticas de denotar el contenido. Por ejemplo, en la ecuación (1) hay que entender los símbolos P como *plus*, es decir, como nuestro “más”; el símbolo [como el cuadrado del número —aludiendo al área de un cuadrado con lado I—; la línea _ como la misma variable I —que corresponde a nuestra x— y el símbolo \diamond como la igualdad. El segundo ejemplo es más cercano a cómo se lee en la actualidad la ecuación en italiano. Estos ejemplos muestran que hace 500 años no había consenso sobre cómo anotar las matemáticas. Fue un proceso largo, un desarrollo de siglos en el cual, poco a poco, se establecieron ciertas convenciones. Por ejemplo, el símbolo de igualdad que usamos hoy día (=) lo utilizó por primera vez el matemático inglés, Robert Recorde, en 1557.

En los dos ejemplos anteriores de ecuaciones con la notación actual, el contexto es im-

plícito. Son diferentes de la ecuación de Einstein en que los símbolos no tienen significados extra, sino que se sobreentienden. Puesto que sólo aparece la letra x , además de los números y los símbolos de suma e igualdad, interpretamos que x también representa a un número. Plantean, entonces, la pregunta ¿existirá un número que al sustituir en vez de x , cumpla la ecuación? Eso sería resolver la ecuación: a veces se puede y otras no, como veremos más adelante en otras partes de este libro.

Se puede decir que, en gran medida, las matemáticas fueron tan prolíficas a partir del siglo XVIII gracias a una simbología y notación más simple, consensuada entre la comunidad de los matemáticos. Se pudieron expresar y comunicar mejor; además, resulta que una buena notación a veces ayuda a entender. En particular, la física moderna es absolutamente impensable sin el uso de las matemáticas y eso no quiero decir números, sino conceptos, simbología, notación y métodos involucrados.

El uso del lenguaje simbólico es un cuchillo de doble filo: por un lado, hace extremadamente eficiente la notación y el manejo de conceptos pero, por otro, constituye un obstáculo serio para entenderlos. Sería un error imperdonable pensar que las matemáticas *sólo* son fórmulas. Más bien, las matemáticas son lo que está escondido en las fórmulas y la mejor manera de explicar aquello, es decir, lo escondido, es a veces justo a través de éstas. Cualquier otra manera de intentarlo es más complicada y tortuosa.

La política que adoptamos en este libro es tratar de evitar el formalismo riguroso y procurar transmitir las matemáticas mismas. Hemos hecho un gran esfuerzo por omitir fórmulas innecesarias y por llegar a lo que está en el fondo mediante vías alternas, sin transitar por la simbología. Que esto sólo fue posible en relativamente pocos casos es consecuencia de la complejidad de las matemáticas mismas. Muchas veces la fórmula es el camino de comunicación menos malo y cuando ya se tiene familiaridad con su uso, se convierte en el camino más directo. Insistimos en esta advertencia pues sabemos de las múltiples dificultades. Este libro no se escribió para *aprender* matemáticas —creemos que esto sólo se puede lograr mediante un intercambio más dinámico y activo—, se escribió para orientar sobre los alcances de esta ciencia, sobre su significado en nuestra cultura y, en particular, sobre su inserción en la tecnología y en la vida del siglo XXI.



EL PORQUÉ DE LAS MATEMÁTICAS

TEMA

1

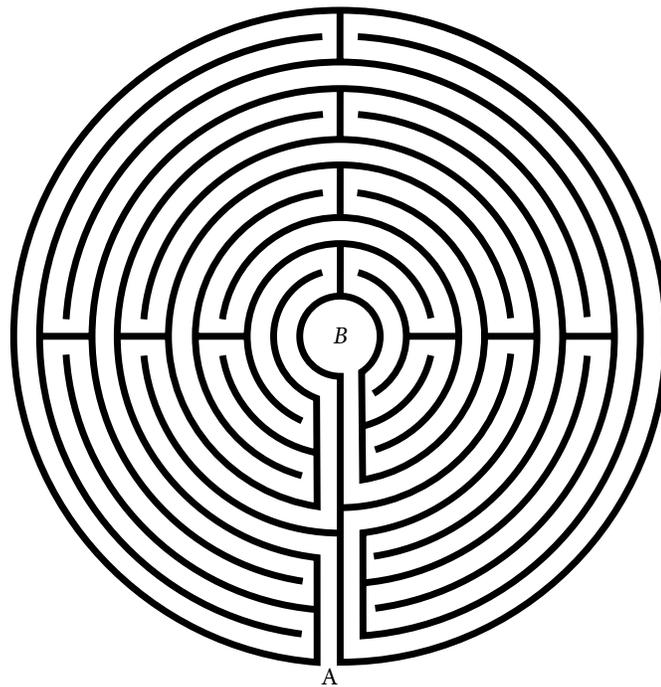


Figura 1.1 Un laberinto medieval donde no es posible perderse. Éste es el origen de los laberintos modernos con bifurcaciones y que presentan un reto para entrar y salir.

1.1 INTRODUCCIÓN

Como se explica en la introducción a la materia, este libro se divide en tres partes que corresponden a las principales fuentes de motivación para la creación matemática: la actividad humana, la naturaleza y las propias matemáticas. El propósito de este primer capítulo es ilustrar, con ejemplos sencillos, dichas fuentes de motivación matemática.

Sin embargo, es conveniente aclarar que hay una fuente de creación matemática que antecede a las otras y sin la cual serían estériles: la irresistible atracción del ser humano por los retos de todo tipo, en especial, los intelectuales. Esta atracción no se limita a un grupo especial o selecto de personas, como los matemáticos, sino que se puede reconocer en cualquier persona, por ejemplo, cuando en un café o en el metro se resuelve como entretenimiento uno de los *sudokus* impresos a diario en los periódicos. ¿Qué es lo que lleva a una persona a intentar, durante largas horas, resolver estos retos? ¿Qué gana con ello? Quienes nunca lo

han hecho les puede parecer extraño, pero lo único que se obtiene es la satisfacción de haberlo resuelto, de lograr algo con la mente.

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

Figura 1.2 Las reglas del sudoku son sencillas: las celdas deben llenarse con las cifras del 1 al 9 de tal manera que no se repita una de ellas en la misma fila, columna o cuadrado pequeño.

Hace mucho tiempo que los periódicos imprimen juegos de entretenimiento o desafíos mentales, como los crucigramas. Unos cien años atrás estaban de moda los problemas de ajedrez, hace unos decenios el cubo de Rubik y ahora, los *sudokus*. Los periódicos están al corriente de que a muchos de sus lectores les encantan estos juegos y de que hay quienes los compran sólo por ellos, por consiguiente, siempre tienen secciones que los contienen.

Antes de los *sudokus* o de otros desafíos ya mencionados, muchas personas dedicaban su tiempo libre a encontrar **cuadrados mágicos** que son arreglos de números en los que se obtiene la misma suma en cada fila y columna. Podemos ver un ejemplo en el grabado *Melancolía* de Alberto Durero, arriba y a la derecha, bajo la campana.

La figura 1.4 es un cuadrado mágico más grande que el del cuadro de Durero donde la suma de cada fila y columna es de 65.



Figura 1.3 Melancholia, de Durero.

9	22	20	1	13
21	5	7	24	8
12	19	2	14	18
17	4	25	3	16
6	15	11	23	10

Figura 1.4 Cuadrado mágico de 5x5.

Los cuadrados mágicos eran conocidos desde el año 650 a.C. por los matemáticos chinos y, alrededor del siglo VII d.C. por los árabes. Aparecen también en las culturas de India y Persia y, en cada una de ellas, se les atribuyen distintos poderes como el de atraer la suerte.

Podemos construir fácilmente un cuadrado mágico de tamaño tres por tres. Primero determinamos cuál debe ser la suma de los números en cada fila y columna. Si usamos los números del 1 al 9 la suma es:

$$1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 = 45$$

Como esta suma debe distribuirse en tres renglones, cada uno debe sumar 15. Si en alguna celda ponemos el número más grande, o sea el 9, debemos sumarle solamente 6 en las otras dos celdas de su misma fila y su misma columna. La única forma de sumar 6 con los números restantes son: $1 + 5$ y $2 + 4$, ya que $3 + 3$ repite un número. Usamos estas dos combinaciones para llenar la fila y la columna donde ya habíamos colocado el 9 y tenemos que:

9	1	5
2		
4		

En las cuatro celdas restantes debemos colocar los cuatro números que faltan: 3, 6, 7 y 8. Buscaremos ahora la ubicación del número 7. No puede estar en la misma fila que el 4 dado que $4 + 7 = 11$ y nos faltarían justo otros 4 para alcanzar la suma 15. De la misma manera, no puede estar en la columna del 1, pues faltarían 7. Por ello, sólo queda un lugar para el 7: debe estar en la columna del 5 y la fila del 2. Ahora, es fácil terminar de rellenar el cuadrado mágico. Se deja este ejercicio para que el lector lo termine.

¿Qué pasos seguimos para construir el cuadrado mágico de 3 por 3? No se probaron muchas distribuciones para ver, si de casualidad, una funcionaba. Tampoco seguimos una estrategia ordenada para evaluar todos los posibles arreglos sin repetir uno, hasta encontrar el que buscábamos. Simplemente, empleamos el razonamiento lógico, ningún otro procedimiento puede llevarnos al resultado. Cuando logramos resolver un problema por deducción lógica, sentimos la misma satisfacción que un niño al poner la última pieza de un rompecabezas.

Usemos este cuadrado mágico para formar un *sudoku*: si intercambiamos filas o columnas obtenemos otros cuadrados mágicos de 3 por 3 que podemos acomodar en el tablero, por ejemplo, de la siguiente manera:

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

Figura 1.5 Sudoku que contiene nueve cuadrados mágicos.

Este *sudoku* terminado consiste en nueve cuadrados mágicos; si además elegimos, por ejemplo, la esquina superior izquierda de cada bloque, obtenemos un arreglo de tres por tres números que, a su vez, forman otro cuadrado mágico. Lo mismo ocurre para cualquier otra celda que elijamos en cada bloque.

Se pueden hacer cuadrados mágicos de todos los tamaños: tres por tres, cuatro por cuatro, cien por cien o tres millones por tres millones. También de uno por uno, que consiste en una única celda, pero como es el ejemplo más sencillo, no presenta mucho interés. No obstante, hay un tamaño imposible para los cuadrados mágicos. ¿Cuál es? Dejamos este reto al lector.

Los cuadrados mágicos no tuvieron mayor trascendencia en la historia ni en la ciencia. En un principio, quizá se emplearon con la esperanza de encontrar algo verdaderamente mágico pero, con el tiempo, se convirtieron en un simple pasatiempo. Sin embargo, cautivaron por igual a hombres y mujeres de distintas épocas y culturas, al evidenciar la fascinación de la humanidad por los retos intelectuales.

Si bien no todas las personas responden a estos desafíos, hay algo en la naturaleza humana que impulsa a plantearse problemas e intentar resolverlos. Este impulso es característico de la actividad matemática, aunque las matemáticas son mucho más que eso según veremos en las siguientes secciones.

1.2 LA ACTIVIDAD HUMANA



Figura 1.6 Mapa de Mercator de 1569. Mapamundi de gran utilidad para los navegantes diseñado por el matemático alemán Gerardus Mercator. Hasta la fecha esta sigue siendo la manera más usada de presentar la superficie de la Tierra en el plano.

Cualquier alumno debe haberse preguntado alguna vez en clase: ¿esto para que me va a servir? ¿Dónde están las matemáticas en la vida cotidiana? Sorprendentemente, basta con usar el teléfono celular o una computadora para estar cerca de las matemáticas... si se aprende a encontrarlas.

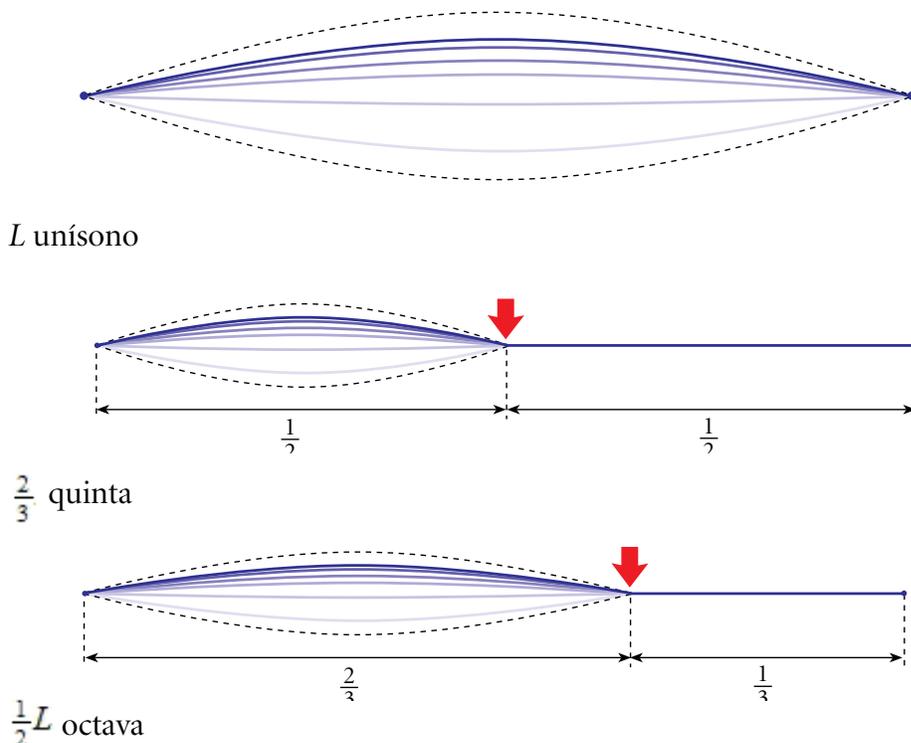
En esta sección presentamos tres problemas que pueden resultar divertidos y, hasta cierto punto, mostrar cómo las matemáticas se usan para resolver problemas de la vida cotidiana. El primero tiene que ver con la música, el segundo con la arqueología y el tercero con medir la profundidad de un pozo.

1.2.1 Música y matemáticas

La relación entre la música y las matemáticas era ya conocida por las culturas caldea, egipcia, babilónica y china; sin embargo, hasta la escuela pitagórica, en la Grecia del siglo VI a.C., fue cuando estas actividades humanas quedaron unidas para siempre por la teoría de la cuerda vibrante.

Si se hace vibrar una cuerda, el sonido que produce depende de su longitud, su grosor y su tensión. El tono es más agudo conforme la cuerda se acorta. Lo que se descubrió en la escuela pitagórica es que, al dividir la cuerda, hay proporciones que producen sonidos más agradables que otros. Por ejemplo, si se divide la cuerda justo a la mitad, la vibración tiene un tono de una octava mayor que la producida por la longitud original. Por ello, la octava juega un papel fundamental en nuestra comprensión de la música.

Los pitagóricos establecieron cuatro intervalos o relaciones entre las longitudes de las cuerdas que producían las únicas consonancias permitidas, es decir, aquellos sonidos que podían escucharse simultáneamente con un efecto agradable. Para producir todas las notas musicales sólo se tienen estos cuatro intervalos y sus combinaciones. El papel fundamental de las fracciones en la música era, sin duda, una de las razones por las cuales Pitágoras consideraba que la esencia de la realidad sólo podía expresarse por medio de números.



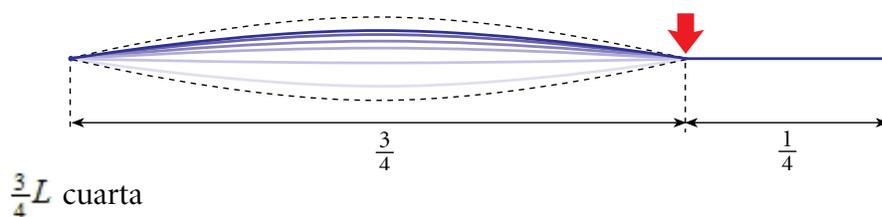


Figura 1.7 Intervalos entre las longitudes de las cuerdas. (incluye todas las imágenes anteriores).

La teoría de la cuerda vibrante se puede resumir de la siguiente manera: si se tensa una cuerda y se hace vibrar, emite un sonido de un tono; si se hace vibrar la mitad de la cuerda, el tono aumentará un octavo; si se hace vibrar dos tercios de la cuerda, el tono estará un quinto por encima del que produjo la cuerda entera; si se hace vibrar tres cuartos de la cuerda original el tono estará una cuarta por arriba del tono original.

Al vibrar, una cuerda o cualquier otro cuerpo transmite su vibración al aire que la rodea; estas alteraciones en la densidad del aire se propagan en forma de ondas y cuando llegan a nuestros oídos, las percibimos como sonido.

Una de las características más importantes de una onda sonora es su frecuencia, que se mide en hertzios —un hertzio equivale a una oscilación por segundo. El oído de un niño sano percibe sonidos de 12 a 20 000 hertzios pero, al envejecer, este rango disminuye, especialmente para la percepción de sonidos agudos.

Para darnos una idea diremos que un bajo, es decir, la voz humana más grave, canta en el rango de 80 a 300 hertzios y una soprano, con voz muy aguda, alcanza un rango entre 220 a 1 000 hertzios.

Las notas musicales son sonidos puros en los que únicamente está presente una frecuencia. La relación que existe entre la frecuencia de distintos sonidos es muy importante y se muestra en la siguiente tabla:

Do	Re	Mi	Fa	Sol	La	Si	Do
1	$\frac{9}{8}$	$\frac{5}{4}$	$\frac{4}{3}$	$\frac{3}{2}$	$\frac{5}{3}$	$\frac{15}{8}$	2

Figura 1.8 Frecuencia relativa de los intervalos a partir del do.

La razón de las frecuencias se llama intervalo o distancia y es el cociente entre las frecuencias. Por ejemplo, el intervalo entre fa y la es $\frac{4}{3} + \frac{5}{3} = \frac{4}{5}$, igual al intervalo entre do y mi.

Si silbamos con los labios, el sonido es casi de una sola frecuencia, pero si hablamos o se toca un instrumento, se producen varias frecuencias a la vez. Usualmente hay una frecuencia predominante, la principal, y otras secundarias de menor intensidad que tienen cierta relación matemática con la frecuencia principal y se llaman armónicos. La intensidad con la cual se emiten los armónicos hace que los instrumentos musicales suenen diferentes entre sí.

Aunque los pitagóricos nunca hablaron de armónicos, determinaron que las cuerdas de longitudes con razones 1 : 2 y 2 : 3 producían combinaciones de sonidos muy agradables y, a partir de estas proporciones, construyeron una escala musical.

Los pitagóricos fueron los primeros en establecer la música como una disciplina matemática, una de las siete fundamentales que los jóvenes tenían que estudiar en la escuela. Las relaciones que establecieron hace más de 2 500 años entre las matemáticas y la música, son vigentes y siguen presentes en cualquier sala de conciertos. Una vez que un descubrimiento

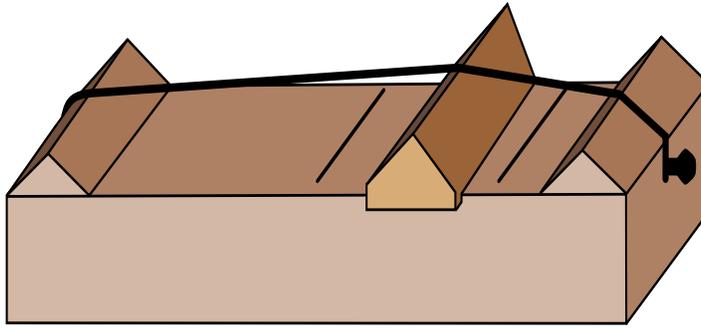


Figura 1.9 Los pitagóricos usaban un instrumento, el monocordio, para producir una escala todavía usada en la actualidad y llamada pitagórica diatónica.

matemático se ha establecido, queda para siempre en la vida cotidiana. ¡Que no nos extrañe que los violines se afinen hoy día como los habrían afinado los miembros de la escuela de Pitágoras!

1.2.2 La pirámide de los nichos

En marzo de 1785, don Diego Ruiz, recolector de impuestos del gobierno español, buscaba sembradíos clandestinos de tabaco en la selva tropical a las afueras de Papantla, Veracruz. No sabemos si halló lo que buscaba, pero lo que sí encontró fue una pirámide de grandes dimensiones que hoy forma parte del sitio arqueológico conocido como El Tajín. Hasta 1938 empezó la limpieza y reconstrucción del lugar, formado por varios edificios y canchas de juegos de pelota. La peculiaridad del sitio es una pirámide que tiene, en cada uno de sus costados, numerosos nichos a partir de los cuales toma su nombre. Se dice que tiene 365 nichos, tantos como los días del año, por lo que se piensa que la pirámide cumplió alguna función similar a la de un calendario.



Figura 1.10 La foto muestra el estado de la pirámide en 1913, antes de que fuera reconstruida, por el costado este, el único que tiene una rampa construida encima de la estructura antigua. A causa de esta rampa, el conteo de los nichos no ha sido fácil, pero varios arqueólogos afirman que la pirámide nunca tuvo 365 nichos.

¿Cómo se pueden acomodar 365 nichos en una pirámide? Esa pregunta se la hicieron con toda seguridad los totonacas antes de empezar con la construcción de la pirámide.

Aunque acomodar nichos en una pirámide no sea un problema de nuestra vida cotidiana, sí lo fue, al menos, para los sacerdotes y sabios encargados de aquella construcción. Si bien las matemáticas del bachillerato son muy pocas veces útiles de manera directa en la vida cotidiana, nos permiten asomarnos de una forma distinta a nuestra cultura: mediante una comprensión más profunda de la tecnología que nos rodea, nos hacen partícipes de los logros de la humanidad.

Volviendo al problema de acomodar nichos, primero hay que observar que 365 no es divisible entre 4, es decir, no es posible acomodar el mismo número de nichos en cada costado tal que en total sumen 365. Pero el número 365 es divisible entre 5: $365 = 5 \cdot 73$. Entonces habría que construir una pirámide con cinco lados y de ese tipo, no hay en México; con seguridad, esto no les hubiera parecido una buena idea.

Si se coloca un nicho en la punta quedan 364 nichos para ser acomodados en los cuatro costados de una pirámide auténtica, como la de El Tajín. Dado que $364 = 4 \cdot 91$, habría 91 nichos en cada lado. La distribución de los nichos en cada lado de la pirámide se puede hacer de muchas maneras; por ejemplo, se podrían acomodar 40 nichos en el nivel más bajo, 30 en el que sigue y 21 en el tercero y último, pero quedaría una pirámide muy ancha con sólo 3 niveles. Además, pasaría que del nivel más ancho al del medio se disminuye el número de nichos en 10, mientras que del medio al último se disminuye en 9. Esto ocasionaría que los nichos en el último nivel tendrían que ser más pequeños que el resto o que la pirámide no tuviera un borde regular. Dicho lo anterior, sería mejor que de nivel en nivel siempre disminuyera el mismo número de nichos. Por ejemplo, podríamos pensar en la sucesión 3, 5, 7, 9, 11, 13, 15, 17 para el número de nichos en cada nivel. Esto nos daría una pirámide regular con $3 + 5 + 7 + 9 + 11 + 13 + 15 + 17 = 80$ nichos. Si agregamos otro nivel, éste tendrá 19 nichos y se excedería de los 91 que deben ser. Al parecer, la distribución de 365 nichos en una pirámide no es un problema tan sencillo.

Las sucesiones de este estilo se llaman **sucesiones aritméticas** y han sido estudiadas. Entre las más famosas está, por ejemplo, aquella que empieza con uno y aumenta siempre de uno en uno: $1 + 2 + 3 + 4 + 5 + 6 + 7 + 8$. Se cuenta que un maestro de matemáticas de Alemania pidió a sus alumnos calcular la suma de los primeros 100 números para tenerlos ocupados por un buen rato; uno de ellos entregó su respuesta después de un minuto mientras los demás necesitaron casi la hora completa. La gran sorpresa fue que la primera respuesta era de las pocas correctas, resuelta por Carl Friedrich Gauss, un joven que el mundo reconocería después como uno de los matemáticos más influyentes de todos los tiempos.

No es que Gauss calculara muy rápido, sino que usó un truco que él inventó: si se suman los números en pares, tomándolos de los extremos así:

$$(100 + 1) + (99 + 2) + (98 + 3) + \dots + (53 + 48) + (52 + 49) + (51 + 50)$$

Entonces, cada par suma 101 y hay 50 pares. Así que el resultado es:

$$101 \cdot 50 = 5050$$



Figura 1.11 Carl Friedrich Gauss (1777-1855).

Los números siguientes:

$$\begin{aligned}
 1 &= 1 \\
 1 + 2 &= 3 \\
 1 + 2 + 3 &= 6 \\
 1 + 2 + 3 + 4 &= 10 \\
 1 + 2 + 3 + 4 + 5 &= 15 \\
 1 + 2 + 3 + 4 + 5 + 6 &= 21
 \end{aligned}$$

se llaman **números triangulares**. También el 91 es triangular:

$$1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 + 11 + 12 + 13 + 14 = 91$$

Pero una pirámide con los nichos acomodados de esta manera, tendría 14 niveles y sería muy empinada pues se vería así:

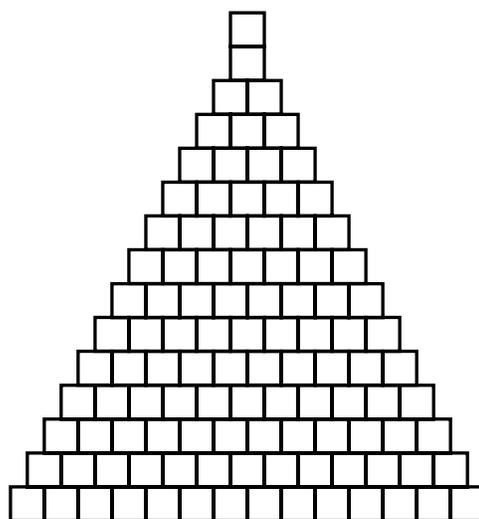


Figura 1.12 Pirámide de 91 nichos en 14 niveles.

No olvidemos que el último nicho se añadió en la cima —el nicho que se quitó de 365 para obtener $364 = 4 \cdot 91$. Para encontrar una solución mejor, quizá conviene no tratar de trabajar con ejemplos específicos sino con todos los ejemplos a la vez.

Las sucesiones aritméticas no siempre empiezan con el número 1. Si se usa una variable, se puede decir que la sucesión empieza con a . Luego, en vez de aumentar por 1, se puede decir que aumenta siempre en b , otra variable. Y se puede usar una tercera variable para denotar el número de niveles: n . La sucesión, entonces, empieza así:

$$a, a + b, a + b + b$$

donde el tercer número de nuestra sucesión es $a + 2b$, el cuarto es $a + 3b$, el quinto es $a + 4b$, y el enésimo es $a + (n - 1)b$. La sucesión quedaría así:

$$a, a + b, a + 2b, a + 3b, a + 4b, \dots, a + (n - 2)b, a + (n - 1)b.$$

Con el argumento del joven Gauss, se puede calcular la suma de estos números y se obtiene la siguiente fórmula:

$$\frac{(2a + (n - 1)b) \cdot n}{2}$$

El problema consiste en encontrar números a , b y n tales que $\frac{(2a + (n - 1)b) \cdot n}{2} = 91$. Si se multiplican ambos lados de esta ecuación por 2 se obtiene que:

$$(2a + (n - 1)b) \cdot n = 182$$

El lado izquierdo de esta ecuación es un producto y uno de sus factores es n . Por lo tanto el lado derecho debe ser divisible entre n . Como $182 = 2 \cdot 91 = 2 \cdot 7 \cdot 13$, se pueden probar diferentes opciones para el valor de n . Si se busca una pirámide con no demasiados niveles hay solamente una opción para asignar un valor a n y esa es $n = 7$. Entonces $n - 1 = 6$ y se puede sustituir en la ecuación original, que queda así:

$$(2a + 6b) \cdot 7 = 26 \cdot 7$$

Entonces $2a + 6b$ debe ser igual a 26. Nuevamente hay varias posibles soluciones —que son $a = 1, b = 4$; $a = 4, b = 3$; $a = 7, b = -2$ y $a = 10, b = 1$ —, pero la que más se acerca a nuestro problema es $a = 4$ y $b = 3$. La sucesión que buscamos es:

$$4, 7, 10, 13, 16, 19, 22$$

y la pirámide tiene entonces la siguiente forma:

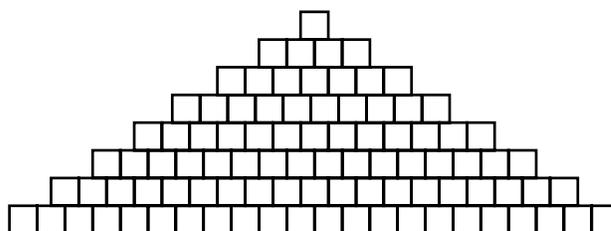


Figura 1.13 Pirámide de 91 nichos en 8 niveles.

Es interesante comparar esta solución con la idea original de los totonacas antes de que hubieran añadido la rampa. La siguiente imagen muestra el lado opuesto al de la rampa que además, es el lado mejor conservado.



<Cuidado: esta imagen tiene copyright de John Barreiro. Debe buscarse alguna similar>

Figura 1.14 Lado oeste de la Pirámide de los Nichos.

Se observa sólo una diferencia con nuestra sucesión: el nivel superior tiene un nicho más. No sabemos cómo llegaron los totonacas a la solución, tal vez simplemente por ensayo y error, al intentar una y otra vez. El análisis que se llevó a cabo aquí requiere de ciertos conceptos y estrategias: necesita generalizar, atacar todas las opciones a la vez al usar **variables**. Esto es un proceso común en matemáticas. George Pólya, un matemático húngaro, decía que a veces es más fácil resolver muchos problemas en conjunto que uno solo a la vez porque al ver la generalidad se puede entender mejor. De esta manera es claro que la matemática ayuda a resolver problemas de nuestra cultura, algunos de importancia crucial, como en este caso.

1.2.3 ¿Cómo se puede medir la profundidad de un pozo?

Un modelo matemático son una o varias fórmulas o ecuaciones, escritas en lenguaje matemático, que describen un proceso o un sistema de procesos. Los modelos matemáticos se usan para predecir o simular o, a veces, sólo para describir lo que pasa en la realidad. Hoy en día, todas las ciencias, incluso las sociales, modelan sus problemas con matemáticas. Por ejemplo, por medio de un modelo matemático se puede tratar de predecir cómo va a mutar un virus, cómo va a cambiar el clima en una región, se puede determinar la altura de un puente e incluso describir algún fenómeno económico y social.

Aquí presentamos un problema muy sencillo pero a la vez muy interesante: cómo medir la profundidad de un pozo. En realidad la idea fundamental no es que el lector entusiasmado vaya midiendo pozos por todos lados, sino que vea cómo, en principio, se hace un modelo matemático.

Para medir la profundidad de un pozo se pueden usar distintas estrategias. Una forma podría ser echar al pozo una cuerda con una piedra amarrada a un extremo. Después de oír que la piedra toca el agua, habría que sacar la cuerda y medirla. No parece ser un método muy práctico porque se necesita una cuerda muy larga. Además, el pozo puede ser profundo y oscuro, de manera que no es posible saber cuándo la cuerda llega al fondo.

Otra posibilidad es usar un cronómetro. Se puede medir el tiempo que transcurre entre que se deja caer la piedra y se escucha el sonido de la piedra al chocar en el agua. ¿Cómo se

relaciona el tiempo que tarda la piedra en llegar al agua con la profundidad del pozo? En otras palabras, ¿cómo relacionar tiempo con distancia? Es claro que cuanto más tiempo transcurre, más profundo es el pozo, pero ¿cómo se puede obtener una medida de profundidad a partir de una medida de tiempo?



Figura 1.15 El pie depende de la foto que se inserte.

Cuando se suelta una piedra en un pozo, como sólo actúa sobre ella la fuerza de gravedad, decimos que cae en caída libre. La fórmula de caída libre es:

$$d = \frac{1}{2}gt^2$$



d es la distancia vertical, en este caso la profundidad del pozo
 $g = 9.81 \frac{m}{s^2}$ es la constante gravitacional, esto es, la aceleración con la que cae la piedra
 t es el tiempo; que en este caso es el que transcurre entre que se suelta la piedra y ésta golpea el fondo del pozo

Por ejemplo, en una caída de 3 segundos (3 s), la fórmula nos daría una profundidad de:

$$d = \frac{1}{2}g(3 \text{ s})^2 = \frac{1}{2} \left(9.81 \frac{m}{s^2} \right) (9 \text{ s}^2) = 44.145m$$

Es decir, un poco más de 44 metros. Éste es un primer modelo, el más sencillo, pero no el más exacto.

Como el pozo es oscuro, no se puede ver cuándo la piedra llega al fondo, solamente se escucha. Tomemos en cuenta, además, que el sonido tarda un cierto tiempo en llegar a nosotros; hay que buscar una manera de medir el tiempo que pasa desde que la piedra choca en el agua y el sonido sube por el pozo y llega a nosotros.

Como la velocidad del sonido (v) es: $v = 340 \frac{\text{m}}{\text{s}}$, el tiempo que el sonido tarda en subir por el pozo puede calcularse con la fórmula:

$$v = \frac{d}{t}$$

de la que puede despejarse la variable t :

$$t = \frac{d}{v}$$

v es la velocidad; en este caso la del sonido: $340 \frac{\text{m}}{\text{s}}$.

d es la distancia; en este caso es la que el sonido recorre hasta llegar a nosotros, es decir, la profundidad del pozo.

t es el tiempo; en este caso, en que el sonido tarda en subir por el pozo hasta llegar al oído del observador que escucha.



¡Cuidado! Aquí se ha vuelto a usar la variable t para denotar al tiempo, aunque ahora se trata de un tiempo diferente. En la caída, la variable t representaba el tiempo que tarda la piedra en caer y ahora representa el tiempo que tarda el sonido en subir desde el fondo hasta el escucha. Esto es algo que ocurre con frecuencia en matemáticas: el significado de una variable depende del contexto en el cual se usa. Para evitar confusiones, distinguiremos estos dos tiempos con un subíndice: t_{caida} para el tiempo de caída y t_{sonido} para el del sonido.

Si el pozo realmente tuviera 44.1 metros de profundidad el sonido tardaría 0.13 segundos en recorrerlo. No es mucho, pero un segundo modelo del problema podría ser restarle este tiempo a los tres segundos que tardamos en escuchar la caída, lo que dejaría $3 - 0.13 = 2.87$ segundos para la caída libre, lo cual resultaría en:

$$d = \frac{1}{2}g(2.87 \text{ s})^2 = \frac{1}{2}\left(9.81 \frac{\text{m}}{\text{s}^2}\right)(8.2369 \text{ s}^2) = 40.4019945\text{m}$$

Esta segunda aproximación es más exacta, pero no del todo correcta, pues tomamos el tiempo del sonido para una profundidad que no es la que obtenemos. El problema radica en que no se conoce ni el tiempo de la caída ni el tiempo que tarda en subir el sonido por separado, lo único que se conoce es que el tiempo (t) desde que se suelta la piedra hasta que se escucha el golpe es la suma de ambos:

$$t = t_{caida} + t_{sonido} \quad (1)$$

Las otras fórmulas importantes, que relacionan la distancia con los tiempos, son:

$$d = \frac{1}{2}gt_{caida}^2 \quad (2)$$

$$\frac{d}{v} = t_{sonido} \quad (3)$$

Lo que se busca es una fórmula que relacione la distancia d directamente con el tiempo total t . Para obtenerla se requiere cierta comodidad con el manejo del álgebra. El lector que no esté familiarizado con ella puede simplemente observar esto como un ejemplo particular. De la ecuación (1) se obtiene que:

$$t_{sonido} = t - t_{caida}$$

Ahora se puede sustituir t_{sonido} en la ecuación (3) por la expresión $t - t_{\text{caída}}$, lo que nos da una nueva ecuación:

$$t - t_{\text{caída}} = \frac{d}{v}$$

De esta ecuación se puede despejar la variable $t_{\text{caída}}$ y se obtiene $t_{\text{caída}} = t - \frac{d}{v}$. Sustituyendo lo anterior en la ecuación (2):

$$d = \frac{1}{2}g\left(t - \frac{d}{v}\right)^2$$



d es la profundidad del pozo

$g = 9.81 \frac{m}{s^2}$ es la constante gravitatoria

$v = 340 \frac{m}{s}$ es la velocidad del sonido en el aire

t es la suma de los tiempos $t_{\text{caída}}$ y t_{sonido}

Ahora sí, obtuvimos una fórmula general y válida para cualquier pozo. Un tercer y último modelo para nuestro problema original que, admitimos, es una ecuación más complicada que la del primer modelo, donde únicamente se tomó en cuenta el tiempo de caída. Una ecuación compleja es un fenómeno común: entre mejor sea el modelo y más preciso, más complicadas serán las matemáticas involucradas. Así, tenemos que para un tiempo de $t = 3$ segundos se obtiene una profundidad $d = 40.7$ metros.

Resolvamos de poste otro problema relacionado con el anterior. Si en el pozo decidiéramos gritar ¿cuánto tiempo pasaría hasta que se escuchara el eco desde el fondo del pozo?

Partamos de que, en cualquier pozo, el sonido tarda el mismo tiempo en bajar que en subir. Como $t = \frac{d}{v}$, entonces el sonido tarda $2\frac{d}{v}$, donde d es la profundidad del pozo que equivale a $d = 26.16942$ metros y $v = 340 \frac{m}{s}$ es la velocidad del sonido en el aire.

Entonces, $t = \frac{2(26.16942)}{340} = 0.1539$ segundos. ¡Y ya acabamos!

1.3 LAS MATEMÁTICAS DE LA NATURALEZA

En la naturaleza hay fenómenos muy diversos y complejos. Resulta asombroso que el ser humano haya podido desentrañar el mecanismo de algunos de ellos y además, expresarlos con fórmulas matemáticas muy sencillas. En esta sección veremos ejemplos de fórmulas simples que describen el comportamiento de algunos de los fenómenos más importantes de la naturaleza.

Galileo Galilei, un científico italiano del siglo XVI, descubrió, entre otras cosas, el tipo de trayectoria que sigue una piedra al ser lanzada al aire. Para ello, empleó la matemática situándola en el lugar especial dentro de las ciencias, que ocupa todavía en la actualidad. En sus



Figura 1.16 Grabado de Ernst Haeckel, biólogo, filósofo, físico y artista alemán que retrató y nombró a miles de especies. En este grabado se aprecia la belleza matemática de las conchas marinas.

libros, escritos como diálogos entre tres personajes, describe y desarrolla sus ideas sobre cómo puede entenderse la naturaleza. Discute fenómenos como la balanza, la caída libre de los cuerpos y el péndulo. La siguiente cita aparece en su obra maestra *Il saggiaiore* y es de las más importantes y famosas en la historia de la ciencia pues dice que:

Sr. Sarsi, las cosas no son así. La filosofía está escrita en ese grandísimo libro que tenemos abierto ante los ojos, quiero decir, el universo, pero no se puede entender si antes no se aprende a entender la lengua, a conocer los caracteres en los que está escrito. Está escrito en lengua matemática y sus caracteres son triángulos, círculos y otras figuras geométricas, sin las cuales es imposible entender ni una palabra; sin ellos es como girar vanamente en un oscuro laberinto.¹

Hay que aclarar que, en aquel entonces, *filosofía* abarcaba todos los conocimientos, esto es, representaba lo que hoy llamamos *ciencia*. Si se acepta la visión de Galilei, entonces para comprender el Universo son necesarias las matemáticas. Afortunadamente no es necesario estudiar matemáticas avanzadas para poder entender los asuntos más importantes de la naturaleza. Incluso a veces, las matemáticas requeridas para comprender algunas cosas de la naturaleza y su comportamiento son extremadamente simples, como en el caso de la esfera y sus aplicaciones que se muestra continuación.

1.3.1 La esfera

La esfera es una de las figuras geométricas que podemos contemplar frecuentemente a nuestro alrededor.



Figura 1.17 Algunos objetos esféricos.

Los objetos de esta ilustración no son esféricos por pura casualidad. La pelota de tenis tiene que ser esférica para que rebote de manera completamente predecible. El balón de futbol americano, al no ser redondo, rebota en el piso de forma impredecible. Sería imposible jugar tenis con una pelota con la forma de un balón de futbol americano. La bur-

¹ Del original (1979, p. 38): “La filosofia è scritta in questo grandissimo libro che continuamente ci sta aperto innanzi a gli occhi (io dico l’universo), ma non si può intendere se prima non s’impara a intender la lingua, e conoscer i caratteri, ne’ quali è scritto. Egli è scritto in lingua matematica, e i caratteri son triangoli, cerchi, ed altre figure geometriche, senza i quali mezzi è impossibile a intenderne umanamente parola; senza questi è un aggirarsi vanamente per un oscuro laberinto.”

buja de jabón es esférica como consecuencia de la presión uniforme que ejerce sobre sus paredes el aire que contiene y por la resistencia también uniforme del material de que está hecha.

El ojo es esférico para que podamos mirar en muchas direcciones sin tener que voltear la cabeza. La esfera es el único cuerpo que puede girar libremente en todas direcciones aunque esté apoyado en varias partes. Este hecho se aprovecha también en la cabeza del fémur, el hueso humano de mayor tamaño y también en los balines de los rodamientos o valeros.

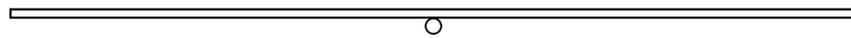
Debido a la función que desempeñan en la naturaleza, al uso para el que fueron diseñados o a las propiedades de los materiales que los forman, muchos objetos tienen forma esférica. De este modo, la esfera es una forma geométrica, y por tanto, matemática, que observamos con frecuencia en el Universo. Es un ente abstracto, un concepto, puesto que muchos objetos de la naturaleza pueden describirse como una esfera. La esfera es uno de esos “caracteres” matemáticos de los que habla Galileo que describen el Universo.

1.3.2 La balanza

Si bien la balanza es una máquina creada por el hombre —que sirve para medir y comparar pesos—, su comportamiento depende de las leyes naturales al aprovechar la gravedad; dicho de otro modo, la balanza nos muestra cómo funciona la gravedad. Por lo anterior, al describir el funcionamiento de la balanza habremos descrito una pequeña parte de la naturaleza.

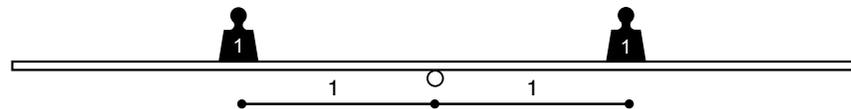
La balanza más sencilla es una tabla equilibrada sobre un rodillo, como se muestra en la figura 1.3.3.

Figura 1.18 Balanza sencilla.



Supongamos que la tabla y el rodillo son muy ligeros, pero firmes. La balanza mantiene su estado de equilibrio si ponemos sobre la tabla dos pesas iguales a la misma distancia en ambos lados del rodillo, por ejemplo, de un kilogramo a un metro de distancia cada una:

Figura 1.19 Balanza equilibrada.



Hay una ley que describe cómo cambia el peso al cambiar la distancia. Sabemos por nuestra experiencia que cuanto más lejos, menor peso se requiere. Esto es lo que nos dice la intuición, pero nos gustaría encontrar una ley precisa que nos diga qué peso, cuántos gramos habría que poner a dos metros del rodillo para equilibrar el peso de un kilogramo que está a un metro del rodillo. El siguiente razonamiento muestra que tiene que ser medio kilo. Podemos imaginarnos el kilogramo dividido en dos pesas iguales puestas sobre una balanza que a su vez reposa en el lugar en donde antes estaba la pesa de un kilo:

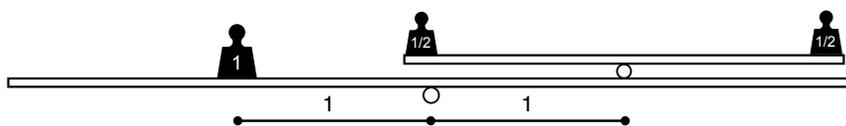


Figura 1.20 La balanza al sustituir el peso derecho por una balanza pequeña.

Como la balanza pequeña está en equilibrio podemos quitar el rodillo y la tabla de arriba —por hipótesis, su peso es despreciable— y apoyar las pesas directamente sobre la tabla grande sin que se altere el equilibrio. Finalmente, la pesa de medio kilo que se apoya justo sobre el rodillo central también puede quitarse.

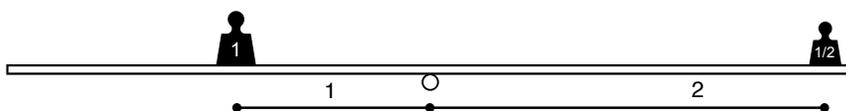


Figura 1.21 Vemos que al quitar la pesa apoyada en el rodillo central, no se influye en el equilibrio de la balanza.

Con ello descubrimos la ley básica de la balanza: al doble de distancia hay que poner la mitad del peso.

La ley general de la balanza se obtiene de este caso al usar consecutivamente el mismo procedimiento de sustituir las pesas por balanzas equilibradas. Por ejemplo, remplacemos ahora la pesa de un kilo del lado izquierdo por una nueva balanza equilibrada con pesas de $\frac{2}{3}$ y $\frac{1}{3}$ de kilo, la primera colocada medio metro a la izquierda y la segunda un metro a la

derecha del lugar que ocupaba la pesa de un kilo, como muestra la siguiente imagen:

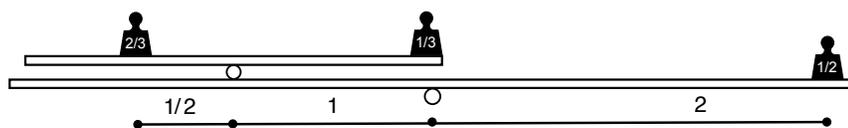


Figura 1.22 La balanza después de sustituir el peso izquierdo por una balanza pequeña.

Al volver a eliminar el rodillo y la tabla que sostiene a estas dos pesas, así como también la que queda sobre el rodillo de la balanza, obtenemos la siguiente balanza en equilibrio:

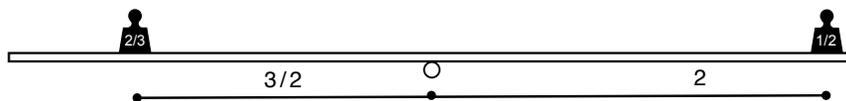


Figura 1.23 La balanza al eliminar el rodillo y la tabla de la balanza pequeña, así como también la pesa del medio.

El lector atento se dará cuenta de que existe una ley detrás de estos números: si la distancia se duplica, el peso se divide a la mitad y si la distancia es $\frac{3}{2}$ de metro entonces el peso es de $\frac{2}{3}$ kilogramo. Podemos escribir esto como una igualdad:

$$\frac{2}{3} \text{ kg} \cdot \frac{3}{2} \text{ m} = \frac{1}{2} \text{ kg} \cdot 2 \text{ m}$$

De esta manera podríamos argumentar para otros pesos y otras distancias. Y siempre encontraríamos que, cuando hay equilibrio, el producto del peso por la distancia al rodillo es igual en ambos lados. La fórmula matemática:

$$p_1 d_1 = p_2 d_2$$

expresa esta ley de manera muy sencilla: nos dice que el producto del primer peso p_1 por su distancia d_1 , al rodillo, es igual al producto del segundo peso p_2 multiplicado por su distancia d_2 hasta el rodillo. En particular, la fórmula se satisface en el caso anterior pues:

$$p_1 = \frac{1}{2} \text{ y } d_1 = 2 \text{ y } p_2 = \frac{2}{3}, d_2 = \frac{3}{2}$$

El lenguaje matemático es a la vez muy condensado y muy expresivo. En una sola fórmula se expresan todas las situaciones de equilibrio de una balanza. La fórmula cubre una infinidad de casos.

El ojo no entrenado no puede entender a primera vista la generalidad y la importancia de esta fórmula, al igual que el ojo de una persona no entrenada en la escritura musical no puede entender una partitura y el que no conoce las letras y sus combinaciones para un sistema lingüístico no puede entender algo escrito. Sin embargo, un poco de entrenamiento matemático permite comprender el significado de los símbolos que aparecen en la fórmula, la relación que expresa entre las cantidades representadas por los símbolos. Como queda claro, unas pocas matemáticas nos llevan muy lejos.

Además de la sencillez de la fórmula obtenida, resulta sorprendente que el razonamiento que nos llevó a descubrirla no requirió de experimentos con objetos reales. Fue suficiente hacerlos mentalmente y argumentar de manera lógica. No siempre es posible descubrir las leyes de la naturaleza por medio de experimentos mentales, sin embargo, estos últimos se utilizan muchas veces para acercarnos a ellas.

Mientras el caso de la esfera nos muestra un acercamiento directo a la naturaleza, el de la balanza nos revela una ley más profunda que se expresa mediante una fórmula matemática que llegamos a descubrir mediante razonamientos lógicos.

A continuación, presentaremos el caso de la gravedad que nos lleva a un tipo de fenómenos aún más complejos, en los que no basta la razón para descubrir las leyes naturales sino que hacen falta, además, experimentos que den información numérica que no podríamos obtener por pura lógica, pero en los que, las leyes que los rigen, siguen pudiéndose expresar con fórmulas matemáticas sencillas.

1.3.3 La gravedad

En la sección anterior se llegó a una fórmula matemática para expresar una ley de la naturaleza: la de la balanza. Con ello queda ilustrado que las matemáticas pueden ser útiles para describir fenómenos naturales sencillos. Existen fenómenos mucho más complejos en los cuales las matemáticas no sólo son útiles sino verdaderamente imprescindibles. Uno de ellos es el de la gravedad. Comenzaremos por su aspecto más simple, la caída de los cuerpos.

Sabemos por nuestra experiencia diaria que los cuerpos caen. Si soltamos una taza ésta va a dar al suelo y probablemente se rompa. Desde épocas muy remotas se sabe lo anterior, pero no fue sino hasta 1590 cuando Galileo Galilei describió en términos matemáticos precisos el fenómeno de la **caída libre** —cuyo nombre expresa que nada se interpone en ella— de los cuerpos. Galileo demostró mediante la observación, la experimentación y el razonamiento que, cuando cae un cuerpo, su velocidad aumenta constantemente. Así, la velocidad del cuerpo que cae es proporcional al tiempo que ha transcurrido desde que comenzó a caer. La constante de proporcionalidad se llama la aceleración de la gravedad, se

denota por g , y hoy sabemos que su valor es de $9.81 \frac{m}{seg^2}$. Esto quiere decir que la velocidad de un cuerpo en caída libre aumente en $9.81 \frac{m}{seg}$ cada segundo. En fórmula, lo anterior se escribe así:

$$v = gt$$

donde v es la velocidad del cuerpo, t es el tiempo transcurrido desde que comenzó su caída y g es la constante de la aceleración de la gravedad.

Galileo demostró que, como consecuencia de esta ley de movimiento según la cual la velocidad aumenta constantemente con el tiempo, la distancia recorrida por el cuerpo en su caída aumenta de manera cuadrática con respecto al tiempo, es decir:

$$d = \frac{1}{2}gt^2$$

donde d es la distancia que un cuerpo ha recorrido en su caída, t es el tiempo que ha transcurrido desde que comenzó a caer y g es la constante de la aceleración de la gravedad.

El siguiente esquema muestra una pelota que cae de un edificio de 80 metros y una gráfica con las posiciones de la pelota cada dos décimas de segundo.

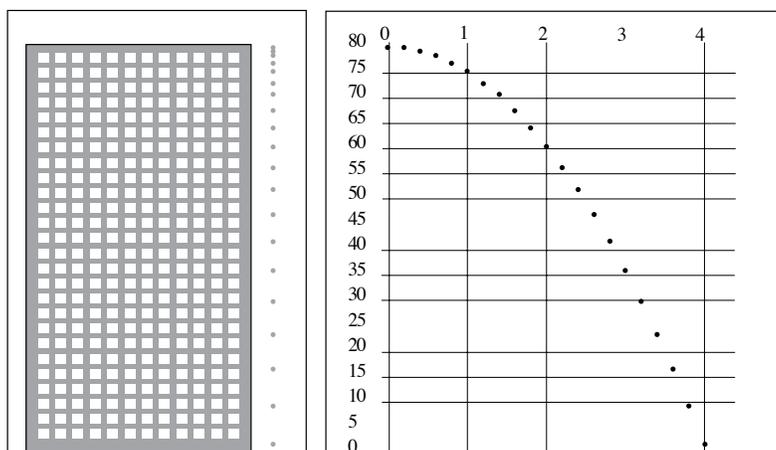


Figura 1.24 A la izquierda se presentan las distintas posiciones de la pelota cada cierta fracción de tiempo (1/4 de segundo). A la derecha, se observa la gráfica de la altura contra el tiempo.

La forma que une los puntos de la gráfica es una curva conocida ya por los antiguos griegos a la que llamaron **parábola**. En la gráfica se observa que la pelota cae aproximadamente 5 metros durante el primer segundo, el triple de 5 durante el siguiente segundo, cinco veces esa misma cantidad durante el tercer segundo y siete veces la cantidad durante el cuarto segundo. Es decir, la distancia recorrida sigue la sucesión 1, 3, 5, 7... de los números impares. Por tanto, la distancia total que ha caído en cada segundo es la suma de los primeros impares: 1, 1+3, 1+3+5, 1+3+5+7... que son los cuadrados de los números enteros 1, 4, 9, 16, etc. Esta observación podemos escribirla como:

$$d = 4.9 \cdot t^2$$

que corresponde a la fórmula de Galileo $d = \frac{1}{2}gt^2$ ya que $\frac{1}{2}g \cong 4.9$.

Por ejemplo, la distancia que recorre un cuerpo en caída libre en el doble de un lapso de tiempo es cuatro veces la que recorrió en el lapso original. La distancia recorrida en el triple de un lapso de tiempo es igual a nueve veces la distancia original. Esta fórmula representa la ley del movimiento de caída libre. No es una ley lineal sino cuadrática: la velocidad depende linealmente del tiempo, pero la distancia recorrida es proporcional al cuadrado del tiempo.

No es una ley complicada, pero podríamos considerar que es ligeramente más compleja que la que describe a la balanza pues en aquella no aparecen cantidades elevadas al cuadrado y en ésta sí.

Galileo demostró también que la trayectoria descrita por un objeto lanzado al aire es una *parábola*, esa curva que habían estudiado dos mil años antes los griegos por razones enteramente diferentes; de hecho, la usaron para intentar resolver el problema —meramente matemático— de la duplicación del cubo, que consistía en encontrar el lado de un cubo que tuviera el doble del volumen de otro cubo dado.

Si regresamos a la parábola como trayectoria de una piedra o bala lanzada, lo sorprendente es que Galilei obtuvo esta descripción no de la observación directa sino mediante una argumentación lógica. Para ello tuvo que introducir la relatividad del movimiento. Su ejemplo es famoso: si dejamos caer una piedra desde una torre, ésta cae por una trayectoria vertical al piso. Pero si dejamos caer la piedra desde un mástil de un barco que navega en una dirección, ¿dónde caerá la piedra?, ¿junto al mástil o detrás de éste porque cae de forma vertical y mientras tanto el barco avanza? El hecho es que cae junto al mástil. La explicación de Galileo es que el marinero que deja caer la piedra cree que cae vertical porque él también se mueve, pero para un espectador en la playa la piedra no sólo cae sino que al soltarla, se impulsa en la dirección del barco. De esta manera, Galileo descubre que las velocidades pueden descomponerse en una componente vertical y otra horizontal.

El movimiento de caída libre no parecía tener nada que ver con el movimiento de los planetas, hasta que uno de los más grandes genios de la historia de la ciencia, Isaac Newton, pensó en relacionarlos. El resultado de esas investigaciones constituye uno de los mayores logros científicos, quizá la hazaña intelectual más importante de todos los tiempos. Lo más sorprendente es que está al alcance de cualquier estudiante de bachillerato dispuesto a familiarizarse con la manipulación de algunas expresiones algebraicas no muy complicadas. Para acercarnos al tema es necesario que repasemos los conocimientos que la humanidad tenía del movimiento planetario en la época de Newton.

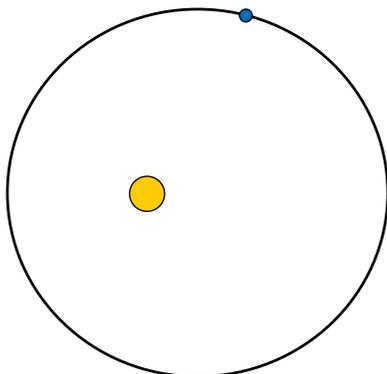


Figura 1.25 Ilustración de los planetas en órbita alrededor del Sol.

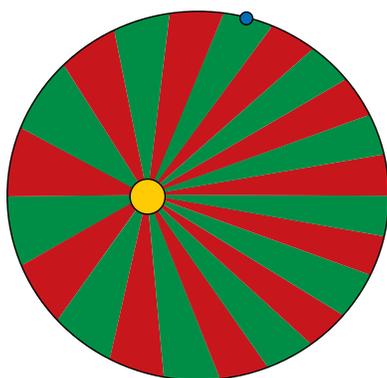
Johannes Kepler fue un contemporáneo de Galileo que dedicó la mayor parte de su vida al estudio de las órbitas de los planetas, es decir, a las trayectorias que describen los planetas alrededor del Sol. Kepler era un hombre profundamente religioso y convencido de que Dios dispuso a los planetas en perfecta armonía describiendo figuras geométricas sencillas. Gracias a las mediciones precisas de Tycho Brahe, después de muchos intentos fallidos con circunferencias y óvalos, le fue posible describir las órbitas celestes mediante la elipse, curva que al igual que la parábola había sido estudiada por los griegos por sus propiedades puramente geométricas. Kepler formuló sus hallazgos principalmente en dos de sus libros: *Astronomía nova* (Nueva astronomía, 1609) y *Harmonices Mundi* (La armonía de los mundos, 1619) en los que describe el movimiento de los planetas en tres leyes, llamadas las leyes de Kepler del movimiento planetario.

Leyes de Kepler del movimiento planetario

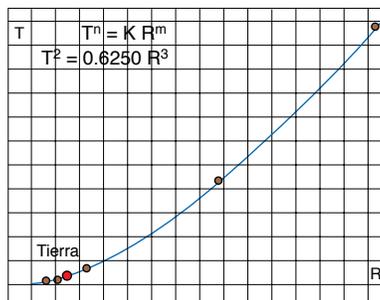
1. Los planetas giran alrededor del Sol en órbitas elípticas, donde el Sol es uno de los focos.



2. Los radios-vectores de los planetas respecto al Sol barren áreas iguales en tiempos iguales.



3. Los cuadrados de los periodos de revolución de los planetas son proporcionales a los cubos de los radios medios de sus órbitas.



Conviene hacer algunas aclaraciones sobre estas leyes. La primera, aunque afirma que las órbitas de los planetas son elípticas, en realidad se trata de elipses con una excentricidad muy pequeña, lo que quiere decir que son casi circulares. La segunda nos indica que los planetas se mueven con mayor velocidad cuando están más cerca del Sol que cuando están más lejos. Finalmente, la tercera, que puede resultar muy extraña, no es más que el descubrimiento de que los radios de las órbitas y sus tiempos de revolución no son arbitrarios sino que guardan una relación funcional que no es lineal ni cuadrática, más bien algo intermedio: el cuadrado del tiempo es proporcional al cubo del radio:

$$\frac{T^2}{R^3} = K$$

donde K es un número fijo para todo el sistema planetario, el mismo para todos los planetas.

Es importante observar que la tercera ley le llevó a Kepler doce años más que las otras para descubrirla y representa una relación muy precisa entre la velocidad con la que giran los planetas y sus distancias al Sol. ¿Cómo es que se da esta relación aparentemente tan extraña? No hay ninguna relación aparente entre los radios de las órbitas y, por ejemplo, los tamaños de los planetas. ¿Qué hace que sí la haya entre los periodos y los radios? ¿Y por qué es a la vez tan simple y tan extraña?

Kepler propuso un modelo matemático del movimiento de los planetas pero no una explicación. Creía que estas leyes provenían de la mano de Dios y no requerían mayor comprensión.

Así quedaron las cosas por varios años hasta que, en 1687, se publicó el libro *Philosophiæ Naturalis Principia Mathematica* de Isaac Newton, matemático, físico y astrónomo inglés. En este libro, Newton explica que se pueden deducir las tres leyes de Kepler a partir de leyes más fundamentales y simples.

La idea principal de Newton es que debe haber una fuerza que mantiene a los planetas en órbitas alrededor del Sol y esa misma fuerza debe ser la responsable del movimiento de caída libre de los cuerpos sobre la Tierra.

Antes de llegar a esta idea Newton, al continuar el trabajo de Galileo, había establecido sus tres famosas leyes:

Leyes del movimiento de Newton

- 1] *Todo cuerpo permanece en estado de reposo o de movimiento rectilíneo uniforme a menos que haya una fuerza externa que lo modifique.*
- 2] *Cuando una fuerza actúa sobre un cuerpo, éste adquiere una aceleración constante proporcional a la magnitud de la fuerza e inversamente proporcional a la masa del cuerpo.*
- 3] *A toda acción corresponde una reacción igual y en sentido contrario.*

En símbolos, la segunda ley de Newton se escribe así:

$$a = \frac{F}{m}$$

donde a es la aceleración que recibe el cuerpo, m es su masa y F es la fuerza que actúa sobre él. Para Newton era claro que si los cuerpos caen es debido a que hay una fuerza que los atrae hacia el suelo y esta fuerza debe ser proporcional a la masa del cuerpo ya que todos los cuerpos, independientemente de su masa, caen con la misma aceleración g . La Tierra atrae a los cuerpos sobre su superficie con una fuerza $F = mg$, donde m es la masa del cuerpo, aunque en forma coloquial decimos “el peso” del cuerpo. Con este razonamiento, Newton define el concepto de la masa de un cuerpo y lo relaciona con su peso por medio de la constante g ; a la vez, diferencia ambos conceptos al definir a la masa como una propiedad intrínseca del cuerpo, mientras que considera al peso como una propiedad del cuerpo en relación con la Tierra, es decir, representa la fuerza con la que la Tierra lo atrae.

Al razonar con ayuda de la simetría y apoyado en su tercera ley del movimiento, Newton piensa que si la Tierra atrae a los cuerpos que se encuentran cerca de su superficie con una fuerza proporcional a la masa de los mismos, entonces debe haber otra fuerza, igual y en sentido contrario, con la que cada cuerpo atrae a la Tierra y esta fuerza, lógicamente, también debe ser proporcional a la masa de la Tierra. Dicha fuerza debe ser la responsable de la

caída de los cuerpos hacia la Tierra y también de las trayectorias de los planetas que, en lugar de ser líneas rectas, giran alrededor del Sol porque éste los atrae. Ésta es la gran idea de Newton: hay una fuerza de atracción entre todos los cuerpos, que aumenta proporcionalmente con la masa de cada cuerpo.

A partir de las ideas anteriores y por deducción matemática, Newton logró establecer que la fuerza de atracción entre dos cuerpos de masas m y M que se encuentran a una distancia R debe tener la forma:

$$F = G \frac{Mm}{R^2}$$

donde G es una constante universal, es decir, es la misma para todos los cuerpos. Ésta es la famosa Ley de la gravitación universal de Newton. Dice que la fuerza de gravedad con la que se atraen dos cuerpos de masas M y m es proporcional al producto de sus masas e inversamente proporcional al cuadrado de la distancia R que los separa. G se llama la **constante de la gravitación universal** y puede calcularse a partir de la aceleración de la gravedad g y el radio r y la masa m de la Tierra. Esta fórmula expresa una ley de la naturaleza de manera compacta y precisa. Las matemáticas son enormemente económicas: con sólo ocho símbolos —cinco letras, un número, una igualdad y una raya de quebrados— se expresa todo este profundo conocimiento. Al final de esta sección hay una deducción de esta fórmula a partir de la tercera ley de Kepler.

Newton logró demostrar que las tres leyes del movimiento planetario de Kepler son meras consecuencias matemáticas de la Ley de la gravitación universal y de sus tres leyes básicas del movimiento de los cuerpos. Este hecho constituye una explicación científica completa del movimiento planetario. Una ley universal que explica los movimientos de todos los cuerpos terrestres y celestes de manera unificada y clara. Probablemente no hay otro descubrimiento científico de mayor trascendencia y belleza que éste. El lector interesado en estas deducciones podrá seguirlas en el Tema 3, en la sección sobre “Espacio, tiempo y movimiento”.

Para llegar al descubrimiento de la Ley de la gravitación universal, Galileo, Kepler, Newton y muchos científicos más tuvieron que luchar contra las ideas preconcebidas que imperaban en aquella época, algunas de carácter religioso y otras apoyadas en antiguas tradiciones, pero todas haciendo una apología irracional de la ignorancia y la intolerancia. A nosotros, seres humanos del siglo XXI, nos parece natural que una misma ley de la naturaleza explique fenómenos terrestres y celestes pero debemos recordar que en aquel entonces se pensaba que lo que pasaba en la Tierra era algo de naturaleza muy distinta a lo que pasaba en el cielo, el cual se identificaba, no con lo material sino con lo divino. Según Aristóteles, el cielo y sus ingredientes como el Sol, la Luna, las estrellas y los planetas, estaba compuesto por una materia completamente distinta a las terrenales. Las ideas de Galileo, Kepler y Newton fueron por tanto muy revolucionarias: la atracción de la gravedad actúa de la misma manera en el cielo y en la Tierra. La misma ley explica por un lado, el movimiento de los planetas alrededor del Sol, el de la Luna alrededor de la Tierra y la caída de los cuerpos sobre la superficie terrestre y predice que, en cualquier astro celeste, debe haber una fuerza que hace caer a los cuerpos que se encuentran sobre su superficie. La atracción gravitatoria actúa en la Tierra y en el cielo, en el Sistema solar y entre estrellas y galaxias, siempre de la misma manera.

Parece haber una relación muy íntima entre la naturaleza y las matemáticas, tal como Galileo Galilei lo hiciera notar en la nota que aparece al principio de esta sección y que escribió muchos años antes del descubrimiento de Newton. Quizá esa idea suya inspiró a Newton e inspira a los científicos, cuya misión es estudiar a la naturaleza al tratar de descu-

brir sus leyes y expresarlas, de preferencia, en lenguaje matemático, porque cuando esto se logra se obtiene un conocimiento que es a la vez profundo y sencillo.

Ésta es una de las grandes fuerzas inspiradoras y generadoras de las matemáticas.

Como la ley de la gravitación universal es uno de los grandes logros culturales, en este primer capítulo quisimos explicar la deducción de la fórmula —a pesar de que el álgebra requerida para ello rebasa el nivel de las matemáticas usadas hasta ahora— que se encuentra en el recuadro a continuación. Ya se mencionó que en los libros de matemáticas hay partes más difíciles de leer; ésta es una de ellas y por consiguiente, se puede omitir, si el lector siente que no entiende, para reanudar la lectura a partir de la siguiente sección.



¿De dónde sale la fórmula $F = G \frac{Mm}{R^2}$? A continuación se presenta una deducción de la Ley de la gravitación universal a partir de la tercera ley de Kepler. Esta deducción requiere de algunas manipulaciones algebraicas a las que tal vez el lector no esté acostumbrado.

Del estudio del movimiento circular uniforme se sabe que para que un cuerpo de masa m se mueva con velocidad constante v en una trayectoria circular de radio R es necesario que actúe sobre él una fuerza —llamada fuerza centrípeta— dirigida hacia el centro de la trayectoria y de magnitud:

$$F = m \frac{v^2}{R} \quad (1)$$

La demostración de este hecho podrá consultarse en el tema 3, como ya se mencionó. La velocidad de un planeta cuya órbita es aproximadamente circular de radio R y cuyo periodo de revolución es T , debe ser igual al perímetro de la órbita $2\pi R$ entre T , es decir:

$$v = \frac{2\pi R}{T}$$

por lo tanto, sustituyendo esta expresión para la velocidad en la fórmula (1) obtenemos que la fuerza con la que el Sol atrae al planeta debe ser:

$$F = m \frac{\left(\frac{2\pi R}{T}\right)^2}{R} = m \frac{4\pi^2 R}{T^2}$$

Esta última fórmula se obtuvo al sustituir v en la primera, por su expresión $\frac{2\pi R}{T}$, y tras realizar algunas operaciones algebraicas como elevar al cuadrado un producto y simplificar fracciones.¹ En la última igualdad utilizamos la tercera ley de Kepler para sustituir T^2 por su equivalente KR^3 . Así, obtenemos:

$$F = m \frac{4\pi^2 R}{KR^3} = m \frac{4\pi^2}{KR^2}$$

lo cual nos dice que la fuerza de gravedad debe depender inversamente del cuadrado de

¹ La sustitución es un método común al trabajar con ecuaciones: éstas expresan igualdades y si una parte es igual a otra, la podemos reemplazar sin alterar la ecuación. Las operaciones algebraicas son sólo cambios de forma que no alteran tampoco la igualdad. El lector que no esté aún familiarizado con estos procedimientos tendrá algunas dificultades para entender todos los detalles en una primera lectura, pero poco a poco se familiarizará con ellos hasta llegar a comprenderlos perfectamente, pues en realidad son simples manipulaciones numéricas con números representados por letras.

la distancia R entre el planeta y el Sol. Si definimos $G = \frac{4\pi^2}{KM}$ donde M es la masa del Sol, podemos escribir esta igualdad como:

$$F = G \frac{Mm}{R^2}$$

Que es precisamente la fórmula a la que queríamos llegar.

1.4 LAS PROPIAS MATEMÁTICAS

Las matemáticas son una fuente inagotable de preguntas y problemas pues se construyen “a base de” entender, que en cierto sentido, es disipar dudas. Saber dudar, hacerse buenas preguntas, es parte integral del que-hacer matemático.

Aunque su materia de trabajo sean ideas abstractas, las matemáticas tienen una enorme solidez basada en su posible comunicación, de una generación a otra, sin alteraciones. Lo que nuestros antepasados entendieron por “dos”, independientemente del vocablo que hayan usado para referirse a esta idea, es lo mismo que entendemos ahora. Y cuando dieron nombre al “cuatro”, la idea que se usa hoy como estereotipo de lo obvio y fácil “dos más dos son cuatro”, les ayudó a entender y formular el concepto “más”. Los niños de cada generación reviven este proceso de entender significados y acceder a nuevos niveles de comprensión que compartimos todos. Esa propiedad de las matemáticas de reconstruirse en la mente de cada ser humano se debe a que se arman con una lógica implacable. No son arbitrarias sino naturales y están basadas en el razonamiento. Lo que aquí llamamos “entender”, ese momento de iluminación, o respiro, en el que todas las piezas caen en su lugar, ese “¡ah, sí!” que hemos experimentado todos, es lo que solidifica y nos permite compartir, más allá de tiempos y culturas, ese mundo etéreo de las matemáticas.

Entender, en el sentido de razonamiento y no en la autoritaria de “¡entiende, niño!”, está muy relacionada con dudar o cuestionar. Cuando entendemos algo es porque resolvimos, aunque no necesariamente de manera explícita, alguna duda. Por consiguiente, es natural que al solucionar un problema o acceder a un nuevo nivel de entendimiento, surjan nuevas dudas o preguntas. Trabajar en problemas que surgen de las propias matemáticas, es decir, pensar en ellos, tratar de resolverlos o clarificar las interrogantes asociadas que surgen a su alrededor, ha sido uno de los grandes generadores del conocimiento matemático. Hoy en día, este proceso ocurre cotidianamente en la investigación y es, probablemente, el motor más prolífico de desarrollo que tienen las matemáticas. Veamos un ejemplo de lo anterior, al que puede asociarse el nacimiento de las matemáticas como ciencia.

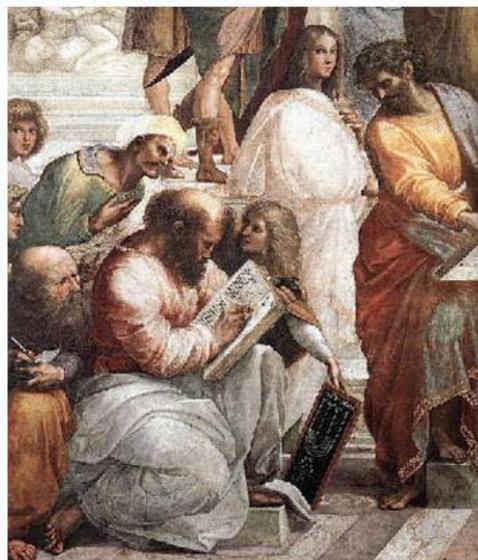
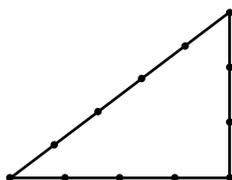


Figura 1.26 Detalle de La escuela de Atenas, del pintor italiano Rafael, que muestra a Pitágoras, fundador de la primera escuela de matemáticas que se enfocaba a los números y sus aplicaciones en el arte y la música.

1.4.1 Pitágoras, Fermat y Wiles

En Babilonia y Egipto se desarrolló el conocimiento para construir edificaciones y trazar ciudades, el cual perdura hasta nuestros días. Uno de los puntos básicos de esta tecnología es poder trazar ángulos rectos, pues da lugar a formas racionales de dividir los espacios, y para lograrlo se usaba el siguiente método. Al trazar un triángulo cuyos lados miden 3, 4 y 5 unidades de longitud —no importa si son metros, pies, o una vara cualquiera—, se forma un ángulo recto entre los lados que miden 3 y 4. Este mismo método lo usan en la actualidad nuestros albañiles para encontrar la “escuadra” con más precisión de lo que da una pequeña regleta: toman un “reventón” —así le llaman a los hilos con que se guían para la construcción— de 12 metros ($3 + 4 + 5 = 12$), lo tensan en los puntos correspondientes y ahí, entre los lados de 3 y 4 unidades, obtienen su esquina perfecta.



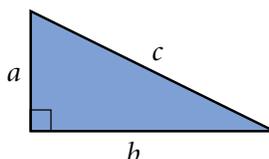
Esta manera de trazar triángulos rectángulos se manejaba como conocimiento empírico, es decir, práctico y corroborado por la experiencia, hasta que los griegos exhibieron lo que hacía tan especial a la terna de números 3, 4 y 5. Para esto, deben haberse preguntado algo como: ¿habrá otras ternas de números con la propiedad de formar ángulos rectos?, ¿qué las distingue?

La respuesta a la pregunta sobre qué ternas forman ángulos rectos es sorprendente: tiene que ver con áreas. Si nos fijamos en los tres cuadrados cuyos lados son los números dados, la suma de las áreas de los dos cuadrados chicos tiene que ser igual al área del cuadrado grande. Cualquier albañil babilonio estará de acuerdo en que con 25 losetas cuadradas se pueden cubrir un cuadrado de 5 por 5, o bien uno de 3 por 3 y uno de 4 por 4. Con la notación de nuestros días, podríamos escribir lo anterior como:

$$3^2 + 4^2 = 9 + 16 = 25 = 5^2$$

Se sabe que los babilonios ya habían observado esta relación numérica y, además, de que conocían otros ejemplos. Pero que esto no fuera una casualidad sino la razón profunda de que el método para trazar ángulos rectos funcionara, es un salto a otro nivel de entendimiento. Una pregunta abstracta, intrínsecamente matemática, sobre la relación entre una terna de números (3, 4 y 5) y un hecho geométrico, el que con ellos se construye un ángulo recto, da lugar al célebre Teorema de Pitágoras, uno de los primeros en su género y un logro de la humanidad equiparable al descubrimiento del fuego o de la rueda. El descubrimiento de la primera “verdad” general y no trivial, como que dos más dos son cuatro, que no depende de la fe o la experiencia, sino que es producto del razonamiento puro. Una verdad que tiene razón de ser más allá de la cultura o idiosincrasia desde la cual se le contemple. Una verdad que, además, rige a la realidad que nos circunda y sugiere que dicha realidad puede comprenderse por medio de la razón.

Figura 1.27 En un triángulo rectángulo se les llama **catetos** a los dos lados que forman el ángulo recto e **hipotenusa** al tercer lado.



El Teorema de Pitágoras dice que las ternas de números a , b , c , ordenadas de manera creciente y que corresponden a las longitudes de los lados de triángulos rectángulos son las que cumplen la relación:

$$a^2 + b^2 = c^2 \quad (1)$$

Pitágoras lo enunció en términos de áreas y para nuestros propósitos no es relevante cuál fue su motivación o argumentación original, lo importante es que dio el gran salto a un enunciado general y demostrable. Una prueba del teorema se obtiene al colocar, de dos maneras diferentes, cuatro triángulos rectángulos con catetos a y b en un cuadrado de lado $a + b$, como en la Figura 1.4.3. Lo que dejan de cubrir estos cuatro triángulos en una de ellas son dos cuadrados de lados a y b , con área total $a^2 + b^2$; y, en la otra, lo que no cubren es un cuadrado cuyo lado es la hipotenusa c .

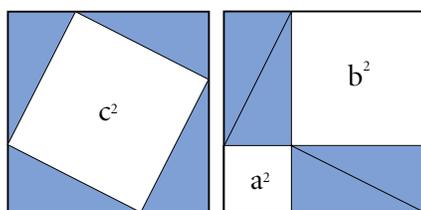


Figura 1.28 Dos maneras de colocar cuatro triángulos rectángulos en un cuadrado de lado la suma de sus catetos.

De la ecuación (1), al considerar que la terna 3, 4, 5 la satisface, surgen dos preguntas interesantes. La primera es si habrá otras ternas de números naturales que la cumplan, además de las obvias que son múltiplos de la original como por ejemplo 6, 8, 10, es decir, si hay otros triángulos rectángulos cuyos lados sean enteros. La respuesta es que sí. Por ejemplo, la siguiente “terna pitagórica” primitiva, que quiere decir que no es múltiplo de otra, es 5, 12, 13. La escuela que se formó alrededor de Pitágoras y continuó los trabajos después de su muerte, alrededor del año 507 a.C., obtuvo familias infinitas de estas ternas. Demostraron, por ejemplo, que cualquier número impar es parte de una de ellas. Sin embargo, no fue hasta el siglo XIII que Fibonacci encontró un método para listar todas las ternas pitagóricas.

La segunda pregunta que surge de la ecuación (1) es más general y ya no tiene una connotación geométrica tan evidente. No obstante, tiene una historia que ilustra el punto remarcado en esta sección. Si cambiamos el exponente 2 por uno más grande 3, 4, 5, 6, ..., que podemos denotar por n , se obtiene la ecuación

$$a^n + b^n = c^n. \quad (2)$$

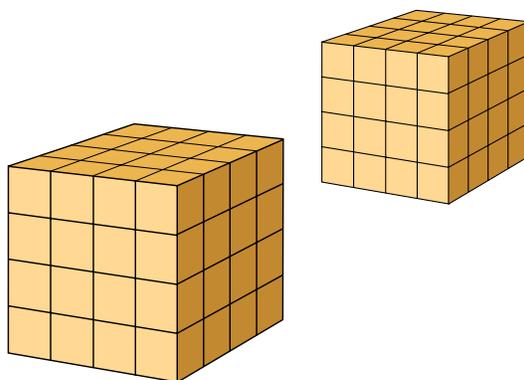


Figura 1.29 Dos cubos de lados enteros. ¿Con sus cubitos se podrá armar otro?

Para $n = 3$, esta ecuación tiene aun un significado geométrico. ¿Se puede partir un cubo formado por cubitos unitarios en dos cubos también formados por cubitos unitarios? Es una pregunta que nos invita a buscar una solución: ¿ $2^3 + 3^3$ será igual a 4^3 ? No, porque $8 + 27 = 35$ es diferente de 64. Pero quizás haya otros números para los cuales sí es cierto... o quizá, no. O bien podemos encontrar una terna de números y resolvemos la pregunta, o bien habrá que demostrar que esto es imposible.

La pregunta más general puede plantearse así: ¿la ecuación $a^n + b^n = c^n$ tiene soluciones enteras?, es decir, ¿existen números a, b, c enteros tales que $a^n + b^n = c^n$? Este problema se atribuye a Pierre de Fermat en el siglo XV, aunque es muy probable que se hubiera planteado antes dado lo natural que es.

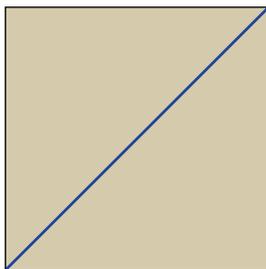
Cuenta la leyenda que Fermat escribió en el margen de un libro que había encontrado una demostración muy elegante de que sólo para $n = 2$ había soluciones enteras, resultado que se refiere a las ternas pitagóricas. Dicha demostración “elegante” nunca apareció, por lo que se hizo famoso el problema como el “Teorema de Fermat”. En realidad no era —todavía— un teorema sino una *conjetura*: la intuición profunda de algo que debe ser cierto, acompañada de fuertes indicios de que es cierto, pero que carece de una demostración. Fermat sí dejó una demostración para el caso $n = 4$, es decir, de que para $n = 4$ la ecuación (2) no tiene ninguna solución entera. A lo largo de los siglos siguientes se demostró la imposibilidad para otros casos como $n = 3, 5$ y 7. También se demostró para algunas familias infinitas de potencias n . Cada paso añadía fundamento y “glamour” a la conjetura.

No fue sino hasta 1995 cuando se concluyó la demostración del Teorema de Fermat. El matemático inglés Andrew Wiles dio los pasos finales y es a quien se atribuye la demostración. Quedará para los siglos venideros como el “Teorema de Fermat-Wiles”. Pero, como decía Newton, para ver lejos, Wiles se paró sobre los hombros de gigantes. En este caso, los gigantes son innumerables pues se desarrollaron áreas completas de las matemáticas, como la teoría algebraica de los números y la geometría algebraica, cuyas ideas y conceptos son la base de la demostración. Éste es un ejemplo célebre de un problema intrínsecamente matemático, sencillo de plantear, que fue semillero de grandes desarrollos. Ejemplo claro de cómo las propias matemáticas generan preguntas que dan lugar a nuevas matemáticas.

1.4.2 Los números irracionales y el espacio euclidiano

Volviendo a la antigua Grecia, la experiencia de haber descubierto que se pueden obtener resultados generales con demostraciones formales, como el Teorema de Pitágoras, dio un impulso enorme a las matemáticas. Se puede decir que con ello se da su banderazo de salida. La recién creada “escuela pitagórica” plantea muchos problemas y avanza en la sistematización de las matemáticas. Para sus miembros, fue como descubrir un continente inexplorado con tesoros y recompensas detrás de cada loma: teoremas y más teoremas, demostraciones y construcciones. Los hizo sentir que habían encontrado el camino del entendimiento y la iluminación por medio de la razón; experimentaban colectivamente el éxtasis de la creatividad intelectual y se constituyeron como una secta. Los pitagóricos llegaron a creer que *la estructura del universo era aritmética y geométrica*. Y lo creyeron con tal fervor que, cuando un joven de su escuela demostró con sus propios métodos que había una distancia que no se podía expresar mediante números, tomaron una actitud sobrecohedora que revela lo profundo de su convicción. Decidieron que se trataba de un error de Dios y que, como escuela, tenían la obligación de guardarle el secreto. Juraron sobre su vida no revelar nunca ese desliz en la “creación” del Universo.

El argumento del joven que reveló el “fallo de las divinidades” se conoce ahora como el descubrimiento de los números irracionales o, más concretamente, de que la raíz cuadrada de dos es irracional. La concepción que tenían los pitagóricos de la “aritmética” es que los números se expresan como razón entre dos naturales, que son de la forma $\frac{p}{q}$ con p y q enteros; éstos son lo que ahora conocemos como *números racionales*, porque se expresan como una “razón”, a pesar de que en México les llamamos “quebrados”. Por otro lado, en la geometría surge de manera muy natural, una longitud, la de la diagonal de un cuadrado *unitario*, con lado 1.



Como dicha diagonal es la hipotenusa de un triángulo rectángulo con catetos de longitud 1, por el Teorema de Pitágoras, al elevar a este número al cuadrado se obtiene 2, pues $1^2 + 1^2 = 1 + 1 = 2$. A este número lo llamamos ahora “raíz de 2” y lo denotamos $\sqrt{2}$. Para los griegos era el lado de un cuadrado con área 2. Suponer que $\sqrt{2}$ se expresa como quebrado o fracción lleva a una contradicción —lo que los pitagóricos explicaron como “un error de Dios”. Veamos cómo llegaron a esa contradicción.

Si suponemos que $\sqrt{2}$ es una fracción, entonces deberían existir dos números enteros p y q , tales que:

$$\sqrt{2} = \frac{p}{q}$$

Puede considerarse que p y q no son ambos pares —cada quebrado se puede escribir así ya que las fracciones $\frac{p}{q}$, pueden simplificarse hasta que el numerador o el denominador no sea par; por ejemplo, cuando simplificamos $\frac{12}{16} = \frac{6}{8} = \frac{3}{4}$, el numerador ya no es par. Al multiplicar ambos lados de la igualdad anterior por q se obtiene la nueva igualdad:

$$\sqrt{2} \cdot q = p$$

Si ahora se elevan al cuadrado ambos miembros, como $(\sqrt{2} \cdot q)^2 = 2 \cdot q^2$, se observa que:

$$2 \cdot q^2 = p^2$$

Como el lado izquierdo es par, ya que 2 es uno de sus factores, también lo es el lado derecho, es decir, p^2 es un número par. Obsérvese que el cuadrado de un número impar siempre es impar y que el cuadrado de un número par, además de ser par es divisible por 4. Así que como p^2 es par, debe ser divisible por 4. Por tanto, se sigue que q^2 también es par y, entonces, también q es par. Esto contradice lo que supusimos sobre p y q al principio de nuestro argumento: que no eran ambos pares. Esta contradicción nos indica que la hipótesis de que $\sqrt{2}$ podía escribirse como una fracción, como un quebrado, tiene que ser falsa. En otras palabras, $\sqrt{2}$ no puede ser un número racional y, por definición, es irracional.

De las propias matemáticas, de la interacción de dos de sus disciplinas —la aritmética y la geometría— y con razonamientos matemáticos, surgió un descubrimiento inesperado. Y con él, una andanada de preguntas y problemas: ¿habrá más números irracionales?, ¿cómo los expresamos o trabajamos con ellos?, etc. Éste es otro ejemplo histórico de cómo las matemáticas se motivan a sí mismas para crecer e indagar continuamente.

Por último, veamos un ejemplo de cómo algunos desarrollos matemáticos tienen consecuencias inesperadas y profundas tanto en la comprensión de la naturaleza como en la actividad humana.

El ímpetu que se dio al desarrollo de las matemáticas en la antigua Grecia alcanza un clímax con la aparición, alrededor del año 300 a.C., de los *Elementos de Euclides*. Son una serie de libros en los que se sistematizan los resultados y los conocimientos de geometría que se habían obtenido hasta ese momento, y en los que se establece el **método axiomático**. Nuestras ideas abstractas de punto, línea, plano, distancia, ángulo, área, volumen, etc., se fundamentan en esa obra, que se usó como libro de texto por más de dos milenios. En ella, se define el modelo teórico de lo que subyace al espacio en que vivimos, que se llama aún el **espacio euclidiano** y es el escenario de la geometría de nuestro entorno. Suponemos desde entonces que, sobre un espacio euclidiano de esencia geométrica tridimensional mínima, limpia y vacía, está transcurriendo lo que ocurre en el mundo y que ahí se hospeda y se mueve la materia. Es decir, suponemos que el mundo físico que nos incluye es euclidiano. Por mucho tiempo se pensó que el mundo físico que nos incluye era euclidiano. Sin embargo, a principios del siglo XX, Albert Einstein mostró que esa no era una buena concepción de nuestro Universo pues la materia deforma el espacio y, al usar las grandes escalas de la astronomía, éste deja de ser euclidiano. Sin embargo, espacios no tan grandes son —con bastante buena aproximación— euclidianos y por ello, tiene sentido hablar de planos, de triángulos y de círculos.

Emparentadas íntimamente con los círculos están las elipses. Un círculo visto de lado es una elipse; de alguna manera, las elipses son círculos “apachurrados”. En muchos casos en los que nuestros ojos ven una elipse —vasos, tazas, llantas, etc.—, nuestra mente lo interpreta como un círculo porque el cerebro “sabe” que al cambiar el punto de vista la figura es, efectivamente, un círculo. Los griegos se dieron a la tarea de estudiarlas y, en particular, de dar una definición precisa de ellas. Encontraron dos. La primera tiene que ver con cómo trazarlas en el plano. Son el lugar geométrico de los puntos cuya suma de sus distancias a dos puntos fijos, llamados focos, es constante. Así que para trazar una elipse en un jardín tendríamos que hacer lo siguiente: se clavan dos estacas —que serían los focos—, se les amarra una cuerda holgada cuya longitud es la constante que se menciona en la definición y, luego, al ir tensándola en todas las direcciones, se dibuja la elipse. A partir de esta definición queda claro que el círculo es un caso límite de elipse que corresponde a cuando los dos focos coinciden en su centro y entonces la constante es el doble del radio.

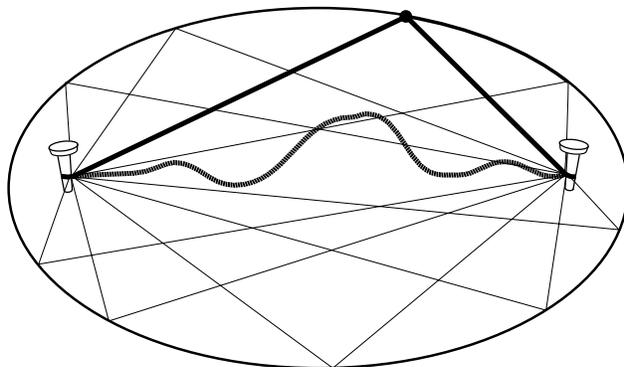


Figura 1.29

La segunda definición tiene que ver con la percepción de círculos y necesita de la tercera dimensión pues se sale del plano. Si levantamos el centro de un círculo perpendicularmente al plano en el que “vive” —digamos que horizontal— y luego consideramos todas las líneas que van de este punto al círculo, obtenemos un cono circular. Las elipses se obtienen al cortar este cono con planos inclinados cercanos al horizontal —la dirección en la que se inclina el plano, corresponde a aquella en la que se “alarga” el círculo. Nuestra percepción de círculos se explica con el proceso inverso: un cono elíptico cortado por otro plano origina el “círculo” real.

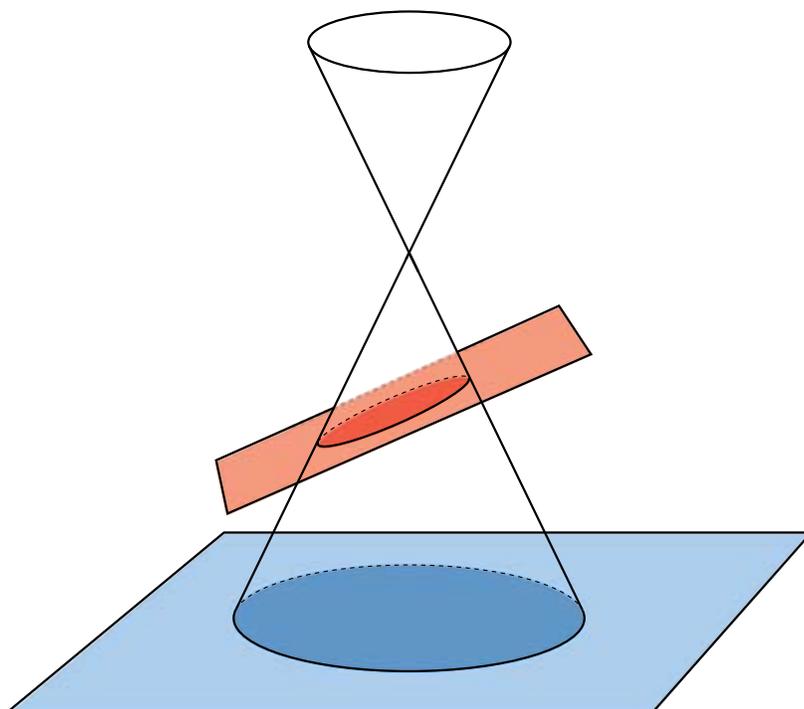


Figura 1.30

Si inclinamos más de la cuenta el plano —y ya que estamos en la teoría, no hay nada que nos detenga, por lo que sería natural hacerlo—, se obtienen dos nuevos tipos de curvas: las parábolas y las hipérbolas. De aquí el nombre de **Secciones Cónicas** que dieron los griegos, con la participación destacada de Apolonio de Perga, a estos tres tipos de curvas planas. Obtuvieron descripciones de estas curvas como lugares geométricos en términos de distancias, similares a la que vimos para las elipses. Se maravillaron con los resultados que obtuvieron y que dejaron como legado a la posteridad. Hicieron matemáticas por sí mismas que parecían muy alejadas de la “realidad”.

Quién hubiera pensado en ese entonces que, dos milenios después, esas mismas curvas sirvieran para entender fenómenos físicos como la trayectoria de los planetas alrededor del Sol —que son elipses con el Sol en uno de sus focos—, o las trayectorias de las balas o proyectiles —que describen parábolas. O bien, dentro de las propias matemáticas, como cuando Descartes puso coordenadas al plano y las secciones cónicas reaparecen como soluciones de ecuaciones cuadráticas con dos incógnitas. O, más aún, que las hipérbolas serían fundamentales para diseñar los lentes ópticos y las parábolas para las telecomunicaciones del presente. Esa “magia” que se ha dado una y otra vez en la historia, ese hecho insólito de que las matemáticas desarrolladas por sí mismas reaparecen una y otra vez en la naturaleza y en la actividad humana, y se interconectan con otros fenómenos, al parecer independientes unos de otros, es uno de los grandes misterios de las matemáticas.

1.5 CONCLUSIÓN

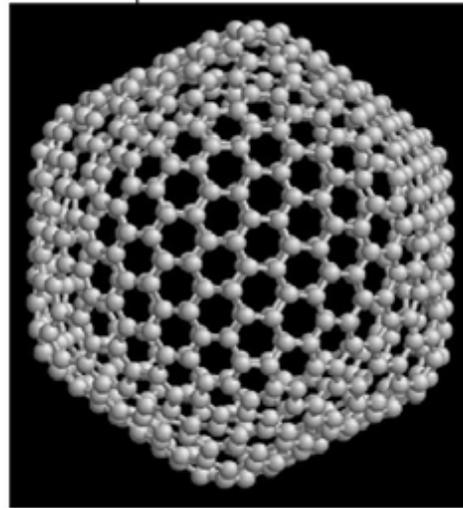


Figura 1.31 Fullerenos icosaédrico, un arreglo de 540 puntos que forman hexágonos y pentágonos. Esta forma estructura moléculas que se producen artificialmente y se estudian por sus aplicaciones en la medicina.

Los ejemplos de las secciones anteriores muestran al lector que las fuentes de los problemas matemáticos son variadas pero pueden clasificarse de manera natural en los tres grandes grupos mencionados: la actividad humana, la naturaleza y las matemáticas mismas.

A pesar de la variedad de sus orígenes, las matemáticas forman un cuerpo de conocimientos perfectamente coherente e interconectado. Los números se relacionan con las figuras geométricas y ambos, con las expresiones algebraicas. Todos los aspectos de las matemáticas se relacionan entre sí y forman una unidad de extraordinaria coherencia, belleza y poder.

Pero ¿qué son las matemáticas? El último capítulo del libro enfrentará esta pregunta que también es un reto para la mente humana, en este caso, de tipo filosófico. De momento, partiendo de los ejemplos presentados en este capítulo, daremos una respuesta tentativa. Las matemáticas son el destilado de la actividad intelectual; aquello que aunque provenga de problemas concretos resulta ser una verdad general necesaria y absoluta que trasciende los casos particulares que motivaron su descubrimiento. Las matemáticas son un producto de la observación, la reflexión y el trabajo humanos, pero sus verdades tienen tal fuerza que parecen ser parte de un mundo absoluto, cerrado y perfecto que existiera independientemente del hombre.

La capacidad que tienen las matemáticas para representar el mundo físico nos inclina a creer que hay una relación íntima, que aún no comprendemos bien, entre la naturaleza y las matemáticas. Sin embargo no podemos afirmar esta idea con total seguridad.

Toda la actividad matemática, independientemente de donde se inicie, se realiza apoyada en la irrefrenable curiosidad del hombre. Es sorprendente la fuerza que da al ser humano esta curiosidad y la cantidad de energía que está dispuesto a emplear para satisfacerla. Parece ser una característica genética que, como casi todas las de una especie, probablemente esté ligada a su capacidad para sobrevivir, lo mismo que la capacidad de oponer el dedo pulgar a los otros, caminar erguido y seguir el instinto sexual. Resolver crucigramas, hacer *sudoku*, jugar ajedrez o *bridge*, leer novelas de misterio y resolver acertijos son algunos ejemplos de actividades intelectuales cotidianas que los seres humanos realizamos por gusto sin que nadie nos obligue. Actividades que requieren esfuerzo, difíciles, pero que pueden proporcionar mucha satisfacción.

En la película de Stanley Kubrick, *2001 Odisea del espacio*, el mono, al darse cuenta de que puede usar un hueso como herramienta para defenderse y atacar, celebra su descubri-

miento al golpear a diestra y sinies- tra con él y, al lanzarlo al aire en su euforia, en pleno vuelo se transfor- ma en una nave espacial. Esta famo- sa, poética y emocionante escena ilustra la satisfacción del descubri- miento y sugiere que con ello se ini- cia un proceso de superación, que lo llevará a conquistar el planeta y posiblemente el Universo. Es una metáfora perfecta de la evolución humana y su relación con el descu- brimiento y la invención.

Nuevos descubrimientos renue- van la experiencia y repiten la emo- ción. El fenómeno acaba por con- vertirse en una característica genética de la especie que le ayuda en su progreso. Con el paso de muchas generaciones, el placer por el descubrimiento se extiende a todos los ámbitos de la vida: el trabajo, el juego, el deporte, la lectura, etc. La innovación produce placer. El descu- brimiento produce placer. La superación de un reto produce placer.

El hombre sabe que al esforzar su intelecto puede resolver problemas y descubrir verdades, dicha actividad le produce una satisfacción extraordinaria, equiparable a la de encon- trarse en el pico más alto de una montaña o al placer erótico, cuando esa pulsión, vital al fin y al cabo, lo lleva a la cúspide, al orgasmo. El placer de superar retos es, probablemente, lo que más ha contribuido al desarrollo de las matemáticas y a lo que las matemáticas deben su dinamismo y vitalidad.



Figura 1.32 El descubrimiento del hueso como herramienta en la película de Stanley Kubrick.

MATEMÁTICAS DE LA ACTIVIDAD HUMANA

TEMA

2



Figura 2.1 Niña jugando con ábaco | © Latin Stock México.

2.1 INTRODUCCIÓN

Nuestra vida moderna está llena de avances tecnológicos a tal grado que es difícil imaginarla sin ellos. Menos conocido, en cambio, es el hecho de que gran parte de esta tecnología se basa en desarrollos matemáticos, tanto de siglos pasados como de la actualidad. Hoy día, las matemáticas están escondidas sobre todo en los aparatos electrónicos y por ello es difícil percatarse de la importancia que tienen para el funcionamiento de éstos.

Ésta es una característica de la modernidad y no siempre fue así. Nuestra era actual nace a partir de una “época mecánica”, en la cual dominaba la preocupación sobre conceptos de la física que, desde Galilei, se formularon en el lenguaje de las matemáticas. Si retrocedemos todavía más, encontramos que aun las primeras culturas avanzadas, como los babilonios o los egipcios, tenían conocimientos matemáticos que usaron para gran provecho en la resolución de problemas de la vida diaria. Entre estos primeros conocimientos matemáticos es-

tá lo que hoy se conoce como el “Teorema de Pitágoras”, que sirvió para establecer cuándo un ángulo es recto.

Aunque las matemáticas que se tratan en este tema tienen su origen en la vida diaria, no es fácil encontrarlas en forma cotidiana. En efecto, las matemáticas que se verán aquí no son útiles en el sentido de que no aportan algo para ir de compras al mercado o arreglar un automóvil; al igual que la historia, por ejemplo, nos permiten entender mejor nuestra cultura y, por lo tanto, ser partícipes de ella.

Este tema no presenta un desarrollo histórico sino, más bien, uno conceptual: comienza al revisar el concepto de número y diversas aplicaciones; después versa sobre conceptos más geométricos como la medición de áreas o volúmenes. En seguida, el tema recorre los inicios de un campo que se formalizó sólo hasta el siglo XVIII y que hoy se conoce como el “cálculo integral”. Al final, se revisa incluso cómo las matemáticas llegaron a dominar el azar.

2.2 NÚMEROS PARA CONTAR

¿Alguien está haciendo “cuernos”? No; así muestran los chinos el número seis. Hay diferentes convenciones sobre cómo se cuentan los números con las manos. Nosotros usamos el índice y el dedo medio para indicar el dos, mientras los europeos usan el índice y el pulgar. Contar es una de las habilidades que aprendemos desde muy pequeños y nos distingue de muchos animales, aunque no de todos; se sabe, por ejemplo, que los cuervos pueden contar bastante bien hasta 6 o 7.

Se dice que contar es una de las primeras actividades humanas: aunque todos contamos con los dedos, no se hace igual en todos lados. Precisamente por ello puede parecer sorprendente que, hoy en día, la gran mayoría de las culturas usan sólo los números arábigos para denotar cantidades. ¿Qué es lo que hace tan particularmente especiales a los números arábigos? ¿Por qué se impusieron sobre otras maneras de denotar? Preguntas como esta y otras se responderán en este apartado. Además, queremos evidenciar que la forma en que denotamos los números incide en la facilidad para hacer operaciones como sumar, restar o multiplicar. Descubriremos que la invención del cero juega un papel muy importante en todo lo anterior. El cero no sólo es la expresión numérica de la ausencia o la nada, sino que posibilita el reutilizar las mismas cifras para denotar unidades, decenas, centenas, miles o millones, como veremos más adelante.

El libro *Discorsi* de Galilei fue escrito en el año MDCXXXVIII, como se puede apreciar en la figura 2.3. Si bien ya no estamos familiarizados con esta manera de escribir los números —conocidos como números romanos—, para leer la fecha en la portada del libro hay que saber los valores de cada símbolo:

$$M = 1000, \quad D = 500, \quad C = 100, \quad L = 50,$$

$$X = 10, \quad V = 5, \quad I = 1.$$

Entonces, podemos conocer, sin mayor dificultad, que el año de impresión es 1638.



Figura 2.2 Una seña que, según el contexto cultural, puede tener dos significados muy distintos.

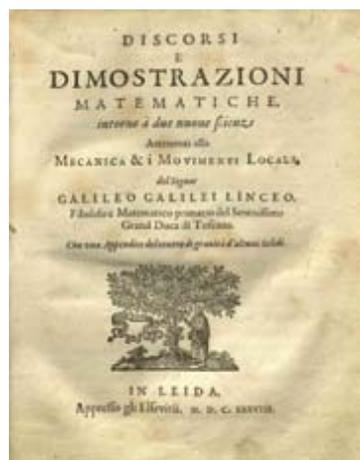
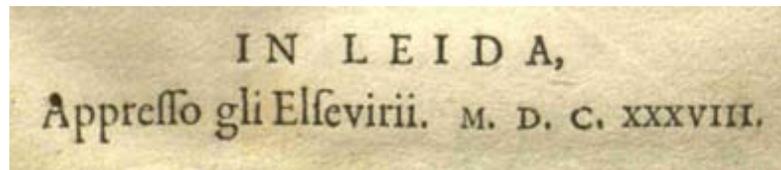


Figura 2.3 Portada de los *Discorsi* de Galileo Galilei.



Más difícil es traducir el año de impresión del libro *De revolutionibus orbium coelestium* de Copérnico, que es MDXLIII. Uno podría pensar que es igual a la suma de $1000 + 500 + 10 + 50 + 3 = 1563$ pero, en realidad, fue impreso en 1543. Lo que sucede en este caso es que hay que aplicar una regla para el uso de los números romanos: restar un signo de menor valor cuando antecede a uno de mayor valor. Como el X está antes del L, se resta 10 a 50 y por ello, se obtiene la cifra de 40.

Se complica el asunto si tratamos de multiplicar con números romanos. Por ejemplo:

$$\text{XIX por XLVII} = ?$$

¡Claro! X por XLVII es CDLXX, ya que simplemente se cambian los signos al multiplicar por diez. Así, lo que debemos calcular es:

$$\text{CDLXX} - \text{XLVII} + \text{CDLXX}$$

Sin embargo, ahora empezamos con algunas dificultades. Por ejemplo, $\text{CD} + \text{CD}$ no es CCDD ni CDCD , sino DCCC . Luego, $\text{LXX} + \text{LXX}$ no es LLXXXX , sino CXL . Los dos resultados anteriores se deben sumar y después, restarles XLVII. Entonces, la operación sería:

$$\text{DCCC} + \text{CXL} - \text{XLVII}$$

Se observa que el segundo sumando es $\text{C} + \text{XL}$, del cual se resta $\text{XL} + \text{VII}$. De esta manera, podemos quitar el término XL en ambos lados:

$$\text{DCCC} + \text{C} - \text{VII}$$

Ahora bien, el cien —es decir, C— tiene que prestar X para poder restar VII. Usamos que $\text{C} = \text{XC} + \text{X}$. Además, $\text{X} - \text{VII} = \text{III}$. Finalmente, el resultado de la operación es:

$$\text{DCCC} + \text{XC} + \text{III} = \text{DCCCXCIII}.$$

Hicimos este cálculo sin transformar los números a nuestro sistema decimal, es decir, sin usar que $\text{XIX} = 19$ y $\text{XLVII} = 47$. El algoritmo que se enseña en la secundaria nos proporciona el resultado sin problema alguno:

$$19 \times 47 = 893.$$

$$\begin{array}{r} 19 \\ \times 47 \\ \hline 133 \\ 76 \\ \hline 893 \end{array}$$

Figura 2.4 Ábaco.

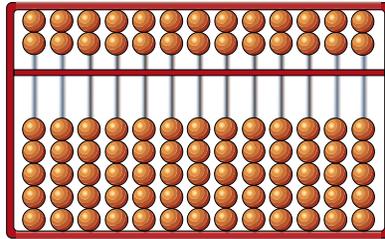
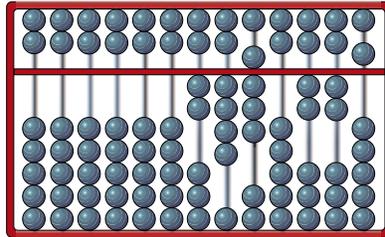


Figura 2.5 Ábaco.



El ábaco que se muestra en la figura 2.4 se basa también en el sistema decimal. Como se observa, tiene varias hileras usualmente divididas en dos partes: una de cinco canicas y otra de sólo dos. Las primeras valen uno —diez, cien, mil, etc.— mientras las de dos valen cinco —cincuenta, quinientos, cinco mil, etcétera.

En la figura 2.5 se muestra el número 2480225. Alguien experimentado en el uso del ábaco es muy rápido al efectuar sumas y, razonablemente rápido al hacer multiplicaciones. Cabe mencionar que en el fondo, emplea los mismos procedimientos que nos enseñan a nosotros.

Dichos procedimientos son muy naturales. Por ejemplo, si queremos sumar 493 más 2865, es claro que la suma tendrá: $3 + 5 = 8$ unidades,

$9 + 6 = 15$ decenas, $4 + 8 = 12$ centenas y $2 + 0 = 2$ miles. Pero 15 decenas son una centena y 5 decenas. Así que se debe aumentar el número de centenas por uno: tenemos 5 decenas y $12 + 1 = 13$ centenas. Éstas son a la vez un millar y 3 centenas. Tenemos entonces 3 millares, 3 centenas, 5 decenas y 8 unidades, es decir 3358. En el procedimiento que se aprende en la primaria, el razonamiento anterior se repite una y otra vez hasta que se hace en automático, de tal manera que ya no haya que pensar en él cada vez.

Con la multiplicación sucede algo similar: se hace de manera muy automatizada. En el fondo, hay una necesidad de hacerlo justo así y no de otra manera. Por ejemplo, si se quiere multiplicar 19 por 47 —el ejemplo que hicimos con los números romanos—, debemos calcular siete veces 19 y luego sumar cuarenta veces 19. Pero 7 por 19 se puede calcular como 7 por 9 y sumarle 7 por 10. De esta forma se obtiene que:

$$7 \times 19 = 7 \times 9 + 7 \times 10 = 63 + 70 = 133 \quad (1)$$

Por otro lado, cuarenta veces 19 no es otra cosa que multiplicar 19 primero por cuatro y luego por 10. Sin embargo, como multiplicar por diez implica añadir un cero, resulta que:

$$4 \times 19 = 4 \times 9 + 4 \times 10 = 36 + 40 = 76$$

y por lo tanto,

$$40 \times 19 = 760. \quad (2)$$

Al sustituir las operaciones de (1) y (2) tenemos que:

$$19 \times 47 = 19 \times 40 + 19 \times 7 = 760 + 133$$

y esto se resuelve con el procedimiento de la adición: $760 + 133 = 893$. Si procedemos a multiplicar siguiendo el esquema aprendido no hacemos otra cosa que seguir esta lógica.

Lo principal es que no importa cuán grandes sean los números, el procedimiento siempre funciona —aunque a veces sea muy tedioso y preferamos tomar una calculadora de bolsillo para resolverlo. Aunque estas herramientas nos son muy útiles, también es relevante comprender que el mecanismo que gobierna el procedimiento aprendido es una absoluta necesidad. En el fondo de cada multiplicación siempre están las tablas de multiplicar del

uno hasta el nueve. Desde esta perspectiva, no hay nada de sorprendente en el hecho de que todos los niños del planeta tienen que aprenderse las tablas de multiplicar de igual manera.

Si denotáramos nuestros números como los romanos tendríamos serios problemas para hacer cálculos. Ahora entendemos por qué se dice que los progresos científicos de los romanos fueron muy limitados, precisamente porque no usaban un buen sistema de notación.

2.3 NÚMEROS PARA MEDIR



Figura 2.6 Personificación de la justicia¹ por Luca Giordano (1634-1705), donde vemos como atributos la espada y la balanza. Esta última, simboliza la medida en el ámbito de la ética y es también una herramienta en los negocios. Las balanzas se conocen desde la Antigüedad, a través de los egipcios. < http://commons.wikimedia.org/wiki/File:Luca_Giordano_013.jpg >

Medir y contar son actividades distintas. Contamos piedras pero no contamos leche, contamos sillas y caballos pero no contamos la longitud de una mesa ni el tamaño de un rancho. De una cantidad de leche *medimos* su volumen, mientras que de una mesa y un rancho *medimos* su longitud y su tamaño, respectivamente.

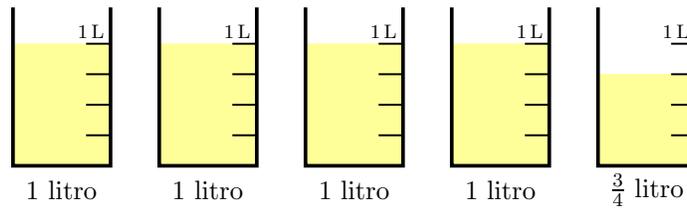
Para poder medir algo es necesario antes definir las **unidades** en las que se va a expresar lo que se mide. No tiene sentido decir que tenemos 5 de leche, que la mesa mide 2 o que el tamaño de un rancho es 24. Es necesario indicar las unidades que expresan los resultados de nuestra medición. Así, diríamos que tenemos 5 litros de leche, que la mesa mide 2 metros y el rancho, 24 hectáreas.

Los números para contar son muchos —de hecho, una infinidad— y aun así, no nos sirven para medir. Lo que medimos no suele constar de un número *exacto* de unidades sino que, frecuentemente, hay que recurrir a fracciones de estas unidades para dar el resultado de la medición. Por ejemplo, la vaca dio “cuatro litros y tres cuartos” de leche, la mesa mide “un metro y ochenta y cinco” centímetros, y el rancho tiene “veinticuatro punto tres” hectáreas.

¹ *Apologia dei Medici* (detalle); fresco en la Galería Palazzo Medici-Riccardi, Florencia (1684-1686).

Los números que utilizamos para contar 1, 2, 3, ... se llaman **enteros positivos** mientras que los que usamos para medir se conocen como **fracciones**. Las fracciones son cocientes de dos enteros positivos. Por ejemplo, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{5}{7}$ y $\frac{723}{45}$ son fracciones.

Figura 2.7 Al decir que, cuando ordeñamos una vaca obtuvimos 4 litros y 3 cuartos de leche, indicamos que la parte adicional —a los primeros 4 litros que obtuvimos— consta de 3 de las cuatro partes del siguiente litro —que hubiera sido el quinto—, mismo que no llegamos a completar quizá porque la vaca alimentó antes a su ternero.



Aun cuando podamos contar algunas cosas, por ejemplo, las fresas o las uvas que compramos en un mercado, preferimos medir su peso y expresarlo, no con un número exacto de fresas o uvas, sino como medio kilo de fresas y un kilo de uvas. Observemos que, a veces, no pluralizamos el sustantivo de aquello que compramos por peso, sino que decimos medio kilo de fresa y un kilo de uva, lo cual indica que estamos ignorando, a propósito, el número exacto de objetos y nos concentramos en su peso, que es normalmente una cantidad fraccionaria.

Las unidades que usamos para medir no siempre fueron las mismas. En la Antigüedad, las necesidades del comercio llevaron a los hombres a establecer unidades para medir peso, longitud, área y volumen. En distintas civilizaciones se utilizaron distintas unidades de medición. Cuando el comercio comenzó a crecer cruzando fronteras, se hizo necesario establecer equivalencias entre las unidades de medida de diferentes culturas. Estas equivalencias subsisten hasta nuestros días; por ejemplo, entre el llamado **sistema inglés** y el **sistema métrico decimal** donde un pie equivale a 30.5 centímetros, una milla a 1 609 metros y una libra a 454 gramos.

En general, los sistemas de pesas y medidas de todas las culturas utilizan fracciones exactas de una unidad para expresar la medida de cualquier cantidad menor que dicha unidad. Así, un pie es la tercera parte de una yarda y una pulgada, la doceava parte de un pie. Estas relaciones, como dijimos, se expresan con fracciones. Por ejemplo, al usar la notación convenida internacionalmente para pies y pulgadas —*ft* e *in*, respectivamente— podemos escribir:

$$1 \text{ ft} = \frac{1}{3} \text{ yarda}$$

$$1 \text{ in} = \frac{1}{12} \text{ ft}$$

Para facilitar el comercio internacional, hace un par de siglos casi toda la humanidad se puso de acuerdo en hacer uso de un sistema de medición estándar, el sistema métrico decimal, donde las distancias se miden en metros y decenas o centenas de metros, y otras unidades de longitud que siempre son potencias de 10 de un metro —por ejemplo, $1000 = 10^3$ metros son un kilómetro— cuando se trata de distancias muy grandes. Para las pequeñas se usan unidades que representan fracciones de metros y que caben exactamente 10^n —10 “a la n veces”, o bien, 10 elevado a la enésima potencia, donde n es cualquier número— veces en un metro, como es el caso de los centímetros —que son $\frac{1}{100} = \frac{1}{10^2} \text{ m}$ —, los milímetros — $\frac{1}{1000} = \frac{1}{10^3} \text{ m}$ — o las micras — $\frac{1}{1000000} = \frac{1}{10^6} \text{ m}$. Lo mismo se hace con las unidades de peso y con algunas otras, como las de temperatura.

Aun en los países donde impera el sistema métrico decimal —casi todos los del mundo—, las unidades de tiempo y angulares no utilizan —estrictamente— un sistema decimal. La definición original de las unidades de tiempo se basa en lo que tarda la Tierra en girar desde que el Sol está en su punto más alto hasta que ocurre lo mismo al día siguiente,

es decir, en un día. En particular, un segundo es la $60 \times 60 \times 24 = 86\,400$ —“ochenta y seis mil cuatrocientosava”— parte de un día. Si ocurriese un cataclismo —como que un meteorito muy grande chocara con la Tierra— que disminuyera un poco la velocidad de rotación del planeta, la medida del día cambiaría. Sin embargo, ya hay relojes muy precisos basados en las vibraciones moleculares de cristales con los cuales seguiríamos sabiendo lo que es un segundo, aunque ya no coincidiera con la 86 400 parte del día. En efecto, una hora es la veinticuatroava parte de un día, un minuto es la sesentava parte de una hora y un segundo la sesentava parte de un minuto. Estas unidades de medida definidas como una fracción de otras más grandes, se expresan por medio de fracciones propias, es decir, fracciones con numerador igual a uno. Usemos la notación convenida para horas y segundos con afán de mostrar lo anterior:

$$\begin{aligned} 1 \text{ h} &= \frac{1}{24} \text{ día} \\ 1 \text{ min} &= \frac{1}{60} \text{ h} \\ 1 \text{ s} &= \frac{1}{60} \text{ min} \end{aligned}$$

Para comunicar resultados simples de medición, es necesario que sepamos operar con fracciones y decimales. Por ejemplo, si nos dicen que una mesa mide seis pies y cinco pulgadas, la mayoría de nosotros no tenemos una idea clara de lo que esto significa y, si quisiéramos recortar un vidrio para ponerlo sobre la mesa, necesitaríamos convertir el dato al sistema métrico decimal —en el que tenemos nuestra cinta métrica—. ¿Cómo sabemos cuánto mide en metros y centímetros esa mesa? Podemos hacer la conversión de varias maneras, siendo la más conveniente aquella que requiera menos investigación de nuestra parte. Si no sabemos cuántos centímetros mide una pulgada y sí conocemos —conviene basarnos únicamente en— la equivalencia entre pies y centímetros que —sabemos— es de 30.5 cm por pie. Entonces, primero expresamos la medida de la mesa en pies. Como vimos, una pulgada es $\frac{1}{12}$ de pie, la mesa mide $6 + \frac{5}{12} = \frac{77}{12}$ pies. Para conocer la medida en centímetros, debemos multiplicar por el número de centímetros que hay en un pie —que es 30.5—, por tanto, la mesa mide $\frac{77}{12} \times 30.5 \text{ cm}$.

Hace años, en este punto hubiéramos desarrollado en un papel las operaciones; hoy en día recurrimos a una calculadora y, en cualquier caso, obtenemos que la mesa mide 195.7083333... cm; al aplicar el redondeo, concluimos que la mesa mide un metro y 96 centímetros, lo cual se expresa como 1.96 m o como 196 cm. Si fuera necesaria mayor precisión podríamos decir que la mesa mide un metro, noventa y cinco centímetros y siete milímetros o 195.7 cm.

El simple ejercicio anterior nos muestra cómo, por las necesidades del comercio y la actividad humana en general, fue necesario desarrollar los métodos para operar con fracciones y decimales, es decir, la **aritmética**. La necesidad de medir y comparar medidas llevó al hombre, primero, a desarrollar la aritmética de las fracciones y, posteriormente, a descubrir la conveniencia de utilizar la **notación decimal** para representar los resultados de una medición, así como la de adoptar un sistema básicamente decimal de unidades de pesas y medidas.

De esta forma, el ser humano medianamente instruido debe saber cómo realizar operaciones con fracciones y decimales ya que, sin contar con esta habilidad y aunque posea la mejor calculadora del mundo, le será imposible entender y comunicar resultados básicos de mediciones. No saber aritmética equivale a ser medio analfabeta. Afortunadamente, casi todas las personas aprenden la aritmética de las fracciones —llamadas quebrados hace algu-



<Imagen 8, de wikipedia>
>dia>

Figura 2.8 Modelo concreto para el metro —unidad del sistema métrico decimal resguardada en París— en una barra de platino-iridio.

nos años— en la escuela primaria que, también, es por fortuna obligatoria en todos los países civilizados. Las operaciones básicas de la aritmética de fracciones y decimales consisten en sumar, restar, multiplicar y dividir —tanto fracciones como decimales—, y en convertir una fracción a decimal y un decimal a una fracción. Las calculadoras pueden ayudarnos a obtener estas conversiones y realizan bastante bien las cuatro operaciones básicas con decimales, pero no operan con fracciones; esto debe saber hacerlo la persona si quiere apoyarse en la calculadora. Por ejemplo, en el problema de convertir la medida de la mesa de pies y pulgadas a centímetros, fue necesario saber expresar las cinco pulgadas como $\frac{5}{12}$ de pie y saber que había que sumar $6 + \frac{5}{12}$ para obtener el número de pies de la mesa como una fracción. Ya planteada la suma hay dos opciones: sumar fracciones y obtener $\frac{77}{12}$, como hicimos en el párrafo anterior, o bien, convertir antes la fracción $\frac{5}{12}$ a decimales y operar con ellos. Ambos caminos son correctos, aunque el primero mantiene la exactitud por más tiempo, de manera que el resultado final, aunque sea un valor aproximado, da lugar a menor incertidumbre que el que obtenemos mediante el segundo camino. De cualquier forma, para efectuar la conversión correctamente, el poseedor de una calculadora tiene que entender lo que va a hacer para plantear las operaciones antes de realizarlas, y esto requiere de un conocimiento en el uso de las fracciones y de la conversión entre fracciones y decimales. En realidad, aunque éstos se consideren temas elementales, tienen suficientes sutilezas para que valga la pena repasarlos. Para conveniencia del lector —que pudiera tener alguna carencia en estos conocimientos básicos—, presentamos aquí un resumen de estas sutilezas en el tema de las fracciones y los decimales.

En primer lugar varias fracciones distintas pueden representar un mismo número. Por ejemplo, $\frac{1}{2}$ representa lo mismo que $\frac{2}{4}$ y que $\frac{7}{14}$. De hecho, dados una fracción $\frac{m}{n}$ y cualquier entero positivo k , la fracción $\frac{m \times k}{n \times k}$ representa el mismo número que $\frac{m}{n}$. Los números representados como fracciones se llaman **números racionales**. Lo que acabamos de observar muestra que un número racional no es “una” fracción sino muchas, todas las que dan lugar al mismo cociente. Los matemáticos dicen a veces cosas como ésta: los números racionales son las clases de equivalencia de todas las fracciones $\frac{m}{n}$ tales que, para dos elementos $\frac{m_1}{n_1}$ y $\frac{m_2}{n_2}$ de una clase, existen enteros k_1 y k_2 que cumplen:

$$\begin{aligned} m_1 \times k_1 &= m_2 \times k_2 \\ n_1 \times k_1 &= n_2 \times k_2 \end{aligned}$$

El lector no tiene que leer ni entender estas “precisiones” para saber lo que es un número racional, basta que lo piense como el cociente de una fracción y sepa que si dos fracciones dan lugar al mismo cociente, entonces ambas representan el mismo número racional. Por ejemplo, $\frac{6}{8}$ y $\frac{15}{20}$ son dos fracciones que representan el mismo número racional; también podemos decir que estas fracciones son iguales —pues aunque sean obviamente distintas por sus numeradores y sus denominadores diferentes, los cocientes sí son iguales— y se pueden escribir $\frac{6}{8} = \frac{15}{20}$. Muchas veces es conveniente representar un número racional por su fracción más simple, que es precisamente aquella en la que el numerador y el denominador no tienen divisores comunes; por ejemplo, $\frac{6}{8}$ y $\frac{15}{20}$ pueden representarse por la fracción $\frac{3}{4}$. Una fracción que no tiene divisores comunes —como $\frac{3}{4}$ —, se llama irreducible.

Y... ¿qué es el cociente de dos enteros?, ¿qué es un número racional? Para no recurrir a un lenguaje técnico, es conveniente identificar a los números racionales como las expresiones decimales que se obtienen al dividir el numerador de una fracción por su denominador. En efecto, toda fracción da lugar a una expresión decimal. Para obtenerla, basta dividir el numerador entre el denominador. Algunas fracciones, como $\frac{3}{8}$, tienen una expresión decimal finita o cerrada:

$$\begin{array}{r}
 0.375 \\
 8 \overline{) 3} \\
 \underline{30} \\
 60 \\
 \underline{40} \\
 0
 \end{array}$$

Por otro lado, toda expresión decimal finita puede escribirse como una fracción. Por ejemplo, $56.3849 = \frac{563849}{10000}$.

Sin embargo, hay fracciones muy simples como $\frac{1}{3}$ o $\frac{1}{9}$ cuyas expresiones decimales son infinitas, aunque ¡atención! Existe una característica que define las expresiones decimales de las fracciones: si no son finitas, entonces son periódicas. Veamos algunos ejemplos de expresiones decimales que se obtienen de fracciones:

$$\begin{aligned}
 \frac{1}{3} &= 0.33333\dots \\
 \frac{1}{9} &= 0.11111\dots \\
 \frac{3}{7} &= 0.428571428571428571\dots
 \end{aligned}$$

He aquí el cálculo de la última división:

$$\begin{array}{r}
 0.428571428571\dots \\
 7 \overline{) 3} \\
 \underline{30} \\
 20 \\
 \underline{60} \\
 40 \\
 \underline{50} \\
 10 \\
 \underline{30} \\
 20 \\
 \underline{60} \\
 40 \\
 \underline{50} \\
 10 \\
 3\dots
 \end{array}$$

Observemos que, a partir de donde el tres aparece como residuo por segunda vez, los dígitos se repiten. La expresión decimal —no finita— de una fracción es periódica cuando se divide un número entre el denominador de la fracción y se obtiene un residuo menor al denominador y, al repetir las divisiones, eventualmente vamos a obtener uno de los residuos que obtuvimos antes —a partir de ese momento, como es lógico, los resultados de la división se repiten. No importa cuán grande sea el denominador, sólo puede haber un número finito de residuos al dividir por él y, por tanto, en algún momento el procedimiento va a repetir el resultado.

Finalmente, vamos a mostrar con un ejemplo cómo puede obtenerse la fracción que corresponde a una expresión decimal periódica. Por ejemplo, si $x = 0.539825398253982\dots$ es el número cuya expansión decimal es periódica, entonces:

$$100000x = 53982.539825398253982\dots$$

Y si hacemos la resta $100000x - x$, obtenemos:

$$\begin{array}{r}
 100000x = 53982.539825398253982\dots \\
 - \quad \quad x = \quad \quad 0.539825398253982\dots \\
 \hline
 99999x = 53982
 \end{array}$$

Por ello, al despejar se tiene que:

$$x = \frac{53982}{99999}.$$

Este truco se puede repetir cada vez que la expresión decimal es periódica. Para corroborar lo anterior, hagamos la división:

$$\begin{array}{r} 0.5398253982\dots \\ 99999 \overline{) 53982} \\ \underline{539820} \\ 398250 \\ \underline{982530} \\ 825390 \\ \underline{253980} \\ 539820 \\ \underline{398250} \\ 982530 \\ \underline{825390} \\ 253980 \\ \underline{53982} \dots \end{array}$$

Aunque la expresión decimal de un racional sea infinita, la información puede mostrarse con un número finito de cifras al indicar cuál parte se repite de manera interminable con una raya horizontal. Así, se puede escribir:

$$\frac{1}{3} = 0.\overline{3} \text{ o } \frac{1}{7} = 0.\overline{142857}.$$

Así, se ve que los números racionales son exactamente aquellos números cuya expresión en decimales resulta periódica. Con los números racionales se puede medir con cualquier exactitud requerida; sin embargo, como se verá en la siguiente sección, hay otros números que se consideran más una necesidad filosófica que práctica.

2.4 NÚMEROS PARA EXPRESAR LO CONTINUO

Figura 2.9 En una locomotora, la energía del vapor se traduce continuamente en el movimiento lineal del pistón que, a su vez, genera el movimiento circular en las ruedas y hace avanzar al tren.



Si observamos fenómenos de la naturaleza, es claro que no dan brinco. Lo mismo ocurre en sucesos cotidianos, por ejemplo, si el pistón de un motor se mueve hacia arriba y hacia abajo pasa por todos los estados intermedios. Por lo anterior, conviene buscar un concepto de número que exprese esta continuidad. Estos números se llaman **reales** y se entienden como números decimales con una precisión infinita. En este apartado se verá cómo los matemáticos lograron encontrar dichos números bajo el concepto de continuidad. Vale la pena hacer una advertencia de antemano: estos números son una necesidad filosófica pero no una práctica; las computadoras actuales pueden simular fenómenos continuos de manera

asombrosa, aunque se basan en una precisión limitada. ¿Qué es lo que sucede en una calculadora de bolsillo cuando realizamos la operación $0.4^{0.7}$? Esta y otras preguntas se resolverán más adelante.

Sabemos que no todo número es racional, es decir, una fracción; por ejemplo, $\sqrt{2}$ no es una fracción, como se explicó con anterioridad. Por otro lado, no toda fracción tiene una expresión decimal finita; de hecho, son racionales justamente aquellos números cuya expansión decimal es periódica.

Sin embargo, hay números con expresión decimal infinita como:

$$5.101001000100001000001\dots$$

donde después de un 1 hay un bloque de ceros, pero cada vez, este bloque tiene un cero de más. Esta expresión no puede ser periódica. Por lo tanto, dado lo que se argumentó con anterioridad, este número —al igual que $\sqrt{2}$ — no es un racional.

Si ponemos todos los números racionales como puntos sobre la **recta numérica**, dejarían huecos y, por eso, no bastan y se requiere rellenar los espacios vacíos entre ellos. Esto puede parecer algo extraño pero los racionales dan una precisión alta, de hecho, arbitrariamente alta. Para tener un ejemplo concreto, veamos cómo se mueve un pistón de un cilindro de una locomotora de vapor.

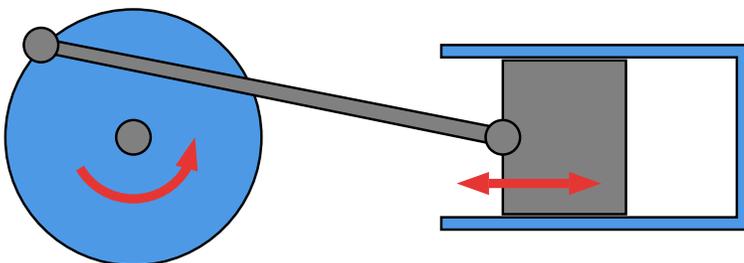


Figura 2.10 Esquema del movimiento de un pistón.

El pistón se encuentra en la parte baja delantera, a la altura de las ruedas, y se mueve horizontalmente por la presión del vapor en el cilindro como se muestra en la figura 2.10. En cada momento se puede determinar su posición y expresarla con un número —si fijamos bien las unidades de medida y un punto de referencia hacia donde se mide—; por ejemplo, al medir en centímetros la posición del pistón hasta el inicio del cilindro. Es evidente que las mediciones deben arrojar todos los números desde un mínimo hasta un máximo posible. No hay huecos porque el pistón no da brinco. Por ello, los números racionales no bastan; se requieren forzosamente todos los números **reales**: aquellos números con una expansión decimal.

Los números cuya expansión decimal no se vuelve periódica llenan los huecos en la recta numérica. Ésta se debe imaginar como una recta graduada, es decir, con marcas que indican la ubicación de los números.

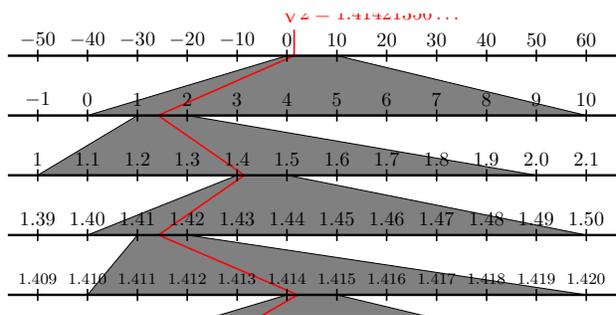


Figura 2.11 Esquema de la recta numérica.

Si se interseca una circunferencia, con centro en el origen y radio 2, con la diagonal entre los dos ejes de coordenadas, entonces, los dos puntos de intersección son $(\sqrt{2}, \sqrt{2})$ y $(-\sqrt{2}, -\sqrt{2})$. Esto quiere decir que los puntos de intersección no tienen coordenadas racionales y así, ésta es la segunda razón para considerar a los números reales como concepto matemático.

2.5 ¿CÓMO CALCULAR DE MANERA EFICIENTE?



Figura 2.13 La regla de cálculo, tan de moda en la primera mitad del siglo XX, ahora está en desuso por la aparición de las calculadoras de bolsillo. Ambas utilizan la misma técnica para simplificar los cálculos: los logaritmos descubiertos alrededor de 1620.

El principio del siglo XVII es el tiempo de Galilei, de la aceptación del método científico, el tiempo de Tycho Brahe y de Kepler, de los telescopios que permitían una observación más precisa de la bóveda celeste y el tiempo de Blaise Pascal, quien elaboró una de las primeras máquinas que podían sumar y multiplicar. En aquel entonces también aparecen innovaciones importantes al elaborar los relojes de precisión. En resumen, la primera mitad del siglo XVII es una era donde la mecánica fina se vuelve importante para la ciencia y la sociedad.

Sin embargo, la mecánica fina requería también unos cálculos cada vez más precisos. Fue a inicios del siglo XVII cuando Jost Bürgi —matemático y relojero suizo— descubrió los logaritmos al intentar facilitar su trabajo con los cálculos de precisión —que eran bastante tediosos en aquel entonces—, también descubiertos y publicados por John Napier, una década después. Hay muchas historias parecidas en las que se habla de que el tiempo es “maduro” para cierto descubrimiento pues, simultáneamente, varias personas encuentran lo mismo de manera independiente. Hoy en día el mundo reconoce a Napier como el descubridor de los **logaritmos** y, tal vez, ello sea correcto dado que fue él quien divulgó la idea y así, la puso al servicio de la humanidad.

A continuación veremos en qué consisten estos logaritmos y por qué son tan útiles para facilitar el trabajo con los cálculos. Empezaremos como lo hicieron Bürgi y Napier, al considerar sucesiones algebraicas y geométricas, que se muestran respectivamente a continuación:

$$2, 5, 8, 11, 14, \dots$$

$$3, 9, 27, 81, \dots$$

Aquellos que gusten de los retos intelectuales se pueden preguntar cómo sigue cada una de estas sucesiones, es decir, ¿cuál es el siguiente número? Para ello hay que indagar sobre el patrón que rige la sucesión. En nuestro caso no es difícil: en la primera, el aumento de un número a otro es siempre el mismo y cada vez la sucesión crece en tres unidades mientras que, en la segunda, cada nuevo número se obtiene al multiplicar el anterior por 3.

Veamos la segunda sucesión más de cerca: empieza con 3 y luego este 3 se multiplica repetidamente por 3. Podemos escribir la sucesión de la siguiente manera:

$$3, 3^2, 3^3, 3^4, \dots$$

En este caso, la simple observación que da lugar al descubrimiento de Bürgi y Napier es el constatar que la sucesión de los exponentes:

$$1, 2, 3, 4, \dots$$

es una **sucesión algebraica**. El gran problema de los cálculos en aquella época no fue la suma, sino la multiplicación. ¿Cómo multiplicar 3^5 con 3^7 ? Consideremos que $3^5=243$ y $3^7=2\ 187$. Ahora, podríamos calcular el producto pero puede ser provechoso observar que 3^5 es el producto de cinco 3 y 3^7 , el producto de siete, por lo que:

$$3^5 \cdot 3^7 = 3^{5+7} = 3^{12}$$

Hasta aquí parece que no hemos ganado nada, sólo hicimos una adición, sumamos los exponentes. Ahora bien, si pudiéramos representar más números como potencias del 3 podríamos rápidamente hacer más multiplicaciones al efectuar una única suma.

En efecto, lo anterior se hizo pero no con el número base 3 sino con el 10, dado que nuestro sistema es decimal. Aquí se muestran algunos valores:

$$\begin{aligned} 1 &= 10^0 \\ 2 &= 10^{0.301} \\ 3 &= 10^{0.477} \\ 4 &= 10^{0.602} \\ 5 &= 10^{0.699} \\ 6 &= 10^{0.778} \\ 7 &= 10^{0.845} \\ 8 &= 10^{0.903} \\ 9 &= 10^{0.954} \\ 10 &= 10^1 \end{aligned}$$

Así, por ejemplo, $2 \cdot 3 = 10^{0.301} \cdot 10^{0.477} = 10^{0.301+0.477} = 10^{0.778} = 6$. ¡Claro! Sabemos multiplicar 2 por 3 más velozmente en nuestra cabeza, pero este simple ejercicio explica el funcionamiento.

En aquel tiempo se empezaron a elaborar libros enteros que contenían tablas de logaritmos, como se muestra en la figura 2.14.

Elevar al cuadrado significa duplicar el exponente: lo vemos si comparamos los exponentes que corresponden a los números 2, 1 y 8, que respectivamente son 0.301, 0.602 y 0.903. De manera similar, obtener la raíz de 5 es ahora fácil: hay que dividir el exponente correspondiente entre 2,

lo que es lo mismo que:

$$\sqrt{5} = 10^{0.699 \div 2} = 10^{0.3495}$$

Con una tabla de logaritmos se puede buscar rápidamente el valor, que es 2.236. Para multiplicar 61.235 por 5.961 se buscan los exponentes correspondientes que son, precisamente, los logaritmos de dichos números en la tabla. Entonces, se encuentra $\log 61.235 = 1.787$ —que equivale a decir que $61.235 = 10^{1.787}$ — y $\log 5.961 = 0.775$. Por lo tanto, su producto satisface que:

<Imagen 14, hay que buscarla>

<ENTRA IMAGEN QUE MUESTRA UNA PÁGINA DE UNA TABLA DE LOGARITMOS>

Gr.	num	Signif.	Logarithm	Diferencia	Antilogarithm	Signif.
0	1	00	0000000		1	00000000
1	2	301	3010300		2	30103000
2	3	477	4771213		3	47712130
3	4	602	6020600		4	60206000
4	5	699	6989700		5	69897000
5	6	778	7781513		6	77815130
6	7	845	8450980		7	84509800
7	8	903	9030890		8	90308900
8	9	954	9542425		9	95424250
9	10	000	0000000		10	00000000
10	20	301	3010300		20	30103000
11	30	477	4771213		30	47712130
12	40	602	6020600		40	60206000
13	50	699	6989700		50	69897000
14	60	778	7781513		60	77815130
15	70	845	8450980		70	84509800
16	80	903	9030890		80	90308900
17	90	954	9542425		90	95424250
18	100	000	0000000		100	00000000
19	200	301	3010300		200	30103000
20	300	477	4771213		300	47712130
21	400	602	6020600		400	60206000
22	500	699	6989700		500	69897000
23	600	778	7781513		600	77815130
24	700	845	8450980		700	84509800
25	800	903	9030890		800	90308900
26	900	954	9542425		900	95424250
27	1000	000	0000000		1000	00000000

Figura 2.14 Tabla de logaritmos.

$$\log(61.235 \cdot 5.961) = 1.787 + 0.775$$

y obtenemos un cálculo sencillo que se puede hacer casi en la cabeza cuyo resultado es 2.562.

Buscando en las tablas de logaritmos se encuentra que $10^{2.562} = 364.8$, lo cual se acerca bastante bien al resultado correcto, que es $61.235 \cdot 5.961 = 365.021835$. Para tener mayor precisión, se requerían libros cada vez más gruesos con tablas de logaritmos más precisos.

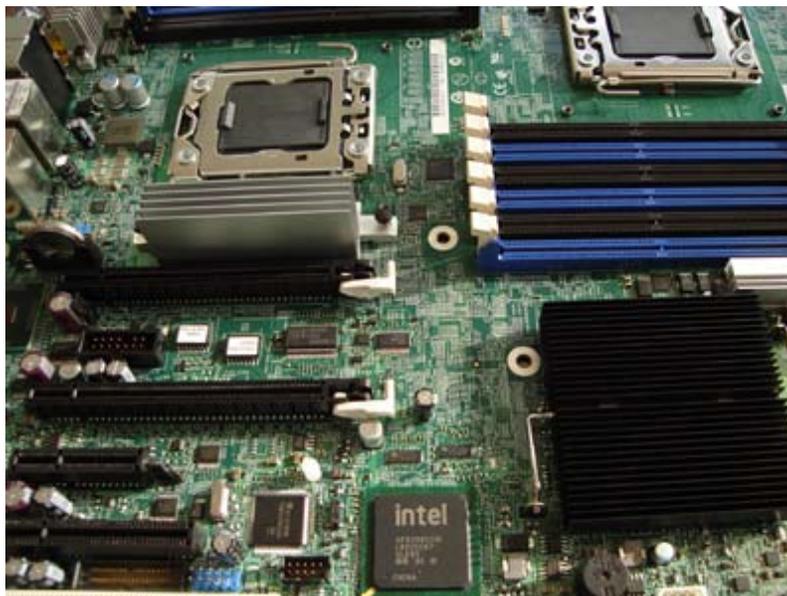
Ya en 1624, Henry Briggs publicó la primera tabla donde reportó los logaritmos de los primeros 20 mil números naturales con una precisión de 14 dígitos decimales. A finales del siglo XVII, se publicaron varios libros con logaritmos en los que se calculaba hasta con 6 dígitos de precisión.

La regla de cálculo que se muestra en la figura 2.13 fue inventada poco después del descubrimiento de los logaritmos. El principio se basa en que ambos lados muestran escalas logarítmicas. Así, con un simple deslizamiento, se podía leer directamente el resultado de una multiplicación hasta, al menos, con un dígito de precisión sin tener que hacer cálculo alguno.

La regla de cálculo desapareció rápidamente con la aparición de las calculadoras de bolsillo, que usan en su arquitectura binaria el principio de los logaritmos y lo combinan con rutinas eficientes para calcular, tanto los logaritmos como sus inversos, los exponenciales, para no almacenar tablas grandes de información.

El ejemplo de los logaritmos muestra uno de los grandes logros en el que se aprecia cómo la humanidad usa las matemáticas para simplificar el quehacer de calcular. Al mismo tiempo, muestra que todos estos descubrimientos e inventos están presentes en nuestra cultura de manera oculta, en este caso, en las calculadoras de bolsillo.

2.6 LOS NÚMEROS DE LA COMPUTACIÓN



<Imagen 15, la foto es nuestra>

Figura 2.15 Las computadoras han revolucionado nuestra vida de una manera muy determinante. Algunas de sus capacidades son almacenar y procesar grandes cantidades de información. Sorprendentemente, todo lo que sucede dentro de una computadora se realiza con sólo dos estados físicos de algún material que, matemáticamente, se representan como el 0 y el 1. Aquí se muestra la "tarjeta madre" de una computadora moderna.

Las computadoras pueden hacer operaciones numéricas a una altísima velocidad y ésta es una de sus aplicaciones más importantes. Pero, ¿cómo maneja los números una computadora? En esta sección veremos la forma en que las computadoras "almacenan" los números y qué tipo de números pueden utilizar.

Hay distintas representaciones de los números en las computadoras, la mayoría utiliza el **sistema binario** o de base 2, que es el más natural en su medio, pues los sistemas de almacenamiento de información están basados en grandes cantidades de imanes microscópicos, cada uno de los cuales puede estar sólo en dos “estados” o posiciones —digamos, para simplificar, que pueden estar orientados hacia arriba o hacia abajo—. Al guardar y acomodar información en un dispositivo de memoria, se ordenan algunos de esos microimanes hacia arriba y otros hacia abajo. Un sistema tal consta, entonces, de muchos elementos de memoria, cada uno de los cuales, al presentar estos dos estados, puede almacenar lo que se llama un bit de información. Un bit es la información requerida para saber si una cosa es o no es. Repasemos este concepto: supongamos que el “sí” lo representamos con un imán orientado hacia arriba y el “no” con un imán orientado hacia abajo. El “sí” o la orientación hacia arriba, puede a su vez representarse numéricamente con un 1, y el “no” o la orientación hacia abajo, con un 0. Así, por ejemplo, si tenemos ocho de estos microelementos de memoria podemos representar su estado con un conjunto ordenado de ocho dígitos binarios o sea con ocho números cada uno de los cuales es 0 o 1. Por ejemplo:

01100101
00000001
11110000
10101010,
etcétera,

son posibles estados de este conjunto ordenado de elementos de memoria o bits. Si el bit tiene un 1, decimos que está encendido y, si tiene un 0, decimos que está apagado. Un conjunto ordenado de ocho bits se llama byte. Los bytes se usan mucho en la computación porque las primeras computadoras se hicieron para operar sobre bytes, que se ha conservado como el elemento mínimo de información que se usa en ellas.

Supongamos ahora que deseamos usar un byte para representar un número. ¿Cómo lo hacemos? ¿Cuántos números distintos podríamos representar con un byte? Lo natural es utilizar el sistema de numeración binario que consiste en asignar al primer bit del byte el valor 1 si está prendido —y cero si no lo está—; al segundo, el valor 2; al tercero, el $4 = 2^2$; al cuarto, el $8 = 2^3$; al siguiente, el $16 = 2^4$ y así sucesivamente. Recordemos que $2^1 = 2$ y $2^0 = 1$. Por analogía con la notación decimal, consideramos a los bits ordenados de derecha a izquierda. Así, por ejemplo, un byte con la configuración 01100101 representa al número 101 según se muestra a continuación:

$$\begin{aligned} 01100101 &= 0 \times 2^7 + 1 \times 2^6 + 1 \times 2^5 + 0 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \\ &= 2^6 + 2^5 + 2^2 + 2^0 \\ &= 64 + 32 + 4 + 1 = 101 \end{aligned}$$

Cabe agregar que el 101 es el ciento uno y está en representación decimal y no binaria, es decir, lo interpretamos como:

$$101 = 1 \times 10^2 + 0 \times 10^1 + 1 \times 10^0$$

Ahora vamos a representar el número 213 en forma binaria como ejercicio. Para ello, hay que comenzar por descomponerlo en una suma de potencias de 2:

Como 213 es mayor que $128 = 2^7$, usaremos 2^7 . Nos falta agregar $213 - 128 = 85$, que es mayor que $64 = 2^6$, por lo que también sumaremos 2^6 . Nos falta incluir $85 - 64 = 21$, que es

menor que $32 = 2^5$ pero mayor que $16 = 2^4$, por lo que no agregaremos 2^5 , pero sí 2^4 . Finalmente, pongamos lo que nos falta: $21 - 16 = 5 = 2^2 + 2^0$. Por lo tanto, la descomposición del número 213 que buscábamos es $2^7 + 2^6 + 2^4 + 2^2 + 2^0$, que nos lleva a la representación binaria 1101011.

Las computadoras manejan los números enteros en formato binario como el que se acaba de mostrar, excepto que, en general, el primer bit se usa para indicar el signo del número —es decir, si el primer bit está “apagado”, el número es positivo, si en cambio, está “encendido”, el número es negativo. Uno de los formatos más utilizados en las computadoras para los números enteros es de 16 bits, llamado *short* en el lenguaje Java. En un entero de tipo *short*, el primer bit se utiliza para el signo y los 15 restantes para el valor absoluto del número. Así, por ejemplo, la expresión binaria:

10000000 1101011

representa el número -213 . Con 16 bits se pueden representar números enteros entre $-2^{15} = -32\,768$ y $2^{15} - 1 = 32\,767$. En efecto, con 15 bits, el número más grande que puede representarse es aquel en el que todos los bits están encendidos, por lo tanto es $2^{14} + 2^{13} + \dots + 2^1 + 2^0 = 2^{15} - 1 = 32\,767$. Para los negativos se puede aprovechar un número más pues el cero está representado entre los no negativos y no hace falta repetirlo, pero no explicaremos la razón de este detalle. La forma más usada para expresar enteros en lenguaje Java es con el tipo *int* que usa 32 bits, con los cuales se pueden representar los enteros entre $-2^{31} = -2\,147\,483\,648$ y $2^{31} - 1 = 2\,147\,483\,647$.

La representación de los números reales es bastante más complicada. En lenguaje Java se emplea la norma IEEE 754—estándar internacionalmente aceptado y usado casi universalmente—. En esta norma hay dos formatos de datos, llamados *float* y *double*, que son los más utilizados. Ambos son “representaciones de punto flotante”, el primero se denomina de precisión simple y usa 32 bits, mientras que el segundo, que es de doble precisión, usa 64 bits. También hay uno de cuádruple precisión que utiliza 128 bits. Aunque el más usado es el de 64 bits, para simplificar aquí la presentación, describiremos solamente el de 32 bits.

De los 32 bits que se usan para representar un número real del tipo *float*, el primero se usa para el **signo**, los ocho siguientes para el **exponente** y los 23 restantes para la **mantisa**, como se muestra a continuación al representar el número -231.125 :

1 100001100 11001110010000000000000
signo exponente mantisa

La fórmula para obtener el número en cuestión a partir de la representación consiste en, primero, obtener los valores enteros positivos del signo, el *exponente* y la *mantisa* a partir de sus representaciones decimales. Así:

$$\begin{aligned} \text{signo} &= 1 \\ \text{exponente} &= 2^7 + 2^2 + 2^1 = 128 + 4 + 2 = 134 \\ \text{mantisa} &= 2^{22} + 2^{21} + 2^{18} + 2^{17} + 2^{16} + 2^{13} = 6\,758\,400 \end{aligned}$$

Y luego, al aplicar la fórmula:

$$\text{número} = (1 - 2^{\text{signo}}) \times \left(1 + \frac{\text{mantisa}}{2^{23}}\right) \times 2^{\text{exponente} - 127}$$

que, en este caso, nos lleva al resultado:

$$\begin{aligned}
 \text{número} &= (1 - 2) \times \left(1 + \frac{6758400}{2^{23}}\right) \times 2^{134-127} \\
 &= -1.8056640625 \times 2^7 \\
 &= -231.125
 \end{aligned}$$

Estas representaciones de punto flotante están diseñadas no sólo para guardar los números dentro de las computadoras, sino para que éstas puedan realizar las operaciones básicas con gran eficiencia. Los procesadores numéricos de las computadoras son los que se encargan de realizar dichas operaciones. Afortunadamente, las computadoras hacen todo esto con mucha facilidad y rapidez. Lo importante es que tengamos conciencia de que la representación de números reales dentro de las computadoras dista mucho de ser equivalente al concepto matemático de número real, más bien, se limita a utilizar sólo algunos números —todos racionales con expansión binaria finita— y a brindarnos buenas aproximaciones de resultados de los cálculos que les pedimos hacer. Por ejemplo, el número π representado en formato de doble precisión es equivalente al valor decimal:

3.141592653589793

mientras que sabemos que el verdadero valor de π tiene una representación decimal infinita.

2.6 MEDIR LO INALCANZABLE

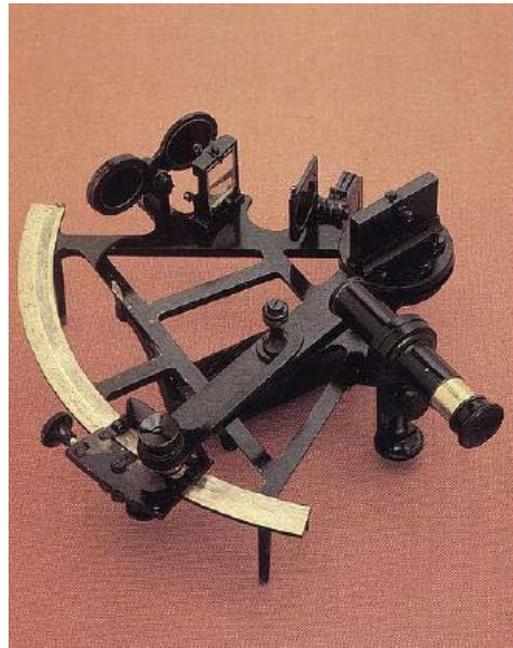


Figura 2.16 La imagen muestra un sextante, herramienta para medir la latitud en alta mar usando el Sol. Casi desde que aparecieron sobre la Tierra, los seres humanos han extendido su percepción y medición más allá de lo que directamente alcanzan al usar, por un lado herramientas sofisticadas como el sextante y por otro, el razonamiento matemático como se descubrirá en esta sección.

Medir distancias pequeñas es muy fácil. Basta una cinta métrica y si con ella no es suficiente, se pueden poner marcas, hacer varias mediciones en serie y luego sumar. También, usando geometría básica se pueden medir distancias enormes o inaccesibles; aquí se hablará de estos métodos para medir distancias. El principio fundamental en que se basan es el de la semejanza de triángulos. Dos triángulos son **semejantes** si sus tres ángulos coinciden. Tienen entonces la misma forma y hay una constante que relaciona los lados de uno con los del otro, o bien, las proporciones entre lados correspondientes son iguales. Así que con pocos datos de un triángulo grande podemos obtener los demás.

Veamos primero un ejemplo de un método que usan los maestros de obras para aproximar distancias. Supongamos que tenemos el brazo extendido hacia enfrente y el pulgar levantado. Al cerrar un ojo y luego el otro, el dedo parece “saltar” o “brincar” de lugar en el fondo, que llamaremos “pared”. Lo que pasa es que cada ojo “proyecta” al pulgar en un punto distinto de la pared y, entonces, se forman dos triángulos que comparten el vértice en el dedo. Se arma un triángulo chico con los dos ojos, y el otro, grande, con los dos puntos de la pared donde se proyecta el pulgar. Si estamos frente a la pared, estos dos triángulos son **isósceles** y semejantes —en el vértice del pulgar tienen el mismo ángulo.

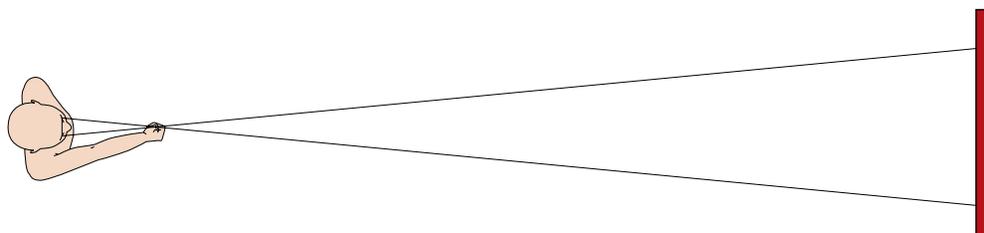


Figura 2.17 Esquema de los triángulos isósceles formados al usar el pulgar como vértice.

Como en el triángulo chico la proporción de la altura a la base —del brazo extendido a la distancia entre los ojos— es, aproximadamente, de 10 a 1, en el triángulo grande se cumple la misma relación. De aquí que, si sabemos que la distancia a la pared es, más o menos, de treinta metros, entonces el “brinco” del dedo en la pared es de aproximadamente tres metros. O bien, si el dedo brincó lo que mide un coche pequeño —alrededor de cuatro metros—, este debe estar como a una distancia de cuarenta metros.

Así se pueden medir distancias inaccesibles con semejanza de triángulos. Es sorprendente, pero fueron ideas igual de sencillas las que permitieron dar una primera estimación del tamaño de la Tierra y de la Luna. La primera estimación del diámetro de la Tierra la hicieron los griegos. Eratóstenes notó que, en la ciudad de Asuán —en Egipto—, la luz del Sol entraba de lleno a los pozos al mediodía del solsticio de verano —junio 21—, cuando las sombras llegan a su mínima longitud en el hemisferio norte. Esto sucede porque dicha ciudad está casi en el Trópico de Cáncer; que es el paralelo más al norte donde la luz del Sol puede caer o incidir verticalmente —literalmente “a plomo” — y lo hace justo en el solsticio de verano.

El cálculo de Eratóstenes se basó en medir el ángulo con el que inciden los rayos del Sol al mediodía del solsticio de verano en Alejandría, que está al norte de Asuán. Este ángulo resultó ser de $\frac{1}{50}$ de la vuelta completa — 2π radianes o 360° —. Así que, al multiplicar por 50 la distancia entre estas dos ciudades, se obtiene una aproximación de la circunferencia de la Tierra y, por consiguiente, su diámetro al dividir entre una aproximación de π .

Lo impresionante es que, con los métodos para medir ángulos y distancias de aquella época, el error en el cálculo fuera pequeño. No se tiene certeza del cálculo preciso de Eratóstenes pues, en sus escritos, la unidad de medida de longitud que usó fueron los **estadios** y en la actualidad, persiste la discusión histórica de a cuánto equivalen. Eratóstenes consideró la distancia de Asuán a Alejandría de 5000 estadios. De aquí, la circunferencia de la Tierra resulta de $5000 \times 50 = 250\,000$ estadios. Con el valor máximo que se tiene de un estadio, que es de 196 m, obtenemos un total de 19000 km y con el mínimo, 157 m, serían 39250. El valor medio de la circunferencia de la Tierra que se estima hoy día es 10000 km, así que Eratóstenes andaba muy cerca.

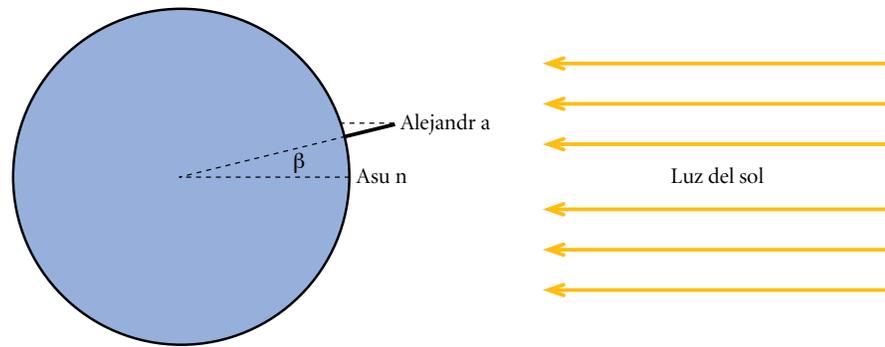
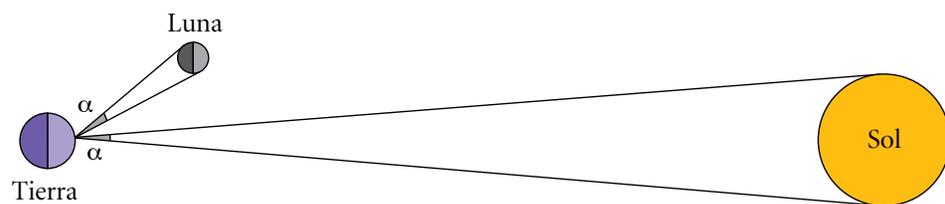


Figura 2.18 Método de Eratóstenes para medir el radio de la Tierra.

Para Eratóstenes, medir la circunferencia de la Tierra era un reto intelectual, “ciencia pura”. Casi dos mil años después, cuando Colón planeaba su viaje hacia el oriente navegando en dirección opuesta, el mismo asunto se convirtió en cuestión de vida o muerte —pues la cantidad de víveres que necesitaba para la travesía dependía de la distancia a recorrer—. Por suerte, el cálculo de Colón era erróneo, pues pensaba que la Tierra era más pequeña de lo que en realidad es y se lanzó a la famosa aventura aunque otro error canceló al primero: se le atravesó un continente insospechado en el camino y los víveres le alcanzaron, aunque él siempre creyó que había llegado a su destino.

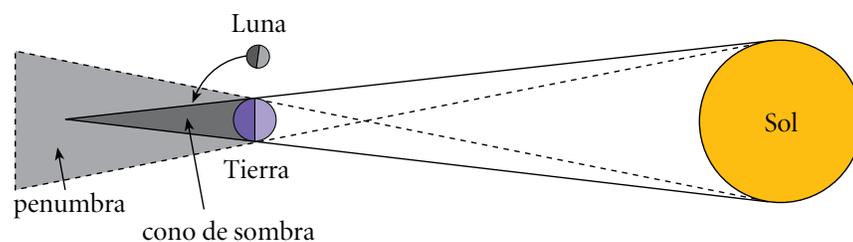
Otras mediciones astronómicas que hicieron los griegos fueron la de la distancia de la Tierra a la Luna y el tamaño de la Luna. Tuvieron ayuda de una enorme coincidencia: la Luna y el Sol tienen —en apariencia— casi el mismo tamaño, es decir, el ángulo de nuestro ojo a los dos bordes del Sol o a los dos bordes de la Luna es aproximadamente el mismo. Y este hecho se corrobora en los eclipses solares cuando la Luna se interpone entre el Sol y nosotros.

Figura 2.19 El disco del Sol y el disco de la Luna son en apariencia del mismo tamaño, vistos desde la Tierra, como se demuestra en los eclipses solares.



Además, Aristarco de Samos para medir la distancia a la Luna, se basó en los datos de los otros eclipses: los lunares. En ellos, la Luna entra en el cono de sombra que produce la Tierra.

Figura 2.20 Esquema de un eclipse lunar en el que no se mantienen las proporciones.



El cono de sombra se crea porque el Sol se ve como un disco en el cielo y entonces la luz que de él nos llega tiene pequeñas variaciones en el ángulo. El cono de sombra es donde toda su luz queda bloqueada. Se puede observar el cono de sombra que produce un dedo en un día soleado alejándolo del piso hasta una altura de un metro o más. Muy cerca del suelo

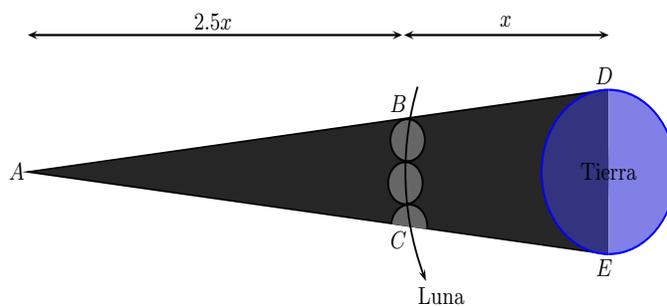
se forma una sombra con bordes bien definidos. Pero a medida que lo alejamos, sus bordes se vuelven difusos o desenfocados: son la penumbra, donde si bien parte de la luz que viene del Sol se bloquea, algo de ella pasa. El cono de sombra del dedo, o de una moneda, es “semejante” al cono de sombra de la Tierra, y entonces lo podemos medir.

Los griegos calcularon en forma experimental que el cono de sombra del Sol tiene una proporción aproximada de altura a base de 108 a 1, es decir, la longitud del cono de sombra es aproximadamente 108 veces el diámetro de la Tierra. De aquí, considerando que Sol y Luna tienen el mismo diámetro aparente, se obtiene que si d_{Luna} denota al diámetro de la Luna, entonces:

$$x = 108 \cdot d_{Luna} \tag{3}$$

es la distancia de la Tierra a la Luna.

En un eclipse lunar, la Luna entra primero a la zona de penumbra y luego, al cono de sombra de la Tierra. Por el tiempo que tarda la Luna en cruzar este cono, se puede estimar que la Luna cabe más o menos 2.5 veces en el cono de sombra, es decir, en el lugar donde la Luna cruza al cono, este mide 2.5 veces el diámetro de la Luna. Con ello ya juntamos las mediciones necesarias para poder determinar el diámetro de la Luna y la distancia que tiene de nosotros; cabe remarcar que todas estas mediciones se obtuvieron a partir de observaciones realizadas desde la Tierra.



<Imagen 21, hacer una similar, guardando las proporciones>

Figura 2.21 Esquema de un eclipse lunar.

En la figura 2.21 tenemos dos triángulos isósceles semejantes: el cono de sombra $\triangle ADE$ es semejante al triángulo $\triangle ABC$. Por otro lado, el triángulo $\triangle ABC$ tiene base $2.5 \cdot d_{Luna}$ y, por ello, la altura —horizontal en este caso— del triángulo $\triangle ABC$ es $2.5x$. Por lo tanto, la altura del triángulo $\triangle ADE$ es $2.5x + x = 3.5x$. Como sabíamos que esta altura es 108 veces el diámetro de la Tierra, entonces:

$$3.5x = 108 \cdot d_{Tierra}$$

de donde podemos despejar x , que es la distancia del centro de la Tierra al de la Luna:

$$x = \frac{108}{3.5} d_{Tierra}$$

También obtenemos que el diámetro de la Luna es $d_{Luna} = \frac{1}{3.5} d_{Tierra}$, por (3).

Es claro que en los métodos que acabamos de describir hay una considerable posibilidad de error en las mediciones. Los estimados básicos —108, 2.5 y que los diámetros aparentes de Luna y Sol coinciden—, vienen de Aristarco. Si en vez del diámetro de la Tierra que él usó, empleamos el que se estima en la actualidad —de $d_{Tierra} = 13\,000\text{ km}$ —, nuestras fórmulas darían 401 142 km para la distancia de la Tierra a la Luna, y 3 714 km para el diámetro lunar. Los estimados actuales son 384 403 km y 3 474 km respectivamente, que dan errores del 4% y el 6% para el método de Aristarco. Esto demuestra el poder de la geometría

euclidiana elemental y la increíble precisión con que los griegos hicieron sus mediciones, así como el enorme poder del razonamiento abstracto.

2.8 LA MEDICIÓN DE LA TIERRA

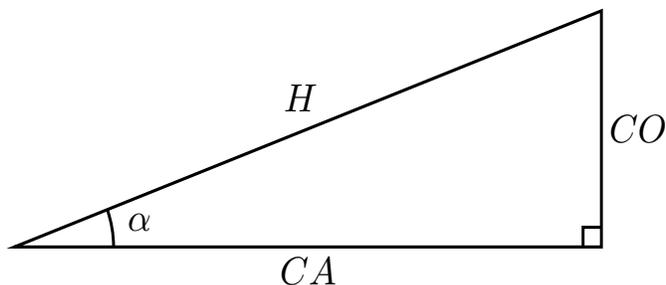


Figura 2.22 Un billete de diez marcos alemanes —antes de la introducción del euro— muestra a Carl Friedrich Gauss, matemático alemán. Gauss fue tal vez el matemático más importante de todos los tiempos; a él se deben numerosos desarrollos en las matemáticas que hizo para fines prácticos. Por ejemplo, inventó el heliógrafo, un instrumento para aumentar la precisión en las mediciones que dirigió en el país de Hannover.

A principios del siglo XIX no se sabía cuál era la montaña más alta del planeta ni se conocía el tamaño de la Tierra con precisión. La medida conocida como **metro** la establecieron los franceses con la idea de que fuera la 10 000-ésima parte de la distancia del Ecuador al Polo norte. Casi al mismo tiempo, los ingleses iniciaron el *Great Trigonometrical Survey* —Gran proyecto de topografía trigonométrica— para medir con precisión toda la India, colonia británica en aquel entonces, pues se pretendía tener una visión más realista del territorio que ocupaba el imperio Británico en Asia y, a la vez, tener una medida más precisa del tamaño de la Tierra. Este proyecto consumió grandes recursos, duró casi todo el siglo XIX y permitió establecer la altura del monte Everest, nombrado así en honor al coronel que estuvo a cargo del proyecto —aunque finalmente ya no vio la montaña pues se quedó ciego. Fue hasta ese momento que se supo que el Everest, a 8 850 metros sobre el nivel del mar, era más alto que los Andes.

En esta sección veremos los principios básicos del proceso para medir la superficie terrestre conocido como “triangulación” —basado en la medición y el cálculo de triángulos, por ejemplo, entre tres picos de montañas. Claro que estos triángulos, en general, no serán ni **rectángulos** —que tienen un ángulo de 90° —, ni isósceles —con dos o tres lados iguales—, sino que son triángulos arbitrarios. Para aclarar la relación que existe entre los lados y los ángulos, tenemos que empezar primero con triángulos rectángulos.

En un triángulo rectángulo, a los dos lados que forman el ángulo recto se les llama **catetos** y al lado opuesto, **hipotenusa**. Los ángulos de un triángulo rectángulo quedan determinados por cualquiera de sus ángulos no rectos, pues entonces, la medida del otro ángulo corresponde a lo que falta para 90° —recordemos que los tres ángulos internos de un triángulo suman 180° . Si elegimos uno de los ángulos no rectos y lo llamamos α , podemos diferenciar a los dos catetos. Al lado que forma α con la hipotenusa, se le llama **cateto adyacente** y se le denota CA ; al lado opuesto al ángulo α , **cateto opuesto** y se le denota CO y, finalmente, a la hipotenusa con H .



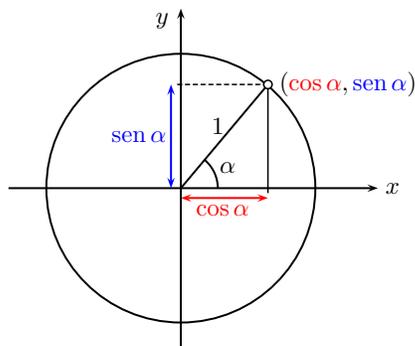
<Imagen 23, hacer una idéntica>

Figura 2.23 Triángulo rectángulo.

Las **funciones trigonométricas** son las proporciones entre los lados de los triángulos rectángulos. Las básicas son el **coseno** y el **seno**, definidas como:

$$\cos \alpha = \frac{CA}{H}, \quad \text{sen } \alpha = \frac{CO}{H}$$

Observemos que cuando la hipotenusa mide 1 —que es igual a decir que $H = 1$ —, el coseno y el seno son, precisamente, lo que miden los catetos. Así, podemos pensar que son las coordenadas cartesianas de un punto en el círculo unitario, es decir, el círculo de radio 1 con centro en el origen. En esta manera de ver al coseno y al seno, los puntos del círculo unitario quedan parametrizados por el ángulo con el eje de las x , como se muestra en la figura 2.24.



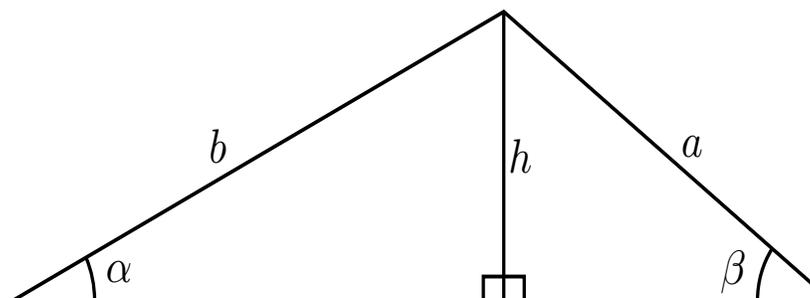
<Imagen 24, hacer una idéntica>

Figura 2.24 Las coordenadas de un punto en el círculo unitario son el coseno y el seno del ángulo correspondiente.

Al pensar en el seno y el coseno como coordenadas, es natural extender su definición a cualquier ángulo con los signos correspondientes a los de los cuatro cuadrantes.

Planteemos ahora un problema más complicado: medir la altura de una montaña sin subir en ella. No podemos usar triángulos rectángulos pues no podemos acceder al interior de la montaña, pero sí se pueden hacer mediciones suficientes desde afuera para determinar un triángulo del que se conozcan dos de sus ángulos y un lado. Luego, utilizando la llamada **ley de los senos**, que veremos a continuación, se pueden calcular los otros dos lados.

En un triángulo cualquiera, tracemos una de sus alturas y llamémosla h ; a los ángulos opuestos a ella, denotémoslos α y β , mientras que los lados opuestos —en el triángulo original—, serán a y b , respectivamente.



<Imagen 25, hacer una idéntica>

Figura 2.25 Un triángulo con una altura h .

El seno de α se puede expresar como $\text{sen } \alpha = \frac{h}{b}$ de donde podemos despejar h :

$$h = b \text{ sen } \alpha$$

De manera análoga, obtenemos que $h = a \text{ sen } \beta$. Hemos descrito a h de dos maneras diferentes, lo que nos da la igualdad: $b \text{ sen } \alpha = a \text{ sen } \beta$, que también se puede escribir como:

$$\frac{\text{sen } \alpha}{a} = \frac{\text{sen } \beta}{b}$$

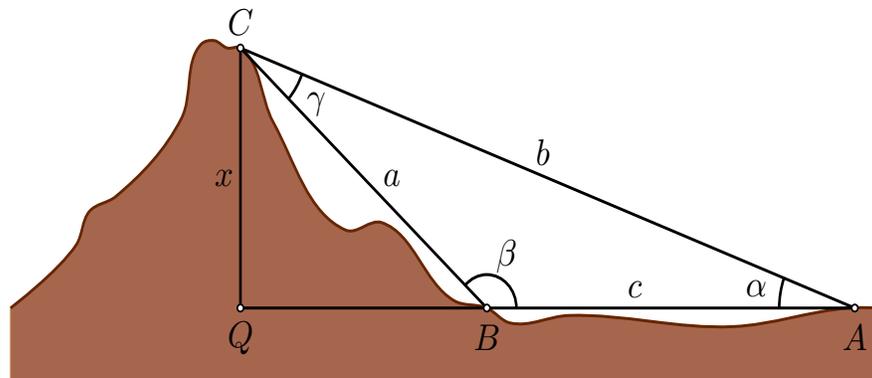
Esta ecuación es el principio básico de la ley de los senos, aunque la manera más común de presentarla es igualando los inversos e incluyendo al tercer ángulo γ y a su lado opuesto, c . Para hacerlo, hay que realizar el mismo razonamiento con alguna de las otras dos alturas. De tal manera, para cualquier triángulo se cumple el que la proporción de los lados con respecto a los senos de sus ángulos opuestos sea la misma. Al expresar la ley de los senos en forma de ecuaciones tendríamos:

$$\frac{a}{\text{sen } \alpha} = \frac{b}{\text{sen } \beta} = \frac{c}{\text{sen } \gamma}$$

Regresemos ahora al problema de la montaña para aplicar la ley de los senos. Llamemos C a un punto del pico visible. Desde dos puntos A y B en la falda, se pueden medir los ángulos correspondientes α y β del triángulo $\triangle ABC$. Supongamos que también conocemos la distancia c entre A y B .

<Imagen 26,
hacer una similar,
guardando las
proporciones>

Figura 2.26 Triángulos para medir la altura de una montaña.



Por la ley de los senos, se obtienen las distancias a y b :

$$a = \frac{\text{sen } \alpha}{\text{sen } \gamma} \cdot c \quad \text{y} \quad b = \frac{\text{sen } \beta}{\text{sen } \gamma} \cdot c$$

donde γ es el ángulo en C , que se obtiene a partir de α y β , pues los tres suman 180° . Conociendo b —y suponiendo que A y B están al mismo nivel—, podemos calcular la altura de C usando la definición del seno:

$$x = b \text{ sen } \alpha.$$

Con las mismas ideas que hemos considerado en esta sección, funciona el método de triangulación: primero, se mide una longitud con mucha precisión llamada “base”; después se miden los ángulos hacia un punto muy visible. Al usar la ley de los senos se pueden calcular las distancias a los extremos de la base. Como las herramientas de medición permiten

medir el ángulo de elevación por separado del ángulo de giro, también es posible calcular la altura del punto visible. Después, el equipo se transporta a este punto y divisa la base y otros puntos nuevos. Así, la medición avanza por el terreno midiendo la forma de manera precisa. Gauss condujo así la medición del Hannover y a finales del siglo XVIII, con este método, se determinó que el Popocatepetl no era la montaña más alta de México, sino el Pico de Orizaba.

2.9 LA PIRÁMIDE TRUNCADA

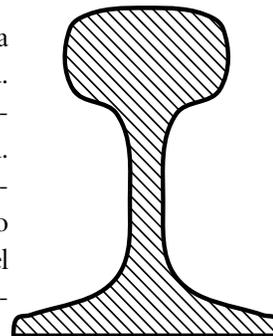


<Imagen 27, fotografía similar, se debe ver toda la pirámide>

Figura 2.27 Las pirámides de Gizeh son unos de los monumentos más impresionantes de las civilizaciones de la Antigüedad. Su forma geométrica reúne tanto estabilidad como belleza.

El ámbito de la producción industrial a principios del siglo XIX demandaba herramientas de medición cada vez más sofisticadas. Entre los hombres que trataron de cumplir esta demanda se encontraba el matemático suizo, Jakob Amsler-Laffon, quien fundó su propia fábrica de instrumentos de medición. Uno de sus más apreciados inventos fue el “integrador”: un arreglo de varas, ruedas y puntas que medía cualquier área como, por ejemplo, la sección transversal de un riel de hierro mostrada en la figura 2.28. Para medir el área se tenía que colocar el integrador sobre la hoja de dibujo y, luego, pasar una punta por el borde de la figura. Una rueda integrada al aparato registraba continuamente el área, que podía leerse al final sin problemas.

Antes de esta invención era sumamente difícil obtener una buena aproximación para el área de una figura así de compleja. Los instrumentos de Amsler se basaban de manera crucial en desarrollos matemáticos, entre ellos, el cálculo diferencial e integral. La importancia de estas herramientas no debe subestimarse: fueron fundamentales para la industria, en particular, la del acero —una buena aproximación del área de la sección transversal del riel proporciona una estimación certera del costo del material, según su peso por metro.



<Imagen 28, hacer algo similar>

Figura 2.28 Dibujo de una sección transversal de un riel.

Es interesante ver que fue, justamente, la insistencia para calcular áreas y volúmenes de figuras —como la circunferencia— o cuerpos —como la esfera o la pirámide— lo que permitió desarrollar, poco a poco, una teoría más general que finalmente aportó herramientas matemáticas aplicables a situaciones de la vida cotidiana, como el cálculo del área de un riel

que ya hemos revisado. En esta sección y en las tres subsecuentes, se exponen varios de estos cálculos con todo detalle pues, así, se logra comprender mejor las bases del cálculo integral.

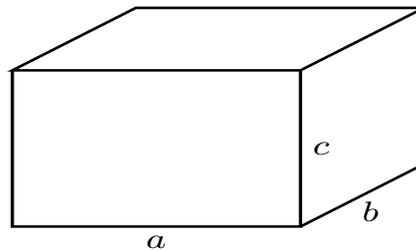
Tratar de calcular el área de figuras y el volumen de cuerpos ha llamado la atención a muchas personas desde la Antigüedad. El ejemplo más famoso es, sin duda, el del área y la circunferencia de un círculo. Después de revisarlo en esta sección, sabremos cómo se puede calcular el volumen y la superficie de otros cuerpos geométricos y, llegaremos a ver los fundamentos del cálculo integral en la sección 2.12.

En particular, en esta sección se verán diferentes fórmulas que expresan los volúmenes de cuerpos como prismas y pirámides. Estos cuerpos geométricos sencillos están delimitados por **polígonos** o figuras planas que, a la vez, están delimitados por segmentos rectos. Después, se usará esta información para determinar el volumen de una **pirámide truncada** —a la cual se le cortó un pedazo de la punta—. Además, se pondrá un énfasis particular en la lectura de tales fórmulas para explicar cómo las propiedades de éstas se reflejan en propiedades geométricas.

Todos sabemos calcular el volumen de una caja con lados a , b y c : $V = abc$, es decir, el producto de los tres lados que terminan en el vértice de la caja.

<Imagen 29, hacer una similar, guardando las proporciones>

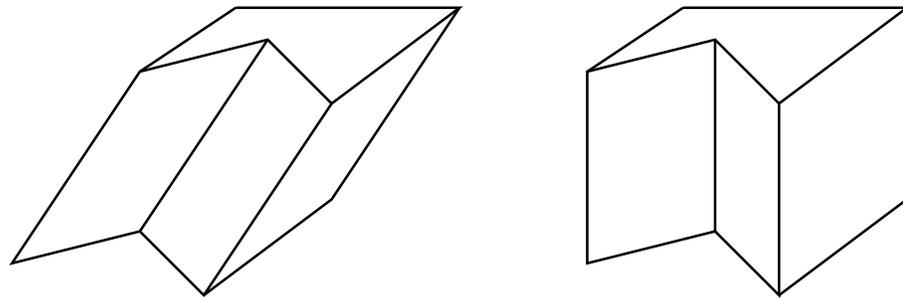
Figura 2.29 Una caja que, en matemáticas, tiene el nombre rimbombante de paralelepípedo rectangular.



La caja tiene un nombre matemático horrible, se llama **paralelepípedo rectangular** y es un caso particular de un **prisma**. Los prismas se obtienen al trasladar un polígono en una dirección como se muestra en el siguiente dibujo.

<Imagen 30, hacer una similar, guardando las proporciones >

Figura 2.30 Dos prismas con bases congruentes y alturas iguales.

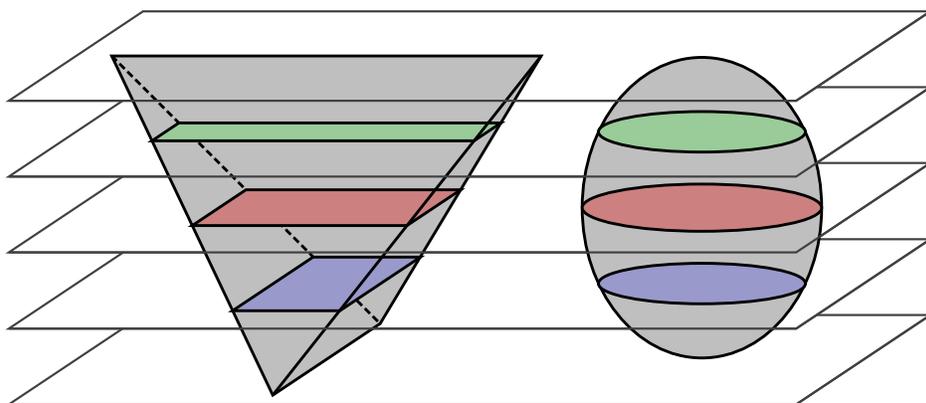


Del lado derecho vemos un prisma vertical. La fórmula del volumen de un prisma es sencilla y se parece mucho a la del rectángulo:

$$\text{volumen de prisma} = \text{base} \times \text{altura}$$

sólo que aquí, “base” significa el área que se traslada y la “altura” es la distancia entre la tapa superior y la inferior. Cuidado: si al generar el prisma la base no se trasladó en dirección vertical, la altura no será la longitud de lo que se trasladó, sino menos.

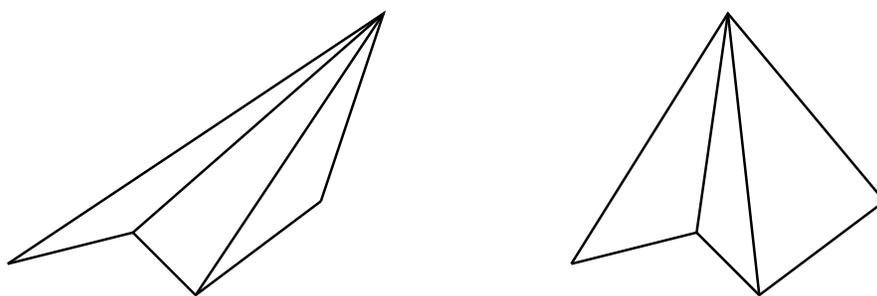
Se debe observar que la dirección del traslado no importa; siempre y cuando se tenga la misma altura, se tendrán los mismos volúmenes. Esto se debe al **principio de Cavalieri**: dos cuerpos que tienen cortes de áreas iguales con cada plano paralelo a la superficie sobre las que están puestos, poseen volúmenes iguales.



<Imagen 31, hacer algo similar, guardando las proporciones>

Figura 2.31 Ilustración del principio de Cavalieri: los dos cuerpos tienen el mismo volumen dado que sus áreas de intersección con cada plano paralelo son iguales.

¿Cuál será entonces el volumen de una pirámide? El cuerpo se construye uniendo cada punto de la base con un punto, que es la punta de la pirámide.



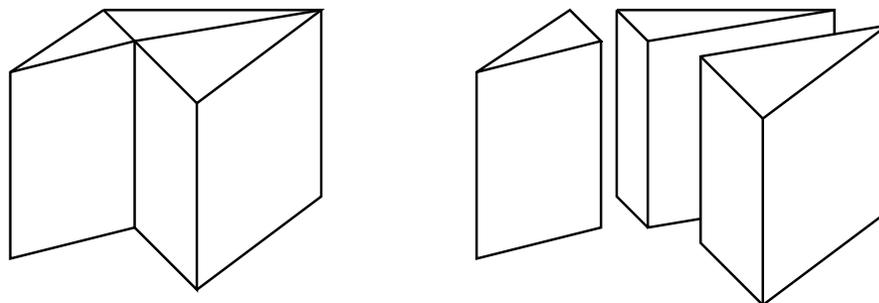
<Imagen 32, hacer algo similar, guardando las proporciones>

Figura 2.32 Dos pirámides con bases congruentes y alturas iguales.

La pirámide se relaciona con el prisma como lo hace el triángulo con el paralelogramo. Se habla de una **analogía**, es decir, una similitud entre dos relaciones en diferentes ámbitos. Las analogías son un fuerte motor de ideas, no sólo en matemáticas, sino también en el lenguaje, por ejemplo, donde se usan para aclarar una cierta relación.

Por esta analogía, podríamos pensar que el volumen de una pirámide se calcula de manera parecida a como se obtiene el área de un triángulo: base por altura entre dos. Sin embargo, con las analogías hay que tener cuidado pues no siempre todo se traduce uno a uno de un lado al otro, la relación es usualmente mas complicada.

Cada prisma —o pirámide— se puede dividir en prismas —o pirámides, respectivamente— con bases triangulares y la misma altura, como se observa en la figura 2.33.



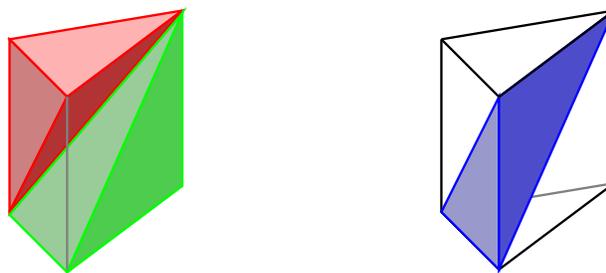
<Imagen 33, hacer algo similar, guardando las proporciones>

Figura 2.33 Partición de un prisma cualquiera en prismas triangulares.

Por lo anterior, basta entender cómo se calcula el volumen de una pirámide con base triangular. Es fácil colocar dos pirámides con la misma base y altura en un prisma que también, tiene la misma base y altura como lo muestra la siguiente figura.

<Imagen 34, hacer algo similar, guardando las proporciones>

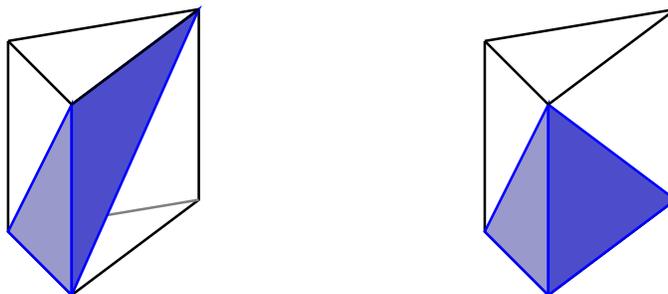
Figura 2.34 Partición de un prisma triangular en tres pirámides triangulares.



Como se observa, dos pirámides no son suficientes para llenar un prisma, así que la primera idea crecida sobre la tierra de la analogía no puede ser correcta. No obstante, el espacio sobrante es, a la vez, una pirámide —la base puede ser, por ejemplo, el costado vertical derecho y la altura la vemos entonces acostada en dirección horizontal.

<Imagen 35, hacer algo similar, guardando las proporciones>

Figura 2.35 Aplicación del principio de Cavalieri.



Por el mismo principio de Cavalieri, podemos mover hacia abajo la punta, que se encuentra arriba en la tapa, ya que no modificamos ni la base ni la altura. Aquí con “la base de la pirámide” nos referimos a la cara azul claro de la figura 2.35 y con “la altura”, a la perpendicular —recordemos que ninguna de estas medidas depende de la dirección—. Obtenemos, entonces, que el resto también tiene el mismo volumen que buscamos. En resumen: dentro del prisma caben tres pirámides y cada una tiene el volumen de una pirámide con la misma altura y base que el prisma. Concluimos entonces que:

$$\text{volumen de pirámide} = \frac{\text{base} \times \text{altura}}{3}.$$

Ahora, podemos revisar la analogía de nuevo y descubrimos que sí hay una relación muy asombrosa: el 2 en el denominador se cambió a un 3 al pasar del triángulo a la pirámide. Estos números se explican en términos de la dimensión de las figuras: el triángulo es plano y tiene dimensión dos, mientras que la pirámide tiene volumen y, por ello, dimensión tres.

Recordemos que se dividió la base de cualquier pirámide en triángulos. Ahora queremos ver si podemos deducir la fórmula del volumen de cualquier pirámide. En efecto, si la base original B se dividió en t partes triangulares — B_1, B_2, \dots, B_t — entonces se tiene que:

$$B = B_1 + B_2 + \dots + B_t,$$

lo que expresa que el área total de la base se obtiene al sumar las áreas de cada uno de los t triángulos. En forma similar:

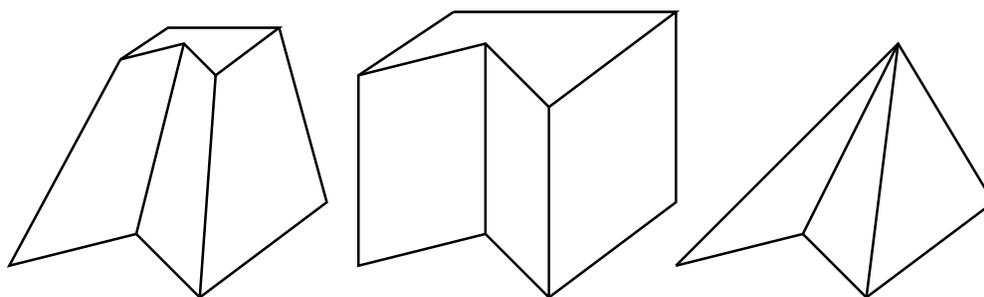
$$V = V_1 + V_2 + \dots + V_t$$

Ahora podemos sustituir cada uno de los volúmenes V_1, V_2, \dots, V_t por $\frac{B_1 \cdot h}{3}, \frac{B_2 \cdot h}{3}, \dots, \frac{B_t \cdot h}{3}$. Así, se obtiene que:

$$\begin{aligned}
 V &= \frac{B_1 \cdot h}{3} + \frac{B_2 \cdot h}{3} + \dots + \frac{B_t \cdot h}{3} \\
 &= \frac{(B_1 + B_2 + \dots + B_t) \cdot h}{3} \\
 &= \frac{B \cdot h}{3}
 \end{aligned}$$

al usar la factorización en la segunda ecuación. En resumen, la fórmula que obtuvimos para pirámides con base triangular vale, en general, para cualquier pirámide.

Ahora se tienen ya conocimientos suficientes para atacar la fórmula del volumen de una pirámide truncada. Antes de entrar en cálculos y argumentaciones involucradas, observemos que la fórmula que buscamos debe “contener” la del prisma y la pirámide, como “casos particulares”.



<Imagen 36, hacer algo similar, guardando las proporciones>

Figura 2.36 El prisma y la pirámide son casos particulares de la pirámide truncada.

Si la base y la tapa son iguales, se trata de un prisma mientras que, si la tapa se contrae hacia un punto, resulta una pirámide. La fórmula que buscamos debe entonces “generalizar” las dos fórmulas anteriores y unirlas en una sola. Además, debe ser **simétrica** en la base y la tapa, es decir, si volteamos la pirámide cortada se intercambian la base y la tapa, pero el volumen no cambia; por ello, dicho intercambio no debe alterar la fórmula.

Para simplificar los cálculos que haremos a continuación, supondremos ahora que la pirámide truncada tiene como base un cuadrado de lado a , como tapa, un cuadrado de lado b y la altura, será denotada siempre por h .

Si la pirámide trunca se completa, obtenemos una altura H . El volumen es de $\frac{a^2 \cdot H}{3}$, dado que la base tiene área a^2 . Lo que se cortó al truncar es, también, una pirámide con base de área b^2 y altura $H - h$. De esta manera, la primera fórmula sería:

$$V = \frac{a^2 \cdot H}{3} - \frac{b^2 \cdot (H - h)}{3} = \frac{a^2 - b^2}{3} H + \frac{b^2}{3} h \tag{4}$$

La altura H se obtiene por semejanza con:

$$\frac{H}{a} = \frac{H - h}{b}$$

de donde, al despejar tenemos que:

$$H = \frac{a}{a - b} h$$

Si sustituimos H por la expresión $\frac{a}{a - b} h$ en la fórmula (4):

$$V = \frac{a^2 - b^2}{3} \frac{a}{a - b} h + \frac{b^2}{3} h ,$$

lo cual puede simplificarse en:

$$V = \frac{a^2 + ab + b^2}{3} h \quad (5)$$

Ahora, si $a = b$ resulta que $a^2 + ab + b^2 = 3a^2$ y entonces, $V = a^2 \cdot h$ como lo esperábamos. Si, en cambio $b = 0$, entonces $V = \frac{a^2 \cdot h}{3}$ que también es lo que esperábamos. Se puede observar la simetría entre a y b : si intercambiamos estas dos variables obtenemos $V = \frac{b^2 + ba + a^2}{3} h$, lo mismo que se obtuvo en (5).

Si bien trabajamos con cuadrados para la base y la tapa, ¿cómo habría que generalizar esta última fórmula a una pirámide truncada general? Primero, observemos que $a^2 = B$ es el área de la base y $b^2 = T$ es el área de la tapa. Pero, ¿cómo habría que interpretar el término ab ? Como $a = \sqrt{B}$ y $b = \sqrt{T}$, entonces $ab = \sqrt{BT}$ y a partir de ahí, encontramos que:

$$\text{volumen de una pirámide} = \frac{B + \sqrt{BT} + T}{3} \cdot h,$$

aun cuando la base y la tapa tengan una forma distinta. Todo este razonamiento se hizo a partir del principio de Cavalieri que se analizará en la sección 2.11, con mayor detalle.

2.10 EL NÚMERO π Y LA CUADRATURA DEL CÍRCULO

<Imagen 37, hacer algo idéntico>

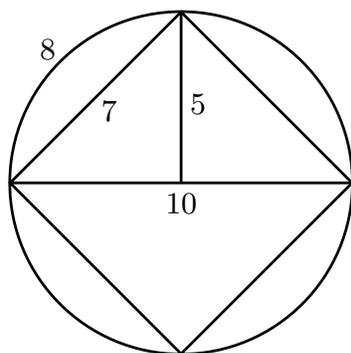
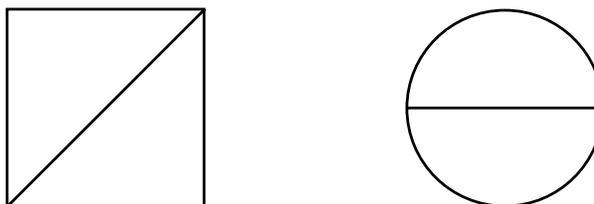


Figura 2.37 Edwin J. Goodwin, un aficionado a las matemáticas, propuso fijar tanto el valor de la raíz cuadrada de 2 como el valor de π —sus aproximaciones eran de 3.2 para π y de 1.4 para $\sqrt{2}$ —. La propuesta fue aceptada por unanimidad por The House of Representatives del estado de Indiana, en 1897. Éste es, tal vez, el intento más célebre de establecer una certeza científica por medio de la ley. El Senado de Indiana, advertido de la falsedad, pospuso la decisión indefinidamente.

Civilizaciones tan antiguas como la babilónica, la egipcia, la china y la hindú, reconocieron que había ciertas relaciones entre las dimensiones de algunas figuras geométricas que se mantenían constantes aunque su tamaño variara, es decir, que eran independientes del tamaño de la figura.

<Imagen 38, hacer algo similar, guardando las proporciones>

Figura 2.38 Un cuadrado con una de sus diagonales y un círculo con uno de sus diámetros.



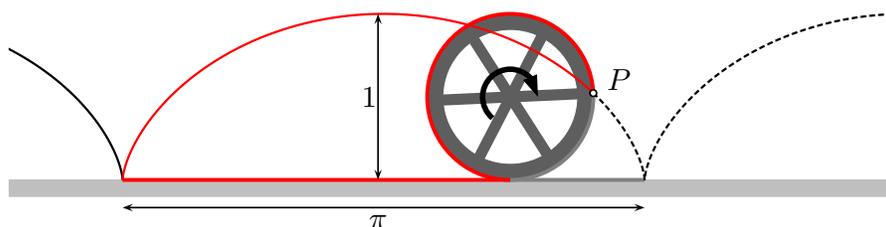
Por ejemplo, la diagonal de un cuadrado mide un poco menos que una y media veces su lado, independientemente de si el cuadrado es grande o pequeño. También el perímetro de una circunferencia es poco más que tres veces su diámetro, sin importar si la circunferencia es grande o minúscula como una retina. Se trata de relaciones, razones o proporciones fijas—que no se alteran aunque varíe el tamaño de las figuras—, que sólo dependen de su for-

ma. Estas relaciones permiten calcular, por ejemplo, la diagonal de una plaza cuadrada cuyos lados miden 80 metros. A través de sucesivas mediciones podríamos comprobar que la relación entre la diagonal y el lado de un cuadrado es mayor que $1 + \frac{3}{8}$ y menor que $1 + \frac{1}{2}$. Por lo tanto, la diagonal de la plaza sería mayor que 110 metros y menor que 120 metros. Pero... ¿cuánto mide exactamente la diagonal de la plaza? Para responderlo, tendríamos que saber cuántas veces es mayor la diagonal de un cuadrado que su lado, es decir, tendríamos que conocer con exactitud la relación entre estas dos dimensiones.

El concepto de número que tenía el hombre de la Antigüedad se limitaba a las fracciones —cocientes de números enteros—, demasiado primitivas para representar cantidades —como los números irracionales— que no pueden ser expresadas mediante fracciones. En particular, fueron los pitagóricos quienes, desde entonces, descubrieron que esta razón o proporción entre la diagonal y el lado de un cuadrado no puede expresarse mediante una fracción, así que no tuvieron otro remedio que hacerlo con algún símbolo y estimar su valor usando fracciones. Hoy en día representamos dicha relación como $\sqrt{2}$ y sabemos que $1.414213 < \sqrt{2} < 1.414214$.

De manera análoga, la relación entre el perímetro de una circunferencia y su diámetro, tampoco puede expresarse con fracciones. De hecho, como veremos más adelante, se trata de una relación aún más compleja. Las civilizaciones antiguas intentaron expresarla a partir de fracciones cada vez más exactas, pero eran sólo meras aproximaciones de algo cuya definición, como número, escapaba al lenguaje matemático de la época. No fue sino hasta el esplendor de la civilización helénica cuando se llegó a un concepto y una definición precisas de dicha relación, razón o proporción, que hoy llamamos π e identificamos con los famosos dígitos 3.1416.

Los babilonios usaron durante un tiempo al 3 como una aproximación práctica; más adelante y al mejorarla, adoptaron el valor $3 + \frac{1}{8}$, equivalente a 3.125. El papiro Rihnd —que data de 1650 a.C.— incluye una aproximación aún más cercana, equivalente a 3.16049. Estas estimaciones eran el resultado de cálculos de perímetros de figuras formadas por segmentos —más o menos parecidas al círculo— y que representaban valores útiles; sin embargo, no respondían a un concepto matemático bien definido.



<Imagen
39, hacer algo
idéntico>

Figura 2.39 π como el recorrido de una rueda de diámetro unitario.

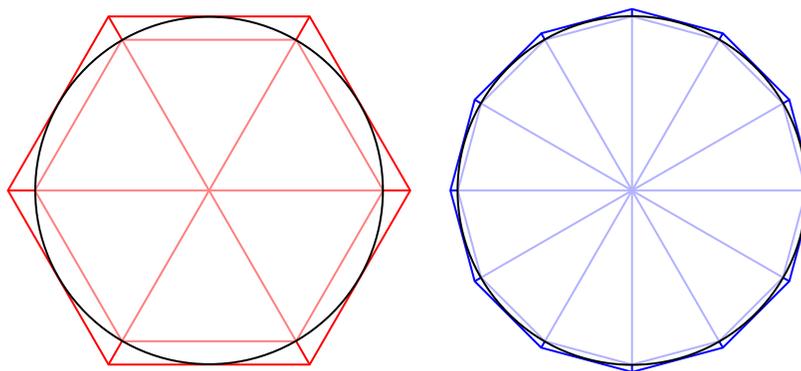
La definición correcta y el primer cálculo de esta relación como un resultado matemático se deben a Arquímedes de Siracusa —siglo III a.C.—. Él demuestra, apoyado en los conocimientos geométricos de la época que, efectivamente, la relación entre el perímetro y el diámetro de la circunferencia es independiente del tamaño de la misma —y se mantiene aunque el tamaño cambie—, y en seguida procede a realizar un cálculo riguroso de la relación demostrando que se encuentra entre $3 + \frac{10}{71}$ y $3 + \frac{1}{7}$. El uso de la letra π —inicial que denota perímetro en griego— es mucho más reciente y proviene del matemático galés William Jones, quien la usó por primera vez en 1706. Esta notación se popularizó luego a través de los trabajos de Leonhard Euler.

Lo relevante del trabajo de Arquímedes no son los valores concretos que obtuvo, sino el método que inventó para ello y con el que abrió la posibilidad de obtener el valor de π con tanta precisión como se desee, aunque el proceso pueda ser lento por la dificultad de los cálculos requeridos. Es importante tomar en cuenta que, para obtener su resultado, Arquímedes tuvo que realizar varias estimaciones de fracciones y raíces cuadradas —cuya expresión decimal era entonces desconocida— en forma de laboriosas desigualdades entre fracciones. Actualmente, el método de Arquímedes para estimar π puede utilizarse perfectamente en una computadora para obtener estimaciones bastante más precisas que las del célebre siracusano.

El método de Arquímedes para definir y calcular π consiste en comparar el perímetro de la circunferencia con los de dos polígonos regulares, uno inscrito y otro circunscrito a ella, como se muestra en la figura.

<Imagen 40, hacer algo idéntico, guardando las proporciones>

Figura 2.40
Aproximaciones del perímetro de una circunferencia mediante los perímetros de polígonos regulares, de 6 y 12 lados, inscritos y circunscritos.



Al utilizar resultados de la geometría euclidiana, Arquímedes demuestra que el perímetro de la circunferencia debe ser mayor que el del polígono inscrito y menor que el del circunscrito. También demuestra que la diferencia entre ambos puede hacerse tan pequeña como se desee al considerar polígonos con un número muy grande de lados. Estas ideas contienen no sólo el germen del concepto de límite utilizado en el cálculo moderno sino, también, la manera de tratarlo con absoluto rigor lógico.

Analicemos una circunferencia de radio 1 y consideremos π como la mitad de su perímetro. Entonces, el lado del hexágono inscrito es $L = 1$ y, por el teorema de Pitágoras, su apotema —la distancia del centro a cada lado— es $A = \sqrt{1 - (\frac{1}{2})^2}$. Por semejanza de triángulos, el lado del hexágono circunscrito mide $\frac{1}{A} = \frac{1}{A}$. De ahí se obtienen las siguientes desigualdades:

$$3 < \pi < 3 \cdot \frac{1}{A}$$

<Imagen 41, hacer algo idéntico, habrá que ponerle ciertas etiquetas de fórmulas como A, 1/2, etc.>

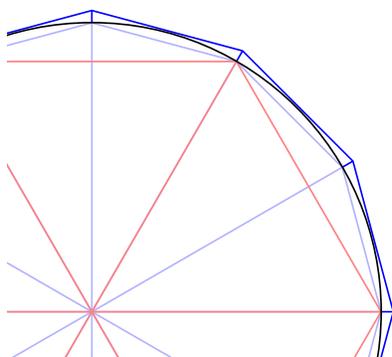


Figura 2.41 Refinamiento de la aproximación pasando de un polígono de 6 lados a uno de 12, es decir, de un hexágono a un dodecágono.

Si ahora se consideran los dodecágonos regulares —inscrito y circunscrito— y se usa nuevamente el teorema de Pitágoras, resulta que el lado del dodecágono inscrito es $L_1 = \sqrt{(\frac{1}{2})^2 + (1 - A)^2}$, mientras que su apotema es $A_1 = \sqrt{1 - (\frac{L_1}{2})^2}$; por semejanza de triángulos se obtiene que el lado del dodecágono circunscrito es $\frac{L_1}{A_1}$. Al sumar los lados, se obtienen las siguientes desigualdades:

$$3 \cdot 2 \cdot L_1 < \pi < 3 \cdot 2 \cdot \frac{L_1}{A_1}$$

Si se siguen construyendo polígonos regulares de 24, 48, 96, ... lados, y cada vez se divide el ángulo interno entre dos, se tienen las desigualdades:

$$3 \cdot 2^n \cdot L_n < \pi < 3 \cdot 2^n \cdot \frac{L_n}{A_n}$$

donde L_n es el lado del polígono regular con $6 \cdot 2^n$ lados y $A_n = \sqrt{1 - (\frac{L_n}{2})^2}$ es su apotema. El valor de L_n se puede obtener recursivamente —que quiere decir “repetir indefinidamente la aplicación de”— mediante la fórmula $L_{n+1} = \sqrt{(\frac{L_n}{2})^2 + (1 - A_n)^2}$, consecuencia directa del teorema de Pitágoras. Estas fórmulas ofrecen un algoritmo para calcular aproximaciones de π con cualquier grado de precisión. La siguiente tabla muestra las desigualdades —que se obtienen al usar este procedimiento— expresadas con 6 decimales:

Número de lados de los polígonos	Estimación de π
$6 = 6 \cdot 2^0$	$3.000000 < \pi < 3.464102$
$12 = 6 \cdot 2^1$	$3.105828 < \pi < 3.215391$
$24 = 6 \cdot 2^2$	$3.132628 < \pi < 3.159660$
$48 = 6 \cdot 2^3$	$3.139350 < \pi < 3.146087$
$96 = 6 \cdot 2^4$	$3.141031 < \pi < 3.142715$
$192 = 6 \cdot 2^5$	$3.141452 < \pi < 3.141874$
$384 = 6 \cdot 2^6$	$3.141557 < \pi < 3.141663$
$768 = 6 \cdot 2^7$	$3.141583 < \pi < 3.141611$
$1536 = 6 \cdot 2^8$	$3.141590 < \pi < 3.141598$
$3072 = 6 \cdot 2^9$	$3.141592 < \pi < 3.141594$

Las desigualdades correspondientes al polígono de 96 lados son ligeramente mejores a las que obtuvo Arquímedes con el mismo polígono: $3 + \frac{10}{71} < \pi < 3 + \frac{1}{7}$ —aproximadamente $3.140845 < \pi < 3.142857$, con precisión de seis decimales. La aproximación que se obtiene con el polígono de 3072 lados corresponde a la que obtuvo en el año 263 d.C. el matemático chino Lui Hui.

Esfuerzos sucesivos en el cálculo han alcanzado más y más precisión. Primero fueron manuales y después, se hicieron usando computadoras cada vez más potentes. En la actualidad, el récord en el cálculo de dígitos lo tiene Daisuke Takahashi con más de dos y medio billones de cifras decimales. Cabe mencionar que estos esfuerzos no tienen ninguna importancia respecto al significado de π , sólo se hacen por el prestigio que da a los programadores y constructores de computadoras el llegar a ellos.

En cambio, es muy interesante saber que el número π aparece en una gran cantidad de resultados matemáticos, aparentemente, sin relación alguna con la circunferencia. Por ejemplo, se sabe que:

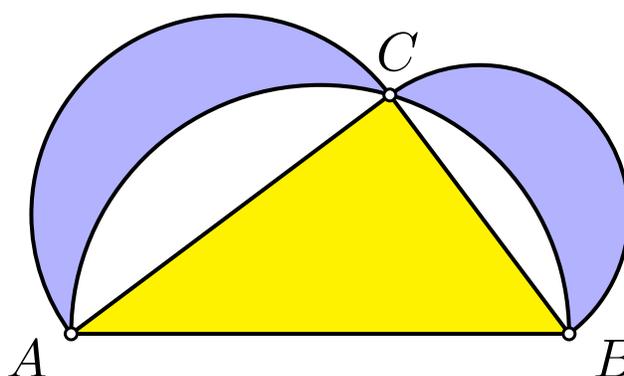
$$\begin{aligned} \frac{\pi}{4} &= \frac{1}{1} - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \dots \\ \frac{\pi^2}{6} &= \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \dots \\ \sqrt{\pi} &= \int_{-\infty}^{\infty} e^{-x^2} dx \end{aligned}$$

Arquímedes no sólo define y calcula π sino que usa su valor para expresar el área de la circunferencia y, también, el de volúmenes y áreas de conos, cilindros y esferas. Se trata de los primeros resultados de cálculos matemáticos precisos de áreas de figuras delimitadas por curvas y de superficies curvas que no pueden aplanarse sin alterar su área. La geometría de Euclides permitía obtener las áreas de regiones delimitadas por rectas, pero al intentar obtener las áreas de regiones con fronteras curvas se experimentaron grandes dificultades. Fue Arquímedes quien señaló el camino adecuado para resolverlas con un método que, a la larga, daría lugar a una de las teorías y herramientas matemáticas más poderosas: el cálculo integral.

Para comprender la importancia de esta parte del trabajo de Arquímedes conviene repasar los problemas de “cuadraturas” que se plantearon los matemáticos griegos. “Cuadrar” una figura era encontrar un cuadrado —un rectángulo o un triángulo— cuya área fuera igual a la de la figura dada. El problema de la cuadratura del círculo se hizo famoso a través de los siglos al encontrar que resultaba imposible hacerlo únicamente con los métodos y elementos de la geometría de Euclides, es decir, sólo con regla y compás. Un caso especial de cuadratura que sí se obtuvo con los métodos euclidianos fue el de las llamadas lunas de Hipócrates de Quíos —en el siglo V a.C.—, quien demostró que las figuras obtenidas por defecto de los semicírculos construidos sobre los catetos de un triángulo rectángulo ABC , y el semicírculo que tiene como diámetro la hipotenusa y pasa por el ángulo recto, tienen igual área que el triángulo ABC .

<Imagen
42, hacer imagen
idéntica>

Figura 2.42 El área de las dos lunas de Hipócrates es igual a la del triángulo rectángulo.

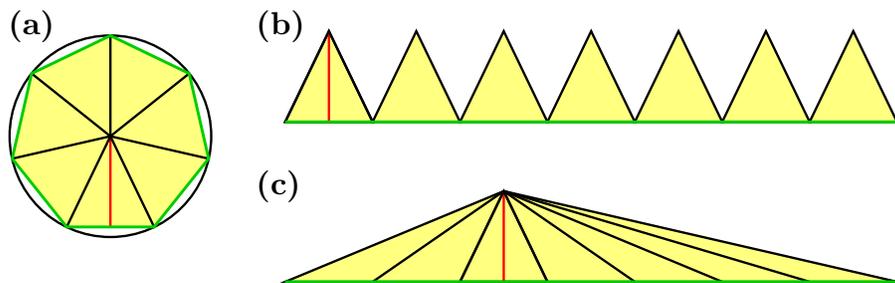


El resultado anterior es consecuencia directa del Teorema de Pitágoras aplicado, no a los cuadrados construidos sobre los catetos y la hipotenusa, sino a los semicírculos. La suma de las áreas de los semicírculos —que tienen como diámetros a los catetos— es igual al área del semicírculo —cuyo diámetro es la hipotenusa—. La conclusión de Hipócrates se infiere observando que las lunas son el resultado de restar de los semicírculos pequeños las mismas regiones que hay que restar al semicírculo grande para obtener el triángulo.

Esta ingeniosa deducción dio esperanzas a los matemáticos griegos para poder obtener, mediante trucos ingeniosos, la cuadratura de otras figuras delimitadas por curvas, en particular la del círculo. Sin embargo fue imposible, como finalmente logró demostrar Ferdinand Lindemann en 1882. Los métodos euclidianos consisten fundamentalmente de construcciones —hechas con regla y compás— en correspondencia con operaciones algebraicas que incluyen la extracción de raíces cuadradas. Los números algebraicos son los que pueden obtenerse mediante este tipo de operaciones —a partir de los enteros—, mientras que se llama **trascendentes** a los que no se pueden obtener de esta manera. Lindemann demostró que π es un número trascendente y, por lo tanto, la cuadratura del círculo no puede obtenerse usando regla y compás.

Sin embargo, dos mil años antes Arquímedes vio que había otro camino para resolver las cuadraturas: uno mucho más general pero que requería de un concepto que él mismo

no definió completamente, aunque contribuyó a crear, el de **límite**. Veamos en forma resumida cómo obtiene Arquímedes el área de un círculo de radio R .



<Imagen 43, imagen idéntica, sin colores>

Figura 2.43 El área del polígono inscrito es igual a la del triángulo, cuya base es el perímetro del polígono y cuya altura es el apotema.

Si se inscribe un polígono regular de n lados en un círculo de radio R , su área es la de n triángulos isósceles de base L y altura A , es decir, $n \frac{LA}{2}$, donde L es la longitud del lado del polígono y A , el apotema. A medida que n aumenta, el área del polígono tiende a la del círculo y, al mismo tiempo, nL tiende al perímetro $2\pi R$ mientras el apotema A tiende al radio R . Por lo tanto:

$$n \frac{LA}{2} = \frac{nL \cdot A}{2} \text{ tiende a } \frac{2\pi R \cdot R}{2} = \pi R^2$$

cuando n tiende a ∞ ,

lo cual demuestra que el área del círculo de radio R es πR^2 . Este argumento recurre al concepto moderno de “tender al límite”, pero Arquímedes hizo su demostración utilizando desigualdades similares a las que usó para el cálculo de π .

En la siguiente sección se estudiará cómo aparece π en el cálculo de las áreas y volúmenes de cilindros, conos y esferas.

2.11 Cilindros, conos y esferas

Arquímedes de Siracusa —sin duda el matemático más importante de la Antigüedad— quiso que, como único epitafio en su tumba —encontrada efectivamente con esa inscripción en la isla de Sicilia—, se dibujara una esfera inscrita en un cilindro junto con la relación que había descubierto entre las áreas y los volúmenes de estas figuras. Consciente de ser un gran matemático, Arquímedes se ufanaba de ello y ese resultado suyo era del que más orgulloso se sentía. Y no es para menos. Desde la escuela primaria sabemos las fórmulas para calcular las superficies y volúmenes de cilindros, conos y esferas, pero nunca nos explican cómo se obtienen, sólo nos dan las fabulosas recetas. A continuación repasaremos esas fórmulas, explicando cómo se obtiene cada una de ellas.



<Imagen 44, tiene que ser una fotografía de la medalla>

Figura 2.44 Busto de Arquímedes en la medalla Fields, reconocimiento internacional que se otorga cada cuatro años por descubrimientos sobresalientes en matemáticas.

<Imagen 45, similar guardando las proporciones>

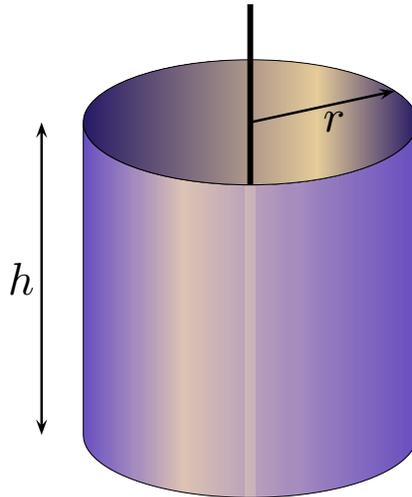


Figura 2.45 La superficie del cilindro con radio r y altura h es $2\pi rh$.

Para el caso del cilindro, la fórmula se justifica con facilidad. Si se corta la figura a lo largo de una de sus alturas y se extiende hasta dejarla plana, se obtiene un rectángulo cuya base es el perímetro $2\pi r$ de la base del cilindro, y cuya altura es la misma del cilindro: h y por lo tanto, su superficie es $2\pi r h$.

Un cono circular recto se caracteriza por tener una base circular y el vértice situado sobre la recta perpendicular a la base, que pasa por su centro. Las aristas de estos conos tienen todas la misma longitud.

<Imagen 46, similar guardando las proporciones>

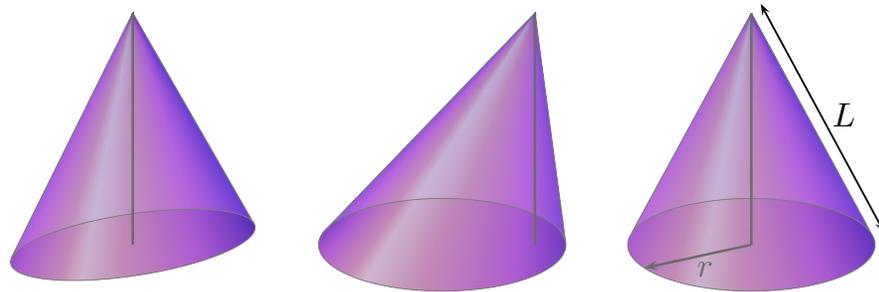


Figura 2.46 De izquierda a derecha, se muestra un cono, un cono circular y un cono circular recto.

Un cono circular recto de lado L puede construirse a partir de un sector circular de radio L y, por lo tanto, su área es igual a la del sector circular, o sea, la mitad del perímetro por el radio.

<Imagen 47, idéntica>

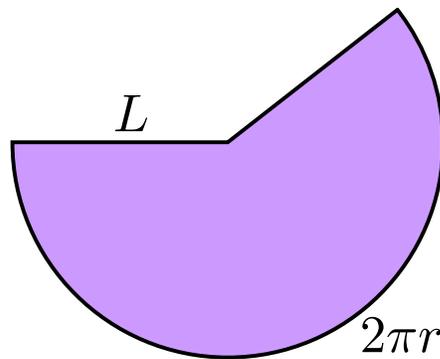


Figura 2.47 Sector circular obtenido al extender un cono.

Como el perímetro es $2\pi r$, donde r es el radio de la base circular del cono y L , el radio, entonces el área del cono es $S = \pi r L$.

<Imagen 48, similar guardando las proporciones>

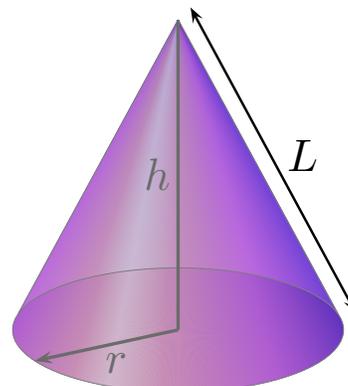
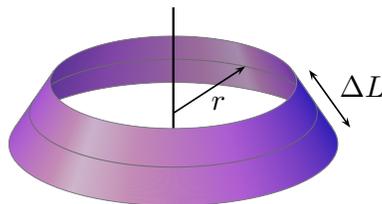


Figura 2.48 Superficie del cono de lado L y con radio de la base igual a r : $S = \pi r L$.

Para estudiar la superficie de una esfera, Arquímedes la aproxima por medio de sectores cónicos. Un sector de cono circular recto es la parte del cono comprendida entre dos planos paralelos a su base, como se muestra en la figura 2.49.



<Imagen 49, similar guardando las proporciones>

Figura 2.49 Sector de cono circular recto, de radio r y ancho ΔL .

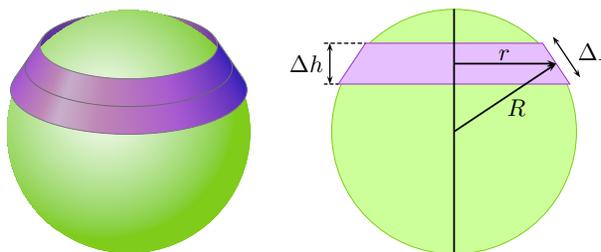
El área ΔS de un sector cónico con radio medio r y ancho ΔL es igual al perímetro de la circunferencia media por el ancho, es decir:

$$\Delta S = 2\pi r \cdot \Delta L$$

Este importante resultado puede obtenerse intuitivamente sumando las áreas de una infinidad de trapecios de ancho **infinitesimal** y altura ΔL . La suma de los anchos medios de los trapecios sería igual al perímetro de la circunferencia media $2\pi r$. También puede obtenerse mediante un cálculo exacto considerando al sector cónico como la diferencia de dos conos: uno cuyo radio de la base es $r + \frac{\Delta r}{2}$ y su lado es $L + \frac{\Delta L}{2}$ y otro, cuyo radio de la base es $r - \frac{\Delta r}{2}$ y su lado es $L - \frac{\Delta L}{2}$. Al aplicar la fórmula para las áreas de ambos conos y restarlas, se obtiene:

$$\begin{aligned} \Delta S &= \pi \cdot \left(r + \frac{\Delta r}{2}\right) \cdot \left(L + \frac{\Delta L}{2}\right) - \pi \cdot \left(r - \frac{\Delta r}{2}\right) \cdot \left(L - \frac{\Delta L}{2}\right) \quad (6) \\ &= \pi \cdot \left(2L \frac{\Delta r}{2} + 2r \frac{\Delta L}{2}\right) = \pi \cdot (L\Delta r + r\Delta L) \end{aligned}$$

Por semejanza de triángulos se tiene que $\frac{\Delta r}{\Delta L} = \frac{r}{L}$ y, por lo tanto, $L\Delta r = r\Delta L$. Al sustituir en (6) tenemos $\Delta S = 2\pi r \cdot \Delta L$, que es lo que se deseaba demostrar.



<Imagen 50, similar guardando las proporciones>

Figura 2.50 Sector de cono circular recto, de radio r , ancho ΔL y altura h , que es tangente a una esfera de radio R .

Si un sector cónico de radio medio r y ancho ΔL es tangente a una esfera de radio R —precisamente a lo largo de la circunferencia media, como muestra la figura 2.50—, entonces el área ΔS del sector cónico es igual a:

$$2\pi R \cdot \Delta h$$

donde Δh es la altura del sector. En efecto, para demostrar esto basta ver que, por semejanza de triángulos,

$$\frac{r}{R} = \frac{\Delta h}{\Delta L} \text{ y por tanto } r \cdot \Delta L = L \cdot \Delta h$$

Estamos ya en condiciones de calcular el área de una esfera tal como lo hizo Arquímedes. Para hacerlo, ahora nosotros, cubrimos la esfera con sectores cónicos tangentes a ella, como se ilustra en la siguiente figura:

<Imagen 51, similar guardando las proporciones>

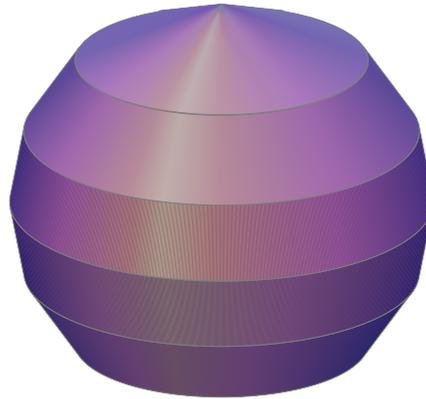


Figura 2.51 Esfera cubierta por sectores cónicos tangentes a ella.

Ya que todos los sectores cónicos son tangentes a la esfera de radio R , sus áreas son $\Delta S_i = 2\pi r \cdot \Delta h_i$, donde Δh_i son las alturas de los diferentes sectores cónicos. La suma de todas estas áreas es igual a:

$$\sum \Delta S_i = 2\pi R H,$$

donde H es la altura total de la cubierta de la esfera. Obsérvese que $2R < H$, y que H puede hacerse tan cercana a $2R$ como se desee utilizando una cubierta suficientemente fina. Por lo tanto, el área de la esfera satisface la desigualdad:

$$S \leq 4\pi R^2$$

De manera análoga, al **inscribir** en la esfera los sectores cónicos, se demuestra la desigualdad contraria: $4\pi R^2 \leq S$ y, por consiguiente,

$$S = 4\pi R^2$$

En otras palabras, el área de la esfera es igual al cuádruple del área de uno de sus círculos máximos. Otra manera de interpretar el resultado es diciendo que el área de la esfera es igual a la del mínimo cilindro que la contiene, sin contar las tapas, o bien, que el área de la esfera es $\frac{2}{3}$ del área del cilindro que la contiene, si se incluyen las tapas.

<Imagen 52, similar guardando las proporciones>

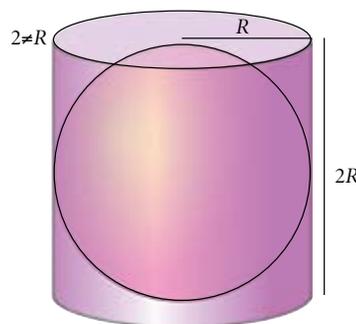
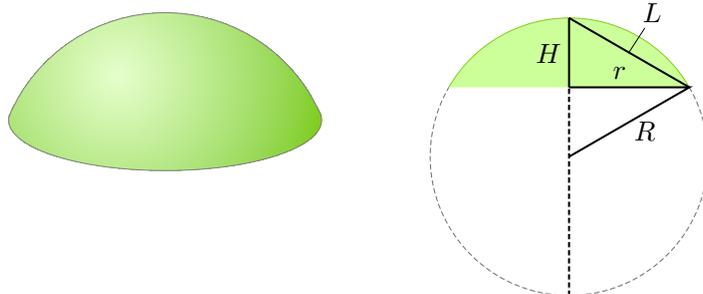


Figura 2.52 Superficies S_E de la esfera y S_C del cilindro: $S_E : S_C :: 2 : 3$.

Ésta es la figura y la relación que debían aparecer como epitafio en la tumba de Arquímedes, de acuerdo con sus propios deseos y también, de acuerdo con el testimonio de Cicerón —que visitó la tumba del gran matemático dos siglos después de su muerte—. Como veremos más adelante, la relación de $\frac{2}{3}$ se da entre los volúmenes de las mismas figuras.

El área de un casquete esférico, que consiste en la parte de una esfera que se encuentra arriba de un plano horizontal —como se muestra en la figura 2.11.10—, es igual a la de la circunferencia, cuyo radio es la distancia L del polo norte a cualquiera de los puntos de la orilla, es decir, $S = \pi L^2$.



<Imagen 53, similar guardando las proporciones>

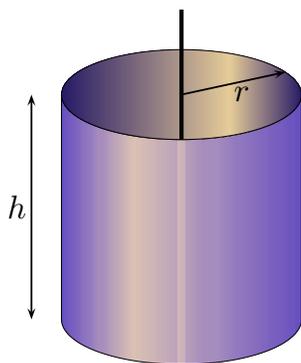
Figura 2.53 Un casquete esférico es la parte de la superficie de una esfera, cortada por un plano que pasa arriba del ecuador.

La prueba de este resultado se hace cubriendo la superficie esférica con sectores cónicos, como se hizo con anterioridad para toda la esfera. Así, se obtiene que $S = 2\pi RH$, donde H es la altura de la superficie esférica. Por el teorema de Pitágoras:

$$L^2 = H^2 + r^2 = H^2 + R^2 - (R - H)^2 = 2RH.$$

Sustituyendo esta igualdad en la fórmula anterior se obtiene el resultado anunciado: $S = \pi L^2$.

Pasemos ahora al cálculo de los volúmenes del cilindro, el cono y la esfera. El caso del cilindro es muy sencillo pues todas sus secciones paralelas a la base son círculos del mismo radio —esto es, tienen la misma área $A = \pi r^2$ — y, por lo tanto, el volumen es el producto de esta área por la altura h , es decir:



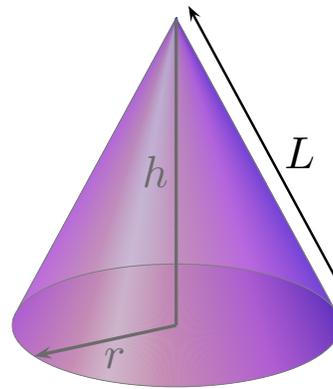
<Imagen 54, similar guardando las proporciones>

Figura 2.54 Volumen del cilindro de altura h y radio r : $V = \pi r^2 h$.

Para obtener el volumen del cono, se recurre a pirámides inscritas y circunscritas, como muestra la figura 2.55 y se utiliza el resultado conocido de que el volumen de una pirámide es igual a la tercera parte del producto del área de la base por la altura.

<Imagen 55, similar guardando las proporciones>

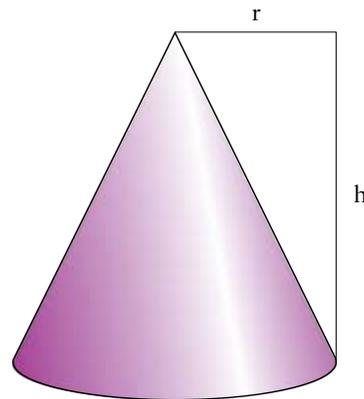
Figura 2.55 Cono con una pirámide inscrita y otra circunscrita.



Esto permite demostrar que lo mismo sucede para un cono, como se muestra en la figura 2.56.

<Imagen 56, similar guardando las proporciones>

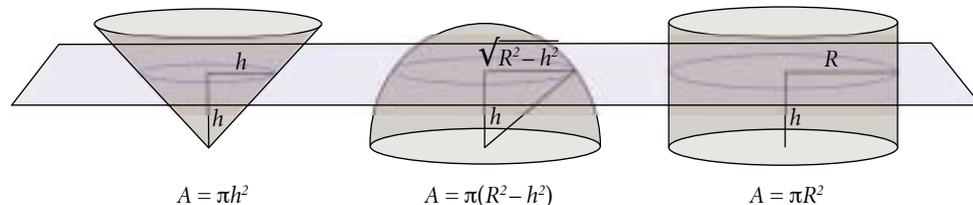
Figura 2.56 Volumen del cono de altura h y radio r :
 $V = \frac{1}{3}\pi r^2 h$.



Para obtener el volumen de la esfera, Arquímedes utilizó un truco muy ingenioso que consiste en comparar las secciones que se obtienen al cortar por un plano horizontal a un cono, una semiesfera y un cilindro, como se muestra en la figura 2.11.13.

<Imagen 57, idéntica>

Figura 2.57 Cortes de un cono, un hemisferio y un cilindro por un plano.



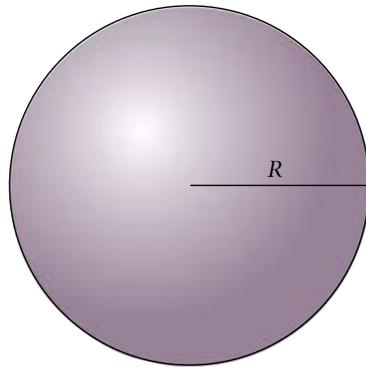
La suma de las áreas de las secciones sobre el cono y la semiesfera es igual al área de la sección del cilindro, y esto sucede para todos los cortes. Por consiguiente, se puede deducir que el volumen del cono más el de la semiesfera es igual al del cilindro, es decir, si denotamos por V el volumen de la semiesfera, entonces:

$$\frac{1}{3}\pi R^3 + V = \pi R^3$$

Por lo tanto,

$$V = \frac{2}{3}\pi R^3$$

y en consecuencia, el volumen de la esfera es el doble:

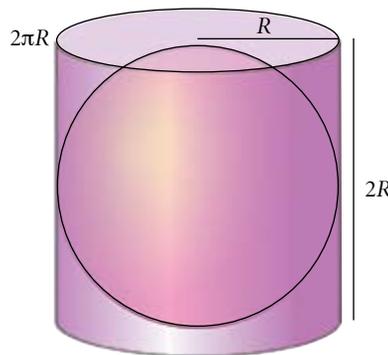


<Imagen 58, similar guardando las proporciones>

Figura 2.58 Volumen de la esfera de radio R :
 $V = \frac{4}{3}\pi R^3$.

Comparando el volumen de la esfera con el del cilindro de radio R y altura $2R$ —que vale $2\pi R^3$ —, observamos que:

$$\frac{\frac{4}{3}\pi R^3}{2\pi R^3} = \frac{\frac{4}{3}}{2} = \frac{4}{6} = \frac{2}{3}$$



<Imagen 52, similar guardando las proporciones>

Figura 2.59 Volúmenes V_E de la esfera y V_C del cilindro: $V_E : V_C :: 2:3$.

y entonces, se cumple la misma relación $2 : 3$ que se obtuvo para las superficies.

2.12 LA CUADRATURA DE LA PARÁBOLA Y EL MÉTODO DE DEMOSTRACIÓN POR INDUCCIÓN

La cuadratura de la parábola, al igual que el cálculo del volumen de la esfera, se debe al genio de Arquímedes. Aquí la obtendremos recurriendo a un procedimiento general para obtener áreas de regiones delimitadas por curvas, esencia del cálculo integral. Este método nos llevará, por un camino diferente al de la sección anterior, hasta obtener los volúmenes del cono y la esfera.

Para encontrar la cuadratura de la parábola necesitaremos la igualdad:

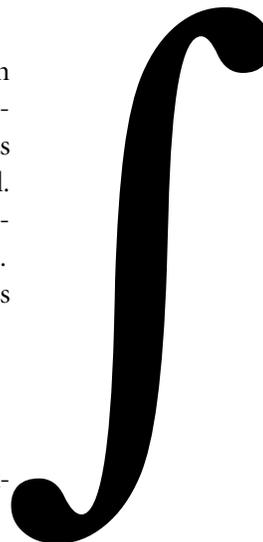
$$1 + 2^2 + \dots + N^2 = \frac{N(N + 1)(2N + 1)}{6}$$

que es válida para todo entero positivo N . Otra manera de expresar esta igualdad es usando la notación de suma:

$$1 + 2^2 + \dots + N^2 = \sum_{n=1}^N n^2$$

<Imagen 59, hacer imagen similar>

Figura 2.60 La integración es un procedimiento para aproximar un área o un volumen mediante sumas de pequeñas partes. Es un proceso matemático tan importante que tiene su propio símbolo: a primera vista se parece a una de las ranuras en un cello, pero proviene de una letra S estilizada que, a la vez, recuerda la suma.



por lo que la fórmula puede también, escribirse así:

$$\sum_{n=1}^N n^2 = \frac{N(N+1)(2N+1)}{6}$$

Esta fórmula suele demostrarse por el llamado **método de inducción** que consta de dos pasos:

1. Se comprueba que la fórmula es válida para $N = 1$.
2. Suponiendo que la fórmula es válida para algún entero positivo N , se prueba que también será válida para el siguiente entero $N + 1$.

El primer paso se cumple porque, para $N = 1$, ambos lados de la igualdad valen 1, en efecto:

$$\sum_{n=1}^1 n^2 = 1 \text{ y } \frac{N(N+1)(2N+1)}{6} = \frac{1 \cdot (1+1) \cdot (2+1)}{6} = \frac{1 \cdot 2 \cdot 3}{6} = 1$$

Ahora, supongamos que la igualdad se cumple para algún entero positivo N y probemos que entonces también se cumple para $N + 1$.

$\sum_{n=1}^{N+1} n^2 = \sum_{n=1}^N n^2 + (N+1)^2$	se separa el último término de la suma
$= \frac{N(N+1)(2N+1)}{6} + (N+1)^2$	se usa la hipótesis de inducción
$= \frac{N(N+1)(2N+1) + 6(N+1)^2}{6}$	se saca común denominador
$= \frac{(N+1)(N \cdot (2N+1) + 6N+6)}{6}$	se factoriza $(N+1)$
$= \frac{(N+1)(2N^2+7N+6)}{6}$	se desarrolla el segundo factor del numerador
$= \frac{(N+1)(N+2)(2N+3)}{6}$	se factoriza el segundo factor del numerador
$= \frac{(N+1)((N+1)+1)(2(N+1)+1)}{6}$	se reescribe para exhibir la fórmula

Esto completa la demostración por inducción.

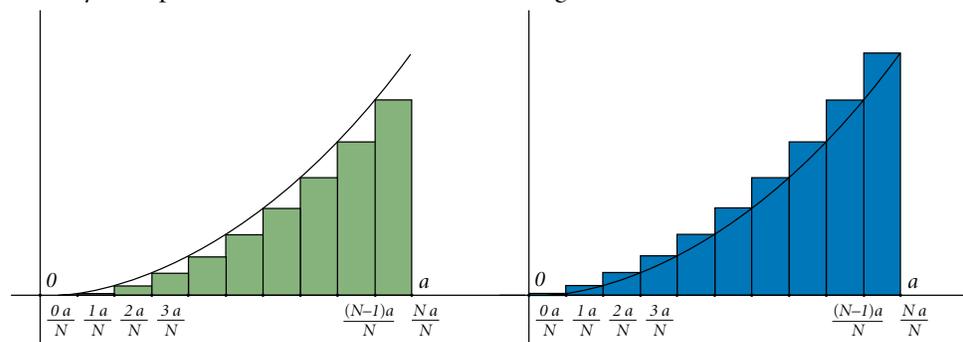
Cuadrar la parábola significa encontrar el área encerrada por ella y una recta. Simplificaremos el proceso al pasar directamente a obtener el área debajo de una parábola específica, la que corresponde a la gráfica de $y = x^2$. Denotemos por A al área bajo la parábola entre 0 y a .

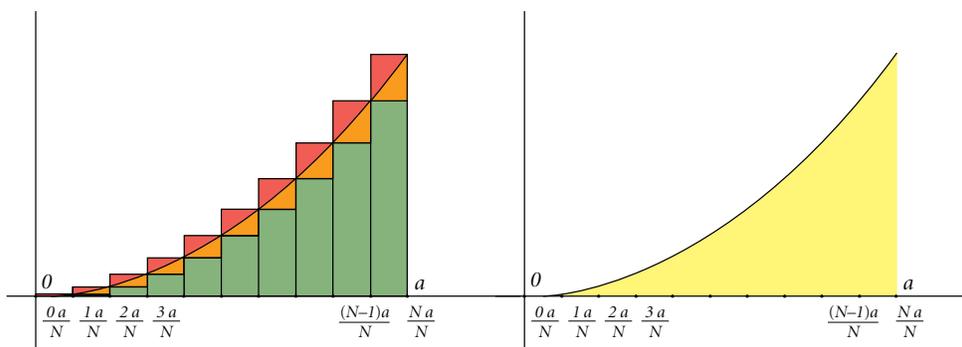
Vamos a demostrar que $A = \frac{a^3}{3}$. Para ello, dividimos el intervalo $[0, a]$ en N partes iguales y dibujamos dos gráficas formadas por rectas horizontales —una por debajo de la parábola y otra, por arriba—, como se ilustra en la figura 2.61.

<Imagen 60, idéntica, debe ir a la izquierda>

<Imagen 61, idéntica, debe ir a la derecha>

Figura 2.61
Aproximaciones por abajo y por arriba del área bajo la parábola.





<Imagen 62, idéntica, debe ir a la izquierda>
 <Imagen 63, idéntica, debe ir a la derecha>

Figura 2.62 Aproximación por ambos lados y el área bajo la parábola.

Entonces, A es mayor que el área de los rectángulos verdes y menor que el área de los rectángulos rojos. La suma de las áreas de los rectángulos verdes es:

$$\left(\frac{a}{N}\right)^2 \cdot \frac{a}{N} + \left(2\frac{a}{N}\right)^2 \cdot \frac{a}{N} + \dots + \left((N-1)\frac{a}{N}\right)^2 \cdot \frac{a}{N} = \frac{a}{N} \cdot \sum_{n=1}^{N-1} \left(n\frac{a}{N}\right)^2$$

y la de los rectángulos rojos:

$$\left(\frac{a}{N}\right)^2 \cdot \frac{a}{N} + \left(2\frac{a}{N}\right)^2 \cdot \frac{a}{N} + \dots + \left(N\frac{a}{N}\right)^2 \cdot \frac{a}{N} = \frac{a}{N} \cdot \sum_{n=1}^N \left(n\frac{a}{N}\right)^2$$

Por lo tanto, se tienen las siguientes desigualdades para el área A bajo la parábola:

$$\begin{aligned} \frac{a}{N} \cdot \sum_{n=1}^{N-1} \left(n\frac{a}{N}\right)^2 &< A < \frac{a}{N} \cdot \sum_{n=1}^N \left(n\frac{a}{N}\right)^2 \\ \frac{a^3}{N^3} \left(\sum_{n=1}^{N-1} n^2\right) &< A < \frac{a^3}{N^3} \left(\sum_{n=1}^N n^2\right) \\ \frac{a^3}{N^3} \left(\sum_{n=1}^N n^2 - N^2\right) &< A < \frac{a^3}{N^3} \left(\sum_{n=1}^N n^2\right) \end{aligned}$$

Al aplicar la fórmula que demostramos anteriormente por inducción y que nos da la suma de los primeros cuadrados, obtenemos las siguientes desigualdades:

$$\begin{aligned} \frac{N(N+1)(2N+1) - 6N^2}{6N^3} \cdot a^3 &< A < \frac{N(N+1)(2N+1)}{6N^3} \cdot a^3 \\ \left(\frac{2N^3 - 3N^2 + N}{6N^3}\right) \cdot a^3 &< A < \left(\frac{2N^3 + 3N^2 + N}{6N^3}\right) \cdot a^3 \\ \left(\frac{1}{3} - \frac{1}{2N} + \frac{1}{N^2}\right) \cdot a^3 &< A < \left(\frac{1}{3} + \frac{1}{2N} + \frac{1}{N^2}\right) \cdot a^3 \end{aligned}$$

Finalmente, como estas desigualdades se cumplen para cualquier entero positivo N , tomando “uno” suficientemente grande podemos hacer que, tanto el lado izquierdo como el derecho, estén tan cercanos a $\frac{a^3}{3}$ como queramos. Y entonces:

$$A = \frac{a^3}{3}$$

que es lo que queríamos demostrar.

El método anterior es un proceso infinito de aproximación que nos permite obtener un resultado exacto, y contiene la esencia del cálculo integral que Arquímedes ya usaba en el siglo III a.C. y, que Newton y luego Leibniz —al formalizarlo—, generalizaron casi dos mil años más tarde. En la notación moderna del cálculo integral, inventada por Leibniz, este resultado se expresa así:

$$\lim_{N \rightarrow \infty} \frac{a}{N} \cdot \sum_{n=1}^{N-1} \left(n \frac{a}{N}\right)^2 = \int_0^a x^2 dx$$

y debe leerse como “la integral de cero a a de x^2 con respecto a x ”. La integral se convierte en el límite de cualquier suma cuando el número de particiones N tiende a infinito —y por lo tanto, el ancho de las particiones tiende a cero.

Intuitivamente, la integral se interpreta como una “suma infinita” de cantidades “infinitamente pequeñas” o bien, “infinitesimales”. Aunque esta manera de hablar no es matemáticamente rigurosa, se utiliza con frecuencia como una forma abreviada de indicar el proceso con el que se llega —sí, rigurosamente—, al resultado.

El método de integración usado para alcanzar la cuadratura de la parábola nos permite obtener fácilmente, como corolarios, dos resultados ya conocidos: el volumen del cono y el de la esfera.

Para encontrar el volumen del cono cuya altura es H y cuya base tiene radio R , lo partimos en una “infinitud” de discos —cilindros circulares— de grosor “infinitesimal” dh y radio $r = \frac{R}{H}h$, donde h varía de 0 a H . Entonces, el volumen del cono es la “suma infinita” de los volúmenes de todos esos discos. En la notación de Leibniz, lo anterior se escribe como:

$$V = \int_0^H \pi r^2 dh = \pi \frac{R^2}{H^2} \int_0^H h^2 dh$$

Salvo por la constante $\pi \frac{R^2}{H^2}$ y un cambio en el nombre de la variable — x por h —, esta es la misma integral que se usó para obtener la cuadratura de la parábola, de donde sabemos que:

$$\int_0^H h^2 dh = \frac{H^3}{3}$$

En consecuencia, el volumen del cono es:

$$V = \pi \frac{R^2}{H^2} \cdot \frac{H^3}{3} = \frac{1}{3} \pi R^2 H$$

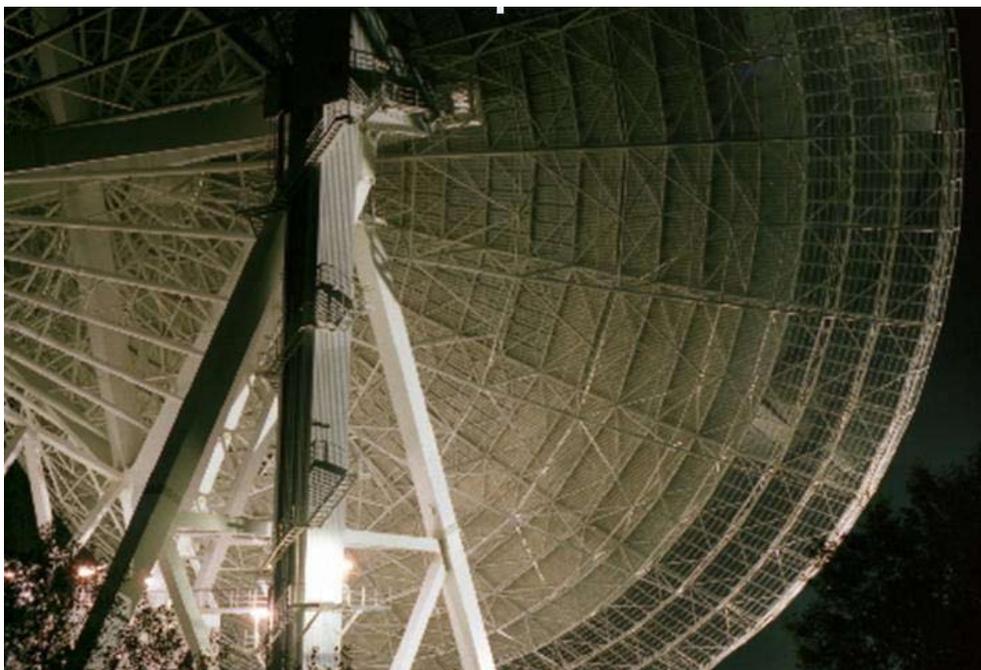
En forma análoga, partiendo del resultado de que el área de una superficie esférica de radio r es $4\pi r^2$, el cálculo del volumen de la esfera de radio R es equivalente a la cuadratura de la parábola $y = 4\pi x^2$ entre 0 y R . En efecto, la esfera se puede considerar como la suma de una infinitud de superficies esféricas de grosor infinitesimal dr y de área —variable— $4\pi r^2$ con r entre 0 y R . Por consiguiente, el volumen de la esfera puede calcularse también como una “suma infinita” de cantidades “infinitesimales”:

$$V = \int_0^R 4\pi r^2 dr$$

Y de nuevo, salvo por la constante 4π y un cambio en el nombre de la variable — x por r —, esta “suma infinita” de “infinitesimales”, que hoy llamamos *integral*, es la misma que se usa para cuadrar la parábola. Por lo tanto, el volumen de la esfera de radio R es:

$$V = \frac{4\pi R^3}{3}$$

2.13 LAS CÓNICAS Y SU USO



<Imagen 64, fotografía similar y avisar para ajustar el pie>

<Effelsberg.jpg hay que sustituirlo>

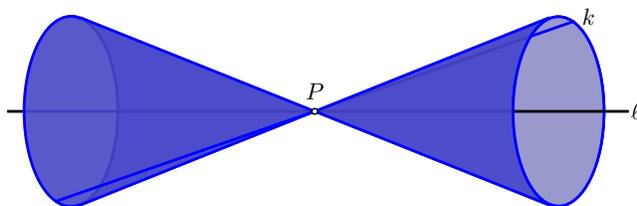
Figura 2.63 La antena de Effelsberg, un radiotelescopio de 100 metros de diámetro es la más grande en Europa. Estos telescopios se diseñan para registrar señales electromagnéticas con frecuencias de 10 MHz a 100 MHz, lo que corresponde a longitudes de onda de 3 a 30 metros. Por ello, es posible usar mallas de metal para formar la superficie del reflector que concentra las ondas en el receptor. Para que las reflexiones lleguen a un solo punto —donde está el receptor— es necesario que la antena tenga una forma particular: que sea parabólica.

2.13.1 Secciones de un cono

Las cónicas son curvas en el plano que aparecen en diferentes contextos —como en la naturaleza y en la tecnología— y están presentes en muchos objetos de uso diario —por ejemplo, en los faros de los automóviles. Su popularidad se debe a un efecto similar que se describió en la esfera de la sección 1.3.1: cumplen muchas propiedades y éstas las definen de manera definitiva. En esta sección revisaremos varios de esos aspectos.

Se cree que el primer estudio sobre cónicas lo hizo Menaechmus, para resolver el problema de la duplicación del cubo —véase también sección 4.2. Aquí empezaremos un poco distinto y tomaremos un primer acercamiento a las cónicas según el origen de su nombre: las **secciones cónicas**, es decir, las secciones con un cono. Cabe agregar que, en dicho acercamiento, se usa el espacio para definir estas curvas.

Primero, se construye un doble cono: Se fija una recta l , el **eje**, un punto P sobre l y un ángulo α , tomando en cuenta que $0^\circ < \alpha < 90^\circ$. Entonces, consideramos todas las rectas que pasan por P e inciden en l con un ángulo α . Podemos pensar que tomamos una de estas rectas, la llamamos k y la rotamos alrededor de l . Lo que obtenemos es un doble cono como se observa en la figura 2.64.



<Imagen 65, similar guardando las proporciones>

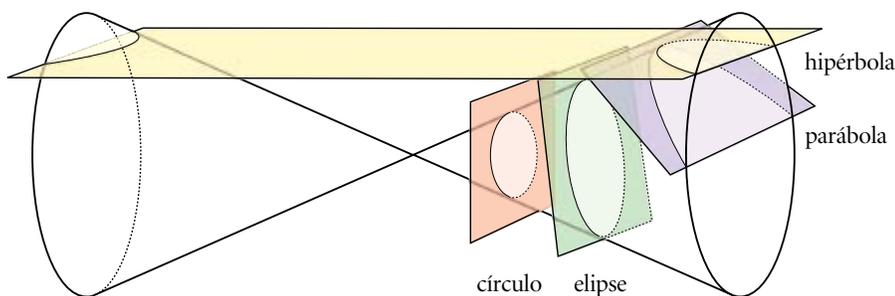
Figura 2.64 Doble cono que se obtiene como resultado al rotar una recta.

Una **sección cónica** es la intersección de un doble cono con un plano y en la figura 2.13.3 se muestran varios casos típicos. Veamos primero los posibles casos, cuando el plano pasa por P : la sección cónica podrá ser un punto, una recta o un par de rectas que se intersectan. Estos casos se llaman cónicas **degeneradas**. Más interesante es cuando el plano no pasa por P y obtenemos cuatro casos distintos:

1. Si el plano es perpendicular al eje l , entonces la cónica resultante será una **circunferencia**.
 2. Si el plano se inclina un ángulo mayor que α pero menor de 90° , entonces la cónica será una **elipse**.
 3. Si el plano se inclina exactamente por el ángulo α , es decir, si algunas de las rectas que definen el doble cono es paralela al plano, entonces la sección resultará ser una **parábola**.
- Si la inclinación del plano con respecto a l es menor que α , se obtendrá una **hipérbola**. La hipérbola no es conexa, sino que consiste de dos **ramas**.

<Imagen 66, necesitamos explicar al ilustrador>
<falta 120pt ENTRA IMAGEN DE LOS CUATRO CORTES DE UN CONO (F2)>

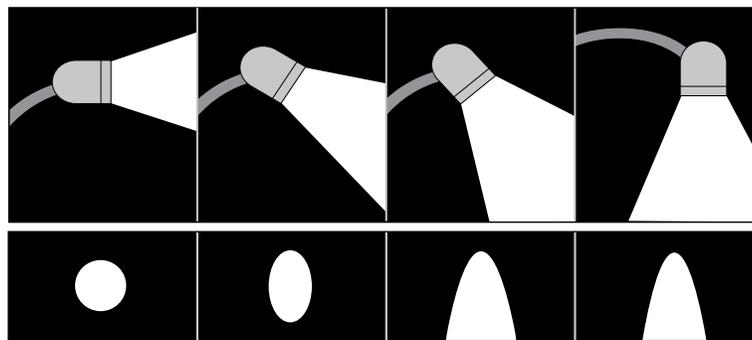
Figura 2.65 Diferentes secciones de un cono.



Podemos ver secciones cónicas en diferentes lugares. Una lámpara de mesa con una sombrilla de apertura circular proyectará —sobre la mesa o una pared cercana— luz en ciertas áreas mientras otras partes, quedarán oscuras. La curva divisoria entre luz y sombra será siempre una cónica. La razón para que ocurra lo anterior es sencilla: la mesa o la pared es plana y la luz, se emite en un cono.

<Imagen 67, necesitamos explicar al ilustrador>
<falta 100pt ILUSTRACION DE DIFERENTES POSICIONES DE LA LAMPARA (F3)>

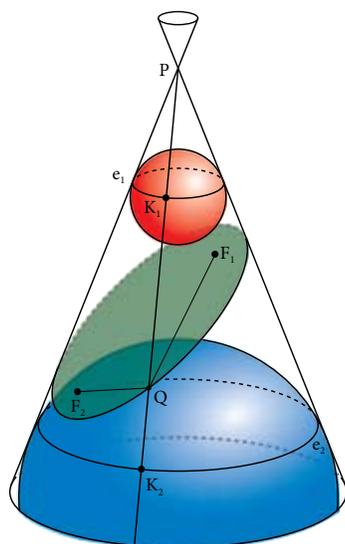
Figura 2.66 La proyección de luz genera diferentes cónicas.



Otro ejemplo son los “spots” empotrados en el techo de muchos cines, que proyectan luz sobre una pared vertical formando parte de una hipérbola.

2.13.2 Las cónicas como lugares geométricos

Fue idea del matemático belga, Germinal Pierre Dandelin, insertar dos esferas en el cono que tocan el plano de intersección dado. La figura 2.67 muestra un caso donde la cónica es una elipse.



<Imagen 68, necesitamos explicar al ilustrador>

<falta 140pt ILUSTRACION DE LAS ESFERAS DE DANDELIN EN ELIPSE, PARABOLA E HIPERBOLA (F4)>

Figura 2.67 Las esferas de Dandelin para el caso de la elipse.

Todas las tangentes de una esfera —desde un punto fijo— tienen la misma longitud. Las dos esferas tocan al cono en dos circunferencias, e_1 y e_2 —donde cada una está formada por puntos que se encuentran a la misma distancia de P —. Por ello, para cada punto Q de la elipse, la distancia $K_1 K_2$ es la misma — K_1 y K_2 son los puntos de las circunferencias e_1 y e_2 que están sobre la recta PQ —. Como QK_1 y QK_2 son tangentes de la misma esfera, miden lo mismo. En forma similar, QK_2 mide lo mismo que QF_2 , por lo tanto:

$$d(Q, F_1) + d(Q, F_2) = d(Q, K_1) + d(Q, K_2) = d(K_1, K_2) = \text{const} \quad (7)$$

Los dos puntos F_1 y F_2 son los **focos** de la elipse. La propiedad de la elipse como lugar geométrico se enuncia como sigue: La **elipse** es el **lugar geométrico** de todos los puntos Q , tales que $d(Q, F_1) + d(Q, F_2) = h$.

Lo anterior quiere decir que la condición $d(Q, F_1) + d(Q, F_2) = d$ define a todos los puntos de la cónica. Esta propiedad de la elipse se conoce como “la construcción del jardinero”, pues se puede trazar una elipse con un hilo amarrado en los cabos a dos postes en el piso —mientras que el hilo esté flojo— al tensar el hilo hacia afuera.

<Imagen 69, similar manteniendo las proporciones>

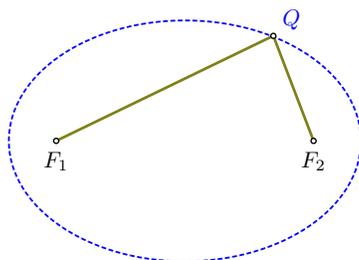


Figura 2.68 La construcción del jardinero de la elipse.

El concepto de **lugar geométrico** es importante en la geometría, dado que es un objeto geométrico definido como un conjunto de puntos que satisfacen alguna propiedad determinada. La siguiente lista muestra algunos lugares geométricos:

1. Dado un punto M y una distancia d , el lugar geométrico de todos los puntos que están a una distancia d de M es una circunferencia con centro M y de radio d .
2. Dados dos puntos A y B , el lugar geométrico de todos los puntos P tal que $d(P, A) = d(P, B)$ es una recta —la mediatriz— del segmento AB .
3. Dadas dos rectas a y b que se intersecan en P , el conjunto de puntos Q , tal que $d(P, a) = d(P, b)$, son dos rectas perpendiculares —la unión de las dos bisectrices de a y b .

4. Dado un segmento AB , el lugar geométrico de puntos C , tal que $\angle ACB = 90^\circ$, es una circunferencia con diámetro AB .
5. Dadas dos circunferencias que no se intersecan c_1, c_2 , el lugar geométrico de todos los puntos P , tal que la tangente de P a c_1 tenga la misma longitud que la tangente de P a c_2 , es una recta.

El argumento para la hipérbola es similar, sólo que ahora es una diferencia:

$$d(Q, F_2) - d(Q, F_1) = d(Q, K_2) - d(Q, K_1) = d(K_1, K_2) = \text{const}$$

Consecuentemente se tiene la siguiente caracterización: La **hipérbola** es el lugar geométrico de todos los puntos Q , tal que $|d(Q, F_1) - d(Q, F_2)| = h$.

El valor absoluto de la diferencia se requiere para obtener ambas ramas de la hipérbola. En el caso de la parábola sólo hay una esfera de Dandelin y la descripción geométrica como lugar geométrico se obtendrá más tarde.

A la mitad del siglo XX, se empezó a instalar un sistema de ubicación en alta mar que se llama LORAN —el nombre viene del inglés “long range navigation” — donde se emiten señales desde puntos fijos en la costa. Un barco recibe estas señales en tiempos distintos y puede, a partir de la diferencia de tiempo, calcular la diferencia de distancia. Se sabe entonces que el barco se encuentra sobre una hipérbola. Con las señales de tres emisores se puede calcular una ubicación con buena precisión. El sistema LORAN dejará de operar durante el año 2010, dado que el GPS —del inglés “global positioning system” — otorga una posición mucho más precisa.

2.13.3 La excentricidad

Cada esfera de Dandelin toca al cono en una circunferencia e , que está en un plano A perpendicular al eje l . Si se interseca A con el plano B definido por la cónica, se obtiene una línea d llamada **directriz**. En consecuencia, una elipse y una hipérbola tienen dos directrices, la parábola tiene una y la circunferencia, ninguna.

<Imagen 70, necesitamos explicar al ilustrador>

<falta 120pt ILUSTARCIÓN DE LA DIRECTRIZ EN EL CASO DE LA ELIPSE (F5)>

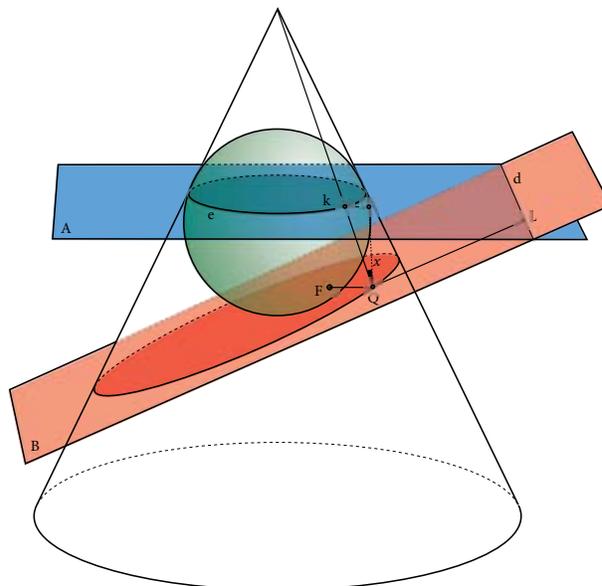


Figura 2.69 Directriz para el caso de la elipse.

Ya vimos que cada punto Q de la cónica satisface $d(Q, F) = d(Q, K)$, donde K es el punto de intersección de e con la recta PQ . Sea x la distancia del punto Q al plano A . Por lo tanto, se tiene que $x = \cos(\alpha) d(Q, K)$ —aquí recordamos que α es el ángulo entre la recta PQ y el eje l . Sea L el punto de la directriz más cercana al punto Q —en consecuencia, QL y d son perpendiculares—, entonces $x = \cos(\beta) d(Q, L)$, donde β es el ángulo entre l y B . De ahí obtenemos que:

$$d(Q, F) = \epsilon d(Q, d), \quad \epsilon = \frac{\cos(\beta)}{\cos(\alpha)}.$$

El valor c es la **excentricidad** de la cónica. La siguiente tabla muestra la excentricidad para los diferentes casos.

Cónica	ángulo entre l y B	excentricidad
circunferencia	$\beta = 90^\circ$	$\epsilon = 0$
elipse	$90^\circ > \beta > \alpha$	$0 < \epsilon < 1$
parábola	$\beta = \alpha$	$\epsilon = 1$
hipérbola	$\beta < \alpha$	$\epsilon > 1$

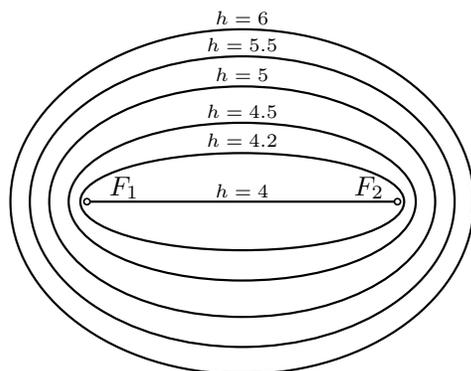
Lo anterior nos proporciona una descripción de la parábola como lugar geométrico: La parábola es el lugar geométrico de todos los puntos Q , tal que $d(Q, F) = d(Q, d)$, donde el punto F es el **foco** y la recta d es la directriz de la parábola.

Los planetas orbitan alrededor del Sol siguiendo trayectorias elípticas. Esto fue un hallazgo importante para la cultura y más aún si uno ve que las elipses son casi circunferencias —se trata de elipses con excentricidad muy baja. Por ejemplo, la órbita de la Tierra tiene una excentricidad de $\epsilon = 0.0167$, mientras la de Marte es de $\epsilon = 0.0933$.

2.13.4 La forma de una antena

Recordemos que la elipse es el lugar geométrico de los puntos que tienen una constante suma de distancias h a los dos focos. De esta manera, el plano se divide en **curvas de nivel** para diferentes valores de h .

Veamos cómo esta división del plano en curvas de nivel tiene una consecuencia inesperada: cualquier rayo que sale de uno de los focos, es reflejado en la elipse —como si fuera un espejo— y llega justo al otro foco. Esto se conoce como la propiedad de la reflexión de la elipse.



<Imagen 71, similar manteniendo las proporciones>

Figura 2.70 Partición del plano en una infinidad de curvas de nivel.

Para entender esta propiedad, consideremos la elipse definida para un valor h . Ahora bien, si Q es un punto de la elipse, entonces definimos como b a la recta que pasa por Q e interseca a las rectas F_1Q y F_2Q con el mismo ángulo. Es decir, b es el espejo para que el rayo

F_1Q se refleje en Q hacia el punto F_2 , como se observa en la figura 2.13.8. Luego definimos al punto L como el reflejo de F_1 en b . Por ello, se tiene que $d(Q, F_2) = d(Q, L)$ y entonces:

$$d(F_1, Q) + d(Q, L) = d(F_1, Q) + d(Q, F_2) = h.$$

Falta demostrar que b es tangente a la elipse, es decir, que cualquier punto R de b —con $R \neq Q$ — está afuera de la elipse, o dicho de otra manera, que $d(R, F_1) + d(R, F_2) > h$. Por lo anterior, sigue que $d(R, F_2) = d(R, L)$ por la desigualdad del triángulo en ΔF_1LR , el lado F_1L tiene menor longitud que la suma de los lados F_1R y RL . Si juntamos ambos argumentos obtenemos que:

$$d(R, F_1) + d(R, F_2) = d(F_1, R) + d(R, L) > d(F_1, Q) + d(Q, L) = h,$$

lo cual muestra que b es tangente a la elipse.

La propiedad de reflexión se usa en la práctica, por ejemplo, en hornos especiales que tienen una forma de elipsoide —un cuerpo que se obtiene al rotar una elipse en el espacio por el eje F_1F_2 — donde en un foco se coloca una fuente de calor y en el otro, el objeto a calentar. Otro caso ocurre en el Desierto de los Leones, donde la bóveda elipsoidal de la “Capilla de los susurros” tiene el siguiente efecto: lo que susurra una pareja en uno de sus focos, se escucha perfectamente bien en el otro.

<Imagen 72, similar manteniendo proporciones>

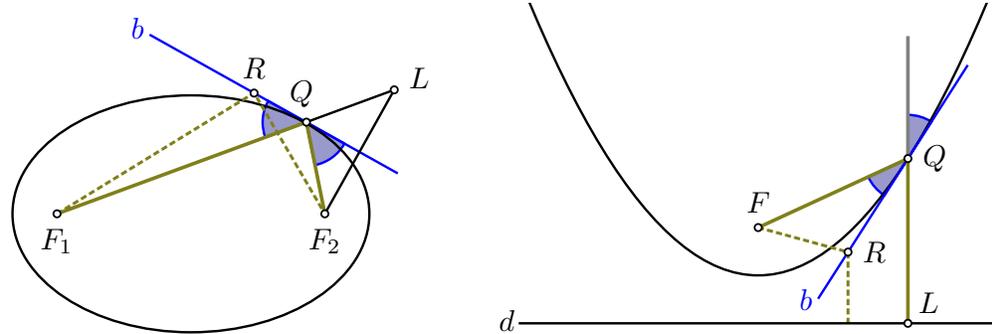


Figura 2.71 La propiedad de reflexión en la elipse y la parábola.

De manera similar se puede ver que en la parábola hay una propiedad de reflexión. Si Q es un punto de la parábola, dibujemos a la recta b que pasa por Q tal que incluya los mismos ángulos con FQ y QL —donde F es el foco y L , el pie de la perpendicular a la directriz a por Q — como se observa en la figura 2.71. Sólo hay que demostrar que b es tangente para terminar de concluir que el reflejo de FQ en la parábola es perpendicular.

El hecho de que b sea tangente se verifica de manera sencilla: para cualquier punto R de b —con $R \neq Q$ —, se tiene que $d(F, R) = d(L, R) > d(L', R)$ donde L' es el pie de R en a . Por lo tanto, R está bajo la parábola.

<Imagen 73, similar manteniendo proporciones>

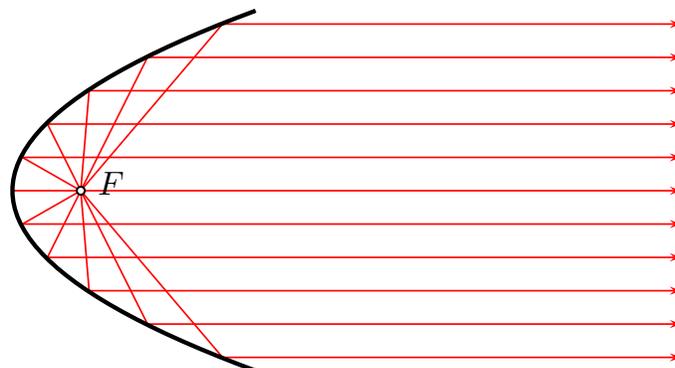


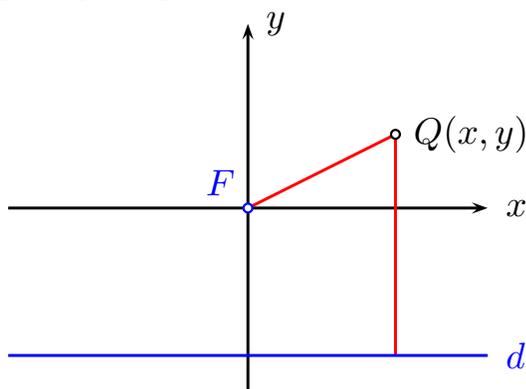
Figura 2.72 Propiedad de reflexión de la parábola.

Esta particularidad de la parábola se aplica en los faros de los automóviles que son paraboloides, es decir, tienen la forma que se obtiene al rotar una parábola en su eje de simetría. El foquito luminoso se coloca en el foco del paraboloide y, de esta manera, se obtiene un haz de luz que se dispersa muy poco y puede alumbrar lejos. El mismo efecto pero a la inversa se usa en las antenas de astronomía y también en las que reciben señales de satélites para la televisión. En ambos casos se reflejan ondas que llegan prácticamente paralelas en un paraboloide que los hace converger en el foco donde se encuentra el receptor. Dicha manera de agrupar los rayos tiene el efecto de multiplicar la señal, es decir, de aumentar su intensidad.

2.13.5 Ecuaciones de segundo grado

La última parte de la descripción de las cónicas es la analítica, es decir, queremos describir las cónicas con ecuaciones. Para ello usamos la descripción de las cónicas como lugares geométricos, mediante un foco y la directriz correspondiente. Por consiguiente, la cónica es el conjunto de puntos Q tal que se satisface la ecuación: $d(Q, F) = \epsilon d(Q, d)$ donde ϵ es la excentricidad, F es un foco y d , la directriz correspondiente.

Podemos elegir nuestro sistema de coordenadas de manera que los cálculos sean lo más sencillos posible. Una manera de hacerlo es al poner el origen en el foco y exigir que la directriz sea paralela al eje de coordenadas x . Entonces, calculamos las dos distancias $d(Q, F) = \sqrt{x^2 + y^2}$ y $d(Q, d) = y + k$, donde $k = d(F, d)$.



<Imagen 74, similar manteniendo proporciones>

Figura 2.73 Distancias al foco y a la directriz de un punto Q.

Al sustituir estas expresiones en (7) y elevar al cuadrado, se obtiene:

$$x^2 + y^2 = \epsilon^2 y^2 + 2\epsilon^2 ky + \epsilon^2 k^2 .$$

Como ϵ y k son números fijos para cada cónica dada, se puede escribir esta ecuación como:

$$x^2 + (1 - \epsilon^2)y^2 + (-2\epsilon^2 k)y + (-\epsilon^2 k^2) = 0. \tag{8}$$

Si definimos $a = 1$, $b = 0$, $c = 1 - \epsilon^2$, $d = 0$, $e = -2\epsilon^2 k$, $f = -\epsilon^2 k^2$, entonces vemos que la expresión (8) es un caso particular de la **ecuación general de segundo grado en dos variables**:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0.$$

Debe resaltarse que ya demostramos que cada cónica puede escribirse de la forma $a = 1$, $b = 0$, $d = 0$.

No es tan claro que cualquier ecuación de segundo grado con dos variables define una cónica. En efecto, se puede demostrar que siempre define una cónica, pero la demostración es bastante larga y engorrosa.

La idea es cambiar el sistema de coordenadas de tal manera que la ecuación se simplifica lo suficiente como para reconocer que, en efecto, se trata de una cónica. Resaltamos aquí que, en ciertos sistemas de coordenadas, las ecuaciones de las cónicas toman una forma muy especial. Por ejemplo, cualquier ecuación de la forma:

$$y = ax^2 + bx + c,$$

con $a \neq 0$, define una parábola. Podemos cambiar el sistema de coordenadas para escribirla en forma aún más sencilla, como:

$$y = ax^2.$$

Para la elipse, se puede encontrar un sistema de coordenadas de manera que la ecuación toma la forma:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

y la hipérbola se representa como:

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1.$$

2.13.6 Cónicas en la física

Galilei argumentó que el movimiento sobre la superficie de la Tierra encuentra en la dirección vertical hacia la Tierra una aceleración constante y, que esto, resulta en una **función cuadrática**:

$$z(t) = at^2 + bt + c \quad (9)$$

para la altura $z(t)$ y en función del tiempo t . Sin embargo, la dirección horizontal no experimenta aceleración alguna. Se trata de un movimiento uniforme y la posición cambia de manera lineal:

$$h(t) = dt + e.$$

Si $d \neq 0$, entonces es posible despejar la variable t que luego se puede sustituir en (9). Lo que resulta es una ecuación de la forma:

$$z = a't^2 + b'h + c',$$

donde a' , b' y c' son coeficientes reales. Como vimos antes, esta expresión corresponde a una parábola. El movimiento sobre la superficie terrestre —de cualquier cuerpo pesado y extenso— es, en muy buena aproximación, parabólico —no toma en cuenta la fricción del

aire, el hecho de que la Tierra sea redonda ni el que “vertical” en un lugar determinado no es paralelo a “vertical” en otro, por lo que no se trata de una ecuación en un sistema de coordenadas ortonormales —ejes verticales con la misma escala.

Con más trabajo se puede ver que la primera ley de Newton tiene como consecuencia el que cualquier cuerpo que orbita un cuerpo pesado —que atrae todos los cuerpos hacia sí— tiene que viajar a lo largo de una trayectoria, que resulta ser una cónica. Como la única cónica que describe una trayectoria cerrada es la elipse —y la circunferencia, como un caso particular—, se obtiene que los planetas giran en elipses alrededor del Sol.

Por último, vale la pena repasar lo que se hizo en esta sección. Las cónicas se definieron como secciones cónicas; después se usaron las esferas de Dandelin para deducir la descripción de las cónicas como lugares geométricos —los griegos ya conocían estas propiedades, aunque Dandelin vivió en el siglo XIX. En la actualidad se usa el argumento de Dandelin porque es más “elegante”, es decir, más corto que los argumentos que dieron los griegos para concluir lo mismo. Al final, esbozamos el tratamiento de las cónicas desde el punto de vista de la geometría analítica, usando ecuaciones. La geometría analítica tiene su origen en el siglo XVII con trabajos de Descartes y de Fermat —quien demostró que cualquier ecuación de segundo grado con dos variables describe una cónica.

Como se puede ver el discurso dio unos saltos para adelante y para atrás en el tiempo y no siguió el desarrollo histórico. Esto es completamente intencional: un tratamiento histórico sería otra cosa. Pero, ¿por qué habrá introducido Dandelin sus esferas si ya se conocían las propiedades? En matemáticas, la historia nunca acaba, siempre se puede repensar un tema y verlo desde otro ángulo. Los objetos de las matemáticas están tan intensamente interrelacionados que, a veces, resulta que lo que era la causa se transforma en la consecuencia.

2.14 PROBABILIDAD Y ESTADÍSTICA, CALCULANDO EL AZAR

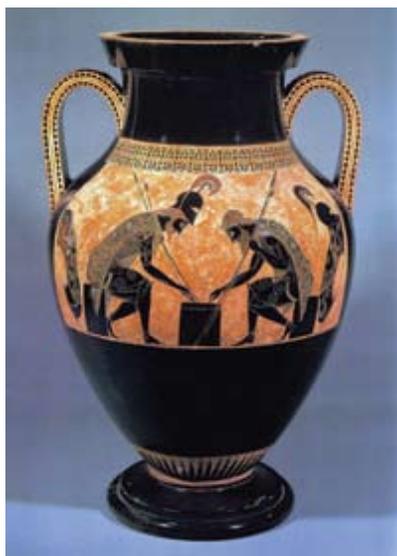


Figura 2.74 Jarrón romano.

En la vida diaria hay acontecimientos que obedecen a causas conocidas y que podemos predecir, mientras otros son resultado de la casualidad. De los primeros se ocupan las ciencias naturales, de los segundos, la probabilidad. Que una moneda lanzada al aire describa una parábola y caiga al suelo, es algo totalmente predecible de lo que se encarga la física. Los rebotes consecuentes a la caída y su posición final en el suelo —con una de sus caras hacia arriba— es algo, en cambio, extraordinariamente difícil de anticipar. Por lo tanto, cada vez que lanzamos la moneda el resultado es fortuito, fruto de la casualidad, del azar. Dentro del

margen de lo que sí sabemos es que, aproximadamente la mitad de las veces, la moneda caerá con una cara hacia arriba y el resto, con esa cara hacia abajo. Este conocimiento es justo el que compete al estudio de la **probabilidad**. Estos fenómenos casuales o azarosos, se llaman **aleatorios**.

<Imagen 76, fotografía similar>



Figura 2.75 Tique, diosa de la suerte.

En las civilizaciones antiguas se creía que algún dios se ocupaba de determinar los resultados de los sucesos aleatorios y, con ello, del destino. Se asocia a la diosa griega Tique —en griego, Τυχη significa suerte— con esa actividad, mediante la cual decidía la fortuna de las personas y las ciudades. El nombre romano de Tique es, precisamente, Fortuna. El caos reinante en diversas épocas históricas se atribuía a una intensa labor de la caprichosa diosa, considerada hija de Hermes y Afrodita. Según Libanius, el reconocido retórico, el templo dedicado a Tique en Alejandría era uno de los más impresionantes del mundo helénico. ¡Cómo no edificarlo grandioso y rendirle adoración a alguien que controlara el resultado de todos los fenómenos aleatorios y que, por ende, tendría un poder casi absoluto sobre el destino de cada individuo! Todas las casualidades serían resul-

tado de su antojo. Con seguridad, si alguien así existiera, todos querríamos contarnos entre sus protegidos.

El hombre moderno no cree ya en una deidad que determine los fenómenos aleatorios; sabe que el azar es impredecible e incontrolable, y que todo cuanto puede hacer con él es tratar de comprenderlo, sin buscar causas donde no las hay.

2.14.1 Necesidad de una teoría de la probabilidad

La teoría de la probabilidad es la rama de las matemáticas que se ocupa del azar desde el punto de vista cuantitativo. Sus inicios, que ocurren en el siglo XVII, son increíblemente recientes para ser un estudio que abarca fenómenos tan importantes y frecuentes. Quizá, el tardío surgimiento de la probabilidad como interés dentro de las matemáticas, se deba a que el hombre se resistía a creer que hubiera incidentes verdaderamente impredecibles, o a que es difícil imaginar cómo decir algo sobre situaciones gobernadas por el azar. En su larga historia intelectual, el ser humano detentaba logros tan grandes como llegar a predecir los eclipses, descubrir las leyes matemáticas del sistema solar, entender cómo se reproducen los seres vivos... Al aceptar la existencia de fenómenos verdaderamente aleatorios, parecía darse por vencido —acaso, claudicar— en la tarea de conocer la naturaleza.

En realidad, hay dos razones para ocuparse de la probabilidad: una práctica y otra filosófica. La primera es que muchas de las cosas que ocurren a nuestro alrededor no son predecibles; por ejemplo, para predecir el resultado del lanzamiento de una ficha, sería necesario conocer su impulso, la altura alcanzada, su aceleración de caída, la velocidad de giro en el momento del lanzamiento, en fin, poseer los detalles exactos y precisos de todo lo anterior y de la superficie sobre la que va a caer, incluyendo las propiedades elásticas del material del que está hecha; habría que conocer incluso la velocidad del aire en cada momento del

recorrido. Recabar estos datos es poco menos que imposible; además, aunque los tuviéramos, el problema matemático que habría que resolver sería de una extraordinaria complejidad. Nada justificaría tanto esfuerzo. Así que definitivamente claudicamos y aceptamos tratar el lanzamiento de fichas como un fenómeno aleatorio.

La razón filosófica es que la ciencia del siglo XX —concretamente la física cuántica— descubrió que el azar existe, no sólo como producto de una capacidad limitada de cálculo, sino como un hecho natural en sí mismo: es imposible determinar —simultáneamente— y con total exactitud, la posición y la velocidad de una partícula elemental como el electrón, o bien, predecir cuándo un átomo de una sustancia radiactiva va a decaer al emitir energía y convertirse en un átomo diferente. No se trata de incompetencia humana o falta de instrumentos adecuados, sino que ésa es parte esencial de la naturaleza misma de dichas partículas. No es aquí donde analizaremos este asunto, pero conviene señalarlo para que el lector sepa que, en la actualidad, ya no se concibe al mundo físico como determinista, al contrario, se sabe que, inexorable e irremediabilmente, el azar interviene en él.

Cualquiera de las razones expuestas con anterioridad es lo suficientemente importante para aceptar a la probabilidad como una rama legítima del conocimiento; aunque el uso práctico motivó su estudio un tiempo antes. De hecho, los fenómenos aleatorios que primero ocuparon la mente de los matemáticos fueron los juegos de azar.

Escrito alrededor de 1560, encontramos el libro *Liber ludo aleae* sobre los juegos de azar de Girolamo Cardano; tenemos después la correspondencia —de 1654 a 1660— entre Pierre de Fermat y Blaise Pascal, que se considera como el inicio del cálculo de probabilidades.



<Imagen 77 de Pascal, cualquiera>

<Imagen 78 de Fermat, cualquiera>

Figura 2.76 Blaise Pascal y Pierre de Fermat.

Aparentemente, el Chevalier de Méré, un jugador empedernido e inteligente, planteó un problema a Pascal que motivó aquella famosa relación epistolar entre él y Fermat. El problema en cuestión trataba sobre cómo sería justo repartir el dinero de las apuestas en un juego de azar si éste se suspendía antes de que hubiera un ganador. La solución requería de un cálculo preciso de las probabilidades que tenía cada participante de ganar el juego en caso de que hubiera continuado. En dicha correspondencia —es importante señalarlo— se establece por vez primera que es posible determinar las probabilidades con absoluta precisión bajo supuestos claros sobre la naturaleza y las reglas del juego en cuestión. Es decir, que hay una única manera lógica de determinar las probabilidades que cada jugador tiene de ganar el juego y éstas, pueden emplearse para repartir de una manera justa la apuesta.

2.14.2 El problema con el que se inicia el cálculo de probabilidades

Supongamos que dos personas deciden participar en un juego que consiste en lanzar cinco veces una ficha al aire. Cada una aporta 50 fichas. Si salen más caras o soles que cruces o águilas, el primer jugador se llevará todas las fichas; en caso contrario, se las llevará el segundo jugador. Cuando se han lanzado 3 fichas y se ha visto que 2 son cara y 1 cruz, tiene que suspenderse el juego. ¿Cuántas fichas debe llevarse cada jugador?

Está claro que al parar quien tenía mayor probabilidad de ganar era el primer jugador, y por lo tanto parece justo que él se lleve más fichas que el segundo. Un árbitro imparcial podría decidir que el primer jugador se lleve todas las fichas ya que en el momento de suspenderse el juego, él llevaba ventaja pero al segundo jugador esta decisión le parece injusta porque aún tenía posibilidades de ganar. El árbitro puede decidir también que, como el juego se suspendió, cada jugador recupere sus 50 fichas; el primer jugador protesta argumentando que esa repartición hubiera sido justa si el juego nunca hubiera comenzado, o se hubiera suspendido cuando se habían tirado dos fichas —una cara y la otra cruz—, pero que en la situación actual se está ignorando la ventaja que evidentemente él llevaba.

A estas alturas parece más justo decidir que cada quien se lleve un número de fichas proporcional a la probabilidad que tenía de ganar en el momento en que se detuvo el juego. Esto requiere esclarecer el concepto de probabilidad y encontrar cómo calcularla numéricamente, y es a lo que Fermat y Pascal se abocaron en su correspondencia. Para abordar la situación sería necesario superar acaloradas discusiones y, finalmente, acordar algunos puntos. Uno es que el resultado parcial de dos caras y una cruz no tiene ningún valor predictivo sobre los dos siguientes lanzamientos —pueden caer tanto cara como cruz. Lo siguiente es cómo debería calcularse la probabilidad de que el vencedor fuera el primer jugador. Para ello, se calcula el número de casos que lo llevarían a ganar y el resultado se divide entre el número total de casos posibles —esto es válido sólo bajo la hipótesis de que todos los casos posibles sean igualmente probables. Siguiendo esta línea de argumentación, los dos siguientes lanzamientos darían un total de cuatro posibles resultados:

cara	cara
cara	cruz
cruz	cara
cruz	cruz

Como se observa, 3 de ellos darían la victoria al primer jugador y sólo 1 al segundo. Por tanto, la probabilidad de que el vencedor sea el primer jugador es de $\frac{3}{4}$, mientras que la del segundo jugador es de $\frac{1}{4}$, lo cual significa que uno debe llevarse 75 fichas y el otro, 25. Sin embargo, si las reglas del juego cambian un poco, el problema se complica considerablemente: cabe mencionar que en matemáticas se define la probabilidad de cualquier evento como un número entre 0 y 1 que, si se multiplica por 100, representa una estimación del porcentaje de casos favorables respecto al total.

Supongamos que hay tres jugadores —A, B y C— que ponen la misma cantidad de fichas al jugar a los dados. Deciden otorgarse puntos de acuerdo con la siguiente regla: el jugador A gana un punto cuando el dado cae en 1 o 2, el jugador B lo gana cuando el dado cae en 3 o 4, y el jugador C si cae en 5 o 6. El juego termina cuando algún jugador ha ganado tres puntos y se lleva todas las fichas.

Supongamos que, en un momento dado, A tiene dos puntos y B y C tienen un punto cada uno y deciden suspender el juego. ¿Cómo repartir el lote de fichas? Habría que considerar todas las opciones posibles para cada una de las siguientes tiradas y calcular las pro-

babilidades de que cada jugador resulte ser el vencedor en esa jugada. La siguiente tabla muestra los posibles resultados de los siguientes tres lanzamientos; la letra A, B o C indica para quién fue el punto en esa tirada.

<Fórmula 271, en forma de tabla, ojo con las negritas>

5a. tirada	A			B			C		
6a. tirada	A	B	C	A	B	C	A	B	C
7a. tirada	A B C	A B C	A B C	A B C	A B C	A B C	A B C	A B C	A B C

Figura 2.77 Con negritas se indica la letra del ganador para cada tirada.

Ahora hay que contar cuántos de estos 27 casos dan la victoria a cada jugador. Para facilitar la cuenta se ha marcado con negritas cuando el jugador, en ese paso, es ya el ganador. Todos los casos debajo de una letra marcada con negritas le darían la victoria al mismo jugador y de hecho, en la práctica, esas jugadas ya no se llevarían a cabo. El resultado es:

- 17 dan la victoria a A
- 5 dan la victoria a D
- 5 dan la victoria a C

Por lo tanto, al suspenderse el juego A debe llevarse $\frac{17}{27}$ fichas del lote, mientras que B y C tendrán $\frac{5}{27}$ fichas cada uno. En números: si cada jugador hubiera apostado 9 fichas —el lote completo sería de 27 fichas—, A debería quedarse con 17, B con 5 y C con las otras 5.

Llegar a esta solución llevó a Pascal y Fermat —dos de los matemáticos más reconocidos de la historia— un esfuerzo considerable, pero su relevancia es crucial ya que señaló el camino hacia un método general para calcular las probabilidades que puede explicarse de manera muy sencilla:

La probabilidad de un evento es el número de casos favorables dividido por el número total de casos, siempre y cuando, todos los casos sean igualmente probables.

Este es el principio básico que sirve como fundamento al cálculo de probabilidades aplicado a los juegos de azar y a muchas otras situaciones en las que es posible construir un modelo basado en casos equiprobables, es decir, que tienen la misma probabilidad. En algunas situaciones concretas, llevar a cabo este cálculo puede ser extremadamente complicado pero la regla siempre es clara. Quien desee calcular probabilidades debe entender bien este principio.

2.14.2 El modelo matemático general de la probabilidad

Entre el siglo XVII y el XX se realizaron muchos avances en la teoría de la probabilidad, pero no fue sino hasta 1930 cuando el trabajo del famoso matemático ruso Andrei Nikolaievich Kolmogórov la fundó sobre bases sólidas, aprovechando un desarrollo del análisis matemático llamado **Teoría de la medida**. La idea intuitiva de este concepto es muy simple. Pensemos en la masa contenida en cada región del espacio. Hay regiones grandes que pueden contener poca masa, y otras pequeñas que pueden contener mucha, como un trozo de plomo. La función matemática que asocia cada región del espacio con la masa que contiene es un ejemplo de una **medida**. Una medida tiene la propiedad de ser **aditiva en conjuntos ajenos**, es decir, si A y B son dos regiones ajenas del espacio, la masa contenida en A y B es la masa contenida en A más la masa contenida en B. En símbolos, lo anterior se escribe así:

$$m(A \cup B) = m(A) + m(B) \text{ si } A \cap B = \emptyset$$

donde $A \cup B$ se lee “A **unión** B” y representa el conjunto de todos los puntos del espacio que están en A o B, mientras que $A \cap B$ se lee “A **intersección** B” y representa los puntos que están tanto en A como en B. \emptyset es el conjunto vacío, pues no tiene ningún elemento, por lo que la igualdad $A \cap B = \emptyset$ significa que no hay ningún punto que esté simultáneamente en A y B, o bien, que el conjunto de puntos en A y B es vacío.

La notación de conjuntos presentada brevemente en el párrafo anterior, junto con el

<Imagen 79, similar>

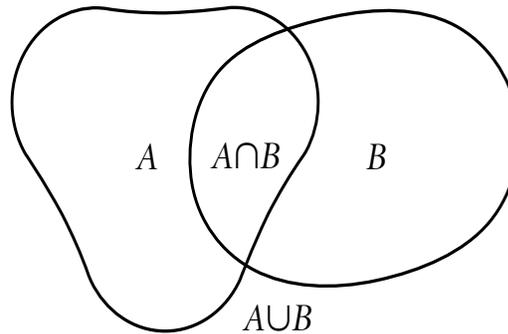


Figura 2.78 Conjuntos que se intersecan.

<Imagen 80, similar>

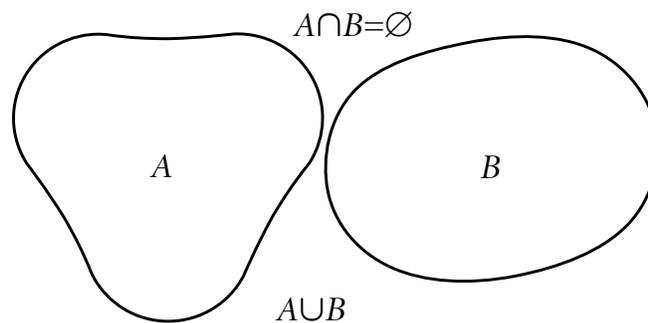


Figura 2.79 Conjuntos ajenos, es decir, con intersección vacío.

concepto de medida, constituyen la herramienta idónea para trabajar la teoría de la probabilidad.

Un **espacio de probabilidad** o **espacio muestral** se define como un conjunto Ω —la letra griega conocida como “omega”— que se interpreta como el de todos los posibles resultados de un experimento; una familia Σ —“sigma”— de subconjuntos de Ω se interpreta como los eventos del experimento cuyas probabilidades se conocen o pueden ser calculadas; y una medida p es una función que asigna, a cada evento E de Σ , un número entre 0 y 1. $p(E)$ se interpreta como la probabilidad del evento E, y debe asignar el valor de 0 al conjunto vacío y 1 al total, es decir, $p(\emptyset) = 0$ y $p(\Omega) = 1$.

Por ejemplo, en el caso más simple del lanzamiento de una ficha, Ω consta de dos elementos: *cara* y *cruz*. Σ son todos los subconjuntos de Ω : el total Ω , el vacío Φ , el que contiene sólo a *cara* y el que contiene sólo a *cruz*. La medida de probabilidad p asigna a estos eventos los valores de 1, 0, $\frac{1}{2}$ y $\frac{1}{2}$, respectivamente.

Para el caso de conjuntos finitos, se especifica un conjunto como la lista de todos sus elementos encerrada entre llaves. Por ejemplo, el evento que ofrece *cara* como el resultado del experimento, se denota por $\{cara\}$, mientras que el evento imposible se denota por el conjunto vacío Φ , que también puede escribirse también como $\{\}$. Al aprovechar estas convenciones, la descripción completa del modelo matemático del experimento aleatorio cuando se lanza una ficha, puede presentarse por la terna (Ω, Σ, p) , cuyos elementos quedan totalmente especificados a continuación:

$$\begin{aligned}\Omega &= \{\text{cara, cruz}\} \\ \Sigma &= \{\{\text{cara, cruz}\}, \{\}, \{\text{cara}\}, \{\text{cruz}\}\}. \\ p(\{\text{cara, cruz}\}) &= 1 \\ p(\{\}) &= 0 \\ p(\{\text{cara}\}) &= \frac{1}{2} \\ p(\{\text{cruz}\}) &= \frac{1}{2}\end{aligned}$$

El caso del lanzamiento de un dado queda descrito por la terna (Ω, Σ, p) , donde $\Omega = \{1, 2, 3, 4, 5, 6\}$ y los eventos son todos los subconjuntos de Ω , es decir, Σ consta de cada uno de estos subconjuntos de Ω —que son muchos, como: $\{\}, \{3\}, \{4, 5, 6\}, \{2, 3\}, \{1, 2, 3, 4, 6\}$ etc. La probabilidad p asigna a cada evento el número de sus elementos entre seis. Por ejemplo:

$$p(\{2, 3, 5, 6\}) = \frac{4}{6} = \frac{2}{3}.$$

Este modelo es aplicable a situaciones donde los posibles resultados de un experimento no son equiprobables, aunque en estos casos, para estimar los valores de las probabilidades habría que recurrir a métodos estadísticos o a suposiciones cuantitativas que podrían no estar bien justificadas.

Veamos el caso de un dado cargado: puede tener la característica de que los números 2, 3, 4 y 5 y aparezcan con probabilidad $\frac{1}{6}$ pero, en cambio, el 6 aparece con probabilidad $\frac{2}{9}$ y el 1 con probabilidad de $\frac{1}{9}$. Como:

$$\frac{1}{9} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{2}{9} = \frac{4}{6} + \frac{3}{9} = \frac{2}{3} + \frac{1}{3} = 1,$$

entonces definimos p de manera que $p(E)$ se calcula sumando tantos $\frac{1}{6}$ como números entre 2 y 5 contenga, más $\frac{1}{9}$ si contiene al 1 y $\frac{2}{9}$ si contiene al 6. Por ejemplo:

$$p(\{3, 4, 6\}) = \frac{2}{6} + \frac{2}{9} = \frac{6+4}{18} = \frac{10}{18} = \frac{5}{9}.$$

Puede comprobarse que esta es una buena definición y que podría modelar el experimento aleatorio de lanzar un dado cargado, siempre y cuando las probabilidades de salir del 1 y el 6 fueran, respectivamente, $\frac{1}{9}$ y $\frac{2}{9}$.

Este modelo de espacio probabilístico puede aplicarse a una gran cantidad de situaciones prácticas mucho más complejas que las descritas en los párrafos anteriores, incluyendo casos en los que el espacio muestral Ω de posibles resultados sea infinito.

2.14.3 Probabilidad condicional, eventos independientes y variables aleatorias

Sean A y B dos eventos en cualquier espacio probabilístico. Supongamos que al realizar el experimento nos informan que el evento A ocurrió. ¿Eso nos dice algo acerca del evento B? En general sí. Por ejemplo, consideremos estos dos eventos en el experimento de lanzar un dado:

A = “el resultado es par” y
B = “el resultado es mayor que 3”.

La probabilidad de A es claramente $\frac{1}{2}$ y la de B, también es $\frac{1}{2}$. Pero si alguien nos informa que A ocurrió, ¿seguirá B teniendo la misma probabilidad? La respuesta es no. Saber que

el resultado es par implica que ya sólo quedan tres posibilidades: 2, 4 y 6 y, como éstas son equiprobables, la probabilidad de que el resultado sea mayor que 3 es ahora $\frac{2}{3}$.

¿Entonces la probabilidad del evento B no estaba bien definida porque antes era $\frac{1}{2}$ y ahora es $\frac{2}{3}$? En realidad lo que sucede es que saber la ocurrencia del evento A nos lleva a una nueva situación probabilística, a un nuevo espacio de probabilidad. El espacio muestral en la nueva situación ya no es $\{1, 2, 3, 4, 5, 6\}$ sino $\{2, 4, 6\}$. Como estos cambios de espacio probabilístico se dan mucho en el estudio de la probabilidad, se ha desarrollado un concepto que permite referirse a estas situaciones nuevas manteniendo el lenguaje de la situación original, lo cual resulta cómodo en la práctica. Se trata del concepto de **probabilidad condicional**. El evento de que ocurra B cuando se garantiza la ocurrencia de A, se denomina “B dado A” y se denota por $B|A$. Las probabilidades de estos eventos son: $p(A) = \frac{2}{3}$, $p(B) = \frac{1}{3}$ y $p(B|A) = \frac{2}{3}$. La probabilidad condicional puede definirse así:

$$p(B|A) = \frac{p(A \cap B)}{p(A)}$$

y se puede interpretar como otro espacio de probabilidad donde el espacio muestral es A. Observemos que esta relación se cumple en el ejemplo anterior. En efecto, $p(A \cap B) = \frac{1}{3}$, pues $A \cap B = \{4, 6\}$ y, por tanto:

$$\frac{2}{3} = p(B|A) = \frac{p(A \cap B)}{p(A)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

El concepto intuitivo de probabilidad condicional corresponde de manera adecuada a la definición formal.

La probabilidad condicional es uno de los dos conceptos claves de la teoría de la probabilidad. El otro es el de **eventos independientes**, que se describe a continuación.

Consideremos otra vez el experimento de lanzar un dado pero ahora con dos nuevos eventos A y B:

A = “el resultado es impar” y

B = “el resultado es 3 o 4”.

Entonces $p(A) = \frac{1}{2}$ y $p(B) = \frac{1}{3}$. Saber que A ocurre nos dice que las únicas posibilidades son 1, 3 y 5, todas igualmente probables; por tanto, $p(B|A) = \frac{1}{3}$. En este caso la probabilidad de B y la de “B dado A” son iguales. En estas circunstancias se dice que B es independiente de A. En términos de probabilidades, decimos que B es independiente de A si $p(B|A) = p(B)$. Usando la definición de probabilidad condicional obtenemos que si B es independiente de A, entonces:

$$p(A \cap B) = p(A) p(B)$$

o en forma equivalente:

$$p(A \cap B) = p(A) p(B)$$

Esto muestra que el concepto de eventos independientes es simétrico, es decir, B es independiente de A si y sólo si A es independiente de B. Por tal motivo, se habla de eventos independientes sin especificar cual de ellos es independiente del otro. Como corolario obtenemos que si A y B son eventos independientes, entonces $p(B|A) = p(B)$ y $p(A|B) = p(A)$.

Se dice que dos eventos son **mutuamente excluyentes** si la ocurrencia de cualquiera de ellos implica que la ocurrencia del otro es imposible. Por ejemplo, en el experimento de lanzar un dado y observar el número en la cara de arriba, los siguientes eventos $A = \text{“el número obtenido es par”}$ y $B = \text{“el número obtenido es 5”}$ son mutuamente excluyentes. Que dos eventos sean mutuamente excluyentes equivale a que los conjuntos de resultados que los representan son ajenos, es decir, dos eventos A y B son mutuamente excluyentes si y sólo si $A \cap B = \emptyset$. Es importante observar que este concepto no es probabilístico sino únicamente lógico y que la probabilidad no interviene en su definición; sin embargo, con frecuencia se comete el error de confundirlo con la independencia de eventos explicada anteriormente.

Una **variable aleatoria** es una función real definida en el espacio muestral de un espacio probabilístico (Ω, Σ, p) , en otras palabras, es una función:

$$X: \Omega \rightarrow \mathbb{R}$$

Las variables aleatorias, a pesar de ser funciones, se suelen denotar con letras X, Y, \dots mayúsculas en lugar de con una f . Se dice que dos variables aleatorias X_1 y X_2 son independientes si los eventos que se pueden definir con una y otra son siempre independientes, es decir, si eventos como $\{\omega \in \Omega : a_1 \leq X_1(\omega) \leq b_1\}$ y $\{\omega \in \Omega : a_2 \leq X_2(\omega) \leq b_2\}$ son independientes.

Intuitivamente, dos variables aleatorias son independientes si el saber algo de una de ellas no nos da ninguna información acerca de la otra. Por ejemplo, en el experimento de lanzar dos dados, saber el resultado de uno de ellos no nos da información acerca del otro. Las variables aleatorias X_1 y X_2 definidas como el número en la cara superior del primer dado y el mismo número pero del segundo dado, respectivamente, son variables independientes.

Las variables independientes juegan un papel fundamental en el **Teorema del límite central**, uno de los resultados más importantes y útiles de la probabilidad y que se aplica en la **estadística**.

2.14.4 El lanzamiento de canicas a una pared y la distribución normal

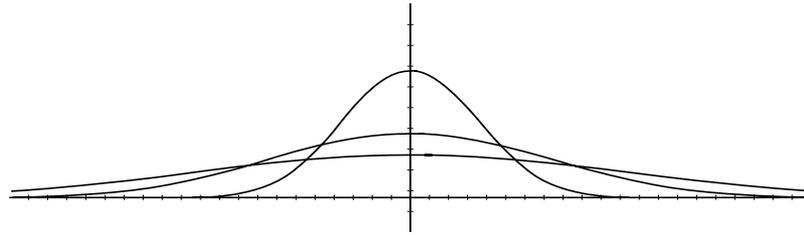
Una muestra sencilla de un experimento aleatorio —donde el espacio muestral sea infinito— es el lanzar rodando por el suelo una canica hacia un punto en la base de una pared, y registrar, para cada lanzamiento, el punto en el que la canica toca la pared. Si los lanzamientos los hace siempre la misma persona, es posible construir un modelo matemático que se aproxime bastante a la realidad. Comencemos haciendo algunas simplificaciones: supongamos que las canicas llegarán con igual probabilidad a la izquierda y a la derecha del blanco —que está ubicado en el punto b . En segundo lugar, supondremos que es más probable que caigan en un intervalo A cercano a b , que en otro intervalo B de la misma longitud pero alejado de b . Necesitamos una función que sea simétrica respecto al origen, que sea mayor a medida que nos acercamos a b y menor mientras nos alejamos, y que el área bajo su gráfica sea igual a 1. Todas estas propiedades las satisfacen las funciones definidas para cualquier número positivo σ .

A la función $N_{b,\sigma}(x)$ se le llama de distribución normal con media b y **varianza** σ^2 . También se le conoce como la función gaussiana —en honor a Gauss— o distribución de campana, debido a la forma de su gráfica parecida al perfil de una campana. En la figura 2.80 se muestran tres ejemplos de estas funciones con un mismo valor de $b = 0$ y distintos valores de σ .

$$N_{b,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-b)^2/2\sigma^2}$$

<Imagen 81, similar manteniendo proporción>

Figura 2.80 Distribuciones normales con varianzas distintas.



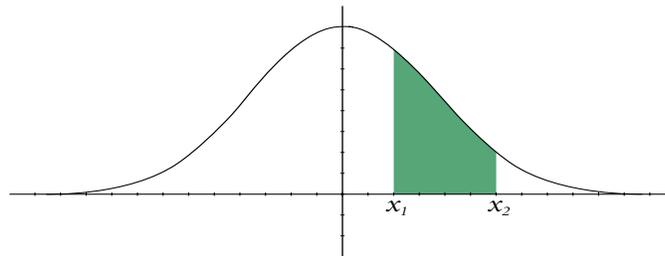
La gráfica con varianza pequeña de $\sigma = 0.5$ corresponde a un lanzador de canicas con buen tino, la de varianza $\sigma = 1$ a un jugador medio, y la de varianza $\sigma = 1.5$ a uno bastante errático. Supongamos que el jugador de nuestro experimento tiene buen tino, por lo que su varianza es $\sigma = 0.5$. La probabilidad $p([x_1, x_2])$ del evento que consiste en que la canica caiga en el intervalo $[x_1, x_2]$ está dada por el área bajo la gráfica de la curva normal $N_{b,\sigma}(x)$ entre x_1 y x_2 —sombreada en la figura 2.81—, lo cual se representa con la integral definida:

$$p([x_1, x_2]) = \int_{x_1}^{x_2} N(b, \sigma, x) dx$$

<Imagen 82, similar manteniendo proporción>

Figura 2.81 Interpretación gráfica —que se representa como el área bajo la curva— de la probabilidad de que la observación caiga entre x_1 y x_2 .

y se ilustra como el área bajo la curva en la figura 2.81.



Calcular estas integrales es difícil, antes se hacía usando tablas y hoy en día puede recurrirse a la computadora. La clave es que nuevamente pudimos aplicar el modelo general. En este caso, la medida de probabilidad \mathbf{P} se definió mediante una integral definida, es decir, como el área bajo una curva.

Las medidas de probabilidad que se utilizan en las diferentes aplicaciones de la teoría llegan a definirse incluso en espacios de dimensión infinita —como ocurre cuando se modelan los llamados **procesos estocásticos**— en los que cada elemento del espacio muestral puede ser, por ejemplo, una curva.

2.14.5 La ley de los grandes números

Quizá el resultado más general e importante de la teoría de la probabilidad es el que se denomina **Ley de los grandes números** y afirma que, a la larga, la frecuencia observada de las ocurrencias de un evento se acerca o converge a la probabilidad del mismo evento. Esta convergencia es lo que da sentido y aplicabilidad a la teoría de la probabilidad. Sin embargo, el acercamiento puede ser muy lento y esa lentitud es causa de no pocos errores en la interpretación de los resultados. Las observaciones realizadas casi siempre parecen indicar que la realidad difiere mucho del modelo probabilístico, que no hay una buena relación entre las frecuencias observadas —los cocientes entre el número de observaciones favora-

bles entre el total de observaciones— y las probabilidades. En muchas ocasiones, el resultado de un experimento aleatorio parece ser el de uno muy diferente al que estamos aplicando.

Por ejemplo, al realizar 10 veces el experimento de lanzar diez dados hemos obtenido los siguientes resultados:

1 2 3 2 4 4 3 4 4 2
 6 4 2 1 5 2 3 4 3 1
 5 2 2 5 6 4 3 5 5 6
 2 2 2 2 2 2 6 3 1 4
 6 1 2 1 4 1 5 1 3 2
 2 5 1 3 2 3 1 5 6 5
 4 3 3 6 2 4 6 5 2 1
 3 4 4 2 2 2 2 6 4 5
 1 2 1 6 3 1 6 6 5 5
 1 1 2 1 3 2 2 6 5 2

<Imagen 83, idéntica>

Figura 2.82 Resultados de cien lanzamientos de un dado.

En la primera línea no aparecen ni el 5 ni el 6, lo cual haría creer que se trata de los resultados de un dado tetraédrico —de cuatro lados— y no cúbico. La cuarta línea parece indicar que el 2 es altamente probable. Las frecuencias de aparición de cada número son:

$$\begin{aligned} \text{fr}(1) &= \frac{17}{100} \\ \text{fr}(2) &= \frac{28}{100} \\ \text{fr}(3) &= \frac{14}{100} \\ \text{fr}(4) &= \frac{14}{100} \\ \text{fr}(5) &= \frac{14}{100} \\ \text{fr}(6) &= \frac{13}{100} \end{aligned}$$

y parecen indicar un fuerte sesgo hacia el 2, como si el dado estuviera “cargado” favoreciendo al 2 sobre los otros números. Si hiciéramos otra vez el mismo experimento podríamos observar ciertas peculiaridades, distintas en cada ocasión.

Si repetimos muchas veces el experimento y calculamos las frecuencias para 10 000, 1 000 000, 100 000 000 observaciones, veremos que difieren cada vez en menos de $\frac{1}{6}$. No obstante, para llegar a observar claramente esta convergencia sería necesario recurrir a muchísimas observaciones.

En general, dos motivos contribuyen a la impresión de que las muestras resultantes de un experimento aleatorio no son tan aleatorias como se esperaría. Primero, la mente humana está entrenada para detectar irregularidades y, al observar un campo aleatorio, lo regular es precisamente lo que resalta. El segundo es que en muestras relativamente pequeñas es altamente probable que haya alguna desviación importante.

Se sabe de experimentos famosos en que el investigador manipuló ligeramente los datos para que señalaran con mayor claridad el resultado que se deseaba probar. En realidad, para probar una hipótesis en situaciones aleatorias es necesario recurrir a los métodos de la estadística, diseñados para tomar en cuenta la ley de los grandes números y, a pesar de la lentitud de la convergencia, se pueden obtener resultados confiables.

Para aplicar la teoría de la probabilidad a la vida cotidiana se requiere tener muy clara la lentitud con la que las frecuencias convergen a las probabilidades y también que, aunque los eventos poco probables ocurren escasas veces, hay muchos —por lo tanto la probabilidad

de observar alguno de ellos es bastante alta. Para aclarar este último punto, recurramos al clásico ejemplo conocido como la paradoja del cumpleaños.

2.14.6 La paradoja del cumpleaños

¿Qué probabilidad hay de que en un auditorio con 50 personas, dos cumplan años el mismo día? La mayoría creemos que esta probabilidad es muy baja, pero un cálculo preciso nos muestra que, en realidad, es alta y se aproxima a 0.97.

¿Por qué esta discrepancia entre la intuición y el cálculo preciso? Pues porque solemos confundir esta pregunta con otra: ¿Cual es la probabilidad de que **yo** encuentre entre esas 50 personas otra con mi mismo cumpleaños? La respuesta a esta segunda pregunta es, aproximadamente, de $\frac{49}{365} \approx 0.13$. Si en el auditorio hubiera 367 personas, la probabilidad de que dos cumplan años el mismo día sería 1. La probabilidad de que entre 50 personas, dos hayan nacido en la misma fecha se calcula de la siguiente manera. Primero, se obtiene la probabilidad p_n de que n personas tengan diferentes fechas de nacimiento. No entraremos en detalles, pero una reflexión cuidadosa lleva a entender que el resultado debe ser:

$$p_n = \frac{365}{365} \cdot \frac{364}{365} \cdot \dots \cdot \frac{365 - n + 1}{365}$$

En particular, para $n = 50$, $p_{50} = 0.03$ y por tanto, la probabilidad de que en el auditorio de 50 personas dos tengan el mismo cumpleaños es de $1 - p_{50} \approx 0.97$. La siguiente tabla enlista los valores de las probabilidades para distintas cantidades de personas en el auditorio:

$$\begin{aligned} 1 - p_{10} &\approx 0.14 \\ 1 - p_{20} &\approx 0.44 \\ 1 - p_{30} &\approx 0.73 \\ 1 - p_{40} &\approx 0.90 \\ 1 - p_{50} &\approx 0.97 \\ 1 - p_{60} &\approx 0.99 \end{aligned}$$

En la teoría de la probabilidad es muy frecuente encontrar resultados como el de la paradoja del cumpleaños, que contradicen nuestra intuición ingenua. Este hecho demuestra que la probabilidad es un asunto muy delicado que no debe dejarse a la intuición de personas sin experiencia en sus cálculos.

2.14.7 El teorema del límite central

La distribución normal presentada en el ejemplo anterior aparece con mucha frecuencia en la teoría de la probabilidad y ello se debe a que es la forma natural de muchas distribuciones de probabilidad, precisamente lo que afirma uno de los grandes teoremas de la teoría, el **Teorema del límite central**.

Para explicar este teorema, debemos definir lo que se entiende por una variable aleatoria. Toda función definida en el espacio muestral que satisfaga ciertas condiciones mínimas —que aquí no tiene sentido especificar— es una **variable aleatoria**. Por ejemplo, en el experimento de lanzar una ficha al aire, la función que vale 1 cuando el resultado es *cara*, y 0 cuando es *cruz*, es una variable aleatoria. En el experimento de lanzar un dado, la función

que da el número que aparece en la cara superior del dado, es una variable aleatoria. En el experimento de lanzar canicas hacia un punto de una pared, la función que asigna a cada resultado del experimento la posición x de la canica al chocar con la pared, es también una variable aleatoria.

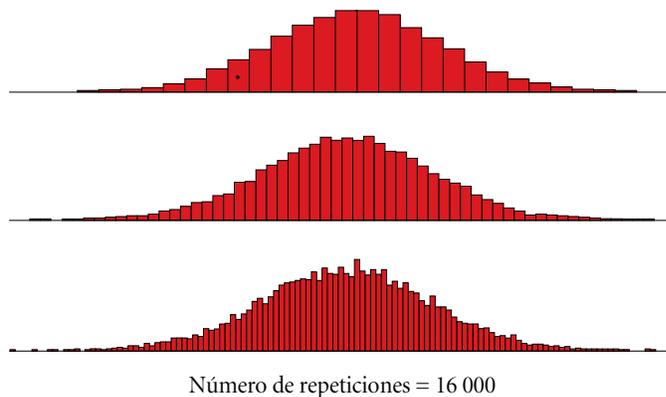
En general, una variable aleatoria es una interpretación numérica —muchas veces parcial e incompleta— del resultado de un experimento, es decir, es una manera numérica pero indirecta de observar los resultados. A las variables aleatorias se les podría llamar **observables**, como de hecho se hace en el ámbito de la mecánica cuántica, pues son formas de “observar” los resultados de un experimento aleatorio.

El enunciado preciso del teorema del límite central es complicado, pero el caso que nos interesa aquí es el que se refiere a la distribución de probabilidad de una variable aleatoria, aplicada a diferentes realizaciones de un mismo experimento. En el ejemplo del lanzamiento de un dado, si usamos la variable aleatoria X —valor del número que aparece en la cara superior del dado—, al repetir muchas veces el experimento vamos a encontrar que toma el valor de:

$$\frac{X_1 + \dots + X_n - nb}{a\sqrt{n}}$$

y al graficar las frecuencias con las que aparece cada valor de este “promedio”, entonces la gráfica de las frecuencias se parece cada vez más a la distribución normal.

Para visualizar este resultado, observemos las gráficas de la figura 2.83 en las que se han marcado con segmentos rojos las frecuencias de los valores obtenidos de los “promedios” de las primeras 5, 20 y 80 observaciones, respectivamente.



<Imagen 84 similar, manteniendo proporciones y texto de abajo: Número de repeticiones = 16000>

Figura 2.83 Histogramas que corresponden a un número grande de observaciones, con conjuntos de 5, 20 y 80 dados, respectivamente.

Vemos que el parecido de estos **histogramas** con la distribución normal es muy grande. Las observaciones aleatorias tienden a comportarse, en promedio, como una distribución normal. Debido a lo anterior, la distribución normal tiene muchas aplicaciones en la estadística.

Muchas veces se elabora la hipótesis de que una variable aleatoria cualquiera tiene una distribución normal y ésta se determina, sin muchos miramientos, a partir de una muestra de tamaño N que ha dado observaciones x_1, x_2, \dots, x_N , tomando como media μ al promedio de los valores x_i y como varianza σ^2 , al promedio de las cantidades $(x_i - \mu)^2$, es decir, se supone que la distribución de la variable aleatoria es:

$$N_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

donde:

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$$

En realidad μ y σ^2 , calculadas de esta manera, son la media y la varianza de la muestra x_1, x_2, \dots, x_N y no de la distribución real. Identificar estos promedios con la media y la varianza de la distribución real es un abuso que, aunque ocasionalmente pueda ser útil, es causa de no pocos errores de interpretación.

Para obtener estimaciones adecuadas de las distribuciones de probabilidad de una variable aleatoria a partir de un conjunto de datos, hay que recurrir a los métodos especializados de la estadística.

2.14.8 La estadística

La estadística es el “arte” de obtener información a partir de datos conocidos. Permite medir y cuantificar la confianza de las conclusiones sobre un cierto problema, así como el tamaño del error que se comete. Si bien se puede definir en pocas palabras, en realidad consiste de un cúmulo enorme de conocimientos y recomendaciones encaminados a la necesidad de tomar decisiones con base en información incompleta —cosa que ocurre en casi todas las situaciones de la vida real—, de manera que se pueda hacer con plena conciencia de lo que la información disponible dice y también, de lo que no dice.

La estadística se divide en dos grandes ramas, la **estadística descriptiva** y la **estadística inferencial**.

La estadística descriptiva consta de todos los métodos recomendados para presentar la información —en general, de forma gráfica— fácilmente comprensible y evitar posibles interpretaciones erróneas. Todos conocemos los diferentes tipos de tablas y gráficas que se muestran en la prensa y en todos los demás sitios donde se presenta información al público —algunos ejemplos en las figuras 2.84 a 2.86.

<Imagen 85, similar manteniendo proporciones y textos, salvo recuadro con flecha>

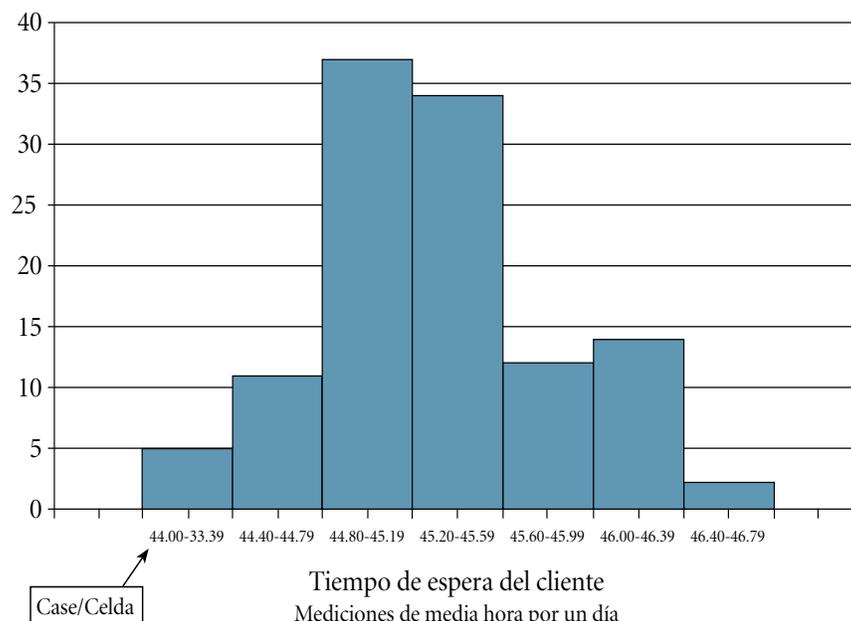
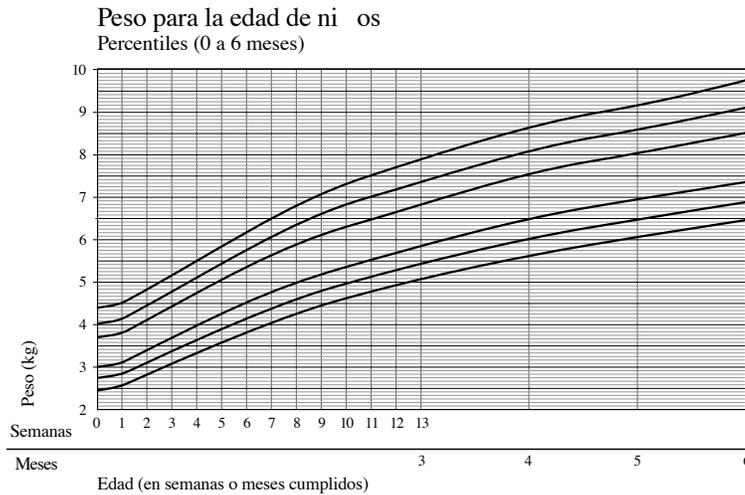
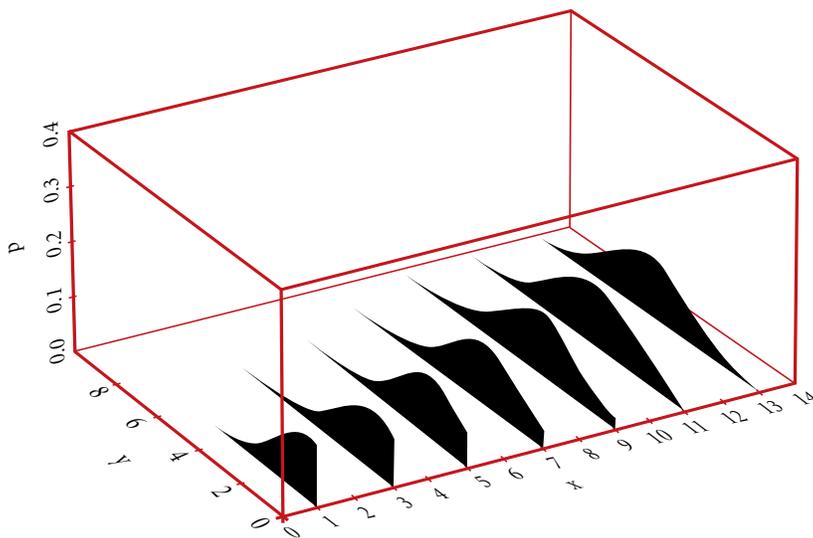


Figura 2.84 Histograma.



<Imagen 86, similar manteniendo proporciones y texto>

Figura 2.85 Gráfica con la evolución del peso de los bebés.



<Imagen 87, similar manteniendo proporciones y texto>

Figura 2.86 Gráfica en tres dimensiones.

Quienes publican gráficas estadísticas no siempre tienen el cuidado de seguir las recomendaciones de los expertos al elaborarlas y eso puede causar pérdidas de calidad en la información que se intenta transmitir. Estos problemas no son de carácter matemático, más bien caen en el ámbito de cómo se transmite la información. En particular, la estadística descriptiva hace muy poco uso —a veces ninguno— de la probabilidad o las matemáticas, a excepción de cálculos más o menos elementales.

En cambio, la estadística inferencial sí tiene un carácter matemático y depende en forma muy importante de modelos probabilísticos. Los métodos de la estadística inferencial son muchos y muy variados, así que aquí no vamos siquiera a intentar describirlos. Únicamente daremos un ejemplo, en el área de la medicina, para exhibir las ideas básicas que se aplican a la obtención de datos y el análisis estadístico de un conjunto de datos, además del tipo de conceptos matemáticos que para ello se emplean. Cabe mencionar que el desarrollo de los diferentes métodos de inferencia estadística ha dado lugar a profundas investigaciones matemáticas y existe, incluso, una rama de investigación —llamada estadística matemática— que se ocupa de este tipo de cuestiones.

Un planteamiento que se hacen frecuentemente los médicos es cómo saber si un nuevo medicamento es mejor que otro —utilizado por algún tiempo y con resultados conocidos—. Para ello, hay que hacer pruebas, de hecho, hay que “diseñar un experimento”. El diseño de experimentos es una parte esencial, pues la forma en que tomamos los datos determina el o los modelos estadísticos que se pueden utilizar para su posterior análisis. Al di-

señar un experimento es necesario cumplir algunas reglas y tomar varias decisiones importantes que determinarán, a fin de cuentas, la validez y confiabilidad de las conclusiones a las que se arribe. Cuando se diseñan experimentos con seres humanos —y más cuando se trata de su salud— es imprescindible que los individuos involucrados acepten participar, después de entender perfectamente en qué consiste el experimento y los riesgos que, como sujetos de estudio, corren.

Después, es necesario definir el alcance del experimento, lo que se llama la **muestra**. ¿Utilizaremos sólo sujetos enfermos de aquello que el medicamento en cuestión pretende curar? Supongamos que sí —primera decisión—. ¿Cómo se realizará el experimento? Si damos el medicamento a todos los sujetos de la muestra ¿qué vamos a medir para saber si el nuevo medicamento funciona mejor que el anterior? Digamos que mediremos la temperatura en cada individuo, antes de suministrar el medicamento y un tiempo después de haberlo suministrado, cualquiera que sea el medicamento suministrado. Con esta información podemos calcular el cambio de temperatura en cada individuo.

Aquí surge otro problema: el lograr que la muestra no sea sesgada al obtener, por ejemplo, una muestra de pacientes particularmente fuertes y que responden mejor a cualquier medicina que la media de los pacientes, o bien, una muestra de pacientes enfermos de varios días y que tienden a curarse más rápidamente que los que acaban de contraer la enfermedad, o al revés. Hay métodos recomendados para resolver este problema los cuales, normalmente, requieren que la elección de los sujetos se haga aleatoriamente —no aceptando, por ejemplo, sólo a los que se autopropongan para el experimento que, quizá, son gente con más ánimo y más fuertes que la media o personas que ya llevan varios días enfermas y están dispuestas a probar cualquier cosa.

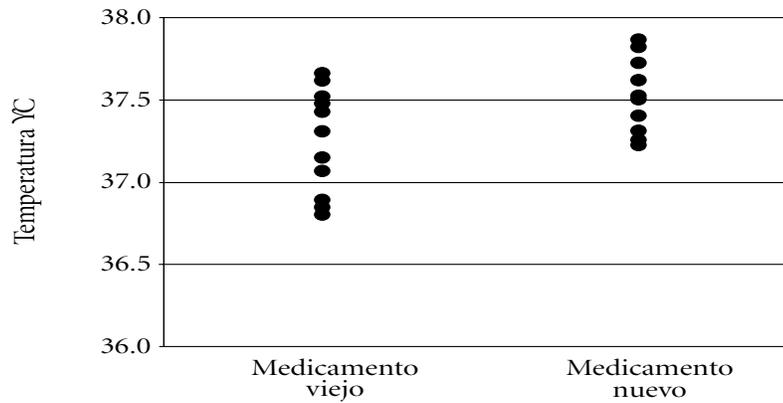
Además, ¿cuántas personas involucramos en el experimento? Cuanto más grande sea la muestra, más confiables serán los resultados, pero también será más caro el experimento y se tendrá que someter a un mayor número de personas a un tratamiento del cual aún se ignoran las posibles consecuencias negativas. Por lo anterior, el tamaño de la muestra es de los primeros problemas a resolver. Existen métodos estadísticos específicos para poder estimar el tamaño de muestra mínimo que puede darnos resultados suficientemente confiables de acuerdo con criterios más o menos universalmente aceptados. En particular, este es uno de los aportes importantes de la estadística a la investigación, pues nos permite obtener inferencias válidas y con cierto nivel de confianza y precisión al disminuir el costo de la investigación.

En un experimento de este tipo se consideran muchos otros detalles, pero vayamos a otra cosa. Supongamos que ya tenemos una muestra no sesgada de N enfermos que han aceptado responsablemente participar en el experimento y que lo llevaremos a cabo al administrar el medicamento a los enfermos, en periodos de tiempo preestablecidos, y al registrar su temperatura un número preestablecido de horas después.

Se pueden identificar tres aspectos fundamentales del diseño experimental: repetición, aleatorización y control. La teoría y los métodos estadísticos nos dicen cómo resolver estos principios en el momento de diseñar un experimento.

El número de repeticiones está determinado por la confiabilidad y precisión que deseamos del experimento, la forma de aleatorizar está definida por las características de nuestros individuos —para obtener una comparación justa entre ambos medicamentos— y, finalmente, los conocimientos biológicos junto con los del diseño experimental nos dicen cuáles factores tenemos que controlar y cómo hacerlo.

Ahora pasemos al análisis de los resultados y a realizar la inferencia deseada: ¿es mejor el nuevo medicamento? Recordemos que debemos llegar a una conclusión a partir de los datos del experimento. Supongamos que, con los resultados experimentales obtenidos, se genera la siguiente gráfica mostrada en la figura 2.87.



<Imagen 88, idéntica>

Figura 2.87 Gráfica de temperaturas con el medicamento viejo y el nuevo.

¿Podemos saber cuál de los dos medicamentos es mejor a partir de la gráfica? A simple vista, parecería que el nuevo no logró disminuir tanto la temperatura de los enfermos como el viejo pero, si miramos atentamente, veremos que hay una acumulación de varios resultados del nuevo medicamento —alrededor de los 37.25 grados— que parece estar cerca de la media de temperatura para el medicamento viejo. Por consiguiente, no resulta obvio cual de los dos medicamentos es mejor. Para obtener una respuesta clara debemos recurrir a una metodología específica que se denomina **prueba de hipótesis**.

Para hacer dicha prueba se plantea una hipótesis, que llamaremos **hipótesis nula** y la denotaremos como H_0 ; la prueba consiste en determinar una regla de decisión que nos llevará a “rechazar” o a “no rechazar” dicha hipótesis. Obsérvese que una hipótesis no se acepta, sólo se puede “rechazar” o “no rechazar”. La situación que describe esta regla de decisión se muestra en el siguiente cuadro:

Inferencia	H_0 en realidad es cierta	H_0 en realidad es falsa
Rechazamos la hipótesis H_0	Error tipo I con probabilidad	Inferencia correcta
No rechazamos la hipótesis H_0	Inferencia correcta	Error tipo I con probabilidad

Las probabilidades α y β indicadas en la tabla tienen significados precisos e importantes. $1-\alpha$ es el **grado de confianza** que nos da la prueba, mientras que $1-\beta$ se llama la **potencia** de la prueba. Para poder explicar esos conceptos en el caso del experimento de los medicamentos y comprender cómo α y β están relacionados con los datos, vamos a introducir la llamada distribución t de Student.

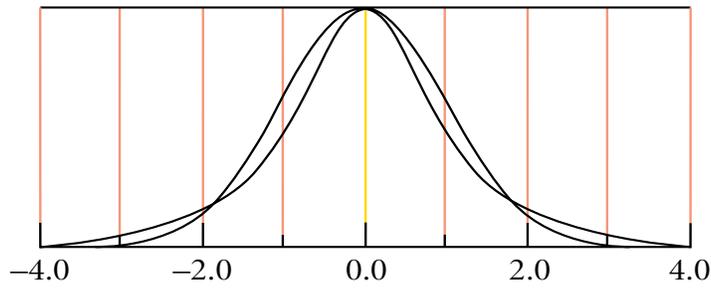
William S. Gosset fue un estadístico que publicó sus trabajos bajo el seudónimo de Student, debido a que la destilería Guinness —en donde trabajaba— había prohibido a sus empleados publicar artículos con cualquier tipo de contenido, debido al secreto industrial. Esta anécdota explica por qué una de las herramientas más utilizadas en la estadística se llama la distribución t de Student y no de Gosset —como sería de esperar ya que Gosset fue su creador. La distribución de Student aparece en esta situación: supongamos que tenemos una población que tiene una distribución normal con media μ y varianza σ^2 ; ahora consideremos una muestra de dicha población, de tamaño n , con media m y varianza s . Entonces la distribución de la variable aleatoria que se muestra a continuación:

$$t = \frac{m - \mu}{s / \sqrt{n}}$$

es la llamada distribución t de Student con $n - 1$ grados de libertad. Para valores grandes de n , la t de Student se parece mucho a la distribución normal, pero es precisamente en situaciones donde la muestra no es muy grande cuando esta distribución resulta muy útil, como en el ejemplo de los medicamentos.

<Imagen 89, similar manteniendo proporciones y textos>

Figura 2.88 La *t* de Student tiene las colas más pesadas que la normal y mientras más grados de libertad tiene, más se aproxima a la normal.



En este caso, es necesario usar la distribución *t* con $n_N + n_V - 2$ grados de libertad, donde n_N y n_V son los tamaños de las muestras para el medicamento nuevo y para el viejo, respectivamente, pues no se conocen las varianzas poblacionales y hay necesidad de estimarlas —no olvidemos que, si éstas fueran conocidas, se usaría la distribución normal—. Los grados de libertad de la *t* corresponden al número de observaciones independientes usadas para estimar la varianza poblacional. El procedimiento a seguir para completar la prueba de hipótesis, consiste en calcular la estadística de prueba:

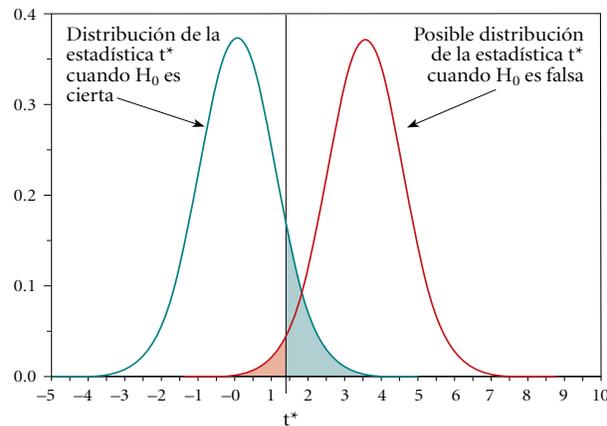
$$t^* = \frac{(\bar{X}_N - \bar{X}_V) - (\mu_N - \mu_V)}{\sqrt{\left(\frac{1}{n_N} + \frac{1}{n_V}\right) \frac{(n_N - 1)\hat{\sigma}_N^2 + (n_V - 1)\hat{\sigma}_V^2}{(n_N + n_V - 2)}}$$

al utilizar las dos muestras de individuos —los que tomaron el medicamento viejo y los que tomaron el nuevo. Si la hipótesis nula H_0 es cierta, esta estadística de prueba tendrá una distribución $t_{n_N + n_V - 2}$, donde intuitivamente esperaríamos que el valor de t^* fuera cercano a cero —ya que esta distribución estaría centrada en el cero y los valores con mayor densidad de probabilidad son cercanos a él. Si la hipótesis H_0 fuera falsa, la estadística t^* tendría una distribución que estaría a la derecha de la que se describió anteriormente y sería simétrica alrededor de un número mayor que cero. Mientras mayor sea la diferencia entre la efectividad de los medicamentos, esta gráfica estará más alejada del cero, α se iría haciendo más pequeña y $1 - \beta$, más grande.

En la figura 2.89 se muestran las gráficas de dichas distribuciones *t*. En el eje *x* se mide la diferencia de temperaturas entre ambos medicamentos, mientras la línea vertical representa el valor obtenido para la estadística de prueba t^* .

<Imagen 90, idéntica>

Figura 2.89 A la izquierda la distribución *t* correspondiente cuando H_0 es cierta y a la derecha la correspondiente una posible distribución cuando H_0 es falsa.



En otros experimentos puede convenir usar otras distribuciones —la normal, la χ^2 , etc.— para las pruebas de hipótesis, o bien, para otros procedimientos estadísticos. Lo que

es invariable es el hecho de que los resultados estadísticos siempre llevan alguna probabilidad de error. La estadística puede hacer recomendaciones pero, siempre, con cierto grado de confianza, nunca con certeza absoluta, aunque es una poderosa herramienta en la toma de decisiones en muchos campos de la actividad humana, como la industria, el comercio, la política y la ciencia. Por ello, su aplicación no debe tomarse a la ligera y conviene recurrir a los especialistas cuando las decisiones que de ella dependen son muy importantes.

LAS MATEMÁTICAS EN LA NATURALEZA

TEMA 3



Figura 3.1

3.1 INTRODUCCIÓN

Desde los inicios de la civilización, la comprensión de la naturaleza ha sido una de las preocupaciones fundamentales de los seres humanos. Es innegable que, a medida que el hombre ha ido entendiendo cómo funciona la naturaleza, ha logrado también mejorar su calidad de vida. Toda ciencia —en el fondo— está basada en principios físicos que subyacen en fenómenos tan diversos como la circulación sanguínea en un animal o las reacciones entre las sustancias químicas.

Durante el desarrollo de la ciencia, las matemáticas fueron vinculándose primero en forma tímida y después de una manera cada vez más clara e íntima. Por ejemplo, en la mecánica que es el estudio del movimiento de los cuerpos en función de las fuerzas que actúan sobre ellos. La mecánica es una teoría científica que no puede concebirse, ni siquiera expresarse con un mínimo de claridad, sin las matemáticas. Más adelante, otros campos de la física se desarrollaron siguiendo el mismo patrón de relación íntima con las matemáticas. De hecho, muchas de las ramas de las matemáticas se desarrollaron para poder dar cuenta de algunos fenómenos físicos, por ejemplo, el cálculo vectorial del que se hablará más adelante. El que la física se haya desarrollado en íntima relación con las matemáticas no parece ser un hecho casual o fortuito, más bien, las leyes de la física aparentan que son, realmente, de carácter matemático. Eugene Wigner, Premio Nobel de Física, dice que la ciencia no ha podido explicar por qué el universo tiene naturaleza matemática y eso es una grave laguna en el conocimiento humano. Pero no hay duda de que para comprender el universo se necesitan las matemáticas. De hecho, así como la creatividad de un pintor puede quedar limitada por una falta de habilidad en el manejo del pincel, así la del científico puede estar limitada si carece de una sólida formación matemática.

Los aspectos más interesantes de la relación entre las matemáticas y la física requieren de conceptos matemáticos avanzados, la mayoría de los cuales, no están al alcance del lector al cual nos dirigimos. No hemos querido ocultar esta complejidad, sino que hemos tratado de explicar en palabras simples los conceptos más importantes, mostrando al mismo tiempo la simbología matemática que se necesita para expresarlos correctamente —y que a más de un lector podrían asustar. Recomendamos al lector enfrentar este tema con tranquilidad y sentido del humor. Algunas fórmulas le parecerán incomprensibles, pero la intención de los autores no es espantarlo sino mostrarle el tipo de simbología que se usa en las matemáticas avanzadas y ofrecerle una explicación intuitiva de su significado. Para los físicos y matemáticos estas fórmulas con símbolos extraños manifiestan una belleza inaudita por su capacidad de expresar leyes de la naturaleza de una manera compacta y precisa, pero incluso su aspecto visual les resulta estéticamente atractivo. Éste es el caso en particular con las ecuaciones de Maxwell que aparecen como adornos en las playeras y tazas de café de ambientes académicos.

Este tema inicia con la simetría, uno de los aspectos matemáticos más importantes y a la vez, más fáciles de comprender de la naturaleza por su atractivo visual. Continúa con una descripción de las matemáticas relacionadas con el estudio del espacio y el movimiento, así como los grandes logros de la mecánica que permitieron explicar matemáticamente el movimiento planetario. Después, se muestra la versatilidad de la aplicación de las matemáticas en diversos temas de ciencias, distintos a la física. Finalmente, se plantea cómo la física y las matemáticas continuaron desarrollándose de la mano en temas de teoría electromagnética y física moderna y contemporánea.



Figura 3.2 Eugene Paul Wigner (1902-1995) fue un físico y matemático húngaro que recibió el Premio Nobel de Física en 1963 debido a sus contribuciones a la teoría del núcleo atómico y de las partículas elementales, en especial, por el descubrimiento y aplicación de los importantes principios de simetría.

3.2 LA SIMETRÍA EN LA NATURALEZA



Figura 3.3 El Nautilus captura la imaginación. Da la idea de movimiento, de crecimiento, de armonía, de perfección en el diseño, de simetría. Con unos cuantos trazos, nos lleva a pensar en lo infinitamente pequeño y lo infinitamente grande.

3.2.1 Algunos objetos simétricos

La simetría es una noción fundamental en el arte. Interviene en nuestra idea subjetiva de algo bello, pues surge de la relación que guardan las partes entre ellas y con el todo. Además,

en la naturaleza hay una multitud de ejemplos donde aparece la simetría, no por consideraciones estéticas, sino por eficiencia o economía en el diseño. Más aún, ha resultado ser un elemento indispensable al considerar los modelos más avanzados de la realidad física; es una especie de principio básico en la naturaleza cuya expresión formal es, ineludiblemente, matemática.

La inmensa mayoría de los animales vertebrados tiene simetría bilateral o de espejo. Esto quiere decir que, al partirlos por un plano imaginario, las dos mitades —izquierda y derecha— se corresponden como si estuvieran reflejadas en un espejo, no de manera estricta, sino en el diseño básico o ideal de la estructura ósea y muscular. Y esto es así pues presenta enormes ventajas evolutivas en el desplazamiento, el equilibrio, el control y el desarrollo —con la mitad de los datos y una instrucción “repítase en espejo”, se obtiene el todo.

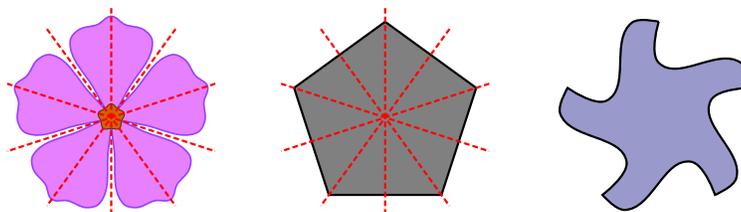
También abundantes en los seres vivos, aunque más escasas, existen simetrías con mayor complejidad. Por ejemplo, en muchas flores no sólo hay un espejo de simetría, sino varios.



Figura 3.4 De izquierda a derecha: flores con simetría de tres, cuatro, cinco y seis rotaciones y reflexiones, respectivamente.

Figura 3.5 Una flor con cinco pétalos tiene la misma simetría que un pentágono regular. Pero una “estrella ninja” de cinco picos sólo comparte con ellas las rotaciones y tiene la simetría de un pentágono con lados orientados cíclicamente.

En el esquema simple de una flor de la figura 3.5, aparecen cinco posibles espejos. La simetría de su diseño es la de un pentágono regular: al reflejarlo en cualquiera de esas cinco líneas, regresa a su lugar.



Cabe aclarar que *reflejar*, en el sentido matemático, es la transformación del plano o del espacio en sí mismo que intercambia los dos lados de una recta o un plano, llamado *espejo*, y mantiene las distancias con respecto a él. En el mundo real, un espejo actúa sobre una mitad del espacio y lo hace aparecer —como visto a través de una ventana— reflejado. Pero una reflexión matemática, actúa sobre ambos lados, intercambiándolos.

Composición e inversos de simetrías

Además de las cinco reflexiones, en la simetría pentagonal aparecen cinco rotaciones por ángulos múltiplos de $\frac{2\pi}{5}$ o 72° . Hay figuras planas que sólo tienen estas rotaciones como simetrías, como se muestra a la derecha de la figura 3.4. En la figura 3.6 se observa cómo podemos distinguir estas rotaciones; para ello, se rotularon las esquinas con números por lo que ahora es posible ver cómo se mueve la esquina número 1 bajo estas rotaciones.

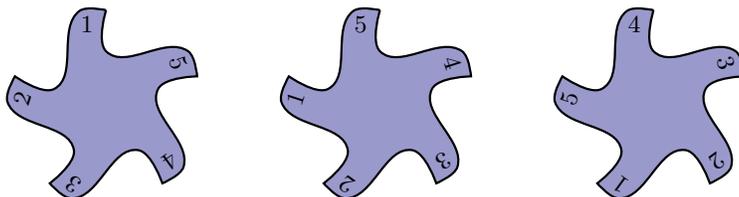


Figura 3.6 Rotaciones de una estrella ninja.

A la izquierda se muestra la posición inicial, en el centro está la posición después de rotar 72° y, a la derecha, después de moverse dos veces según dicho ángulo. Hay algo importante que debe observarse aquí —aunque parezca muy simple— y que es absolutamente crucial: la rotación de la estrella *ninja* a la derecha en la figura 3.6 puede obtenerse como una *rotación* de 144° o *por dos rotaciones consecutivas* de 72° . Esto se puede reformular de una manera más general: dos simetrías, es decir, dos movimientos que dejan invariante a una figura, pueden efectuarse uno tras otro para obtener, de nuevo, una simetría.

La figura 3.7 muestra la importancia del orden en que se efectúan dos simetrías. A la izquierda se numeran los pétalos y se muestra una rotación ρ y una reflexión σ —estas letras se leen “rho” y “sigma”, respectivamente—. La flor del centro muestra el efecto de primero reflejar (σ) y después rotar (ρ) —esto se expresa en notación de funciones como $\rho \circ \sigma$; y se lee “rho después de sigma”— y en la tercera imagen se muestra el efecto de primero rotar y después reflejar ($\sigma \circ \rho$). Por ejemplo, si rotamos con σ , el pétalo 1 queda en el lugar del 2 —podríamos escribir $\rho(1) = 2$, que se lee “rho de 1, o aplicado a 1, es igual a 2”—, y si después reflejamos con ρ el pétalo 1 queda finalmente en el lugar donde antes estaba el pétalo 5 —pues $\sigma(2) = 5$ y le da sentido a la notación $(\sigma \circ \rho)(1) = \sigma(\rho(1)) = \sigma(2) = 5$.

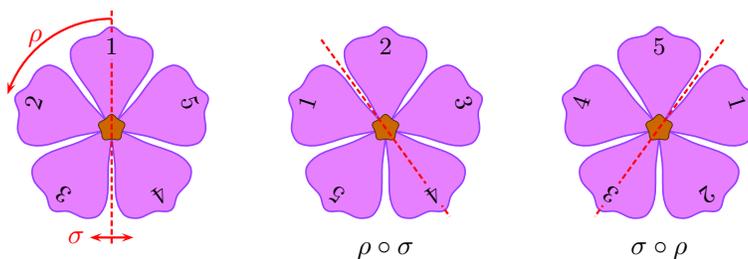


Figura 3.7 Composición de una reflexión (σ) y una rotación (ρ) en sus diferentes órdenes. En el centro, primero σ y luego ρ ; en el extremo derecho, primero ρ y después σ .

Se observa en la figura 3.7 —viendo el efecto en los pétalos numerados— que $\rho \circ \sigma$ es lo mismo que reflejar en el eje que pasa por el pétalo marcado con el número 4, mientras $\sigma \circ \rho$ equivale a reflejar en el eje que pasa por el pétalo 3. El proceso de efectuar primero una simetría y luego otra se llama *composición*. Y ahora sucede algo raro, si componemos la reflexión σ —de la figura 3.7— consigo misma, obtenemos un movimiento que no hace nada: cada pétalo regresa a su lugar. Al principio cuesta trabajo considerar al “no hacer nada” como un movimiento, pero es claramente un movimiento del plano que no altera las distancias. A este movimiento se le llama *identidad*, dado que funciona como el número 1 al multiplicar: deja todo como estaba.

También, cada simetría tiene su *inversa*: un movimiento como una reflexión o una rotación siempre tiene una reflexión —o rotación— correspondiente, que deshace el efecto

del movimiento original; para las reflexiones son ellas mismas. En general, dada una simetría, la que deshace su efecto se conoce como su inversa, muy similar al hecho de que $\frac{1}{2}$ es el inverso de 2, pues al multiplicar por $\frac{1}{2}$ se deshace el efecto de multiplicar por 2.

Por estas razones vemos que la identidad es una simetría importante y, sin ella, no podríamos componer dos simetrías cualesquiera ni podríamos considerar la formación de inversos. Las simetrías de una figura forman una estructura que se llama *grupo*. Los grupos tienen un elemento distinguido que es la identidad; además, sus elementos se pueden componer y se permite formar inversos.

El concepto detrás de la simetría

La simetría de muchas galaxias es sencilla: además de la identidad solamente hay una rotación de 180° ; muchas orquídeas también sólo tienen una simetría más, que es la reflexión en un plano vertical que atraviesa la planta a la mitad, como se observa en la figura 3.8.



Figura 3.7 Del lado izquierdo, la galaxia NGC 1365 con dos brazos principales; del lado derecho, una orquídea.

En ambos casos, hay dos simetrías: la identidad —que se abrevia con *id*— y otra que llamamos ρ y que tiene la propiedad de que:

$$\rho \circ \rho = \text{id}.$$

Las otras posibles composiciones son:

$$\rho \circ \text{id} = \rho, \quad \text{id} \circ \rho = \rho \quad \text{y} \quad \text{id} \circ \text{id} = \text{id}.$$

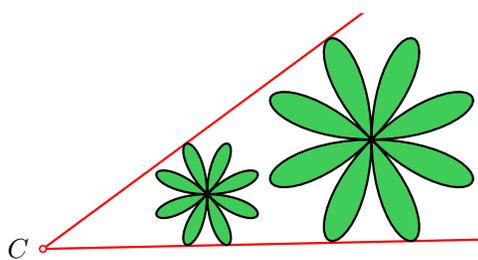
Estas últimas igualdades siempre se tienen pues *id* es como el 1 de la multiplicación: deja todo como estaba.

En conclusión, el grupo de simetría de la galaxia con dos brazos es “muy similar” al grupo de simetrías de la orquídea: tienen el mismo número de elementos y además, si éstos se denotan con *id* y ρ , entonces se componen de la misma manera. En matemáticas, se dice que dos grupos son *isomorfos* cuando, literalmente, tienen la misma estructura. A menudo, se identifican grupos isomorfos como si fueran iguales, en forma muy similar a como se pueden identificar un número de frijoles con un número de puntos de fichas de dominó o con la abstracción: el número. El grupo en abstracto es el que está detrás de la *representación* o *realización* concreta.

Las simetrías del Nautilus

El *Nautilus*, cuya imagen inició esta sección, no tiene simetrías en el sentido de que no hay movimientos del plano como reflexiones o rotaciones que lo dejen invariante, excepto —claro está— la identidad. Sin embargo, el *Nautilus* tiene una forma bella y, usualmente, nuestros ojos no mienten: donde hay belleza, hay cierta simetría. En efecto, así es: el *Nautilus* tiene muchas simetrías pero para verlas hay que relajar la noción de “transformación que deja invariante a la figura”, a que signifique no sólo *movimientos rígidos*, es decir, transformaciones del plano que no alteran las distancias.

En el *Nautilus* necesitamos también dilataciones, es decir, transformaciones que agrandan o achican. Para tener una transformación bien definida se requiere establecer el centro de la dilatación y el factor. En la figura 3.9 vemos el ejemplo de una dilatación o, como también se les dice, *homotecia*.



<Imagen 11>

Figura 3.9 El efecto de una dilatación con factor 2.

La espiral del Nautilus tiene la propiedad de que es invariante bajo ciertas dilataciones. En la figura 3.10 se muestra una espiral particular: sobre un rayo que sale del centro —marcado con rojo— se numeraron los puntos de la espiral con $\dots, C_{-1}, C_0, C_1, \dots$, por lo tanto, es una sucesión cuyo índice son los enteros. En esta espiral cada punto C_n está al doble de distancia al centro que el punto anterior C_{n-1} . Esto no depende de la dirección del rayo rojo y es igualmente cierto para otro rayo como, por ejemplo, el azul marcado en la misma figura. Por ello, una dilatación con factor 2 deja a la espiral invariante. El grupo que se forma con estas dilataciones es isomorfo al grupo de los enteros bajo la adición: al número entero n le hacemos corresponder aquella dilatación que envía el punto C_0 al punto C_n

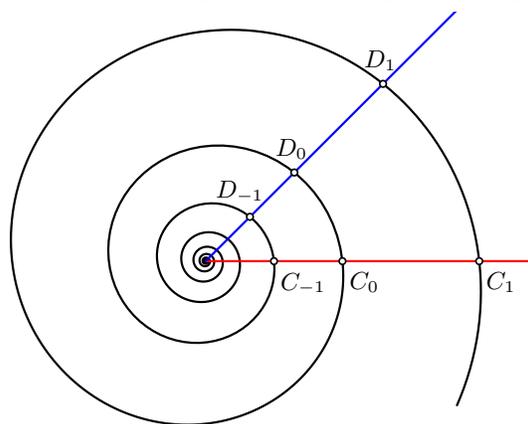


Figura 3.10 Una espiral logarítmica invariante bajo dilataciones con factor de potencias de 2.

. Pero éstas no son todas las simetrías de la espiral porque podemos combinar dilataciones con rotaciones. Si primero giramos de manera tal que el rayo rojo coincida con el azul y luego dilatamos por un factor apropiado, obtenemos otra simetría que manda C_0 a D_0 . El grupo de simetría de la espiral es isomorfo al grupo de los números reales positivos bajo la multiplicación: a cada número real positivo corresponde una simetría que está compuesta por una dilatación con este factor y una rotación apropiada. Para referencia posterior, usaremos $\mathbb{R}_{>0}$ para denotar a este grupo.

Si observamos de nuevo la imagen del *Nautilus*, destacan las paredes divisorias entre las cámaras internas. Si buscamos simetrías de la espiral que “respeten” también las paredes, entonces solamente ciertos factores de dilatación serían permitidos. Estos factores son de la forma ρ^i , es decir, son múltiplos enteros —con exponentes positivos o negativos— de un sólo factor ρ . Este grupo de simetría es como el grupo de los enteros bajo la suma.

Simetría de fórmulas

A primera vista, la simetría es un concepto que aparece meramente de la geometría. Veremos ahora que lo anterior no es sino un punto de vista demasiado limitado. Para ello, consideramos la ecuación:

$$xy - 1 = 0. \quad (1)$$

Esta ecuación es invariante bajo el intercambio de las dos variables x, y —dado que $yx = xy$. En consecuencia, si el par de números (x, y) satisface la ecuación (1), entonces también lo hace el par (y, x) . Esta transformación corresponde geoméricamente a una reflexión del plano en la diagonal $x = y$, así que esperamos que el conjunto de soluciones sea simétrico respecto a dicha reflexión. Denotemos a este intercambio como σ :

$$\sigma: \begin{cases} x \mapsto y, \\ y \mapsto x. \end{cases}$$

El conjunto de soluciones (x, y) describe una hipérbola con los ejes de coordenadas como asíntotas, según se muestra en la figura 3.11. En la misma figura se indica la recta $x = y$ con color rojo. Es —como esperábamos— un eje de simetría de la hipérbola azul. También se dibujó el otro eje de simetría $y = -x$.

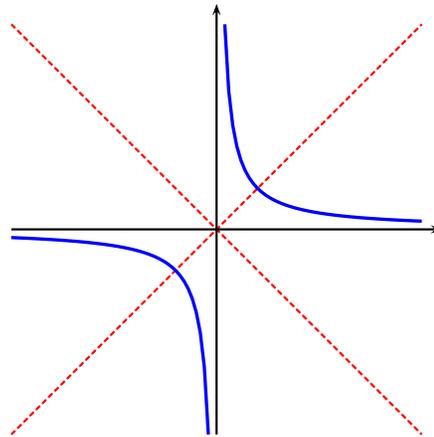


Figura 3.11 La hipérbola descrita por la ecuación $xy - 1 = 0$.

Reflejar en la recta $y = -x$ tiene el efecto de hacer que un punto con coordenadas (x, y) sea enviado al punto $(-y, -x)$, pues intercambia los ejes y les cambia orientación.

En otras palabras, la reflexión geométrica corresponde al intercambio de variables con cambio de signo:

$$\rho: \begin{cases} x \mapsto -y, \\ y \mapsto -x. \end{cases}$$

Dado que las simetrías forman un grupo, podemos componer las dos reflexiones y ob-

tenemos una rotación de 180° con centro en el origen del sistema de coordenadas. Esta operación corresponde al cambio de signo τ :

$$\tau = \rho \circ \sigma: \begin{cases} x \mapsto -x, \\ y \mapsto -y. \end{cases}$$

Junto con la identidad, hemos encontrado cuatro simetrías: id , σ , ρ , τ ; la siguiente tabla muestra como se componen estos elementos. Por ejemplo, en el segundo renglón y tercera columna —contando en lo blanco— está τ , que es el resultado de componer ρ con σ —justo los encabezados amarillos correspondientes. A esta tabla se le llama *tabla de multiplicación* y un grupo queda definido por ella.

\circ	id	ρ	σ	τ
id	id	ρ	σ	τ
ρ	ρ	id	τ	σ
σ	σ	τ	id	ρ
τ	τ	σ	ρ	id

Figura 3.12 Tabla de multiplicación que define un grupo.

Al grupo con esta tabla de multiplicación se le conoce como el *grupo de Klein*. Otra realización de este grupo se obtiene al considerar entre las permutaciones de cuatro elementos 1, 2, 3, 4, sólo aquellas que permuten dos pares: por ejemplo, intercambiar 1 con 2 y al mismo tiempo 3 con 4; esta permutación se denota por $(1\ 2)(3\ 4)$. Además de la identidad sólo hay 3 permutaciones así y éstas están determinadas por cómo se permute el número 1:

$$\rho = (1\ 2)(3\ 4), \quad \sigma = (1\ 3)(2\ 4), \quad \tau = (1\ 4)(2\ 3).$$

Para calcular una composición, como $\rho \circ \sigma$, es útil dibujar un diagrama como el que se muestra en la figura 3.12. Debe recordarse que $\rho \circ \sigma$ denota la permutación que se obtiene al aplicar, primero σ y luego, ρ , contrario a la lectura de la expresión de izquierda a derecha. Con este diagrama se verifica sin problema que $\rho \circ \sigma = \tau$ y, de manera similar, se puede comprobar que las cuatro permutaciones id , ρ , σ y τ satisfacen la misma tabla de multiplicación que las simetrías de $xy = 1$.

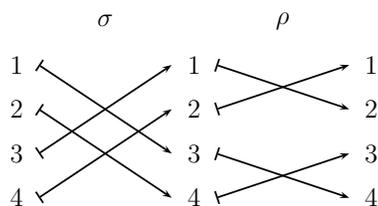


Figura 3.13 Diagrama que muestra cómo se permutan los elementos bajo la composición $\rho \circ \sigma$.

Antes de regresar a la hipérbola, vale la pena observar que el grupo de Klein también se realiza en el espacio como las simetrías rotacionales de la pelota de tenis.



Figura 3.14 Las rotaciones que dejan al diseño de una pelota de tenis en su lugar forman un grupo isomorfo al de Klein.

En la hipérbola, y como vimos con el *Nautilus*, no tenemos que restringirnos a las simetrías formadas por movimientos del plano que conservan la distancia. Para cada número real positivo, la transformación:

$$M_a: \begin{cases} x \mapsto ax, \\ y \mapsto \frac{1}{a}y \end{cases}$$

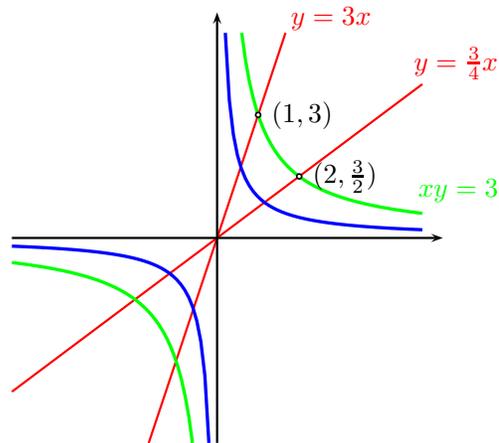
también deja invariante la ecuación de la hipérbola (1). El efecto geométrico de esta transformación es que envía un punto (x, y) al punto $(ax, \frac{1}{a}y)$. En particular, la recta $y = xy$ se envía a la recta $y = \frac{1}{a^2}x$. Por ejemplo, para $a = 2$, la transformación M_2 envía la recta $y = mx$ a la recta $y = \frac{m}{4}x$, mientras que todas las hipérbolas de la forma $xy = c$ se mantienen fijas.

El grupo formado por las transformaciones M_a es isomorfo al grupo $\mathbb{R}_{>0}$ ya que las transformaciones se componen según la siguiente ley:

$$M_a \circ M_b = M_{ab}.$$

Éste es el mismo grupo que el de las dilataciones del *Nautilus*. Si permitimos que a también puede ser negativo, es decir, cualquier número real excepto el cero, entonces obtenemos un nuevo grupo y, si consideramos también el intercambio de x y y , obtenemos uno todavía más grande.

Figura 3.15 Ejemplo de cómo actúa la transformación M_2 en el plano: el punto $(1, 3)$ es enviado al punto $(2, \frac{3}{2})$. Toda la recta $y = 3x$ es enviada a la recta $y = \frac{3}{4}x$, mientras la hipérbola $xy = 3$ permanece—como conjunto—fija.



De esta manera hemos visto que un mismo objeto puede exhibir diferentes grupos de simetría y que el mismo grupo puede aparecer en diferentes objetos.

Simetría conceptual

Existen palabras como *oso*, *rayar* o *somos* que pueden leerse de adelante hacia atrás o al revés y significan lo mismo. Estas palabras se llaman *palíndromos*. También hay frases que son palíndromos:

Anita lava la tina.

Los palíndromos tienen una simetría especial: son invariantes bajo la acción de voltearse. No se trata de una simetría geométrica, sino de cambiar de lugar letra por letra. Si aplicamos dos veces esta simetría obtenemos la identidad. El grupo es isomorfo al grupo de simetría de la galaxia y de la orquídea.

Es usual que recordemos la fórmula para calcular el área de un triángulo como “base por altura sobre dos”. Pero, ¿cuál de los tres lados es la base? ¿El que está más abajo? Estrictamente, un triángulo no tiene una base, sino tres. Cualquiera de los tres lados puede fungir como base y “la altura” será, entonces, la altura correspondiente. Por ello, la fórmula mencionada anteriormente no es simétrica en las tres variables, sino que requiere una elección: hay que elegir uno de los tres lados como base.

Si tenemos un triángulo con lados de 6, 9 y 7 centímetros, no podríamos calcular con esta fórmula el área de manera directa. Por otro lado, es claro que el triángulo queda completamente definido al especificar sus tres lados y que debería ser posible calcular su área. Lo anterior se puede resolver con trigonometría, según se hizo en la sección 2.7 o con la *fórmula de Herón*, que dice que:

$$A = \sqrt{s \cdot (s - a) \cdot (s - b) \cdot (s - c)} \quad (2)$$

es el área del triángulo con lados a , b y c si:

$$s = \frac{a + b + c}{2}$$

es el *semiperímetro*, es decir, la mitad del perímetro del triángulo. La fórmula de Herón es simétrica en los tres lados. Si permutamos los lados obtenemos una expresión que se reduce, al usar la conmutatividad de la multiplicación, a (2). Por otro lado, esta fórmula se puede evaluar simplemente: en nuestro ejemplo el semiperímetro es $s = 11$ cm y por lo tanto:

$$\begin{aligned} A &= \sqrt{11 \text{ cm} \cdot (11 \text{ cm} - 6 \text{ cm}) \cdot (11 \text{ cm} - 9 \text{ cm}) \cdot (11 \text{ cm} - 7 \text{ cm})} \\ &= \sqrt{440} \text{ cm}^2 \end{aligned}$$

Vale la pena reflexionar sobre cuáles de las fórmulas que usualmente se aprenden en bachillerato son simétricas. Por ejemplo, el Teorema de Pitágoras trata de triángulos rectángulos, cuyos lados a , b y c no son simétricos. Uno de ellos es la hipotenusa, los otros dos son los catetos. Si c es la hipotenusa, entonces el Teorema de Pitágoras afirma que:

$$c^2 = a^2 + b^2,$$

una fórmula que “es simétrica en a y b ”, es decir, podemos intercambiar a y b sin alterar la fórmula. Si el triángulo no es rectángulo hay una fórmula similar que toma en cuenta el valor del ángulo opuesto al lado c , que según la convención de Euler se denota por γ . Esta fórmula se llama *ley del coseno* y dice que:

$$c^2 = a^2 + b^2 - 2ab \cos(\gamma)$$

El coseno de 90° es cero, por lo que vemos que la ley de coseno se especializa en el Teorema de Pitágoras, o dicho de otra manera, la ley del coseno contiene como un caso particular al Teorema de Pitágoras. Nuevamente, la ley del coseno es simétrica en a y b .

Si f y g son dos funciones, entonces la derivada de la suma o del producto es simétrica en f y g :

$$\begin{aligned} (f + g)' &= f' + g' \\ (f \cdot g)' &= f \cdot g' + f' \cdot g \end{aligned}$$

donde denotamos la derivada de una función con una prima, es decir, f' es la derivada de f —aquí no importa la definición precisa de derivada sino las propiedades que debe heredar de las que tienen las funciones. Esta simetría corresponde a que $f + g = g + f$ y $f \cdot g = g \cdot f$, es decir, la conmutatividad de la suma y la multiplicación de funciones. Pero la derivada de la composición de funciones no es simétrica en f y g :

$$(g \circ f)' = (g' \circ f) \cdot f'$$

y ello corresponde al hecho de que la composición de funciones no es conmutativa: si $f(x) = x^2$ y $g(x) = 2x$, entonces $(f \circ g)(x) = (2x)^2$ mientras que $(g \circ f)(x) = 2x^2$, que no es lo mismo.

Simetría en la física

Las leyes de la física toman en cuenta conceptos como el espacio, el tiempo, la materia o la electricidad. Para expresar las cantidades se suelen usar coordenadas. Por ejemplo, la segunda Ley de Newton, dice que la fuerza que actúa sobre un cuerpo es igual al producto de su masa por la aceleración y se suele escribir como $F = ma$, donde F es la fuerza, m la masa y a la aceleración.

Pero en realidad se trata de una ley en el espacio: la fuerza y la aceleración son *vectores*, es decir tienen dirección y longitud. Si esto lo escribimos en nuestro sistema de coordenadas obtenemos tres ecuaciones:

$$\begin{aligned} F_x &= ma_x \\ F_y &= ma_y \\ F_z &= ma_z \end{aligned}$$

donde F_x, F_y, F_z son los componentes del vector F y similarmente a_x, a_y, a_z del vector a .

En principio, no es claro que estas leyes no cambien si tomamos otro sistema de coordenadas; por ejemplo, si lo rotamos por un ángulo α alrededor del eje de coordenadas z , la transformación es:

$$\begin{aligned} x' &= \cos(\alpha)x - \text{sen}(\alpha)y \\ y' &= \text{sen}(\alpha)x + \cos(\alpha)y \\ z' &= z \end{aligned}$$

donde x', y' y z' significan las coordenadas en el sistema girado. De forma similar se calculan los componentes del vector de la fuerza y la aceleración. Por ejemplo:

$$\begin{aligned} F_{x'} &= \cos(\alpha)F_x - \text{sen}(\alpha)F_y \\ &= \cos(\alpha)ma_x - \text{sen}(\alpha)ma_y \\ &= m(\cos(\alpha)a_x - \text{sen}(\alpha)a_y) \\ &= ma_{x'} \end{aligned}$$

De manera semejante, se obtiene $F_{y'} = ma_{y'}$ y $F_{z'} = ma_{z'}$. Con ello se muestra que la segunda Ley de Newton es invariante bajo rotaciones alrededor del eje de coordenadas z . Con argumentos similares se puede mostrar también la invariancia bajo rotaciones alrede-

dor de otros ejes y, por lo tanto, de una combinación de ellos. Éste es un ejemplo sencillo de una de las motivaciones que llevaron a Albert Einstein a proponer la Teoría de la Relatividad: que las leyes de la física deben ser las mismas en cualquier parte del universo y en todo momento; que al expresarlas como ecuaciones, estas deben ser simétricas respecto al movimiento en el espacio y en todo momento.

La simetría juega un papel importante en la física y, sobre todo en el siglo XX, llegó a ser una consideración sustancial para generar expectativas y hasta teorías enteras. Una simetría de cierta ecuación, la de Dirac, condujo a la predicción de la existencia del *positrón*, la anti-partícula del electrón. El concepto de grupo, en particular su abstracción, no sólo abrió camino a las matemáticas en el siglo XIX sino también a la física durante el siglo XX. Tiene un papel central en la frontera de la física, en el mundo cuántico de lo muy pequeño y en el relativista de lo muy grande. Se espera, incluso, que guíe la ruta de la buscada “Teoría del Todo”, una unificación de esas dos ramas claves de la física moderna. Para finalizar, hay que resaltar que el concepto de grupo, ese que subyace al de simetría y a las más modernas teorías físicas, surgió a principios del siglo XIX al atacar un problema de las mismas matemáticas: el resolver la ecuación de quinto grado.

3.3 ESPACIO, TIEMPO Y MOVIMIENTO

La relación entre la naturaleza y las matemáticas es tan estrecha que muchas veces resulta difícil separarlas. Por ejemplo, ¿qué tanto la geometría es un estudio del espacio físico en el que nos encontramos? Con sus conceptos y teoremas, la geometría proviene del esfuerzo de muchas generaciones para adquirir un modelo teórico que represente fielmente al espacio en que vivimos y sus propiedades. ¿Esto es física o matemáticas? Más bien, ambas.

La concepción moderna de las matemáticas que dominó el pensamiento durante el siglo XX, está basada en la idea de que éstas no tratan con ninguna realidad sino con objetos abstractos, las relaciones entre ellos —que se postulan como axiomas— y las consecuencias lógicas que pueden deducirse de estos últimos. Si aceptamos esta idea, la geometría, como parte de las matemáticas, no podría tratar sobre el espacio. Pero a lo largo de muchos siglos, el hombre intentó modelar y comprender su entorno apoyándose para ello en las matemáticas y concibiéndolas como el lenguaje adecuado para explicar las propiedades y el comportamiento del mundo que le rodea.

¿Existen los puntos? Según la geometría de Euclides, un punto es aquello que no tiene partes. ¿Alguien ha visto alguna vez un punto; no una pequeña mancha en el papel —que por muy pequeña que sea, consta de una zona del papel manchada por la tinta—, sino un punto de verdad que no conste de otras cosas? ¿Existen una recta y un círculo perfectos? Estas preguntas se las hacían los griegos y Platón las respondía diciendo que tales objetos matemáticos “perfectos” como el punto, la recta y el círculo, no son parte del mundo de los sentidos sino que habitan un mundo ideal al que el ser humano sólo tiene acceso mediante el pensamiento. Los objetos reales que señalamos diciendo que son puntos, rectas y circunferencias, no son más que ejemplos materiales imperfectos de aquellos objetos ideales.

El formalismo llegó aún más lejos que la geometría de Euclides en la construcción de una teoría de las matemáticas desligada del mundo material, que se revisa en la sección 4.7, al especificar que las matemáticas tratan con objetos abstractos ajenos al mundo material, es decir, que esos objetos ni siquiera deben intentar representarse por dibujos o algún otro medio, y que las palabras punto, recta y circunferencia, bien podrían cambiarse por otras como silla, corcho y sal, sin alterar en absoluto el contenido de la geometría, siempre y cuando los objetos abstractos designados por ellas guarden entre sí las relaciones postuladas, es decir, los axiomas.

Quizá el punto de vista formalista representa el sentir mayoritario de la comunidad académica matemática y, sin embargo, no explica la increíble eficacia de las matemáticas para modelar la realidad, sus componentes y sus leyes, como se verá en la sección 3.5. Tampoco explica la gran utilidad de las matemáticas para representar, plantear y resolver problemas prácticos en el ámbito de la actividad humana en general. El formalismo da cuenta de una parte de la actividad matemática pero ignora por completo las relaciones de las matemáticas con las ciencias y su utilidad social. En este capítulo incursionaremos en el mundo fascinante de las relaciones entre las matemáticas y el mundo físico, entendido como el espacio en el que vivimos, sus propiedades y las de los cuerpos materiales que lo habitan, el movimiento, las leyes que lo rigen, las fuerzas de la naturaleza y los conceptos matemáticos que el hombre ha desarrollado —específicamente— para modelar ese mundo físico.

3.3.1 El espacio

En general, modelamos el mundo en que vivimos como un espacio euclidiano de tres dimensiones. Aunque la Teoría de la Relatividad apunta en la dirección de que el espacio físico puede no ser euclidiano —podría tener cuatro dimensiones, ser curvo y quizás hasta ser finito—, nuestro modelo cotidiano del espacio físico es euclidiano de tres dimensiones e infinito. ¿Qué significa esto? Pues que la posición de cada objeto queda perfectamente especificada mediante tres números x , y , z , a partir de un sistema de coordenadas que podemos fijar arbitrariamente en cualquier sitio, con cualquier orientación y con cualquier unidad de medida.

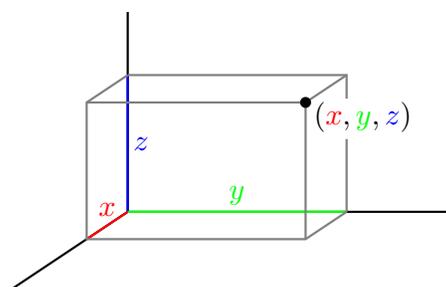


Figura 3.16 Un punto en el espacio se determina por sus coordenadas x , y , z , que son tres números reales.

Por supuesto, los números x , y , z que dan la posición de un punto, dependen de dónde se haya colocado el origen del sistema, cómo se haya orientado y qué unidad se haya elegido. Pero una vez definidas estas variables, la posición de cada punto queda determinada de manera inequívoca mediante una terna ordenada (x, y, z) de números. Que este modelo del espacio sea “correcto” o sólo una buena aproximación a la realidad es una cuestión que corresponde resolver a la física, pero el modelo es claro y se ajusta perfectamente a la realidad cotidiana y a la mayor parte del ámbito científico. Sólo haría falta modificar este modelo utilizando el de la Teoría de la Relatividad, cuando se trata con cuestiones del espacio astronómico donde las distancias, las velocidades y las fuerzas pueden ser mucho mayores que aquellas con las que lidiamos cada día aquí en la Tierra.

La idea de los sistemas de coordenadas proviene de René Descartes, que en su libro *La geometría* propone este modelo para el espacio físico. Descartes no hizo más que plantear el modelo explícitamente pues éste ya existía en la mente humana y se usaba diariamente desde la Antigüedad. Para que el modelo cartesiano quede claro es necesario suponer que contamos con un sistema de numeración que permite asignar números a las coordenadas de cada punto del espacio. En particular, esto requiere que contemos con un modelo de la recta que permita asignar a cada uno de sus puntos, un número.

El continuo espacio-tiempo: los números reales

Para modelar el continuo se utiliza el sistema numérico de los **números reales** que, si bien fue utilizado implícitamente por los matemáticos griegos, no se definió con precisión hasta el siglo XIX, muchos años después de que Descartes propusiera su modelo del espacio. Los hombres de ciencia y el público en general ya tenían —desde mucho tiempo atrás— una idea intuitiva del continuo, de que el espacio no puede ser modelado con algo discreto y finito sino que hace falta considerar que entre dos puntos hay una infinidad de otros puntos y que, por lo tanto, cada punto debe tener un tamaño nulo pues, de otra manera, se podría cubrir un intervalo con un número finito de puntos. Las ideas sobre el continuo provienen de la Antigüedad y siempre causaron controversias intelectuales, como las famosas paradojas de Zenón, según las cuales sería imposible el movimiento. Por ejemplo, una de ellas dice que:

Una persona no puede recorrer una cierta distancia, porque primero debe llegar a la mitad de ésta, antes a la mitad de la mitad y, antes aún, debería recorrer la mitad de la mitad de la mitad y así, indefinidamente.

De este modo, en teoría, una persona no puede recorrer una distancia, aunque la experiencia muestra que el movimiento sí es posible.

La paradoja se resuelve eliminando la hipótesis —implícita en el lenguaje— de que para realizar una infinidad de pasos es necesario un tiempo infinito. Todos los pasos pueden llevarse a cabo en un tiempo finito, porque cada uno es menor que el anterior y la suma de todos ellos es una cantidad finita. Por ejemplo, la suma infinita de un medio de una unidad, más la mitad de un medio de la unidad, más la mitad de la mitad de un medio de la unidad, etc., es igual a una unidad. En símbolos, se representa como:

$$\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = 1 \text{ y gráficamente como:}$$

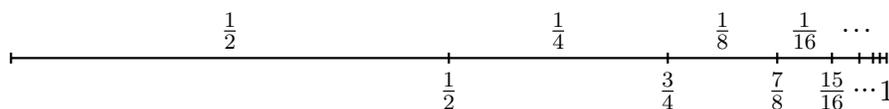


Figura 3.17 Suma de $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots = 1$.

Un poco de álgebra nos permite demostrar que:

$$\begin{aligned} \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots + \frac{1}{2^N} &= \frac{1}{2^1} + \frac{1}{2^2} + \dots + \frac{1}{2^N} \\ &= \frac{1}{2} \frac{1 - \frac{1}{2^N}}{1 - \frac{1}{2}} \\ &= 1 - \frac{1}{2^N} \end{aligned}$$

y, por lo tanto, todas las sumas $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots + \frac{1}{2^N}$ están acotadas por 1 y tienden a 1 cuando N tiende a infinito.

Como ya se dijo en el tema 2, concebimos a la línea recta como algo continuo, sin huecos, donde en cada intervalo existe una infinidad de puntos y donde, entre cada dos puntos, existen otros. Para representar todos los puntos de una línea recta usamos los números reales. Por lo general, representamos al continuo como una recta horizontal, elegimos un punto en ella que llamamos *origen* e identificamos con el cero y marcamos otro punto a la derecha que identificaremos con el uno. La distancia entre el origen y este segundo punto se toma como la unidad para medir distancias.

Figura 3.18 Unidad para medir distancias.



Dada esta construcción, podemos asignar un número real a cada punto P de la recta. Para ello, contamos el número de unidades a la derecha o a la izquierda que pueden colocarse desde el origen hasta el punto y asignamos este número entero, 2 en el caso ilustrado en la figura, como primera aproximación en la representación que haremos del punto P .

Figura 3.19 Primera aproximación en la representación que haremos del punto P .



A continuación, dividimos el intervalo de longitud uno en el que se encuentra el punto —en este caso, entre 2 y 3— en diez partes iguales y contamos cuántas de esas partes están a la izquierda del punto y así, definimos el primer decimal. En este caso, obtenemos 2.4 como segunda aproximación a P . Luego dividimos el intervalo de un décimo de unidad en el que se encuentra P , en diez partes iguales y contamos cuántas caben antes del punto, con lo cual definimos el segundo decimal. Continuamos este proceso indefinidamente para completar la definición del número real que representa al punto P .

La descripción contenida en el párrafo anterior es la clave del concepto de número real. Dice que el proceso continúa *indefinidamente*, es decir, que nunca termina. De acuerdo con nuestro lenguaje ordinario, esto parecería indicar que nunca terminaremos de definir el número. Sin embargo, con las instrucciones dadas el número ya queda perfectamente definido pues nos permiten escribir una aproximación decimal del punto tan precisa como deseemos y eso es precisamente lo que entendemos por número real. Preocuparse porque esta definición no ofrece una expresión infinita no nos lleva a ningún lado. En general, un punto de la recta requeriría, para ser representado mediante una expresión decimal, de una cadena infinitamente larga de dígitos. Sin embargo, como por lo general esto es imposible, no tiene caso intentarlo; para considerar que el número está definido, basta describir el proceso mediante el cual se puede encontrar cada uno de sus decimales.

La representación del continuo espacial mediante números nos obliga a la creación del concepto de número real como un proceso infinito de aproximación y no como un resultado que puede exhibirse explícitamente. Se trata de algo incómodo al principio, pero perfectamente correcto desde un punto de vista lógico y, por tanto, algo a lo que la mente humana puede habituarse y manipular sin ambigüedades. La creación de los números reales permite aplicar las matemáticas no sólo a la descripción del espacio físico, sino también a la del tiempo y al estudio del movimiento de los cuerpos en el espacio, lo cual puede considerarse como el origen de la ciencia cuantitativa.

Al igual que el espacio, el tiempo también requiere del concepto del continuo y se beneficia del concepto de número real. En efecto, la observación del movimiento nos indica una continuidad en el flujo del tiempo. Cuando algo se mueve no pasa instantáneamente de una posición a otra, sino que va cambiando continuamente su posición, pasando por todos los puntos intermedios y haciéndolo en instantes intermedios entre el inicio y el final. Es perfectamente concebible un mundo en que el tiempo fuera discreto, donde todo ocurriera como en los fotogramas de una película. Sin embargo, nuestra experiencia indica que si hubiera alguna granularidad en el tiempo, ésta sería tan pequeña que no podríamos distinguirla y, dado que el espacio parece continuo y ya nos acostumbramos a esa idea, resulta natural concebir al tiempo como un continuo y representarlo con números reales.

Para medir el tiempo también se elige un origen, es decir, un evento o acontecimiento que ocurre en un instante preciso, por ejemplo, el nacimiento de Cristo o el momento en que damos el disparo de salida para una competencia deportiva. También se requiere de una unidad, que debe ser el tiempo que pasa entre dos eventos repetibles. Hay una unidad

muy natural para el tiempo que casi todas las civilizaciones han tomado como unidad: el día o el lapso que tarda el Sol desde que asoma por el horizonte, hasta que vuelve a asomar por el mismo lugar al día siguiente. Las otras unidades de tiempo que se utilizan son fracciones enteras del día. La hora es la veinticuatroava parte del día, el minuto la sesentava parte de la hora, etcétera. Conviene observar que estas unidades de tiempo sólo tienen sentido en la Tierra; unos extraterrestres, si existieran, usarían con seguridad unidades diferentes para medir el tiempo y las distancias espaciales, simplemente porque no nos hemos puesto de acuerdo con ellos en una medida común. Las unidades son arbitrarias pero imprescindibles para aritmetizar el espacio y el tiempo. Las unidades de medida que puedan usar otros pueblos o civilizaciones, tendrán necesariamente una equivalencia con las nuestras. Quizá algunos extraterrestres usen como unidad del tiempo una que para nosotros equivalga, por ejemplo, a 32.27 horas y quizás algún “pueblo bárbaro” se empeñe en usar una unidad de distancia que equivalga a 1 609 de nuestros metros.

3.3.3 El movimiento

Los objetos pueden moverse, esto quiere decir que en cada momento, el cuerpo puede ocupar una posición diferente. Gracias a que podemos describir la posición y el tiempo con números, también podemos describir matemáticamente el movimiento. Consideremos un cuerpo relativamente pequeño, que suele llamarse partícula, que se mueve de una posición a otra en un intervalo de tiempo. Sean t_1 y t_2 los instantes de tiempo al inicio y al final del movimiento, y sean (x_1, y_1, z_1) y (x_2, y_2, z_2) las posiciones inicial y final de la partícula.

Esta situación puede representarse matemáticamente mediante una función del tiempo al espacio, concretamente, una función que asocia a cada instante de tiempo t una posición (x, y, z) . Una función así puede representar no sólo el inicio y el final del movimiento, sino todo el recorrido. Llamemos \vec{r} a la función que asigna a cada instante de tiempo t del intervalo $[t_1, t_2]$, la posición (x, y, z) de la partícula en ese instante. Entonces, esta función debe cumplir ciertas condiciones. En primer lugar $\vec{r}(t_1)$ debe ser (x_1, y_1, z_1) y $\vec{r}(t_2)$ debe ser (x_2, y_2, z_2) . En segundo lugar, la función \vec{r} debe ser **continua**, es decir, no debe dar saltos sino llevar el punto (x_1, y_1, z_1) al punto (x_2, y_2, z_2) a lo largo de una trayectoria continua, como el arco que se muestra en la figura.

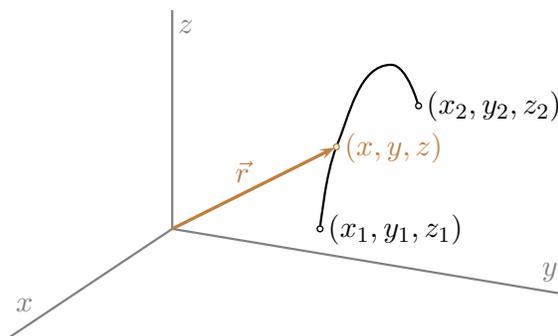


Figura 3.20 Una trayectoria en el espacio se determina por una función de tres coordenadas $\mathbf{r}(t) = (x(t), y(t), z(t))$, donde la variable t representa al tiempo.

Una función así puede construirse con tres funciones continuas $x(t)$, $y(t)$, $z(t)$ reales de variable real, definidas en el intervalo $[t_1, t_2]$. De esta manera es posible definir con toda precisión el movimiento de una partícula a través del espacio. La función —vectorial— \vec{r} , o bien las funciones escalares $x(t)$, $y(t)$, $z(t)$, determinan completamente el movimiento y se llaman ecuaciones paramétricas, en las que el parámetro es el tiempo.

Por ejemplo, el movimiento de una piedra lanzada desde el origen puede representarse con las siguientes funciones:

$$\begin{aligned}x(t) &= v_x \cdot t \\y(t) &= v_y \cdot t \\z(t) &= v_z \cdot t + \frac{g \cdot t^2}{2}\end{aligned}$$

donde t representa el tiempo transcurrido desde que se lanzó la piedra y v_x , v_y y v_z son las componentes de la velocidad inicial en las direcciones de los ejes coordenados. Otro ejemplo, serían las funciones:

$$\begin{aligned}x(t) &= r \cdot \cos(\omega t) \\y(t) &= r \cdot \text{sen}(\omega t) \\z(t) &= 0\end{aligned}$$

que representan un movimiento circular uniforme, a lo largo de una circunferencia de radio r que yace en el plano horizontal, de una partícula que gira a razón de ω radianes por unidad de tiempo.

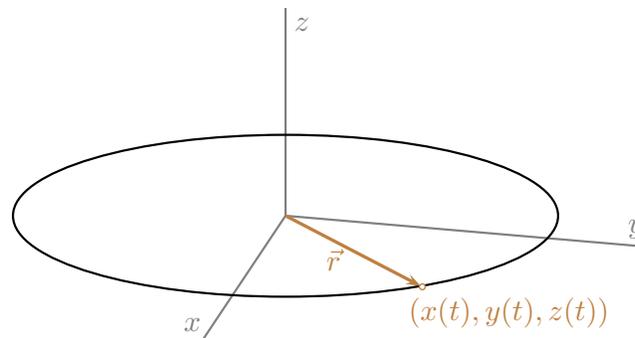


Figura 3.21 Trayectoria de un movimiento circular uniforme en el plano ω .

3.3.4 La velocidad y el concepto de derivada

Si conocemos las ecuaciones paramétricas del movimiento de un cuerpo, ¿podemos determinar su velocidad en cada instante? Esta pregunta, aparentemente sencilla, dio lugar a que Isaac Newton creara el cálculo diferencial. La velocidad de un cuerpo se define como el cociente de la distancia que cubre dividido por el tiempo que tarda en recorrerla. Ésta es una definición adecuada de velocidad para un cuerpo que se mueve con velocidad constante a lo largo de una línea recta. Sin embargo, para un cuerpo que se mueve a lo largo de una curva, el concepto de velocidad es más complicado por dos razones. En primer lugar, necesitamos que la velocidad indique no sólo cuán rápido se mueve sino en qué dirección lo hace; en segundo lugar, la velocidad puede ser distinta en distintos momentos durante el movimiento, por lo tanto, hace falta definir el concepto de velocidad instantánea. El concepto de velocidad requiere de una definición precisa y rigurosa que satisfaga la noción intuitiva que tenemos de ella.

La manera en que se define la **velocidad instantánea** $\mathbf{v}(t)$ en el instante t , es con un límite, específicamente, como el límite de los cocientes del desplazamiento del cuerpo en un intervalo de tiempo $[t, t + h]$, dividido entre el tamaño h de ese intervalo, cuando h tiende a cero. En símbolos, lo anterior se escribe así:

$$\mathbf{v}(t) = \lim_{h \rightarrow 0} \frac{\vec{r}(t+h) - \vec{r}(t)}{h}$$

Newton llamaba a esto la *flujo*n de \vec{r} ; hoy día se le llama la *derivada* de \vec{r} con respecto a t .

Dada una función cualquiera f de una variable x , la **derivada** de f en el punto x se denota por $f'(x)$ o por $\frac{df}{dx}$ y se define como:

$$f'(x) = \frac{df}{dx} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h},$$

siempre y cuando este límite exista. La definición es muy general y aplica cuando los valores de f son números reales, pero también cuando son vectores en el espacio de tres dimensiones e incluso, puede extenderse a funciones con valores en espacios mucho más generales —llamados espacios vectoriales topológicos.

El concepto de derivada de una función tiene dos aplicaciones principales, una es el que generó el propio concepto de derivada, es decir, la velocidad instantánea, y la otra es el que está relacionado con la gráfica de una función f , que tiene valores reales. La derivada de la función en un punto x viene a ser la *pendiente de la recta tangente a la gráfica de la función* en el punto $(x, f(x))$, como se indica en la figura 3.22.

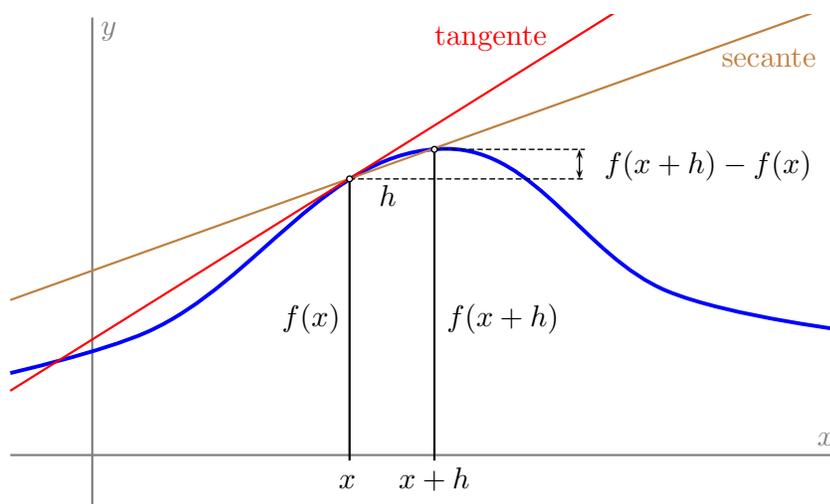


Figura 3.22 La derivada es la pendiente de la recta tangente a la gráfica de f en el punto $(x, f(x))$.

Esto se debe a que los cocientes:

$$\frac{f(x+h) - f(x)}{h}$$

son iguales a la pendiente de la recta —llamada secante—, que pasa por los puntos $(x, f(x))$ y $(x+h, f(x+h))$.

Para obtener las derivadas de las funciones, hay que calcular los límites correspondientes. Calculemos, por ejemplo, la derivada de x^3 :

$$\begin{aligned} \frac{dx^3}{dx} &= \lim_{h \rightarrow 0} \frac{(x+h)^3 - x^3}{h} \\ &= \lim_{h \rightarrow 0} \frac{3x^2h + 3xh^2 + h^3}{h} \\ &= \lim_{h \rightarrow 0} (3x^2 + 3xh + h^2) \\ &= 3x^2 \end{aligned}$$

Aunque el cálculo de la derivada de una función no suele ser muy complicado si uno tiene claro el concepto de límite, es conveniente memorizar las derivadas de las funciones más frecuentes, como las que aparecen en la siguiente tabla:

$$\begin{aligned}\frac{d c}{d x} &= 0 \\ \frac{d x}{d x} &= 1 \\ \frac{d x^2}{d x} &= 2x \\ \frac{d x^n}{d x} &= nx^{n-1}, \\ \frac{d \operatorname{sen}(x)}{d x} &= \cos(x) \\ \frac{d \cos(x)}{d x} &= -\operatorname{sen}(x)\end{aligned}$$

También es importante saber las reglas que se aplican al derivar sumas y productos de funciones. Sean f y g dos funciones cualesquiera, entonces:

$$\begin{aligned}\frac{d(cf)}{d x} &= c \frac{d f}{d x}, \text{ para cualquier constante } c \\ \frac{d(f+g)}{d x} &= \frac{d f}{d x} + \frac{d g}{d x} \\ \frac{d(fg)}{d x} &= f \frac{d g}{d x} + g \frac{d f}{d x}\end{aligned}$$

Con estas pocas fórmulas y reglas se pueden obtener las derivadas de muchas funciones. Los cursos de cálculo suelen dedicar bastante atención a que los alumnos aprendan a obtener derivadas. En ocasiones, se usan otras pocas fórmulas que aparecen en las tablas de derivación de los libros. Sabiendo derivar se pueden resolver muchos problemas como, por ejemplo, saber la velocidad instantánea de una partícula dada su ecuación de movimiento. En particular, si regresamos al ejemplo del movimiento circular uniforme descrito por las ecuaciones paramétricas:

$$\begin{aligned}x(t) &= r \cdot \cos(\omega t) \\ y(t) &= r \cdot \operatorname{sen}(\omega t) \\ z(t) &= 0\end{aligned}$$

podemos obtener las tres coordenadas del vector velocidad derivando las ecuaciones con respecto a t :

$$\begin{aligned}v_x(t) &= -r\omega \cdot \cos(\omega t) \\ v_y(t) &= r\omega \cdot \operatorname{sen}(\omega t) \\ v_z(t) &= 0\end{aligned}$$

y, si volvemos a derivar con respecto a t , obtenemos las componentes de lo que se llama el vector **aceleración**:

$$\begin{aligned}a_x(t) &= -r\omega^2 \cdot \operatorname{sen}(\omega t) \\ a_y(t) &= -r\omega^2 \cdot \cos(\omega t) \\ a_z(t) &= 0\end{aligned}$$

La aceleración y no la velocidad es lo que se relaciona directamente con la fuerza que se aplica a un cuerpo, esto es, el contenido de la segunda ley de Newton. Es interesante observar que el vector aceleración \vec{a} en este caso es proporcional al vector de posición \vec{r} , concretamente:

$$\vec{a} = -\omega^2 \vec{r}$$

En otras palabras, la aceleración de un cuerpo en movimiento circular uniforme es hacia el centro de su trayectoria circular y de magnitud igual al radio multiplicado por ω^2 , el cuadrado de la velocidad angular. Por lo tanto, si un cuerpo sigue un movimiento circular uniforme, es porque hay una fuerza \vec{F} —llamada *fuerza centrípeta*— que lo impulsa hacia el centro de la trayectoria:

$$\vec{F} = -m \cdot \omega^2 \vec{r}$$

Recordemos al lector que en el tema 1 de este libro hay una deducción de la tercera ley de Kepler basada en este resultado.

De manera similar se pueden obtener muchos resultados útiles sobre el movimiento de los cuerpos utilizando las derivadas.

3.3.5 La integral y el Teorema Fundamental del Cálculo

La segunda ley de Newton nos permite encontrar la trayectoria de una partícula si sabemos las fuerzas que actúan sobre ella. Sin embargo, esto no siempre resulta fácil. Poder resolver este tipo de problemas fue una de las principales motivaciones de Newton para desarrollar el cálculo integral. Si conocemos la fuerza que actúa sobre una partícula, por la segunda ley de Newton, sabemos cuáles son los valores de la aceleración, es decir, de la derivada de la velocidad con respecto al tiempo. Para obtener la velocidad de la partícula necesitamos resolver el problema inverso a la derivación, es decir, necesitamos encontrar la velocidad como una función del tiempo cuya derivada sea la aceleración que conocemos. El proceso inverso de la derivación se llama **integración**. Los matemáticos de la época de Newton y Leibniz, con ellos incluidos, se dieron cuenta de que la integración tenía mucho que ver con el cálculo de áreas y por ese motivo resulta conveniente comenzar el estudio de la integración con el concepto de área bajo la gráfica de una función, que se denomina **integral definida**.

La integral definida de una función entre dos puntos de la recta real a y b , se define mediante un límite:

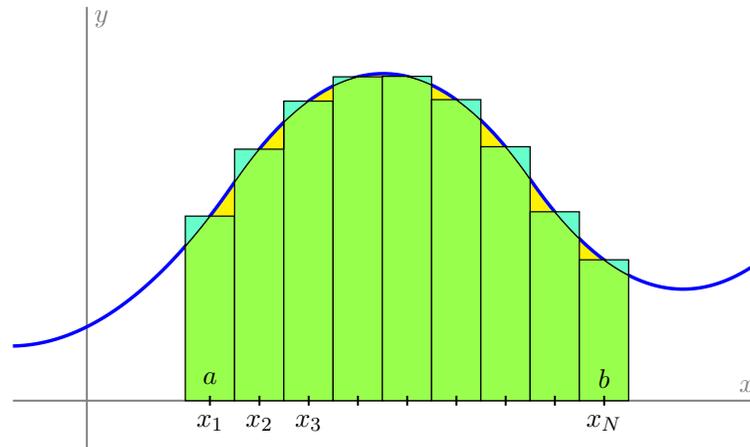
$$\int_a^b f(x)dx = \lim_{N \rightarrow \infty} \sum_{n=1}^N f(x_i) \Delta x \quad (3)$$

donde N es un número entero positivo, $\Delta x = \frac{b-a}{N}$ y los valores x_i son, por ejemplo, los puntos $a + (i + \frac{1}{2})\Delta x$ que están en medio de los intervalos de ancho Δx y donde también, se ha usado el símbolo \int llamado integral. La expresión $\int_a^b f(x)dx$ se lee “la integral entre a y b de f de x ”. Obsérvese que cuando N tiende a infinito, Δx tiende a cero y viceversa; en símbolos, $N \rightarrow \infty$ si y sólo si $\Delta x \rightarrow 0$.

El límite de (3) mide el área bajo la gráfica de $f(x)$ entre los puntos a y b , como se ilustra en la figura 3.23.

De la misma manera que el concepto de la derivada está íntimamente relacionado con su significado geométrico de ser la pendiente de la recta tangente a la gráfica de la función, la integral definida entre dos puntos a y b está íntimamente relacionada con el área delimitada por la gráfica, el eje x y los puntos a y b . De hecho, el área bajo la curva se define mediante la integral definida. Ambos conceptos, el de tangente a una curva y el de área bajo la curva sólo pueden definirse de manera rigurosa recurriendo a los conceptos de derivada e integral definida, respectivamente.

Figura 3.23 Aproximación del área bajo la curva mediante una función escalonada.



Dada una función $f(x)$ puede definirse otra con el área bajo la gráfica de la función entre un punto fijo a y una variable x . En símbolos, esta función de $F(x)$ se escribe así:

$$F(x) = \int_a^x f(s)ds$$

Las funciones definidas de esa manera tienen mucha importancia en las matemáticas. Por ejemplo, en la teoría de la probabilidad, si $f(x)$ representa la densidad de probabilidad de una variable aleatoria, entonces $F(x)$ es la probabilidad de que la variable tome valores en el intervalo $[a, x]$. En forma geométrica, la función $F(x)$ definida así representa el área bajo la gráfica de f desde a hasta x y, si se concibe como una función del extremo superior de integración x , puede derivarse al igual que cualquier otra función de x , sólo que en este caso el resultado es muy interesante:

$$\frac{dF}{dx} = f(x)$$

es decir, la derivación y la integración son procesos inversos uno del otro. Este es uno de los resultados más importantes de la historia de las matemáticas pues ha permitido hacer cálculos exactos de muchos procesos límite que, de otra manera, sólo hubieran podido obtenerse en forma aproximada. El nombre de cálculo que se da al estudio de las derivadas, las integrales y su íntima relación, se debe precisamente a que dicho estudio ofrece una metodología muy poderosa para realizar cálculos en campos muy diversos como la física, la geometría y la probabilidad.

El **Teorema Fundamental del Cálculo** exhibe la relación íntima que hay entre las integrales y las derivadas. Muestra que los procesos de integración y derivación son cada uno el inverso del otro. El teorema tiene dos versiones:

1] La derivada de la función $F(x) = \int_a^x f(s)ds$ es igual a $f(x)$. Es fácil entender por qué esto es cierto si aplicamos la definición de derivada a la función $F(x)$:

$$\begin{aligned} F'(x) &= \lim_{h \rightarrow 0} \frac{\int_a^{x+h} f(s)ds - \int_a^x f(s)ds}{h} \\ &= \lim_{h \rightarrow 0} \frac{\int_x^{x+h} f(s)ds}{h} = f(x) \end{aligned}$$

en donde hemos interpretado el primer numerador como una diferencia de áreas y lo hemos escrito como una integral definida entre x y $x + h$, y luego, hemos aprovechado el hecho de que cuando h es muy pequeño, $\int_x^{x+h} f(s)ds$ es casi igual a $f(x)h$.

La segunda forma del Teorema Fundamental dice que:

2] Si la derivada de una función $F(x)$ es otra función $f(x)$, entonces $\int_a^b f(x)dx = F(b) - F(a)$. Para entender intuitivamente por qué ocurre esto, conviene remplazar la integral por una suma ya muy cercana a ella, digamos $\sum_{n=1}^N f(x_i)\Delta x$, y sustituir $f(x_i)$ por el cociente:

$$\frac{F(x_i) - F(x_{i-1})}{\Delta x}$$

donde $x_i = a + i\Delta x$. La suma puede escribirse entonces como:

$$\sum_{n=1}^N f(x_i)\Delta x = \sum_{n=1}^N F(x_i) - F(x_{i-1}) = F(b) - F(a)$$

puesto que los términos $F(x_i)$ se cancelan por aparecer una vez con signo positivo y otra con signo negativo, así que sólo sobreviven los correspondientes a los extremos a y b .

Volvamos a considerar las funciones que se definen usando la integral así: $F(x) = \int_a^x f(x)dx$. Diferentes puntos fijos de inicio, digamos a_1 y a_2 , dan lugar a diferentes funciones $F_1(x) = \int_{a_1}^x f(s)ds$ y $F_2(x) = \int_{a_2}^x f(s)ds$ pero tales funciones difieren entre sí sólo por una constante, que es precisamente igual al área bajo la gráfica de la función entre los puntos a_1 y a_2 . Las funciones cuya derivada es $f(x)$ se llaman **primitivas** o **antiderivadas** de f . También se dice de ellas que son la **integral indefinida** de f , lo cual se escribe así:

$$\int f(x)dx$$

Las **tablas de integrales** muestran las funciones primitivas, antiderivadas o integrales indefinidas de varias funciones. La tabla siguiente es una pequeña muestra, donde c es una constante arbitraria:

$$\begin{aligned} \int 1dx &= x + c \\ \int xdx &= \frac{x^2}{2} + c \\ \int x^n dx &= \frac{x^{n+1}}{n+1} + c \\ \int \text{sen}(x)dx &= \text{cos}(x) + c \\ \int \text{cos}(x)dx &= -\text{sen}(x) + c \end{aligned}$$

Las tablas de integrales pueden obtenerse de las tablas de derivadas. Cada fórmula de derivación nos da una de integración. Dado que es mucho más fácil calcular las derivadas que las integrales cuando partimos de las definiciones como límites, el Teorema Fundamental nos resuelve de manera muy general un problema. Nos permite usar las tablas de integración para realizar cálculos que de otra manera serían mucho más complicados. Ésta es la razón por la que el cálculo es una herramienta tan poderosa. Por ejemplo, si intentáramos calcular el área bajo la gráfica de $f(x) = \text{cos}(x)$ entre $-\frac{\pi}{2}$ y $\frac{\pi}{2}$, sin el Teorema Fundamental tendríamos que calcular un límite muy difícil. Pero gracias al Teorema Fundamental sólo necesitamos encontrar la antiderivada de $f(x) = \text{cos}(x)$, la cual sabemos que es $\text{sen}(x)$, y por lo tanto:

$$\int_{-\pi/2}^{\pi/2} \text{cos}(x)dx = \text{sen}\left(\frac{\pi}{2}\right) - \text{sen}\left(-\frac{\pi}{2}\right) = 1 - (-1) = 2$$

Aunque no todos los cálculos que se realizan usando la integral son tan sencillos, el Teorema Fundamental permite muchas veces llegar a resultados exactos que sin él serían poco menos que imposibles.

Newton y las leyes de Kepler, una nueva concepción del Universo

Como ya vimos previamente, los conceptos del cálculo están íntimamente relacionados con la física. La velocidad no puede definirse con rigor sin la derivada mientras que el cálculo de la trayectoria de un cuerpo requiere del uso de los poderosos métodos del cálculo.

La primera y más impactante aplicación del cálculo a la física fue la deducción matemática de las leyes del movimiento de los planetas, es decir, de las leyes de Kepler, a partir de la segunda ley de Newton:

$$\vec{F} = m\vec{a}$$

así como de la ley de la gravitación universal, en forma vectorial:

$$F = G \frac{mM}{r^2}$$

Usando los métodos del cálculo se puede demostrar que, si suponemos que la fuerza que actúa sobre un cuerpo celeste es la fuerza gravitatoria del Sol, entonces su trayectoria será una curva cónica —elipse, parábola o hipérbola— que tendrá al Sol en uno de sus focos. Los planetas, como giran alrededor del Sol, no pueden tener órbitas parabólicas o hiperbólicas, por lo tanto sus órbitas son elípticas: segunda ley de Kepler. También, a partir de la segunda ley de Kepler se obtiene que el radio vector del Sol al cuerpo barrerá áreas iguales en tiempos iguales. La tercera ley de Kepler se puede deducir de igual forma con absoluto rigor: los cuadrados de los tiempos de revolución de los planetas alrededor del Sol son proporcionales a los cubos de los semiejes mayores de sus órbitas elípticas.

No vamos a mostrar cómo se usa el cálculo para deducir las leyes de Kepler, tal demostración puede encontrarse en los libros avanzados. Pero es importante que quede claro al lector que se trata de resultados de una extraordinaria importancia pues llevan al hombre a cambiar su concepción del universo, le hacen dar el paso definitivo entre la superstición y la ciencia. Y en este paso las matemáticas jugaron un papel central.

El universo parece comportarse siguiendo unas leyes de naturaleza matemática y todos los detalles del movimiento de los cuerpos pueden deducirse matemáticamente de estas leyes. Éste fue uno de los descubrimientos científicos más importantes de la historia, si no es que el más importante. Se trata de un descubrimiento que pone a la razón y en particular a las matemáticas en el gobierno de la naturaleza, con lo que la existencia de deidades que se encargan de decidir todo lo que ocurre, deja de ser lógicamente necesaria. El universo tiene leyes racionales, leyes que se expresan con fórmulas matemáticas y que determinan todo lo que ocurre. La religión, que se había adueñado por muchos siglos del poder sobre el universo recibió un duro golpe cuando se descubrió que las maravillas de la naturaleza podían explicarse, en principio, a partir de fórmulas y por lo tanto, no era necesario recurrir a un sacerdote —supuestamente en contacto privilegiado con los dioses— para cambiar el curso de los acontecimientos, sino que se podía actuar directamente sobre las causas para producir los efectos deseados. Dejan de ser necesarios intermediarios para influir sobre el mundo, ahora ya sólo bastan la ciencia y la tecnología.

A medida que los descubrimientos de Newton se fueron popularizando y alcanzaron a las personas interesadas en la política, se convirtieron en herramientas de liberación. La Ilustración y por tanto la Revolución francesa, madre de todas las revoluciones sociales posteriores, deben mucho a la concepción del universo que nos legó Newton. El cálculo juega un papel fundamental en esta historia, los propios conceptos que permitieron entender el universo son entes matemáticos.

La siguiente sección puede dar una idea al lector de por qué la segunda ley de Newton y la ley de la gravitación universal llevan a las leyes de Kepler. No se trata de una deducción rigurosa en la que habría que usar el cálculo, sino de un divertimento geométrico que hace plausible la deducción.

3.4 LAS ÓRBITAS CELESTES



Figura 3.24 Richard Feynman, físico estadounidense que ganó el Premio Nobel por su trabajo sobre la unión de dos áreas de la física: la mecánica cuántica y la electrodinámica. Fue un hombre con extraordinario talento no sólo en física como se puede constatar en libros autobiográficos.

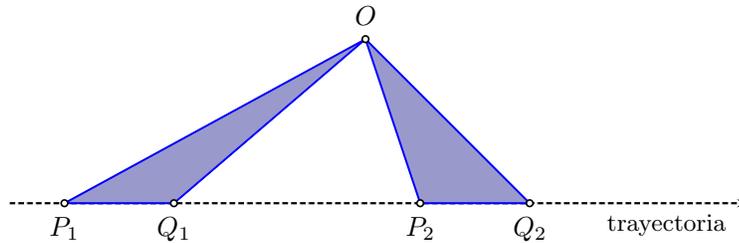
Como revisamos en la sección 1.3, Newton desarrolló el cálculo diferencial para estudiar el movimiento, el fenómeno de cambios continuos, y explicó de esta manera, con base en principios básicos, las leyes de Kepler. Aquí haremos algo distinto, pero no menos interesante: se discutirá cómo se pueden hacer verosímiles las leyes de Kepler a partir de principios geométricos. El desarrollo se basa en una idea del físico Richard Feynman. Es importante resaltar que se trata de consideraciones verosímiles y no de una demostración exacta; por ello, en casi todos lados tendremos aproximaciones —y usaremos el símbolo \approx para denotarlo— y no igualdades exactas.

Un cuerpo que se mueve en el espacio sin ninguna atracción describe una trayectoria recta y se traslada distancias iguales en tiempos iguales. Es decir, en cada momento avanza la misma distancia respecto a un observador que se piensa inmóvil. Podemos observar que, en este caso, se cumple lo que enuncia la segunda ley de Kepler: el radio vector del Sol a un planeta barre áreas iguales en tiempos iguales.

Para cualquier punto del espacio O , el área del triángulo P_1Q_1O es igual al área del triángulo P_2Q_2O si P_1, P_2 marcan dos posiciones en la trayectoria y Q_1, Q_2 son las posiciones respectivas después de un intervalo Δt de tiempo.

Ahora analizamos un caso más complejo en el que, en algún momento, un asteroide

Figura 3.25 Ilustración de la segunda ley de Kepler en el movimiento uniforme.

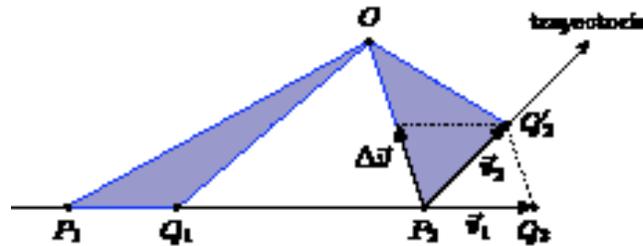


golpea el cuerpo, lo que produce un cambio casi instantáneo en la dirección del movimiento, es decir, en la velocidad. Obviamente, el cambio en la velocidad depende de la “intensidad” con la que golpeó el cuerpo y de la dirección del golpe. Pero aquí no nos interesa demasiado analizar el mecanismo del golpe, sino el efecto en la velocidad. Podemos simplificar el fenómeno y pensar que, en este momento, la velocidad se altera por la diferencia $\Delta\vec{v}$, es decir, la velocidad posterior al impacto \vec{v}_2 se calcula a partir de la velocidad anterior \vec{v}_1 :

$$\vec{v}_2 = \vec{v}_1 + \Delta\vec{v}$$

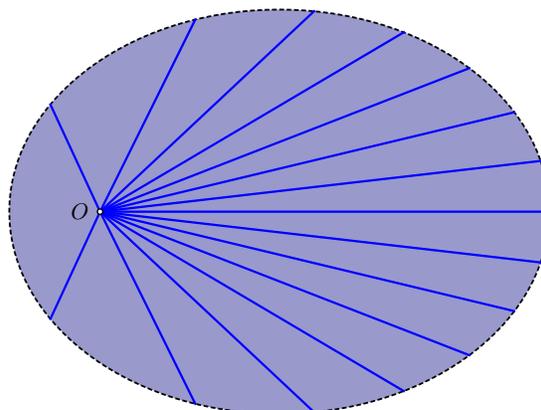
Sabemos que las velocidades se suman formando el paralelogramo correspondiente, según se ve en la figura 3.26. De nuevo, nos preguntamos si es cierto que se cumple la segunda ley de Kepler para cualquier observador, es decir, si para cualquier punto O se tiene que el vector que une O con el planeta barre áreas iguales en tiempos iguales. La respuesta es negativa pues no se cumple para cualquier punto O , pero sí lo hace para puntos que están sobre la línea que marca el vector $\Delta\vec{v}$ desde el punto P_2 del impacto (véase figura 3.26).

Figura 3.26 En el punto P_2 , el asteroide golpea al planeta y hay una desviación en la trayectoria. El cambio de la velocidad $\Delta\vec{v}$ se suma con la velocidad anterior \vec{v}_1 para dar como resultado la velocidad posterior $\vec{v}_2 = \vec{v}_1 + \Delta\vec{v}$. Esta suma se forma al usar el paralelogramo con lados $\Delta\vec{v}$ y \vec{v}_1 .



Para simplificar el dibujo supusimos que el impacto se realizó justo en la posición P_2 . En vez de que el cuerpo siga hacia la posición Q_2 , éste ahora se encuentra en Q'_2 después del mismo lapso Δt . Los dos triángulos P_2Q_2O y $P_2Q'_2O$ comparten el lado OP_2 y, además, tienen la misma altura —respecto a OP_2 — dado que $Q_2Q'_2$ es paralelo a OP_2 , porque $\Delta\vec{v}$ está dirigido a O en el momento del impacto.

Figura 3.27 La partición de una órbita en intervalos de tiempos iguales, o lo que es lo mismo, en áreas “barridas” iguales.



En las órbitas celestes se supone que la masa del Sol es muy grande y, por lo tanto, éste no se mueve y sólo atrae a los planetas hacia sí mismo. En realidad, también el Sol es atraído hacia los planetas pero esta atracción tiene un efecto pequeño sobre él y, en una primera aproximación, no hay que tomarlo en cuenta. Si nos imaginamos la fuerza gravitacional ejercida por el Sol sobre los planetas como una sucesión de pequeños golpes, cada uno de estos golpes produce en un cambio de velocidad dirigido hacia el Sol. Por lo que vimos antes con el asteroide, esto tiene como consecuencia que el radio del Sol a la Tierra barre áreas iguales en tiempos iguales, según se muestra en la figura 3.27. Lo único que se usó hasta este momento en el argumento fue que tenemos una *atracción central*, es decir que la fuerza siempre se dirige al mismo punto O .

En realidad, la fuerza es continua y se ejerce todo el tiempo, pero imaginar que se trata de pequeños impactos nos permite aplicar la geometría y arribar a la segunda ley de Kepler de manera intuitiva y clara sin tener que usar herramientas matemáticas más avanzadas.

Ahora haremos una división diferente de la órbita. La dividimos en partes, llamadas sectores, tal que cada una de ellas incluya el mismo ángulo $\Delta\alpha$ en el punto O , como se muestra en la figura 3.28.

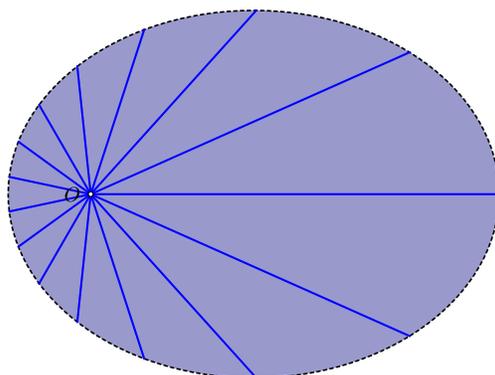


Figura 3.28 La ilustración muestra una partición de una órbita en sectores del mismo ángulo O .

Por lo que vimos con anterioridad, el área es proporcional al tiempo transcurrido en la órbita. Regresamos de nuevo al punto de vista de que el cuerpo recibe impulsos instantáneos en ciertos momentos, que ahora ya no serán después de cierto tiempo sino cada vez que se cubre cierto ángulo. Esto permitirá emplear nuevamente conceptos de la geometría. La figura 3.29 muestra cómo cambian las velocidades $\vec{v}_0, \vec{v}_1, \vec{v}_2 \dots$

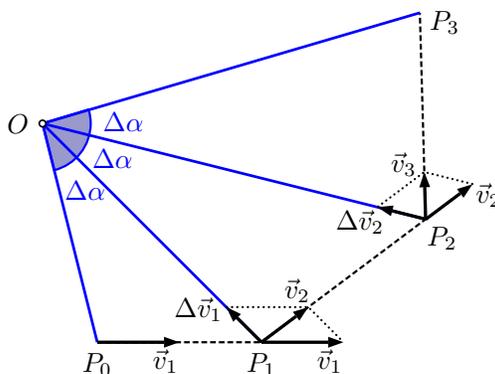


Figura 3.29 Los cambios de velocidades son constantes $|\Delta\vec{v}_1| \approx |\Delta\vec{v}_2|$.

Veremos que los cambios de velocidades $\Delta\vec{v}_1 \approx \vec{v}_2 - \vec{v}_1, \Delta\vec{v}_2 \approx \vec{v}_3 - \vec{v}_2, \dots$ son iguales si estos cambios ocurren cada vez que se gira la órbita por el mismo ángulo. En otras palabras, se tiene $|\Delta\vec{v}_1| \approx |\Delta\vec{v}_2|$, donde las barras verticales indican que se considera solamente la longitud del vector, pero no la dirección. La aproximación tiende a ser una igualdad conforme el ángulo $\Delta\alpha$ se hace pequeño.

Para ello, necesitaremos por primera vez que la fuerza gravitacional que provoca la atracción sea proporcional al inverso del cuadrado de la distancia. Según la ley fundamental de Newton, la fuerza \vec{F} es igual a la masa por la aceleración:

$$\vec{F} = m\vec{a}$$

donde m es la masa del cuerpo y \vec{a} es la aceleración. Esta última es el cambio de velocidad. En nuestro modelo de golpes, la aceleración en el punto P_1 es:

$$\vec{a} \approx \frac{\Delta\vec{v}_1}{\Delta t_1}$$

donde Δt_1 es el tiempo que transcurre para que llegue el cuerpo del punto P_0 al punto P_1 . Como la masa es constante y la fuerza proporcional al cuadrado de la distancia $r_1 = OP_1$, se obtiene que:

$$\frac{|\Delta\vec{v}_1|}{\Delta t_1} \propto \frac{1}{r_1^2}.$$

Aquí, el símbolo \propto expresa la proporcionalidad, esto es que:

$$\frac{|\Delta\vec{v}_1|}{\Delta t_1} \approx k \frac{1}{r_1^2}, \quad \frac{|\Delta\vec{v}_2|}{\Delta t_2} \approx k \frac{1}{r_2^2}, \quad \frac{|\Delta\vec{v}_3|}{\Delta t_3} \approx k \frac{1}{r_3^2}, \dots \quad (4)$$

siempre con la misma constante de proporcionalidad k .

Por otro lado, el área ΔA_1 del triángulo OP_0P_1 es $\Delta A_1 \approx \frac{1}{2r_1^2} \Delta\alpha$, donde se usa que $\text{sen}(\Delta\alpha) \approx \Delta\alpha$ es una buena aproximación para ángulos pequeños. En particular, el área ΔA_1 es proporcional al cuadrado de la distancia r_1 , si el ángulo $\Delta\alpha$ se mantiene constante. Pero, por lo que vimos antes, esta área es también proporcional al tiempo Δt_1 . En conclusión, si $\Delta\alpha$ es constante, entonces:

$$\frac{\Delta t_1}{r_1^2} \approx \frac{\Delta t_2}{r_2^2} \approx \frac{\Delta t_3}{r_3^2} \approx \dots$$

y de aquí sigue que a partir de las ecuaciones (4), tenemos que:

$$|\Delta\vec{v}_1| \approx |\Delta\vec{v}_2| \approx |\Delta\vec{v}_3| \approx \dots$$

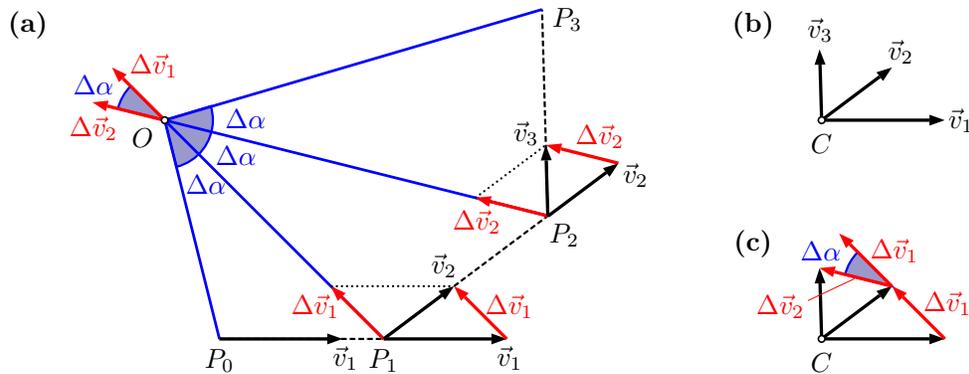


Figura 3.30 Construcción del hodograma.

Ahora podemos dibujar el diagrama de las velocidades que se llama *hodograma*. Éste se obtiene al trasladar todas las velocidades que tiene el cuerpo a lo largo de su trayectoria de manera paralela, tal que todos los vectores empiecen en un mismo punto C , como se observa en la parte (b) de la figura 3.30. Se debe notar que los vectores \vec{v}_1 en la parte (a) y (b)

son paralelos. El vector que une la punta de \vec{v}_1 con la de \vec{v}_2 es $\Delta\vec{v}_1$, dado que $\vec{v}_2 \approx \vec{v}_1 + \Delta\vec{v}_1$, como se observa en la parte baja de la figura 3.30(a). En forma similar, $\Delta\vec{v}_2$ une la punta de \vec{v}_2 con la de \vec{v}_3 .

Finalmente, a partir de todo lo que se ha argumentado se puede llegar a la siguiente conclusión: los vectores $\Delta\vec{v}_1, \Delta\vec{v}_2, \dots$ son todos igual de largos y además, difieren en su dirección siempre por el mismo ángulo $\Delta\alpha$, según se puede observar si se trasladan hasta el punto O ; véase parte (a) en la figura 3.30. El resultado se muestra en la parte (c) de la misma figura. La conclusión es que la aproximación del hodograma formado por las puntas de los vectores $\vec{v}_1, \vec{v}_2, \vec{v}_3, \dots$ es un polígono regular, dado que su contorno está formado por los vectores $\Delta\vec{v}_1, \Delta\vec{v}_2, \dots$ que, como ya dijimos, son igual de largos y siempre incluyen el mismo ángulo.

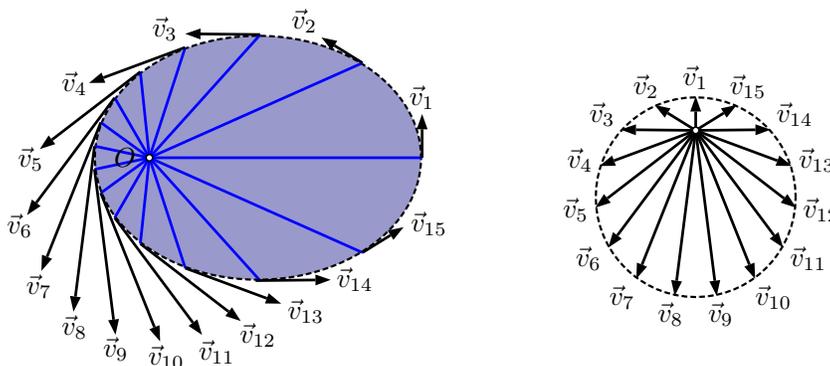


Figura 3.31 El hodograma de una órbita elíptica.

Al hacer más pequeño el ángulo $\Delta\alpha$, el polígono regular se aproximará a una circunferencia. Concluimos que el hodograma es una circunferencia si el cuerpo se mueve en un campo de atracción central con una fuerza que es inversamente proporcional al cuadrado de la distancia. Es importante observar que el punto C —el “origen” del hodograma— no es el centro de la circunferencia. Esto se puede ver en la figura 3.31. Pero todavía no hemos demostrado que las órbitas sean elípticas y la figura 3.31 es, meramente, una ilustración.

A cada punto P de la órbita le corresponde un punto del hodograma P' , que se marca por la punta del vector de velocidad \vec{v}_P , que es la velocidad que tiene el cuerpo en el punto P . Por ejemplo, si P es el punto más a la derecha de la órbita en la figura 3.31, entonces P' es el punto más arriba en el hodograma.

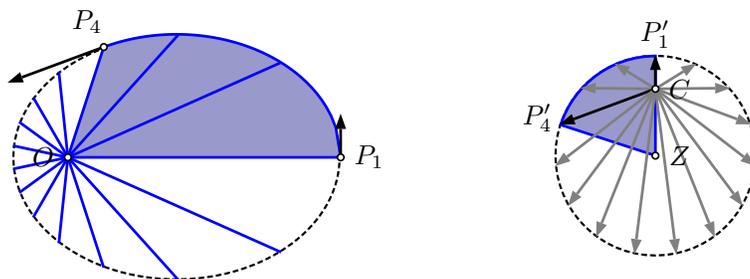


Figura 3.32 La correspondencia de ángulos β .

Hemos visto que si dividimos la órbita en sectores de ángulos iguales, los puntos correspondientes en el hodograma forman un polígono regular que tiene también sectores de ángulo iguales $\Delta\alpha$, si los sectores se miden con el centro Z del hodograma y no con el punto C (véase figura 3.32).

Finalmente, reunimos todos los elementos para poder llegar a la conclusión de que la órbita es realmente una cónica. Para ello, giramos el hodograma 90° y empalmamos el punto Z con el punto central O , como se muestra en la figura 3.33. Esto se hace con el efecto de empalmar los dos ángulos considerados anteriormente en la figura 3.32. Ahora se conside-

ra que el punto P' recorre el hodograma. Se traza la *mediatriz* t de CP' y se interseca con ZP' para obtener un punto P'' , según se aprecia en la figura 3.33.

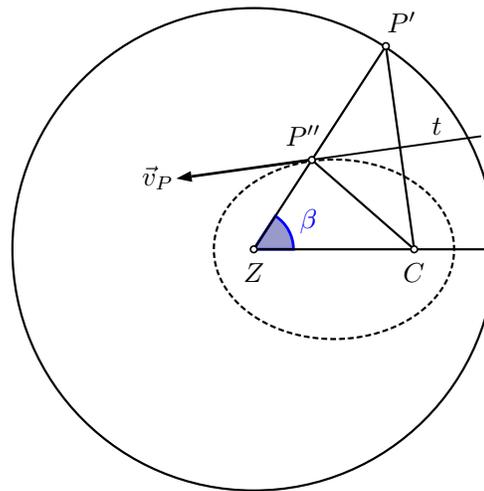


Figura 3.33 La órbita es una cónica.

Como P' recorre una circunferencia con centro Z , la distancia $d(Z, P')$ no cambia. Pero $d(Z, P') = d(Z, P'') + d(P'', P')$ y, como t es mediatriz de CP' , se tiene que $d(P'', P') = d(P'', C)$. Por lo tanto:

$$d(Z, P'') + d(P'', C) = d(Z, P')$$

que es constante. En consecuencia, el punto P'' se encuentra en una elipse con los focos C y Z . Además, como se puede verificar en la sección cónicas, la mediatriz t es tangente a la elipse.

El girar 90° el hodograma tiene otro efecto grato: la velocidad \vec{v}_P del cuerpo en el punto P es paralela a t , dado que la velocidad \vec{v}_P se representa en el hodograma original —antes de girar— por el vector CP' . Hemos encontrado una trayectoria para el cuerpo: la elipse. El punto P'' se mueve sobre una elipse y la velocidad \vec{v}_P es, en cada momento, tangente a la trayectoria de P'' .

Pero, en principio, estamos haciendo una barbaridad. El punto P'' pertenece a un diagrama de velocidades que es incomparable con un diagrama de posiciones si no se fija una escala común. Si invertimos el argumento, se puede aprovechar y decir que podemos fijar una escala tal que, en un momento, la posición P coincida con la de P'' . Entonces, a partir de este momento, P y P'' se moverán al unísono: ambos se encuentran sobre ZP' por la igualdad de los ángulos β en los dos diagramas y ambos tienen, también, el mismo vector de velocidad. Como P'' se mueve por una elipse, lo mismo hace P .

Hemos llegado al final de la conclusión. Se quería demostrar que las trayectorias siempre son cónicas y encontramos lo que ocurre en el caso de las elipses, que se hizo para simplificar el argumento. Lo que se supuso sin argumento alguno es que el punto C está en el interior del hodograma. No obstante, habría que considerar realmente tres casos: cuando C está en el interior —esto nos da una elipse—, cuando C está sobre el hodograma —que nos dará una parábola— y finalmente, cuando C está afuera —que produce una hipérbola—. En cualquiera de estos casos se argumenta de manera similar, pero como la parábola y la hipérbola son curvas abiertas, corresponden a cuerpos que pasan al centro de atracción una sola vez y jamás regresan. Consecuentemente, las órbitas de los planetas y los cometas son elipses.

La argumentación anterior no es del todo sencilla, pero es interesante ver que es posible obtener una intuición geométrica del porqué de lo elíptico de las órbitas y no delegar nues-

tra comprensión a un cálculo diferencial que requiere familiaridad con herramientas más avanzadas, como los vectores.

Con ello, concluimos la argumentación geométrica que hace verosímiles las primeras dos leyes de Kepler a partir de dos principios: que hay una atracción central y que ésta es proporcional al inverso del cuadrado de la distancia. Esto es uno de los más grandes logros científicos de la humanidad.

3.5. LAS ECUACIONES QUE MODELAN EL MUNDO

<Imagen 36>

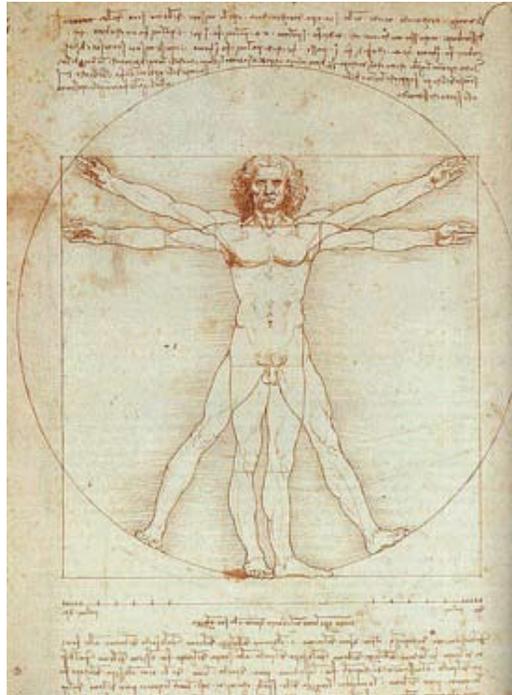


Figura 3.34 El hombre de Vitruvio es un famoso trabajo de Leonardo da Vinci realizado en uno de sus diarios, alrededor de 1492. El dibujo representa una figura masculina desnuda —inscrita en un círculo y un cuadrado— y en dos posiciones sobreimpresas de pies y manos. Este modelo es un estudio de las proporciones del cuerpo humano que utiliza la razón áurea, descrita en la sección 4.2.

3.5.1 Primeros ejemplos

En el contexto matemático, un modelo es una herramienta que facilita la descripción, representación, explicación, predicción, discusión o evolución particular de un objeto, proceso, sistema o fenómeno. En esta sección veremos diferentes herramientas matemáticas que pueden usarse para la modelación y que hay una que, claramente, es la más potente: la ecuación diferencial. Antes de llegar a ello, veremos primero unos ejemplos que muestran la gran versatilidad de los conceptos.

Un número es impar si no es divisible entre dos. La suma de dos impares es par, como se ve, por ejemplo, en $3 + 5 = 8$. Pero si quisiéramos hacer una demostración de ello, necesitamos tener una expresión que nos modele el número impar en general. Ésta existe y es un término de la forma:

$$2n + 1$$

donde n es cualquier entero. Para diferentes valores de n obtenemos diferentes números impares, es más, obtenemos todos los impares y sólo impares. Con este modelo es fácil comprobar que, en general, la suma de dos impares —digamos $2m + 1$ y $2n + 1$ — es par:

$$(2m + 1) + (2n + 1) = 2m + 2n + 2 = 2(m + n + 1),$$

ya que el lado derecho es un múltiplo de 2.

Veremos ahora cómo una *ecuación* puede modelar una situación interesante. El principio que impulsa los cohetes espaciales se basa en la conservación del impulso: si en el espacio vacío —donde no actúa la gravedad— un astronauta lanza un objeto, se impulsa en la dirección opuesta. El cohete trae combustible y, al quemarlo, expulsa gas a gran velocidad v_e —la velocidad de emisión— al espacio, lo que impulsa al cohete. La *ecuación de Tsiolkovski* (5) dice el aumento en la velocidad Δv que puede alcanzar un cohete al quemar todo su combustible. La ecuación relaciona la masa total inicial m_{inicial} y la masa final m_{final} después de consumir el combustible, con las velocidades Δv y v_e .

$$\Delta v = v_e \cdot \ln \left(\frac{m_{\text{inicial}}}{m_{\text{final}}} \right) \quad (5)$$

En esta ecuación aparece la función \ln de logaritmo natural. Por ejemplo, si el 80% de la masa total es combustible, entonces:

$$\frac{m_{\text{inicial}}}{m_{\text{final}}} = \frac{100\%}{20\%} = 5$$

y como $\ln(5)$ es aproximadamente 1.6, el cohete alcanzará 1.6 veces la velocidad de emisión. La velocidad de expulsión puede alcanzar hasta los 4 400 metros por segundo en motores que usan hidrógeno y oxígeno; con ello, el cohete alcanzará una velocidad de más de 25 mil kilómetros por hora. Para un cálculo realista hay que tomar en cuenta que el cohete sale de la Tierra y usa gran parte del combustible para alejarse de ella, es decir, para elevarse en contra de la fuerza de gravedad. Para ese cálculo necesitaríamos otra ecuación.

Como vemos, la ecuación puede ser muy útil para calcular ciertas cantidades, pero al mismo tiempo tiene sus limitaciones —en este caso sólo sirve para el espacio vacío, así que su uso real siempre será aproximado—. Lo anterior es la típica característica de un modelo: permite hacer un cálculo que, usualmente, es simplificado y, a partir de él, hacer predicciones. Toda la aeronáutica espacial se basa en modelos que tienen que ser muy buenos, dado que los experimentos reales son muy costosos.

3.5.2 Un modelo para la radiactividad

La radiactividad es un proceso espontáneo que ocurre en la naturaleza cuando en el núcleo de un átomo ocurre un cambio. En el núcleo se encuentran partículas como los protones, con carga positiva, y los neutrones, que no tienen carga y que estabilizan a los protones. El número de protones determina las características al reaccionar con otros átomos. Es por lo anterior por lo que los átomos se clasifican en *elementos*, como el oxígeno y el carbono que tienen, respectivamente, 8 y 6 protones. El número de neutrones puede variar. Hay átomos de carbono con 6, 7 u 8 neutrones. Esto da un total de 12, 13 y 14, respectivamente, partículas en el núcleo y por ello, se habla de carbono 12, carbono 13 y carbono 14. Se dice que son diferentes *isótopos* de carbono.

El carbono 14 es radiactivo, es decir, puede decaer en cualquier momento, mientras el carbono 13 y el 12 son *isótopos estables*. En el carbono 14 puede suceder que uno de los neutrones se desintegre en un protón, que quede atrapado en el núcleo, y un electrón y un antineutrino que se emitan hacia el exterior del núcleo. En este cambio, el núcleo se queda con 7 protones y 7 neutrones, por lo que el átomo ahora es de nitrógeno. La cantidad de energía

de las partículas que se emiten durante el decaimiento se puede medir y es lo que le dio su nombre al fenómeno: actividad en el espectro de radio, es decir, se registró la radiactividad por la radiación que emite.

Nadie puede predecir cuando un átomo de carbono 14 se desintegra pues lo hace de manera espontánea, de repente. Lo único que se puede decir es que en 5730 años hay 50% de probabilidad de que suceda y otro tanto de que no. Más sorprendente aún es que, si no decayó en estos 5730 años, tiene 5730 años para otra vuelta con una oportunidad para decaer o no del 50% – 50%. En otras palabras, los átomos de carbono 14 si no decaen, no envejecen.

Si en un momento aislamos una gran cantidad de átomos de carbono 14 y esperamos 5730 años encontraremos que más o menos la mitad de átomos ha decaído mientras la otra mitad sigue igual. ¿Qué encontraremos otros 5730 años después? La respuesta correcta no es que ahora la otra mitad también decayó y se convirtió en nitrógeno, sino que de los átomos de carbono que sobrevivieron los primeros 5730 años, ahora la mitad también decayó. En conclusión sólo un cuarto de los átomos originales queda sin decaer y después del tercer periodo de 5730 años, será sólo un octavo.

En nuestro ambiente hay una concentración baja de carbono 14. ¿Cómo es posible si estos isótopos decaen? ¿No tendrían que haber decaído ya todos, después de los millones de años de existencia en la Tierra? Es cierto que continuamente decaen pero, por otro lado, también se generan nuevos átomos de carbono 14 en la atmósfera alta a causa de la radiación solar. De esta manera, la concentración total se mantiene estable en nuestra atmósfera y también en los organismos pues, cada vez que respiran intercambiando gases, incorporan a su estructura estos átomos. Cuando el ser vivo muere, el intercambio cesa y los átomos de carbono 14 que quedaron integrados en su estructura hasta este momento, decaen en los milenios posteriores.

Lo anterior hace posible calcular en un resto orgánico el tiempo que transcurrió desde la muerte del organismo que lo originó al medir la concentración de carbono 14 que resta en el tejido, un método conocido como *fechamiento con carbono 14*. Para ver cómo funciona dicho método, veremos primero cómo depende de la concentración de carbono 14 según el tiempo —medido en años— que transcurre después de la muerte del organismo. Lo que buscamos es una *función* f que modele dicha concentración. De esta función sabemos que $f(0) = 100\%$ y que $f(5730) = 50\%$. Por lo que vimos antes, tenemos que $f(2 \cdot 5730) = 25\%$. Cada 5730 años, la concentración se divide a la mitad, así que $f(3 \cdot 5730) = 12.5\%$. Si escribimos los porcentajes como números tenemos:

$$\begin{aligned} f(0 \cdot 5730) &= 1 \\ f(1 \cdot 5730) &= \frac{1}{2} \\ f(2 \cdot 5730) &= \frac{1}{4} \\ f(3 \cdot 5730) &= \frac{1}{8} \\ f(n \cdot 5730) &= \frac{1}{2^n} \end{aligned}$$

En la última línea escribimos una expresión general. Si ahora sustituimos $t = n \cdot 5730$ y observamos que $\frac{1}{2^n} = 2^{-n}$, obtenemos la fórmula:

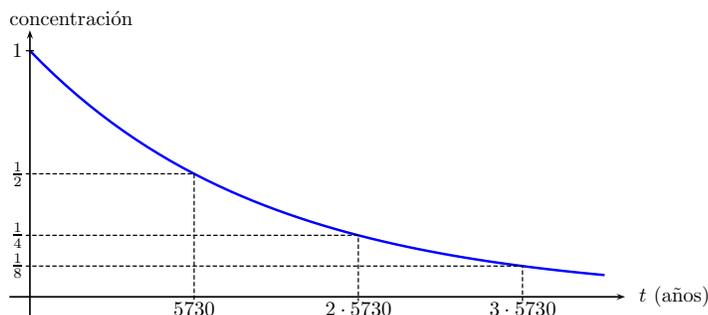
$$f(t) = 2^{-\frac{1}{5730}t}$$

Si, por ejemplo, en un hueso se encuentra una concentración de $14.2\% = 0.142$, entonces debemos resolver la ecuación:

$$0.142 = 2^{-\frac{1}{5730}t}$$

en la variable t . La figura 3.35 muestra la gráfica de esta función.

Figura 3.35 La gráfica de la función que modela la concentración de átomos de carbono 14 en el tiempo, después de la muerte del organismo.



Por ello, es conveniente reescribir el lado derecho y usar la base e —el número de Euler— en vez de 2, ya que la función $\exp(x) = e^x$ tiene un inverso conocido, que es el *logaritmo natural* $\ln(x)$: si $y = e^x$ entonces $x = \ln(y)$. Véase también el siguiente recuadro.



El número e es definido como el siguiente límite:

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$$

La función $\exp(x) = e^x$ tiene la importante propiedad de que es su propia derivada:

$$\frac{d \exp(x)}{dx} = \exp(x)$$

y su inverso se llama *logaritmo natural* y se denota por $\ln(x)$.

Tomamos $y = 2^{-\frac{1}{5730}}$ y obtenemos un número:

$$\ln(2^{-\frac{1}{5730}}) = -0.00012097$$

Entonces, si usamos la letra griega “lambda” para denotar a este número —es decir, $\lambda = 0.00012097$ —, obtenemos:

$$f(t) = e^{-\lambda t} \quad (6)$$

y la ecuación que debemos resolver es:

$$0.142 = e^{-\lambda t}$$

o, con el *logaritmo natural*:

$$\ln(0.142) = -\lambda t,$$

es decir:

$$t = \frac{\ln(0.142)}{-\lambda} = \frac{-1.952}{-0.00012097} = 16136$$

Lo cual quiere decir que el organismo del que proviene el hueso murió hace aproximadamente 16 136 años.

La función (6) se llama *función exponencial* y se emplea con frecuencia para modelar fenómenos de la naturaleza. Lo que la función modela en este caso es el comportamiento de la concentración del carbono 14 en el tiempo; pero esta concentración no es otra cosa que la fracción de la cantidad de átomos en el tiempo entre la cantidad inicial.

En la vida real, esta función da brincos, dado que siempre hay un número entero de átomos en cada momento. Nos deberíamos preguntar si realmente es permisible usar una función continua para un fenómeno discreto, más aún, en la realidad la función depende del azar, dado que no se puede predecir cuándo los átomos decaerán. Para entender por qué la función (6) es extremadamente buena para describir el fenómeno real debe tomarse en cuenta que la cantidad de átomos que conforman la concentración es realmente enorme.

De todos los átomos de carbono en nuestro cuerpo sólo uno de un billón —un millón de millones— es un átomo de carbono 14, por lo que en un gramo de carbono hay, aproximadamente, 40 mil millones de átomos de carbono 14. Por ello, es muy improbable que no sea más o menos la mitad que decae en 5730 años. Si sólo tuviéramos 2 átomos, la probabilidad de que la mitad decaiga es del 50%, mientras hay 25% de probabilidad de que ninguno decaiga y otros 25% de probabilidad de que ambos decaigan. Con 40 mil millones de átomos hay una probabilidad de 99.9953% de que entre 19 999 600 000 y 20 000 400 000 átomos decaigan. Por ello, la función continua aproxima muy bien el fenómeno discreto y probabilístico y, el método de fechamiento con carbono 14 se puede emplear con éxito desde su desarrollo, en 1949, a partir de un equipo encabezado por el físico-químico estadounidense Willard Frank Libby.

3.5.3 Modelos para el crecimiento poblacional

En el año 2000 había, por cada mil habitantes de la Tierra, 22 nacimientos y 9 muertos, es decir, un crecimiento de 13 personas por cada mil. ¿Cuánto tiempo transcurrirá hasta que se duplique la población si el crecimiento fuera constante a este ritmo? Es un error pensar que pasarán $\frac{1000}{13} \approx 77$ años. Para entenderlo, simplificamos y digamos que pasa un cierto tiempo T hasta que la población se duplique. ¿En cuánto tiempo se habrá cuadruplicado? El tiempo necesario es sólo el doble de T ya que en cada periodo T se duplicará por lo tanto, sólo se necesitan dos periodos. El posible error consiste en pensar que en cada año se *suma* la misma cantidad, cuando en realidad en cada año se *multiplica* por el mismo factor.

El factor de crecimiento anual λ se obtiene así: después de un año habrá por cada 1000, unas 13 personas más. Así que:

$$\lambda = \frac{1013}{1000} = 1.013,$$

entonces, $1013 = \lambda \cdot 1000$. Después de dos años la población habrá alcanzado λ^2 veces y, después de tres años, λ^3 veces la población original. En la figura 3.36 se muestra cómo crece λ^n conforme aumenta n .

Figura 3.36 Evaluación de la función 1.013^n para diferentes valores de n .

n	1.014^n	n	1.014^n
1	1.013	10	1.138
2	1.026	20	1.295
3	1.040	30	1.473
4	1.053	40	1.676
5	1.067	50	1.908
6	1.081	60	2.171
\vdots	\vdots	\vdots	\vdots

De la tabla se ve que se requieren entre 50 y 60 años para que se duplique la población mundial si sigue creciendo al mismo ritmo que lo hizo durante el año 2000. Este tipo de crecimiento se conoce como *exponencial*. A corto plazo, el modelo es muy exitoso como se puede observar en la figura 3.37: la estimación no rebasa nunca el 1% de error. El modelo que se usó fue:

$$V(n) = V(0) \cdot \lambda^n \quad (7)$$

donde $\lambda = 1.013$ y $V(n)$ es la población en el año $2000 + n$. Nuevamente, se trata de una función exponencial, pero ahora creciente.

Figura 3.37 La columna de en medio muestra los datos de las Naciones Unidas, la columna a la derecha exhibe la estimación basada en la hipótesis de que el crecimiento porcentual es constante de año en año, e igual al del año 2000. Los datos fueron redondeados a miles.

Año	Población mundial	Estimación
2000	6 115 367 000	6 115 367 000
2001	6 194 886 000	6 194 867 000
2002	6 274 302 000	6 275 400 000
2003	6 353 658 000	6 356 980 000
2004	6 432 978 000	6 439 621 000
2005	6 512 276 000	6 523 336 000
2006	6 591 548 000	6 608 139 000
2007	6 670 801 000	6 694 045 000
2008	6 750 062 000	6 781 068 000
2009	6 829 360 000	6 869 222 000
2010	6 908 688 000	6 958 522 000

El crecimiento exponencial es ilimitado —en el modelo, la población crece y crece, rebasando cualquier límite— y, por lo tanto, inadecuado para predicciones a largo plazo: hay que tomar en cuenta que nuestros recursos son limitados, aún más, el espacio es limitado o, de manera todavía más determinante: la materia en la Tierra es infinita y hay un tiempo para el cual el modelo predice que la población rebasará aun estos límites, lo que es absurdo. Esto se conoce como la *catástrofe de Malthus*, llamada así por el inglés Thomas Robert Malthus, quien publicó seis ediciones de *Ensayo sobre el principio de la población* entre 1798 y 1826.

Por ello, en 1838, el matemático belga Pierre François Verhulst propuso un modelo que tomaba en cuenta lo limitado de los recursos y consideraba que el mundo puede sostener un máximo número de personas K —aunque este máximo sea desconocido. Para explicar este modelo, reformulamos primero el crecimiento exponencial (7) en términos de ecuaciones diferenciales:

$$\frac{dV(t)}{dt} = r \cdot V(t) \quad (8)$$

donde $r = \ln(\lambda) \frac{1}{\text{año}}$; esta reformulación se justifica en el siguiente recuadro.



El número de personas es una función en el tiempo $V(t)$ y, en un tiempo Δt , la población crecerá ΔV . Por ejemplo, si Δt es un año, entonces $\Delta V = 0.013 \cdot V(0)$, si tomamos el crecimiento durante el año 2000 como indicador. Es decir, el crecimiento es proporcional a la población con un factor de 0.013, para lapsos de un año. Por ello, tenemos:

$$V(t + \Delta t) - V(t) = \Delta V = (\lambda - 1)V(t)$$

donde $\lambda = 1.013$. Si ahora tomamos medio año $\frac{\Delta t}{2}$, entonces tendremos un crecimiento de:

$$V(t + \frac{\Delta t}{2}) - V(t) = (\sqrt{\lambda} - 1)V(t).$$

Más general, para la n -ésima parte de un año $\frac{\Delta t}{n}$ se tiene que:

$$V(t + \frac{\Delta t}{n}) - V(t) = (\sqrt[n]{\lambda} - 1)V(t).$$

Obviamente, si crecemos el valor de n , este incremento disminuye y se acerca a 1. Pero si consideramos el cociente:

$$\frac{V(t + \frac{\Delta t}{n}) - V(t)}{\frac{\Delta t}{n}} = n(\sqrt[n]{\lambda} - 1) \frac{V(t)}{\Delta t} \tag{9}$$

pasa algo interesante: del lado izquierdo el numerador y el denominador se acercan simultáneamente al valor 0.

Al tomar el límite para $n \rightarrow \infty$, obtenemos en (9) la derivada $\frac{dV(t)}{dt}$, mientras del lado derecho, el cociente $\frac{V(t)}{\Delta t}$ no depende de n . Por otro lado, el límite se calcula de la siguiente manera:

$$\lim_{n \rightarrow \infty} n(\sqrt[n]{\lambda} - 1) = \ln(\lambda) = 0.0129,$$

como se explica en seguida. Para calcular $\lim_{n \rightarrow \infty} n(\sqrt[n]{\lambda} - 1)$, se hace primero un cambio de variable $n = \frac{1}{m}$ y se observa que $\sqrt[n]{\lambda} - \lambda^n = \lambda^m$:

$$\lim_{n \rightarrow \infty} n(\sqrt[n]{\lambda} - 1) = \lim_{m \rightarrow 0} \frac{\lambda^m - 1}{m}.$$

Esta última fracción se puede interpretar como la derivada de la función λ^x en el punto $x = 0$:

$$\lim_{m \rightarrow 0} \frac{\lambda^m - 1}{m} = \left. \frac{d\lambda^x}{dx} \right|_{x=0} = \left. \frac{d e^{\ln(\lambda) \cdot x}}{dx} \right|_{x=0}$$

que, según la regla de derivación de la composición de funciones, se calcula en:

$$\lim_{m \rightarrow 0} \frac{\lambda^m - 1}{m} = \ln(\lambda) \cdot e^{\ln(\lambda) \cdot x} \Big|_{x=0} = \ln(\lambda).$$

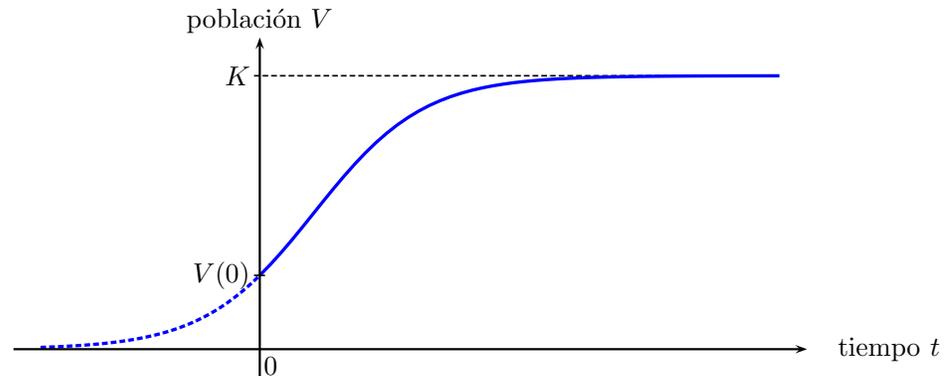
Si ponemos $r = \ln(\lambda) \frac{1}{n \Delta t}$, entonces obtenemos finalmente la siguiente reformulación (9) de la ecuación (8), como se había anunciado.

La modificación de (8) que propone Verhulst es introducir un término correctivo negativo $\frac{r}{K} \cdot V(t)^2$:

$$\frac{dV(t)}{dt} = r \cdot V(t) - \frac{r}{K} \cdot V(t)^2. \quad (10)$$

Este término de corrección depende del cuadrado de $V(t)$ y esto tiene una buena razón: el número de encuentros en una población depende aproximadamente del cuadrado del número de integrantes, si se supone que estos se mueven por el espacio disponible. Con los contactos, se aumenta propagación de enfermedades, lo que disminuye la población —por ello, el término aparece con signo negativo.

Figura 3.38 El modelo de Verhulst muestra el desarrollo de la población $V(t)$ en el tiempo t de manera cualitativa, es decir, sin especificar valores concretos para r y K , ya que estos se desconocen para la población humana.



El término de corrección tiene el efecto deseado sobre el modelo: el desarrollo en el tiempo deja que $V(t)$ se acerque a K pero sin sobrepasarlo, como se observa en la figura 3.38 y justo como se deseaba.

3.5.4 Un modelo para la propagación de un virus

Existen muchos modelos matemáticos usados para la prevención de la salud y la estimación de cuán contagiosa es una determinada enfermedad infecciosa. Por ejemplo, se puede modelar cómo un virus —como el de la influenza— se disemina en una comunidad. En este caso se toma en cuenta que los enfermos entran en contacto con personas sanas y las contagian con el agente infeccioso. Veamos el planteamiento del modelo en forma sencilla: sea $I(t)$ el número de personas que se han contagiado con la enfermedad como función del tiempo t y $S(t)$, el número de personas que aún no se han enfermado. Finalmente, también habrá que considerar el número de personas $R(t)$ que ya no se infectarán ni podrán infectar a otros: son aquellas personas que se han recuperado de la enfermedad o que han muerto a causa de ella.

Parece razonable suponer que, la rapidez $\frac{dS(t)}{dt}$ a la que las personas sanas se enferman, es proporcional al número de encuentros o interacciones entre estos dos grupos de personas. También es razonable suponer que el número de interacciones es proporcional al número de enfermos $I(t)$ y al número de personas sanas $S(t)$, es decir, proporcional al producto $I(t) \cdot S(t)$, entonces:

$$\frac{dS(t)}{dt} = -k \cdot I(t)S(t), \quad (11)$$

donde k es la constante de proporcionalidad. El signo negativo indica que el número de personas sanas disminuye. Hay dos factores que influyen en el cambio del número de infec-

tados: por un lado, aumenta por los nuevos infectados y, por otro, disminuye por los que se curaron o murieron. En el modelo se supone que el número de personas que deja de ser infeccioso es proporcional al número de infectados. La siguiente ecuación da cuenta de ello:

$$\frac{dI(t)}{dt} = k \cdot I(t)S(t) - \ell \cdot I(t), \quad (12)$$

Nuevamente, debe observarse los signos de los dos sumandos. Para finalizar, la tercera ecuación explicará el cambio en el número de personas que dejaron de ser infecciosas.

$$\frac{dR(t)}{dt} = \ell \cdot I(t), \quad (13)$$

Se observa que contraer la enfermedad da inmunidad a la persona, es decir, la persona ya no se puede infectar nuevamente. Las tres ecuaciones (11), (12) y (13) forman el modelo que discutimos aquí. Se trata de un modelo continuo para un fenómeno discreto: los números considerados siempre serán enteros ya que no hay fracciones de personas.

Sólo la práctica puede darnos indicaciones sobre los parámetros k y ℓ para una población dada y un virus determinado. La correcta estimación de estos parámetros puede conducir a una predicción sobre el porcentaje de población que debe vacunarse para impedir una epidemia.

Las campañas de vacunación masiva son proyectos de la Organización Mundial de la Salud (OMS) y de muchos departamentos de salud pública. Una característica importante de las enfermedades infecciosas es la reproducción básica del factor R_0 , que permite estimar el número de infecciones secundarias que un individuo infectado producirá durante el tiempo en que está enfermo. Para casi cualquier enfermedad infecciosa, los modelos de ecuaciones diferenciales pueden llevar al cálculo del R_0 básico al usar información específica —tal como la tasa de transmisión del patógeno, la duración del periodo infeccioso y la tasa promedio de muerte en la población—. El R_0 proporciona un estimado de la situación de la infección, por ejemplo, de cómo se puede diseminar en una población sana: si el R_0 es menor a uno, la enfermedad no persistirá en la población, pero si es mayor a uno, entonces la enfermedad posee el potencial para diseminarse en forma de epidemia o de volverse endémica. Un caso interesante para ejemplificar lo anterior es el de la viruela, actualmente erradicada del planeta, a pesar de que tenía un R_0 de 3 a 5 y requirió vacunar alrededor del 70% u 80% de la población —el último caso de viruela se reportó en 1977—. La OMS está tratando de erradicar también la polio, que tiene un R_0 de 5 a 7, lo cual significa que, aproximadamente, se requiere inmunizar entre un 80% y 86% de la población mundial para lograr desaparecer la infección.

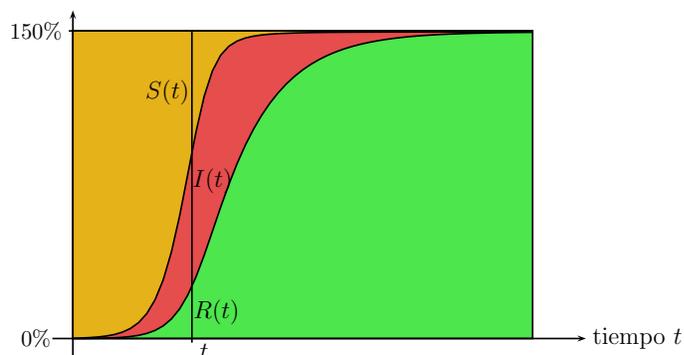


Figura 3.39 El desarrollo cualitativo de las tres cantidades, “sanos” $S(t)$, “infectados” $I(t)$ y “recuperados” $R(t)$, en el modelo de propagación de virus. En cada momento t , la suma de los tres números debe ser el total de la población, es decir, igual a 100%.

Los modelos permiten estudiar y comprender cómo es que ocurren los fenómenos que no podemos observar fácilmente o manipular directamente. Por ello, son de fundamental importancia para las ciencias. El mayor éxito entre los modelos han sido las ecuaciones diferenciales, de las cuales aquí dimos algunas muestras.

3.6 EL CAMPO Y LOS VECTORES



Figura 3.40 La atracción gravitatoria establece una relación cuantitativa entre dos objetos con masa, como el Sol y la Tierra.

¿Cómo actúa la fuerza gravitatoria? ¿Cómo es que el Sol ejerce una fuerza sobre un planeta que se encuentra a miles de kilómetros de distancia? Resulta aún más sorprendente este hecho si consideramos que esa fuerza ejercida a gran distancia, a juzgar por sus efectos en el movimiento de los astros, parece tomar valores perfectamente definidos con gran precisión a cada momento. No sabemos cómo responder a esas preguntas, pero hay una idea de los físicos que resulta muy útil para tranquilizar nuestras angustias filosóficas. La idea es que el Sol y todo cuerpo con masa en realidad no está localizado en la pequeña zona en la que está comprendida su parte material, su masa, sino que hay algo no material que es también parte del Sol: el campo de fuerza gravitatoria, y esa parte está distribuida en todo el Universo. ¿Qué es este campo? En lugar de decir que es sólo una abstracción matemática, adoptamos la postura más científica que consiste en decir que tiene una realidad física perfectamente clara y bien definida, lo único que no posee, y que estamos acostumbrados a pensar que sí poseen los entes físicos, es masa. Decimos que el campo gravitatorio existe pues se manifiesta físicamente ejerciendo fuerza sobre los objetos con masa.

El estudio de la electricidad impulsó aún más la idea del campo, en este caso, el llamado campo eléctrico. A continuación se describe la notación matemática de vectores y campos vectoriales que se generaron alrededor de la electrostática y resultaron útiles para profundizar en el estudio del electromagnetismo, la mecánica de los fluidos y muchos otros temas de la física.

3.6.1 La electrostática y el concepto de campo vectorial

El campo eléctrico es una función vectorial $\mathbf{E}(x, y, z)$ que se define como la fuerza que actuaría sobre una carga eléctrica unitaria colocada en el punto (x, y, z) . Se dice que una función es vectorial si a cada punto de su dominio se le asocia un vector con dirección, magnitud y sentido. El campo vectorial \vec{E} se puede representar mediante tres componentes escalares o numéricas E_x, E_y, E_z , una a lo largo de cada eje de coordenadas. En símbolos, se puede representar al campo eléctrico así:

$$\vec{E} = E_x \vec{i} + E_y \vec{j} + E_z \vec{k}$$

o más explícitamente, así:

$$\vec{E}(x, y, z) = E_x(x, y, z) \vec{i} + E_y(x, y, z) \vec{j} + E_z(x, y, z) \vec{k}$$

donde $\vec{i}, \vec{j}, \vec{k}$ son vectores unitarios —de magnitud 1— en las direcciones de los ejes x, y, z , respectivamente. El lector debe observar que los vectores se escriben en los textos impresos con letra **negrita**. En los manuscritos los vectores se indican poniendo una raya o una flecha encima del nombre del vector.

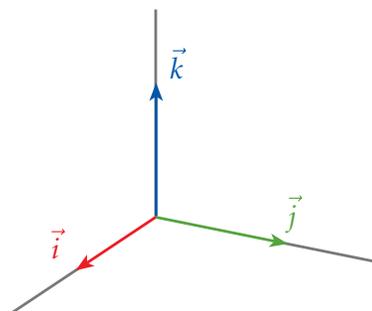


Figura 3.41 Los vectores unitarios $\vec{i}, \vec{j}, \vec{k}$.

El comportamiento del campo eléctrico está definido por la Ley de Coulomb y la ley de superposición. La Ley de Coulomb dice que el campo eléctrico producido por una carga eléctrica Q en un punto del espacio es proporcional a la magnitud de la carga e inversamente proporcional al cuadrado de la distancia, y su dirección coincide con la del vector que va de la carga al punto. El sentido del campo es hacia la carga si ésta es negativa y, en sentido opuesto si es positiva. Suponiendo que la carga eléctrica Q se encuentra en el origen, que $\vec{r} = x\vec{i} + y\vec{j} + z\vec{k}$ es el vector de posición del punto (x, y, z) , que $r = |\vec{r}|$ es la distancia del punto (x, y, z) al origen y que \vec{u} es el vector unitario que apunta en la dirección del punto (x, y, z) en sentido opuesto al origen, la Ley de Coulomb puede expresarse así:

$$\vec{E} = \frac{Q}{r^2} \vec{u} = \frac{Q}{r^3} \vec{r} = \frac{Q}{r^3} (x\vec{i} + y\vec{j} + z\vec{k})$$

donde la segunda igualdad se justifica porque:

$$\vec{u} = \frac{\vec{r}}{r}.$$

La ley de superposición dice que el campo eléctrico producido por dos o más cargas eléctricas es la suma de los producidos por cada una de ellas. Por ejemplo, el campo eléctrico correspondiente a una carga positiva Q colocada en el punto $(-a, 0)$ y una negativa $-Q$ en el punto $(a, 0)$ tiene esta expresión:

$$E = \frac{Q}{r_1^3} \left((x+a)\vec{i} + y\vec{j} + z\vec{k} \right) - \frac{Q}{r_2^3} \left((x-a)\vec{i} + y\vec{j} + z\vec{k} \right)$$

$$\text{donde } r_1 = \sqrt{(x+a)^2 + y^2 + z^2} \text{ y } r_2 = \sqrt{(x-a)^2 + y^2 + z^2}.$$

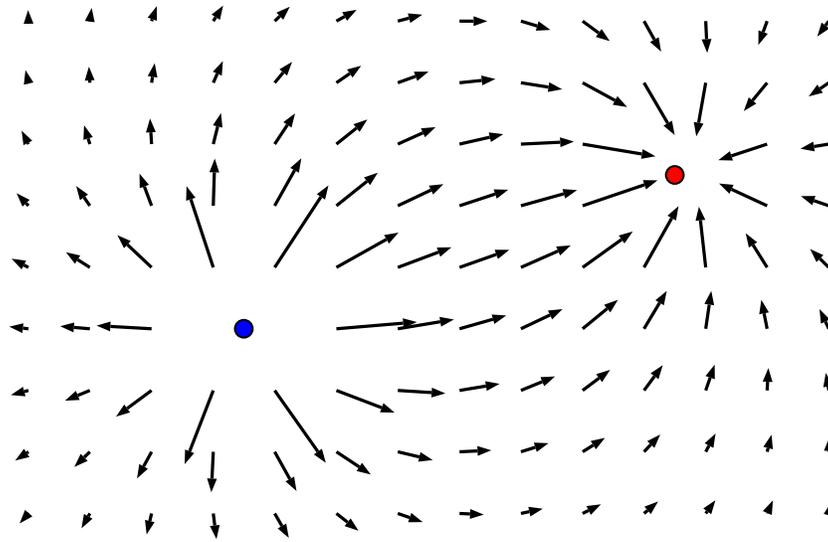


Figura 3.42 Ilustración del campo eléctrico debido a una carga eléctrica positiva (azul) y una negativa (roja).

El estudio de la electrostática es prácticamente imposible sin el uso de estas nuevas herramientas matemáticas que hemos venido introduciendo a lo largo de los párrafos anteriores y que constituyen lo que se llama vectores y campos vectoriales. A continuación, se presenta un resumen de los principales conceptos del álgebra de vectores y el cálculo vectorial. Se trata de una presentación muy escueta cuyo único objetivo es ofrecer al lector una primera idea de la potencia expresiva de estas herramientas matemáticas que, como se verá en la siguiente sección, llevaron a grandes descubrimientos en relación al electromagnetismo y las ondas electromagnéticas, esenciales para la tecnología de las comunicaciones modernas.

Álgebra vectorial

Dados dos vectores $\vec{a} = a_1\vec{i} + a_2\vec{j} + a_3\vec{k}$ y $\vec{b} = b_1\vec{i} + b_2\vec{j} + b_3\vec{k}$, y un escalar —o sea, un número— c se definen cuatro operaciones algebraicas:

- 1] Suma de dos vectores: $\vec{a} + \vec{b} = (a_1 + b_1)\vec{i} + (a_2 + b_2)\vec{j} + (a_3 + b_3)\vec{k}$.
- 2] Multiplicación de un vector por un escalar: $c\vec{a} = ca_1\vec{i} + ca_2\vec{j} + ca_3\vec{k}$.
- 3] Producto escalar de dos vectores (también llamado producto punto):
 $\vec{a} \cdot \vec{b} = a_1b_1 + a_2b_2 + a_3b_3$.
- 4] Producto vectorial (o producto cruz) de dos vectores:
 $\vec{a} \times \vec{b} = (a_2b_3 - a_3b_2)\vec{i} + (a_3b_1 - a_1b_3)\vec{j} + (a_1b_2 - a_2b_1)\vec{k}$.

Las primeras dos operaciones corresponden a lo que se llama álgebra lineal y las dos segundas a lo que propiamente es el álgebra vectorial. Los productos entre vectores son de gran utilidad por sus propiedades geométricas que pueden demostrarse como teoremas a partir de la definición de dichas operaciones, que describimos a continuación.

El producto escalar o producto punto $\vec{a} \cdot \vec{b}$ es un número —un escalar— igual al pro-

ducto de la magnitud de cualquiera de ellos por la proyección del otro sobre el primero o, equivalentemente: $\vec{a} \cdot \vec{b} = |\vec{a}||\vec{b}| \cos(\theta)$, donde θ es el ángulo que forman los dos vectores.

El producto vectorial de dos vectores es un vector que apunta en la dirección perpendicular a ambos, el sentido está dado por lo que se llama la regla de la mano derecha, que dice: el producto vectorial $\vec{a} \times \vec{b}$ apunta hacia donde lo hace el pulgar de la mano derecha cuando el dedo índice apunta se alinea con \vec{a} y el dedo medio con \vec{b} . Finalmente, la magnitud de $\vec{a} \times \vec{b}$ es igual al producto de las magnitudes de \vec{a} y \vec{b} por el seno del ángulo entre ellos:

$$|\vec{a} \times \vec{b}| = |\vec{a}||\vec{b}| \operatorname{sen}(\theta)$$

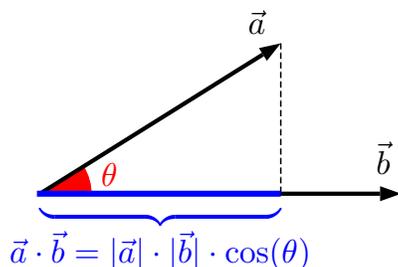


Figura 3.43 Producto escalar.

Cálculo vectorial

Los conceptos de derivada e integral que se desarrollaron para funciones de una variable, se generalizan de distintas maneras a funciones de varias variables y a campos vectoriales. Aquí nos interesa, las generalizaciones de esos conceptos a funciones escalares y vectoriales de tres variables por ser éstas de gran utilidad en el estudio de los fenómenos físicos. El Teorema Fundamental del Cálculo tiene versiones vectoriales, muy interesantes en sí mismas y muy útiles en la física.

Estudiaremos tres conceptos que corresponden vagamente al de derivada: el **gradiente**, la **divergencia** y el **rotacional**. El gradiente se aplica a campos escalares, es decir, a funciones escalares $f(x, y, z)$ definidas en alguna región del espacio de tres dimensiones. Un ejemplo de función escalar es la temperatura $T(x, y, z)$. El concepto de gradiente corresponde al vector que apunta en la dirección y sentido del máximo crecimiento de la función y cuya magnitud es igual a la razón de cambio del valor de la función en esa dirección. Un teorema del cálculo vectorial nos dice que el gradiente $\operatorname{grad}(f)$, de una función $f(x, y, z)$, que también se denota como ∇f , puede calcularse fácilmente usando las derivadas parciales de:

$$f : \operatorname{grad}(f) = \nabla f = \frac{\partial f}{\partial x} \vec{i} + \frac{\partial f}{\partial y} \vec{j} + \frac{\partial f}{\partial z} \vec{k}$$

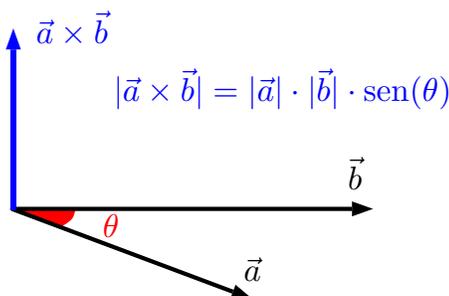
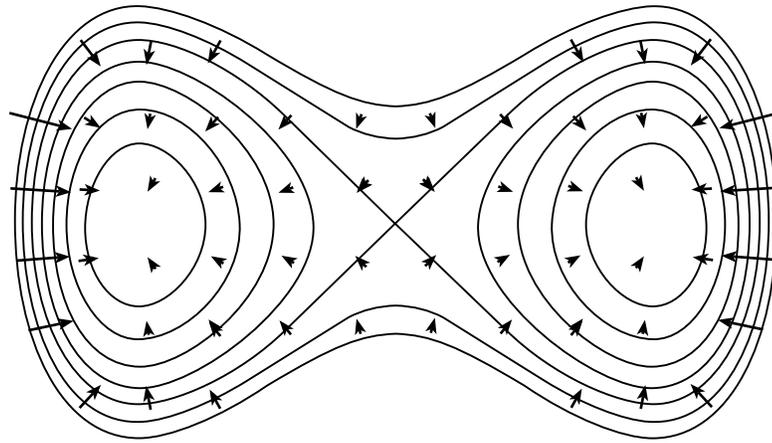


Figura 3.44 Producto vectorial.

Figura 3. 45 El gradiente de un campo escalar mide la dirección del mayor ascenso: a cada punto le asocia un vector que apunta en la dirección en la cual el campo escalar aumenta en mayor grado. La imagen muestra una función escalar con líneas de nivel: cada línea marca un valor constante de la función. Las flechas indican el gradiente en varios lugares. El gradiente asocia a un campo escalar un campo vectorial.



Para definir los conceptos de divergencia y rotacional es conveniente definir antes dos conceptos de integración para campos vectoriales: la integral de línea a lo largo de una curva y el flujo o integral de superficie. Definiremos estos conceptos utilizando la idea intuitiva de que la integral es una suma infinita de elementos infinitamente pequeños. Lo importante es entender qué representan los elementos que se suman y, por tanto, lo que representa cada una de esas integrales. No nos va a preocupar en este libro de la formalización matemáticamente precisa de los conceptos —que por supuesto se hace usando límites—, basta decir que tal formalización puede lograrse y se puede encontrar en casi todos los libros especializados en cálculo —o análisis— vectorial.

La integral de línea del campo vectorial \vec{F} a lo largo de una curva C se denota por:

$$\int_C \vec{F} \cdot d\vec{r}$$

y representa la suma infinita de elementos infinitamente pequeños $\vec{F} \cdot d\vec{r}$ a lo largo de la curva. Aquí, $d\vec{r}$ representa el vector de desplazamiento infinitesimal a lo largo de la curva —en realidad, tangente a ella—. Si pensamos que \vec{F} es una fuerza, cada elemento $\vec{F} \cdot d\vec{r} = |\vec{F}| \cdot |d\vec{r}| \cos(\theta)$ representa el trabajo realizado por esa fuerza \vec{F} a lo largo del elemento $d\vec{r}$ de la misma —recuerda: *trabajo = fuerza \times distancia*; la multiplicación por $\cos(\theta)$ corresponde a tomar la proyección de \vec{F} a lo largo de la tangente a la curva. Por tanto, la integral de línea puede interpretarse como el trabajo realizado por \vec{F} sobre una partícula que se mueve a lo largo de la curva C . Cuando C es una curva cerrada, es decir, un ciclo, entonces la integral de línea se escribe así:

$$\oint_C \vec{F} \cdot d\vec{r}$$

y se llama la circulación de \vec{F} a lo largo de la curva cerrada C .

El flujo de un campo vectorial \vec{F} a través de una superficie S se define como la integral de superficie y se denota como:

$$\iint_S \vec{F} \cdot d\vec{S},$$

que es la suma de una infinidad de elementos infinitesimales $\vec{F} \cdot d\vec{S}$ cubriendo toda la superficie S . Los elementos $\vec{F} \cdot d\vec{S}$ son el producto escalar del vector \vec{F} por el vector $d\vec{S}$ que se define como perpendicular a la superficie y con magnitud igual a la del área que cubre. Las superficies se consideran orientadas y la orientación de una superficie es, precisamente, el sentido del vector perpendicular a ella. La mejor manera de interpretar el flujo de un

campo vectorial a través de una superficie es pensando que el campo vectorial corresponde a la velocidad de un fluido. Entonces el flujo es la cantidad de fluido que atraviesa la superficie por unidad de tiempo.

El concepto de divergencia $\text{div}(\vec{F})$ o $\vec{\nabla} \cdot \vec{F}$ se aplica a un campo vectorial $\vec{F} = F_x \vec{i} + F_y \vec{j} + F_z \vec{k}$. La divergencia se define como el siguiente límite —si existe—:

$$\text{div}(\vec{F}) = \vec{\nabla} \cdot \vec{F} = \lim_{V \rightarrow 0} \frac{1}{V} \iint_S \vec{F} \cdot d\vec{S}$$

donde V representa un volumen encerrado por una superficie S . El límite se toma haciendo el volumen cada vez más pequeño, colapsándolo al punto para el que se está calculando la divergencia. Si \vec{F} representa el campo gravitatorio, resulta que la divergencia es proporcional a densidad de masa ρ :

$$\vec{\nabla} \cdot \vec{F} = \rho$$

En particular, en el vacío no hay masa y, por lo tanto, la divergencia del campo gravitatorio en el vacío es cero. Algo similar ocurre para el campo eléctrico:

$$\vec{\nabla} \cdot \vec{E} = \rho$$

donde ρ representa en este caso la densidad de carga eléctrica. No es difícil demostrar que la divergencia de un campo vectorial puede calcularse con las derivadas parciales de sus componentes, específicamente:

$$\text{div}(\vec{F}) = \vec{\nabla} \cdot \vec{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$$

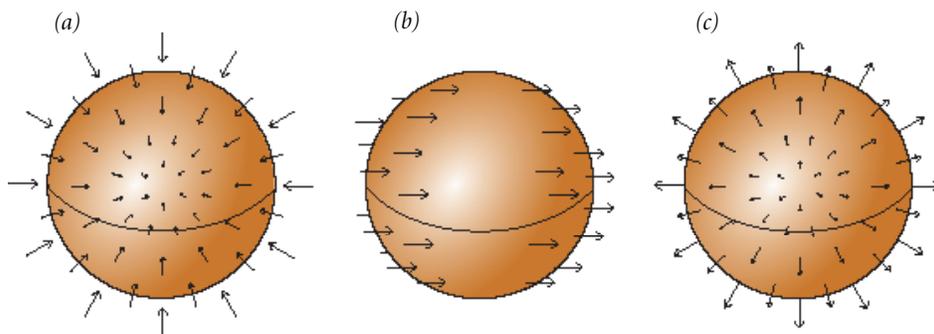


Figura 3. 46 La divergencia de un campo vectorial es una medida para medir flujos que salen de un lugar. En los tres ejemplos, en (a) se tiene una divergencia positiva, en (b) es cero y en (c) es negativa. La medida es local: se puede medir en cada punto al calcular el flujo que sale de una pequeña esfera que rodea este punto y luego se reduce el radio de esta esfera hacia cero en un proceso límite. Así la divergencia convierte un campo vectorial en un campo escalar.

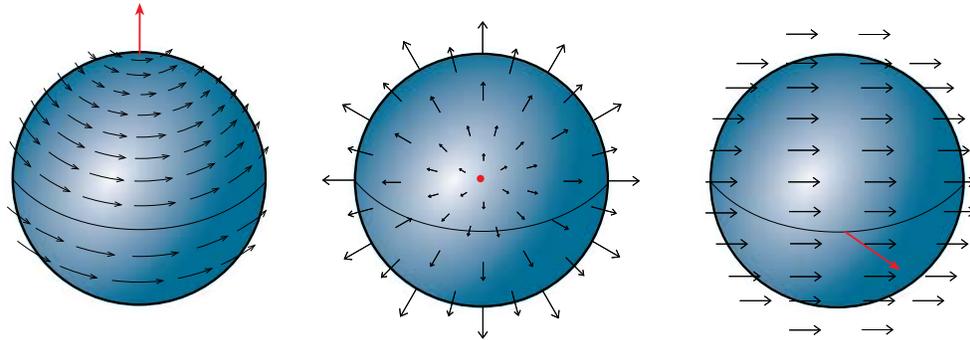
El concepto de rotacional $\text{rot}(\vec{F}) = \nabla \times \vec{F}$ también se aplica a un campo vectorial $\vec{F} = F_x \vec{i} + F_y \vec{j} + F_z \vec{k}$. Se define como otro campo vectorial que tiene la siguiente propiedad para cualquier vector unitario \vec{n} :

$$\text{rot}(\vec{F}) \cdot \vec{n} = (\vec{\nabla} \times \vec{F}) \cdot \vec{n} = \lim_{C \rightarrow 0} \frac{1}{L(C)} \oint_C \vec{F} \cdot d\vec{r}$$

donde C es una curva cerrada sobre un plano perpendicular a \vec{n} que en el límite se encoge o se colapsa a un punto. Se puede demostrar que el rotacional de un campo vectorial puede calcularse con las derivadas parciales de sus componentes, específicamente así:

$$\text{rot}(\vec{F}) = \vec{\nabla} \times \vec{F} = \left(\frac{\partial F_y}{\partial z} - \frac{\partial F_z}{\partial y} \right) \vec{i} + \left(\frac{\partial F_z}{\partial x} - \frac{\partial F_x}{\partial z} \right) \vec{j} + \left(\frac{\partial F_x}{\partial y} - \frac{\partial F_y}{\partial x} \right) \vec{k}$$

Figura 3.47 La rotacional de un campo vectorial mide el giro alrededor de un punto. En el ejemplo (a) la rotación apunta hacia arriba, en (b) es cero y en (c) apunta hacia adelante. Si se imagina el campo vectorial como la velocidad de un líquido o un gas y se coloca una pelota pequeña con el centro fijo dentro de este líquido entonces la rotacional apunta en la dirección del eje en el cual gira la pelota. La rotacional asigna a un campo vectorial otro campo vectorial.



Hay tres famosos e importantes teoremas del cálculo vectorial en tres dimensiones que enunciamos a continuación.

a) Teorema de Gauss:

$$\iiint_V \nabla \cdot \vec{F} = \iint_S \vec{F} \cdot d\vec{S}$$

donde V es un volumen encerrado por la superficie S , lo cual se denota por $S = \partial V$.

b) Teorema de Stokes:

$$\iint_S \nabla \times \vec{F} \cdot d\vec{S} = \oint_C \vec{F} \cdot d\vec{r}$$

donde S es una superficie cuya frontera es la curva C , lo cual se denota por $C = \partial S$.

c) Teorema del potencial: Si $\nabla \times \vec{F} = 0$ (o equivalentemente, si $\oint_C \vec{F} \cdot d\vec{r} = 0$ para toda curva cerrada C) entonces existe un campo escalar $f(x, y, z)$, llamado el potencial de \vec{F} , tal que:

$$\vec{F} = \nabla f = \text{grad}(f)$$

y viceversa, si un campo vectorial es el gradiente de un potencial, entonces su rotacional es cero y las integrales de línea sobre curvas cerradas son cero.

Los dos primeros teoremas son una consecuencia más o menos directa de las definiciones de divergencia y el rotacional, respectivamente. El último es un resultado profundo que en particular nos dice que el campo gravitatorio y el campo electrostático son gradientes de un potencial. La función potencial de estos campos resulta muy útil en los cálculos. El campo magnético, en cambio, no tiene esta propiedad. El campo magnético no es el gradiente de un potencial, su rotacional no es cero y hay curvas a lo largo de las cuales la circulación del campo magnético es distinta de cero.

La siguiente sección explica las interesantes leyes del electromagnetismo, su importancia en la ciencia moderna y cómo las matemáticas jugaron un papel fundamental en el descubrimiento de las ondas electromagnéticas, que son la base de las comunicaciones modernas.

El campo electromagnético

Entender el funcionamiento planetario a través de un modelo matemático que asignó números a ciertas entidades físicas —para luego relacionarlos en fórmulas precisas y claras—, fue un gran logro del intelecto humano. Este hallazgo, la teoría de la gravitación universal de Newton, hizo pensar a los científicos, durante dos siglos, que habían encontrado uno de los secretos más íntimos de la naturaleza. A lo largo del siglo XIX se estudiaron la electrici-

dad y el magnetismo, y se halló entre ambos fenómenos una correspondencia que también puede describirse perfectamente mediante fórmulas matemáticas que relacionan los campos eléctricos y magnéticos. El descubrimiento de Ørsted —las corrientes eléctricas generan campos magnéticos—, y el de Faraday —sobre la creación de corrientes eléctricas con el movimiento de imanes—, dieron origen a toda una nueva rama de la física que a la vez, reforzó la idea de que el universo se rige por leyes matemáticas.

Uno de los mayores logros científicos que se deben directamente a las matemáticas, fue el descubrimiento de las ondas electromagnéticas. Al poner en limpio las leyes del electromagnetismo y tomar en cuenta la tendencia de la naturaleza a exhibir simetrías, Maxwell llegó a un sistema de ecuaciones que describe perfectamente los fenómenos electromagnéticos conocidos hasta entonces. Obtuvo una representación matemática sumamente elegante y compacta del electromagnetismo: las ecuaciones de Maxwell; a la vez, descubrió la inducción mutua de los campos eléctrico y magnéticos, es decir, el hecho de que las variaciones en el tiempo del campo magnético producen uno eléctrico, y, a la vez, las variaciones en el tiempo del campo eléctrico producen otro magnético. Lo anterior se traduce matemáticamente en la Ecuación de Onda, conocida con anterioridad. Las soluciones de esta ecuación son ondas, es decir, perturbaciones en un medio que viajan por el espacio. Las fórmulas matemáticas del electromagnetismo señalaron que si de verdad este fenómeno podía describirse con esas ecuaciones, entonces deberían existir ondas electromagnéticas. Más aún, las ecuaciones de Maxwell definían con absoluta precisión la velocidad de transmisión de dichas ondas, aproximadamente $300\,000 \frac{\text{km}}{\text{s}}$. Hoy sabemos que ésta es la velocidad de la luz.



<Imagen 47, esta es de Internet, no está muy buena. Habría que buscar.>

Figura 3.48 Max Karl Ernst Ludwig Planck (1858-1947), galardonado con el Premio Nobel de Física en 1918, fue un físico alemán considerado el fundador de la teoría cuántica.

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= \frac{\rho}{\epsilon_0} \\ \vec{\nabla} \cdot \vec{B} &= 0 \\ \vec{\nabla} \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} \\ \vec{\nabla} \times \vec{B} &= \mu_0 \vec{J} + \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t} \end{aligned}$$

Figura 3.49 Las ecuaciones de Maxwell en forma diferencial.

$$\begin{aligned} \vec{\nabla} \cdot \vec{E} &= 0 \\ \vec{\nabla} \cdot \vec{B} &= 0 \\ \vec{\nabla} \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} \\ \vec{\nabla} \times \vec{B} &= \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t} \end{aligned}$$

Figura 3.50 Las ecuaciones de Maxwell en el vacío, que exhiben la simetría de la naturaleza con mayor claridad que las otras ecuaciones.

$$\begin{aligned} \frac{1}{c^2} \frac{\partial^2 \vec{E}}{\partial t^2} &= \frac{\partial^2 \vec{E}}{\partial x^2} + \frac{\partial^2 \vec{E}}{\partial y^2} + \frac{\partial^2 \vec{E}}{\partial z^2} \\ \frac{1}{c^2} \frac{\partial^2 \vec{B}}{\partial t^2} &= \frac{\partial^2 \vec{B}}{\partial x^2} + \frac{\partial^2 \vec{B}}{\partial y^2} + \frac{\partial^2 \vec{B}}{\partial z^2} \end{aligned}$$

Figura 3.51 Ecuaciones de onda para los campos eléctrico y magnético, que se obtienen manipulando las ecuaciones de Maxwell en el vacío. Sus soluciones son ondas que se propagan con velocidad $c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$.

Motivado por el trabajo de Maxwell, Hertz intentó descubrir si tales ondas de verdad existían y logró producirlas en un punto y registrarlas en otro, es decir, logró transmitir di-

chas perturbaciones de un lugar a otro. Esas ondas se llamaron por un tiempo ondas hertzianas y muy pronto quedó perfectamente claro que correspondían a las que Maxwell había previsto con sus ecuaciones. La primera aplicación de las ondas electromagnéticas fue la radio, y ésta fue el paso inicial de la revolución tecnológica de las comunicaciones, basada en ese gran descubrimiento al que se llegó con unas fórmulas matemáticas.

No es pues sorprendente que parte de la humanidad esté enamorada de la idea de que al mundo lo rigen un conjunto de fórmulas matemáticas, lo cual ha funcionado muy bien en muchos ámbitos de la ciencia, sobre todo en las diversas ramas de la física. En innumerables ocasiones, la creación de un modelo matemático para representar un fenómeno de la naturaleza ha llevado a importantes descubrimientos más allá de lo estrictamente vinculado con el fenómeno en cuestión. La fórmula de la atracción gravitatoria ayudó no únicamente a comprender cómo se mueven los planetas, sino también a descubrir cuerpos celestes desconocidos hasta entonces.

Este resultado matemático —las ecuaciones de Maxwell— para el electromagnetismo llevó no sólo a la aplicación de las ondas electromagnéticas en el mundo de las comunicaciones sino que, además, el haber entendido la interrelación entre la electricidad y el magnetismo guió el avance tecnológico desde finales del siglo XIX.

Una de sus consecuencias es el motor eléctrico. Incontables usos de la energía eléctrica son posibles gracias al electromagnetismo. Las computadoras y todo aquello que llamamos “nuevas tecnologías” dependen, en un altísimo porcentaje, de

nuestro descubrimiento del electromagnetismo. La enorme cantidad de información, que se acumula hoy en discos de computadora y se transmite vía internet, es almacenada en medios magnéticos y se extrae y transmite usando procesos electromagnéticos. La medicina moderna ha alcanzado diagnósticos antes insospechados a través de la resonancia magnética que es posible, precisamente, gracias a la manipulación del electromagnetismo. Incluso las nuevas generaciones de trenes de pasajeros llevarán grandes electroimanes y levitarán sobre vías, que son, a su vez, electroimanes. Los ciclotrones, utilizados en la investigación de los secretos de la materia y el origen del universo, aceleran las partículas haciéndolas colisionar unas con otras, por medio de electroimanes.

La predicción de las ondas electromagnéticas es quizá la aplicación más importante en toda la historia de las matemáticas. Los efectos de este descubrimiento en la forma de vida de la humanidad han sido espectaculares.

La física cuántica

Todas las ciencias utilizan en mayor o menor medida las matemáticas, aunque la física es sin duda la que mayor provecho ha sacado de ellas. Cuanto más alejada se encuentra una ciencia de los fenómenos físicos, más difícil resulta utilizar en ella las matemáticas clásicas y más necesario se hace el recurrir a métodos estadísticos. Podría creerse que la física es determinista y que, sólo al alejarnos de ella, es necesario introducir el elemento de azar que reflejaría cierto grado de ignorancia. Sin embargo, durante la primera mitad del siglo XX se

<Imagen 48>

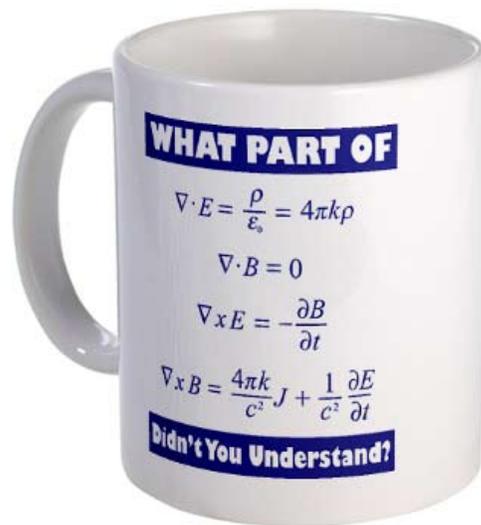


Figura 3.52 Las ecuaciones de Maxwell son tan elegantes que se han vuelto populares entre los estudiantes y las escriben en playeras y tazas de café.

realizaron descubrimientos extraordinarios en la física, en concreto en la estructura atómica de la materia, que condujeron a la necesidad de crear un modelo matemático, intrínsecamente aleatorio, para explicar el comportamiento del mundo atómico y subatómico.

Como suele ocurrir, un hallazgo de esta relevancia, propició la creación de una nueva rama de la física, llamada esta vez: mecánica cuántica. Niels Bohr, Heisenberg, Schrödinger y otros físicos llegaron a la conclusión de que si había un modelo matemático para el comportamiento de los electrones alrededor del núcleo en un átomo, éste debería incluir elementos de azar. De hecho, existe algo intrínsecamente aleatorio en la naturaleza. No se trata de simple falta de datos lo que nos lleva a introducir el elemento de azar sino que, en realidad, las partículas subatómicas tienen un comportamiento esencialmente aleatorio. Por ejemplo, al lanzar una partícula sobre una pared con dos pequeños huecos, puede que choque sin pasar por los huecos o, también,

puede ser que pase por ellos, pero el que pase o no, no depende de manera determinista de la posición y velocidad inicial de la partícula, como ocurriría en el campo de la mecánica newtoniana que se aplica a los cuerpos macroscópicos. En el caso de las partículas elementales es imposible determinar con precisión absoluta la posición y velocidad inicial de una de ellas, es más: cuanto mejor definida esté su posición, menos definida estará su velocidad y viceversa. Esto se llama el *principio de incertidumbre* de Heisenberg y tiene una formulación matemática muy precisa:

$$\Delta x \cdot \Delta p \geq \frac{h}{4\pi}$$

donde Δx es la incertidumbre en la posición, Δp la incertidumbre en la cantidad de movimiento íntimamente relacionado con la velocidad, y h es una constante, llamada la constante de Planck. El concepto de incertidumbre es equivalente al de desviación estándar que se usa en la estadística. El principio de incertidumbre dice que ninguna de las dos incertidumbres Δx o Δp puede ser cero y que, si una de ellas es muy pequeña, la otra deberá ser muy grande. Schrödinger encontró una ecuación que describe la distribución de probabilidad de que un electrón —de un átomo— se encuentre en algún punto del espacio alrededor del núcleo:

$$i \frac{h}{2\pi} \frac{\partial \Psi}{\partial t} = \frac{h^2}{8\pi^2 m} \Delta \Psi + V \Psi$$

La ecuación de Schrödinger para la amplitud de onda Ψ .

$|\Psi|^2$ es la densidad de probabilidad de la posición de una partícula.

La ecuación de Schrödinger, además, resulta ser una ecuación de onda, similar a la que describe el campo electromagnético, por lo cual surge la posibilidad de que los electrones de un átomo y las partículas elementales en general, tengan un comportamiento ondulatorio. De Broglie predice que las partículas podrían comportarse como ondas, al final se demuestra experimentalmente que en verdad sucede así: al cruzar por una rendija, los electrones exhiben un comportamiento típico de las ondas: interferencia.

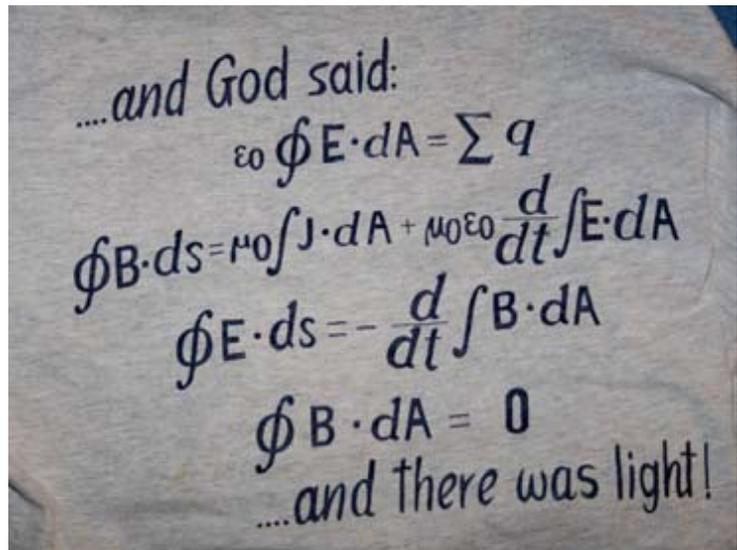


Figura 3.53 Las ecuaciones de Maxwell en forma integral.

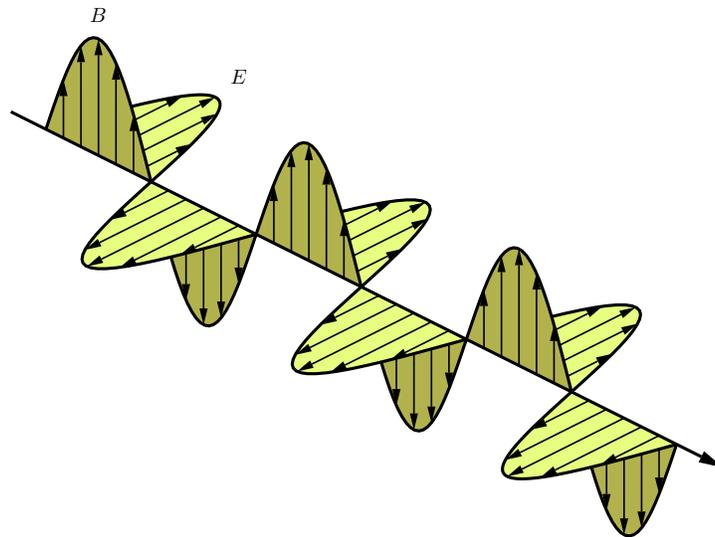


Figura 3.54 (falta pie)

La distribución de electrones que han pasado a través de una rendija se parece mucho a la de la luz que pasa por una mirilla, similitud por la cual la ciencia estaba convencida de que la luz es un fenómeno ondulatorio. Se pensaba que la luz consistía probablemente en ondas electromagnéticas, pues la velocidad de estas ondas, de acuerdo con las ecuaciones de Maxwell, era exactamente la misma que se había medido para la luz: $300\,000\frac{\text{km}}{\text{s}}$.

De nuevo, al crear un modelo matemático para describir un fenómeno se llega a un descubrimiento científico: la materia exhibe un comportamiento ondulatorio. Ya con anterioridad el trabajo de Planck sobre la radiación del cuerpo negro, y el de Einstein sobre el efecto fotoeléctrico, habían sugerido que la luz, considerada en aquel momento como ondas electromagnéticas, se comportaba como si estuviera constituida por pequeñas partículas llamadas “cuantos”. Ahora se deducía que lo contrario también era cierto, las partículas de las que estaba constituida la materia, se comportaban como ondas. “¿Qué onda?”, se preguntaría asombrado un joven mexicano... Por fin, ¿la luz es onda o partícula? y la materia, ¿es onda o partícula? La conclusión de todas estas interrogantes e investigaciones es increíble: la luz y también la materia, son ambas onda y partícula. Hay una realidad dual. Quizás es difícil comprenderlo —igual que fue difícil entender el concepto del continuo y los límites— pero esta dualidad pasó a ser un principio fundamental de la física que, apoyado en fórmulas matemáticas, permite describir el comportamiento aleatorio de las partículas-ondas de la materia y la luz.

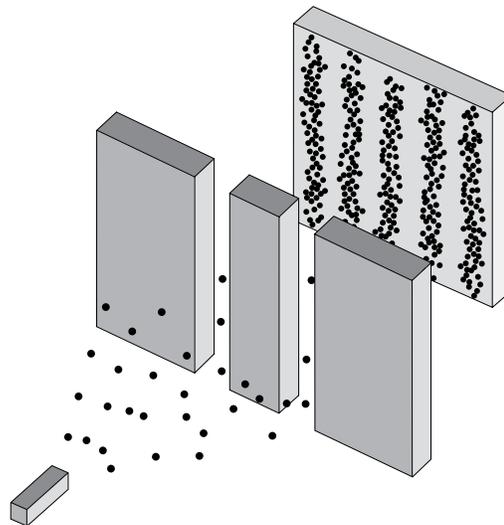


Figura 3.55 Esquema del experimento de la doble rendija para electrones en el que se observa el fenómeno ondulatorio de interferencia.

La física subatómica tiene aún muchos enigmas y cuestionamientos, y lleva años buscando su Santo Grial, es decir, una teoría unificada que, en caso de existir, consistiría en un conjunto de fórmulas matemáticas capaces de describir el comportamiento de la materia-energía, relacionándola con todas las interacciones de la naturaleza, de las cuales hoy se conocen cuatro: la gravedad, el electromagnetismo, las interacciones nucleares débiles y las interacciones nucleares fuertes.

Albert Einstein y la Teoría de la Relatividad

El lector pensará que es exageración, pero aún hay otra rama de la física que se creó al definir un modelo matemático. Dicho modelo supuso predicciones sorprendentes que fueron verificadas después mediante observaciones y experimentos. Se trata de la Teoría de la Relatividad de Einstein.

En 1905, antes de que la física cuántica se desarrollara, uno de los más grandes genios de la humanidad, Albert Einstein —que entonces era un modesto y joven empleado de la oficina de patentes de la ciudad suiza de Zúrich—, publicó sus trabajos sobre tres inquietantes temas de la física de ese momento. Uno de ellos fue una brillante explicación del efecto fotoeléctrico, que consistía en cómo un haz de luz al incidir sobre ciertos materiales —fotoeléctricos— lograba que los electrones saltaran de esos materiales y pasaran a una placa con carga eléctrica positiva frente a ellos, generando de esa manera una corriente eléctrica. El efecto era conocido y no resultaba sorprendente que la luz, consistente en ondas electromagnéticas que transportan energía, interactuara con partículas de carga eléctrica como los electrones. Lo inaudito era el comportamiento cuantitativo del fenómeno. Por ejemplo, en el caso de cierto material con luz roja, los electrones no saltaban aunque ésta fuera de gran intensidad, mientras que con muy poca luz azul sí lo hacían.

Planck en su teoría sobre la radiación del cuerpo negro había señalado que la luz podía comportarse como si la luz estuviera constituida por partículas cuya energía era proporcional a su frecuencia: $E = h\nu$, donde ν es la frecuencia y h es la llamada constante de Planck. A esas partículas de luz se les llamó “cuantos” aunque ahora se conocen como fotones. Einstein tuvo la idea de explicar la radiación del cuerpo negro aprovechando la idea de los cuantos de Planck: como la frecuencia de la luz azul es casi el doble que la de la luz roja, los cuantos de luz azul tenían mucha más energía que los de luz roja y por eso podían expulsar a los electrones de su órbita alrededor del átomo, mientras los de luz roja no. Esta sencilla explicación fue una de las ideas generadoras de la mecánica cuántica.

Otro de los trabajos publicados por Einstein, en 1905, contenía un modelo matemático que explicaba, cuantitativamente, el comportamiento del llamado movimiento browniano. Antes de este trabajo las causas del fenómeno estaban claras: se trataba de colisiones entre las moléculas del fluido y otras partículas más grandes, que se hallaban en suspensión. Lo paradójico de esta explicación es que la partícula en suspensión era demasiado grande en relación a las moléculas, para que el efecto individual de éstas al chocar pudiera producir el movimiento observado. El modelo matemático de Einstein resolvía esta aparente paradoja pues demostraba que las trayectorias quebradas descritas por la partícula en suspensión eran el resultado de muchísimos choques de las moléculas. La explicación de Einstein está muy relacionada con la ley débil de los grandes números que nos dice que las variables aleatorias se acercan a sus valores esperados, pero lo hacen muy lentamente. A pesar de la importancia evidente de los dos trabajos mencionados, el tercero es el que mayor fama dio al sabio alemán y el que tiene un gran peso en la historia de intimidad entre la física y las matemáticas. Se trata del trabajo inaugural de la Teoría de la Relatividad.

A Einstein le preocupaba un detalle acerca de lo que se sabía sobre la luz y los electrones: por un lado parecía que la velocidad de la luz era una constante universal, esto podía deducirse de las ecuaciones de Maxwell e incluso había un experimento —el de Michelson y Morley— cuyos resultados eran comprensibles sólo pensando que dos observadores que se mueven uno con respecto al otro, obtendrían el mismo valor si midieran la velocidad de la luz. Sin embargo, era obvio que estos mismos observadores medirían velocidades diferentes para cualquier objeto en movimiento. Por ejemplo, si dicho objeto se encontrara en reposo para uno, para el otro estaría moviéndose con la misma velocidad que hay entre ambos. En el caso de la luz no pasa esto. La luz, entendida como onda electromagnética, no podría estar en reposo con respecto a uno de los observadores: como para ambos las ecuaciones del campo electromagnético son exactamente las mismas y, de acuerdo con ellas, las ondas electromagnéticas propagándose en el vacío tendrían que moverse a una velocidad específica de $300\,000 \frac{\text{km}}{\text{s}}$, resulta que ambos observadores deben medir la misma velocidad para cualquier rayo de luz. Esto parece una paradoja, sin embargo Einstein, al dar más importancia a la lógica que a la intuición —que depende de nuestra limitada experiencia como ínfimos habitantes del universo— se dio cuenta de que la verdad era que la velocidad de la luz es efectivamente la misma para todos a quienes ilumine y la midan; en cambio las distancias, las masas y el tiempo, medidos por diferentes observadores, sí cambian.

$$\begin{aligned}t' &= \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}} \\x' &= \frac{x - vt}{\sqrt{1 - v^2/c^2}} \\y' &= y \\z' &= z\end{aligned}$$

Figura 3.56
Transformaciones de
Lorentz.

Las transformaciones de Lorentz son consecuencia lógica de que la velocidad de la luz es la misma para dos observadores que se mueven a velocidad constante, uno respecto al otro.

Esta conclusión, impecable desde el punto de vista lógico y tan contradictoria con nuestras experiencias vitales, puede representarse perfectamente mediante un modelo matemático y, como ocurrió en muchos otros casos donde se llegó a un modelo matemático de algún fenómeno de la naturaleza, el modelo conocido como la Teoría de la Relatividad, predice fenómenos que la ciencia acaba por verificar mediante la observación.

Para ejemplificar lo anterior, conviene aclarar que existen dos teorías de Einstein sobre la Relatividad. Las conclusiones del estudio de Einstein sobre los cuerpos que se mueven con velocidad constante unos con respecto a otros se conocen como la Teoría de la Relatividad Especial. La Teoría de la Relatividad General desarrollada algunos años más tarde por el propio Einstein, incorpora fuerzas, aceleraciones y la atracción gravitatoria. En ella, Einstein encuentra que todo funciona como si la presencia de cuerpos con masa deformara el espacio y estas deformaciones son, precisamente, las causantes de las aceleraciones y curvaturas en las trayectorias de los cuerpos.

La Teoría de la Relatividad General, predice que la órbita de Mercurio —y en realidad, la de todos los planetas— no es en verdad una elipse sino que es casi una elipse pero con una precisión, es decir, el eje de la elipse va girando poco a poco. Como el fenómeno es más intenso cuanto más cerca está el planeta del Sol, es claramente observable sólo en Mercurio. La teoría de la relatividad también predice que la luz, al pasar cerca de un cuerpo masivo, podría deflectarse, es decir, que la atracción gravitatoria puede actuar sobre un rayo de luz. Esto lleva a considerar que la luz se comporta como si tuviera masa, una nueva sugerencia

de que luz y materia no son cosas diferentes sino manifestaciones distintas de un mismo ente. Pero, sin duda, la consecuencia de la Teoría de la Relatividad que más ha impactado a la sociedad está representada por la famosa fórmula:

$$E = mc^2$$

que dice que una masa m puede convertirse en una enorme cantidad de energía. La energía que puede llegar a generarse a partir de un objeto material es proporcional a su masa y la constante de proporcionalidad es el cuadrado de la velocidad de la luz, que es un número muy, muy grande.

Esta fórmula anuncia un hecho angustiante: un solo gramo de materia podría generar suficiente energía para destruir la Tierra. La ciencia y la tecnología han unido esfuerzos para explotar esta importante relación entre masa y energía, produciendo primero, las bombas atómicas y de hidrógeno, que convierten materia en energía de una manera violenta; después, creando los reactores nucleares que generan energía eléctrica de una manera controlada,

a partir de materiales radiactivos y, posteriormente, al controlar la fisión nuclear —que se produce cuando dos átomos de hidrógeno se unen para formar uno de helio— con una pérdida insignificante de masa que libera una gran cantidad de energía. El control de la fisión nuclear representa una de las pocas oportunidades del hombre para obtener una fuente limpia de energía y sustituir el consumo de petróleo y otras formas contaminantes de producción energética.



Figura 3.57 Timbre de correos soviético de 1979, con Albert Einstein y la famosa fórmula $E = mc^2$.

Conclusión

No es de extrañar que los científicos sigan intentando descubrir fórmulas matemáticas que describan aquellos fenómenos de la naturaleza que aún no han podido ser comprendidos y dominados. Están convencidos de que cualquier avance en la representación matemática de la naturaleza volverá a traer consecuencias importantes para la vida y hasta para la supervivencia de la humanidad. En particular, la posibilidad de controlar la fisión nuclear y aprovecharla para generar energía sin producir contaminantes, el poder almacenarla y transmitirla hasta donde se necesite, quizá dependa de que se resuelvan algunos enigmas de la ciencia y de que la solución se exprese en forma de ecuaciones matemáticas.

Albert Einstein dedicó gran parte de su vida a intentar crear una teoría unificada del campo gravitacional y el electromagnético y no lo logró. El descubrimiento de otras dos fuerzas de la naturaleza llamadas nucleares —la interacción débil y la fuerte— complicó más el panorama sobre un posible modelo matemático integral y unificado de las fuerzas de la naturaleza, pero el sueño de Einstein continúa vigente y se cree que es posible realizarlo a pesar de que las mentes más brillantes del mundo le han dedicado un peculiar y extraordinario esfuerzo durante años sin tener éxito. Desde que el hombre descubrió que su sola capacidad de raciocinio podría ser suficiente para comprender el mundo, se ha propuesto explicarlo todo mediante leyes racionales expresadas con fórmulas matemáticas. La teoría unificada sería el mayor logro imaginable en esa dirección; es, como ya dijimos, el Santo Grial de la ciencia contemporánea y nadie concibe esta teoría sin la participación de las matemáticas.

Las matemáticas se vinculan íntimamente con las leyes de la naturaleza. No sabemos por qué, pero es un hecho evidente dados los muchos ejemplos de esta estrecha relación que se encuentran en la historia de la ciencia. Quizá sea cierto aquello que decía Galileo de que la naturaleza es un libro abierto escrito en caracteres matemáticos. Es probable que el propio Galileo nunca imaginara que su verdad atravesaría los siglos alcanzando un sentido cada vez más profundo.

LAS MATEMÁTICAS DE LAS MATEMÁTICAS

TEMA

4

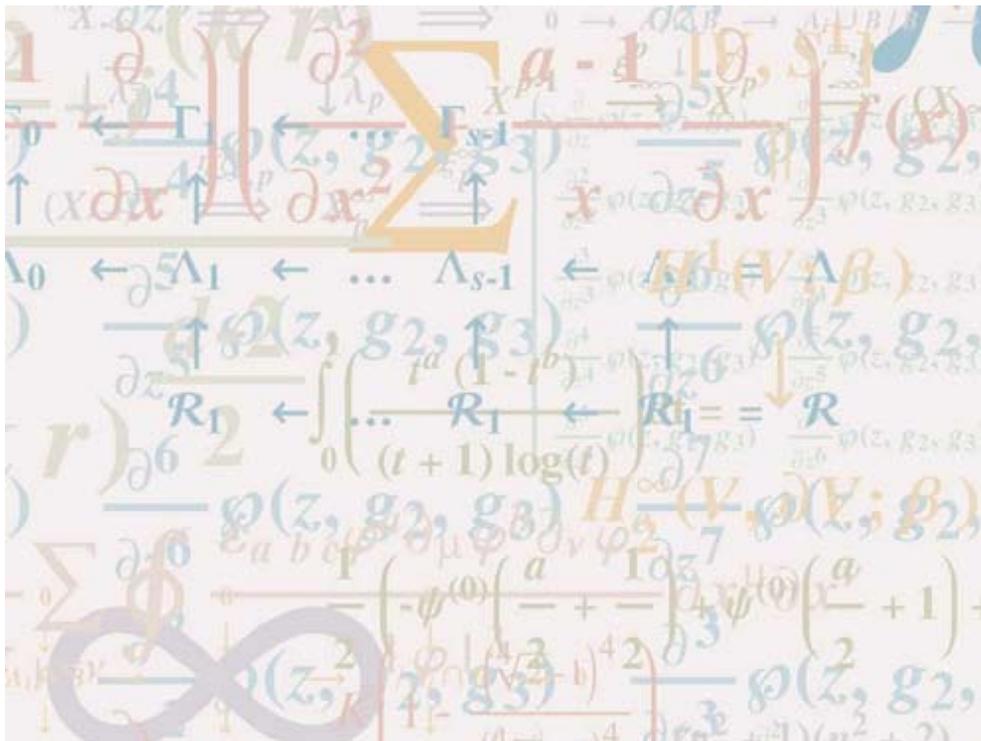


Figura 4.1 Se confunde con frecuencia a las matemáticas con simbología complicada. Pero en realidad la buena simbología es tan sólo una herramienta que ayuda a asimilar y trabajar con ideas profundas, bellas, naturales y con su razón de ser.

4.1 INTRODUCCIÓN

Una de las fuentes de inspiración más fértiles para la creación de las matemáticas son ellas mismas. Muchos de sus desarrollos más espectaculares surgen de mirarse obsesivamente al ombligo, es decir, del trabajo para resolver problemas que ellas mismas plantean. Siendo por naturaleza abstractas, parecería que este ensimismamiento las conduciría a una espiral que se aleja irremediabilmente de la realidad. Pero no es así. La historia ha demostrado una y mil veces que matemáticas creadas en la estratósfera de la abstracción se convierten en la herramienta para entender o resolver problemas de otra ciencia o área muy concreta de la actividad humana. El ejemplo más clásico son las

cónicas de los griegos que reaparecen como las órbitas planetarias con Kepler, pero hay innumerables más: el MP3 y los DVD, Google, la relatividad general o la mecánica cuántica por citar algunos otros ejemplos famosos de este fenómeno en que desarrollos abstractos de las matemáticas se aplican tiempo después.

Los matemáticos creemos, no como fe sino como método de trabajo, que esto sucede así porque hay *naturalidad* en las matemáticas y que éstas se descubren; que están *ahí* como la realidad lo está para las otras ciencias pues da la sensación de que su *naturalidad* es algo cercano a la naturaleza —y la historia lo indica—. De tal manera que uno de los encantos de hacer matemáticas es ese contrapunto lúdico que se establece entre crear y descubrir. Henry Poincaré, uno de los matemáticos más importantes hace un siglo, decía que “el científico no estudia la naturaleza porque sea útil; lo hace porque se deleita en ello y se deleita en ello porque es bella”. Quizá no todos los científicos estén de acuerdo con él hoy día, pero los matemáticos seguro que sí, e incluyen, como Poincaré lo hacía, a su materia de trabajo en el vocablo *naturaleza*.

Las secciones o apartados que se aglutinan en este capítulo responden a motivaciones, o describen desarrollos, que surgen de las propias matemáticas; de su dinámica interna. Pedimos entonces al lector que se muerda la lengua si lo asaltan las preguntas ¿y esto para qué sirve? o ¿en qué me será útil? Debe ser condescendiente al empezar cada apartado y dar por válida la motivación que se presenta, pues como fenómeno cultural así es como se han desarrollado en gran medida las matemáticas y vale la pena conocerlas como tal; tratar de apreciar por qué han cautivado a mentes tan extraordinarias.

En este capítulo iniciamos con un tópico muy clásico que es la razón áurea, pero poniendo énfasis en su interés matemático más que en el estético. Se sigue con secciones que versan en áreas como la combinatoria, la teoría de conjuntos, el álgebra y la geometría, con enfoques diversos pero siempre partiendo de los cuestionamientos que dieron origen a algunos de sus desarrollos. Después, damos un giro y revisamos aspectos de los fundamentos de las matemáticas así como de sus límites, para intentar transmitir que a partir de cuestionamientos aparentemente filosóficos también surgen matemáticas profundas. Concluimos con una breve exposición de la vitalidad y el crecimiento explosivo de las matemáticas contemporáneas.

4.2 RAZÓN ÁUREA (Y FIBONACCI)

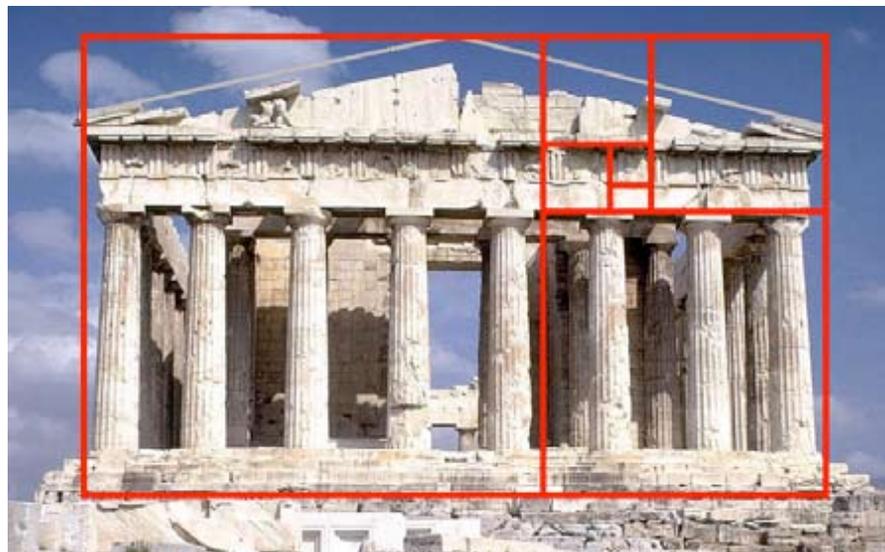


Figura 4.2 El Partenón en la Acrópolis de Atenas, Grecia. Su diseño arquitectónico es una oda a la razón áurea.

No es claro quién descubre la razón áurea y cuándo; pero es un hecho que los griegos la usaron de manera prominente en su arte y arquitectura. El nombre que seguimos usando para ella, razón o proporción “dorada”, y a veces también se le llama “divina”, viene desde entonces. Uno de los ejemplos emblemáticos, es la fachada del Partenón en la Acrópolis que está en Atenas, Grecia: todo el diseño arquitectónico está basado en la razón áurea, como lo demuestra la imagen de la página anterior. Pero desde mucho antes se le conocía, porque en las grandes pirámides de Egipto también se le utiliza. La gran Pirámide de Giza, muy cerca de El Cairo, tiene en su diseño central un triángulo cuya proporción de la base a la hipotenusa es áurea.

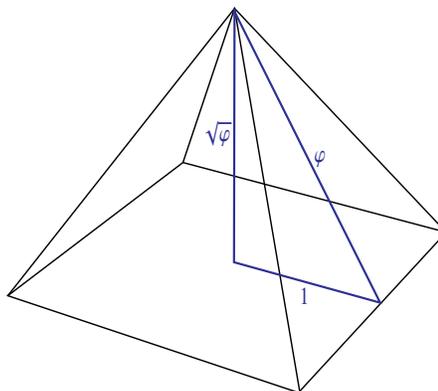


Figura 4.3 Pirámide de Giza y un esquema de su triángulo central.

Pitágoras fue un gran estudioso de la razón áurea. Observó que las proporciones humanas tienen una estrecha relación con ella; observación que Leonardo da Vinci también documentó en el Renacimiento. Pero, ¿qué es la razón áurea?

Un rectángulo es *áureo* cuando al quitarle el cuadrado más grande posible (pegado a una arista corta), el rectángulo que queda tiene la misma proporción que el original.

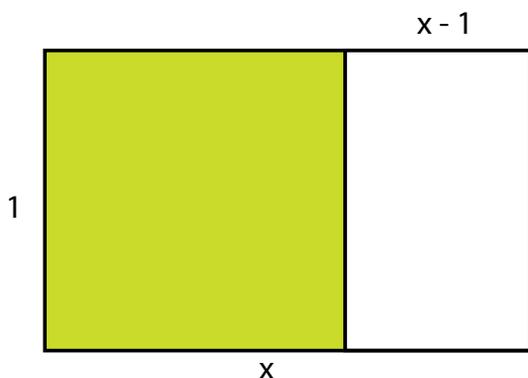


Figura 4.4 Rectángulo áureo.

Veremos que esta propiedad da lugar a una ecuación cuadrática. Como sólo nos interesa la proporción, podemos suponer que el lado chico del rectángulo es 1. Denotemos por x (la incógnita por excelencia) al lado grande. De tal manera que el cuadrado máximo es de lado 1, y el rectángulo que sobra tiene lados $x - 1$ (el chico) y 1 (el grande). Si éste tiene la misma proporción que el original, se cumple la ecuación:

$$\frac{1}{x} = \frac{x - 1}{1},$$

pues cada lado de la ecuación es la razón de lado chico a lado grande.

Multiplicando por x , se obtiene:

$$1 = x(x - 1) = x^2 - x ,$$

que equivale a la ecuación cuadrática:

$$x^2 - x - 1 = 0 . \quad (1)$$

Se resuelve con la fórmula del “chicharronero” que nos da:

$$x = \frac{1 \pm \sqrt{5}}{2} .$$

Hay dos soluciones a la ecuación (1). Cuando usamos el signo $-$, como $\sqrt{5} > 1$, nos da un número negativo. Así que la solución que nos interesa es con el signo $+$, y podemos definir:

$$\varphi = \frac{1 + \sqrt{5}}{2} \approx 1.6180339887499... . \quad (2)$$

donde ya estamos usando la notación común de llamar φ (es la letra griega que se lee “fi”) al número áureo. Su proporción con el 1 es la razón áurea. Hemos demostrado que un rectángulo con lados a y b (con $a < b$) es áureo si:

$$\frac{b}{a} = \varphi .$$

Vale la pena hacer notar que la otra solución de la ecuación (1) es $-\varphi^{-1}$.

4.2.1 El pentagrama místico

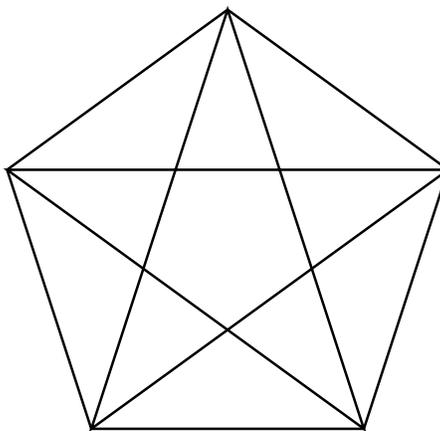


Figura 4.5 Pentágono regular y sus diagonales (el pentagrama).

Uno de los hechos que emocionaba a los griegos sobre la razón áurea, que les daba la seguridad de que su uso era trascendente y no casual, es que apareciera también en el pentágono regular: es la proporción de la diagonal al lado. Que estas dos longitudes, que surgen de un contexto tan diferente, también guarden la proporción divina, produce un estremecimiento o algo parecido al goce estético de estar ante el Partenón o ante un cuadro de Leonardo. Además, como veremos a continuación, en el razonamiento que conduce a este hecho, las piezas caen con tal precisión en el rompecabezas que es inevitable sentir “el toque mágico” de las matemáticas. No extraña, pues, que la escuela pitagórica haya adoptado al pentagrama como su símbolo, y que le hayan conferido poderes místicos.

Consideremos un pentágono regular con sus cinco diagonales (figura 4.5); a veces, se le llama el *pentagrama*. Supongamos que sus lados miden 1 y veremos a continuación que el número áureo φ es efectivamente la longitud de las diagonales.

Observemos primero que todos los ángulos que aparecen en el pentagrama son múltiplos de $\pi/5$ (léase “pi- quintos”) o 36 grados (aquí, es mejor medir ángulos con radianes pues los argumentos se basan en que cinco de estos ángulos dan justo media vuelta, es decir, π radianes o 180 grados). Para verlo, considérense los ángulos que se forman en un vértice del pentágono, al trazar segmentos a los otros cuatro vértices y añadir su tangente al círculo que los contiene como en la figura 4.6 a). Que los cinco ángulos que se forman son iguales, se sigue de un teorema general sobre los ángulos inscritos en un círculo, pues los cinco arcos en que se divide este último son iguales. De este hecho, y de que los ángulos de un triángulo suman π , se pueden obtener todos los ángulos que aparecen en el pentagrama en la figura 4.6 b).

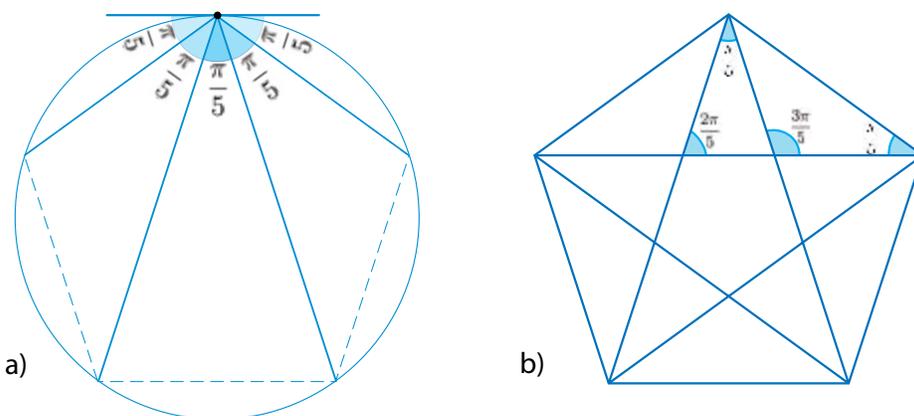


Figura 4.6 a) Los cinco ángulos en un vértice de un pentágono regular. b) Hay tres tipos de ángulos en el pentagrama.

Si se alargan dos lados no consecutivos del pentágono, el AC y el BD en la figura 4.7, se intersectan en un nuevo punto E.

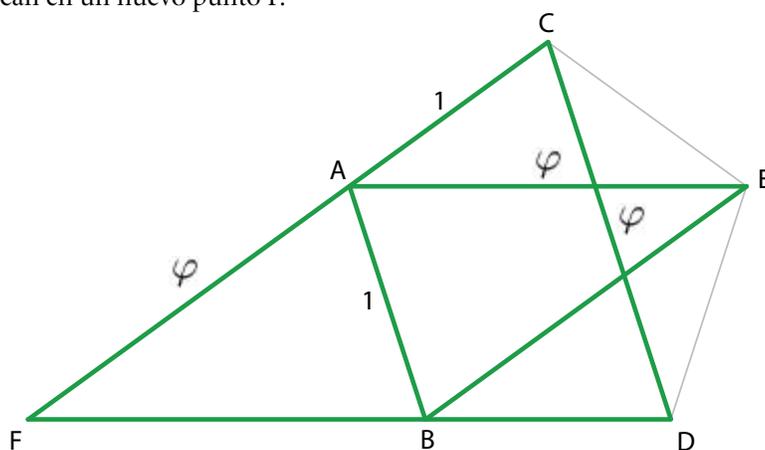


Figura 4.7 El pentágono regular con dos de sus lados alargados y tres diagonales.

Los triángulos isósceles ABF y ABE son iguales, pues su base, que mide 1, coincide y tienen los mismos ángulos (uno de $\pi/5$ y dos de $2\pi/5$, que se lee “dos piquintos”). Esto nos dice que los segmentos AF y BF miden φ (como la diagonal AE). Entonces, de la clara semejanza de los triángulos FAB y FCD, se obtiene que:

$$\frac{\varphi}{1} = \frac{\varphi + 1}{\varphi},$$

que es equivalente a la ecuación:

$$\varphi^2 = \varphi + 1 . \quad (3)$$

Ésta es otra forma de escribir la ecuación (1) cuya solución positiva es el número áureo φ . Así que hemos demostrado que la diagonal de un pentágono regular está en proporción áurea con su lado.

La ecuación (3) es conocida como la *ecuación áurea*. Recordando que $\varphi^0 = 1$ y multiplicándola por φ^{n-2} se obtiene:

$$\varphi^n = \varphi^{n-1} + \varphi^{n-2}$$

para cualquier n positiva o negativa. En particular, se obtiene que $\varphi = 1 + \varphi^{-1}$ y que $1 = \varphi^0 = \varphi^{-1} + \varphi^{-2}$, de tal manera que todos los segmentos del pentagrama son potencias de φ como se aprecia en la figura 4.8.

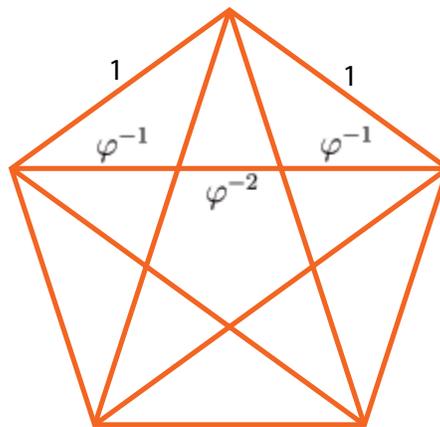


Figura 4.8 Todos los segmentos en el pentagrama son potencias de φ .

Los dos tipos de triángulos isósceles que aparecen en el pentagrama son conocidos como *triángulos áureos*. El *alto*, que usamos en la demostración, tiene dos ángulos de $2\pi/5$ y uno de $\pi/5$; y el *chaparro* tiene dos ángulos de $\pi/5$ y uno de $3\pi/5$. En el pentagrama hay cinco copias (una por arista) del triángulo áureo alto con lados 1 y φ , y diez copias del chaparro (una por vértice tanto del pentágono grande como del chiquito en el centro). Aparecen también, como sus intersecciones, copias de éstos en escalas áureas hacia abajo: diez altos y cinco chaparros con razón de semejanza φ^{-1} a los originales; y, por último, cinco altos en la segunda escala áurea hacia abajo, es decir, con base φ^{-2} , que son los picos de la estrella.

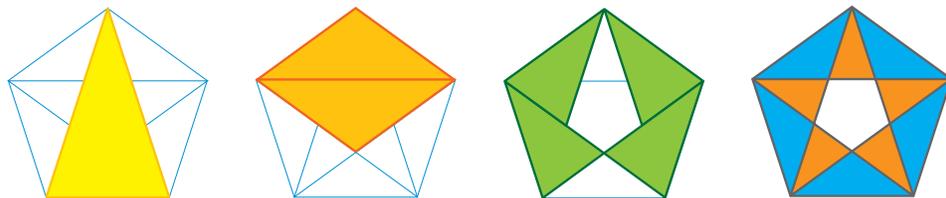


Figura 4.9 Triángulos áureos en el pentagrama.

Aunque el pentagrama y los triángulos áureos han sido estudiados, apreciados y usados por milenios, siguen dando de qué hablar. Recientemente, alrededor de 1975, el físico matemático inglés Roger Penrose descubrió una familia de mosaicos íntimamente relacionados con ellos que ahora se conocen como *mosaicos de Penrose*. Se construyen con dos piezas: los *papalotes* y las *dagas* que, a su vez, se construyen pegando dos triángulos áureos, altos y chaparros respectivamente. En el entendido de que cada vez que se pegan dos piezas, los extremos de las *curvas de Amman* coinciden (véase figura 4.10), se

puede llenar todo el plano. Pero siempre da lugar a mosaicos no periódicos, es decir, que no se pueden trasladar para regresar a sí mismos. Ésta es la propiedad que ha atraído la atención de físicos y químicos en los últimos 35 años, pues están relacionados con los llamados cuasicristales. Y, por supuesto, han sido también un deleite para los matemáticos pues tienen propiedades muy interesantes; una de ellas es que la proporción de papalotes a dagas siempre es áurea.

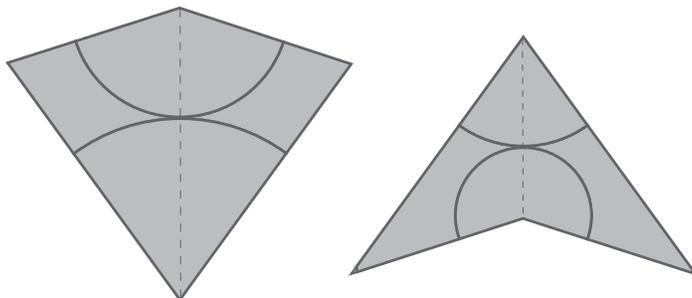


Figura 4.10 El papalote y la daga con los que se arman los mosaicos de Penrose. Los arcos de círculo, cuyos extremos parten las aristas en proporción áurea, deben coincidir al pegar dos piezas por una arista.



Figura 4.11 El mosaico de Penrose en la sala de matemáticas de Universum, realizado por el artista Juan Sandoval. Uno de los más hermosos en el mundo.

4.2.2 La sucesión de Fibonacci

Como sucede a menudo con conceptos fundamentales de las matemáticas, la razón áurea surge en varios contextos que, en primera instancia, parecen independientes. Uno de ellos es la famosa sucesión de Fibonacci:

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, \dots,$$

donde cada término es la suma de los dos anteriores. Es decir, si llamamos F_n al n -ésimo (léase “enésimo”) término de la sucesión, el que le sigue está definido por:

$$F_{n+1} = F_n + F_{n-1}. \quad (4)$$

En dicha ecuación (4) hay que suponer que $n > 1$ y declarar que los dos primeros son: $F_1 = F_2 = 1$.

Fibonacci, cuyo nombre real era Leonardo de Pisa, la introduce en 1202 para estudiar la reproducción de los conejos, en un modelo muy simple que es precursor de lo que ahora se llama biología matemática o biomatemática. Aunque ahora sabemos que muchos siglos antes, los matemáticos de la India conocían esta sucesión y la usaban.

Lo que nos interesa en este momento de la sucesión de Fibonacci, es que la razón entre términos consecutivos se aproxima a la razón áurea. Los primeros casos de estos

cocientes son: $1/1 = 1$, $2/1 = 2$, $3/2 = 1.5$, $5/3 = 1.666\dots$, $8/5 = 1.6$, $13/8 = 1.625$, $21/13 = 1.6153\dots$, $34/21 = 1.61904\dots$, ... Claramente se aproximan a $\varphi = 1.6803$ alternadamente por abajo y por arriba. Pero, al usar la notación moderna de límites, nos interesa ver que:

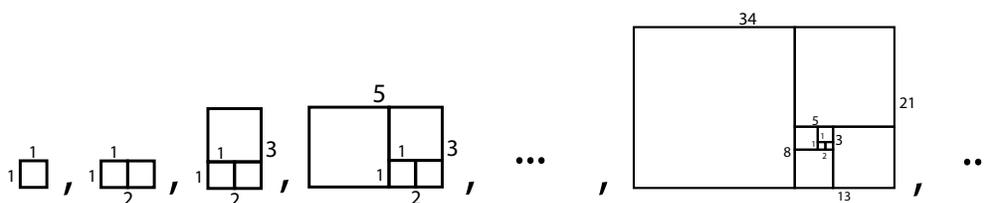
$$\lim_{n \rightarrow \infty} \frac{F_n}{F_{n-1}} = \varphi. \quad (5)$$

Esta ecuación se lee “el límite cuando n tiende a infinito de *efe-ene sobre efe-ene-menos-1* es igual a la razón áurea”; o bien “ F_n/F_{n-1} tiende a φ cuando n tiende a infinito”. Y quiere decir que entre más grande sea n , F_n/F_{n-1} aproxima mejor a φ .

Para verlo, recordemos nuestra primera definición de razón áurea quitándole un cuadrado a un rectángulo. Podemos usarla al revés: pegándole un cuadrado a un rectángulo áureo se obtiene un rectángulo más grande pero con la misma proporción. Este proceso de pegar cuadrados se puede aplicar a cualquier rectángulo e iterarse; en esa iteración aparecerá la sucesión de Fibonacci.

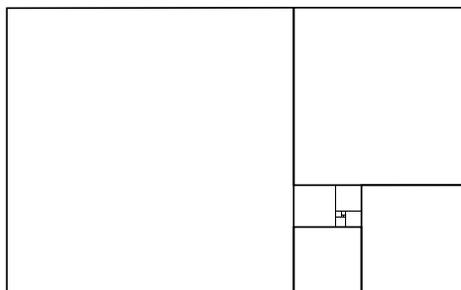
Si empezamos con un cuadrado unitario, 1×1 : al pegarle un cuadrado obtenemos un rectángulo de 1×2 ; luego le pegamos un cuadrado de 2×2 y obtenemos un rectángulo de 2×3 ; al pegarle un cuadrado 3×3 , obtenemos un rectángulo de 3×5 y así, sucesivamente (figura 4.12). Este proceso da una sucesión de rectángulos cuyos lados son términos consecutivos de la sucesión de Fibonacci.

Figura 4.12 Sucesión de rectángulos que se obtiene pegando cuadrados.



En la figura 4.12, nos saltamos tres términos y tuvimos que reducir el octavo (un rectángulo de 21×34) para que cupiera. Pero no importa, porque sólo nos interesan las proporciones, no los tamaños explícitos. Si de nuevo nos saltamos tres términos obtendríamos un rectángulo de 144×233 , que al volver a reducir es casi indistinguible del áureo, pues el error, evidente en los primeros términos, se ha hecho muy pequeño. De hecho, $|233/144 - \varphi| < 0.000022$.

Figura 4.13: Término 12 de la sucesión de rectángulos.



El argumento de que el “error se hace más pequeño” conforme avanzamos en la sucesión de rectángulos, convence intuitivamente. Pero muchos matemáticos, aunque no rebatirían la veracidad de la ecuación (5), pedirían una prueba más contundente; intuirían que se puede tener más control en cómo se da la aproximación para argumentar mejor un hecho que sucede en el infinito. Puede parecer superfluo, pero esta insistencia en demostrar las cosas es lo que da solidez a las matemáticas.

Si generalizamos, al tiempo que recapitamos en el argumento, obtendremos esa prueba. Por desgracia, puede ser que rebase el nivel de este libro pero, por las ideas que usa, puede resultar interesante y la incluimos.

Un rectángulo queda determinado por sus dos lados, que podemos codificar en una pareja ordenada de números (x, y) . El proceso de pegarle un cuadrado es, según lo que hemos hecho, cambiar esta pareja por la pareja $(y, x + y)$. Así, que si empezamos con la pareja $(x, y) = (1, 1)$, la que da origen a la sucesión de Fibonacci, y le aplicamos reiterativamente esta regla de cambio nos da las parejas $(1, 2), (2, 3), (3, 5), (5, 8), \dots$ que son las parejas de términos sucesivos. Pero esto nos da una regla para una *transformación* del plano cartesiano en sí mismo:

$$(x, y) \mapsto (y, x + y) . \tag{6}$$

Al entender geoméricamente esta transformación, obtendremos la prueba que buscamos.

El rectángulo áureo está representado por la pareja $(1, \varphi)$. Veamos qué le hace la transformación:

$$(1, \varphi) \mapsto (\varphi, 1 + \varphi) = (\varphi, \varphi^2) = \varphi(1, \varphi) ,$$

donde hemos usado la propiedad básica de la razón áurea ($\varphi^2 = \varphi + 1$), y denotamos por $t(x, y)$ a la pareja (tx, ty) . Si pensamos en términos de vectores: al *vector áureo* $(1, \varphi)$, la transformación simplemente lo alarga en proporción áurea al vector $\varphi(1, \varphi) = (\varphi, 1 + \varphi)$; como era de esperarse. A un vector con esta propiedad (de ser cambiado de escala) se le llama *vector propio* de la transformación. El otro vector propio es su ortogonal:

$$(-\varphi, 1) \mapsto (1, 1 - \varphi) = (1, -\varphi^{-1}) = -\varphi^{-1}(-\varphi, 1) .$$

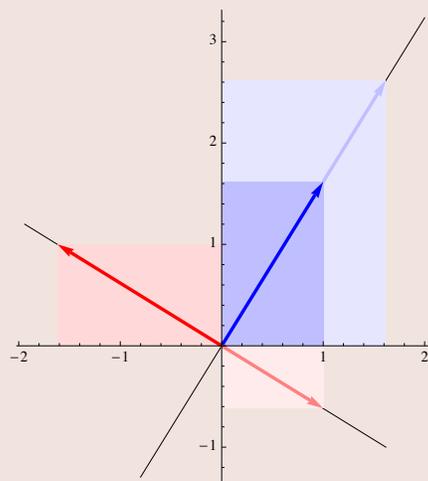


Figura 4.14 Qué le hace la transformación 6 a los vectores áureos $(1, \varphi)$ y $(-\varphi, 1)$.

Pero en este caso, el factor de escalamiento, o *valor propio*, $-\varphi^{-1}$, es negativo y de magnitud menor que 1. Éste será el meollo del asunto. Lo negativo, le invierte orientación a $(-\varphi, 1)$, y lo de magnitud menor que 1: ¡lo encoge!

Podemos entender la transformación $(x, y) \mapsto (y, x + y)$ en dos pasos. Primero, refleja en la recta de pendiente φ (lo negativo de $-\varphi^{-1}$), y después cambia de escalas: en la *dirección áurea* (del vector $(1, \varphi)$) hay un alargamiento áureo, de factor φ ; y en su dirección ortogonal, un encogimiento áureo inverso, de factor φ^{-1} . De tal manera que si aplicamos la transformación a cualquier punto (x, y) , se cambia de lado de la recta de pendiente φ pero se acerca a ella. Si, además, le pedimos que no esté en la recta de pendiente $-\varphi^{-1}$ e iteramos este proceso, se va hacia el infinito en la dirección de $(1, \varphi)$ o de $-(1, \varphi)$ (dependiendo de en qué lado estaba al principio) pues en cada paso se acerca a la recta en una proporción áurea inversa. Esto puede darnos el “control” que se necesita para precisar la aproximación, pero nos saltamos las fórmulas.

Al generalizar, hemos demostrado más de lo que pretendíamos. Casi cualquier sucesión tipo Fibonacci, es decir, dada por (4) pero con F_1 y F_2 arbitrarios, cumple (5). Pues hay que tomar el punto $(x, y) = (F_1, F_2)$ e iterar la transformación (6), para obtener la sucesión correspondiente. Sólo hay dos casos especiales. Cuando $F_1/F_2 = -\varphi^{-1}$, pues empezariamos con un punto en la recta de pendiente $-\varphi^{-1}$ y entonces las iteraciones se mantienen ahí pero tienden al $(0, 0)$; o bien, cuando $F_1 = F_2 = 0$ y la sucesión es constante 0.

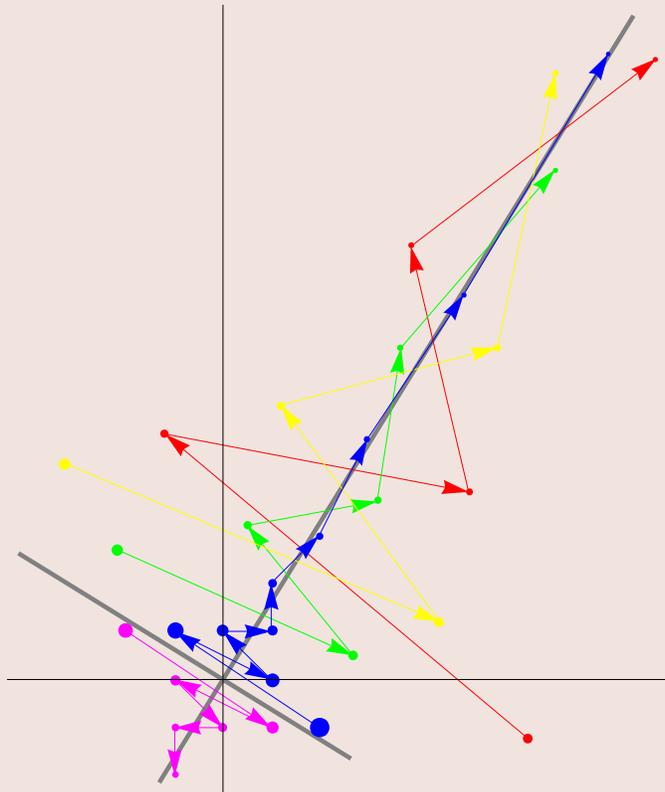


Figura 4.15 Iteraciones de (6) en diferentes puntos. Los azules pasan por la sucesión de Fibonacci.

En particular, hemos indicado cómo demostrar lo que queríamos de la sucesión de Fibonacci (la ecuación (5)). Se explica también por qué nuestros primeros cálculos se alternaban de menor a mayor que φ , y además, que en el proceso de añadir cuadrados a cualquier rectángulo, se obtiene como límite uno de proporciones áureas, pues la pareja (x, y) con que arrancaríamos la iteración tiene entradas positivas.

4.3 DE KÖNIGSBERG A GOOGLE



Figura 4.16 El suizo Leonard Euler (1707-1783) es considerado uno de los más grandes matemáticos de todos los tiempos.

4.3.1 Los puentes de Königsberg

La ciudad de Königsberg en la antigua Prusia —que ahora se llama Kaliningrado y está en Rusia— se hizo célebre entre los matemáticos por un problema que resolvió Euler a principios del siglo XVIII. El problema era más bien un divertimento social entre los habitantes de una ciudad que se enorgullecía de sus puentes. Pero dio lugar a un desarrollo teórico que en la actualidad tiene una enorme importancia en muchas áreas de la actividad humana. El problema, juego o desafío consistía en dar un paseo por la ciudad de Königsberg de tal manera que se cruzaran sus siete puentes sin repetir ninguno. El mapa del Königsberg de aquella época está en la figura 4.17.

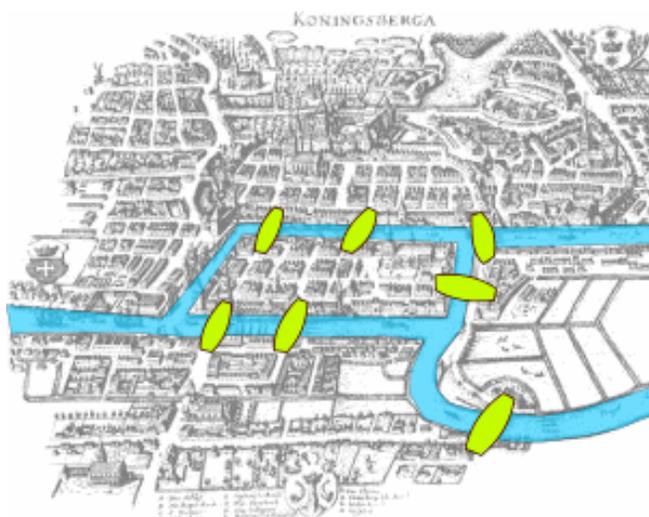


Figura 4.17: Los siete puentes de Königsberg alrededor de 1730.

El problema se hizo famoso porque no obstante lo sencillo de su planteamiento, nadie daba con una respuesta. La solución de Euler no es diseñar un paseo con las características que se exigen, sino demostrar y hacer evidente que es imposible hacerlo. Para entender su argumento, conviene abstraer la información esencial del problema. Son cuatro los bloques de tierra, que si representamos con pequeños círculos, que llamaremos *vértices*, unidos por aristas que representan a los puentes, se obtiene la *gráfica* de la figura 4.18.

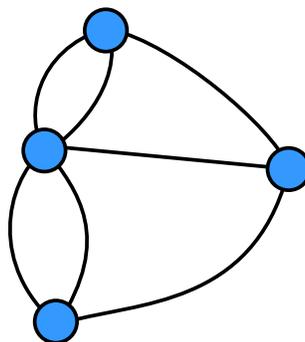


Figura 4.18 Gráfica de los siete puentes de Königsberg.

En lo que se refiere al problema, diseñar un paseo en esta gráfica, en un plano de Königsberg o echárselo a pie, en bici o en carroza por la ciudad, son equivalentes. Concéntrenos entonces en la gráfica, aunque sea válido hacer referencia a los otros dos enfoques. Un *paseo* consiste en una sucesión alternada de vértices y aristas que empieza y termina en vértices, en el entendido de que cada arista de la sucesión está entre los dos vértices que une —al pasear en la ciudad, cada puente comunica dos bloques de tierra—. Es decir, un paseo equivale a dibujar sobre la gráfica con un lápiz sin levantarlo nunca.

Esto remite a un problema común entre los niños de secundaria de la ciudad de México en los años cincuenta y sesenta, conocido como la *firma del diablo*. Consistía en dibujar un cuadrado con sus dos diagonales de un solo trazo: sin levantar el lápiz y sin dibujar sobre lo dibujado. Se insistía, para avivar la curiosidad, en que el diablo sí podía hacerlo.

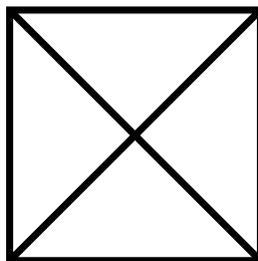


Figura 4.19 La firma del diablo.

Si ponemos un vértice en las cuatro esquinas y uno en el cruce del centro, el problema es equivalente al de los puentes de Königsberg; lo que ha cambiado es la *gráfica* en la cual se debe diseñar un paseo. Hemos *generalizado* el problema de los puentes de Königsberg a una *familia* de problemas: en cualquier gráfica se puede plantear. Y sucede muy frecuentemente que esto de generalizar ayuda a la solución.

La observación básica de Euler es que cada vez que en un paseo en una gráfica pasamos por un vértice, se usan dos de sus aristas —por la que entramos y por la que salimos. De tal manera que si un vértice —vale la pena ponerle un nombre para referirnos a él en específico, y qué más simple que una letra—, llamémosle v —para pensar en vértice—, no es aquel donde iniciamos o concluimos un paseo que no repite aristas, entonces habremos usado un número par de las aristas que inciden en v . ¿Cuántas? Justo el doble de las veces que pasamos por v . Revisemos la gráfica de Königsberg para concluir que el paseo que se pide en el problema es imposible. En cada vértice inciden un número impar de aristas —hay un vértice con 5 aristas y tres con 3—; ninguno de ellos puede ser intermedio en un paseo que no repite aristas y las usa todas. Y el mismo argumento se aplica a la firma del diablo: hay un vértice con cuatro aristas —el del centro—, pero hay cuatro —las esquinas— con tres aristas. Faltan candidatos a ser vértices intermedios. ¡Ni el mismísimo diablo puede firmarlo!

4.3.2. Paseos eulerianos

Cuando se usa una y otra vez la misma descripción o frase, los matemáticos acostumbramos ponerle un nombre para que el discurso sea más fluido. En honor al primero que los estudió, se les llama *paseos eulerianos* a los paseos en una gráfica donde no se repiten aristas y además pasan por todas ellas, como los que estábamos buscando en Königsberg o en la firma del diablo. Puede haber de dos tipos: *cerrado* cuando el paseo empieza y acaba en el mismo vértice o *abierto* cuando éstos son distintos. El otro concepto que se repite insistentemente es el del número de aristas que inciden en un vértice: llamémosle su *grado*. Podemos ahora limpiar y resumir el argumento de Euler.

Si una gráfica tiene un paseo euleriano cerrado, entonces todos sus vértices tienen grado par; las aristas donde se empieza y se acaba se pueden aparear para dar ahí el grado par. Pero el argumento da para más: si una gráfica tiene un paseo euleriano abierto, entonces todos sus vértices tienen grado par excepto el del principio y el del final —la arista con que empezamos y con la que acabamos no tienen con quién aparearse.

Euler, siendo un gran matemático, no se quedó satisfecho con su ingeniosa solución al problema de los puentes de Königsberg. Definió en general el concepto de *gráfica* como un conjunto de vértices (o nodos) junto con un conjunto de aristas, cada una de las cuales une (o *incide* con) dos vértices. Y se hizo la pregunta inversa: ¿será cierto que una gráfica que tiene todos los vértices de grado par admite un paseo euleriano cerrado? Y, si tiene únicamente dos vértices de grado impar, ¿tendrá un paseo euleriano abierto? Demostró que así es. Por ejemplo, si en Königsberg se construyera otro puente, cualquiera que sea, ya habría paseos eulerianos abiertos. O bien, cuando el diablo firma, pasa dos veces por una arista exterior pero lo hace con tal precisión y rapidez que no se nota el repaso.

Vale la pena reconstruir el argumento de Euler porque es indicativo del área de las matemáticas que ahora se llama *combinatoria* o *matemática discreta*.

Primero, observemos que el caso de dos vértices de grado impar se puede reducir al otro añadiendo una nueva arista entre los dos vértices especiales.

Dada una gráfica G —de nuevo será importante ponerle un nombre sencillo—, con todos sus vértices de grado par, iremos construyendo un paseo euleriano cerrado poco a poco. Pensemos que la tenemos dibujada con lápiz en un papel con circulitos de vértices y segmentos —rectos o curvos, pero simples— como aristas. Escogemos uno de sus vértices, llamémoslo v_0 —léase “v-cero”— y con un plumón rojo vamos recorriendo a G : armando paso a paso un paseo, con la única precaución de que no se vale pasar por una arista ya pintada de rojo (figura 4.20).

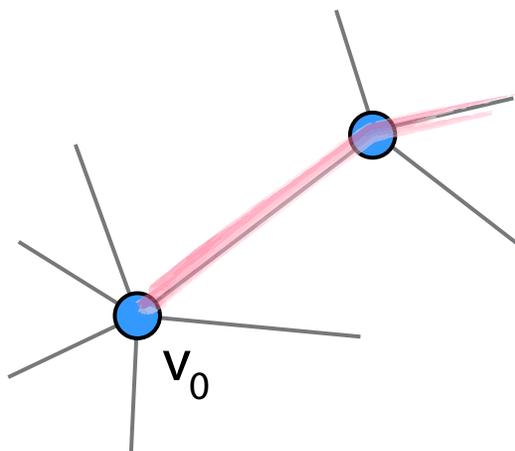


Figura 4.20
Inicio del paseo.

Cuando pasamos por algún vértice se marcan dos de sus aristas, así que de un vértice que no sea v_0 siempre podemos salir porque su grado es par: al llegar a él hay un número impar de aristas rojas y por lo tanto queda alguna en lápiz, por ella salimos y lo volvemos a dejar con “grado rojo” —y “grado en lápiz” — par. Entonces, la única manera en que nos atoremos y no podamos seguir adelante es regresando a v_0 y además cuando el resto de sus aristas ya son rojas (fig. 4.21).

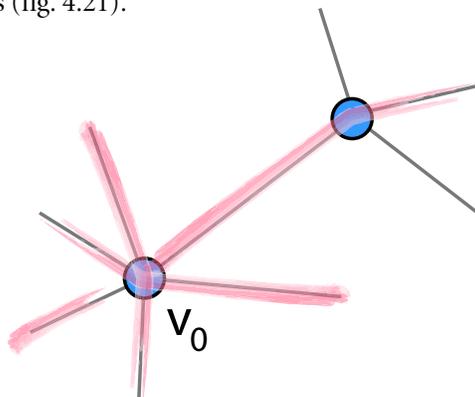
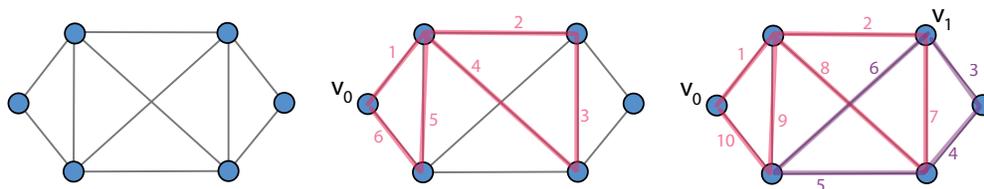


Figura 4.21 Final de la primera etapa.

En ese punto ya no tenemos por donde salir. Hemos construido un paseo cerrado que empieza y termina en v_0 y usa todas las aristas que inciden en él. Si toda la gráfica G ya es roja: acabamos, pues estamos de suerte y, al primer intento, encontramos el paseo euleriano que buscábamos. Si no es así, quedan en G algunas aristas en lápiz y mejoraremos nuestro intento, pero usándolo como base.

Volvemos a recorrer el primer paseo —justo el mismo recorrido— hasta el momento en que llegemos a algún vértice v_1 con aristas en lápiz. Hacemos ahí una pausa. Notemos que hay un número par de aristas en lápiz en v_1 —ya que hay un número par de rojas— y que en cada otro vértice también hay un número par de aristas en lápiz —que puede ser 0—. Salimos de v_1 por alguna de estas aristas en lápiz y seguimos y seguimos pintando más aristas de rojo. Por la misma razón que en el caso anterior este deambular sólo puede terminar regresando a v_1 cuando todas sus aristas ya son rojas. Aquí concluye la pausa y retomamos el camino de nuestro primer paseo que termina en v_0 . Este nuevo paseo, basado en v_0 , incluye más aristas rojas que el primero.

Figura 4.22 En una gráfica con todos los vértices de grado par: construcción de un paseo euleriano en dos etapas. La numeración en las aristas corresponde al orden del paseo.



Si no hemos pasado con el plumón rojo por todas las aristas de G , repetimos el mismo proceso pero ahora con nuestro paseo extendido como base y así, seguimos una y otra vez. En cada vuelta ampliamos el paseo y pintamos más aristas. Y para demostrar que este proceso termina en algún momento, que el paseo se vuelve euleriano, tenemos que suponer dos cosas que no habíamos hecho explícitas. Primero, que la gráfica G —tanto vértices como aristas— es finita, lo cual sobreentendíamos al pintarla en un papel y obliga a que tiene que llegar el momento en que el paseo ya no pasa por ningún vértice con aristas en lápiz. Y segundo, que G es *conexa*, es decir, que no consiste en dos o más pedazos aislados, porque entonces nuestro proceso sólo pinta de rojo a *la componente conexa* en que vive v_0 , nos da un paseo euleriano en ella, pero lo demás, como no tiene manera de comunicarse con v_0 a través de paseos, se queda pintada en lápiz, el plumón rojo nunca llega ahí.

4.3.3. La Red y la red

Éste fue el primer resultado de *teoría de las gráficas*, que ahora es un campo muy activo de las matemáticas. Tanto en sí mismo como por sus aplicaciones a muchas áreas del quehacer humano, pues las gráficas son objetos matemáticos que modelan muy diversas situaciones. Por ejemplo, *la Red*: ese universo informático, también conocido como internet, que está cambiando a pasos agigantados al mundo y la manera de interactuar de la humanidad. Puede pensarse que cada página web es un *nodo* o vértice de una gráfica; gigantesca y cambiante segundo a segundo, pero finita a fin de cuentas.

Y si una página web u incluye un *enlace* —o *link*— a otra página web v , se representa por una arista que va del vértice u al vértice v . Tiene la estructura de una *gráfica dirigida*, en la cual las aristas tienen *dirección*, principio y fin, van *de* un vértice *a* otro. Llamemos *la red*, con minúsculas, a esta gráfica dirigida. Se pierde en ella mucha de la información que hay en cada uno de los nodos, la información concreta de la página web en cuestión —así como en la gráfica de los puentes de Königsberg se pierde la información que hay en el mapa y en éste se pierde la información de la ciudad—. Pero la red modela algo fundamental de la Red. Representa su “conectividad” en abstracto.

En el último lustro del siglo xx, el estudio de la red como gráfica llevó a dos estudiantes de la Universidad de Stanford a diseñar el buscador Google que se ha convertido en la página web más importante y socorrida de La Red. Entre los ingredientes básicos de este buscador está el que como en cualquier gráfica, sus vértices adquieren un orden natural, llamado *espectral*, que ya se había estudiado. No tenemos herramientas para describirlo, pero con base en ese orden espectral es como se recorren los vértices de la red para hacer una búsqueda en la Red y a este método debe su éxito Google. El punto es señalar otra historia de cómo de un divertimento intelectual, surgen conceptos y objetos interesantes para los matemáticos que los estudian *per se*, y siglos después se aplican y cambian nuestro mundo.

4.4 LA CONQUISTA DEL INFINITO

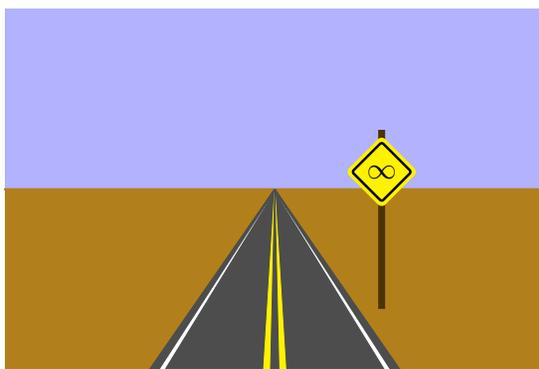


Figura 4.23 Cuando vemos cómo se pierde una carretera en el horizonte, nos acercamos a la forma de imaginar el infinito. Como el infinito fue por mucho tiempo, simplemente lo contrario de lo finito, no fue fácil aclarar el concepto y no fue sino hasta finales del siglo xix, cuando esto se resolvió.

4.4.1 Cómo medir lo infinito

Aristóteles había escrito que *el todo es más que la suma de sus partes*. Euclides, medio siglo después, lo acortó y dijo que *el todo es mayor que la parte*. Fue Galilei quien por primera vez observó que estas afirmaciones podían fallar si se trataba del infinito. En su libro *Consideraciones y demostraciones matemáticas sobre dos nuevas ciencias* dice, por boca de Simplicio, que “seguramente la infinitud de una línea corta es menor que la infinitud de

una línea más larga”; la respuesta de Salviati es que “eso son dificultades que resultan al tratar de comprender, con nuestro intelecto finito, algo infinito y que se atribuyen al infinito propiedades que se conocen de lo que es finito, pero que esto no es válido”.

El ejemplo que expone Salviati es interesante. En ello, argumenta que podemos elevar al cuadrado cada número y obtener un *cuadrado perfecto*, es decir, un cuadrado cuya raíz cuadrada es un número natural. De esta manera, se muestra que los números naturales y los números cuadrados *se corresponden uno a uno*. Lo anterior choca con nuestro sentido común, entrenado por el mundo de lo finito, de que una parte —en este caso, los números cuadrados— son sólo una parte de todos los números naturales. A este tipo de problemas se le conoció como *paradojas del infinito* y no se sabía en un inicio cómo lidiar con ellos ya que el infinito fue, por mucho tiempo, lo contrario de lo finito.

El problema de considerar conjuntos, es decir, dos colecciones de objetos, consiste en establecer un criterio claro para compararlos. Si se trata de conjuntos infinitos como, por ejemplo, el conjunto de números naturales, entonces no podemos contarlos: el proceso de contar —decir número tras número señalando un objeto tras otro hasta llegar al número asignado como último objeto que es el número de objetos en el conjunto— no funciona porque no hay un último objeto.

Por consiguiente, hay dos alternativas. La primera es la contención. Podríamos decir que los números cuadrados perfectos son menos que los números naturales porque los primeros están contenidos como una parte propia del segundo conjunto. Pero con este criterio no podemos comparar el conjunto de tres manzanas con el de dos peras, ya que peras no son manzanas y el conjunto de peras no es un *subconjunto*, es decir, una parte, del conjunto de manzanas. Vemos que el número es una abstracción y, al contar, hacemos una asignación uno a uno de los objetos con los primeros números.

La segunda alternativa consiste en tomar la correspondencia de uno a uno como criterio de comparación. Las consecuencias son sorprendentes. El hombre que trabajó con detalle esta idea fue Georg Cantor, matemático alemán, y lo hizo precisamente con esta manera de comparar que ya Galilei había propuesto. Con ello, concluimos que, en efecto, hay la misma cantidad de números cuadrados perfectos que de números naturales. Si entre dos conjuntos A y B no es posible establecer una correspondencia uno a uno pero sí es posible establecer una correspondencia del conjunto A con una parte del conjunto B , entonces diremos que A es de menor *cardinalidad* que B . Con este concepto vemos sin problema que el conjunto de dos peras es de menor cardinalidad que el conjunto de tres manzanas.

Los conjuntos se dividen en diferentes cardinalidades formando *clases de equivalencia* de conjuntos que tienen la misma cardinalidad, es decir, conjuntos entre los cuales es posible establecer una correspondencia uno a uno. Los cardinales finitos son aquellos que corresponden a los números naturales. Después, hay un primer cardinal infinito que corresponde al conjunto de los números naturales. Cantor denotó a este cardinal con \aleph_0 y se lee “alef cero” —aleph por ser la primera letra del alfabeto hebreo y cero, porque es el primer cardinal infinito—. Los conjuntos de cardinalidad se llaman también *numerables*, ya que una correspondencia con los números naturales indica una manera de numerar estos elementos. El elemento que corresponde al número 1 será el primero en la lista, el que corresponde al 2, el segundo y así, sucesivamente.

Cualquier conjunto infinito contiene una parte que tiene cardinalidad \aleph_0 , que se puede ver de la siguiente manera. Como el conjunto es infinito, podemos empezar a hacer una lista de elementos sin repetir ninguno. Esta lista siempre es finita y, por lo tanto, siempre se puede extender. De esta manera se obtiene un listado $\ell_1, \ell_2, \ell_3, \dots$ infinito con cardinalidad \aleph_0 .



Figura 4.24 Georg Cantor (1845-1918) fue un matemático alemán al que se le atribuye la invención de la teoría de conjuntos, base de las matemáticas modernas.

Lo anterior se puede aplicar justo a los números naturales. Si quitamos el primer número, seguimos con un conjunto infinito $\ell_1 = 2, \ell_2 = 3, \ell_3 = 4, \dots$ de cardinalidad \aleph_0 . Así, obtenemos una correspondencia uno a uno, como se observa en la figura 4.25, donde se muestra la correspondencia con flechas dobles.

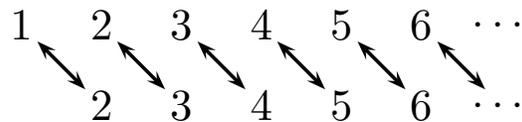


Figura 4.25 La correspondencia uno a uno de los números naturales con una de sus partes.

De hecho, cada conjunto infinito tiene la misma cardinalidad que una de sus partes y sólo los conjuntos infinitos tienen esta propiedad. En efecto, si A es un conjunto infinito, especificamos un subconjunto $L = \{\ell_1, \ell_2, \dots\}$ y denotamos con $B = A \setminus L$ lo que queda de A al remover los elementos de L . Ahora, sea $A' = A \setminus \{\ell_1\}$, es decir, quitamos únicamente el elemento ℓ_1 y establecemos una correspondencia uno a uno entre A y A' , como en la figura 4.25 entre L y $L \setminus \{\ell_1\}$ y la identidad entre B y B .

4.4.2 Diferentes infinitos

Cantor descubrió que hay más que un tipo de infinitos. Demostró que los números enteros—incluyendo los negativos— tienen la misma cardinalidad que los naturales y que, también los racionales, tienen la misma cardinalidad. Pero los números reales, aquellos que vimos en la sección 2.3 para expresar lo continuo, tienen otra cardinalidad, una que es mayor que \aleph_0 .

El argumento que dio Cantor se conoce hoy como el *argumento diagonal* y se basa en la contradicción, es decir, supone que los números reales son numerables. Si éste fuera el caso sería posible hacer una lista completa—aunque infinita— de los números reales. A continuación, se demuestra que hay un número que no está en la lista, lo que constituye una contradicción a la suposición de que la lista era completa pues contenía todos los números reales. Como en la argumentación no hay error alguno, se concluye que la suposición fue falsa y así queda demostrado que los reales no son numerables.

Veamos ahora cómo construir el número real que no está en la lista. Para ello, se usa

la misma lista que se suponía completa y se escribe cada número en su expansión digital. Después, se construye el número en su expansión decimal al usar los dígitos que se encuentran en la diagonal de la lista. Observemos cómo podría empezar la lista —dado que los dígitos de la diagonal juegan un papel importante, se marcaron con color rojo:

$$\begin{aligned}
 n_1 &= 4.151784200\dots \\
 n_2 &= 1.332699051\dots \\
 n_3 &= 0.067001004\dots \\
 n_4 &= 4.223991063\dots \\
 n_5 &= 8.010207501\dots \\
 n_6 &= 0.005141699\dots \\
 n_7 &= 8.112086032\dots \\
 n_8 &= 8.888888888\dots \\
 n_9 &= 3.141592653\dots
 \end{aligned} \tag{7}$$

De esta manera, obtenemos una sucesión de números:

$$1, 3, 7, 9, 0, 1, 0, 8, 3, \dots \tag{8}$$

Ahora bien, se construye un nuevo número x de manera que el dígito en posición n sea diferente del dígito en la misma posición de la sucesión (8). Esto se puede hacer, por ejemplo, de forma que cada dígito 1 se convierte en 0 y los otros dígitos en 1. En nuestro caso, obtendríamos:

$$x = 0.011110111\dots$$

Entonces, encontramos que $x \neq n_1$ ya que difieren en el primer dígito, que $x \neq n_2$ ya que difieren en el segundo y así, sucesivamente. La conclusión es que x no es ningún número de la lista y la lista fue incompleta. Hemos llegado a la contradicción que buscábamos.

Cantor compara diferentes conjuntos y es capaz de mostrar que la cardinalidad de los dos conjuntos es la misma:

$$\{(x, y) \mid y = 0, 0 \leq x \leq 1\} \quad \text{y} \quad \{(x, y) \mid 0 \leq y \leq 1, 0 \leq x \leq 1\} .$$

Los dos conjuntos describen el lado y la superficie de un cuadrado. Es decir, el conjunto de puntos de un segmento es igual que la cardinalidad de toda el área. Eso fue muy sorprendente. Con el uso de los *conjuntos potencia* se puede demostrar aun más: hay una infinidad de “diferentes infinitos”.



El conjunto potencia consiste de todos los subconjuntos de un conjunto dado. Por ejemplo, si $X = \{1, 2, 3\}$, entonces los subconjuntos de X son:

$$\{\}, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\} .$$

Las “llaves” $\{ \}$ indican la colección del conjunto. Entre ellas, se enlistan los elementos del conjunto, por ejemplo $\{1\}$ que es el conjunto que sólo contiene al elemento 1 mientras $\{ \}$ denota al *conjunto vacío*, es decir, aquel conjunto que no contiene ningún elemento. El conjunto potencia de un conjunto X se denota con (X) . Por ejemplo:

$$(\{1, 2, 3\}) = \{ \{ \}, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\} \}$$

Esto se ve muy complicado en la notación, pero no es tan difícil. Los elementos de (X) son conjuntos a la vez y, por ello, aparecen las llaves dentro de otras llaves.

Ahora se puede demostrar que la cardinalidad de (X) es siempre mayor que la cardinalidad de X . En efecto, se trata nuevamente de un argumento de contradicción. Se supone que X y (X) tienen la misma cardinalidad y que existe una correspondencia uno a uno entre ellos. Escribiremos esta correspondencia como una función:

$$e: X \rightarrow (X).$$

Entonces, para cada elemento $x \in X$, tenemos que $e(x) \subset (X)$ es un elemento de (X) y, por lo tanto, un subconjunto de X . Ahora nos podemos preguntar si el elemento x pertenece o no al subconjunto $e(x)$. Dividimos los elementos de X en dos subconjuntos S y N y decimos que $x \in S$ si x sí es un elemento de $e(x)$ y, por el contrario, si x no es un elemento de $e(x)$, entonces $x \in N$.

Como la función e establece una correspondencia uno a uno, hay un elemento n que corresponde al conjunto N , es decir $e(n) = N$. Luego, nos preguntamos si n es un elemento de $e(n)$ o si no lo es. Veamos, si $n \in e(n)$ entonces n es uno de los elementos de S y no de N . Por lo tanto, $n \notin N = e(n)$, que es justo lo contrario de $n \in e(n)$. Contrariamente, si $n \notin e(n) = N$ entonces $n \in S$ y, por definición, tenemos que $n \in e(n)$, ¡otra vez una contradicción! Esto muestra que seguro hay una contradicción si X y (X) tienen la misma cardinalidad.

Claro es que $X \rightarrow (X), x \mapsto \{x\}$ es una función *inyectiva*, es decir, diferentes elementos se envían a diferentes elementos. Finalmente, esto muestra que la cardinalidad de X es menor que la cardinalidad de (X) .

Con los conjuntos de potencia se muestra que hay una sucesión infinita de diferentes infinitos, cada vez más grandes.

4.4.3 Números ordinales y números cardinales

Ya vimos que hay muchos infinitos, y con los conjuntos potencia, sabemos que podemos encontrar cardinalidades cada vez más grandes. Se escribe $|X|$ para denotar la cardinalidad del conjunto X . En efecto, se puede mostrar que cualesquiera dos conjuntos X y Y siempre son *comparables*, es decir, que sólo hay tres opciones posibles: $|X| < |Y|$ o $|X| = |Y|$ o $|X| > |Y|$. En otras palabras, si las cardinalidades de X y Y no son iguales, entonces o el conjunto X tiene la misma cardinalidad que un subconjunto propio de Y o al revés, el conjunto Y tiene la misma cardinalidad de un

subconjunto propio de X . Esto no es para nada obvio y Cantor tuvo que introducir los *ordinales*; primero, demostró que cualesquiera dos conjuntos ordenados, siempre se pueden comparar. Como consecuencia los números cardinales, es decir, las cardinalidades siempre se pueden comparar en forma muy similar a como se comportan los números naturales o reales.

Dado que los cardinales se pueden comparar, Cantor pudo definir el cardinal \aleph_1 como el primer número cardinal que es mayor que \aleph_0 . El número cardinal \aleph_2 es el que sigue después de \aleph_1 y así, sucesivamente. Lo que, en cambio, no quedaba nada claro era la cardinalidad del conjunto de los números reales. Es casi seguro que Cantor pensara que los reales tienen la cardinalidad \aleph_1 , lo que hoy se conoce como la *hipótesis del continuo*. Pero unos 50 años después de él, se descubrió que esta pregunta es *independiente* en el sistema de axiomas de Zermelo-Fraenkel. Como consecuencia, se puede hacer matemáticas con la hipótesis de que los reales tienen cardinalidad \aleph_1 o con la hipótesis de que la cardinalidad de los reales es mayor que \aleph_1 .

Lo anterior debería ser algo inquietante pues afirma que el conjunto de los números reales no se conoce bien, ni siquiera en lo que se refiere a sus subconjuntos. No se puede determinar si hay o no un subconjunto X de números reales que tenga una cardinalidad intermedia entre los naturales y los reales o si tal subconjunto no existe.

4.4.4 La base formal de las matemáticas

Incluso en tiempos de Cantor, un conjunto era simplemente una colección de objetos que tenía la única restricción de que a partir de un conjunto siempre se debía poder decidir si algún objeto pertenecía o no a dicho conjunto. Con esta definición tan amplia se llegó rápidamente a contradicciones, entre las cuales la que formuló Russell es la más famosa.

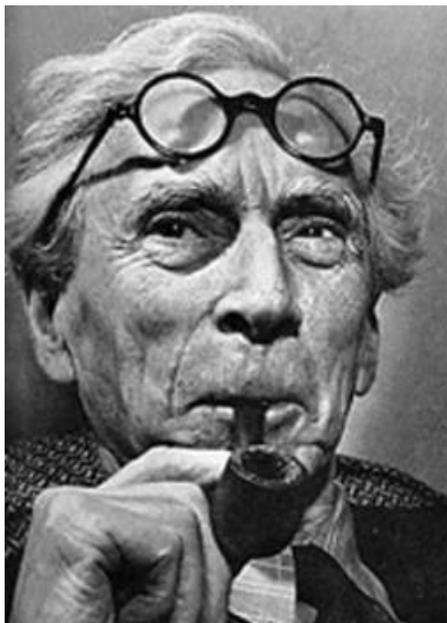


Figura 4.26 Bertrand Russell (1872-1970) fue un matemático, filósofo y escritor británico que también se dedicó al pacifismo y a defender los derechos de las mujeres.

Russell dice que conoce un pueblo donde el único barbero afeita a todos los hombres que no se afeitan por sí mismos. A primera vista, esto suena totalmente plausible. La formulación implica que cada hombre es afeitado: si no lo hace él solo, entonces visita al barbero. Con los hombres del pueblo se forman dos subconjuntos: los que

se afeitan solos y los que visitan al barbero. Ahora, ¿a cuál subconjunto pertenece el barbero?

Al hacer las conclusiones lógicas se deduce rápidamente que, de cualquier manera, se llega a una contradicción. Con ello, Russell muestra que los conjuntos no se pueden definir de una manera tan laxa. En efecto, la definición que hoy opera para un conjunto es más restrictiva y evita que se generen este tipo de contradicciones.

Y aunque la discusión sobre cuál debe ser exactamente la base de las matemáticas no se ha concluido, es un hecho que los conjuntos son los que forman la base moderna para las matemáticas. A partir de los conjuntos, se forman todos los demás conceptos desde un punto de vista más formal. Como ejemplo, veremos de qué manera el concepto de *función* se basa en el de conjuntos. Una función tiene un *dominio* A y un *codominio* B y “asigna” a cada elemento de A , un elemento de B . Ahora, esto se transforma en el lenguaje abstracto de conjuntos. Para ello, partimos de dos conjuntos A y B y definimos primero el concepto de *producto cartesiano* $A \times B$, que consiste en los pares ordenados (a, b) con $a \in A$ y $b \in B$. Formalmente, también se debe definir *par ordenado* con base en conjuntos, esto lo indicamos al lector interesado en el siguiente recuadro.

Si a y b son dos elementos, entonces $\{a, b\}$ es el conjunto que contiene los dos elementos a y b . Consecuentemente, tenemos $\{a, b\} = \{b, a\}$ ya que en un conjunto no hay orden alguno y sólo se puede saber si un elemento le pertenece o no. En matemáticas, se suele denotar con (a, b) al *conjunto ordenado* que contiene a a como primer elemento y a b como segundo. Esto se puede realizar de la siguiente manera. Se define (a, b) como el conjunto:

$$\{ \{a\}, \{b, \{\}\} \},$$

es decir, el conjunto que tiene los dos elementos: $\{a\}$ y $\{b, \{\}\}$.

El primero de estos dos conjuntos tiene un solo elemento —que es a — mientras el segundo conjunto tiene dos elementos: b y el conjunto vacío. De esta manera se pueden recuperar los dos elementos sin confundirse: vemos la cardinalidad de los dos elementos y concluimos, a partir de ello, cuál es el primer elemento y cuál el segundo.

Con el concepto del par ordenado podemos formar ternas $(a, b, c) = (a, (b, c))$, cuartetos o n -adas con n elementos.

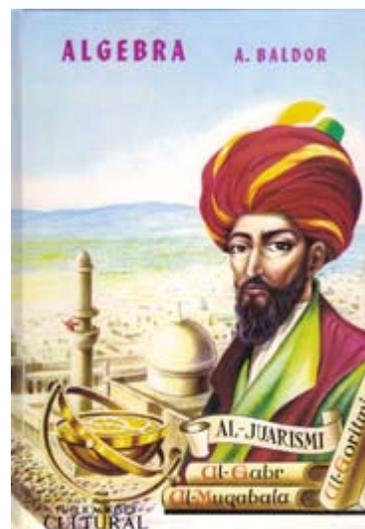
Un subconjunto $\Sigma \subset A \times B$ —esto se lee “sigma subconjunto de A cruz B ” — se llama *gráfica* si para cada $a \in A$, existe un $b \in B$ —uno y solamente uno— tal que $(a, b) \in \Sigma$. La condición realiza lo que antes se expresó: “la función asigna a cada elemento $a \in A$, un elemento $b \in B$ ”. Finalmente, una *función* es un triple (A, B, Σ) donde A, B son conjuntos y $\Sigma \subset A \times B$ es una gráfica.

Los conjuntos no sólo sirvieron para liberar y aclarar el concepto del infinito sino, también, para dar formalidad a todos los conceptos matemáticos en la actualidad. En este sentido, han sido el gran artículo de moda matemática para todo el siglo xx.



4.5 EL DESARROLLO DEL ÁLGEBRA

Figura 4.27 La imagen muestra el libro de álgebra del docente cubano, Aurelio Baldor. Centrado en la aritmética y el cálculo con variables es, sin duda, el libro más consultado de álgebra en América Latina. La portada presenta al matemático persa Muhammad ibn Musa Al-Juarismi, que escribió un libro llamado *Al-Kitab al-Jabr wa-l-Muqabala* y que significa “Compendio de cálculo por el método de completado y balanceado”.



4.5.1 La aritmética

La *aritmética* se encarga de la operación de números; es la rama más antigua de las matemáticas y, además de la geometría, la que más desarrollo ha generado con el pasar de los siglos.

Al principio, los fines de la aritmética eran sólo prácticos; por ejemplo, enseñaba cómo hacer cálculos correctamente y, además, de manera eficiente.

No debe olvidarse, como ya se mencionó en el capítulo 2, que la operación con números requiere de considerable destreza si uno no cuenta con herramientas auxiliares como un ábaco en tiempos pasados, una regla de cálculo hace 50 años, o una calculadora de bolsillo en la actualidad.

La representación de los números en un *sistema posicional* —como el decimal de los árabes— fue fundamental para el desarrollo de *algoritmos*, con los cuales se pudieron efectuar las operaciones básicas.

Fueron precisamente los griegos quienes establecieron las primeras propiedades de los números —los pitagóricos estaban particularmente fascinados con estos números—. Ellos llamaron número *n perfecto* a aquellos donde la suma de sus divisores propios, es decir, los divisores menores que dicho número, resultaba justamente *n*. Por ejemplo, 6 es perfecto: sus divisores propios son 1, 2 y 3 que suman 6. Pero 5 no es perfecto, ya que sólo tiene un divisor propio que es el uno. Los números que sólo tienen al 1 como divisor propio se llaman *primos*. Por ejemplo, 5 es primo y 6 no lo es.

Los griegos mostraron que cada número positivo se descompone en factores primos y que esta descomposición es esencialmente única, pues diferentes descomposiciones sólo difieren en el orden. Por ejemplo $24 = 2 \cdot 2 \cdot 2 \cdot 3 = 2 \cdot 3 \cdot 2 \cdot 2$. Al hecho de que se puede y que es único, se le conoce como el *teorema fundamental de la aritmética*.

Además, los griegos demostraron que hay un número infinito de primos. Esto lo demostraron por contradicción, es decir, se supone lo contrario y con ello se llega a una contradicción. En efecto, si solamente hubiera un número finito estos primos se podrían poner en una lista y numerarlos como p_1, p_2, \dots, p_N . Entonces, se formaría el número:

$$x = p_1 \cdot p_2 \cdot \dots \cdot p_N + 1.$$

Este número no puede ser divisible entre p_1 porque si no, también $x - p_1 p_2 \dots p_N = 1$ sería divisible entre p_1 , lo cual no puede ser dado ya que $p_1 > 1$. De manera similar, se muestra que x tampoco es divisible por ninguno de los primos p_2, \dots, p_N . En conclusión, x no es producto de primos, lo que contradice el teorema fundamental de la aritmética. Hemos encontrado una contradicción.

Una revisión minuciosa de la argumentación muestra que no se ha cometido ningún error y, por lo tanto, no puede ser cierta la suposición de que hay sólo un número finito de primos.

4.5.2 El largo nacimiento de la notación algebraica moderna

Desde los más antiguos documentos que reportan la enseñanza de las matemáticas, destacan problemas formulados para buscar algún número que cumpla con cierta propiedad. Uno de estos documentos es el papiro Rhind, que se considera una colección de problemas con los cuales se enseñaban las matemáticas en Egipto, 2 000 años antes de nuestra era. El problema 26 del papiro Rhind es:

Una cantidad y una cuarta de sus partes son 15, ¿cuál es esta cantidad?

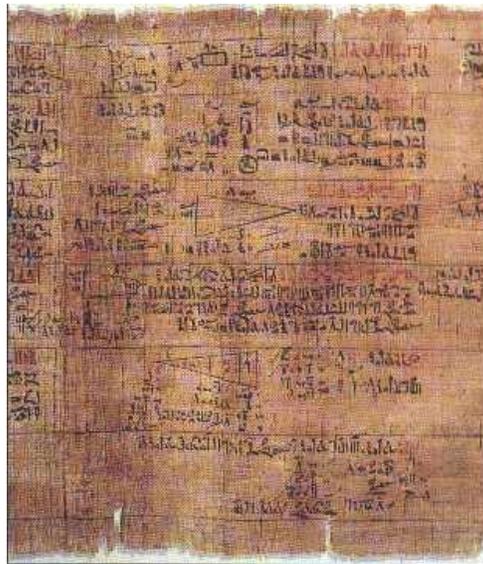


Figura 4.28 El papiro de Ahmes, también conocido como de Rhind es un documento de contenido matemático que data aproximadamente del 1680 a.C. y se le atribuye al escriba Aahmes, fue encontrado en el siglo XIX; entre las ruinas de Luxor, en Egipto. Actualmente, reside en el Museo Británico de Londres.

Este problema se podría modelar hoy con la siguiente ecuación lineal $x + \frac{x}{4} = 15$. Sin embargo, ésta es una notación moderna. En India, usaban la palabra *ya* —de *yavat tavat*— para llamar a la incógnita principal, mientras que los nombres de colores se utilizaban para denotar otras variables. Al-Juarismi usaba “shai” para denotar incógnitas, que después se tradujo al latín en “res” o “causa” y que, en italiano, se transformó en “cosa”. Los alemanes copiaron el sonido y lo escribieron “coss” y, por ello, durante cierto tiempo a los matemáticos se les conocía como los “cosistas”. Hasta aquí se empleaban palabras y sílabas cuyo uso variaba en el tiempo y el espacio, es decir, se utilizaba el lenguaje normal para expresar problemas o ecuaciones. Por ejemplo, Leonardo de Pisa escribió en 1202:

El cubo y siete cosas menos 5 cuadrados es igual a la raíz de la cosa más 6.

Para no confundirnos, aclaramos que se trata de la siguiente ecuación:

$$x^3 + 7x - 5x^2 = \sqrt{x + 6}$$

En un manuscrito de 1485, Nicolas Chuquet usó la notación 5^3 para expresar lo que hoy escribiríamos como $5x^3$.

En 1590, François Viète, un francés, usó consonantes para constantes y vocales para incógnitas, pero escribía, por ejemplo, “a cubum” para lo que hoy denotaríamos como a^3 . Descartes, en 1637, combinaba estas ideas aunque con cambios: usaba las últimas letras del alfabeto para denotar incógnitas y las primeras para constantes. Él hubiera escrito:

$$x^3 + 7x - 5x^2 \propto \sqrt{x + 6},$$

ya que el símbolo de igualdad que usamos en la actualidad todavía no se había establecido.

Aunque la matemática no se reduce a la notación, ésta sí clarifica y ayuda al pensamiento. Más aún, posibilita conceptos completamente nuevos. Si escribimos $x^2, x^3, x^4, x^5, \dots$ para expresar las potencias de x , podemos estar más tentados de pensar en expresiones como x^{-2} o $x^{\frac{1}{2}}$, es decir, la notación puede sugerirnos una *generalización*, cosa que no pueden los nombres “censo”, “cubus”, “censo de censo” y “primo relato” que usaba, por ejemplo, Niccolo Tartaglia y se basa en una tabla amplia de Luca Pacioli de finales del siglo xv.



Figura 4.29 El libro *Trattato di Numeri et Mesure* de Niccolo Fontana Tartaglia fue escrito en 1556 y es un tratado extenso en el que se encuentra buena parte de la notación usada en la época.

Además, la notación sugiere nuevas leyes de manera más fácil. Por ejemplo, $x^2 \cdot x^3 = x^5$ se hubiera escrito como “ce.cu-p°.r°” ya que se usaban abreviaturas como “ce” para “censo”, “p°.r°” para “primo relato”, o la línea para indicar la igualdad. Pero era imposible formular la generalización: $x^a \cdot x^b = x^{a+b}$. De esta manera se ve que la notación sí influye en las ideas, en cómo hacer matemáticas y en cómo, una notación adecuada, propicia el desarrollo mientras que una complicada, lo inhibe. La notación algebraica se estabiliza más o menos a la mitad del siglo xvii y se puede decir que es una de las grandes aportaciones a las matemáticas y a la ciencia, en general.

4.5.3 Ecuaciones lineales, cuadráticas, cúbicas y de cuarto grado

El método de resolución que se indica en el papiro Rhind fue enseñado hasta principios del siglo xx. En un documento conservado del segundo siglo antes de nuestra era se considera el siguiente problema:

Para 3 gavillas de una buena cosecha, 2 de una mediana cosecha y una de una mala cosecha se reciben 39 Tou. Para 2 gavillas de una buena cosecha, 3 de una mediana cosecha y una de una mala cosecha se reciben 34 Tou. Para 1 gavillas de una buena cosecha, 2 de una mediana cosecha y 3 de una mala cosecha se reciben 26 Tou. ¿Cuánto se recibe para cada gavilla de una buena, de una mediana y de una mala cosecha?

Hoy diríamos que en este problema se trata de plantear un sistema de ecuaciones lineales. El método que se indica en el documento chino es el que hoy conocemos como *algoritmo de Gauss*; no obstante Gauss vivió dos milenios después de Fang Cheng, el autor del documento. Pero, la historia rara vez hace justicia al atribuir los descubrimientos y tiende a dar a quien ya tiene.

Problemas como los que se han presentado sugirieron la invención de variables, una hazaña que tardó varios siglos y tuvo muchos altibajos. Los matemáticos árabes, en particular después de la obra de Al-Khwárizmi, introdujeron propiamente el *álgebra elemental* como la conocemos hoy día, comprendiendo el concepto de ecuación algebraica y de polinomios, la solución numérica de ecuaciones y la construcción geométrica de soluciones.

Al-Khwárizmi nació alrededor de 790, vivió y trabajó en Bagdad y murió, aproximadamente, con sesenta años. Su libro contenía una clasificación de ecuaciones cuadráticas y cómo resolverlas de manera geométrica. Los tipos de ecuaciones que trataba Al-Khwárizmi se leen en la notación moderna como:

- | | | |
|--------------------|--------------------|--------------------|
| 1. $ax^2 = bx$ | 2. $ax^2 = b$ | 3. $ax = b$ |
| 4. $ax^2 + bx = c$ | 5. $ax^2 + c = bx$ | 6. $bx + c = ax^2$ |

Hay que notar que sólo consideraba soluciones y coeficientes positivos. Luego, dividió estas ecuaciones entre el coeficiente a , con el fin de obtener “formas normales”. Para cada uno de estos tipos deriva una fórmula de solución usando ejemplos y argumentaciones geométricas. Por ejemplo, la argumentación geométrica para resolver la ecuación $x^2 + q = px$ se basa en la figura 4.30.

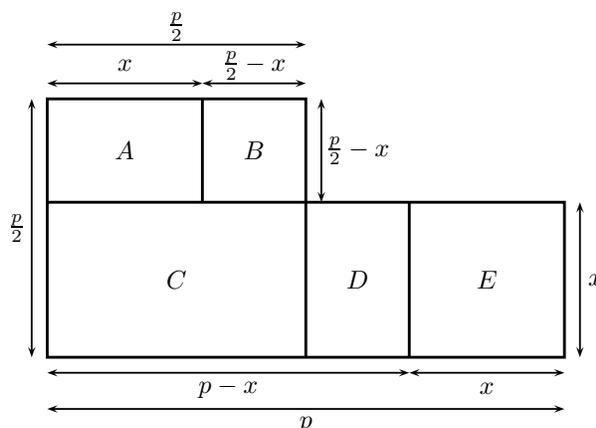


Figura 4.30 La figura geométrica usada en la demostración de la fórmula de solución por Al-Juarismi.

En la figura se ve que el rectángulo C tiene la mitad de ancho que CDE juntos. Por ello, $C = D + E$ si con los mismos símbolos se denotan las áreas. Además $C + D = (p - x)x = q$, según la ecuación que se quiere resolver, que es $x^2 + q = px$. Finalmente, se observa que A y D son congruentes. Entonces, podemos calcular:

$$B = \left(\frac{p}{2}\right)^2 - A - C = \left(\frac{p}{2}\right)^2 - D - C = \left(\frac{p}{2}\right)^2 - q,$$

es decir,

$$\left(\frac{p}{2} - x\right)^2 = \left(\frac{p}{2}\right)^2 - q$$

lo que ahora, se puede resolver fácilmente pues:

$$\begin{aligned}\frac{p}{2} - x &= \sqrt{\left(\frac{p}{2}\right)^2 - q} \\ x &= \frac{p}{2} - \sqrt{\left(\frac{p}{2}\right)^2 - q}\end{aligned}$$

De esta manera, se comprende mucho mejor el título de la obra que se traduce como “Compendio de cálculo por el método de completado y balanceado”, ya que el argumento anterior muestra cómo se balancean los términos para lograr expresar la incógnita en un cuadrado.

Después de Al-Khwárizmi hubo muchos otros matemáticos árabes, de los cuales destacaron, principalmente, el matemático persa Al-Karaji quien usó los monomios $x, x^2, x^3 \dots$ y $\frac{1}{x}, \frac{1}{x^2}, \frac{1}{x^3}, \dots$ sin representación geométrica y después, Omar Khayyam quien clasificó y resolvió la ecuación cúbica al usar intersecciones de secciones cónicas.

Los conocimientos de los árabes se difundieron poco a poco en Europa, durante los siglos XII, XIII y XIV; se sabe también que a finales del siglo XV, las principales obras de griegos como Euclides, Apolonio, Diofanto, Herón y Arquímedes, entre otros, ya se conocían. Fue precisamente en esta época cuando Tartaglia, Cardano y Ferrer desarrollaron la solución general de la ecuación algebraica de tercer y cuarto grado que detallamos en la sección 4.11.

Los ejemplos mostrados en esta sección ponen de manifiesto que el álgebra básica se desarrolló en diferentes regiones del planeta, de manera independiente.

Lo que sigue es una de las historias más asombrosas y que sólo ocurrió una vez, en Europa.

4.5.4 Nuevos horizontes

El cambio al siglo XVI trajo dos grandes personajes: Galileo Galilei, quien liberó la ciencia de las ataduras de la escolástica y dio a las matemáticas su lugar moderno, y René Descartes, quien inauguró el pensamiento autónomo frente a la fe e inició una unión entre las dos ramas principales de las matemáticas en aquel tiempo, es decir, la geometría y el álgebra. Esta unión se conoce hoy como geometría analítica y consiste en el uso de coordenadas en el plano y el espacio y en la representación de objetos geométricos mediante ecuaciones. La geometría analítica tuvo mucho impacto tanto para la geometría como para el álgebra.

En 1770, aparece el libro *Instrucciones completas para el álgebra* de Leonhard Euler, que reúne la aritmética y la teoría de ecuaciones, y está escrito con tanta claridad que fue reeditado múltiples veces. En él, ya aparecen los números imaginarios, pero no fue

sino hasta unos veinte años después que el concepto de *números complejos*, finalmente, se formalizó.

Más o menos al mismo tiempo, aparece la primera demostración rigurosa del *teorema fundamental del álgebra* que afirma que cualquier ecuación algebraica de grado positivo con coeficientes racionales tiene, al menos, una solución compleja. Es importante resaltar que el teorema sólo afirma la existencia en abstracto sin dar ninguna solución concreta, es decir, el resultado asegura que hay al menos una solución, pero no ayuda a encontrarla. Por ello, la búsqueda de fórmulas para las soluciones cobra aún más importancia.

No obstante que las ecuaciones lineales, cuadráticas, cúbicas y cuárticas podían resolverse con fórmulas, para la ecuación de quinto grado no se conocía ninguna fórmula. Fue un joven sueco, Niels Henrik Abel, quien demostró que dicha empresa es imposible. Vale la pena enfatizar que el teorema fundamental del álgebra afirma que las soluciones existen, mientras que el resultado de Abel implica que dichas soluciones no se pueden encontrar con una fórmula usando sólo las operaciones básicas y raíces, como se mencionó en la sección 4.9.



Figura 4.31 Sello postal fabricado en honor a Leonhard Euler conmemorando los 300 años de su nacimiento.

Sin embargo, la demostración es técnica y hoy se suele usar la teoría de Galois, llamada así por otro joven, el francés Evariste Galois quien vislumbró una teoría completa para tratar en general la problemática de resolución de ecuaciones algebraicas. Llegamos a un parteaguas en la historia: aunque lo que demuestra la teoría de Galois no es novedoso —Abel ya lo había demostrado unos años antes—, dicha teoría es la que desató realmente el álgebra en su versión moderna.

4.5.5. El álgebra moderna

Galois estudia cómo ciertas permutaciones de las raíces de un polinomio, es decir, las soluciones de una ecuación algebraica, corresponden a ciertos *campos intermedios*, véase sección 4.9. En lo anterior, hay dos nociones abstractas: el *grupo* y el *campo*. La formalización de estas nociones fue tardía.

Hoy se define al grupo como un conjunto G junto con una *operación binaria*: $\mu : G \times G \rightarrow G$, es decir, una función del producto cartesiano $G \times G$ al conjunto G con ciertas propiedades. El *producto cartesiano* $G \times G$ es, por definición, el conjunto de pares (a, b) de elementos a, b de G . En otras palabras, la función μ asigna a cada par de elementos un elemento de G .

Un ejemplo de lo anterior, es la suma de números enteros que asigna a cada par de números, su suma; en este caso se denota $\mu(a, b) = a + b$.

La multiplicación y la exponencial son también otras funciones y se denotan $\mu(a, b) = ab$ y $\nu(a, b) = a^b$, respectivamente.

Para que G sea un grupo debe satisfacerse la *propiedad de asociatividad*, que se escribe de la siguiente manera si $\mu(a, b) = ab$: para todo a, b, c en G se tiene $(ab)c = a(bc)$. La segunda propiedad es en la que se pide que exista un *elemento neutro* e , esto es, un elemento que satisfaga $ea = a = ae$, para todo elemento a de G . La tercera y última propiedad que se pide es que, para cada elemento a de G existe un *inverso multiplicativo*, es decir, un elemento b tal que $ab = e = ba$, donde e es el elemento neutro.

Debe observarse que no se requiere que $ab = ba$ para todo a y b de G . Si esta propiedad también se satisface, se dice que el grupo es *conmutativo*, o también se dice que el grupo es *abeliano* en honor a Abel, quien se dio cuenta de la importancia de la conmutatividad para poder resolver una ecuación con radicales.

Algunos ejemplos de grupos conmutativos son los números enteros con la adición o los números racionales distintos de cero con la multiplicación como operación binaria. Un grupo no conmutativo es el conjunto de todas las permutaciones del conjunto $I = \{1, 2, 3, 4\}$, es decir, todas las funciones biyectivas de I en I . La operación binaria en este caso está dada por la composición de funciones.

La segunda noción es la de *campo*. Un campo es —en el lenguaje moderno— un conjunto con dos operaciones que suelen denotarse como suma y multiplicación, tal que la suma es asociativa, conmutativa, admite elemento neutro 0 e inversos, mientras que la multiplicación es asociativa, conmutativa admite elemento neutro 1 y cada elemento no cero tiene inverso multiplicativo. Ejemplos de campos así son los números racionales, los reales o los complejos, pero también campos como $(\sqrt{2})$, esto es, los números reales que tienen la forma $a + b\sqrt{2}$ con a y b racionales.

Hoy día hay teorías completas que se llaman *teoría de grupos* y *teoría de campos*, y que se dedican al estudio de cada una de estas dos nociones. Una vez que se han formulado dichas nociones, es fácil obtener nuevas estructuras.

Por ejemplo, un *anillo con unidad* se define en forma muy similar a un campo: es un conjunto con dos operaciones binarias denotadas como suma y multiplicación, con la única diferencia de que en la multiplicación no se pide ni conmutatividad ni existencia de inversos. Ejemplos de anillos son los polinomios $[x]$, con coeficientes racionales en la incógnita x , que se pueden sumar y multiplicar. El polinomio constante 1 es la unidad multiplicativa, pero ningún polinomio de grado positivo tiene un inverso multiplicativo.

Aunque la noción de anillo es una generalización de la de campo, la *teoría de anillos* es muy distinta a la teoría de campos, pues se requieren herramientas diferentes para trabajarlas. Una noción importante para la teoría de anillos es la de un *ideal*. Un ideal I de un anillo A es un subconjunto $I \subseteq A$ que contiene el 0, es cerrado bajo la suma, es decir $i_1 + i_2$ es un elemento de I para cada i_1, i_2 de I . Además, el inverso aditivo de cada elemento de I es, de nuevo, un elemento de I y, con respecto a la multiplicación, se tiene que para cada a de A y cada i de I , también los elementos ai, ia pertenecen a I . Para campos, el concepto de ideal no es interesante ya que siempre hay sólo dos ideales: todo el campo y $\{0\}$, el ideal cero.

Para anillos e ideales hay muchos ejemplos naturales. Por ejemplo, los polinomios de grado mayores o iguales que 3 forman un ideal de $[x]$, o más general, para cada polinomio f de $[x]$ los múltiplos de f , es decir, los polinomios gf —donde g es cualquier polinomio— forman un ideal que se denota por (f) .

Un ideal se llama *primo* si satisface la siguiente propiedad: si un producto a_1a_2 pertenece a I , entonces, al menos uno de los dos factores pertenece a I . Veamos esto en el ejemplo del anillo Z de los números enteros con la suma y multiplicación usual. El ideal

(5) consiste en todos los múltiplos de 5 y es un ideal primo: si $a_1 a_2$ es un múltiplo de 5, entonces a_1 o a_2 tienen que ser múltiplos de 5. En cambio, el ideal (6) no es primo ya que el producto 2×3 está en (6) pero ninguno de los factores está en (6). Se ve que el ideal (f) es primo exactamente cuando el número f es primo o es cero.



Figura 4.32 Amalie Emmy Noether (1882-1935) fue una matemática alemana destacada por sus contribuciones al álgebra.

La formalización de estos conceptos y la abstracción del contexto de las ecuaciones tardó más de un siglo y culminó a principios del siglo xx. En este proceso incidieron muchos matemáticos, como el inglés William Rowan Hamilton y los alemanes Ernst Eduard Kummer, Emmy Noether —una de las mujeres matemáticas más destacadas— y Emil Artin, entre muchos otros.

En 1930, el matemático van der Waerden publicó un libro llamado *Álgebra moderna*, en donde, por primera vez, se definen todos estos conceptos —y muchos más— en el lenguaje que se sigue usando hasta hoy y que se basa en la teoría de conjuntos, establecida a principios del siglo xx.

4.5.6 Álgebra lineal

Hemos dejado la solución de ecuaciones lineales desde la mención de que los chinos ya conocían un método general que se llama hoy el *método de Gauss*, lo que puede interpretarse correctamente como que en estos 2 000 años, no hubo avance sustancial. La situación cambia radicalmente durante la segunda mitad del siglo xix. Matemáticos como Arthur Cayley y Hermann Günther Grassmann lucharon por la noción de *espacio vectorial*, que se hace respecto a un campo k . Hoy, se define un espacio vectorial como un grupo abeliano V junto con una función:

$$k \times V \rightarrow V \quad (9)$$

tal que se satisfacen una serie de propiedades, que relacionan la suma en V con la suma y el producto en k . Por ejemplo, se quiere que:

$$\begin{aligned} (a + b)v &= av + bv \\ a(v + w) &= av + aw \end{aligned}$$

para todo a, b de k y todo v, w de V . Estas ecuaciones se ven como leyes de distributividad. Las otras propiedades que deben satisfacerse son:

$$\begin{aligned} (ab)v &= a(bv) \\ 1_k v &= v \\ 0_k v &= 0_V \end{aligned}$$

para todo a, b de k y todo v de V . La primera se parece a una asociatividad, la segunda establece que el elemento neutro 1_k del campo k satisface una propiedad similar en V respecto a la multiplicación con escalares (9), mientras que la tercera relaciona los elementos neutros aditivos de k y de V .

Si tomamos el ejemplo de Fang Cheng, entonces el sistema de ecuaciones:

$$\begin{aligned}3x + 2y + 1z &= 39 \\ 2x + 3y + 1z &= 34 \\ 1x + 2y + 3z &= 26\end{aligned}$$

expresa una relación entre las tres variables x, y, z . En el siglo XIX se empezó a pensar muy diferente sobre ello. Al considerar la matriz de coeficientes:

$$M = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 2 & 3 \end{bmatrix}$$

se considera una transformación $R^3 \rightarrow R^3$, del espacio de tres dimensiones en sí mismo. Aquí, R denota al conjunto de números reales que debemos pensar en su forma geométrica como recta real, mientras R^3 denota $R \times R \times R$, que es el producto cartesiano de tres números reales.

También se formalizó la noción de *dimensión* como el mínimo número de coordenadas que requiere en un espacio vectorial para describir cualquier punto en él, en completa concordancia con nuestro sentido común para la dimensión en la geometría. Es decir, la dimensión es el mínimo número N tal que existen elementos v_1, \dots, v_N de V con la propiedad de que cualquier elemento w de V se puede escribir como una *combinación lineal* de v_1, \dots, v_N con coeficientes en el campo k , esto es, que existen a_1, \dots, a_N de k tal que:

$$w = a_1v_1 + a_2v_2 + \dots + a_nv_n.$$

Se puede demostrar que estos coeficientes están determinados de manera única. Los números a_1, \dots, a_N son las *coordenadas* de w en la *base* v_1, \dots, v_N . Por ello, basta fijar el *vector* (a_1, \dots, a_N) en k^N y lo que se obtiene es una identificación —una *biyección*, en términos matemáticos— de los elementos de V con los elementos de k^N mediante las coordenadas en la base v_1, \dots, v_N .

Con el álgebra lineal se crean fuertes herramientas para tratar todo tipo de fenómenos lineales en el álgebra y aquellos fenómenos geométricos que son lineales, es decir, que sólo tienen que ver con *líneas* y *planos* y su generalización en dimensiones mayores.

4.5.7 Algebraización

Con la teoría de conjuntos, véase sección 4.3, y las estructuras algebraicas a la mano, el siglo XX emprendió una formalización de las matemáticas —o mejor dicho— una *algebraización*. Amplias áreas de las matemáticas se revisaron y se basaron en el álgebra y se puede decir que, en algunos casos, la euforia era tal que se pecó de exageración.

Veamos aquí en qué sentido se “algebraizó” la geometría, de nuevo y una vez más, después de la geometría analítica de Descartes. Lo que surgió a partir de entonces se

llama hoy *geometría algebraica* y debe bastarnos como ejemplo. Mencionamos que existe también una topología algebraica, que un grupo de matemáticos franceses —que publicaron bajo el pseudónimo de Nicolas Bourbaki— algebraizaron el análisis y que la teoría algebraica de los números es de lo más moderno y más difícil.

En la geometría algebraica no sólo se tratan planos y rectas, sino *curvas* y *superficies*, aunque éstas se piensen dentro de algún espacio vectorial; por ejemplo, el C^n , es decir, el espacio de números complejos de dimensión n . Por consiguiente, los objetos que interesan son curvos dentro de un espacio bien conocido. Se exige que la curva o la superficie esté definida por una o varias ecuaciones algebraicas. Por ejemplo:

$$x_1^2 + x_2^2 + x_3^2 - 1 = 0$$

describe una esfera en el espacio tridimensional, con centro en el origen $(0, 0, 0)$ y radio 1, mientras que el sistema de ecuaciones: t

$$\begin{aligned} x_1^2 + x_2^2 + x_3^2 - 1 &= 0 \\ x_3 &= 0 \end{aligned}$$

describe una circunferencia con centro en el origen y radio 1, que está en el plano de coordenadas x_1, x_2 .

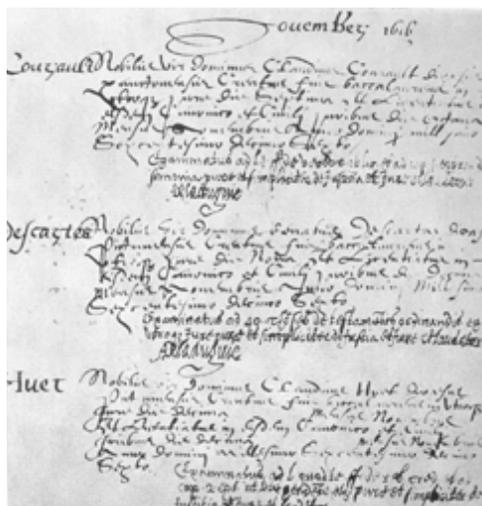


Figura 4.33 Registro de graduación de 1616 de René Descartes (1596-1650), en el Collège Royal Henry-Le-Grand, La Flèche. Descartes fue un filósofo y matemático francés, reconocido por sus grandes contribuciones a las ciencias naturales.

En lo que sigue, sea T un conjunto finito de polinomios en $C[x_1, \dots, x_N]$, o lo que es lo mismo, T es un conjunto de polinomios en las incógnitas x_1, \dots, x_N con coeficientes en el campo de los números complejos C . Con $Z(T)$ denotamos al conjunto:

$$Z(T) = \{x \in C^N \mid f(x) = 0, \text{ para cada } f \text{ en } T\},$$

es decir, el conjunto $Z(T)$ consiste en las raíces comunes de todos los polinomios en f . Se escribe $Z(f)$ si $T = \{f\}$.

Conversamente, para cada conjunto algebraico Y de C^N definimos:



$$I(Y) = \{f \in C[x_1, \dots, x_N] \mid f(y) = 0, \text{ para cada } y \text{ de } Y\}.$$

Se observa que si $T_1 \subseteq T_2$, entonces $Z(T_1) \supseteq Z(T_2)$ y si $Y_1 \subseteq Y_2$ entonces $I(Y_1) \supseteq I(Y_2)$. Esto es que las dos funciones Z, I invierten la contención de conjuntos, que se parece a la correspondencia de Galois, véase sección 4.9. Además, Y siempre es un subconjunto de $Z(I(Y))$ y se puede mostrar que $Y = Z(I(Y))$, se cumple en general.

Es fácil ver que $I(Y)$ es un ideal de $C[x_1, \dots, x_N]$ y por ello, no necesariamente se tiene $T = I(Z(T))$, ya que no cualquier subconjunto T de $C[x_1, \dots, x_N]$ es un ideal. Además, $I(Y)$ es un ideal *radical*, eso quiere decir que si f^n pertenece a $I(Y)$, entonces f pertenece a $I(Y)$.

Los subconjuntos de C^N que son de la forma $Z(T)$ para algún T se llaman *conjunto algebraico*. Se dice que un conjunto algebraico es *irreducible* si no puede escribirse como la unión de dos subconjuntos propios que son algebraicos. Por ejemplo, $Z(x_1, x_2)$ consiste en los dos ejes de coordenadas $x_1 = 0$ y $x_2 = 0$ y no es irreducible dado que $Z(T) = Z(x_1) \cup Z(x_2)$. Sin embargo, los ejemplos de la esfera y la circunferencia que dimos anteriormente, sí son irreducibles.

Un *teorema de Hilbert* afirma que hay una correspondencia entre los conjuntos algebraicos de C^N y los ideales radicales de $C[x_1, \dots, x_N]$. Además, bajo esta correspondencia, los conjuntos algebraicos irreducibles corresponden a los ideales radicales que son primos.

Esto muestra que el estudio de los objetos geométricos que están dados por ecuaciones algebraicas se puede traducir a un lenguaje algebraico. Se critica a la geometría algebraica porque ya no es geométrica, sin embargo, algebraizaciones como ésta han tenido un fuerte impacto en el avance de las matemáticas.

4.5.8 El gran proyecto de clasificación de los grupos simples

Después de su nacimiento, los grupos se hicieron abstractos, es decir, los matemáticos empezaron a estudiarlos sin pensar específicamente en grupos de Galois y en su conexión con la resolución de ecuaciones y permutaciones de raíces. De la misma manera, se empezaron a estudiar los grupos finitos, es decir, los grupos que tienen un número finito de elementos. Como los números enteros mayores que 1 se descomponen en un producto de primos, también los grupos finitos se “descomponen”, pero en *grupos simples*. Es decir, así como los primos son la pieza clave para los números, los grupos simples lo son al concepto de grupo.

Lo anterior muestra el gran interés de la comunidad matemática por conocer todos los grupos simples que hay, es decir, de obtener un resultado de *clasificación*. En 1954, en el Congreso Internacional de Matemáticas en Amsterdam, Richard Brauer anunció una idea sobre cómo podría emprenderse esta clasificación, lo cual desató una actividad frenética que tardó más de dos décadas en resolverse e involucró alrededor de cien matemáticos que publicaron sus hallazgos y avances en más de diez mil páginas.

El trabajo concluyó alrededor de 1980, cuando se obtuvo una lista completa de los grupos finitos que son simples. Esta lista consta de tres familias infinitas. Una de ellas es la

de los *grupos alternantes*: para cada entero $n \geq 5$, el grupo A_n se forma con las permutaciones que se pueden escribir como composición de un número par de transposiciones, véase sección 4.9. Además, hay 26 grupos *esporádicos*, es decir, grupos que no aparecen en familias, cuyos nombres son muy llamativos: el *Monster* es el grupo esporádico más grande y tiene:

808 017 424 794 512 875 886 459 904 961 710 757 005 754 368 000 000 000

elementos. Sin embargo, ya durante la escritura de la demostración surgieron dudas sobre la validez del resultado. La demostración de que éstos son todos los grupos simples es extremadamente difícil y, por ello, Daniel Gorenstein, Richard Lyons y Ron Solomon emprendieron una segunda etapa: el proyecto de revisión, que aún no ha concluido aunque ya se publicaron 4 de los 12 tomos planeados para la demostración. Lamentablemente, Gorenstein murió en 1992, situación que frenó el proyecto.

El teorema de clasificación de los grupos simples es un gran logro de la mente humana y completamente único en la historia. Nunca antes los científicos se habían reunido para producir una certeza en un estilo casi industrial. El logro es comparable con la Gran Pirámide de Giza, que también sólo fue posible por el esfuerzo colectivo. El resultado en sí ha sido muy exitoso: muchas veces que se quiere demostrar una propiedad para grupos finitos, en general se puede mostrar para los simples porque éstos se conocen y luego, se extiende el resultado a todos los grupos usando el teorema de Jordan-Hölder.

Sin embargo, este ejemplo muestra también la gran dificultad a la cual se enfrentan los matemáticos hoy día: la complejidad de los argumentos puede volverse tan grande que es casi imposible avanzar aún más en esta dirección. Lo anterior es un ejemplo de las limitaciones que existen para el avance científico. Aún más sorprendente es que el grupo llamado *monster* está relacionado con la teoría de cuerdas; esta teoría de la física aún no se ha concluido, pues los cálculos involucrados son extremadamente difíciles. Una vez más se puede apreciar lo que Eugene Wigner llamó “la inexplicable eficacia de las matemáticas”: es realmente sorprendente que este lenguaje abstracto sea tan útil en la descripción de las leyes que gobiernan nuestro universo.

4.6 ¿QUÉ ES LA GEOMETRÍA HOY?

Las nociones geométricas elementales residen en algo más primordial que el intelecto humano, pues los animales también las manejan. Con ellos compartimos un espacio —el *espacio físico*—, así como las dificultades, obstáculos y peligros que surgen al moverse en él. De esta manera, la contundencia del espacio que habitamos por medio de las reglas elementales —e implacables— que lo rigen es la fuente de nuestra intuición geométrica. Sin embargo, no fue tarea fácil precisar esas reglas y, eso sí, es obra del intelecto humano.

Se puede pensar en la geometría euclidiana como el primer modelo exitoso del espacio físico. Con base en unos cuantos *axiomas* o postulados se construye una teoría en la que se pueden ir demostrando ciertas afirmaciones, llamadas teoremas, por medio de razonamientos lógicos precisos. Con estos axiomas se trata de definir los elementos básicos —pero etéreos— del espacio, que son precisamente los puntos y las líneas, al indicar las relaciones que mantienen. Por ejemplo, con el axioma “por dos puntos pasa una única línea” se expresa la idea de que en las líneas hay una manera preferente para ir de un punto a otro y que, con ellas, se miden las distancias.

Figura 4.34 Salto de un tigre en pos de una pequeña presa. Aunque el tigre no mida los metros, sino la fuerza muscular que emplea para saltar, tiene que considerar distancias y velocidades; con base en ello, decide trayectorias para escapar de predadores o atrapar presas. De la eficacia de estas decisiones intuitivas depende la supervivencia: en sus cálculos, el tigre también tomó en cuenta el inminente aterrizaje.



Además de describir con unos cuantos axiomas el espacio que habitamos, Euclides también describe y sienta las bases para estudiar algo mucho más abstracto: el *plano euclidiano* que tiene sólo dos dimensiones y sus símiles en el mundo real son las superficies lisas como paredes, pisos, papeles y pizarrones. A diferencia de ellos, Euclides se extiende indefinidamente en algo que parece más fácil de intuir para el espacio: nosotros nos sentimos inmersos en este espacio y sabemos por experiencia que nunca veremos sus “límites”, pero a los “planos reales” los vemos desde fuera y siempre se acaban. El plano euclidiano es el primer ejemplo de lo que hoy los matemáticos llamamos *espacios*. Sí, en plural porque hay muchos. Sabemos bien que, en el sentido físico no existe el plano euclidiano, que es algo abstracto de lo cual podemos decir muchas cosas y que al estudiarlo, obtenemos herramientas para modelar y controlar al espacio físico. Pero también plantea problemas intrínsecos, que aunque parezcan alejados de la realidad, a la vuelta de la historia nos ayudan a entenderla.

4.6.1 El quinto postulado

De los postulados con que Euclides define el plano, el quinto se hizo famoso pues parece no ser tan elemental como los demás y hasta se sospechaba que era teorema, más que axioma. Por más de dos milenios los matemáticos intentaron demostrarlo usando sólo a los otros cuatro, pero no pudieron. Y no pudieron porque no se puede aunque esto quedó claro y bien establecido sólo hasta el siglo XIX.

Hay varias maneras equivalentes de enunciar al quinto postulado. La más usual es el *axioma de las paralelas*: dada una línea ℓ y un punto P fuera de ella, se puede trazar una única línea que pasa por P y es paralela a ℓ .

Existen dos formas de negar este axioma. La primera es que no existe la paralela y la segunda, que por el punto P no pasa sólo una sino muchas paralelas a ℓ . En la primera negación se peca de escasez y en la segunda, por exceso. Ambas suposiciones, al establecerlas como quinto axioma, son válidas y dan lugar a las *geometrías no euclidianas*. La primera negación del quinto postulado equivale a que no existe el paralelismo: cualquier par de rectas se intersecan como sucede en el *plano elíptico*, íntimamente relacionado con la geometría proyectiva. Por otra parte, en el *plano hiperbólico* se cumple que por un punto pasa más de una línea paralela a otra lejana.

Hacia principios del siglo XIX, varios matemáticos, entre los que destacan János Bolyai y Nikolai Ivanovich Lobachevsky, se convencieron de que existía el plano hiperbólico en el sentido matemático —pues aunque fuera algo abstracto, estaba muy bien definido por un sistema de axiomas—. Entonces, era cierto que se podían hacer razonamientos y demostrar teoremas como antes lo hizo Euclides y que estos resultados se iban aglomerando con una nueva lógica y elegancia interna. Por ejemplo, se podía demostrar que, en el plano hiperbólico, la suma de los ángulos internos de un triángulo siempre era menor que π —o 180 grados— pero, además, que lo que le faltaba para π era justamente su área. Entre más chica el área de un triángulo, más se parece a uno euclidiano; pero por el otro extremo, los triángulos hiperbólicos siempre tienen un área menor a π ; ¡algo sorprendente y bonito!

Estos últimos resultados se los achacan algunos historiadores a Gauss, pero no quiso publicarlos por temor al descrédito. De hecho, el trabajo de Bolyai y Lobachevsky —que sí publicaron— fue atacado y despreciado por muchos como algo insensato y sin sentido ni fundamentos en nuestra “realidad”. Sin embargo, hacia finales del siglo XIX se descubrieron *modelos* del plano hiperbólico dentro de la propia geometría euclidiana y, entonces, no hubo más remedio que aceptar su existencia. El más simple de estos modelos es el de Klein, que consiste en el interior de un disco —sin su frontera— y las líneas son los segmentos o cuerdas de este disco (figura 4.35).

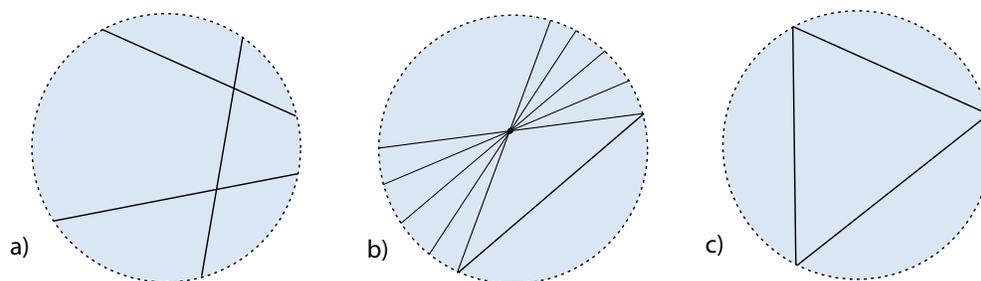


Figura 4.35 a) El modelo de Klein del plano hiperbólico. b) Muchas paralelas a una línea dada, pasan por un punto fuera de ella. c) Un triángulo ideal de área π ; se le llama así pues sus vértices no están en el plano hiperbólico.

En el modelo de Klein del plano hiperbólico, las distancias y los ángulos no corresponden con los euclidianos. Están dados por fórmulas complejas que expresan con precisión cómo dos puntos que nos parecen cercanos a la frontera, hiperbólicamente están muy lejos. Así que una barra rígida —hiperbólica— se hace chica —euclidianamente— conforme se le acerca al borde, como se verá en los ejemplos de la figura 4.40.

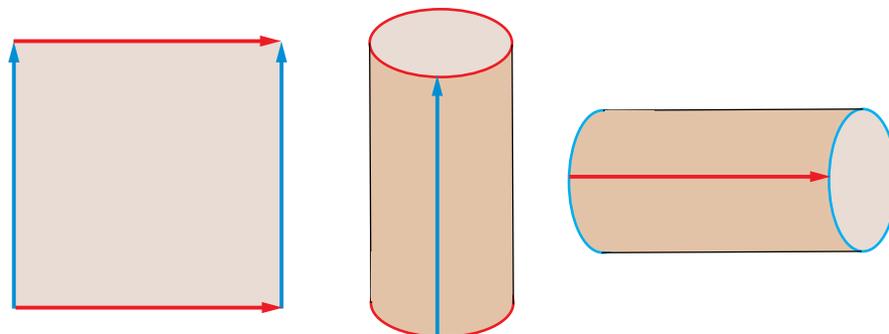
Con la aceptación de estas nuevas geometrías, la matemática se desprende de la realidad inmediata como único objeto de estudio y aclara el papel de la ciencia como proveedor de modelos abstractos para entenderla. De hecho, esta nueva libertad creativa proporciona herramientas para cuestionar nuestra concepción de la realidad, pues Gauss, como buen científico, quiso medir los ángulos de triángulos muy grandes para ver si realmente sumaban π , o lo que es lo mismo, si a gran escala éramos euclidianos o hiperbólicos. Sin embargo, el control en los errores de medición no le permitieron decidir y esa duda sigue en pie.

4.6.2 El toro plano y el plano elíptico

Para describir un modelo del plano elíptico, nos conviene considerar primero un espacio en el que muchos han jugado. Entre los primeros juegos de computadora, hay uno en el que una navecita en la pantalla se mueve cambiando su dirección con las flechas y acelerando —por propulsión a chorro— con la barra espaciadora o alguna otra tecla. Con

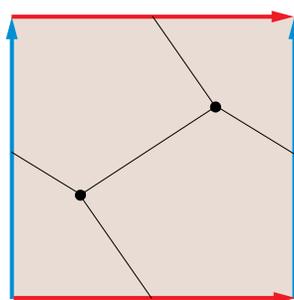
un impulso, la nave viaja inercialmente en línea recta pero, para que no choque con los bordes, el programa la hace aparecer del lado opuesto y viajando en la misma dirección. Así que si vamos a salir por arriba, aparecemos por abajo y si salimos por la derecha reaparecemos por la izquierda. El espacio donde se mueve la nave, y las rocas que hay que destruir a balazos, se llama el *toro plano* y se muestra en la figura 4.36.

Figura 4.36 Para obtener el toro plano se identifican las aristas opuestas de un cuadrado, de manera que su borde o frontera aparente ya no es tal. En el espacio físico y con el cuadrado de papel, se pueden identificar dos de las aristas para formar un cilindro, pero no las cuatro sin arrugar el papel o cambiarle su geometría.



No es difícil ver que el toro plano es un espacio de dos dimensiones donde cada punto tiene una pequeña vecindad igual que si estuviera en el plano euclidiano; no importa dónde ande la navecita, “siente” que está en un plano. Este espacio también presenta ciertas nociones geométricas como distancias, ángulos y trayectorias inerciales o líneas, pero no cumple todos los axiomas de una geometría. Por ejemplo, no es cierto que haya una única línea entre dos puntos, como se muestra en la figura 4.37. Además, no se le puede sumergir en el espacio euclidiano de tres dimensiones sin deformar algunas de sus propiedades geométricas, de lo que hablaremos más adelante. Por el momento, nos interesa para usar la posible familiaridad personal con él como símil de nuestra siguiente definición.

Figura 4.37: En el toro plano, hay varias posibilidades para unir dos puntos por un segmento.



El *plano elíptico* se obtiene a partir de media esfera, al identificar los puntos opuestos en su borde o frontera. Para fijar ideas, pensemos que la media esfera es el hemisferio norte, es decir, que el borde es el ecuador y está horizontal. Si una navecita viaja en el plano elíptico, al llegar al borde va de bajada pero reaparece por la posición diametralmente opuesta y de subida. También podemos hacer que la nave viaje inercialmente a lo largo del ecuador y, en este caso, sólo la veríamos a la mitad, mientras que la otra mitad estaría justo del lado opuesto y viajando —en la misma dirección y con la misma velocidad—; después de “media” vuelta, la navecita estaría en el mismo lugar pero sus lados se habrían intercambiado.

Las líneas del plano elíptico son los círculos máximos de la esfera o, mejor dicho, sus mitades; son las trayectorias inerciales que seguiría la nave. Parecen salir y acabar en puntos opuestos del ecuador, pero éstos son en realidad el mismo punto. Ahora sí, tenemos una geometría “a la Euclides” con todas las de la ley, donde en cada par de puntos pasa una única línea y, al seguirla, se obtiene la trayectoria más corta entre ellos. A diferencia de los planos euclidiano e hiperbólico, las líneas en el plano elíptico no son infinitas sino

circulares, en el sentido de que al viajar en ellas se regresa eventualmente al mismo lugar. Y se cumple que no hay paralelismo: cualquier par de rectas se interseca en un único punto.

Parecería que los puntos del ecuador —y que éste como línea—, son diferentes a los demás. Pero no es así: para descubrirlo, imaginemos que la media esfera es parte de una esfera completa de la que sólo vemos el hemisferio superior porque está incrustada en el plano opaco del ecuador —algo similar a las bolitas que, a veces, se usan como ratón—. Si la esfera gira libre y lentamente, los puntos que van desapareciendo debajo del plano son remplazados —y ahora representados— por sus correspondientes puntos antípodas que emergen por abajo y en el lado opuesto del ecuador, conforme a la regla de pegado de este espacio. En dicho caso, vemos al ecuador como un medio círculo cualquiera y decimos que el plano elíptico es *homogéneo*, pues podemos moverlo rígidamente para llevar un punto hasta cualquier otro al girar la esfera; lo anterior quiere decir que todos los puntos son iguales y tienen pequeñas vecindades equivalentes a las de la esfera o, en otras palabras, que el espacio es localmente esférico y su geometría se “hereda” de la esfera.

Al girar un poco la esfera, observamos que sale un “gajo” o *sector angular*, cuya área es proporcional al ángulo de giro. Si suponemos que el radio de la esfera es 1, entonces su área total es 4π y, por lo tanto, el área del plano elíptico es la mitad o 2π . Así, el área de un sector angular de ángulo α es 2α .

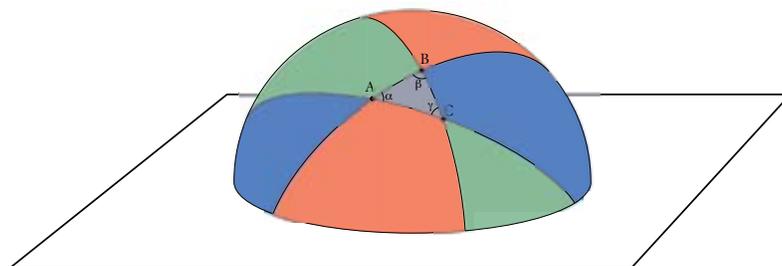
Consideremos un triángulo en el plano elíptico con ángulos internos α , β y γ ; y sea A su área. Si pintamos los tres sectores angulares con pintura, se pinta todo; pero el triángulo queda pintado con tres manos, mientras que el resto sólo con una. Al contabilizar la pintura total que se usó, se obtiene:

$$2\alpha + 2\beta + 2\gamma = 2\pi + 2A ,$$

es decir, que:

$$\alpha + \beta + \gamma = \pi + A .$$

Como el área A es positiva, ello implica que la suma de los ángulos de un triángulo en el plano elíptico es mayor que π , además de que el exceso sobre π , es justo su área. Lo mismo vale para triángulos en la esfera.



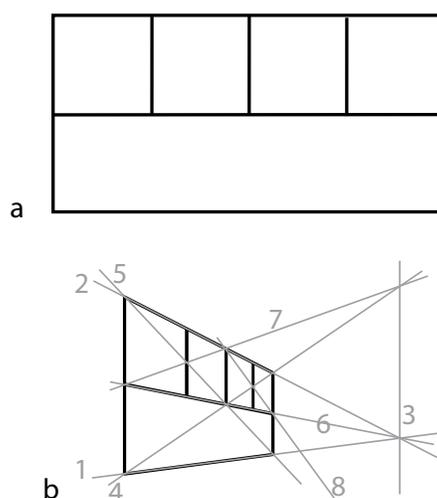
4.6.3 El plano proyectivo

En el Renacimiento, los artistas plásticos acabaron de establecer los principios técnicos de la *perspectiva*. La idea básica es que un cuadro representa mejor la realidad si se le concibe como una ventana —a través de la cual se ve el paisaje— en la que se plasma lo que el pintor ve. Geométricamente, lo anterior equivale a *proyectar* al plano del lienzo, lo visible —representado con su color— desde un punto fijo: el ojo del pintor que será

posteriormente el del observador. Entonces resulta que las escalas de las cosas se hacen más chicas conforme están más lejos, y que líneas rectas que en la realidad son paralelas, deben dibujarse como líneas que concurren a un punto llamado su *punto de fuga*.

Las líneas reales se proyectan en líneas al lienzo, pues son la intersección del lienzo con el plano que une la línea real con el ojo. Resulta que los puntos de fuga de todos los haces paralelos en un plano forman una línea: su línea al infinito; por ejemplo, la *línea al infinito* del piso es el *horizonte* en el lienzo, al cual se fugan todas las líneas horizontales. Supongamos que queremos dibujar en perspectiva un edificio que se encuentra enfrente y a nuestra izquierda. El plano paralelo a su fachada que pasa por nuestro ojo, intersecta al lienzo en su línea al infinito: cualquier cosa en el plano de la fachada se verá a la izquierda de esta línea, como se aprecia en la figura 4.38.b.

Figura 4.38 a) Una fachada sencilla en un rectángulo de 2×4 . b) Si se dibuja a ojo el contorno del rectángulo —en este caso, con las verticales aún verticales—, la fachada se reconstruye con trazos de líneas auxiliares en el orden de la numeración. Si se alargan las líneas 5 y 8, concurren en un punto de la línea 3 —que es la vertical en la intersección de 1 y 2— y corresponde a la línea al infinito en el plano de (a), donde 5 y 8 son paralelas.



Al proyectar planos a planos desde un punto, resulta natural y teóricamente indispensable considerar que sus líneas al infinito constan de un punto de fuga por cada haz paralelo. Al plano euclidiano con un nuevo punto para cada dirección, en el que concurren las paralelas correspondientes y forman una nueva línea —al infinito— se le llama *plano proyectivo*; tiene un grupo de transformaciones asociado, las *transformaciones proyectivas*, que son las que preservan líneas. Un ejemplo es el de la figura 4.38, que manda la fachada en (a) hacia su perspectiva en (b). En este caso, basados en que las líneas van en líneas, sólo se hicieron los trazos necesarios para reconstruir la fachada, pero la transformación está definida en todo el plano, es decir, se puede decidir cuál debe ser la imagen de cualquier punto. En términos de proyecciones y volviendo a pensar en el paradigma del pintor, en la figura 4.39, la fachada es parte de un plano (a) y el lienzo es parte de otro plano (b); al pintor sólo le interesa lo que está enfrente pero, matemáticamente, se considera todo: lo que está atrás del pintor se proyecta del otro lado de la línea al infinito, de (a) vista en (b).

El plano elíptico y el plano proyectivo están en correspondencia natural. Para entender lo anterior, se considera al plano tangente a la esfera en su polo norte y, después, se proyecta desde el centro de la esfera. Las líneas se corresponden pues en ambos casos son intersecciones de planos por el centro de proyección. Los puntos en el ecuador de la esfera van hacia el punto correspondiente en la dirección de la línea al infinito del plano proyectivo.

De hecho, a estos dos planos se les piensa como el mismo espacio y lo que los distingue es el grupo de transformaciones que se consideran: en el plano elíptico se juega sólo con las transformaciones rígidas y, entonces, se puede hablar de distancias, ángulos

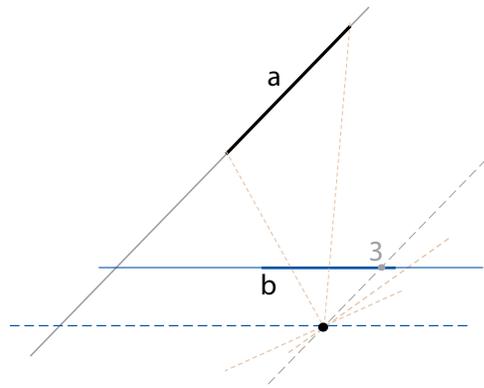


Figura 4.39 Vista superior de la transformación proyectiva de la figura 4.38 considerada como proyección de un edificio real (a) en el lienzo (b). La línea al infinito de (a) se proyecta en la línea 3 en (b), que aquí se ve como un punto.

y áreas; en el proyectivo se usan todas aquellas transformaciones proyectivas o *colineaciones*, que de hecho, son muchas.

Aunque aquí hemos hablado de la geometría proyectiva en forma sintética, también se puede trabajar analíticamente —es decir, con coordenadas— y ésta, es hoy día una de las herramientas básicas para el despliegue de la realidad virtual en las computadoras.

4.6.4 Los espacios multidimensionales

Se considera a Girard Desargues (1591-1661) como el padre de la geometría proyectiva, pues sentó sus bases y demostró los primeros teoremas. Sin embargo, éstos no fueron retomados por otros matemáticos durante los dos siglos siguientes, quizá porque su contemporáneo, René Descartes (1596-1650), atrajo los reflectores de la historia al “coordenatizar” el plano euclidiano. Descartes estableció la correspondencia entre parejas ordenadas de números reales y puntos en el plano euclidiano; con ello, nace la *geometría analítica* y la posibilidad de usar nuevos métodos en la geometría. Más aún, también se abre la puerta para trabajar en otras dimensiones.

Si se describe el plano con parejas de números y el espacio con ternas, al considerar todas las cuartetos de números reales, tendremos el espacio de dimensión cuatro. Estos espacios se denotan R^2 , R^3 y R^4 respectivamente. No hay que pararse en cuatro: para cualquier número n se puede definir el *espacio euclidiano de n dimensiones* como el conjunto de n -adas ordenadas de números reales:

$$R^n = \{(x_1, x_2, \dots, x_n) \mid x_i \in R\} .$$

Aunque no se pueda ver o se dude de su existencia en el sentido físico, matemáticamente está ahí, para ser explorado. Por ejemplo, se pueden definir líneas, segmentos y distancias de manera que se extiende naturalmente a los conceptos euclidianos para dimensiones 2 y 3.

A mediados del siglo XIX, Bernhard Riemann amplía, aún más, la noción de “espacio”. Ciertos subconjuntos de R^N son especiales, pues localmente se parecen a o se pueden modelar como R^n , para alguna $n < N$; se les llama ahora variedades de dimensión n . Por ejemplo, las superficies lisas — $n = 2$ — en el espacio — $N = 3$ —, son *variedades de dimensión 2*; o bien, los puntos que equidistan de uno dado en R^N son las esferas de dimensión $N - 1$. Y, por vivir en R^N , heredan una geometría explícita.

Se puede medir la distancia entre dos puntos como la longitud más corta de las trayectorias que los unen dentro de la variedad; y a estas trayectorias más cortas o eficientes para viajar se les llama *geodésicas*. Por ejemplo, si tomamos una esfera en R^3 o en R^N , sus

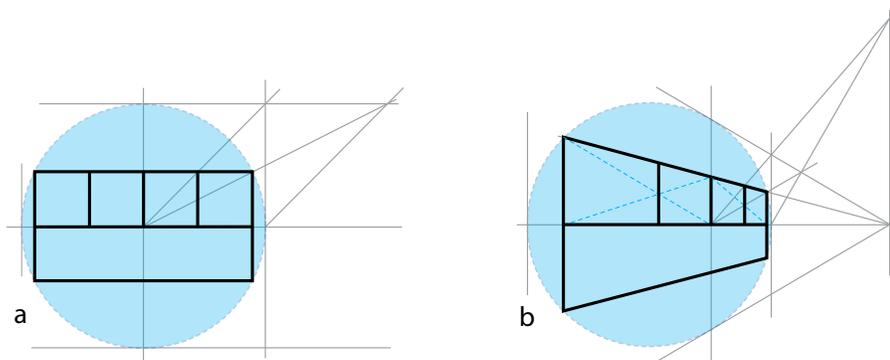
geodésicas son los círculos máximos o intersecciones de planos que pasan por su centro. Riemann observa que ahí hay muchísimos espacios donde se puede hacer geometría, y establece las bases y herramientas para hacerla: lo que ahora llamamos *geometría riemanniana*. Cuando a principios del siglo xx, Albert Einstein le demuestra al mundo que el universo debe concebirse como algo de dimensión cuatro y, además, no euclidiano sino con *curvatura* producida por la masa, tenía ya la herramienta matemática para hablar de ello.

Otro personaje importante en la geometría del siglo xix fue Sophus Lie. Estudió los grupos de transformaciones de los espacios euclidianos multidimensionales y observó que son variedades de grupos continuos llamados *grupos de Lie*. El ejemplo más sencillo es el círculo, pensado como las rotaciones del plano alrededor del origen y, a su vez, como variedad de dimensión 1. Los grupos de Lie han resultado ser una herramienta indispensable para la física del siglo xx, sobre todo para la física cuántica.

Hacia el final del siglo xix, cuando la geometría se convulsionó y expandió en múltiples direcciones, Felix Klein trató de resumir la definición de geometría diciendo que “es el estudio de los invariantes de un espacio bajo un grupo escogido de sus transformaciones”. Pongamos un ejemplo de esta idea, retomando el modelo del plano hiperbólico del propio Klein. El espacio dado es el interior del disco —como se muestra en la figura 4.35—; si lo pensamos dentro del plano proyectivo —que contiene al euclidiano—, podemos considerar todas las transformaciones proyectivas que lo dejan en su lugar: éste es el grupo de *transformaciones hiperbólicas*. Entonces, resulta que cualquier punto del interior se puede mover a cualquier otro; que el espacio es homogéneo y que los “invariantes”, como distancia y ángulo, se pueden construir a partir del grupo.

Para ver un ejemplo de una transformación hiperbólica, además de las obvias, las rotaciones, supongamos que la fachada de la figura 4.38.a) está inscrita en el círculo del modelo de Klein y queremos trasladarla —como figura hiperbólica— en su línea media horizontal. Basados en que las líneas tangentes al círculo deben ir a líneas tangentes al mismo, pues éste se queda, como conjunto, en su lugar —ésta es la definición de transformación hiperbólica—; una vez fijado el punto donde va el centro de la fachada, con unos cuantos trazos más, se obtiene su traslación hiperbólica, como se observa en la figura 4.40.

Figura 4.40 Una traslación en la línea hiperbólica que corresponde al diámetro horizontal del disco, deja a las dos tangentes verticales al círculo en su lugar, y por lo tanto manda líneas verticales en líneas verticales.



4.6.5 Topología

En el cambio de siglo del xix al xx, ya se tenía claro que había una geometría aún más libre, donde los espacios se pudieran deformar sin romper su “continuidad” y se sentaron las bases formales para hacerlo: la *topología*. La intuición básica en esta área de las matemáticas es dejar de lado la noción rígida de distancia pero mantener la de vecindad de los puntos, cambiar la cercanía estricta y rígida por una más laxa y flexible.

Veamos el ejemplo del toro plano. Lo teníamos como un cilindro de papel en R^3 —fi-

gura 4.36—, donde había que identificar sus dos extremos circulares —esto es, unirlos punto a punto—; no nos atrevimos a hacer tal cosa, pues el papel o la pantalla de la computadora regían su geometría, importante en ese momento. Pero si lo imaginamos de un material elástico —idealmente flexible— como espacio topológico, podemos doblarlo y deformarlo, poco a poco, hasta identificar las dos bocas y obtener la superficie de una dona o una llanta, en la cual todas las identificaciones prescritas ya están hechas, según muestra la figura 4.41.

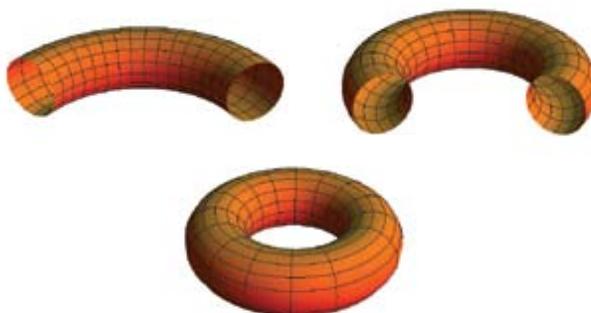


Figura 4.41 Al identificar las dos bocas de un cilindro, se obtiene el toro.

Se le llama el *toro*, sin apellidos. El toro plano es un toro topológico con una estructura extra: la variedad riemanniana que le da rigidez. La topología es algo más elemental o general, pero llegar a ella es, técnicamente, más complicado. Por eso, históricamente se dio después. Otro ejemplo indicativo es un cuadrado y un disco en el plano, donde es fácil concebir una deformación gradual de uno en el otro: aunque geoméricamente sean distintos, topológicamente son equivalentes.

A la topología, que estudia los espacios en esta generalidad plástica, se le puede considerar como “la geometría del siglo xx”, aunque tal nombre es injusto con la gran cantidad de áreas especializadas en las que se ha desarrollado la geometría.

Uno de los problemas que rigió el desarrollo de la topología en el siglo xx fue planteado por Henry Poincaré en 1905 y se le conoce como la *conjetura de Poincaré*. Ésta tiene que ver con una caracterización elegante de la esfera de dimensión 3, entre todas las variedades de dimensión 3 —o 3-variedades—. En la década de los ochenta, William Thurston demostró que la conjetura de Poincaré sería consecuencia de otra conjetura aún más osada: la de *geometrización*, que dice, a grandes rasgos, que cualquier 3-variedad topológica se puede partir en pedazos, cada uno modelado sobre una de 8 posibles “geometrías” tridimensionales —un ejemplo en dimensión 2 es el toro plano, que está modelado sobre el plano euclidiano—.

En 2002, Grigori Perelman demostró la conjetura de geometrización, aunque todavía son muy pocos los matemáticos que entienden dicha demostración a cabalidad aunque la idea base sea comprensible. Una variedad riemanniana tiene una manera natural de homogeneizarse o de “acomodarse” y si la dejamos fluir —en su *flujo de Ricci*— tenderá a adquirir una estructura geométrica como la que previó Thurston. Por ejemplo, si empezamos con el toro de la figura 4.41, con la estructura riemanniana o geometría que se hereda de vivir en R^3 , los círculos horizontales del centro “querrían” crecer y los de afuera, decrecer para parecerse entre ellos, pero los verticales se quedarían rígidos porque ya son iguales. Hacer todo esto a la vez en R^3 es imposible; sin embargo, en abstracto o considerando una dimensión más, en R^4 , este toro tendería naturalmente a convertirse plácidamente en el toro plano.

En fin, los matemáticos ya tienen un catálogo de posibles formas que puede tener un universo tridimensional. Toca a los cosmólogos, como alguna vez lo intentó Gauss, decidir cuál se aproxima más a la “real”.

4.7 ¿CÓMO FUNDAMENTAR LAS MATEMÁTICAS?



Figura 4.42 El pensador, realizada en bronce en 1880 por Auguste Rodin, es una de las esculturas más famosas. A pesar de que en su origen Rodin buscaba representar a Dante frente a las puertas del infierno, se considera que simboliza al hombre —en sobria meditación— mientras se debate ante un poderoso dilema interno.

“Reparte 10 sacos de cebada entre 10 hombres de modo que la diferencia entre la parte de cada hombre y la de sus vecinos sea $\frac{1}{8}$ de saco.” Este problema es uno más que se plantea en el papiro de Rhind.

Alrededor del año 1700 a.C., Ahmés —un escribano egipcio— elaboró un texto en el que se recopilaban ochenta problemas matemáticos, todos ellos relacionados con situaciones concretas de la vida cotidiana, como la medición de la Tierra o la repartición de bienes. Este papiro fue encontrado en 1858 en Luxor, Egipto, por un anticuario escocés que se llamaba Henry Rhind y en 1863, luego de ser entregado al Museo Británico, se le llamó el *papiro de Rhind*.

Hace miles de años, en las civilizaciones babilonia y egipcia, las matemáticas estaban orientadas, fundamentalmente, a resolver problemas prácticos. No fue sino hasta varios siglos más tarde cuando los matemáticos griegos dieron a las matemáticas un carácter completamente distinto. Para ellos, no bastaba que el conocimiento matemático resolviera problemas concretos que se comprobaran a través de la experiencia, sino que buscaban la generalización de los resultados y su deducción, a partir de otros. Por consiguiente, organizaron las matemáticas como un sistema deductivo donde el único método válido para obtener resultados era el método axiomático.

El ejemplo más importante en las matemáticas griegas es la obra *Los elementos* de Euclides, quien vivió en Alejandría hacia el año 300 a.C. En este libro —compuesto por trece capítulos— Euclides recopiló y organizó en forma deductiva gran parte de la geometría y aritmética conocida en su época.

Los griegos hicieron de las matemáticas una ciencia teórica y deductiva en la que, para justificar una afirmación, no es suficiente probarla al comprobar que se cumple en casos particulares —aunque sean muchísimos—, sino que es necesario establecerla como un teorema.

Un teorema es una afirmación que se demuestra por medio de la argumentación lógica y de forma deductiva a partir de otros teoremas que, a la vez, ya han sido demostrados. Como las demostraciones no pueden ser indefinidas, se parte de ciertos principios tan claros y evidentes que se pueden asumir sin necesidad de demostrarlos: las definiciones, los postulados y los axiomas. El método axiomático consiste, a grandes rasgos, en lo anterior y con ello, los griegos dieron un gran paso hacia la fundamentación de las matemáticas.

Un filósofo griego que no podemos dejar de mencionar aquí es Aristóteles, que nació en el 384 a.C. en Macedonia y que propuso el razonamiento deductivo a partir de silogismos como el que se muestra a continuación:

- Todos los hombres son mortales,
- Sócrates es hombre,

por lo tanto,

- Sócrates es mortal.

El nombre de Aristóteles estará siempre ligado al de su maestro, Platón, quien a la vez era discípulo de Sócrates y estudioso y admirador de Pitágoras. Es decir, los nombres de los filósofos, matemáticos y físicos griegos están entrelazados en la búsqueda de una estructura que le diera rigor a las formas de pensamiento matemático.

Figura 4.44Aristóteles.

Debemos a Euclides el método axiomático, lo que constituye un gran paso en la historia de las matemáticas, pues introduce la manera moderna o actual de hacerlas.

En un principio, el método axiomático era intuitivo, es decir, los axiomas y los postulados se establecían como evidentes —y por lo tanto no era necesario demostrarlos— porque así se veían en la realidad y las deducciones que se hacían a partir de ellos eran, también, bastante intuitivas.

Poco a poco, y con el paso de los siglos, los sistemas axiomáticos se fueron haciendo más abstractos y, entonces, lo importante no era ya que los axiomas y los teoremas se adecuaran a la realidad sino que formaran *un cuerpo de afirmaciones consistente*: libre de contradicciones. En otras palabras, que mediante los axiomas no pudieran demostrarse a la vez una afirmación y su negación, por ejemplo: “a es igual a b” y, al mismo tiempo, que también se demostrara que “a no es igual a b”.

¿Cómo saber que un sistema de axiomas no lleva a contradicciones, es decir, cómo saber que, a partir de los axiomas y de las reglas lógicas que usamos para hacer deducciones, no se pueden deducir un teorema y su negación? Aunque a primera vista pueda parecer increíble, esta pregunta ha ocupado a varios filósofos y matemáticos a lo largo de los siglos.

La sistematización del pensamiento matemático fue dando lugar a preguntas muy importantes —complejas y entrelazadas entre sí— que hoy día siguen vigentes, acerca de los fundamentos de las matemáticas:

- ¿Las matemáticas se crean o se descubren?
- ¿Cómo se define el concepto de verdad en matemáticas
- ¿Cómo fundamentar correctamente un edificio intelectual tan sofisticado como las matemáticas?

Estos cuestionamientos dieron lugar a un gran debate durante la segunda mitad del siglo XIX y primera del XX.

Para empezar esta historia, explicaremos las dos posturas que existen sobre la esencia de los objetos matemáticos: aquella que sostiene que se construyen a través de procedimientos intuitivos y aquella que, por el contrario, afirma que los objetos matemáticos se conciben o se entienden como parte de una totalidad existente al margen de quien los estudia, es decir, el matemático. Esta segunda postura se conoce como platonismo matemático y ha sido muy relevante en el desarrollo de las matemáticas modernas durante el siglo XX. El afirmar que los objetos matemáticos gozan de una existencia real —análoga en algún sentido a la existencia de los objetos físicos— permite concebir un “universo matemático” independiente de las técnicas o los procedimientos que se usen para estudiarlo.

Son muchos los intentos a lo largo de la historia por mecanizar el razonamiento matemático y probar que estaba libre de contradicciones, es decir, por exhibir fundamentos sólidos sobre los cuales podía erigirse.

El trabajo más importante en este campo comenzó a finales del siglo XIX, cuando los objetos matemáticos empezaron a definirse en términos de conjuntos. Por consiguiente, se intentó definir los números, las funciones y los distintos espacios en términos de conjuntos, con el fin de tener un sistema de reglas para deducir los teoremas a partir de conjuntos pequeños de axiomas y lograr, así, un sistema matemático riguroso.

La teoría de conjuntos se desarrolló con el trabajo de Cantor —un matemático alemán que vivió de 1845 a 1918— aunque varios matemáticos ya habían trabajado en la orientación “conjuntista” de las matemáticas. La teoría de conjuntos tiene sus raíces en el análisis real, en la teoría de series trigonométricas y, en particular, en la representación de funciones discontinuas a través de series de Fourier y en la caracterización de los conjuntos de puntos de discontinuidad.

Cuando Cantor descubrió, en 1873, que el conjunto de los números reales no es numerable, es decir, que no puede ponerse en correspondencia uno a uno con el conjunto de los números naturales, la concepción que hasta entonces se tenía del infinito, cambió radicalmente. Surgieron muchos conceptos nuevos a partir del trabajo de Cantor, como el de conjunto bien ordenado, número ordinal, número cardinal y número transfinito. Cada vez era más fácil caracterizar a los objetos matemáticos en términos de conjuntos.

Pero con la teoría de conjuntos llegaron también las paradojas. A mediados del siglo XIX, ya era bastante grande el grupo de matemáticos y filósofos muy importantes que sostenían una discusión sobre los fundamentos de las matemáticas. El objetivo era eliminar

las paradojas y discutir qué era o no aceptable en las matemáticas. Por ejemplo, el *axioma de elección* o la *hipótesis del continuo* tenían que ser discutidos.

Las paradojas nacían de asumir que cualquier colección de objetos —tangibles o pensables— que satisficiera una propiedad era un conjunto. Y así había una correspondencia entre *conjuntos* y *propiedades*, pero Bertrand Russell —filósofo y matemático galés— mostró que al menos había una propiedad, $X \notin X$, que no determina ningún conjunto si no se quiere caer en una tremenda contradicción.

Una manera de replantear este problema es con la paradoja que ya habíamos revisado anteriormente: “el barbero de un pueblo afeita a todas las personas del pueblo que no se afeitan a sí mismas y sólo a esas”, ¿debe el barbero afeitarse a sí mismo?

Sea A el conjunto de todas las personas que no se afeitan a sí mismas, entonces podemos hacer la siguiente argumentación: el barbero pertenece a A si y sólo si no se afeita a sí mismo, si y sólo si lo afeita el barbero, si y sólo si se afeita a sí mismo, si y sólo si no pertenece al conjunto A . Entonces el barbero es elemento del conjunto A si y sólo si no es elemento del conjunto A .

Pasa lo mismo al definir el conjunto de todos los conjuntos que no se pertenecen a sí mismos y preguntarse, si dicho conjunto pertenece a sí mismo. Con esta paradoja, Russell planteó la imposibilidad de asociar un conjunto a cada propiedad.

Para resolver la paradoja, Russell desarrolló la *teoría de tipos* en la que la idea básica es establecer tipos o clases que, a la vez, contienen tipos o clases de una jerarquía inferior.

La necesidad de resolver las paradojas tuvo como consecuencia determinar que el camino más sólido para desarrollar las matemáticas era establecer una **axiomática** para la teoría de conjuntos, similar a la que se tenía para la geometría.

Hubo varias propuestas: la de Zermelo-Fraenkel y la de von Neumann-Bernays-Gödel. Por otra parte, el matemático alemán David Hilbert (1862-1943), que en 1899 había publicado una *axiomática* para la geometría euclidiana —más completa que la que hiciera el propio Euclides—, desarrolló los conceptos necesarios para el estudio de las propiedades formales de las teorías axiomáticas en lo que él llamó *teoría de prueba*.

La idea de Hilbert era genial pues se trataba de conseguir una formalización completa de las matemáticas, en particular de la teoría de conjuntos o de la aritmética, de manera que al unir la lógica con una interpretación formal de una teoría matemática concreta, se pudiera formalizar cualquier afirmación de dicha teoría. Esto quiere decir que cualquier afirmación podría formalizarse como una sucesión finita de fórmulas abstractas cuyos símbolos, aislados y sin interpretación, carecerían de significado por sí mismos, evitando así la aparición de paradojas.

Hasta antes del siglo XIX, la intuición era la forma de saber si un sistema axiomático pudiera llevar a contradicciones o no. Euclides formuló sus axiomas para la geometría tomando como modelo la realidad física. Pero, conforme los sistemas axiomáticos se fueron haciendo más abstractos, la intuición y la experiencia dejaron de ser la forma adecuada de verificar si dichos sistemas eran consistentes o no.

Así que una de las cosas más importantes al trabajar con un sistema axiomático es determinar si es consistente o no. Y esta pregunta se hace y se fue haciendo extensiva a todas las matemáticas, a lo largo de la historia: ¿están las matemáticas construidas sobre bases y fundamentos sólidos? ¿Cómo saber que las matemáticas no contienen ninguna contradicción en su seno? Se trataba, a fin de cuentas, de garantizar que cualquier afirmación verdadera pudiera ser demostrada dentro del sistema.

Figura 4.46
Giuseppe Peano.

Un personaje muy importante en la sistematización rigurosa de las matemáticas es Giuseppe Peano, matemático y filósofo italiano que vivió de 1858 a 1932. Peano hizo un análisis muy serio del proceso demostrativo de las matemáticas: estableció la formulación axiomática de la aritmética a través de un conjunto de cinco axiomas que hoy se conocen como los axiomas de Peano y que definen a los números naturales, en términos de la teoría de conjuntos. Los axiomas de Peano son:

1. El número 0 es natural.
2. Si n es un número natural, entonces $s(n)$ —el sucesor de n — también es un número natural.
3. El número 0 no es sucesor de ningún número natural.
4. Si para dos números naturales m y n sucede que $s(m) = s(n)$, es decir, sus sucesores son iguales, entonces $m = n$.
5. Si dada una propiedad p , sucede que:
 - a) el número 0 cumple la propiedad p ,
 - b) cada vez que un natural n cumple la propiedad p , ocurre que $s(n)$ también cumple la propiedad p ,
 entonces todos los números naturales cumplen la propiedad p .

Este último axioma se conoce como el principio de inducción.

Durante la segunda mitad del siglo XIX y la primera del XX, el “edificio matemático” fue adquiriendo cada vez más rigor. Fueron muchísimos los matemáticos y filósofos que participaron en la discusión sobre los fundamentos de las matemáticas, con posturas a veces antagónicas y otras veces, complementarias.

Y entre estas posturas pueden distinguirse tres que son cruciales y en las que las personas de las que hemos venido hablando pueden ubicarse: el logicismo, el intuicionismo y el formalismo.

El logicismo fue iniciado por Gottlob Frege, un filósofo y matemático alemán que vivió de 1848 a 1925, pero sus máximos exponentes fueron el galés Bertrand Russell (1872-1970) y el inglés Alfred N. Whitehead (1861-1974). El objetivo principal de los logicistas era mostrar que las matemáticas clásicas eran una parte de la lógica, así que la pregunta ¿están las matemáticas libres de contradicciones? se replanteaba como ¿está la lógica libre de contradicciones? Russell y Whitehead escribieron, entre muchas otras obras, *Principia Mathematica*, y en ella intentaron mostrar que todas las matemáticas conocidas hasta ese momento se podían derivar de un sistema axiomático que ellos proponían en sus *Principia*. Sin embargo, nunca pudieron probar que la lógica estaba libre de contradicciones.

Figura 4.47 Whitehead.

Figura 4.48 Luitzen
Egbertus Jan Brouwer.

El intuicionismo comenzó alrededor de 1908 con un matemático holandés llamado Luitzen Egbertus Jan Brouwer (1881-1966). Mientras que los logicistas jamás habían dudado de la solidez de las matemáticas, los intuicionistas cuestionaron —desde el principio— las matemáticas clásicas y propusieron reconstruirlas por completo, partiendo de cero.

El formalismo existía ya en el siglo XIX, pero, sin duda, su exponente más importante es David Hilbert. Los formalistas propusieron que era necesario formalizar todos los sistemas axiomáticos.

Es interesante ver que, tanto los logicistas como los formalistas, intentaron formalizar todas las ramas de las matemáticas por diferentes razones. Los primeros querían usar la formalización para mostrar que ese campo pertenecía a la lógica; los segundos, para probar matemáticamente que en esa rama no había contradicciones.

Russell y Whitehead construyeron un sistema lógico intentando que de él pudiese derivar toda la matemática: ése fue el sistema que expusieron en sus *Principia*. ¿Era posible deducir todas las verdades matemáticas a partir de ese sistema? ¿Cómo podían tener la certeza de que ese sistema era consistente, es decir, que no podían deducirse contradicciones a partir de él? Este último cuestionamiento era tan importante que Hilbert lo propuso a los matemáticos como una de las preguntas abiertas más importantes para ser contestada en el siglo XX.

El intuicionismo asumió como suyas las críticas que emergieron frente al carácter abstracto de las matemáticas. Con Brouwer se estructuró una visión sobre la naturaleza de las matemáticas, presente también entre los matemáticos decimonónicos como Kronecker y Baire. Los intuicionistas se colocaban en un terreno opuesto, de alguna manera, al formalismo y al logicismo, pues para ellos era indispensable recurrir a la intuición.

Mientras que los logicistas elevaban la lógica a una categoría muy alta, para los intuicionistas se trataba tan sólo de un instrumento absolutamente accesorio.

Los intuicionistas no trataban de probar la consistencia de la matemática sino de hacer “matemática verdadera”, apegada a la intuición introspectiva.

Si bien los intuicionistas argumentaban —permanentemente— que no estaban preocupados por los fundamentos de las matemáticas, siempre trabajaron por hacer estable el edificio matemático en el que se movían. Las ideas en la filosofía de las matemáticas y, en particular, en torno a los fundamentos, nunca han sido blancas o negras; todas fueron entretejiéndose.

Figura 4.49
Kronecker y Braire.

La historia no tuvo un final feliz. En 1931, Kurt Gödel mostró al mundo en su artículo “Sobre proposiciones formalmente indecidibles en *Principia Mathematica* y sistemas relacionados” que no es posible fundamentar las matemáticas a través de un sistema formal que sea consistente y completo; dicho de otra manera, ningún sistema formal que contenga a la aritmética puede probar su propia consistencia, como se verá en la sección 4.11.

Gödel evidenció una limitación fundamental del método axiomático al probar que, para cualquier conjunto consistente en axiomas que contenga a los axiomas de la aritmética, existen afirmaciones verdaderas que no pueden deducirse a partir de dicho conjunto.

En matemáticas, hay muchos enunciados que parecen evidentes para los que no se ha encontrado un contraejemplo y, sin embargo, no han podido ser demostrados. Un ejemplo es la *conjetura de Golbach*, que afirma que todo número par mayor que dos es la suma de dos números primos.

El debate sobre los fundamentos sobrepasó la necesidad de elegir alguna de las posturas enfrentadas. Legó, tanto a matemáticos como a filósofos, el conocimiento de hasta dónde podía llegarse en la formalización de teorías matemáticas, a través de la lógica, y puso de manifiesto la diferencia entre la postura *constructivista* y la concepción formalista —que considera a los objetos matemáticos como existentes—, que no estaba tan clara antes del siglo XIX. Esta definición ha sido una de las contribuciones más importantes del debate sobre los fundamentos.

Si bien el logicismo y el formalismo no lograron fundamentar totalmente las matemáticas, el desarrollo en el campo de la filosofía de las matemáticas ha tenido repercusiones muy importantes en otros ámbitos. El trabajo realizado por estos filósofos y matemáticos ha contribuido en otros campos como la lingüística, el análisis del razonamiento deductivo, la informática y la computación.

4.8 ¿QUÉ SE PUEDE MEDIR?

Siempre que medimos algo hay que indicar la *unidad* en la que se mide. Decir que el objeto “pesa 4” carece de sentido cuando no se aclara si se trata de kilogramos, gramos, toneladas, libras u onzas. Las medidas utilizadas han cambiado durante la historia y buena parte de las que usamos en la actualidad se establecieron en Francia, después de la Revolución y a principios del siglo XIX. El metro, por ejemplo, se estableció como la diez milésima parte de la distancia del Ecuador al Polo norte. Lo importante es que cualquier medida fija puede servir como unidad.

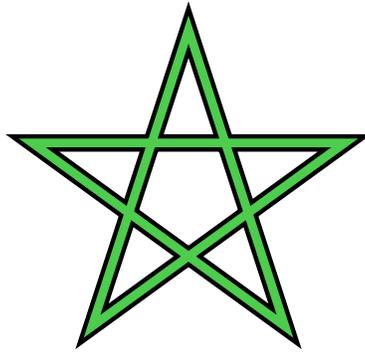


Figura 4.50 El pentagrama es una estrella de cinco puntas que se forma con cinco segmentos rectos. Fue un símbolo que representaba la perfección para los pitagóricos. Entre las diferentes longitudes aparece la razón áurea, a la que hoy día se asocian significados muy diversos según la cultura.

Si tenemos un segmento de 1.2 cm de longitud y otro de $\frac{10}{3}$ cm, podemos usar un tercero como unidad, de $\frac{1}{30}$ cm de longitud, y medir los dos primeros: el segmento de 1.2 cm de largo es 36 veces la unidad, mientras que el segmento de $\frac{10}{3}$ cm es 100 veces la unidad. Por lo tanto, encontramos una unidad de medida para expresar la longitud de ambos segmentos como un múltiplo *entero* de esta unidad. Estas ideas se atribuyen a los pitagóricos, la escuela que fundó Pitágoras en el siglo v a. C. Dos segmentos cualesquiera que se pueden medir con una misma unidad los llamaron *commensurables*. Los pitagóricos partieron tácitamente de la suposición de que cada dos segmentos son commensurables.

Para los pitagóricos, los números enteros eran centrales pues, con ellos, se expresaba la naturaleza, por ejemplo, en las relaciones de las notas con la música. Grande fue su asombro al descubrir que hay segmentos que no se pueden medir con una sola unidad; tal es el caso del lado de un cuadrado y su diagonal, que no son commensurables. La leyenda cuenta que Hippasus, el pitagórico que hizo este descubrimiento, fue echado por la borda del barco cuando comunicó su hallazgo.

Figura 4.51 Hippasus.

En la sección 1.4, expusimos una demostración de este hecho: $\sqrt{2}$ no es racional y, de ahí, sigue la incommensurabilidad del lado y la diagonal: si hubiera una unidad u para expresar la longitud del lado de un cuadrado $\ell = mu$ —donde m es un entero— y también la diagonal $\sqrt{2}\ell = nu$ —con n , un entero— entonces $\sqrt{2} = \frac{n}{m}$ es racional, lo cual es una contradicción pues sabemos que $\sqrt{2}$ no es racional.

La existencia de segmentos *incommensurables* es, tal vez, la primera ocurrencia en las matemáticas de una demostración de *algo que no existe*: no existe y nunca existirá una unidad para medir el lado y la diagonal de un cuadrado con números enteros. En este tema se tratarán varios hallazgos matemáticos y cada uno de ellos afirmará que algo no existe de manera determinante. A la vez, cada uno tiene que ver con una pregunta importante y una teoría completa que hubo que desarrollar para poder llegar a la conclusión de la no-existencia.

4.9 ¿QUÉ SE PUEDE RESOLVER?

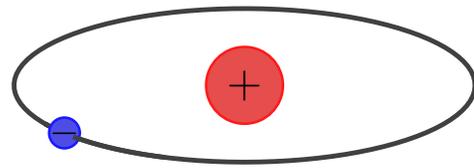
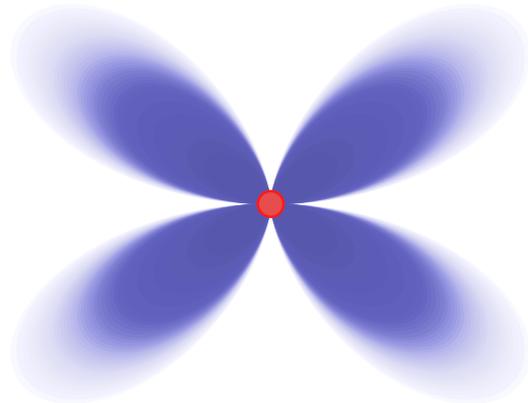


Figura 4.52 A partir de la tercera década del siglo XX, se cambió la manera de pensar sobre los átomos. Los electrones ya no son partículas que orbitan alrededor de un núcleo sino que se encuentran dispersados como nubes de probabilidad alrededor del núcleo. Como pueden estar en diferentes estados, influyen en la forma de la nube probabilística. La imagen de abajo muestra la probabilidad de encontrar un electrón en un átomo de hidrógeno en el estado ($l = 2, m = 1$).



4.9.1 Limitación a ecuaciones algebraicas

Ya los hindúes se daban a la tarea de proponer y resolver ecuaciones desde el siglo VIII a.C. En una ecuación aparecen *variables* y *constantes*, que se enlazan con operaciones para formar los dos *términos* en ambos lados de la ecuación. Por ejemplo, en la ecuación:

$$x + 3 = 6, \quad (10)$$

x es una variable, los números 3 y 6 son constantes y hay una sola operación involucrada, que es la suma. Nuestras constantes serán números enteros o racionales o, a veces, reales.

Resolver la ecuación significa encontrar los valores para las variables, de manera que la ecuación sea válida. Si sustituimos $x = 2$ en la ecuación 10, ésta se transforma en:

$$5 = 6$$

que, evidentemente, es falso. En cambio, si sustituimos $x = 3$, obtenemos la ecuación:

$$6 = 6$$

que es correcta. Por ello, $x = 3$ es solución de la ecuación 10 y $x = 2$, no lo es.

También hay ecuaciones que involucran *funciones*. Por ejemplo:

$$\cos(\pi x) = x^2 \quad (11)$$

es una ecuación que involucra dos funciones: coseno y elevar al cuadrado. Sin embargo, estas dos funciones no juegan el mismo papel, pues podemos entender al cuadrado de x como el producto de x con x misma, lo que equivale a decir que $x^2 = x \cdot x$. En cambio,

la función coseno no se puede reducir a una expresión sencilla que sólo involucre las operaciones básicas de sumar, restar, multiplicar y dividir.

Las *ecuaciones algebraicas* son aquellas que sólo involucran variables, constantes y operaciones básicas. Veamos algunos ejemplos sencillos.

Si en 10 cambiamos los valores de las constantes, cambiamos la ecuación pero no su estructura. De manera más general, podemos considerar ecuaciones que tienen la forma:

$$ax + b = 0, \quad (12)$$

donde a y b son constantes dadas. Si en una ecuación se indican constantes con una variable, se habla de *parámetros*. La ecuación 12 se llama *ecuación lineal* y podemos resolverla siempre que a no sea cero, en cuyo caso la solución es:

$$x = -\frac{b}{a}.$$

Un poco más complicada es la *ecuación cuadrática*:

$$ax^2 + bx + c = 0, \quad (13)$$

donde x es la variable mientras que a , b y c son los parámetros, es decir, constantes. Esta ecuación puede tener hasta dos soluciones:

$$x = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \quad x = \frac{-b - \sqrt{b^2 - 4ac}}{2a}. \quad (14)$$

Estas fórmulas se obtienen con un truco para completar cuadrados y que se explicará con detalle más adelante.

El número de soluciones depende del valor del *discriminante* $D = b^2 - 4ac$, es decir, del valor de lo que está dentro de la raíz cuadrada. Si $D = 0$, entonces sólo hay una solución: $x = -\frac{b}{2a}$. Cuando $D > 0$, las dos soluciones 14 son distintas.

Un caso interesante es cuando $D < 0$ pues, a partir de la fórmula, se tendría que calcular la raíz cuadrada de un número negativo. Pero no existe número real λ tal que $\lambda = \sqrt{-2}$, ya que al elevar al cuadrado se obtendría $\lambda^2 = -2$. ¡Es imposible tener el lado izquierdo positivo y el derecho, negativo! Por ello, podríamos decir que la ecuación 13 no tiene solución. También podríamos tratar de extender la noción de los números e incluir las raíces de números negativos. Aunque sea algo bastante raro, sorprendentemente, funciona muy bien. La humanidad tardó siglos en aceptar estos nuevos números que, en la actualidad, son muy comunes. Los números negativos son un concepto difícil, así que para entenderlos, iremos despacio.

4.9.2 El cálculo con “el número” $i = \sqrt{-1}$

Todo parte del supuesto de que existe un número i , que llamaremos *unidad imaginaria* —por razones históricas—, tal que $i^2 = -1$. Lo que buscamos es extender los números reales a un nuevo conjunto <más grande, llamado el de los *números complejos*, de manera muy similar a como se extienden los números naturales a los enteros, al añadir los negativos. Así, debemos poder formar nuevos números —por lo pronto, los marcamos con negritas para distinguirlos— como:

$$\mathbf{x} = 2 + \mathbf{i}, \quad \mathbf{y} = 3 - 2\mathbf{i}.$$

Con ellos, debemos poder calcular productos:

$$\mathbf{x} \cdot \mathbf{y} = (2 + \mathbf{i})(3 - 2\mathbf{i}).$$

Pero entonces, necesitamos que los nuevos números satisfagan las mismas leyes que los reales, respecto a la adición y el producto. En este caso, podríamos seguir:

$$\begin{aligned} \mathbf{x} \cdot \mathbf{y} &= (2 + \mathbf{i})(3 - 2\mathbf{i}) \\ &= 6 + 2(-2\mathbf{i}) + \mathbf{i}3 + \mathbf{i}(-2\mathbf{i}) && \text{por distributivo} \\ &= 6 + 2(-2\mathbf{i}) + 3\mathbf{i} + (-2\mathbf{i})\mathbf{i} && \text{por conmutativo de la multiplicación} \\ &= 6 + (-4)\mathbf{i} + 3\mathbf{i} + (-2)\mathbf{i}^2 && \text{por asociativo de la multiplicación} \\ &= 6 + (-4 + 3)\mathbf{i} + (-2)(-1) && \text{por distributividad y } \mathbf{i}^2 = -1 \\ &= 8 - \mathbf{i} && \text{por conmutatividad de la adición} \end{aligned}$$

Como se observa en el ejemplo anterior, obtuvimos de nuevo la forma $a + b\mathbf{i}$, donde a y b son dos números reales. Esto, en efecto, siempre es así y podemos hacer lo mismo que en el ejemplo para ver que:

$$(a + b\mathbf{i}) \cdot (c + d\mathbf{i}) = (ac - bd) + (ad + bc)\mathbf{i} \quad (15)$$

La suma es más fácil:

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= (2 + \mathbf{i}) + (3 - 2\mathbf{i}) \\ &= (2 + 3) + (\mathbf{i} - 2\mathbf{i}) && \text{por conmutatividad y asociatividad de la adición} \\ &= 5 - \mathbf{i} \end{aligned}$$

En general, se tiene que:

$$(a + b\mathbf{i}) + (c + d\mathbf{i}) = (a + c) + (b + d)\mathbf{i}.$$

También podemos restar:

$$(a + b\mathbf{i}) - (c + d\mathbf{i}) = (a - c) + (b - d)\mathbf{i},$$

y para ver que podemos dividir, tenemos que hacer un par de maniobras adicionales. Dividir entre 2 es lo mismo que multiplicar por $\frac{1}{2}$, esto es, por el multiplicativo inverso de 2. En general, esperamos que:

$$\mathbf{x} \div \mathbf{y} = \mathbf{x} \cdot \frac{1}{\mathbf{y}},$$

y como sabemos multiplicar gracias a la fórmula 15, es suficiente calcular el inverso multiplicativo de cualquier número complejo con forma $\mathbf{y} = c + d\mathbf{i}$. Para ello, observamos que:

$$(c + d\mathbf{i}) \cdot (c - d\mathbf{i}) = (c^2 - d(-d)) + 0\mathbf{i} = c^2 + d^2,$$

es un número real que no es cero si $\mathbf{y} \neq 0 + 0\mathbf{i} = 0$. Por ello, si $\mathbf{y} \neq 0$ tenemos que:

$$(c + d\mathbf{i}) \cdot \left(\frac{c}{c^2 + d^2} - \frac{d}{c^2 + d^2}\mathbf{i} \right) = \left(c \cdot \frac{c}{c^2 + d^2} - d \left(-\frac{d}{c^2 + d^2} \right) \right) + 0\mathbf{i} = 1.$$

Por consiguiente, se obtiene la siguiente fórmula para la formación de inversos multiplicativos:

$$\frac{1}{c + d\mathbf{i}} = \frac{c}{c^2 + d^2} - \frac{d}{c^2 + d^2}\mathbf{i}.$$

Para resumir, debemos considerar que todo parte de dos supuestos:

- Existe un número i , que llamaremos *unidad imaginaria*, tal que $i^2 = -1$.
- El número i y todos los nuevos números satisfacen las mismas leyes que involucran la adición y la multiplicación que los números reales.

Entonces, los números complejos son de la forma $a + b\mathbf{i}$, con a y b reales y que cumplen las siguientes leyes:

$(a + b\mathbf{i}) + (c + d\mathbf{i}) = (a + c) + (b + d)\mathbf{i}$	adición	16
$(a + b\mathbf{i}) \cdot (c + d\mathbf{i}) = (ac - bd) + (ad + bc)\mathbf{i}$	multiplicación	17
$\mathbf{0} = 0 + 0\mathbf{i}$	el cero (neutro aditivo)	18
$\mathbf{1} = 1 + 0\mathbf{i}$	el uno (neutro multiplicativo)	19
$-(a + b\mathbf{i}) = -a - b\mathbf{i}$	inverso aditivo	20
$\frac{1}{c + d\mathbf{i}} = \frac{c}{c^2 + d^2} - \frac{d}{c^2 + d^2}\mathbf{i}$	inverso multiplicativo	21

4.9.3 Representación geométrica de los números complejos

Todo lo anterior está muy bien, sólo que es muy abstracto. ¿Cómo imaginarse un número complejo? Para los números reales tenemos la recta real como objeto geométrico, que nos da una cierta posibilidad de imaginación e intuición. Para los números complejos hay también un modelo geométrico llamado el *plano complejo*, idea de Jean-Robert Argand —un suizo aficionado a las matemáticas—, donde cada número complejo $\mathbf{x} = a + b\mathbf{i}$ está dado por dos números reales a y b , que podemos imaginar como coordenadas.

El eje que se dibujó horizontalmente es la recta real, mientras que el eje vertical o imaginario contiene los múltiplos de \mathbf{i} . Ahora, debemos aclarar cómo se representan la adición y la multiplicación geoméricamente.

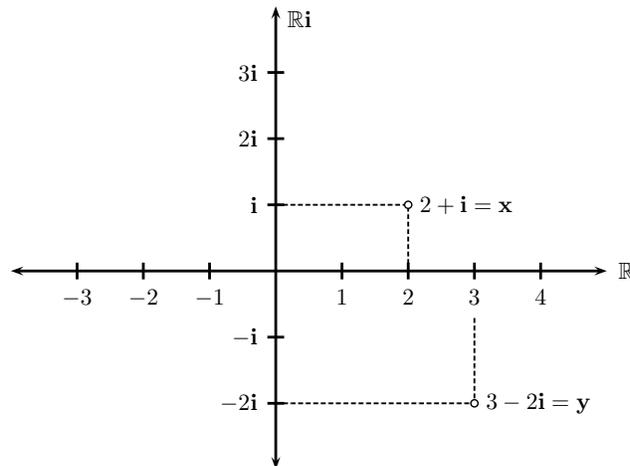


Figura 4.54: El plano complejo.

De la fórmula 16, observamos que la adición se obtiene como la cuarta esquina de un paralelogramo, con sus otras tres esquinas $\mathbf{0}$, \mathbf{x} y \mathbf{y} . Formar geoméricamente el inverso aditivo no es otra cosa que la reflexión en el punto $\mathbf{0}$. Para entender el significado de la multiplicación, tenemos primero que trabajar algebraicamente. Dado que cada número complejo \mathbf{x} se encuentra en el plano, se puede representar en *coordenadas polares* al indicar, por un lado, la distancia al origen r que también se llama el *valor absoluto* de \mathbf{x} y, por el otro, el ángulo $\angle \mathbf{1}-\mathbf{0}-\mathbf{x}$, es decir, el ángulo que forman los puntos $\mathbf{1}$, $\mathbf{0}$ y \mathbf{x} .

El número complejo con coordenadas polares r y φ es $\mathbf{x} = r(\cos(\varphi) + \text{sen}(\varphi)\mathbf{i})$. Las conversiones entre $\mathbf{x} = a + b\mathbf{i}$ y la forma polar se calculan con las siguientes fórmulas:

$$\begin{aligned} a &= r \cos(\varphi), & r &= \sqrt{a^2 + b^2}, \\ b &= r \text{sen}(\varphi), & \varphi &= \tan^{-1} \left(\frac{b}{a} \right). \end{aligned}$$

Entonces podemos reescribir la fórmula de multiplicación 17 de la siguiente manera:

$$\begin{aligned} r(\cos(\varphi) + \text{sen}(\varphi)\mathbf{i}) \cdot s(\cos(\psi) + \text{sen}(\psi)\mathbf{i}) &= rs(\cos(\varphi)\cos(\psi) - \text{sen}(\varphi)\text{sen}(\psi)) + \\ &+ rs(\cos(\varphi)\text{sen}(\psi) + \text{sen}(\varphi)\cos(\psi))\mathbf{i} \\ &= rs(\cos(\varphi + \psi) + \text{sen}(\varphi + \psi)\mathbf{i}) \end{aligned}$$

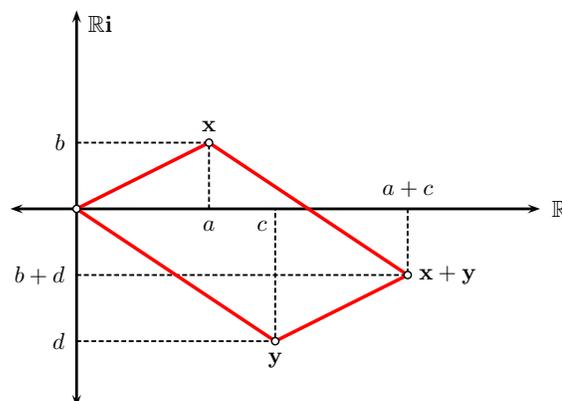


Figura 4.55: La suma de dos números complejos.

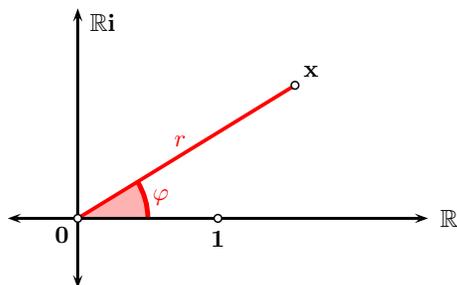


Figura 4.56 Coordenadas polares para números complejos.

Esto significa que, al multiplicar dos números complejos se multiplican sus valores absolutos y se suman sus ángulos.

Por último, regresamos al punto de partida. Al inicio del discurso, nos dejamos llevar hacia los números complejos con la finalidad de obtener raíces de números reales negativos. Ahora, queremos ver si podemos sacar la raíz cuadrada de cualquier número complejo o si es necesario extender nuestra noción de números una y otra vez.

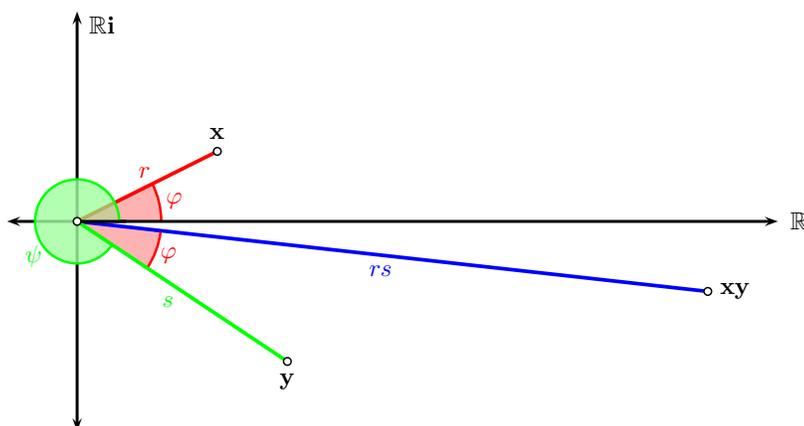


Figura 4.57 El producto de dos números complejos.

Si queremos sacar la raíz cuadrada de un número complejo \mathbf{x} , debemos encontrar otro número complejo \mathbf{y} tal que $\mathbf{y} \cdot \mathbf{y} = \mathbf{x}$. Con la intuición geométrica, esto ya no es difícil: usamos coordenadas polares para \mathbf{x} —digamos que tiene valor absoluto r y ángulo φ —. Entonces, \mathbf{y} debe tener valor absoluto \sqrt{r} y ángulo $\frac{\varphi}{2}$. Observamos que siempre —salvo cuando $\mathbf{x} = 0$ — hay dos soluciones: una es el inverso negativo de la otra.

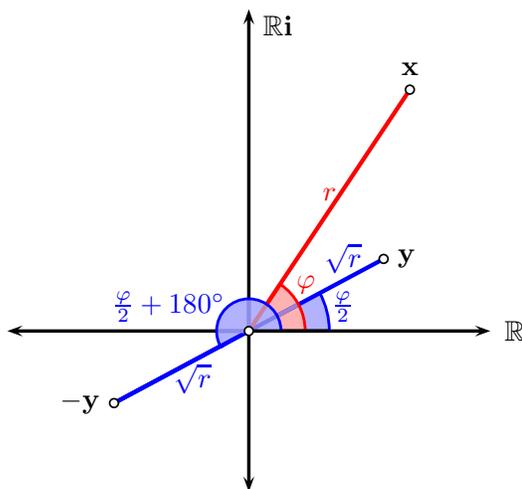


Figura 4.58 Las dos raíces de un número complejo.

Al recapitular, tenemos que los números complejos son expresiones de la forma $a + bi$, donde a y b son números reales que pueden sumarse y multiplicarse, restarse y dividirse. Por lo anterior, los números complejos forman un conjunto que los matemáticos llaman un *campo* y además, tienen una representación geométrica: el plano complejo.

Todavía queda la duda de si estos números complejos son algo más que una invención medio rara para poder sacar la raíz cuadrado de cualquier número real. ¿Qué relevancia tienen? ¿Se aplican en algún lugar o son un mero juguete matemático? Hasta el principio del siglo xx, los números complejos no tenían gran interés fuera de las matemáticas. Sin embargo, en los años veinte del siglo xx se forjó la *mecánica cuántica*, una teoría que logra describir con gran precisión los fenómenos de las interacciones de partículas atómicas. Esta teoría usa los números complejos de una manera esencial: sin ellos, no se podría formular adecuadamente pues explica, en forma contundente, el sistema de los elementos químicos que antes se había formulado con base en observaciones.

4.9.4 La ecuación cuadrática con coeficientes complejos

Antes nos desviamos considerablemente de nuestro objetivo principal: estudiar cómo podría responderse la pregunta ¿qué ecuación se puede resolver? Primero, la restringimos a las ecuaciones algebraicas. Y, hasta ahora, sólo llegamos a la conclusión de que con los números complejos podemos resolver cualquier *ecuación cuadrática* 13 con parámetros a, b y c reales. No obstante, vimos que podemos sacar la raíz cuadrada de cualquier número complejo, lo cual nos podría dar la idea de cuestionar si sería posible resolver 13, aún cuando los parámetros a, b, c sean *números complejos*, como en:

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

En efecto, lo anterior es cierto y se usa exactamente la misma fórmula 14 que antes. La razón se debe a que esta fórmula se deriva de manera *algebraica* sin hacer ninguna suposición, excepto que $a \neq 0$. Consideremos esta deducción y pensemos que a, b, c son números complejos. Primero, se divide la ecuación entre a :

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0.$$

Luego, se suma el término $\frac{b}{4a^2}$ en ambos lados y se observa que $x^2 + \frac{b}{a}x + \frac{b}{4a^2}$ es un cuadrado perfecto:

$$\begin{aligned} x^2 + \frac{b}{a}x + \frac{b}{4a^2} + \frac{c}{a} &= \frac{b}{4a^2} \\ \left(x + \frac{b}{2a}\right)^2 + \frac{c}{a} &= \frac{b}{4a^2} \\ \left(x + \frac{b}{2a}\right)^2 &= \frac{b}{4a^2} - \frac{c}{a} = \frac{b - 4ac}{4a^2} \end{aligned}$$

En la última fila, se restó el término $\frac{c}{a}$ en ambos lados. Ahora se puede sacar la raíz. Recordemos aquí que ello siempre se puede, dado que $\frac{b}{4a} - \frac{c}{a}$ es un número complejo y, por lo tanto, siempre hay dos soluciones, salvo cuando $\frac{b}{4a^2} - \frac{c}{a} = 0$.

$$x + \frac{b}{2a} = \pm \sqrt{\frac{b - 4ac}{4a^2}} = \frac{\sqrt{b - 4ac}}{2a}$$

$$x = \pm \frac{\sqrt{b - 4ac}}{2a} - \frac{b}{2a} = \frac{-b \pm \sqrt{b - 4ac}}{2a}$$

De nuevo, obtuvimos la fórmula de solución 14 y de nada importó que los coeficientes **a**, **b** y **c** sean complejos.

4.9.5 Las ecuaciones de tercer y cuarto grado

No sólo la ecuación cuadrática, sino también la ecuación algebraica de *tercer grado* que se representa como:

$$ax^3 + bx^2 + cx + d = 0, \quad (22)$$

admite una fórmula de solución. Lo mismo sucede con la ecuación algebraica de *cuarto grado*:

$$ax^4 + bx^3 + cx^2 + dx + e = 0.$$

Aquí ya no resaltamos que los coeficientes pueden ser números racionales, reales o complejos, pues el tratamiento que daremos no depende de ello.

La resolución de la ecuación de tercer grado se le atribuye a Niccolò Tartaglia en 1530. La historia cuenta que sólo después de considerables insistencias por parte de Gerolamo Cardano, Tartaglia le confió su secreto bajo la promesa de nunca publicarlo. Unos diez años después, Lodovico Ferrari —un estudiante de Cardano— encontró la solución para la ecuación de cuarto grado. En 1545, el propio Cardano publicó ambos resultados en un libro que hoy se conoce como las *fórmulas de Cardano*.

A continuación, daremos una breve idea de la solución de la ecuación de tercer grado. Primero, se hacen dos maniobras para reducir el número de parámetros involucrados, al dividir la ecuación 22 entre *a*:

$$x^3 + b'x^2 + c'x + d' = 0,$$

donde $b' = \frac{b}{a}$, $c' = \frac{c}{a}$, $d' = \frac{d}{a}$. Después, se sustituye:

$$x = y - \frac{b'}{3}, \quad (23)$$

y se obtiene que:

$$\left(y - \frac{b'}{3}\right)^3 + b' \left(y - \frac{b'}{3}\right)^2 + c' \left(y - \frac{b'}{3}\right) + d' = 0,$$

que al expandir, se convierte en:



$$y^3 - 3y^2 \frac{b'}{3} + 3y \left(\frac{b'}{3}\right)^2 - \left(\frac{b'}{3}\right)^3 + b'y^2 - 2b'y \frac{b'}{3} + b' \left(\frac{b'}{3}\right)^2 + c'y - c' \frac{b'}{3} + d' = 0$$

$$y^3 + \underbrace{\left(c' - \frac{b'^2}{3}\right)}_{=p} y + \underbrace{\left(\frac{2b'^3}{27} - \frac{b'c'}{3} + d'\right)}_{=q} = 0$$

Figura 4.59
Gerolamo Cardano.

Con las definiciones de p y q como hemos indicado, reducimos el número de parámetros a sólo dos. Ahora, sólo tenemos que resolver la ecuación:

$$y^3 + py + q = 0. \quad (24)$$

El truco consiste en suponer que tenemos una solución y escribirla como una diferencia $y = r - s$. Al sustituir, se obtiene que:

$$r^3 - 3r^2s + 3rs^2 - s^3 + pr - ps + q = 0 \quad (25)$$

$$(r^3 - s^3 + q) + (r - s)(p - 3rs) = 0$$

Ahora, tratemos de encontrar los números r y s , tales que tenemos simultáneamente que ambos sumandos en 25 son cero:

$$r^3 - s^3 + q = 0, \quad (26)$$

$$(r - s)(p - 3rs) = 0. \quad (27)$$

Revisemos primero la segunda ecuación. Como tenemos un producto que es cero, uno de los dos factores también debe ser cero. Si $r - s = 0$, entonces $y = 0$. Pero si $y = 0$ es una solución, entonces $q = 0$ y la ecuación 24 se puede escribir como $(y^2 + p)y = 0$, donde las otras dos soluciones son $\pm\sqrt{-p}$. Por lo tanto, en caso de que $y = 0$ sea solución, no tenemos que seguir.

El caso $y \neq 0$ es más interesante, pues obtenemos de 26 que:

$$3rs = p,$$

por lo que $s = \frac{p}{3r}$. Al sustituir s por $\frac{p}{3r}$ en (27), se obtiene:

$$r^3 - \frac{p^3}{27r^3} - q = 0.$$

Si multiplicamos por r^3 y sustituimos $R = r^3$, obtenemos la ecuación:

$$R^2 + qR - \frac{p^3}{27} = 0$$

que es cuadrática para la variable R y cuyas soluciones son:

$$R = \frac{-q \pm \sqrt{q^2 + \frac{4p^3}{27}}}{2} = -\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}$$

Además podemos fijar un signo, como veremos pronto. Digamos que $R = -\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}$. Entonces:

$$r = \sqrt[3]{\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}, \quad s = \sqrt[3]{\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} \tag{28}$$

donde la expresión para s se obtiene de 27: simplemente al sustituir las dos expresiones para r y s del lado izquierdo de 27 y después se simplifica, para obtener el lado derecho. Así, hemos encontrado una solución de 24:

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

Si quisiéramos encontrar la solución para la ecuación de tercer grado 22, tendríamos que sustituir y en 23 para obtener x .

Observemos que en 28 tenemos tres soluciones de $r^3 = R$: si tenemos una solución r , también αr y $\alpha^2 r$ son soluciones, donde $\alpha = \cos(120^\circ) + i\sin(120^\circ)$.

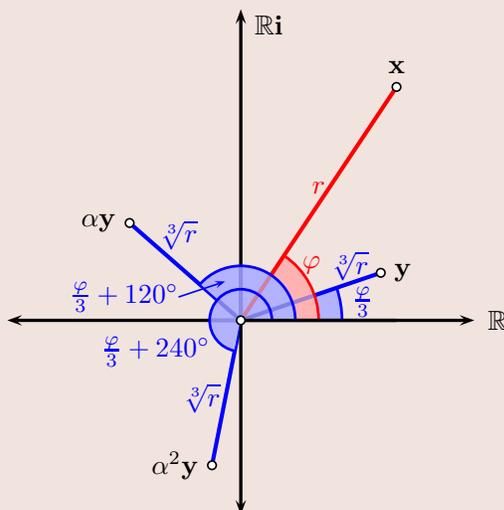


Figura 4.60
Soluciones
de $r^3 = R$.

Por lo tanto, las otras dos soluciones son:

$$y_2 = \alpha \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} - \alpha^2 \sqrt[3]{\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$$

$$y_3 = \alpha^2 \sqrt[3]{\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} - \alpha \sqrt[3]{\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$$

Las fórmulas para la ecuación de cuarto grado son aún más complicadas y ni siquiera las indicaremos.

Queremos ahora ocuparnos de las ecuaciones de grado mayor que cuatro. Posterior a Cardano, los matemáticos intentaron reiteradamente resolver ecuaciones de grado cinco y encontrar una fórmula de solución, pero siempre fracasaron. En 1799, Paolo Ruffini dió ideas sobre cómo demostrar que tal fórmula no puede existir. En 1824, Niels Hendrik Abel, un matemático sueco, llenó los “huecos” de la demostración de Ruffini.

4.9.6 El teorema fundamental del álgebra

Figura 4.61 Niels
Hendrik Abel.

Hay un punto delicado aquí: la demostración de Abel no afirma que *no hay solución*, sino que *no hay una fórmula de solución*. Ya desde el siglo xvii se pensaba que la ecuación algebraica de grado n tiene n soluciones que pueden repetirse. Para hacer más preciso lo anterior: hay n números complejos y_1, y_2, \dots, y_n tal que:

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_2 x^2 + a_1 x + a_0 = a_n \cdot (x - y_1) \cdot (x - y_2) \cdot \dots \cdot (x - y_n).$$

Así, si quisiéramos resolver la ecuación:

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

tendríamos que resolver:

$$a_n \cdot (x - y_1) \cdot (x - y_2) \cdot \dots \cdot (x - y_n) = 0, \quad (29)$$

lo que resulta fácil dado que un producto de factores es cero sólo si uno de sus factores lo es. Por ello —y como $a_n \neq 0$ —, se tiene que $x - y_i = 0$ para algún i , así que $x = y_i$ para algún i . Lo anterior muestra que sí tiene sentido decir que una solución es “doble” o “múltiple”. El resultado formal se llama el Teorema fundamental del álgebra y dice que: *Cada ecuación algebraica:*

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

con coeficientes a_0, a_1, \dots, a_n números complejos tiene n soluciones complejas, si se cuentan posibles repeticiones.

Por ejemplo, la ecuación $x^2 - 2x + 1 = 0$ tiene las dos soluciones: 1 y 1 ya que $x^2 - 2x + 1 = (x - 1)(x - 1)$.

La primera demostración correcta es de Jean-Robert Argand, en 1806. Gauss hizo

varias demostraciones de este resultado a lo largo de su vida. El Teorema fundamental del álgebra muestra que siempre hay soluciones complejas pero no es constructivo, es decir, no da ningún método para encontrarlas. Es un punto delicado: el resultado afirma la existencia de soluciones sin exhibir ninguna de ellas. Sólo dice que existen, nada más. Por ello, se percibe la gran importancia de fórmulas para calcular las soluciones de manera explícita. Sin embargo, el Teorema de Abel-Ruffini dice que éstas no se pueden encontrar con una fórmula si $n \geq 5$. Lo anterior quiere decir que las soluciones existen y pueden encontrarse de manera *aproximativa*, es decir, hasta cualquier precisión de dígitos, pero si $n \geq 5$, *no es posible* sustituir los coeficientes a_0, a_1, \dots, a_n de la ecuación 29 en alguna fórmula para obtener de golpe las soluciones.

Figura 4.62 Ruffini.

4.9.7 Polinomios, raíces y simetrías

La demostración de Abel-Ruffini tiene una desventaja: es técnica y dificulta el entender las razones por las que ecuaciones de grado 1, 2, 3 y 4 se resuelven con una fórmula, pero las de grado 5 —y mayores— no se pueden resolver de esta manera. En la actualidad, se usa una teoría basada en las ideas de Evariste Galois —un matemático francés contemporáneo de Abel— para demostrarlo. Sin embargo, antes de llegar a la *teoría de Galois*, tendremos que preparar unas nociones sobre polinomios.

Los términos de la forma:

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_2 x^2 + a_1 x + a_0 \quad (30)$$

se llaman *polinomios* en la variable x con coeficientes a_0, a_1, \dots, a_n . Si $a_n \neq 0$ entonces n es el *grado* de $f(x)$.

En lo que sigue, será importante considerar polinomios *sobre diferentes campos*: si todos los coeficientes pertenecen al campo K de los números racionales, diremos que $f(x)$ es un *polinomio racional*. Los polinomios racionales forman un conjunto $[x]$ donde es posible sumar y multiplicar, también hay elementos neutros aditivos y multiplicativos e inversos aditivos, pero no existen inversos multiplicativos. Se dice que $[x]$ es un *anillo*. De manera similar si K es un campo, denotamos con $K[x]$ al anillo de polinomios con coeficientes en K . Un caso importante son los números complejos: $K = \mathbb{C}$.

Si en la expresión 30 se sustituye la variable x por un número s , entonces se obtiene un número que se denota por $f(s)$. Un número s que satisface $f(s) = 0$ se llama *raíz* del polinomio $f(x)$. De hecho la notación $f(x)$ sugiere que se trata de una función y eso es cierto: se trata de la función polinomial que se obtiene al sustituir diferentes valores para x . Si s es raíz de $f(x)$, entonces $f(x)$ se puede factorizar:

$$f(x) = g(x) \cdot (x - s),$$

es decir, existe un polinomio $g(x)$ con grado menor que $f(x)$ y factor lineal de $x - s$, tal que su producto es $f(x)$. Pero los coeficientes del polinomio $g(x)$ serán expresiones que pueden involucrar a la raíz s como se explica en el siguiente ejemplo.



Por ejemplo, si $f(x) = x^3 - x + 1$ y s es alguna de sus raíces entonces se puede hacer la división de polinomios $(x^3 - x + 1) \div (x - s)$, que se explica a continuación.

Para hacer la operación de $f(x) = x^3 - x + 1$ entre $x - s$, se divide el monomio de $f(x)$ de mayor grado entre el monomio de mayor grado de $x - s$; en este caso hay que dividir x^3 entre x y se obtiene x^2 . Luego, se forma la resta $f_1(x)$:

$$\begin{aligned} f_1(x) &= (x^3 - x + 1) - x^2 \cdot (x - s) \\ &= x^3 - x + 1 - x^3 + sx^2 \\ &= sx^2 - x + 1 \end{aligned}$$

Después se procede en forma similar: el monomio de $f_1(x)$ de mayor grado es sx^2 , por lo que al dividirlo entre x , tenemos sx . Esta vez, se resta $(sx) \cdot (x - s) = sx^2 - s^2x$ de $f_1(x)$ y se obtiene:

$$\begin{aligned} f_2(x) &= (sx^2 - x + 1) - sx \cdot (x - s) \\ &= sx^2 - x + 1 - sx^2 + s^2x \\ &= (s^2 - 1)x + 1, \end{aligned}$$

cuyo monomio de grado mayor es $(s^2 - 1)x$. Nuevamente, al dividirlo entre x se obtiene $(s^2 - 1)$ y al calcular la resta:

$$\begin{aligned} f_3(x) &= ((s^2 - 1)x + 1) - (s^2 - 1) \cdot (x - s) \\ &= (s^2 - 1)x + 1 - (s^2 - 1)x + (s^2 - 1)s \\ &= s^3 - s + 1. \end{aligned}$$

Ahora hay que observar que $f_3(x) = s^3 - s + 1 = f(s) = 0$, lo que muestra que la división se hizo sin resta. Se obtuvo la división:

$$(x^3 - x + 1) : (x - s) = x^2 + sx + (s^2 - 1)$$

El teorema fundamental del álgebra tiene una interpretación alterna: cada polinomio $f(x) \in \mathbb{C}[x]$ se puede escribir como producto:

$$a_n \cdot (x - s_1) \cdot (x - s_2) \cdot \dots \cdot (x - s_n), \quad (31)$$

donde s_1, \dots, s_n son las soluciones de la ecuación 29. En otras palabras, cada polinomio con coeficientes complejos se descompone en *factores lineales*. Por ejemplo:

$$x^2 + 1 = (x - \mathbf{i}) \cdot (x + \mathbf{i}), \quad (32)$$

dado que $\pm \mathbf{i}$ son las dos soluciones de la ecuación $x^2 + 1 = 0$. El polinomio $x^2 + 1 \in \mathbb{Q}[x]$ no se descompone en factores lineales pues $\pm \mathbf{i}$ no son racionales, pero $x^2 + 1 \in \mathbb{Q}[x]$ sí

se descompone como 32. La descomposición en 31 es única en el siguiente sentido: el factor a_n es único y los elementos s_1, \dots, s_n son únicos salvo permutación, es decir, dos representaciones 31 —si son iguales— sólo pueden distinguirse en el orden en que se escriben los factores $(x - s_1), \dots, (x - s_n)$.

Hemos llegado a un punto crucial: ¿cuál es el número \mathbf{i} y cuál es $-\mathbf{i}$? Recordemos que antes simplemente postulamos su existencia y vimos que podíamos hacer cálculos con números de la forma $a + b\mathbf{i}$. El número \mathbf{i} es una solución de la ecuación $x^2 + 1 = 0$, mientras que la otra es $-\mathbf{i}$. Pero no hay ninguna manera de saber cuál es cuál, sólo sabemos que una solución es la inversa aditiva de la otra. Si otra persona escoge $\mathbf{j} = -\mathbf{i}$ como solución, entonces obtiene una copia igual de números complejos. Lo anterior se puede decir todavía de otra manera: hay una *simetría* en los números complejos, donde la función:

$$a + b\mathbf{i} \mapsto \overline{a + b\mathbf{i}} = a - b\mathbf{i}$$

se llama *conjugación*. La conjugación es un *automorfismo*, es decir, es biyectiva y conserva las operaciones de adición, multiplicación y la formación de inversos:

$$\begin{aligned}\overline{\mathbf{x} + \mathbf{y}} &= \overline{\mathbf{x}} + \overline{\mathbf{y}}, \\ \overline{\mathbf{x} \cdot \mathbf{y}} &= \overline{\mathbf{x}} \cdot \overline{\mathbf{y}}, \\ \overline{-\mathbf{x}} &= -\overline{\mathbf{x}} \\ \overline{1/\mathbf{x}} &= 1/\overline{\mathbf{x}}.\end{aligned}$$

Los números fijos bajo la conjugación son los números reales, dado que $a + b\mathbf{i} = \overline{a + b\mathbf{i}}$ implica que $b = 0$.

Sin embargo, en la realidad no necesitamos los números reales para estudiar la ecuación $x^2 + 1 = 0$. Los coeficientes son números enteros. El campo *más pequeño* (respecto a la contención de conjuntos) que contiene a los coeficientes 0, 1 es \mathbb{Q} , el campo de los números racionales. Por ejemplo, el campo más pequeño que contiene los coeficientes y las soluciones $\pm\mathbf{i}$ es el campo K , cuyos elementos son de la forma $a + b\mathbf{i}$ donde $a, b \in \mathbb{Q}$. Se dice que K es el *campo de descomposición* del polinomio $x^2 + 1$ sobre \mathbb{Q} , dado que $x^2 + 1$ se descompone en K factores lineales 32 y K es el campo más chico con esta propiedad.

En general, fijamos un polinomio $f(x)$ como en 30, con coeficientes $a_0, a_1, \dots, a_n \in \mathbb{Q}$ y definimos K como el *campo de descomposición* de $f(x)$, es decir, el campo más pequeño que contiene \mathbb{Q} y todas las soluciones de la ecuación $f(x) = 0$. Por ejemplo, si $f(x) = x^3 - 2$ es el polinomio a descomponer, entonces $\sqrt[3]{2}$ es una raíz y $f(x)$ se descompone como:

$$x^3 - 2 = (x - \sqrt[3]{2})(x^2 + \sqrt[3]{2}x + \sqrt[3]{2}^2), \quad (33)$$

en $K[x]$, donde K es el campo:

$$K = \mathbb{Q}(\sqrt[3]{2}) = \{b_0 + b_1\sqrt[3]{2} + b_2\sqrt[3]{4} \mid b_0, b_1, b_2 \in \mathbb{Q}\}.$$

Con ayuda de la fórmula general para resolver ecuaciones cuadráticas y después de algunas transformaciones algebraicas, se obtiene que el segundo factor de 33 se puede escribir como:

$$x^2 + \sqrt[3]{2}x + \sqrt[3]{2}^2 = (x - \rho\sqrt[3]{2})(x - \rho^2\sqrt[3]{2}),$$

donde $\rho = \frac{-1+\sqrt{3}}{2}$ es una *tercera raíz primitiva de la unidad*, es decir, una solución de la ecuación $x^3 = 1$ que no es solución de una ecuación $x^p = 1$ para $p < 3$. Las tres raíces de $f(x) = x^3 - x - 1$ son entonces $z_1 = \sqrt[3]{2}$, $z_2 = \rho\sqrt[3]{2}$ y $z_3 = \rho^2\sqrt[3]{2}$.

En seguida, denotamos con $\text{Gal}(K : \mathbb{Q})$ al conjunto de todos los automorfismos de K que fijan \mathbb{Q} punto por punto, es decir $\varphi(z) = z$ para cada $z \in \mathbb{Q}$ y cada $\varphi \in \text{Gal}(K : \mathbb{Q})$. Por construcción, $\text{Gal}(K : \mathbb{Q})$ es un grupo, dado que podemos componer dos automorfismos que fijan \mathbb{Q} para obtener otra vez un automorfismo que fija \mathbb{Q} ; la identidad de K siempre es un elemento de $\text{Gal}(K : \mathbb{Q})$ y un inverso de un elemento de $\text{Gal}(K : \mathbb{Q})$ es, nuevamente, un elemento de $\text{Gal}(K : \mathbb{Q})$.

El grupo $\text{Gal}(K : \mathbb{Q})$ se llama *grupo de Galois* de K sobre \mathbb{Q} . En general, un grupo es un conjunto con una operación binaria —como una multiplicación, adición, composición o cualquier otra operación binaria— que tiene elemento neutro e inverso.

Cada elemento φ de $\text{Gal}(K : \mathbb{Q})$ fija los coeficientes del polinomio $f(x)$, es decir, $\varphi(a_i) = a_i$ dado que $a_i \in \mathbb{Q}$. Por otro lado, el automorfismo φ induce una función entre los polinomios en $K[x]$:

$$b_mx^m + \dots + b_1x + b_0 \mapsto \varphi(b_m)x^m + \dots + \varphi(b_1)x + \varphi(b_0).$$

Esta función también conserva suma, multiplicación, elementos neutros e inversos aditivos de $K[x]$, dado que φ es un automorfismo que fija los elementos de K . Esto tiene una fuerte implicación si aplicamos φ a la ecuación:

$$f(x) = a_nx^n + \dots + a_1x + a_0 = a_n \cdot (x - s_1) \cdots \cdots (x - s_n),$$

entonces, el lado izquierdo no cambia. El lado derecho, $a_n \cdot (x - \varphi(s_1)) \cdots \cdots (x - \varphi(s_n))$ es una factorización de $f(x)$. Como ésta es única, concluimos que φ induce una *permutación* de las raíces s_1, \dots, s_n .

Por otro lado, si $\varphi, \psi \in \text{Gal}(K : \mathbb{Q})$ inducen la misma permutación en las raíces s_1, \dots, s_n , entonces $\varphi = \psi$ —esto sigue de que K es el campo más pequeño que contiene \mathbb{Q} y s_1, \dots, s_n , un argumento que aquí no detallamos. Como sólo hay un número finito de permutaciones de s_1, \dots, s_n concluimos que $\text{Gal}(K : \mathbb{Q})$ es un grupo *finito*. Por ejemplo, si $f(x) = x^2 + 1$ entonces $\text{Gal}(K : \mathbb{Q})$ contiene dos elementos: la conjugación y la identidad.

Si $f(x) = x^4 + 1$ entonces los cuatro números:

$$z_1 = \frac{\sqrt{2}}{2}(1 + \mathbf{i}), \quad z_2 = \frac{\sqrt{2}}{2}(1 - \mathbf{i}), \quad z_3 = -z_1, \quad z_4 = -z_2,$$

son raíces de $f(x)$, como se comprueba al sustituir y dado que $f(x)$ tiene grado 4, no puede tener más raíces. El campo de descomposición K contiene, por lo tanto, estos cuatro números pero también al cuadrado de z_1 :

$$z_1^2 = \frac{2}{4}(1 + 2\mathbf{i} + \mathbf{i}^2) = \frac{1}{2}(1 + 2\mathbf{i} - 1) = \mathbf{i}$$

y además, la suma:

$$z_1 + z_2 = \sqrt{2}.$$

Por lo tanto, K contiene números de la forma $a_0 + a_1\sqrt{2} + a_2\mathbf{i} + a_3\sqrt{2}\mathbf{i}$, donde

los coeficientes a_0, a_1, a_2 y a_3 son números racionales. No es difícil ver que, en efecto, K consiste exactamente en estos números: hay que ver que el conjunto formado por dichos números es cerrado bajo la multiplicación e inversos multiplicativos.

Veamos ahora cómo los elementos del grupo $\text{Gal}(K : \mathbb{Q})$ permutan las raíces: si por ejemplo $\varphi(z_1) = z_3$, entonces $\varphi(z_3) = \varphi(-z_1) = \varphi(-1)\varphi(z_1) = -z_3 = z_1$, ya que $\varphi(-1) = -1$. Además, $\varphi(\mathbf{i}) = \varphi(z_1^2) = \varphi(z_1)^2 = (z_3)^2 = (-z_1)^2 = \mathbf{i}$. Entonces, $\varphi(z_1) = \frac{1}{2}\varphi(\sqrt{2}) + \frac{1}{2}\varphi(\sqrt{2})\mathbf{i}$ que debe ser igual a $-z_1 = -\frac{1}{2}\sqrt{2} - \frac{1}{2}\sqrt{2}\mathbf{i}$. Lo anterior muestra que $\varphi(\sqrt{2}) = -\sqrt{2}$. Con ello se concluye que $\varphi(z_3) = -z_3 = z_4$ y, por lo tanto, φ intercambia z_1 con z_3 y z_2 con z_4 . De manera similar se puede mostrar que el automorfismo ψ que satisface $\psi(z_1) = z_2$ intercambia z_1 con z_2 y φ con z_4 , y que el automorfismo η que satisface $\eta(z_1) = z_4$ intercambia z_1 con z_4 y z_2 con z_3 . Con la identidad —que es el automorfismo que deja todo fijo— se obtiene un grupo de 4 automorfismos.

Se observa que $\varphi(\mathbf{i}) = \mathbf{i}$, $\psi(\sqrt{2}) = \sqrt{2}$ y $\eta(\sqrt{2}\mathbf{i}) = \sqrt{2}\mathbf{i}$. El conjunto de todos los números y que satisfacen $\varphi(y) = y$, forma un campo L que contiene \mathbb{Q} y que es contenido en K . Se llama el *campo de puntos fijos* de φ . Más general, a los campos que contienen \mathbb{Q} y que están contenidos en K se les llama *campos intermedios* y juegan un papel importante en el resto de esta sección. El campo de puntos fijos de φ es justamente el campo $\mathbb{Q}(\mathbf{i})$, que se llama el campo de los números de Gauss, en honor a Carl Friedrich Gauss. El campo de los puntos fijos de ψ es $\mathbb{Q}(\sqrt{2})$ y el de η es $\mathbb{Q}(\sqrt{2}\mathbf{i})$.

Figura 4.63 Galois.

La teoría de Galois

La teoría de Galois parte de un polinomio *irreducible* $f(x) \in \mathbb{Q}[x]$, es decir, de un polinomio que no se puede factorizar como $g(x) \cdot h(x)$, donde $g(x)$ y $h(x)$ son polinomios *no constantes* en $\mathbb{Q}[x]$. Se denota con K el campo de descomposición de $f(x)$, es decir el campo K más pequeño que contiene \mathbb{Q} .

La teoría de Galois establece una relación entre dos conjuntos. Por un lado, están los *subgrupos* de $\text{Gal}(K : \mathbb{Q})$, es decir, los subconjuntos de $\text{Gal}(K : \mathbb{Q})$ que son cerrados bajo la multiplicación y la formación de inversos —y que siempre contienen la identidad como elemento—. Por otro, están los campos *L intermedios*, es decir, los campos L que satisfacen $\mathbb{Q} \subset L \subset K$. La correspondencia se da por dos funciones r y s :

$$\{H \mid H \text{ es subgrupo de } \text{Gal}(K : \mathbb{Q})\} \xrightleftharpoons[s]{r} \{L \mid L \text{ es campo intermedio } \mathbb{Q} \subset L \subset K\}$$

Es decir, r asigna a cada subgrupo de $\text{Gal}(K : \mathbb{Q})$ un campo intermedio y conversamente, la función s asigna a cada campo intermedio uno de estos subgrupos. La función r se define por:

$$r(H) = \{z \in K \mid \forall \varphi \in H, \varphi(z) = z\},$$

es decir, $r(H)$ es el conjunto de números $z \in K$ que permanecen fijos bajo cualquier φ de H . Por otro lado, se define que:

$$s(L) = \{\varphi \in \text{Gal}(K : \mathbb{Q}) \mid \forall z \in L, \varphi(z) = z\},$$

es decir, $s(L)$ es el conjunto de automorfismos $\varphi \in \text{Gal}(K : \mathbb{Q})$ tal que cualquier $z \in L$ permanece fijo bajo φ . No es difícil ver que r y s son funciones inversas una de la otra.

Las contenciones se invierten bajo r y s : si $H \subset H'$ son subgrupos de $\text{Gal}(K : \mathbb{Q})$, entonces $r(H) \supset r(H')$ y viceversa, si $L \subset L'$ son dos campos intermedios, entonces $s(L) \supset s(L')$. Al subgrupo $H = \{\text{id}\}$ que contiene solamente la identidad y que es el subgrupo más pequeño, corresponde el campo intermedio más grande, que es K mismo, mientras que al subgrupo más grande, que es $\varphi \in \text{Gal}(K : \mathbb{Q})$, corresponde el campo intermedio más pequeño que es \mathbb{Q} .

Así, en el caso del polinomio $f(x) = x^4 + 1$ se ve que los cálculos anteriores ejemplifican la teoría de Galois, como se ve en la Figura 4.64.

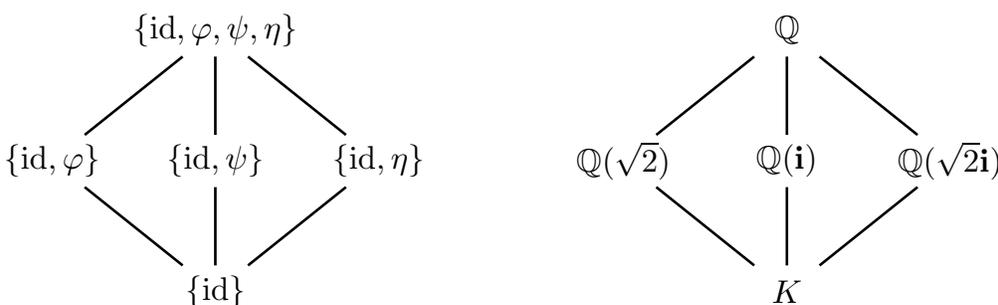


Figura 4.64: Del lado izquierdo, la correspondencia entre subgrupos del grupo de Galois y del derecho, los campos intermedios.

Se puede mostrar que el polinomio $f(x) = x^5 + x + 1$ es irreducible con cinco raíces distintas s_1, \dots, s_5 y el grupo de Galois $\text{Gal}(K : \mathbb{Q})$ es todo el grupo de permutaciones de las cinco raíces s_1, \dots, s_5 . En general, para cada n es posible encontrar un polinomio irreducible $f(x) \in \mathbb{Q}[x]$ de grado n tal que el grupo $\text{Gal}(K : \mathbb{Q})$ es todo el grupo de permutaciones de las n raíces distintas.

Por ello, es importante estudiar los subgrupos del grupo de permutaciones de n elementos distintos s_1, \dots, s_n . Dicho estudio es el inicio de la teoría de grupos que se estudian por sí mismos desde entonces, sin conexión con el origen de la resolución de ecuaciones. El grupo de permutaciones tiene una estructura que no depende de cuáles sean los n elementos a permutar. Por consiguiente, las podemos elegir como los números $1, 2, \dots, n$ para facilitar la notación. El grupo resultante se denota por S_n y se llama *grupo simétrico*. A continuación, resumimos algunos de los hechos de la teoría de grupos que serán de relevancia en lo que sigue.

Una *transposición* es una permutación que deja fijos todos los elementos, salvo dos que intercambia. Si i y j son los dos elementos intercambiados, la transposición se denota por $(i j)$. Cada permutación se puede escribir como composición de transposiciones. El número de transposiciones usadas en una composición no es invariante, por ejemplo, se tiene que:

$$(12)(23)(12) = (13).$$

Sin embargo, la paridad del número de transposiciones usadas sí es una invariante: una composición de 3 transposiciones no se puede escribir como composición de un

número par de transposiciones. Por ello, el conjunto de composiciones A_n de un número par de transposiciones forma un subgrupo, llamado el *grupo alternante*.

El subgrupo A_n es un *subgrupo normal* de S_n , es decir, para cada $\alpha \in A_n$ y cada $\sigma \in S_n$ se tiene $\sigma^{-1}\alpha\sigma \in A_n$. Si $n = 4$, entonces A_4 tiene un subgrupo normal a la vez, que es el *subgrupo de Klein*:

$$K = \{\text{Id}, (12)(34), (13)(24), (14)(23)\}.$$

Éste es justamente el grupo de Galois $\text{Gal}(K : \mathbb{Q})$ si K es el campo de descomposición del polinomio $f(x) = x^4 + 1$, el ejemplo que hemos seguido más de cerca hasta ahora.

Aquí se debe tener cuidado: K es *subgrupo no normal* de S_4 pero sí es normal en A_4 . El grupo K tiene, a la vez, tres subgrupos normales: $Z_1 = \{\text{Id}, (12)(34)\}$, $Z_2 = \{\text{Id}, (13)(24)\}$ y $Z_3 = \{\text{Id}, (14)(23)\}$. Si H es subgrupo normal de H' , se denota por $H \triangleleft H'$. Con ello tenemos la siguiente sucesión de subgrupos normales:

$$\{\text{Id}\} \triangleleft Z_1 \triangleleft K \triangleleft A_4 \triangleleft S_4.$$

Ésta es una *sucesión de composición* de S_4 donde ya no es posible insertar más subgrupos en medio sin tener igualdad. Si escribimos la cardinalidad en lugar de los grupos, obtenemos la siguiente sucesión de divisores $1 \triangleleft 2 \triangleleft 4 \triangleleft 12 \triangleleft 24$. Los factores entre dos números sucesivos son primos.

En la teoría de grupos es posible formar cocientes de subgrupos si éstos son normales y, además, la cardinalidad de estos grupos cocientes es justo el factor de los números consecutivos correspondientes. Más aún, cualquier grupo con un número primo de elementos es *conmutativo*. Los grupos conmutativos se llaman también *abelianos*, en honor a Abel, quién enfatizó la importancia de este hecho para la resolución de las ecuaciones.

Las siguientes cadenas muestran sucesiones de composición de S_2 y S_3 :

$$\{\text{Id}\} \triangleleft A_2 \triangleleft S_2, \quad \{\text{Id}\} \triangleleft A_3 \triangleleft S_3.$$

Las cardinalidades son $1 \triangleleft 2 \triangleleft 4$ y $1 \triangleleft 3 \triangleleft 6$, respectivamente. En consecuencia, S_2 , S_3 y S_4 tienen sucesiones de composición con factores abelianos. También S_5 y, más general, S_n para $n \geq 5$ tiene una sucesión de composición:

$$\{\text{Id}\} \triangleleft A_n \triangleleft S_n,$$

pero A_n *no es conmutativo*. Los únicos subgrupos normales de A_n son $\{\text{Id}\}$ y A_n . Es fácil ver que, A_n para $n \geq 5$, no es conmutativo. La importancia de esta diferencia reside en lo que se explica a continuación.

La teoría de Galois establece una muy buena correspondencia entre los subgrupos de $\text{Gal}(K : \mathbb{Q})$ y los campos intermedios $\mathbb{Q} \subset L \subset K$: el subgrupo $H \subset H'$ es normal en H' si y solamente si $L = r(H)$ es una *extensión normal* de $L' = r(H)$, es decir, cada $g(x) \in L'[x]$ que tiene una raíz en L , tiene todas las raíces en L . Además, en este caso podemos considerar el cociente H'/H . Este cociente es *cíclico* —es decir, es isomorfo al grupo dado por $\{1, \dots, n\}$ bajo la adición módulo n — si y solamente si L es una *extensión cíclica* de L' , es decir existe un elemento $a \in L'$ tal que L es el campo de descomposición del polinomio $x^n - a \in L'[x]$. En otras palabras, todos los elementos de L

se pueden escribir como combinaciones de $1, \sqrt[n]{a}, \sqrt[n]{a^2}, \dots, \sqrt[n]{a^{n-1}}$ con coeficientes en L .¹

Si se junta toda esta información, se obtiene que el hecho de que $\text{Gal}(K : \mathbb{Q})$ tiene una sucesión de composición con factores abelianos corresponde a que las soluciones de la ecuación $f(x) = 0$ se pueden escribir con números racionales, las cuatro operaciones básicas y raíces. El grupo S_n no es soluble, es decir, no tiene una sucesión de composición con factores abelianos. Consecuentemente, un polinomio $f(x)$ cuyo grupo de Galois $\text{Gal}(K : \mathbb{Q})$ es isomorfo a S_n , no será *soluble por radicales*, lo cual significa que sus raíces no se podrán escribir usando los coeficientes, las operaciones básicas y raíces cuadradas. Como hemos visto, hay siempre polinomios con esa propiedad. Así se establece una visión más profunda de la imposibilidad de poder dar una fórmula de solución para las ecuaciones algebraicas de grado mayor que 4.

Es una coincidencia triste que, tanto Abel como Galois, murieran a temprana edad. Niels Henrik Abel nació en Noruega en 1802 pero era muy pobre y sólo gracias a una beca pudo viajar a Europa para presentar sus brillantes ideas en los centros de investigación matemática. Viajó a Berlín y luego a París, pero no obtuvo una recepción favorable ni logró despertar el interés de los líderes matemáticos de aquel tiempo, como Gauss, Legendre o Cauchy. En París contrajo tuberculosis y murió a la edad de 27 años, dos días antes de que llegara la carta de Berlín que le informaba que había conseguido un puesto de profesor.

Figura 4.65: Sello conmemorativo por los 100 años de la muerte de Evariste Galois (1811-1832), matemático francés que murió con tan sólo veinte años en un duelo. Años más tarde, otro matemático alemán —Felix Klein— diría: En Francia apareció hacia 1800 una nueva estrella de inimaginable brillo en el firmamento de las matemáticas: Evariste Galois.



Galois nació en 1811, en Francia, y murió a los 22 años en un duelo en el que había retado a un oficial, por motivos desconocidos. Vivió en tiempos turbios y difíciles. Lo que se sabe de su corta vida habla de que era una persona rebelde. Fue expulsado de la escuela, estuvo en prisión y lideró protestas. La noche anterior al duelo, escribió una última carta donde reportó y resumió frenéticamente sus ideas matemáticas, a sabiendas de que la muerte lo esperaba al principio del día siguiente. A Galois debemos la idea de considerar varias permutaciones de las raíces a la vez y, con ello, toda la teoría de grupos.

Hemos recorrido un camino largo y sinuoso. Primero, necesitábamos delimitar la pregunta y llegamos a la noción de ecuaciones algebraicas. Expandimos la noción de los números reales a los números complejos para poder resolver cualquier ecuación cuadrática. Después, consideramos la ecuación de tercer grado y vimos que, por el teorema fundamental del álgebra, la ecuación algebraica de cualquier grado positivo siempre tiene soluciones. El problema consiste en encontrar una fórmula para las soluciones a partir

¹ Esto da la idea correcta, pero habría que precisarla: sólo es correcta si se cambia el campo \mathbb{Q} por un campo k que contiene \mathbb{Q} y todas las soluciones de $x^n - 1 = 0$. Cabe mencionar que lo anterior no afecta la cuestión de ser soluble por radicales, es decir, de poder escribir una raíz de un polinomio con las operaciones básicas y raíces cuadradas.

de los coeficientes que aparecen en la ecuación. Finalmente, descubrimos que tal fórmula no existe por el teorema de Abel-Ruffini y que la teoría de Galois aporta una explicación profunda de este hecho. Esta teoría requiere de considerable abstracción pero es, indudablemente, una de las joyas del álgebra. Así concluimos el resumen sobre una de las cuatro preguntas fundamentales que consideraremos en este capítulo.

4.10 ¿QUÉ SE PUEDE CONSTRUIR?



Figura 4.66: El compás permite medir distancias y trazar circunferencias. En esta ilustración del Codex Vindobonensis 2554, una Biblia ilustrada, el compás es la herramienta usada por Dios en el acto de creación. Con ello se expresa la importancia del compás como herramienta geométrica y de la geometría, como disciplina del pensamiento.

4.10.1 Delimitación de la pregunta

Los científicos griegos estaban fascinados por la posibilidad de realizar construcciones con las herramientas más básicas: la regla y el compás. La regla sirve simplemente para trazar pedazos de rectas mientras el compás sirve para trazar circunferencias. ¿Por qué en muchos de sus estudios se habrán restringido a estas dos herramientas? ¿Qué tienen en particular la recta y la circunferencia?

La circunferencia con centro C y radio r es el conjunto de puntos que están a distancia r del punto C . Las circunferencias miden así lo que está cerca o lejos de un punto y, en ellas, se encuentra codificada la distancia. En la recta es diferente pues los griegos sólo consideraban pedazos finitos de rectas: *segmentos*. Un segmento marca el camino más corto entre sus dos extremos. ¿Cómo habrá entonces que interpretar la prolongación de un segmento sobre una de sus dos extremidades? La recta marca cómo seguir “la ruta marcada” sin desviarse hacia la derecha o la izquierda.

Sobre superficies curvas como una esfera o una dona ya no podemos hablar de circunferencias o rectas, pero sí podemos hablar del conjunto de puntos que están a una distancia dada de un punto C o del camino más corto entre dos puntos. En el toro nos damos cuenta rápidamente de que la noción del camino “más corto” no siempre se puede prolongar sobre los extremos. Por ello, mejor se habla de las *geodésicas* que son las curvas que tienen la propiedad de que para puntos cercanos sobre ellas, sí marcan los caminos más cortos.

4.10.2 Los problemas clásicos

La circunferencia y la recta son entonces conceptos muy fundamentales en la geometría plana. Alrededor de 300 a.C. se escribió *Los elementos*, un libro de gran influencia, en donde se reunió gran parte de lo que se sabía dentro de un marco común, que es la construcción con regla y compás. La fascinación griega por los problemas geométricos data de muchos años atrás. Algunos de ellos se hicieron tan famosos que ahora se llaman “clásicos” y son los siguientes:

- *Duplicación del cubo: hallar con regla y compás la longitud del lado de un cubo que tiene el doble del volumen de un cubo cuyo lado esté dado.*
- *Trisección del ángulo: encontrar con regla y compás el tercio de un ángulo dado.*
- *Construcción de polígonos regulares: construir con regla y compás la longitud de un lado de un polígono regular con n lados inscrito en una circunferencia dada.*
- *Cuadrar un círculo: hallar con regla y compás la longitud del lado de un cuadrado que tenga la misma área que una circunferencia cuyo radio sea conocido.*

La expresión “hallar con regla y compás” merece una aclaración: lo que se busca es una lista de instrucciones para construir lo indicado al usar solamente la regla —con la cual se puede trazar una recta por dos puntos ya conocidos— y el compás —con el cual se puede construir la circunferencia con centro dado que pasa por un punto ya conocido. Nuevos puntos sólo se obtienen como intersección de dos objetos —rectas o circunferencias— ya construidas. Para hallar la circunferencia inscrita en un triángulo dado, no se pueden dibujar una serie de circunferencias probando con diferentes centros y radios para aproximarse, poco a poco, a la solución; lo que debe indicarse es cómo se obtienen el centro y el radio para cualquier triángulo dado, no sólo uno en particular.

En *Los Elementos* se procede paso a paso al explicar cómo resolver tareas cada vez más complicadas, pero no se resuelve ninguno de los problemas clásicos. Cada uno de estos problemas tiene una historia larga e interesante de muchos intentos fallidos. En la primera mitad del siglo XIX, se demuestra que los primeros tres de estos problemas no tienen solución y en la segunda mitad del mismo siglo, se concluye que tampoco la cuadratura del círculo tiene solución. Para ello, intervinieron matemáticos como Carl Friedrich Gauss, Evariste Galois, Pierre Wantzel y Ferdinand von Lindemann.

Figura 4.67 Pierre Wantzel y Ferdinand von Lindemann.

El adjetivo “que no tiene solución” tiene un carácter definitivo: no hay y no habrá nunca. No es que los matemáticos no han buscado lo suficientemente bien —lo hicieron por más de dos mil años y, a veces, hasta con una dedicación feroz— sino que no puede haber solución si uno se restringe a la regla y al compás como las únicas herramientas de construcción. Desde la época de Euclides se conocen varias soluciones a los primeros tres

problemas si se elimina la restricción de hacerlo sólo con regla y compás; por ejemplo, con una herramienta que traza parábolas. Con la herramienta que se muestra en la figura 4.68 se puede trisecar un ángulo con facilidad.

La herramienta consiste en varas y articulaciones en forma similar al compás. Para trisecar un ángulo $\angle EFG$ se coloca el punto C en F y, luego, se ajustan las dos manecillas A y B sobre la recta FE y FG , respectivamente. Las puntas P y Q indican entonces la ubicación de los rayos que trisecan el ángulo dado.

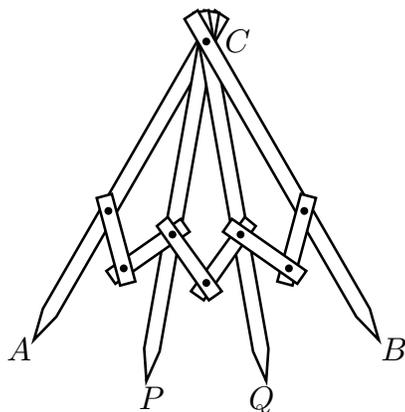


Figura 4.68 Herramienta de construcción para trisecar ángulos.

La dificultad de los problemas clásicos reside en la restricción de las herramientas a, únicamente, regla y compás. En lo que sigue, daremos una idea sobre cómo es posible llegar a la conclusión de que no existe solución a los problemas clásicos. Para ello, usaremos los *números complejos* de la sección 4.2. Cada construcción comienza con unos datos iniciales, como una circunferencia o un triángulo dado. Empezaremos aquí con el mínimo absoluto: a partir de dos puntos dados.

4.10.3 El plano complejo como modo algebraico

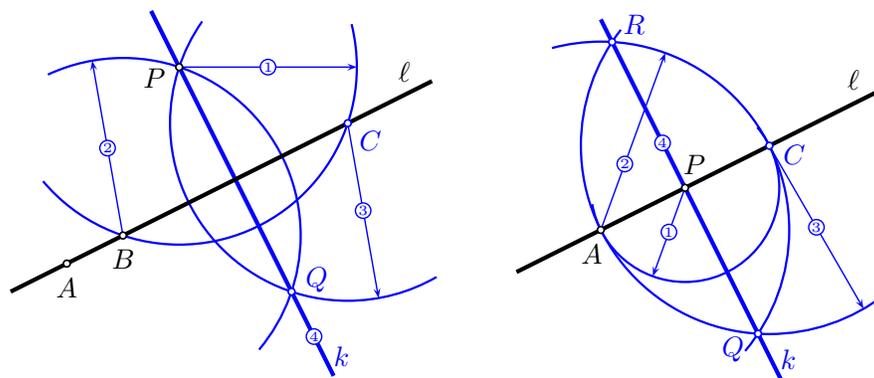
El punto clave es usar, como modelo para el plano, el *plano complejo* véase sección 4.2. Podemos pensar que los dos puntos dados son los números 0 y 1. ¿Qué otros números del plano complejo se pueden construir a partir de estos dos? A continuación, veremos que el conjunto de puntos \mathcal{C} que se puede construir forma —visto cómo números complejos— un *campo*, es decir, los números en \mathcal{C} se pueden sumar, restar, multiplicar y dividir. Antes de ello, observemos que se pueden construir perpendiculares con regla y compás si damos un punto P y una recta ℓ , pues se debe construir la perpendicular k a ℓ que pasa por P .

Como la figura 4.69 sugiere, hay que distinguir el caso en que P pertenece a ℓ del caso en que no pertenece. La figura indica los cuatro pasos en ambos casos.

También podemos construir *paralelas*: la paralela a la recta ℓ que pasa por P se obtiene al construir la perpendicular a k que pasa por P , donde k es la perpendicular a ℓ que pasa por P . Por la interpretación geométrica de la suma de números complejos véase figura 4.2, es posible construir el número $\mathbf{x} + \mathbf{y}$ si ya se tienen construidos \mathbf{x} , \mathbf{y} con anterioridad: ahora sólo necesitamos intersectar la paralela a $0\mathbf{y}$ por \mathbf{x} con la paralela a $0\mathbf{x}$ por \mathbf{y} . En otras palabras, si \mathbf{x} , \mathbf{y} pertenecen a \mathcal{C} , entonces también $\mathbf{x} + \mathbf{y}$ lo hará.

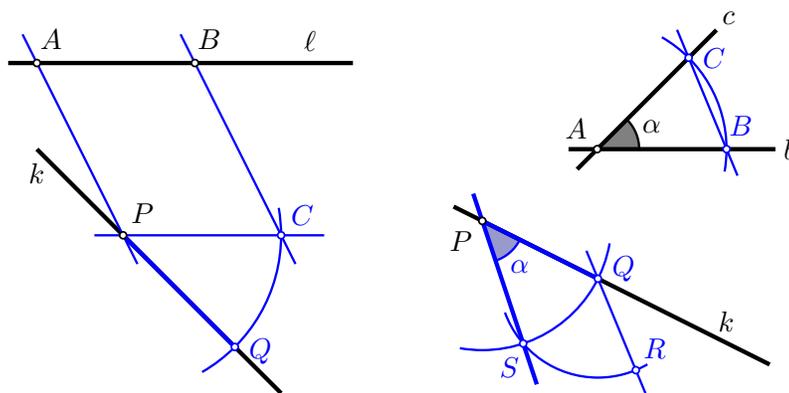
Por otro lado, podemos construir $-\mathbf{x}$ si ya está construido \mathbf{x} , dado que $-\mathbf{x}$ es simplemente el reflejo de \mathbf{x} en el punto 0. Consecuentemente, si \mathbf{x} , \mathbf{y} pertenecen a \mathcal{C} , entonces también $-\mathbf{y}$ y, por lo tanto, también $\mathbf{x} - \mathbf{y} = \mathbf{x} + (-\mathbf{y})$.

Figura 4.69 Indicación de cómo construir perpendiculares. A la izquierda para puntos P fuera y a la derecha, para puntos P sobre la recta ℓ .



Antes de revisar la multiplicación, analicemos el compás como herramienta. Un compás puede levantarse de la hoja de papel y, sin cambiar su apertura, trazar una circunferencia en otro lugar con el mismo radio. Parece que el compás es una herramienta más poderosa de lo que pensábamos —que, simplemente, construir rectas y circunferencias a partir de puntos ya construidos—. Pero no es así: podemos usar solamente rectas y circunferencias para implementar toda la funcionalidad del compás, como lo muestra la figura 4.70 a la izquierda.

En esta figura se *transportó* la distancia entre los puntos A y B a la recta k , partiendo del punto dado P . Como resultado, se obtiene el punto Q sobre k , que está a la misma distancia que B de A . Del lado derecho se muestra cómo se pueden transportar ángulos. Si se quiere transportar el ángulo α entre dos rectas b y c , al punto P a partir de la recta k , entonces, primero se construye el punto Q —al transportar la distancia AC a P —, luego el punto R —al transportar la distancia CB a una paralela a BC por Q — y, finalmente, se interseca la circunferencia con centro P que pasa por Q con la circunferencia con centro Q que pasa por R , para obtener el punto S y $\angle QPS = \angle CAB = \alpha$.

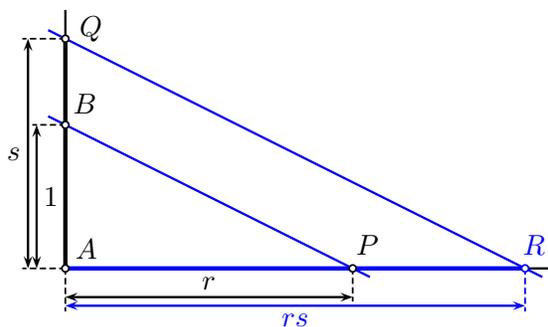


4.70 Indicaciones de cómo transportar distancias y ángulos de un lugar a otro sin el uso del compás.

Claro es que los números enteros son construibles si se transporta, por ejemplo, el segmento s con extremos 0 y 1 a la derecha de 1 , para encontrar el punto 2 . Luego se transporta s a la derecha de 2 , para obtener 3 .

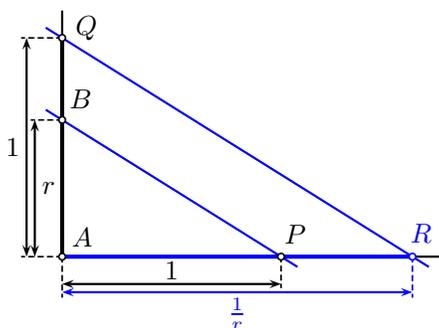
Veamos ahora la multiplicación. Según la interpretación geométrica que se observa en la figura 4.42, debemos sumar los ángulos y multiplicar los valores absolutos. Ambas cosas son bastante fáciles de lograr con regla y compás: los ángulos se suman con el método para transportar explicado arriba y el producto de los valores absolutos se obtiene por semejanza, como en la figura 4.71.

A continuación, veamos la última de las cuatro operaciones básicas: la división. Como en la resta, es suficiente mostrar que podemos construir con regla y compás el inverso



4.71 Construcción para encontrar el producto rs de dos números reales r y s positivos.

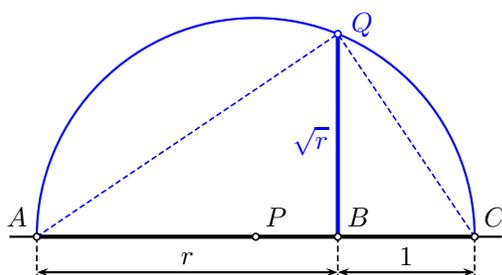
multiplicativo de un número construible \mathbf{z} dado. Con tal fin usamos la representación geométrica: debemos construir el negativo del ángulo de \mathbf{z} y el inverso multiplicativo $\frac{1}{r}$, de su valor absoluto r . Esto se hace de nuevo por semejanza, como lo indica la figura 4.72.



4.72 Construcción para encontrar el inverso multiplicativo $\frac{1}{r}$ de un número real r positivo.

Como consecuencia, podemos construir todos los números racionales, es decir, $\mathbb{Q} \subset \mathcal{C}$, si \mathbb{Q} denota al conjunto de los números racionales.

Por último, observemos que si un número complejo \mathbf{z} es construible, entonces también lo serán sus dos raíces cuadradas $\pm\sqrt{\mathbf{z}}$; este hecho se suele abreviar diciendo que los números construibles son “cerrados bajo la operación de sacar la raíz cuadrada”. Otra vez recurrimos a la interpretación geométrica de sacar raíces, según lo hicimos en la figura 4.58. Debemos dividir el ángulo de \mathbf{z} en dos —que es fácil con regla y compás— y obtener la raíz cuadrada del valor absoluto de \mathbf{z} , como se indica la figura 4.73.



4.73 Construcción para encontrar la raíz cuadrada \sqrt{r} de un número real r positivo.

En efecto, por semejanza de los triángulos ΔABQ y ΔQBC , sigue que la altura $x = BQ$ debe satisfacer $x : r = 1 : x$, es decir, $x^2 = r$ y por ello $x = \sqrt{r}$. Con ello, hemos llegado al primer resultado importante de esta sección:

Los números construibles \mathcal{C} forman un *campo* que es cerrado bajo la operación de sacar raíces cuadradas.

4.10.4 Descripción alterna del campo de los números construibles

Del teorema anterior se deduce que cualquier número complejo que se puede escribir como una expresión que involucre solamente números racionales, las cuatro operaciones básicas y raíces cuadradas es construible. Denotamos por \mathcal{N} al conjunto de los números que se pueden escribir así. Por ejemplo:

$$\frac{8 - 2\sqrt{3}}{\sqrt{-2} - \sqrt{\frac{4}{7}}} - \sqrt{\frac{3}{17} - \frac{1 + 4\sqrt{\frac{171}{3}}}{\sqrt{2} - \sqrt{-2}}}$$

es uno de los números en \mathcal{N} . Hemos demostrado que cada número en \mathcal{N} pertenece a \mathcal{C} , es decir, $\mathcal{N} \subset \mathcal{C}$.

Ahora veamos que cada número construible es, en efecto, un número en \mathcal{N} o lo que es lo mismo, que $\mathcal{C} \subset \mathcal{N}$ y, por ello, $\mathcal{C} = \mathcal{N}$. Demostraremos:

$$\text{si } \mathbf{z} = a + b\mathbf{i} \in \mathcal{C}, \text{ entonces } a, b \in \mathcal{N}. \quad (34)$$

De ahí seguirá que $\mathbf{z} = a + b\sqrt{-1}$ pertenece a \mathcal{N} , es decir $\mathcal{C} \subset \mathcal{N}$.

Para ello, primero observamos que si un número complejo $\mathbf{x} = a + b\mathbf{i}$ es construible, entonces también lo son la parte real a y la parte imaginaria b , y viceversa: si a y b son dos números reales construibles entonces también $a + b\mathbf{i}$ es construible. La parte real y la parte imaginaria de los números complejos sirven como las dos coordenadas en el plano. Lo anterior nos permite elaborar argumentos con técnicas de geometría analítica.

Los números que son construibles se organizan en niveles \mathcal{C}_N . En el primer nivel \mathcal{C}_1 están solamente los dos números iniciales 0 y 1. En el segundo nivel \mathcal{C}_2 se encuentran los números del primer nivel más aquellos otros que se puedan obtener a partir de intersecciones de rectas y circunferencias definidas por números del nivel anterior. Por ejemplo, $\frac{1}{2} + \frac{\sqrt{3}}{2}\mathbf{i}$ es un número del nivel 1 —es una de las dos intersecciones de la circunferencia con centro 0 que pasa por el punto 1 con la circunferencia con centro 1 que pasa por el punto 0—. Inductivamente, los puntos del nivel N son los puntos del nivel $N - 1$ más aquellos puntos que se pueden obtener como intersección de rectas o circunferencias definidas a partir de puntos del nivel anterior.

Como cada punto construible se obtiene por una sucesión finita de pasos de construcción, tenemos que para cada $\mathbf{z} \in \mathcal{C}$ existe un N tal que $\mathbf{z} \in \mathcal{C}_N$. Ahora hemos preparado las bases para una *demonstración por inducción*: demostramos 34 por inducción sobre N , es decir demostramos que:

$$\text{si } \mathbf{z} = a + b\mathbf{i} \in \mathcal{C}_N, \text{ entonces } a, b \in \mathcal{N} \quad (35)$$

para los diferentes N , uno tras otro. El inicio de la inducción es cuando $N = 1$. Como $\mathcal{C}_1 = \{0, 1\}$, tenemos que $\mathcal{C}_1 \subset \mathcal{N}$ y con ello, 35 para $N = 1$.

Ahora, si suponemos que ya demostramos 35 para $N = L - 1$ y argumentamos que (35) también satisface para $N = L$.

Sea $\mathbf{z} = x + y\mathbf{i} \in \mathcal{C}_L$. Entonces, existen dos objetos ℓ y k —rectas o circunferencias—, dadas por puntos en \mathcal{C}_{L-1} : el objeto ℓ está dado por $\mathbf{P} = a + b\mathbf{i}$ y $\mathbf{Q} = c + d\mathbf{i}$ mientras que el objeto k está definido por los puntos $\mathbf{Q}\mathbf{P}' = a' + b'\mathbf{i}$ y $\mathbf{Q}' = c' + d'\mathbf{i}$. Si ℓ es la recta que pasa por los puntos \mathbf{P} y \mathbf{Q} , entonces ℓ se representa por la ecuación:

$$(a - c)y + bc = (b - d)x + ad, \quad (36)$$

y si ℓ es la circunferencia con centro \mathbf{P} que pasa por \mathbf{Q} , entonces ℓ se representa por la ecuación:

$$(x - a)^2 + (y - b)^2 = (c - a)^2 + (d - b)^2. \quad (37)$$

Ecuaciones similares definen al objeto k . Ahora tenemos que distinguir tres casos.

Caso 1: cuando ℓ y k son rectas. Aquí, las ecuaciones que representan ℓ y k definen un sistema de dos ecuaciones con dos incógnitas x y y . La solución está dada por expresiones que involucran los números $a, b, c, d, a', b', c', d'$ y las cuatro operaciones básicas, en concreto:

$$x = \frac{(a' - c')(bc - ad) - (a - c)(b'c' - a'd')}{(a' - c')(b - d) - (a - c)(b' - d')},$$

$$y = \frac{(b' - d')(ad - bc) - (b - d)(a' - d' - b'c')}{(b' - d')(a - c) - (b - d)(a' - c')}.$$

Como $a, b, c, d, a', b', c', d'$ están en \mathcal{N} por hipótesis de inducción —porque $\mathbf{P}, \mathbf{Q}, \mathbf{P}', \mathbf{Q}' \in \mathcal{C}_{L-1}$ —, tenemos que $x, y \in \mathcal{N}$ por definición de los números que pertenecen a \mathcal{N} .

Caso 2: cuando ℓ es recta y k es una circunferencia, similar al caso en que k es recta y ℓ circunferencia. Si $b \neq d$, se puede despejar x de la ecuación 36 y sustituirlo en la ecuación que representa a k . Con ello se obtiene una ecuación cuadrática en y de la forma:

$$Ay^2 + By + C = 0 \quad (38)$$

donde:

$$A = (a - c)^2 + (b - d)^2$$

$$B = 2(a - c)[bc - ad - a'(b - d)] + 2(b - d)^2b'$$

$$C = [bc - ad - a'(b - d)]^2 + (b - d)^2[b'^2 - (c' - a')^2 - (d' - b')^2]$$

Nuevamente, $a, b, c, d, a', b', c', d' \in \mathcal{N}$ por hipótesis de inducción y, con ello, $A, B, C \in \mathcal{N}$ por definición de la forma de los números que pertenecen a \mathcal{N} . Por lo tanto, también la solución de 4.38:

$$y = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \quad (39)$$

pertenece a \mathcal{N} . La sustitución de 39 da una ecuación cuadrática en x con coeficientes en \mathcal{N} . Consecuentemente, x pertenece a \mathcal{N} .

Si $b = d$, entonces se puede despejar y de la ecuación 36 y se obtiene que:

$$y = \frac{ad - bc}{a - c} = b,$$

que está en \mathcal{N} por hipótesis de inducción. La sustitución en la ecuación que representa k da:

$$(x - a')^2 - (b - b')^2 = (c' - a')^2 + (d' - b')^2,$$

que es cuadrática en x con coeficientes en \mathcal{N} . Por lo tanto, las soluciones están en \mathcal{N} .

Caso 3: cuando ℓ y k son *circunferencias*. En vez de trabajar con el sistema de las dos ecuaciones que representan ℓ y k , se trabaja con el sistema de ℓ y m donde m es la diferencia entre ℓ y k :

$$\begin{aligned} m &: (x - a)^2 - (x - a')^2 + (y - b)^2 - (y - b')^2 = (c - a)^2 - (c' - a')^2 + (d - b)^2 - (d' - b')^2, \\ m &: (2a - 2a')x + (2b - 2b')y + (a^2 - a'^2 + b^2 - b'^2) = (c - a)^2 - (c' - a')^2 + (d - b)^2 - (d' - b')^2 \end{aligned}$$

La segunda línea muestra la ecuación después de expandir y simplificar el lado izquierdo. Se ve que es lineal en las variables x, y por lo que se puede proceder como en el caso 2. La ecuación m representa la recta que pasa por los dos puntos de intersección de ℓ y k , o la tangente común si ℓ y k son tangentes en Z .

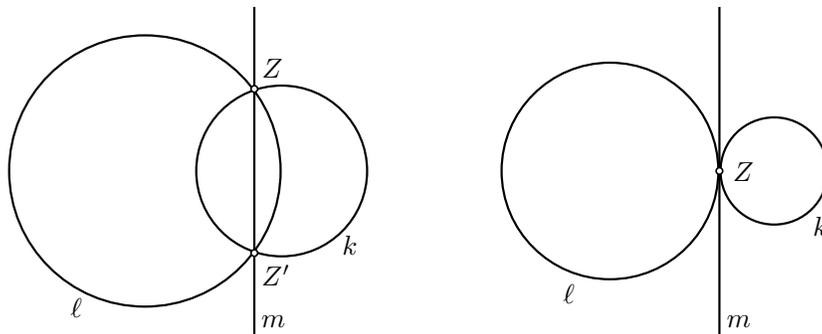


Figura 4.74 Definición de la recta m de dos circunferencias ℓ y k que se intersecan.

Con todo lo anterior, demostramos el segundo resultado importante: los números construibles son precisamente aquellos que se pueden escribir a partir de números racionales usando un número finito de las cuatro operaciones básicas y raíces cuadradas.

4.10.5 Sobre la imposibilidad de resolver los problemas clásicos

Con los dos resultados anteriores demostrados, tenemos las bases para indicar los razonamientos para deducir que los cuatro problemas clásicos no se pueden resolver.

En la descripción dada del Teorema 4.10, vemos que cualquier número construible es un *número algebraico*, es decir, un número que es solución de alguna *ecuación algebraica* con coeficientes racionales como se vio en la sección 4.9.

En 1882, Ferdinand von Lindemann demostró que el número π es *transcendental*, o lo que es lo mismo, que no es solución de ninguna ecuación algebraica con coeficientes racionales. Si se quiere cuadrar el círculo, lo que se busca es el encontrar la longitud del lado de un cuadrado que tiene la misma área que la de un círculo dado, por ejemplo, el círculo con radio 1, que posee un área de π , como se vio en la sección 2.9. Por lo tanto, se busca construir la longitud $\sqrt{\pi}$. Si pudiéramos construir esta longitud, entonces también podríamos hacerlo con su cuadrado, es decir π , lo que es imposible por el Teorema de Lindemann. Así se muestra que no es posible cuadrar el círculo.

La imposibilidad de cuadrar el círculo es la última de las cuatro demostraciones, pero la indicamos primero dado que, conceptualmente, es la más clara. Veamos ahora la duplicación del cubo.

Si tenemos un cubo dado con aristas de longitud ℓ , entonces el volumen es de ℓ^3 . El cubo que tiene el doble de volumen presenta aristas de longitud $\sqrt[3]{2}\ell$. Para hallar esta longitud con regla y compás, debemos construir $\sqrt[3]{2}$ dado que podemos multiplicar. Sin embargo, según el Teorema 4.10 sólo podemos construir raíces cuadradas, no cúbicas. Pero esto no es suficiente: ¿por qué no es posible que alguna expresión que sólo involucra raíces cuadradas sea igual a $\sqrt[3]{2}$? Para resolver lo anterior, necesitamos algunas herramientas de la teoría de *campos*.

En la sección 4.9 consideramos campos intermedios como $\mathbb{Q} \subset K \subset \mathbb{C}$, es decir, campos K contenidos en los números complejos \mathbb{C} y que también, contienen los números racionales \mathbb{Q} . Un campo así surgió como *campo de descomposición* de algún polinomio $f(x)$ y siempre tiene una *dimensión sobre \mathbb{Q}* . La dimensión de campos generaliza la idea del “número de coordenadas que se usan para describir una ubicación”: en la recta basta una coordenada, en el plano se requieren dos y en el espacio tres. Para usar coordenadas se requiere de un origen y de ejes.

Por ejemplo, en los números complejos tenemos dos ejes: el real y el imaginario. Cualquier número complejo $a + bi$ se expresa con dos coordenadas reales a y b . En otras palabras, cualquier número complejo es combinación con coeficientes reales de los números complejos 1 e i .

Si consideramos ahora una situación de un campo de descomposición K sobre \mathbb{Q} , entonces fijamos $k_1, \dots, k_d \in K$ y consideramos *combinaciones lineales*, como en:

$$q_1 k_1 + q_2 k_2 + \dots + q_n k_n, \quad (40)$$

con coeficientes $q_1, \dots, q_d \in \mathbb{Q}$. La *dimensión* de K sobre \mathbb{Q} , es entonces el mínimo número d tal que existen elementos $k_1, \dots, k_d \in K$ llamados *base*, con la propiedad de que cualquier elemento de K es combinación lineal 40 para algunos coeficientes q_1, \dots, q_d . Esta dimensión se denota por $\dim_{\mathbb{Q}} K$. Se demuestra que cualesquiera dos bases, siempre tienen la misma cardinalidad. Además, si $\mathbb{Q} \subset L \subset K$ entonces:

$$\dim_{\mathbb{Q}} K = \dim_{\mathbb{Q}} L \cdot \dim_L K \quad (41)$$

Una consecuencia del Teorema 4.10 es que la dimensión sobre \mathbb{Q} de cualquier campo K que consiste en números construibles es una potencia de 2, dado que en cada paso, al aumentar una raíz cuadrada, se aumenta por un factor de 2. El campo de descomposición L de $f(x) = x^3 - 2$, que contiene la raíz $\sqrt[3]{2}$, tiene dimensión 3 sobre \mathbb{Q} , por lo tanto, una base es:

$$1, \quad \sqrt[3]{2}, \quad \sqrt[3]{2}^2 = \sqrt[3]{4}, \quad (42)$$

y, por lo tanto, cualquier base tiene cardinalidad 3. Si suponemos que $\sqrt[3]{2}$ es un número construible, entonces este número pertenece a un campo K que tiene dimensión 2^t sobre \mathbb{Q} para algún t . Pero si K contiene $\sqrt[3]{2}$, entonces también la contiene su cuadrado y, por lo tanto, la base 42. Por consiguiente, tenemos ahora las siguientes inclusiones de campos $\mathbb{Q} \subset L \subset K$, pero $\dim_{\mathbb{Q}} L = 3$ y $\dim_{\mathbb{Q}} K = 2^t$, lo que contradice la igualdad 41.

Todo lo anterior muestra que $\sqrt[3]{2}$ no es construible y, por lo tanto, no es posible

construir sólo con regla y compás el lado de un cubo que tenga el doble de volumen de un cubo dado.

Un argumento muy similar muestra que los otros dos problemas clásicos tampoco tienen solución con regla y compás: si queremos trisecar un ángulo α conocido debemos construir $\cos(\frac{\alpha}{3})$ o $\sin(\frac{\alpha}{3})$. Para ello, requerimos de una identidad trigonométrica que se deriva de la fórmula de Euler:

$$e^{\varphi i} = \cos(\varphi) + \sin(\varphi)i \quad (43)$$

donde e es el *número de Euler* —otro número trascendente— cuyo valor aproximado es 2.7182818284. Al elevar 43 al cubo, se obtiene:

$$\cos(3\varphi) + \sin(3\varphi)i = (\cos^3(\varphi) - 3\cos(\varphi)\sin^2(\varphi)) + (3\cos^2(\varphi)\sin(\varphi) - \sin^3(\varphi))i.$$

Si comparamos las coordenadas, tenemos, después de sustituir $\sin^2(\varphi) = 1 - \cos^2(\varphi)$ y $\cos^2(\varphi) = 1 - \sin^2(\varphi)$, respectivamente, las siguientes identidades para el triple de ángulo:

$$\cos(3\varphi) = 4\cos^3(\varphi) - 3\cos(\varphi) \quad (44)$$

$$\sin(3\varphi) = -4\sin^3(\varphi) + 3\sin(\varphi) \quad (45)$$

Podemos usar estas fórmulas para calcular $\cos(\frac{\alpha}{3})$, si sustituimos $\varphi = \frac{\alpha}{3}$ en 44:

OJO, FALTA FÓRMULA (4.46)

Dado que el ángulo α ya está construido, podemos hallar $c = \cos(\alpha) = \cos(3\varphi)$ con regla y compás, y debemos resolver la ecuación 46 en la incógnita $x = \cos(\frac{\alpha}{3})$, es decir, tenemos que resolver la ecuación cúbica:

$$x^3 - \frac{3}{4}x - \frac{c}{4} = 0$$

En la sección 4.2 calculamos la solución y vimos que había que calcular raíces cúbicas. Si no se trata de ángulos muy particulares, $\cos(\frac{\alpha}{3})$ no será construible. Por ejemplo, $\cos(20^\circ)$ no es construible.

Por ello, el problema de la trisección de ángulos no se puede resolver con regla y compás. Por la misma razón, tampoco se pueden construir polígonos con cualquier número de lados pues, por ejemplo, para construir un polígono con 9 lados, tendríamos que construir primero el $\cos(20^\circ)$, que es imposible como acabamos de ver.

Así concluimos el recuento de una larga historia de búsqueda por la solución de algo que nunca fue posible. Los problemas clásicos no tienen solución mientras se mantenga la restricción de que las herramientas de construcción sean regla y compás, únicamente. Para comprender la imposibilidad, es necesario salir del ambiente de las construcciones: es necesario formular qué quiere decir el construir con regla y compás en un lenguaje que permite una vista distinta y que, en nuestro caso, fue el lenguaje algebraico con el uso de las herramientas de la teoría de campos. Ver que algo es imposible requiere de cierta distancia y reflexión y, en esta situación particular, una considerable cantidad de abstracción. Por ello, no debe sorprender que muchos aficionados han seguido y siguen intentando resolver alguno de los problemas y presentan construcciones ingeniosas para cuadrar el círculo, todas ellas falsas como ahora sabemos.

4.11 ¿QUÉ SE PUEDE DEMOSTRAR?



Figura 4.75 Las tres letras Q.E.D. son la abreviatura del latín “quod erat demonstrandum”, que se traduce como “lo que se quería demostrar”. Con ellas se indica el fin de una demostración. Hoy pueden sustituirse también por un cuadrado “□”, cuyo uso lo inició Paul Halmos. Las demostraciones juegan un papel central en las matemáticas. Con ellas se establece certidumbre y, a la vez, sirven para comunicar las matemáticas.

4.11.1 El sistema axiomático

En esta sección queremos indagar un poco sobre las siguientes preguntas: ¿en qué consiste una demostración?, ¿cómo se demuestra algo?, ¿cómo se establece la veracidad de una afirmación? y, finalmente, ¿se puede demostrar todo lo que es verdadero?

Lo que hoy se llama “el Teorema de Pitágoras” se conoció más de 1 000 años antes de Pitágoras, que vivió en el siglo v a.C. A diferencia de las culturas anteriores, los griegos no se dieron por satisfechos con la validez en ejemplos aislados: percibieron la generalidad y la necesidad de una explicación, argumentación o *demostración* general.

Al principio, sólo fueron resultados sueltos como el Teorema de Tales, los teoremas de triángulos isósceles y el Teorema de Pitágoras. Pero, alrededor de 300 a.C., Euclides escribió un libro realmente extraordinario conocido como *Los Elementos*, donde reunió en 13 tomos gran parte de lo que se sabía en aquel entonces de matemáticas. El logro de Euclides no fue la tarea de simplemente reunirlos y juntarlos, sino de idear una forma de organización sin precedente que hoy se llama *sistema axiomático*.

Figura 4.76 Pitágoras o Tales de Mileto.

La idea de un sistema axiomático consiste en dar al principio los *axiomas*, es decir, aquellas afirmaciones que se aceptan como verdaderas sin cuestionar y después derivar, mediante deducción, todas las demás afirmaciones. Los axiomas deben ser elegidos con cuidado. En particular, se debe evitar que alguno de los axiomas se pueda deducir a partir

de los otros. Más importante aún, los axiomas deben ser *consistentes*, es decir, no debe ser posible deducirlos a partir de una afirmación y a la vez, también de su negación.

Los axiomas formulados en *Los elementos* se dividen en dos grupos. Primero, vienen cinco “Postulados”:

1. Por dos puntos diferentes sólo se puede trazar una línea recta.
2. Todo segmento rectilíneo se puede prolongar indefinidamente.
3. Con un centro y un radio dados sólo se puede trazar una circunferencia.
4. Todos los ángulos rectos son iguales.
5. Si una recta corta en forma transversal a otras dos formando ángulos internos del mismo lado de la transversal cuya suma es menor que 180 grados, entonces las dos rectas prolongadas lo suficiente se intersectarán de ese mismo lado.

Después, siguen cinco “Nociones comunes”. La primera de ellas dice que “Cosas iguales a una tercera son iguales entre sí”, mientras que la quinta dice “El todo es mayor que la parte”.

Se ve que los axiomas que se enuncian como “nociones comunes” mencionan afirmaciones que, posiblemente, nunca reflexionaremos tan detenidamente mientras que las afirmaciones que se enuncian como postulados tienen un carácter más específico, dado que tratan de la geometría en particular.

Estos axiomas son precedidos por las 23 “definiciones” que tratan de aclarar qué se entiende por “punto” o “recta”. Por ejemplo, la primera definición dice que “un punto es lo que no tiene partes” y la segunda, “una línea es una longitud sin anchura”. Desde un punto de vista moderno, estas definiciones sobran. Son los axiomas los que deben aclarar la relación que tienen los objetos entre sí. Es decir, los objetos como “punto” y “recta” se definen por sus propiedades, enunciadas en los mismos axiomas.

Veamos ahora la primera *proposición* que dice que se puede construir un triángulo equilátero sobre cualquier segmento dado. En seguida se da la demostración:

Sea AB el segmento dado. [P.1]

Así pues, se ha de dibujar sobre la recta AB un triángulo equilátero.

Dibujar el círculo BCD , con centro A y radio AB . [P.3]

Dibujar también el círculo ACE , con centro B y radio AB . [P.3]

A partir del punto C , que es intersección de los dos círculos, dibujar el segmento CA . [P. 1]

A partir del punto C , que es intersección de los dos círculos, dibujar el segmento CE . [P. 1]

Dado que el punto A es el centro del círculo BCD , AC es igual a AB . [D.15]

Dado que el punto B es, a la vez, el centro del círculo ACE , BC es igual a AB . [D. 15]

Como BC es igual a BA y AC es igual a BA , entonces BC es igual a AC . Ahora bien, las cosas iguales a una misma cosa son también iguales entre sí, por lo tanto, CA es también igual a CB y entonces CA , AB y BC son iguales entre sí. [N.C.1]

Por lo tanto, el triángulo ABC es equilátero y ha sido construido sobre el segmento dado AB .

En cada paso se indica la definición, el axioma o la noción común que se usa. Esto

refleja la honradez al argumentar: se trata de no suponer nada, de explicar paso a paso la razón de cada argumento para transitar de las premisas —el segmento inicial dado— a la conclusión —el triángulo equilátero sobre este segmento inicial.

Con *Los elementos* se estableció la forma de trabajo de las matemáticas como disciplina donde se deducen las afirmaciones, en forma similar a un gran edificio que se construye poco a poco. Es por ello por lo que *Los elementos* fue el gran modelo a seguir. Por ejemplo, en 1537 se imprimió el libro *La Nova Scientia* de Niccolò Tartaglia sobre la balística, redactado en el mismo estilo. Esto nos muestra la gran aceptación que tuvo el método de Euclides. Sin embargo, Tartaglia determina como *Suposición II* —del segundo libro— el axioma de que un cuerpo que se mueve fuera de la perpendicular describe una trayectoria primero rectilínea y después, curva —como arco de circunferencia—, aunque este axioma no corresponde a la realidad, según lo que Galilei demostraría —tiempo después— acerca de las trayectorias parabólicas.

En consecuencia, queda claro que la elección de los axiomas es crucial. Pero no sólo eso: también a la hora de argumentar puede ser —no obstante lo mucho que se trató de evitar— que se usen algunas intuiciones que no se estipularon previamente. Por ejemplo, en la demostración del primer postulado de Euclides se asume la existencia de la intersección de dos circunferencias, resultando un punto de intersección C . ¿Cómo se sabe que tal punto existe? Actualmente, se sabe que aquí hay un problema en la argumentación y que se requiere del *Teorema de curvas de Jordan* para tener la certeza de que tal C exista.

Esto muestra que los estándares de rigor en las demostraciones matemáticas han cambiado con el tiempo. No obstante los puntos débiles que se encontraron en *Los elementos*, este libro marcó un camino a seguir de honestidad en la argumentación. No fue sino hasta el principio del siglo xx cuando se dio otro avance sustancial en el rigor matemático.

4.11.2 La teoría de conjuntos como base para las matemáticas

Figura 4.77 Galilei.

Galilei se dio cuenta de que la quinta noción común, “El todo es mayor que la parte”, no era siempre correcta cuando se trataba de una infinidad de cosas. En 1638, se imprimió su libro *Discorsi* en el cual explica que para cada número natural hay un cuadrado perfecto —el cuadrado de un entero como 4, 9 o 100— y viceversa, para cada cuadrado perfecto obtenemos un número natural al sacar la raíz. Galilei concluye que hay, por consiguiente, tantos números naturales como cuadrados perfectos, no obstante que los cuadrados perfectos son sólo una parte de todos los números naturales.

Con ello, Galilei usó una noción para comparar el “tamaño” de dos conjuntos, lo que hoy se llama *correspondencia uno a uno* o *biyección* y que se introduce, por primera vez, en 1903. Galilei no persigue más esta idea y lo que encontró ha quedado simplemente como una de las paradojas del infinito. Fue Georg Cantor quien enfrentó estas paradojas de frente, al estudiar los *conjuntos* formalmente y para ello, usó los *números trascendentes*

(véase también sección 4.9). Los conjuntos no tuvieron una fácil aceptación al principio; en particular, Leopold Kronecker se oponía fervientemente.

Al principio, un conjunto era simplemente una colección de cosas. Lo único que se pedía era que, de cada “objeto”, siempre se podía decidir si pertenecía o no a dicho conjunto. Relativamente pronto se dieron cuenta de que la noción, así de general, producía paradojas a su vez, como la que hoy se llama *paradoja de Russell*. Para este caso se considera E como el conjunto de todos los conjuntos que no se contienen como elemento y que, al usar el símbolo de pertenencia “ \in ”, se podría escribir:

$$E = \{X \mid X \notin X\}.$$

La paradoja surge entonces al preguntar si $E \in E$ o $E \notin E$. En el primer caso, E es un conjunto X con la propiedad $X \notin X$, una contradicción con $E \in E$. En el segundo caso, E satisface la condición $X \notin X$ y, por lo tanto, $E \in E$, lo cual es otra vez una contradicción.

Figura 4.78
Zermelo y Fraenkel.

Fue necesario un trabajo adicional de Ernst Zermelo y Abraham Fraenkel, matemáticos alemanes en las primeras dos décadas del siglo xx, para salvar la idea de los conjuntos y redefinirlos en forma completamente nueva. Además, Hilbert —quien era “formalista” y apreciaba la precisión y el rigor que surgió de la teoría de conjuntos a la matemática— acuñó la frase: “Nadie debe poder expulsarnos del paraíso que nos creó Cantor”.

Durante la primera mitad del siglo xx tuvo lugar una gran formalización en la matemática que, además, se extendió prácticamente por toda la disciplina. Con este espíritu debe verse también la obra *Principia Mathematica*, de Alfred North Whitehead y Bertrand Russell. En este libro, los autores querían establecer de una vez por todas, las bases axiomáticas para todas las matemáticas. Los autores publicaron tres tomos cubriendo “sólo” teoría de conjuntos, números cardinales, números ordinales y números reales; para ello, necesitaron aproximadamente 2 000 páginas. El cuarto tomo sobre geometría ya no lo escribieron por agotamiento.

Hoy día, todas las matemáticas se basan —de manera formal— en la teoría de conjuntos, el sistema de axiomas *ZFC*, que quiere decir el sistema de axiomas de Zermelo-Fraenkel, incluyendo el axioma de elección.

4.11.3 El programa de Hilbert

En 1920, David Hilbert propuso un ambicioso programa para subsanar los problemas de los cimientos de las matemáticas, que hoy se conoce, simplemente, como el *programa de Hilbert*.



Figura 4.79
David Hilbert en 1912.

Los puntos principales del programa de Hilbert eran:

1. Una formalización de todas las matemáticas, es decir, formular todas las expresiones matemáticas en un lenguaje formal y manipularlas de acuerdo con reglas fijas y bien definidas.
2. Demostrar que el sistema es consistente, es decir, que no es posible encontrar contradicciones dentro de él.
3. Demostrar que el sistema es completo, lo cual quiere decir que cada afirmación matemática que sea correcta puede ser demostrada dentro del mismo sistema.
4. Demostrar que existe un algoritmo para decidir la veracidad o falsedad de cualquier afirmación matemática.

Como veremos en esta sección, el programa de Hilbert fracasó: era demasiado ambicioso. No es que fracasara porque no hubiera suficientes esfuerzos vertidos en él, sino porque había obstáculos efinitivos para lograrlo. Lo que Hilbert quería, no se puede hacer pero para comprenderlo más a fondo, es necesario entender mejor el primer punto: la formalización en sí.

Las matemáticas se escriben con símbolos que, en ocasiones, son muy raros. Por ejemplo, se suele usar el símbolo \vee para la conjunción lógica “o”. La cadena de símbolos $p \vee q$ expresa entonces “ p o q ”, donde p y q son dos afirmaciones cualesquiera. En forma similar, se usa \wedge para expresar la conjunción “y”, el símbolo \neg para expresar “no”. Por ejemplo:

$$(p \wedge q) \vee \neg q \quad (47)$$

quiere decir “se tiene p y q o se tiene no q ”. Esto tiene el mismo significado que “si p entonces q ”, que se suele escribir como $p \Rightarrow q$. No se puede saber si $p \Rightarrow q$ es verdadero o falso, sin conocer la validez de p y q . Si p y q son verdaderos, entonces $p \wedge q$ también es verdadero, mientras $\neg q$ es falso. En consecuencia, 47 es verdadero.

La expresión $p \wedge q \Rightarrow p$ siempre es verdadera, sin importar la validez de p y q . Por lo tanto, si sabemos que $p \wedge q$ es verdadero, podemos sustituir esta expresión por p y obtenemos una nueva cadena que es verdadera. Similarmente, podemos sustituir $p \wedge q$ por $q \wedge p$, es decir, $p \wedge q \Rightarrow q \wedge p$ es siempre verdadera.

Un sistema formal consiste en varias partes:

- Un conjunto finito de símbolos, llamado *alfabeto*.
- Un conjunto —usualmente infinito— de fórmulas bien definidas. Cada una de estas

fórmulas es una cadena de símbolos del alfabeto. Se espera que exista un procedimiento para decidir si una cadena de símbolos está bien definida o no.

- Un conjunto finito de *axiomas*. Cada axioma debe ser una fórmula bien definida.
- Unas *reglas de inferencia* que explican las maneras permitidas para obtener fórmulas bien definidas a partir de otras.

Un *sistema lógico* asigna valores de “verdadero” o “falso” a cada fórmula bien definida.

Hagamos un ejemplo sencillo. El alfabeto es $A = \{1, +, =\}$. Una cadena de símbolos está bien formada si contiene exactamente un símbolo $=$, además, cada símbolo $+$ y cada símbolo $=$ está encerrado entre dos símbolos 1. Por ejemplo:

$$111 + 11 + 1 = 111111 \quad (48)$$

es una fórmula bien definida. Hay un único axioma que es $1 = 1$. Las reglas de inferencia son las siguientes, las variables p, q, r que aparecen son cadenas en los símbolos 1, $+$ tal que cada $+$ está encerrado entre dos 1:

1. Si $pq = r$, entonces $p + q = r$.
2. Si $p = q$, entonces $p1 = q1$.
3. Si $p = q$, entonces $q = p$.

Si queremos podemos interpretar este sistema formal como el de la adición de números enteros positivos representados en el sistema unitario, es decir, simplemente indicando el número por la cantidad de símbolos 1. Habría que interpretar la ecuación 48 como $3 + 2 + 1 = 6$. Con estas reglas de inferencia es posible deducir cualquier expresión matemáticamente correcta entre dos sumas o números enteros positivos. Las cadenas bien formadas que no se pueden deducir son *falsas* si las interpretamos en este sentido y las que sí se pueden deducir son *verdaderas*.

Del axioma $1 = 1$ derivamos $111111 = 111111$, al usar varias veces la segunda regla de inferencia. Con la primera regla de inferencia, $p = 111$ y $q = 111$, obtenemos $111 + 111 = 111111$ y, si la aplicamos nuevamente con $p = 111 + 11$ y $q = 1$, obtenemos 48. Formalmente, una *demonstración* es una sucesión de fórmulas bien formadas, donde se indica en cada paso si es axioma o cómo se obtiene de una fórmula bien definida anterior. Nuestro ejemplo es sencillo, la demostración es:

$$1 = 1 \xrightarrow{I.2} 11 = 11 \xrightarrow{I.2} 111 = 111 \xrightarrow{I.2} 1111 = 1111 \xrightarrow{I.2} 111111 = 111111 \xrightarrow{I.2} 111111 = 111111 \xrightarrow{I.2} 111111 = 111111 \xrightarrow{I.1} 111 + 111 = 111111 \xrightarrow{I.1} 111 + 11 + 1 = 111111$$

4.11.4 El Teorema de Gödel

Las nociones de sistemas formales nos dan herramientas para hablar de lo que se puede demostrar dentro de un sistema de axiomas dado. Un sistema así se llama *completo* si, para cada fórmula bien formada f se puede demostrar f o $\neg f$ usando sólo las reglas de inferencia. Es posible *a priori* o al menos pensable que existen sistemas donde hay fórmulas bien formadas f de las cuales ni f ni tampoco $\neg f$ se pueden derivar usando nada más que las reglas de inferencia. En este caso f es *independiente* de los axiomas dados y es posible aumentar el conjunto de axiomas por f o por $\neg f$.

Hay varios casos conocidos de axiomas que son independientes de otros como, por ejemplo, el quinto postulado de Euclides mencionado anteriormente, que es independiente de los otros cuatro postulados. Si se sustituye por su negación, se obtienen las *geo-*

metrias no-euclidianas. También la hipótesis del continuo es independiente de los axiomas ZFC; esta hipótesis afirma que cada subconjunto de \mathbb{R} , el conjunto de los números reales, tiene la misma cardinalidad que \mathbb{R} o que \mathbb{Q} , pero no hay nada intermedio.

Lo que pedía Hilbert en su programa era dar un conjunto de axiomas que establecieran las bases para todas las matemáticas, como se hacía en el libro *Principia Mathematica* para la aritmética y luego, demostrar que este sistema de axiomas era completo. Pero eso no es posible, según lo comprobó Kurt Gödel, un matemático que nació en 1906 en Austria y murió en 1978 en Estados Unidos.

El resultado que demostró se llama hoy el *Teorema de Gödel* y afirma que un sistema axiomático consistente y “suficientemente poderoso para describir los números enteros” es necesariamente incompleto. Consecuentemente, en cualquier sistema formal que incluya la aritmética como se hace en *Principia Mathematica*, siempre habrá fórmulas bien definidas que no se puedan derivar ni tampoco su contrario. En otras palabras, siempre habrá afirmaciones en las que no podremos decidir si son verdaderas ni falsas.

La idea central de Gödel consiste en codificar cualquier fórmula bien formada como un número, lo cual se puede hacer de varias maneras. Lo importante es que se puede recuperar sin problema la fórmula a partir del número. La siguiente tabla muestra posibles códigos para los símbolos de un alfabeto sencillo.

Símbolo	Código	Significado
0	1	cero
S	2	función sucesora, (SS0 significa el número 2)
=	3	relación de igualdad
<	4	relación de orden
+	5	operación de adición
×	6	operación de multiplicación
(7	paréntesis abre
)	8	paréntesis cierra
∨	9	operador lógico “o”
∧	10	operador lógico “y”
¬	11	operador lógico “no”
x	12	una variable
'	13	apóstrofe (se usa para crear más variables: x', x'', ...)
∀	14	cuantificador “para cada”
∃	15	cuantificador “existe”



Figura 4.80 Posibles códigos para los símbolos de un alfabeto sencillo.

La codificación de una fórmula con k símbolos se hace con los primeros k números primos $2, 3, 5, \dots, p_k$ —aquí usamos p_i para denotar al i -ésimo primo. El primo p_i se eleva a la potencia c_i , si c_i es el código del i -ésimo símbolo. Por ejemplo, la fórmula:

$$\forall x \forall x' (x + x' = x' + x),$$

tiene 16 símbolos y se codifica en el siguiente número:

$$2^{14} \cdot 3^{12} \cdot 5^{14} \cdot 7^{12} \cdot 11^{13} \cdot 13^7 \cdot 17^{12} \cdot 19^5 \cdot 23^{12} \cdot 29^{13} \cdot 31^3 \cdot 37^{12} \cdot 41^{14} \cdot 43^5 \cdot 47^{12} \cdot 53^8,$$

Para recuperar la fórmula a partir del número, hay que hacer la descomposición en factores primos y contar cuántas veces aparece cada uno de ellos. Esto puede ser muy tedioso de hacer en ejemplos concretos; el número de arriba es:

99 729 218 036 602 955 429 934 818 571 840 953 217 751 753 892 397 683 751 686
555 355 842 609 377 305 551 005 904 156 314 767 308 829 381 973 185 267 401 685
826 988 220 470 759 865 745 426 481 243 147 842 988 402 164 767 156 563 454 278
100 000 000 000 000.

¡Un número con 194 cifras! Pero aquí no nos interesa cuán práctico es sino, simplemente, que es posible.

También las demostraciones dentro del sistema formal se pueden expresar como número. Recordemos que una demostración es una cadena de fórmulas bien formadas como:

$$F_1, F_2, \dots, F_l \quad (49)$$

donde las primeras fórmulas son axiomas y cada fórmula siguiente es una consecuencia de las anteriores, al aplicar alguna de las reglas de deducción. La última, es decir F_l , es la fórmula que se demostró. Usando los códigos de Gödel para cada una de las fórmulas, obtenemos una cadena de números:

$$m_1, m_2, \dots, m_l.$$

El código asociado a la demostración 49 se forma al elevar los primeros primos a las potencias m_1, \dots, m_l :

$$3^{m_1} \cdot 5^{m_2} \cdot \dots \cdot p_{l+1}^{m_l},$$

es decir, se forma muy parecido al código de las fórmulas bien formadas con el único cambio de que se empieza con el primo $p_2 = 3$ y no con $p_1 = 2$. Este pequeño cambio tiene la ventaja de poder distinguir, inequívocamente, entre los códigos de fórmulas y los códigos de demostraciones: los primeros siempre serán pares y los segundos, impares.

Por ello, ahora podemos considerar la siguiente función:

OJO, ESTA FÓRMULA ESTÁ INCOMPLETA

Lo importante de esta función es que es posible evaluarla en la práctica: si el número m no corresponde al código de fórmulas bien formada, el valor de $\text{DEM}(d, m)$ es cero y lo mismo vale si d no corresponde a una sucesión de fórmulas bien formadas. En caso contrario, d corresponde a F_1, \dots, F_l y m a una fórmula F . Si $F \neq F_l$, entonces nuevamente $\text{DEM}(d, m) = 0$. Y también es posible verificar si cada F_i es axioma o consecuencia de F_1, \dots, F_{i-1} .

Lo sorprendente es que encontramos una función de dos números naturales que nos indica si el primer número corresponde a una demostración de lo que corresponde al segundo número. Además, esta función se puede expresar en nuestro sistema formal, es decir, la expresión:

$$\forall x(\text{DEM}(x, m) = 0) \tag{50}$$

se puede escribir explícitamente con los símbolos dados en la tabla de arriba. Lo escribimos para mejor legibilidad aunque muy comprimido: tan sólo el número m se tendría que escribir como:

$$\underbrace{SS \cdots S0}_{m \text{ veces}} \tag{51}$$

y DEM se expandirá a una expresión muy grande. Con ello, hemos expresado dentro de nuestro sistema el hecho de que una fórmula bien formada F con código m no es demostrable pues no existe código de Gödel que represente una sucesión de fórmulas bien definidas.

Sin embargo, en 50 entró en juego el número específico m . Por ello, se considera mejor la fórmula bien definida:

$$\forall x(\text{DEM}(x, x') = 0) \tag{52}$$

que tiene una variable libre, a saber x' . Esta variable la podemos sustituir por diferentes valores para obtener diferentes fórmulas bien definidas.

Es crucial que la sustitución de una variable libre, como x' , en una fórmula pueda hacerse de manera mecánica con los códigos de Gödel. Es decir, la aplicación que asigna al código de 52 el código de 50 es, a la vez, algo que se puede codificar dentro del sistema formal. Será un procedimiento complicado: habrá que ubicar todos los primos p_i dentro del código de 52 que ocurren con una multiplicidad de 12 —el código de x — tal que la multiplicidad de p_{i+1} sea 13 —el código del apóstrofe— y la de p_{i+2} no sea 13 y, luego, recorrer los primos para poder insertar en lugar de los dos símbolos x' , los $m + 1$ símbolos $SS \cdots S0$.

Representaremos con $\text{SUB}(k, 2^{12} \cdot 3^{13}, \ell)$ el código de Gödel de la fórmula que se obtiene al sustituir la variable libre x' —que tiene el código $2^{12} \cdot 3^{13}$ — por el valor ℓ en la fórmula cuyo código es k . Repetimos: $\text{SUB}(k, 2^{12} \cdot 3^{13}, \ell)$ es un número y se puede expresar dentro del sistema formal. Lo que aquí importa es que la formulación de SUB dentro del sistema formal no depende de los valores exactos de k, ℓ . Así será posible usar variables en vez de k y ℓ .

Ahora, juntemos todo y consideremos la siguiente fórmula con variable libre x' :

$$\forall x(\text{DEM}(x, \text{SUB}(x', 2^{12} \cdot 3^{13}, x')) = 0), \tag{53}$$

como fórmula bien formada, lo cual significa que $\text{SUB}(x', 2^{12} \cdot 3^{13}, x')$ no tiene demostración dentro del sistema formal. Como fórmula bien formada, 53 tiene un código de Gödel n , un número natural —que, sin duda, será gigantesco—. Sustituimos ahora la variable x' por el código n :

$$\forall x(\text{DEM}(x, \text{SUB}(n, 2^{12} \cdot 3^{13}, n)) = 0) \quad (54)$$

Lo que obtenemos es una fórmula bien formada que ya no tiene variable libre. ¿Qué significa? Ya lo sabemos: que $\text{SUB}(n, 2^{12} \cdot 3^{13}, n)$ no tiene demostración dentro del sistema formal. Pero $\text{SUB}(n, 2^{12} \cdot 3^{13}, n)$ es la fórmula que se obtiene al sustituir en 53 —que es la fórmula cuyo código es n — la variable x' por n . Es decir, $\text{SUB}(n, 2^{12} \cdot 3^{13}, n)$ no es otra cosa que 4.54, que dice que $\text{SUB}(n, 2^{12} \cdot 3^{13}, n)$ no es demostrable.

El círculo se cerró: 54 es una fórmula bien formada que afirma que ella misma no es demostrable. Falta ver que ni 54 ni su negación se pueden deducir a partir de los axiomas. En efecto, si existiera una demostración F_1, \dots, F_l con código d de 4.54, entonces:

$$\text{DEM}(d, \text{SUB}(n, 2^{12} \cdot 3^{13}, n)) = 1$$

dado que $\text{SUB}(n, 2^{12} \cdot 3^{13}, n)$ es 54. Esto nos da una demostración de la fórmula:

$$\exists x(\text{DEM}(x, \text{SUB}(n, 2^{12} \cdot 3^{13}, n)) = 1). \quad (55)$$

Pero como 55 es justo la negación de 54, demostramos una afirmación y su contradicción, cosa que es imposible si el sistema formal es consistente, lo que supusimos desde el principio. Así, llegamos a la conclusión de que no puede haber demostración de la fórmula 54 dentro del sistema y de que tampoco puede haber demostración de la negación de $\text{SUB}(n, 2^{12} \cdot 3^{13}, n)$, pues si hubiera una demostración con código d , entonces:

$$\text{DEM}(d, \neg \text{SUB}(n, 2^{12} \cdot 3^{13}, n)) = 1$$

y, por la consistencia, encontraríamos que no puede haber demostración de $\text{SUB}(n, 2^{12} \cdot 3^{13}, n)$, es decir, dedujimos dentro del sistema la fórmula 4.54, que no es otra cosa que $\text{SUB}(n, 2^{12} \cdot 3^{13}, n)$ mismo. Nuevamente, tenemos una contradicción con la consistencia.

Así es como Gödel llega a la conclusión de que cualquier sistema que sea lo suficientemente poderoso para poder operar con número naturales, siempre contendrá afirmaciones que no son demostrables, ni tampoco su negación.

Con ello, concluimos este viaje por la lógica. En resumen, cualquier formalización que incluye la aritmética es necesariamente incompleta. Por lo tanto, cualquier formalización de *todas* las matemáticas, necesariamente, es incompleta. En este sentido, los puntos 2 y 3 del programa de Hilbert no son compatibles: si el sistema es consistente es necesariamente incompleto.

Lo anterior puede parecer desalentador, ya que con ello se derrumbó el ambicioso programa de Hilbert. Por otro lado, se puede ver como una invitación para hacer matemáticas, ya que estas no se pueden realizar completamente dentro de sistemas formales. La demostración misma del Teorema de Gödel es una joya de razonamiento que no se da en un sistema formal, sino que se razona sobre ellos. Visto de esta manera podríamos interpretar este resultado como una confirmación de la importancia para la creatividad del intelecto humano dentro de las matemáticas.

4.12 Y... ¿SI TODO QUEDARA DESCUBIERTO?



Figura 4.82: La Sociedad Matemática Mexicana eligió como logotipo una cinta torcida en forma de cuadrado, una figura que aparenta un cuerpo tridimensional pero que es imposible de realizar. ¿Es una metáfora para lo que representan las matemáticas? ¿Será que aun lo que parece imposible se puede lograr en las matemáticas? La Sociedad Matemática Mexicana se compone actualmente de varios cientos de matemáticos y matemáticas, investigadores y profesores.

Es común escuchar expresiones de sorpresa y hasta de incredulidad respecto al quehacer de un matemático en nuestros días: “pero ¿qué van a investigar?, ¿la tabla del 18?, ¿los números?” Es cierto, los números han estado ahí desde siempre y es la asociación más frecuente a la palabra “matemáticas”. Pero, aunque hay muchas preguntas respecto a los números que nadie ha podido contestar aún, las matemáticas no se reducen sólo a ellos.

En la historia ha habido tres etapas en las cuales las matemáticas prosperaron particularmente bien: la primera en la antigua Grecia, la segunda en el siglo XVIII y la tercera es ahora. Nunca hubo una comunidad tan grande de científicos que se dedicaran a las matemáticas como ahora, ni tampoco existía la actual producción frenética de resultados.

La revista *Mathematical Reviews* se imprime cada mes y cada uno de sus tomos tiene alrededor de mil páginas. En ellas se encuentran las reseñas, es decir, los resúmenes elaborados por matemáticos de los artículos que se publicaron hace poco. Cada mes se reseñan alrededor de cinco mil artículos y libros, lo que equivale, aproximadamente, a 50 mil páginas de matemáticas nuevas cada mes. Este volumen de conocimiento es tan grande que es absolutamente imposible leerlo todo y, por ello, las matemáticas se han dividido en áreas y en ramas.

La ramificación de las matemáticas no es estática, sino que se encuentra en permanente desarrollo. Tan sólo el catálogo actual de los grandes temas requiere de diez páginas para imprimirse. El cúmulo de información generada por estos científicos es tan grande que no resulta fácil, para los propios matemáticos, mantenerse actualizados.

A lo anterior, debe añadirse que los artículos de investigación publicados por los matemáticos —como reporte de sus investigaciones— no son de fácil lectura. Leer uno de estos artículos puede costar días y, a veces, hasta meses de trabajo para comprender cabalmente lo que se reportó. La frontera del conocimiento avanza lentamente y, al mismo tiempo, se esparce y se ramifica cada vez más.

Por ello, para mantenerse informados, los matemáticos suelen reunirse con frecuencia en conferencias de especialistas que se dedican a alguna de las ramas específicas de las matemáticas. Por otro lado, también se asocian en sociedades como la Sociedad Matemática Mexicana, para mantenerse en contacto permanente.

¿Y no tendrá esta actividad frenética, de repente, su culminación cuando todo esté descubierto? Hay que detenerse un momento para reflexionar acerca de qué significaría que, llegado un momento, ya no hubiera nada que estudiar. Esto sólo podría suceder si todos los problemas del mundo quedaran resueltos, si toda la naturaleza fuera entendida y se contestaran todas las preguntas matemáticas. Pero, como caja de Pandora, cada pregunta que se contesta genera una, tres o más nuevas preguntas. Un fenómeno exponencial que a fin de cuentas, está muy lejano de poner un fin a las actividades matemáticas por falta de preguntas. Quizá podría llegar por falta de interés. Pero las sociedades exitosas han encontrado que sus logros dependen cada vez más del desarrollo de las ciencias, y las economías emergentes lo son por estar promoviéndolas. El interés social en las matemáticas también crece. ¿Y el individual? Que decayera sería como pensar que dejara de haber músicos, poetas o artistas pues, al igual que los matemáticos, expresan fibras muy profundas del espíritu humano. Queda claro que no se le ve el fin.

Se podría pensar que las preguntas estudiadas por los matemáticos son cada vez más sofisticadas, cada vez más retiradas de lo que se pueda comprender y, en cierta medida, eso es cierto si no fuera por algunos desarrollos que, a finales del siglo xx, resolvieron unas preguntas muy viejas. Entre ellas, la conjetura de Fermat que sólo concierne a los números y que ya hemos mencionado: la ecuación $a^n + b^n = c^n$ sólo tiene soluciones de números enteros positivos si $n \leq 2$. Esta conjetura debida a Fermat, que vivió en el siglo xvii no se demostró sino hasta 1995 por Andrew Wiles.

Figura 4.83 Fermat.

Actualmente hay siete problemas que se llaman “los problemas del milenio”. Para cada uno de ellos, el Instituto Clay ofrece un millón de dólares a aquel que logre resolverlo. Esto quiere decir que la solución se tiene que publicar en una revista internacional y tiene que sostenerse sin problema durante los siguientes dos años, después de su publicación. Durante este tiempo, se supone, los especialistas revisarán la prueba para detectar posibles errores o huecos en la argumentación. Por ejemplo, la primera prueba de Andrew Wiles presentó un hueco que requirió dos años más de trabajo.

Figura 4.84 Wiles.

En conclusión, las matemáticas se encuentran en una de las eras más prósperas y donde no hay señales de que todo lo que se puede resolver, pronto esté resuelto. Las matemáticas son un campo de investigación activo que da servicio a todas las ciencias al proveer un lenguaje abstracto pero tremendamente útil, al mismo tiempo que se desarrollan por intereses generados por las propias matemáticas y sirven, de manera casi inexplicable, como lenguaje para describir la naturaleza. Por todo lo anterior, vale la pena acercarse a ellas, tratar de adquirir conocimientos profundos en ellas al familiarizarse con sus métodos y su sensibilidad para enfrentar diversos problemas.