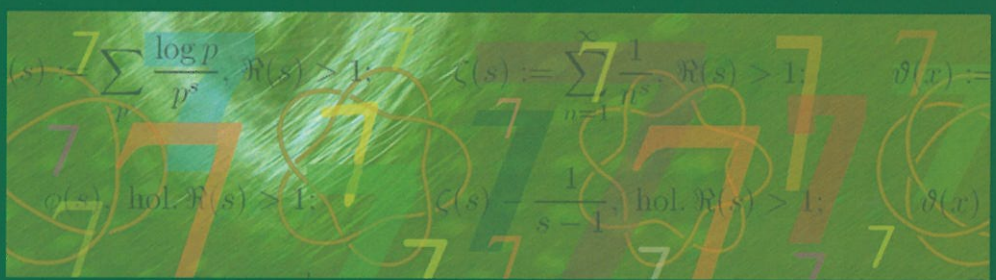
AULA DE CIÈNCIA
I CULTURA, 25

Els set problemes del mil·lenni

IV Cicle Ferran Sunyer i Balaguer



Jordi Quer (ed.)

Albert Atserias
Pilar Bayer
José Ignacio Burgos
María Teresa Lozano
Ignasi Mundet
Josep Pla
Jordi Quer
Anna Rio
Xavier Tolsa

OBRA SOCIAL
Caixa Sabadell

AULA DE CIÈNCIA
I CULTURA, 25

Els set problemes del mil·lenni

Jordi Quer (ed.)

Albert Atserias
Pilar Bayer
José Ignacio Burgos
María Teresa Lozano
Ignasi Mundet
Josep Pla
Jordi Quer
Anna Rio
Xavier Tolsa

Catalogació en publicació de la
FUNDACIÓ CAIXA SABADELL

Els set problemes del mil·lenni : IV Cicle Ferran Sunyer i Balaguer / Jordi Quer (ed) ; [autors] Albert Atserias ... [et al.]. — Sabadell : Obra Social Caixa Sabadell, 2006. — (Aula de ciència i cultura, 25)

Conté: P VS. NP (i el 10è problema de Hibert) / Albert Atserias ... [i altres conferències]

I. Obra Social Caixa Sabadell
II. Cicle Ferran Sunyer i Balaguer
III. Quer, Jordi ed.
IV. Atserias, Albert

1. Matemàtica-Assaigs, conferències, etc.
51(04)

Edita: Fundació Caixa Sabadell, 2007
Carrer d'en Font, 25, 08201 - Sabadell
Tel.: 93 725 95 22 A/e: *obrasocial@caixasabadell.org*

Disseny i producció: *Addenda*, s.c.c.l., www.addenda.es

Impressió: Winihard

Imprès sobre paper ecològic

ISBN: 978-84-95166-67-8

Dipòsit legal: B. 4.821-2007

Índex

Pròleg

Entre el 15 de febrer i el 17 de març de 2005 va tenir lloc la quarta edició del cicle de conferències Ferran Sunyer i Balaguer. Aquestes sèries de conferències de contingut matemàtic per a no especialistes són una iniciativa cultural de la Fundació Caixa Sabadell i el Centre de Recerca Matemàtica que es va iniciar l'any 1999; des d'aleshores s'han organitzat un cop cada dos anys a la seu de la Fundació, a Sabadell. En l'edició de 2005 es va decidir dedicar al cicle a la recerca en matemàtiques, tot presentant una mostra d'alguns dels problemes oberts (o *conjectures*) als quals dediquen actualment els seus esforços els investigadors, explicant la seva importància per a la pròpia matemàtica, i també la seva relació amb altres disciplines científiques: física, enginyeria, computació, meteorologia, tecnologies de la informació i les comunicacions, etc.

Entre l'amplíssim ventall de conjectures a triar, la nostra opció va ser centrarnos en una sel·lecció presentada l'any 2000 amb motiu de l'Any Mundial de les Matemàtiques per l'Institut Clay sota el nom de *Problemes del Mil·lenni*. Es tracta d'una llista de set problemes que pertanyen a diverses àrees de les Matemàtiques, proposats per un grup d'investigadors de reconegut prestigi internacional, que són una mostra representativa dels reptes més importants per a la recerca que s'està fent actualment en Matemàtiques. A més, l'Institut Clay —una fundació privada dels EUA que té per objecte promoure la recerca en Matemàtiques— va oferir un premi d'un milió de dòlars a qui aconsegueixi resoldre cadascun d'aquests problemes. Val a dir que hores d'ara l'únic de tots set que ha estat resolt és el conegut com a *conjectura de Poincaré*. Quan va tenir lloc el cicle ja es comentava que la prova d'aquesta conjectura podia deduir-se d'una sèrie de tres articles que el matemàtic rus Grigori Perelman havia fet públics a través d'Internet, i des d'aleshores l'estudi detallat del contingut d'aquests articles per part dels especialistes ha portat al convenciment general que la prova és correcta. En reconeixement per aquesta contribució, Perelman va ser un dels guardonats amb la Medalla Fields en el Congrés Internacional de Matemàtics celebrat l'agost de 2006 a Madrid, distinció que va rebutjar; aquest rebuig és un cas singular en la història d'aquests guardons, els de més prestigi dins de les matemàtiques. L'Institut Clay preveu un període de dos anys a partir de la publicació d'una demostració abans d'entregar el premi, per tal de permetre l'escrutini i comprovació de tots els detalls. Ara per ara és una incògnita si Perelman acceptarà o no el milió de dòlars quan arribi el moment.

En aquesta quarta edició del cicle Ferran Sunyer i Balaguer, es va dedicar una sessió a cadascun dels set problemes del mil·lenni, a càrrec d'un investigador especialista en el tema, i la sèrie es va completar amb altres tres conferències, fins a un total de deu, on es van discutir altres temes relacionats amb el seu fil conductor: la recerca matemàtica i la resolució de conjectures. En tractar-se de conferències destinades a un públic no especialista, els conferenciants van haver de fer un esforç considerable per tal d'evitar tant com fos possible tots els tecnicismes que hauria comportat una descripció rigorosa dels problemes, i substituir-los per idees intuïtives que permetessin als assistents la comprensió de les idees fonamentals que hi ha en el fons de cadascun d'ells. Breus descripcions del contingut de les conferències, així com el material de suport utilitzat en les presentacions, es poden trobar al web del Centre de Recerca Matemàtica (http://www.crm.cat).

A més d'aquest material, cada conferenciant va redactar un article amb una versió ampliada d'allò que va exposar en la sessió corresponent del cicle, i que és el contingut del recull que ara editem. Sis dels nou capítols de què consta aquest volum estan destinats a presentar problemes del mil·lenni: conjectura de Birch i Swinnerton-Dyer, conjectura de Hodge, relació entre les classes de complexitat P i NP, conjectura de Poincaré, hipòtesi de Riemann i teoria quàntica de Yang-Mills. En els altres tres es parla de la relació d'un d'aquests problemes amb la Criptografia, de l'important paper que han tingut sempre les conjectures dins de la història de les matemàtiques, i de dues conjectures —els problemes de Vitushkin i de Painlevé— que van ser resoltes per un jove matemàtic català, el qual és l'autor del capítol corresponent. Per problemes tècnics, ha estat impossible incloure en aquest recull l'article corresponent a la conferència sobre el problema del mil·lenni que tracta de la resolució de les equacions de Navier-Stokes, impartida pel professor de la Universitat Autònoma de Barcelona Xavier Mora.

Esperem que aquesta publicació sigui útil a totes les persones interessades a tenir una panoràmica sobre les fronteres actuals del coneixement matemàtic a través d'algunes de les grans preguntes que preocupen els més destacats investigadors.

Com a coordinador del cicle, vull agrair al director del Centre de Recerca Matemàtica, Manuel Castellet, i a la Fundació Caixa Sabadell haver-me confiat l'organització d'aquesta quarta edició, a tots els conferenciants per haver acceptat el repte de la sempre difícil tasca de fer arribar a un públic general les idees que hi ha al darrere de teories i conceptes d'un alt nivell d'abstracció i dificultat, i a tots els inscrits al cicle pel seu interès i la seva participació.

Jordi Quer

P *vs.* NP (i el 10è problema de Hilbert)

Albert Atserias

Prof. Titular d'Universitat. Dept. Llenguatges i sistemes informàtics.
Universitat Politècnica de Catalunya

1. INTRODUCCIÓ

L'any 1900, David Hilbert va proposar vint-i-tres problemes irresolts de les matemàtiques del moment com a grans reptes per al segle XX. A dia d'avui, alguns d'aquests problemes es consideren resolts satisfactòriament i d'altres es mantenen, com se sol dir, "oberts". D'entre els problemes resolts en trobem un amb una solució molt singular i sorprenent que volem destacar.

La formulació original del problema és la següent [3]:

> *Determinar un procés mitjançant el qual es pugui establir, en un nombre finit d'operacions, si una equació diofàntica amb una quantitat arbitrària d'indeterminades i coeficients enters té solució en els enters.*

Aquest enunciat es coneix com el 10è problema de Hilbert.

La solució final va ser trobada l'any 1970 per un jove matemàtic rus, de nom Matijasevich, culminant el treball dels seus predecessors americans Davis, Putnam i Robinson (vegeu Matijasevich, 1993). En poques paraules, la solució al 10è problema de Hilbert és la següent:

No n'hi ha cap, de procés.

I això és una solució satisfactòria? El cert és que, amb els matisos necessaris per fer que tant la pregunta com la resposta siguin enunciats pròpiament matemàtics, ho és. A més, és probable que el mateix Hilbert l'hagués considerada satisfactòria. I no tant perquè Hilbert intuís per on anava la resposta, sinó més aviat al contrari, perquè ni el mateix Hilbert podia imaginar-se que la solució requeriria el descobriment d'un fructífer nou camp de la matemàtica: la teoria de la computabilitat.

Però, què té a veure el 10è problema de Hilbert amb la qüestió P *vs.* NP i els problemes del mil·lenni ? Per esbrinar-ho, us convidem a llegir la transcripció d'una hipotètica conversa entre un metòdic matemàtic d'inicis del segle XX, de nom V, i una brillant matemàtica de finals del segle XX, de nom P. En V tot just acaba d'aterrar a l'any 1983

venint directament de l'any 1938, per alguna estranya drecera de l'espai-temps, i ha sentit dir que Yuri Matijasevich va demostrar l'any 1970 que el 10è problema de Hilbert no té solució.

2. PRIMER CONTACTE

V: Què vol dir que no n'hi ha cap, de procés? No ho entenc Dra. P. Això no sembla que es pugui establir matemàticament sense definir prèviament què vol dir el terme *procés* del mateix enunciat del 10è problema de Hilbert.

P: Evidentment, Dr. V; té tota la raó. Qualsevol pregunta matemàtica requereix que els conceptes involucrats estiguin definits amb el rigor propi de la nostra ciència. En el cas del 10è problema de Hilbert, un *procés* s'interpreta simplement com un procediment efectiu, o computable, en el sentit de Gödel, Turing o Church.

V: Conec molt bé el concepte de *funció computable* i no entenc com el problema de les equacions diofàntiques s'hi pot encabir. Les funcions computables són funcions de $N \to N$ i per tant prenen simples nombres naturals com a arguments i retornen simples nombres naturals com a resposta.

P: Doncs també deu saber, per tant, que qualsevol seqüència de nombres naturals es pot representar unívocament per un únic nombre natural, oi? Per exemple, la seqüència de nombres naturals 4,5,4,3 es pot representar unívocament pel nombre natural $2^4 \cdot 3^5 \cdot 5^4 \cdot 7^3$ que s'obté d'elevar els quatre primers nombres primers 2, 3, 5 i 7 a les potències indicades per la seqüència.

V: Sí, és clar. Això es pot fer gràcies al Teorema Fonamental de l'Aritmètica, segons el qual qualsevol nombre natural admet una única representació com a producte de nombres primers. Aquest truc ja el va fer servir Gödel. On vol anar a parar amb això?

P: Doncs, fixi's-hi bé: una equació diofàntica amb incògnites $X_1, \ldots, X_n$ i coeficients enters no és més que una expressió de la forma

$$\sum_{i=1}^{r} \alpha_{0,i} \prod_{k=1}^{n} X_k^{\alpha_{k,i}} = \sum_{i=r+1}^{s} \alpha_{0,i} \prod_{k=1}^{n} X_k^{\alpha_{k,i}}$$

on cada $\alpha_{k,i}$ és un nombre natural.

V: Això és pràcticament la definició.

P: Això mateix. Per tant, una equació diofàntica es pot representar mitjançant la seqüència de nombre naturals següent:

$$n,\ r,\ s,\ \alpha_{0,1}, \ldots, \alpha_{n,1}, \ldots, \alpha_{0,s}, \ldots, \alpha_{n,s}.$$

Per exemple, la famosa equació diofàntica $X^4 + Y^4 - Z^2$ es pot escriure de manera equivalent així

$$1 \cdot X^4 Y^0 Z^0 + 1 \cdot X^0 Y^4 Z^0 = 1 \cdot X^0 Y^0 Z^2,$$

i per tant es pot representar mitjançant la seqüència

$$3, 2, 3, 1, 4, 0, 0, 1, 0, 4, 0, 1, 0, 0, 2,$$

que al seu torn es pot codificar com el nombre $2^3 \cdot 3^2 \cdot 5^3 \cdot 7^1 \cdot 11^4 \cdot 13^0 \cdot 17^0 \cdot 19^1 \cdot 23^0 \cdot 29^4 \cdot 31^0 \cdot 37^1 \cdot 41^0 \cdot 43^0 \cdot 47^2$.

V: Ja ho entenc. Així doncs, la questió del 10è problema de Hilbert consisteix a determinar la computabilitat de la funció $H_{10} : N \to N$ on: si x és un nombre natural tal que la seva factorització única en nombres primers representa una seqüència de nombres naturals que representa una equació diofàntica amb solucions en els enters, aleshores $H_{10}(x) = 1$ i altrament $H_{10}(x) = 0$. És això?

P: Sí, és això. I Matijasevich, seguint la drecera marcada per Davis, Putnam i Robinson, va aconseguir demostrar que H_{10} de fet *no és* computable. Per tant, si entenem que el *procés* que demanava Hilbert es formalitza mitjançant el concepte de funció computable que coneixem, la solució al 10è problema de Hilbert és precisament que no n'hi ha cap; com deiem abans.

V: Que curiós. No em puc imaginar de cap de les maneres com s'ho va fer Matijasevich per aconseguir demostrar això.

P: Si ho vol, Dr. V, el convido a venir demà i li explico una mica per sobre com ho va fer. A més, fer-ho em permetrà introduir-lo a un dels problemes oberts més interessants per al proper segle: P *vs.* NP.

V: Vindré amb molt de gust. Curiosament, aquest matí he sentit una conversa entre dos estudiants que parlaven d'un tal «problema P *vs.* NP». Diu que té relació amb el 10è de Hilbert? Demà m'ho explica.

3. SOLUCIÓ AL 10È PROBLEMA DE HILBERT

P: Bon dia, Dr. V. Va trobar el camí fins a la residència d'investigadors ahir a la nit?

V: No podríem dir que el vaig trobar de seguida. Però l'ambient que es respira pels voltants és d'allò més interessant i no em va faltar distracció...

P: Si li sembla, tornem al 10è problema de Hilbert i li explico què té a veure amb el meu camp d'expertesa. Oi que em va dir que coneix el concepte de *funció computable*? Recordem-lo, de totes maneres.

La Dra. P *s'aixeca del seu seient, s'acosta a la pissarra del seu despatx, agafa un tros de guix i es posa a dibuixar l'esquema de la figura 1.*

Faré servir la definició de Turing, que al cap i a la fi és equivalent a les altres. Turing es va inventar un model de màquina, avui dia anomenada *màquina de Turing*, i és precisament això el que veu en aquest esquema: tenim un conjunt finit d'*estats* $Q = \{q_0, q_1 \cdots\}$, i un *estat actual*, en el dibuix q_3. Tenim una *cinta* dividida en infinites *cel·les*, on cada cel·la pot contenir 0, 1 o ␣ (espai en blanc). I tenim un *capçal* que permet modificar el contingut d'una cel·la, i que es pot desplaçar una posició a l'esquerra o a la dreta en funció de l'estat actual i del símbol que conté la cel·la del capçal. És així com coneix vostè el model de màquina que va proposar Turing?

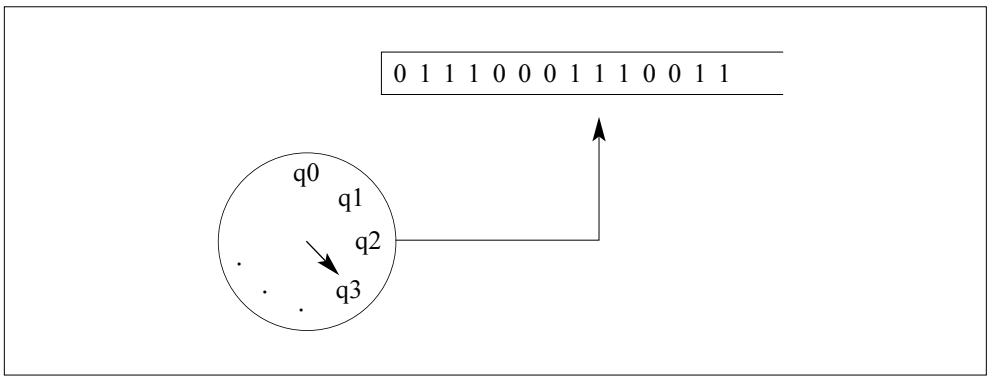


Figura 1. Esquema d'una màquina de Turing.

V: Potser amb una nomenclatura diferent, però sí. Aquesta és la mateixa definició que conec jo. Jo tinc per costum especificar una màquina d'aquestes com una 5-tupla $M=(Q,q_0,q_+,q_-,\delta)$, on Q és el conjunt finit d'estats, q_0, q_+ i q_- són estats especials anomenats *estat inicial*, *estat acceptador* i *estat rebutjador*, respectivament, i δ és la funció de transició

$$\delta: Q \times \{0,1,\sqcup\} \to Q \times \{0,1,\sqcup\} \times \{-1,+1\},$$

on $\delta(q, a) = (q', a', m)$ indica que «si l'estat actual és q i el símbol del capçal és a, aleshores el següent estat és q', el símbol de sota el capçal se substitueix per a', i el capçal es mou a l'esquerra o a la dreta segons si $m=-1$ o $m=+1$ respectivament.»

P: Perfecte. Així doncs tenim la mateixa definició en ment. Direm, doncs, que una funció $f: N \to N$ és *computable* si existeix una màquina de Turing M que, iniciada amb la representació binària d'un nombre natural x continguda a la cinta, i iniciada a

l'estat inicial q_0, arriba a l'estat q_+ i, quan ho fa, el contingut de la cinta és precisament la representació binària de $f(x)$.

V: Representació binària? Què vol dir amb això?

P: Sí, home, representació binària vol dir «representació en base 2». Per exemple, el setè nombre primer en representació binària és 10001, i en representació decimal és 17.

V: I per què es complica tant la vida? N'hi ha prou a iniciar el còmput amb x zeros consecutius continguts a la cinta i exigir que el còmput acabi amb $f(x)$ zeros consecutius continguts a la cinta. No és més senzill?

P: M'oblidava que vostè acaba d'arribar de l'any 1938. No li falta raó quan argumenta que x zeros consecutius són una representació tan vàlida de x com la seva representació binària. De fet, si ho pensa un moment, veurà que x zeros consecutius són la representació *unària* de x. El problema és la longitud de la representació: fixi's bé que la representació unària de x requereix x símbols, mentre que la representació binària de x només en requereix $\lceil \log_2(x+1) \rceil$.

V: I quina importància té, això? La representació decimal de x només en requereix $\lceil \log_{10}(x+1) \rceil$, que és un nombre encara més petit de símbols. Encara no veig quin és el problema amb la representació «unària», com diu vostè.

P: El problema és que la representació unària no és eficient, tenint en compte que disposem de dos símbols 0 i 1. Evidentment, la representació decimal seria encara més eficient, però això requeriria tenir deu símbols per a la cinta de la nostra màquina de Turing, i el resultat ve a ser el mateix.

V: Però, al cap i a la fi, disposar de dos símbols, en comptes d'un o deu, és només una decisió arbitrària que ha près vostè, no?

P: Entenc què vol dir. Potser hauria d'haver començat per aquí. Resulta que les realitzacions físiques de la màquina de Turing que s'han dissenyat fins al moment només són tecnològicament viables (o útils) si es disposa d'un nombre de símbols b petit però més gran que 1. Si el nombre de símbols fos massa gran, seria un maldecap tecnològic realitzar físicament la funció de transició; i, si només hi hagués un símbol possible, la cinta necessària per representar nombres astronòmicament grans com ara 10^{50} hauria de ser tan gran que la realització física seria impossible. I posats que el nombre de símbols sigui petit, s'escull $b=2$ perquè la diferència entre les longituds de les representacions en base 2 o qualsevol altra base fixada més gran que 1 és només d'una constant multiplicativa independent de x.

Però també hi ha una raó més pràctica; dos símbols es poden representar mitjançant dos estats físics elementals: *on/off*, o encès/apagat.

V: Bé, bé; tant és. Al cap i a la fi només és una qüestió de convenció sobre com es representen els nombres naturals. Els humans fa molts anys que fem servir la base 10 i vostè em vol convèncer que al final del segle XX la tecnologia ens farà canviar a base 2. Tant és. Hi ha un noi al MIT que també està capficat amb la base binària. Acaba d'escriure la seva tesi de màster. Com es diu... Claude Shannon, crec.

P: Sí, és clar, en Shannon; conec la seva feina. Vull que entengui, però, que a part d'una convenció, la representació en base unària és *exponencialment* més llarga i, per tant, ineficient si es disposa almenys de dos símbols.

Aclarit aquest punt, tornem, doncs, a les funcions computables i el 10è problema de Hilbert. La clau de tot plegat va ser adonar-se que el conjunt de solucions d'una equació diofàntica permet codificar molta informació. Per exemple, l'equació diofàntica $X - Y^2 = 0$ té solució si, i només si, X és un quadrat perfecte.

De la mateixa manera, l'equació diofàntica $X_1 \cdot Y_1 + X_2 \cdot Y_2 - 1 = 0$ té solució si, i només si, X_1 i X_2 són nombres relativament primers.

En general, Davis, Putnam i Robinson van demostrar, amb molt d'enginy i esforç, que si f és una funció computable qualsevol, aleshores existeix un polinomi exponencial $p_f(X_1, X_2, Y_1, \ldots, Y_m)$ amb coeficients enters tal que l'equació diofàntica exponencial $p_f(X_1, X_2, Y_1, \ldots, Y_m) = 0$ té solució si, i només si, $f(X_1) = X_2$.

V: M'imagino que un polinomi exponencial és com un polinomi normal on, a més, tenim dret a incloure factors del tipus 2^X o X^Y, on X i Y són incògnites.

Però no estavem parlant d'equacions diofàntiques? A què ve això d'equacions diofàntiques exponencials?

P: Sí, té raó; ara hi tornem. Però fixi's que el resultat de Davis, Putnam i Robinson és suficient per concloure que un problema semblant al 10è de Hilbert no té solució.

V: Evidentment, podem concloure que no hi ha cap procés computable per determinar si una equació diofàntica exponencial té solució. Si n'hi hagués, llavors per cada funció computable f hi hauria un procés computable per determinar si $f(x) = 1$ per cada x donat. Però això sabem que no és possible, com va demostrar el mateix Turing amb la seva funció universal. Potser sí que és veritat que això s'acosta a una solució al 10è problema de Hilbert, però no crec que els grecs consideressin equacions diofàntiques exponencials... Li agraïria que tornéssim a les equacions diofàntiques *sense* exponencials.

P: D'acord, d'acord, una mica de paciència. Fixi's que si es pogués trobar un polinomi $p_{\exp}(X_1, X_2, X_3, Y_1, \ldots, Y_r)$ de coeficients enters tal que l'equació diofàntica $p_{exp}(X_1, X_2, X_3, Y_1, \ldots, Y_r) = 0$ té solució si, i només si, $X_1^{X_2} = X_3$, aleshores podríem eliminar tots els exponencials de qualsevol equació diofàntica exponencial $q(Z_1, \ldots, Z_m) = 0$. Pensi-hi un moment.

V: Vejam... El que podem fer és substituir un factor de q que tingui la forma $Z_i^{Z_j}$ per una nova incògnita, digue'm-ne $Z_{i,j}$, i obtenir d'aquesta manera un polinomi exponencial $q'(Z_1, \ldots, Z_m, Z_{i,j})$ amb una exponencial menys. Llavors, l'equació $q(Z_1, \ldots, Z_m) = 0$ té solució si, i només si, l'equació

$$(q'(Z_1, \ldots, Z_m, Z_{i,j}))^2 + (p_{\exp}(Z_i, Z_j, Z_{i,j}, Y_1, \ldots, Y_r))^2 = 0$$

té solució. Interessant...

P: D'això se'n diu reduir el problema de les equacions diofàntiques exponencials al problema de les equacions diofàntiques normals. Doncs aquí és on apareix Matijasevich.

Ell va ser qui va construir el polinomi $p_{\exp}$ i en va demostrar la propietat necessària. La demostració va requerir força teoria de nombres.

Demà m'agradaria explicar-li què és això del problema P *vs.* NP i què té a veure amb el 10è problema de Hilbert. Ara em veig obligada a deixar-lo per atendre qüestions administratives del departament.

4. EL PROBLEMA P *vs.* NP: EXPLICACIÓ INFORMAL

P: Bon dia, Dr. V. Permeti'm que comenci amb un petit joc.

Suposem que el vull convèncer que l'equació diofàntica

$$3 \cdot X^2 \cdot Y \cdot Z^2 - 3 \cdot Z^2 \cdot Y^2 - X^4 - 20 = 0$$

té solució. Què hauria de fer?

V: Bé doncs... doni'm una solució per tal que la verifiqui i m'ho creuré.

P: Exacte. En aquest cas la meva prova que l'equació té solució és $X=2$, $Y=3$, $Z=2$. Per la seva part, per comprovar-ho només haurà de fer unes quantes operacions aritmètiques. Ja ho ha comprovat? Suposem ara, però, que intento convèncer-lo que l'equació diofàntica

$$Z^4 + Y^4 - Z^2 = 0$$

no té cap solució tret de la trivial $X=Y=Z=0$. Com podria convèncer-lo?

V: En aquest cas ho té fàcil. Només m'ha d'explicar la demostració que va donar Fermat. Però crec que ja sé on vol anar a parar. Demostrar que una equació diofàntica no té cap solució pot arribar a ser molt i molt complicat. En canvi, demostrar que una equació té alguna solució sempre és molt senzill; una qüestió molt diferent és com trobar aquesta solució.

P: Aquest és un problema etern de les matemàtiques: enunciats existencials *vs.* enunciats universals. De fet, es desprèn del Teorema d'Incompletesa de Gödel i de la solució al 10è problema de Hilbert que existeixen equacions diofàntiques sense solució per a les quals no existeix cap demostració d'insatisfactibilitat amb les regles habituals de la lògica i els axiomes de Peano, per exemple. Oi que fa ràbia que passin aquestes coses?

V: Doncs sí, la veritat. Però coneixent el Teorema d'Incompletesa de Gödel, no em sorprèn. A mi m'atabalaria encara més que existissin enunciats simples i demostra-

bles per als quals la demostració més curta sigui tan gran que un humà fos incapaç de llegir-la, o, pitjor encara, que hi hagi demostracions curtes però ningú sigui capaç de trobar-les.

P: Caram, Dr. V, quina intuició que té! El problema P *vs.* NP tracta precisament d'això. Abans de donar la definició, però, necessitaré alguns exemples més. Consideri els grafs següents.

La Dra. P *torna a aixecar-se i s'acosta a la pissarra per dibuixar els grafs de la figura 2.*

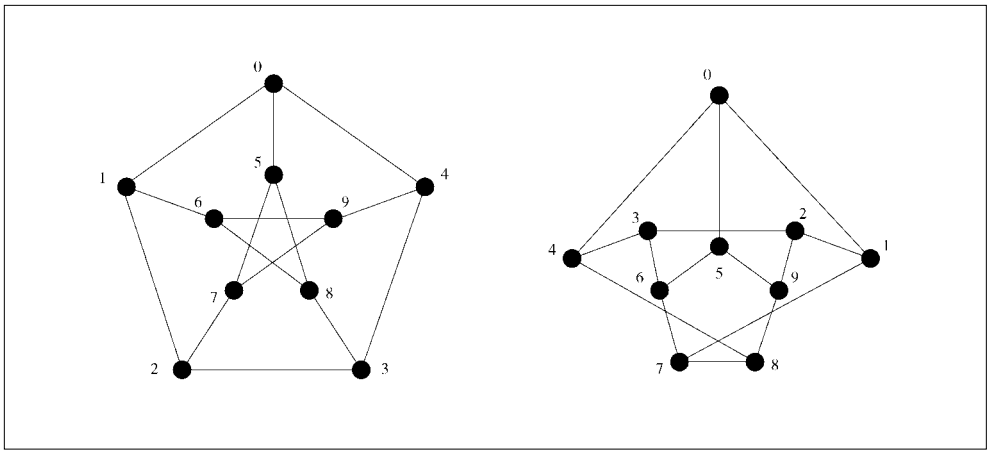


Figura 2. Grafs dibuixats a la pissarra.

Suposem que el vull convèncer que els dos grafs són el mateix, és a dir, isomorfs. Oi que n'hi ha prou que li proporcioni la bijecció entre nodes que ho fa evident?

$$0 \mapsto 0 \qquad 1 \mapsto 4$$
$$2 \mapsto 3 \qquad 3 \mapsto 2$$
$$4 \mapsto 1 \qquad 5 \mapsto 5$$
$$6 \mapsto 8 \qquad 7 \mapsto 6$$
$$8 \mapsto 9 \qquad 9 \mapsto 7$$

Vostè ho pot verificar en un tres i no res. En canvi, trobar la bijecció sembla una tasca molt més difícil; en general, no sembla que hi hagi millors maneres que recórrer les $10! = 3628800$ possibilitats.

Un altre exemple. Suposem que el vull convèncer que el següent sistema d'equacions quadràtiques sobre els enters mòdul dos té solució en els enters mòdul dos:

$$x_0 + x_1 + x_2x_3 + x_4 \equiv 0$$
$$x_1 (x_2 + x_4 + x_6 + x_{10}) + x_0 \equiv 0$$
$$x_1 (x_3 + x_5 + x_7 + x_9 + x_{11} + x_{13}) + x_4 + x_6 \equiv 1$$
$$(1 + x_8)(1 + x_{10}) + x_{12} + x_8(x_7 + x_9) \equiv 0$$
$$x_2 + x_3 (1 + x_4) + x_5x_7 \equiv 1$$
$$x_8 + x_{11} + x_{12} + x_{12} \equiv 1$$
$$x_{13}x_1 + x_{12}x_2 + x_{11}x_3 + x_{10}x_4 \equiv 1$$

N'hi ha prou que li proporcioni la solució

$$x_0 = 0 \qquad x_1 = 1$$
$$x_2 = 1 \qquad x_3 = 1$$
$$x_4 = 0 \qquad x_5 = 1$$
$$x_6 = 1 \qquad x_7 = 1$$
$$x_8 = 0 \qquad x_9 = 1$$
$$x_{10} = 0 \qquad x_{11} = 1$$
$$x_{12} = 1 \qquad x_{13} = 1$$

Una vegada més, la solució es pot verificar molt ràpidament. En canvi, si vol trobar la solució vostè mateix, probablement li calgui recórrer de manera més o menys sistemàtica les $2^{14} = 16384$ possibilitats i verificar-les.

V: En aquests dos exemples que m'ha donat, una solució sempre és més o menys curta en relació al nombre de graus de llibertat del problema en qüestió. La dificultat està en el fet que hi ha moltíssimes possibles solucions, de fet una quantitat exponencial respecte del nombre de graus de llibertat, i no tenim manera de descartar-ne gaires. És això?

P: Sí, és això. Òbviament els meus exemples són petits. Però imagini's què passaria si els grafs tinguessin 200 nodes cadascun, o el sistema d'equacions tingués 150 variables i 300 equacions, com sol passar en problemes d'enginyeria real.

Fixi's, però, que no sempre ens cal recórrer a una quantitat exponencial de possibles solucions. Per exemple, en un sistema d'equacions lineals podem trobar una solució per eliminació gaussiana. En aquest cas, l'àlgebra lineal ens ajuda a trobar la solució molt més ràpidament. Malauradament, no tenim una teoria de l'àlgebra «quadràtica» que ens permeti fer el mateix amb sistemes d'equacions quadràtiques.

V: L'entenc. I, llavors, el problema P *vs.* NP quin és?

P: En poques paraules, el problema P *vs.* NP consisteix a determinar si costa el mateix trobar solucions que verificar-les, encara i quan el conjunt de solucions pot ser exponencialment gran respecte del nombre de variables del problema.

V: Què vol dir que costi el mateix?

P: Doncs que hi hagi un mètode tan eficient per trobar solucions com per verificar-les. Després de dinar li ensenyo les definicions.

5. SOBRE L'EFICIÈNCIA

P: Permeti'm que introdueixi una mica de notació. Per un nombre natural x, faré servir $|x|$ per denotar la longitud de la seva representació en binari. Per tant, $|x| = \lceil \log_2(x+1) \rceil$.

V: Insisteix amb la notació binària. Confio que sigui important.

P: Recordi la definició de computabilitat d'una funció $f : N \to N$: diem que f és computable si existeix una màquina de Turing M que, quan rep un nombre natural x escrit en binari, produeix $f(x)$ també en binari, en un nombre finit de passos. Considerem ara el nombre de passos $t_M(x)$ que fa la màquina M quan rep x com a entrada. Si la màquina fa massa passos per produir la sortida, com per exemple 2^{2^x} passos, serà bastant inútil, no troba? A la pràctica, hi ha raons per creure que la tecnologia mai serà capaç d'implementar les màquines de Turing de manera que executin més de $|x|^4$ passos en un temps físicament raonable, quan x és molt gran.

V: Em semblen molt pocs $|x|^4$ passos. El mínim nombre de passos exigible és $|x|$ perquè l'entrada, ella mateixa, ja té aquesta quantitat de símbols. Per tant, $|x|^4$ passos sembla que sigui només una mica més del que és imperativament necessari.

P: A veure si això el convenç. Fixi-s'hi bé. Suposem que un pas elemental d'una màquina de Turing triga 10^{-6} segons a executar-se. Estem d'acord que això és un temps fantàsticament petit per a un procés físic complex com el d'una transició d'una màquina de Turing?

V: Sí.

P: Considerem ara dues màquines, una que fa $|x|^2$ passos amb entrada x, i una que en fa $2^{|x|}$.

La Dra. P *s'aixeca una vegada més i escriu la següent taula a la pissarra:*

x	$\|x\|^2$	$2^{\|x\|}$
2^{20}	0,4 ms	1,0 segon
2^{30}	0,9 ms	17,9 minuts
2^{40}	1,6 ms	12,7 dies
2^{50}	2,5 ms	35,7 anys
2^{60}	3,6 ms	36,6 segles

Fixi's que el creixement del temps de la primera màquina és molt més suau. Es diu que és *escalable*. En canvi, el segon *explota*, com es diu col·loquialment. Per cert, recordi que entrades de l'ordre de 2^{60} no són excessivament grans. Al cap i a la fi, només són 60 símbols en binari que fem servir per codificar un objecte finit qualsevol, com ara una equació diofàntica, un graf, o un sistema d'equacions.

V: Aquesta taula és del tot elemental, Dra. P. Però he de reconèixer que és molt convincent.

P: Doncs bé; siguem una mica generosos i postulem que una funció f és eficientment computable si existeix una màquina de Turing M que la computa de manera que

$$t_M(x) \leq |x|^c$$

per a alguna constant $c > 0$ independent de x. En aquest cas es diu que M computa f en temps polinòmic perquè el nombre de passos està afitat per un polinomi en el nombre de símbols de l'entrada. La gràcia principal de considerar polinomis de grau arbitrari és que llavors les funcions computables en temps polinòmic estan tancades per composició. Això fa que la definició sigui robusta i fàcil de treballar.

V: Molt bé. Això formalitza el concepte de funció eficientment computable. En particular, això permet formular preguntes del tipus: existeix una funció eficientment computable que, donat un sistema d'equacions sobre els enters mòdul dos, trobi una solució o indiqui que no n'hi ha cap? A priori, no és gens obvi que un algorisme d'aquest tipus pugui existir perquè el conjunt de possibles solucions és exponencialment gran respecte del nombre de variables del sistema. Espero, però, que el mètode de l'eliminació gaussiana proporcioni un algorisme eficient en el cas lineal.

P: Efectivament. La seva intuició no deixa de sorprendre'm. De fet, l'eliminació gaussiana ho aconsegueix en uns n^3 passos elementals, on n és el nombre de variables del sistema lineal. Per tant, l'algorisme és polinòmic respecte del nombre de símbols per representar l'entrada que serà, certament, com a mínim n.

V: Entenc. Com sempre, m'imagino que haurem fixat alguna codificació raonable dels sistemes d'equacions com un nombre natural en binari. De fet, ara començo a entendre perquè és tant important l'eficiència de les codificacions. Si la codificació fos més llarga del que és estrictament necessari, com en el cas de la representació unària dels nombres naturals, aleshores podríem disposar de molt més temps de càlcul per a una deficiència artificial de la codificació i no intrínsica del problema.

P: M'alleuja que hi reflexioni. Posats a fer, fixi's que la representació de l'altre dia per seqüències finites de nombres naturals $a_1, a_2, \ldots, a_m$ com a productes de potències de primers $p_1^{a_1} \cdot p_2^{a_2} \cdots p_m^{a_m}$ no és eficient perquè és molt semblant a la unària.

Més valdria representar els nombres en binari amb un nombre de bits fixat però prou llarg per representar el més gran de tots, i concatenar-los afegint-hi un 1^m0 al principi per saber quants en tenim. Per exemple, la codificació de 4, 5, 4, 3 seria la concatenació de 11110, 100, 101, 100, i 011.

V: Tot i que em sembla entendre-ho, m'avorreix aquesta obsessió que teniu al final del segle XX per fer servir les representacions més eficients possibles. M'esperaria que una teoria robusta no hagués de fixar-se en aquests detalls.

P: De fet, la teoria és robusta sempre que es permetin un mínim de dos símbols. De fet, els detalls de codificació no són pràcticament mai un problema. És qüestió de

pràctica i sentit comú. Tornem ara a les definicions. Plantegem-nos aquests problemes de cerca de solucions en abstracte. Sigui

$$R \subseteq N \times N$$

una relació binària tal que si $(x,y) \in R$, aleshores $|y| \leq |x|^d$ per a alguna constant $d > 0$ independent de x. Per raons òbvies, es diu que R és una relació *equilibrada* polinòmicament. Haurem de pensar que R relaciona les entrades d'un problema amb les seves solucions. En l'exemple anterior, R relacionaria els sistemes d'equacions mòdul dos amb les seves solucions. Per tant, és exigible que es pugui determinar de manera eficient, és a dir, en temps polinòmic respecte de la longitud de l'entrada, si un parell donat (x,y) pertany a R. En altres paraules, suposem que verificar solucions de R és computacionalment fàcil. En aquest cas direm que R és *decidible en temps polinòmic*.

Ara podem plantejar-nos el *problema d'existència* associat a una relació binària R que sigui equilibrada i decidible en temps polinòmic: donada una entrada x, determinar si existeix un $y \leq 2^{|x|^d}$. El conjunt de possibles y pot ser $2^{|x|^d}$, que és exponencial respecte de la longitud $|x|$ de l'entrada x. Per tant, un algorisme de cerca exhaustiva sobre tots els possibles y no és eficient. El problema P *vs.* NP és, doncs, el següent:

> *Determinar si existeixen relacions $R \subseteq N \times N$, equilibrades i decidibles en temps polinòmic, per a les quals el problema d'existència associat a R no és computable en temps polinòmic.*

És clar que si R és una d'aquestes relacions en què el problema d'existència associat no és eficientment computable, aleshores, amb més raó, el problema de cerca de solucions de R no serà eficientment computable.

V: Ja ho entenc. M'imagino que els exemples dels grafs isomorfs i els sistemes d'equacions quadràtiques mòdul dos són bons candidats de problemes difícils. El problema és demostrar que *qualsevol* algorisme que els resolgui requereix fer un nombre de passos que no està afitat per cap polinomi en la longitud de l'entrada.

P: Sí. El proper dia li explico d'on vénen els noms P i NP, i al següent alguna cosa més sobre aquests candidats.

6. ORIGEN DE P *vs.* NP I DEFINICIÓ FORMAL

V: Vam dir que avui m'explicaria d'on vénen els noms P i NP.

P: Això mateix. És molt senzill: P ve de «Polynomial time» en anglès. De fet, P és una classe de conjunts. Per una màquina de Turing M, direm que M accepta l'entrada

x si la màquina arriba a l'estat q_+ quan s'inicia amb x escrit en binari a la cinta. Sigui $L(M)$ el conjunt de les entrades acceptades per M. Definim P com la classe dels conjunts A per als quals existeix una màquina de Turing M tal que $L(M) = A$ i $t_M(x) \leq |x|^c$ per a alguna constant $c > 0$ independent de x. Això defineix P. D'altra banda, les sigles NP provenen de «Non-deterministic Polynomial time», també en anglès.

La Dra. P *fa una pausa per veure la cara que fa el Dr.* V.

V: Expliqui'm això de l'indeterminisme. Què fem, filosofia, ara?

P: Anem a pams. El model de màquina de Turing és perfectament determinista: a cada configuració no terminal de la cinta i l'estat actual, la següent configuració està unívocament determinada. De fet, no podia ser d'una altra manera. De què serviria una màquina amb un comportament imprevisible?

V: De res.

P: Bé doncs, donem-li un significat a una màquina que a cada configuració no terminal té una o més opcions per on continuar. Imaginem-nos que la màquina té més d'una funció de transició i, cada cop que la màquina es troba davant d'una alternativa, es *multiplica* en còpies paral·leles, una per cada opció. Sembla una bajanada, oi?

V: Doncs sí. Sembla el discurs d'un defensor de la mecànica quàntica...

P: No va gaire desencaminat, Dr. V; però no subestimi els seus col·legues físics, faci'm cas. Suposarem, però, a diferència de l'indeterminisme quàntic, que les còpies paral·leles no poden interferir entre si. Direm que la màquina accepta l'entrada x si alguna de les còpies arriba a l'estat acceptador q_+. Com amb les màquines deterministes, sigui $L(M)$ el conjunt de les entrades acceptades per la màquina indeterminista M. Definirem el nombre de passos $t_M(x)$ de M com el màxim dels nombres de passos de les còpies produïdes quan s'inicia el còmput amb l'entrada x. Finalment, definim NP com la classe dels conjunts A per als quals existeix una màquina de Turing indeterminista tal que $L(M) = A$ i $t_M(x) \leq |x|^c$ per a alguna constant $c > 0$ independent de x.

V: M'imagino que el problema P *vs.* NP consisteix a determinar si $\mathrm{P} = \mathrm{NP}$ o $\mathrm{P} \neq \mathrm{NP}$. És això?

P: Correcte. Fixi's que $\mathrm{P} \subseteq \mathrm{NP}$, perquè qualsevol màquina determinista és un cas particular d'una d'indeterminista.

V: Sí, sí, és clar. Però escolti una cosa, Dra. P.

Aquesta formulació de P *vs.* NP no s'assembla de res a l'anterior! M'està dient que són el mateix problema?

P: Evidentment. Deixi'm que ho raoni. Suposem primer que per qualsevol relació R, equilibrada i decidible en temps polinòmic, podem resoldre el problema d'existència associat a R en temps polinòmic. Veurem que sota aquesta hipòtesi, $\mathrm{NP} \subseteq \mathrm{P}$ i, per tant, $\mathrm{P} = \mathrm{NP}$. Per qualsevol màquina indeterminista M tal que $t_M(x) \leq |x|^d$, podem definir una relació R així: $(x, y) \in R$ si y codifica el còmput d'una còpia de M que acaba a l'estat q quan s'inicia amb l'entrada x. No és difícil veure que R és equilibrada i decidible en

temps polinòmic. Per hipòtesi, doncs, el seu problema d'existència associat es pot resoldre en temps polinòmic. Però el problema d'existència de R és precisament el problema de determinar si la màquina indeterminista M accepta o no. Per tant, $\mathrm{NP} \subseteq \mathrm{P}$.

Per veure l'altra implicació, suposem que $\mathrm{NP} \subseteq \mathrm{P}$ i sigui R una relació qualsevol, equilibrada i decidible en temps polinòmic. És prou evident que el problema existencial associat a R es pot resoldre amb una màquina indeterminista així: durant els $|x|^d$ primers passos, fem que la màquina es multipliqui en dues còpies a cada pas, generant un total $2^{|x|^d}$ còpies, una per cada candidat a solució y. A continuació, només cal fer que cada còpia verifiqui el seu candidat a solució, de manera determinista, en temps polinòmic, i fer acceptar les còpies que realment corresponen a una solució.

Per tant, el problema d'existència associat a R pertany a NP i, per hipòtesi, a P. Per tant el podem resoldre en temps polinòmic.

V: Si ho he entès bé, aquest segon argument aprofita l'indeterminisme per «provar» totes les possibles solucions en paral·lel. És evident que això no es pot dur a terme a la pràctica.

P: Precisament la qüestió P *vs.* NP està plantejant si hi ha alguna manera d'evitar simular totes les còpies paral·leles. No deixa de ser curiós que en l'exemple del sistema d'equacions lineals siguem capaços de fer-ho gràcies a l'eliminació gaussiana, un mètode inventat al segle XIX.

V: Sí que és curiós...

7. TEOREMA DE COOK: NP-COMPLETESA

P: Avui m'agradaria tornar sobre els exemples de l'altre dia. Recordem-los. El primer consistia a determinar si dos grafs són isomorfs: donats dos grafs G i H, determinar si existeix una bijecció h del conjunt de vèrtexs de G al conjunt de vèrtexs de H, de manera que $\{u,v\}$ és una aresta de G si, i només si, $\{h(u), h(v)\}$ és una aresta de H. Per qualsevol esquema de codificació raonable, és clar que aquest problema pertany a NP. El segon exemple consistia en determinar si un sistema d'equacions quadràtiques mòdul dos té solució. També pertany a NP. Permeti'm una pregunta. Aprecia alguna relació entre els dos problemes?

V: A part del fet que ambdós problemes pertanyen a NP, a simple vista no en veig cap, de relació.

P: D'acord. M'agradaria argumentar que, de fet, hi ha una relació bastant directa. El que veurem és que el primer es pot transformar en el segon. Fixi's bé. El que volem fer és tranformar un parell de grafs G i H en un sistema d'equacions quadràtiques $S(G,H)$ de manera que G i H siguin isomorfs si, i només si, $S(G,H)$ té solució. Vol veure-ho?

V: Sí.

P: Per cada vèrtex u de G i cada vèrtex v de H, sigui $x_{u,v}$ una variable que pren valors a Z/2Z, els enters mòdul dos. El significat intuitiu de $x_{u,v} = 1$ és que la possible bijecció entre G i H, si n'hi ha, envia u a v. Em segueix?

V: Sí.

P: Ara ens caldrà imposar equacions que forcin que un mateix vèrtex u no s'envii a dos vèrtexs diferents v i v' simultàniament. Per tant, volem que

$$x_{u,v}\, x_{u',v'} \equiv 0 \tag{1}$$

per cada vèrtex u de G i cada parell de vèrtexs v i v' de H amb $v \neq v'$. També necessitarem imposar que si $\{u,u'\}$ és una aresta de G però $\{v,v'\}$ no és una aresta de H, alesho-res o bé u no s'envia a v, o bé u' no s'envia a v'. Per tant, volem que

$$x_{u,v}\, x_{u',v'} \equiv 0 \tag{2}$$

per a cada aresta $\{u,u'\}$ de G i cada parell $\{v,v'\}$ que no és una aresta de H. Òbviament, també imposarem aquesta equació per cada parell $\{u,u'\}$ que no és una aresta de G i cada aresta $\{v,v'\}$ de H. Finalment, ens caldrà especificar que cada u s'envia com a mínim a un v i que cada v rep algun u. De moment ens contentarem de fer-ho amb equacions no quadràtiques: per cada vèrtex u de G imposarem l'equació

$$\prod_v (1 - x_{u,v}) \equiv 0, \tag{3}$$

i per a cada vèrtex v de H imposarem l'equació

$$\prod_u (1 - x_{u,v}) \equiv 0. \tag{4}$$

Vol comprovar que és correcte?

V: A veure... si volem que el producte $\prod_v (1 - x_{u,v})$ sigui congruent amb 0, necessàriament algun $x_{u,v}$ haurà de prendre el valor 1, i per tant no tots seran 0. Per tant, tot u s'envia a algun v, i el mateix per l'altre equació, tot v rep algun u. D'acord, això fa el que volem. Però això no és quadràtic, com vostè diu.

P: Ara ho fem quadràtic. Sigui $\{v_1, \ldots, v_n\}$ una enumeració dels vèrtexs de H. Per cada cada vèrtex u de G i cada $i \in \{0, \ldots, n\}$ introduïm una nova variable $z_{u,i}$. Ara imposem l'equació quadràtica

$$z_{u,i} \equiv z_{u,i-1}(1 - x_{u,v_i}). \tag{5}$$

Per $i = 0$ imposarem l'equació $x_{u,v} \equiv 1$. Amb aquestes restriccions, és clar que

$$\prod_v (1 - x_{u,v}) \equiv z_{u,n}.$$

Per tant, podem substituir l'equació (3) per les equacions quadràtiques (5) i l'equació

$$z_{u,n} \equiv 0. \tag{6}$$

Si fem el mateix per a l'equació (4), haurem aconseguit un sistema quadràtic d'equacions mòdul dos que té solució si, i només si, G i H són isomorfs. Fixi's que la transformació és eficient, és a dir, es pot dur a terme en temps polinòmic.

V: Ara entenc què volia dir quan em preguntava si apreciava alguna relació entre els problemes. Amb aquesta transformació, si sabéssim resoldre sistemes d'equacions quadràtiques mòdul dos en temps polinòmic, també sabriem resoldre el problema de l'isomorfisme de grafs en temps polinòmic.

P: Exacte. D'això se'n diu fer una reducció d'un problema a l'altre, de la mateixa manera que Matijasevich va reduir el 10è problema de Hilbert per a equacions diofàntiques exponencials al 10è problema de Hilbert per a equacions diofàntiques normals. Se'n recorda?

V: Sí que ho recordo. Però deixi'm dir-li una cosa. No em sorprèn gens que hàgim estat capaços d'expressar problemes de grafs de manera algebraica. Descartes ja va expressar problemes de geometria en termes algebraics.

P: És veritat. El que potser és una mica més sorprenent és el que es coneix com teorema de Cook, segons el qual *qualsevol* problema que pertanyi a NP es pot transformar, de manera eficient, al problema de resoldre sistemes d'equacions quadràtiques mòdul dos. De fet, Cook no va considerar sistemes d'equacions directament, sinó fórmules booleanes en forma normal conjuntiva, però ve a ser el mateix en una àlgebra diferent.

V: Qualsevol problema que pertanyi a NP s'hi pot reduir? Això és més interessant perquè vol dir que el conjunt de solucions d'un sistema quadràtic en els enters mòdul dos pot codificar còmputs eficients d'una màquina de Turing. És anàleg al fet que el conjunt de solucions d'una equació diofàntica també pot codificar còmputs, eficients o no, d'una màquina de Turing.

P: Aquesta és l'analogia exacta. Més tard hi tornarem. Ara, però, deixi'm que enunciï el teorema de Cook amb precisió:

Teorema (de Cook 1971 [1], versió sistemes quadràtics) Per a tot A que pertanyi a NP, existeix una funció f computable en temps polinòmic, i tal que per a tota entrada x tenim que $x \in A$ si, i només si, $f(x)$ codifica un sistema d'equacions quadràtiques en els enters mòdul dos que té solució.

Una conseqüència immediata del teorema de Cook és que per demostrar $P = NP$ *n'hi ha prou* a trobar un algorisme eficient per resoldre sistemes quadràtics mòdul dos. De la mateixa manera, per demostrar $P \neq NP$ és suficient i necessari demostrar que no es poden resoldre sistemes quadràtics mòdul dos en temps polinòmic.

V: En certa manera ens podem oblidar de la classe NP. Tota la seva dificultat es troba en un sol problema!

P: En certa manera sí, i això el fa interessant. El teorema de Cook, a més, suggereix una definició. Direm que un conjunt B és NP-complet si B pertany a NP i per qualsevol altre problema A que pertanyi a NP existeix una reducció de A a B com en el teorema de Cook; és a dir, existeix una funció f computable en temps polinòmic tal que $x \in A$ si, i només si, $f(x) \in B$. El teorema diu que el problema de les equacions quadràtiques en els enters mòdul dos és NP-complet.

La Dra. P. *agafa un llibre de tapes toves, negres i molt gastades de la llibreria. El llibre porta per títol:* Computers and Intractability. A Guide to the Theory of NP-Completeness *i els autors són M.R. Garey i D.S. Johnson* [2].

Miri's aquest llibre. Aquí té una col·lecció de més d'un centenar de problemes NP-complets de molt diverses àrees de les matemàtiques i l'enginyeria: des de determinar el número cromàtic d'un graf, passant per problemes de lògica proposicional, fins a problemes de planificació en enginyeria.

V: Tots ells són, doncs, equivalents?

P: Sí. Demostri que un de sol és computable en temps polinòmic i haurà demostrat que tots ho són. Demostri que un de sol no és computable en temps polinòmic i haurà demostrat que cap no ho és. Hi ha qui creu que el fet que hi hagi tantíssims problemes NP-complets de tantes àrees diferents és l'evidència més flagrant que $P \neq NP$. Però això és discutible...

V: Li estic molt agraït per les seves explicacions, Dra. P. Malauradament, demà serà el meu últim dia a la ciutat i abans de marxar m'agradaria que parléssim de l'analogia amb el 10è problema de Hilbert que abans hem aparcat. Podríem parlar-ne demà?

P: I tant. Fins demà, doncs.

8. ANALOGIA, ESTÈTICA, I COMIAT

V: La nostra conversa es va iniciar pel meu interès en la resolució del 10è problema de Hilbert i hem acabat parlant del Teorema de Cook. Durant la conversa ens vam adonar d'una analogia entre les dues coses. Faci'm cinc cèntims de la relació exacta.

P: A vista d'ocell i amb la perspectiva històrica, la solució al 10è problema de Hilbert es pot resumir de la següent manera. 1) Turing identifica un problema no computable, i 2) Davis, Putnam, Robinson i Matijasevich demostren que el problema no computable de Turing es pot reduir al problema de determinar si una equació diofàntica té solució. La conclusió és que el problema de les equacions diofàntiques no és computable.

Però, de fet, el problema que Turing va identificar és molt més que un problema no computable; Turing va identificar un problema *complet per als problemes d'existència* en general. M'explicaré. Sigui

$$R \subseteq N \times N$$

una relació binària. Aquest cop no exigirem que R sigui equilibrada. Suposem que es pugui determinar mitjançant una funció computable si un parell donat (x,y) pertany a R. En aquest cas direm que R és decidible. Com en el cas de les relacions equilibrades, ens podem plantejar el problema d'existència associat a R, però sense exigir que la solució y estigui afitada. El problema d'existència associat a R és, doncs, aquest: donada una entrada x, determinar si existeix un y tal que $(x,y) \in R$. Fins aquí tot és anàleg, oi? El que Turing va identificar és una relació binària i decidible tal que el seu problema d'existència associat no és computable.

V: I quina és aquesta relació binària?

P: Doncs la de la funció universal de Turing: la relació binària que relaciona màquines de Turing M i les entrades x amb els còmputs acceptadors. Però el més interessant del cas és que la relació binària de Turing és com un problema *complet*, és a dir, qualsevol problema d'existència associat a una relació decidible s'hi pot reduir.

V: Per tant, Turing va demostrar l'anàleg al Teorema de Cook per relacions decidibles generals.

P: Més aviat al contrari, amic meu. En Cook va demostrar l'anàleg al resultat de Turing per a relacions binàries equilibrades i decidibles en temps polinòmic.

V: Sí, és clar, perdoni. Viatjar en el temps provoca aquests lapsus.

P: Però seguim amb el 10è problema de Hilbert. Donat que Davis, Putnam, Robinson i Matijasevich van demostrar que el problema de Turing es redueix al problema de les equacions diofàntiques, en deduïm que qualsevol problema d'existència associat a una relació decidible es redueix al problema de les equacions diofàntiques.

V: Interessant. De fet, el mateix problema de determinar si una equació diofàntica té solució és un problema d'existència associat a una relació decidible. Per tant, l'analogia és molt clara: el problema de les equacions diofàntiques és *complet per als problemes d'existència* generals. Però l'analogia hauria estat més estètica si el problema NP-complet de Cook fos més semblant al de les equacions diofàntiques.

P: Cert. Fixi's, però, que el problema de les equacions diofàntiques es pot reduir al problema de determinar si un sistema d'equacions diofàntiques quadràtiques té solució. Només hem d'aplicar el truc de les variables auxiliars per reduir el grau dels monomis.

V: ... molt astut. Així i tot, hauria estat més bonic mantenir intacte el problema clàssic dels grecs i canviar el de Cook.

P: Això mateix devien pensar l'Adleman i el Manders l'any 1978. Se'n recorda que el vaig avisar que hi havia problemes NP-complets per donar i vendre? Obri sius-

plau el llibre de Garey and Johnson per la pàgina 250. Quin problema hi veu, al final de pàgina?

V: A veure... 248, pàgina 250. Equacions diofàntiques quadràtiques: donada una equació diofàntica de la forma $aX^2 + bY = c$ amb incògnites X i Y, determinar si té solució. Suposo que el fet que l'equació sigui quadràtica deu garantir que, si hi ha solucions, alguna estigui afitada polinòmicament... Aquesta analogia sí que és estèticament satisfactòria: el problema de les equacions diofàntiques generals és complet per als problemes d'existència generals, i el problema de les equacions diofàntiques quadràtiques és NP-complet, és a dir, complet per als problemes d'existència amb solucions petites i verificadors eficients. Només falta un petit detall per acabar de lligar-ho tot: demostrar que $P \neq NP$.

REFERÈNCIES

1. S. Cook. «The complexity of theorem proving procedures». A: *3rd Annual ACM Symposium on the Theory of Computing*, pàg. 151-158, 1971.
2. M.R. Garey; D.S Johnson. *Computers and Intractability.* W.H. Freeman and Company, 1979.
3. D. Hilbert. «Mathematical problems». *Bulletin of the American Mathematical Society*, 8: 437-479, 1901-1902.
4. Yu.V. Matijasevich. *Hilberts's Tenth Problem*. MIT Press, 1993.

La hipòtesi de Riemann

Pilar Bayer
Facultat de Matemàtiques
Universitat de Barcelona

INTRODUCCIÓ

La hipòtesi de Riemann, formulada de passada per Bernhard Riemann en una memòria de 1859 dedicada a l'estudi dels nombres primers, afirma que tots els zeros no trivials de la funció zeta estan situats sobre una mateixa recta. La seva demostració és un vell somni, encara no acomplert, de la comunitat matemàtica.

La funció zeta és, de lluny, l'eina analítica més important de què es disposa per a l'estudi dels nombres primers. Una demostració de la hipòtesi de Riemann indicaria que la ment humana ha estat capaç d'endinsar-se en la comprensió de les misterioses lleis que regeixen la distribució d'aquests nombres.

L'interès de la hipòtesi de Riemann fou ja destacat per David Hilbert l'any 1900. La recerca de la seva demostració formava part del problema vuitè de la coneguda llista de vint-i-tres problemes oberts que presentà a l'*International Congress of Mathematicians*, celebrat a París.

Més de 10 bilions de zeros de la funció zeta calculats fins avui, tots alineats en la recta crítica, i nombroses investigacions sobre el tema corroboren, d'una banda, l'extraordinària intuïció de Riemann i, de l'altra, l'enorme dificultat del problema.

En incloure la hipòtesi de Riemann en la llista dels *Set problemes del mil·lenni*, l'Institut Clay es fa ressò d'una pregunta omnipresent en nombroses llistes de problemes oberts.

1. ANTECEDENTS DE LA FUNCIÓ ZETA DE RIEMANN

Un precedent de la funció zeta de Riemann es troba en la sèrie harmònica

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots$$

El nom d'aquesta sèrie obeeix al fet que cadascun dels seus termes és la mitjana harmònica dels dos termes contigus:

$$\frac{1}{n} = \frac{2\,\dfrac{1}{n-1}\,\dfrac{1}{n+1}}{\dfrac{1}{n-1}+\dfrac{1}{n+1}}$$

Nicolas d'Oresme (1323-1382), en el text *Questiones super geometricam Euclidis*, donà la següent demostració rigorosa de la divergència de la sèrie harmònica: les sumes

$$1+\frac{1}{2}>\frac{1}{2}+\frac{1}{2},\qquad \frac{1}{3}+\frac{1}{4}>\frac{1}{4}+\frac{1}{4},\qquad \frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8}>\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}$$

són totes més grans que 1/2, i així fins a l'infinit. Per tant,

$$1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots=\infty.$$

Pietro Mengoli (1626-1686) s'interessà per la suma dels inversos dels quadrats. Leonhard Euler (1707-1783) en proporcionà el valor correcte:

$$\sum_{n=1}^{\infty}\frac{1}{n^2}=1+\frac{1}{2^2}+\frac{1}{3^2}+\frac{1}{4^2}+\cdots=\frac{\pi^2}{6}.$$

Figura 1. Leonhard Euler (1707-1783).
(Fotografia procedent de www.gap.dcs.st-and.ac.uk/~history/Mathematicians)

La sèrie, dita també harmònica,

$$\sum_{n=1}^{\infty} \frac{1}{n^s}, \quad s \in \mathbb{R}, \quad s > 1,$$

fou avaluada per Euler en tots els enters positius parells, obtenint el resultat

$$\sum_{n=1}^{\infty} \frac{1}{n^{2m}} = (-1)^{m+1} \frac{(2\pi)^{2m}}{2} \frac{B_{2m}}{(2m)!}, \quad m \geq 1,$$

on B_{2m} són els nombres de Bernoulli, que són racionals.

De bon començament, es va veure que la sèrie harmònica proporcionava informació sobre els nombres primers. A partir de la igualtat

$$1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots = \left(1 - \frac{1}{p^s}\right)^{-1},$$

i en tenir en compte que tot nombre natural $n > 1$ descompon de manera única com a producte de nombres primers, Euler obtingué l'expressió de la sèrie harmònica en forma de producte infinit:

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p} \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots\right) = \prod_{p} \left(1 - \frac{1}{p^s}\right)^{-1},$$

on el producte s'estén a tots els primers. Euler utilitzà la descomposició anterior per donar una demostració alternativa del teorema d'Euclides sobre l'existència d'infinits nombres primers. Suposem que el conjunt P dels nombres primers fos finit:

$$\mathrm{P} = \{p_1, \cdots, p_N\}.$$

Es tindria que

$$\prod_{i=1}^{N} \left(1 - \frac{1}{p_i^s}\right)^{-1} = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Però, per a $s = 1$, el terme de l'esquerra representa una quantitat finita, mentre que el de la dreta representa una quantitat infinita. La contradicció posa de manifest que P és un conjunt infinit.

Euler no dubtava a escriure:

$$\log \frac{1}{1-1} = \log \infty = 1 + \frac{1}{2} + \frac{1}{3} + \cdots$$

Tenint present que

$$\log \sum_{n=1}^{\infty} \frac{1}{n} = - \sum_{p} \log \left(1 - \frac{1}{p}\right) = \sum_{p} \{ \frac{1}{p} + \mathcal{O}(p^{-2}) \} = \sum_{p} \frac{1}{p} + \mathcal{O}(1),$$

obté que

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \cdots \quad = \log \log (\infty)$$

Amb això, Euler posà de manifest que la sèrie dels inversos dels primers divergeix, però molt més lentament que la sèrie harmònica.

L'any 1874, Franz Mertens (1840-1927) demostrà que

$$\sum_{p \geq x} \frac{1}{p} = \log (\log x) + \mathcal{O}\left(\frac{1}{\log x}\right),$$

d'acord amb Euler. Atès que

$$\log (\log x) = \int_{1}^{\log x} \frac{du}{u} = \int_{e}^{x} \frac{1}{v} \frac{dv}{\log v},$$

podem sospitar que la densitat dels primers al voltant d'un enter n serà $\frac{1}{\log n}$. Equivalentment,

$$\text{Prob } \{n \text{ primer}\} \sim \frac{1}{\log n} \cdot$$

Una demostració rigorosa del fet anterior constitueix el teorema dels nombres primers, que no s'obtingué fins transcorreguts més de 150 anys.

Extenses taules de primers calculades per Adrien-Marie Legendre (1752-1833) i Carl Friedrich Gauss (1777-1855) conduïren a l'estudi de la funció que compta els primers per sota d'una quantitat donada:

$$\pi(x) := \sum_{p \leq x} 1.$$

Així,

$\pi(10^2) = 25$, $\pi(10^3) = 168$, $\pi(10^4) = 1\,229$,
$\pi(10^5) = 9\,592$, $\pi(10^6) = 74\,498$, $\pi(10^7) = 664\,579$, etc.

Legendre conjecturà que

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

Gauss conjecturà que

$$\pi(x) \sim \mathrm{Li}(x), \quad x \to \infty,$$

on la funció Li denota el logaritme integral:

$$\mathrm{Li}(x) := \int_2^x \frac{du}{\log u}.$$

Com que

$$\mathrm{Li}(x) := \frac{x}{\log x} + \mathcal{O}\left(\frac{x}{\log^2 x}\right),$$

les dues conjectures són equivalents a la conjectura

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

la prova de la qual constituiria el teorema dels nombres primers.

Entre les primeres contribucions a l'estudi dels nombres primers destaquen les de Panufty Lvovich Txebixev (1821-1894). Txebixev demostrà el *postulat de Bertrand*, segons el qual entre un nombre natural $n \geq 1$ i el seu doble sempre hi ha un primer:

$$(n, 2n] \cap P \neq \varnothing.$$

En particular, si p_r denota el primer r-èsim, es té que $p_{r+1} < 2p_r$. Txebixev provà, també, que

$$0.92129 \leq \liminf_{x \to \infty} \frac{\pi(x)}{x} \log x \leq \limsup_{x \to \infty} \frac{\pi(x)}{x} \log x \leq 1.056,$$

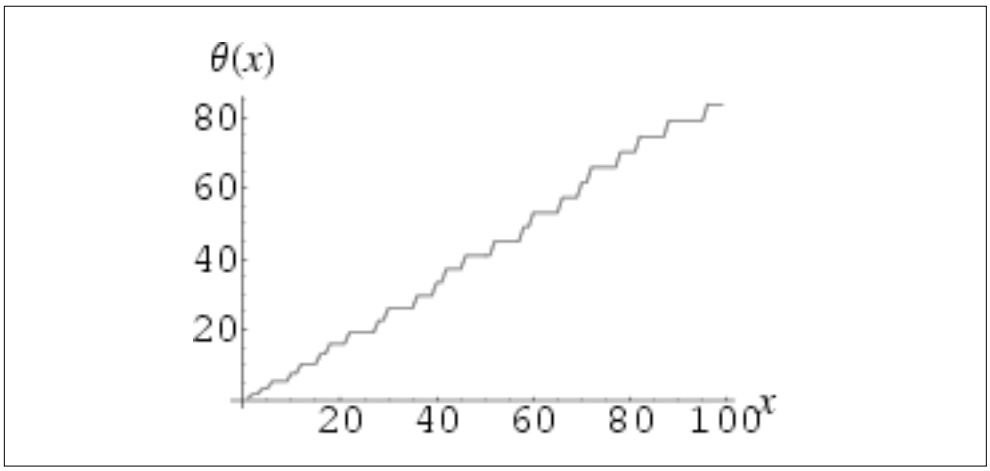


Figura 2. La primera funció de Txebixev. (mathworld.wolfram.com/ChebyshevFunctions.html)

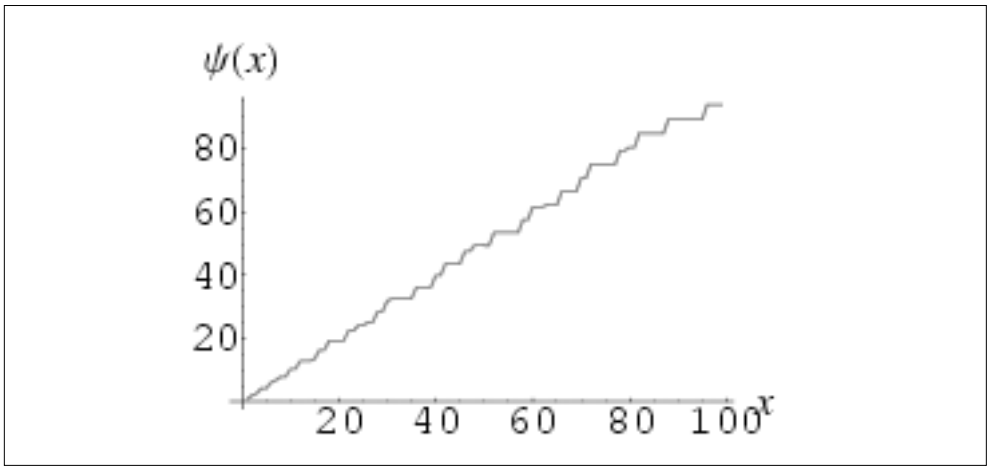


Figura 3. La segona funció de Txebixev ψ(x). (mathworld.wolfram.com/ChebyshevFunctions.html)

i, que si existeix $\lim_{x\to\infty} \frac{\pi(x)}{x} \log x$, aleshores és igual a 1. Per portar a terme aquests estudis, Txebixev introduí dues funcions: la primera funció de Txebixev, ϑ, compta els primers p afectats del pes log (p); la segona funció de Txebixev, ψ, compta les potències dels primers p^m afectades del mateix pes:

$$\vartheta(x) := \sum_{p\le x} \log p, \qquad \psi(x) := \sum_{p^m\le x} \log p,$$

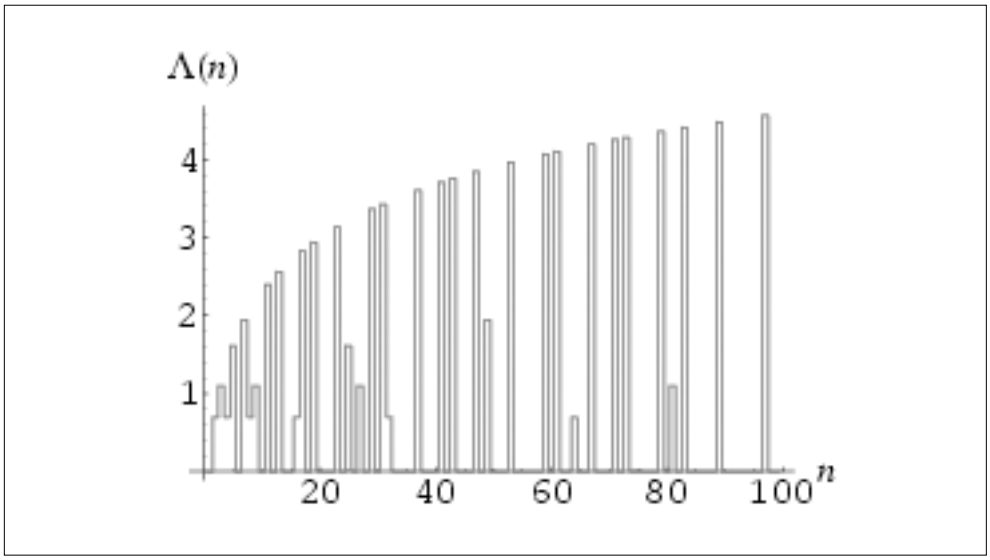


Figura 4. La funció Λ de Mangoldt. (mathworld.wolfram.com/ChebyshevFunctions.html)

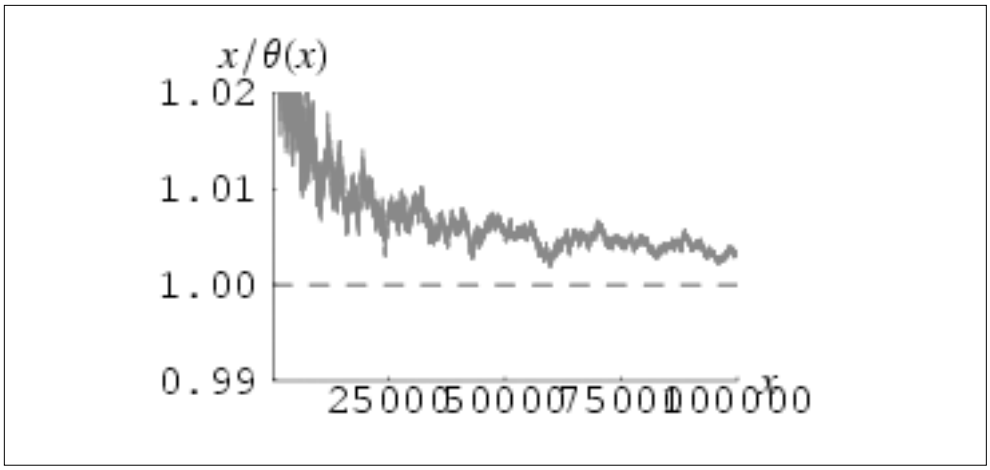


Figura 5. El quocient $x/\vartheta(x)$. (mathworld.wolfram.com/ChebyshevFunctions.html)

Si utilitzem la primera funció de Txebixev, el teorema dels nombres primers admet les formulacions següents:

$$\vartheta(x) \sim x, \quad x \to \infty, \qquad \psi(x) \sim x, \quad x \to \infty.$$

Més endavant, Hans von Mangoldt (1824-1868) introduiria la funció Λ

$$\Lambda(n) := \begin{cases} \log p & \text{si } n = p^m, \\ 0 & \text{altrament,} \end{cases}$$

amb la qual cosa

$$\psi(x) = \sum_{n \leq x} \Lambda(n).$$

2. LA FUNCIÓ ZETA DE RIEMANN

La memòria de Bernhard Riemann (1826-1866) *Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse* [Ri1859], el manuscrit de la qual no arriba a les sis pàgines, representà un canvi de paradigma en l'estudi dels nombres primers. En aquest treball, Riemann posà de manifest que les lleis que regeixen la distribució dels nombres primers depenen en gran part del comportament de la sèrie harmònica, quan aquesta s'estén a una funció de variable complexa.

Donat un nombre complex $z \in \mathbb{C}$, denotarem per a $\Re(z)$, $\Im(z)$ la seva part real i la seva part imaginària, respectivament.

Figura 6. Bernhard Riemann (1826-1866).
(Fotografia procedent de www.gap.dcs.st-and.ac.uk/~history/Mathematicians)

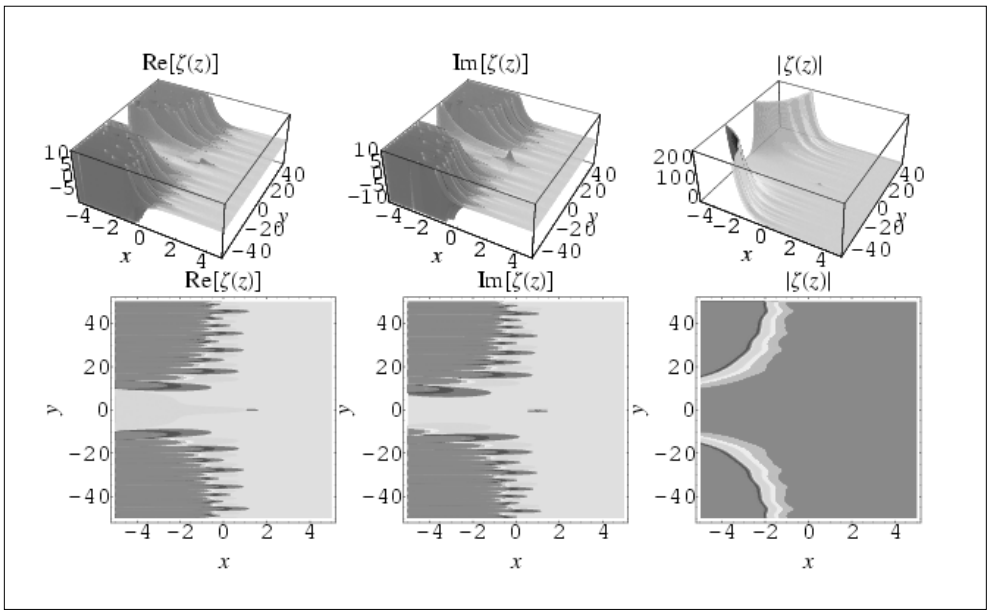


Figura 7. Part real, mòdul de la part imaginària, i mòdul de la funció ζ de Riemann. (mathworld.wolfram.com)

La funció zeta de Riemann es defineix a partir de la sèrie harmònica, però interpretada com a una funció de variable complexa:

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

La funció així definida és analítica per a $\Re(s) > 1$. La representació integral anterior permet la prolongació de la funció zeta en una funció meromorfa de $\mathbb{C}$. La funció estesa, $\zeta(s)$, té un únic pol en $s = 1$, que és simple i de residu igual a 1. En completar la funció zeta amb factors convenients (en $p = \infty$), s'obté la funció

$$\hat{\zeta}(s) := \frac{1}{2} s\,(s-1)\,\pi^{-\frac{s}{2}}\,\Gamma\left(\frac{s}{2}\right)\,\zeta(s).$$

En deformar convenientment el camí d'integració, Riemann obté l'equació funcional:

$$\hat{\zeta}(s) = \hat{\zeta}(1-s), \qquad \text{per a tot } s \in \mathbb{C}.$$

Aquesta equació funcional posa de manifest que el comportament de $\zeta(s)$ per a $\Re(s) < 0$ està determinat pel comportament de $\zeta(s)$ per a $\Re(s) > 1$ i, per tant, pel producte d'Euler corresponent. Els valors de $\zeta(s)$ que aporten més informació són els situats a la banda crítica, definida per:

$$0 \leq \Re(s) \leq 1.$$

L'eix de simetria de la banda crítica, o sigui la recta

$$\Re(s) = \frac{1}{2},$$

rep el nom de *recta crítica*.

L'equació funcional permet avaluar la funció zeta en els enters negatius, en els quals pren valors racionals. Concretament,

$$\zeta(-2m+1) = -\frac{B_{2m}}{2m}, \qquad \zeta(-2m) = 0, \qquad \text{per a tot } m \geq 1.$$

Les fórmules anteriors són equivalents a les fórmules donades per Euler. Fixem-nos que la funció zeta s'anul·la en els enters negatius parells; aquests zeros són els anomenats *zeros trivials*.

Riemann proporciona un esbós de la fórmula del producte

$$\hat{\zeta}(s) = \hat{\zeta}(0) \prod_{\rho} \left(1 - \frac{s}{\rho}\right),$$

en la qual ρ recorre els zeros de $\zeta(s)$ a la banda crítica.

La fórmula del producte per a la funció zeta sense completar s'escriu

$$\hat{\zeta}(s) = \frac{e^{[\log(2\pi) - 1 - \gamma/2]s}}{2(s-1)\Gamma\left(1 + \frac{s}{2}\right)} \prod_{\rho} \left(1 - \frac{s}{\rho}\right) e^{\frac{s}{\rho}},$$

on γ és la constant d'Euler-Mascheroni:

$$\gamma = \lim_{n \to \infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log n\right) = 0.577216...$$

Se sospita que aquesta constant és un nombre irracional, però la seva irracionalitat no ha estat provada.

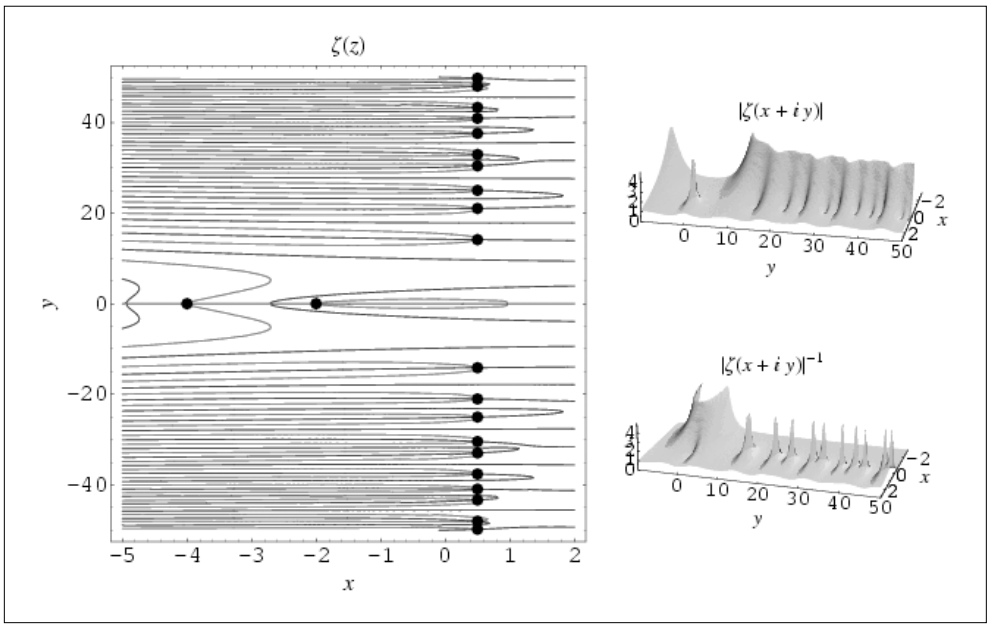


Figura 8. Primers zeros de la funció zeta de Riemann. (mathworld.wolfram.com)

Els primers zeros no trivials de la funció zeta foren calculats a mà pel propi Riemann, però no els inclogué en la memòria esmentada:

$$\rho_1 = \frac{1}{2} \pm i\,14.13..., \qquad \rho_2 = \frac{1}{2} \pm i\,21.02..., \qquad \rho_3 = \frac{1}{2} \pm i\,25.01...,$$

Riemann establí que la funció zeta té infinits zeros a la banda crítica, i donà una estimació del nombre de zeros a la banda crítica d'altura fitada per T:

$$N(T) := \#\{\rho : \zeta(\rho) = 0,\ 0 < \Im(\rho) \le T\} \sim \frac{T}{2\pi} \log\left(\frac{T}{2\pi e}\right).$$

Així, a l'altura T, l'espaiat mitjà entre els zeros de ζ hauria de ser asimptòticament equivalent a $2\pi/\log T$.

Escrivim, d'acord amb Riemann:

$$\xi(t) := \pi^{-\frac{s}{2}}\,\Gamma\left(\frac{s}{2}\right)\zeta(s), \qquad s = \frac{1}{2} + it.$$

Riemann fa el comentari següent sobre els zeros de $\xi(t)$.

> [...] *es ist sehr wahrscheinlich, dass alle Wurzeln reell sind. Hiervon wäre allerdings ein strenger Beweis zu wünschen; ich habe indes die Aufsuchung desselben nach einigen flüchtigen vergeblichen Versuchen vorläufig beiseite gelassen, da er für den nächsten Zweck meiner Untersuchung entbehrlich schien.*
>
> Riemann [Ri1859]

> [...] és molt probable que totes les arrels siguin reals. Per descomptat que seria desitjable tenir-ne una demostració rigorosa; però, després d'alguns intents fugissers realitzats en va, he deixat de banda la recerca d'una prova d'aquesta mena, atès que el fet semblava prescindible per a l'objectiu immediat de la meva investigació.
>
> Riemann [Ri1859]

Es coneix com a *hipòtesi de Riemann* l'afirmació segons la qual tots els zeros no trivials de la funció zeta de Riemann es troben sobre la recta crítica:

$$\text{(HR)} \qquad \zeta(\rho) = 0, \quad 0 \le \Re(\rho) \le 1 \quad \Rightarrow \quad \Re(\rho) = \frac{1}{2}\,.$$

El resultat central obtingut per Riemann en la seva memòria és una fórmula explícita per al càlcul de $\pi(x)$. A partir del desenvolupament

$$\log \zeta(s) = -\sum \log\,(1-p^{-s}) = \sum p^{-s} + \frac{1}{2}\sum p^{-2s} + \frac{1}{3}\sum p^{-3s} + \ldots\,,$$

vàlid per a $\Re(s) > 1$, Riemann escriu

$$\log \zeta(s) = s\int_0^\infty \Pi(x)\, x^{-s-1} dx, \qquad \Re(s) > 1,$$

on

$$\Pi(x) := \pi(x) + \frac{1}{2}\,\pi(\sqrt[2]{x}) + \frac{1}{3}\,\pi(\sqrt[3]{x}) + \cdots$$

Fixem-nos que la suma

$$\pi(x) = \sum_{n \geq 1} \frac{\mu(n)}{n} (x^{1/n}),$$

on μ denota la funció de Möbius, conté únicament un nombre finit de termes no nuls. Per obtenir la fórmula explícita, Riemann utilitza inversió de Fourier, càlcul de residus, i la fórmula d'inversió de Möbius. La fórmula explícita en qüestió és

$$\Pi(x) := \mathrm{Li}(x) - \sum_{\Im(\rho)>0} (\mathrm{Li}(x^{\rho}) + \mathrm{Li}(x^{1-\rho})) + \int_x^{\infty} \frac{dt}{t(t^2-1)\log t} + \log \hat{\zeta}(0),$$

vàlida per a tot $x>1$.

La fórmula explícita de Riemann permet connectar la quantitat de primers inferiors a una quantitat donada amb els zeros de la funció zeta. El mateix Riemann observa que en la fórmula

$$\pi(x) - \sum_{n \geq 1}^{N} \frac{\mu(n)}{n} \mathrm{Li}(x^{1/n}) = \sum_{n \geq 1}^{N} \sum_{\rho} \mathrm{Li}(x^{\rho/n}) + \mathcal{O}(1)$$

Es detecten tres tipus de termes:

1. Termes que no creixen quan x creix: inclosos en $\mathcal{O}(1)$.
2. Termes que creixen quan x creix: $\mathrm{Li}(x^{1/n})$.
3. Termes que creixen en valor absolut quan x creix, però que oscil·len en signe: $\mathrm{Li}(x^{\rho/n})$.

Riemann designa amb el nom de *periòdics* els termes oscil·latoris. I interpreta que són els causants de les fluctuacions observades experimentalment en el nombre de primers dels diferents intervals.

Si escrivim els zeros no trivials de la funció zeta en la forma

$$\rho_n := \frac{1}{2} + i\gamma_n,$$

és clar que la hipòtesi de Riemann és certa si, i només si,

$$\Im(\gamma_n) = 0, \qquad \text{per a tot } n.$$

En derivar, s'obté que

$$\Pi'(x) = \frac{1}{\log x} \left(1 - \frac{1}{x(x^2-1)}\right) - \frac{2}{\sqrt{x}\,\log x} \sum_{\Re(\gamma_n)>0} \frac{\cos(\Re(\gamma_n)\log x)}{x^{\Im(\gamma_n)}}.$$

Cada terme del sumatori representa una contribució oscil·latòria en les fluctuacions de la densitat dels nombres primers. A mesura que x creix, les fluctuacions es fan més petites. En una analogia acústica, el físic Sir Michael Berry i el matemàtic Jon Keating interpreten les freqüències γ_n com a reminiscències de la descomposició d'un so musical en els seus constituents harmònics. Aquests autors proporcionen la interpretació següent de la hipòtesi de Riemann:

> La hipòtesi de Riemann és certa si, i només si, en els primers hi ha música.
>
> M.V. Berry, J.P. Keating [Be-Ke2000]

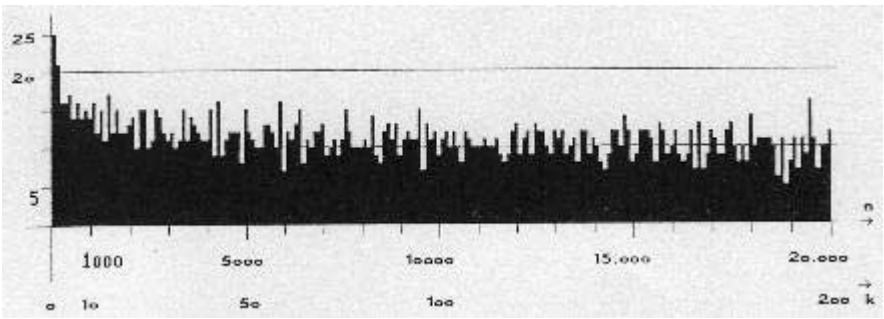


Figura 9. Primers en intervals. (imatges fotografiades)

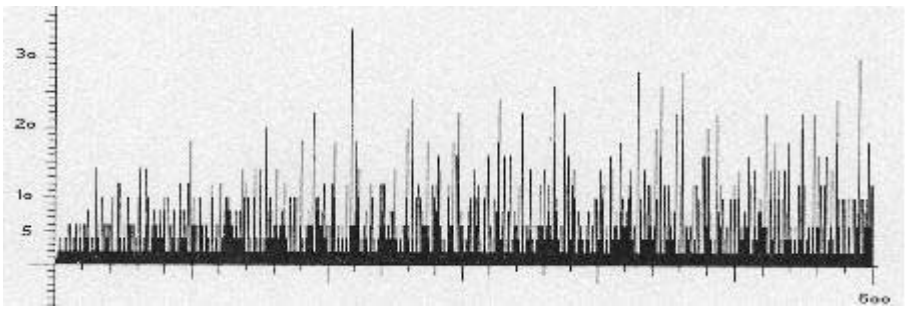


Figura 10. Espaiat entre primers consecutius. (imatges fotografiades)

La memòria de Riemann exercí i continua exercint una influència molt remarcable. Jacques Hadamard (1865-1963) afirmà que havia necessitat tres dècades per comprendre'n el contingut i que havia pogut demostrar totes les afirmacions de Riemann, llevat d'una: la hipòtesi de Riemann.

L'any 1892, Hadamard obtingué la descomposició de $\zeta(s)$ en forma de producte infinit, tal com havia predit Riemann:

$$\hat{\zeta}(s) = \hat{\zeta}(0) \prod_{\rho} \left(1 - \frac{s}{\rho}\right),$$

on ρ descriu els zeros de $\zeta(s)$ a la banda crítica i el producte es realitza en l'ordre creixent de $|\Im(\rho)|$.

Figura 11. Jacques Hadamard (1865-1963). (Fotografia procedent de www.gap.dcs.st-and.ac.uk/~history/Mathematicians)

L'any 1894, von Mangoldt demostrà el resultat, ja enunciat també per Riemann, sobre el nombre de zeros complexos d'altura fitada de la funció zeta:

$$N(T) = \frac{T}{2\pi} \log\left(\frac{T}{2\pi e}\right) + \mathcal{O}(\log T).$$

La fórmula explícita de Riemann fou reescrita i provada per von Mangoldt en la forma

$$\psi(x) = x - \sum_{\rho} \frac{x^{\rho}}{\rho} - \frac{1}{2} \log\left(1 - \frac{1}{x^2}\right) - \log(2), \quad x > 1.$$

L'any 1914, Godfrey Harold Hardy (1877-1947) demostrà que $\xi(t)$ té infinits zeros reals.

L'any 1942, Hardy i John Edensor Littlewood (1885-1977) demostraren que

$$N(T) \geq KT \log(T), \qquad T \gg 0,$$

on $K > 0$ és una constant.

3. EL TEOREMA DELS NOMBRES PRIMERS

L'any 1896, Jacques Hadamard (1865-1963) i Charles Jean de la Vallée Poussin (1866-1962) demostraren de manera independent el teorema dels nombres primers:

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

Figura 12. Charles de la Vallée Poussin (1866-1962). (Fotografia procedent de www.gap.dcs.st-and.ac.uk/~history/Mathematicians)

El seu mètode consistí en l'estimació dels termes oscil·latoris de la fórmula explícita de Riemann. Un punt essencial en les demostracions d'ambdós autors estableix que la funció zeta no s'anul·la en els extrems de la banda crítica:

$$\zeta(1 + it) \neq 0, \quad \text{per a tot } t \in \mathbb{R}.$$

La demostració de la Vallée Poussin proporciona, a més, una fita del terme d'error:

$$|\pi(x) - \mathrm{Li}(x)| = \mathcal{O}(xe^{-\sqrt{c \log(x)}}).$$

Més generalment, es pot demostrar que l'absència de zeros a l'esquerra de la recta $\Re(s)=1$ controla el terme d'error en el teorema dels nombres primers. Sigui $1/2 < \alpha < 1$. Si $\zeta(s) \neq 0$ per a $\alpha < \Re(s) < 1$, aleshores $\vartheta(x) - x = \mathcal{O}(x^{\alpha}(\log x)^2)$. Recíprocament, si $\vartheta(x) - x = \mathcal{O}(x^{\alpha}(\log x)^{\beta})$, aleshores $\zeta(s) \neq 0$ per $\alpha < \Re(s) < 1$. Per tant, la hipòtesi de Riemann és certa si, i només si,

$$|\pi(x) - \mathrm{Li}(x)| = \mathcal{O}(x^{1/2} \log x).$$

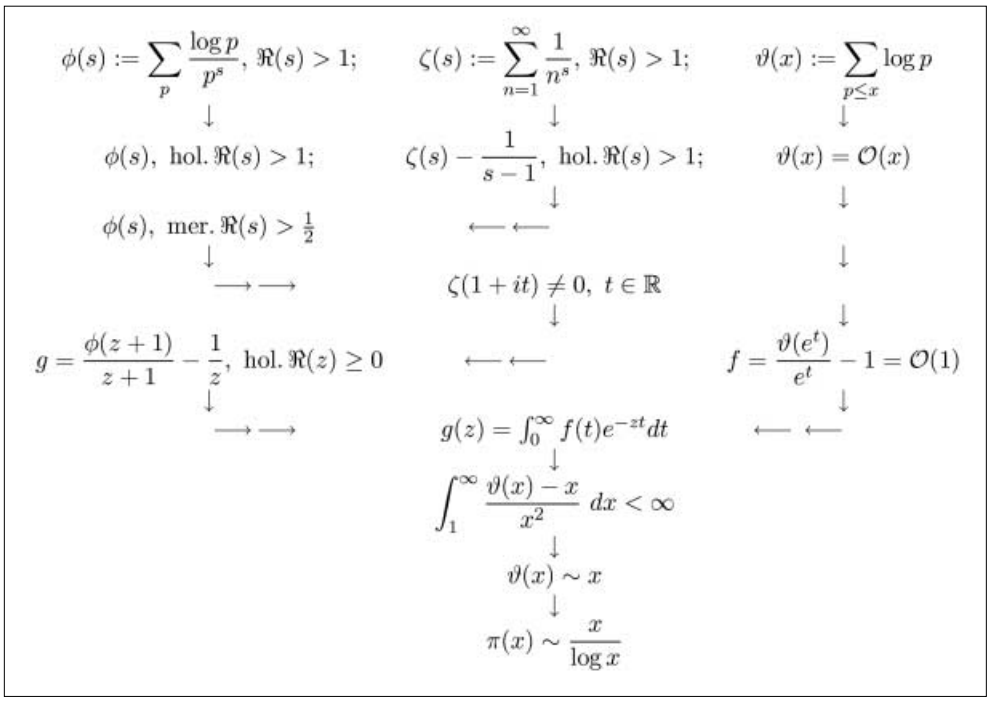


Figura 13. Esbós del teorema dels nombres primers.

Una simplificació apreciable en la demostració del teorema dels nombres primers es produí l'any 1980, gràcies a un teorema de Donald J. Newman. Newman parteix d'una funció $f(t)$ fitada en $(0,\infty)$, integrable localment, amb transformada de Laplace

$$g(z) := \int_0^\infty f(t)e^{-zt}dt,$$

ben definida i analítica per a $\Re(z) > 0$. Demostra que si la funció $g(z)$ admet una prolongació analítica a l'entorn de cada punt de l'eix imaginari, $\Re(z) = 0$, aleshores la integral

$$\int_0^\infty f(t)\,dt,$$

existeix com a integral impròpia, i és igual a $g(0)$. Mitjançant aquest teorema, el teorema dels nombres primers (sense terme d'error) es pot demostrar seguint el recorregut indicat en el mapa de la figura 13 (*pàg. anterior*).

4. LA CONJECTURA DE HILBERT-PÓLYA

L'anomenada *conjectura de Hilbert-Pólya* forma part d'una creença popular. Proposa una interpretació espectral dels zeros no trivials de la funció zeta de Riemann. Més concretament, hauria d'existir un operador lineal Δ els valors propis del qual es relacionessin amb els zeros no trivials de la funció zeta tal com s'indica:

$$\begin{array}{lcl} \text{zero de } \zeta & \leftrightarrow & \text{valor propi de } \Delta \\ \rho = \frac{1}{2} + i\gamma & \leftrightarrow & \lambda = \rho(1-\rho) = \frac{1}{4} + \gamma^2. \end{array}$$

L'operador Δ seria positiu i s'escriuria com a $\Delta = D(1-D)$, essent $i\,(D - \frac{1}{2})$ un operador lineal autoadjunt.

Atès que el nombre de zeros de la funció zeta de Riemann és infinit, de ser certa l'atribució que es fa a Hilbert de la interpretació espectral dels zeros de la funció zeta, Hilbert hauria estat pensant en un espai vectorial complex, de dimensió no finita, dotat d'un producte hermític, complet (en un espai de Hilbert!) i en certs operadors d'espectre discret positiu.

Es diu que Hilbert anomenà *espectre* el conjunt dels valors propis d'un operador hermític per analogia amb les línies, anomenades espectre, produïdes per les freqüències de radiació dels àtoms. Més endavant, la mecànica quàntica confirmaria aquesta

interpretació: les línies espectrals es corresponen amb els valors propis d'operadors hermítics que proporcionen el hamiltonià de sistemes mecànics-quàntics. [Max Born (1882-1970), Werner Heisenberg (1901-1976) i John von Neumann (1903-1957) estudiaren amb Hilbert a Göttingen.]

Tom M. Apostol:
Where are the zeroes of zeta of s?

Where are the zeros of zeta of s?
G.F.B. Riemann has made a good guess,
They're all on the critical line, said he,
And their density's one over 2\pi log t.

This statement of Riemann's has been like trigger
And many good men, with vim and with vigor,
Have attempted to find with mathematical rigor,
What happens to zeta as mod t gets bigger.

The efforts of Landau and Bohr and Cramer,
And Littlewood, Hardy and Titchmarsh are there,
In spite of their efforts and skill and finesse,
(In) locating the zeros there's been no success

In 1914 G.H. Hardy did find,
An infinite number that lay on the line,
His theorem however won't rule out the case,
There might be a zero at some other place.

Let P be the function pi minus li,
The order of P is not know for x high,
If square root of x times log x we could show,
Then Riemann's conjeture would surely be so.

Related to this another enigma,
Concerting the Lindelof function mu (sigma)
Which measures the growth in the critical trip,
On the number of zeros it gives us a grip.

But nobody knows how this function behaves,
Convexity tells us it can have no waves,
Lindelof said that the shape of this graph,
Is constant when sigma is more than one-half.

Oh, where are the zeros of zeta of s?
We must know exactly, we cannot just guess,
In order to strengthen the prime number theorem,
The path of integration must not get too near ´em.

Figura 14. Tom M. Apostol: *On són els zeros de* $\zeta(s)$?

La interpretació espectral dels zeros de la funció zeta ha resultat certa en un context molt diferent: per a certes funcions zeta sorgides de l'estudi de varietats algebraiques definides sobre cossos finits. A tota varietat X, algebraica, projectiva, llisa, definida sobre el cos finit $\mathbb{F}_q$ de $q = p^f$ elements, on p és un primer, se li pot associar una funció zeta de variable complexa,

$$Z(X, q^{-s}),$$

que té cura del nombre de punts de la varietat en el cos $\mathbb{F}_q$ i en totes les seves extensions algebraiques. Aquests punts poden ésser interpretats com a punts fixos de l'anomenat *automorfisme de Frobenius*:

$$X(\mathbb{F}_{q^n}) = X(\overline{\mathbb{F}}_{q^n})^{\mathrm{Frob}_q{}^n}.$$

La funció $Z(X, q^{-s})$ és una funció racional en q^{-s}; satisfà una equació funcional; i satisfà l'anàleg de la hipòtesi de Riemann. Aquests resultats foren obtinguts per Emil Artin (1898-1962), Helmut Hasse (1898-1979), André Weil (1906-1998), Alexander Grothendieck i Pierre Deligne, entre d'altres.

La hipòtesi de Riemann, provada en el context de les varietats algebraiques sobre cossos finits per Deligne, està en la base dels principals resultats obtinguts els darrers anys en l'estudi d'equacions diofantines. Els espais vectorials predits per Hilbert provenen aquí de la cohomologia ℓ-àdica de la varietat, que proporciona espais vectorials en característica zero de dimensió finita. L'operador és l'automorfisme de Frobenius operant en la cohomologia. Deligne aconsegueix la demostració de la hipòtesi de Riemann en característica p a partir d'una fórmula de traces (de Lefschetz) anàloga a la fórmula explícita de Riemann. L'acció d'un grup de monodromia associat a certes famílies de varietats és, també, un ingredient essencial de la prova.

Figura 16. George Pólya (1887-1985).
(Fotografia procedent de www.gap.dcs.st-and.ac.uk/~history/Mathematicians)

Les dificultats vençudes en característica p per provar la hipòtesi de Riemann no són, però, directament traslladables al cas de la característica zero, on els espais vectorials implicats són de dimensió no finita.

Una altra funció zeta, introduïda per Atle Selberg l'any 1957 per a superfícies de Riemann convenientment uniformitzades, satisfà, també, la hipòtesi de Riemann. Per provar aquest resultat, Selberg demostrà prèviament una fórmula de traces per al grup especial lineal $\mathrm{SL}(2,\mathbb{R})$. Els resultats de Selberg han estat objecte de generalitzacions diverses. En aquest cas, l'operador és l'operador de Laplace-Beltrami i els espais vectorials en què opera són de dimensió no finita.

Saunders MacLane:
What Tom Apostol Didn't Know

Andre Weil has bettered old Riemann's fine guess,
By using a fancier zeta of s.
He proves that the zeros are where they should be,
Provided the characteristic is p.

There's a moral to draw from this long tale of woe,
That every young genius among you should know.
If you tackle a problem and seem to get stuck,
Just take it mod p and you'll have better luck.

What fraction of zeros on the line will be found,
When mod t is kept below some given bound?
Does the fraction, whatever, stay bounded below,
As the bound on mod t is permitted to grow?

The efforts of Selberg did finally barnish,
All fears that the fraction might possibly vanish,
It stays bounded below, which is just as it should,
But the bound he determined was not very good.

Norm Levinson managed to show, better yet,
At two-to-one odds it would be a good bet,
If over a zero you happen to trip,
It would be on the line and not just in the strip.

Levinson tried in a classical way,
Weil brought modular means into play,
Atiyah then left and Paul Cohen quit,
So now there's no proof at all that will fit.

But now we must study this matter anew,
Serre points out manifold things it makes true.
A medal might be the reward in this quest,
For Riemann's conjeture is surely the best.

Figura 17. Saunders McLane: *El que Tom Apostol no sabia*.

5. ZEROS I SIMETRIES

Retornem al cas estàndard de la funció zeta de Riemann. L'any 1973, Hugh L. Montgomery publicà un important treball dedicat a l'estudi del comportament de l'espaiat dels zeros de la funció zeta de Riemann a la banda crítica. Montgomery suposa certa la hipòtesi de Riemann. Defineix la funció

$$F(\alpha,T) := \frac{2\pi}{T\log T} \sum_{0<\gamma,\,\gamma'\le T} T^{i\alpha(\gamma-\gamma')}\, w(\gamma-\gamma'),$$

on $w(u) := 4/(4+u^2)$ és un pes i

$$\rho = \tfrac{1}{2} + i\gamma, \quad \rho' = \tfrac{1}{2} + i\gamma', \quad \gamma,\gamma' \in \mathbb{R},$$

denoten zeros no trivials de $\zeta(s)$.

Siguin $\alpha \in \mathbb{R}$, $T \ge 2$. Montgomery demostra que

1. $F(\alpha,T) = F(-\alpha,T) \ge -\varepsilon$ per a tot α, sempre que $T > T_0(\varepsilon)$.
2. $F(\alpha,T) = [1+o(1)]T^{-2\alpha}\log T + \alpha + o(1)$, quan $T \to \infty$, si $\alpha \in [0,1)$, i uniformement per a $0 \le \alpha \le 1-\varepsilon$.

D'ací es dedueix que el nombre de zeros simples $\rho = \frac{1}{2} + i\gamma$ tals que $\gamma \in \mathbb{R}, 0 < \gamma \leq T$, és més gran o igual que

$$\left(\frac{2}{3} + o(1)\right) \frac{T}{2\pi} \log T.$$

I, alhora, que existeix una constant computable $\lambda < 1$ tal que

$$\liminf_n = (\gamma_{n+1} - \gamma_n) \log \frac{\gamma_n}{2\pi} \leq \lambda,$$

on $0 < \gamma_1 \leq \gamma_2 \leq \cdots$ són les parts imaginàries dels zeros en el semiplà superior complex.

Els resultats anteriors conduïren Montgomery a formular l'anomenada *conjectura de correlació de parells*, segons la qual

$$\lim_{T\to\infty} \frac{1}{N(T)} \sum_{\substack{0<\gamma,\,\gamma'\leq T \\ \frac{2\pi\alpha}{\log T} \leq \gamma-\gamma' \leq \frac{2\pi\beta}{\log T}}} 1 = \int_0^\beta \left(1 - \left(\frac{\sin\pi x}{\pi x}\right)^2\right) dx,$$

per a $0 < \alpha < \beta < \infty$. En normalitzar l'espaiat entre dos zeros consecutius, definint

$$\delta_n := (\gamma_{n+1} - \gamma_n) \frac{\log(\gamma_n/(2\pi))}{2\pi}; \qquad \sum_{n=N+1}^{N+M} \delta_n = M + \mathcal{O}(\log NM),$$

la conjectura anterior es formula tot dient que

$$\frac{1}{N} \#\left\{(n,k) : 1 \leq n \leq N,\ k \geq 0,\ \sum_n^{n+k} \delta_j \in [\alpha,\beta]\right\} \sim$$

$$\int_\alpha^\beta \left(1 - \left(\frac{\sin\pi x}{\pi x}\right)^2\right) dx,$$

per a $N \to \infty$.

Estem, per tant, davant d'un possible teorema central del límit que governaria l'espaiat normalitzat entre zeros consecutius de la funció $\xi(t)$; la densitat de probabilitat essent donada per

$$1 - \left(\frac{\sin\pi x}{\pi x}\right)^2.$$

D'entrada, la conjectura anterior és molt sorprenent, ja que implicaria en particular que els espaiats petits entre els zeros de la funció zeta són molt poc freqüents. En canvi, el nombre de primers en intervals petits ha estat observat experimentalment i es conjectura que es distribuirien segons una llei de Poisson.

Es conta que l'any 1973, en el decurs d'un te preparat a Princeton després d'una exposició de Montgomery, Freeman Dyson, que no havia assistit a la conferència de Montgomery, li endevinà exactament la densitat de probabilitat que regiria la distribució dels espaiats normalitzats dels zeros de la funció zeta de Riemann. El motiu era que el mateix tipus de distribució es trobava en l'estudi dels conjunts gaussians unitaris (GUE), en la teoria de matrius aleatòries, que ell havia estudiat.

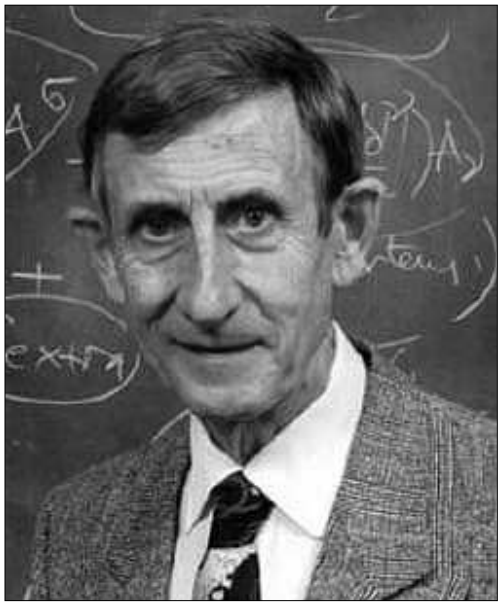


Figura 18. Freeman Dyson.
(Fotografia procedent de www.gap.dcs.st-and.ac.uk/~history/Mathematicians)

L'estudi dels espaiats dels valors propis de matrius aleatòries es remunta a Eugene Wigner (1902-1995), qui s'ocupà d'aquestes qüestions en relació amb l'estudi estadístic dels nivells d'energia de grans nuclis. Les lleis estadístiques que governen aquests espaiats són sensibles al tipus de simetria i són diferents segons que es tracti de matrius ortogonals (GOE), unitàries (GUE), o simplèctiques (GSE).

Considerem el grup unitari en dimensió N

$$U(N) := \{A \in M_N(\mathbb{C}) : AA^* = 1_N, \text{ on } A^* := (\overline{a_{j,i}})\}.$$

Figura 19. Hugh L. Montgomery.

Tota matriu $A \in U(N)$ és diagonalitzable i el seu espectre consta de valors propis de mòdul 1. Escrivim

$$\text{Spec}\,(A) = \{e^{i\theta_1}, \ldots, e^{i\theta_N}\}, \quad 0 \le \theta_1 \le \cdots \le \theta_N < 2\pi.$$

L'observació de Dyson fa referència a que els espaiats normalitzats $\hat{\gamma} - \hat{\gamma}'$, on

$$\zeta(\rho) = 0, \quad \rho = \frac{1}{2} + i\,\gamma, \quad \hat{\gamma} := \frac{\gamma \log \gamma}{2\pi},$$

es comportarien asimptòticament com els espaiats convenientment normalitzats $\hat{\theta} - \hat{\theta}'$ dels valors propis d'una matriu unitària aleatòria:

$$\hat{\theta} := \frac{N}{2\pi}\theta.$$

Més generalment, donades una successió $\mathcal{T}$ de N-ples

$$T_N = \{t_1 < t_2 < \cdots < t_N\},$$

en la qual suposarem que l'espaiat mitjà és igual a 1, i una família de funcions test

$$f \in \mathcal{S},$$

podem considerar diferents funcions de densitat:

(1) $W(x)$, la densitat de nivell n, cal que satisfaci

$$\lim_{N\to\infty} \sum_{\substack{(i_1,\ldots,i_n)\\ i_j\le N,\ i_j\neq i_k}} f(t_{i_1},\ldots,t_{i_n}) = \int_{\mathbb{R}^n} f(x)\,W(x)\,dx.$$

(2) $\mu_k(x)$, la densitat de l'espaiat k-èsim, cal que satisfaci

$$\lim_{N\to\infty} \sum_{i\le N-1} f(t_{i+k}-t_i) = \int_0^\infty f(x)\,\mu_k(x)\,dx.$$

S'introdueix, també, la funció de correlació de parells:

$$R_2(A)\,[a,b] := \frac{\#\left\{j\neq k : \frac{N}{2\pi}\,(\theta_j-\theta_k)\in[a,b]\right\}}{N}.$$

En la teoria de matrius aleatòries, es demostra que existeixen mesures μ_k(GUE), R_2(GUE) en $U(N)$ (i similars en els altres grups clàssics) tals que

$$\lim_{N\to\infty} \int_{U(N)} D(\mu_k(A)\,,\,\mu_k(\mathrm{GUE}))\,dA = 0,$$

$$\lim_{N\to\infty} \int_{U(N)} |\,R_2(A)\,[a,b] - R_2(\mathrm{GUE})\,[a,b]\,|\,dA = 0,$$

essent

$$R_2(\mathrm{GUE})[a,b] := \int_a^b \left(1-\left(\frac{\sin\pi x}{\pi x}\right)^2\right) dx$$

i denotant per a $D(\nu_1,\nu_2)$ la discrepància dita de Kolmogorov-Smirnov.

La conjectura de Montgomery i l'observació de Dyson han originat un treball numèric intents per posar a prova la seva validesa. Andrew Odlyzko fou un dels primers

Figura 20. Andrew Odlyzko. (Font: Internet)

a interessar-se per aquest tipus de qüestions. Odlyzko dirigí una carta a Pólya preguntant-li per una possible interpretació espectral dels zeros de la funció zeta de Riemann:

> Benvolgut professor Pólya,
>
> He sentit dir en diverses ocasions que vosté i Hilbert havien conjecturat, de manera independent, que els zeros de la funció zeta de Riemann es corresponen amb els valors propis d'un operador hermític autoadjunt. Podria proporcionar-me'n alguna referència? Podria dir-me, també, quan es va formular aquesta conjectura, i quin fou el motiu que en el seu moment l'induí a la seva consideració? [...]
>
> Andrew Odlyzko, 8 de desembre del 1981

La resposta de Pólya a Odlyzko fou la següent:

> Benvolgut professor Odlyzko,
>
> Moltes gràcies per la seva carta del 8 de desembre. Únicament li puc dir el que em passà a mi. Vaig fer una estada a Göttingen que finalitzà a començaments del 1914. Vaig intentar aprendre teoria analítica de nombres amb Landau. Un dia em va preguntar: «Vosté, que té coneixements de física, coneix alguna raó física per la qual la hipòtesi de Riemann pogués ser certa?» Aquest seria el cas, vaig respondre, si els zeros no trivials de la funció ξ *estiguessin connectats amb un proble-*

> *ma físic* de manera que la hipòtesi de Riemann fos equivalent al fet que tots els valors propis del problema físic fossin reals. Mai vaig publicar aquesta observació, però d'alguna manera esdevingué coneguda i encara es recorda.
>
> George Pólya, 3 de gener de 1982

En un treball numèric impressionant, Odlyzko estudià l'espaiat consecutiu de 70 milions de zeros de la funció zeta obtinguts a partir del zero 10^{22}, i la correlació de parells entre 8 milions de zeros a partir del zero 10^{20}. Els resultats experimentals es comparen amb els resultats teòrics en les figures 22 i 23. Permeten enunciar la:

Llei de Montgomery-Odlyzko (deduïda de l'observació empírica). La distribució de l'espaiat (convenientment normalitzat) entre els zeros no trivials consecutius de la funció zeta de Riemann és idèntica, en sentit estadístic, a la distribució de l'espaiat dels valors propis d'un conjunt unitari gaussià.

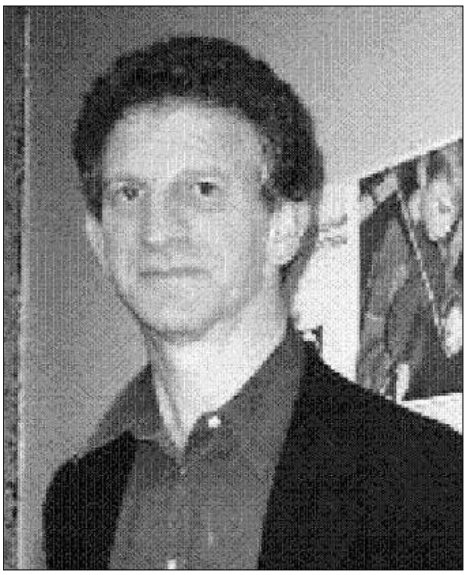


Figura 21. Peter Sarnak. (Font: Internet)

Nicolas Katz i Peter Sarnak han estès l'estudi dels espaiats dels zeros a altres funcions zeta, i a famílies d'aquestes funcions, posant de manifest que s'obtenen simetries diferents segons les famílies considerades. Tots aquests experiments numèrics han despertat un alt interès, tant en l'àmbit de les matemàtiques com en l'àmbit de la física, en tant que corroboren el punt de vista de Hilbert i Pólya vers una interpretació espectral

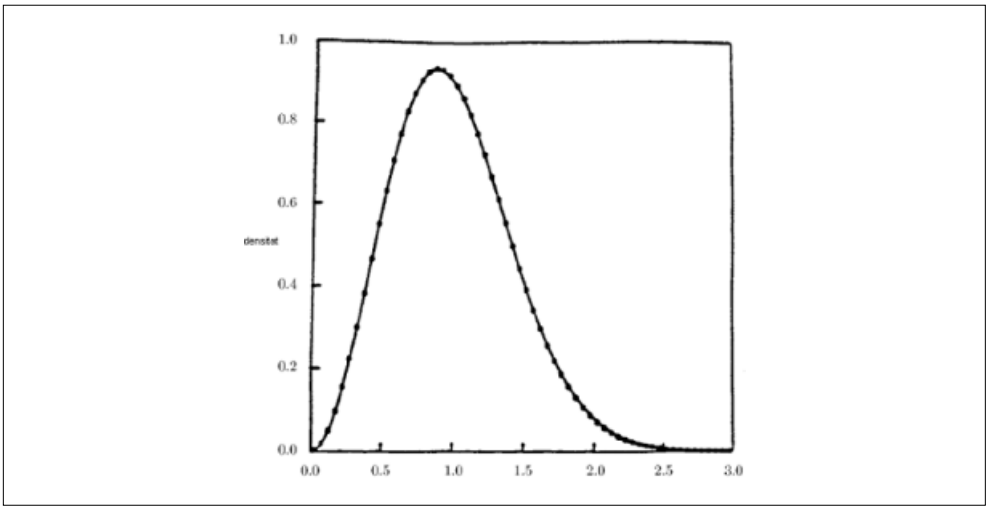


Figura 22. Densitat de $\delta_n = \hat{\gamma}_{n+1} - \hat{\gamma}_n$, per a $10^{20}+1 \leq n \leq 10^{20}+70\times 10^6$, vers μ_1(GUE).

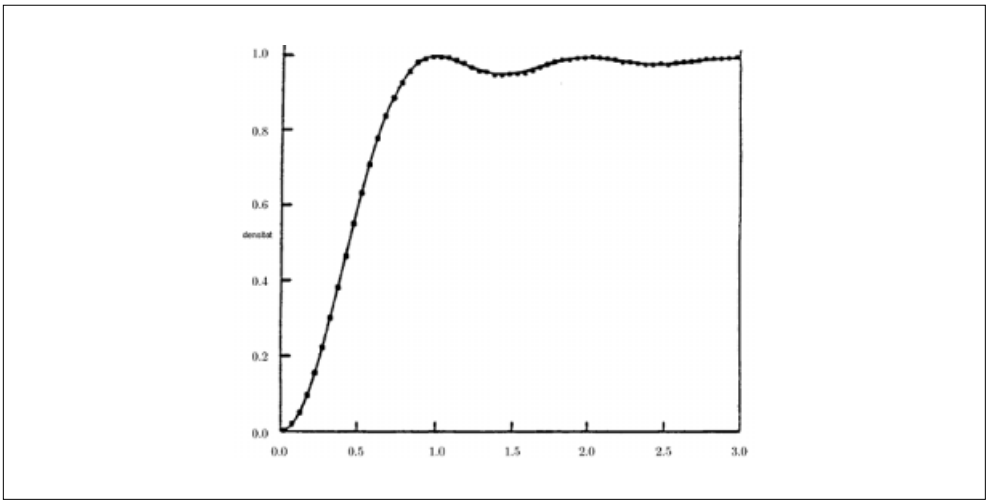


Figura 23. Correlació de parells $\hat{\gamma}_j - \hat{\gamma}_k$, $10^{20}+1 \leq j,\ k \leq 10^{20}+8\times 10^6$, vers R_2 (GUE).

dels zeros de la funció zeta de Riemann (i d'altres de similars), i permeten deduir propietats de simetria de l'operador cercat.

A la taula 1 donem compte d'algunes funcions zeta, de l'àmbit al qual pertanyen i d'algunes de les seves propietats més notables. Hi incloem els noms d'alguns dels seus investigadors.

Taula 1: **Una selecció de funcions zeta i funcions *L***

Funcions zeta i funcions *L* aritmètiques	
Riemann	$\zeta(s)$: enters; nombres primers; equació funcional; fórmula explícita
Dirichlet	$L(\chi, s)$: cossos ciclotòmics; progressions aritmètiques; equació funcional; fórmula explícita
Dedekind	$\zeta(K, s)$: cossos de nombres; ideals primers; equació funcional; fórmula explícita
Artin	$L(\rho, s)$: representacions de Galois; distribució d'ideals primers; fórmula explícita
p-àdiques	Leopoldt; Iwasawa; cossos de nombres abelians i no abelians
Funcions zeta i funcions *L* aritmètico-geomètriques	
Hasse-Weil	$Z(X, q^{-s})$: varietats, cossos finits; fórmula de Lefschetz; equació funcional; **HR**
Serre	$\zeta(X, s)$, $L(X, s)$ Hasse-Weil; teoria de Hodge; esquemes aritmètics
Grothendieck	$L(M, s)$ motius
p-àdiques	varietats algebraiques sobre cossos de nombres; motius
Funcions *L* automorfes	
Hecke	$L(\psi, s)$: grups de classes d'ideles; equació funcional
Jacquet	$L(\pi, s)$: Hecke; Shimura; grups de Lie; equació funcional (cas estàndard); traces
p-àdiques	representacions automorfes de grups de Lie *p*-àdics
Funcions zeta i funcions *L* espectrals	
Selberg	superfícies de Riemann; grups kleinians; geodèsiques tancades; fórmula de traces; **HR**
Connes	grups de Lie adèlics; C^*-sistemes dinàmics; fórmula de traces (?); HR (?)
Stark	Terras; grafs finits; circuits tancats; **HR** per a grafs de Ramanujan
Funcions zeta i funcions *L* algebraiques	
Epstein	formes quadràtiques; no satisfan HR
Godement	Hey; Tamagawa; àlgebres simples
Lustig	automorfismes de grups lliures;
Funcions zeta i funcions *L* dinàmiques	
Nielsen	espais topològics; sistemes dinàmics; òrbites periòdiques
Ruelle	Smale; difeomorfismes de varietats compactes; fluxos; òrbites periòdiques
Deninger	foliacions; fórmula de Lefschetz
Carlitz	autòmates
Gutzwiller	Hawking; sistemes quàntics; cosmologia quàntica; teoria quàntica de camps

El treball d'Odlyzko tingué una repercussió en la conjectura de Mertens. Si denotem per

$$M(x) := \sum_{n \le x} \mu(n),$$

Mertens havia conjecturat que

$$|M(x)| < x^{1/2}, \qquad \text{per a tot } x > 1.$$

El treball numèric portat a terme per Odlyzko en l'estudi dels zeros de la funció zeta li permeté concloure que la conjectura de Mertens és falsa, atès que

$$\limsup_{x \to \infty} M(x) x^{-1/2} > 1.06.$$

D'altra banda, és sabut que la hipòtesi de Riemann és equivalent a

$$|M(x)| = \mathcal{O}(x^{1/2+\varepsilon}), \qquad \text{per a tot } \varepsilon > 0.$$

6. EVIDÈNCIA DE LA HIPÒTESI DE RIEMANN

L'any 1932, Siegel portà a terme un acurat estudi dels manuscrits de Riemann. Siegel opinà que la llegenda segons la qual Riemann hauria obtingut els seus resultats mitjançant idees d'una gran generalitat i sense utilitzar les eines formals de l'anàlisi, no era sostenible. Segons Siegel, la força de la tècnica analítica de Riemann es posava de manifest especialment en la seva derivació i transformació de la sèrie semiconvergent per a $\zeta(s)$.

Els darrers anys han conegut un avenç impressionant en els algoritmes de verificació de la hipòtesi de Riemann. Les eines més importants que s'han anat succeint, i els seus temps d'execució són:

1. La fórmula de sumació d'Euler-MacLaurin: $\mathcal{O}(n^{2+\varepsilon})$,
2. La fórmula integral de Riemann-Siegel: $\mathcal{O}(n^{3/2+\varepsilon})$,
3. El mètode de Odlyzko-Schönhage: $\mathcal{O}(n^{1+\varepsilon})$.

En els càlculs, és essencial la fórmula

$$N(T) - 1 = \frac{1}{2\pi i} \int_{\partial \mathcal{R}} \frac{\zeta'}{\zeta}(s)\, ds,$$

que permet obtenir el nombre de zeros en un interval mitjançant una integració. Els avenços en el disseny d'algoritmes han permès no solament la verificació de la hipòtesi de Riemann en intervals molt amplis sinó, a més, obtenir els seus zeros amb una gran precisió.

Primers zeros de la funció zeta de Riemann:

n	ρn
1	$\frac{1}{2} + i$ 14.13472514173469379045...
2	$\frac{1}{2} + i$ 21.02203963877155499262...
3	$\frac{1}{2} + i$ 25.01085758014568876321...
4	$\frac{1}{2} + i$ 30.42487612585951321031...
5	$\frac{1}{2} + i$ 32.93506158773918969066...
⋮	⋮

Alguns zeros de la funció zeta de Riemann:

n	ρn
10^{20}	$\frac{1}{2} + i$ 15 202 440 115 920 747 268.62902990...
$10^{22} - 1$	$+ i$ 1 370 919 909 931 995 308 226.490240...
10^{22}	$\frac{1}{2} + i$ 1 370 919 909 931 995 308 226.627511...
$10^{22} + 1$	$\frac{1}{2} + i$ 1 370 919 909 931 995 308 226.680160...
⋮	⋮

Investigador	**any**	**nombre de zeros**
Gram	1903	15
Backlund	1914	79
Hutchinson	1925	138
Tichmarsh *et al*.	1935	1 041
Turing	1952	1 054
Lehmer	1956	25 000
Meller	1958	35 337
Lehman	1966	250 000
Rosser *et al.*	1968	3 500 000
Brendt	1979	81 000 000
te Riele *et al*.	1986	1 500 000 000
van de Lune	2001	10 000 000 000
Wendeniwski	2004	900 000 000 000
Gourdon *et al*.	2004	10 000 000 000 000

La xarxa ZetaGrid (http://www.zetagrid.net), dedicada a la verificació de la hipòtesi de Riemann, aplega més de 11.000 estacions de treball i 4.000 usuaris. Es calculen més de 10^9 zeros de la funció zeta per dia (tothom hi és convidat!). La hipòtesi de Riemann ha estat comprovada per als 10 primers bilions de zeros.

En el context del caos quàntic aritmètic, s'han proposat alguns models físics per a l'estudi de $\zeta(s)$. Des del punt de vista de la física, fórmules explícites de Riemann, Weil, Selberg, Gutwiller han estat interpretades com a lligams entre propietats clàssiques i propietats quàntiques de sistemes mecànics-quàntics. Aquestes interpretacions tenen cura de: funcions de partició en mecànica estadística, oscil·ladors bosònics amb energies log *p*, plasmes de *quarks-gluons*, condensats de Bose-Einstein, etc.

En la interpretació espectral de la funció zeta, hi ha qui creu que l'operador de Riemann podria ser la quantització d'un hamiltonià clàssic. Les òrbites periòdiques clàssiques tindrien longitud log *p*. La dinàmica de Riemann seria caòtica i amb *mass gap*. No seria invariant per inversió de la fletxa del temps. Tindria instantons de períodes múltiples de $i\pi$. En resum, i segons Marcus du Sautoy (2003):

> És evident que els zeros de Riemann són vibracions, però no sabem què és el que vibra.

REFERÈNCIES

[Be-Ke2000] BERRY, M.V.; KEATING, J.P.: «The Riemann zeros and eigenvalue

asymptotics», *SIAM Rev.* 41 (1999), núm. 2, p. 236-266 (electronic). MR1684543 (2000f:11107).

[Co1999] CONNES, Alain: «Trace formula in noncommutative geometry and the zeros of the Riemann zeta function», *Selecta Math.* (N.S.) 5 (1999), núm. 1, p. 29-106. MR1694895 (2000i:11133).

[De1995] DENINGER, Christopher: «Higher order operations in Deligne cohomology», *Invent. Math.* 120 (1995), núm. 2, p. 289-315. MR1329043 (96f:11085).

[De1998] DENINGER, Christopher: «Some analogies between number theory and dynamical systems on foliated spaces», *Proceedings of the International Congress of Mathematicians*, vol. I (Berlin, 1998). Doc. Math. 1998, Extra vol. I, p. 163-186 (electrònic). MR1648030 (99g:11084).

[De-Si2001] DENINGER, Christopher; SINGHOF, Wilhelm: «A note on dynamical trace formulas», *Dynamical, spectral, and arithmetic zeta functions* (San Antonio, TX, 1999), p. 41-55, Contemp. Math., 290, Amer. Math. Soc., Providence, RI, 2001. MR1868467 (2002k:58054).

[duSa2003] DU SAUTOY, Marcus: *The music of the primes. Searching to solve the greatest mystery in mathematics*. Harper-Collins Publishers, Nova York, 2003. X+335 p. ISBN: 0-06-621070-4. MR2060134.

[Eu1737] EULER, L.: «Variae observationes circa series infinitas», *Commentarii academiae scientarium Petropolitanae* 9 (1737), p. 160-188.

[Eu1748] EULER, L.: *Introduction in analysin infinitorum*, Lausana. *Introducción al análisis de los infinitos*. Traducción de J.L. Arantegui; anotado por A.J. Durán. SAEM «Thales» 2000.

[Ka2001] KATZ, Nicholas M.: «*L*-functions and monodromy: four lectures on Weil II», *Adv. Math.* 160 (2001), núm. 1, p. 81-132. MR1831948 (2002c:11066).

[Ka2002] KATZ, Nicholas M.: «Twisted *L*-functions and monodromy», *Annals of Mathematics Studies*, 150. Princeton University Press, Princeton, NJ, 2002. VIII+249 p. ISBN: 0-691-09150-1; 0-691-09151-X. MR1875130 (2003i:11087).

[Ka-Sa1999] KATZ, Nicholas M.; SARNAK, Peter: «Zeroes of zeta functions and symmetry», *Bull. Amer. Math. Soc.* (N.S.) 36 (1999), núm. 1, p. 1-26. MR1640151 (2000f:11114).

[Ka-Sa2000] KATZ, Nicholas M.; SARNAK, Peter: *Random matrices, Frobenius eigenvalues, and monodromy*, American Mathematical Society Colloquium Publications, 45. American Mathematical Society, Providence, RI, 1999. XII+419 p. ISBN: 0-8218-1017-0. MR1659828 (2000b:11070).

[Ke-Sna2003] KEATING, J.P.; SNAITH, N.C.: «Random matrices and *L*-functions. Random matrix theory», *J. Phys. A* 36 (2003), núm. 12, p. 2.859-2.881. MR1986396 (2004d:11090).

[La-Od1987] LAGARIAS, J.C.; ODLYZKO, A.M.: «Computing $\pi(x)$: an analytic method», *J. Algorithms* 8 (1987), núm. 2, p. 173-191. MR0890871 (88k:11095).

[La-Fr2001] LAPIDUS, Michel L.; VAN FRANKENHUYSEN, Machiel: «Dynamical, spectral,

and arithmetic zeta functions» (San Antonio, TX, 1999), *Contemp. Math.*, 290, Amer. Math. Soc., Providence, RI, 2001. MR1868472 (2003b:37011).

[Mo1972] MONTGOMERY, Hugh: «Corrélations dans l'ensemble des zéros de la fonction zéta», *Séminaire de Théorie des Nombres, 1971-1972* (Univ. Bordeaux I, Talence), Exp. núm. 19, 9 p. Lab. Théorie des Nombres, Centre Nat. Recherche Sci., Talence, 1972. MR0389782 (52:10613).

[Mo1973] MONTGOMERY, H.L.: «The pair correlation of zeros of the zeta function», *Analytic number theory* (Proc. Sympos. Pure Math., Vol. XXIV, St. Louis Univ., St. Louis, Mo., 1972), p. 181-193. Amer. Math. Soc., Providence, R.I., 1973. MR0337821 (49:2590).

[Ne1998] NEWMAN, Donald J.: *Analytic number theory*. Graduate Texts in Mathematics, 177. Springer-Verlag, New York, 1998. VIII+76 p. ISBN: 0-387-98308-2. MR1488421 (98m:11001).

[Od1987] ODLYZKO, A.M.: «On the distribution of spacings between zeros of the zeta function», *Math. Comp.* 48 (1987), núm. 177, p. 273-308. MR0866115 (88d:11082).

[Od1999] ODLYZKO, A.M.: «The 10^{22}-nd zero of the Riemann zeta function», *Dynamical, spectral, and arithmetic zeta functions* (San Antonio, TX, 1999), p. 139-144, Contemp. Math., 290, Amer. Math. Soc., Providence, RI, 2001. MR1868473 (2003h:11109).

[Od-Sch1988] ODLYZKO, A.M.; SCHÖUMLÄNHAGE, A.: «Fast algorithms for multiple evaluations of the Riemann zeta function», *Trans. Amer. Math. Soc.* 309 (1988), núm. 2, p. 797-809. MR0961614 (89j:11083).

[Ri1859] RIEMANN, G.F.B.: «Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse», *Monatsberichte der Berliner Akademie* (1859), p. 671-680.

[Se1956] SELBERG, A.: «Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series», *J. Indian Math. Soc.* (N.S.) 20 (1956), p. 47-87. MR0088511 (19,531g).

[We1979-I] WEIL, André: *Œuvres scientifiques. Collected papers*, vol. I (1926-1951). Springer-Verlag, Nova York - Heidelberg, 1979. XVIII+574 p. ISBN: 0-387-90330-5 MR0537937 (80k:01067a).

[We1979-II] WEIL, André: *Œuvres scientifiques. Collected papers*, vol. II (1951-1964). Springer-Verlag, Nova York - Heidelberg, 1979. XII+561 p. ISBN: 0-387-90330-5 MR0537935 (80k:01067b).

[We1979-III] WEIL, André: *Œuvres scientifiques. Collected papers*, vol. III (1964-1978). Springer-Verlag, Nova York - Heidelberg, 1979. XII+465 p. ISBN: 0-387-90330-5. MR0537936 (80k:01067c).

[Wei] WEISSTEIN, E.W.: «Chebyshev Functions; Mangoldt Functions; Riemann Zeta Function Zeros», *MathWorld* —A Wolfram Web Resource. http://mathworld.wolfram.com.

La conjetura de Hodge

José I. Burgos Gil

INTRODUCCIÓN

La conjetura de Hodge fue propuesta como problema en el Congreso Internacional de Matemáticas celebrado en Cambridge (Massachusetts, USA) en 1950 [8]. La conjetura de Hodge es un problema geométrico y predice una relación muy estrecha entre la geometría, el álgebra, la topología y el análisis.

Este problema es, sin duda, el más abstracto y el más alejado de las matemáticas cotidianas de los siete problemas del milenio. Éste es un hecho sorprendente si tenemos en cuenta que la geometría, junto con la aritmética, es una de las ramas mas antiguas de la matemática. El capítulo sobre la conjetura de Birch y Swinnerton-Dyer proporciona un ejemplo de un problema moderno en aritmética. En él, el lector puede comprobar que, aunque las técnicas usadas en aritmética han evolucionado considerablemente, los problemas de los que se ocupa, su núcleo, es el mismo que hace 2.500 años: el estudio de los números naturales y de sus propiedades.

Por el contrario, lo que hoy en día consideramos geometría está muy alejado de la idea clásica de geometría. Los conceptos con los que trabaja la geometría moderna son abstracciones construidas sobre abstracciones durante más de 2.000 años. Por ello, para poder entender en qué consiste esta conjetura es necesario hacer un recorrido por la historia de la geometría.

1. LA GEOMETRÍA CLÁSICA

El origen de la geometría se remonta a las antiguas civilizaciones de Mesopotamia, Egipto, India y China. En un principio, la geometría es una ciencia práctica. Los problemas de los que se ocupa son el cálculo de longitudes, de áreas y de volúmenes. Estos conocimientos se pueden aplicar a la predicción de la cantidad de material necesario en una construcción, al cálculo de la capacidad de una vasija o a la extensión de un campo cultivado.

Veamos, a modo de ejemplo, cómo se trata el problema del área de un círculo en estas civilizaciones.

En el *Papiro Matemático de Rhind* [3], fechado hace aproximadamente 3.650 años, encontramos el siguiente problema: «Ejemplo de un campo redondo de diámetro 9.

¿Cuál es el área? Extrae 1/9 del diámetro. El resto es 8. Multiplica 8 por 8. El resultado es 64. Por tanto, el área es 64». Este procedimiento nos da un valor aproximado de π de $256/81 = 3{,}160...$

Los *Sulvasutras* son una colección de versos de la época védica de la India fechados hace aproximadamente 2.600 años, aunque reflejan una tradición oral muy anterior. En ellos encontramos el siguiente método para calcular el área del círculo: «Si deseas convertir un círculo en un cuadrado, divide el diámetro en 8 partes, y una de estas partes en 29 partes; de estas 29 retira 28 y la sexta parte (de la restante) menos la octava parte (de la sexta parte)». Estos cálculos equivalen a tomar un valor de π igual a 3,0883.

Los babilonios y los chinos trataron el problema del área del círculo de forma distinta. En ambas civilizaciones encontramos la fórmula $A = (C/2)(d/2)$, donde A es el área del círculo; C, la longitud de la circunferencia, y d, el diámetro. Aunque este resultado no resuelve el problema de encontrar la constante de proporcionalidad π, nos proporciona una íntima relación entre dos problemas en principio distintos: el área del círculo y la longitud de la circunferencia.

Por tanto, la geometría aparece junto a las primeras grandes civilizaciones humanas. Sin embargo, la primera edad de oro de la geometría como ciencia abstracta se alcanza en la civilización griega y helenística y tiene un punto culminante con los *Elementos* de Euclides. Este trabajo, que consta de 13 libros, fue escrito hace 2.300 años en Alejandría, en el actual Egipto, y es, tras la Biblia, el libro que ha aparecido en más ediciones. Ha sido traducido a multitud de idiomas y todavía puede encontrarse en prensa [6]. Las biografías de muchos matemáticos famosos indican que el trabajo de Euclides fue la motivación que los impulsó a convertirse en matemáticos.

En la geometría griega encontramos los conceptos de *axioma*, verdad *a priori* que no se demuestra, y *teorema*, afirmación que se demuestra a partir de los axiomas. Sin embargo, en la geometría griega no existe una clara distinción entre lo que es un sistema axiomático y lo que es la realidad.

Uno de los objetos principales de estudio de la geometría griega y que tomaremos como punto de partida, es lo que hoy conocemos como el *plano euclídeo*. Implícita en la noción de plano euclídeo está la elección de qué transformaciones del plano están permitidas. En este caso, las transformaciones permitidas son los giros y las traslaciones, es decir, los desplazamientos. Desde un punto de vista moderno, decimos que la geometría euclídea estudia las propiedades que son invariantes respecto a los desplazamientos: ángulos, longitudes y áreas.

Para hacer interesante el plano euclídeo hay que poblarlo de objetos, de «formas». Los objetos básicos de la geometría euclídea son los puntos y las rectas, con los que podemos formar ángulos, triángulos y cuadriláteros. Una figura un poco más elaborada es la circunferencia, que se define en términos de distancia: una circunferencia es el conjunto de puntos que están a una distancia fija de un punto llamado *centro*. Con estos objetos podemos realizar toda la geometría de regla y compás, es decir, aquellas construcciones que se pueden realizar únicamente trazando circunferencias y rectas.

Pero los matemáticos griegos consideraron también otro tipo de curvas como las cónicas (elipses, parábolas e hipérbolas) [1] o la tractriz (una curva de grado 3). Por ejemplo, los problemas de trisecar un ángulo o de duplicar un cubo, que no pueden ser resueltos con regla y compás, admiten una solución muy elegante si se permite el uso de cónicas generales y no únicamente circunferencias.

2. GEOMETRÍA PROYECTIVA

En nuestra excursión por la historia de la geometría vamos a utilizar la teoría de la intersección como hilo conductor.

Empezaremos con una simple observación. En el plano euclídeo, dos puntos distintos determinan una única recta. De forma dual, dos rectas distintas determinan un único punto de intersección. ¡No siempre! Dos rectas paralelas no se cortan y, por tanto, no determinan ningún punto. De alguna forma, la existencia de rectas paralelas puede verse como una anomalía del espacio en el que estamos trabajando, y nos podemos preguntar si es posible la construcción de un nuevo espacio donde dos rectas distintas siempre se corten en un punto, y que sea lo más parecido posible al plano euclídeo.

Al mirar un conjunto de rectas paralelas entre sí que no sean, a su vez, paralelas a la retina, tenemos la sensación de que todas ellas se cortan en un «punto de fuga» como en la figura 1. Es decir, si proyectamos dos rectas paralelas sobre un plano que no sea

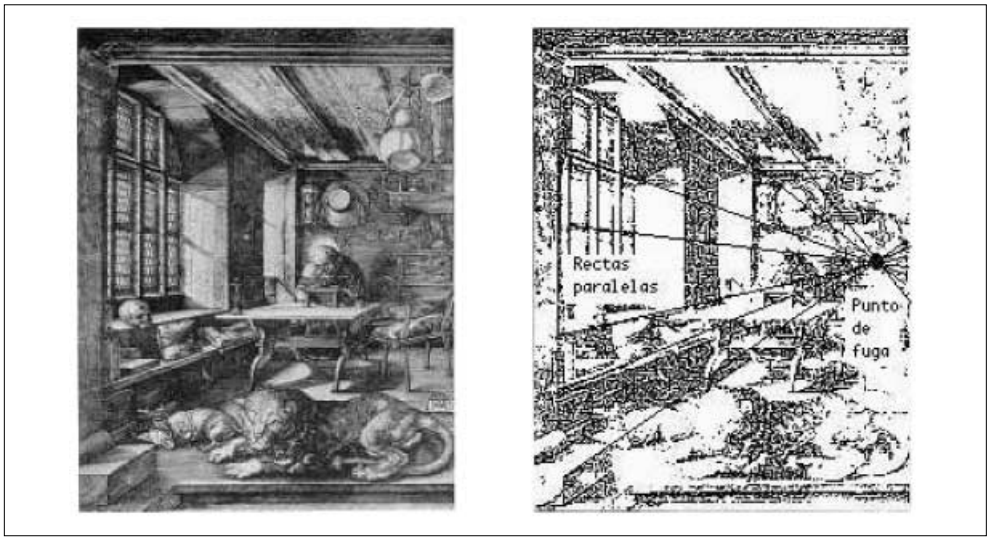


Figura 1. San Jerónimo en su celda (Durero).

paralelo a ellas, la propiedad de paralelismo se pierde. La geometría proyectiva estudia precisamente las propiedades que son invariantes respecto a proyecciones y, por tanto, en geometría proyectiva no hay diferencia entre rectas paralelas y rectas no paralelas. El plano proyectivo es el espacio de la geometría proyectiva y podemos construirlo a partir del plano euclídeo añadiendo todos los posibles puntos de fuga.

En primer lugar estudiaremos el caso de dimensión uno, es decir, construiremos la recta proyectiva real, que denotamos $\mathbb{RP}^1$. Dado que todas las rectas contenidas en una recta dada son paralelas entre sí, para construir la recta proyectiva real bastará añadir un único punto de fuga a la recta real $\mathbb{R}$. Por tanto, podemos poner $\mathbb{RP}^1=\mathbb{R}\cup\{\infty\}$. Otra forma de entender la recta proyectiva real es identificarla con el espacio que parametriza el conjunto de rectas del plano euclídeo que pasan por un punto dado.

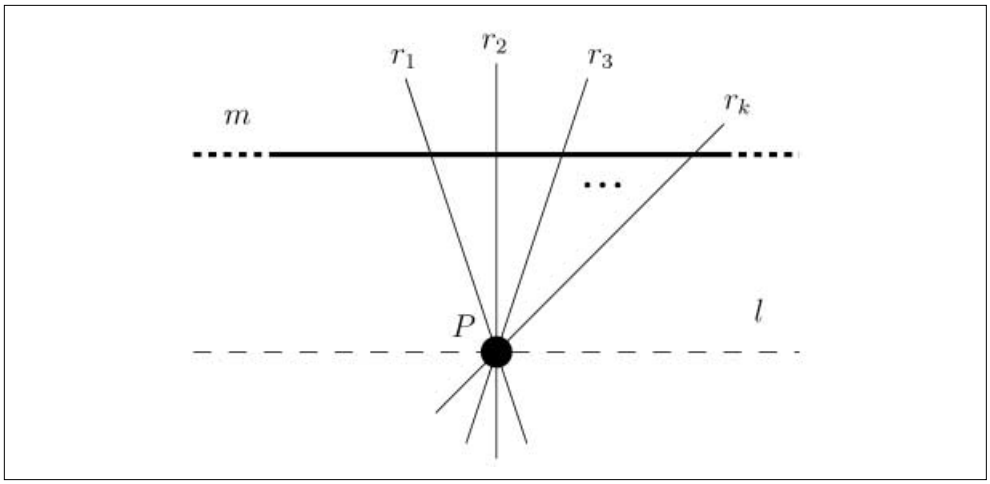


Figura 2. La recta proyectiva.

Este proceso está ilustrado en la figura 2. Para parametrizar el conjunto de rectas del plano que pasan por el punto P, trazamos una recta auxiliar m. Cada recta r_1, r_2, r_3, etc., corresponde a un punto de m. Excepto la recta l, que no corta a m y a la que le corresponde el punto que hemos añadido en el infinito.

El plano proyectivo lo podemos construir de forma análoga como el conjunto de rectas del espacio tridimensional que pasan por un punto fijo P, como se ilustra en la figura 3. Para ello, trazamos un plano auxiliar π que no contenga a P. A cada recta por P, por ejemplo r_1, le corresponde un punto de π, en este caso Q_1. Pero cualquier recta que esté contenida en el plano π', paralelo a π por P, no cortará a π. Por tanto, tenemos que añadir tantos puntos de fuga como rectas contenidas en π pasan por P. Es decir, tenemos que añadir toda una recta proyectiva real en el infinito.

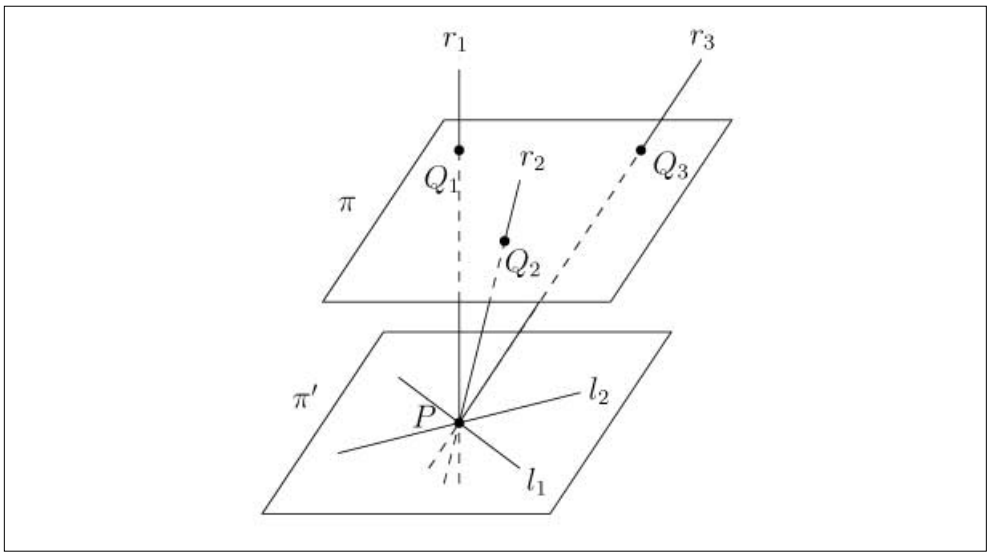


Figura 3. El plano proyectivo.

Con la construcción del plano proyectivo hemos unificado la teoría de intersección de rectas. Dos rectas distintas siempre se cortan en un punto. Podemos ahora estudiar la intersección de otras curvas. El siguiente ejemplo es la intersección de una recta y de una circunferencia. En el plano euclídeo, una recta y una circunferencia pueden cortarse en dos puntos, pero también en un solo punto si la recta es tangente a la circunferencia, o incluso pueden no tener ningún punto en común. A pesar de haber añadido toda una recta de puntos nuevos, no hay ninguna diferencia entre realizar la intersección en el plano euclídeo o en el plano proyectivo. Es decir, en la teoría de intersección de curvas más generales aparecen fenómenos nuevos que no se pueden explicar mediante los puntos en el infinito. Para poder entender lo que ocurre tenemos que acercar el álgebra a la geometría.

3. COORDENADAS CARTESIANAS

El uso de coordenadas en geometría fue introducido de forma independiente por Fermat y, sobre todo, por Descartes [14] en la primera mitad del siglo XVII. El sistema de coordenadas usual se denomina *cartesiano* en honor a este último. Un sistema de coordenadas cartesianas del plano consiste en la elección de un origen de coordenadas O y de dos ejes ortogonales OX y OY. A cada punto P del plano se le hace corresponder

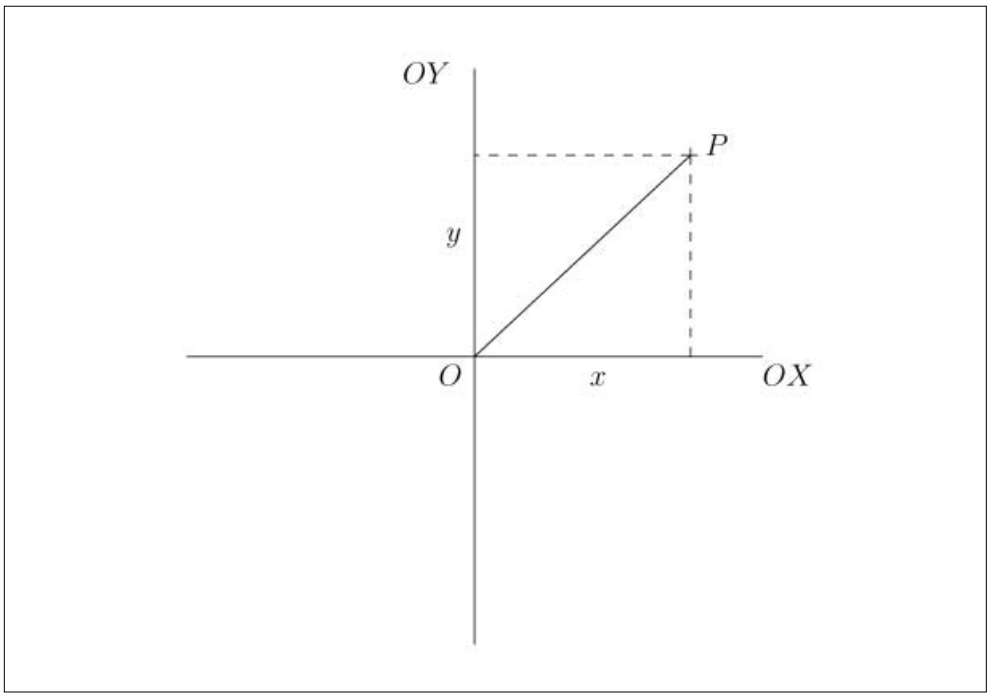


Figura 4. Coordenadas cartesianas.

un par de números reales (x, y) donde el número x es la longitud (con signo) de la proyección ortogonal del segmento OP sobre el eje OX y el número y es la longitud de la proyección ortogonal del segmento OP sobre el eje OY. Esta construcción está representada en la figura 4.

Análogamente, los puntos del espacio tridimensional se pueden representar por una terna de números (x, y, z).

Una vez introducidas las coordenadas cartesianas, una curva se puede describir mediante una ecuación, y la intersección de dos curvas se obtiene resolviendo simultáneamente las ecuaciones de ambas curvas.

Veamos ahora el problema de intersectar una recta y una circunferencia desde el punto de vista algebraico. En coordenadas cartesianas, una recta tiene por ecuación $Ax + By + C = 0$, mientras que la ecuación de una circunferencia es $(x-p)^2 + (y-q)^2 = r^2$. Para resolver simultáneamente ambas ecuaciones, supongamos que $B \neq 0$. Podemos despejar y de la ecuación de la recta y substituirla en la ecuación de la circunferencia. Una vez simplificada nos quedará una ecuación de segundo grado de la forma $ax^2 + bx + c = 0$ que posee

$$\begin{cases} \text{dos soluciones si } b^2 - 4ac > 0, \\ \text{una solución doble si } b^2 - 4ac = 0, \\ \text{ninguna solución si } b^2 - 4ac < 0. \end{cases}$$

Vemos en este ejemplo que hay dos tipos de fenómenos distintos. El primero es que hay puntos de intersección que se deben contar más de una vez. Es decir, los puntos de intersección tienen una multiplicidad. El segundo fenómeno está relacionado con que el cuadrado de un número real siempre es positivo y, por tanto, determinadas ecuaciones algebraicas como $x^2 = -1$ no tienen solución.

4. NÚMEROS COMPLEJOS

Para poder encontrar las soluciones de ecuaciones algebraicas que no tienen solución real se introducen los números complejos. Estos números, que aparecen ya en los trabajos de Bombelli en el Renacimiento, no fueron completamente aceptados como entidades matemáticas de pleno derecho hasta los trabajos de Peacock, Hamilton y Gauss.

El ingrediente principal necesario para construir los números complejos es la unidad imaginaria i, que satisface la ecuación $i^2 = -1$. Los números complejos son números de la forma $a + bi$ con a y b números reales. Por tanto, el conjunto de los números complejos se puede representar mediante un plano.

Una vez introducidos los números complejos, toda ecuación de segundo grado tiene o bien una solución doble o bien dos soluciones distintas. Lo que es una particularidad de los números reales es que todas las ecuaciones algebraicas de una variable se pueden resolver completamente con números complejos. Éste es el contenido del teorema fundamental del álgebra, que fue enunciado por primera vez por A. Girard en 1629, y que fue demostrado por Gauss de cuatro formas distintas en 1799, 1815, 1816 y 1848.

Teorema 4.1 (Teorema fundamental del álgebra). *El número de soluciones (contadas con su multiplicidad) de un polinomio con coeficientes complejos es igual al grado del polinomio.*

5. COORDENADAS PROYECTIVAS

A partir de las coordenadas cartesianas en el espacio euclídeo tridimensional podemos diseñar unas coordenadas para el plano proyectivo, que, recordemos, habíamos identificado con el conjunto de rectas del espacio que pasan por un punto dado. Para

parametrizar el conjunto de rectas que pasan por el punto O podemos hacer la siguiente observación. Dos puntos P y Q, distintos de O, con coordenadas (x,y,z) y (x',y',z'), están en la misma recta que pasa por O si, y sólo si, existe un número real λ tal que $(x,y,z)=(\lambda x,\lambda y,\lambda z)$. Por tanto, los puntos del plano proyectivo se pueden representar por ternas de números reales $(x:y:z)\neq(0:0:0)$ con la relación de equivalencia $(x:y:z)\sim(\lambda x:\lambda y:\lambda z)$. Los puntos del plano euclídeo original son los puntos de la forma $(x:y:1)$ mientras que la recta en el infinito está formada por los puntos $(x:y:0)$.

Una de las ventajas de usar el álgebra para representar objetos geométricos es que dejamos de estar limitados por las tres dimensiones del espacio usual. Los puntos de un espacio proyectivo real de dimensión n se pueden representar por secuencias de $n+1$ números reales $(x_0:\cdots:x_n)\neq(0:\cdots:0)$ con la relación de equivalencia $(x_0:\cdots:x_n)\sim(\lambda x_0:\cdots:\lambda x_n)$.

Desde el punto de vista algebraico, introducir los números complejos no presenta ninguna dificultad. Los puntos del espacio proyectivo complejo de dimensión n, que denotaremos $\mathbb{CP}^n$, se pueden representar por secuencias de $n+1$ números complejos $(x_0:\cdots:x_n)\neq(0:\cdots:0)$ con la relación de equivalencia $(x_0:\cdots:x_n)\sim(\lambda x_0:\cdots:\lambda x_n)$.

Desde el punto de vista geométrico, dado que los números complejos forman un plano, que tiene dimensión 2, el espacio proyectivo complejo de dimensión compleja n tiene dimensión real $2n$. Por ejemplo, la recta proyectiva compleja tiene el aspecto de una esfera de dimensión 2, mientras que el plano proyectivo complejo tiene dimensión real 4 y, por tanto, es difícil de imaginar.

En el espacio proyectivo, los objetos geométricos vendrán dados por ecuaciones polinómicas en las coordenadas. Sin embargo, dada la ambigüedad en la elección de coordenadas (distintas secuencias de números representan el mismo punto), las ecuaciones polinómicas tienen que ser homogéneas, es decir, todos los términos tienen que tener el mismo grado. Por ejemplo, una curva plana proyectiva compleja vendrá dada por un polinomio homogéneo de grado d en tres variables. La recta euclídea $Ax+By+C=0$ tiene ecuación proyectiva $Ax+By+Cz=0$, mientras que la circunferencia euclídea $(x-p)^2+(y-q)^2=r^2$ tendrá ecuación proyectiva $(x-pz)^2+(y-qz)^2=r^2z^2$.

Con el plano proyectivo complejo hemos alcanzado nuestro objetivo de encontrar un espacio donde la teoría de intersección sea lo más uniforme posible. Esto queda reflejado por el teorema de Bezout, que es una generalización geométrica del teorema fundamental del álgebra.

Teorema 5.1 (Teorema de Bezout). *Dos curvas de $\mathbb{CP}^2$, de grados n y m sin componentes comunes se cortan en exactamente nm puntos contados con multiplicidad.*

6. PLANO EUCLÍDEO Y PLANO PROYECTIVO

Llegados a este punto el lector puede pensar que los matemáticos son un poco tramposos. Teníamos un problema: la teoría de interseción en el plano euclídeo, que no tenía una solución satisfactoria. Al no encontrar la solución adecuada hemos cambiado las reglas del juego, complicando además el problema añadiendo puntos en el infinito y números complejos.

Por el contrario, el punto de vista matemático es que el plano proyectivo complejo es un objeto mucho más simple que el plano euclídeo, en el sentido de que es mucho más uniforme y con propiedades mucho más definidas. Además podemos recuperar el plano euclídeo a partir del plano proyectivo como explicaremos a continuación.

Los números complejos tienen una operación fundamental denominada *conjugación* y que denotaremos por σ. Esta operación actúa cambiando la unidad imaginaria por su opuesto

$$\sigma(a+bi) = a-bi.$$

Los números reales son precisamente aquellos números que no cambian al aplicarles la conjugación. Esta operación se puede extender al plano proyectivo complejo (o incluso al espacio proyectivo complejo de cualquier dimensión) y recuperamos el plano proyectivo real como el conjunto de puntos invariantes por conjugación compleja.

Los puntos del infinito forman una recta L del plano proyectivo. Los puntos del plano euclídeo son los que no pertenecen a L.

Por último, las propiedades métricas del plano euclídeo se pueden codificar mediante una cónica Ω contenida en L. Así el plano euclídeo se recupera como una cuaterna $(\mathbb{CP}^2, \sigma, L, \Omega)$.

Dadas ahora dos curvas en el plano euclídeo de grados n y m, por el teorema de Bezout sabemos que en el plano proyectivo complejo encontraremos nm puntos de intersección. Estos puntos los podemos clasificar de la siguiente forma: los puntos invariantes por σ son reales y tendrán un reflejo en el plano euclídeo. Los puntos que no son invariantes por σ vendrán en parejas de puntos conjugados y no tendrán un reflejo en el plano euclídeo. Los puntos de intersección invariantes por σ se clasifican a su vez dependiendo de si pertenecen a L o no. Los puntos que no son de L son los puntos de intersección propios, mientras que los puntos de L corresponderían a direcciones asintóticas comunes.

Es decir, hemos dividido el problema de calcular la intersección de dos curvas reales en dos problemas más pequeños. Primero calculamos la intersección en el plano proyectivo complejo, para lo que disponemos de un teorema que nos asegura cuantas soluciones debemos encontrar. En segundo lugar, clasificamos las soluciones obtenidas para obtener las soluciones propias.

7. VARIEDADES ALGEBRAICAS

Hemos visto que se puede hacer geometría en diferentes espacios. Hemos empezado con el plano euclídeo y hemos llegado a los espacios proyectivos complejos. Pero no tenemos que parar aquí, podemos extender la geometría al estudio de otros espacios. Como subconjuntos de los espacios proyectivos complejos, encontramos a las variedades algebraicas (proyectivas complejas), que son las soluciones de sistemas de ecuaciones polinómicas homogéneas. Un primer ejemplo lo hemos visto en el caso de las curvas planas, que son el conjunto de soluciones de una única ecuación. En geometría algebraica, las propias variedades algebraicas alcanzan el estatus de espacio «abstracto» y son, a su vez, el espacio ambiente de una nueva geometría. Dentro de las variedades algebraicas, las que tienen una estructura local más sencilla son las variedades lisas. Una variedad lisa de dimensión n, localmente, se parece a $\mathbb{C}^n$. En la figura 5 se puede observar una curva compleja lisa (una cúbica de ecuación $y^2 = x^3 + x$ tendría este aspecto) y una curva singular (una cúbica nodal de ecuación $x^3 = x^2 - y^2$).

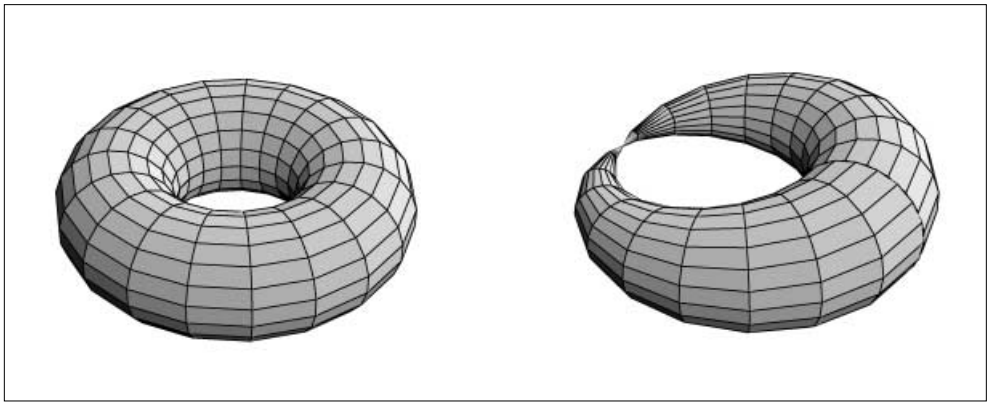


Figura 5. Una curva lisa y una curva con un punto singular.

El espacio en el que queremos hacer geometría, y del cual habla la conjetura de Hodge, es una variedad algebraica lisa compleja y proyectiva cualquiera. Las figuras interesantes dentro de esta variedad son sus subvariedades algebraicas, es decir, aquellas formas que podemos describir mediante ecuaciones polinómicas. Combinaciones lineales formales de subvariedades algebraicas se denominarán *ciclos algebraicos*. Nuestro objetivo es entender la forma de estas variedades algebraicas y la estructura de los ciclos algebraicos.

Dado que queremos trabajar en dimensión arbitraria, necesitamos desarrollar herramientas sofisticadas, una de las cuales es la topología algebraica.

8. TOPOLOGÍA

La topología se origina con Euler y el problema de los puentes de Könisberg. La topología se puede describir como la geometría del caucho, en la que se permite deformar los objetos. De una manera un poco más formal podemos decir que la topología estudia las propiedades que son invariantes respecto de aplicaciones continuas.

La topología combinatoria nace con Poincaré [13]. La idea básica de la topología combinatoria es que podemos construir un espacio complicado mediante una colección de bloques básicos sencillos. La combinatoria de esta construcción nos permite entender la estructura del espacio.

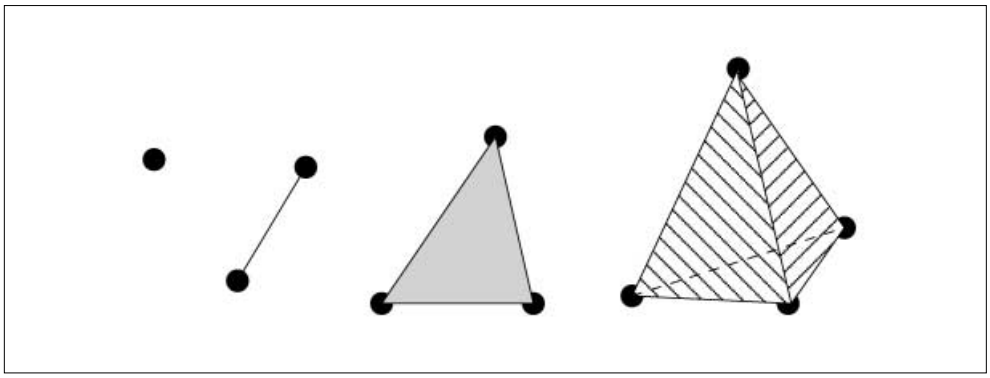


Figura 6. Símplices de dimensión 0, 1, 2 y 3.

Los bloques básicos se denominan *símplices*. Matemáticamente, un simplex estándar de dimensión n es el subconjunto de $\mathbb{R}^{n+1}$ formado por los puntos $(x_0 : \cdots : x_n)$ tales que

$$x_0 + \cdots + x_n = 1,\ x_i > 0,\ i = 0, \ldots, n.$$

Un simplex es una aplicación continua del simplex estándar en un espacio topológico. En la figura 6 están representados símplices de dimensiones 0 a 3.

Una combinación lineal formal de símplices se denomina una *cadena*. A cada simplex le corresponde una cadena de dimensión inferior denominada su *frontera*. En la figura 7 se ilustran las fronteras de algunos símplices. Es importante tener en cuenta que cada componente de la frontera tiene un signo. Este signo permite que símplices adyacentes tengan alguna componente de la frontera común, pero con distinto signo. De esta forma, al calcular la frontera de la suma de ambos símplices, la componente común se cancela.

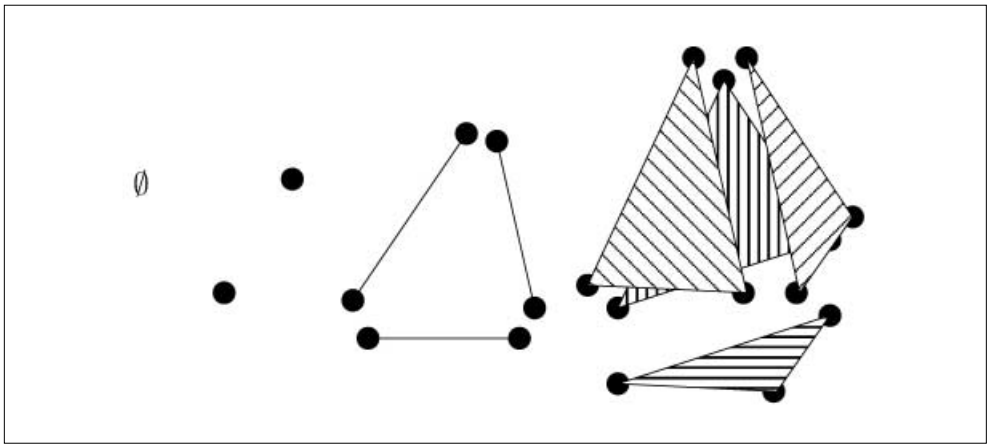


Figura 7. Frontera de los símplices de dimensión 0, 1, 2 y 3.

Una cadena cuya frontera es nula, es decir, tal que todas las componentes de la frontera se han cancelado entre sí, se denomina un *ciclo*. Diremos que dos ciclos son equivalentes si su diferencia es la frontera de alguna cadena. En particular un ciclo es equivalente a cero si es la frontera de una cadena.

En la figura 8 vemos un ciclo de dimensión uno en una esfera que se puede obtener como la frontera de una cadena de dimensión dos.

El conjunto de clases de ciclos de dimensión n de un espacio X se denomina *el grupo de homología singular* de X y se denota $H_n(X, \mathbb{Z})$. El conjunto de los números enteros $\mathbb{Z}$ que aparece en la notación indican que estamos tomando combinaciones

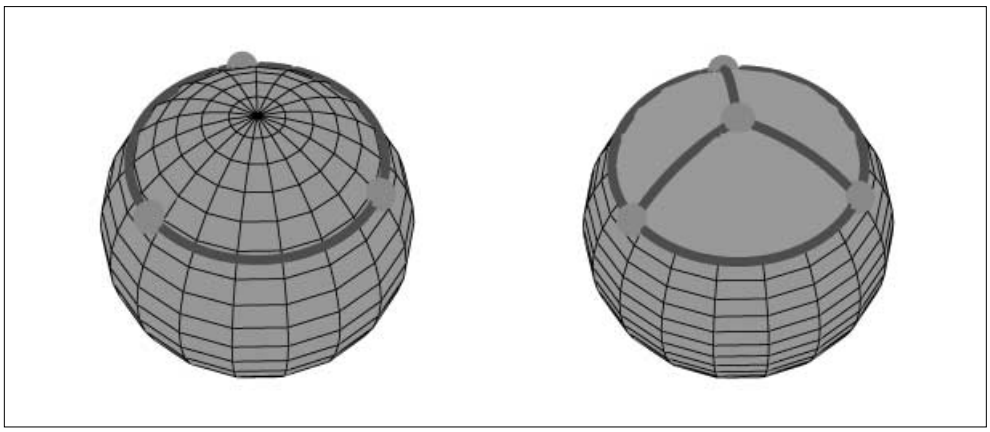


Figura 8. Un ciclo en una esfera.

lineales formales con coeficientes enteros. De la misma forma podríamos usar otros coeficientes, como los números racionales.

En la figura 9 está representado un toro (la superficie de una rosquilla o una curva elíptica compleja lisa) con un 1-ciclo que no es equivalente a cero. Se puede comprobar que el primer grupo de homología de la esfera es nulo: todos los 1-ciclos son equivalentes a cero, mientras que el primer grupo de homología del toro tiene dos generadores independientes. Es decir, la topología nos permite distinguir ambos tipos de superficies.

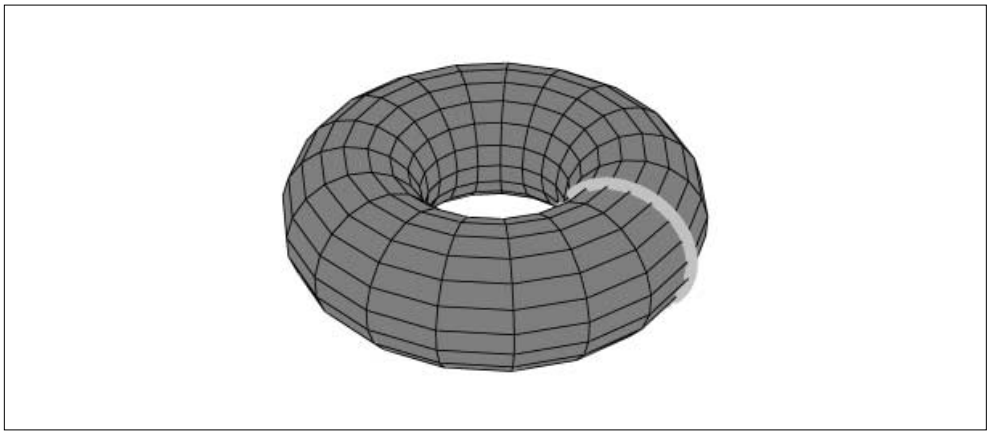


Figura 9. Un toro con un 1-ciclo que no es equivalente a cero.

9. COHOMOLOGÍA

Además de los grupos de homología singular presentados en la sección anterior, existe la noción dual de *grupos de cohomología singular*. Estos grupos se denotan $H^r(X, \mathbb{Z})$. Si X es una variedad compleja de dimensión n, estos grupos pueden ser distintos de cero únicamente para $0 \leq r \leq 2n$. Una de las ventajas que tiene la cohomología es que se puede definir un producto

$$H^r(X, \mathbb{Z}) \otimes H^s(X, \mathbb{Z}) \longrightarrow H^{r+s}(X, \mathbb{Z}).$$

En el caso de las variedades proyectivas lisas hay una relación muy estrecha entre la homología y la cohomología.

Teorema 9.1 (Dualidad de Poincaré). *Sea* X *una variedad proyectiva compleja lisa e irreducible de dimensión compleja n. Entonces existe un isomorfismo natural*

$$H^r(X,\mathbb{Z}) \to H_{2n-r}(X,\mathbb{Z}).$$

En particular se tiene una identificación natural

$$H^{2n}(X,\mathbb{Z}) \to \mathbb{Z}.$$

Este teorema nos permite interpretar la teoría de intersección de forma puramente topológica. Sea X una variedad compleja de dimensión n. Sean Y y Z dos subvariedades algebraicas de dimensión (compleja) r y s. Supongamos que $r+s=n$ y que ambas subvariedades se cortan en un número finito de puntos. Por ejemplo, X puede ser el plano proyectivo y las subvariedades Y y Z ser dos curvas sin componentes comunes. Las subvariedades Y y Z definen clases de homología de dimensión (real) $2r$ y $2s$, respectivamente. Por el teorema de dualidad definen clases de cohomología de grados $2n-2r$ y $2n-2s$. El producto de estas clases tiene grado $2n-2r+2n-2s=2n$. Como hemos identificado $H^{2n}(X,\mathbb{Z})$ con $\mathbb{Z}$ este producto en un número entero. Este número es precisamente el número de puntos de intersección contados con su multiplicidad.

Al trabajar con la cohomología es conveniente emplear la *codimensión* en lugar de la dimensión. Una subvariedad de dimensión r dentro de una variedad de dimensión n se dice que tiene codimensión $n-r$. Como en el caso de la dimensión hay que distinguir entre la codimensión compleja y la codimensión real.

10. EL PROBLEMA DE HODGE

Una variedad algebraica tiene una estructura muy sutil. Por un lado, las funciones algebraicas son muy pocas en comparación con todas las funciones continuas. Por otro, para que dos variedades algebraicas sean isomorfas es necesario obtener una aplicación algebraica entre ellas que posea una inversa algebraica. Si consideramos el espacio topológico subyacente y nos olvidamos de la estructura algebraica estamos simplificando considerablemente el problema. Pongamos por caso que queremos clasificar las curvas planas lisas. Es fácil ver que todas las rectas proyectivas son isomorfas entre sí y cualquier cónica lisa es también isomorfa algebraicamente a una recta proyectiva. Lo siguiente que podríamos preguntarnos es si una cúbica lisa, es decir, una curva elíptica, puede ser isomorfa a una recta proyectiva. Sin necesidad de hacer ninguna consideración algebraica podemos concluir que no, puesto que el espacio subyacente a una recta proyectiva compleja es una esfera y el subyacente a una curva elíptica es un toro, y ya hemos visto que los grupos de homología permiten distinguir entre una esfera y un toro. Es decir, una recta proyectiva y una curva elíptica son topológicamente distintas.

Por el contrario, si queremos comparar dos curvas elípticas entre sí, la topología no nos ayuda, pues desde el punto de vista topológico todos los toros son isomorfos. Por el contrario, desde el punto de vista algebraico existe un continuo de curvas elípticas distintas.

Hemos visto que un problema algebraico, como la clasificación de las variedades algebraicas, se puede descomponer en dos. Primero clasificamos las variedades desde el punto de vista topológico y dentro de cada tipo topológico clasificamos las variedades algebraicamente.

Éste es el punto de vista que defendió S. Lefschetz a principios del siglo XX. Lefschetz observó que los invariantes discretos que los matemáticos del siglo XIX habían asociado a las variedades algebraicas eran en realidad de naturaleza topológica. Incluso la teoría de intersección admite una interpretación topológica como hemos comentado anteriormente. Lefschetz además demostró muchos teoremas sobre la estructura topológica de las variedades algebraicas (véase, por ejemplo, [9]). En palabras del propio Lefschetz: «Mi destino ha sido clavar el arpón de la topología algebraica en el cuerpo de ballena de la geometría algebraica».

Estamos ya en situación de enunciar el problema de Hodge, al que la conjetura de Hodge pretende dar respuesta.

El problema de Hodge. *Identificar qué ciclos topológicos se pueden realizar algebraicamente.*

De forma más precisa, el problema de Hodge pretende identificar los ciclos topológicos que pueden representarse como combinación lineal de ciclos algebraicos.

La solución propuesta por Hodge a este problema hace intervenir el análisis.

11. INTEGRACIÓN

Las formas diferenciales son una generalización del concepto de función y fueron introducidas por E. Cartan (véase, por ejemplo, [2]) a principios del siglo XX para definir la integración en variedades diferenciables. No vamos a entrar en la definición de formas diferenciales, pero sí que mencionaremos algunas de sus propiedades más importantes.

Las formas diferenciales se clasifican por su grado. Las formas diferenciales de grado cero son las funciones diferenciables.

Existe un producto exterior que es compatible con el grado. El producto de una forma de grado r por una forma de grado s tiene grado $r+s$.

Una forma ω de grado r se puede «integrar» sobre una cadena (diferenciable) c de dimensión r. Esta integral se escribe

$$I = \int_c \omega.$$

Existe una diferencial exterior que a cada forma ω de grado r le hace corresponder una forma $d\omega$ de grado $r+1$. La diferencial exterior es una generalización de los conceptos de gradiente, rotacional y divergencia que encontramos en el análisis vectorial.

El teorema de Stokes, que es una generalización del teorema fundamental del cálculo, relaciona la derivada exterior de una forma diferencial con la frontera de una cadena.

Teorema 11.1 (Stokes).

$$\int_c d\omega = \int_{\partial c} \omega\,.$$

El teorema de Stokes insinúa la existencia de una relación profunda entre las formas diferenciales y la homología. Esta relación fue hecha explícita por Georges de De Rham (véase, por ejemplo, [5]). Decimos que una forma diferencial ω es *cerrada* si cumple $d\omega = 0$. Diremos que es *exacta* si existe otra forma diferencial η tal que $d\eta = \omega$. Dos formas cerradas se dice que son *cohomólogas* si su diferencia es exacta. Ésta es una relación de equivalencia. El grupo de clases de equivalencia de formas cerradas se denomina el *grupo de cohomología de De Rham* que denotaremos $H^*_{dR}(X, \mathbb{R})$. La relación precisa entre las formas diferenciales y la cohomología singular nos la da el teorema de De Rham:

Teorema 11.2 (De Rham). *Si* X *es una variedad diferenciable, se tiene un isomorfismo natural*

$$H^*_{dR}(X, \mathbb{R}) \to H^*(X, \mathbb{R}).$$

Este teorema se puede entender como un teorema sobre las propiedades globales de algunas ecuaciones en derivadas parciales. Si fijamos una forma diferencial ω, la ecuación $\omega = d\eta$ es un sistema de ecuaciones en derivadas parciales. Localmente (en un disco de $\mathbb{R}^n$), para que un sistema de ecuaciones diferenciales tenga solución, hay que pedirle una condición de compatibilidad: el ser integrable. Precisamente, la condición para que el sistema de ecuaciones sea integrable es $d\omega = 0$. Es decir, la ecuación $\omega = d\eta$ es integrable si, y sólo si, ω es cerrada. Por tanto, si ω es cerrada siempre podremos encontrar soluciones locales a esta ecuación. Lo que nos dice el teorema de De Rham es que la obstrucción a extender estas soluciones locales para obtener una solución global en una variedad es puramente topológica y está caracterizada por los grupos de homología.

El teorema de De Rham nos proporciona otra forma de interpretar el producto de intersección. Esta vez en términos del análisis. Sea X otra vez una variedad compleja proyectiva y lisa de dimensión compleja n y sean Y y Z dos subvariedades de dimensiones complejas r y s, con $n = r + s$. Supongamos que Y y Z se cortan en un número finito de puntos. Por el teorema de De Rham, la clase de cohomología de Z está representada por una forma diferencial ω_Z de grado $2n-2s = 2r$. Por tanto, podemos integrar esta forma a lo largo de Y. Entonces el número de puntos de intersección contados con su multiplicidad es exactamente la integral

$$Y \cdot Z = \int_Y \omega_Z.$$

12. LA CONTRIBUCIÓN DE HODGE

En el plano real $\mathbb{R}^2$, una función f se denomina *armónica* si cumple la ecuación

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0.$$

El operador $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$ se denomina la *laplaciana*. En una variedad diferenciable provista de una métrica (una variedad riemanniana) se puede generalizar la laplaciana para que actúe sobre las formas diferenciales. Una forma diferencial ω tal que $\Delta\omega = 0$ se denomina una forma diferencial *armónica*.

Teorema 12.1 (Hodge [7]). *Si* X *es una variedad riemanniana compacta, en cada clase de cohomología de De Rham existe una única forma diferencial armónica.*

Este teorema tiene muchas consecuencias en geometría compleja. Una de ellas afirma que, en una variedad proyectiva compleja y lisa, los grupos de cohomología de De Rham con coeficientes complejos se pueden descomponer en piezas:

$$H^n(X,\mathbb{C}) = \bigoplus_{p+q=n} H^{p,q}(X).$$

Esta descomposición en piezas, que se denomina *la estructura de Hodge* de la cohomología, refleja la estructura compleja de la variedad y puede proporcionar información muy sutil. Por ejemplo, si X es una curva (con la condición técnica adicional de no ser hiperelíptica), el primer grupo de cohomología con coeficientes complejos, $H^1(X,\mathbb{C})$, tiene tres estructuras adicionales:

1. El subgrupo de clases de cohomología con coeficientes enteros

$$H^1(X,\mathbb{Z}) \subset H^1(X,\mathbb{C}).$$

2. La descomposición $H^1(X,\mathbb{C}) = H^{1,0} \oplus H^{0,1}$.

3. El producto de intersección $H^1(X,\mathbb{Z}) \otimes H^1(X,\mathbb{Z}) \to \mathbb{Z}$.

A diferencia del tamaño de este grupo, que nos proporciona únicamente una clasificación de las curvas en tipos topológicos, las tres estructuras anteriores, en su conjunto, proporcionan suficiente información para recuperar completamente la curva algebraica.

13. LA CONJETURA DE HODGE

Si X es una variedad compleja proyectiva lisa de dimensión n e Y es una subvariedad de dimensión r y, por tanto, de codimensión $p = n - r$, entonces la clase de cohomología de Y pertenece a la pieza $H^{p,p}$ de $H^{2p}(X,\mathbb{C})$. Esta observación lleva a definir los ciclos de Hodge como aquellos ciclos con coeficientes racionales que caen en las piezas de tipo (p,p):

$$\mathrm{Hod}^p(X) = H^{2p}(X,\mathbb{Q}) \cap H^{p,p}(X).$$

Por fin estamos en condiciones de enunciar la conjetura de Hodge.

Conjetura 13.1. *Todo ciclo de Hodge es cohomólogo a una combinación lineal racional de ciclos algebraicos [8].*

Veamos un ejemplo de clases de cohomología a las que se les podría aplicar la conjetura de Hodge. Si X es una variedad algebraica proyectiva, lo mismo es cierto para el producto cartesiano $X \times X$. Dentro de este producto hay una subvariedad distinguida, la diagonal, que es el conjunto de puntos de la forma (x,x). La diagonal no es más que la gráfica de la aplicación «identidad» y está representada en la figura 10.

El hecho de que $X \times X$ es un producto introduce una estructura extra en su cohomología: la descomposición de Künneth.

$$H^n(X \leftrightarrow X,\mathbb{Q}) = \bigoplus_{r+s=n} H^r(X,\mathbb{Q}) \otimes H^s(X,\mathbb{Q})$$

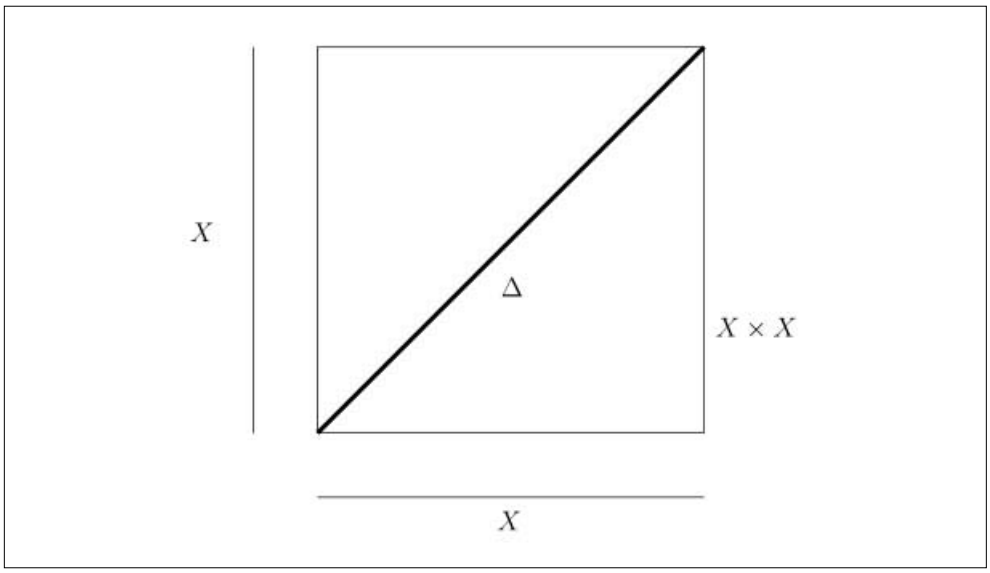


Figura 10. La diagonal.

La diagonal define una clase de cohomología que se descompone en una suma de clases bajo la descomposición de Künneth. Cada una de estas clases es un ciclo de Hodge. Por tanto, si la conjetura de Hodge fuera cierta, estas clases, estas piezas en las que se descompone la diagonal, se podrían realizar mediante una combinación lineal racional de ciclos algebraicos. Actualmente no se sabe si estas clases se pueden o no realizar como ciclos algebraicos.

Pero, ¿qué casos de la conjetura de Hodge son conocidos? El único resultado completamente general sobre la conjetura de Hodge fue demostrado por Lefschetz y es anterior a la conjetura de Hodge. De hecho, el resultado de Lefschetz es el inspirador de la conjetura de Hodge. El teorema (1,1) de Lefschetz afirma que si X es una variedad proyectiva compleja lisa de dimensión (compleja) n, entonces la conjetura de Hodge se cumple para las clases de homología de dimensión (real) $2n-2$. Es decir, para las subvariedades de codimensión uno. Los ciclos algebraicos obtenidos a partir de subvariedades de codimensión uno se denominan *divisores*. Otro teorema de Lefschetz implica que, si la conjetura de Hodge es cierta para ciclos de dimensión $2p$, entonces también es cierta para ciclos de dimensión $2n-2p$. Esto implica que la conjetura de Hodge es cierta para toda variedad de dimensión menor o igual a tres. Por tanto, los primeros casos interesantes aparecen en dimensión compleja 4 y, por tanto, en dimensión real 8.

Otros resultados generales se refieren a grassmannianas e hipersuperficies.

Las grassmannianas y las variedades de banderas son una familia de variedades que generalizan de alguna forma los espacios proyectivos. Estas variedades tienen toda

su cohomología generada por ciclos algebraicos. Por tanto cumplen trivialmente la conjetura de Hodge.

Si X es una hipersuperficie del espacio proyectivo $\mathbb{P}^{n+1}$ (es decir una subvariedad de codimensión uno) y, por tanto, tiene dimensión compleja n, entonces X cumple la conjetura de Hodge en todas las dimensiones reales excepto quizás en la dimensión real media n.

Aparte de estos casos, se conocen diversos ejemplos concretos. Por ejemplo, toda variedad unirreglada (una variedad que puede recubrirse mediante curvas racionales) de dimensión 4 satisface la conjetura de Hodge. Por el contrario, Mumford ha construido un ejemplo de variedad abeliana de dimensión 4 para la que no se conoce si la conjetura de Hodge es cierta.

Hasta que no dispongamos de una demostración o de un contraejemplo no quedará zanjada la cuestión de si la conjetura de Hodge es cierta o no.

En realidad, hemos visto que hay muy pocos indicios en favor de esta conjetura. El único resultado general es el teorema (1,1) de Lefschetz, que trata del caso de divisores (subvariedades que localmente se describen mediante una única ecuación). En geometría algebraica el caso de los divisores es siempre mucho más simple que el caso general, con propiedades muy específicas. Por tanto, inferir resultados generales a partir del comportamiento de los divisores es muy arriesgado. Como hemos comentado, aparte del teorema (1,1) de Lefschetz, sólo se conocen algunos casos particulares.

Por ejemplo, André Weil ha sido muy crítico con la proliferación de la palabra *conjetura* en las matemáticas actuales. Respecto a la conjetura de Hodge ha dicho:

«La pregunta propuesta por la *conjetura de Hodge* es muy natural... Desgraciadamente, a pesar de la palabra *conjetura*, no hay, que yo sepa, ni la sombra de una razón para creerla; se rendiría un servicio a los geómetras si se pudiera zanjar la cuestión mediante un contraejemplo».

14. GENERALIZACIONES Y APLICACIONES

Una variedad compleja proyectiva es sólo una de las facetas que puede presentar una variedad algebraica. Por ejemplo, si una variedad está definida por un conjunto de polinomios con coeficientes enteros, podemos, por un lado, considerar la variedad compleja asociada, pero también podemos reducir los coeficientes de las ecuaciones módulo un primo p y considerar la variedad así obtenida. Ésta es una variedad definida sobre el cuerpo finito $\mathbb{Z}/p\mathbb{Z}$.

Además de la cohomología singular y de la cohomología de De Rham que hemos comentado, para tener en cuenta todas estas facetas de las variedades algebraicas, se han desarrollado muchas otras teorías de cohomología, como la cohomología étale. Todas estas teorías de cohomología, a pesar de tener un origen muy distinto, tienen propieda-

des muy similares. Para explicar estas similitudes Grothendieck postuló la existencia de una categoría de «motivos» algebraicos. Ésta debería ser una categoría abeliana y tensorial y proporcionar una teoría de cohomología universal. Todas las demás teorías de cohomología no serían más que un reflejo de esta teoría universal. De esta forma, si se demuestra un teorema en la categoría de motivos, automáticamente este teorema estará demostrado en cualquier teoría de cohomología.

Sin embargo, la construcción y la demostración de las propiedades necesarias de la categoría de motivos están bloqueadas por nuestra incapacidad para producir suficientes ciclos algebraicos. Por ejemplo, uno de los ingredientes necesarios para establecer las propiedades de la categoría de motivos es que las componentes de Künneth de la diagonal sean algebraicas. Ésta es una de las conjeturas estándard de Grothendieck y ya hemos visto que es consecuencia de la conjetura de Hodge. Por tanto, si la conjetura de Hodge fuera cierta, dispondríamos de herramientas muy potentes para el estudio de las variedades algebraicas. Hay que tener en cuenta, sin embargo, que la teoría de motivos no necesita de toda la conjetura de Hodge, sino únicamente de algunas consecuencias que están englobadas en las llamadas *conjeturas estándar* de Grothendieck.

La teoría de motivos sugiere que debe de haber un análogo de la conjetura de Hodge para las otras facetas de las variedades algebraicas. En 1963 J. Tate [15] enunció una versión aritmética de la conjetura de Hodge en el que el papel de la estructura de Hodge se reemplaza por la acción del grupo de Galois. De esta conjetura no se conoce ni siquiera si es cierto el caso de divisores (el análogo al teorema (1,1) de Lefschetz). El hecho de que la conjetura de Hodge y la conjetura de Tate son dos caras de la misma moneda se refleja, entre otros, en los siguientes ejemplos.

Si A es una variedad abeliana definida sobre un cuerpo finitamente generado $L \subset \mathbb{C}$, entonces la conjetura de Tate para A implica la conjetura de Hodge para A [12]. En la dirección contraria, si la conjetura de Hodge es cierta para todas las variedades abelianas de tipo CM sobre $\mathbb{C}$, entonces la conjetura de Tate es cierta para todas las variedades abelianas definidas sobre la clausura algebraica de un cuerpo finito [11].

El lector interesado puede encontrar más información sobre la conjetura de Hodge en la página web del Instituto Clay [4], donde, además de un artículo de Deligne sobre la conjetura de Hodge, se puede encontrar el vídeo de una conferencia de D. Freed sobre el tema. También puede consultar el extenso estudio realizado por J. D. Lewis [10] respecto al estado actual de la conjetura de Hodge.

REFERENCIAS

[1] APOLLONIUS, *Conics* (trad. de R.C. Taliaferro), Dana Densmore, Editor, 1998.

[2] E. CARTAN, *Les systèmes différentiels extérieurs et leurs applications geométriques*, Actualités Sci. Ind., n.º 994, Hermann, 1945.

[3] A.B. CHACE y otros, *The Rhind Mathematical Papyrus*, Oberlin, Ohio, 1927-29.
[4] http://www.claymath.org/millennium/.
[5] G. DE RHAM, *Variétés différentiables. Formes, courants, formes harmoniques*, Hermann, 1955.
[6] EUCLID, *The thirteen books of the elements* (trad. de T.L. Heath), Dover, 1956.
[7] W.V.D HODGE. *The theory and applications of harmonic integrals*, Cambridge University Press, 1941.
[8]—«The topological invariants of algebraic varieties», *International Congress of Mathematicians*, Amer. Math. Soc., 1950, p. 182-192.
[9] LEFSCHETZ, S. *L'analysis situs et la géométrie algébrique*, Gauthier-Villars, 1950.
[10] J.D. LEWIS, «A survey of the hodge conjecture», *CRM Monographs*, vol. 10, Am. Math. Soc., 1999.
[11] J.S. MILNE, «Lefschetz motives and the Tate conjecture», *Compositio Math.* 117 (1999), p. 45-76.
[12] PJATECKII-SAPIRO, «Interrelations between the Tate and Hodge conjectures for abelian varieties», *Math. URSS-Sb* 14 (1971), p. 615-625.
[13] H. POINCARÉ, «Analysis situs», *J. Ec. Polyt.* 1 (1895), p. 1-121.
[14] D.E. SMITH y M.L. LATHAM, *The geometry of René Descartes,* Dover, 1954.
[15] J. TATE, *Algebraic cycles and poles of zeta functions*, Arithmetic Algebraic Geometry (O.F.G. Schilling, ed.), Harper & Row, 1965, p. 93-111.

La conjetura de Poincaré. Agujeros y esferas

María Teresa Lozano Imízcoz
Universidad de Zaragoza

RESUMEN

Este artículo[1] trata de divulgar un viejo problema que ha merecido el honor de ser considerado uno de los problemas del milenio. Es conocido como *conjetura de Poincaré*. Es un problema que data de 1904. En esa fecha Poincaré afirmó que «Una 3-variedad simplemente conexa es la esfera tridimensional», pero no dio la demostración de este resultado por considerar que era demasiado larga. En otras palabras menos técnicas, esta conjetura dice que «un espacio ambiente sin agujeros es la esfera».

La búsqueda de una demostración rigurosa ha contribuido al desarrollo espectacular que ha tenido durante el siglo XX la rama de las matemáticas denominada topología. En este artículo, que recoge el contenido de la conferencia impartida en el Auditori de Caixa Sabadell, se hablará de su autor y de la génesis de este problema, dando las nociones elementales de topología necesarias para entenderlo. Se comentarán algunos métodos utilizados para demostrarlo y su situación actual.

HENRI POINCARÉ

Henri Poincaré nació el día 29 de abril de 1854 en Nancy (Francia). Era hijo de Léon Poincaré, profesor de medicina en la universidad. Estudió en el Lycée de su ciudad natal, que hoy lleva su nombre, donde permaneció durante once años (1862-1873), y destacó 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]

1. En enero de 2004 tuve la oportunidad de impartir la primera conferencia sobre este tema dentro de la Jornada Poincaré que organizó la Facultad de Matemáticas y Estadística de la Universidad Politécnica de Cataluña. Este artículo contiene en parte material desarrollado en aquella ocasión y en su publicación. Su inclusión también aquí pretende contribuir a una mayor difusión del tema.

2. Carta de Elliot à Liard a un amigo en 1872: «J'ai dans ma classe á Nancy, un monstre de mathématiques, c'est Henri Poincaré».

Después estudió en L'Ècole Polytechnique y en L'Ècole des Mines y se graduó en ingeniería de minas. Su tesis doctoral la realizó, bajo la dirección de Charles Hermite, en ecuaciones diferenciales, y la defendió en París en 1879.

Fue profesor en la Universidad de Caen, la Facultad de Ciencias de París, la Sorbonne y L'Ècole Polytechnique. También trabajó durante algunos años como ingeniero en el Ministerio francés de Obras Públicas como responsable en parte del desarrollo del ferrocarril. Se casó con Poulain d'Andecy y tuvo cuatro hijos. Murió en París, el 17 de julio de 1912 de una embolia a la edad de 58 años.

En su trabajo era una persona metódica, dedicaba todos los días unas horas determinadas a investigar en su mesa de trabajo. Conocemos su método particular de trabajo, porque él mismo lo analiza y comenta en sus escritos. Una vez analizado un problema, no continuaba demasiado tiempo pensando exclusivamente en él, sino que pasaba a otro tema. Pensaba que su subconsciente seguía trabajando, de manera que cuando volviera a pensar conscientemente en un problema anterior sus ideas habrían madurado. Es evidente que este método le resultó muy eficiente, porque su producción científica fue extraordinaria. Otras de sus cualidades eran su buena memoria, su intuición, su visión espacial y su enorme poder de abstracción.

Figura 1. Fotografía de H. Poincaré.

Se puede decir que dedicó su vida a la investigación, y fue uno de los últimos sabios integrales, capaces de abarcar muchos aspectos de las matemáticas, la física e incluso la filosofía. Entre sus importantes contribuciones a la ciencia, destacaré sólo algunas: su descubrimiento y estudio de las *funciones automorfas*, que él llamó *fuchsianas* en honor de Fuchs, aunque este nombre no gustó a Klein. Cuenta Poincaré que durante un par de semanas estuvo tratando de probar metódicamente que no podían existir tales funciones (funciones complejas analíticas invariantes bajo ciertos grupos de transformaciones), sin conseguir lo propuesto. En una noche de insomnio, debida al café consumido la tarde anterior, las ideas acudieron fluidamente a su mente, de manera que en la mañana siguiente sólo tuvo que sentarse a escribir los resultados.

Otro importante concepto propuesto por Poincaré fue el de *punto homoclínico*, que le llevó, casi casualmente, al descubrimiento de la teoría del Caos. Esto sucedió con motivo de la memoria sobre el problema de los tres cuerpos que escribió y presentó al concurso que en 1887 el rey Oscar II de Suecia y Noruega convocó para celebrar su 60 cumpleaños. Poincaré ganó el concurso. La memoria tenía un error, descubierto por Phragmén, que era el editor de la revista *Acta Mathematica*, y su corrección le llevó al descubrimiento de la teoría del Caos. La memoria corregida se publicó en 1890. La teoría del Caos como área de investigación se desarrolló a partir de 1963, año en que Edward Lorenz encontró un sistema determinístico caótico usando un modelo atmosférico sencillo.

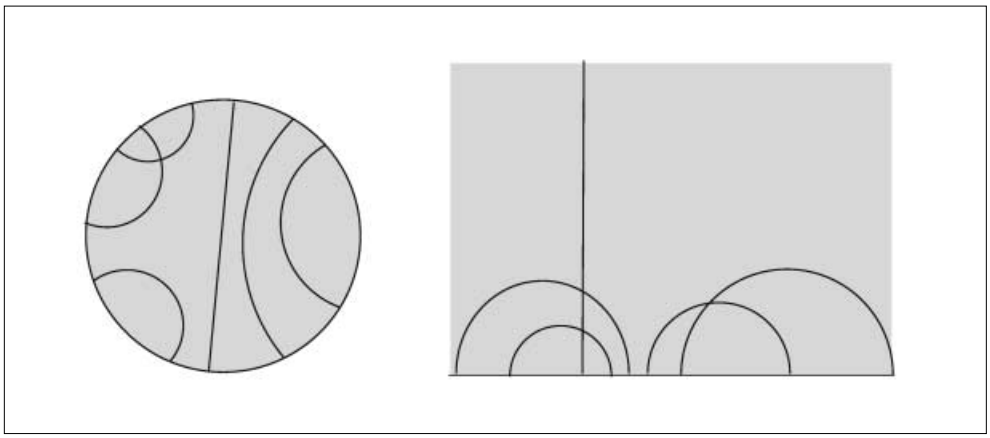


Figura 2. Modelos del plano hiperbólico.

Le debemos también los *modelos del plano hiperbólico* que llevan su nombre: el disco y el semiplano superior. Las geodésicas en estos modelos son circunferencias ortogonales al borde del modelo. Se trata de modelos conformes, en los que los ángulos

euclídeos son los ángulos hiperbólicos. Véase figura 2. El modelo del disco fue utilizado por el pintor y arquitecto Escher en algunos de sus célebres obras. Véase figura 3.

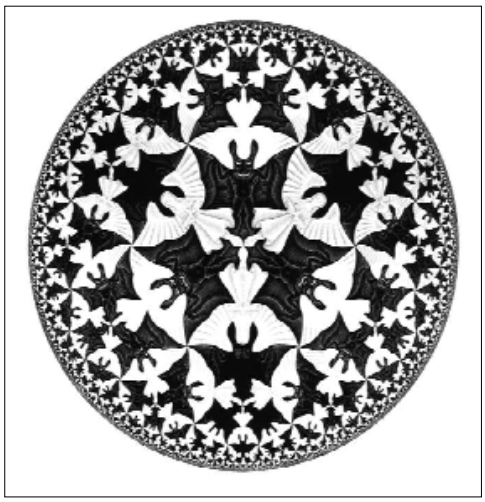


Figura 3. Cielo e infierno de Escher.

Trabajó también en diversos campos de matemática aplicada, como teoría cualitativa de ecuaciones diferenciales, en mecánica celeste, mecánica de fluidos, electromagnetismo, teoría de la relatividad.

Escribió tres libros sobre filosofía de la ciencia: *Ciencia e hipótesis*, *El valor de la ciencia* y *Ciencia y método*. *Últimos pensamientos* es su obra póstuma.

Sus estudios y descubrimientos en tantas áreas de la ciencia han hecho que sea considerado como el primer universalista después de Gauss, y quizás el último. Sus obras completas [14], publicadas por Gauthier-Villars (1916-1956), ocupan once volúmenes.

Es el creador de la rama de las matemáticas que hoy llamamos *topología* y que él denominó *Analysis sitûs*. Su estudio comprende doce artículos que publicó entre 1892 y 1912. He aquí la definición dada por Poincaré [15]:

> *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 aquéllas de las que se ocupa el Analysis sitûs.»

Su célebre conjetura, que hoy nos ocupa, ha tenido una importancia excepcional para el desarrollo de esta disciplina. A ella dedicamos el resto del artículo.

TOPOLOGÍA

Esta rama de las matemáticas nace como una abstracción de la geometría. Los objetos de estudio de esta ciencia son espacios maleables y dúctiles. Dos de estos espacios son equivalentes si uno se obtiene del otro por una deformación continua que en ningún caso contenga cortes o soldaduras. Imaginemos un objeto de plastilina que podemos moldear a nuestro gusto sin llegar a agujerearlo, ni romperlo, ni realizar autopegados. Desde el punto de vista topológico este objeto será en cada momento equivalente al de partida. Tampoco importan sus dimensiones. Para un topólogo es lo mismo una colina que una montaña. El pintor Salvador Dalí actuó como un topólogo cuando pintó sus relojes blandos. En la actualidad existen relojes digitales blandos, inspirados en la obra de Dalí, como el de la siguiente fotografía.

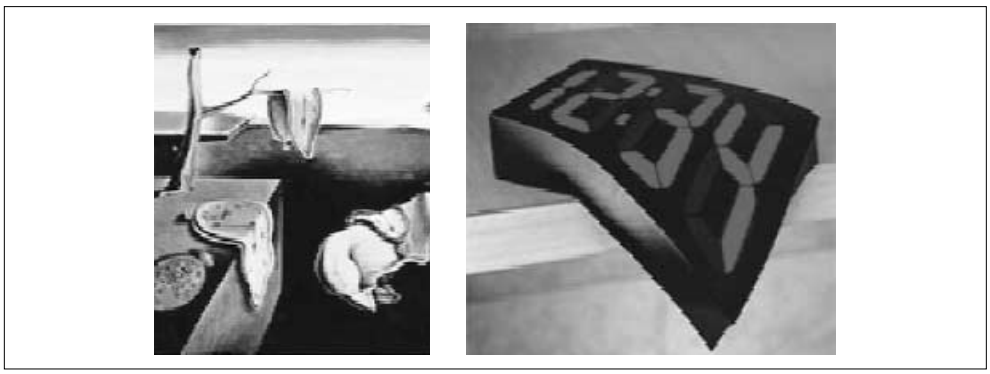


Figura 4. Fragmento de un cuadro de Dalí y fotografía de un reloj diseñado por Ross McBride.

Con este concepto de equivalencia, llamada *homeomorfismo*, idea novedosa a principios del siglo XX, no es fácil asegurar cuándo dos espacios X e Y son o no equivalentes (*homeomorfos*), es decir, si pertenecen o no a la misma clase de equivalencia. Para facilitar el proceso de clasificación se utilizan *invariantes*. Los invariantes son objetos matemáticos (números, grupos, polinomios...) asociados a cada espacio, de manera que tienen el mismo valor para cada uno de los espacios pertenecientes a una misma clase de equivalencia. Por ejemplo, el número de componentes conexa de un espacio es un invariante numérico.

La rama de la topología que estudia los invariantes algebraicos es la topología algebraica, también iniciada por Henri Poincaré. Los invariantes son una herramienta útil para distinguir espacios. Si I es un invariante que asocia a un espacio X un valor distinto al que asocia a un espacio Y, $I(X) \neq I(Y)$, se puede asegurar que X e Y no son espacios equivalentes, pero si el valor del invariante es el mismo para los dos espacios, $I(X) = I(Y)$, no se puede asegurar que se trate de espacios equivalentes (a no ser que el invariante sea un invariante completo).

La bondad de un invariante depende de la dificultad que conlleva su cálculo y de su poder separador. Ambas cualidades suelen estar en relación inversa. Cuanto mayor es el poder separador que tiene un invariante, más difícil suele resultar su cálculo.

El grupo fundamental

El grupo fundamental de un espacio X es un invariante algebraico introducido por Poincaré. En cierto sentido da idea de la existencia de ciertos agujeros en un espacio, aquéllos que se podrían rellenar con un disco. Se llama también *primer grupo de homotopía* o *grupo de Poincaré* y se denota en la actualidad por $\pi_1(X, x_0)$, donde la letra π hace honor a su creador Poincaré y el subíndice 1 obedece a que, generalizando la idea de construcción, se han definido nuevos invariantes conocidos como grupos de homotopía de orden $n > 1$, $\pi_n(X, x_0)$. Estos invariantes también están asociados a agujeros, pero éstos necesitarían bolas de dimensión n para ser rellenados.

Un lazo en un espacio X basado en un punto x_0 es una aplicación continua $\alpha : [0,1] \to X$ con x_0 como punto inicial y final $\alpha(0) = \alpha(1) = x_0$. La orientación en el intervalo unidad induce una orientación en la imagen del lazo. Fijado un punto x_0 en un espacio X, el conjunto de lazos basados en ese punto x_0 se puede distribuir en clases de equivalencia, de manera que dos lazos son equivalentes cuando uno se puede deformar en el otro de manera continua a través de una familia de lazos intermedios basados todos ellos en el mismo punto base. Uno debe pensar en un lazo α como un hilo elástico orientado con los extremos fijos en el punto base y obligado a moverse siempre dentro del espacio en cuestión. Las posiciones que puede adoptar este hilo partiendo de una posición inicial α constituyen lazos equivalentes al primero y pertenecen a la misma clase, $[\alpha]$. Dos lazos, α y β, se pueden multiplicar dando como resultado otro lazo, $\alpha \star \beta$, sin más que considerar uno a continuación del otro, es decir, uniendo el extremo final del hilo que representa al primer lazo α con el extremo inicial del hilo que representa al segundo β. Véase figura 5. Esta operación es compatible con la relación de equivalencia y dota al conjunto de clases de equivalencia de lazos en un espacio X basados en un punto x_0,

$$\pi_1(X, x_0) = \{[\alpha] | \alpha : [0,1] \to X \text{ continua}, \ \alpha(0) = \alpha(1) = x_0\}$$

de una estructura de grupo.

$$\begin{array}{ccc} \pi_1(X, x_0) \times \pi_1(X, x_0) & \rightarrow & \pi_1(X, x_0) \\ ([\alpha], [\beta]) & \rightarrow & [\alpha \star \beta] \end{array}$$

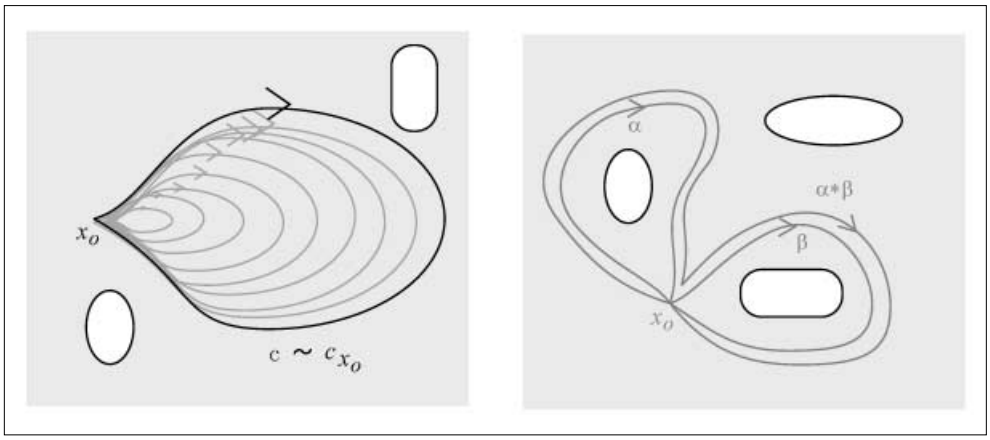


Figura 5. Lazo equivalente al constante y producto de lazos.

El elemento neutro en este grupo es la clase del lazo constante. El elemento inverso, $[\alpha]^{-1}$, es la clase que tiene como representante el lazo opuesto a α, que es el mismo lazo pero recorrido en el sentido inverso. Se observa que un lazo equivalente al camino constante está representado por un hilo elástico que puede recogerse dentro del espacio al punto base, barriendo en este proceso una porción de superficie. En cambio un lazo que no es equivalente al lazo constante, es decir, que representa un elemento no nulo del grupo, no puede contraerse dentro del espacio a x_0. Este lazo abraza un agujero en el espacio X que impide su deformación.

Homología

Un invariante más débil que el grupo fundamental es el primer grupo de homología. Está formado por clases de homología de 1-ciclos (un número finito de lazos orientados libres, posiblemente con repeticiones). Un lazo libre en un espacio X es la imagen de una circunferencia. Podemos imaginar un hilo elástico orientado con sus extremos identificados que se puede mover dentro del espacio. En este caso no existe la restricción de estar sujeto al punto base. La equivalencia se denomina *homología*. Dos 1-ciclos, c_1 y c_2, son homólogos ($c_1 \sim c_2$) si $c_1 - c_2$ bordea una porción de superficie, donde el signo (–) cambia la orientación a los lazos libres que afecta. El producto de dos clases es la clase representada por la suma de sendos 1-ciclos representantes. Esta operación es compati-

ble con la relación de homología y dota al conjunto de clases de homología en un espacio X, $H_1(X)$ de una estructura de grupo.

$$\begin{array}{ccc} H_1(X)\times H_1(X) & \rightarrow & H_1(X) \\ ([c_1], [c_2]) & \rightarrow & [c_1+c_2] \end{array}$$

El elemento neutro en este grupo es la clase de homología del lazo constante. El primer grupo de homología de un espacio coincide con el abelianizado del grupo fundamental. Un 1-ciclo representa al elemento neutro si es un conjunto finito de lazos orientados que bordean conjuntamente una superficie.

LAS ESFERAS

La primera definición de esfera es de origen geométrico. La esfera de dimensión n, S^n es el conjunto de vectores unitarios del espacio euclídeo de dimensión $n+1$, E^{n+1}.

$$S^n = \{(x_1, x_2, \ldots, x_{n+1}) \mid \sum_{i=0}^{n+1} x_i^2 = 1\}$$

En la figura 6 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 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.

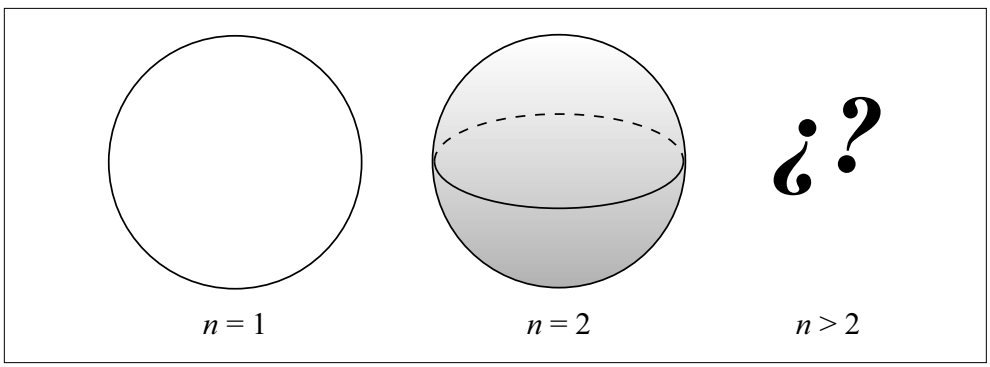


Figura 6. Esferas.

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 $f_x: U_x \to D^n$ 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. Véase figura 7.

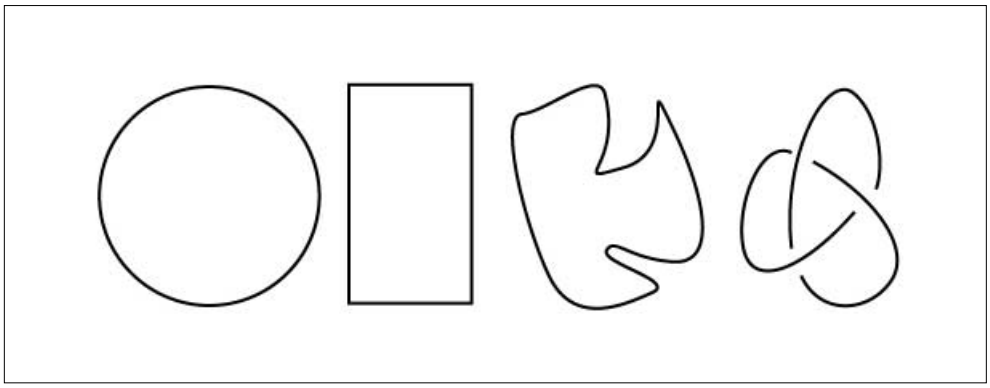


Figura 7. Circunferencias (topológicas).

La superficie esférica es la esfera de dimensión 2. En la figura 8 se dibujan algunas superficies cerradas. 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 cualquiera de los dos invariantes antes citados: el grupo fundamental o el primer grupo de homología.

R1. *La esfera bidimensional es la única superficie cerrada cuyo grupo fundamental es trivial:* $\pi_1(S^2, x_0) = 0$. *Es decir, cada camino simple cerrado se contrae a un punto.*

R2. *La esfera bidimensional es la única superficie cerrada cuyo primer grupo de homología es trivial:* $H_1(S^2) = 0$. *Un camino cerrado en la esfera* S^2 *es el borde de una porción de superficie.*

Podemos decir que la esfera bidimensional es la única superficie cerrada sin agujeros bidimensionales.

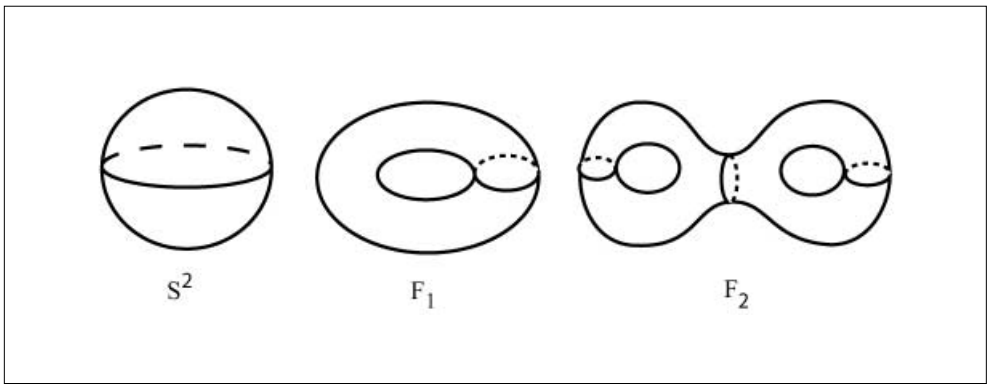


Figura 8. Superficies orientables cerradas.

LA CONJETURA DE POINCARÉ

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, como superficie con homología trivial, **R2** antes citada, escribió que también la esfera S^3 es la única variedad de dimensión 3 en la que toda curva simple cerrada bordea una superficie [12].

Cuatro años más tarde, en 1904, él mismo publicó en el quinto complemento al *Analysis sitûs* [13], 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. Se trata de una variedad en la que cada punto es un dodecaedro inscrito en la esfera unidad S^2. Observamos (figura 9) que un dodecaedro inscrito está determinado por la posición de un vértice (un punto P de S^2) y una arista incidente en ese vértice (un vector unitario en P). Es decir, un punto de $ST(S^2) = SO(3)$, fibrado esférico tangente de la esfera S^2, que es un vector tangente unitario en la esfera S^2, determina un único dodecaedro inscrito. Pero varios puntos del esférico tangente de la esfera S^2 determinan el mismo dodecaedro, exactamente 60 (3 aristas por cada uno de los 20 vértices).

Esta variedad de dodecaedros inscritos, M_d, llamada *variedad dodecaedral de Poincaré*, es de dimensión 3, cociente del esférico tangente de la esfera por un grupo finito, $(SO(3)/I_{60}) \simeq M_d$, donde I_{60} es el grupo de simetrías del icosaedro (=grupo de isometrías del dodecaedro).

La teoría elemental de espacios recubridores proporciona una manera de calcular el grupo fundamental de un espacio M. Se trata de considerar su espacio recubridor uni-

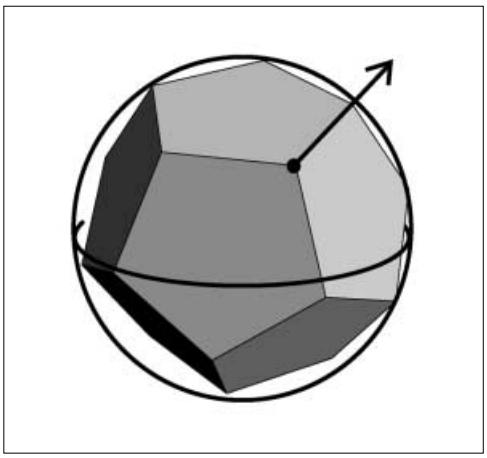


Figura 9. Un punto de la variedad dodecaedral.

versal $p_u : U \to M$, y recordar que el grupo fundamental es isomorfo al grupo de superposiciones que actúa libremente en U dando como cociente M. En el caso de la variedad dodecaedral M_d, es fácil observar que su espacio recubridor universal es la esfera S^3, que es el espacio recubridor universal (de dos hojas) del fibrado esférico tangente de la esfera S^2. Entonces, el grupo fundamental, que coincide con el grupo de superposiciones del espacio recubridor universal

$$\begin{array}{ccc} S^3 & \xrightarrow{p_u} & M_d \simeq (SO(3)/I_{60}) \simeq (S^3/BI_{120}) \,, \\ {\scriptstyle 2:1}\searrow & & \nearrow \\ & SO(3) & \end{array}$$

es el grupo binario icosaedral BI_{120}, grupo perfecto (su abelianizado es trivial) de 120 elementos. El primer grupo de homología de un espacio coincide con el abelianizado del grupo fundamental. Por tanto, el primer grupo de homología de M_d es el abelianizado del grupo binario icosaedral, que es trivial. Esto significa que cualquier curva cerrada en M_d es homóloga a cero, es decir, cualquier curva cerrada es el borde de alguna superficie compacta. Sin embargo, la variedad M_d no es homeomorfa a la esfera S^3 puesto que ambas variedades tienen grupos fundamentales no isomorfos.

Por la dualidad del dodecaedro y del icosaedro, la variedad dodecaedral de Poincaré es también el espacio formado por los icosaedros inscritos en una esfera, por lo que también se denomina *variedad icosaedral de Poincaré*.

La variedad M_d es el resultado de pegar cada par de caras opuestas de un dodecaedro por giro a derecha de $2\pi/10$. También se obtiene como espacio recubridor cíclico de cinco hojas de S^3 ramificado sobre el trébol, y como resultado de hacer cirugía 1 en el trébol.

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 estudio ha sido objeto de muchos topólogos durante todo un siglo. El enunciado preciso de la conjetura de Poincaré es:

CP. *Una variedad tridimensional cerrada con grupo fundamental trivial es homeomorfa a la esfera tridimensional.*

El problema en otras dimensiones

La n-esfera S^n es el conjunto de vectores unitarios del espacio euclídeo $\mathbb{E}^{n+1}$, luego cabe plantearse este problema en cualquier dimensión $n>1$.

El resultado **R1** es la solución positiva a la conjetura en dimensión 2.

En dimensiones superiores a 3 se debe adecuar el enunciado, igual que sucedió en el paso de dimensión 2 a dimensión 3, porque la condición de ser variedad simplemente conexa es una condición insuficiente para caracterizar la esfera en dimensiones más altas. Existen n-variedades cerradas simplemente conexas ($n>3$) que no son homeomorfas a la esfera de dimensión n, por ejemplo la variedad producto $S^2\times S^{n-2}$. La topología algebraica proporciona invariantes algebraicos que generalizan al grupo fundamental. Se trata de los grupos de homotopía π_i cuyos elementos son clases de equivalencia de aplicaciones basadas de esferas basadas (S^i, z_0) en la variedad basada (M, m_0), ($i=1$ es el grupo fundamental). Dos espacios X, Y son del mismo *tipo de homotopía* si tienen todos sus respectivos grupos de homotopía isomorfos: $\pi_i(X, x_0) \cong \pi_i(Y, y_0), \forall i \in \mathbb{N}$. El enunciado de la conjetura en dimensión n es:

CPn: *Toda n-variedad cerrada del tipo de homotopía de la esfera S^n es homeomorfa a la esfera S^n.*

La conjetura se resolvió primero para dimensión $n>4$. A partir de 1960 varios matemáticos probaron por diferentes métodos distintas versiones de la Conjetura de Poincaré CPn. En dimensión menor o igual que 3 es indiferente trabajar con variedades topológicas, combinatorias o diferenciables, pero esto no sucede en dimensión superior. La categoría de variedades utilizada y sus correspondientes métodos es lo que distingue las diversas demostraciones. Citemos a Smale [19], Stallings [20] y Wallace [23] y enunciamos el teorema de Stephen Smale que se refiere al caso diferenciable. El hecho de que estos resultados contribuyeran a que Stephen Smale recibiera la medalla Fields en 1966 da idea de la importancia del tema.

Teorema (Smale). *Toda esfera homotópica diferenciable M^n, $n>4$ es homeomorfa a la esfera S^n. La variedad M^n es el resultado de identificar por su borde dos bolas n-dimensionales mediante un difeomorfismo.*

La prueba en dimensión 4 fue obtenida veinte años más tarde por Michael Freedman [F1]. En el mismo artículo clasificó todas las 4-variedades cerradas y simplemente conexas.

Teorema (Freedman). *Dos 4-variedades simplemente conexas son homeomorfas si, y sólo si, ambas tienen el mismo producto cup en cohomología de dimensión* 2, $\beta : H^2 \times H^2 \to H^4$ *y el mismo invariante de Kirby-Siemenmann* κ.

Este teorema tiene como corolario la Conjetura de Poincaré en dimensión 4, porque si M^4 es una esfera homotópica, su segundo grupo de cohomología es cero $H^2(M^4) = 0$, y $\kappa(M^4) = 0$, como para la esfera S^4, luego el teorema asegura que M^4 y S^4 son homeomorfas. Por éste y otros importantes resultados Michael Freedman recibió también una medalla Fields en 1986.

CLASIFICACIÓN DE 3-VARIEDADES

La conjetura de Poincaré es parte de un problema fundamental en topología: la clasificación de 3-variedades. A semejanza de la clasificación de las superficies, conocida desde el siglo XIX, desearíamos clasificar las 3-variedades dando una lista completa sin repeticiones. Los métodos de trabajo en este campo han sido diversos.

Desde el nacimiento de la topología, se han venido utilizando métodos topológicos y combinatorios que han sido capaces de demostrar importantes resultados y también de clasificar muchas familias de 3-variedades.

Cada 3-variedad M^3 admite una descomposición única (salvo el orden) en suma conexa de variedades primas [6, 7].

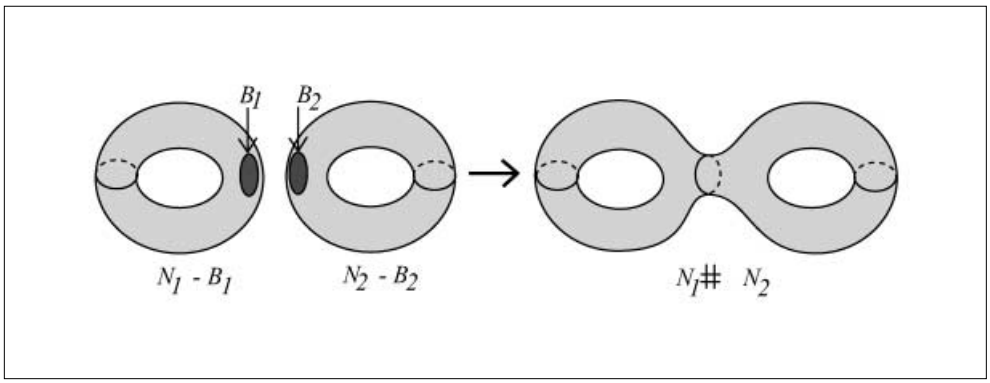


Figura 10. Suma conexa de dos superficies.

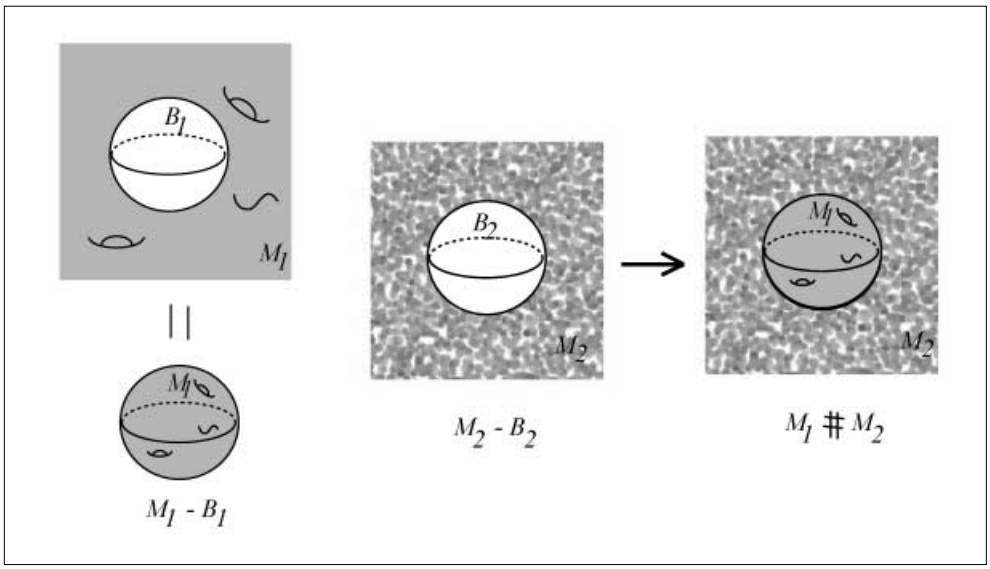


Figura 11. Suma conexa de dos 3-variedades.

$$M^3 = M_1 \# M_2 \# \ldots \# M_r$$

Recordemos que la suma conexa de dos n-variedades N_1 y N_2 conexas se obtiene de la siguiente forma: En cada una de las dos variedades N_i se quita el interior de una bola B_i^n contenida en el interior de la variedad. El resultado en cada caso es una variedad, $N_i \setminus B_i^n$, con una nueva esfera, S_i de dimensión n–1, en su borde. Se identifica $N_1 \setminus B_1^n$ y $N_2 \setminus B_2^n$ a lo largo de esas nuevas componentes del borde mediante un homeomorfismo, y se obtiene la variedad n-dimensional suma conexa de N_1 y N_2. La figura 10 muestra la suma conexa de superficies cerradas y la figura 11 muestra la suma conexa de 3-variedades.

Una n-variedad es prima si no es suma conexa de dos variedades distintas de la esfera S^n. Es claro que para clasificar 3-variedades es suficiente clasificar las 3-variedades primas.

Un proceso estándar para demostrar la conjetura de Poincaré siguiendo métodos topológicos y combinatorios consiste en:

1. Describir un procedimiento que construya todas las 3-variedades.
2. Determinar transformaciones de los datos que no cambian la 3-variedad.
3. Suponer que se trata de una 3-variedad con grupo fundamental trivial y, aplicando las transformaciones anteriores, demostrar que es la esfera.

Existen muchos procedimientos de construcción de las 3-variedades cerradas orientables, casi siempre definidos por analogía con los existentes en superficies. Citamos algunos métodos que construyen todas las 3-variedades:

- Cirugía en un nudo.
- Unión de dos cuerpos sólidos (diagramas de Heegaard).
- Poliedro con caras identificadas.
- Espacio recubridor de la esfera ramificado sobre nudos y enlaces:
  - Número de hojas fijo: 3 hojas.
  - Nudo o enlace fijo: nudos y enlaces universales (4_1, enlace de Borromeo...).

El estudio de la conjetura de Poincaré en cada uno de los procedimientos anteriores ha dado lugar a nuevos enunciados equivalentes de la conjetura y, en ocasiones, a demostraciones de la conjetura... en las que siempre, hasta el momento, se ha encontrado un error o un paso no justificado.

En la década de 1970 las técnicas de trabajo se enriquecen con el uso de estructuras riemannianas en 3-variedades, siempre generalizando a dimensión 3 lo que se conoce en dimensión 2. Es bien conocido que cada superficie cerrada soporta una estructura riemanniana de curvatura constante, aquélla que tiene su espacio recubridor universal (la esfera S^2, el plano euclídeo E^2 o el plano hiperbólico H^2) del que es cociente por la acción de un subgrupo discreto de isometrías. En el último cuarto del siglo XX, los matemáticos R. Riley [16] y W.Thurston [21, 22] demuestran independientemente la existencia de una estructura hiperbólica en el exterior de un nudo en S^3. Éste es el comienzo de una interesante contribución de la geometría al estudio de las 3-variedades.

LA CONJETURA DE GEOMETRIZACIÓN

Éste es un resultado propuesto por W. Thurston y que se conoce con el nombre de *Conjetura de geometrización de 3-variedades*:

CG. *Cada 3-variedad prima M es geométrica o sus piezas simples son geométricas.*

En este enunciado aparecen los conceptos de *pieza simple* y de *variedad geométrica* en el sentido de Thurston, que describo a continuación.

De la misma manera que una 3-variedad cerrada se descompone por un número finito de esferas encajadas en suma conexa de 3-variedades primas, se puede considerar la descomposición de una variedad prima por toros encajados incompresibles. Un toro (F_1) es incompresible si no se puede comprimir, es decir, si ninguna curva simple cerra-

da esencial en F_1 es homótopa a un punto en la variedad. Por tanto, el grupo fundamental del toro incompresible se inyecta en el grupo fundamental de la variedad. En una variedad prima M existe una familia maximal finita (que puede ser vacía) de toros incompresibles. Esta familia maximal se denomina *jerarquía* y divide la variedad en un conjunto finito de 3-variedades abiertas que son las *piezas simples* de M [5, 4].

Una *geometría X en el sentido de Thurston* es una 3-variedad riemanniana homogénea unimodular y simplemente conexa, donde homogénea significa que la situación métrica local es la misma en todos los puntos, es decir, que dados dos puntos de X siempre existe una isometría que lleva un punto al otro. La propiedad de ser unimodular es que X admite un grupo discreto de isometrías con cociente compacto.

Su lista se reduce a las 8 geometrías siguientes, que agrupamos en tres tipos:

1. Curvatura seccional constante: (positiva) esférica S^3, (cero) euclídea E^3, (negativa) hiperbólica H^3.

2. Geometrías producto: $S^2\times\mathbb{R}$, $H^2\times\mathbb{R}$.

3. Productos torcidos: Nil (grupo de Heisenberg, fibrados no triviales por circunferencias sobre un toro), Sol (Fibrados sobre la circunferencia con toros como fibras, y monodromía Anosov), $\widetilde{SL(2,\mathbb{R})}$ (recubridor universal de $SL(2,\mathbb{R})$, esférico tangente del plano hiperbólico). Estas tres últimas geometrías son grupos de Lie simplemente conexos con una métrica invariante a izquierda.

Una interesante descripción de estas geometrías se encuentra en el artículo de Peter Scott [17]. Obsérvese que de estas geometrías sólo es compacta S^3, la de curvatura seccional positiva.

Una 3-variedad compacta es *geométrica* si su interior es X/Γ y tiene volumen finito, donde X es una geometría y Γ es un subgrupo discreto de isometrías de X que actúa libremente. Se dice que está modelada por la geometría X.

Por otra parte las variedades compactas que soportan una acción del grupo de la circunferencia S^1 son las llamadas *variedades de Seifert*, que fueron definidas y clasificadas totalmente por H. Seifert en un célebre artículo [18] publicado en 1933. Las variedades geométricas modeladas por S^3, E^3, $S^2\times\mathbb{R}$, $H^2\times\mathbb{R}$, Nil y $\widetilde{SL(2,\mathbb{R})}$ son variedades de Seifert.

Los cocientes con volumen finito de las 8 geometrías citadas son piezas simples de 3-variedades cerradas. Lo que significa que las 8 geometrías son necesarias para geometrizar las 3-variedades. Lo que realmente dice la CG es que esas 8 geometrías son suficientes para geometrizar todas las piezas simples de todas las 3-variedades primas.

Existen dos maneras de unir 3-variedades geométricas para obtener variedades compuestas. La primera (*suma conexa*) es quitar una bola en cada una de ellas e identificar las dos esferas que han quedado como borde. La segunda, que une variedades con borde, es identificar sendos componentes del borde por un homeomorfismo. Estas uniones es lo que ahora se conoce como *agujeros de gusano* que conectan piezas geo-

métricas. En la suma conexa se trata de un agujero de gusano con sección esférica, y en el segundo caso un agujero de gusano que tiene como sección la superficie que sirve de pegado.

Las *variedades de grafo* son el resultado de identificar variedades de Seifert por toros contenidos en su borde.

Teorema. *La conjetura de geometrización implica la conjetura de Poincaré.*

En efecto. Supongamos que la conjetura de geometrización es cierta y sea M^3 una 3-variedad cerrada simplemente conexa. Entonces M^3 es suma conexa de un número finito de 3-variedades primas

$$M^3 = M_1 \# M_2 \# \ldots \# M_r ,$$

donde cada sumando M_i es una 3-variedad cerrada y simplemente conexa. Como el grupo fundamental de M_i es trivial, la 3-variedad prima M_i no contiene toros incompresibles. Entonces, la conjetura de geometrización implica que M_i es geométrica además de compacta, cerrada y simplemente conexa, luego es la esfera S^3. Por tanto, cada sumando de M^3 es S^3, es decir

$$M^3 = S^3 \# S^3 \# \ldots \# S^3 \simeq S^3.$$

Se deduce también fácilmente de la conjetura de geometrización que una 3-variedad cerrada tiene grupo fundamental finito si, y sólo si, tiene una métrica de curvatura constante positiva.

Otro enunciado de la conjetura de geometrización es el siguiente:

Cada 3-variedad cerrada prima es geométrica, o existe una familia de toros disjuntos incompresibles que la divide en dos partes, la gruesa y la fina. La gruesa es unión de variedades hiperbólicas, y la fina es una variedad de grafo.

Hacia una solución usando el flujo de Ricci

En los últimos años, se ha introducido en el estudio de las 3-variedades un método que deforma estructuras riemannianas.

El *flujo de Ricci* fue ideado por Hamilton [2] para variar de manera diferenciable la métrica en una n-variedad tendiendo hacia una estructura más homogénea. Su definición es la siguiente.

Definición. Sea M una n-variedad cerrada. La familia de estructuras riemannianas diferenciables,

$$\{g(t) \mid t \in [0,\mathrm{T})\},$$

es flujo de Ricci si satisface

$$g'(t) = -2\mathrm{Ric}\,(g(t)),$$

donde $\mathrm{Ric}\,(g(t))$ es el tensor de Ricci de la métrica $(g(t)$.

En el mismo artículo [2], Hamilton demostró el siguiente importante resultado.

Teorema. *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 este contexto, tomando como punto de partida cualquier métrica riemanniana en una 3-variedad, el flujo de Ricci debería producir una familia uniparamétrica de métricas riemannianas que converge a una de las métricas consideradas en la conjetura de Thurston.

Pero en el estudio de tal flujo pueden presentarse singularidades. Otro problema difícil es estudiar la existencia y naturaleza de la variedad riemanniana límite.

Hamilton no resolvió totalmente el problema pero realizó grandes avances. En [3], desarrolló este programa para demostrar la conjetura de geometrización, consiguiendo un éxito parcial, puesto que para evitar singularidades del flujo necesitaba imponer condiciones a las métricas y al flujo. También consiguió analizar la variedad límite. Enunciamos a continuación alguno de sus resultados.

Teorema. [3] *Si el flujo de Ricci en una variedad riemanniana cerrada M^3 tiene solución $\forall t > 0$ y el tensor curvatura de la solución normalizada tiene norma acotada, entonces la conjetura de geometrización es cierta para M^3.*

La variedad M^3 con la métrica $(t \to \infty)$ se descompone por toros incompresibles en $M_{gruesa} \cup M_{fina}$, donde M_{gruesa} es unión de variedades hiperbólicas completas de volumen finito y M_{fina} es variedad de grafo.

Por sus resultados acerca del flujo de Ricci en variedades Richard Hamilton recibió en noviembre de 2003 un premio de investigación del Instituo Clay de Matemáticas.

En los últimos años, el matemático ruso Perelman ha desarrollado técnicas para solventar las dificultades del programa de Hamilton. Ha definido un concepto de *flujo de Ricci con cirugías* que incorpora la utilización de cirugías controladas para resolver las singularidades. Hay un tipo de singularidades que se pueden producir en regiones donde la curvatura crece, cuando colapsan 2-esferas, pero antes de ocurrir el colapso la

solución consiste en realizar cirugía: cortar por la esfera en cuestión y pegar sendas bolas a lo largo de las nuevas esferas en el borde que se han producido al cortar. En la nueva variedad, posiblemente disconexa, se continúa considerando el flujo de Ricci. Este proceso necesita utilizar una serie de delicados parámetros para normalizar el flujo de partida, y para decidir los momentos en que se realiza la cirugía. Otro punto crucial es el estudio del límite, puesto que quizá puedan existir infinitas singularidades. El trabajo de Perelman sobre el tema está contenido en tres artículos, [9, 10, 11] 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. John Morgan ha escrito recientemente un clarificador artículo [8] sobre el programa de Hamilton y las aportaciones de Perelman, a quien atribuye un tremendo avance en el conocimiento del flujo de Ricci en 3-variedades.

También son muchos los matemáticos que esperan cautelosamente la publicación en una revista científica de los resultados de Perelman, antes de asegurar que el problema está resuelto. 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.

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] M. Freedman, «The topology of four-dimensional manifolds», *J. Differential Geom.* 17 (1982), núm. 3, p. 357-453.

[2] R.S. Hamilton, «Three-manifolds with positive Ricci curvature», *J. Differential Geom.* 17 (1982), núm. 2, p. 255-306.

[3]— «Non-singular solutions of the Ricci flow on three-manifolds», *Comm. Anal. Geom.* 7 (1999), núm. 4, p. 695-729.

[4] W. Jaco y P.B. Shalen, «Seifert fibered spaces in 3-manifolds», *Geometric topology* (Proc. Georgia Topology Conf., Athens, Ga., 1977), Academic Press, New York, 1979, p. 91-99.

[5] K. Johannson, «Homotopy equivalences of 3-manifolds with boundaries», *Lecture Notes in Mathematics*, vol. 761, Springer, Berlin, 1979.

[6] H. Kneser, «Geschlossene flächen in dreidimensionalen mannigfaltigkeiten», *Jahresber. Deutsch. Math.* 38 (1929), p. 248-260.

[7] J. Milnor, «A unique decomposition theorem for 3-manifolds», *Amer. J. Math.* 84 (1962), p. 1-7.

[8] J.W. MORGAN, «Recent progress on the Poincaré conjecture and the classification of 3-manifolds», *Bull. Amer. Math. Soc.* (N.S.) 42 (2005), núm. 1, p. 57-78.
[9] G. PERELMAN, «Finite extintion time for the solutions to the Ricci flow on certain three-manifolds», arXiv.math.DG/0307245 (17 de julio, 2003).
[10]— «Ricci flow with surgery on three-manifolds», arXiv.math.DG/0303109 (10 de marzo de 2003).
[11]— «The entropy formula for the Ricci flow and its geometric applications», arXiv.math.DG/0211159 (11 de noviembre de 2002).
[12] H. POINCARÉ, «Second complement à l'analysis sitûs», *Proc. London Math. Soc.* 32 (1900), núm. supl., p. 277-308.
[13]— «Cinquième complement à l'analysis sitûs», Rendiconti Circolo mat. Palermo 18 (1904), p. 45-110.
[14]— *Oeuvres de Henri Poincaré*, t. I-XI, Gauthier-Villars, París, 1916-1956.
[15]— «Analyse de ses travaux scientifiques», *Acta mathematica* 38 (1921), p. 36-135.
[16] R. RILEY, «A quadratic parabolic group», *Math. Proc. Cambridge Philos. Soc.* 77 (1975), p. 281-288.
[17] P. SCOTT, «The geometries of 3-manifolds», *Bull. London Math. Soc.* 15 (1983), núm. 5, p. 401-487.
[18] H. SEIFERT, «Topology of 3-manifolds fibered spaces», *Acta. Math.* 60 (1933), p. 147-288.
[19] S. SMALE, «Generalized Poincaré's conjecture in dimensions greater than four», *Ann. of Math.* (2) 74 (1961), p. 391-406.
[20] J.R. Stallings, «Polyhedral homotopy-spheres», *Bull. Amer. Math. Soc.* 66 (1960), p. 485-488.
[21] W.P. THURSTON, «Three-dimensional manifolds, Kleinian groups and hyperbolic geometry», *Bull. Amer. Math. Soc.* (N.S.) 6 (1982), núm. 3, p. 357-381.
[22]— «*Three-dimensional geometry and topology*. Vol. 1, Princeton Mathematical Series, vol. 35, Princeton University Press, Princeton, NJ, 1997, Editat per Silvio Levy.
[23] A.H. WALLACE, «Modifications and cobounding manifolds», Canad. J. Math. 12 (1960), p. 503-528.

Teoria Quàntica de Yang-Mills

Ignasi Mundet i Riera

1. ENUNCIAT DEL PROBLEMA I INTRODUCCIÓ

Un dels set *problemes del mil·lenni* proposats pel Clay Mathematics Institute l'any 2000 té el següent enunciat:

> Demostreu que per a tot grup de Lie simple i compacte G la teoria de Yang-Mills sobre $\mathbb{R}^4$ amb grup d'estructura G es pot quantitzar i té un *mass gap* $\Delta > 0$.

Es tracta d'un problema propi de la física matemàtica. La seva solució seria, sense cap dubte, un gran avenç de la física, ja que permetria entendre aspectes crucials de la teoria quàntica de camps que avui dia estan envoltats de misteri. Pot sorprendre, doncs, que un problema d'aquest tipus aparegui en una llista de problemes de matemàtiques. Ara bé, una ullada a la història permet concloure que la gran majoria de conceptes i resultats crucials a les matemàtiques s'han obtingut gràcies a la interacció amb les ciències naturals, especialment la física. Vegem-ne alguns exemples.

— La mecànica clàssica va ser una de les principals motivacions que tenien Leibniz i Newton per desenvolupar el càlcul diferencial (per exemple, per donar sentit a la noció de velocitat); posteriorment, els treballs de Hamilton sobre mecànica varen suposar el naixement de la geometria simplèctica, mentre que la geometria de contacte va néixer de l'òptica.

— Estudiant la propagació de la calor, Fourier va crear la seva teoria sobre funcions periòdiques; també des de bon principi l'estudi de l'operador laplacià va estar motivat per la seva rellevància en la modelització de fenòmens físics, com la mateixa propagació de la calor.

— Per formalitzar la mecànica quàntica es van desenvolupar intensament l'anàlisi funcional i la representació de grups.

— La teoria general de la relativitat d'Einstein va suposar un impuls fortíssim en el desenvolupament de la geometria riemanniana.

S'hi podrien afegir molts d'altres exemples. Amb aquesta perspectiva, una de les grans motivacions del problema de quantitzar Yang-Mills és que la seva solució suposarà necessàriament descobrir nous conceptes matemàtics que, amb tota seguretat, condui-

ran a nous resultats profunds i il·luminaran resultats que ja es coneixen però que encara no és comprenen gaire bé. Es podria donar una llarga llista de resultats matemàtics conjecturats gràcies a les intuicions i nocions de la teoria quàntica de camps que posteriorment han estat demostrats rigorosament pels matemàtics, i és raonable pensar que quan aquesta teoria física rebi una fonamentació matemàtica rigorosa aquests resultats i les relacions entre ells s'entendran d'una manera més profunda.

L'objectiu principal d'aquestes pàgines és explicar informalment alguns exemples de resultats matemàtics obtinguts gràcies a la teoria quàntica de camps. Paral·lelament, mirarem d'explicar, molt breument, els conceptes que apareixen a l'enunciat del problema del mil·lenni. El lector trobarà bastants articles sobre el problema de quantitzar la teoria de Yang-Mills. A aquells que tenen una certa formació matemàtica o de física teòrica, els recomanem llegir l'article de Jaffe i Witten [JW] amb l'enunciat oficial del problema, així com les notes de Douglas [Dou]. Una altra referència excel·lent és [W3]. Finalment, per a una introducció a un nivell potser més assequible i decididament més orientada cap a la física, es pot consultar l'article d'Asorey [As].

Agraïments. Voldria agrair a L. Álvarez Cónsul, T. Gómez i A. Ibort tantes i tan interessants converses sobre matemàtiques i física.

2. FORCES FONAMENTALS

Comencem recordant quines són les que avui dia es consideren forces fonamentals de la natura.

— L'electromagnetisme. *És responsable dels fenòmens electromagnètics: electricitat, magnetisme, llum, etc., així com dels fenòmens químics.*

— La força feble. *Dóna lloc a la radioactivitat: el decaïment β o desintegració de protons.*

— La força forta. *És la força nuclear, que manté units els quarks que formen protons i neutrons, i uneix els protons i neutrons que formen els nuclis dels àtoms.*

— La gravetat. *Fa caure les pomes de la pomera.*

El *model estàndard* (del qual direm quatre paraules més endavant) té en compte les tres primeres forces, i està basat enterament en la teoria de Yang-Mills. No se sap com combinar-lo (a nivell quàntic) amb la gravetat.

3. FIBRATS VECTORIALS I CONNEXIONS

En aquesta secció, farem un repàs a tota velocitat de les nocions geomètriques de fibrat, connexió i curvatura. Aquestes nocions, que se solen enquadrar dins de les teories *gauge* (*gauge* vol dir «calibració» en anglès), apareixeran tant en la formulació de la teoria de Yang-Mills com en la majoria de les aplicacions geomètriques d'aquesta teoria.

3.1. Fibrats vectorials

La idea intuitiva de què és un fibrat vectorial sobre un espai topològic M és la d'una família d'espais vectorials parametritzats per M. Més exactament, es defineix un fibrat vectorial com el resultat d'enganxar una col·lecció de fibrats trivials sobre oberts de M. Per exemple, per construir un fibrat vectorial real prenem un recobriment per oberts $M = \bigcup_i U_i$ i definim

$$E = \left(\bigcup_i U_i \times \mathbb{R}^k \right) / \sim,$$

on per cada $x \in U_i \cap U_j$ s'identifiquen

$$U_i \times \mathbb{R}^k \ni (x, v) \sim (x, \rho_{ij}(v)) \in U_j \times \mathbb{R}^k$$

usant aplicacions $\rho_{ij} : U_i \cap U_j \to GL(k, \mathbb{R})$. Aquestes aplicacions han de satisfer $\rho_{jk} \circ \rho_{ij} = \rho_{ik}$ per cada terna d'índexs i, j, k i a més ρ_{ii} ha de ser la identitat. Si substituïm el cos $\mathbb{R}$ per $\mathbb{C}$, obtenim un fibrat vectorial complex.

3.1.1. En funció de la categoria on treballem hi ha diverses nocions de fibrat vectorial:

— Si M és un espai topològic i les aplicacions $\{\rho_{ij}\}$ són contínues, aleshores E és un fibrat vectorial topològic.

— Si M és una varietat diferencial i les aplicacions $\{\rho_{ij}\}$ són diferenciables, aleshores E és un fibrat vectorial diferencial.

— Finalment, si M és una varietat complexa i $\{\rho_{ij}\}$ són holomorfes, aleshores E és un fibrat vectorial holomorf (aquesta situació només es pot donar quan considerem fibrats complexos).

3.1.2. Exemples de fibrats vectorials:

— $M \times \mathbb{C}^n$ és el fibrat trivial complex de rang n sobre M.

— La banda de Möbius és un fibrat real de rang 1 sobre el cercle i *no és trivial*; si ho fos, seria orientable.

— El fibrat tangent $T \to \mathbb{C}P^1$ és un fibrat complex holomorf de rang 1 que no és trivial com a fibrat topològic (teorema de la bola peluda).

— Els fibrats $T \oplus T^*$ i $\mathbb{C}P^1 \times \mathbb{C}^2$ sobre $\mathbb{C}P^1$ són isomorfs com a fibrats diferencials però no com a fibrats holomorfs.

3.2. Connexions

Sigui $E \to M$ un fibrat diferencial complex. Una secció (diferenciable) σ de E és una funció que assigna a cada $p \in M$ un vector $\sigma(p) \in E_p$ i que, vista com una aplicació entre varietats diferencials $M \to E$, és una aplicació diferenciable. L'espai de seccions de E és un espai vectorial sobre $\mathbb{C}$ i es denota per $C^\infty(E)$.

Una connexió a E és una aplicació $\mathbb{C}$-lineal

$$d_A : C^\infty(E) \to C^\infty(T^*M \otimes E)$$

que satisfà la regla de Leibniz: si $\sigma \in C^\infty(E)$ és una secció i $f : M \to \mathbb{C}$ una funció diferenciable, aleshores

$$d_A(f\sigma) = df \cdot \sigma + f\, d_A\sigma$$

Intuïtivament, una connexió ens dóna una manera d'identificar fibres de E *infinitament properes*. Tot fibrat té infinites connexions (això es demostra usant una partició de la unitat) però *a priori* no n'hi ha cap de canònica: el conjunt de connexions és un espai afí sobre $C^\infty(T^*M \otimes \mathrm{End}\, E)$, que és un espai vectorial de dimensió infinita.

3.2.1. Exemples de connexions

Sigui E el fibrat trivial $M \times \mathbb{C}^n$. Una connexió a E és equivalent a un element

$$\alpha \in C^\infty(T^*M \otimes \mathrm{End}\, \mathbb{C}^n).$$

Un tal α es correspon amb la connexió d_A que envia una secció $\sigma : \mathrm{M} \to \mathbb{C}^n$ de E a

$$d_A\sigma = d\sigma + \alpha\sigma$$

A més, totes les connexions a E són d'aquesta mena.

Com que, per definició, els fibrats vectorials són localment trivials, tota connexió localment es pot escriure com $d_A = d + \alpha$.

3.3. Connexions hermítiques i estructures holomorfes

Sigui $E \to M$ un fibrat complex diferencial. Direm que E és un fibrat hermític si cada fibra de E té una estructura d'espai vectorial hermític, de manera que si u i v són seccions diferenciables de E la funció

$$\langle u, v \rangle : M \to \mathbb{C}$$

que associa a cada punt $p \in M$ el producte hermític $\langle u(p), v(p) \rangle$ és diferenciable. Llavors direm que una connexió d_A és hermítica si se satisfà

$$d\langle u, v \rangle = \langle d_A u, v \rangle + \langle u,\ d_A v \rangle.$$

Hi ha una correspondència bijectiva

$$\left\{ \begin{array}{c} \text{connexions hermítiques} \\ \text{al fibrat hermític } E \end{array} \right\} \quad \Leftrightarrow \quad \left\{ \begin{array}{c} \text{estructures holomorfes} \\ \text{al fibrat diferencial } E \end{array} \right\}.$$

3.4. Curvatura i transport paral·lel

Si $\pi : E \to M$ és un fibrat vectorial dotat d'una connexió d_A, la curvatura de d_A és una 2-forma

$$F_A \in C^\infty(\Lambda^2 T^* M \otimes \operatorname{End} E)$$

definida per la connexió $d_A = d + \alpha$ al fibrat trivial $M \times \mathbb{C}^n$ com

$$F_A = d\alpha + \tfrac{1}{2}\,[\alpha, \alpha];$$

ara bé, com que qualsevol connexió es pot escriure localment com $d + \alpha$, la fórmula anterior ens permet definir la curvatura de qualsevol connexió.

Una manera d'entendre intuïtivament la noció de curvatura és relacionant-la amb el transport paral·lel. Per qualsevol camí diferenciable a M el transport paral·lel dóna un isomorfisme d'espais vectorials entre la fibra al punt inicial del camí i la fibra al punt final. Concretament, donat un camí $\gamma : [0,1] \to M$ i un vector $v \in E_{\gamma(0)}$, hi ha un únic camí $\Gamma : [0,1] \to E$ que aixeca γ (és a dir, $\pi \circ \Gamma = \gamma$), té derivada nul·la ($d_A \Gamma = 0$) i comença a v ($\Gamma(0) = v$). Llavors es defineix la imatge de v per transport paral·lel al llarg de γ com el vector $\Gamma(1) \in E_{\gamma(1)}$.

La curvatura de d_A mesura com canvia el transport paral·lel al llarg d'un camí γ quan deformem γ lleugerament preservant els punts inicial i final. Hom comprova que

per entendre aquesta variació és suficient conèixer el transport paral·lel al llarg de petits camins tancats, que és precisament el que ens dóna la curvatura: si $p \in M$ és un punt i u, v són vectors ortogonals de norma 1 dins l'espai tangent T_pM, l'endomorfisme $F_A(p)(u \wedge v) \in \mathrm{End}\, E_p$ mesura el transport paral·lel al llarg d'un camí que segueix el perímetre d'un quadrat amb vèrtex p i costats paral·lels a ϵu, ϵv, on $\epsilon \to 0$.

3.5. El grup gauge

Una transformació *gauge* d'un fibrat $E \to M$ associa a cada $p \in M$ un isomorfisme $E_p \to E_p$. Dit d'una altra manera, una transformació *gauge* és una secció del fibrat d'endomorfismes $\mathrm{End}\, E = E^* \otimes E$ que admet una d'inversa. Si E té una estructura hermítica, es demana a les transformacions *gauge* que la preservin.

El conjunt de transformacions *gauge* de E és el grup *gauge* $\mathcal{G}$ de E. El grup $\mathcal{G}$ actua a l'espai de connexions $\mathcal{A}$ de E per *pull back*: si d_A és una connexió i g és una transformació *gauge* la connexió $g^* d_A$ actua sobre una secció $\sigma \in C^\infty(E)$ com

$$g^* d_A \sigma = g \circ d_A \circ g^{-1} \sigma .$$

Hom comprova que si en una trivialització local s'escriu $d_A = d + \alpha$, aleshores

$$g^* d_A = d + g\alpha g^{-1} - g^{-1} d_A g .$$

D'aquí es pot deduir que la curvatura satisfà:

$$F_{g^*A} = g F_A g^{-1} .$$

4. LES EQUACIONS DE MAXWELL

4.1. Electromagnetisme

Les equacions de Maxwell relacionen el camp elèctric E, el camp magnètic B, la càrrega elèctrica ρ_{em} i el corrent elèctric j_{em}. Si ens mirem E i B com a camps vectorials a l'espai $\mathbb{R}^3$ i ρ_{em} i j_{em} com a funcions a $\mathbb{R}^3$, les equacions prenen la forma

$$\begin{aligned} \nabla \cdot E &= \rho_{em} && \text{llei de Gauss} \\ \nabla \times E &= -\partial B / \partial t && \text{llei de Faraday-Lenz} \\ \nabla \cdot B &= 0 && \text{absència de càrrega magnètica} \\ \nabla \times B &= j_{em} + \partial E / \partial t && \text{llei d'Ampère-Maxwell,} \end{aligned}$$

en les quals, com és habitual al càlcul diferencial en dimensió tres, el símbol ∇ denota el vector $(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$ i els productes escalar div $E = \nabla \cdot E$ i tensorial rot $E = \nabla \times E$ es defineixen aplicant formalment les nocions corresponents per parelles de camps vectorials.

La segona i la tercera equació de Maxwell impliquen que existeix un camp vectorial A (anomenat *potencial vector*) i una funció V (anomenada *potencial escalar*) tals que

$$B = \nabla \times A, \qquad E = -\nabla V - \partial A / \partial t.$$

Els potencials A no són únics: si prenem una funció χ i substituïm

$$A \mapsto A + \nabla \chi, \qquad V \mapsto V - \partial \chi / \partial t$$

se segueixen satisfent les equacions anteriors. L'observació clau és que la parella (V, A) es comporta com els coeficients d'una connexió en un fibrat de rang 1 quan apliquem la transformació *gauge* e^{χ}.

Per concretar d'una manera senzilla aquesta darrera observació, convé que canviem lleugerament el nostre punt de vista. D'una banda, hem d'agrupar l'espai i el temps considerant l'espai de Minkowski:

$$\mathbb{R}^4 = \mathbb{R} \times \mathbb{R}^3 = \text{temps} \times \text{espai},$$

amb la mètrica de Lorentz $g = dt^2 - dx^2 - dy^2 - dz^2$. D'altra banda, cal pensar E com una 1-forma i B com a una 2-forma. Aleshores

$$F := -dt \wedge E + B$$

és la curvatura de la connexió $d + (V, A)$, i les equacions de Maxwell esdevenen:

$$dF = 0 \qquad d^* F = (\rho_{em}, j_{em})$$

La primera equació prové de la segona i tercera equació de Maxwell i és la fórmula de Bianchi, que sempre se satisfà (cosa que no vol dir que es puguin ignorar la segona i tercera equació de Maxwell, ja que les hem usades prèviament per arribar a aquesta formulació). De l'operador d^* en parlem tot seguit.

4.2. Teoria de Hodge

A la segona equació hem usat l'operador d^*, l'adjunt formal de l'operador d, que es pot definir en tota generalitat: si (M, g) és una varietat diferenciable amb una mètrica (no necessàriament definida), aleshores l'operador

$$d^*\colon \Omega^k(M) \to \Omega^{k-1}(M)$$

està caracteritzat per la propietat que, per a qualsevol $\alpha \in \Omega^k(M)$ i $\beta \in \Omega^{k-1}(M)$, se satisfà la igualtat

$$\int_M \langle d\beta, \alpha\rangle = \int_M \langle \beta, d^*\alpha\rangle$$

L'estudi d'aquest operador d^* en el context riemannià, però motivat fortament pel seu paper en l'electromagnetisme, va portar Hodge a descobrir i demostrar (llevat d'alguns detalls tècnics completats per Weyl) el següent teorema.

Teorema. (Hodge) *Si g és definida positiva i M és compacta, aleshores*

$$\{\alpha \in \Omega^k(M) \mid d\alpha = 0 = d^*\alpha\} \simeq H^k(\mathrm{M};\mathbb{R}).$$

Aquest teorema s'ha d'entendre des de la perspectiva que aporta el teorema de Rham, que afirma que la cohomologia del complex $(\Omega^*(M), d)$ és isomorfa a la cohomologia singular de M a coeficients reals. El que ens diu llavors el teorema de Hodge és que hi ha una manera canònica de triar dins cada classe de cohomologia un representant: l'única forma que a part de ser tancada és *cotancada* (és a dir, tal que l'operador d^* s'hi anul·la). Aquestes formes s'anomenen *harmòniques*.

El teorema de Hodge es pot veure com una de les primeres aplicacions a la geometria de les idees de la teoria de Yang-Mills, concretament de la teoria abeliana de rang 1 (és a dir, l'electromagnetisme).

Aquest teorema ha jugat un paper fonamental en el desenvolupament de la geometria de tota la segona meitat del segle XX (les seves implicacions en la geometria algebraica, per exemple, són importantíssimes). No és estrany, doncs, que les aplicacions obtingudes de la teoria de Yang-Mills no abeliana (que s'han donat unes quantes dècades més tard) hagin estat també extraordinàriament fructíferes.

4.3. Teoria de la relativitat especial

Una de les conseqüències més importants a la física de les equacions de Maxwell va ser el descobriment de la teoria de la relativitat especial. El primer principi de relativitat enunciat a la física prové de Galileu.

> Principi de relativitat de ***Galileu***: *dos observadors que es mouen a velocitat constant l'un en relació amb l'altre obtindran els mateixos resultats en qualsevol experiment mecànic*.

Dit en llenguatge matemàtic modern, les lleis de la física són invariants sota transformacions afins de l'espai temps (que identifiquem amb $\mathbb{R}^4$ —però de moment no hi prenem cap mètrica de Lorentz). Aquest principi s'adiu amb la intuició que tenim de les lleis físiques a partir de les observacions més elementals: per exemple, segons la mecànica clàssica una partícula que no estigui accelerada es mou en línia recta, i efectivament una transformació afí de $\mathbb{R}^4$ envia línies rectes a línies rectes.

De bon principi ja es va veure que les equacions de Maxwell no satisfan el principi de relativitat de Galileu, és a dir, no són invariants per afinitats. L'aportació de Lorentz va ser entendre quin és el grup de simetries de les equacions de Maxwell.

> **Lorentz**: *El grup de simetries de les equacions de Maxwell és el grup de Poincaré, generat per les transformacions de Lorentz* $SO(1, 3)$ *i les translacions.*

Òbviament, Lorentz no va anomenar $SO(1, 3)$ (el grup de transformacions lineals de $\mathbb{R}^4$ que preserven la mètrica de Lorentz) d'aquesta manera, i ni tan sols el va denotar amb aquest símbol.

L'any 1905 (és a dir, fa cent anys) Einstein va formular el principi de relativitat especial.

> Principi de relativitat d'**Einstein** (1905): *Si en dos sistemes de referència les lleis de la mecànica són les mateixes, aleshores també les lleis de l'electrodinàmica i l'òptica hi coincideixen.*

Cosa que, combinada amb l'observació de Lorentz, implica que les simetres de la mecànica no són les afinitats de $\mathbb{R}^4$, sinó les del grup de Poincaré. Proposant aquest principi, Einstein va iniciar una de les revolucions més importants a la física del segle XX. És ben sabut que la teoria de la relativitat té implicacions que semblen contradictòries amb la nostra experiència quotidiana (per exemple, que tant la massa d'un objecte com el pas del temps depenguin de la velocitat). Tot i això, la teoria de la relativitat inclou, en un cert sentit, la mecànica clàssica, com una bona aproximació a velocitats petites.

5. FÍSICA QUÀNTICA

5.1. Mecànica quàntica

La segona gran revolució en la física del segle XX va ser la introducció de la mecànica quàntica. La mecànica quàntica és una *correcció* de la mecànica clàssica, que descriu sistemes amb un nombre finit de graus de llibertat a escales subatòmiques. Mentre que a la mecànica clàssica (en el formalisme hamiltonià) es descriuen els estats del sistema amb els punts d'una varietat simplèctica, a la mecànica quàntica els estats es corresponen amb els subespais de dimensió 1 d'un espai de Hilbert H. I mentre que clàssicament els observables són funcions definides a la varietat dels estats, en la mecànica quàntica els observables són operadors autoadjunts a H i en general no commuten, sinó que es tenen fórmules del tipus

$$[\hat{p}_i, \hat{x}_j] = -i\hbar\delta_{ij},$$

on $\hbar$ és la constant de Plank, $\hat{p}_i$ és la coordenada i-èsima del moment i $\hat{x}_j$ és la coordenada j-èsima de la posició. Aquesta falta de commutativitat és el que fa que la mecànica quàntica funcioni d'una manera radicalment diferent de la mecànica clàssica, ja que clàssicament els observables commuten. De fet, per obtenir la mecànica clàssica com un límit de la mecànica quàntica cal fer tendir $\hbar$ cap a zero.

En l'estat $\psi \in H$, el valor esperat d'un observable $\hat{X}$ es defineix com

$$\langle \psi \,|\hat{X}|\, \psi \rangle = \langle \psi \hat{X} \psi \rangle.$$

Mentre que en la mecànica clàssica el resultat d'una observació està completament determinat per l'estat en què es troba el sistema (ja que no és més que el valor que pren l'observable al punt donat per l'estat), en la mecànica quàntica (almenys en la seva interpretació més habitual) el resultat d'un experiment és aleatori i segueix una llei de probabilitat. Des d'aquest punt de vista, el valor esperat es correspon amb la noció de la teoria de probabilitats.

Per acabar, l'evolució d'un sistema quàntic ve descrita per l'equació de Schrödinger:

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{\mathcal{H}}\psi\,,$$

on $\hat{\mathcal{H}}$ és l'operador hamiltonià. L'equació de Heisenberg, equivalent a la de Schrödinger, diu que l'evolució d'un observable $\widehat{X_t}$ al llarg del temps ve donada per:

$$i\hbar \frac{\partial \widehat{X}_t}{\partial t} = [\widehat{X}_t, \hat{\mathcal{H}}],$$

Aquestes dues formulacions són anàlogues a les que es tenen en mecànica clàssica: si (X, ω) és la varietat d'estats, i $h : X \to \mathbb{R}$ és el hamiltonià, podem descriure l'evolució del sistema des de dos punts de vista. Primerament, com el flux generat pel camp Z definit per la propietat $dh = \imath_Z \omega$ (aquest és l'anàleg de l'equació de Schrödinger). O, anàlogament a l'equació de Heisenberg, si $f : X \to \mathbb{R}$ és un observable, la variació de f quan fem evolucionar el sistema ve descrita per $\partial f / \partial t = \{f, h\}$, on $\{\cdot,\cdot\}$ denota el claudàtor de Poisson.

5.2. Teoria quàntica de camps

Quantitzar una teoria de camps com l'electromagnetisme (que té infinits graus de llibertat) presenta dificultats molt més grans que la mecànica quàntica. Part del problema del mil·lenni consisteix a resoldre d'una manera matemàticament satisfactòria aquestes dificultats en el cas de les teories de Yang-Mills. Tot i que avui dia s'està molt lluny d'una solució completa del problema, al llarg del segle XX s'ha estudiat amb molta profunditat la teoria quàntica de camps, tant des d'un punt de vista físic com matemàtic. Pel que fa a la física, hi ha un seguit de resultats i de tècniques (la majoria de les vegades fonamentades en arguments heurístics) que permeten fer càlculs d'una manera molt efectiva i precisa. L'exemple més espectacular són les prediccions obtingudes en l'electrodinàmica quàntica (quantització de l'electromagnetisme), com per exemple el càlcul del moment magnètic de l'electró, que s'adiu extraordinàriament bé amb les observacions experimentals. Pel que fa als resultats matemàtics, vegeu-ne alguns exemples a la secció 10 d'aquest article.

Des d'un punt de vista conceptual, hi ha dues maneres d'entendre des de la física la teoria quàntica de camps, segons que se segueix el model hamiltonià o el lagrangià. Per a cada una d'aquestes dues possibilitats hi ha una teoria corresponent, i un dels problemes més bàsics consisteix a demostrar que els dos punts de vista són equivalents. A grans trets, el punt de vista lagrangià consisteix a assumir que una teoria està descrita a nivell clàssic per un lagrangià $\mathcal{L}$ (de tal manera que les *solucions* del sistema són els punts crítics de $\mathcal{L}$) i definir el valor esperat de l'observable quàntic $\widehat{X}$ corresponent a l'observable clàssic X com:

$$\langle \psi \,|\hat{X}|\, \psi \rangle = \int X(\phi)\, e^{i\mathcal{L}(\phi)/\hbar} \mathcal{D}\phi \tag{5.1}$$

La integral de la dreta és sobre l'espai de tots els camps, i part de la dificultat de definir la teoria quàntica de camps consisteix a donar un sentit a aquesta integral (ja que en l'espai de camps té dimensió infinita i no té cap mesura natural).

Una de les tècniques que ha tingut més èxit consisteix a suposar que $\mathcal{L}$ és una petita pertorbació d'un lagrangià lliure (és a dir, sense interaccions), definir per decret la integral en el cas lliure (habitualment, inspirant-se en els valors de les integrals gaussianes a $\mathbb{R}^n$ i utilitzant certs procediments canònics per definir —*regularitzar*— els determinants d'operadors en espais de Hilbert), i, finalment, expressar les integrals de la nostra teoria com a pertorbacions de les de la teoria lliure (és a dir, fent un desenvolupament de Taylor en dimensió infinita). Els termes que apareixen en el desenvolupament pertorbatiu de les integrals se solen organitzar utilitzant una eina combinatòria anomenada *diagrames de Feynman*. Per acabar, cal trobar una manera de sumar totes les contribucions donades pels diagrames de Feynman. Aquest últim pas sol ser extraordinàriament subtil, i sovint cal recórrer a una tècnica anomenada *renormalització* per cancel·lar d'una manera coherent les divergències que apareixen en la majoria dels diagrames de Feynman. Es tracta d'un procediment que pot semblar a primera vista arbitrari, però la gran sorpresa és que els resultats obtinguts seguint aquesta recepta coincideixen amb molta precisió amb els experiments.

5.3. Fase estacionària

La integral (5.1) es defineix sense dificultat quan la considerem sobre una varietat de dimensió finita M i les funcions $\mathcal{L}$ i X tenen un comportament adequat a l'infinit. El teorema de l'aproximació estacionària permet aproximar el valor de la integral en termes del comportament de $\mathcal{L}$ i X prop dels punts crítics de $\mathcal{L}$. Suposem, per simplificar, que M és compacta i de dimensió $2l$, i que la funció $\mathcal{L} : \mathrm{M} \to \mathbb{R}$ és una funció de Morse. Sigui $X : \mathrm{M} \to \mathbb{R}$ una funció diferenciable. Aleshores

$$\int_{\mathrm{M}} X(p)\, e^{it\mathcal{L}(x)} dp = \sum_{p \in M_0} \left(\frac{2\pi}{t}\right)^{l} e^{\pi i \sigma(Hp)/4} a_p X(p)\, e^{it\mathcal{L}(p)} + O(\mathrm{t}^{-l-1}),$$

on $M_0 = \{\text{punts crítics de } \mathcal{L}\}$, H_p és el Hessià de $\mathcal{L}$ al punt p, $\sigma(H_p)$ és la signatura de H_p i a_p és igual a $|\det H_p|^{1/2}$ (vegeu per exemple [BeGeV]). Un anàleg d'aquesta fórmula en un context de dimensió infinita (on M és un espai de camps) es fa servir per donar sentit als primers termes del desenvolupament pertorbatiu de la integral (5.1.). De la mateixa definició es dedueix que les solucions clàssiques donen lloc a les contribucions més significatives de la integral, anàlogament a com la mecànica clàssica és una primera aproximació de la mecànica quàntica.

5.4. Teoria de Yang-Mills i model estàndard

Yang i Mills van proposar els anys 50 d'usar teories *gauge* no abelianes (és a dir, fibrats de rang>1) per modelar fenòmens no relacionats amb l'electromagnetisme.

A nivell clàssic, el lagrangià més senzill associa a una connexió el *funcional de Yang-Mills*:

$$YM(A) := \int \|F_A\|^2.$$

La idea de Yang i Mills va ser represa anys després com un model de les forces feble i forta, i ha esdevingut clau en el model estàndard. El mòdel estàndard, on s'unifiquen les forces forta, feble i electromagnètica, és una teoria *gauge* amb grup d'estructura

$$G = S^1 \times SU(2) \times SU(3).$$

Això vol dir que un dels camps de la teoria és una connexió en un fibrat hermític de rang 6 (que de fet es compatible amb una descomposició del fibrat com a suma de fibrats de rangs 1, 2 i 3). A aquest camp s'hi afegeixen d'altres camps, típicament seccions de fibrats associats, per incloure en el model els diversos tipus de partícules.

Hi ha moltes evidències experimentals que recolzen la validesa del model estàndard, especialment a l'electrodinàmica quàntica. Però encara queden molts problemes a resoldre, especialment un de fonamental: entendre la quantització del model, incloent-hi els aspectes no pertorbatius. Un problema més concret és comprendre de quina manera les partícules que *transmeten* la força feble i forta poden adquirir massa en la teoria quantitzada. La dificultat prové del fet que clàssicament aquestes partícules (el lagrangià de les quals és el de Yang-Mills) no contenen cap terme de massa; per tant a nivell clàssic es comporten com si tinguessin massa 0. Ara bé, experimentalment s'ha observat que aquestes partícules (tant a la força feble com a la forta, en contrast amb l'electromagnetisme) sí que tenen massa. La resposta que s'ha donat a aquesta dificultat en el cas de la força feble és l'anomenat *mecanisme de Higgs* o *trencament de simetria*. Aquest mètode és consistent amb moltes de les observacions, llevat d'un aspecte crucial: fins ara mai s'ha observat el bosó de Higgs, la partícula que teòricament està a l'origen d'aquest trencament de simetria. Pel que fa a la força forta, una explicació parcial ve donada per l'anomenada llibertat asimptòtica descoberta per Gross, Wilczek i Politzer, que van rebre l'any 2004 el premi Nobel per aquest descobriment.

5.5. El mass gap

En general, l'espectre del hamiltonià H en una teoria quàntica de camps està contingut a $[0, \infty)$. Es diu que hi ha un *mass gap* si

$$m := \inf \operatorname{Spec} H \setminus \{0\} > 0,$$

i el nombre *m* s'anomena la *massa del sistema*. A l'electromagnetisme la massa és 0, ja que els fotons no tenen massa. En canvi, es conjectura que la teoria quàntica de Yang-Mills no abeliana a dimensió 4 té un *mass gap*. Això explicaria el fet que els gluons (les partícules que *transmeten* la força forta) estiguin dotats de massa i que la força forta sigui de curt abast, i ajudaria a entendre el fenomen de la llibertat asimptòtica.

Aquí acabem l'explicació de les nocions que apareixen a l'enunciat del problema del mil·lenni . En la resta de seccions explicarem algunes de les aplicacions que han trobat les idees de la teoria de Yang-Mills no abeliana a la geometria, ressaltant especialment aquelles que provenen de les contribucions directes de físics teòrics (notablement, de Witten).

6. CONNEXIONS SOBRE SUPERFÍCIES DE RIEMANN

Sigui Σ una superfície de Riemann compacta. Els fibrats vectorials *diferencials* sobre Σ es classifiquen pel seu rang $r \in \mathbb{N}$ i el seu grau $d \in \mathbb{Z}$. En canvi, donats (r, d), existeixen moltes classes d'isomorfisme de fibrats *holomorfs* amb rang r i grau d.

6.1. La jacobiana

Un exemple il·lustratiu ve donat pels fibrats de rang 1, anomenats *fibrats de línia*. El conjunt de les classes d'isomorfisme de fibrats de línia de grau 0 és la jacobiana de Σ, $\mathrm{Jac}(\Sigma)$. Una propietat important d'aquest conjunt és que té una estructura natural de varietat complexa, compatible amb l'estructura de grup donada pel producte tensorial de fibrats de línia (de fet, com que Σ sempre té una estructura de corba algebraica, la jacobiana $\mathrm{Jac}(\Sigma)$ és, de fet, una varietat algebraica).

6.2. Teoria de Hodge i la jacobiana

Una aplicació bastant senzilla del teorema de Hodge és la demostració que hi ha un homeomorfisme natural

$$\mathrm{Jac}(\Sigma) \simeq \mathrm{Hom}(\pi_1(\Sigma), S^1). \tag{6.2}$$

Per tant, topològicament, $\mathrm{Jac}(\Sigma)$ és el producte cartesià de $2g$ còpies de S^1. Però l'estructura de varietat complexa a $\mathrm{Jac}(\Sigma)$ depèn de l'estructura complexa de Σ. De fet, el teorema de Torelli ens diu que $\mathrm{Jac}(\Sigma)$, juntament amb una certa estructura addicional (la polarització), permet recuperar Σ.

6.3. L'espai de mòduli de fibrats

Sigui $\mathcal{F}(r,d)$ el conjunt de classes d'isomorfisme de fibrats holomorfs de rang r i grau d sobre Σ. Motivats pels resultats sobre la jacobiana ens podem preguntar si el conjunt $\mathcal{F}(r,d)$ té alguna estructura natural de varietat complexa (aquí natural s'ha d'entendre com compatible amb la noció de família de fibrats vectorials).

La resposta a aquesta pregunta la va donar Mumford l'any 1962 (vegeu [MFK]).

> **Mumford** (1962): *Si* r *i* d *són primers entre si, el conjunt de classes d'isomorfisme de fibrats poliestables dins* $\mathcal{F}(r,d)$*, que es denota per* $\mathcal{M}(r,d)$*, té una estructura natural de varietat complexa compacta.*

Un fibrat holomorf $\mathcal{E}$ és estable si per qualsevol subfibrat seu no trivial $\mathcal{F}\subset\mathcal{E}$ se satisfà $\frac{\deg\mathcal{F}}{\mathrm{rk}\mathcal{F}}<\frac{\deg\mathcal{E}}{\mathrm{rk}\,\mathcal{E}}$. Es diu que $\mathcal{E}$ és poliestable si $\mathcal{E}=\mathcal{E}_1\oplus\cdots\oplus\mathcal{E}_k$, cada $\mathcal{E}_j$ és estable i a més $\deg\mathcal{E}_j/\mathrm{rk}\mathcal{E}_j$ és el mateix per qualsevol j. De fet, la condició que r i d siguin primers entre si no és gaire important. Quan no es dóna, l'espai $\mathcal{M}(r,d)$ té estructura de varietat complexa singular. Una propietat molt important de la noció de poliestabilitat és que, en un cert sentit, genèricament tots els fibrats holomorfs són poliestables (és a dir, els que no ho són formen un conjunt molt petit). L'espai $\mathcal{M}(r,d)$ s'anomena l'*espai de mòduli de fibrats*.

L'any 1965 Narasimhan i Seshadri [NS] van demostrar un teorema que generalitza a rang superior l'homeomorfisme (6.2).

> **Narasimhan-Seshadri** (1965): *Existeix un homeomorfisme entre* $\mathcal{M}(r,d)$ *i el conjunt de classes d'isomorfisme de connexions unitàries projectivament planes en un fibrat diferencial de tipus* (r,d) *sobre* Σ*.*

Aquest resultat permet concloure, per exemple, que

$$\mathcal{M}(r,d)\simeq \mathrm{Hom}(\pi_1(\Sigma),U(r))/U(r), \tag{6.3}$$

on al quocient de la dreta fem actuar $U(r)$ sobre $\mathrm{Hom}(\pi_1(\Sigma),U(r))$, a través de l'acció adjunta de $U(r)$ sobre si mateix. Una altra conclusió és que l'estructura topològica de $\mathcal{M}(r,d)$ només depèn de la topologia de Σ, exactament com ocorria amb la jacobiana.

L'any 1983 Donaldson [D], continuant una recerca començada per Atiyah i Bott (vegeu més endavant) va donar una nova demostració del teorema de Narasimhan-Seshadri usant el funcional de Yang-Mills.

A diferència del que passava amb l'homemorfisme (6.2), la identificació (6.3) entre l'espai de mòduli $\mathcal{M}(r,d)$ i l'espai de classes de representacions unitàries del grup

fonamental no permet calcular immediatament la topologia de l'espai de mòduli, ja que l'espai de representacions és força complicat topològicament.

6.4. Nombres de Betti de $\mathcal{M}(r,d)$

Un primer pas en l'estudi de la topologia de $\mathcal{M}(r,d)$ va ser el càlcul dels seus nombres de Betti. Els primers a aconseguir-ho van ser Harder i Narasimhan [HN], que van utilitzar les famoses conjectures de Weil poc abans demostrades per Deligne. Una segona demostració va ser donada per Atiyah i Bott usant mètodes absolutament diferents (però relacionats d'una manera misteriosa amb els de Harder i Narasimhan, vegeu [AB]). La idea clau usada per Atiyah i Bott és la següent:

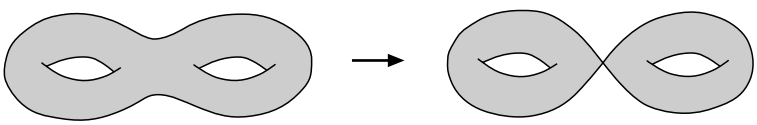
Atiyah-Bott (1982): *Sigui $\mathcal{A}$ l'espai de connexions unitàries en un fibrat donat sobre una superfície de Riemann. El funcional de Yang-Mills*

$$\mathcal{A} \ni A \mapsto YM(A) := \int \| F_A \|^2$$

és una aplicació de Bott-Morse a l'espai de connexions equivariant sota l'acció del grup gauge $\mathcal{G}$.

6.5. Cohomologia de $\mathcal{M}(r,d)$

Un problema posterior va ser calcular l'estructura d'anell de la cohomologia a coeficients racionals de $\mathcal{M}(r,d)$. El primer a calcular-la, l'any 1992, va ser Witten, que es va basar en la quantització de la teoria de Yang-Mills a dimensió 2 estudiada l'any 1974 per Migdal (i que no presenta ni de lluny les dificultats que apareixen en pujar la dimensió). Simultàniament, Witten va calcular la funció de partició de la teoria quàntica de Yang-Mills en dimensió 2. Una idea bàsica en els arguments de Witten és un procés de degeneració de la superfície de Riemann on una certa corba es contreu fins a col·lapsar en un punt:



Amb aquest procés es redueixen els càlculs a un mecanisme inductiu sobre el gènere de la superfície. Els arguments donats per Witten no eren completament rigorosos des del punt de vista matemàtic, i va caldre esperar fins a l'any 1998 perquè Jeffrey i Kirwan [JK] demostressin rigorosament els càlculs de Witten.

7. TEORIA DE DONALDSON

Un altre camp on les idees de Yang-Mills han tingut una influència espectacular és la geometria en dimensió 4, notablement a través de la teoria de Donaldson. Sigui (X, g) una varietat compacta de dimensió 4, sigui E un fibrat sobre X i $\mathcal{A}$ l'espai de connexions unitàries a E. Les connexions en les quals es minimitza el funcional de Yang-Mills

$$\mathcal{A} \ni A \mapsto \int \| F_A \|^2$$

s'anomenen *instantons*. L'espai de mòduli d'instantons és el quocient

$$\mathcal{M} = \{ \text{ instantons } \} / \mathcal{G} ,$$

on $\mathcal{G}$ denota el grup *gauge* de E. L'espai $\mathcal{M}$ admet una topologia i una compactificació natural, i la idea clau de Donaldson (vegeu per exemple [DK]) és que la topologia de $\mathcal{M}$ conté informació sobre la geometria diferencial de X. Combinant un dels primers teoremes obtinguts per Donaldson usant instantons amb resultats topològics de Freedman, Gompf [G] va demostrar que

> *existeix una varietat diferencial* M *que és homeomorfa a* $\mathbb{R}^4$ *però que no és difeomorfa a* $\mathbb{R}^4$ *amb la seva estructura diferencial habitual.*

Aquest resultat és encara més espectacular quan es té present que, per tot k, diferent de 4, qualsevol varietat diferencial homeomorfa a $\mathbb{R}^k$ és difeomorfa a $\mathbb{R}^k$. Posteriorment, Taubes va demostrar que $\mathbb{R}^4$ admet una quantitat no numerable d'estructures diferencials diferents!

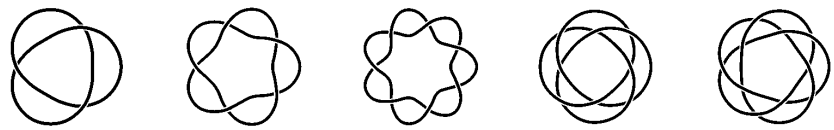
La teoria de Donaldson és una eina molt potent, especialment per estudiar la geometria diferencial de les varietats compactes de dimensió 4. Per exemple, ha estat molt útil en l'estudi de les superfícies complexes des del punt de vista diferencial. A la següent secció explicarem una aplicació en un camp aparentment allunyat de la dimensió 4: la teoria de nusos.

7.1. Aplicació: teoria de nusos

Recordem que en el llenguatge matemàtic per un nus s'entén un *embedding* de la circumferència S^1 a l'espai $\mathbb{R}^3$. Es diu que dos nusos són equivalents quan hi ha un camí d'*embeddings* que uneix un nus amb l'altre de manera que la curvatura dels *embeddings*, a mesura que passem de l'un a l'altre, estigui constantment acotat superiorment (aques-

ta condició tècnica permet evitar que estirem la circumferència per dos punts fins a fer col·lapsar una porció del nus, com se sol fer en el que quotidianament entenem per un nus). Un dels problemes més naturals (a part del problema fonamental de classificar els nusos) és calcular el nombre de deslligament d'un nus, que es defineix com el mínim nombre de modificacions que cal fer al nus per convertir-lo en el trivial. Per modificació entenem el següent: si en un diagrama del nus hi ha dos trams que s'intersequen, de manera que un passa per sobre de l'altre, modifiquem el nus a l'interior dels trams de manera que el diagrama quedi idèntic fora del creuament i que el tram que passava per sobre ara passi per sota. Tot nus es pot convertir en el trivial fent un cert nombre de modificacions, però calcular el nombre de deslligament és en general un problema molt complicat.

Una família molt important de nusos són els anomenats *nusos tòrics* $T(p,q)$, que s'obtenen recorrent un torus $S^1 \times S^1$ fent variar els angles a velocitats constants p i q. Per exemple:



$T(2,3)$ $T(2,5)$ $T(2,7)$ $T(3,4)$ $T(3,5)$

Malgrat que des de molts punts de vista els nusos tòrics són relativament senzills, no és gens trivial calcular-ne el nombre de deslligament (excepte en casos particularment senzills com, per exemple, $T(2, 3)$, que té nombre de deslligament 1). Usant la teoria de Donaldson, Kronheimer i Mrówka [KM1] van aconseguir calcular exactament el nombre de deslligament de qualsevol nus tòric.

Teorema. (Kronheimer i Mrówka, 1993) *El nombre de deslligament de* $T(p,q)$ *és*

$$\frac{(|p|-1)(|q|-1)}{2}$$

L'any 2003 Rasmussen va redemostrar el teorema amb mètodes molt més elementals, usant homologia de Khovanov. Aquesta homologia, però, també té els seus orígens a la física, concretament a través del polinomi de Jones. De manera que tant una demostració com l'altra parteixen en el seu inici de nocions provinents de la física!

Val a dir que una altra font importantíssima d'influència de la física en la teoria de nusos (gràcies, un cop més, als treballs de Witten) és la teoria de Chern-Simons, íntimament relacionada amb la teoria de Yang-Mills.

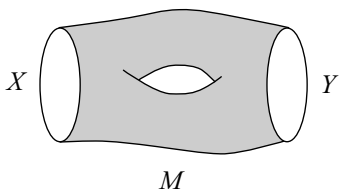
8. TQFT

L'any 1988 Witten va donar una interpretació física dels invariants de Donaldson, veient-los com a observables d'una teoria quàntica de camps *torçada*. A principis de l'any 1994, Witten va aconseguir calcular els invariants de Donaldson de superfícies Kaehler usant arguments físics i partint de la seva interpretació dels invarians obtinguda l'any 1988. Una idea clau és que sobre superfícies Kaehler hi ha una deformació de la teoria de Donaldson a una teoria que posseix un *mass gap*. En termes geomètrics això implica que es pot estudiar la teoria de Donaldson en el límit quan la mètrica és reescalada per una factor que tendeix a ∞. En aquest límit, la teoria es pot modelar amb una teoria sobre $\mathbb{R}^4$ amb la mètrica plana. Remarcablement, les fórmules obtingudes per Witten usant mètodes físics coincideixen amb els resultats que Kronheimer i Mrówka varen obtenir un any abans amb mètodes matemàtics.

Atiyah va reelaborar les idees de Witten en la noció de TQFT (*topological quantum field theory*). En termes generals, una TQFT és un functor

$$\mathcal{F} : \mathrm{Cob}\,(d+1) \longrightarrow \mathrm{Vect}$$

que satisfà uns certs axiomes. Aquí, $\mathrm{Cob}\,(d+1)$ és la categoria dels cobordismes $(d+1)$-dimensionals (objectes: varietats compactes de dimensió d; morfismes: cobordismes compactes de dimensió $d+1$), i Vect denota la categoria dels espais vectorials. Per exemple, el cobordisme



dóna un morfisme d'espais vectorials $\mathcal{F}\,(M) : \mathcal{F}\,(X) \longrightarrow \mathcal{F}\,(Y)$. La noció de TQFT és un marc general on entren molts dels nous invariants geomètrics descoberts a les últimes dècades gràcies a la influència de la física.

Construir exemples no trivials de TQFT és un problema molt subtil. Únicament les TQFT amb $d = 1$ se saben classificar completament (es corresponen, a través d'una construcció geomètrica, amb les anomenades *àlgebres de Frobenius*). La teoria de Chern-Simons-Witten dóna lloc a una TQFT en $d = 2$, i la teoria de Donaldson dóna lloc, combinada amb l'homologia de Floer, a un exemple de TQFT en $d = 3$.

9. TEORIA DE SEIBERG-WITTEN

El novembre de 1994 Witten [W2] va combinar la interpretació física dels invariants de Donaldson amb càlculs *no pertorbatius* obtinguts juntament amb Seiberg usant una dualitat electromagnètica conjecturada per Montonen i Olive, i d'aquesta manera va arribar a les anomenades *equacions de Seiberg-Witten*, que permeten redemostrar molts dels resultats obtinguts prèviament en la teoria de Donaldson. A més a més, la teoria de Seiberg-Witten té un grup *gauge* abelià, cosa que elimina moltes de les dificultats tècniques a la teoria de Donaldson.

Witten va conjecturar una relació entre els invariants de Seiberg-Witten i els de Donaldson, que encara no ha estat completament demostrada matemàticament. Tot i això, hi ha resultats de Feehan i Leness (seguint un programa proposat per Pidstrigach i Tyurin) que s'acosten bastant a la solució de la conjectura de Witten.

9.1. Aplicació: la propietat P *per als nusos*

Si $K \subset S^3$ és un nus (és a dir, la imatge d'un *embedding* de S^1 dins S^3), fer una cirurgia de Dehn al llarg de K consisteix a treure un entorn tubular de K (que és un torus sòlid) i tornar-lo a enganxar identificant les vores amb un difeomorfisme arbitrari. Un problema clàssic sobre cirurgia, proposat per Bing, era saber si el resultat d'una cirurgia no trivial al llarg d'un nus no trivial podia donar lloc a una varietat simplement connexa. En els últims anys s'havia arribat a respondre (negativament) de manera quasi completa aquesta pregunta, però hi havia un cas que quedava obert (concretament, quan el coeficient de la cirurgia és 1). L'any 2003 Kronheimer i Mrówka [KM2] van ser capaços de resoldre aquest cas.

> **Teorema.** (Kronheimer i Mrówka, 2003) *Si $K \subset S^3$ és un nus no trivial, el resultat de fer una cirurgia de Dehn al llarg de K només és simplement connex si la cirurgia és trivial.*

Kronheimer i Mrówka van combinar resultats recents en teories *gauge*, geometria simplèctica i teoria de foliacions, en particular els resultats parcials de Feehan i Leness sobre la conjectura de Witten. Notem que per un teorema de Gordon i Lücke de l'any 1989 se sap que el resultat d'una cirurgia no trivial no és mai l'esfera, de manera que el teorema de Kronheimer i Mrówka és un corol·lari de la conjectura de Poincaré.

10. QFT CONSTRUCTIVA

En les últimes seccions hem estat parlant d'aplicacions de la teoria de Yang-Mills a la geometria. Per acabar aquest article donem alguns exemples de teories quàntiques de camps, més senzilles que Yang-Mills a dimensió 4, que s'han pogut definir rigorosament.

Un primer exemple són les teories escalars amb potencial $\lambda\phi^4 + \phi^2$ ($\lambda = 0$ és una teoria lliure), que s'han estudiat en dimensions 2 i 3. En alguns casos s'ha provat l'existència de *mass gap*.

Pel que fa a quantitzar Yang-Mills en dimensió 2, ja hem mencionat el treball de Migdal de l'any 1974, on se soluciona essencialment el problema. Recordem que aquesta solució és utilitzada per Witten per obtenir fórmules que descriuen l'estructura d'anell de la cohomologia de l'espai de mòduli de connexions planes en una superfície.

Finalment, pel que fa a Yang-Mills en dimensió 3 s'han obtingut alguns resultats parcials, especialment per les teories on el funcional de Yang-Mills es pertorba amb l'addició d'un terme de Chern-Simons. Hom pensa que hi ha d'haver *mass gap* independentment de la pertorbació.

REFERÈNCIES

[As] M. Asorey, «Teoría cuàntica de Yang-Mills, la generación de la masa», *Problemas del milenio*, ed. per L.J. Boya, Monografías de la Real Academia de Ciencias de Zaragoza (2004).

[AB] M. Atiyah, R. Bott, «The Yang-Mills equations over Riemann surfaces», *Phil. Trans. R. Soc. Lond.* A 308 (1982), p. 523-615.

[BeGeV] N. Berline, E. Getzler, M. Vergne, *Heat kernels and Dirac operators*, Grundlehren der Mathematischen Wissenschaften 298, Springer-Verlag.

[D] S.K. Donaldson, «A new proof of a theorem of Narasimhan and Seshadri», *J. Diff. Geom.* 18 (1983), p. 269-278.

[Dou] M.R. Douglas, «Report on the status of the Yang-Mills millenium prize problem», al web: http://www.claymath.org/millennium/Yang-Mills{_}Theory/ym2.pdf}.

[DK] S.K. Donaldson, P.B. Kronheimer, *The geometry of four-manifolds*, Oxford Mathematical Monographs, Oxford Science Publications, The Clarendon Press, Oxford University Press, Nova York, 1990.

[G] R.E. Gompf, «Three exotic $\mathbb{R}^4$'s and other anomalies», *J. Differential Geom.* 18 (1983), núm. 2, p. 317-328.

[HN] G. Harder, M.S. Narasimhan, «On the cohomology groups of moduli spaces of vector bundles on curves», *Math. Ann.* 212 (1974/75), p. 215-248.

[JW] A. JAFFE, E. WITTEN, «Quantum Yang-Mills theory», al web: http://www.claymath.org/millennium/Yang-Mills_Theory/Official_Problem_Description.pdf.

[JK] L. JEFFREY, F. KIRWAN, «Intersection theory on moduli spaces of holomorphic bundles of arbitrary rank on a Riemann surface», *Ann. of Math.* (2) 148 (1998), núm. 1, p. 109-196.

[KM1] P.B. KRONHEIMER, T. MROWKA, «Gauge theory for embedded surfaces I», *Topology* 32 (1993), núm. 4, p. 773-826.

[KM2] «Witten's conjecture and property P», *Geom. Topol.* 8 (2004), p. 295-310

[MFK] J. MUMFORD, D. FOGARTY, F. KIRWAN, *Geometric Invariant Theory*, 3rd edition, Ergebnisse der Math., Springer-Verlag (1994), Nova York.

[NS] M.S. NARASIMHAN, C.S. SESHADRI, «Stable and unitary vector bundles on a compact Riemann surface», *Ann. of Math.* 82 (1965), p. 540-564.

[W1] WITTEN, Edward, «Two-dimensional gauge theories revisited», *J. Geom. Phys.* 9 (1992), núm. 4, p. 303-368.

[W2] E. WITTEN, «Monopoles and four-manifolds», *Math. Res. Lett.* 1 (1994), núm. 6, p. 769-796.

[W3] E. WITTEN, «Physical law and the quest for mathematical understanding», *Bull. A.M.S.* 40 (2003), p. 21-30.

Dels problemes de Hilbert als problemes del mil·lenni[1]

Josep Pla i Carrera
Facultat de Matemàtiques. Universitat de Barcelona

1. INTRODUCCIÓ

Després de l'exitosa conferència —tantes vegades esmentada— de David Hilbert (1862-1943), llegida a París el dimecres 8 d'agost de 1900, davant la comunitat matemàtica durant la celebració del Segon Congrés Internacional de les Matemàtiques, no hi ha gaire més a dir sobre la importància indiscutible, essencial i irrenunciable dels problemes com a objecte i objectiu del quefer matemàtic, tant individualment com col·lectiva. Ell mateix en una síntesi conclusiva diu:

> La ciència matemàtica és un tot indivisible, un organisme la força vital del qual té com a condició indispensable la indisolubilitat de les parts. Efectivament, sigui quina sigui la diversitat de les matèries de la nostra ciència en relació amb els detalls, no deixa de sorprendre'ns l'equivalència dels processos lògics, el parentiu que hi ha entre les idees en el conjunt de la ciència així com nombroses analogies en els dominis diversos. Observem encara això: com més es desenvolupa una teoria matemàtica, la seva exposició hi guanya en harmonia i unitat, i es descobreixen més i més relacions entre la teoria i d'altres branques de la ciència que, fins aleshores, li eren alienes.
>
> Així doncs, malgrat que les Matemàtiques estenguin els dominis, no perden mai el caràcter unitari, ans al contrari, cada cop se'ns ofereixen amb una evidència més gran.[2]

1. Conferència llegida a la seu de la Caixa de Sabadell. Sabadell, el 15 de febrer de 2005. D'antuvi, voldria agrair molt efusivament a Jordi Quer l'amabilitat i confiança que m'ha demostrat en convidar-me a participar en aquesta exposició dels problemes més rellevants del segle XXI, oferint-me la possibilitat de fer una reflexió històrica i epistemològica de la importància que els problemes tenen en el quefer matemàtic tant en el vessant docent com en el vessant de la recerca més estricta. I també vull fer-la extensiva a la Caixa de Sabadell per haver acceptat la proposta d'en Jordi.

Finalment, em sento obligat amb l'amiga i docta col·lega Pilar Bayer i Isant per la lectura pacient que ha fet del text i per les observacions que m'ha proporcionat per millorar-ne la qualitat i la precisió. (Començada a escriure: 15 de desembre de 2004. Acabada d'escriure: 20 de gener de 2005.)

2. HILBERT, D. (1900b), a GRAY, J.J. (2000), edició castellana de 2003, 310.

Tot això està d'acord amb un paràgraf anterior:

> Sovint, la raó per la qual no aconseguim resoldre un problema matemàtic és que encara no hem assolit un punt de vista prou general des del qual el problema se'ns mostri com una baula simple d'una cadena de problemes de la mateixa naturalesa. Però una vegada aconseguim assolir aquest punt de vista, no solament el problema es fa més abordable, sinó que, a més, ens trobem en possessió d'un mètode aplicable a d'altres problemes de la mateixa espècie.[3]

Malgrat les diferències conceptuals que els separen —que, des d'un punt de vista epistemològic, són moltes i importants— la resolució de problemes com a metodologia per comprendre i desenvolupar la matemàtica també la defensa el gran matemàtic francès del segle XIX, Henri Poincaré (1854-1912). Així en l'obra *Science et méthode* (1908) hi llegim:

> El mètode vertader per preveure l'esdevenidor de la matemàtica passa per estudiar la seva història i l'estat actual. Per als matemàtics no és un procediment poc professional? Estem acostumats a extrapolar, que no és altra cosa que un mitjà de deduir l'esdevenidor del passat i del present. A més, atès que sabem el que val tot això, no hi ha cap perill de desil·lusionar-nos pel que fa a l'abast dels resultats que ens proporciona.[4]

I afegeix:

> En èpoques passades hi ha hagut profetes de la desgràcia. Repetien una vegada i una altra, molt convençuts, que tots els problemes susceptibles de ser resolts ja ho havien estat, de resolts, i que, després d'ells, l'única cosa que quedaria per fer fóra posar-se a plegar la collita. Feliçment l'exemple del passat ens tranquilitza. Moltes vegades hem cregut que havíem resolt tots els problemes o, almenys, havíem fet l'inventari dels que impliquen una solució. Però més tard, el sentit de la paraula *solució* s'ha eixamplat i els problemes irresolubles s'han convertit en els de més interès, i se n'han plantejat de nous que mai no s'havien somniat.[5]

Tanmateix, Hilbert va encara més lluny, si és possible. Després de repassar alguns problemes que ell considera clàssics i abans d'exposar els vint-i-tres de la conferència —problemes que planteja amb l'esperança que siguin resolts durant el segle XX—, afir-

3. HILBERT, D. (1900b), a GRAY, J.J. (2000), edició castellana de 2003, 270.
4. POINCARÉ, H. (1908), edició castellana de 1962, 45.
5. POINCARÉ, H. (1908), edició castellana de 1962, 45.

ma amb contundència una mena de postulat relatiu a la capacitat creativa de l'activitat matemàtica, com quelcom propi de la seva naturalesa. Diu:

> Aquesta convicció de la possibilitat de resoldre qualsevol problema matemàtic és, per a tots nosaltres, en la nostra feina, un incentiu d'allò més poderós. En tot moment sentim que ressonen dins nostre les paraules: «Heus aquí el problema. Busca-li la solució. Pots aconseguir-la amb el raonament pur. Mai dels mais, un matemàtic es veurà forçat a dir *Ignoràbimus*».[6]

I, a més, unes quantes línies abans, ens ha fet notar que la resolució d'un problema pot portar a una *solució negativa* que, de fet, estableix la seva pròpia impossibilitat.[7]

Diu:

> Els antics ens han proporcionat els primers exemples d'aquesta mena de demostracions: les demostracions d'impossibilitat [...]. En la matemàtica moderna, la qüestió de la impossibilitat juga un paper molt preponderant.[8]

Però la tasca del paper que juguen els problemes en el món de l'ensenyament de la matemàtica —o de l'aprenentatge— la dugué a terme, a mitjan segle xx, George Pólya (1887-1985), que en féu una metodologia didàctica que ha generat, d'aleshores ençà, molta literatura al si dels diferents corrents de pensament que es preocupen d'establir les tècniques més idònies per transmetre i assimilar els coneixements.

Les seves reflexions han portat a la línia d'aprenentatge de la matemàtica coneguda amb el nom de *Solving Problems*.

D'altra banda, hem vist créixer, cada cop més, una recerca matemàtica, basada fonamentalment en la resolució de problemes.[9] Així, per exemple, ments modernes i genials com les de Paul Erdös (1913-1996) i John Horton Conway (1937-), entre molts d'altres, han posat de manifest d'una manera clara i indiscutible que la necessitat de

6. Hilbert, D. (1900b), a Gray, J.J. (2000), edició castellana de 2003, 272. L'èmfasi, llevat el de la paraula *Ignoràbimus*, és meu.

7. En aquest text observem una gran concordança amb el text de Poincaré. Tanmateix, no hem de confondre els problemes la resolució dels quals és impossible, perquè val el seu negat, amb aquells altres que són «irresolubles», els quals, per a Hilbert, no existien. En parlarem al § 4.

8. Hilbert, D. (1900b), a Gray, J.J. (2000), edició castellana de 2003, 271.

9. En èpoques pretèrites també s'ha fornit una recerca matemàtica important basada en els problemes. Un personatge paradigmàtic en aquest sentit fou, sense cap mena de dubte, en el món de l'aritmètica, Pierre de Fermat. Recordem, de passada, que Fermat s'afeccionà a l'aritmètica després d'haver llegit l'*Aritmètica* (1621) de Claude-Gaspart Bachet de Méziriac (1581-1638), una traducció llatina, anotada, de les obres de Diofant d'Alexandria (segle iii). Bé, doncs, aquest eminent erudit francès s'havia interessat ja per la matemàtica lúdica a *Problèmes plaisantes et déléctables, qui se font par les nombres* (1612).

resoldre problemes —i diria més, de plantejar-ne— és molt més que una simple actitud lúdica. És una eina indispensable per fer avançar el pensament matemàtic, suggerir intuïcions, motivar recerques, etc.

Per a tots ells, és una *heurística autèntica* de la matemàtica, si considerem el terme *heurística de la matemàtica* en el sentit més primigeni i simple: una bona comprensió del procés matemàtic i alhora una bona comprensió de l'acte cognitiu del mateix procés creatiu.[10]

En resum, davant del soroll —a vegades eixordador— de pedagogs lletraferits i de didactes nostrats i algun de forani, ens adonem que, de fet, el camí real —i, potser, reial—[11] que cal seguir per fer matemàtiques és quelcom tan simple com l'habilitat de resoldre problemes i la capacitat de plantejar-ne. Simplificant podríem dir que aprendre matemàtiques és aprendre a resoldre problemes i crear matemàtiques és ser capaç de plantejar problemes nous, rellevants i interessants.

L'èmfasi d'aquesta exposició —que vol ser la porta d'entrada a l'anàlisi dels set problemes del mil·lenni—[12] recaurà a fer veure que els corrents subterranis emergeixen arreu en els moments i situacions més inesperats. Aquesta intercomunicació íntima és una de les qualitats que considero més remarcables de la matemàtica i, si no la tenim en compte, no podrem entendre realment amb tota la profunditat possible allò que realment és el quefer matemàtic.[13] Nosaltres ara, en aquesta reflexió col·lectiva, mirarem d'entendre-ho, o si més no de copsar-ho, tot basant-nos en alguns dels problemes que Hilbert menciona o proposa al text.

10. En aquest context voldria recordar, i recomanar, els textos magnífics d'Otto Toeplitz, *Die Entwicklung der infinitesimalrechnung* (1947) i *Von Zahlen und Figuren* (1923), aquest darrer escrit juntament amb Hans Rademacher. I també val la pena retenir l'excel·lent obra de Heinrich Dörrie (Dörrie, H. 1958). Molt més moderna és la manera com John Stillwell (Stillwell, J. 1989) ens apropa a la història de la matemàtica, no d'una manera històrica rigorosa sinó a través del desenvolupament més o menys evolutiu de certs conceptes i problemes. D'altres intents, en aquest mateix sentit, però més divulgatius, són, per exemple, Laubenbacher, R. i Pengelley, D. (1999).

11. Com recordareu, Arquimedes respon al rei de Sicília, Hieró II: «No hi ha camins específics per als reis per aprendre la matemàtica», en el sentit que el fet de ser rei no és cap privilegi —ni tampoc, en principi, cap demèrit— per aprendre matemàtiques. Però jo ara aquí uso aquest terme en un sentit ben diferent: en el sentit de «camí principal».

12. Vegeu la conferència de Smale, a l'Institut Fields, en ocasió del seixanta aniversari d'Arnold, Smale, S. (1998).

13. Posa de manifest la discussió mai acabada i mai aclarida de si cal dir «matemàtica» o «matemàtiques» per referir-se a aquesta disciplina.

2. EL TEOREMA DELS NOMBRES PRIMERS

Un dels problemes que Hilbert esmenta, tot de passada, és el que fa referència al *teorema dels nombres primers*. De fet, es tracta de saber quina és la quantitat de nombres primers $\leq n$, relativament a n.

Aquest problema és realment interessant perquè posa de manifest d'una manera molt clara aquesta característica tan pregona i interessant de la matemàtica: hi ha una comunicació —i una comunió— entre allò que, per qüestions de taxonomia, es considera com a diferent i, moltes vegades, disjunt.

Els nombres primers en la tradició pitagòrica

No cal indicar la importància dels nombres primers en la teoria de la divisibilitat, recollida en els llibres VII, VIII i IX dels *Elements* d'Euclides (325 aC-265 aC).[14] S'hi estableix:

> Tot nombre s'obté a partir dels nombres primers, de forma única i, de nombres primers, n'hi ha infinits.[15]

Nicòmac de Gerasa (~60-~120) ofereix, a l'*Aritmètica*, un mètode per determinar tots els nombres primers $\leq n$. Es coneix amb el nom de *garbell d'Eratòstenes* perquè, segons diu, fou Eratòstenes de Cirene (276 aC-194 aC) el primer a establir-lo.[16] No és gaire efectiu, però, durant molt de temps, fou l'únic que permetia trobar els nombres primers $\leq n$.

14. *Elements*, VII, proposicions 30, 31, 32, a VERA, F. (1970), 840-841.

15. Són, respectivament, les proposicions 14 i 20 del llibre IX dels *Elements* (VERA, F. 1970, 851-852; 853-854.) S'ha discutit, a bastament, si Euclides havia establert la unicitat de la descomposició en primers. El seu enunciat (IX 14: «el nombre més petit mesurat per diversos nombres primers no té d'altres nombres primers que el mesurin que aquests») és certament ambigu: En l'expressió «diversos nombres primers», Euclides preveu la possibilitat que puguin repetir-se, o no? Tot depèn de la interpretació que donem a aquesta expressió. Això no obstant, Euclides disposa de totes les eines conceptuals necessàries per fer-ho, en particular del *lema de Gauss*: «si p és un nombre primer que divideix $m \times n$, però no divideix m, aleshores necessàriament divideix n». Tanmateix, passarien molts segles abans que, amb el treball d'Euler de 1737, no es produís un avenç significatiu en aquest sentit.

16. Conservem el dos i passem ratlla de dos en dos; conservem el tres i passem ratlla de tres en tres; conservem el cinc i passem ratlla de cinc en cinc; conservem el set i la passem de set en set; atès que ara els nombres no barrats són $>\sqrt{50}$, ja hem acabat. Tots els nombres que queden sense ratlla són primers i ≤ 50.

	2	3	4̸	5	6̸	7	8̸	9̸	10̸
11	12̸	13	14̸	15̸	16̸	17	18̸	19	20̸
21̸	22̸	23	24̸	25̸	26̸	27̸	28̸	29	30̸
31	32̸	33̸	34̸	35̸	36̸	37	38̸	39̸	40̸
41	42̸	43	44̸	45̸	46̸	47	48̸	49̸	50̸

Aquesta mentalitat pitagòrica, basada en els nombres primers, la trobem encara en el darrer dels problemes plantejats per Euclides: la determinació dels nombres perfectes (parells).[17]

Un nombre natural P és perfecte parell si és de la forma $P = 2^{n-1}(2^n - 1)$, sempre que $M_n = 2^n - 1$ sigui primer.[18]

Segles més tard, Euler demostraria que aquesta condició és necessària.[19]

La distribució i determinació dels nombres primers

Una qüestió que havia de sorgir més tard o més d'hora era la recerca de mètodes estàndards per determinar els nombres primers.[20] Sense conèixer els nombres primers és, a més, del tot impossible descompondre un nombre en factors primers.[21]

Una primera manera d'apropar-se a aquest problema fóra conèixer una mica millor l'aspecte que pot tenir la taula que s'obté després d'aplicar el mètode d'Eratòstenes.

(1) Hi ha forats molt grans entre primers consecutius?
(2) Hi ha moltes parelles de nombres primers bessons?[22]

17. «Un nombre natural *P és perfecte* si, i només si, és igual a la suma de les seves parts pròpies» (*Elements*, VII, definició 22, VERA, F. 1970, 831).

18. Proposició 36 del llibre IX dels *Elements* (VERA, F. (1970), 857-860).

19. EULER, L. (1849), obra pòstuma.

20. Recordem que, durant segles, hom va creure que $2^n - 1$ és primer si, i només si, n és primer. Hudalrichus Regius (~1535) va adonar-se que $2^{11} - 1$ és compost. L'única manera d'entendre-ho, atesa la petitesa de $2^{11} - 1 = 2047 = 23 \times 87$, és el fet de no disposar d'un sistema numèric àgil dotat d'algorismes de càlcul efectius, quelcom que es resoldria amb els numerals indoaràbigs.

21. Cal indicar, com veurem més endavant, que descompondre un nombre en factors primers és un problema d'una gran complexitat. Tanmateix, en 1643, Pierre de Fermat (1601-1650) va trobar un criteri de factorització realment simple: «un nombre és compost si, i només si, és una diferència de quadrats». És a dir, si existeixen $x, y \in \mathbb{N}, n = x^2 - y^2$. Malgrat la simplicitat aparent, quan els nombres són grans, és molt poc operatiu. Leonhard Euler el milloraria imposant simplement que existissin $x, y \in \mathbb{N}, n \equiv x^2 - y^2$ (mòd n). En 1931 Derrick H. Lehmer (1905-1991) i R.E. Powers s'adonaren que era possible recórrer a les convergents del desenvolupament de $\sqrt{n}$ en fracció contínua, atès que $q_k^2 - np_k^2 = (-1)^{k-1}\, Q_{k-1}$. Aleshores el mètode es simplifica si aconseguim trobar un Q_{k-1} que sigui quadrat perfecte. (Vegeu, per exemple, ROSEN, K.H. 2000, 477-479).

22. Recordem que, si p i $p + 2$ són primers, p senar, són *primers bessons*.

Pel que fa a la primera qüestió, sabem que hi ha forats tan grans com vulguem entre dos nombres primers consecutius.[23] La segona qüestió —«l'existència d'infinites parelles de nombres primers bessons»— encara ara és una *conjectura*.

En 1837, P.G. Leujene Dirichlet (1805-1859) establia el teorema següent:

> En tota progressió aritmètica $a,\ a+d,\ a+2d,\ldots,\ a+(n-1)d,\ldots,$ amb a i d primers entre si, hi ha una infinitat de nombres primers.[24]

Al costat d'aquests avenços hi ha qüestions obertes, com ara:

> (1) La *conjectura de Goldbach*: Tot nombre natural, diferent del dos, és la suma de dos nombres primers.
> (2) La *conjectura de Bertrand*: Per a cada nombre natural n, hi ha un nombre primer p que $n<p<2n$?
> (3) Hi ha infinits primers de la forma n^2+1?
> (4) Hi ha sempre un nombre primer entre n^2-n i n^2? I entre n^2 i n^2+n?
> (5) Hi ha infinits nombres primers de Fermat?
> (6) Hi ha infinits nombres primers de Mersenne?, etc.[25]

23. Sigui $n\in\mathbb{N}$. Aleshores els nombres $n!+2$, $n!+3,\ldots,$ $n!+n$ són nombres compostos.

24. DIRICHLET, P.G. LEUJENE (1837). Ja havia estat conjecturat per Euler i Legendre. De fet, generalitza els teoremes: «hi ha infinits nombres primers (1) de la forma $4k+1$ i (2) de la forma $4k-1$».

25. La conjectura de Goldbach la proposà Christian Goldbach (1690-1764) a Euler, en una carta del 7 de juny de 1742, amb l'esperança que l'eminent matemàtic la resoldria (EULER, L.-GOLDBACH, C. 1965, 127).

La conjectura de Bertrand, proposada per Joseph Bertrand (1822-1900) —que la verificà fins al 3.000.000—, fou demostrada per Pafnuty Lvovich Chebyshev (1821-1894) en 1850. En 1931, Robert Breusch la millorà establint que, per a cada $n\geq 48$, hi ha un nombre primer entre n i $\frac{9n}{8}$.

En 1659, en una carta adreçada a Pierre de Carcavi (1600-1684), Pierre de Fermat afirmava que tots els nombres de la forma $F_n=2^{2^n}+1$ són primers. En 1732, Euler en demostrà l'error veient simplement que $F_5=4\,294\,967\,297=641\times 6\,700\,417$ és compost. D'aleshores ençà tots els que s'han pogut tractar ($6\leq n\leq 16$ i $n=18, 19, 21, 23, 36, 38, 39, 55, 63, 73$) han resultat compostos i avui hom creu que tots els nombres F_n, llevat dels casos $n=1, 2, 3, 4$, són compostos. La primeralitat dels nombres de la forma $M_n=2^n-1$ és molt important per conèixer els nombres perfectes parells. Però és força complicada atesa la grandària d'aquest tipus de nombres. En 1644, en l'obra *Cogitata Physico-Mathematica*, el pare mínim Marin Mersenne (1588-1648) afirmava que, quan $p=2, 3, 5, 7, 13, 17, 19, 31, 67, 127$, i 257, M_p és primer i, per a tot altre nombre primer $p\leq 257$, M_p és compost. Avui sabem la gran dificultat que comportà escatir si la *llista de Mersenne* era correcta, aclarir-ne els errors i completar-ne les llacunes. No s'assolí fins a mitjan segle xx, amb l'ajut de les primeres grans computadores. (Vegeu, per exemple, BURTON, D.M. (1985), 478-483.) Per aquesta raó els nombres M_n reben el nom de *nombres de Mersenne*.

El *teorema de Wilson* és un teorema senzill que caracteritza la *primeralitat* d'un nombre natural:

> Un nombre p és primer si, i només si, divideix $(p-1)!+1$.[26]

Atès el creixement de $n!$,[27] el criteri de Wilson és molt poc apropiat per determinar la primeralitat de p.[28]

Encara podem considerar un altre camí per conèixer els nombres primers:

> Determinar una funció $f(k)$ que els generi.[29]

Però Joseph-Louis Lagrange (1736-1813) demostraria que no hi ha cap funció polinòmica amb una variable que proporcioni solament nombres primers.[30]

El 1975, G.L. Miller (1942-) va establir una definició que, d'alguna manera, generalitzava la pseudoprimeralitat:

> Sigui n un nombre senar amb $n-1=2^S t$, amb t senar. Diem que n passa el test de Miller per a la base b, si s'esdevé una d'ambdues possibilitats:
> (1) $b^t \equiv 1$ (mòd n),
> (2) $b^{2^j t} \equiv 1$ (mòd n) per a algun $0 \le j \le s$.

26. A *Meditationes Algebraicæ* (1770), Edward Waring (1734-1798) l'atribueix a John Wilson (1741-1793). Aquest teorema ja era conegut per Leibniz, però fou demostrat, per primera vegada, en 1773, per Lagrange.

27. Recordem la *fórmula de Stirling* $n! \simeq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ quan y és gran—demostrada en 1730, simultàniament i independent, per Abraham De Moivre (1667-1754) i per James Stirling (1692-1770) —palesa el caràcter exponencial de $n!$. És un d'aquests resultats en què l'anàlisi aporta un resultat d'aproximació d'una qüestió de tipus aritmètic.

28. D'altra banda, el *teorema petit* de Fermat (FERMAT, P. DE (1640)).

> Si $\langle a,p \rangle = 1$, $a^{p-1}-1$ és divisible per p,

és vàlid quan p és primer, però no caracteritza els nombres primers. Permet introduir la definició següent: Un nombre natural n és pseudoprimer (per a la base b) quan $b^{p-1}-1$ és divisible per p. El més petit nombre pseudoprimer (per a la base 2), no primer, és el 341.

29. El més interessant fóra que $f(1) = 2$, $f(2) = 3, \ldots$, $f(k) = p_k, \ldots$, on p_k és el k-èsim nombre primer. Leonhard Euler observà que $f(n) = n^2 - n + 41$ proporciona nombres primers per a cada $n < 41$, atès que $f(41) = 41^2$ no és primer. També el polinomi quadràtic $f(n) = n^2 - 79n + 1601$ propociona nombres primers, per a tot $n < 80$.

30. En aquesta línia cal esmentar el resultat de W.H. Mills de 1947: «Hi ha un nombre real C que $E(n) = [C^{3^n}]$ ni genera nombres primers per a tot $n \in \mathbb{N}$». No es coneix amb exactitud quina és la constant C, però se sap que $C \simeq 1{,}3064$.

I en deduí un *test de primalitat*,[31] tot observant que:

> Si n és un nombre primer i b no és múltiple de n, aleshores n passa el test de Miller per a la base b.
> Si n és un nombre natural senar compost, aleshores n passa el test de Miller per a, com a màxim, $\frac{n-1}{4}$ bases $b, 1 \le b \le n-1$.
> Tot això fa que:
> (1) Si n no passa una base b petita, és compost.
> (2) Si n passa el test de Miller per a $k \ge \frac{n-1}{4}$ bases, és primer.

Un incís doble

1. Quan Christiaan Huygens (1629-1695) va conèixer Wilhelm Gottfried Leibniz (1642-1716), li va proposar que trobés la suma dels inversos dels nombres triangulars i dels nombres quadrats.[32] Leibniz calculà, amb facilitat, el valor de la suma dels inversos dels nombres triangulars.[33] La suma dels inversos dels nombres quadrats

$$\zeta(2) = \frac{1}{1} + \frac{1}{4} + \cdots + \frac{1}{n^2} + \cdots = \frac{\pi^2}{6}$$

seria calculada per Leonhard Euler (1707-1783).[34] A més, s'interessà per la funció

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \frac{1}{1} + \frac{1}{2^s} + \frac{1}{3^s} + \cdots + \frac{1}{n^s} + \cdots$$ [35]

31. Proporciona un criteri probabilístic molt bo atès que, si n és compost, passa el test de Miller per a una base elegida a l'atzar amb probabilitat $(\frac{1}{4})^k$ i el passa per a k bases elegides a l'atzar amb una probabilitat k que, quan $k = 20$, és aproximadament 0,000000000001. (Vegeu DAVIS, D.M. (1993), 245-248.)

32. Com ja va demostrar Nicole Oresme (1323-1382) al segle XIV, la suma dels inversos dels nombres naturals —és a dir, la sèrie harmònica $1+\frac{1}{2}+\cdots+\frac{1}{n}\cdots$— és divergent. Els nombres triangulars i quadrats són, respectivament: $t_n = \frac{n(n-1)}{2}$, i $q_n = n^2$, $n \in \mathbb{N}$.

33. De fet, observà que $t_n = \frac{n(n-1)}{2} = \frac{2}{n-1} - \frac{2}{n}$. La resta és fàcil.

34. EULER, L. (1748), 168.

35. EULER, L. (1737). Quan $s \le 1$, la sèrie és divergent i $\zeta(s)$ no està definida; i, quan $s>1$, sí. A més va observar que $\zeta(s) = \frac{1}{1-(\frac{1}{2})^s} \times \frac{1}{1-(\frac{1}{3})^s} \times \cdots \times \frac{1}{1-(\frac{1}{p})^s} \times \cdots$, amb p primer. Això posa de manifest un cert lligam entre la funció zeta i els nombres primers i també l'enorme dependència que la funció té de l'infinit. (Vegeu BAYER I ISANT, P. (1984), 91-101; 106-115.) Recordem que Euler usà aquest fet per veure, d'una forma ben novedosa, l'existència d'infinits primers. Suposava que només n'hi havia una quantitat finita N i demostrava que $\prod_{k=1}^{n} \left(1-\frac{1}{p_k^s}\right)^{-1} = \sum_{k=1}^{\infty} \frac{1}{n^s}$. Per a $s = 1$, el

Ha entrat en escena la famosa *funció zeta*. L'ha introduït Euler i som aproximadament a l'any 1740.

2. Una fracció $\frac{m}{n} \in (0, 1)$, amb $\langle m, n\rangle = 1$, és una *fracció pròpia*. Per a cada nombre natural n, la *successió de Farey* d'ordre n, F_n, és la successió de totes les fraccions pròpies de denominador $\leq n$, amb la fracció $\frac{1}{1}$, en ordre creixent. Per exemple, les successions de Farey F_4 i F_7 són:

$$F_4 : \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{1}{1}; \quad F_7 : \frac{1}{7}, \frac{1}{6}, \frac{1}{5}, \frac{1}{4}, \frac{2}{7}, \frac{1}{3}, \frac{2}{5}, \frac{3}{7}, \frac{1}{2}, \frac{4}{7}, \frac{3}{5}, \frac{2}{3}, \frac{5}{7}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \frac{6}{7}, \frac{1}{1}.$$

No sabem qui fou el primer matemàtic que es preocupà per les successions de Farey, però les trobem ja en l'obra de C. Haros de 1802.[36] En 1816, John Farey (1766-1826) demostrà alguns dels resultats que, sense demostració, havia establert Haros. Alguns dels resultats d'Haros estableixen:

> (1) Si considerem tres fraccions successives $\frac{a}{d}$, $\frac{b}{e}$, $\frac{c}{f}$, aleshores $\frac{b}{e} = \frac{(a+c)}{(d+f)}$.
> (2) Dos termes successius, $\frac{a}{d}$, $\frac{b}{e}$, de F_n, compleixen $bc - ad = 1$.[37]

Ara, per a cada natural n, considerem el nombre $A(n)$ que compta els termes de la successió de Farey F_n.[38] Dividim el segment (0, 1) en $A(n)$ parts iguales. Els punts de divisió són $\frac{1}{A(n)}, \frac{2}{A(n)}, \ldots, \frac{A(n)-1}{A(n)}$. Els termes de la successió F_n no estan pas igualment distanciats. Per això resulta que no coincideixen necessàriament amb els punts de separació $\frac{k}{A(n)}$, $k = 0, \ldots, n-1$. Sigui d_k el valor amb què el k-èsim terme de la successió F_n difereix de $\frac{k}{A(n)}$ i $D(n)$ la suma de tots els valors $d_1, d_2, \ldots, d_{A(n)-1}$.[39]

---

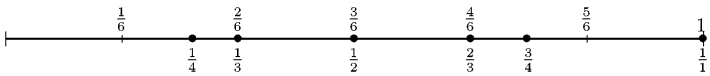
membre de l'esquerra fóra finit, mentre que el de la dreta és necessàriament infinit. No pot ser, doncs, que només hi hagi un nombre finit de primers. A banda de tot això, l'expressió anterior permet demostrar que la probabilitat que hi ha que, en agafar a l'atzar dos nombres naturals, siguin primers entre si, és $\frac{1}{\zeta(2)} = \frac{6}{\pi^2}$. Només cal observar que un nombre n no és múltiple de p amb probabilitat $\frac{1}{p}$.

36. Fou, però, Augustin-Louis Cauchy (1789-1857) qui les atribuí a Farey.

37. Aquests dos resultats —que podem constatar amb els exemples anteriors— són dependents: n'hi ha prou a demostrar-ne un per tenir la validesa de l'altre. (Vegeu FAREY, J. (1816), 386.)

38. Així $A(4)=6$, $A(5)=10$ i $A(7)=18$.

39. En el cas F_4, resulta que A(4) = 6. Els punts de divisió són $\frac{1}{6}$, $\frac{2}{6}$, $\frac{3}{6}$, $\frac{4}{6}$, $\frac{5}{6}$. Per tant,



$d_1 = \frac{1}{4} - \frac{1}{6} = \frac{1}{12}$, $d_2 = \frac{1}{3} - \frac{2}{6} = 0$, $d_3 = \frac{1}{2} - \frac{3}{6} = 0$, $d_4 = \frac{2}{3} - \frac{4}{6} = 0$, $d_5 = \frac{5}{6} - \frac{3}{4} = \frac{1}{12}$.

Per tant $D(4) = \frac{1}{12} + 0 + 0 + 0 + \frac{1}{12} = \frac{1}{6}$.

En un article de l'any 1924, J. Franel i E. Landau van investigar el comportament de la funció $D(n)$ quan n creix sobre els nombres naturals. Van considerar la proposició següent:[40]

> Si $r \in (\frac{1}{2}, +\infty)$, hi ha alguna constant C que, per a cada $n \in \mathbb{N}$, $D(n) < Cn^r$?

Ja hi tornarem!

El teorema dels nombres primers

Malgrat totes les qüestions, aparentment desordenades del comportament del conjunt dels nombres primers,[41] hi ha un cert ordre que podem posar de manifest a través de la funció $\pi(n)$, introduïda per Adrien-Marie Legendre (1752-1833) a *Essai sur la théorie des nombres* (1798).[42] La funció $\pi(n)$ compta els nombres primers $\leq n$ i Legendre observa que $\pi(n) \simeq \mathrm{Le}(n)$, on $\mathrm{Le}(n) = \frac{n}{\ln n - 1{,}083\,66}$.[43]

El jove Carl Friedrich Gauss (1777-1855), quan tenia quinze o setze anys, s'interessà per aquesta funció i observà que $\pi(n)$ es pot aproximar directament amb el nombre $\frac{n}{\ln n}$ i també amb el nombre Li(n), donat per la integral: $\mathrm{Li}(n) = \int_2^n \frac{1}{\ln n}\,dx$.[44]

La taula següent dóna els valors d'aquestes funcions.

n	$\pi(n)$	Le(n)	$\frac{n}{\ln n}$	Le(n)
1.000	168	172	145	178
10.000	1.229	1.231	1.086	1.246
100.000	9.592	9.588	8.686	9.630
1.000.000	78.498	78.534	72.382	78.628
10.000.000	664.579	665.138	620.420	664.918
100.000.000	5.761.455	5.769.341	5.428.681	5.762.209

40. Franel, J.-Landau, E. (1924).

41. Recordem les paraules d'Euler de 1751: «Els matemàtics han intentat fins avui, sense èxit, descobrir alguna mena d'ordre en la progressió dels nombres primers, i els resultats obtinguts ens porten a creure que és un misteri que l'esperit humà mai no aconseguirà de resoldre. Per convèncer-nos del que dic n'hi ha prou a contemplar les taules dels nombres primers —que alguns homes han aconseguit d'establir fins al nombre cent-mil— i ens adonarem, de seguida, que no hi ha cap ordre ni cap regla».

42. Vegeu l'article de Javier Echevarría, Echevarría, J. (1992).

43. Legendre ho trobà observant simplement les taules dels nombres naturals i dels nombres primers ≤ 400.000 i elegint la constant perquè l'aproximació fos prou acurada.

44. Recordem que això no és altra cosa que l'àrea limitada per la corba inversa de la funció logaritme neperià entre les abcisses 2 i n. És la funció *logaritme integral*. (Vegeu Gauss, C. F. (1849.)

De l'observació directa de la taula podríem deduir, d'una banda, que Li(n) aproxima $\pi(n)$ millor que no pas les altres dues i, d'una altra, que aquesta aproximació de $\pi(n)$ es fa sempre amb escreix. En 1896, Charles de la Vallée Poussin demostraria que la primera observació és certa, a partir d'un cert valor de n endavant. En canvi, l'any 1914, el matemàtic anglès John Edensor Littlewood (1885-1977) establiria la falsedat de la segona, tot mostrant que la diferència $\mathrm{Li}(n)-\pi(n)$ canvia de positiu a negatiu una *infinitat* de vegades quan n recorre els nombres enters positius.[45]

En 1896, els matemàtics francesos Jacques Hadamard (1865-1963) i C. de la Vallée Poussin (1886-1962), independentment, van demostrar el teorema següent:

> Quan n creix el valor $\frac{n}{\ln n}$ s'apropa més i més a $\pi(n)$.[46]

Aquest resultat és el que actualment es coneix amb el nom de *teorema dels nombres primers*. Cal que indiquem que el treball d'ambdós matemàtics francesos es fonamentava en els resultats obtinguts per l'alemany G.F. Bernhard Riemann (1826-1866) en 1859, en l'article «Über die Anzahl der Primzahlen unter einer gegebenen Grösse».

Aquest article —l'únic que va dedicar a la teoria de nombres— ha estat un article d'una importància difícil de descriure, però el que sí podem afirmar és que constitueix la base de la *teoría analítica dels nombres*, en la qual hom recorre als resultats de l'anàlisi per resoldre problemes d'aritmètica.

La hipòtesi de Riemann

Riemann estén la funció $\zeta(s)$ al camp dels nombres complexos.[47] Aquesta funció complexa rep el nom de *funció zeta de Riemann*.

És relativament fàcil veure que, quan $s=-2,-4,-6$, la funció $\zeta(s)$ s'anul·la. És a dir, els nombres enters parells negatius són *zeros* de la funció zeta de Riemann.

---

45. Tanmateix, les fites de n perquè la diferència canviï de signe per primera vegada són molt grans. I es desconeix la primera vegada que això passa exactament.

46. HADAMARD. J. (1896) i VALLÉE POUSSIN, CH. DE LA (1896).

47. Això no significa solament atribuir a s valors complexos $a+bi$, sinó que significa fer una extensió analítica de la funció, quelcom que pertany a l'anàlisi complexa. La fórmula que usà Riemann per estendre la funció zeta fou

$$\zeta(s)=\frac{\Pi(-s)}{2\pi i}\int_C \frac{(-x)^s}{e^x-1}\frac{dx}{x},$$

on la integral s'agafa damunt el camí C que va de ∞ vers 0, aturant-se ben a prop de 0, fa una volta al voltant del zero per tornar a 1 damunt de l'eix positiu de les abscisses. A més, $\Pi(s)=\lim_{N\to\infty}\frac{N!}{(s+1)(s+2)\cdots(s+N)}(N+1)^s$, per a tot enter no negatiu s.

La qüestió és:

> Hi ha d'altres zeros en el camp $\mathbb{C}$ dels nombres complexos?

La resposta és:

> N'hi ha una infinitat i *tots* tenen la part real entre 0 i 1.[48]

Bé, doncs, Riemann afirmà:

> Tots els zeros complexos de la funció $\zeta(s)$ tenen la part real igual a $\frac{1}{2}$; és a dir, són de la forma $s=\frac{1}{2}+bi$.[49]

Això és el que hom coneix com la *hipòtesi de Riemann* i encara avui és un problema obert.[50]

De fet, com veié Riemann, hi ha una relació molt íntima entre els zeros de la funció $\zeta(s)$ i la funció $\pi(n)$. Foren precisament aquests lligams els que precisaren Hadamard i De la Vallée Poussin per establir la seva demostració. Ara bé, només van precisar nexes vertaders i ben establerts i no pas de la hipòtesi de Riemann.

Atesa la complexitat del problema que hem encetat —del qual us parlarà a bastament la professora i amiga Pilar Bayer— l'única cosa que us diré és la següent:

> La hipòtesi de Riemann és equivalent a la pregunta que ens hem fet sobre la funció $D(n)$ generada per les successions de Farey.[51]

Hem vist, doncs, com indicàvem en començar, el lligam íntim que hi ha entre un problema aparentment aritmètic i l'anàlisi complexa.[52]

---

48. Si $s = a + bi$, amb a i b reals, la part real és a i la part complexa és b.

49. RIEMANN, G.F.B. (1859), traducció francesa de 1898, 168.

50. Tanmateix, l'any 1914 Godfrey Harold Hardy (1877-1947) va demostrar que, a la recta $x=\frac{1}{2}$, hi ha una infinitat de zeros de la funció $\zeta(s)$.

51. FRANEL, J. i LANDAU, E. (1924).

52. Hi ha una aproximació de $\pi(n)$ que inclou la funció $\zeta(n)$ de Riemann i està donada per mitjà d'una suma infinita:

$$R(n)=1+\sum_{k=1}^{\infty}\frac{1}{k\zeta(k+1)}\frac{(\ln n)^k}{k!}$$

3. IGNORÀBIMUS

Malgrat l'afirmació de David Hilbert (1862-1943) que mai, davant d'un problema matemàtic, haurem de dir «n'ignorem la resposta», Kurt Gödel (1906-1978) va demostrar que això no era així i que, efectivament, hi ha problemes dels quals ignorem si són certs i si són falsos.

Els antecedents o Programa de Hilbert

Com és ben sabut, el 1899 van aparèixer els *Grundlagen der Geometrie*, on Hilbert aconseguia, després de més de vint-i-tres segles, fixar l'axiomàtica de la geometria euclidiana.[53] Ho feia basant-se en el formalisme. Parafrasejant el que digué Hilbert una vegada, podem expressar-ho així:

> El que importa no són els objectes que intervenen en els axiomes —«podrien ser taules, cadires i gerres de cervesa». El que importa són els lligams que hi ha entre els objectes, i aquests lligams vénen imposats pels axiomes.

Sorgeixen aleshores diverses qüestions —d'índole lògica— que Hilbert anirà precisant a partir de 1915.[54] De fet són qüestions relatives a les propietats del sistema axiomàtic de la teoria.[55]

És una aproximació molt bona, com podem veure a la taula següent:

n	$\pi(n)$	$R(n)$
100 000 000	5 761 455	5 761 552
200 000 000	11 078 937	11 079 090
300 000 000	16 252 325	16 252 355
400 000 000	21 336 326	21 336 185
500 000 000	26 355 867	26 355 517
600 000 000	31 324 706	31 324 622
700 000 000	36 252 931	36 252 719
800 000 000	41 146 179	41 146 248
900 000 000	46 009 215	46 009 949
1 000 000 000	50 847 534	50 847 455

53. N'existeix una traducció castellana, força deficient, però que ha estat reeditada recentment, amb una introducció prou interessant de José Manuel Sànchez Ron. Per a una visió detallada de la situació de l'anàlisi dels fonaments de la matemàtica a l'època que ens ocupa, podeu consultar l'excel·lent recopilació de textos realitzada per Paolo Mancosu i, més divulgatiu, Dou, A. (1970).

54. Vegeu la tesi de Joan Rosselló, Rosselló, J. (2003), tercera part.

55. Hilbert creu fermament que, per a cada part essencial de la matemàtica, és possible establir una axiomàtica *ad hoc*.

Cal, en definitiva, que els axiomes siguin complets, independents i consistents.

La *completesa* imposa que el sistema sigui apte per demostrar tot allò que és vertader; és a dir, si σ designa un enunciat formal, la teoria ha de permetre que puguem demostrar σ o $\neg\sigma$, atès que una d'ambdues sentències ha de ser necessàriament vertadera.[56] És aquesta la raó per la qual Hilbert, en l'axiomàtica de la geometria —i també en la dels nombres reals—,[57] imposi l'axioma de completesa.[58]

La *independència* imposa que no introduïm axiomes superflus, és a dir, que siguin deduïbles de la resta.[59]

La *consistència* és la pedra de toc del sistema. Diu Hilbert:

> En la demostració de la no-contradicció dels axiomes (dels nombres reals) hi veig, al mateix temps, la prova de l'*existència* del conjunt dels nombres reals...[60]
>
> Cada vegada que usem el mètode axiomàtic se'ns fa present la necessitat d'establir una prova de la propietat de la no-contradicció.[61]
>
> Hi ha una condició única (en tot sistema axiomàtic), efectivament; però alhora és completament necessària i a ella està íntimament lligada l'aplicació del mètode dels elements ideals, i és *la demostració que el sistema està lliure de contradicció*.[62]

Per això no ens ha d'estranyar gens ni mica que, als *Grundlagen der Geometrie*, dediqui molt d'esforç a establir que els axiomes de la geometria són consistents o no contradictoris. S'iniciava així el camí de la reducció dels axiomes de la geometria als axiomes numèrics —primer als dels nombres reals i després als dels nombres naturals. En definitiva, s'obria, d'una forma incipient, la necessitat d'establir la consistència de l'aritmètica de Peano.

És aleshores quan, en l'obra de Hilbert, apareix el que s'anomena la *teoria de la demostració*, que consisteix bàsicament en l'anàlisi matemàtica de la teoria formal. I

56. Fixat el context d'interpretació, una sentència és vertadera o falsa. Mai no és ambigua.

57. HILBERT, D. (1900b), edició castellana de 1996, 244-249.

58. Aquest i l'arquimedianitat són axiomes de naturalesa diferents dels altres perquè no limiten l'àmbit dels ens formals als objectes sinó que precisen de col·leccions d'objectes.

59. Als *Grundlagen der Geometrie* dedica una part del capítol II a constatar la independència dels axiomes euclidians de la geometria.

60. HILBERT, D. (1900), edició castellana de 1996, 248. En revisions posteriors, Hilbert afegeix una nota: «Aquesta demostració exigeix essencialment mètodes de raonament nous i constitueix un problema capital de la meva teoria de la demostració» (i remet als seus treballs sobre la teoria de la demostració de 1925, 1928).

61. HILBERT, D. (1928), edició castellana de 1996, 299.

62. HILBERT, D. (1925), edició castellana de 1996, 285. L'èmfasi és del mateix Hilbert.

Hilbert imposa que aquesta anàlisi o metateoria ha de ser de «caràcter finitari», com la mateixa teoria.[63] Això, però, no es pot fer fora «del paradís creat per Cantor i del qual mai ningú no ens expulsarà».[64] Tanmateix, el paradís de Cantor «axiomatitzat per Zermelo presenta també el problema de la no contradicció».[65]

De fet, a aquesta metodologia cal afegir el *concepte semàntic de veritat*, introduït per Alfred Tarski (1902-1983) en el seu famós article de 1931 —«Sobre el concepte de veritat en els sistemes deductius formalitzats»—[66], que obria la porta a la teoria de models. Bàsicament fa referència a la validesa d'un objecte formal en una estructura algèbrica, en la qual interpretem d'alguna manera els símbols formals —relacionals i funcionals— per mitjà de relacions i funcions de l'estructura, les constants com a objectes fixos i les variables com objectes arbitraris de l'estructura.

Així l'expressió formal —de primer ordre— es converteix en una expressió que fa referència als objectes de l'estructura que pot ser vertadera o falsa. Si esdevé vertadera, diem que l'estructura n'és un *model*.[67] La validesa o falsedat al si de l'estructura l'hem d'establir, tanmateix, dins la teoria axiomàtica de conjunts, que és allà on, si no és contradictòria, en paraules de Hilbert, «existeix» l'objecte ideal.[68]

Disposem, doncs, de dues menes de validesa matemàtica: la *sintàctica* —que depèn de l'axiomàtica del sistema— diu que és valid allò que s'ha demostrat, i la *semàntica* —que depèn del model de l'axiomàtica—, diu que és vàlid allò que, un cop interpretat, és vertader en el model.

En aquest context és on apareix Kurt Gödel i on elabora la seva tesi doctoral establint el famós *teorema de completesa* de les teories formals de primer ordre:[69]

> Les sentències vàlides sintàcticament i les sentències vàlides semànticament, en tots els models possibles del sistema axiomàtic, són les mateixes; és a dir, la validesa sintàctica i la validesa semàntica universal coincideixen.[70]

63. En síntesi: els raonaments —com les demostracions formals— han de ser finits.

64. Hilbert, D. (1925), edició castellana de 1996, 273.

65. Hilbert, D. (1925), edició castellana de 1996, 311-312.

66. Tarski, A. (1972), I, 157-269.

67. $\forall x \exists y (x * y = e)$, la podem llegir en $\mathbb{Z}$ fent $e = \mathbf{0} \in \mathbb{Z}$ i interpretant $*$ com la suma + de $\mathbb{Z}$. És vertadera! En canvi, si la interpretem en $\mathbb{Z}$, però fem $e = \mathbf{1} \in \mathbb{Z}$ i interpretem $*$ com la multiplicació $\cdot$, és falsa.

68. No entraré pas a reproduir la discussió que mantingueren Hilbert i Gottlob Frege (1848-1925) sobre «que és primer la no-contradicció del sistema axiomàtic o l'existència d'un model?». Primer ha de ser no contradictori —diu Hilbert— i n'existirà un model; no, primer ha d'existir el model —diu Frege— i aleshores el sistema és no contradictori. Vegeu el paràgraf «Algunes reaccions als *Grundlagen* de Hilbert» i la nota 56 de la introducció de Sànchez Ron. Tanmateix, en la teoria de models de la lògica de primer ordre, es pot demostrar que «una teoria és consistent si, i només si, té un model en el sentit de Tarski».

69. Vegeu Pla, J. (1991), 80-84.

70. Gödel, K. (1930), a Gödel, K. (1989), 23-37.

La qüestió és, doncs:

> És possible, com creia Hilbert, axiomatitzar tot allò que és vàlid en un únic model?

Una altra qüestió que suggereix Hilbert és que els teoremes d'una teoria formal són, d'alguna manera, decidibles. És el que es coneix com el *Entscheidungsproblem*.

El teorema d'incompletesa de Gödel

Arribem així al nucli de la qüestió relativa a l'«Ignoràbimus» de Hilbert. Tot consisteix a veure simplement que allò que és veritat en $\mathfrak{N}=\langle\mathbb{N};\mathbf{0};\mathbf{s};+;\cdot;\leq\rangle$ no ho podrem mai axiomatitzar en primer ordre, si imposem que el conjunt d'axiomes sigui *raonable*.[71]

Per aconseguir-ho, Gödel distingeix amb tota claredat els tres llenguatges que entren en joc en tot problema d'anàlisi de la demostració d'una teoria formal. En el cas associat al model $\mathfrak{N}$, els llenguatges són el català, $\mathfrak{N}$ i el llenguatge formal. En síntesi,[72]

el llenguatge formal	el llenguatge del model $\mathfrak{N}$	el català
$\forall v_1\exists v_2(v_1+v_2=\mathbf{0})$	$\forall m\in\mathbb{N}\exists n\in\mathbb{N}(m+n=0)$	cada nombre natural té un oposat natural

L'habilitat de Gödel consisteix a passar del llenguatge formal a la intrepretació lingüística catalana que, a través de la *gödelització*,[73] proporciona funcions i relacions del conjunt $\mathbb{N}$. Per exemple, a la sentència $\sigma:=\forall v_1\exists v_2(v_1+v_2=\mathbf{0})$ li correspon un nombre natural n_σ. Aleshores l'expressió «σ és una sentència» la convertim en «$n_\sigma\in S$, on S és el subconjunt de $\mathbb{N}$, format pels nombres de totes les sentències».[74]

La qüestió ara és la següent:

71. Per conjunt raonable, vegeu la pàgina 146.

72. Vegeu PLA, J. (2002), 195-203.

73. Atribuint un nombre natural adequat a cada símbol formal i lògic, podem atribuir un nombre natural a les expressions lògiques —fórmules, sentències, demostracions, fórmula substituïda, etc.

74. Tot teorema σ d'una certa teoria formal té una demostració formal $\sigma_1,\ldots,\sigma_{n-1}$. Disposem, doncs, de parelles de nombres naturals $\langle n, d_n\rangle$, formades pel nombre n d'una sentència σ i pel nombre d_n d'una de les demostracions de la sentència. Totes aquestes parelles $\langle n, d_n\rangle$ formen una relació binària $Dem\subseteq\mathbb{N}\times\mathbb{N}$, definida per

$Dem=\{\langle n, d_n\rangle: n \text{ i } d_n \text{ són els nombres d'una sentència } \sigma \text{ i d'una demostració de } \sigma\}$.

Les expressions en català referents a tot allò que fem en el procés de construcció lògica, les podem expressar dins del sistema formal? És a dir, podem, d'alguna manera, parlar dels subconjunts, relacions i funcions de $\mathbb{N}$ que hagin anat sorgint, usant el llenguatge formal?[75]

La resposta és afirmativa i això és el que m'interessa de retenir. Sigui **P** la teoria axiomàtica de primer ordre de l'aritmètica de Peano.[76] Aleshores tenim el concepte de *conjunt*, *relació* i *funció representables*:

Sigui A un subconjunt de $\mathbb{N}$. Diem que A és *representable* en la teoria axiomàtica **P** si, i només si, existeix una fórmula $\varphi(v_1)$, amb una sola variable lliure, que, si abreujem el terme $\boldsymbol{s}^{n)}\boldsymbol{s}(\mathbf{0})$ per $\boldsymbol{n}$, compleix

$\vdash_{\mathbf{P}}\varphi(\boldsymbol{n})$ —és a dir, $\varphi(\boldsymbol{n})$ és demostrable en la teoria **P**—, si $n\in A$,

$\vdash_{\mathbf{P}}\neg\varphi(\boldsymbol{n})$ —és a dir,$\neg\varphi(\boldsymbol{n})$ és demostrable en la teoria **P**—, si $n\notin A$.[77]

La qüestió és, aleshores:

És possible identificar aritmèticament els subconjunts representables de $\mathbb{N}$?

La resposta de Gödel és afirmativa.

Un conjunt $A\subseteq\mathbb{N}\to\mathbb{N}$ és **P**-*representable* si, i només si, la funció 1_A és recursiva.[78]

Cal, doncs, el concepte de funció recursiva.

75. Atenció! El llenguatge formal que estem usant —el de l'aritmètica de primer ordre— només pot descriure un subconjunt per mitjà d'una fórmula amb una variable lliure $\varphi(v)$, etc. Atès que, en un llenguatge lògic, normalment, solament disposem d'una quantitat numerable de fórmules, només permetrà parlar, al si del llenguatge formal, d'una quantitat numerable de subconjunts de $\mathbb{N}$.

76. Els axiomes són els de Peano inclòs el d'inducció. Ara bé, en el llenguatge formal de l'aritmètica no podem parlar de subconjunts de $\mathbb{N}$. Solament podem parlar de fórmules $\varphi(v_1)$, amb una variable lliure. Vegeu la nota anterior.

77. GÖDEL, K. (1934), edició castellana de 1989, 183. Ens limitarem als subconjunts perquè les relacions essencialment es defineixen de forma anàloga. Òbviament, en el cas de les funcions, cal afegir-hi la funcionalitat de la fórmula.

78. Pel que fa al teorema d'incompletesa només calen les funcions recursives primitives. (Vegeu PLA, J. (2002), 232-241.)

Una funció $f : A \subseteq \mathbb{N} \to \mathbb{N}$ és *recursiva* si, i només si, pertany a la classe de les funcions generades per les constants, la funció següent, les projeccions, tancada per les operacions de composició, recursió i minimalització.[79]

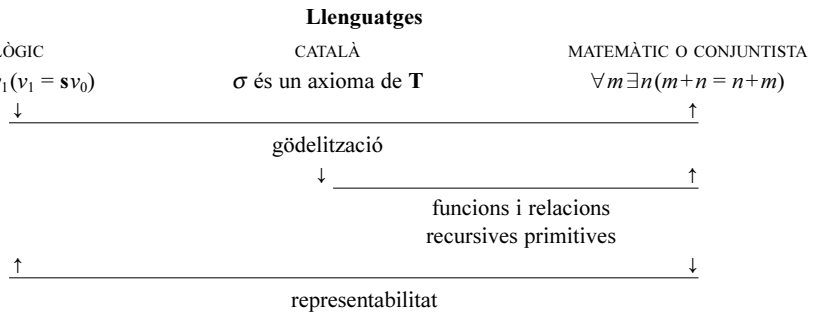
Amb aquestes eines, Gödel demostra finalment el *teorema d'incompletesa* que estableix formalment que, al si de les teories formals de primer ordre, hi ha «Ignoràbimus»:

Si **P** és consistent, hi ha una sentència σ_g que és indemostrable en **P** i la seva negada $\neg\sigma_g$, també.[80]

A més, si afegim la sentència indecidible σ a la teoria **P**, la teoria $\mathbf{P}+\sigma$ també és incompleta.[81] Aquest teorema —que Gödel estableix per a l'aritmètica de Peano de primer ordre— val també en la teoria de conjunts de Zermelo-Fraenkel i, atès que aquest és l'àmbit natural actual de resolució de problemes, la matemàtica té situacions en les quals cal dir «Ignoràbimus».[82]

A més d'això, Gödel indica que els raonaments metalògics que s'han hagut de fer per provar el teorema d'incompletesa poden ser reproduïts al si de la teoria **P** de Peano, i alhora és possible introduir una sentència $\sigma_{\text{Consistència}}$ que, al si de la teoria, estableix la seva consistència. Aleshores:

79. GÖDEL, K. (1934), edició castellana de 1989, 168-169; 183-186. Tot això podem sintetitzar-ho gràficament de la forma següent:



80. GÖDEL, K. (1931), edició castellana de 1989, 85-87.

Pensem el que passa, per exemple, amb el postulat de les paral·leles respecte de la resta de postulats de la geometria, o amb l'axioma de l'elecció o de la hipòtesi del continu respecte de la resta de postulats de la teoria de conjunts Zermelo-Fraenkel.

81. Per assolir una teoria completa cal una axiomàtica no raonable. És a dir, el conjunt de nombres de Gödel dels axiomes no és recursiu o, equivalentment, no és decidible. En concret, donada una sentència formal no sabem si és un axioma o no i, a més, no disposem de cap mètode per saber-ho.

82. GÖDEL, K. (1931), edició castellana de 1989, 86-87.

Si **P** és consistent, $\nvdash_{\mathbf{P}} \sigma_{\text{Consistència}}$.[83]

Així Gödel resolia negativament el segon problema de Hilbert de 1900:

Demostrar que els sistemes formals de primer ordre no són contradictoris; és a dir, amb un nombre finit de passos lògics basats en els axiomes del sistema formal no podem arribar a contradicció.[84]

4. LA DECIDIBILITAT

Amb la definició de funció recursiva s'obria la porta cap a un problema matemàtic nou: el *problema de la decidibilitat*.

Quan podem dir que coneixem un conjunt A de nombres naturals?[85]

Quan parlem de conjunts, sembla que allò que importa és conèixer-los. Però:

La resposta no és unívoca. Hi ha dues maneres diferents de donar un conjunt de nombres naturals. La primera consisteix a disposar d'un algorisme que ens digui si un cert element $n \in \mathbb{N}$, pertany o no al conjunt A. En direm els *conjunts decidibles* o *raonables*. La segona, més feble, consisteix a disposar d'un algorisme que vagi construint, un després d'un altre, els elements del conjunt A. Són els *conjunts construïbles* o *llistables*. De fet, tot conjunt decidible és llistable. La qüestió és saber si hi ha conjunts llistables que no siguin decidibles.

Per tal de comprendre millor aquestes definicions, introduirem un cert formalisme. Sigui A un subconjunt de $\mathbb{N}$. Suposem que disposem d'una certa funció $f: \mathbb{N} \to \mathbb{N}$ que $A = \{f(n) : n \in \mathbb{N}\}$. De fet, aquest conjunt és un conjunt construïble. La funció f construeix $f(1)$, després $f(2)$, després $f(3)$, etc.[86] Mai, però, no el tenim totalment construït. En coneixem el criteri de formació, però no coneixem el conjunt en la seva totalitat.

83. GÖDEL, K. (1931), edició castellana de 1989, 87. De fet, tot rau a veure que, si **P** és consistent, $\vdash_{\mathbf{P}} \sigma_{\text{Consistència}} \to \sigma_g$.

84. HILBERT, D. (1900b), edició castellana de 1996, 274-276. Val la pena indicar que la primera demostració del teorema fou obtinguda per Paul Bernays (1888-1977), un deixeble de Hilbert, i publicada a HILBERT, D. i BERNAYS, P. (1939), 285-328.

85. Cal recordar el desastre lògic que suposà per Gottlob Frege (1848-1925) el convenciment que una «propietat defineix un conjunt»?

86. Pensem, per exemple, en el conjunt dels nombres primers.

Suposem, en canvi, que disposem d'un criteri —d'un *mètode* o *algorisme*— que permeti respondre «sí» o «no» a la pregunta:

> Per a cada nombre natural $n \in \mathbb{N}$, n pertany al conjunt A?

Aquesta pregunta la podem matematitzar usant la *funció característica del conjunt A*, és a dir, la funció 1_A, definida per

$$1_A(n) = \begin{cases} 1, & \text{si } n \in A, \\ 0, & \text{si } n \notin A. \end{cases}$$

Aleshores diem que el conjunt A és *decidible*, quan la funció 1_A és computable; és a dir, hi ha un mètode matemàtic que permet conèixer-la.

Sorgeix aleshores la qüestió:

> Què hem d'entendre per funció computable? Què és un mètode o sistema de computació? És possible definir matemàticament la idea intuïtiva d'algorisme o de computació?

Tots nosaltres, des de menuts, hem après alguns algorismes de càlcul. Per exemple, l'algorisme que serveix per multiplicar dos nombres naturals expressats en el sistema de numeració indoaràbic, l'algorisme d'Euclides per determinar el màxim comú divisor de dos nombres naturals, etc. Però, de fet, els matemàtics no van disposar d'una definició d'algorisme fins que, l'any 1936, Alan Turing (1912-1954) va introduir la *màquina de Turing*.

La idea de màquina de Turing és realment senzilla. Una màquina de Turing $\mathcal{M}$ consta d'una cinta il·limitada, dividida en cel·les individuals. Cada cel·la conté un pal com ara $|$, o bé resta buida.[87] Té un cap lector que, en cada moment de la computació, llegeix el contingut d'una única cel·la. Finalment, consta d'una capsa negra que pot trobar-se en un estat intern q_i, $i = 0, 1, 2, \ldots, s$, d'una col·lecció finita d'estats interns possibles.

La màquina de Turing $\mathcal{M}$ actua de la manera següent. En la cinta s'escriuen k nombres naturals, separats entre si per una cel·la buida. La resta de cel·les estan buides. El cap lector es col·loca a la dreta del pal $|$ escrit damunt de la cinta més a la dreta de tots. L'estat intern inicial és l'estat q_0. En cada moment de la computació, la màquina pot fer cinc accions, d'acord amb l'estat intern en què es troba i amb què llegeix el cap lector en la cel·la en què es troba col·locat. Aquestes accions són:

87. No s'ha de confondre el nombre 0 amb una cel·la buida. A una cel·la buida no hi ha cap nombre. Si volem representar el nombre 0, ho hem de fer usant pals. N'usem un de sol. Això fa que, en una màquina de Turing, un nombre natural n es representi usant $n+1$ pals.

— Borra el que hi ha escrit, deixant la cel·la buida (B), i es col·loca en un dels estats interns possibles.

— Borra el que hi ha escrit i escriu un pal (|), i es col·loca en un dels estats interns possibles.

— Es mou una cel·la cap a la dreta i es col·loca en un dels estats interns possibles.

— Es mou una cel·la cap a l'esquerra i es col·loca en un dels estats interns possibles.

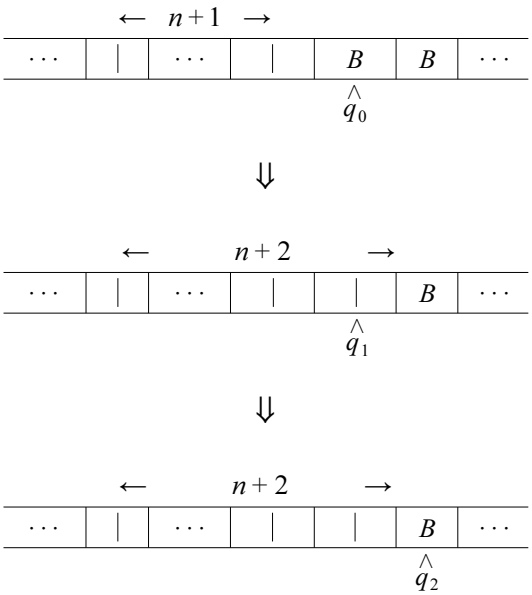
— S'atura, perquè no pot seguir actuant.

La millor manera de comprendre el significat i el comportament d'una màquina de Turing és donar-ne exemples senzills. De fet, una màquina de Turing és una matriu finita.

1. La màquina de Turing que descriu la funció de *pas al següent*. Considerem la matriu

$$\begin{array}{cccc} q_0 & B & | & q_1 \\ q_1 & | & D & q_2. \end{array}$$

Aquesta màquina actua damunt una cinta que contingui el nombre natural n —és a dir, damunt la cinta amb $n+1$ pals— quan el cap lector es troba llegint la primera cel·la buida a la dreta del | que hi ha més a la dreta en estat intern inicial q_0 així:



situació inicial

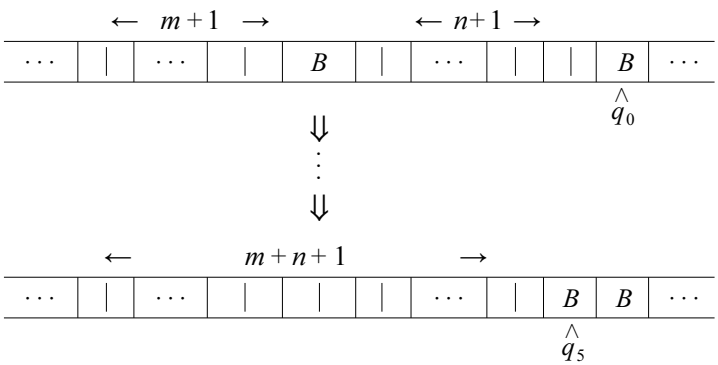
computació primera

computació segona

2. Considerem la màquina que *suma dos nombres naturals*. És la màquina:

q_0	B	E	q_1
q_1	$\vert$	E	q_1
q_1	B	$\vert$	q_2
q_2	$\vert$	D	q_2
q_2	B	E	q_3
q_3	$\vert$	B	q_4
q_4	B	E	q_4
q_4	$\vert$	B	q_5

L'apliquem als nombres naturals m i n separats per un blanc, amb el cap lector a la cel·la de la dreta del darrer $\vert$ de la dreta de tots, i en estat intern q_0.



3. Tanmateix, hi ha màquines de Turing que no paren mai i, per tant, no computen res de res. Per exemple, la màquina descrita per la matriu

$$q_0 \;\; B \;\; B \;\; q_0.$$

La màquina de Turing és l'eina que permet introduir el concepte de *funció computable*.

> Una funció parcial $f : \mathbb{N} \to \mathbb{N}$ és *computable de Turing* si hi ha una màquina de Turing que, col·locada en estat incial q_0, a la cel·la buida a la dreta del pal de la dreta que simbolitza el número natural n, s'atura precisament quan, a la cinta, hi ha escrit el nombre $f(n)$, si $f(n)$ està definit, i no s'atura mai més, si $f(n)$ no està definit.[88]

88. TURING, A. (1936).

Ara ja podem descriure amb tota mena de rigor els subconjunts de $\mathbb{N}$ que va usar Gödel en la demostració dels seus teoremes i que, en treballs posteriors, va generalitzar.

Un subconjunt A de $\mathbb{N}$ és *recursiu* quan la seva funció característica 1_A és computable amb una màquina de Turing.

En canvi, un subconjunt A de $\mathbb{N}$ és *recursivament enumerable* (r.e.) quan és el recorregut $\{f(1), f(2), \ldots, f(n), \ldots\}$ d'una funció f computable amb una màquina de Turing.[89]

Un resultat realment important és el següent:

Hi ha subconjunts de $\mathbb{N}$ que són recursivament enumerables, però en canvi no són recursius.[90]

Per exemple, si s'enumeren tots els teoremes de la teoria de Peano **P** d'una forma adequada, s'obté un conjunt de nombres naturals. Concretament, el conjunt

$$\text{Teoremes}(\mathbf{P}) = \{n \in \mathbb{N} : n \text{ és el número d'un teorema de la teoria } \mathbf{P}\}.$$

Aquest conjunt és recursivament enumerable, però en canvi no és recursiu. Dit d'una altra manera, el conjunt de teoremes de la teoria **P** no és decidible. Aquí s'estableix, en forma negativa, un altre problema que Hilbert havia plantejat: l'*Entscheidungsproblem*, segons el qual els teoremes d'una teoria eren un conjunt decidible.[91]

Turing va introduir la *màquina universal de Turing U*. La màquina U és una matriu que davant una parella de nombres naturals m, n, actua de la forma següent:

Si m és el nombre de Gödel d'una màquina de Turing M, aleshores M actua sobre n. El resultat $U(m,n)$ és, per tant, el resultat $M(n)$.

Si m no és el nombre de Gödel de cap màquina de Turing, $U(m, n)$ val zero.

La màquina universal U, si disposa de prou temps, computa el que computaria qualsevol altra màquina de Turing M.[92]

Una altra qüestió que plantegen les màquines de Turing és el *problema de l'aturada*.

89. Els conjunts recursivament enumerables els va introduir Emile Post (1944).
90. Post, E. (1944).
91. Aquesta impossibilitat fou establerta per Alonzo Church (1903-1995) en 1936.
92. Les màquines de Turing són matrius finites i admeten una gödelització.

La idea és la següent. Enumerem totes les màquines de Turing: M_1, M_2, M_3, …, M_k, … Definim la funció $h : \mathbb{N} \to \mathbb{N}$ de la forma següent:

$$h(m, n) = \begin{cases} 1, & \text{si la màquina de Turing } M_m \text{ s'atura davant l'entrada } n; \\ 0, & \text{altrament.} \end{cases}$$

Aquesta funció —que es coneix amb el nom de *funció d'aturada*— no és computable amb una màquina de Turing.[93] Les màquines de Turing ens proporcionen, doncs, un problema que no és decidible.[94]

S'ha resolt una gran quantitat de problemes però, alhora, se n'han plantejat de nous.

La complexitat algorísmica: una altra mena d'Ignoràbimus

Un cop definit matemàticament el concepte d'algorisme, podem saber amb una certa precisió quan un problema és resoluble i quan no ho és. Ara la qüestió és:

> Quan un problema és resoluble, ho és efectivament? És a dir, la computació teòrica és realitzable en un temps raonable?

Imaginem dos problemes molt vinculats, però que, de fet, des del punt de vista de la computació algorísmica, són força diferents.

1. El primer consisteix a multiplicar dos nombres primers p_1 i p_2. El temps —el nombre de passos que ha de fer la màquina de Turing que multiplica dos nombres— és relativament anàleg a la complexitat de les dades introduïdes.

2. Imaginem, en canvi, que ens donen un nombre natural N del qual sabem que és el producte de dos nombres primers, i ens demanen els factors que el componen. El temps que cal és molt gran comparat amb el nombre N.[95] És possible donar productes de

93. Una exposició divulgativa amena i brillant la trobareu a DAVIS, M. (2000), edició castellana de 2002.

94. El *problema de les paraules*, que és, de fet, un problema referent a grups finits generats, és un altre problema indecidible.

95. L'any 1977, Martin Gardner (1914-) va demanar la descomposició en producte de dos nombres primers del nombre

N = 114 381 625 757 888 867 669 235 779 976 146 612 010 218 296 721 242 362 562 561 842 935 706 935 245 733 897 830 597 123 563 958 705 058 989 075 147 599 290 026 879 543 541.

dos factors primers que la seva descomposició, usant en la tasca tots el ordinadors del món, necessités més temps que l'edat de l'univers.[96]

Aquests dos exemples ens permeten de distingir els problemes **P** i **NP**. Per entendre millor els problemes **NP** introduirem el *problema del viatjant de comerç*. Aquest viatjant ha de visitar, per exemple, 50 indrets. L'ordre amb què els visiti no importa gens, sempre que faci totes les visites. El que sí interessa, per tal d'evitar esforços innecessaris i de minimitzar costos, és que el recorregut final sigui mínim, evitant, a més, passar dues vegades pel mateix indret. Aquest és un problema finit. El viatjant, en principi, té només 50! possibilitats. Les calculem totes i, per a cada una d'elles, avaluem la longitud del trajecte. Ara bé, la funció factorial $n!$ és exponencial; és a dir, creix més de pressa que la funció 2^n, però menys de pressa que la funció n^n.[97] Això fa que la computació dels casos possibles sigui de tipus exponencial i no pas de tipus polinòmic. Concretant:

> Una màquina de Turing treballa *en temps polinòmic* quan existeixen dos enters fixos C i k que, a partir de dades de longitud n, el càlcul s'efectua, com a màxim, en Cn^k etapes i això per a cada dada n.
>
> Una màquina de Turing treballa *en temps exponencial* quan requereix 2^n, 3^n, n^n, o $n!$ etapes per acabar la computació, quan les dades tenen una longitud n.[98]

Els problemes de temps polinòmic generen la classe **P**. Hi ha una altra classe de problemes, la classe **NP**, que conté la classe **P**. La formen els problemes *de temps polinòmic no determinista*.[99] Si usem màquines de Turing deterministes,[100] la suma i la multiplicació de dos nombres naturals és una operació que s'efectua en un temps polinòmic. En canvi, el còmput del problema del viatjant s'efectua en un temps exponencial.[101]

Van passar disset anys —fou el 26 d'abril de 1994— abans que un equip de sis-cents voluntaris aconseguís trobar-ne un dels factors. Aquest nombre, però, és petit. És de l'ordre 10.129.

96. També podríem plantejar-nos el problema de *criptar*, del qual us parlarà amb tota mena de detalls Anna Rio. Vegeu el text divulgatiu de Pla, J. (2000), 130-144.

97. Recordem la fórmula de Stirling que dóna una aproximació de $n!$ (Vegeu la nota 27.)

98. Amb rigor, quan parlem de problemes que fan referència a nombres naturals, la longitud indica el nombre de dígits dels nombres; en canvi, el nombre de passos fa referència al nombre de bits de les computacions.

99. Són problemes que poden ser resolts en temps polinòmic usant màquines de Turing no deterministes —és a dir, l'estat intern i el que hi ha a la cel·la llegida pel cap lector *no determinen* l'acció següent. La màquina té capacitat de decisió a l'atzar.

100. Com les que hem descrit a la pàgina 16.

101. El més curiós d'aquest problema és que, per a un autèntic viatjant de comerç, recórrer 25 localitats no és pas una qüestió impensable, ben al contrari.

Tanmateix, el problema del viatjant de comerç podria ser polinòmic. Avui, però, això és una qüestió oberta.[102]

El problema que consisteix a descompondre el nombre N en el producte de dos factors primers és també de tipus **NP**. L'algorisme mira a l'atzar els números p i q i verifica si $p \times q = N$. Tindrem una resposta en temps polinòmic, i una de les conjectures a l'atzar donarà la solució correcta.

La qüestió que es planteja és saber si $\mathbf{P} \neq \mathbf{NP}$.

La dificultat rau en el fet que caldria provar que un cert problema de la classe **NP** no es pot resoldre en temps polinòmic. I això obligaria a considerar tots els algorismes que permeten resoldre el problema i veure que cada un d'ells és ineficaç. Un avenç en aquesta línia l'aportà Stephen Cook (1943-) l'any 1971 en un treball relatiu a l'eficàcia dels algorismes.[103] Va aconseguir demostrar que era altament improbable que certs problemes de tipus **NP** fossin resolubles amb un algorisme que els resolgués en temps polinòmic. De fet, va provar que una certa subclasse de problemes **NP** és **NP**-completa. És a dir, si un d'aquest problemes és resoluble en temps polinòmic —és de tipus **P**— aleshores qualsevol altre problema de tipus **NP** també ho és, de tipus **P**.[104]

Malgrat tots aquests avenços, ara com ara, la qüestió central —saber si $\mathbf{NP} = \mathbf{P}$, o si $\mathbf{NP} \neq \mathbf{P}$— és una qüestió oberta. No s'ha trobat cap demostració, ni directa ni indirecta, que permeti donar una resposta. No sabem si la igualtat $\mathbf{NP} = \mathbf{P}$ és certa o falsa, o si, com succeeix amb la hipòtesi general del continu, és independent de la teoria formal de conjunts. És un dels misteris no resolts de la matemàtica del segle XX. De tot això us en parlarà molt més àmpliament Albert Atserias.

Si la hipòtesi de Riemann (generalitzada) és certa, aleshores per a cada enter compost n hi ha una base b, amb b primer i $< 2(\ln n)^2$, respecte de la qual n falla el test de Miller.[105]

102. Tanmateix és de temps polinòmic no determinístic.

103. COOK, S. (1971).

104. D'altres problemes **NP**-complets són, per exemple, entre d'altres, el *colorejament de mapes*: «És possible colorejar un mapa amb tres colors de manera que mai dos països veïns tinguin el mateix color?». El *residu quadràtic*: «Donats els enters positius a, b, c, amb $a<b$, existeix un enter positiu $x<c$ que $x^2 \equiv a$(mòd b)?». Les *equacions diofàntiques quadràtiques*: «Donats els enters positius a, b, c, existeixen enters positius x i y que $ax^2 + by = c$? (Vegeu DEVLIN, K. (1998), 238.)

105. La hipotèsi de Riemann generalitzada és tècnicament més elaborada que la que establí Riemann. Fa referència als zeros de la funció zeta dels cossos de nombres. (Vegeu SERRE, J.-P. (1981).)

Això fa que —amb la hipòtesi de Riemann— disposem d'un test de primalitat de tipus **P**.[106]

5. EL PROBLEMA DIOFÀNTIC

Amb tot aquest utillatge podem donar resposta a un altre dels problemes plantejats per Hilbert el 1900, el *problema diofàntic*:

> Donada una equació diofàntica amb un nombre arbitrari d'incògnites i amb coeficients numèrics enters racionals, *idear un procediment amb el qual puguem saber, amb un nombre finit de passos, si l'equació admet una solució en enters racionals*.[107]

Aquí, davant la dificultat enorme que presenta la resolució concreta de les equacions diofàntiques, Hilbert demana un mètode que digui, d'avant mà, si té solució o no.[108]

Els primers passos cap a la resolució negativa del problema els van donar Martin Davis (1926-) i Julia Robinson (1919-1985) el 1950 i 1952, respectivament.

> Un conjunt $D \subseteq \mathbb{N}$ és *diofàntic* si existeix un polinomi $P(X_0, X_1, \ldots, X_m)$ que $n \in D$ si, i només si, $P(n, X_1, \ldots, Xm) = 0$ té solucions enteres positives.[109]

Davis va establir, a més, la caracterització següent:

> Tot conjunt r.e. $E \subseteq \mathbb{N}$ es caracteritza per:
> $n \in E$ si, i només si, $\exists y \forall k \leq y\ P(n, y, k, X_1, \ldots, X_m) = 0$
> té solucions $x_1, \ldots, x_m \in \mathbb{N}$.

106. Vegeu DAVIS, D.M. (1993), 246.

107. HILBERT, D. (1900), edició castellana de 1996, 87.

108. En el cas de les equacions polinòmiques, el teorema fonamental de l'àlgebra garanteix l'existència de solucions en $\mathbb{C}$. Una qüestió ben diferent és resoldre-les.

109. Això val també per a relacions ℓ-àries, $R(n_1, \ldots, n_\ell)$. Els polinomis que les defineixen són de la forma $P(Y_1, \ldots, Y_\ell, X_1, \ldots, X_m) = 0$.

En resulta que, si les relacions diofàntiques són tancades per quantificació universal fitada, els conjunts diofàntics i els conjunts r.e. *coincideixen*.

Julia Robinson introdueix els conjunts exponencial diofàntics i demostra que coincideixen amb els conjunts r.e.[110] S'obté el *teorema de Julia Robinson*:

> Els conjunts exponencial diofàntics i els conjunts r.e. coincideixen.[111]

L'objectiu és, doncs, demostrar que

> la funció exponencial $w = u^v$ és diofàntica,

ja que aleshores els conjunts r.e. coincidiran amb els conjunts diofàntics.[112]

Julia Robinson dóna una condició suficient perquè això es compleixi:

> Perquè les funcions exponencials diofàntiques siguin diofàntiques, n'hi ha prou que existeixi una relació diofàntica $D(u,v)$ que
> (1) $D(u,v)$ implica $v \le u^u$;
> (2) per a tot k existeix $D(u,v)$ que $v > u^k$.[113]

En 1971, Yuri Matijasevicz (1947-) estableix el teorema següent:

> La condició de Robinson és vertadera: l'exponencial $w = v^u$ és diofàntica.[114]

Així queda establert definitivament el problema deu de Hilbert, en la forma següent:

---

110. Un conjunt ED és *exponencial diofàntic* si, i només si, existeix un polinomi $P(X_0, X_1, \ldots, X_m)$ que $n \in ED$ si, i només si, $P(n, x_1^{y_1}, \ldots, x_m^{y_m}) = 0$ té solucions enteres positives. (Val també per a les relacions.)

111. Val també per a relacions.

112. És a dir, cal veure que hi ha un polinomi, amb els coeficients enters, $P(Y_1, Y_2, Y_3, X_1, \ldots, X_n)$, que $p = m^n$ si, i només si, l'equació diofàntica $P(m, n, p, X_1, \ldots, X_n) = 0$ té solucions enteres $x_1, \ldots, x_n$.

113. ROBINSON, J. (1952). Veure que la funció exponencial $w = u^v$ és diofàntica es redueix, doncs, a establir la condició de Robinson.

114. MATIJASEVICZ, Y. (1971). Estableix la hipòtesi de Julia Robinson, considerant el conjunt: $D = \{\langle u,v\rangle : v = a_{2u} \wedge u \ge 2\}$, on a_i són els nombres de la successió de Fibonaci $a_0 = 1$, $a_1 = 1$, $a_{i+2} = a_i + a_{i+1}$.

No hi ha cap algorisme que permeti decidir si una equació diofàntica arbitrària és resoluble en $\mathbb{N}$.[115]

D'altres problemes lligats amb els anteriors

A la conferència de 1900, Hilbert, de passada o entre els vint-i-tres problemes, n'indica d'altres que estan lligats amb el caràcter recursivament enumerable o diofàntic de certes relacions. Són:

1. *El teorema darrer de Fermat*. Per a cada $n \in \mathbb{N}$, l'equació diofàntica

$$(X+1)^{n+3} + (Y+1)^{n+3} = (Z+1)^{n+3}$$

no té solucions enteres positives.[116] Així:

Afirmar que l'equació $(X+1)^{n+3} + (Y+1)^{n+3} = (Z+1)^{n+3}$ no té solucions enteres positives equival a afirmar que existeix una equació polinòmica $P(n, X, Y, Z, X_1, ..., X_k) = 0$ que tampoc no en té.

2. La *conjectura de Goldbach*. Tot nombre parell $2a+4$ és la suma de dos nombres primers senars. Formalment,

$$\forall a \exists p_1 \exists p_2 \big((2a = p_1 + p_2) \wedge \mathit{Primer}\,(p_1)\ \mathit{Primer}\,(p_2)\big).^{117}$$

La conjectura de Goldbach dóna una família paramètrica d'equacions diofàntiques, de paràmetre a. Ara bé, la negació, en canvi, estableix que

$$\forall z < a\ \exists x\ \exists y\ \big(z + 2 = (x+2)(y+2) \vee (2a + 4 - z) = (x+2)(y+2)\big),$$

115. MATIJASEVICZ, Y. (1971). La idea és la següent: enumeren els polinomis amb una gödelització recursiva g i considerem el conjunt $H = \{n \in \mathbb{N} : P(X_1, \ldots, X_m\} = 0$ té solució en $\mathbb{N}\}$, on $n = g\,(P(X_1, \ldots, X_m))$. Ara agafem un conjunt r.e., però no recursiu T. Existeix un polinomi $Q\,(X_0, X_1, \ldots, X_m)$ que defineix T. Així, $n \in T$ si, i només si, $Q_n(X_1, \ldots, X_m) = 0$ té solució en N, on $Q_n(X_1, \ldots, X_m) := Q\,(n, X_1, \ldots, X_m)$. D'on: $n \in T$ si, i només si, $g\,(Q_n(X_1, \ldots, X_m)) \in H$. Això és $1_T = 1_H$ o g. Per tant, H no és decidible perquè 1_T no és recursiva.

116. És una equació exponencial diofàntica i, per tant, diofàntica.

117. Observem que la propietat numèrica Primer (p) és diofàntica: Primer (p) si, i només si,

$\forall q_1 \le \sqrt{p}\ \forall q_2 \le \sqrt{p}\ \big((p \neq q_1 \times q_2) \wedge (q_1 > 1 \vee q_2 > 1)\big)$; o bé si, i només si, $\exists q\,(p \times q = (p-1)!+1)$.

Ambdues caracteritzacions són diofàntiques, atès que la quantificació universal acotada i la quantificació existencial són aplicables a relacions diofàntiques i donen relacions diofàntiques.

la qual és una equació diofàntica $D(x_0, \ldots, x_k) = 0$ amb solucions.

La conjectura de Goldbach estableix que l'equació diofàntica anterior no té solucions.

3. La *hipòtesi de Riemann*. D'acord amb les aportacions de M. Davis, Y. Matijasevicz i J. Robinson,[118] la negació de la hipòtesi de Riemann equival a l'existència de nombres k, ℓ, m i n que:

(1) $n \geq 600,\ m > 0$;
(2) $\forall y < \forall n\,((y+1) \mid m)$ [dóna un comú múltiple de 1, 2, …, n];
(3) $\forall y < m\,\big(y = 0 \vee \exists x < n\,((x+1) \nmid y)\big)$ [dóna el mcm de 1, 2, …, n];

i les tres que provenen de la substitució de ln per logexp:[119]

(4) $\operatorname{logexp}(m-1, \ell)$; (5) $\operatorname{logexp}(n-1, k)$; (6) $(\ell - n)^2 > 4n^2k^2$.

Totes són diofàntiques.

La hipòtesi de Riemann equival a la irresolubilitat d'una certa equació diofàntica.[120]

4. El *conjunt dels nombres primers*. En 1960 Hilary Putnam (1926-) establia el teorema següent, que té una certa semblança amb les equivalències que hem donat per als conjunts recursius i r.e.

Tot conjunt diofàntic $D \subseteq \mathbb{N}$ és el recorregut positiu d'un polinomi.
En concret, $n \in D$ si, i només si, $\exists x_1 \in \mathbb{Z} \ldots\ \exists x_k \in \mathbb{Z}\big(P(x_1, \ldots, x_k) = n\big)$.[121]

---

118. Davis, M., Matijasevicz, Y. i Robinson, J. (1976).

119. La funció $\operatorname{logexp}(a, b)$, definida per $\exists x\big(x > b+1 \wedge (1+\frac{1}{x})^{xb} \leq a+1 \wedge (1+\frac{1}{x})^{xb}\big)$, evita el ln.

120. La demostració és relativament senzilla. El més delicat és veure que la hipòtesi de Riemann es pot donar en termes de la funció ψ de Chebyshev: $\psi(n) = \ln\big(\operatorname{mcm}(1, 2, \ldots, n)\big)$.

121. Putnam, H. (1960). Una implicació és evident atès que $P(x_1, \ldots, x_k) = m$ és una relació recursiva. Per establir l'altra implicació, suposem que D és diofàntic. Aleshores existeix un polinomi $Q(X_0, X_1, \ldots, X_k)$ que $m \in D$ si, i només si, $m > 0 \wedge \exists x_1 \ldots \exists x_k\,\big(Q(m, x_1, \ldots, x_k) = 0\big)$. Considerem el polinomi $P(X_0, X_1, \ldots, X_k) = X_0\,\big(1 - Q^2(X_0, X_1, \ldots, X_k)\big)$. Aquest polinomi compleix la condició que volem establir.

← Si $m \in D$, hi ha $x_1, \ldots, x_k \in \mathbb{N}$ que $Q(m, x_1, \ldots, x_k) = 0$ i $P(m, x_1, \ldots, x_k) = 0$.

→ Si $m = P(n, x_1, \ldots, x_k) > 0$, per a la definició del polinomi P, $1 - Q^2(n, x_1, \ldots, x_k) > 0$ i $n > 0$. Per tant, $Q(n, x_1, \ldots, x_k) = 0$, ja que, altrament, $1 - Q^2(n, x_1, \ldots, x_k) \leq 0$.

Segons establí Julia Robinson, la funció $n!$ és exponencial diofàntica i aleshores el teorema de Wilson[122] diu que el conjunt dels nombres primers és diofàntic. És, doncs, la imatge positiva d'un polinomi.[123] Però, quin és aquest polinomi? La resposta data de 1976. És el *teorema de J. Jones, D. Sato, H. Wada* i *D. Wiens*. Donen un polinomi de grau 25 amb 26 variables. És el següent:[124]

$$
\begin{aligned}
(k+2)\ \{ & 1-(wz+h+j-q)^2 \\
& -((gk+2g+k+1)(h+j)+h-z)^2 \\
& -(2n+p+q+z-e)^2 \\
& -(16\,(k+1)^3\,(k+2)(n+1)^2+1-f^2)^2 \\
& -(e^3(e+2)(a+1)^2+1-o^2)^2 \\
& -((a^2-1)y^2+1-x^2)^2 \\
& -(16r^2y^4(a^2-1)+1-u^2)^2-(((a+u^2(u^2-a))^2-1)(n+4dy)^2+1-(x+cu)^2)^2 \\
& -(n+\ell+v-y)^2 \\
& -((a^2-1)\ell^2+1-m^2)^2 \\
& -(ai+k+1-\ell-i)^2 \\
& -(p+\ell(a-n-1)+b(2an+2a-n^2-2n-2)-m)^2 \\
& -(q+y(a-p-1)+s(2ap+2a-p^2-2p-2)-x)^2 \\
& -(z+p\ell(a-p)+\ell(2ap+2a-p^2-1)-pm)^2\}.
\end{aligned}
$$

6. COM A CONCLUSIÓ

Succintament, hem exposat els detalls que van portar a la hipòtesi de Riemann, les nocions que van ser necessàries per poder demostrar els teoremes d'incompletesa de Gödel que posen de manifest que l'«Ignoràbimus» de Hilbert no és correcte, de quina manera aquests conceptes nous plantejaven també problemes nous —com ara la complexitat algorísmica i el problema $\mathbf{P} \neq \mathbf{NP}$—; després ens hem endinsat en les idees bàsiques de la resolució negativa del problema diofàntic —els conjunts recursivament enumerables i els conjunts diofàntics coincideixen—, i finalment hem vist com lligaven aquests conjunts i alguns del problemes que havien anat apareixent en l'exposició i també amb el teorema darrer de Fermat.

122. Vegeu la pàgina 134.

123. Hi ha un polinomi $P(X_1, \ldots, X_n)$ que, per a $x_1, \ldots, x_n \in \mathbb{Z}$, si $P(x_1, \ldots, x_n) = p > 0$, aleshores p és primer i, a més, tot primer s'aconsegueix d'aquesta forma.

124. És una traducció pacient dels termes involucrats en la definició diofàntica del conjunt dels nombres primers. Pot sorprendre d'entrada que allò que caracteritza els nombres primers sigui un producte de dos factors. De fet el segon val 1, quan $k+2>0$.

Hi ha, tanmateix, una qüestió que em sembla que val la pena que posi de relleu, malgrat que ja l'he insinuat abans. Quan va aparèixer l'axioma de l'elecció i els matemàtics es van adonar de la seva importància i necessitat van haver de renunciar a un resultat que semblava desitjable:[125] «Tot subconjunt de $\mathbb{R}$ és mesurable de Lebesgue».[126] Després dels resultats de metamatemàtica segons els quals «l'axioma de l'elecció és independent de la resta d'axiomes de la teoria ZF de conjunts», hom podia haver imposat la seva negació i haver introduït una teoria de conjunts no-cantoriana.[127] Això no obstant, haurien hagut de renunciar al teorema de Tychonoff.[128]

Bé doncs, ens hem trobat amb dos resultats —la hipòtesi de Riemann i el problema $\mathbf{P} \neq \mathbf{NP}$— que, ara com ara, no sabem si són vertaders o falsos. De la seva acceptació —i d'una manera particular de la hipòtesi de Riemann— se'n dedueixen molts altres resultats desitjables. És a dir, al si de la teoria de conjunts ZFC+HR tenim una matemàtica molt més rica que al si de la teoria ZFC. Això no obstant, sabem si la HR és compatible amb la resta d'axiomes i si n'és independent? Podríem tenir teories de conjunts riemannianes i teories de conjunts no riemannianes, de la mateixa manera que tenim geometries riemannianes i geometries no riemannianes?

És possible demostrar la HR dins la teoria ZFC o dins la teoria ZFC+HGC?[129] I, en un àmbit molt més limitat i, per tant, molt més complex: problemes diofàntics com ara el teorema darrer de Fermat, la conjectura de Göldbach són demostrables al si de la teoria de Peano de primer ordre?

Tot això són divagacions que faig perquè veieu que hi ha una mena de problemes que lliguen amb alguns dels que va suggerir Hilbert i que, malgrat que no són a la llista dels més notables per al segle XXI, són problemes actualment oberts. Però potser el més adequat fóra que acabés l'exposició amb les mateixes paraules amb què Hilbert va acabar la seva conferència de 1900:

> Em limitaré a fer notar fins a quin punt és característic de la nostra ciència el fet que cada progrés efectiu porta aparellat el descobriment de mitjans auxiliars més rigorosos i més simples que, alhora que faciliten la comprensió de problemes anteriors i porten a la desaparició d'alguns dels desenvolupaments precedents per

125. I que alguns matemàtics creien que era vertader.

126. Usant l'AC, Giuseppe Vitali (1875-1932) donava un subconjunt no mesurable de Lebesgue. (Vegeu PLA, J. (1991), 367-368.)

127. A més es podia fer acceptant una negació particular —l'axioma de determinació— que fa vàlid l'enunciat de teoria de la mesura anterior. (Vegeu PLA, J. (2005).)

128. Perquè ambdós —teorema de Tychonoff i AC— són equivalents. (Vegeu PLA, J. (1991), 365-367.) Val la pena indicar que la situació és molt més greu pel que fa a la HGC que, segons el propi Cantor, té una validesa molt discutible, en el benentès que acceptar-lo ens obliga a acceptar d'altres resultats que no semblen els més desitjables.

129. Totes aquestes preguntes són també vàlides per a la hipòtesi $\mathbf{P} \neq \mathbf{NP}$.

la seva inutilitat, permeten orientar-nos en totes les branques de les Matemàtiques d'una manera molt més fàcil que en qualsevol altra ciència.

El caràcter unitari de la Matemàtica en constitueix l'essència. Efectivament, les Matemàtiques són el fonament de totes les altres ciències naturals exactes. A fi que, al segle que ve, compleixin totalment aquest objectiu tan elevat hauran de ser cultivades per mestres genials i per nombrosos joves inflamats per un zel noble.[130]

7. APÈNDIX. UNA APROXIMACIÓ A LA DEMOSTRACIÓ DEL TEOREMA DELS NOMBRES PRIMERS

Atès que hem centrat una gran part de l'exposició en el *teorema dels nombres primers*, em sembla adequat veure una demostració —potser fóra millor dir-ne una mostració— dels fets en què es basa.[131]

Volem veure que efectivament $\pi(n) \sim \frac{n}{\ln n}$. Per aconseguir-ho comencem fent la hipòtesi de l'*existència* d'una llei matemàtica que descrigui la distribució dels nombres primers en el sentit següent:

> Per a valors grans de n, la funció $\pi(n)$ s'apropa a la integral $\int_2^n W(x)\,dx$, on $W(x)$ és una funció que amida la «densitat» dels nombres primers.[132]

130. HILBERT, D. (1900b), a: GRAY, J.J. (2000), edició castellana de 2003, 310-311.

131. Ens hem inspirat en l'exposició que hi ha a COURANT, R. i ROBBINS, H. (1941), cinquena edició castellana de 1979, 493-496. Tanmateix, a la conferència de Pilar Bayer, hi trobareu la «demostració d'Euler» i el quadre dels lligams que menen a la demostració per contradicció.

132. Agafem l'índex inferior 2 perquè, per a $x<2$, $\pi(x) = 0$. Per precisar-ho millor, siguin x, Δx dos nombres grans, però entre els quals l'ordre de magnitud de x superi el de Δx (per exemple $\Delta x = \sqrt{x}$). Un cop fet això podem dir que la distribució que suposem als nombres primers és de manera que el nombre de primers de l'interval $[x, x+\Delta x]$ és aproximadament igual a $W(x) \times \Delta x$ i, a més, que $W(x)$ és una funció de x la variació de la qual és tan lenta que la integral $\int_2^n W(x)\,dx$ pot ésser substituïda per una aproximació rectangular sense que variï el valor asimptòtic. És delicat demostrar l'existència d'aquesta funció de densitat, però un cop feta la hipòtesi de la seva existència el *càlcul* de la funció és força senzill.

Sabem que, per a valors grans de n, $\ln n! \sim n \cdot \ln n$.[133] Ara donarem una segona fórmula per aproximar $\ln n$ relacionada amb els nombres primers. Mirem el nombre de vegades que un nombre primer p es troba com a factor en el producte $1 \times 2 \times \ldots \times n = n!$. Sigui $[a]_p$ el màxim enter k per al qual p^k divideix a. Aleshores, $[a \times b]_p = [a]_p \times [b]_p$.[134] Per tant,

$$[n!]_p = [1]_p + [2]_p + [3]_p + \cdots + [n]_p.$$

Els termes de la successió 1, 2, 3,…, n, divisibles per p^k són p^k, $2p^k$, $3p^k$… El seu nombre, quan n és gran, val aproximadament $N_k \sim \frac{n}{p^k}$. Aleshores, el nombre M_k de termes de la successió anterior divisibles per p^k, però no divisibles per p^{k+1}, és igual a $N_k - N_{k+1}$. Per tant,

$$\begin{aligned} [n!]\mathrm{p} &= M_1 + 2M_2 + 3M_3 + \cdots \\ &= (N_1 - N_2) + 2(N_2 - N_3) + 3(N_3 - N_4) + \cdots \\ &= N_1 + N_2 + N_3 + \cdots \\ &= \tfrac{n}{p^n} + \tfrac{n}{p^2} + \tfrac{n}{p^3} + \cdots = \tfrac{n}{p-1}.^{135} \end{aligned}$$

De tot això en resulta que, quan n és gran, el valor de $n!$ s'obté, de forma aproximada, com el producte de totes les expressions $p^{\frac{n}{p-1}}$, per a tots els primers $p < n$. Tenim, doncs, l'expressió següent:

$$\ln n! \sim \sum_{p<n} \frac{n}{p-1} \ln p.$$

Si comparem aquesta expressió de $\ln n!$ amb l'expressió asimptòtica d'abans i substituïm n per x, obtenim:

$$\ln x \sim \sum_{p<x} \frac{\ln p}{p-1}. \tag{7.1}$$

133. Fem $P_n = \ln 2 + \ln 3 + \cdots + \ln n$. Considerem les sumes dels rectangles inscrits a la corba $\ln x$, $x \geq 1$, de base unitat, i dels circumscrits. Comparem-les amb la integral $\int_1^{n+1} \ln x \, dx$. Per a la primera suma, tenim $\int_1^{n+1} \ln x \, dx \sim (n+1)\ln(n+1) - (n+1) + 1$ i, per a la segona, $\int_1^n \ln x \, dx \sim n \ln n - n + 1$.

Així doncs,

$$n \ln n - n + 1 < P_n < (n+1)\ln(n+1) - n.$$

Ara ho dividim tot per $n \ln n$, fem que $n \to \infty$ i hem acabat.

134. És una conseqüència immediata del fet que la descomposició en nombres primers és única.

135. Totes les igualtats són aproximades.

Ara ve el pas decisiu, que consisteix a obtenir una expressió asimptòtica en funció de $W(x)$ per al segon membre de l'equació (7.1). Com que x és gran, podem subdividir l'interval $[2, n]$, amb $x = n$, en un nombre, també gran r, de subintervals, elegint els punts $\xi_1 = 2, \xi_2, \ldots, \xi_r = x$, amb els increments $\Delta\xi_j = \xi_{j+1} - \xi_j$. A cada un dels subintervals poden existir nombres primers, però tots els que hi ha al subinterval j-èsim valen ξ_j. Atesa la hipòtesi relativa a la funció $W(x)$, existeixen aproximadament $W(\xi_j) \cdot \xi_j$ nombres primers en el subinterval j-èsim. Per tant, la suma del segon membre de l'equació (7.1) és aproximadament igual a

$$\sum_{j=1}^{r+1} W(\xi_j) \frac{\ln \xi_j}{\xi_j - 1} \cdot \Delta\xi_j.$$

Si canviem aquesta suma finita per la integral que l'aproxima, obtenim una conseqüència plausible de l'equació (7.1):

$$\ln x \sim \int_2^x W(x) \frac{\ln \xi}{\xi - 1} d\xi. \tag{7.2}$$

Ara ja podem determinar la funció desconeguda $W(x)$. D'antuvi substituïm el signe ~ pel signe =; després derivem membre a membre, respecte de x, usant el teorema fonamental del càlcul:

$$\frac{1}{x} = W(x) \frac{\ln x}{x - 1}, \text{ i, en conseqüència, } W(x) = \frac{x - 1}{x \ln x}. \tag{7.3}$$

En començar hem suposat que $\pi(x)$ és aproximadament igual a $\int_2^x W(x)\,dx$. Per tant, $\pi(x)$ s'obté per la integral

$$\int_2^x \frac{x - 1}{x \ln x} dx.^{136} \tag{7.4}$$

En conseqüència, la integral (7.4) serà asimptòticament igual a la integral

$$\int_2^x f'(x)\,dx = f(x) - f(2) = \frac{x}{\ln x} - \frac{2}{\ln 2}.$$

136. Observem que la derivada de la funció $f(x) = \frac{x}{\ln x}$ és $f'(x) = \frac{1}{\ln x} - \frac{1}{(\ln x)^2}$. Però, per a valors grans de x, les expressions $\frac{1}{\ln x} - \frac{1}{(\ln x)^2}$ i $\frac{1}{\ln x} = \frac{x-1}{x \ln x}$ són aproximadament iguals.

Podem despreciar, quan x és gran, el valor de $\frac{2}{\ln 2}$ i obtenim el resultat esperat:

$$\pi(x) \sim \frac{x}{\ln x}\ .^{137}$$

BIBLIOGRAFIA

BACHET DE MÉZIRIAC, Claude-Gaspart (1612). *Problèmes plaisantes et déléctables, qui se font par les nombres.* París.

BAYER I ISANT, Pilar (1984). «Sèries infinites», a *Leonhard Euler (1707-1783)*. *Butlletí de la Societat Catalana de Ciències Físiques, Químiques i Matemàtiques*. Segona època, II, 4, p. 75-127.

BURTON, David M. (1983). *The History of Mathematics. An Introduction. Allyn and Bacon*, Inc. Boston.

CHURCH, Alonzo (1936). «An unsolvable problem of elementary number theory». *The American Journal of Mathematics*, 58, p. 345-363. Reimprès a DAVIS, M. (1965), p. 89-101.

—(1936a). «A note on the Entscheindugsproblem». *Journal of Symbolic Logic*, 1, p. 40-41; correcció en *ibid*, p. 101-102. Reimprès a DAVIS, M. (1965), p. 110-114.

COOK, Stephen (1971). «The Complexity of Theorem Proving Procedures». *ACM SIGACT Symposium on the Theory of Computing.*

COURANT, Richard i ROBBINS, Herbert (1941). *What is mathematics?* Oxford University Press. Nova York. Traducció castellana de Luis Bravo Gala, *¿Qué es la matemática?*, cinquena edició. Aguilar. Barcelona, 1979.

DAVIS, Donald M. (1993). *The Nature and Power of Mathematics.* Princeton University Press. Princeton.

DAVIS, Martin (1973). «Hilbert's Problem Is Unsolvable». *The American Mathematical Montly*, 80, p. 233-269.

—(2000) *The Universal Computer*. W.W. Norton & company, cop. Nova York, 2000. Traducció castellana de Ricardo García Pérez, *La computadora universal*. Debate. Madrid, 2002.

137. Val la pena que insistim en els fets següents: és possible fer, amb tota mena de correcció, els passos que porten a l'equació (7.1), a la igualtat asimptòtica entre aquesta suma i la integral (7.2), i els que permeten passar de (7.2) a (7.3). El punt difícil és, com dèiem, el que permet establir l'existència de la funció de densitat $W(x)$. Tanmateix, un cop feta la hipòtesi, com hem vist, el seu càlcul és senzill. En definitiva, el punt clau de la demostració del teorema dels nombres primers és l'existència de la funció de distribució $W(x)$.

DAVIS, Martin; MATIJASEVICZ, Yuri; ROBINSON, Julia. (1976). «Hilbert's Tenth Problem. Diophantine equations: positive aspects of a negative solution», a: *Mathematical Developments Arising from Hilbert Problems*, vol. 28 de *Proceedings of Symposia in Pure Mathematics*, p. 323-378. Providence. Rhode Island. American Mathematical Society.

DEVLIN, Keith (1988). *Mathematics: The New Golden Age*. Penguin Books. Londres.

DIRICHLET, P.G. Leujene (1837) «Beweis des Satzes, dass jede unbegrenzte arithmetische Progression, deren erstes Glied und Differenz ganze Zahlen ohne gemeinschaftlichen Factor sind, unendliche viele Primzahlen enthält». *Abh. Akad. Wiss. Berlin*, p. 45-81, a *Werke*, I, p. 307-312.

DÖRRIE, Heinrich (1958) *Triumph der Mathematik: Hundert berühmte Probleme aus zwei Jahrtausenden mathematischer Kultur.* Physica-Verlag. Würzburg. Traducció anglesa de David Antin, *100 Great Problems of Elementary Mathematics. Their History and solution*. Dover Publications, Inc. Nova York, 1965.

DOU, Alberto (1970). *Fundamentos de la Matemàtica*. Labor. Barcelona.

EDWARDS, H.M. (1974). *Riemann's zeta function*. Academic Press. Nova York.

ECHEVARRÍA, Javier (1992). «Observations, Conjectures and Problems in Number Theory: the History of Prime Number Theorem», a: *The Space in Mathematics*, p. 248-278. De Gruyter. Berlín.

EULER, Leonhard (1732). «Observationes de theoremate quodam Fermatiano aliisque ad numeros primos spectantibus», a: *Opera Omnia*, 2, p. 1-5.

—(1737) «Variæ observationes circa series infinitas», a: *Opera*, 14, p. 216-244.

—(1748) *Introductio in Analysin infinitorum*, 2 volums. Lausanne, a *Opera Omnia*, 8. Traducció castellana: *Introducción al análisis del infinito*. Sevilla, 2000.

—(1751) «Découverte d'une loi tout extraordinaire des nombres par rapport à la somme de leurs diviseurs». Bibliotècque imperiale 3, p. 10-31, a: *Opera Omnia*, 2, p. 241-262.

—(1849) «De numeris amicabilis», a: *Comm. Arith.*, 2, p. 627-636.

EULER, Leonhard i GOLDBACH, Christian (1965). *Briefwechsel*. A.P. Juskevic und E. Winter. Berlín.

FAREY, John (1816). «On a Curious Property of Vulgar Fractions». *London, Edimburg and Dublin Philosophical Magazine*, 47, p. 385-387.

FERMAT, Pierre (1640) «Carta a Frenicle», a *Œuvres*, 2, p. 206.

—(1643) «Fragment d'una carta de Fermat» (probablement adreçada a Mersenne), a *Œuvres*, 2, p. 256-258.

—(1659) «Carta a Carcavi d'agost de 1659», a *Œuvres*, 2, p. 431-436.

FRANEL, J. i LANDAU, Edmund (1924). «Les suites de Farey et le problème des nombres premiers». Gottinger Nachr, p. 198-206.

GAUSS, Carl Friedrich (1799). «Beiträge zu Theorie der algebraischen Gleichungen», a: *Werke*, III, p. 72-102.

—(1801) *Disquisitiones Arithmeticæ*. Leipzig, a: *Werke*, I. Traducció catalana de Griselda Pascual, *Disquisicions aritmètiques*. IEC. Barcelona 1996.

—(1849) «Carta a Enke, del 24 de desembre de 1849». a: *Werke*, II, p. 444-447. Königlichen Gesellschaft der Wissenschaften zu Göttingen.

GÖDEL, Kurt (1930) «Die Vollständigkeit der Axiome des logischen Funktionalkalküls». *Monatshefte für Mathematik und Physik*, 37, p. 349-360. Traducció castellana a GÖDEL, K. (1981), edició de 1989, p. 23-37.

—(1931) «Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme». *Monatshefte für Mathematik und Physik*, 38, p. 173-198. Traducció castellana a GÖDEL, K. (1981), edició de 1989, p. 53-87.

—(1934) «On Undecidable Propositions of Formal Mathematical Systems». Notes de S. C. Kleene i J. B. Rosser. The Institute for Advanced Study. Princeton. Traducció castellana a GÖDEL, K. (1981), edició de 1989, p. 53-87.

—(1981) *Kurt Gödel. Obras completas*. Madrid. Alianza. Reeditat el 1989.

GRAY, Jeremy J. (2000). *The Hilbert Challenge*. Oxford University Press. Oxford. Traducció castellana de Javier García Sanz, *El reto de Hilbert*. Drakontos. Crítica. Barcelona 2003.

HADAMARD, Jacques (1896). «Sur la distribution des zeros de la fonction $\zeta(s)$ et ses consequences arithmétiques». *Bull. Soc. Math. France*, 24, p. 199-220.

HARDY, Godfrey Harold (1914). «Sur les zeros de la fonction $\zeta(s)$ de Riemann». *Comptes Rendues de la Société des Sciences de Paris*, 158, p. 1.012-1.014, a *Collected Papers*, 2, p. 6-9.

HILBERT, David (1899). *Grundlagen der Geometrie*, a: Hilbert, D. (1966), p. 1-135.

—(1900a) «Über den Zahlbegriff», a: HILBERT, D. (1966), p. 244-249.

—(1900b) «Mathematische Probleme». *Archiv für Mathematik und Physik*, 1, p. 44-63, p. 213-237. Traducció castellana de Javier García a: GRAY, J.J. (2000), p. 263-314.

—(1925) «Über das Unendliche», a: HILBERT, D. (1966), p. 264-287.

—(1928a) «Die Grundlagen der Mathematik», a: HILBERT, D. (1966), p. 288-309.

—(1928b) «Probleme der Grundlegung der Mathematik», a: HILBERT, D. (1966), p. 310-319.

HILBERT, DAVID; BERNAYS, PAUL (1934). *Grundlagen der Mathematik*, vol. I. Berlín.

—(1939). *Grundlagen der Mathematik*, vol. II. Berlín.

JONES, J.; SATO, D.; WADA, H.; WIENS, D. (1976). «Diophantine representation of the Set of Prime Numbers». *American Mathematical Montly*, 83, p. 449-464.

LAGRANGE, Joseph-Louis (1773). «Démonstration d'un théorème nouveau concernant les nombres premiers», a: *Œuvres de Lagrange*, III, p. 425-438. Gauthier-Villers. París 1869.

LAUBENBACHER, Reinhard i PENGELLEY, David (1999). *Mathematical Expeditions*. Springer-Verlag. Nova York.

LEHMER, Derrick H. i POWER, R.E. (1931). «On factoring large numbers». *Bulletin of the American Mathematical Society*, 37, p. 770-776.

LITTLEWOOD, John Edensor (1914). «Sur la distribution des nombres premiers». *Comptes Rendues de la Société des Sciences de Paris*, 158, p. 1.869-1.872.

MANCOSU, Paolo (1998). *From Brouwer to Hilbert*. Oxford University Press. Oxford.
MATIJASEWICZ, Yuri (1993). *Hilbert's Tenth Problem*. MIT Press. Massachusetts.
MERSENNE, Marin (1944). *Cogitata Physico-Mathematica*. París.
MILLER, G.L. (1976). «Riemann hypothesis and tests for primality». *Journal of computer and systems Science*, 13, p. 300-317.
MOIVRE, Abraham de (1730). *Miscellanea Analytica*. Londres.
PEANO, Giuseppe (1889). *Arithmetices principia, nova methodo exposita*. Bocca. Torino. Traducció castellana a: *Los principios de la aritmética expuestos según un nuevo método*. Pentalfa Ediciones. Oviedo 1979.
PLA I BRUNET, Joaquim (ed.) (2002). *10 impactes de la ciència del segle XX*. Eumo. Vic. Traducció castellana dels autors, *10 impactos de la ciencia del siglo veinte*. Fondo de Cultura Económica. Mèxic, 2002.
PLA I CARRERA, Josep (1991). *Lliçons de lògica matemàtica*. PPU. Barcelona.
—(2000) «Matemàtiques, naturalesa i espionatge». *Novembre matemàtic*. Sant Feliu de Guíxols.
—(2000) «Alguns resultats de la matemàtica del segle XX», a: PLA I BRUNET, J. (2000), p. 261-302.
—(2002) *Notes per a unes classes de teoria de models i decidibilitat*. Universitat de les Illes Balears. Palma.
—(2005) «El platonisme en matemàtiques i la teoria de conjunts». Reial Acadèmia de Doctors. Barcelona. Pendent de publicació.
POINCARÉ, Henri (1908). «L'Avenir des mathématiques». *Rendiconti del Circolo Matematico di Palermo*, 26, p. 152-168. Traducció castellana al capítol II de *Ciencia y método*. Espasa-Calpe. Madrid 1962.
POST, Emil (1944). «Recursively enumerable sets of positive integers and their decision problema». *Bulletin of the American Mathematical Society* (2), 50, p. 284-316. Reimprès a DAVIS, M. (1965), p. 304-337.
PUTNAM, Hilary (1960 «An unsolvable problem in arithmetic». *Journal of Symbolic Logic*, 25, p. 220-232.
RADEMACHER, Hans i TOEPLIZ, Erna (1923). *Von Zahlen und Figuren*. Traducció castellana d'Eva Rodríguez Halffter i Manuel Gregori Sousa, *Números y figuras*. Alianza Editorial. Madrid 1970.
REY PASTOR, Julio i BABINI, José (1984). *Historia de la matemática*, dos volums. Gedisa. Barcelona.
RIEMANN, Bernhard (1859). «Über die Anzahl der Primzahlen unter einer gegebenen Grösse». *Monatsberichte Berliner Akad.*, 671-68, a: *Werke*, p. 144-155. Traducció francesa de L. Langel, *Œuvres mathématiques de Riemann*, p. 165-176. Gauthier-Villars. París 1898.
ROBINSON, Julia (1952). «Existencial definability in arithmetic». *Transactions of American Mathematical Society*, 72, p. 437-449.
ROSEN, Kenneth H. (2000). *Elementary Number Theory*. Addison-Wesley. Massachusetts.

ROSSELLÓ, Joan (2003). *Lògica i fonaments 1850-1920. Un estudi comparatiu de les contribucions del corrent algèbric i logicista a la lògica contemporània*. Tesi doctoral. Universitat de Barcelona.

SERRE, Jean-Pierre (1981). «Quelques applications du théorème de densité de Chebotarev». *Pub. Math. IHES*, 54, p. 123-201.

SMALE, Steve (1998). «Mathematical Problems for the Next Century». *The Mathematical Intelligencer*, 20(2), p. 7-15.

STILLWELL, John (1999). *Mathematics and its History*. Springer-Verlag. Nova York.

STIRLING, James (1730). *Methodus Differentialis*. Londres.

STRUIK, Dirk J. (1969). *A Source Book in Mathematics: 1200-1800*. Harvard University Press. Cambridge. Massachusetts.

TARSKI, Alfred (1931). «Der Wahrheitsbegriff in der formalisierten Sprachen». *Studia philosophica*, 1. p. 261-405, a: TARSKI, A. (1972), I, p. 157-269.

—(1972) *Logiques, sémantique, métamathématique*, 2 volums. Gauthier-Villars. París.

TOEPLIZ, Otto (1947). *Die Entwicklung der infinitesimalrechnung*. Editat per Gottfried Köthe. Traducció anglesa de Luise Lange, *The Calculus. A Genetic Approach*. The University of Chicago Press. Chicago i Londres.

TURING, Allan Mathison (1936). «On computable numbers, with an application to the Entscheidungsproblem». *Proceedings of the London Mathematical Society. Second Series*, 42, p. 230-265.

—(1937) «On computable numbers, with an application to the Entscheidungsproblem. A correction». *Proceedings of the London Mathematical Society. Second Series*, 43, p. 544-546.

—(1939) «Systems logics based on ordinals». *Proceedings of the London Mathematical Society. Second Series*, 45, p. 161-228.

VALLÉE POUSSIN, Charles de la (1896). «Recherches analytiques sur la théorie des nombres premiers (première partie)». *Ann. Soc. Sci.* Brussel·les, 20, p. 183-256.

VERA, Francisco (1970). *Los científicos griegos*, dos volums. Aguilar. Madrid.

WARING, Edward (1770). *Meditationes Algebraicæ*. Londres.

La conjectura de Birch i Swinnerton-Dyer

Jordi Quer

1. EQUACIONS DIOFÀNTIQUES

Les equacions diofàntiques són equacions polinòmiques a coeficients enters de les quals es volen saber les solucions donades per nombres enters. La taula següent, per exemple, conté una llista d'equacions diofàntiques que per una raó o altra han jugat un paper important en la història de les matemàtiques.

Algunes d'aquestes equacions es discutiran amb més o menys detall més endavant.

Equació	**Nom/Origen**	**Solució**
$aX+bY=c$	Equació lineal	Euclides (s.IV aC)
$X^2+Y^2=Z^2$	Ternes pitagòriques	Babilònia (s.XIX aC)
$aX^2+bY^2=cZ^2$	Cònica racional	Legendre (1780)
$X^2-aY^2=1$	Equació de Pell	Lagrange (1771)
$X^3-n^2XZ^2=Y^2Z$	Nombres congrus (s.X)	Tunnell (1983)
$X^n+Y^n=Z^n$	Equació de Fermat (1620)	Wiles (1995)
$X^4+Y^4+Z^4=W^4$	Quàrtica d'Euler (1720)	Elkies (1988)
$X^n+1=Y^m$	Equació de Catalan (1856)	Mihailescu (2003)

Quan es planteja la resolució d'una equació diofàntica es poden demanar diferents nivells de precisió pel que fa a què es vol saber de les solucions. En ordre creixent de dificultat, la resolució completa d'una equació diofàntica consisteix a:

1. Decidir si té o no solució (deixant de banda possibles solucions òbvies, si n'hi ha).
2. Dir quantes solucions té. En cas que en tingui infinites, dotar-les si és possible d'alguna estructura que en faciliti l'estudi.
3. Trobar totes les solucions o donar un procediment que permeti calcular-les.

En aquesta secció es discutiran algunes de les equacions de la taula anterior, fent èmfasi en aspectes i en tècniques de resolució que juguen un paper important en la conjectura de Birch i Swinnerton-Dyer.

Equacions lineals. La solució de les equacions lineals en dues variables

$$aX + bY = c, \qquad a,b,c \in \mathbb{Z} \qquad ab \neq 0$$

és molt senzilla, i s'explica a molts cursos d'àlgebra elemental de primer curs universitari. Per a aquesta equació els tres problemes que tot just hem esmentat admeten una resposta ben simple:

1. L'equació té alguna solució si, i només si, el màxim comú divisor $d = (a,b)$ dels coeficients de les variables divideix el terme independent c.

2. En cas que en tingui, sempre en té infinites, les quals es poden posar en bijecció de manera natural amb el conjunt dels nombres enters.

3. Es pot trobar una solució de l'equació amb una versió estesa de l'algorisme d'Euclides.

A partir d'una solució particular trobada d'aquesta manera, diguem (x_0, y_0), s'obtenen totes les solucions amb la fórmula següent:

$$x = x_0 + \frac{b}{d}t, \qquad y = y_0 - \frac{a}{d}t,$$

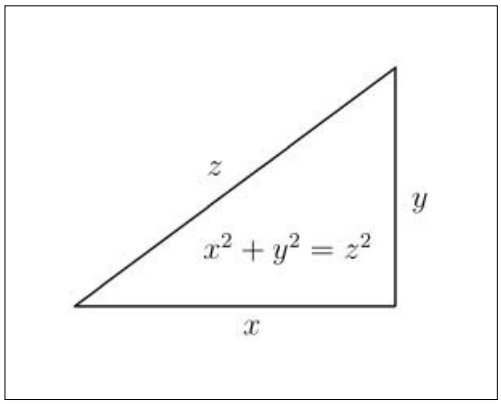
on el paràmetre t recorre el conjunt dels nombres enters.

Amb tècniques del mateix tipus es poden estudiar i resoldre les equacions diofàntiques lineals de més de dues variables, i fins i tot els sistemes de qualsevol nombre d'equacions lineals a coeficients enters.

Ternes pitagòriques. Es considera l'equació diofàntica

$$X^2 + Y^2 = Z^2.$$

El teorema de Pitàgores diu que cada solució en nombres positius d'aquesta equació es correspon amb els tres costats d'un triangle rectangle. Tenint en compte que les solucions segueixen sent-ho en canviar els signes, resoldre l'equació diofàntica equival a trobar tots els triangles rectangles tals que la mida de cadascun dels tres costats és un nombre enter. És per això que les solucions de l'equació s'anomenen



ternes pitagòriques. Per exemple, una solució de l'equació és la terna (3,4,5). De fet, de ternes pitagòriques n'hi ha infinites.

Utilitzant només la factorització única dels nombres enters en producte de primers, i amb arguments de divisibilitat, és un exercici senzill arribar a l'expressió

$$x=u^2-v^2, \qquad y=2uv, \qquad z=u^2+v^2 \tag{1}$$

que, quan els paràmetres u i v recorren el conjunt dels nombres enters, proporciona totes les solucions de l'equació que ens ocupa, juntament amb l'expressió anàloga obtinguda intercanviant les fórmules que donen x i y.

Una tauleta babilònica datada cap al 1800-1900 aC, que es coneix amb el nom de *Plimpton 322*, conté una llista de ternes pitagòriques, escrita en signes de l'escriptura cuneïforme, amb enters d'una mida tal que difícilment poden ser fruit d'una recerca numèrica, la qual cosa fa pensar que els babilonis ja coneixien un mètode per generar solucions del tipus de la parametrització anterior, és a dir, que sabien «resoldre» l'equació.

Si (x,y,z) és una terna pitagòrica, aleshores també ho és (dx,dy,dz) per a tot nombre enter d. Les solucions (x,y,z) de

$$X^2+Y^2=Z^2 \tag{2}$$

es diuen *primitives* quan els tres nombres x,y,z no tenen cap factor comú. Com que a partir de les solucions primitives es pot obtenir tota la resta, n'hi ha prou a trobar les que ho són. Donada una solució primitiva (x,y,z), els nombres racionals x/z i y/z són solució de l'equació

$$X^2+Y^2=1 \tag{3}$$

i, recíprocament, si (r,s) és una solució d'aquesta equació en nombres racionals (fraccions)

$$r,s\in\mathbb{Q}=\left\{\frac{a}{b}\ :a,b\in\mathbb{Z},\ b\neq 0\right\},$$

escrivint-los com una fracció reduïda (sense factors comuns al numerador i denominador), aquests nombres r i s seran de la forma

$$r=\frac{x}{z}, \qquad s=\frac{y}{z},$$

on el denominador ha de ser el mateix a tots dos, el qual es comprova amb un argument simple de divisibilitat. Per tant, a partir de la solució (r,s) s'obté una terna pitagòrica primitiva (x,y,z).

Aquesta correspondència entre solucions enteres reduïdes de l'equació (2) i solucions racionals de l'equació (3) és un fet que es dóna sempre que es tracta amb equa-

cions diofàntiques homogènies: aquelles en què els graus dels monomis són tots iguals. Per exemple, a la taula del començament de la secció totes les equacions són homogènies excepte la primera (lineal), l'equació de Pell i l'equació de Catalan. Per tant, estudiar equacions diofàntiques homogènies equival a trobar solucions racionals de les equacions que s'obtenen en deshomogeneitzar respecte a una de les variables.

La determinació de totes les solucions racionals de l'equació (3), és a dir, dels punts de la circumferència unitat amb coordenades nombres racionals, es pot dur a terme amb el procediment següent: els punts (x, y) del pla que satisfan l'equació són els punts de la circumferència de radi 1 centrada a l'origen. Es consideren totes les rectes que passen pel punt $(-1, 0)$, determinades per la seva pendent t. Cadascuna d'aquestes rectes talla aquesta circumferència en dos punts: el mateix punt $(-1, 0)$ i un altre punt, que es pot calcular com la intersecció entre la circumferència i la recta, que té coordenades

$$(x, y) = \left(\frac{1-t^2}{1+t^2} \, , \, \frac{2t}{1+t^2} \right).$$

Clarament, aquestes coordenades són nombres racionals si, i només si, la pendent t de la recta considerada és un nombre racional, de manera que les solucions racionals de (3) queden parametritzades en termes d'un paràmetre racional t. En escriure el paràmetre t com a quocient de nombres enters $t = u/v$, es recupera parametrització (1) de les ternes pitagòriques.

Equació de Legendre. Es considera l'equació diofàntica

$$aX^2 + bY^2 = cZ^2$$

on els coeficients són nombres enters diferents de zero i, canviant el signe a tots dos costats si cal, es pot suposar que c és positiu. Anàlogament al cas anterior, totes les seves solucions s'obtenen a partir de solucions primitives, i la solució trivial $(0, 0, 0)$ ni tan sols es considera, de manera que les solucions primitives amb $z \neq 0$ es corresponen amb els punts racionals de la cònica

$$aX^2 + bY^2 = c. \tag{4}$$

Naturalment, per tal que una equació com aquesta tingui solucions racionals n'ha de tenir de reals, i això només és possible si algun dels dos nombres a i b és positiu.

Per estudiar aquesta mena d'equacions són molt útils les congruències: si n és un enter positiu, dos enters a i b es diuen congruents mòdul n quan la seva diferència és divisible per n. En classificar els nombres enters per la seva classe de congruència

mòdul n, s'obté un conjunt finit, que admet com a representants els nombres enters a de l'interval $0 \leq a < n$. Aquest conjunt de classes de congruència mòdul n acostuma a denotar-se $\mathbb{Z} / n\mathbb{Z}$.

És clar que tota solució d'una equació diofàntica

$$f(X_1, X_2, \ldots, X_n) = 0$$

també és solució de la congruència corresponent per a tot mòdul

$$f(X_1, X_2, \ldots, X_n) \equiv 0 \pmod{n},$$

i estudiar una equació com aquesta sempre es pot reduir a provar tots els possibles valors de les variables, ja que de classes de congruència n'hi ha només un nombre finit. Aquesta observació trivial proporciona un mètode per demostrar que una equació diofàntica no té solucions. Per exemple, l'equació

$$X^2 + Y^2 = 3Z^2$$

no té solucions no trivials. En efecte, tota solució no trivial donaria lloc, dividint pel màxim comú divisor dels seus coeficients, a una solució primitiva. Es considera la congruència

$$X^2 + Y^2 \equiv 3Z^2 \pmod{4}.$$

Tenint en compte que el quadrat d'un enter parell és $\equiv 0 \pmod 4$ i el quadrat d'un enter senar és $\equiv 1 \pmod 4$, totes les solucions d'aquesta congruència es donen només amb valors parells de totes tres variables, i, per tant, no poden ser reducció de cap solució primitiva de l'equació diofàntica, d'on es dedueix que aquesta no pot tenir solucions no trivials. Amb un argument semblant també es pot demostrar la inexistència de solucions primitives estudiant la congruència mòdul 3. Tota solució primitiva (x, y, z) de l'equació diofàntica ha de tenir les coordenades x, y no divisibles per 3 (si una d'elles és divisible per 3, l'altra també ho és, aleshores $3z^2$ és divisible per 9 i, per tant, z és divisible per 3, en contradicció amb la hipòtesi de primitivitat). Tenint en compte que el quadrat d'un nombre no divisible per 3 és sempre $\equiv 1 \pmod 3$ s'arriba a la contradicció

$$x^2 + y^2 \equiv 1 + 1 \equiv 2 \not\equiv 0 \equiv 3z^2 \pmod{3}.$$

Cap al 1780, Legendre va demostrar que per a les equacions que ens ocupen es compleix el recíproc

Teorema 1.1 (Legendre): *L'equació diofàntica* $aX^2+bY^2=cZ^2$ *té solució no trivial si, i només si, la congruència corresponent té solució no trivial per a tot mòdul (i els signes dels coeficients* a, b, c *permeten que tingui solució real).*

Podria semblar que el teorema no és gaire útil, ja que per assegurar l'existència de solució s'han de comprovar infinites congruències, una per a cada valor del mòdul $n=2, 3, 4...$ En canvi, un estudi una mica més detallat permet observar que n'hi ha prou a estudiar congruències per al conjunt finit de mòduls que divideixin el nombre $8abc$, i que això ja garanteix l'existència de solució per a tots els mòduls i, per tant, solució entera.

Les equacions diofàntiques amb aquesta propietat:

> L'equació té solució entera si, i només si, té solució en nombres reals i la congruència corresponent té solució per a tot mòdul n,

es diu que satisfan el *principi de Hasse* (s'ha d'afegir «no trivial» en tots els casos quan es tracta d'equacions homogènies).

El principi de Hasse també es pot aplicar a solucions racionals d'equacions polinòmiques, i aleshores es refereix a les solucions enteres no trivials de les equacions diofàntiques homogènies corresponents. El teorema de Legendre assegura, doncs, que les equacions (4) satisfan el principi de Hasse; és a dir, que tota cònica racional satisfà el principi de Hasse. De fet, se sap que tota equació diofàntica de grau 2 en qualsevol nombre de variables el satisfà.

Els primers exemples d'equacions que no satisfan aquest principi els va donar Selmer cap al 1930. Per exemple, la cúbica

$$3X^3+4Y^3=5Z^3$$

no té solucions enteres no trivials i, en canvi, la congruència corresponent té solucions no trivials per a tot mòdul n.

Les còniques (4) són versions «torçades» de la circumferència unitat (3), en el sentit que són corbes isomorfes a aquesta sobre la clausura algebraica dels nombres racionals (o, si es vol, sobre els nombres complexos). Les transformacions $X \mapsto \sqrt{a/c}X$, $Y \mapsto \sqrt{b/c}\,Y$ són un isomorfisme amb coeficients complexos que, en general, no són racionals, i de fet les corbes (4) no són en general isomorfes a la circumferència sobre els nombres racionals. Així, el teorema de Legendre es pot enunciar dient que totes les torçades de la circumferència satisfan el principi de Hasse. A la conjectura de Birch i Swinnerton-Dyer hi juga un paper molt important el problema anàleg, però on en comptes de la circumferència unitat es consideren les torçades d'una corba el·líptica: algunes torçades de la corba el·líptica satisfan el principi de Hasse i d'altres no el satisfan.

Aquestes darreres, les corbes torçades d'una corba el·líptiques que no satisfan el principi de Hasse, són les responsables que actualment no se sàpiguen resoldre en general les equacions diofàntiques cúbiques (donades per corbes el·líptiques).

Últim teorema de Fermat. Es tracta, potser, del problema matemàtic que ha tingut més ressò fora de la comunitat matemàtica en tota la història. Cap al 1620 Pierre de Fermat va escriure al marge del seu exemplar de l'*Aritmètica* de Diofant que l'equació

$$X^n + Y^n = Z^n$$

no té cap solució entera no trivial per a exponents $n \geq 3$. També va escriure que tenia una demostració d'aquest fet, però el marge era massa estret per poder explicar-la. Tot i que Fermat va donar una demostració correcta per a exponent $n = 4$, tots els experts opinen que s'enganyava sobre la potència dels seus mètodes per resoldre el problema en general.

Aquest problema ha exercit una influència notable en el desenvolupament de la teoria de nombres en els darrers segles. Molts dels matemàtics més brillants d'aquesta disciplina han intentat resoldre'l sense èxit, tot i que en molts casos l'estudi del problema els ha portat a elaborar noves teories i desenvolupar eines i tècniques que han enriquit el coneixement matemàtic.

A la dècada de 1980 es va obrir una nova via per atacar el problema quan Frey, Serre, Ribet i altres van aconseguir lligar la seva solució a la d'una conjectura sobre corbes el·líptiques i corbes modulars feta per Shimura i Taniyama cap al 1995.

L'any 1993, en una sèrie de tres conferències a la Universitat de Cambridge, Andrew Wiles va esbossar una demostració llarga i complexa d'una part de la conjectura de Shimura-Taniyama suficient per deduir-ne l'últim teorema de Fermat.

Tot i que la demostració proposada inicialment contenia un error irreparable en un càlcul, durant l'any següent, i amb la contribució del seu estudiant Richard Taylor, Wiles va aconseguir reemplaçar i simplificar alguns arguments de manera que la demostració no depengués del càlcul erroni. Finalment, l'any 1995 la demostració es va publicar després d'haver superat l'escrutini dels experts.

La conjectura de Shimura-Taniyama juga també un paper fonamental en la conjectura de Birch i Swinnerton-Dyer, tal com s'explicarà més endavant, ja que garanteix la prolongació analítica de les L-sèries i les corbes el·líptiques a l'esquerra de l'abscissa de convergència, sense la qual l'enunciat de la conjectura de Birch i Swinnerton-Dyer ni tan sols tindria sentit.

Conjectura de la quàrtica d'Euler. Tot i que en general les conjectures s'acaben demostrant en la direcció en què es van formular, també hi ha casos en què els seus enunciats resulten ser falsos. Per exemple, Euler va conjecturar el 1780 que l'equació diofàntica

$$X^4 + Y^4 + Z^4 = T^4$$

no té cap solució no trivial. Aquesta conjectura va ser refutada l'any 1987 per Noam Elkies, que va trobar la solució

$$2682440^4 + 15365639^4 + 18796760^4 = 20615673^4$$

amb una construcció geomètrica sofisticada que consisteix a submergir una corba el·líptica dins de la superfície que defineix la quàrtica anterior. De fet, Elkies aconseguí demostrar molt més: l'equació té en realitat infinites solucions no trivials, i la donada és només una de les més senzilles.

Conjectura de Catalan. El matemàtic belga Eugène Catalan va conjecturar el 1855 que els únics nombres enters consecutius que són potències (d'exponent >1) de nombres enters són

$$8 = 2^3 \quad \text{i} \quad 9 = 3^2.$$

En termes d'equacions diofàntiques això equival a dir que les equacions

$$X^n + 1 = Y^m$$

no tenen cap altra solució entera per a exponents $n, m \geq 2$ que la solució donada, corresponent a les coordenades $(x, y) = (2, 3)$ per a exponents $n = 3$, $m = 2$. La solució d'aquest problema ha estat obtinguda recentment pel matemàtic polonès Preda Mihailescu l'any 2003, amb una aplicació enginyosa de tècniques de teoria dels nombres algebraics i la teoria de cossos de classes que es remunten a finals del segle XIX i principis del segle XX.

2. CORBES EL·LÍPTIQUES

Una *corba el·líptica E* sobre un cos *K* és una corba de gènere 1 que tingui algun punt amb coordenades a *K*. Llevat d'isomorfisme, tota corba el·líptica es pot donar[1] amb una equació cúbica homogènia de la forma

1. Quan *K* és de característica 2 o 3 l'equació general d'una corba el·líptica és una mica més complicada.

$$E : Y^2Z = X^3 - aXZ^2 - bZ^3, \qquad a,b \in K, \qquad 4a^3 - 27b^2 \neq 0. \tag{5}$$

Aquesta equació sempre admet la solució $(x,y,z) = (0,1,0)$, amb coordenades al cos K, que representa l'únic punt projectiu de la corba a l'hiperplà $Z = 0$. En deshomogeneitzar l'equació respecte a la variable Z s'obté l'equació afí

$$E : Y^2 = X^3 - aX - b,$$

que és la manera habitual de representar les corbes el·líptiques. En treballar amb aquesta equació, els punts de la corba el·líptica amb coordenades a K són els donats per

$$E(K) = \{(x,y) \in K^2 \mid y^2 = x^3 - ax - b\} \cup \{\infty\}, \tag{6}$$

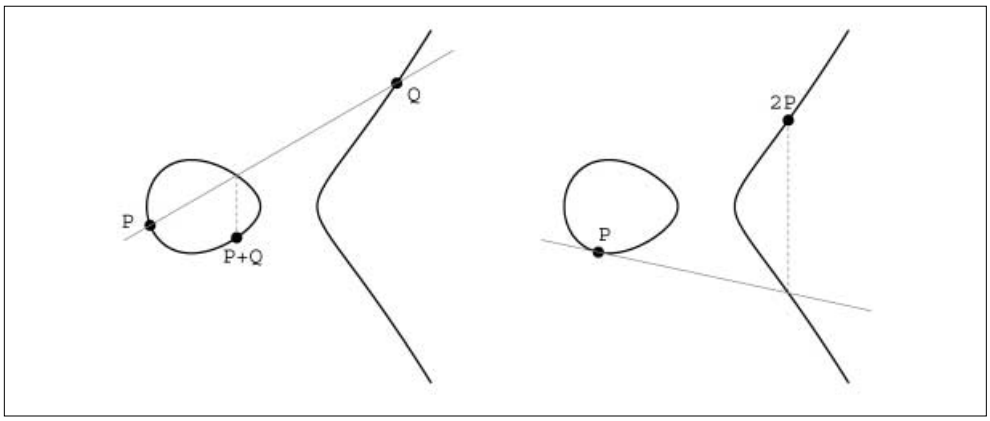
on el símbol ∞ denota el punt projectiu que s'ha quedat «a l'infinit» en passar de l'equació projectiva (5) a l'afí (6); els altres punts (x,y) són els punts afins de la corba.

La conjectura de Birch i Swinnerton-Dyer s'ocupa de les corbes el·líptiques sobre el cos $\mathbb{Q}$ dels nombres racionals. De la mateixa manera que els punts racionals de la circumferència de radi 1 corresponen a ternes pitagòriques primitives, els punts racionals d'una corba el·líptica definida sobre $\mathbb{Q}$ corresponen a solucions primitives de l'equació diofàntica (5).

La propietat fonamental de les corbes el·líptiques és que els seus punts tenen una estructura de grup natural que es pot definir per procediments geomètrics. En primer lloc s'observa que a cada punt afí (x,y) li correspon un punt $(x,-y)$ de manera que tots dos són simètrics respecte a l'eix de les x. L'estructura de grup es defineix de la manera següent:

1. L'element neutre és el punt ∞, i l'invers d'un punt afí (x,y) és el seu simètric $(x,-y)$.
2. Mètode de la corda. Si $P = (x,y)$ i $Q = (x',y')$ són dos punts afins diferents no simètrics l'un de l'altre, aleshores la recta que passa per tots dos talla la corba E en un tercer punt afí, i es defineix $P + Q$ com el simètric d'aquest tercer punt de tall.
3. Mètode de la tangent. Si $P = (x,y)$ és un punt afí que no és simètric d'ell mateix, aleshores la recta tangent a E que passa per aquest punt (amb multiplicitat almenys 2) talla la corba en un tercer punt afí, i es defineix $2P = P + P$ com el simètric d'aquest tercer punt de tall.

Els diagrames següents mostren el mètode de la corda i la tangent a una corba el·líptica sobre $\mathbb{R}$:



Una propietat fonamental que té aquesta estructura de grup és que, en sumar punts de la corba amb coordenades al cos K, s'obté un punt suma que també té coordenades a aquest mateix cos.

Per exemple, es considera la corba el·líptica d'equació afí

$$E : Y^2 = X^3 - 2X.$$

A part del punt de l'infinit es troben de seguida els dos punts de coordenades racionals $P=(2,2)$ i $Q=(0,0)$. El punt Q és simètric d'ell mateix i, per tant, $2Q=\infty$. En aplicar el mètode de la corda i la tangent a aquests punts es troben els altres punts

$$P=(2,2) \qquad Q=(0,0)$$

$$2P=\left(\frac{9}{4},\frac{-21}{8}\right) \qquad P+Q=(-1,1)$$

$$3P=(338,\,6214) \qquad 2P+Q=\left(\frac{-8}{9},\frac{-28}{27}\right)$$

$$4P=\left(\frac{12769}{7056},\frac{900271}{592704}\right) \qquad 3P+Q=\left(\frac{-1}{169},\frac{239}{2197}\right).$$

Si se segueix operant-los es veu que les coordenades dels punts que es van obtenint de seguida són nombres racionals amb numeradors i denominadors molt grans; per exemple

$$15P=\left(\frac{197970893765498628138595401}{126910706410112355831302500},\frac{1175500403585227043473605276082780828149}{1429708147909445625334240738443241375000}\right).$$

Es pot demostrar que tots els punts racionals d'aquesta corba són de la forma nP o de la forma $nP+Q$, per $n \in \mathbb{Z}$.

En termes de grups, això equival a dir que el grup dels punts racionals $E(\mathbb{Q})$ és finitament generat (per P i Q), i que de fet és isomorf al producte cartesià $\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ d'un grup cíclic infinit generat per P i un grup cíclic de dos elements generat per Q.

El resultat més important de què es disposa sobre els punts racionals de corbes el·líptiques és que en tots els casos n'hi ha prou amb un conjunt finit de punts per obtenir-los tots aplicant reiteradament el mètode de la corda i la tangent, igual que passa amb l'exemple que s'acaba de veure. Es tracta del

Teorema 2.1 (Mordell, 1920). *El grup* $E(\mathbb{Q})$ *dels punts racionals d'una corba el·líptica és finitament generat.*

Tot grup abelià finitament generat es pot descompondre en dos trossos: una part «de torsió» (finita) formada pels elements del grup que són d'ordre finit, i una part que és un «grup lliure» isomorf a un producte cartesià $\mathbb{Z}^r = \mathbb{Z} \times \cdots \times \mathbb{Z}$ de r còpies del grup $\mathbb{Z}$.

Per tant, es té una expressió

$$E(\mathbb{Q}) \simeq E(\mathbb{Q})_{\text{tors}} \times \mathbb{Z}^r, \tag{7}$$

de manera que tot punt de $E(\mathbb{Q})$ s'escriu de manera única de la forma $P_0 + n_1P_1 + \cdots n_rP_r$, $n_i \in \mathbb{Z}$, amb P_0 un punt de torsió i on els punts P_i formen una base de la part lliure de torsió.

Definició 2.2 (Rang). *El rang de la corba el·líptica* E *és el nombre* r *a la descomposició (7).*

Observi's que el conjunt $E(\mathbb{Q})$ de solucions racionals de (6) és finit si, i només si, el rang de la corba és $r = 0$, i també que el càlcul del grup dels punts racionals $E(\mathbb{Q})$ d'una corba el·líptica es redueix a calcular la seva part de torsió (un conjunt finit de punts) i calcular r punts que siguin base de la part lliure de torsió, ja que un cop calculats aquests tot altre punt es pot obtenir (de manera única) aplicant el mètode de la corda i la tangent.

Pel que fa a la part de torsió, el problema està resolt de manera molt satisfactòria. D'una banda, hi ha algorismes molt simples i ràpids que permeten trobar tots els punts de $E(\mathbb{Q})_{\text{tors}}$. De l'altra banda, en un cèlebre article de l'any 1978, Barry Mazur va demostrar que només hi ha quinze grups finits diferents que puguin ser el grup de torsió d'una corba el·líptica sobre $\mathbb{Q}$: els grups cíclics d'ordre k per a $1 \le k \le 10$ i per a $k = 12$, i els grups que són producte d'un cíclic d'ordre 2 per a un cíclic d'ordre $2k$, per a $1 \le k \le 4$.

En canvi, la part lliure de torsió sembla molt més difícil d'estudiar i, avui dia:

— No es coneix cap algorisme general per calcular el rang r d'una corba el·líptica (i encara menys generadors P_i). Tot i això, donada una corba, hi ha mètodes per trobar fites superiors per al rang r —anomenades *tècniques de descens* en honor a Pierre de Fermat— que, en alguns casos, permeten arribar a calcular el grup $E(\mathbb{Q})$ completament.

— No se sap quins valors pot prendre el rang d'una corba el·líptica. Es coneixen famílies infinites de corbes per a cada rang $0 \le r \le 11$ i exemples particulars de corbes amb rangs encara més grans: el rècord actual (de l'any 2000) és una corba de rang 24. Es conjectura que hi ha corbes amb rang arbitràriament gran.

3. LA L-SÈRIE D'UNA CORBA EL·LÍPTICA

Tots els objectes aritmètico-geomètrics interessants tenen associades funcions zeta i/o L-sèries. Es tracta de funcions d'una variable complexa s definides per sumes infinites del tipus

$$\sum_{n=1}^{\infty} \frac{a_n}{n^s}, \qquad a_n \in \mathbb{C},$$

anomenades *sèries de Dirichlet*. Aquestes sèries són convergents només en un semiplà dels complexos situat a la dreta d'una recta vertical d'abscissa $\rho \in \mathbb{R}$, anomenada *abscissa de convergència*. En general admeten prolongació analítica (o meromorfa) a tot el pla complex i gairebé sempre la part més interessant de la funció es troba a l'esquerra de l'abscisa de convergència, a llocs on la funció no es pot calcular a partir de la sèrie de Dirichlet que la defineix sinó que cal alguna forma alternativa (moltes vegades expressions integrals) que en doni la prolongació analítica.

L'exemple més conegut és el de la funció zeta de Riemann, que és la funció zeta associada al cos $\mathbb{Q}$ dels nombres racionals, i que ve donada per la sèrie de Dirichlet

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

amb abscissa de convergència $\rho = 1$. Admet prolongació meromorfa a tot el pla complex amb un únic pol en el punt $s = 1$.

El més interessant d'aquesta funció és el seu comportament al voltant de la recta vertical d'abscissa $1/2$ (a l'esquerra de l'abscissa de convergència), ja que els zeros de la funció en aquesta regió determinen la distribució dels nombres primers (vegeu l'article sobre la hipòtesi de Riemann en aquest mateix volum).

A tota corba el·líptica E sobre el cos dels nombres racionals se li associa una L-sèrie $L(E; s)$ de la manera següent. Es parteix d'una equació afí per a la corba de la forma

(6) on els coeficients a, b es pot suposar que són nombres enters. Per a cada nombre primer p sigui N_p el nombre que compta les solucions de la congruència corresponent:

$$N_p = |\{(x,y) \in (\mathbb{Z}/p\mathbb{Z})^2 : y^2 \equiv x^3 - ax - b \pmod{p}\}| + 1$$

on el $+1$ es posa per tenir en compte el punt de l'infinit. Un dels resultats més importants sobre corbes el·líptiques, conjecturat per Artin i demostrat per Hasse a la seva tesi els anys 30, diu que aquest nombre és aproximadament $1+p$, amb error fitat per $2\sqrt{p}$. És a dir, que per a tot nombre primer p es compleix la desigualtat

$$1+p-2\sqrt{p} \le N_p \le 1+p+2\sqrt{p}.$$

A partir d'aquests nombres de solucions es defineix la L-sèrie de la corba E de la manera següent:

$$L(E;s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s} = \prod_p \left(1 - \frac{1+p-N_p}{p^s} + \frac{p}{p^{2s}}\right)^{-1}$$

on el producte de la dreta s'estén sobre tots els nombres primers[2]. Els coeficients a_n de la sèrie de Dirichlet que defineix la L-serie $L(E;s)$ són doncs nombres enters que depenen del nombre de solucions de la congruència obtinguda en reduir la corba mòdul n. Per exemple, la L-sèrie de la corba el·líptica d'equació $Y^2 = X^3 - 2X$ considerada anteriorment comença així:

$$L(E;s) = 1 + \frac{-4}{5^s} + \frac{-3}{9^s} + \frac{-4}{13^s} + \frac{-2}{17^s} + \frac{11}{25^s} + \frac{-4}{29^s} + \frac{12}{37^s} + \frac{-10}{41^s} + \frac{12}{45^s} + \frac{-7}{49^s} + \cdots$$

L'afitació de Hasse dels nombres N_p té com a conseqüència immediata que l'abscissa de convergència de $L(E;s)$ és $\rho = 3/2$, de manera que la L-sèrie d'una corba el·líptica està definida només per a valors de la variable que estiguin al semiplà a la dreta de la recta vertical que passa pel punt $1{,}5 = 3/2$.

Hasse va conjecturar que la funció $L(E;s)$ admet prolongació analítica a tot el pla complex. Fins ara l'única manera que s'ha trobat de demostrar aquesta prolongació analítica ha estat utilitzar la modularitat de les corbes el·líptiques. La modularitat ha estat una de les propietats més estudiades pels teoristes de nombres les darreres dècades. Les corbes modulars formen una família $X_1(N)$ de corbes definides sobre $\mathbb{Q}$ parametritzades per un enter $N \ge 1$, anomenat el *conductor de la corba*; intentar donar aquestes corbes

2. Els primers p que divideixen el discriminant de la corba $4a^3 - 27b^2$ requereixen modificacions al factor corresponent que aquí, per simplificar, no es tenen en compte.

amb equacions és poc útil ja que les equacions «naturals» tenen grau que augmenta ràpidament amb N i coeficients extraordinàriament complicats. En canvi, la manera natural de treballar amb aquestes corbes, que és en el fons el motiu pel qual són interessants, és tenir en compte el fet que són espais de mòduli d'objectes aritmètico-geomètrics relacionats amb corbes el·líptiques.

La conjectura de Shimura-Taniyama, formulada cap al 1955, prediu que tota corba el·líptica E definida sobre $\mathbb{Q}$ «és modular», on ser modular vol dir, simplement, que hi ha un morfisme (no constant) $X_1(N) \to E$ d'alguna corba modular $X_1(N)$ en la corba E. És relativament fàcil comprovar que tota corba el·líptica modular satisfà la conjectura de Hasse, i aquest camí de la modularitat és l'única manera que es coneix d'obtenir la prolongació analítica de la L-sèrie d'una corba el·líptica.

Fins a mitjan anys 90 la modularitat només s'havia demostrat per a una col·lecció finita de classes d'isomorfisme de corbes el·líptiques. L'any 1995 es va produir un avenç espectacular quan Andrew Wiles va publicar la seva demostració del fet que totes les corbes el·líptiques d'un cert tipus anomenades *semiestables* són modulars, del qual es dedueix l'últim teorema de Fermat. Més endavant les tècniques de Wiles s'han anat perfeccionant i aplicant a famílies més i més grans de corbes fins que finalment l'any 2000, i gràcies a la contribució de diversos matemàtics, entre ells alguns estudiants de Wiles, s'ha obtingut la demostració completa de la modularitat de tota corba el·líptica sobre $\mathbb{Q}$. En particular, doncs, avui sabem que la L-sèrie de tota corba el·líptica sobre els racionals admet prolongació analítica a tot $\mathbb{C}$.

4. LA CONJECTURA DE BIRCH I SWINNERTON-DYER

Bryan Birch (Burton-on-Trend, 1931) i Peter Swinnerton-Dyer (Ponteeland, 1927) van estudiar tots dos matemàtiques al Trinity College (Cambridge). Birch va llegir la tesi a Cambridge l'any 1958, dirigida per J.W.S. Cassels. Swinnerton-Dyer va ser estudiant de Littlewood; en acabar la tesi a mitjan anys 50 va passar un any a Chicago treballant amb André Weil, i després va tornar a Cambridge.

A finals dels anys 50 tots dos es van trobar treballant al laboratori de càlcul de la Universitat de Cambridge pel mateix motiu: no haver trobat una feina en «matemàtiques pures». Aquest laboratori és un organisme pioner en la utilització de computadors electrònics: els computadors EDSAC[3] I i EDSAC II es van posar en marxa, respectivament, els anys 1948 i 1958. Va ser amb aquest segon que Birch i Swinnerton-Dyer van fer els càlculs que els van portar a enunciar la conjectura.

3. Electronic Delay Storage Automatic Calculator.

Birch i Swinnerton-Dyer van decidir estudiar el grup de punts racionals $E(\mathbb{Q})$ de les corbes el·líptiques. Per començar van desenvolupar un procediment per intentar el càlcul d'aquest grup basat en la idea mateixa de la demostració del teorema de Mordell, procediment conegut amb el nom de *2-descens*. La idea és la següent: a cada corba el·líptica E sobre els racionals se li pot associar una col·lecció finita de corbes de gènere 1 sobre els racionals anomenades *2-recobriments de la corba de partida*. Aquestes corbes admeten una equació quàrtica afí de la forma

$$Y^2 = aX^4 + bX^3 + cX^2 + dX + e, \qquad a, b, c, d, e \in \mathbb{Z} \tag{8}$$

tal que les congruències corresponents tenen solució per a tot mòdul, però poden tenir o poden no tenir un punt (una solució) racional. És a dir, són corbes que poden satisfer o no el principi de Hasse.

El conjunt de tots aquests 2-recobriments es pot dotar d'estructura de grup abelià de manera natural, i s'obté així l'anomenat *2-grup de Selmer* $Sel^{(2)}(E/\mathbb{Q})$ de la corba E sobre $\mathbb{Q}$. Es tracta d'un grup amb una estructura molt senzilla: és un producte cartesià d'uns quants grups cíclics d'ordre 2. El subconjunt format pels 2-recobriments que tenen un punt racional (els que sí satisfan el principi de Hasse) forma un subgrup del grup de Selmer, que és isomorf a $E(\mathbb{Q})/2E(\mathbb{Q})$. El quocient corresponent s'anomena *2-component del grup de Tate-Shafarevich* i es denota $Sha^{(2)}(E/\mathbb{Q})$. Es té, doncs, una successió exacta de grups

$$0 \to E(\mathbb{Q})/2E(\mathbb{Q}) \to \mathrm{Sel}^{(2)}(E/\mathbb{Q}) \to \mathrm{Sha}^{(2)}(E/\mathbb{Q}) \to 0.$$

Per calcular el grup $E(\mathbb{Q})$, i en particular el rang de la corba, n'hi ha prou a saber calcular el grup de l'esquerra d'aquesta successió. El grup de Selmer $\mathrm{Sel}^{(2)}(E/\mathbb{Q})$ es pot calcular, tot i que moltes vegades aquest càlcul sol resultar prou llarg i complicat. Una part important del primer dels dos articles on Birch i Swinnerton-Dyer van enunciar la conjectura es dedica a descriure un algorisme per portar a terme aquest càlcul del grup de Selmer en un nombre finit de passos, algorisme que Birch va implementar per poder fer càlculs amb l'EDSAC.

La part difícil per al càlcul de $E(\mathbb{Q})$, que encara no està resolta avui dia, consisteix a calcular el grup de Tate-Shafarevich.

Els seus elements representen corbes del tipus (8) que no satisfan el principi de Hasse, i en general no se sap demostrar que una corba com aquestes no té punts racionals. En canvi, el recíproc és trivial si s'està de sort: n'hi ha prou a trobar un punt racional. Així, sempre que en calcular el grup de Selmer $\mathrm{Sel}^{(2)}(E/\mathbb{Q})$ siguem capaços de trobar un punt a cadascun dels 2-recobriments que en formen part, haurem demostrat que el grup de Tate-Shafarevich $\mathrm{Sha}^{(2)}(E/\mathbb{Q})$ és trivial per a la corba en qüestió, i a sobre aquests punts ens permetran calcular tot el grup $E(\mathbb{Q})$.

A la pràctica és relativament freqüent que el grup $Ш^{(2)}(E/\mathbb{Q})$ sigui trivial, i, per tant, si es busca prou s'acabaran trobant punts racionals a tots els 2-recobriments. El problema apareix quan després de molt buscar no es troben punts a algunes d'aquestes corbes (8), ja que mai s'estarà segur de si el motiu és que no n'hi ha (i, per tant, es tracta d'elements de Tate-Shafarevich) o que encara no ha buscat prou i resulta que sí que hi ha punts racionals, però tots tenen coordenades molt «complicades», fora de la mida en què s'ha buscat.

El primer que van fer Birch i Swinnerton-Dyer va ser programar aquest algorisme i executar-lo sobre uns quants milers de corbes el·líptiques. En alguns casos no els va ser possible calcular el rang de la corba, ja que hi havia 2-recobriments on no trobaven punts, i van despreciar les corbes corresponents. En canvi, en molts altres trobaven un punt racional a cada 2-recobriment i aleshores havien calculat exactament el rang de la corba. D'aquesta manera van elaborar taules amb algunes dotzenes de corbes el·líptiques de rangs $r = 0, 1, 2$ i fins i tot van trobar alguna corba de rang $r = 3$.

A continuació, i inspirant-se en treballs de Siegel on es fan càlculs anàlegs per a formes quadràtiques, van començar a buscar alguna relació entre el rang de la corba i el nombre N_p de solucions mòdul p de les congruències corresponents.

La idea subjacent és que si una corba té molts punts racionals (rang gran) aquests punts, en reduir mòdul p, produiran un nombre especialment elevat de solucions de la congruència (naturalment, sempre dins de l'interval de Hasse). Concretament, van estudiar per a cadascuna de les corbes de les quals havien pogut calcular el rang el creixement de la funció

$$f(P) = \prod_{p<P} \frac{N_p}{p},$$

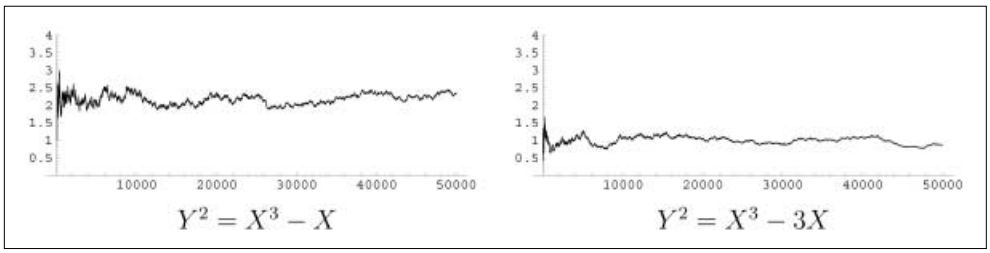
que en certa manera mesura el nombre de solucions de la congruència mòdul tots els primers menors que una fita donada, donant més pes a les solucions mòdul els primers més petits. Estudiant aquestes funcions van observar experimentalment el comportament següent:

1. Per a corbes de rang $r = 0$ la funció $f(P)$ es manté fitada i el seu valor oscil·la al voltant d'un nombre positiu C.
2. Per a corbes de rang $r > 0$ la funció $f(P)$ tendeix a infinit, i la rapidesa amb què tendeix depèn del rang com

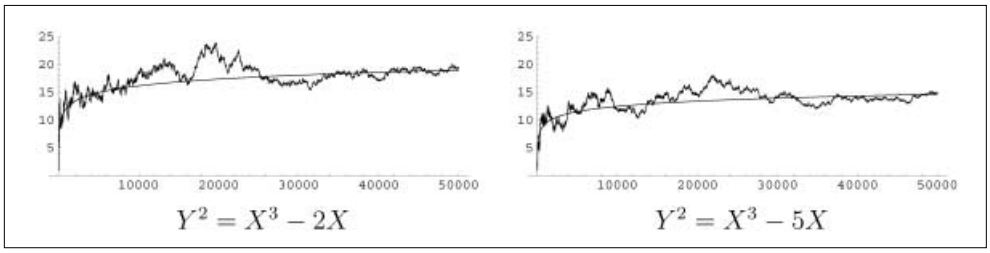
$$f(P) \sim C(\log P)^r \qquad \text{si} \qquad P \to \infty$$

per a alguna constant positiva C (que depèn de la corba).

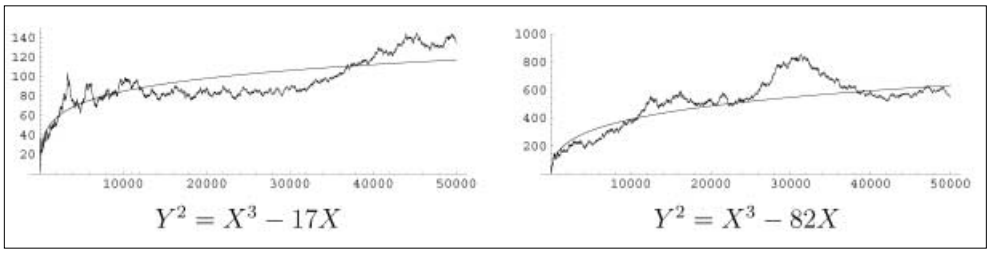
Per exemple, a continuació es dibuixen les gràfiques de $f(P)$ per a dues corbes de rang zero



dues corbes de rang 1



i corbes de rangs 2 i 3, respectivament



on per a les corbes de rang positiu s'ha dibuixat també la gràfica de la funció $C(\log P)^r$ amb un valor de C adequat.

Tot i que Birch i Swinnerton-Dyer esperaven trobar alguna relació, el fet que aquesta fos tan evident i precisa els va sorprendre i els va convèncer que havien topat amb alguna cosa important. Una simple ullada a la definició de la funció $f(P)$ i a l'expressió com a producte infinit que defineix la L-sèrie $L(E;s)$ de la corba, mostra que la funció $f(P)$ tendeix formalment (és a dir, sense tenir en compte la convergència) cap a l'invers del valor de la funció L en el punt $s = 1$. Per tant, el fet que aquest valor tendeixi a infinit s'hauria de traduir en $L(E;1) = 0$, i la rapidesa de creixement hauria d'estar relacionada amb l'ordre d'aquest zero.

Això motivà la

Conjectura 4.1 (Birch i Swinnerton-Dyer). *El rang d'una corba el·líptica* E *sobre els racionals coincideix amb l'ordre del zero de la seva* L*-serie en el punt* s = 1

$$\operatorname{rank} E(\mathbb{Q}) = \operatorname{ord}_{s=1} L(E;s).$$

S'ha de tenir en compte que en el moment en què es va formular aquesta conjectura no era encara segur que el nombre que apareix a la dreta de la igualtat estigués definit, ja que la prolongació analítica de la funció $L(E;s)$ a l'esquerra de la seva abscissa de convergència $\rho = 3/2$ era només conjectural.

A continuació Birch i Swinnerton-Dyer van estudiar la constant C a què tendeix $f(P)$ en el cas de rang zero, i van observar que, conjecturalment, aquesta constant empaqueta informació d'invariants tant aritmètics com geomètrics de la corba. Poc després John Tate va generalitzar aquesta conjectura a corbes de rang arbitrari proposant la fórmula següent per al primer coeficient del desenvolupament de Taylor de la funció $L(E;s)$ al voltant del punt $s = 1$

Conjectura 4.2 (Birch i Swinnerton-Dyer forta). *Suposant la conjectura anterior,*

$$\frac{L^{(r)}(E;1)}{\mathrm{r!}} = \frac{\Omega_E \,|\, \mathrm{Sha(E/\mathbb{Q})} \,|\, \mathrm{Reg(E/\mathbb{Q})} \prod c_p}{|\mathrm{E}(\mathbb{Q})_{\mathrm{tors}}|^2}.$$

Dels invariants de la corba E que apareixen a la part dreta en aquesta fórmula, n'hi ha un que presenta dificultats i té un interès especial. Es tracta del grup $\mathrm{Sha}(E/\mathbb{Q})$, anomenat *grup de Tate-Shafarevich de la corba*, que està format per totes les classes d'isomorfisme sobre $\mathbb{Q}$ de corbes isomorfes a E sobre els nombres complexos que no satisfan el principi de Hasse. El grup $\mathrm{Sha}^{(2)}(E/\mathbb{Q})$ que obstrueix el càlcul del rang a través d'un 2-descens, del qual s'ha parlat anteriorment, és només la 2-torsió del grup de Tate-Shafarevich tot sencer.

En el moment de formular la conjectura no s'havia aconseguit demostrar que el grup $\mathrm{Sha}(E/\mathbb{Q})$ fos finit ni tan sols per una sola corba el·líptica E. Fins als anys 80 no es va aconseguir demostrar la finitud de Sha per a algunes corbes (Rubin i Kolyvagin). Actualment es coneix aquesta finitud per a corbes tals que la seva L-sèrie no s'anul·li al punt $s = 1$ o bé s'hi anul·li amb multiplicitat 1.

Amb relació a aquesta situació i a la fórmula de la conjectura forta, John Tate va escriure: «*This remarkable conjecture relates the behavior of a function L at a point where it is not at present known to be defined to the order of a group* Sha *which is not known to be finite!*» (*Invent. Math*. vol 23, 1974, pàg. 198).

El problema del mil·lenni es refereix a la conjectura 4.1. És a dir, per obtenir el milió de dòlars «només» cal demostrar que el rang d'una corba el·líptica racional coin-

cideix sempre amb l'ordre del zero de la seva L-sèrie en el punt $s=1$, i no cal demostrar la fórmula que prediu el valor del primer coeficient de Taylor no nul en aquest punt.

Tot i això, hi ha moltes generalitzacions de les conjectures 4.1 i 4.2 en direccions diverses: per a corbes el·líptiques sobre cossos de nombres, per a varietats abelianes sobre $\mathbb{Q}$ o també sobre cossos de nombres, per a corbes el·líptiques sobre cossos de funcions, o generalitzacions que relacionen rangs de grups de cicles algebraics amb ordres d'anul·lació de certes funcions.

Fins avui s'han publicat essencialment dos grans resultats sobre la conjectura, deixant de banda la modularitat de les corbes el·líptiques que permet la prolongació analítica, i que té com a conseqüència el fet que l'enunciat mateix de la conjectura tingui sentit. El primer resultat val només per a corbes el·líptiques que tenen multiplicació complexa (sobre els racionals, de corbes com aquestes n'hi ha només 11 classes d'isomorfisme diferents), i assegura que quan la corba té algun punt d'ordre infinit aleshores la seva L-sèrie s'anul·la:

Teorema 4.3 (Coates-Wiles, 1973). *Si* E *és una corba el·líptica sobre els racionals amb multiplicació complexa, aleshores*

$$\mathrm{Rang}\,(\mathrm{E}(\mathbb{Q})) \geq 1 \;\Rightarrow\; L(E;1)=0.$$

El segon és una combinació de resultats de Gross i Zagier de l'any 1983 i de Kolyvagin dels anys 90 i és vàlid per a corbes el·líptiques modulars (actualment ja sabem que totes ho són, de modulars, però en el moment d'enunciar aquests resultats la modularitat semblava estar encara molt lluny).

El resultat assegura que si l'ordre de la L-sèrie és zero o 1 aleshores el grup de Tate-Shafarevich és finit (aquesta és la part de Kolyvagin) i es compleix la conjectura forta (això es demostra utilitzant punts de Heegner i és el treball de Gross-Zagier).

Teorema 4.4 (Gross-Zagier, 1983, Kolyvagin, 1990). *Si* E *és una corba el·líptica sobre els racionals tal que* $\mathrm{ord}_{s=1} L(E;s) \leq 1$ *(o sigui, zero o u), aleshores el grup de Tate-Shafarevich de la corba és finit, i la corba satisfà la conjectura forta.*

5. UNA APLICACIÓ: NOMBRES CONGRUS

Un manuscrit àrab del segle X planteja el problema següent:

> Determinar els nombres enters n que són àrea d'un triangle rectangle amb tots tres costats racionals.

Els nombres amb aquesta propietat es diuen *congrus*. Per exemple 5, 6 i 7 ho són, ja que són l'àrea dels triangles de costats

$$\left(\frac{20}{3}, \frac{3}{2}, \frac{41}{6}\right), \qquad (3,4,5) \qquad \left(\frac{24}{5}, \frac{31}{12}, \frac{337}{60}\right),$$

respectivament. En canvi, ja Fermat va demostrar que 1 no és congru i, de fet, per a $n \leq 12$ els únics enters congrus són els tres anteriors.

Saber si un enter n és congru és equivalent a decidir sobre la resolubilitat del sistema d'equacions

$$a^2+b^2=c^2, \qquad ab=2n, \qquad a,b,c \in \mathbb{Q},$$

en què la primera indica que (a,b,c) són els tres costats d'un triangle rectangle i la segona que l'àrea d'aquest triangle és el nombre n. Partint d'una solució (a,b,c) els nombres racionals $x=(c/2)^2$ i $y=c(b^2-a^2)/8$ són un punt racional de la corba el·líptica d'equació

$$E_n : Y^2=X^3-n^2X$$

amb $y \neq 0$ i, recíprocament, és un exercici senzill comprovar que tot punt racional de la corba E_n amb coordenada $y \neq 0$ permet fabricar una solució del problema invertint el procés anterior. D'altra banda, el grup de torsió $E_n(\mathbb{Q})_{\text{tors}}$ està format per només 4 punts: el de l'infinit i els tres punts (0,0), (n,0), (–n,0) amb coordenada y igual a zero. Per tant, el problema de saber si un enter positiu n és o no congru és equivalent a saber si la corba E_n té algun punt que no sigui de torsió; és a dir,

n és congru si, i només si, E_n té rang $r>0$.

Aplicant els teoremes de Coates-Wiles i de Gross-Zagier, que s'han citat a la secció anterior (les corbes E_n són de multiplicació complexa i, en particular, als anys 80 ja se sabia que són modulars) Jerry Tunnell va aconseguir demostrar el criteri següent.

Teorema 5.1 (Tunnell, 1983). *El nombre enter positiu* n *senar (i lliure de quadrats) és congru si, i només si, el nombre de solucions enteres de l'equació*

$$2X^2+Y^2+8Z^2=n$$

és el doble del nombre de solucions enteres de

$$2X^2+Y^2+32Z^2=n.$$

Es tracta d'un criteri fàcil d'aplicar, ja que per comptar el nombre de solucions d'aquestes equacions n'hi ha prou a recórrer conjunts finits de valors x, y, z. Hi ha un criteri anàleg per nombres n parells. En canvi, un cop decidit que el nombre és congru, trobar efectivament un triangle rectangle que el tingui per àrea pot no ser fàcil. Per exemple, si $n = 157$ les equacions anteriors no tenen solució cap de les dues; per tant, n és congru. En canvi, el triangle rectangle de costats racionals més «senzill» d'àrea 157 té hipotenusa

$$\frac{224403517704336969924557513090667486316094847204 1}{891233226892885958802553517896716357001648083 0}$$

i s'obté a partir d'un punt d'ordre infinit de la corba E_{157}.

REFERÈNCIES

[1] B. J. Birch; H. P. F. Swinnerton-Dyer. «Notes on Elliptic Curves». *J. Reine Angew. Math*. 212, (1963), p. 7-25.

[2] B. J. Birch; H. P. F. Swinnerton-Dyer. «Notes on Elliptic Curves II». *J. Reine Angew. Math.* 218, (1965), p. 79-108.

[3] J. Coates, A. Wiles. «On the Conjecture of Birch and Swinnerton-Dyer». *Invent. Math.* 39(3), (1977), p. 223-251.

[4] B. H. Gross; D. B. Zagier. «Heegner points and derivatives of L-series». *Invent. Math.* 84, (1986), p. 225-320.

[5] V. A. Kolyvagin. «Euler systems», *The Grothendieck festschrift, Prog. in Math., Birkhäuser* (1990).

Criptografia, complexitat i geometria

Anna Rio
Departament de Matemàtica Aplicada II.
Universitat Politècnica de Catalunya

1. INTRODUCCIÓ

El tema que inicialment ens ocupa, la criptografia, té lligams amb gairebé totes les branques de la matemàtica i la informàtica. L'objectiu d'aquesta exposició és mostrar que la teoria de la complexitat computacional ha passat a ser un ingredient essencial de la criptografia actual i que la geometria aritmètica subministra problemes interessants de cara al disseny de protocols criptogràfics. Per tant, no es tracta de presentar cap problema per a la recerca futura sinó d'analitzar un dels problemes del mil·lenni, $\mathbf{P \neq NP}$, des del punt de vista de la seva implicació en la seguretat criptogràfica dels sistemes que es troben en funcionament.

De tota manera, pel context del cicle de conferències en el qual ens trobem, resulta natural fer algunes reflexions sobre el futur de la criptografia. Amb motiu del canvi de mil·lenni es van escriure alguns articles sobre aquest tema i m'ha semblat oportú plantejar aquí algunes de les qüestions que es tracten i algunes de les opinions expressades.

1.1. Criptografia del mil·lenni?

En l'article «Ultimate Cryptography [5]», Whitfield Diffie, que citarem més endavant com a pioner de la criptografia de clau pública, comença afirmant que la idea de predir com serà la criptografia d'aquí a mil anys és quelcom que intimida, ja que es tracta d'especular sobre si hi haurà necessitat de guardar secrets i la manera de fer-ho. La criptografia actual és un elaborat edifici computacional, matemàtic i electrònic; però els conceptes comunament acceptats dins aquesta teoria tot just van aparèixer a la segona meitat del segle XX i eren impossibles d'imaginar l'any 1901. Així doncs, conclou que tenim unes possibilitats bastant minses de predir les *capacitats* que demanarem a la criptografia del futur, o les que aquesta ens oferirà.

Suggereix al final de l'article que potser la qüestió més fonamental és la següent: l'any 3001, *hi haurà* algun tipus de criptografia? Deixem per a la reflexió tres aspectes que planteja al voltant d'aquesta qüestió:

— probablement, millorar les comunicacions porta l'espècie humana cap a una organització molt més unida;

— en un entorn altament integrat encara hi hauria seguretat, però la criptografia podria no ser un mecanisme adient;

— al cos humà trobem autenticació de tipus criptogràfic en el sistema immunològic, però no xifratge de missatges,

i que cadascú opini sobre si al tercer mil·lenni la criptografia serà merament un record llunyà.

1.2. Criptografia del segle?

Si limitem al segle XXI les nostres pretensions de futuròlegs, potser aviat ens trobarem amb l'afirmació que aquesta ha d'ésser l'*era quàntica*, així com el segle XX ha estat l'era de la informació i el segle XIX, la de les màquines. Tal com diu Fausto Montoya a la seva conferència «La criptografía cuántica, ¿realidad o ficción?» [15], una gran quantitat d'articles anuncien que la criptografia quàntica serà la criptografia del futur i els més atrevits proclamen que la criptografia clàssica ha mort i només es pot considerar segura la seva variant quàntica. Els seus arguments es basen en el fet que l'aparició d'ordinadors quàntics faria més fàcil la cerca exhaustiva, la cerca de col·lisions per a funcions *hash* o la factorització de nombres enters (cf. [18]).

Cal tornar, doncs, a l'únic sistema incondicionalment segur conegut: el xifratge de Vernam, també anomenat *one-time pad*, satisfà les condicions de secret perfecte de Shannon (cf. [17]).

missatge	00011 01111 01101 00101
$\oplus$	
clau	11011 00101 01011 00110
$\downarrow$	
criptograma	11000 01010 00110 00011
$\oplus$	
clau	11011 00101 01011 00110
$\downarrow$	
missatge	00011 01111 01101 00101

En aquest criptosistema, emissor i receptor comparteixen una clau secreta que és una cadena aleatòria de bits de la mateixa longitud del missatge i d'un sol ús. Aquesta darrera restricció fa que en les condicions actuals no es pugui utilitzar en infraestructures de comunicació per a grans quantitats d'usuaris o grans fluxos d'informació. La crip-

tografia quàntica consisteix a fer una *distribució quàntica de claus* que faci pràctic aquest xifratge de Vernam. Un cop emissor i receptor disposen de la clau, ja poden fer servir un «XOR clàssic» i un canal de transmissió clàssic.

L'autor citat situa la criptografia quàntica actual com un camp molt interessant d'investigació en física, però encara no com un candidat sòlid a prendre el relleu a curt termini. De tota manera, cal estar molt atents als progressos d'aquest camp, perquè de mica en mica es van batent rècords en la velocitat i l'abast de les transmissions quàntiques. Projectes com

— SECOQC (*Development of a Global Network for Secure Communication based on Quantum Cryptography*) endegat a Àustria l'abril de 2004 i finançat amb 11,4 milions d'euros (http://www.secoqc.net/index.html) per la Unió Europea;

— *Quantum cryptography and entanglement* de l'Institut de Ciències Fotòniques (http://www.icfo.es),

i molts d'altres, poden fer que en poc temps haguem passat de la ficció a la realitat.

1.3. Criptografia de la dècada?

Si reduïm encara més l'àmbit d'aquesta especulació sobre el desenvolupament de la criptografia, trobem l'article de Ueli Maurer «Cryptography 2000±10» [13] que considera que el principal tema de recerca immediata és l'afebliment de les hipòtesis en què es basa la seguretat dels criptosistemes actuals. L'autor classifica aquestes hipòtesis en les tres categories següents:

— Hipòtesis d'*intractabilitat computacional* de certs problemes matemàtics. Atès que no es coneixen proves útils i generals de la dificultat computacional de cap problema, la criptografia depèn per complet d'aquestes hipòtesis.

— Hipòtesis de *confiança* en certes autoritats, entitats i components del sistema.

— Hipòtesis *físiques* sobre el canal. Computació i comunicació són processos físics més que objectes matemàtics idealitzats i, com a tals, involucren soroll, efectes quàntics i altres factors d'incertesa.

L'objectiu del disseny criptogràfic prudent és fer explícites aquestes hipòtesis i tan febles com sigui possible. Reduir les hipòtesis necessàries i els requeriments de confiança per assolir un cert objectiu de seguretat és una qüestió primordial de la recerca actual i futura, perquè els components crucials per a l'emergent infraestructura d'informació global hauran d'ésser sistemes que operin de manera fiable i segura en un entorn potencialment molt advers. La societat de la informació dependrà fortament de la seguretat de sistemes criptogràfics, en particular dels esquemes de signatura digital. Per això,

la cripotanàlisi és una font potencial de desastres. Per exemple, podria ser que en els propers anys la descoberta i publicació d'un algoritme ràpid de factorització arribés a significar un cert col·lapse de l'economia mundial, llevat que siguem extremadament curosos en el disseny de protocols, la tria d'algoritmes i la legislació.

1.4. Com guanyar un milió de dòlars?

Per acabar, si ens fixem en l'aspecte més pragmàtic de l'etiqueta «problema del mil·lenni», és a dir, en la seva remuneració econòmica, i cerquem una recompensa similar per a problemes específicament criptogràfics, és molt probable que el primer lloc que visitem sigui http://rsasecurity.com/rsalabs/challenges, la pàgina web dedicada als *reptes de factorització RSA*. %No hi ha un AES-Challenge

Hi trobarem una taula com aquesta

Resolt RSA-576	$10,000	RSA-896	$75,000
Resolt RSA-640	$20,000	RSA-1024	$100,000
RSA-704	$30,000	RSA-1536	$150,000
RSA-768	$50,000	RSA-2048	$200,000

que ens indica la mida en bits del nombre enter que cal factoritzar (sabent que és producte de dos primers) i el premi que s'atorga al primer que ho aconsegueix. La resolució del RSA-576 va ser anunciada el desembre de 2003 per J. Franke. El premi màxim de 200.000 dòlars s'obtindria factoritzant el nombre següent, que té 617 xifres decimals:

25195908475657893494027183240048398571429282126204032027777137836043662020707595556264018525880784406918290641249515082189298559149176184502808489120072844992687392807287776735971418347270261896375014971824691165077613379859095700097330459748808428401797429100642458691817195118746121515172654632282216869987549182422433637259085141865462043576798423387184774479207399342365848238242811981638150106748104516603773060562016196762561338441436038339044149526344321901146575444541784240209246165157233507787077498171257724679629263863563732899121548314381678998850404453640235273819513786365643912120103971228221207203 57.

Tot i que la remuneració oferta per l'empresa RSA Security queda lluny del milió de dòlars, és més que probable que si la factorització d'aquest nombre s'aconseguís gràcies a la descoberta d'un bon algoritme de factorització, aleshores la repercussió a tots els nivells aniria molt més enllà que el simple fet d'aconseguir aquesta quantitat de diners.

2. CRIPTOGRAFIA

Fins després de la II Guerra Mundial la criptografia era un art d'ús pràcticament limitat als àmbits militars i diplomàtics, concentrat quasi exclusivament en l'encriptació de missatges, que utilitzava sistemes *ad hoc* sense gairebé cap justificació teòrica de la seva seguretat. En contraposició, la criptografia moderna està considerada com una ciència amb multitud d'aplicacions en els sistemes d'informació i comunicacions, implementada mitjançant elaborats protocols criptogràfics dissenyats conjuntament per matemàtics i informàtics, amb criteris cada cop més precisos i estrictes pel que fa als conceptes d'eficiència i seguretat.

2.1. Criptografia moderna

Una brevíssima cronologia del naixement d'aquesta ciència moderna podria ser la següent:

1938. El govern britànic instal·la a Bletchey Park la Foreign Office's Code and Cipher School.

1940. S'incorpora *Alan Turing*. Durant la II Guerra Mundial, deu mil persones hi treballen. El repte per a matemàtics i criptoanalistes és trencar el criptosistema *Enigma*, que utilitzava l'exèrcit alemany.

1944. Es posa en funcionament *Colossus*, la primera computadora electrònica programable.

Tot i l'enorme simplificació que hem fet, serveixi aquest apunt per constatar una interacció curiosa: d'una banda, la criptologia es troba en el naixement dels ordinadors i de la ciència informàtica, ja que la necessitat de fer cerques exhaustives inabastables va forçar la construcció de *màquines calculadores*; de l'altra, l'aparició dels ordinadors i de la transmissió d'informació usant nous canals de comunicació va canviar radicalment les demandes criptogràfiques i les tècniques criptoanalítiques.

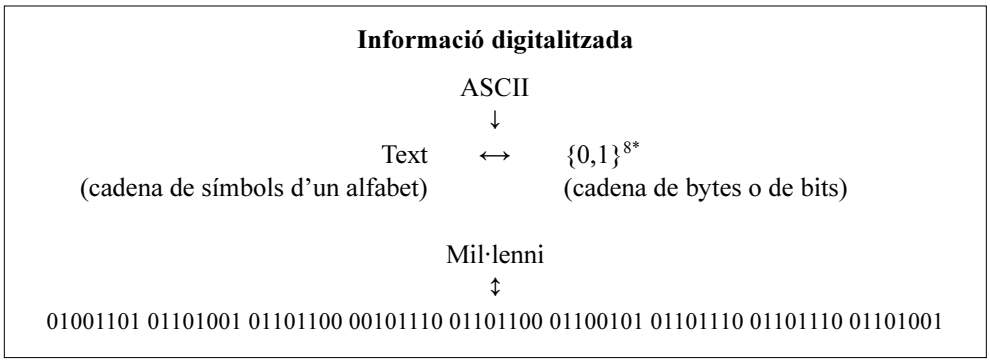
De les moltes contribucions a la transició de l'antic art de la criptografia cap a la ciència moderna que considerem avui en dia, és imprescindible esmentar dues peces clau. En primer lloc, tenim el treball de C.E. Shannon, que després d'haver establert la teoria matemàtica de la informació, l'any 1949 va publicar l'article «Communication theory of secrecy systems» [17]. Shannon considera el llenguatge com un procés estocàstic que proporciona cadenes de símbols de l'alfabet corresponent. Partint de les hipòtesis bàsiques de clau d'un sol ús i limitació a un atac sobre text xifrat, un sistema criptogràfic és perfectament segur si el text i el criptograma són estadísticament independents, és a dir, si la probabilitat d'un missatge *a posteriori*, conegut el criptograma, és igual a la seva probabilitat

a priori. En termes d'entropia, la informació sobre el missatge aportada pel criptograma és nul·la, independentment de temps i recursos computacionals. Basant-se en aquest concepte d'entropia, a l'article esmentat Shannon demostra dos teoremes fonamentals. D'una banda, per tal de tenir un sistema perfectament segur és condició necessària que la longitud de la clau sigui més gran o igual que la longitud del missatge. De l'altra, existeixen sistemes perfectament segurs, en concret, el xifratge de Vernam ho és.

En segon lloc, cal ressaltar l'article de Diffie i Hellman «New Directions in Cryptography» [6] que estableix els conceptes fonamentals de la criptografia de clau pública, incloent-hi la producció de signatures digitals, i posa les bases per a l'ús massiu de protocols criptogràfics. La criptografia de clau pública és una invenció força paradoxal, ja que fa possible transmetre informació secreta a través d'un canal insegur sense necessitat d'acordar prèviament una clau. L'altre gran tret característic és que proporciona un anàleg electrònic de la firma escrita, la qual cosa obre les portes a qüestions de tan ampli abast com ara el comerç electrònic o la realització de tràmits i gestions de forma no presencial.

Ciència multidisciplinar

L'objectiu de la criptografia moderna és proporcionar seguretat a les tecnologies de la informació i la comunicació, el conjunt de tecnologies usades per processar, transmetre i emmagatzemar informació. Segons la definició del diccionari, el concepte d'informació s'associa al contigut, fent abstracció de la representació concreta que adopta. Això no obstant, en aquest context tecnològic actual, considerem la informació *digitalitzada*, i automàticament una peça d'informació s'identifica amb una cadena de bits. Els bits s'agrupen normalment en grups de vuit, anomenats *octets* o *bytes*. Per exemple, cadascun dels caràcters i símbols usats en textos del nostre entorn lingüístic es guarda en un octet usant el codi ASCII (*American Standard Code for Information Interchange*).



Els problemes que concerneixen aquesta ciència moderna de tractament d'informació digitalitzada es troben majoritàriament en tres àrees generals: matemàtiques, complexitat computacional i tecnologia informàtica.

Alguns autors consideren que dissenyar un criptosistema consisteix a triar un problema matemàtic *difícil* i usar-lo com a base per construir un sistema segur. La definició de la dificultat d'un problema matemàtic cau fora de l'àmbit de les matemàtiques, però en qualsevol cas és lògic que aquest plantejament genèric, unit a la importància adquirida per les aplicacions criptogràfiques, hagi impulsat la investigació matemàtica, i el cert és que podem trobar propostes d'utilització criptogràfica en àrees de recerca matemàtica molt diverses. Si tornem a pensar que l'objecte que manipularem a la pràctica és una cadena de bits i que el tractament serà de tipus informàtic, aviat ens trobarem situats de manera natural en una estructura aritmètica: una paraula binària de longitud k és un element del cos finit F_{2^k} i, si $k \leq \lfloor \log_2 n \rfloor$, una paraula binària de longitud k és la representació en base 2 d'un enter no negatiu i menor que n, és a dir, d'un element de Z/nZ.

El sistema criptogràfic més àmpliament comercialitzat utilitza com a problema matemàtic difícil la factorització de nombres enters. Al seu davant, i en situació emergent, tenim una família de sistemes que basen la seguretat en problemes de logaritme discret en grups algèbrics. Els més consolidats treballen amb corbes el·líptiques definides sobre cossos finits i la recerca és molt intensa per a altres casos de varietats abelianes. Així, els problemes matemàtics d'aplicació criptogràfica més immediata pertanyen principalment al camp de la geometria aritmètica, l'àrea de les matemàtiques que tracta la resolució de sistemes d'equacions algèbriques sobre estructures aritmètiques.

Pot ser que tradicionalment la matemàtica hagi mostrat una actitud molt permissiva envers el tema de la resolubilitat d'un problema, tenint molt poc en compte l'eficiència. Però en el moment de considerar les aplicacions criptogràfiques d'un problema matemàtic es fa necessari definir quantitativament els conceptes abstractes de *fàcil* i *difícil*. La teoria de la complexitat es dedica a aquesta tasca, estudiant la complexitat intrínseca dels problemes. L'objectiu és entendre l'efecte de limitar els recursos computacionals naturals (especialment el temps) en el conjunt dels problemes computacionals que es poden resoldre (usant una màquina de Turing com a model de «calculadora»). La teoria de la complexitat és una ciència molt jove que pot experimentar grans transformacions en el futur immediat. Ara com ara, el problema del mil·lenni **P** vs. **NP** és el seu gran repte.

En el terreny de la tecnologia informàtica, durant dècades s'ha considerat vàlida la *llei de Moore*, que afirma que el cost computacional es divideix per la meitat cada 18 mesos (en aquest temps el nombre de transistors incorporats en un xip es duplica). En aquest moment és possible abordar tasques criptoanalítiques que requereixin aproximadament 2^{64} operacions i es consideren segurs els sistemes per als quals no s'ha trobat cap mètode d'atac que requereixi menys de 2^{128} operacions.

Si assumim la validesa de la llei de Moore, les claus de 128 bits dels sistemes de clau secreta es mantindrien segures durant tot el segle XXI i en arribar el proper mil·lenni estaríem treballant amb claus de 730 bits. És clar que aquesta llei no pot continuar vigent indefinidament, ja que topa amb les limitacions del món físic: *com de petites poden arribar a ser les unitats de càlcul?* D'altra banda, també podria quedar invalidada per nous

models de càlcul, com ara la computació en superposició quàntica. Atès que la cerca exhaustiva sempre és una possibilitat, la criptografia ha de tenir molt present la potència computacional del moment i les perspectives de futur, i fixar en funció d'això les mides de claus previsiblement segures durant un cert període de temps.

Estandardització

Per tal que l'intercanvi d'informació xifrada funcioni adequadament en grans xarxes de comunicació resulta imprescindible fixar uns protocols criptogràfics estandarditzats. No cal dir que els estàndards de major influència arreu del món són els que fixa el govern dels Estats Units, una tasca de la qual s'encarrega el National Institute of Standards and Technology (NIST). Els estàndards relatius a les tecnologies de la informació es publiquen mitjançant *Federal Information Processing Standards Publications* (FIPS-PUB).

A continuació mostrem el funcionament d'un sistema criptogràfic bàsic de xifratge i signatura de missatges, indicant els estàndards vigents que hi fan referència. Cal remarcar que l'única part no estandarditzada és el xifratge amb sistemes de clau pública. El sistema RSA que s'indica és el sistema més comunament utilitzat.

Sistema criptogràfic
Xifratge del missatge usant un algoritme de clau secreta
FIPS-PUB 197: **AES**
Xifratge de la clau secreta usant un algoritme de clau pública (RSA)
Càlcul del *hash* (resum) del missatge FIPS-PUB 180-2: **SHA-1, SHA-224, SHA-256, SHA-384, SHA-512**
Signatura del *hash* usant un algoritme de signatura digital FIPS-PUB 186-2: **DSA, RSA, ECDSA**

El primer estàndard de la criptografia de clau secreta va ser el DES (*Data Encryption Standard*), un sistema simètric de xifratge en bloc que xifra blocs de 64 bits usant claus de 56 bits. Dissenyat per la companyia IBM, va ser adoptat l'any 1976 com a estàndard per a la transmissió i emmagatzematge d'informació no classificada. Les primeres reticències a la seva ratificació van sorgir ja l'any 1987, no pas per defectes de l'algoritme, sinó per considerar massa petita la mida de la clau. El 17 de juliol de 1998, la Electronic Frontier Foundation va presentar el seu *DES craker*, que podia trencar el DES utilitzant la cerca exhaustiva en un temps mitjà de 4.5 dies. Només sis mesos després ja es podia fer en menys de 23 hores. Finalment, l'any 2001 va ser substituït per un nou algoritme; però encara no ha desaparegut completament, ja que és encara vigent el TDEA (Triple DES), que dobla la longitud efectiva de la clau (112 bits) al preu de triplicar el nombre d'operacions de xifratge.

Un altre sistema de clau secreta força rellevant és RC5, usat per Nestcape per implementar el seu sistema de seguretat en comunicacions SSL (*Secure Socket Layer*), que ara mateix s'utilitza en gran quantitat de sistemes de comerç electrònic. L'any 1998, els laboratoris RSA van presentar el sistema RC6 a la primera fase de selecció de l'estàndard que havia de substituir el DES. D'un total de 15 candidats, 5 van passar a la segona fase i l'any 2001 l'algoritme RIJNDAEL, dissenyat per Joan Daemen i Vincent Rijmen, de la Universitat Catòlica de Leuven (Bèlgica), es converteix en el nou estàndard, ara anomenat AES (*Advanced Encryption Standard*). És un algoritme simètric de bloc de 128 bits i clau de 128, 192 o 256 bits. Les transformacions es fan sobre la *matriu d'estat*, una matriu 4×4, els coeficients de la qual són *bytes*, i les operacions es fan al cos finit F_{2^8} i a l'espai vectorial de dimensió 4 sobre aquest cos.

Tots aquests criptosistemes de clau secreta basen la seva seguretat en la *confusió i difusió* de la informació. Un objectiu important per a la recerca en criptografia simètrica seria dissenyar un algoritme de xifratge per al qual la seguretat pogués demostrar-se rigorosament en un model de computació raonable.

2.2. Criptografia de clau pública

L'any 1976, Whitfield Diffie i Martin Hellman van provocar una revolució en el camp de la criptografia en publicar «New Directions in Cryptography», [6] l'article que va introduir la idea de la criptografia de clau pública (o asimètrica), concretada en una proposta d'esquema d'intercanvi de claus.

DIFFIE-HELLMAN (1976)

Intercanvi de claus a través d'un canal insegur
(canal autenticat però no confidencial)

A i B acorden (públicament) un grup G i un element $x \in G$

A genera un nombre enter a, envia $x^a \longrightarrow$ B

B genera un nombre enter b, envia $x^b \longrightarrow$ A

A calcula $(x^b)^a$, B calcula $(x^a)^b$

CLAU = aquest element comú $x^{ab} \in G$

L'eficiència d'aquest protocol depèn únicament de l'existència d'un bon algoritme d'exponenciació a G.

$$EXP_G \in P?$$

I el secret es basa en la dificultat de resoldre el que s'anomena *problema de Diffie-Hellman*: coneguts x^a, $x^b \in G$, calcular $x^{ab} \in G$.

$$DHP_G \notin P?$$

En un sistema de clau pública cada usuari té dues claus, una pública i una privada. La pública es distribueix, no de manera secreta sinó obertament, de manera que tothom pot enviar missatges al seu propietari. La privada es manté en secret i s'utilitza per desxifrar els missatges rebuts. Així esdevé possible establir comunicació xifrada entre dues parts sense intercanvi previ de claus. Com més gran és la població que participa en el sistema de comunicació, més avantatjosa resulta aquesta propietat. A més, aquests sistemes tenen una altra característica que està adquirint gran rellevància: permeten la signatura (digital) de documents i l'autenticació de signatures. Però també s'ha de dir que els sistemes de clau pública són significativament més lents que els de clau secreta i, en conseqüència, a la pràctica s'utilitzen conjuntament amb algoritmes simètrics en sistemes híbrids.

La seguretat dels sistemes de clau pública rau en el secret de la clau privada. Tot i conèixer la fórmula matemàtica per obtenir-la a partir de la clau pública, ha d'ésser impossible fer-ho (en un temps raonable); és a dir, ens hem de trobar davant d'un problema computacionalment intractable: *no existeix* un algoritme eficient per resoldre'l. A l'espera de resultats provinents de la teoria de la complexitat, se substitueix la no-existència per la no-coneixença, en casos en què sembla raonable fer-ho.

La paradoxal funcionalitat asimètrica d'un esquema de clau pública és difícil d'aconseguir i requereix estructura matemàtica substancial. No hi ha cap principi general de disseny per trobar aquesta estructura i per això fins al moment només s'han proposat un nombre limitat de sistemes de clau pública.

3. COMPLEXITAT

A través de la criptografia de clau pública, el problema **P** *vs.* **NP** de la teoria de la complexitat es troba present a la nostra vida quotidiana. Certs problemes pertanyents a la classe **NP** per als quals hom conjectura que no es troben a la classe **P** són la base per dissenyar protocols criptogràfics que s'incorporen a caixers automàtics, telèfons mòbils, pàgines web per al comerç electrònic o els tràmits amb l'administració, etc. Pel que fa a aplicacions futures que es perfilen com a imminents, podem esmentar els sistemes de votació electrònica o el DNI digital.

Els sistemes bàsics de clau pública per als quals existeixen realitzacions comercials són el sistema RSA i els sistemes tipus ElGamal, entre els quals s'inclouen els algoritmes de signatura digital DSA i ECDSA. La seguretat del criptosistema RSA es basa en la intractabilitat del problema de factorització de nombres enters. La dels sistemes tipus ElGamal, en la intractabilitat del problema del logaritme discret. Naturalment, no tenim proves que cap d'aquests problemes sigui en realitat intractable. I potser tampoc tenim moltes esperances de veure aquestes proves en un futur immediat, perquè per alguna raó l'anomenem *problema del mil·lenni*. Per això, per tal de confiar en la seguretat que ofereixen aquests sistemes criptogràfics, cal confiar en el treball i l'experiència col·lectiva de matemàtics i informàtics que han dedicat molts esforços a intentar trobar algoritmes eficients per resoldre aquests problemes.

3.1. Factorització

Atès que el problema de la factorització de nombres enters s'ha investigat des de fa més de 2.000 anys, podem considerar que hi ha evidència històrica de la seva dificultat. De tota manera, no és fins a l'aparició del sistema RSA que trobem una gran motivació per dur a terme una intensa i extensa recerca al voltant d'aquest problema, que tant pot dirigir-se a la cerca de millors algoritmes per resoldre'l com a la cerca d'una demostració rigorosa de la seva dificultat intrínseca.

El criptosistema RSA
(Rivest-Shamir-Adleman, 1978)

GENERACIÓ DE CLAUS

Cada usuari tria dos nombres primers p i q

— calcula $n = pq$
— calcula $\varphi(n) = (p-1)(q-1)$ i tria e tal que $\text{mcd}(e,\varphi(n)) = 1$
— calcula $d = e^{-1} \bmod \varphi(n)$

dóna a conèixer la clau *pública* $\{n, e\}$ i guarda la clau *privada* d

EXEMPLE: CLAUS RSA AMB MÒDUL n DE 1024 BITS

p = 84997172489332578490384442745470933566648021786255802843114129
106338875770084181315383045268895951414271920538102125284234191
352753916481505630202163787 73

→

q = 124987075137699682010130344784719887401171766209184429254814007267157805668200318061133138657163417406550755050927232487278984454650243654196318594606447674

n = 106235479844162313112623622376706062835748388959562335916243821142638516264459989006286900617903692802910893730009874328562357789700487670277809790099753313862825239708084105047965267520607476005127653021190353575336378429508937242498528860769971373863559801544240476211228445080625179919835128519096472330891

e = 65537

d = 175943344710398362226592123117745335614876026331246409514459074214901270356355756393218795383787277672581113042331381656501797679083432743823389453551844338427178964159288543712996418572646972868464784112970105225369155085004191787247418860205873925732969157121794069965324678045086645158380639036124493990 57

Vegem primer com s'utilitzen aquestes claus per xifrar/desxifrar i signar/autenticar. Després ens ocuparem d'analitzar l'eficiència i la seguretat computacional d'aquest sistema.

Si hom vol enviar un missatge a l'usuari que té clau pública (n, e), el primer que cal fer és dividir el missatge (cadena de bits) en blocs de longitud $\lfloor \log_2 n \rfloor$, que representen enters menors que n, és a dir, elements de $\mathbf{Z}/n\mathbf{Z}$. Per a cadascun d'aquests blocs $m \in \mathbf{Z}/n\mathbf{Z}$, el criptograma corresponent és

$$c = m^e \bmod n.$$

El teorema d'Euler, que afirma que si $(x, n) = 1$, aleshores $x^{\varphi(n)} \equiv 1 \bmod n$, garanteix que un cop el propietari de les claus (n, e, d) ha rebut el criptograma c, només ha de fer una altra exponenciació per recuperar el bloc de missatge:

$$m = c^d \bmod n$$

El sistema RSA de signatura digital, que utilitzen per exemple l'Agència Catalana de Certificació (www.catcert.net) o la Fábrica Nacional de Moneda y Timbre (www.fnmt.es), té el mateix fonament teòric que el sistema de xifratge: si l'usuari A vol signar un missatge m destinat a l'usuari B, el que farà serà calcular $h(m)$, un resum (*hash*) del missatge (per exemple, amb l'algoritme SHA-256 obtindria un resum de 256 bits) i llavors, primer calcularà la seva rúbrica

$$r = h(m)^{d_A} \mod n_A$$

i després la signatura digital destinada a B

$$s = r^{e_B} \mod n_B.$$

Aquest usuari B autentica la signatura del missatge m per part de A fent els dos passos inversos

$$s^{d_B} \mod n_B = r$$
$$r^{e_A} \mod n_A = h(m)$$

i comparant el resultat obtingut amb el resultat d'aplicar el mateix algoritme de resum al missatge rebut prèviament.

Dues són les qüestions elementals que s'han de plantejar davant de qualsevol proposta de sistema criptogràfic:

1. Totes les operacions involucrades en el seu funcionament es poden dur a terme mitjançant algoritmes eficients?
2. En què es fonamenta la seguretat?

En la generació de claus RSA, les dues operacions aritmètiques que es duen a terme són un càlcul de màxim comú divisor i un càlcul d'invers modular. Ambdós es poden fer eficientment mitjançant l'algoritme d'Euclides estès, de complexitat quadràtica en la longitud del mòdul n.

Algoritme d'Euclides estès: $O(\ell(n)^2)$

mcd $(\varphi(n),e) = 1$	$\varphi(n)\,x + ed = 1$	
mcd (19872, 343)	1	0
mcd (343, 19872 mod 343)	0	1
mcd (321, 343 mod 321)	1	–57
mcd (22, 321 mod 22)	–1	58
mcd (13, 22 mod 13)	15	–869
mcd (9, 13 mod 9)	–16	927
mcd (4, 9 mod 4)	31	–1796
mcd (1, 4 mod 1) = mcd (1,0) = **1**	$x = -78$	$\boldsymbol{d = 4519}$

L'operació RSA, la que cal fer cada cop que es vol xifrar o desxifrar, és l'exponenciació modular d'enters $< n$. Un algoritme que permet fer-la de manera eficient és l'algoritme de quadrats successius, de complexitat cúbica.

Algoritme de quadrats successius: $O(\ell(n)^3)$

$23 = 10111 = 2\,(2\,(2\,(2\,(0 + 1) + 0) + 1) + 1) + 1$

$a \in Z/nZ$

$a^{23} = ((((1 * a)^2 * 1)^2 * a)^2 * a)^2 * a$

Prenent $e = 65537 = 2^{16} + 1 = 1000000000000001$, el primer de Fermat que utilitzen moltes implementacions del sistema RSA com a exponent públic comú per a tots els usuaris, el xifratge de missatges es pot fer amb només setze quadrats i un producte a $\mathbf{Z}/n\mathbf{Z}$: $m^{65537} = ((((((((((((((m)^2)^2)^2)^2)^2)^2)^2)^2)^2)^2)^2)^2)^2)^2)^2 * m$.

Hem deixat per al final la primera de les funcions necessàries per posar en marxa un sistema RSA: la generació de nombres primers de longitud fixada ℓ. Atès que la densitat dels nombres primers és aproximadament $1 / ln(n)$, si disposem d'un algoritme eficient per resoldre el problema decisional de la primeritat d'un nombre enter, aleshores un mètode eficient consisteix a començar per un nombre senar aleatori de longitud ℓ i anar fent execucions successives de l'algoritme fins a obtenir la resposta *cert*.

El 8 d'agost de 2002, M. Agrawal, N. Kayal i N. Saxena van anunciar el seu treball «Primes is in P» [2], on presenten un algoritme polinòmic determinista que decideix si un nombre enter és primer o compost. L'algoritme està basat en el resultat següent: si $a \in \mathbf{Z}$, $n \in \mathbf{N}$, $n \geq 2$ i $\mathrm{mcd}(a, n) = 1$, aleshores n és primer si, i només si, $(X + a)^n \equiv X^n + a \bmod n$. La idea és avaluar els dos costats de la congruència mòdul un polinomi de la forma $X^r - 1$, amb un r petit convenientment triat. La complexitat d'aquest algoritme és $O^{\sim}(\ell(n)^{15/2})$.

Tot i haver col·locat definitivament el problema de la primeritat a la classe P, l'algoritme proposat encara té una complexitat massa elevada per a la demanda criptogràfica d'eficiència. A l'espera que aquesta complexitat disminueixi significativament, els protocols criptogràfics que requereixen la generació de nombres primers segueixen usant majoritàriament algoritmes probabilístics. El més comunament utilitzat és el *test de Miller-Rabin*, un algoritme de Monte-Carlo basat en el petit teorema de Fermat: si n és primer i $\mathrm{mcd}(p, b) = 1$, aleshores $b^{p-1} \equiv \bmod p$, i en el fet que, si p és primer, aleshores $X^2 = 1 \bmod p$ té exactament dues solucions: 1 i $p - 1$.

Pel que fa a la seguretat del sistema RSA, la qüestió principal és l'obtenció de la clau privada d a partir de les dades públiques n, e. És clar que el càlcul és possible si el mòdul n es pot factoritzar. Atesa la creença general en la dificultat del problema de la factorització, el que es voldria és provar l'equivalència computacional d'ambdós problemes.

Test de Miller Rabin
per a un enter senar N **i una base** $b \in (Z/NZ)^*$
$O(\ell(N)^3)$

$N - 1 = 2^t N_0$

$x_0 = b^{N_0} \mod N \qquad x_{i+1} = x_i^2 \mod N$

$x_0 = 1$ o bé $x_i = N - 1$ per a algun $i \in \{0,1, \ldots, t-1\}$

$\Downarrow$

N primer (Probabilitat d'error < 1/4)

D'una banda, és fàcil veure que el càlcul de $\varphi(n)$ sí té la mateixa dificultat computacional que la factorització de n, atès que coneguts n i $\varphi(n)$ és suficient resoldre l'equació

$$x^2 + (\varphi(n) - n - 1) \quad x + n = 0$$

per trobar els factors p i q de n. De l'altra, si es coneix l'exponent secret d, usant que $ed - 1$ és múltiple de $\varphi(n)$ i que l'1 té quatre arrels quadrades mòdul n, el càlcul de $\mathrm{mcd}(x^{(ed-1)/2} - 1, n)$ per a valors aleatoris de $x \in (Z/nZ)^*$ és un algoritme probabilístic de factorització de n. Finalment, cal esmentar que Coron i May (IACR ePrint 2004/208) han presentat recentment el primer algoritme determinista polinòmic de factorització de n, amb la qual cosa quedaria provada l'equivalència computacional abans esmentada.

Un altre objectiu criptoanalític és el d'obtenir el missatge a partir del criptograma. Atès que $c = m^e \mod n$, el problema que es presenta és un càlcul d'arrels e-èsimes mòdul n. Aquest problema també es conjectura computacionalment equivalent a la factorització d'enters, però encara no es coneix cap demostració d'això.

De fet, Boneh i Venkatesan [3] van demostrar que si s'utilitza un exponent petit (com, per exemple, $e = 3$), aleshores el problema d'extracció d'arrels no pot ser equivalent al de la factorització; és més fàcil. Per això, cal utilitzar exponents com ara 65537 o més grans.

Com hem vist, la seguretat del sistema criptogràfic RSA està completament basada en la hipòtesi que la factorització d'enters és un problema computacional que no pertany a la classe **P**. Des que es va inventar el sistema RSA, s'han produït dos avenços significatius en el terreny dels algoritmes de factorització de nombres enters: la descoberta de l'*algoritme de garbell quadràtic* (QS, Quadratic Sieve) l'any 1982 i la descoberta del l'*algoritme del garbell del cos de nombres* (NFS, Number Field Sieve) l'any 1990. Tot i ser conegut que el garbell del cos de nombres era més ràpid que el garbell quadràtic, no

va ser fins al 1996 que va ser acceptat com a mètode superior a la pràctica i es va convertir en el campió dels algoritmes de factorització.

Evolució de la factorització

Nombre	*Bits*	*Data*	*Algoritme*
$10^{71}-1$	236	1984	QS
RSA-100	330	Abril 1991	QS
RSA-110	364	Abril 1992	QS
RSA-120	397	Juny 1993	QS
RSA-129	425	Abril 1994	QS
RSA-130	430	Abril 1996	NFS
RSA-140	463	Febrer 1999	NFS
RSA-155	512	Agost 1999	NFS
RSA-576	576	Desembre 2003	NFS
$11^{281}+1$	582	Maig 2005	NFS

L'algoritme NFS té dues parts principals: l'etapa de garbell i l'etapa de càlcul matricial. La primera etapa, on es fa la major part de la feina, pot paral·lelitzar-se de manera eficient per tal d'involucrar una xarxa d'estacions de treball. Al contrari, les tècniques existents per a l'etapa de càlcul matricial sembla que només funcionen correctament quan es treballa en un únic ordinador. L'objectiu final de tot plegat és buscar una congruència del tipus $x^2 \equiv y^2 \bmod n$ de manera que el càlcul de mcd $(x-y, n)$ proporcioni un factor no trivial de n. La seva complexitat esperada és subexponencial

$$\mathcal{O}(e^{(1.923+o(1))(\log n)^{1/3}(\log \log n)^{2/3}}).$$

En la factorització del RSA-155 (un nombre de 512 bits) van participar científics de sis països i es van utilitzar unes 300 estacions de treball per a l'etapa de garbell, que es va completar en tres mesos i mig. Després, en 40 dies es va realitzar l'etapa de càlcul matricial. El dia 2 de desembre de 2003 J. Franke, de la Universitat de Bonn, va fer circular per correu electrònic l'anunci de la factorització del primer dels nombres de la llista de reptes plantejats per la companyia RSA Security. Ni en el seu anunci ni posteriorment s'han donat detalls exactes del procés de factorització, però el que sí quedava clar és que l'equip humà i el nombre d'ordinadors involucrats era molt més reduït que en el cas previ. En la mateixa línia, aquest equip acaba de fer pública la factorització d'un nombre de 663 bits.

Factorització del RSA-200 (Maig 2005)

RSA-200 =
27997833911221327870829467638722601621070446786955428537560009929326128400107609345671052955360856061822351910951365788637105954482006576775098580557613579098734950144178863178946295187237869221823983

p =
35324619344027701212726049781984643686711974001976250236493034687761212536794232000585474956528088349

q =
79258699544783330333470858414800596877379758573642199607343303414557678728181521353814093047401854675

Mida en bits dels mòduls RSA per xifrar claus AES

Any	*AES-128*	*AES-192*	*AES-256*
2010	2942	7426	14645
2020	3296	8042	15574
2030	3675	8689	16538

De fet, J. Franke *et al* ja han escrit un treball titulat «SHARK: A Realizable Special Hardware Sieving Device for Factoring 1024-bit Integers», i tot sembla indicar que els mòduls RSA de 1.024 bits no romandran segurs durant gaire temps. Si això passa, caldrà revisar les estimacions de Lenstra i Verheul [12] segons les quals l'any 2002 un RSA de 1024 bits podia considerar-se molt segur i el mateix nivell de seguretat es tindria l'any 2023 amb un RSA de 2.048 bits.

En un altre estudi, Lenstra [11] també estima les mides de les claus RSA necessàries per protegir les claus dels sistemes simètrics amb el mateix nivell de seguretat que si aquestes claus no es transmetessin per cap canal insegur. Per exemple, si l'any 2010 volem que el treball computacional necessari per factoritzar un mòdul RSA sigui equivalent a la cerca exhaustiva a l'espai de claus de 128 bits, caldria que aquest mòdul fos de 2.942 bits.

La incògnita no consisteix només a esbrinar com d'acurades resulten finalment aquestes prediccions, sinó que també ens hem de plantejar si en un futur més o menys immediat simplement s'anirà augmentant la mida de les claus en funció dels avenços algorítmics i RSA continuarà essent el sistema criptogràfic de clau pública amb una major implantació o bé assistirem al seu relleu.

3.2. Logaritme discret

Malgrat que el protocol de Diffie i Hellman que va donar origen a la criptografia de clau pública estava basat en la dificultat del problema del logaritme discret, van haver de passar gairebé deu anys fins que l'any 1985 es va publicar el treball de T. ElGamal [7] que va establir l'interès criptogràfic del problema i va impulsar l'estudi de la seva complexitat. Ara mateix, el problema del logaritme discret en un grup finit és el problema matemàtic més útil per basar la criptografia de clau pública.

Ja el primer estàndard de signatura digital, anunciat pel National Institute of Standards and Technology l'any 1991 i publicat el 1994 en el FIPS PUB 186, especificava com a algoritme de signatura digital (DSA) una modificació de l'esquema de signatura d'ElGamal.

DSA (Digital Signature Algorithm)

Paràmetres públics i comuns per a un grup d'usuaris

Un primer p de 1.024 bits
Un primer q de 160 bits, divisor de $p-1$
$g = x^{(p-1)/q} \bmod p$, generador del subgrup d'ordre q a F_p^*

Claus privada i pública d'un usuari

r un enter mòdul q aleatori o pseudoaleatori
$u = g^r \bmod p$

Paràmetre (secret) que cal generar per a cada signatura

r un enter mòdul q aleatori o pseudoaleatori

Signatura d'un missatge m

$f_1 = (g^k \bmod p) \bmod q$
$f_2 = k^{-1} (SHA(m) + f_1, r) \bmod q$
La signatura que acompanya el missatge és (f_1, f_2) (320 bits)

Verificació d'una signatura (f_1, f_2)

$w = (f_2)^{-1} \bmod q$
$w_1 = SHA(m)\ w \bmod q$
$w_2 = f_1\ w \bmod q$
$v = (g^{w_1} u^{w_2} \bmod p) \bmod q$
Acceptar si $v = f_1$

Per al bon funcionament d'aquest esquema de signatura digital cal disposar d'algoritmes eficients per al càlcul d'inversos modulars i per dur a terme l'exponenciació al grup $\mathbf{F}_p^*$. Com ja hem comentat abans, l'algoritme d'Euclides estès i l'algoritme de quadrats successius donen resposta satisfactòria a aquesta demanda. Si analitzem la seguretat, novament hem de considerar la qüestió d'obtenir la clau privada r a partir de la pública $u = g^r \bmod p$, i llavors estarem plantejant un problema de logaritme discret.

Sigui $G = \langle g \rangle$ un grup cíclic d'ordre n. L'exponenciació

$$\begin{array}{rcl} \exp_g: \mathbf{Z}/n\mathbf{Z} & \longrightarrow & G \\ a & \mapsto & g^a \end{array}$$

és un isomorfisme de grups. El *problema del logaritme discret* consisteix a fer explícit l'isomorfisme invers

$$\log_g: \mathrm{G} \longrightarrow \mathbf{Z}/n\mathbf{Z},$$

és a dir: donats $G = \langle g \rangle$ d'ordre n i $b \in G$, trobar $a \in \mathbf{Z}/n\mathbf{Z}$ tal que $b = g^a$.

Els sistemes criptogràfics basats en aquest problema funcionen sota la hipòtesi que es tracta d'un problema que no pertany a la classe de complexitat **P**. Però a diferència del problema de la factorització de nombres enters, aquí tenim una multitud de problemes computacionals, atès que hi ha dependència del grup G. La situació precisa és que no es coneix cap algoritme de complexitat polinòmica per resoldre'l que sigui general, És a dir, que no faci ús de cap propietat específica del grup.

Els primers sistemes de xifratge i signatura basats en el problema del logaritme discret, com DSA, treballaven en el grup multiplicatiu d'un cos finit $\mathbf{F}_p$. Malauradament, en aquest cas sembla haver-hi un lligam entre el problema del logaritme discret i la factorització d'enters pel que fa a la complexitat. Es tracta d'un lligam no establert per cap resultat teòric, però el cert és que s'han anat produint avenços algorítmics simultanis: molts algoritmes desenvolupats per a un dels problemes s'han pogut modificar per aplicar-los a l'altre.

Pensant en la seguretat, seria desitjable tenir molta més diversitat. És lògic pensar que en la tria de DSA com a estàndard de signatura hi pot haver influït una qüestió de patents de RSA, però també que hi ha un component de prudència a no basar-ho tot en un únic sistema i intentar mitigar les conseqüències d'un possible trencament. Malgrat aquesta voluntat de diversificar i de molts intents no reeixits d'utilitzar criptogràficament diferents problemes matemàtics, gairebé tres dècades després de l'aparició de la criptografia de clau pública tots els sistemes en què es confia estan basats o bé en el problema de la factorització d'enters o bé en el del logaritme discret en un grup finit.

Per a la majoria de grups cíclics $G = \langle g \rangle$, hi ha una manera eficient d'obtenir una representació canònica única dels elements. En aquest cas, hi ha diversos algoritmes per calcular el logaritme discret amb $|G|^{1/2}$ operacions aproximadament, entre els quals destaquen l'algoritme determinista *Baby-step*, *Giant-step* de Shanks i l'algoritme probabi-

lístic *Rho* de Pollard. El fet que durant dècades no s'hagi produït cap progrés substancial en aquest terreny, ja que s'ha avançat en la paral·lelització dels algoritmes però no s'ha millorat essencialment la complexitat, és un fet considerat important que ha portat a conjecturar que en absència d'altra estructura al grup cíclic G d'ordre primer p, el càlcul del logaritme discret a G té complexitat exponencial $\mathcal{O}(\sqrt{p})$.

Però els sistemes criptogràfics no es plantegen en un grup abstracte, sinó en un de ben concret. Per exemple, una versió d'un algoritme de complexitat exponencial que permeti la computació en paral·lel és rellevant a la pràctica per a tots aquells grups per als quals no es disposi d'un algoritme de complexitat subexponencial.

En els grups on es té d'alguna manera una noció de factorització en irreductibles és possible utilitzar l'algoritme *Index-Calculus*. Diversos matemàtics van descobrir aquest mètode, però sembla que la primera referència correspon a Kraitchik, que va escriure sobre ell a les seves «Recherches sur la theorie des nombres» publicades l'any 1924. L'algoritme té clar parentiu amb els algoritmes QS i NFS per a la factorització de nombres enters, amb una etapa de garbell i una etapa d'àlgebra lineal, que ara no és binària sinó mòdul el cardinal del grup G. Aquestes etapes precedeixen el càlcul del logaritme de l'element en concret.

Algoritme *Index-Calculus*

1. Triar una base de factors $B = \{p_1, p_2, \ldots, p_m\} \subset \mathrm{G}$
2. (Garbell) Produir relacions $g^{k_i} = \prod p_j^{e_{ij}}$
   fins que rang (e_{ij}) = rang $(e_{ij} \mid k_i) = m$
3. (Àlgebra lineal) Amb $A = (e_{ij})$ i $v = (k_i)$, resoldre $AX \equiv v \bmod |\mathrm{G}|$
   Solució: $X = (\log p_1, \log p_2, \ldots, \log p_m)$
4. Obtenir una relació $bg^K = p_1^{e_1} p_2^{e_2} \ldots p_m^{e_m}$
   $$\log b = -K + \sum e_i \log p_i$$

En analitzar aquest algoritme, ni que sigui superficialment, veiem que la mida de la base de factors s'ha d'optimitzar per equilibrar convenientment el temps per trobar relacions i el temps per resoldre el sistema lineal. I en el pas 2 cal considerar la probabilitat de trobar una d'aquestes relacions en el grup G. Si prenem com a grup $\mathbf{F}_p^*$, el grup multiplicatiu del cos finit de cardinal primer p, els elements es representen de manera canònica com a enters menors que p. Llavors, la noció de factorització al grup és naturalment la factorització de nombres enters i la base de factors estarà formada pels nombres primers (de $\mathbf{Z}$) menors que un certa fita. Per a aquest grup, l'algoritme *Index-Calculus* té complexitat subexponencial $\mathcal{O}(e^{c(\log p)^{1/2}(\log\log p)^{1/2}})$.

Rècord de Joux i Lercier a $\mathbf{F}_p^*$ (2001)

$p = \lfloor 10^{119}\pi \rfloor + 207819$
$= 31415926535897932384626433832795028841971693993751058209749445923078164062862089986280348253421170679821480865132843 8483$

$g = 2$

$b = \lfloor 10^{119} e \rfloor$
$= 27182818284590452353602874713526624977572470936999595749669676277240766303535475945713821785251664274274663919320030599 2$

$b = g^{\wedge} 26211228068581138763600862203819182737039076852065697424303538038219347876743601868144980494084037374164145286473076508 2$

Si es vol treballar al grup multiplicatiu d'un cos finit de cardinal $q = p^n$, llavors es pot fixar un isomorfisme $\mathbf{F}_q \simeq \mathbf{F}_p[X]/(f)$, amb $f \in \mathbf{F}_p[X]$ polinomi irreductible de grau n, amb la qual cosa els elements del grup $\mathbf{F}_q^*$ es poden representar com a polinomis a coeficients en $\mathbf{F}_p$ de grau menor que n. Atès que $\mathbf{F}_p[X]$ és un anell factorial, es té també una noció natural de factorització i s'obtenen algoritmes de complexitat subexponencial treballant amb bases de factors formades per polinomis mònics irreductibles de grau menor que una certa fita.

L'algoritme *Index-Calculus* per calcular el logaritme discret en un cos finit $\mathbf{F}_q$ admet una formulació més general que posa de manifest el seu lligam amb l'algoritme *Number Field Sieve* per a la factorització de nombres enters. Es consideren morfismes d'anells $\phi : R \to \mathbf{F}_q$, on R és un domini de Dedekind, tals que g i b pertanyen a la imatge. Les bases de factors estan formades per ideals primers de R, i en l'etapa de garbell cal buscar elements de R tals que l'ideal (principal) que generen factoritzi sobre la base de factors. Pel fet de treballar amb ideals, apareix una obstrucció, donada pel grup V/K^{*q-1}, on K és el cos de fraccions de R, K^{*q-1} és el grup de les potències $(q-1)$-èsimes de K^* i V està format pels elements $\alpha \in K^*$ tals que, per a tots els ideals primers $\wp$ de R, es té que $\mathrm{ord}_\wp(\alpha)$ és divisible per $q-1$. Amb tries adients de l'anell R i controlant l'obstrucció, s'obtenen algoritmes de complexitat subexponencial amb exponent 1/3, la mateixa complexitat que el millor algoritme de factorització, per al logaritme discret al grup $\mathbf{F}_q^*$, però amb condicions sobre $q = p^n$, que essencialment són $n < (\log p)^{1/2}$ o $n > (\log p)^2$. Arribar a aquesta complexitat per als valors intermedis és un problema obert, com també ho és l'existència d'algoritmes per al logaritme discret de complexitat subexponencial, però no en la mida del grup sinó en la del més gran primer que divi-

deixi el cardinal del grup. Aquests serien la mena d'algoritmes que podrien trencar l'estàndard DSA de signatura digital.

4. GEOMETRIA

Excepte pel petit escull de la presència d'una suma en el càlcul de f_2, l'algoritme de signatura digital DSA pot formular-se en qualsevol grup, i això obre les portes al fet que la matemàtica vagi oferint diferents grups a la criptografia. Fins ara els grups que s'han proposat per a usos criptogràfics inclouen grups multiplicatius de cossos finits de característica 2, subgrups del grup multiplicatiu de l'anell d'enters mòdul un primer, grups d'unitats de l'anell d'enters mòdul un compost, grups de punts de corbes el·líptiques definides sobre cossos finits, jacobianes de corbes hiperel·líptiques definides sobre cossos finits i grups de classes de cossos quadràtics imaginaris.

A priori, per proposar un grup només cal que l'operació es pugui implementar de manera eficient i que el logaritme discret sigui considerat intractable. Com ja hem dit, aquesta consideració pot dependre de certes propietats del grup concret. En tot cas, hi ha una àmplia família de grups que queda eliminada per l'algoritme de Pohlig-Hellman: aquells que tenen un cardinal que factoritza com a producte de primers *petits*.

Algoritme de Pohlig-Hellman
$G = \langle g \rangle$ **d'ordre** p^v

Entrada: $b \in G$
Sortida: $a \bmod p^v$ tal que $b = g^a$

Calcular $\zeta = g^{p^{v-1}}$ (arrel p-èsima de la unitat)
Calcular $\zeta^2, \zeta^3, \ldots, \zeta^{p-1}$

El logaritme en base g es redueix al logaritme en base ζ:
$b^{p^{v-1}} = g^{p^{v-1}a} = \zeta^{a_0}$, on a_0 és el primer p-dígit de a
$(bg^{-a_0})^{p^{v-2}} = \zeta^{a_0}$, on a_1 és el segon p-dígit de a
Successivament es determinen els v dígits de a en base p

Si podem emmagatzemar la taula de potències de ζ i accedir-hi per fer cerques, aleshores el logaritme discret a un grup cíclic g de cardinal p^v es pot resoldre amb $\mathcal{O}(v^2) = \mathcal{O}((\log |G|)^2)$ operacions de grup.

Si $G = \langle g \rangle$ és un grup de cardinal $n = \Pi p_i^{v_i}$, aleshores el grup $G_i = \langle g^{n/p_i} \rangle$ té cardinal $p_i^{v_i}$ i l'element $b_i = b^{n/p_i}$ pertany a G_i. Les resolucions dels respectius problemes de logaritme discret i del sistema de congruències $a \equiv a_i \bmod p_i^{v_i}$ dóna lloc a un algoritme eficient per al logaritme discret. Però, com ja hem observat, això només és factible si els primers p_i són prou petits com per permetre guardar les taules d'arrels p_i-èsimes de la unitat. Així, la conclusió a què arribem és que són grups d'utilitat criptogràfica aquells grups finits que tenen una operació implementable de manera eficient i un cardinal divisible per algun primer *gran*.

4.1. Corbes el·líptiques

L'any 1985, Koblitz [10] i Miller [14] van proposar independentment utilitzar criptogràficament el grup de punts d'una corba el·líptica definida sobre un cos finit. El principal avantatge sobre els sistemes que utilitzen el grup multiplitiu d'un cos finit (o la factorització d'enters) és l'absència d'algoritmes de complexitat subexponencial.

Una corba el·líptica definida sobre un cos K és una cúbica no singular donada per una *equació de Weierstrass*

$$Y^2 + a_1 XY + a_3 Y = X^3 + a_2 X^2 + a_4 X + a_6 \qquad a_i \in K,$$

juntament amb un punt especial, anomenat *punt de l'infinit*, que podem escriure en coordenades projectives $O = (0 : 1 : 0)$.

Normalment, les corbes el·líptiques que s'utilitzen en criptografia són corbes definides sobre $K = \mathbf{F}_{2^n}$ o bé sobre $K = \mathbf{F}_p$ amb p primer.

Llei de grup

$$(x_1, y_1) + (x_2, y_2) = (x_3, y_3)$$

$$\mu = y_1 - \lambda x_1$$

$$\lambda = \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1} & x_1 \neq x_2 \\ \dfrac{3x_1^2 + 2a_2 x_1 + a_4 - a_1 y_1}{2y_1 + a_1 x_1 + a_3} & x_1 = x_2 \end{cases}$$

$$x_3 = \lambda^2 + a_1 \lambda - a_2 - x_1 - x_2$$
$$y_3 = -(\lambda + a_1) x_3 - \mu - a_3$$

El mètode geomètric de la *corda-tangent* —tres punts sumen O si, i només si, estan alineats— expressat algebraicament proporciona al conjunt de punts $E(\overline{K})$ una estructura de grup abelià, amb el punt O com a element neutre.

Podem observar que si L és un cos tal que $K \subseteq L \subseteq \overline{K}$, aleshores $E(L)$ és un subgrup de $E(\overline{K})$. I si l'aritmètica del cos L es pot implementar de manera eficient, aleshores l'operació del grup $E(L)$ també. Amb tot això veiem que si E és una corba el·líptica definida sobre un cos finit $\mathbf{F}_q$, llavors $E(\mathbf{F}_q)$ és un grup abelià finit amb *operació eficient*. Quan es tria un cos $\mathbf{F}_p$ l'aritmètica base és l'aritmètica dels enters mòdul p i quan es tria un cos $\mathbf{F}_{2^n}$, tot i que es pot considerar aritmètica modular de polinomis de $\mathbf{F}_2[X]$, és molt més habitual treballar amb vectors binaris i usar bases normals per implementar el producte.

L'altra qüestió que cal prendre en consideració abans d'utilitzar criptogràficament un grup $E(\mathbf{F}_q)$ és la mida dels primers que divideixen el seu cardinal. La primera informació de què disposem és que el cardinal es troba dins de l'anomenat *interval de Hasse*:

$$|E(F_q)| \in [q + 1 - 2\sqrt{q}\,,\, q + 1 + 2\sqrt{q}\,].$$

Per tant, sabem que la mida del grup és de l'ordre de la mida del cos finit $\mathbf{F}_q$. Però necessitem conèixer exactament el cardinal i la seva factorització en primers. De fet, un cop calculat $|E(\mathbf{F}_q)|$ s'usa un test de primalitat per a aquest cardinal, o pot ser dividit per algun nombre petit, ja que només interessen les corbes el·líptiques de cardinal quasi-primer.

Si prenem una corba E definida sobre $\mathbf{F}_p$ per a la qual sabem calcular el cardinal, llavors podem considerar també que la corba està definida sobre qualsevol extensió $\mathbf{F}_{p^m}$ i disposem d'una fórmula per al cardinal: $|E(\mathbf{F}_{p^m})| = 1 + p^m - \alpha^m - \overline{\alpha}^m$, on α, $\overline{\alpha}$ són les arrels complexes del polinomi $T^2 - a_p T + p$ i $a_p = p + 1 - |E(\mathbf{F}_p)|$.

Corba K-163 del FIPS 186-2

$$E/\mathbf{F}_{2^{163}} \quad Y^2 + XY = X^3 + X^2 + 1$$

$$E(F_2) = \{O, (0,1)\} \Rightarrow a_2 = 2 + 1 - 2 = 1$$

$$T^2 - T + 2 = (T - \alpha)(T - \overline{\alpha}) \Rightarrow \alpha = \tfrac{1 + i\sqrt{7}}{2}$$

$$t = \alpha^{163} + \overline{\alpha}^{163} = -4835703278458516698824704$$

$$|E(\mathbf{F}_{2^{163}})| = 1 + 2^{163} - t$$

$$= 2 \cdot 5846006549323611672814741753598448348329118574063$$

Ara bé, si volem usar corbes el·líptiques definides sobre un cos $\mathbf{F}_p$, amb p un primer gran, o definides sobre un cos $\mathbf{F}_q$ però no sobre cap subcòs, aleshores no tenim cap fórmula senzilla per calcular el cardinal del grup. Com es fa, doncs, per trobar un exemple com aquest?

Corba P-192 del FIPS 186-2

$E/\mathbf{F}_p \quad Y^2 = X^3 - 3X^2 + B$

$p = 2^{192} - 2^{64} - 1$
$= 6277101735386680763835789423207666416083908700390324961279$

$B = 64210519e59c80e70fa7e9ab72243049feb8deecc146b9b1$ (hexadecimal)

$|E(\mathbf{F}_p)|$ = primer
$= 6277101735386680763835789423176059013767194773182842284081$

L'any 1985, Schoof [16] va presentar un algoritme polinòmic per calcular el nombre de punts $\mathbf{F}_q$-racionals d'una corba el·líptica definida sobre el cos finit $\mathbf{F}_q$, amb total generalitat. L'algoritme calcula el valor $a_q = q + 1 - |E(\mathbf{F}_q)|$ mòdul suficients primers ℓ, usant que $T^2 - a_q T + q$ mod ℓ és el polinomi característic de l'endomorfisme de Frobenius operant al mòdul de ℓ-torsió $E[\ell]$, de manera que el càlcul es fa amb aritmètica de polinomis a $\mathbf{F}_q[X]$ mòdul el polinomi de ℓ-divisió, que té grau $(\ell^2-1)/2$. La seva complexitat $\mathcal{O}((\log q)^8)$ el fa bastant difícil d'usar a la pràctica per als valors que interessen a la criptografia, que són $q > 2^{160}$. Modificacions posteriors que incorporen idees d'Elkies i Atkin han donat lloc a l'algoritme SEA, de complexitat $\mathcal{O}((\log q)^{4+\varepsilon})$. Disposant d'un algoritme general com aquest que funcioni eficientment, es pot fixar el cos finit en què es vol treballar i anar seleccionant corbes el·líptiques (coeficients d'equacions de Weierstraß) de forma aleatòria fins que el cardinal satisfaci la condició de quasi-primalitat desitjada.

Paral·lelament han sorgit mètodes alternatius de selecció de corbes el·líptiques aptes per a la criptografia. El *mètode de la multiplicació complexa* permet triar el cardinal de la corba abans de construir la corba explícitament. Per tant, s'analitza l'interval de Hasse a la recerca d'un cardinal adient i després es busca l'equació de la corba. Aquest mètode és eficient sempre que les mides q del cos finit i $q + 1 - a_q$ del grup de punts de la corba siguin tals que el cos quadràtic imaginari $Q(\sqrt{a_q^2 - 4q})$ tingui nombre de classes petit. Per acabar, donada una corba sobre un cos finit (de característica 2 o 3), els *mètodes p-àdics* utilitzen l'aixecament canònic, una corba sobre un cos local de característica zero amb anell d'endomorfismes isomorf al de la corba donada. Els coeficients de l'equació de Weierstraß d'aquest aixecament tenen informació suficient per calcular la traça del Frobenius, que determina el cardinal.

El cas és que la criptografia amb corbes el·líptiques es troba ja en una fase de plena incorporació a la tecnologia. El FIPS PUB 186-2, publicat el 27 de gener de 2000, inclou com a estàndard de signatura digital el *Elliptic Curve Digital Signature Algorithm* (ECDSA), l'anàleg amb corbes el·líptiques de l'algoritme DSA.

ECDSA (Elliptic Curve Digital Signature Algorithm)

Paràmetres públics i comuns per a un grup d'usuaris

Un cos finit $\mathbb{F} = \mathbf{F}_p$ o $\mathbf{F}_{2^n}$
Una corba el·líptica E definida sobre $\mathbb{F}$
Un punt $P \in E(\mathbb{F})$ tal que $|\langle P \rangle| = q$ primer

Claus privada i pública d'un usuari

r un enter mòdul q aleatori o pseudoaleatori
$Q = rP$

Paràmetre (secret) que cal generar per a cada signatura

k un enter mòdul q aleatori o pseudoaleatori

Signatura d'un missatge m

$kP = (x_1, y_1)$
$f_1 = x_1 \bmod q$
$f_2 = k^{-1} (SHA\ (m) + f_1\ r) \bmod q$
La signatura que acompanya el missatge és (f_1, f_2)

Verificació d'una signatura (f_1, f_2)

$w = (f_2)^{-1} \bmod q$
$w_1 = SHA\ (m)w \bmod q$
$w_2 = f_1\ w \bmod q$
$w_1 P + w_2 Q = (x_0, y_0)$
Acceptar si $x_0 \bmod q = f_1$

Arribats a aquest punt, ha de quedar clar que la complicació que suposa implementar la suma de punts d'una corba el·líptica en lloc de l'aritmètica entera elemental o el problema afegit que suposa haver de calcular el cardinal del grup, són coses que val la pena assumir si tenim en compte l'avantatge que ofereixen aquests criptosistemes: *no es coneix cap algoritme de complexitat subexponencial per al càlcul de logaritme discret en corbes el·líptiques*. L'algoritme Index-Calculus no es pot fer servir perquè no s'ha

trobat cap bona noció de base de factors: com es trien punts P_i tals que molts punts de la corba es puguin escriure fàcilment com a $\sum n_i P_i$? I tampoc es coneix fins ara cap algoritme dissenyat especialment per a aquesta mena de grups. De manera que avui dia l'únic que es pot fer per atacar un criptosistema el·líptic és usar un algoritme de complexitat exponencial, com ara el rho de Pollard. Per aquesta raó, es pot aconseguir la mateixa seguretat que amb un DSA o un RSA usant claus de mida més petita. El premi de 100.000 dòlars que atorga l'empresa *Certicom* a qui resolgui el repte ECC2K-358, un logaritme discret a un grup $E(\mathbf{F}_p)$ amb p primer de 358 bits, és una mostra que aquest problema es cotitza millor que la factorització de nombres enters.

4.2. Varietats abelianes

Amb les corbes el·líptiques plenament incorporades a la criptografia, la qüestió sorgeix de manera natural: per què no utilitzar altres grups d'origen geomètric? El tema es troba en un moment d'intensa recerca i només donarem aquí unes pinzellades molt breus.

Els candidats naturals a iniciar aquest camí de generalització eren les jacobianes de corbes hiperel·líptiques. Una corba hiperel·líptica de gènere g, definida sobre $\mathbf{F}_q$, és

$$C : Y^2 + h(X)\,Y = f(X)$$

no singular, amb f i h polinomis de $\mathbf{F}_q[X]$ tals que deg $f = 2g + 1$ i deg $h \leq g$, i la seva *Jacobiana* Jac $(C) = \mathrm{Div}_0(C)/\mathrm{Pr}(C)$ és una varietat abeliana de dimensió g. Per al seu grup de divisors $\mathbf{F}_q$-racionals cal plantejar-se inicialment les dues qüestions bàsiques: eficiència de l'operació i càlcul del cardinal. Cantor [4] va proposar l'any 1987 un algoritme que opera a la jacobiana de manera eficient, usant una representació polinomial de divisors $D = \mathrm{div}\,(a, b)$. Pel que fa als cardinals, tenim també per començar un interval de Hasse $[q^g + 1 - 2q^{g/2}, q^g + 1 + 2q^{g/2}]$, de manera que $|\mathrm{Jac}\,(C)(\mathbf{F}_q)| = \mathcal{O}(q^g)$, i s'han anat adaptant satisfactòriament els algoritmes del cas el·líptic per al càlcul exacte.

Ara bé, des del moment que disposem d'una representació polinomial per als elements del grup, tenim de manera natural una noció d'irreductibilitat i de factorització. En el context de les corbes hiperel·líptiques, el paper d'irreductibles el fan els *divisors primers*, aquells divisors $D = \mathrm{div}\,(a, b)$ tals que a és un polinomi irreductible. Per determinar la descomposició en divisors primers d'un divisor $D = \mathrm{div}\,(a, b)$ s'ha de factoritzar el polinomi a com a producte $a_1^{e_1}\, a_2^{e_2} \ldots a_L^{e_L}$ de polinomis mònics irreductibles i prendre $b_i = b \bmod a_i$. Aleshores, el divisor descompon $D = \sum \mathrm{e}_i\,\mathrm{div}\,(a_i, b_i)$. Les bases de factors estaran formades per polinomis mònics irreductibles de grau menor que una certa fita.

Adleman, DeMarrais i Huang [1] van descriure el primer algoritme Index-Calculus per resoldre el logaritme discret a la jacobiana d'una corba hiperel·líptica. Amb la condició $\log q \leq (2g+1)^{0{,}98}$, la complexitat del seu algoritme és subexponencial. Les corbes de gènere gran quedaven descartades. Gaudry [8] va ser el primer a presentar un

algoritme Index-Calculus hiperel·líptic que suposava conegut el cardinal del grup. Gaudry analitza el seu algoritme suposant fixat el gènere g i deixant variar q, el cardinal del cos, i obté una complexitat esperada $\mathcal{O}(g^3q^2\log^2 q + g^2\, g!q\, \log^2 q)$. Si $g > 4$, aquest algoritme millora la complexitat dels algoritmes genèrics per al càlcul de logaritmes discrets. Enge i Gaudry van donar després una versió de l'algoritme per a corbes de gènere gran. Si $g > \vartheta \log q$ per a alguna constant positiva ϑ, obtenen complexitat subexponencial. I, per acabar, una versió millorada de l'algoritme de Gaudry proposada per Thériault l'any 2003 supera també els algoritmes genèrics sempre que $g \geq 3$, suggerint que les corbes hiperel·líptiques de gènere 3 podrien no ser tan segures com s'havia cregut prèviament. En resum, sembla que les úniques corbes hiperel·líptiques que poden ser d'utitilitat criptogràfica són les de gènere 2.

De tota manera, aquests resultats que han anant eliminant progressivament les corbes hiperel·líptiques de les propostes criptogràfiques constructives s'han intentat aprofitar per fer una proposta destructiva. La tècnica de la *restricció de Weil*, proposada per Frey, permet traslladar el problema del logaritme discret des d'una corba el·líptica definida sobre un cos $\mathbf{F}_{q^n}$ cap a una varietat abeliana de dimensió n definida sobre el cos $\mathbf{F}_q$. Si aquesta varietat abeliana té com a subvarietat una jacobiana d'una corba hiperel·líptica i les imatges dels punts involucrats en el problema de logaritme discret es troben en aquesta jacobiana, aleshores tenim un possible mètode més eficient que el de Pollard, fins i tot potser de complexitat subexponencial, per resoldre el problema. Això és el que es coneix com a *atac de Gaudry, Hess i Smart* (cf. [9]). Però el cas és que aquest atac no ha afectat gaire les corbes el·líptiques que s'utilitzen a la criptografia: no s'aplica a les corbes el·líptiques definides sobre un cos primer $\mathbf{F}_p$ i, d'altra banda, s'ha demostrat que tampoc s'aplica al cas $q = 2$ i n primer, el més comú en les implementacions.

Sigui com sigui, la recerca en aquest camp no s'atura, ans al contrari es diversifica amb propostes de tota mena. Com a mostra de les darreres tendències suggerim aquests tres temes:

— Criptografia amb corbes de gènere 3 no hiperel·líptiques.
— Criptografia basada en aparellaments $G \times G \to H$.

Weil $E(\mathbf{F}_q)[\ell] \times \mathrm{E}(\mathbf{F}_q)[\ell] \to \mu_\ell$
Tate $E(K)[\ell] \times E(K)/\ell\, E(K) \to K^*/K^{*\ell}$
Tate-Lichtenbaum $\mathrm{Jac}(C)(\mathbf{F}_q)[\ell] \times \mathrm{Jac}(C)(\mathbf{F}_q)/\ell\, \mathrm{Jac}(C)(\mathbf{F}_q) \to \mu_\ell$.

— Criptografia amb grups de Brauer.

Amb ells posem el punt final d'aquesta panoràmica, però de ben segur que són només mostres molt petites de tot el que la matemàtica ofereix i continuarà oferint en el futur a la criptografia.

REFERÈNCIES

[1] L. ADLEMAN, J. DEMARRAIS i M. HUANG, «A subexponential algorithm for discrete logarithms over the rational subgroup of the jacobians of large genus hyperelliptic curves over finite fields», *Algorithmic Number Theory, Lecture Notes in Computer Science* 877: p. 28-40, 1994.

[2] M. AGRAWAL, N. KAYAL i N. SAXENA, «Primes is in P», *Annals of Mathematics* 160 (2): p. 781-793, 2004.

[3] D. BONEH and R. VENKATESAN, «Breaking RSA may not be equivalent to factoring», *Advances in Cryptology-Eurocrypt'98, Lecture Notes in Computer Science*, 1403: p. 59-71 (1998)

[4] D. CANTOR, «Computing in the Jacobian of a Hyperelliptic Curve», *Mathematics of Computation*, 48: p. 95-101, 1987.

[5] W. DIFFIE, «Ultimate Cryptography», *Comunications of the ACM*, 44 (3): p. 84-87, 2001.

[6] W. DIFFIE i M. HELLMAN, «New Directions in Cryptography», *IEEE Transactions on Information Theory* 22 (6): p. 644-654, 1976.

[7] T. ELGAMAL, «A public key cryptosystem and a signature scheme based on discrete logarithms», *IEEE Transactions on Information Theory* 31: p. 469-472, 1985.

[8] P. GAUDRY, «An algorithm for solving the discrete log problem on hyperelliptic curves», *Advances in Cryptology-Eurocrypt 2000, Lecture Notes in Computer Science* 1807: p. 19-34, 2000.

[9] P. GAUDRY, F. HESS, i N.P. SMART, «Constructive and destructive facets of Weil descent on elliptic curves», *Journal of Cryptology*, 15 (1): p. 19-46, 2002.

[10] N. KOBLITZ, «Elliptic curve cryptosystems», *Mathematics of Computation*, 48: p. 203-209, 1987.

[11] A. LENSTRA, «Unbelievable security: Matching AES security using public key systems», *Advances in Cryptology-Asiacrypt 2001, Lecture Notes in Computer Science*, 2248: p. 67-86, 2001.

[12] A. LENSTRA i E. VERHEUL, «Selecting cryptographic key sizes», *Public Key Cryptography-Proceedings of PKC 2000, Lecture Notes in Computer Science*, 1751: p. 446-465, 2000.

[13] U. MAURER, «Cryptography 2000±10», *Informatics: 10 Years Back, 10 Years Ahead. Lecture Notes in Computer Science*, 2000: p. 63-85, 2001.

[14] V. MILLER, «Uses of Elliptic Curves in Cryptography», *Advances in Cryptology-Crypto'85, Lecture Notes in Computer Science*, 218: p. 417-426, 1986.

[15] F. MONTOYA, «La criptografía cuántica, ¿realidad o ficción?» [en línia]: http://www.iec.csic.es/fausto/publica/Montoya04.pdf

[16] R. SCHOOF, «Elliptic curves over finite fields and the computation of square roots mod *p*», *Mathematics of Computation*, 44: p. 483-494, 1985.

[17] C.E. SHANNON, «Communication theory of secrecy systems», *Bell System Technical Journal,* 28: p. 656-715, 1949.
[18] P.W. SHOR, «Algorithms for Quantum Computation: Discrete logarithms and factoring», *SIAM J. Computing*, 26: p. 1.484-1.509, 1997.

Els problemes de Vitushkin i de Painlevé*

Xavier Tolsa

1. INTRODUCCIÓ

En aquesta conferència parlarem dels anomenats *problemes de Painlevé i de Vitushkin*, i de la seva recent resolució, on han jugat un paper clau diversos matemàtics catalans. Agraeixo als organitzadors del IV Cicle Ferran Sunyer i Balaguer, dedicat als set problemes del mil·lenni, el fet d'haver-me convidat a impartir aquesta conferència en aquest cicle. Remarquem, però, que, tot i que els problemes de Painlevé i de Vitushkin han despertat l'interès de nombrosos matemàtics durant les darreres decàdes, aquests problemes no formen part dels set problemes del mil·lenni (malauradament per a la nostra economia).

Painlevé i Vitushkin, a més de grans matemàtics, varen ser dues persones excepcionals. Painlevé va néixer a Paris l'any 1863 i morí el 1933. Fou un matemàtic de gran prestigi i fama durant el seu temps. Per exemple, pels seus treballs va obtenir el «Grand Prix des Sciences Mathématiques», el «Prix Bordin», el «Prix Poncelet», i el 1900 va ser escollit membre de la secció de geometria de l'Académie des Sciences. És especialment recordat per algunes aportacions a l'estudi i classificació de diverses equacions diferencials. Però també va estar molt interessat en problemes relacionats amb la física i l'enginyeria aeronàutica. Com a detall anecdòtic, comentarem que va ser el primer passatger en un vol de Wilbur Wright. I, a més a més de tot això, va ser un polític important: va ser Primer Ministre francès dos cops, i també ministre de les Invencions i ministre de la Guerra en plena Primera Guerra Mundial. Sembla, però, que va ser millor matemàtic que polític...

Vitushkin va néixer a Moscou el 1931 i ha mort recentment: el 2004. Malgrat un accident que li va causar una ceguesa total durant la seva adolescència, fou un gran matemàtic. Va ser alumne de tesi de Kolmogorov, i abans fins i tot de llicenciar-se en matemàtiques ja havia fet importants aportacions al problema 13 de Hilbert (sobre la impossibilitat de resoldre l'equació general de setè grau mitjançant funcions de dues variables). Va treballar en àrees tan allunyades com l'anàlisi complexa (amb una o diverses variables), o la teoria de la informació, i va interessar-se també en qüestions d'enginyeria electrònica i d'acústica. L'any 2004, poc abans de la seva mort, havia estat

*Basat en la conferència pronunciada al IV Cicle Ferran Sunyer i Balaguer, «Els set problemes del mil·lenni». Treball finançat parcialment pels projectes MTM2004-00519 i SGR-2001-00431.

nomenat doctor *honoris causa* per la Universitat Autònoma de Barcelona, honor que podia afegir a d'altres tals com ser membre de l'Acadèmia Russa de Ciències o tenir el Premi de l'Estat Soviètic...

Els problemes dels quals parlarem en aquesta conferència rauen en la intersecció de l'anàlisi complexa i l'anàlisi de Fourier. A grans trets, podríem dir que l'Anàlisi complexa estudia les funcions complexes de variable complexa: $f : \mathbb{C} \to \mathbb{C}$ (això gairebé és una tautologia). D'altra banda, l'anàlisi de Fourier estudia les funcions amb tècniques provinents o inspirades per la descomposició en sèries de Fourier.

Els problemes de Painlevé i de Vitushkin són problemes de matemàtica fonamental, i jo no en conec cap possible aplicació a la «vida pràctica». En qualsevol cas, si volguéssim una justificació per al seu estudi, sempre podríem dir que el tipus anàlisi de Fourier utilitzat per atacar-los és imprescindible per a l'estudi d'equacions en derivades parcials que regeixen diferents lleis físiques (eq. d'ones, eq. de la calor...) i és a la base de desenvolupaments tecnològics recents tals com tomografies, escàners, tècniques de compressió d'imatges (JPEG, JPEG2) i sons, etc.

2. SINGULARITATS EVITABLES PER A FUNCIONS ANALÍTIQUES ACOTADES: EL PROBLEMA DE PAINLEVÉ

Recordem que, donat un obert $\Omega \subset \mathbb{C}$, una funció $f : \Omega \subset \mathbb{C} \to \mathbb{C}$ és analítica en Ω si el límit $\lim_{w \to z} \frac{f(w)-f(w)}{w-z}$ existeix per a tot z de Ω. És a dir, si f és derivable en sentit complex en Ω.

Un teorema clàssic de Riemann afirma que si f és analítica i acotada en un entorn d'un punt z_0, aleshores f es pot estendre analíticament al punt z_0. Observem que això és completament diferent del que passa en el cas real. Per exemple, sigui f la funció idènticament nul·la en $(-\infty, 0)$ i que val 1 en $[0, +\infty)$. Clarament, f és derivable fora de l'origen, però no es pot estendre de manera que també sigui derivable (ni tan sols contínua) a l'origen.

Diem que un conjunt compacte E del pla complex és evitable (o singularitat evitable) per a les funcions analítiques acotades si, donat un obert Ω qualsevol que contingui E, tota funció analítica i acotada en $\Omega \setminus E$ es pot estendre analíticament a tot Ω. Així, pel teorema de Riemann, un punt és una singularitat evitable per a les funcions analítiques acotades. Utilitzant la fórmula integral de Cauchy, no és gaire difícil demostrar que en la definició anterior ens podem limitar a considerar només un entorn obert Ω de E fixat. En altres paraules, obtenim una definició equivalent de singularitat evitable si prenem, per exemple, $\Omega = \mathbb{C}$.

Vegem alguns exemples. Es pot demostrar (mitjançant el teorema de Riemann i el teorema de categoria de Baire) que si E és un compacte donat per una unió numerable de punts, aleshores és evitable. D'altra banda, un disc tancat no és evitable. Per veure-ho,

podem considerar la funció $f(z) = \frac{1}{z-z_0}$, on z_0 és el centre del disc. Aquesta funció és analítica i acotada fora del disc, per`o no es pot estendre al propi disc de manera que es mantingui analítica i acotada (ja que pel principi de prolongació analítica, hauria de coincidir amb $1/(z-z_0)$ en $\mathbb{C}\setminus\{z_0\}$).

Si E és un compacte simplement connex diferent d'un punt, també és no evitable. Això es pot comprovar fàcilment considerant la transformació conforme $f\colon \mathbb{C}_\infty\setminus E \to D(0,1)$, on $\mathbb{C}_\infty$ denota l'esfera de Riemann i $D(0,1)$ el disc unitat. Finalment, si E és un compacte connex arbitrari i diferent d'un punt, un raonament anàleg mostra que E també és no evitable.

Observem que, directament de la definició, es dedueix que si E i F són compactes, E és no evitable i $E \subset F$, aleshores F també és no evitable. Com a conseqüència d'aquest fet i del darrer exemple, tenim que els compactes evitables per a les funcions analítiques acotades són totalment disconnexos.

De la discussió anterior n'hauríem de treure la conclusió que si E és «gran» en algun sentit, aleshores és no evitable, mentre que si és «petit» llavors és evitable. L'anomenat *problema de Painlevé* (proposat pel matemàtic francés Painlevé a finals del segle XIX) consisteix a trobar una caracterització en termes mètrics i/o geomètrics de les singularitats evitables per a les funcions analítiques acotades.

3. LA CAPACITAT ANALÍTICA I LA PREGUNTA DE VITUSHKIN

La capacitat analítica d'un compacte E és

$$\gamma(E) = \sup |f'(\infty)|,$$

on el suprem es pren sobre totes les funcions $f\colon \mathbb{C}_\infty\setminus E \to \mathbb{C}$ analítiques tals que $|f| \leq 1$ en $\mathbb{C}\setminus E$, essent $f'(\infty) = \lim_{z\to\infty} z(f(z) - f(\infty))$. Notem que aquestes funcions tenen una singularitat evitable a l'inifinit, i per tant $f(\infty)$ està ben definit. Podeu comprovar fàcilment que $f'(\infty)$ coincideix amb el coeficient a_{-1} de la sèrie de Laurent de f centrada en el zero.

La noció de capacitat analítica fou introduïda per Ahlfors [Ah] l'any 1947. Ahlfors va demostrar que E és evitable per a les funcions analítiques acotades si, i només si, $\gamma(E) = 0$. Aquest resultat ens proporciona una caracterització molt útil de les singularitats evitables. Ara bé, això no és una solució del problema de Painlevé ja que aquesta caracterització no és en termes mètrics o geomètrics (tot i que la denominació *termes mètrics o geomètrics* és poc precisa), perquè a la definició de *capacitat analítica* apareix un suprem sobre totes les funcions analítiques complint unes certes propietats.

De fet, la definició de capacitat analítica és una mica «fosca» i complicada. Per

exemple, encara és un problema obert saber si es compleix la següent desigualtat per a qualsevol parella de compactes $E, F \subset \mathbb{C}$:

$$\gamma(E \cup F) \leq \gamma(E) + \gamma(F)?$$

Una versió una mica més dèbil d'aquesta desigualtat és la següent:

Pregunta de Vitushkin. *Existeix una constant universal* C *tal que*

$$\gamma(E \cup F) \leq C\,(\gamma(E) + \gamma(F)) \tag{3.1}$$

per a qualsevol parella de compactes $E, F \subset \mathbb{C}$*?*

Si la resposta és afirmativa, es diu que γ és semiadditiva.

La semiadditivitat de la capacitat analítica és una pregunta plantejada per Vitushkin en la dècada de l'any 1960 (vegeu [Vi]), que va estudiar diversos problemes d'aproximació racional uniforme en compactes del pla complex i va demostrar alguns criteris d'aproximació que s'expressen en termes de la capacitat analítica. Vitushkin va provar que una resposta afirmativa a la pregunta de la semiadditivitat permetria comprendre millor com són els compactes on és vàlida l'aproximació racional uniforme.

4. RELACIÓ ENTRE LES MESURES DE HAUSDORFF I LA CAPACITAT ANALÍTICA

Estudiar el problema de Painlevé equival a estudiar les propietats mètriques i geomètriques de la capacitat analítica.

Per parlar d'algunes d'aquestes propietats cal saber com mesurar conjunts en el pla complex, i per això necessitem definicions addicionals.

Donat un conjunt $A \subset \mathbb{C}$ i $s > 0$, denotem

$$\mathcal{H}^s_\varepsilon(A) = \inf \sum_i \operatorname{diam}(A_i)^s,$$

on l'ínfim es pren sobre tots els recobriments $A \subset \bigcup_i A_i$ tals que $\operatorname{diam}(A_i) \leq \varepsilon$. La mesura de Hausdorff s-dimensional de A és

$$\mathcal{H}^s(A) = \lim_{\varepsilon \to 0} \mathcal{H}^s_\varepsilon(A) = \sup_{\varepsilon > 0} \mathcal{H}^s_\varepsilon(A).$$

La mesura de Hausdorff 1-dimensional $\mathcal{H}^1$ també s'anomena longitud (i evidentment coincideix amb la noció intuitiva que hom té de longitud), i la mesura 2-dimensional coincideix amb la mesura d'àrea.

Per a qualsevol conjunt $A \subset \mathbb{C}$, existeix un nombre $s_0 \in [0,2]$ tal que $\mathcal{H}^s(A) = \infty$ si $s < s_0$, i $\mathcal{H}^s(A) = 0$ si $s > s_0$. Aquest valor s_0 rep el nom de *dimensió de Hausdorff* de A, i es denota per $\dim(A)$. Per exemple, un punt té dimensió 0, una corba de longitud finita (també anomenada *rectificable*) té dimensió 1, i un conjunt d'àrea positiva té dimensió 2. Notem que també existeixen conjunts de dimensió no entera. De fet, per a qualsevol $s \in [0,2]$ es pot construir un conjunt amb dimensió de Hausdorff s. Per a més informació sobre les mesures de Hausdorff, podeu consultar [Fa] o [Ma].

La relació entre les mesures de Hausdorff i la capacitat analítica és la següent:

— Si $\dim(E) > 1$, llavors $\gamma(E) > 0$ (que equival a dir que E no és evitable).
— $\gamma(E) \leq \mathcal{H}^1(E)$. En particular, si $\dim(E) < 1$, llavors $\gamma(E) = 0$ i E és evitable.

La primera afirmació es dedueix d'un resultat clàssic en teoria del potencial, el Lema de Frostman; mentre que la segona afirmació va ser demostrada (amb un altre llenguatge) per Painlevé fa aproximadament un segle. Observem que aquestes propietats són una formulació precisa de la idea que els conjunts petits són evitables, i els grans no ho són.

A partir de les relacions entre mesures de Hausdorff i capacitat analítica, veiem que la dimensió de Hausdorff crítica és 1. El cas en què la dimensió de E és 1 és delicat, i hom està temptat de pensar que E és no evitable si, i només si, la seva longitud és estrictament positiva. Això és el que es conjecturava fins que Vitushkin, a principis dels anys 1960, en va donar un contraexemple. Vitushkin va construir un compacte de longitud positiva i finita (i, per tant, amb dimensió de Hausdorff 1) que era evitable.

Un exemple una mica posterior (però més senzill) d'un compacte E amb $0 < \mathcal{H}^1(E) < \infty$ i $\gamma(E) = 0$ és l'anomenat *conjunt de Cantor planar 1/4*, que denotarem per $E_{1/4}$. Aquest conjunt es construeix de la manera següent: Considerem un quadrat Q^0 (per definició, «quadrat» vol dir «quadrat tancat») de costat igual a 1. En la primera etapa de la construcció substituïm Q^0 per quatre quadrats Q_i^1, $i = 1, \ldots, 4$, de costat 1/4 continguts en Q^0, de forma que cada quadrat Q_i^1 contingui un vèrtex diferent de Q^0. Anàlogament, en la segona etapa substituïm cada quadrat Q_i^1 per quatre quadrats disjunts continguts en Q_i^1, de forma que cadascun d'ells contingui un dels vèrtexs de Q_i^1. Així, haurem construit setze quadrats Q_j^2 de costat 1/16. Procedim inductivament (vegeu la figura 1), de manera que en cada etapa n tenim 4^n quadrats disjunts Q_i^n de costat 4^{-n}. Denotem $E_n = \bigcup_{i=1}^{4n} Q_i^n$. El conjunt de Cantor 1/4 és $E_{1/4} = \bigcap_{n=1}^{\infty} E_n$.

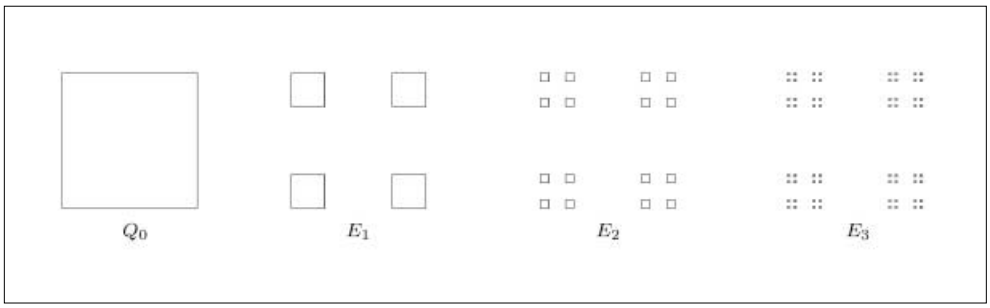


Figura 1. El quadrat Q_0 i els conjunts E_1, E_2 i E_3 que apareixen en les etapes 1, 2 i 3 de la construcció del conjunt de Cantor planar 1/4.

Clarament $E_{1/4}$ és un compacte no buit (perquè és una intersecció de compactes encaixats). A més a més, per a cada $n \geq 0$ es compleix

$$\sum_{i=1}^{4n} \ell(Q_i^n) = 1,$$

on $\ell(Q_i^n)$ denota la longitud del costat de Q_i^n. D'aquest fet es dedueix que $\mathcal{H}^1(E_{1/4}) \leq \sqrt{2}$, i amb una mica més de feina no és gaire difícil comprovar que també $\mathcal{H}^1(E_{1/4}) \geq 1$. Que $E_{1/4}$ és evitable (*i.e.*, $\gamma(E_{1/4}) = 0$) és una mica més complicat de veure, i va ser demostrat per Garnett [Ga] i Ivanov (vegeu [Iv], p.153) independentment.

El conjunt de Cantor planar 1/4 té longitud positiva i finita, igual que una corba de longitud positiva i finita. Ara bé, un té capacitat analítica nul·la i l'altra no. Quina propietat geomètrica és responsable d'aquest fet? La rectificabilitat.

Un conjunt s'anomena *rectificable* si està contingut en una unió numerable de corbes rectificables (*i.e.*, de longitud positiva i finita). D'altra banda, si la intersecció d'un conjunt amb tota corba rectificable té longitud zero, diem que el conjunt és *purament no rectificable*. Una definició equivalent és la següent: un conjunt és purament no rectificable si no té subconjunts de longitud positiva que siguin rectificables.

A la primera meitat del segle XX, Besicovitch va demostrar que tot conjunt de longitud positiva i finita es pot posar com a unió d'un rectificable i d'un purament no rectificable, i també va provar que els conjunts rectificables i els purament no rectificables tenen propietats geomètriques molt diferents en quant a l'existència de tangents, densitats, projeccions... (vegeu [Ma] o [Fa] per a més detalls).

El conjunt de Cantor planar 1/4 és purament no rectificable (això es pot deduir del fet que les seves projeccions sobre els eixos horitzontal i vertical tenen longitud nul·la), i també els compactes construits per Vitushkin que tenien capacitat analítica nul·la i longitud positiva i finita. Per aquesta raó, Vitushkin va formular la següent conjectura a principis de la dècada dels 1960.

Conjectura de Vitushkin. *Un compacte del pla complex amb longitud positiva i finita té capacitat analítica nul·la si, i només si, és purament no rectificable.*

5. LA TRANSFORMADA DE CAUCHY, CURVATURA I RECTIFICABILITAT

La fórmula integral de Cauchy és una eina fonamental en l'anàlisi complexa. Donat un obert Ω (que suposem simplement connex per simplicitat) i una corba tancada $\Gamma \subset \Omega$ (prou bona), la fórmula integral de Cauchy diu que si Γ dóna una volta al voltant de $w \in \Omega$ (*i.e.*, l'índex de w respecte Γ és 1) i f és analítica en Ω, llavors

$$f(w) = \int_\Gamma \frac{1}{z-w}\, f(z)\, \frac{dz}{2\pi i}.$$

Introduïm ara el concepte de *transformada de Cauchy* d'una mesura. Donada una mesura de Borel finita μ en $\mathbb{C}$, la transformada de Cauchy de μ és la funció

$$C\mu(w) = \int \frac{1}{z-w}\, d\mu(z).$$

Derivant sota el signe integral es comprova fàcilment que, fora del suport de μ, $C\mu$ és una funció analítica. Sobre el suport de μ pot no ser-ho, i de fet pot passar que la integral en la definició de $C\mu$ no sigui absolutament convergent. Esmentem però que $\int \frac{1}{|z-w|}\, d\mu(z)$ és absolutament convergent per a quasi tot $w \in \mathbb{C}$ (respecte a la mesura de Lebesgue) i, per tant, $C\mu(w)$ està ben definit per a gairebé tot $w \in \mathbb{C}$.

Tot i que, en general, quan parlem de mesures ens referim a mesures positives, la noció de transformada de Cauchy també té sentit per a mesures complexes. Així, donada f analítica en Ω, si considerem la mesura $f(z)\, \frac{dz\lfloor\Gamma}{2\pi i}$, podem reescriure la fórmula integral de Cauchy de la manera següent:

$$f(w) = C\left(f(z)\, \tfrac{dz\lfloor\Gamma}{2\pi i}\right)(w),$$

per a Γ i w apropiats.

Recordem que el problema de Painlevé té a veure amb l'existència o no de funcions analítiques acotades. La transformada de Cauchy és una eina molt útil, entre d'altres raons perquè ens proporciona una manera fàcil de «fabricar» funcions analítiques a partir de mesures. El problema fonamental rau a conèixer quan podrem aconseguir que la transformada de Cauchy d'una mesura sigui una funció acotada.

Siguem ara una mica més precisos. Donat un compacte $E \in \mathbb{C}$, si existeix una mesura μ no nul·la suportada en E tal que $C\mu$ és una funció acotada en $\mathbb{C} \setminus E$, alesho-

res E és una singularitat no evitable, ja que (denotant per $\|C\mu\|_{\infty,\mathbb{C}\setminus E}$ el suprem de $|C\mu|$ sobre $\mathbb{C}\setminus E$),

$$\gamma(E) \geq \frac{|(C\mu)'(\infty)|}{\|C\mu\|_{\infty,\mathbb{C}\setminus E}} = \frac{\mu(E)}{\|C\mu\|_{\infty,\mathbb{C}\setminus E}} > 0.$$

En la darrera igualtat, el fet que $|(C\mu)'(\infty)|$ coincideix amb la massa total $\mu(E)$ es dedueix fàcilment a partir de les definicions.

Es diu que la transformada de Cauchy està acotada en $L^2(\mu)$, o que és un operador acotat en $L^2(\mu)$, si es verifica la desigualtat

$$\int |C(f\,d\mu)|^2\,d\mu \leq A\int |f|^2\,d\mu.$$

per a tota funció f, essent A una constant independent de f. Del problema de la definició de f sobre el suport de μ preferim no parlar-ne gaire. Per simplicitat, podeu pensar que $C\mu$ està definida sobre el suport de μ com a valor principal.

L'estudi de l'acotació en $L^2(\mu)$ de la transformada de Cauchy va ser un problema central en l'anàlisi harmònica durant les dècades dels anys 1970 i 1980. Coifman, McIntosh i Meyer [CMM] (generalitzant un resultat previ de Calderón [Ca]) van demostrar que si hom pren com a mesura μ la longitud sobre el graf d'una funció Lipschitz de $\mathbb{R}$ en $\mathbb{R}$, aleshores la transformada de Cauchy és acotada sobre $L^2(\mu)$. Utilitzant un argument de dualitat, aquest resultat implica que tot subconjunt de longitud positiva d'una corba rectificable té capacitat analítica positiva. Això equival a dir que si $E\subset\mathbb{C}$ té longitud finita i capacitat analítica nul·la, aleshores E és purament no rectificable, la qual cosa demostra que una de les implicacions de la conjectura de Vitushkin és certa.

De la discussió anterior queda clar que hi ha una certa relació entre capacitat analítica, acotacions L^2 de la transformada de Cauchy i rectificabilitat. El concepte de curvatura d'una mesura, introduït l'any 1995 per Mark Melnikov [Me], és fonamental per aclarir aquestes relacions. Donats tres punts $x,y,z\in\mathbb{C}$ diferents dos a dos, sigui $\mathbb{R}(x,y,z)$ el radi de la circumferència que passi per x,y,z (amb $R(x,y,z)=\infty$ si els punts són alineats). La curvatura de la mesura μ és

$$c^2(\mu) := \iiint \frac{1}{R(x,y,z)^2}\,d\mu(x)\,d\mu(y)\,d\mu(z).$$

Melnikov va demostrar que si μ està suportada en un compacte E i $\mu(B(x,r))\leq r$ per a tota bola $B(x,r)\subset\mathbb{C}$ (a partir d'ara direm que μ té *creixement lineal* si verifica aquesta propietat), aleshores

$$\gamma(E) \geq C \ \frac{\mu(E)^2}{\mu(E) + c^2(\mu)}, \tag{5.1}$$

on $C > 0$ és una constant universal. Observeu que si E suporta una mesura amb creixement lineal i curvatura finita aleshores $\gamma(E) > 0$ i E és no evitable.

La connexió de la curvatura amb la transformada de Cauchy prové de la la següent identitat, descoberta per Melnikov i Verdera [MV]:

$$\int |C(\mu)|^2 \, d\mu = \tfrac{1}{6} \, c^2(\mu) + O(\mu), \tag{5.2}$$

per a tota mesura μ amb creixement lineal, essent $O(\mu)$ un terme «d'error» tal que $|O(\mu)| \leq 100\mu(\mathbb{C})$ (aquí, la constant 100 no és òptima). De la fórmula anterior i de la desigualtat 5.1 es dedueix fàcilment que si E suporta una mesura μ amb creixement lineal tal que la transformada de Cauchy està acotada en $L^2(\mu)$, llavors E és no evitable (en realitat, això ja era conegut abans de la introducció de la noció de curvatura d'una mesura).

La curvatura d'una mesura conté informació geomètrica del seu suport. Per exemple, és immediat comprovar que una mesura està suportada en una recta si, i només si, la seva curvatura és nul·la.

Ens plantegem ara la següent pregunta: Quines propietats sobre el suport d'una mesura μ podem deduir a partir d'estimacions sobre la seva curvatura $c^2(\mu)$? Hom pot pensar que si $c^2(\mu)$ és petita, aleshores μ «tendirà a concentrar-se sobre rectes», en algun sentit. Això és cert si μ és cert si μ és la mesura de longitud sobre un conjunt de longitud finita. El resultat precís, demostrat per David i Léger [Lé], és el següent:

Teorema 5.1. *Sigui $E \subset \mathbb{C}$ un compacte de longitud finita. Si $c^2(\mathcal{H}^1_{|E}) < \infty$, aleshores E és rectificable.*

Insistim en el fet que aquest resultat només afecta conjunts de longitud finita i la curvatura de la mesura de longitud. Hi ha mesures, com per exemple la mesura de Lebesgue planar sobre un disc, tals que la seva curvatura és finita però el seu suport no està contingut en un conjunt rectificable.

6. LA SOLUCIÓ DE LA CONJECTURA DE VITUSHKIN

L'any 1997 Guy David [David2] va provar que la conjectura de Vitushkin és certa:

Teorema 6.1. *Si $E \subset \mathbb{C}$ és un compacte amb longitud finita, aleshores É és evitable (equivalentment $\gamma(E) = 0$) si, i només si, E és purament no rectificable.*

Remarquem que per a compactes AD regulars (que ara no definirem), la conjectura de Vitushkin va ser demostrada prèviament per Mattila, Melnikov i Verdera a [MMV].

La implicació del teorema anterior demostrada per Guy David és

$$E \text{ purament no rectificable} \quad \Rightarrow \quad \gamma(E) = 0 \tag{6.1}$$

la qual cosa equival a veure que, si E és no evitable ($\gamma(E) > 0$), llavors existeix $F \subset E$ de longitud finita rectificable. L'altra implicació del teorema ja era coneguda (i ha estat comentada a la secció 5).

El camí per demostrar 6.1 és el següent. Sigui E no evitable amb longitud finita. Això equival a dir que existeix una funció analítica i acotada en $\mathbb{C} \setminus E$ no constant. La primera part de la demostració consisteix a provar (a partir de l'existència d'aquesta funció, que podem suposar que coincideix amb la transformada de Cauchy d'una mesura complexa suportada en F) que aleshores existeix un subconjunt $F \subset E$ de longitud positiva on la transformada de Cauchy és acotada respecte de la mesura de longitud. Aquest és un problema d'anàlisi harmònica força delicat que, pel qui sàpiga una mica d'anàlisi harmònica, direm que es redueix bàsicament a provar un teorema de tipus $T(b)$ sense cap condició de doblament. Aquesta part de la demostració és la que es troba a [Da] (i també a [DM], en part). Una demostració més senzilla posterior és a [NTV].

En la segona part de la demostració de 6.1 es tracta de veure que si la transformada de Cauchy és acotada per a la mesura de longitud sobre un conjunt F de longitud finita, aleshores F és rectificable. Aquí és on intervé la curvatura: de rfeqmv es dedueix que la curvatura de la mesura de longitud sobre F és finita, i aleshores pel teorema 5.1, F és rectificable.

El teorema 6.1 és un resultat molt important. Observeu, però, que es refereix únicament als conjunts de longitud finita. Encara no hem parlat de cap criteri de tipus mètric/geomètric que ens serveixi per decidir si els compactes de longitud inifinita són evitables o no. D'aquesta qüestió i de la semiadditivitat de la capacitat analítica en parlarem a la secció següent.

7. ELS CONJUNTS DE LONGITUD INFINITA I LA SEMIADDITIVITAT DE LA CAPACITAT ANALÍTICA

Ja hem vist que per als compactes de longitud finita la rectificabilitat és el factor essencial per poder decidir si un compacte és evitable o no. Per als conjunts de longitud

infinita no és així. Hi ha compactes E de longitud infinita que són no evitables i tals que la intersecció de E amb tota corba rectificable té longitud zero (equivalentment, E és purament no rectificable). Vegem-ne un exemple: considerem un conjunt anàleg al conjunt de Cantor planar 1/4, però de manera que la raó entre quadrats de generacions consecutives sigui 1/3 en lloc de 1/4. Aquest conjunt, que podem anomenar *conjunt de Cantor planar 1/3*, té dimensió de Hausdorff $\log 4/\log 3 > 1$ i, per tant, no és evitable. Però, en canvi, té la propietat que interseca tota corba rectificable en conjunts de longitud zero (això es pot deduir del fet que les projeccions d'aquest conjunt sobre els eixos de coordenades coincideixen amb el familiar conjunt ternari de Cantor i, per tant, tenen longitud zero).

Per a compactes arbitraris a [To1] s'ha demostrat el següent resultat:

Teorema 7.1. *Sigui $E \subset \mathbb{C}$ compacte. Aleshores,*

$$\gamma(E) \approx \sup \mu(E), \tag{7.1}$$

on el suprem es pren sobre totes μ suportades en E, amb creixement lineal tals que $c^2(\mu) \leq \mu(E)$.

La notació $A \approx B$ significa que existeix una constant universal $C > 0$ tal que $C^{-1}A \leq B \leq CA$. Recordeu que diem que μ té creixement lineal si $\mu(B(x,r)) \leq r$ per a tota bola $B(x,r) \subset \mathbb{C}$.

En el teorema 7.1, la desigualtat $\gamma(E) \geq C^{-1} \sup \mu(E)$ (on $C > 0$ és una constant universal) és una conseqüència senzilla de la desigualtat rfmark. Per demostrar l'altra desigualtat del teorema (*i.e.* que $\gamma(E) \leq C \sup \mu(E)$) cal provar l'existència d'una mesura μ suportada en E amb creixement lineal, amb $c^2(\mu) \leq \mu(E)$ i tal que $\mu(E) \approx \gamma(E)$. La part complicada de tot això és el control de la curvatura $c^2(\mu)$. No és gaire difícil comprovar, utilitzant rfeqmv, que si aconseguim construir μ suportada en E amb creixement lineal, amb $\mu(E) \approx \gamma(E)$, i tal que la transformada de Cauchy sigui acotada en $L^2(\mu)$ (amb constants universals), haurem resolt el problema.

Recordeu que per demostrar la conjectura de Vitushkin (teorema 6.1) calia fer una cosa anàloga. S'havia de trobar $F \subset E$ tal que la transformada de Cauchy fos acotada en L^2 respecte a la mesura $\mu := \mathcal{H}^1_{|F}$ (suposant $\gamma(\mathrm{E}) > 0$). Ara bé, la demostració del teorema 7.1 té algunes complicacions addicionals. Per exemple, no es pot prendre μ igual a la longitud sobre cap subconjunt F de longitud finita (ja que pot ocórrer que $\gamma(F) = 0$ per tot $F \subset E$ amb $\mathcal{H}^1(F) < \infty$). De fet, si un intenta estendre d'una forma naïf els arguments utlitzats per a la demostració del teorema 6.1, les constants involucrades en l'estimació 7.1 exploten. Una de les idees noves utilitzades en la demostració del teorema 7.1 a [To1] que impedeix aquesta «explosió» és una mena d'inducció «en escales».

Una conseqüència immediata del teorema 7.1 és el següent resultat (conjecturat prèviament per Melnikov el 1995):

Corol·lari 7.2. *Sigui $E \subset \mathbb{C}$ compacte. E és no evitable si, i només si, existeix una mesura no nul·la suportada en E amb creixement lineal i curvatura finita.*

Observeu que aquest corol·lari proporciona una caracterització dels conjunts no evitables en termes de la curvatura de mesures, i que la noció de curvatura té un caràcter marcadament geomètric. Tanmateix, aquest resultat té l'inconvenient que parla de l'existència d'una certa mesura, però no de com construir o com poder decidir si existeix aquesta mesura. Remarquem, però, que en el cas en què E és connex, P. Jones (vegeu [Pa], capítol 3) ha mostrat un mètode constructiu per obtenir l'esmentada mesura μ amb curvatura finita.

Tot i que la caracterització de les singularitats evitables donada pel corol·lari 7.2 no és tan simple com voldríem, aquest resultat augmenta enormement el nostre coneixement sobre les propietats de les singularitats evitables i sobre la capacitat analítica. A tall d'exemple, abans de la demostració del teorema 7.1 no se sabia si la classe del conjunts evitables era invariant per a bijeccions afins del pla complex. Així, donada l'aplicació $\varphi : \mathbb{C} \to \mathbb{C}$ definida per $\varphi(x, y) = (x, 2y)$ (amb $x, y \in \mathbb{R}$) hom ignorava si era certa la següent equivalència:

$$E \text{ evitable} \iff \varphi(E) \text{ evitable}$$

Ara, es pot comprovar fàcilment que aquest resultat és una conseqüència immediata del corol·lari 7.2. De fet, recentment s'ha demostrat a [To3] que si φ és un difeomorfisme en $\mathbb{C}$ de classe C^1 (en realitat, és suficient que sigui bilipschitz), aleshores 7.2 és cert.

D'altra banda, el teorema 7.1 té una altra conseqüència important: resol el problema de la semiadditivitat plantejat per Vitushkin.

Corol·lari 7.3. *La capacitat analítica és semiadditiva. És a dir, existeix una constant universal C tal que per a qualssevol compactes $E, F \subset \mathbb{C}$,*

$$\gamma(E \cup E) \leq C(\gamma(E) + \gamma(F)).$$

Aquest corol·lari es dedueix del fet que el terme sup $\mu(E)$ que apareix a la dreta de l'estimació 7.1 es comporta de forma semiadditiva. Tal com hem comentat anteriorment, aquest resultat té aplicacions a certs problemes d'aproximació racional uniforme sobre compactes del pla complex. A [To2] es mostra una d'aquestes aplicacions (en concret, a la resolució de l'anomenada *conjectura de la frontera interior*).

REFERÈNCIES

[Ah] AHLFORS, L. «Bounded analytic functions», *Duke Math. J.* 14 (1947), p. 1-11.

[Ca] CALDERÓN, A.P. «Cauchy integrals on Lipschitz curves and related operators», *Proc. Nat. Acad. Sci.* USA 74 (1977), p. 1.324-1.327.

[CMM] COIFMAN, R.R., MCINTOSH, A. i MEYER, Y., «L'intégrale de Cauchy définit un opérateur borné sur L^2 pour les courbes lipschitziennes», *Ann. of Math.* (2)116 (1982), p. 361-387.

[Da] DAVID, G. «Unrectifiable 1-sets have vanishing analytic capacity», *Revista Mat. Iberoamericana* 14(2) (1998), p. 369-479.

[DM] DAVID, G. i MATTILA, P. «Removable sets for Lipschitz harmonic functions in the plane», *Rev. Mat. Iberoamericana* 16(1) (2000), p. 137-215.

[Fa] FALCONER, K. *The geometry of fractal sets*, Cambridge Univ. Press, 1985.

[Ga] GARNETT, J.B. «Positive length but zero analytic capacity», *Proc. Amer. Math. Soc.* 24 (1970), p. 696-699.

[Iv] IVANOV, L.D. *On sets of analytic capacity zero* in «Linear and Complex Analysis Problem Book 3» (part II), *Lectures Notes in Mathematics* 1574, Springer-Verlag, Berlin, 1994, p. 150-153.

[Jo] JONES, P.W. «Rectifiable sets and the travelling salesman problem», *Invent. Math.* 102 (1990), p. 1-15.

[Lé] LÉGER, J.C. «Menger curvature and rectifiability», *Ann. of Math.* 149 (1999), p. 831-869.

[Ma] MATTILA, P. «Geometry of sets and measures in Euclidean spaces», *Cambridge Stud. Adv. Math.* 44, Cambridge Univ. Press, Cambridge, 1995.

[MMV] MATTILA, P., MELNIKOV, M.S. i VERDERA, J. «The Cauchy integral, analytic capacity, and uniform rectifiability», *Ann. of Math.* (2) 144 (1996), p. 127-136.

[Me] MELNIKOV, M.S. «Analytic capacity: discrete approach and curvature of a measure», *Sbornik: Mathematics* 186(6) (1995), p. 827-846.

[MV] MELNIKOV, M.S. i VERDERA, J. «A geometric proof of the L^2 boundedness of the Cauchy integral on Lipschitz graphs», *Internat. Math. Res. Notices* (1995), p. 325-331.

[NTV] NAZAROV, F., TREIL, S. i VOLBERG, A. *How to prove Vitushkin's conjecture by pulling ourselves up by the hair*, Preprint (2000).

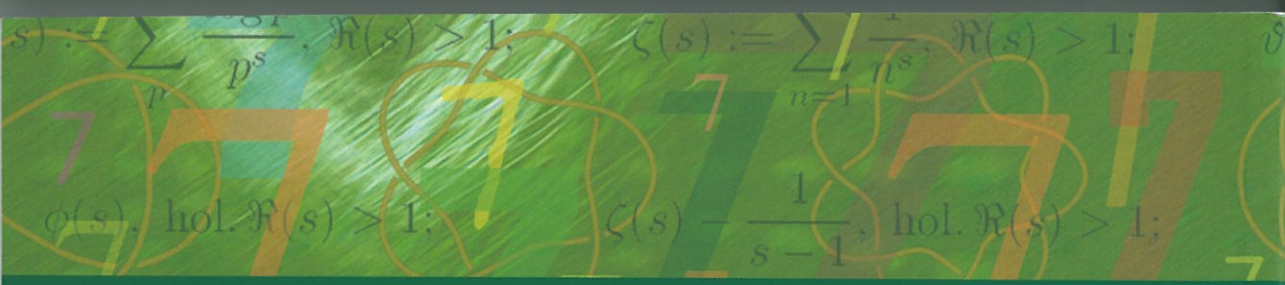
[Pa] PAJOT, H. «Analytic capacity, rectifiability, Menger curvature and the Cauchy integral», *Lecture Notes in Math.* 1799 (2002), Springer.

[To1] TOLSA, X. «Painlevé's problem and the semiadditivity of analytic capacity», *Acta Math.* 190:1 (2003), p. 105-149.

[To2] TOLSA, X. «The semiadditivity of continuous analytic capacity and the inner boundary conjecture», *Amer. J. Math.* 126 (2004), p. 523-567.

[To3] TOLSA, X. *Bilipschitz maps, analytic capacity, and the Cauchy integral* Preprint (2003). Apareix a *Ann. of Math.*

[Vi] VITUSHKIN, A.G. «The analytic capacity of sets in problems of approximation theory», *Uspeikhi Mat. Nauk.* 22(6)(1967), 141-199 (Russian); a: *Russian Math. Surveys* 22 (1967), p. 139-200.



El cicle de conferències Ferran Sunyer i Balaguer és una iniciativa de la Fundació Caixa Sabadell i el Centre de Recerca Matemàtica. L'edició de 2005 es va dedicar a la recerca en Matemàtiques, tot presentant una mostra d'alguns dels problemes oberts (o *conjectures*) als quals els investigadors dediquen actualment els seus esforços.

El fil conductor de les conferències va ser *els set problemes del mil·lenni*. Es tracta d'una selecció de problemes que pertanyen a diverses àrees de les Matemàtiques, que són una mostra representativa dels reptes més importants per a la recerca que s'està fent actualment. L'institut Clay —una fundació privada dels EUA que té per objecte promoure la recerca en Matemàtiques— ofereix un premi d'un milió de dòlars a qui aconsegueixi resoldre cadascun d'aquests problemes.

En aquesta quarta edició del cicle Ferran Sunyer i Balaguer es va dedicar una sessió a cadascun dels set problemes del mil·lenni, a càrrec d'un investigador especialista en el tema, i la sèrie es va completar amb altres tres conferències, fins a un total de deu, on es van discutir altres temes relacionats amb la recerca matemàtica i la resolució de conjectures. En tractar-se de conferències destinades a un públic no especialista, els conferenciants van haver de fer un esforç considerable per tal d'evitar tant com fos possible tots els tecnicismes que hauria comportat una descripció rigorosa dels problemes, i substituir-los per idees intuïtives que permetessin als assistents la comprensió de les idees fonamentals que hi ha en el fons de cadascun d'ells.

Aquest volum és un recull d'articles redactats pels ponents del cicle amb una versió ampliada del contingut de les seves conferències. Esperem que aquesta publicació sigui útil a totes les persones interessades a tenir una panoràmica sobre les fronteres actuals del coneixement matemàtic a través d'algunes de les grans preguntes que preocupen els més destacats investigadors.

CENTRE DE RECERCA MATEMÀTICA CRM

INSTITUT D'ESTUDIS CATALANS MCMVII

OBRA SOCIAL Caixa Sabadell

ISBN 978-84-95166-67-8