MONOGRAFÍAS
DE LA
REAL ACADEMIA
DE CIENCIAS
Exactas
Físicas
Químicas y
Naturales
DE
ZARAGOZA

N.º 26

Problemas del Milenio

Luis Joaquín BOYA (Editor)

2004

Índice

Los Problemas Matemáticos del Milenio

En una conferencia pública en París el 24 de Mayo del año 2000 el Clay Mathematics Institute de Boston (USA) anunció siete premios de un millón de dólares cada uno a quienes resolviesen, a satisfacción de la comunidad matemática internacional, siete célebres problemas matemáticos que permanecían sin solución en esas fechas y, que a juicio de un selecto comité de profesionales, estaban entre los más difíciles e importantes de la matemática en ese momento. En el comité figuraban Arthur Jaffe (Harvard), presidente que fué de la American Mathematical Society, y actualmente presidente-fundador del Clay M. Institute, y los medalla Fields, Michael Atiyah (Cambridge), Edward Witten (Princeton) y Alain Connes (París). Entre los proponentes de problemas concretos figuraban, además, los conocidos matemáticos Enrico Bompieri, John Milnor y Andrew Wiles. Se hizo coincidir el anuncio con el centenario de la presentación de los "Problemas de Hilbert" a que nos referiremos enseguida.

En diversos lugares del mundo, y en particular en la Universidad de Texas en Austin, se celebró el acontecimiento con diversas conferencias por especialistas sobre todos estos problemas. Las conferencias de Texas pueden consultarse en video *on line* en la dirección: `http://www.claymath.org/annual_meeting/2000_Millennium_Event/Video/`

En una reunión de la Real Academia de Ciencias de Zaragoza en octubre de 2003 se acordó que la Academia organizase unas conferencias donde especialistas nacionales explicarían al público científico universitario zaragozano el significado de esos problemas y el estado actual de su posible solución; a esos efectos la Sección de Matemáticas de la Facultad de Ciencias de la Universidad colaboró activamente con la Academia en la selección de los conferenciantes y en el acceso a los mismos. Las siete Conferencias tuvieron lugar en la Facultad de Ciencias de Zaragoza en el Otoño/Invierno de 2003/04 y contaron con una gran asistencia de público. Se pidió a los conferenciantes que preparasen una versión escrita de sus intervenciones, que aparecería como una edición especial de la Revista de la Academia, y éste es el número monográfico que ahora se presenta.

En conversaciones con los profesores de la Sección de Matemáticas se acordó dejar de lado los dos problemas más esotéricos propuestos por el Clay M. Institute, y reemplazarlos por otros dos problemas importantes, que habían sido resueltos recientemente, y que tenían quizá más interés para el estudioso actual. En estas Actas se incluyen también naturalmente exposiciones de estos dos problemas.

El anuncio del Clay M. Institute en París en el año 2000 fue en conmemoración de los 23 célebres problemas propuestos por el gran matemático alemán David Hilbert con motivo del Segundo Congreso Internacional de Matemáticas (París, agosto 1900); él los presentó como un reto para la matemática del siglo entrante y, en efecto, la mayor parte de esos problemas se han ido resolviendo a lo largo del pasado siglo veinte.

El primer International Congress of Mathematics (ICM) había tenido lugar en Zürich en 1897; actualmente las reuniones del ICM tienen lugar cada cuatro años; el congreso del año 2002 tuvo lugar en Beijing (China), y el de 2006 está previsto en Madrid, siendo la primera vez que este congreso tenga lugar en España.

Aunque no es cuestión de comentar en detalle los problemas de Hilbert, digamos que en su lectura de presentación en Paris figuraron solamente diez, pero que en la redacción escrita subsiguiente aparecieron 23; hay otro problema que Hilbert había comentado en París pero que no figuró en la numeración final (la conjetura de Fermat, véase luego) y se habla también de un "24" problema de Hilbert. Casi todos estos problemas han sido esencialmente resueltos en el siglo que ha transcurido, como hemos dicho, alguno inmediatamente (M. Dehn resolvió ya en 1905 el Problema Tres de Hilbert, "Sobre la congruencia de tetraedros"), otros estaban enunciados de forma insuficientemente precisa ("Axiomática de la Física", Problema Seis (H.)), pero uno destaca por encima de todos, porque no sólo arranca de 1859, sino que sigue aun figurando como asignatura pendiente fundamental ¡al comenzar el año 2005! Es, como el lector habrá quizá adivinado, el Problema Ocho (H.): la Hipótesis de Riemann (sobre la localización de los ceros complejos de la función $\zeta(s)$. Naturalmente, ese problema figura también como Problema Uno[1] en la lista del Clay M. Institute, ha sido incorporado por indicación de E. Bompieri, y es debidamente expuesto en estas Actas por una especialista zaragozana, Catalina Calderón. Si la conjetura de Riemann es cierta, se dispone de una fórmula asintóticamente exacta para la ley de distribución de los números primos, resultado que va bastante más allá del teorema de los números primos de J. Hadamard y C. de la Vallée-Poussin (1896).

Comentaremos ahora brevemente las alteraciones, es decir, los dos problemas Clay no incluidos, y el resto de los mismos.

Problema Seis (Clay M. Institute). "La conjetura de Birch y de Swinnerton-Dyer". La conjetura se refiere a los puntos racionales de ciertas curvas elípticas. Tiene relación con la función Zeta de Riemann, con el último teorema de Fermat (del que se habla enseguida), etc.

[1]Seguimos en esta Introducción el orden de numeración de los Problemas Clay propuesto por K. DEVLIN en "The Millennium Problems" (ver la Bibliografía al final de esta Introducción), que nos parece más adecuada que la ordenación original establecida, un poco erráticamente, por el Clay M. Institute.

Problema Siete (Clay M. Institute). "La conjetura de Hodge". El lector conocerá la figura del matematico inglés W. Hodge siquiera sea sólo por el operador "estrella" de Hodge, dual de la diferencial exterior en geometría diferencial, o el diamante de Hodge, ordenación de los números de Betti dobles, en geometría algebraica. He aquí el enunciado preciso de la conjetura, en el terso lenguaje de las matemáticas:

> *Toda forma diferencial armónica (dentro de un cierto tipo) en una variedad algebraica proyectiva no singular es una combinación racional de clases de cohomología de ciclos algebraicos.*

Completamos ahora la lista de los problemas tratados. El Problema Dos (Clay M. Ins.) lleva por título "La teoría de Yang-Mills y la Hipótesis de Masa no Nula"; es un tema genuino y capital de Física Teórica, propuesto por Jaffe y Witten, y desarrollado aquí por Manuel Asorey, del Departamento de Física Teórica de la Universidad de Zaragoza. El Problema Tres (Clay M. Ins.) es el único problema que hace referencia directa a los ordenadores: "El problema P versus el NP" y fue desarrollado por Elvira Maldonado, del Centro Politécnico Superior de la Universidad de Zaragoza. La motivación física domina también el Problema Cuatro (Clay M. Ins.): "Las Ecuaciones de Navier-Stokes", cuya brillante exposición es debida a Juan Luis Vázquez, de la Universidad Autónoma de Madrid.

La conferencia final en Zaragoza corrió a cargo de María Teresa Lozano, de nuestra Universidad y Académica, una especialista mundial en la Conjetura de Poincaré, el Problema Cinco (Clay M. Ins.). El tópico es especialmente caliente por las pretensiones del ruso Gregory Perelman, que parece haber resuelto la conjetura en sentido positivo, aunque aún no hay un veredicto definitivo por parte de la comunidad matemática. La conjetura de Poincaré dice que si una variedad cerrada de tres dimensiones es simplemente conexa, es homeomorfa a la esfera S^3.

Es de destacar que la técnica de resolución del ruso, originalmente debida a R. Hamilton (1982), está basada en la ecuación del grupo de renormalización, una conocida herramienta de la teoría cuántica de campos en física (Gell-Mann; Wilson).

Y ahora viene una pequeña explicación sobre los dos problemas que han sustituído a los Problemas (Clay) Seis y Siete; a saber, la clasificación de los grupos finitos simples y el último teorema de Fermat.

Clasificación de los grupos finitos simples. Desde que Evariste Galois en 1829 (Galois reescribió varias veces su trabajo fundamental, ya que las primeras versiones se las perdieron; la última en la noche víspera de su muerte en duelo por una cocotte, el 31 de mayo de 1832) determinó la relación entre la estructura del grupo de simetría $G \subset S_n$ del

conjunto de las raíces de un polinomio $P_n(z)$ y la solubilidad del mismo por radicales, la estructura de los grupos finitos ha sido estudiada a fondo por legiones de matemáticos, comenzando por C. Jordan hacia 1870 y siguiendo por F. Klein, G. Burnside y otros aun en el s. XIX. En particular, se conocen de antiguo dos familias infinitas de grupos finitos simples, a saber los grupos cíclicos de orden primo $\mathbb{Z}_p$ (se anima al lector a que compruebe que son los únicos grupos Abelianos finitos simples), y el grupo de las permutaciones pares (o grupo alternante A_n cuando $n > 4$) cuyo orden es $n!/2$, que es el resultado de Galois.

Tras más de un siglo, y la colaboración de cientos de matemáticos, el resultado final, es decir, la lista de todos los grupos finitos simples se obtuvo completa hacia 1983, como el magnífico relato de Javier Otal (Universidad de Zaragoza) explica en detalle. Hay cuatro familias infinitas y 26 grupos esporádicos, de momento muy enigmáticos. Naturalmente, hay uno máximo, el llamado Monstruo, M o el Gigante Amigo (a veces F_1) de Fischer-Griess; su orden es

$$\#M \approx 8 \cdot 10^{54}$$

y hay un M-módulo (= representación compleja irreducible) de dimension 196883, número que es el primer término del desarrollo en serie de Laurent de una funcion modular . . . . Todo este asunto forma parte de las conjeturas de McKay, probadas en parte por Borchers, que estamos lejos de entender aun en su totalidad.

El último teorema de Fermat. La conjetura que P. de Fermat dijo haber probado, pero que no publicó, dice que no hay solución entera a la ecuación diofántica $x^n + y^n = z^n$ para $n > 2$. No podía faltar en nuestra colección, por la gran difusión que ha tenido en los medios la demostración encontrada por A. Wiles en 1993, que se probó ser incompleta, privando así a su autor de la medalla Fields (que exige la edad del recipendario por debajo de los 40), pero que fue completada poco después por él mismo en colaboración con R. Taylor (1996). La charla de Fernando Montaner (Universdiad de Zaragoza) sobre la conjetura de Fermat no ha sido posible incluirla en esta colección, y aparecerá verosimilmente como artículo normal en otro número de la Revista de la Academia.

Queremos destacar, para terminar, la gran relación con la Física que tienen casi todos estos problemas del milenio. Por ejemplo, los problemas Cuatro (Navier-Stokes) y el Dos (Yang-Mills) arrancan directamente de la física. El Problema Tres (P vs. NP) trata sobre ordenadores, con gran aplicación práctica en criptografía. Otros, como el Uno (Riemann) y el Cinco (Poincaré) se relacionan directamente con problemas físicos muy importantes, como la distribución estadística de los niveles energéticos de los núcleos atómicos complejos (Wigner-Dyson) y la integración funcional sobre membranas (Polyakov), respectivamente. La Física ha sido siempre fuente de inspiración para las matemáticas, y no olvidemos que los tres matemáticos más grandes de la historia, según consenso, que son Arquímedes, Newton y Gauss, hicieron grandes contribuciones a la Física.

Agradecimientos

Quiero agradecer aquí, primero, el gran esfuerzo realizado por los conferenciantes, incluyendo la redacción de sus charlas. La ayuda específica recibida, entre otros, de José F. Cariñena y Julio Abad de Física Teórica, así como de Mariano Gasca de Matemática Aplicada y Jesús Bastero y Alberto Elduque por el Departamento de Matemáticas. El Presidente de la Academia, Horacio Marco Moll, ha acogido con gran entusiasmo la iniciativa y ha inaugurado y clasurado el ciclo. La ayuda institucional de la Facultad de Ciencias, personificada por el Decano Antonio Elipe es difícil de reconocer suficientemente. A Elipe es de agradecer, además, la cuidadosa labor de edición de los artículos de esta colección, y a Cariñena la revisión de esta Introducción.

Bibliografía.-

De la extensa biblografía sobre los Problemas de Hilbert y del Clay M. Institute seleccionamos lo siguiente:

1. Benjamin H. YANDELL. "The Honors Class: Hilbert's Problems and Their Solvers". A.K. Peters ed. Natick (MA), USA 2002

2. Keith DEVLIN. "The Millennium Problems: The Seven Greatest Unsolved Mathematical Puzzles of Our Time". Basic Boooks, N.Y. 2002

3. Jeremy J. Gray. "El reto de Hilbert: Los 23 problemas que desafiaron a la matemática". Crítica, Barcelona 2003

4. El website del Clay M. Insitute con el anuncio y descripción particular de cada problema es `www.claymath.org`

   La literatura sobre la conjetura de Riemann es enorme; destaquemos las siguientes referencias recientes:

5. "Riemann Selecta". Edición y estudio por José FERREIRÓS. Edición del CSIC, Madrid 2001. Contiene el artículo fundamental de Riemann de 19859 sobre la distribución de números primos en alemán y en español.

6. Marcus du SAUTOY. "The Music of the Primes: Searching to Solve the Greatest Mystery in Mathematics". Harper, N.Y. 2003

7. Karl SABBAGH. "The Riemann Hypothesis: The Greatest Unsolved Problem in Mathematics". Farrar, N.Y. 2002

8. John DERBYSHIRE. "Prime Obsession: Bernhard Riemann and the Greatest Unsolved Problem in Mathematics". J. Henry Press, Washington D.C., 2003.

Señalemos el libro-fuente sobre la solución de A. Wiles de la conjetura de Fermat:

9. Simon SINGH. "Fermat´s Last Theorem: The story of a Riddle that confounded the World´s Greatest Minds for 358 Years". Fourth State, London 1997.

La conjetura de Poincaré está en el punto de mira de todos los matemáticos a raíz de los trabajos del ruso Gregory Perelman, como bien describe Maite Lozano en su conferencia. Algunas referencias técnicas recientes son:

10. John MILNOR. "Towards the Poincaré conjecture and the Classification of 3-Manifolds". Notices of the Am. Math. Soc. **50**, 1226-1233 (2003).

11. Michael T. ANDERSON. "Geometrization of 3-Manifolds via the Ricci Flow". Notices of the Am. Math. Soc. **51**, 184-193 (2004)

12. John W. MORGAN. "Recent Progress on the Poincaré conjecture and the classification of 3-manifolds". Bulletin of the Am. Math. Soc. **42**, 52-78 (2005)

Zaragoza, el día de Inocentes, 2004

Luis Joaquín BOYA

Departamento de Física Teórica, Universidad de Zaragoza
Real Academia de Ciencias de Zaragoza

Monografías de la Real Academia de Ciencias de Zaragoza. **26:** 1–30, (2004).

The Riemann Hypothesis

C. Calderón

Departamento de Matemáticas. Facultad de Ciencia y Tecnología

Universidad del País Vasco. 48080 Bilbao, Spain

mtpcagac@lg.ehu.es

Abstract

Riemann proved (see [53], [37] and preferably Edwards' book [15]), that $\zeta(s)$ has an analytic continuation to the whole plane apart from a simple pole at $s = 1$. Moreover, he showed that $\zeta(s)$ satisfies the functional equation

$$\pi^{-s/2}\Gamma(s/2)\zeta(s) = \pi^{-(1-s)/2}\Gamma((1-s)/2)\zeta(1-s).$$

Riemann uses the functional equation to deduce an approximate formula for $\pi(x) = \#\{p \leq x; p \text{ prime}\}$: that is

$$\pi(x) \sim Li(x) + \sum_{n=2}^{N} \frac{\mu(n)}{n} Li(x^{\frac{1}{n}}), \quad N > \log x/\log 2.$$

In the paper (see [53]), he stated also that $\zeta(s)$ had infinitely many non trivial roots and that it seemed probable that they all have real part $1/2$.

1 Distribution of prime numbers

Among the positive integers there is a sub-class of special importance, that is, the class of primes. Some questions about primes are: a) How many prime numbers are there? and b) How are the prime numbers distributed?

The answer to the question of how many prime numbers are there is given by the theorem of Euclid (*Elements*, Book 9, Prop. 20): *There exist infinitely many prime numbers.* In the proof of Euclid, to prove the theorem, it will suffice to prove that if $\{p_1, p_2, \cdots, p_n\}$ is any finite set of primes, then we can find a prime number that is not in the set.

As usual $\mathbb{Z}$ will denote the ring of integers, $\mathbb{Q}$ the field of rational numbers, $\mathbb{R}$ the field of real numbers and $\mathbb{C}$ the field of complex numbers. Also we shall denote by $\mathbb{N}$ the set of positive (natural) integers, not including 0.

In 1737, Euler proved the existence of an infinity of primes by a new method, which shows moreover that the series $\sum_{p_n} p_n^{-1}$ is divergent. Euler's work is based on the idea of using an identity in which the primes appear on one side but not on the other. Stated formally his identity is

$$\prod_p \{1 - p^{-s}\}^{-1} = \sum_{n=1}^{\infty} n^{-s}, \quad (s > 1) \tag{1}$$

where p ranges over all prime numbers. This formula results from expanding each of the factors on the left

$$(1 - x)^{-1} = 1 + x + x^2 + \dots, \quad x = p^{-s}.$$

As their product is a sum of terms of the form $(p_1^{\alpha_1} \cdots p_r^{\alpha_r})^{-s}$, where $p_1 \neq p_2 \neq \cdots \neq p_r$, the formula (1) is deduced by using the fundamental theorem of arithmetic: *Each natural number $n > 1$ can be decomposed uniquely, up to order of the factors, as a product of prime numbers.*

From an old theorem of Nicole Oresme (1360), we have that: *The harmonic series $\sum_{n=1}^{\infty} 1/n = \zeta(1)$ is divergent.*

Function $\pi(x)$. We introduce a function $\pi(x)$ which has become universally accepted and means the number of primes not exceeding x. If $p_1 = 2, p_2 = 3, \cdots$, and p_n denote the nth prime number, then for each integer n we have $\pi(p_n) = n$. It follows from Euclid's theorem that $\pi(x) \to \infty$ as $x \to \infty$.

Legendre. Experimentally, Legendre conjectured in 1798 and again in 1808 the hypothesis that

$$\pi(x) \sim \frac{x}{\log x - A(x)}, \quad \lim_{x \to \infty} A(x) = 1,08366\dots$$

Gauss. Gauss [20], in a letter to the astronomer Enke in 1849, stated that he had found in his early years (at age 17, in 1792), that the number $\pi(x)$ of primes up to x and the integral

$$\int_2^x \frac{dt}{\log t}$$

are asymptotically equal. However, Gauss does not mention Euler's formula and he gives no analytic basis for the approximation, which he presents only on the basis of extensive computations. So, Gauss conjectured that

$$\pi(x) \sim \frac{x}{\log x}, \quad x \to \infty$$

This conjecture, now a theorem, is called the "Prime Number Theorem".

Chebyshev. The first serious work on the function $\pi(x)$ is due to the Russian mathematician Chebyshev. On May 24, 1848, Chebyshev read at the Academy of St. Petersburg his

first memoir on the distribution on prime numbers, later published in 1850. He proved, using elementary methods, that for every $\epsilon > 0$ there exists $x_0 > 0$ such that if $x > x_0$, then

$$(C' - \epsilon)\frac{x}{\log x} < \pi(x) < (C + \epsilon)\frac{x}{\log x}$$

where $C' = 2^{1/2}3^{1/3}5^{1/5}30^{-1/30} < 1$, $C = (6/5)C'$. Moreover, Chebyshev showed that if the limit

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log x}$$

exists, it must be equal to 1. He deduced also, that Legendre's approximation of $\pi(x)$ cannot be true, unless $1,08366$ is replaced by 1.

Dirichlet. In 1837, Dirichlet proved his famous theorem of the existence of infinitely many primes in any arithmetic progression $n \equiv h \pmod{k}$, with h and k positive co-primes integers. The main novelty in his proof consisted in making use of characters modulo k, that is, homomorphisms from the group $(\mathbb{Z}/k\mathbb{Z})^*$ of invertible residues $\pmod{k}$ into the multiplicative group of complex numbers of modulus 1, (see Tenenbaum [59] cap II.8, for all these facts).

Riemann. Riemann considers the zeta-function defined by the generalized harmonic series $\zeta(s) = \sum n^{-s}$, where the letter s denotes a complex variable $s = \sigma + it$ $\sigma > 1$ its real part and t its imaginary part. If $\sigma > 1$ we can represent $\zeta(s)$ by an absolutely convergent infinite product, namely

$$\zeta(s) = \prod_p \{1 - p^{-s}\}^{-1}. \tag{2}$$

The Riemann hypothesis appears in a paper over the number of primes less than a given magnitude, published in 1859 by Riemann (see-preferably- Edwards' book [15], also [53], or [37]). It is the only work that Riemann published in number theory, but most of Riemann's ideas have been incorporated in later to the work of many mathematicians.

In 1859 Dirichlet died and Riemann was appointed to the chair of mathematics at Göttingen. A few days later Riemann was elected to the Berlin Academy of Sciences. He had been proposed by Kummer, Borchardt and Weierstrass. Riemann, newly elected to the Berlin Academy of Sciences had to report on their most recent research and he sent a report on *Über die Anzahl der Prinzahlen unter einer gegebener Gröse*. Riemann style is extremely difficult, his paper is extremely condensed and in particular [53], is probably a summary of very extensive researches which he never found the time to expound at grater length. The main purpose of the paper was to give estimates for the number of primes less than a given number. In this paper he stated that $\zeta(s)$ had infinitely many non trivial roots and that it seemed probable that they all have real part 1/2. Many of

the results which Riemann obtained in this paper were later proved by Hadamard, de la Vallée-Poussin, Hardy and von Mangoldt.

2 Analytic continuation

The development of complex analysis was a central preoccupation of Riemann's and so it comes as no surprise that from the beginning Riemann considered the zeta-function as an analytic function. He first showed that $\zeta(s)$ had an analytic continuation to the complex plane as a meromorphic function which has only one singularity, a simple pole of residue 1 at $s = 1$.

Riemann derives his functional equation for $\zeta(s)$ from the representation of the gamma-function as an integral, for which Riemann still used the symbol Π

$$\Gamma(s) = \Pi(s-1) = \int_0^\infty e^{-u} u^{s-1} du. \tag{3}$$

The gamma-function defined by (3) has meromorphic continuation to all of $\mathbb{C}$, and is analytic except at $s = 0, -1, -2, \cdots, -n, \cdots$, where it has simple poles.

If n is a positive integer, we have, on writing nx for u

$$\frac{\Gamma(s)}{n^s} = \int_0^\infty e^{-nx} x^{s-1} dx.$$

Hence

$$\Gamma(s)\zeta(s) = \sum_{n=1}^{\infty} \int_0^\infty e^{-nx} x^{s-1} dx = \int_0^\infty \frac{x^{s-1} dx}{e^x - 1}, \quad Re(s) > 1$$

if the inversion of the order of summation and integration can be justified. Next he considers the integral

$$I(s) = \int_C \frac{(-x)^{s-1}}{e^x - 1} dx$$

where the contour C starts at infinity on the positive real axis, encircles the origin once in the positive direction (counterclockwise), excluding the points $\pm 2i\pi n, n \in \mathbb{N}$ and returns up the positive real axis to $+\infty$. In the many-valued function $(-x)^{s-1} = e^{(s-1)\log(-x)}$ the $\log(-x)$ is determined in such a way that it is real for negative values of x.

Thus Riemann deduced that for each $s \neq 1$

$$\zeta(s) = 2^s \pi^{s-1} \sin(\pi s/2) \Gamma(1-s) \zeta(1-s) \tag{4}$$

(see the proof in [59] pag. 139 or [60], Chap. II, §2.3). The equation (4) gives a relation between $\zeta(s)$ and $\zeta(1-s)$ which, by making use of properties of the gamma-function, can be formulated as the statement that $\Gamma(s/2)\pi^{-s/2}\zeta(s)$ remains unchanged when s is replaced by $1 - s$. Indeed, on writing $\pi n^2 x$ for u in $\Gamma(s/2)$ we have

$$\Gamma(s/2)\pi^{-s/2}\zeta(s) = \int_0^\infty \theta(x) x^{s/2-1} dx, \quad (\sigma > 1)$$

where $\theta(x) = \sum_{n=1}^{\infty} e^{-n^2\pi x}$, and using the equation

$$2\theta(x) + 1 = x^{-1/2}\,[2\theta(1/x) + 1]$$

for the theta-function of Jacobi, one obtains the relation

$$\Gamma(s/2)\pi^{-s/2}\zeta(s) = \frac{1}{s(s-1)} + \int_1^{\infty} \theta(x)(x^{(s/2)-1} + x^{-(1+s)/2})dx.$$

The last integral is convergent for all values of s, and so the formula holds, by analytic continuation, for all values of $s \neq 1$. Now the right-hand side is unchanged if s is replaced by $1 - s$. Hence we can write the functional equation (4) in the more symmetric form

$$\pi^{-s/2}\Gamma(s/2)\zeta(s) = \pi^{-(1-s)/2}\Gamma((1-s)/2)\zeta(1-s). \tag{5}$$

(see [59], [11] or [15]). The symmetry of the functional equation and the poles at $s = 0$, and $s = 1$ of $\pi^{-s/2}\Gamma(s/2)\zeta(s)$, suggests the introduction of the function

$$\xi(s) = \frac{1}{2}s(s-1)\pi^{-s/2}\Gamma(s/2)\zeta(s) \tag{6}$$

that is an entire function that non vanishing in $Re(s) > 1$. Then from (5) and (6) we have that the function $\xi(s)$ verifies the functional equation

$$\xi(s) = \xi(1-s). \tag{7}$$

it also has no zeros in $Re(s) < 0$, apart from the trivial zeros at $s = -2, -4, -6, \ldots$. Thus all the zeros have their real parts between 0 and 1.

Riemann considered $\xi(s)$ with $s = 1/2 + it$ and stated that there is a product representation

$$\xi(t) = \xi(0)\prod_{\alpha}\left(1 - \frac{t^2}{\alpha^2}\right) \tag{8}$$

with zeros α of $\xi(t)$ that correspond to the critical zeros ρ of the zeta-function. Hadamard studied ([21]) entire functions and their representations as infinite products. One consequence is that the product formula (8) is valid.

Theorem. (Riemann) *The zeta function $\zeta(s)$ is analytic in the whole complex plane except for a simple pole at $s = 1$ with residue 1. It satisfies the functional equation (4).*

Riemann hypothesis. *The non trivial zeros of the function $\zeta(s)$ have real part equal to 1/2.* In 1900 Hilbert included the resolution of Riemann hypothesis as the 8th problem of his 23 problems for mathematicians of the twentieth century to work on. In the later years of the nineteenth century several mathematicians had developed Riemann's work to the point that Hadamard and C. de la Vallée Poussin independently of one another in 1896 proved the Prime Number Theorem.

3 Riemann's explicit formula

Riemann gave an explicit formula for $\pi(x)$ in terms of the complex zeros of $\zeta(s)$. In his derivation of the formula, he uses complex integrals and the Cauchy residue theorem. The starting point is the product representation (2) of $\zeta(s)$ in $Re(s) > 1$. Thus, the relation (2) implies that $\log \zeta(s) = -\sum \log(1 - p^{-s})$. Expanding $\log(1 - p^{-s})$ and summing we obtain

$$\log \zeta(s) = \sum_p \sum_n (1/n) p^{-ns}, \quad Re(s) > 1. \tag{9}$$

With the assistance of these methods (as Riemann says), the number of prime numbers that are smaller than x can now be determined. Let $F(x)$ be equal to this number when x is not exactly equal to a prime number; but let it be greater by $1/2$ when x is a prime number, so that, for any x at which there is a jump in the value in $f(x)$

$$F(x) = \frac{F(x+0) + F(x-0)}{2}$$

and states: If in the identity (9) one now replaces p^{-ns} by $s \int_{p^n}^{\infty} x^{-s-1} dx$, $(Re(s) > 1)$ for each $n \in \mathbb{N}$, then one obtains

$$\frac{\log \zeta(s)}{s} = \int_0^{\infty} f(x) x^{-s-1} dx, \quad Re(s) > 1 \tag{10}$$

if one denotes

$$F(x) + (1/2)F(x^{1/2}) + (1/3)F(x^{1/3}) + \ldots$$

by $f(x)$. If we think of $\zeta(s)$ as given and $f(x)$ as required, then (10) represents an integral equation to be solved for $f(x)$.

He applies Fourier inversion to (10) to obtain that

$$f(x) = \frac{1}{2\pi i} \int_{a-i\infty}^{a+i\infty} \log \zeta(s) x^s \frac{ds}{s}, \quad (a > 1). \tag{11}$$

First, Riemann integrates by parts to obtain

$$f(x) = -\frac{1}{2\pi i} \cdot \frac{1}{\log x} \int_{a-i\infty}^{a+i\infty} \frac{d}{ds} \left(\frac{\log \zeta(s)}{s} \right) x^s ds, \quad (a > 1) \tag{12}$$

From the relations (6) and (8), the following formula for $\log \zeta(s)$ is obtained

$$\log \zeta(s) = \log \xi(0) + \textstyle\sum_\alpha \log \left(1 + \frac{(s-1/2)^2}{\alpha^2} \right) - \log \Gamma(s/2+1) + (s/2) \log \pi - \log(s-1). \tag{13}$$

and substituting (13) in (12) one obtains his result

$$f(x) = Li(x) - \sum_{Im\rho > 0} [Li(x^\rho) + Li(x^{1-\rho})] + \int_x^{\infty} \frac{dt}{t(t^2-1)\log t} + \log \xi(0) \tag{14}$$

which is the Riemann's formula except that, as noted by Edwards in [15], $\log \xi(0)$ equals $-\log 2$, and

$$Li(x) = \lim_{\epsilon \to 0} \left(\int_0^{1-\epsilon} \frac{dt}{\log t} + \int_{1+\epsilon}^{x} \frac{dt}{\log t} \right) = \int_2^x \frac{dt}{\log t} + 1,04 \cdots$$

To deduce the Rieman's formula for $\pi(x)$, that is, for the number of primes less than any given magnitude x, we observe that the functions $f(x)$ and $\pi(x)$ are related by the formula

$$f(x) = \pi(x) + \frac{1}{2}\pi(x^{\frac{1}{2}}) + \frac{1}{3}\pi(x^{\frac{1}{3}}) + \cdots + \frac{1}{n}\pi(x^{\frac{1}{n}}) + \cdots \tag{15}$$

But if $x^{1/n} < 2$ for any given x and n sufficiently large, then $\pi(x^{\frac{1}{n}}) = 0$. Thus the series in (15) is finite. Riemann inverts this relation (15) by means of the Möbius inversion formula (see §10.9 [15]) to obtain

$$\pi(x) = \sum_{n=1}^{\infty} \frac{\mu(n)}{n} f(x^{\frac{1}{n}}). \tag{16}$$

From (16) and (14) one obtains the Riemann's formula for $\pi(x)$

$$\pi(x) \sim Li(x) + \sum_{n=2}^{\infty} \frac{\mu(n)}{n} Li(x^{\frac{1}{n}}). \tag{17}$$

Riemann made in their paper the following assertions related to zeros of $\zeta(s)$:

(1) There are infinitely many complex zeros of $\zeta(s)$ which lie in the critical strip. Zeros of the zeta-function in the critical strip are denoted $\rho = \beta + i\gamma$. Let $T > 0$, and let $N(T)$ denote the number of zeros of the function $\zeta(s)$ in the region $0 \leq Re(s) \leq 1$, $0 < Im(s) \leq T$. That is

$$(T) = \#\{\rho = \beta + i\gamma \mid 0 \leq \beta \leq 1, 0 < \gamma \leq T\}. \tag{18}$$

According to Riemann, the number of zeros with imaginary parts between 0 and T is about

$$\frac{T}{2\pi}\log(\frac{T}{2\pi}) - \frac{T}{2\pi}.$$

(2) Riemann stated also, that the number of zeros of the zeta-function between this limits, in the critical line $Re(s) = 1/2$, is "about" the same.

(3) If ρ is a complex zero of $\zeta(s)$, then the series $\sum \mid \rho \mid^{-2}$ converges and the series $\sum \mid \rho \mid^{-1}$ diverges.

(4) The entire function $\xi(s) = s(s-1)\pi^{-(1/2)s}\Gamma(s/2)\zeta(s)$ can be written as Weierstrass' product.

(5) All complex zeros of $\zeta(s)$ lie on the critical line $Re(s) = 1/2$.

(6) The relation

$$f(x) = Li(x) - \sum_{\rho} Li(x^{\rho}) + \int_{x}^{\infty} \frac{dt}{t(t^2-1)\log t} - \log 2,\ x > 1$$

where

$$Li(x^{\rho}) = Li(e^{\rho \log x}),\ Li(e^{w}) = \int_{-\infty+iv}^{u+iv} \frac{e^z}{z} dz,\ w = u + iv,\ \sum_{\rho} Li(x^{\rho}) = \lim_{T\to+\infty} \sum_{|\rho|\leq T} Lix^{\rho}$$

hods true.

A simpler variant of his formula is

$$\psi(x) = \sum_{n\leq x} \Lambda(n) = x - \sum_{\rho} \frac{x^{\rho}}{\rho} - \log 2\pi - \frac{1}{2}\log(1 - x^{-2}),$$

valid for x not a prime power ($x \neq p^m$). Note that the Prime Number Theorem is equivalent to $\psi(x) \sim x,\ x \to \infty$ and that $|x^{\rho}| = x^{\beta}$; thus it was necessary to show that $\beta < 1$ in order to conclude of Gauss's conjecture, that is, the Prime Number Theorem.

Hypothesis (1), (3), (4) were proved by Hadamard [21]. The estimate of $N(T)$ and (6) were proved by Mangoldt [45].

4 Complex zeros of the zeta function

In order to prove the convergence of the product (8) Riemann needed to investigate the distribution of roots of $\xi(s)$, then he begins to observing that if $Re(s) > 1$, then by the Euler product, $\zeta(s) \neq 0$ (because a convergent infinite product can be zero only if one of its factors is zero). Moreover if $s = 1 + it$ we have the following result

Theorem (Hadamard-de la Vallée Poussin). *If $t \neq 0$, then $\zeta(1 + it) \neq 0$.*

As consequence, the zeta function has no zeros in the closed half plane $Re(s) \geq 1$.

If $Re(s) < 0$, then $Re(1 - s) > 1$, the right-side in the functional equation (5) is not zero, so the zeros must be exactly the poles of $\Gamma(s/2)$. Then, the zeros of $\zeta(s)$ are

1. Simple zeros at the points $s = -2, -4, -6, \cdots$, which are called the trivial zeros

2. Zeros in the critical strip consisting of the complex zeros $\rho = \beta + i\gamma$ with $0 \leq Re(\rho) \leq 1$. In 1914, Hardy proved that there are infinitely many roots ρ of $\zeta(s) = 0$ on the line $Re(s) = 1/2$. (see [24] or [11]).

Since $\zeta(\overline{s}) = \overline{\zeta(s)}$, the zeros of $\zeta(s)$ lie symmetrically with respect to the real axis, so, it suffices to consider the zeros in the upper half of the critical strip. It is common to list the complex zeros $\rho_n = \beta_n + i\gamma_n$, with $\gamma_n > 0$, in order of increasing imaginary parts as $\gamma_1 \leq \gamma_2 \leq \gamma_3 \leq \cdots$. Here zeros are repeated according to their multiplicity.

Van de Lune, te Riele and Winter [44] have determined that the first 1.500.000.001 complex zeros of $\zeta(s)$ are all simple, lie on the critical line, and have imaginary part with $0 < \gamma < 545.439.823,215$.

(i) Let $T > 0$, and let $N(T)$ denote the number of zeros of the function $\zeta(s)$ in the region $0 \leq Re(s) \leq 1$, $0 < Im(s) \leq T$. That is

$$N(T) = \#\{\rho = \beta + i\gamma \mid 0 < \beta \leq 1, 0 < \gamma \leq T\}.$$

By the functional equation and the argument principle, one obtains

$$N(T) = \frac{T}{2\pi}\log(\frac{T}{2\pi e}) + \frac{7}{8} + S(T) + O(\frac{1}{T}) \tag{19}$$

where $S(T) = \frac{1}{\pi}\arg\zeta(\frac{1}{2}+iT) \ll \log T$ with the argument obtained by continuous variation from 2 to $2 + iT$, and thence to $1/2 + iT$, along straight lines. It was conjectured by Riemann and proved by von Mangoldt. This is the so called Riemann-von Mangoldt formula, (the sign $\ll$ of I.M. Vinogradov is taken, as usual, in the sense O).

If the Riemann hypothesis (henceforth **RH** for short) is true, then it is known that

$$S(T) = \frac{1}{\pi}\arg\zeta(\frac{1}{2} + iT) \ll \frac{\log T}{\log\log T}. \tag{20}$$

(ii) Now, let $N_0(T)$ be the zero counting function

$$N_0(T) = \#\{\rho = 1/2 + i\gamma \mid 0 < \gamma \leq T\},$$

that it, $N_0(T)$ counts the number of zeros of $\zeta(s)$ in the critical line, up to height T. The **RH** is equivalent to $N(T) = N_0(T)$ for all T.

(iii) Let $N(\sigma, T)$ the function which counts the number of zeros of $\zeta(s)$ in the critical strip up to height T, and to the right of σ-line.

A method to approach Riemann's hypothesis is to estimate for any given σ, $1/2 \leq \sigma \leq 1$, the number $N(\sigma, T)$ of zeros $\rho = \beta + i\gamma$ of $\zeta(s)$, with $\sigma \leq \beta$ and $0 < t \leq T$ (where T is sufficiently large):

$$N(\sigma, T) = \#\{\rho = \beta + i\gamma \mid \beta \geq \sigma, 0 < \gamma \leq T\}$$

$1/2 \leq \sigma \leq 1$. In this case, the **RH** is equivalent to the property $N(1/2, T) = 0$ for all T.

Estimates for $N(\sigma, T)$ may be written in the form

$$N(\sigma, T) \ll T^{a(\sigma)(1-\sigma)} \log^b T, \quad b \geq 0. \tag{21}$$

In view of formula (19) one has $a(\sigma)(1-\sigma) = 1$ and $b = 1$ in (21) for $0 \leq \sigma \leq 1/2$, while $a(\sigma)(1-\sigma) \leq 1$ for $\sigma > 1/2$ and $a(\sigma)(1-\sigma)$ is non increasing.

The hypothesis $a(\sigma) \leq 2$ in (21) is known as "the density hypothesis". A. Selberg [57] has proved that

$$N(\sigma, T) \ll T^{1-\frac{1}{4}(\sigma-\frac{1}{2})} \log T \tag{22}$$

uniformly for $1/2 < \sigma \leq 1$. The **RH** is equivalent to $N(\sigma, T) = 0$ for $\sigma > 1/2$.

Gaps between zeros. The most detailed study of the zeros on the critical line involves estimates for the difference $\gamma_{n+1} - \gamma_n$ between the ordinates of consecutive zeros $s = \frac{1}{2} + i\gamma$, $\gamma \in \mathbb{R}$. For a long time, the result of Hardy and Littlewood (1918) that $\gamma_{n+1} - \gamma_n \ll \gamma_n^{1/4+\epsilon}$ was the best. Later it was superseded, with the use of finer methods by Moser, Balasubramanian, Karatzsuba, Ivic [31].

It is known that the average size of $\gamma_{n+1} - \gamma_n$ is $\sim 2\pi/\log \gamma_n$. Thus if

$$\lambda := \limsup_{n\to\infty} (\gamma_{n+1} - \gamma_n) \frac{\log \gamma_n}{2\pi}, \quad \mu := \liminf_{n\to\infty} (\gamma_{n+1} - \gamma_n) \frac{\log \gamma_n}{2\pi}$$

we have $\mu \leq 1 \leq \lambda$, and it is known (unconditionally) that $\mu < 1 < \lambda$. It is conjectured that $\mu = 0$ and $\lambda = \infty$, but assuming the generalized **RH** Conrey-Ghosh- Gonek [9] proved only that $\lambda > 2.68$. Also under the **RH** they had proved $\mu < 0.5172$.

If (20) holds, then from the estimate (19) one obtains

$$N(T + H) - N(T) > 0, \text{ for } H = C/\log\log T, C > 0, T \geq T_0.$$

Hence assuming the **RH** the following estimate holds

$$\gamma_{n+1} - \gamma_n \ll \frac{1}{\log\log \gamma_n.}$$

Instead of working with the complex zeros of $\zeta(s)$ on the critical line $Re(s) = 1/2$, it is convenient to introduce the function

$$Z(t) = \chi^{-1/2}(\frac{1}{2} + it)\zeta(\frac{1}{2} + it)$$

where $\chi(s)$ is given by $\chi(s) = 2^s \pi^{s-1} \sin(\pi s/2)\Gamma(1-s)$. As $\chi(s)\chi(1-s) = 1$ and $\overline{\Gamma(s)} = \Gamma(\overline{s})$, it follows that $\mid Z(t) \mid = \mid \zeta(\frac{1}{2} + it) \mid$, $Z(t)$ is even, and $\overline{Z(t)} = Z(t)$. Hence $Z(t)$ is real if t is real and the zeros of $Z(t)$ correspond to the zeros of $\zeta(s)$ on the critical line $Re(s) = 1/2$.

5 Consequences of the Riemann hypothesis

There are many equivalent forms for the Riemann hypothesis. The study and classification of these assertions sharpens our understanding of the problem. Writing $\Xi(t) = \xi(1/2+it)$, and as consequence of the functional equation (7), we obtain $\Xi(t) = \Xi(-t)$ which is real for real t an even function of t. Then **RH** is the assertion to all zeros of $\Xi(t)$ are reals.

Riemann obtained that $\Xi(t)$ verifies the relation

$$\Xi(t) = 4\int_1^\infty \frac{d[x^{3/2}\theta'(x)]}{dx} x^{-1/4}\cos\left(\frac{t}{2}\log x\right) dx \tag{23}$$

where $\theta(x) = \sum_{n=1}^\infty e^{-n^2\pi x}$. We can write (23) in the form

$$\Xi(t) = 2\int_0^\infty \phi(u)\cos ut\ du$$

where $\phi(u) = 2\sum_{n=1}^\infty (2n^4\pi^2 e^{9t/2} - 3n^2\pi e^{5t/2})e^{-\pi n^2 e^{2t}}$. A necessary and sufficiently condition to of zeros of $\Xi(t)$ are reals is

$$\int_{-\infty}^\infty \int_{-\infty}^\infty \phi(\alpha)\phi(\beta)e^{i(\alpha+\beta)x}e^{(\alpha-\beta)y}(\alpha-\beta)^2 d\alpha d\beta \geq 0$$

for all x, y reals (see Chap X and XIV of [60]).

A. Speiser [58], proved that **RH** is equivalent to the non-vanishing of the derivate $\zeta'(s)$ in the left-half of the critical strip $0 < \sigma < 1/2$.

The order of zeta-function on the critical line. Hardy and Littlewood proved that $\zeta(1/2+it) \ll t^{\frac{1}{4}+\epsilon}$ for every positive ϵ, and $t \geq t_0 > 0$ (since $\overline{\zeta(1/2+it)} = \zeta(1/2-it)$, t may be assumed to be positive. H. Weyl improved the bound to $t^{1/6+\epsilon}$, Bombieri-Iwaniec obtained 9/56, instead to 1/6, A. Watt improved the bound to 89/560 and M.N. Huxley ([29]) improved the bound to

$$\zeta(1/2+it) \ll t^{\frac{89}{570}+\epsilon}.$$

Lindelöf conjectured that

$$\zeta(1/2+it) \ll_\epsilon t^\epsilon, \quad t \geq t_0 > 0$$

for every positive ϵ. In 1912, Littlewood proved that the Lindelöf hypothesis is true if the **RH** is true. The converse theorem cannot be made.

For the function $S(T) = \frac{1}{\pi}\arg\zeta(\frac{1}{2}+iT)$ we have that the **LH** is equivalent to

$$\int_0^T S(t)dt = o(\log T),\ (T\to\infty)$$

while the **RH** implies much more

$$\int_0^T S(t)dt \ll \frac{\log T}{\log\log T},\ (T\to\infty).$$

The order of zeta-function on the 1-line. E.C. Titchmarsh, was proved ([60] theorems 8.5 and 8.8), that each one of the inequalities

$$|\ \zeta(1+it)\ |> A\log\log t,\quad 1/\ |\ \zeta(1+it)\ |> A\log\log t$$

is satisfies for some arbitrary large values of t, if A is a suitable constant and considers the question of how large the constant can be in the two cases.

On **RH** we have

$$e^\gamma \le \limsup_{t\to\infty}\frac{\zeta(1+it)}{\log\log t} \le 2e^\gamma$$

$$\frac{6}{\pi^2}e^\gamma \le \limsup_{t\to\infty}\frac{1/\zeta(1+it)}{\log\log t} \le \frac{12}{\pi^2}e^\gamma$$

where γ is Euler's constant, (see Titchmarsh [60] 8.9).

Riemann hypothesis and distribution of primes. We know that the zeta function was introduced as an analytic tool for studying prime numbers and some of the most important applications of the zeta functions belong to prime number theory. The equivalence between the error term in the Prime Number Theorem and the real part of the zeros of $\zeta(s)$ is as follow

$$\pi(x) = \int_2^x \frac{dt}{\log t} + O(x^{\theta+\epsilon})$$

or

$$\psi(x) = \sum_{p^m\le x}\log p = x + O(x^\theta\log^2 x)$$

is equivalent to $\zeta(s)\neq 0$ for $Re(s)>\theta$, $1/2\le\theta<1$ (Th. 12.3 [31]). If we use the strongest zero-free region then we deduce the results

$$\pi(x) - \int_2^x \frac{dt}{\log t} \ll x^{1/2}\exp(-C(\log x)^{3/5}(\log\log x)^{-1/5})$$

$$\psi(x) - x \ll x^{1/2}\exp(-C(\log x)^{3/5}(\log\log x)^{-1/5})$$

but the Riemann hypothesis is equivalent to each of the following statements related to prime numbers

(1) $\pi(x) = \int_2^x \frac{dt}{\log t} + O(\sqrt{x}\log x)$. [35]

(2) $\psi(x) = \sum_{p^m\le x}\log p = x + O(\sqrt{x}\log^2 x)$, $x>0$.

In the problem related to the difference of two consecutive prime numbers, Pilts conjectured that for every $\epsilon>0$, $p_{n+1}-p_n \ll p_n^\epsilon$ where p_n denote the nth prime number, but this

has never been proved. On the other hand, it is known that the relation $p_{n+1}-p_n \ll \log p_n$ is certainly not true.

(3) The **RH** is equivalent to $p_{n+1} - p_n \ll p_n^{1/2} \log p_n$.

The Möbius function. The Euler product formula for $\zeta(s)$ implies that

$$\frac{1}{\zeta(s)} = \prod_p \{1 - p^{-s}\} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}, \quad s = \sigma + it, \sigma > 1.$$

The coefficient $\mu(n)$ is known as Möbius function.

It has the property $\sum_{d|n} \mu(d) = 1,\ (n = 1)$ and 0 if $n > 1$, where $d|n$ means that d is a divisor of n. Thus $\mu(n)$ verifies

$$\mu(1) = 1, \mu(n) = -\sum_{d|n, d<n} \mu(d), (n > 1).$$

Each one of the following properties is equivalent to the **RH** (see [60])

(4) $M(x) = \sum_{n \le x} \mu(n) \ll x^{1/2+\epsilon}$, for all $\epsilon > 0$,

(5) the series $\sum_{n=1}^{\infty} \mu(n)/n^{-s}$ is convergent, and its sum is $1/\zeta(s)$, for every s with $\sigma > 1/2$.

The function w(n). Let $\omega(n)$ be the number of prime factors of the positive integer n counted without multiplicity, and $q \in \mathbb{N}$. D. Wolke proved that the **RH** is true if and only if

$$\sum_{n \le x} \omega^q(n) - x \sum_{0 \le j \le \frac{1}{2} \log x} R_{jq}(\log\log x)(\log x)^{-j} \ll x^{1/2+\epsilon}$$

holds for every $\epsilon > 0$, where $R_{jq}(t)$ are certain polynomials such that $deg\ R_{0q} \le q$, and $deg\ R_{jq} \le q - 1,\ (j \ge 1)$ (see [63]).

Sum of divisors. Let $\sigma(n)$ denote the sum of divisors of n, that is $\sigma(n) = \sum_{d|n} d$. G. Robin (see[54]) proved that **RH** is true, if and only if

$$\sigma(n) < e^{\gamma} n \log\log n$$

for large n, where γ is the Euler's constant. J.C. Lagarias [39], proved that under the **RH**

$$\sigma(n) < H_n + e^{H_n} \log H_n, \quad H_n = \sum_{j=1}^{n} 1/j,\ (n \ge 2).$$

Bertrand's postulate. In 1852, Chebyshev proved Bertrand's postulate according to which each interval $(n, 2n]$, $n \ge 1$, contains at least one prime. O. Romare, Y. Saoutier

[55] under the **RH** proved that for all $x \geq 2$ in the interval $(x - \frac{8}{5}\sqrt{x}\log x, x]$ there exist a prime number.

Pythagorean triangles. The positive integers a, b, c are said to form a primitive Pythagorean triples (a, b, c) if $a \leq b$, a, b, c are coprimes and $a^2 + b^2 = c^2$. Let $P(x)$ denote the number of primitive Pythagorean triangles, whose area is less than x. In 1955, J. Lambek y L. Moser [Pacific J. Math. 5(1955), 73-83, MR 16, 796h] proved that $P(x) = c_1 x^{1/2} + O(x^{1/3})$. Muler, Nowak and Menzer ([50]), assuming the **RH** proved that $P(x) = c_1 x^{1/2} + c_2 x^{1/3} + R(x)$, with $R(x) \ll x^{127/560+\epsilon}$. The error term is connected with the Möbius function and then with the zeros of the zeta function. W. Zhai [70] under the **RH** deduced

$$R(x) \ll x^{127/616}(\log x)^{963/308}, \ 127/616 = 0.2061\ldots$$

A Divisor problem. Let a, b be positive coprime integers such that $1 \leq a < b$, and let $\Delta_{a,b}(x)$ be the error term for the asymptotic formula of

$$D^*(a, b; x) = \sum_{m^a n^b \leq x, \ (m,n)=1} 1, \ (1 \leq a < b, \ (a, b) = 1.$$

Lu y Zhai [43], proved under the **RH** that

$$\Delta_{a,b}(x) \ll x^{\frac{a+6b}{(2a+7b)(a+b)}+\epsilon}, \ b \leq \frac{3a}{2}; \quad \Delta_{a,b}(x) \ll x^{\frac{a+2b}{(2a+2b)(a+b)}+\epsilon}, \ b > \frac{3a}{2}.$$

The primitive circle problem. Let

$$P(x) = \sum_{m^2+n^2 \leq x, (m,n)=1} 1$$

and let $\Delta(x) = P(x) - (6/\pi)x$. The "primitive circle problem" is to obtain an upper bound for $\Delta(x)$. The best known unconditional result is

$$\Delta(x) \ll x^{1/2} \exp(-c(\log x)^{3/5}(\log\log x)^{-1/5}).$$

Jie Wu ([68]), under the Riemann hypothesis proved that $\Delta(x) \ll x^{221/608+\epsilon}$.

Powerful numbers. Let $k \geq 2$ be a fixed integer and let N_k denote the set of all positive integers n with the property that if a prime $p \mid n$, then $p^k \mid n$. Thus the set of powerful numbers N_k contains 1 and the numbers whose canonical representation is $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$, $\alpha_i \geq k$ for all $i = 1, 2, \cdots, r$. We put $f_k(n) = 1$ if $n \in N_k$ and $f_k(n) = 0$ if $n \notin N_k$ and the Dirichlet series representation is

$$F_k(s) = \sum_{n=1}^{\infty} f_k(n) n^{-s} = \prod_p (1 + \frac{p^{-ks}}{1 - p^{-s}}), \ Re(s) > 1/k,$$

then $f_k(n)$ is the characteristic function of N_k, (see E. Kratzel [36] Chap. 7). The problem is to obtain an estimate for the number $N_k(x)$ of k-full integers not exceeding x:

$$N_k(x) = \sum_{n \le x} f_k(n), \quad k \ge 2$$

the main term is deduced from the residues of $F_k(s)$ at the simples poles $s = 1/u$, $u = k, k+1, \cdots, 2k-1$ and we have the following relation

$$N_k(x) = \sum_{t=k}^{2k-1} c_{t,k} x^{1/k} + \Delta_k(x), \; c_{t,k} = \operatorname*{Res}_{s=1/t} F_k(s)/s.$$

Let $\gamma_k = \inf\{\rho_k \mid \Delta_k(x) \ll x^{\rho_k}\}$. A. Ivić showed that we would have under the assumption of the truth of Lindelöf hypothesis, $\gamma_k \le 1/2k$, but in special cases, it is possible to prove much more. If $k = 2$, we have

$$F_2(s) = \frac{\zeta(2s)\zeta(3s)}{\zeta(6s)}.$$

Let $N_2(x)$ be denote the number of square-full integers $n \le x$, and $\Delta_2(x)$ the error term. The best unconditional O-estimate is

$$\Delta_2(x) = N_2(x) - \frac{\zeta(3/2)}{\zeta(3)} x^{\frac{1}{2}} - \frac{\zeta(2/3)}{\zeta(2)} x^{\frac{1}{3}} \ll x^{\frac{1}{6}} \exp(-c(\log x)^{\frac{3}{5}} (\log\log x)^{-\frac{1}{5}}),$$

essentially due to P. T. Bateman and E. Grosswald [3] see Theorem 14.4 and page 438 in A. Ivić's book [31].

Under the **RH**, J. Wu [67] proved that $\Delta_2(x) \ll x^{(12/85)+\epsilon}$, for any $\epsilon > 0$, which can be compared with the complementary estimate $\Delta(x) = \Omega(x^{1/10})$, due to R. Balasubramanian, K. Ramachandra and M. V. Subbarao (Acta Arith. 50 (1988), no. 2, 107-118). Bateman and Grosswald also noted that the statement $\Delta_2(x) \ll x^{\alpha}$, for any constant exponent $\alpha < 1/6$, is equivalent to some quasi-Riemann hypothesis, i.e. to the assertion that $\zeta(s) \ne 0$ for all s with $\sigma > \sigma_0$ and $\sigma_0 < 1$.

A similar situation arises in case of cube-full integers.

Farey series and the RH. A sequence (x_n) of real numbers is uniformly distributed (mod 1) if and only if for every Riemann-integrable function f on $[0, 1]$ one has

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n \le N} f(\{x_n\}) = \int_0^1 f(x) dx$$

where $\{x\}$ is the fractional part of x, ([10], Th. 3 Cahp. VIII). Let $F_x = F_{[x]}$ denote the sequence of all irreducible fractions with denominators $\le x$, arranged in increasing order of magnitude, that is

$F_x = \{r_k = a_k/b_k; 0 < a_k \leq b_k \leq x, (a_k, b_k) = 1, k = 1, 2, \cdots, \phi(x)\}$, for any $x \geq 1$, where $\phi(x) = \sum_{n \leq x} \varphi(n)$, $(\varphi(n) = \#\{k \leq n; (k, n) = 1\}$ is the Euler function. $F_{[x]}$ is called the Farey series of order x, (Note that the Farey series is not a series at all but a finite sequence).

Since the Farey fractions are uniformly distributed (mod 1) we have that

$$\lim_{N \to \infty} \frac{1}{\phi(N)} \sum_{q \leq N} \sum_{\substack{a=1 \\ (a,q)=1}}^{q} f(a/q) = \int_0^1 f(x)dx \tag{24}$$

for every Riemann-integrable function f on $[0, 1]$. Relation (24) suggests the problem of estimation of the error term

$$E_f(N) = \sum_{q \leq N} \sum_{\substack{a=1 \\ (a,q)=1}}^{q} f(a/q) - \phi(N) \int_0^1 f(x)dx.$$

Franel discovered a connection between Farey series and the Riemann hypothesis. He proved ([17]) that the estimation

$$\sum_{k \leq \phi(x)} \left(r_k - \frac{k}{\phi(x)} \right)^2 \ll_\epsilon x^{2(\beta-1)+\epsilon}$$

for every $\epsilon > 0$ is equivalent to $\zeta(s) \neq 0$ for $Re(s) > \beta$. Thus

RH is truth if and only if

$$\sum_{k \leq \phi(x)} \left(r_k - \frac{k}{\phi(x)} \right)^2 \ll x^{-1+\epsilon}.$$

Another version (Landau *Vorlesungen ii,* 167-177), is that the **RH** is equivalent to the statement

$$\sum_{k \leq \phi(x)} \mid r_k - \frac{k}{\phi(x)} \mid \ll x^{\frac{1}{2}+\epsilon},$$

for all $\epsilon > 0$ as $x \to \infty$.

M. Mikolás, (see [46]-[49]) proved that the **RH** is equivalent to the relation

$$\sum_{k=1}^{\phi(x)} f(r_k) - \phi(x) \int_0^1 f(u)du \ll x^{1/2+\epsilon}$$

for all $\epsilon > 0$, where $f(u)$ can be $\sin(\lambda u)$, $\cos(\lambda u)$, $(|\lambda| \neq \pi)$ or a polynomial of degree ≤ 3. Huxley [28] has generalized Franel's theorem to the case of Dirichlet L-functions and Fujii [18] gives another equivalence with the **RH** in terms of the Farey series.

Kanemitsu and Yoshimoto [33] showed that each one of the estimates are equivalent to the **RH**

$$\sum_{r_k\leq 1/3}\left(r_k-\frac{h(1/3)}{2\phi(x)}\right)\ll x^{1/2+\epsilon},\ \sum_{r_k\leq 1/4}\left(r_k-\frac{h(1/4)}{2\phi(x)}\right)\ll x^{1/2+\epsilon}$$

where $h(t)=\sum_{r_k\leq t}1$. Moreover, Kanemitsu and Yoshimoto obtain conditions equivalents for a great class of functions (see also [34], [69]).

The Riesz sum. M. Riesz (see [60]) considers the function

$$f(x)=\sum_{n=1}^{\infty}\frac{(-1)^{n+1}x^n}{(n-1)!\zeta(2n)}.$$

An application of the calculus of residues gives

$$f(x)=\frac{i}{2}\int_{a-i\infty}^{a+i\infty}\frac{x^s ds}{\Gamma(s)\zeta(2s)\sin(\pi s)}=\frac{i}{2\pi}\int_{a-i\infty}^{a+i\infty}\frac{\Gamma(1-s)x^s ds}{\zeta(2s)},$$

where $1/2<a<1$. Taking $a>1/2$ it follows that $f(x)\ll x^{1/2+\epsilon}$. On the **RH** we can move the line of integration to $a=1/4+\epsilon$ and obtain the estimation $f(x)\ll x^{1/4+\epsilon}$. Conversely, by Mellin's inversion formula

$$\frac{\Gamma(1-s)}{\zeta(2s)}=-\int_0^{\infty}f(x)x^{-s-1}dx$$

and if $f(x)\ll x^{1/4+\epsilon}$ holds, then the integral converges uniformly for $\sigma\geq\sigma_0>1/4$, the analytic function represented is regular for $\sigma>1/4$ and the truth of the **RH** follows.

Hardy and Littlewood in 1918 [60] stated that **RH** holds if and only if

$$\sum_{n=1}^{\infty}\frac{(-x)^n}{n!\zeta(2n+1)}\ll x^{-1/4},\ x\to\infty.$$

Nyman-Beurling criterion. The Riemann zeta function on may see as a Mellin transform in the critical strip by means of the following formula

$$\frac{\zeta(s)}{s}=-\int_0^{\infty}\rho(1/x)x^{s-1}dx, 0<\sigma<1$$

where $\rho(x)$ is the fractional part function $\rho(x)=x-[x]$.

B. Nyman and A. Beurling had the idea that it should be possible to translate the Riemann hypothesis into a property of $\rho(x)$. Beurling's linear space $N_{(0,1)}$ consists of all functions

$$x\to f(x)=\sum_{k=1}^{n}a_k\rho(\frac{\theta_k}{x}),\ 0<\theta_k\leq 1$$

where the a_k are constants such that, $\sum_{k=1}^{n}a_k\theta_k=0$. For $1<p<\infty$, let B^p be the closure of B in $L^p(0,1)$. In his thesis, Nyman [51] proved the following theorem:

Theorem. *The Riemann hypothesis is true if and only if $N_{(0,1)}$ is dense in $L^2(0,1)$.*

Beurling gives in his paper [4] a generalization of Nyman's theorem

Theorem. *Let $1 < p < \infty$. The following properties are equivalent.*

(i) $\zeta(s)$ has no zeros in the half plane $Re(s) > 1/p$

(ii) $N_{(0,1)}$ is dense in $L^p(0,1)$

(iii) The characteristic function $\chi_{(0,1)}$ is in the closure of $N_{(0,1)}$ in $L^p(0,1)$.

L. Baez-Duarte [1], stated a new version of Nyman-Beurling criterion:

Theorem. *Let $M(y)$ be denote the summatory function of $\mu(n)$, and let*

$$G_n(x) = \int_1^n M(\theta)\rho(\frac{1}{\theta x})\frac{d\theta}{\theta}.$$

(1) If $\lambda(x) = \chi_{(0,1)}(x)\log x$, then $\parallel G_n - \lambda \parallel_p \to 0$, as $n \to \infty$ implies that $\zeta(s) \neq 0$ para $Re(s) > 1/p$. (2) If $\zeta(s) \neq 0$ for $Re(s) > 1/p$ then $\parallel G_n - \lambda \parallel_r \to 0$, as $n \to \infty$ whenever $r \in (1,p)$.

Riemann ξ-function and positivity criterion. The Riemann ξ-function (6), is an entire function of order one which is real-valued on the real axis and satisfies the relation $Re\left(\frac{\xi'(s)}{\xi(s)}\right) > 0$ when $Re(s) > 1$. The Riemann hypothesis is equivalent to the positivity condition

$$Re\left(\frac{\xi'(s)}{\xi(s)}\right) > 0, \text{ when } Re(s) > 1/2.$$

(Hinkkanen [27]). J.C. Lagarias [38] defined an arbitrary discrete set Ω in $\mathbb{C}$ which represents the set of zeros of an entire function $f_\Omega(z)$ counted with multiplicity. Lagarias call admissible to the set Ω if complex conjugate zeros ρ and $\overline{\rho}$ occur with the same multiplicity, and the zeros satisfy the convergence condition

$$\sum_{\rho\in\Omega} \frac{1+ \mid Re(\rho) \mid}{1+ \mid \rho \mid^2} < \infty.$$

Theorem. ([38]) *Let Ω be an admissible zero of set in $\mathbb{C}$. Then the following conditions are equivalent*

(1) All elements $\rho \in \Omega$ have $Re(\rho) \leq \theta$

(2) The function $f'_\Omega(s)/f_\Omega(s)$ satisfies the positivity condition

$$Re(f'_\Omega(s)/f_\Omega(s)) > 0, \text{ for } Re(s) > \theta.$$

If Ω are the non trivial zeros of $\zeta(s)$, and

$$h_{\mathbb{Q}}(\sigma) = \inf\{Re\left(\frac{\xi'(\sigma+it)}{\xi(\sigma+it)}\right) : -\infty < t < \infty\},$$

then under the **RH** Lagarias proved that $h(\sigma) = \xi'(\sigma)/\xi(\sigma)$, for $\sigma > 1/2$

This result is improved by R. Garunkstis [19], who obtained that if $\zeta(s) \neq 0$ for $\sigma > a$, $1/2 \leq a < 1$ then $h(\sigma) = \xi'(\sigma)/\xi(\sigma)$ for $\sigma > a$.

Xian-Jin Li (see [40]) showed that a necessary and sufficient condition for the non trivial zeros of the Riemann zeta function to lie on the critical line is that $\lambda_n = ((n-1)!)^{-1}(d^n/ds^n)(s^{n-1}\log\xi(s))\mid_{s=1}$ is non negative for every positive integer. Thus the **RH** holds if and only if $\lambda_n \geq 0$, for each $n = 1, 2, 3, \ldots$.

He also showed that an identical result applies to the Riemann hypothesis for the Dedekind zeta-function $\zeta_K(s)$ of a number field.

The number λ_n can be written in terms of the complex zeros of $\zeta(s)$ as

$$\lambda_n = \sum_{\rho}(1-(1-1/\rho)^n)$$

where the sum over ρ is understood as $\sum_\rho = \lim_{T\to\infty}\sum_{|Im(\rho)\leq T}$.

Bombieri y Lagarias showed that Li's criterion follows as a consequence of a general set of inequalities for an arbitrary multiset of complex numbers ρ and therefore is not specific to zeta function (see [5]).

Weil-Bombieri and RH. In 1952, A. Weil [62] gives a generalization of the Riemann explicit formula and in the same paper Weil proved that from his formula one can define a quadratic functional whose positivity is equivalent to the Riemann hypothesis.

E. Bombieri [6], studies the Weil explicit formula. First of all, he provides a very clear proof of the formula which relates the values of a smooth function f summed over the primes to the sum of its Mellin transform $\tilde{f}$ summed over the complex zeros of the Riemann zeta-function.

Explicit Formula. Define $f^*(x) = f(1/x)/x$. If $f(x)$ is any smooth complex-valued function with compact support in $(0,\infty)$ and Mellin transform

$$\tilde{f}(x) = \int_0^\infty f(x)x^{s-1}dx$$

then

$$\sum_\rho \tilde{f}(\rho) = \int_0^\infty f(x)dx + \int_0^\infty f^*(x)dx - \sum_{n=1}^\infty \Lambda(n)(f(n)+f^*(n))$$
$$-(\log 4\pi + \gamma)f(1) - \int_1^\infty \left(f(x)+f^*(x)-\frac{2}{x}f(1)\right)\frac{x}{x^2-1}dx.$$

Bombieri next proves a strong version of Weil's criterion for the Riemann hypothesis. One takes a function $f(x) = g * \overline{g}^*$ in the Weil formula, where $*$ is the multiplicative convolution of a function g and its transpose conjugate $\overline{g}^*$.

Theorem (Bombieri). *The Riemann Hypothesis holds if and only if*

$$\sum_{\rho} \tilde{g}(\rho)\tilde{\overline{g}}(1-\rho) > 0$$

for every complex-valued $g(x) \in C_0^\infty((0,\infty))$, not identically 0.

6 Generalizations of the Riemann hypothesis

Let $\{a(n)\}_{n=1}^\infty$ be a sequence of complex numbers, and $\sum_{n=1}^\infty a(n)n^{-s}$, $(s \in \mathbb{C})$ the associated Dirichlet series. When $a(n) \equiv 1$ we have the Riemann zeta function.

L-functions. Let k be a fixed positive integer. A Dirichlet character χ of conductor k is a completely multiplicative function on the integers: that is $\chi(mn) = \chi(m)\chi(n)$ for every pair of integers m, n, periodic with period k, $\chi(n+k) = \chi(n)$, and such that $\chi(n) = 0$ if $(n,k) > 1$. The principal character verifies $\chi(n) = 1$ if $(n,k) = 1$ and $\chi(n) = 0$ if $(n,k) > 1$.

One can then define a L-function associated to $\chi(n)$ by

$$L(s,\chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$$

and as $\chi(n)$ is bounded, we have that the series is absolutely convergent for $Re(s) > 1$. More than this is true, however: the series for $L(s,\chi)$, (χ non principal) converges for $Re(s) > 0$. One has also an expansion as an Euler product

$$L(s,\chi) = \prod_p (1 - \chi(p)p^{-s})^{-1}.$$

It was introduced by Dirichlet for solving the problem of existence of infinity prime numbers on arithmetic progressions.

Dirichlet's theorem. *There are infinitely many primes of the form $kt + \ell$ if and only if ℓ and k are coprime integers $(\ell, k) = 1$.* (For instance, see Ellison Mendes-France's book for a proof [16]).

The Dirichlet series simplest after $\zeta(s)$ is the Dirichlet L-function, $L(s,\chi_3)$, for the non trivial character of conductor 3:

$$L(s,\chi_3) = \sum_{r=0}^{\infty} \{(3r+1)^{-s} - (3r+2)^{-s}\}.$$

This can be written as an Euler product

$$L(s,\chi_3)=\prod_{p\equiv 1 \pmod 3}\{1-p^{-s}\}^{-1}\prod_{p\equiv 2 \pmod 3}\{1+p^{-s}\}^{-1}.$$

The L-function $L(s,\chi_3)$ satisfies the functional equation

$$\xi(s,\chi_3)=(\pi/3)^{-(s+1)/2}\Gamma((s+1)/2))L(s,\chi_3)=\xi(1-s,\chi_3).$$

It is expected to have all of its nontrivial zeros on the critical line $1/2$.

One defines a character of conductor 4 as $\chi_4(n)=0$ if n is even; and $\chi_4(n)=(-1)^{(n-1)/2}$ if n is odd. Then

$$L(s,\chi_4)=\sum_{n=1}^{\infty}\frac{\chi_4(n)}{n^s}=1-\frac{1}{3^s}+\frac{1}{5^s}-\frac{1}{7^s}+\cdots,\ Re(s)>1.$$

This can be written as an Euler product

$$L(s,\chi_4)=\prod_{p\equiv 1 \pmod 4}\{1-p^{-s}\}^{-1}\prod_{p\equiv 3 \pmod 4}\{1+p^{-s}\}^{-1}.$$

The L-function $L(s,\chi_4)$ also satisfies an analogue functional equation.

Theorem. *Let χ be a non principal character, primitive of conductor k, ($k>1$). The L-function $L(s,\chi)$ has an analytic continuation to the complex plane which is an entire function of s. It verifies a functional equation*

$$\xi(s,\chi)=E(\chi)\xi(1-s,\bar{\chi})$$

where

$$\xi(s,\chi)=\left(\frac{\pi}{k}\right)^{-\frac{s+a}{2}}\Gamma(\frac{s+a}{2})L(s,\chi)$$

and $a=0$ if $\chi(-1)=1$; $a=1$ if $\chi(-1)=-1$. The bar means complex conjugation and

$$E(\chi)=k^{-1/2}\sum_{r=1}^{k-1}\chi(r)e^{\frac{2\pi i}{k}}.$$

For the proof, see for instance [11].

Since $L(s,\chi)\neq 0$ for $\sigma>1$, $\xi(s,\chi)\neq 0$ and $\xi(s,\overline{\chi})\neq 0$ for $\sigma>1$. By the functional equation we obtain that $\xi(s,\chi)\neq 0$ for $\sigma<0$ so that all the zeros of $\xi(s,\chi)$ must lie in the strip $0\leq\sigma\leq 1$.

Corollary. *If $a=0$, the function $L(s,\chi)$ has simple zeros at $s=0,-2n,n\in\mathbb{N}$. If $a=1$, then $L(s,\chi)$ has simple zeros at $s=-(2n-1),n\in\mathbb{N}$. These are called the trivial zeros.*

As in the case of $\zeta(s)$, one can deduce the existence of an infinity of non-real zeros of $L(s,\chi)$ either by using to the theory of entire functions, or by an estimate of the Riemann-von Mangoldt type. One can prove that if $N(T)$ denotes the number of zeros $\{\rho = \beta + i\gamma, 0 < \beta < 1, 0 < \gamma \leq T\}$ of $L(s,\chi)$ then

$$N(T) = \frac{T}{2\pi}\log T + AT + O(\log T)$$

where A is a constant which depends on k, the modulus of character χ.

The theorem of the existence of an infinity of zeros of the critical line, proved by Hardy for the $\zeta(s)$, has been extended to a general class of Dirichlet series, including the L-functions, by E. Hecke.

Generalized Riemann Hypothesis. The conjecture is that all the zeros of L-functions are situated on the critical line.

Let $\left(\frac{d}{n}\right)$ denote Kronecker's extension of the Legendre symbol for quadratic residues, for $n = 1, 2, \cdots$. It is known that if d is the discriminant of a quadratic number field (that is, d is square-free and $d \equiv 1 \pmod 4$; or $d = 4N$ where $N \equiv 2$ or $3 \pmod 4$, and square-free), and $\chi_d(n) = \left(\frac{d}{n}\right)$, then χ_d is a real primitive Dirichlet character modulus $\mid d \mid$. Let

$$L(s,\chi_d) = \sum_{n=1}^{\infty} \chi_d(n) n^{-s}$$

be the associated Dirichlet L-function. It is an entire function (for $d \neq 1$) which satisfies the functional equation

$$\xi(s,\chi_d) = \left(\frac{\pi}{\mid d \mid}\right)^{-s/2} \Gamma\left(\frac{s+a_d}{2}\right) L(s,\chi_d) = \xi(1-s,\chi_d)$$

where $a_d = 1$ if $d < 0$ and $a_d = 0$ if $d > 0$.

Zeta functions of number field. For a number field K of degree n, the zeta function of K is

$$\zeta_K(s) = \sum_{m=1}^{\infty} \frac{F_K(m)}{m^s},$$

where $F_K(m)$ is the number of ideals whose norm is precisely m. This sum converges for complex s with $Re(s) > 1$. Often $\zeta_K(s)$ is referred to as the Dedekind zeta function. When $K = \mathbb{Q}$, is just the Riemann zeta function: $\zeta_{\mathbb{Q}}(s) = \zeta(s)$.

The function $\zeta_K(s)$ has an analytic continuation to the entire complex s-plane except for a first order pole at $s = 1$. The residue of $\zeta_K(s)$ at $s = 1$ is

$$Res_{s=1}\zeta_K(s) = \frac{2^{r_1+r_2}\pi^{r_2} hR}{w\sqrt{\mid D \mid}}$$

where r_1 is the number of real conjugate fields of K, $2r_2$ is the number of complex conjugate fields of K, n is the degree of K, h is the class-number of K, R is the regulator of K, w is the number of roots of unity in K, and D is the discriminant of K. There is also a functional equation relating values at $s = 1$ to values at $1 - s$,

$$\xi_K(s) = \xi_K(1-s)$$

where

$$\xi_K(s) = \left(\frac{\mid D \mid}{2^{2r_2}\pi^n}\right)^{s/2} \Gamma^{r_1}(s/2)\Gamma^{r_2}(s)\zeta_K(s)$$

The function $\zeta_K(s)$ has no zeros in $Re(s) > 1$. It has trivial zeros with $Re(s) < 0$ of order r_2 at $-1, -3, -5, \ldots$; of order $r_1 + r_2$ at $-2, -4, -6, \ldots$; and one zero of order $r_1 + r_2 - 1$ at $s = 0$. All other zeros are in the critical strip. (See [65] chap. 6: *Galois Theory, Algebraic Number Theory, and Zeta Functions* from H.M. Stark).

Extended Riemann Hypothesis. The Extended Riemann Hypothesis is the assertion that the non trivial zeros of Dedekind zeta function of any algebraic number field lie on the critical line.

Some consequences.

(1) Given the **GRH**, the Siegel-Walfisz theorem (see Th-5 [59], cap II.8) can be improved to

$$\psi(x; h, k) = \sum_{\substack{n \leq x \\ n \equiv h \pmod{k}}} \Lambda(n) = \frac{x}{\varphi(k)} + O(\sqrt{x}\log^2 x),\ k \leq x.$$

(2) Also we have

$$\pi(x; h, k) = \sum_{\substack{p \leq x \\ p \equiv h \pmod{k}}} 1 = \frac{x}{\varphi(k)\log x} + O(\sqrt{x}\log x).$$

(3) D. Wolke (see [64]), obtained the following equivalence for L-functions: Let $\chi_0, \chi_1, \chi_2, \ldots$ be the sequence of all Dirichlet characters (in which the principal character χ_0 occurs only once), ordered which increasing moduli, and let $\{\alpha_k\}_{k\in\mathbb{N}}$ a sequence of real numbers $\alpha_k \to 0$ for $k \to \infty$ which, in addition, satisfies the condition

$$\sum_{k\in\mathbb{N}} \alpha_k \log(k+1) \mid \log \alpha_k \mid < \infty.$$

Define the multiplicative function $f : \mathbb{N} \to \mathbb{C}$ by

$$\sum_{n=1}^{\infty} f(n)n^{-s} = (\zeta(s))^{-1} \prod_{k=1}^{\infty} (L(s, \chi_k))^{\alpha_k},\ Re(s) > 1$$

then Wolke proved that the **GRH** in true iff

$$\sum_{n\leq x} f(n) \ll x^{1/2+\epsilon}, \quad for\ all\ \epsilon > 0.$$

(4) *The Goldbach conjecture.* The three prime Goldbach conjecture states that every odd integer ≥ 9 is a sum of three odd primes. The conjecture was proved under the **GRH** by Deshouillers-Effinger-te Riele-Zinoviev [14].

Quasi Riemann hypothesis. The term "Quasi Riemann hypothesis" related to a zeta-function is used to mean that this zeta-function has no zeros in a half-plane $\sigma > \sigma_0$ for some $\sigma_0 < 1$.

Modified Generalized Riemann Hypothesis. Say that an L-function satisfies the Modified Great Riemann Hypothesis, if its zeros are all on the critical line or the real axis.

The function $\omega(n) = \sum_{p|n} 1$ has "normal" order $\log\log n$ and a famous theorem of Turan bounds the variance of this distribution. Now let $f_a(n)$ be the smallest positive integer m such that $a^m \equiv 1 \pmod{n}$, with a relatively prime to n. Saidak ([56]) obtained results of Turan type for the function $\omega(f_a(n))$, thus assuming a quasi-Riemann hypothesis for the Dedekind zeta functions of certain nonabelian number fields, he proves that for each squarefree integer $a \geq 2$,

$$\sum_{p\leq x(p,a)=1} \Big(\omega(f_a(p)) - \log\log p\Big)^2 \ll \pi(x)\log\log x$$

$$\sum_{n\leq x(n,a)=1} \Big(\omega(f_a(n)) - \frac{1}{2}(\log\log n)^2\Big)^2 \ll x(\log\log x)^3.$$

The Ramanujan zeta function. Let

$$L(s,\tau) = \sum_{n=1}^{\infty} \frac{\tau(n)}{n^s}, \quad \sigma > \frac{13}{2} \tag{25}$$

be the L-function attached to the Ramanujan function $\tau(n)$. The function $\tau(n)$ is multiplicative (proved by Mordell, 1917). The order of magnitude for $\tau(n)$ is $|\tau(n)| \leq n^{11/2}d(n)$, which was conjectured by Ramanujan in 1916 and proved by P. Deligne 60 years later [13].

The Euler product corresponding to $L(s,\tau)$ is

$$L(s,\tau) = \prod_p (1 - \tau(p)p^{-s} + p^{11-2s})^{-1}, \quad (\sigma > \frac{13}{2})$$

and the functional equation, becomes

$$(2\pi)^{-s}\Gamma(s)L(s,\tau) = (2\pi)^{s-12}\Gamma(12-s)L(12-s,\tau).$$

It has the trivial zeros $s = 0, -1, -2, -3, \cdots$ but no other zeros for $\sigma \leq \frac{11}{2}$ and $\sigma \geq \frac{13}{2}$.

The critical strip $0 < \sigma < 1$ corresponding to the function $\zeta(s)$ is the strip $\frac{11}{2} < \sigma < \frac{13}{2}$. The analytic continuation of $L(s,\tau)$ has an infinity of zeros on the critical line $Re(s) = 6$ and the Riemann hypothesis for $L(s,\tau)$ asserts that all its complex zeros lie on that line.

The Davenport-Heilbronn zeta function. This function was introduced by H. Davenport and H. Heilbronn as

$$f(s) = 5^{-s}(\zeta(s,1/5) + \tan\theta\zeta(s,2/5) - \tan\theta\zeta(s,3/5) - \zeta(s,4/5)) \tag{26}$$

where

$$\zeta(s,a) = \sum_{n=0}^{\infty}(n+a)^{-s} \quad (0 < a \leq 1),\ Re(s) > 1$$

is the Hurwitz zeta function defined for $Re(s) \leq 1$ by analytic continuation. For $\theta = \arctan((\sqrt{10-2\sqrt{5}}-2)/(\sqrt{5}-1))$ it can be shown (see [12] or [60] cap. X), that $f(s)$ satisfies the functional equation

$$f(s) = 2\cdot 5^{-s+\frac{1}{2}}(2\pi)^{s-1}\Gamma(1-s)\cos(\frac{\pi s}{2})f(1-s).$$

Moreover $f(s)$ has an infinity of zeros on the line $\sigma = 1/2$ and it also has an infinity of zeros in the half-plane $Re(s) > 1$, ($\zeta(s) \neq 0$ in $Re(s) > 1$). Thus the function $f(s)$ of Davenport and H. Heilbronn, satisfies a functional equation similar to the functional equation for $\zeta(s)$, but has no Euler product, and the analogue of the Riemann hypothesis is FALSE for $f(s)$.

7 Remarks.

The work of Riemann on the distribution of primes is thoroughly studied in Edwards' book [15], that I recommend strongly (a jewel), but a moderate knowledge of elementary number theory and complex analysis is required of the reader. I suggest for example to study the Tenenbaum's book (also the book of Ellison-Mendés). Another important text is the great classic book of E.C. Titchmarh with an extensive treatment of the theory of $\zeta(s)$ up to 1951 and its second edition revised by D.R. Heath-Brown (see Chapters XIII-XV for Riemann Hypothesis). A. Ivić [31] develops the theory of the Riemann zeta-function and some applications, and the results proved in this text are unconditional, that is, they do not depend on any hitherto unproved hypothesis.

There are so many results and conjectures having to do with the Riemann hypothesis that it is very difficult to mention all of them. The reader interested on Riemann hypothesis can consult the Mathematical Reviews with about 1600 references (up today).

References

[1] Baez-Duarte, L. *New versions of the Nyman-Beurling criterion for the Riemann Hypothesis.* Int. J. Math. Math. sci. 31 (2002), 7, 387-406.

[2] Balazard, M. *Completeness problems and the Riemann hypothesis; An annotated bibliography.* Collection: Number theory for the Milleniunm (2002), 21-48.

[3] Bateman, P.T.; Grosswald, E. *On a theorem of Erdos and Szekeres.* Illinois-J.-Math. 2 (1958), 88-98.

[4] Beurling, A. *A closure problem related to the Riemann zeta-function.* Proc. Nat. Acad. Sci. U.S.A. (1955), 41, 312-314.

[5] Bombieri, E. Lagarias, J.C. *Complements to Li's criterion for the Riemann hypothesis.* J. Number Theory 77, 274-287, (1999).

[6] Bombieri, E. *Remarks on Weil's quadratic functional in the theory of prime numbers, I.* Att. Accad Naz. Licei (9) Mat. Appl. 11 (2000) n-3, 183-233.

[7] Calderón, C. *La función zeta de Riemann.* Rev. Real Academia de Ciencias de Zaragoza. **57**, (2002), 67-87.

[8] Conrey,J.B. *The Riemann Hypothesis.* Not. Amer. Math. Soc. 50 (2003),no. 3, 341-353.

[9] Conrey,J.B.-Ghosh, A.-Gonek,S.M. *Large gaps between zeros on the zeta function.* Mathematika, 33 (1986) 2, 212-338 (1987).

[10] Chandrasekharan, K. *Introduction to analytic number theory.*Springer Verlag New York 1968.

[11] Chandrasekharan, K. *Arithmetical functions.* Springer Verlag New York 1969.

[12] Davenport, H.-Heilbrom, H. *On the zeros of certain Dirichlet series, I, II.*J. London Math. Soc. 11 (1936), 181-185, 307-312.

[13] Deligne, P. *La conjeture de Weil I.* I.H.E.S. Publ. Math. 43 (1974), 273-307.

[14] Deshouillers, J.M.-Effinger, G.-te-Riele, H.-Zinoviev, D. *A complete Vinogradov* 3*-primes theorem under the Riemann hypothesis.* Electron Res. Announc. Amer. Math. Soc. 3 (1997), 99-104.

[15] Edwards, H.M. *Riemann's zeta function* Acad. Press. New York. 1974.

[16] Ellison, W. J. and Mendès-France, M. *Les nombres premiers.* Hermann Paris (1975).

[17] Franel, J. *Les suites de Farey et le probléme des nombres premiers.* Göttinger Nachrichten (1924), 198-201.

[18] Fujii, A. *A remark on the Riemann hypothesis.* Comment. Math. Univ. St. Paul. 29 (1980), 195-201.

[19] Garunkstis, R. *On a positivity property of the Riemann ξ-function* Lithuanian Math. Journal 42 (2002) n-2, 140-145.

[20] Gauss, C.F. *Carta a Enke fechada el 24 de Diciembre de 1849,* "Werke" Vol.II, pp. 444-447. Königlichen Gesellschaft der Wissenchaften zu Göttingen.

[21] Hadamard, J. *Étude sur les propriétés des fonctions entiéres et en particulier d'une fonction considerée par Riemann.* J. Math. Pures Appl. (4) 9, 171-215 (1893)

[22] —— *Sur la distribution des zéros de la fonction $\zeta(s)$ et ses consécuences arithmétiques.* Bull. Soc. Math. France 24, 199-220 (1896).

[23] Hardy, G.H. *Comptes Rendus,* Acad. Sci. Paris, 158 (1914), 1012-1014.

[24] —— *Sur les zéros de la function $\zeta(s)$ de Riemann.* C.R. Acad. Sci. Paris 158, 1012-1014 (1914).

[25] Hardy, G.H. and Wright, E.M. *An Introduction to the theory of numbers.* Oxford. Clarendon Press. 1979.

[26] Hardy, G.H.-Littlewood, J.E. *Contributions to the theory of the Riemann zeta function and the distributions of primes.* Acta Math. 41, 19-196 (1918).

[27] Hinkkanen, A. *On functions of bounded type.* Complex Var. Th. Appl. Int. Jour. 34 (1997) no. 1-2, 119-139.

[28] Huxley, N.M. *The distribution of Farey points I.* Acta Arith. 18 (1971), 281-287.

[29] —— *Exponential sums and the Riemann zeta function IV.* Proc. London Math. Soc. (1993) (3) 66, 1-40

[30] Ingham, A. *The distribution of prime numbers.* Cambridge University Press, Cambridge, 1932.

[31] Ivić, A. *The Riemann zeta function.* Jhon Wiley, New York 1985.

[32] —— *On some results concerning the Riemann hypothesis.*Collection: Analytic number theory London Math. Soc. Lecture Notes Ser. 247, 139-167, (1997).

[33] Kanemitsu, S- Yoshimoto, M. *Farey series and the Riemann hypothesis.* Acta Arith. 75 (1996) n-4, 351-374.

[34] Kanemitsu,-Shigeru; Yoshimoto,-Masami, *Farey series and the Riemann hypothesis. III.* International Symposium on Number Theory (Madras, 1996). Ramanujan-J. 1 (1997), no. 4, 363-378.

[35] Koch, H. von. *Sur la distribution des nombres premiers.* Acta Math. 24 (1901), 159-182

[36] Kratzel, E. *Lattice points,* VEB Deutscher Verlag, Berlin 1988.

[37] Laugwitz, D. *Bernhard Riemann 1826-1866. Turning points in the conception of Mathematics.* Translated by Abe Shenitzer, Birkhäuser, Boston Basel Berlin, 1999.

[38] Lagarias, J.C. *On a positivity property of the Riemann ξ function.* Acta Arith. 89, (1999),3, 217-234.

[39] —— *An elementary problem equivalent to the RiemannHypothesis.*Amer. Math. Montly, 109 (2002) 6, 534-543.

[40] Li, Xian Jin. *The positivity of a sequence of numbers and the Riemann hypothesis.* J. Number Theory 65 (1997) n-2, 325-333.

[41] Littlewood, J.E. *Quelques consécuences de l'hypothèse que la fonction $\zeta(s)$ de Riemann n'a pas de zéros dans la demi-plan $Re(s) > 1/2$.* Comptes Rendus de l'Acad. des Sciences (Paris) 154 (1912), 263-266.

[42] Littlewood, J.E. *On the zeros of Riemann zeta-function.* Proc. Camb. Phil. Soc. 22 (1924), 295-318.

[43] Lu, G.S. , Zhai, W. *A special divisor problem.* Acta Math. Sinica 45 (2002) 5, 991-1004.

[44] van de Lune,J. - te Riele, H.J.J.-Winter, D.T. *On the zeros of the Riemann zeta function IV.* Math. Comp. 46 (1986) n-174, 667-681.

[45] Mangoldt, H. von *Zur Verteilung der Nullstellen der Riemannschen Funktion $\xi(t)$* Math. Annalen, 60 (1905), 1-19.

[46] Mikolas, M. *Sur l'hypothese de Riemann.* C.-R.-Acad.-Sci.-Paris 228, (1949). 633–636

[47] —— *Farey series and their connection with the prime number problem. I. (Errata: EA 13,1138)* Acta-Univ.-Szeged.-Sect.-Sci.-Math. 13, (1949). 93–117

[48] —— *Farey series and their connection with the prime number problem. II.* Acta-Sci.-Math.-Szeged 14, (1951). 5–21

[49] —— *An equivalence theorem concerning Farey series.* Mat.-Lapok 2, (1951). 46–53

[50] Muler, W. , Nowak, W.G. and Menzer, H. *On the number of primitive Pythagorean triangles.* Ann. Sci. Math. Quebec 12 (1988) 2, 263-273.

[51] Nyman, B. *On the one-dimensional translation group abd semi-group in certain functions spaces. Ph.D.* Thesis Upsala, 1950.

[52] Ribenboim, P. *The book of prime number records.* Springer-Verlag, 1989.

[53] Riemann, B. *Über die Anzahl der Prinzahlen unter einer gegebener Gröse.* Mo- nastsber. Akad. Berlin, 671-680, (1859).

[54] Robin, G. *Large values of the sum of divisors function and the Riemann hypothesis.* J. Math. Pures Appl. (9) 63 (1984) n-2, 187-213.

[55] Romare, O. and Saoutier, Y. *Short effective intervals containig primes.* J. Number Theory, 98 (2003), 1, 10-33

[56] Saidak, F. *The normal number of prime factors of $f_a(n)$.* J.-Ramanujan-Math.-Soc. 17 (2002), no. 1, 19-33.

[57] Selberg, A. *Contributions to the theory of the Riemann zeta function.* Arch. Math. og. Naturv. B, 48 (1946) n-5.

[58] Speiser, A. *Geometrisches zur Riemannschen Zetafuncktion.* M.A. 110 (1934), 514-521.

[59] Tenenbaum, G. *Introduction to analytic and probabilistic number theory.*Cambridge Univ. Press 46 (1995).

[60] Titchmarsh, E.C. *The theory of the Riemann Zeta-Function.* Clarendon Press. Oxford 1988. Revised by D.R. Heath Brown.

[61] Vallée-Poussin, Ch. de la *Recherches analytiques sur la theorie des nombres. La función $\zeta(s)$ de Riemann et les nombres premiers en general.* Annales de la Soc. Sci. de Bruxelles 20, (1896), 183-256.

[62] Weil,A. *Sur les "formules explicites" de la theorie des nombres premiers.* Comm.-Sem.-Math.-Univ.-Lund 1952, (1952). Tome Supplementaire, 252-265.

[63] Wolke, D. *On the number theoretic function w(n).* Acta Arith. 55 (1990), 4, 323-331.

[64] —— *On a certain class of multiplicative functions.* J. Number Theory 37 (1991),n-3, 279-287.

[65] Waldschmidt, M. ; Moussa, P. ; Luck, J.M. ; Itzykson, C. *From Number theory to Physics.* Springer-Verlag,1989.

[66] Walfisz, A. *Weylsche Exponentiansummen in der Neueren Zahlentheorie.* VEB Deutscher Verlag, Berlin 1963.

[67] Wu, J. *On the distribution of square-full integers.* Arch. Math. (2001) 233-240.

[68] —— *On the primitive circle problem.* Monatsh.-Math. 135 (2002), no. 1, 69-81.

[69] Yoshimoto, M. *Farey series and the Riemann hypothesis. IV.* Acta-Math.-Hungar. 87 (2000), no. 1-2, 109-119. Part II: Acta Math. Hungar. 78 (1998), no. 4, 287-304.

[70] Zhai, Wen-Guang, *On the number of primitive Pythagorean triangles.* Acta Arith. 105 (2002) 4, 387-403.

Monografías de la Real Academia de Ciencias de Zaragoza. **26:** 31–56, (2004).

La ecuación de Navier-Stokes. Un reto físico-matemático para el siglo XXI

Juan Luis Vázquez

Departamento de Matemáticas. Univ. Autónoma de Madrid

Resumen

Examinamos en estas notas el reto matemático de las ecuaciones de Navier-Stokes en el marco de Los Problemas Clay y concedemos importancia al hecho de que un problema de índole intelectual pura tenga relación con una problemática que afecta a la Física, a la Ingeniería y a la vida diaria de la Sociedad. En el terreno de las matemáticas puras, que es aquel en que se juega el reto, intentamos explicar cuál es la dificultad que ha eludido a algunas de las mejores mentes del mundo científico por siglo y medio. En concreto, planteamos el problema bajo el punto de vista de los problemas de explosión o *blow-up*.

1 Retos matemáticos. Los "Problemas del Milenio"

Las matemáticas tienen múltiples facetas, desde la construcción de sofisticadas teorías intelectuales a la modelización del mundo real, de Pitágoras a Newton, de Gauss a Einstein, etc. Hay en las mejores matemáticas una tensión permanente entre el arte y la utilidad, entre las capacidades de crear y descubrir en el reino matemático y las de explicar y controlar el mundo que nos rodea[1]. Pero las matemáticas son una cultura con muchas aspectos y matices y uno de los ingredientes que más fascina a los profesionales es el reto de *los problemas abiertos*: el largo discurrir, a veces arduo, a veces tranquilo, de la construcción de una teoría matemática se va encrespando en dificultad según se avanza y en casos frecuentes (y al parecer de los matemáticos, afortunados) cristaliza en una dificultad muy específica, una dura roca, que pide a gritos el concurso a una mente excelente o la combinación de varias de tales mentes, para que ataquen al monstruo, lo pongan a nuestros pies y nos permitan así seguir avanzando.

[1]Una detallada discusión de esta dualidad puede encontrarse en [32].

El año 1900 fue un año extraordinario para esa parte de la Humanidad (mínima, pero entrañable para nosotros) que se apasiona por los *grandes problemas abiertos matemáticos*. En efecto, en ese año el gran matemático alemán David Hilbert planteó en el Congreso Internacional de París sus famosos 23 problemas que tuvieron en el mundo matemático del siglo XX tanta o más resonancia que las tesis de Lutero en el mundo norteuropeo[2].

Al cumplirse un siglo de este notable hecho, diversas iniciativas pretenden dar la réplica al gran hombre, cf. por ejemplo los libro de Arnold-Atiyah-Lax-Mazur [2], la lista de Smale en ese volumen, o el libro de Engquist-Schmid [10]. El miércoles 24 de mayo de 2000 se anunció en el Collège de France de París el Conjunto de los 7 problemas matemáticos que constituyen los *Millennium Prize Problems*, patrocinados por el *Clay Mathematics Institute*. Recordando a Hilbert, pretendía reflejar 7 de los más importantes problemas abiertos de la ciencia matemática al comienzo del nuevo siglo[3]. Estos problemas recorren las diversas áreas las matemáticas puras y aplicadas y son

1. P versus NP (Teoría de la computación)

2. Conjetura de Hodge (Geometría algebraica)

3. Conjetura de Poincaré (Geometría y topología)

4. Hipótesis de Riemann (Teoría de números y Análisis)

5. Existencia de Yang-Mills y salto de masa (Física teórica)

6. Existencia y regularidad para las ecuaciones de Navier-Stokes (Mecánica de Fluidos y EDPs)

7. Conjetura de Birch y Swinnerton-Dyer (Geometría aritmética algebraica).

Por el cuidado en la selección de problemas, por la seriedad con que procede la Fundación y en vista de la reacción habida en los cuatro años pasados, esta lista parece destinada a ser famosa e influyente. De acuerdo con los tiempos de optimismo y expansión que corren para las matemáticas, la lista incluye problemas abiertos importantes en temas variados tanto de la matemática pura como de la aplicada. La computación teórica, ese hijo aventajado que le ha surgido a las matemáticas en siglo XX figura con "su problema".

Nosotros nos centraremos en el problema 6, que une física de medios continuos, ecuaciones en derivadas parciales, análisis funcional y un número prometedor de aplicaciones a cálculos de la vida diaria. Su resolución haría justicia a la visión de la matemática como herramienta básica de la ciencia y la ingeniería y haría un gran favor a la popularidad del matemático teórico en el "mundo real"; pues si seguimos la máxima de que

[2]Para una referencia a estos problemas ver por ejemplo [15].

[3]La resolución de cada problema valdría al autor un premio de 1 millón de dólares. Toda la información sobre el premio y los problemas se puede obtener en la dirección `http://www.claymath.org/prize_problems`.

las Matemáticas son el lenguaje en que se piensa la Ciencia, sería prudente que la comunidad matemática diera una cierta prioridad a tener el tal lenguaje listo y reluciente en los campos en que las ciencias llaman a nuestra puerta.

La ecuación de Navier-Stokes en el portal del Clay Mathematics Institute

Waves follow our boat as we meander across the lake, and turbulent air currents follow our flight in a modern jet. Mathematicians and physicists believe that an explanation for and the prediction of both the breeze and the turbulence can be found through an understanding of solutions to the Navier-Stokes equations. Although these equations were written down in the 19th Century, our understanding of them remains minimal. The challenge is to make substantial progress toward a mathematical theory which will unlock the secrets hidden in the Navier-Stokes equations.

`http://www.claymath.org/millennium/Navier-Stokes_Equations/`

2 Qué es un fluido. Realidad e idealización

Un fluido es un **medio continuo**, es decir un agregado que se mueve (se deforma) en forma continua al transcurrir el tiempo, t, y forma un todo continuo en el espacio $\mathbf{x} = (x_1, x_2, x_3)$. Pensamos en tal medio como compuesto de partículas puntuales. No hay en ello ninguna objección de tipo matemático; en los últimos siglos las matemáticas se han inclinado frecuentemente por el estudio de magnitudes continuas frente a las discretas, y en tal hipótesis se basan la geometría diferencial, las ecuaciones diferenciales y una gran parte de los procesos estocásticos. Aunque no el cálculo numérico, evidentemente.

Ahora bien, la mecánica es una ciencia física que pretende describir el comportamiento de los cuerpos (sólidos, líquidos, gases o plasmas) y apoya por tanto su formulación matemática en la *experiencia* y la *teoría*. A este respecto, el concepto de medio continuo es una abstracción que, estrictamente hablando, está en contra de una teoría incontestable y ampliamente verificada, la teoría atómica, que describe la realidad a escala inferior al nanómetro (10^{-9}m; por ejemplo, el radio del átomo más pequeño, el de hidrógeno, mide alrededor de medio angström, $0,5 \times 10^{-10}$m). Un matemático a la usanza clásica tiene tendencia a resolver tal situación rechazando de plano al candidato que tropieza con una tal contradicción. Pues bien, la teoría de los fluidos no acepta tal rechazo. Se trata por el contrario de construir una teoría matemática que sirva de *modelo* a una parcela de

la Realidad; un modelo renuncia a la categórica exactitud y ha de ser juzgado por una parte desde el punto de vista *matemático*, en que se tiene en cuenta la belleza, extensión y profundidad de las matemáticas originadas; y por otra parte desde el punto de vista *físico*, por su eficacia en reflejar y en permitirnos *intuir* y *conocer* la realidad subyacente, *explicar* su funcionamiento observado y *predecir* su evolución futura. Hoy día, en el período dorado de la ciencia *computacional*, añadiríamos como esencial la capacidad de *calcular* y *controlar* eficazmente en base a este modelo.

La aproximación del medio continuo resulta ser tan efectiva que se olvida con frecuencia de que se trata de un modelo. Es con todo importante tener en cuenta las hipótesis de partida. Así, la consideración del fluido como un medio continuo se basa en que éste consiste en un agregado de partículas en movimiento caótico y que la distancia característica de este movimiento, que recibe el nombre técnico de "recorrido libre medio entre colisiones", λ, es mucho menor que las longitudes experimentales, que tomamos típicamente como mayores de 10^{-5} cm, de forma que sólo percibimos un cierto promedio de los procesos individuales entre partículas. Ahora bien, en ocasiones (piénsese en los gases enrarecidos de la materia interestelar) el recorrido libre medio puede ser mucho mayor, la hipótesis del continuo cesa de ser válida y no quedará más remedio que recurrir a teorías "más detalladas" que tengan en cuenta los movimientos moleculares (como la teoría cinética de gases). Precisamente, una de las líneas más activas de la investigación matemática actual es la obtención de las leyes del medio continuo como límite de las teorías cinéticas.

Una vez establecido que trabajamos en escalas muy superiores al recorrido libre medio de las partículas podemos olvidar el fino detalle de su movimiento individual y ver en torno a cada punto del espacio $\mathbf{x}$ y para cada instante t un *volumen elemental representativo*, δV, de tamaño *mesoscópico*[4], es decir mucho mayor que λ y mucho menor que las longitudes macroscópicas en las que deseamos trabajar. Este volumen elemental, que se denomina también *partícula fluida*, es considerado como un medio continuo y homogéneo; en él se define una velocidad media del movimiento de ese elemento, que será para nosotros la *velocidad* puntual en este punto e instante, $\mathbf{u}(\mathbf{x}, t)$. Para decirlo en forma más matemática, admitimos que existe un valor límite de los promedios cuando δV se hace muy pequeño en la escala intermedia, es decir es muy pequeño pero aún muy por encima de la escala atómica. Del mismo modo, se habla de las demás magnitudes macroscópicas, como la *densidad*, que es la masa por unidad de volumen en el sentido de límite antedicho, y la *presión*, que es la fuerza normal por unidad de área ejercida por el fluido sobre una superficie ideal inmersa en él o rodeándolo. Esta magnitud tiene una evidente explicación física, por ejemplo en un gas encerrado en un recipiente, como el efecto neto de las colisiones de las partículas individuales reales sobre la superficie de las paredes. A estas

[4]del griego, *mesos*, medio, *skopein*, mirar; intermedio entre macroscópico y microscópico.

tres magnitudes básicas se unirán otras en el curso del estudio, como *temperatura, energía interna, entropía, viscosidad,...* según el modelo sea más o menos complejo. La existencia de estos valores medios para las magnitudes fundamentales en cada partícula fluida es lo que constituye la hipótesis de continuidad del medio.

3 Ecuaciones fundamentales de los fluidos

Una vez identificado el tema de estudio, con sus aproximaciones admitidas y las variables que describen el sistema, el modelizador ha de proceder a escribir las leyes que relacionan a esas variables y nos permitirán predecir el funcionamiento del sistema. Siguiendo a Newton [26], estas leyes son diferenciales. Al involucrar el espacio y el tiempo son ecuaciones en derivadas parciales, EDPs. Siendo las variables que describen el sistema varias, se tratará de un sistema de ecuaciones. Finalmente veremos que, para poner la guinda al pastel, las ecuaciones son no lineales. Llegaremos pues a un Sistema de Ecuaciones en Derivadas Parciales de Evolución No Lineales, que son uno de los temas en donde está la frontera del saber matemático en nuestros días, tres siglos después de Newton.

Las leyes fundamentales son las siguientes: ley de conservación de la masa y ley de conservación de la cantidad de movimiento. Las introducimos a continuación. El lector que, en su prisa por conseguir el premio, se interese sólo por Navier-Stokes puede proceder a la sección 5.

3.1 Ley de Conservación de la Masa

Esta ley enuncia matemáticamente el principio según el cuál estamos describiendo un fenómeno de transporte de partículas que no se crean ni se destruyen. ¡Los cálculos que siguen son muy sencillos! La variable fundamental es la densidad $\rho(\mathbf{x}, t)$. En la formulación más geométrica, llamada lagrangiana, la ley dice

$$\frac{d}{dt}(\rho\, J) = 0, \tag{3.1}$$

donde J es el jacobiano de la deformación que sucede entre el momento $t = 0$ y el momento t en la situación de las partículas y d/dt indica la derivada a lo largo de las trayectorias que tiene como fórmula

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla = \frac{\partial}{\partial t} + \sum_{1}^{3} u_i \frac{\partial}{\partial x_i} \tag{3.2}$$

en función de las derivadas parciales usuales; $\mathbf{u} = (u_1, u_2, u_3)$ es la velocidad, que va a ser en un momento la variable fundamental. Así pues, si J es la medida de la *expansión de*

volumen a lo largo de una trayectoria, como la masa se conserva, (3.1) simplemente dice que densidad × volumen = constante.

En un artículo fundamental titulado "Principes généraux du mouvement des fluides" y publicado en 1755, Leonhard Euler tradujo esta ley de conservación de masa mediante el cálculo de la derivada en t de J:

$$\frac{dJ}{dt} = J.(\text{div } \mathbf{u}). \tag{3.3}$$

Créanlo o consulten la demostración en [33]. Con este lema se deduce que

$$\frac{d}{dt}(\rho J) = \frac{d\rho}{dt}J + \rho J(\nabla \cdot \mathbf{u}) = 0. \tag{3.4}$$

Escribimos $\nabla \cdot \mathbf{u} = \text{div } \mathbf{u}$. Como $J \neq 0$ por razones físicas evidentes, se tiene la versión en derivada total respecto al tiempo:

$$\frac{d\rho}{dt} + \rho \nabla \cdot \mathbf{u} = 0. \tag{3.5}$$

Aún podemos transformar la derivada total en parcial usando (3.2), llegando así a la fórmula: $\partial\rho/\partial t + \mathbf{u} \cdot \nabla\rho + \rho(\nabla \cdot \mathbf{u}) = 0$, que finalmente da

$$\boxed{\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0} \tag{3.6}$$

Esta es la forma llamada euleriana de la ley de conservación de masa. Nótese que es no lineal, pues contiene un término producto.

3.2 Ley de conservación de la cantidad de movimiento

La LCCM describe la dinámica del medio fluido. Comienza como un capítulo de la mecánica newtoniana afirmando que la variación de la cantidad de movimiento se debe a la acción de fuerzas,

$$\rho \frac{d\mathbf{u}}{dt} = \mathbf{f}_e(\mathbf{x}, t) + \mathbf{f}_c(\mathbf{x}, t). \tag{3.7}$$

No hay gran novedad en el término $\mathbf{f}_e$ que es la fuerza debida a campos externos, como el gravitatorio. La particularidad de los fluidos reside en la *fuerza de contacto* $\mathbf{f}_c$. Identificar sus componentes llevó siglo y medio y en la tarea participaron Johann y Daniel Bernoulli y L. Euler que describieron la componente de presión como

$$\mathbf{f}_p = -\nabla p.$$

Digno de mención es Augustin Cauchy que añadió el análisis del concepto de tensor de esfuerzos como forma general del efecto de contacto, en 1822. En los decenios que siguen varios prominentes científicos identificaron el efecto que debe añadirse al gradiente de

presión para obtener el conjunto de fuerzas de contacto. Entre ellos la posteridad ha seleccionado los nombres de Claude-Louis Navier que propuso en 1822 la fórmula del efecto viscoso [25], y sir George Gabriel Stokes, que culminó en 1845 la modelización con una deducción racional y matemáticamente elegante, al uso actual [29]. Según estos autores, en los fluidos usuales, llamados newtonianos, el esfuerzo de contacto toma la forma de una fuerza viscosa, de la forma

$$\mathbf{f}_v(\mathbf{x},t) = \lambda\nabla(\nabla\cdot\mathbf{u}) + \mu\Delta\mathbf{u}.$$

Este es un hito histórico de la modelización matemática de los problemas de la Física. Aparecen nuevas variables o parámetros cuyo significado físico ha de ser examinado: la presión $p(\mathbf{x},t)$ es una variable reconocida como relevante desde la Antigüedad. Los parámetros de λ y μ describen la viscosidad y podemos suponerlos en primera aproximación constantes medibles que dependen del fluido. Poniendo todo junto, llegamos a la ley

$$\boxed{\rho\left(\frac{\partial\mathbf{u}}{dt} + \mathbf{u}\cdot\nabla\mathbf{u}\right) = -\nabla p + \lambda\nabla(\nabla\cdot\mathbf{u}) + \mu\Delta\mathbf{u} + \mathbf{f}_e(\mathbf{x},t)} \tag{3.8}$$

Nótese que el término no lineal $\mathbf{u}\cdot\nabla\mathbf{u}$ proviene del paso de derivadas totales en tiempo a derivadas parciales (derivada de la función de función). Se llama término convectivo o de transporte y en la coordenada i vale

$$(\mathbf{u}\cdot\nabla\mathbf{u})_i = \sum_{j=1}^{3} u_j\frac{\partial u_i}{dx_j}. \tag{3.9}$$

Puede parecer una simple complicación técnica pero no es sólo eso: su nolinealidad, aunque sea solo cuadrática, es la razón de que las soluciones de la ecuación puedan en principio desarrollar singularidades y los matemáticos no han logrado decidir si este fenómeno ocurre o no tras incesantes esfuerzos teóricos y computacionales durante todo el pasado siglo. Explicar tal hecho será el punto culminante de estas notas en la sección 8.

El sistema formado por las dos leyes anteriores, (3.6) y (3.8), contiene cuatro ecuaciones, la LCM escalar y la LCCM vectorial, e implica a cinco variables, la densidad ρ, las tres componentes de la velocidad $\mathbf{u}$ y la presión p. Es pues indeterminado y necesita una o varias nuevas leyes que son las que hacen intervenir el balance de energía e involucran nuevas variables como la temperatura. La temperatura es fundamental en la descripción ajustada de muchos flujos reales, como los atmosféricos o marinos, como todo el mundo sabe. Concluimos pues que la descripción de los fluidos "reales" implica modelos de una notable envergadura matemática, aún hoy día difíciles de abordar, incluso a nivel computacional. Fenómenos climatológicos de gran interés, como *El Niño,* escapan aún casi completamente a la capacidad de explicación de los modelos matemáticos y más aún a la predicción. Referimos a los textos clásicos de Batchelor [3] o Landau-Lipshitz [19] para

una introducción a los sistemas completos de los fluidos reales. Quien se interese por un punto de vista más matemático puede consultar el libro de Chorin y Marsden [7] o el curso del autor [33][5].

4 Las ecuaciones de Euler. Fluidos perfectos

De las dificultades del "modelo matemático completo" de los fluidos eran conscientes los precursores, los Bernoulli y Euler, en el siglo XVIII, y propusieron hallar condiciones razonables que simplificaran el problema y lo redujeran a un problema susceptible de ser analizado matemáticamente. La reducción tomó dos vías: la primera, considerar fluidos que no se comprimen, llamados *fluidos incompresibles*; la segunda considerar fluidos que no sufren efectos viscosos, llamados *fluidos perfectos*.

4.1 Incompresibilidad

Examinemos la primera de estas simplificaciones. La condición de fluido incompresible nos dice que el factor de expansión J debe ser constante igual a 1 por lo que la LCM en su primera versión lagrangiana (3.1) dice que $d\rho/dt = 0$, mientras que el lema de Euler dice que div $\mathbf{u} = 0$. En total, la hipótesis de incompresibilidad nos lleva a mejorar la ley de conservación de masa en forma de dos condiciones

$$\frac{d\rho}{dt} = 0, \quad \nabla \cdot \mathbf{u} = 0. \tag{4.10}$$

Podemos simplificar aún un poco más la situación añadiendo la hipótesis de homogeneidad de la densidad. Basta con pedir homogeneidad espacial $\rho = \rho(t)$ para obtener de $d\rho/dt = 0$ que $\partial\rho/dt = 0$, o sea que ρ debe ser constante tanto en espacio como en tiempo. Ello es muy conveniente pues hace desaparecer ρ como variable del sistema, que pasa a tener tantas ecuaciones como incógnitas. Tan radical simplificación es con todo aceptable en los estudios hídráulicos, en oceanografía y muchas veces en las cuestiones atmosféricas.

4.2 Ecuaciones de Euler. Fluidos ideales

Tratemos ahora de simplificar las fuerzas de contacto. Euler supuso que podíamos partir del estudio de fluidos que son sensibles a la presión, pero no a los llamados esfuerzos cortantes, es decir, a lo que llamaríamos arrastre de capas contiguas. En ese caso ponemos simplemente $\mathbf{f}_c = \mathbf{f}_p = -\nabla p$ y la ecuación dinámica nos queda

$$\rho\left(\frac{\partial \mathbf{u}}{dt} + \mathbf{u} \cdot \nabla \mathbf{u}\right) + \nabla p = \mathbf{f}_e(\mathbf{x}, t). \tag{4.11}$$

[5]Nóte el lector curioso que, como se describe la última referencia, existen también diversos fluidos no newtonianos con viscosidades dadas por leyes más complejas, que son de gran aplicación en la industria moderna.

Si a esta ecuación vectorial unimos la incompresibilidad tenemos

$$\nabla \cdot \mathbf{u} = 0. \tag{4.12}$$

Éstas son las cuatro ecuaciones que gobiernan la evolución de las incógnitas $\mathbf{u}$ y p. Nos queda aún ligarlas con la evolución de ρ que viene regida por la ley $d\rho/dt = 0$. Pero es cómodo y usual suponer homogeneidad con lo que ρ es una constante. Un fluido perfecto, incompresible y homogéneo se llama *fluido ideal*. El sistema de Euler de los fluidos ideales consiste en las leyes (4.11) y (4.12). Se suele poner $\rho = 1$ para simplificar.

El sistema de Euler - SE en lo que sigue - es un sistema de ecuaciones en derivadas parciales de primer orden no lineal. El método más natural para los sistemas de tipo hiperbólico es el método de características, y ello se adapta bien a la ecuación dinámica salvo por la no linealidad. Es bien conocido que incluso las ecuaciones diferenciales ordinarias pueden generar discontinuidades en tiempo finito. En todo caso, como sucede en todas las EDPs, la resolución del SE exige de condiciones iniciales y de contorno adecuadas para que de pueda identificar una solución única.

Tosio Kato probó en 1967 el siguiente teorema de existencia y unicidad de solución global clásica en dimensión dos de espacio; global quiere decir que existe para todo $t > 0$, clásica que todas las derivadas que aparecen en las ecuaciones son funciones continuas y las ecuaciones se satisfacen en todo punto, [17].

Teorema 1 *Sea Ω un dominio acotado del plano con frontera Γ compuesta de $m+1$ curvas cerradas simples regulares Γ_0, $\Gamma_1, \cdots, \Gamma_m$, de las que Γ_0 rodea a todas las demás y éstas no se contienen unas a otras. Denotemos por Q_T el cilindro espacio-temporal $\Omega \times [0,T]$, $T > 0$. Sea $\mathbf{f}(\mathbf{x},t)$ un campo de fuerzas de la clase de Hölder $C^{1+\alpha,0}_{x,t}(\overline{Q}_T)$, para un $0 < \alpha < 1$, y sea $\mathbf{u}_0(\mathbf{x})$ un dato de velocidad inicial en la clase $C^{1+\alpha}(\overline{\Omega})$, que es además solenoidal, $\nabla \cdot \mathbf{u}_0 = 0$.*

Entonces existen un par de funciones, $\mathbf{u}(\mathbf{x},t)$, $p(\mathbf{x},t)$, que satisfacen el sistema SE *en el sentido clásico, siendo continuas en $\overline{Q}_T$, clausura de Q_T, tanto ellas como todas sus derivadas que aparecen en las ecuaciones. Además, $\mathbf{u}$ satisface la condición de contorno*

$$\mathbf{u} \cdot \mathbf{n} = 0 \qquad \text{en } \Gamma, \tag{4.13}$$

así como la condición inicial

$$\mathbf{u}(\mathbf{x},0) = u_0(\mathbf{x}) \qquad \text{para } \ \mathbf{x} \in \Omega. \tag{4.14}$$

Por último, $\mathbf{u}$ es única y p es única salvo adición de una función arbitraria del tiempo.

Nos interesa saber si un resultado similar es cierto en tres dimensiones de espacio. Se sabe que el resultado es cierto si admitimos que el intervalo de definición de la solución

en el tiempo sea pequeño (de tamaño dependiendo de los datos). Es lo que se llama problema local en el tiempo. Se plantea entonces el resultado de existencia y unicidad de una solución clásica definida para todo $t > 0$, es decir, una *solución global*. Queda así formulado el **Problema Abierto de las Ecuaciones de Euler.**

Aunque este problema no forma parte de la Lista de Clay, es considerado por la comunidad matemática de tanto interés como el de las ecuaciones de Navier-Stokes que discutiremos a continuación. En este momento no se tiene una hipótesis mayoritariamente compartida sobre una u otra de las opciones del problema abierto.

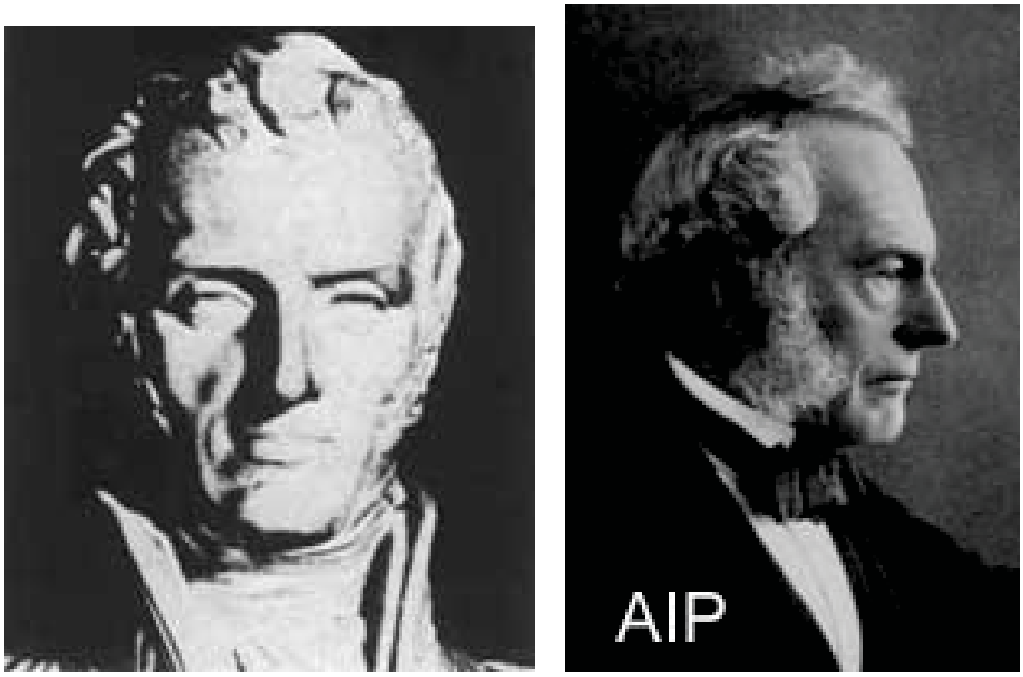
5 Las ecuaciones de Navier-Stokes. Fluidos viscosos

La extremada simplificación inherente a los fluidos perfectos, que no admiten el arrastre lateral (¡contradiciendo a Newton y a la realidad!), fue notada desde sus comienzos por los precursores y puesta muy de relieve por D'Alembert. Aunque según esa teoría, un barco flotaría en el agua, ¡sin embargo los aviones no volarían, grave defecto que retrasó el comienzo de la ciencia aeronáutica! Remediar esta situación con "un modelo de nivel superior" nos ha llevado de la mano de Cauchy, Navier y Stokes a considerar en la sección 3 fluidos más realistas que incluyen efectos de viscosidad. Cuando se impone la incompresibilidad y se supone $\rho = 1$, el sistema de Navier-Stokes SNS, toma la forma

$$\boxed{\begin{array}{c} \dfrac{\partial \mathbf{u}}{dt} + \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla p + \nu \Delta \mathbf{u} + \mathbf{f}_e(\mathbf{x}, t) \\ \nabla \cdot \mathbf{u} = 0. \end{array}} \tag{5.15}$$

Siguiendo la costumbre, usamos la notación $\nu = \mu/\rho$ para el parámetro que caracteriza la propiedad de viscosidad de cada fluido "real". El operador $\mathbf{u} \cdot \nabla \mathbf{u}$ viene dado en coordenadas por la expresión (3.9) con suma de $j = 1$ a n ($n = 3$ en el problema físico). Este es el sistema de EDPs al que queríamos llegar y que figura en el anuncio del Instituto Clay. Necesitamos datos adicionales, iniciales y /o de contorno, dependiendo del dominio donde se plantee el problema y del tipo de datos que tengan interés para las aplicaciones.

En el último siglo y medio, estas ecuaciones han pasado el test de la aplicación siendo utilizadas por físicos e ingenieros con notable éxito en muy diversos campos, entre ellos la hidráulica, la meteorología y la aeronáutica, y su rango de validez está bien establecido. Pertenecen ya, junto a las ecuaciones de Newton, Schrödinger y Maxwell, a las ecuaciones básicas de la Física.



Claude-Louis Navier (1785-1836) y George Gabriel Stokes (1819-1903)

5.1 Problemas matemáticos

No hay ninguna objección matemática a que el problema de construcción de soluciones y de su unicidad y regularidad se plantee dentro de un espacio $\mathbb{R}^n$ de un número de dimensiones $n \geq 1$, siendo el caso $n = 3$ el interesante para la ciencia aplicada y el $n = 1$ trivial.

Contexto del Problema de Cauchy[6]. El dominio espacial es todo $\mathbb{R}^3$, o en general, todo $\mathbb{R}^n$, $n = 1, 2, ...$ Se plantean pues las ecuaciones (5.15) para $\mathbf{x} \in \mathbb{R}^n$, $t > 0$. Se añaden datos iniciales

$$\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^n. \tag{5.16}$$

Se supone que $\mathbf{u}_0$ es un campo vectorial de clase C^∞ y de divergencia nula en $\mathbb{R}^n$. El campo $\mathbf{f}$, también regular, representa la acción de las fuerzas externas como la gravedad, pero no es esencial para el tema que nos ocupa en este momento. Dado que el problema se plantea en el espacio infinito, razones de coherencia con la fisica y de comodidad matemática sugieren someter a los datos y a la solución a condiciones de decrecimiento rápido del tipo

$$|\partial_x^\alpha u_{0i}(\mathbf{x})| \leq C_{\alpha,K}(1 + |\mathbf{x}|)^K \qquad \forall \alpha, \forall K \tag{5.17}$$

$$|\partial_x^\alpha \partial_t^m \mathbf{f}_i(\mathbf{x})| \leq C_{\alpha,m,K}(1 + |\mathbf{x}| + t)^K \qquad \forall \alpha, \forall m, \forall K \tag{5.18}$$

[6]Seguimos aquí en lo esencial la exposición de Charles Fefferman en la presentación del problema para el Clay Institute, [11].

Declaramos admisibles las soluciones clásicas $\mathbf{u}, p \in C^\infty(\mathbb{R}^n \times (0,T))$ tales que

$$\int u^2(\mathbf{x},t)\, d\mathbf{x} < \infty \qquad \forall t,$$

lo que significa que la energía cinética es acotada, una condición muy razonable desde el punto de vista de la mecánica y también de las EDPs.

El **problema de decisión** se formula como sigue: decidir cuál de los dos enunciados responde a la realidad.

(PSNS-1) Problema de existencia y regularidad. *Tomamos ν constante positiva y $\mathbf{f} = 0$ y suponemos que $\mathbf{u}_0$ satisface las condiciones de regularidad, divergencia nula y decaimiento rápido en el infinito enumeradas. Demostrar que existen funciones $\mathbf{u}$ y $p \in C^\infty(\mathbb{R}^n \times (0,\infty))$, $\mathbf{u}$ de energía finita, que resuelven el sistema* SNS *en el sentido clásico y $\mathbf{u}$ toma el dato inicial $\mathbf{u}_0$.*

(PSNS-2) Problema de colapso de la solución. *Tomamos $\nu > 0$. Encontrar un dato inicial $\mathbf{u}_0$ y una función $\mathbf{f}$ con las condiciones enumeradas, tales que no existe una solución clásica $(\mathbf{u}, p)$ del sistema* SNS *con condiciòn inicial* (5.16).

A este problema alternativo se une el problema de unicidad:

(PSNS-3) Problema de unicidad. *Dada una teoría de existencia de soluciones físicamente aceptable (como las soluciones débiles de Leray de la próxima sección o las soluciones obtenidas computacionalmente), demostrar que la solución es única durante todo el tiempo de existencia.*

Contexto de Cauchy-Dirichlet. El dominio espacial es un abierto conexo Ω de $\mathbb{R}^3$, o en general, de $\mathbb{R}^n$, $n = 1, 2, ...$ En general se toma acotado y de borde regular. Se imponen condiciones de no deslizamiento en el borde, $\mathbf{u} = 0$ en $\Gamma = \partial\Omega$. Problemas similares a los anteriores se plantean.

Contexto con condiciones periódicas. En un intento de enfocar la atención de los investigadores sobre las dificultades esenciales el siguiente problema más artificial se admite como marco del desafío: se toma como dominio un cubo $\Omega = \Pi_i(0, ll_i)$ y se supone que $\mathbf{u}_0$ y $\mathbf{f}$ son funciones suaves en el cierre de Ω que se extienden por periodicidad a todo $\mathbb{R}^n$. Se pide entonces que las funciones extendidas sean C^∞. Se buscan soluciones clásicas y periódicas. Los problemas (PSNS-1) y (PSNS-2) se formulan mutatis mutandis.

5.2 Resultados parciales

Se trata pues de un Problema de Decisión. Lo mismo que en el caso de las Ecuaciones de Euler, el problema ha sido decidido en dimensión $n = 2$, y la respuesta es: la opción (PSNS-1) es cierta, ver Ladyzhenskaya [18].

El problema completo en $n = 3$ ha resistido todo los intentos hasta ahora. Centrándose en el problema de Cauchy que es el más interesante, varios casos parciales están decididos:

(i) Si sustituimos $T = \infty$ por un tiempo pequeño entonces (PSNS-1) es cierto. Ello fue ya demostrado por Jean Leray en sus artículos fundamentales de 1933-34.

(ii) Si $\mathbf{u}_0$ es pequeño en un sentido a precisar, o si es fuertemente oscilante, (PSNS-1) es cierto.

(iii) Si (PSNS-2) fuese cierto y una solución no pudiese ser continuada más allá de un tiempo $T > 0$, entonces la velocidad se debe hacer infinita cuando $t \to T$.

Este fenómeno se llama técnicamente *blow-up* en inglés o explosión en castellano y de él hablaremos en detalle en la Sección 7, pues es nuestra estrella invitada matemática.

6 Soluciones débiles del sistema SNS. La obra de Leray

Jean Leray hizo en los años 1933-34 una contribución fundamental al problema SNS. Tras obtener existencia de solución clásica para datos regulares durante un pequeño intervalo de tiempo $(0, T)$, se encontró con el problema de que no le era posible controlar *a priori* el crecimiento de la velocidad y sus derivadas al avanzar el tiempo, lo cual arruinaba la esperanza de construir una solución global. Enfrentado a esta dificultad, optó por el procedimiento ya seguido por Hilbert en el tratamiento del problema de Dirichlet para el operador laplaciano y planteó el problema en el marco de las llamadas *soluciones débiles* en los espacios de energía que hoy llamamos de Sóbolev, ver [20, Le3].

6.1 Solución débil

La idea es simple: si $\mathbf{u}$ es una solución clásica de SNS definida en $Q = \mathbb{R}^n \times (0, T)$ y $\theta = (\theta_i(\mathbf{x}, t))$ es un campo solenoidad, de clase C^2 y con soporte compacto, multiplicando la ecuación de Navier-Stokes por θ e integrando por partes se tiene

$$\int\int_Q \{\langle \mathbf{u}, \frac{\partial \theta}{\partial t}\rangle + \sum_i u_i u_j \frac{\partial \theta_i}{\partial x_j} + \nu\langle \mathbf{u}, \Delta\theta\rangle + \langle \mathbf{f}, \theta\rangle\}\, d\mathbf{x}dt = 0. \tag{6.19}$$

Denotamos por $\langle \mathbf{u}, \mathbf{v}\rangle$ el producto escalar en $\mathbb{R}^n$ para más claridad. Nótese que el término de presión desaperece, $\int\int p(\nabla \cdot \theta)\, d\mathbf{x}dt = 0$, dado que hemos elegido las funciones test θ de divergencia nula. Podemos ahora testear la condición de divergencia nula de forma similar multiplicando la ecución $\nabla \cdot \mathbf{u} = 0$ por una función test φ de clase C^2 y con soporte compacto para dar

$$\int\int_Q \langle \mathbf{u}, \nabla_x \varphi\rangle\, d\mathbf{x}dt = 0. \tag{6.20}$$

Definición. Toda función $\mathbf{u} = (u_i)$ localmente integrable en espacio y tiempo y tal que (6.19), (6.20) se satisfacen para todo campo vectorial test θ y escalar φ con las propiedades

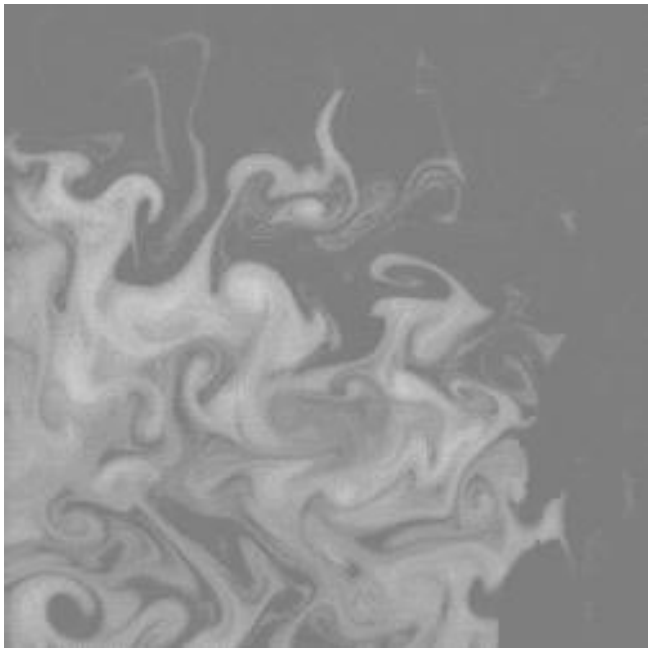
enumeradas, se llama una solución débil del sistema SNS[7].

Por lo tanto una solución clásica es débil, pero una solución débil sólo satisface una serie de tests y podría ser un objeto más general.

6.2 Programa "débil"

A partir de años 50 del pasado siglo, y gracias en parte a la obra previa de Leray y a causa del trabajo sobre el SNS, los matemáticos han abrazado con entusiamo la mentalidad de las soluciones débiles y han contagiado este entusiasmo al mundo de la computación a través de los elementos finitos. Una idea fundamental en las EDPs del siglo XX ha sido pues la de construir soluciones débiles de una serie de problemas; se demuesta luego la unicidad de tales objetos matemáticos; en un tercer paso se trata de probar (mediante técnicas a veces muy sofisticadas) que tales soluciones débiles son en realidad soluciones en el sentido clásico. El programa no siempre cumple este objetivo, pues en algunos casos las soluciones débiles pueden no ser regulares (contiene conjuntos de singularidad) o no ser únicas (como sucede en las leyes de conservación de la dinámica de gases estudiadas por Riemann); en ese caso es preciso añadir condiciones de selección (condiciones de entropía en el caso citado).

6.3 Singularidades y turbulencia



Turbulencias en un fluido

Leray fue capaz de construir soluciones débiles del sistema SNS de energía cinética finita, $\int \mathbf{u}^2(\mathbf{x}, t)\, d\mathbf{x}$ finito para todo t, pero la regularidad de tales objetos se resistió en dimensión $n = 3$. Tampoco fue posible probar en dimensión $n = 3$ la unicidad de tales soluciones, que es *otro problema fundamental abierto*. La presencia de singularidades fue conjeturada por Leray y le sirvió como posible explicación del fenómeno físico de la **turbulencia**. Según esta hipótesis, incluso para datos regulares las soluciones en tres dimensiones pueden desarrollar en un tiempo finito singula-

[7]Esta formulación es la usual hoy día, actualizada respecto a la original de Leray.

ridades en la forma de puntos donde la vorticidad $\overrightarrow{\omega} = \mathrm{rot}\,(\mathbf{u})$ se hace infinita.

La teoría de soluciones débiles ha sido luego elaborada por matemáticos como E. Hopf, O. A. Ladyzhenskaya, J. Serrin, J. L. Lions, G. Prodi, T. Kato, R. Temam y otros muchos[8].

7 Qué es un problema de explosión o blow-up

Analicemos un momento el problema que se nos presenta con las posibles singularidades en un marco más general, lo cual permitirá obtener una visión más amplia de la interrelación de los problemas de Euler y Navier-Stokes con otros problemas.

7.1 El mundo no lineal y sus peculiaridades

Numerosos procesos de las ciencias aplicadas se modelan por medio de sistemas de ecuaciones de evolución que involucran operadores diferenciales como los arriba vistos. El tratamiento matemático tiene como objeto obtener problemas bien propuesto, para lo cual se añaden datos iniciales y de contorno, se suministra un adecuado marco funcional y eventualmente se imponen condiciones de compatibilidad. Problema bien propuesto quiere decir para el matemático que existe una solución en el marco descrito, es única y depende continuamente de los datos (estabilidad).

Las teorías matemáticas clásicas involucran operadores lineales para los que existe hoy día una enorme teoría matemática desarrollada en el marco del Análisis Funcional. Por suerte existen importantes teorías físicas que se modelan en forma lineal en su rango de aplicación usual, como son la teoría electromagnética y la teoría de propagación del calor. Sin embargo, muchos otros modelos importantes son no lineales, y entre ellos se cuentan la teoría de la relatividad y la de los fluidos. Se ha comprobado que tales teorías tienden a una dificutad matemática mayor y que exhiben un número de propiedades que no se dan en los modelos lineales. Además, se ha visto que estos nuevos fenómenos o propiedades reflejan aspectos esenciales de la realidad que se pretendía describir, por lo que volver la vista al mundo lineal, más sereno y regular, no resuelve nada, salvo como primera aproximación.

Una de las más notables propiedades que distinguen el mundo no lineal es precisamente la que nos ocupaba al final de la sección precedente, a saber, la posibilidad de que datos

[8]Franceses, italianos, alemanes, ingleses, irlandeses como Stokes, norteamericanos, japoneses y rusos; pueblos ilustres a través de sus científicos. Es triste observar cómo hasta hace nada las matemáticas españolas no aparecían en la ciencia mundial. El autor confía que en una futura Lista de los Fluidos del Siglo XXI la situación cambie sustancialmente.

perfectamente regulares den lugar a una evolución que (i) está bien propuesta en el sentido matemático para tiempos pequeños, (ii) en un determinado tiempo la solución clásica deja de existir pues se genera una singularidad. Nótese que pueden existir singularidades en problemas lineales, pero éstas deben ser ya patentes en la regularidad de los datos o coeficientes adecuadamente examinados. En cambio, en los problemas no lineales, las singularidades surgen del mecanismo interno a la ecuación, incluso a partir de datos y coeficientes extremadamente regulares.

La forma más simple en que se observan singularidades espontáneas en un problema de evolución es aquel en que la variable o variables tienden a infinito cuando el tiempo se acerca a un valor finito $T > 0$. Esto es lo que se llama fenómeno de *blow-up* o explosión.

7.2 Blow-up para ecuaciones diferenciales ordinarias

El contexto más elemental en que se observa el blow-up es la teoría de ecuaciones diferenciales ordinarias (EDOs). Y el ejemplo más simple, y a la vez enormemente representativo lo suministra la *ecuación del crecimiento cuadrático*: se considera una variable real escalar $Y = Y(t)$ que obedece a la ley

$$Y_t = Y^2, \;\; t > 0; \quad Y(0) = a. \tag{7.21}$$

Si el dato inicial es $a > 0$, se sigue immediatamente que existe una única solución definida en un intervalo temporal $0 < t < T$ con $T = 1/a$, y dada por la fórmula

$$Y(t) = \frac{1}{T - t}. \tag{7.22}$$

Vemos pues que la evolución está descrita por una función regular para $t < T$. Cuando $t \to T^-$ (límite por la izquierda), vemos que la solución explota, $Y(t) \to \infty$. No solo eso, también sabemos cual es la tasa de crecimiento cerca de la explosión, $Y(t) = O((T-t)^{-1})$. Este será para nosotros el *ejemplo elemental de explosión*.

Arrancando de este ejemplo afortunado (por simple y representativo), los matemáticos han extendido el concepto de explosión y han realizado estudios de cuándo, cómo y dónde sucede en toda una serie de diferentes problemas y contextos de la matemática y la ciencia aplicada. En general se trata de que una o más de las variables de un sistema se hagan infinitas al acercarse a un tiempo finito T, el tiempo de explosión, que impide que la solución pueda ser continuada globalmente en el tiempo, al menos en el sentido original. En algunos casos la explosión sucede en una derivada de una variable del sistema. Una tal explosión de derivada se admite como "explosión del sistema" si impide la demostración de existencia de solución más allá de T.

Una primera extensión del ejemplo elemental de explosión la proporcionan las EDOs de la forma $Y_t = Y^p$ con $p > 1$ (caso superlineal), que el lector no tendrá dificultad en

integrar. Pero no explota si $0 < p < 1$, caso en el que problema es la falta de unicidad. Nosotros nos encontraremos más adelante con el caso $p = 3$ en el estudio del SNS. Más generalemte, podemos considerar la ODE

$$Y_t = f(Y), \tag{7.23}$$

con f positivo y continuo; la condición de Osgood

$$\int_1^\infty ds/f(s) < \infty \tag{7.24}$$

es necesaria y suficiente para que la solución que arranca con dato inicial positivo $Y(0) = a$ explote en tiempo finito. Siguiendo el camino de generalización, podemos considerar sistemas $Y_t = f(t, Y)$ con variable vectorial $u \in \mathbb{R}^n$. Entonces, tenemos blow-up por las mismas razones si f es superlineal con respecto a Y para $|Y|$ grande.

Resumiendo, el estudio de las ODEs proporciona ejemplos y técnicas básicos para la teoría matemática de los fenómenos explosivos. Cuando se extienden a otros ámbitos, no siempre se encontrarán fórmulas explícitas como las anteriores, pero el matemático halla al menos útiles suficientes para resolver los problemas de decisión e incluso para estimar cuándo, dónde y cómo explota la solución. Tal tarea se ha cumplido con éxito en los últimos decenios en un número importante de ecuaciones y sistemas de ecuaciones en derivadas parciales de la Física Matemática. Desgraciadamente, no es aún el caso en los problemas de Euler y Navier-Stokes.

8 Blow-up o no blow-up, esa es la cuestión

Un buen artículo de matemáticas ha de tener algún "cálculo de verdad". Veamos ahora cuales son los cálculos básicos de la teoría débil para el SNS y porqué las cosas se tuercen en dimensión tres y no en dos. Una teoría débil suele basarse en tres ingredientes: un buen marco funcional, un procedimiento de aproximación por problemas resolubles y estimaciones a priori sobre el comportamiento de las eventuales soluciones y de sus aproximantes que permitan pasar al límite en el proceso.

8.1 Marco funcional

El marco se establece primero para la variación espacial a tiempo fijo. Las velocidades han de ser campos vectoriales $\mathbf{u}(t) \in L^2(\mathbb{R}^n)$ con gradientes $\nabla\mathbf{u}(t) \in L^2(\mathbb{R}^n)^n$ y divergencia nula (en el sentido débil (6.20))[9]. Tales funciones vectoriales forman el espacio V. Si no pedimos que los gradientes sean funciones sino solo quizá distribuciones tenemos el espacio

[9]Siguiendo la tradición en problemas de evolución escribimos a veces $\mathbf{u}(t)$ para abreviar $\mathbf{u}(\mathbf{x}, t)$.

H. Nótese que V es un subespacio cerrado del espacio de Sobolev clásico $(H^1(\mathbb{R}^n))^n$, mientras $H \subset (L^2(\mathbb{R}^n))^n$. Ambos espacios V y H son espacios de Hilbert. Introducimos las notaciones

$$(\mathbf{u}, \mathbf{v}) = \int \langle \mathbf{u}, \mathbf{v} \rangle \, dx, \qquad ((\mathbf{u}, \mathbf{v})) = \sum_i \int \left\langle \frac{\partial \mathbf{u}}{\partial x_i}, \frac{\partial \mathbf{v}}{\partial x_i} \right\rangle \, dx, \tag{8.25}$$

para elementos de V (y de H en su caso). Usamos además las notaciones

$$|\mathbf{u}|^2 = (\mathbf{u}, \mathbf{u}), \qquad \|\mathbf{u}\|^2 = ((\mathbf{u}, \mathbf{u})). \tag{8.26}$$

Podemos ya formular las soluciones débiles dentro de estos espacios. Así, multiplicando formalmente la ENS por una función vectorial $\mathbf{v} \in V$ e integrando queda la identidad variacional, básica en lo que sigue:

$$\left(\frac{d\mathbf{u}(t)}{dt}, \mathbf{v} \right) + \nu((\mathbf{u}(t), \mathbf{v})) + b(\mathbf{u}(t), \mathbf{u}(t), \mathbf{v}) = (\mathbf{f}(t), \mathbf{v}) \tag{8.27}$$

Aquí el término de transporte $\mathbf{u}(t) \cdot \nabla \mathbf{u}(t)$ de la ecuación de NS da de sí tras integración por partes $b(\mathbf{u}(t), \mathbf{u}(t), \mathbf{v})$, un término no lineal que hemos de vigilar. Hay varios resultados técnicos que estiman su influencia. He aquí el lema básico:

Lema 2 *La fórmula*

$$b(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \sum_i \int u_i \frac{\partial v_j}{\partial x_i} w_j \, dx \tag{8.28}$$

define una forma trilineal acotada en $V \times V \times V$.

8.2 Problemas aproximados. Problema de Stokes.

Son diversos y forman la parte técnica de la disciplina. Referimos al lector a los libros de Temam [30, 31] o de Constantin-Foias [9]. El punto de apoyo más importante consiste en hacer un anális completo del problema reducido en que se suprime el término no lineal y la ecuación queda en la forma

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla p = \nu \Delta \mathbf{u} + \mathbf{f}_e(\mathbf{x}, t), \qquad \nabla \cdot \mathbf{u} = 0. \tag{8.29}$$

Este problema se llama Problema de Stokes. He aquí el resultado que se obtiene utilizando los métodos del Análisis Funcional (a final de cuentas, el teorema de Lax-Milgram).

Teorema 3 *Para todo dato inicial* $\mathbf{u}_0 \in H$ *y toda función* $\mathbf{f} \in L^2(0, T : L^2(\mathbb{R}^n)^n)$ *existe una única* $\mathbf{u} \in L^2(0, T : V)$ *que es solución débil del sistema de Stokes y tal que* $\mathbf{u}$ *es continua en* $t \in [0, T)$ *con valores en* V' *(el espacio dual de* V*) y se toma el dato inicial* $\mathbf{u}_0$.

Una vez obtenido $\mathbf{u}$ no es difícil obtener p salvo constantes espaciales. El estudio estacionario previo a la resolución del problema pasa por definir el laplaciano como un operador A autoadjunto, no acotado y no negativo con dominio

$$D(A) = \{\mathbf{u} \in H : \Delta\mathbf{u} \in H\}.$$

8.3 Estimaciones no lineales

Las dificultades están pues en la parte lineal de la formulación variacional descrita en (8.27). Veamos las estimaciones a nivel formal.

• Poniendo $\mathbf{v} = \mathbf{u}(t)$ en la formulación débil tenemos

$$\frac{1}{2}\frac{d}{dt}|\mathbf{u}(t)|^2 + \nu\|\mathbf{u}(t)\|^2 = (\mathbf{f}(t), \mathbf{u}(t)) \leq \|\mathbf{f}\|_{V'}\|\mathbf{u}(t)\|, \tag{8.30}$$

pues se demuestra fácilmente que $b(\mathbf{u}(t), \mathbf{u}(t), \mathbf{u}(t)) = 0$. Integrando (8.30) se obtiene la ley de conservación de la energía, que para $\mathbf{f} = \mathbf{0}$ toma en la forma usual

$$\frac{1}{2}|\mathbf{u}(t)|^2 + \nu\int_0^T \|\mathbf{u}(t)\|^2\, dt = \frac{1}{2}|\mathbf{u}(0)|^2, \tag{8.31}$$

donde el término integral describe la energía disipada por el sistema. El término no lineal $\mathbf{u}\cdot\nabla\mathbf{u}$ no tiene pues influencia a este nivel y obtenemos un control a priori para las soluciones

$$|\mathbf{u}(t)| \leq C_1, \qquad \int_0^T \|\mathbf{u}(t)\|^2\, dt \leq C_2, \tag{8.32}$$

con constantes dependientes de las normas de $\mathbf{u}_0$ y $\mathbf{f}$ pero no de T. Leray utilizó estas estimaciones para construir su teoría débil.

• Trabajando de nuevo formalmente, suponemos que $\mathbf{u}(t) \in D(A)$ y multiplicamos por $\Delta\mathbf{u}(t)$. Tras integrar por partes obtenemos esta vez

$$\left(\frac{d\mathbf{u}(t)}{dt}, A\mathbf{u}(t)\right) + \nu((\mathbf{u}(t), A\mathbf{u}(t))) + b(\mathbf{u}(t), \mathbf{u}(t), A\mathbf{u}(t)) = (\mathbf{f}(t), A\mathbf{u}(t)). \tag{8.33}$$

Tras algunas manipulaciones la relación puede ser escrita como

$$\frac{1}{2}\frac{d}{dt}\|\mathbf{u}(t)\|^2 + \nu|A\mathbf{u}(t)|^2 + b(\mathbf{u}(t), \mathbf{u}(t), A\mathbf{u}(t)) \leq \frac{1}{\nu}|\mathbf{f}|^2 + \frac{\nu}{4}|A\mathbf{u}(t)|^2. \tag{8.34}$$

¡El término no lineal no desaparece ahora!

• **Final feliz para $n = 2$.** Seguimos en dimensión dos estimándo el término no lineal mediante las inmersiones de Sobolev en la forma:

Lema 4 *Si $n = 2$ y $\mathbf{u} \in V$, $\mathbf{v} \in D(A)$, $\mathbf{w} \in H$, entonces*

$$|b(\mathbf{u}, \mathbf{v}, \mathbf{w})| \leq C|\mathbf{u}|^{1/2}\|\mathbf{u}\|^{1/2}\|\mathbf{v}\|^{1/2}|A\mathbf{v}|^{1/2}|\mathbf{w}|$$

Usando este lema, llegamos a

$$\frac{d}{dt}\|\mathbf{u}(t)\|^2 + \frac{3}{2}\nu|A\mathbf{u}(t)|^2 \leq \frac{2}{\nu}|\mathbf{f}|^2 + C|\mathbf{u}(t)|^{1/2}\|\mathbf{u}(t)\||A\mathbf{u}(t)|^{3/2} \tag{8.35}$$

Usando la desigualdad de Young acotamos el último término por

$$\frac{2}{\nu}|A\mathbf{u}(t)|^2 + C'|\mathbf{u}|^2\|\mathbf{u}\|^4.$$

Llegamos pues a

$$\frac{d}{dt}\|\mathbf{u}(t)\|^2 + \nu|A\mathbf{u}(t)|^2 \leq \frac{2}{\nu}|\mathbf{f}|^2 + C''|\mathbf{u}(t)|^2\|\mathbf{u}(t)\|^4. \tag{8.36}$$

Este es el momento importante: la estimación obtenida, junto con las ya obtenidas (8.32), permite probar mediante una técnica llamada "desigualdad de Gronwall" que

$$\|\mathbf{u}(t)\| \leq C_3, \qquad \int_0^T |A\mathbf{u}(t)|^2\,dt \leq C_4, \tag{8.37}$$

con constantes dependientes de las normas de $\mathbf{u}_0$ y $\mathbf{f}$, pero no de T. La primera de ellas es una estimación de la norma L^2 de los gradientes de velocidad (en particular de la vorticidad) que es uniforme en el tiempo. La segunda controla derivadas segundas en L^2 del espacio-tiempo.

Estas estimaciones son suficientes para actuar en los problemas aproximados, que usualmente son aproximaciones finito-dimensionales del tipo llamado Galerkin, pasar al límite y demostrar la existencia de solución global regular para datos regulares.

Es muy interesante ver un segundo como se aplica el "truco Gronwall". Llamemos $Y(t) = 1 + \|\mathbf{u}(t)\|^2$. Entonces la estimación (8.36) implica que

$$\boxed{Y'(t) \leq C(t)Y(t)} \tag{8.38}$$

una inecuación diferencial ordinaria (IDO) de tipo cuadrático, muy parecida a nuestro ejemplo elemental de explosión $u' = u^2$, salvo por el signo de desigualdad, que va en la dirección adecuada y no causa problemas, y por el coeficiente

$$C(t) = C''|\mathbf{u}(t)|^2\|\mathbf{u}(t)\|^2,$$

del cual sabemos que es integrable en el tiempo debido a (8.32). Gronwall nos dice que en esas circunstancias $Y(t)$ está acotado superiormente independientemente del tiempo.

• **Final no feliz para $n = 3$.** En dimensión tres estimamos el término no lineal mediante las inmersiones de Sobolev de forma menos efectiva:

Lema 5 *Si $n = 3$ y $\mathbf{u} \in V$, $\mathbf{v} \in D(A)$, $\mathbf{w} \in H$, entonces*

$$|b(\mathbf{u}, \mathbf{v}, \mathbf{w})| \leq C\|\mathbf{u}\|\|\mathbf{v}\|^{1/2}|A\mathbf{v}|^{1/2}|\mathbf{w}|$$

Manipulaciones como las anteriores conducen a una IDO de la forma

$$\boxed{Y'(t) \leq Y^3(t)} \tag{8.39}$$

que no excluye la explosión en tiempo finito. Omitimos los detalles pues nos parece que ya hemos abusado de la atención del lector, pero cf. [31].

9 En caso de haber explosión

Enfrentados a la posibilidad de explosión en tiempo finito, los investigadores han querido saber pormenores sobre tal fenómeno si llegara a producirse. Hay dos tipos de resultados relacionados con este tema

9.1 Geometría del conjunto singular

Se supone que $\mathbf{f}$ es regular (o nula) y $\mathbf{u}_0$ regular y se define el *conjunto singular* E de la solución débil $\mathbf{u}$ como el conjunto de puntos $(\mathbf{x}, t)$ tales que $\mathbf{u}$ no es acotada en ningún entorno de $(\mathbf{x}, t)$. En caso $\mathbf{u}$ fuese acotada en un entorno no es difícil probar que $\mathbf{u}$ es C^∞ en ese entorno.

En 1976 V. Scheffer introdujo ideas de teoría geométrica de la medida para estimar el conjunto E. Esta estimación fue mejorada por L. A. Caffarelli, R. Kohn y L. Nirenberg en 1982. Definen medidas de Hausdorff parabólicas $\mathcal{P}_r$ en el espacio-tiempo y concluyen que para toda solución con unas condiciones de crecimiento razonables

$$\mathcal{P}_{5/3}(E) = 0. \tag{9.40}$$

Existe una prueba simplificada por Lin [21] de este resultado, que muchos consideran el más importante tras los trabajos de Leray. Dado que la medida parabólica cuenta el tiempo como espacio al cuadrado (9.40) impide singularidades que se propaguen a lo largo de líneas de la forma $\mathbf{x} = \varphi(t)$. Concluimos: un conjunto singular, de existir, es un conjunto ralo y raro.

9.2 Formas y tasas de divergencia

Otra posibilidad de dar luz al problema es la de investigar qué pasaría en caso de explosión con diversas cantidades relevantes, como la vorticidad, y cuáles podrían ser las tasas y perfiles de explosión. En el primer aspecto los trabajos más notables se refieren al problema gemelo de Euler; podemos citar el famoso artículo de Beale-Kato-Majda en 1984 [4], que dice que el supremo espacial de la vorticidad ha de diverger cuando se integra en tiempo.

En cuanto a los perfiles, tras mucho especular con perfiles autosemejantes, hoy se buscan perfiles mucho más complejos.

10 Explosión para Ecuaciones en Derivadas Parciales

10.1 Blow-up y combustión

El estudio del blow-up no se ha encontrado con tantas dificultades en otros tipos de EDPs y ha adquirido notable madurez en algunas áreas. Quizá el área más estudiada sean las ecuaciones de Reacción Difusión, que en la forma más simple se escriben

$$u_t = \Delta u + f(u), \tag{10.41}$$

donde f es como en la sección 7. Añadimos al fenómeno evolutivo de la ODE la complicación de la estructura espacial (la dependencia de $\mathbf{x}$). La motivación aplicada viene de una disciplina relacionada con los fluidos, la Teoría de la Combustión. Refiero al lector a los libros de Bebernes and Eberly [5] y Samarskii et al. [27], o al reciente artículo survey de V. Galaktionov y el autor [12], donde se organiza el examen el campo desde el punto de vista de la llamada "Lista de Preguntas": en qué problemas se da el blow-up, cuándo ocurre, dónde, cómo, a qué velocidad diverge la solución, es posible continuar la solución y el problema de describir las avalanchas térmicas. A lo que se añade el problema computacional con sus problemas de estabilidad anexos.

10.2 Blow-up y el Problema Clay número 3

Nuestro último tema se refiere a un desarrollo reciente y más bien espectacular. El Problema 3 de la lista Clay llama a resolver la conjetura de Poincaré sobre la estructura de las 3-variedades. Una forma de ataque en el intento de clasificación ha sido el hacer evolucionar superficies riemannianas mediante un "motor" relacionado con su curvatura. R. Hamilton [13, 14] propuso utilizar el flujo de Ricci, lo que le condujo a un problema de singularidades en tiempo finito, es decir a un problema de blow-up. La aparición no es casual: estas ecuaciones se parecen algo a los casos aquí mencionados de Navier-Stokes y Reacción-Difusión en el sentido de que todas ellas son *versiones no lineales de la ecuación del calor* $u_t = \Delta u$.

Las noticias recibidas a lo largo de 2004 apuntan a que David Perelman, matemático ruso, ha resuelto la conjetura de Poincaré (o está muy cerca de hacerlo) [1]. Lo que confirma el extraordinaria importancia de los problemas de blow-up en las matemáticas actuales[10].

[10]Nota de lectura: el análisis de singularidad hecho por Perelman se basa en la autosemejanza.

11 Comentarios finales

Terminamos estas notas con algunas cuestiones que suscita este reto matemático.

- ¿Se resolverá el problema de Navier-Stokes en los próximos años?

Hay opiniones para todos los gustos. Charles Fefferman termina así su report para el Clay Institute: *"Los métodos estándar de las EDPs parecen inadecuados para resolver el problema. Probablemente necesitamos nuevas y profundas ideas"*. Pues ya lo saben, busquen nuevas ideas.

- ¿Tendrá consecuencias prácticas la solución del reto?

Las opiniones son también de lo más diverso a este respecto: muchos investigadores limitan su respuesta al campo de la matemática pura y para ellos los grandes retos matemáticos son la sal de la profesión, al acicate para elaborar nuevas y profundas teorías, como ha sucedido con el Teorema de Fermat, cuya utilidad *inmediata* no puede ser menor para la vida diaria (pero nos puede dar una sorpresa).

Este no es el caso de las ecuaciones de los fluidos, que intervienen en aspectos cruciales de la vida moderna, de los que destacamos: la meteorología, el estudio del clima, la aeronáutica, la oceanografía, la hidráulica, el estudio del flujo sanguíneo, la explotación de los recursos de gas y petróleo y el control de la contaminación. En todos estos campos el esfuerzo computacional que se está haciendo es enorme y continuo. Y es preciso señalar que no solo se trata evidentemente de los modelos incompresibles sino que incluye los modelos compresibles, los fluidos no newtonianos y los fluidos en medios porosos que estudia el autor de las presentes notas. Los protagonistas del esfuerzo práctico son conscientes de la falta de una adecuada comprensión teórica de los sutiles mecanismos que subyacen a los complicados fenómenos observados.

¿Cambiará pues nuestra comprensión teórica del SNS este panorama? Lo dejo a la consideración del amable lector.

Este texto tiene su origen en una conferencia impartida en la Academia de Ciencias de Zaragoza el 15 de enero de 2004.

Referencias

[1] M. Anderson. *Geometrization of 3 - manifolds via the Ricci Flow,* Notices Amer. Math. Soc, **51**, 2 (2004), pp. 184–193 *(Breve e intensa presentación del problema de Poincaré y la obra de D. Perelman).*

[2] V. Arnold, M. Atiyah, P. Lax, B. Mazur, "Mathematics: Frontiers and Perspectives", AMS Publications, 2000.

[3] G.K. Batchelor, "An Introduction to Fluid Dynamics", Cambridge Univ. Press, 1967.

[4] J. T. Beale, T. Kato, A. Majda, *Remarks on the breakdown of smooth solutions for the 3-D Euler equations*. Comm. Math. Phys. 94 (1984), no. 1, 61–66.

[5] J. Bebernes, D. Eberly. "Mathematical Problems from Combustion Theory", Appl. Math. Sci. **83**, Springer-Verlag, New York, 1989.

[6] L. A. Caffarelli, R. Kohn, L. Nirenberg, *Partial regularity of suitable weak solutions of the Navier-Stokes equations*, Comm. Pure Applied Maths., **35** (1982), pp. 771-831.

[7] A.J. Chorin, J.E. Marsden, "A Mathematical Introduction to Fluid Mechanics", Springer-Verlag, 1980.

[8] P. Constantin, *Some open problems and research directions in the mathematical study of fluid dynamics,* in [10], pages 353–360.

[9] P. Constantin, C. Foias, "Navier Stokes equations", Chicago Lectures in Mathematics, Univ. of Chicago press, Chicago, 1988.

[10] B. Engquist (Editor), W. Schmid (Editor), "Mathematics Unlimited - 2001 and Beyond", Springer Verlag, Berlin, 2001.

[11] C. Fefferman. Clay Mathematics Institute, Millenium Problems. Official problem description, http://www.claymath.org/millennium/Navier-Stokes_Equations/.

[12] V. A. Galaktionov, J. L. Vázquez. *The problem of blow-up in nonlinear parabolic equations*, Discrete Contin. Dynam. Systems A **8**, 2 (2002), 399–433. (A Special Issue: *Current Developments in PDE,* Guest Editors: Carlos Conca, Manuel del Pino, Patricio Felmer, and Raúl Manásevich) (Proceedings of the Summer Course in Temuco, Chile, jan. 1999).

[13] R. Hamilton. *Three manifolds of positive Ricci curvature*, J. Differential Geom. **17** (1982), 255–306.

[14] R. Hamilton. *The formation of singularities in the Ricci flow*, Surveys in Differential Geometry, vol. 2, International Press, 1955, pp. 7–136.

[15] "Mathematical Developments arising from Hilbert Problems", Proceedings of Symposia in Pure Mathematics, XXVIII, Amer. Math. Soc, Providence, 1976.

[16] A. Jackson, *Mathematical challenges of the XXI century*, Notices Amer. Math. Soc., vol. **47**, no 10 (2000), pp. 1271-1273.

[17] T. Kato, *On classical solutions of the two-dimensional non-stationary Euler equation*, Archive Rat. Mech. Anal. **25** (1967), pp. 188–200.

[18] O.A. Ladyzhenskaya, "The mathematical theory of Viscous Incompressible flow", Gordon and Breach, 1969.

[19] L.D. Landau, E.M. Lifshitz, "Mecánica de Fluidos", Reverté, Barcelona, 1991.

[20] J. Leray, (L1) *Étude de diverses équations non linéaires et de quelques problèmes que pose l'hydrodynamique*, Jour. Math. Pures Appl. **12** (1933), pp. 1–82.

(L2) *Essai sur les mouvements plans d'un liquide visqueux que limitent des parois*, Jour. Math. Pures Appl. **13** (1934), pp. 331–418.

(L3) *Essai sur le mouvement d'un liquide emplissant l'espace*, Acta Math. **63** (1934), pp. 193–248.

(L4) "Oeuvres scientifiques", Tome II, Équations aux dérivées partielles réelles et mécanique des fluides. Reedición SMF, 1998.

[21] F. H. Lin, *A new proof of the Caffarelli-Kohn-Nirenberg theorem*, Comm. Pure Appl. math. **51**, (1998), 241-257.

[22] J. L. Lions, "Quelques méthodes de résolution des problèmes aux limites non linéaires", Dunod, Paris, 1969.

[23] P.L. Lions, "Mathematical models in fluid mechanics", 2 volúmenes, Oxford Univ. Press, Oxford, 1996/1998.

[24] A. Majda, A. Bertozzi, "Vorticity and incompressible flows", Cambridge Univ. Press, 2002.

[25] C.L.M.H. Navier, *Mémoire sur les lois du mouvement des fluides,* Mém. Acad. Sci. Inst. France **6** (1822), 380–440.

[26] I. Newton, "Philosophiae Naturalis Principia Mathematica", Pepys, London, 1687. En castellano: "Principios matemáticos de la Filosofía Natural", Alianza Ed., Madrid, 1987.

[27] A. A. Samarskii; V. A. Galaktionov; S. P. Kurdyumov; A. P. Mikhailov. "Blow-up in problems for quasilinear parabolic equations", Nauka, Moscow, 1987 (in Russian). English transl.: Walter de Gruyter, Berlin, 1995.

[28] V. Scheffer, *Turbulence and Hausdorff dimension,* in "Turbulence and the Navier-Stokes Equations", Lecture Notes in Math. **565**, Springer Verlag, 1976, pp. 94–112.

[29] G.G. Stokes, *On the theories of internal friction of fluids in motion*, Trans. Cambridge Philos. Soc. **8** (1845).

[30] R. Temam, "Navier-Stokes equations", North-Holland, New York, 1979.

[31] R. Temam, "Navier-Stokes equations and Nonlinear functional analysis", SIAM, Philadelphia, 1983.

[32] J. L. Vázquez *The importance of Mathematics in the development of Science and Technology*, Boletín Soc. Esp. Mat. Aplicada, no 19, 2001, pg. 69–112. Versión española retocada: *"Matemáticas, Ciencia y Tecnología: una relación profunda y duradera*, `http://www.uam.es/personal_pdi/ciencias/jvazquez/`

[33] J. L. Vázquez, "Fundamentos matemáticos de la Mecánica de Fluidos", Notas Curso Doct. UAM, 2003. `http://www.uam.es/personal_pdi/ciencias/jvazquez/coursejlv.html`

Dirección:

Juan Luis Vázquez, Dpto. de Matemáticas,

Univ. Autónoma de Madrid, 28049 Madrid, España

Tel. 34-91-3974935, FAX 34-91-3974889

Correo electrónico: `juanluis.vazquez@uam.es`

`htpp://www.uam.es/juanluis.vazquez`

Monografías de la Real Academia de Ciencias de Zaragoza. **26:** 57–68, (2004).

P versus NP

Elvira Mayordomo

Departamento de Informática e Ingeniería de Sistemas
Universidad de Zaragoza
María de Luna 1, 50018 Zaragoza, SPAIN.
elvira@unizar.es

1 Introduction and history

In 1971 Cook first explicitly formulated the P versus NP conjecture. This is indeed the youngest of the seven Millennium problems, though I strongly believe that it won't be the first, not even the second of them to be solved. I'll go back to this daring statement, but let me start with a brief introduction to the problem and a summary of its history.

The term P, or polynomial time, refers to the class of (decisional) problems that can be efficiently solved by an algorithm. NP, for nondeterministic polynomial time, is the class of (decisional) problems for which "solutions" or certificates are efficiently verifiable. Therefore the question of whether P is equal to NP means whether solving is harder than verifying. Yet in other words, we want to know whether there is always an efficient alternative to brute force search. Formal definitions will be given in the next section.

Historically, Cook [9] and Levin [20] first defined the class NP and proved the existence of complete problems for NP, that is, problems such as SAT for which the question "Is SAT solvable efficiently?" is equivalent to "Is P equal to NP?". Karp [18] demonstrated that many familiar problems were complete for NP.

The root of this work can be traced back to the thirties and Computability or Recursion Theory originated by Turing, Church and Gödel. Computability theory is the immediate precursor of Computational Complexity, that Hartmanis, Lewis and Stearns [15, 29] and other started with their classification of languages and functions in terms of the time and space needed to compute them. Cobham [6] and Edmonds [11] in the 1960's introduced the notion of polynomial-time computation (see also previous work by Von Neumann [30]). Before Cook's and Levin's definition of the class NP, the P versus NP question appeared somewhat in the papers of Edmonds, Levin, Yablonski and in a letter from Gödel to Von Neumann [28].

In this paper I will give three different and equivalent formulations of the P vs. NP question, starting from the most natural one that connects it with verification algorithms. The second statement is in terms of the efficient solution of a particular problem and the last one uses the concept of Probabilistic Checkable Proofs, that is, probabilistic verification of a proof by checking a very small portion of it. I will also mention important connections of the P vs. NP problem with finite model theory and propositional proof systems. Finally, I will explore the ongoing research lines and the consequences of the two possible solutions of the problem. Notice that the mathematical consequence of P=NP, by using Occam razor principle, is that "we can then find proofs of theorems that have reasonably length proofs, say in under 100 pages. A person who proves P=NP would walk home from the Clay Institute not with one million-dollar check but with seven" (from Lance Fortnow in [13], see also [7]).

This historical introduction has been possible through detailed information provided in [7, 28, 27, 5].

Disclaimer. This chapter has been deliberately written for a nonexpert audience. For the purpose of clarity, formal and exact definitions have been often sacrificed. The references given intend to provide a minimal number of pointers for an interesting reader to start a more detailed study, and finally many important topics have been left out for space reasons but can be found in [3, 26].

2 Initial formulation of the problem

We start with the definition of the class P. A *decisional problem* is a problem in which each instance can have one of two possible solutions, that is, a Yes/No answer question such as "Given a natural number n, is n prime?" Formally a decisional problem $\Lambda = (I, R)$ is a set of data I codified over a finite alphabet, and a property R; the problem is stated as "Given x in I, does $R(x)$ hold?". In general problems, the solution to each input can take values in a larger set, for instance for the problem of computing the square root of a given number. Here we will only work with decisional problems, and will often drop the term "decisional".

We divide the set of all decisional problems in *complexity classes*, according to the resources used by an algorithm solving each problem. By the term *algorithm* we mean a finite set of instructions on any of a number of related models of computation, e.g., the Turing Machine or the Random Access Machine (this last one is an idealization of our everyday computers).

P is the class of problems that can be solved by an algorithm in time bounded above by a polynomial on the size of the input, that is, if a problem Λ is in P this means that

there are constants c, k such that the time needed for input x is at most $c|x|^k$, where $|x|$ is the size of the input, that is, the length of its codification. When we say time we mean the total number of steps taken by the algorithm. This definition of P is robust over the choice of a reasonable computation model.

An example of a problem in P is the boolean formula evaluation problem; given a well written boolean formula F with variables and boolean operators AND, OR, and NOT (such as the formula $(x \vee \neg y) \wedge z$), and given an assignment α that sets each of the variables as TRUE or FALSE (such as $x = y =$TRUE, $z =$FALSE for the above formula), does the formula F with assignment α evaluate to TRUE? (Usually when a formula F with an assignment α evaluates to TRUE we say that α satisfies the formula F). An algorithm for this problem working in polynomial time would be to first substitute each variable by the corresponding value and then simplify.

P is usually identified with the class of feasible problems, or problems that can be solved in practice. Evidently a polynomial-time bound can be huge both because of the multiplicative constant and the degree, but there are two practical reasons why we think of P as the efficiently solvable problems. Natural problems that are known to be in P have a very reasonable polynomial time bound, with degree at most 4, currently the extreme case is the problem of primality testing, very recently known to be solvable in polynomial time [1] for which the best known polynomial bounds have degree 6, but in this case I am convinced it is just a matter of time that this algorithm is improved. The second reason is that for natural problems not known to be in P, the best known algorithms take exponential time for some inputs, therefore the intermediate time bounds between polynomial and exponential don't seem to happen in practice.

The class NP can be defined in terms of verification algorithms or verifiers. A *verifier* V for a problem $\Lambda = (I, R)$ is an algorithm that takes two inputs, an input x in I and a *certificate* π. A certificate π is a candidate for a proof that $R(x)$ holds, that is, a witness of a Yes answer. The verifier V on input x, π outputs Yes if it can show that $R(x)$ holds using π, and No otherwise. The goal is that any Yes input has at least one certificate that convinces the verifier, whereas no certificate is valid for a No input. Formally, V verifies the problem Λ if for each x such that $R(x)$ holds there is a certificate π such that V with input x, π outputs Yes, and for every x such that $R(x)$ does not hold V outputs No on input x, π, for every π.

NP is the class of problems that can be verified by an algorithm running in time bounded by a polynomial on the size of its first input, that is, the input of the problem itself.

An example of a problem in NP is SAT or satisfiability of boolean formulae, that is, given a well written boolean formula F with operators AND, OR, and NOT, is there an

assignment α that satisfies F (that is, an assignment that makes the formula evaluate to TRUE)? A certificate for the verifier would be an assignment, and the algorithm only needs to check that the certificate satisfies the formula, which can be done in polynomial time as we have seen before.

It is easy to see that every problem in P is also in NP, the corresponding verification algorithm just ignores the certificate and then uses the polynomial time algorithm that solves the problem. So P$\subseteq$ NP, but does the other containment hold? Is P=NP? The question is thus a matter of solving vs. verifying a possible solution, we want to know if solving is really harder or less efficient than just verifying. We can also state our main question in terms of exhaustive search. It is clear that if a problem is in NP, then it can be solved by a brute force search of a certificate that satisfies the verifier, but such a procedure would take exponential time in many cases. Can we do better?, that is, can brute force search be replaced by an efficient (polynomial time) algorithm in all cases?

3 NP-completeness: a second formulation

In the previous section we gave a statement in terms of abstract problems, we want to know if each polynomial-time verifiable problem is also solvable in polynomial time. Here we have a formulation in terms of a particular problem, we see that P vs. NP is equivalent to the question of whether SAT, the Satisfiability problem, is in P.

In this approach we need the concept of a *reduction* between two problems. A problem A reduces to a problem B (denoted $A \leq_{\mathrm{m}}^{\mathrm{p}} B$) if there is a polynomial time algorithm that transforms each x, input of A, into an input of B, $f(x)$, that has exactly the same solution for B as x had for problem A. This means that if we have an efficient algorithm solving B we can combine it with a reduction from A to B and in this way we solve A efficiently. Notice that if B is in P and A reduces to B then A is also in P by the last argument. Equivalently, if A is not in P and A reduces to B then B is not in P.

We can therefore establish a partial ordering among the problems in NP by using the reductions $\leq_{\mathrm{m}}^{\mathrm{p}}$. Cook [9] proved in 1971 that the problem SAT is a maximum for this order, that is, every problem in NP reduces to SAT. Therefore it is enough to know whether SAT is in P. If SAT is in P then every problem that reduces to SAT is in P, therefore P=NP. If SAT is not in P then P$\neq$ NP, because SAT is in NP.

This property of SAT representing in a very strong sense the behaviour of the whole class NP holds for many other interesting problems. We say that a problem C is *complete* for NP, or NP-complete, if every problem in NP reduces to C. Therefore answering the question of whether C is in P, for a particular NP-complete problem C, would settle P vs. NP.

The list of known NP-complete problems includes important ones from virtually every area in science and engineering (see [14]) so in this sense the P vs. NP open question could be settled by a researcher from virtually any topic. After thirty years of unfruitful hard work in this direction, I don't think the analysis of the complexity of a particular problem will solve it. I explore in the next two sections alternative formulations that give a flavour of the robustness and strength of the problem statement and therefore of the foreseen difficulty of a solution.

4 Connections to logic: finite model theory and propositional proof systems

coNP is the class of complements of problems in NP, that is, A is in coNP if there is a problem B in NP such that x has solution Yes for A exactly when x has solution No for B.

If P=NP then clearly NP=coNP because the class P is closed under complementation of its problems. But the hypothesis NP $\neq$ coNP is stronger than P $\neq$ NP. It is plausible (although widely conjectured to be false) that P $\neq$ NP but NP=coNP.

In this section I will very briefly explore connections of the NP vs. coNP question and logic. The interested reader can find more details in [19] and [10, 21].

A *tautology* is a boolean formula that is true for any assignment of the variables. We denote as TAUT the set of all tautologies. A *propositional proof system* is a function f from proofs to tautologies that can be computed in polynomial time, where a *proof* is just a finite string of symbols. π is a *proof of* Ψ if $f(\pi) = \Psi$. An important question in propositional logic is whether there exists a propositional proof system such that every tautology has a polynomial size proof, that is, whether every true statement has a short proof that can be efficiently checked.

In fact it is known that it is enough to consider propositional proof systems that are based on modus ponens (formally, they are extensions of a SF, Frege system with substitutions). See [19] for all definitions.

Cook and Reckhow proved in [8] that NP=coNP is equivalent to the existence of a propositional proof system such that every tautology has a polynomial size proof.

In the rest of this section I will briefly sketch a different connection of logic and complexity theory. Fagin [12] proved that a problem $\Lambda = (I, R)$ is in NP if R can be written as an existential second-order formula with existential second order quantifiers.

I will not define first order or second order here (see for instance [10]) but I will give a few simple examples of their corresponding expressive power. Assume for a minute that our inputs are graphs (a graph is a set of vertices V and a binary relation E on V, the

edges). We can express with first order that a vertex u_0 is isolated:

$$\forall x \neg E(u_0, x) \wedge \neg E(x, u_0)$$

that is, the connectives $\wedge, \vee, \neg$ are allowed, as well as universal and existential quantifiers on vertices. I can express 2-colorability, that is, there is a way to colour all vertices with two colours such that adjacent vertices don't have the same colour as follows:

$$(\exists P)(\forall x)(\forall y)(E(x, y) \rightarrow (P(x) \leftrightarrow \neg P(y)))$$

Notice that I have quantified over relations here, that is, I am using second order logic. The reader can try to express 3-colorability with second-order.

More recently, P has also been characterized using the expressive power of a more sophisticated logic, first order with fixed point operators. The fixed point operator allows us to iterate the relation "there is an edge from vertex u to vertex v" to "vertices u and v are connected (by a path of any length)"

$$\begin{aligned} \phi^1(x, y) &\equiv E(x, y) \\ \phi^{m+1}(x, y) &\equiv \phi^m(x, y) \vee \exists z(E(x, z) \wedge \phi^m(z, y)) \end{aligned}$$

The iteration of this process, ϕ^∞, is the fixed point of E.

The question of whether P=NP is thus equivalent to a question on logic expressibility, namely whether existential second order and first order with fixed point express the same properties.

5 Probabilistically Checkable Proofs

In this section we characterize the class NP with a generalization of verification algorithms called Probabilistically Checkable Proofs. This result was proven in 1992 [2] and received a wide attention from the Computational Complexity community for two reasons. On the one hand this definition of NP had dramatically different properties from the previously known characterizations, in the sense that this one does not relativize as I will explain below. On the other hand many negative approximation results were derived from it, meaning that it was proven that no approximated solution to many optimization problems exists unless P=NP.

We consider probabilistic algorithms, that is, algorithms that have access to random bits, which means that at any point of its execution the algorithm can request a random bit and receive it in unit time, the bit being 1 with probability 1/2.

We define probabilistic verifiers which are verifiers such as those defined in section 2 (that is, the input of the verifier is the original input x and a certificate π) but with the additional power of probabilistic algorithms.

So now the verification algorithm on a fixed input x and certificate π can give different outputs because they depend on the random bits received, so we need to relax the notion of a verifier being correct for a problem.

A probabilistic verifier V is valid for a problem $\Lambda = (I, R)$ if for each input $x \in I$, if $R(x)$ holds then there is a certificate π such that V outputs Yes with probability 1 (the probability is taken over the random bits produced); if $R(x)$ does not hold then for any certificate π, the probability of V giving output Yes is smaller than 0.1

$$\begin{aligned} R(x) &\Rightarrow \exists \pi \Pr(V(x, \pi)) = 1 \\ \text{NOT } R(x) &\Rightarrow \forall \pi \Pr(V(x, \pi)) < 0.1 \end{aligned}$$

Notice that for Yes inputs there is a certificate for which the verifier always gives a correct answer, whereas for No inputs and any certificate the output can be wrong with a small probability.

Polynomial-time probabilistic verifiers are very powerful if we allow them to use polynomially random bits and to have full access to the certificate, in fact they correspond in this case to exponential time verifiers and the complexity class NEXP, that is, the exponential time analogous to nondeterministic polynomial time NP.

This is the reason why we introduce two parameters restricting the amount of randomness and the access to the certificate. The certificate π can be read one symbol at a time by requesting the symbol in a particular position of π and getting it in unit time, that is, a direct access mechanism. We can now restrict the number of symbols in the certificate that are actually read.

It is clear that if we only restrict the number of random bits use to 0 we get exactly the class NP, because with that restriction only we get our original polynomial time verifiers. But what happens if we use randomness? Can we probabilistically verify without having to read the whole certificate? This would correspond to the idea of quickly checking (very long) candidate proofs of a theorem; correct proofs should be accepted but there is a small probability of accepting a false proof.

From 1990 to 1992 several results appeared in this line, first showing that a polylogarithmic number of both random bits and certificate access were sufficient to capture NP, and then getting conditions that were both sufficient and necessary. See for instance [23] for the whole story. The best known result [2] is that NP is exactly the class of problems that can be solved by probabilistic verifiers using a constant number of certificate bits (that is, the same number of bits for all inputs and certificates) and a logarithmic number of random bits, this is the complexity class $\mathrm{PCP}(\log n, 1)$.

These probabilistic verifiers are thus very restricted, they check a very small amount of the certificate and get a correct answer with high probability. The result was proven for the NP-complete SAT, by a beautiful arithmetization technique that transforms each boolean assignment into a linear function. Our P versus NP problem is now transformed into the question of whether logarithmically many random bits and a constant number of times of certificate access can do more than a regular polynomial-time algorithm.

This is the first known form of the problem that does not relativize, which was very celebrated by the researchers in Computational Complexity. Why were they so happy? Assume that we live in a world were the solution of a particular problem A is given for free. This means that at any point an algorithm can ask for the solution of A on a particular input y an get an exact solution in unit time. This is called having A as an oracle. We can now define the class of polynomial-time solvable problems in this world A, denoted as P^A, and similarly for verifiers. The standard separation techniques that were used in attacking the P versus NP question were all known to relativize, meaning that they only give results that are independent of the oracle, that is, results that hold in any possible world A. But this is useless because it is known that there are oracles for which $\mathrm{P}^A = \mathrm{NP}^A$ and other for which $\mathrm{P}^A \neq \mathrm{NP}^A$, so techniques that relativize will never solve the question. The equality NP=PCP$(\log n, 1)$ is known not to hold for some oracles and therefore proofs that rely on this characterization of NP do not relativize, so there is some hope that they can obtain stronger results than those of known relativizable techniques.

6 Research directions

In this section we list research areas created and/or highly motivated by the P vs. NP question.

Computational Complexity [3] defines different complexity classes such as P and NP in terms of different computing resources. The main open question in this context is the separation of different complexity classes, that is, whether they have the same problems. Typical tools are reductions, such as $\leq_{\mathrm{m}}^{\mathrm{p}}$, that can vary according to the resources used in the computation of the reduction itself and the access the reduction gives to the problem it reduces to, for instance in the case of $A \leq_{\mathrm{m}}^{\mathrm{p}} B$ the reduction accesses a single input of the problem B and the solution to this input is exactly the solution for the original input of A.

Besides the more specific directions we mention below, there is a rich variation of techniques in Computational Complexity, starting with the classical diagonalization and counting techniques and including quantitative approaches such as resource-bounded measure [22] and highly nonclassical ones such as quantum computing complexity classes [24].

An important question in complexity is whether having access to a source of random bits can make computation substantially quicker. There are several **probabilistic complexity classes** corresponding to different allowed computation errors in this context, for instance one-sided or two-sided errors for the case of decisional problems. BPP is the class of problems that can be solved in polynomial time with a source of random bits and allowing an exponentially small on both possible cases, Yes and No instances. It is open whether BPP=P and in this case a positive answer is not ruled out by many complexity theorists [25].

Average case complexity considers a probability distribution on the inputs of a problem and measures the time needed to solve it as an average value, as opposed to the worse case complexity we use in the definition of P, for instance, where we consider the time needed for the slowest input of each length. We are not looking for algorithms that solve quickly all instances of a problem but for those that work well on average [31].

The size of **Boolean circuits** [32, 3] that solve a problem is another complexity measure. This is a nonuniform computational model, since a different circuit is needed for each input size. The advantage of these simple models is that lower bounds have been obtained, at least for small bounds, and it seems plausible that this work direction can give more powerful results (not with the known techniques though). It is known that each problem in P has polynomial-size circuits so showing that a single problem in NP does not have polynomial size circuits would separate these two classes.

Approximation algorithms is a very active area of research that has used the PCP characterization of NP (see section 5) for negative results. They consider the optimization problems corresponding naturally to many NP-complete problems, for example MAXCLIQUE is the problem of computing the size of the maximum complete subgraph, which is the optimization problem corresponding to NP-complete CLIQUE (CLIQUE is the problem of deciding whether a graph has a complete subgraph of a given size).

The behaviour of these NP-complete based optimization problems is very varied in terms of how well they can be approximately solved. Some of them can be solved in polynomial time with an exponential error whereas for other just solving the problem with a constant error would imply P=NP.

Descriptive complexity [17] explores the characterization of complexity classes in terms of logic expressibility, in the line of Fagin's characterization of NP ([12], section 4). Immerman [16] characterized P, NL (nondeterministic logarithmic space) and other complexity classes and proved that NL=coNL using logic techniques. There is a parallel line of research dealing with algebraic characterizations of complexity classes [4].

Proof complexity [19] explores the connection we introduced in section 4 between

NP=coNP and the existence of short proofs for tautologies. The idea is to prove superpolynomial lower bounds for the length of proofs in propositional proof systems of increasing complexity, in order to end up obtaining the result for every propositional proof system.

7 Consequences

Most complexity researchers believe that P is different from NP. In fact much more is expected to hold, the assumptions used in cryptography include that integer factoring cannot be done in polynomial time, which implies P≠NP, and in fact it is even assumed (for instance in DES) that factoring cannot be done in polynomial time for "many" integers. Since multiplication is feasibly computable, factoring can be formulated as inverting a polynomial time computable function. The existence of a polynomial time computable function that cannot be inverted in polynomial time is crucial in modern cryptography and is conjectured to be a much stronger hypothesis than P≠NP.

As explained above, a feasible algorithm for an NP-complete problem (therefore showing P=NP) would mean the end of DES and most currently used cryptographic protocols, with devastating financial and military consequences. But not all consequences would be negative. Consider the problem X of deciding whether an input T, p corresponds to a valid theorem T and a prefix of a formal proof of T, p, where a formal proof is detailed enough to be checked by a computer. This problem X is in NP, so if P=NP it can be solved in polynomial time, and this would mean a breakthrough in mathematics. For any theorem which has a proof of reasonable length we can efficiently find such a proof!

Although I strongly believe this is not the case (I am sure that P is different from NP), the possibility of walking home with all seven Clay Institute prize checks, as Lance Fortnow says in [13], is definitely an incentive for the other direction.

References

[1] M. Agrawal, N. Kayal, and N. Saxena. Primes is in P, 2002.

[2] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998.

[3] J. L. Balcázar, J. Díaz, and J. Gabarró. *Structural Complexity I (second edition)*. Springer-Verlag, Berlin, 1995.

[4] D. A. Mix Barrington and D. Thérien. Finite monoids and the fine structure of NC1. *Journal of the ACM*, 35:941–952, 1988.

[5] Daniel Pierre Bovet and Pierluigi Crescenzi. *Introduction to the Theory of Complexity.* Prentice Hall, 1994.

[6] A. Cobham. The intrinsic computational difficulty of functions. In *Proceedings of the 1964 International Congress for Logic, Methodology, and Philosophy of Science*, pages 24–30, 1964.

[7] S. Cook. The P versus NP problem. Manuscript.

[8] S. Cook and R. Reckhow. The relative efficiency of propositional proof systems. *Journal of Symbolic Logic*, 44:36–50, 1979.

[9] S. A. Cook. The complexity of theorem proving procedures. In *Proceedings of the Third ACM Symposium on the Theory of Computing*, pages 151–158, 1971.

[10] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory.* Springer-Verlag, 1999.

[11] J. Edmonds. Minimum partition of a matroid into independent subsets. *J. Res. Nat. Bur. Standards Sect. B*, 69:67–72, 1965.

[12] R. Fagin. Generalized first-order spectra and polynomial-time recognizable sets. *Complexity of Computation, SIAM-AMS Proceedings*, 4:43–73, 1974.

[13] L. Fortnow. My computational complexity web log, may 25, 2004.

[14] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-completeness.* W.H. Freeman and Company, San Francisco, 1979.

[15] J. Hartmanis and R.E. Stearns. On the computational complexity of algorithms. *Transactions of the American Mathematical Society*, 117:285–306, 1965.

[16] N. Immerman. Languages that capture complexity classes. *SIAM Journal on Computing*, 16:760–778, 1987.

[17] N. Immerman. *Descriptive Complexity.* Springer-Verlag, 1999.

[18] R. M. Karp. Reducibility among combinatorial problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–104. Plenum Press, 1972.

[19] J. Krajicek. *Bounded Arithmetic, Propositional Logic and Complexity Theory.* Cambridge University Press, 1996.

[20] L. A. Levin. Universal sequential search problems. *Problems of Information Transmission*, 9:265–266, 1973.

[21] L. Libkin. *Elements of Finite Model Theory.* Springer-Verlag, 2004.

[22] J. H. Lutz. The quantitative structure of exponential time. In L. A. Hemaspaandra and A. L. Selman, editors, *Complexity Theory Retrospective II*, pages 225–254. Springer-Verlag, 1997.

[23] E. W. Mayr, H. J. Prmel, and A. Steger. *Lectures on Proof Verification and Approximation Algorithms*, volume 1367 of *Lecture Notes in Computer Science*. Springer-Verlag, 1998.

[24] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[25] N. Nisan and A. Wigderson. Hardness vs randomness. *Journal of Computer and System Sciences*, 49:149–167, 1994.

[26] Christos H. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1994.

[27] J.E. Savage. *Models of Computation: Exploring the Power of Computing*. Addison-Wesley, 1998.

[28] M. Sipser. The history and status of the P versus NP question. In *Proceedings of the 24th Annual ACM Symposium on the theory of Computing*, pages 603–618, 1992.

[29] R. E. Stearns, J. Hartmanis, and P.M. Lewis. Hierarchies of memory limited computations. In *Proc. 6th Annual Symp. on Switching Circuit Theory and Logical Design*, pages 179–190, 1965.

[30] J. von Neumann. A certain zero-sum two-person game equivalent to the optimal assignment problem. In H.W. Kahn and A.W. Tucker, editors, *Contributions to the Theory of Games II*. Princeton University Press, 1953.

[31] J. Wang. Average-case computational complexity theory. In L. Hemaspaandra and A. Selman, editors, *Complexity Theory Retrospective II*, pages 295–328. Springer-Verlag, 1997.

[32] I. Wegener. *The Complexity of Boolean Functions*. John Wiley & Sons, 1987.

Monografías de la Real Academia de Ciencias de Zaragoza. **26:** 69–88, (2004).

Teoría Cuántica de Yang-Mills. la Generación de la Masa

Manuel Asorey

Dpto. de Física Teórica, Universidad de Zaragoza, Zaragoza (Spain)

asorey@saturno.unizar.es

Resumen

A pesar de constituir la pieza central del paradigma de la física de interacciones fundamentales basado en teorías relativistas cuánticas de campos, las teorías gauge no abelianas presentan a los 50 años de su descubrimiento por Yang y Mills numerosos interrogantes que afectan incluso a su propia consistencia. La importancia de resolver alguno de estos problemas ha impulsado al Instituto Clay a considerarlos como unos de los retos matemáticos del nuevo milenio. En esta nota repasamos diversos aspectos de las teorías de Yang-Mills y la formulación precisa del problema seleccionado por el Instituto Clay como merecedor de su más distinguido galardón.

1 Introducción

Matemáticos de todas las generaciones han enfatizado la importancia de ciertos problemas cuya resolución les es esquiva. El reto por la resolución de problemas simultáneamente reconocidos como importantes y difíciles por matemáticos de prestigio, ha contribuido a impulsar la investigación en áreas matemáticas que se tornan prioritarias por su relevancia para resolver los desafíos planteados. Por su importancia histórica, destaca la selección hecha por Hilbert al comienzo del siglo XX. Entre los 23 problemas elegidos por Hilbert para el Congreso Internacional de Matemáticas de París de 1900 se encuentra en el número seis uno sorprendente: Tratamiento Matemático de los Axiomas de la Física (*Mathematische Behandlung der Axiome der Physik*)[1]. El sexto problema de Hilbert es demasiado vago en su formulación, aunque en su propia descripción Hilbert parece restringirlo a los ámbitos de mecánica y la teoría de la probabilidad, como para pretender encontrar una solución precisa. Un análisis del desarrollo posterior del estudio de este problema puede encontrase en el libro de Corry [2] y en el artículo reciente de Rañada [3]. El enfoque axiomático las Matemáticas propuesto por Hilbert sufrió un duro embate con

los resultados de Gödel [4]. Sin embargo sus ideas influyeron en la Física de una manera asombrosa gracias al gran influjo directo e indirecto de Hilbert en el mundo académico alemán. Los fundamentos de la mecánica cuántica fueron establecidos en los años treinta a partir de unos postulados o axiomas que todavía perduran en los manuales europeos [5, 6, 7, 8]. Más tarde, en los años sesenta la incipiente teoría relativista de los campos cuánticos que constituye la síntesis de la relatividad especial de Einstein con la Mecánica Cuántica, comenzó a formularse de forma axiomática y llegó acuñarse el término: teoría axiomática de campos, para describir un campo de investigación que involucró a muy destacados físicos teóricos de todo el mundo [9, 10, 11].

Al hilo de las celebraciones del inicio del tercer milenio, el Instituto Matemático Clay de Cambridge (USA) instituyó un galardón para premiar a los matemáticos que resolviesen los siete problemas más destacados pendientes de solución [12]. Entre los siete problemas vuelven a aparecer algunos directamente vinculados a la Física. Entre ellos destaca el conocido como el problema de la masa en las teoría de Yang-Mills. De los siete problemas del Milenio seleccionados por el Instituto Clay es el más directamente vinculado a la Física contemporánea y el más desconocido para la comunidad matemática. Es sin duda el problema más difícil de formular de los seleccionados porque involucra conceptos de frontera de la física y matemáticas cuya simple formulación requiere varios manuales. En efecto el planteamiento del problema requiere elementos de las teorías físicas de la relatividad especial y la mecánica cuántica al mismo tiempo que campos de la matemática como la teoría de probabilidades, geometría diferencial y análisis funcional. A lo largo de esta reseña pretendemos dar una posible formulación lo más simplificada posible de este problema.

Sin entrar en los detalles técnicos del problema, que veremos más adelante hay una manera sencilla e intuitiva de comprender el problema en términos puramente físicos. Desde finales de los años sesenta existe una teoría fundamental que explica a la perfección la teoría de la interacciones fuertes responsables de la estabilidad del núcleo atómico. Esta teoría recibe la denominación de Cromodinámica Cuántica y su elemento esencial consiste en la descripción de la propagación relativista de dicha interacción fuerte a través de una partícula transmisora virtual conocida como gluón. El gluón juega en el mundo nuclear un papel análogo al del fotón en el mundo de las interacciones electromagnéticas. Ambas se propagan a la velocidad de la luz, sin embargo existen dos diferencias esenciales entre las mismas. El fotón posee una realidad experimental que nuestro ojos detectan en cada instante, sin embargo del gluón sólo observamos sus efectos secundarios. La otra gran diferencia estriba en que el fotón es una partícula sin masa lo que permite que se propague más lejos lo que da un alcance infinito a la interacción electromagnética y un gran tamaño, en términos de distancias fundamentales, al átomo y las moléculas. La

interacción fuerte generada por los gluones sin embargo es de corto alcance y no va más allá del núcleo atómico. Esto sugiere que el gluón o sus partículas derivadas responsables de la interacción fuerte poseen en realidad una masa no nula. El explicar este fenómeno en términos de la cromodinámca cuántica, es decir a partir de primeros principios es el objeto del problema Clay. Desde el punto de vista puramente físico esto explicaría porqué los protones y neutrones del núcleo atómico que son tan pesados (m(protón)= 938 MeV, m(neutrón)=940 MeV) [1] mientras que sus constituyentes materiales más fundamentales, tres quarks, son muy ligeros (menos de 20 Mev en total). El resto de la masa debe provenir de la energía de interacción generada por los gluones que pasa de esta forma a constituir el elemento fundamental de las partículas nucleares. La explicación del fenómeno aunque no incumbe al Instituto Clay es de gran interés en la física fundamental de altas energías.

2 Teoría de Yang-Mills

El nacimiento de las teorías de Yang-Mills, una de las grandes invenciones teóricas de la ciencia contemporánea, surge como fruto de una idea abstracta teórica generada a lo largo de medio siglo de estudios sobre la estructura profunda del electromagnetismo y la gravitación.

Inmediatamente después de que Einstein formulase la teoría relativista de la gravitación en la que la interacción gravitatoria pasa de ser una mera acción a distancia en el universo Newtoniano a ser una interacción transmitida por ondas, similares a las electromagnéticas, que viajan también a la velocidad de la luz, comenzaron a vislumbrarse más características comunes entre ambas interacciones. La más destacada es que ambas son de largo alcance. La primera extiende sus dominios hasta los confines del átomo para las partículas elementales y la segunda hasta los confines del Universo. Esta naturaleza de ambas interacciones radica en que las dos partículas responsables del transporte de la interacción: el fotón y el gravitón no poseen masa. Desde un punto de vista aparentemente más formal ambas comparten una nuevo tipo simetría: la invariancia gauge. La primera observación de esta fenómeno parte de Weyl que en su intento de unificar ambas interacciones en una sola, utiliza como elemento guía la existencia de esta simetría gauge. ¿En qué consiste este nuevo principio?.

En el electromagnetismo la simetría gauge tiene como consecuencia física la conservación de la carga eléctrica. En el caso gravitatorio el resultado análogo implica la conservación del momento y la energía. En el formalismo covariante relativista (c=1) el campo electromagnético es descrito por un campo vectorial tetra-dimensional A_μ cuya primera componente $A_0 = -\phi$ corresponde al potencial escalar del campo eléctrico $E = -\vec{\nabla}\phi - \partial_t \vec{A}$

[1] MeV=Mega electrón voltio=1,8 . 10^{-27} gramos

y cuyas tres últimas al potencial vector $\vec{A}$ que genera el campo magnético $\vec{B} = \vec{\nabla} \times \vec{A}$. Si la fuente del campo electromagnético A_μ es un campo complejo Ψ la teoría posee una invariancia bajo la siguiente transformación conjunta

$$\Psi \to \xi\Psi; \quad A_\mu \to A_\mu - i\xi^* \partial_\mu \xi \tag{1}$$

donde ξ es una función compleja unimodular

$$\xi(x) \in U(1), \quad |\xi(x)| = 1$$

si la carga eléctrica se conserva y viceversa.

En ausencia de materia dicha simetría es evidente. En efecto, la dinámica de las ondas electromagnéticas viene gobernada por la acción

$$S = \frac{1}{4e^2} \int F_{\mu\nu} F^{\mu\nu} \tag{2}$$

donde

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \qquad \mu, \nu = 0, 1, 2, 3 \tag{3}$$

es el tensor electromagnético formado por los campos eléctrico $E_i = F_{0i}$ y magnético $B_i = F_{jk}$ (con i, j, k diferentes y ordenados de forma acorde con las permutaciones cíclicas de la terna $1, 2, 3$) que quedan invariantes bajo la transformación (1). Aunque puede parecer una simetría ficticia debido al empeño en expresar la dinámica en función del potencial electromagnético, esto no es así. Los potenciales son necesarios para la cuantización de las partículas materiales y la propia cuantización de la interacción electromagnética.

En el análisis desarrollado por Weyl [13] la simetría gauge proviene del hecho de que la carga eléctrica es una noción local, definida en cada punto del espacio-tiempo y que su definición en términos de los campos fundamentales de la materia debe permanecer invariante bajo el cambio de sistema de referencia que se adopte en cada punto del espacio-tiempo para medir estos campos. En este sentido multiplicar por la fase [2] $\xi(x)$ los campos materiales puede considerarse como realizar un giro bidimensional asociado en cada punto del espacio-tiempo x a un cambio de sistema de referencia realizado en ese mismo punto de las coordenadas eléctricas internas (complejas) de la materia descrita por la función de estado Ψ. Este nuevo tipo de simetría se conoce con nombre de simetría gauge (anglicismo[3] derivado del original alemán eich, jauge en francés). El campo electromagnético representa un elemento necesario para comparar esos sistemas de referencia en dos puntos alejados.

[2] En un principio Weyl consideró esta transformación gauge como una dilatación del campo, pero enseguida resultó evidente que dicha interpretación no era correcta

[3] Algunos autores, fundamentalmente americanos, utilizan las palabras castellanas calibre o aforo para referirse a esta nueva simetría

El campo electromagnético proporciona el elemento de orientación base para determinar como el sistema de referencia elegido en uno de los puntos x se traslada al punto x' cuando se sigue un camino determinado para viajar de x a x'. La noción de transporte paralelo es la idea necesaria compatible con los principios de la relatividad y el positivismo implícito que subyace en su formulación para describir cualquier tipo de interacción.

La manera en la que realiza esa comparación es fijando cual es el transporte paralelo del valor campo Ψ en el punto x al punto x'. Esto queda determinado por la solución de la ecuación diferencial

$$\dot{\gamma^\mu}\partial_\mu\Psi = i\dot{\gamma^\mu}A_\mu\Psi$$

a lo largo de la curva γ que une x a x'.

De forma análoga el campo gravitatorio proporciona el elemento necesario para para comparar sistema de referencia espacio-temporales de un punto a otro. ¿Cómo pueden comparar sus resultados dos observadores que estén en dos puntos diferentes del espacio-tiempo?. En ausencia de gravitación la relatividad especial nos dice que mediante una transformación de Poincaré, sin embargo en presencia de gravitación la comparación debe realizarse de acuerdo con un camino γ elegido para viajar de x a x'. La teoría del transporte paralelo de los sistemas de referencia espacio-temporales fue desarrollada por el matemático Levi-Civita [14] e intensamente utilizada por Weyl, Einstein, Cartan y otros en la búsqueda de una teoría relativista unificada del electromagnetismo y la gravitación.

En el año 1954 Yang y Mills publicaron un trabajo [15] en el que introducían una nueva teoría como propuesta para el fundamento de la teoría de las interacciones fuertes del núcleo atómico. Es bien conocido que los elementos básicos del núcleo lo constituyen protones y neutrones. También se conocía que ambas partículas se comportaban de forma similar bajo las interacciones fuertes nucleares. La simple idea de Yang y Mills fue postular que puesto que estas interacciones debían respectar la simetría de intercambiar un protón por un neutrón en realidad deberían ser invariantes por cualquier rotación intermedia en el plano formado por los campo cuánticos asociados al protón y al neutrón (simetría de isospín). Como estos campos son complejos (esencia básica de la física cuántica) dicha rotación debe ser compleja y el grupo de estas rotaciones es el de matrices unitarias unimodulares $SU(2)$

$$\begin{gathered}\Psi = \begin{pmatrix} p \\ n \end{pmatrix} \rightarrow \xi \begin{pmatrix} p \\ n \end{pmatrix} \\ \xi = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SU(2), \quad \det\xi = ad - cb = 1\end{gathered} \tag{4}$$

La parte genial de su razonamiento fue hacer que esta simetría fuese no sólo global sino también local (simetría local de isospín),es decir, demandar que la teoría fuese no sólo invariante bajo un cambio global rígido del concepto de protón o neutrón sino incluso bajo

un cambio local del mismo en el que la rotación varía de punto a punto. Como hemos visto en el caso electromagnético esta nueva simetría gauge requiere la introducción de una interacción por un campo gauge que mantenga esta simetría y sirva como referencia para comparar los conceptos de protón y neutrón en puntos puntos separados. Es decir, postularon que la nueva teoría debería ser invariante bajo la transformación conjunta

$$\begin{aligned} \Psi = \begin{pmatrix} p \\ n \end{pmatrix} \rightarrow \xi(x) \begin{pmatrix} p \\ n \end{pmatrix} \\ A_\mu \rightarrow A_\mu - i\xi^\dagger(x)\partial_\mu \xi(x) \end{aligned} \tag{5}$$

donde ahora la transformación ξ puede variar de punto a punto.

La dinámica de la interacción puede obtenerse de forma análoga al caso electromagnético a partir de una acción similar

$$\begin{aligned} S = \frac{1}{2g^2} \int \mathrm{tr}\, F_{\mu\nu}\, F^{\mu\nu} \quad A_\mu \in \mathrm{Lie}\, SU(N) \\ F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu] \end{aligned} \tag{6}$$

La misma idea habia sido explorada previamente por Klein y Pauli quienes habían descartado la idea puesto que requería que la interacción fuerte fuese de largo alcance al estar basada en la transmisión por una partícula similar al fotón y descrita por el campo gauge A_μ y los hechos experimentales muestran lo contrario: dicha interacción no sale del núcleo atómico. Yang y Mills conscientes de este problema lo mencionan al final de su artículo pero lanzan su teoría pensando que alguna solución implícita contendría del mismo.

Aunque la teoría de Yang y Mills fracasó en su intento de describir la interacciones fuertes en su formulación original, la idea germinó una década más tarde en la teoría de las interacciones nucleares débiles formulada por Glashow, Salam y Weinberg. En ellas el carácter de corto alcance se logra por un mecanismo basado en un campo auxiliar que genera una masa para el campo gauge A_μ. A finales de los sesenta volvió a retomarse la teoría original de Yang-Mills para describir la interacción fuerte, pero ahora en vez de la simetría de sabor del protón y neutrón se consideró una simetría nueva de color descubierta por Gell-Mann que se basa en rotaciones del espacio de los quarks (constituyentes elementales de los nucleones) y que como son tres pasa a ser de tipo $SU(3)$ en vez de $SU(2)$ original de Yang-Mills. Dicha teoría tuvo un éxito inmediato a partir de nuevos datos experimentales en la física de interacciones fuertes a muy altas energías que indican que los quarks en el interior de los nucleones se mueven casi libremente. La teoría constituye lo que se conoce con el nombre de Cromodinámica y se considera la teoría básica de la interacción fuerte. Sin embargo ésta presenta dos problemas inexplicados: ¿cómo se genera la masa de la partícula gauge puesto que la interacción es de corto alcance? y ¿por qué los quarks no pueden observarse libremente fuera de los nucleones que componen los núcleos atómicos?. Los dos problemas están íntimamente relacionados y podría decirse

que son dos caras de la misma moneda: un fenómeno escondido en la dinámica cuántica no lineal de la ingenua y simple teoría formulada en 1954 por Yang y Mills. La solución del primero de esos problemas será galardonado por el Instituto Clay, el segundo es de vital importancia para la física teórica fundamental.

3 El problema de la masa

Una manera sencilla de ver que la interacción descrita por teorías de campos gauge es de largo alcance es que su acción (6) es invariante bajo cambios de escala en las medidas de longitudes y tiempos. En otras palabras la constante que mide la fortaleza del acoplo g que aparece como un prefactor no posee dimensiones desde el punto de vista espacio-temporal, y por tanto permanece constante bajo dilataciones del espacio-tiempo, lo que explica porque esta teoría puede describir al menos de forma clásica una interacción que posea una escala que da cuenta del alcance espacialmente acotado de la interacción. Éste es en esencia el problema que la teoría cuántica debe resolver y que resultará premiado por el Instituto Clay.

Por otra parte esta invariancia conforme de la teoría clásica ha hecho que el estudio de las soluciones clásicas de la teoría proporcione información muy valiosa acerca de la topología y estructura diferenciable del espacio-tiempo. Ésta es la vía que condujo a Donaldson a probar un famoso teorema acerca de la existencia de diferentes estructuras diferenciables en el espacio-tiempo de Minkowski [16] que el valió la consecución de la prestigiosa medalla Field.

Ahora bien es sabido que al cuantizar un sistema clásico algunas de las simetrías pueden quebrarse y desaparecer en el correspondiente sistema cuántico. Si la simetría clásica de dilataciones de la teoría de Yang-Mills desapareciese en el mundo cuántico no habría ningún problema para que la teoría generase una masa no nula que fuese no sólo responsable de su corto alcance sino también del confinamiento de los quarks.

Sin embargo la cuantización de la teorías gauge no es sencilla. La rutina de cuantización seguida con gran éxito en los sistemas atómicos se enfrentó a un gran problema cuando trató de cuantizar el campo electromagnético. La teoría comenzó a plagarse de predicciones divergentes lo que llevó a uno de sus fundadores Dirac a sombríos pensamientos pesimistas acerca de toda la teoría [4]. Sin embargo dichas dificultades fueron resueltas mediante un proceso que se conoce con el nombre de renormalización cuyo fundamento

[4] "Parece ser que hemos seguido hasta donde es posible el desarrollo lógico de las ideas de la mecánica cuántica tal y como se conocen hoy en día. Teniendo en cuenta que las dificultades son de carácter muy profundo, únicamente pueden ser superadas por un cambio drástico de los fundamentos de la teoría, probablemente tan drástico de como el paso dado de la teoría de las órbitas de Bohr a la mecánica cuántica actual"[17].

estriba en que los parámetros que medimos de los sistemas cuánticos no se corresponden con los parámetros que aparecen en la teoría clásica. Así la carga eléctrica elemental observada no coincide con el parámetro desnudo e que aparece en el Lagrangiano (2). Una vez aceptado este principio no hay ninguna razón para que la carga eléctrica presente en la acción (2) no tenga una dependencia (*renormalización*) en el parámetro auxiliar de control de las divergencias (*regularización*) de forma que el resultado final sea finito. El único requisito es que la predicción surgida de la teoría cuántica sea finita y no dependa de la forma en que este parámetro regulador es introducido. Ahora bien como contrapartida esta solución al problema lleva implícitamente acompañada una dependencia de la carga observada con la escala de energías. En el caso de la electrodinámica esta dependencia viene dada a orden dominante por el flujo del grupo de renormalización

$$E\,\partial_{_E} e^2 = \frac{e^4}{6\pi^2}\log E \tag{7}$$

lo que implica que dicha carga crece con la energía, o lo que es lo mismo al acercarse a la carga. La constante de integración que aparece en la resolución de la ecuación diferencial ordinaria (7) introduce una escala fundamental E_0 en la teoría que rompe la invariancia de escala de la teoría clásica.

La solución en el caso de Yang-Mills no fue tan sencilla. Hasta que Faddeev y Popov [18] no encontraron la necesidad de apoyarse en campos fantasma (sin realidad física) para resolver las dificultades técnicas del método tradicional de cuantización no pudo comenzarse el camino seguido con éxito en caso del electrodinámica clásica. En este esquema pudo comprobarse de forma perturbativa que el mecanismo de renormalización funciona de forma similar al caso electromagnético, aunque sin la necesidad de otros campos materiales dado que el propio campo gauge autointeracciona consigo mismo. La variación de la contante de acoplo con la energía [19, 20]

$$E\,\partial_{_E} g^2 = -\frac{11g^4}{12\pi^2}\log E \tag{8}$$

es en este caso inversa a la de la electodinámica. La carga g disminuye con la energía de forma que a cortas distancias explica el comportamiento casi libre de los quarks en el interior de un nucleón. Esta propiedad puesta de manifiesto por Gross, Politzer y Wilczek en 1973 y mereció la concesión del premio Nobel de Física este mismo año 2004. Sin embargo, aunque la constante de integración que surge de la ecuación (8) rompe con la invariancia de escala, los cálculos perturbativos de altas energías no proporcionan ninguna información sobre el mecanismo de generación de masa y confinamiento que domina el comportamiento de la teoría a bajas energías.

La escala de energía E_0 que surge de la resolución de (8) no sólo rompe la simetría conforme clásica sino también separa dos regímenes de comportamiento de la teoría. Para

energías superiores $E > E_0$ donde asintóticamente existe libertad de movimiento de los quarks, son válidas las predicciones obtenidas por los métodos perturbativos que son genéricos para todas las teorías de campos. Para energías inferiores $E < E_0$ la interacción se vuelve tan fuerte que es capaz de impedir que los quarks abandonar los nucleones y los métodos perturbativos se vuelven ineficaces para analizar el comportamiento de la teoría. En la jerga técnica los especialistas distinguen a los dos regímenes con nombres más sugerentes como libertad ultravioleta y esclavitud infrarroja, respectivamente.

El problema de la masa por lo tanto requiere el desarrollo de nuevos métodos matemáticos que den cuenta de los efectos no perturbativos y que serán especiales para cada teoría, en este caso para la teoría de Yang-Mills no abeliana.

4 Regularización de la Teoría de Yang-Mills

Cualquier intento de construcción rigurosa de la teoría cuántica debe resolver en primer lugar el problema de las divergencias ultravioletas. Para ello debe partirse de una formulación ligeramente modificada de la teoría que produzca sólo resultados finitos y que en un cierto límite renormalizado conduzca a una teoría finita con todas las propiedades exigibles a una teoría relativista de campos cuánticos.

En definitiva el problema se diseccionado en dos partes. La primera consiste en encontrar una teoría regularizada sin divergencias, mientras que la segunda, que es la realmente difícil de analizar, trata de encontrar un procedimiento de tomar el límite ultravioleta de forma que se recupere la teoría cuántica sin divergencias.

Desde un punto de vista muy simplificado el problema que se plantea es como tratar de definir el área de una superficie curva. En primer lugar hay que encontrar una aproximación a la superficie por un mosaico formado pequeñas teselas planas y calcular una aproximación al área. A continuación el área se obtendrá como límite al hacer tender el tamaño de las facetas a cero. Este método consiste en la generalización del método de Riemann para definir la integral de una superficie. Este sencillo problema tiene dos dificultades. En primer lugar, la elección de la forma de las teselas es fundamental. Una elección inadecuada puede producir una definición de área con propiedades indeseadas. En este sentido, la elección de la forma triangular para las teselas es la óptima. En segundo lugar si la superficie es complicada el cálculo del límite puede ser muy costoso y desde luego no estar al alcance de métodos analíticos.

En el caso de Yang-Mills la dificultad es infinitamente superior. El calificativo no es exagerado. En efecto, a esos dos problemas se une que la dimensión de la superficie es infinita, lo que requiere una renormalización en el proceso de tomar el límite.

Aunque la mecánica cuántica fue formulada por Heisenberg en el formalismo Hamilto-

niano, Feynman, inspirado por Dirac, encontró una formulación en el formalismo Lagrangiano que ha resultado ser más eficaz para cuantizar las teorías de campos. El método de Feynman se basa en que todos los efectos observables de la teoría cuántica pueden obtenerse a partir de funciones de correlación de una integral funcional extendida al dominio de los campos clásicos de la exponencial de la acción clásica de la teoría, i.e.

$$\int \delta A \ \mathrm{e}^{\frac{i}{\hbar}S(A)} \tag{9}$$

Obviamente, la notación de la expresión (9) es una puramente formal porque δA no puede designar una generalización inexistente de la integración de Lebesgue ordinaria en dimensión finita.

Aparte, de las divergencias ultravioletas previstas en el proceso de construcción de la integral funcional (9) un nuevo tipo de divergencias aparecen debido a la gran invariancia gauge de la acción (5). En efecto, existe un conjunto de dimensión infinita de campos gauge que dan el mismo valor a la acción $S(A)$. Este problema puede resolverse proyectando la integral (9) a una integral definida exclusivamente en el espacio $\mathcal{M}$ de las clases de campos equivalentes bajo transformaciones gauge. Este espacio $\mathcal{M}$ que se conoce como *espacio de órbitas gauge* es una variedad de dimensión infinita con una geometría y topología altamente no triviales responsables de fenómenos físicos exclusivos de las teorías gauge, como son la existencia de anomalías cuánticas [21, 22] y una familia uniparamétrica de teorías cuánticas de Yang-Mills inequivalentes conocidas como teorías de vacío θ [23, 24].

Una vez más se muestra acertada la analogía con el cálculo del área de una superficie mencionada al comienzo de la sección.

El problema de la regularización de la teoría de Yang-Mills fue durante años un quebradero de cabeza, pero afortunadamente en la actualidad está resuelto satisfactoriamente. La relevancia de la simetría gauge para la consistencia física de la teoría y de la integral (9) hace que cualquier modificación de la misma tendente a eliminar las divergencias debe ser muy cuidadosa con la conservación de esta simetría. Este requerimiento unido a la no linealidad de la simetría gauge hizo que desde el primer momento la regularización de la teoría de Yang-Mills fuese un problema a añadir a los usuales en teorías cuánticas de campos. Desde un punto de vista perturbativo el problema se resolvió satisfactoriamente con el descubrimiento de la regularización dimensional [25, 26]. Sin embargo desde un punto de vista no perturbativo el problema continuó durante varios años más. La solución surgió de dos vías distintas. Por un lado se encontraron regularizaciones que mantienen la continuidad del espacio-tiempo como se había hecho con las teorías de campos más tradicionales ya sea desde el punto de vista de Feynman[27, 28] o desde el punto de vista de Schwinger [29]. Por otro, si se introduce la discretización del espacio-tiempo es posible

regularizar la teoría de una forma reticular más radical [30]. Ventajas del primer tipo de regularización incluyen la conservación explícita de las simetrías relativistas inherentes al espacio-tiempo continuo, mientras que las regularizaciones de tipo reticular aunque las violan, permiten una aproximación numérica más eficiente a la teoría.

En ambos casos se torna necesario realizar un giro en el planteamiento del problema. Además de los expuestos la integral (9) presenta un problema adicional. El integrando es una fase pura lo que no parece facilitar la convergencia de la misma. La solución a este problema genérico del enfoque de Feynman de la cuantización de sistemas clásicos se consigue considerando la extensión analítica a tiempos imaginarios de la acción y todos los observables físicos. Esta propuesta fue seriamente introducida por Symanzik [31] quien mostró como recuperar todos los elementos de la teoría cuántica a partir de su formulación Euclídea. Dicha formulación fue intensivamente utilizada por Wilson [32] para establecer una conexión entre la Teoría de Campos Cuánticos y la Mecánica Estadística permitiendo a ambas aprovechar métodos y técnicas previamente desarrolladas en la otra. Con este enfoque la integral funcional (9) se convierte en

$$\int \delta A \; \mathrm{e}^{-\frac{1}{\hbar} S_\epsilon(A)} \tag{10}$$

donde $S_e(A)$ denota la extensión de la acción (6) para tiempos imaginarios. Como el exponente del integrando es una magnitud negativa hay más posibilidades de conseguir la convergencia de la integral. Efectivamente, esto es así en los dos esquemas de regularización mencionados. Por simplicidad nos limitaremos a describir el correspondiente a la regularización reticular.

Si introducimos una discretización del espacio-tiempo Euclídeo este se convierte en un retículo de puntos situados en los vértices de una familia infinita de hipercubos tetradimensionales que llenan todo el espacio-tiempo (Figura 1). Si tomamos las aristas de todos los cubos con igual longitud a el retículo será muy regular. La invariancia relativista en el espacio-tiempo Euclídeo se reduce a invariancia bajo translaciones y rotaciones en cuatro dimensiones. Obviamente, el reticulado rompe esta simetría pero la esperanza es recuperarla en el límite en que la longitud de las aristas de los cubos básicos tienda a ceros, $a \to 0$.

En campos materiales la regularización de la teoría en el retículo Euclídeo [32] se consigue simplemente restringiendo los campos a sus valores en los vértices del retículo, sustituyendo las derivadas por diferencias entre esos valores en vértices contiguos y las integrales ordinarias en el espacio-tiempo por sumas a todos los vértices. La integral funcional se convierte simplemente en el producto de las integrales a todos los valores de los campos en cada punto del retículo.

En el caso de Yang-Mills como siempre hay una gran diferencia. El hecho de que

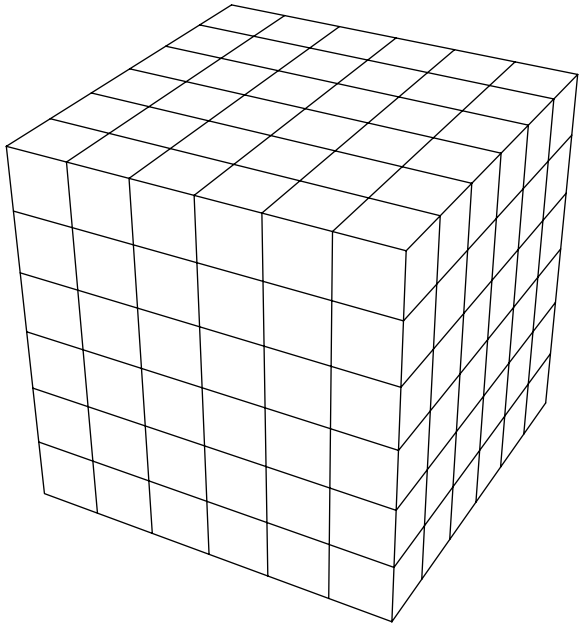


Figura 1.— Retículo espacio-temporal regulador del comportamiento de la teoría de Yang-Mills a cortas distancias

el campo de Yang-Mills esté asociado a un campo gauge, que es objeto geométrico que establece una conexión de referencia entre puntos conectados por un camino que los une mediante el transporte paralelo, obliga que la correspondiente descripción en el retículo no sea la ordinaria de los campos de materia. La descripción más adecuada consiste en asociar un operador unitario a cada arista básica de los hipercubos del reticulado espacio-temporal. El transporte paralelo a lo largo de un camino formado por la unión de aristas contiguas se obtiene por el producto ordenado de los operadores correspondientes a las aristas elementales que lo componen.

La regularización reticular de los campos de Yang-Mills se formula de forma explícita asignando a cada vértice del retículo una coordenada x y a cada arista elemental que parte de ese punto en la dirección positiva de los ejes de coordenadas un índice $\mu = 1, 2, 3, 4$ que indica de que eje se trata, a cada plano elemental que arranca de x dos índices μ, ν con $\mu < \nu$ que indican de que plano se trata. Finalmente cada cara de los hipercubos elementales queda únicamente determinada por la especificación de su origen x y tres índices μ, ν, σ con $\mu < \nu < \sigma$ y el hipercubo correspondiente por su vértice básico x. El campo gauge viene descrito por la familia de elementos $U_\mu(x)$ del grupo $SU(2)$ asociados a cada arista (x, μ) . La acción regularizada viene dada por la suma a todos las planos elementales $P_{\mu,\nu}(x)$ de los hipercubos elementales de la traza de los productos ordenados de los valores del campo gauge en las cuatro aristas que la bordean (Figura 2), i.e.

$$S_\epsilon = \frac{1}{4g^2} \sum_x \sum_{\mu<\nu} \left[2 - \Re \operatorname{Tr} U_\mu(x)\, U_\nu(x+\mu)\, U_\mu^\dagger(x+\mu+\nu)\, U_\nu^\dagger(x)\right] \tag{11}$$

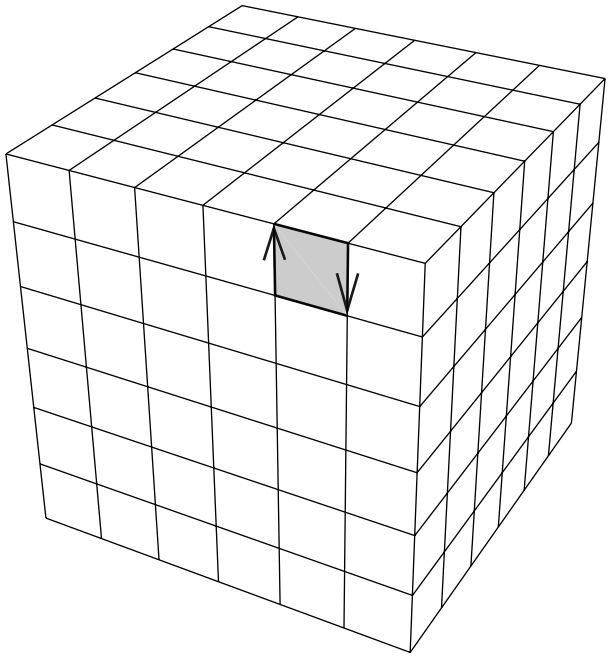


Figura 2.— La acción de Wilson se define a partir de las contribuciones de las cara fundamentales de los hipercubos del retículo

donde $x+\mu$ denota el vértice contiguo a x en la dirección μ y $\Re$ la parte real del número complejo que le sigue.

La integral funcional se completa con la definición de la medida de integración sobre los elementos del grupo en cada arista elemental. Si parametrizamos las matrices de cada arista

$$U = \begin{pmatrix} u_0 + iu_3 & u_1 + iu_2 \\ u_1 - iu_2 & u_0 - iu_3 \end{pmatrix} \tag{12}$$

por cuatro parámetros reales u_0, u_1, u_2, u_3 con $u_0^2+u_1^2+u_2^2+u_3^2 = 1$, la medida de integración de de Haar viene explícitamente dada por

$$dU = du_0\, du_1\, du_2\, du_3\ \delta(u_0^2 + u_1^2 + u_2^2 + u_3^2 - 1) \tag{13}$$

En definitiva la integración funcional (9) viene expresada en la regularización reticular como ($\hbar = 1$)

$$\mathcal{Z} = \prod_{x,\mu} \int dU_\mu(x)\, \mathrm{e}^{-S_\epsilon(U)} \tag{14}$$

Si el volumen espacio temporal es finito la integral es de dimensión finita y convergente. El problema estriba en como conseguir que los promedios de los observables físicos permanezcan finitos cuando la dimensión del retículo se hace infinita (límite termodinámico) y sobre todo cuando la longitud de las aristas básicas del mismo tiende a cero $a \to 0$ (límite continuo).

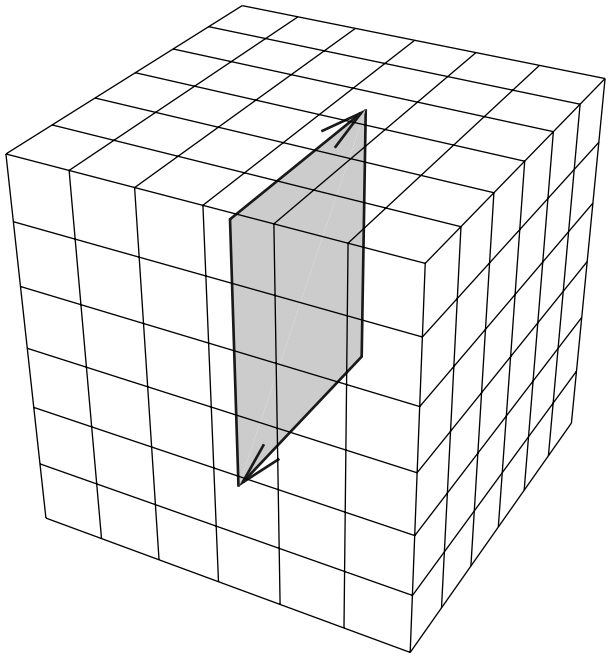


Figura 3.— Bucle de Wilson cuya tensión determina el carácter confinante de la teoría

A partir de dichos promedios deberían poder obtenerse un par de parámetros básicos que corresponden a dos observables físicos la tensión de confinamiento σ y el salto de masa m. Por diversos razonamientos se puede ver que ambos están relacionados con el comportamiento asintótico de dos promedios. La tensión de confinamiento σ viene dada por el comportamiento asintótico del valor esperado del bucle de Wilson

$$\sigma = -\lim_{L\to\infty} \frac{1}{L^2} \left[\log \prod_{x,\mu} \int dU_\mu(x)\, \mathrm{e}^{-S_\epsilon(U)}\, \mathrm{Tr} \prod_{x,\mu\in C} U_\mu(x) - \log \mathcal{Z} \right]. \tag{15}$$

donde C es el contorno de un cuadrado plano formado por L^2 caras planas de hipercubos del retículo (Figura 3).

La masa m es la diferencia de energías entre los dos estados con menos energía de la teoría. En términos de la regularización en el retículo viene definida por el comportamiento asintótico de la función de correlación (Figura 4)

$$m = \lim_{L\to\infty} \frac{1}{L} \log \quad \Big[\mathcal{Z}^{-1} \textstyle\prod_{x,\mu} \int dU_\mu(x)\, \mathrm{e}^{-S_\epsilon(U)}\, \mathcal{P}_{\mu,\nu}(x)\, \mathcal{P}_{\mu,\nu}(x+L\sigma) \tag{16}$$

$$- \mathcal{Z}^{-2} \left(\textstyle\prod_{x,\mu} \int dU_\mu(x)\, \mathrm{e}^{-S_\epsilon(U)}\, \mathcal{P}_{\mu,\nu}(x) \right)^2 \Big]^{-1} \tag{17}$$

donde

$$\mathcal{P}_{\mu,\nu}(x) = \mathrm{Tr}\; U_\mu(x)\, U_\nu(x+\mu)\, U_\mu^\dagger(x+\mu+\nu)\, U_\nu^\dagger(x)$$

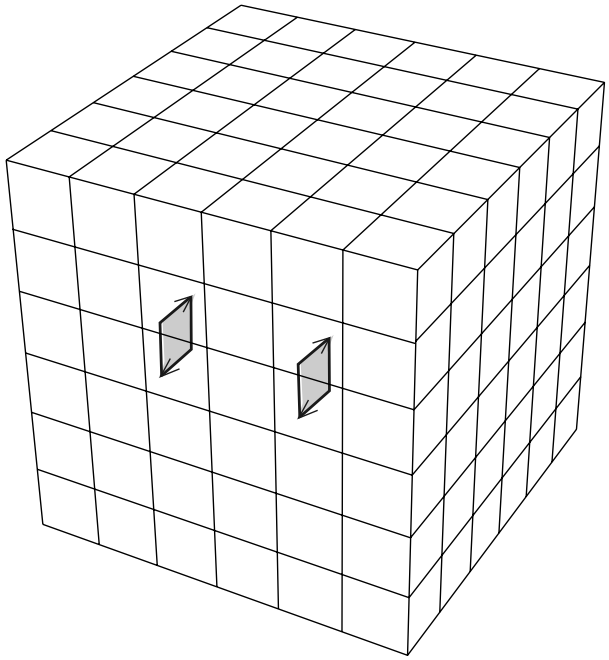


Figura 4.— Función de correlación de dos caras elementales cuyo decaimiento con la distancia determina la masa de la teoría de Yang-Mills

5 El Desafío Matemático de Yang-Mills

El problema de la masa de Yang-Mills puede reducirse a demostrar la siguiente conjetura.

Conjetura de Yang-Mills: *Los límites que definen la tensión de confinamiento σ y el salto de masa m existen y son positivos. El cociente de ambas magnitudes posee un límite finito y positivo cuando la constante de acoplamiento g de la acción (11) tiende a cero*

$$0 < \lim_{g \to \infty} \frac{m(g)}{\sqrt{\sigma(g)}} < \infty. \tag{18}$$

No cabe duda que el problema crudamente planteado en versión reticular es muy difícil de resolver. No se trata de calcular el valor de dicho límite sino solamente que existe es finito y positivo. Pero aun así sin una poderosa estrategia el análisis el problema es inabordable.

La estrategia más prometedora es la basada en el método del grupo de renormalización introducido por Wilson [32] y por el que fue galardonado con el premio Nobel. Esta estrategia consiste en promediar de forma organizada a los valores del campo gauge en solamente algunas aristas del retículo de modo que las integrales que definen las magnitudes m y σ quedan reducidas a integrales similares pero dependientes solamente en los campos de las aristas restantes que a su vez forman un retículo del mismo tipo pero con aristas dobles. Repetir el procedimiento conduce a un proceso iterativo cuyo control

permite establecer cotas sobre los valores de m y σ. Lo interesante es que en cada etapa del proceso de promedios iterativos se obtiene una acción efectiva del mismo tipo que la acción original S_ϵ pero con una constante de acoplamiento mayor. Este cambio puede controlarse mediante la teoría de perturbaciones y los términos restantes de la acción efectiva también pueden acotarse mediante cotas estables bajo las recurrencias del método. En esto consiste el método del grupo de renormalización introducido por Wilson.

El éxito del mismo depende en gran medida en la elección adecuada de las aristas que se promedian y en el control analítico que pueda obtenerse sobre los términos adicionales que aparecen en la acción efectiva.

El método del grupo de renormalización ha conseguido implementarse en otras teorías similares en espacio-tiempos de más baja dimensión [33] y también de forma parcialmente satisfactoria [34] en la propia teoría de Yang-Mills en tres dimensiones (dos dimensiones espaciales y una temporal)[5].

Sin embargo en tres dimensiones espaciales el método presenta innumerables dificultades que no han permitido alcanzar ni siquiera mínimos resultados esperanzadores [35]. Ahora bien la formulación del problema de Yang-Mills en un retículo abre también la posibilidad de utilizar métodos numéricos que no sólo permiten iluminar posibles vías de solución analítica sino que proporcionan resultados de interés para confirmar la validez de la teoría en el mundo de las interacciones fuertes en los regímenes de bajas energías [36]. De acuerdo con los resultados numéricos el valor límite de la masa es [37]

$$\frac{m}{\sqrt{\sigma}} = 3.844 \pm 0.061,$$

lo que está de acuerdo con la conjetura y anima a seguir intentado encontrar una demostración analítica. De hecho los resultados numéricos muestran que la teoría posee un rico espectro de masas (Figura 5).

Desde un punto de vista más exigente la demostración de la existencia de una masa finita no nula en el límite ultravioleta (18) no basta para demostrar la consistencia de la teoría de Yang-Mills como teoría cuántica de campos. Haría falta demostrar la recuperación de la invariancia relativista en límite continuo. En la formulación euclídea esta se reduce a la simetría rotacional. También haría falta probar que las funciones de correlación que intervienen en la definición de la masa (17) satisfacen una desigualdad importante conocida como condición de postividad de Osterwalder-Schrader [38]. Dicha desigualdad refleja en el formalismo Euclídeo el carácter unitario de la evolución temporal de la teoría cuántica. Desde un punto de vista físico habría que demostrar también que la interacción gauge de Yang-Mills con SU(3) confina los quarks de la cromodinámica cuántica y que estos también adquieren masa no nula posiblemente por el mecanismo de rotura de simetría.

[5]En una dimensión espacial la teoría de Yang-Mills es exactamente soluble pero trivial.

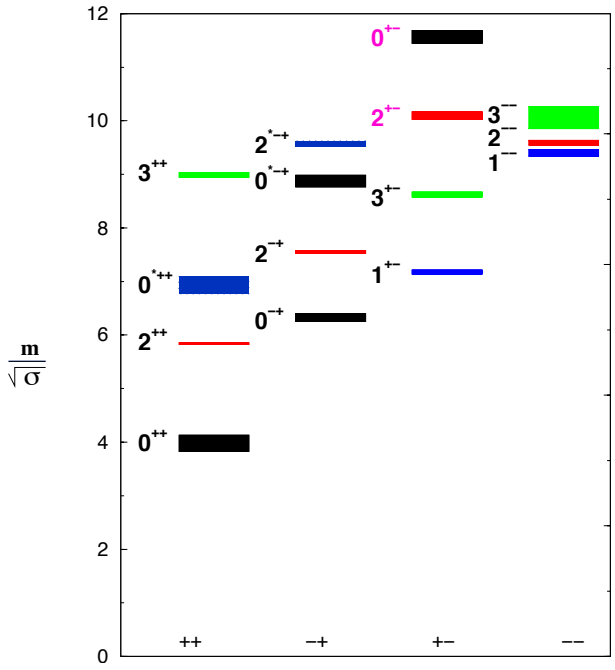


Figura 5.— Valores de los cocientes $\frac{m}{\sqrt{\sigma}}$ para las masas m de los distintos compuestos de gluones de la teoría de Yang-Mills

No obstante muchos de nosotros, entre los que incluyo los miembros del Instituto Clay, aceptaríamos con gran admiración y respeto, como un gran hito teórico, la verificación de la simple conjetura de Yang-Mills (18) en términos puramente analíticos.

Como todo problema matemático de primera linea no puede alcanzar su popularidad hasta que no haya habido varias falsas reclamaciones de resolución, el caso de Yang-Mills posee ya cierta notoriedad también desde este punto de vista. Existen varios intentos de resolución claramente fallidos [39, 40].

El problema de la masa de Yang-Mills constituyó durante más de una década mi el objetivo fundamental de mis pesquisas. Cuando se convocó el galardón Clay yo ya había renunciado con enorme frustración a resolverlo. Espero que la iniciativa Clay anime a otros investigadores más jóvenes y con más recursos a aproximarse a su resolución.

En mi opinión faltan muchas décadas o siglos hasta conseguir una completa resolución. En este sentido me atrevo a conjeturar que si en 2100 algún matemático célebre o institución prestigiosa hacen pública una lista de problemas del siglo, en ella no faltará el problema de la masa en la teoría de Yang-Mills y quizás se le una otro más difícil todavía, el problema de la consistencia de la teoría de gravitación cuántica.

Agradecimientos

A Luis J. Boya por su interés en ver materializado el interesante ciclo de conferencias sobre los problemas Clay del Milenio. Este trabajo está parcialmente financiado por los proyectos del MECD nº FPA2003-1252 y DGA Grupo Teórico de Altas Energías.

Referencias

[1] D. Hilbert, *Mathematicsche Probleme*, Göttinger Nachritten (1900) 253-297

[2] L. Corry, *David Hilbert and the Axiomatization of Physics (1898-1918)*, Kluwer, Dordrecht (2004)

[3] M. F. Rañada, *David Hilbert, Hermann Minkowski, la Axiomatización de la Física y el problema número seis*, Gaceta RSME, **6** (2003) 641-664

[4] K. Gödel, *Über Formal Unentscheidbare Sätze der Principia Mathematica und Verwandter Systeme, I*, Monatshefte für Math. u. Physik, **38** (1931)173-198

[5] J. von Neumann, *Fundamentos Matemáticos de la Mecánica Cuántica*, Ed. CSIC, Madrid (1949)

[6] A. Messiah, *Mecánica Cuántica*, Technos, Madrid (1965)

[7] C. Cohen-Tannoudji, B. Diu, F. Laloë, *Mécanique quantique*, Vols. I and II, Hermann, Paris (1973)

[8] A. Galindo and P. Pascual, *Mecánica Cuántica*, Alhambra, Madrid (1978)

[9] R. Streater , A. Wightman, PCT, *Spin and statistics and all that*, W. A. Benjamin, New York (1964)

[10] B. Simon, *The $P(\phi)_2$ Euclidean (Quantum) Field Theory*,Princeton University Press, Princeton (1974)

[11] J. Glimm, A. Jaffe, *Quantum Physics*, 2nd edition, Springer , Berlin (1987)

[12] Millenium Prize Problems of Clay Mathematics Institute: http://www.claymath.org/prizeproblems

[13] H. Weyl, *Gravitation und Elektrizität*, Sitzunsber. Preuss. Akad. Wiss. **26** (1918)465-478

[14] T. Levi-Civita, *Nozione de parallelismo in una varietà qualunque e conseguente specificazione geometrica della curvatura Riemanniana*, Rend. Circ. Mat. Palermo, **42** (1917) 173-201

[15] C. N. Yang, R. L. Mills, *Conservation of Isotopic Spin and Isotopic Gauge Invariance* , Phys. Rev. **96** (1954) 191-195

[16] S.K. Donaldson, *Anti self-dual Yang-Mills connections over complex algebraic surfaces and stable vector bundles*, Proc. London Math. Soc, **50** (1985)1-26

[17] P. A. M. Dirac, *Principios de Mecánica Cuántica*, Ariel, Barcelona (1958)

[18] L. D. Faddeev, V. N. Popov, *Feynman diagrams for the Yang-Mills field* , Phys. Lett. **B 25** (1967)30

[19] H. D. Politzer, *Reliable perturbative results for strong interactions?*, Phys. Rev. Lett. **30** (1973), 1346-1349.

[20] D. Gross, F. Wilczek, *Ultraviolet behavior of non-abelian gauge theories*, Phys. Rev. Lett. **30** (1973) 1343-1346

[21] S. Adler, *Axial-Vector Vertex in Spinor Electrodynamics*, Phys. Rev. **177** (1969) 2426-2438

[22] J. Bell, R. Jackiw, *A PCAC puzzle:* $\pi_0 \to \gamma\gamma$ *in the Sigma Model*, Nuovo Cimento **60A**(1969)47 9

[23] R. Jackiw, C. Rebbi, *Vacuum Periodicity in a Yang-Mills Quantum Theory* Phys. Rev. Lett. **37** (1976) 172-175

[24] C. G. Callan, Jr. R. F. Dashen, D. J. Gross, *The structure of the gauge theory vacuum* Phys. Lett. **B 66** (1977) 375-381 ; *Toward a theory of the strong interactions* Phys. Rev. **D 17** (1978) 2717-2763

[25] C. G. Bollini, J. J. Giambiagi, *Dimensional renormalization: The number of dimensions as a regularizing parameter*, Nuovo Cimento **B12** (1972) 20

[26] G. 't Hooft, M. Veltman, *Regularization and renormalization of gauge fields*, Nucl. Phys. **B 44** (1972) 189-213.

[27] L. D. Faddeev, A. Slavnov, *Gauge fields: Introduction to quantum theory*, Benjamin-Cummings (1980)

[28] M. Asorey, F. Falceto, *Geometric Regularization of Gauge Theories*, Nucl. Phys. **B 327** (1989) 427

[29] M. Asorey, P. K. Mitter, *Regularized, continuum Yang-Mills process and Feynman-Kac functional integral*, Commun. Math. Phys. **80** (1981) 43

[30] K.G. Wilson, *Confinement of quarks* , Phys. Rev. **D 10** (1974) 2445-2459

[31] K. Symanzik, *Euclidean quantum field theory, in local quantum theory*, Ed. R. Jost, Academic Press, New York (1969) 152-226

[32] K. G. Wilson, *The Renormalization group: critical phenomena and the Kondo problem*, Rev. Mod. Phys. **47** (1975) 773

[33] K. Gawedzki, A. Kupiainen, *A regourous block spin approach to massless lattice theories*, Comm. Math. Phys. **77** (1980) 31-64

[34] T. Balaban, *Ultraviolet stability of three-dimensional lattice pure gauge field theories* , Comm. Math. Phys. **102** (1985) 255-275

[35] T. Balaban, *Renormalization group approach to lattice gauge field theories I, and II: generation of effective actions in a small field approximation and a coupling constant renormalization in 4D*, Comm. Math. Phys. **109** (1987) 249-301; Comm. Math. Phys. **119** (1989) 243

[36] M. Creutz, *Monte Carlo study of quantized SU(2) gauge theory*, Phys. Rev. **D21** (1980) 2308-2315

[37] B. Lucini, M. Teper, *SU (N) gauge theories in four dimensions: exploring the approach to* $N = \infty$,JHEP **0106** (2001) 050

[38] K. Osterwalder and R. Schräder, *Axioms for Euclidean Green's functions*, Com- mun. Math. Phys. **31** (1973), 83-112; Commun. Math. Phys. **42** (1975), 281-305.

[39] E.T. Tomboulis, *Permanent confinement in four-dimensional non-abelian gauge theory*, Phys. Rev. Lett. **50** (1983) 885; **ArXiv:**hep-lat/0409019 (2004)

[40] K.-I. Kondo, *A proof of quark confinement in QCD* **ArXiv:**hep-lat/9808186 (1998)

Monografías de la Real Academia de Ciencias de Zaragoza. **26:** 89–104, (2004).

The Classification of the Finite Simple Groups: An Overview *

Javier Otal †

Departamento de Matemáticas

Universidad de Zaragoza, Zaragoza (Spain)

e-mail: otal@unizar.es

Abstract

We briefly survey on the classification of finite simple groups.

A.M.S. Classification: 20D05, 20D06, 20D08, 20-02.

1 Introduction

The Theory of Groups is one of the oldest and most established branches of Algebra. Nowadays, the concept of group appears not only in the Theory of Groups itself, Algebra and other areas of Mathematics but in different parts of Science, as Physics, Chemistry, Chrystallography, and even Arts. In these areas, the concept of group arises as a kind of measure of the symmetry of a certain configuration to give deep information about it (see [31, 52]).

As it is well-known, different numerical systems form group under a certain law. For example, integers under sum, rationals under sum, non-zero rationals under multiplication, and so on. This is not the right origin of the concept of group. Rather we should take the point of view that this concept is the abstraction of common ideas of some areas (see [37]) as the following ones:

(1) Geometry at the beginning of the 19^{th} Century (*Erlangen Programm*)

(2) Number Theory at the end of 18^{th} Century (*Modular Arithmetic*)

(3) Algebraic Equations at the end of the 18^{th} Century (*Permutations*)

linked to important mathematicians as *Klein, Gauss, Lagrange, Ruffini, Cauchy* and many others. Special efforts are within the work of Galois in the Theory of Algebraic Equations,

*This article is based on a lecture given at a conference held at the Faculty of Sciences of Zaragoza (Spain) in 2004, organized by the Faculty and the Royal Academy of Sciences of Zaragoza

†This research was supported by Proyecto BFM2001-2452 of CICYT (Spain) and Proyecto 100/2001 of Gobierno de Aragón (Spain)

now known as *Galois Theory* (see [30, 46]), as from this work the ideas of "group of an algebraic equation", "normal subgroup", ... are introduced. After some additional work on groups as *permutation groups* (i.e. *Sylow* [45]), the abstract concept of group appears as a major contribution of *Cayley, von Dyck, Kronecker,* and specially *Burnside* (see [38]). In fact, the Theory of Groups came of age with the book by Burnside *Theory of groups of finite order* published in 1897 (see [11]).

1.1 Definitions

The modern definition of a group is usually given in the following way (see [18, 32, 42]). *A group* G is a non-empty set endowed with a binary operation $G \times G \longrightarrow G$ which assigns to every ordered pair of elements $x, y \in G$ a unique element of G (called *the product of* x *and* y) denoted by xy satisfying the following properties:

(1) *Associative law*: if $x, y, z \in G$ then $x(yz) = (xy)z$.
(2) *Identity*: there is an element $1 \in G$ such that $1x = x1 = x$ for all $x \in G$.
(3) *Inverse*: if $x \in G$ there is an element $x^{-1} \in G$ such that $xx^{-1} = x^{-1}x = 1$.

If further G satisfies

(4) *Commutative law*: if $x, y \in G$ then $xy = yx$.

G is said to be *an abelian group* (after *Abel*).

A non-empty subset H of G is said to be *a subgroup* of G, if $xy^{-1} \in H$ for every $x, y \in H$. We indicate this by $H \leq G$. Note that H itself is a group and H and G has the same identity. If $X \subseteq G$, *the subgroup generated by* X is the intersection $\langle X \rangle$ of all subgroups of G containing X. Actually, if $X^{-1} = \{x^{-1} \mid x \in X\}$, it is not hard to see that

$$\langle X \rangle = \{x_1 \ \cdots \ x_r \mid r \geq 1, \ x_i \in X \cup X^{-1}\}.$$

An *isomorphism* between two groups G_1 and G_2 is a bijective map

$$f : G_1 \longrightarrow G_2$$

such that $f(xy) = f(x)f(y)$ for all $x, y \in G_1$. It is said that G_1 and G_2 are *isomorphic* or *have the same type of isomorphy* and denoted $G_1 \cong G_2$.

Let G be a group. If $x \in G$ and $n \in \mathbb{Z}$, *the* n^{th}*–power* x^n *of* x is

$$x^n = \begin{cases} x \ \cdots \ x \ (n \text{ times}, & \text{if } n > 0 \\ 1, & \text{if } n = 0 \\ x^{-1} \ \cdots \ x^{-1} \ (|n| \text{ times}, & \text{if } n < 0 \end{cases}$$

Clearly, $x^n x^m = x^{n+m}$ and $(x^n)^m = x^{nm}$ for every $x \in G$ and $n, m \in \mathbb{Z}$. The least $n > 0$ such that $x^n = 1$ is called *the order of* x. If such an n does not exist, x is said to

have *infinite order*. A group G is said to be *finite* if the underlying set G is a finite set; in this case, the cardinal of G is called *the order of* G and denoted by $|G|$. A non-finite group is called *infinite*. If G is a finite group and $x \in G$ then the order of x is finite and divides $|G|$.

1.2 Examples

(1) $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{Q} \setminus \{0\}, .)$, $(\mathbb{R}, +)$, $(\mathbb{R} \setminus \{0\}, .)$ $(\mathbb{C}, +)$, $(\mathbb{C} \setminus \{0\}, .)$ are examples of infinite abelian groups.

(2) If $n > 1$ is given, *the complex n^{th}–roots of unity* under multiplication, *the integers module n* under sum and *the rotations fixing a regular n-gon* under composition are examples of isomorphic abelian groups of order n. The common type of isomorphy is called *the (finite) cyclic group of order n* and will be denoted by C_n. On the other hand, it can be shown that a finite abelian group is a cartesian product of cyclic groups (see [42])

(3) In general, the plane movements fixing a regular n-gon under composition form a non-abelian group whose type of isomorphy is called *the dihedral group* D_n of order $2n$. This group can be generated by a rotation and an axis symmetry and contains the n rotations fixing the n-gon, that is $C_n \leq D_n$.

(4) A bijection σ of $\{1, \cdots, n\}$ into itself is called *a permutation* on n cyphers. If $n \geq 3$, the permutations on n cyphers under composition of maps compose a finite non-abelian group of order $n!$ called *the symmetric group* of degree n and denoted S_n. Clearly, $D_n \leq S_n$. The n–tuple $(\alpha(1), \cdots \alpha(n))$ contains a number of inversions whose parity is an invariant ruled out by the rule of signs. If follows that the even permutations form a subgroup A_n of the symmetric group called *the alternating group* of degree n. $|A_n| = n!/2$.

(5) Let K be a field and let $n \geq 2$. The invertible matrices

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

whose entries are all in K form a group under multiplication called *the general linear group* on K of dimension n and denoted by $GL(n, K)$. This group can also be thought as the group of all invertible linear operators of a vector space V over K of dimension n and in this case is denoted by $GL(V)$.

If K is a finite field then $GL(n, K)$ is finite. If $|K| = q = p^r$ (p a prime), we denote it by $GL(n, q)$ instead of $GL(n, K)$. We have

$$|GL(n, q)| = (q^n - 1)(q^n - q) \cdots (q^n - q^{n-1}).$$

A subgroup $G \leq GL(K)$ is usually known as *a linear group*. For example, *the special linear group* $SL(n, K)$ composed for all matrices whose determinant is 1 is a linear group.

This group is *the commutator subgroup* of $GL(n, K)$ as it can be generated by the elements of the form $x^{-1}y^{-1}xy,\ x, y \in GL(n, K)$. This remark will be very important in the constructions of classical groups.

On the other hand, it can be shown that *any finite group is linear* as the so-called regular representation or Cayley representation shows ([42]).

1.3 Normal Subgroups and Simple Groups

A subgroup H of a group G defines an equivalence relation on G by

$$x \sim_l y \Longleftrightarrow x^{-1}y \in H.$$

The equivalence class of x is the subset $xH = \{xh \mid h \in H\}$, called *the left coset* of x module H. Similarly,

$$x \sim_r y \Longleftrightarrow yx^{-1} \in H$$

is another equivalent relation and Hx is *the right coset* of an element $x \in G$ module H. Obviously left and right cosets of a group module a subgroup can be different. However, if G is a finite group, since $|xH| = |H| = |Hx|$, the number of left and right cosets coincide. This number is called *the index* of H in G and denoted by $|G : H|$. We have $|G| = |H||G : H|$ (theorem of *Lagrange*).

A subgroup N of a group G is said to be *normal* if $xN = Nx$ for all $x \in G$ (equivalently, if N is G*–invariant*, that is $x^{-1}Nx = N$ for every $x \in G$) and will be denoted by $N \trianglelefteq G$. For example, $C_n \trianglelefteq D_n$ and $A_n \trianglelefteq S_n$. In an arbitrary group G, the trivial subgroup $\langle 1\rangle$ and G itself are always normal subgroups of G. Actually G is said to be a *simple group* if $\langle 1\rangle$ and G are the only normal subgroups of G. The cyclic group C_p of prime order p is simple as well as the alternating group A_n for $n \geq 5$.

If N is a normal subgroup of a group G then the underlying quotient set consisting of the cosets of G module Hcan be endowed with a law of group given by

$$(Nx)(Ny) = Nxy.$$

This group is called *the quotient group* of G by N and denoted by G/N. If G is finite, by Lagrange's theorem, $|G/N| = |G|/|N|$.

2 The Classification of the Finite Simple Groups

Throughout this section, *group will mean finite group.*

As we mentioned above, a fair consequence of Lagrange's theorem establishes that a cyclic group of prime order is simple. In a certain sense, this fact was probably managed

by *Klein*. On the other hand, the simplicity of the alternating groups is closely linked to the unsolvability by radicals of the algebraic equations of degree $n \geq 5$ (see [46]). Other simple groups, as *the Mathieu groups* (the first five *sporadic* groups), were constructed within the XIX Century (see [33, 34, 35]) and other simple groups appeared at first in an isolated way. In fact, there is no a precise date in which the Classification of the Finite Simple Groups (shortly, the CFSG) began. Despite this, some highlighting facts happened (especially around 1955-1958) and they stimulated very much the classification. It is worth to mentioning that its proof is not the usual proof of several theorems as it runs to more or less between 10000 and 15000 journal pages, spread across some 500 separate articles by more than 100 mathematicians, almost all written between 1950 and the early 1980's. Moreover, it was not until the 1970's that a global strategy was developed for attacking the complete classification problem, while, in addition, new simple groups were being discovered especially the twenty-one remainder *sporadic* groups. The full theorem was not even possible in precise form until 1980. See [22, 36] for more details.

From the above comments, it is easily understood that the complexity of that proof will obey to concern ourselves to some facts around the CSFG, which allow to have a more precise idea about it.

2.1 Composition Factors

Let G be a group. *A composition series of* G is a chain of subgroups of G

$$\langle 1\rangle = G_0 \trianglelefteq G_1 \trianglelefteq \cdots G_i \trianglelefteq G_{i+1} \trianglelefteq \cdots G_n = G$$

such that, for every $0 \leq i \leq n-1$, G_i is a normal subgroup of G_{i+1} and G_{i+1}/G_i is a simple group. The set

$$\{G_n/G_{n-1}, \cdots, G_{i+1}/G_i, \cdots, G_1/G_0\}$$

are called *the composition factors of the series*. Since G has finitely many subgroups only, it is immediate that G possesses composition series. We have

Theorem 2.1.1 *(Jordan-Hölder, see [42, Theorem 7.9]) Two composition series of a group have isomorphic composition factors.*

Thus, every group G determines a finite family of simple groups, which say a lot about the structure of the group. The work of *Galois* in characterizing the solvability by radicals of algebraic equations by means of the group of them (see [46]) allowed to introduce the concept of *soluble* (or *solvable*) *group*, which can be characterized by means of composition factors. Actually a group G is soluble if and only if the composition factors of G are cyclic of prime order. It follows that a simple group G is soluble if and only if is abelian, what

happens if and only if G is cyclic of prime order. Apparently solvable groups are placed opposite of non-abelian simple groups and so are the problems considered between these two type of groups. Note that there exist non-solvable groups that are not simple, for example the symmetric groups S_n, $n \geq 5$.

Another important consequence of the theorem of Jordan-Hölder is worth to be mentioned. At a first sight one could think that the knowledge of all simple groups would describe all groups in a similar way that a natural number is described as the product of finitely many primes. Unfortunately, the analogy is false as non-isomorphic groups can have equal composition factors. For example, the cyclic group of order 4 C_4 and the 4–group of *Klein* $C_2 \times C_2$ are not isomorphic but have $\{C_2, C_2\}$ as composition factors. However, the question of the knowledge of simple groups and the CFSG appear to be very natural.

2.2 Towards the Classification

As we mentioned above, the cyclic groups of prime order are the only abelian simple groups. In 1955, Burnside raised the following question.

Conjecture *There are no non-abelian simple groups of odd order.*

This conjecture is shown to be true by the celebrated theorem of Feit-Thompson [17], one of the achievements of this theory, which runs over about 255 pages of the Pacific Journal in Mathematis.

Theorem 2.2.1 *A group of odd order is soluble.*

Corollary 2.2.2 *A non-abelian simple group has even order divisible by* 4.

Previously to this, Richard Brauer (Berlin 1901 – Velmont MA 1977), one of the fathers of the CFSG had proposed an inductive treatment of the problem, of which the theorem of Feit-Thompson is probably the first success. With the aid of his student K.A. Fowler, Brauer had characterized the simplicity of some linear groups, but the fundamental contribution of Brauer and Fowler to this problem can be deduced to their fundamental work on groups of even order ([10]). Let G be a group. *An involution of* G is an element $i \in G$ of order 2. An elemental result of the Theory of Groups asserts that every group of even order has involutions (see [42]). On the other hand, if $x \in G$, *the centralizer of* x *in* G is

$$C_G(x) = \{g \in G \mid gx = xg\}.$$

Now, we can state the theorem of Brauer-Fowler [10].

Theorem 2.2.3 *Let G be a simple group of even order, and let $i \in G$ be an involution. Then $C_G(i) \neq G$, and if $|C_G(i)| = m$ then $|G| \leq (\frac{1}{2}m(m+1))!$.*

This result raises the possibility of characterizing simple groups G in terms of the structure of the centralizer of an involution of G, which is a group of smaller order than G. In fact, the following consequence is immediate.

Corollary 2.2.4 *Let H be a group of even order with an involution $j \in Z(H) = C_H(H)$. Then there are at most finitely many types of simple groups G having an involution i such that $C_G(i) \cong H$.*

As a consequence, Brauer develop a procedure to classify simple groups, which is introduced to the mathematical community with the occasion of the International Congress of Mathematics held at Amsterdam in 1954.

Method of centralizers of involutions

Start with a known non-abelian finite simple group S and an involution $u \in S$, and let $K = C_S(u)$. Consider now simple non-abelian groups G having an involution i such that $C_G(i) \cong S$. There are only finitely many types of such groups G, one of which is S itself by construction. Actually, if there is only one type, we have $G \cong S$, and then a characterization theorem for S has been established in terms of the structure of the centralizer of one of its involutions, a group of smaller order than S. If the attempt fails because there are groups not isomorphic to S among the groups G, there may be previously unknown simple groups among the groups G.

This procedure has been a source of discovery of several new finite simple groups. More details in [18, 19, 20, 42, 43].

2.2.1 Classical Groups and Groups of Lie Type

The so-called *classical groups* are (finite) groups belonging to three big families, namely

- *Linear groups*
- *Symplectic and Orthogonal groups*
- *Unitary groups*

We have already defined the general linear group $GL(n, q)$ and the special linear group $SL(n, q)$, which is its commutator subgroup. By definition, *the projective special linear group* $L(n, q) = PSL(n, q) = SL(n, q)/Z$ is the quotient group of the special linear group by the normal subgroup formed by their scalar matrices. With the exception of lower cases ($n = 2$ and $q = 2, 3$), $L(n, q)$ is know to be simple. *The symplectic and orthogonal groups* can be defined in a similar way starting of subgroups of $GL(n, q)$ consisting of the matrices leaving invariant a given non-degenerate alternating bilinear form or a quadratic form, respectively, on the underlying n–dimensional vector space on which $GL(n, q)$ naturally acts. Then we consider the corresponding commutator subgroup and the quotient groups

of these by their scalar matrices. This procedure leads to families of groups $PSp(n,q)$ and $P\Omega^{\pm}(n,q)$, most of which are simple groups. *The unitary groups* $U(n,q) = PSU(n,q)$ are constructed following a similar procedure; in this case, the starting group is the subgroup of $GL(n,q)$ consisting of the matrices that are invariant by the automorphism α of $GL(n,q)$ given by

$$\alpha(M) = ((\overline{M})^t)^{-1}.$$

The classical groups are known to be analogues of the complex Lie groups A_n, B_n, C_n and D_n (the interested reader can see [15, 51] for details on this and related topics). Although finite analogues of the exceptional Lie groups were constructed by *Dickson* by the early part of the XXthcentury, it was not until 1955 that Chevalley [12] showed by a general Lie-theoretic argument that there exist finite analogues $\mathcal{L}(q)$ of every semisimple complex Lie group $\mathcal{L}(q)$ associated to a finite field with q elements. In particular, he proved the existence of the five families $G_2(q)$, $F_4(q)$, $E_6(q)$, $E_7(q)$ and $E_8(q)$ of exceptional simple groups of Lie type. At the same time, Tits [47] was giving geometric constructions of several of these famu1es. These groups $\mathcal{L}(q)$, with $\mathcal{L} = A_n,\ B_n,\ C_n,\ D_n,\ G_2,\ F_4,\ E_6,\ E_7$ and E_s, are called *the untwisted groups of Lie type* or *the Chevalley groups*. The more technical construction of *the twisted groups* followed soon after (see details in the book by Steinberg [44]).

2.2.2 The Sporadic Groups

The sporadic groups are the 26 simple groups that do not fit into any of the four infinite families of simple groups we have just described (i.e., the cyclic groups of prime order, alternating groups of degree at least five, twisted and untwisted Lie-type groups). They have been constructed in several ways, usually as the automorphisms of a geometric configuration. The smallest sporadic group is the Mathieu group, which has order 7920, and the largest is *the monster group*, which has order 808017424794512875886459904961710757005754368000000
The orders of the sporadic groups given in increasing order are 7920, 95040, 175560, 443520, 604800, 10200960, 44352000, 50232960, ... A summary of sporadic groups can be found in [1, 15, 50, 51]. We only recall their names.

1. *The Mathieu groups*: $M_{11},\ M_{12},\ M_{22},\ M_{23},\ M_{24}$
2. *The Janko groups*: $J_1,\ J_2,\ J_3,\ J_4$
3. *The Conway groups*: $Co_1,\ Co_2,\ Co_3$
4. *The Fischer groups*: $Fi_{22},\ Fi_{23},\ Fi_{24}$
5. *Other groups*: $HS,\ McL,\ He,\ Ru,\ Suz,\ O'N,\ Th,\ Ly,\ HN$

Higman-Sims, McLaughlin, Held, Rudvalis, O'Nan, Suzuki, Thompson, Lyons, Harada-Norton, respectively.

6. *The monster of Fischer-Gries*: M

7. *The baby monster*: F

2.3 The Theorem of Classification

We state the theorem of the classification.

Theorem 2.3.1 *Let G be a finite simple group. Then G is isomorphic to one group of the following*

(1) A cyclic group of prime order;

(2) An alternating group of degree at least five;

(3) A twisted Lie-type group;

(4) An untwisted Lie-type group; or

(5) A sporadic group.

3 After the Classification of the Finite Simple Groups

The classification of the finite simple groups was an ongoing organic process, whose progress in the last decades of the XX^{th} Century since the Odd Order Theorem (the theorem of Feit-Thompson) was extraordinary. Many significant questions and conjectures wee suddenly accessible thanks to the completion of the Classification and its consequences. Surveys on some of this work are available in [19, 20, 36]. Moreover new results were proved by checking the CFSG. It would be impossible to tell in short the numerous progress made.

On the other hand, quite a bit of recent research in finite group theory has developed in response to problems from other areas of mathematics. Here we briefly mention some of the active areas of research (see [43]).

- *Representation Theory.*
- *Maximal Subgroups and Primitive Permutation Representations.*
- *Theory of Graphs.*
- *Field Theory.*
- *Geometry and Topology.*
- *Model Theory.*

Nowadays, within the Theory of Finite Groups, there is not an objective having the importance that the CFSG had. The group-theoretical questions that are now object of study deals with several aspects of *composite groups*.

3.0.1 The Revision Project

The process of *revision* of the classification was for years inextricably associated with the name of Helmut Bender, who began the creative revision of the Odd Order Theorem and successive attempts to explain other different pieces of this research are due to Glauberman, Enguerhard, Goldschmidt, Aschbacher and others. In 1982, Gorenstein, who together with Brauer took an overview of the whole project and steered it to a successful conclusion, launched a *revision project* in which he was joined by Lyons and Solomon. This project is intended to complement the work of the other revision efforts to yield a new and complete proof of the Classification. To realize this, they conceive a project of a dozen books, *the GLS series* (of which five have already appeared: [22, 23, 24, 25, 26]), which Gorenstein was not be able to see a cause of his death in 1992. We simply present here the statement of the Revised Theorem. The status of the project and its interfaces with other revision efforts are available in the GLS series. See also [2] to learn more on the status of the revision project.

A $\mathcal{K}$–group is a finite group H such that if S is a simple quotient of a subgroup of H then S is isomorphic to one of the simple groups listed in Theorem 2.3.1. Here the letter "$\mathcal{K}$" is used as an abbreviation for "known". A group G is said to be *$\mathcal{K}$–proper* if every proper subgroup of G is a $\mathcal{K}$–group. With this terminology, the statement of the Classification Theorem is

Theorem 3.0.2 *If G is a $\mathcal{K}$–proper finite simple group then G is a $\mathcal{K}$–group.*

The proof is an application of the technique known as *minimal counterexample*, that is it is supposed that the result is false and it is chosen a $\mathcal{K}$–proper finite simple group G that is not a $\mathcal{K}$–group with least order in order to arrive to a contradiction.

Clearly the term "$\mathcal{K}$–group" is only used as part of the proof of Theorem 3.0.2. Indeed, that theorem states that *every finite simple group is isomorphic to a known simple group* and it follows that *every finite group is a $\mathcal{K}$–group.*

3.0.2 Monsterology

The monster group M is the most popular among the sporadic groups. It was conjectured in 1972 by Fischer in 1972 and constructed by Griess in 1982; details on its uniqueness were checked by Norton around 1983. It is a group of rotations of a complex space of dimension 196883 whose order ($\simeq 10^{54}$) is bigger than the number of elementary particles of Jupiter!

It is worth to mentioning that many of the properties of the monster group have been discovered before it was constructed. The monster remains the single most tantalizing simple group, with apparent (but as yet mysterious) connections to Kac-Moody Lie theory, quantum field theory, modular functions, and congruence subgroups of $SL(2,\mathbb{Z})$. It is well known how to show that a finite group generated by involutions acts analytically on a Riemann surface. It would be interesting to understand higher-dimensional analytic representations of finite (simple) groups, which would clarify the connections between the Monster (and its subgroups) and classical elliptic modular functions.

A concrete relationship between these ideas was carry out in the work of Borcherds. Richard E. Borcherds is a professor of Mathematics at Berkeley CA, which was awarded by a 1999 Fields Medal by his proof of a conjecture of Conway and Norton ([13, 14, 16]), which establishes a connection between the dimensions of the irreducible representations of the monster group and certain modular functions of weight 0 ([4, 5, 6, 7]). To do this, he invented *the vertex algebras* a generalization of commutative rings with derivation, which are closely linked to elements of *the String Theory* ([5]). The connection produces the so-called *Moonshine* that are a very popular object of study (see [3, 8, 13]).

3.0.3 Infinite Groups

Would it be possible to classify infinite simple groups? The general answer to this question is no because the existence of the so called *Tarski's Monsters*, which are infinite groups with all proper subgroups finite. These groups were constructed by Ol'shanskii [39].

Theorem 3.0.3 *For every odd prime p, there exists an infinite group G whose proper subgroups are cyclic of order p.*

Clearly, this G is simple, can be generated by two elements and all non-trivial elements of G have order p. The groups of Ol'shanskii disprove many conjectures, as the general problem of Burnside, avoid the classification of infinite groups under several criteria, for example, simplicity, and indicate that *a finiteness condition* has to be imposed in classification's problems of infinite groups to avoid their presence.

However, something can be done in the frame of infinite groups and we will give a brief notice about (see also [40]). A group G is said to be *locally finite* if finitely generated subgroups of G are finite. The question of classifying simple infinite locally finite groups was initiated in 1964 by Kegel [28]. Examples, constructions and properties of locally finite groups are available in [29] and also in [41]. In fact, it is very easy to construct infinite locally finite simple groups. For example, for every $n \geq 1$, we embed the symmetric group S_n in the symmetric group S_{n+1} simply adding the cypher $n+1$ in such a way that the latter is fixed and the others move as in the lower symmetric group. This embedding

keeps the parity of the number of inversions so that it induces an inclusion $A_n \longrightarrow A_{n+1}$. Transforming the embeddings in inclusions, we have an infinite chain of subgroups

$$A_1 \leq A_2 \leq \cdots A_5 \leq \cdots A_n \leq A_{n+1} \leq \cdots$$

whose union is an infinite locally finite simple group (called *the restricted infinite alternating group*). We refer to [29] to find more details.

It is rather easy to see that the question can be reduced to the study of countable groups. Indeed, *a group G is simple if and only if it has a local system of subgroups* (i.e. G is the union of an upper directed set of subgroups) *that are countable and simple.* The structure of countable locally finite simple groups is

Theorem 3.0.4 *Let G be a countably infinite locally finite simple group. Then there exists an ascending chain of subgroups of G*

$$G_1 \leq G_2 \leq \cdots \leq G_n \leq G_{n+1} \cdots \leq$$

such that the following conditions are satisfied:

(1) For each $n \geq 1$, G_n contains a maximal normal subgroup M_n such that $G_n \cap M_{n+1} = \langle 1 \rangle$; and

(2) $G = \bigcup_{n \geq 1} G_n$.

Independently, Borovik [9], Hartley and Shute [27] and Thomas [48, 49] showed a fundamental result that can be enounced as follows.

Theorem 3.0.5 *Suppose that G is a group having a countable local system of finite subgroups $\{G_n\}$ such that there exists a Chevalley functor $C(d, --)$ satisfying $G_n = C(d, F_n)$, where F_n is a finite field. Then there exists a locally finite field F such that $G \cong C(d, F)$.*

This essentially asserts that the countably infinite locally finite simple groups are the infinite locally finite versions of the types described in Theorem 2.3.1, that is the locally finite versions of cases (2)-(3)-(4).

A fairly consequence is the following result.

Corollary 3.0.6 *An infinite simple periodic linear group is a group of type Lie-Chevalley over a locally finite field.*

References

[1] Aschbacher, M. Sporadic groups. Cambridge UP, Cambridge 1994.

[2] Aschbacher, M. *The Status of the Classification of the Finite Simple Groups.* Notices of the Amer. Math. Soc. **51** (2004), 736–740.

[3] Bayer, P. *Monstres, cordes, fantasmes i clars de lluna.* Butlletí Soc. Catalana de Mat. **14** (1999), 9–30.

[4] Borcherds, R.E. *Monstrous moonshine and monstrous Lie superalgebras.* Invent. Math. **109** (1992), 405-444.

[5] Borcherds, R.E. *Sporadic Groups and String Theory.* In *First European Congress of Maths. Paris 1992.* Vol. I, pp. 421-431, Birkhäuser, 1992.

[6] Borcherds, R.E. *Automorphic forms on $O_{s+2,2}(\mathbb{R})^+$ and generalized Kac-Moody algebras.* Proceed. ICM94 Zurich **Vol. II**, pp. 744-752, Birkhäuser, 1995.

[7] Borcherds, R.E. *Automorphic forms on $O_{s+2,2}(\mathbb{R})^+$ and infinite products.* Invent. Math. **120** (1995), 161-213.

[8] Borcherds, R.E., Ryba, A.J.E. *Modular Moonshine II.* Duke Math. J. **83** (1996), 435-459.

[9] Borovik, A.V. *Periodic linear groups of odd characteristic.* Soviet Mat. Dokl. **34** (1982), 484–486.

[10] Brauer, R.D., Fowler, K.A. *On groups of even order.* Math. Ann. **62** (1955), 565–538.

[11] Burnside, W. Theory of groups of finite order, 2d ed. Cambridge UP, Cambridge 1911.

[12] Chevalley, C. *Sur certains groupes simples.* Tohoku Math. J. **7** (1955), 14–66.

[13] Conway, J.H. *Monsters and Moonshine.* The Math. Intelligencer **2** (1980), 165-171.

[14] Conway, J.H. *A simple construction for the Fischer-Griess monster group.* Invent. Math. **79** (1980), 513-540.

[15] Conway, J.H., Curtis, R.T., Norton, S.P., Parker, R.A., Wilson, R.A. Atlas of Finite Groups: Maximal Subgroups and Ordinary Characters for Simple Groups. Clarendon Press, Oxford (England) 1985.

[16] Conway, J.H., Norton, S.P. *Monstrous moonshine.* Bull. London Math. Soc. **11** (1979) 308-339.

[17] Feit, W., Thompson, J.G. *Solvability of groups of ood order.* Pacific J. Math. **13** (1963), 775–1029.

[18] Gorenstein, D. Finite groups. Harper Row, New York 1968.

[19] Gorenstein, D. Finite simple groups: an introduction to their classification. Plenum Press, New York 1982.

[20] Gorenstein, D *Classifying the finite simple groups.* Bull. Amer. Math. Soc. **14** (1986), 198.

[21] Gorenstein, D., Lyons, R. *The local structure of finite groups of characteristic 2 type.* Memoirs Amer. Math. Soc. **276** (1983).

[22] Gorenstein, D, Lyons, R., Solomon, R. The Classification of the Finite Simple Groups. Amer. Math. Soc., Providence RI 1994.

[23] Gorenstein, D, Lyons, R., Solomon, R. The Classification of the Finite Simple Groups. Amer. Math. Soc., Providence RI 1996.

[24] Gorenstein, D, Lyons, R., Solomon, R. The Classification of the Finite Simple Groups. Amer. Math. Soc., Providence RI 1998.

[25] Gorenstein, D, Lyons, R., Solomon, R. The Classification of the Finite Simple Groups. Amer. Math. Soc., Providence RI 1999.

[26] Gorenstein, D, Lyons, R., Solomon, R. The Classification of the Finite Simple Groups. Amer. Math. Soc., Providence RI 2002.

[27] Hartley, B., Shute, G. *Monomorphisms and direct limits of finite groups of Lie type.* Quart. J. Math. Oxford Ser. **35** (1984), 49–71.

[28] Kegel, O.H. *Über einfache, lokal endlichen Gruppen.* Math. Z. **95** (1967), 169–195.

[29] Kegel, O.H., Wehrfritz, B.A.F. Locally finite groups. North-Holland, Amsterdam 1973.

[30] Kieman, B.M. *The development of Galois Theory from Lagrange to Artin.* Archiv History Exact Sciences **8** (1971), 40–154.

[31] Kleiner, I. *The evolution of group theory: a brief survey.* Maths. magazine **59** (1986), 195–215.

[32] Ledermann, W. Introduction to the theory of groups of finite order. Oliver and Boyd, Edinburgh and London; Interscience Publ. Inc., New York 1949.

[33] Mathieu, E. *Mémoire sur le nombre de valeurs que peut acquérir une fonction quand on y permut ses variable de toutes les formes possibles.* Crelle, J. **5** (1860), 9–42.

[34] Mathieu, E. *Mémoire sur l'etude de fonctions de plusieurs quantitiés, sur la manière des formes et des substitutions qui les laissent invariables.* Crelle, J. **6** (1861), 241–323.

[35] Mathieu, E. *Sur la fonction cinq fois transitive des 24 quantitiés*. Crelle, J. **18** (1873), 25–46.

[36] McKay, J. ed. Finite groups: Coming of age. Amer. Math. Soc. Contemporary Mathematics, Vol. 45, Providence RI 1982.

[37] O'Connor, J.J., Robertson, E.F. *The Developmnent of group theory.* `htpp://www.gap.dcs.st-and.uk/~history/HistTopics/Development_group_theory.html`

[38] O'Connor, J., Robertson, E.F. *The abstract group concept.* `htpp://www.gap.dcs.st-and.uk/~history/HistTopics/ Abstract_groups.html`

[39] Ol'shanskii, A.Yu. *An infinite group with subgroups of prime orders*. Izv. Akad. Nauk SSSR Ser. Mat. **44** (1980), 309–321.

[40] Otal, J. *Grupos simples infinitos y localmente finitos*. In *Contribuciones Matemáticas en honor de Luis Vigil*. Eds. M. Alfaro et al (ISBN 84600-3440-2). Departamento de Teoría de Funciones. Universidad de Zaragoza. Zaragoza 1984.

[41] Robinson, D.J.S. Finiteness conditions and soluble generalized groups. Springer, Berlin 1972.

[42] Rose, J. A course in the theory of groups. Cambridge UP, Cambridge 1978.

[43] Solomon, R. *On Finite Simple Groups and their Classification*. Notices Amer. Math. Soc. **42** (1995), 231–239.

[44] Steinberg, R. Lectures on Chevalley groups. Yale University, New Haven 1968.

[45] Sylow, P.L. *Théorèmes sur les groupes de substitution*. Math. Ann. **5**(1872), 584–594.

[46] Stewart, I.N. Galois Theory, Third Edition. Chapman Hall/CRC Mathematics Series. Boca Raton, Florida, 2003.

[47] Tits, J. *Groupes simples et géométries associeés*. Proc. Int. Congr. Math. Stockholm (1962), 197–221.

[48] Thomas, S. *An identification theorem for the locally finite non-twisted Chevalley groups*. Arch. Math. **40** (1983), 21–23.

[49] Thomas, S. *The classification of the simple periodic linear groups*. Arch. Math. **41** (1983), 103–116.

[50] Weisstein, E.W. *Sporadic Groups*. From *MathWorld: A Wolfram Web Resource.* `http://mathworld.wolfram.com/SporadicGroup.html`

[51] Wilson, R.A. ATLAS of Finite Group Representation.

[52] Wüssing, H. The Genesis of the Abstract Group Concept. Cambridge, Massachussets, 1984.

Monografías de la Real Academia de Ciencias de Zaragoza. **26:** 105–112, (2004).

La Conjetura de Poincaré. Caracterización de la esfera tridimensional

María Teresa Lozano Imízcoz

Académica Numeraria

Departamento de Matemáticas, Universidad de Zaragoza, 50009 Zaragoza, España

Resumen

La Conjetura de Poincaré ocupa el quinto lugar en la lista de los siete Problemas del Milenio considerados por el Instituto Clay de Matemáticas con sede en Cambridge (Massachuses).

En el año 2004 se ha celebrado el centenario del enunciado de esta célebre Conjetura. También se ha conmemorado el 150 aniversario del nacimiento del gran matemático Henri Poincaré, quien la propuso.[1] Es un problema de Topología que ha dado lugar, directa o indirectamente, a un desarrollo espectacular de esta rama de las Matemáticas durante todo un siglo. Aunque su enunciado es sencillo, la dificultad de su resolución ha dado lugar a la búsqueda de nuevos métodos cada vez que los conocidos hasta ese momento se consideraban insuficientes para alcanzar una solución.

Su creador: Henri Poincaré

Dedicaré unas líneas a glosar la figura del insigne matemático que originó la cuestión que, con una manera de pensar e investigar novedosa para la época, (su intuición geométrica era extraordinaria) aportó importantes avances en el desarrollo de algunas ramas de las matemáticas, en particular de la Topología.

Henri Poincaré nació el día 29 de abril de 1854 en Nancy (Francia). Era hijo de Léon Poincaré, profesor de medicina en la universidad. Destacó en el Lycée de Nancy, donde estudió durante once años (1862-1873), en varias materias, pero sobre todo en matemáticas, según el testimonio de uno de sus profesores que lo define como un monstruo de las matemáticas.[2] En 1873 entró en École Polytechnique. Tras graduarse en 1875, continuó estudios en L'École des Mines. Su tesis doctoral la realizó, bajo la dirección de Charles Hermite, en ecuaciones diferenciales, defendiéndola en París en 1879. Fue profesor en la Universidad de Caen, la Facultad de Ciencias de París, la Sorbonne y la École Polytechnique. Murió en París, el día 17 de julio de 1912, a los 58 años.

Se puede decir que dedicó su vida a la investigación, siendo uno de los últimos sabios integrales, capaces de abarcar muchos aspectos de las matemáticas, la física e incluso la filosofía.

[1] Por estas razones ha habido diversos eventos de conmemoración en los que he tenido la oportunidad de participar. El contenido de esta conferencia recoge material que también fue expuesto en otros foros. En particular en la Real Academia de Ciencias Exactas, Físico-Químicas y Naturales. Su publicación también aquí pretende contribuir a una mayor divulgación del tema.

[2] Carta de Elliot à Liard a un amigo en 1872: "J'ai dans ma classe à Nancy, un monstre de mathématiques, c'est Henri Poincaré".

Figura 1.— Fotografía de H. Poincaré

Su contexto: Analysis sitûs (Topología)

Poincaré es el creador de la topología, que él llamó Analysis sitûs, rama de las matemáticas que se ocupa de caracterizar las propiedades de los objetos que permanecen tras una deformación continua, sin roturas ni pegados. Es algo así como una geometría blanda. Esta manera de pensar sólo es posible en una mente con una extraordinaria capacidad de abstracción espacial. He aquí la definición dada por Poincaré en [10]:

El Analysis sitûs es la ciencia que nos hace conocer las propiedades cualitativas de las figuras geométricas no sólo en el espacio ordinario sino en espacios de más de tres dimensiones.

Más adelante explica este grado de abstracción como la que realizamos en el arte de la geometría: *razonar bien sobre figuras mal realizadas. Las proporciones de las figuras pueden ser alteradas, pero sus elementos no pueden ser trastocados y deben conservar su posición relativa. En otras palabras, las propiedades cuantitativas no son importantes, sino que se deben respetar las propiedades cualitativas, es decir precisamente aquellas de las que se ocupa el Analysis sitûs.*

La esfera topológica

La primera definición de esfera es de origen geométrico. La esfera de dimensión n, S^n, es el conjunto de vectores unitarios del espacio Euclideo de dimensión $n+1$, E^{n+1}.

$$S^n = \left\{ (x_1, x_2, ..., x_{n+1}) | \sum_{i=0}^{n+1} x_i^2 = 1 \right\}$$

En la Figura 2 dibujamos la esfera S^n como una figura geométrica. Así, la esfera S^1 es la circunferencia y la esfera S^2 es la superficie esférica, pero tenemos dificultades cuando tratamos de pintar S^n, $n > 2$. La

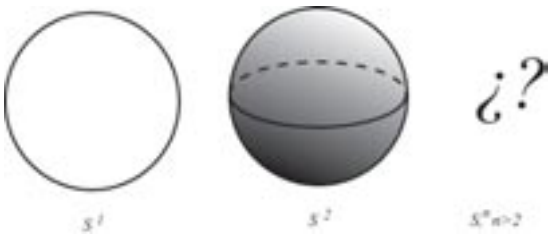


Figura 2.— Esferas.

razón es que no es posible incrustar toda la esfera S^n, $n > 2$ en una porción del espacio tridimensional, que es nuestro campo visual, de la misma manera que no es posible incrustar (sin autointersecciones) la superficie esférica en una porción de plano.

Observamos que una esfera S^n es un espacio con una notable propiedad local: cualquier punto, x, tiene un entorno, U_x, que es como un disco de dimensión n, D^n. En lenguaje topológico, este entorno es *homeomorfo* a un disco (existe una aplicación biyectiva continua $f_x : U_x \longrightarrow D^n$ con inversa continua). Todo espacio con esta propiedad local recibe el nombre de *variedad de dimensión n*. Considerar la esfera S^n únicamente como variedad de dimensión n es olvidar la estructura rígida que le da la geometría, conservando las demás propiedades cualitativas. Así desde el punto de vista topológico, *una curva simple cerrada en el plano o en cualquier espacio de dimensión superior es una esfera* S^1. Esta propiedad caracteriza la esfera de dimensión 1.

La caracterización de la esfera de dimensión 2 con una propiedad intrínseca que la distinga de las otras variedades de dimensión 2 (superficies), se consigue con un invariante algebraico definido y estudiado por Poincaré, que es conocido como grupo fundamental o de Poincaré. Cada espacio topológico X con un punto distinguido x_0, tiene asociado un grupo algebraico, $\pi_1(X, x_0)$, cuyos elementos son clases de equivalencia de caminos que empiezan y terminan en el punto x_0. Dos caminos son equivalentes si se puede deformar uno en el otro de forma continua manteniendo fijos los extremos.

$$\pi_1(X, x_0) = \{[\alpha] \Big| \alpha : [0, 1] \longrightarrow X \text{ continua}, \alpha(0) = \alpha(1) = x_0\}.$$

La composición de dos elementos del grupo es la clase que resulta de recorrer un representante a continuación de otro.

$$\begin{array}{ccc} \pi_1(X, x_0) \times \pi_1(X, x_0) & \longrightarrow & \pi_1(X, x_0) \\ ([\alpha], [\beta]) & \rightarrow & [\alpha \star \beta] \end{array}$$

La esfera es la única superficie cerrada cuyo grupo fundamental es el grupo trivial: $\pi_1(S^2, x_0) = 0$. Es decir, cada camino cerrado se contrae a un punto. *Un camino cerrado en la esfera* S^2 *es el borde de un disco inmerso (una porción de superficie)*, que es el área barrida durante la contracción a un punto.

Origen de la cuestión

La conjetura de Poincaré es una caracterización de la esfera S^3. En 1900, Henri Poincaré, por analogía con la caracterización de la esfera S^2, antes citada, escribió que también la esfera S^3 es la única variedad de dimensión 3 en la que toda curva cerrada bordea una superficie [8].

Cuatro años más tarde, en 1904, él mismo publicó en el quinto complemento al Analysis situs [9], un contraejemplo a esta caracterización. En el artículo describe una variedad de dimensión 3, hoy conocida como esfera homológica de Poincaré, en la que toda curva simple cerrada bordea una superficie, pero no

es homeomorfa a la esfera tridimensional. De hecho, esta variedad tiene un grupo fundamental de 120 elementos y su recubridor universal es la esfera S^3. Se puede definir como el conjunto de dodecaedros (o icosaedros) inscritos en una esfera bidimensional. El artículo termina asegurando que la propiedad que caracteriza la esfera tridimensional es la de tener grupo fundamental trivial. La última frase de este escrito es: *Mais cette question nous entrainerait trop loin.* Nada más exacto. Su investigación ha sido objeto de estudio de muchos topólogos durante todo un siglo. El enunciado preciso de la Conjetura de Poincaré es:

> *Una variedad tridimensional cerrada con grupo fundamental trivial es homeomorfa a la esfera tridimensional.*

Parece una sencilla afirmación y es difícil de imaginar un contraejemplo, pero las demostraciones detalladas que se han ido produciendo en el siglo XX, han resultado incompletas o erróneas.

Visualizando la esfera tridimensional

Para representar la esfera tridimensional como un espacio topológico que podamos comprender e imaginar fácilmente, se utilizan varios procedimientos. Antes de mencionar algunos, recordemos ciertos elementos que se definen en S^3 por analogía con la esfera bidimensional.

Polo Norte $N = (0, 0, 0, 1)$

Polo Sur $S = (0, 0, 0, -1)$

Hemisferio Norte $H_N = \{(x_1, x_2, x_3, x_4) | x_4 \geq 0\}$

Hemisferio Sur $H_S = \{(x_1, x_2, x_3, x_4) | x_4 \leq 0\}$

Ecuador $E = H_N \cap H_S = \{(x_1, x_2, x_3, 0)\} \cong S^2$

Unión de dos bolas tridimensionales

Es claro que cada hemisferio es una bola tridimensional, puesto que

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1 \Leftrightarrow x_1^2 + x_2^2 + x_3^2 = 1 - x_4^2 \leq 1.$$

Ambos hemisferios tienen el ecuador como parte común, luego podemos representar S^3 como dos bolas tridimensionales pegadas por su esfera borde, es decir dos bolas identificadas por el borde mediante un homeomorfismo. Cualquier homeomorfismo (pegado) entre las esferas bordes de las dos bolas produce el mismo resultado (Figura 3).

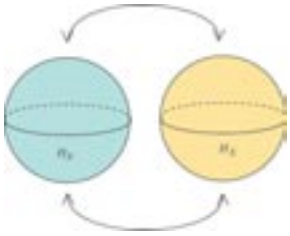


Figura 3.— Dos bolas identificadas.

Proyección estereográfica

La proyección estereográfica desde el norte de la esfera al hiperplano ecuatorial

$$\begin{array}{ccc} e : S^3 \setminus N & \longrightarrow & \mathbb{R}^3 \\ (x_1, x_2, x_3, x_4) & \to & \left(\dfrac{x_1}{1-x_4}, \dfrac{x_2}{1-x_4}, \dfrac{x_3}{1-x_4} \right) \end{array}$$

es un homeomorfismo (aplicación biyectiva continua con inversa continua). Si identificamos la imagen con nuestro espacio ambiente, podemos pensar que la esfera tridimensional es su compactificación con un punto en el infinito. En esta representación el polo sur, S, es el origen de coordenadas, el hemisferio sur es la bola unidad, cuyo borde, la esfera bidimensional unidad, es el ecuador, y el exterior de la bola unidad se corresponde con el hemisferio sur, entorno del punto del infinito que representa al polo norte (ver Figura 4).

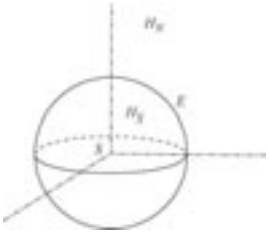


Figura 4.— $\mathbb{R}^3 \simeq S^3 \setminus N$.

Unión de dos toros sólidos

Si se considera un toro sólido ($D^2 \times S^1$) no anudado en $\mathbb{R}^3$, su complementario en $\mathbb{R}^3 \cup \infty \cong S^3$ es también un toro sólido. El pegado se realiza por la superficie tórica del borde identificando el meridiano de un toro con una longitud del otro toro.

Una bola con su borde identificado por reflexión en el ecuador

Si partimos de la representación de S^3 como dos bolas pegadas por su esfera borde y hacemos primero la identificación de un disco, el resultado es homeomorfo a una bola en la que se debe identificar la esfera borde por reflexión en una línea. Ver Figura 5.

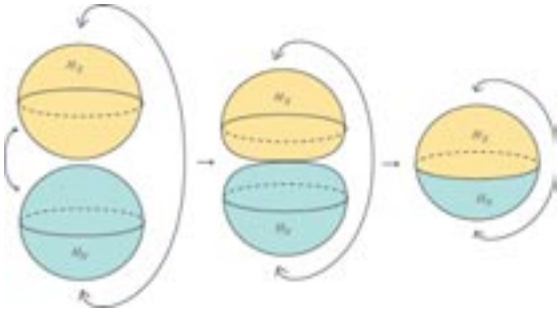


Figura 5.— Una bolas con autoidentificación.

Un dodecaedro con caras identificadas

En la Figura 6 se ha dibujado la intersección de un dodecaedro regular centrado en el origen con el primer octante. El dodecaedro total es el resultado de reflejar en los tres planos coordenados la porción

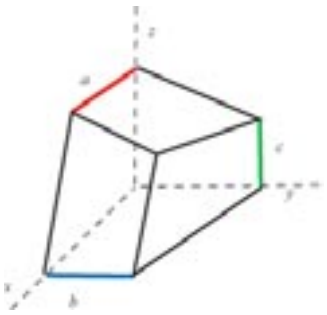


Figura 6.— Una porción del dodecaedro.

dibujada en la Figura 6. Cada cara del dodecaedro tiene un lado marcado con las letras a, b ó c. Si se identifican las caras por giro en el lado marcado, el resultado es S^3, y las líneas marcadas forman el enlace conocido como anillos de Borromeo. En la Figura 7 se ha dibujado este proceso en varias etapas. En 7.1 se muestran las identificaciones a realizar. En 7.2 ya se han identificado las caras que tienen en común las aristas con etiqueta a, que han quedado en el interior del sólido. En 7.3 se ha identificado las caras laterales dejando las aristas marcadas con c como una curva cerrada en el interior; a la vez las aristas marcadas con b se han convertido en una curva cerrada en la esfera borde del sólido, que es una bola en la que se debe identificar el borde por reflexión en b, y eso produce la esfera S^3. En la Figura 4 se han dibujado sólo las curvas correspondientes a las aristas marcadas.

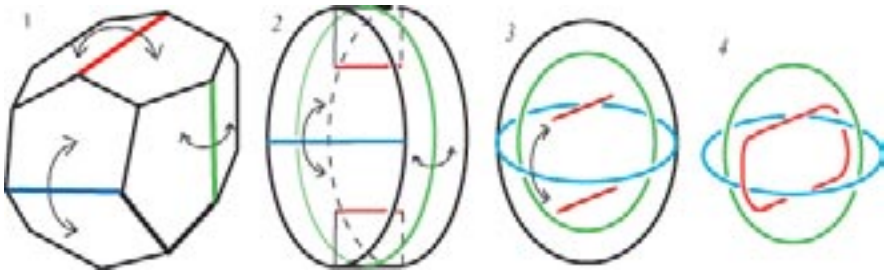


Figura 7.— Identificación en el dodecaedro.

Hacia una solución con métodos topológicos

Durante todo el siglo XX se han utilizado diversos métodos para resolver el problema. En numerosas ocasiones ha habido matemáticos que han presentado a la comunidad matemática una demostración. Pero en un tiempo más o menos breve, otros matemáticos han encontrado algún error o algún detalle no demostrado. Esto da idea de la complejidad del problema. Realmente no es difícil imaginar caminos de demostración, pero lo complicado es comprobar cada paso de la prueba.

En lo que sigue vamos a indicar, con un ilustrativo y simplificado ejemplo, los pasos a seguir para demostrar la conjetura.

En primer lugar fijamos un procedimiento para construir o visualizar todas las 3-variedades cerradas. Existen muchos procedimientos para hacerlo, generalizando los existentes para 2-variedades. Uno de ellos, que se deduce del Teorema 1.1 de [3], es el siguiente:

Teorema 1 *Toda 3-variedad cerrada es unión de un número finito de dodecaedros pegados por sus caras, de manera que cada arista pertenece a uno, dos ó cuatro dodecaedros.*

Los datos en este caso son:

1. n el número de dodecaedros, que corresponde al número de hojas del espacio recubridor en el Teorema

2. los pegados de las caras, que vienen dados por la monodromía del espacio recubridor.

En segundo lugar, como desgraciadamente no conocemos un procedimiento que asocie a cada 3-variedad un único conjunto de datos (las 3-variedades cerradas no están totalmente clasificadas), hay que estudiar las transformaciones en los datos que no modifican la variedad que representan. En nuestro caso hay que estudiar las jugadas que cambian el número de hojas o la monodromía, sin modificar el espacio recubridor.

En tercer lugar hay que relacionar el grupo fundamental de la variedad con los datos que la representan. Para nosotros esto es factible algorítmicamente usando la relación existente entre espacios recubridores y grupos fundamentales.

A continuación, se establecen los datos que corresponden a una 3-variedad cerrada simplemente conexa. Por último, mediante el uso de jugadas que cambian los datos pero no cambian la 3-variedad, se trata de llegar a unos datos que correspondan a la 3-esfera.

Como se puede deducir de este ejemplo (inconcluso hasta el momento), el camino es largo y con sutiles detalles a comprobar.

N.B.: El camino esbozado, es una simplificación del método general aplicado a la representación de toda 3-variedad cerrada como orbifold hiperbólica que cubre a la orbifold hiperbólica $B_{4,4,4}$ (con la 3-esfera como espacio subyacente y los anillos de Borromeo como singularidad de orden 4).

Hacia una solución usando el flujo de Ricci

En los últimos años, se han introducido en el estudio de las 3-variedades topológicas, métodos que utilizan estructuras Riemannianas.

Es bien conocido que cada superficie cerrada admite una estructura Riemanniana de curvatura constante, aquella que tiene su espacio recubridor universal. Las superficies orientables de género mayor o igual que 2, y las no orientables de género mayor o igual que 3, tienen una estructura hiperbólica (curvatura contante negativa); el toro y la botella de Klein tienen una estructura Euclídea (curvatura 0); y la 2-esfera y el plano proyectivo tiene una estructura Riemanniana de curvatura constante positiva. Entonces se deduce que una superficie cerrada simplemente conexa con una estructura Riemanniana de curvatura constante positiva, es necesariamente la 2-esfera.

La analogía en dimensión 3, sugiere un camino de demostración de la conjetura de Poincaré. Se trata de demostrar que toda 3-variedad cerrada simplemente conexa posee una estructura Riemanniana de curvatura constante positiva y, por tanto, es la 3-esfera.

El flujo de Ricci, fue ideado por Hamilton en [2] para variar de manera diferenciable la métrica en la variedad tendiendo hacia una estructura más homogénea. Su definición es la siguiente. Sea M una 3-variedad cerrada. La familia de estructuras Riemannianas diferenciables, $\{g(t)|t \in [0,T)\}$, es flujo de Ricci si satisface $g'(t) = -2Ric(g(t))$, donde $Ric(g(t))$ es el tensor de Ricci de la métrica $g(t)$.

En el mismo artículo, Hamilton demostró el siguiente resultado

Teorema 2 *Si M^3 es una variedad Riemanniana cerrada cuyo tensor de Ricci es definido positivo, entonces la variedad colapsa a un punto bajo el flujo de Ricci. Si se considera el flujo normalizado (volumen constante) converge a una variedad con curvatura constante positiva.*

Entonces, para demostrar la Conjetura de Poincaré es suficiente probar que toda 3-variedad cerrada simplemente conexa admite una estructura Riemanniana cuyo tensor de Ricci es definido positivo.

En los últimos años, el matemático ruso Perelman ha usado el flujo de Ricci para demostrar la Conjetura de Geometrización, propuesta por el gran matemático Thruston, en la que confiere a cada pieza simple de cada 3-variedad una estructura Riemanniana. La Conjetura de Poincaré es consecuencia de la Conjetura de Geometrización. El trabajo de Perelman sobre el tema está contenido en tres artículos, [5, 7, 6] disponibles a través de la red informática de comunicaciones. Sus resultados no han sido todavía publicados en una revista científica, pero son numerosos los matemáticos que han emitido una opinión positiva de su contenido, aunque también son muchos los que esperan cautelosamente su publicación. Esta prevención se justifica por la evolución de las anteriores pruebas anunciadas periódicamente, algunas de las cuales estuvieron vigentes bastante tiempo hasta que, finalmente, fueron desechadas por ser incompletas o erróneas. Para más información sobre el tema recomendamos los artículos recientes [4] y [1].

En la actualidad, podemos decir que el premio prometido por el Instituto Clay para este problema del milenio no ha sido todavía otorgado a ningún matemático. Por tanto debe ser considerado como un problema abierto.

Referencias

[1] Michael T. Anderson. Geometrization of 3-manifolds via the Ricci flow. *Notices Amer. Math. Soc.*, 51(2):184–193, 2004.

[2] R. S. Hamilton. Three-manifolds with positive Ricci curvature. *J. Differential Geom.*, 17(2):255–306, 1982.

[3] H. M. Hilden, M. T. Lozano, J. M. Montesinos, and W. C. Whitten. On universal groups and three-manifolds. *Invent. Math.*, 87(3):441–456, 1987.

[4] John Milnor. Towards the Poincaré conjecture and the classification of 3-manifolds. *Notices Amer. Math. Soc.*, 50(10):1226–1233, 2003.

[5] G. Perelman. The entropy formula for the ricci flow and its geometric applications. *preprint math.DG/0211159*, 2002.

[6] G. Perelman. Finite extinction time for the solutions to the Ricci flow on certain three-manifolds. *preprint math.DG/0307245*, 2003.

[7] G. Perelman. Ricci flow with surgery on three-manifolds. *preprint math.DG/0303109*, 2003.

[8] Henri Poincaré. Second complément a "l'analysis situs". *Proc. London Math. Soc.*, 32:277–308, 1900.

[9] Henri Poincaré. Cinquieme complément a "l'analysis situs". *Rendiconti Circolo mat. Palermo*, 18:45–110, 1904.

[10] Henri Poincaré. Analyse de ses travaux scientifiques. *Acta Math.*, 38:36–135, 1921.

ÚLTIMOS NÚMEROS PUBLICADOS

25.– "Actas de las VI Jornadas de Mecánica Celeste". Jesús Palacián y Patricia Yanguas (Editores). (2004).

24.– "Aspectos ecológicos y culturales del dinamismo rural". Pedro Montserrat Recorder. (2003).

23.– "Los saberes científico y popular en torno a las plantas del Pirineo Aragonés". Luis Villar Pérez. (2003).

22.– "Técnicas analíticas y numéricas en dinámica orbital. Actas de las V Jornadas de Mecánica Celeste". Manuel Palacios y Antonio Elipe (Editores). (2003).

21.– "La extinción de las especies biológicas. Construcción de un paradigma científico". Leandro Sequeiros San Román. (2002).

20.– "Multivariate Approximation and Interpolation with Applications". Mariano Gasca (Editor). (2002).

19.– "Matemáticas en el año 2000". (2001).