

Desarrollo del pensamiento matemático¹

José Luis Abreu y Michael Barot

28 de septiembre de 2017

¹En construcción

Índice general

1. Resumen histórico	7
1.1. Introducción	7
1.2. Los orígenes (3000 - 350 aec)	8
1.3. Las matemáticas griegas, época de oro (350 - 100 aec)	8
1.4. Alejandría, siglos I al IV	9
1.5. Del siglo V al Renacimiento (siglo XVI)	9
1.6. Siglo de oro, el XVII	10
1.7. Expansión en el siglo XVIII	11
1.8. Consolidación en el siglo XIX	11
1.9. Siglos XX y XXI	12
1.10. Presente y futuro de las Matemáticas	12
Bibliografía	12
2. Los números	13
2.1. Los sistemas de numeración	13
2.2. Numerología y Teoría de números.	21
2.3. La irracionalidad de la raíz cuadrada de dos	22
2.4. La Teoría de Números	25
2.5. El estudio de los números naturales en Euclides	28
3. Geometría: origen y fundamentos	33
3.1. Los orígenes del pensamiento geométrico	33
3.2. El pensamiento geométrico en la antigua Grecia	37
3.3. Los fundamentos de la geometría	50
3.4. Los axiomas de conexión e incidencia	53
3.5. La geometría proyectiva	53
3.6. La perspectiva	64
3.7. La geometría hiperbólica	74
3.8. Los axiomas de orden y el de Euclides	87
3.9. Los axiomas de congruencia y el de Arquímedes	90
4. La geometría clásica	93
4.1. Propiedades del círculo	94
4.2. Propiedades del triángulo	98

4.3.	Construcciones con regla y compás	103
4.4.	Los problemas de Apolonio	106
4.5.	Las secciones cónicas según Apolonio	110
5.	Geometría y Medición	121
5.1.	El uso de los números para cuantificar la realidad	121
5.2.	Longitud, área y volumen	124
5.3.	El método de exhaustión	134
5.4.	La cuadratura de la parábola	139
5.5.	El número π a través de los siglos	142
5.6.	El volumen y la superficie de la esfera	158
5.7.	Cálculos de distancias celestes	160
5.8.	La trigonometría	168
5.9.	La medición fina de la Tierra	185
5.10.	Herramientas de precisión	194
6.	El álgebra	195
6.1.	El álgebra como disciplina independiente	195
6.2.	El camino a una notación algebraica	205
6.3.	Teoría de Ecuaciones	214
6.4.	Los Números complejos y las funciones complejas	225
6.5.	El álgebra lineal	225
6.6.	Álgebra moderna, álgebra abstracta	238
7.	Las coordenadas	241
7.1.	La Geometría con Coordenadas	241
7.2.	Ecuaciones cuadráticas	242
7.3.	Vectores	243
7.4.	Curvas y superficies	243
7.5.	Transformaciones proyectivas	243
7.6.	La dimensión	243
8.	Cálculo, funciones y límites	253
8.1.	El Cálculo Diferencial e Integral	253
8.2.	El cálculo	258
8.3.	Las funciones en el cálculo	262
8.4.	Desarrollo del concepto de límite	270
8.5.	Leyes del movimiento Planetario	279
8.6.	Expansión del Cálculo	279
8.7.	Ecuaciones diferenciales	279
8.8.	Series de Potencias	279
8.9.	Formalización del Cálculo	279
8.10.	Métodos numéricos en el Cálculo	280
9.	La probabilidad	281
9.1.	La Invención de la probabilidad	281
9.2.	El cálculo combinatorio	281

9.3. La ley de los grandes números	281
9.4. El Teorema del límite central y la campana de Gauss	281
9.5. Formalización de la Teoría de Probabilidad	282
9.6. La estadística	282
10.Otros desarrollos	283
10.1. Introducción	283
10.2. Teoría de gráficas	283
10.3. Geometría de la forma: la Topología	283
10.4. Las matemáticas discretas	283
10.5. Los problemas de Hilbert	283
10.6. Las corrientes fundacionistas	284
10.7. El Teorema de Gödel y el de Turing	284
10.8. La Computación como rama de las Matemáticas	297
10.9. El papel de las matemáticas en la Física	299
11.La perspectiva actual	301
11.1. Introducción	301
11.2. Irrupción de la computadora y el Internet	301
11.3. ¿Revaloración del Cálculo?	301
11.4. Ejemplos de usos de las matemáticas contemporáneas	301
11.5. Usos actuales de las matemáticas tradicionales	301
11.6. ¿Hacia dónde van las matemáticas?	301
11.7. Los problemas del milenio	301
11.8. ¿Y los fundamentos?	301
Índice de Referencias de las Figuras	303
Índice de Personas	307

Capítulo 1

Resumen histórico

1.1. Introducción

La tradición nos dice que los primeros desarrollos matemáticos se hicieron en el antiguo Egipto, en Sumeria y Babilonia hace unos cuatro o cinco mil años. En realidad los primeros elementos de las matemáticas surgen mucho antes con el desarrollo del lenguaje. Por ejemplo las palabras que se usan para contar pequeños conjuntos de objetos, e incluso la manera de representar esos números, aunque fuese muy rudimentaria, por ejemplo haciendo marcas en un hueso, tantas como ovejas hay en un rebaño, surge mucho antes de que las culturas del oriente mediterráneo se establecieran. Surgen como parte integral del lenguaje humano. Keith Devlin (Ref. 1), un conocido investigador en el tema del pensamiento matemático piensa que las habilidades matemáticas son las mismas que las lingüísticas, ambas se ocupan de abstracciones, la única diferencia es que las de las matemáticas tienen un grado de abstracción mayor que las del lenguaje ordinario. Pero las relaciones entre los conceptos abstractos que manejan las matemáticas son, si no más simples, por lo menos más claras que las que ocurren entre las abstracciones más mundanas como podrían ser las personalidades de los miembros de una tribu o las plantas y sus productos o los animales y sus costumbres, asuntos con los que hay que lidiar en la vida diaria.

A lo largo de este libro nos adherimos a la idea de que las matemáticas constituyen un lenguaje especializado en el manejo de conceptos abstractos entre los que hay unas relaciones bien definidas. De hecho el nombre mismo de las Matemáticas proviene del griego MATHEMATIA que significa aquello que se puede entender racionalmente. Y entender algo significa tener claras las relaciones que operan entre los elementos de un sistema y poseer la capacidad de deducir que otras relaciones también se cumplen a partir de las primeras mediante un proceso de deducción lógica. Esto es, a grandes rasgos, lo que se entiende por un sistema axiomático. Sin embargo no deseamos reducir las matemáticas a sistemas axiomáticos ya que éstos suponen una claridad ab-initio de las relaciones entre los conceptos abstractos de que se ocupan, mientras que en el desarrollo de las matemáticas la formación de conceptos y la clarificación de las relaciones entre ellos y la búsqueda de las más básicas a partir de las cuales las demás pueden deducirse, suele ser un proceso complejo que lleva tiempo y esfuerzo. Consideramos que estos procesos son parte integrante del quehacer matemático y que describir a las Matemáticas como un conjunto de sistemas axiomáticos deja de lado lo que

puede ser, quizás, su aspecto más importante.

Las matemáticas constituyen una de las actividades humanas más refinadas. Fueron desarrolladas en muchas culturas diferentes por gente que observaba el mundo en que vivía, tratando de reconocer orden y coherencia, con la esperanza de predecir y tal vez controlar eventos futuros.

1.2. Los orígenes (3000 - 350 aec)

Las matemáticas nacieron cuando el hombre primitivo tuvo necesidad de contar y evolucionaron como un lenguaje escrito hace unos cinco mil años cuando los creadores de calendarios calcularon el inicio de las estaciones e hicieron marcas en tabletas o papiros para registrar esos datos. Siglos después, el escolar renacentista Galileo Galilei escribió que el gran libro de la naturaleza está abierto ante nuestros ojos y la filosofía verdadera está escrita en él, pero no podemos leerlo sin antes aprender el lenguaje y los símbolos con los que está escrito; está escrito en lenguaje atemático y los símbolos son triángulos, círculos y otras figuras geométricas.

Pero los símbolos matemáticos más antiguos no fueron figuras geométricas sino marcas en tabletas de barro babilonias que se usaban como calendarios o jeroglíficos cavados en piedra en los templos mayas. Con el desarrollo del comercio, la necesidad de medir y contar acompañó a las caravanas de viajeros comerciantes.

La medición de la tierra, de donde la Geometría toma su nombre, tuvo que satisfacer las necesidades de la irrigación y el cálculo de impuestos en el antiguo Egipto y en Mesopotamia. Al construir las pirámides y tener que orientarlas debidamente exactamente al norte o al oriente, los arquitectos egipcios descubrieron relaciones entre triángulos, rectángulos y círculos. En la antigua Alejandría, la Geometría y la Trigonometría se usaron para medir el tamaño de la Tierra y las distancias al Sol y a la Luna.

Grandes desarrollos ocurrieron cuando los antiguos griegos como TALES DE MILETO (620 - 546 aec) o PITÁGORAS (569 - 475 aec) y sus discípulos llamados “los pitagóricos” trataron a los números y las figuras geométricas como abstracciones puras. Estos juguetes mentales generaron las bases para toda la ciencia y tecnología con que contamos en la actualidad. ¿Cómo ocurrió esto?

1.3. Las matemáticas griegas, época de oro (350 - 100 aec)

El desarrollo de los conceptos de área y volumen como entes cuantificables por medio de los números, y del conceptos de razón, también expresable mediante números llevan al inicio de la época de oro de las Matemáticas griegas en las que se descubren el teorema de los triángulos semejantes (los triángulos semejantes tienen sus lados proporcionales) y el Teorema de Pitágoras. También se formaliza el número π como la razón del perímetro de un círculo a su diámetro y se obtienen buenas aproximaciones de su valor. Éstas y otras ideas importantes que siguieron desarrollándose a lo largo de estas líneas culminaron en el siglo XVII en la creación del Cálculo dando un impulso sin precedentes al desarrollo de la Ciencia y la Tecnología.

Sin duda el libro más importante en toda la historia de las Matemáticas y uno de los más importantes en la historia de la humanidad es “Los elementos” de EUCLIDES DE ALEJANDRÍA (325 - 265 aec), escrito alrededor del año 300 aec. La aparición de esta obra marca un momento decisivo en la

historia de la humanidad. Es una admirable colección de trece libros que contienen casi todo el conocimiento matemático de su tiempo, incluyendo los trabajos de EUDOXO DE CNIDO (390 - 337 aec), uno de los tres matemáticos más importantes de la antigüedad griega, junto con ARQUÍMEDES DE SIRACUSA (287 - 212 aec) y APOLONIO DE PERGA (262 - 190 aec). Bien presentado y organizado en orden lógico, es un logro cultural extraordinario que ha influido en el pensamiento científico de la humanidad durante más de dos mil años. Asentó las bases del método deductivo, mostrando que el intelecto humano puede descubrir verdades matemáticas utilizando únicamente la razón y la lógica. “Los Elementos” se escribió originalmente en griego sobre rollos de papiro, pero fue traducido a muchos idiomas y publicado en incontables ediciones.

Entre los logros más importantes que se atribuyen a EUDOXO están la teoría de las proporciones (Libro V de Los elementos), la fórmula para el volumen de una pirámide y el método de exhaustión para calcular o estimar áreas y volúmenes de regiones acotadas por curvas y superficies que no son planos.

Entre los principales logros de ARQUÍMEDES se consideran su estimación de π , la cuadratura de la parábola (utilizando en ambos casos el método de exhaustión) y los cálculos del volumen y la superficie de la esfera. El método de exhaustión también lo utilizó en China, de manera independiente, Liu Hui en el siglo III aec para encontrar el área del círculo, logrando obtener un valor de π más preciso que el de Arquímedes.

A APOLONIO se le conoce principalmente por su exhaustivo y elegante trabajo sobre las secciones cónicas, que solo podrá ser superado casi dos milenios más tarde en el siglo XVII con la geometría analítica de FERMAT y DESCARTES.

1.4. Alejandría, siglos I al IV

Durante el apogeo del Imperio romano que controló la cuenca del mar Mediterráneo desde unos años antes de Cristo hasta el siglo IV, el desarrollo de las matemáticas decayó. Esta época está dominada por personajes que más que desarrollar conceptos, ideas y métodos, lo que hacían era comentar o recopilar los conocimientos de sus predecesores. Sin embargo la ciudad griega de Alejandría, fundada por ALEJANDRO MAGNO (356 - 323 aec) en la desembocadura del río Nilo, en Egipto, se convirtió en la capital y refugio cultural de la época y especialmente durante los siglos II al V hubo importantes desarrollo en matemáticas. Entre los personajes más notables de esta ciudad y en esta etapa (no olvidemos que Euclides vivió en Alejandría, pero mucho antes, alrededor del 300 aec) que hicieron contribuciones a las matemáticas destacan MENELAO DE ALEJANDRÍA CLAUDIO PTOLOMEO (100 - 170), DIOFANTO DE ALEJANDRÍA (210 - 290 aprox), PAPPUS (siglos III y IV), TEÓN DE ALEJANDRÍA (335 - 405) y su hija HIPATIA (370 - 415). Entre quienes hicieron principalmente comentarios a las obras de sus predecesores, sobre todo a las de EUCLIDES y ARQUÍMEDES, destaca sobre todos PROCLO (412 - 485).

1.5. Del siglo V al Renacimiento (siglo XVI)

De manera independiente, en Asia los matemáticos indios y chinos crearon sus propios símbolos matemáticos, incluidos los números indo-arábigos que usamos en la actualidad. De las manos de matemáticos como AL-JUARISMI (780 - 850) y OMAR JAYAM (1048 - 1131), surgieron nuevos

métodos matemáticos que hoy conocemos como álgebra, de acuerdo con su nombre árabe que significa reintegración o recomposición.

En China, el siglo V, ZU CHONGZHI usó lo que más tarde sería llamado la teoría de los indivisibles por el matemático italiano Bonaventura Cavalieri para encontrar el volumen de una esfera.

Cerca del año 1000, el matemático islámico ALHACÉN fue el primero en derivar la fórmula para la suma de la cuarta potencia de una progresión aritmética, usando un método a partir del cual es fácil encontrar la fórmula para la suma de cualquier potencia integral de mayor orden. Esto sentó las bases para obtener más adelante el área bajo las gráficas de las funciones polinomiales.

En el siglo XI, el polímata chino SHEN KUO desarrolló ecuaciones que se encargaban de integrar. En el siglo XII, el matemático indio, BHASKARA II, desarrolló una derivada temprana representando el cambio infinitesimal, y describió una forma temprana del llamado teorema de Rolle.

También en el siglo XII, el matemático persa SHARAF AL-DIN AL-TUSI descubrió la derivada de la función cúbica, un importante acontecimiento que abre el camino hacia el cálculo diferencial.

En el siglo XIV, MADHAVA DE SANGAMAGRAMA, en conjunto con otros matemáticos y astrónomos de la escuela Kerala de matemáticas y astronomía, al sur de la India, describieron casos especiales de lo que hoy se conocen como las series de Taylor.

Las rutas comerciales llevaron a Europa la notación indo-arábiga y los algoritmos para realizar operaciones con ella, desplazando rápidamente el incómodo sistema de números romanos que se había usado durante casi toda la edad media. Es importante reconocer la contribución de LEONARDO DE PISA (FIBONACCI) (1170 - 1240) en este proceso.

Esto generó un fuerte interés en el álgebra de los árabes que culminó con la creación del álgebra simbólica en la que se utilizan letras para denotar cantidades desconocidas o variables genéricas. La creación de la notación algebraica fue impulsada principalmente por FRANÇOIS VIÈTE (1540 - 1603) y dio un nuevo impulso al desarrollo de las matemáticas que por fin encontraron un lenguaje simbólico que además de simplificar la notación de las ecuaciones y fórmulas, ayudaba a manipularlas con una eficiencia nunca antes experimentada.

1.6. Siglo de oro, el XVII

Pronto las ideas renacentistas se unieron a la antigua Geometría griega con lo que PIERRE DE FERMAT (1601 - 1665) y RENÉ DESCARTES (1596 - 1650) crearon la Geometría Analítica.

Ya con la notación algebraica y la Geometría analítica en el saco de herramientas, los matemáticos del siglo XVII, es decir, de la generación de ISAAC NEWTON (1643 - 1727) y GOTTFRIED LEIBNIZ (1646 - 1716), estuvieron en posición de domar los procesos continuos como el movimiento y el cálculo de áreas y volúmenes de figuras arbitrarias. Nace así el Cálculo diferencial e integral que será aprovechado por las ciencias, en especial por la Física, para modelar sus leyes y procesos y predecir el comportamiento de muchos sistemas, usando casi siempre los conceptos de función, límite, derivada e integral.

1.7. Expansión en el siglo XVIII

Durante el siglo XVIII la actividad matemática se ocupó principalmente de extender sus conceptos y sus aplicaciones, especialmente las del Cálculo. En este último aspecto destacan los nombres de los matemáticos suizos JACOB BERNOULLI (1654 - 1705), JOHANN BERNOULLI (1667 - 1748) y LEONHARD EULER (1707 - 1783), los primeros dos hermanos, el tercero discípulo del segundo y todos ellos procedentes de Basilea, que dedicaron sus vidas a profundizar en el estudio del cálculo, las funciones, los límites, las sucesiones infinitas y las ecuaciones diferenciales y a aplicar estas ideas a la modelación de muchos problemas científicos.

A partir del siglo XVIII las Matemáticas no han dejado de avanzar, creando nuevos conceptos, nuevas disciplinas, encontrando nuevos retos y nuevas aplicaciones. Casi todas las actividades humanas del mundo moderno están rodeadas de procesos matemáticos. Desde la Física teórica que confía a las matemáticas todos sus modelos del mundo material, pasando por la Biología que las necesita para modelar muchos aspectos del comportamiento de las poblaciones de animales y plantas, hasta los sistemas informáticos que las usan en sus sofisticados algoritmos.

1.8. Consolidación en el siglo XIX

Desde principios del siglo XIX comenzaron a aparecer signos de una preocupación por los fundamentos, por entender las raíces más profundas de las matemáticas. Una de estas manifestaciones fue la preocupación por determinar si todas las ecuaciones polinomiales tenían o no soluciones, se vio que sí las tenían, siempre y cuando se aceptara que estuvieran definidas en los complejos. Esto dio mayor impulso a los números complejos, que también se analizaron detalladamente, se formalizaron e incluso se creó una fructífera teoría de funciones de variable compleja que complementa el álgebra de polinomios complejos y extiende incluso el Cálculo. SIR WILLIAM ROWAN HAMILTON (1805 - 1865) dedicó una buena parte de su vida intelectual a resolver el problema de la existencia o no de números más generales que los complejos, llegando a la conclusión de que si se quería extender el concepto de número más allá de los complejos, era necesario prescindir de la conmutabilidad en la multiplicación y así descubrió los cuaternios.

Otra manifestación del interés en los fundamentos es la preocupación por el quinto postulado de Euclides en la Geometría que llevó al descubrimiento de su independencia y de la existencia de geometría no euclidianas, es decir, que no satisfacen el quinto postulado. Esta preocupación llevó hacia finales del siglo a una formulación axiomática rigurosa, por parte de DAVID HILBERT (1862 - 1943), de la Geometría euclidiana que sentó las bases para el concepto moderno de sistema axiomático. Quizás la mayor preocupación de los matemáticos del siglo XIX fueron los fundamentos del Cálculo. A pesar de sus innegables éxitos en las aplicaciones, el Cálculo sufría de un grave problema, usaba infinitesimales, entes mal definidos que promovían discusiones interminables acerca de su existencia misma. Los matemáticos sintieron necesidad de aclarar estos conceptos y la solución vino de la mano de matemáticos como AUGUSTIN LOUIS CAUCHY (1789 - 1857), KARL WEIERSTRASS (1815 - 1897) y RICHARD DEDEKIND (1831 - 1916) que acabaron desterrando el concepto y uso de los infinitesimales, sustituyendo en cambio el de límite. Así se desarrolló una extensa y detallada teoría sobre las funciones, sus límites y sus derivadas que dejaron perfectamente aclarados los conceptos y los fundamentos del Cálculo, aunque a costa de perder la simplicidad original con la que en el siglo XVIII se enfrentaban los problemas usando los infinitesimales de manera intuitiva.

1.9. Siglos XX y XXI

Las Matemáticas siguieron avanzando en el siglo XX de una manera impresionante. Lo hicieron creando nuevos conceptos, nuevas disciplinas, encontrando nuevos retos y nuevas aplicaciones y analizando los rincones más recónditos de sus fundamentos.

Hoy en día casi todas las actividades humanas del mundo moderno están rodeadas de procesos matemáticos. Desde la Física teórica que confía en las matemáticas todos sus modelos del mundo material, pasando por la Biología que necesita a las matemáticas para modelar muchos aspectos del comportamiento de las poblaciones de animales y plantas, hasta los cálculos que hace un GPS para determinar su posición exacta sobre la superficie de la Tierra o los cálculos que hacen las computadoras para dibujar a los personajes y objetos en las películas hechas por computadora, sin olvidar los sofisticados algoritmos que Google para encontrar los sitios de interés que pueden ser de mayor interés para quien hace una búsqueda o los que utilizan las redes sociales para ofrecer un funcionamiento útil y eficiente a sus usuarios, utilizan matemáticas.

Sin embargo es preocupante comprobar que al mismo tiempo los estudiantes de todo el mundo, en su mayoría temen a las matemáticas y tratan de huir de ellas eligiendo carreras en función del menor uso que hacen de las matemáticas. En opinión de los autores esto se debe a que en general la enseñanza de las matemáticas no aprovecha sus logros históricos para motivar y guiar el aprendizaje de la materia.

1.10. Presente y futuro de las Matemáticas

Durante toda la historia, las matemáticas han sido un espejo de la civilización. Desde aquellas marcas primitivas que se usaron para anotar el inicio de las estaciones, las matemáticas fueron evolucionando convirtiéndose en un lenguaje cada vez más útil y preciso. Quienes hicieron descubrimientos importante en relación con el movimiento de los cuerpos celestes, como Kepler y Newton, la convirtieron en un lenguaje aún más poderoso que se convirtió en la lingua franca de las Ciencias. Este hecho unido a la creciente utilización de las matemáticas en muchos aspectos de la vida moderna, debe ser motivo suficiente para que los jóvenes se interesen en ellas. Uno de los objetivos de este libro es proporcionar material que ayude a motivar el interés en las matemáticas echando mano de los detalles más apasionantes de su desarrollo histórico y mostrar que coinciden precisamente con lo que la sociedad espera que los jóvenes aprendan de matemáticas, a pesar de que los planes y programas de estudio muchas veces no lo reflejan.

Bibliografía

Sir Thomas Heath: A history of greek mathematics, Vol II, Dover 2014.

Keith Devlin: The Mathematics Gene, Editor YEAR.

Wikipedia en español, inglés y alemán.

Capítulo 2

Los números

Una primera extensión del lenguaje natural

2.1. Los sistemas de numeración

Cuando una cultura se desarrolla forma una organización más compleja con divisiones de trabajo. Se intercambian bienes, se registran cantidades, un sentido para la historia des desarrolla y se anotan fechas. Por ello no debe sorprender que hay diferentes sistemas de anotar números: esta idea se tuvo que inventar varias veces. En lo que sigue veremos los sistemas de enumeración de algunas culturas. Nos interesa en particular, si el sistema facilitó el cálculo o más bien lo limitó.

En casi todos los sistemas de numeración se observan repeticiones, la excepción de los que presentamos aquí es el de los griegos. El uso repetitivo de un símbolo disminuye el número de símbolos usados y así facilita el empleo ya que no hay que memorizar tantos de ellos. Al final de esta sección haremos una comparación y un análisis sobre el uso para hacer cálculos.

La Cuneiform

Lo usaron diferentes culturas en asia occidental, entre ellos los sumerios y los babilonios. Con una herramienta sencilla, como cuña, se grababa en arcilla húmeda. Luego se horneaba la tablilla. La cuneiforma se desarrolló a partir de pictogramas. La escritura se desarrolló alrededor de 3300 a.C. como pictogramas y se transformó en una forma cada vez más abstracta. Hay tablillas con la forma que cuneiforma que presentaremos abajo de alrededor de 2200 a.C.

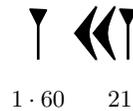
Para denotar números, se usaban dos símbolos sencillos. Para el Uno: un triángulo con punta hacia abajo, que a veces se alargaba con una raya y para el Diez un ángulo abierto hacia la derecha.



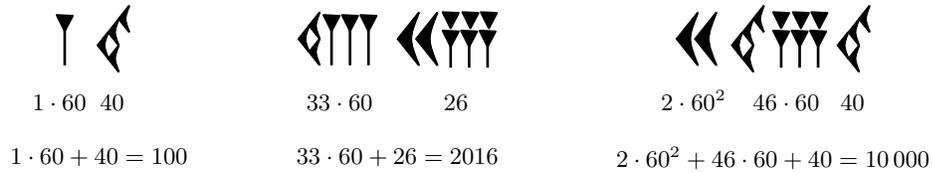
Números hasta el 59 se obtenían por repetición, por ejemplo



Números más grandes que 59 se escribían en forma posicional en base 60. Por ejemplo, el número $81 = 60 + 21$ se escribe como se muestra en la siguiente figura.

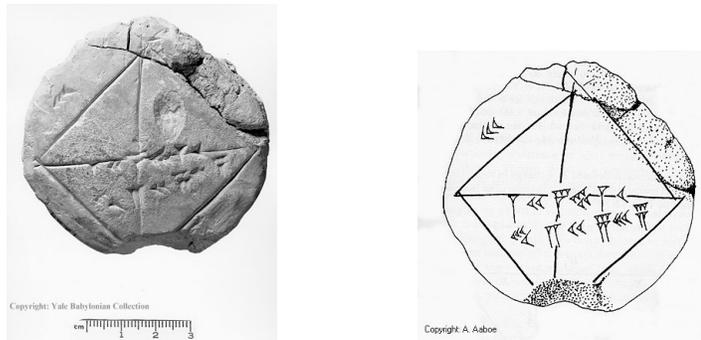


El primer 1 está en la segunda posición e indica que hay una vez 60 y luego se suman los 21 del segundo grupo de símbolos. Más representaciones en cuneiforme de números se ven en la siguiente figura.



Los babilonios también usaron sesentavos y el número en la figura arriba a la izquierda también podía significar $1 + 40 \cdot 60^{-1} = \frac{5}{3}$. El punto decimal se tenía que concluir del contexto y no se indicaba en la notación.

La siguiente tablilla, que se llama YBC7289 de alrededor de 1800 a.C., muestra un cuadrado con la diagonal.



La anotación sobre la diagonal es interesante. Si se interpreta con un punto decimal después del primer 1:

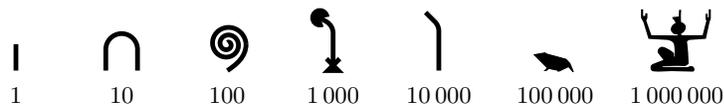
1	$24 \cdot 60^{-1}$	$51 \cdot 60^{-2}$	$10 \cdot 60^{-3}$
$1 + 24 \cdot 60^{-1} + 51 \cdot 60^{-2} + 10 \cdot 60^{-3} = 1.41421296 \dots \approx \sqrt{2}$			

se obtiene una aproximación extremadamente buena de raíz de 2.

De los babilonios heredamos la división del tiempo: la hora en 60 minutos, el minuto en 60 segundos. Además también la medición de ángulos en 360 grados.

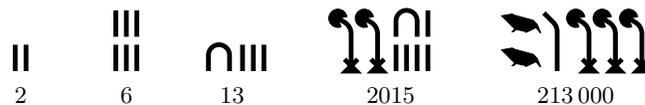
Los jeroglíficos egipcios

Los egipcios usaron un sistema que se basaba en pocos símbolos y posiblemente muchas repeticiones. Se basaban en la base 10: por cada potencia de diez tenían un símbolo distinto, tal como se muestra en la siguiente figura.



El símbolo del 1 es totalmente abstracto, sobre el significado del 10 hay diversas opiniones, pero el 1000 es representado por una planta de agua como un lirio. Vemos en estos símbolos la transformación que transcurrieron con los jeroglíficos: al principio eran imágenes y por el frecuente uso se abstrajeron cada vez más. El uno, por ser el más usado es una simple raya, mientras el símbolo del millón sigue siendo una imagen.

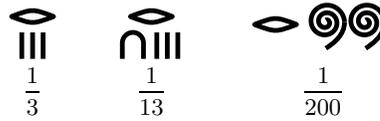
Para representar números se repetían los símbolos las veces que eran necesarias.



La siguiente ilustración muestra una parte de los anales de Thoutmosis III del año 1437 a. C. que indica que en el momento el rey contaba con 4622 ganados.



El sistema no era posicional. Por ello necesitaron una idea diferente para representar fracciones. Estas las expresaron como suma de fracciones unitarias, es decir con numerador 1. Para indicar que se trataba de una fracción y no de un número escribían el símbolo de un ojo encima o a un lado del número. La siguiente figura muestra algunas fracciones unitarias en la denotación egipcia.



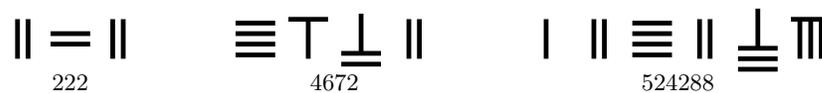
Los egipcios tenían símbolos especiales para las fracciones $\frac{2}{3}$, $\frac{1}{2}$ y en general para fracciones unitarias cuyo denominador es una potencia de 2.

Las varillas de los chinos

En China se desarrolló un sistema posicional en base diez. Se encontró en tumbas del período de los Reinos Combatientes (alrededor desde 480 a.C hasta 220 a.C.). El sistema usa un sólo símbolo: una varilla. Para números hasta 9 una varilla vertical significa 1 y se repetía hasta por máximo 4 veces. Una varilla horizontal significa 5 y se escribía encima de las varillas verticales.



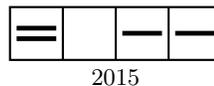
El sistema era posicional, pero el significado de horizontal y vertical cambiaba de posición en posición. En la posición de las decenas una varilla horizontal significaba 1 decena mientras una vertical representaba 5 decenas. La siguiente figura muestra algunos números.



No tenían un símbolo para representar el cero, pero como se muestra en la siguiente ilustración, se usaba una rejilla para indicar las posiciones.

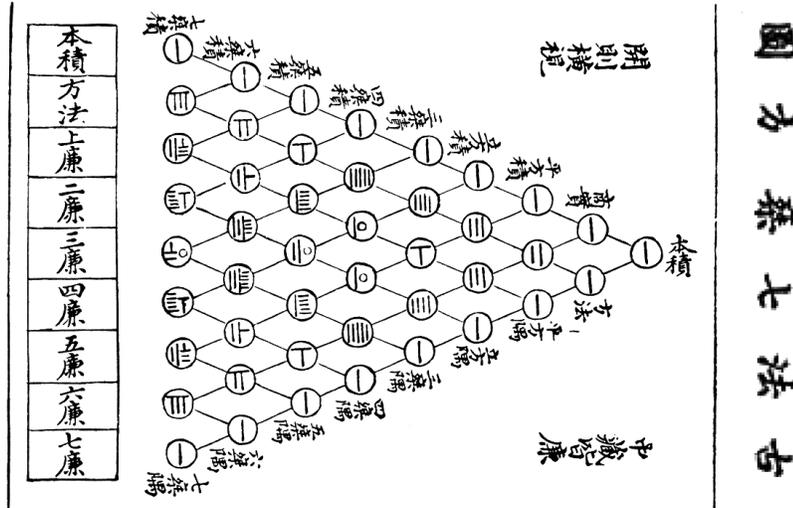


Así, el número 2015 se escribía de la siguiente manera.



La adición con varillas es muy sencillo ya que estas se pueden juntar físicamente. En este sentido, el sistema de las varillas era más práctico que nuestro sistema decimal actual. En la dinastía Song (del año 960 al 1279) se usaban las varillas con eficacia: se tenían algoritmos para llevar a cabo las cuatro operaciones básicas, se conocían los números negativos y las fracciones. La siguiente ilustración, del

año 1303, muestra lo que nosotros conocemos por “triángulo de Pascal” (era conocido siglos antes en China). Los números se representaron con varillas, pero ya se puede ver un pequeño circulito para indicar el cero y que ya no siempre se cambiaba de dirección de las varillas de posición en posición.



Este sistema numérico de varillas se abandonó por el uso del ábaco.

Letras de los griegos

Los griegos es la cultura, que descubrió la argumentación, el sistema axiomático, la observación en medicina, la filosofía, la democracia, la poesía y el teatro tenían el peor sistema de numeración. Adoptaron sus letras de los fenicios, igual que los judíos y con ello también el sistema de numeración. Usaron muchas de sus letras como símbolos para expresar números. Para distinguirlos de las letras las adornaron con un acento, es decir α denota la letra y α' denota el número 1. La siguiente tabla muestra los símbolos que usaron, tres de ellos ya no se usan hoy: digrama, copa y sampi.

α	β	γ	δ	ϵ	ζ	η	θ
(alfa)	(beta)	(gama)	(delta)	(épsilon)	(digama)	(dseta)	(theta)
1	2	3	4	5	6	7	8
ι	κ	λ	μ	ν	ξ	\omicron	π
(iota)	(kappa)	(lambda)	(mi)	(ni)	(xi)	(ómicron)	(pi)
10	20	30	40	50	60	70	80
ρ	ς	τ	υ	ϕ	χ	ψ	ω
(rho)	(sigma)	(tau)	(ípsilon)	(fi)	(ji)	(psi)	(omega)
100	200	300	400	500	600	700	800
							λ
							(sampi)
							900

Con ello podían representar números hasta 999. Por ejemplo

ι'	$\phi\lambda\beta'$	$\omega\pi\eta'$
11	532	888

Para números más grandes usaron un símbolo de acento en la línea base. Por ejemplo 2015 se representaba por $\beta\epsilon'$. No obstante que su sistema de numeración era poco útil para hacer operaciones matemáticas, investigaron sobre los números y se sirvieron de la geometría para hacerlo: así descubrieron que cada número podía factorizarse en primos, que había un número infinito de primos y que existían números que no eran fraccionarios que hoy se llaman irracionales.

Más tarde, alrededor de 140 a.C. se sirvieron de un sistema posicional en base de 60, posiblemente adoptado de los babilonios, para denotar fracciones menores que 1. Para ello usaron un símbolo especial para el 0. Esto lo usaron para hacer cálculos astronómicos.

Letras de los romanos

Los romanos usaron algunas letras para denotar números. La siguiente figura muestra los cinco símbolos que se usaron en la antigüedad.

I	V	X	L	C	[D	M]
1	5	10	50	100	[500	1000]

Las letras I, X y C se podían repetir hasta 4 veces. Sólo más tarde se añadieron las letras D para denotar 500 y M para denotar 1 000. También se introdujo la regla que un símbolo de menor valor a la izquierda de uno de mayor valor significa que el primero se tenía que restar del segundo: IX denota el número 9 y CM el número 900. Aquí hay algunos ejemplos.

XXIII	LIX	XCVI	DCCCLXXXVIII	MCMXLV	MMXV
23	59	96	888	1945	2015

El sistema romano usa pocos símbolos y es parecido al de los egipcios, solo que no se podían denotar números muy grandes. Esto no se hizo hasta en la edad medieval. Los números romanos se encuentran con frecuencia para denotar años y los números en los relojes.

Los jeroglíficos de los mayas

Los mayas usaron sólo dos símbolos para denotar números: un punto y una raya que representaban el número 1 y 5 respectivamente. Estos símbolos lo repitieron para representar números hasta 19. La siguiente figura muestra los primeros números en la escritura maya.

	•	••	•••	••••
0	1	2	3	4
<hr style="width: 100%;"/>	•	••	•••	••••
5	6	7	8	9
<hr style="width: 100%;"/> <hr style="width: 100%;"/>	•	••	•••	••••
10	11	12	13	14
<hr style="width: 100%;"/> <hr style="width: 100%;"/> <hr style="width: 100%;"/>	•	••	•••	••••
15	16	17	18	19

Los mayas usaron un sistema posicional en base de veinte. La siguiente figura muestra algunos números en este sistema numérico.

				12×20^4	$=$	1 920 000		
				0×20^3	$=$	0		
	3×20^2	$=$	1200		19×20^2	$=$	7 600	
	0×20^1	$=$	0		8×20^1	$=$	160	
	16×20^0	$=$	16		11×20^0	$=$	11	
			1216				1 927 771	

Los mayas usaron este sistema de numeración alrededor de 300 a. C. hasta la llegada de los conquistadores españoles para anotar fechas. Como anotaron eventos astronómicos, como eclipses solares, fue posible descifrar que su conteo de días iniciaba el 11 de agosto de 3114 a. C. en el calendario Gregoriano proléptico. Contaban día por día en un sistema casi-vigésimal, dado que había una excepción al uso de la base 20: un día se llama *k'in*, 20 *k'in* hacen un *winal*, pero 18 *winal* hacen un *tun*, luego 20 *tun* hacen un *k'atun*, 20 *k'atun* un *b'ak'tun* y hay dos niveles más. La excepción tiene el efecto que un *tun* consiste de 360 días, lo que es cercano a un año.

En un sistema posicional se requiere un cero, que indica que en esta posición no hay ningún número. Los mayas usaron diferentes símbolos. La primera piedra que usa un símbolo data alrededor de 300 d. C. y por ello es el primer testimonio del cero, tres siglos antes de los árabes.

Los símbolos de los hindues y árabes

Durante los siglos I al IV d. C. los hindues desarrollaron el sistema decimal que hoy usamos, sólo que los símbolos que usaron fueran otros. Este sistema lo adoptaron los árabes y contaron con un símbolo para el cero. Los símbolos se transformaron con el uso y por la adopción en otras las culturas y no eran estandarizados.

La numeración moderna

En el siglo X europeos empiezan a estudiar textos árabes y adoptan de la numeración los símbolos del 1 al 9. Más tarde se adopta también el cero y los símbolos toman su forma moderna en el siglo XV, cuando los libros se empiezan a imprimir con letras móviles. El punto decimal se usa por primera vez en 1492 en la obra *Compendio del Abaco* del italiano Francesco Pellos, pero tomó alrededor de 100 años más hasta que se adoptara ampliamente. Hoy día se usa en algunas países también la coma para separar la parte entera del fraccionario.

Comparación

Para denotar números se inventaron diferentes sistemas, muchas veces con rasgos similares. Con excepción de la numeración de los griegos, todos los sistemas usan algún tipo de repetición. Esto permite reducir la cantidad de símbolos. Podemos ver que en diferentes culturas se usaron ideas muy similares. Distinguiamos dos clases de sistemas de numeración:

- Los símbolos tienen un valor fijo: egipcios, griegos y romanos.
- Los símbolos cambian su valor según la posición donde se encuentran: esto es el principio del sistema posicional. Lo usaron: los sumerios y babilonios, los chinos, los mayas y finalmente los hindues y árabes.

El sistema posicional tiene claras ventajas: es posible representar números arbitrariamente grandes y las operaciones básicas como la adición o la multiplicación se pueden hacer siguiendo un algoritmo fijo. Los sistemas cuyos símbolos un valorfijo no tienen estas propiedades, pero a cambio son bastante fáciles de leer.

Es interesante ver que los sumerios y babilonios, los chinos y los mayas usaron muy pocos símbolos básicos que repetían para representar “una cifra”, es decir, el número en una sola posición.

También es interesante comparar los usos que los testimonios históricos dejan entrever. Los babilonios y egipcios y también los romanos usaron los números sobre todo para fines mercantiles: para contar cantidades de ganado o de bienes en las bóvedas. Los babilonios también los usaron para fines astronómicos: el sistema sexagésimal de los babilonios es el origen de la división del ángulo completa en 360 grados.

Los mayas usaron los números para un calendario preciso. Los chinos, hindues y árabes usaron los números de manera más abstracta: la multiplicación y la división de números es una idea que no se basa en una aplicación directa de la vida cotidiana en aquellos tiempos.

No tiene sentido multiplicar números como 4622 por otro número cualquiera como 7765, si uno denota el número de vacas y otro el número de árboles. Es decir, las operaciones aritméticas en general sólo tienen sentido, si se entiende el número como concepto abstracto, sólo así se pueden motivar las operaciones de multiplicación y división entre todos los números.

Ejercicio. Para comprender mejor las ventajas y desventajas de los distintos sistemas de enumeración sugerimos al lector representar los números 832 y 57 y llevar a cabo la adición y la multiplicación en varios sistemas, sin recurrir al sistema decimal.

Bibliografía

Georges Ifrah: *Historia universal de las cifras*, Editorial Espasa Calpe, 2009.

Johannes Lehmann: *So rechneten Ägypter und Babylonier*, Urania Verlag 1994.

Johannes Lehmann: *So rechneten Griechen und Römer*, Urania Verlag 1994.

Hans Lenz: *Universalgeschichte der Zeit.*, Marix Verlag, 2013.

David Freidel, Linda Schele, Joy Parker: *Maya Cosmos: three thousand years on the shaman's path*, 1993.

Wikipedia en español, inglés y alemán.

2.2. Numerología y Teoría de números.

La teoría de números nace alrededor de 600 A.C. Cuando Pitágoras y sus discípulos comenzaron a estudiar las propiedades de los enteros positivos. Sabían que los tonos musicales estaban relacionadas con razones de números enteros. Por ejemplo si un tono tiene 440 vibraciones por segundo, el que tiene 880 vibraciones por segundo es el mismo tono, solo que una octava más alto. Claro que los pitagóricos no tenían manera de contar el número de vibraciones por segundo de una nota musical, pero sí observaron que si una cuerda producía una nota, al hacer vibrar la misma cuerda a la mitad de su longitud, se producía la misma nota, una octava más alta. Y si se presionaba la cuerda para reducir su tamaño a $3/4$ o $2/3$, los sonidos que se obtenían eran armónicos, mientras que si se presionaba a distancias intermedias que no correspondían a fracciones simples de la longitud original, podían obtenerse tonos disonantes. Los pitagóricos creían que todo en el universo tenía secretos relacionados con razones de números enteros y que si pudieran descubrir estos secretos entonces todo sería claro y podrían escuchar una gran armonía: la música de las estrellas.

Para buscar esos secretos, los pitagóricos clasificaron a los enteros positivos en varios tipos de números:

los pares 2, 4, 6, 8 10, 12, 14, 16, ...

los impares 1, 3, 5, 7, 9, 11, 13, 15, ...

los primos 2, 3, 5, 7, 11, 13, 17, 19, 23, ...

y los compuestos 4, 6, 8, 9, 10, 12, 14, 15, 16, 18, 20, ...

Los griegos sabían que todo número compuesto es producto de primos. Euclides demostró que hay una infinidad de números primos. Hubieran estado encantados si hubieran sabido que los números primos a la larha iban a utilizarse en criptografía y en la codificación y decodificación de mensajes que se enviarían al espacio exterior.

También hicieron otras clasificaciones:

Números perfectos como el $6=1+2+3$ y $28= 1+2+4+7+14$, que son iguales a la suma de todos sus factores enteros. ¿Existen números perfectos que sean impares? Acomodando objetos en patrones poligonales, generaron números que llamaron

triangulares, 1, 3, 6, 10, 21, ... cuadrados, 1, 4, 9, 16, 25, 36, ...

pentagonales, 1, 5, 12, 22, 35, ... hexagonales, 1, 6, 15, 28, 45, ...

etcétera.

Siglos más tarde Pierre Fermat descubrió uno de los profundos secretos de la teoría de números: que todo número entero positivo puede ser escrito como la suma de a lo más tres números triangulares, y de a lo más cuatro números cuadrados y de a lo más cinco números pentagonales, etcétera. Por ejemplo:

$30 = 10 + 10 + 10$ suma de tres números triangulares

$30 = 1 + 4 + 9 + 16$ suma de cuatro números cuadrados

$30 = 1 + 1 + 1 + 5 + 22$ suma de cinco números pentagonales

$30 = 1 + 1 + 1 + 6 + 6 + 15$ suma de seis números hexagonales

Las matemáticas y el mundo, una relación mágica o científica. El interés de los pitagóricos en los números y sus propiedades representa quizás la manifestación más antigua de las matemáticas puras. Se trata de entender por deducción las propiedades de unos entes abstractos (los números) que han adquirido una existencia propia, independiente de sus humildes orígenes como utensilios para ayudar en el quehacer cotidiano. Un hecho que es interesante notar, es que esta primera aparición del interés matemático puro surge unido al misticismo, se atribuyen propiedades mágicas a ciertos números. Cabe preguntar qué tanto ayudó esta tendencia la misticismo a desarrollar las matemáticas. Aparte de esta etapa de los pitagóricos, caen en esta misma línea de misticismo la astrología de Ptolomeo y los esfuerzos de Kepler para explicar el sistema solar con los poliedros regulares. Son intentos de entender el mundo usando las matemáticas pero de una manera poco científica, como esperando descubrir algunas decisiones divinas a las que podemos tener acceso solo porque coinciden con relaciones entre números o entre cuerpos geométricos. Es como suponer que Dios es matemático y se comunica con nosotros a través de los objetos matemáticos pero sin que haya relaciones de causa efecto en los caprichos que llevan a organizar “matemáticamente” el mundo. Esta tendencia mística o mágica hacia entender el mundo a través de las matemáticas contrasta notablemente con la postura más científica de Galileo en donde ya las leyes de la naturaleza están ante nosotros como un libro abierto, no hace falta adivinar, basta leerlo, pero antes hay que conocer las matemáticas y su lenguaje. La gran diferencia es que en esta postura galileana, no hay magia, no hay capricho ni casualidad, es una relación fáctica. El mundo tiene un comportamiento matemático. La idea de Dios deja de ser necesaria, aunque tampoco es necesariamente incompatible.

2.3. La irracionalidad de la raíz cuadrada de dos

Los pitagóricos creían que todo en el universo estaba relacionado con razones de enteros, es decir, que todo, en particular todas las distancias) se podía medir con fracciones. Sin embargo al intentar calcular la diagonal de un cuadrado de lado igual a la unidad, descubrieron que esto no era cierto. Era imposible representar la longitud de esa diagonal con una fracción. Hé aquí un par de demostraciones de este hecho.

Los pitagóricos creían que todo podía representarse con números. Por ejemplo, descubrieron que las notas y melodías producidas eran armónicas, y sonaban bien (en su concepción) cuando había relaciones numéricas enteras simples entre las distancias de los huecos de las flautas o las longitudes de las cuerdas. Una cuerda reducida a la mitad de tamaño, produce un sonido que es una octava más alta que el tono original, y las combinaciones sonoras agradables para el gusto occidental guardan relaciones numéricas simples como $\frac{2}{3}$, $\frac{3}{4}$, $\frac{4}{5}$. Hoy en día decimos que se representan con fracciones

sencillas. Así, también creían que todos los segmentos que se pudieran construir en la geometría, guardaban relaciones fraccionarias entre ellos.

Sin embargo, un día descubrieron una contradicción en esa creencia. La diagonal de un cuadrado no podía guardar una relación fraccionaria con el lado del cuadrado. Fue un descubrimiento que sacudió sus creencias y que, a la larga, les llevaría a construir una noción más precisa y completa de la longitud, cosa que lograron sin ampliar el concepto de número, que es lo que se hace en la actualidad.

He aquí una demostración de que la razón entre la diagonal y el lado de un cuadrado no puede ser una fracción. Supongamos que la relación entre la diagonal y el lado del cuadrado es $\frac{m}{n}$, donde m y n son enteros positivos. Sin pérdida de generalidad podemos suponer que m y n son primos entre sí, ya que al dividir m y n por todos sus factores comunes, la fracción $\frac{m}{n}$ no se altera. Esta suposición equivale a decir que, usando cierta longitud como unidad, la diagonal del cuadrado mide m y su lado mide n . Del teorema de Pitágoras resulta que $m^2 = n^2 + n^2$, es decir, $m^2 = 2n^2$ y por tanto m^2 es un número par. Pero se sabe que sólo los números pares tienen cuadrados pares, por lo que m debe ser par. Por tanto $m = 2k$ para algún entero positivo k . Entonces $m^2 = 4k^2 = 2n^2$. Dividiendo por 2 obtenemos: $2k^2 = n^2$, por tanto n^2 es par y también lo es n . Hemos concluido entonces que tanto m como n son pares, lo cual es una contradicción, pues habíamos postulado que m y n no tenían múltiplos comunes. Esto demuestra que es contradictorio suponer que la razón entre la diagonal y el lado de un cuadrado es una fracción.

Por el teorema de Pitágoras sabemos que el cuadrado construido sobre la diagonal de un cuadrado de lado 1 es 2 y por ello decimos que esa diagonal mide $\sqrt{2}$. La demostración anterior prueba que $\sqrt{2}$ no es una fracción, en lenguaje moderno, $\sqrt{2}$ no es un número racional. Hoy en día decimos que

$$\sqrt{2} \text{ es un número irracional}$$

Pero para poder decir esto fueron necesarios 20 siglos de maduración intelectual que llevaron a la humanidad a concebir o inventar, poco a poco y culminando el proceso a finales del siglo XIX, los *números reales* que es el sistema numérico que permite incluir a los números irracionales.

Este descubrimiento (de que la razón entre la diagonal y el lado de un cuadrado no es una fracción) destruyó completamente la creencia pitagórica de que todo en el universo se podía expresar con razones de enteros positivos. De hecho, el descubrimiento produjo una gran crisis en el ámbito de los pitagóricos, al grado de que quisieron mantenerlo en secreto e incluso hay rumores de quien lo descubrió, o lo publicó por vez primera, Hipaso de Metaponto (siglo V - siglo IV aec), murió en un naufragio, quizás provocado por sus propios compañeros en venganza por haber destruido la teoría en la que habían basado todo el conocimiento; o que se suicidó para expiar su culpa buscando otra vida en la que resolver el problema que había creado.

Los pitagóricos habían basado toda su filosofía en que todo en el mundo era número. Su descubrimiento de que la música y los números estaban relacionados les entusiasmó de tal manera que creyeron ver en ello una señal mística de que todo en el mundo era así: relaciones entre números. El hecho de que la relación entre dos magnitudes tan simples como el lado y la diagonal de un cuadrado no cumplieran con esta ley, trastornaba toda su filosofía. De hecho la invalidaba. No supieron ver que ese hecho constituía un nuevo reto que hubieran podido enfrentar ampliando el concepto de número. Y es que ellos veían a los números (enteros positivos) como algo dado al hombre junto con su propia existencia, posiblemente por alguna divinidad, en lugar de verlos como una creación humana que, como tal, era perfectamente ampliable. Lamentablemente esa creencia, la de la existencia

a priori de los números enteros positivos, sigue vigente hoy en día entre la gente común, lo cual le dificulta mucho apropiarse de las matemáticas modernas. Darse cuenta de que tanto los números enteros positivos como los racionales, los reales, que incluyen a los irracionales, y los complejos son creaciones humanas, abre la puerta de entrada al mundo de las matemáticas modernas.

Quien enfrentó por primera vez y resolvió parcialmente este conflicto, fue el gran matemático Eudoxo de Cnido (390 - 337 aec), creando una aritmética de segmentos que permitía operar con ellos, con absoluta claridad y precisión, sin asignarles valores numéricos como longitudes. Decimos que lo resolvió parcialmente porque no generalizó el concepto de número sino que desarrolló una manera de hablar y manipular de esas magnitudes que podían ser inconmensurables, sin asignarles valores numéricos. Sin embargo el paso que le faltó no era de carácter técnico sino puramente filosófico. Para llegar a la solución moderna sólo faltaba definir esas magnitudes como números, que es, en el fondo, lo que hicieron dos milenios más tarde personajes como Weierstrass, Dedekind y Cauchy para definir los números reales. Pero el hecho de que este paso hubiera tardado dos milenios en darse nos da una clara indicación de que se trata de un salto conceptual muy importante y nada fácil. Es pues, natural, que su aprendizaje cause tantas dificultades a los estudiantes del mundo moderno.

Los pitagóricos no podían aceptar que raíz cuadrada de 2 es un número ni una razón entre números, pero tampoco podían negar que era la medida de la diagonal del cuadrado de lado uno, lo decía su teorema más sólido. Los problemas lógicos generados por el descubrimiento de la irracionalidad de $\sqrt{2}$ fueron resueltos por EUDOXO DE CNIDO, uno de los más grandes matemáticos de la antigüedad griega, de una manera muy ingeniosa, creando lo que se conoció como la *Teoría de las proporciones*, que EUCLIDES presenta en el Libro V de *Los elementos*. En esta teoría se construyen *proporciones* que pueden usarse en lugar de los números y que incluyen a la razón entre la diagonal y el lado del cuadrado, o sea, a $\sqrt{2}$. Sin embargo, a cambio, estas proporciones, es decir estos *nuevos números*, no siempre pueden representarse por fracciones. Muchos de ellos no son fracciones, y aún así pueden usarse para realizar operaciones con ellos mediante un *álgebra geométrica* que se presenta en el capítulo de Geometría.

Siglos después, a finales del XIX, una construcción fue utilizada por Dedekind (y otra por Cauchy) para formalizar el concepto de número real, que es el que abarca tanto a los racionales como a los irracionales. También poco más o menos en la misma época David Hilbert reorganiza los fundamentos de la Geometría usando el Teorema de Pitágoras como fundamento. Así que el Teorema de Pitágoras es una de las piezas más importantes en la historia de las matemáticas. No solo es uno de los primeros teoremas descubiertos por varias civilizaciones, de los primeros en ser demostrados lógicamente y rigurosamente, sino también fue pieza clave en el descubrimiento de los números irracionales y se convirtió en el fundamento esencial de las matemáticas modernas. Otra pieza clave en esta historia, igualmente importante, tiene que ver no con triángulos rectángulos sino con círculos, se trata del número π .

Las matemáticas y el mundo: ¿una relación mágica o científica?

El interés de los pitagóricos en los números y sus propiedades representa quizás la manifestación más antigua de las matemáticas puras. Se trata de entender, por deducción lógica, las propiedades de unos entes abstractos (los números) que han adquirido una existencia propia, independiente de sus humildes orígenes como utensilios para ayudar en el quehacer cotidiano. Un hecho que es interesante notar, es que esta primera aparición del interés matemático puro surge unido al

misticismo, se atribuyen propiedades mágicas a ciertos números. Cabe preguntar qué tanto ayudó esta tendencia al misticismo a desarrollar las matemáticas. Aparte de esta etapa de los pitagóricos, caen en esta misma línea de misticismo la astrología de PTOLOMEO y los esfuerzos de KEPLER para explicar el sistema solar con los poliedros regulares. Son intentos de entender el mundo usando las matemáticas pero de una manera poco científica, como esperando descubrir algunas decisiones divinas a las que podemos tener acceso solo porque coinciden con relaciones entre números o entre cuerpos geométricos. Es como suponer que Dios es matemático y se comunica con nosotros a través de los objetos matemáticos pero sin que haya relaciones de causa efecto en los caprichos que llevan a organizar “matemáticamente” el mundo. Esta tendencia mística o mágica hacia entender el mundo a través de las matemáticas contrasta notablemente con la postura más científica de GALILEO GALILEI en donde ya las leyes de la naturaleza están ante nosotros como un libro abierto, no hace falta adivinar, basta leerlo, pero antes hay que conocer el lenguaje en que están escritas: las matemáticas. La gran diferencia es que en la postura galileana, no hay magia, no hay capricho divino ni casualidad, es una relación fáctica. El mundo tiene un comportamiento matemático. La idea de Dios deja de ser necesaria, aunque no por ello se necesariamente incompatible.

2.4. La Teoría de Números

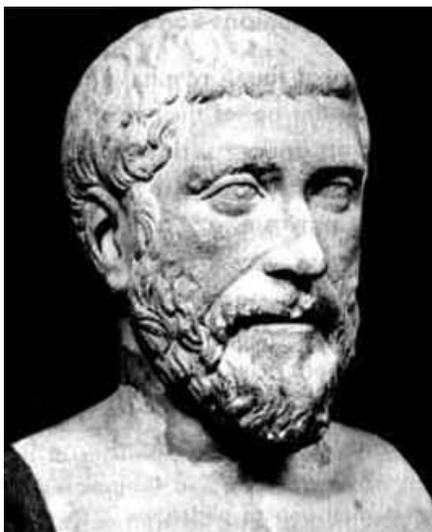
Números se ocupaban en todas las culturas avanzadas, ver la sección 2.1 sobre los diferentes sistemas de numeración. Pero fue hasta en los griegos que se empezaron a estudiar los números por si mismo, es decir en propiedades que tienen los números respecto a las operaciones que son la adición y la multiplicación.

Un número era según TALES una colección de unidades. Por ello no se consideraba al 1 un número. Y ARISTÓTELES dijo que el 1 no debería considerarse número ya que es la unidad y sirve para medir. Y lo que se mide y con lo que se mide no puede ser lo mismo. También había el pensamiento de que el 2, la *díada*, es la característica de ser par, por lo que el 2 a veces tampoco se consideraba un número, sino una característica de números.

De ello se ve como los griegos empezaron a estudiar las propiedades y aunque difieren en las definiciones, encontraron muchos resultados de la teoría de números.

Los pitagóricos

La teoría de números nace alrededor de 550 a.e.c. cuando PITÁGORAS DE SAMOS (ca. 569–475 a.e.c) y sus discípulos comenzaron a estudiar las propiedades de los enteros positivos.



PITÁGORAS DE SAMOS

Sobre la vida de PITÁGORAS se sabe poco. Nació en la isla de Samos cerca de la costa de la actual Turquía. Su padre era aristócrata o negociante. Se mudó a Croton, en Sicilia donde fundó un tipo de escuela. Sobre la característica de la escuela hay opiniones divergentes entre los historiadores. Algunos piensan que era una escuela abierta parecida a la escuela de PLATÓN, otros opinan que era muy cerrada como una secta en las cuales los discípulos tenían que vivir según reglas estrictas. El problema es que las fuentes que hoy se tienen de ella, son romanos y datan alrededor de 500 años posteriores. Hay indicios que aceptaban hombres igual que mujeres entre sus discípulos y que eran vegetarianos.

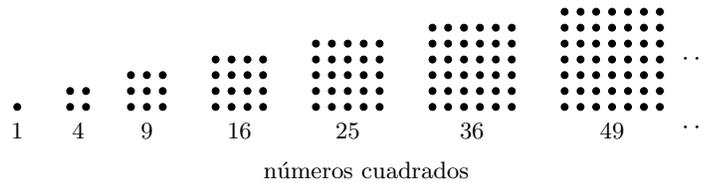
De lo que se cuenta de la escuela de PITÁGORAS es que se interesaban en las matemáticas, en particular en los números y la geometría y con la relación de estos conceptos con el cosmos. Reducido al máximo, el lema de los pitagóricos era: *todo es números*. Hay indicios que su pensamiento todavía estaba dominado mucho por la mística, es decir en creencias de interrelaciones y conexiones que sólo se asumen. De lo que quedó mucha evidencia es que fue una escuela sumamente influyente, poco tiempo después de fundar la escuela, la hermandad de los pitagóricos dominaban la ciudad y pronto una parte importante de la extensión de los griegos. Hacia el final de su vida, expulsaron a Pitágoras de Croton. Él y sus discípulos se fueron al exilio. Influyeron el desarrollo de las ciencias de manera profunda.

Podemos ver la manera cómo relacionaban la música con los números. Sabían que los tonos musicales estaban relacionadas con razones de números enteros. Por ejemplo si un tono tiene 440 vibraciones por segundo, el que tiene 880 vibraciones por segundo es el mismo tono, solo que una octava más alto. Claro que los pitagóricos no tenían manera de contar el número de vibraciones por segundo de una nota musical, pero sí observaron que si una cuerda producía una nota, al hacer vibrar la misma cuerda a la mitad de su longitud, se producía la misma nota, una octava más alta. Y si se presionaba la cuerda para reducir su tamaño a $\frac{3}{4}$ o $\frac{2}{3}$, los sonidos que se obtenían eran armónicos, mientras que si se presionaba a distancias intermedias que no correspondían a fracciones simples de la longitud original, podían obtenerse tonos disonantes. Los pitagóricos creían que todo en el universo tenía secretos relacionados con razones de números enteros y que si pudieran descubrir

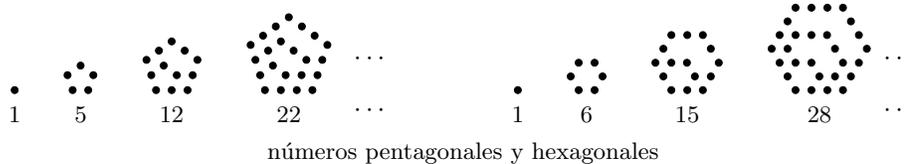
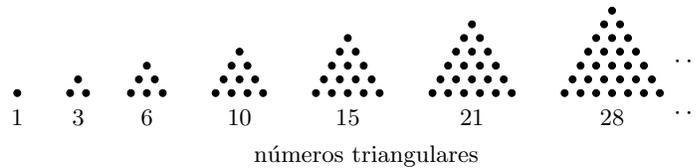
estos secretos entonces todo sería claro y podrían escuchar una gran armonía: la música de las estrellas.

Para buscar esos secretos, los pitagóricos clasificaron a los enteros positivos en varios tipos de números: distinguían entre *pares*, esto son los números que se pueden separar en dos partes iguales, e *impares*, en los que esta separación no es posible.

A parte de los números *cuadrados*



consideraban también números *triangulares*, *pentagonales* y *hexagonales* etcétera.



2.5. El estudio de los números naturales en Euclides

Se piensa que el material que trata el estudio de los números naturales en *Los Elementos* de EUCLIDES, esto son los tomos VII, VIII y IX, se deben a ARQUITAS DE TARENTO (430–360). Esto porque se encontró un fragmento de Arquitas exhibiendo una demostración que es esencialmente idéntica a una demostración en Euclides.

Veamos ahora algunos resultados expuestos en *Los Elementos*. EUCLIDES trató los números como segmentos que son medibles por una unidad dada. Por ello el lenguaje usado en los EUCLIDES es algo difícil de leer. Por esta razón usaremos un lenguaje y una notación moderna para dar las demostraciones, pero trataremos de apegarnos a la idea original lo más posible.

Prop. VII.1: *Dados dos números desiguales y restando sucesivamente el menor del mayor, si el que queda no mide nunca al anterior hasta que quede una unidad, los números iniciales serán primos relativos.*

Demostración. Es por reductio ad absurdum, es decir, se supone que si hay un número $e > 1$ que mide ambos números dados a y b . Sea $a > b$. Entonces también $a - b$ es divisible entre e . Esto se puede repetir las veces que se quita el menor número del mayor. Cada vez, la diferencia que se obtiene también será divisible entre e . Como la última diferencia que se obtiene es la unidad, e debe dividir la unidad, que es absurdo ya que $e > 1$.

Comentarios. En la demostración original de Euclides, se lleva a cabo el quitar el menor número del mayor tres veces de manera explícita como para dar un ejemplo.

El método de quitar el menor número del mayor – llámemosle *quita-e-intercambia* para poder referirnos a él –, era conocido antes. Así se encontraron por primera vez que hay segmentos que no son *commensurables*, ver la Sección ??, como por ejemplo el lado y la diagonal en un pentágono regular o en un cuadrado.

Prop. VII.2: *Dados dos números que no son primos relativos, hallar el máximo número que divide a los dos números (el máximo común divisor).*

Demostración. Sean $a > b$ los dos números dados. Si b divide a a , entonces b es divisor común y también es máximo ya que no puede haber uno mayor que b .

Si b no divide a a entonces se quita repetitivamente el menor número del mayor. Sea e el último número que se obtiene. Entonces e tienen que ser mayor que 1, sino a y b serían primos relativos por la Prop. VII.1. Así que $e > 1$ es un divisor común de a y b .

Luego demuestra por reductio ad absurdum que e es el máximo de los divisores comunes, muy similar a la demostración de la Prop. VII.1.

Comentario. En esencia el argumento aquí es igual que en la Prop. VII.1, pero EUCLIDES tiene que diferenciar los dos casos porque el resultado, hoy lo llamaríamos *máximo común divisor*, en la Prop. VII.1 es la unidad (que no es un número) mientras en la Prop VII.2 es un número.

Este resultado se conoce hoy como *Algoritmo de EUCLIDES*. Los algoritmos son procedimientos para calcular algo o para obtener algún resultado. Los algoritmos más simples hoy en día, gracias a la notación algebraica, pueden expresarse mediante fórmulas. Por ejemplo, el algoritmo para calcular el área de un triángulo es $\frac{\text{base} \times \text{altura}}{2}$, una fórmula, y para calcular la superficie de una esfera es $\frac{4}{3}\pi \cdot r^2$, otra fórmula.

Pero no todos los algoritmos pueden expresarse mediante fórmulas cerradas. El cálculo del máximo

común divisor de dos números enteros positivos fue quizás el primer algoritmo en la historia que no puede describirse mediante una fórmula.

Veamos ahora cómo funciona el Algoritmo de ECULIDES en un ejemplo: buscamos el máximo común divisor de $a = 263$ y $b = 98$. Denotamos el par como $(273, 98)$. Ahora siempre quitamos el menor del mayor.

$$\begin{array}{ll}
 (273,98) & 273 - 98 = 175 \\
 (175,98) & 175 - 98 = 77 \\
 (98,77) & 98 - 77 = 21 \\
 (77,21) & 77 - 21 = 56 \\
 (56,21) & 56 - 21 = 35 \\
 (35,21) & 35 - 21 = 14 \\
 (21,14) & 21 - 14 = 7 \\
 (14,7) & 14 - 7 = 7 \\
 (7,7) & 7 - 7 = 0.
 \end{array}$$

El máximo común divisor de 273 y 98 es 7. Este algoritmo es fácil de implementar y además es muy rápido. Muchas veces se acorta el número de pasos al dividir el mayor entre el menor y tomando el resto en cada paso. En el ejemplo anterior se tendría

$$\begin{array}{ll}
 (273,98) & 273 = 2 \cdot 98 + 77 \\
 (98,77) & 98 = 1 \cdot 77 + 21 \\
 (77,21) & 77 = 3 \cdot 21 + 14 \\
 (21,14) & 21 = 1 \cdot 14 + 7 \\
 (14,7) & 14 = 2 \cdot 7 + 0.
 \end{array}$$

La importancia práctica de este algoritmo radica en que es rápido, el resultado puede obtenerse realizando pocas operaciones, aún cuando los números involucrados puedan ser muy grandes.

Prop. VII.31: *Cada número compuesto se divide por algún número primo.*

Demostración. Sea a el número compuesto dado. Como a es compuesto, es divisible por algún número $b > 1$. Si b es primo entonces ya demostramos la afirmación. Si b no es primo, entonces es compuesto y por lo tanto existe algún divisor $c > 1$. Si c es primo, entonces c divide a b y como b divide a a entonces c divide a a y demostramos la afirmación. Si c no es primo entonces se puede seguir de la misma manera. Finalmente se encontrará un número primo que divide a .

Prop. VII.32: *Cada número (mayor que 1) o es primo o se divide por algún número primo.*

Demostración. Sea $a > 1$ un número. Entonces a es primo o a es compuesto. En el segundo caso sigue de la Prop. VII.31 que a es divisible entre un número primo.

Comentario. Como consecuencia se obtiene que todo número es producto de primos.

Prop. IX.20: *Hay más números primos que cualquier número dado de números primos.*

Reformulación moderna. Hay una infinidad de números primos. En aquel entonces no se habló nunca de una “infinidad” de cosas.

Demostración. Sean p_1, \dots, p_n los números primos que están dados. Consideramos el producto $x = p_1 \cdot p_2 \cdot \dots \cdot p_n$ y su sucesor $y = x + 1$. Por la Prop. VII.32 el número y es o primo o divisible por un primo. Si y es primo, entonces, como $y > p_i$ para cualquier $i = 1, \dots, n$, hay al menos un número primo más que los que estaban dados. Si y es divisible entre un primo, entonces sea z un primo que divide y .

En el tomo VIII de *Los Elementos* habla de sucesiones geométricas. Esto son sucesiones, en los cuales la proporción entre dos números consecutivos es constante. Por ejemplo 3, 12, 48, 172 o 32, 8, 2, $\frac{1}{2}$ son sucesiones geométricas de cuatro números. En tiempos de los griegos no había sucesiones infinitas, sino cada sucesión tenía un número finito de miembros.

Prop. IX.36: *Si se forma una sucesión geométrica a partir de la unidad en donde los términos están en la proporción 1:2 y si la suma de todos estos números es un número primo entonces es el producto de esta suma con el último número de la sucesión es un número perfecto.*

Reformulación moderna. Si $p = 1 + 2 + 2^2 + \dots + 2^{n-1}$ es primo, entonces $x = 2^{n-1}p$ es perfecto.

Demostración. La suma $1 + 2 + 4 + \dots + 2^{n-1}$ es igual a $p = 2^n - 1$ y se supone que p es primo. El número es entonces $x = (2^n - 1) \cdot 2^{n-1} = 2^{n-1}p$. Sus divisores son

$$1, 2, \dots, 2^{n-2}, 2^{n-1}, p, 2p, \dots, 2^{n-2}p, 2^{n-1}p = x.$$

La suma alícuota s es

$$\begin{aligned} s &= \underbrace{1 + 2 + \dots + 2^{n-2} + 2^{n-1}}_{=2^n - 1 = p} + \underbrace{p + 2p + \dots + 2^{n-2}p}_{(2^{n-1} - 1)p} \\ &= 2^{n-1}p \\ &= x, \end{aligned}$$

lo que demuestra que x es perfecto.

Comentario. LEONHARD EULER demostró que cada número perfecto que es par tiene que ser de esta forma.

Desarrollos en el renacimiento y después

Siglos más tarde PIERRE FERMAT (1601–1665) descubrió uno de los profundos secretos de la teoría de números: que todo número entero positivo puede ser escrito como la suma de a lo más tres números triangulares, y de a lo más cuatro números cuadrados y de a lo más cinco números pentagonales, etcétera.

Por ejemplo:

$30 = 10 + 10 + 10$	suma de tres números triangulares
$30 = 1 + 4 + 9 + 16$	suma de cuatro números cuadrados
$30 = 1 + 1 + 1 + 5 + 22$	suma de cinco números pentagonales
$30 = 1 + 1 + 1 + 6 + 6 + 15$	suma de seis números hexagonales

Pero este resultado no lo demostró. En 1770 JOSEPH LOUIS LAGRANGE demostró el resultado para cuadrados. En 1796 CARL FRIEDRICH GAUSS demostró el caso para números triangulares. La versión general la demostró AUGUSTIN LOUIS CAUCHY finalmente en 1813. Esto muestra una característica de las matemáticas: muchas veces sólo por el esfuerzo conjunto de varias personas es posible avanzar en terrenos difíciles.

Fermat también pensaba que había encontrado una fórmula para primos. Pensó que $F_n = 2^{(2^n)} + 1$ es primo para cualquier entero $n \geq 0$. En efecto

$$\begin{aligned}F_0 &= 2^{2^0} + 1 = 2^1 + 1 = 3, \\F_1 &= 2^{2^1} + 1 = 2^2 + 1 = 5, \\F_2 &= 2^{2^2} + 1 = 2^4 + 1 = 17, \\F_3 &= 2^{2^3} + 1 = 2^8 + 1 = 129, \\F_4 &= 2^{2^4} + 1 = 2^{16} + 1 = 65\,537\end{aligned}$$

todos son primos, pero

$$F_5 = 2^{2^5} + 1 = 2^{32} + 1 = 4\,294\,967\,297 = 641 \cdot 6\,700\,417$$

no lo es, como demostó LEONHARD EULER como 100 años después. Hoy se sabe que F_5, \dots, F_{32} no son primos, pero no se sabe de F_{33} . El número F_{33} es simplemente demasiado grande, para poder decidirlo. No se sabe si hay más *Números de Fermat*, es decir, más números de la forma $F_n = 2^{(2^n)} + 1$ que sean primos.

Hay una serie de enunciados sobre números naturales, que son fáciles de enunciar pero muy difíciles de demostrar.

El último Teorema de FERMAT: *La ecuación $a^n + b^n = c^n$ no tiene soluciones en los números naturales si $n > 2$.*

Fue formulado en 1637 por PIERRE FERMAT y demostrado en 1995 por ANDREW WILES.

La Conjetura de GOLDBACH: *Cada número par mayor que 2 es suma de dos primos.*

Fue formulado en 1742 por CHRISTIAN GOLDBACH y no se ha podido demostrar.

La Conjetura de CATALAN: *Solo existen dos potencias de números naturales que están a distancia 1: esto son 3^2 y 2^3 .*

Fue formulado en 1844 por EUGÈNE CHARLES CATALAN y demostrado en 2002 por PREDĂ MIHĂILESCU.

La Conjetura de primos gemelos: *Hay un número infinito de pares de primos que están a distancia 2 como 11 y 13 o 41 y 43.*

En 1849 ALPHONSE DE POLIGNAC formuló una conjetura más general: que para cada número natural k hay un número infinito de pares de primos p y p' tales que $p' - p = 2k$. Para $k = 1$ se obtiene la Conjetura de primos gemelos. No se ha podido demostrar.

Primos de MERSENNE: *Hay un número infinito de primos de MERSENNE, esto son primos de la forma $2^n - 1$. No es difícil ver que si $2^n - 1$ es primo, entonces n tiene que ser primo, ya que*

$$2^{ab} - 1 = (2^a - 1)(1 + 2^a + 2^{2a} + 2^{3a} + \dots + 2^{(b-1)a}).$$

Si se podría demostrar que hay un número infinito de primos de MERSENNE, entonces por la Prop. IX.36 habría un número infinito de números perfectos.

Perfectos impares: *¿Existen números perfectos impares?* No se sabe hasta la fecha si esto es cierto o falso.

Hay muchas preguntas más que están abiertas. Algunas involucran también conceptos mucho más recientes como por ejemplo la *Función zeta de RIEMANN*. Se define como

$$\zeta(s) = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \dots$$

para cualquier número complejo s . Se sabe que $\zeta(s) = 0$ para cualquier número entero negativo, esto se llaman los “zeros triviales” de la función ζ . La siguiente conjetura conforma uno de los problemas de milenio, cuya resolución está dotado de mundial reconocimiento además de un premio de un millón de dólares estadounidenses.

Hipótesis de RIEMANN: *Los zeros no triviales de ζ tienen parte real de $\frac{1}{2}$.* Propuesto en 1859 por BERNHARD RIEMANN. Sigue abierto desde entonces.

Si se podría demostrar la Hipótesis de RIEMANN, se obtendrá como consecuencia una estimación precisa sobre el número de primos que son menores que un número dado. Sea $\Pi(k)$ el número de primos p con $p \leq k$. Por ejemplo $\Pi(20) = 8$ ya que los primos p con $p \leq 20$ son 2, 3, 5, 7, 11, 13, 17, 19. Se sabe que para grandes k se tiene $\Pi(k) \approx \frac{k}{\ln(k)}$, más preciso, que

$$\lim_{k \rightarrow \infty} \frac{\Pi(k)}{\frac{k}{\ln(k)}} = 1.$$

Si la Hipótesis de RIEMANN sería cierta, entonces se podría estimar la diferencia entre las dos funciones.

La dificultad de encontrar la factorización en primos de un número natural hoy se aprovecha para encriptar y desencriptar mensajes. Se basa en la teoría de números y el hecho que es posible encriptar a partir del producto pq de dos primos p y q pero sólo es posible desencriptar si se conocen los dos factores. De esta manera, teoría de números fue en un inicio el estudio del mundo, se volvió un campo de estudio muy alejado de la realidad y en la era digital se ha vuelto nuevamente un área aplicado e intensamente usado.

Capítulo 3

Geometría: origen y fundamentos

El estudio de las propiedades de nuestro espacio

La experiencia cotidiana del espacio que habitamos, las formas más comunes de los objetos que co-habitan con nosotros en ese espacio, las simetrías que exhiben, es decir las propiedades que conservan ante ciertas transformaciones como reflexiones, rotaciones, traslaciones, expansiones y contracciones, etcétera, lleva de manera natural a conceptos como la de igualdad o congruencia de algunas figuras, la semejanza de otras, el paralelismo y la perpendicularidad entre líneas rectas y el reconocimiento de figuras especiales con simetrías muy particulares como el cuadrado, y la circunferencia. El estudio de las propiedades y relaciones entre todos estos conceptos es lo que constituye la Geometría euclidiana o Geometría clásica. A pesar de que la palabra geometría en griego significa medición de la tierra (queriendo decir medición de terrenos), la medición es solo un aspecto particular de la geometría, del que nos ocupamos a fondo en el siguiente capítulo, y que se distingue de los temas que trataremos en éste en que aquel requiere de un sistema de numeración. En este capítulo nos ocupamos de los aspectos de la geometría que no requieren de un sistema de numeración, aspectos que los antiguos griegos desarrollaron magistralmente.

3.1. Los orígenes del pensamiento geométrico

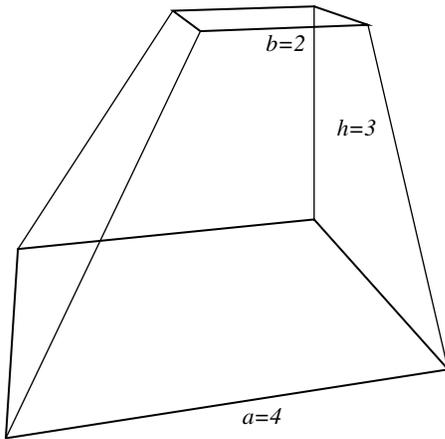
Los rudimentos de un pensamiento pre-geométrico aparecen en todas las grandes culturas de la antigüedad como la egipcia, la babilonia y la china, pero en forma muy rudimentaria y siempre vinculado a las actividades de medición y cuantificación, como en los casos de longitud, área y volumen. De hecho la palabra Geometría, según sus raíces griegas significa “medición de la tierra”. La cultura Sumeria dejó testimonios de sus intereses y conocimientos en miles de tablillas de arcilla y hay muy poco en ellas que pueda identificarse con una actividad matemática, y mucho menos geométrica. En la cultura de Babilonia, cuyos avances en cuestiones de aritmética y álgebra son indudables, las manifestaciones geométricas son muy escasas. Debido a las grandes obras que realizaban como presas y canales de riego, se enfrentaban a la necesidad de calcular cuántos tabiques

iban a necesitar o cuánto tiempo iba a llevar la excavación de un canal y para ello tenían que estimar áreas y volúmenes, pero en general solo lo hacían por interés administrativo y no parece que hayan desarrollado conceptos ni ideas propiamente geométricas. No usaron conceptos geométricos reconocibles ni siquiera en sus cálculos astronómicos, a pesar de que fueron capaces de algunos muy precisos y sofisticados que equivalen a aproximaciones polinomiales.

Para encontrar ideas que podamos considerar corresponden al *pensamiento geométrico* hay que volver la mirada a la cultura egipcia. Y aún allí encontramos muy poco. En los papiros egipcios aparecen varios ejemplos de uso de las matemáticas en relación con algunas figuras geométricas, en especial con las pirámides, pero relacionados siempre con un interés práctico como el cálculo de un volumen. Por ejemplo en el papiro Rhind, que data de aproximadamente 1650 a.e.c y está basado en otro aún más antiguo, contiene algunos problemas relacionados con la geometría. El ítem 48 sugiere el siguiente procedimiento para calcular el área de una circunferencia: *Restar al diámetro una novena parte y elevar al cuadrado* lo que equivalente a la fórmula $A = \left(\frac{8}{9}d\right)^2 = \frac{64}{81}d^2$ para el área de un círculo de diámetro d . Se trata de una *cuadratura del círculo* pues $\frac{8}{9}d$ es el lado de un cuadrado cuya área es *casi* igual a la del círculo de diámetro d . En particular, podríamos pensar que esta fórmula propone un valor de $\pi = 4(64/81) = 256/81 = 3.16049\dots$, que es una aproximación bastante aceptable para aquella época. Sin embargo no hay nada que indique que esto sea algo más que un procedimiento para estimar el área de un círculo. No hay evidencias de un razonamiento deductivo ni de que los egipcios a esas alturas tuvieran una noción clara del número π como *la razón del perímetro al diámetro en una circunferencia* o como *de la razón del área de un círculo al cuadrado de su radio*. En la sección 5.5 se analiza el significado de esta fórmula con más detenimiento.

En el ítem 41 del papiro Rhind, el escriba calcula el volumen de un granero cilíndrico multiplicando el área de la base circular, es decir la del cuadrado cuyo lado es $\frac{8}{9}$ del diámetro por la altura. El procedimiento sería totalmente correcto si hubiera tenido la fórmula exacta para el área del círculo. El ítem 42 ofrece un método alternativo para llegar al mismo resultado: calcula el área de un cuadrado de lado $\frac{4}{3}$ del diámetro de la base y la multiplica por $\frac{2}{3}$ de la altura. El resultado es el mismo porque dos tercios del cuadrado de $\frac{4}{3}$ es igual a $\frac{32}{27}$ y por tanto igual a $\frac{64}{81}$. El hecho de querer ofrecer esta segunda alternativa sugiere un interés en proporcionar un procedimiento que requiere menos operaciones y usa números más sencillos. No parece desear explicar por qué funciona la fórmula, ésta se ofrece tal cual. Solo se trata de ayudar al futuro usuario a realizar los cálculos con mayor facilidad. Esta actitud no es reprochable, después de todo, realizar las operaciones para completar esos cálculos no era tarea fácil cuando se usaba una notación numérica como la egipcia. Seguramente cualquier simplificación en las operaciones era muy útil y apreciada.

El papiro de Moscú, transcrito durante la décima tercera dinastía a partir de originales que se remontan más o menos a 1850 a.e.c., contiene un ejemplo muy revelador del nivel de las matemáticas egipcias de aquella época. Se trata del ítem 14, que ofrece el cálculo del volumen de una pirámide truncada. El papiro incluye un esquema que nos permite intuir que se trata de un *tronco* de pirámide como el que representa figura adjunta.



Como de costumbre, el escriba procede a realizar el cálculo indicando paso a paso las operaciones que hay que hacer para llegar al resultado. En ellas claramente se utilizan los números 2, 4 y 3 como representantes genéricos de los lados de las bases cuadradas y la altura, que hoy día denotaríamos, por ejemplo, por a , b y h , respectivamente. Este hecho lo reconocemos en que los datos son números muy sencillos y diferentes entre sí, lo cual facilita la identificación de lo que representan y permite usar el texto como guía para calcular el volumen de otras pirámides como ésta pero con diferentes valores de a , b y h . El procedimiento descrito por el escriba equivale, en nuestro lenguaje algebraico, a la fórmula

$$\text{Volumen} = \frac{(a^2 + ab + b^2)h}{3}$$

que es totalmente correcta. No se trata solo de una buena aproximación sino de un resultado que sigue siendo válido tres mil años después de que se escribió, además de completamente general, pues no depende de que los datos guarden alguna relación especial entre ellos, como por ejemplo $a = 2b$, aunque así suceda con los del ejemplo concreto.

¿Significa esto que los egipcios supieron deducir racionalmente la fórmula del volumen de la pirámide truncada? Tal deducción, de cualquier manera que se haga, involucra muchos conocimientos previos. Por ejemplo, una manera de obtenerla consiste en conocer (habiéndola deducido antes) la fórmula para calcular el volumen de una pirámide de base cuadrada:

$$\text{Volumen} = \frac{b^2h}{3}$$

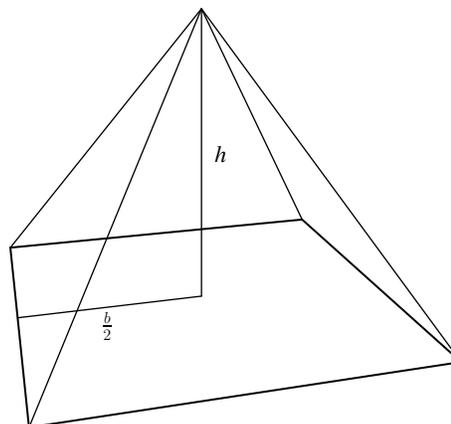
donde b es el lado de la base y h la altura, es decir la distancia perpendicular a la base desde el vértice superior; luego utilizar un razonamiento geométrico como el de concebir la pirámide truncada como la diferencia de dos pirámides completas, la más pequeña contenida en la mayor. (Probablemente los egipcios conocían la fórmula para calcular el volumen de una pirámide completa, pero no sabemos si la obtuvieron mediante un razonamiento o de una manera más o menos empírica.) Por otro lado, el desarrollo algebraico no es inmediato puesto que no se tiene el dato de la altura de la pirámide grande que hay que trincar, sino que hay que obtenerlo a partir de los valores a , b y h de la pirámide truncada, por ejemplo utilizando semejanza de triángulos. Finalmente hay que realizar una simplificación algebraica que involucra una factorización nada trivial.

Ejercicio. Invitamos al lector a realizar este ejercicio para que pueda apreciar las dificultades involucradas en la deducción, sobre todo tomando en cuenta la carencia de un lenguaje algebraico como el nuestro.

¿Pudieron haber realizado todo ese trabajo deductivo los egipcios? Es posible. Lo que está claro es que no tenemos ningún documento que lo compruebe. Y sin embargo también es evidente que sabían muchas cosas sobre geometría y es difícil imaginar que no practicaran algún tipo de razonamiento lógico cuando pensaban en ellas. Nuestra conclusión es que los egipcios posiblemente realizaron algunos razonamientos geométricos lógico-deductivos, pero no supieron o no vieron la conveniencia

de escribirlos, aunque también cabe la posibilidad de que los hayan asentado en papiros que están perdidos y, posiblemente, nunca aparezcan.

Otro ejemplo que parece indicar alguna evidencia de un pensamiento geométrico abstracto en los antiguos egipcios son los ítems 56 y 57 del papiro Rhind en los que se define el *seqed* de una pirámide como la razón de la mitad del lado de la base a la altura (en el esquema: $seqed = \frac{b}{2} : h$). Es un claro ejemplo de un concepto abstracto (el de razón) y de su aplicación a la pendiente de un plano inclinado. Ofrece una manera de manipular la pendiente (el *seqed* resulta ser la razón inversa de la que hoy llamamos pendiente) que se usa para obtener la altura de una pirámide conociendo su lado y su *seqed*. Esto último demuestra un manejo de la semejanza de triángulos y uno de los primeros ejemplos de lo que luego vino a llamarse trigonometría.



No es de extrañarse que cuando Tales de Mileto, y más tarde Pitágoras, viajaran por Egipto, aprendieran este tipo de geometría que, aunque muy básica, ya utilizaba abstracciones como la razón de dos magnitudes (evidenciada en el concepto de *seqed*), y posiblemente algún razonamiento deductivo, como el que sugiere el cálculo del volumen de la pirámide trunca. Al regresar de sus viajes por Egipto y Mesopotamia, probablemente reflexionaron sobre esos conceptos e ideas y así iniciaron el camino hacia el pensamiento geométrico que tanto ha contribuido al progreso de la humanidad.

Es cierto que babilonios, egipcios, chinos, indios y tal vez otros pueblos de la antigüedad descubrieron, de manera independiente unos de otros, lo que hoy se identifica como el Teorema de PITÁGORAS. Hay muestras (no siempre claras por cierto¹), de que conocían algunas “ternas pitagóricas”, es decir ternas de números enteros a, b, c tales que $a^2 + b^2 = c^2$ (como 3, 4 y 5) y sabían que al formar un triángulo cuyos lados constituyeran una de ellas, éste resultaba ser un triángulo rectángulo. Incluso es posible que utilizaran esas ternas en la edificación de monumentos y edificios como un método para construir paredes perpendiculares entre sí. Sin embargo no hay nada registrado en sus legados anteriores a la época de oro de las matemáticas griegas que indique que conocieran una demostración del teorema (no siquiera de que se plantearan que tal demostración fuera posible, conveniente o necesaria).

Todo esto nos lleva a afirmar que el pensamiento geométrico como tal nació en la antigua Grecia alrededor de siglo V a.e.c.

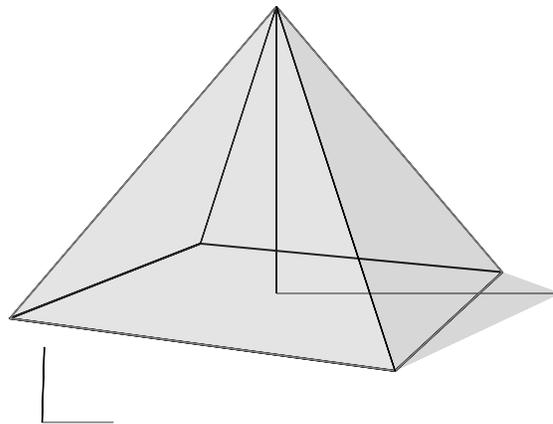
¹La famosa tableta Primpton 322 que durante muchos años se creyó que mostraba una lista de elementos de algunas ternas pitagóricas, ha sido interpretada de manera diferente y más plausible como los resultados de la solución algebraica del problema algebraico $x - \frac{1}{x} = n$ para diferentes valores de n

3.2. El pensamiento geométrico en la antigua Grecia

Los orígenes del pensamiento geométrico

Contamos con muy pocos documentos que revelen los detalles de la geometría griega desde su origen en el siglo VII a.e.c. atribuido a Tales de Mileto (625-546 a.e.c.) hasta el siglo III a.e.c. en que aparecen *Los elementos* de Euclides. Lo que sabemos es que muchos filósofos se ocuparon de las Matemáticas y especialmente de la Geometría y que ésta última aparentemente se desarrolló bastante alejada de las aplicaciones prácticas, aunque nadie niega que seguramente fue usada en ellas. Pero los documentos de la época revelan si no un desdén, sí una separación entre el pensamiento geométrico y la actividad productiva en general.

El detalle más antiguo del que tenemos referencia en el que se puede apreciar un pensamiento geométrico deductivo es el relato del procedimiento que Tales de Mileto sugirió para calcular la altura de una pirámide: esperar el momento en que la sombra de una vara vertical mide lo mismo que la vara, entonces la sombra de la pirámide mide lo mismo que su altura. Éste es un claro ejemplo de pensamiento deductivo. Se construyen, imaginariamente, dos triángulos, uno usando la vara y su sombra y el otro usando el vértice de la pirámide y su sombra.



Como los rayos del sol son paralelos, estos triángulos son semejantes y como el de la vara es isósceles, también lo es el otro. Para que este procedimiento puede llevarse realmente a cabo, hay que resolver algunos problemas prácticos. Medir la sombra de la pirámide tiene la dificultad de que no hay acceso directo al sitio donde estaría la proyección perpendicular al suelo de su vértice. Si en el momento en que la sombra de la vara mide lo mismo que ella los rayos del sol fueran paralelos a una de las caras de la pirámide este problema sería fácilmente soluble, pero no es un caso muy probable. Lo que queremos señalar es que, independientemente de si el procedimiento se puede realizar en la práctica, o no, la idea es válida, señala cómo, en principio, se puede estimar una magnitud inaccesible a partir de otras que sí son accesibles, utilizando las matemáticas y la razón.

Sabemos que los pitagóricos (Pitágoras 569-475 a.e.c.) veían al mundo como racional, aprensible, a través de los números. Hay que admitir que aunque en cierto sentido estas ideas se adelantan a la visión científica de Galileo Galilei (1564-1642) acerca de que la naturaleza es comprensible solamente

a través de las matemáticas, la idea pitagórica era probablemente menos científica y algo inclinada al misticismo. Por ejemplo, se dice que quisieron ocultar el descubrimiento de los incomensurables porque no encajaban en su visión del mundo, lo cual indicaría una actitud más dogmática que científica. Desde su punto de vista, que consistía en creer que todo conocimiento podía expresarse en términos numéricos, la razón entre la diagonal y el lado de un cuadrado *debería* tener la forma m/n con m y n números enteros positivos. Al descubrir ellos mismos que eso no era posible, su mundo intelectual entró en una crisis de la que no supieron sacarlo y decidieron ocultar su descubrimiento, no sabemos si por vergüenza o solo para evitar el desprestigio que podría acarrearles.

Recordemos que uno de los grandes descubrimientos de los pitagóricos fue que dos cuerdas de igual grosor y tensión producían sonidos que combinados resultaban agradables al oído humano, cuando sus longitudes estaban en relaciones “racionales”, es decir de la forma m/n con m y n siendo números enteros pequeños como $3/5$ o $5/8$. Este descubrimiento, sin duda importantísimo, probablemente les llevó a una exaltación que les llevó a postular que todo en el mundo debía poderse expresar con relaciones racionales. Lamentablemente, a pesar de que estas ideas fomentaron en ellos el estudio de las matemáticas, llegando a realizar contribuciones importantes en ellas, en sí no son ideas precisamente racionales. Más bien al contrario, la filosofía pitagórica era mística y dogmática. Y aún así, constituye el origen de todo el pensamiento matemático y racional.

Sin embargo, uno de los logros más importantes del pensamiento racional, logro inequívocamente matemático y geométrico, el llamado *Teorema de Pitágoras*, lleva el nombre de un filósofo místico y ecologista que no sabemos si lo demostró.

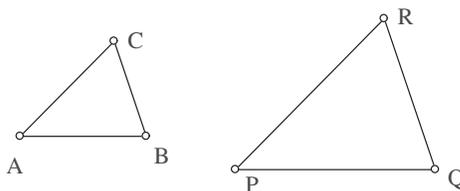
La Geometría surge en la antigua Grecia con dos teoremas igualmente importantes y útiles. El que dice que *los lados de dos triángulos semejantes son proporcionales*, que al menos en una de sus primeras versiones se atribuye a Tales de Mileto (625-546 a.e.c.) y el que dice que *los cuadrados contruidos sobre los catetos de un triángulo rectángulo, equivalen al cuadrado construido sobre su hipotenusa*, atribuido a Pitágoras (569-475 a.e.c.). Al primero nos referiremos como la *Ley de las proporciones para triángulos semejantes* y al segundo con su nombre tradicional: *Teorema de Pitágoras*.

Independiente de a quien o a quiénes debieran atribuirse estos teoremas, el hecho es que son los dos resultados propiamente matemáticos y geométricos más antiguos de la historia y también los más importantes y más usados en las demostraciones de otros teoremas.

El teorema de Pitágoras y los conceptos de área y semejanza

Un hecho interesante es que el *Teorema de Pitágoras* se puede demostrar fácilmente a partir de la *Ley de las proporciones*, como veremos a continuación.

Sea ABC un triángulo rectángulo donde AB es la hipotenusa. Entonces $AB^2 = AC^2 + BC^2$.



- 1) ABC y PQR son semejantes (se entiende que AB corresponde a PQ , BC a QR y CA a RP).
- 2) $AB/PQ = BC/QR = CA/RP$ (los cocientes interpretados como razones)
- 3) $AB \cdot QR = BC \cdot PQ$, $BC \cdot RP = CA \cdot QR$ y $AB \cdot RP = CA \cdot PQ$.

Este resultado nos permite transitar de manera transparente entre las expresiones

$$\frac{AB}{PQ} = \frac{BC}{QR}$$

y

$$AB \cdot QR = BC \cdot PQ$$

como acostumbramos a hacerlo en el álgebra elemental, es decir, “pasando los denominadores de la primera al otro miembro de la igualdad, pero multiplicando en lugar de dividiendo”, y viceversa. La mayor parte del “álgebra” que se usa en la geometría euclidiana consiste en aplicar esta equivalencia que, para los antiguos griegos, siempre se interpretaba así: lo primero es una proporción (igualdad entre dos razones), lo segundo significa que (las áreas de) dos rectángulos son iguales y ambas cosas son equivalentes, es decir, que cualquiera de ellas se cumpla implica que la otra también se cumple.

En las Matemáticas modernas los conceptos de distancia o longitud, área, razón y ángulo se trivializan. Simplemente se les asignan valores numéricos (números reales, claro) e interpretamos sus relaciones en términos de operaciones aritméticas (o algebraicas). Esta manera de proceder deja de lado la naturaleza geométrica de estos conceptos que es independiente de cualquier sistema de numeración. Los antiguos griegos manejaron estos conceptos de una manera más sutil y profunda, probablemente porque su sistema de numeración era deplorable.

No es casualidad que, como veremos más adelante, en los primeros cuatro libros de *Los elementos* de Euclides (360-280 a.e.c.) no se realizan operaciones numéricas ni se hacen manipulaciones algebraicas. No es sino hasta el libro V en el que se desarrolla la teoría de las proporciones, debida a Eudoxio de Cnidos (390-337 a.e.c.), donde se empiezan a manejar abstracciones que pueden identificarse más o menos como números reales. Este hecho que hoy pudiera parecer extraño e innecesario, se debe al cuidado que pusieron los antiguos griegos en no caer en el error de sus predecesores pitagóricos de asignar alegremente números a los segmentos presuponiendo que siempre se podían medir con fracciones (razones entre números enteros) usando alguna unidad adecuada. La razón por la que hemos pospuesto la geometría de la medición al siguiente capítulo es para crear conciencia en el lector de que los conceptos *no numéricos* de longitud, área, volumen, ángulo y razón existen independientemente de los números y constituyen la esencia de la geometría euclidiana.

El descubrimiento de la razón y la emancipación del pensamiento

Los desarrollos matemáticos anteriores a los de la antigua Grecia a partir de los siglos VII y VI a.e.c. fueron, aparentemente, inconscientes y poco racionales. Antes de esa época la humanidad había elaborado abstracciones como los números y algunas figuras geométricas, y los había utilizado en las actividades cotidianas como la agricultura, el comercio y la construcción de viviendas, templos y obras públicas como canales y muros para proteger sus ciudades o incluso todo su territorio, como es el caso de la gran muralla china. Pero no es sino hasta la época de la Grecia clásica que aparece el pensamiento matemático como una actividad consciente de sí misma. No podemos afirmar que en las civilizaciones anteriores a la griega clásica no hubiera razonamiento lógico, solamente sabemos que no hay evidencia escrita o documentada de tal razonamiento. En cambio, como atestiguan Los elementos de EUCLIDES, los trabajos de ARQUÍMEDES y APOLONIO y el *Almagest* de CLAUDIO PTOLOMEO, los griegos fueron especialistas en dejar evidencia clara e incontestable de sus razonamientos. Es por ello que solemos decir que los griegos descubrieron la razón e inventaron el razonamiento abstracto y lógico. De hecho para apoyar esta afirmación, basta recurrir a los escritos de EUCLIDES y ARQUÍMEDES.

Lo que los antiguos matemáticos griegos descubrieron es que, si se establecen con claridad las relaciones entre algunos conceptos abstractos, llamémoslas premisas, de ellas se pueden deducir otras relaciones cuya validez es innegable, siempre y cuando quien las juzgue acepte como válidas las premisas. En ese punto de la historia el ser humano descubre la razón y su extraordinario poder. Este paso requiere de concebir que algunas verdades pueden ser evidentes (por ejemplo que por dos puntos dados pasa una y solo una línea recta) y que otras, menos evidentes (como el Teorema de Pitágoras) pueden obtenerse de ellas por medio del pensamiento racional.

La cultura griega difería en muchos aspectos de las culturas babilonia y egipcia. La diferencia fundamental se ve ya en el ejemplo de *Tales de Mileto* (ca. 624–546 a.e.c.), quien viajó para conocer ambas culturas y aprender de ellas. Los egipcios creían que un ente divino, que llamaron *Inum* y que tenía forma de un carnero, “traía” las inundaciones anuales del Nilo de alguna manera. Tales en cambio trató de explicarlas por fuertes vientos que soplaban y embalsaban el agua.

Aunque la explicación de TALES es errónea, tiene un aspecto revolucionariamente nuevo: trata de explicar un fenómeno de la naturaleza por la naturaleza misma. Su discípulo *Anaximandro de Mileto* (ca. 610–547 a.e.c.), ya contaba con otra explicación, según la cual las inundaciones se provocaban por el derretido de la nieve en montañas lejanas.

En ello podemos observar dos características:

1. Las diferentes explicaciones se discuten. Así nace el diálogo abierto y científico.
2. Los nuevos pensadores quieren saber *por el conocimiento*, sin preguntarse por la utilidad del conocimiento.

Tales también es conocido en las matemáticas: de él se conoce el Teorema de Tales. Las fuentes transmiten que TALES encontró cuatro teoremas y que dió demostraciones para ellos. Es interesante ver que todos son Teoremas muy sencillos y lo característico de ello es que dió demostraciones aún para afirmaciones tan plausibles como que cada diámetro divide el círculo en dos partes iguales. Se puede discernir el esfuerzo de ya no aceptar nada sin cuestionarlo, de tratar de reflexionar sobre todo y de argumentar aún lo obvio.

Con ello empezó un gran desarrollo: se sustituye la creencia (ciega) en autoridades por un trato crítico con explicaciones y argumentaciones. Así fue como comenzó la emancipación del pensamiento

de sus ataduras místicas y autoritarias.

¿Por qué ocurre el descubrimiento de la actividad matemática racional precisamente en la antigua Grecia? Quizás porque sus habitantes tenían pocos prejuicios que les impidieran absorber ideas y conocimientos de culturas ajenas y construir su propio pensamiento con entera libertad. De hecho lograron vivir (al menos quienes eran considerados ciudadanos, los esclavos no) en un ambiente de libertad propicio para explorar sin freno alguno el poder de la razón para incrementar el conocimiento y la sabiduría. Descubrieron que no era necesario obtener revelaciones divinas o provenientes de los poderosos para llegar a saber algo. Tuvieron libertad de seleccionar los conocimientos más sólidos y útiles que les brindó la experiencia de culturas anteriores como la egipcia y la babilonia, y elaborarlos usando esa herramienta libre, gratuita y (casi) universal que es la razón.

Cabe aclarar que en la antigua Grecia las matemáticas incluían originalmente, además de la geometría y la aritmética, también la música, la geología y otros campos que los griegos consideraban que podían estudiarse racionalmente. Poco a poco esos campos fueron considerándose independientes y las matemáticas se redujeron a incluir solo aquellas áreas que trataban con las abstracciones más puras, que en aquella época eran los números y las figuras geométricas. Siglos después se incorporan a las matemáticas muchas otras abstracciones (como el movimiento, las fuerzas, los campos eléctricos y magnéticos, el azar, las estructuras lógicas, algebraicas, topológicas, etcétera), que constituyen el vasto campo actual de las Matemáticas.

El método axiomático

Un patrón conocido de argumentación es reducir la veracidad de una afirmación a la de otras, que ya han sido aceptadas, ya sea por convicción o por convención. ¿Pero en qué se basa la veracidad de esas otras afirmaciones? ¿En otras? El proceso de reducción sin fin se parece a la pregunta repetitiva del niño “¿y eso por qué es así?” que con cada explicación dada se puede volver a plantear de manera casi automática.

Por ello los griegos vieron que es necesario (o al menos conveniente) aceptar algunas afirmaciones como verdaderas sin cuestionarlos. Esas afirmaciones, que llamaron *axiomas* y *postulados*, conforman el fundamento de todo lo que se deduce de ellas. La distinción entre axiomas y postulados se debe a ARISTÓTELES (384–322 a.e.c.): los axiomas son generales y los postulados se refieren a la temática especial de la que trata el texto. Estas axiomas y postulados deberían ser afirmaciones que sean clara e inmediatamente evidentes por sí mismas.

La idea fundamental del método axiomático es que se exhiba con toda claridad qué es lo que se supone sin cuestionamiento (es decir los axiomas y los postulados) para poder así tener absoluta claridad de lo que se puede deducir de ello usando únicamente la razón. Si alguien está dispuesto a aceptar estos fundamentos que se llaman axiomas y postulados, entonces está obligado a aceptar como verdaderos todos los enunciados que se demuestran a partir de ellos. El método axiomático es en realidad un acto de honestidad y claridad.

El método axiomático en Euclides

300 años después de TALES vivió EUCLIDES (ca. 360–280 a.e.c.). Sobre su vida nos han llegado muy pocas noticias, pero su obra más importante, “*Los Elementos*”, en la que expone el *método axiomático* como ningún otro, se convirtió en el modelo indiscutible del rigor metódico que conduce

al conocimiento racional. Los 13 tomos de su obra no contenían, en general, matemáticas nuevas, gran parte de sus enunciados ya era conocidos en aquella época. Lo importante fue que formuló esos conocimientos utilizando el método axiomático con todo rigor. Durante más de dos milenios *Los Elementos* ha sido el libro más estudiado por los matemáticos y hombres de ciencia de todas las épocas y culturas. Tal es su importancia.

Los Elementos consta 5 POSTULADOS y 5 NOCIONES COMUNES que se enuncian en el libro I y de muchas DEFINICIONES y PROPOSICIONES que se distribuyen a lo largo de los 13 libros. Los POSTULADOS y las NOCIONES COMUNES constituyen los AXIOMAS (en griego $\alpha\chi\iota\omega\mu\alpha$ significa “aquello que parece correcto”). Las definiciones sirven para establecer el tema de estudio y dar nombres a sus elementos.

El Libro I comienza con 23 DEFINICIONES y enuncia a continuación los axiomas, es decir, los 5 POSTULADOS y las 5 NOCIONES COMUNES, en los que se basa el resto de la obra. Es importante notar que los 5 postulados requieren para ser enunciados de algunas definiciones, pues son afirmaciones sobre *puntos, líneas, ángulos y círculos*, conceptos cuyo significado se intenta aclarar en las definiciones.²

DEFINICIONES

1. Un punto es lo que no tiene partes.
2. Una línea es una longitud sin anchura.
3. Los extremos de una línea son puntos.
4. Una línea recta es aquella que yace por igual respecto de los puntos que están en ella.
5. Una superficie es aquello que sólo tiene longitud y anchura.
6. Los extremos de una superficie son líneas.
7. Una superficie plana es aquella superficie que yace por igual respecto de las líneas que están en ella.
8. Un ángulo plano es la inclinación mutua de dos líneas que se encuentran una a otra en un plano y no están en línea recta.
9. Cuando las líneas que comprenden el ángulo son rectas, el ángulo se llama rectilíneo.
10. Cuando una línea recta que está sobre otra hace que los ángulos adyacentes sean iguales, cada uno de los ángulos es recto, y la recta que está sobre la otra se llama perpendicular a la otra.
11. Un ángulo obtuso es el que es mayor que uno recto.
12. Un ángulo agudo es el que es menor que uno recto.
13. Un límite es lo que es extremo de algo.
14. Una figura es aquello que está contenido por cualquier límite o límites.
15. Un círculo es una figura plana comprendida por una sola línea (llamada circunferencia) de tal modo que todas las rectas dibujadas que caen sobre ella desde un punto de los que están dentro de la figura son iguales entre sí.
16. Y el punto se llama centro del círculo.

²Como puede verse, algunas definiciones, especialmente las primeras, resultan torpes y no logran del todo su objetivo. Este hecho lo reconoció con claridad Hilbert por lo que en su versión de los fundamentos de la geometría algunos términos como punto, recta y plano aparecen sin justificación alguna y son precisamente los axiomas los que los definen estableciendo sus propiedades (Ver 3.3).

17. Un diámetro de un círculo es una recta cualquiera que pasa por el centro y que acaba en ambas direcciones en la circunferencia del círculo; esta línea recta también divide el círculo en dos partes iguales.
18. Un semicírculo es la figura comprendida entre el diámetro y la circunferencia cortada por él. El centro del semicírculo es el mismo que el del círculo.
19. Figuras rectilíneas son aquellas que están comprendidas por líneas rectas, triláteras las comprendidas por tres, cuadriláteras las comprendidas por cuatro y multiláteras las comprendidas por más de cuatro.
20. De los triángulos, el equilátero es el que tiene los tres lados iguales; isósceles el que tiene dos lados iguales y uno desigual; y escaleno el que tiene los tres lados desiguales.
21. De los triángulos, triángulo rectángulo es el que tiene un ángulo recto, obtusángulo el que tiene un ángulo obtuso y acutángulo el que tiene los tres ángulos agudos.
22. De los cuadriláteros, cuadrado es el que tiene los lados iguales y los ángulos rectos; rectángulo el que es rectangular pero no equilátero; rombo el que es equilátero, pero no tiene los ángulos rectos; y romboide el que tiene dos lados y dos ángulos opuestos iguales, pero ni es equilátero ni sus ángulos son rectos. Los otros cuadriláteros se llaman trapecios.
23. Rectas paralelas son aquellas que, estando en un mismo plano y siendo prolongadas indefinidamente en ambos sentidos, no se encuentran una a otra en ninguno de ellos.

Como puede verse, estas definiciones coinciden casi exactamente con las que usamos en la actualidad, con las excepción de la *línea*, que en la actualidad corresponde a *curva* y *línea recta* que en la actualidad llamamos *segmento*.

POSTULADOS

1. Se puede trazar una línea recta de un punto a cualquier otro.
2. Se puede prolongar cualquier línea recta acotada de manera indefinida.
3. Se puede trazar un círculo con cualquier centro y cualquier radio.
4. Todos los ángulos rectos son iguales entre sí.
5. Si cuando una línea recta intersecta a otras dos hace que los ángulos interiores del mismo lado sean menos que dos ángulos rectos, entonces las dos líneas rectas, al prolongarse, se intersectan de ese lado.

NOCIONES COMUNES

1. Cosas que son iguales a una tercera son iguales entre sí.
2. Si uno suma iguales a iguales, los totales son iguales.
3. Si uno subtrae iguales de iguales, los restantes son iguales.
4. Cosas que coinciden una con otra son iguales entre sí.
5. El total es más grande que cada una de sus partes.

Después se enuncian y se demuestran las PROPOSICIONES una a una. En cada paso se indica lo que se usa, sea esto un postulado o una proposición ya demostrada. La lógica del método axiomático consiste en separar claramente entre lo que se acepta sin discusión y lo que se obtiene como consecuencia. Si uno tiene duda que una de las consecuencias sea correcta, debe ser porque se tiene una duda sobre alguno de los pasos de la deducción, o porque se duda de alguno de los postulados.

Las PROPOSICIONES SON DE DOS TIPOS:

1. Afirmaciones cuya validez se demuestra a partir de los postulados y de proposiciones anteriores.
2. Enunciados en los que se describe algo que se desea construir y en los que se describe un procedimiento de construcción se demuestra es que tal procedimiento produce el resultado enunciado.

Los siguientes libros comienzan a veces con nuevas definiciones que establecen un lenguaje adecuado para referirse a ciertos conceptos. En algunas ocasiones hay definiciones intercaladas en el libro entre un grupo y otro de proposiciones, pero siempre lo que presenta el grueso del libro son proposiciones con sus enunciados destacados y las demostraciones respectivas, frecuentemente ilustradas con dibujos o esquemas.

Es importante notar que a lo largo de los 13 libros solo aparecen nuevas definiciones y nuevas proposiciones, pero nunca nuevas nociones comunes o nuevos postulados. Las nociones comunes pertenecen al ámbito de la lógica general. Los 5 postulados, en cambio, son afirmaciones relativas al tema de la geometría que se ofrecen sin demostración y en las que deberán basarse las demostraciones de todas las afirmaciones posteriores. Es decir, todo el conocimiento geométrico se fundamenta en estos 5 postulados. Cabe hacer notar que a lo largo de los dos milenios posteriores a *Los elementos*, estos postulados fueron sometidos a un fuerte escrutinio por parte de los matemáticos y no siempre salieron airosos. Se descubrieron varios huecos y fue necesario enmendar los detalles del método de Euclides (Ver 3.3). Pero el concepto del método se preserva hasta nuestros días y sigue siendo el fundamento del conocimiento racional.

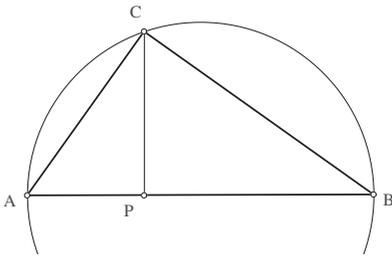
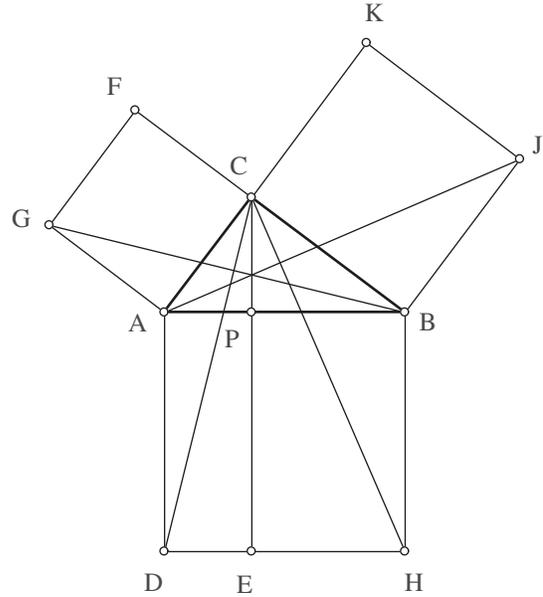
Contenido matemático de *Los Elementos* de EUCLIDES

El contenido de esta obra es siempre de carácter matemático, pero no trata únicamente de geometría, también hay temas de aritmética y teoría de números, aunque es verdad que los temas geométricos dominan la mayor parte de la obra y que la teoría de números se presenta en términos que podrían considerarse casi geométricos.

Los libros I, II, III y IV tratan de geometría plana

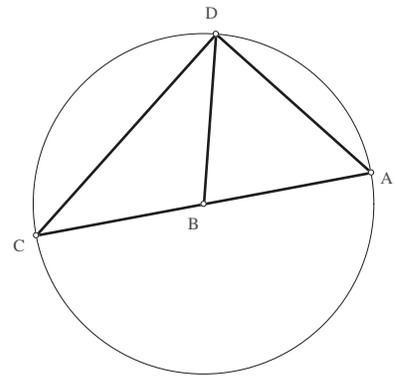
El libro I, después de establecer las definiciones y los axiomas, presenta proposiciones básicas de la geometría plana, algunas de las cuales constituyen construcciones “con regla y compás”, como la proposición 1: *Construir un triángulo equilátero sobre un segmento*, o la 10: *Dividir un segmento en dos partes iguales*. La proposición 5 dice: *Los ángulos en la base de un triángulo equilátero son iguales*. La proposición 36 dice: *Los paralelogramos que están sobre bases iguales y entre las mismas paralelas, son iguales entre sí* (es decir, tienen áreas iguales) y por tanto justifica la fórmula $base \times altura$ para el área del paralelogramo.

La 38 dice: *Los triángulos que están sobre bases iguales y entre las mismas paralelas, son iguales entre sí.* y la 41: *Si un paralelogramo tiene la misma base que un triángulo y ambos están entre las mismas paralelas, entonces el paralelogramo es el doble que el triángulo* justifican la fórmula $\frac{\text{base} \times \text{altura}}{2}$ para el área de un triángulo. Las dos últimas proposiciones del libro I, la 47 y la 48, constituyen el Teorema de Pitágoras y su inverso. En resumen, este libro fija los primeros principios que son una especie de explicitación del lenguaje y las normas que se usan en el razonamiento deductivo y desarrolla gradualmente los conceptos no numéricos de área, ángulo recto y perpendicularidad, y concluye con una de las demostraciones más bellas, aunque ciertamente no la más simple, del Teorema de Pitágoras. La figura adjunta muestra el esquema correspondiente a la proposición 47.



Se dice que el libro II trata de álgebra geométrica pues sus proposiciones pueden interpretarse como identidades algebraicas tales como $a(b+c+\dots) = ab+ac+\dots$ o $(2a+b)^2 + b^2 = 2(a^2 + (a+b)^2)$. Se trata de un álgebra de segmentos que son figuras, y no números. Su última proposición, la 14: *Construir un cuadrado igual a una figura rectilínea dada* muestra una manera de definir un segmento que es la raíz cuadrada del producto de otros dos, es decir, realiza una operación algebraica con segmentos, que son entes geométricos.

El libro III presenta los círculos y sus propiedades. Por ejemplo, cómo encontrar el centro de un círculo (proposición 1), intersecciones de círculos y círculos tangentes (proposiciones 10-13), trazar una recta tangente a un círculo (proposición 17), los ángulos inscritos y su relación con los ángulos centrales, incluyendo el llamado segundo teorema de Tales (figura adjunta) que afirma que el ángulo desde el cual se ve un diámetro desde un punto de la circunferencia es recto (proposiciones 20-22) y los teoremas sobre la potencia de un punto respecto a un círculo (proposiciones 35-37).



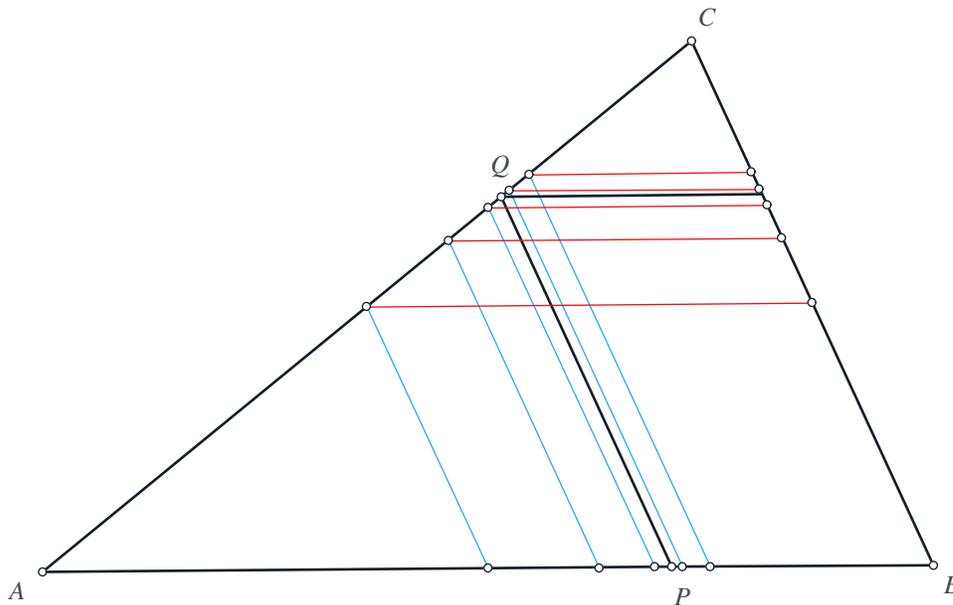
En el libro IV se construyen el incírculo (proposición 4) y el circuncírculo (proposición 5) de un

triángulo y también los polígonos regulares de 4, 5, 6 y 15 lados. La joya de este libro son las proposiciones 10 y 11 en las que se construye el pentágono regular. En la proposición 16 se construye el pentadecágono regular utilizando las construcciones del triángulo y el pentágono regulares inscritos en un mismo círculo.

Los libros V y VI tratan la teoría de las proporciones y la semejanza de triángulos

El libro V presenta la teoría de las proporciones, debida a EUDOXIO de Cnidos (390-337 a.e.c.), uno de los matemáticos más importantes de la antigüedad griega junto con ARQUÍMEDES de Siracusa y APOLONIO de Pérgamo. Éste es uno de los libros más difíciles de leer de *Los elementos* en la actualidad. En él se presenta el concepto de la razón de dos magnitudes (las magnitudes pueden visualizarse como longitudes de segmentos, por ejemplo), que hoy día no utilizamos ya que las representamos simplemente con números reales. Se trata de magnitudes no cuantificadas que permiten trabajar con entidades cuyos valores deberían ser continuos (no discretos) como la longitud, y las medidas de los ángulos, las área y los volúmenes, sin usar números irracionales.

En el fondo, la teoría de las proporciones es un sustituto de los números reales positivos que fue necesario desarrollar porque los antiguos griegos no tenían un sistema de numeración que permitiera siquiera concebir a los irracionales como números. Fue Eudoxio quien hizo este desarrollo y lo hizo con gran elegancia y rigor. Cabe mencionar que también se debe a Eudoxio el llamado método de exhaustión para definir y a veces calcular las áreas de figuras con fronteras no rectilíneas y los volúmenes de cuerpos cuyas superficies no son planas. Fue Arquímedes quien obtuvo los resultados más importantes de la antigüedad con este método, como la cuadratura de la parábola, el volumen y la superficie de la esfera y su estimación del número π (Ver 5.4, 5.5 y 5.6). Pero fue Eudoxio quien sentó las bases teóricas para realizar estos cálculos. Es probable que desarrollase la teoría de las proporciones precisamente para sustentar su método de exhaustión.



<http://descartes.matem.unam.mx/ejemplos/ConGeo2D/>

El libro VI aplica la teoría de las proporciones desarrollada en el libro V a la semejanza de triángulos. Uno de los resultados poco conocidos de este libro y que sin embargo es muy útil es la proposición 3 en la que se demuestra que la bisectriz del vértice de un triángulo divide al lado opuesto en la misma razón en la que están sus lados adyacentes. El resultado más conocido y más usado de este libro es parte de la proposición 4: *En los triángulos equiángulos, los lados que comprenden los ángulos iguales son proporcionales*. Solemos referirnos a este resultado diciendo simplemente que los *lados correspondientes de triángulos semejantes son proporcionales*. La mayor parte de este libro profundiza en lo que hoy llamamos “criterios de semejanza”.

En la enseñanza moderna de la geometría los resultados del libro VI son unos de los más conocidos y utilizados, pero suelen verse haciendo caso omiso de la teoría de las proporciones del capítulo V, abusando de la ventaja que tenemos de poder tratar las longitudes de segmentos como números reales y de utilizar el álgebra básica para operar con ellos. Por ejemplo, si los triángulos ABC y DEF son semejantes, interpretamos la proporcionalidad de sus lados como las igualdades de los cocientes $\frac{AB}{DE} = \frac{BC}{EF} = \frac{CA}{FD}$ y no tenemos ningún empacho en deducir de aquí, por álgebra elemental, que $AB \cdot EF = BC \cdot DE$ y que $\frac{AB}{BC} = \frac{DE}{EF}$. Para los antiguos griegos estos resultados estaban lejos de ser tan triviales como hoy los vemos. Tales cocientes y productos no existían para ellos, no eran números sino razones y no podían operar con ellas tan fácilmente como lo hacemos ahora que los interpretamos como números reales. Los resultados equivalentes a estas manipulaciones algebraicas tan simples fueron para ellos resultados del desarrollo cuidadoso de la teoría de las proporciones, que les permitió avanzar mucho en la geometría a pesar de no contar con los números reales.

Los libros VII, VIII, IX y X tratan la teoría de números

Los libros VII, VIII y IX tratan teoría de los números enteros positivos y comparten una sola lista de definiciones que se encuentra al principio del libro VII. El libro VII trata la divisibilidad de los números (enteros positivos) y los números primos. Presenta el “algoritmo de Euclides” para encontrar el máximo común divisor y el mínimo común múltiplo. Las proposiciones 30 y 32 constituyen lo que hoy llamamos el Teorema fundamental de la aritmética: *Todo entero positivo se puede escribir como el producto de números primos de una manera única*. Aunque Euclides no podía enunciarlo así pues en su época no se concebía un producto de más de tres números.

Los libros VIII y IX tratan de proporciones entre números enteros positivos y sus cuadrados y cubos. El libro VIII trata también las sucesiones geométricas. El libro IX demuestra que hay una infinidad de números primos (proposición 20), calcula la suma de una serie geométrica (proposición 35) y presenta la construcción de números perfectos (proposición 36).

El libro X es por mucho el más voluminoso. Trata de “magnitudes inconmensurables” (es decir números irracionales) y utiliza los conocimientos desarrolladas en los libros V, VI, VII, VIII y IX, así como las ideas del método de exhaustión para relacionarlas con los números.

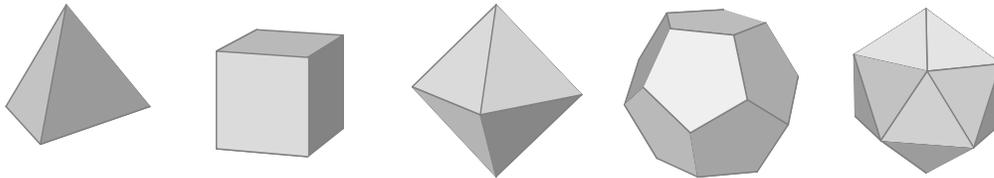
Es interesante leer partes de estos cuatro libros y comprobar lo difícil que era hablar de los temas que tratan sin el beneficio de la notación algebraica que hoy disfrutamos. En gran medida los libros VII, VIII, IX y X son los más anticuados de *Los elementos*, no porque sus resultados no sean válidos e importantes, sino porque están presentados en un lenguaje que hoy en día resulta sumamente pesado.

Los libros XI, XII y XIII tratan la geometría del espacio

El libro XI trata, en particular el tema de los puntos, las rectas, los planos en el espacio y las figuras sólidas.

El libro XII trata del volumen de algunas figuras sólidas como prismas, pirámides y conos, y la razón que guardan los volúmenes esas figuras con respecto a sus magnitudes lineales.

Finalmente el libro XIII es un tratamiento de los *poliedros regulares*, también llamados *sólidos platónicos*.



El mayor logro de este libro es la construcción, podríamos decir que “con regla y compás”, del tetraedro, el octaedro, el dodecaedro y el icosaedro regulares. Lo que en realidad se construye con regla y compás son los lados y los ángulos entre los lados y entre las caras de estas figuras para que queden inscritas en una esfera dada.

Conclusión

Los elementos de EUCLIDES constituyen una exposición bien ordenada y muy completa de los conocimientos matemáticos que se tenían en la época de Euclides. No deben considerarse únicamente como un libro que expone el método axiomático. La combinación de ambos aspectos: el método axiomático y los conocimientos matemáticos, son lo que hacen de *Los elementos* una obra tan importante. Las obras de los matemáticos anteriores a Euclides no sobrevivieron. Probablemente esto se debe a que resultaban innecesarias al contar con la excelente exposición de *Los elementos*. Pero se sabe que tales obras existieron y que ya contenían casi todos los conocimientos expuestos en la de Euclides.

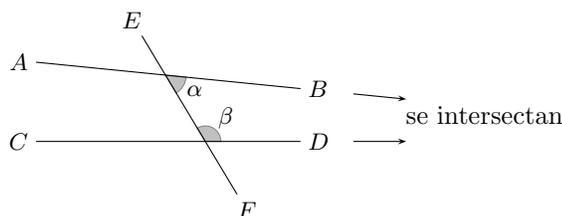
A lo largo de este capítulo y del siguiente se presentará la mayor parte de los contenidos geométricos que aparecen en *Los elementos*, así como varios desarrollos posteriores. Primero haremos una presentación de los fundamentos de la geometría, más apegada a Hilbert que a Euclides, pero difiriendo también de este último autor en algunos aspectos. A continuación seguiremos un programa para desarrollar la geometría partiendo de tales fundamentos separándola en dos partes: aquello que no tiene que ver con áreas y volúmenes, que es lo que se trata en este capítulo, y todo aquello que tiene que ver con la medida de segmentos, áreas y volúmenes, que es lo que se trata en el siguiente.

3.3. Los fundamentos de la geometría

El quinto postulado

Los primeros cuatro postulados de *Los Elementos* de EUCLIDES satisfacen el requerimiento de que los axiomas deben ser afirmaciones evidentes. Pero eso no pasa con el quinto postulado.

El enunciado del quinto postulado no es inmediatamente evidente. Para empezar, hay que leerlo dos o tres veces para lograr entender lo que quiere decir. Habla de un segmento recto que intersecta a otros dos, llamemos al primero EF . A los otros dos, que son cortados por EF , los llamamos AB y CD .



El postulado parte de la premisa que los ángulos interiores de un lado, que en la figura adjunta se llaman α y β , suman menos que dos ángulos rectos. En esta situación, el postulado afirma que las dos rectas AB y CD se intersectan posiblemente después de prolongarlas del lado donde están estos dos ángulos.

Es claro que el quinto postulado no tienen la elegancia de los anteriores, no nos parece tan evidente como el de que por cualesquiera dos puntos se puede trazar el segmento recto que los une. Euclides mismo no usó el quinto postulado hasta la Proposición 27. Parece que él mismo no estaba muy satisfecho con él.

Esto causó mucha controversia. Muchos matemáticos creyeron que el quinto postulado era en realidad una consecuencia de los primeros cuatro y trataron de demostrarlo, pero nunca pudieron. Dos mil años después de Euclides se demostró que era imposible demostrarlo y se descubrió una nueva geometría en la que el quinto postulado no se cumple (ver 3.7). Este hecho nos hace apreciar aún más el trabajo de Euclides pues fue capaz de distinguir que este postulado era necesario para su geometría. Tal vez incluso intuyó que sin él la geometría sería imposible, o muy diferente.

Este hecho propició una revisión muy detallada de los fundamentos de la geometría que culmina en el libro de David Hilbert (1862.1943) *Grundlagen der Geometrie (Fundamentos de la geometría)* publicado en 1899 que no solo proporciona una nueva versión depurada de los axiomas de la geometría sino que ejemplifica la versión moderna del método axiomático.

El método axiomático en Hilbert

Perspectiva educativa

Es innegable que, en su origen, el pensamiento racional busca primero deducir unas verdades de otras y mucho más tarde intenta encontrar primeros principios o axiomas a partir de los cuales se puedan deducir las demás verdades. La búsqueda de primeros principios surge como una necesidad

cuando el conjunto de verdades inter-conectadas por la razón debe ordenarse y documentarse, como ocurrió en el trabajo de Euclides. Allí es donde surge la necesidad de decidir por dónde empezar. En la práctica y en la génesis de las matemáticas la búsqueda de primeros principios es posterior al establecimiento de los conceptos, sus propiedades y los teoremas más importantes. No conviene olvidar esto para que al enseñar matemáticas no ignoremos ese paso pre-axiomático que es la etapa más vital, interesante, motivadora y formadora de la actividad matemática. Es la etapa en la que se exploran las primeras ideas y se descubren relaciones interesantes entre los conceptos abstractos que se están creando o aprendiendo. Cuando al enseñar matemáticas se presentan al alumno directamente los axiomas, no se puede esperar que responda con entusiasmo y menos aún que disfrute de los primeros teoremas que se pueden deducir directamente de ellos, los cuales suelen ser bastante sosos. En una presentación estrictamente axiomática, los teoremas más interesantes, que son el objetivo de toda la construcción, no aparecen sino al final de la presentación.

Creemos que es muy inconveniente esconder al estudiante la motivación originaria, el descubrimiento de las relaciones entre unos conceptos aún mal definidos pero que tienen significado por su cercanía con la problemática concreta que los impulsa a nacer. Conviene ofrecerle la oportunidad de redescubrir todo, en la medida de lo posible, sin llegar a una parálisis, y guiarlo para profundizar poco a poco, tanto hacia las aplicaciones como hacia los fundamentos. Los sistemas axiomáticos pulidos son necesarios para los matemáticos profesionales. Al ciudadano medio le conviene conocer las abstracciones que son significativas para él, y entender sus consecuencias importantes, sin necesidad de pasar por las sutilezas de los primeros principios y los detalles de una estructura lógica muy pulida pero que quizás carece de naturalidad y significado inmediato. Esto no debe confundirse con dar la espalda al pensamiento riguroso que debe ser siempre el aspecto central del aprendizaje de las matemáticas, pero éste tampoco debe confundirse con la deducción a partir de primeros principios. El pensamiento riguroso puede apreciarse y aprenderse mejor si se experimenta como una de las partes del proceso de descubrimiento, abstracción y formalización.

El método axiomático en el mundo moderno

Pero tarde o temprano hay que “coger el toro por los cuernos” y decidir cuáles serán los axiomas de una teoría matemática. Los antiguos griegos pensaban que los primeros principios (como los postulados y las nociones comunes de Euclides) deberían ser *axiomas*, que para ellos quería decir *afirmaciones cuya verdad es evidente*. Sin embargo la experiencia del quinto postulado obligó a los matemáticos a ser más cuidadosos con el concepto de primeros principios y es así como David Hilbert en sus *Fundamentos de la geometría* no los concibe como verdades evidentes sino como *propiedades de las abstracciones que ellos mismos ayudan a determinar*. Por ejemplo, el que *por dos puntos pasa una y solo una línea recta* es un axioma, pero no porque sea un hecho evidente sino porque los puntos y las rectas son entes abstractos que tienen, entre otras, esta propiedad. Hilbert modifica radicalmente el concepto de axioma eliminando de él el requerimiento de ser evidente y lo convierte en definición parcial de aquello de lo que afirma algo. Al mismo tiempo elimina algunas definiciones, como las de punto, recta y plano, que sin duda eran un poco ridículas en la versión de Euclides, y las sustituye precisamente por el conjunto de axiomas que hablan de ellas, convirtiéndose así los axiomas en los determinantes los conceptos abstractos cuyas propiedades describen.

Sin duda la idea del método axiomático de Hilbert es mucho más clara y precisa que la de Euclides, pero a cambio tiene que pagar un precio: su aplicabilidad al mundo real no está garantizada. Hilbert libera a las matemáticas de la necesidad de responder al mundo real, pero a cambio las equipa de una

estructura casi perfecta. La utilidad de un sistema axiomático hilbertiano depende de si hay alguna interpretación en el mundo real de sus conceptos que se comporte de manera idéntica o al menos cercana al modelo abstracto que el sistema define. Pero un sistema axiomático no tiene la obligación de representar ninguna realidad. Los sistemas axiomáticos se obtienen verdades absolutas, pero no hablan de la realidad sino de entes abstractos.

Los requerimientos que Hilbert pide a un sistema axiomático es que los axiomas sean *consistentes*, es decir que de ellos no se pueda deducir una contradicción; *independientes*, es decir que ninguno de ellos pueda deducirse a partir de los otros; y *completos*, es decir toda proposición que pueda expresarse en el lenguaje del sistema axiomático es *decidible*, o sea que ella, o bien su negación, debe poder deducirse de los axiomas. Los dos primeros requisitos son factibles y Hilbert mismo lo ilustró con claridad en su libro. Sin embargo el tercer requerimiento, el de que el sistema sea *completo*, no es factible. Kurt Gödel (1906-1978) demostró en 1931 que en cualquier sistema axiomático suficientemente robusto como para describir a los números naturales, hay proposiciones *indecidibles*, es decir, proposiciones verdaderas que no pueden deducirse de los axiomas.

Los axiomas de la geometría de Hilbert

A pesar de la imposibilidad de contar con un sistema axiomático completo, los axiomas de la geometría postulados por Hilbert son muy interesantes y mal que bien constituyen el enfoque más claro y riguroso que tenemos en la actualidad para el estudio de la geometría. Desafortunadamente el libro de Hilbert trata solo el asunto de los fundamentos y logra demostrar que sus axiomas son consistentes e independientes, que es lo mejor que se puede lograr. También resulta claro que son suficientes para reproducir todos los resultados de la geometría euclidiana. En este capítulo usaremos los axiomas de Hilbert para desarrollar la geometría euclidiana, aunque lo haremos siguiendo un programa distinto al del libro de Hilbert ya que nuestro propósito es exponer de manera ordenada y racional los resultados más importantes de la geometría euclidiana y no tanto hacer un escrutinio de sus fundamentos.

Hilbert plantea cinco grupos de axiomas.

GRUPOS DE AXIOMA DE HILBERT

1. Axiomas de *conexión e incidencia*.
2. Axiomas de *orden*.
3. Axioma de *Euclides* o *de las paralelas*.
4. Axiomas de *congruencia*.
5. Axioma de *Arquímedes* o de continuidad.

A diferencia de cómo hace su presentación Hilbert, aquí, después de presentar los axiomas de *conexión e incidencia* vamos a extraer de ellos algunos resultados importantes, antes de proceder a enunciar los otros grupos de axiomas.

3.4. Los axiomas de conexión e incidencia

Como ya aclaramos anteriormente, Hilbert no intenta definir los términos básicos como *punto*, *recta* y *plano* sino que adopta como definiciones de ellos precisamente las propiedades que los axiomas de conexión e incidencia describen. Lo único que menciona antes de enunciar estos axiomas es que se trata de entes abstractos que se van a denotar así: los puntos con letras mayúsculas A, B, \dots ; las rectas con letras minúsculas a, b, \dots y los planos con letras griegas α, β, \dots .

AXIOMAS DE CONEXIÓN E INCIDENCIA

1. Dos puntos distintos A y B siempre determinan completamente una recta a . En tal caso decimos que $a = AB$ (y también que $a = A \vee B$).
2. Dos puntos distintos de una recta siempre la determinan completamente. Es decir, si $A \vee B = a$ y $A \vee C = a$, entonces también $B \vee C = a$.
3. Tres puntos A, B, C que no están en una misma recta siempre determinan completamente un plano α . En ese caso escribimos $ABC = \alpha$ y también $A \vee B \vee C = \alpha$.
4. Tres puntos que están en un mismo plano siempre lo determinan completamente.
5. Si dos puntos A, B de una recta a están en un plano α entonces todos los puntos de a están también en α .
6. Si dos planos α y β tienen un punto A en común, entonces tienen al menos otro punto B en común.
7. En toda recta hay al menos dos puntos. En todo plano hay al menos tres puntos que no están en una misma recta. Y en el espacio hay al menos cuatro puntos que no están en un mismo plano.

He aquí algunas consecuencia inmediatas de estos axiomas. Dejamos las demostraciones como ejercicio para el lector.

CONSECUENCIAS INMEDIATAS DE LOS AXIOMAS DE CONEXIÓN E INCIDENCIA

1. Dos rectas distintas tienen un punto en común o bien no tienen ningún punto en común.
2. Dos planos tienen una recta en común o bien no tienen ningún punto en común.
3. Un plano y una recta que no está contenida en él tienen un punto en común o bien no tienen ninguno en común.
4. Una recta y un punto que no está en ella determinan un único plano que los contiene.
5. Dos rectas que tienen un solo un punto en común determinan un único plano que las contiene.

3.5. La geometría proyectiva

Los axiomas de conexión e incidencia describen aspectos básicos pero importantes de nuestro espacio. La mejor muestra de esto está en la perspectiva (ver 3.6). Todo lo que visualizamos en proyecciones planas (cine, tv y videojuegos), así como las obras de arte de la pintura, usan la teoría de la perspectiva, que se basa en los axiomas de conexión e incidencia.

Es sorprendente comprobar que el estudio de las consecuencias de los axiomas de conexión e incidencia, que hoy llamamos *Geometría proyectiva*, los más sencillos de toda la geometría, tuvo su

apogeo a principios del siglo XIX y no, como podría pensarse, en la antigüedad. Aunque los antiguos griegos y sobre todo los matemáticos de Alejandría estudiaron algunos aspectos de la geometría proyectiva, no llegaron a separar los aspectos propiamente proyectivos de los de medición.

Quien tuvo por primera vez la visión clara de que había muchos aspectos de la geometría que dependían exclusivamente de las propiedades de los puntos y las rectas, fue el matemático alemán KARL GEORG CHRISTIAN VON STAUDT (1798 - 1867), colaborador de GAUSS. Sin embargo a lo largo de la historia muchos matemáticos contribuyeron con ideas importantes al desarrollo de la Geometría proyectiva. El más importante en la antigüedad es, sin duda PAPPUS DE ALEJANDRÍA quien enunció por vez primera uno de los resultados más importantes de esta geometría. Pero podemos encontrar contribuciones anteriores como en el caso de MENELAO. Quizás la primera contribución moderna deba considerarse la de DESARGUES que aunque en la actualidad está bien reconocida, por muchos años fue ignorada por el establecimiento matemático. Y es que no fue fácil para el mundo matemático reconocer que hay muchos resultados de importancia fundamental en la geometría que nada tienen que ver con los aspectos métricos. Renunciar a los aspectos métricos implica, para empezar, renunciar a una de las figuras geométricas más apreciadas, el círculo. En Geometría proyectiva no hay círculos. Sin embargo la Geometría proyectiva está íntimamente relacionada con muchos resultados que identificamos como relativos al círculo. Y es que esos resultados en realidad aplican a todas las curvas cónicas, las cuales, o sorpresa, sí son parte de la Geometría proyectiva. Sí, el hecho de que los círculos no sean elementos proyectivos, pero las curvas cónicas sí, es uno de los aspectos más sorprendentes de la Geometría proyectiva y quizás el que más ha contribuido a que ésta sea malentendida por la mayoría de los matemáticos que poco conocen de ella.

No es posible en este trabajo de carácter general hacer una presentación que haga justicia a la Geometría proyectiva. Solamente intentaremos señalar sus aspectos más representativos y a describir su importancia como fundamento del resto de la Geometría.

Comenzaremos por mostrar tres de los resultados más típicos y sencillos de la Geometría proyectiva, que pueden considerarse consecuencias más o menos directas de los axiomas de conexión e incidencia.

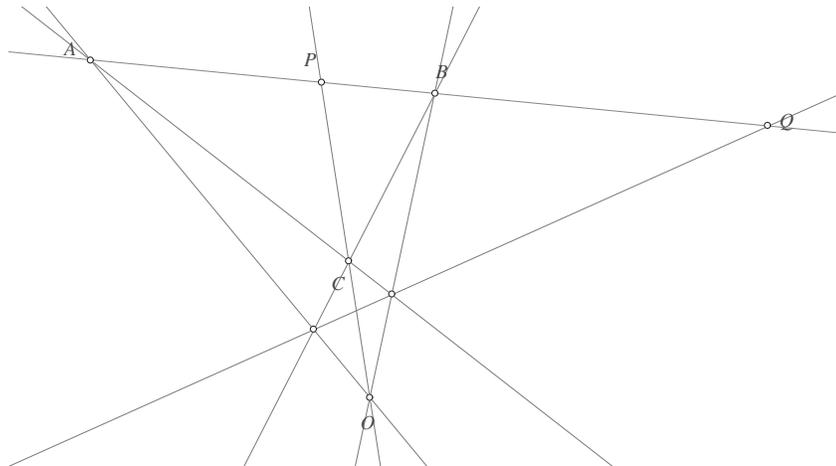
CONSECUENCIAS NO TRIVIALES DE LOS AXIOMAS DE CONEXIÓN E INCIDENCIA

1. Teoremas sobre cuartetos armónicos.
2. Teorema de Desargues.
3. Teorema de Pappus.

Una característica muy interesante de estos teoremas es que, aunque pueden enunciarse como afirmaciones sobre puntos y rectas del plano, sus demostraciones más directas y que no requieren más que de los axiomas de conexión e incidencia, pasan por el espacio.

Teorema de las cuartetos armónicos

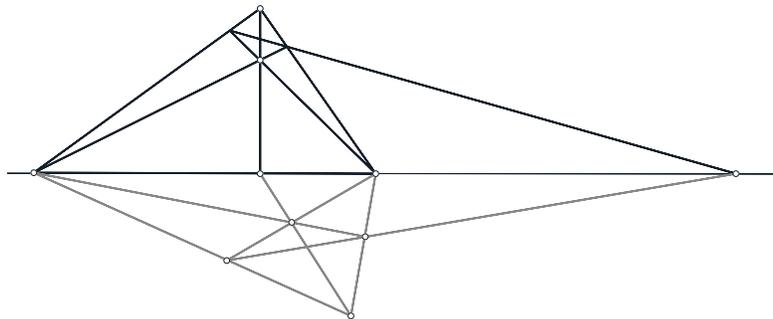
Teorema. Dados dos puntos A y B y otro punto P alineado con ellos, construyamos un cuarto punto Q , en la misma línea, como se ilustra en la figura, donde el punto O es cualquiera que no esté en AB y C es un punto arbitrario de OP . Entonces el punto Q es siempre el mismo, es decir, es independiente de O y de C . Q se llama el *armónico conjugado de P respecto a AB* .



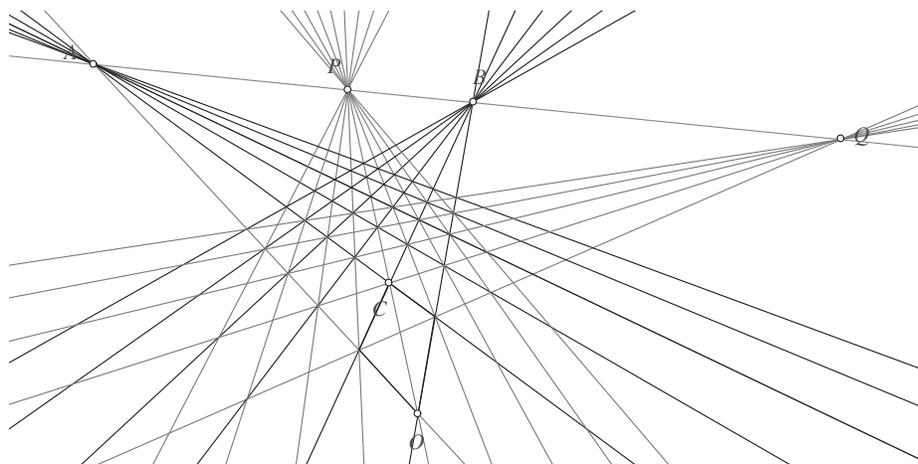
©3D

Imaginemos la recta que contiene a los puntos A , B y P como yaciendo en un plano horizontal, digamos el suelo. Realicemos la construcción del armónico Q de P respecto a AB en un plano distinto (por ejemplo vertical). Proyectemos al plano original (es decir al suelo) desde un punto fuera de ambos planos. Podemos pensar esa proyección como la *sombra* de la construcción. Es claro que esa sombra constituye otra construcción del armónico Q de P respecto a AB y lleva al mismo resultado porque A , P , B y Q no cambian con la proyección.

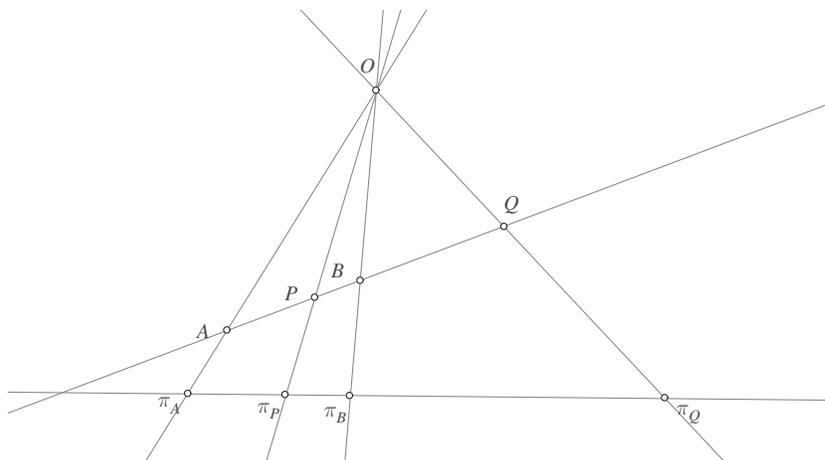
◦



El teorema de las cuartet armónicas está íntimamente relacionado con la perspectiva, como la construcción que aquí presentamos sugiere.



La proyección de una cuarteta armónica desde un punto O a otra recta es también una cuarteta armónica. Es decir, si A, B, P, Q es una cuarteta armónica, O es un punto que no está en la recta AB y r es una recta distinta de AB que no contiene a O , entonces la cuarteta $\pi_A = OA \wedge r$, $\pi_B = OB \wedge r$, $\pi_P = OP \wedge r$, $\pi_Q = OQ \wedge r$ formada por las proyecciones de los puntos A, B, P, Q , es armónica.



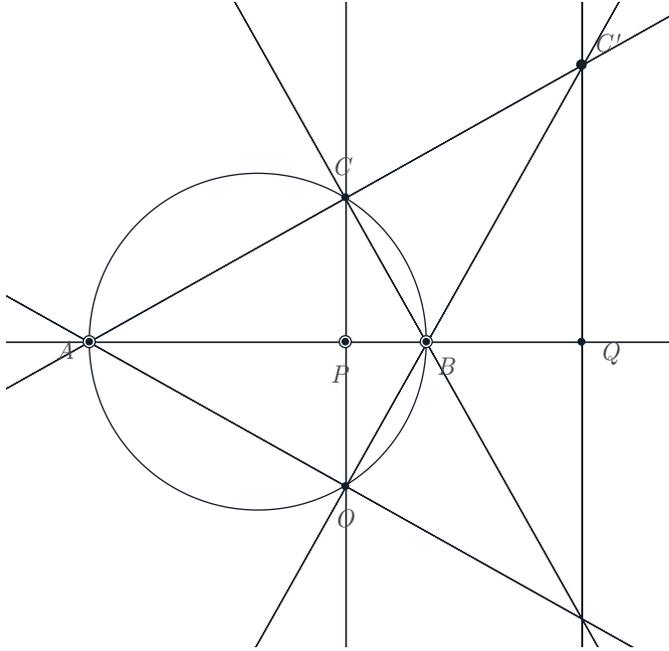
Dejamos las demostraciones formales de estos teoremas como ejercicios para el lector.

Es interesante observar que este teorema tan elemental no fue reconocido por el mundo matemático antes del siglo XIX. La noción de cuartetos armónicos surgió en la historia vinculada a una expresión en términos de tamaños de segmentos. Aún en la actualidad, si se le pregunta a un matemático qué es una cuarteta armónica, lo más probable es que responda que se trata de una cuarteta de puntos $ABPQ$ alineados que satisface la relación

$$\frac{AP}{PQ} = -\frac{AQ}{QB} \quad (3.1)$$

Ésta es una propiedad numérica de las cuartetos armónicos, que efectivamente es cierta cuando se introduce el concepto de distancia o al menos el de igualdad entre razones. Sin embargo utilizar esta

relación como la definición de cuarteta armónica oscurece en gran medida su carácter mucho más básico. La siguiente figura muestra una configuración de construcción de una cuarteta armónica con la cual es fácil deducir la propiedad 3.1.



Se construye el círculo con diámetro AB y se eligen O y C en la perpendicular a AB que pasa por P y se completa la construcción del armónico conjugado Q de P respecto a AB . Por semejanza de triángulos resulta que

$$\frac{AP}{AQ} = \frac{PC}{QC'} = \frac{PB}{BQ}$$

y por tanto

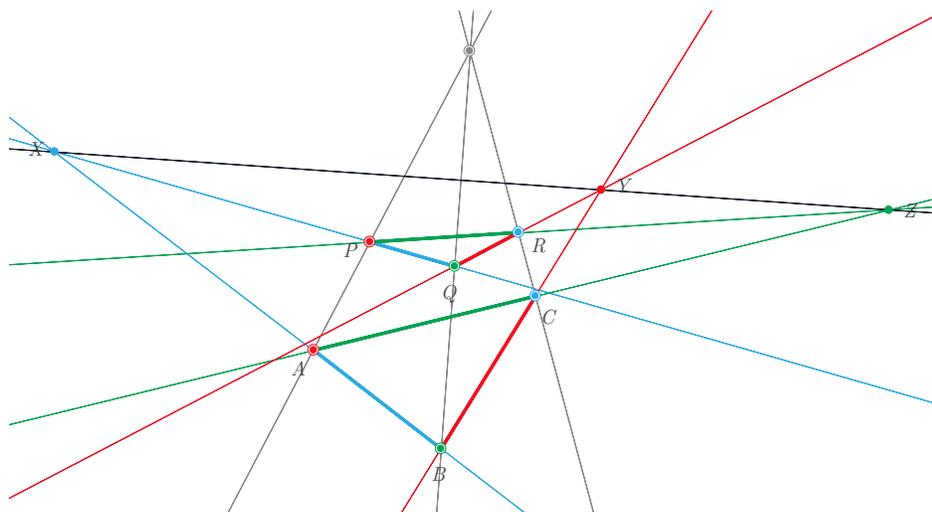
$$\frac{AP}{PB} = \frac{AQ}{BQ} = -\frac{AQ}{QB}$$

que es precisamente la fórmula de armonicidad 3.1.

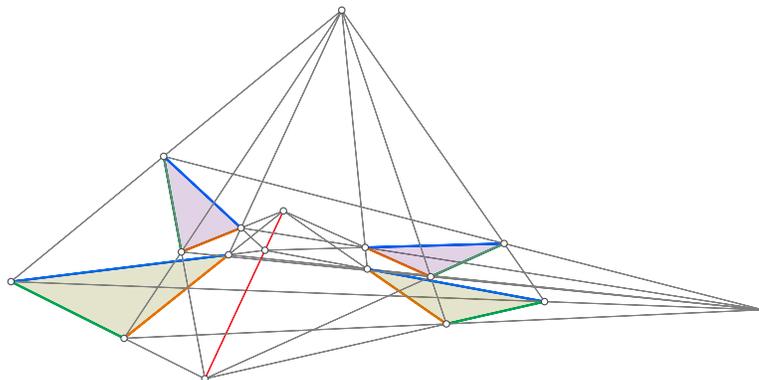
Pero hay que observar que para obtenerla fue necesario construir una circunferencia con centro en el punto medio de AB , una perpendicular a AB por P y usar las igualdades entre razones que la semejanza de triángulos proporciona, todo lo cual se aleja claramente de los conceptos básicos contemplados en los axiomas de conexión e incidencia.

El teorema de Desargues fue enunciado y demostrado por vez primera por el matemático francés GIRARD DESARGUES (1591 - 1661), por lo cual se le considera uno de los fundadores de la Geometría proyectiva.

Se dice que dos triángulos A, B, C y P, Q, R están en perspectiva si las rectas AP, BQ y CR coinciden en un punto. En ese caso, se dice que los vértices A y P son correspondientes, al igual que B y Q y también C y R . Análogamente, se dice que los lados AB y PQ son correspondientes, al igual que BC y QR y también CA y RP . En la siguiente figura los puntos correspondientes y los segmentos correspondientes están pintados en colores iguales.



El teorema de Desargues dice que si los triángulos A, B, C y P, Q, R están en perspectiva, entonces las intersecciones de sus lados correspondientes están alineadas, es decir, $X = AB \cap PQ$, $Y = BC \cap QR$ y $Z = CA \cap RP$ son colineales. Es fácil demostrar este teorema usando sólo argumentos de incidencia, mediante un *levantamiento a 3D*, como el que la siguiente figura ilustra.

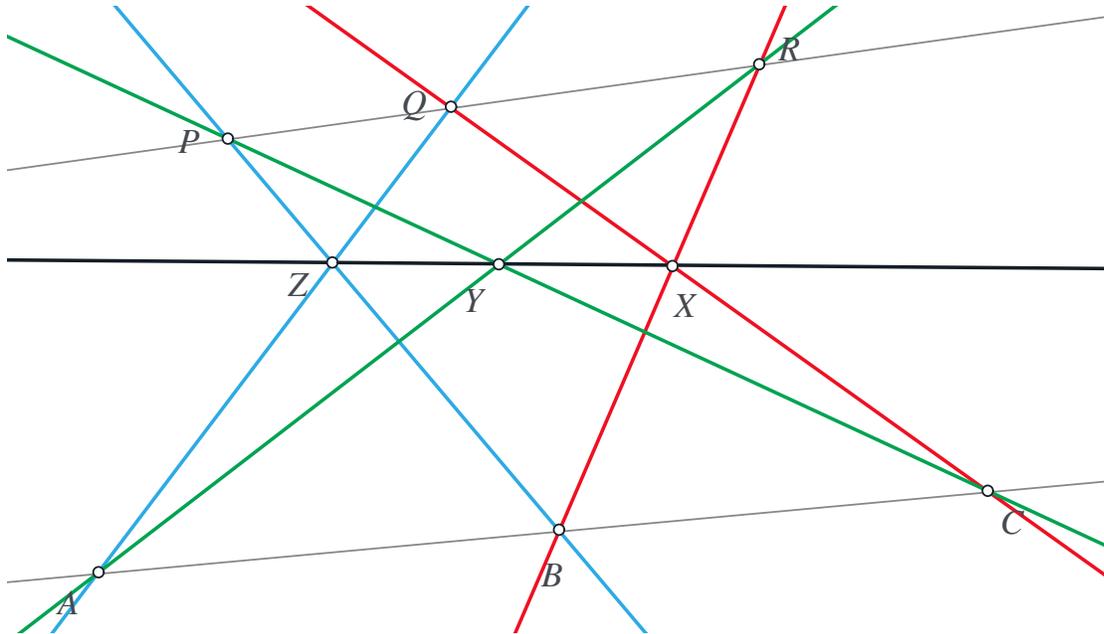


Cuando las intersecciones de los lados correspondientes de dos triángulos están alineadas, se dice que los triángulos están en *perspectiva axial*. La perspectiva normal se llama también *perspectiva polar*. Es fácil ver, utilizando los levantamientos a 3D, que también el inverso del teorema de Desargues se cumple. Así que podemos decir que dos triángulos están en perspectiva polar si y sólo si están en perspectiva axial.

El teorema de Pappus

Sean A, B, C y P, Q, R dos ternas alineadas de puntos. Coloreamos A y P en rojo, B y Q en verde y C y R en azul. Unimos por rectas los puntos de cada terna con los de la otra que tienen distinto color y hagámoslo con rectas del color distinto al de los dos puntos. Sean X, Y y Z las intersecciones de las rectas rojas, verdes y azules, respectivamente. El teorema de Pappus afirma

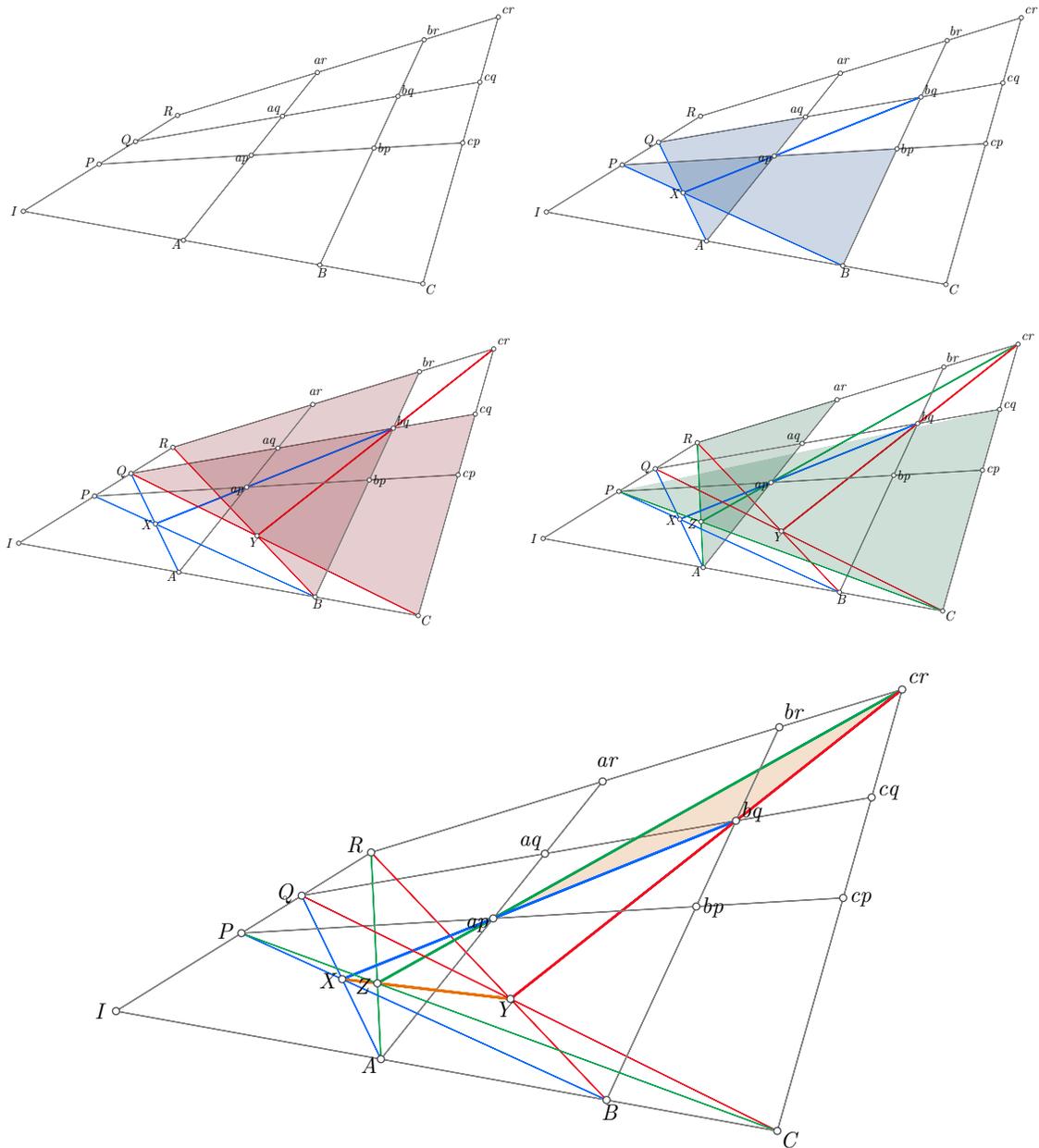
que X , Y y Z están alineados.



Este teorema también se puede demostrar mediante un levantamiento a $3D$, pero la demostración es más complicada que la del Teorema de Desargues. Pappus de Alejandría (290-350) enunció este teorema pero se desconoce cómo lo demostró (si es que lo demostró). Se dice que Blaise Pascal (1623-1662), con apenas 16 años, demostró un teorema más potente que el de Pappus, que dice afirma la colinealidad de ciertas intersecciones de las rectas que pasan por dos ternas de puntos que se encuentren sobre una curva cónica. Lamentablemente la obra del adolescente Pascal se perdió y tampoco sabemos cómo era esa demostración.

Aunque muchas veces se afirma lo contrario, el teorema de Pascal no es realmente más general que el de Pappus, pues su propio enunciado habla de curvas cónicas, concepto que no tiene sentido con solo los axiomas de conexión e incidencia.

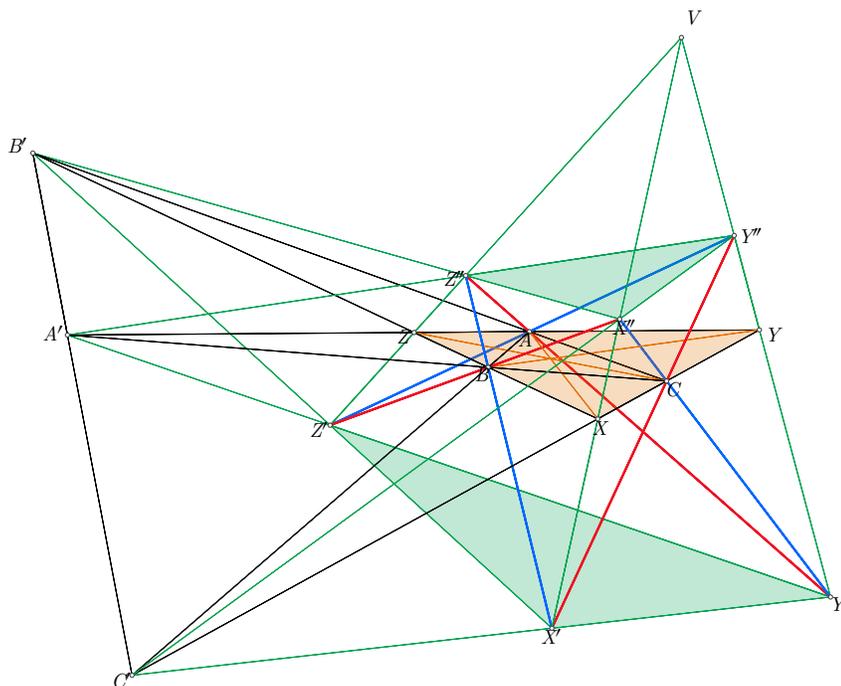
A principios del siglo XVIII Germain Pierre Dandelin (1794-1847) publica por fin una demostración del teorema de Pascal que procede mediante un levantamiento a $3D$ y que, cuando se especializa al teorema de Pappus, utiliza únicamente los axiomas de conexión e incidencia. Utiliza un teorema sobre superficies generadas por tres rectas en el espacio que no se tocan, que puede consultarse haciendo clic aquí: [©3D](#).



La demostración de Dandelin procede más o menos así. Primero se construye la estructura de la primera figura definiendo un punto ap fuera del plano $\alpha = AC \vee PR$, y se trazan las rectas indicadas, de manera que las que pasan por A , B y C no se toquen, las que pasan por P , Q y R tampoco se toquen, pero en cambio cada una de las primeras toque a todas las segundas. Puede verse que los puntos X , Y , Z son precisamente las intersecciones de los lados del triángulo $ap \vee bq \vee cr$ con el plano α , por lo cual necesariamente están alineados.

Las curvas cónicas en la geometría proyectiva

Teorema. Sean X, Y y Z tres puntos y sean $A \in YZ, B \in ZX$ y $C \in XY$. Sea V un punto fuera del plano $\pi = X \vee Y \vee Z$ y sea $X' \in VX$ distinto de V y de X . Definamos ahora $Y'' = X'C \wedge VY, Z' = Y''A \wedge VZ, X'' = Z'B \wedge VX, Y' = X''C \wedge VY$ y $Z'' = Y'A \wedge VZ$.



Entonces $Z'' = X'B \wedge VZ$ si y sólo si XA, YB y CZ son concurrentes.

DEMOSTRACIÓN. Supongamos primero que $Z'' = X'B \wedge VZ$. Entonces los triángulos $XYZ, X'Y'Z''$ y $X''Y''Z'$ están en perspectiva y, como A y B son dos de las intersecciones de sus lados correspondientes, la tercera será $C' = XY \wedge X'Y' = XY \wedge X''Y'' = XY \wedge AB$. Análogamente se puede ver que $A' = YZ \wedge Y'Z' = YZ \wedge Y''Z'' = YZ \wedge BC$ y $B' = ZX \wedge Z'X' = ZX \wedge Z''X'' = ZX \wedge CA$. Por otro lado los triángulos $X'Y'Z'$ y $X''Y''Z''$ están en perspectiva desde V y, por el teorema de DESARGUES, las intersecciones de sus lados correspondientes, que son precisamente A', B' y C' , están alineadas. Esto prueba que los triángulos XYZ y ABC están en perspectiva axial y por el inverso del teorema de DESARGUES también lo están en perspectiva polar, lo cual quiere decir que XA, YB y CZ son concurrentes. En la figura el punto en que concurren es O .

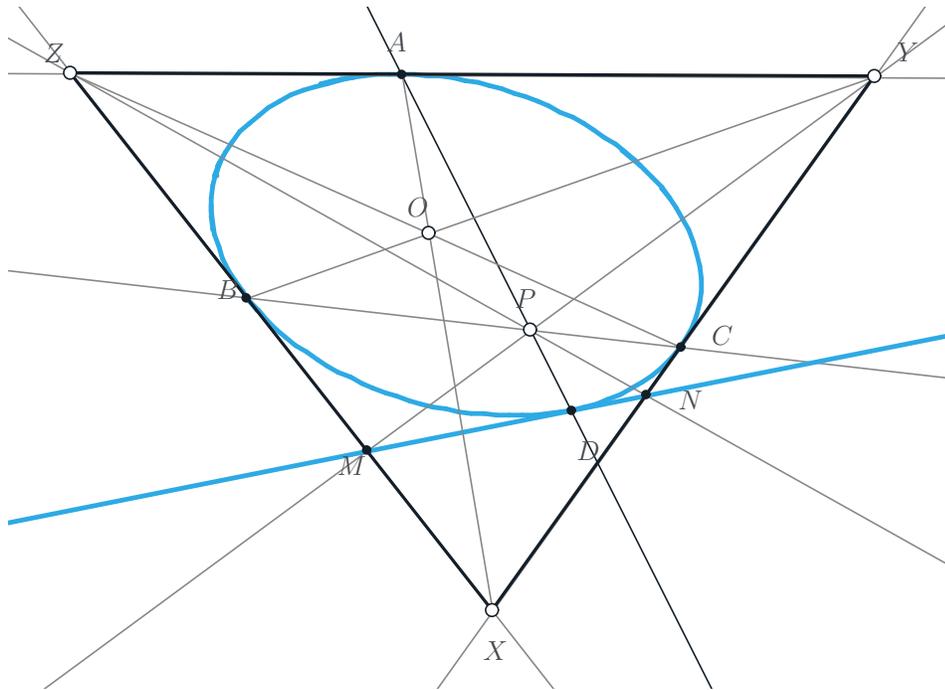
Para demostrar el inverso supongamos que XA, YB y CZ son concurrentes. Sea O el punto en que concurren. Entonces los triángulos XYZ y ABC están en perspectiva desde O . Sean A', B' , y C' los armónicos conjugados de A, B y C , respecto a YZ, ZX y XY , respectivamente. Sea $X' \in VX$ diferente de V y de X y sea X'' su armónico conjugado con respecto a VX . Construimos ahora $Y' = X''C \wedge VY$ y $Y'' = X'C \wedge VY$, por lo que van a ser armónicos conjugados uno del otro con respecto a VY , ya que C y C' lo son respecto a XY . Ahora construimos $Z' = Y''A \wedge VZ$ y $Z'' = Y'A \wedge VZ$. Z' y Z'' son armónicos conjugados uno del otro, ahora con respecto a VZ , ya que

A y A' lo son respecto a YZ . Finalmente, como B y B' son armónicos conjugados uno del otro con respecto a XZ , resulta que $B = Z_1X_2 \wedge X'Z''$. Esto demuestra que $Z'' = X'B \wedge VZ$. QED.

Gracias a este teorema sabemos que dado un triángulo XYZ y puntos $A \in YZ$, $B \in ZX$ y $C \in XY$ tales que XA , YB y ZC son concurrentes, entonces hay tres rectas (rojas) que pasan, una por A , otra por B y otra por C y otras tres rectas (azules) de las que también, una pasa por A , otra por B y otra por C , tales que cada una de las rojas intersecta a una de las azules, cada una de las azules intersecta a una de las rojas y ni las rojas ni las azules se intersectan entre sí. Además, la recta roja y la recta azul que pasan por A o por B o por C son armónicas conjugadas una de la otra con respecto a cierto punto V y al plano $X \vee Y \vee Z$. Las tres rectas rojas o las tres azules generan una superficie doblemente reglada (la misma, por el "axioma del equipal"). Probaremos que tal superficie se puede generar a partir de nuestra construcción de la curva cónica que dimos al principio e esta sección, y que su intersección con el plano $X \vee Y \vee Z$ es precisamente esa curva.

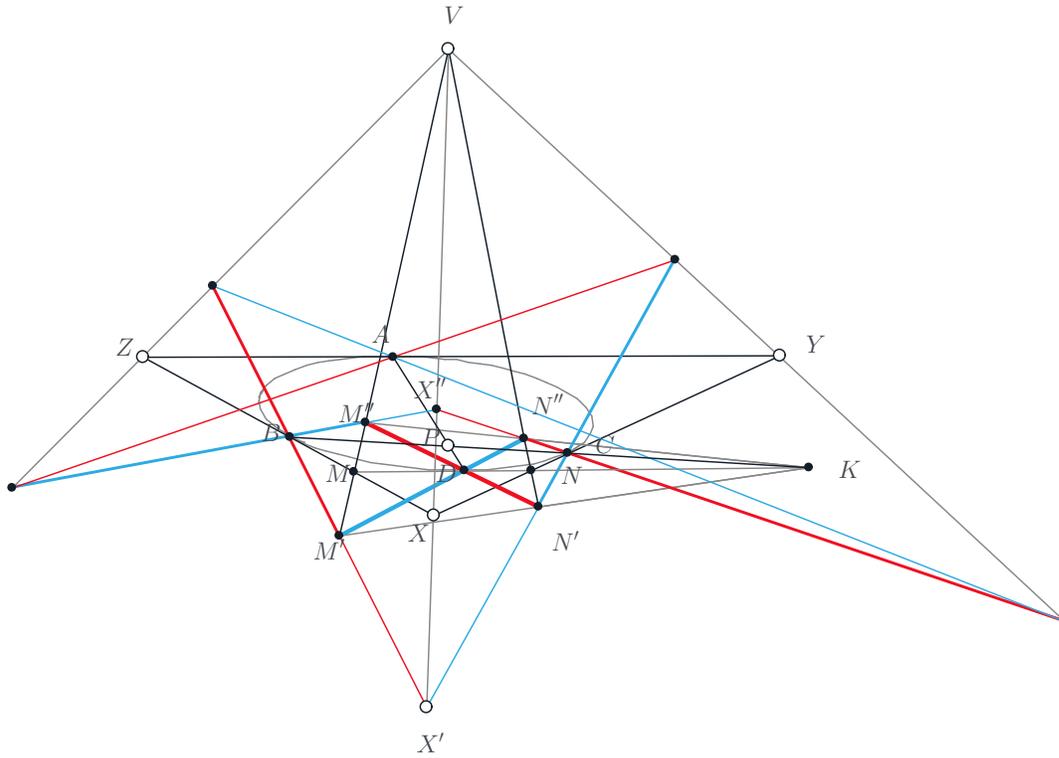
Es posible definir las curvas cónicas utilizando únicamente las propiedades de incidencia de los puntos y las rectas. Hay varias formas de hacerlo, mostraremos aquí una de ellas que, al combinarla con el teorema anterior permite demostrar que estas curvas son las intersecciones de planos con superficies doblemente regladas.

Sean X , Y y Z los vértices de un triángulo. Sea O un punto de $X \vee Y \vee Z$ que no esté en ninguno de los lados del triángulo. Sean $A = OX \wedge YZ$, $B = OY \wedge ZX$, $C = OZ \wedge XY$. Sea $P \in AB$ y sean $M = PY \wedge ZX$ y $N = PZ \wedge XY$. Finalmente, sea $D = AP \wedge MN$. Entonces, cuando P recorre la recta AB , el punto D genera una curva cuya tangente en D es precisamente MN .



Se puede ver que la curva así definida pasa por A , por B y por C y es tangente en esos puntos a las rectas YZ , ZX y XY , respectivamente. A las curvas así definidas les llamaremos en esta sección *curvas cónicas*. Cuando se introducen axiomas de congruencia, es posible demostrar que estas

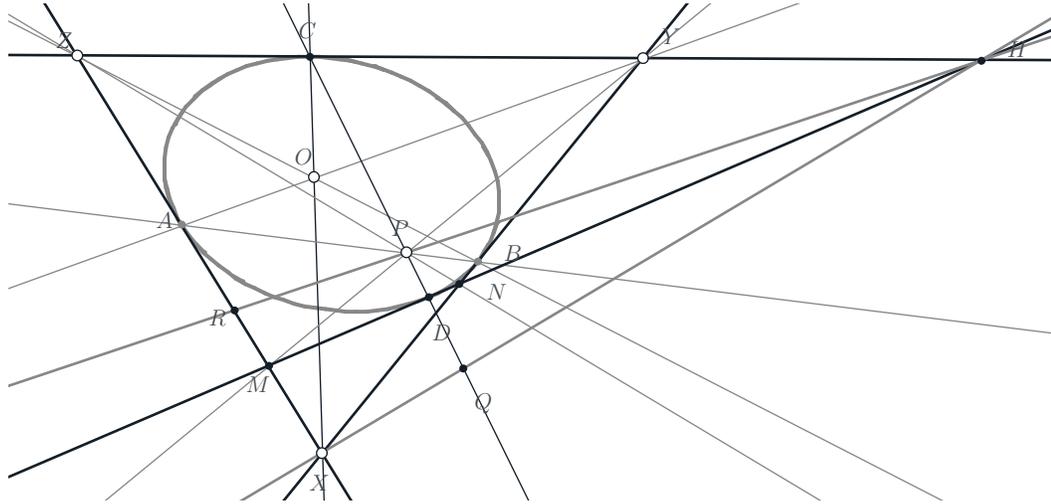
curvas coinciden con la que en la literatura clásica se llaman también “curvas cónicas”, de ahí la elección del nombre. Sin embargo, nuestra definición no requiere de la noción de *congruencia* ni de la de *círculo* o *cono circular*, como ocurre con la definición clásica.



Ampliamos ahora la construcción para obtener una superficie doblemente reglada cuya intersección con el plano $X \vee Y \vee Z$ es precisamente esta curva. Sea $M' = BX' \wedge VM$ y sea $N'' = DM' \wedge VN$. Sea $N' = CX' \wedge VN$ y sea $M'' = DN' \wedge VM$. Sea $K = M'N' \wedge M''N''$. Por construcción, $M'N'M''N''$ es un cuadrilátero cuyas diagonales se intersectan en D y sus lados en V y K . Las rectas $M'N''$ (azul) y $M''N'$ (roja) pertenecen a las familias de rectas que generan a la superficie doblemente reglada que se definió en la construcción anterior. De hecho la curva generada por D es precisamente la intersección de dicha superficie con el plano $X \vee Y \vee Z$. La existencia de esta superficie doblemente reglada nos regala el teorema de Pascal para estas curvas.

Demostraremos ahora que las curvas que hemos definido tienen ciertas propiedades de armonía.

Sea $H = MN \cap YZ$, sea $Q = PC \cap XY$ y se $R = PH \cap ZX$. Es fácil ver que X es el armónico conjugado de R con respecto a ZM (basta observar el cuadrilátero $MNYZ$). Como $ZMRX$ es una cuarteta armónica y su proyección desde H a la recta PC es la cuarteta $CDPQ$, esta última también es una cuarteta armónica. En particular, D es el armónico conjugado de C con respecto a PQ .

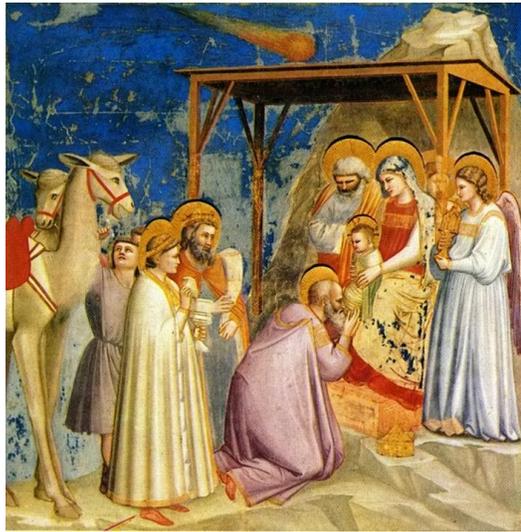


Estos resultados nos permiten definir las curvas cónicas, proyectivamente, como la intersección de una superficie doblemente reglada con un plano.

3.6. La perspectiva

La conquista de la realidad en la pintura

Los griegos tenían gran estimación por el retrato del cuerpo humano. En las estatuas más apreciadas se pueden apreciar los músculos y se dejan entrever los huesos, todo refleja una natura idealizada, en búsqueda de la forma ideal, divina. En las pinturas sobre los jarrones hay más representaciones alegóricas de sucesos mitológicos, hay movimientos plasmados, el paisaje juega un papel secundario, solo de adorno.



A la izquierda: la estatua de Hermes del escultor Praxíteles, alrededor de 340ac. A la derecha el cuadro “Adoración de los Reyes Magos” de Giotto di Bondone del siglo XIV.

Con el declive de la cultura griega, el arte se puso más al servicio de nuevas religiones como el cristianismo o el islam, donde la expresión realista pasaba a segundo plano ante una necesidad de poder expresar la nueva concepción del mundo, la importancia de la creencia, y los relatos de la nueva mitología alrededor del surgimiento de las creencias. Así se aprecia el contraste en las dos imágenes anteriores. En el cuadro de GIOTTO DI BONDONE (1266–1337), la expresión simbólica se antepone a la representación natural, aunque se aprecia el uso de un fondo más elaborado, de una casa que por deber sobresalir a las personas se alargó mucho en lo vertical. La imagen muestra un momento de la mitología cristiana, el nacimiento de Jesús Cristo y la por ello provocada llegada de reyes, llamados “magos”. Todo en el cuadro está en la función de este suceso y no de un retrato realista.

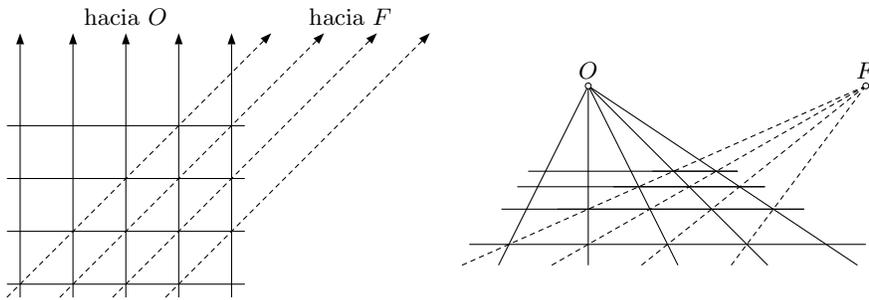
Con el descubrimiento de los antecedentes griegos en Europa, se abrió nuevo camino en el arte también. En el siglo XV se descubrieron los principios cómo había que pintar los objetos para que aparecieran más realistamente en relación con su ubicación. Se le atribuye a FILIPPO BRUNELLESCHI (1377–1446) el haber encontrado este secreto a principios el siglo XV, que es la *perspectiva*. El siguiente cuadro muestra una escena ficticia de personajes de la antigua Grecia y del renacimiento en Italia de aquel entonces congregados en el interior de un pabellón imaginario. Lo pintó RAFFAEL DI SANZIO (1483–1520) en los años 1509 y 1510 para el Papa Julio II sobre el muro en sus cuartos en el vaticano.



Raffael: “La escuela de Atenas”, estancias del Vaticano.

Se puede apreciar como el cuadro da la sensación del espacio, nos parece natural y realista porque el pintor aplicó sus conocimientos de la perspectiva.

El principio básico de la perspectiva es que rectas, que en realidad son paralelas, aparecen en un dibujo como si se fueran a “unir en el infinito”, se intersectan en el “punto de fuga”. Con este principio se puede calcular cómo se achican los objetos conforme se alejan del observador. Basta con usar una rejilla de cuadrados en el piso, de tal manera que unas líneas son paralelas a la línea de vista principal que une el ojo del observador con el centro O del cuadro a pintar. Como estas líneas son paralelas concurren en el cuadro en el punto de fuga O mientras las diagonales, que también son paralelas en realidad, se unen en el cuadro en otro punto F .



Rayos paralelos en realidad concurren en la imagen.

Al principio, los pintores italianos del renacimiento guardaron estos conocimientos como secretos de cofradía. Usar los puntos de fuga permitía estimar qué tanto se tenían que achicabar los objetos conforme se alejan del observador. El primer libro sobre la perspectiva “De picture” escribió LEON BATTISTA ALBERTI (1404–1472) en 1436. Después hay una serie de libros, con enfoques diferentes.

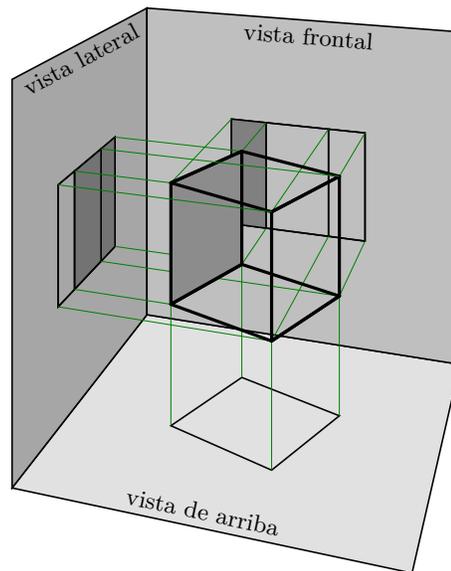
Unos se dirigen más al arquitecta o pintor, otros más a “teóricos”. Mencionamos dos: “De prospectiva pingendi” (ca. 1470) de PIERO DELLA FRANCESCA (1420–1492); “De Divina Proportione” (en 1509) de LUCA PACIOLI (1445–1517), ilustrado por LEONARDO DA VINCI (1452–1519), aunque el enfoque del último es menos sobre perspectiva que sobre proporciones.

Construcciones a partir de planos auxiliares

El uso de planos auxiliares fue un gran avance ya que permitió construir objetos geométricos en perspectiva. Los planos auxiliares son usualmente un plano horizontal, al cual se proyecta verticalmente cada punto. Si P es un punto en el espacio, a su proyección sobre el plano horizontal la denotaremos como P' . Un segundo plano es vertical y la proyección de P a ese plano la denotaremos como P'' .

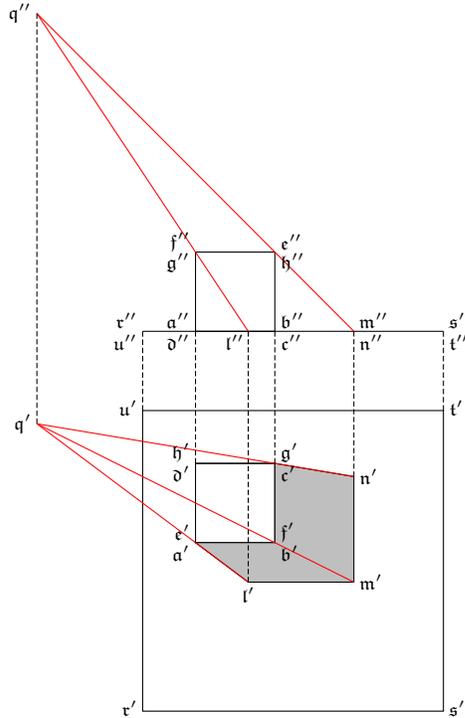
ALBERTO DURERO (1471–1528) publicó un libro cuyo título en alemán es *Underweysung der mesung mit dem zirckel un richtscheyt* y que se suele abreviar en español por “Instituciones de Geometría”. Apareció en 1525, dos décadas después de su viaje a Venecia, donde conoció a los pintores italianos del renacimiento Tiziano y Giorgione.

En este libro se explica la manera de usar diferentes planos, que también se llaman *vistas*, para construir una proyección central. Las vistas se obtienen al proyectar con luz paralela a un plano que es ortogonal a la dirección de la luz. Así se obtiene la vista de arriba, la vista frontal y la vista lateral de un objeto, tal como se muestra en el siguiente dibujo.



Durero empieza con un dado que coloca sobre una superficie cuadrada y plana y una luz que ilumina, tal como se muestra en la siguiente ilustración. Los cuatro puntos inferiores del dado son $abcd$, los superiores son $efgh$. Así que vemos en el plano visto desde arriba el punto e encima del punto a . Este plano se ve en la parte de abajo en la siguiente ilustración. Diferimos aquí de Durero, ya que denotamos las proyecciones verticales de estos puntos como e' y a' respectivamente. En

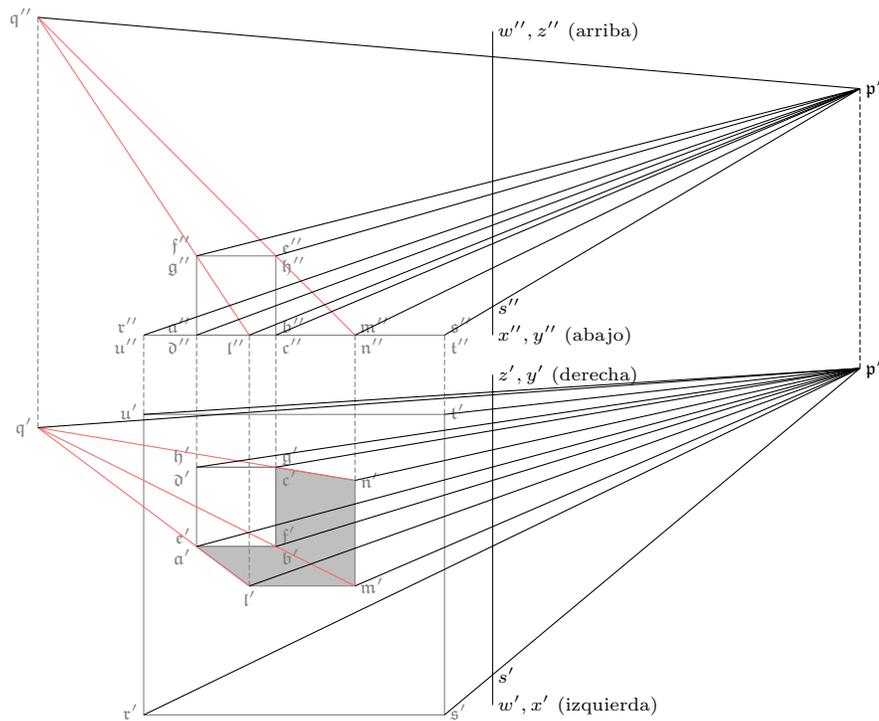
la vista de frente, que vemos en la parte superior, vemos el punto a delante del punto d y sus proyecciones horizontales a este plano vertical las denotamos con dos apóstrofes, es decir como a'' y d'' respectivamente.



La construcción de una sombra.

Una luz en q , cuya ubicación se puede ver en los dos planos como q' y q'' , ilumina la escena y produce una sombra sobre la superficie plana. Sea l la sombra de la esquina c . En la parte superior vemos la ubicación de l'' como la intersección de la recta $q''c''$ con la proyección de la superficie $rstu$ que se ve como línea horizontal $r''s''$. La recta qc se ve en el plano de abajo como $q'e'$. Sobre ella tiene que encontrarse l' . Pero las dos proyecciones l' y l'' de un mismo punto siempre se unen por rectas paralelas a las dos proyecciones de cualquier otro punto, por ejemplo $q'q''$, se obtiene l' con facilidad. De la misma manera se construye las dos proyecciones en ambos planos de las sombras m y n de las esquinas f y g respectivamente.

Ahora, Durero añade el ojo del espectador, es decir el punto p , desde donde se observa la escena. Indica las dos posiciones p' y p'' en ambos planos. Se une p' con cada punto del plano abajo y p'' con cada uno del plano arriba. Ahora se intersectan estas líneas con el plano del lienzo. El lienzo es un rectángulo $wxyz$ que vemos como línea en ambos planos. En la parte de arriba vemos la altura de un punto, por ejemplo s del borde inferior del lienzo, en la parte de abajo vemos la distancia al borde derecho.



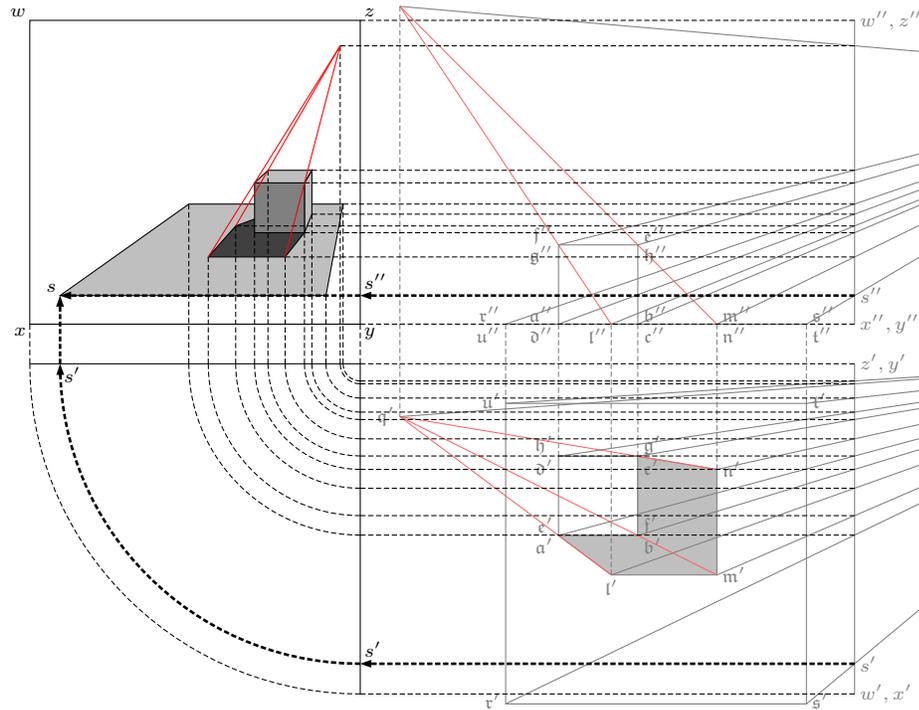
La proyección de los puntos a un lienzo en los dos planos.

Durero indica que se debe tomar una nueva hoja y medir estas distancias con un compas y pasarlas a la nueva hoja, y aunque las instrucciones son correctas, son poco claras ya que no las exhibe geoméricamente a partir de la construcción ya hecha. Preferimos aquí seguir con el mismo dibujo. Para ello observamos un punto, por ejemplo la esquina s de la superficie cuadrada. En el cuadro este se verá como el punto s y las proyecciones s' y s'' en los dos planos ya están a nuestra disposición.

Movemos de manera paralela las dos proyecciones del lienzo hacia la izquierda para contar con espacio vacía, ver la siguiente figura. Ahí se rota la proyección del lienzo por extremo derecho por 90° . De esta manera, los lados izquierdos y derechos de la imagen ya quedarán en su lugar. Luego se cruza la recta vertical por s' con la recta horizontal por s'' para obtener s .

Así se puede proceder con todos los puntos para obtener el dibujo completo.

Durero manejaba el uso de estos planos magistralmente. En una obra destinada al diseño y las proporciones del cuerpo humano, muestra dibujos que obtiene al introducir nuevos planos que no son paralelos a ninguna de las aristas de un cubo.



La construcción de la imagen a partir de los puntos en las dos proyecciones del lienzo.

El descubrimiento de los principios de la perspectiva, produjo una revolución en la pintura. Cuando antes se retrataban sobre todo motivos alrededor de la religión, donde se representaban ideas, después se retrataban cada vez más motivos de la naturaleza, de una manera realista. Como en una carrera, los pintores compitieron en querer llegar a la expresión más realista posible.

¿Por qué las proyecciones de rectas paralelas concurren?

En su libro Durero también muestra el grabado de la siguiente figura. Es una manera de componer una imagen punto a punto sin la necesidad de recurrir a construcciones geométricas. El ojo desde el cuál se retrata el cuadro se encuentra en la pared en la ubicación del anillo por el cual pasa la cuerda. Durero dice que lo que se ve sólo se ve en línea recta y no curva y de dos objetos en línea recta se ve lo que está más cerca del ojo.

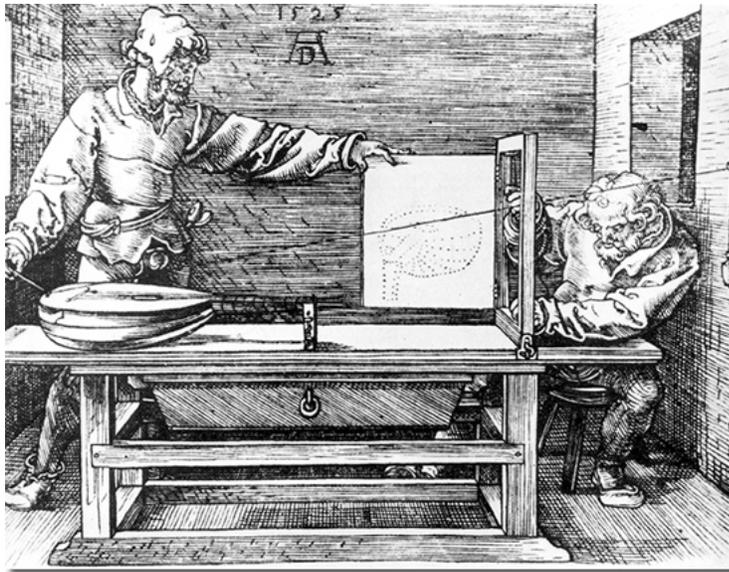


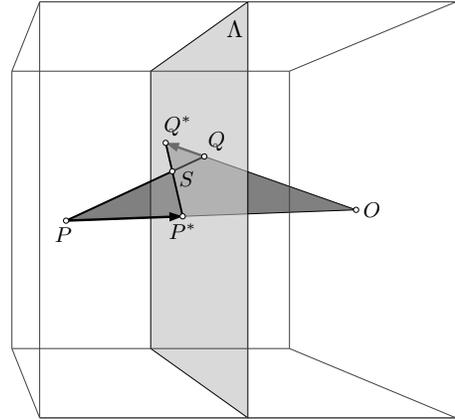
Ilustración de Durero, de como se puede medir el punto en la imagen, al cual es proyectado un punto de la realidad (sobre el borde de la mandolina), si este se une con un punto fijo (el punto en la pared por donde pasa la cuerda).

Esta idea, que un punto real P , el punto en la imagen P' que representa P y el ojo O son siempre *colineales*, es decir se encuentran sobre una recta, es a la vez sencilla y muy poderosa. De ello se puede inferir que rectas paralelas en la imagen o son paralelas o concurren en un punto.

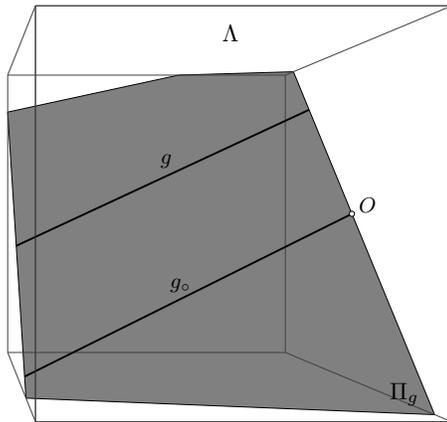
Parece que la primera demostración de este hecho se encuentra en “*Perspectivae libri sex*” (1600) de GUIDOBALDO DEL MONTE (1545–1607). Su primera demostración es bastante enredada, así que preferimos dar una más moderna. Para ello usaremos la visión de GIRARD DESARGUES (1591–1661) en 1639: el de considerar rectas doblemente infinitas y no segmentos acotados como hacían los antiguos griegos.

Primero aclaramos la situación: el lienzo, o *plano de proyección* se llama Λ . El ojo, o *centro de proyección* se llamará O . Un punto P se proyecta a P^* , es decir P^* es la intersección de la recta OP con Λ . Como se muestra en nuestra ilustración del lado P puede estar – visto desde O – detrás de Λ o delante de Λ como es el punto Q . El segmento PQ es proyectado al segmento P^*Q^* .

El punto de intersección S del segmento PQ con Λ es igual a su proyección, es decir se tienen $S = S^*$.



Para seguir necesitamos una observación sencilla.



Una recta g que no pasa por O junto con el punto O definen un plano en el espacio. En lo que sigue denotaremos a este plano por Π_g , ya que – como el centro de proyección O lo consideramos fijo – solo depende de la recta g .

Este plano Π_g contiene la recta g_o , que es la recta paralela a g que pasa por el punto O .

La observación es: g_o está contenido en Π_g .

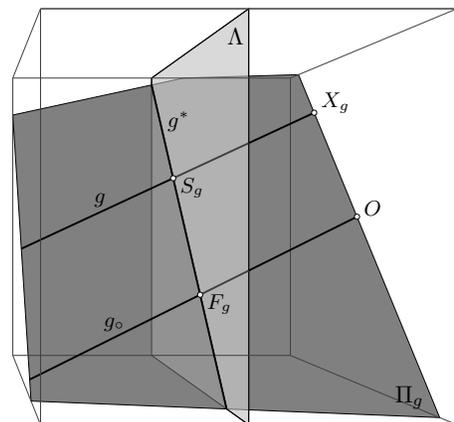
Ahora podemos seguir.

Sea g una recta que no pasa por O . Consideramos el caso en donde g no es paralelo a Λ y por lo tanto intersecta Λ en un punto, que llamamos S_g .

Si P es un punto de g , entonces la recta PO se encuentra en el plano Π_g . Por lo tanto también P' se encuentra en Π_g . Esto muestra que la proyección g^* de la recta g al plano Λ es la intersección de Λ con Π_g .

Como el plano Π_g contiene la recta g_o , también contiene al punto de intersección F_g de g_o con Λ . Con ello tenemos dos puntos de g^* : el punto S_g y el punto F_g .

La proyección de g es entonces la recta $g^* = S_gF_g$.



Ahora concluimos con facilidad: si h es paralela a g , entonces $h_{\circ} = g_{\circ}$ y por ello $F_g = F_h$. Esto muestra que g^* y h^* se intersectan en el punto $F_g = F_h$, como lo hará la proyección de cualquier otra recta paralela a g . Por lo tanto, F_g es el *punto de fuga* de g , es decir, tiene la propiedad de que la proyección de cualquier recta paralela a g pasa por él.

Si g es paralelo a Λ , la recta g_{\circ} no intersecta a Λ y la proyección g^* es paralela a g . Si h es paralelo a g entonces h^* es paralela a h y por lo tanto h^* y g^* son paralelos. Se trató de asignarle a ellos un “punto de fuga al infinito, en dirección de g ” y se decía, que las rectas paralelas se intersectaban “al infinito”.

Hay un punto que no se proyecta a Λ : es el punto X_g , que es la intersección de g con un plano por O que es paralela a Λ . Por otro lado hay un punto de la recta g^* que no se obtiene como proyección de algún punto de g : este es el punto F_g . No se puede obtener por que g y g_{\circ} son paralelos. Si observamos el proceso inverso: de proyectar g^* desde O a la recta g pasa algo similar: F_g no se puede proyectar y X_g no se puede obtener como imagen. Y como se verá más tarde, en la Sección 3.5 sobre la geometría proyectiva, X_g se puede considerar el “punto de fuga” de g^* al igual que F_g es el punto de fuga de g .

Es sorprendente que los matemáticos se tardaran alrededor de 300 años en formular lo que estaba a punto de ser descubierto en aquel momento. ¿Si los puntos que se alinean con el ojo se ven como un solo punto, por qué no identificarlos? Sólo mucho más tarde dieron este paso. Claro, en un principio faltaba el descubrimiento de la utilidad de coordenadas. Con coordenadas, la formulación es así:

Se consideran las ternas (x, y, z) diferentes del origen $(0, 0, 0)$ – que es la ubicación del ojo – y se establece una relación de equivalencia: $(x, y, z) \sim (x', y', z')$ si existe un número $\lambda \neq 0$ tal que

$$\begin{aligned}x' &= \lambda x, \\y' &= \lambda y, \\z' &= \lambda z.\end{aligned}$$

Esto es equivalente a decir que los tres puntos (x, y, z) , (x', y', z') y $(0, 0, 0)$ son colineales, es decir se encuentran sobre una recta. Se fija un plano Σ en el espacio que no pasa por el ojo, por ejemplo $z = 0$. Este plano juega el papel del lienzo. Muchas clases de equivalencia de puntos contienen un representante en Σ . Por ejemplo el punto $(3, -4, 5)$ es equivalente a $(\frac{3}{5}, -\frac{4}{5}, 1)$, un punto que se encuentra en Σ . Solo los puntos que se encuentran en $z = 0$ no se pueden identificar con puntos en Σ .

Una recta ℓ en el espacio se compone de puntos. Usualmente casi todos de estos puntos se pueden – al usar la relación de equivalencia – identificar con puntos en Σ . Lo que obtenemos en Σ es una recta ℓ' , que se puede ver como la intersección de Σ con el plano que contiene a ℓ y al ojo $(0, 0, 0)$.

Todavía se hicieron algunos avances en el siglo XVII, debidos sobre todo a GIRARD DESARGUES (1591–1661). Consideró rectas infinitas, y consecuentemente doble conos y no solo conos. Así consideró el cilindro como un (doble) cono con el vértice al infinito. Investigó las secciones cónicas desde un punto de vista general: estaba interesado en propiedades comunes a todas las secciones cónicas. Su obra inspiró a FERMAT, DESCARTES y PASCAL pero en su mayoría los matemáticos no recibieron bien su obra, que cayó en el olvido durante mucho tiempo.

3.7. La geometría hiperbólica

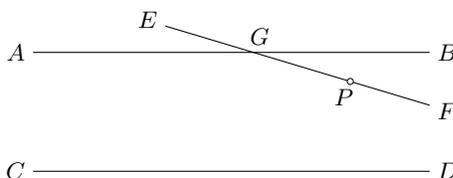
Los argumentos deductivos permiten estudiar cuáles teoremas pueden deducirse de algunos supuestos sin utilizar otros. El quinto postulado de EUCLIDES siempre fue sospechoso de quizás no ser indispensable, es decir, que pudiera haber alguna manera de demostrarlo a partir de los otros, aunque se les tuvieran que agregar otros postulados con tal de que fueran más fáciles de aceptar por la intuición. Hubo muchos intentos de eliminar de una u otra manera el famoso e incómodo quinto postulado, pero cuanto más se trabajaba en ello y con más rigor se analizaba el problema, más parecía ser independiente de los otros. Esta situación llevó a varios matemáticos de finales del siglo XVIII y principios del XIX a concebir la posibilidad de que se pudiera definir una o varias geometrías que no cumplieran con el quinto postulado, es decir, que lo negaran explícitamente. Esto llevó eventualmente a la creación de modelos geométricos claros en los que lo que se llamaba recta era sustituido por otro tipo de curva de tal manera que o bien hubiera más de una paralela a otra por un punto fuera de ella, o bien no hubiera ninguna.

Intentos de demostrar el quinto postulado

Recordamos de la Sección 3.2 que EUCLIDES basó el libro *Los Elementos* en 5 postulados (ver página 44). De estos postulados todos fueron aceptados excepto uno, que causó controversia, el quinto postulado o “axioma de paralelas”. En comparación con los demás no fue considerado igual de intuitivo, evidente o natural. Por ello muchos matemáticos, aunque reconocieron el gran trabajo de EUCLIDES, lo criticaron por admitir entre sus axiomas tal afirmación.

La lista de los matemáticos que intentaron demostrar el postulado de las paralelas, a partir de los cuatro primeros, es larga y comprende varios griegos, árabes y europeos.

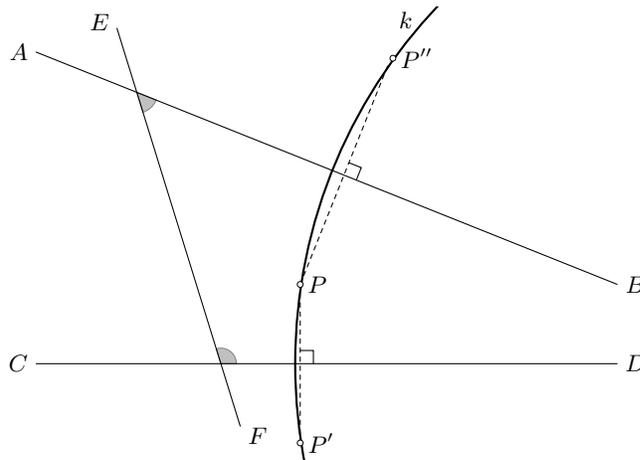
Veamos uno de estos intentos: PROCLO (410–485) argumentó así: Sean AB y CD dos paralelas, es decir dos segmentos rectos que aún al prologarse no se intersectan. Luego sea EF una recta que intersecta AB en el punto G y de tal manera que F se encuentra entre AB y CD , ver la siguiente figura.



Ahora se supone que un punto P recorre de G hacia F y más allá por la prolongación. Entonces la longitud de la perpendicular a AB por P se hará más grande y rebasará cualquier cota definida. Por consecuente EF tiene necesariamente, si se prolonga lo suficiente, que intersectarse con CD .

El error en la argumentación anterior es que se asume tácitamente que de dos rectas que no se intersectan como AB y CD , los puntos de AB están todos a la misma distancia de CD . Esto es cierto en la geometría euclídea, pero no se puede demostrar sin usar el quinto postulado. De hecho, esta propiedad es equivalente al quinto postulado. Por ello el argumento de PROCLO no muestra que el quinto postulado sigue de los cuatro anteriores, sino sólo que el quinto postulado demuestra a sí mismo.

Veamos un segundo intento de demostrar que el quinto axioma es consecuencia de los primeros cuatro. El argumento es de FARKAS BOLYAI (1775–1856). El argumento es el siguiente: A partir de la situación de la premisa del quinto postulado, es decir de tener dos rectas AB y CD con una transversal EF tales que los ángulos interiores suman menos que dos rectos, se puede construir un punto P tal que los reflejos P' resp P'' de P en AB resp. CD no son colineales, es decir que no se hallan sobre una recta.

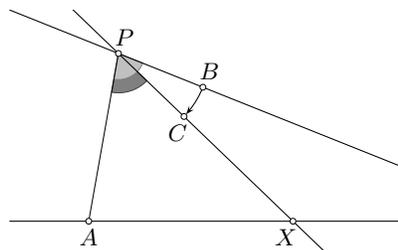


Luego el argumento de FARKAS BOLYAI concluye: sea k la circunferencia por P , P' y P'' y sea U el centro de esta circunferencia. Dado que AB resp. CD es la mediatriz de PP' resp. de PP'' , esta pasa por U . Por ello AB y CD se intersectan en U . Por ello AB y CD se intersectan.

El fallo de esta demostración consiste en asumir que cualquier triángulo tiene una circunferencia circunscrita. Curiosamente, esta demostración la publicó en 1851, tres décadas después de que su hijo publicó su trabajo, que ahora es considerado como uno de las dos obras de descubrimiento de una nueva geometría.

Los errores en estas demostraciones surgen por una parte de confusiones entre diferentes conceptos de paralelismo. Como *paralela* a una recta g por un punto P podemos entender

- a) una recta por P que no intersecta a g ,
- b) el lugar geométrico de los puntos que tienen la misma distancia de g que P , o
- c) la recta que es perpendicular a la perpendicular a g en P .
- d) una recta PB que no intersecta a g , tal que cualquier recta PC con $\sphericalangle APC < \sphericalangle APB$ intersecta a g (A es un punto en g).



Las cuatro definiciones son equivalentes en la geometría de EUCLIDES, pero para poder demostrarlo se usa el axioma de las paralelas. En lo que sigue usaremos “paralelo” para indicar que dos rectas no tienen punto en común.

Otro error en los intentos surge de asumir propiedades que son equivalentes al quinto postulado:

- e) La suma de ángulos en cualquier triángulo es 180° .
- f) De cada triángulo se puede construir uno semejante en cualquier tamaño.
- g) Cualquier triángulo tiene una circunferencia circunscrita.

Suponer uno de los conceptos de paralelo equivalente a otro o asumir uno de las propiedades equivale a suponer que el quinto postulado es válido. Muchas veces se sustituye el quinto postulado por el equivalente *axioma de Playfair*:

Si P es punto que no está en la recta g entonces hay exactamente una recta por P que no interseca a g .

Cambio de perspectiva

Entre los intentos de derivar el quinto postulado de los demás, hubo varios que trataron una demostración *reduccio ad absurdum*: suponían que el quinto postulado era falso y trataron de hallar una contradicción. El primero que hizo esto con mucha insistencia, fue SACCHERI, el segundo JOHANN HEINRICH LAMBERT (1728–1777). Ambos abandonaron sus intentos, después de desarrollar una serie de consecuencias de los primeros cuatro postulados y el contrario del quinto, sin encontrar una contradicción clara. Lo que encontraron fueron propiedades de una geometría que después se conoce como “no euclidiana”. No encontraron una contradicción porque no había ninguna.

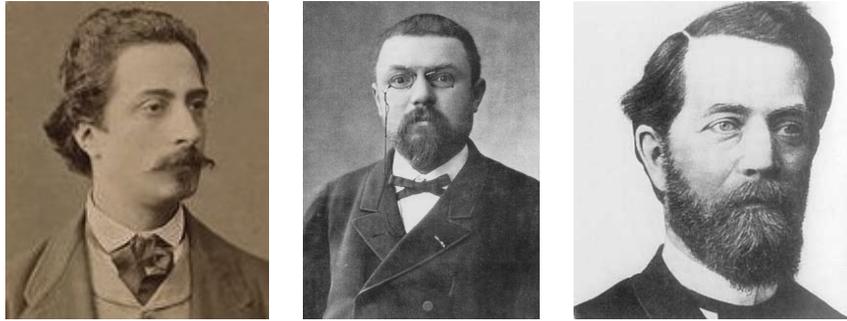
Casi un siglo más tarde hubo nuevamente dos personas que suponían los primeros cuatro postulados y el contrario del quinto. Sólo que ahora no buscaban una contradicción, sino trataron de explorar las consecuencias porque pensaron que una geometría diferente a la euclidiana era posible. Esos dos hombres eran NIKOLAI IVANOVICH LOBACHEVSKI (1792–1856) y JANÓS BOLYAI (1802–1860). Publicaron sus hallazgos de manera independiente, es decir sin saber del otro, en 1826 resp. en 1831. Hubo un tercer matemático que pensó en aquel entonces que otra geometría era posible: CARL FRIEDRICH GAUSS (1777–1855). Se sabe por sus diarios. GAUSS no publicó nada, porque temía la reacción de la comunidad matemática y con cierta justificación: la obra de LOBACHEVSKI y BOLYAI no recibió el respeto y la distinción que merecía.



JOHANN HEINRICH LAMBERT NIKOLAI LOBACHEVSKI JANÓS BOLYAI CARL FRIEDRICH GAUSS

Los modelos promueven la aceptación

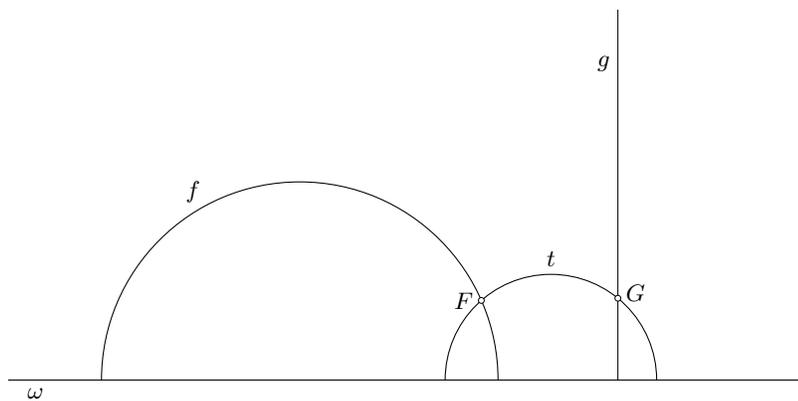
En 1868 EUGENIO BELTRAMI (1835–1900) publicó un trabajo en el cuál mostraba que la geometría expuesto en los trabajos de LOBACHEVSKI y BOLYAI es libre de contradicciones, si es que la geometría euclidiana lo era. Para ello describió tres modelos, que ahora se llaman curiosamente el modelo de “disco de KLEIN”, el modelo de “disco de POINCARÉ” y el modelo del “semiplano de POINCARÉ”, en honor al alemán FELIX KLEIN (1849–1925) y HENRI POINCARÉ (1854–1912).



EUGENIO BELTRAMI HENRI POINCARÉ FELIX KLEIN

En estos modelos se explicarán qué habra que entender bajo “punto” o “recta”. Es decir define lo que son los objetos geométricos importantes en el modelo de una manera inesperado y lo describe en términos de la geometría euclidiana. Una recta en el modelo no tendrá que ser recta en el sentido usual, es decir en el sentido de la geometría euclidiana. Los conceptos importantes que explicaremos son lo que es un “punto”, una “recta”, un “ángulo” y la “reflexión” en una “recta”.

Para describir el modelo del semiplano de POINCARÉ se fija una recta euclidiana ω en el plano. Esta divide al plano en dos semiplanos, de los cuales a uno de ellos llamaremos “superior”. Los puntos del modelo serán los puntos del semiplano superior, sin los puntos de la recta euclidiana ω . Las rectas del modelo son rayos o semicircunferencias del plano euclidiano que tocan a ω en un ángulo recto. La siguiente figura muestra ω , dos rectas paralelas del modelo f, g – una es un semicírculo, la otra un rayo como objeto euclidiano – y una tercera recta del modelo t que intersecta las anteriores en los puntos F y G respectivamente.



Los ángulos en el modelo se miden como los ángulos normales, es decir usando tangentes en el punto

de intersección.

Si g es una recta del modelo que es un rayo en el plano euclidiano, entonces la reflexión en el modelo es la reflexión en la recta que determina g , en el sentido de la geometría euclidiana.

Para describir lo que es una reflexión en una recta como f arriba, será importante recordar las propiedades de la inversión en un círculo c :

- fija todos los puntos sobre la circunferencia de c .
- intercambia el interior del c (menos el centro) con el exterior de c ,
- conserva ángulos (que se piden entre curvas con las tangentes en el punto de intersección),
- la imagen de un círculo o una recta es una circunferencia o una recta,

Sea ahora f una recta del modelo, que es un semicírculo. Denotaremos con P' el punto que se obtiene al invertir el punto P . De las propiedades de la inversión mencionados sigue que

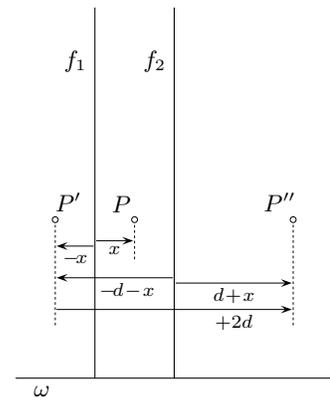
- se tiene $\omega' = \omega$,
- los puntos del semiplano “superior” van a dar a puntos del semiplano “superior”,
- la parte interior del disco que está en el semiplano “superior” da a la parte exterior del semiplano “superior”.

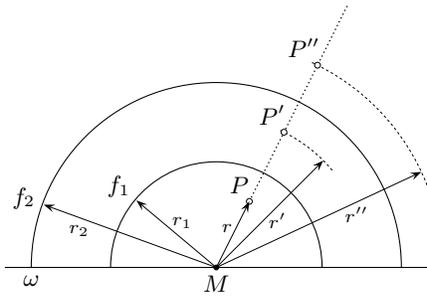
Estas propiedades indican que la inversión en f puede servir como reflexión en la recta f pensado en el modelo. A saber, las propiedades de la inversión y las que arriba se mencionaron se traducen en propiedades de la reflexión en la recta f pensado en el modelo:

- los puntos del espejo f quedan fijos,
- los dos lados de la recta se intercambian,
- los ángulos del modelo no cambian al reflejar,
- la imagen de una recta del modelo, es decir de un rayo o un semicírculo del plano euclidiano que toca ω en un ángulo recto, es nuevamente un rayo o un semicírculo que toca ω en un ángulo recto, es decir nuevamente una recta del modelo.

La distancia en el modelo se puede dar usando una fórmula, pero es suficiente y hasta más útil para nosotros aquí, decir que dos segmentos del modelo tienen la misma longitud si hay una serie de reflexiones en rectas tal que un segmento es la imagen del otro segmento bajo estas reflexiones.

Veremos dos casos particulares de rectas paralelas f_1 y f_2 en el modelo: en el primero las dos rectas son rayos en el plano euclidiano, en el segundo son dos semicírculos concéntricos. En ambos casos vemos un punto P “arriba” de ω . Denotamos con P' el reflejo de P en f_1 y con P'' el reflejo de P' en f_2 . Nos interesa analizar qué es la función que manda P en P'' . En el primer caso es una traslación euclidiana paralelo a ω por lo doble de la distancia entre los dos rayos.



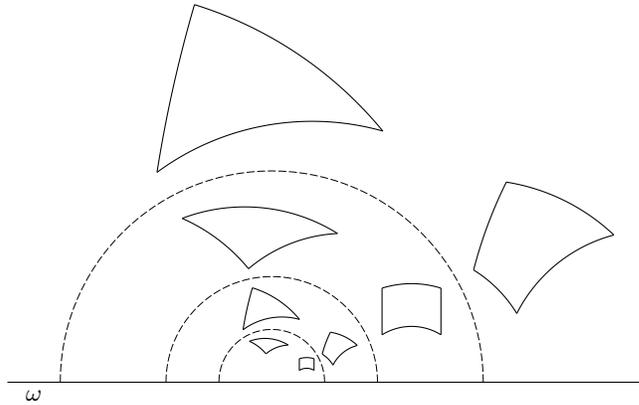


El segundo caso es más interesante. Sea r , r' y r'' la distancia euclidiana de P , P' resp. P'' del centro común M de f_1 y f_2 . Entonces se tiene

$$r' = \frac{r_1^2}{r} \quad \text{y por ello} \quad r'' = \frac{r_2^2}{r'} = \frac{r_2^2}{r_1^2} \cdot r.$$

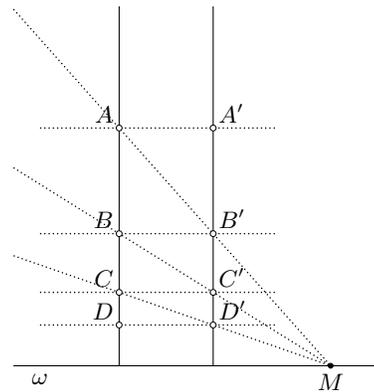
Esto muestra que la composición de las dos reflexiones en estas dos rectas del modelo es en el plano euclidiano una homotecia con centro M y factor $\frac{r_2^2}{r_1^2}$.

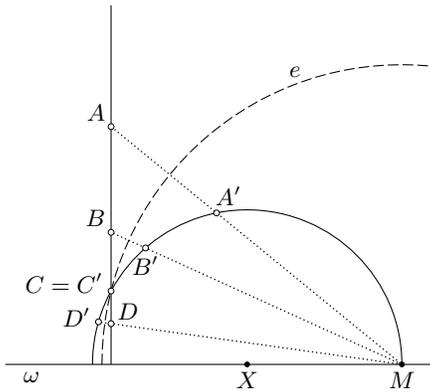
La siguiente figura muestra un triángulo y un cuadrilátero que se reflejaron tres veces. Se observa como las figuras se hacen euclidianamente pequeñas al acercarse a ω . Pero hay que recordar que los cuatro triángulos y los cuatro cuadriláteros son congruentes entre si en el modelo.



Esto permite marcar con facilidad puntos sobre rectas f_1 y f_2 del modelo que sean rayos en el plano euclidiano, ver la ilustración a la derecha. Como la traslación euclidiana horizontal es una congruencia en el modelo, se tiene que las siguientes distancias en el modelo sean iguales: $AB = A'B'$, $BC = B'C'$ y $CD = C'D'$.

Por otro lado, la homotecia con centro en M es también una homotecia en el modelo, por lo que las siguientes distancias en el modelo también son iguales: $AB = B'C'$ y $BC = C'D'$, ver la ilustración a la izquierda. Por ello se tiene para las distancias en el modelo $AB = BC = CD$.

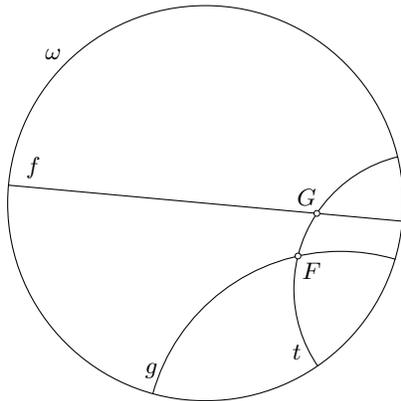




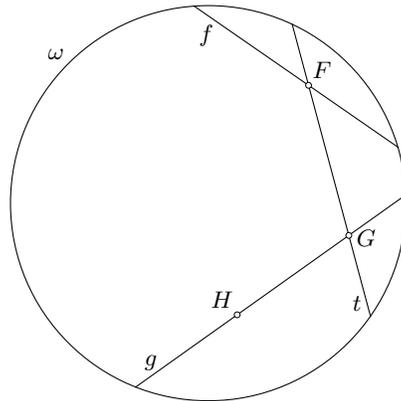
Otro caso interesante es comparar distancias sobre dos rectas del modelo que en el plano euclidiano sean distintos, es decir, una f_1 que es rayo y la otra f_2 que es un semicírculo. Sea M uno de los puntos donde f_2 toca ω . La inversión en un círculo con centro M intercambia f_1 y f_2 . Por ello $AB = A'B'$, $BC = B'C'$ y $CD = C'D'$, por lo que de $AB = BC = CD$ sigue que $A'B' = B'C' = C'D'$.

Los dos casos muestran: con un número finito de pasos igual de grandes en el modelo no es posible llegar a ω , es decir en el modelo, los puntos de ω (que no forman parte del modelo) están infinitamente lejos. Con ello se demostró el postulado que cualquier segmento recto en el modelo se puede prolongar tanto como se quiere. Dejamos al lector la tarea de verificar de también se satisface el postulado de que por cualquier dos puntos del modelo se define de manera única una recta del modelo.

Hay otros dos modelos famosos. Uno de ellos se obtiene por una inversión: Si se coloca un círculo euclidiano ε en el semiplano inferior y se invierte en ε entonces el semiplano superior se convierte en el interior de un disco ω' . Esto da el modelo de disco de POINCARÉ. En él los puntos del modelo son los puntos euclidianos en el interior de ω' . Las rectas del modelo son los diámetros euclidianos de ω o los arcos de circunferencias euclidianas que tocan a ω' en un ángulo recto. La siguiente figura de la izquierda muestra la circunferencia ω' , dos rectas paralelas f, g y una tercera recta t que intersecta las anteriores en los puntos F y G respectivamente.



El disco de POINCARÉ



El disco de KLEIN

A la derecha se muestra el modelo del disco Klein. Para definirlo se fija igualmente una circunferencia euclidiana ω . En este modelo, los puntos son los puntos del plano euclidiano en el interior de ω , las rectas son cuerdas rectas en el sentido euclidiano. Los ángulos son más difíciles de medir que en los otros dos modelos. La figura anterior muestra a la derecha ω , dos rectas paralelas f, g y

una tercera recta t que intersecta las anteriores en los puntos F y G respectivamente.

Cada uno de los modelos tiene ventajas y desventajas. Usamos el semiplano de Poincaré ya que es más fácil exponer algunas características del modelo, pero cualquiera de los tres modelos explica la misma geometría. Estos modelos ayudaron mucho en que la nueva geometría se aceptara. También sirven para darse cuenta de algunas de las propiedades peculiares y muy diferentes que en la geometría euclidiana.

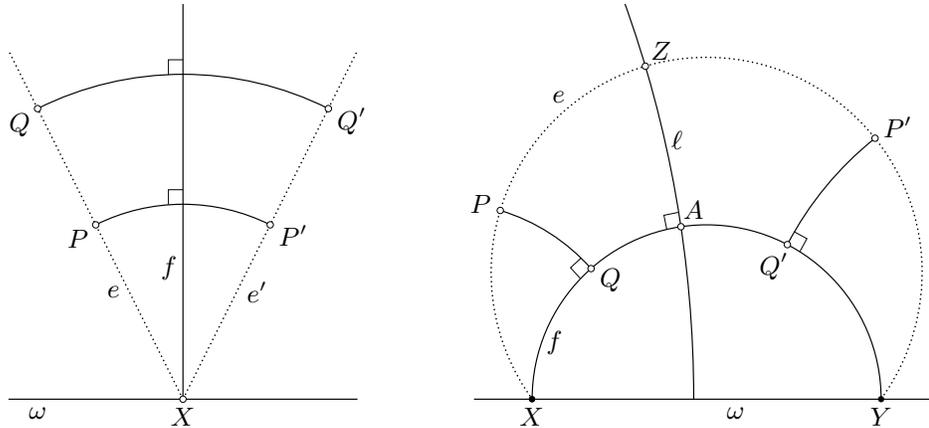
Algunas propiedades de la geometría hiperbólica

En lo que sigue veremos que se cumplen los primeros cuatro postulados de Euclides, más no el quinto. Veremos varias propiedades de la geometría descrito en este modelo, una geometría que ahora se conoce como *hiperbólica*. Estas propiedades son a veces bastante diferentes y a veces bastante similares de las que conocemos de la geometría euclidiana. Sugerimos al lector siempre comparar con los enunciados análogos de la geometría euclidiana.

Primero nos ocuparemos de la pregunta del lugar geométrico de todos los puntos que están en el modelo a la misma distancia de una recta dada que un punto dado. Resulta que este lugar geométrico no es una recta del modelo. Por ello necesitamos darle un nuevo nombre: se llama *equidistante*.

P1: *Una equidistante a una recta es un rayo euclidiano o un arco de circunferencia euclidiano.*

Sea f la recta y P el punto que están dados. Si f es un rayo en el plano euclidiano, que toca a ω en el punto X , entonces este lugar geométrico consiste de dos rayos de X , ver la siguiente figura a la izquierda.



Ahora consideramos el caso donde la recta del modelo f es un semicirculo del plano euclidiano. Como en el caso anterior, la distancia de un punto hacia una recta se mide con la perpendicular. Ahora se piensa un punto A móvil sobre la recta f y en ella erigido la perpendicular l . El reflejo en l manda a f en una recta f' que pasa por A y es perpendicular en A a l . Esto muestra $f = f'$. Dado ello, Q' tiene que ser un punto de f y $P'Q'$ perpendicular a f .

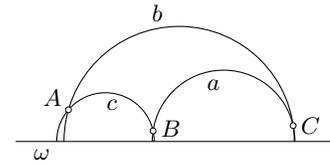
Ahora bien, sea e la circunferencia del plano euclidiano que pasa por los puntos X y Y donde f toca ω y además por P . Sea Z el punto de intersección de e con l . La reflexión en l intercambia

X y Y y fija Z . Por lo tanto manda e en sí mismo: $e' = e$. Esto muestra P' es la intersección de e con la recta perpendicular del modelo en Q a f . Pero el mismo argumento muestra que todos los puntos de e están a la misma distancia de f que P en el modelo. Por ello e y su reflejo e' en f conforman el lugar geométrico de todos los puntos que en el modelo están a la misma distancia de f que P .

Debe notarse que este lugar geométrico, se conforma por dos *equidistantes*, esto son en el plano euclidiano, arcos de circunferencia que pasan por X y Y , los puntos donde toca f a ω . Estas equidistantes no son rectas en el modelo ya que no tocan a ω en un ángulo recto, sino son curvas.

Mostraremos algunas propiedades más de esta geometría, que se conoce ahora como geometría hiperbólica.

Hay triángulos cuya suma de ángulos es muy pequeña, como lo muestra la siguiente figura a la derecha.



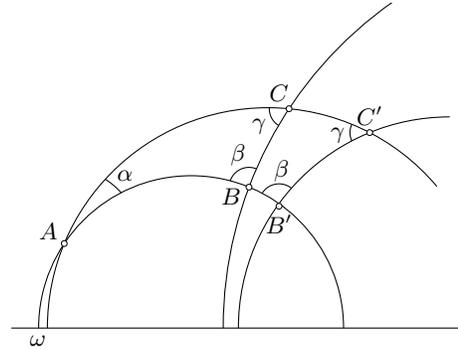
P2: Cada triángulo del modelo tiene una suma de ángulos menor a dos rectos.

Esta propiedad se puede demostrar con la ayuda de la inversión, pero aquí nos saltaremos este argumento y nos concentraremos en sus consecuencias. De ello sigue que cada cuadrilátero tiene una suma de ángulos menor a cuatro rectos. Pero hay una consecuencia mucho más sorprendente:

P3: Dos triángulos que tienen los mismos ángulos en el modelo son congruentes en el modelo.

Supongamos lo contrario: que haya dos triángulos ABC y $A'B'C'$ que tengan los mismos ángulos, pero que no sean congruentes. Podemos suponer que $A = A'$ sino reflejamos uno de ellos en la mediatriz de AA' .

Luego podemos suponer que AB y $A'B' = AB'$ definan la misma recta, sino reflejamos uno de ellos en la bisectriz del ángulo $\sphericalangle BAB'$. Ahora hay dos casos posibles: o C y C' están del mismo lado que $AB = A'B'$ o no los son. En el primer caso coinciden las rectas AC y $A'C' = AC'$, dado que los dos triángulos tienen al mismo ángulo α , en el segundo coinciden después de reflejar uno de los triángulos en AB . Obtenemos entonces la situación de la figura a la derecha. Pero esto es imposible ya que el cuadrilátero $AA'B'B$ tendría una suma de ángulos igual a cuatro rectos, que es imposible. Por lo tanto ABC y $A'B'C'$ tienen que ser congruentes.



En la demostración anterior se usa un hecho interesante en sí mismo. Al igual que en la geometría euclidiana, se pueden llevar cualesquiera dos triángulos congruentes uno en el otro por medio de a lo más tres reflexiones:

P4: Cualquier congruencia es composición de reflexiones y se necesita a lo máximo tres reflexiones para una congruencia dada.

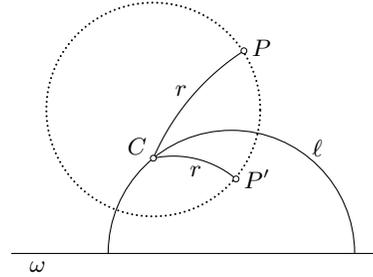
Por ello las reflexiones son como las piezas principales de las congruencias. Las traslaciones que

nos suelen ser muy familiares de la geometría euclidiana, son a cambio en la geometría hiperbólica bastante diferentes y los trataremos al final de este apartado.

Al igual como se averiguó las equidistantes como lugar geométrico, se puede averiguar qué es una circunferencia en el modelo, es decir el lugar geométrico de todos los puntos en el modelo que están a la misma distancia de un punto C que un punto dado P .

P5: *Un círculo en el modelo es un círculo euclidiano que está contenido completamente en el semiplano superior.*

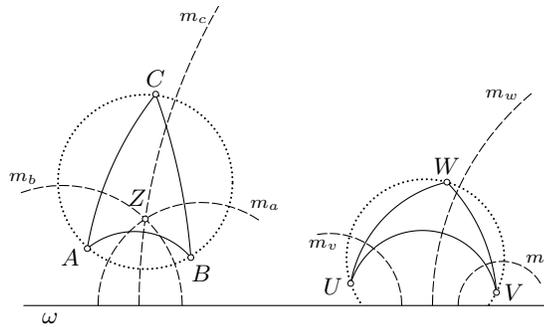
Se considera una recta ℓ que pasa por C y en ello se refleja P para obtener otro punto P' del modelo que está a la misma distancia, ver la siguiente figura a la derecha. Al girar ℓ alrededor de C se obtiene que una circunferencia en el modelo es una circunferencia en el sentido euclidiana que está completamente contenido en el semiplano superior.



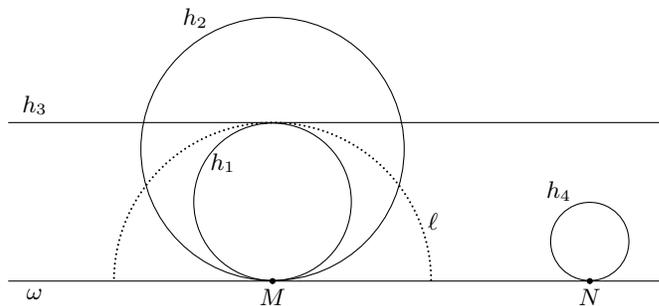
Cualquier tres puntos del modelo son –en el plano euclidiano– o colineales o cocíclicos, es decir se encuentran sobre una circunferencia. Pero esta circunferencia no tiene que ser completamente comprendido en el semiplano superior.

P6: *Hay triangulos del modelo que no tienen circunferencia circunscrita*

En la figura del lado se muestran dos triangulos ABC y UVW del modelo, uno tienen una circunferencia en el modelo y la otra no, sus puntos están sobre una equidistante y las tres mediatrices m_u , m_v y m_w no se intersectan.

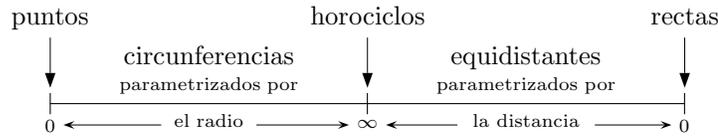


Círculos euclidianos que están completamente dentro del semiplano superior son círculos del modelo – aunque los centros no coinciden. Círculos euclidianos que cortan a ω en dos puntos se pueden interpretar como equidistantes a la recta que tiene estos dos puntos como puntos *límite*. Pero, ¿qué sería un círculo euclidiano que toca ω en un punto dentro del modelo? Resulta que esto es otro objeto interesante, se llama *horociclo*. La siguiente figura muestra cuatro horociclos: h_1 , h_2 , h_3 y h_4 .



Que la recta euclidiana h_3 , paralela a ω también es un horociclo del modelo, sigue de que h_3 se obtiene al reflejar h_1 en la recta ℓ . Como las homotecias con centro sobre ω son congruencias en el modelo, todos los horociclos son congruentes. Como los puntos y las rectas, de horociclos solo hay una clase bajo la conjugación.

Una equidistante determina la recta del modelo a la que es equidistante, ya que determina los dos puntos límites. Por ello, las equidistantes están parametrizadas por la distancia a la recta que determinan, similar como los círculos están parametrizados por el radio. Los horociclos se pueden ver como el caso límite de círculos con radio infinito o como equidistantes que están infinitamente lejos de su rectas. Así que los objetos en la geometría hiperbólica se pueden agrupar de la siguiente manera.

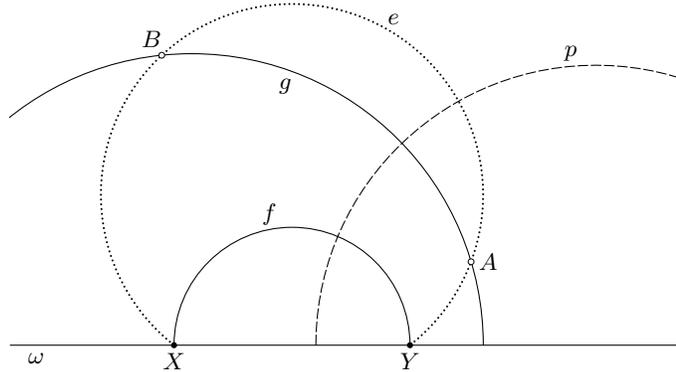


Esto muestra que la geometría hiperbólica es mucho más interesante. En la geometría euclidiana horociclos y equidistantes coinciden con las rectas.

Veamos finalmente cómo se comportan las traslaciones en el modelo de la geometría hiperbólica. Dos rectas en el modelo pueden intersectarse en un punto o pueden tener un punto límite en común. Pero también es posible un tercer caso: que no se intersectan, ni tienen punto límite en común.

P7: *Dos rectas que no se intersectan y no tienen punto límite en común, tienen una única recta perpendicular en común, es decir una recta que es perpendicular a ambas rectas dadas.*

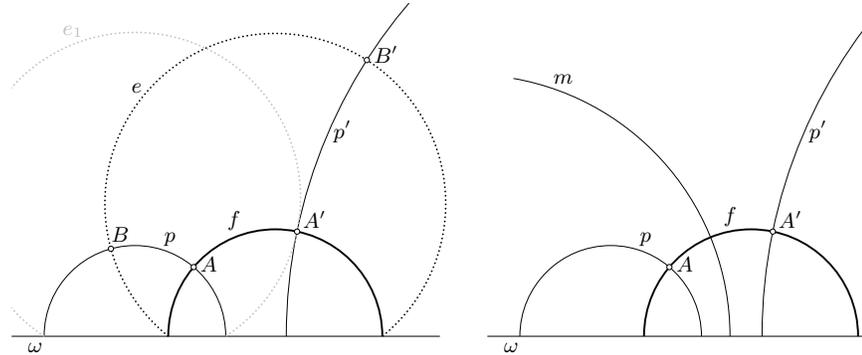
Para ver esto llamamos las dos rectas dadas f y g . Elegimos un punto A en g . Sea e la equidistante a f por A y B el segundo punto de intersección de e con g , ver la figura del lado. Finalmente sea p la mediatriz de AB .



La argumentación es la siguiente: la reflexión en p intercambia A y B . Por ello fija la recta $g = AB$. Esto implica que g y p se intersectan en un ángulo recto. Entonces cualquier círculo por A y B se intersecta con p en un ángulo recto. De ahí que p y e se intersectan en un ángulo recto y por ello la reflexión en p fija también la equidistante e e intercambia X y Y , los dos puntos límites de e y f . Por ello la reflexión en g también fija f y por ello p es perpendicular a f .

Para definir una traslación en la geometría hiperbólica es necesario especificar a lo largo de qué recta g se traslade y qué tan lejos. Esto se puede hacer al especificar un punto A de recta g y su imagen A' bajo la traslación. Dado que una traslación conserva ángulos debe enviar la perpendicular p a g en A a la perpendicular p' a g en A' . Un punto B de p se moverá a un punto de p' y este estará a la misma distancia de g que B . En otras palabras, B y B' están sobre una equidistante a g , ver la siguiente imagen. La siguiente figura a la izquierda muestra la situación. Se aprecia también que B

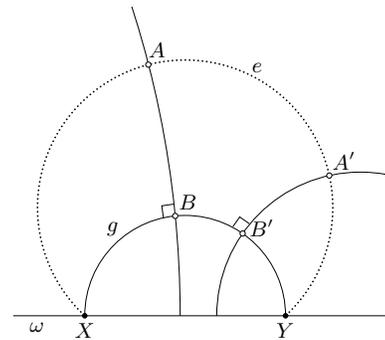
se mueve más lejos que A : obsérvese para ello la equidistante e_1 en gris que indica los puntos que están a la misma distancia de p que lo es A' . Se observa que B' es más lejos.



Una manera práctica de definir la traslación es como la composición de dos reflexiones: primero en la perpendicular p y luego en la mediatriz m de AA' , ver la siguiente imagen anterior a la derecha. Claramente esta composición mueve A en A' y traslada todos los puntos sobre f a lo largo de esta recta. Como consecuencia los demás puntos se mueven por equidistantes a f . Denotaremos a $\tau_{A,A'}$ a la traslación definida de esta manera.

P8: Hay una infinidad de traslaciones que mueven un punto dado en otro punto dado.

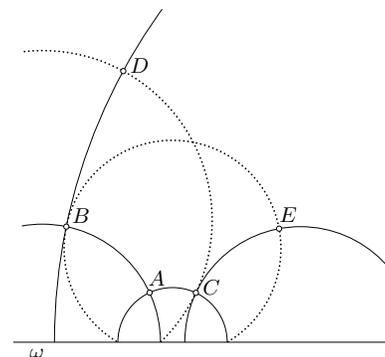
Sean A y A' dos puntos dados. Hay una infinidad de traslaciones que mueven A en A' . Para verlo, basta dibujar una equidistante e cualquiera por A y A' . Luego sea g la recta que define e – es decir la recta a la cual e es equidistante y B resp. B' la intersección de las perpendiculares a g por A resp. A' , ver la ilustración a la derecha. Entonces la traslación $\tau_{B,B'}$ que mueve a lo largo de la recta BB' el punto B en B' moverá A en A' , pero no será la misma traslación que $\tau_{A,A'}$.



En la geometría euclidiana traslaciones *conmutan*, es decir no importa en qué orden se ejecutan, el resultado será lo mismo. En la geometría hiperbólica esto no es así.

P9: Las traslaciones no conmutan

Para ver esto consideramos tres puntos A , B y C del modelo de tal manera que el ángulo $\sphericalangle BAC$ es recto –esto no es muy importante pero muestra el efecto claramente. Luego consideramos las composiciones de las traslaciones $\tau_{A,B}$ y $\tau_{A,C}$, ver la imagen anterior a la derecha. Veremos a donde se mueve el punto A después de efectuar las dos traslaciones. Claro, $\tau_{A,B}$ mueve A a B . La segunda traslación $\tau_{A,C}$ mueve B al punto E . Por otro lado, si primero efectuamos $\tau_{A,C}$, el punto A se mueve a C y luego este se mueve a D bajo $\tau_{A,B}$. Por lo tanto las dos posibles composiciones no son los mismos.



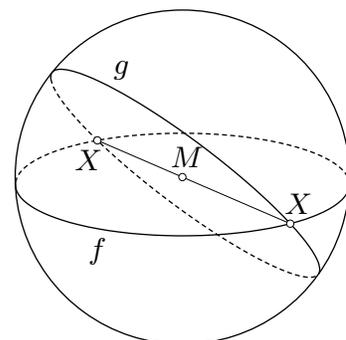
La geometría elíptica

En base al descubrimiento de la geometría hiperbólica, los matemáticos se dieron cuenta que hay otra geometría, una tercera a parte de la euclidiana y la hiperbólica: la geometría *elíptica*. En ésta el quinto postulado es también falso. Se conocía un modelo de esta geometría desde hace mucho tiempo pero no se interpretaba como un modelo de una geometría “plana”: se trata de una geometría sobre la esfera.

En este modelo se toma una esfera euclidiana. Un punto del modelo es un par de puntos euclidianos que son antípodas en la esfera. Una recta del modelo es un círculo máximo de la esfera, es decir un círculo euclidiano que se obtiene al intersectar la esfera con un plano que pasa por su centro.

La figura muestra dos rectas f y g del modelo que se intersectan en un punto X . Los ángulos en el modelo se miden como los euclidianos entre las tangentes en los puntos de intersección.

No es difícil mostrar que en este modelo no hay rectas paralelas.



La importancia del descubrimiento

El descubrimiento de la geometría hiperbólica marcó un punto de transición en la historia de las matemáticas. Antes nadie dudaba que al hacer geometría uno investigaba las propiedades del espacio que nos rodea. Después, cuando ya no era claro cuál es la geometría de nuestro espacio, las matemáticas se habían despedido de describir la realidad, se tuvieron que limitar a estudiar las diferentes posibilidades que existen. Cuál es realmente la geometría de nuestro espacio es una cuestión de la física y no de las matemáticas.

De esta manera se aclaró el papel que juegan las matemáticas dentro del tejido de las ciencias: las matemáticas proveen un lenguaje inmensamente versátil, pero no deciden cuál de los ejemplos que se pueden estudiar es el que mejor describe nuestra realidad.

Hay otro aspecto que se aclaró con el descubrimiento de las nuevas geometrías: no tiene sentido tratar de aclarar qué es un punto o qué es una recta. En vez de ello se deben buscar aclarar las relaciones que tienen unos objetos con otros. En vez de buscar definiciones explícitas hay que dar definiciones implícitas. Los objetos se definen únicamente por sus relaciones con los otros objetos. Este es el punto de vista que adoptó DAVID HILBERT (1862–1943) al formular sus axiomas de la geometría y es lo que quiso indicar cuando dijo que en vez de “punto”, “recta” y “plano” se debe poder decir también “tarro”, “silla” y “mesa”. Para hacer geometría lo único que se necesitan son *las relaciones entre los objetos que se describen en los axiomas*.

El estudio de la geometría hiperbólica ayuda a afinar nuestra visión de la geometría euclidiana: hace reflexionar sobre cuestiones que muchas veces tomamos por garantizadas. Es similar al hecho de que el aprendizaje de una segunda lengua lleva a las personas a reflexionar sobre su propia lengua y a conocerla mejor.

3.8. Los axiomas de orden y el de Euclides

Los axiomas de orden de Hilbert son éstos.

AXIOMAS DE ORDEN

1. Si A, B, C son puntos de una línea recta y B está entre A y C , entonces B también está entre C y A .
2. Si A y C son dos puntos de una línea recta, hay al menos un punto B entre A y C y al menos un punto D tal que C está entre A y D .
3. Dados tres puntos en una línea recta, siempre hay uno y solo uno de ellos que está entre los otros dos.
4. Dados cuatro puntos en una recta siempre se pueden nombrar como A, B, C, D de manera que B esté entre A y C y también entre A y D , y que además C esté entre A y D y también entre B y D .
5. Sean A, B, C tres puntos que no están en una misma recta y sea a una recta en el plano $A \vee B \vee C$ que no pasa por ninguno de esos puntos. Entonces si la recta a pasa por un punto entre A y B , también pasará por un punto entre B y C o bien por un punto entre A y C .

Al conjunto de puntos que están entre dos puntos dados A y B se les llama *el segmento AB* . A veces, para distinguir entre la recta definida por dos puntos A y B y el segmento definido por ellos mismos conviene denotar la primera por $A \vee B$ y al segundo por AB . Con estas herramientas se pueden definir también las curvas poligonales (o líneas quebradas) y los polígonos simples, los primeros como cadenas de segmentos que no se tocan y los segundos como curvas poligonales simple, excepto porque el primer extremo del primero coincide con el segundo extremo del último.

CONSECUENCIAS DE LOS AXIOMAS DE ORDEN

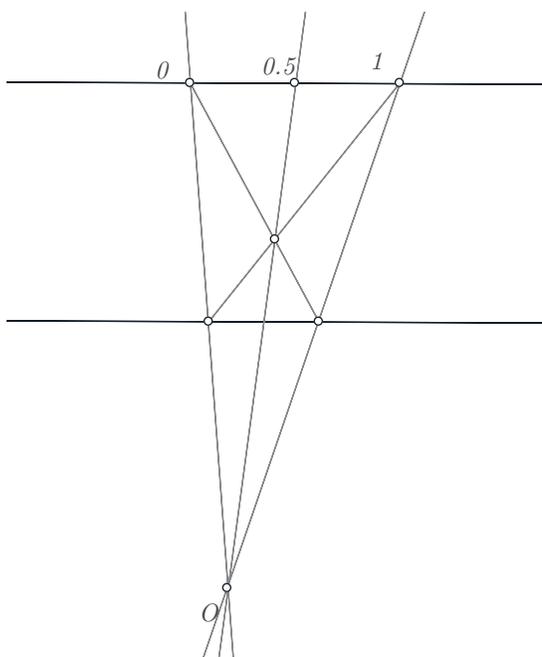
1. Entre dos puntos de una recta hay una infinidad de otros puntos.
2. Cualquier número finito de puntos A, B, \dots, K de una recta siempre se pueden acomodar de manera que B esté entre A y C y entre A y D , ... y entre A y K ; C entre B y D y entre B y E , ... y entre B y K , etc.
3. Toda línea recta a en un plano lo divide en dos regiones tales que: Para cada punto A de una de ellas y otro punto B de la otra, hay un punto entre A y B que está en a . Por otro lado, Dos punto A y A' determinan un segmento AA' que no contiene ningún punto de a .
4. Todo polígono simple cuyos vértices se encuentran en un plano α , divide al plano en dos regiones, una *interior* y otra *exterior* tales que si A es un punto de la primera y B uno de la segunda, cualquier curva poligonal que los una tiene forzosamente un punto en el polígono; por otro lado, dados dos puntos A y A' en la región interior y otros dos puntos B y B' en la exterior, hay curvas poligonales que unen a A con A' y a B con B' que no tienen ningún punto en común con el polígono.
5. Todo plano α divide al espacio en dos regiones tales que todo punto A de una de ella y todo punto B de la otra define un segmento AB que tiene un punto en el plano α ; y por otro lado el segmento definido por un par de puntos A y A' en una de ellas no tiene ningún punto en común con el plano α .

El tercer grupo de axiomas consta de uno solo y es una reformulación del famoso *quinto postulado* de

Los *elementos* de Euclides, sin el cual, como podrá verse en la sección 3.7, es posible construir otras geometrías diferentes de la clásica o euclidea, pero igualmente interesantes y que en principio pueden ser útiles en la ciencia.

AXIOMA DE LAS PARALELAS O DE EUCLIDES

1. Sean r una recta y A un punto en un plano α tales que A no esté en r . Entonces existe una y solo una recta p en α que pasa por A y que no tiene ningún punto en común con r . A la recta p se le llama *la paralela a r en el plano α que pasa por A* , o simplemente, cuando el plano está sobreentendido, *la paralela a r por A* .



La existencia de una paralela es equivalente a la de un punto C , entre otros dos cualesquiera A y B , cuyo armónico conjugado no existe (también se dice que el armónico conjugado de C respecto a AB está en el infinito). El punto C se puede definir como el punto medio del segmento AB .

En esta figura a los puntos A y B se les ha llamado 0 y 1 y al punto C se le ha llamado 0.5. La figura ilustra el hecho de que la existencia de una paralela es equivalente a la de un punto (el 0.5) cuyo armónico conjugado no existe o está en el infinito.

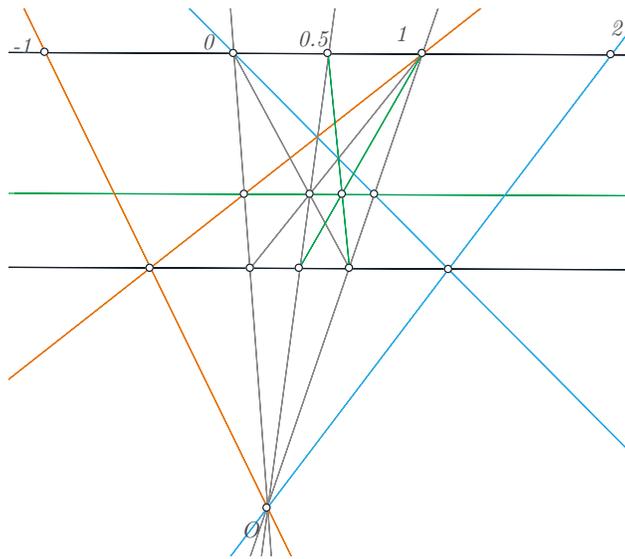
El punto O se ha construido como la intersección de dos rectas, una que pasa por 0 y por un punto de la paralela y otra que pasa por 1 y por otro punto de la paralela. En cierto sentido, el punto 0.5 es el armónico conjugado del punto al infinito.

Continuando con esta construcción se puede llegar otro teorema, mucho más ambicioso.

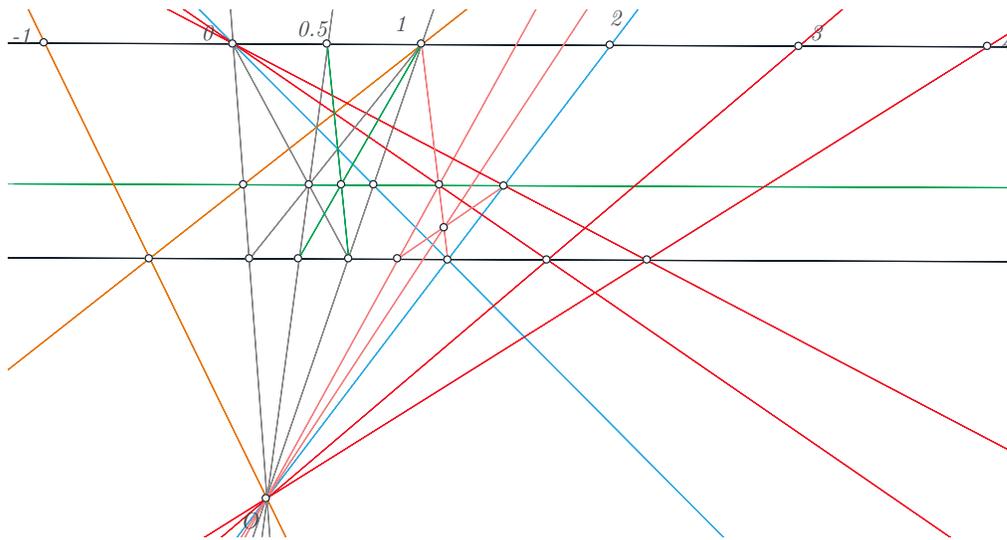
Teorema. *Dados dos puntos 0 y 1 en una recta y una paralela a ella, es posible asignar a cada punto de la primera una representación binaria o decimal única (en general infinita) que lo represente.*

Como veremos, para demostrar este teorema hace falta, además de los axiomas de orden y de las paralelas, probar que cualquier punto es “alcanzable” mediante un número finito de pasos y esto a su vez requiere de alguna versión del axioma de continuidad o de Arquímedes. Sin embargo, lo interesante del resultado es que no requiere de los axiomas de congruencia, que sólo hacen falta cuando se quiere comparar segmentos que no están en la misma rectas. La construcción se ilustra en la siguiente sucesión de figuras.

Después de haber construido el 0.5 como se indicó anteriormente, se construye el 2 como se muestra con las rectas azules, y el -1 como se hace con las rectas color naranja. Para construir más enteros se repiten las operaciones descritas anteriormente. Por ejemplo con las rectas rojas se obtienen los puntos 3 y 4. Así se puede seguir indefinidamente hasta rebasar el punto que queremos representar (aquí se usa el axioma de Arquímedes que se presenta en siguiente sección). Para las fracciones que corresponden a un punto entre dos enteros, se procede como se hizo para 0.5, partiendo el segmento que lo contiene en dos. Encontramos el punto medio de ese segmento y nuestro punto quedará en uno de dos nuevos segmentos.



Las rectas color de rosa muestran cómo se pueden encontrar los puntos 1.5 y 1.75. Procedemos así por inducción hasta que nuestro punto sea uno de los puntos medios construidos o indefinidamente si esto nunca sucede.³



Los axiomas de congruencia permiten definir el concepto de *rayo*, como la parte de una recta que queda *a un lado* de uno de sus puntos y se dice que el rayo *emana* de ese punto. A su vez el concepto de rayo permite definir el de *ángulo* como una pareja (no ordenada) de dos rayos h y k que *emanan de un mismo punto* O , el cual se denota por $\angle(h, k)$.

³Que esto produzca representaciones diferentes para puntos distintos también depende del axioma de Arquímedes.

3.9. Los axiomas de congruencia y el de Arquímedes

Los axiomas de congruencia

Los axiomas de congruencia especifican la idea de *desplazamiento* o transformación rígida (como traslación, rotación y reflexión) y la propiedad que estas transformaciones conservan que se identifica con la *congruencia*.

AXIOMAS DE CONGRUENCIA

1. Si A y B son puntos de una recta a , y si A' es un punto de la misma o de otra recta a' , entonces de un lado de A' y sobre la recta a' hay un punto B' tal que el segmento AB es congruente con el segmento $A'B'$. Esta relación se escribe en símbolos como $AB \equiv A'B'$. Por supuesto un segmento es congruente con si mismo: $AB \equiv AB$.
2. Si $AB \equiv A'B'$ y $A'B' \equiv A''B''$ entonces $AB \equiv A''B''$ (transitividad de la congruencia)
3. Si AB y BC son dos segmentos de una recta a que no tienen ningún punto en común, excepto B y $AB \equiv BC$, y si $A'B'$ y $B'C'$ son dos segmentos de una recta a' que no tienen ningún punto en común, excepto B' y $A'B' \equiv B'C'$, entonces $AC \equiv A'C'$.
4. Sea $\angle(h, k)$ un ángulo en un plano α y sea h' un rayo en otro plano α' . Entonces existe uno y solo un rayo k' en α' que emana del mismo punto que h' tal que $\angle(h, k)$ es congruente con $\angle(h', k')$, lo cual se denota por $\angle(h, k) \equiv \angle(h', k')$. Todo ángulo es congruente consigo mismo, es decir, $\angle(h, k) \equiv \angle(h, k)$.
5. Si $\angle(h, k) \equiv \angle(h', k')$ y $\angle(h', k') \equiv \angle(h'', k'')$ entonces $\angle(h, k) \equiv \angle(h'', k'')$.
6. Si en dos triángulos ABC y $A'B'C'$ se dan las congruencias⁴:

$$AB \equiv A'B'; BC \equiv B'C'; \angle BAC \equiv \angle B'A'C'$$

entonces también se cumplen estas otras congruencias:

$$\angle ABC \equiv \angle A'B'C'; \angle ACB \equiv \angle A'C'B'$$

A partir de estos axiomas se pueden demostrar varias propiedades básicas como que la congruencia tanto de segmentos como de ángulos es *reflexiva* (Si $AB \equiv A'B'$ entonces $A'B' \equiv AB$ y si $\angle(h, k) \equiv \angle(h', k')$ entonces $\angle(h', k') \equiv \angle(h, k)$) y se pueden definir los conceptos de *ángulos complementarios* y de *ángulo recto*, que es aquel que es congruente con su ángulo complementario. En particular el que todos los ángulos rectos son congruentes es un teorema, aunque para Euclides éste era un postulado.

A partir de estos axiomas, especialmente del 6º, se pueden deducir los criterios conocidos de congruencia de triángulos.

Por supuesto, los axiomas de congruencia, en especial el 3º y el 5º sirven para definir la suma de segmentos colineales y la de ángulos coplanares y para demostrar el teorema de que los ángulos alternos internos entre paralelas son iguales y el de que los ángulos internos de un triángulo suman dos rectos.

Las demostraciones de estos teoremas se dejan como ejercicios para el lector. En caso de dudas sugerimos consultar el libro de HILBERT *Los fundamentos de la geometría*.

⁴ $\angle BAC$ se define como $\angle(h, k)$ donde h y k son los rayos que emanan de A y el primero está formado por la parte de $A \vee B$ que está del lado de B y el segundo por la parte de $A \vee C$ que está del lado de C

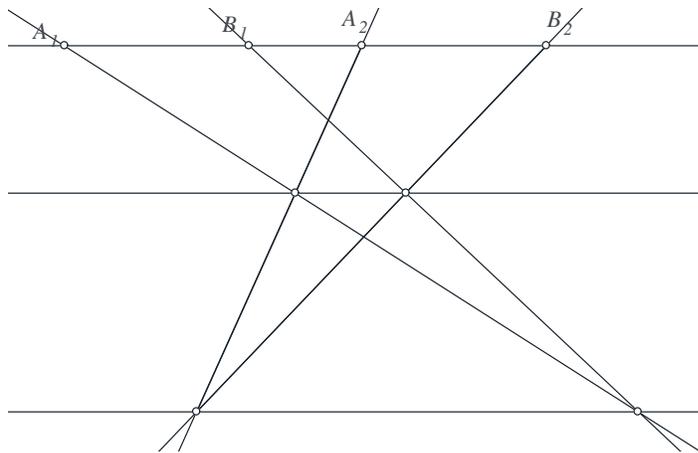
El axioma de continuidad o de Arquímedes

Para enunciar ese axioma, que es el que permite introducir en la geometría la idea de continuidad, es necesario definir el concepto de “*igualdad*” de dos segmentos de una recta. Hay dos opciones, definir la igualdad como *congruencia*, lo cual requiere la adopción del cuarto grupo de axiomas, o bien definirla utilizando alguna construcción que se apoye únicamente en los dos o tres primeros grupos de axiomas.

AXIOMA DE CONTINUIDAD O DE ARQUÍMEDES

1. Sea A_1 un punto arbitrario de una recta entre dos puntos A y B de ella misma. Sean A_2, A_3, \dots los puntos de la recta tales que A_1 está entre A y A_2 y AA_1 es igual a A_1A_2 , A_2 está entre A_1 y A_3 y AA_1 es igual a A_2A_3 , etc. Entonces hay algún punto A_n tal que B está entre A y A_n .

Para completar la identificación de los puntos de una recta con las expansiones decimales o binarias como se describió en la sección 3.8, claramente conviene adoptar la segunda de las opciones arriba mencionadas para definir la *igualdad* entre segmentos. No haremos aquí tal construcción sino que la dejaremos al lector como un *ejercicio avanzado*. La siguiente figura sugiere una posible definición para la *igualdad de dos segmentos* $A_1B_1 = A_2B_2$ basada en el axioma de las paralelas.



Este trabajo equivale a una *construcción de los números reales*. Como es bien sabido, cualquier construcción de los números reales requiere del axioma de continuidad o de Arquímedes.

Cabe mencionar algo más en este punto y es que al definir las operaciones entre segmentos, ya sea que la igualdad entre ellos se defina como congruencia o con una construcción como la que sugerimos en el párrafo anterior, la conmutatividad del producto requiere del teorema de Pappus. El hecho de que éste pueda demostrarse a partir de los axiomas de conexión e incidencia permite, en principio, realizar la construcción de los números reales de manera geométrica y sin necesidad de utilizar los axiomas de congruencia.

Capítulo 4

La geometría clásica

Llamamos Geometría clásica a todo el cuerpo de conocimientos geométricos que se deducen racionalmente a partir, ya sea de los postulados de Euclides o de los axiomas de Hilbert. Con ambos enfoques encontramos los mismos teoremas y las mismas construcciones con regla y compás. Incluimos en lo clásico tanto la geometría plana como la del espacio. La geometría deja de ser clásica cuando utiliza coordenadas y álgebra y se convierte en Geometría analítica, o cuando se aplica a superficies curvas y se convierte en Geometría diferencial. Cuando se incorporan las rectas y puntos al infinito la geometría se convierte *proyectiva*, pero se sigue considerando como clásica. La negación del 5º postulado de Euclides y el desarrollo de la *geometría hiperbólica*, puede incluirse como clásico pues no requiere de nuevas herramientas matemáticas, sino que procede exactamente con las mismas, de hecho con algo menos. Sin embargo muchos autores no la incluyen en lo clásico.

La mayor parte de la Geometría clásica que conocemos, es decir los teoremas y las construcciones con regla y compás, ya se conocían en la época de Euclides. Por eso su libro sigue siendo tan valioso. Sin embargo a lo largo de los siguientes 2000 años se han ido encontrando nuevos resultados. Por ejemplo se pueden mencionar los teoremas de *Pappus* y *Menelao* (siglo IV), la recta de Euler (siglo XVIII), la circunferencia de los 9 puntos, las esferas de Dandelin y la construcción del heptadecágono regular (siglo XIX), al igual que Los fundamentos de la Geometría de Hilbert. Incluso fue hasta el siglo XX que se demostró que no es posible partir cualquier poliedro en otros de manera que puedan ensamblarse en un cubo (1903). En este capítulo presentamos varios de estos temas que se engloban dentro de lo que se llama Geometría clásica, algunos antiguos, algunos más recientes).

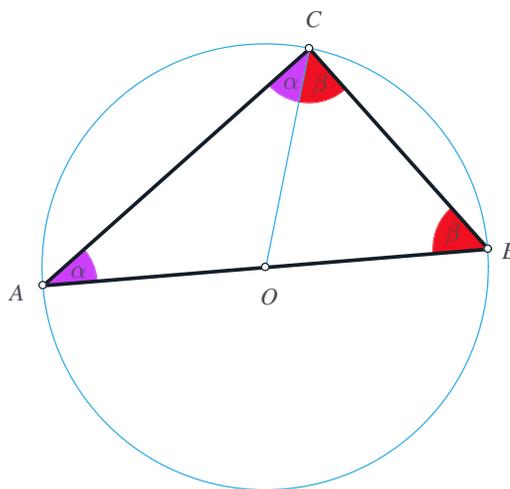
Es importante mencionar que la presentación se beneficia de la *geometría dinámica*, tecnología surgida a finales del siglo XX y que hoy en día es muy eficaz para reactivar el interés y la comprensión de la geometría clásica. Si el lector visualiza este texto en formato *pdf* desde una computadora o un dispositivo móvil conectados a internet, mediante un clic sobre muchas de las imágenes de este capítulo (y algunas del anterior), podrá acceder a las versiones interactivas con las que se prepararon. No solo eso, sino que podrá ensayar sus propias construcciones y redescubrir algunos de los resultados, encontrar nuevas maneras de construir o demostrar algo y, ¿por qué no?, descubrir algunos resultados desconocidos.

4.1. Propiedades del círculo

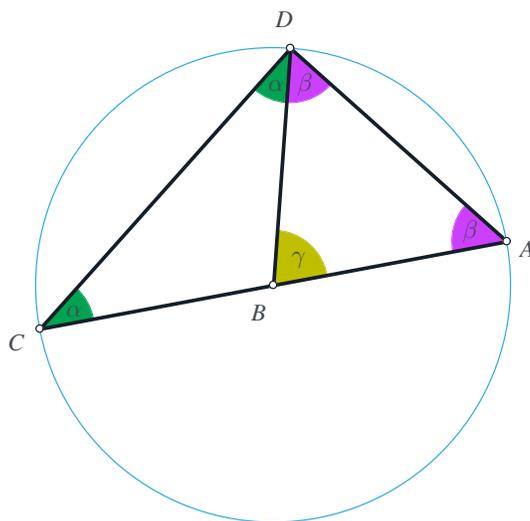
En esta sección demostraremos varios teoremas sobre círculos y ángulos.

Teorema. Si un triángulo tiene un lado en un diámetro de una circunferencia y el vértice opuesto a dicho lado sobre la circunferencia, el ángulo en dicho vértice será siempre recto.

Demostración. Consideremos un triángulo inscrito en una circunferencia y con uno de los lados (AB) como un diámetro de la circunferencia. Sea O el centro de la circunferencia. Trazamos el segmento OC , con lo que se forman los triángulos AOC y COB . AOC es isósceles pues AO y CO , al ser radios de la circunferencia, son iguales. Lo mismo se puede decir del triángulo COB . Como el triángulo AOC es isósceles, $\sphericalangle OAC = \sphericalangle OCA$. Llamemos a dicho ángulo α . Igualmente a $\sphericalangle OCB = \sphericalangle OBC$, y lo llamaremos β . La suma de ángulos internos de un triángulo es 180° . Para nuestro triángulo, $\alpha + \alpha + \beta + \beta = 180^\circ$. Es decir, dos ángulos α más dos ángulos β dan un ángulo llano, por lo que un ángulo α más un β debe dar un ángulo recto. Q.E.D.

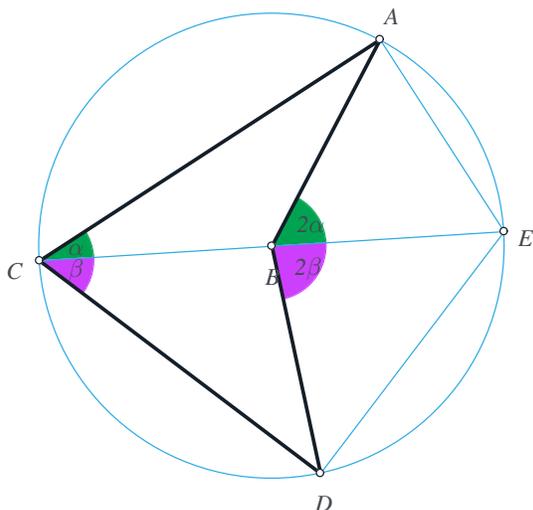


Teorema. El ángulo central subtendido por un arco desde un punto es el doble que el central subtendido por el mismo arco. Caso 1: el punto C y uno de los extremos A del arco AD son diametralmente opuestos.



Demostración. Tenemos un ángulo inscrito a una circunferencia $\alpha = \sphericalangle ACD$ y el ángulo central correspondiente es $\gamma = \sphericalangle ABD$. El triángulo BDC es isósceles pues BC y BD son iguales al ser ambos radios de la circunferencia. Así, los ángulos de la base de dicho triángulo son ambos α . Entonces, por un lado, el complemento de $\sphericalangle DBC$ es γ , pero como la suma de los ángulos internos del triángulo BDC es un ángulo llano, el complemento de $\sphericalangle DBC$ también es igual a 2α . Por tanto $\gamma = 2\alpha$. Q.E.D.

Teorema. El ángulo central subtendido por un arco desde un punto es el doble que el inscrito y subtendido por el mismo arco. Caso 2: el ángulo desde el punto incluye al centro. (En este caso buscamos la relación entre el ángulo $\sphericalangle ABD$ y el $\sphericalangle ACD$ cuando el diámetro de la circunferencia que pasa por C queda entre AC y DC .)

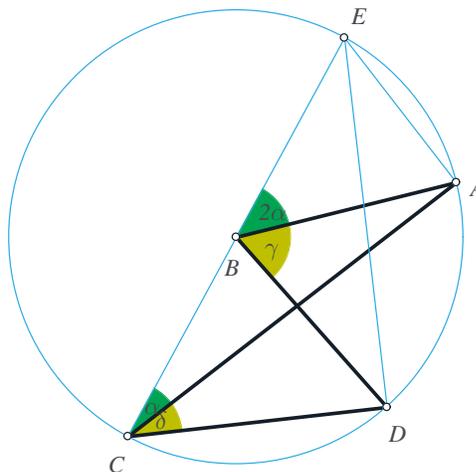


Demostración. Trazamos el diámetro en cuestión. Llamamos E al punto diametralmente opuesto a C . Trazamos el segmento auxiliar AE . Notamos que los triángulos $\sphericalangle ABC$ y $\sphericalangle ABE$ corresponden al caso inicial para el que ya se demostró que el ángulo central es el doble que el inscrito. Marcamos los ángulos central e inscrito como 2α y α , respectivamente. Podemos hacer lo mismo agregando el segmento auxiliar DE . Así, los triángulos DBC y DBE se comportan igual. Marcamos el ángulo central y el inscrito ahora como 2α y β , respectivamente. Así, el nuevo ángulo central $\sphericalangle ABD$ es $2\alpha + 2\beta$, y el nuevo ángulo inscrito $\sphericalangle ACD$ es $\alpha + \beta$, por lo que se sigue respetando que el ángulo central subtendido por un arco (en este caso el AD) es el doble que el inscrito subtendido por el mismo arco. Q.E.D.

Teorema. El ángulo central subtendido por un arco desde un punto es el doble que el inscrito y subtendido por el mismo arco. Caso 3: el ángulo desde el punto no incluye al centro. (En este caso el diámetro queda fuera del ángulo inscrito y del central.)

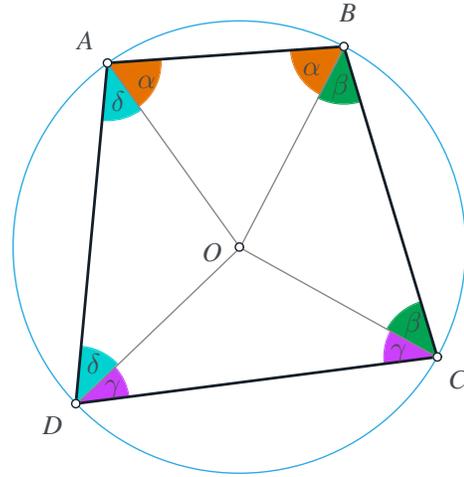
Demostración. Trazamos primero el diámetro que pasa por C y llamamos E al punto diametralmente opuesto a C . Trazamos el segmento AE auxiliar y notamos que los triángulos ABE y ACE son el central e inscrito para el caso inicial (cuando CE forma un diámetro). Así pues, podemos marcar el central como 2α y el inscrito como α . Igualmente, si trazamos el segmento auxiliar DE , los triángulos DBE y DCE caen en el mismo caso, y $\sphericalangle DBE = 2\sphericalangle DCE$. No los marcamos para no ensuciar la figura, pero denominémoslos 2β y β , respectivamente. Sea γ el ángulo central $\sphericalangle ABD$ subtendido por el arco AD . Notamos que $2\alpha + \gamma = 2\beta$. Sea δ el ángulo inscrito $\sphericalangle ACD$ subtendido por AD . Notamos que $\alpha + \delta = \beta$. Así, $\gamma = 2\beta - 2\alpha$ y $\delta = \beta - \alpha$. De ahí que $\gamma = 2\delta$. Q.E.D.

Esto completa el teorema para todos los casos.

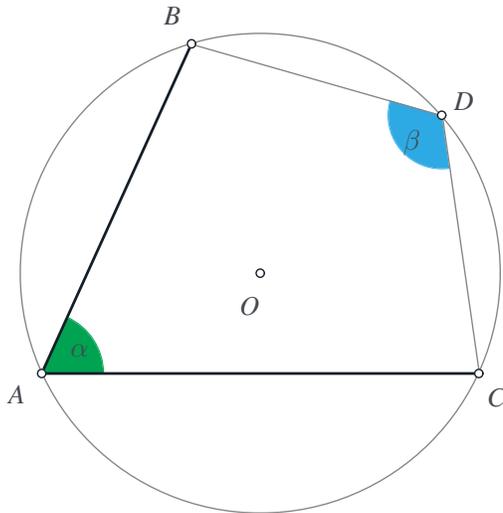


Teorema. Los ángulos opuestos de un cuadrilátero cíclico suman dos rectos.
(Cuadrilátero cíclico aquel en el que todos sus vértices están en una circunferencia.)

Demostración. Unimos los puntos con el centro O del círculo, con lo que formamos cuatro triángulos internos que comparten el vértice O . Observamos el triángulo AOB . Notamos que es isósceles ya que dos de sus lados son iguales (los que corresponden a radios de la circunferencia). Por tanto los ángulos entre esos lados y el tercero son iguales. Los llamamos α y los trazamos del mismo color. Análogamente, los otros triángulos comparten el vértice O y también son isósceles. Y, por lo mismo, sus ángulos correspondientes son iguales. Se trazan ahora todos esos ángulos. Sabemos que la suma de ángulos internos de un cuadrilátero es 360° . Así que $2\alpha + 2\beta + 2\gamma + 2\delta = 360^\circ$. De donde $\alpha + \beta + \gamma + \delta = 180^\circ$. Q.E.D.



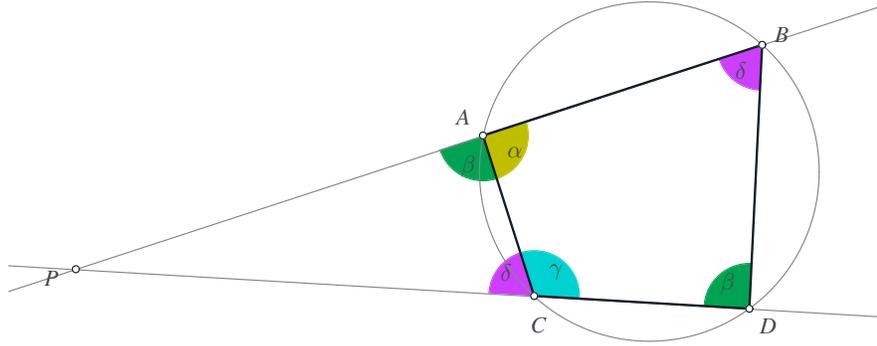
Teorema. Un ángulo inscrito en una circunferencia siempre es igual independientemente de cómo se mueva el punto que lo forma, siempre y cuando se mueva de un mismo lado de los puntos que flanquean los segmentos. Cuando rebasa esa cota su valor cambia a 180° menos el otro.



Consideremos 3 puntos A , B y C en una circunferencia. Trazamos los segmentos AB y AC y llamamos α al ángulo entre ellos. Agregamos un punto D sobre la circunferencia pero opuesto a A . Este punto se puede generar como la intersección de la mediatriz de B y C con la circunferencia, pero que es opuesta a A . Por el momento, no se muestra dicha construcción, sino sólo el punto D . Trazamos los segmentos BD y CD para completar el cuadrilátero cíclico, y trazamos el ángulo β . Por la propiedad de los cuadriláteros cíclicos, $\alpha + \beta = 180^\circ$. Si movemos A del mismo lado de B y C , D se mantiene fijo y β es constante, por lo que α también lo es. Q.E.D.

Nótese que esta propiedad también pudo haber sido demostrada mediante la propiedad del ángulo central e inscrito.

Teorema. Sea P un punto que no está en una circunferencia dada. Trácese dos rectas que pasan por P y que tocan a la circunferencia, la primera en los puntos A y B y la segunda en los puntos C y D . Entonces $PA \cdot PB = PC \cdot PD$. Si la segunda recta es tangente a la circunferencia y el punto de tangencia es C , entonces $PA \cdot PB = PC^2$. El producto constante $PA \cdot PB$ se llama *la potencia del punto P respecto a la circunferencia*.

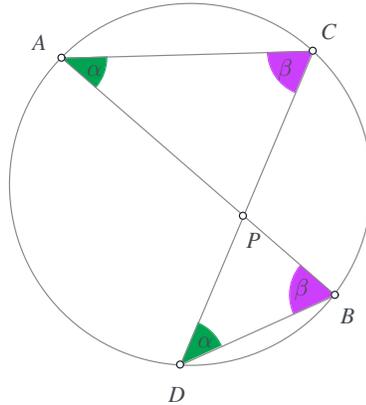


Demostración para el caso en que el punto P está fuera de la circunferencia.

Sean $\alpha, \beta, \gamma, \delta$ como en la figura anterior. Por la propiedad de los cuadriláteros cíclicos sabemos que $\alpha + \beta = 180^\circ$. Por tanto el ángulo suplementario a α es igual a β y lo pintamos del mismo color. Por el mismo argumento, el ángulo δ es el suplementario de γ . Los triángulos PBD y PCA son semejantes pues comparten dos ángulos internos. Entonces, $\frac{PB}{PC} = \frac{PD}{PA}$. Pero sabemos que esta igualdad es equivalente a $PA \cdot PB = PC \cdot PD$, que es lo que se quería demostrar.

Demostración para el caso en que el punto P está dentro de la circunferencia.

Sean α y β los ángulos de la siguiente.



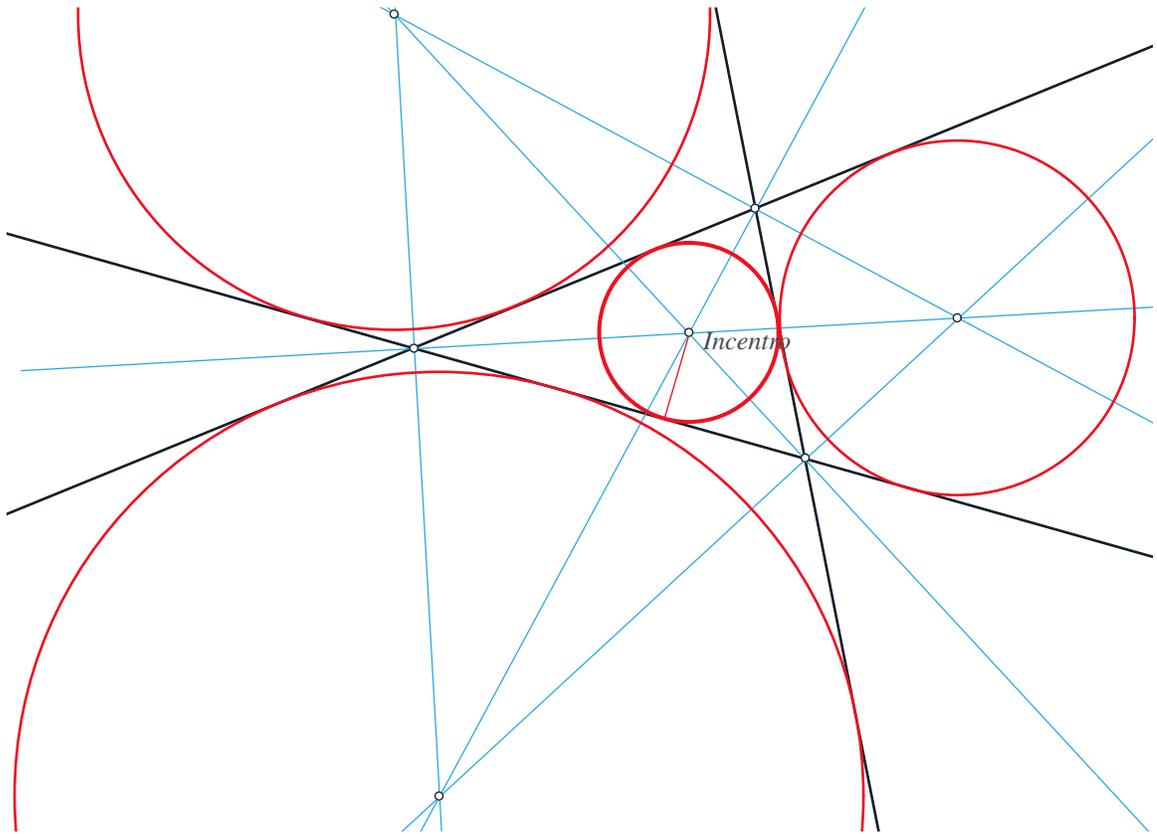
Resulta evidente que los triángulos ACP y DBP son semejantes, por lo tanto $\frac{PA}{PD} = \frac{PC}{PB}$, que es equivalente a $PA \cdot PB = PC \cdot PD$. Q.E.D.

4.2. Propiedades del triángulo

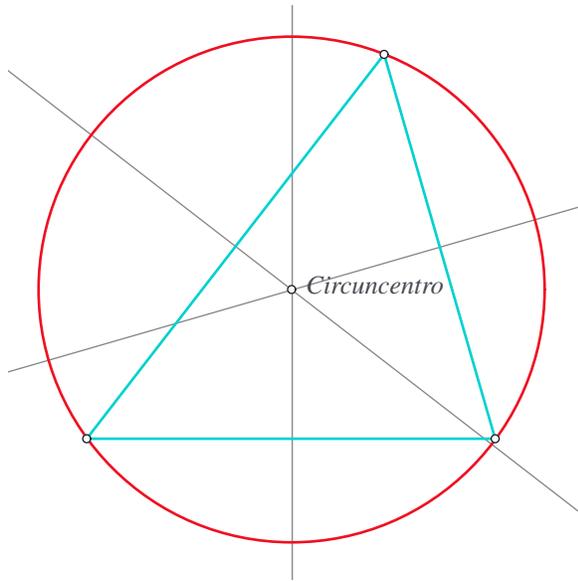
En esta sección enunciamos varios resultados relativos al triángulo. Demostraremos algunos de ellos y daremos alguna idea de cómo demostrar los otros. En estos últimos casos dejamos los detalles de las demostraciones como ejercicios para el lector.

1. Construcción de la circunferencia *inscrita* en un triángulo, es decir la que está dentro del triángulo y es tangente a sus lados; y de las circunferencias *excríticas*, que son las que son tangentes a las prolongaciones de los lados y quedan fuera del triángulo.

Los centros de estas circunferencias son los puntos en los que coinciden las bisectrices o sus perpendiculares (azules). Los radios son las distancias de los centros a las rectas dadas (ver puntos color naranja que son las proyecciones ortogonales de los centros a las rectas dadas).

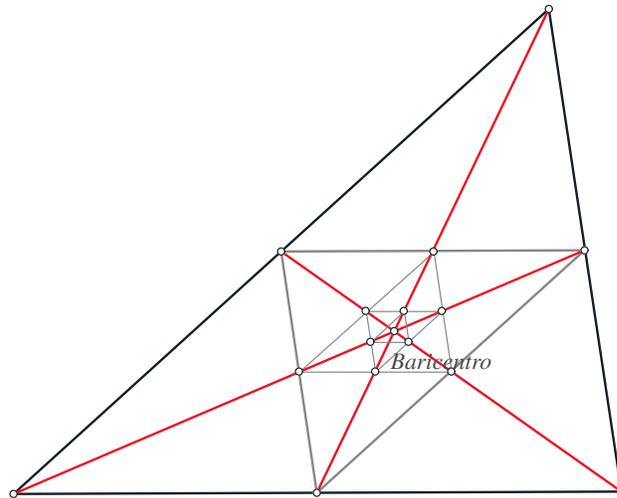


El incentro es el punto donde coinciden las bisectrices y es también el centro del círculo inscrito en un triángulo

2. Construcción del *circuncírculo*¹

El punto en el que coinciden dos de las tres mediatrices (la mediatriz de dos puntos es la recta de los puntos que equidistan de ellos) de un triángulo está a la misma distancia de los tres vértices y por tanto la otra mediatriz pasa por él. Por tanto la circunferencia con centro en ese punto, que llama *circuncentro*, pasa por los tres vértices del triángulo y es la circunferencia circunscrita al triángulo.

3. El baricentro (es decir, el punto donde coinciden las medianas²) es el punto al que convergen los triángulos anidados que se forman tomando como vértices los puntos medios de los lados del anterior. La demostración de que las medianas concurren en un punto se deja como ejercicio.



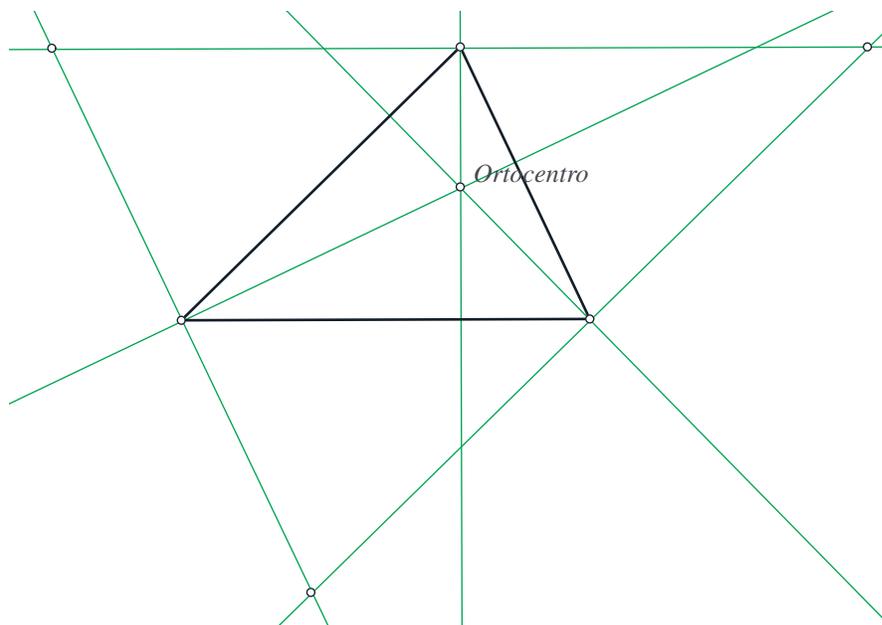
Las medianas de un triángulo coinciden en el baricentro

¹El *circuncírculo* es la circunferencia circunscrita a un triángulo, es decir, la que pasa sus tres vértices.

²las *medianas* son las rectas que pasan por los vértices de un triángulo y por los puntos medios de sus lados opuestos.

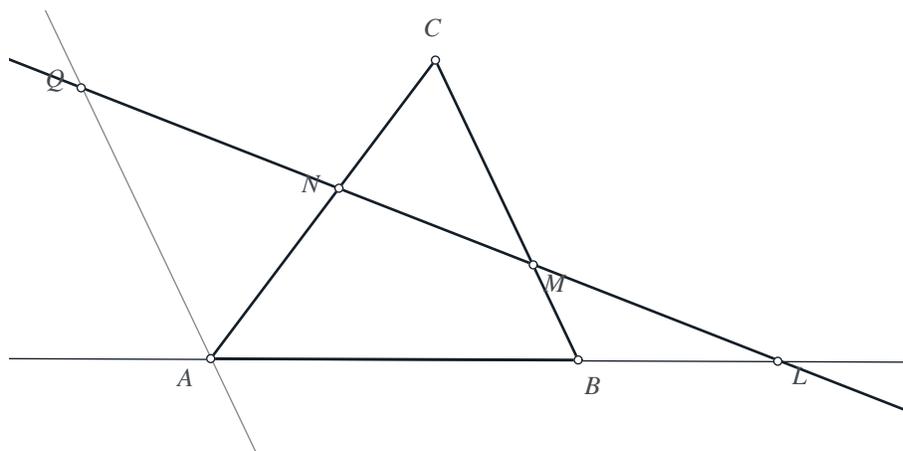
4. Las alturas de un triángulo coinciden en un punto que se llama el ortocentro del triángulo.

Demostración. Las alturas de un triángulo son las mediatrices del triángulo formado por las paralelas a cada uno de sus lados por los vértices opuestos a ellos. Por tanto son concurrentes. Q.E.D.



5. *Teorema de Menelao.* Si una recta corta a los lados AB , BC y CA del triángulo ABC en L , M y N , respectivamente, entonces³

$$\frac{AL}{LB} \cdot \frac{BM}{MC} \cdot \frac{CN}{NA} = -1$$



Demostración. Sea Q la intersección de la recta con la paralela a BC por A . Entonces, por semejanza

³Los segmentos se consideran con signo positivo si van en sentido contrario a las manecillas del reloj, y negativo si van en el sentido de las manecillas del reloj. En el dibujo todos son positivos excepto LB , que es negativo.

de triángulos:

$$\frac{AL}{AQ} = \frac{BL}{BM}; \frac{AQ}{CM} = \frac{NA}{NC}$$

que equivalen a $\frac{AL}{AQ} \cdot \frac{BM}{BL} = 1$ y $\frac{AQ}{CM} \cdot \frac{NC}{NA} = 1$. Multiplicando ambas razones obtenemos:

$$\frac{AL}{AQ} \cdot \frac{BM}{BL} \cdot \frac{AQ}{CM} \cdot \frac{NC}{NA}$$

. Simplificando:

$$\frac{AL}{BL} \cdot \frac{BM}{CM} \cdot \frac{NC}{NA} = 1$$

y

$$\frac{AL}{LB} \cdot \frac{BM}{MC} \cdot \frac{CN}{NA} = -1$$

QED.

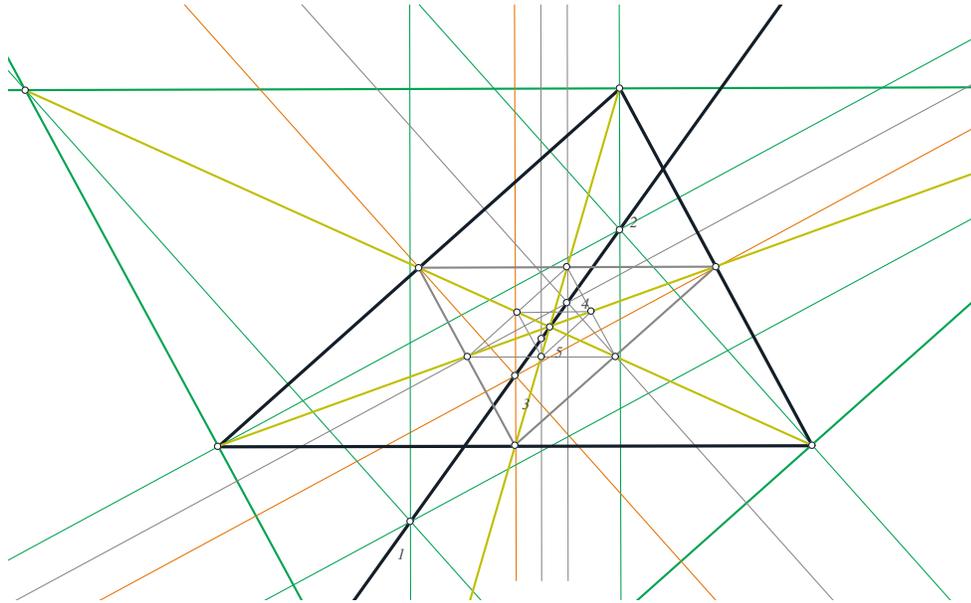
6. *Teorema de Ceva.* Si tres rectas que pasan por los vértices A, B, C de un triángulo cortan a los lados opuestos en los puntos L, M, N entonces

$$\frac{AL}{LB} \cdot \frac{BM}{MC} \cdot \frac{CN}{NA} = 1$$

. La demostración de este teorema puede hacerse a partir del teorema de Menelao, cambiando uno de los puntos L, M o N por su armónico conjugado respecto a los vértices con los que es colineal. Se dejan los detalles como ejercicio.

7. La recta de Euler. El circuncentro, el ortocentro y el baricentro de un triángulo están alineados. La recta que los une se llama la recta de Euler.

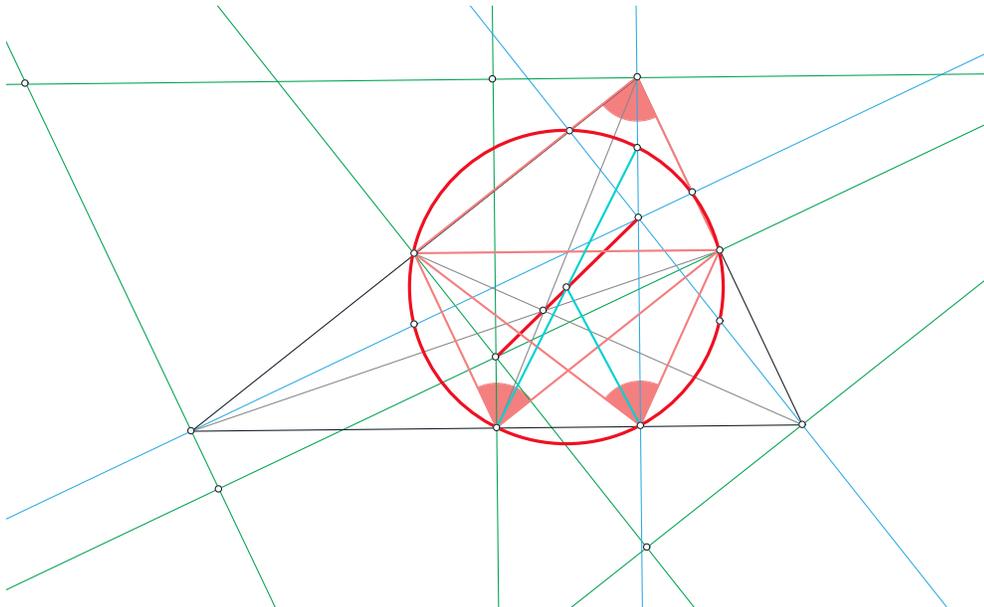
Demostración. Los ortocentros y los circuncentros de todos los triángulos anidados que se van formando tomando como vértices los puntos medios de los lados del anterior, están todos alineados. Por tanto el baricentro, que es el límite de todos ellos también debe estar sobre la misma recta.



La recta de Euler contiene al baricentro, al ortocentro y al circuncentro de todos los triángulos anidados.

8. El círculo de los 9 puntos. El llamado *círculo de los 9 puntos* es uno que pasa por los puntos medios de los lados, por los pies de las alturas y por los puntos medios entre los vértices y el ortocentro.

La recta de Euler (roja) une el circuncentro (verde) con el ortocentro (azul). En ella yace también el baricentro (gris). El punto medio entre el circuncentro y el ortocentro es centro de un círculo de los 9 puntos. La demostración se basa en que los tres ángulos marcados en la figura son iguales.



4.3. Construcciones con regla y compás

La geometría clásica o euclidiana intentaba realizar todas sus construcciones con regla y compás. Los antiguos griegos creían que todas las construcciones geométricas deberían realizarse exclusivamente con esos medios. Sin embargo hubo tres construcciones geométricas que, por más que lo intentaron, nunca pudieron resolver con regla y compás. Fueron éstas:

PROBLEMAS QUE NO SE PUEDEN RESOLVER CON SOLO REGLA Y COMPÁS

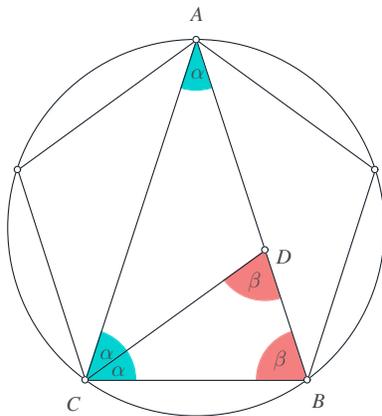
1. *Cuadratura del círculo*: Construir un cuadrado con la misma área de un círculo dado.
2. *Trisección de un ángulo*: Dado cualquier ángulo otro que sea su tercera parte.
3. *Duplicación del cubo*: Dado (el lado de) un cubo, construir (el lado de) otro cuyo volumen sea el doble que el del primero.

En el siglo XVIII se demostró que las tres son imposibles de realizar con sólo regla y compás. Esto llevó a desarrollos de gran trascendencia para el *Álgebra moderna*.

Un ejemplo interesante de construcción con regla y compás que no es precisamente trivial, pero que sí se puede hacer, es la del pentágono regular. Euclides realiza la construcción en el libro IV, Proposición 11, así que no utiliza la semejanza de triángulos, desarrollada hasta el libro VI, para justificarla. Sin embargo, dado que el pentágono regular está íntimamente relacionado con la *razón áurea*, resulta enriquecedor realizar una construcción en la que ésta aparezca de manera explícita, como la que exhibimos en esta sección. Antes definiremos y estudiaremos brevemente la *razón áurea*.

La razón áurea

Hay varias maneras de definir la razón áurea γ . Aquí lo haremos como *la razón entre la diagonal y el lado de un pentágono regular*. De acuerdo con la figura, $\gamma = \frac{AB}{BC}$.



Demostremos que γ satisface la ecuación

$$\gamma^2 = 1 + \gamma$$

De las propiedades de los ángulos inscritos en una circunferencia se deduce que $\angle ABC = \angle BCA = 2\angle CAB$. Sea D la intersección de la bisectriz de $\angle BCA$ con AB . Es fácil ver que $\angle CDB = \angle DBC = \beta$ y $\angle CAD = \angle DCA = \alpha$. Por tanto los triángulos $\triangle ABC$ y $\triangle CDB$ son semejantes y $CD = DA$, $\frac{CD}{DB} = \frac{AB}{BC} = \gamma$ y $CD + DB = AB$. Dividiendo la segunda igualdad por DB obtenemos

$$\gamma + 1 = \frac{AB}{DB} = \frac{AB}{BC} \frac{BC}{DB} = \gamma \frac{CD}{DB} = \gamma^2$$

Es fácil ver que la única solución positiva de esta ecuación cuadrática es $\gamma = \frac{1+\sqrt{5}}{2}$. A veces conviene expresar la ecuación para γ en la forma: $\gamma = 1 + \frac{1}{\gamma}$.

Otra definición de la razón áurea es con tres puntos alineados A, B, D que satisfacen: $AB = AD + DB$ y $\frac{AB}{AD} = \frac{AD}{DB}$. Euclides define esta relación diciendo que AB es *cortado en extrema y media razón* por el punto D . Es decir, que el segmento completo AB es cortado por D en dos partes de manera que la razón del segmento total AB a la parte mayor AD es igual a la razón de la parte mayor AD a la menor DB . (Ver *Los elementos*, Definición VI.3.)

Construcción del pentágono regular inscrito a una circunferencia

La construcción del pentágono regular es una de las hazañas importantes de las matemáticas griegas. Aparece como Proposición 11 del libro IV de *Los elementos* de EUCLIDES, antes del desarrollo de la *Teoría de las proporciones*, lo que quiere decir que EUCLIDES no utiliza en ella el concepto de razón. Sin embargo, dado que el pentágono regular está íntimamente relacionado con la razón áurea, resulta enriquecedor realizar una construcción en la que ésta aparezca de manera explícita.

Consideremos un círculo con centro en O . Sin pérdida de generalidad podemos suponer que su radio es 1. Sean P y Q dos puntos de la circunferencia tales que OP y OQ son perpendiculares. Sea M el punto medio de OQ . Sea R la intersección de la bisectriz de $\angle OMP$ con OP . Afirmamos que las intersecciones S y T de la paralela a OQ por R , junto con P constituyen tres vértices consecutivos del pentágono regular que queremos construir. Los otros dos vértices pueden hallarse como intersecciones de las mediatrices de PS y PT con la circunferencia.

Demostración. Por VI.3. de Los elementos, $\frac{RP}{OR} = \frac{MP}{MO}$ y por el Teorema de Pitágoras aplicado a $\triangle MOP$,

$$\frac{RP}{OR} = \frac{\sqrt{1 + 1/4}}{\frac{1}{2}} = \frac{\sqrt{\frac{5}{4}}}{\frac{1}{2}} = \frac{\frac{\sqrt{5}}{2}}{\frac{1}{2}} = \sqrt{5}$$

Pero $RP = 1 - OR$ y $1 - OR = \sqrt{5}OR$. Por lo tanto

$$1 = (1 + \sqrt{5})OR; OR = \frac{1}{2\gamma}$$

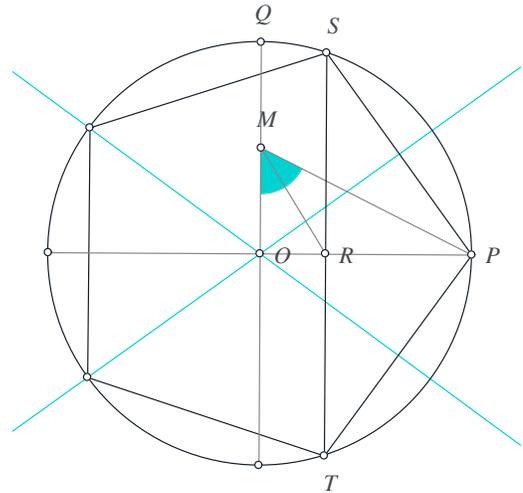
Usando el teorema de Pitágoras en $\triangle PRS$, obtenemos $PS^2 = RS^2 + RP^2$, y poniendo todo en términos de OR ,

$$PS^2 = (1 - OR^2) + (1 - OR)^2 = 2(1 - OR)$$

Calculemos ahora el cuadrado de la razón $\frac{ST}{PS}$:

$$\frac{ST^2}{PS^2} = \frac{4RS^2}{PS^2} = \frac{2(1 - OR^2)}{(1 - OR)} = 2(1 + OR) = 2 + \frac{1}{\gamma} = 1 + \left(1 + \frac{1}{\gamma}\right) = 1 + \gamma = \gamma^2$$

(ya que $1 + \frac{1}{\gamma} = \gamma$ y $1 + \gamma = \gamma^2$). Por tanto, $\frac{ST}{PS} = \gamma$. Esto demuestra que ST es la diagonal y PS el lado del polígono regular inscrito en la circunferencia. Q.E.D.



Construcción del heptadecágono regular inscrito a una circunferencia

A continuación presentamos una imagen con la construcción con regla y compás del heptadecágono (polígono de 17 lados) regular realizada por Carl Friederich Gauss en 1796 a la edad de 19 años. Por algo le llamaban el *Príncipe de la matemáticas*. Fue la primera construcción con regla y compás nueva de un polígono regular que se hacía en 2000 años. Gauss no solo hizo esta construcción sino que determinó cuáles polígonos regulares se pueden construir con regla y compás y cuáles no.

Los únicos que se que se pueden construir con regla y compás son aquellos cuyo número se lados es

$$3, 4, 5, 6, 8, 10, 12, 15, 16, 17, \dots, n, \dots$$

para n de la forma

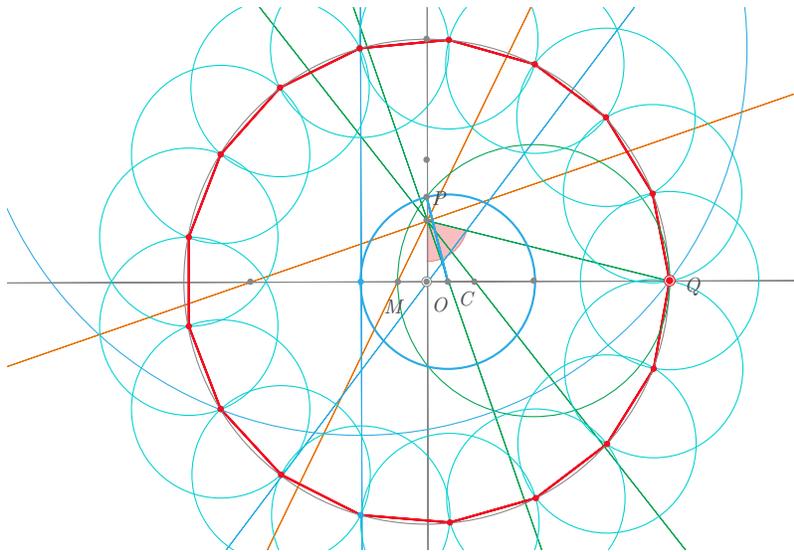
$$n = 2^k p_1 p_2 \dots p_j$$

donde las p_i son *primos de Fermat*, es decir números de la forma $1 + 2^{2^m}$ que son primos. Los primeros primos de Fermat son:

$$5, 17, 257, 65537$$

En particular los números 9, 11, 13 y 14 no tienen la forma requerida, por lo que los polígonos regulares con esos números de lados no se pueden construir con regla y compás.

Brevemente, la construcción del heptadecágono regular procede tomando un punto P a una distancia de $\frac{1}{4}$ del centro O , y otro punto Q que sea un extremo del diámetro perpendicular a OP . Q será uno de los vértices del heptadecágono. Se dividen en 4 el ángulo OPQ y así se define un punto C en ese mismo diámetro.

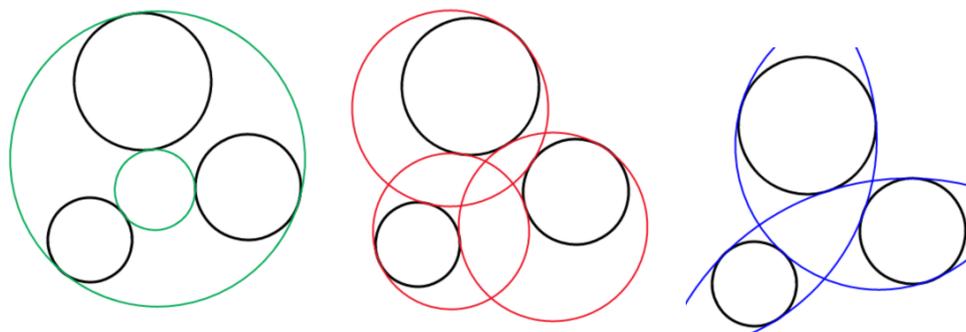


Se toma el punto M , sobre el mismo diámetro, que está a un ángulo de 45° de PC y se traza el círculo con centro en el punto medio entre M y Q y que pasa por M y Q . Se traza otro círculo con centro en el punto C que pase por la intersección de OP con el anterior. La perpendicular a OQ por su intersección con este último círculo cortará al círculo original en dos puntos del heptadecágono deseado. A partir de estos dos y de Q es fácil construir los otros vértices del heptadecágono.

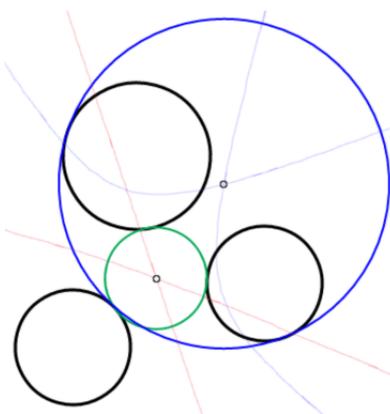
Omitimos la justificación de la construcción, que es bastante complicada.

4.4. Los problemas de Apolonio

Apolonio de Perga (262-190 a.e.c) fue uno de los más grandes matemáticos griegos. Su trabajo más importante fue un tratado sobre las secciones cónicas, parte de cuyo contenido se presenta en 4.5. Otra de sus contribuciones importantes fue un tratado sobre *las tangencias* que, entre otras cosas, contenía la solución del problema de encontrar un círculo tangente a otros tres círculos dados. Lamentablemente este libro se perdió. Sabemos de que el libro existió y tenemos idea de su contenido gracias a los comentarios del matemático alejandrino Pappus del siglo IV. En general hay 8 soluciones posibles, como muestra la siguiente figura en la que los círculos dados son los negros y los tangentes a ellos son 2 verdes, 3 rojos y 3 azules.

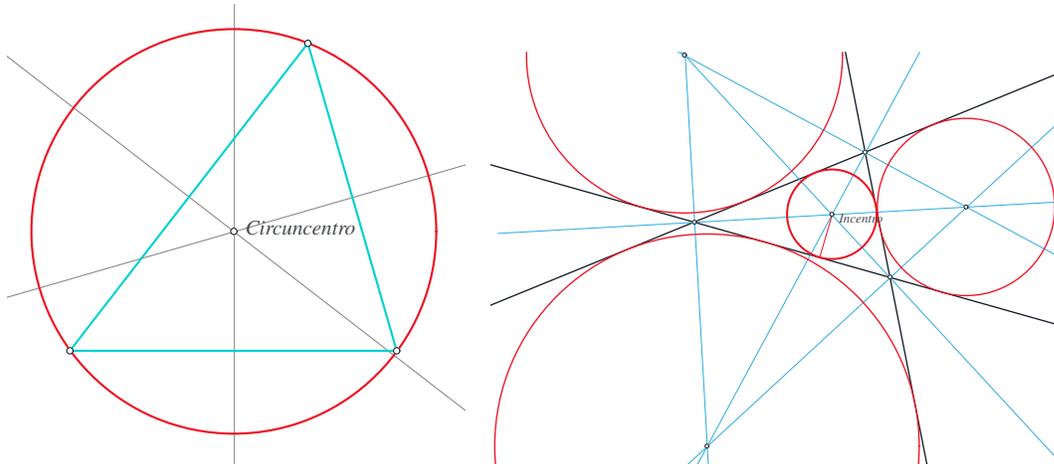


El problema es fácilmente soluble utilizando geometría analítica. Se trata de encontrar las intersecciones de unas hipérbolas definidas como los lugares geométricos de los puntos cuyas distancias a los centros de dos de los círculos, más o menos sus radios, son iguales. Las intersecciones de esas hipérbolas son los centros de las circunferencias buscadas. Los radios son la distancia a cualquiera de los centros más o menos el radio correspondiente.

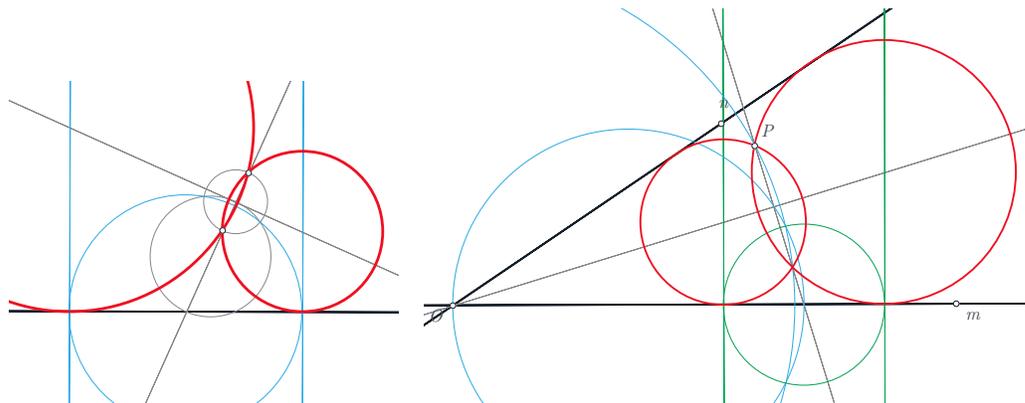


No sabemos si la solución de Apolonio utilizaba hipérbolas. Es posible. Tal vez fue resolver éste y otros problemas similares que eran difíciles de atacar con sólo regla y compás haya sido una de sus motivaciones para estudiar las secciones cónica. Lo interesante del problema es que sí se puede resolver con regla y compás, aunque el camino es complicado. François Viète (1540-1603) encontró una solución reduciendo el problema a otros más sencillos y combinándolas para obtener el resultado deseado. La idea de Viète dio lugar a lo que se conoce como *los 10 problemas de Apolonio*.

Los problemas se pueden plantear así: Para tres objetos cualesquiera que pueden ser puntos, rectas o círculos, encontrar los círculos que pasan por los puntos y son tangentes a las rectas y a los círculos. A cada uno de estos problemas se le asigna una combinación de las letras P , L , C para especificarlo, ordenadas siempre en el orden indicado. Así el problema de Apolonio PPP consiste en encontrar el círculo que pasa por tres puntos dados y el problema PLC consta de encontrar los círculos que pasan por un punto dado y son tangentes a una recta y a un círculo dados. El problema original de Apolonio se convierte así en el CCC . Algunos de estos problemas son muy sencillos de resolver. Por ejemplo, el PPP se resuelve encontrando el circuncentro de los tres puntos. El problema LLL tiene 4 soluciones, que son la circunferencia inscrita y las tres circunferencia excritas.

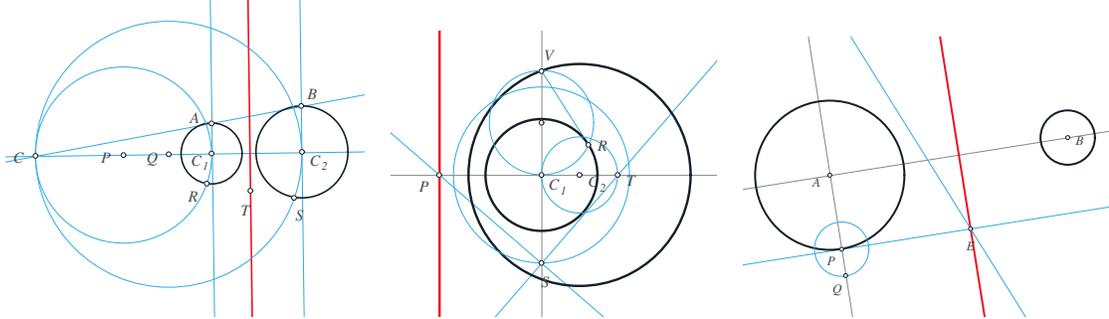
Soluciones a los problemas PPP y LLL

Las siguientes figuras ilustran las ideas principales (con un clic en las imágenes se accede a la versión interactiva con explicaciones). Las demostraciones de que éstas construcciones logran el objetivo deseado se dejan como ejercicios para el lector.

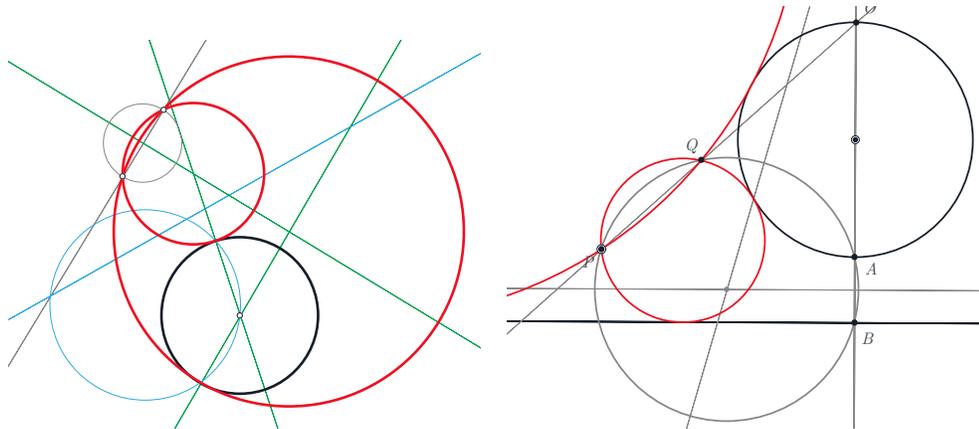
Soluciones a los problemas PPL y PLL

Hay un par de ideas que no han sido descritas en este texto y que son indispensables para resolver los siguientes casos: el *eje radical de dos circunferencias* y la operación de *inversión respecto a un círculo*.

El *eje radical* de dos circunferencias es una recta cuyos puntos son los centros de las circunferencias ortogonales a ellas. La siguiente figura muestra distintas maneras de construir el *eje radical* (que consta de los puntos que tienen la misma potencia respecto a un círculo) con regla y compás. Las primeras dos son complementarias. La tercera es general, funciona en todos los casos.



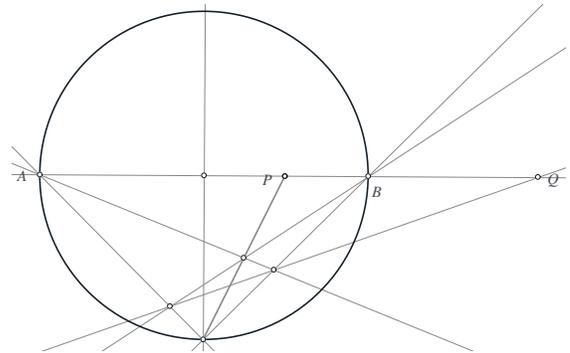
Armados con esta herramienta podemos enfrentar el problema de Apolonio *PPC*. El problema *PLC* es bastante complejo. En el fondo lo que hay que hacer es encontrar las intersecciones de dos parábolas cuyos ejes son ortogonales a la recta dada. La solución que ofrecemos hace precisamente eso utilizando solamente regla y compás.



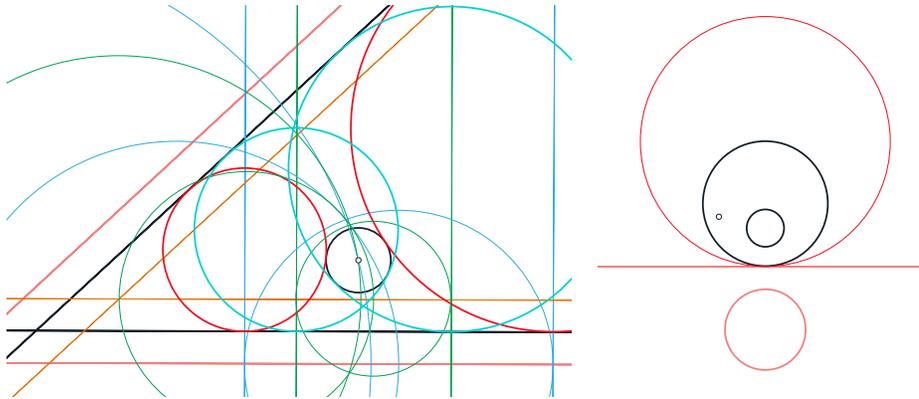
Soluciones a los problemas *PPC* y *PLC*

La solución de algunos de los siguientes problemas utiliza el concepto de *inversión*.

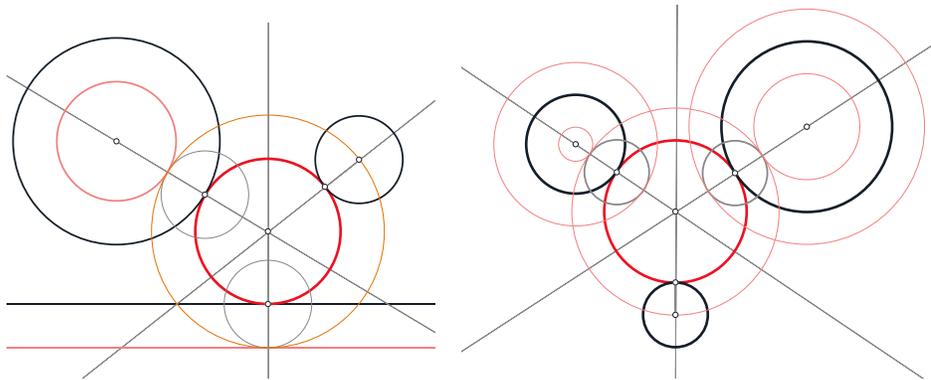
El *inverso* de un punto respecto a un círculo puede definirse como el armónico conjugado del punto respecto al diámetro que lo contiene. Lo interesante de esta transformación llamada inversión, es que transforma circunferencias en circunferencias, las que pasan por el centro, que las transforma en rectas. Además, aplicar dos veces la misma inversión regresa todos los puntos (excepto el centro) a su posición original.



Los siguiente problemas, el *LLC* y el *PCC* no se resuelven directamente sino que se reducen a otros ya resueltos utilizando esencialmente dos trucos: para el *LLC* se contrae la circunferencia a un punto recorriendo paralelamente las rectas por una distancia igual al radio, con lo que el problema se reduce al caso *PLL*, ya resuelto. Para el problema *PCC* se realizan una *inversión* respecto a una circunferencia adecuada que transforma una de las circunferencias dadas en una recta, con lo que se reduce al problema *PLC* ya resuelto.

Soluciones a los problemas *LLC* y *PCC*

Finalmente el problema *LCC* se resuelve reduciendo una de las circunferencia, la más pequeña, a un punto; la otra se reduce a una con el mismo centro y con el radio igual a la diferencia de los radios originales y la recta se traalada paralelamente una distancia igual al radio menor, con lo que el problema se reduce al caso *PLC*, ya resuelto. Finalmente el caso *CCC* se reduce al *PCC* reduciendo los radios de las tres circunferencias por el del más pequeño.

Soluciones a los problemas *LCC* y *CCC*

Esto completa la solución de los 10 problemas de Apolonio únicamente con regla y compás, más o menos como lo hizo François Viète en el siglo XVI.

La importancia de los problemas de Apolonio radica en que ayudaron a generar nuevas construcciones con regla y compás, como la del eje radical y la inversión, promoviendo avances importantes en la geometría.

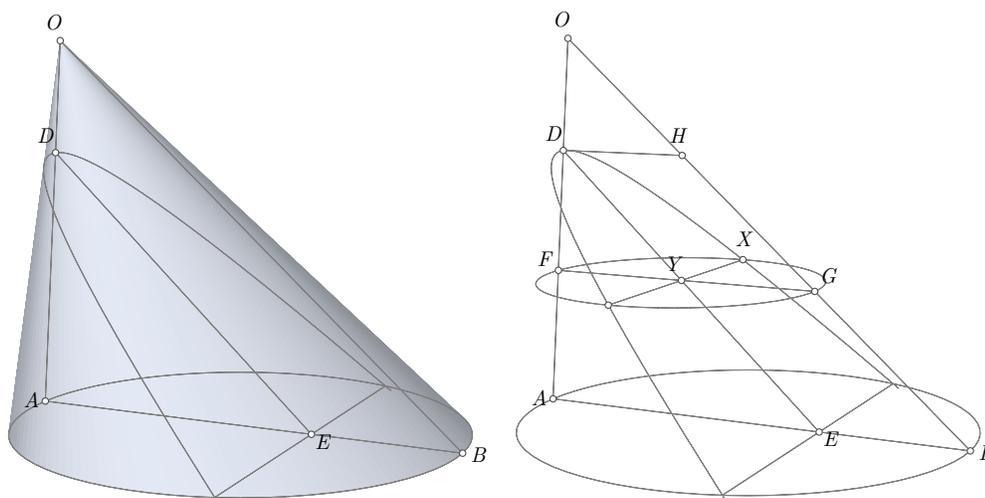
4.5. Las secciones cónicas según Apolonio

Las curvas cónicas pueden generarse cortando un cono circular con un plano. Estas curvas fueron estudiadas en la antigua Grecia, especialmente por uno de sus grandes matemáticos: APOLONIO DE PERGA (262-190 a.e.c.), tanto por sus interesantes propiedades de reflexión como por su procedencia, que las vincula íntimamente con el círculo, considerado por ellos como la figura perfecta. Casi dos mil años más tarde, resultaron de gran utilidad para comprender y describir el movimiento planetario (Leyes de Kepler).

Veamos algunos de los resultados más conocidos de APOLONIO.

La parábola

A la curva que se forma como intersección de un cono circular y un plano que es paralelo a una de sus generatrices y no pasa por el vértice, Apolonio la llamó *parábola*. Veamos cómo trata APOLONIO a la parábola.



Sea O el vértice del cono. El plano de la parábola y el que contiene a la base del cono se cortan en una recta. Sea AB el diámetro de la base del cono que es perpendicular a ella. Sean D y E los puntos de intersección del plano de la parábola con OA y AB , respectivamente. Entonces OB es paralela a DE y es precisamente la generatriz del cono a la que el plano de la parábola es paralelo.

Sea X un punto de la parábola. Consideremos el plano paralelo a la base del cono que pasa por X . Sean F , G y Y sus intersecciones con OA , DE y OB , respectivamente. Entonces F , G y X están en una circunferencia de la que FG es un diámetro; Y está en FG y YX es perpendicular a FG . Del teorema en 4.1 sobre la potencia de un punto respecto a un círculo, obtenemos

$$YX^2 = FY \cdot YG. \quad (4.1)$$

Es claro que $YG = DH = EB$ es constante, es decir, no depende de X . Por semejanza de triángulos vemos que $FY = DY \cdot \frac{AB}{OB}$. Por tanto,

$$YX^2 = DY \cdot \frac{AB}{OB} \cdot YG. \tag{4.2}$$

Si definimos $q = \frac{AB}{OB} YG$, que es una constante, la igualdad 4.2 se puede escribir como

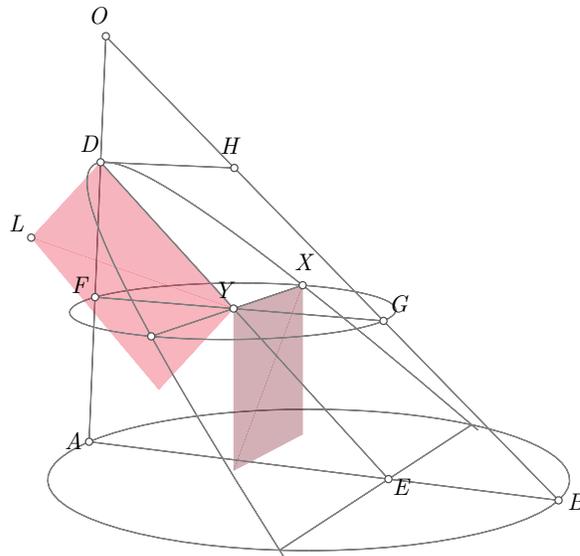
$$YX^2 = DY \cdot q \tag{4.3}$$

Finalmente, si definimos $x = YX$ y $y = DY$, esta relación adquiere un aspecto muy conocido:

$$x^2 = qy,$$

el de la ecuación que identificamos como “de la parábola” en los cursos de geometría analítica.

Apolonio construye un rectángulo de lados DY y q (en la figura, $DL = q$) e interpreta su resultado como la igualdad (de las áreas) del cuadrado de lado YX y el rectángulo de lados DY y DL .

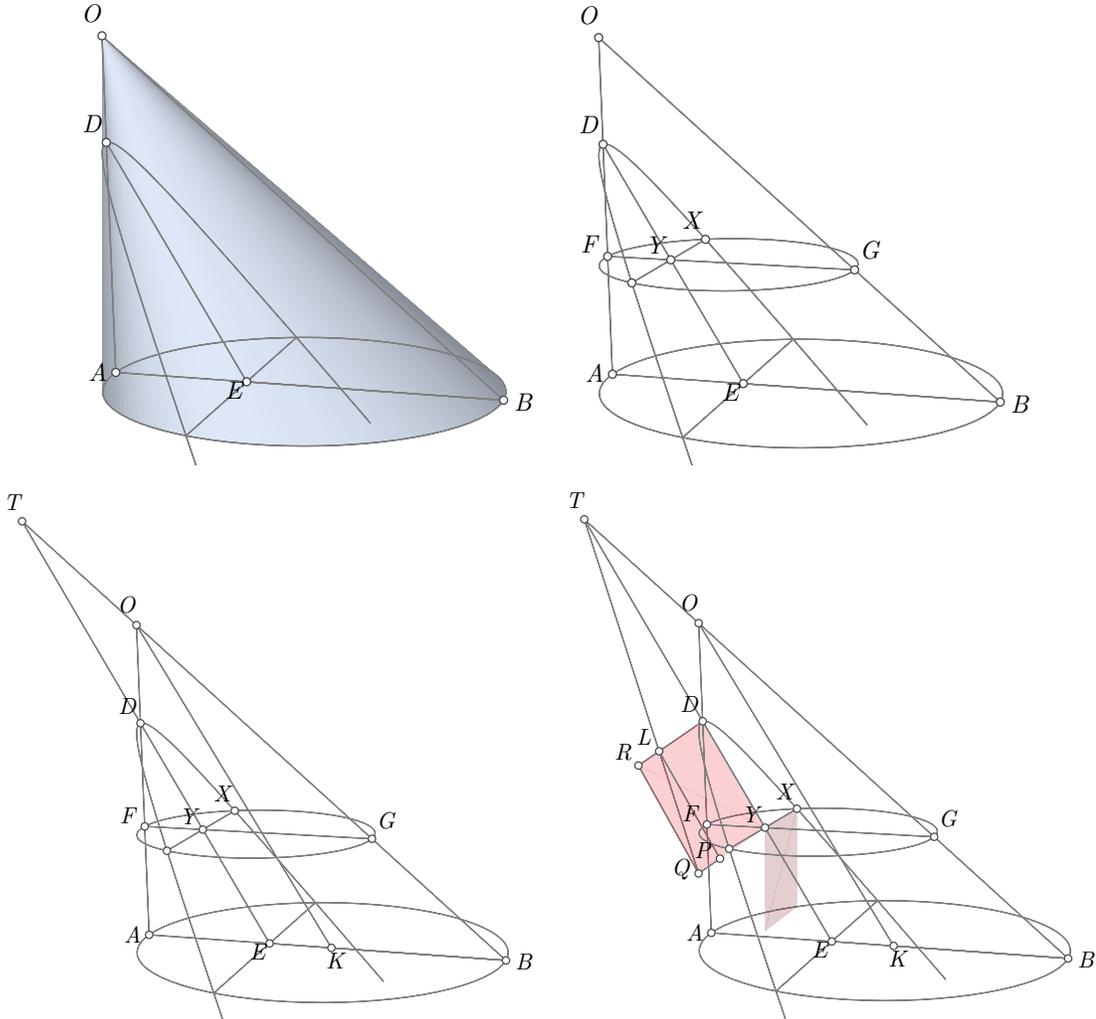


Para los antiguos griegos esto era necesario pues sólo podían interpretar la multiplicación de segmentos como rectángulos. Hoy en día este detalle, aunque resulte visualmente atractivo, nos parece innecesario porque aceptamos (sin cuestionarla mucho) la idea de que todo segmento tiene una longitud que se expresa como un número real. Pero en la antigua Grecia tal cosa no era posible pues, por un lado, sólo concebían los números fraccionarios (racionales) y, por otro, sabían que no siempre era posible asignar longitudes fraccionarias simultáneamente a dos segmentos, como por ejemplo, al lado y a la diagonal de un cuadrado (ver 2.3).

La hipérbola

Consideremos un cono circular oblicuo con vértice O cuya base es una circunferencia. Sea AB un diámetro de la circunferencia. Cortemos el cono con un plano perpendicular a OAB . Sean D y E

los puntos de intersección de ese plano con OA y AB , respectivamente. Supondremos que DE tiene mayor inclinación que OB , que es el caso en que el corte que produce a una hipérbola.



Consideremos un plano cualquiera que sea paralelo a la base del cono y cuya intersección con ED es el punto Y . Sean F y G las intersecciones de OA y OB con ese plano, respectivamente. Sea X la intersección del cono con de la perpendicular a FG , en ese plano, que pasa por Y . Sea T la intersección de DE con OB y sea K en AB tal que OK es paralela a DE . Sabemos que $YX^2 = FY \cdot YG$. Por semejanza de triángulos: $\frac{FY}{DY} = \frac{AK}{OK}$ y $\frac{YG}{YT} = \frac{KB}{OK}$. Por tanto $YX^2 = DY \cdot YT \cdot AK \cdot \frac{KB}{OK^2}$. Si definimos $DL = DT \cdot AK \cdot \frac{KB}{OK^2}$, podemos escribir $YX^2 = DY \cdot DL \cdot \frac{YT}{DT}$. Como $YT = DT + DY$, esto nos da

$$YX^2 = DY \cdot \frac{DT + DY}{DT} \cdot DL = \left(DY + \frac{DY^2}{DT}\right) \cdot DL$$

Sustituyendo YX por x y DY por y , esta igualdad se puede escribir

$$x^2 = \left(y + \frac{y^2}{DT}\right) \cdot DL$$

que es la ecuación de una hipérbola en coordenadas oblicuas.

Apolonio en su demostración construye un rectángulo sobre DY cuyos lados son DY y $DL = DT \cdot AK \cdot \frac{KB}{OK^2}$. Sea P la intersección de TL con la paralela a DL por Y y sea R la intersección de DL con la paralela a PL por Q . Lo anteriormente demostrado nos dice que $YX^2 = DR \cdot DY$. Apolonio llamó “hipérbola.” a esta curva porque en este caso $DR > DL$ (*hiper* sugiere que algo sobra, concretamente a DR le sobra algo respecto a DL).

Apolonio dio mucha importancia a este hecho, quizás porque para los griegos la multiplicación de dos segmentos sólo se podía entender como (el área de) un rectángulo. Así pues, Apolonio describe la hipérbola con la igualdad de (las áreas de) un rectángulo y un cuadrado.

La elipse

A la curva que se obtiene cuando se corta un cono circular con un plano que sólo toca a uno de los mantos del cono Apolonio la llamó elipse. El tratamiento que hace Apolonio de la elipse y casi idéntico al de la hipérbola, salvo por un signo.

Consideremos un cono circular oblicuo con vértice O cuya base es una circunferencia. Sea AB un diámetro de la circunferencia. Cortemos el cono con el plano perpendicular a OAB que pasa por DE . Supondremos que DE tiene menor inclinación que OB , que es el caso en que el corte que produce a una elipse.

Consideremos un plano cualquiera que sea paralelo a la base del cono y cuya intersección con ED es el punto Y . Sean F y G las intersecciones de OA y OB con ese plano, respectivamente. Sea X la intersección del cono con de la perpendicular a FG , en ese plano, que pasa por Y . Sea T la intersección de DE con OB y sea K en AB tal que OK es paralela a DE . Sabemos que $YX^2 = FY \cdot YG$. Por semejanza de triángulos: $\frac{FY}{DY} = \frac{AK}{OK}$ y $\frac{YG}{YT} = \frac{KB}{OK}$. Por tanto $YX^2 = DY \cdot YT \cdot AK \cdot \frac{KB}{OK^2}$. Si definimos $DL = DT \cdot AK \cdot \frac{KB}{OK^2} \cdot DL$, podemos escribir $YX^2 = DY \cdot \frac{YT}{DT}$. Como $YT = DT - DY$, esto nos da

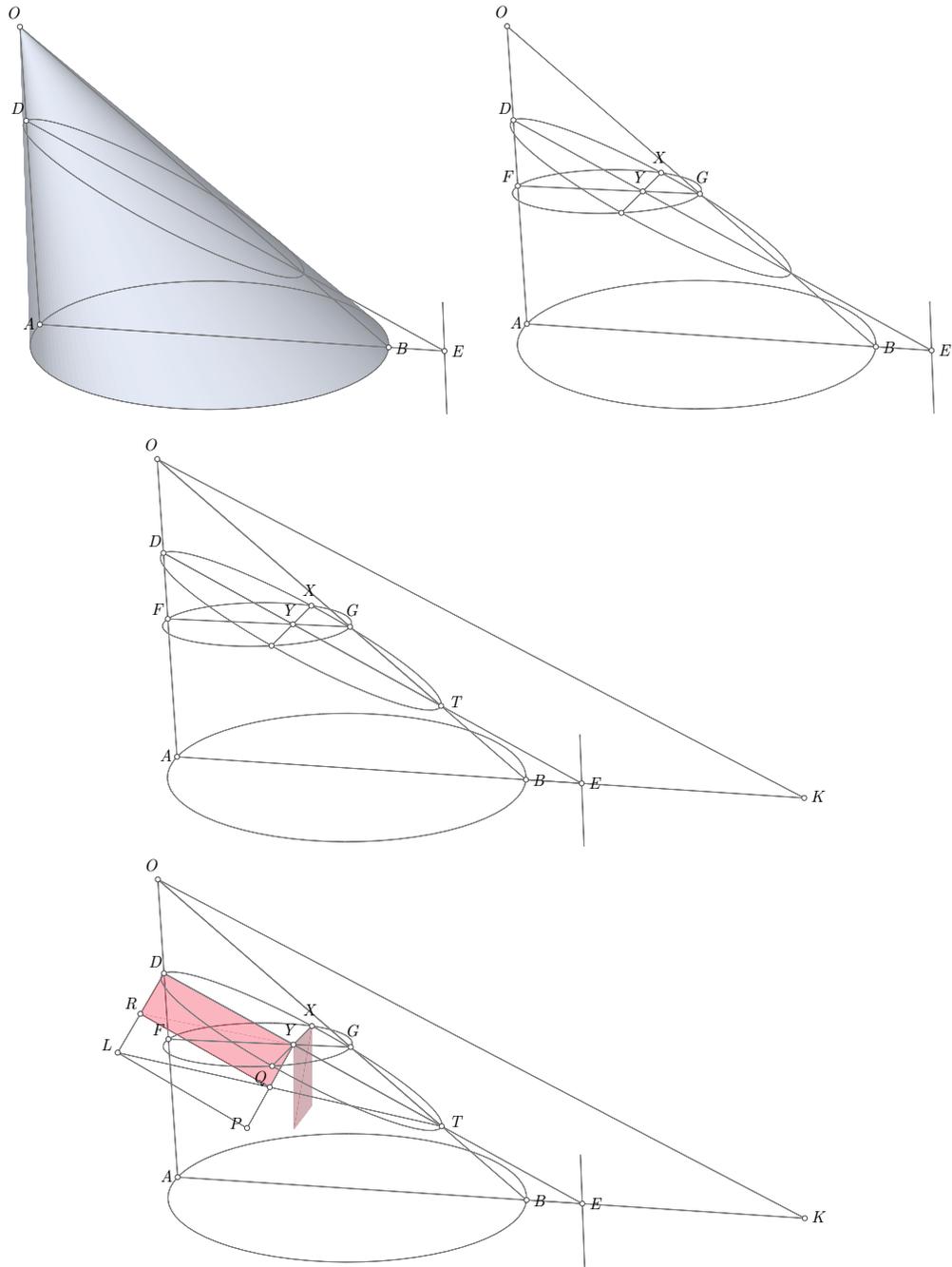
$$YX^2 = DY \cdot \frac{DT - DY}{DT} \cdot DL = \left(DY - \frac{DY^2}{DT}\right) \cdot DL$$

Sustituyendo YX por x y DY por y , esta igualdad se puede escribir

$$x^2 = \left(y - \frac{y^2}{DT}\right) \cdot DL$$

que es la ecuación de una elipse en coordenadas oblicuas.

Igual que en el caso de la hipérbola, Apolonio construye un rectángulo sobre DY cuyos lados son DY y $DL = DT \cdot AK \cdot \frac{KB}{OK^2}$.



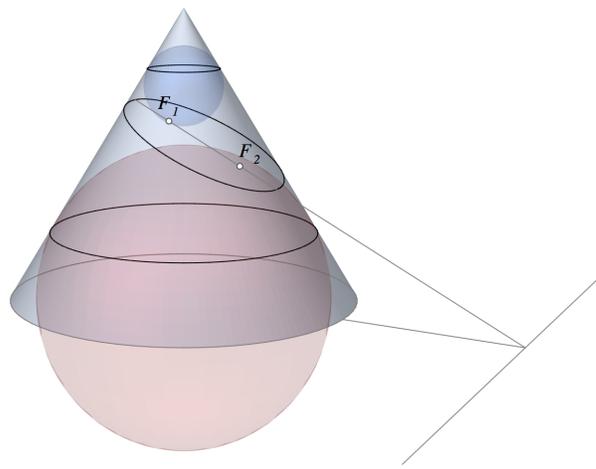
Sea P La intersección de TL con la paralela a DL por Y y sea R la intersección de DL con la paralela a PL por Q . Lo anteriormente demostrado nos dice que $YX^2 = DR \cdot DY$. Apolonio llamó “elipse” a esta curva porque en este caso $DR < DL$ (a DR le falta algo para llegar a DL).

Los resultados de Apolonio hablan esencialmente de *ecuaciones* que las secciones cónicas satisfacen, aunque ni el concepto ni el término de *ecuación* no se conocían en aquella época. Para los antiguos griegos, los resultados de Apolonio, definían relaciones geométricas entre algunos elementos de estas curvas, que a la postre, en el siglo XVII acabaron identificándose como ecuaciones. Además de este trabajo, Apolonio realiza un estudio detallado de otras propiedades de las secciones cónicas, en particular deduce sus propiedades de reflexión.

Las esferas de Dandelin

Las esferas de Dandelin, inventadas 2000 años después del trabajo de Apolonio por el matemático francés GERMINAL PIERRE DANDELIN (1794-1847), muestran de manera gráfica y clara que, en el caso de los *conos circulares rectos*, es decir, aquellos cuyo eje es perpendicular a la base, la elipse tiene la propiedad de que la suma de las distancias de sus puntos a dos puntos llamados focos es constante y también de que la diferencia de las distancias de los puntos de una hipérbola a sus focos es constante.

Las *esferas de Dandelin* se definen para un plano que corta a un cono circular recto como aquellas que son tangentes tanto al cono como al plano. En los casos en que el corte producido por el plano en el cono es una elipse o una hipérbola, hay dos de estas esferas y sus puntos de tangencia con el cono son los focos de la sección cónica correspondiente. En el caso de la parábola hay una sola esfera de Dandelin y su punto de tangencia con el plano es el foco de la parábola.



La demostración es muy sencilla. En el caso de la elipse que se ilustra en la figura, desde cualquier punto de la curva, la suma de las distancias a los puntos F_1 y F_2 es igual a la distancia entre los círculos de tangencia de las esferas con el cono, o sea que es constante. En el caso de la hipérbola la distancia entre los círculos de tangencia es la diferencia entre las distancias a F_1 y F_2 .

A veces se dice que el método de las esferas de Dandelin sustituye al trabajo de Apolonio, pero esto no es cierto. Las esferas de Dandelin sólo funcionan en el caso de conos circulares *rectos*, mientras que el trabajo de Apolonio aplica a los cortes de planos para todos los conos circulares, sean o no rectos.

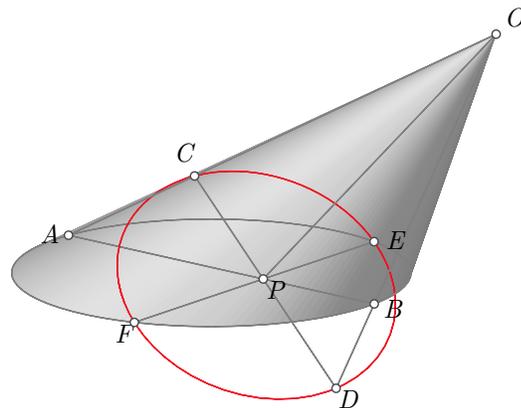
4.5.1. Polos y polares

Definimos la recta polar de un punto P dentro respecto a un círculo como la perpendicular al armónico conjugado de Q de P respecto al diámetro del círculo que pasa por P .

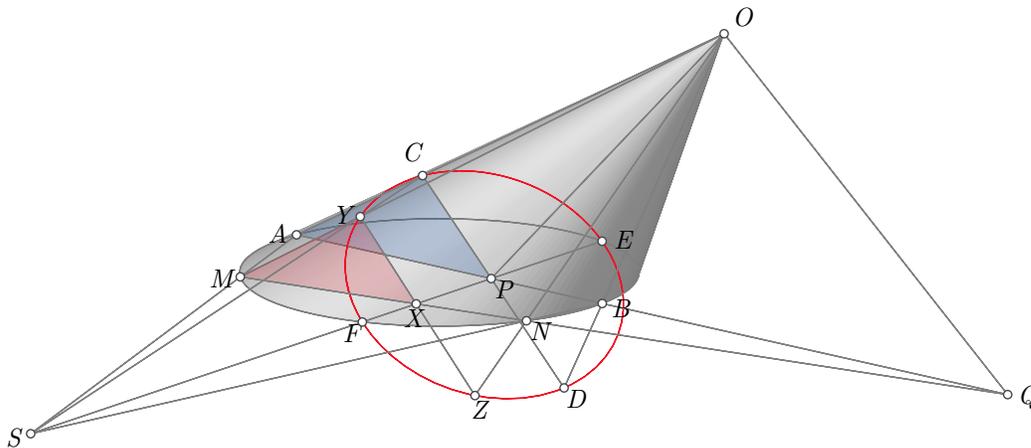
4.5.2. La perspectiva del círculo

Una de las cuestiones que llevó a los griegos a estudiar las curvas cónicas fue el deseo de comprender *la perspectiva del círculo*. ¿Qué es lo que el ojo ve cuando mira una circunferencia? A menos que la mire desde un punto en la perpendicular al plano que la contiene y que pasa por el centro de ella, no verá un círculo sino un óvalo. ¿Este óvalo es realmente una elipse? ¿Qué entendemos por elipse? La definición tradicional es que una elipse es la curva que se forma cuando un plano corta a un cono formando una curva acotada. Sin embargo esta definición no nos aporta una información vital: ¿qué simetrías tiene esta curva? El siguiente desarrollo nos lo aclara.

Consideremos un cono circular oblicuo. Sin pérdida de generalidad supondremos que su base es una circunferencia horizontal y su vértice un punto O que no está ni en la perpendicular al centro de la base ni en el plano que la contiene. Sean A y B las intersecciones de la circunferencia base del cono con el plano perpendicular a ella y que pasa por el vértice O y por el centro de la base. Sea P el punto de intersección de la bisectriz de AOB con AB . Sean F y E los puntos de intersección del plano ortogonal a OP por P con la base del cono. Y sean C y D las intersecciones de ese mismo plano con OA y OB , respectivamente. La curva $CFDE$ es lo que el ojo ve al mirar la circunferencia desde O . Veremos que $CFDE$ es una elipse.



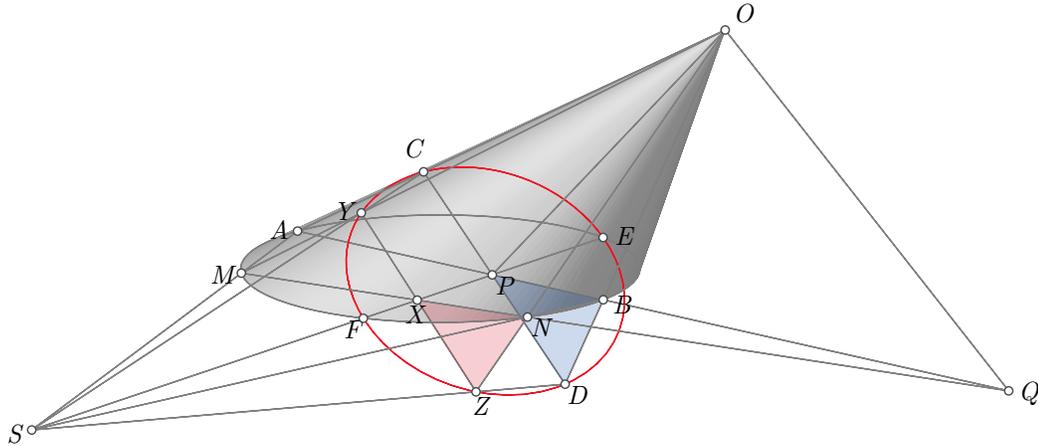
Sea Q el punto de intersección de AB con la paralela a CD por O . Sea X un punto de EF y sean M y N las intersecciones de QX con la base del cono.



Sean Y y Z las intersecciones del plano PCF con OM y ON , respectivamente. Observemos que

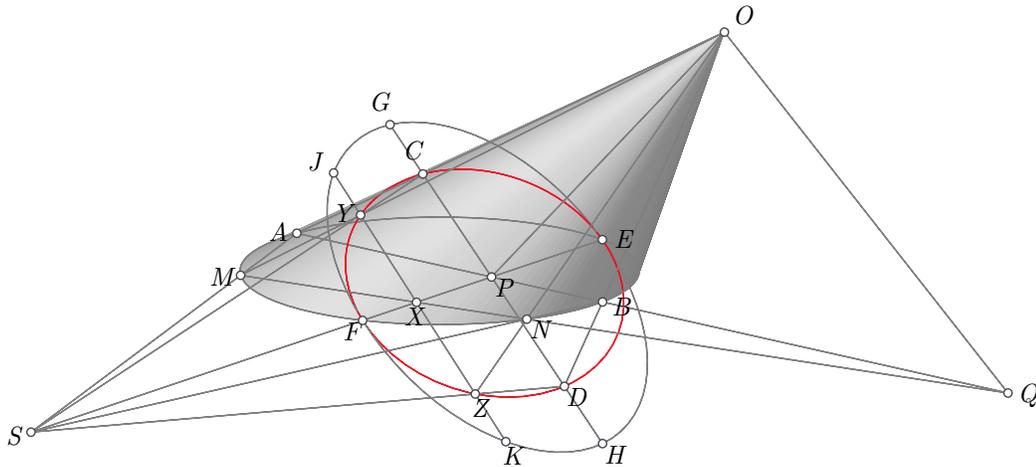
YZ es paralela a CD . Vamos a demostrar que $XY = XZ$, y por lo tanto OEF es un plano de simetría de nuestro cono oblicuo.

EF es la polar (ver Sección ??) de Q respecto a la base del cono. Por tanto AM y BN intersectan a EF en un mismo punto S . Los triángulos MXY y APC están en perspectiva axial porque AC y MY se intersectan en O , AP y MX en Q y PC y XY son paralelos. Por el inverso del Teorema de Desargues, también están en perspectiva polar y por tanto C y Y están alineados con S .



Análogamente, D y Z están alineados con S porque XNZ y PBD están en perspectiva axial. Como SPC y SPD son triángulos congruente y YZ es paralelo a CD , también SXY y SXZ son congruentes y por tanto $XY = XZ$.

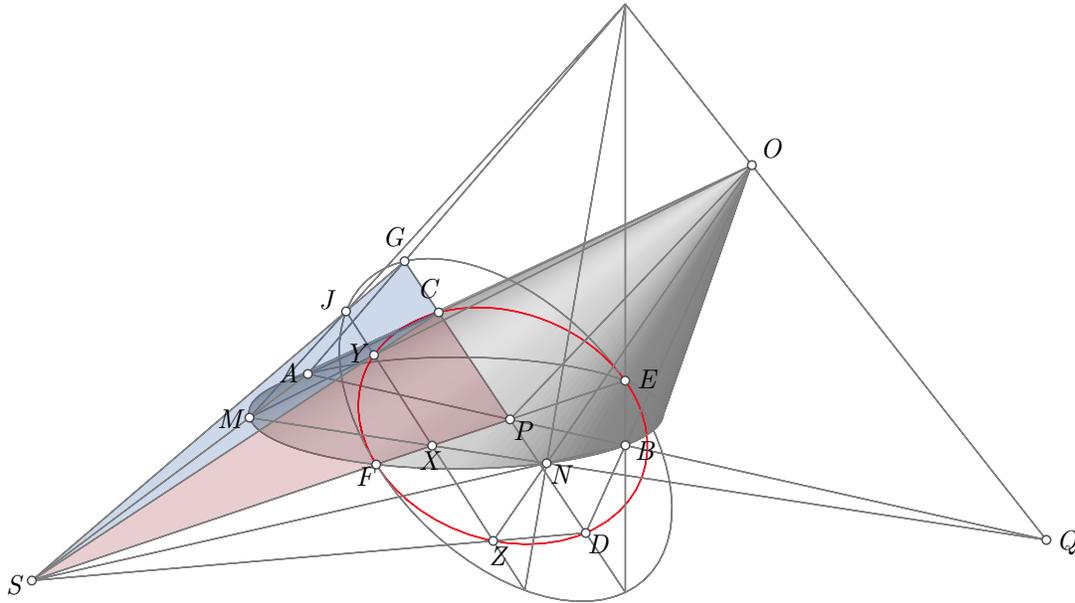
Ahora probaremos que la curva $ECFD$ es una contracción de la circunferencia con centro en P y radio PF , que yace en el plano ortogonal a OP .



Sean G y H las intersecciones de PC con la nueva circunferencia. Sean J y K las intersecciones de

XY con esa misma circunferencia. Sea R el vértice del cono que contiene a las dos circunferencias: la base del cono y EGF . R es la intersección de AG con HB y está alineado con O y Q .

De hecho OQ es la polar de EF respecto a la esfera que contiene a las dos circunferencias. Por tanto, MXH y APG están en perspectiva axial y polar. En particular, los puntos S , H y G están alineados.



Como SXY y SPC son triángulos semejante y también lo son SXJ y SPG , resulta que $\frac{XY}{XJ} = \frac{PC}{PG}$. Esto demuestra que la curva $ECFD$ es una contracción de la circunferencia $EGFH$ en la dirección PC , por el factor $\frac{PC}{PG}$. Esto es lo que coincide con nuestra intuición de lo que debe ser una elipse: la contracción de una circunferencia.

Capítulo 5

Geometría y Medición

Estrategias para cuantificar el espacio

La mayor parte de las aplicaciones de la geometría a la vida productiva se realizan al intentar calcular tamaños, distancias, áreas o volúmenes de objetos del mundo material cuya forma puede describirse en términos de figuras geométricas. En estos casos resulta imprescindible poder asignarles valores numéricos. De esto se ocupa la geometría cuando se combina con la *medición*.

En este capítulo exploramos la relación entre la geometría clásica y la necesidad de medir cosas que se pueden identificar con algunas partes de figuras geométricas. Primero se investigan las relaciones entre los conceptos geométricos de longitud, área y volumen con sus medidas numéricas y más adelante se presentan algunos casos de cálculos específico junto con los métodos que para realizarlos se han desarrollado a través de los siglos.

Cuando las nociones de medida se intentan aplicar a figuras geométricas formadas por curvas que no están acotadas por líneas rectas, o por cuerpos que no están acotados por superficies planas, la medición comienza a necesitar de procesos infinitos. Los esfuerzos de Eudoxo de Cnido (390 - 337 aec) y Arquímedes de Siracusa (287 - 212 aec) en esta dirección representan los primeros avances en lo que dos milenios más tarde, culminaría con la creación del *cálculo*, gracias al trabajo de matemáticos de los siglos XVII y XVIII como Isaac Newton (1643 - 1727), Gottfried Leibniz (1646 - 1716), Jakob Bernoulli (1654 - 1705), Johann Bernoulli (1667 - 1748) y Leonhard Euler (1707 - 1783).

5.1. El uso de los números para cuantificar la realidad

Las necesidades del comercio, el cobro de impuestos, la construcción y muchas otras actividades que realizaban los miembros de las diversas culturas de la antigüedad requirieron no sólo la creación de los números sino también formas de cuantificar, asignando valores numéricos, a ciertos aspectos de la realidad como tamaño, peso, duración, sin lo cual su utilidad era escasa. Este tipo de actividad ha sido probablemente la mayor fuente de conceptos e ideas matemáticas desde la antigüedad hasta

nuestros días. Esta fue, sin duda, una de las primeras actividades humanas que requirió la aplicación de un pensamiento matemático avanzado.

Los registros que tenemos en tabletas y papiros muestran un interés en realizar cálculos que permitieran asignar valores numéricos a ciertas *abstracciones de utilidad práctica*. En particular exhiben el conocimiento de algunos procedimientos y fórmulas para calcular *áreas* y *volúmenes*, a veces de manera exacta y a veces sólo aproximada. Por ejemplo, hay en ellas estimaciones de lo que hoy llamamos *el número π* , que eran útiles en fórmulas o procedimientos que se usaban para calcular *áreas* de terrenos y *volúmenes* de pirámides y contenedores. No hay evidencias de que se hubiera llegado a esas estimaciones, fórmulas y procedimientos mediante razonamientos deductivos, lo cual apunta a un pensamiento matemático muy primitivo. Esta actividad matemática surge en la antigüedad como una *tecnología* asociada a la cuantificación, es decir, a la asignación de valores numéricos a ciertas abstracciones de interés práctico. Los métodos de cálculo se desarrollan para obtener unos valores a partir de otros aprovechando las propiedades de los números y sus operaciones, las cuales aparecen como algo natural en esos contextos. La notación numérica tan distinta en las diversas culturas atestigua que estamos ante desarrollos independientes que obedecieron a necesidades prácticas similares y dieron lugar a sistemas equivalentes para denotar a los números y realizar operaciones con ellos. Podríamos decir que se trata de una concepción de las matemáticas como una tecnología para cuantificar abstracciones de utilidad práctica. Aún no aparece, al menos de manera explícita, la deducción.

Con el tiempo, esta actividad se fue haciendo cada vez más racional, justificando de manera rigurosa cada fórmula y procedimiento que se inventaba, de manera que acabó integrándose por completo a la actividad matemática deductiva inaugurada por los antiguos griegos con la geometría, aunque sigue dando mayor importancia a la utilidad práctica que al aspecto teórico.

Los métodos de cálculo de los babilonios y los egipcios

Hemos visto que los antiguos babilonios y egipcios ya tenían sorprendentes conocimientos prácticos de matemáticas y que incluso habían desarrollado formas de enseñarlos a través de la escritura en tabletas de barro y papiros, respectivamente. Tales conocimientos se transmitían a través de ejemplos concretos, algunos de los cuales ya se describieron en la Sección 3.1. El siguiente ejemplo, que aparece en una de las tabletas babilonias, ilustra la capacidad que tenían para describir algoritmos de cálculo y realizar operaciones con su extraordinaria notación numérica. El ejemplo enseña cómo calcular la *longitud* de uno de los catetos de un triángulo rectángulo cuya hipotenusa y el otro cateto se conocen. Se trata de una viga que primero se apoya verticalmente en un muro y luego su parte inferior es separada de la pared, de manera que la superior descende una cierta distancia. Para reconocer el significado del contenido de la tableta hay que recordar que los babilonios usaban el sistema sexagesimal y que en su notación $a; b, c$ representa al número $a + b\frac{1}{60} + c\frac{1}{60^2}$. El ejemplo se ve así:

Una viga 0;30	tiene la longitud $30 \frac{1}{60} = \frac{1}{2}$
De arriba bajó 0;6	eso equivale a $6 \frac{1}{60}$
Abajo, cuánto se alejó?	
0;30 elévalo al cuadrado, 0;15 lo ves	$(30 \frac{1}{60})^2 = 900 \frac{1}{60^2} = \frac{15}{60}$
0;6 de 0;30 réstalo, 0;24 lo ves	$30 \frac{1}{60} - 6 \frac{1}{60} = 24 \frac{1}{60}$
0;24 elévalo al cuadrado, 0;9,36 lo ves	$(24 \frac{1}{60})^2 = \frac{576}{3600} = 9 \frac{1}{60} + 36 \frac{1}{60^2}$
0;9;36 de 0;15 réstalo, 0;5,24 lo ves	$15 \frac{1}{60} - (9 \frac{1}{60} + 36 \frac{1}{60^2}) = 5 \frac{1}{60} + 24 \frac{1}{60^2}$
0;5,24 tiene qué como raíz? 0;18 lo es	$(18 \frac{1}{60})^2 = 324 \frac{1}{60^2} = 5 \frac{1}{60} + 24 \frac{1}{60^2}$
0;18 se alejó.	

En las fuentes antiguas, como las tablillas de arcilla que se encontraron en Babilonia, no hay ningún indicio de una *justificación* lógica de por qué esta serie de instrucciones funciona *siempre*, es decir aún si se toman otros números de los que se dan en el ejemplo.

Es interesante ver que en las fuentes egipcias y babilonias se encontraron dos métodos diferentes de calcular el volumen de un *tronco*, es decir una pirámide cuadrada a la cual se le quitó la punta con un corte paralelo a la base.

Si se traducen estos dos métodos a nuestro lenguaje algebraico, resultan las siguientes dos fórmulas:

$$V = \frac{1}{2}(a^2 + b^2)h \quad \text{Babilona (ca. 1900 a.e.c.)}$$

$$V = \frac{1}{3}(a^2 + ab + b^2)h \quad \text{Egipto (ca. 1850 a.e.c.)}$$

En ambas a y b son los lados de las dos bases y h es la altura. Está claro que las fórmulas no pueden ser ambas correctas.

Ejercicio. Sugerimos al lector determinar rigurosamente si alguna de las dos fórmulas es correcta; en qué casos, si los hay, alguna de ellas puede ser correcta y, si alguna no es correcta, encontrar qué tan buenas o malas aproximaciones produce. Este ejercicio le ayudará a apreciar el nivel matemático de la cultura egipcia.

Como veremos en este capítulo, cuantificar conceptos como longitud, área, volumen, peso y tiempo, es decir aquello que no es *discreto*, sino *continuo*, no fue fácil. Una de las razones por las que se encontraron dificultades para lograrlo es que la humanidad no tenía, en lenguaje natural, una noción suficientemente amplia de *número*. Carecía de lo que hoy llamamos *números reales*, que son los que se necesitan para cuantificar lo que es continuo. En cierto sentido, este capítulo es una historia del desarrollo paralelo de las ideas del *continuo* y de los *números reales*.

Recordemos que la cultura clásica griega, que es de la que parten estas ideas, encontró muy temprano en su historia que había longitudes incommensurables, es decir, que los números fraccionarios, que son los que habían llegado a manejar las civilizaciones más antiguas como la babilonia y la egipcia, eran insuficientes para cuantificar lo continuo. La respuesta inicial ante este problema, como veremos en las primeras secciones del capítulo, fue desarrollar los conceptos de manera geométrica, evitando tanto como era posible su cuantificación. De hecho, *Los elementos* de Euclides logran una presentación muy completa de los conceptos de longitud, área y volumen sin asignarles valores numéricos.

5.2. Longitud, área y volumen

Introducción

El volumen de un cuerpo depende de su tamaño y de su forma. Por ejemplo, un cubo tiene mayor volumen que una esfera cuyo diámetro es igual al lado del cubo. Esto lo sabemos porque la esfera cabe dentro del cubo y no lo llena. Tenemos una idea intuitiva de lo que llamamos *volumen* y sabemos que, independientemente de cómo lo midamos, debe tener ciertas propiedades. Concretamente:

- (a) Si un cuerpo A cabe en otro cuerpo B , el volumen de A es menor que el de B .
- (b) Si A y B son cuerpos ajenos, el volumen de la unión de A y B es la suma de los volúmenes de A y B .

Al hacer estas afirmaciones suponemos inconscientemente que el volumen de un cuerpo es un número. (a) dice que uno de esos números es menor que otro y (b) habla de sumar dos de esos números.

Lo mismo que con el volumen ocurre con las longitudes de segmentos y las áreas de figuras planas.

Sin embargo Euclides, en su libro *Los elementos*, hace un tratado muy completo de longitudes, áreas y volúmenes sin asignar valores numéricos a tales conceptos. De hecho ni siquiera introduce palabras equivalentes a longitud, área y volumen.

¿Cómo es eso posible?

Simplemente crea un concepto de igualdad diciendo que dos segmentos son *iguales* si se pueden superponer exactamente, es decir, si uno de ellos se puede mover rígidamente (mediante traslaciones y giros) hasta superponerlo exactamente sobre el otro (en el lenguaje moderno: si son congruentes).

En el caso de las figuras planas dice que dos de ellas son iguales si una se puede cortar en partes que puedan acomodarse, mediante movimientos rígidos (sin deformarlas), hasta cubrir exactamente la otra.

La educación matemática actual tiene una fuerte tendencia a tratar los conceptos de longitud, área y volumen en términos estrictamente numéricos, lo cual propicia un ocultamiento de algunos aspectos geométricos muy profundos de estos conceptos y redundante en una presentación superficial del tema.

Usemos el Teorema de Pitágoras para ejemplificar este hecho. El teorema de Pitágoras dice, según Euclides, que

***En un triángulo rectángulo
los cuadrados contruidos sobre los catetos
son iguales al cuadrado construido sobre la hipotenusa***

En este enunciado los *cuadrados* son figuras geométricas. Lo que quiere decir es que hay una manera de cortar esos dos *cuadrados* cuyos lados son los catetos y acomodar las partes de manera que cubran exactamente el *cuadrado* cuyo lado es la hipotenusa. El enunciado habla de figuras geométricas y de operaciones geométricas consistentes en cortes y movimientos rígidos de los pedazos. En ningún momento habla de números asignados a esos cuadrados ni de sumarlos para compararlos con el

número asignado al otro.

En cambio, el enunciado que se usa en general en la educación matemática actual es:

***En un triángulo rectángulo,
la suma de los cuadrados de los catetos
es igual al cuadrado de la hipotenusa***

Los *cuadrados* que se mencionan en este enunciado son *números* que se obtienen al elevar al cuadrado otros números (multiplicándolos por sí mismos) que son las longitudes de los catetos. Y la *suma* se entiende como una operación aritmética, suma de números, y no como una operación geométrica.

Sin duda la creación de la Geometría Analítica y en general la aritmetización de la geometría fue un gran logro de la humanidad. Pero al reemplazar la enseñanza de la geometría clásica con su forma aritmetizada se ha perdido profundidad, se han empobrecido los conceptos y ha quedado un estudio que privilegia el cálculo por encima de la comprensión conceptual.

Es perfectamente posible desarrollar los conceptos matemáticos de longitud, área y volumen sin necesidad de asignarles valores numéricos. Para ello es necesario tratarlos como propiedades de los objetos geométricos, que se relacionan unas con otras. Por ejemplo, se pueden comparar y agregar y, cuando uno de ellos está contenido en otro, se puede tomar su diferencia.

En las próximas secciones nos apartamos del tratamiento numérico con la intención de recuperar, al menos en espíritu, el enfoque euclidiano, y exhibir con claridad los aspectos geométricos de la medición. El tratamiento numérico irá apareciendo poco a poco, de manera parecida a como ocurrió históricamente.

5.2.1. Los conceptos de longitud, área y volumen

En el antiguo Egipto, se calculaba el volumen de pirámides y otros cuerpos geométricos, tanto para estimar la cantidad de grano que podían guardar en sus trojes, como la cantidad de material y tiempo que iba a requerir una construcción.

Desarrollaron el concepto de volumen como una generalización del de área, y dejaron claro que el volumen de un prisma es el producto de la superficie de la base por la altura. Esto no se limita a prismas rectos sino que aplica a los oblicuos. Este hecho es una generalización, en tres dimensiones, de la relación ya conocida entre en un paralelogramo y el rectángulo con la misma base y la misma altura.

Los egipcios supieron calcular el volumen de las pirámides, tomando la tercera parte del área de la base por la altura. Se desconoce si tenían argumentos racionales para deducir esta fórmula, pero la usaron incluso para calcular el volumen de pirámides truncadas.

Observaron la analogía entre

$$\text{Area} = \frac{\text{base} \times \text{altura}}{2}$$

para el cálculo del área de un triángulo y

$$\text{Volumen} = \frac{\text{área de la base} \times \text{altura}}{3}$$

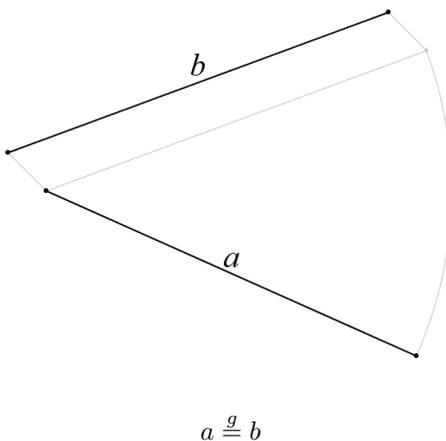
en el cálculo del volumen de una pirámide.

Como veremos más adelante, la primera demostración conocida de esta fórmula la dio Eudoxo de Cnido (390 - 337 aec). Pero para llegar a esto, conviene antes hacer una presentación conceptual de los conceptos de longitud, área y volumen que nos acerque al pensamiento griego de aquella época en relación a estos conceptos.

Longitud de segmentos

Partiremos de las nociones de *segmento recto*, que llamaremos simplemente *segmento* y de *transformación rígida*, entendida como aquella que no cambia ni el tamaño ni la forma de las cosas. Cada segmento queda determinado por dos puntos que se denominan los extremos del segmento. Los segmentos pueden estar sobre una línea recta, en un plano o en el espacio. Un segmento determina una única línea recta, la que determinan los puntos extremos del segmento, según los dos primeros postulados de Euclides.

Decimos que *dos segmentos a y b tienen la misma longitud* si se puede aplicar una transformación rígida a uno de ellos de manera que lo superponga exactamente sobre el otro. Para ello basta que la transformación rígida lleve los dos extremos del primer segmento a los dos extremos del segundo. Cuando dos segmentos a y b tienen la misma longitud, podemos escribir $a \stackrel{g}{=} b$.¹

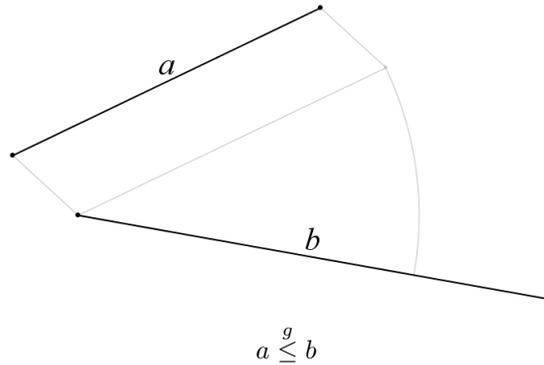


Es fácil demostrar que $\stackrel{g}{=}$ es una *relación de equivalencia*, es decir, que:

- para todo segmento a , $a \stackrel{g}{=} a$
- si a y b son dos segmentos tales que $a \stackrel{g}{=} b$ entonces $b \stackrel{g}{=} a$
- si a , b y c son tres segmentos tales que $a \stackrel{g}{=} b$ y $b \stackrel{g}{=} c$, entonces $a \stackrel{g}{=} c$.

También podemos comparar las longitudes de dos segmentos a y b . Decimos que la longitud de a es menor que la de b si es posible hallar una transformación rígida que lleve a a a estar contenido en b sin que los dos extremos de a coincidan con los de b . Cuando la longitud del segmento a es menor que la del segmento b podemos escribir $a \stackrel{g}{\leq} b$

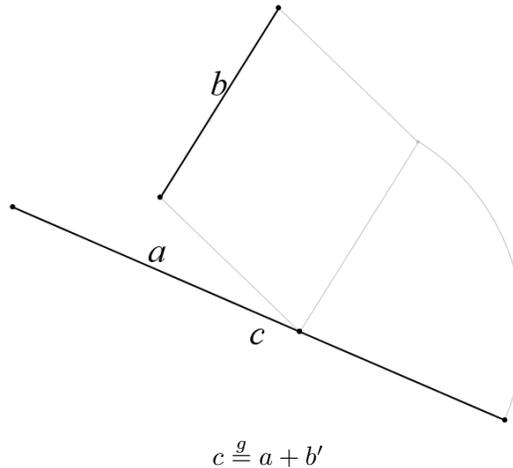
¹La letra g sobre los signos de igualdad y desigualdad indicará en esta sección que no se trata de relaciones numéricas sino geométricas



Es fácil ver que \lesssim es una relación transitiva, es decir, que si $a \lesssim b$ y $b \lesssim c$ entonces $a \lesssim c$.

Bajo ciertas circunstancias dos segmentos se pueden sumar. Concretamente, si dos segmentos a y b se encuentran sobre una misma línea recta y son ajenos excepto por un extremo de a coincide con uno de b , entonces el segmento c determinado por los dos extremos libres (o sea los que no coinciden), se puede llamar *la suma* de a y b , y se puede escribir: $a + b \stackrel{g}{=} c$.

En general no es posible sumar segmentos. Sin embargo, si a y b son dos segmentos cualesquiera, se puede construir un segmento b' sobre la línea recta determinada por a tal que $b' \stackrel{g}{=} b$ que tenga la misma longitud que b y quede en posición de poderse sumar con a . Tal construcción se puede hacer con regla y compás como se indica en la siguiente figura.



Podemos definir una “función” *Long* que asigne a cada segmento un ente abstracto que podemos llamar *longitud*, entendida no como un número sino como

la clase de equivalencia de todos los segmentos a' tales que $a' \stackrel{g}{=} a$

Con esta definición queda claro que la longitud $Long(a)$ de un segmento a **no es un número** pero podemos hablar de ella absoluta precisión, como veremos a continuación.

Todos los segmento que constan de un sólo punto, es decir, cuyos extremos coinciden, tiene la misma longitud. En efecto, uno de estos segmentos puede hacerse coincidir exactamente con cualquier otro

de ellos mediante una traslación. A estos segmentos los llamamos *segmentos nulos*. La longitud de los segmentos nulos es única y la consideramos como el *cero aditivo* de las longitudes.

Hemos establecido una manera de hablar de las longitudes de los segmentos sin necesidad de asignarles valores numéricos. Podemos comparar las longitudes de dos segmentos, saber si son iguales o si una es menor que la otra y hasta podemos sumar las longitudes de dos segmentos. Todo ello sin asignarles valores numéricos.

A continuación enunciamos las propiedades más importantes de la longitud de segmentos. Dejamos al lector los detalles de las demostraciones, algunas de las cuales utilizan propiedades de las transformaciones rígidas que aún no hemos especificado y que están íntimamente relacionadas con los axiomas referentes a la homogeneidad del espacio.

Propiedades de la longitud de segmentos

Sean a, b, c, \dots segmentos cualesquiera.

Simetría: Si $Long(a) = Long(b)$, entonces $Long(b) = Long(a)$.

Transitividad: Si $Long(a) = Long(b)$ y $Long(b) = Long(c)$, entonces $Long(a) = Long(c)$.

Conmutabilidad: $Long(a) + Long(b) = Long(b) + Long(a)$.

Cero aditivo: Si o es un segmento nulo (que es vacío o bien consta de un solo punto) entonces $Long(a) + Long(o) = Long(a)$ para cualquier segmento a .

Tricotomía: Sucede una y sólo una de estas tres posibilidades:

- (1) $Long(a) < Long(b)$
- (2) $Long(b) < Long(a)$
- (3) $Long(a) = Long(b)$.

Transitividad de la desigualdad: Si $Long(a) < Long(b)$ y $Long(b) < Long(c)$, entonces $Long(a) < Long(c)$.

Propiedad arquimediana: Dados dos segmentos a y b existe un entero n tal que $Long(a) < nLong(b)$.

Notas

1) Podemos extender la definición de $nLong(a)$ al caso $n = 0$ definiendo $0Long(a) = Long(o)$ donde o es cualquier segmento nulo.

2) Se puede definir la resta o diferencia $Long(b) - Long(a)$ de longitudes, pero sólo en el caso en que $Long(a) < Long(b)$.

3) En *Los elementos*, Euclides hace esencialmente lo mismo que hemos hecho aquí, pero sin hablar de longitud, sino que dice que *dos segmentos son iguales* si uno se puede superponer sobre el otro. Euclides y sus contemporáneos no distinguieron entre el segmento y su longitud. Hablaban de ambas cosas como si fueran la misma. En nuestra presentación, en cambio, dos segmentos son iguales únicamente si ocupan el mismo sitio, es decir, si tienen los mismos extremos. Si no, pueden tener la misma longitud, pero no ser iguales.

Armados ya con el concepto abstracto, no numérico, de la longitud de un segmento, podremos desarrollar en las siguientes secciones los de área y volumen, y más adelante el de longitud de una curva.

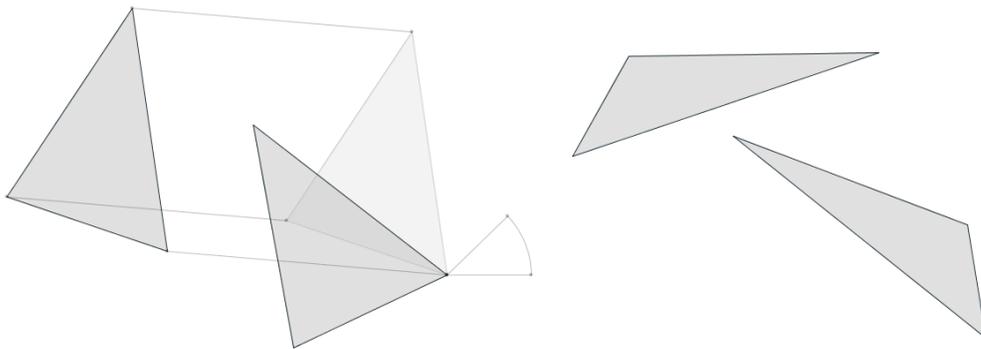
Es interesante observar que el concepto no numérico de longitud, que es esencialmente el que usa la geometría euclidiana y que hemos formalizado con la función $Long$, puede dar lugar al concepto numérico de longitud que se usa en la actualidad con sólo asumir la existencia de los números reales, como veremos en el siguiente párrafo.

Dado un segmento a podemos definir $2Long(a)$ como $Long(a + a')$ para un segmento a' tal que $a' \cong a$ que se encuentre en posición de ser sumado con a , es decir, que sea ajeno con a , excepto por un extremo de cada uno de ellos. Podemos extender esta definición y dar significado a $nLong(a)$ para cualquier entero positivo n . Igualmente se puede extender la definición a $qLong(a)$ para cualquier número racional $q = m/n$ y finalmente, por continuidad, a $rLong(a)$ para cualquier número real r . Si se adopta un segmento cualquiera u como “unidad”, se puede demostrar, usando el axioma de Arquímedes, que para cada segmento a , existe un único número real r para el cual $Long(a) = rLong(u)$. Así es posible definir una función que asigne a cada segmento a el número real tal que $Long(a) = rLong(u)$. Ésta es la función que se usa en la actualidad para hablar de longitudes de segmentos. Lo que hemos intentado mostrar en esta sección es que es posible hablar de longitud sin usar números, y eso es esencialmente lo que hacían los antiguos griegos cuando hablaban de igualdad de segmentos. Lo mismo ocurriría cuando hablaban de igualdad de figuras y cuerpos, queriendo significar la igualdad de sus áreas o sus volúmenes, como veremos a continuación.

Área de figuras planas

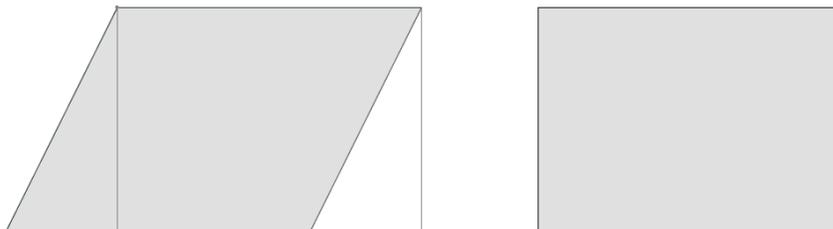
Decimos que dos figuras planas tienen la misma área si:

(a) podemos superponer exactamente una sobre la otra mediante traslaciones, rotaciones o reflexiones respecto a una recta.



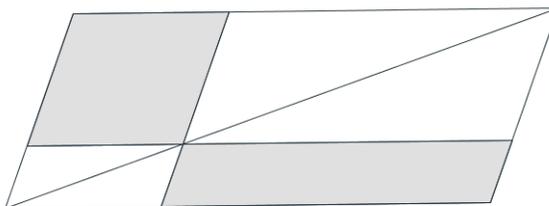
Ejemplos de triángulos con áreas iguales, según (a)

(b) podemos partir una de ellas y mover los pedazos hasta cubrir exactamente la otra



Ejemplo de dos paralelogramos con áreas iguales, según (b)

(c) podemos formarlas a partir de figuras con la misma área quitando de ellas partes que también tienen la misma área.



Ejemplo de dos paralelogramos con áreas iguales, según (c)

Nota: Las reflexiones de una figura plana respecto a un eje pueden considerarse rotaciones en el espacio tridimensional. Sin embargo cuando más adelante hablemos de cuerpos en el espacio, las reflexiones respecto a un plano no podrán considerarse rotaciones, a menos que éstas fueran en un espacio de 4 dimensiones. Por ello resulta más conveniente pensar en las reflexiones respecto a un eje como transformaciones en el plano que conservan el área.

Nota: Las traslaciones, rotaciones y reflexiones se llaman *transformaciones rígidas o de congruencia* y forman un grupo no conmutativo respecto a la operación de composición.

Si A y B tiene la misma área, según los criterios (a), (b) y (c), entonces escribimos:

$$A \stackrel{g}{=} B$$

El criterio (b) nos plantea un problema que es importante resolver de entrada. ¿Qué significa partir una figura? Tomemos el ejemplo de un triángulo que partimos mediante un corte recto que pasa por uno de sus vértices y por el punto medio del segmento opuesto. El segmento que determina el corte ¿en cual de las dos partes queda? ¿En una, en la otra, en ambas o en ninguna? Esta duda se relaciona también con la de si una figura polígona plana incluye o no su orilla. Lo primero que hay que aclarar es que estos problemas surgen porque pensamos en las figuras planas como subconjuntos del plano formados por puntos y, dada una figura, que concebimos como un conjunto de puntos, cada punto del plano debe estar en ella o en su complemento.

La comparación y la suma de áreas puede tratarse de manera análoga a como se hizo con las longitudes.

Decimos que el área de A es menor o igual que la de B si es posible hallar una partición de A tal que sus pedazos puedan meterse (mediante transformaciones de congruencia) dentro de B . Cuando

el área de A es menor o igual que la de B podemos escribir:

$$A \stackrel{g}{\leq} B$$

De momento nos limitaremos al estudio de figuras poligonales. La orilla o frontera de una figura poligonal es un polígono (segmentos “encadenados”) cerrado.

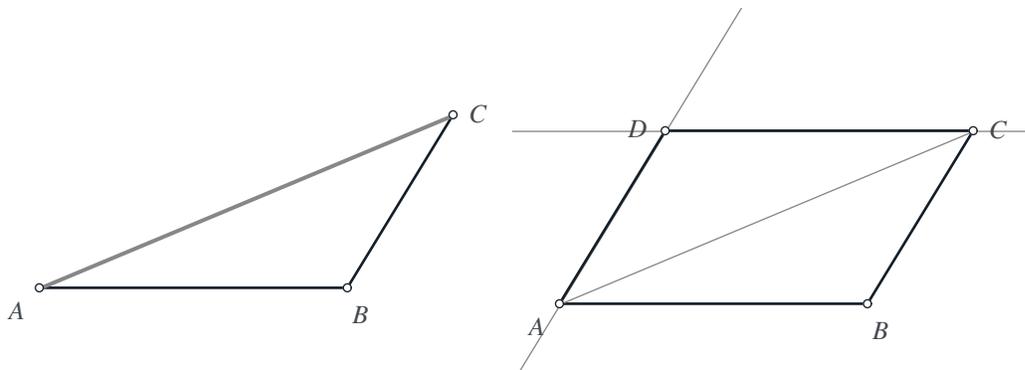
De la misma manera en que se hizo en la sección anterior para la función *Long* aplicada a segmentos, en esta sección podemos definir la función $Area(A)$ para una figura poligonal como *la clase de equivalencia* de todas las figuras poligonales A' tales que $A' \stackrel{g}{=} A$. Así, si A y B son dos figuras planas ajenas, la figura C que consta de ambas tiene área igual a la suma de las áreas de A y B . Es decir:

$$Area(A) + Area(B) = Area(C)$$

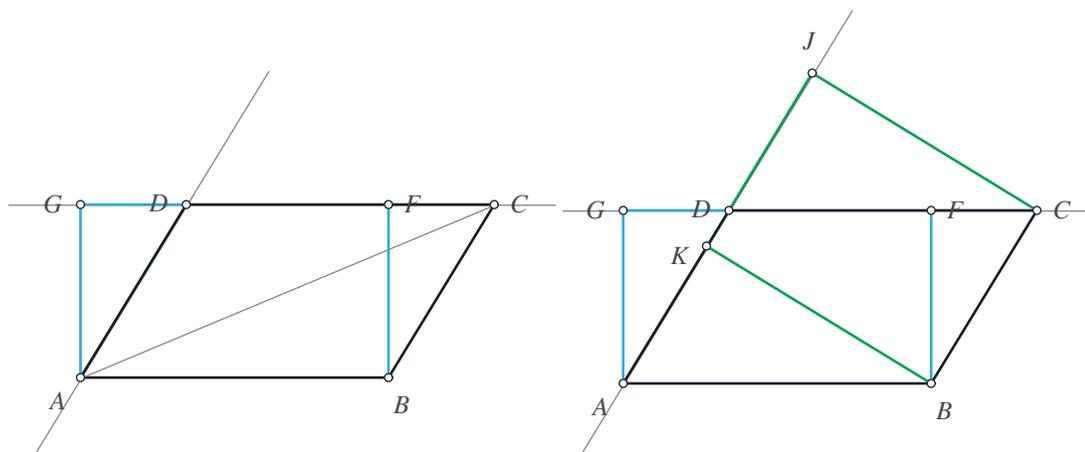
Nota: Igual que con los segmentos, es posible definir para las figuras poligonales una función numérica si se asume la existencia de los números reales y se adopta alguna figura poligonal como “la unidad”. En el caso de las áreas es costumbre usar como unidad un cuadrado cuyo lado es congruente con la unidad adoptada para medir longitudes.

Área del triángulo

¿Cuál es el área del triángulo ABC ? Podemos decir que es “la mitad” de la del paralelogramo $ABCD$. ¿Pero cuál es el área del paralelogramo $ABCD$?



Si F es la intersección de CD con su perpendicular por B , decimos que el área del paralelogramo $ABCD$ es $AB \cdot BF$. Lo justificamos porque si cortamos el rectángulo BCF y lo trasladamos a la posición ADG , como se indica en la figura, obtenemos un rectángulo de lados AB y BF , al cual asignamos $AB \cdot BF$ como área, por definición.



Pero igualmente podemos construir el rectángulo $BCJK$ y decir que el área del paralelogramo es $BC \cdot BK$. ¿Cuál es la buena? ¿Son iguales las áreas de $ABFG$ y $BCJK$? Los triángulos BCF y ABK son congruentes y esto implica que $\frac{BF}{BC} = \frac{BK}{AB}$, que es equivalente a $AB \cdot BF = BC \cdot BK$.²

Vemos que para tener una medida de área para triángulos y paralelogramos es necesario usar la ley de las proporciones de los triángulos semejantes. Pero también podríamos demostrar esta ley postulando la existencia de una medida del área de las figuras poligonales, la cual asigna como área a un rectángulo el producto de sus lados y asigna el mismo valor a dos figuras cuando una de ellas puede obtenerse de la otra cortando una parte y pegándola en otro sitio.

Esta situación, en la que podemos formular la teoría de áreas a partir de la semejanza de triángulos o la teoría de los triángulos semejantes a partir de la teoría de áreas, es típica de los modelos matemáticos de la física. Otro ejemplo, muy apreciado por los físicos, es la equivalencia de las formulaciones de la mecánica clásica a partir de las leyes de Newton o a partir de las leyes de conservación o a partir del principio de mínima acción.

Áreas de polígonos

Todo rectángulo puede transformarse en un cuadrado mediante cortes y traslaciones sin superposición. Este hecho, que demostraremos en la siguiente escena, permite demostrar que toda figura

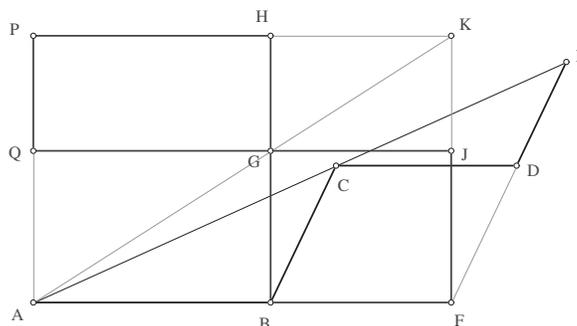
²

Recordemos la discusión presentada en la sección 3.2 relacionada con esta figura, en la que se exhibe la equivalencia entre la semejanza de triángulos y el concepto de área para rectángulos, lo cual ayuda a justificar la equivalencia de las igualdades:

$$\frac{AB}{CD} = \frac{BC}{DE}$$

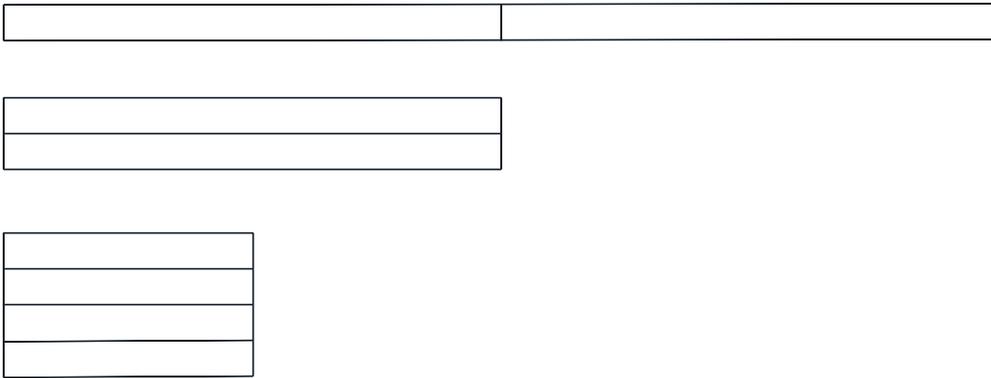
y

$$AB \cdot DE = BC \cdot CD$$

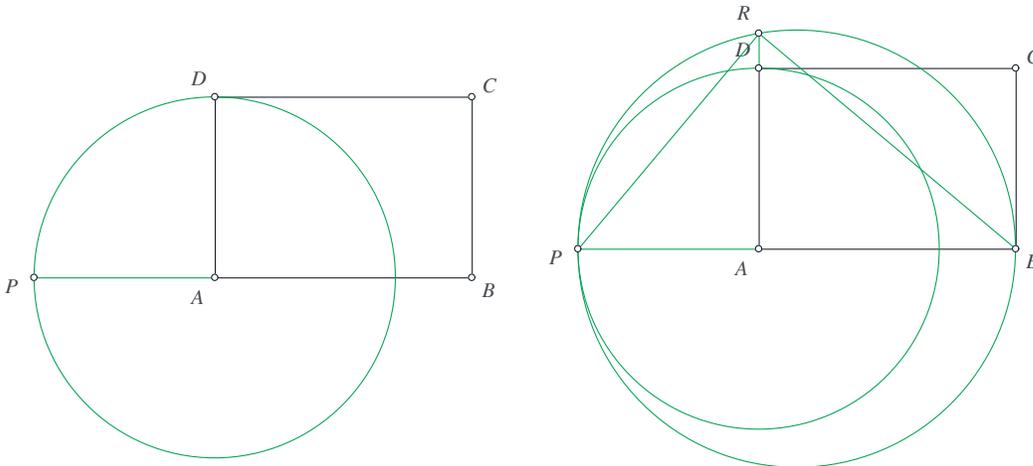


poligonal se puede transformar, mediante cortes y traslaciones sin superposición, en un cuadrado, con lo cual se le puede asignar como medida de su área la del cuadrado. Esto permite completar fácilmente la teoría de áreas de figuras poligonales planas.

La ilustración prueba este lema: Todo rectángulo se puede transformar, mediante cortes y traslaciones sin superposición, en otro cuyo lado más largo no sea mayor que el doble del otro. Si un rectángulo, como el de la figura, tiene un lado más largo que el doble del otro, podemos partirlo a la mitad por su lado más largo, tantas veces como sea necesario hasta llegar a uno en el que el lado más largo no sea mayor que el doble del otro. (En este paso se usa el axioma de Arquímedes o de continuidad.)

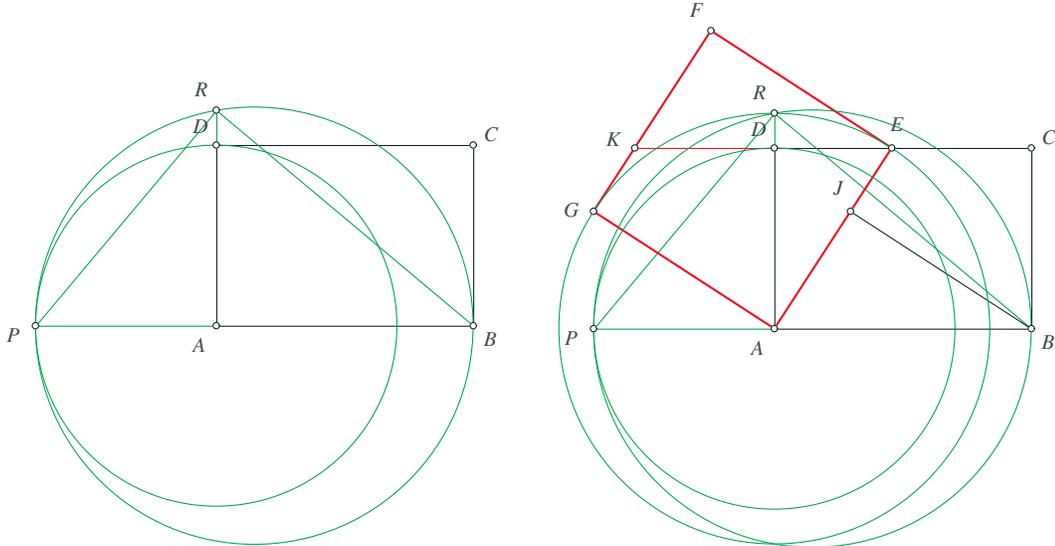


Pero con esta operación no podemos transformar un rectángulo arbitrario en un cuadrado. Para lograr esto hace falta un corte particular, como se muestra en las siguientes figuras.



Probaremos que un rectángulo cuyo lado más largo no es mayor que el doble del otro, se puede transformar en un cuadrado mediante cortes y traslaciones sin superposición. Sea $ABCD$ un rectángulo cuyo lado más largo es AB y éste es menor que el doble de BC . Sea P la intersección de AB con la circunferencia de centro A y radio $BC = AD$. Entonces $PA = BC$.

Sea R la intersección de AD con la circunferencia de diámetro PB . El ángulo PRB es recto (esto se demuestra en la unidad sobre el círculo) y los triángulos PAR y RAB son semejantes. Por lo tanto $\frac{PA}{AR} = \frac{AR}{AB}$, que equivale a $PA \cdot AB = AR^2$. Y como $PA = BC$, se tiene que $AB \cdot BC = AR^2$.



Sea E la intersección de DC con la circunferencia de centro A y radio AR . Entonces $AE = AR$ y el cuadrado de lado AE tiene la misma área que el rectángulo $ABCD$. Pero además $ABCD$ se puede transformar en el cuadrado $AEFG$. Para ello definimos J como la proyección de B sobre AE y K como la intersección de CD con FG . Trasladando el triángulo ABJ a la posición KEF y el cuadrilátero $JBCE$ a la posición $GADK$ se convierte $ABCD$ en $AEFG$.

Dejamos como ejercicio para el lector la demostración de que una figura poligonal se puede transformar en un cuadrado mediante cortes y traslaciones sin superposición. Para ello se sugiere considerar las siguientes proposiciones:

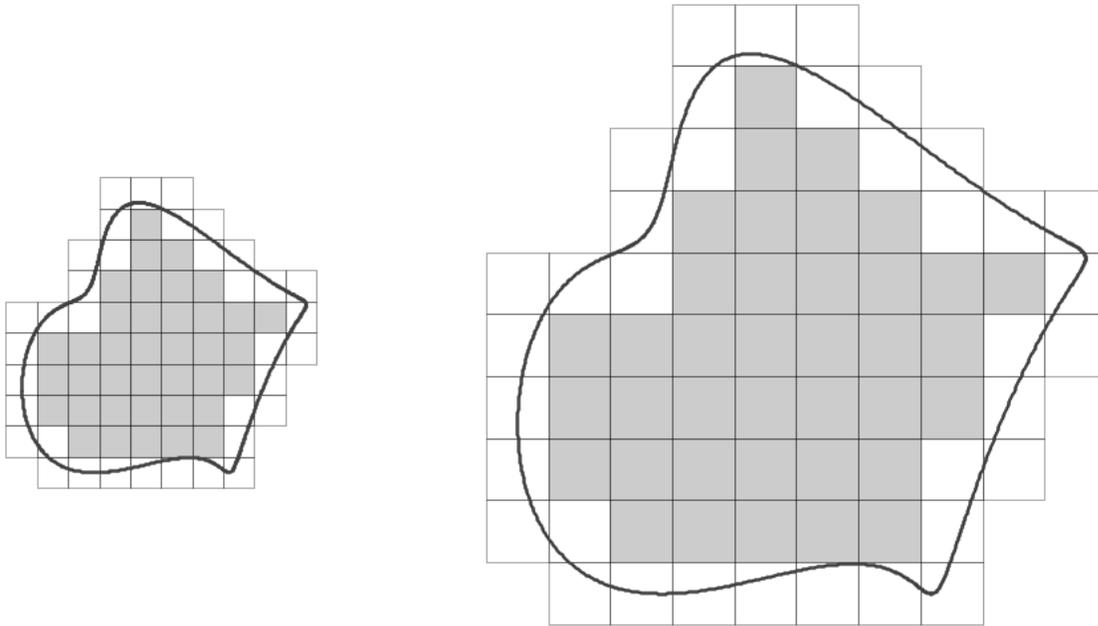
- 1) Toda figura poligonal se puede partir en triángulos.
- 2) Todo triángulo se puede transformar en un rectángulo.
- 3) Un cuadrado se puede transformar en un rectángulo con un lado dado.

5.3. El método de exhaustión

Áreas de otras figuras

Para obtener el área de una figura plana que no es poligonal, es decir, que su frontera puede incluir partes curvas, hace falta recurrir al método de exhaustión de Eudoxio que, a grandes rasgos, se puede describir así: el área de una figura plana se puede encontrar como el "límite" de las áreas de una sucesión de figuras poligonales contenidas en ella, si cualquier punto del interior de la figura cae en alguna de las figuras de la sucesión.

El resultado más importante con respecto a las áreas las figuras planas que no son poligonales, y que es fácil de demostrar siguiendo la idea sugerida en la siguiente figura,



es que si dos figuras planas F_1 y F_2 son semejantes y la razón de homotecia de (las dimensiones lineales de) F_2 a (las de) F_1 es 2, entonces

$$Area(F_2) = 4Area(F_1)$$

Este resultado puede hallarse en el libro XII de *Los elementos* de Euclides. Ambos cálculos se presentan más adelante en este mismo capítulo.

La demostración es una aplicación simple de la idea descrita en el método de exhaustión. Los resultados más importantes que se obtuvieron en la antigua Grecia utilizando el método de exhaustión son la *cuadratura de la parábola* y el cálculo del *área del círculo* y ambos deben su realización a Arquímedes.

Volumen

El volumen de un cuerpo se define de manera análoga al área de una figura plana.

Aquí hay que demostrar algunos teoremas sobre volúmenes de prismas. Por ejemplo:

- (a) Si las áreas de las bases de dos prismas son iguales y las alturas también son iguales, entonces los dos prismas tienen el mismo volumen.
- (b) Si dos prismas tienen alturas iguales y el área de la base de uno de ellos es el doble que la del otro, entonces el volumen del primero es el doble que el del segundo.
- (c) Si un prisma cuya base es un paralelogramo se corta en dos prismas triangulares mediante un

plano que pasa por una diagonal de la base y es paralelo a las aristas del prisma que no son de las bases, entonces los prismas resultantes tienen igual volumen y éste, por lo tanto, es la mitad del volumen del prisma original.

El resultado más importante que usaremos, y es fácil de demostrar siguiendo la idea equivalente al caso del área, es que:

(d) Si dos cuerpos C_1 y C_2 son semejantes y la razón de homotecia de (las dimensiones lineales de) C_2 a (las de) C_1 es 2, entonces

$$Vol(C_2) = 8Vol(C_1).$$

Armados con estos resultados se puede enfrentar uno de los problemas más importantes del cálculo de volúmenes de poliedros: el volumen de una pirámide, que puede reducirse al de un tetraedro. La conocida fórmula

$$\text{Volumen} = \frac{\text{área de la base} \times \text{altura}}{3} \quad (5.1)$$

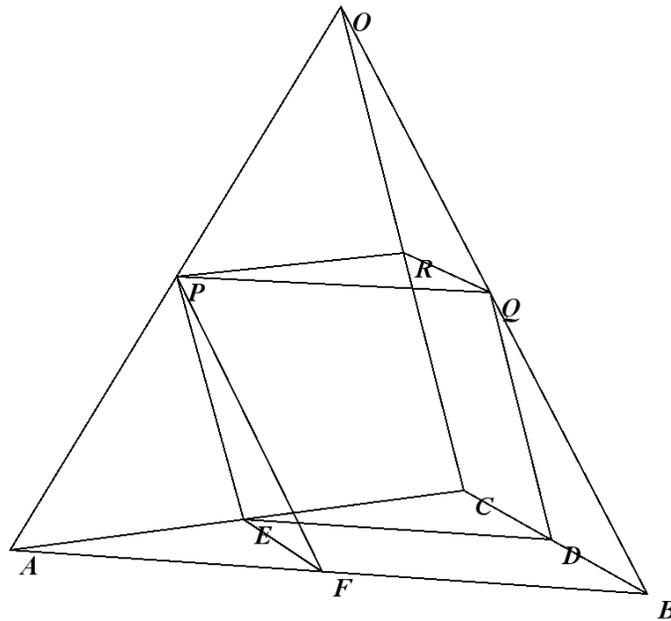
es análoga a la del área del triángulo: $\text{Área} = \frac{\text{base} \times \text{altura}}{2}$, pero es bastante más difícil de obtener. Sabemos que los egipcios la conocían, pero no hay ninguna prueba de que tuvieran alguna demostración de ella. A continuación presentamos su demostración, que se debe a Eudoxo de Cnido (390 - 337 aec) y aparece en el libro XII de Los elementos de Euclides, donde además se generaliza a las pirámides y al cono.

El volumen de un tetraedro

Sea $ABCO$ un tetraedro.

Sean D, E, F los puntos medios de BC, CA y AB , respectivamente.

Sean P, Q, R los puntos medios de OA, OB y OC , respectivamente.



Sea V el volumen del prisma con base triangular ABC y lado OA . Sea T el volumen del tetraedro $ABCO$. Demostraremos que $T = \frac{V}{3}$

El tetraedro puede partirse en cuatro cuerpos: dos tetraedros $AFFP$ y $PQRO$, semejantes al original con razón de homotecia $\frac{1}{2}$; y dos prismas triangulares cuyas bases son DBQ y EDC .

El volumen de cada uno de los tetraedros $AFFP$ y $PQRO$ es $\frac{T}{8}$ y el de cada uno de los prismas triangulares de bases DBQ y EDC es $\frac{V}{8}$. Por tanto:

$$T = 2\frac{T}{8} + 2\frac{V}{8}$$

de donde resulta que

$$T = \frac{V}{3}$$

Es importante reconocer que uno de los hechos más importantes que se utilizan en esta demostración es que el volumen de un cuerpo cuyas dimensiones lineales son el doble de las de otro, tiene un volumen igual a ocho veces el del segundo (que probablemente aprendieron los griegos en relación con el famoso problema de duplicar el cubo) y esto se demuestra usando el método de exhaustión.

Nota: En realidad Eudoxio demostró la fórmula $V = \frac{abc}{6}$ para un tetraedro rectangular utilizando el método de exhaustión y un *proceso recursivo*. Es la demostración que puede hallarse en *Los elementos* de Euclides. La que presentamos aquí no requiere del proceso recursivo pero sí del método de exhaustión. Sin embargo, el proceso recursivo que utilizó Eudoxo probablemente inspiró a Arquímedes para resolver el problema de la cuadratura de la parábola, como se ve más adelante.

A partir de este resultado no es difícil demostrar la fórmula 5.1:

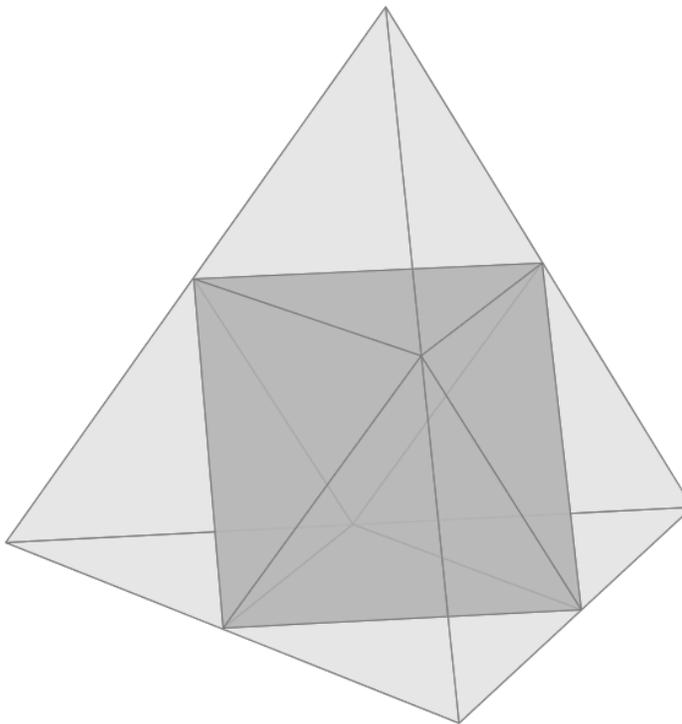
$$\text{Volumen} = \frac{\text{área de la base} \times \text{altura}}{3}$$

para el volumen de cualquier pirámide con base poligonal.

Ejercicio. Demostrar la fórmula 5.1 para el volumen de una pirámide y calcular los volúmenes de un tetraedro, un octaedro y un icosaedro regulares, descomponiendo estos poliedros en pirámides y hallando en cada caso las áreas de sus bases y sus alturas.

Volúmenes de poliedros

Ejemplo. El volumen del octaedro inscrito en un tetraedro, con lados la mitad de los del tetraedro, tiene la mitad de volumen porque resulta de quitarle al tetraedro cuatro tetraedros iguales, cada uno con la octava parte de volumen del original.



Este resultado es válido tanto si el tetraedro inicial es regular como si es irregular. Lamentablemente, este procedimiento para obtener volúmenes de poliedros no es muy general. A diferencia del caso plano, no siempre es posible descomponer un poliedro en partes que puedan recomponerse en un cubo. Este resultado apenas se demostró en el siglo XX. Fue el tercer problema de los que planteó David Hilbert en el congreso de matemáticos del año 1900.

El tercer problema de Hilbert

En el año 1900 David Hilbert presentó a la comunidad matemática 23 problemas que aún no se habían podido resolver para esa fecha. Uno de ellos, el tercero, era la pregunta de si con los poliedros sucedería lo mismo que con los polígonos en el sentido de que toda figura poligonal se puede partir en otras las cuales puedan recomponerse en un cuadrado. Es decir, ¿es cierto que todo cuerpo en el espacio formado por poliedros se puede partir en otros que puedan recomponerse en un cubo?

La formulación original del problema era ésta:

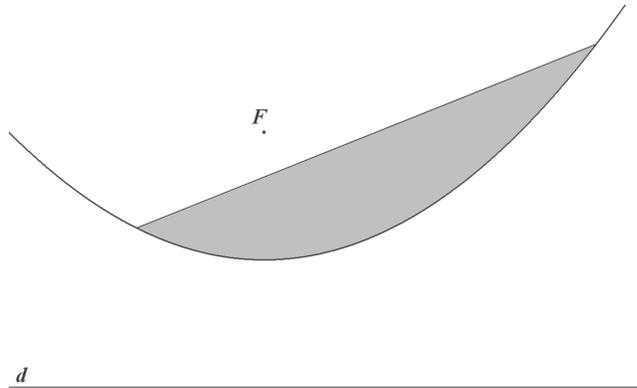
Tercer problema de Hilbert. Dados dos poliedros de igual volumen, ¿es posible cortar el primero en una cantidad finita de piezas poliédricas que puedan ser ensambladas de modo que quede armado el segundo?

Éste fue el primero de los 23 problemas que se resolvió. En 1903 MAX DEHN (1878-1952) encontró un invariante ante seccionamientos de poliedros que no es el mismo, por ejemplo, para un tetraedro regular que para un cubo. Por tanto, por mucho que partamos el tetraedro, sus partes tendrán siempre el mismo valor del invariante esas partes no podrán juntarse para formar un cubo.

El invariante de Dehn es lo que se llama el *producto tensorial* de las longitudes de las aristas por el ángulo definido por las caras que se unen en cada arista.

5.4. La cuadratura de la parábola

Sin duda la aplicación estelar del método de exhaustión de Eudoxio la realizó Arquímedes al calcular el área de un sector de parábola. Un *sector de parábola* es la región acotada por una parábola y una recta que la corta en dos puntos.



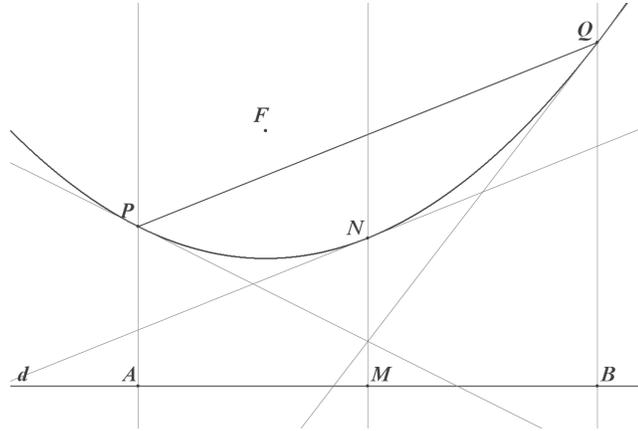
El resultado de Arquímedes puede expresarse de manera muy simple y general:

El área de un sector de parábola es $\frac{2}{3}$ del mínimo paralelogramo que lo contiene.

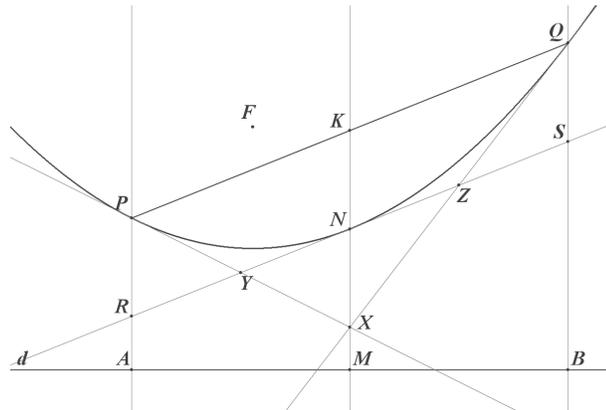
Arquímedes produjo al menos dos demostraciones de este resultado. Aquí nos ocuparemos únicamente de la que se apoya en el método de exhaustión. Esta demostración se basa en una propiedad de la parábola que describimos a continuación. Sea F el foco de una parábola y sea d su directriz. Sean P y Q dos puntos sobre la parábola y sea N el punto sobre la parábola cuya proyección

ortogonal M sobre la directriz d es el punto medio de las proyecciones ortogonales A y B sobre d de P y Q , respectivamente. Entonces:

La tangente a la parábola en N es paralela a PQ



Para demostrar esta propiedad definimos X como el punto de intersección de las tangentes a la parábola en P y Q . Sabemos que X está sobre MN porque su distancia a A es la misma que a B , ya que las tangentes a las parábolas en P y Q son las mediatrices de FA y FB , respectivamente. Esta última es una propiedad básica de las tangentes a una parábola que se el lector no, conviene que la demuestre como **ejercicio**.

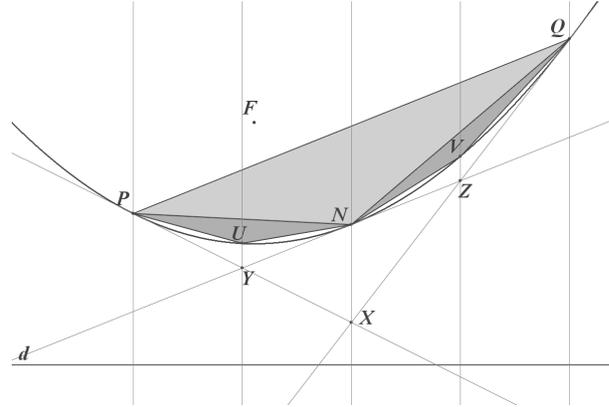


Utilizando el mismo argumento puede verse que el punto Y , definido como la intersección de las tangentes a la parábola en P y N , también se encuentra a la mitad entre las paralelas AP y MN . Por tanto los triángulos PRY y XNY son congruentes y $PR = NX$ – donde R y S son los puntos de intersección de la tangente a la parábola en N con AP y BQ , respectivamente. Análogamente, $NX = QS$. Por lo tanto $PR = QS$. Esto completa la demostración de que la tangente en P es paralela a PQ .

De hecho, como $PR = KN = QS$, donde K es el punto medio de PQ , resulta que N es el punto medio de KX y por lo tanto el área del triángulo PNQ es la mitad de la del triángulo PXQ . También es claro que el área del triángulo PNQ es la mitad de la del paralelogramo $PRSQ$.

Arquímedes demostró que el área del sector de parábola PNQ es igual a $\frac{4}{3}$ del área del triángulo PNQ , o lo que es igual, $\frac{2}{3}$ del área del paralelogramo $PRSQ$.

La demostración de Arquímedes procede por exhaustión, agregando primero el triángulo PNQ y luego dos triángulos PUN y NVQ donde U y V son los puntos sobre la parábola que están también sobre las perpendiculares a d en los puntos medios de AM y MB , respectivamente. Y así se continúa dividiendo cada uno de los sectores de parábola en dos, y agregando sus triángulos inscritos.



El área a_0 del triángulo PNQ se puede calcular explícitamente:

$$a_0 = \frac{AB \cdot KN}{2}$$

Las áreas de los dos triángulos PUN y NVQ , sumadas, tienen el valor:

$$a_1 = 2 \frac{\frac{AB}{2} \cdot \frac{NK}{4}}{2} = \frac{AB \cdot NK}{8} = \frac{a_0}{4}$$

Procediendo por iteración, los cuatro triángulos siguientes tienen un área conjunta:

$$a_2 = 4 \frac{\frac{AB}{4} \cdot \frac{NK}{4^2}}{2} = \frac{AB \cdot NK}{2 \cdot 4^2} = \frac{a_0}{4^2}$$

En general:

$$a_n = 2n \frac{\frac{AB}{2^n} \cdot \frac{NK}{4^n}}{2} = \frac{AB \cdot NK}{2 \cdot 4^n} = \frac{a_0}{4^n}$$

Así, agragando todos estos términos, llegamos a una serie infinita para expresar el área s del sector de parábola PNQ :

$$s = a_0 + a_1 + a_2 + \dots = \left(1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} \dots\right) a_0 = t \cdot a_0$$

donde

$$t = 1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} \dots$$

Para calcular esta suma infinita observamos que

$$t = 1 + \frac{1}{4} \left(1 + \frac{1}{4} + \frac{1}{4^2} + \dots\right) = 1 + \frac{1}{4} t$$

Por tanto, despejando t de esta sencilla ecuación, obtenemos $t = \frac{4}{3}$.

Sustituyendo en $s = t \cdot a_0$ obtenemos:

$$s = \frac{4}{3}a_0 = \frac{4}{3} \frac{AB \cdot NK}{2} = \frac{2}{3}AB \cdot NK$$

que es lo que queríamos demostrar.

5.5. El número π a través de los siglos

Una de las contribuciones más notables de Arquímedes al pensamiento matemático fue su manera de estimar el valor del número π , o más precisamente, de calcular el perímetro de la circunferencia. Para hacer justicia a este trabajo vamos a presentarlo enmarcado en un relato detallado de las distintas maneras en que se ha enfrentado este problema antes y después de Arquímedes, desde las primeras fórmulas usadas en Babilonia y Egipto hasta los trabajos que probaron que π no es un número racional y los métodos que se usan en la actualidad para estimar con muchísima precisión su valor. En toda esta historia la intervención de Arquímedes es fundamental y marca un hito en la historia del pensamiento matemático porque muestra cómo realizar estimaciones tan precisas como uno necesite, de cantidades que no pueden calcularse con total exactitud. En el fondo el trabajo de Arquímedes respecto al número π marca el nacimiento de lo que hoy llamamos Análisis Matemático.

El número π en la antigüedad

Las culturas antiguas sabían que el perímetro de una circunferencia es un poco mayor que el triple de su diámetro y que el área de un círculo es algo mayor que tres veces la del cuadrado de lado igual al radio. Está más o menos claro que sabían que esas razones eran constantes, pero no que supieran que son iguales. De hecho las estimaciones que usaban para una y otra no siempre eran las mismas.

Los babilonios usaban unas veces 3 y otras $3\frac{1}{8} = 3.125$. No está claro si uno de estos valores era para la razón de los perímetros y el otro para la de las áreas o si usaban 3 en los ejercicios escolares y $3\frac{1}{8}$ cuando querían mayor precisión. No se sabe cómo obtuvieron esas aproximaciones. Nada hay que indique un intento para justificarlas. Es probable que las obtuvieran experimentalmente (por ejemplo rodeando una columna cilíndrica con una cuerda o midiendo las marcas que deja una protuberancia en el perímetro de una rueda).

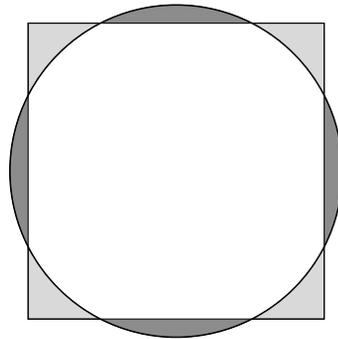
El número π entendido como *la razón entre el perímetro y el diámetro de una circunferencia* o como *la razón entre el área de un círculo y el cuadrado de su radio*, no aparece en las primeras civilizaciones, pero en cambio casi todas ellas conocían fórmulas para calcular, de manera aproximada, los perímetros y las áreas de los círculos. Un ejemplo es la fórmula

$$A = (8/9d)^2$$

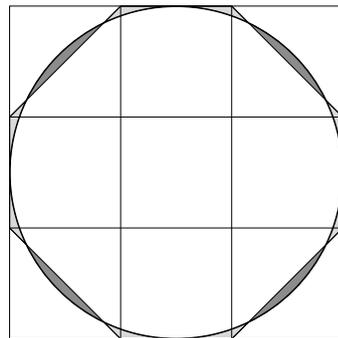
que usaban los egipcios para el área de un círculo de diámetro d . Como puede apreciarse por su aspecto, esta fórmula es más bien una “cuadratura del círculo” (solo aproximada, por supuesto), es decir, exhibe un cuadrado (de lado $8/9 d$) cuya área es más o menos igual a la del círculo. Esta

fórmula es una muestra clara de lo que interesaba a los egipcios: tener un método simple de calcular, de manera aproximada, el área de un círculo. Muchos otros ejemplos de los ítems encontrados en los papiros egipcios que datan de más de 1500 años A.C. Muestran que su interés era tener procedimientos o fórmulas para calcular algo. Se desconoce si el procedimiento era producto de un razonamiento o fue encontrado de manera empírica y su justificación residía únicamente en que funcionaba bien en la práctica. Justificar procedimientos de cálculo por su utilidad y no por su racionalidad, ha subsistido hasta nuestros días y constituye una de las maneras no ortodoxa de hacer matemáticas y que muchos justifican por los beneficios prácticos que aporta.

En su afán por justificar racionalmente las fórmulas de la antigüedad, los historiadores de las matemáticas han inventado a veces argumentos factibles que hubieran podido haber usado quienes las inventaron. Por ejemplo, la figura muestra un círculo con diámetro d y un cuadrado con lado $\frac{8}{9}d$. A simple vista puede apreciarse que la diferencia entre sus áreas es pequeña. El exceso en las esquinas del cuadrado parece compensarse con los sectores del círculo que le sobresalen. De acuerdo con esto, el área del círculo (πr^2 donde $r = \frac{d}{2}$) es “casi” igual a $(\frac{8}{9}d)^2$. Tomando $d = 2$ (o sea $r = 1$) obtenemos una estimación para π de $(\frac{8}{9}2)^2 = 256/81$, que en decimales nos da $\pi \simeq 3.16$.



El razonamiento es éste: si se parte el cuadrado circunscrito al círculo de diámetro d en nueve cuadrados, cada uno de lado $\frac{1}{3}d$, como muestra la figura.



Salta a la vista que el octágono irregular en la figura debe tener un área muy parecida a la del círculo. Los 8 excesos alrededor de sus vértices parecen compensarse con los 4 de la circunferencia sobre los lados diagonales del hexágono. El área del octágono es igual a 7 veces la de los cuadrados

de lado $\frac{d}{3}$, o sea

$$7\left(\frac{d}{3}\right)^2 = \frac{7}{9}d^2 = \frac{63}{81}d^2$$

que es un valor muy cercano a $\frac{64}{81}d^2$. Pero $\frac{64}{81}$ tiene la ventaja de ser el cuadrado exacto de $\frac{8}{9}$ y proporciona una *cuadratura* aproximada del círculo que ayuda a simplificar los cálculos.

Que la razón entre la circunferencia y el diámetro es la misma para todos los círculos parecía ser conocimiento común en Babilonia y Egipto. La primera demostración documentada de que la razón del área del círculo al cuadrado de su radio es constante corresponde a Euclides en el libro XII de sus Elementos. Sin embargo Euclides no proporciona el valor de dicha razón. Y es que esta constante fundamental de la naturaleza que denotamos por π es difícil de calcular, además de que la notación numérica de los griegos no fomentaba el análisis numérico ³.

El tratamiento de Arquímedes del número π

La primera estimación razonada de π , con cotas bien definidas, la realizó Arquímedes (Siracusa, Sicilia 287 – 212 a.e.c.). Arquímedes fue a la vez físico, ingeniero, inventor, astrónomo y, sobre todo, matemático. Se le considera uno de los científicos más importante de la antigüedad clásica. Pero por lo que más se le admira es por sus trabajos matemáticos, que son los únicos que él mismo consideró suficientemente importantes como para escribirlos y legarlos a la posteridad.



«Domenico-Fetti Archimedes 1620» de Domenico Fetti.

Arquímedes murió durante el sitio de Siracusa (214–212 a.e.c.), ejecutado por un soldado romano. Se dice que el soldado le ordenó levantarse y seguirlo, pero Arquímedes se mantuvo absorto en sus pensamientos dibujando en la arena, lo cual irritó al soldado que, enfadado, le asestó un golpe de espada, a pesar de que tenía órdenes de no hacerle daño.

³Los primeros en usar la letra π para denotar esta constante fueron los matemáticos ingleses Oughtred (1647), ISAAC BARROW (1664) y DAVID GREGORY (1697).



«Cara de la medalla Fields, Archimedes» de Stefan Zachow.

Sus ideas, sus métodos y el rigor de sus demostraciones siguen siendo hoy motivo de asombro y admiración, la cual ha quedado patente en el hecho de que la medalla Fields, que es el premio más importante de las matemáticas, lleva su efigie.

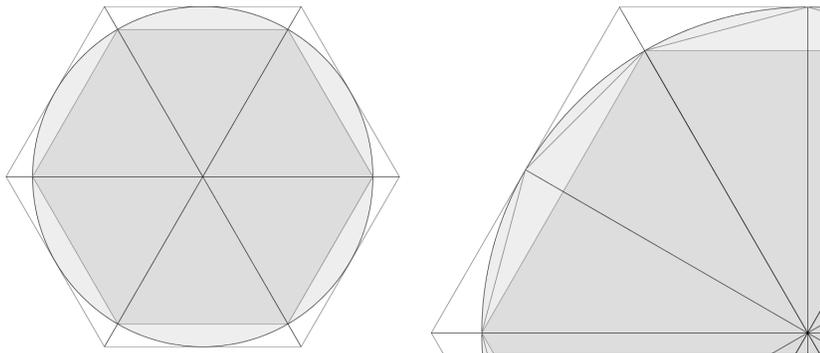
Si p_n y P_n representan los perímetros de los polígonos regulares inscrito y circunscrito de $6 \cdot 2^n$ lados, entonces, según Arquímedes

$$\frac{p_n}{2} < \pi < \frac{P_n}{2}$$

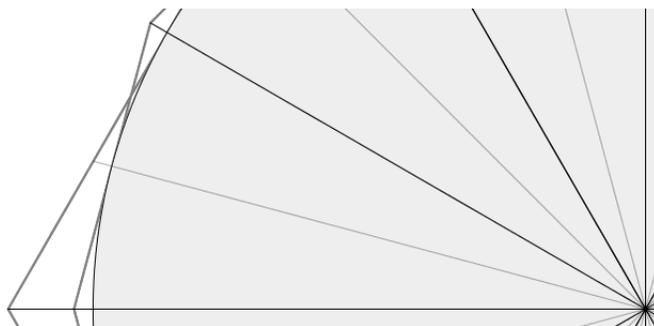
Esto nos dice que el número π se encuentra en todos los intervalos

$$\left[\frac{p_n}{2}, \frac{P_n}{2} \right]$$

Además Arquímedes da argumentos geométricos para demostrar que $\{p_n\}$ es creciente y $\{P_n\}$ es decreciente. Lo primero es bastante evidente porque dos segmentos consecutivos del polígono inscrito de $6 \cdot 2^{n+1}$ lados une los extremos de un solo lado del de $6 \cdot 2^n$ lados.



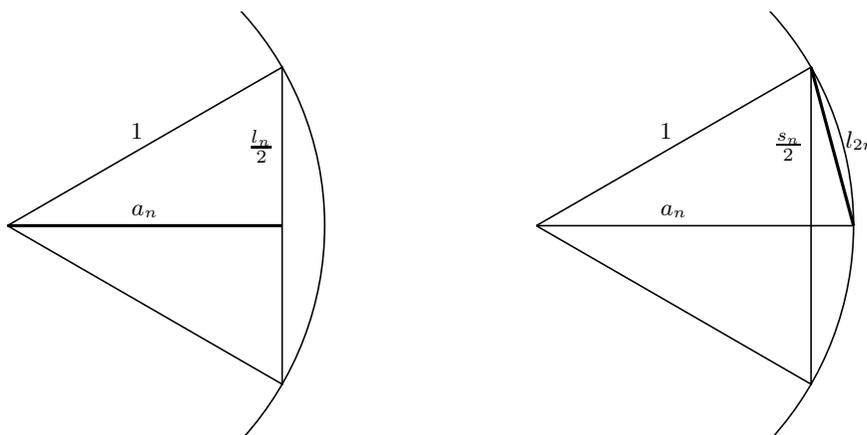
Para lo segundo basta observar que en la siguiente figura



el lado del polígono anterior puede partirse en dos partes, una de ellas igual al lado del polígono siguiente y la otra claramente mayor por poder convertirse en el lado opuesto a un ángulo obtuso en un triángulo en el que uno de los lados adyacentes al ángulo obtuso es precisamente el lado del polígono anterior.

Si probáramos que la longitud de esos intervalos tiende a cero cuando n tiende a infinito, obtendríamos una manera de estimar el valor de π con tanta precisión como quisiéramos. Para ellos bastaría tomar polígonos con un número de lados n suficientemente grande. Para lograr esto Arquímedes calculó explícitamente esos perímetros para los polígonos regulares inscritos y circunscritos de $6 \cdot 2^n$ lados y demostró que su diferencia tiende a cero cuando n tiende a infinito.

Sea l_n el lado del polígono inscrito de $6 \cdot 2^n$ lados y a_n el apotema correspondiente. Entonces, usando el Teorema de Pitágoras, obtenemos



$$a_n = \sqrt{1 - \left(\frac{l_n}{2}\right)^2} \quad (5.2)$$

$$l_{n+1} = \sqrt{\left(\frac{l_n}{2}\right)^2 + (1 - a_n)^2} \quad (5.3)$$

El lado L_n del polígono circunscrito de $6 \cdot 2^n$ lados puede hallarse por triángulos semejantes a partir de l_n y a_n :

$$L_n = \frac{l_n}{a_n} \quad (5.4)$$

Sumando la mitad de los lados de los polígonos, Arquímedes obtiene estas cotas para π que permiten estimar su valor con la precisión que uno quiera

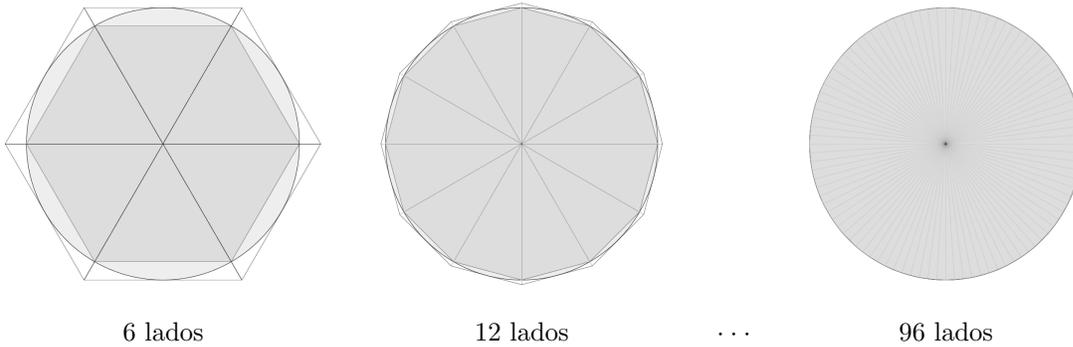
$$\frac{p_n}{2} = 3 \cdot 2^n \cdot l_n < \pi < 3 \cdot 2^n \cdot L_n = \frac{P_n}{2}$$

ya que la diferencia entre estos valores es

$$3 \cdot 2^n (L_n - l_n) = 3 \cdot 2^n \cdot L_n (1 - a_n)$$

que tiende a cero cuando n tiende a infinito porque $3 \cdot 2^n \cdot L_n$ está acotado (por ejemplo por $3 \cdot 2^1 \cdot L_1$ que es el semi-perímetro del hexágono circunscrito) y obviamente a_n tiende a 1 cuando n tiende a infinito.

Las fórmulas recursivas (5.2) y (5.3) involucran raíces cuadradas, que son difíciles de calcular manualmente, y aún con los excelentes algoritmos de nuestro sistema de decimal, hoy no haríamos estos cálculos sin contar con una calculadora. Las raíces cuadradas de números racionales suelen ser irracionales y sólo podemos expresarlas de manera aproximada en nuestro sistema decimal.



A pesar de estas dificultades, Arquímedes, con algoritmos aritméticos menos eficientes que los nuestros, logró obtener buenas cotas racionales para el número π . Usando los polígonos inscrito y circunscrito de 96 lados obtuvo

$$3 + \frac{10}{71} = \frac{223}{71} < \pi < \frac{220}{70} = 3 + \frac{10}{70}$$

Estas cotas dan 2 cifras decimales de precisión

$$3.1408 < \pi < 3.1429 \quad (5.5)$$

Con una calculadora hoy podemos obtener

$$3.1410 < \pi < 3.1428 \quad (5.6)$$

Observemos que el promedio pesado con $\frac{2}{3}$ de la cota inferior y $\frac{1}{3}$ de la superior es 3.1416, que es una excelente aproximación para π con cuatro cifras decimales de precisión. Esto no es una casualidad, como veremos más adelante.

El método de Arquímedes fue aplicado por Ptolomeo alrededor del año 100, hasta alcanzar un polígono de $1536 = 6 \cdot 2^8$ lados, obteniendo la estimación $\pi \simeq 317/120 \simeq 3.1416$ que es el valor que usamos todavía en las escuelas.

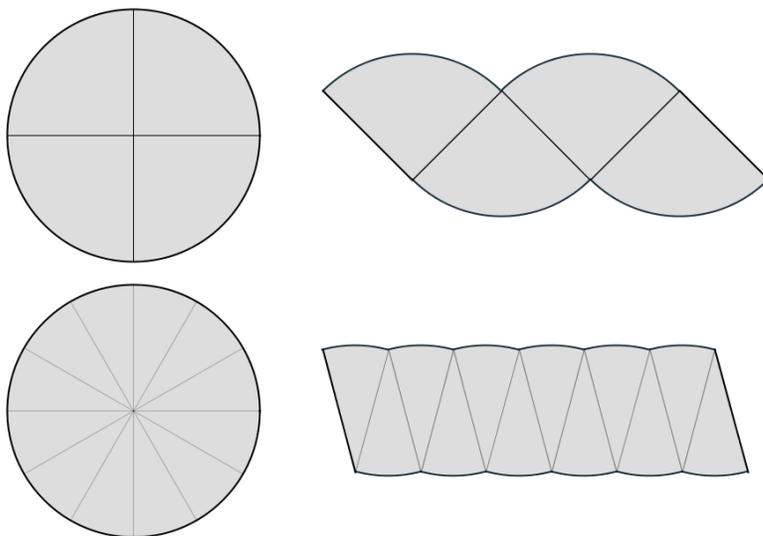
Algunos siglos más tarde y de manera independiente, los chinos usaron el mismo método para obtener una aproximación aún más precisa $\pi \simeq 355/113$, que en decimales corresponde a $\pi \simeq 3.141592$ y es exacta en seis cifras decimales. Este resultado constituye una hazaña extraordinaria de cálculo, dadas las limitaciones aritméticas de aquella época, y se puede considerar que dio a los chinos el récord mundial durante mil años. No fue sino hasta finales de la edad media europea que el uso de los números arábigos permitió mejorar los métodos aritméticos y obtener mejores estimaciones de π . Como veremos más adelante, Arquímedes pudo haber obtenido la misma aproximación que los chinos y con casi las mismas operaciones aritméticas con que logró la suya, que fue mucho más modesta.

El área del círculo

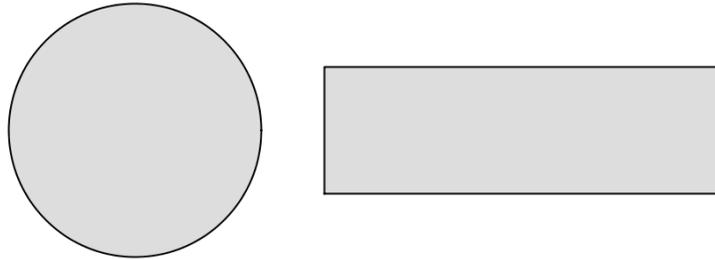
Euclides demostró que la razón entre el perímetro de la circunferencia y su diámetro es constante y también que la razón entre el área del círculo y el cuadrado de su radio es constante. Sin embargo no demostró que ambas constantes son iguales. Quien logra demostrarlo es, para no variar, Arquímedes. Lo hace unos 50 años después de que Euclides escribiera sus *Elementos*. Este resultado es el que nos permite escribir las fórmulas $p = 2\pi r$ para el perímetro de la circunferencia y $A = \pi r^2$ para el área del círculo de radio r .

Las siguientes figuras ilustran el método que usó Arquímedes para obtener el área de un círculo en términos del número π , que de hecho consiste precisamente en encontrar la relación entre el área, el perímetro y el radio:

$$A = \frac{p \cdot r}{2}.$$



Lo que hizo fue dividir el círculo en un número par de sectores y acomodar la mitad de ellos alineados y paralelos, tocándose por un punto, y la otra mitad de la misma manera, pero invertidos, encajados exactamente en los primeros. Luego fue aumentando el número de sectores y observó que la figura que se obtiene, y que por construcción siempre tiene la misma área que el círculo, se va convirtiendo en un rectángulo cuyo lado es la mitad del perímetro y cuya altura es el radio.



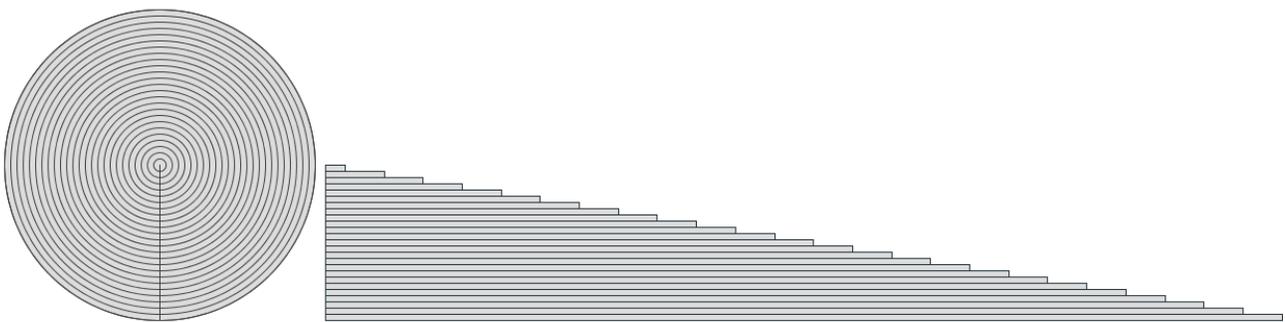
Así, resulta claro que el área del círculo es la mitad del perímetro por el radio. Dado que el perímetro, por definición del número π , es $2\pi r$, esto demuestra que el área del círculo es πr^2 . Cabe mencionar que JOHANNES KEPLER (1571 - 1630) utilizó una idea similar para demostrar que el volumen V y la superficie S de una esfera de radio r guardan la relación

$$V = \frac{S \cdot r}{3}.$$

Sugerimos al lector demostrar esta fórmula mediante la idea de partir la superficie de la esfera en muchos pequeños triángulos e interpretarlos como las bases de pirámides con un vértice en el centro de la esfera y cuyas alturas se aproximan al radio de la esfera a medida que los lados de los triángulos tienden a cero.

Otra manera de demostrar la fórmula para el área del círculo de radio r es partiéndolo en muchos anillos concéntricos del mismo espesor y desenrollarlos, sin estirarlos ni encogerlos, hasta dejarlos rectos y horizontales. Se forma así una figura muy cercana al triángulo de base $2\pi r$ y altura r , cuya área es, por supuesto, πr^2 . Cuantos más anillos se usen, mejor es la aproximación.

Así se consigue ontiene demostración de que el área del círculo de radio r es πr^2 .



Este argumento aparece en un texto medieval del matemático y astrónomo judío-catalán Abraham Bar Hiyya (o Abraham Iudaeus Savasorda) (Barcelona, siglo XII, 1065-1070 – 1136 ó 1145), precursor de la escuela de Toledo.



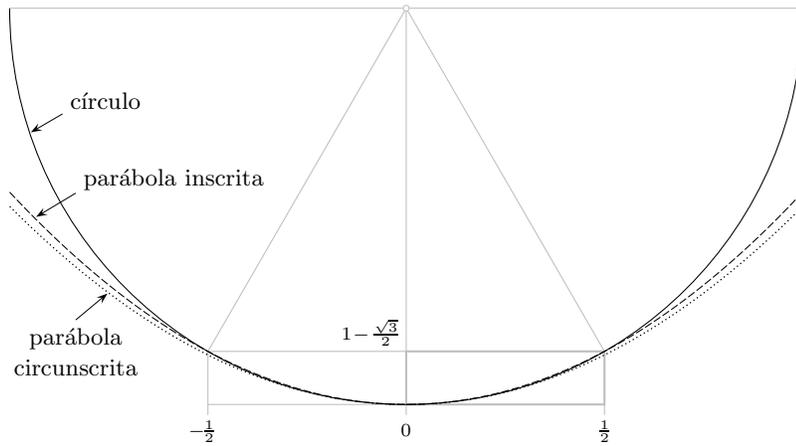
«Sefer tzuret, abraham bar»

Photo of Exhibit at the Diaspora Museum, Tel Aviv - en:Beit Hatefutsot.

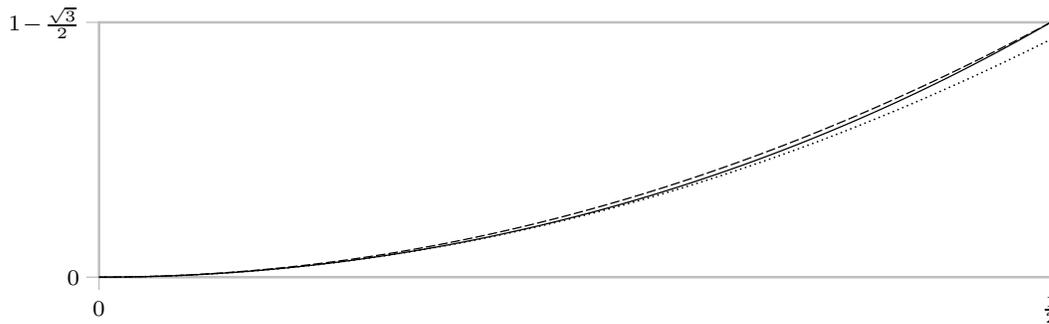
Ya que que π puede obtenerse como la razón del área del círculo al cuadrado de su radio, resulta que es igual al área del círculo unitario (el de radio 1). Así calcularemos π de aquí en adelante.

El número π como el área del círculo unitario

Usando la misma idea de Arquímedes de los polígonos de $6 \cdot 2^n$ lados, pero ahora buscando cotas para sus áreas en lugar de sus perímetros, se llega a estimaciones similares a las que se obtienen con perímetros. Pero usar áreas tiene ventajas. Nos permite aprovechar otro de los importantes resultados de Arquímedes: *la cuadratura de la parábola* (ver 5.4), es decir, que el área de un sector de parábola es igual a $\frac{2}{3}$ de la del paralelogramo o rectángulo más pequeño que la contiene. El promedio pesado con $\frac{2}{3}$ de la cota inferior más $\frac{1}{3}$ de la exterior, que como antes vimos da una estimación muy buena de π , nos sugiere usar parábolas para obtener mejores resultados.



Consideremos dos parábolas con sus ejes sobre la mediatriz del triángulo formado por el centro del círculo y los extremos de uno de los hexágonos inscritos, ambas con vértice en la intersección de la mediatriz con la circunferencia. Sea la primera de ellas la que corta a la circunferencia en los extremos del lado elegido del hexágono y sea la segunda la que tiene el foco en el punto medio del radio que coincide con la mediatriz. Se puede demostrar que la segunda es exterior al círculo y la primera es interior, entre los extremos del lado del hexágono. Las tres curvas coinciden en el vértice de las parábolas.



Se puede obtener una cota inferior para el área del círculo unitario como 6 veces el área del triángulo de base 1 y apotema $a = \sqrt{1 - (\frac{1}{2})^2} = \frac{\sqrt{3}}{2}$ más $\frac{2}{3}$ del rectángulo de base 1 y altura $1 - a$. O sea que

$$\pi > 6\left(\frac{a}{2} + \frac{2}{3}(1 - a)\right) = 6\left(\frac{\sqrt{3}}{4} + \frac{2}{3}\left(1 - \frac{\sqrt{3}}{2}\right)\right) > 3.141104 \quad (5.7)$$

Si colocamos el origen en el vértice de las parábolas y hacemos coincidir el eje y con los ejes de las parábolas de manera que el centro del círculo quede en el punto $(0, 1)$, la ecuación de nuestra parábola exterior resulta ser $y = \frac{x^2}{2}$. Sumando las áreas del triángulo y el rectángulo y restando la que está bajo la parábola, obtenemos

$$\pi < 6\left(\frac{a}{2} + (1 - a) - \frac{1}{3}\left(\frac{0.5^2}{2}\right)\right) = 6\left(\frac{\sqrt{3}}{4} + \left(1 - \frac{\sqrt{3}}{2}\right) - \frac{1}{3}\left(\frac{0.5^2}{2}\right)\right) < 3.142603 \quad (5.8)$$

Reuniendo (5.7) y (5.8) llegamos a

$$3.141104 < \pi < 3.142603$$

Esta estimación es ligeramente mejor que la de Arquímedes correspondiente al polígono de 96 lados, pero ésta se obtuvo con muchas menos operaciones aritméticas y, entre ellas, sólo una raíz cuadrada, la de 3.

La misma idea se puede aplicar a los polígonos de $6 \cdot 2^n$ lados aprovechando las fórmulas recursivas (5.2) y (5.3) para a_n y l_n que se usaron en el método de Arquímedes. Con esos valores las cotas para π son

$$6 \cdot 2^n \left(\frac{a_n}{2} + \frac{2}{3} \cdot \frac{l_n^2}{2} \right) < \pi < 6 \cdot 2^n \left(\frac{a_n}{2} + (1 - a_n) - \frac{1}{3} \cdot \frac{l_n^2}{2} \right)$$

En el caso $n = 4$, que corresponde al polígono de 96 lados, esto da 7 dígitos de precisión

$$3.14159264 < \pi < 3.14159267$$

que son 5 más que los 2 que dio el método de Arquímedes.

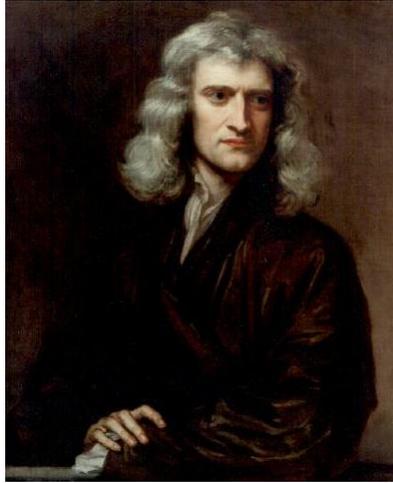
Resulta bastante extraño que Arquímedes no haya usado esta idea, siendo que él mismo obtuvo la cuadratura de la parábola. Tal vez su trabajo sobre la cuadratura de la parábola haya sido posterior al del cálculo de π .

Aunque este camino nos permite calcular muchos decimales de π , tiene el gran inconveniente de las raíces cuadradas, que nos obligan a realizar operaciones complejas con resultados inexactos.

Si se hacen promedios pesados con $\frac{3}{5}$ de las cotas inferiores y $\frac{2}{5}$ de las superiores, se obtienen aproximaciones con un 50% más de cifras decimales correctas. Esto sugiere que podría haber métodos aún mejores. Uno de ellos lo descubrió Isaac Newton. En la siguiente sección lo redescubriremos nosotros mismos usando argumentos geométricos y del cálculo diferencial e integral.

Cálculo de π según Newton

El gran científico Sir Isaac Newton (1643 - 1727) encontró un método para calcular π que es mucho más eficiente que cualquiera de los que le precedieron.



«Sir Isaac Newton (1643-1727)» de Sir Godfrey Kneller
 - <http://www.phys.uu.nl/vgent/astrology/images/newton1689.jpg>.
 Disponible bajo la licencia Dominio público via Wikimedia Commons -
[https://commons.wikimedia.org/wiki/File:Sir_Isaac_Newton_\(1643-1727\).jpg](https://commons.wikimedia.org/wiki/File:Sir_Isaac_Newton_(1643-1727).jpg)
[#/media/File:Sir_Isaac_Newton_\(1643-1727\).jpg](#)

Su resultado puede escribirse así

$$\pi = 3 \sum_{n=0}^{\infty} \frac{(2n)!}{16^n \cdot (n!)^2 \cdot (2n+1)} \quad (5.9)$$

Esta fórmula expresa π en términos de una serie rápidamente convergente de números racionales en la que cada término tiene una expresión explícita. De ella podemos obtener un algoritmo recursivo muy sencillo para cada término de la serie a partir del que le precede. Efectivamente, definiendo

$$c_n = \frac{(2n)!}{16^n \cdot (n!)^2}$$

la fórmula (5.9) se puede escribir así

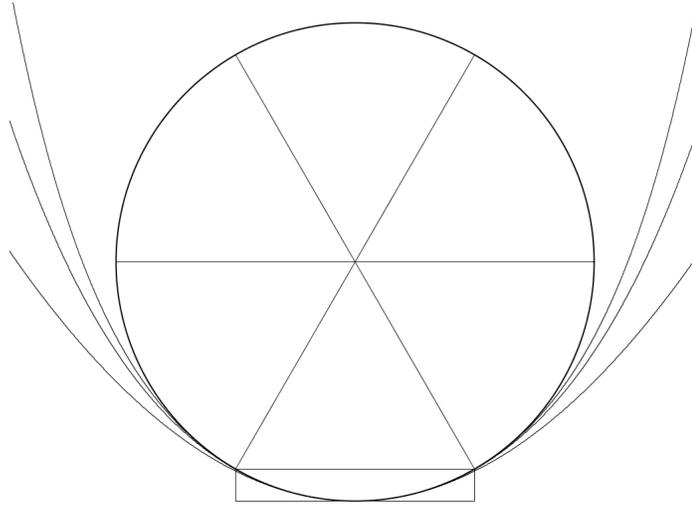
$$\pi = 3 \sum_{n=0}^{\infty} \frac{c_n}{2n+1} \quad (5.10)$$

donde $c_0 = 1$ y, para $n = 1, 2, \dots$,

$$c_n = c_{n-1} \frac{(2n-1)(2n)}{16n^2} = c_{n-1} \frac{(2n-1)}{8n} \quad (5.11)$$

La serie (5.10) brinda una manera sencilla y muy rápida de obtener buenas aproximaciones de π . De hecho, con este método se pueden obtener 16 decimales de π con sólo 22 términos. Manualmente este cálculo puede hacerse en una hora, en minutos usando una calculadora y en fracciones de segundo con una computadora.

Para obtener este resultado calcularemos π como el área del círculo unitario y usaremos el hexágono inscrito y la aproximación a una función mediante una serie de potencias, concepto que surgió apenas en el siglo XVII gracias, entre otros, al propio Newton.



Aproximando $1 - \sqrt{1 - x^2}$ por $\sum_{n=0}^N a_n x^n$, $N = 1, 2, 3, \dots$

El área puede calcularse así

$$\pi = 6(T + R - 2C) \quad (5.12)$$

donde:

T es el área del triángulo equilátero de lado 1: $T = \frac{\sqrt{1 - (\frac{1}{2})^2}}{2}$,

R es el área del rectángulo de base 1 y altura $1 - \sqrt{1 - (\frac{1}{2})^2}$ y

C es el área bajo la gráfica⁴ de la función $1 - \sqrt{1 - x^2}$ entre 0 y $\frac{1}{2}$.

Entonces

$$T = \frac{1}{2} - \frac{1 - \sqrt{1 - (\frac{1}{2})^2}}{2} \quad R = 1 - \sqrt{1 - (\frac{1}{2})^2} \quad C = \int_0^{\frac{1}{2}} 1 - \sqrt{1 - x^2} dx$$

Definiendo $f(x) = 1 - \sqrt{1 - x^2}$, estas fórmulas pueden escribirse así

$$T = \frac{1}{2} - \frac{f(\frac{1}{2})}{2} \quad R = f(\frac{1}{2}) \quad C = \int_0^{\frac{1}{2}} f(x) dx$$

Sustituyendo en (5.12) obtenemos

$$\pi = 6\left(\frac{1}{2} - \frac{f(\frac{1}{2})}{2} + f(\frac{1}{2}) - 2 \int_0^{\frac{1}{2}} f(x) dx\right)$$

que se simplifica a

$$\pi = 3\left(1 + f\left(\frac{1}{2}\right) - 4 \int_0^{\frac{1}{2}} f(x) dx\right) \quad (5.13)$$

La función $f(x)$ se puede escribir como

$$f(x) = \sum_{n=1}^{\infty} a_n x^{2n} \quad (5.14)$$

⁴El origen se colocó una unidad abajo del centro para que la gráfica coincida con la circunferencia.

donde los coeficientes a_n , que más adelante obtendremos usando la serie binomial, son positivos. Sustituyendo (5.14) en (5.13)

$$\pi = 3 \left(1 + \sum_{n=1}^{\infty} a_n \left(\frac{1}{2} \right)^{2n} - 4 \int_0^{\frac{1}{2}} \sum_{n=1}^{\infty} a_n x^{2n} dx \right)$$

Calculando la integral término a término

$$\pi = 3 \left(1 + \sum_{n=1}^{\infty} a_n \left(\frac{1}{2} \right)^{2n} - 4 \sum_{n=1}^{\infty} a_n \frac{\left(\frac{1}{2} \right)^{2n+1}}{2n+1} \right)$$

Reuniendo las dos sumas y sacando como factores comunes a $a_n \left(\frac{1}{2} \right)^{2n}$

$$\pi = 3 \left(1 + \sum_{n=1}^{\infty} a_n \left(\frac{1}{2} \right)^{2n} \left(1 - 4 \frac{\frac{1}{2}}{2n+1} \right) \right)$$

Lo cual podemos reescribir así

$$\pi = 3 \left(1 + \sum_{n=1}^{\infty} a_n \frac{1}{2^{2n}} \left(1 - \frac{2}{2n+1} \right) \right)$$

Usando $2n+1$ como denominador común, podemos simplificar y escribir

$$\pi = 3 \left(1 + \sum_{n=1}^{\infty} a_n \frac{2n-1}{4^n \cdot (2n+1)} \right) \quad (5.15)$$

Para obtener los coeficientes a_n desarrollamos $(1-x^2)^{\frac{1}{2}}$ en serie binomial

$$f(x) = 1 - (1-x^2)^{\frac{1}{2}} = 1 - 1 + \frac{1}{2}x^2 + \frac{\frac{1}{2} \cdot \frac{1}{2}}{2!}x^4 + \frac{\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{3}{2}}{3!}x^6 + \dots$$

Simplificando

$$f(x) = \frac{1}{2}x^2 + \frac{\frac{1}{2} \cdot \frac{1}{2}}{2!}x^4 + \frac{\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{3}{2}}{3!}x^6 + \dots$$

De lo cual se puede ver que $a_1 = \frac{1}{2}$ y, para $n > 1$

$$a_n = \frac{1 \cdot 3 \cdot \dots \cdot (2n-3)}{2^n n!} = \frac{1 \cdot 3 \cdot \dots \cdot (2n-1)}{2^n \cdot n! \cdot (2n-1)}$$

Multiplicando numerador y denominador por $2 \cdot 4 \cdot \dots \cdot 2n$ obtenemos

$$a_n = \frac{(2n)!}{(2 \cdot 4 \cdot \dots \cdot 2n) \cdot 2^n \cdot n! \cdot (2n-1)} = \frac{(2n)!}{2^n \cdot n! \cdot 2^n \cdot n! \cdot (2n-1)}$$

y finalmente

$$a_n = \frac{(2n)!}{4^n \cdot (n!)^2 \cdot (2n-1)} \quad (5.16)$$

para $n = 1, 2, \dots$. Sustituyendo (5.16) en (5.15)

$$\pi = 3 \left(1 + \sum_{n=1}^{\infty} \frac{(2n)!}{4^n \cdot (n!)^2 \cdot (2n-1)} \cdot \frac{(2n-1)}{4^n \cdot (2n+1)} \right)$$

Lo cual puede escribirse finalmente como

$$\pi = 3 \left(1 + \sum_{n=1}^{\infty} \frac{(2n)!}{16^n \cdot (n!)^2} \cdot \frac{1}{(2n+1)} \right) \quad (5.17)$$

Como la fórmula para los coeficientes en (5.17) es válida para $n = 0$ y da 1, esto nos lleva a (5.9), que es lo que queríamos demostrar.

Cálculos de π con muchos decimales

Utilizando (5.10) y (5.11) en una computadora con aritmética de alta precisión se puede obtener la siguiente aproximación de π con 600 cifras decimales de precisión.

$\pi = 3.$

141592 653589 793238 462643 383279 502884 197169 399375 105820 974944
 592307 816406 286208 998628 034825 342117 067982 148086 513282 306647
 093844 609550 582231 725359 408128 481117 450284 102701 938521 105559
 644622 948954 930381 964428 810975 665933 446128 475648 233786 783165
 271201 909145 648566 923460 348610 454326 648213 393607 260249 141273
 724587 006606 315588 174881 520920 962829 254091 715364 367892 590360
 011330 530548 820466 521384 146951 941511 609433 057270 365759 591953
 092186 117381 932611 793105 118548 074462 379962 749567 351885 752724
 891227 938183 011949 129833 673362 440656 643086 021394 946395 224737
 190702 179860 943702 770539 217176 293176 752384 674818 467669 405132

El cálculo tarda unos 25 segundos.

Para obtener 1,200 decimales se requieren unos 3 minutos, para 7,200 decimales, 4 horas y para 12,000 decimales, 22 horas.

Actualmente se conocen métodos aún más eficientes con los que se han llegado a calcular miles de millones de cifras decimales de π . Esos cálculos tardan meses en realizarse. Hace mucho tiempo había interés en esto porque no se sabía si π era racional y se intentaba descubrir un patrón que se repitiera, en la esperanza de descubrir la fracción que lo representaba exactamente. Pero desde que se descubrió que π no puede ser un número racional (siguiente sección), este tipo de esfuerzos perdió relevancia porque para la mayoría de los cálculos que hay que realizar, basta conocer unos 16 decimales de π , que es lo que ofrece la aritmética de una computadora personal. Para cálculos más delicados, un centenar de decimales serían más que suficientes. Así que en realidad el cálculo de π con miles o millones de cifras decimales no tienen ninguna utilidad práctica, se trata de un mero desafío. Algunos investigadores tienen interés en estudiar la distribución estadística de los dígitos en una expansión larga de π , por ejemplo, en saber si todos aparecen con la misma probabilidad. Con la riqueza de las matemáticas es imposible saber si ese tipo de investigaciones dará alguna vez un resultado interesante.

La irracionalidad de π

Durante siglos los matemáticos se preguntaron si π sería o no racional. Si lo fuera habría una periodicidad en sus decimales. Esto fomentó el interés de obtener muchos decimales de π . A medida que se calculaban más y más se iban buscando patrones que se repitieran. Pero nunca se encontró uno. En el siglo XVIII, el matemático suizo JOHANN HEINRICH LAMBERT dio la primera demostración de que nunca se encontraría tal patrón porque π es un número irracional. A continuación presentamos una demostración de la irracionalidad de π publicada por IVAN NIVEN de Purdue University en 1946, la cual requiere únicamente de conceptos elementales de cálculo.

La demostración procede por contradicción. Supongamos que $\pi = \frac{a}{b}$, donde a y b son enteros positivos. Definamos

$$f(x) = \frac{x^n(a-bx)^n}{n!}$$

y

$$F(x) = f(x) - f^{(2)}(x) + f^{(4)}(x) - \dots + (-1)^n f^{(2n)}(x)$$

donde n es un entero positivo cualquiera. Entonces $n!f(x)$ es un polinomio con coeficientes enteros de grado no mayores que n , así que $f(x)$ y sus derivadas $f^{(i)}(x)$ toman valores enteros en $x = 0$ y en $x = \frac{a}{b} = \pi$. Por cálculo elemental sabemos que

$$\frac{d}{dx} [F'(x) \operatorname{sen}(x) - F(x) \operatorname{cos}(x)] = [F''(x) + F(x)] \operatorname{sen}(x) = f(x) \operatorname{sen}(x)$$

y por tanto

$$\int_0^\pi f(x) \operatorname{sen}(x) dx = [F'(x) \operatorname{sen}(x) - F(x) \operatorname{cos}(x)]_0^\pi = F(\pi) + F(0) \quad (5.18)$$

$F(\pi) + F(0)$ es un entero porque $f^{(i)}(\pi)$ y $f^{(i)}(0)$ son enteros para todo i . Pero para $0 < x < \pi$

$$0 < f(x) \operatorname{sen}(x) < \frac{\pi^n a^n}{n!}$$

así que el valor de la integral (5.18) es un entero positivo y sin embargo es menor que $\pi \frac{(\pi a)^n}{n!}$ para todo entero positivo n . Pero $\frac{c^n}{n!}$ tiende a cero cuando n tiende a infinito, lo cual nos lleva a una contradicción. Por lo tanto es imposible que π sea racional.

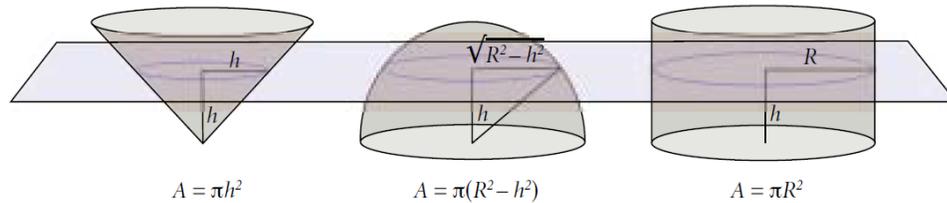
Ahora se sabe que el número π no solo no es racional sino que tampoco es algebraico, es decir, que no es solución de alguna ecuación algebraica con coeficientes racionales. Los números reales que no son algebraicos se llaman trascendentes. La primera demostración de que π es un número trascendente la dio CARL LOUIS FERDINAND VON LINDEMANN en 1882.

El enorme interés que se puede apreciar a través de lo expuesto en esta sección por comprender al número π y los grandes esfuerzos realizados por estimar con gran precisión su valor, se deben a la enorme importancia que ha tenido para el desarrollo de las matemáticas. No sólo el perímetro y el área del círculo están íntimamente relacionados con el número π sino también el volumen y la superficie de la esfera guardan una íntima relación con este número, como demostró el mismo Arquímedes.

5.6. El volumen y la superficie de la esfera

Arquímedes obtuvo el volumen y la superficie de una esfera y demostró que el volumen de la esfera es igual a $2/3$ del volumen del mínimo cilindro que la contiene. También demostró que la superficie de la esfera es igual a $2/3$ de la superficie de ese mismo cilindro. Estaba tan orgulloso de este resultado que pidió se inscribiera en su tumba. Gracias a que aparente se cumplió su deseo, fue posible localizar su tumba, que estuvo perdida durante varios siglos.

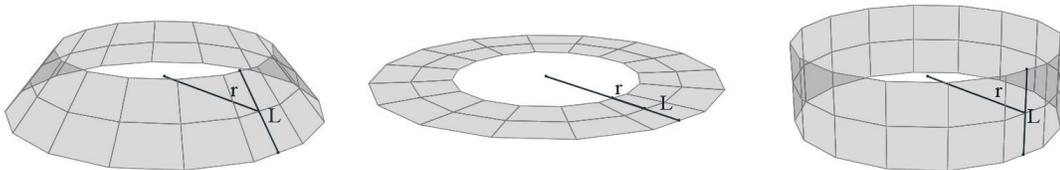
El cálculo del volumen de la esfera utiliza el Teorema de Pitágoras de una manera muy ingeniosa. Consideremos tres cuerpos: media esfera, un cono circular recto y un cilindro circular. Las bases del cono y el cilindro tiene el mismo radio que la esfera ambos tiene altura también igual al radio. Lo que Arquímedes demuestra es que el volumen de la semiesfera más el del cono es igual al del cilindro. Como ya sabía que el volumen del cono es $\frac{1}{3}$ del del cilindro, resulta que el de la semiesfera debe ser $\frac{2}{3}$ del del cilindro.



Sea R el radio común de la semiesfera y de las bases del cono y el cilindro. Coloquemos los tres cuerpos apoyados en un plano horizontal, con los ejes del cono y el cilindro perpendiculares al plano el vértice del cono a la altura de la base de la semiesfera.

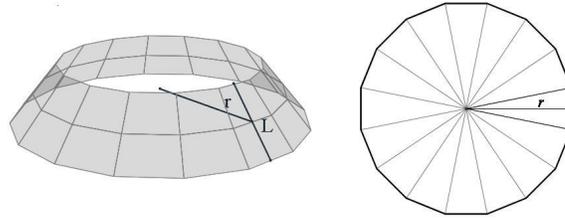
Ahora cortamos los tres cuerpos por un plano horizontal a una altura h de la base de la semiesfera. Los cortes del plano en el cono y en la semiesfera son círculos de radios h y $\sqrt{R^2 - h^2}$ respectivamente. Por tanto la suma de sus áreas es $\pi h^2 + \pi(R^2 - h^2) = \pi R^2$, que es el área del corte en el cilindro. Como esto ocurre para todos los planos horizontales, lo que queríamos probar resulta evidente. Esta idea fue utilizada sistemáticamente por BONAVENTURA CAVALIERI (1598 - 1647) calcular los volúmenes de otros cuerpos a partir de las áreas de sus secciones paralelas a un plano, por lo que en la actualidad se reconoce como el *Principio de Cavalieri*.

El cálculo de la superficie de la esfera depende de la siguiente observación. Consideraremos sectores de pirámides cuyas bases son polígonos regulares. Entonces sus caras son trapecios iguales. Todos los sectores de piramidales que tienen el mismo perímetro medio y la misma longitud lateral L , tienen la misma superficie.

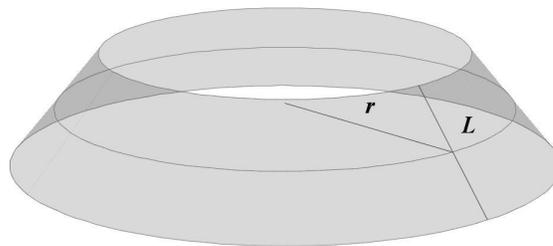


Esto se debe a que el área de los trapecios que los forman depende sólo de L y de la anchura media. Por tanto el área del sector piramidal depende sólo de la longitud del polígono cuyos vértices son

los puntos medios de las aristas laterales. De hecho el área del sector de pirámide es PL , donde P es el perímetro del polígono.



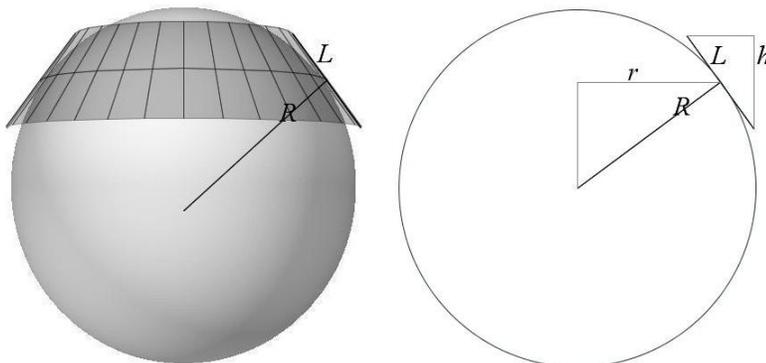
Como esto sucede independientemente del número de lados que tengan las orillas del sector piramidal, lo mismo va a ocurrir con los sectores cónicos.



Los que tengan el mismo radio medio r y el mismo tamaño de sus elementos laterales L , tendrán la misma superficie: $2\pi rL$.

Para calcular la superficie de la esfera Arquímedes la cubre con N sectores de cono como el de la figura. La superficie de la cubierta será igual a

$$\sum_{n=1}^N 2\pi r_n L_n$$



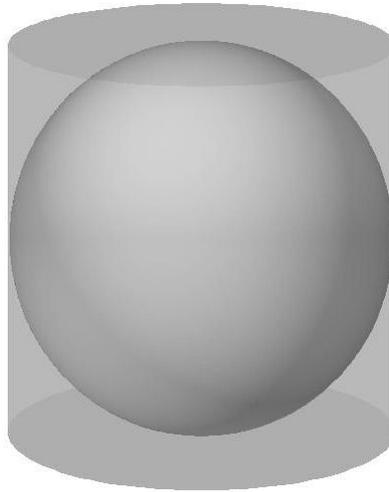
Luego observa que, por semejanza de triángulos:

$$\sum_{n=1}^N 2\pi r_n L_n = \sum_{n=1}^N 2\pi R h_n = 2\pi R \sum_{n=1}^N h_n = 2\pi R H$$

donde H es la altura total de la cubierta. A medida que la cubierta se hace más fina, H tiende a $2R$ y por lo tanto, la superficie de la esfera debe ser

$$4\pi R^2$$

En realidad el argumento de Arquímedes es más riguroso, acota la superficie de la esfera por una sucesión que converge a $4\pi R^2$ desde arriba y por otra que converge al mismo valor desde abajo. Su método satisface los criterios de rigor más estrictos.



La superficie del cilindro circunscrito a la esfera es la del rectángulo cuyos lados son el perímetro $2\pi R$ y la altura $2R$, más las de las dos tapas, que son círculos de radio R :

$$2\pi R \cdot 2R + 2\pi R^2 = 6\pi R^2$$

Por lo tanto la razón entre la superficie de la esfera y la del cilindro circunscrito a ella es $\frac{2}{3}$, la misma que entre el volumen de la esfera y el del cilindro circunscrito. Éste es el resultado del que Arquímedes estaba más orgulloso, por lo que pidió se inscribiese en su tumba.

5.7. Cálculos de distancias celestes

Ideas sobre el cosmos

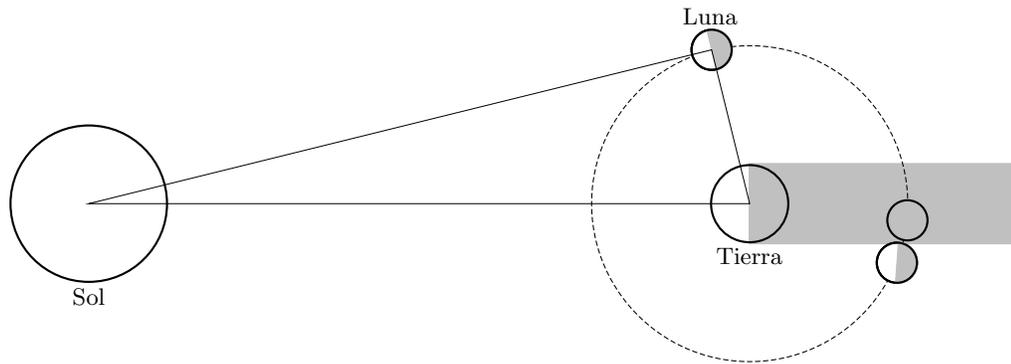
Para poder medir distancias entre objetos celestes es necesario tener una concepción de estos como cuerpos físicos. Mientras se piensa que el Sol es una rueda de fuego que es jalado por una yunta

uno no puede pensar en medir la distancia hacia el Sol. El requisito es entonces que el pensamiento se haya liberado en cierta medida de las explicaciones místicas. Esto inició con TALES DE MILETO (c. 625–547 a.C.), quién – por ejemplo – trató de explicar las inundaciones anuales del Nilo por vientos fuertes. Una explicación que ciertamente es equivocada, pero trata de explicar un fenómeno físico, y en esto se diferencía mucho de las demás explicaciones en aquel entonces, por razones de la misma naturaleza. ANAXIMANDRO (c. 610–547 a.C.), considerado como “alumno” de TALES, dió una explicación diferente para estas inundaciones que era igualmente equivocada. Este pequeño ejemplo muestra como empezó una era diferente: se busca explicar los fenómenos de la naturaleza por la misma naturaleza. Con ello se puede especular y discutir sobre las razones de los fenómenos y se empieza dudar de las explicaciones dogmáticas basadas en creencias.

Para retornar al tema de esta sección: para poder medir distancias entre objetos celestes, es necesario tener primero una concepción natural del cosmos. La palabra “cosmos” origina de la palabra griega para “adornar” ya que se pensaba que la bóveda celeste estaba adornada con el Sol, la Luna, las estrellas y los planetas.

TALES pensaba que la Tierra flota sobre agua, sin decirnos sobre qué se apoyaba el agua. Los hindues a cambio opinaban que la Tierra reposaba sobre una tortuga, esta sobre un elefante, y este último sobre una hoja de Lotus. Nos podemos preguntar en qué se apoyaba la hoja. ANAXIMANDRO fue el primero que resolvió el “problema de soporte”, es decir la cuestión, qué soportaba la Tierra. ANAXIMANDRO pensó que la Tierra flotaba en el aire, aunque la pensaba todavía en forma de un cilindro. Fue PITÁGORAS (ca. 569–475 a.e.c.) el primero que pensó la Tierra como una esfera. ANAXÁGORAS, él que afirmó que la Luna y el Sol eran rocas esféricas y que la Luna reflejaba la luz del Sol. ARISTÓTELES dió varios argumentos en favor de la forma esférica de la Tierra misma. Dos de ellos son: que más en el Norte o en el Sur se podían ver nuevos astros, que en eclipses lunares, cuando la Luna entra a la sombra que proyecta la Tierra, la sombra siempre se delimita por un arco de circunferencia. La sombra de la Tierra es redonda, por lo que se concluye que la Tierra misma lo es también.

El primero que desarrolló un sistema celeste, que colocaba el Sol en el centro fue ARISTARCO DE SAMOS (c. 310–230 a.C.). En este sistema se pensó a la Tierra y a los cinco planetas en aquel entonces conocidos: Mercurio, Venus, Marte, Jupiter y Saturno como esferas que giran alrededor del Sol. Veremos en esta sección primero las ingeniosas ideas de Aristarco: cómo estimó los tamaños relativos entre la Tierra, la Luna y el Sol y las distancias entre ellos. Sus estimaciones daban que el Sol era varias veces más grande que la Tierra y por ello colocó a él en el centro del universo y no la Tierra. Eso fue son casi dos mil años antes de NICOLÁS COPÉRNICO (1473–1543), que acredita de manera justa el sistema heliocéntrico a Aristarco.



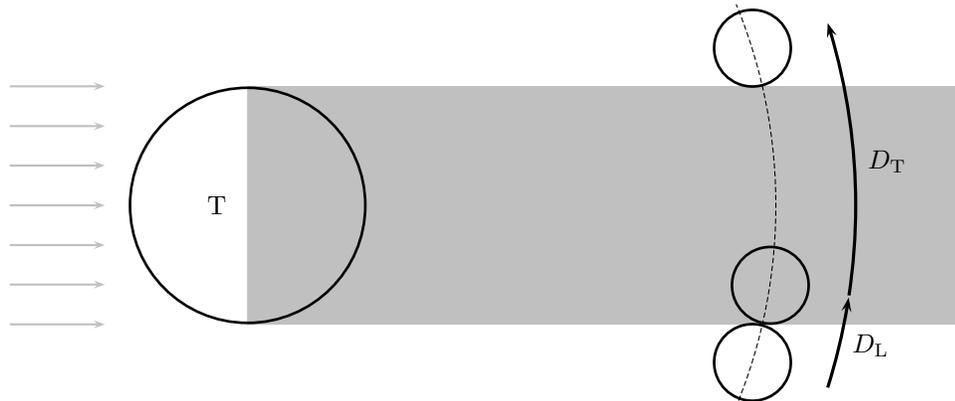
Después veremos la idea de ERATÓSTENES DE CYRENE (c. 275.–194 a.C.), de medir el tamaño de la Tierra y finalmente una idea de GALILEO GALILEI (1564–1642) para medir la altura de las montañas en la Luna.

Todas estas ideas tienen varios aspectos en conjunto: primero, se basan en mediciones que se hicieron en una sola posición de la Tierra, segundo, se mide una distancia inalcanzable de manera indirecta y tercero, las ideas son muy sencillas pero no por eso menos ingeniosas.

Tamaño relativo entre la Tierra y la Luna

ARISTARCO comparó el diámetro de la Tierra D_T con el diámetro de la Luna D_L observando un eclipse lunar. En un eclipse lunar la Tierra se encuentra en medio del Sol y la Luna y esta última pasa en su órbita la sombra que proyecta la Tierra. En esta posición se observa una Luna llena y durante la eclipsis la aparición de un disco lleno empieza reducirse poco a poco hasta que la Luna quede casi totalmente oscurecida.

ARISTARCO comparó dos tiempos: por un lado, el tiempo t_L que tardaba la Luna desde el primer momento que empieza oscurecerse hasta que quede totalmente oscurecido. Por otro lado observó el tiempo t_T , que tardó en volverse como luna llena completo después de haberse oscurecido por completo. La situación se muestra esquemáticamente en la siguiente figura.



Como la Luna se mueve más o menos con la misma velocidad todo el tiempo, se tiene que

$$\frac{D_T}{D_L} \approx \frac{t_T}{t_L}.$$

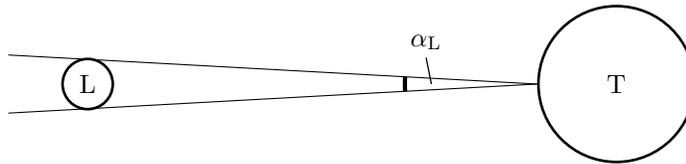
ARISTARCO estimó o midió que $\frac{t_T}{t_L} \approx 0.36$. En aquel entonces no había herramientas de medición muy precisas, lo que tal vez explica la diferencia con el valor que hoy se tiene: $\frac{t_T}{t_L} \approx 0.27$.

ARISTARCO obtuvo entonces la estimación

$$D_L \approx 0.36D_T.$$

Tamaño relativo entre el diámetro de la Luna y la distancia hacia ella

El Sol y la Luna se ven desde la Tierra bajo un cierto *diámetro angular*. Ambos cuerpos celestes, el Sol y la Luna, tienen aproximadamente el mismo diámetro angular, pero sólo aproximadamente, ya que ambos varían ligeramente durante el año, ya que la distancia a la Tierra varía.



ARISTARCO midió un diámetro angular de ambos cuerpos celestes de 2° . En realidad, el diámetro angular es aproximadamente 0.53° . La medición se puede hacer con una moneda chica: se tapa el Sol, o lo que para el ojo es más agradable, la Luna con la moneda con el brazo extendido y se le pide a otra persona que mida la distancia a de la moneda al ojo. Luego se mide el diámetro de la moneda d . Por semejanza se tiene

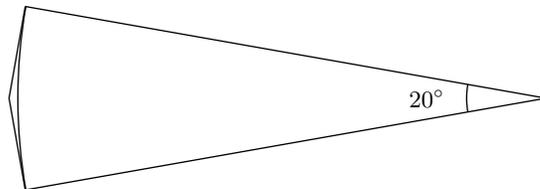
$$d : a = D_L : A_{TL},$$

donde A_{TL} denota la distancia de la Tierra a la Luna.

Si se toma la medición de Aristarco, entonces se tiene

$$\frac{D_L}{A_{TL}} \approx \frac{2^\circ}{360^\circ} 2\pi,$$

en buena aproximación, ya que la longitud del arco es casi igual al diámetro para ángulos chicos, como se muestra en la siguiente ilustración con un ángulo mucho más grande: 20° .



Dado que el diámetro angular del Sol y la Luna son casi iguales se tiene

$$\frac{A_{TL}}{A_{TS}} \approx \frac{D_L}{D_S},$$

donde D_S denota el diámetro del Sol y A_{TS} la distancia de la Tierra al Sol.

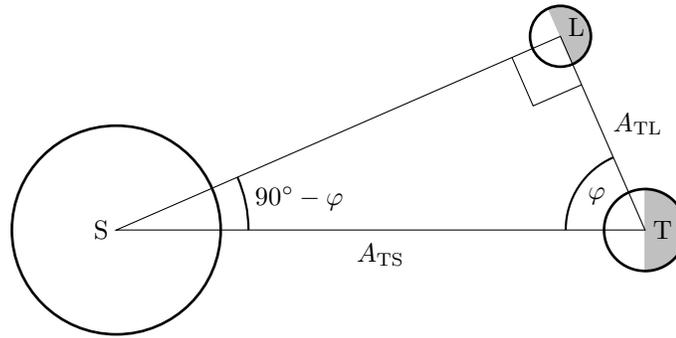
Si se toma a $\pi \approx 3$, entonces se obtiene con las mediciones de ARISTARCO las siguientes estimaciones:

$$A_{TL} \approx \frac{360^\circ}{2^\circ \cdot 2\pi} D_L \approx 30 \cdot D_L \approx 10.8 \cdot D_T,$$

$$A_{TS} \approx \frac{360^\circ}{2^\circ \cdot 2\pi} D_S \approx 30 \cdot D_S.$$

Tamaño relativo entre la distancia hacia el Sol y la Luna

Para poder relacionar todas las distancias, ARISTARCO requirió una idea más. Para ello se imaginó la constelación de Tierra, Luna y Sol cuando es media Luna. En esta situación se ve la Luna dividida en dos mitades: una iluminada y otra oscura, luz y sombra, dividida por una línea circular. El plano que contiene esta línea circular pasa por el ojo del observador en la Tierra. El triángulo LTS formada por los tres cuerpos, tiene en la Luna L un ángulo recto. En este momento, ARISTARCO midió el ángulo



La idea de ARISTARCO es muy buena, pero a la vez también muy difícil de llevar a cabo con precisión. Medir un ángulo chico es bastante fácil, pero medir un ángulo muy cercano a 90° es bastante difícil si no se cuenta con herramientas de medición de precisión. Además es difícil determinar el momento preciso cuando sucede que el ángulo SLM es exactamente un ángulo recto. ARISTARCO midió un ángulo de $\varphi \approx 87^\circ$, cuando en realidad es $\varphi = 89.95^\circ$. Eso condujo a que el Sol y la distancia hacia ella se estimó muy por debajo del valor real.

Con la aproximación de ángulos chicos se tiene ahora

$$\frac{A_{TL}}{A_{TS}} \approx \frac{90^\circ - \varphi}{360^\circ} 2\pi.$$

Con la medición de ARISTARCO se obtiene entonces

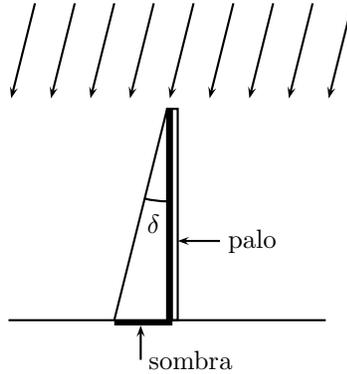
$$A_{TS} \approx 20 \cdot A_{TL} \approx 216 \cdot D_T$$

$$D_S \approx \frac{1}{30} \cdot A_{TS} \approx 7.2 \cdot D_T$$

Se ve que las relaciones están bien observados: el Sol es más grande que la Tierra, la Luna más chica, y el Sol más lejos que la Luna. Eso explica por qué ARISTARCO colocó en su sistema el Sol en el centro del universo. Si ARISTARCO hubiera tenido unas herramientas e medición, entonces habría obtenido estimaciones muy buenas con las sencillas, pero ingeniosas ideas de relacionar los tamaños. Con ello concluimos ahora lo que se refiere al trabajo de ARISTARCO.

El tamaño de la Tierra

La primera medición del tamaño de la Tierra se atribuye a ERATÓSTENES (276–194 a. C.). Escuchó que el día del solsticio de verano un palo vertical en Siena (una ciudad que hoy se llama Asuan, en Egipto) no deja sombra, es decir el Sol se encuentra justo en el Cenit. Siena se encuentra casi exactamente en el Sur de Alejandría, lugar en el cual se encontraba ERATÓSTENES. Él midió entonces el ángulo δ entre un palo vertical y la línea recta entre la punta del palo y su sombra, es decir el ángulo de los rayos de Sol con la vertical en Alejandría.



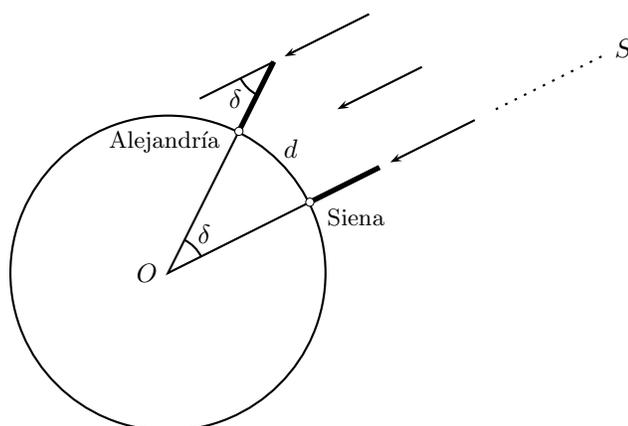
ERATÓSTENES midió que el ángulo δ era $\frac{1}{50}$ de un ángulo completo, es decir

$$\delta = \frac{1}{50} 360^\circ = 7.2^\circ.$$

Si se conoce la distancia d entre Alejandría y Siena, entonces se puede calcular la circunferencia U_T de la Tierra por

$$\frac{\delta}{360^\circ} = \frac{d}{U_T}.$$

Esto lo explica la siguiente figura.



Para estimar la distancia entre Alejandría y Siena, ERATÓSTENES se basó en los relatos de los comerciantes que contaban que recorrían la distancia en caravanas en 50 días y estimaban que se trasladaban como 100 “estadios” cada día. El “estadio” es una medida antigua y hoy no se sabe bien cuánto valía, pero se piensa que debería ser entre 157 m y 211 m.

La distancia entre Alejandría y Siena se estimaba entonces en 5000 estadios y por ello la circunferencia terrestre por

$$U_T = 50 \cdot 5000 \text{ estadios} = 250\,000 \text{ estadios},$$

es decir entre 39 250 km y 52 750 km, una estimación muy buena si uno piensa en la sencillez de la medición.

Más tarde se perdió el conocimiento de ERATÓSTENES y Cristobal Colón emprendió su viaje hacia la India pensando que la Tierra era mucho más chica. Si hubiera conocido la medición de ERATÓSTENES, tal vez nunca hubiera zarpado. Se volvió la imaginación nuevamente más antropocentrista y con ello geocentrista.

La altura de las montañas en la Luna

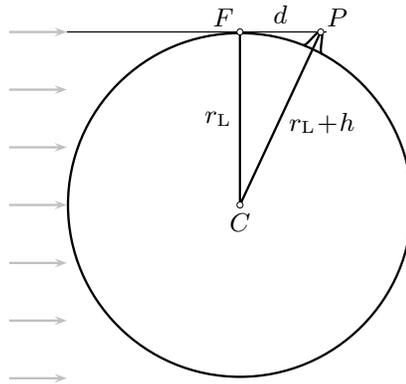
Muy influyente para la Edad Media eran los libros de Aristóteles. En ciertos asuntos, él se basaba en PLATÓN. Así, al menos en parte, en su cosmología. Según una especulación de PLATÓN, la materia está compuesta por triángulos que forman sólidos regulares. Así se explicó la existencia de los cuatro elementos y el quinto sólido (el dodecaedro) lo atribuyeron al cosmos, a la composición de las estrellas, la Luna, el Sol y los planetas.

Por ello, cuenado GALILEO GALILEI usa por primera vez, en enero de 1610, de observar el cielo nocturno con un telescopio sus descubrimientos chocan con toda esta teoría, ya que observó que Jupiter tenía cuatro lunas, que la vía láctea estaba compuesta por miles de estrellas, que el Sol tenía manchas y que la superficie de la Luna no estaba lisa sino presentaba irregularidades, ver las siguientes dos ilustraciones. La izquierda es de GALILEI misma, la derecha es moderna.



GALILEI interpretó estas irregularidades como las sombras que proyectan montañas. Midió la altura de estas montañas al observar la media luna de la siguiente manera: Los rayos de Sol iluminan los picos más altas aunque las valles ya se oscurecieron por completo. Por ello se puede observar los picos más altos más allá que el diámetro. En la ilustración arriba se señaló el diámetro y los picos más altos iluminados del lado oscuro. GALILEI estimó que la distancia de los picos que todavía se iluminan distan un ventésimo del diámetro de la Luna.

La siguiente ilustración muestra un esquema de la situación y explica el cálculo correspondiente.



Con la estimación de la distancia $d \approx \frac{1}{10}r_L$ se puede calcular la hipotenusa del triángulo rectángulo CPF :

$$r_L + h = \sqrt{r^2 + d^2} \approx \sqrt{r^2 + \frac{1}{100}r^2} = \sqrt{1.01}r \approx 1.005r.$$

Por ello se obtiene

$$h = 0.005r,$$

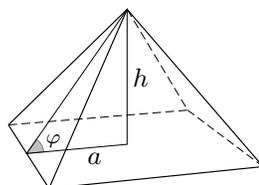
lo que equivale a una altura de $h = 8268$ m. GALILEI afirma que en la Tierra no hay montaña de semejante altura, pero esto hay que atribuir al hecho que las mediciones de la Tierra apenas se empiezan a hacer unos 100 años después y el Himalaya se midió a mediados del siglo XIX. Actualmente se sabe que las montañas más altas en la Luna se erigen unos 5500 m sobre la superficie. El cálculo realizado por Galileo dio un valor correcto en cuanto al orden de magnitud.

5.8. La trigonometría

Antecedentes

Con *trigonometría* nos referimos hoy día al cálculo de longitudes y ángulos en triángulos. La etimología es clara: *gon* significa en griego “ángulo”, *tri* significa “tres” y *metría* significa “medición”, así que el *trigon* es el triángulo, y la trigonometría es la medición de los triángulos. Pero la palabra es una creación de la edad media: aparece por primera vez en un libro de 1590.

Los inicios de lo que hoy llamamos trigonometría son muy antiguos, provienen de los egipcios: ellos calculaban el *sequed* de pirámides, un número ínimamente relacionado con nuestra noción de *pendiente*: el *sequed* es el número de palmas (anchos de mano) horizontales para subir 7 palmas verticales.



En la figura anterior, el *sequed* de la pirámide es $s = \frac{7a}{h}$.

Hoy podemos ver el *sequed* –si ignoramos el factor 7 en la definición– como la cotangente del ángulo φ entre los planos que forman un costado de la pirámide y su base. Pero esto es una interpretación moderna y da una falsa impresión. En realidad el *sequed* sólo utiliza la semejanza de triángulos, no relaciona los catetos con el ángulo, como lo hacen las modernas funciones trigonométricas.

La trigonometría, como muchas áreas de las matemáticas se desarrolló para poder llevar a cabo ciertos cálculos con precisión y eficacia. En este caso, el promotor principal del desarrollo fue la astronomía. Se pensaba que los objetos celestes se movían en círculos alrededor de la Tierra y los astrónomos querían poder calcular sus posiciones y hacer predicciones de los futuros eclipses. Para ello era necesario poder relacionar los ángulos con las proporciones entre los lados correspondientes de los triángulos.

Trigonometría esférica

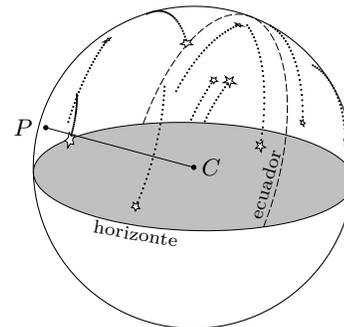
La trigonometría que conocemos hoy es la trigonometría plana. Es interesante saber que su origen está en la trigonometría esférica, que se desarrolló mucho antes. Los griegos como muchas culturas observaban el cielo y crearon diferentes sistema para explicar los movimientos de los cuerpos celestes.

Al observar el cielo nocturno desde un llano, uno se percató de que las estrellas se mueven respecto a los árboles o las montañas. Nuestra vista nos muestra una bóveda celeste delimitada por el horizonte.

No es difícil observar una regularidad en este movimiento: las estrellas se mueven como se estuvieran pegadas sobre una esfera gigante, el *orbe celeste*, que gira lentamente por un eje que pasa por el punto C en donde está el observador.

El movimiento es de 15° a la hora, por lo que el orbe celeste da una vuelta completa al día.

Hay una dirección que queda fija: el *polo celeste* P , que es el punto del orbe celeste por el cual este gira. El plano perpendicular al eje de rotación intersecta el orbe en el *ecuador celeste*.

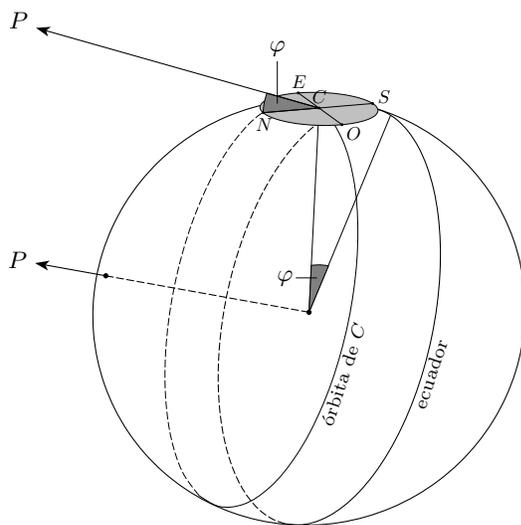
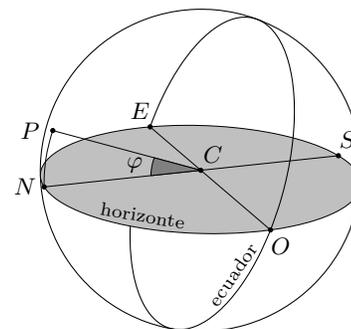


Una estrella cerca del polo celeste casi no se mueve. En el hemisferio norte hay una estrella muy cercana a este punto ideal fijo: la *estrella polar*. En el hemisferio sur no hay una estrella muy luminosa cerca, pero una constelación, llamada *Cruce* (o Cruz del Sur) que permite la determinación del polo celeste visible con facilidad.

Dado que el polo magnético de la Tierra se encuentra muy cercano al polo norte de la rotación de la Tierra, la estrella polar en el hemisferio norte sirve para ubicarse tanto como una brújula.

El orbe celeste gira de tal manera que las nuevas estrellas aparecen en el este y otras desaparecen en el oeste.

Si cambiamos en nuestro pensamiento el punto de vista y pensamos que el orbe celeste es fijo y la Tierra se rota por su eje, entonces hay una explicación fácil para el ángulo λ del polo celeste P sobre el horizonte: $\varphi = \sphericalangle PCN$.



El ángulo φ es la *latitud* del punto C donde se encuentra el observador, es decir el ángulo entre la recta CZ , donde Z es el centro de la Tierra y el plano que contiene el ecuador terrestre.

La figura lateral muestra la situación. La esfera es ahora la Tierra y el disco gris es el horizonte con los cuatro puntos cardinales. Dado que las estrellas están muy lejos, el eje de rotación de la Tierra es en buena medida paralelo a la recta que une el espectador con el polo celeste o la estrella polar.

En el primer paso de simplificar la situación, se puede uno imaginar de estar en el polo norte. Ahí las estrellas siempre están a la misma altura. El horizonte coincide con el ecuador celeste. Dicho de

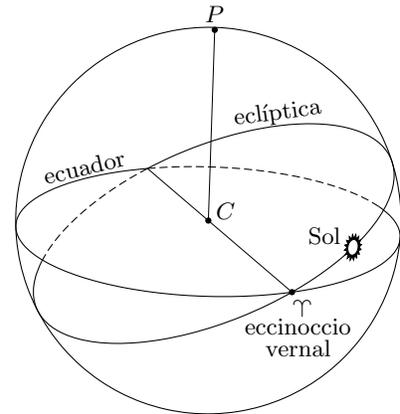
otro modo: podemos así ignorar la rotación de la Tierra sobre su propio eje.

Ahora veremos el movimiento aparente del Sol. Si el Sol fuera opaco, podríamos ver las estrellas a un lado del Sol, al igual que lo podemos cuando está la Luna. Dado que el Sol es mucho más luminoso que las estrellas, no podemos verlas de día. Pero podemos calcular la posición del Sol en el orbe celeste.

Resulta que el Sol se mueve muy despacio por el orbe celeste. Un poco menos que 1° cada día. Después de un año completó una vuelta. Claro, el origen de este movimiento aparente es que la Tierra se mueve alrededor del Sol y por ello cambian las estrellas, que están muchísimo más lejos, atrás del Sol.

El Sol sigue una círculo máximo que se llama *eclíptica* y que está inclinado por 23.44° respecto al ecuador. Recordemos: el ecuador resulta por el movimiento de la Tierra por su propio eje, la eclíptica por el movimiento de la Tierra alrededor del Sol.

El eje de rotación de la Tierra está inclinado por 23.44° respecto al plano en donde se mueve la Tierra alrededor del Sol.



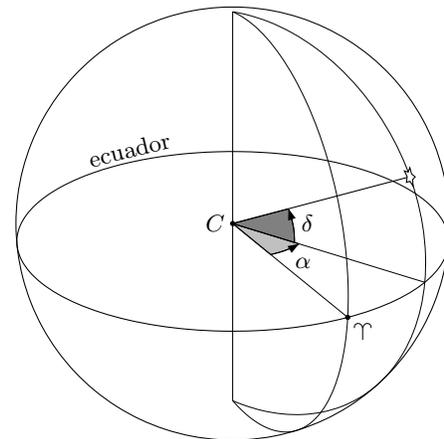
Los dos planos, el que contiene el ecuador y el que contiene la eclíptica se intersectan en una recta. Cuando el Sol está en uno de los dos puntos de esta recta, en la Tierra ocurren los *equinoccios* una vez en primavera, el Sol está entonces en el punto Υ , y otra vez en otoño. En cada uno de los dos equinoccios, el día y la noche son igual de largos en toda la tierra. Dado que el eje de rotación está inclinada respecto a la eclíptica, la Tierra expone medio año más el hemisferio norte y la otra mitad más el hemisferios sur. Esta desigualdad produce las estaciones en el año.

La ubicación de las estrellas en el orbe celeste se suele dar con respecto al ecuador. Se dan dos coordenadas: la declinación δ y la ascensión recta α .

La *declinación* δ es el ángulo que la estrella se eleva sobre el ecuador celeste, visto desde el centro C . Para estrellas en el hemisferio norte y sur la latitud es positiva resp. negativa.

La *ascensión recta* α es el ángulo que se mide a partir del equinoccio de primavera Υ en sentido matemático, es decir en contra de las manecillas del reloj si vemos el orbe desde arriba.

Las dos coordenadas, declinación y ascensión recta, conforman las coordenadas ecuatoriales.

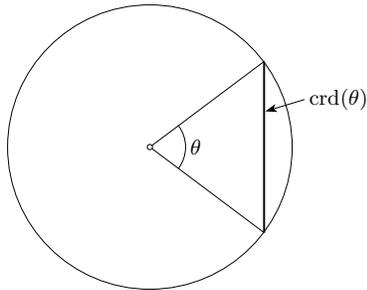
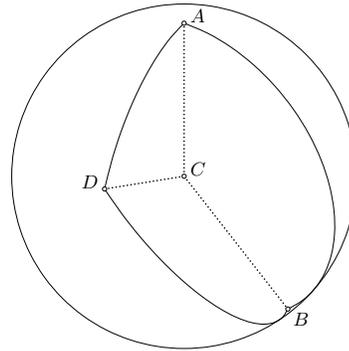


En nuestro sistema solar, todos los planetas se mueven en planos que tienen poca inclinación respecto al plano en el que se mueve la Tierra. Por ello, los planetas se ven desde la Tierra cercano a la eclíptica, en una cinta que va desde 5° abajo a 5° arriba. Esta cinta alrededor de la eclíptica se llama *zodiaco*.

Para dar la posición de un planeta se dan por ello coordenadas con respecto a la eclíptica. Estas coordenadas se llaman *latitud* y *longitud*. Calcular de unas coordenadas a las otras constituye uno de los principales problemas de la trigonometría esférica. Véamos un problema sencillo: la longitud del Sol indica el día del año en el cual estamos. Se debe calcular la declinación y la ascensión recta del Sol a partir de su longitud λ .

El primero que escribió un libro para resolver estos problemas es MENELAO DE ALEJANDRÍA (ca. 70–130). De su vida se sabe poco. PTOLOMEO registró observaciones que hizo MENELAO en Roma en el año 98. Escribió el libro *Sphaerica* en el que introduce el concepto de triángulo esférico.

Los lados de un triángulo esférico son arcos de círculos máximos. La figura muestra un triángulo esférico ABD . El punto C es el centro de la esfera.



Para entender lo que ahora se llama el Teorema de MENELAO, hay que saber que los griegos calculaban las cuerdas en función del ángulo que las subtende.

$\text{crd}(\theta)$ denota la cuerda que se ve desde el centro del círculo bajo un ángulo θ .

En el tercer tomo de su libro demuestra una proposición, que en la actualidad se conoce como el Teorema de MENELAO.

Se intersecta el triángulo con un arco de círculo máximo y obtiene así los puntos E sobre BD , F sobre AD y G sobre AB .

El Teorema de MENELAO dice lo siguiente:

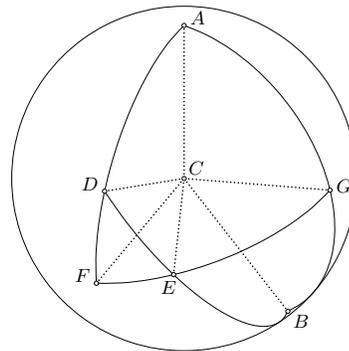
$$\frac{\text{crd}(2\angle ACG)}{\text{crd}(2\angle GCB)} = \frac{\text{crd}(2\angle ACF)}{\text{crd}(2\angle FCD)} \cdot \frac{\text{crd}(2\angle DCE)}{\text{crd}(2\angle ECB)}$$

Hoy en día usaríamos la función seno, en vez de las cuerdas y el Teorema se ve así:

$$\frac{\text{sen}(\angle ACF)}{\text{sen}(\angle FCD)} \cdot \frac{\text{sen}(\angle DCE)}{\text{sen}(\angle ECB)} \cdot \frac{\text{sen}(\angle BCG)}{\text{sen}(\angle GCA)} = 1.$$

Similarmente demuestra

$$\frac{\text{crd}(2\angle ACB)}{\text{crd}(2\angle ACG)} = \frac{\text{crd}(2\angle DCB)}{\text{crd}(2\angle DCE)} \cdot \frac{\text{crd}(2\angle FCE)}{\text{crd}(2\angle FCG)}. \tag{5.19}$$

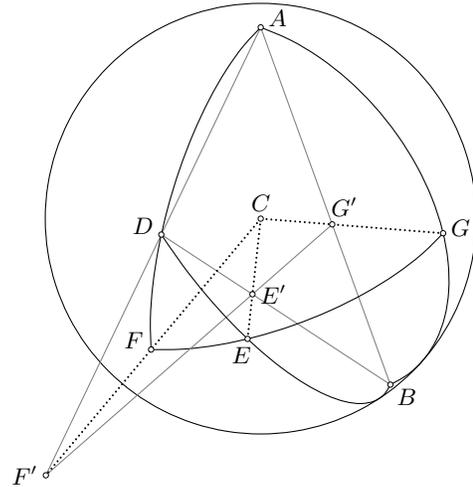


Para demostrarlo, considera el plano por los puntos del triángulo original ABD y proyecta los otros tres puntos desde C a este plano. Se obtienen así los puntos E' , F' y G' .

MENELAO usa sin demostrar la afirmación análoga en el plano:

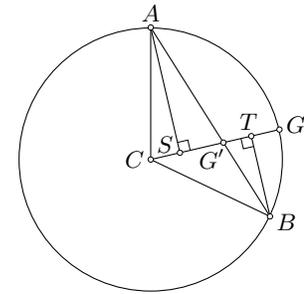
$$\frac{AG'}{G'B} = \frac{AF'}{F'D} \cdot \frac{DE'}{E'B}.$$

El hecho de que la usó sin demostrar indica que ya se conocía antes. Pero como es la primera fuente donde aparece, se considera hoy también como Teorema de MENELAO. La demostración es fácil y elemental al usar homotecias.



Luego usa que $\frac{AG'}{G'B} = \frac{\text{crd}(2\angle ACG)}{\text{crd}(2\angle GCB)}$. Esto se ve al considerar el círculo que pasa por A , G' y B : Se erigen perpendiculares a CG por A y por B , entonces se obtienen triángulos semejantes ASG' y BTG' . Por ello se tiene

$$\frac{AG'}{G'B} = \frac{AS}{BT} = \frac{2AS}{2BT} = \frac{\text{crd}(2\angle ACG)}{\text{crd}(2\angle GCB)}.$$



Similarmente se obtiene $\frac{AF'}{F'D} = \frac{\text{crd}(2\angle ACF)}{\text{crd}(2\angle FCD)}$ y también $\frac{DE'}{E'B} = \frac{\text{crd}(2\angle BCG)}{\text{crd}(2\angle GCA)}$. Con ello queda demostrado el Teorema de MENELAO.

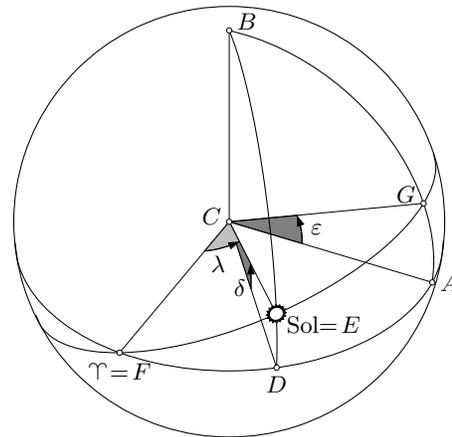
Véamos ahora cómo se aplica este último resultado para determinar la declinación δ del Sol a partir de su latitud λ .

Para ello se usa el polo norte celeste B , el Sol será el punto E y $F = \Upsilon$ el punto vernal de primavera. El círculo máximo por F y D es el ecuador y aquel por F y E la eclíptica. Además se usa el círculo máximo que es perpendicular al ecuador y a la eclíptica y que intersecta a estos en A y G respectivamente.

Entonces se tiene la situación del Teorema de MENELAO. Nótese que $\angle ACG = \varepsilon = 23.44^\circ$ y que $\angle ACB = \angle FCG = \angle DCB = 90^\circ$. Al sustituirlo en (5.19) se obtiene

$$\frac{\text{crd}(180^\circ)}{\text{crd}(2\varepsilon)} = \frac{\text{crd}(180^\circ)}{\text{crd}(2\delta)} \cdot \frac{\text{crd}(2\lambda)}{\text{crd}(180^\circ)},$$

de donde se obtiene:



$$\text{crd}(2\delta) = \frac{\text{crd}(2\varepsilon) \cdot \text{crd}(2\lambda)}{2r}.$$

En la notación moderna esto es

$$\text{sen}(\delta) = \text{sen}(\varepsilon) \text{sen}(\lambda).$$

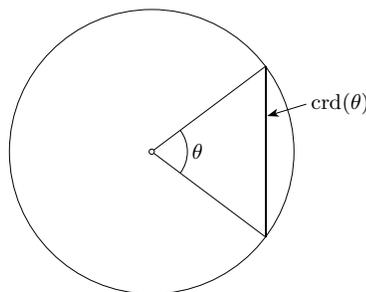
Las primeras tablas trigonométricas

Del primero que se sabe que construyó algo parecido a lo que hoy llamamos *tablas trigonométricas* fue HIPARCO DE NICEA (190–120 a.e.c). A HIPARCO se le considera uno de los más grandes astrónomos de la antigüedad. Desarrolló parte de la trigonometría esférica, la cual consideraremos más adelante en este capítulo, y se le conoce por haber descubierto la precesión de los equinoccios, es decir que el eje de rotación de la Tierra no está fijo sino que gira lentamente. Sin duda HIPARCO tenía una extrema destreza matemática.



HIPARCO DE NICEA

Se sabe que calculó una tabla de cuerdas, es decir, para diferentes ángulos θ , la longitud de la cuerda que abarca el ángulo θ en un círculo de cierto radio.



$\text{crd}(\theta)$ denota la cuerda que se ve desde el centro del círculo bajo un ángulo θ .

Lamentablemente casi todos los trabajos de HIPARCO se perdieron. Se piensa que la siguiente tabla podría ser una buena reconstrucción de su tabla de cuerdas.

θ	$\text{crd}(\theta)$	θ	$\text{crd}(\theta)$	θ	$\text{crd}(\theta)$	θ	$\text{crd}(\theta)$	θ	$\text{crd}(\theta)$
0°	0	$37^\circ 30'$	2210	75°	4186	$112^\circ 30'$	5717	150°	6641
$7^\circ 30'$	450	45°	2631	$82^\circ 30'$	4533	120°	5954	$157^\circ 30'$	6743
15°	897	$52^\circ 30'$	3041	90°	4862	$127^\circ 30'$	6166	165°	6817
$22^\circ 30'$	1341	60°	3438	$97^\circ 30'$	5169	135°	6352	$172^\circ 30'$	6861
30°	1780	$67^\circ 30'$	3820	105°	5455	$142^\circ 30'$	6511	180°	6876

Hay varios aspectos que deben observarse. Los cálculos se hacían sobre un círculo con un radio muy grande. Podemos obtener el radio de ese círculo usando $r = \text{crd}(60^\circ)$, que arroja el valor $r = 3438$. Los valores de $\text{crd}(\theta)$ aparecen redondeados a números enteros. La tabla no resulta muy útil para fines prácticos ya que la diferencia entre dos ángulos consecutivos en ella es de $7^\circ 30'$, que es un intervalo demasiado grande para realizar cálculos con buena precisión. Probablemente se usaba combinada con interpolaciones lineales.

No sabemos cómo llegó HIPARCO a estos valores. Sobre el número 3438 hay una especulación: Si se subdivide el ángulo completo de 360° en minutos se obtiene 21 600 minutos y

$$\frac{21\,600}{2\pi} \approx 3\,438.$$

Esto permite comparar la longitud de arcos medidos en minutos directamente con la longitud de las cuerdas. Por ejemplo, la cuerda correspondiente al arco de $60^\circ = 3600'$ tiene una longitud de 3438, mientras el arco tiene una longitud de 3600.

Del primero que tenemos indicios claros de cómo calculó los valores en una tabla de cuerdas es de CLAUDIO PTOLOMEO, (98–162).



CLAUDIUS PTOLOMEO

Se sabe poco de la vida de PTOLOMEO. Trabajó como astrónomo en Alejandría y elaboró un atlas de las estrellas. Para ello retomó datos de HIPARCO. Se sabe que hizo observaciones del año 127 al 141. El nombre Claudio, en latín Claudius, indica que su familia, de ascendencia griega, era respetada por considerárseles ciudadanos romanos. Escribió un libro cuyo título original era “La Compilación Matemática”, pero pronto se le llamó “La Gran Compilación”. Al traducirlo al árabe se le nombró “al-majisti” y hoy lo conocemos como el “Almagest”. En 13 tomos describe el movimiento del Sol, la Luna y los planetas conocidos en aquel entonces, todo ello basado en una teoría geocéntrica, que

perduró hasta que Copernico redescubriera la teoría heliocéntrica de ARISTARCO.

También escribió un libro llamado “Geografía” en donde describe, en 8 tomos, cómo hacer un mapa y da la ubicación de los lugares más importantes del mundo conocido entonces, con sus longitudes y latitudes. Por mucho tiempo, la suya fue la única manera conocida de hacer un mapamundi.

Para elaborar una mejor tabla de cuerdas, PTOLOMEO usa un resultado que ahora lleva su nombre: el *Teorema de PTOLOMEO*, aunque es muy posible que éste fuera conocido desde antes: En un cuadrilátero $ABCD$ inscrito en un círculo, el producto de las diagonales $AC \cdot BD$ es igual a la suma de los productos de los lados opuestos $AB \cdot CD + BC \cdot DA$, es decir

$$AC \cdot BD = AB \cdot CD + BC \cdot DA.$$

El Teorema de PITÁGORAS resulta un caso particular de este teorema, al tomar el cuadrilátero como rectángulo. El Teorema de PTOLOMEO se puede demostrar de manera elemental, como veremos a continuación.

Sea E sobre la diagonal AC aquel punto que satisfaga $\sphericalangle CBE = \sphericalangle DBA$. De ahí se sigue que $\sphericalangle CBD = \sphericalangle EBA$, que son los ángulos indicados con un arco sencillo en la figura.

Además se tiene que $\sphericalangle BAC = \sphericalangle BDC$, dado que son ángulos sobre la cuerda BC . Estos ángulos se muestran con arcos dobles en la figura. Similarmente se tiene $\sphericalangle ACB = \sphericalangle ADB$, dado que son ángulos sobre la cuerda AB .

De ahí se sigue que los triángulos BAE y BDC son semejantes y también lo son BCE y BDA , ya que coinciden en los tres ángulos.

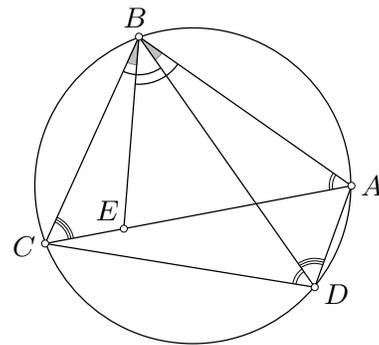
Por tanto

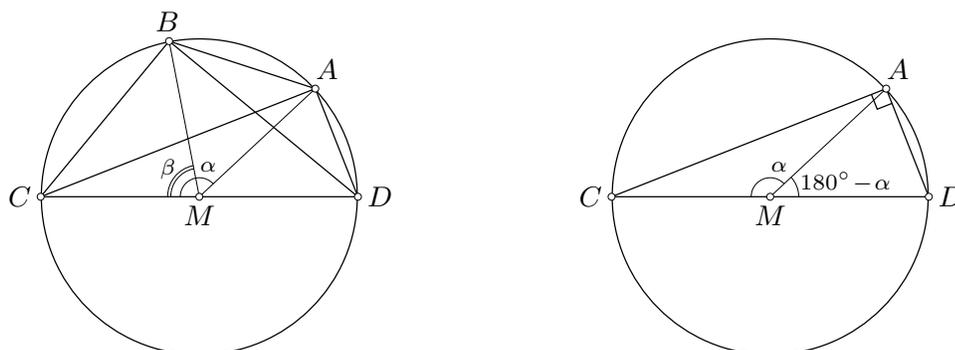
$$\frac{AB}{AE} = \frac{DB}{DC} \quad \text{y} \quad \frac{BC}{CE} = \frac{BD}{DA}$$

de donde

$$\begin{aligned} AB \cdot CD &= BD \cdot AE \\ BC \cdot DA &= BD \cdot CE \\ AB \cdot CD + BC \cdot DA &= BD \cdot (AE + CE) = BD \cdot AC. \end{aligned}$$

Con este Teorema, PTOLOMEO podía calcular la cuerda correspondiente a la diferencia de dos ángulos cuyas cuerdas ya conocía, es decir, si ya había calculado $a = \text{crd}(\alpha)$ y $b = \text{crd}(\beta)$ entonces podía obtener $\text{crd}(\alpha - \beta)$. Para ver esto consideramos un diámetro CD de un círculo con centro M y colocamos A y B sobre la circunferencia de tal manera que $\sphericalangle CMA = \alpha$ y $\sphericalangle CMB = \beta$, como se puede ver en la figura del lado izquierdo.





Se tiene ahora que

$$\begin{aligned} AB &= \text{crd}(\alpha - \beta), & AC &= \text{crd}(\alpha), & BC &= \text{crd}(\beta), \\ CD &= 2R, & DA &= \text{crd}(180^\circ - \alpha), & BD &= \text{crd}(180^\circ - \beta). \end{aligned}$$

Nótese que $\text{crd}(180^\circ - \alpha)$ y $\text{crd}(180^\circ - \beta)$ se pueden calcular con el Teorema de PITÁGORAS – véase la figura anterior a la derecha – como sigue:

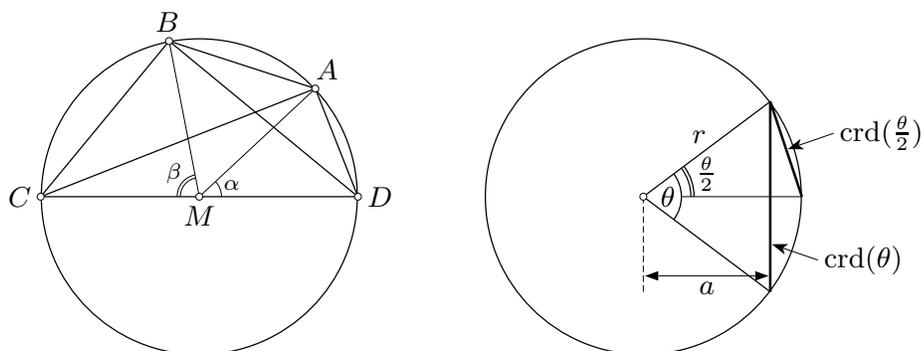
$$\text{crd}(180^\circ - \alpha) = \sqrt{4r^2 - \text{crd}(\alpha)^2}.$$

De la relación de PTOLOMEO se sigue que

$$\text{crd}(\alpha - \beta) = AB = \frac{AC \cdot BD - BC \cdot DA}{CD} = \frac{\text{crd}(\alpha) \text{crd}(180^\circ - \beta) - \text{crd}(\beta) \text{crd}(180^\circ - \alpha)}{2r},$$

donde el lado derecho se puede calcular cuando se conocen $\text{crd}(\alpha)$ y $\text{crd}(\beta)$.

Un argumento completamente similar muestra que se puede calcular $\text{crd}(180^\circ - (\alpha + \beta))$, y por lo tanto $\text{crd}(\alpha + \beta)$, si se conocen $\text{crd}(\alpha)$ y $\text{crd}(\beta)$. Para ello, considerando la siguiente situación



se obtiene la fórmula

$$\text{crd}(180^\circ - (\alpha + \beta)) = AB = \frac{\text{crd}(180^\circ - \alpha) \text{crd}(180^\circ - \beta) - \text{crd}(\alpha) \text{crd}(\beta)}{2r}.$$

Además, desde la época de ARQUÍMEDES, se sabía cómo calcular el apotema de una cuerda y de ahí la cuerda de la mitad del ángulo dado – ver la figura anterior derecha. Este cálculo queda así:

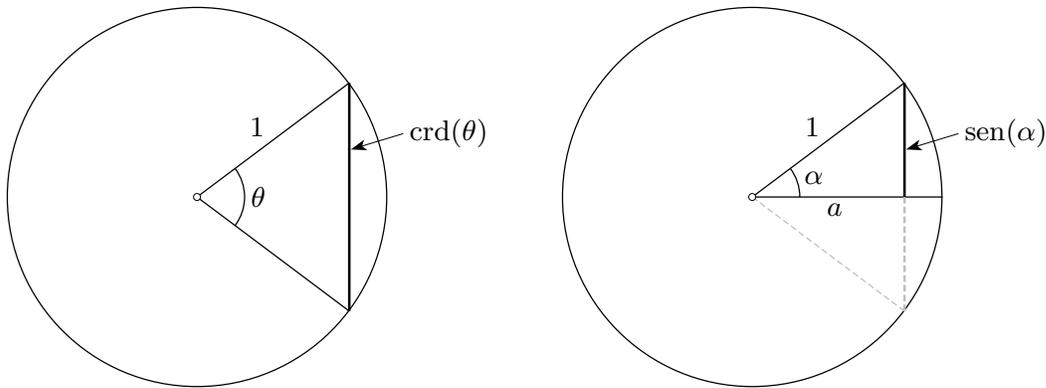
$$a = \sqrt{r^2 - \left(\frac{\text{crd}(\theta)}{2}\right)^2},$$

$$\text{crd}\left(\frac{\theta}{2}\right) = \sqrt{\left(\frac{\text{crd}(\theta)}{2}\right)^2 + (1-a)^2}.$$

Los griegos conocían $\text{crd}(72^\circ)$ y $\text{crd}(60^\circ)$. A partir de esto podían obtener $\text{crd}(18^\circ)$ y $\text{crd}(15^\circ)$. Con la fórmula para la sustracción de ángulos podían calcular $\text{crd}(3^\circ)$ y con la fórmula para la adición podían obtener $\text{crd}(k \cdot 3^\circ)$ para cualquier número natural $k = 0, 1, \dots, 60$.

Finalmente PTOLOMEO estimó $\text{crd}(1^\circ)$ y elaboró después una tabla de cuerdas para los múltiplos de medio grado, es decir para los ángulos $0^\circ, 30', 1^\circ, \dots, 179^\circ 30', 180^\circ$. Su tabla daba las cuerdas en el sistema sexagesimal, no en enteros como la de Hiparco.

En la India se empezaron a calcular tablas del seno a partir del siglo V. El primero en hacerlo fue ARYABHATA (ca. 476–550). Se elaboraron sofisticados algoritmos para calcular muchos valores a partir de unos pocos, similarmente a lo que vimos para la tabla de cuerdas de PTOLOMEO. La relación entre las dos tablas es bastante directa.



Se tiene que

$$\text{crd}(\theta) = 2 \text{sen}\left(\frac{\theta}{2}\right) \quad \text{y} \quad \text{sen}(\alpha) = \frac{1}{2} \text{crd}(2\alpha).$$

Por ello realmente no importa cuál de las dos tablas se usa. Hoy día ya sólo se conoce el seno, las cuerdas han pasado al olvido. Además se definieron nuevas funciones como el coseno

$$\cos(\alpha) = a, \quad \text{y} \quad \text{vers}(\alpha) = 1 - a$$

donde a es el apotema en la figura de arriba a la derecha. Si reemplazamos las cuerdas por los senos en la fórmula

$$\text{crd}(\alpha - \beta) = \frac{\text{crd}(\alpha) \text{crd}(180^\circ - \beta) - \text{crd}(\beta) \text{crd}(180^\circ - \alpha)}{2r}$$

y usamos $\varphi = \frac{\alpha}{2}$, $\psi = \frac{\beta}{2}$ y $r = 1$, obtenemos:

$$2 \text{sen}(\varphi - \psi) = \frac{2 \text{sen}(\varphi) 2 \cos(\psi) - 2 \text{sen}(\beta) 2 \cos(\alpha)}{2}$$

porque $\text{crd}(180^\circ - \beta) = 2 \text{sen}(\frac{180^\circ - \beta}{2}) = 2 \text{sen}(90^\circ - \psi) = 2 \text{cos}(\psi)$. De ahí se sigue que

$$\text{sen}(\varphi - \psi) = \text{sen}(\varphi) \text{cos}(\psi) - \text{sen}(\beta) \text{cos}(\alpha),$$

que es la fórmula conocida del seno de la diferencia de dos ángulos. De la fórmula para $\text{crd}(\alpha + \beta)$ se obtiene análogamente

$$\text{cos}(\varphi + \psi) = \text{cos}(\varphi) \text{cos}(\psi) - \text{sin}(\varphi) \text{sin}(\psi).$$

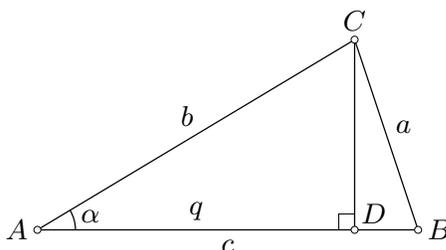
Ley de los senos y Ley del coseno

Estas dos leyes forman parte esencial de la trigonometría moderna. Pero son de procedencia relativamente reciente.

La ley del coseno aparece por primera vez en *Los Elementos* de EUCLIDES como Proposición II.13 en la siguiente forma:

En los triángulos acutángulos, el cuadrado del lado que subtiende al ángulo agudo es menor que [la suma de] los cuadrados de los lados que comprenden el ángulo agudo en dos veces el rectángulo comprendido por uno de los lados del ángulo agudo sobre el que cae la perpendicular y la [recta] interior cortada por la perpendicular hasta el ángulo agudo.

La redacción se aclara con un dibujo.



Sea α un ángulo agudo del triángulo ABC . En el texto, el cuadrado del lado que subtiende el ángulo agudo, a^2 , es igual a la suma de los cuadrados de los lados adyacentes, esto es $b^2 + c^2$, disminuido por el doble de un rectángulo. Un lado de ese rectángulo es uno de los lados adyacentes, tomamos c , y el otro es lo que la altura corta hacia el ángulo agudo, es decir $q = AD$ en el dibujo anterior. Escrito como una fórmula esto queda así

$$a^2 = b^2 + c^2 - 2cq.$$

Este resultado no se considera parte de la trigonometría, es sólo una comparación entre áreas. Sin embargo, si se observa que $q = b \text{cos}(\alpha)$, da lo que hoy llamamos Ley del coseno

$$a^2 = b^2 + c^2 - 2bc \text{cos}(\alpha).$$

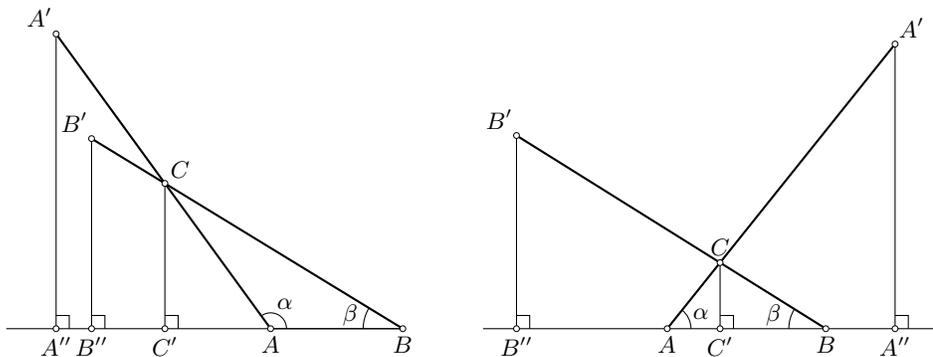
La ley de los senos aparece explícitamente en la literatura en el libro *Tratamiento sobre el cuadrilátero* de NASIR AL-DIN AL-TUSI (1201–1274). Al-Tusi nació en Tus, una ciudad en el norte de lo que hoy es Iran. Su tío le enseñó lógica y física a una edad temprana. Se mudó a Nishapur y allí estudió filosofía, medicina y matemáticas. Vivió en un tiempo de guerra: los mongoles invadían

grandes espacios, sembraban la destrucción y cometían terribles crueldades. AL-TUSI aceptó una invitación a unirse a los “asesinos”, una secta islámica con fuerza militar que defendía las fortalezas en esa región. En esa época al-Tusi escribió varios textos sobre filosofía, lógica, matemáticas y astronomía. Cuando los mongoles bajo el mando de Hulegu destruyeron el castillo Alamut, incorporaron a al-Tusi en sus tropas como asesor científico. Así, AL-TUSI acompañó a Hulegu en la conquista de Bagdad y le propuso construir un observatorio. En 1262 se completó el observatorio y AL-TUSI empezó a elaborar tablas muy precisas sobre el movimiento de los planetas. Tradujo varios textos del griego al árabe.



NASIR AL-DIN AL-TUSI

La demostración de la Ley de los senos de AL-TUSI es como sigue. Primero se prolonga AC sobre A' de tal manera que AA' mida r . Luego se erige una perpendicular $A'A''$ a AB por A' con pie A'' . Por definición se tiene $A'A'' = \text{sen}(\alpha)$. Lo mismo se hace con B y se obtiene $B'B'' = \text{sen}(\beta)$.



Esto es válido aún si uno de los ángulos es obtuso – como se ve en la ilustración anterior izquierda – dado que $\text{sen}(\varphi) = \text{sen}(180^\circ - \varphi)$. Sea además CC' la perpendicular a AB por C . Entonces los triángulos $AC'C$ y $AA''A'$ son semejantes y por ello se tiene $A'A'' : A'A = CC' : CA$, es decir $\text{sen}(\alpha) \cdot AB = r \cdot CC'$. De manera similar $BC'C$ y $BB''B'$ son semejantes, por lo que $\text{sen}(\beta) \cdot BC = r \cdot CC'$. Al igualar los lados derechos se obtiene $\text{sen}(\alpha) \cdot AB = \text{sen}(\beta) \cdot BC$, es decir

$$\frac{\text{sen}(\alpha)}{\text{sen}(\beta)} = \frac{BC}{AC}.$$

La demostración de AL-TUSI es similar a la moderna, ambas se basan en calcular la altura CC' de dos maneras, usando los triángulos rectángulos $AC'C$ y $BC'C$. Pero AL-TUSI primero amplía estos triángulos para que su hipotenusa mida r . Esto deja entrever que $\text{sen}(\varphi)$ no era para él una proporción sino una longitud en un círculo de radio r .

Una vez que la comprensión de las funciones seno y coseno queda claramente establecida – sea esto en el sentido de hoy en un círculo de radio 1, o como se hizo antes en un círculo con un radio grande como $r = 3438$ – resulta inmediato interpretar el enunciado de la Proposición II.13 de *Los Elementos* de Euclides como lo que en la actualidad conocemos como la Ley del coseno.

Se reconoce a Hiparco como padre de la trigonometría porque fue el primero en concebir la importancia de conocer las longitudes de las cuerdas en función del ángulo. Con eso la astronomía obtuvo una herramienta muy poderosa. Pero fue AL-TUSI, quien escribió el primer libro sobre trigonometría y con ello la elevó a un asunto de interés matemático, independiente de la astronomía.

Mejoramiento de las tablas

En la Edad Media se reanudaron los esfuerzos para calcular tablas cada vez más precisas del seno. Las herramientas eran esencialmente las mismas que las que ya vimos en funcionamiento en manos de PTOLOMEO. Para mantener las fórmulas lo más sencillas posible y cercanas a lo que se conoce hoy día, adoptamos esas herramientas aquí con $r = 1$, aunque en la Edad Media esto no era común:

- (a) Calcular $\cos(\varphi)$ a partir de $\text{sen}(\varphi)$ según la fórmula

$$\cos(\varphi) = \sqrt{1 - \text{sen}(\varphi)^2}.$$

- (b) Calcular $\text{sen}(\frac{\varphi}{2})$ a partir de $\text{sen}(\varphi)$, según (a) y la fórmula

$$\text{sen}(\frac{\varphi}{2}) = \sqrt{\frac{1 - \cos(\varphi)}{2}}.$$

- (c) Calcular $\text{sen}(\varphi + \psi)$ a partir de $\text{sen}(\varphi)$ y $\text{sen}(\psi)$, según la fórmula

$$\text{sen}(\varphi + \psi) = \text{sen}(\varphi) \cos(\psi) + \cos(\varphi) \text{sen}(\psi).$$

- (d) Calcular $\text{sen}(\varphi - \psi)$ a partir de $\text{sen}(\varphi)$ y $\text{sen}(\psi)$, según la fórmula

$$\text{sen}(\varphi - \psi) = \text{sen}(\varphi) \cos(\psi) - \cos(\varphi) \text{sen}(\psi).$$

Veamos cómo RICHARD OF WALLINGFORD (1292–1336) describió cómo calcular $\text{sen}(1^\circ)$ con la precisión que uno desee. Se parte de que uno conoce $\text{sen}(3^\circ)$. Para ello se usa (d).

1. A partir de $\text{sen}(3^\circ)$ se calcula $\text{sen}(\frac{1}{16}3^\circ)$ usando cuatro veces (b).
2. Se calcula $\text{sen}(5 \cdot \frac{3}{16}^\circ) = \text{sen}(\frac{15}{16}^\circ)$ usando varias veces (c).
3. Se calcula $\text{sen}(\frac{1}{4} \cdot \frac{3}{16}^\circ) = \text{sen}(\frac{3}{64}^\circ)$ usando (b) dos veces.
4. Se calcula $\text{sen}(\frac{3}{64}^\circ + \frac{15}{16}^\circ) = \text{sen}(\frac{63}{64}^\circ)$ usando (c).

5. Se calcula $\text{sen}\left(\frac{1}{4} \cdot \frac{3}{64}^\circ\right) = \text{sen}\left(\frac{3}{256}^\circ\right)$ usando (b) dos veces.
 6. Se calcula $\text{sen}\left(\frac{3}{256}^\circ + \frac{63}{64}^\circ\right) = \text{sen}\left(\frac{255}{256}^\circ\right)$.

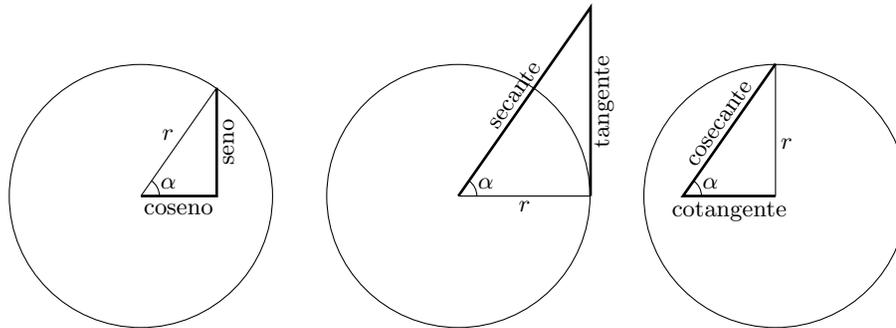
Este proceso se puede continuar tantas veces uno quiera. En este proceso se usa que

$$\frac{3}{4^{n+1}} + \frac{4^n - 1}{4^n} = \frac{3}{4^{n+1}} + \frac{4 \cdot (4^n - 1)}{4^{n+1}} = \frac{3 + 4^{n+1} - 4}{4^{n+1}} = \frac{4^{n+1} - 1}{4^{n+1}}.$$

Conforme n aumenta, se obtienen cada vez mejores aproximaciones de 1.

En 1596 se publicó el libro *Opus Palatinum de Triangulis* (Obra palatina sobre triángulos) de GEORG JOACHIM DE PORRIS (1514–1574), también conocido como RHETICUS, completado 22 años después de la muerte de Rheticus por su discípulo LUCIUS VALENTON OTHO. Rheticus nació en Feldkirch, en el extremo Oeste de Austria, estudió en Zúrich y después con COPÉRNICO y es considerado su único discípulo.

Elaboró tablas de las seis funciones trigonométricas que indicamos en el siguiente esquema, aunque no usó estos nombres.



Él y su discípulo usaron un radio $r = 10\,000\,000\,000$ y calcularon los valores de las seis funciones en intervalos de $10'' = \frac{1}{360}^\circ$. La obra consta de 1400 páginas tamaño folio.

Una implementación en las computadoras: el método CORDIC

La función seno es una función *trascendente*, es decir, es imposible calcular $\text{sen}(x)$ con exactitud en un número finito de pasos, salvo para valores especiales como por ejemplo $x = 0$.

Con la invención de las computadoras en los años 50 del siglo pasado, se produjo una revolución en cómo se calcula. Las tablas de logaritmos, seno y tangente desaparecieron y se hicieron común el uso de calculadoras de bolsillo. Pero ¿cómo calcula una calculadora el valor de $\text{sen}(20^\circ)$?

Una idea sería expresar $\text{sen}(x)$ como una serie de potencias:

$$\text{sen}(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \frac{x^{11}}{11!} + \dots,$$

donde $n!$ es la *factorial* de n , que está definida como el producto de los primeros n números naturales, es decir, $n! = 1 \cdot 2 \cdot \dots \cdot (n-1) \cdot n$. Es importante que aquí se use *radianes*, es decir la longitud del arco de un sector con ángulo x y radio 1.

Como $\sin(\pi - x) = \sin(x)$ y $\sin(x + 2\pi) = \sin(x)$ es suficiente calcular $\sin(x)$ para $-\frac{\pi}{2} \leq x \leq \frac{\pi}{2}$. Por ello basta que se tome los primeros 6 sumandos en la serie arriba para obtener una exactitud de 8 dígitos.

Pero para las calculadoras, llevar a cabo una multiplicación es “costoso”, es decir cuesta mucho tiempo. Con el *esquema de Horner* de WILLIAM GEORGE HORNER (1789–1837) se reduce el número de multiplicaciones considerablemente:

$$\begin{aligned}\sin(x) &\approx x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \frac{x^{11}}{11!} \\ &= x \cdot \left(1 - x^2 \cdot \left(\frac{1}{3!} - x^2 \cdot \left(\frac{1}{5!} - x^2 \cdot \left(\frac{1}{7!} - x^2 \cdot \left(\frac{1}{9!} - \frac{1}{11!}x^2\right)\right)\right)\right)\right)\end{aligned}$$

Hay que calcular x^2 una vez y luego hay que hacer 5 multiplicaciones. Los valores de $\frac{1}{3!}, \dots, \frac{1}{11!}$ se pueden calcular de antemano y guardar como constantes.

Pero en 1959, aún eso era demasiado lento y se buscaba una manera de hacerlo sin multiplicación alguna, excepto multiplicación o división por 2. En las computadoras y calculadoras de bolsillo, los números representan en el *sistema binario*. En este sistema el número decimal 10 se expresa como 1010_2 ya que

$$10 = 1 \cdot 2^3 + 0 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0.$$

El número 0.2 se expresa como $0.00110011\overline{0011}_2$. El subíndice 2 indica que se trabaja en base 2 y no en la base 10.

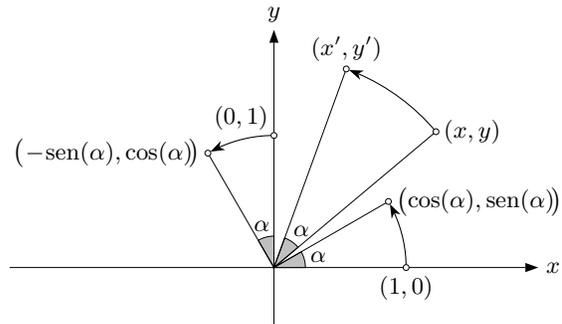
En el sistema binario la multiplicación o la división con 2 no cuesta trabajo, ya que es simplemente es movimiento a la izquierda resp. a la derecha de todos los dígitos.

En 1959 el ingeniero Jack E. Volder desarrolló un algoritmo que lograra calcular $\sin(x)$ sin hacer una sola multiplicación excepto con 2. El método fue bautizado CORDIC que es un acrónimo de coordinate rotation digital computer.

El nombre expresa parte de la idea: se calcula la rotación en coordenadas.

Se tiene que al rotar por el ángulo α en contra de las manecillas del reloj el punto $(1, 0)$ da al punto $(\cos(\alpha), \sin(\alpha))$ y el punto $(0, 1)$ se rota al punto $(-\sin(\alpha), \cos(\alpha))$. Un punto en general (x, y) se rota en (x', y') donde

$$\begin{aligned}x' &= \cos(\alpha)x - \sin(\alpha)y, \\ y' &= \sin(\alpha)x + \cos(\alpha)y.\end{aligned}$$



Esto se puede expresar de manera compacta al usar el lenguaje de matrices:

$$\underbrace{\begin{bmatrix} x' \\ y' \end{bmatrix}}_{p'} = \underbrace{\begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}}_{M_\alpha} \cdot \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_p.$$

Algunas propiedades que serán importantes:

- Las rotaciones se pueden acumular: $M_\beta M_\alpha = M_{\alpha+\beta}$.
- Las rotaciones son lineales, en particular se tiene: $M_\alpha(\lambda p) = \lambda(M_\alpha p)$, para cada punto p y cada número λ .

Volder conjuntó varias ideas para lograr calcular el valor de $\text{sen}(\alpha)$ sin usar multiplicaciones excepto con potencias de 2:

- En vez de M_α se usa

$$R_\alpha = \frac{1}{\cos(\alpha)} M_\alpha = \frac{1}{\cos(\alpha)} \begin{bmatrix} 1 & -\tan(\alpha) \\ \tan(\alpha) & 1 \end{bmatrix};$$

- Se usan ángulos especiales $\varphi_0, \varphi_1, \dots$ para los cuales se tiene $\tan(\varphi_k) = \frac{1}{2^k}$. Estos ángulos se llaman *ángulos de CORDIC* y se pueden calcular de antemano y guardar en constantes. Para un ángulo de CORDIC φ_k se tiene $\cos(\varphi_k) = \cos(\tan^{-1}(\frac{1}{2^k})) = \frac{1}{\sqrt{1+(\frac{1}{2^k})^2}} = \kappa_k$

$$R_k = \frac{1}{\kappa_k} \begin{bmatrix} 1 & -\frac{1}{2^k} \\ \frac{1}{2^k} & 1 \end{bmatrix}.$$

Por las propiedades mencionadas se tiene

$$R_k R_\ell = \frac{1}{\kappa_k \kappa_\ell} \begin{bmatrix} 1 & -\frac{1}{2^k} \\ \frac{1}{2^k} & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2^\ell} \\ \frac{1}{2^\ell} & 1 \end{bmatrix}.$$

- El ángulo α del cual se debe calcular el valor $\text{sen}(\alpha)$, se aproxima como

$$\alpha \approx \varepsilon_0 \varphi_0 + \varepsilon_1 \varphi_1 + \varepsilon_2 \varphi_2 + \dots + \varepsilon_n \varphi_n,$$

donde $\varepsilon_k = \pm 1$ para $k = 0, 1, \dots, n$. Además, el número n se puede calcular de antemano, ya que de él depende la precisión del valor que se calculará.

Si se juntan estas ideas se obtiene

$$\begin{bmatrix} \cos(\alpha) \\ \text{sen}(\alpha) \end{bmatrix} \approx \frac{1}{\kappa_0 \kappa_1 \kappa_2 \dots \kappa_n} \begin{bmatrix} 1 & -\varepsilon_n \frac{1}{2^n} \\ \varepsilon_n \frac{1}{2^n} & 1 \end{bmatrix} \dots \begin{bmatrix} 1 & -\varepsilon_1 \frac{1}{2^1} \\ \varepsilon_1 \frac{1}{2^1} & 1 \end{bmatrix} \begin{bmatrix} 1 & -\varepsilon_0 \frac{1}{2^0} \\ \varepsilon_0 \frac{1}{2^0} & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Se calcula $K_n = 1/\kappa_0 \kappa_1 \dots \kappa_n$ de antemano y entonces se tiene

$$\begin{bmatrix} \cos(\alpha) \\ \text{sen}(\alpha) \end{bmatrix} \approx \begin{bmatrix} 1 & -\varepsilon_n \frac{1}{2^n} \\ \varepsilon_n \frac{1}{2^n} & 1 \end{bmatrix} \dots \begin{bmatrix} 1 & -\varepsilon_1 \frac{1}{2^1} \\ \varepsilon_1 \frac{1}{2^1} & 1 \end{bmatrix} \begin{bmatrix} 1 & -\varepsilon_0 \frac{1}{2^0} \\ \varepsilon_0 \frac{1}{2^0} & 1 \end{bmatrix} \begin{bmatrix} K_n \\ 0 \end{bmatrix}.$$

Falta explicar cómo se calculan los signos $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_n$. Se pone

$$\varepsilon_0 = \begin{cases} 1, & \text{si } \alpha \geq 0, \\ -1, & \text{si } \alpha < 0, \end{cases}$$

y luego de manera iterativo

$$\varepsilon_{k+1} = \begin{cases} 1, & \text{si } \alpha \geq \varepsilon_0\varphi_0 + \dots + \varepsilon_k\varphi_k, \\ -1, & \text{si } \alpha < \varepsilon_0\varphi_0 + \dots + \varepsilon_k\varphi_k. \end{cases}$$

El siguiente pseudo-código muestra el algoritmo

```

cordic(alpha):
  a=0
  x=K_n
  y=0
  for k=0 to n
    if alpha < a then
      e=-1
    else
      e=1
    endif
    a=a+e·φ_k
    x=x-e·2-k·y
    y=y+e·2-k·x
    k=k+1
  endfor
  cos(alpha)=x
  sen(alpha)=y

```

La idea de Volder luego se amplió a que puede calcular también exponenciales, logaritmos y otras funciones trascendentales. Fue un éxito de tal manera que hoy día ya no nos damos cuenta de cómo los aparatos que nos rodean llevan a cabo los cálculos de manera eficiente y confiable. De alguna manera, las matemáticas se alejaron de nosotros al ser empacados en el interior íntimo de los aparatos.

Bibliografía para la sección de trigonometría esférica

Glen van Brummelen: The mathematics of the heavens and earth, the early history of trigonometry. Princeton University Press, 2009.

J. L. Berggren: Episodes in the Mathematics of medieval Islam, Springer, 1986.

Scriba, Schreiber: 5000 Jahre Geometrie, Springer 2000.

Lambacher, Schweizer: Ebene Trigonometrie, Klett 1968

MacTutor

Wikipedia inglés, alemán y español

David Maletinsky: Wie rechnen Taschenrechner?, Mentorierte Arbeit, ETH Zürich, 2009

5.9. La medición fina de la Tierra

Hace mucho tiempo, el área de un campo se medía por el tiempo que requería sembrarlo o por la cantidad de semillas requeridos. Poco a poco se vio la necesidad de poder medir con mayor precisión áreas o distancias por ejemplo entre ciudades. Este desarrollo se hizo a la par con un desarrollo en los instrumentos de medición y el desarrollo en las matemáticas, entre un desarrollo en precisión de las tablas de funciones trigonométricas.

Para las matemáticas fue un desarrollo por exigencias externos, muy mundanos, en el tejido de intereses militares, disputas territoriales, donde los científicos involucrados eran relojeros y fabricantes de instrumentos de precisión entre otros. En su conjunto, dió un conocimiento preciso del tamaño y la forma de la Tierra y aclaró que las montaña más alto no es un Pico en los Andes sino uno en Nepal, un país cerrado a extreanjeros en aquel entonces.

Al inicio, la medición de longitudes

El primer instrumento para medir distancias fue un *odómetro* en forma de un carrito, que al jalarlo transmitió el girar de las ruedas mediante un mecanismo sencillo a una rueda que giraba lentamente y así registraba la distancia recorrida. Los romanos medían de esta manera la longitud de sus calzadas de su vasta red viario romano.

La siguientes dos ilustraciones muestran un odómetro griego y uno romano, ambos reconstrucciones que se hicieron para exhibirse en un museo.



Es claro, que estos herramientas de medición no estaban pensados para hacer mediciones en terrenos poco accesibles. Para ello se requerían otras ideas.

Los primeros mapas usaron latitudes y longitudes

Fue el griego CLAUDIO PTOLOMEO (100–170) quien tuvo una idea novedosa de hacer un mapa: bastaba medir de muchos lugares la latitud y la longitud y luego colocar en una rejilla estos ligares para indicar la mayor información que da un mapa. Aquí se debe advertir que hay una diferencia muy importante entre las mediciones de latitud y longitud: la de la latitud se puede hacer con relativa facilidad, al tomar la elevación del Sol sobre el horizonte, mientras que la longitud es muy difícil de medir.

No sobrevivió ningún mapa original de PTOLOMEO, pero en la edad media se usaron las instrucciones de su libro y su tabla de latitudes y longitudes para reconstruir uno.

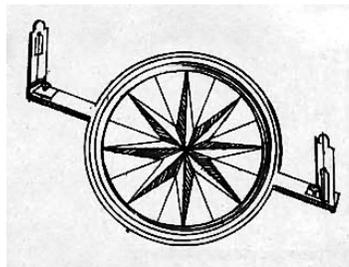


La ilustración anterior muestra una reconstrucción de Johannes de Armsshein en 1482. Pocos años después Cristobal Colón emprendió su viaje hacia el Oeste, en búsqueda de un nuevo camino hacia la India. Tener buenos mapas de los Mares era un necesidad inmediata. Estos mapas se hicieron con el método de PTOLOMEO, es decir midiendo la latitud y la longitud tan bueno como se podía.

Mediciones de ángulos

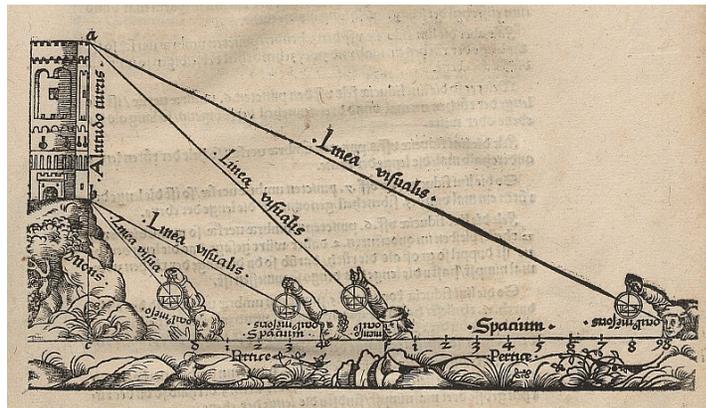
Sobre terreno terrestre, a diferencia que en altamar, hay muchos puntos marcados que se pueden tomar como referencia, y que permiten una medición más precisa. El suelo es fijo y no se tambalea como la cubierta de un barco. Por ello es posible nivelar un instrumento hasta que esté horizontal y medir con calma.

La gravedad da una referencia a lo vertical y por ello se miden ángulos de dos tipos: horizontales y de elevación. Las siguientes ilustraciones muestran dos instrumentos para ello: uno para medir ángulos horizontales y otro para medir ángulos de elevación.



ESTAS IMAGENES HAY QUE REHACER, PERO NO TUVE TIEMPO

Más tarde, sobre todo a partir del siglo XVIII, se combinaron las dos herramientas en uno: en el *teodolito*. La siguiente ilustración de 1536 es de un libro del alemán *Johannes Stöffler* y muestra el uso del *grafómetro* para medir ángulos de elevación.



Quelle: Deutsche Fotothek

Estas ideas se deben a los árabes del siglo X, sobre todo a *Ibn al-Saffar*.

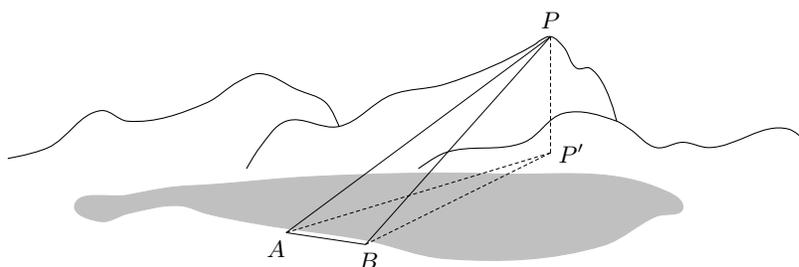
La triangulación

El matemático holandés GEMMA FRISIUS (1508-1555) publicó en 1533 en la nueva edición de su libro *Cosmographia* el método de la *triangulación*. Consiste en colocar en el terreno triángulos imaginarios con las esquinas en puntos bien visibles como picos de colinas o montañas y luego ir a varias de estos puntos y medir los ángulos horizontales y de elevación hacia los otros puntos visibles. Un lado de alguno de los triángulos debe ser horizontal y medirse con mucha precisión. La distancia horizontal hacia los otros puntos y la diferencia en la altura se puede calcular con trigonometría.



GEMMA FRISIUS

Para aclarar el método de triangulación veremos un ejemplo sencillo. Se quiere medir la altura de una montaña que es visible desde el borde de un lago. Se colocan al borde de un lago dos palos a la distancia de 500 metros. Denotamos estos dos puntos con A y B y el pico de la montaña con P . En A se mide el ángulo horizontal $\alpha = \sphericalangle BAP'$ y en B el ángulo horizontal $\beta = \sphericalangle ABP'$, donde P' denota el punto verticalmente debajo P , que está a la misma altura de A y B .



Por otro lado se miden en ambos puntos el ángulo de elevación hacia P : en A se mide $\varepsilon_A = \sphericalangle P'AP$ y en B se mide $\varepsilon_B = \sphericalangle P'BP$.

Ahora se puede calcular: en el triángulo ABP' se tiene el ángulo $\gamma = \sphericalangle AP'B = 180^\circ - \alpha - \beta$. Luego se usa el Teorema del Seno: de

$$\frac{AB}{\text{sen}(\gamma)} = \frac{BP'}{\text{sen}(\alpha)} = \frac{AP'}{\text{sen}(\beta)}$$

se obtiene la longitud de AP' y de BP' . Finalmente se considera el triángulo $AP'P$ que es rectángulo. Se calcula PP' con la función tan:

$$\tan(\varepsilon_A) = \frac{PP'}{AP'} \quad \text{y también} \quad \tan(\varepsilon_B) = \frac{PP'}{BP'}$$

Es importante notar, que no se tuvo que hacer ninguna medición en el pico de la montaña: sólo se hicieron mediciones en A y B . También debe notarse que un pequeño error en la medición de la línea de base o en los ángulos puede producir considerables errores en el resultado final. Además hay que recordar que en aquel entonces no se contaba con calculadoras de bolsillo, sino todos los cálculos se tenían que hacer a mano.

El primero que hizo una medición útil con el método de la triangulación fue el matemático y físico holandés WILLEBRORD VAN ROIJEN SNELL (Snellius, 1580–1626) alrededor de 1615. Trató de medir la longitud de un arco de un meridiano. Pero cometió algunos errores en la medición y en los cálculos, por lo que su resultado no estaba muy preciso.

La triangulación era un método que alcanzó gran popularidad: los duques, marqueses, archiduques y reyes querían conocer las extensiones de sus territorios y tener buenos mapas para fines militares. Las primeras triangulaciones se hicieron en el siglo XVII.

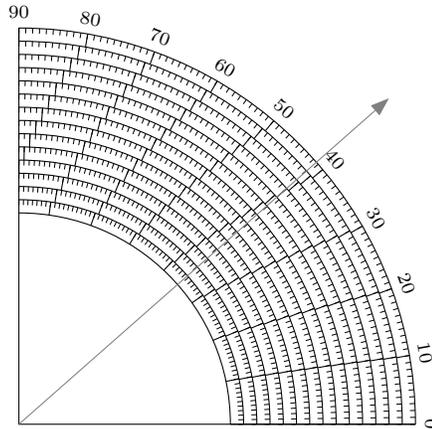
Mejoras en las escalas de medición

Los primeros instrumentos estaban poco elaborados con escalas poco finas. Por ello la precisión estaba limitada. La necesidad de aumentar la precisión desató una serie de inventos que mejoraron gradualmente los instrumentos. Estudios sobre las funciones trigonométricas permitieron calcular tablas de las funciones seno, coseno y tangente de mayor precisión. El uso de telescopios en los teodolitos aumentó nuevamente la precisión de medición que requirió tener escalas todavía más finas y tablas más precisas.

Veamos algunos de estas mejoras. El portugués PEDRO NUNES (1502-1578) inventó una mejora en poder leer escalas con mayor precisión. Para medir ángulos dejó gravar varios anillos concen-

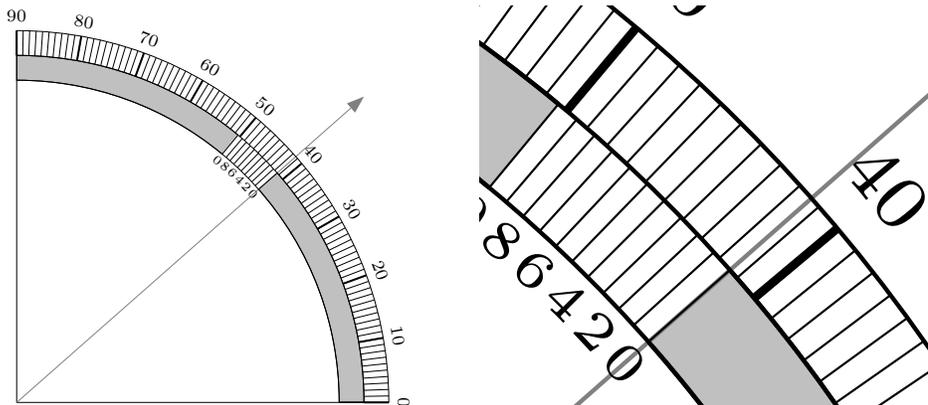
tricos, el más grande con una división de 90 divisiones el segundo interior con 89 divisiones y así sucesivamente con uno menos.

La siguiente ilustración muestra una escala de Nonius.



Las adicionales escalas – a parte de la principal que da la división en 90 – permite una determinación mucho mejor en el ángulo medido (que está indicado con una flecha gris). Pero los cálculos involucrados no son de todo fáciles. Sin embargo, TYCHO BRAHE (1564-1601) los usaba y elogió a PEDRO NUNES.

Mucho mejor por ser más fácil de entender y usar es la escala de VERNIER, que es una pequeña escala que se desliza en otra, nombrado según el francés PIERRE VERNIER (1580-1637). En la siguiente ilustración se muestra una escala de VERNIER. Para leerla, se busca cuál de sus marcas coincide mejor con una marca de la escala principal. En el ejemplo de la ilustración es la marca 4 de la escala de VERNIER. Eso quiere decir que el siguiente dígito es 4, es decir el ángulo es aproximadamente 41.4° .



El funcionamiento es sencillo: en la escala de Vernier las marcas están a la distancia 0.9 de la distancia entre las marcas principales de los grados. Supongamos que el ángulo realmente fuera 41.4° . Si avanzamos n marcas en la escala de VERNIER, entonces avanzamos por un ángulo de $n \cdot (1 - 0.1)^\circ$ en la escala de grados afuera. La pregunta que hay que contestar es ¿para cuál n la suma de 41.4° y $n \cdot (1 - 0.1)^\circ$ es un entero? Esta es la condición para que la marca de Vernier

coincida con una marca de la escala de grados. La respuesta es fácil: hay que buscar un número entero tal que

$$41.4 + n \cdot (1 - 0.1) = (41 - n) + (4 - n) \cdot 0.1$$

sea un número entero. Claramente para $n = 4$ esto es el caso, y $n = 4$ es la única solución entre los posibles valores $n = 0, 1, \dots, 9$. Por ello, hay que leer dónde coincide la marca de Vernier con la de grados y la leyenda en la escala de Vernier dará el siguiente dígito después del punto decimal.

Mejoras en los cálculos

Las tablas de las funciones trigonométricas se hicieron cada vez con mayor precisión, con 4, luego con 5 hasta con 8 dígitos. En los cálculos hay que dividir frecuentemente dos valores de Senos, por ejemplo $AP' = AB \cdot \frac{\text{sen}(\beta)}{\text{sen}(\gamma)}$, en el ejemplo de arriba. Por ello el matemático escocés JOHN NAPIER (1550–1617) tuvo en 1614 la idea de tabular los logaritmos de los valores de seno, es decir hacer una table que muestra $\log(\text{sen}(\varphi))$ para muchos valores de φ .

Más que diez años antes, el relojero suizo JOST BÜRGI (1552–1632) tuvo la idea de tabular una función exponencial y así también una logarítmica. Tenía estas tablas en el año 1600 y después de tabular la función Seno, pero no lo publicó hasta el año 1620. Veamos su idea. Tomo el número $q = 1.0001$ y calculó sucesivamente las potencias q^n . Esto es relativamente fácil, porque

$$q \cdot z = 1.0001 \cdot z = z + \frac{1}{10\,000}z,$$

es decir cada potencia es la anterior más un 10 000-avo del anterior: q^{n+1} se calcula como $q^{n+1} = q^n + \frac{1}{10\,000}q^n$. Así, BÜRGI calculó todas las potencias desde q^0 hasta que $q^n > 10$. Esto sucede para $n = 23\,028$. Además averigua que

$$q^{23\,027.0022} < 10 < q^{23\,027.0023}$$

Con sus tablas tiene una herramienta potente de simplificar los cálculos: para multiplicar dos números A y B , los escribe de la forma,

$$A = q^m \cdot 10^a, \quad \text{resp.} \quad B = q^n \cdot 10^b,$$

donde m y n los obtiene al buscar en su tabla. El producto $A \cdot B$ es de la forma $A \cdot B = q^{m+n} \cdot 10^{a+b}$ y el resultado de q^{m+n} lo obtiene por ver nuevamente en su tabla. De esta manera, Bürgi redujo la multiplicación a la adición, que a mano es mucho más fácil de llevar a cabo.

Con las tablas de BÜRGI trabajaba JOHANNES KEPLER (1571–1630) cuando trataba de encontrar la forma de las órbitas celestes. Los relojes y demás instrumentos que construía BÜRGI eran superiores en precisión e inalcanzados en 100 años después. Era el primero en construir un reloj con segundero.

Mejoras en las herramientas de medición

Durante 60 años del siglo XVIII se midió la extensión del reinado de Francia. A cargo estaban cuatro generaciones de una familia: CASSINI. El resultado era un atlas a la escala 1 : 86 400 con una precisión desconocido hasta entonces.

Durante el siglo XIX se llevaron a cabo múltiples proyectos de medición. En los teodolitos se usaron telescopios, para tener una medición aún mejor. El matemático alemán CARL FRIEDRICH GAUSS (1777–1855) fue encargado de medir el reinado de Hannover. Empezó en 1821 y terminó en 1840. Para ello inventó un instrumento nuevo: el *heliotropo*, que reflejaba la luz del Sol en una dirección determinada, para hacer mejor visible este punto desde uno lejano.

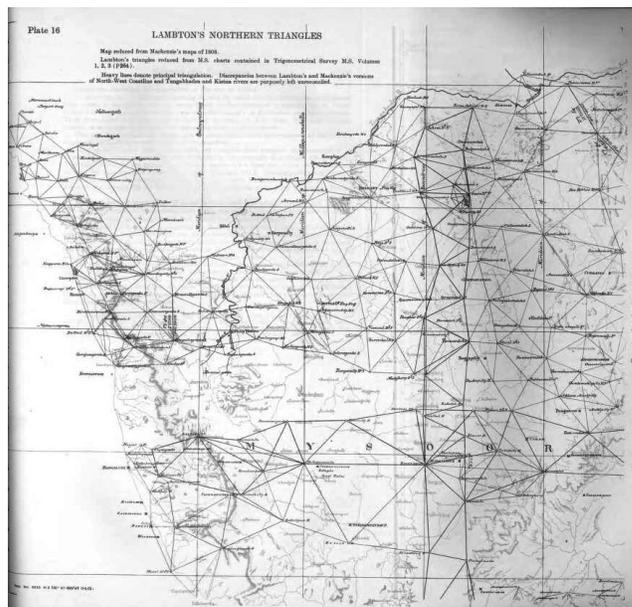
Las mejoras en las herramientas demandó una nueva mejora en las escalas de medición. El inglés JESSE RAMSDEN contribuyó a ello: usó –probablemente como el primero– el torno para cortar tuercas con filetes muy precisos. Estos los usaba para construir unas “máquinas de dividir” de alta precisión: que marcaban en un arco líneas de división a espacios iguales con alta precisión.

Más tarde se empezaron a usar microscopios para leer estas escalas tan finamente gravados.

El proyecto “The Great Survey”

La triangulación más grande que se llevó a cabo en el siglo XIX tenía lugar en la India. Empezó en 1802 con la medición de la línea de base en Madras: una línea recta de 12 km de longitud que se medía varias veces con una cadena calibrada. A partir de esta medida el resto de los cálculos se hicieron con mediciones de ángulos. Sólo de vez en cuando se volvió a medir una línea para disminuir los errores de medición y de cálculo.

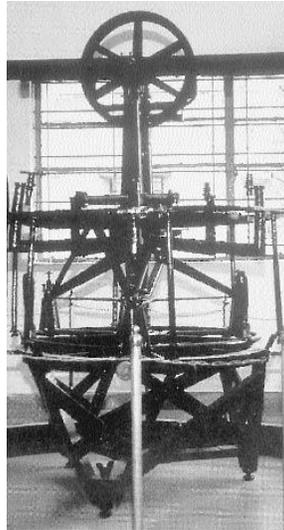
El proyecto estaba planeado para 5 años, pero duró más que 60. Los costos absorbía una gran parte de las ganancias de la *Compañía Británica de las Indias Orientales*, una sociedad de comerciantes ingleses, que por una carta real obtuvieron grandes privilegios en la India. Finalmente el proyecto fue trasladado directamente bajo la custodia de la corona inglesa.



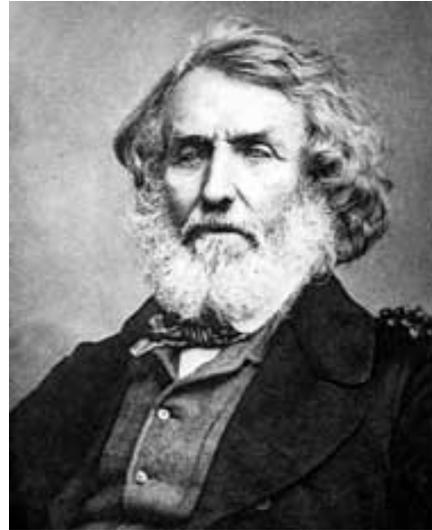
Parte de la triangulación de Lambton

Al principio el proyecto estaba a cargo de WILLIAM LAMBTON (1753–1823). La ilustración anterior muestra una parte del Sur de India, que fue triangulada por LAMBTON. La siguiente ilustración

muestra el teodolito que usó LAMBTON. Pesó media tonelada y fue necesario transportarlo entre doce hombres. Varias veces se averió y tuvo que ser reparado, que retrasó el proyecto ya que en la India no había fabricantes de instrumentos.



El teodolito de Lambton



George Everest

Después de la muerte de LAMBTON, se encargó su asistente GEORGE EVEREST (1790–1866) de dirigir el proyecto. Desistió de querer triangular toda la India y propuso la idea de sólo avanzar en guirnaldas de Sur a Norte y de Este a Oeste. Una ilustración más abajo muestra estas guirnaldas.

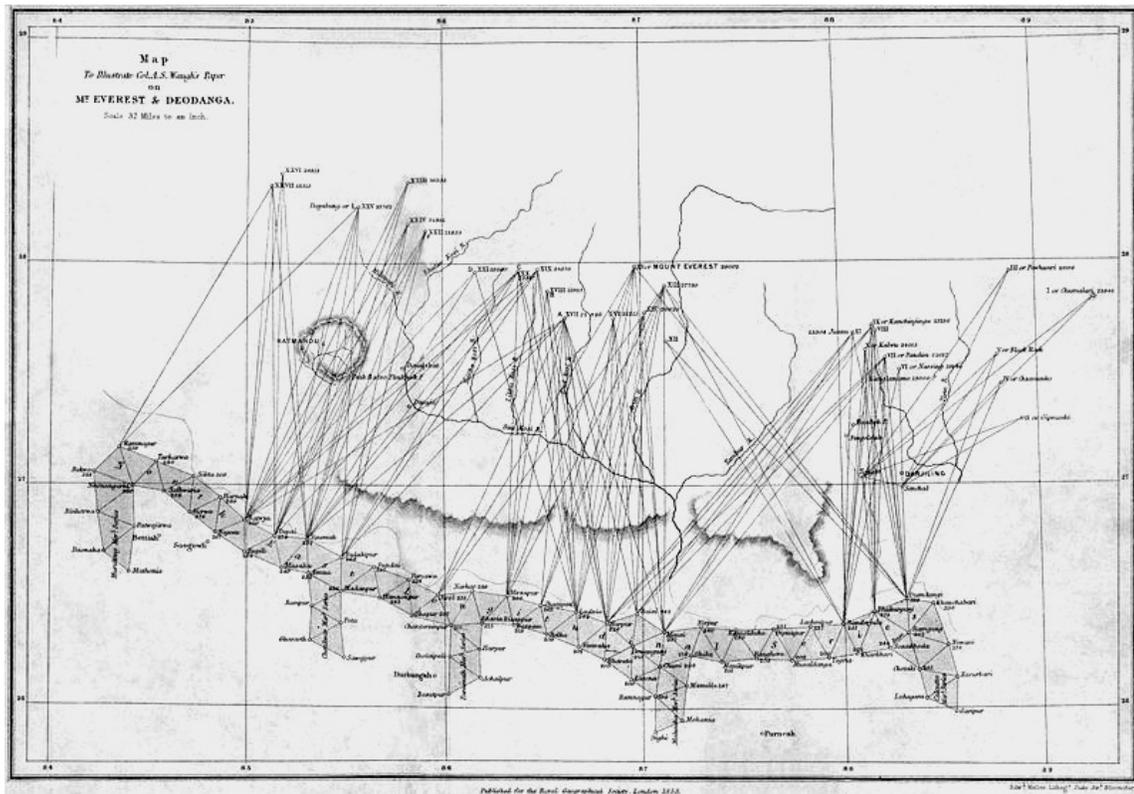
Renovó los instrumentos y consiguió que un fabricante de instrumentos viniera a la India. Dado que las colinas estaban a menudo en las manos de unos caciques locales dejó erigir torres de madera y cuando de día estaba demasiado brumosa, dejó medir de noche usando unas lámparas enormes.

Bajo su dirección trabajaban finalmente 700 hombres. Disponía de cuatro elefantes, 30 caballos y 42 camellos. Después de muchos obstáculos alcanzó las faldas del Himalaya, 2400 kilómetros en el norte. En el año 1843 se retiró.

Dado que el gobierno de Nepal no daba acceso a su territorio, los trabajos entre 1845 y 1850 tuvieron que hacerse atravesando la jungla húmeda en el Sur. Esto costaba la vida de muchos hombres. En un solo año murieron 40 trabajadores indios y de los agrimensores ingleses se enfermó la mitad de fiebre y murió ahí mismo o años más tarde. Se midieron los picos más altos de la cordillera de Himalaya desde lejos. Los triángulos de la triangulación se muestran en abajo.

RADHANATH SIKDAR (1813 – 1870), que EVEREST originalmente había contratado para hacer cálculos, ya había avanzado a ser *Chief Computer of the Survey of India*, determinó que el pico XV con 8840 metros sobre el nivel del mar estaba el más alto de todos. Dado que no se podía saber cómo nombraba el pueblo nepalés este pico, lo bautizó *Mount Everest*, en honor a su jefe retirado.

Las mediciones de la India fueron usados más tarde para calcular la forma de la tierra: los meridianos medidos de Sur a Norte y los paralelos medidos de Este a Oeste permitieron saber que Tierra tiene forma de elipsoide.



Triangulación desde las faldas del Himalaya avisando los picos

5.10. Herramientas de precisión

relojes, la regla de cálculo, integradores analógicos.

Capítulo 6

El álgebra

La gran utilidad y versatilidad de las variables

6.1. El álgebra como disciplina independiente

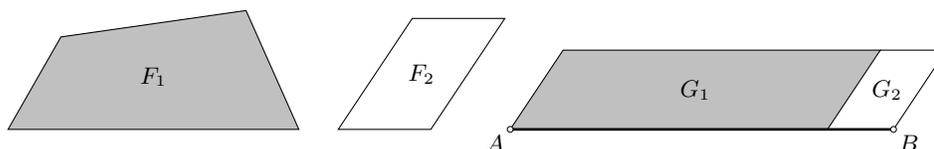
Antecedentes

Hay historiadores que interpretan ciertos enunciados de *Los Elementos* de EUCLIDES como resolución geométrica de una ecuación, que en la notación moderna tendría la forma

$$ax^2 + c = bx \tag{6.1}$$

donde los coeficientes a, b, c son positivos. El enunciado original es el siguiente.

Prop. VI.28: *Aplicar a una recta dada un paralelogramo igual a una figura rectilínea dada deficiente en una figura paralelograma semejante a una dada; pero es necesario que la figura rectilínea dada no sea mayor que el paralelogramo construido a partir de la mitad y semejante al defecto.*



Explicación. Dados un polígono F_1 (la figura rectilínea), un paralelogramo F_2 y un segmento AB (la recta dada) hay que construir sobre una parte de AB un paralelogramo G_1 del mismo área que F_1 de tal manera que, si se extiende G_1 por G_2 (el defecto) a un paralelogramo sobre AB , la parte añadida, G_2 , sea semejante a F_2 .

Análisis de la condición adicional. En la última frase, Euclides formula una condición bajo la cual se debe hacer la construcción. Empecemos a analizarla a ella primero.

Primero colocamos al paralelograma dado F_2 en el extremo B del segmento dado, ver la siguiente ilustración.

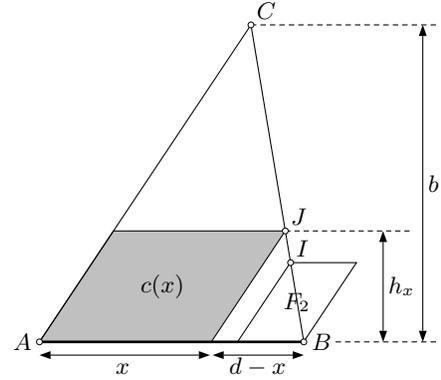
Para que el defecto G_2 sea semejante a F_2 , el punto J opuesto a A del paralelograma, debe encontrarse sobre la diagonal BI de F_2 .

Sea x una parte del segmento dado AB de longitud d . Sobre x erigimos un paralelograma cuya esquina opuesta a A se encuentra sobre BI . El área $c(x)$ de este paralelograma es una función de x . Por la semejanza se tiene

$$h_x : (d - x) = b : d.$$

De ello se calcula $h_x = \frac{b}{d}(d - x)$ y de ahí

$$c(x) = \frac{b}{d}(d - x)x.$$



Así vemos que $c(x)$ es una función cuadrática de x que alcanza su máximo para $x = \frac{d}{2}$. Se tiene

$$c\left(\frac{d}{2}\right) = \frac{b}{d} \cdot \frac{d}{2} \cdot \frac{d}{2} = \frac{bd}{4},$$

que es la mitad del área del triángulo ABC , o el área del paralelograma semejante a F_2 erigido sobre la mitad de d . En conclusión: el área $c(x)$ no puede nunca ser mayor que este valor. Euclides ya sabía que la construcción sólo tiene una solución si esta condición se cumple.

Reformulación moderna. Usamos el lenguaje moderno. Lo que se da es el área c de la figura F_1 , una proporción $a = \frac{h_2}{d_2}$ entre la altura h_2 del paralelograma F_2 y su lado d_2 y además el lado $d = AB$.

Lo que se busca es la base x de G_1 y su altura y . Por las condiciones se deben satisfacer

$$x \cdot y = c \quad \text{y} \quad \frac{y}{d - x} = \frac{h_2}{d_2} = a.$$

De la segunda ecuación se obtiene $y = a \cdot (d - x)$ y por ello de la primera que $a \cdot (d - x) \cdot x = c$, de donde obtenemos

$$ad \cdot x = ax^2 + c$$

que es (6.1) cuando $b = ad$, que es la altura de un paralelograma semejante a F_2 con base d .

De manera similar se puede interpretar la Prop. VI.29 como una manera geométrica de encontrar una solución de la ecuación cuadrática de la forma

$$ax^2 = bx + c, \tag{6.2}$$

donde los coeficientes a , b y c son positivos.

Si EUCLIDES realmente pensó en resolver ecuaciones con la Prop. VI.28 o Prop. VI.29 no queda tan claro. Al menos en los Elementos nunca se habla de ello. La distinción entre (6.1) y (6.2) – que hoy podríamos ver equivalentes – reside en el hecho de que los griegos pensaron todos los números como longitudes, áreas o proporciones de ellos y por lo tanto siempre como números positivos.

ARQUÍMEDES se propuso contestar la pregunta a qué altura h había que cortar una esfera para que los volúmenes de las dos partes, V_1 y V_2 , tengan una proporción dada $\frac{V_1}{V_2} = \frac{p}{q}$. De ello sigue primero que $\frac{V_1}{V_1+V_2} = \frac{p}{p+q}$ y como $V_1 + V_2 = \frac{4\pi}{3}r^3$ se obtiene al resolver por V_1 que

$$V_1 = \frac{p}{p+q} \cdot \frac{4\pi}{3}r^3.$$

ARQUÍMEDES ya conocía el volumen del segmento V_1 :

$$V_1 = \frac{\pi h^2}{3}(3r - h),$$

por lo que obtiene una ecuación cúbica de la forma

$$\begin{aligned} \frac{\pi h^2}{3}(3r - h) &= \frac{p}{p+q} \cdot \frac{4\pi}{3}r^3 \\ h^2(3r - h) &= \frac{4p}{p+q} \cdot r^3. \end{aligned}$$

Esto es una ecuación cúbica en $x = h$ de la forma

$$x^2 \cdot (a - x) = c, \tag{6.3}$$

donde a y c son positivos. La puede reescribir como

$$x^2 = \frac{c}{a - x}$$

y por ello considera la parábola e hipérbola que hoy en la notación moderna se expresan por las ecuaciones

$$y = x^2 \quad \text{resp.} \quad y = \frac{c}{a - x}.$$

En un trabajo sutil ARQUÍMEDES logra dar condiciones en función de a y c para que estas dos curvas se tocan en un punto, se intersectan en dos o no se intersectan. De esta manera resuelve la ecuación (6.3).

Esto nos muestra que ARQUÍMEDES sabía resolver ecuaciones de tercer grado geoméricamente, pero no parece que dedicó gran trabajo al estudio general de las ecuaciones, sino que logró resolver un caso particular, que le interesaba por una motivación en especial.

Otro es el caso de DIOFANTO DE ALEJANDRÍA. Las opiniones acerca del trabajo de DIOFANTO están divididas. Algunos lo ven como un gran genio que planteó problemas de aspectos muy diversos y que siempre dió una solución. Otros a cambio piensan que – como EUCLIDES – DIOFANTO fue más bien un copilador de trabajos hechos por otros.

Véamos dos ejemplos de problemas que consideró DIOFANTO.

Ejercicio II.8: *Se busca escribir un cuadrado dado como suma de dos cuadrados. Sea 16 el cuadrado dado.*

Solución. Si una de las soluciones es $s_1 = x^2$ entonces el otro es $s_2 = 16 - x^2$. Ahora Diofanto propone ver este segundo cuadrado como $(2x - 4)^2$. El -4 está elegido de tal manera que cancele el 16 en s_2 . Se obtiene

$$16 - x^2 = 4x^2 - 16x + 16$$

que se simplifica a

$$\begin{aligned} 5x^2 &= 16x \\ x &= \frac{16}{5}, \end{aligned}$$

Por ello $s_1 = \left(\frac{16}{5}\right)^2$ y $s_2 = 16 - s_1 = \frac{400}{25} - \frac{256}{25} = \frac{144}{25} = \left(\frac{12}{5}\right)^2$.

Ejercicio IV.1: Resuelve la ecuación $x^3 + y^3 = z^2$.

Solución. Nuevamente, Diofanto propone que $y = 2x$ y $z = 6x$ porque entonces

$$\begin{aligned} x^3 + 8x^3 &= 36x^2 \\ 9x^3 &= 36x^2 \\ x &= 4, \end{aligned}$$

de donde $y = 8$ y $z = 24$.

Lo que es interesante observar, es que DIOFANTO no parte de un problema concreto como ARQUÍMEDES. Sino parece tener más bien un enfoque lúdico. Propone problemas que muchas veces tienen una infinidad de soluciones y luego indica cómo se puede encontrar una de ellas. Para hacerlo asume cierta relación entre las variables que no está forzado pero que permite llegar a una solución.

Hay claros indicios de que DIOFANTO usó números negativos, pero no es hasta muchos siglos después que esto se llegara a aceptar por la comunidad científica. Los trabajos de Diofanto no se retoman hasta en el renacimiento, cuando se traducen al latín. Así inspiraron a FERMAT, EULER y LEGENDRE, entre otros, a realizar trabajos propios que llegaron a ser desarrollos importantes en la teoría de números.

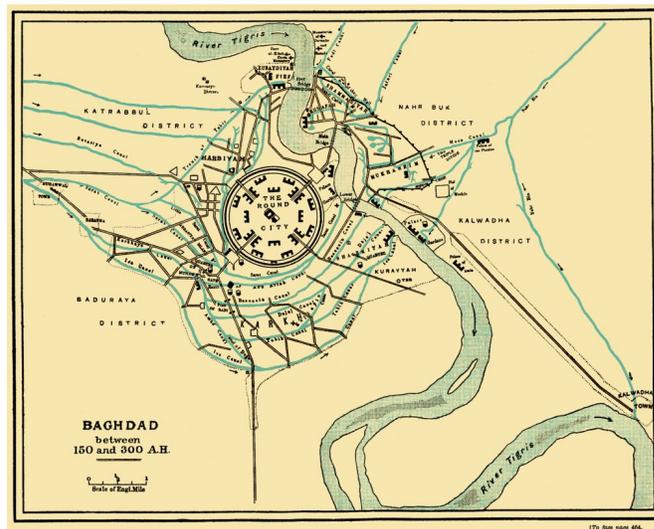
La expansión del islam

Mahoma nació en el año 570 en Meca y murió 632 en Medina. A la edad de 40 años dijo que se le había aparecido el ángel Gabriel y que le había revelado la palabra de Dios. Empezó a distribuir sus creencias y juntó durante su vida un gran número de adeptos. Fue perseguido y siguió la invitación de la ciudad Yathrib, que pronto cambió su nombre a “Medina al Nabi” (ciudad del profeta) que se abrevió a Medina. Entre los ciudadanos de Meca y de Medina se ensancharon las hostilidades y se enfrentaron dos ejércitos. Gracias a la astucia de los generales musulimes de Medina, el ejército mucho mayor en número no pudo conquistar la ciudad Medina. Después de un año se firmó un acuerdo de paz. Mahoma siguió proclamando sus creencias y dos años después el ejército de Medina tomó la ciudad de Meca sin resistencia considerable.

Después de la muerte de Mahoma, su suegro reclamó el liderazgo como califa y abatió varios levantamientos. Se siguieron otros tres califas que expandieron el reinado y las creencias a una extensión enorme. El último fue asasinado en el año 661 y otro clan familiar, llamados Omeya y descendiente de Mahoma, tomó el poder y aseguró las fronteras del reinado. Los Omeya introdujeron una monarquía única para todo el reinado y persiguieron una política de expansión que finalmente fue frenado en 732 por Carlos Martel, el líder del pueblo de los francos. El territorio en este momento era más grande que del imperio romano había sido en su máximo extensión. Pero los omeyas excluyeron a todos los externos de su familia del poder y eran reacios a la ciencia.

Por ello hubo disturbios. Otro clan familiar, los abasí, un linaje del bisabuelo de Mahoma, se levantó y en una serie de batallas acabó con el ejército de los Omeyas. El líder de los abasí, Abu al-Abbas invitó a los Omeyas a un festín y los dejó matar. Sólo un Omeya logró sobrevivir, huyó a España. El reinado de los abasí era largo y duradero. Se diferencian de los Omeyas es que recurrieron a la sabiduría del pueblo del pueblo persa, que era mucho más avanzado.

El sucesor de Abu al Abbas fue al-Mansur. En el grupo de asesores se encontraron no sólo musulimes sino también persas, cristianos y judíos. En 762 fundó Bagdad, el nuevo centro del reinado que ya 50 años después era la ciudad más grande del mundo con más de un millón de habitantes. En el centro de la ciudad se construyó “la ciudad redonda”, una fortaleza circular con dos muros enormes de 10 kilómetros de perímetro, separados por una cequia nutrida por el Tigris. El muro estaba interrumpido por cuatro portales. Adentro estaba el palacio del califa al-Mansur y los edificios del gobierno. El pueblo vivió afuera.



El inicio del calendario musulmán es el año 622, cuando Mahoma emprendió su viaje de Meca a Medina, llamado *hégira*. El año 150 AH (para *Anno Hegirae*) se refiere entonces el año 772, es decir 10 años después de la fundación de Bagdad.

La casa de la sabiduría

Bajo al-Mansur, la vida cultural en Bagdad era pulsante, atrayente, diferentes religiones coexistieron y se toleraban. Esto promovió un desarrollo cultural. El énfasis inicial estaba dedicada a la interpretación del Corán pero bajo los sucesivos califas se tornó hacia la medicina, la astronomía, las matemáticas y la filosofía. Se empezaron a traducir las fuentes griegas al árabe y al persa.

Esta época, de la traducción duró unos 200 años y fue promovido por el mismo califa, pero también por mecenas que pagaban considerables montos por las traducciones. Así, la profesión de traductor fue un trabajo lucrativo. El quinto califa de los abasí, Harun al-Rashid, fundó durante los años de su reinado del año 786 a 809 un centro que se llamó *la casa de la sabiduría*, copiando la idea de la mucho más antigua academia de Gundishapur de los persas. Estos dos centros competían, pero

la casa de la sabiduría atrajo varios científicos de Gundishapur y esta se cerró un siglo y medio después.

Bajo el reinado de al Ma'mun, el hijo de Harun al-Rashid, la Casa de la sabiduría aumentó su esplendor y su reconocimiento atrayendo científicos de muchos lugares. Inicialmente se seguía el trabajo de traducción, pero se empezaba también a general ciencias propios ahí. Se desarrollaron las matemáticas y la astronomía a partir de los conocimientos griegos. La Casa de la sabiduría constaba de un observatorio, una academia, una excelente biblioteca y un hospital.

Por la Casa de la sabiduría Bagdad fue por siglos el centro científico más importante en la región árabe. En el islam, el desarrollo de las ciencias siempre fue unido a las creencias religiosas. La siguiente cita es de Mahoma mismo: *La tinta del erudito es más sagrada que la sangre del mártir*. Los siglos IX y X se pueden considerar como el renacimiento en la cultura islámica. En comparación: los musulmanes tardaron como tres siglos desde su profeta hasta el inicio del renacimiento, el mundo cristiano se tradó como 15 siglos.

El estudio de ecuaciones cuadráticas por al-Juarizmi

ABU ABDALLAH MUHAMMAD IBN MUSA AL-JWARIZMI, conocido también como AL-JUARIZMI (ca. 780–850 e.c) fue uno de los matemáticos que trabajaron en la casa de la sabiduría.

Por petición de al-Ma'mun escribió el libro *Kitab al-mukhtasar fi hisab al-gabr wa l-muqabala* (que quiere decir “La obra sobre el cálculo mediante recuperación y compensación”). Las dos palabras importantes en el título son *gabr* y *muqabala*. El término *gabr* en tratados matemáticos significa “añadir” y se refiere a “añadir términos iguales en ambos lados de una ecuación” o también “multiplicar ambos lados por el mismo número”, mientras *muqabala* significa “quitar” y se refiere a “restar de ambos lados de una ecuación cosas iguales”.

La palabra *al-gabr* se convirtió en “álgebra” y el significado en aquel entonces era “resolver ecuaciones”.

La imagen muestra el título de una copia del año 863.

El libro se escribió con la intención de ayudar en la resolución de problemas diarios. Después de una exposición sobre ecuaciones lineales y cuadráticas, contiene partes sobre cálculos de testamentos y la división de herencias.

AL-JUARIZMI considera manipulaciones de ecuaciones. Para dar un ejemplo, usaremos la notación moderna: convierte

$$x^4 = 40x - 4x^2$$

por la adición de $4x^2$, es decir por *al-gabr*, en

$$5x^2 = 40x,$$

una ecuación que sabe resolver.

AL-JUARIZMI considera ecuaciones lineales y cuadráticas y las separa en casos distintos, que con la



notación moderna se pueden escribir así:

$$\begin{array}{lll} \text{I. } x^2 = px, & \text{II. } x^2 = q, & \text{III. } x = q, \\ \text{IV. } x^2 + px = q, & \text{V. } x^2 + q = px, & \text{VI. } x^2 = px + q. \end{array}$$

AL-JUARIZMI da en cada caso la solución en palabras y luego en una sección a parte da una demostración geométrica para cada caso. Usa ejemplos con números específicos, pero indica que los métodos que usa se pueden usar en cada caso. Además hace un análisis sobre el número de soluciones.

En el caso I. solo considera la solución $x = p$ ya que el cero no es considerado una solución. Este punto de vista perduró hasta el siglo XVII. Como ilustración veremos el caso VI. Considera el ejemplo

Cuadrado equivale a tres raíces y el número cuatro.

Con “raíz” se refiere a la raíz del cuadrado. Si este último es x^2 entonces la raíz es x . En nuestra notación moderna lo escribiríamos así:

$$x^2 = 3x + 4.$$

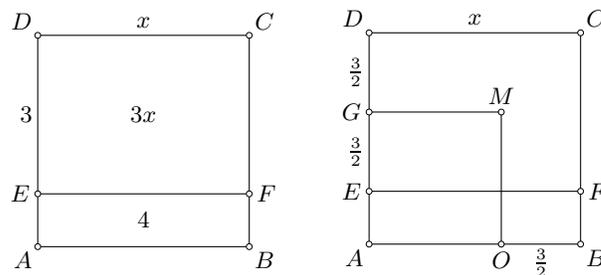
AL-JUARIZMI primero da la solución en palabras. Se toma la mitad del número de raíces, es decir $\frac{3}{2}$ y esto se multiplica por sí mismo, se produce $\frac{9}{4}$. A $\frac{9}{4}$ suma 4, esto da $\frac{25}{4}$, de donde se saca la raíz cuadrada, que es $\frac{5}{2}$. A $\frac{5}{2}$ suma la mitad del número de raíces, o $\frac{3}{2}$, esto da 4. Esto es la raíz del cuadrado y el cuadrado mismo es 16.

No tenemos problemas en traducir esto en una fórmula con nuestro lenguaje moderno:

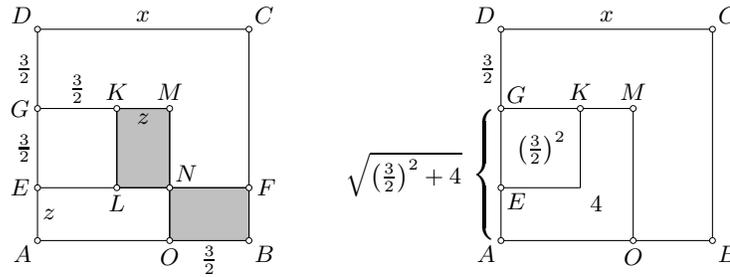
$$x = \sqrt{\left(\frac{p}{2}\right)^2 + q} + \frac{p}{2}. \quad (6.4)$$

Véamos ahora la demostración. Seguiremos la idea de AL-JUARIZMI, pero en un lenguaje más moderno.

Empieza con el cuadrado $ABCD$ cuyo lado x se desconoce. Este equivale a $3x + 4$. Supone que esta división EF ya está dado.



Divide el lado ED por G a la mitad y completa AG a un cuadrado $AOMG$. Por ello GE y GD miden $\frac{3}{2}$. Como $ABCD$ y $AOMG$ son cuadrados GD y OB miden lo mismo, por lo que $OB = \frac{3}{2}$, ver la segunda imagen arriba. Luego completa el segmento EG a un cuadrado $EGLK$, ver la siguiente imagen a la izquierda. Como $AOMG$ y $AOMG$ son cuadrados se tiene que AE mide lo mismo que KM . Consecuentemente, los dos rectángulos resaltados $OBFN$ y $NMKL$ son congruentes.



Por ello el hexágono $AOMKLE$ tiene área 4 y el cuadrado $AOMG$ tiene área $(\frac{3}{2})^2 + 4$, su lado se calcula sacando la raíz, como se indica en la imagen arriba a la derecha. El lado del cuadrado se calcula por lo tanto como

$$x = \sqrt{\left(\frac{3}{2}\right)^2 + 4} + \frac{3}{2}.$$

Con ello termina la demostración ya que la fórmula (6.4) explica su fórmula de manera evidente.

Después de explicar las fórmulas que hoy llamamos binomios, da ejemplos para cada uno de los seis casos. Para el caso VI, este es el siguiente:

Se multiplica $\frac{1}{3}x$ por $\frac{1}{4}x$ y se obtiene x más 24 unidades.

Luego da una explicación: “Primero se debe observar que al multiplicar $\frac{1}{3}x$ por $\frac{1}{4}x$ se obtiene $\frac{1}{2}$ de $\frac{1}{6}x^2$, que es igual a $x + 24$ unidades. La multiplicación de $\frac{1}{2}$ de $\frac{1}{6}x^2$ por 12 da el cuadrado completo [es decir x^2]. Similarmente, la multiplicación de $x + 24$ unidades da $12x + 288$ que es igual a x^2 . Por ello, se toma la mitad de las raíces y esto se multiplica por sí mismo. Luego se suma este producto a 288, se obtiene 324. Toma la raíz de ello, es decir 18 y suma la mitad de las raíces. Se obtiene que la raíz es 24. Entonces este problema te llevó al caso VI, en el cual se trató el tipo donde raíces y números son iguales a un cuadrado.”

Podemos ver varias características en las cuales AL-JUARIZMI difiere del tratamiento que hemos visto de los griegos:

- Procede de manera sistemática y clasifica los posibles casos.
- Para cada caso hace un estudio completo: indica un primer ejemplo, la demostración basada en este ejemplo y un ejemplo adicional.
- Para AL-JUARIZMI, los números 3 y 4 pueden significar longitud, área o simplemente número. No piensa los números como ligados a una dimensión.
- Se argumenta geoméricamente de acuerdo a una operación algebraica.
- Contrario a lo que se puede encontrar en Euclides, AL-JUARIZMI parte en su argumentación de la solución. Esto se ve en la demostración que se indicó: parte de que ya se tiene el cuadrado cuyo lado es desconocido. Después infiere hábilmente cómo se puede calcular el lado.

El método de AL-JUARIZMI se repitió literalmente o de manera semejante por todos los matemáticos conocidos hasta hoy en los países islámicos y aún influyó los científicos de Europa como FIBONACCI o GEROLAMO CARDANO.

El estudio de ecuaciones cúbicas por Omar Jayam

OMAR JAYAM (1048–1131 e.c.) nació en Nishapur y estudió ahí mismo. Nishapur queda hoy en el Noreste de Iran y en aquel entonces era una región turbada por que el tribu de selyúcidas invadían el norte de Iran. A la edad de 32 fue a Samarcanda (hoy en el sureste de Uzbekistán). Ahí, JAYAM escribió su obra sobre ecuaciones cúbicas *Tratado sobre Demostraciones de Problemas de Álgebra*, del cual se hablará en seguida.

Toghril Beg, el fundador de la dinastía selyúcida, declaró la ciudad de Isfahán (en el centro del Iran de hoy) como capital de su reino. Uno de sus nietos lo gobernaba e invitó a JAYAM a construir un observatorio en Isfahán, llamada que este aceptó. Con ello atrajo muchos científicos a Isfahán. Fue un período fructífero de JAYAM.

En su tratado clasificó ecuaciones lineales, cuadráticas y cúbicas en 25 casos que pueden tener soluciones positivas. En la notación moderna, estas son, con dos términos:

$$\begin{array}{lll} \text{I. } x = d, & \text{II. } x^2 = d, & \text{III. } x^3 = d, \\ \text{IV. } x^2 = cx, & \text{V. } x^3 = bx^2, & \text{VI. } x^3 = cx, \end{array}$$

con tres términos

$$\begin{array}{lll} \text{VII. } x^2 + cx = d, & \text{VIII. } x^2 + d = cx, & \text{IX. } x^2 = cx + d, \\ \text{X. } x^3 + bx^2 = cx, & \text{XI. } x^3 + cx = bx^2, & \text{XII. } x^3 = bx^2 + cx, \\ \text{XIII. } x^3 + cx = d, & \text{XIV. } x^3 + d = cx, & \text{XV. } x^3 = cx + d, \\ \text{XVI. } x^3 + bx^2 = d, & \text{XVII. } x^3 + d = bx^2, & \text{XVIII. } x^3 = bx^2 + d, \end{array}$$

y finalmente de cuatro términos

$$\begin{array}{lll} \text{XIX. } x^3 + bx^2 + cx = d, & \text{XX. } x^3 + bx^2 + d = cx, & \text{XXI. } x^3 + cx + d = bx, \\ \text{XXII. } x^3 = bx^2 + cx + d, & \text{XXIII. } x^3 + bx^2 = cx + d, & \text{XXIV. } x^3 + cx = bx^2 + d, \\ \text{XXV. } x^3 + d = bx^2 + cx. & & \end{array}$$

Para cada caso da un argumento geométrico para encontrar la solución. Para las ecuaciones cúbicas, es decir, aquellas que involucran un término x^3 , JAYAM intuía que no era posible construir las soluciones con regla y compás. Su intención era correcta, pero esto no se demostró hasta en el siglo XIX. El dió en estos casos una solución usando intersecciones de cónicas.

Véamos el caso XIX. es decir el primer caso con cuatro términos. JAYAM lo describe así:

Un cubo más cuadrados más lados es igual a un número.

Lo primero que hace JAYAM es que homogeniza los coeficientes. Es decir, define $p = \sqrt{c}$ y $q = \frac{d}{c}$, por lo que $c = p^2$ y $d = cq = p^2q$. Así que trata de resolver la ecuación

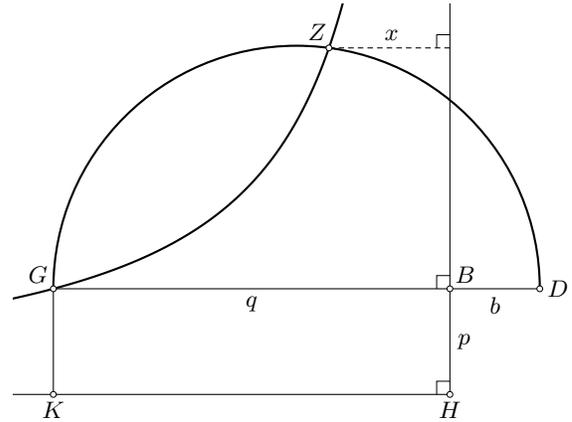
$$x^3 + bx^2 + p^2x = p^2q, \quad (6.5)$$

donde b , p y q son números positivos que puede interpretar como longitudes de segmentos.

Ahora comienza la siguiente construcción. Empezamos con el segmento BG de longitud q y lo prolongamos por BD de longitud b . En un ángulo recto a GD erige en B la perpendicular BH de longitud p .

Completamos GBH a un rectángulo $GBHK$. Erija el semicírculo con diámetro GD y traza la hipérbola por G que tiene a KH y HB como asíntotas.

Las dos cónicas trazadas, el círculo y la hipérbola, se intersectan en G y en otro punto, que llamaremos Z .



En lo que sigue demuestra que la distancia x de Z a la asíntota HB es una solución de la ecuación (6.5). Para ello usa que Z se encuentra sobre las dos cónicas.

Primero erige perpendiculares por Z a las dos asíntotas. Así obtiene el punto A sobre HB y T sobre HK . Además define L como el punto de intersección entre BG y ZT .

Ahora usa el hecho que Z está sobre la hipérbola. De esto sigue que los dos rectángulos $HBGK$ y $HAZT$ tienen el mismo área. Como comparten el rectángulo $HBLT$ los dos restos $TLGK$ y $BLZA$ también son iguales en área entre sí, es decir

$$ZL \cdot LB = GL \cdot LT$$

o con la abreviación $h = ZL$ se tiene $hx = (q - x)p$.

Por ello se tiene

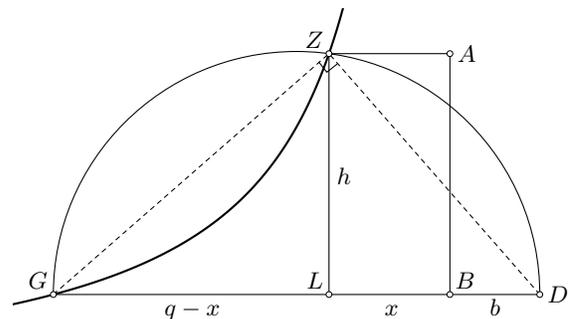
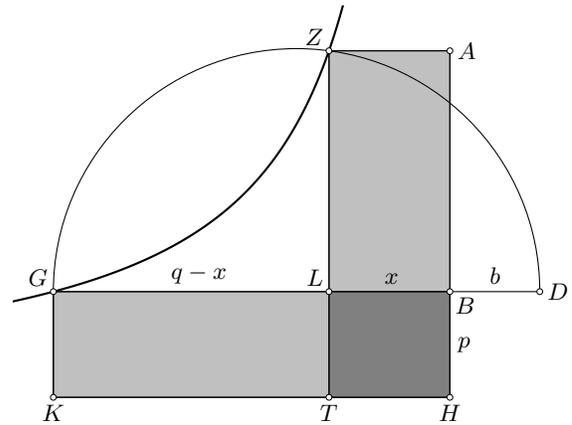
$$\frac{h}{q - x} = \frac{p}{x} \quad \text{y por eso también} \quad \frac{h^2}{(q - x)^2} = \frac{p^2}{x^2}.$$

Ahora usa el hecho que Z se encuentra sobre la circunferencia. De ello sigue que el triángulo GDZ es rectángulo y de ahí que

$$h^2 = (q - x)(b + x).$$

Al sustituir esto, OMAR JAYAM obtiene

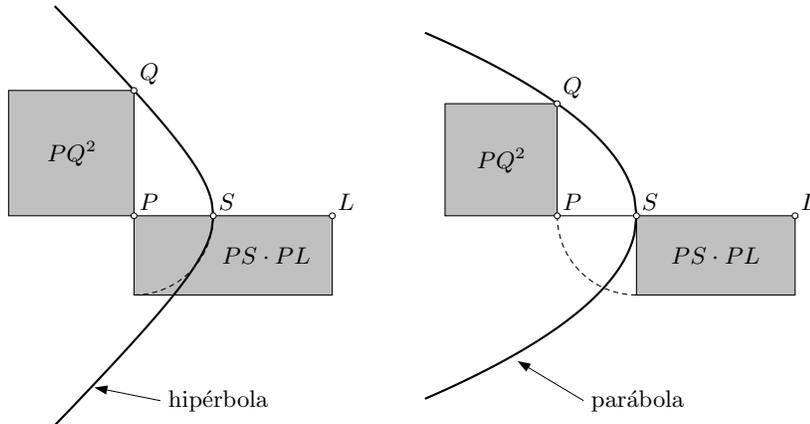
$$\frac{b + x}{q - x} = \frac{p^2}{x^2}$$



Al multiplicar por los denominadores obtiene $(b+x)x^2 = p^2(q-x)$ y después de sumar p^2x finalmente

se consigue (6.5).

Se ve que OMAR JAYAM usa secciones cónicas, en el ejemplo arriba, el círculo y la hipérbola. En muchos libros de texto, se trata de reproducir sus argumentos pero se usan ecuaciones para las curvas en un sistema de coordenadas. Esto no lo hizo OMAR JAYAM, ya que esta idea se la debemos a RENÉ DESCARTES y a PIERRE FERMAT, pero es unos 500 años posterior a OMAR JAYAM. En su trabajo JAYAM usa argumentos del siguiente estilo: por el trabajo de APOLONIO sabe que las dos áreas de rectángulos son iguales:



El trabajo de OMAR JAYAM lamentablemente no se conoció en Europa hasta el siglo XIX y aún en el mundo árabe no era bien conocido. Esto no lo desmerita de ninguna manera. Es un avance importante aunque JAYAM mismo sabía que la meta era conseguir un método algebraico, es decir una manera de obtener la solución a partir de los coeficientes dados.

Bibliografía

B. L. van der Waerden: A History of Algebra, From al-Khwarizmi to Emmy Noether, Springer, 1985.

Dietmar Herrmann: Die antike Mathematik, Springer, 2015.

Alten, et al.: 4000 Jahre Algebra, Springer, 2003.

Jim al-Khalili: Pathfinders. The Golden Age of Arabic Science. Allen Lane, The Penguin Group, London 2010.

6.2. El camino a una notación algebraica

El camino hacia la notación algebraica, que hoy usamos, fue un camino muy largo. Al principio, los matemáticos expresaron los problemas y sus soluciones con palabras o con la geometría. Sólo hasta los siglos XV y XVI se abrió en Europa paso hacia la una notación algebraica moderna.

Contar esta parte de la historia de las matemáticas es nada sencillo. El problema es que casi ningún texto, sobre todo de los griegos, se tiene en su original. En la vasta mayoría sólo se tienen copias

y a veces sólo copias de copias. Los textos originalmente fueron escritos en papiro y sólo más tarde se empezó a usar el papel. Las transcripciones no siempre eran completamente fieles al original, se introdujeron abreviaciones o comentarios. A veces se tradujo del griego al árabe y de este idioma al latín. Todo esto hace difícil una interpretación certera de cómo nació la notación algebraica.

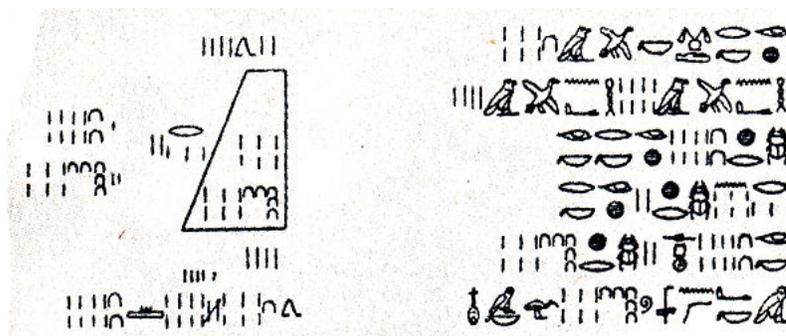
No obstante de esta dificultad, sin duda se puede ver en la historia el siguiente proceso: al principio, los problemas se formulaban con palabras y las soluciones también se daban con palabras. Se habla a veces de “matemáticas retóricas”. Así es la matemática antigua de los babilonios, los egipcios, los chinos y gran parte de los griegos.

Después se empieza a tener un nombre para el número que soluciona un problema dado. Se ve un proceso de abstracción: primero “número” (en DIOFANTO), a “cosa”, que luego se abrevia con “*c*”, “*co.*” o “*ρ*” (por ejemplo en REGIOMONTANO) hasta que después se puede usar cualquier letra como variable, como *x*, *y* o *α*, que es el uso moderno.

Hay abundante literatura sobre este tema. Aquí sólo queremos dar una idea del desarrollo general sin entrar en demasiados detalles y por ello nos restringimos en la discusión a pocos ejemplos, que fueron muy importantes y en donde se pueden ver desarrollos importantes y al mismo tiempo carencias en comparación con el sistema algebraico moderno.

6.2.1. Problemas y soluciones verbales

Para ver cómo se formulaban los egipcios lo que hoy día llamaríamos “fórmula”, veamos el problema 14 del “papiro de Moscú” que data aproximadamente del año 1850 a.C. El papiro está escrito con símbolos hieráticos. La transcripción a jeroglifos se muestra en la siguiente imagen.



El texto en la columna derecha se traduce al español más o menos de la siguiente manera:

“Se pide que calcules el volumen de una pirámide truncada que tiene una altura vertical de 6, una base de 4 de ancho y una tapa de 2 de ancho: debes cuadrar el 4 y obtienes 16; debes multiplicar 4 y 2 y obtienes 8; debes cuadrar 2 y obtienes 4. Debes sumar 16 con 8 y con 4, obtienes 28. Debes tomar $\frac{1}{3}$ de 6, obtienes 2. Multiplica ese 4 con el 28, obtienes 56. Ve, es 56 y eso es el resultado correcto.”

Se ve que es una receta que se podría aplicar también a pirámides truncadas de otras medidas. Nosotros estamos acostumbrados al uso de variables. Si denotamos con *h* la altura, con *a* el ancho de la base y con *b* el ancho de la tapa. Vemos a qué fórmula para el Volumen *V* de la pirámide truncada con estas medidas corresponde la receta formulado en este papiro: es sumar a^2 con ab y

b^2 y luego multiplicar esta suma con $\frac{1}{3}h$, es decir

$$V = \frac{1}{3}h(a^2 + ab + b^2).$$

¡Esto es en efecto la fórmula correcta para el volumen de una pirámide truncada! Pero la explicación que se encuentra en el papiro no contiene variables, sino se usan números distintos para la altura, y los dos anchos. El cálculo se explica usando estos números y en mayor generalidad.

6.2.2. Argumentación geométrica

Del griego EUCLIDES no hay testimonio de un contemporáneo. Para la historia de las matemáticas fue un hombre muy importante. Se tomó la tarea de formalizar casi todo lo que se sabía de geometría en aquel entonces en un sistema axiomático. La obra es conocida como “Los Elementos”. Consta de 13 rollos de pergamino, que hoy se llaman 13 libros. Esto es un fragmento antiguo:



El fragmento data de entre 75 y 125 d.C. y fue encontrado en Egipto en excavaciones que se hicieron en 1896 y 1897. Aunque la letra es difícil de leer, se puede ver que sólo se usaban mayúsculas. Eso fue costumbre en tiempo de los griegos. La transcripción a minúsculas se hizo en el siglo IX en los países árabes. El fragmento contiene el enunciado de la Proposición 5 del segundo libro. Este libro prueba resultados sobre el área de rectángulos.

La Proposición 5 en español dice:

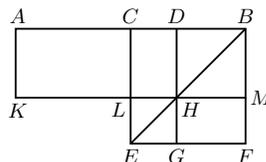
Si un segmento recto es dividido en partes iguales y en partes desiguales, entonces el rectángulo contenido en los segmentos desiguales junto con el cuadrado sobre el segmento entre los dos puntos de división es igual al cuadrado sobre la mitad.

Euclides sigue y aclara lo que se enuncia:

Sea AB un segmento dividido en C en partes iguales y en D en partes desiguales. Digo que el rectángulo contenido en AD , DB junto con el cuadrado sobre CD es igual al cuadrado sobre CB .

Luego sigue la demostración:

Sea $CEFB$ el cuadrado sobre CB [Prop I.46] y sea BE unido. Por D se dibuje DG paralelo a CE o BF por H se dibuje KM paralelo a AB o EF y por A sea AK paralelo a CL o BM [Prop I.31].



Entonces como el complemento CH (se refiere al rectángulo con esta diagonal, es decir $CLHD$) es igual al complemento HF [Prop I.43], sea DM añadido a todos: por ello CM es igual DF (se refiere al área de los rectángulos con esas diagonales, es decir $CLMB$ y $DGFB$).

Como AC es igual a CB , CM es igual a AL [Prop I.36], entonces AL es igual a DF . Sea CH añadido a ambos: por ello AH es igual al gnomon $CLHGFB$.

Pero AH es el rectángulo contenido en AD , DB ya que DH es igual a DB . Por ello el gnomon es igual al rectángulo AD , DB .

Sea LG que es igual a cuadrado sobre CD añadido a ambos, entonces el gnomon y LG son igual al rectángulo AD , DB y el cuadrado CD .

Pero el gnomon y LG son todo el cuadrado $CEFB$ sobre CB y por ello el rectángulo contenido en AD , DB junto con el cuadrado CD es igual al cuadrado sobre CB .
Q.E.D.

Ahora comparamos esta demostración con una argumentación que usa la notación algebraica: Se introducen las variables $x = AC$, $y = CD$ y con ello se tiene $AD = x + y$, $DB = x - y$. El área del rectángulo con los lados AD , DB junto con el área del cuadrado sobre CD es igual a

$$AD \cdot DB + CD^2 = (x + y)(x - y) + y^2 = x^2 - y^2 + y^2 = x^2 = CB^2,$$

que es área del cuadrado sobre CB .

EUCLIDES no podía dar esta demostración porque no tenía la notación algebraica a su disposición.

6.2.3. El nacimiento de la noción de variable

Dos milenios después en Grecia, DIOFANTO DE ALEJANDRIA hizo mayores contribuciones al desarrollo de la aritmética. De hecho hoy existe un área de las matemáticas que se llama “ecuaciones diofantinas” en honor a su labor. Escribió 13 tomos sobre la aritmética, de los cuales los últimos 3 se perdieron. De su vida no se sabe prácticamente nada, ni siquiera sus fechas de nacimiento o muerte. Sólo se sabe que vivió entre 100 a.C y 350 d.C y que trabajó en Alejandria. Su obra la escribió en griego, pero hay muchos historiadores que dudan que fue de origen griego y que suponen que fue cristiano, descendiente de la dinastía que reinaba en aquel entonces en Egipto.

Para ver cómo DIOFANTO formuló sus problemas consideramos el planteamiento VIII del libro II, que se muestra en la ilustración abajo con una traducción al latín y un comentario de FERMAT (que vendrá a ser la conjetura de FERMAT). Es de una impresión de 1670.

QVÆSTIO VIII.

PROPOSITVM quadratum diuidere in duos quadratos. Imperatum fit vt 16. diuidatur in duos quadratos. Ponatur primus 1 Q. Oportet igitur 16 - 1 Q. æquales esse quadrato. Fingo quadratum à numeris quotquot libuerit, cum defectu tot vnitatum quod continet latus ipsius 16. esto à 2 N. - 4. ipse igitur quadratus erit 4 Q. + 16. - 16 N. hæc æquabuntur vnitatibus 16 - 1 Q. Communis adiciatur vtriusque defectus, & à similibus auferantur similia, sient 5 Q. æquales 16 N. & fit 1 N. $\frac{16}{5}$ Erit igitur alter quadratorum $\frac{16}{5}$ alter verò $\frac{4}{5}$ & vtriusque summa est $\frac{20}{5}$ seu 16. & vterque quadratus est.

ΤΟΝ ἑπιτετραγώνου τετραγώνου διελθὲν εἰς δύο τετραγώνους. ἐπιτετραγώνου δὴ τὸ 16 διελθὲν εἰς δύο τετραγώνους. καὶ τετραγώνου ὁ κερστος δυναμικῶς μιας. δώσει ἄρα μονάδας 16 λείψει δυναμικῶς μιας ἴσως ἢ τετραγώνου. πλάσσει τὸ τετραγώνον ἀπὸ 2 N. ἔστω δὴ πρὸς λείψει ποσῶν μὲ ὅσων ἔστιν ἡ τὸ 16 μὲ πλάσσει. ἔστω 5 β. λείψει μὲ δ. αὐτὸς ἄρα ὁ τετραγώνος ἔσται δυναμικῶς δ' μὲ 16 λείψει 5 16. ταῦτα ἴσα μονάδας 16 λείψει δυναμικῶς μιας. κοινὴ κερσκέειδω ἡ λείψει, καὶ ἀπὸ οὐκείων ὁμοία. δυναμικῶς ἄρα ἔσται ἀριθμῶς 16. καὶ γίνεται ὁ κερσκέειδω 16. πέμπτων. ἔσται ὁ μὲν σὺν εἰκοσπέμπτων. ὁ δὲ μετὰ εἰκοσπέμπτων. Ἐοὶ δύο σωτηριεὺς ποιῶν ἢ εἰκοσπέμπτων, ἢ μιν μονάδας 16. καὶ ἔσιν ἑσάτερος τετραγώνου.

OBSERVATIO DOMINI PETRI DE FERMAT.

Cubum autem in duos cubes, aut quadratoquadratum in duos quadratoquadratos & gener. liter nullam in infinitum ultra quadratum potestatem in duos eiusdem nominis fas est diuidere cuius rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.

En este planteamiento considera escribir un número cuadrado como suma de dos cuadrados. El lenguaje que usa requiere de interpretación: DIOFANTO usa la palabra *aritmōs* (αριθμος) para una “longitud”, es decir, el número buscado. Define *aritmōs* por: *lo que contienen un número indefinido de unidades*. Esto ya es muy cercano a lo que hoy llamaríamos una *incógnita*, sólo que se restringe a una longitud.

DIOFANTO usó dos palabras para cuadrado: *tetrágono* (τετραγων), de *tetra* cuatro y *gon*: ángulo; cuadrángulo, es decir caudrilátero, pero aquí quiere decir cuadrado, y *dinamis* (δυναμς). Usó *tetrágono* para indicar el cuadrado de un número concreto y *dinamis* para indicar el cuadrado de un número que se busca, es decir de *aritmōs*. Hoy escribiríamos *aritmōs* como x y *dinamis* como x^2 .

En el texto de DIOFANTO	en greigo	Explicación	notación moderna
aritmōs	αριθμος	número o longitud buscado	x
tetrágono	τετραγων	un cuadrado dado	$16, x^2, \dots$
dinamis	δυναμς	el cuadrado del número buscado	x^2

Una traducción al español del método de DIOFANTO se da en la siguiente tabla, donde del lado derecho se añadieron unas explicaciones con notaciones modernas para mayor legibilidad.

El texto de Diofanto	Explicaciones en notación moderna
<p>A dividir un tetrágono en dos tetrágonos</p> <p>Sea 16 el tetrágono a dividir en dos tetrágonos</p> <p>Pongamos el primer tetrágono como 1 dinamis, entonces el segundo tetrágono es 16 menos 1 dinamis</p> <p>Por ello 16 unidades menos 1 dinamis debe ser un tetrágono.</p> <p>(*) Formo este tetrágono a partir de cualquier número de aritmoi menos tantas unidades que hay en el lado de 16 unidades.</p> <p>Sea esto 2 aritmoi menos 4 unidades.</p> <p>El tetrágonos será 4 dinamis más 16 unidades menos 16 aritmoi.</p> <p>Estos son igual a 16 menos 1 dinamis.</p>	<p>Formulación general: se debe escribir un cuadrado como Suma de dos cuadrados</p> <p>Un ejemplo para ilustrar el procedimiento</p> <p>Si el primer cuadrado es x^2 el segundo es $16 - x^2$</p> <p>El segundo cuadrado es $y^2 = 16 - x^2$</p> <p>(*) Se escribe $y = ax - 4$ ya que $4^2 = 16$.</p> <p>Se toma $a = 2$, es decir $y = 2x - 4$.</p> <p>Con ello se obtiene $y^2 = 4x^2 + 16 - 16x$.</p> <p>Se obtiene $16 - x^2 = 4x^2 + 16 - 16x$.</p>
<p>Suma a ambos aldós los términos faltantes y resta igual de igual.</p> <p>Entonces 5 dinamis igualan 16 aritmoi y el aritmoi será 16 quintos.</p>	<p>Suma en ambos lados los términos que aparecen con un signo negativo en la ecuación: Se debe sumar x^2 y $16x$. Resta 16 de ambos lados.</p> <p>Se obtienen la ecuación $5x^2 = 16x$ y de ahí $x = \frac{16}{5}$.</p>
<p>Uno será 256 veintiquintos y el otro 144 veintiquintos.</p> <p>Su suma es 400 veintiquintos, es decir 16, y cada uno es un tetrágono.</p>	<p>Se obtiene $x^2 = \frac{256}{25}$ y $y^2 = \frac{144}{25}$.</p> <p>Se tienen $x^2 + y^2 = \frac{256}{25} + \frac{144}{25} = \frac{400}{25} = 16$.</p>

El punto crucial es (*): en donde expresa una variable por medio de la otra: lo hace de tal manera que el cuadrado tanga la misma cantidad de unidades. Por ello resta 4 de ax . Así obtiene

$$y^2 = (ax - 4)^2 = a^2x^2 + 16 - 8ax.$$

Y al igualarlo con $y^2 = 16 - x^2$ obtiene

$$a^2x^2 + 16 - 8ax = 16 - x^2.$$

De ahí obtiene

$$(a^2 + 1)x^2 = 8ax$$

y finalmente

$$x = \frac{8a}{a^2 + 1}.$$

El método es general, pero lo exhibe a través del ejemplo $a = 2$. Así se obtienen un número infinito de soluciones. Estas soluciones son fracciones. Vale mencionar que DIOFANTO es el primero de que se tienen testimonio de que consideró fracciones como números.

Según Thomas Heath, DIOFANTO abrevia *aritmói* en algunos textos con *ar* en griego, es decir $\alpha\rho$ y esta combinación de letras lo escribe ligado, de tal manera que aparece como una nueva letra. La imagen que dimos arriba no aparece la abreviación, sino la palabra completa, pero esto se debe a que no es el texto original, sino una impresión hecha en el siglo XVII en Francia.

Diofanto usó en ciertos textos también abreviaciones para potencias de la incógnita:

Nombre	Transcripción	Símbolo de Diofanto	Símbolo moderno
$\delta\nu\nu\alpha\mu\iota\varsigma$	<i>dinamis</i>	Δ^Υ	x^2
$\kappa\nu\beta\omicron\varsigma$	<i>cubos</i>	K^Υ	x^3
$\delta\nu\nu\alpha\mu\omicron\delta\nu\nu\alpha\mu\iota\varsigma$	<i>dinamodinamis</i>	$\Delta^\Upsilon\Delta$	x^4
$\delta\nu\nu\alpha\mu\omicron\kappa\nu\beta\omicron\varsigma$	<i>dinamocubos</i>	ΔK^Υ	x^5
$\kappa\nu\beta\omicron\kappa\nu\beta\omicron\varsigma$	<i>cubocubos</i>	$K^\Upsilon K$	x^6

También usó símbolos para denotar los recíprocos de estas potencias, es decir para $\frac{1}{x}$, $\frac{1}{x^2}$ etcétera. No usó potencias mayores que la sexta. Queda claro que en DIOFANTO se ve un avance tremendo con respecto a la formulación de los egipcios. DIOFANTO es de los griegos sin duda, el que ha llegado más lejos en aplanar el camino hacia una simbología algebraica.

Pero también deja entrever dificultades de esta notación. La ley

$$x^m \cdot x^n = x^{m+n}$$

es inconcebible con esta notación. Para cada potencia hay un símbolo particular y no se ve la generalidad de “potencia” en sí. DIOFANTO no usó símbolo para denotar la adición.

En este mismo estilo se escribieron las matemáticas en los países árabes en los cuales se preservaron los conocimientos y más tarde, sobre todo en el siglo IX y X d.C. se recolectaron textos de las civilizaciones anteriores y se tradujeron del griego y de otros idiomas al árabe. A parte se hicieron grandes desarrollos sobre todo en la resolución de ecuaciones algebraicas, sobre todo de ecuaciones cuadráticas y cúbicas.

Con el renacimiento en Europa, al principio sobre todo en Italia, se tradujeron estos textos del árabe al latín y se descubrieron las raíces griegas. Se hizo una escuela “abaquista”, que tenía por fin el eficiente y con ello rápido cálculo con (o sin) el abaco. Muy famoso e influyente fue la escuela alemana que promovió la aritmética. Pero ninguna de estas corrientes lograron un cambio serio en cómo se denotaba el álgebra. La incógnita se denotaba de una u otra manera y sus potencias de otra.

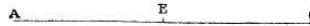
Poco a poco se estableció el uso de ciertos símbolo para denotar las operaciones. Por ejemplo, en 1489 apareció por primera vez el símbolo $+$ para denotar la adición en un texto de JOHANNES WIDMANN (ca. 1460–1498).

6.2.4. Se establece una notación común

Poco a poco se estableció una simbología entre los matemáticos. Aquí veremos dos fragmentos del siglo XVII. El primero es la impresión de una carta que escribió PIERRE FERMAT (1601/1607–1665) a su colega MARIN MERSENNE (1588–1648) en 1629.

Exemplum subijcimus : *Sit recta AC (fig. 91) ita dividenda in E ut rectangulum AEC sit maximum.*

Fig. 91.



Recta AC dicatur B . Ponatur pars altera ipsius B esse A : ergo reliqua erit $B - A$, et rectangulum sub segmentis erit B in $A - Aq.$, quod debet inveniri maximum. Ponatur rursus pars altera ipsius B esse $A + E$: ergo reliqua erit $B - A - E$, et rectangulum sub segmentis erit

$$B \text{ in } A - Aq. + B \text{ in } E - A \text{ in } E \text{ bis} - Eq.,$$

quod debet adæquari superiori rectangulo

$$B \text{ in } A - Aq.$$

Demptis communibus,

$$B \text{ in } E \text{ adæquabitur } A \text{ in } E \text{ bis} + Eq.,$$

et, omnibus per E divisis,

$$B \text{ adæquabitur } A \text{ bis} + E.$$

Elidatur E ,

$$B \text{ æquabitur } A \text{ bis}.$$

Igitur B bifariam est dividenda ad solutionem propositi; nec potest generalior dari methodus.

En esta carta FERMAT explica el principio de encontrar un máximo o un mínimo de una función polinomial en un ejemplo. Se trata de dividir el segmento AC en un punto E de tal manera que el rectángulo con los lados AE y EC tenga un área máximo. FERMAT usa “in” para denotar la multiplicación, los símbolos $+$ y $-$ como nosotros, indica cuadrados con un “q” y la multiplicación con 2 con un “bis”.

FERMAT pone $b = AC$ y $a = AE$ con lo que $EC = b - a$, sólo que FERMAT usa para todo letras mayúsculas. Ahora sigue FERMAT así: Si cambiamos a por $a + \varepsilon$ dividimos b en $a + \varepsilon$ y $b - a - \varepsilon$. El área sería $(a + \varepsilon)(b - a - \varepsilon) = ba - a^2 + b\varepsilon - 2a\varepsilon - \varepsilon^2$, esto aparece como

$$B \text{ in } A - Aq. + B \text{ in } E - A \text{ in } E \text{ bis} - Eq.$$

Este área debe ser igual al área del primer rectángulo (con la división en a y $b - a$), es decir a

$ba - a^2$. Esto es la nueva idea de FERMAT: dado que E está puesto de tal manera que $AE \times EC$ sea máximo, el área casi no cambia al variar E un poco. Así que

$$ba - a^2 \approx ba - a^2 + b\varepsilon - 2a\varepsilon - \varepsilon^2$$

de donde obtiene

$$2a\varepsilon + \varepsilon^2 \approx b\varepsilon.$$

Esta ecuación divide entre ε y luego “elimina” ε , es decir, en el lenguaje moderno: hace el límite $\varepsilon \rightarrow 0$. Con ello obtiene $b = 2a$, es decir, el cuadrado de todos los rectángulos con perímetro dado, aquel con mayor área.

FERMAT ya usa letras para denotar diferentes variables. Algunos símbolos ya se establecieron, así por ejemplo para denotar la adición y sustracción, pero no para potencias o para la igualdad.

Veamos ahora un pedazo del apéndice “Geometrie” del famoso “Discours de la méthode” de RENÉ DESCARTES (1596–1650), que fue publicado en 1637.

LIVRE PREMIER 301

ainsi en les demellant, qu'il n'en demeure qu'une seule, esgale a quelque autre, qui soit connue, ou bien dont le quarré, ou le cube, ou le quarré de quarré, ou le surfolide, ou le quarré de cube, &c. soit esgal a ce, qui se produit par l'addition, ou soustraction de deux ou plusieurs autres quantités, dont l'une soit connue, & les autres soient composées de quelques moyennes proportionnelles entre l'unité, & ce quarré, ou cube, ou quarré de quarré, &c. multipliées par d'autres connues. Ce que j'explique en cete sorte.

$x \propto b$. ou
 $x^2 \propto -ax + bb$. ou
 $x^3 \propto +ax^2 + bbx - c$. ou
 $x^4 \propto ax^3 - cx^2 + d$. &c.

C'est a dire, x , que ie prens pour la quantité inconnue, est esgalé a b , ou le quarré de x est esgal au quarré de b moins a multiplié par x . ou le cube de x est esgal à a multiplié par le quarré de x plus le quarré de b multiplié par x moins le cube de c . & ainsi des autres.

En este apéndice explica cómo se pueden usar coordenadas y que con ellos las curvas se expresan mediante ecuaciones. Es el nacimiento de la geometría analítica. Veamos cómo DESCARTES usa ya una notación algebraica muy similar a la nuestra. Si escribimos el símbolo “ \propto ” como una

igualdad “=”, entonces las cuatro líneas en medio se ven así:

$$\begin{aligned}z &= b. \\z^2 &= -az + bb. \\z^3 &= +az^2 + bbz - c^3. \\z^4 &= az^3 - c^3z + d^4\end{aligned}$$

y luego describe en francés lo que significan estas ecuaciones. En español:

“es decir, z , que uso como la cantidad desconocida, es igual a b , o el cuadrado de z es igual al cuadrado de b menos a multiplicado por z . O el cubo de z es igual a a multiplicado por el cuadrado de z más el cuadrado de b multiplicado por z menos el cubo de c , y así los demás.

También usa los símbolos $+$ y $-$ y escribe la multiplicación al escribir dos variables juntos. Lo nuevo es que escribe las potencias con variables como bases y números como exponentes. Sólo falta el símbolo de igualdad.

6.2.5. Conclusión

Con DESCARTES la transformación de la notación casi llega a su fin. Sólo poco se añadió después de él. Un siglo después, la notación es prácticamente idéntico a la de hoy día y uno puede leer los textos en original, claro si uno entiende el idioma en el cual están escritas.

Con esta notación es posible hacer operaciones a nivel notacional, sin tener que razonar usando la geometría. Por ejemplo: la notación de potencias con el exponente como un número “elevado” permite expresar leyes de potencias. De esta manera, la notación algebraica crea condiciones para un desarrollo impresionante en las matemáticas.

Bibliografía

Jean Christianidis: The way of Diophantus: Some clarifications on Diophantus' method of solution. *Historia math.* doi:10.1016/j.hm.2006.10.003

Albrecht Heffer: On the nature and origin of algebraic symbolism. *New Perspectives on Mathematical Practices*. 2008.

Sir Thomas Heath: A history of greek mathematics, Vol II, Dover 2014. Wikipedia en español, inglés y alemán.

6.3. Teoría de Ecuaciones

6.3.1. Primeros intentos algebraicos

En la Sección 6.1 vimos que OMAR JAYAM (1048–1131 e.c.) supo resolver ecuaciones cúbicas con el uso de secciones cónicas, es decir hipérbolas, parábolas y círculos. Aquí veremos cómo el estudio

de las ecuaciones *algebraicas* progresó. Ejemplos de ecuaciones algebraicas son

$$\begin{aligned} 3x &= 4 \\ 3x^2 &= x + 1 \\ x^3 + x^2 + 2x + 7 &= 0 \\ 2x^4 + x^3 + 1 &= 0, \end{aligned}$$

o más general ecuaciones de la forma

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

en donde x es la incógnita y $a_0, a_1, \dots, a_{n-1}, a_n$ los coeficientes, donde sólo se pide que $a_n \neq 0$. El número n es el *grado* de la ecuación. Para $n = 1, 2, 3$ la ecuación se llama lineal, cuadrática o cúbica, respectivamente.

El trabajo de JAYAM no fue conocido en Europa hasta el siglo XIX, por lo que retomamos el desarrollo con el estudio de las ecuaciones cuadráticas de AL-JUARIZMI (ca. 780–850 e.c). Su obra fue traducido en el siglo XII en España al castellano y también latín y así se hizo accesible en Europa desde el siglo XIII.

El cambio de JAYAM, quien dió las soluciones como construcciones geométricas, AL-JUARIZMI dió soluciones *algebraicas*, es decir indicó como hallar las soluciones a partir de los coeficientes. Por ejemplo, para ecuaciones de la forma

$$x^2 = px + q$$

indicó que las soluciones se obtienen al tomar la mitad de p y elevar ésta al cuadrado, luego sumar a este cuadrado q y sacar la raíz de esta suma y finalmente sumar la mitad de p . Con la notación moderna esto se escribe como

$$x = \sqrt{\left(\frac{p}{2}\right)^2 + q} + \frac{p}{2}.$$

Es un punto de vista completamente diferente: si se suma el cuadrado de $\frac{p}{2}$ con q , no puede uno pensar en objetos geométricos: no tiene sentido sumar un área con una longitud, ya que tienen dimensiones diferentes. Esto es el punto *algebraico* que aquí queremos seguir en su desarrollo.

El descubrimiento de los textos árabes y griegos en la región que los españoles reconquistaron de los moros, en particular en la ciudad Toledo, inició una ola de traducciones que propagó el redescubrimiento de los conocimientos en Europa. Al principio sólo se tradujeron, a veces primero del árabe al español y después de este idioma al latín, más tarde en el siglo XII y XIII hubo los primeros trabajos originales en Europa que superaron los avances traducidos.

Uno de los matemáticos más notables en esta época fue LEONARDO DI PISA (1170–1240), quien también se conoce como FIBONACCI, que es una abreviación de *filius* (hijo) y *Bonaccio* (el nombre de su padre).



Un retrato que se le atribuye a FIBONACCI (1170–1250)

De MacTutor

En 1202 FIBONACCI escribió un libro con el nombre *Liber Abaci* (libro sobre el Abaco), que es más bien un libro sobre la aritmética en donde también presenta el sistema decimal con las cifras árabes (aunque él las atribuye a los Hindues). En este libro presenta muchos problemas con sus soluciones, algunos de ellos tomados desde AL-JUARIZMI, otros propios suyos. Es el primer libro que rebasa claramente el alcance de que se conoció de los árabes, tanto en en la variedad de los problemas presentados, la creatividad de los métodos de resolución como la claridad de la exposición.

En otro tratado llamada *flos* (flor) se encuentra el tratamiento de FIBONACCI la ecuación

$$x^3 + 2x^2 + 10x = 20.$$

Después de dividir entre 10 obtiene

$$x + \frac{x^2}{5} + \frac{x^3}{10} = 2$$

de lo que concluye que la solución tiene que ser menor que 2. Recordamos que en aquel entonces sólo se consideraron soluciones positivas. Por otro lado, $x = 1$ no es una solución, cosa que se comprueba fácilmente al sustituirlo. Por ello la solución no puede ser un número natural. Después de dar más información sobre una posible solución (que no puede ser racional, ni raíz cuadrada de un número racional) da un valor aproximado en el sistema sexagesimal

$$1 + \frac{22}{60} + \frac{7}{60^2} + \frac{42}{60^3} + \frac{33}{60^4} + \frac{4}{60^5} + \frac{40}{60^6},$$

sin explicación de cómo llegó a ello, pero que tiene una exactitud de diez cifras decimales. El resultado muestra también lo difícil que fue la introducción del sistema decimal: aún uno de sus propagadores usaba el sistema sexagesimal en un ejemplo concreto.

Entre los manuscritos de los siglos XIV y XV hay varios tratados sobre álgebra, pero poco avance sustancial. En uno de ellos, de autor desconocido, se da para la ecuación

$$x^3 + px^2 + qx = r$$

una manera de encontrar la solución, que corresponde en la notación de hoy a la siguiente fórmula:

$$x = \sqrt[3]{\left(\frac{q}{p}\right) + r} - \frac{q}{p}.$$

La fórmula es falsa en general, pero correcta en el caso que $p^2 = 3q$, y en este caso fue empleada.

LUCA PACIOLI, con el nombre completo FRA LUCA BARTOLOMEO DE PACIOLI (1447–1517) comentaba en su obra *Summa de arithmetica geometria proportioni et proportionalità* que para las ecuaciones cúbicas nadie hasta entonces ha podido dar una regla general para encontrar las soluciones. Su obra fue importante y muy conocida menos por el contenido, que no rebasaba textos anteriores, sino porque era de los primeros libros sobre aritmética que se imprimió.

6.3.2. El turbio descubrimiento

Queremos dar aquí el relato de las circunstancias que condujeron a la publicación final en 1545 de la fórmula general para resolver una ecuación cúbica de la forma

$$x^3 + px = q. \tag{6.6}$$

Esta ecuación se llama hoy ecuación cúbica *reducida*. El relato es algo largo, pero se puede obtener de los detalles una buena impresión sobre la manera cómo se llevó a cabo el trabajo de un matemático en aquel entonces.

SCIPIONE DEL FERRO (1465–1526) era el primero en encontrar una regla para resolver la ecuación cúbica reducida (6.6), pero nunca la publicó. Probablemente también supo cómo reducir otras ecuaciones cúbicas a pocos casos. Dado que no se trabajaba con números negativos había que considerar casos distintos.

SCIPIONE DEL FERRO era docente en la Universidad de Bologna desde 1496 hasta su muerte. Antes de morir entregó a su yerno HANNIBAL NAVE un pequeño cuaderno de notas que contenía la solución. NAVE era también matemático y tomó el puesto de DEL FERRO. Al menos un discípulo más de DEL FERRO conoció la solución también: ANTONIO MARIA FIOR.

La segunda persona que descubrió la solución fue NICCOLÒ FONTANA también conocido como NICCOLÒ TARTAGLIA (ca. 1500–1557). Nació en condiciones pobres. Su padre murió cuando tuvo 6 años.



NICCOLÒ TARTAGLIA (ca. 1500–1557)

De la Wikipedia

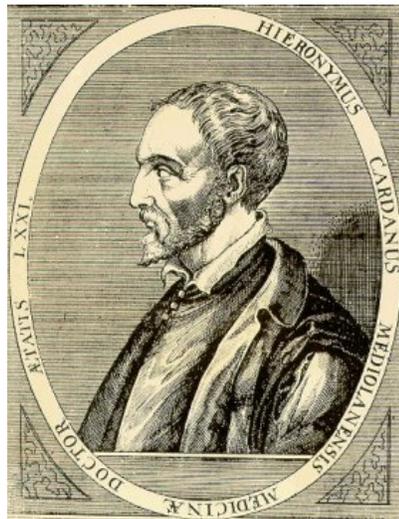
A los 12 años tropas franceses tomaron la ciudad de Brescia en el norte de Italia y mataron a 46 000 personas. TARTAGLIA quien se había refugiado en una iglesia, sufrió varios golpes de sable en la cara y fue dado por muerto por los soldados. Así y con el cuidado de su madre sobrevivió, pero su cara era desfigurada y padecía de tartamudeo, lo que le dio su apodo *Tartaglia*, el tartamudo.

Era un buen autodidacta. Aprendió a leer y todas sus habilidades matemáticas por su propia cuenta. Se ganó su vida como calculista para negociantes y maestro privado. TARTAGLIA encontró cómo resolver ecuaciones de la forma

$$x^3 + mx^2 = n.$$

En 1533 se dió una competencia entre FIOR y TARTAGLIA. Cada uno había que entregar 30 problemas con sus soluciones a un notario. Será ganador quien, después de 30 días habría resuelto más problemas. TARTAGLIA entregó a FIOR problemas de la forma $x^3 + mx^2 = n$, que este no pudo resolver. FIOR entregó a TARTAGLIA de la forma $x^3 + px = q$. Según el relato del mismo TARTAGLIA, encontró la solución solo horas antes de que finalizara la competencia, y que resolvió los 30 problemas en sólo 2 horas. La competencia terminó 30 : 0 a favor de TARTAGLIA.

Las competencias eran una forma de hacerse fama y por ello era importante mantener sus conocimientos en secreto. Lo de la fama le funcionó a TARTAGLIA: le despertó el interés en GIROLAMO CARDANO (1501–1576), quien era médico, filósofo y matemático, había viajado y quien era reconocido por curar al arzobispo John Hamilton, escuchó de ello en Milán.



GIROLAMO CARDANO (ca. 1501–1576)

De la Wikipedia

CARDANO trató de resolver el problema él mismo, pero fracasó. A través de un intermedario, dejó preguntar a TARTAGLIA, si quería publicar su solución en un libro que estaba escribiendo. Pero este lo negó. Entonces CARDANO escribió directamente a TARTAGLIA y dejó entrever, que ya había debatido la brillantez de TARTAGLIA con el gobernador Alfonso d'Avalos y que lo invitaba.

Así, Tartaglia viajó a Milán, porque esperaba obtener un empleo bien remunerada por el gobernador. Pero, éste estuvo de viaje. CARDANO agasajó a TARTAGLIA tan bien, que este reveló su secreto bajo la promesa de nunca publicarlo y anotarlo sólo en clave. TARTAGLIA regresó en la misma noche, profundamente descontento. Sentía que había hecho un error en pasar su secreto a CARDANO. Este cumplió su promesa. El libro fue publicado sin la fórmula de TARTAGLIA.

Pero con este nuevo conocimiento CARDANO y su discípulo LODOVICO FERRARI (1522–1565) pudieron resolver todos los casos posibles de ecuaciones cúbicas, y un poco más tarde aún la ecuación de cuarto grado. En el año 1543 viajaron los dos a Bologna y se enteraron de NAVE, que DEL FERRO supo resolver este tipo de ecuaciones unos 30 años antes de Tartaglia. Por ello CARDANO ya no se sintió obligado a respetar su promesa y publicó 1545 la solución completa de las ecuaciones de tercer y cuarto grado en el libro *Artis magnaë sive de regulis algebraicis liber unus* (abreviado como *Ars magna*). En este libro se reconoce las contribuciones de DEL FERRO, TARTAGLIA y FERRARI explícitamente

Cuando TARTAGLIA se enteró estaba furioso y desarrolló odio hacia CARDANO. Publicó un libro en donde realataba su punto de vista y reto a CARDANO varias veces a una competencia. Este lo negó y TARTAGLIA y FERRARI disputaron con creciente agresividad y mutuos insultos durante un año sin dar resultado alguno. En 1548 TARTAGLIA obtuvo un muy buen ofrecimiento de un puesto en su ciudad natal, Brescia. En el mismo año, se dió una competencia entre FERRARI y TARTAGLIA en Milán. TARTAGLIA tenía experiencias en semejantes competencias, mientras FERRARI era joven e inexperto. Sin embargo, después de un día, TARTAGLIA se dió cuenta, de que la habilidad de FERRARI superaba la suya y abandonó la ciudad en la noche. Así FERRARI fue electo ganador. Como consecuencia, a TARTAGLIA ni siquiera fue pagada el curso que había impartida en la universidad.

Murió enardecido, pobre y sin fama: las fórmulas se conocen hoy como *Fórmulas de Cardano*.

6.3.3. Descubriendo la solución

Queremos ahora ver cómo es posible – con sólo dos ideas – llegar a la solución de la ecuación cúbica reducida (6.6).

La primera idea es que se escribe la solución como una diferencia:

$$x = a - b.$$

Al sustituir en (6.6) y expandir se obtiene

$$\begin{aligned}(a - b)^3 + p(a - b) &= q \\ a^3 - 3a^2b + 3ab^2 - b^3 + p(a - b) &= q \\ a^3 - b^3 + (p - 3ab)(a - b) &= q\end{aligned}\tag{6.7}$$

Con la primera idea aparentemente todo se complica, pero se introduce un cierto grado de libertad: hay muchos pares a y b tales que $a - b$ sea igual a un número dado, en nuestro caso, la solución buscada.

La segunda idea es hacer de la ecuación (6.7) dos ecuaciones:

$$\begin{aligned}a^3 - b^3 &= q & (6.8) \\ (p - 3ab)(a - b) &= 0 & (6.9)\end{aligned}$$

Hay que detenerse un poco ya que la idea está un poco rara: si supieramos que (6.8) y (6.9) estuvieran correctas, entonces pudieramos concluir que (6.7) también se cumpliera. Pero la implicación inversa es falsa. De (6.7) no tiene que seguir que necesariamente se cumplan también (6.8) y (6.9). Resulta que la segunda idea junto con la primera es fructífera: la primera idea introduce un grado de libertad tal que se puede resolver (6.8) y (6.9) al mismo tiempo.

De (6.9) sigue que $p = 3ab$ o que $a = b$. Lo último no nos interesa porque en este caso $x = 0$ y con ello $q = 0$, por lo que la ecuación original sería $x^3 + px = 0$. Esta ecuación claramente tiene las tres soluciones $x = 0, x = \sqrt{p}, x = -\sqrt{p}$. Por ello concluimos que $p = 3ab$ o $p^3 = 27a^3b^3$. Si sustituimos $A = a^3, B = b^3$ obtenemos entonces el siguiente sistema de ecuaciones

$$A = B + q \quad \text{y} \quad AB = \left(\frac{p}{3}\right)^3.$$

De lo que resulta la ecuación cuadrática

$$\begin{aligned}(B + q)B &= \left(\frac{p}{3}\right)^3 \\ B^2 + qB - \left(\frac{p}{3}\right)^3 &= 0.\end{aligned}$$

Esta ecuación era fácil de resolver. Una solución es:

$$B = \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3} - \frac{q}{2}.$$

De ahí sigue

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

Al sacar las raíces cúbicas se obtienen $a = \sqrt[3]{A}$ y $b = \sqrt[3]{B}$ y finalmente $x = a - b$:

$$x = \sqrt[3]{\sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3} + \frac{q}{2}} - \sqrt[3]{\sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3} - \frac{q}{2}}. \quad (6.10)$$

Esta es la fórmula de solución para la ecuación (6.6).

6.3.4. Expansión de la noción de número

Un ejemplo del libro *Ars magna* de GIROLAMO CARDANO es

$$x^3 + 6x = 20.$$

Notamos primero que ya se completó la algebraización, en el sentido que la incógnita no es pensado como un ente geométrico: x no puede ser pensado como la longitud de un segmento, porque sino del lado izquierdo se sumaran un volumen con una longitud, una expresión que carece de sentido geométrico. La incógnita x es pensado como un *número abstracto*.

El descubrimiento de la solución de la ecuación cúbica reducida (6.6) cae en el tiempo que se empezaron a aceptar los números negativos. En el *Ars magna* de CARDANO no se aceptan estos números abstractos y por ello se tienen que considerar muchos casos. Por ejemplo, para la ecuación

$$x^3 = px + q \quad (6.11)$$

se daba otra receta para encontrar la solución, que equivale a la fórmula

$$x = \sqrt[3]{\sqrt{\left(\frac{q}{2}\right)^2 - \left(\frac{p}{3}\right)^3} + \frac{q}{2}} - \sqrt[3]{\sqrt{\left(\frac{q}{2}\right)^2 - \left(\frac{p}{3}\right)^3} - \frac{q}{2}}.$$

Para nosotros es más fácil reescribir la ecuación 6.11 como

$$x^3 - px = q$$

y luego sólo se tiene que cambiar el signo de p en la solución (6.10). En el libro *Arithmetica integra* el autor MICHAEL STIEFEL (1487–1567), un monje alemán, usó ya en 1544 (un año antes de la aparición de *Ars magna*) números negativos para evitar la distinción de casos en la solución de la ecuación cuadrática. STIEFEL llamó a los números negativos *números absurdos*.

promueve una reflexión sobre la noción de número. Por un lado, no tardaron

Veamos un ejemplo de la ecuación (6.11)

$$x^3 = 6x + 4.$$

En este caso existen tres soluciones: $x = -2$, $x = 1 - \sqrt{3}$ y $x = 1 + \sqrt{3}$, de las cuales la última es positiva y se hubiera aceptada como solución aún en aquellos tiempos. Pero $\frac{p}{3} = 2$ y $\frac{q}{2} = 2$ y la fórmula de solución da

$$x = \sqrt[3]{\sqrt{2^2 - 2^3} + 2} - \sqrt[3]{\sqrt{2^2 - 2^3} - 2} = \sqrt[3]{\sqrt{-4} + 2} - \sqrt[3]{\sqrt{-4} - 2},$$

una expresión que contiene la raíz cuadrada de un número negativo. El conocimiento de casos como este abrió paso al descubrimiento y la aceptación de los números complejos. El nombre de *números imaginarios* para los números complejos, que acuñó RENÉ DESCARTES, refleja el escepticismo que se tenía en aceptar expresiones como $\sqrt{-4}$ como “números”. Se veían como construcciones auxiliares para llegar a una solución, pero no como números mismos.

Sin embargo, el uso de estas “construcciones auxiliares” mostró que se podía calcular sin problema con ellos. Con ello, y con el paso del tiempo, se empezaron a aceptar hasta llegar al punto de vista moderno: la pregunta si un número existe o qué es la naturaleza de un número no son preguntas importantes, sino preguntas aristotélicas, que no son fértiles. Lo que importa es que se puede calcular con los números. O dicho todavía de otra manera: un número no se debe ver como un ente aislado, sino en conjunto con todos los números con los cuales se puede operar con los cuatro operaciones básicas.

6.3.5. Abstracción: anillos y campos

A principios del siglo XX se llegó a la posición moderna respecto al concepto de “número”. Veamos primero la estructura de los números enteros. Es posible sumarlos y multiplicarlos. Además existe el 0 que es aquel número que satisface $a + 0 = a = 0 + a$ para cualquier número entero a . Se llama el *elemento neutro* respecto a la adición. Igualmente existe un elemento neutro para la multiplicación: es el 1 ya que $a \cdot 1 = a = 1 \cdot a$ para cualquier número entero a . Respecto al cero existen *inversos*, es decir, para cada número entero a existe un número b que satisface $a + b = 0$. Normalmente escribimos b como $-a$. Respecto a la multiplicación, no existen siempre inversos. Por ejemplo, $2 \cdot x = 1$ no tiene solución en los números enteros.

Tomemos ahora un punto de vista escéptico: no aceptamos a números negativos así que en lo que sigue “número” siempre significará “número natural positivo”. Ahora consideramos ecuaciones de la forma

$$x + a = b \tag{6.12}$$

donde x es la incógnita y a, b son parámetros, *que son números naturales positivos*. Rápidamente nos daremos cuenta que no siempre existe una solución dentro de los números conocidos, que son *números naturales positivos*, por ejemplo, si $a = 5$ y $b = 2$ entonces ningún número natural positivo satisface la ecuación $x + 5 = 2$. Sabemos que -3 es “la solución”, pero tenemos que ponernos en la situación de un matemático griego o de la Edad Media para comprender la dificultad de entender el concepto de los números negativos.

En lo que sigue deberíamos adoptar una posición similar: trataremos ser tercios y no aceptaremos “nuevos números”, hasta que no nos hayan explicado bien que es lo que son.

Con esta mentalidad veremos de nuevo nuestro ejemplo:

$$x + 5 = 2. \tag{6.13}$$

Lo que podemos decir es que si tuvieramos una solución de (6.13), entonces también tendríamos una solución de

$$x + 9 = 6 \tag{6.14}$$

ya que (6.14) se obtuvo de (6.13) sumando de ambos lados el número 4. Además, si tenemos ahora

¿Cómo sería la adición de estas soluciones? De nuevo seguimos el camino del ejemplo anterior:
De

$$\begin{aligned}x_{a,b} + a &= b \\ x_{c,d} + c &= d\end{aligned}\tag{6.20}$$

inferimos al sumar que $x_{a,b} + x_{c,d} + (a + c) = (b + d)$, es decir

$$x_{a,b} + x_{c,d} = x_{a+c,b+d}.\tag{6.21}$$

Para el producto $x_{a,b} \cdot x_{c,d}$ calculamos

$$\begin{array}{rcl}x_{a,b} + a = b & & | \cdot x_{c,d} \\ x_{a,b} \cdot x_{c,d} + a \cdot x_{c,d} = b \cdot x_{c,d} & & | +ac + bc \\ x_{a,b} \cdot x_{c,d} + a(x_{c,d} + c) + bc = b(x_{c,d} + c) + ac & & | \text{sustituir } x_{c,d} + c = d \\ x_{a,b} \cdot x_{c,d} + (ad + bc) = (ac + bd)\end{array}$$

es decir la regla de la multiplicación es

$$x_{a,b} \cdot x_{c,d} = x_{ad+bc,ac+bd}\tag{6.22}$$

En conclusión, si fingimos que una solución de $x + a = b$ existe, podemos calcular muy bien con ella. Podemos sumar usando (6.21) y multiplicar usando (6.22).

Finalmente llegamos al punto en donde ya no importa si existen o no estos números, podemos calcular con ellas y eso es lo que nos importa. Entonces podemos **definir el cero como la solución de la ecuación $x + a = a$ donde a es cualquier número**. De manera similar definimos -1 como la solución de $x + (a + 1) = a$ y en general $-n$ como la solución de $x + (a + n) = a$, donde a es cualquier número. **Así damos luz a los enteros:**

$$\dots, -7, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, \dots$$

La construcción de los números enteros a partir de los naturales es similar a la construcción de los números racionales a partir de los enteros. Se consideran ecuaciones de tipo

$$p \cdot y = q,$$

donde p y q son números enteros y $p \neq 0$. Una solución la llamamos $y_{p,q}$. Dado que $y_{p,q}$ es solución de $cp \cdot y = cq$ para cada entero $c \neq 0$ se tiene $y_{cp,cq} = y_{p,q}$. Si escribimos $y_{p,q}$ como el símbolo $\frac{q}{p}$ entonces esto equivale a decir que $\frac{q}{p} = \frac{cq}{cp}$, la equivalencia de fracciones. De manera similar podemos descubrir las reglas para operar con las fracciones.

Similarmente a escribir $\frac{q}{p}$ para $y_{p,q}$ podríamos haber escrito el símbolo $(b - a)$ en vez de $x_{a,b}$. Esto no lo hicimos porque solemos caer muy rápidamente en la tentación de usar o pensar en los números negativos a los cuales ya nos hemos acostumbrados.

También los complejos los podemos introducir como pares de números reales (r, s) . Definimos la suma y el producto como

$$\begin{aligned}(r, s) + (u, v) &= (r + u, s + v) \\ (r, s) \cdot (u, v) &= (ru - sv, rv + su)\end{aligned}$$

y podemos ver que con estos pares se puede calcular perfectamente bien. Podemos ver que $(0, 0)$ es el elemento neutro de la adición y $(1, 0)$ el de la multiplicación. No hay nada místico con el par $\mathbf{i} = (0, 1)$. No nos tenemos que preguntar si existe o no el símbolo \mathbf{i} y después de ver que podemos calcular bien con él, lo podemos aceptar como “número”.

6.3.6. El teorema fundamental del álgebra

Hemos visto en la sección anterior que el querer resolver ecuaciones cada vez más complicadas ha conducido a una expansión del concepto de número pero también del esclarecimiento del mismo concepto. La pregunta que ahora queremos tocar es que: si hay algún fin esta expansión. Hasta ahora ha sido exitoso la idea de considerar ecuaciones que no se pueden resolver en un conjunto dado de números para poder expandir el concepto de números. Pero, ¿siempre será posible seguir así expandir la noción de número?

La respuesta a esta pregunta es negativa, si se consideran – como hasta ahora se ha hecho – sólo ecuaciones *algebraicas*, es decir ecuaciones de tipo

$$x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \dots + a_2x^2 + a_1x + a_0 = 0, \quad (6.23)$$

con $n > 0$, donde los coeficientes $a_0, a_1, a_2, \dots, a_{n-1}, a_{n-1}$ pertenecen a un campo de números ya conocidos.

El teorema fundamental del álgebra dice: aunque los coeficientes $a_0, a_1, a_2, \dots, a_{n-1}, a_{n-1}$ se eligen libremente en el campo de los números complejos, la ecuación (6.23) siempre tendrá una solución en los números complejos. Por ello no será posible expandir con esta idea más allá de los números complejos. El campo de los números complejos constituye una piedra final a este proceso.

Bibliografía

Ars Magna or the rules of algebra, de *Girolamo Cardano*, traducido por T. Richard Wittmer.

6.4. Los Números complejos y las funciones complejas

6.5. El álgebra lineal

El algoritmo de Fang-Cheng

El siguiente ejercicio se encuentra en un libro chino titulado *Nueve capítulos sobre el arte matemático* y que data del siglo I. En él se usa una medida de capacidad antigua “Dou” para el trigo.

Tres gavillas de una buena cosecha, dos gavillas de una cosecha mediana y una gavilla de una mala cosecha dan 39 Dous.

Dos gavillas de una buena cosecha, tres gavillas de una cosecha mediana y una gavilla de una mala cosecha dan 34 Dous.

Una gavilla de una buena cosecha, dos gavillas de una cosecha mediana y tres gavillas de una mala cosecha dan 26 Dous.

¿Cuántos Dous resultan de una gavilla de buena cosecha, cuántos de una gavilla de una cosecha mediana y cuántos de una gavilla de una mala cosecha?

El libro en el que aparece este problema trata nueve tipos de problemas matemáticos, uno de ellos es el de resolver ecuaciones lineales simultáneas. Después de dar la solución indica cómo se puede llegar a ella. Los chinos solían hacer sus cálculos con palillos de bambú. Los acomodaban en un arreglo rectangular, que llamaban *Fang Cheng*. Así es cómo explican el procedimiento para llegar a la solución: primero se llena un arreglo donde la columna derecha corresponde a la primera condición descrita en el enunciado – hay que recordar que en chino se lee de derecha a izquierda, así que para ellos la columna a la derecha es la *primera*.

1	2	3
2	3	2
3	1	1
26	34	39

Ahora se multiplican los números de la columna de en medio por el número de gavillas de buena cosecha (columna derecha), es decir, por tres. Se obtiene la tabla que se muestra a la izquierda en la siguiente figura. Después se resta la columna del lado derecho de la de en medio tantas veces como sea necesario hasta que ya no queden gavillas de buena cosecha en ella:

1	2	3	→	1	6	3	→	1	3	3	→	1		3
2	3	2		2	9	2		2	7	2		2	5	2
3	1	1		3	3	1		3	2	1		3	1	1
26	34	39		26	102	39		26	63	39		26	24	39

De manera similar se trata la columna a la izquierda:

1		3	→	3		3	→			3
2	5	2		6	5	2		4	5	2
3	1	1		9	1	1		8	1	1
26	24	39		78	24	39		39	24	39

Luego se multiplica la columna izquierda por el número de gavillas de cosecha media de indica la columna de en medio, es decir, por 5, y se resta ésta última a la columna izquierda varias veces, hasta que no queden en ella gavillas de media cosecha:

		3			3			3			3											
4	5	2	→	20	5	2	→	15	5	2	→	10	5	2	→	5	5	2	→		5	2
8	1	1		40	1	1		39	1	1		38	1	1		37	1	1		36	1	1
39	24	39		155	24	39		171	24	39		147	24	39		123	24	39		99	24	39

Esto muestra que 36 gavillas de mala cosecha dan 99 Dous, por lo que una gavilla de mala cosecha da $\frac{99}{36} = \frac{11}{4} = 2\frac{3}{4}$ Dous. Luego, restando $2\frac{3}{4}$ a 24, se llega a la conclusión de que 5 gavillas de cosecha media dan $21\frac{1}{4} = 20\frac{5}{4}$ Dous, es decir, una gavilla de cosecha mediana da $4 + \frac{1}{4}$ Dous. Finalmente llegan a la conclusión de que una de buena cosecha da $9 + \frac{1}{4}$ Dous por gavilla.

Ejercicio. Se sugiere al lector completar las operaciones necesarias para obtener el resultado.

Se ve que el algoritmo que usaron los chinos en este caso no es otra cosa que el que solemos llamar *algoritmo de eliminación de Gauss* en honor al matemático CARL FRIEDRICH GAUSS que vivió de 1777 a 1855. Podríamos decir que los chinos hace 2000 años trabajaron con *matrices*, es decir la esencia de la ecuación. Hoy en día voltearíamos la matriz y la escribiríamos de la siguiente manera:

3	2	1	39
2	3	1	34
1	2	3	26

porque así aparecen los coeficientes en las ecuaciones:

$$\begin{cases} 3x + 2y + 1z = 39 \\ 2x + 3y + 1z = 34 \\ 1x + 2y + 3z = 26. \end{cases} \quad (6.24)$$

El desarrollo del álgebra lineal en Europa

En Europa se desarrolló el álgebra lineal con mayor profundidad en los siglos XVIII y XIX. En 1545 GIROLAMO CARDANO dio las fórmulas de solución para sistemas de dos ecuaciones lineales en dos incógnitas. Éstas fórmulas se pueden obtener al llevar a cabo el algoritmo de Fang Cheng para dos ecuaciones lineales en dos incógnitas en general. Hoy día se llaman *la regla de Cramer* en honor al matemático GABRIEL CRAMER (1704–1752), quien a la edad de 18 recibe el título de doctor y con 20 se vuelve profesor de la universidad de Ginebra en Suiza.



GABRIEL CRAMER (1704–1750)

CRAMER formula 1750 la solución general para n ecuaciones lineales en n incógnitas: Si el sistema es

$$\begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \dots + a_{1,n}x_n = b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \dots + a_{2,n}x_n = b_2 \\ \vdots \\ a_{n,1}x_1 + a_{n,2}x_2 + \dots + a_{n,n}x_n = b_n \end{cases} \quad (6.25)$$

se considera el determinante $\det(A)$ de la matrix de coeficientes

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{bmatrix}.$$

Si $\det(A) \neq 0$ entonces la solución del sistema es

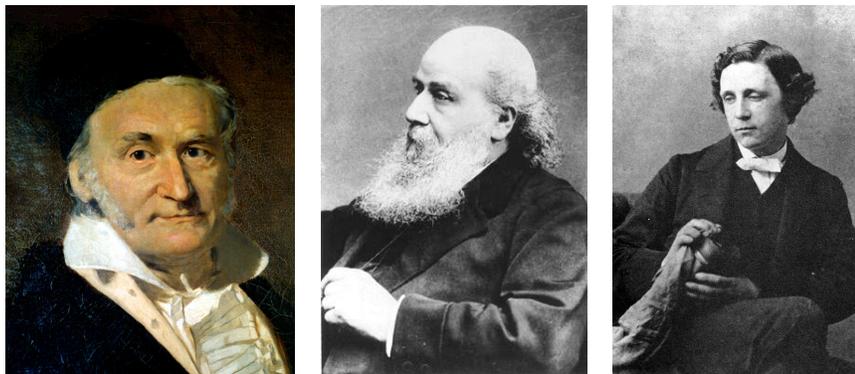
$$x_1 = \frac{\det(A_1)}{\det(A)}, \quad x_2 = \frac{\det(A_2)}{\det(A)}, \quad \dots \quad x_n = \frac{\det(A_n)}{\det(A)},$$

donde A_i es la matrix que se obtiene de A al sustituir la i -ésima columna por b , el vector con las entradas b_1, \dots, b_n . El concepto de *determinante* la introdujo GOTTFRIED WILLHELM LEIBNIZ aunque bajo otro nombre y el resultado de CRAMER reanudó el interés en los determinantes: en 1764 ÉTIENNE BÉZOUT y en 1771 ALEXANDRE-THÉOPHILE VANDERMONDE publicaron resultados sobre determinantes. PIERRE-SIMON LAPLACE afirmó que estos métodos no eran practicables para calcular determinantes de manera eficiente. En 1772 LAPLACE publica el método de desarrollo de determinantes (por filas o columnas).



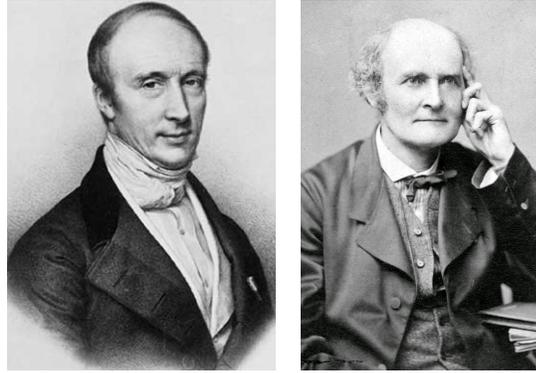
GOTTFRIED WILLHELM LEIBNIZ, ÉTIENNE BÉZOUT y PIERRE-SIMON LAPLACE

Hasta este momento todavía no se usan los términos *determinante* o *matriz*. Apenas en el año 1801 CARL FRIEDRICH GAUSS introduce el nombre de *determinante* y más tarde GAUSS muestra cómo se puede obtener la solución de un sistema de ecuaciones lineales de manera sistemático transformando las ecuaciones sin tener que calcular determinantes. Pero se refiere al método como “ampliamente conocido”. De cualquier manera, el nombre de este procedimiento queda asociado a su nombre hasta hoy en día. El término *matriz* lo introdujo JAMES JOSEPH SYLVESTER en 1850. SYLVESTER trabajó con el caso general en donde el número de ecuaciones y el número de incógnitas no tienen que coincidir. Por ello tiene que extraer del arreglo rectangular de coeficientes subdeterminantes. De ahí el nombre de la *matriz*: es el arreglo del que todo proviene. El matemático CHARLES LUTWIDGE DODGSON, mejor conocido bajo su seudónimo LEWIS CARROLL trató de cambiar el nombre de *matriz* a *bloque*, pero fracasó en ello como ahora sabemos.



CARL FRIEDRICH GAUSS, JAMES JOSEPH SYLVESTER y CHARLES LUTWIDGE DODGSON

Trabajos muy influyentes sobre determinantes los produjo AUGUSTIN-LOUIS CAUCHY en 1812 y sobre matrices ARTHUR CAYLEY en 1858.



AUGUSTIN-LOUIS CAUCHY y ARTHUR CAYLEY

Así vimos pasar más que 100 años de desarrollo. Es interesante que al principio – si dejamos de lado los conocimientos de los chinos – está la fórmula de solución general y no el método de la eliminación de GAUSS que es más practicable porque requiere de llevar a cabo mucho menos operaciones.

Lo ideal: un caso raro

El algoritmo, llamáse *algoritmo de Fang Cheng* o *algoritmo de Gauss*, produce una cierta forma de matriz, que se llama *escalonada*. Veamos esto en un ejemplo concreto.

Empezamos con el siguiente sistema de 4 ecuaciones de 5 incógnitas:

$$\begin{cases} x_1 + 2x_3 + 3x_5 = 1 \\ 2x_2 + x_3 - x_4 - x_5 = 4 \\ 2x_1 + 2x_2 + 5x_3 + 9x_5 = 8 \\ x_1 + 2x_2 + 3x_3 + x_4 + 10x_5 = 9 \end{cases} \quad (6.26)$$

Representamos el sistema por la matriz

$$M = \left[\begin{array}{ccccc|c} 1 & 0 & 2 & 0 & 3 & 1 \\ 0 & 2 & 1 & -1 & -1 & 4 \\ 2 & 2 & 5 & 0 & 9 & 8 \\ 1 & 2 & 3 & 1 & 10 & 9 \end{array} \right]$$

La división vertical sirve para acordarnos que del lado derecho están los constantes. Restamos dos veces la primera fila de la tercera y una vez la primera de la última, y obtenemos

$$\left[\begin{array}{ccccc|c} 1 & 0 & 2 & 0 & 3 & 1 \\ 0 & 2 & 1 & -1 & -1 & 4 \\ 0 & 2 & 1 & 0 & 3 & 6 \\ 0 & 2 & 1 & 1 & 7 & 8 \end{array} \right]$$

El efecto es que en la primera columna solo hay un coeficiente diferente de cero. Ahora restamos una vez la segunda fila de la tercera y de la cuarta fila:

$$\left[\begin{array}{ccccc|c} 1 & 0 & 2 & 0 & 3 & 1 \\ 0 & 2 & 1 & -1 & -1 & 4 \\ 0 & 0 & 0 & 1 & 4 & 2 \\ 0 & 0 & 0 & 2 & 8 & 4 \end{array} \right]$$

Ahora la segunda y tercera fila solo ceros a partir de la segunda entrada. Por ello nos fijamos en la cuarta columna. Restamos la tercera fila dos veces de la cuarta.

$$\left[\begin{array}{ccccc|c} 1 & 0 & 2 & 0 & 3 & 1 \\ 0 & 2 & 1 & -1 & -1 & 4 \\ 0 & 0 & 0 & 1 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

Las columnas 1, 2 y 4 se llaman *pivotes*. Podemos obtener ceros arriba de las últimas entradas no cero en los pivotes. Para ello sumamos la tercera fila a la segunda:

$$\left[\begin{array}{ccccc|c} 1 & 0 & 2 & 0 & 3 & 1 \\ 0 & 2 & 1 & 0 & 3 & 6 \\ 0 & 0 & 0 & 1 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

Finalmente multiplicamos la segunda fila por $\frac{1}{2}$ para obtener un 1 en el segundo pivote:

$$\left[\begin{array}{ccccc|c} 1 & 0 & 2 & 0 & 3 & 1 \\ 0 & 1 & \frac{1}{2} & 0 & \frac{3}{2} & 3 \\ 0 & 0 & 0 & 1 & 4 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

Esta es la forma *escalonada*. Se obtiene de la matriz original M por medio de transformaciones de filas. Dado que la cuarta fila es cero, sólo cofifica tres ecuaciones:

$$\begin{cases} x_1 + 2x_3 + 3x_5 = 1 \\ x_2 + \frac{1}{2}x_3 + \frac{3}{2}x_5 = 3 \\ x_4 + 4x_5 = 2 \end{cases} \quad (6.27)$$

Este sistema de ecuaciones tiene las mismas soluciones que (??) ya que sólo se sumaron y restaron unas filas de otras y se multiplicó una fila por un número distinto de cero. De la forma (??) es fácil de indicar la solución: los valores de x_3 y x_5 se pueden elegir libremente y después se obtienen los

variables de los pivotes:

$$\begin{cases} x_1 = 1 - 2x_3 - 3x_5 \\ x_2 = 3 - \frac{1}{2}x_3 - \frac{3}{2}x_5 \\ x_4 = 2 - 4x_5 \end{cases} \quad (6.28)$$

Con ello el conjunto de solución es $\{(1 - 2x_3 - 3x_5, 3 - \frac{1}{2}x_3 - \frac{3}{2}x_5, x_3, 2 - 4x_5, x_5) \mid x_3, x_5 \in \mathbb{R}\}$.

La reducción de filas se puede hacer totalmente mecánico, por lo que este método es ideal para implementaciones en algoritmos. Es un caso *ideal* en las matemáticas: la solución siempre se puede describir por completo y de manera exacta. Además hay un método claro y eficiente para llegar a ella. Hay muy pocas situaciones en matemáticas así, normalmente en matemáticas es todo al revés: no se puede describir la solución de manera exacta y no hay un camino claro para llegar a ello.

La linearización

No todas las situaciones son tan ideales como las de los sistemas de ecuaciones lineales. Por ello se trata de reducir casos más complejos a este caso ideal. Veremos primero unos ejemplos con una sola incógnita y después un ejemplo con varias ecuaciones no lineales. En cada uno de los casos se reduce el problema a una ecuación lineal o a un sistema de ecuaciones lineales.

Empecemos con un ejemplo sencillo: una aproximación de $\sqrt{2}$. Uno se puede preguntar ¿por qué esto se tiene que aproximar? Se obtiene el valor fácilmente con una calculadora de bolsillo. Pero esto es sólo trasladar la pregunta, porque uno debe preguntarse entonces ¿cómo lo hace la calculadora para calcular la raíz cuadrada de un número? La *linearización* nos da una herramienta poderosa para calcular con pocos pasos aproximaciones buenas. La idea central de este método es que a partir de una aproximación dada se puede encontrar una aproximación mejor.

El número $x = \sqrt{2}$ una de las dos soluciones de la ecuación

$$x^2 = 2. \quad (6.29)$$

Partimos de alguna aproximación para $\sqrt{2}$, por ejemplo $a = 1$. Esta aproximación no tiene que ser particularmente buena, ya que el mismo procedimiento nos dará cada vez mejores aproximaciones. Ahora consideramos la diferencia δ entre $x = \sqrt{2}$ y la aproximación a y entonces $x = a + \delta$. Sustituimos entonces x en (6.29) por $a + \delta$ y obtenemos:

$$a^2 + 2a\delta + \delta^2 = 2. \quad (6.30)$$

Si la aproximación es más o menos buena, el error δ será más o menos pequeño, y δ^2 será aún más pequeño. La linearización consiste en olvidar de la ecuación (6.30) el término δ^2 :

$$a^2 + 2a\delta \approx 2. \quad (6.31)$$

Así obtenemos ya no una ecuación exacta sino solo una aproximación a 2 del lado izquierdo. Sustituimos (6.31) por una igualdad con una nueva variable d en vez de δ :

$$a^2 + 2ad = 2. \quad (6.32)$$

entonces esta ecuación es lineal en d y la podemos resolver:

$$d = \frac{2 - a^2}{2a}.$$

En la espera que la diferencia de los dos lados de (6.32) sea pequeña tenemos $d \approx \delta$ y por lo tanto

$$a_1 = a + d = a + \frac{2 - a^2}{2a} \quad (6.33)$$

será una mejor aproximación que a . Ponemos esto a prueba para calcular una aproximación de $\sqrt{2}$ a partir de $a = 1$:

$$\begin{aligned} a_1 &= 1 + \frac{2 - 1^2}{2 \cdot 1} = 1.5 \\ a_2 &= 1.5 + \frac{2 - 1.5^2}{2 \cdot 1.5} = 1.41\bar{6} \\ a_3 &= 1.41\bar{6} + \frac{2 - 1.41\bar{6}^2}{2 \cdot 1.41\bar{6}} \approx 1.414216 \end{aligned}$$

un resultado que después de solo tres pasos ya tiene una exactitud de 4 cifras decimales.

Este método corresponde al algoritmo de HERÓN DE ALEXANDRIA. Según él se parte de un valor inicial $a = 1$ y luego se forma un rectángulo de área 2 que tiene un lado igual a $a = 1$ El otro lado es entonces 2 y el promedio de estos dos es

$$a_1 = \frac{a + \frac{2}{a}}{2} = \frac{a^2 + 2}{2a} = a + \frac{2 - a^2}{2a},$$

que es la misma fórmula que (6.33). La idea de HERÓN es que el promedio de ambos lados del rectángulo con área 2 es más cerca de $\sqrt{2}$ que cualquiera de sus dos lados.

Veamos ahora un segundo ejemplo de ecuación, donde el problema de encontrar una solución es más serio:

$$x^5 - x - 1 = 0. \quad (6.34)$$

Se puede mostrar que no existe una fórmula general para resolver ecuaciones de grado mayor que 4 y en particular que la solución de (6.33) no se puede escribir de ninguna manera con raíces. Lo único que nos queda es tratar de aproximar una solución. Usaremos la linearización: a partir de una primera estimación a consideramos el error $\delta = x - a$ y con ello $x = a + \delta$. Por ello, la sustitución en (6.33) da

$$(a + \delta)^5 + (a + \delta) - 1 = 0.$$

Ahora seguimos el procedimiento de arriba: expandimos esta ecuación, olvidamos sumandos en donde δ aparece en una potencia mayor o igual que 2 y sustituimos δ por d y resolvemos por esta variable:

$$\begin{aligned} a^5 + 5a^4\delta + 10a^3\delta^2 + 10a^2\delta^3 + 5a\delta^4 + \delta^5 + a + \delta - 1 &= 0 \\ a^5 + 5a^4\delta + a + \delta - 1 &\approx 0 \\ a^5 + 5a^4d + a + d - 1 &= 0 \\ d &= -\frac{a^5 + a - 1}{5a^4 + 1} \end{aligned}$$

Con ello obtenemos a partir de una aproximación a la siguiente aproximación

$$a_1 = a - \frac{a^5 + a - 1}{5a^4 + 1}. \quad (6.35)$$

Si partimos de $a = 1$ obtenemos con una aproximación de 6 decimales

$$a_1 = 0.833333, \quad a_2 = 0.764382, \quad a_3 = 0.755025, \quad a_4 = 0.754878$$

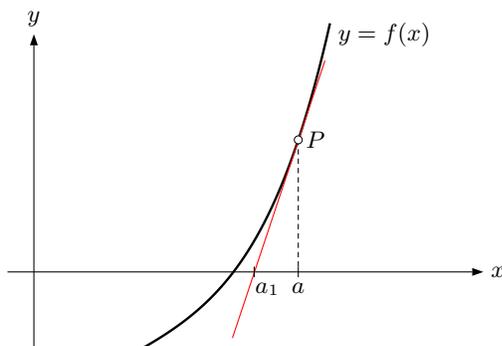
y $a_k \approx a_4$ para $k > 4$ con 6 decimales de exactitud.

Ejercicio. Para obtener una experiencia propia con la linearización se recomienda al lector resolver la ecuación cuadrática $x^2 + x = 1$ por la linearización y luego comprobar con la fórmula de solución para ecuaciones cuadráticas que las soluciones son correctas. Para obtener ambas soluciones con la linearización se tiene que partir de diferentes aproximaciones iniciales.

El método de linearización es equivalente al algoritmo de Newton para encontrar los ceros de una función f . En nuestro caso la función está dada por

$$f(x) = x^5 + x - 1.$$

A partir de una aproximación inicial a se calcula la tangente t a la gráfica de f en el punto $P(a, f(a))$ y se toma el punto de intersección de t con el eje x como la siguiente aproximación a_1 .



La tangente t tiene pendiente $f'(a)$ y pasa por el punto P y por ello su ecuación es

$$y - f(a) = f'(a) \cdot (x - a).$$

Dado que queremos calcular la intersección con el eje x se debe tener $y = 0$ por lo que

$$\begin{aligned} f'(a) \cdot (x - a) &= -f(a) \\ x - a &= \frac{-f(a)}{f'(a)} \\ x &= a + \frac{f(a)}{f'(a)}, \end{aligned}$$

que es justamente (6.35) en el caso $f(x) = x^5 + x - 1$ ya que $f'(a) = 5a^4 + 1$.

El sistema de posicionamiento global

La situación

Cuando se trata de *calcular la propia posición* con el uso del *sistema de posicionamiento global*, en breve GPS, por las letras en inglés, se presenta el siguiente problema: conocidos son las posiciones de tres emisoras y las distancias a ellas, y desconocido es la propia ubicación del receptor. Las emisoras del GPS son satélites que circulan alrededor de la Tierra y emiten una señal todo el tiempo que permiten calcular su posición en un momento con mucha precisión. La distancia a ellos se puede obtener al calcular el lapso de tiempo que pasa desde la emisión de la señal hasta que ésta se recibe en el receptor. Al final del texto daremos unos detalles más que muestran que el cálculo que realmente hay que hacer es sumamente más complicado y que aquí sólo tratamos un caso simplificado.

Resumimos lo que se sabe y lo que se desconoce:

Conocido:	La posición de tres emisoras	$(x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3)$
	La distancia de la propia posición a estas emisoras:	r_1, r_2, r_3
Desconocido:	La propia posición	(x_P, y_P, z_P)

Una sistema de ecuaciones exactas

El problema se puede escribir como un sistema de ecuaciones *no-lineales*:

$$(x_1 - x_P)^2 + (y_1 - y_P)^2 + (z_1 - z_P)^2 = r_1^2 \quad (6.36)$$

$$(x_2 - x_P)^2 + (y_2 - y_P)^2 + (z_2 - z_P)^2 = r_2^2 \quad (6.37)$$

$$(x_3 - x_P)^2 + (y_3 - y_P)^2 + (z_3 - z_P)^2 = r_3^2 \quad (6.38)$$

Esto equivale a intersectar tres esferas en el espacio. Hay que esperar que existen dos soluciones, pero en la práctica no se resuelve este sistema, sino este se lineariza. La razón es la medición de las distancias se tiene que corregir porque la atmósfera altera la velocidad con la que viajan las señales y por ello el tiempo que dura. La linearización tiene la ventaja adicional que es fácil de programar.

La aproximación

En vez de resolver el sistema (6.36), (6.37), (6.38) de manera directa, se parte de la situación que uno cuenta con una aproximación Q de la posición actual P , por ejemplo la última posición calculada. Ahora se expresa la posición actual (x_P, y_P, z_P) como la posición aproximada (x_Q, y_Q, z_Q) más un desconocido error

$$x_P = x_Q + \Delta x, \quad y_P = y_Q + \Delta y, \quad z_P = z_Q + \Delta z. \quad (6.39)$$

Esto se puede sustituir en las ecuaciones (6.36) a (6.38). Calculamos aquí el resultado de la sustitución en (6.36):

$$\begin{aligned} (x_1 - x_Q - \Delta x)^2 + (y_1 - y_Q - \Delta y)^2 + (z_1 - z_Q - \Delta z)^2 &= r_1^2 \\ (x_1 - x_Q)^2 - 2(x_1 - x_Q)\Delta x + (\Delta x)^2 + (y_1 - y_Q)^2 + \\ -2(y_1 - y_Q)\Delta y + (\Delta y)^2 + (z_1 - z_Q)^2 - 2(z_1 - z_Q)\Delta z + (\Delta z)^2 &= r_1^2 \end{aligned} \quad (6.40)$$

La suma

$$s_1^2 = (x_1 - x_Q)^2 + (y_1 - y_Q)^2 + (z_1 - z_Q)^2$$

es el cuadrado de la distancia conocida s_1 entre la posición aproximativa (x_Q, y_Q, z_Q) y la posición del primer satélite (x_1, y_1, z_1) . La suma

$$\Delta^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$$

es el cuadrado de la distancia Δ entre la posición desconocida actual (x_P, y_P, z_P) y la posición aproximativa (x_Q, y_Q, z_Q) . La ecuación (6.40) se puede simplificar a

$$-2(x_1 - x_Q)\Delta x - 2(y_1 - y_Q)\Delta y - 2(z_1 - z_Q)\Delta z + \Delta^2 = r_1^2 - s_1^2 \quad (6.41)$$

Como Δ^2 se considera pequeño en contraste con las demás distancias se omite de la ecuación (6.41). Así ya no se tiene una ecuación exacta sino una aproximativa:

$$(x_1 - x_Q)\Delta x + (y_1 - y_Q)\Delta y + (z_1 - z_Q)\Delta z \approx \frac{s_1^2 - r_1^2}{2}. \quad (6.42)$$

Se *linearizó* la ecuación, ya que se quitó el término cuadrático. De manera análogo se obtienen las aproximaciones:

$$(x_2 - x_Q)\Delta x + (y_2 - y_Q)\Delta y + (z_2 - z_Q)\Delta z \approx \frac{s_2^2 - r_2^2}{2}, \quad (6.43)$$

$$(x_3 - x_Q)\Delta x + (y_3 - y_Q)\Delta y + (z_3 - z_Q)\Delta z \approx \frac{s_3^2 - r_3^2}{2}. \quad (6.44)$$

Las tres aproximaciones se deben resolver *simultáneamente*. Si se sustituye el símbolo de aproximación por uno de igualdad en (6.42) a (6.44), entonces se obtiene el siguiente sistema de ecuaciones lineales:

$$(x_1 - x_Q)\Delta x' + (y_1 - y_Q)\Delta y' + (z_1 - z_Q)\Delta z' = \frac{s_1^2 - r_1^2}{2}, \quad (6.45)$$

$$(x_2 - x_Q)\Delta x' + (y_2 - y_Q)\Delta y' + (z_2 - z_Q)\Delta z' = \frac{s_2^2 - r_2^2}{2}, \quad (6.46)$$

$$(x_3 - x_Q)\Delta x' + (y_3 - y_Q)\Delta y' + (z_3 - z_Q)\Delta z' = \frac{s_3^2 - r_3^2}{2}. \quad (6.47)$$

En ello se sustituyeron los errores verdaderos Δx , Δy , Δz por (desconocidos) errores aproximados: $\Delta x'$, $\Delta y'$, $\Delta z'$. Recordamos:

Conocido: La posición de tres emisoras $(x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3)$
 una posición aproximada (x_1, y_Q, z_Q)
 La distancia de la posición aproximada a las emisoras: s_1, s_2, s_3

Desconocido: Los errores aproximativos $\Delta x', \Delta y', \Delta z'$

El sistema de ecuaciones (6.45)–(6.47) es *lineal* en los errores aproximativos $\Delta x', \Delta y'$ y $\Delta z'$. Por ello se puede resolver sin problema.

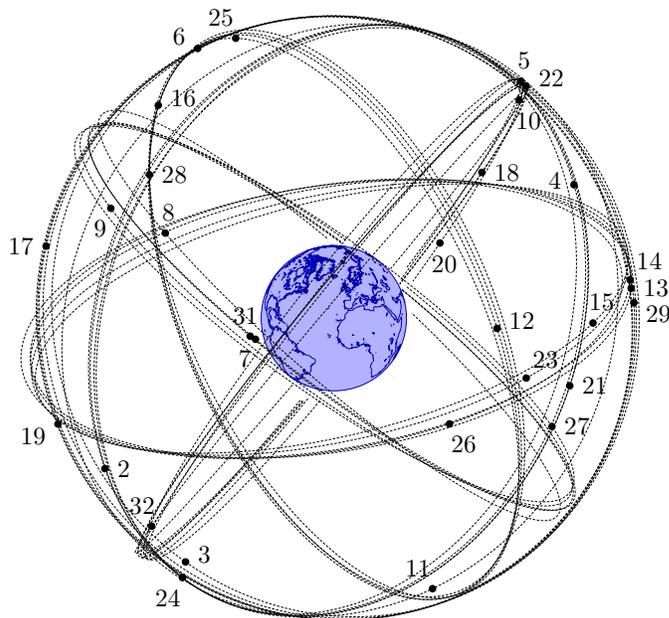
Con ello se determinan los errores aproximativos. Así se obtiene una nueva posición aproximativa Q_2 :

$$Q_2(x_Q + \Delta x', y_Q + \Delta y', z_Q + \Delta z'),$$

Después se itera la aproximación un par de veces para obtener una buena aproximación. Con ello se determina posición P .

Detalles acerca del GPS

El GPS se basa en un sistema de satélites que rodean la Tierra de tal manera que desde cualquier punto siempre se puede ver al menos 4 de ellos. La siguiente figura muestra la posición de los satélites a escala el día 8 de junio de 2011 a la hora 19:50:24. En este momento hubo 30 satélites. Los números en la figura muestra el número identificador de los satélites en el GPS.



Los satélites son relativamente pesados: la generación más moderna pesa alrededor de dos toneladas cada uno. Este peso se debe en parte a los tres o cuatro relojes atómicos que se encuentran en su interior. Estos relojes permiten una medición del tiempo muy precisa. Esta precisión es necesaria ya que la velocidad de la luz c es tan enormemente grande:

$$c = 299\,792\,458 \text{ m/s.}$$

Si se quiere medir la posición a un metro de precisión se tiene que medir el tiempo con la precisión correspondiente: la luz transcurre un metro en el tiempo $\frac{1\text{m}}{c} = 0.000\,000\,003\,335 \text{ s.}$

Los satélites emiten una señal constante que se repite cada 30 segundos y que indica el tiempo actual, la posición y datos adicionales. En la Tierra estos datos se pueden recibir. Pero los receptores usualmente no cuentan con un reloj atómico sino de un reloj de cuarzo que es en comparación poco preciso. Por ello no es posible medir el tiempo t_i que pasa desde la emisión hasta la recepción de manera directa. Lo que se hace es introducir una cuarta variable: Δt , el error en la medición del tiempo por el reloj de cuarzo. Por ello la distancia r_i no es conocida sino se tiene que sustituir por

$$r_i = c \cdot (t_i + \Delta t)$$

Esto explica por qué se requiere recibir la señal de cuatro satélites y no de tres: se tiene que linearizar a un sistema de cuatro incógnitas y por ello se requiere de cuatro ecuaciones.

Para lograr una medición con la precisión de un metro, esto aún no basta. La velocidad de la luz es menor en el aire y mayor en el vacío. El tiempo que transcurre desde la emisión hasta la recepción no se puede convertir en una distancia de inmediato. Responsable es la atmósfera y en particular la capa superior llamada ionósfera. Para complicar las cosas más, hay que tomar en cuenta la forma actual de la ionósfera: esta varía y los satélites mandan entre sus datos también coeficientes para corregir estos efectos.

El ejemplo del GPS muestra: en parte se puede comprender cómo se logra simplificar los cálculos que se hacen dentro de nuestros aparatos a diario, en parte usan unas matemáticas muy avanzadas.

6.6. Álgebra moderna, álgebra abstracta

Importancia histórica y cultural, que significó la simbología algebraica que se usa en la actualidad.

La importancia de las propiedades, o axiomas básicos, de los diversos sistemas numéricos se reconoce hasta hace relativamente poco tiempo. En ello influyó mucho que en algunas abstracciones superiores como los polinomios los vectores, las matrices y las transformaciones, se definen de manera natural operaciones que cumplen propiedades similares a las de los números. Estudiar las estructuras que forman los objetos abstractos con sus operaciones es el objeto del Álgebra superior.

El concepto de grupo. Las matrices tienen una estructura similar en la que, en el caso de matrices de tres o más dimensiones, el producto (o la composición) no es conmutativo.

Las estructuras algebraicas se usan en las matemáticas más avanzadas con mucho provecho y que en particular la Teoría de las ecuaciones polinomiales pudo completarse gracias a la aplicación de estas estructuras. También la demostración de que ciertas construcciones como la trisección de un ángulo, la duplicación del cubo y la cuadratura del círculo no pueden realizarse con regla y compás se hizo utilizando estructuras algebraicas especiales.

Consolidación del álgebra abstracta en el siglo XX: (Artin y Noether). El álgebra moderna. El proyecto de clasificación de grupos simples: un proyecto de investigación que rompe límites y pone en duda lo que es una demostración.

La algebraización de varias áreas (Geometría, Teoría de números, Topología,...): un fenómeno del siglo XX.

Capítulo 7

Las coordenadas

El puente entre la geometría y el álgebra

7.1. La Geometría con Coordenadas

En el siglo XVII, unos cien años después de los trabajos de Dal Ferro, Tartaglia, Cardano y Ferrari, dos grandes matemáticos de la época: Pierre Fermat y René Descartes, inventaron la Geometría Analítica, uniendo los nuevos métodos del álgebra con los antiguos de la Geometría, a través del uso de coordenadas para representar los puntos y las figuras geométricas.

El método consistía en introducir sistemas de coordenadas en el plano y describir las figuras geométricas, como soluciones de ecuaciones con esas coordenadas como variables. Por ejemplo, si el sistema de coordenadas es ortogonal (es decir el eje x y el eje y son perpendiculares), que es como suelen utilizarse hoy en día los sistemas de coordenadas, la ecuación algebraica

$$x^2 + y^2 = r^2$$

representa una circunferencia con centro en el origen. Esto es gracias al Teorema de Pitágoras aplicado al triángulo rectángulo de catetos x, y e hipotenusa r pues nos asegura que la igualdad

$$x^2 + y^2 = r^2$$

se satisface si y solo si el punto (x,y) está a una distancia r del origen (0,0). y por tanto en la circunferencia con centro en el origen y radio r.

Este método de representación de curvas mediante ecuaciones algebraicas resultó tremendamente eficaz para resolver problemas que no se habían podido resolver con otros métodos y de hecho generó una gran cantidad de curvas nuevas que antes ni siquiera se podían describir. Muchos siglos antes, alrededor de 200 A.C, Apolonio de Perga escribió un tratado muy detallado sobre las secciones cónicas, que son las curvas que se obtienen al cortar un cono con un plano. Dichas curvas reciben los nombres de elipse, parábola e hipérbola según el plano con que se corta el cono tenga menor, igual o mayor inclinación que sus elementos, es decir, las rectas que forman al cono. La Geometría

Analítica permitió demostrar que las curvas cónicas corresponden exactamente a las ecuaciones de segundo grado en x y y , es decir a las ecuaciones de la forma

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

Donde a , b , c , d , e , y f son constantes.

Apolonio había estudiado sobre todo las llamadas propiedades de reflexión de las cónicas, que son útiles por ejemplo en la construcción de espejos reflectores. Pero las curvas cónicas resultaron ser mucho más útiles de lo que su campeón Apolonio pudo jamás imaginar pues a la larga se descubrió que las trayectorias de los cuerpos que se lanzan al aire son parábolas (cuando la resistencia del aire es despreciable debido a la forma y peso del proyectil). Pero más importante aún, las trayectorias de los planetas alrededor del sol son elipses y las de los cometas pueden ser elipses, parábolas e incluso hipérbolas. Johannes Kepler analizó con gran detalle la trayectoria del planeta Marte y tras años de cálculos detallados y muchas hipótesis sobre la naturaleza de la órbita rechazadas, llegó a la conclusión de que la órbita era una elipse con el Sol en uno de sus focos. Poco tardó en demostrar Kepler que lo mismo ocurría para los demás planetas. Gracias a este descubrimiento y a las otras dos leyes del movimiento planetario que descubrió Kepler: que el radio vector de un planeta al Sol barre áreas iguales en tiempos iguales y que los cuadrados de los tiempos de revolución de los planetas son proporcionales a los cubos de los semiejes mayores de sus órbitas, Newton pudo realizar sus grandes descubrimientos sobre las leyes del movimiento de los cuerpos y a unificar así el estudio del movimiento de los cuerpos terrestres y los celestes al demostrar que ambos respondían a las mismas leyes.

Sin duda la Geometría Analítica resultó una herramienta importantísima para el avance de la Ciencia a partir del siglo XVII.

La Geometría analítica como tecnología

La fusión entre la geometría y el álgebra que realizaron Descartes y Fermat ofreció una herramienta tecnológica sin precedente que permitía resolver muchos problemas geométricos con métodos nuevos y de maneras mucho más eficientes. Para ilustrar esto baste un ejemplo: el llamado problema de Apolonio consistente en encontrar los círculos tangentes a otros tres círculos dados. La manera clásica de enfrentar este problema usando regla y compás, requirió de gran astucia por parte de Apolonio. De hecho no se conocen sus soluciones y las que se pueden ofrecer en la actualidad usando regla y compás son bastante elaboradas a pesar de utilizar conceptos avanzados como la inversión respecto a un círculo. Usando la geometría analítica la solución es el punto de intersección de dos hipérbolas y por tanto puede encontrarse resolviendo un sistema de ecuaciones cuadráticas.

7.2. Ecuaciones cuadráticas

Clasificación de cónicas

Clasificación de cuadráticas

7.3. Vectores

7.4. Curvas y superficies

la hipocicloide

la catenaria

la cardioide

la cycloide

etc

concepto de curvatura

por ecuación, o parametrizadas

7.5. Transformaciones proyectivas

7.6. La dimensión

El concepto de la *dimensión* es a la vez fácil de entender intuitivamente pero muy difícil de definir con rigor. Partimos en esta sección de una noción intuitiva y mostramos cómo ésta se vio en apuros debido a algunos descubrimientos realizados a finales del siglo XIX. Después presentamos dos maneras de poner la definición de la dimensión sobre bases más firmes.

El número de coordenadas

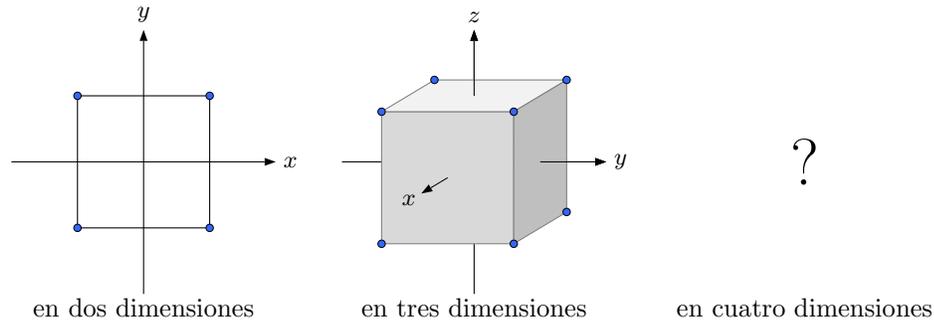
Una noción intuitiva de dimensión es el número de direcciones en las cuales nos podemos mover. Por ejemplo, si vamos en una carretera sólo nos podemos mover en una dimensión, es decir hacia delante o hacia atrás pero siempre a lo largo de la carretera. En una ciudad hay más libertad, nos podemos mover hacia delante o hacia atrás y independientemente de esto a la derecha o a la izquierda, es decir nos podemos mover en dos dimensiones. Dentro de un edificio además podemos subir y bajar, así que nos podemos mover en tres dimensiones.

Con coordenadas esto se expresa de manera sencilla: podemos decir que la dimensión es el número de coordenadas que se requieren para determinar la ubicación de un punto. Éste fue el punto de vista de BERNHARD RIEMANN. Con esta definición, una recta tiene dimensión uno, un plano dimensión dos y el espacio dimensión tres. Pero también el círculo tiene dimensión uno, ya que basta dar un ángulo para especificar la ubicación de un punto en él. La superficie de una esfera, o de la Tierra tiene dimensión dos, ya que un punto en ella se puede ubicar con dos coordenadas, la latitud y la longitud.

Una de las grandes ventajas de esta definición es que nada nos impide aumentar el número de coordenadas. Por ejemplo podemos describir un espacio de cuatro dimensiones utilizando cuatro coordenadas (x_1, x_2, x_3, x_4) . Veamos esto un poco más de cerca con una analogía. Empezamos en el plano y consideramos un cuadrado con sus cuatro esquinas en $(1, 1)$, $(1, -1)$, $(-1, 1)$ y $(-1, -1)$.

Esto se puede abreviar diciendo que las esquinas se encuentran en $(\pm 1, \pm 1)$, donde los signos se pueden elegir libremente. El cuadrado está delimitado por los cuatro lados $x = \pm 1$ y $y = \pm 1$.

El cubo es el cuerpo análogo en tres dimensiones a lo que es el cuadrado en dos dimensiones. Tiene los siguientes 8 vértices $(\pm 1, \pm 1, \pm 1)$. Sus caras son cuadrados que se encuentran en los planos $x = \pm 1$, $y = \pm 1$ y $z = \pm 1$. La cara $x = 1$ contiene a los vértices $(1, \pm 1, \pm 1)$, y la que está determinada por $y = -1$ contienen a los vértices $(\pm 1, -1, \pm 1)$.

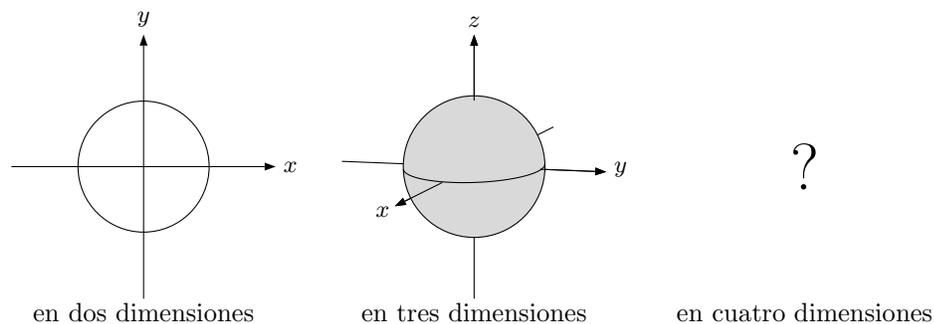


Ahora estamos listos para determinar lo que es el objeto análogo en cuatro dimensiones, que se llama *hipercubo*. No es posible representarlo como objeto de nuestro espacio que solo tiene tres dimensiones, pero es posible describir sus propiedades, porque podemos manipular sus coordenadas.

Por analogía inferimos que el hipercubo tiene 16 esquinas $(\pm 1, \pm 1, \pm 1, \pm 1)$. Sus *hipercaras* se encuentran en los *subespacios* de tres dimensiones que están descritos por las ecuaciones $x_1 = \pm 1$, $x_2 = \pm 1$, $x_3 = \pm 1$ y $x_4 = \pm 1$. Por ejemplo en el subespacio $x_4 = 1$ se encuentra la hipercara con los vértices $(\pm 1, \pm 1, \pm 1, 1)$. Estos vértices forman un cubo normal de tres dimensiones. Concluimos que el hipercubo está delimitado por 8 cubos. Como podemos ver, es perfectamente posible estudiar los objetos en cuatro o más dimensiones, aunque no los podamos visualizar.

Ejercicio. Para desarrollar mayor familiaridad con los espacios de dimensiones superiores a tres, sugerimos al lector calcular el número de vértices e hipercaras que tiene el *hipercubo* de 5 dimensiones.

Otra figura prominente en dos dimensiones es la circunferencia. Si su centro es el origen $(0, 0)$ y su radio es 1, entonces la circunferencia se “describe” por la ecuación $x^2 + y^2 = 1$, es decir, consiste de todos los puntos cuyas coordenadas (x, y) satisfacen la ecuación. Otra forma de describir esta circunferencia es como lugar geométrico de los puntos que distan 1 del origen. Lo análogo a la circunferencia en tres dimensiones es la superficie de una esfera, que se describe por la ecuación $x^2 + y^2 + z^2 = 1$.



Por analogía inferimos que el objeto análogo a un círculo o una esfera en cuatro dimensiones se describe con la ecuación

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1.$$

Aunque no podamos visualizar este objeto, si podemos trabajar con él. Por ejemplo podemos encontrar sus puntos de intersección con una recta, un plano o un *hiperplano*, es decir un subespacio de tres dimensiones. Esto quita todo lo misterioso que podría sugerir la expresión “la cuarta dimensión”. Con ello las matemáticas brindan una pequeña contribución a la desmitificación apoyando a la razón y al pensamiento crítico.

La noción de dimensión encuentra dificultades

Hasta 1873 la idea de RIEMANN, de que la dimensión es el número de coordenadas que se necesitan para describir la ubicación de un punto, era ampliamente aceptada. Pero ese año GEORG CANTOR descubre que es posible establecer una correspondencia biunívoca entre los elementos del intervalo $I = [0, 1] = \{x \mid 0 \leq x \leq 1\}$ y los del cuadrado $I \times I = \{(y, z) \mid 0 \leq y, z \leq 1\}$. Sus ideas eran muy novedosas y fueron rechazadas por muchos de sus contemporáneos. Por ello sus resultados tardaron cinco años en publicarse.

Veremos la idea clave de CANTOR. Cada número $x \in I$ se puede escribir en el sistema decimal como

$$x = \sum_{k=1}^{\infty} a_k 10^{-k},$$

con $0 \leq a_k \leq 9$, en otras palabras, a_k es el k -ésimo dígito en el desarrollo decimal de x . Si ahora tenemos un par de coordenadas $(x, y) \in I \times I$, escribimos ambos números en su desarrollo decimal

$$x = \sum_{k=1}^{\infty} a_k 10^{-k}, \quad y = \sum_{k=1}^{\infty} b_k 10^{-k},$$

lo que nos da dos sucesiones a_k y b_k . A partir de ellas construimos otra sucesión c_k alternando los miembros de a_k y b_k , más preciso:

$$c_{2k-1} = a_k, \quad y \quad c_{2k} = b_k$$

para cada número natural k . Esta nueva sucesión define un número

$$z = \sum_{k=1}^{\infty} c_k 10^{-k}.$$

De esta manera obtenemos una función $f: I \times I \rightarrow I$. Por ejemplo si $x = \frac{1}{7}$ y $y = \frac{2}{11}$ entonces tenemos

$$x = 0.142857142857\overline{142857} \quad y = 0.181818181818\overline{18}$$

y con ello

$$z = 0.114821885178114821885178.$$

Ésta es la idea general de CANTOR, pero lamentablemente no se puede formalizar debido a un problema técnico: la representación decimal de algunos números en I no es única. Por ejemplo se tiene

$$0.432 = 0.43199999\overline{9}.$$

Por ello tenemos que ser más cuidadoso. Los únicos números que tienen dos representaciones son aquellos que son diferentes de cero y tienen un desarrollo decimal finito. Cada uno de estos números lo podemos escribir de otra manera con una cola infinita de cifras 9. Para simplificar la manera de hablar diremos que un desarrollo decimal

$$\sum_{k=1}^{\infty} a_k 10^{-k}$$

es *admisibile* si $a_k = 0$ para todo $k \geq 1$ o si existe un número infinito de índices k tal que $a_k \neq 0$. Por lo que dijimos, cada número $x \in I$ tiene una única manera de presentarlo en un desarrollo decimal admisible.

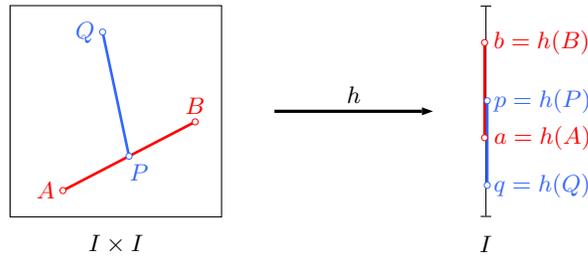
Ahora definimos $f: I \times I \rightarrow I$ como arriba. La única diferencia es que primera nos fijamos de representar ambos números, x y y , por un desarrollo decimal admisible. Entonces $z = f(x, y)$ también está dado por un desarrollo decimal admisible: si el desarrollo decimal de z es no es finita entonces es admisible y si es finito entonces ambos desarrollos de x y de y tienen que terminar en ceros, es decir finitos. Pero esto sólo es posible si $x = 0 = y$ en cuyo caso $z = 0$, por lo que el desarrollo decimal de z es admisible. Esto tiene una implicación fuerte: la función f es inyectiva: si $f(x, y) = f(x', y')$ entonces ambos desarrollos decimales son admisibles y por tanto idénticos, por lo que $x = x'$ y $y = y'$.

Esto va en contra de nuestra intuición: el argumento de CANTOR muestra que existe una función inyectiva que mete todo el cuadrado dentro del intervalo. Lamentablemente no se puede corregir la construcción tal que f sea una biyección y de hecho CANTOR da una construcción mucho más compleja para demostrar que sí existe una biyección. Nosotros daremos aquí satisfechos con la función inyectiva f de arriba.

El resultado de CANTOR muestra que la noción intuitiva de RIEMANN no es nada clara. El trabajo de CANTOR muestra que basta una coordenada x en I para describir la ubicación de un punto en el cuadrado.

El artículo de CANTOR produjo una actividad frenética de entender el concepto de dimensión de mejor manera. Se alegaba que la construcción de CANTOR *no* daba una función *continua*. En efecto, la biyección que construye CANTOR trata los puntos como aislados, es como si primera hace polvo al intervalo y luego reacomoda el polvo para cubrir el cuadrado. Esto muestra que la dimensión debe tomar en cuenta la distancia entre los puntos, o como hoy diríamos la topología. La dimensión no es un concepto que podemos atribuir sólo a conjuntos. Tal vez, se pensaba, era posible salvar la idea Riemann, al demostrar que no era posible definir una función continua y biyectiva entre espacios “de diferentes dimensiones”, en el sentido de Riemann.

En efecto, en el mismo año 1878, EUGEN NETTO logró demostrar que esto no era posible. Su demostración era sorprendentemente sencillo: asume que sí existe una función biyectiva y continua $h: I \times I \rightarrow I$. Luego fija dos puntos distintos A y B en $I \times I$ y considera el segmento AB y su imagen bajo h . Dado que el segmento AB no es todo $I \times I$, la imagen $h(AB)$ tampoco puede ser todo. Por otro lado AB contiene un número infinito de puntos, por lo que $h(AB)$ tienen que ser más que un punto. Finalmente: AB es conexo, por lo que $h(AB)$ también tienen que serlo. No quede de otra para $h(AB)$ que ser el segmento ab , sie $a = h(A)$ y $b = h(B)$.

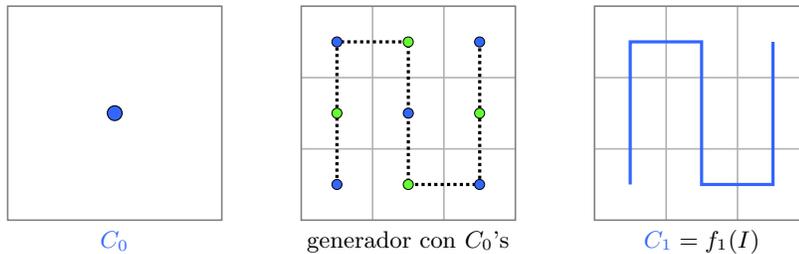


Luego se toma un punto P en el interior de AB y otro Q que no se encuentra sobre la recta que contiene AB y considera el segmento PQ . ¿Qué se sabe de la imagen $h(PQ)$? Por las propiedades de h se tiene que $p = h(P)$ está en el interior de ab y $q = h(Q)$ está fuera de ab . Pero como vimos antes, $h(PQ)$ es el segmento pq por lo que contiene uno de los dos extremos a o b . Esto contradice la biyectividad de h ya que PQ contienen ni A ni B .

Con ello NETTO llega a una contradicción y demuestra que no existe función $h: I \times I \rightarrow I$ que es al mismo tiempo continua y biyectiva.

El descubrimiento de CANTOR motivó a GIUSEPPE PEANO de buscar una función continua $f: I \rightarrow I \times I$ que es aunque no biyectiva al menos fuera suprayectiva. En 1890 tuvo éxito. Esta es su construcción:

Empieza con un cuadrado con su centro marcado. En el siguiente paso subdivide este cuadrado en 9 cuadrados del mismo tamaño y marca ahí los centros. Aunque en este paso todavía no es importante, usamos desde aquí dos colores para los centros: azul y verde. Luego une los centros de la siguiente manera: empieza abajo a la izquierda y sube verticalmente hasta arriba en la columna izquierda, luego baja en la columna en medio y finalmente sube en la columna derecha.

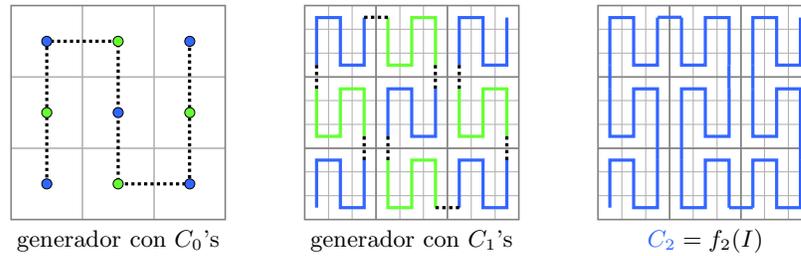


Esto da la función $f_1: I \rightarrow I \times I$. El trazo de la curva tiene 8 segmentos de longitud $\frac{1}{3}$. Por ello se tiene

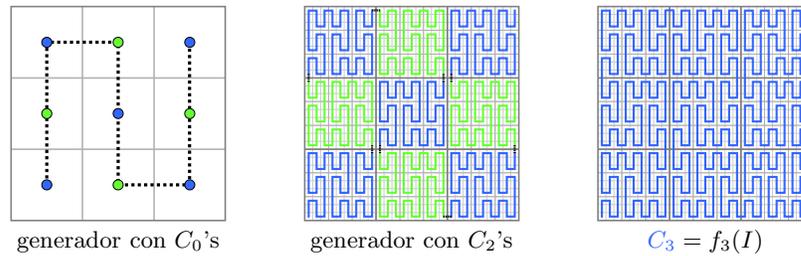
$$f_1(x) = \begin{cases} \left(\frac{1}{6}, \frac{1}{6} + \frac{10}{3}x\right), & \text{para } 0 \leq x \leq \frac{2}{8} \\ \left(\frac{1}{6} + \frac{8}{3}\left(x - \frac{2}{8}\right), \frac{5}{6}\right), & \text{para } \frac{2}{8} \leq x \leq \frac{3}{8} \\ \left(\frac{3}{6}, \frac{5}{6} - \frac{10}{3}\left(x - \frac{3}{8}\right)\right), & \text{para } \frac{3}{8} \leq x \leq \frac{5}{8} \\ \left(\frac{3}{6} + \frac{8}{3}\left(x - \frac{5}{8}\right), \frac{5}{6}\right), & \text{para } \frac{5}{8} \leq x \leq \frac{6}{8} \\ \left(\frac{5}{6}, \frac{1}{6} + \frac{10}{3}\left(x - \frac{6}{8}\right)\right), & \text{para } \frac{6}{8} \leq x \leq 1 \end{cases}$$

Se ve: es posible describir la función explícitamente, pero no es muy instructivo. Es mejor imaginarse que el intervalo I se estira de longitud 1 a la longitud $\frac{8}{3}$ y después se dobla y se acomoda sobre la curva azul. Para el siguiente paso usamos ahora los colores: en cada punto azul y en cada punto verde una copia a escala de la curva C_1 y en cada punto verde una copia de la curva verde. Ésta se obtiene de la azul C_1 al reflejarla por un eje horizontal – o un eje vertical que da los mismo. Luego se unen

los extremos entre azul y verde en el orden que indica el generador. Así se obtiene la curva C_2 y con ello la función f_2 .



Para obtener f_3 se procede de manera similar: en los puntos azules del generador se colocan copias a escala de C_2 y en los puntos verdes se copian los reflejos de C_2 y luego se unen en el orden que indica el generador. Esto define la curva C_3 y con ello la función f_3 .



Así se procede para obtener una familia infinita de funciones $f_k: f_1, f_2, f_3, \dots$. Ahora PEANO define la *función límite* $f: I \rightarrow I \times I$ por

$$f(x) = \lim_{k \rightarrow \infty} f_k(x).$$

Primero debemos reflexionar si tal límite tienen sentido. Dado que la función f_k estira el intervalo I de tal manera que tenga la longitud de la curva C_k , siempre el primer noveno de I , es decir $[0, \frac{1}{9}]$ corresponderá al cuadradito $[0, \frac{1}{9}] \times [0, \frac{1}{9}]$, es decir, el que se encuentra a la izquierda abajo. El siguiente noveno $[\frac{1}{9}, \frac{2}{9}]$ corresponderá al cuadradito encima de este y así consecutivamente. Pero esto se puede iterar. El intervalito $[0, \frac{1}{81}]$ corresponderá a partir de $k > 1$ al minicuatradito $[0, \frac{1}{81}] \times [0, \frac{1}{81}]$, el intervalito $[\frac{3}{81}, \frac{4}{81}]$ a $[\frac{1}{81}, \frac{2}{81}] \times [\frac{2}{81}, \frac{3}{81}]$. Así que podemos indicar que los puntos del intervalo $[\frac{a}{9^m}, \frac{a+1}{9^m}]$ quedarán ubicados bajo cualquier f_k con $k > m+1$ en un cuadradito $[\frac{u}{9^m}, \frac{u+1}{9^m}] \times [\frac{v}{9^m}, \frac{v+1}{9^m}]$. Por un resultado de análisis sobre la convergencia *uniforme*, la función límite f es continua. Por otro lado es también sobre, ya que cualquier punto (y, z) de $I \times I$ existe un x en I tal que la sucesión $f_1(x), f_2(x), \dots$ converge hacia (y, z) .

Con ello PEANO demuestra que es posible parametrizar los puntos de un cuadrado de manera continua por una sola coordenada. Algunos parámetros describirán el mismo punto del cuadrado, por ejemplo el punto $(\frac{1}{9}, \frac{1}{9})$ será aproximado por cuatro direcciones distintas, una vez en cada uno de los cuatro cuadraditos que se unen ahí.

La pregunta sobre la noción de dimensión se hace virulento de nuevo: ¿Se puede decir que la dimensión de un objeto geométrico es el número de coordenadas necesarias para describir la ubicación de un punto en este objeto? Más bien no: la función f de PEANO “recorre” los puntos del cuadrado uno tras otro.

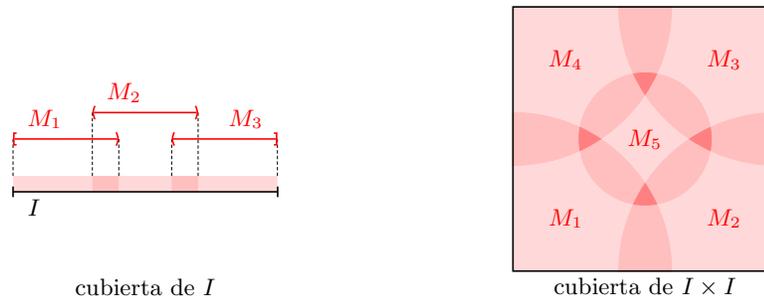
Nuevas definiciones

La situación acerca de la comprensión de qué es la dimensión quedaba insatisfactoria. Por ello hubo varios intentos de dar una mejor definición. Dado que los matemáticos consideraban diferentes situaciones, es decir objetos geométricos con propiedades muy diferentes, sus definiciones resultaban muy diversos. Así que hoy día contamos con diferentes definiciones que en muchos casos, es decir para objetos como la recta o el plano dan el mismo número como dimensión. Sin embargo, dado que las definiciones se enfocan en diferentes propiedades, es posible dar objetos geométricos en los cuales estas dimensiones difieren.

Aquí solo veremos dos de los posibles definiciones de dimensión: la *dimensión de cubiertas de Lebesgue* y la *dimensión de Hausdorff*.

Empezamos con la *dimensión de Lebesgue*, que se debe a los dos matemáticos: HENRI LEBESGUE y EDUARD ČECH. Para describir la dimensión de un objeto geométrico X , se consideran *cubiertas*, esto es una familia de conjuntos M_j (donde j recorre algún conjunto de índices J) de tal manera que cada punto $x \in X$ pertenece a al menos un conjunto M_j . Los conjuntos deben ser *abiertos* como un segmento sin sus puntos extremos o el interior de un disco.

Veamos dos ejemplos de cubiertas, una del intervalo I y otra del cuadrado $I \times I$.



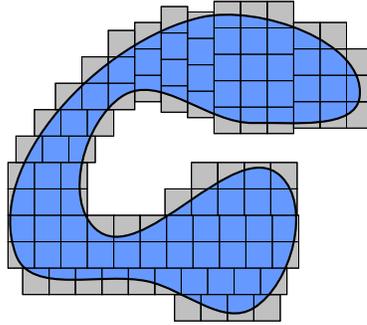
cubierta de I

cubierta de $I \times I$

Lo que se puede ver: para cubrir el segmento con segmentos abiertos es necesario que haya puntos $x \in I$ que se encuentran en *dos* de los conjuntos M_j , mientras que en la cubierta del cuadrado encontraremos puntos $(y, z) \in I \times I$ que se encuentran en *tres* de los conjuntos que cubren.

Para cada punto x del objeto geométrico X podemos contar el número de conjuntos cubrientes M_j que contengan x . La dimensión topológica es algo como el máximo de estos números menos uno. Esta no es una definición correcta, ya que se busca una definición que sea independiente de la cubierta, pero indica la esencia de la dimensión de Lebesgue. Dejamos estos detalles para ver la segunda posibilidad de definir “dimensión”.

Para ello veamos primero una idea de cómo se podría aproximar el contenido de un área A de una figura curva, como la azul en la siguiente figura.



Podríamos contar el número N de cuadrados de tamaño $\ell \times \ell$ que necesitamos para cubrir la figura por completo. Una aproximación para el área es entonces

$$A_2(\ell) = N \times \ell^2$$

Como N depende del tamaño ℓ escribimos $N = N(\ell)$. Ahora, si hacemos más pequeño a ℓ el número $N(\ell)$ aumentará, pero tendríamos menor área que sobra en los bordes, es decir nuestro error disminuirá. Así que podemos concluir que obtenemos A como límite

$$A_2 = \lim_{\ell \rightarrow 0} A_2(\ell)$$

Podemos usar este método para medir el área de un segmento. Ha leído bien: el *área* de un segmento.



Si L es la longitud del segmento, entonces necesitaríamos aproximadamente $N(\ell) = \frac{L}{\ell}$ cuadrados. Por ello obtenemos $A_2(\ell) = N(\ell) \cdot \ell^2 = L\ell$ y

$$A_2 = \lim_{\ell \rightarrow 0} A_2(\ell) = \lim_{\ell \rightarrow 0} N\ell = 0,$$

es decir el área es cero – como lo esperábamos. Podemos medir la longitud del segmento si en vez de calcular $A_2(\ell) = N(\ell) \cdot \ell^2$ calculamos $A_1(\ell) = N(\ell) \cdot \ell = L$ y tomamos ahí el límite. Obtendremos

$$A_1 = \lim_{\ell \rightarrow 0} A_1(\ell) = L,$$

que también es lo que esperábamos. El cubrimiento con cuadrados sirve entonces también para medir la longitud. Lo único que hay que hacer es tomar el exponente 1 de ℓ . Podemos poner a prueba esta definición y medir la longitud de un cuadrado de tamaño $L \times L$. Si el tamaño de los cuadraditos es ℓ , necesitaríamos aproximadamente $N(\ell) = \frac{L^2}{\ell^2}$ cuadraditos para cubrir el cuadrado grande. Por ello su longitud es

$$A_1 = \lim_{\ell \rightarrow 0} N(\ell) \cdot \ell = \lim_{\ell \rightarrow 0} \frac{L^2}{\ell} = \infty$$

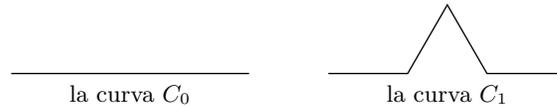
como los espábamos.

La dimensión de Hausdorff, llamado así por el matemático FELIX HAUSDORFF, se define como el número d tal que el límite

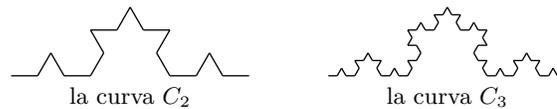
$$A_d = \lim_{\ell \rightarrow 0} N(\ell) \cdot \ell^d$$

sea finita pero no cero. Con esta definición, el segmento tiene dimensión 1 y el cuadrado tiene dimensión 2.

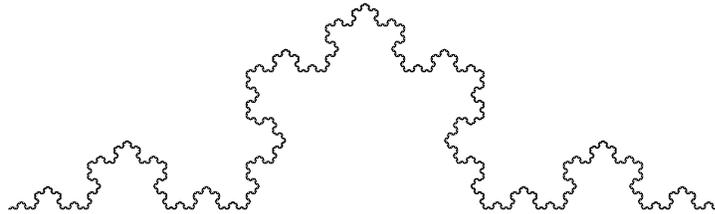
Veamos ahora una curva más rara. Se llama *curva de Koch* en honor al matemático sueco HELGE VON KOCH y se define parecido a la curva de PEANO como límite. La curva C_0 es un segmento de longitud 1, mientras C_1 se obtiene al dividir C_0 en tres partes iguales, sustituir el tercio de medio por los otros dos lados de un triángulo equilátero.



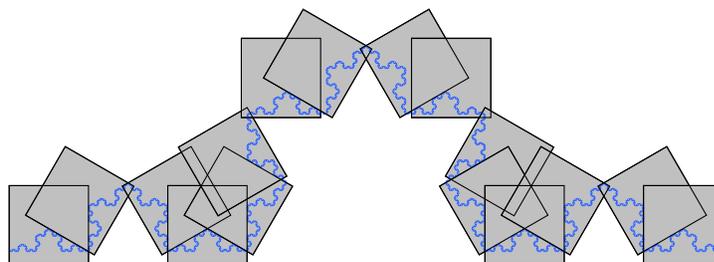
En cada paso, cada segmento se sustituye por una copia semejante de C_1 .



La curva de Koch C es límite que se obtiene por este proceso.



Esta curva tiene propiedades interesantes, como por ejemplo la *autosemejanza*, es decir, la propiedad que una parte de ella es semejante a toda la figura. Trataremos de determinar la dimensión de Hausdorff de este objeto. Por ello lo cubrimos primero con $N(\frac{1}{3}) = 4$ cuadrados de longitud $\frac{1}{3}$. Si pasamos a cuadraditos de longitud $(\frac{1}{3})^2$ entonces necesitamos $N((\frac{1}{3})^2) = 4^2$ de ellos.



Con 4^2 cuadraditos de longitud $(\frac{1}{3})^2$ se cubre la curva de Koch.

Más generalmente, se requieren 4^k cuadraditos de longitud $\ell = (\frac{1}{3})^k$ para cubrir la curva de Koch. Para que ℓ disminuya a cero se requiere que k crezca al infinita. La longitud de C es entonces

$$A_1 = \lim_{k \rightarrow \infty} 4^k \left(\frac{1}{3}\right)^k = \lim_{k \rightarrow \infty} \left(\frac{4}{3}\right)^k = \infty.$$

El área de C es

$$A_2 = \lim_{k \rightarrow \infty} 4^k \left(\frac{1}{3}\right)^2 k = \lim_{k \rightarrow \infty} \left(\frac{4}{9}\right)^k = 0.$$

La curva de Koch tiene área 0 y longitud infinita. Su dimensión no es ni 1 ni 2. Para que el límite

$$\lim_{k \rightarrow \infty} 4^k \left(\frac{1}{3}\right)^k d = \lim_{k \rightarrow \infty} \left(\frac{4}{3^d}\right)^k$$

sea finito, pero no cero se requiere

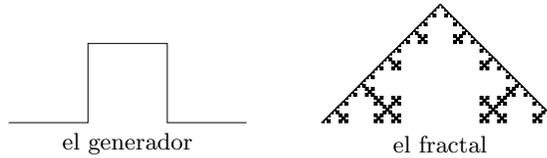
$$\frac{4}{3^d} = 1,$$

por lo que $3^d = 4$ y

$$d = \log_3(4) \approx 1.262.$$

Hemos encontrado la dimensión de curva de Koch: es 1.262, un número entre 1 y 2. Como la dimensión no es entera, se llama a veces también *dimensión fractal*.

Ejercicio. Para entender mejor la dimensión fractal se recomienda al lector calcular la dimensión de Hausdorff de la curva que se obtiene si se itera sustituir cada segmento por una copia semejante del trazo que se muestra en la siguiente figura.



Capítulo 8

Cálculo, funciones y límites

Herramientas para modelar el cambio continuo

8.1. El Cálculo Diferencial e Integral

¿Qué es el cálculo?

El cálculo infinitesimal, cálculo de infinitesimales o simplemente “Cálculo” constituye una parte muy importante de las matemáticas modernas. Incluye el estudio de las funciones, los límites, las derivadas, las integrales, las series infinitas y las ecuaciones diferenciales. El cálculo se puede definir como el estudio del cambio en el tiempo, de la misma manera que la geometría es el estudio del espacio.

El cálculo infinitesimal tiene amplias aplicaciones en la ciencia y la ingeniería y se usa para resolver problemas para los cuales el álgebra por sí sola es insuficiente. Este cálculo se construye con base en el álgebra, la trigonometría y la geometría analítica y como ya se mencionó, incluye dos campos principales, cálculo diferencial y cálculo integral, que están relacionados por el teorema fundamental del cálculo. En matemáticas más avanzada, el cálculo se suele llamar análisis matemático y puede definirse como el estudio de las funciones.

La historia del cálculo

ISAAC NEWTON y GOTTFRIED LEIBNIZ son los nombres más famosos vinculados al desarrollo del cálculo. La aplicación más importante en la historia del cálculo es la formulación y solución de las leyes del movimiento y de la gravitación universal, y se deben a Newton.

Durante toda la historia, desde la antigüedad remota hasta el siglo XVII los matemáticos trataron de descubrir maneras de calcular el área de las figuras planas delimitadas por curvas, En el siglo IV A.C. Eudoxo inventó el método de exhaución que consistía en ir llenando la figura cuya área se desea calcular, con polígonos simples como rectángulos y triángulos, cuyas áreas son fáciles de calcular, acomodando figuras cada vez más pequeñas para ir cubriendo los huecos que quedaban. La teoría

es que si el procedimiento se continúa indefinidamente, se obtiene el área deseada. Pero los procesos infinitos no son fáciles de manejar. Arquímedes logró utilizarlos de una manera increíblemente eficaz y rigurosa para obtener el área acotada por una parábola y una recta, el área de un círculo y los volúmenes y superficies de cilindros, conos y esferas. Sobre todo su trabajo para calcular la superficie de una esfera es una joya en elegancia y rigor matemático. Pero el método de exhaustión no dio resultados más allá de los obtenidos por Arquímedes, hasta que el desarrollo de la Geometría Analítica renovó el interés en él. La posibilidad de describir a las curvas con ecuaciones permitía aplicar el método de exhaustión de una forma simplificada, analizando los problemas de cálculo de áreas y volúmenes utilizando ecuaciones algebraicas. Este método se desarrolló poco a poco hasta formar una nueva disciplina llamada Cálculo Integral.

Al mismo tiempo, las nociones de velocidad y aceleración que gracias a las leyes de Newton permitían analizar matemáticamente el movimiento de los cuerpos, llevaron a la creación del concepto de derivada, ligado al de velocidad instantánea o razón de cambio instantáneo, lo que da lugar a otra nueva rama de las matemáticas: el Cálculo Diferencial, que permite obtener la velocidad y la aceleración de un cuerpo conociendo su posición en cada instante.

El cálculo antes del siglo XVII

En la antigüedad griega se introdujeron algunas de las ideas del cálculo integral, pero no se desarrollaron de una manera sistemática, aunque sí de manera bastante rigurosa. En el cálculo de áreas y volúmenes, la función básica del cálculo integral puede ser rastreada en el tiempo hasta los papiros matemáticos de Moscú que datan del año 1890 a. C, en los que un egipcio calculó satisfactoriamente el volumen del tronco de una pirámide, sin embargo esto se hizo sin dar argumentos racionales para justificar el método de cálculo.

De la escuela de los matemáticos griegos, Eudoxo usó el método exhaustivo, el cual prefiguraba el concepto de límite, para calcular áreas y volúmenes, mientras que Arquímedes (287 - 212 aec) desarrolló más allá su idea inventando un método heurístico, denominado exhaustión, que se asemeja al cálculo ya bastante al cálculo integral, con lo cual logró obtener el volumen y la superficie de la esfera. El método de exhaustión fue usado en el siglo III aec y de manera independiente en China por LIU HUI para encontrar el área de un círculo, obteniendo una mejor aproximación que la de Arquímedes, gracias a la superior aritmética de los chinos.

En China, el siglo V, ZU CHONGZHI usó lo que más tarde sería llamado la teoría de los indivisibles por el matemático italiano Bonaventura Cavalieri para encontrar el volumen de una esfera.

Cerca del año 1000, el matemático islámico ALHACÉN fue el primero en derivar la fórmula para la suma de la cuarta potencia de una progresión aritmética, usando un método a partir del cual es fácil encontrar la fórmula para la suma de cualquier potencia integral de mayor orden. Esto sentó las bases para obtener más adelante el área bajo las gráficas de las funciones polinomiales.

En el siglo XI, el polímata chino SHEN KUO desarrolló ecuaciones que se encargaban de integrar. En el siglo XII, el matemático indio, BHASKARA II, desarrolló una derivada temprana representando el cambio infinitesimal, y describió una forma temprana del llamado teorema de Rolle.

También en el siglo XII, el matemático persa SHARAF AL-DIN AL-TUSI descubrió la derivada de la función cúbica, un importante acontecimiento que abre el camino hacia el cálculo diferencial.

En el siglo XIV, MADHAVA DE SANGAMAGRAMA, en conjunto con otros matemáticos y astrónomos

de la escuela Kerala de matemáticas y astronomía, al sur de la India, describieron casos especiales de lo que hoy se conocen como las series de Taylor.

Las rutas comerciales llevaron a Europa la notación indo-arábiga y los algoritmos para realizar operaciones con ella, desplazando rápidamente el incómodo sistema de números romanos que se había usado durante casi toda la edad media. Es importante reconocer la contribución de LEONARDO DE PISA (FIBONACCI) (1170 - 1240) en este proceso.

En la época moderna, descubrimientos independientes relacionados con el cálculo se estaban llevando a cabo por la matemática japonesa del siglo XVII, gracias al aporte de matemáticos como Seki Kōwa, quien expandió el método exhaustivo.

En Europa, el trabajo fundacional fue un tratado del matemático italiano Bonaventura Cavalieri, quien argumentó que los volúmenes y áreas deberían ser calculados como las sumas de los volúmenes y áreas de delgadas secciones infinitesimales. Estas ideas eran similares a las expuestas en el trabajo El método de los teoremas mecánicos de Arquímedes, el cual estuvo perdido hasta principios del siglo XX. El trabajo de Cavalieri no fue bien recibido por la comunidad científica ya que se observó que a veces sus métodos pueden llevar a resultados erróneos, y porque las cantidades infinitesimales que introdujo no parecían estar definidas con suficiente claridad.

El cálculo en el siglo XVII

El estudio formal del cálculo combinó los infinitesimales de Cavalieri con el cálculo de diferencias finitas desarrollado en Europa más o menos al mismo tiempo. La combinación fue lograda por John Wallis, Isaac Barrow y James Gregory, probando estos últimos el teorema fundamental del cálculo integral cerca del año 1675.

La regla del producto y la regla de la cadena, la noción de derivada de mayor orden, las series de Taylor, y las funciones analíticas fueron introducidas por Isaac Newton en una notación idiosincrásica que usó para resolver problemas de física, principalmente de mecánica. En sus publicaciones, Newton reformuló sus ideas para acomodar el idioma matemático de la época, reemplazando cálculo con infinitesimales por argumentos geométricos equivalentes, los cuales estaban más allá de reproches. Usó los métodos del cálculo para resolver el problema del movimiento planetario, la forma de la superficie de un fluido rotante, y se refirió a lo achatada que es la tierra por los polos, así como a muchos otros problemas, los cuales discutió en su Principia mathematica. En otro trabajo, desarrolló una serie de expansiones para las funciones, incluyendo las potencias fraccionarias e irracionales. Estaba claro que Newton entendía perfectamente los principios de las series de Taylor. Sin embargo no publicó muchos de sus descubrimientos. En su tiempo los sistemas infinitesimales tenían mala reputación.

Las ideas del Cálculo fueron descubiertas independientemente por Gottfried Wilhelm Leibniz, quien además las sistematizó de manera magistral, arrebatando a Newton gran parte del mérito por el descubrimiento del cálculo, gracias a su magnífica organización del tema a su excelente y sugestiva notación. Leibniz fue acusado de plagio por Newton, pero se le reconoce sin ninguna duda como inventor independiente del cálculo y como gran contribuyente a su formalización y organización. Su principal contribución fue el proveer un conjunto de reglas claras para la manipulación de cantidades infinitesimales, permitiendo el cómputo de derivadas de segundo orden y de orden superior, y estableciendo la regla del producto y regla de la cadena en su forma diferencial e integral. A diferencia de Newton, Leibniz le puso mucha atención al formalismo y a menudo le dedicaba varios

días a determinar los símbolos apropiados para cada concepto.

La controversia entre Newton y Leibniz sobre la invención del cálculo

Usualmente se acredita tanto a Leibniz como a Newton con la invención del cálculo. Newton fue el primero en aplicar el cálculo a la física general y Leibniz desarrolló la notación usada en cálculo hasta nuestros días. Las ideas principales que ambos Newton y Leibniz estipularon fueron las leyes de diferenciación e integración, las segundas derivadas, las derivadas de orden superior, y la noción de una aproximación de series de potencias, conocida actualmente como las series de Taylor.

Cuando Newton y Leibniz primero publicaron sus resultados, hubo gran controversia sobre qué matemático (y por ende qué país) merecía el crédito por la invención de esta disciplina. Newton llegó primero a sus resultados, pero Leibniz publicó primero. Newton acusó a Leibniz de robar sus ideas de sus notas inéditas, las cuales Newton había compartido con unos cuantos miembros de la Royal Society. Esta controversia dividió a los matemáticos de habla inglesa de los matemáticos continentales por varios años, causando un retraso de las matemáticas inglesas. Un cuidadoso examen de los papeles de ambos matemáticos demuestra que ellos llegaron a sus resultados independientemente, con Leibniz empezando primero con la integración y Newton con la diferenciación. Hoy, se les da crédito a ambos matemáticos por desarrollar el cálculo independientemente. Fue Leibniz, sin embargo, quien le dio el nuevo nombre a su disciplina. Newton llamó su cálculo el *método de las fluxiones*. La simbología usada por Newton, tal como \dot{x} (derivada primera), \ddot{x} (derivada segunda), no fue muy afortunada y retrasó el avance del cálculo en Gran Bretaña. En cambio, la simbología de Leibniz fue manejable y apuntaló el progreso del cálculo en Europa.

Desde los tiempos de Leibniz y Newton, muchos matemáticos han contribuido al desarrollo del cálculo. En el siglo XIX, el cálculo comenzó a ser planteado más rigurosamente por matemáticos como Cauchy, Riemann y Weierstrass. También fue en este período que las ideas del cálculo fueron generalizadas al espacio euclidiano y al plano complejo. Lebesgue generalizó la noción de la integral de tal manera que virtualmente cualquier función tenga una integral, mientras que Laurent Schwartz extendió la diferenciación casi de la misma manera.

La importancia del cálculo

En la actualidad el cálculo es un tema omnipresente en la mayoría de los programas de educación superior y en las universidades. Los matemáticos alrededor del mundo continúan contribuyendo al desarrollo de esta disciplina, la cual ha sido considerada como uno de los logros más grandes del intelecto humano. El desarrollo de las ecuaciones diferenciales, inconcebibles sin el cálculo, ha jugado un gran papel de cambio cualitativo en la ciencia y la tecnología, comparable con el control del fuego en la época primitiva, las ecuaciones diferenciales son un salto enorme para la ciencia.

Mientras que algunas ideas del cálculo fueron desarrolladas tempranamente en las matemáticas griegas, chinas, indias, islámicas y japonesas, el uso moderno del cálculo comenzó en Europa, durante el siglo XVII, cuando Isaac Newton y Gottfried Leibniz construyeron con base al trabajo de antiguos matemáticos los principios básicos de esta disciplina. El desarrollo del cálculo fue constituido con base en los conceptos de movimiento instantáneo y el área bajo las curvas.

Las aplicaciones del cálculo diferencial incluyen cálculos que involucran velocidad, aceleración, la pendiente de una recta tangente a una curva y optimización. Las aplicaciones del cálculo integral

están en cálculos que incluyen elementos de área, volumen, centro de masa, longitud de arco, trabajo y presión. Aplicaciones más avanzadas incluyen series de potencias y series de Fourier. El cálculo puede ser usado para computar la trayectoria de una nave acoplándose a una estación espacial o la cantidad de nieve en una calzada para coches.

El cálculo es también usado para obtener un entendimiento más preciso de la naturaleza del espacio, el tiempo y del movimiento. Por siglos, matemáticos y filósofos lucharon con paradojas que involucraban la división por cero o sumas de series infinitas de números. Estas preguntas surgen en el estudio del movimiento y área. El antiguo filósofo griego Zenón dio varios ejemplos famosos de tales paradojas. El cálculo provee herramientas que pueden resolver tales paradojas, especialmente los límites y las series infinitas.

Límites vs infinitesimales. Los fundamentos del cálculo.

En matemáticas, los fundamentos se refieren al desarrollo riguroso de un tema a partir de axiomas y definiciones precisas. El obtener un fundamento riguroso para el cálculo ocupó a los matemáticos durante la mayor parte del siglo que siguió a Leibniz y Newton y es aún un área activa en el siglo XXI. Los fundamentos del cálculo fueron objeto de diversas especulaciones filosóficas e interpretaciones informales, la falta de rigor y laxitud con que fueron afrontados ciertos problemas de fundamentación contribuyeron a la crisis de los fundamentos de las matemáticas.

Sin embargo, ya durante el siglo XIX se empezó a trabajar en una aproximación rigurosa para los fundamentos del cálculo. El más usual hoy en día es el concepto de límite definido en la continuidad de los números reales (el concepto de límite es esencialmente un concepto topológico). Una alternativa es el análisis no estándar, en el cual el sistema de números reales aumentado con infinitesimales y números infinitos, como en la concepción original de Newton y Leibniz. Los fundamentos del cálculo son incluidos en el campo del análisis real, el cual contiene las definiciones completas y pruebas matemáticas de los teoremas del cálculo, así como también generalizaciones tales como la teoría de la medida y la teoría de distribuciones.

El cálculo se desarrolló mediante la manipulación de “cantidades pequeñas”. Históricamente, el primer método para lograr eso se basaba en los infinitesimales. Estos son objetos que pueden ser tratados como números pero que son, en algún sentido, “infinitamente pequeños”. Tratándose de números, éstos serían puntos que no son cero, pero que tienen una distancia cero del número ‘cero’. Desde este punto de vista, el cálculo es una colección de técnicas para manipular infinitesimales. Este punto de vista perdió terreno en el siglo XIX porque era difícil lograr una noción precisa del infinitesimal. El concepto cobró fuerza nuevamente en el siglo XX con la introducción del análisis no estándar y del “análisis infinitesimal suave” (del inglés *smooth infinitesimal analysis*), los que proporcionaron fundamentos sólidos para la manipulación de infinitesimales.

En el siglo XIX, los infinitesimales fueron reemplazados por los límites. Los límites describen el valor de una función en un cierto valor de entrada en términos de sus valores en un punto cercano. Capturan el comportamiento a pequeña escala, como los infinitesimales, pero usan el sistema ordinario de los números reales. En este contexto, el cálculo es una colección de técnicas usadas para la manipulación de ciertos límites. Los infinitesimales son reemplazados por números muy pequeños y el comportamiento infinitamente pequeño de la función es encontrado mediante el comportamiento límite para números cada vez más pequeños. Los límites son fáciles de poner en fundamentos, y por esta razón son usualmente considerados como el acercamiento estándar al cálculo.

Ambos cálculos, el diferencial y el integral fueron el resultado de muchas ideas acumuladas durante siglos, de pensadores como Eudoxo, Arquímedes, Nicole Oresme, Galilei, Kepler, Descartes, Fermat, Cavallieri, Torricelli, e Isaac Barrow, maestro de Newton. Apoyados en las ideas acumuladas durante tantos siglos, Isaac Newton y Wilhelm Leibniz, casi simultáneamente y de manera independiente uno del otro, dieron origen al Cálculo Diferencial e Integral. A Newton se deben sobre todo las aplicaciones de esta rama de las matemáticas a la física, sobre todo al movimiento de los cuerpos. A Leibniz se deben sobre todo la notación que se usa en el cálculo para denotar derivadas e integrales, que no es un logro menor. Pero a ambos se debe la unificación de ambas ramas, el Cálculo Diferencial e Integral con la aplicación sistemática del llamado Teorema Fundamental del Cálculo para generar una tecnología que permite resolver problemas de integración utilizando el proceso inverso a la derivación.

8.2. El cálculo

El cálculo diferencial

Línea tangente en $(x, f(x))$. La derivada $f'(x)$ de una curva en un punto es la pendiente de la línea tangente a la curva en ese punto. El cálculo diferencial es el estudio de la definición, propiedades, y aplicaciones de la derivada de una función, o lo que es lo mismo, la pendiente de la tangente a lo largo de su gráfica. El proceso de encontrar la derivada se llama derivación o diferenciación. Dada una función y un punto en su dominio, la derivada en ese punto es una forma de codificar el comportamiento a pequeña escala de la función cerca del punto. Encontrando la derivada de una función para cada punto en su dominio, es posible producir una nueva función, llamada la *función derivada* o simplemente *la derivada* de la función original. En lenguaje técnico, la derivada es un operador lineal, el cual toma una función y devuelve una segunda función, de manera que para cada punto de la primera, la segunda proporciona la pendiente de la recta tangente a la gráfica en ese punto.

Para entender la derivada es importante familiarizarse con su notación. Un símbolo común para la derivada de una función es una marca parecida a un acento o apóstrofo llamada símbolo primo. Así la derivada de f es f' (pronunciado “efe prima”). En lo siguiente la segunda función es la derivada de la primera:

$$\text{si } f(x) = mx + b, \text{ entonces } f'(x) = m$$

$$\text{si } f(x) = x^2, \text{ entonces } f'(x) = 2x$$

$$\text{si } f(x) = \text{sen}(x), \text{ entonces } f'(x) = \text{cos}(x)$$

$$\text{si } f(x) = \log(x), \text{ entonces } f'(x) = \frac{1}{x}$$

$$\text{si } f(x) = e^x, \text{ entonces } f'(x) = e^x$$

Si la variable de la función representa tiempo, entonces la derivada representa el cambio de la función con respecto del tiempo. Por ejemplo, si f es una función que tiene al tiempo como variable y sus valores dan la posición de una pelota en cada momento, entonces la derivada de f determina cuánto está cambiando en el tiempo la posición de la pelota, esto es, la velocidad de la pelota.

El cálculo integral y su teorema fundamental

El cálculo integral es el estudio de las definiciones, propiedades, y aplicaciones de dos conceptos relacionados, la integral indefinida y la integral definida. El proceso de encontrar el valor de una integral es llamado integración. En lenguaje técnico, el cálculo integral estudia dos operadores lineales relacionados.

La integral indefinida es la antiderivada, es decir, la operación inversa de la derivada. La función F es una integral indefinida de la función f cuando f es una derivada de F . (El uso de mayúsculas y minúsculas para distinguir entre la función y su integral indefinida es común en el cálculo).

La integral definida transforma una función definida en un intervalo $[a, b]$ en un número que se denota por

$$\int_a^b f(x)dx$$

y que es el valor del área entre una curva definida por $y = f(x)$ y el eje- x , en el intervalo $[a, b]$. La definición técnica de la integral se realiza como el límite de sumas de áreas de rectángulos, que se llaman sumas de Riemann:

$$\int_a^b f(x)dx = \lim_{\|P\| \rightarrow 0} \sum_{n=1}^N f(x_n) \Delta x_n$$

donde P representa una partición del intervalo $[a, b]$ en subintervalos de longitudes $\{\Delta x_n\}_{n=1}^{n=N}$, cada x_n es un punto del intervalo de longitud Δx_n y $\|P\|$ es la *norma de la partición* P que se define como el máximo de los valores Δx_n .

El teorema fundamental del cálculo establece que la diferenciación y la integración son operaciones inversas. Más precisamente, relaciona los valores de las antiderivadas para definir las integrales. Ya que es normalmente más fácil computar una antiderivada que aplicar la definición de una integral definida, el teorema fundamental del cálculo provee una forma práctica de computar integrales definidas. También puede ser interpretado como una declaración precisa del hecho de que la diferenciación es la inversa de la integración.

El teorema fundamental del cálculo establece: Si una función f es continua en el intervalo $[a, b]$ y si F es una función cuya derivada en el intervalo (a, b) es f , entonces

$$\int_a^b f(x) dx = F(b) - F(a)$$

Así entonces, para cada x en el intervalo (a, b) , es cierto que:

$$\frac{d}{dx} \int_a^x f(t) dt = f(x)$$

Este hecho, descubierto tanto por Newton como Leibniz, quienes basaron sus resultados en el trabajo previo de Isaac Barrow, fue clave para la masiva proliferación de resultados analíticos luego que su trabajo fuese conocido. El teorema fundamental provee un método algebraico para calcular

muchas integrales definidas – sin necesidad de calcular límites – mediante el encuentro de fórmulas apropiadas para las antiderivadas.

Otra versión de Teorema nos dice que si a partir de una función $f(x)$ creamos otra $F(x)$ definiéndola como el área bajo la gráfica de $f(x)$ entre un punto fijo a y una variable x , entonces $F(x)$ se puede derivar, es decir, se pueden calcular las pendientes de su gráfica y están dadas por la función original $f(x)$. En la notación de LEIBNIZ,

$$\text{si } F(x) = \int_a^x f(x)dx, \text{ entonces } \frac{dF}{dx} = f(x)$$

Como ya dijimos, este resultado permite obtener las áreas bajo muchas curvas mediante el procedimiento de representarlas como funciones que son las derivadas de otras, de manera que para calcular el área bajo la gráfica de ellas basta tomar la diferencia de las segundas entre los dos extremos del intervalo de integración: Si $f(x) = \frac{dF}{dx}$, entonces la integral de a a b de $f(x)$ es igual a $F(b) - F(a)$. Y esto crea un método muy general para obtener las áreas bajo curvas, para lo cual basta tener una larga lista de funciones y sus derivadas, es decir una tabla de derivadas o, vista en orden inverso, una tabla de integrales. Cabe aclarar que el proceso de obtener las derivadas de muchas funciones es relativamente simple, gracias a unos cuantos resultados básicos que se pueden aplicar recursivamente para cubrir muchos casos de interés.

El aprendizaje del cálculo

Conocer y dominar el cálculo consiste en lo siguiente:

- 1) Descubrir cómo obtener el área bajo la gráfica de una función escalonada (es decir que es constante en intervalos consecutivos) sumando términos de la forma $f(\xi_i)(x_{i+1} - x_i)$ donde $f(\xi_i)$ es el valor constante de f entre x_i y x_{i+1} .
- 2) Utilizando una función $f(x)$ que es la derivada de otra función $F(x)$ descubrir el Teorema Fundamental del Cálculo que le permite obtener el área bajo la gráfica de $f(x)$ entre $x = a$ y $x = b$ como

$$\int_a^b f(x)dx = F(b) - F(a)$$

. Para ello aproximar la función $f(x)$ en el intervalo $[a, b]$ por una función escalonada con valores

$$f(\xi_i) = \frac{F(x_{i+1}) - F(x_i)}{x_{i+1} - x_i}$$

en los intervalos de la partición $a = x_0 < x_1 < x_2 < \dots < x_n < x_{n+1} = b$. El área bajo la gráfica de esta función es

$$\sum_{i=0}^n f(\xi_i)(x_{i+1} - x_i) = \sum_{i=0}^n F(x_{i+1}) - F(x_i) = F(b) - F(a)$$

Tomando particiones cada vez más finas se puede concluir que

$$\int_a^b f(x)dx = F(b) - F(a)$$

Esto demuestra el Teorema Fundamental del Cálculo que permite calcular fácilmente el área bajo la gráfica de una función $f(x)$ si se conoce su antiderivada $F(x)$, es decir una función tal que

$$F'(x) = \frac{dF}{dx} = f(x)$$

3) Conocer el concepto de integral indefinida o antiderivada y saber que dos antiderivadas de una función sólo difieren en una constante, por lo cual en general éstas se denotan como:

$$\int f(x)dx = F(x) + c$$

donde c es una constante arbitraria.

4) Entender el significado histórico de este descubrimiento. El Cálculo se convirtió, a partir del siglo XVIII, en una herramienta matemática esencial que permitía resolver problemas como el cálculo de áreas de figuras curvas, volúmenes de todo tipo de cuerpos, longitudes de todo tipo de curvas, trayectorias de partículas sujetas a fuerzas conocidas y otras cantidades de interés científico como el trabajo realizado por una fuerza, o la media y la desviación estándar de una distribución de probabilidad, que anteriormente eran intratables o ni siquiera podían definirse con precisión.

Gracias al *Teorema Fundamental del Cálculo* que relaciona la integral con la derivada, el Cálculo Diferencial e Integral se convirtió en una poderosa herramienta que ha sido por siglos el lenguaje común de las Ciencias y además abrió una nueva era de las matemáticas participando íntimamente en el desarrollo de varias nuevas áreas como la teoría de Ecuaciones Diferenciales, el Análisis Real y Complejo, el Análisis Funcional y la Geometría Diferencial. Los detalles de todos los desarrollos descritos brevemente en este capítulo, así como algunos de los desarrollos posteriores al Cálculo podrán encontrarse en el resto de los capítulos del libro.

Las ecuaciones diferenciales y las aplicaciones del cálculo

Las ecuaciones diferenciales relacionan a una función con sus derivadas, y son omnipresentes en las ciencias pues se utilizan para plantear modelos cuantitativos del comportamiento de diversos sistemas.

La Espiral logarítmica de la concha del Nautilus es una clásica imagen usada para representar el crecimiento y cambio relacionados con el cálculo. El cálculo es usado en cada rama de las ciencias físicas y de informática, estadística, ingeniería, economía, negocios, medicina, demografía y en otras áreas donde un problema pueda ser modelado matemáticamente y una solución óptima sea deseada.

La física hace un particular uso del cálculo; todos los conceptos en la mecánica clásica están relacionados a través del cálculo. La masa de un objeto, el momento de inercia de los cuerpos, la energía total de un objeto dentro de un campo conservativo, todo esto se define usando el cálculo. Un ejemplo históricamente más impactante del uso del cálculo en la física son las leyes del movimiento de Newton, donde se usa expresamente la tasa de cambio de la posición para definir la velocidad y la tasa de cambio de la velocidad para definir la aceleración y finalmente para enunciar la segunda ley del movimiento de Newton: la tasa de cambio de momentum (el momentum es el producto e la masa por la velocidad) de un cuerpo es igual a la fuerza que actúa sobre el cuerpo. Incluso la expresión común de la segunda ley de Newton como fuerza = masa \times aceleración involucra el cálculo

diferencial porque la aceleración debe entenderse como la razón de cambio, o la derivada respecto al tiempo, de la velocidad.

En la electricidad y el magnetismo, el cálculo se usa para describir los campos electromagnéticos. Gracias a tal descripción Maxwell descubrió las ondas electromagnéticas ya que las ecuaciones del campo electromagnético pueden manipularse para obtener una ecuación de onda. También la Teoría de la relatividad general de Einstein y la mecánica cuántica se expresan en el lenguaje del cálculo y las ecuaciones diferenciales.

El cálculo también puede ser usado en conjunto con otras disciplinas matemáticas. Por ejemplo, puede ser usado con el álgebra lineal para encontrar la mejor aproximación lineal para un conjunto de puntos en un dominio. También puede ser usado en la teoría de la probabilidad para determinar la probabilidad de una variable continua al azar desde una densidad de función asumida.

El teorema de Green, el cual establece la relación entre una integral lineal alrededor una simple curva cerrada C y una doble integral sobre el plano de región D delimitada por C , es aplicado en un instrumento conocido como planímetro, el cual es usado para calcular el área de una superficie plana en un dibujo. Por ejemplo, puede ser usado para calcular la cantidad de área que toma una piscina cuando se bosqueja el diseño de un pedazo de propiedad.

En geometría analítica, el estudio de los gráficos de funciones, el cálculo es usado para encontrar puntos máximos y mínimos, la tangente, así también como para determinar la concavidad y los puntos de inflexión.

La química también usa el cálculo para determinar los ritmos de las reacciones y el decaimiento radioactivo.

En la medicina, el cálculo puede ser usado para encontrar el ángulo de ramificación óptimo de vaso sanguíneo para maximizar el flujo.

En economía, el cálculo permite determinar el beneficio máximo por medio del costo marginal y del ingreso marginal.

El cálculo también puede ser usado para encontrar soluciones aproximadas para ecuaciones, usando métodos como por ejemplo el método de Newton, la iteración de punto fijo y la aproximación lineal. Por ejemplo, las naves espaciales usan una variación del método de Euler para aproximar trayectorias curvas dentro de entornos de gravedad cero.

8.3. Las funciones en el cálculo

Todo el lenguaje del cálculo se basa en el concepto de función. Concretamente, se basa en las llamadas funciones reales de variable real, que son aquellas cuyas gráficas pueden representarse en un plano cartesiano como lugares geométricos consistentes en aquellos puntos (x, y) del plano para los cuales $y = f(x)$.

La geometría analítica de Fermat y Descartes había preparado el terreno para la creación del concepto de función como se describió en el párrafo anterior. Además la propia geometría analítica proporcionó muchos ejemplos concretos de funciones que pudieron aplicarse directamente al cálculo. Para empezar, los de función lineal $y = ax + b$ y función cuadrática $y = ax^2 + bx + c$ sirven al cálculo como puntos de partida para comprobar la validez de sus métodos en unos casos en que las derivadas y las integrales pueden encontrarse fácilmente.

Por ejemplo, si $y = mx$ entonces la razón de cambio $\frac{\Delta y}{\Delta x}$ es igual a m en cualquier parte de la gráfica. Y es que la recta tangente a la gráfica de la función $f(x) = mx$ coincide exactamente con la recta $y = mx$. En otras palabras, si $f(x) = mx$, entonces $f'(x) = m$. Por otro lado, en este caso la región entre la curva y el eje- x es siempre un trapecio y su área puede calcularse directamente:

$$\int_a^b f(x)dx = \frac{f(a) + f(b)}{2}(b - a) = \frac{(mb + ma)(b - a)}{2} = \frac{mb^2}{2} - \frac{ma^2}{2}$$

De manera que si se define la función $F(x) = \frac{mx^2}{2}$, se tiene que

$$\int_a^b f(x)dx = F(b) - F(a)$$

Finalmente, es fácil comprobar que $F'(x) = mx$. La gráfica de $y = \frac{mx^2}{2}$ es una parábola con foco en el punto $(0, \frac{1}{2m})$ y directriz $y = -\frac{1}{2m}$. Sea $(a, \frac{ma^2}{2})$ un punto cualquiera de esta parábola. La recta tangente a la parábola en ese punto es la mediatriz de los puntos $(a, -\frac{1}{2m})$ y $(0, \frac{1}{2m})$ (el foco) y por tanto es ortogonal a la recta que los une, la cual tiene pendiente $-\frac{1}{ma}$. Por tanto la pendiente de la recta tangente a la gráfica de $y = \frac{mx^2}{2}$ en el punto $(a, \frac{ma^2}{2})$ es ma . Esto demuestra que la derivada de $f(x) = \frac{mx^2}{2}$ es $f'(x) = mx$.

Otra manera de obtener la recta tangente en el punto $(a, \frac{ma^2}{2})$ es considerar directamente la ecuación de la mediatriz de los puntos $(0, \frac{1}{2m})$ y $(a, -\frac{1}{2m})$. Su ecuación es: $x^2 + (y - \frac{1}{2m})^2 = (x - a)^2 + (y + \frac{1}{2m})^2$, que al desarrollarse y tras simplificar los cuadrados de x , y y $\frac{1}{2m}$ (los tres desaparecen) queda

$$-\frac{2y}{2m} = -2ax + a^2 + \frac{2y}{2m}$$

que puede escribirse como

$$y = max - \frac{ma^2}{2}$$

que es la ecuación de una recta con pendiente ma , lo que constituye otra prueba de que la derivada de $f(x) = \frac{mx^2}{2}$ es $f'(x) = mx$.

Pero hay otra prueba de este hecho que muestra un método más general para obtener la derivada de una función y es la que corresponde a lo que propiamente se considera el cálculo diferencial. La idea es ésta: la pendiente de la recta tangente a la gráfica $y = f(x)$ de una función cualquiera $f(x)$ en el punto $(x, f(x))$ es el límite de la pendiente de la recta (secante) que pasa por los dos puntos cercanos $(x, f(x))$ y $(x + h, f(x + h))$, cuando h tiende a cero. En símbolos esto lo escribimos así

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

A pesar de que aparentemente esta definición de la derivada es más difícil de entender que alguna de las versiones alternativas más intuitivas, resulta ser la única de carácter completamente general. Y

además, su aplicación para obtener derivadas es más sencilla que los métodos que nos proporciona la geometría analítica y que solo aplican a funciones muy particulares, como es el caso anterior de $f(x) = \frac{mx^2}{2}$.

Efectivamente, si aplicamos la definición a esta función vemos que debemos obtener el límite cuando $h \rightarrow 0$ de

$$\frac{\frac{m(x+h)^2}{2} - \frac{mx^2}{2}}{h}$$

Desarrollando el binomio en el numerador, simplificando los términos en x^2 que se anulan y dividiendo término a término por h , se tiene

$$\frac{\frac{m(x+h)^2}{2} - \frac{mx^2}{2}}{h} = \frac{\frac{mx^2}{2} + \frac{2mxxh}{2} + \frac{mh^2}{2} - \frac{mx^2}{2}}{h} = \frac{mxxh + \frac{mh^2}{2}}{h} = mx + \frac{mh}{2}$$

Así resulta muy fácil ver que

$$\lim_{h \rightarrow 0} \frac{\frac{m(x+h)^2}{2} - \frac{mx^2}{2}}{h} = \lim_{h \rightarrow 0} \left(mx + \frac{mh}{2} \right) = mx$$

y por tanto, que la derivada de $f(x) = \frac{mx^2}{2}$ es $f'(x) = mx$.

Las dos grandes ventajas del método del cálculo diferencial para obtener derivadas son que se puede aplicar a muchas funciones y que sus resultados son muy generales, de manera que solo hay que aplicarlos una vez a ciertas funciones y, utilizando algunas reglas fáciles de aprender y de justificar, resulta muy fácil obtener las derivadas de un gran número de funciones.

La tecnología del cálculo diferencial tiene dos aspectos:

- (1) Encontrar las derivadas de las principales funciones de uso común como las potencias de x , las funciones trigonométricas, la función exponencial y la función logaritmo.
- (2) Obtener las reglas para derivar funciones que se forman por medio de sumas, multiplicaciones y divisiones de otras funciones, por la composición de funciones y por la definición de una función inversa.

En las dos secciones siguientes se presentan, en orden inverso, y de manera resumida, estos dos aspectos del cálculo diferencial. Todo esto se hizo lo hicieron los creadores del cálculo diferencial sin tener una definición precisa del concepto de límite. Y es que como se verá, para los propósitos planteados, los límites que se calculan son bastante sencillos y pueden obtenerse sin necesidad de profundizar mucho en el concepto de límite. Sin embargo, el de límite es un concepto tan fundamental para el desarrollo de las matemáticas contemporáneas, que su comprensión es imprescindible, por lo que se estudiará su desarrollo detenidamente en las secciones posteriores.

Reglas para derivar las funciones formadas por operaciones con otras

A continuación presentamos las reglas de derivación para las funciones definidas mediante operaciones de otras funciones. Demostraremos sólo dos de estas reglas, las que resultan un poco más

complicadas de probar. Casi todas ellas las enunció y demostró GOTTFRIED LEIBNIZ. Al enunciar la reglas usaremos las dos notaciones, la de f' y la de $\frac{df}{dx}$ para denotar las derivadas.

Regla de la suma

$$(f + g)'(x) = f'(x) + g'(x) \text{ equivalentemente: } \frac{d(f + g)}{dx} = \frac{df}{dx} + \frac{dg}{dx}$$

Regla del producto

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x) \text{ equivalentemente: } \frac{d(fg)}{dx} = \frac{df}{dx}g + f\frac{dg}{dx}$$

Regla de la cadena para composición de funciones:

$$(f(g))'(x) = f'(g(x))g'(x) \text{ equivalentemente: } \frac{d(f(g))}{dx} = \frac{df}{dx}(g)g'$$

La demostración de la regla de la cadena se hace dividiendo y multiplicando expresión cuyo límite hay que obtener por $g(x + h) - g(x)$:

$$\frac{f(g(x + h)) - f(g(x))}{h} = \frac{f(g(x + h)) - f(g(x))}{g(x + h) - g(x)} \frac{g(x + h) - g(x)}{h}$$

Si la función $g(x)$ es constante en una vecindad de x , la división no es válida, y en ese caso está claro que $f(g(x + h)) - f(g(x))$ es igual a 0 para h suficientemente pequeña y por tanto en ese caso tanto la derivada de $g(x)$ como la de $f(g(x))$ son cero y por tanto la regla de la cadena se cumple de manera trivial en ese caso. Si $g(x)$ no es constante en una vecindad de x , entonces $g(x + h)$ es un número cercano a $g(x)$ (aquí, claro, se está suponiendo que g es una función bien portada, es decir, continua y derivable) y por tanto el primer factor del segundo término de la igualdad anterior tiende a $f'(g(x))$ cuando $h \rightarrow 0$. El segundo factor, por definición, tiene a $g'(x)$ cuando $h \rightarrow 0$. Con esto queda demostrada la regla de la cadena.

Regla de la función inversa

$$(f^{-1})'(x) = \frac{1}{f'(f^{-1}(x))}$$

Ésta última regla se obtiene derivando ambos lados de la expresión $f(f^{-1}(x)) = x$, que es la que define a la función inversa. Al aplicar la regla de la cadena al miembro izquierdo y derivar directamente el derecho, se obtiene:

$$f'(f^{-1}(x))f^{-1}(x) = 1$$

La regla se obtiene despejando $f^{-1}(x)$ de esta igualdad.

La regla de la división de funciones

$$\left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2} \text{ equivalentemente: } \frac{d}{dx}\left(\frac{f}{g}\right) = \frac{\frac{df}{dx}g - f\frac{dg}{dx}}{g^2}$$

es muy útil también y puede obtenerse aplicando sucesivamente la regla del producto a $f\frac{1}{g}$, luego la regla para derivar potencias, que aparece en la siguiente subsección, y la regla de la cadena. Éste es un magnífico ejemplo de cómo funciona la tecnología del cálculo diferencial. Dejamos los detalles como ejercicio para el lector.

Derivadas de las funciones comunes

Derivadas de las potencias x^p

La fórmula

$$\frac{dx^p}{dx} = px^{p-1}$$

para derivar potencias x , es decir, funciones de la forma $f(x) = x^p$ es muy general. En realidad, aplica a cualquier número real p , pero sólo vamos a demostrarla para valores enteros, tanto positivos como negativos de p y para valores fraccionarios de $p = \frac{m}{n}$ con m y n números enteros arbitrarios. La demostración se basa en la fórmula binomial:

$$(a + b)^n = a^n + na^{n-1}b + \frac{n(n-1)}{2!}a^{n-2}b^2 + \frac{n(n-1)(n-2)}{3!}a^{n-3}b^3 + \dots + b^n$$

Si aplicamos la fórmula binomial al primer sumando del numerador de la expresión cuyo límite hay que calcular para obtener la derivada de x^n , tendremos:

$$\frac{(x+h)^n - x^n}{h} = \frac{x^n + nx^{n-1}h + \frac{n(n-1)}{2!}x^{n-2}h^2 + \frac{n(n-1)(n-2)}{3!}x^{n-3}h^3 + \dots + h^n - x^n}{h}$$

Cancelando x^n con $-x^n$ y dividiendo por h término a término, obtenemos

$$\frac{(x+h)^n - x^n}{h} = nx^{n-1} + \frac{n(n-1)}{2!}x^{n-2}h + \frac{n(n-1)(n-2)}{3!}x^{n-3}h^2 + \dots + h^{n-1}$$

Todos los términos de esta suma, excepto el primero, tienen como factor alguna potencia positiva de h , por lo tanto todos ellos, salvo el primero, tienden a 0 cuando $h \rightarrow 0$. Por lo tanto

$$\lim_{h \rightarrow 0} \frac{(x+h)^n - x^n}{h} = nx^{n-1}$$

lo cual demuestra la fórmula $\frac{dx^p}{dx} = px^{p-1}$ para el caso en que p es un entero positivo.

Para el caso en que p es un entero negativo $p = -n$ aplicamos la regla del producto a la función $f(x) = x^n x^{-n} = 1$ y obtenemos

$$nx^{n-1}x^{-n} + x^n \frac{dx^{-n}}{dx} = 0$$

igualdad de la cual podemos despejar $\frac{dx^{-n}}{dx}$, con lo que obtenemos:

$$\frac{dx^{-n}}{dx} = -\frac{nx^{n-1}x^{-n}}{x^n} = -nx^{-n-1}$$

lo cual demuestra nuestra fórmula para el caso $p = -n$.

Para demostrar el caso en que $p = \frac{1}{n}$ aplicamos la regla de la cadena a la función

$$f(x) = (x^{\frac{1}{n}})^n = x$$

obteniendo

$$n(x^{\frac{1}{n}})^{n-1} \frac{dx^{\frac{1}{n}}}{dx} = 1$$

Despejando $\frac{dx^{\frac{1}{n}}}{dx}$ obtenemos

$$\frac{dx^{\frac{1}{n}}}{dx} = \frac{1}{nx^{1-\frac{1}{n}}} = \frac{1}{n}x^{\frac{1}{n}-1}$$

que demuestra la fórmula para el caso $p = \frac{1}{n}$.

Finalmente, para el caso $p = \frac{m}{n}$ basta aplicar la regla de la cadena a la función $f(x) = (x^{\frac{1}{n}})^m = x^{\frac{m}{n}}$. Dejamos los detalles de este último paso como ejercicio para el lector.

Es importante observar que la demostración de cada caso casi siempre se realiza utilizando un caso anterior y la aplicación de una varias de las reglas de derivación para funciones formadas con operaciones de otras funciones. La clave está, claro en la inventar estas funciones de manera adecuada para que nos permitan obtener la derivada buscada, en general, despejándola de una ecuación a la que se llega aplicando las reglas mencionadas.

Derivadas de las funciones trigonométricas

Las fórmula de derivación para el seno y el coseno

$$\frac{d\text{sen}(x)}{dx} = \text{cos}(x) \quad \text{y} \quad \frac{d\text{cos}(x)}{dx} = -\text{sen}(x)$$

se obtiene aplicando las fórmulas para el seno y el coseno de una suma

$$\begin{aligned} \text{sen}(\alpha + \beta) &= \text{sen}(\alpha)\text{cos}(\beta) + \text{cos}(\alpha)\text{sen}(\beta) \\ \text{cos}(\alpha + \beta) &= \text{cos}(\alpha)\text{cos}(\beta) - \text{sen}(\alpha)\text{sen}(\beta) \end{aligned}$$

a las expresiones cuyos límites hay que calcular para obtener las derivadas correspondientes

$$\frac{\text{sen}(x+h) - \text{sen}(x)}{h} = \frac{\text{sen}(x)\text{cos}(h) + \text{cos}(x)\text{sen}(h) - \text{sen}(x)}{h} = \text{sen}(x)\frac{\text{cos}(h) - 1}{h} + \text{cos}(x)\frac{\text{sen}(h)}{h}$$

y

$$\frac{\cos(x+h) - \cos(x)}{h} = \frac{\cos(x)\cos(h) - \operatorname{sen}(x)\operatorname{sen}(h) - \cos(x)}{h} = \cos(x)\frac{\cos(h) - 1}{h} - \operatorname{sen}(x)\frac{\operatorname{sen}(h)}{h}$$

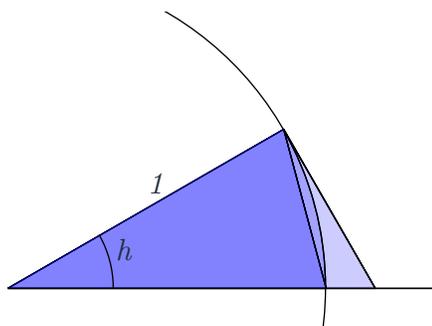
Por tanto, para obtener las derivadas del seno y el coseno basta demostrar que

$$\lim_{h \rightarrow 0} \frac{1 - \cos(h)}{h} = 0 \quad (8.1)$$

y

$$\lim_{h \rightarrow 0} \frac{\operatorname{sen}(h)}{h} = 1 \quad (8.2)$$

Estos resultados se pueden demostrar comparando las tres áreas sombreadas en este esquema



$$\frac{\operatorname{sen}(h)}{2} < \frac{h}{2} < \frac{\tan(h)}{2}$$

Si multiplicamos por 2 y dividimos por $\operatorname{sen}(h)$

$$1 < \frac{h}{\operatorname{sen}(h)} < \frac{1}{\cos(h)}$$

Invertimos la desigualdad y obtenemos

$$\cos(h) < \frac{\operatorname{sen}(h)}{h} < 1$$

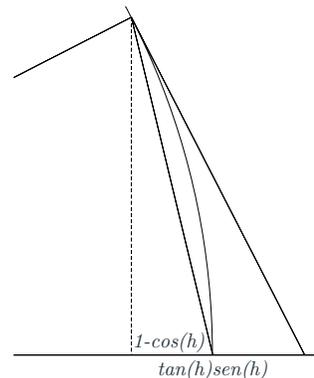
Esto demuestra (8.2). Por otro lado, comparando segmentos sobre el eje x , se ve que

$$1 - \cos(h) < \tan(h)\operatorname{sen}(h)$$

Dividiendo por h :

$$\frac{1 - \cos(h)}{h} < \tan(h)\frac{\operatorname{sen}(h)}{h} < \tan(h)$$

lo cual demuestra (8.1).



Derivadas de las funciones logaritmo y exponencial

Hay varias maneras de definir las funciones exponencial y logaritmo. Una de ellas es definir la exponencial aritméticamente, que aparentemente es la más natural. Para esto se parte de

$$a^m = a \cdot a \cdot \dots \cdot a \quad (m \text{ factores})$$

para m entero positivo. A continuación se define $b = a^{\frac{m}{n}}$ como el número para el cual $b^n = a^m$. Luego se define para una fracción p , $b = a^{-p}$ como el número para el cual $b \cdot a^p = 1$. Finalmente, se comprueba que $a^{p+q} = a^p a^q$ para cualquier par de fracciones p y q .

Pero aquí aparece una dificultad. ¿Cómo definir, por ejemplo, $a^{\sqrt{(2)}}$ o a^π ? Definir a^b para números irracionales b requiere de realizar una extensión de la definición de exponenciación. Se expresa b como el límite de una sucesión de números racionales $\{p_n\}$ tal que $b = \lim_{n \rightarrow \infty} p_n$, y *definir*

$$a^b = \lim_{n \rightarrow \infty} a^{p_n}$$

Para justificar esta definición hace falta demostrar que la exponenciación es una función *continua* en los racionales. Dejamos los detalles al lector. Finalmente, se puede ver que

$$\frac{da^x}{dx} = \lim_{h \rightarrow 0} \frac{a^{x+h} - a^x}{h} = a^x \lim_{h \rightarrow 0} \frac{a^h - 1}{h}$$

Para completar la construcción de la función exponencial por este camino, se define el número e como aquel para el cual

$$\lim_{h \rightarrow 0} \frac{e^h - 1}{h} = 1$$

En particular,

$$\lim_{n \rightarrow \infty} \frac{e^{\frac{1}{n}} - 1}{\frac{1}{n}} = 1$$

Si definimos $e_n = (1 + \frac{1}{n})^n$, vemos que para todo n ,

$$\frac{e_n - 1}{\frac{1}{n}} = 1$$

lo cual sugiere que podríamos definir

$$e = \lim_{n \rightarrow \infty} e_n = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$$

Finalmente se define la función exponencial como

$$\exp(x) = e^x$$

Todos estos pasos pueden justificarse rigurosamente, y se hace en libros de análisis y en algunos libros de cálculo.

Otro camino que se puede seguir para definir la función exponencial es usar el límite

$$\exp(x) = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n$$

este camino presenta una dificultad de inicio ya que aunque a partir de ella resulta natural definir el número e como

$$e = \exp(1) = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n$$

nada nos garantiza que

$$\exp(x) = e^x$$

Otra definición equivalente de la función exponencial es

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

que presenta el mismo problema de entrada que la anterior.

En cualquiera de las opciones que hemos mencionado hasta ahora, la definición de la función logaritmo se hace como la función inversa de la exponencial. Es decir,

$$\log(x) = y \text{ significa que } x = \exp(y)$$

Finalmente, hay una manera de definir primero la función logaritmo como

$$\log(x) = \int_1^x \frac{1}{z} dz$$

y entonces definir la función exponencial como la función inversa del logaritmo, es decir,

$$\exp(x) = y \text{ significa que } x = \log(y)$$

Explorar todos estos caminos enriquece nuestra comprensión de estas dos importantes funciones.

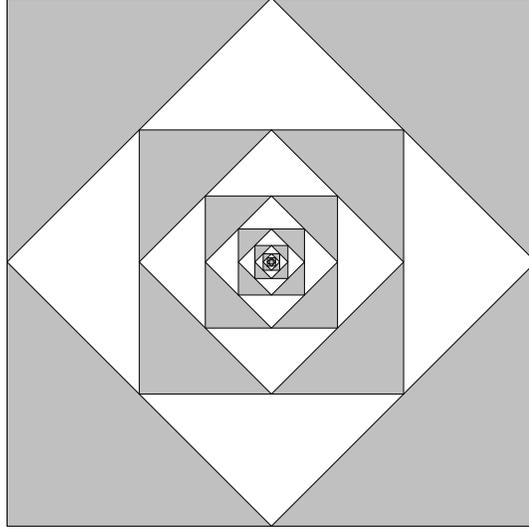
8.4. Desarrollo del concepto de límite

En lo que sigue el concepto de *límite* jugará un papel importante. En el Capítulo 5 se vieron algunos ejemplos donde se usó la idea de la aproximación: en la Sección 5.5 se vió cómo se puede aproximar la longitud de una circunferencia por polígonos inscritos y circunscritos, en la Sección 5.6 se obtuvo la superficie de la esfera por aproximación y en la Sección 5.4 se obtuvo el área de un segmento de parábola, también por aproximación.

Aquí se quiere retomar esta idea, analizarla y generalizarla a todo tipo de sucesiones. Veamos ahora algunos ejemplos de sucesiones.

Ejemplos de sucesiones convergentes y divergentes

En la siguiente figura se muestra un cuadrado, que incluye algunos triángulos sombreados. Deseamos calcular el contenido A del área sombreada suponiendo que el lado del cuadrado grande es 1.



Calculemos esta área paso a paso. Vamos a juntar las áreas de los triángulos de igual tamaño para luego sumarlas todas. El área de los primeros cuatro es la mitad del cuadrado mayor, es decir $F_1 = \frac{1}{2}$. La de los siguiente cuatro triángulos es a su vez la mitad del cuadrado de lado $\frac{1}{2}$, es decir es $\frac{1}{2}(\frac{1}{4})$. Con esto obtenemos la siguiente aproximación para A que es $F_2 = \frac{1}{2} + \frac{1}{2}(\frac{1}{4})$. Luego juntamos los cuatro triángulos del siguiente tamaño y obtenemos una nueva aproximación $F_3 = \frac{1}{2} + \frac{1}{2}(\frac{1}{4}) + \frac{1}{2}(\frac{1}{4})^2$. De esta manera obtenemos cada vez mejores aproximaciones para A :

$$F_n = \frac{1}{2} \left[1 + \frac{1}{4} + \left(\frac{1}{4}\right)^2 + \dots + \left(\frac{1}{4}\right)^{n-1} \right]. \quad (8.3)$$

La pregunta inicial acerca del área A se reformula ahora como: a qué valor tiende F_n cuando n se hace más y más grande? El paréntesis del lado derecho tiene la forma

$$t_n = 1 + x + x^2 + x^3 + \dots + x^{n-1} \quad \text{con } x = \frac{1}{4}.$$

Si se multiplica t_n por $1 - x$ se obtiene

$$\begin{aligned} t_n \cdot (1 - x) &= (1 + x + x^2 + x^3 + \dots + x^{n-1}) \cdot (1 - x) \\ &= 1 + x + x^2 + x^3 + \dots + x^{n-1} + \\ &\quad - x - x^2 - x^3 - \dots - x^{n-1} - x^n \\ &= 1 - x^n. \end{aligned}$$

De ahí que

$$t_n = \frac{1 - x^n}{1 - x}.$$

En el caso del área sombreado de arriba se tiene $x = \frac{1}{4}$ por lo que

$$F_n = \frac{1}{2} \cdot \frac{1 - \left(\frac{1}{4}\right)^n}{1 - \frac{1}{4}} = \frac{1}{2} \cdot \frac{1 - \left(\frac{1}{4}\right)^n}{\frac{3}{4}}. \quad (8.4)$$

Como $\left(\frac{1}{4}\right)^n$ tiende a cero cuando n se crece, el límite al que tiende F_n es:

$$A = \frac{1}{2} \cdot \frac{1 - 0}{\frac{3}{4}} = \frac{1}{2} \cdot \frac{4}{3},$$

es decir

$$A = \frac{2}{3}.$$

Veamos de nuevo la diferencia de las dos representaciones de F_n como se dan en las ecuaciones (8.3) y (8.4). En la ecuación (8.3) se aproxima A por una suma y de antemano no es evidente qué valor se obtendrá. En cambio, en (8.4), se transformó la expresión algebraicamente de manera que se puede ver bien a qué tiende cuando n se hace grande. Se podría decir que la expresión (8.4) está “madura para obtener el límite”, pero (8.3) no.

El ejemplo anterior es algo especial en el sentido de que de antemano se sabe que la sucesión F_n tiene que converger: por un lado crece cada vez que n aumenta porque se añade más área, por otro la suma no puede crecer demasiado dado que el área sombreado debe ser menor que 1. En este tipo de ejemplos la noción de *convergencia* es intuitivamente clara. De hecho, durante mucho tiempo a lo largo de la historia de las matemáticas, el desarrollo de esta noción fue guiado por la intuición.

Hay muchos casos en los que la intuición también nos permite determinar si la sucesión converge o no. Por ejemplo la sucesión a_n :

$$a_1 = 1, \quad a_2 = 2, \quad a_3 = 3, \quad a_4 = 4, \quad \dots, \quad a_n = n, \quad \dots$$

crece indefinidamente, es decir rebasa cualquier cota, y por lo tanto no converge. Por otro lado, es evidente que la sucesión b_n :

$$b_1 = 1, \quad b_2 = \frac{1}{2}, \quad b_3 = \frac{1}{3}, \quad b_4 = \frac{1}{4}, \quad \dots, \quad b_n = \frac{1}{n}, \quad \dots$$

converge a 0. Finalmente la sucesión c_n :

$$c_1 = 1, \quad c_2 = 3, \quad c_3 = 1, \quad c_4 = 3, \quad \dots, \quad c_n = 1 + (-1)^n, \quad \dots$$

no converge porque oscila indefinidamente entre dos valores fijos.

Pero no siempre es tan fácil determinar si una sucesión converge o no. Veamos el siguiente ejemplo:

$$d_1 = 1, \quad d_2 = 1 + \frac{1}{2}, \quad d_3 = 1 + \frac{1}{2} + \frac{1}{3}, \quad \dots, \quad d_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}, \quad \dots$$

Utilizando una computadora se obtiene

$$d_{10\,000} \approx 9.79, \quad d_{20\,000} \approx 10.48, \quad d_{30\,000} \approx 10.89.$$

Aparentemente la sucesión no va a crecer mucho más. Pero ¿nos basta esto para concluir que la sucesión converge?

No olvidemos que la sucesión d_n se construye sumando cada vez más términos. Así que se requiere de un argumento más sólido para determinar si la sucesión converge o no. Como la sucesión crece, es claro que sólo hay dos situaciones posibles: o la sucesión es acotada y entonces converge a algún valor, o no es acotada y tiende a infinito.

El siguiente argumento demuestra que la sucesión d_n diverge. Construimos una segunda sucesión s_n sustituyendo algunos sumandos de d_n por otros más pequeños, de manera que $s_n \leq d_n$ para todo n . Sustituimos el sumando $\frac{1}{k}$ por $\frac{1}{2^p}$ cuando k satisface $2^{p-1} < k < 2^p$. Por ejemplo

$$\begin{aligned} d_{10} &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9} + \frac{1}{10}, \\ s_{10} &= 1 + \frac{1}{2} + \underbrace{\frac{1}{4} + \frac{1}{4}}_{=\frac{1}{2}} + \underbrace{\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}}_{=\frac{1}{2}} + \frac{1}{16} + \frac{1}{16}. \end{aligned}$$

Por ello se tiene para $n = 2^k$

$$d_{2^k} \geq s_{2^k} = 1 + k \cdot \frac{1}{2},$$

lo que muestra que la sucesión d_n eventualmente rebasa cualquier cota dada y por lo tanto no converge.

El estilo informal del siglo XVIII

Los ejemplos anteriores muestran dos cosas: por un lado en algunos casos se pueden usar argumentos intuitivos para determinar si una sucesión converge y por otro lado que debemos ser muy cuidadosos porque a veces la intuición puede malaconsejarnos. En el siglo XVIII se pudieron determinar los límites de muchas sucesiones utilizando la intuición y el ingenio. Por ejemplo JOHANNES BERNOULLI infirió que

$$\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \dots = 1. \quad (8.5)$$

¿Como lo hizo? Utilizó la igualdad

$$\frac{1}{k} - \frac{1}{k+1} = \frac{1}{k(k+1)},$$

con la que puede escribir

$$\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \dots = \left(1 + \frac{1}{2} + \frac{1}{3} + \dots\right) - \left(\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots\right)$$

que es la diferencia de dos infinitos, pero observó que todos los sumandos se cancelan mutuamente excepto el 1, lo que prueba (8.5). ¿Es válido este argumento? Este argumento no pasaría la exigencia de rigor de hoy día. Sin embargo produjo el resultado correcto.

Ejercicio. Para profundizar en este tema sugerimos al lector producir un ejemplo en el que un argumento similar llevaría a un error. Esto muestra que hace falta un argumento más riguroso para demostrar (8.5). Sugerimos al lector producir tal argumento.

Un resultado sorprendente de esta época, que se debe a LEONHARD EULER, es el siguiente:

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}. \quad (8.6)$$

La manera como llegó a ello muestra a la vez su genio y su atrevimiento para descubrir resultados.

Dado que

$$\frac{1}{(k+1)^2} < \frac{1}{k(k+1)}$$

concluyó que

$$\frac{1}{1} + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \dots < 1 + \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \frac{1}{3 \cdot 4} + \dots = 2$$

por lo que la serie de los recíprocos de los cuadrados está acotada y por lo tanto converge. Para obtener el límite observó que si una ecuación polinomial tiene término constante igual a 1 entonces el coeficiente del término lineal es la suma de los recíprocos de sus raíces, con el signo invertido. Por ejemplo para el grado dos, la ecuación polinomial se puede escribir así

$$(x-a)(x-b) = 0,$$

donde las raíces son a y b . Si expandimos el producto y dividimos entre ab obtenemos

$$\frac{1}{ab}x^2 - \left(\frac{1}{a} + \frac{1}{b}\right)x + 1 = 0,$$

que es una ecuación polinomial con las mismas soluciones y con término constante 1. El coeficiente del término lineal es en efecto la suma de los recíprocos de las raíces, con el signo invertido. Esto se puede generalizar a polinomios de cualquier grado.

Ahora EULER parte de la ecuación $\sin(x) = 0$ y desarrolla $\sin(x)$ en serie de potencias:

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots = 0.$$

Al dividir entre x obtiene

$$1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \dots = 0$$

y al sustituir $x^2 = y$

$$1 - \frac{y}{3!} + \frac{y^2}{5!} - \frac{y^3}{7!} + \dots = 0. \quad (8.7)$$

Como las raíces de $\sin(x) = 0$ son $x = \pm k\pi$, donde k recorre los números naturales, al dividir entre x se pierde la solución $x = 0$ y al sustituir $x^2 = y$ obtiene la ecuación (8.7) cuyas raíces son $y = k^2\pi^2$, con k natural.

Ahora manipula la ecuación (8.7) como si fuera un polinomio. Dado que el término constante es 1, el coeficiente del término lineal, es decir $-\frac{1}{3!} = \frac{1}{6}$, debería ser igual a la suma de los recíprocos de las raíces, con el signo invertido:

$$-\frac{1}{6} = -\left(\frac{1}{\pi^2} + \frac{1}{4\pi^2} + \frac{1}{9\pi^2} + \frac{1}{16\pi^2} + \dots\right).$$

Al multiplicar por $-\pi^2$ obtiene (8.6).

Vemos dos aspectos importantes en el desarrollo anterior. Por un lado, JOHANNES BERNOULLI y LEONHARD EULER y sus contemporáneos descubrieron resultados muy importantes al manipular

“sumas infinitas” de manera ingeniosa. Por otro lado extiraban, sin justificación alguna, la validez de las propiedades de las sumas finitas a las infinitas. Hoy día sabemos que hacer esto es algo muy delicado y que puede conducir a errores. Curiosamente nunca se ha descubierto que EULER haya cometido un error manipulando sumas infinitas. Tenía una intuición muy profunda que le permitía actuar con gran atrevimiento sin caer nunca en el error. Lamentablemente no todos somos como Euler. Por ello necesitamos navegar por las aguas más seguras que nos brinda el rigor desarrollado en siglos posteriores.

Podemos resumir esta lección de historia diciendo que con la intuición y el atrevimiento se descubre y con el rigor se demuestra. Para hacer matemáticas conviene aprovecharlo todo.

El camino hacia una definición rigurosa de límite

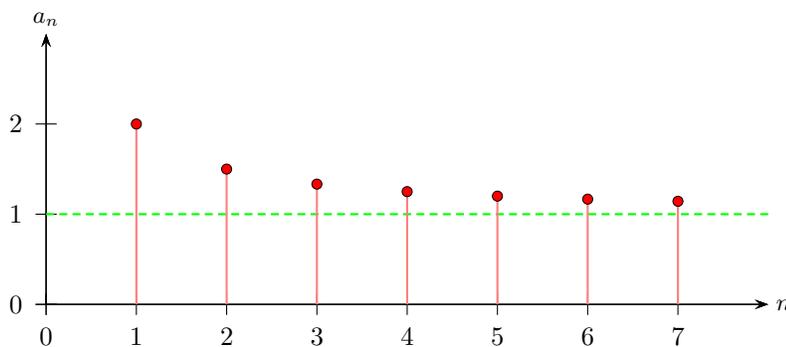
Aunque a lo largo de los dos siglos en los que se trabajó con una noción intuitiva de límite se obtuvieron muchos resultados importantes, durante ese tiempo también se encontraron ejemplos que desafiaban la intuición. Por eso algunos matemáticos del siglo XIX llegaron a la conclusión de que era conveniente aclarar el concepto de límite y definirlo de manera rigurosa. Esto no fue un camino fácil.

A continuación vamos a recrear el proceso que llevó a una definición rigurosa de límite basándonos en algunas formulaciones históricas e inventando otras que ayudan a dar continuidad al desarrollo. En el proceso usaremos algunas sucesiones creadas artificialmente para señalar las debilidades de los sucesivos intentos de definición. La definición que buscamos debe poder aplicarse a cualquier caso, aún al más artificial, es decir, debe poder determinar, para cualquier sucesión que se nos ocurra, si ésta converge o diverge.

Representaremos las sucesiones gráficamente. Por ejemplo, a la sucesión

$$\frac{2}{1}, \frac{3}{2}, \frac{4}{3}, \frac{5}{4}, \dots \quad (8.8)$$

la representamos con este dibujo:



Intuitivamente es claro que esta sucesión *tiende a 1*, o que *converge a 1*. Observemos que ambas palabras, *tiende* y *converge*, reflejan un pensamiento *dinámico*. En lo que sigue veremos que esta descripción dinámica, que es tan natural y sugerente, es muy engañosa y será necesario reemplazarla por una más *estática*.

A continuación daremos una lista de formulaciones tentativas de lo que quiere decir que una sucesión *tiende* a 1, junto con una crítica que muestra que la formulación aún no capta el sentido correctamente.

1. **La sucesión se acerca cada vez más a 1.** Esto significa que la distancia de a_n a 1 *es cada vez más pequeña*, es decir que disminuye en todos los pasos.

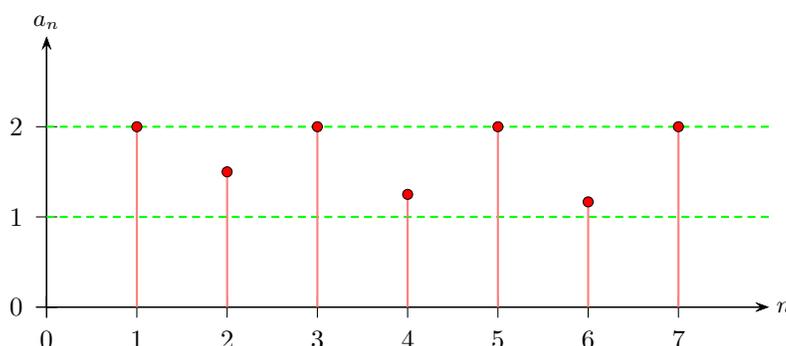
Crítica: La distancia de a_n a 0.5 también disminuye en cada paso, y sin embargo es claro que la sucesión no converge a 0.5.

2. **La distancia de a_n a 1 se vuelve arbitrariamente pequeña.** Esta formulación ya es más acertada que la anterior, sin embargo aún es insuficiente.

Crítica: Consideremos ahora la sucesión a_n definida por:

$$a_n = \begin{cases} 1 + \frac{1}{n}, & \text{si } n \text{ es par,} \\ 2, & \text{si } n \text{ es impar.} \end{cases}$$

Su representación gráfica es:



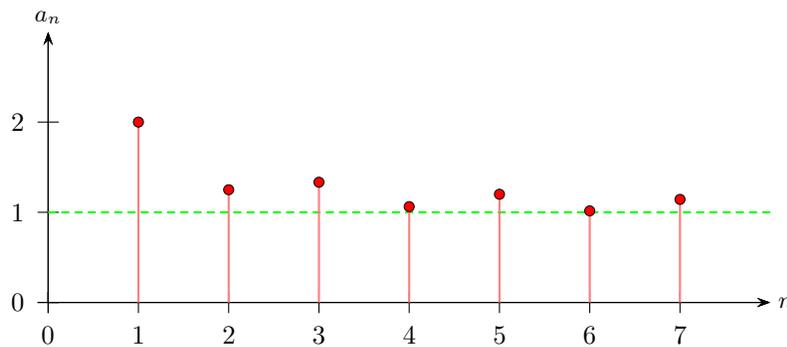
Se puede decir, y no sin razón, que en esta sucesión la distancia de a_n a 1 se hace arbitrariamente pequeña para muchos miembros de la sucesión, pero no para todos. Para los índices impares la distancia es siempre igual a 1. De hecho esta sucesión no tiene *un* límite, en todo caso tiene *dos*. Aún no hemos captado bien una parte importante del concepto de límite.

3. **La distancia de a_n a 1 disminuye cada vez más y se hace arbitrariamente pequeña.** Aquí se conjuntan las ideas de los intentos 1 y 2 y con ello evitamos la crítica a la definición anterior.

Crítica: Según esta definición la siguiente sucesión no converge:

$$a_n = \begin{cases} 1 + \frac{1}{n^2}, & \text{si } n \text{ es par,} \\ 1 + \frac{1}{n}, & \text{si } n \text{ es impar.} \end{cases}$$

Mostramos aquí la representación gráfica de esta sucesión:



Se trata de una sucesión que converge a 1, pero que no cumple la propiedad expresada en 3, ya que la distancia de a_n a 1 no disminuye conforme el índice n aumenta. En una sucesión que converge es posible que la distancia al valor límite aumente temporalmente y más tarde disminuya.

4. **Para índices N suficientemente grandes, la distancia entre a_N y 1 se hace arbitrariamente pequeña y después no vuelve a rebasar esta distancia.** Lo último quiere decir que la distancia de a_n a 1 para $n > N$ nunca es mayor que la de a_N a 1.

Crítica: En esta formulación hay un problema: para cada índice N el miembro a_N tiene una distancia fija a 1. Por ello no tiene sentido decir que si escogemos N suficientemente grande, a_N estará “arbitrariamente” cerca de 1, ya que es un valor fijo. Un valor que está “arbitrariamente” cerca de otro es idéntico a éste.

Ahora llegamos al punto clave: en todos los intentos anteriores se ha tratado de definir el límite por cómo se comporta la distancia de a_n a 1 cuando n crece. En el último tratamos de describir qué pasa con la *distancia* a partir de un *índice* N “suficientemente grande”. En la siguiente formulación invertimos el orden entre *índice* y *distancia*: tratamos de describir cómo deben ser los índices para que la distancia sea suficientemente pequeña.

5. **Para cada distancia a 1 que consideremos, hay un índice N tal que los miembros a_n con $n > N$ distan de 1 menos que ella.** Esta es la definición correcta. Para llegar a ella tuvimos que fijar primero una distancia y después dar una cota del índice tal que todos los elementos de la sucesión a partir de esa cota distarán de 1 menos que la distancia dada.

Es una definición compleja y formularla solamente con palabras resulta torpe y difícil de leer. Es un caso típico en donde la introducción de una variable, aquí ε para “la distancia dada”, ayuda a presentar la idea en forma más concisa y legible. Así es como se llega a la definición que se encuentra en los libros de texto.

Definición. Una sucesión a_n converge a 1 si

$$\text{para cada } \varepsilon > 0 \text{ existe } N \text{ tal que } |a_n - 1| < \varepsilon \text{ para todo } n \geq N.$$

Esta formulación la dió BERNARD BOLZANO en 1817 en palabras y después KARL WEIERSTRASS de manera formal con la ε , tal como la dimos arriba.

Observemos que esta formulación es meramente estática, en ella ha desaparecido por completo el aspecto dinámico. La expresión “la sucesión tiende a 1” se sigue usando porque refleja la noción intuitiva que tenemos del concepto de límite. Pero su uso tiene ahora el significado, más riguroso,

de la formulación de BOLZANO y WEIERSTRASS.

Puntos de acumulación, sucesiones de Cauchy y números reales

Se dice que a es un *punto de acumulación* de una sucesión s_n si en cualquier intervalo abierto que contenga a hay un número infinito de miembros de la sucesión. La definición formal de *punto de acumulación* la dio primero GEORG CANTOR en 1872, pero una década antes KARL WEIERSTRASS ya había usado el concepto, de manera implícita, en sus estudios sobre la convergencia.

Hay dos maneras de formular que una sucesión converge a a basadas en el concepto de *punto de acumulación* :

a es el único punto de acumulación de la sucesión.

Cada intervalo abierto que contiene a a , contiene a todos los puntos de la sucesión, excepto a un número finito de ellos.

La siguiente formulación, que se debe a AUGUSTIN LOUIS CAUCHY, expresa cuando una sucesión converge sin mencionar en ningún momento el valor al que converge.

Si para cada distancia $\varepsilon > 0$ existe una cota N tal que para todos los $m, n \geq N$ se tiene $|s_m - s_n| \leq \varepsilon$, entonces la sucesión converge.

Las sucesiones que tienen esta propiedad, es decir para las cuales “dada una distancia $\varepsilon > 0$, existe una cota N tal que para todos los $m, n \geq N$ sucede que $|s_m - s_n| \leq \varepsilon$ ” se llaman *sucesiones de Cauchy*, de manera que la formulación anterior se puede abreviar diciendo que “las sucesiones de Cauchy convergen”.

A veces los miembros de una sucesión pertenecen a un conjunto M , como por ejemplo los números racionales $M = \mathbb{Q}$, en donde es posible hablar de la distancia entre sus elementos y de las sucesiones de Cauchy formadas con sus elementos, pero para los cuales los límites de estas sucesiones de Cauchy no siempre existen como miembros del conjunto. Éste es el caso de las sucesiones de números racionales que convergen a valores irracionales, como por ejemplo la sucesión $s_n = \left(1 + \frac{1}{n}\right)^n$, cuyos miembros son números racionales, pero cuyo límite es el número de EULER e , que es irracional y no pertenece a M . Las sucesiones de Cauchy permiten determinar números que no se encuentran en M y así ampliar el conjunto M . La cerradura \overline{M} de un conjunto M se define como el conjunto de todos sus puntos de acumulación. Es fácil ver que para cada punto de acumulación de M se puede construir una sucesión de Cauchy que converja a él. En otras palabras, la cerradura de un conjunto M consta de todos los límites de las sucesiones de Cauchy formadas con elementos de M . Esta idea es la que se utiliza para definir el conjunto de los números reales como $\mathbb{R} = \overline{\mathbb{Q}}$, es decir, como la cerradura del conjunto de los números racionales. No desarrollaremos formalmente la construcción de los números reales, que involucra detalles sutiles como el concepto de “clases de equivalencia de sucesiones de Cauchy”, pero en realidad la idea central de tal construcción ya está expuesta.

Conclusiones

Las matemáticas no siempre se desarrollan con el mismo rigor. Sobre todo al principio del desarrollo de una nueva área, cuando se aprovecha el uso de conceptos que apenas se están analizando, el rigor es de segunda importancia. Mucho más atractivos parecen los logros que se pueden obtener con esos nuevos conceptos y sólo más tarde empieza una fase de reflexión sobre sus fundamentos. Esto es muy

marcado en el desarrollo de toda el área que hoy llamamos “análisis” y se ejemplifica perfectamente con la historia del concepto de límite.

8.5. Leyes del movimiento Planetario

Copérnico, Kepler, Newton

8.6. Expansión del Cálculo

El Cálculo diferencial e integral como tecnología.

8.7. Ecuaciones diferenciales

como herramienta de modelación de procesos y fenómenos. Ejemplo significativo: el oscilador armónico forzado y amortiguado.

8.8. Series de Potencias

Una herramienta útil en sitios insospechados. Euler.

8.9. Formalización del Cálculo

Descubrimiento de funciones monstruosas: necesidad de formalizar el cálculo. El análisis real, complejo y funcional.

Cauchy y Weierstrass

Cantor y los números infinitos

Una función inyectiva del cuadrado $(0, 1) \times (0, 1)$ al intervalo $(0, 1)$

Representaremos a los números reales del intervalo abierto $I = (0, 1)$ con un desarrollo binario de manera que la representación de cada uno sea única. Lo haremos evitando las representaciones que tiene sólo ceros a partir de algún índice. Si $x \in I$, entonces $x \in (0, \frac{1}{2}]$ o $x \in (1/2, 1)$. En el primer caso tomamos $a_1 = 0$ y en el segundo $a_1 = 1$. Si en el primer caso además $x = \frac{1}{2}$ entonces definimos $a_i = 1$ para todo $i > 1$. Para los siguientes pasos procedemos por inducción. Supongamos que ya tenemos definidos los elementos a_i para $i < n$ pero aún no han quedado definidos los elementos a_i para $i \geq n$. Entonces $x \in (\frac{2k}{2^n}, \frac{2k+1}{2^n}]$ o $x \in (\frac{2k+1}{2^n}, \frac{2k+2}{2^n})$ para algún entero no negativo $k < 2^n - 1$. En el primer caso definimos $a_n = 0$ y en el segundo dos $a_n = 1$. Y si en el primer caso además $x = \frac{2k+1}{2^n}$,

entonces definimos $a_i = 1$ para todo $i > n$. Esto define una sucesión a_n de ceros y unos para cada $x \in I$ tal que

$$x = \sum_{n=1}^{\infty} \frac{a_n}{2^n}$$

La representación es única porque si dos representaciones de este tipo a_n y b_n no son idénticas, deben diferir en algún índice, es decir, $a_k \neq b_k$ para algún índice k . Entonces alguno de los dos elementos a_k y b_k es igual a 0 el otro es igual a 1. Podemos suponer, sin pérdida de generalidad, que $a_k = 1$ y $b_k = 0$. La única manera de que ambas sumas infinitas pudieran ser iguales, es decir, de que

$$\sum_{n=1}^{\infty} \frac{a_n}{2^n} = \sum_{n=1}^{\infty} \frac{b_n}{2^n}$$

es que $b_j = 1$ y $a_j = 0$ para todo $j > k$. Pero esto último es imposible porque ninguna de las representaciones que definimos tiene ceros en la cola.

Ahora definimos la siguiente función $f : I \times I \rightarrow I$. Para cada $(x, y) \in I \times I$, tomamos las representaciones binarias únicas de x y y . Supongamos que

$$x = \sum_{n=1}^{\infty} \frac{a_n}{2^n}; \quad y = \sum_{n=1}^{\infty} \frac{b_n}{2^n}$$

Con ellas definimos $c_{2n-1} = a_n$ y $c_{2n} = b_n$ para todo entero positivo n y definimos

$$f(x, y) = \sum_{n=1}^{\infty} \frac{c_n}{2^n}$$

La función así definida es inyectiva porque si $(x', y') \in I \times I$ no es igual a (x, y) , entonces $x \neq x'$ o $y \neq y'$. En cualquiera de los dos casos resultaría que la representación de $f(x', y')$ tendría algún índice diferente de la correspondiente representación de $f(x, y)$ y por tanto $f(x, y) \neq f(x', y')$.

La existencia de esta función nos dice que, en cierto sentido, el conjunto $(0, 1) \times (0, 1)$ tiene al menos tantos elementos como $(0, 1)$, y esto contradice nuestra primera intuición.

8.9.1. Sumas de Riemann y teoría de la medida

8.10. Métodos numéricos en el Cálculo

Capítulo 9

La probabilidad

El orden del azar

9.1. La Invención de la probabilidad

La probabilidad se puede cuantificar: Pascal, Fermat y los juegos de azar.

Conceptos básicos para describir el azar: espacio muestral, variable aleatoria, esperanza y distribución de probabilidad.

9.2. El cálculo combinatorio

9.3. La ley de los grandes números

La ley de números grandes: la intuición vs. La cruda realidad.

9.4. El Teorema del límite central y la campana de Gauss

Conocer la distribución normal o campana de Gauss:

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Experimentar cómo ésta aparece de manera natural como la distribución de resultados de algunos comportamientos probabilísticos, por ejemplo, la caminata aleatoria o los errores de una medición. Conocer la característica forma de campana de su gráfica, y entenderá el papel que juegan la media μ y la desviación estándar σ .

Demostrar que el área bajo su gráfica es 1 (lo cual es necesario para que sea una distribución de probabilidad) mediante una aplicación del cálculo del volumen bajo la gráfica de la función de dos variables $e^{-\frac{x^2}{2\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}$ en todo el plano xy , usando coordenadas polares:

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx \int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma^2}} dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2\sigma^2}} dx dy = \int_0^{2\pi} \int_0^{\infty} e^{-\frac{\rho^2}{2\sigma^2}} \rho d\rho d\theta$$

Esta integral se puede evaluar fácilmente y su valor es $2\pi\sigma^2$. Por lo tanto

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx = \sqrt{2\pi}\sigma$$

lo cual, con un sencillo cambio de variable, demuestra que

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$$

para cualquier valor de la media μ .

Descubrir experimentalmente, por medio de simulaciones, que la distribución de los resultados de, por ejemplo, una caminata aleatoria, converge a la distribución normal (Teorema del límite central) a medida que el tamaño de los ensayos crece.

Saber que Carl Friedrich Gauss fue quien descubrió la distribución normal y dio la primera demostración del Teorema del límite central que asegura que las variables aleatorias independientes se distribuyen de manera normal, es decir, de acuerdo con la campana de Gauss. También saber que en las aplicaciones prácticas, cuando se desconoce la distribución de alguna variable aleatoria, como por ejemplo la altura de los miembros de una población, se suele adoptar la hipótesis de que tiene una distribución normal

La formalización de Kolmogorov

Retomando la teoría de la medida Kolmogorov formaliza la Teoría de la Probabilidad.

Aplicaciones del azar

Muestreos, pruebas de hipótesis, decaimiento radioactivo.

El azar como herramienta de cálculo: algoritmos randomizados.

9.5. Formalización de la Teoría de Probabilidad

Kolmogorov etc

9.6. La estadística

Capítulo 10

Otros desarrollos

Algunos ejemplos de logros más recientes

10.1. Introducción

10.2. Teoría de gráficas

Contiene la fórmula de Euler.

10.3. Geometría de la forma: la Topología

Cuando concebimos la Geometría como el estudio de aquellas propiedades de las figuras y cuerpos que se conservan antes ciertas transformaciones, e imaginamos transformaciones distintas a las rígidas, las homotecias y las proyecciones, es natural considerar deformaciones que solo tengan la restricción de no romper dichas figuras o cuerpos. Esta idea da origen a la Topología. Por ejemplo, un cuerpo en tres dimensiones que puede obtenerse de una esfera deformándola de cualquier manera, pero sin hacerle un agujeros, se dice que es topológicamente equivalente a la esfera.

10.4. Las matemáticas discretas

10.5. Los problemas de Hilbert

Ejemplo: el tercer problema.

10.6. Las corrientes fundacionistas

El programa de formalización de Hilbert. Un éxito: Los fundamentos de la Geometría. La visión de Félix Klein sobre el conflicto entre el programa de Hilbert y la enseñanza. El límite de la formalización: Gödel y Turing.

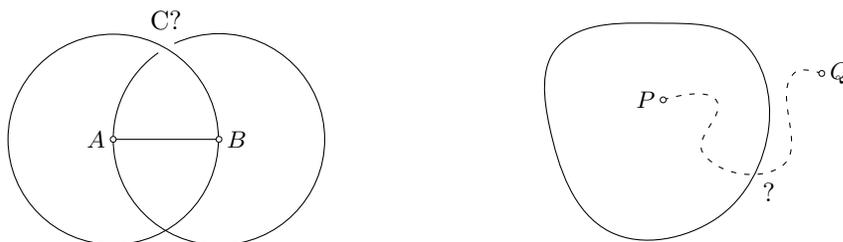
Preocupación por los fundamentos: Intuicionismo, Logicismo y Formalismo.

10.7. El Teorema de Gödel y el de Turing

10.7.1. El camino hacia los sistemas formales

Uno de los más grandes logros de la cultura griega fue el desarrollo del método axiomático, paradigma de la honestidad intelectual, que consiste en poner sobre la mesa todo lo que se supone de antemano (los axiomas) y de separarlo de las consecuencias (los teoremas) de esos supuestos. A los griegos les debemos que las matemáticas se convirtieran en una ciencia deductiva, en donde todas las afirmaciones se organizan en un tejido argumentativo. El libro *Los Elementos* de EUCLIDES DE ALEXANDRIA exhibe el método axiomático de manera ejemplar, tan ejemplar que fue por mucho tiempo el ideal del rigor en la argumentación.

Pero a lo largo de la historia se descubrieron algunas fallas en las argumentaciones de EUCLIDES. Por ejemplo, ya en la primera proposición, donde quiere demostrar que sobre cualquier segmento se puede construir un triángulo equilátero, asume algo que no ha dicho en los axiomas. Su argumentación es la siguiente: Sea dado el segmento AB ; se construye ahora la circunferencia con centro A que pasa por B y, por otro lado, la circunferencia con centro en B que pasa por A (esto se justifica por el segundo postulado). Con C , la intersección de estas circunferencias, se tiene que ABC es equilátero. El hueco en la demostración es éste: ¿por qué existe el punto de intersección C ?



Podríamos intentar demostrar la existencia de la intersección así:

Hay una parte de la primera circunferencia (la que tiene centro en A) que está dentro de la segunda (la que tiene centro en B) y otra parte de la primera que está afuera de la segunda, por tanto debe haber algún punto de la primera que esté en la segunda.

Pero este argumento utiliza un “hecho” que no está incluido en los postulados de EUCLIDES y que no hemos deducido de ellos. En el método axiomático no nos permitimos usar de repente algo de nuestra experiencia que no hemos deducido de los postulados y que no tampoco hemos declarado como uno de ellos.

Hoy se sabe que éste es un punto delicado. Hay un resultado llamado el *Teorema de la curva de*

JORDAN, en honor a CAMILLE JORDAN (1838 – 1922) que dice lo siguiente:

El borde de un área conexas divide al plano en dos partes: interior y exterior; y no es posible unir un punto en el interior con uno en el exterior por una curva continua sin cruzar el borde.

Este resultado se puede deducir, pero para ello se requieren axiomas adicionales. O sea que el punto de intersección sí existe, la primera proposición de Euclides es demostrable, pero no es consecuencia de los postulados de EUCLIDES, sino que requiere de un conjunto más completo de axiomas.

Éste y otros fallos indicaron a los matemáticos que tenían que ser mucho más cuidadosos a la hora de definir los axiomas de lo que lo fue EUCLIDES. A finales del siglo XIX en la comunidad matemática quedaba claro que era conveniente contar con un sistema axiomático para cada área de las matemáticas y que había que formular la noción de “demostración” de manera muy estricta. Más aún, los conceptos involucrados deberían definirse únicamente a través de los propios axiomas. Por ejemplo:

para cada dos “puntos” existe exactamente una “recta” que incide en ellos

define parcialmente tanto el concepto de punto como el de recta.

Hubo logros importantes en la axiomatización de aquel tiempo: la lógica la fundamentó FRIEDRICH LUDWIG GOTTLIB FREGE en 1879; la aritmética se axiomatizó gracias a las contribuciones de RICHARD DEDEKIND en 1888 y de GIUSEPPE PEANO en 1889; la geometría euclidiana la axiomatizó DAVID HILBERT en 1889, la teoría de conjuntos la axiomatizaron ERNST ZERMELO y ABRAHAM FRAENKEL entre 1908 y 1922 y finalmente, la axiomatización de los fundamentos de las matemáticas (los números cardinales y ordinales y los números reales) se debe al trabajo de ALFRED NORTH WHITEHEAD y BERTRAND RUSSELL realizado entre 1909 y 1913.

Con todo esto, en las primeras décadas del siglo XX el panorama de las matemáticas sugería que era posible, y tal vez conveniente, llevar a las matemáticas a un estado de total perfección. DAVID HILBERT, que era entonces el personaje más influyente entre los matemáticos, formuló un programa de trabajo ambicioso y optimista que pretendía lograr que todas las matemáticas siguieran un camino absolutamente certero y seguro. De lejos, las matemáticas parecen un lenguaje abstracto que usa símbolos de su propia invención. Tal vez, planteó HILBERT, las matemáticas son – o deben ser – precisamente eso: *un juego consistente en manipular símbolos abstractos*. Aunque éste ya no es el punto de vista de la mayoría de los matemáticos actuales, resulta formativo perseguir tal idea hasta sus últimas consecuencias, porque al hacerlo se llegan a descubrir sus limitaciones intrínsecas. Tales limitaciones son las de las propias matemáticas cuando éstas se conciben bajo el programa de HILBERT.

HILBERT propuso describir el lenguaje de las matemáticas como un sistema de símbolos carentes de significado que se pueden encadenar. Debe quedar claro:

- (1) cuáles símbolos son los que se pueden usar,
- (2) cuándo una cadena de ellos está bien formada (es decir es sintácticamente correcta),
- (3) cuáles son las cadenas bien formadas que se aceptan como válidas desde el principio y
- (4) cuáles son las reglas para deducir la validez de una cadena a partir de la de otras.

Con ello pedía un gran sacrificio: privar a las matemáticas de cualquier significado. Esto era un paso necesario, para obtener la certidumbre absoluta. Así surgió la idea de un *sistema formal*.

Dado que la intención era de desarrollar sistemas formales que podían representar ciertas partes

de las matemáticas, éstos deberían incluir cierta porción de lógica. Por ejemplo, debería ser posible hablar de la negación de una cadena dentro de este sistema.

Hay un par de exigencias que son bastante naturales acerca de sistemas formales. Por un lado se espera formular axiomas que son *independientes* unos de otros, es decir que ninguno de ellos sea deducible de los demás. Ésta no es una exigencia imprescindible, pero es deseable desde una posición intelectual pues no se quiere imponer como axioma una afirmación que se pueda obtener como consecuencia de otros axiomas.

Aún más importante es la exigencia de *consistencia*, es decir no debería ser posible derivar un enunciado y también su negación. Esto también se expresa a veces diciendo que el sistema debe estar *libre de contradicciones*.

Otra exigencia que es deseable es que dado un enunciado sintácticamente correcto, éste debería ser derivable, o bien, su negación debería ser derivable. A un sistema que cumple esta propiedad se le llama *completo* o más específicamente, *completo por negación*. En un sistema de este tipo las cadenas bien formadas se dividen en dos clases: la de aquellas que son derivables y la de aquellas cuya negación es derivable. Si el sistema es también *consistente*, las dos clases son ajenas.

Desde el punto de vista de la lógica tradicional la exigencia de que un sistema formal sea *completo por negación* es muy natural. ¿Si los objetos tienen propiedades o no las tienen, por ejemplo, un número es primo o no es primo, un triángulo es rectángulo o no es rectángulo; por qué en las afirmaciones esto debería ser diferente? Hay una infinidad de primos o no hay una infinidad de primos; las mediatrices de un triángulo inciden en un punto o no inciden en un punto. La expectativa en aquel entonces era que para cualquier afirmación que se pudiera formular de manera sintácticamente correcta, debería poder decidirse si era correcta o no, y en el sistema formal “ser correcto” significa “ser derivable a partir de los axiomas usando las reglas derivación”.

El optimismo de HILBERT llegaba aún más lejos: él esperaba que en algún momento sería posible encontrar una “mecanismo” para poder decidir si una cadena es derivable o no. Un sistema en el que tal mecanismo existe se llama *decidible*.

Estas exigencias para los sistemas formales de las matemáticas hoy día se conocen genéricamente como el *programa de HILBERT*.

PROGRAMA DE HILBERT PARA LOS SISTEMAS FORMALES DE LAS MATEMÁTICAS

- que los axiomas sean independientes
- que los axiomas sean consistentes
- que el sistema sea completo por negación
- que el sistema sea decidible

Debido a los trabajos de KURT GÖDEL y ALAN TURING, publicados en 1931 y 1936 respectivamente, y a algunos otros publicados posteriormente, sabemos que el programa de HILBERT no se puede realizar en los sistemas formales que pudieran representar a la lógica y a los números naturales, es decir a las matemáticas básicas.

GÖDEL demostró que si se exige la *consistencia* en un sistema, éste no puede ser *completo por negación*, es decir, en cualquier posible sistema axiomático para los números naturales siempre es posible definir un enunciado sobre ellos que no se pueden derivar de los axiomas y que tampoco su negación se puede derivar de los axiomas.

Por otro lado, TURING demostró que es imposible diseñar un mecanismo que pueda determinar para cualquier enunciado si éste es o no derivable a partir de los axiomas.

Así que hoy sabemos que el programa de HILBERT era demasiado optimista y que no se puede realizar íntegramente. A lo más que podemos aspirar es a tener sistemas formales en que los axiomas sean independientes y consistentes (como es el caso en *Los fundamentos de la Geometría* de HILBERT), pero no podemos aspirar a tener un sistema formal que sea *completo por negación* o que sea *decidible*.

El resto de esta sección presenta de manera bastante precisa los resultados de GÖDEL y TURING y da una idea simplificada de los métodos que usaron para demostrarlos.

10.7.2. Los sistemas antes de 1931

Para poder explicar el contenido de los trabajos de KURT GÖDEL y de ALAN TURING es necesario conocer el grado de avance en los fundamentos de las matemáticas en el año 1930.

En las matemáticas, la lógica suele jugar un papel fundamental. Por ello, lo primero que haremos es presentar un sistema formal con el cual se puede modelar las maniobras lógicas. Esto se conoce como la *lógica de proposiciones*. Esta lógica es fundamental, pero no es todo, ya que queremos expresar algo sobre los objetos, sean estos puntos y rectas, o números. Para ello presentaremos un sistema lógico que es muy general y en muchos casos es posible expresar la parte de las matemáticas que nos interesa en ella, se llama *lógica de predicados*. Finalmente presentaremos una lógica de predicados en particular que nos servirá para poder manipular números naturales.

En la *lógica de proposiciones* se trata de poder expresar los conectores lógicos más básicos. Consideramos las afirmaciones

A = “López está enfermo.” y **B** = “Hernández fue de vacaciones.”

Cada una de ellas puede ser verdadera o falsa. Ahora podemos conectarlos como por ejemplo:

C = “López no está enfermo.”
D = “López está enfermo y Hernández no fue de vacaciones.”

Lo importante en ello, que podemos deducir si estos últimos son verdaderos o falsos, a partir de la veracidad de las afirmaciones arriba. Por ejemplo, si sabemos **A** es verdadera y **B** falsa sabemos que **C** es falsa y **D** es verdadera. Con las abreviaciones **A**, **B** podemos escribir **C** y **D** como sigue del lado izquierdo.

C = no A	C = \neg A
D = A y (no B)	D = A \wedge (\neg B)

Del lado derecho se denotaron las mismas construcciones pero se usaron los símbolos de lógica correspondientes: “ \neg ” para la *negación* (de lo que sigue del lado derecho) y “ \wedge ” para la *conjunción* (de lo que está del lado izquierdo y derecho). La siguiente tabla muestra algunos de los símbolos más:

símbolo	nombre	uso	significado
\neg	negación	$\neg \mathbf{A}$	“no \mathbf{A} ”
\wedge	conjunción	$\mathbf{A} \wedge \mathbf{B}$	“ \mathbf{A} y \mathbf{B} ”
\vee	disyunción	$\mathbf{A} \vee \mathbf{B}$	“ \mathbf{A} o \mathbf{B} ”
\rightarrow	implicación	$\mathbf{A} \rightarrow \mathbf{B}$	“si \mathbf{A} entonces \mathbf{B} ”
\leftrightarrow	equivalencia	$\mathbf{A} \leftrightarrow \mathbf{B}$	“ \mathbf{A} si y sólo si \mathbf{B} ”
\Leftrightarrow	anti-equivalencia	$\mathbf{A} \Leftrightarrow \mathbf{B}$	“o \mathbf{A} o \mathbf{B} ”

Podemos ver \mathbf{A}, \mathbf{B} como variables y \mathbf{C}, \mathbf{D} como fórmulas que se construyen a partir de las variables. En el sistema formal que expresa la lógica de proposiciones se parte de unas variables y se consideran las fórmulas que se pueden obtener a partir de las variables al usar los conectores de lógica.

De particular interés son las fórmulas que son verdaderas siempre, sin importar qué valor tienen las variables. De una fórmula así se dice que es *general* o *tautológica*. Un ejemplo: la fórmula

$$\mathbf{A} \rightarrow (\mathbf{B} \rightarrow \mathbf{A})$$

es tautológica. Para ello veamos la siguiente tabla en donde V significa “verdadero” y F significa “falso”.

\mathbf{A}	\mathbf{B}	$\mathbf{B} \rightarrow \mathbf{A}$	$\mathbf{A} \rightarrow (\mathbf{B} \rightarrow \mathbf{A})$
V	V	V	V
V	F	V	V
F	V	F	V
F	F	V	V

Pero la lógica de proposiciones no es suficiente para expresar afirmaciones en matemáticas. En muchas afirmaciones se expresa algo para una generalidad de objetos. Por ejemplo:

“Esta función tiene al menos una raíz.”
 “Para todo x se tiene $x^2 \geq 0$.”

Para ello se requieren de cuantificadores que expresan “para todos” y “existe algún”. Estos cuantificadores se introducen en la *lógica de predicados* y sus símbolos son \forall y \exists . Esta es una lógica más compleja en la cual se diferencia entre variables, funciones y predicados. Las variables son aquellos símbolos del sistema formal, que después serán sustituidos por objetos como números o puntos de un plano. Las funciones aceptan un número definido de argumentos (este número se llama *aridad*) y darán otra vez un objeto. Por ejemplo $\text{sum}(x, y) = x + y$ es una función de aridad 2 para los números, $\text{medio}(x, y) =$ “punto medio del segmento xy ” es una función de aridad 2 para puntos en el plano.

Los predicados también aceptan un número definido de argumentos pero darán un valor de veracidad. Por ejemplo $\text{menor}(x, y)$ es un predicado de aridad 2 para números naturales: es verdadero si x es menor que y y falso en los demás casos. Se suele expresar un predicado como conjunto

$$\{(x, y) \mid x < y\} \subset \mathbb{N}^2,$$

donde \mathbb{N} es el conjunto de números naturales y \mathbb{N}^2 es el conjunto de pares ordenados de números naturales. Un predicado para puntos en un plano puede ser $\text{col}(x, y, z)$ que es verdadero si x, y, z se

encuentran sobre una recta y falso, si esto no es el caso. Como conjunto:

$$\{(x, y, z) \mid \text{si } x, y, z \text{ se encuentran sobre una recta}\} \subset \mathcal{P}^3,$$

si \mathcal{P} denota el conjunto de puntos en un plano. Así podríamos considerar las expresiones

$$\begin{aligned} \forall \mathbf{x} \exists \mathbf{y} \text{ menor}(\mathbf{x}, \mathbf{y}), \\ \forall \mathbf{x} \forall \mathbf{y} \text{ col}(\mathbf{x}, \mathbf{y}, \text{medio}(\mathbf{x}, \mathbf{y})), \end{aligned}$$

la primera para números naturales y la segunda para puntos en un plano. Ambas afirmaciones son verdaderas: para todo número natural x existe un número natural y tal que x sea menor que y . Para todos los puntos x, y en un plano, el punto medio de x y y se encuentra con x, y sobre una recta.

Ahora, como sistema formal, la *lógica de predicados* no provee sentido para sus variables, funciones y predicados. Como sistema formal sólo se ve el lado de sintaxis. Los símbolos en las fórmulas expresan la relación que hay entre las variables, funciones y predicados involucrados. Para darle sentido hay que considerar una *interpretación*. Para ello hay que fijar el conjunto U , de donde se pueden tomar los valores para las variable (por ejemplo los números naturales). La interpretación le asignará entonces a cada variable un elemento de U , a cada símbolo de función \mathbf{f} una función concreta $U^k \rightarrow U$ (donde k es la aridad de \mathbf{f}), y a cada predicado \mathbf{p} un subconjunto de U^m (donde m es la aridad de \mathbf{p}).

Dados U y la interpretación I se puede hablar si una fórmula es verdadera o falsa. La interpretación introduce la *semántica* ya que fija el significado de cada símbolo. Sólo con la semántica podemos hablar de verdadero o falso. A nivel de la sintaxis sólo se puede hablar de si una fórmula es derivable o no.

Finalmente podemos llegar a explicar de qué trata el sistema formal PA, la *aritmética de PEANO*. Es una lógica de predicados con cuatro símbolos para funciones

$$0, \mathbf{s}, +, \times$$

El “0” tiene aridad 0, el “ \mathbf{s} ” tiene aridad 1 y “+”, “ \times ” tienen aridad 2. Además hay un solo símbolo de predicado “=” y este es de aridad 2. En la *interpretación estándar*, el conjunto base son los números naturales: $U = \mathbb{N}$ (que incluye el cero). Además, “0” es la función constante 0, “ \mathbf{s} ” es la función sucesor y “+” y “ \times ” se interpretan como la adición y la multiplicación.

El sistema PA cuenta con los axiomas de la lógica de predicados y además algunos adicionales para determinar las relaciones entre las cuatro funciones y el predicado. Por ejemplo se tienen los axiomas (entre otros):

Axioma	Significado
$\sigma = \tau \rightarrow \mathbf{s}(\sigma) = \mathbf{s}(\tau)$	si dos números son iguales, entonces también sus sucesores son iguales
$\neg(0 = \mathbf{s}(\sigma))$	0 no es sucesor de nada.
$\sigma + 0 = \sigma$	0 es el elemento neutro de la adición.
$\sigma \times 0 = 0$	0 anula todo en la multiplicación.

Además se cuenta con el axioma de inducción que dice lo siguiente:

$$\varphi(0) \rightarrow (\forall x(\varphi(x) \rightarrow \varphi(\mathbf{s}(x))) \rightarrow \forall x \varphi(x))$$

Sin embargo, también se puede enunciar de esta otra manera que puede resultar más familiar:

$$(\varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(\mathbf{s}(x)))) \rightarrow \forall x \varphi(x)$$

Es un ejercicio de la lógica de proposiciones ver que las dos fórmulas

$$\varphi_1 = [\mathbf{A} \rightarrow (\mathbf{B} \rightarrow \mathbf{C})] \quad \text{y} \quad \varphi_2 = [(\mathbf{A} \wedge \mathbf{B}) \rightarrow \mathbf{C}]$$

son equivalentes, en el sentido de que para cualquier valor de de las variables **A**, **B**, **C** siempre φ_1 y φ_2 tienen el mismo valor de veracidad. Aquí como también más adelante escribimos fórmulas entre corchetes grandes si las queremos abreviar con algún nombre, como arriba con φ_1, φ_2 .

Una fórmula tautológica con cuantificadores con un predicado **p** de aridad dos es

$$\forall \mathbf{x} \mathbf{p}(\mathbf{x} \mathbf{x}) \rightarrow \forall \mathbf{x} \exists \mathbf{y} \mathbf{p}(\mathbf{x} \mathbf{y}).$$

Si leemos su significado resulta evidente que no importa cómo interpretamos el predicado, ni cuál sea el conjunto base U , siempre será válido.

Con el sistema formal PA es posible expresar las afirmaciones sobre números naturales en cadenas de símbolos. Podemos hacer teoría de números de esta manera. Una demostración formal es una derivación en el sistema formal: una sucesión de fórmulas, en donde cada uno es un axioma o sigue con las reglas de deducción de las fórmulas anteriores.

Introducimos la siguiente notación (en calidad de abreviación y no como parte de un sistema formal):

$\vdash \varphi$	la fórmula φ es <i>derivable</i>
$\nvdash \varphi$	la fórmula φ <i>no</i> es <i>derivable</i>
$I \models \varphi$	la fórmula φ es <i>verdadera</i> en la interpretation I
$I \not\models \varphi$	la fórmula φ es <i>falsa</i> en la interpretation I

10.7.3. Los logros de Gödel y Turing

Sobre la lógica de proposiciones se conoce el siguiente resultado: Cada fórmula es derivable si y solo si es tautológica, es decir, si es verdadera en cualquier interpretación, en forma abreviado: para toda fórmula φ se tiene

$$\vdash \varphi \Leftrightarrow \text{para todas las interpretaciones } I \text{ se tiene } I \models \varphi$$

Como hay fórmulas que para ciertas interpretaciones son verdaderas y para otras no, esto implica que la lógica de proposiciones no es completo por contradicción. En forma abreviado:

$$\text{existe una fórmula } \varphi \text{ tal que } \nvdash \varphi \text{ y también } \nvdash \neg\varphi.$$

En la lógica de proposiciones el programa de HILBERT se cumple en lo que se puede esperar: hay también un método de decidir si una fórmula es tautológica: sólo se tienen que elaborar una tabla de veracidad. En esta tabla se elabora una línea por cada caso posible de asignar valores de veracidad a las variables que aparecen en la fórmula. Luego se calcula el valor de verdad de la fórmula en cada caso. Si en cada caso, la fórmula es verdadera, entonces es una fórmula tautológica. Por lo que dijimos anteriormente, es derivable. Este proceso se podrá llevar a cabo de manera algorítmica en un número finito de pasos y así es posible decidir de cada fórmula si es derivable o no.

Ahora consideramos la lógica de predicados. Un resultado importante acerca de la lógica de predicados lo obtuvo KURT GÖDEL en su trabajo de doctorado (a la edad de 23 años!): si una fórmula en la lógica de predicados es tautológica, entonces es derivable. A este resultado se le llama *el teorema de completitud de GÖDEL*. El revés, es decir que el sistema es correcto, es mucho más fácil: los axiomas son tautológicos y las reglas de derivación hacen de una fórmula tautológica otra vez una fórmula tautológica, por lo que toda fórmula derivable será tautológica. Así que también en la lógica de predicados se tiene para cada fórmula

$$\vdash \varphi \iff \text{para todas las interpretaciones } I \text{ se tiene } I \models \varphi.$$

Pero para la lógica de predicados no existe un método para poder decidir si una fórmula es derivable o no. Esto es parte de lo que ALAN TURING DEMOSTRÓ en su trabajo de 1936. Para ello, TURING tuvo que formalizar primero el concepto de qué es un algoritmo.

La situación en el sistema de la aritmética de PEANO (abreviado por PA) es la siguiente. Si se supone que PA es consistente entonces es incompleto por negación, es decir que existe una fórmula φ tal que ni φ ni $\neg\varphi$ es derivable. Esto es *el primer teorema de incompletitud de GÖDEL* – en su *versión sintáctica*, ya que en ningún momento se habla de una interpretación. La demostración la completó JOHN BARKLEY ROSSER en 1936, cinco años después de la publicación del artículo de GÖDEL.

También existe una *versión semántica del primer teorema de incompletitud*. Para ello se considera la interpretación estándar I_{st} . El teorema afirma que si PA es correcto ($\vdash \varphi$ implica siempre $I_{st} \models \varphi$) entonces es incompleto, es decir existe φ con $I_{st} \models \varphi$ pero $\not\vdash \varphi$.

Además GÖDEL demostró que en el sistema PA no puede demostrar su consistencia, es decir es posible formular una afirmación cuyo significado es “el sistema PA es consistente” y que que es imposible de derivarlo dentro de PA. Este teorema se conoce como *el segundo teorema de incompletitud de GÖDEL*.

Es consecuencia del trabajo de ALAN TURING que el sistema PA *no es decidible*, es decir si tenemos una fórmula φ no es posible encontrar un algoritmo que arroje después de un número finito de pasos con certeza la respuesta si se tiene $I_{st} \models \varphi$ o $I_{st} \not\models \varphi$.

10.7.4. Un bosquejo de la demostración del primer teorema de incompletitud de GÖDEL

Esta frase fue escrita en español.

La primera frase de esta sección fue escrita en español, por ello afirma una verdad. La segunda oración, o sea, la que antecede a ésta, se refiere a la primera, pero la primera se refiere a sí misma. Por ello existe una *autoreferencia*. Esta se da porque se mezclan dos niveles:

- (α) Por un lado, la primera oración está compuesta por letras y éstas fueran impresas en papel o se muestran como pixeles en la pantalla. En este sentido la oración existe de manera *física*.
- (β) Por otro lado – y en otro nivel – sabemos interpretar el significado de las palabras de esta oración y lo podemos comprender. Según su contenido hace una afirmación sobre la primera oración de esta sección (en el sentido α).

La autoreferencia puede producir unas situaciones raras. Consideremos la siguiente frase:

Esta oración es falsa.

Si es verdadera (en el sentido β), entonces dice que la oración (en el sentido α) es falsa. Y si es falsa (β), entonces la oración (α) es falsa, por lo que está escrita (α) es correcto (β), esto es una contradicción. Hicimos esta excursión en la autoreferencia porque en el fondo, el primer teorema de incompletitud de GÖDEL se demuestra utilizando una autoreferencia. GÖDEL logró demostrar que es posible construir una fórmula en el sistema PA que dice

Yo no soy demostrable en el sistema PA.

Para ello una fórmula tienen que poder hacer una afirmación sobre otra fórmulas (que luego será ella misma). Esto lo logró GÖDEL con un truco ingenioso, que ahora lleva su nombre: es la *gödelización*. El asigna a cada fórmula, o más general a cada cadena de símbolos un número que identifica ésta de manera única. Para ello asigna a cada símbolo un número. La siguiente tabla muestra cómo se codifica un símbolo σ por un número $\gamma(\sigma)$:

σ	¬	∧	∨	→	∀	∃	()	0	s	+	×	=	x	y	z	...
$\gamma(\sigma)$	1	3	5	7	9	11	13	15	17	19	21	23	25	2	4	6	...

Luego, se asigna a cada cadena τ su *número de Gödel* $\Gamma(\tau)$ de la siguiente manera: Si τ es la cadena de símbolos

$$\tau = \left[\sigma_1 \sigma_2 \cdots \sigma_l \right]$$

le asigna el número

$$\Gamma(\tau) = p_1^{\gamma(\sigma_1)} \cdot p_2^{\gamma(\sigma_2)} \cdot \dots \cdot p_l^{\gamma(\sigma_l)},$$

donde p_i es el i -ésimo primo, es decir $p_1 = 2, p_2 = 3, p_3 = 5$ etcétera. Un ejemplo: Si τ es la fórmula

$$\tau = \left[\exists \mathbf{y} (\mathbf{y} \times \mathbf{y} = \mathbf{x}) \right]$$

entonces su *número de GÖDEL* es

$$\Gamma(\tau) = 2^{11} \cdot 3^4 \cdot 5^{13} \cdot 7^{23} \cdot 11^4 \cdot 13^{25} \cdot 17^2 \cdot 19^{15}$$

que tiene el valor 2 512 096 701 302 153 075 469 729 639 198 104 299 667 955 742 226 983 930 909 625 454 216 780 380 422 500 000 000 000.

Dado que la factorización en primos es única, se puede reconstruir la fórmula a partir de su número de Gödel. Además, diferentes fórmulas tienen números de Gödel diferentes.

En nuestro ejemplo, τ es una cadena que tiene una variable libre que es \mathbf{x} . Esto lo expresamos escribiendo $\tau(\mathbf{x})$ en vez de τ . Ahora bien, el significado de τ es “ \mathbf{x} ” es un número cuadrado. El número de Gödel de \mathbf{x} como cadena de un solo símbolo es $\Gamma(\mathbf{x}) = 2^{\gamma(\mathbf{x})} = 2^2 = 4$, que es un número cuadrado. No podemos sustituir \mathbf{x} por 4, ya que 4 no es un símbolo del sistema formal, pero podemos

usar la cadena

$$\bar{4} = \left[\mathbf{s}(\mathbf{s}(\mathbf{s}(\mathbf{s}(0)))) \right].$$

Así que

$$\tau(\overline{\Gamma(\mathbf{x})}) = \left[\tau(\mathbf{s}(\mathbf{s}(\mathbf{s}(\mathbf{s}(0)))) \right]$$

es verdadero y significa que \mathbf{x} tiene un número de Gödel que es un cuadrado. De esta manera, una fórmula puede expresar propiedades de otras fórmulas.

Para la demostración del teorema de incompletitud se requiere de más: una derivación en el sistema es una cadena de fórmulas con la propiedad que cada una es un axioma o se obtiene de los anteriores por medio de una regla de derivación. Para poder hablar de ellos, se consideran sucesiones (finitas) de fórmulas. Si v es una sucesión así

$$v = \tau_1 \tau_2 \dots \tau_j$$

entonces se le asigna el número de Gödel $\hat{\Gamma}(v)$:

$$\hat{\Gamma}(v) = p_1^{\Gamma(\tau_1)} \cdot p_2^{\Gamma(\tau_2)} \cdot \dots \cdot p_n^{\Gamma(\tau_j)}.$$

Es claro que $\hat{\Gamma}$ tiene las mismas propiedades que Γ , es decir a partir del número se puede reconstruir la sucesión de fórmulas y diferentes sucesiones tienen diferentes números de Gödel asociados.

En el argumento de GÖDEL las fórmulas con una sola variable juegan un papel importante. Todas estas fórmulas las ordenamos por su número de Gödel. Así obtenemos una lista $\psi_1(\mathbf{x}), \psi_2(\mathbf{x}), \dots$ de fórmulas. Los números de Gödel $n_i = \Gamma(\psi_i(\mathbf{x}))$ de estas fórmulas satisfacen $n_1 < n_2 < \dots$

En lo que sigue nos interesan las fórmulas que obtenemos cuando sustituimos la única variable libre \mathbf{x} en la fórmula $\psi_i(\mathbf{x})$ por \bar{n}_i , la cadena que representa el número de Gödel n_i , es decir nos interesa

$$\psi_i(\bar{n}_i).$$

Veamos esto en nuestro ejemplo de τ arriba. Se tiene $\tau = \psi_i$ para algún i (que no conocemos) y su número de Gödel es $n_i = 2\ 512\ 096\ 701\ 302\ 153\ 075\ 469\ 729\ 639\ 198\ 104\ 299\ 667\ 955\ 742\ 226\ 983\ 930\ 909\ 625\ 454\ 216\ 780\ 380\ 422\ 500\ 000\ 000\ 000$. Ahora

$$\psi_i(\bar{n}_i) = \left[\exists \mathbf{x}_2 (\mathbf{x}_2 \times \mathbf{x}_2 = \underbrace{\mathbf{s}(\mathbf{s}(\dots \mathbf{s}(\mathbf{s}(0))))}_{n_i \text{ veces "s"}} \underbrace{\dots}_{n_i \text{ veces "}"}) \right].$$

El número de Gödel $m = \Gamma(\psi_i(\bar{n}_i))$ de $\psi_i(\bar{n}_i)$ es

$$m = 2^{11} \cdot 3^4 \cdot 5^{13} \cdot 7^{23} \cdot 11^4 \cdot 13^{25} \cdot \underbrace{17^{19} \cdot 19^{13} \cdot \dots \cdot p_{5+2n_i}^{19} \cdot p_{6+2n_i}^{13}}_{\text{los exponentes 19, 13 se repiten } n_i \text{ veces}} \cdot p_{7+2n_i}^{17} \cdot \underbrace{p_{8+2n_i}^{15} \cdot \dots \cdot p_{7+3n_i}^{15}}_{\text{el exponente 15 se repite } n_i \text{ veces}} \cdot p_{8+3n_i}^{15}$$

que es un número verdaderamente gigantesco.

Consideramos la relación $R_{\text{gdl}} \subset \mathbb{N}^2$:

$$(m, n) \in R_{\text{gdl}} \quad \text{si y solo si} \quad \begin{array}{l} n = n_i \text{ es el número de Gödel de una fórmula } \psi_i \text{ y} \\ m \text{ es el número de Gödel de una sucesión de fórmulas} \\ \text{que es una derivación de la fórmula } \psi_i(\bar{n}_i) \end{array}$$

El trabajo fuerte de GÖDEL es haber descrito cómo es posible construir una fórmula φ_{gdl} en dos variables libres, con las siguientes propiedades:

$$\text{si } (m, n) \in R_{\text{gdl}} \Rightarrow \vdash \varphi_{\text{gdl}}(\overline{m}, \overline{n}), \quad (10.1)$$

$$\text{si } (m, n) \notin R_{\text{gdl}} \Rightarrow \vdash \neg \varphi_{\text{gdl}}(\overline{m}, \overline{n}). \quad (10.2)$$

Casi ya llegamos al final de la construcción. Consideramos ahora

$$\psi_j(\mathbf{x}) = \left[\forall \mathbf{y} \neg \varphi_{\text{gdl}}(\mathbf{y}, \mathbf{x}) \right]. \quad (10.3)$$

Una fórmula en una variable libre. Por ello aparece en nuestra lista ψ_1, ψ_2, \dots . Esto explica por qué la escribimos como ψ_j : el número j es el índice en la lista y n_j es su número de Gödel. Ahora, finalmente podemos indicar cuál es la fórmula crucial:

$$\Psi := \left[\psi_j(\overline{n_j}) \right].$$

Demostraremos que $\not\vdash \Psi$ y también $\not\vdash \neg \Psi$, ambos por contradicción. Empecemos:

- Suponemos primero que $\vdash \Psi$.

Entonces existe una derivación v de Ψ . Sea $m = \hat{\Gamma}(v)$ el número de Gödel de la derivación v . Dado que Ψ es de la forma $\psi_j(\overline{n_j})$ tenemos que $(m, n_j) \in R_{\text{gdl}}$. Por (10.1) tenemos

$$\vdash \varphi_{\text{gdl}}(\overline{m}, \overline{n_j}) \quad (10.4)$$

Por otro lado, nuestra hipótesis $\vdash \Psi$ significa

$$\vdash \forall \mathbf{y} \neg \varphi_{\text{gdl}}(\mathbf{y}, \overline{n_j})$$

De ahí se deriva al poner \mathbf{y} igual a \overline{m} :

$$\vdash \neg \varphi_{\text{gdl}}(\overline{m}, \overline{n_j}) \quad (10.5)$$

Así que encontramos en (10.4) y (10.5) una fórmula que ella y su negación se puede derivar. Esto contradice la consistencia. Por ello nuestra hipótesis fue falsa y concluimos $\not\vdash \Psi$.

- Suponemos ahora que $\vdash \neg \Psi$.

Dado que el sistema PA es consistente, no es posible que se tiene $\vdash \Psi$. Así que $\not\vdash \Psi$. Esto quiere decir que para cada número $m \in \mathbb{N}$ se tiene: m no es el número de Gödel de una derivación de $\psi_j(\overline{n_j})$. Por ello $(m, n_j) \notin R_{\text{gdl}}$ para todo $m \in \mathbb{N}$.

Por (10.5) sigue que

$$\vdash \neg \varphi_{\text{gdl}}(\overline{m}, \overline{n_j}) \quad \text{para todo } m \in \mathbb{N} \quad (10.6)$$

Lamentablemente esto *no* implica que la fórmula $\forall \mathbf{y} (\neg \varphi_{\text{gdl}}(\mathbf{y}, \overline{n_j}))$ sea derivable. Por ello GÖDEL tuvo que asumir que el sistema PA es ω -consistente, es decir que se cumple:

$$\text{Si } \vdash \zeta(\overline{m}) \text{ para todo } m \in \mathbb{N} \text{ entonces } \vdash \forall \mathbf{y} (\zeta(\mathbf{y})).$$

En palabras: si para cada número natural m hay una derivación de $\zeta(\overline{m})$ entonces también existe una derivación de la fórmula $\forall \mathbf{y} \zeta(\mathbf{y})$.

Bajo la hipótesis de la ω -consistencia de PA se puede concluir de (10.6) que

$$\vdash \forall \mathbf{y} \neg \varphi_{\text{gdl}}(\mathbf{y}, \overline{n_j}) \quad (10.7)$$

Una comparación con la definición (10.3) muestra que (10.7) afirma

$$\vdash \psi_j(\overline{n_j}), \quad (10.8)$$

es decir $\vdash \Psi$. Esto, junto con la hipótesis $\vdash \neg\Psi$, contradice con la consistencia de PA. Por ello la hipótesis fue errónea. Concluimos $\not\vdash \neg\Psi$.

Esto concluye la demostración: Ψ es una fórmula tal que $\not\vdash \Psi$ y $\not\vdash \neg\Psi$. Por ello el sistema PA no es completo por negación.

Es importante mencionar que en ningún momento usamos la interpretación estándar. La argumentación se desarrolla a nivel sintáctica. La hipótesis adicional de la ω -consistencia es algo molesto. Afortunadamente JOHN BARKLEY ROSSER pudo – con una modificación que ahora se conoce como *el truco de ROSSER* – ajustar la construcción y la prueba para que esta hipótesis no se requiere. Omitimos aquí la modificación, porque el argumento original muestra mejor la autoreferencia y la demostración modificada es más técnica.

El resultado puede parecer técnico y sin importancia, pero esto es falso. El resultado desató un terremoto que derrumbó a cenizas el sueño de poder formalizar las matemáticas. Además estimuló a otros matemáticos a investigar a fondo las limitaciones de las matemáticas.

10.7.5. Un bosquejo de la demostración del resultado principal de TURING

ALAN TURING se dedicó a estudiar la decidibilidad de sistemas formales. Recordamos que un sistema es *decidible*, si es posible encontrar un método que decide en un número finito de pasos si una fórmula dada es derivable o no.

El problema que enfrentó TURING es que de antemano no es nada claro qué significa “un método”. Parar poder investigar la decidibilidad, TURING tuvo primero que aclarar el concepto de “algoritmo”. La propuesta de TURING es un concepto que hoy se llama en su honor *máquina de Turing*. Como antecedentes de máquinas existían máquinas a escribir y también se controlaban telares mecánicos con cintas perforadas para producir patrones en la tela. Más allá es especulación lo que le llevó al joven TURING a concebir una idea así de sencilla y poderosa.

Una *máquina de Turing* consta de un conjunto finito de *estados* y puede cambiar de estado. En cada momento uno de estos estados es el actual. Hay uno en el cual inicia y uno en el que termina. Además es capaz de leer y escribir sobre una cinta infinita una *letra* a la vez. Solo hay un número finito de letras y en la cinta debe haber una información finita. Esto es, toda la cinta está llena de la misma letra, salvo un número finito de posiciones. Aquí siempre tomaremos 0 como esta letra que juega un papel especial.

El funcionamiento tienen que ser fijo: dado la letra que lee y el estado actual, escribirá un letra (que puede ser la misma que leyó), activa el siguiente estado y mueve la cabeza de lectoescritura por una posición a la izquierda o a la derecha. Esto indica también el funcionamiento de la máquina de Turing: trabaja en turnos.

Un poco más formal: tenemos un conjunto finito de letras λ_j que conforma el *alfabeto* A :

$$A = \{\lambda_1, \dots, \lambda_m\}$$

y un conjunto finito E de estados ε_i :

$$E = \{\varepsilon_1, \dots, \varepsilon_n\}$$

y una lista de instrucciones de la forma

estado actual, letra leída | nuevo estado, letra que escribe, movimiento de la cabeza

una instrucción por cada par $(\varepsilon_i, \lambda_j)$. En otras palabras, se tiene una función

$$E \times A \longrightarrow E \times A \times M,$$

donde $M = \{\leftarrow, \rightarrow\}$ es el conjunto de posibles movimientos de la cabeza.

La siguiente lista define una máquina de Turing:

Alfabeto: $A = \{0, 1\}$
Estados: $E = \{a, b, c\}$, estado inicial: a , estado final: c
Instrucciones: $(a, 0) \mapsto (b, 1, \rightarrow)$,
 $(a, 1) \mapsto (c, 1, \rightarrow)$,
 $(b, 0) \mapsto (a, 1, \leftarrow)$,
 $(b, 1) \mapsto (a, 0, \leftarrow)$.

Es más fácil visualizar la máquina en un esquema como en la siguiente ilustración. Los estados

Ahora la dejamos andar: inicia en el estado a y le la letra 0. Entonces se cambiará al estado b , escribirá la letra 1 y moverá la cabeza a la derecha. La situación se muestra en la segunda fila de la siguiente ilustración. Ahí mismo se ve también cómo se desarrollará el proceso.

El último renglón muestra que se llegó al estado c que es el estado final. Por ello la máquina de Turing detendrá. Habrá concluido “su cálculo”.

No parece ser una máquina muy útil y ciertamente no lo es, pero el concepto de máquina de Turing es tremendamente versátil. Es a tal grado versátil, que ahora se toma como equivalente a ‘algoritmo’. Más preciso: se piensa que toda función de los número naturales que intuitivamente se puede algoritmizar, se puede calcular con una máquina de Turing. Esto es el contenido de la *Tesis de CHURCH*. Esta tesis no se puede demostrar ya que “intuitivamente algoritmizable” no es un término bien definido.

Aún TURING en su primer trabajo se dio cuenta de que bastaba con una sola de sus máquinas, una máquina *universal*. Esta podría *interpretar* el funcionamiento de cualquier otra. Veamos el argumento un poco más a detalle. El primer paso consiste en reducir el alfabeto de una máquina a $\{0, 1\}$. Más preciso: para cualquier máquina de Turing existe una con el alfabeto $\{0, 1\}$ que logra calcular exactamente lo mismo.

Veamos esto en un ejemplo.

Alfabeto: $A = \{0, C, D, E\}$
Estados: $E = \{a, b, c\}$, estado inicial: a , estado final: c
Instrucciones: $(a, 0) \mapsto (a, C, \rightarrow)$,
 $(a, 1) \mapsto (c, 1, \rightarrow)$,
 $(b, 0) \mapsto (a, 1, \leftarrow)$,
 $(b, C) \mapsto (a, D, \leftarrow)$.

Supongamos que la máquina tiene sólo dos estados: a , que es el inicial y b que es el final. El alfabeto es $\{0, C, D, E\}$. Ahora enumeramos en binario estas letras: representamos 0 como 00, C por 01, D por 10 y E por 11.

El argumento no es muy complicado y contiene dos pasos. Primero, el alfabeto siempre se puede reducir a $\{0, 1\}$: si el alfabeto tiene más letras, las enumeramos en el sistema binario y luego ampliamos el número de estados. Segundo, toda máquina con el alfabeto $\{0, 1\}$ se puede codificar de una manera bien definida usando un alfabeto con varios marcadores en la cinta

, es decir si tenemos una máquina M con el alfabeto A se puede construir una máquina máquina universal

el problema de detenerse

implicaciones

10.8. La Computación como rama de las Matemáticas

De Turing a Von Neuman.

Máquinas para calcular.

El desarrollo de las computadoras y la expansión de la capacidad para calcular y para desarrollar y programar algoritmos complejos.

Tanto el almacenamiento como la transmisión de la información, requieren que ésta se codifique de manera adecuada. Así ha sido desde el inicio de la civilización. El lenguaje hablado es un método de codificar información, y el escrito también. El invento de la imprenta hizo que la información, codificada en lenguaje escrito, pudiera reproducirse masivamente y de manera bastante económica, lo cual tuvo una gran influencia en el desarrollo y distribución del conocimiento a partir del siglo XV. Posteriormente, el hombre ha aprendido a codificar información de tal manera, que ésta pueda ser recuperada y procesada por máquinas. Al principio se trataba de patrones para los telares, luego datos de nóminas y actualmente prácticamente toda la información se puede codificar. No sólo la información numérica o textual, sino que también las imágenes y el sonido se codifican para guardarse en sistemas de almacenamiento digital, de forma que pueda ser recuperada por máquinas como los reproductores de mp3, discos DVD y Blu-ray, y ser transmitida a través de las redes de comunicación digital.

Es importante conocer estos hechos y las maneras en que se codifica la información en la actualidad. En la era digital toda la información se codifica, en última instancia, en el sistema binario. ya que

ésta se almacena en memorias de silicio que tienen la propiedad de poder orientar magnéticamente algunas de sus moléculas en una de dos posibles direcciones, lo cual permite almacenar la información equivalente a un bit, es decir uno de dos estados.

Por una convención de orden práctico se suelen utilizar grupos de bits como los bytes, consistentes en 8 bits, o pares de bytes que permiten almacenar hasta $256 = 2^8$ o $65536 = 2^{16}$ posibilidades, respectivamente. Esto permite codificar texto de manera que cada letra, dígito o símbolo de puntuación quede representado por un byte o un par de bytes. La codificación de números reales se hace de manera aproximada utilizando casi siempre 8 bytes (64 bits) para cada número real. En este último caso, como en muchos otros, la codificación no es perfecta, puede haber pérdida, que en este caso es de precisión. Cuando la información se codifica de esta manera se dice que se *codifica en formato digital*.

El formato digital permite codificar todo tipo de información, por ejemplo la música de un CD se concibe como un desplazamiento (de la membrana auditiva por ejemplo) que varía en el tiempo. Esto se *digitaliza*, primero usando las $65536 = 2^{16}$ posibilidades de un par de bytes para representar el desplazamiento, y luego se toman 44,000 muestras del desplazamiento cada segundo. Lo mismo se hace con las imágenes, las cuales se conciben como funciones definidas en una red de pixeles con tres o cuatro valores que son las cantidades de rojo, verde, azul y tal vez opacidad, de cada pixel. Cada uno de estos cuatro valores se representa con un byte. Por ello, la imagen en una pantalla de 1920×1080 pixeles (que se conoce como full HD) puede tener más de 8 millones de bytes o 64 millones de bits.

Debido a que el almacenamiento y la transmisión de la información tienen un costo asociado, es importante también poder comprimir la información, para lo cual se han desarrollado métodos generales de compresión de archivos como el famoso formato *.zip*.

Tanto el sonido como las imágenes requieren cantidades exorbitantes de información para codificarse y por eso se han desarrollado métodos especiales de compresión que utilizan ideas matemáticas de aproximación de funciones (concretamente el análisis de Fourier). Así es posible codificar este tipo de información, con cierta pérdida, pero reduciendo considerablemente la cantidad de almacenamiento requerido. Este ahorro también es importante para la transmisión. Sin estos métodos no existiría hoy en día ni la telefonía móvil, ni la música en mp3, ni las películas en DVD y Blu-ray, o serían unas 10 veces más costosas. Además, sería casi imposible transmitir este tipo de información a través de las redes de comunicación, excepto entre puntos con conexiones de mucho mayor ancho de banda, como las que tenemos hoy en día.

Hay otro tipo de codificación que sirve para esconder la información. El proceso de codificar la información con el propósito de esconderla se llama encriptación. La idea es natural y se ha usado mucho. Hace más de 2000 años Julio César ya usaba mensajes encriptados para comunicarse con sus oficiales. Durante la Segunda Guerra Mundial, gran parte de la información secreta se transmitía por radio o telégrafo, y era fácil de interceptar si no se enviaba encriptada. Por tal motivo se desarrollaron sofisticados sistemas de encriptación, que tuvieron gran protagonismo y dieron trabajo a muchos investigadores matemáticos como Alan Turing, que se dedicaron a romper los códigos de encriptación del enemigo.

Adquirir conciencia de que la codificación es un proceso matemático del que hay que conocer sus fundamentos para poder entender el mundo de información digital que habita.

Los principios básicos de la computación distinguiendo con claridad los conceptos de hardware y software, lo que es un procesador, la memoria RAM y la que se usa para almacenar archivos.

Nociones básicas de un sistema operativo y destreza para aprovecharlo, en particular para el almacenamiento y recuperación de archivos.

Nociones básicas de programación: tipos y estructuras de datos, y los elementos de control de flujo como las condicionales, repetición y recursión, usando algún lenguaje de programación popular y de uso general, por ejemplo, JavaScript. (No se recomienda recurrir a lenguajes especiales pues el objetivo es que adquiera una herramienta de programación que pueda serle útil en el futuro.)

Tanto la idea de estructuras de datos como los elementos básicos de control de flujo pueden aprenderse programando tres o cuatro ejemplos sencillos como: calcular el factorial de un número, calcular el máximo común divisor de dos números, ordenar un conjunto de datos, y programar dibujos a partir de elementos simples como segmentos y arcos.

La complejidad de algoritmos.

Generadores aleatorios.

Popularización y refinamiento de la computadora y los lenguajes de programación.

Demostraciones humanas vs demostraciones mecánicas. Teorema de los 4 colores.

10.9. El papel de las matemáticas en la Física

La mecánica cuántica y la Teoría de la relatividad.

La idea de un espacio físico de más de tres dimensiones resulta chocante a la mente humana. Nuestra experiencia cotidiana dice que nos movemos en un espacio de 3 dimensiones.

Capítulo 11

La perspectiva actual

- 11.1. Introducción
- 11.2. Irrupción de la computadora y el Internet
- 11.3. ¿Revaloración del Cálculo?
- 11.4. Ejemplos de usos de las matemáticas contemporáneas
- 11.5. Usos actuales de las matemáticas tradicionales
- 11.6. ¿Hacia dónde van las matemáticas?
- 11.7. Los problemas del milenio
- 11.8. ¿Y los fundamentos?

Índice de Referencias de las Figuras

Índice de Personas

- Čech, Eduard, 249
- Abraham Bar Hiyya, 149
- Al-Juarismi, 9
- al-Juarizmi, **200**, 200–202, **215**, 215, 216
- al-Saffar, Ibn, 187
- al-Tusi, Nasir al-Din, **178**, 180
- Alberti, Leon Battista, 66
- Anaxágoras, 161
- Anaximandro, 161
- Apolonio, 242
- Apolonio de Pérgamo, 47
- Apolonio de Perga, 9, 41, 106, 110, 241, 242
- Apolonio de Perge, 205
- Aristóteles, 25, 42, 166
- Aristarco de Samos, **161**, 162–164, 175
- Arquímedes de Siracusa, 9, 41, 47, 121, 144, 145, 147, 148, 150, 152, 160, 177, 197, 198, 254, 258
- Arquimedes de Siracura, 135
- Arquitas de Tarento, 28
- Aryabhata, 177
- Bürgi, Jost, 190
- Bézout, Étienne, 228
- Barrow, Isaac, 144, 258
- Beltrami, Eugenio, **77**
- Bernoulli, Jacob, 11
- Bernoulli, Jakob, 121
- Bernoulli, Johann, 11, 121
- Bernoulli, Johannes, 273, 274
- Bolyai, Farkas, **75**
- Bolyai, Janós, 76, 77
- Bolzano, Bernard, 277, 278
- Brahe, Tycho, 189
- Brunelleschi, Filippo, 65
- Cantor, Georg, 245–247, 278
- Cardano, Gerolamo, 202, 241
- Cardano, Girolamo, 218, 219, 221
- Cassini, 190
- Catalan, Eugène Charles, 31
- Cauchy, Augustin Louis, 11, 31, 278
- Cauchy, Augustin-Louis, 229
- Cavalieri, Bonaventura, 158, 258
- Cayley, Arthur, 229
- Claudio Ptolomeo, 9
- Copérnico, Nicolás, 161, 181, 279
- Cramer, Gabriel, **227**, 228
- da Vinci, Leonardo, 67
- Dandelin, Germinal Pierre, 59
- Dandelin, Germinal Pierre , 115
- David Hilbert, 50
- Dedekind, Richard, 11
- Dehn, Max, 139
- Del Ferro, Scipione, 241
- del Ferro, Scipione, **217**, 217, 219
- del Monte, Guidobaldo, 71
- Desargues, Girard, 57, 71, 73
- Descartes, Réne, 213, 214, 241, 242, 258
- Descartes, René, 222
- Descartes, René, 10, 205
- di Pisa, Leonardo, 215, 216

- Diofanto de Alejandría, 9, 197, 198, 206, 208, 209, 211
 Dodgson, Charles Lutwidge, 229
 Durero, Alberto, 67

 Eratóstenes, 162, 165, 166
 Euclides, 284
 Euclides de Alejandría, 8, 27–29, 40–42, 45, 49, 50, 74, 76, 123, 135, 148, 178, 195–197, 207
 Eudoxio de Cnidos, 40, 47
 Eudoxio de Samos, 134
 Eudoxo de Cnido, 9, 24, 121, 126, 136, 253, 258
 Euler, Leonhard, 11, 30, 31, 121, 273–275
 Everest, George, 192

 Fermat, Pierre, 30, 31, 205, 208, 212, 213, 241, 242, 258, 281
 Fermat, Pierre de, 10
 Ferrari, Ludovico, 219, 241
 Ferro, Scipione del, 241
 Fibonacci, 202, 215, 216
 Fior, Antonio Maria, 217, 218
 Fontana, Niccoló, **217**, 219
 Francesca, Piero della, 67
 Frisius, Gemma, 187

 Gödel, Kurt, 286, 287, 291
 Galilei, Galileo, 25, 37, 162, 166, 167, 258
 Gauss, Carl Freidrich, 229
 Gauss, Carl Friedrich, 31, 54, 76, 191, 227, 281, 282
 Giotto di Bondone, 65
 Goldbach, Christian, 31
 Gregory, David, 144

 Hamilton, Sir William Rowan, 11
 Hausdorff, Felix, 250
 Herón de Alexandria, 233
 Hilbert, David, 11, 86, 90, 138, 139, 285
 Hiparco de Nicea, **173**, 173, 174, 180
 Hipatia, 9
 Horner, William George, 182

 Ibn al-Saffar, 187
 Jayam, Omar, 9, **203**, 203–205, 214, 215
 Jordan, Camille, 285

 Kepler, Johannes, 25, 110, 190, 242, 258, 279
 Klein, Felix, 77
 Koch, Helge, 251
 Kolmogorov, Andrei Nikolajewitsch, 282
 Kurt Gödel, 52

 Lagrange, Joseph Louis, 31
 Lambert, Johann Heinrich, **76**, 157
 Lambton, William, 191, 192
 Laplace, Pierre-Simon, 228
 Lebesgue, Henri, 249
 Leibniz, Gottfried Wilhelm, 121, 258, 260
 Leibniz, Gottfried Willhelm, 228
 Leonardo de Pisa, Fibonacci, 10, 255
 Lindemann, Carl Louis Ferdinand von, 157
 Lobachevski, Nikolai, 76, 77

 Menelao de Alejandría, 171, 172
 Menelao de Alejandría (70 - 140), 9
 Mersenne, Marin, 31, 212
 Mihăilescu, Preda, 31
 Monte, Guidobaldo del, 71

 Napier, John, 190
 Nave, Hannibal, 217
 Netto, Eugen, 246, 247
 Newton, Isaac, 121, 152, 242, 258, 279
 Niven, Ivan, 157
 Nonius, *véase* Nunes, Pedro
 Nunes, Pedro, 188, 189

 Oresme, Nicole, 258

 Pacioli, Luca, 67, 217
 Pappus de Alejandría, 9, 54, 59
 Pascal, Blaise, 59, 281
 Peano, Giuseppe, 247, 248
 Phaidos, 27
 Pitágoras, 8

- Pitágoras de Samos, 21, **25**, 26, 27, 37, 38, 161
- Platón, 26, 166
- Poincaré, Henri, 77
- Polignac, Alphonse de, 31
- Proclo, 74
- Ptolomeo, Claudio, 22, 25, 41, 147, 171, 174–177, 180, 185
- Raffael Sanzio, 65
- Ramsden, Jesse, 191
- Regiomontano, 206
- Rheticus, Georg Joachim, **181**
- Riemann, Bernhard, 32, 243, 246
- Rosser, John, 291
- Saccheri, Giovanni, 76
- Savasorda, *véase* Abraham bar Hiia
- Sikdar, Radhanath, 192
- Snell, Willebrord van Roijen, *véase* Snellius
- Snellius, 188
- Stiefel, Michael, 221
- Sylvester, James Joseph, 229
- Tales de Mileto, 25, 37, 38, 41, 42, **161**
- Tartaglia, 241
- Tartaglia, Niccoló, 219
- Tartaglia, Niccolò, **217**, 218, 219
- Teón de Alejandría, 9
- Thymarides, 27
- Torricelli, Evangelista, 258
- Turing, Alan, 286, 287, 295
- Vandermonde, Alexandre-Théophile, 228
- Vernier, Pierre, 189
- Viète, François, 10, 106
- Vinci, Leonardo da, 67
- von Lindemann, Carl Louis Ferdinand, 157
- von Staudt, Karl Georg Christian, 54
- Wallingford, Richard of, 180
- Weierstrass, Karl, 11, 278
- Weiserstrass, Karl, 277
- Wiles, Andrew, 31