MATEMÀTIQUES: COMENCEU PER AQUÍ

Jaume Aguadé

pàgina (gairebé) en blanc

MATEMÀTIQUES: COMENCEU PER AQUÍ

Jaume Aguadé

ISBN: 978-84-09-62739-4

L'autor d'aquesta obra és Jaume Aguadé Bover
Jaume.Aguade@uab.cat
jaume@jaumeaguade.cat

Versió 1.3.1

4 de juny de 2024
162è aniversari del naixement de **Teresa Claramunt**,
pionera del feminisme obrerista i anarcosindicalista;
empresonada, torturada, expulsada del país;
reivindicadora de l'educació de les dones
com a eina de l'alliberament femení i anticapitalista.

Aquesta obra es pot descarregar del **Dipòsit Digital de la UAB**:
https://ddd.uab.cat

Portada: fragment d'un enrajolament del pla amb *rajoles de Wang* (vegeu la pàgina 44) format per només 11 tipus de rajoles diferents, el conjunt de rajoles més petit que pot enrajolar el pla però no pot fer-ho de manera periòdica.

Índex

Preàmbul

Aquest llibre recull les lliçons que he impartit en una assignatura que es deia *Fonaments de les Matemàtiques*, a primer curs del grau de matemàtiques i del doble grau de matemàtiques i física a la Universitat Autònoma de Barcelona, en el període 2021–2024 però, de fet, el camí fins arribar a aquest llibre ha sigut força més llarg i potser va començar en uns cursos que vaig impartir a la Facultat de Filosofia i Lletres o en uns cursos sobre geometria axiomàtica que faig fer, fa moltíssims anys, al cinquè curs de la Llicenciatura de Matemàtiques. Tanmateix, en el rerefons d'aquest llibre hi podreu intuir algunes de les meves obsessions —en el millor sentit d'aquesta paraula— sobre les matemàtiques que a mi m'agraden i sobre la manera de fer, entendre i comunicar les matemàtiques que a mi m'agrada.

Quan el Departament de Matemàtiques va incloure —just en el primer any del pla d'estudis— l'assignatura obligatòria de *Fonaments*, l'objectiu que es perseguia anava molt més enllà de *donar servei* a les altres assignatures, en el sentit d'estalviar-los la feina d'haver de parlar, per exemple, de la demostració per inducció o per reducció a l'absurd, del concepte de funció injectiva, del signe d'una permutació, dels nombres complexos o del teorema xinès del residu. Aquest llibre parla d'aquests temes, evidentment, però els seus objectius són força més ambiciosos i consisteixen en voler transmetre als alumnes que emprenen els estudis de matemàtiques, d'una banda, quines són les grans idees transversals a totes les matemàtiques —simetria, estructura, equivalència, abstracció, generalització, analogia, computabilitat, per citar-ne unes poques— i d'una altra donar-los a conèixer quina és l'estructura pròpia del discurs matemàtic —com s'escriuen les matemàtiques— i ajudar-los en la difícil tasca d'adquirir el rigor imprescindible i la creativitat necessària del matemàtic professional.

D'acord amb aquests objectius, podríem dir que la tria de quins temes es tracten i quins no té una importància molt relativa: no ens preocupen gaire els continguts concrets —gairebé tots ells s'explicaran amb més profunditat en assignatures posteriors— perquè, si el que volem és transmetre les gran idees i els principis bàsics de com s'estructura el discurs, qualsevol tema ens pot ser útil per fer-ho. Tanmateix, cal matisar una mica aquesta afirmació i exposar alguns principis que he tingut en compte a l'hora de fer la tria de continguts:

- Crec que cal començar parlant de **lògica** i aquest llibre en parla amb més

profunditat del que és habitual en els primers cursos de grau. És cert que la majoria dels matemàtics hem après els fonaments de la lògica que utilitzem en la nostra feina sense haver seguit cap assignatura específica, però alguns de nosaltres pensem que en vam aprendre massa poca i la vam aprendre malament. Començar el camí amb una sòlida base lògica sense mistificacions, que no amagui els problemes, que no es redueixi a l'utilitarisme, em sembla imprescindible per al futur matemàtic.

- No he tingut cap dubte en incloure en aquest text un capítol sobre teoria de **grups**. Si estic dient que l'objectiu és mostrar a l'estudiant algunes de les idees centrals de les matemàtiques, seria imperdonable que l'estructura de grup, la més fonamental de totes, la més transversal, la que codifica la idea de simetria, la deixéssim per a un (hipotètic) curs posterior.

- He tingut molt en compte el vessant **cultural** extraordinari de les matemàtiques en la història de la humanitat. En conseqüència, he inclòs força temes que, sincerament, haurien de formar part del bagatge mínim de qualsevol persona culta: dels axiomes de Peano a la conjectura de Russell, del teorema dels nombres primers a la criptografia RSA, de la trisecció de l'angle a la fórmula d'Euler, i molts més.

- Finalment —ja ho veureu si us decidiu a llegir aquest llibre— m'he deixat guiar també per principis estètics i lúdics. He tingut la sort de trobar-me amb estudiants excel·lents, realment apassionats per les matemàtiques, que m'han permès gaudir enormement d'explicar el que em venia més de gust mentre ells, n'estic força segur, també experimentaven l'immens plaer —que els matemàtics coneixem prou bé!— del coneixement matemàtic clarament estructurat i rigorosament fonamentat.

L'autor d'aquesta obra que teniu a les mans —o, més probablement, a la pantalla— ja té més de setanta anys i, per tant, ha vist amb els seus propis ulls el canvi extraordinari que les matemàtiques i l'ofici de matemàtic han sofert en tots aquests anys. Ara hi ha matemàtiques arreu de la nostra vida, i la feina dels matemàtics és més diversa que mai. Ens podem preguntar, doncs, si l'ensenyament de les matemàtiques ha pogut adaptar-se a aquest canvi i si la immensa *tecnologia* matemàtica que ara ens envolta encara pot sentir alguna necessitat de llibres com aquest. La meva resposta a la primera pregunta és *«ens hem anat adaptant, però ho podríem haver fet millor»* i en aquest llibre he intentat deliberadament fer-ho millor. La meva resposta a la segona pregunta és *«sí, absolutament»*, i aquest és un tema que fins i tot he debatut a l'aula, amb els meus estudiants de primer curs. A diferència del que passava fa anys, quan hi havia una *via única* d'accés a les matemàtiques, ara els camins són molt diversos i molts d'ells no comencen pel que entenem com a *fonaments*. En canvi, les grans idees que estan presents en llibres com aquest segueixen sent fonamentals i fins i tot instrumentals en els desenvolupaments més actuals. En conclusió —i això és una intuïció meva que podria ser errònia— *tothom* que vulgui entrar en el món de les matemàtiques pot treure profit de **començar per aquí**.

Part I:

Lògica, nombre i estructura del discurs matemàtic

L'essència de la matemàtica rau en la utilització del que es coneix com a *raonament deductiu* en el qual el coneixement —la *veritat* matemàtica— només ho és quan es pot deduir a partir d'altres veritats —establertes prèviament o postulades explícitament— a través d'unes regles vàlides d'inferència. La lògica és, doncs, *el principi del principi* de les matemàtiques i aquest text, en conseqüència, ha de començar amb l'estudi de la lògica, ni que sigui la mínima quantitat de lògica imprescindible per a la fonamentació de la sintaxi del llenguatge de les matemàtiques.

Estudiarem la *lògica proposicional*, la part més elemental de la lògica, i seguirem amb una breu introducció a la *lògica de primer ordre*. Fins aquí, podríem estar estudiant matemàtiques o qualsevol altra disciplina basada en la lògica, però a continuació entrarem en la definició axiomàtica dels nombres naturals —els *axiomes de Peano*, d'una importància històrica immensa— i ens enfrontarem, per primera vegada, amb el vertigen de l'*infinit*, component inevitable de la matemàtica, i amb els *teoremes de Gödel* que posen límits insuperables a allò que podem demostrar.

També repassarem quina és l'estructura bàsica del discurs matemàtic: definicions, axiomes, teoremes, demostracions, contraexemples...

Foto: Gottlob Frege, 1849–1925

1 Lògica proposicional

La part més bàsica, més elemental, de la lògica és la **lògica proposicional**. Sobre ella es fonamenta tota la resta de la lògica, tot el raonament que anomenem *lògico-deductiu*, la matemàtica, la ciència, la filosofia, el dret, els llenguatges de programació, l'estructura dels ordinadors... De fet, hauríem de dir que s'hi fonamenta *la major part* de tot això, perquè hi ha també altres lògiques —intuïcionistes, constructivistes, polivalents, quàntiques, difuses,...— que s'aparten de la lògica proposicional que estudiarem ara. D'altra banda, també veurem més endavant que la lògica proposicional és clarament insuficient per desenvolupar tot el que acabem d'esmentar i caldrà anar més enllà amb sistemes lògics que inclouran la lògica proposicional però seran molt més potents.

Comencem, doncs, a parlar de lògica proposicional i la primera cosa que observem és que per parlar de lògica necessitem un llenguatge en el qual ens entenguem —i aquest llenguatge no pot ser el de la lògica proposicional perquè encara no el tenim. Distingirem, doncs, entre el llenguatge de la lògica proposicional i el *metallenguatge* que utilitzem per parlar de lògica proposicional. Admetre aquest fet és inevitable.[1]

La lògica proposicional —més exactament, la versió més simple de la lògica proposicional que estudiarem ara— conté aquests ingredients:

- **Proposicions**, que designarem amb lletres A, B, C... i suposarem que en tenim en quantitat il·limitada, en el sentit que sempre podrem trobar, si ens cal, una nova lletra que no haguem utilitzat abans. Si ens preguntem què són i què signifiquen les proposicions, la resposta és aquesta:

  > **La lògica no s'ocupa del significat dels seus termes (semàntica) sinó que només s'interessa per com aquests termes es poden encadenar —escriure'n un a continuació d'un altre— vàlidament (sintaxi).**

  En cada aplicació de la lògica proposicional a un camp de raonament concret, les proposicions seran frases d'aquest camp que tenen, com veurem més endavant, un valor de veritat/fals ben definit: *«8 és un nombre primer»,*

[1] No estem cometent cap incorrecció si comencem ara mateix a *numerar* les pàgines d'aquest llibre quan encara no hem *definit* els nombres naturals (ho farem a la pàgina 62).

«el daltonisme és hereditari», «el delicte de furt està tipificat a l'article 234 del codi penal», «Sòcrates és mortal».

- **Connectors lògics.** N'hi ha cinc:[2]

$$\wedge \qquad \vee \qquad \neg \qquad \Rightarrow \qquad \Leftrightarrow$$

Què signifiquen cadascun d'aquests connectors? La resposta és la mateixa d'abans:

> **La lògica no s'ocupa del significat dels seus termes (semàntica) sinó que només s'interessa per com aquests termes es poden encadenar vàlidament (sintaxi).**

Tanmateix, les regles d'utilització d'aquests connectors —que veurem més endavant— ens mostraran que funcionen, aproximadament, com les paraules

i, o, no, implica, si i només si,

respectivament.

- **Parèntesis**. En podríem prescindir (vegeu l'exercici **I**.9) però fan que les fórmules siguin més fàcils de llegir. A la pràctica, sovint s'ometen quan no hi ha risc de confusió.

- **Fórmules ben fetes** (FBF) que són successions finites de proposicions, connectors i parèntesis que segueixen aquestes tres úniques regles sintàctiques:

  1. Si A és una proposició, A és una FBF.
  2. Si ϕ i ψ són FBF, aleshores

$$\neg\phi \qquad (\phi \wedge \psi) \qquad (\phi \vee \psi) \qquad (\phi \Rightarrow \psi) \qquad (\phi \Leftrightarrow \psi)$$

     són FBF.
  3. Totes les FBF s'obtenen aplicant repetidament les dues regles anteriors.

Per exemple, $\neg(A \Rightarrow (B \vee \neg C))$ és una FBF però $\neg A \neg B \Rightarrow)$ no ho és. És possible decidir si una fórmula està ben feta o no.

[2] Podríem utilitzar-ne menys perquè alguns d'aquests es poden definir a partir dels altres, com veurem més endavant (exercici **I**.17), però tot és més clar si els utilitzem tots cinc. Per exemple, tota la lògica proposicional es podria escriure amb un únic connector, el que s'anomena, en el llenguatge de la informàtica, la *porta lògica* **NOR**, equivalent a $\neg(A \vee B)$. Cal observar que els dos últims connectors de la nostra llista de cinc s'escriuen, en els textos de lògica, en la forma $\longrightarrow$ i $\longleftrightarrow$, però a les matemàtiques el símbol $\longrightarrow$ indica una aplicació entre dos conjunts i és per aquest motiu que els matemàtics prefereixen representar els dos últims connectors lògics amb els símbols $\Rightarrow$ i $\Leftrightarrow$.

- **Veritat i Fals.** Cada proposició de la teoria pot tenir associat un *valor de veritat* —direm que tenim una *interpretació*— que pren només un dels dos valors *Veritat* (V) o *Fals* (F). Si ens preguntéssim què vol dir Veritat i què vol dir Fals en el context de la lògica proposicional, la resposta seria la mateixa d'abans: *«la lògica no s'ocupa del significat...»*. En cada aplicació concreta de la lògica proposicional —cada interpretació— Veritat/Fals tindran significats precisos.

- **Taules de veritat.** Un fet crucial de la lògica proposicional és aquest:

  > *A partir dels valors de veritat de les proposicions $A, B, C \ldots$ que apareixen en una FBF ϕ podem assignar, de manera unívoca i computable, un valor de veritat a ϕ.*

  Per arribar a aquest principi cal que fixem com es comporten cadascun dels cinc connectors respecte del valor de veritat de les proposicions que hi intervenen. Això es fa fixant les **taules de veritat** de cada connector. Aquestes taules de veritat ens assenyalaran, també, quina és la interpretació —*metalògica*, és clar— d'aquests connectors.

Ja tenim tots els ingredients de la lògica proposicional. Ara només ens resta dir quines són les taules de veritat dels cinc connectors.

$\neg\phi$

Aquest és el connector més senzill d'explicar perquè actua —fins un cert punt— com ho fa la paraula **«no»** en català:[3] $\neg\phi$ té el valor de veritat oposat al de ϕ. La taula de veritat és

ϕ	$\neg\phi$
V	F
F	V

$\phi \wedge \psi$

El funcionament d'aquest connector també és senzill d'explicar perquè actua de manera similar a la paraula **«i»**. La taula de veritat és

[3] Aquestes semblances que indicarem entre un connector lògic i una paraula d'un llenguatge natural —en aquest cas, el català— s'han d'utilitzar amb molt de compte. La riquesa i la flexibilitat dels idiomes naturals —i, en conseqüència, el seu nivell d'ambigüitat— són immenses i, per tant, aquestes semblances tenen només un limitat valor heurístic. És absurd pensar que cada vegada que pronunciem, per exemple, la paraula «no», podem substituir-la automàticament pel símbol lògic ¬. Com veurem, en alguns altres connectors la distància entre connector i paraula és encara més gran.

ϕ	ψ	$\phi \wedge \psi$
V	V	V
V	F	F
F	V	F
F	F	F

Per il·lustrar el funcionament correcte d'aquest connector en el context de la lògica podem utilitzar aquest acudit:

> *Tres estudiants de lògica entren en un bar. El cambrer els pregunta:*
> *—Cervesa per a tothom?*
> *El primer estudiant contesta:*
> *—No ho sé.*
> *El segon estudiant contesta:*
> *—No ho sé.*
> *El tercer estudiant contesta:*
> *—Sí!*
> *El cambrer, que també sap lògica, ho entén perfectament i serveix una cervesa a cadascun dels tres estudiants.*[4]

$\phi \Leftrightarrow \psi$

Aquest connector també té un funcionament senzill i força intuïtiu: $\phi \Leftrightarrow \psi$ és Veritat quan ϕ i ψ tenen el mateix valor de Veritat/Fals, i és Fals en cas contrari. Normalment, es tradueix per l'expressió **«si i només si»**. També es pot traduir com a **«ϕ és condició necessària i suficient per a ψ»**.

$\phi \vee \psi$

Aquest connector actua de manera similar a la paraula **«o»** però aquí cal anar amb compte[5] perquè la paraula «o» té diversos significats en el llenguatge ordinari —inclusiu, exclusiu, disjuntiu, alternatiu, equivalent[6]— i la seva utilització en aquests llenguatges naturals es presta a diverses interpretacions mentre que a la lògica —i, per tant, a les matemàtiques— el sentit de $\phi \vee \psi$ ha de ser sempre el mateix, clar i inequívoc. Simplificant, el «o» de les matemàtiques és sempre un «o inclusiu». La taula de veritat és

[4] He après aquest acudit a *An Introduction to Formal Methods for Philosophy Students*, de Thomas Forster.

[5] A l'exercici I.13 veurem un exemple on la paraula «i» es tradueix en el connector lògic $\vee$.

[6] Si busquem la conjunció «o» al diccionari de l'IEC hi trobarem només dos significats: (a) *expressa una alternativa* i (b) *denota equivalència*, i cap dels dos reflecteix exactament l'ús lògic del símbol $\vee$.

ϕ	ψ	$\phi \vee \psi$
V	V	V
V	F	V
F	V	V
F	F	F

Per il·lustrar el funcionament correcte d'aquest connector en el context de la lògica podem utilitzar aquest acudit:

> *Un estudiant de lògica va amb ascensor. L'ascensor s'atura a un pis, s'obre la porta, i una persona que esperava l'ascensor pregunta:*
> *—Puja o baixa?*
> *L'estudiant contesta:*
> *—Sí, evidentment!*

$\phi \Rightarrow \psi$

Finalment, aquest connector, que es llegeix **«ϕ implica ψ»** o també **«si ϕ, aleshores ψ»**, és el més problemàtic de tots perquè, si bé és cert que té un funcionament que s'assembla una mica al de la paraula *«implica»*, resulta que aquesta paraula s'utilitza en el llenguatge ordinari en diversos sentits que no tenen gaire a veure amb el sentit lògic del connector $\Rightarrow$. Aquí l'estudiant que comença a estudiar matemàtiques ha de ser molt curós. La taula de veritat és aquesta:

ϕ	ψ	$\phi \Rightarrow \psi$
V	V	V
V	F	F
F	V	V
F	F	V

Observem que si ϕ i ψ són Veritat, $\phi \Rightarrow \psi$ és també Veritat, cosa que no passa quan utilitzem la paraula *implica* en el llenguatge ordinari. Encara més, si ϕ és Fals, aleshores $\phi \Rightarrow \psi$ és Veritat **sigui quin sigui** ψ. Per il·lustrar la diferència que hi ha entre l'ús lògic de $\Rightarrow$ i l'ús de *implica* en el llenguatge natural fem aquestes consideracions:

- Imaginem dues persones: A que es diu Andreu i B que es diu Berta. Considerem aquestes afirmacions:
  - *(A es diu Andreu)* $\Rightarrow$ *(B es diu Berta)*. Això és cert a la lògica proposicional, però generalment no es considera que sigui cert en el llenguatge

ordinari. Segurament, si l'Andreu fes aquesta afirmació la Berta li contestaria que no, que ella no es diu pas Berta perquè ell es digui Andreu. El motiu d'aquesta discrepància és que, en el llenguatge ordinari, la paraula *implica* duu associada una certa connotació de causa/efecte **que el connector lògic $\Rightarrow$ no té**.

- *(A es diu David)* $\Rightarrow$ *(B es diu Berta)*. Això també és cert a la lògica proposicional, però generalment no es considera que sigui cert en el llenguatge ordinari.
- *(A es diu David)* $\Rightarrow$ *(B es diu Elisa)*. Això també és cert a la lògica proposicional, però no és clar que es consideri correcte en el llenguatge ordinari.
- Finalment, *(A es diu Andreu)* $\Rightarrow$ *(B es diu Elisa)* és Fals i aquí sí que, en el llenguatge ordinari, també es consideraria que és fals.

- Una bona manera d'entendre el sentit lògic de $A \Rightarrow B$ és imaginar-ho com un **contracte** en el qual la persona que afirma $A \Rightarrow B$ es compromet a verificar B si li donen una verificació de A. Posem un exemple: V (el venedor) diu a C (el comprador) que li pot vendre una bicicleta per 300€. Sigui

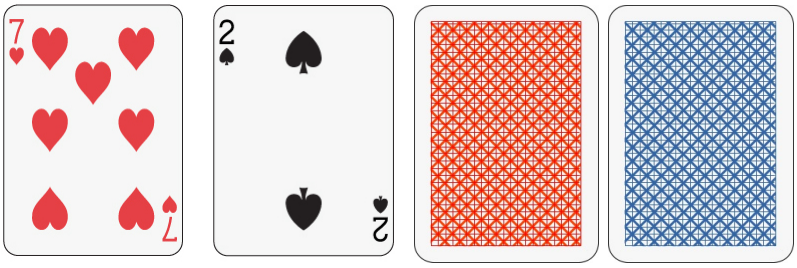
  A : C entrega 300€ a V.

  B : V entrega una bicicleta a C

  Quan V diu a C que li ven una bicicleta per 300€, està establint un contracte $A \Rightarrow B$. Què pot passar?

  1. Si C entrega 300€ a V i V entrega la bicicleta a C, s'ha complert el contracte $A \Rightarrow B$.
  2. Si C no entrega 300€ a V, s'ha complert el contracte $A \Rightarrow B$, faci el que faci V, és a dir, tant si li entrega la bicicleta (la regala) com si no l'hi entrega.
  3. Si C entrega 300€ a V, i V no entrega la bicicleta a C, en aquest cas **i només en aquest cas** no s'ha complert el contracte $A \Rightarrow B$.

- El psicòleg britànic Peter Wason (1924–2003) va crear un test relacionat amb la idea d'implicació que va esdevenir molt popular. En aquest test es mostren quatre cartes a una persona, d'aquesta manera:



  i es fa aquesta afirmació

Les cartes de número parell tenen el dors de color vermell.

Es demana que la persona decideixi si aquesta afirmació és certa o falsa donant la volta al mínim de cartes necessari per arribar al resultat demanat. L'estudiant de fonaments hauria de fer-se aquest test per saber del cert si realment ha interioritzat suficientment el significat lògic de $A \Rightarrow B$.

- Podem il·lustrar el fet que si A és Fals aleshores $A \Rightarrow B$ és Veritat, per a qualsevol B, amb aquesta anècdota:

  Un estudiant de lògica parla amb un amic que li diu que no entén com a partir d'una falsedat es pot deduir qualsevol cosa, i l'amic posa aquest repte a l'estudiant: «demostra'm que si $4 = 5$ aleshores jo sóc Napoleó.» *L'estudiant contesta:* «És molt senzill: si $4 = 5$, restant tres a cada costat tenim que $1 = 2$. Aleshores, com que Napoleó i tu sou dues persones, Napoleó i tu sou una persona.»[7]

Un cop fixades les taules de veritat dels diversos connectors lògics ja podem escriure la taula de veritat de qualsevol FBF i, per tant, si tenim assignats valors Veritat/Fals a cadascuna de les proposicions $A, B, C, \ldots$ que apareixen a una FBF ϕ, tenim assignat, de manera unívoca, un valor Veritat/Fals a ϕ.

Tautologies i contradiccions

En el context que estem ara, una **tautologia** és una FBF que té valor de Veritat, siguin quins siguin els valors de veritat de les proposicions que hi apareixen. Una **contradicció** és una FBF que té valor de Fals, siguin quins siguin els valors de veritat de les proposicions que hi apareixen. Calculant les taules de veritat, és molt senzill comprovar la validesa d'aquestes afirmacions:

- Si ϕ és una tautologia, $\neg\phi$ és una contradicció, i viceversa.
- Si ϕ és una tautologia, $\psi \Rightarrow \phi$ també ho és.
- Si θ és una contradicció, $\theta \Rightarrow \psi$ és una tautologia.
- $\phi \vee \neg\phi$ és una tautologia. Es tracta del famós *tertium non datur* —el «tercer exclòs»—, una veritat lògica històricament controvertida que, tanmateix, en el sistema formal de la lògica proposicional és una tautologia: té valor de Veritat per tota FBF ϕ. Aleshores, la seva negació $\phi \wedge \neg\phi$ és una contradicció.

[7] Tanmateix, aquesta història no s'adiu gaire amb el que estem explicant ara perquè l'estudiant de lògica, per tal de convèncer el seu amic, li mostra una *cadena de raonaments* que comença en A i acaba en B. Està bé com a mètode per convèncer un incrèdul, però aquesta cadena de raonaments és realment innecessària perquè el fet que si A és fals $A \Rightarrow B$ és cert és un principi de la lògica que no té res a veure amb cap cadena causal.

- $\phi \Leftrightarrow ((\phi \wedge \psi) \vee (\phi \wedge \neg\psi))$ és una tautologia.
- $\phi \Rightarrow (\psi \Rightarrow \phi)$ és una tautologia.
- $((\phi \Rightarrow \psi) \wedge \phi) \Rightarrow \psi$ és una tautologia. Es tracta de la traducció a la lògica proposicional del més famós dels sil·logismes clàssics: el *modus ponendo ponens*: «si ser home implica ser mortal i s'és home, aleshores s'és mortal».
- Donada una FBF, sempre podem decidir si és una tautologia, una contradicció, o cap de les dues coses (vegeu l'exercici I.16).

Fórmules equivalents

Dues FBF diem que són *equivalents* —escriurem $\phi \equiv \psi$— quan tenen la mateixa taula de veritat.[8] Algunes equivalències senzilles són aquestes:

- $\neg\neg\phi \equiv \phi$, la llei de la doble negació.[9]
- Les propietats de *commutativitat, associativitat* i *distributivitat* d'alguns connectors:

$$\begin{array}{l}(\phi \Leftrightarrow \psi) \equiv (\psi \Leftrightarrow \phi)\\ (\phi \wedge \psi) \equiv (\psi \wedge \phi)\\ (\phi \vee \psi) \equiv (\psi \vee \phi)\end{array} \qquad \begin{array}{ll}(\phi \wedge \psi) \wedge \rho \equiv \phi \wedge (\psi \wedge \rho) & \phi \vee (\psi \wedge \rho) \equiv (\phi \vee \psi) \wedge (\phi \vee \rho)\\ (\phi \vee \psi) \vee \rho \equiv \phi \vee (\psi \vee \rho) & \phi \wedge (\psi \vee \rho) \equiv (\phi \wedge \psi) \vee (\phi \wedge \rho)\end{array}$$

- Les famoses *lleis de De Morgan*, que usem constantment en la negació d'un «i» i en la negació d'un «o»:

$$\neg(\phi \wedge \psi) \equiv (\neg\phi \vee \neg\psi)$$

$$\neg(\phi \vee \psi) \equiv (\neg\phi \wedge \neg\psi)$$

  En particular, $A \wedge B \equiv \neg(\neg A \vee \neg B)$ i això ens demostra que podríem prescindir del connector $\wedge$.
- $(\phi \Rightarrow \psi) \equiv (\neg\psi \Rightarrow \neg\phi)$, la llei del contrarecíproc, de la que parlarem més endavant (pàgina 33).
- $(\phi \Rightarrow \psi) \equiv (\psi \vee \neg\phi)$. Aquesta equivalència ens diu que podríem haver prescindit del connector $\Rightarrow$ i és molt útil per, en cas de dubte, substituir el conflictiu $\Rightarrow$ per dos connectors més intuïtius com són $\vee$ i $\neg$.

[8] Podria semblar que $\phi \equiv \psi$ és el mateix que $\phi \Leftrightarrow \psi$, però són coses diferents. $\phi \Leftrightarrow \psi$ és una FBF però $\phi \equiv \psi$ no ho és perquè $\equiv$ no és un connector lògic vàlid. Afirmar $\phi \equiv \psi$ és el mateix que dir que la FBF $\phi \Leftrightarrow \psi$ és una tautologia.

[9] Aquesta llei de la doble negació, com hem dit que també passava amb la llei del tercer exclòs, no s'accepta a alguns tipus de lògiques.

- $(\phi \Leftrightarrow \psi) \equiv ((\phi \Rightarrow \psi) \wedge (\psi \Rightarrow \phi))$, que ens diu que també hauríem pogut haver prescindit de $\Leftrightarrow$ perquè és equivalent a la doble implicació.

Per demostrar que aquestes equivalències són correctes, n'hi hauria prou amb calcular les taules de veritat dels dos termes de l'equivalència i comprovar que coincideixen. Els matemàtics —els lògics, els informàtics, els filòsofs... tothom que usi diàriament les eines del raonament lògico-deductiu— tenen sempre al cap aquestes equivalències i les utilitzen constantment sense necessitat de pensar-hi. Al mateix temps, estan especialment entrenats en detectar immediatament les violacions de les normes de la lògica proposicional que van trobant al seu voltant.

Fem una observació final. Els matemàtics mai no fan servir,[10] en la seva pràctica professional, els símbols $\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$ sinó que usen les paraules i la sintaxi dels llenguatges naturals. Però el fet que les matemàtiques s'escriguin en llenguatges naturals no impedeix que, en l'estructura dels raonaments, es segueixin escrupolosament les normes de la lògica proposicional que estem estudiant —de manera elemental— en aquest capítol.

Reducció a l'absurd

Hi ha un cert mètode de demostració molt habitual a matemàtiques que és el que s'anomena *reducció a l'absurd*. En aquest tipus de demostració es comença admetent precisament el contrari del que es vol demostrar i es procedeix fins que s'arriba a una **contradicció**, de la qual (s'afirma que) es dedueix la conclusió que volíem. Trobarem molts exemples al llarg del curs.

La demostració per reducció a l'absurd s'ha comparat amb el *gambit* dels escacs que és aquella jugada on, per aconseguir un avantatge sobre l'oponent se li ofereix el sacrifici d'una peça. En la demostració per reducció a l'absurd el matemàtic, per aconseguir demostrar una certa conclusió, ofereix no una peça sinó *totes les matemàtiques* a canvi.[11]

L'esquema d'una demostració per reducció a l'absurd és aquest:

- Suposem que volem demostrar, a partir d'una hipòtesi A, una conclusió B. Per tant, la demostració comença dient *«suposem A»*.
- A continuació diem *«per reducció a l'absurd, suposem $\neg B$»*.

[10] Hi ha excepcions, és clar. Per exemple, en els treballs de lògica o a la pissarra o en un esborrany quan es vol guanyar temps, o en un cas on hi hagi un enunciat mol complicat i es vulgui evitar que la imprecisió present en els llenguatges naturals pugui dur a confusió.

[11] *«Reductio ad absurdum, which Euclid loved so much, is one of a mathematician's finest weapons. It is a far finer gambit than any chess play: a chess player may offer the sacrifice of a pawn or even a piece, but a mathematician offers the game.»* G.H. Hardy, *A Mathematicians Apology*.

- A partir de A i $\neg B$ es van traient conclusions fins que s'arriba a una contradicció. Per exemple, s'arriba a demostrar $C \wedge \neg C$.
- Un cop s'ha trobat una contradicció, diem que el teorema ja està demostrat.

Tornarem a parlar d'aquest mètode de demostració més endavant, però ara és el moment d'explicar que la validesa de la «demostració per reducció a l'absurd» emana de la lògica proposicional que acabem d'estudiar. Com? En el procés de reducció a l'absurd hem arribat a demostrar

$$(A \wedge \neg B) \quad \Rightarrow \quad (C \wedge \neg C).$$

Com és que d'aquí podem deduir $A \Rightarrow B$? Recordem que si D és falsa, $D \Rightarrow E$ és certa, sigui qui sigui E. Per tant, tenim

$$(A \wedge \neg B) \quad \Rightarrow \quad (C \wedge \neg C) \quad \Rightarrow \quad B.$$

Ara, pel principi del *tercer exclòs*, $B \vee \neg B$ és sempre cert, és una tautologia. Per tant,

$$A \Rightarrow A \wedge (B \vee \neg B) \Rightarrow (A \wedge B) \vee (A \wedge \neg B) \Rightarrow B \vee B \Rightarrow B.$$

És un fet important que la validesa de la demostració per reducció a l'absurd necessita el principi del tercer exclòs i, per tant, no és vàlida en sistemes lògics on no s'admeti aquest principi.[12]

[12] Quina part de les matemàtiques es pot establir sense usar mai la demostració per reducció a l'absurd, és a dir, sense admetre el *tertium non datur*? Aquesta és una pregunta molt rellevant que no podem abordar en un text elemental com aquest.

2 | Lògica de primer ordre

És evident que amb la lògica proposicional del tema anterior no en tenim prou per començar a fer matemàtiques —ni gairebé res més— i si volem fonamentar les matemàtiques cal ampliar la lògica proposicional a un àmbit molt més ric. Per exemple, analitzem aquest raonament sobre el nombre $c = 10223 \times 2^{31172165} + 1$:

A: *Tot primer de la forma* $4k + 1$ *és suma de dos quadrats (Fermat, 1640).*

B: c *és primer (Szabolcs* et al.*, 2016) de la forma* $4k + 1$.

C: *Existeixen enters* a, b *tals que* $c = a^2 + b^2$.

En el raonament matemàtic, és clar que $A \wedge B \Rightarrow C$ és cert, però només amb la lògica proposicional no tenim cap manera de demostrar-ho perquè, en l'àmbit de la lògica proposicional, les proposicions A, B i C no tenen cap relació entre elles. A l'enunciat anterior hi trobem termes que no apareixen a la lògica proposicional, com ara *tot, existeix, és*. Hi apareixen també **predicats**, és a dir, propietats P que, aplicades a un objecte x, donen proposicions $P(x)$ que són certes o falses.

La lògica de primer ordre[1] que estudiem en aquest capítol —de manera força més superficial de com hem estudiat la lògica proposicional, molt més senzilla— és l'àmbit on aquest tipus d'arguments que volem fer troben la seva expressió formal. La lògica de primer ordre també es coneix com a **lògica de predicats** i, a més de tot el que hi ha a la lògica proposicional conté, com veurem, predicats, quantificadors, variables, constants, etc. Per exemple, l'argument anterior, escrit el el llenguatge de la lògica de primer ordre, tindria aquesta forma:

$$A:\ \forall x \left(P(x) \Rightarrow Q(x)\right), \qquad B:\ P(c), \qquad C:\ Q(c)$$

on P seria el predicat *«és primer de la forma* $4k + 1$*»* i Q seria el predicat *«és suma de dos quadrats»*. Aleshores, efectivament, $(A \wedge B) \Rightarrow C$ seria cert segons les lleis d'aquesta lògica més potent. De fet, la immensa majoria del corpus de les matemàtiques es pot fonamentar en la lògica de primer ordre que és, per tant, el sistema lògic que utilitzen majoritàriament els matemàtics en el seu dia a dia. En conseqüència, l'aprenent de matemàtiques ha d'adquirir l'habilitat de poder

[1] El fet de dir-ne *de primer ordre* suggereix que hi ha una lògica més enllà d'aquesta, que seria *de segon ordre*. Efectivament, en parlarem una mica més endavant.

expressar tots els seus raonaments en el llenguatge de la lògica de primer ordre, si cal.[2] Donem un parell d'exemples de traducció d'afirmacions matemàtiques escrites en un llenguatge natural al llenguatge formal de la lògica de primer ordre:

- Considerem, en els nombres naturals, els predicats $P(x)$: x és primer; $S(x)$: x és senar. El fet que tot nombre primer diferent de dos és senar s'escriuria així:
$$\forall x \left((\neg(x = 2) \wedge P(x)) \Rightarrow S(x)\right).$$
- Considerem, en l'àmbit de les funcions reals, els predicats $C(f)$: f és contínua; $D(f)$: f és derivable. El fet que tota funció derivable és contínua però no tota funció contínua és derivable s'escriuria així:
$$\left(\forall x\,(D(x) \Rightarrow C(x))\right) \wedge \left(\exists y\,(C(y) \wedge \neg D(y))\right).$$
- L'exemple de l'inici d'aquest capítol seria una aplicació del *modus ponendo ponens*:
$$\forall y \left(\left(\forall x\,(P(x) \Rightarrow Q(x)) \wedge P(y)\right) \Rightarrow Q(y)\right).$$

Després d'aquests exemples, expliquem quins són els ingredients de la lògica de primer ordre.

- Tot el que ja tenim a la lògica proposicional segueix essent vàlid.
- **Quantificadors:** $\exists$ i $\forall$, amb el significat intuïtiu de *existeix* i *per a tot*, respectivament.[3]
- **Variables:** $x, y, z, \ldots$ en quantitat il·limitada.
- **Predicats:** $P(x), Q(x), \ldots$. També de diverses variables $H(x, y, z), \ldots$.
- **Constants.** Cada sistema formal basat en la lògica de primer ordre pot tenir les seves constants pròpies. Per exemple, *Sòcrates* seria una constant en els sil·logismes clàssics, 0 seria una constant en la teoria dels nombres naturals, etc.
- **Funcions** d'una o diverses variables, que assignen objectes de la teoria a altres objectes de la teoria.
- **Axiomes.** Cada sistema formal basat en la lògica de primer ordre tindrà els seus axiomes, és a dir, FBF a les que s'associa el valor Veritat. Implícita o explícitament sempre s'inclou un axioma que afirma que el domini del sistema no és buit (per exemple, amb l'axioma $\exists x\,(x = x)$).

[2] Diem *«si cal»* perquè, com ja hem explicat en el capítol anterior, les matemàtiques s'escriuen utilitzant, majoritàriament, els llenguatges naturals, però en qualsevol moment ha de ser possible transcriure el text al llenguatge de la lògica de primer ordre.

[3] Igual com passava amb la lògica proposicional, la utilització de *dos* quantificadors està motivada per la comoditat. N'hi hauria prou amb un de sol perquè $\forall x\, A \equiv \neg\exists x\, \neg A$.

- **Igualtat.** Els sistemes formals basats en la lògica de primer ordre tenen, generalment, un predicat d'igualtat (de dues variables) que escrivim $a = b$. S'han de complir aquestes propietats:
  - La igualtat ha de ser una *relació d'equivalència*. És a dir, ha de complir les propietats
    1. (Reflexiva) $\forall x\,(x = x)$.
    2. (Simètrica) $\forall x\,\forall y\,((x = y) \Rightarrow (y = x))$.
    3. (Transitiva) $\forall x\,\forall y\,\forall z\,(((x = y) \wedge (y = z)) \Rightarrow (x = z))$.
  - S'ha de complir la **llei de Leibnitz**, és a dir, la il·limitada substituïbilitat d'una cosa per qualsevol altra que sigui igual a ella:

    Si $a = b$ i ϕ és una FBF que contingui a, aleshores[4]

    $$\phi(\cdots a \cdots) \equiv \phi(\cdots b \cdots)$$

    És a dir, si $a = b$, podem substituir arreu a per b sense que s'alteri el valor de Veritat/Fals.[5]

Tots aquests ingredients d'una teoria formal de primer ordre han de complir unes normes sintàctiques que en aquesta introducció elemental que estem estudiant no explicitarem completament.

[4] Si volguéssim escriure aquesta llei en un llenguatge formal ens veuríem obligats a escriure una cosa semblant a aquesta

$$\forall\phi\,\big((a = b) \Rightarrow (\phi(\cdots a \cdots) \Leftrightarrow \phi(\cdots b \cdots))\big).$$

El problema està en que **aquesta expressió no és vàlida a la lògica de primer ordre**. La diferència fonamental entre la lògica de primer ordre i la lògica de segon ordre —que no estudiarem aquí— és que a la lògica de primer ordre els quantificadors només poden actuar sobre les variables del sistema i no sobre els predicats ni sobre les FBF. Si ens volem mantenir dins de la lògica de primer ordre la solució és substituir la llei de Leibnitz per una *successió infinita de lleis, recursivament enumerables*, uns conceptes en els que no entrarem, una llei per a cada FBF, tenint en compte que les FBF formen un conjunt recursivament enumerable. Es diu que la llei de Leibnitz no és un axioma sinó que és un *esquema d'axiomes* que representa, doncs, una quantitat numerable d'axiomes.

[5] Ara que parlem de la llei de Leibnitz que ens diu què «significa» la igualtat a matemàtiques, és un bon moment per fer aquest comentari. Tothom sap quines són les primitives de la funció x: $\int x = x^2/2 + C$. Suposem ara que tenim $x = t^3$, que és una suposició perfectament vàlida. Aplicant la llei de Leibnitz que, recordem-ho, representa una propietat essencial de la igualtat matemàtica, podem substituir x per t^3 a la integral anterior i obtenim $\int t^3 = t^6/2 + C$. Però també sabem que $\int t^3 = t^4/4 + C'$ i això ens porta immediatament a una contradicció. Què és el que hem fet malament? On és l'error? Certament, no en la llei de Leibnitz, que és un principi fonamental que està per sobre de tot. Certament, no en els càlculs que hem fet. On és l'error, doncs? La conclusió a què arribem és que escriure $\int x$ no és vàlid perquè contradiu la llei de Leibnitz i és per això que escrivim $\int x\,dx$, una notació, recordem-ho, deguda a Leibnitz. I ara ens adonem, també, que incloure dx no és una simple *notació*, sinó una cosa mot més fonamental: no integrem funcions sinó que integrem uns altres objectes matemàtics que s'anomenen *formes diferencials*.

Dos exemples senzills

Com a exemple de com és un sistema formal fonamentat en la lògica de primer ordre podem parlar de la **teoria de grups**. Aquests són els ingredients de la teoria de grups (notació additiva):[6]

- **Predicats**: cap (només la igualtat).
- **Constants**: una constant designada amb el símbol 0 que anomenarem *zero*.
- **Funcions**: una funció de dues variables $S(x, y)$ que escriurem $x + y := S(x, y)$ i anomenarem *suma*.
- **Axiomes**: tres axiomes:
  1. (element neutre): $\forall x \big((x + 0) = (0 + x) = x\big)$.
  2. (associativitat): $\forall x \forall y \forall z \big(x + (y + z) = (x + y) + z\big)$.
  3. (inversos): $\forall x \exists y \, (x + y = y + x = 0)$.

D'aquesta manera, la teoria de grups és un sistema formal dins de la lògica de primer ordre, amb un nombre finit d'axiomes.

Un segon exemple podria ser la teoria dels **conjunts parcialment ordenats**:

- **Predicats**: Un únic predicat de dues variables que denotem $x < y$.
- **Constants i funcions**: cap.
- **Axiomes**: dos axiomes:
  1. (antireflexivitat) $\forall x \, \neg(x < x)$.
  2. (transitivitat) $\forall x \forall y \forall z \Big(\big((x < y) \wedge (y < z)\big) \Rightarrow (x < z) \Big)$.

Sintaxi dels quantificadors

Hem dit que no explicitaríem totes les normes sintàctiques de la lògica de primer ordre, però sí que n'explicitarem algunes que s'utilitzen constantment i cal tenir ben clares:

- Si una FBF conté una variable x aquesta variable pot estar lligada a un quantificador o no. En el primer cas direm que x és una variable *lligada* i en el segon cas direm que és una variable *lliure*. Si una FBF conté variables lliures, es diu que és una fórmula *oberta* i la podem entendre com

[6] Parlarem d'aquest concepte de grup, força més a fons, en la tercera part d'aquestes notes.

un predicat perquè el seu valor de veritat dependrà dels valors assignats a les seves variables lliures. Els teoremes de les matemàtiques sempre són FBF *tancades*: totes les variables que hi apareixen estan lligades. Quan una variable està ligada, és intercanviable amb qualsevol altra variable:

$$\forall x P(x) \equiv \forall y P(y), \quad \exists x P(x) \equiv \exists y P(y).$$

- L'ordre dels quantificadors és important. D'una banda, dos $\forall$ o dos $\exists$ seguits es poden commutar

  $$\forall x\, \forall y\, P(x,y) \equiv \forall y\, \forall x\, P(x,y); \quad \exists x\, \exists y\, P(x,y) \equiv \exists y\, \exists x\, P(x,y).$$

  En canvi, $\forall$ i $\exists$ **no es poden permutar**

  $$\forall x\, \exists y\, P(x,y) \not\equiv \exists y\, \forall x\, P(x,y).$$

  Pensem, per exemple, en aquestes dues frases que tenen significats ben diferents:

  - *A cada partit hi ha un jugador que es lesiona.*
  - *Hi ha un jugador que es lesiona a cada partit.*

- La sintaxi dels quantificadors respecte de la negació es pot resumir en aquesta frase: *la negació d'un* $\forall$ *és un* $\exists$ *de la negació i viceversa*. Més formalment:

  $$\neg\forall x\, P(x) \equiv \exists x \neg P(x)$$
  $$\neg\exists x\, P(x) \equiv \forall x \neg P(x).$$

  Així, per exemple, la negació de *plou cada dia* és *hi ha algun dia que no plou*, i la negació de *hi ha algun bar obert* és *tots els bars estan tancats*.

- Algunes propietats distributives dels quantificadors respecte de $\wedge$ i $\vee$ es compleixen i altres no. Concretament:

  $$\forall x\, (P(x) \wedge Q(x)) \equiv \forall x\, P(x) \wedge \forall x\, Q(x)$$
  $$\exists x\, (P(x) \vee Q(x)) \equiv \exists x\, P(x) \vee \exists x\, Q(x)$$

  En canvi,

  $$\forall x\, (P(x) \vee Q(x)) \not\equiv \forall x\, P(x) \vee \forall x\, Q(x)$$
  $$\exists x\, (P(x) \wedge Q(x)) \not\equiv \exists x\, P(x) \wedge \exists x\, Q(x)$$

  Hi ha un cas particular on aquestes equivalències sí que es compleixen, que és quan una de les dues proposicions no depèn de x:[7]

  $$\forall x (P(x) \vee Q) \equiv \forall x\, P(x) \vee Q$$
  $$\exists x\, (P(x) \wedge Q) \equiv \exists x\, P(x) \wedge Q$$

[7] En aquestes fórmules és important haver postulat que el domini no és buit. Per exemple, l'equivalència $\exists x(P(x) \vee Q) \equiv \exists x P(x) \vee Q$ deixa de ser vàlida en un domini buit. Posem un exemple: suposem que el domini del sistema lògic són els unicorns, $P(x)$ és la proposició *«x és rosa»* i Q és la proposició *«els unicorns no existeixen»*. Aleshores, $\exists x P(x) \vee Q$ és Veritat però $\exists x(P(x) \vee Q)$ és Fals.

N'hi ha prou amb considerar qualsevol exemple concret per adonar-se que aquestes són regles sintàctiques que han de ser vàlides. Per exemple, considerem els predicats *elèctric* i *blanc*, referits als cotxes aparcats en una gran esplanada. Si suposem que hi ha com a mínim un cotxe aparcat a l'esplanada, estarem d'acord que

- Dir que tots són blancs i elèctrics és el mateix que dir que tots són blancs i tots són elèctrics.
- Dir que hi ha un cotxe que és blanc o elèctric és el mateix que dir que hi ha un cotxe que és blanc o hi ha un cotxe que és elèctric.
- Dir que tots són blancs o elèctrics no és el mateix que dir que tots són blancs o tots són elèctrics.
- Dir que hi ha un cotxe blanc i elèctric no és el mateix que dir que hi ha un cotxe blanc i hi ha un cotxe elèctric.

- A partir de les normes sintàctiques anteriors podem deduir propietats com aquestes:

$$\forall y\,(\forall x\,A(x) \Rightarrow A(y))$$
$$\forall y\,(A(y) \Rightarrow \exists x\,A(x))$$

Per exemple, la primera propietat es deduiria d'aquesta manera:

$$\forall y\,(\forall x\,A(x) \Rightarrow A(y)) \equiv \forall y\,(A(y) \vee \neg\forall x A(x)) \equiv \forall y\,A(y) \vee \neg\forall x A(x).$$

En la primera equivalència hem convertit $P \Rightarrow Q$ en $Q \vee \neg P$. En la segona, hem aplicat que si Q no conté la variable y, aleshores $\forall y(P(y) \vee Q)$ és el mateix que $\forall y P(y) \vee Q$. Finalment, la darrera fórmula és la llei del tercer exclòs perquè $\forall x P(x)$ és exactament el mateix que $\forall y P(y)$. La segona propietat es dedueix de la primera per la llei del contrarecíproc.

- Tornem ara a l'exemple de l'inici d'aquest capítol —la versió de primer ordre del *modus ponendo ponens*:

$$\forall y\,\Big(\big(\forall x\,(P(x) \Rightarrow Q(x)) \wedge P(y)\big) \Rightarrow Q(y)\Big).$$

En efecte, per la primera fórmula de l'apartat anterior

$$\forall y\,\Big(\big(\forall x\,(P(x) \Rightarrow Q(x))\big) \Rightarrow \big(P(y) \Rightarrow Q(y)\big)\Big).$$

Per tant,

$$\forall y\,\Big(\big(\forall x\,(P(x) \Rightarrow Q(x)) \wedge P(y)\big) \Rightarrow \big(P(y) \Rightarrow Q(y)\big) \wedge P(y)\Big).$$

Pel modus ponendo ponens de la lògica proposicional, tenim

$$\forall y\Big(\big(\big(P(y) \Rightarrow Q(y)\big) \wedge P(y)\big) \Rightarrow Q(y)\Big)$$

que, juntament amb la fórmula anterior, ens dóna la fórmula que volíem.

3 Els nombres naturals: axiomes de Peano i inducció

Es pot dir que les matemàtiques comencen amb els nombres naturals —i també amb la geometria elemental— i, per tant, quan parlem de fonaments de les matemàtiques, més d'hora o més tard cal abordar el tema dels nombres naturals: qualsevol fonamentació de les matemàtiques ha d'incloure l'aritmètica elemental dels nombres naturals. Però quan fem entrar els nombres naturals en el discurs lògic, obrim la porta a la **fascinació** i el **vertigen** de l'INFINIT, que entra a l'edifici que volem construir, de manera inevitable!

Algun matemàtic va dir que *les matemàtiques són la ciència de l'infinit*. En la sèrie

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

els primers 13 nombres, o els primers tretze mil milions de nombres, representen la part trivial de la sèrie; és en els *punts suspensius* on hi ha l'autèntic misteri: l'infinit. Per afirmar, per exemple,

$$3 + 1 = 1 + 3$$

potser no caldrien gaires fonaments lògics. Si, en canvi, volem demostrar que

$$n + 1 = 1 + n$$

per a tot n ja estem parlant d'una afirmació sobre tots els infinits nombres naturals o, si ho volem veure així, estem parlant d'una infinitat de teoremes, un per a cada valor de n, i ara sí que ens podem adonar que ens cal alguna base sòlida sobre la que puguem recolzar el vertigen d'aquestes infinites demostracions.

Per exemple, ens podem fer preguntes com aquestes:

a) $n + 1 = 1 + n$ per a **tots** els nombres naturals n?

b) Per **tot** n, existeix un primer $p > n$?

c) **Tot** nombre parell més gran que 2 és suma de dos primers?

L'afirmació b) es va demostrar fa mil·lennis (proposició 20 del llibre IX dels *Elements* d'Euclides, segle III aC.); c) encara no s'ha pogut demostrar si és cert o fals (conjectura de Goldbach, formulada el 1742 i encara no resolta); a) no es va creure que requeria demostració fins a finals del segle XIX. Fins i tot la idea que realment siguem capaços de demostrar coses com aquestes —en els dos primers exemples, ho hem pogut fet— té alguna cosa de meravella. Si, per exemple, la conjectura de Goldbach fos falsa, hi hauria un nombre parell que no seria suma de dos primers i podríem, en teoria, programar un ordinador perquè anés comprovant la conjectura, de manera successiva, per a cada nombre fins que, amb total seguretat, arribarà un moment —potser ni nosaltres ni el nostre univers hi seran— que trobarà el contraexemple. Però si la conjectura és certa, una comprovació per ordinador mai no ho podrà demostrar per simple inspecció perquè mai no pot arribar a comprovar **tots** els nombres. Però la situació és més complicada que això: encara que existís, per a cada n, una demostració vàlida per aquest n, a partir d'aquestes infinites demostracions, totes (potser) diferents, com podem pensar que es pot arribar a construir una única demostració —de longitud finita, és clar!— que les englobi totes?

Durant segles i segles, l'existència i la consistència lògica dels nombres naturals i les seves propietats bàsiques —la suma, la multiplicació, l'ordenació— es van donar per suposades, indiscutibles. Kronecker, en una frase del 1886 que es va fer famosa, va dir *«Déu ens ha donat els nombres naturals, la resta de les matemàtiques és obra de l'home»*. No és estrany que els éssers humans, enfrontats a la idea d'infinit, acabin invocant alguna divinitat.

A mitjan segle XIX alguns matemàtics i alguns lògics van sentir la necessitat d'establir uns fonaments lògics sòlids per als nombres naturals. En aquell moment, algunes idees innovadores —com la teoria de conjunts de Cantor— van permetre desenvolupaments admirables, però la força d'aquelles noves eines també va comportar l'aparició, al si de les matemàtiques, d'algunes paradoxes, d'algunes contradiccions que van sembrar la llavor del dubte sobre tot l'edifici de les matemàtiques, començant per la seva primera base: $1, 2, 3, 4, \ldots\ldots\ldots$ Principalment, és clar, per la banda de la dreta: «$\ldots\ldots\ldots$».

La manera de fonamentar els nombres naturals hauria de ser a partir d'un *sistema formal dins de la lògica de primer ordre*. Això és el que va fer, entre d'altres, Giuseppe Peano el 1889. Si bé, actualment, la teoria dels nombres naturals s'inclou dins de la teoria de conjunts —ho estudiarem més endavant— els axiomes de Peano segueixen tenint un interès per ells mateixos, si més no com una fita històrica fonamental.

Els axiomes de Peano

Els axiomes de Peano dels nombres naturals —n'hi ha diverses versions, ja en parlarem més endavant— fonamenten tota la teoria en dos principis bàsics:

- El concepte de **successor** d'un nombre, és a dir, el que és una unitat superior.
- El **principi d'inducció**.

Fem ara una exposició de l'axiomàtica de Peano (diferent de la que va proposar Peano: vegeu l'apèndix de la pàgina 27):

- És un sistema formal dins de la lògica de primer ordre amb
  - Una única **constant**, denotada 0.[1]
  - Una única **funció**, denotada S. Traduirem $S(x) = y$ com *«y és el successor de x»*.
  - Un connector d'**igualtat**.
- Tres axiomes:
  - **Axioma 1:** 0 no és successor de cap nombre:
    $$\forall x \neg \big(S(x) = 0\big).$$
  - **Axioma 2:** Si dos nombres tenen el mateix successor, són iguals:
    $$\forall x \forall y \left(\big(S(x) = S(y)\big) \Rightarrow (x = y) \right).$$
  - **Axioma 3 [Principi d'inducció]:**[2] Per tot predicat P es compleix
    $$\big[P(0) \wedge \forall x \left(P(x) \Rightarrow P(S(x))\right)\big] \Rightarrow \forall x\, P(x).$$
    És a dir, si P és un predicat i
    1. $P(0)$ és cert,
    2. si $P(x)$ és cert, aleshores $P(S(x))$ també ho és,

    aleshores, podem afirmar que $P(x)$ és cert per a **tots** els nombres naturals.

És clar que els dos primers axiomes —necessaris, és clar— no invoquen directament l'infinit ni han de suscitar cap dubte. És el tercer axioma el que *fa entrar a la sala un elefant* immens o, millor encara, una infinitat d'elefants. És l'axioma que ens permet, fins a cert punt, «domesticar» l'infinit dels nombres naturals i

[1] Antigament es considerava que el primer nombre natural era 1, és a dir, es considerava que 0 no era un nombre natural. És important recalcar que començar els nombres naturals amb 0 o amb 1 és una decisió arbitrària absolutament intranscendent des del punt de vista dels fonaments de les matemàtiques. Simplement: actualment el natural més petit és el zero perquè moltes de les coses que fem a les matemàtiques —i també a la computació— esdevenen una mica més senzilles si ho fem així.

[2] Aquesta manera d'enunciar el principi d'inducció no forma part de la lògica de primer ordre. Recordem que a la lògica de primer ordre és lícit escriure *«per tot nombre natural»* però no és lícit escriure *«per tot predicat»*. Aquest delicat tema el discutirem més endavant.

es pot entendre, també, com si ens digués que si bé hi ha infinits nombres naturals, tampoc no n'hi ha *tants*, perquè a cadascun d'ells s'hi pot arribar *pas a pas*, començant pel zero i pujant d'unitat en unitat: no hi ha cap nombre natural *inaccessible* al mètode del *pas-a-pas*.

Observem que el principi d'inducció no es pot demostrar a partir dels altres dos axiomes de Peano. N'hi ha prou amb considerar el sistema format pels nombres racionals ≥ 0 amb $S(x) := x + 1$: es compleixen els axiomes 1 i 2 però no l'axioma 3.

El principi d'inducció s'ha anomenant alguna vegada *principi del dòmino*: si tenim una filera de peces de dòmino i estan prou pròximes cadascuna de la següent de manera que si cau una també cau la següent, tenim clar que fent caure la primera cauran **totes**. Aquesta és l'essència del principi d'inducció, una eina poderosa que ens permet tenir la confiança que no és impossible demostrar teoremes que involucrin tots els nombres naturals.

La primera utilització explícita del principi d'inducció s'atribueix a Blaise Pascal però els historiadors han trobat traces de la utilització d'aquest principi en autors antics com l'erudit hebreu del segle XIV Leví ben Guerson —que, per cert, va escriure un comentari a Euclides en el que «demostrava» el famós cinquè postulat— o, fins i tot, Euclides.

L'axioma d'inducció ens permet fer coses com aquestes:

- Demostrar teoremes *per inducció*. Veurem algun exemple a continuació, i molts exemples al llarg del curs.
- Definir funcions *per recursió*. En parlarem amb més detall més endavant.
- Definir la suma, la multiplicació i l'ordre dels nombres naturals i demostrar les seves propietats bàsiques. També ho estudiarem més endavant.
- Demostrar que tot conjunt de nombres naturals té un mínim. Equivalentment, si un predicat no és cert per a tots els nombres naturals, hi ha d'haver un nombre natural mínim per al que el predicat sigui fals.
- Demostrar que els nombres naturals són únics: si dos sistemes formals compleixen els axiomes de Peano anteriors, són essencialment iguals. Es diu que els axiomes de Peano són *categòrics*.

Demostració per inducció

La demostració per inducció és una aplicació directa de l'axioma d'inducció. Aquest tipus de demostració segueix aquest esquema:

1. Es tracta de demostrar un teorema del tipus *«per tot nombre natural n es compleix $P(n)$»*.

2. El primer pas de la demostració consisteix en demostrar que el teorema és cert per $n = 0$.

3. El segon pas consisteix en demostrar que si $P(n)$ és cert per un n, aleshores $P(n+1)$ també és cert. Es diu *«per hipòtesi d'inducció, suposem $P(n)$ cert»*.

4. Arribats en aquest punt, el teorema ja està demostrat.

Hi ha diverses variants d'aquest mètode. Entre elles, aquestes:

- Si volem demostrar que el teorema $P(n)$ és cert per $n \geq k$, podem fer-ho també per inducció, començant no amb $n = 0$ sinó amb $n = k$.

- A l'hora de demostrar $P(n+1)$ ens podem trobar que ens calgui utilitzar no només $P(n)$ sinó $P(k)$ per algun (o per tot) $k \leq n$. És fàcil veure que el principi d'inducció també ens garanteix que aquest mètode —conegut com principi d'inducció **forta** o **completa**— és vàlid:

$$\Big(P(0) \wedge \forall n \left((P(0) \wedge \cdots \wedge P(n)) \Rightarrow P(n+1)\right)\Big) \Rightarrow \forall n\, P(n).$$

  (Vegeu l'exercici **I**.30.)

- El mètode de demostració anomenat del *contraexemple minimal* es justifica per la inducció completa que acabem de comentar: si un cert predicat P és fals per algun n, hi ha d'haver un nombre natural n_0 que sigui un contraexemple minimal, és a dir $P(n_0)$ és fals però $P(k)$ és cert per tot $k < n_0$. Veurem un exemple més endavant i alguns altres exemples al llarg del curs.

- Un altre mètode de demostració que es basa en el principi d'inducció és el que es coneix com a mètode del *descens infinit*. Suposem que volem demostrar que una propietat P es compleix per a tots els nombres naturals. El mètode consisteix en demostrar que si és falsa per a un cert n també és falsa per algun $k < n$. Evidentment, és una reformulació del mètode del contraexemple minimal.

- Si tenim un algorisme que va generant nombres naturals cada vegada més petits $n_0 > n_1 > n_2 > \cdots$ el principi d'inducció ens permet assegurar que l'algorisme, en algun moment, generarà 0 i s'aturarà.

Exemple

Teorema: *$n! > 2^n$ per tot $n \geq 4$.*

Demostració per inducció. És evident que el teorema és cert per $n = 4$. Ara ens cal demostrar que el teorema és cert per a $n + 1$, suposant que sigui cert per a

n.[3] Això és senzill de fer. Tenim

$$(n+1)! = (n+1)\,n! > (n+1)2^n$$

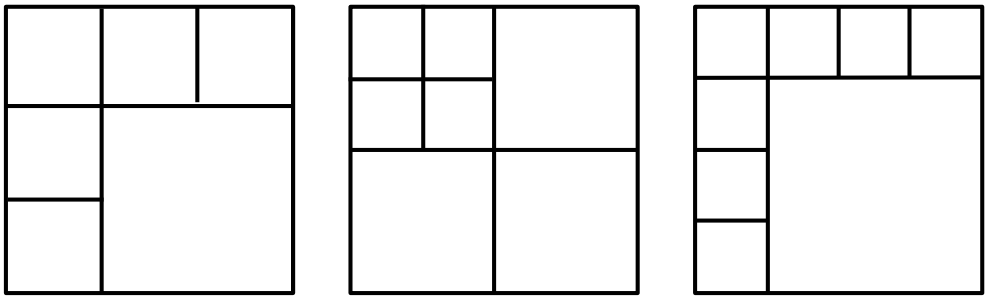
on la desigualtat la podem escriure perquè estem suposant $n! > 2^n$ *«per hipòtesi d'inducció»*. Aleshores, com que $n \geq 4$, tenim que $n+1 > 2$ i, substituint a la fórmula anterior, arribem a $(n+1)! > 2^{n+1}$ i la demostració s'ha acabat.

Exemple

Discutim ara un exemple una mica més atípic.

> **Teorema:** *Per tot $n \neq 0, 2, 3, 5$ podem subdividir un quadrat en n quadrats.*

Comencem observant que el teorema és trivialment cert per $n = 1, 4$. També és fàcil trobar solucions per $n = 6, 7, 8$:



Demostrem ara que el teorema és cert per tot $n > 8$ utilitzant el mètode del *contraexemple minimal* —que és una combinació de reducció a l'absurd i inducció. Suposem que $n_0 > 8$ fos un contraexemple minimal, és a dir, un quadrat no es pot dividir en n_0 quadrats però sí que es pot dividir en k quadrats per tot $8 \leq k < n_0$. En particular, podem dividir el quadrat en $n_0 - 3$ quadrats. Aleshores, sel·leccionem un d'aquests quadrats i el subdividim en quatre parts. D'aquesta manera, haurem subdividit el quadrat inicial en n_0 quadrats, una cosa que havíem dit que era impossible

[3] Suposar que el teorema és cert per a n sembla que sigui una *petitio principii*, és a dir, caure en la fal·làcia de suposar allò mateix que volem demostrar. No és així, de cap manera: no és el mateix demostrar $\forall n\, P(n)$ que demostrar $\forall n \left(P(n) \Rightarrow P(n+1)\right)$.

4 | Els nombres naturals: recursió

Ja hem dit que una de les conseqüències de l'axioma d'inducció és la capacitat de definir funcions de manera recursiva. Aquesta capacitat —que també és molt important en els llenguatges de programació— té una gran utilitat a totes les matemàtiques i, en el cas dels nombres naturals, és imprescindible per definir les operacions de suma i producte.

En primer lloc, enunciem explícitament el teorema d'existència i unicitat de funcions definides recursivament:

> *Donats un nombre natural n_0 i una funció $g(x, y)$ definida sobre els naturals, existeix una única funció f definida sobre els naturals que té aquestes dues propietats:*
>
> 1. $f(0) = n_0$.
> 2. *Per tot n, $f(S(n)) = g(n, f(n))$, on $S(n)$ denota el successor de n.*

Per exemple, la funció $n!$ per $n > 0$ es defineix habitualment així:

$$n! := n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1$$

i amb aquesta expressió tots entenem què volem dir. Però aquesta no és una definició formal perquè els punts suspensius no formen part del sistema formal dels nombres. Si volem una definició formal cal recórrer a fer una definició **recursiva** com aquesta:

1. $0! := 1$
2. $(n+1)! := (n+1)\, n!$

Aleshores, el teorema anterior ens assegura que la funció $n!$ està ben definida per **tot** nombre natural n.

Una primera conseqüència del teorema és la **unicitat** dels nombres naturals. Dit amb més precisió,

> *Si $(N, 0, S)$ i $(N', 0', S')$ són dos sistemes formals que compleixen els axiomes de Peano, hi ha una equivalència entre els dos.*

En efecte, podem definir per recursió una funció $f : N \to N'$ tal que $f(0) = 0'$ i $f(S(n)) = S'(f(n))$ i també una funció $g : N' \to N$ amb propietats anàlogues, inverses una de l'altra.

Sembla clar que la validesa de la definició per recursió hauria de ser una conseqüència de l'axioma d'inducció dels nombres naturals: tenim una funció definida per $n = 0$ i, si la tenim definida per un valor de n, també la tenim definida per $n + 1$. Segons la *idea del dòmino* sembla «evident» que la tenim definida per a tots els nombres naturals. La idea és correcta però la demostració no és senzilla i la deixarem per a un apèndix que es pot deixar de banda en una primera lectura.

Suma, producte i ordre als nombres naturals

Les operacions de suma i producte de nombres naturals es defineixen de manera recursiva. Per simplificar la notació escrivim $n' := S(n)$.

- **Suma.** És l'única funció que compleix, per cada m, aquestes lleis recursives:
  - $m + 0 = m$
  - $m + n' = (m + n)'$
- **Producte.** És l'única funció que compleix, per cada m, aquestes lleis recursives:
  - $m \cdot 0 = 0$
  - $m \cdot n' = (m \cdot n) + m$

El teorema d'existència i unicitat de les funcions definides recursivament ens garanteix que les operacions suma i producte estan unívocament definides per a tots els nombres naturals. A partir de l'existència d'aquestes dues funcions podem demostrar, per inducció, les seves propietats elementals (associativitat, commutativitat, distributivitat). Fem alguna d'aquestes demostracions, com a exemple.

- Per tot n, $0 + n = n$. Demostrem-ho per inducció. Per $n = 0$, és clar perquè sabem que $m + 0 = m$ per tot m. Suposem que $0 + n = n$ per un cert n i demostrem ara que $0 + n' = n'$. Però això també és clar per la definició recursiva de suma i per la hipòtesi d'inducció.
- Per tot n, m, $m' + n = (m + n)'$. Per inducció sobre n. Per $n = 0$:

  $$m' + 0 = m' = (m + 0)'.$$

  Suposem-ho cert per un n i demostrem-ho per al seu successor n':

  $$m' + n' = (m' + n)' = (m + n)'' = (m + n')'.$$

- Ara podem demostrar la propietat commutativa de la suma: per tot n, m, $m + n = n + m$. Fem-ho també per inducció sobre n. El cas $n = 0$ ja l'hem vist abans. Suposem que $m + n = n + m$ i demostrem que $m + n' = n' + m$. Aplicant la hipòtesi d'inducció i l'apartat anterior tenim

$$m + n' = (m + n)' = (n + m)' = n' + m.$$

La relació d'ordre als nombres naturals es defineix a partir de la suma:

$$n \leq m \quad \equiv \quad \exists k\,(n + k = m).$$

Tampoc no és difícil demostrar, a partir de les propietats de la suma, aquestes propietats bàsiques de la relació $\leq$:

1. (Transitivitat) $\forall n, m, k\,\big((n \leq m \wedge m \leq k) \Rightarrow n \leq k\big)$.
2. (Antisimetria) $\forall n, m\,\big((n \leq m \wedge m \leq n) \Rightarrow n = m\big)$.
3. (Totalitat) $\forall n, m\,(n \leq m \vee m \leq n)$.

* * * *

En els dos apèndix que vénen a continuació tractarem uns temes força tècnics i complicats que fan referència al teorema de recursió i a la versió de l'axiomàtica de Peano que vam estudiar en el capítol anterior —que ja vam dir que no era la mateixa que va considerar Peano el 1889.

Apèndix 1: demostració del teorema de recursió

Comencem demostrant el teorema que afirma que podem definir funcions de manera recursiva. La idea de la demostració sembla clara (recordem que n' indica el successor de n):

1. $f(0)$ està definit per $f(0) = n_0$
2. Si $f(n)$ està definit, també $f(n')$ està definit per $f(n') = g(n, f(n))$.
3. *Per tant*, pel principi del dòmino, $f(n)$ està definit per tot n.

El raonament ha de tenir aquesta forma, però la dificultat es troba en aconseguir formular *«està definit»* d'una manera formal. Els detalls són complicats.

En primer lloc, diguem que un predicat de dues variables $P(-,-)$ és **recursiu** si compleix

$$P(0, n_0) \wedge \forall a, b\,\big(P(a, b) \Rightarrow P(a', g(a, b))\big).$$

Clarament, la idea que hi ha al darrere d'aquesta definició és que si disposéssim ja de la funció f podríem definir un predicat P tal que $P(a, b)$ és cert si i només si $f(a) = b$. Aleshores, aquest predicat P seria recursiu, segon la definició que acabem de donar i, en certa manera, P seria el predicat recursiu *minimal*. A continuació definim un nou predicat $P_0(-,-)$ d'aquesta manera:

$P_0(x, y)$ és cert $\Leftrightarrow$
$P(x, y)$ és cert per tot P recursiu.

Una comprovació senzilla ens demostra que P_0 és recursiu.

Lema 1. *Per tot x existeix y tal que $P_0(x, y)$ és cert.*

La demostració també és senzilla, per inducció sobre x.

Lema 2. *Per tot x, l'element y del lema anterior és únic. És a dir*

$$\forall x \left(P_0(x, y) \wedge P_0(x, z) \Rightarrow y = z\right).$$

La demostració també es fa per inducció sobre x, però és una mica més complicada.

Primer pas: $x = 0$. Cal veure que $P_0(0, y) \Rightarrow (y = n_0)$. Suposem $y \neq n_0$ i considerem aquest predicat:

$$P_1(u, v) \text{ és cert } \Leftrightarrow$$
$$P_0(u, v) \text{ és cert i } (u, v) \neq (0, y).$$

Observem que $P_1(0, y)$ és fals. El punt essencial és que P_1 també és un predicat recursiu (demostració relativament senzilla). Per tant, per la manera com hem definit P_0, deduïm que $P_0(0, y)$ no pot ser cert i això acaba la demostració del cas $x = 0$.

Segon pas: suposem que el teorema és cert per $x = n$ i demostrem-lo per $x = n'$. La idea és la mateixa d'abans. Suposem que $P_0(n', m)$ i $P_0(n', s)$ són certs amb $s \neq m$ i definim un nou predicat recursiu P_2:

$$P_2(u, v) \text{ és cert } \Leftrightarrow$$
$$P_0(u, v) \text{ és cert i } (u, v) \neq (n', s).$$

Això ens duu a contradicció.

Per tant, amb aquests dos lemes anteriors ja podem **definir** la funció f que volíem: $f(x)$ és igual a l'únic y tal que $P_0(x, y)$ és cert. Ara cal veure que, amb aquesta definició, la funció f compleix el que volíem. Que $f(0) = n_0$ és clar. Si volem demostrar que $f(n') = g(n, f(n))$ ens cal demostrar això:

$$\forall n\, P_0(n', g(n, f(n))).$$

és a dir, ens cal demostrar que si P és un predicat recursiu qualsevol, aleshores

$$\forall n\, P(n', g(n, f(n))).$$

Aquesta demostració no és difícil, per inducció sobre n. Finalment, la **unicitat** de f també és fàcil de demostrar, per inducció. Això acaba la demostració.

Apèndix 2: els axiomes de Peano i la lògica de primer ordre

Quan vam començar a parlar de la fonamentació formal dels nombres naturals vam dir que voldríem tenir els nombres naturals com un sistema formal dins de la lògica de primer ordre. En canvi, els axiomes que hem enunciat **no estan formulats en lògica de primer ordre!** De fet, el problema es troba només al tercer axioma, l'axioma d'inducció. Aquest axioma, en la formulació que hem donat, comença dient *«per tot predicat P»* i això no és vàlid en la lògica de primer ordre, en la qual els quantificadors poden afectar objectes de la teoria, però no predicats de la teoria. *«Per tot nombre natural n»* és vàlid, però *«per tot predicat P»* no ho és.

En aquest apèndix —que també recomanem que, en una primera lectura, l'estudiant el deixi de banda— volem contestar aquestes preguntes:

- És possible reformular els axiomes de Peano en la lògica de primer ordre?
- Si formulem la teoria dels nombres naturals en la lògica de primer ordre, perdem alguna propietat important dels nombres?
- No seria possible fer una teoria dels nombres naturals que prescindís del delicat axioma d'inducció?

Si ens volem mantenir dins de la lògica de primer ordre podem utilitzar una estratagema similar a la que vam comentar quan discutíem la llei de Leibnitz. En l'enunciat de l'axioma d'inducció, quan diem *«per tot predicat P»* canviem això per *«per tota FBF ϕ»*. Què hi guanyem? Dir *«per tota FBF ϕ»* tampoc no és acceptable a la lògica de primer ordre, però ho és si pensem l'axioma d'inducció com un *esquema d'axiomes* que representa una *infinitat numerable d'axiomes*, un per a cada FBF ϕ. Hem de tenir en compte que les FBF formen una família numerable i es poden anar enunciant recursivament (exercici **I**.14).

En conclusió, hi ha uns axiomes de Peano de primer ordre, *amb una infinitat numerable d'axiomes*. Diguem-ne $\mathcal{P}_1$ d'aquesta teoria de primer ordre dels naturals, i diguem-ne $\mathcal{P}_2$ de la teoria de segon ordre amb els tres axiomes que hem vist anteriorment. Evidentment, l'axioma d'inducció de $\mathcal{P}_1$ és més feble que l'axioma d'inducció de $\mathcal{P}_2$ i això ens fa pensar que potser hi ha teoremes que es poden demostrar a $\mathcal{P}_2$ i no es poden demostrar a $\mathcal{P}_1$. Quins?

Potser seria útil tenir una idea intuïtiva de la diferència que hi ha entre l'axioma d'inducció de $\mathcal{P}_1$ i l'axioma d'inducció de $\mathcal{P}_2$. Pensem en el principi del dòmino. A $\mathcal{P}_2$, el principi del dòmino diu que qualsevol *«cosa»* que es transmeti d'un dòmino al següent, si li passa al primer dòmino, es transmet a **tots**. En canvi, a $\mathcal{P}_1$, el principi del dòmino diu que algunes *«coses»* (les FBF) que es transmetin d'un dòmino al següent, si li passen al primer dòmino, es transmeten a **tots**. Per exemple, imaginem que el fet que un dòmino faci caure el següent per efecte de la gravetat sí que es transmet a tots, però que hi ha altres coses —per exemple, una interacció electromagnètica— que encara que es transmeti de cada dòmino al següent, no arriba a tots. Això voldria dir que hi ha dòminos inaccessibles a la interacció electromagnètica.

A la demostració del teorema de recursió invoquem el principi d'inducció per a un predicat estrany P_0, la definició del qual no és vàlida a la lògica de primer ordre. Efectivament, el teorema de recursió **no es pot demostrar** amb l'axioma d'inducció de primer ordre.

Però la suma i la multiplicació es defineixen de manera recursiva! Resulta que la suma i la multiplicació no es poden definir amb els axiomes de $\mathcal{P}_1$ i, com que una teoria dels nombres naturals ha de contenir les dues operacions fonamentals, és imprescindible incloure la suma i la multiplicació entre les funcions de la teoria (a més de la funció S) i incloure com axiomes les seves fórmules recursives. Així és com ho va fer Peano.

Encara més: la demostració que, essencialment, hi ha uns únics nombres naturals és una conseqüència senzilla del teorema que ens permet definir funcions de manera recursiva. Com que aquest teorema falla a $\mathcal{P}_1$, podria passar que hi hagués diverses teories diferents dels nombres naturals, cosa ben estranya. Efectivament, això passa. Hi ha nombres naturals *«no-estàndard»*, és a dir, nombres naturals que compleixen tots els axiomes de Peano de primer ordre, però que no són cap dels nombres naturals *«clàssics»* 0,1,2,3,...[1]

En resum:

[1] L'existència d'aquests nombres naturals que no són cap dels 1,2,3,... no ens hauria de produir cap malestar perquè, si ens mantenim dins de la lògica de primer ordre, no els veurem mai. Seguint amb l'exemple del dòmino, imaginem que la gravetat és primer ordre i la resta d'interaccions —electromagnètica, quàntica— són segon ordre. Amb una axiomàtica de primer ordre, hi pot haver dòminos més enllà dels que ocupen els llocs 0,1,2,3,... però mentre no utilitzem res més que la gravetat, no notarem que hi són. De fet, la construcció de models no estàndard de l'aritmètica de Peano de primer ordre requereix l'axioma de l'elecció (vegeu la pàgina 64), que és com dir que no els podem construir.

- A la lògica de segon ordre, amb tres axiomes n'hi ha prou per definir, de manera categòrica, la teoria dels nombres naturals i les seves operacions aritmètiques.
- Podem fer una teoria dels nombres naturals que es mantingui dins la lògica de primer ordre, amb una infinitat numerable d'axiomes, incloent com axiomes l'existència de la suma i l'existència de la multiplicació. En aquest cas, la teoria deixa de ser categòrica: no hi ha un únic sistema de nombres naturals.
- El teorema de Ryll-Nardzewski de l'any 1952 afirma que és impossible axiomatitzar els nombres naturals amb una quantitat **finita** d'axiomes.

Fem un comentari final. Probablement, l'estudiant que hagi llegit aquests dos apèndixs haurà trobat que són excessivament tècnics i difícils, i potser s'estranyarà que un concepte tan bàsic com els nombres naturals requereixi aquestes delicades elucubracions. Per apaivagar aquests dubtes cal dir que la fonamentació estàndard de la major part de les matemàtiques —també dels nombres naturals— és la teoria de conjunts i que aquesta teoria, com veurem més endavant, permet *esquivar* les dificultats que hem trobat en aquest capítol en què hem volgut estudiar —per motius històrics i també conceptuals— l'aritmètica de Peano fora de la teoria de conjunts. Si més no, aquests apèndixs poden servir per fer més evident la «comoditat» que ens aporta la teoria de conjunts on, per exemple, és vàlid el teorema de recursió.

5 | El discurs matemàtic

Què fan els matemàtics? Com ho fan? Quina és l'estructura d'un text matemàtic? Què fa que les matemàtiques siguin tan singulars en tot l'àmbit del coneixement humà?

Els matemàtics fan moltes coses diverses. Per exemple, ensenyar i aprendre —aprendre sempre, tota la vida— i moltes altres coses com: modelitzar, intuir, definir, generalitzar, imaginar, abstreure, conjecturar, experimentar, classificar, afirmar, refutar... **comprendre** i **demostrar**: comprendre de manera profunda, i demostrar de manera irrefutable i eterna, com no ho fa cap altra ciència o cap altra activitat humana. En el fons, **crear** coneixement sòlid i perdurable.

Fer matemàtiques es troba a les antípodes de fer càlculs rutinaris o aplicar receptes estàndard per resoldre problemes estàndard. Un matemàtic s'enfronta cada dia a uns problemes que ni sap com atacar ni, probablement, disposa de les eines per fer-ho —crear-les, inventar-les, descobrir-les, forma part de la seva feina. Els matemàtics sempre intenten *fer allò que no saben fer*, allò que potser ningú ha fet abans i ningú no sap com fer.

La matemàtica —és una llàstima que això no sigui gaire conegut fora del món matemàtic— és una de les activitats més **creatives** que hi ha, i també una de les més exigents i implacables: un error, qualsevol error, ensorra tot l'edifici. Quan van preguntar a Hans Freudenthal quines virtuts havia de posseir un bon matemàtic, això és el que va contestar: energia, fantasia, autoconfiança i autocrítica.

El caràcter definitiu i immutable de les veritats matemàtiques ha seduït els pensadors des de les èpoques més antigues. A diferència de les altres ciències, el coneixement matemàtic de fa, diguem, dos mil anys, segueix conservant la seva plena validesa. Plató va dir que la matemàtica[1] tracta d'**allò que sempre és**.[2]

La fascinació per la immutabilitat i la certesa irrefutable de les veritats matemàtiques han influït profundament en tota la història del pensament i és in-

[1] De fet, Plató parla de la geometria.

[2] Als professors de matemàtiques se'ns demana, quan fem la guia docent, que utilitzem bibliografia *actualitzada* i aquesta demanda ens sembla ridícula perquè, per exemple, a la bibliografia d'una assignatura com *geometria lineal* hi apareixen els *Elements* d'Euclides (escrits el segle III aC) o el llibre *Grundlagen der Geometrie* de David Hilbert, escrit fa més de cent anys!

concebible[3] que un filòsof, un historiador o un humanista no conegui ni que sigui mínimament la influència històrica que han tingut els *Elements* d'Euclides, la teoria de conjunts, el problema del cinquè postulat, la resolució de la quíntica, la quadratura del cercle, les màquines de Turing, el problema del continu, la classificació dels grups simples o els teoremes de Gödel, per citar només uns pocs exemples d'extrema importància.

No cal insistir sobre la rellevància de les matemàtiques a l'obra de Plató —que afirma que cal exigir als governants de la república que sàpiguen matemàtiques—, d'Aristòtil o de Pitàgores. Ramon Llull —avançant-se enormement al seu temps— volia fonamentar el coneixement en una pura manipulació lògico-simbòlica; Spinoza va escriure la seva *Ètica a la manera dels geòmetres* perquè així li semblava que esdevenia incontrovertible; Wittgenstein va escriure el seu *Tractatus logico-philosophicus* en un llenguatge de «definicions» i «teoremes» manllevat de les matemàtiques, etc.

Aprendre a estructurar un discurs matemàtic de qualitat requereix, probablement, molts anys de treball, i en aquest capítol ens limitarem a indicar unes primeres pautes.

Definicions

En el seu aspecte més primitiu, donar una definició consisteix en assignar un nom —un nom *nou*, no usat abans per designar cap altra cosa— a un objecte matemàtic o a un predicat. Per exemple, en el context dels axiomes de Peano, l'objecte "dos", que es representa pel símbol "2" es defineix com el nombre natural successor del successor de 0:

$$2 := s(s(0))$$

i, en el context de les funcions reals, dir que una funció és *contínua* és una manera abreujada de dir que la funció compleix aquesta propietat:

$$f \text{ contínua } \equiv \forall x\, \forall \epsilon > 0\, \exists \delta > 0\, \forall y \left(|x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon \right)$$

de manera que, clarament, fem un bon estalvi d'esforç. És a dir, una definició apropiada en un moment apropiat ens ajuda a fer més clar el discurs i també a pensar amb més facilitat.

Però una definició pot ser molt més que això. La definició és una de les eines més poderoses que tenim els matemàtics —principalment els matemàtics. La llibertat que tenim a les matemàtiques per definir allò que vulguem de la manera que vulguem és il·limitada. Per exemple, res no ens impedeix definir, si volem,

$$2^0 := \pi$$

[3] Inconcebible no vol dir *impossible*.

però aquesta definició estúpida ens portaria a tenir fórmules massa complicades[4] mentre que, en canvi, la definició $2^0 := 1$ dóna lloc a fórmules molt més senzilles i naturals.

Una bona definició pot representar un avenç matemàtic important, pot ser un acte de geni com, per exemple, quan Cayley va definir el concepte de grup abstracte o quan Hausdorff va definir l'any 1914 el concepte d'espai topològic. Poincaré va intuir perfectament la idea de l'*homologia* d'una varietat en una sèrie de treballs genials publicats a partir de 1895, però la *definició* abstracta, axiomàtica, d'homologia que van donar Eilenberg i Steenrod cinquanta anys després va representar un pas endavant immens a l'hora d'entendre, clarificar i generalitzar les idees de Poincaré.

Teoremes

Un teorema —*lema*, *proposició*, *corol·lari*, etc.— és una afirmació matemàtica **que es pot demostrar**. L'existència de demostració és **l'únic** criteri de validació a les matemàtiques.

Un teorema es situa sempre dins d'un context formal —generalment basat en la lògica de primer ordre— en el qual hi ha uns objectes d'estudi —sobre la naturalesa dels quals no es diu res—, unes relacions entre aquests objectes i uns axiomes que compleixen els objectes i les relacions.

Per exemple, a la geometria d'Euclides els objectes són els punts i les rectes, les relacions són del tipus *"la recta r passa pel punt P"* i els axiomes són els cinc axiomes d'Euclides (*per dos punts diferents hi passa una única recta*, etc.). Si estudiem un grup abstracte tenim elements del grup $x, y, \ldots$, un element unitat 1 i una multiplicació $x \cdot y$, i els axiomes afirmen que la multiplicació compleix la llei associativa, 1 actua com a element neutre (és a dir $x \cdot 1 = 1 \cdot x = x$ per tot x) i per tot element x existeix un element y tal que $x \cdot y = y \cdot x = 1$. Si, en canvi, estudiem els nombres naturals, podem fer-ho sobre l'estructura formal donada pels axiomes de Peano.

Una **demostració** és una successió **finita** de proposicions $P_1, \ldots, P_n$ on P_n és el teorema que es demostra i cadascuna de les proposicions anteriors és un axioma, un teorema demostrat prèviament, una hipòtesi present a l'enunciat del teorema, o una proposició que es dedueix de les anteriors per alguna de les regles de la lògica.

Però, abans d'intentar demostrar un teorema, s'ha d'**enunciar**. D'on surten, doncs, els teoremes? En l'ensenyament, normalment, a l'estudiant se li proporciona l'enunciat del teorema i se li demana que trobi una demostració que el validi. A la vida real, en canvi, *trobar* quin és el teorema que cal demostrar és una part no gens trivial de l'activitat matemàtica, potser la més creativa i la més

[4] Penseu quina seria la fórmula per a 2^{n-m} si haguéssim definit 2^0 de la manera anterior

emocionant. Cal usar grans dosis d'imaginació, cal experimentar, fer-se preguntes, assajar exemples... és la part de les matemàtiques que més s'assembla a les ciències experimentals.[5]

Tipus de teoremes

La major part de les vegades, un teorema es redueix a una afirmació del tipus

$$A \Rightarrow B,$$

és a dir, A són unes hipòtesis i el teorema ens diu que, si aquestes hipòtesis es compleixen, també es compleix la conclusió B. Dit d'una altra manera, sense utilitzar el connector lògic $\Rightarrow$, diríem: *A és condició suficient per a B.*

Sobre aquest esquema podem considerar quatre variants:

- $A \Rightarrow B$, l'afirmació original.
- $B \Rightarrow A$, l'afirmació **recíproca** que també es pot enunciar dient que *A és condició necessària per a B.* L'afirmació original i la seva recíproca són ben diferents i un matemàtic és una persona especialment entrenada per no confondre-les. No és el mateix dir *tots els cotxes de bombers són vermells* que dir *tots els cotxes vermells són de bombers*.[6]
- $\neg B \Rightarrow \neg A$. Aquesta afirmació s'anomena la **contrarecíproca** de $A \Rightarrow B$ i té la important propietat de ser **equivalent a l'original**, com vam veure quan estudiàvem la lògica proposicional. Per exemple, si un matemàtic hagués de demostrar que *tots els corbs són negres* podria fer-ho demostrant que *totes les coses que no són negres no són corbs*.[7]
- $\neg A \Rightarrow \neg B$, que és el **contrari** de $A \Rightarrow B$ i el contrarecíproc de $B \Rightarrow A$.

Posem un exemple:

1. (original) Si plou, suspenem la sortida.

[5] Un exemple curiós és aquest. La successió de Fibonacci 0, 1, 1, 2, 3, 5, 8, 13,... es coneix i s'estudia des de fa segles, però no va ser fins el 1970 que el matemàtic Iuri Matiasévitx es va «adonar» que els nombres de Fibonacci F_n tenen aquesta propietat curiosa: $F_n^2 \mid F_m \Rightarrow F_n \mid m$. Un cop *descobert* aquest teorema, demostrar-lo no va ser gaire difícil. Aquestes investigacions sobre els nombres de Fibonacci van jugar un paper important en la resolució (negativa) del problema 10 de Hilbert per Matiasévitx i altres matemàtics.

[6] Algú va dir: *és més fàcil que un matemàtic surti al carrer amb les sabates canviades de peu que no pas que confongui $A \Rightarrow B$ amb $B \Rightarrow A$.*

[7] És clar que aquest mètode no és vàlid a les ciències experimentals! (*Paradoxa del corb o de Hempel.*) Aquesta és una de les grans diferències entre les matemàtiques i les ciències experimentals.

2. (recíproc) Si suspenem la sortida, és que plou.
3. (contrarecíproc) Si no suspenem la sortida, és que no plou.
4. (contrari) Si no plou, no suspenem la sortida.

Exemples i contraexemples

Podem demostrar un teorema amb un **exemple**? I refutar-lo amb un exemple? Depèn.

Una afirmació de la forma

tot x que compleix A també compleix B

no es pot demostrar amb un exemple, però si el volem **refutar** aleshores sí que n'hi ha prou amb posar un exemple d'un x que compleixi A i no compleixi B. Direm que hem trobat un *contraexemple*.

Considerem aquesta afirmació (Euler):

Per tot natural n el nombre $n^2 - n + 41$ és primer.

Si volem demostrar que això és cert, un exemple no aporta absolutament res a la demostració. Dir que $37^2 - 37 + 41 = 1373$ és primer no ens acosta a la demostració que $n^2 - n + 41$ sigui primer *per tot* n.[8] En canvi, si l'afirmació d'Euler és falsa (Euler ja sabia que ho era), aleshores sí que n'hi ha prou amb mostrar un contraexemple. Podria ser aquest:[9]

$$998^2 - 998 + 41 = 995047 = 197 \times 5051.$$

D'altra banda, un teorema de la forma

existeix un x que compleix A

queda demostrat si, simplement, donem un exemple.

Cal tenir molt clara aquesta situació que apareix constantment a la pràctica diària de les matemàtiques i que, malauradament, produeix força errors entre els principiants. Per demostrar que tots els habitants d'Enfesta (comarca del Solsonès) presenten peu grec hem de comprovar que tots i cadascun d'ells tenen peu grec; per refutar l'afirmació, n'hi ha prou amb mostrar un únic habitant d'Enfesta que no tingui peu grec. Són circumstàncies extraordinàriament diferents.

[8] Ni tampoc demostra res tenir un ordinador comprovant que $n^2 - n + 41$ és primer per n=1,2,3,... durant mil anys o mil milions d'anys!

[9] Hi ha un contraexemple molt més evident. El sabeu trobar?

Estructura d'una demostració (amb un exemple)

La pregunta *com es demostra un teorema?* no té resposta. No hi ha cap mètode per trobar una demostració —pensem que hi ha conjectures que no s'han convertit en teoremes (o s'hi ha trobat un contraexemple) fins molts anys (o segles) després d'haver sigut formulades. Però un estudiant de matemàtiques ha de tenir clara quina és l'**estructura** que hauria de tenir la demostració que busca. Això s'aprèn amb molts anys de pràctica, però aquí podem donar un primer exemple (n'hi haurà molts més al llarg del curs).

Suposem que volem demostrar aquest teorema:

Tota funció derivable és contínua.

Encara que no sapiguem demostrar el teorema, hi ha algunes coses que cal tenir clares. En primer lloc, cal intentar evitar aquests errors molt freqüents:

a) Mostrar un exemple. Pensar en exemples concrets pot ser útil a nivell heurístic o en un context docent, però no aporta absolutament res a una possible demostració del teorema.

b) Utilitzar com a vàlid allò que volem demostrar: *petitio principii* o raonament circular.

c) Utilitzar alguna propietat que precisament es demostra a partir del que volem demostrar. Aquest error pot arribar a ser subtil i difícil de reconèixer. Com exemples històricament importants tenim la llarga llista de «demostracions» del cinquè postulat d'Euclides, que sempre utilitzaven, en algun punt més o menys amagat de l'argument, una propietat de les rectes que només es pot demostrar amb el cinquè postulat.

d) Intentar demostrar el recíproc o el contrari del teorema que es proposa, pensant que es demostra el teorema original. Sempre podem substituir un teorema pel seu contrarecíproc, que és un teorema equivalent, però no pas pel contrari ni pel recíproc.

e) Utilitzar a la demostració termes que no han estat definits o *«inicialitzats»*. Aquest és un error molt freqüent entre els estudiants de matemàtiques més principiants. Per exemple, si a la demostració apareix un terme u, és imprescindible que, prèviament, s'hagi definit què és u —per exemple, $u := 2k+1$, on k s'ha definit anteriorment o apareix a les hipòtesis del teorema— o bé s'hagi deixat clar que u és un element genèric d'un determinat conjunt —per exemple: *sigui u un nombre real positiu*.

Vegem ara què podem saber sobre l'estructura de la demostració del teorema «*tota funció derivable és contínua*».

- La demostració no inclourà cap exemple. Dir, per exemple, que la funció $f(x) = x^2 + 1$ és derivable i també és contínua no ens aporta res a l'hora de demostrar que **totes** les funcions derivables són contínues.

- Tenim clar com començarà i com acabarà la demostració:

  > **Demostració:** *Sigui f una funció derivable, a continuació demostrarem que f és contínua.*
  >
  > * * *
  >
  > *Per tant, f és contínua, com volíem demostrar.* □

  ... ara hem d'omplir els asteriscs. Observem la presència de la paraula **sigui**, inevitable i ubiqua a totes les demostracions. Tal com hem dit a l'apartat anterior, la frase «*sigui f una funció derivable*» inicialitza f i, d'aquesta manera, ja podem utilitzar f a la demostració.

- Com que hem de demostrar que f és contínua, hem de començar repassant la definició de funció contínua.[10] Per tant,

  > **Demostració:** *Sigui f una funció derivable, a continuació demostrarem que f és contínua. Per demostrar que f és contínua hem de veure que es compleix això:*
  >
  > $$\forall x\, \forall \epsilon > 0\, \exists \delta > 0\, \forall y\, \big(|x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon\big)$$
  >
  > * * *
  >
  > *Per tant, f és contínua, com volíem demostrar.* □

- Ara veiem que hem de demostrar que es compleix una certa desigualtat complicada **per a tot** x **i per a tot** $\epsilon > 0$. Això ja ens diu alguna cosa més sobre com serà la nostra demostració:

[10] És a dir, encara que a l'enunciat del teorema la paraula derivable apareix abans de la paraula contínua, a la demostració hem de començar amb la definició de funció contínua. La definició de funció derivable ja la necessitarem més endavant. Un problema molt estès entre els principiants consisteix precisament en voler començar a demostrar $A \Rightarrow B$ treballant la hipòtesi A en lloc de la conclusió B. Il·lustrem això amb un exemple com el de la venda de la bicicleta (pàgina 7): has posat un anunci on hi diu que et vens una bicicleta per 200€. Ja hem comentat que això és una implicació del tipus $A \Rightarrow B$ on A és *em dónes 200€* i B és *et dono una bicicleta*. Imagineu que et truca una persona dient que està interessada en la bicicleta i demà et vindrà a veure per tancar el tracte. Què fas? No comences a treure conclusions sobre què són 200€, no comptes els diners, no calcules quants bitllets de 20 són.... el que fas, lògicament, és anar a buscar la bicicleta i mirar que estigui a punt per vendre. Quan arribi el client, li mostraràs la bicicleta, li mostraràs que és realment una bicicleta i no pas un paraigua fins que ell et digui que se la queda i aleshores, precisament aleshores, li exigiràs A, és a dir, que et doni els 200€. En una demostració matemàtica del tipus $A \Rightarrow B$, la majoria de vegades, hem d'actuar exactament igual. Hem de començar mirant què ha de passar perquè B sigui cert i *reservar-nos a la màniga* la hipòtesi A per al moment que la necessitem.

Demostració: *Sigui f una funció derivable, a continuació demostrarem que f és contínua. Per demostrar que f és contínua hem de veure que es compleix això:*

$$\forall x\, \forall \epsilon > 0\, \exists \delta > 0\, \forall y \left(|x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon \right)$$

Sigui x un nombre real qualsevol i sigui $\epsilon > 0$ un nombre real positiu qualsevol.

* * *

Per tant, f és contínua, com volíem demostrar. □

- A continuació hem de demostrar que existeix un $\delta > 0$ amb una certa propietat. Per tant, al llarg de la demostració haurem de ser capaços de trobar aquest nombre δ.

  Demostració: *Sigui f una funció derivable, a continuació demostrarem que f és contínua. Per demostrar que f és contínua hem de veure que es compleix això:*

  $$\forall x\, \forall \epsilon > 0\, \exists \delta > 0\, \forall y \left(|x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon \right)$$

  Sigui x un nombre real qualsevol i sigui $\epsilon > 0$ un nombre real positiu qualsevol.

  * * *

  *Definim $\delta :=$ * * **

  * * *

  Per tant, f és contínua, com volíem demostrar. □

- Ara hem de demostrar que per tot y es compleix $|x-y| < \delta \Rightarrow |f(x)-f(y)| < \epsilon$. Per tant,

  Demostració: *Sigui f una funció derivable, a continuació demostrarem que f és contínua. Per demostrar que f és contínua hem de veure que es compleix això:*

  $$\forall x\, \forall \epsilon > 0\, \exists \delta > 0\, \forall y \left(|x - y| < \delta \Rightarrow |f(x) - f(y)| < \epsilon \right)$$

  Sigui x un nombre real qualsevol i sigui $\epsilon > 0$ un nombre real positiu qualsevol.

  * * *

  *Definim $\delta :=$ * * *. Sigui y un nombre real tal que $|x - y| < \delta$. Demostrarem que $|f(x) - f(y)| < \epsilon$.*

  * * *

  Hem vist que $|f(x) - f(y)| < \epsilon$. Per tant, f és contínua, com volíem demostrar. □

- Observeu que encara no hem utilitzat la hipòtesi del teorema que ens diu que la funció f és derivable. Això és normal: les hipòtesis del teorema no apareixen a la demostració fins un cert moment en que trobarem que les necessitem. En aquest exemple que estem considerant, la hipòtesi l'haurem de fer servir per arribar a trobar una bona definició de δ.

Tenir clara aquesta estructura és una condició necessària per arribar a obtenir una demostració. Però, evidentment, no és suficient.

6 | Els teoremes de Gödel

Els famosíssims **teoremes de Gödel** (1930) posen límits a allò que es pot **demostrar** en determinats sistemes formals. La seva importància conceptual va ser immensa. D'alguna manera, ara aquests resultats que van ser revolucionaris han entrat al *domini públic* i la quantitat d'obres de divulgació que en parlen —no sempre prou acuradament— és molt gran. En tot cas, un (futur) matemàtic ha de tenir una visió d'aquests resultats més precisa i més tècnica del que ofereixen la majoria d'obres de divulgació que hem esmentat. Tanmateix, tampoc no podrem *anar fins el final* —cosa que requeriria més temps i més coneixements de lògica dels que tenim— però sí que intentarem donar, en aquest capítol, una visió prou correcta de què diuen exactament els teoremes i quines implicacions tenen.

D'entrada, ens situem en un sistema formal dins la lògica de primer ordre, amb una sèrie d'axiomes. Com hem vist en el cas dels axiomes de Peano de primer ordre, hem d'admetre la possibilitat de tenir *infinits axiomes*, però sempre exigirem que aquests infinits axiomes es puguin enunciar algorísmicament, és a dir, hi hagi un algorisme per anar-los enunciant un darrera l'altre (no volem ser més precisos en aquest aspecte tècnic).

Podem pensar, per exemple, en aquests casos:

- Els nombres naturals amb els axiomes de Peano de primer ordre.
- La teoria de conjunts amb els axiomes de Zermelo-Fraenkel (que veurem més endavant).
- La teoria de grups.
- La geometria d'Euclides amb els seus 5 axiomes (convenientment formalitzats).
- Els nombres reals amb els axiomes que estudiem a càlcul (en la seva versió de primer ordre).
- La teoria de les *màquines de Turing* com a model formal de la teoria de la computació.

Quan tenim una teoria d'aquestes ens podem fer dues preguntes crucials: és consistent? És completa? Discutim què vol dir això.

Consistència d'una teoria

Una teoria diem que és **consistent** quan no hi ha contradiccions, és a dir, quan és **impossible** demostrar $A \wedge \neg A$.[1]

Una teoria **inconsistent** no té cap utilitat perquè ja sabem que si tenim demostrada una contradicció, automàticament *totes* les proposicions són demostrables. Recordem que, per tot B, es compleix que $A \wedge \neg A \Rightarrow B$.

Si, desenvolupant una teoria, arribem a un punt on hem demostrat una contradicció, l'única sortida possible és *debilitar* els axiomes que hem posat.

Exemple. Suposem que en els axiomes dels nombres reals canviem aquest axioma (existència d'inversos)

$$\forall x \neq 0 \; \exists y \; : xy = 1$$

per aquest altre axioma més fort:

$$\forall x \; \exists y \; : xy = 1.$$

La teoria que obtindríem és una teoria contradictòria: aplicant l'axioma a $x = 0$, existiria y tal que $0 = 0y = 1$. Però $1 > 0$ i, per tant, $0 > 0$, en contradicció amb un altre axioma dels nombres reals.

Exemple. Quan Cantor va crear la teoria de conjunts va posar com a axioma que per tot predicat P hi ha un conjunt A format pels x tals que $P(x)$ és cert:

$$\forall P \, \exists A \, \forall x \, (x \in A \Leftrightarrow P(x)).$$

Bertrand Russell va observar que, amb aquest axioma, la teoria de conjunts és inconsistent:

> **Paradoxa de Russell.** *Apliquem l'axioma de Cantor al predicat $x \notin x$: existirà un conjunt A tal que*
>
> $$\forall x \, (x \in A \Leftrightarrow x \notin x).$$
>
> *Si substiuïm x per A, obtenim*
>
> $$A \in A \Leftrightarrow A \notin A$$
>
> *que és una contradicció (observem que $(P \Leftrightarrow \neg P) \equiv (P \wedge \neg P)$). Per tant, la teoria de conjunts de Cantor és contradictòria.*

Per resoldre aquesta paradoxa —que, evidentment, invalidava la teoria de conjunts de Cantor— Russell va desenvolupar la *teoria de tipus* (que no estudiarem) i Zermelo i Fraenkel van reemplaçar l'axioma de Cantor per un axioma més feble:

[1] Recordem que $A \wedge \neg A$ és sempre Fals, però aquí estem parlant de si és demostrable o no és demostrable. És a dir, en una teoria consistent no hi ha cap successió de deduccions lògiques que comenci amb els axiomes i arribi a un teorema que digui $A \wedge \neg A$.

Per a tot conjunt U i per a tot predicat P existeix un conjunt A format pels $x \in U$ tals que $P(x)$ és cert.

En resum, una teoria només és útil si és no contradictòria i, per tant, seria fonamental poder demostrar que les teories que utilitzem a les matemàtiques, principalment les més bàsiques sobre les que es construeix tot l'edifici —els nombres naturals, la geometria elemental, l'aritmètica, la teoria de grups, la teoria dels nombres reals...— estan lliures de contradicció. La pregunta crucial de si l'aritmètica és contradictòria o no ho és apareix al segon lloc de la famosa llista de 23 problemes que David Hilbert va proposar l'any 1900 en el Congrés Internacional de Matemàtics de Paris:

Problema 2 de Hilbert (1900): *demostrar que la teoria dels nombres naturals segons els axiomes de Peano no és contradictòria.*

Teories completes i proposicions indecidibles

Una teoria diem que és **completa** si, per a qualsevol proposició Φ, existeix una demostració de Φ o existeix una demostració de $\neg\Phi$. Com que en una teoria contradictòria totes les proposicions es poden demostrar, les teories contradictòries són completes, d'una manera trivial. Per tant, la pregunta si una teoria és completa o no ho és només té interès en les teories consistents.

Si ni Φ ni $\neg\Phi$ es poden demostrar (en una certa teoria formal), direm que Φ és **indecidible**. L'existència de proposicions indecidibles en una teoria formal no és cap cosa estranya. Posem un parell d'exemples:

- **La teoria de grups.** Recordem que la teoria de grups és un sistema formal amb una funció $+$ (utilitzem per exemple, notació additiva), una constant 0 i tres axiomes: associativitat, element neutre i existència d'inversos. Considerem, en aquesta teoria, la proposició següent

$$\forall x\, \forall y \;\; x + y = y + x,$$

  és a dir, la llei commutativa de la suma. És evident que ni aquesta proposició ni la seva negació es poden demostrar a partir dels axiomes de grup, perquè $\mathbb{Z}$ és un grup commutatiu i les simetries d'un triangle equilàter formen un grup no commutatiu.

- **El cinquè postulat d'Euclides.** Durant segles, molts matemàtics van intentar demostrar el cinquè axioma d'Euclides a partir dels altres quatre, cosa que, si fos possible, demostraria que Euclides va afegir un axioma superflu a la seva teoria de la geometria. A mitjans segle XIX es va construir una geometria que complia tots els axiomes d'Euclides menys el cinquè. Per tant, el cinquè postulat és indecidible a partir dels altres quatre.

Dit això, també hem de dir que hi ha certes teories que, pel seu caràcter fonamental, *sembla* com si haguessin de ser completes. Per exemple, l'aritmètica elemental, representada pels axiomes de Peano. És imaginable que hi hagi una proposició que parli només de nombres naturals, sumes i multiplicacions i que ni ella ni la seva negació siguin demostrables? És possible que hi hagi una proposició de geometria elemental que sigui indecidible a partir dels cinc axiomes d'Euclides (convenientment perfeccionats)? Resulta difícil pensar que hi pugui haver indecidibles en aquestes àrees tan clàssiques, tan fonamentals.

David Hilbert va expressar repetidament l'opinió que, més aviat o més tard, podríem resoldre afirmativament o negativament totes les conjectures i aquesta opinió de caràcter més aviat filosòfic s'ha interpretat en el sentit de dir que Hilbert creia que no hi podia haver proposicions indecidibles a l'aritmètica. Explícitament, a la seva famosa llista de problemes hi trobem:

> **Problema 10 de Hilbert (1900):** *trobar un algorisme que permeti decidir si qualsevol equació diofàntica té solució o no en té.*[2]

Els teoremes de Gödel que explicarem a continuació van demostrar que aquesta pretensió o desig de Hilbert era irrealitzable.

El primer teorema de Gödel

Recordem que estem parlant de sistemes formals $\mathcal{F}$ amb una quantitat *«raonable»* d'axiomes (com hem comentat abans). El primer teorema de Gödel diu:

> **Primer teorema de Gödel (1931):** *Si $\mathcal{F}$ és consistent i prou potent com per contenir l'aritmètica dels nombres naturals, aleshores $\mathcal{F}$ no és complet.*

És a dir,

> *No sabem si l'aritmètica dels naturals està lliure de contradicció, però si ho està, aleshores no és una teoria completa: hi ha alguna proposició sobre els nombres naturals que ni ella ni la seva negació es poden demostrar. Hi ha proposicions* ***indecidibles****.*

[2] Una equació diofàntica és una equació $q(x_1, \ldots, x_n) = 0$ on q és un polinomi en n variables, amb coeficients enters, i només ens interessen les solucions $x_1, \ldots, x_n$ que siguin nombres enters. Un dels exemples més famosos d'equació diofàntica és l'equació de Fermat $x^n + y^n = z^n$. El problema de demostrar que aquesta equació diofàntica, per $n > 2$, no té cap solució en els nombres enters $\neq 0$, va ser durant molts anys un dels problemes oberts més famosos de les matemàtiques.

Fem algunes observacions sobre aquest teorema. Suposem que la teoria dels nombres naturals és consistent (si no ho és... *tanquem la paradeta*: la matemàtica no té res més a dir).

- Sabem que a l'aritmètica hi ha alguna proposició indecidible, diguem-n'hi F. Això deu voler dir que hem sigut massa poc exigents a l'hora d'escriure els axiomes de la teoria. Afegim F als axiomes i tenim una nova teoria en la qual F ja és demostrable. Però... la nova teoria segueix complint les hipòtesis del teorema de Gödel i, per tant, seguim tenint alguna proposició indecidible, diguem-n'hi F'. Podem afegir aquesta nova proposició als axiomes, però ja veiem que així no anirem enlloc: a totes les teories que anirem obtenint encara els podem aplicar el teorema de Gödel i mai no ens podrem desempallegar de les proposicions indecidibles!

  En el moment que hem donat entrada als nombres naturals, a la suma i a la multiplicació —i els hem de deixar entrar perquè sense ells no hi ha matemàtiques— hem d'acceptar que sempre hi haurà un límit a allò que podrem demostrar, sempre hi haurà proposicions indecidibles amb les que podem ensopegar en qualsevol moment de la nostra activitat com a matemàtics. Les teories matemàtiques que incloguin els nombres naturals no poden ser, simultàniament, consistents i completes.

- Ara que sabem que hi ha proposicions indecidibles[3] ens podem preguntar quina probabilitat tenim de trobar-nos-en una algun dia. Per als estudiants de matemàtiques o d'informàtica la resposta és: la probabilitat és del 100%; al llarg dels estudis us aniran explicant que tal o qual afirmació és indecidible. En aquest mateix text en trobarem alguna, més endavant.

  La pregunta és més complicada si afegim la condició que la proposició indecidible tingui un enunciat molt senzill dins de l'aritmètica més elemental. És a dir, voldríem explicar un problema indecidible a *«una persona del carrer»* que només tingui una formació matemàtica bàsica. Aquests exemples podrien servir:

  - **La conjectura de Collatz** és el problema més senzill d'explicar i més difícil de resoldre que hi ha. El 1937 a un postdoc anomenat Lothar Collatz se li va ocórrer aquesta pregunta d'aspecte trivial:

    > *Prenem un nombre natural n qualsevol; si és parell, el dividim per 2; si és senar, el multipliquem per 3 i li sumem 1; repetim el procés indefinidament. Podem demostrar que, en tots els casos, arribarem al nombre 1?*

    Ningú ha resolt aquesta pregunta ni ningú ha demostrat encara (2024) que aquest problema sigui indecidible, però hi ha veus autoritzades[4]

[3] Quan parlem de proposicions indecidibles sempre hauríem de completar la frase dient en quin sistema formal són indecidibles.

[4] Vegeu l'article *On Unsettleable Arithmetical Problems* de John H. Conway a The American Mathematical Monthly, Vol. 120, No. 3 (March 2013), pp. 192–198.

que veuen molt probable que ho sigui. De fet, sí que s'ha demostrat que el problema general de decidir si aquest tipus de problemes tenen solució o no és indecidible.

- **El problema 10 de Hilbert** que hem esmentat abans, sobre un algorisme per decidir si les equacions diofàntiques tenen solució, es va demostrar el 1970 que és indecidible (Matiasévitx, Robinson, Davis, Putnam): *es pot construir una equació diofàntica que no té cap solució, però no es pot demostrar que no en té cap.*[5]
- Fora de l'àmbit de l'aritmètica —però molts d'aquests problemes estan relacionats entre ells— el problema indecidible més famós és el **problema de l'aturada d'una màquina de Turing**. No tenim temps d'entrar en detalls. Diguem que aquest resultat el podem expressar (aproximadament) en aquesta forma: *és impossible crear un programa d'ordinador que analitzi els programes i ens digui si s'aturaran o entraran en un cicle*.
- Evidentment, quan coneixem una proposició indecidible F, ja sabem que qualsevol proposició la demostració de la qual ens porti a una demostració de F, ha de ser també indecidible. Per exemple, el famós *joc de la vida* de Conway o els *dòminos (o rajoles) de Wang* presenten preguntes que, si tinguessin resposta, també tindria resposta el problema de l'aturada d'una màquina de Turing i, per tant, són indecidibles.

Idea de la demostració del primer teorema de Gödel

La demostració del teorema de Gödel no és ni molt llarga ni molt difícil, però aquest text no és el lloc indicat per fer-la amb detall. Ens limitarem, doncs, a donar una idea heurística —molt aproximada— de per on va la solució.

Comencem observant que els llenguatges naturals són tan rics que no és gens difícil fer afirmacions que duen immediatament a contradicció. Posem alguns exemples clàssics:

- **Paradoxa del mentider**: *Tot el que dic és fals*. (S'atribueix a Epimenides, filòsof del segle VI aC.)
- **Paradoxa del barber**: *El barber afaita tothom que no s'afaita ell mateix. Qui afaita el barber?*

[5] Aquí sembla que hi hagi una contradicció entre que diguem que no té cap solució i que diguem que no podem demostrar que no hi ha cap solució. Aquesta contradicció no hi és. Els autors citats demostren que existeix una equació diofàntica $q(x_1, \ldots, x_n) = 0$ tal que la proposició *«l'equació no té solució»* no es pot demostrar i la proposició *«l'equació té solució»* tampoc no es pot demostrar. Per tant, l'equació no té solució perquè si $(x_1, \ldots, x_n)$ fos una solució, automàticament existiria una demostració de *«l'equació té solució»* consistent en substituir els valors $(x_1, \ldots, x_n)$ a l'equació i veure que surt 0.

- *Els guardes d'un pont pregunten a tothom on va, i pengen a la forca els qui menteixen. Arriba una persona i quan li pregunten on va contesta que va a que el pengin a la forca. Què han de fer els guardes?* (Don Quijote de la Mancha, segona part, capítol 52.)
- *Pinotxo diu: ara em creixerà el nas. Què passa a continuació?*
- **Paradoxa de Protàgores i el seu deixeble Euatle**: *El deixeble va acordar amb el seu mestre que li pagaria les lliçons de dret quan guanyés el seu primer cas. No va pagar perquè no es va dedicar a fer d'advocat i Protàgores el va demandar perquè pagués el seu deute. Què podia passar en el judici?*
- **Paradoxa de Grelling-Nelson**: *diem que un adjectiu és heterològic si no s'aplica a ell mateix: monosil·làbic i proparoxíton son heterològics; polisil·làbic i agut no ho són. És heterològic l'adjectiu «heterològic»?*
- *Marca la resposta correcta d'entre aquestes dues: (a) la resposta correcta és la (b); (b) la resposta correcta és la (a).*

Veiem que aquestes paradoxes —i moltes més que hi ha— es basen en frases que són *autoreferents* i són conegudes des de fa molt de temps. No ens ha de sorprendre,[6] doncs, l'existència de sentències indecidibles en els llenguatges naturals. Per exemple:

> **«Teorema»:** *aquest teorema és indecidible.*

és un teorema —de manera informal— clarament indecidible (en una teoria consistent). El que Gödel va fer va ser demostrar, d'una manera molt enginyosa, que una proposició semblant a aquesta es pot arribar a formalitzar utilitzant la lògica de primer ordre i l'aritmètica de $\mathbb{N}$.

El segon teorema de Gödel

Recordem que el problema 2 de Hilbert demanava demostrar que l'aritmètica de Peano és una teoria consistent, és a dir, lliure de contradicció. El segon teorema de Gödel diu que la demanda de Hilbert no es pot dur a terme:

> **Segon teorema de Gödel.** *Si $\mathcal{F}$ és un sistema formal consistent com el del primer teorema, la proposició «$\mathcal{F}$ és consistent» és indecidible.*

[6] Després de tants segles de la paradoxa del mentider, no ens hauria de sorprendre que una frase autoreferent sigui paradoxal, però encara ho fa. El 2021 va circular per les xarxes socials una versió més elaborada d'aquestes paradoxes i va generar força debats. Representava un concursant d'un programa de preguntes i respostes que havia de contestar aquesta pregunta: *Si respons aquesta pregunta a l'atzar, quina probabilitat tens d'encertar la resposta correcta? (a) 25%, (b) 0%, (c) 50%, (d) 25%.*

És a dir, o bé l'aritmètica és inconsistent —i, per tant, absolutament inútil— o bé és consistent —com tothom creu i espera— i, en aquest cas, no hi ha cap demostració que ho demostri.

* * * *

Havíem començat innocentment dient «0, 1, 2, 3,.....» i ara comencem a notar el vertigen del món infinit on hem entrat. A mida que avancem en l'estudi de la matemàtica, aquest vertigen —aquesta fascinació— encara creixerà il·limitadament.

Apèndix: més detalls sobre la demostració del teorema de Gödel

La idea que hem donat sobre la demostració del primer teorema de Gödel és força superficial i és probable que hi hagi estudiants que vulguin tenir una idea més acurada i més tècnica —sense necessitat d'arribar a fer tots els detalls. Com hem dit, es tracta de formular una FBF que, d'alguna manera, *afirmi la seva pròpia indemostrabilitat*. Vegem com es pot arribar a fer.

El **primer pas** consisteix en convertir totes les expressions de la teoria formal en *nombres naturals*. És a dir, codifiquem, d'alguna manera efectivament computable, cada símbol lògic, la constant 0, la funció successor, el signe de sumar, el signe de multiplicar, els quantificadors,... utilitzant nombres naturals. Pel que fa a les variables, observem que es poden codificar totes amb dos nombres: un nombre per a una variable x i un nombre per a un símbol $'$ que ens permeti crear totes les altres variables: $x', x'', x''', \ldots$ A continuació —per exemple, amb un nombre que actuï com a *separador*— podem també assignar un nombre natural a cada fórmula ben feta (FBF) i un nombre a cada demostració. Tot això es pot fer de diverses maneres i la conclusió d'aquest procés és que tindrem una funció **injectiva** i computable $\mathcal{G}$ que assigna a cada FBF i a cada demostració un nombre natural $\mathcal{G}(-)$ que s'anomena el seu *nombre de Gödel*.[7]

En el **segon pas** definim un predicat $Q(n, m)$ de dues variables (que són nombres naturals) que és Veritat quan passen aquestes tres coses:

- $m = \mathcal{G}(F)$ per una certa FBF F (única, és clar).
- F té una *variable lliure* x. És a dir, podem escriure $F(x)$.
- Si $\mathcal{G}(D) = n$, aleshores D **no** és una demostració de $F(m)$.

Observem, doncs, que hem definit un predicat Q de manera que, si $F(x)$ és una FBF amb una variable lliure, l'expressió formal[8]

$$\forall n\, Q(n, \mathcal{G}(F(x)))$$

significa, precisament, que *«no hi ha cap demostració de $F(\mathcal{G}(F(x)))$»*.

A continuació, observem que

$$H := \text{''}\forall n\ Q(n, x)\text{''}$$

és una FBF amb una variable lliure x. Per tant, aquesta fórmula tindrà un nombre de Gödel, diguem-n'hi k:

$$k := \mathcal{G}(H) = \mathcal{G}(\text{''}\forall n\ Q(n, x)\text{''}).$$

[7] Aquest procés, que en temps de Gödel devia semblar força sorprenent, en el temps de la revolució digital ja ens sembla molt més natural.

[8] Quan escrivim $\forall n$ estem utilitzant la hipòtesi del teorema de Gödel que diu que el sistema formal que tenim inclou la teoria dels nombres naturals.

Finalment, sigui U aquesta FBF sense variables lliures:[9]

$$U := \text{"}\forall n\ Q(n,k)\text{"}.$$

Segons hem dit abans, U s'interpreta com *«no hi ha cap demostració de $H(k)$»*. Però $H(k) = U$ i, per tant, U s'interpreta com *«no hi ha cap demostració de U»*. Ja hem aconseguit, doncs, el que ens proposàvem: trobar una expressió formal que afirmi la seva pròpia indemostrabilitat.

Com que partim de la hipòtesi que la nostra teoria formal no és contradictòria, és clar que la FBF U que afirma, precisament, que ella mateixa no és demostrable, no és demostrable. Per acabar la demostració del primer teorema de Gödel hem de veure que $\neg U$ tampoc no és demostrable. Però $\neg U$ és

$$\neg U = \text{"}\exists n_0\ \neg Q(n_0,k)\text{"}$$

i si tinguéssim una demostració de $\neg U$, tindríem una demostració de $\neg Q(n_0,k)$ per un cert n_0 i això, per la manera com hem definit Q, vol dir que n_0 seria el nombre de Gödel d'una demostració de $H(k) = U$, cosa que ja sabem que és impossible.

En tot això que hem dit hi ha detalls que requeririen molta més precisió, però creiem que amb l'explicació anterior l'estudiant ja pot tenir una idea força exacta de com es demostra el primer teorema de Gödel. La idea de fons —expressada d'una manera molt informal— és que l'aritmètica dels nombres naturals és tan potent que permet formular-hi la paradoxa del mentider.

[9] Observeu com estem jugant perillosament —però correctament!— amb l'*autoreferència*!

Exercicis de lògica i nombres naturals

1. Quines de les proposicions següents són la negació de l'afirmació «la solució és 2 o 3»: (a) Ni 2 ni 3 no són la solució; (b) La solució no és 2 o no és 3; (c) La solució no és 2 i no és 3.

2. Escriviu la negació de les proposicions següents:

   (a) Dues rectes diferents sempre es tallen en un únic punt.

   (b) Hi ha un polinomi a coeficients enters que no té arrels reals o, si en té, són totes positives.

   (c) A tots els municipis hi ha alguna persona tots els fills de la qual no han tingut ni el xarampió ni la rubèola.

   (d) Per tot nombre real a existeix un nombre real x tal que per tot nombre real y es compleix que $y > x$ implica $1 < y - a$.[10]

   (e) Si $\sqrt{2}$ és racional, jo sóc Juli Cèsar.

   (f) Si $\sqrt{2}$ és irracional, jo sóc Juli César.

   (g) L'alarma sonarà si s'obre la porta i el botó d'anul·lació no es prem, o si hi ha moviment i no succeeix que el botó d'anul·lació es prem o l'alarma no està activada.

3. Decidiu si les proposicions (a), (b), (d), (e), (f) de l'exercici anterior són certes o falses.

4. Comproveu si aquests raonaments són lògicament correctes:

   (a) Perquè jo dugui el paraigües és necessari que plogui. Quan plou, mai no duc sandàlies. Avui duc sandàlies. Per tant, no està plovent i en conseqüència no duc el paraigües.

   (b) Si baixen els tipus d'interès, la borsa pujarà. Si els tipus d'interès no baixen, aleshores la construcció i el consum privat baixaran. Ara, el consum privat no està baixant. Per tant, és cert que la construcció no està baixant o el consum privat no està baixant. És a dir, és fals que la construcció i el consum privat estiguin baixant. Això vol dir que els tipus d'interès estan baixant i en conseqüència la borsa pujarà.

[10] Aquest exercici i algun altre com el 4.b estan trets del *Cours d'Algèbre* de Roger Godement (1962). La seva inclusió aquí és un petit homenatge a aquell llibre que tant va inspirar els qui van iniciar-se en les matemàtiques al voltant de l'any 1971, en particular l'autor d'aquestes notes.

5. Enuncieu el recíproc, el contrari i el contrarecíproc del teorema de Pitàgores. Quins d'aquests teoremes són certs a la geometria ordinària?

6. Considereu aquestes dues proposicions: (A) 3 és parell; (B) El polinomi x^2+x+1 no té cap arrel real. Considereu les proposicions: A implica B; A implica (no B); (no A) implica B; (no A) implica (no B). Quines d'aquestes quatre proposicions són certes i quines són falses?

7. Escriviu les taules de veritat d'aquestes fórmules de lògica proposicional:

   (a) $A \Rightarrow (B \Rightarrow A)$.

   (b) $\neg(\neg A \vee A)$.

   (c) $(A \vee (B \Leftrightarrow A)) \Rightarrow \neg(C \wedge \neg B)$.

8. Decidiu quines d'aquestes fórmules són tautologies, quines són contradiccions i quines no són ni una cosa ni l'altra:

   (a) $A \Rightarrow (A \Rightarrow A)$.

   (b) $(A \Rightarrow A) \Rightarrow A$.

   (c) $(A \Rightarrow \neg A) \Leftrightarrow (\neg A \Rightarrow A)$.

   (d) $(A \wedge B) \Rightarrow (C \Rightarrow A)$.

   (e) $(A \Rightarrow B) \vee (B \Rightarrow A)$.[11]

   (f) $\big((C \Rightarrow A) \Rightarrow (A \vee B)\big) \vee \big((C \Rightarrow B) \Rightarrow (B \Rightarrow A)\big)$.

   (g) $\big(A \Rightarrow (B \Rightarrow C)\big) \vee \big((A \Rightarrow B) \Rightarrow (A \Rightarrow C)\big)$.

9. (*Lògica proposicional sense parèntesis*) Si modifiquem les regles sintàctiques de la lògica proposicional podem prescindir dels parèntesis —a canvi de fer les expressions molt menys intel·ligibles. La idea consisteix en escriure tots els connectors com ho fem amb el de la negació: abans de les proposicions a les que afecta. Així, escrivim $\vee AB$ en lloc de $(A \vee B)$, $\wedge AB$ en lloc de $(A \wedge B)$, $\Rightarrow AB$ en lloc de $(A \Rightarrow B)$ i $\Leftrightarrow AB$ en lloc de $(A \Leftrightarrow B)$. Per comprovar que realment les dues sintaxis són equivalents, traduïu d'una sintaxi a l'altra aquestes FBF (i decidiu si són tautologies):

   (a) $\Big[\big((A \Rightarrow X) \wedge (B \Rightarrow X)\big) \wedge \big((A \Rightarrow Y) \wedge (B \Rightarrow Y)\big)\Big] \Leftrightarrow \big((A \vee B) \Rightarrow (X \vee Y)\big)$.

   (b) $\Leftrightarrow \Rightarrow A \Rightarrow BC \Rightarrow \wedge ABC$.

10. Utilitzeu el càlcul proposicional per resoldre aquesta endevinalla clàssica de Lewis Carroll:

    (a) Cap poema interessant és impopular entre les persones de bon gust.

[11] És fàcil veure que $(A \Rightarrow B) \vee (B \Rightarrow A)$ és una tautologia, i això pot semblar sorprenent o, fins i tot, absurd. Estem dient que, sempre, una implicació o la seva recíproca són certes? Per exemple, o bé *parell* implica *primer* o bé *primer* implica *parell*?!? Sembla que no, però la solució d'aquest aparent contrasentit es troba quan ens adonem que ara estem en el context de la lògica proposicional i no en el de la lògica de primer ordre. El que tenim clar és que ni tot nombre parell és primer ni tot nombre primer és parell, però això no contradiu el fet que $(A \Rightarrow B) \vee (B \Rightarrow A)$ sigui una tautologia, perquè a la lògica proposicional no hi ha el concepte de *«tot nombre»*. A i B són proposicions com, per exemple, «3 és parell» i «3 és primer» i, certament, la proposició *«3 és parell implica 3 és primer, o 3 és primer implica 3 és parell»* és certa.

(b) Cap poema modern està lliure d'afectació.

(c) Tots els poemes que vostè ha escrit parlen de bombolles de sabó.

(d) Cap poema afectat és popular entre les persones de bon gust.

(e) Cap poema antic parla de bombolles de sabó.

Quina conclusió podeu treure d'aquestes cinc premisses?

11. En un examen de càlcul es demana demostrar que la funció sinus és derivable. Un estudiant contesta *«com que* $\sin 0 = 0$*, això implica que* $\sin(x)$ *és derivable»* i el professor de càlcul li diu que la seva demostració és incorrecta, però l'estudiant afirma que a la classe de fonaments li han explicat que aquesta implicació és correcta. Comenteu la situació.

12. Utilitzeu la lògica proposicional per resoldre aquesta endevinalla de Lewis Carroll:

(a) Els animals s'ofenen si no els faig cas.

(b) Tots els meus animals estan en aquest camp.

(c) Els animals no poden entendre situacions complicades si no han estat ben entrenats.

(d) Cap dels animals que hi ha en aquest camp és un toixó.

(e) Quan un animal s'ofèn, arrenca a córrer udolant.

(f) No faig cas de cap animal que no em pertanyi.

(g) Cap animal que hagi estat ben entrenat arrenca a córrer udolant.

13. Considereu aquestes afirmacions:

(a) Al casino del poble només hi trobes jubilats i anarquistes.

(b) Els que no miren mai la televisió, escolten les notícies a la ràdio.

(c) Qui pot llegir sense ulleres, no està jubilat.

(d) Els anarquistes ni miren mai la televisió ni pengen fotos a Instagram

Traduïu tot això al llenguatge de la lògica proposicional i demostreu (utilitzant només el llenguatge de la lògica proposicional) que *«els clients del casino que no escolten les notícies de la ràdio necessiten ulleres per llegir i miren la televisió»*.

14. Escriviu un programa en el vostre llenguatge preferit que vagi escrivint consecutivament totes les FBF de la lògica proposicional que continguin només les proposicions A, B, C.[12]

15. Escriviu un programa en el vostre llenguatge preferit que, donada una cadena de caràcters formada per lletres i símbols lògics ($\wedge$, $\vee$, $\neg$, $\Rightarrow$) decideixi si és una FBF o no ho és.

[12] El nombre de FBF amb només tres lletres es fa molt gran a partir de $n = 3$, on n és el nombre de vegades que apliquem les regles sintàctiques bàsiques. Per reduir una mica aquest nombre, suprimiu el connector $\Leftrightarrow$ i tingueu en compte la commutativitat dels connectors $\wedge, \vee$. Un cop tingueu totes les fórmules per $n = 3$ (n'hi ha 793.158), reduïu aquest nombre suprimint les fórmules que simplement es diferenciïn en una permutació de les tres lletres A, B, C. En quedaran 448.343, de les quals 121.216 són tautologies.

16. Escriviu un programa en el vostre llenguatge preferit que, donada una FBF, decideixi si és una tautologia o és una contradicció.

17. Un conjunt de connectors lògics es diu que és *complet* quan donada qualsevol taula de veritat, sempre hi ha una FBF que dóna aquesta taula de veritat i només utilitza els connectors lògics del conjunt.

    (a) Demostreu que el conjunt de connectors $\{\wedge, \vee, \neg\}$ és complet.

    (b) Demostreu que el conjunt $\{\wedge, \vee\}$ no ho és.

    (c) En informàtica es consideren les *portes lògiques* NAND i NOR que tenen aquestes taules de veritat:

A	B	NAND(A, B)	NOR(A, B)
V	V	F	F
V	F	V	F
F	V	V	F
F	F	V	V

    Escriviu NAND i NOR a partir dels connectors $\wedge, \vee, \neg$. Demostreu que cadascun dels dos connectors és complet. És a dir, podem fer tota la lògica proposicional amb un únic connector.

18. (a) De vegades s'utilitza el quantificador $\exists!$ que significa *«existeix un únic»*. Doneu una definició de $\exists! x\, P(x)$ a partir dels termes de la lògica de primer ordre.

    (b) Sovint trobem expressions del tipus $\forall x > k\, P(x)$. Escriviu això mateix en el llenguatge formal de la lògica de primer ordre.

19. Escriviu en el llenguatge de la lògica de primer ordre aquesta afirmació: *tot nombre natural primer diferent de 2 és successor d'un múltiple de 2.* (Podeu utilitzar la suma i la multiplicació.)

20. Considerem, als nombres naturals, el predicat $R(x, y)$ que és cert exactament quan $x < y$ i el predicat $S(x, y)$ que és cert exactament quan $x > y$. Decidiu quines d'aquestes fórmules són certes:

$$\text{(a) } \forall x \exists y\, R(x, y); \qquad \text{(b) } \forall x \exists y\, S(x, y).$$

21. Digueu si és certa o falsa aquesta propietat dels nombres naturals:

$$\forall n\, \exists m\, \forall k \left[\big(\exists u\, (n + 1 = ku) \wedge \exists v\, (m = kv) \big) \Rightarrow (k = 1) \right].$$

22. Considerem, en els nombres naturals, el predicat $S(a, b, c)$ que és cert exactament quan $c = a + b$. Demostreu o refuteu aquestes fórmules de lògica de primer ordre

$$\exists x \exists z \big(\neg x = z \wedge (\exists y\, S(y, y, x) \wedge \exists t\, S(t, t, z)) \big).$$

$$\exists x \big(\forall y \exists z\, S(y, z, x) \big).$$

23. Enuncieu, en el llenguatge formal de la lògica de primer ordre dels nombres naturals, aquest teorema de Fermat: *tot primer de la forma* $4k + 1$ *és suma de dos quadrats.* (Podeu utilitzar $+$, $\times$ i un predicat $P(n)$ que signifiqui que n és primer.)

24. Demostreu (usant el llenguatge ordinari de les matemàtiques) aquest teorema sobre els nombres naturals:

$$\forall n\,\exists m\,\forall k\left(\exists t\,\exists s\Big((t<k)\wedge(t<n)\wedge(t>s)\Big)\Rightarrow\Big((m>n)\wedge(m<kn)\Big)\right).$$

25. Considereu aquesta proposició sobre els nombres naturals:

$$\forall n\,\forall m\,\exists t\,\forall k\,\big((n>m+k)\Rightarrow(k\leq t)\big).$$

Demostreu-la o trobeu un contraexemple.

26. Demostreu que aquesta fórmula de la lògica de primer ordre és certa:

$$\Big[\big(\forall x\,(A(x)\Rightarrow B)\big)\wedge\big(\exists y\,A(y)\big)\Big]\Rightarrow B.$$

27. Demostreu que aquesta fórmula de la lògica de primer ordre és certa:

$$\forall x\big(A(x)\Rightarrow B(x)\big)\Rightarrow\big(\forall xA(x)\Rightarrow\forall xB(x)\big).$$

28. Considerem $A := \exists y\forall x P(x,y)$, $B : \forall x\exists y P(x,y)$. De les dues implicacions $A\Rightarrow B$, $B\Rightarrow A$, una és correcta i l'altra no. Demostreu la que és vàlida i trobeu un contraexemple en els nombres naturals de la que no ho és. (Apliqueu els exercicis 26 i 27.)

29. Demostreu, a partir dels tres axiomes de Peano, aquestes dues propietats:

    (a) Tot nombre natural $\neq 0$ és successor d'algun nombre natural.

    (b) Cap nombre natural és successor d'ell mateix.

    (c) Cap nombre natural és el successor del successor d'ell mateix.

30. Demostreu, a partir dels axiomes de Peano, el *principi d'inducció forta* de la pàgina 22.

31. Demostreu aquesta variant del principi d'inducció:

$$\Big(P(0)\wedge P(1)\wedge\forall n\big(P(n)\wedge P(n+1)\Rightarrow P(n+2)\big)\Big)\Rightarrow\forall n\,P(n).$$

32. Doneu una definició recursiva de les *fraccions contínues*

$$\langle a_0,a_1,\ldots,a_n\rangle := a_0+\cfrac{1}{a_1+\cfrac{1}{a_2+\cfrac{1}{\ddots+\cfrac{1}{a_n}}}}$$

33. Afegim als nombres naturals dos elements més designats a, b. Definim $s(a) := b$ i $s(b) := a$. Demostreu que es compleixen els dos primers axiomes de Peano, però no es compleix l'axioma d'inducció.

34. La multiplicació és una suma iterada i l'exponenciació és una multiplicació iterada. Podem continuar aquest procés de construir noves operacions iterant operacions existents. Per exemple, la *tetració* és una exponenciació iterada. Donald Knuth va introduir la notació $a \uparrow^n m$ per aquestes operacions, de manera que $a \uparrow^0 m := am$, $a \uparrow m := a^m$ i, en general $a \uparrow^n m$ consisteix en operar a amb ell mateix m vegades, amb l'operació $\uparrow^{n-1}$. Doneu una definició (doblement) recursiva de $a \uparrow^n m$ i doneu una descripció de $2 \uparrow^4 3$ que només utilitzi l'exponenciació.

35. Demostreu por inducció sobre el natural n que les següents igualtats o afirmacions són certes per tot $n \geq 1$:

    a) $\sum_{k=1}^{n} k = n(n+1)/2$.

    b) $\sum_{k=1}^{n} k^2 = \frac{1}{6} n(n+1)(2n+1)$.

    c) $n(n^2+5)$ és divisible per 6.

    d) $8^n - 3^n$ és divisible per 5.

    e) $2n^3 + 3n^2 + n$ és divisible per 6.

    f) $\sum_{k=1}^{n} 1/[(2k-1)(2k+1)] = n/(2n+1)$.

    g) $\sum_{k=1}^{n} 1/k^2 \leq 2 - 1/n$.

36. Demostreu que aquesta fórmula és vàlida per tot natural $n > 1$:

$$\prod_{i=2}^{n} \left(1 - \frac{1}{i^2}\right) = \frac{n+1}{2n}.$$

37. Estudieu totes les funcions $f : \mathbb{N} \to \mathbb{N}$ que compleixen aquestes dues propietats:

    i) $f(2) = 2$;

    ii) $\forall n \geq 1 \quad f(n+1) = 1 + f(1) + 2f(2) + \cdots + nf(n)$.

38. Recordeu que la successió de Fibonacci és la successió de nombres naturals unívocament definida per les fórmules recurrents $F_0 = 0$, $F_1 = 1$, $F_{i+1} = F_i + F_{i-1}$.

    (a) Demostreu per inducció que es compleix aquesta fórmula:

$$1 + \sum_{i=1}^{n} F_i = F_{n+2} \quad \text{per tot } n > 0.$$

    (b) Definim una successió de nombres naturals K_n per $n > 0$ com: K_n és el nombre de maneres diferents d'escriure n com a suma de 1's i 2's (entenent que, per exemple, $3 = 2 + 1$ i $3 = 1 + 2$ són expressions diferents). Demostreu que, per tot $n > 1$, es compleix $F_n = K_{n-1}$.

39. Considereu aquests nombres reals:

$$\phi := \frac{1+\sqrt{5}}{2}, \quad \psi := \frac{1-\sqrt{5}}{2}, \quad F_n := \frac{\phi^n - \psi^n}{\sqrt{5}},\ n \geq 0.$$

    Demostreu, per inducció, $\phi^n = \phi^{n-1} + \phi^{n-2}$ per tot $n \geq 2$. Deduïu $F_n = F_{n-1} + F_{n-2}$ per tot $n \geq 2$. Demostreu per inducció que $F_n \in \mathbb{N}$ per tot $n \geq 0$.

40. Definim per recursió els nombres naturals T_n:

$$T_0 = 0, \quad T_1 = 1, \quad T_n = 6T_{n-1} - T_{n-2}.$$

Siguin $\alpha > \beta$ les arrels reals de l'equació $x^2 - 6x + 1$.

(a) Demostreu les identitats $\alpha^{n+2} = 6\alpha^{n+1} - \alpha^n$ i $\beta^{n+2} = 6\beta^{n+1} - \beta^n$.

(b) Demostreu per inducció la identitat $T_n = (\alpha^n - \beta^n)/4\sqrt{2}$.

(c) Definim $t_n := (\alpha^n + \beta^n - 2)/4$. Demostreu que $2T_n^2 = t_n(t_n + 1)$ i deduïu que si tenim T_n^2 monedes iguals formant un quadrat, les podem disposar formant un triangle equilàter amb t_n monedes a cada costat (Euler).

41. Siguin $a_0, a_1, \ldots, a_n$ nombres naturals > 0. Demostreu que, per tot $n \geq 2$:

$$(1 + a_1) \cdots (1 + a_n) > 1 + a_1 + \cdots + a_n.$$

42. Dibuixem n rectes sobre un pla. Demostreu per inducció:

(a) El nombre màxim de regions que formen aquestes rectes és $(n^2 + n + 2)/2$.

(b) Les regions obtingudes es poden pintar amb dos colors (blanc i negre, per exemple), de manera que en cada frontera les dues regions limítrofes tinguin colors diferents.

43. Tenim una rajola de xocolata. Quantes vegades l'hem de trencar (seguint les línies rectes marcades a la rajola) si volem separar totes les porcions que conté? (Entenem que no és vàlid trencar a l'hora dues peces disjuntes.)

44. Imaginem un tauler com el d'escacs però amb $2^n > 1$ caselles a cada costat. Una de les caselles del tauler està ocupada per un peó. Disposem de fitxes planes en forma de ▙ cadascuna de les quals ocupa tres caselles. Demostreu que podem omplir tot el tauler (excepte la casella on hi ha el peó).

45. Què en penseu del següent raonament?

Teorema: Tots els alumnes de classe es diuen igual.

Demostració: Ho demostrarem per inducció sobre el nombre n d'alumnes que hi ha a classe. Si $n = 1$, el teorema és trivialment cert. Suposem ara que la classe té n alumnes. Demanem a un alumne que surti fora i tindrem una classe amb $n-1$ alumnes que, per hipòtesi d'inducció, es diran tots igual igual, diguem *Serravinyals*. Així tenim que tota la classe es diu *Serravinyals* excepte, potser, la persona que hem fet sortir. La fem entrar i en fem sortir una altra. De nou tenim una classe amb $n - 1$ alumnes que per inducció es diuen igual, però ja sabem que tots es deien *Serravinyals*, per tant el primer alumne que hem fet sortir també es diu *Serravinyals*.

46. Imaginem aquesta situació: dos jugadors A i B reben cadascun un cert nombre de cartes de la baralla a i b respectivament. Cadascun sap, doncs, quantes cartes té, però no sap quantes cartes té l'altre jugador. Es tracta d'endevinar el valor de $n = a + b$, amb la informació que $n \in \{r, r + k\}$ i procedint d'aquesta manera:

(a) A i B formen un equip: si un d'ells encerta el valor de n, tots dos guanyen i si un d'ells dóna un valor equivocat de n, tots dos perden.

(b) A i B poden pactar una estratègia comuna abans de començar el joc, però no es poden comunicar de cap manera durant el joc.

(c) En primer lloc, el jugador A ha de dir el valor de n o ha de passar.

(d) A continuació, el jugador B ha de dir el valor de n o ha de passar.

(e) Retornem al pas (c).

Podeu trobar una estratègia guanyadora per aquest joc?

pàgina (gairebé) en blanc

Part II:

Teoria de conjunts

Com hem dit manta vegades en aquest text, la major part de les matemàtiques es fonamenten en la teoria de conjunts i en aquesta segona part, amb les eines de lògica que hem adquirit en la part anterior, entrarem de ple en l'estudi de la teoria de conjunts en la seva fonamentació més àmpliament utilitzada, que és la que es basa en els clàssics axiomes de Zermelo–Fraenkel.

Desenvoluparem l'axiomàtica ZFC d'un manera força rigorosa, però sense entrar a fons en els temes excessivament tècnics, impropis d'una obra elemental com aquesta.

Parlarem, és clar, de conjunts, subconjunts, aplicacions, relacions, conjunt quocient, conjunts infinits, els nombres naturals i el continu, inducció i recursió...—instruments i conceptes que els matemàtics utilitzen constantment en el seu quefer quotidià. Adquirir un domini destre d'aquestes eines és fonamental per a qualsevol futur matemàtic.

Foto: Felix Hausdorff, 1868–1942

7 | Els axiomes ZFC

La teoria de conjunts es presenta com una fonamentació de la matemàtica —de la major part de la matemàtica, no de tota— en la qual tots els objectes del sistema formal s'anomenen **conjunts** i entre ells hi ha una única relació primitiva —és a dir, una relació que no es defineix, sinó que es postula— anomenada *pertinença*, designada amb el símbol $\in$.

Intuïtivament, es tracta de pensar els objectes de les matemàtiques com a col·leccions d'altres objectes, de manera que la relació $A \in B$ s'interpreta intuïtivament com que A és un element de la col·lecció (del conjunt) B.

Hem de fer èmfasi en que **tots els objectes de la teoria de conjunts són conjunts**: no hi ha *conjunts i elements*, només hi ha conjunts i un conjunt pot pertànyer a un altre conjunt, del qual direm que és un *element*. La paraula element només es pot utilitzar en una frase del tipus *«A és un element de B»*. Quan escrivim $A \in B$, això vol dir que A és un conjunt, B és un conjunt i A és un element de B.

Normalment, l'estructura formal de la teoria de conjunts que utilitzen la majoria dels matemàtics és la que es basa en els axiomes que van establir Ernst Zermelo i Abraham Fraenkel als inicis del segle XX i es coneix com la teoria ZF (o, amb un axioma extra del que parlarem més endavant, la teoria ZFC). Aquesta axiomàtica ZF és la que estudiarem ara —evitant entrar en excessius detalls tècnics.

La teoria ZF és un sistema formal dins de la lògica de primer ordre, amb una relació d'igualtat que es relaciona amb la de pertinença de la manera següent:

> **Axioma d'extensionalitat:** *si A i B són conjunts, aleshores $A = B$ és equivalent a que, per tot X,*
>
> $$X \in A \Leftrightarrow X \in B.$$

És a dir, dos conjunts amb els mateixos elements són el mateix conjunt. Si per tot X, $X \in A$ implica $X \in B$, direm que A és un *subconjunt* de B i utilitzarem la notació $A \subseteq B$. Si volem indicar que A és un subconjunt de B però $A \neq B$ —direm que és un *subconjunt propi*— podem utilitzar la notació $A \subsetneq B$.

Els primers axiomes de ZF

Ja hem vist un primer axioma de la teoria ZF: l'axioma d'extensionalitat. El pas següent és donar una sèrie d'axiomes que ens permetran construir conjunts. Com que per començar necessitem alguna cosa, prenem com axioma que existeix algun conjunt.[1] Un cop sabem que hi ha algun conjunt, el mètode principal per construir conjunts, que utilitzarem constantment, el dóna aquest axioma:

> **Axioma d'especificació:** *si A és un conjunt i P és un predicat (és a dir, una FBF de la teoria que, aplicada a un conjunt, pot ser veritable o falsa), aleshores existeix un conjunt que té per elements exactament els elements X de A tals que $P(X)$ és cert. La notació que utilitzarem per denotar aquest nou conjunt és aquesta:*
>
> $$\{X \in A \,:\, X \text{ compleix } P\}.$$

Observem que aquest axioma és el que n'havíem dit un *esquema d'axiomes* (pàgina 14), és a dir, un axioma per a cada predicat P. Observem també que l'axioma d'especificació de Zermelo-Fraenkel ens permet esquivar la *paradoxa de Russell* de la que hem parlat abans (pagina 6): per construir un conjunt a partir d'una propietat cal començar amb un conjunt previ. D'aquesta manera, *«el conjunt de tots els conjunts»* o *«el conjunt de tots els conjunts que no es pertanyen a ells mateixos»*, que donen lloc a la paradoxa, no existeixen.

Amb els axiomes anteriors ja en tenim prou per demostrar que existeix un únic conjunt que no té cap element, és a dir, un conjunt A que té la propietat que $x \in A$ és fals per tot x. L'anomenem *el conjunt buit* i el designem per $\varnothing$. Demostrem la seva existència i unicitat.

- **Existència:** Sabem que existeix algun conjunt. Sigui A un conjunt. Definim $\varnothing := \{x \in A : x \neq x\}$. És evident, doncs, que aquest conjunt $\varnothing$ té la propietat que $x \in \varnothing$ és fals per tot x.
- **Unicitat:** Suposem que $\emptyset$ i $\varnothing$ son dos conjunts buits. Pels principis de la lògica proposicional tenim que $x \in \emptyset \Rightarrow x \in \varnothing$ i, per tant, $\emptyset \subseteq \varnothing$. Pel mateix raonament, $\varnothing \subseteq \emptyset$. Per tant, per l'axioma d'extensionalitat, $\emptyset = \varnothing$.

Hi ha **tres axiomes més** que ens permeten construir nous conjunts a partir d'uns altres conjunts donats.

> **Axioma de la parella:** *si A i B són conjunts (podrien ser iguals), existeix un conjunt els elements del qual són exactament A i B. Aquest conjunt es denota $\{A, B\}$.*

[1] Aquest axioma és innecessari si, com comentàvem a la nota de la pàgina 16, a la lògica de primer ordre ja donem per fet que el domini del discurs no pot ser buit.

En particular, si A és un conjunt i apliquem l'axioma de la parella a A, A, obtenim l'existència del conjunt $\{A, A\} = \{A\}$.

L'axioma següent ens diu que si A és un conjunt, tots els subconjunts de A també formen un conjunt, que es designa 2^A o també $\mathcal{P}(A)$.

> **Axioma del conjunt de parts:** *si A és un conjunt, existeix un conjunt $\mathcal{P}(A)$ tal que*
> $$X \in \mathcal{P}(A) \quad \Leftrightarrow \quad X \subseteq A.$$

Finalment, el tercer axioma que ens permet construir nous conjunts és l'axioma que afirma l'existència de la unió de conjunts:

> **Axioma de la unió:** *si X és un conjunt, existeix un conjunt els elements del qual són els x tals que $x \in A$ per algun $A \in X$.*

Si apliquem aquest axioma a un conjunt $X = \{A, B\}$ obtenim el conjunt *unió* de A i B, que es denota $A \cup B$, els elements del qual són els elements de A i els elements de B. Però l'axioma va molt més enllà perquè ens diu que podem considerar la unió de qualsevol conjunt de conjunts[2]: si X és un conjunt, la unió dels seus elements seria l'únic conjunt $\bigcup_{A \in X} A$ que compleix aquesta propietat[3]

$$a \in \bigcup_{A \in X} A \quad \Leftrightarrow \quad \exists A\,(A \in X \wedge a \in A).$$

En general, si tenim conjunts $A_1, \ldots, A_n$, l'axioma de la parella i l'axioma de la unió ens permeten considerar el conjunt $\{A_1, \ldots, A_n\}$. En canvi, amb els axiomes que hem vist fins ara no podem afirmar l'existència de conjunts amb infinits elements. Caldrà un axioma específic que veurem més endavant.

La *intersecció* d'una família no buida de conjunts no necessita cap axioma específic. Donats conjunts A i B podem definir

$$A \cap B := \{x \in A : x \in B\} = \{x \in B : x \in A\}$$

i, més en general, el conjunt

$$\bigcap_{A \in X} A$$

està ben definit sempre que $X \neq \varnothing$.[4] Si $A \cap B = \varnothing$, direm que els conjunts A i B són *disjunts*.

---

[2] L'expressió *conjunt de conjunts* és redundant: tots els conjunts són conjunts de conjunts, perquè ja hem dit que els elements de qualsevol conjunt són conjunts (quina altra cosa podrien ser?). Malgrat això, de vegades utilitzem aquesta expressió, simplement per èmfasi.

[3] Qui seria aquesta unió si $X = \varnothing$?

[4] Qui seria aquests intersecció si $X = \varnothing$?

Hi ha dos axiomes més que són força tècnics i podem ometre en una primera lectura perquè fora dels estudis de teoria de conjunts gairebé no caldrà utilitzar-los mai. Són l'**axioma del reemplaçament**[5] i l'**axioma de fundació**.[6]

Els nombres naturals a ZF

Si la teoria de conjunts ha de ser una fonamentació de la matemàtica, hem de poder definir, amb les eines de la teoria de conjunts que acabem d'introduir, els **nombres naturals**. Recordem que *tot* han de ser conjunts i, per tant, cada nombre natural 0, 1, 2 ... ha de ser un conjunt. Com podem fer-ho? Una idea interessant pot ser definir cada natural n com un conjunt concret amb n elements. Per exemple, 0 seria un conjunt amb zero elements. Com de de conjunt amb zero elements només n'hi ha un, sembla natural definir $0 := \varnothing$. Ara toca definir 1 com un cert conjunt amb un únic element. A partir dels axiomes que hem introduït fins ara, un conjunt concret amb un únic element podria ser $\{\varnothing\}$. Amb aquesta idea, hi ha diverses maneres de definir 0, 1, 2, 3,... Per exemple:

$$
\begin{aligned}
0 :=& \varnothing \\
1 :=& \{\varnothing\} = 0 \cup \{0\} = \{0\} \\
2 :=& \{\varnothing, \{\varnothing\}\} = 1 \cup \{1\} = \{0, 1\} \\
3 :=& \{\varnothing, \{\varnothing\}, \{\{\varnothing, \{\varnothing\}\}\}\} = 2 \cup \{2\} = \{0, 1, 2\} \\
& \ldots\ldots
\end{aligned}
$$

[5] L'**axioma del reemplaçament** (que, novament, és un *esquema d'axiomes*) es pot descriure d'una manera informal dient que si tenim una família de conjunts X_i indexada sobre un conjunt A, és a dir, per cada $i \in A$ tenim ben definit un conjunt X_i, aleshores aquesta família forma un conjunt. Dit d'una altra manera, si ja sabem que A és un conjunt, aleshores és vàlid parlar del conjunt $\{X_i : i \in A\}$ i, en particular, podem considerar la unió dels conjunts d'aquesta família $\bigcup_{i\in A} X_i$. Per exemple, considerem les inclusions naturals de la recta real en el pla real, d'aquest en l'espai tridimensional, etc. Tenim una successió d'inclusions de conjunts $\mathbb{R} \subset \mathbb{R}^2 \subset \mathbb{R}^3 \subset \mathbb{R}^4 \subset \cdots$ que dóna lloc a un conjunt força habitual a les matemàtiques que és la unió de tots aquests conjunts: $\mathbb{R}^\infty := \bigcup_{n\in\mathbb{N}} \mathbb{R}^n$. L'axioma del reemplaçament ens diu que aquesta construcció és correcta (si bé, en aquest exemple concret, hi ha altres maneres de definir $\mathbb{R}^\infty$ sense necessitat d'utilitzar l'axioma del reemplaçament). L'exemple més senzill on necessitem inevitablement l'axioma del reemplaçament per assegurar que tenim un conjunt és $\{\mathbb{N}, \mathcal{P}(\mathbb{N}), \mathcal{P}(\mathcal{P}(\mathbb{N})), \mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N}))), \ldots\}$, un conjunt que, tanmateix, potser no ens caldrà utilitzar mai a la vida.

[6] L'**axioma de fundació** —conegut també com a *axioma de regularitat*— afirma que si A és un conjunt no buit, existeix un $x \in A$ tal que $x \cap A = \varnothing$. Per entendre la importància d'aquest axioma, pensem que implica que $x \in x$ és sempre fals (n'hi ha prou amb aplicar l'axioma a $A = \{x\}$) o, més en general, implica que no hi ha cap cadena infinita de conjunts $x_0 \ni x_1 \ni x_2 \ni x_3 \ni \cdots$ (apliquem l'axioma al conjunt $\{x_i : i \in \mathbb{N}\}$ que és un conjunt per l'axioma del reemplaçament) ni cap *cicle* $x_0 \ni x_1 \ni x_2 \ni \cdots \ni x_n \ni x_0$ i també implica que si $A \in B$, aleshores $B \notin A$ (apliquem l'axioma a $\{A, B\}$). En particular, si A és un conjunt no buit i prenem un element de A i un element d'aquest element de A i així successivament, després d'un nombre finit de passos el procediment s'aturarà i això només pot passar si arribem al conjunt buit. De manera informal, l'axioma de fundació ens diu que tots els conjunts estan formats a partir del conjunt buit.

És a dir, definim zero com el conjunt buit i definim el *successor* d'un natural n com

$$s(n) := n \cup \{n\}.^{7}$$

Ara bé, amb els axiomes que hem vist fins ara no podem demostrar que tots els naturals formen un conjunt. Ens falta un últim axioma:

> **Axioma de l'infinit:** *Existeix un únic conjunt* $\mathbb{N}$ *que conté exactament els nombres 0, 1, 2 ...*[8]

La manera com hem enunciat aquest axioma no és satisfactòria des d'un punt de vista del rigor lògic (els *punts suspensius*!). Expliquem breument com s'escriu aquest axioma d'una manera correcta. En primer lloc, definim el *successor* d'un conjunt

$$s(A) := A \cup \{A\}.$$

Ara definim *conjunt inductiu* com aquell conjunt S tal que $\varnothing \in S$ i $A \in S \Rightarrow s(A) \in S$. Amb aquestes definicions, l'enunciat de l'**axioma de l'infinit** diu: *«existeix un conjunt inductiu»*. Suposem, doncs, que U és un conjunt inductiu. És clar que qualsevol intersecció de subconjunts inductius de U és també un conjunt inductiu. Aleshores, definim $\mathbb{N}$ com la intersecció de *tots* els subconjunts inductius de U. Ja tenim els nombres naturals. Cal fer encara una darrera consideració: aquest conjunt $\mathbb{N}$ l'hem construït a partir de U i cal veure que si comencem amb un altre conjunt inductiu U' també arribem al mateix $\mathbb{N}$. La conclusió és que $\mathbb{N}$ és efectivament un conjunt ben definit, és un conjunt inductiu, i és el *més petit dels conjunts inductius* en el sentit que qualsevol altre conjunt inductiu ha de contenir $\mathbb{N}$ com a subconjunt.

Observem que a la teoria ZFC dels nombres naturals el *principi d'inducció* està incrustat a la mateixa definició de $\mathbb{N}$. Considerem una propietat $P(n)$ que es compleixi per 0 i tal que, per tot n, $P(n) \Rightarrow P(n+1)$. Aleshores, si considerem el subconjunt de $\mathbb{N}$

$$A := \{n \in \mathbb{N} : P(n)\}$$

tenim que A és un subconjunt inductiu de $\mathbb{N}$ i, com que $\mathbb{N}$, per definició, és el més petit dels conjunts inductius, tenim que $A = \mathbb{N}$ i $P(n)$ és cert per tot n.

Si aquests naturals que hem construït a la teoria de conjunts han de ser els naturals *de tota la vida*, hauríem de comprovar que compleixen els **axiomes de**

[7] Cal situar aquestes definicions en el seu context. En primer lloc, diguem que el matemàtic professional *mai* no utilitzarà aquestes definicions per a res. És important que existeixi una definició amb la qual els nombres naturals tinguin les propietats que volem, però és irrellevant quina sigui aquesta definició i, de fet, hi ha diverses definicions possibles. D'atra banda, si comparem, per exemple, aquestes dues fórmules: $2 \in 3$ i $2 + 3 = 5$, trobarem que hi ha una diferència essencial entre elles. La primera és certa en aquest model dels nombres naturals que proposem aquí, mentre que la segona és certa en tots els models dels nombres naturals que puguin existir. De fet, hi ha un punt de vista sobre els fonaments de les matemàtiques que vol prescindir del primer tipus de fórmules.

[8] Se'n diu axioma de l'infinit perquè és el primer axioma que ens permet afirmar l'existència d'un conjunt amb infinits elements.

Peano. Per la manera com els hem definit, els axiomes de Peano són evidents excepte un:

$$s(n) = s(m) \quad \Rightarrow \quad n = m$$

la demostració del qual és una mica enrevessada. Cal començar demostrant aquest lema:

$$m \in n \quad \Rightarrow \quad m \subseteq n.$$

Per veure-ho, considerem el conjunt de tots els nombres naturals per als quals això és cert:

$$S := \{n \in \mathbb{N} : \forall m(m \in n \Rightarrow m \subseteq n)\} \subseteq \mathbb{N}.$$

Si demostrem que S és un conjunt inductiu, com que $\mathbb{N}$ és el més petit dels conjunts inductius, deduirem que $S = \mathbb{N}$ i el lema serà cert. Que $0 \in S$ és evident. Suposem que $k \in S$ i intentem demostrar que $s(k) \in S$. És a dir, suposem que $m \in s(k)$ i intentem demostrar que $m \subseteq s(k)$. Si $m \in s(k) = k \cup \{k\}$, tenim que $m \in k$ o $m = k$. En el primer cas, com que $k \in S$, deduïm que $m \subseteq k \subseteq s(k)$. En el segon cas, tenim $m = k \subseteq s(k)$. En els dos casos, doncs, arribem a la conclusió que volíem.

Un cop demostrat el lema, ja podem demostrar l'axioma de Peano que diu

$$s(n) = s(m) \quad \Rightarrow \quad n = m.$$

Suposem $n \cup \{n\} = s(n) = s(m) = m \cup \{m\}$. Com que $m \in s(m)$, tindrem $m \in n \cup \{n\}$ i, per tant, $n = m$ (amb la qual cosa hem acabat la demostració) o $m \in n$. En aquest cas, el lema anterior ens diu que $m \subseteq n$. Repetint aquest mateix raonament intercanviant els papers de n i m obtenim que $n = m$ o $n \subseteq m$. La doble inclusió $n \subseteq m$, $m \subseteq n$ ens diu que $n = m$.

A partir de tot això que hem dit podem desenvolupar, a la teoria ZF, tota l'aritmètica de Peano incloent, en particular, el principi d'inducció, el teorema de recursió (vegeu l'exercici **II**.24),[9] les definicions de suma, multiplicació i ordre amb les seves propietats bàsiques, etc. Aquesta fonamentació dels nombres naturals a partir de la teoria de conjunts s'ha convertit, doncs, en la fonamentació estàndard, relegant a un segon pla el punt de vista pre-conjuntista de Peano que hem estudiat en capítols anteriors, de manera que l'aritmètica de Peano segueix sent important per als especialistes en fonaments de les matemàtiques, però no per a la gran majoria dels matemàtics, que treballen en el context de la teoria de conjunts de Zermelo-Fraenkel.

[9] Ens podríem preguntar com és que els axiomes de Peano de primer ordre no permeten demostrar el teorema de recursió i, en canvi, els axiomes de ZF, que també són de primer ordre, sí que ho permeten. La resposta es troba en l'*axioma del conjunt de parts*, que és un axioma potentíssim. Per exemple, aquest axioma ens permet escriure *«per tot subconjunt de* $\mathbb{N}$*...»* mentre que la majoria dels subconjunts de $\mathbb{N}$ no es poden definir utilitzant l'axioma d'especificació perquè, com veurem, $\mathbb{N}$ té una quantitat no numerable de subconjunts i, en canvi, només hi ha una quantitat numerable de FBF.

L'axioma de l'elecció

Els axiomes que hem introduït fins ara són, com hem explicat, els de Zermelo i Fraenkel (ZF). Posteriorment —no sense polèmica— es va decidir afegir un nou axioma que s'ha demostrat que és molt útil a l'hora de fer matemàtiques: l'**axioma de l'elecció**. Expliquem en què consisteix aquest axioma.

Suposem, per exemple, que $X = \{A, B, C\}$ és un conjunt amb tres elements i que els conjunts A, B, C són $\neq \varnothing$ i, per simplificar, disjunts dos a dos. Aleshores, podem fer aquesta construcció: com que $A \neq \varnothing$, existeix $a \in A$; com que $B \neq \varnothing$, existeix $b \in B$; com que $C \neq \varnothing$, existeix $c \in C$. Podem, per tant, considerar el conjunt $Y := \{a, b, c\}$ que té, exactament, un element de cadascun dels conjunts que pertanyen a X. És a dir, podem *escollir* elements a, b, c, un de cada conjunt A, B, C. Si en lloc de tres conjunts A, B, C en tinguéssim tres mil, ho podríem fer igual, però el procés seria més llarg. La pregunta és: i si en tenim infinits? Sembla que també podríem escollir —per què no?— un element de cada conjunt i formar el conjunt Y.[10] El problema és que, amb els axiomes de ZF, això no és possible.[11] Si ho volem fer —i en moltes demostracions que ens trobem a les matemàtiques, ens cal fer aquesta construcció consistent en fer *infinites eleccions*— necessitem un nou axioma que, d'alguna manera, ens digui que si tenim un conjunt X i els seus elements A són conjunts no buits, podem triar un element de cadascun dels conjunts A.

Per poder enunciar correctament aquest axioma hem de donar un significat formal a la idea de *triar*. Això es fa amb el concepte de *funció d'elecció*. Aleshores, l'axioma diu això:

> **Axioma de l'elecció:** *Sigui X un conjunt. Suposem que $\varnothing \notin X$. Aleshores, existeix una funció (anomenada funció d'elecció)*
>
> $$f : X \longrightarrow \bigcup_{A \in X} A$$
>
> *tal que, per tot $A \in X$ es compleix $f(A) \in A$.*[12]

[10] Per explicar d'una manera intuïtiva la necessitat de l'axioma de l'elecció s'acostuma a recórrer a l'*exemple de les sabates i els mitjons*. Imaginem un magatzem amb una col·lecció infinita de capses de sabates. Imaginem que l'encarregat del magatzem —que compleix escrupolosament les ordres que rep, però no té lliure albir— rep l'ordre de posar a l'aparador una sabata de cada capsa. L'encarregat pregunta *«quina?»* i si li contesten, per exemple *«la del peu dret»*, ja té prou informació per complir l'ordre rebuda. En canvi, si en lloc de sabates parlem de parells de mitjons, l'ordre de presentar exactament un mitjó de cada parell no es pot dur a terme perquè no podem contestar la pregunta *«quin?»*.

[11] És a dir, l'axioma de l'elecció és *indecidible* a ZF. Gödel va demostrar (1938) que a ZF no es pot demostrar que l'axioma de l'elecció és fals, i Paul Cohen va demostrar (1963) que a ZF tampoc no es pot demostrar l'axioma de l'elecció. Tot això, evidentment, suposant que ZF sigui consistent.

[12] Aquest axioma dóna existència a uns objectes que de cap manera no poden ser construïts ni per un ésser humà ni per un ordinador. Es pot comparar amb postular l'existència d'un *dimoni de*

Quan afegim a ZF l'axioma de l'elecció obtenim la teoria ZFC (ZF+*«choice»*) que es considera que és la fonamentació estàndard de les matemàtiques. L'estudiant, al llarg dels seus estudis de matemàtiques, trobarà un bon nombre d'aplicacions d'aquest axioma de l'elecció. Potser la primera serà el teorema *«tot espai vectorial té alguna base»*, que és un teorema que no es pot demostrar sense l'axioma de l'elecció.

Maxwell a les matemàtiques. Recordem que aquest dimoni hipotètic podia violar les lleis de la termodinàmica i, per exemple, fer que un gas en equilibri es separés en una part calenta i una part freda. L'axioma de l'elecció té un comportament similar al d'aquell dimoniet i ens diu, per exemple, que podem descompondre un pèsol en un nombre finit de peces i tornar-les a ajuntar fent una esfera de la mida del Sol (paradoxa de Banach-Tarski), que existeix una base de $\mathbb{R}$ com espai vectorial sobre $\mathbb{Q}$, que podem reordenar els nombres reals de manera que tot subconjunt no buit tingui mínim, o que hi ha uns nombres naturals no estàndard que compleixen els axiomes de Peano (de primer ordre) però són no numerables. Cadascuna d'aquestes construccions requereix una infinitat no numerable d'eleccions, és a dir, *no es poden fer de manera efectiva*. Tanmateix, l'axioma de l'elecció és útil en el sentit que, si més no, ens diu que no hem d'intentar demostrar que tot subconjunt de $\mathbb{R}^3$ té assignat un volum o que no hi ha nombres naturals als que no es pot arribar pas a pas a partir del zero.

8 Productes, relacions, aplicacions

Els axiomes de la teoria ZFC que hem vist al capítol anterior ens permeten fer algunes de les construccions bàsiques de les matemàtiques —subconjunts, conjunt de parts, unió, intersecció, elecció. Ara estudiarem algunes altres construccions essencials com són el producte cartesià o les funcions.

El producte cartesià de dos conjunts

Si A i B són conjunts, el producte $A \times B$ serà el conjunt format per les parelles ordenades (a, b) amb $a \in A$ i $b \in B$. Hem de definir què és una parella ordenada (que ha de ser un conjunt) i hem de demostrar que formen un conjunt. En una parella ordenada, la propietat que volem garantir és que

$$(a, b) = (a', b') \quad \Leftrightarrow \quad (a = a') \wedge (b = b'). \qquad (*)$$

Com podem **definir** (a, b) de manera que tinguem aquesta propietat? De fet, hi ha diverses maneres de fer-ho i n'hem d'escollir una.[1] Per exemple, si definim

$$(a, b) := \{\{a\}, \{a, b\}\} \in \mathcal{PP}(A \cup B)$$

es pot comprovar (exercici interessant per practicar aquests temes) que es compleix la propietat desitjada $(*)$ i això ens permet definir[2] el producte cartesià de dos conjunts (l'anomenarem simplement «producte»)

$$A \times B := \{(a, b) \in \mathcal{PP}(A \cup B) : a \in A \wedge b \in B\}.$$

Amb aquesta definició també podem parlar del producte d'un conjunt finit[3] (ordenat) de conjunts $A_1 \times \cdots \times A_n$.[4]

[1] A aquesta definició de parella ordenada podem aplicar-li el mateix comentari de la nota 7 del capítol anterior.

[2] Observem que per poder aplicar l'axioma d'especificació cal que totes les parelles ordenades estiguin en un conjunt. Efectivament, per la definició que hem pres és clar que totes pertanyen a $\mathcal{P}\big(\mathcal{P}(A \cup B)\big)$.

[3] També es pot estendre el concepte de producte de conjunts a una família infinita de conjunts, però aquest tema no el tractarem aquest curs.

[4] La definició és recursiva $A_1 \times \cdots \times A_n := A_1 \times (A_2 \times \cdots \times A_n)$.

Relacions

Sovint, tindrem una *relació* entre els elements d'un conjunt. Per exemple, en els nombres naturals tenim una relació d'ordre amb la que, per exemple, $2 < 5$ és cert i $3 < 1$ és fals. O, per exemple, podem considerar la relació

$$n \sim m \quad \Leftrightarrow \quad |n - m| \text{ és múltiple de } 79.$$

Què és, exactament, una **relació** en un conjunt? Com que estem fonamentant-ho tot en la teoria de conjunts ZF, una relació també ha de ser un conjunt. Quin conjunt? Una solució senzilla d'aquesta pregunta és pensar una relació com *el conjunt de totes les parelles d'elements relacionats*. Amb aquesta idea, definim

> ***Definició de relació**: una relació en un conjunt A és un subconjunt $R \subseteq A \times A$. Direm que a està relacionat amb b si $(a, b) \in R$.*

Hi ha un tipus de relacions que juguen un paper important a les matemàtiques (i ja han aparegut anteriorment en aquestes notes): les **relacions d'equivalència**. Una relació (que escriurem amb el símbol $\sim$) en un conjunt A diem que és d'equivalència si compleix aquestes tres propietats:

- **Propietat reflexiva.** Per tot $a \in A$ es compleix $a \sim a$.
- **Propietat simètrica.** Per tot $a, b \in A$ es compleix $a \sim b \Rightarrow b \sim a$.
- **Propietat transitiva.** Per tot $a, b, c \in A$ es compleix $(a \sim b) \wedge (b \sim c) \Rightarrow a \sim c$.

Observem que aquestes tres propietats les compleix la relació d'**igualtat** i, d'aquesta manera, podem pensar en les relacions d'equivalència com una generalització del concepte d'igualtat.

Unes altres relacions que són importants a les matemàtiques són les **relacions d'ordre**. Una relació (que escriurem amb el símbol $\leq$) en un conjunt A direm que és d'ordre si compleix aquestes tres propietats:

- **Propietat reflexiva.** Per tot $a \in A$ es compleix $a \leq a$.
- **Propietat antisimètrica.** Per tot $a, b \in A$ es compleix $(a \leq b) \wedge (b \leq a) \Rightarrow a = b$.
- **Propietat transitiva.** Per tot $a, b, c \in A$ es compleix $(a \leq b) \wedge (b \leq c) \Rightarrow a \leq c$.

Evidentment, si tenim una relació d'ordre $\leq$ també podem considerar la relació *estricta* associada $<$. Hi ha una tercera condició que podem exigir (o no) a una relació d'ordre, que afirma que donats dos elements sempre els podem *comparar*:

- **Totalitat.** Per tot $a, b \in A$ es compleix $(a \leq b) \vee (b \leq a)$.

Si una relació d'ordre compleix aquesta tercera condició, direm que és una relació d'ordre **total**. Si no la compleix, direm que és una relació d'ordre **parcial**. Un conjunt amb una relació d'ordre parcial/total diem que és un *conjunt parcialment/totalment ordenat*. Sovint, d'un conjunt parcialment ordenat se'n diu un **poset**.

Exemples:

- Definim una relació a $\mathbb{N}$ dient que $a \sim b$ si $|a - b|$ és múltiple de 3. No és difícil veure que es tracta d'una relació d'equivalència.
- $\mathbb{N}$ és un conjunt totalment ordenat amb la relació d'ordre habitual.
- Considerem el conjunt dels polinomis amb coeficients nombres racionals i definim $f \leq g$ si el grau de f és menor o igual que el grau de g. No és una relació d'ordre perquè no es compleix la propietat antisimètrica.
- Si A és un conjunt, $\mathcal{P}(A)$ és un poset amb la inclusió com a relació d'ordre.
- A $\mathbb{N}$ definim una relació dient que $n \vdash m$ si n divideix m. És una relació d'ordre parcial.

Quan tenim una relació d'ordre a un conjunt A i un subconjunt $X \subseteq A$ podem considerar diversos conceptes importants:

- **màxim, mínim**. Diem que $x \in X$ és un mínim de X si $x \leq y$ per tot $y \in X$. Diem que $x \in X$ és un màxim de X si $y \leq x$ per tot $y \in X$. Un subconjunt X pot tenir màxim, mínim o no tenir-los. Si en té, són únics.
- **maximal, minimal**. Diem que $x \in X$ és un element minimal de X si no hi ha cap $y \in X$ tal que $y < x$. Diem que $x \in X$ és un element maximal de X si no hi ha cap $y \in X$ tal que $x < y$. Un subconjunt X pot tenir elements maximals, minimals o no tenir-ne. Poden no ser únics.
- **cotes superiors, inferiors**. Una cota superior per a X és un element $a \in A$ tal que $x \leq a$ per tot $x \in X$. Una cota inferior per a X és un element $a \in A$ tal que $a \leq x$ per tot $x \in X$.
- **intervals**. Donats dos elements $a \leq b$ del conjunt ordenat A, definim l'interval
$$[a, b] := \{x \in A : a \leq x \leq b\}.$$
També podem definir intervals del tipus (a, b) —que no hem de confondre amb la *parella ordenada* (a, b)—, $[a, b)$ i $(a, b]$.

La relació d'ordre $n \vdash m$ de l'exemple anterior és útil per entende millor les diferències entre aquests conceptes (vegeu l'exercici **II**.17).

Aplicacions

Un dels conceptes més importants de les matemàtiques és el concepte de **funció** que rep diversos noms segons el context i que en el context de la teoria de conjunts s'anomena **aplicació**.

Tots coneixem aquest concepte. Per exemple, considerem aquesta aplicació (que apareix a la famosa conjectura de Collatz de la que hem parlat abans):

$$f : \mathbb{N} \to \mathbb{N}, \qquad f(n) := \begin{cases} n/2 \text{ si } n \text{ és parell;} \\ 3n+1 \text{ si } n \text{ és senar.} \end{cases}$$

Coses importants a tenir en compte:

- Tenim un domini de definició o també conjunt de sortida, sobre el que està definida l'aplicació. En aquest cas, $\mathbb{N}$.
- Tenim un conjunt d'arribada al qual pertanyen els valors $f(n)$. En aquest cas, $\mathbb{N}$.
- Finalment, tenim una definició de $f(n)$ per a cada valor de n que ens determina, de manera unívoca, quin valor té $f(n)$. Diem que $f(n)$ és la *imatge* de n per l'aplicació f.

En el context de la teoria de conjunts, si volem definir el concepte d'aplicació entre dos conjunts $f : A \to B$ ho farem a través de la *gràfica* de f: direm que una aplicació *de* A *a* B és un subconjunt $f \subseteq A \times B$ que té la propietat que per tot $a \in A$ existeix un únic $b \in B$ tal que $(a, b) \in f$. En aquest cas, escriurem $b = f(a)$ i direm que b és la imatge de a per f. Si sobreentenem f, també escriurem $a \mapsto b$.[5] Observem que, si A i B són conjunts, totes les aplicacions $f : A \to B$ formen un conjunt que s'acostuma a designar B^A o també (preferiblement) $\mathcal{F}(A, B)$.

Repassem ara alguns conceptes importants sobre aplicacions.

- Una propietat essencial de les aplicacions és que, en determinats casos, es poden compondre. Si $f : A \to B$ és una aplicació i $g : B \to C$ és una altra aplicació, podem considerar la seva **composició**

  $$g \circ f : A \to C$$

  definida així

  $$(g \circ f)(a) := g\big(f(a)\big).$$

[5] Cal observar que els símbols $\to$ i $\mapsto$ indiquen coses molt diferents. A més, el fet que $\to$ indiqui una aplicació ens diu que no hauríem d'utilitzar aquest símbol per a cap altra cosa.

- Per a qualsevol conjunt A podem considerar l'anomenada aplicació identitat $I : A \to A$ que ve donada per $I(a) = a$ per tot $a \in A$.
- Sigui $f : A \to B$ una aplicació i suposem que $f(a) = b$. Direm que b és **la imatge** de a i que a és **una antiimatge** de b. Aleshores:
  - Si tot element de B té com a mínim una antiimatge, direm que f és una aplicació **exhaustiva**. De vegades, utilitzarem la notació $A \twoheadrightarrow B$ per indicar que f és una aplicació exhaustiva. De les aplicacions exhaustives de vegades també se'n diu *projeccions*.
  - Si cada element de B té com a màxim una antiimatge, direm que f és una aplicació **injectiva**. Formulat d'una manera equivalent, una aplicació és injectiva si compleix
    $$\big(f(x) = f(y)\big) \;\Rightarrow\; (x = y).$$
    De vegades, utilitzarem la notació $A \rightarrowtail B$ per indicar que f és una aplicació injectiva. De les aplicacions injectives de vegades també se'n diu *injeccions*.[6]
  - Si f és a la vegada injectiva i exhaustiva, diem que és **bijectiva** (o que és una *bijecció*). En aquest cas, cada element de B té una única antiimatge i podem definir una aplicació $g : B \to A$ que és inversa de f, en el sentit que, per tot $a \in A$ es compleix que $g(f(a)) = a$ i per tot $b \in B$ es compleix que $f(g(b)) = b$. La notació tradicional per aquesta aplicació inversa —que és única— és f^{-1}, cosa que, en alguns contextos, pot generar confusió. Cal insistir que aquesta aplicació inversa només existeix si f és bijectiva.
- Algunes aplicacions reben el qualificatiu de **canòniques**. Per exemple:
  - Si tenim un producte de conjunts $A \times B \neq \varnothing$ podem definir dues aplicacions $\pi_A : A \times B \to A$, $\pi_B : A \times B \to B$ per
    $$\pi_A(a, b) := a, \quad \pi_B(a, b) := b.$$

[6] Els conceptes d'aplicació injectiva i aplicació exhaustiva són molt senzills, però també és cert que, quan l'estudiant s'hi enfronta per primera vegada, hi ha el risc que confongui aquests conceptes amb altres que poden semblar similars. Considerem aquestes afirmacions sobre una aplicació $f : A \to B$:

1. Tot element de A té alguna imatge.
2. Tot element de A té una única imatge.
3. Elements de A diferents tenen imatges diferents.
4. Per tot $x, y \in A$, si $x = y$, aleshores $f(x) = f(y)$.
5. Per tot $x, y \in A$, si $f(x) = f(y)$, aleshores $x = y$.
6. Si dos elements de A tenen imatges diferents, són diferents.
7. Tot element de B té una única antiimatge.

1 i 2 són sempre certs, per la pura definició d'aplicació. 3 i 5 són contrarecíprocs un de l'altre i, per tant, són afirmacions equivalents; les dues defineixen el concepte d'aplicació injectiva. 4 és una tautologia, per aplicació directa de la llei de Leibnitz. 6 és el contrarecíproc de 4 i, per tant, és també una tautologia. 7 és equivalent a la definició d'aplicació bijectiva.

Direm que són les **projeccions canòniques** de $A \times B$ sobre A i B, respectivament, perquè, efectivament, són aplicacions exhaustives.

- Si tenim un conjunt B i un subconjunt A podem definir una aplicació $i : A \to B$ per $i(a) := a$. Direm que és la **injecció canònica** de A a B perquè, efectivament, és una aplicació injectiva.

- Si $f : A \to B$ és una aplicació i $b \in B$, podem considerar el conjunt de totes les antiimatges de b. De vegades, aquest conjunt s'anomena «la fibra de f sobre b». És un subconjunt de A que pot ser buit. Malauradament, la notació que utilitza tothom per indicar aquest conjunt indueix a confusió:

$$f^{-1}(b) := \{a \in A \ : \ f(a) = b\} \subseteq A.$$

  Observem, doncs, que $f^{-1}(b)$ està sempre definit, encara que f no sigui bijectiva, i és un subconjunt (potser buit) de A.

- Una aplicació $f : A \to B$ dóna lloc a dues aplicacions

$$f_* : \mathcal{P}(A) \to \mathcal{P}(B)$$
$$f^* : \mathcal{P}(B) \to \mathcal{P}(A)$$

  Aquestes dues aplicacions es defineixen així:

$$f_*(X) := \{b \in B \ : \ \text{existeix } a \in X \text{ tal que } f(a) = b \ \}$$
$$f^*(Y) := \{a \in A \ : \ f(a) \in Y\}$$

  Malauradament, per augmentar la confusió en les notacions, aquestes aplicacions que hem designat provisionalment per f_* i f^*, es designen a la pràctica f i f^{-1}, respectivament. L'estudiant ha d'aprendre a no confondre's amb aquestes notacions tan poc afortunades.

- Utilitzarem sovint aquestes propietats de les aplicacions f_* i f^* (que l'estudiant demostrarà com a exercici):
  - $f(\bigcup_{i\in I} A_i) = \bigcup_{i\in I} f(A_i)$.
  - $f(\bigcap_{i\in I} A_i) \subseteq \bigcap_{i\in I} f(A_i)$ i la igualtat es compleix si f és injectiva.
  - $f^{-1}(\bigcup_{i\in I} B_i) = \bigcup_{i\in I} f^{-1}(B_i)$.
  - $f^{-1}(\bigcap_{i\in I} B_i) = \bigcap_{i\in I} f^{-1}(B_i)$.
  - $f^{-1}(B - Y) = A - f^{-1}(Y)$.
  - $X \subseteq f^{-1}(f(X))$ i la igualtat es compleix si f és injectiva.
  - $f(f^{-1}(Y)) \subseteq Y$ i la igualtat es compleix si f és exhaustiva.

  Observem que, en certa manera, f^{-1} té millors propietats que f.

9 El conjunt quocient

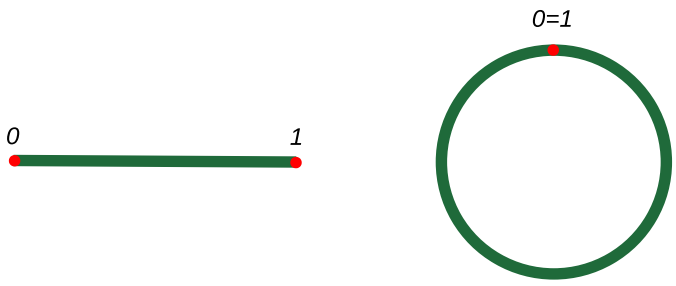
En aquest capítol estudiarem una de les eines més poderoses de les matemàtiques, *la que ens permet considerar com iguals coses que no ho són!*

La força d'aquesta capacitat de fer iguals coses que no són iguals és immensa i l'estudiant de matemàtiques ho anirà descobrint al llarg dels seus estudis. Posem uns exemples informals:

- Si, en els nombres reals, decretem $2\pi = 0$, ja no tenim els nombres reals: tenim els *angles*.

  $$\frac{\pi}{2} \neq \frac{5\pi}{2} \text{ com a nombres reals, però } \frac{\pi}{2} = \frac{5\pi}{2} \text{ com a angles.}$$

- Si considerem l'interval real $[0,1]$ i decretem $1 = 0$, ja no tenim un segment: tenim una *circumferència*.



- Si considerem dos individus iguals quan les seves cèl·lules vermelles tenen a la membrana els mateixos antígens, deixem de tenir individus: tenim *grups sanguinis*. Hem fet una *classificació*.

- Si, en els nombres reals, decretem $24 = 0$ tenim les hores del dia, amb les quals, per cert, podem fer operacions com la suma, perfectament coherents.

- Si en un quadrat (ple), per exemple $[0,1]\times[0,1] \subseteq \mathbb{R}^2$, identifiquem cada punt d'un dels costats amb el punt corresponent del seu costat oposat ($(x,0) = (x,1)$ per tot $x \in [0,1]$), obtenim un *cilindre*. Si a continuació decretem que tots els punts d'un altre costat són el mateix punt ($(0,y) = (0,y')$ per tot $y, y' \in [0,1]$), obtenim *un con*.

- Si en un rectangle (ple), per exemple $[0,2]\times[-1,1]\subset\mathbb{R}^2$, identifiquem cada punt d'un dels costats, no amb el punt corresponent del seu costat oposat sinó amb el seu simètric ($(0,y)=(2,-y)$ per tot $y\in[-1,1]$), obtenim: una *banda de Möbius*.
- Fins aquí hem obtingut coses que ja coneixíem. El procés és més interessant quan obtenim objectes matemàtics *nous*. Per exemple, què són les direccions de l'espai? No ho sabem, però tenim clar que una direcció és el que tenen en comú dues rectes paral·leles. Per tant, si decretem que les rectes paral·leles passen a ser iguals obtenim... les *direccions de l'espai*.

És clar que per poder fer tota aquesta màgia necessitem un fonament teòric sòlid. Aquest fonament vindrà donat per les **relacions d'equivalència**.

Recordem que una relació d'equivalència $\sim$ en un conjunt X és una relació que compleix tres propietats que també compleix la igualtat: reflexivitat, simetria i transitivitat. A partir de X i $\sim$ construirem un nou conjunt que anomenarem $X/\sim$ en el qual la relació $x\sim y$ a X es convertirà en una relació d'igualtat $x=y$ a $X/\sim$.

Suposem, doncs, que X és un conjunt i tenim una relació d'equivalència a X. Per cada $x\in X$, denotem $[x]$ el conjunt de tots els elements de X que estan relacionats amb x:

$$[x]:=\{y\in X : y\sim x\}\subseteq X.$$

Direm que $[x]\subseteq X$ és *la classe d'equivalència* de x o que x és un *representant* de la classe d'equivalència $[x]$. Observem que cada classe d'equivalència és un *subconjunt* de X. Les classes d'equivalència compleixen aquestes propietats (exercici):

1. $[x]=[y]$ si i només si $x\sim y$.
2. La unió de totes les classes d'equivalència és igual a X.
3. Dues classes d'equivalència diferents són disjuntes: $[x]\cap[y]\neq\varnothing\Rightarrow[x]=[y]$.

Doncs bé: els elements del conjunt $X/\sim$ són precisament totes les classes d'equivalència. Més formalment,

$$X/\sim:=\{A\in\mathcal{P}(X) : \exists x\,(x\in X\wedge A=[x])\}.$$

Hem de tenir clar que els elements de $X/\sim$ **no** són elements de X, són *subconjunts* de X. Per exemple, si X és el conjunt de tots els éssers humans —X té, per tant, milers de milions d'elements— i la relació d'equivalència, com en un dels exemples anteriors, és *«tenir el mateix grup sanguini»*, aleshores el conjunt quocient només té quatre elements:

$$X/\sim=\{A,B,AB,O\}$$

on hem designat per A la classe d'equivalència de tots els éssers humans amb grup sanguini A, etc.

Quina relació hi ha entre el conjunt original X i el nou conjunt $X/\sim$ que hem creat? Hi ha una aplicació:

$$\pi : X \to X/\sim$$

que anomenarem *«pas al quocient»* o *«projecció (canònica) sobre el quocient»* perquè, efectivament, és exhaustiva. L'aplicació π es defineix així:

$$\pi(x) := [x] \in X/\sim .$$

En l'exemple dels grups sanguinis, π assignaria a cada individu el seu grup.

Exemple: la banda de Möbius

Considerem el quadrat $Q := [0,2] \times [-1,1]$. És a dir,

$$Q = \{(a,b) \in \mathbb{R}^2 : 0 \leq a \leq 2,\ -1 \leq b \leq 1\}.$$

Abans hem vist que la banda de Möbius s'obté identificant $(0,y) \sim (2,-y)$. Aquesta identificació, per ella mateixa, no és una relació d'equivalència a Q, però no costa gens convertir-la en una relació d'equivalència afegint-hi algunes altres identificacions trivials:

$$(a,b) \sim (a',b') \Leftrightarrow \begin{cases} (a,b) = (a',b') \\ \vee \\ (a=0) \wedge (a'=2) \wedge (b'=-b) \\ \vee \\ (a=2) \wedge (a'=0) \wedge (b'=-b) \end{cases}$$

Aleshores, observem que hi ha dos tipus de classes d'equivalència:

- Si $a \neq 0,2$, la classe d'equivalència $[(a,b)]$ té un únic element.
- En cas contrari, la classe d'equivalència $[(a,b)]$ té dos elements. Per exemple

$$[(0,1/2)] = \{(0,1/2),(2,-1/2)\}.$$

El conjunt quocient $X/\sim$ és, geomètricament, una banda de Möbius.[1]

[1] Observem, però, que aquesta construcció amb un quadrat no es pot fer amb paper i cinta adhesiva, necessitem un rectangle més llarg que ample. És a dir, aquesta banda de Möbius que hem construït és un objecte *abstracte*, no un objecte físic. I està molt bé que sigui així: l'operació de pas al quocient permet construir objectes matemàtics que, en principi, no «viuen» al nostre espai físic tridimensional. Recordem, per exemple, que a l'inici d'aquest capítol vam dir, d'una manera informal, que si a l'interval real $[0,1]$ identifiquem 0 amb 1 obtenim la circumferència. Això és cert, en un cert sentit molt precís, però també és cert que aquesta *circumferència* que hem obtingut no és *«el conjunt de punts del pla a distància 1 d'un punt fixat»*, sinó que és una circumferència *abstracta*.

Definir aplicacions sobre un conjunt quocient

Una situació molt habitual és que ens interessi definir una aplicació sobre un conjunt quocient

$$f : X/\sim \longrightarrow Y.$$

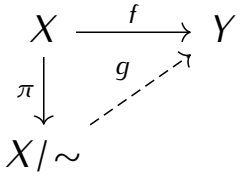
Ho podem fer? Com? Considerem aquests dos exemples.

- Suposem que als nombres reals considerem una relació d'equivalència segons la qual $x \sim x + 2\pi$ per tot x. El conjunt quocient el podem identificar a una circumferència: $S^1 = \mathbb{R}/(x \sim x + 2\pi)$. Com podem definir una funció $f : S^1 \to \mathbb{R}$? La resposta és clara: n'hi ha prou amb definir f sobre els nombres reals i, evidentment, assegurar-se que $f(x + 2\pi) = f(x)$. És a dir, una funció periòdica de període 2π dóna lloc immediatament a una funció sobre la circumferència.
- En l'exemple dels grups sanguinis, si suposem que hi ha diversos tractaments possibles, quan direm que el tractament que rep un individu només depèn del grup sanguini? Doncs quan sempre que dos individus tenen el mateix grup sanguini, reben el mateix tractament.

Escrivim això (que és tan evident!) en un llenguatge matemàtic formal (i deixarà de semblar evident):

> *Sigui $\sim$ una relació d'equivalència sobre un conjunt X i sigui $\pi : X \to X/\sim$ el corresponent pas al quocient. Suposem que tenim una aplicació $f : X \to Y$. Aleshores, la condició necessària i suficient perquè existeixi una aplicació $g : X/\sim \longrightarrow Y$ tal que $g \circ \pi = f$ és*
>
> $$\forall x, y \in X : x \sim y \Rightarrow f(x) = f(y).$$

Un diagrama que és útil per entendre millor la situació és aquest:



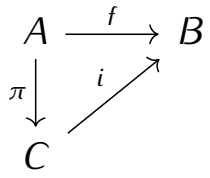
Aquest resultat s'aplica freqüentment a la pràctica matemàtica. Es vol definir una aplicació $g : X/\sim \longrightarrow Y$ i el que es fa és definir una funció $f : X \to Y$. A continuació, es diu la frase *«demostrem que la funció està ben definida»* i aleshores es comprova la condició necessària i suficient del teorema anterior, és a dir, es comprova que $x \sim y \Rightarrow f(x) = f(y)$. Veurem molts exemples al llarg del curs.

Com convertir qualsevol aplicació en injectiva

Suposem que tenim una aplicació $f : A \to B$. Potser no és injectiva. El que és fantàstic és que la podem *«convertir»* en una aplicació injectiva. Exactament, el que volem dir és això:

> **Teorema.** *Sigui $f : A \to B$ una aplicació qualsevol. Podem descompondre f com a composició de dues aplicacions*
>
> $$f = i \circ \pi$$
>
> *de manera que π és exhaustiva i i és injectiva.*

Fent un diagrama s'entendrà millor:

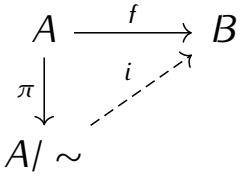


Pensem com podem convertir f en injectiva. Si f no és injectiva, vol dir que hi ha elements diferents x, y que tenen la mateixa imatge. Si *decretem* que, automàticament, x, y passin a ser *iguals*, ja haurem convertit f en una aplicació injectiva. Com podem fer que dues coses diferents com x, y es converteixin en iguals? Resposta: utilitzant el conjunt quocient!

Demostració del teorema. Definim a A la següent relació

$$x \sim y \Leftrightarrow f(x) = f(y).$$

És molt senzill veure que es tracta d'una relació d'equivalència. Sigui $\pi : A \to A/\sim$ el pas al quocient (que és una aplicació exhaustiva). Tenim



Aplicant el que hem vist abans, l'aplicació i existeix i és evident que i és injectiva:

$$i([x]) = i([y]) \;\Rightarrow\; i\pi(x) = i\pi(y) \;\Rightarrow\; f(x) = f(y) \;\Rightarrow\; x \sim y \;\Rightarrow\; [x] = [y].$$

10 | Finit, infinit, infinits

En aquests apunts han aparegut sovint les paraules **finit** i **infinit**, però sempre a nivell de *metallenguatge*: mai hem definit què volen dir, exactament. En aquest capítol tractarem aquestes qüestions importants:

- Què vol dir, exactament, que un conjunt és finit o és infinit?
- Com podem comparar la mida de dos conjunts?
- De la mateixa manera que hi ha conjunts finits més grans que altres, hi ha també «mides» diferents d'infinits? Hi ha infinits més «grans» que altres infinits?

Aquestes preguntes se les va plantejar Georg Cantor cap al 1873 i les respostes que va trobar van fascinar —i també van pertorbar— els matemàtics d'aquella època.

Abans de començar a contestar les preguntes anteriors, és pertinent fer una observació important:

> **«Infinit» no és un objecte matemàtic.**

És a dir, 0, π, $\mathbb{N}$, la funció exponencial, el dodecàedre regular... són objectes matemàtics, però *infinit* (o ∞) **no ho és**. Infinit és un concepte general que apareix sovint a la matemàtica i que, en cada context, tindrà un sentit específic que caldrà fixar clarament. Per exemple, en aquestes fórmules

$$|\mathbb{N}| = \infty, \quad \lim_{x\to\infty}(1 - e^x) = -\infty, \quad \sum_{n=1}^{\infty} \frac{1}{n} = \infty, \quad \int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}$$

hi apareix set vegades el símbol ∞ precedit, en algun cas, del signe menys. En cadascuna de les fórmules, el significat matemàtic del símbol ∞ requereix una explicació, una definició.

Dit això, diguem que l'objectiu d'aquest capítol és donar un sentit precís a expressions del tipus: *el conjunt A és infinit*.

Com comparar dos conjunts sense saber comptar

Imaginem que tenim un ramat de cabres i un ramat d'ovelles. Suposem que no sabem comptar —com realment succeirà quan parlem de conjunts infinits. Com podem saber si tenim la mateixa quantitat de cabres que d'ovelles? Resposta: fem parelles (cabra, ovella) i mirem si sobren cabres o sobren ovelles o no sobra res. D'aquesta manera podem saber, sense haver de comptar, si els dos conjunts tenen la mateixa mida o no.

Aquesta idea tan simple, quan es formula matemàticament com va fer Cantor, obre tot un nou món de possibilitats per tractar el tema de l'infinit i dels infinits.

> *Dos conjunts A i B direm que són* ***equipotents*** *—o que tenen la mateixa* ***cardinalitat*** *o que tenen el mateix* ***cardinal***[1] *— si existeix una aplicació bijectiva $f : A \to B$.*
>
> *Si no existeix cap aplicació bijectiva entre A i B, però sí que existeix una aplicació injectiva $g : A \rightarrowtail B$, direm que el conjunt A té cardinal menor que el conjunt B.*[2]
>
> *Direm que un conjunt A és* ***finit*** *quan no és equipotent a cap subconjunt propi (és a dir, diferent de A). En cas contrari, direm que és* ***infinit***.[3]
>
> *Direm que un conjunt és* ***numerable*** *quan és equipotent a $\mathbb{N}$.*

És així de senzill. Si al conjunt $A = \{0, 3, 6, 7\}$ li traiem ni que sigui un sol element, obtenim un nou conjunt que no es pot posar mai en correspondència bijectiva amb A. És un conjunt finit. El conjunt $\mathbb{N}$, en canvi, és un conjunt infinit perquè, per exemple, l'aplicació $f(n) = 2n$ és una bijecció entre $\mathbb{N}$ i el subconjunt

[1] Hem definit *«tenir el mateix cardinal»* sense haver definit què és el cardinal. Això és perfectament legal. Si existís el conjunt de tots els conjunts, podríem definir el cardinal d'un conjunt com la seva classe d'equivalència respecte de la relació d'equivalència de ser equipotents. És a dir, el cardinal és allò que tenen en comú tots els conjunts equipotents. Però això, és clar, no és una definició matemàtica correcta, perquè el conjunt de tots els conjunts no existeix.

[2] Perquè aquesta noció de «menor que» sigui coherent amb la idea d'ordre hauríem de comprovar que compleix la propietat antisimètrica —les altres propietats d'una relació d'ordre són evidents— o, més exactament, una versió adaptada de la propietat antisimètrica. Hauríem de demostrar que si hi ha una aplicació injectiva $A \to B$ i una aplicació injectiva $B \to A$, aleshores també hi ha una aplicació *bijectiva* $A \to B$ i, per tant, A i B són equipotents. Això és cert i va ser enunciat per Cantor, sense demostració. Es coneix com a *teorema de Cantor-Schröder-Bernstein*.

[3] Aquesta definició va ser proposada per Richard Dedekind l'any 1888 i té l'avantatge que no requereix la teoria dels nombres naturals. Una altra definició de conjunt finit seria aquesta: *un conjunt és finit si és equipotent a algun d'aquests conjunts* $\varnothing$, $\{1\}$, $\{1, 2\}, \ldots, \{1, \ldots, n\}, \ldots$. A la pràctica, utilitzem més freqüentment aquesta segona definició que no pas la de Dedekind, però hem de tenir en compte que demostrar que les dues definicions són equivalents requereix l'axioma de l'elecció.

dels nombres parells. Quan un conjunt es pot posar en bijecció amb un subconjunt propi seu, es tracta d'un conjunt infinit.

Coneixem, doncs, molts conjunts infinits perquè qualsevol conjunt que contingui els nombres naturals ha de ser infinit: $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, les funcions reals, els polinomis, etc. són infinits i el producte cartesià d'un conjunt infinit i un conjunt no buit també és infinit. Alguns d'aquests conjunts infinits són clarament numerables. Per exemple:

- $\mathbb{Z}$ és numerable i el podem numerar —que vol dir posar en correspondència bijectiva amb $\mathbb{N}$— així

$$0, 1, -1, 2, -2, 3, -3, 4, -4, 5, -5, \ldots$$

- $\mathbb{Q}$ també és numerable i aquest fet pot semblar sorprenent al primer moment perquè l'aspecte que tenen els nombres racionals quan els dibuixem sobre una recta en coordenades cartesianes és molt diferent de l'aspecte que tenen els nombres naturals sobre la mateixa recta. A més, sembla talment que hi hagi molts més nombres racionals que nombres naturals. La realitat, però, és que els conjunts $\mathbb{N}$ i $\mathbb{Q}$ són equipotents i hem de dir que tenen la mateixa cardinalitat. Una manera de numerar els nombres racionals és aquesta. Escrivim les fraccions positives en una taula de doble entrada (hi haurà repeticions):

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

  Aleshores, comencem a numerar per l'extrem superior esquerre i seguim fent antidiagonals, de manera que quan ens trobem amb una fracció que representi un nombre racional que ja hem posat a la llista, la saltem:

$$1/1,\ 1/2,\ 2/1,\ 1/3,\ 3/1,\ 1/4,\ 2/3,\ 3/2,\ 4/1,\ 1/4, \ldots$$

  D'aquesta manera, hem numerat tots els racionals positius; afegim ara el zero a l'inici i intercalem positius i negatius com hem fet amb els enters. Tenim numerats tots els nombres racionals.

- Si A i B són conjunts numerables, també $A \times B$ ho és. En efecte, si tenim $A = \{a_0, a_1, a_2, \ldots\}$ i $B = \{b_0, b_1, b_2, \ldots\}$, podem numerar $A \times B$ així:

$$A \times B = \{(a_0, b_0), (a_0, b_1), (a_1, b_0), (a_0, b_2), (a_1, b_1), (a_2, b_0), \ldots\}.$$

Infinits més grans que altres

En primer lloc diguem que l'**infinit numerable**, és a dir, la cardinalitat de $\mathbb{N}$, és l'infinit més petit que hi ha. Dit ben dit, estem dient que

Tot subconjunt de $\mathbb{N}$ és o finit o numerable.

Demostració: Utilitzarem aquests dos fets:

- Tot subconjunt no buit de $\mathbb{N}$ té mínim. Això es demostra fàcilment per inducció (exercici **II**.25).
- Si $0 = a_0 < a_1 < a_2 < \cdots$ és una successió creixent de nombres naturals i $k \in \mathbb{N}$, existeix i tal que $a_i \leq k < a_{i+1}$. La demostració és senzilla (exercici **II**.26).

Demostrem ara el teorema. Suposem que $A \subseteq \mathbb{N}$ és un subconjunt infinit. Hem de demostrar que hi ha una aplicació bijectiva $f : \mathbb{N} \to A$. Definirem f de manera recursiva:[4]

$$f(0) := \min(A)$$
$$f(n+1) := \min(A - \{f(0), \ldots, f(n)\})$$

Observem que, amb aquesta definició, tenim una successió creixent

$$f(0) < f(1) < f(2) < \cdots$$

Això implica que la funció f és injectiva. També és exhaustiva. En efecte, sigui $k \in A$. Aleshores, existirà n tal que $f(n) \leq k < f(n+1)$. Per la definició de $f(n+1)$ veiem que $k \in \{f(0), \ldots, f(n)\}$ i deduïm $k = f(n)$.

A continuació demostrem que hi ha infinits més grans que l'infinit numerable:

$\mathbb{R}$ *no és numerable.*

És a dir, la cardinalitat de $\mathbb{R}$ —que es diu que és la cardinalitat *del continu*— és més gran que la cardinalitat de $\mathbb{N}$. La demostració (deguda a Cantor) és espectacularment senzilla. Pel teorema anterior, n'hi ha prou amb demostrar que $\mathbb{R}$ té algun subconjunt no numerable. Demostrarem per reducció a l'absurd que l'interval $[0, 1)$ no és numerable. Suposem que sí que ho fos i tinguéssim una llista de tots els nombres reals de l'interval, escrits en la seva forma decimal. Tindria aquesta forma:

$$a_0 = 0.a_{01}a_{02}a_{03}a_{04}a_{05}\ldots$$
$$a_1 = 0.a_{11}a_{12}a_{13}a_{14}a_{15}\ldots$$
$$a_2 = 0.a_{21}a_{22}a_{23}a_{24}a_{25}\ldots$$
$$a_3 = 0.a_{31}a_{32}a_{33}a_{34}a_{35}\ldots$$
$$\ldots$$

[4] Les definicions recursives són vàlides a la teoria de conjunts (exercici **II**.24).

Considerem ara el nombre real $b = 0.b_1b_2b_3b_4\ldots \in [0,1)$ definit així:

$$b_i := \begin{cases} 0, & a_{i-1,i} \neq 0 \\ 1, & a_{i-1,i} = 0 \end{cases}$$

És evident que b no pot ser a la llista anterior.

Per tant, hi ha infinits nombres naturals i hi ha infinits nombres reals, però la infinitud dels nombres reals és més gran que la dels nombres naturals. Hi ha infinits encara més grans? I tant!

Per tot conjunt X, la cardinalitat de $\mathcal{P}(X)$ és més gran que la de X.

És a dir, si considerem tots els subconjunts dels reals tenim un infinit encara més gran que el dels reals. La demostració d'aquest teorema és senzillíssima i s'assembla molt a la paradoxa de Russell. Per reducció a l'absurd, suposem que tenim una aplicació bijectiva $f : X \to \mathcal{P}(X)$ de manera que, per cada $x \in X$, $f(x)$ és un subconjunt de X. Considerem aquest subconjunt de X:

$$A := \{x \in X : x \notin f(x)\}.$$

Com que f és exhaustiva, existirà $y \in X$ tal que $A = f(y)$. Preguntem-nos ara si $y \in A$ i arribarem a contradicció.

En particular, el conjunt de subconjunts de $\mathbb{N}$ és no numerable. De fet, $\mathcal{P}(\mathbb{N})$ és equipotent amb $\mathbb{R}$.

La hipòtesi del continu

Hem vist que hi ha una jerarquia de cardinalitats que no s'acaba mai

$$\text{finit } (0,1,2,\ldots) < \text{infinit numerable } (\mathbb{N}) < \text{infinit continu } (\mathbb{R}) < \cdots\cdots$$

i hem vist que entre el finit i el numerable no hi ha cap pas intermedi: tot subconjunt de $\mathbb{N}$ és finit o numerable. Cantor, quan va desenvolupar aquesta teoria dels diversos infinits, es va preguntar

Hi ha algun pas intermedi entre el numerable i el continu?

i es va obsessionar profundament intentant contestar aquesta pregunta, sense aconseguir-ho. És a dir, pot ser que hi hagi un subconjunt dels nombres reals que no sigui ni finit ni numerable ni equipotent als reals?

Hipòtesi del continu: *Tot $A \subseteq \mathbb{R}$ és finit, numerable o equipotent a $\mathbb{R}$.*

Aquesta hipòtesi és.... **Indecidible!**[5] El 1940 Kurt Gödel va demostrar que la hipòtesi del continu és compatible amb ZFC i el 1963 Paul Cohen va demostrar que la negació de la hipòtesi del continu també és compatible amb ZFC. Per tant, podem decidir lliurement si afegim la hipòtesi del continu a la llista d'axiomes de la teoria de conjunts o, en canvi, afegim la negació de la hipòtesi del continu a la llista d'axiomes de la teoria de conjunts. És la mateixa situació que es donava en el cas de l'axioma de l'elecció, però mentre que l'axioma de l'elecció s'accepta de manera força general —perquè és molt útil—, ni la hipòtesi del continu ni la seva negació gaudeixen de la mateixa consideració.

[5] Suposant, és clar, que la teoria de conjunts sigui consistent.

Exercicis de teoria de conjunts

1. Sigui X un conjunt i $A, B \subseteq X$ dos subconjunts. Utilitzem la notació U^c per indicar el complementari a X d'un subconjunt U, és a dir, $U^c := \{x \in X : x \notin U\}$, i recordem que $A \setminus B := A \cap B^c$. Per a cadascuna de les següents afirmacions, doneu una demostració si és correcta o un contraexemple si és falsa.

   (a) $(A \cap B) \cup (A \cap B^c) = A$.
   (b) $A \cap (A^c \cup B) = A \cap B$.
   (c) $A \subseteq B$ si i només si $B^c \subseteq A^c$.
   (d) $A = (A \cap B) \cup (A \setminus B)$.
   (e) $A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$.
   (f) $A \setminus B = B^c \setminus A^c$.

2. Decidiu si les següents afirmacions són certes o falses:

   (a) $\{a, b\} \in \{a, \{a, b\}\}$.
   (b) $\{a, b\} \subseteq \{a, \{a, b\}\}$.
   (c) $\{a, b\} \in \{a, b, \{a, b\}\}$.
   (d) $\{a, b\} \subseteq \{a, b, \{a, b\}\}$
   (e) $\varnothing \subseteq \{\varnothing\}$.
   (f) $\varnothing \in \{\varnothing\}$.
   (g) $\{\varnothing\} \subseteq \varnothing$.
   (h) $\{\varnothing\} \in \{\varnothing\}$.

3. Demostreu que no hi ha cap conjunt els elements del qual siguin tots els conjunts finits. (Cal utilitzar l'axioma de fundació de la pàgina 61.)

4. Demostreu que les parelles ordenades, tal com les hem definit, compleixen que $(a, b) = (a', b')$ si i només si $a = a'$ i $b = b'$. Discutiu si podríem definir una *terna ordenada* (a, b, c) amb la fórmula

$$(a, b, c) := \big\{\{a\}, \{a, b\}, \{a, b, c\}\big\}.$$

5. Demostreu que si F i G són subconjunts de E, aleshores: (a) $F \subseteq G \Leftrightarrow F \cup G = G$; (b) $F \subseteq G \Leftrightarrow F^c \cup G = E$; (c) $F \subseteq G \Leftrightarrow F \cap G = F$; (d) $F \subseteq G \Leftrightarrow F \cap G^c = \varnothing$.

6. Comproveu que si els conjunts A_i $i \in \mathbb{N}$ són finits i no buits, i si $A_i \supseteq A_{i+1}$ per a tot i, aleshores $\bigcap_{i \in \mathbb{N}} A_i \neq \varnothing$. Demostreu que la conclusió pot ser falsa si suposem els conjunts A_i infinits.

7. Sigui E un conjunt, i A i B dos subconjunts. Determineu tots els subconjunts $X \subseteq E$ tals que (a) $A \cup X = B$; (b) $A \cap X = B$.

8. Siguin a, b, c, d conjunts diferents i considerem els conjunts $A = \{a, b, c\}$ i $B = \{a, b, d\}$. Descriviu els conjunts $(A \times A) \cap (B \times B)$ i $A \times \varnothing$.

9. Siguin E, F, G tres conjunts. Demostreu que $(E \times G) \cup (F \times G) = (E \cup F) \times G$.

10. Descriviu els elements de de $\mathcal{P}(\mathcal{P}(\{a, b\}))$ i els de $\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathcal{P}(\varnothing))))$.

11. Donat un conjunt E i un subconjunt $A \subseteq E$, és cert que $\mathcal{P}(A) \cap \mathcal{P}(E \setminus A) = \varnothing$?

12. Siguin X i Y dos conjunts. Demostreu que $X \subseteq Y$ si i només si $\mathcal{P}(X) \subseteq \mathcal{P}(Y)$.

13. Sigui U un conjunt i sigui $A = \mathcal{P}(U)$. Podem considerar una estructura algebraica[6] a A amb dues operacions commutatives, associatives i distributives $a + b := a \cup b$, $ab := a \cap b$, dues constants $0 := \varnothing \in A$, $1 := U \in A$ i una conjugació $a' := U \setminus a$ que és idempotent, és a dir $a'' = a$ per tot $a \in A$. Comproveu aquestes propietats:

    (a) $a + 0 = a$, $a + a = a$, $a1 = a$, $a0 = 0$, $aa = a$.

    (b) $a(a + b) = a$, $a + ab = a$.

    (c) $aa' = 0$, $a + a' = 1$, $(a + b)' = a'b'$, $(ab)' = a' + b'$.

14. Sigui E un conjunt no buit. Si A i B son dos subconjunts de E definim

$$A * B = (A \cap B) \cup (A^c \cap B^c),$$

    on A^c designa el complementari de A en E. Això defineix una operació a $\mathcal{P}(E)$. Traduïu aquesta operació al llenguatge de l'exercici anterior i treballeu, a partir d'ara, en aquell llenguatge. Suposem $a, b, c \in \mathcal{P}(E)$.

    (a) Il·lustreu amb una figura el conjunt $a * b$. Observeu que $a * b = b * a$.

    (b) Demostreu que $a * (b * c) = (a * b) * c$.

    (c) Calculeu $a * 1$, $a * 0$ i $a * a$.

    (d) Demostreu $a * b = a' * b'$.

    (e) Demostreu $(a * b)' = a * b' = a' * b$.

    (f) Resoleu l'equació d'incògnita x: $a * x = b$.

15. Sigui A un conjunt amb n elements. Demostreu que $\mathcal{P}(A)$ té 2^n elements. Per cada $0 \leq k \leq n$, definim $\binom{n}{k}$ com el nombre de subconjunts de k elements de A. Demostreu

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n, \quad \binom{n}{k} = \binom{n}{n-k}.$$

    Demostreu

$$\binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k}$$

    i, a partir d'aquí, justifiqueu el *triangle de Tartaglia* i demostreu per inducció la fórmula clàssica per calcular els nombres combinatoris.

16. Els *nombres de Bell* B_n es defineixen per aquesta fórmula recursiva

$$B_0 = 1, \quad B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k.$$

[6] Aquest tipus d'estructura es coneix com *àlgebra de Boole*.

Una *partició* d'un conjunt finit X és una descomposició $X = A_1 \cup \cdots \cup A_k$ com unió de subconjunts disjunts dos a dos. Demostreu per inducció que B_n és igual al nombre de particions del conjunt $\{1, \ldots, n\}$. (Indicació: classifiqueu les particions de $\{1, \ldots, n+1\}$ segons la mida del subconjunt que contingui $n+1$.)

17. Considereu aquesta relació a $\mathbb{N}$: $n \vdash m$ si n és un divisor de m. Demostreu que és una relació d'ordre parcial. Sigui $A := \mathbb{N} \setminus \{0, 1\}$. Determineu el màxim, el mínim, els elements maximals, els elements minimals i les cotes superiors i inferiors de A, si existeixen.

18. Considerem a $\mathcal{P}(\mathbb{N})$ la relació d'ordre parcial donada per la inclusió. Sigui $A \subsetneq \mathcal{P}(\mathbb{N})$ el conjunt de tots els subconjunts cofinits de $\mathbb{N}$, és a dir, els conjunts tals que els seu complement és finit. Trobeu, si existeixen, el màxim, el mínim, els elements maximals i minimals i les cotes superior i inferiors de A.

19. Comenteu aquest text: *«a la definició de relació d'equivalència, exigir la propietat reflexiva és superflu perquè es dedueix de les propietats transitiva i simètrica. Efectivament, la propietat simètrica en diu que si $x \sim y$ també tenim $y \sim x$ i aleshores, aplicant la propietat transitiva deduïm $x \sim x$».*

20. Hi ha un teorema que diu que «tota successió acotada de nombres reals té una parcial convergent». Definiu «successió» i «parcial».

21. De vegades interessa considerar conjunts (finits) amb elements repetits. S'anomenen *multiconjunts*. Doneu una definició de multiconjunt. Demostreu que els multiconjunts de nombres naturals formen un conjunt.

22. Demostreu les propietats de f i f^{-1} que apareixen a la pàgina 71 i vegeu que, en general, les inclusions no es poden substituir per igualtats.

23. Determineu si les aplicacions següents són injectives, exhaustives o bijectives:

    (a) $f : \mathbb{N} \to \mathbb{N}$, $f(n) = n + 1$.

    (b) $g : \mathbb{Z} \to \mathbb{Z}$, $g(n) = n + 1$.

    (c) $h : \mathbb{R}^2 \to \mathbb{R}^2$, $h((x, y)) = (x + y, x - y)$. (Normalment s'escriu $h(x, y)$.)

    (d) $k : \mathbb{R} \setminus \{1\} \to \mathbb{R}$, $k(x) = (x + 1)/(x - 1)$.

24. Demostreu que, a la teoria de conjunts, és vàlid el teorema de recursió de la pàgina 24. Per fer-ho, seguiu els passos de la demostració de la pàgina 26 substituint «predicats» per «subconjunts de $\mathbb{N} \times \mathbb{N}$».

25. Demostreu, per inducció, que tot subconjunt no buit de $\mathbb{N}$ té mínim.

26. Sigui $0 = a_0 < a_1 < a_2 < \cdots$ una successió creixent de nombres naturals i sigui $k \in \mathbb{N}$. Demostreu que existeix i tal que $a_i \leq k < a_{i+1}$.

27. Utilitzeu el mètode del contraexemple minimal per demostrar que tot natural > 0 és suma de potències de dos diferents.

28. Sigui A un conjunt. Demostreu que hi ha una bijecció entre $\mathcal{P}(A)$ i el conjunt $\mathcal{F}(A, \{0, 1\})$ de totes les aplicacions de A en el conjunt $\{0, 1\}$.

29. Sigui $f : A \to A$ una aplicació. Per cada $n \geq 0$ definim recursivament una aplicació $f^n : A \to A$ per $f^0(a) = a$, $f^{n+1}(a) = f(f^n(a))$ per tot $a \in A$. Demostreu:

    (a) Per tot n, per tot $a \in A$, $f^{n+1}(a) = f^n(f(a))$.

    (b) Si f és injectiva, aleshores f^n és injectiva per tot n.

    (c) Si f és exhaustiva, aleshores f^n és exhaustiva per tot n.

    (d) Si existeix $n > 0$ tal que f^n és injectiva, aleshores f és injectiva.

    (e) Si existeix $n > 0$ tal que f^n és exhaustiva, aleshores f és exhaustiva.

30. Sigui E un conjunt i siguin A i B dos subconjunts de E. Definim:

$$\begin{array}{rcl} f : \mathcal{P}(E) & \longrightarrow & \mathcal{P}(A) \times \mathcal{P}(B) \\ X & \longmapsto & (X \cap A, X \cap B) \end{array}$$

    Demostreu que:

    (a) L'aplicació f és injectiva si i només si $A \cup B = E$.

    (b) L'aplicació f és exhaustiva si i només si $A \cap B = \varnothing$.

    (c) Determineu una condició necessària i suficient sobre A i B per a que f sigui bijectiva. En aquest cas determineu la funció inversa de f.

31. Siguin X i Y dos conjunts i $f : X \to Y$ una aplicació. Demostreu que

    (a) Per a tot $B \subset Y$ es compleix que $f(f^{-1}(B)) = B \cap f(X)$.

    (b) L'aplicació f és exhaustiva si i només si per a tot $B \subseteq Y$ es compleix la igualtat $f(f^{-1}(B)) = B$.

    (c) L'aplicació f és injectiva si i només si per a tot $A \subseteq X$ es compleix $f^{-1}(f(A)) = A$.

    (d) L'aplicació f és bijectiva si i només si per a tot $A \subseteq X$ es compleix $f(A^c) = (f(A))^c$.

32. Siguin X i Y dos conjunts i $f : X \to Y$ una aplicació. Recordem (pàgina 71) que tenim dues aplicacions $f_* : \mathcal{P}(X) \to \mathcal{P}(Y)$ i $f^* : \mathcal{P}(Y) \to \mathcal{P}(X)$. Demostreu que f és injectiva si i només si f_* és injectiva i f és exhaustiva si i només si f^* és injectiva.

33. Siguin $f : X \to Y$, $g : Y \to X$ aplicacions.

    (a) És cert que $g \circ f = \mathrm{id}_X$ implica $f \circ g = \mathrm{id}_Y$?

    (b) Pot existir més d'una aplicació $g : Y \to X$ tal que $g \circ f = \mathrm{id}_X$?

    (c) Pot existir més d'una aplicació $g : Y \to X$ tal que $f \circ g = \mathrm{id}_Y$?

34. Siguin X i Y dos conjunts no buits i $f : X \to Y$ una aplicació. Demostreu que f és injectiva si i només si existeix una aplicació $g : Y \to X$ tal que $g \circ f = \mathrm{id}_X$. Demostreu que f és exhaustiva si i només si existeix una aplicació $g : Y \to X$ tal que $f \circ g = \mathrm{id}_Y$. (Atenció: cal utilitzar l'axioma de l'elecció.)

35. Sigui X un conjunt no buit.

    (a) Si $U \subsetneq X$, considereu l'aplicació $h : \mathcal{P}(X) \to \mathcal{P}(U)$ definida com $h(A) := A \cap U$. És injectiva? És exhaustiva?

(b) Sigui $f : X \to \mathcal{P}(\mathcal{P}(X))$ l'aplicació $f(x) := \{A \subseteq X : x \in A\}$. És injectiva? És exhaustiva?

(c) Sigui $\sim$ una relació d'equivalència a X i considerem l'aplicació de pas al quocient $\pi : X \to X/\sim$. Considereu l'aplicació $\pi^{-1} : \mathcal{P}(X/\sim) \to \mathcal{P}(X)$. És injectiva? És exhaustiva?

36. Siguin X, Y, Z tres conjunts i $f : X \to Y$ i $g : Y \to Z$ dues aplicacions. Decidiu si aquestes implicacions són certes o falses:

(a) $g \circ f$ injectiva $\Rightarrow f$ injectiva.

(b) $g \circ f$ injectiva $\Rightarrow g$ injectiva.

(c) $g \circ f$ exhaustiva $\Rightarrow f$ exhaustiva.

(d) $g \circ f$ exhaustiva $\Rightarrow g$ exhaustiva.

37. Considereu aquesta relació a $\mathbb{N}$:

$$n \sim m \iff \exists\, a, b \in \mathbb{N},\ a, b > 0,\ a^2 n = b^2 m.$$

Demostreu que és una relació d'equivalència. Descriviu les classes $[0]$ i $[1]$.[7] Demostreu que hi ha infinites classes d'equivalència. Què canvia si substituïm $\mathbb{N}$ per $\mathbb{R}$?

38. Sigui X un conjunt i sigui $\sim$ una relació a X. Doneu una definició de «la relació d'equivalència més petita que conté la relació $\sim$». Expliciteu aquesta relació d'equivalència en el cas que $\sim$ sigui: (a) la relació d'ordre habitual de $\mathbb{N}$; (b) la relació $0 \sim 1$ a l'interval $X = [0, 1]$.

39. Quantes relacions d'equivalència es poden definir en un conjunt de n elements? (Utilitzeu l'exercici 16.)

40. (a) Sigui $S^1 := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ la circumferència unitat del pla $\mathbb{R}^2$. Considereu l'aplicació $f : S^1 \to S^1$ donada per $f(x, y) := (x^2 - y^2, 2xy)$. Demostreu que f està ben definida. Escriviu f en coordenades polars. Utilitzeu la teoria del conjunt quocient per descompondre f en $f = gp$ amb p exhaustiva i g bijectiva.

(b) Sigui $\mathcal{R}$ el conjunt de totes les rectes del pla $\mathbb{R}^2$. Definim una relació d'equivalència a $\mathcal{R}$ dient que $p \sim q$ si p i q són rectes paral·leles. Trobeu una bijecció entre el conjunt quocient $\mathcal{R}/\sim$ i la circumferència S^1.

41. A $X = \mathbb{R}^2 - \{(0, 0)\}$ definim

$$(x, y) \sim (x', y') \iff (x = x' \neq 0) \vee (x = x' = 0,\ yy' > 0).$$

Demostreu que és una relació d'equivalència i doneu una descripció del conjunt quocient $X/\sim$.

42. Sigui $\mathcal{F}_n$ el conjunt de totes les FBF de la lògica proposicional que només contenen proposicions del conjunt $X := \{A_1, \ldots, A_n\}$. Considereu la relació d'equivalència $\equiv$ al conjunt $\mathcal{F}_n$ i sigui $F_n := \mathcal{F}_n/\equiv$ el conjunt quocient. Calculeu el nombre d'elements del conjunt F_n. (Utilitzeu l'exercici I.17.)

[7] En aquest exercici cal aplicar les propietats de la descomposició en nombres primers, que estudiarem més endavant.

43. Determineu quins d'aquests conjunts són numerables i quin no ho són:

    (a) Els polinomis $p(x)$ amb coeficients enters.

    (b) El conjunt de punts del pla amb coordenades racionals.

    (c) El conjunt de totes les fórmules ben fetes de la lògica proposicional.

    (d) El conjunt de totes les funcions reals contínues.

    (e) El conjunt de totes les funcions $\{0,1\} \to \mathbb{N}$.

    (f) El conjunt de totes les funcions $\mathbb{N} \to \{0,1\}$.

    (g) El conjunt de tots els subconjunts finits de $\mathbb{N}$.

    (h) El conjunt de punts sobre la circumferència unitat del pla.

    (i) El conjunt de tots els nombres reals algebraics. (Un nombre algebraic és el que és solució d'algun polinomi amb coeficients enters.)

44. Utilitzeu l'últim apartat de l'exercici anterior per demostrar que existeixen infinits nombres reals transcendents. (Un nombre transcendent és el que no és algebraic.)[8]

[8] Aquest exercici proporciona un exemple interessant i famós de *demostració no constructiva*: es demostra que hi ha infinits nombres transcendents sense mostrar-ne ni un de sol.

Part III:

Alguns conceptes de teoria de grups

El concepte de grup, malgrat la seva aparent simplicitat, és un dels més importants de tota la matemàtica perquè codifica la idea de simetria. La seva inclusió en un text de fonaments de la matemàtica —al mateix nivell de, per exemple, l'aritmètica— està plenament justificada.

En aquesta tercera part no pretenem fer una teoria de grups però sí que volem que l'aprenent de matemàtic incorpori aquest concepte, des de l'inici dels seus estudis, en el lloc central que realment li correspon. Per parlar de grups —que entendrem sempre com a *grups de transformacions*– treballarem amb dos exemples principals: els grups de permutacions i els grups de simetries d'un objecte geomètric lineal. A través d'aquests exemples anirem veient el significat de conceptes bàsics com subgrups, homomorfismes, conjugació, representacions, grup quocient, etc.

Foto: Felix Klein, 1849–1925

11 | Permutacions

Una **permutació** de n objectes ordenats ($n = 1, 2, \ldots$) és una reordenació d'aquests objectes. Els objectes poden ser els de qualsevol conjunt però és natural considerar, sense pèrdua de generalitat, que els objectes ordenats són precisament els nombres naturals $\{1, 2, \ldots, n\}$. Una manera rigorosa de definir[1] què vol dir una *reordenació* és dir que és una *aplicació bijectiva*

$$\sigma : \{1, 2, \ldots, n\} \longrightarrow \{1, 2, \ldots, n\}.$$

Per determinar una aplicació d'aquestes n'hi ha prou amb conèixer les imatges de cada nombre de 1 a n. Normalment, això s'indica així:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \cdots & \sigma(n) \end{pmatrix}.$$

Per exemple, si $n = 3$ hi ha exactament 6 permutacions possibles:

$$\sigma_0 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \ \sigma_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \ \sigma_2 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix},$$

$$\sigma_3 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \ \sigma_4 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \ \sigma_5 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}.$$

Comencem amb algunes observacions senzilles:

- L'aplicació identitat $\sigma(i) = i$ per tot $1 \leq i \leq n$ també és una permutació. Sovint la denotarem I.
- Com que les permutacions són aplicacions, podem considerar composicions de permutacions (fixant un valor de n, és clar) i aquesta composició de permutacions compleix la propietat associativa. Per tal de respectar la identificació de les permutacions amb aplicacions, la composició de permutacions es llegeix *«de dreta a esquerra»*. Per exemple:

$$\sigma_2\sigma_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} = \sigma_5.$$

[1] Aquí tenim un bon exemple del que dèiem en un capítol anterior: un aspecte molt important de la feina d'un matemàtic consisteix en trobar bones **definicions**. Tothom té clara la idea de reordenar uns objectes però el que cal fer és convertir aquesta idea en una definició matemàtica rigorosa —evidentment!— que, a més, reflecteixi exactament les característiques que ens interessen en la idea original i que permeti treballar-hi matemàticament amb comoditat.

- En general, la composició de permutacions no és una operació commutativa:

$$\sigma_3\sigma_2 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \sigma_4 \neq \sigma_5 = \sigma_2\sigma_3.$$

- El nombre total de permutacions de n objectes és igual a $n!$ Fins i tot podríem considerar el cas de les permutacions de 0 objectes. Segons la definició, una permutació de zero objectes ha de ser una aplicació bijectiva $\varnothing \to \varnothing$. Hi ha una única aplicació del conjunt vuit en ell mateix i aquest fet justifica que *definim* $0! := 1$.

- Com que les permutacions són aplicacions bijectives, podem parlar de la **inversa** d'una permutació σ. La denotarem per σ^{-1} i, evidentment, es compleix $\sigma\sigma^{-1} = \sigma^{-1}\sigma = I$. Trobar la inversa d'una permutació és ben senzill. Per exemple, en el cas de les permutacions de tres elements que hem descrit abans, observem $\sigma_1^{-1} = \sigma_1$ i $\sigma_4^{-1} = \sigma_5$.

- Amb una mica de paciència podem escriure la *«taula de multiplicar»* de les permutacions de 3 elements.

	σ_1	σ_2	σ_3	σ_4	σ_5
σ_1	I	σ_5	σ_4	σ_3	σ_2
σ_2	σ_4	I	σ_5	σ_1	σ_3
σ_3	σ_5	σ_4	I	σ_2	σ_1
σ_4	σ_2	σ_3	σ_1	σ_5	I
σ_5	σ_3	σ_1	σ_2	I	σ_4

  Hem omès la fila i la columna corresponents a $\sigma_0 = I$ perquè són trivials. Llegim la taula de manera que l'entrada de la fila σ_i i la columna σ_j és la permutació $\sigma_i\sigma_j$. La feina d'escriure aquesta taula queda simplificada pel fet que a cada fila i a cada columna no hi pot haver cap repetició (per què?).

  De la informació que hi ha a la taula destaquem aquest fet:

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^3 = \sigma_5^3 = I.$$

- Un fet de caràcter general que convé tenir present és que $(\sigma\tau)(\tau^{-1}\sigma^{-1}) = I$. Per tant,

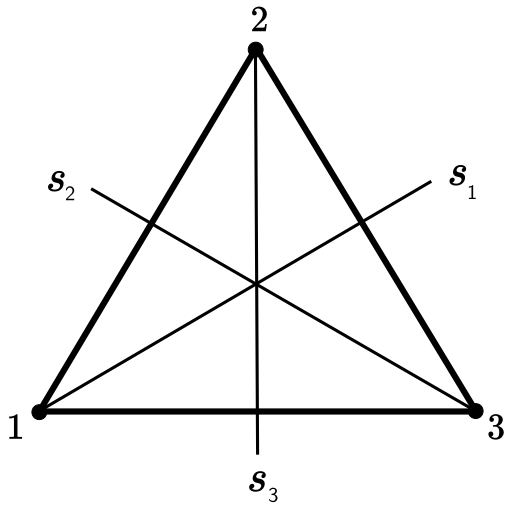
$$(\sigma\tau)^{-1} = \tau^{-1}\sigma^{-1}.$$

  Dit d'una altra manera: *l'invers d'un producte és el producte dels inversos,* ***en l'ordre invers***.[2]

- Les permutacions de 3 objectes $\{I, \sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5\}$ es poden identificar amb les **simetries d'un triangle equilàter**. En efecte, deixant de banda la identitat, un triangle equilàter té aquestes simetries:

[2] ...de la mateixa manera que si volem desfer l'operació de posar-nos els mitjons i a continuació les sabates, ens hem de treure primer les sabates i després els mitjons!

- Les rotacions r_1 i r_2 d'angle 60° i 120° (en sentit positiu), respectivament, amb centre al centre del triangle.
- Les reflexions s_1, s_2, s_3 respecte de les tres altures del triangle.



De fet, cada simetria del triangle permuta els tres vèrtex del triangle i, si numerem els vèrtex com 1,2,3, a cada simetria li correspon una permutació. Concretament:

$$s_1 \mapsto \sigma_1,\ s_2 \mapsto \sigma_2,\ s_3 \mapsto \sigma_3,\ r_1 \mapsto \sigma_5,\ r_2 \mapsto \sigma_4$$

de manera que el producte de dues simetries es correspon amb el producte de les permutacions respectives.

Ordre d'una permutació

Considerem ara permutacions de n objectes, per qualsevol valor de n. Si σ és una permutació, podem considerar aquesta successió infinita de permutacions

$$I, \sigma, \sigma^2, \sigma^3, \sigma^4, \ldots, \sigma^n, \ldots$$

Com que de permutacions diferents només n'hi ha una quantitat finita —exactament $n!$— és clar que en aquesta successió infinita de permutacions hi haurà repeticions $\sigma^k = \sigma^{k+s}$ per alguns valors de k, s. Aleshores, si multipliquem els dos costat d'aquesta igualtat per σ^{-k} obtenim $\sigma^s = I$. És a dir, si una permutació σ la repetim un cert nombre de vegades aconseguirem que els n objectes tornin a estar en el seu ordre natural.

Definim l'**ordre** d'una permutació σ com el mínim $r > 0$ tal que $\sigma^r = I$. Podem fer un parell d'observacions:

- Si l'ordre de σ és r, aleshores $\sigma\sigma^{r-1} = I$ i, per tant, $\sigma^{-1} = \sigma^{r-1}$. Més en general,
$$\sigma^{-j} = \sigma^{r-j}.$$

- Si l'ordre de σ és r, aleshores les permutacions

$$I, \sigma, \sigma^2, \ldots, \sigma^{r-1}$$

formen una llista completa i sense repeticions de totes les potències de σ (entenent $\sigma^0 := I$). En efecte, si tenim una potència σ^n, fem la divisió amb residu $n = kr + s$ amb $0 \leq s < r$ (vegeu la pàgina 131) i, aleshores, és fàcil veure que $\sigma^n = \sigma^s$.

Transposicions

Una permutació és una **transposició** si deixa fixos tots els elements menys dos. Per exemple

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 5 & 4 & 3 \end{pmatrix}$$

és una transposició que, per simplificar, indicaríem amb la notació $(3, 5)$ que posa de manifest quins són els dos objectes que es permuten.

Observem que les transposicions tenen ordre dos i, en conseqüència, la inversa d'una transposició és ella mateixa

$$\tau \text{ transposició } \Rightarrow \tau^2 = I,\ \tau = \tau^{-1}.$$

Cicles

Un altre tipus interessant de permutacions són els **cicles**. Abans de donar la definició de què és un cicle, donarem un exemple: el cicle $\sigma = (2, 3, 1, 4)$ és la permutació tal que

$$\sigma(2) = 3,\ \sigma(3) = 1,\ \sigma(1) = 4,\ \sigma(4) = 2,\ \sigma(i) = i \text{ per tot } i \neq 2, 3, 1, 4.$$

Amb aquest exemple, la idea de què és un cicle és ben clara, però trobar una definició és una mica més complicat.[3] És útil introduir el concepte d'**òrbites** d'una permutació.

> *Sigui σ una permutació de $\{1, \ldots, n\}$. Direm que i, j estan en una mateixa* **òrbita** *si existeix r tal que $\sigma^r(i) = j$. Aquest concepte d'estar en una mateixa òrbita és una relació d'equivalència (exercici). Les òrbites de σ són les classes d'equivalència respecte d'aquesta relació.*

Dit d'una altra manera: l'òrbita de i per una permutació σ és el conjunt **finit**

$$\{i, \sigma(i), \sigma^2(i), \sigma^3(i), \ldots\}.$$

[3] Aquí trobem un altre exemple encara de la importància de les definicions. El concepte de cicle, a partir de l'exemple que hem vist, és molt clar però, com el podem formalitzar en una definició matemàtica?

Considerem, per exemple, aquesta permutació

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 1 & 5 & 7 & 2 & 4 & 6 & 8 \end{pmatrix}.$$

És molt senzill trobar quines són les seves òrbites:

$$\{1, 2, 3, 5\}, \quad \{4, 6, 7\}, \quad \{8\}.$$

Aleshores, un **cicle** és una permutació que té una única òrbita de mida > 1. Dit d'una altra manera: tots els elements que no són fixos pertanyen a la mateixa òrbita. La notació que s'utilitza és escriure els elements de l'òrbita de mida > 1 en l'ordre $i, \sigma(i), \sigma^2(i), \ldots$ per un element qualsevol i de l'òrbita.

Un cicle $\sigma = (a_1, a_2, \ldots, a_r)$ té ordre r. En efecte, com que $\sigma^i(a_1) \neq a_1$ per $1 \leq i < r$, és clar que l'ordre no pot ser menor que r. D'altra banda, és clar que $\sigma^r(a_i) = a_i$ per tot $i = 1, 2, \ldots, r$. Com que, per definició de cicle, tots els altres elements diferents de $a_1, \ldots, a_r$ són fixos per σ, és clar que $\sigma^r = I$ i r és l'ordre de σ.

Com a cas particular, les transposicions són cicles d'ordre dos; i tot cicle d'ordre dos és una transposició. En l'exemple de les permutacions de tres objectes —o, equivalentment, les simetries del triangle equilàter— σ_1, σ_2 i σ_3 són cicles d'ordre dos —és a dir, transposicions— i σ_4, σ_5 són cicles d'ordre 3. En general, hi ha permutacions que no són cicles, per exemple, la permutació σ anterior.

12 El concepte de grup

Les permutacions de n elements que hem estudiat al capítol anterior formen un exemple del que és un dels conceptes més importants de les matemàtiques: el concepte de **grup**. La definició de grup ja ha aparegut més d'una vegada en aquests apunts. Tornem-la a repetir ara.

Un **grup** és una parella formada per un conjunt G i una **operació** entre els elements de G que compleix tres propietats senzilles que veurem a continuació. Una operació a G és una aplicació

$$\Phi : G \times G \longrightarrow G$$

que associa a cada parella ordenada (a, b) d'elements de G un element de G que direm que és el resultat d'operar a amb b.

Generalment, l'operació d'un grup no es designa en la forma $\Phi(a, b)$ sinó que es denota o bé amb el símbol de la suma —direm que estem utilitzant *notació additiva*— o bé amb el símbol del producte (un punt o la simple juxtaposició) —direm que estem utilitzant *notació multiplicativa*.

Expliquem ara quines són les tres propietats que ha de complir l'operació d'un grup. Ho farem per duplicat, en notació additiva i també en notació multiplicativa.

1. Propietat associativa
Per tot $a, b, c \in G$,
$a + (b + c) = (a + b) + c$

2. Element neutre
Existeix un element $0 \in G$ tal que
Per tot $a \in G$,
$a + 0 = 0 + a = a$

3. Inversos
Per tot $a \in G$ existeix $b \in G$
tal que $a + b = b + a = 0$

1. Propietat associativa
Per tot $a, b, c \in G$,
$a(bc) = (ab)c$

2. Element neutre
Existeix un element $1 \in G$ tal que
Per tot $a \in G$,
$a1 = 1a = a$

3. Inversos
Per tot $a \in G$ existeix $b \in G$
$ab = ba = 1$

En el cas que l'operació del grup compleixi també la propietat **commutativa** —és a dir, $a + b = b + a$ o $ab = ba$, tot tot $a, b \in G$— direm que el grup és

commutatiu o *abelià*. La utilització de dues notacions diferents a la teoria de grup és un fet al que convé acostumar-s'hi ben aviat. En aquest capítol i en els següents, anirem canviant sovint i deliberadament d'una notació a l'altra.

Algunes propietats elementals:

- L'element neutre és únic. Suposem que 0 i $0'$ fossin elements neutres. Aleshores $0 = 0 + 0' = 0'$.
- L'invers d'un element és únic. Considerem un element x que tingui dos inversos y i y'. Aleshores

  $$y' = 0 + y' = (y + x) + y' = y + (x + y') = y + 0 = y.$$

  Per tant, podem parlar de l'invers d'un element, sense que hi hagi ambigüitat. L'invers de a es denota $-a$ si utilitzem notació additiva i a^{-1} si utilitzem notació multiplicativa.

El concepte de grup és un concepte abstracte i potser hem de donar alguna indicació sobre per què hem dit que és un dels conceptes més importants de les matemàtiques.

> *Les **simetries** —aquesta paraula tindrà significats concrets segons el context que estiguem estudiant— de qualsevol objecte (matemàtic o de gairebé qualsevol àmbit) formen un grup amb l'operació de composició.*
>
> *Les simetries d'un objecte contenen una gran quantitat d'informació valuosa sobre l'objecte. En alguns casos, pràcticament tota la informació.*
>
> *Si coneguéssim tots els grups també coneixeríem totes les simetries possibles de tots els objectes possibles —d'una partícula subatòmica a tot l'univers, passant pel cub de Rubik, els polítops regulars de dimensió arbitrària, una molècula o qualsevol objecte matemàtic conegut o desconegut.*[1]

Al llarg d'aquest text —i al llarg dels cursos que l'estudiant de matemàtiques anirà fent— hem trobat alguns exemples de grups. Donem ara alguns pocs exemples:

1. Les permutacions de n objectes formen un grup que es coneix com el **grup simètric** i es denota habitualment Σ_n. Si $n > 2$, aquests grups no són commutatius. En canvi, el grup Σ_2, que només té dos elements $\Sigma_2 = \{I, (1, 2)\}$, és commutatiu, òbviament.

[1] En tot problema, molt sovint hi ha un grup que ens pot ajudar a resoldre'l. Modificant lleugerament una frase d'Alexandre Dumas que va esdevenir un tòpic de la novel·la negra, podríem dir: *Il y a un groupe dans toutes les affaires; aussitôt qu'on me fait un rapport, je dis: «Cherchez le groupe!»*.

2. Els enters amb la suma formen un grup. Els racionals diferents de zero amb el producte també. Són grups commutatius.

3. Les simetries de qualsevol figura geomètrica —triangle equilàter, cub, icosàedre,... qualsevol cosa— formen un grup. Aquí *«simetria»* s'entén en el sentit tradicional de la geometria: rotacions i reflexions. Generalment, es tracta de grups no commutatius.

4. La circumferència també és un grup de la manera següent. Escollim un punt A de la circumferència de manera que cada altre punt forma un cert angle amb el punt A i el centre de la circumferència. Aleshores, l'operació del grup és la suma d'angles. És un grup commutatiu.

5. Els vectors de $\mathbb{R}^3$, amb la suma de vectors, formen un grup commutatiu.

6. En canvi, els nombres reals amb la multiplicació no formen un grup perquè l'element 0 no té invers. Tampoc formen grup les matrius amb la multiplicació, perquè dues matrius només es poden multiplicar si tenen mides apropiades i, fins i tot si ens limitem a les matrius $n \times n$, tampoc no tenim un grup perquè hi ha matrius que no tenen inversa.

El grups poden tenir infinits elements —com el grup $\mathbb{Z}$— o tenir només un nombre finit d'elements —com els grups Σ_n. Si un grup té n elements, direm que és un grup d'**ordre** n.[2]

Homomorfismes de grups

Un principi molt general —i molt potent— de les matemàtiques és aquest:

> **Sempre que tenim algun tipus d'estructura ens interessa estudiar les transformacions d'aquesta estructura.**

En el cas de l'estructura de grup, les transformacions que ens interessen són els **homomorfismes de grup** que, dit d'una manera informal, són les aplicacions entre grups que «conserven» l'operació del grup.

Formalment: si G_1, G_2 són grups, un homomorfisme de G_1 a G_2 és una aplicació $f : G_1 \longrightarrow G_2$ que compleix aquesta propietat

$$\text{Per tot } a, b \in G_1 \quad f(a+b) = f(a) + f(b).$$

Aquí hem utilitzat notació additiva. Evidentment, la condició d'homomorfisme en notació multiplicativa s'escriuria[3]

$$\text{Per tot } a, b \in G_1 \quad f(ab) = f(a)\, f(b).$$

[2] El terme *ordre* ja s'ha utilitzar abans, quan parlàvem de l'ordre d'una permutació. Els dos conceptes estan relacionats, però no els hem de confondre un amb l'altre.

[3] També ens podem trobar que un dels grups estigui escrit en notació additiva i l'altre estigui escrit en notació multiplicativa. Caldrà fer les modificacions òbvies a la definició.

Dit d'una altra manera, si f és un homomorfisme de grups, obtenim el mateix resultat si primer operem dos elements i després apliquem f al resultat o si primer els apliquem f i després operem.

$$\left.\begin{matrix} a \\ b \end{matrix}\right\} \mapsto a+b \mapsto f(a+b)$$

$$\left.\begin{matrix} a \mapsto f(a) \\ b \mapsto f(b) \end{matrix}\right\} \mapsto f(a)+f(b)$$

Els homomorfismes de grup bijectius s'anomenen **isomorfismes**, els que són injectius o exhaustius s'anomenen, respectivament, **monomorfismes** i **epimorfismes**.

Exemples

- El grup Σ_3 l'hem estudiat al capítol anterior. Té 6 elements i n'hem escrit la taula de multiplicar. El grup $\Sigma_2 = \{I, \tau\}$ té només dos elements. Considerem aquesta aplicació $f : \Sigma_3 \to \Sigma_2$:

  $$f(I) = f(\sigma_4) = f(\sigma_5) = I,\ f(\sigma_1) = f(\sigma_2) = f(\sigma_3) = \tau.$$

  Amb paciència es pot anar comprovant que és un homomorfisme de grups.

- Al capítol anterior hem construït una aplicació $f : \Sigma_3 \to \mathcal{S}(\Delta)$, on $\mathcal{S}(\Delta)$ és el grup de les simetries del triangle equilàter. Es pot veure sense gaire dificultat que és un homomorfisme de grups. A més, és una aplicació bijectiva. És tracta, doncs, d'un **isomorfisme** de grups. Direm que els dos grups Σ_3 i $\mathcal{S}(\Delta)$ són **isomorfs** i això vol dir que, pel que fa a la teoria de grups, els dos grups tenen exactament les mateixes propietats. En la major part de les situacions, podem considerar que són *el mateix grup*.

Entre les propietats més elementals dels homomorfismes de grups hi ha aquestes:

- La composició d'homomorfismes és homomorfisme.
- Si un homomorfisme és bijectiu —i és, per tant, un isomorfisme—, l'aplicació inversa també és un homomorfisme (i, per tant, un isomorfisme).
- Si f és un homomorfisme, aleshores $f(0) = 0$ (en notació additiva). La demostració és simple: $f(0) = f(0+0) = f(0) + f(0)$ i, restant $f(0)$ als dos costats, $f(0) = 0$.
- Els homomorfismes conserven els inversos. És a dir, si f és un homomorfisme, aleshores (notació multiplicativa) $f(x^{-1}) = \big(f(x)\big)^{-1}$.

Subgrups

Si G és un grup i $A \subseteq G$, direm que A és un **subgrup** de G si A, amb la mateixa operació de G, compleix els axiomes de grup. Això és el mateix que dir que es compleixen aquestes tres propietats (notació multiplicativa):

1. $1 \in A$.
2. Per tot $x, y \in A$ es compleix $xy \in A$.
3. Per tot $x \in A$ es compleix $x^{-1} \in A$.

És fàcil veure que aquestes tres propietats són equivalents a aquestes dues (notació additiva):

1. $0 \in A$.
2. Per tot $x, y \in A$ es compleix $x - y \in A$.

Vegem alguns exemples:

- Si G és un grup (notació additiva) sempre podem considerar els subgrups $\{0\}$ i G. Els subgrups de G diferents de G s'anomenen *subgrups propis*.
- En el grup additiu dels nombres enters $\mathbb{Z}$, els nombres parells formen un subgrup, però els nombres senars no formen un subgrup (encara que li afegim el zero).
- Considerem el grup simètric Σ_3 que hem estudiat abans. Té 6 elements i, per tant, té 32 subconjunts que contenen l'element neutre I. No és difícil —però sí que és avorrit— repassar aquests subconjunts i fer una llista concreta de tots els que compleixen les condicions per ser subgrup. N'hi ha exactament 6:
  - Un subgrup trivial amb un únic element: $\{I\}$.
  - tres subgrups de dos elements: $\{I, \sigma_1\}$, $\{I, \sigma_2\}$, $\{I, \sigma_3\}$.
  - Un únic subgrup amb tres elements: $\{I, \sigma_4, \sigma_5\}$.
  - Σ_3.
- Si $\phi : G_1 \to G_2$ és un homomorfisme de grups, a partir de ϕ podem definir aquests subgrups interessants:
  - $\phi(G_1)$, es a dir, la imatge de G_1 per l'aplicació ϕ. És un subgrup de G_2.
  - $\{g \in G_1 : \phi(g) = 1\} \subseteq G_1$ (notació multiplicativa) és un subgrup de G_1 que s'anomena el **nucli** de l'homomorfisme ϕ i es denota $\ker(\phi)$:

    $$\ker(\phi) := \phi^{-1}(1) = \{g \in G_1 : \phi(g) = 1\}.$$

    Aquest subgrup compleix aquesta propietat interessant:

L'homomorfisme de grups ϕ és ***injectiu*** *si i només si (notació multiplicativa)*

$$\ker(\phi) = \{1\}.$$

La demostració és senzilla. Si ϕ és injectiu i $\phi(g) = 1 = \phi(1)$, ha de ser $g = 1$. Recíprocament, si suposem que $\ker(\phi) = \{1\}$ i que tenim $\phi(g) = \phi(h)$, tindrem $\phi(gh^{-1}) = \phi(g)\phi(h)^{-1} = 1$. Deduïm $gh^{-1} = 1$ i $g = h$, amb la qual cosa ja hem vist que ϕ és una aplicació injectiva.

13 El grup alternat

Després d'un parèntesi per estudiar grups en general, tornem ara a l'estudi de les permutacions, és a dir, a l'estudi del grup simètric Σ_n.

Tot cicle és producte de transposicions

Recordem que els *cicles* són un tipus particular de permutacions: el cicle que denotem $(a_1, \ldots, a_r)$, on tots els a_i són diferents, és la permutació

$$a_1 \mapsto a_2 \mapsto a_3 \mapsto \cdots \mapsto a_{r-1} \mapsto a_r \mapsto a_1.$$

D'altra banda, una *transposició* és simplement un cicle d'ordre 2: (i, j).

Qualsevol cicle d'ordre r es pot expressar com a producte de $r-1$ transposicions, d'aquesta manera:

$$(a_1, \ldots, a_r) = (a_1, a_r)\,(a_1, a_{r-1}) \cdots (a_1, a_3)\,(a_1, a_2).$$

Per comprovar que aquesta igualtat és certa n'hi ha prou amb aplicar els dos termes a cadascun dels elements $a_1, \ldots, a_r$ i veure que obtenim els mateixos resultats.[1] En conclusió, tot cicle es pot escriure com a producte de transposicions.

Tota permutació ($\neq I$) és producte de cicles disjunts

Volem demostrar ara que qualsevol permutació $\sigma \in \Sigma_n$, $\sigma \neq I$, es pot expressar com a producte de cicles disjunts

$$\sigma = c_1 c_2 \cdots c_k,$$

on la paraula *disjunts* fa referència a que cada element de $\{1, \ldots, n\}$ apareix com a màxim a un dels cicles $c_1, \ldots, c_k$. Observem que dos cicles disjunts necessàriament commuten: si $\sigma_1 = (a_i, \ldots, a_r)$ i $\sigma_2 = (b_1, \ldots, b_s)$ i en cap cas tenim $a_i = b_j$, és trivial comprovar que $\sigma_1\sigma_2 = \sigma_2\sigma_1$.

[1] Aquí tenim un bon exemple d'una situació que vam discutir en un capítol anterior. La demostració del teorema és senzillíssima *quan has descobert la fórmula anterior* i, per tant, tota la dificultat consisteix en *descobrir* aquesta fórmula, una dificultat que no és visible a la demostració.

Recordem com havíem definit el concepte de cicle. A partir d'una permutació qualsevol σ, vam definir una relació d'equivalència al conjunt $\{1,\dots,n\}$ de la següent manera: $i \sim j$ si $\sigma^r(i) = j$ per algun r. Aleshores, havíem anomenat **òrbites** de σ cadascuna de les classes d'equivalència. És a dir, si les òrbites són $O_1,\dots,O_s$, tenim

$$\{1,\dots,n\} = O_1 \cup O_2 \cup \cdots \cup O_s$$

on la unió és disjunta perquè dues classes d'equivalència no poden tenir elements en comú. A més, σ no pot transformar elements d'una òrbita en elements d'una òrbita diferent. Dit d'una altra manera, σ permuta els elements dins de cada òrbita O_i segona una permutació que en direm σ_i.

Hi ha dos tipus d'òrbites:

- Les òrbites amb més d'un element són cicles —per definició de cicle. Suposem que són les òrbites $O_1,\dots,O_k$ i, per tant, $\sigma_1,\dots,\sigma_k$ són cicles disjunts.
- Les òrbites amb un únic element corresponen a elements que queden fixos per σ.

Aleshores, és immediat que $\sigma = \sigma_1\sigma_2\cdots\sigma_k$ i el teorema està demostrat.

Aquesta descomposició de qualsevol permutació com a producte de cicles pot semblar una mica teòrica, però el cert és que és absolutament trivial obtenir la descomposició a la pràctica, d'aquesta manera:

1. Prenem un element i qualsevol amb $\sigma(i) \neq i$ i anem aplicant σ fins que tornem a arribar a l'element i. Ja tenim un primer cicle.
2. Si els cicles que tenim ja contenen tots els elements no fixos de $\{1,\dots,n\}$, hem acabat. En cas contrari, prenem un element no fix que no estigui a cap dels cicles que tenim i repetim el procés al punt 2.

Per exemple:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 1 & 5 & 7 & 2 & 4 & 6 & 8 \end{pmatrix} = (1,3,5,2)\,(4,7,6).$$

Combinant els dos teoremes que hem vist a l'apartat anterior —tot cicle és producte de transposicions— i en aquest —tota permutació és producte de cicles— arribem a aquesta conclusió:

Tota permutació de Σ_n ($n > 1$) és producte de transposicions.

El signe d'una permutació

El teorema que acabem de demostrar que ens diu que tota permutació és producte de transposicions no sembla cap gran cosa: és de sentit comú que, a base d'anar permutant objectes de dos en dos podem ordenar n objectes de totes les maneres possibles. Però el mètode que hem utilitzat (òrbites i cicles) sí que té un interès per ell mateix. D'altra banda, el resultat realment interessant sobre les permutacions és el que veurem en aquest apartat.

Hem vist, doncs, que tota permutació es pot descompondre com a producte de transposicions. Direm que les transposicions **generen** el grup simètric Σ_n. Però aquesta descomposició no és única, ni tampoc no és únic el nombre de transposicions que apareixen a una descomposició:

$$(1,2)\,(1,3)\,(2,3) = (1,3).$$

Però, en canvi, sí que es compleix aquest teorema important:

> **Teorema.** *Dues descomposicions d'una permutació en transposicions tenen la mateixa* ***paritat****.*

Quan diem que tenen la mateixa *paritat* volem dir que si una permutació es pot expressar com a producte d'un nombre parell de transposicions, no es pot expressar com a producte d'un nombre senar de transposicions, i si es pot expressar com a producte d'un nombre senar de transposicions, aleshores no es pot expressar com a producte d'un nombre parell de transposicions.

Aquest teorema (que demostrarem a continuació) ens permet definir el concepte de **permutació parella** i **permutació senar** com les que es poden expressar com a producte d'un nombre parell o senar de transposicions, respectivament. també ens permet definir la **signatura** o **signe** d'una permutació com

$$\operatorname{sig}(\sigma) = \begin{cases} +1 & \text{si } \sigma \text{ és parella} \\ -1 & \text{si } \sigma \text{ és senar} \end{cases}$$

Es compleixen aquestes propietats evidents:

- Les transposicions tenen signatura -1. La identitat té signatura $+1$.
- Un cicle d'ordre r té signatura $(-1)^{r-1}$.
- $\operatorname{sig}(\sigma_1\sigma_2) = \operatorname{sig}(\sigma_1)\operatorname{sig}(\sigma_2)$.
- La signatura es pot entendre com un homomorfisme de grups de la manera següent. Observem que el conjunt $\{+1,-1\}$ té una estructura evident de

grup multiplicatiu.[2] Aleshores, la propietat anterior ens diu que $\sigma \mapsto \mathrm{sig}(\sigma)$ és un homomorfisme de grups

$$\mathrm{sig} : \Sigma_n \longrightarrow \Sigma_2.$$

Demostració del teorema

- El pas clau de la demostració consisteix en demostrar això:

  Lema. *És impossible descompondre la identitat com a producte d'un nombre senar de transposicions.*

  La demostració d'aquest lema es fa per reducció a l'absurd i pel mètode del contraexemple minimal. Suposem, doncs, que

  $$I = \tau_1 \cdots \tau_k$$

  és una descomposició de la identitat en un nombre senar k de transposicions i suposem que k és mínim, en el senti que no hi ha cap altra descomposició de la identitat en un nombre senar $k' < k$ de transposicions.

  Suposem que $\tau_k = (a, b)$ i estudiem quines possibilitats hi ha per a τ_{k-1} i quines conseqüències té cadascuna d'elles:

  - $\tau_{k-1} = (a, b)$. En aquest cas, podríem simplificar l'expressió i obtenir una descomposició en un nombre també senar de transposicions, menor que k, cosa impossible.
  - $\tau_{k-1} = (a, c)$, $c \neq a, b$. En aquest cas, fem el canvi $\tau_{k-1}\tau_k = (a, b)(b, c)$.
  - $\tau_{k-1} = (b, c)$, $c \neq a, b$. En aquest cas, fem el canvi $\tau_{k-1}\tau_k = (a, c)(b, c)$.
  - $\tau_{k-1} = (c, d)$, $c, d \neq a, b$. En aquest cas, fem el canvi $\tau_{k-1}\tau_k = (a, b)(c, d)$.

  És a dir, descartant el primer cas, que és impossible, tenim una nova descomposició de la identitat en un nombre senar k de transposicions, però hem aconseguit que l'element a ja no surti a l'última transposició, sinó que surti a la penúltima. Repetint el procés tantes vegades com calgui, l'element a s'anirà desplaçant cap a l'esquerra fins arribar al primer lloc:

  $$I = (a, z)(u, v) \cdots (m, j)$$

  i, a més, a no tornarà a aparèixer en la descomposició. Ara observem que el terme de la dreta mou a i el de l'esquerra no mou a. Això és una contradicció que demostra el lema.[3]

[2] De fet, llevat d'isomorfisme hi ha un únic grup de dos elements. S'utilitzen diverses notacions per designar aquest grup. En notació additiva es designa $\{0, 1\}$ o $\mathbb{Z}/2\mathbb{Z}$; en notació multiplicativa es designa $\{1, -1\}$ o $\{1, \epsilon\}$ o també C_2. Si recordem que el grup Σ_2 té dos elements, també podem designar Σ_2 el grup de dos elements.

[3] On hem utilitzat la hipòtesi que k és senar?

- Ara la demostració del teorema és senzilla. Sigui $\sigma \in \Sigma_n$ i suposem que tenim dues descomposicions en transposicions

$$\sigma = \tau_1 \cdots \tau_k = \widetilde{\tau}_1 \cdots \widetilde{\tau}_s.$$

D'aquí es dedueix

$$\tau_1 \cdots \tau_k \widetilde{\tau}_s \cdots \widetilde{\tau}_1 = I.$$

Per tant $r + s$ és parell i r i s tenen la mateixa paritat.

Un cop tenim ben establert el concepte de signatura d'una permutació podem definir el **grup alternat**:

$$A_n = \{\sigma \in \Sigma_n : \operatorname{sig}(\sigma) = +1\}.$$

És un subgrup de Σ_n. De fet,

$$A_n = \ker(\operatorname{sig} : \Sigma_n \to \Sigma_2).$$

Per exemple, el grup alternat $A_3 \subseteq \Sigma_3$ és

$$A_3 = \{I, \sigma_4, \sigma_5\}.$$

Ens podem preguntar ara **quants elements té el grup alternat** A_n. La resposta s'obté immediatament aplicant aquesta estratagema senzilla. En primer lloc, si $n = 0, 1$, no hi ha cap permutació parella $\neq I$ i el grup alternat és trivial: $A_n = \{I\}$. Si $n > 1$ podem classificar totes les permutacions en parelles i senars. Les parelles formen el subgrup A_n i les senars formen el subconjunt $\mathcal{S}_n := \Sigma_n - A_n$, que clarament no és un subgrup. Considerem la transposició $\tau := (1, 2)$ i considerem l'aplicació $f : A_n \to \mathcal{S}_n$ donada per $f(\sigma) := \tau\sigma$. Es pot comprovar fàcilment que f és una aplicació bijectiva. Per tant, A_n i $\mathcal{S}_n$ tenen el mateix nombre d'elements: hi ha tantes permutacions parelles com permutacions senars. Com que, en total, hi ha $n!$ permutacions, el grup alternat té, exactament, $n!/2$ elements.[4]

[4] Aquest fet és un cas particular d'una propietat general que veurem més endavant (pàgina 108).

14 Conjugació, subgrups normals i nuclis

(En tot aquest capítol utilitzarem notació multiplicativa.)

La conjugació

Suposem que G és un grup i $g \in G$. L'element g ens defineix un homomorfisme $c_g : G \to G$ anomenat la **conjugació** per g:

$$c_g(x) := g^{-1}xg.$$

Que c_g és un homomorfisme de grups és evident:

$$c_g(xy) = g^{-1}xyg = g^{-1}xgg^{-1}yg = c_g(x)c_g(y).$$

També és evident que si G és un grup abelià, c_g és la identitat per tot $g \in G$. De fet, alguns dels temes que tractarem en aquest capítol no tenen interès quan el grup és abelià. Tanmateix, encara que G no sigui abelià, pot passar que c_g sigui la identitat per alguns $g \in G$.

Subgrups normals

Si tenim un homomorfisme de grups $\phi : G \to H$, podem considerar el subgrup $\ker(\phi) \subseteq G$. Aquest subgrup té una particularitat que el fa especial:

$$x \in \ker(\phi) \Rightarrow c_g(x) \in \ker(\phi) \text{ per tot } g \in G.$$

No tots els subgrups tenen aquesta propietat. Per exemple, considerem el subgrup $S := \{I, \sigma_1\} \subseteq \Sigma_3$. Veiem que $c_{\sigma_2}(\sigma_1) = \sigma_2\sigma_1\sigma_2 = \sigma_3 \notin S$.

Els subgrups que tenen la propietat anterior —que consisteix en que cap conjugació fa sortir elements fora del subgrup— s'anomenen **subgrups normals**. Concretament:

Definició. *Un subgrup H d'un grup G direm que és **normal** si per tot $h \in H$ i tot $g \in G$ es compleix $c_g(h) \in H$.*[1]

En particular, com hem dit abans, el nucli de qualsevol homomorfisme $G \to H$ és un subgrup normal de G.

El grup alternat A_n és un subgrup normal del grup simètric Σ_n. Això és trivial perquè

$$A_n = \ker(\text{sig} : \Sigma_n \to \Sigma_2)$$

i ja sabem que els nuclis sempre són subgrups normals.

Tot subgrup normal és un nucli

Hem dit que si tenim un homomorfisme ϕ d'un grup G a un altre grup, el nucli de ϕ és un subgrup normal de G. En aquest apartat demostrarem el teorema recíproc: tot subgrup normal de G és el nucli d'algun homomorfisme de G a algun altre grup. És un teorema molt important i la demostració es fa utilitzant la poderosa eina del *conjunt quocient*. Comencem enunciant amb precisió el teorema.

Teorema. *Sigui H un subgrup normal d'un grup G. Existeix un grup G' i un homomorfisme $\phi : G \to G'$ tal que $H = \ker(\phi)$.*

Farem la **demostració** en tres passos:

- Comencem definint una **relació d'equivalència** a G:

  $$g_1 \sim g_2 \Leftrightarrow g_1^{-1} g_2 \in H.$$

  Comprovem que es tracta realment d'una relació d'equivalència.

  - La propietat reflexiva es compleix trivialment.
  - Propietat simètrica. Si $g_1 \sim g_2$, tindrem $h := g_1^{-1} g_2 \in H$. Aleshores

    $$g_2^{-1} g_1 = (g_1^{-1} g_2)^{-1} = h^{-1} \in H$$

    i $g_2 \sim g_1$.
  - propietat transitiva. Suposem que $g_1 \sim g_2$ i $g_2 \sim g_3$. Aleshores $g_1^{-1} g_3 = g_1^{-1}(g_2 g_2^{-1}) g_3 = (g_1^{-1} g_2)(g_2^{-1} g_3) \in H$ i $g_1 \sim g_3$.

- Com que tenim una relació d'equivalència a G, podem definir el **conjunt quocient** $G/\sim$ que en aquest cas concret denotarem G/H. Recordem que G/H és, en principi, només un *conjunt* i que tenim una aplicació canònica

[1] Observem que ser *normal* no és una propietat intrínseca de H sinó que és una propietat de H com a subgrup de G.

$\pi : G \to G/H$. El punt clau de la demostració consisteix en adonar-se que G/H és també un **grup**.

Definim una operació a G/H d'aquesta manera:

$$[g_1]\,[g_2] := [g_1 g_2].$$

Cal veure que això està ben definit, és a dir cal veure que si $g_1 \sim g_1'$ i $g_2 \sim g_2'$, aleshores $g_1 g_2 \sim g_1' g_2'$. Tenim $g_1' = g_1 h$ amb $h \in H$, $g_2' = g_2 k$ amb $k \in H$. Aleshores, hem de comprovar que l'element

$$l := (g_1 g_2)^{-1}(g_1' g_2') = g_2^{-1} g_1^{-1} g_1 h g_2 k = g_2^{-1} h g_2 k = c_{g_2}(h) k$$

pertany a H. Però, **com que H és un subgrup normal** sabem que $c_{g_2}(h) \in H$ i, per tant, $l \in H$, com volíem demostrar.[2]

Un cop hem vist que la multiplicació de G/H està ben definida caldria comprovar que es compleixen els tres axiomes de grup, però això ja és senzill.

- Finalment, considerem l'aplicació $\pi : G \to G/H$ i observem aquestes dues propietats senzilles de demostrar:
  - π és un homomorfisme de grups.
  - $\ker(\pi) = H$.

Això acaba la demostració del teorema.

El teorema de Lagrange

La construcció que hem explicat a l'apartat anterior té, en el cas dels grups finits, una conseqüència senzilla però molt interessant que es coneix com a *teorema de Lagrange*.

Suposem que tenim un grup G d'ordre n i un subgrup —no cal que sigui normal— d'ordre k. Fem la construcció del conjunt quocient G/H. Si repassem la demostració anterior veurem que el fet que H sigui normal només s'utilitza per veure que G/H és un grup. Admetem, doncs que G/H és un conjunt de r elements, cadascun dels quals és una classe d'equivalència de G. Observem ara que si tenim dues classes d'equivalència $[g_1]$, $[g_2]$, l'aplicació $x \mapsto g_2 g_1^{-1}$ és una bijecció entre $[g_1]$ i $[g_2]$ (exercici **III.**26). Per tant, totes les classes tenen el mateix nombre d'elements. Com que $[1] = H$, totes les classes tenen exactament k elements. Tenim, doncs, un conjunt de n elements subdividit en r subconjunts de k elements i, en conseqüència, $n = rk$. Hem demostrat això:

[2] Observem com en aquesta demostració no hem començat dient *«com que H és normal bla, bla, bla»*, sinó que ens hem guardat el fet que H sigui normal fins el moment en que ens hem vist en l'obligació d'utilitzar-ho. És el que havíem explicat quan parlàvem de l'*estructura d'una demostració* a la pàgina 36.

Teorema de Lagrange *L'ordre de qualsevol subgrup d'un grup finit és un divisor de l'ordre del grup.*

Com convertir un homomorfisme en isomorfisme

Suposem que tenim un homomorfisme entre dos grups $\phi : G \to H$. Pot passar que ϕ no sigui injectiu o no sigui exhaustiu o no sigui cap de les dues coses. Què hi podem fer?

- Recordem que havíem definit el subgrup $\phi(G) \subseteq H$. Aleshores, podem descompondre ϕ com a composició de dues aplicacions

$$G \xrightarrow{\phi'} \phi(G) \xrightarrow{i} H$$

  on ara la primera aplicació ve donada per ϕ i ara ja és exhaustiva i la segona és la inclusió natural $i : \phi(G) \to H$ d'un subgrup en el seu grup. És a dir, qualsevol homomorfisme $\phi : G \to H$ es pot descompondre com $i \circ \phi'$ amb i homomorfisme injectiu i ϕ' homomorfisme exhaustiu.

- Com podem ara «convertir» ϕ en un homomorfisme injectiu? La idea és utilitzar el conjunt quocient, en particular el quocient del grup G pel subgrup $\ker\phi \subseteq G$. Recordem que $\ker\phi$ és un subgrup normal de G i, per tant, podem fer la construcció del *grup quocient* $G/\ker\phi$ que hem estudiat abans. Tindrem la projecció canònica $\pi : G \to G/\ker\phi$. Apliquem ara la propietat fonamental del conjunt quocient que ens permet definir una aplicació $\overline{\phi} : G/\ker\phi \to H$ amb la propietat que $\overline{\phi}\pi = \phi$. Tindrem, doncs, un diagrama d'aplicacions com aquest:

$$\begin{array}{ccc} G & \xrightarrow{\phi} & H \\ {\scriptstyle\pi}\downarrow & & \uparrow{\scriptstyle i} \\ G/\ker\phi & \xrightarrow{\overline{\phi}} & \phi(G) \end{array}$$

És a dir, hem aconseguit descompondre ϕ com a composició de tres aplicacions $\phi = i\,\overline{\phi}\,\pi$ de manera que

- π és un homomorfisme exhaustiu.
- i és un homomorfisme injectiu.
- No és difícil comprovar que $\overline{\phi}$ és un homomorfisme bijectiu. Per tant, és un isomorfisme.

Com a conseqüència d'això obtenim que qualsevol homomorfisme entre dos grups $\phi : G \to H$ dóna lloc a un **isomorfisme**

$$G/\ker\phi \cong \phi(G).$$

Aquest fet —que és molt útil— de vegades s'anomena **teorema de l'isomorfisme**.

Com a exemple d'això, considerem la signatura $\text{sig} : \Sigma_n \to \Sigma_2$ que és un homomorfisme exhaustiu ($n > 1$). Aplicant la tècnica anterior, i com que $\ker(\text{sig}) = A_n$, obtenim un isomorfisme de grups

$$\Sigma_n / A_n \cong \Sigma_2.$$

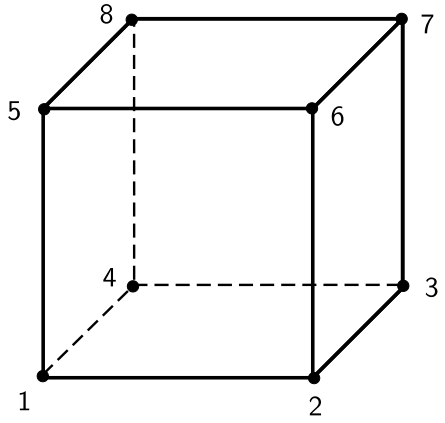
15 El grup del cub

En aquest capítol, com a aplicació de tot el que hem estat estudiant fins ara, discutirem les propietats d'un grup interessant: el grup Q que formen les **simetries d'un cub**. És un grup que també s'estudia a cristal·lografia i que ens servirà per il·lustrar els conceptes de teoria de grups que hem après —i aprendre'n algun altre.

El cub és un políedre regular en l'espai de tres dimensions —un dels cinc **sòlids platònics**. Imaginem-lo centrat a l'origen de $\mathbb{R}^3$ amb els seus vèrtex en els 8 punts de coordenades $(\pm 1, \pm 1, \pm 1)$:

$$V_1 = (1,-1,-1),\ V_2 = (1,1,-1),\ V_3 = (-1,1,-1),\ V_4 = (-1,-1,-1),$$

$$V_5 = (1,-1,1),\ V_6 = (1,1,1),\ V_7 = (-1,1,1),\ V_8 = (-1,-1,1).$$



Quan parlem de les *simetries* d'aquest cub ens referim als moviments de la geometria euclidiana —reflexions i rotacions— que deixen invariant el cub o, dit d'una altra manera, transformen cada punt del cub en un altre punt del cub. Per estudiar el grup que formen aquestes simetries, comencem enumerant-ne algunes que observem:

- La identitat I.
- 3 reflexions s_x, s_y, s_z respecte dels plans $x = 0$, $y = 0$, $z = 0$.

- 6 reflexions respecte de plans que passen per dues arestes oposades del cub. Aquests plans són $x+y=0$, $x-y=0$, $x+z=0$, $x-z=0$, $y+z=0$, $y-z=0$ i les reflexions les denotem s_{x+y}, s_{x-y}, s_{x+z}, s_{x-z}, s_{y+z}, s_{y-z}, respectivament.

- 3 rotacions R_x, R_y, R_z de 90 graus respecte dels eixos de coordenades. Aquestes rotacions tenen ordre 4 i si considerem R_i^2 i R_i^3 tenim en total 9 rotacions.

- 4 rotacions R_{17}, R_{28}, R_{35}, R_{46} de 120 graus respecte de les diagonals del cub. Aquestes rotacions tenen ordre tres i si considerem R_{ij}^2 tenim un total de 8 rotacions.

- 6 rotacions de 180° respecte d'eixos que passen pels centres de dues arestes oposades. Les denotem $R_{1,2}$, $R_{2,6}$, $R_{1,5}$, $R_{5,6}$, $R_{2,3}$ i $R_{6,7}$.

- L'aplicació antipodal T que transforma cada punt de l'espai en el seu simètric respecte de l'origen de coordenades i es pot entendre com la reflexió respecte de l'origen. És una simetria d'ordre 2.

- Les composicions arbitràries de simetries anteriors.

En aquesta llista hi ha, potencialment, infinites simetries, potser n'hi ha de repetides i potser encara hi ha alguna simetria que ens hem oblidat d'incloure. Per tant, encara hem de treballar més per poder dir que coneixem l'estructura del grup Q. Per exemple, entre les simetries anteriors es compleix aquesta igualtat:

$$T\,R_x^2 = s_x.$$

Com podem demostrar això? Com podem determinar exactament quines combinacions de simetries són iguals? Com podem programar un ordinador per fer càlculs explícits amb aquestes simetries? Com podem conèixer quina és l'**estructura** exacta del grup Q?

La resposta a aquestes preguntes ens porta a una de les idees més importants de la teoria de grup: les **representacions**.

Treballar geomètricament amb reflexions i rotacions és complicat. Si volem fer-ho de manera eficient o si volem que un ordinador ho faci per nosaltres, una idea interessantíssima és aquesta:

> *Observem que cada simetria del cub ens permuta els 8 vèrtex del cub. Si a cada simetria li fem correspondre la permutació dels seus 8 vèrtex tenim un homomorfisme de grups*
>
> $$\phi : Q \longrightarrow \Sigma_8.$$
>
> *El nucli de ϕ està format per les simetries que deixen fixos tots els vèrtex, però és evident que si tots els vèrtex del cub són fixos, tots els*

punts del cub són fixos i la simetria és la identitat.[1] *Per tant, ϕ és un homomorfisme injectiu i ens permet fer càlculs amb permutacions (fàcils de programar) en lloc de fer-los amb transformacions geomètriques (difícils de visualitzar i de programar). Diem que ϕ és una representació del grup Q en el grup simètric Σ_8.*[2]

En particular, el grup Q és finit i té com a màxim 8! elements —perquè és (isomorf a) un subgrup de Σ_8. A més, cada simetria del cub la podem codificar amb una permutació de 8 elements. Per exemple:

$$\begin{aligned}\phi(s_x) &= (1\,4)\,(2\,3)\,(5\,8)\,(6\,7)\\ \phi(R_x) &= (1\,2\,6\,5)\,(3\,7\,8\,4)\\ \phi(T) &= (1\,7)\,(2\,8)\,(3\,5)\,(4\,6)\end{aligned}$$

I ara és trivial comprovar que $T\,R_x^2 = s_x$, veient que $\phi(T\,R_x^2) = \phi(s_x)$. La representació del grup Q en el grup Σ_8 ens permet fer càlculs efectius amb les simetries del cub i ens podria portar, en principi, a entendre l'estructura del grup Q.

Però Σ_8 és un grup relativament gran —té 40.320 elements— i hi ha una altra representació de Q força millor. La idea és la següent. Considerem les 4 **diagonals** del cub

$$D_{17},\ D_{28},\ D_{35},\ D_{46}.$$

És clar que cada simetria del cub permutarà aquestes diagonals i, per tant, també tenim una representació

$$\psi : Q \longrightarrow \Sigma_4$$

que assigna a cada simetria la permutació de les diagonals. Per exemple,

$$\psi(s_{x+z}) = (D_{35}\,D_{46}), \quad \psi(s_{x-z}) = (D_{17}\,D_{28}),$$

$$\psi(s_{y-z}) = (D_{35}\,D_{28}), \quad \psi(s_{y+z}) = (D_{17}\,D_{46}),$$

i, com que amb aquestes quatre transposicions ja podem obtenir totes les permutacions de Σ_4, sabem que ϕ és un homomorfisme exhaustiu.

És ψ injectiu? Si no ho és, qui és el nucli? La resposta és que

$$\ker\psi = \{I,\ T\}.$$

Demostració: Suposem que S és una simetria del cub que deixa fixes les diagonals però no és la identitat. Suposem, sense pèrdua de generalitat, que S mou

[1] Això és cert perquè el concepte de simetria que hem escollit només contempla transformacions *lineals* de l'espai, transformacions que sempre converteixen línies rectes en línies rectes. D'aquesta manera, si una simetria deixa fixos dos vèrtex, aleshores deixa fixos tots els punts de la recta que els uneix. Això pot deixar de ser cert en simetries d'un tipus més general i, aleshores, tot el que fem en aquest capítol deixa de ser vàlid.

[2] Un expert en teoria de grups dirà que una representació no és un homomorfisme com ϕ sinó que és una classe d'equivalència d'homomorfismes, però en un text elemental com aquest ens permetem algunes petites simplificacions.

el vèrtex V_7. Però S conserva la diagonal D_{17} i, per tant $S(V_7) = V_1$. Aleshores, les tres arestes que conflueixen en el vèrtex V_7 s'han de convertir en les tres arestes que conflueixen en el vèrtex V_1 i S converteix els vèrtex V_3, V_6, V_8 en els vèrtex V_2, V_5, V_4, en algun ordre. Tornem a aplicar que S conserva les diagonals i ens veiem forçats a acceptar que $S(V_3) = V_5$, $S(V_6) = V_4$ i $S(V_8) = V_2$. En definitiva, S permuta els vèrtex igual que ho fa T i, per tant, $S = T$ i hem acabat la demostració.

Quina és la conclusió?

- $\phi : Q \to \Sigma_4$ és un homomorfisme exhaustiu.
- $\ker \psi = \{I, T\}$

Aplicant ara el teorema que vam veure al capítol anterior sobre convertir un homomorfisme en isomorfisme, arribem a aquesta conclusió:

$$Q/\{I, T\} \cong \Sigma_4$$

i això ens dóna una idea força acurada de quina és l'estructura del grup Q. En particular, quants elements té Q? Podem determinar aquest nombre amb un mètode similar al que vam utilitzar a la pàgina 105 per determinar l'ordre del grup alternat. L'isomorfisme anterior ens diu que si identifiquem $s \sim Ts$, obtenim un grup (el grup simètric Σ_4) amb 24 elements. Cada classe d'equivalència té dos elements. Per tant, Q té 48 elements.

Finalment, ens podem preguntar quins són aquests 48 elements. Abans hem descrit aquests 24 elements de Q:

- I.
- $R_x, R_y, R_z, R_x^2, R_y^2, R_z^2, R_x^3, R_y^3, R_z^3$.
- $R_{17}, R_{28}, R_{35}, R_{46}, R_{17}^2, R_{28}^2, R_{35}^2, R_{46}^2$.
- $R_{1,2}, R_{2,6}, R_{1,5}, R_{5,6}, R_{2,3}$ i $R_{6,7}$.

Calculant les imatges d'aquests elements de Q a Σ_4 veiem que totes aquestes imatges són diferents. Per tant, els elements són diferents i, si hi afegim els 24 elements que s'obtenen multiplicant per T cadascun dels 24 elements anteriors, obtenim les 48 simetries del cub.

Encara podem donar una tercera representació interessant del grup Q. Suposem que tenim tres colors i pintem cada cara del cub d'un color, de manera que les cares oposades tinguin el mateix color. Cada simetria del cub permutarà els tres colors i això ens donarà un homomorfisme de grups

$$\rho : Q \longrightarrow \Sigma_3.$$

Analitzem l'homomorfisme ρ:

- Observem que les reflexions s_x, s_y, s_z no alteren els colors i, per tant, el subgrup generat per aquestes tres reflexions està inclòs dins el nucli de ρ.
- Les reflexions s_x, s_y, s_z, com que els seus *miralls* són perpendiculars, commuten entre elles i, en conseqüència, el grup que generen, denotat $\langle s_x, s_y, s_z \rangle$, és commutatiu. Es tracta d'un grup força senzill: $\Sigma_2 \times \Sigma_2 \times \Sigma_2$ que té 8 elements (exercici 33).
- És molt senzill trobar simetries del cub que permutin els tres colors de qualsevol manera i això ens diu que ρ és exhaustiu.
- Com que Q té 48 elements i Σ_3 en té 6, per un raonament com el que hem fet abans, el nucli de ρ ha de tenir 8 elements i, per tant, ha de coincidir amb $\langle s_x, s_y, s_z \rangle$.

La conclusió és que tenim isomorfismes de grups

$$Q/\langle s_x, s_y, s_z \rangle \cong \Sigma_3, \qquad \langle s_x, s_y, s_z \rangle \cong (\Sigma_2)^3.$$

Exercicis de grups

1. (a) Calculeu $(3,4)\,(4,5)\,(2,3)\,(1,2)\,(5,6)\,(2,3)\,(4,5)\,(3,4)\,(2,3)$.

   (b) Descomponeu les següents permutacions en producte de cicles disjunts:

$$\begin{pmatrix}1&2&3&4&5\\3&4&5&1&2\end{pmatrix}\quad\begin{pmatrix}1&2&3&4&5&6&7\\7&6&1&2&3&4&5\end{pmatrix}\quad\begin{pmatrix}1&2&3&4&5&6&7&8\\6&2&5&7&8&1&3&4\end{pmatrix}$$

2. Considereu la permutació $\sigma \in \Sigma_5$ donada per

$$\sigma := \begin{pmatrix}1&2&3&4&5\\3&5&4&1&2\end{pmatrix}.$$

   (a) Calculeu el signe de σ.

   (b) Calculeu la permutació σ^{1250}.

   (c) Sigui $\tau := (135)(234) \in \Sigma_5$. Determineu la conjugació $c_\tau(\sigma)$.

3. Trobeu la descomposició en producte de cicles disjunts, la signatura, l'ordre i una descomposició en producte de transposicions de les següents permutacions de Σ_{10}:

$$\sigma = \begin{pmatrix}1&2&3&4&5&6&7&8&9&10\\3&7&1&4&2&6&9&8&5&10\end{pmatrix},$$

$$\rho = (10,3,4,1)\,(8,7)\,(4,7)\,(5,6)\,(2,6)\,(2,9).$$

   Calculeu σ^{2021} i ρ^{2022}.

4. Siguin τ_1, τ_2 dues transposicions diferents. Demostreu que la permutació $\tau_1\tau_2$ o bé és un cicle d'ordre tres o bé es pot escriure com a producte de dos cicles d'ordre 3.

5. Sabem que cada permutació es pot escriure com a producte de cicles disjunts. Estudieu quines possibilitats hi ha per a una permutació de Σ_7 o de Σ_5 i, a partir d'aquí, contesteu aquestes preguntes: podem tenir a Σ_7 un element d'ordre 15? Quins són els ordres dels elements de Σ_5?

6. Trobeu tots els elements $\sigma \in \Sigma_5$ que compleixin aquestes igualtats:

   (a) $\sigma\,(1,3,5)^{32} = (2,4,1,3)$.

   (b) $(1,2,5)\,\sigma\,(1,2,5)^{-1} = (2,4)$.

   (c) $\sigma^2 = (1,2)$.

   (d) $\sigma^7 = (1,2,4)$.

(e) $\sigma\,(1,3,2)\,\sigma^{-1} = (2,3)$.

(f) $\sigma^3 = (1,2,4)$.

(En alguns casos caldrà utilitzar els resultats de l'exercici 5.)

7. Els *nombres de Montmort* M_n, $n > 0$ es defineixen com el nombre de permutacions de n elements que no deixen fix cap element. Demostreu aquesta fórmula recursiva:

$$M_1 = 0, \quad M_2 = 1, \quad M_n = (n-1)(M_{n-1} + M_{n-2}).$$

Demostreu per inducció que

$$M_n = n! \sum_{i=0}^{n} \frac{(-1)^i}{i!}.$$

Suposeu que, per un error informàtic, les qualificacions de selectivitat de tot el país es trameten de manera aleatòria als alumnes que hi han participat. Quin és la probabilitat que cap alume rebi les seves pròpies qualificacions?

8. Determineu tots els elements $\sigma \in \Sigma_n$ tals que $\sigma^2 = I$.

9. Siguin $\sigma, \gamma \in \Sigma_n$.

   (a) Calculeu l'ordre i la signatura de $\gamma\sigma\gamma^{-1}$ en funció de l'ordre i la signatura de σ i γ.

   (b) Proveu que si σ és un cicle, $\sigma = (a_1, a_2, \ldots, a_r)$, aleshores

$$\gamma\sigma\gamma^{-1} = (\gamma(a_1), \gamma(a_2), \ldots, \gamma(a_r)).$$

10. Demostreu que la relació de conjugació és una relació d'equivalència. És a dir, en un grup G definim $g_1 \sim g_2$ si existeix $g \in G$ tal que $c_g(g_1) = g_2$ i es demana demostrar que es tracta d'una relació d'equivalència.

11. Demostreu que si

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \cdot & \cdot & \cdot & n \\ b_1 & b_2 & b_3 & \cdot & \cdot & \cdot & b_n \end{pmatrix}$$

llavors $\gamma\sigma\gamma^{-1}$ és la permutació obtinguda aplicant γ a aquesta taula. Feu servir aquest mètode per calcular $\rho\sigma\rho^{-1}$ on σ i ρ són les permutacions de l'exercici 3.

12. Utilitzeu l'exercici 9 per trobar la descomposició en producte de cicles disjunts de $\gamma\sigma\gamma^{-1}$ en termes de la de σ.

13. Donat $\sigma \in \Sigma_n$, el *tipus cíclic* de σ és la n-tupla d'enters $(p_1, p_2, \ldots, p_n)$ on p_k és el nombre de k-cicles en la descomposició en cicles disjunts de σ (entenem que p_1 és el nombre d'elements fixos per σ). Demostreu que dues permutacions de Σ_n són conjugades si i només si tenen el mateix tipus cíclic.

14. Demostreu que tota permutació $\sigma \in A_n$, $n \geq 3$, es pot escriure com a producte de cicles d'ordre 3. (Utilitzeu l'exercici 4.)

15. Denotem els elements (diferents de I) del grup simètric Σ_3 igual que a la pàgina 90. Considereu l'homomorfisme $c : \Sigma_3 \to \Sigma_3$ donat per la conjugació amb σ_4, és a dir $c(g) := \sigma_4^{-1} g \sigma_4$.

(a) Determineu explícitament els elements del conjunt $H := \{g \in \Sigma_3 : c(g) = g\}$.

(b) És H un subgrup de Σ_3? En cas afirmatiu, és commutatiu?, és normal?

16. En un cert truc de màgia[3] el mag mostra un panell com aquest

A						1
B						2
C						3
D						4
E						5

Aleshores, demana la participació d'una parella del públic i diu que els demostrarà que estaven predestinats a trobar-se. Un dels membres de la parella tria lliurement una de les posicions A, B, C, D, E del panell i l'altre membre tria lliurement una xifra $1, 2, 3, 4, 5$ del panell. A continuació, el mag mostra per una cara, successivament, cinc cartolines i demana a la parella que decideixi lliurement a quin dels cinc llocs buits del panell vol col·locar cadascuna de les cartolines. Quan ho han fet, els ofereix la possibilitat de canviar l'ordre de les cartolines i també la seva orientació. Finalment, el mag dóna la volta a cadascuna de les cinc cartolines i tothom pot veure que al darrere de cadascuna hi ha traçats cinc camins entortolligats que van d'esquerra a dreta i, sorprenentment, si comencem per la lletra que ha triat un membre de la parella i seguim el camí que s'ha format, arribem a la xifra que ha triat l'altre membre de la parella. Utilitzeu la teoria dels grups de permutacions per descobrir el funcionament d'aquest efecte màgic.

17. Direm que un cicle de Σ_n és *gegant* si té ordre $k > n/2$. Calculeu, en funció de n i k, el nombre de permutacions que contenen un cicle gegant d'ordre k. Utilitzeu la relació entre la sèrie harmònica i el logaritme per veure que, si n és gran, la probabilitat que una permutació de Σ_n escollida a l'atzar contingui un cicle gegant és $\approx \log(2)$.[4]

18. Considereu aquests grups:

(a) El grup Σ_2.

(b) El grup A_3.

(c) El subgrup C_4 de Σ_4 format per les potències del cicle $(1, 2, 3, 4)$.

(d) El subgrup K_4 de Σ_4 definit $K_4 := \{I, (1, 2), (1, 4), (1, 2)(3, 4)\}$.

Demostreu que tot grup no trivial de menys de 5 elements és isomorf a un i només un dels grups d'aquesta llista.

19. Siguin G_1 i G_2 dos grups (notació multiplicativa). Considerem, al conjunt $G_1 \times G_2$, l'operació

$$(g_1, g_2) \cdot (g_1', g_2') := (g_1 g_1', g_2 g_2').$$

Demostreu que, amb aquesta operació, $G_1 \times G_2$ és un grup.

[3] El podeu veure (abril de 2024) en aquest enllaç: `youtube.com/watch?v=aZpA1k-sDcg` (no llegiu els comentaris perquè poden contenir idees sobre com es fa el truc).

[4] Aquest curiós exercici prové d'un article de John Baez (`math.ucr.edu/home/baez/permutations/permutations_4.html`). Podeu trobar una utilització sorprenent d'aquest càlcul en el vídeo: `youtube.com/watch?v=iSNsgj1OCLA` (enllaços vàlids l'abril del 2024).

20. Quins d'aquests subconjunts de $\mathbb{Z}$ són subgrups i quins no ho són: (a) els múltiples d'un enter fixat m; (b) els enters no negatius; (c) els enters múltiples de m o de n, on m, n són enters diferents; (d) els enters múltiples de m i de n, on m, n són enters diferents; (e) els enters que no són múltiples d'un enter fixat m.

21. Sigui $m > 1$ un enter i sigui A el conjunt de tots els nombres racionals que es poden escriure com una fracció on el denominador no és múltiple de m. És A un subgrup de $\mathbb{Q}$? I si m és un nombre primer? Sigui B el conjunt de tots els nombres racionals que es poden escriure com una fracció on el denominador és una potència de m. És B un subgrup de $\mathbb{Q}$?

22. Sigui A un conjunt amb $n > 0$ elements, $A = \{a_0, \ldots, a_{n-1}\}$. Definiu una estructura (multiplicativa) de grup a A que compleixi que $a_0 = 1$, $a_i = a_1^i$ i $a_1^n = 1$. Aquest grup s'anomena *el grup cíclic d'ordre n*.

23. Sigui $\psi : G \to H$ un homomorfisme de grups i sigui $g \in G$ un element d'ordre r. Què podem afirmar de l'ordre de $\psi(g) \in H$?

24. Considerem $\mathbb{R}$ com a grup additiu amb la suma ordinària de nombres reals i sigui $\mathbb{R}^{>0}$ el subconjunt del nombres reals positius. Demostreu que $\mathbb{R}^{>0}$ és un grup multiplicatiu amb la multiplicació ordinària de nombres reals. Utilitzeu la funció exponencial per demostrar que aquests dos grups (que d'entrada semblen tan diferents) són isomorfs: $\mathbb{R} \cong \mathbb{R}^{>0}$.

25. Considereu els grups simètrics Σ_2, Σ_3. Construïu un homomorfisme exhaustiu $\phi : \Sigma_3 \to \Sigma_2$ i demostreu que és únic.

26. Completeu els detalls de la demostració del teorema de Lagrange (pàgina 108).

27. Considerem aquestes permutacions de Σ_7:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 5 & 6 & 4 & 3 & 2 & 1 & 7 \end{pmatrix}, \ \rho = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 6 & 7 & 1 & 3 & 2 & 5 \end{pmatrix},$$

$$\gamma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 5 & 6 & 7 & 2 & 3 & 1 \end{pmatrix}$$

    (a) Per a σ i ρ, calculeu la descomposició en cicles disjunts, l'ordre i el signe. Calculeu la conjugació $c_\sigma(\rho)$.

    (b) Demostreu que no hi ha cap homomorfisme de grups $\phi : \Sigma_7 \to \Sigma_7$ tal que $\phi(\rho) = \sigma$, però sí que n'hi ha un tal que $\phi(\rho) = \gamma$.

28. Considereu aquest subconjunt del grup alternat A_4:

$$H := \{I, (1,2)(3,4), (1,3)(2,4), (1,4)(2,3)\}.$$

    Demostreu que H és un subgrup de A_4. Demostreu que és un subgrup normal (utilitzeu l'exercici 13). Determineu l'estructura del grup quocient A_4/H.

29. Considereu aquestes sis matrius:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

    Demostreu que formen un grup multiplicatiu isomorf a Σ_3.

30. Sigui G un grup i sigui $\mathcal{A}(G)$ el conjunt de tots els isomorfimes $\phi : G \to G$. Demostreu que $\mathcal{A}(G)$ és un grup amb l'operació de composició. Demostreu que l'aplicació $c : G \to \mathcal{A}(G)$ donada per $g \mapsto c_g$ és un homomorfisme de grups. Demostreu que el nucli de c està format pels elements de G que commuten amb tots els elements de G. Per tant, aquest subgrup de G —que s'anomena el *centre* de G— és un subgrup normal de G. Determineu el centre de Σ_3.

31. En el grup de simetries del cub, escriviu les reflexions s_x, s_{y+z} en funció de rotacions i l'aplicació antipodal T.

32. Considereu aquestes reflexions del cub: s_y, s_{y-z}, s_{x-z}.

    (a) Utilitzeu la representació del grup de simetries del cub en Σ_8 per demostrar aquestes igualtats:

    $$(s_y s_{y-z}^2)^4 = (s_{y-z} s_{x-z})^3 = (s_y s_{x-z})^2 = I$$
    $$s_y s_{x-z} = s_{x-z} s_y$$
    $$(s_y s_{y-z} s_{x-z})^3 = T.$$

    (b) Demostreu que qualsevol simetria del cub es pot expressar com a producte de les reflexions s_y, s_{y-z}, s_{x-z}.

33. Considereu les reflexions s_x, s_y, s_z del cub, i tots els seus productes. Demostreu que obteniu un subgrup abelià de Q, d'ordre 8.

34. Sigui M el conjunt format per totes les matrius 3×3 que compleixen aquestes propietats:

    (a) Només contenen entrades $0, 1, -1$.

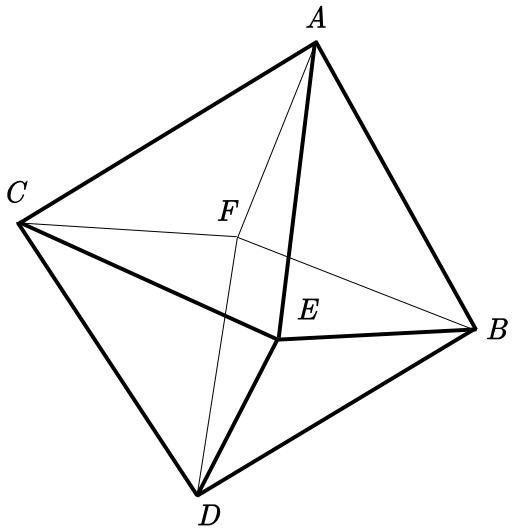
    (b) A cada fila i cada columna només hi ha un terme $\neq 0$.

    Demostreu que M, amb la multiplicació de matrius, és un grup. Trobeu un isomorfisme entre el grup M i el grup Q de simetries del cub. (Utilitzeu l'exercici 32b.)

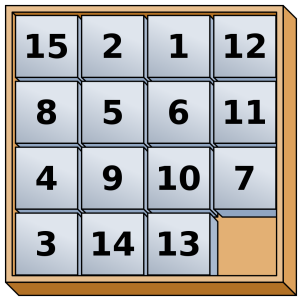
35. Sigui G el grup de simetries de l'octàedre regular. Considereu aquests elements de G:

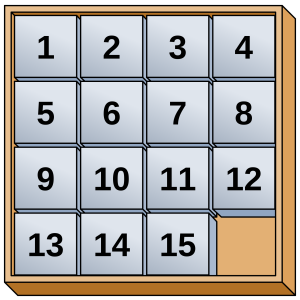
    - R és la rotació de 90° respecte de l'eix que passa per E i F, de manera que $R(B) = A$.
    - S és la reflexió respecte del pla que passa pels punts A, F, E, D.
    - H és la reflexió respecte del pla que passa pels punts E, F, pel punt mig entre A i B i pel punt mig entre C i D.

    Utilitzeu una representació injectiva de G en un grup simètric i demostreu aquestes propietats de G: (a) G té ≤ 720 elements; (b) $(R^{-1}S)^{19} = H$.



36. Utilitzeu la teoria de permutacions per estudiar matemàticament aquests trencaclosques clàssics:





Estudieu el cas general amb $nm - 1$ peces quadrades iguals («fitxes»), cadascuna amb un nombre (diferent) de 1 a $nm - 1$, situades en un rectangle de n files i m columnes amb $n, m \geq 2$. Seguiu aquests passos:

(a) Definiu **moviment elemental** com el fet de desplaçar una fitxa a una casella buida contigua (en horitzontal o en vertical). Un **moviment** és una successió finita de moviments elementals. Un **moviment complet** és un moviment que comença i acaba amb la casella buida a baix a la dreta. Vegeu que els moviments complets formen un grup $\mathcal{M}$.

(b) Definiu un homomorfisme de grups $\phi : \mathcal{M} \to \Sigma_{nm-1}$.

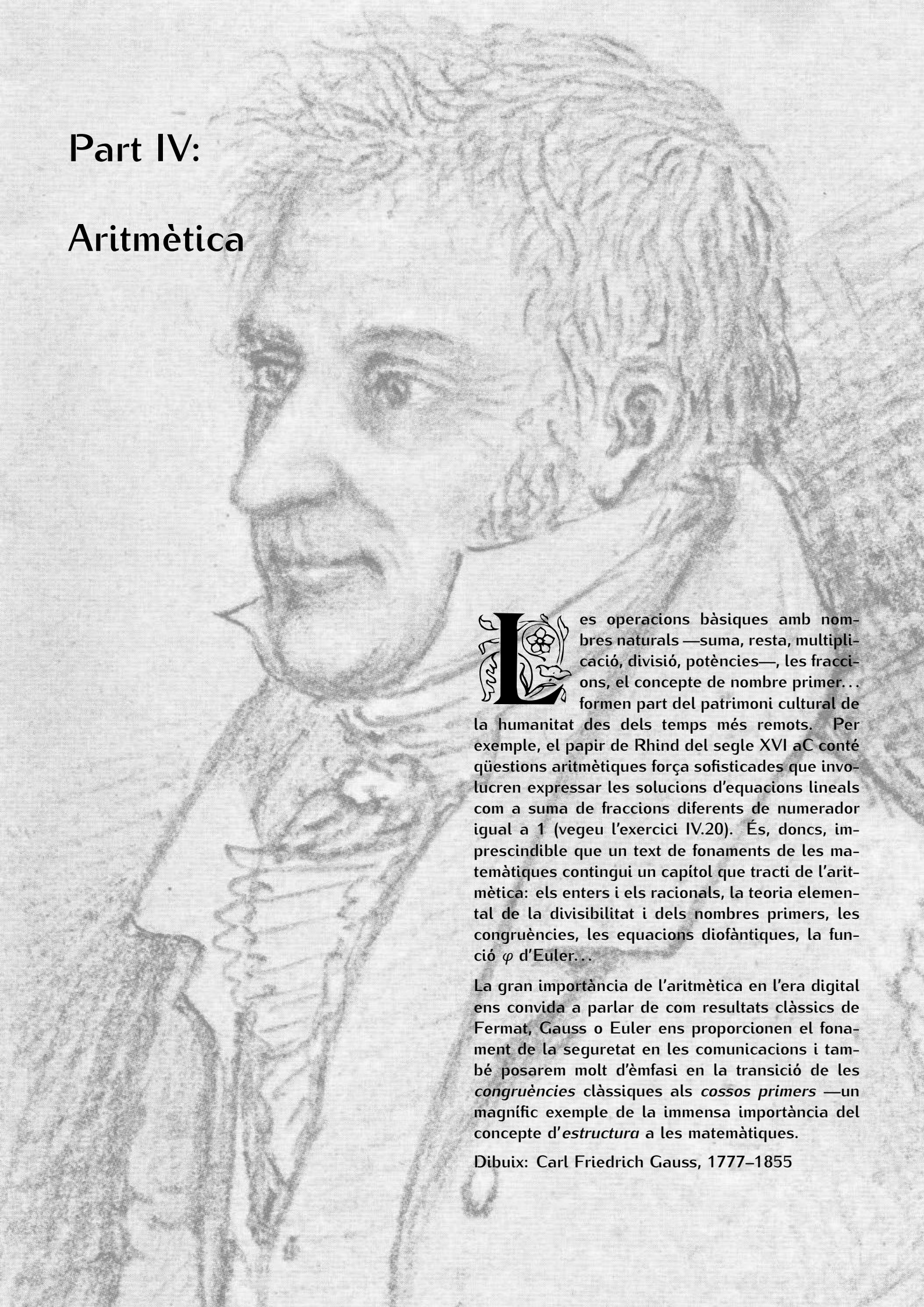
(c) Demostreu que la imatge de ϕ està dins del grup alternat A_{nm-1}.

(d) Demostreu que la imatge de ϕ és igual al grup alternat A_{nm-1}. Feu-ho per inducció sobre n i m i, al mateix temps, construïu un algorisme per resoldre el trencaclosques.

pàgina (gairebé) en blanc

Part IV:

Aritmètica



Les operacions bàsiques amb nombres naturals —suma, resta, multiplicació, divisió, potències—, les fraccions, el concepte de nombre primer... formen part del patrimoni cultural de la humanitat des dels temps més remots. Per exemple, el papir de Rhind del segle XVI aC conté qüestions aritmètiques força sofisticades que involucren expressar les solucions d'equacions lineals com a suma de fraccions diferents de numerador igual a 1 (vegeu l'exercici IV.20). És, doncs, imprescindible que un text de fonaments de les matemàtiques contingui un capítol que tracti de l'aritmètica: els enters i els racionals, la teoria elemental de la divisibilitat i dels nombres primers, les congruències, les equacions diofàntiques, la funció φ d'Euler...

La gran importància de l'aritmètica en l'era digital ens convida a parlar de com resultats clàssics de Fermat, Gauss o Euler ens proporcionen el fonament de la seguretat en les comunicacions i també posarem molt d'èmfasi en la transició de les *congruències* clàssiques als *cossos primers* —un magnífic exemple de la immensa importància del concepte d'*estructura* a les matemàtiques.

Dibuix: Carl Friedrich Gauss, 1777–1855

16 | Els enters i els racionals

Els nombres negatius i les fraccions són extensions dels nombres naturals *«inventades»* amb un mateix objectiu: treballar amb solucions d'operacions que *no tenen solució* en els nombres naturals. La idea de la construcció dels negatius i de les fraccions a partir dels nombres naturals és la mateixa en els dos casos i els podem tractar en paral·lel.

• A $\mathbb{N}$ hi ha **restes** que no es poden fer, com

$$2-3.$$

• Com que no es poden fer, les *deixem indicades*

$$"2-3".$$

• Són expressions *formals* o *simbòliques*.

• En diem *nombres enters*.

• Per evitar confusions, de moment escrivim (a,b) en lloc de $"a-b"$.

• **Problema:** $"2-3"$ ha de ser el mateix que $"4-5"$. És a dir

$$(2,3)=(4,5).$$

• **Solució:** Utilitzem la poderosa idea del **conjunt quocient**.

• A $\mathbb{N}$ hi ha **divisions** que no es poden fer, com

$$2/3.$$

• Com que no es poden fer, les *deixem indicades*

$$"2/3".$$

• Són expressions *formals* o *simbòliques*.

• En diem *nombres racionals*.

• Per evitar confusions, de moment escrivim (a,b) en lloc de $"a/b"$.

• **Problema:** $"2/3"$ ha de ser el mateix que $"4/6"$. És a dir

$$(2,3)=(4,6).$$

• **Solució:** Utilitzem la poderosa idea del **conjunt quocient**.

Construcció de $\mathbb{Z}$ a partir de $\mathbb{N}$

Considerem el conjunt $\mathbb{N} \times \mathbb{N}$ amb aquesta relació:[1]

$$(a, b) \sim (a', b') \Leftrightarrow a + b' = b + a'$$

que és fàcil veure que és una relació d'equivalència. Aleshores, **definim el conjunt del nombres enters** així:

$$\mathbb{Z} := (\mathbb{N} \times \mathbb{N}) / \sim .$$

En aquest conjunt $\mathbb{Z}$ podem definir una operació de **suma** d'aquesta manera:[2]

$$[(a, b)] + [(a', b')] := [(a + a', b + b')]$$

i podem comprovar fàcilment aquestes propietats:

- Està ben definida. Observem que hem definit la suma en un conjunt quocient i, per tant, cal fer la comprovació que si canviem de representants en les classes d'equivalència, la classe d'equivalència del resultat no canvia.
- Aquesta operació de suma dota $\mathbb{Z}$ d'estructura de **grup abelià**. És a dir, la suma compleix les propietats associativa i commutativa, l'element $[(0, 0)]$ actua com a element neutre i l'element invers de $[(a, b)]$ és $[(b, a)]$.
- Hi ha una aplicació natural $\mathbb{N} \to \mathbb{Z}$ definida per $n \mapsto [(n, 0)]$. Aquesta aplicació és injectiva i conserva la suma. Això ens permet *identificar* cada nombre natural n amb el nombre enter $[(n, 0)]$.
- Observem que, si $n \in \mathbb{N}$, es compleix $n + [(0, n)] = 0$. Això ens permet escriure $-n := [(0, n)] \in \mathbb{Z}$.
- També és fàcil veure que els nombres enters són una unió disjunta

$$\mathbb{Z} = \{n \in \mathbb{N} : n > 0\} \cup \{0\} \cup \{-n : n \in \mathbb{N},\ n > 0\}.$$

En el grup abelià $\mathbb{Z}$ també podem definir una operació de **producte**:[3]

$$[(a, b)]\ [(c, d)] := [(ac + bd, ad + bc)]$$

de manera que es compleixen aquestes propietats senzilles:

- El producte està ben definit.

[1] Per què donem precisament aquesta definició i no cap altra? Perquè volem que (a, b) representi $a - b$ i és clar que volem que $a - b = a' - b'$ sigui equivalent a $a + b' = b + a'$.

[2] Per què donem precisament aquesta definició i no cap altra? Perquè volem que (a, b) representi $a - b$ i és clar que volem que $(a - b) + (a' - b') = (a + a') - (b + b')$.

[3] Per què donem precisament aquesta definició i no cap altra? Perquè volem que (a, b) representi $a - b$ i és clar que volem que $(a - b)(c - d) = (ac + bd) - (ab + bc)$.

- El producte compleix les propietats associativa, commutativa i distributiva (respecte de la suma de $\mathbb{Z}$) de manera que $1 \in \mathbb{Z}$ actua com a element neutre multiplicatiu.
- Aquestes operacions de suma i producte doten $\mathbb{Z}$ d'una estructura algebraica que es coneix amb el nom d'**anell** (commutatiu amb unitat).

Construcció de $\mathbb{Q}$ a partir de $\mathbb{Z}$

La construcció dels nombres racionals (o fraccionaris) $\mathbb{Q}$ a partir dels enters $\mathbb{Z}$ s'assembla moltíssim a la construcció de $\mathbb{Z}$ a partir de $\mathbb{N}$ que hem vist a l'apartat anterior.

Considerem el conjunt $\mathbb{Z} \times (\mathbb{Z} - \{0\})$ amb aquesta relació:[4]

$$(a, b) \sim (a', b') \Leftrightarrow ab' = ba'$$

que és fàcil veure que és una relació d'equivalència. Aleshores, **definim el conjunt del nombres racionals** així:

$$\mathbb{Q} := \big(\mathbb{Z} \times (\mathbb{Z} - \{0\})\big) / \sim .$$

En aquest conjunt $\mathbb{Q}$ podem definir una operació de **suma** d'aquesta manera:

$$[(a, b)] + [(c, d)] := [(ad + bc, bd)]$$

i podem comprovar sense gaires dificultats aquestes propietats:

- Està ben definida. Observem que hem definit la suma en un conjunt quocient i, per tant, cal fer la comprovació que si canviem de representants en les classes d'equivalència, la classe d'equivalència del resultat no canvia.
- Aquesta operació de suma dota $\mathbb{Q}$ d'estructura de **grup abelià**. És a dir, la suma compleix les propietats associativa i commutativa, l'element $[(0, 1)]$ actua com a element neutre i l'element invers de $[(a, b)]$ és $[(-a, b)]$.
- Hi ha una aplicació natural $\mathbb{Z} \to \mathbb{Q}$ definida per $a \mapsto [(a, 1)]$. Aquesta aplicació és injectiva i conserva la suma. Això ens permet *identificar* cada nombre enter a amb el nombre racional $[(a, 1)]$.

En el grup abelià $\mathbb{Q}$ també podem definir una operació de **producte** d'aquesta manera:

$$[(a, b)]\ [(c, d)] := [(ac, bd)]$$

de manera que es compleixen aquestes propietats senzilles:

[4] Per què donem precisament aquesta definició i no cap altra? Perquè volem que (a, b) representi a/b i és clar que volem que $a/b = a'/b'$ sigui equivalent a $ab' = ba'$.

- El producte està ben definit.
- El producte compleix les propietats associativa, commutativa i distributiva (respecte de la suma de $\mathbb{Q}$) de manera que $1 \in \mathbb{Q}$ actua com a element neutre multiplicatiu.
- Les propietats de la suma i el producte de $\mathbb{Q}$ el converteixen en un **anell** (commutatiu amb unitat).
- A més, si tenim $q = [(a, b)] \in \mathbb{Q}$ amb $q \neq 0$ podem considerar el nombre racional $q' := [(b, a)] \in \mathbb{Q}$ i observar que $qq' = 1$. Per tant, a $\mathbb{Q}$ tots els elements diferents de zero tenen invers multiplicatiu. Direm que $\mathbb{Q}$ té estructura de **cos**.
- Aquesta última propietat ens permet escriure

$$[(a, b)] = \frac{a}{b} \text{ si } b \neq 0$$

  i recuperar, d'aquesta manera, la notació tradicional dels nombres racionals.

Anell, domini, cos

Hem dit que $\mathbb{Z}$ és un anell i $\mathbb{Q}$ és un cos. Donem ara definicions precises d'aquests dos conceptes —i d'algun altre.

- **Grup commutatiu**. Hi ha una operació $x + y$ que compleix les propietats
  - **associativa**: $\forall x, y, z,\ (x + y) + z = x + (y + z)$.
  - **commutativa**: $\forall x, y,\ x + y = y + x$.
  - **element neutre**: $\exists 0\, \forall x,\ x + 0 = x$.
  - **inversos**: $\forall x\, \exists y,\ x + y = 0$.
- **Anell**. Hi ha dues operacions $x + y$ i xy que compleixen
  - La suma compleix les propietats de grup commutatiu.
  - La multiplicació compleix les propietats:
    - **associativa**: $\forall x, y, z,\ (xy)z = x(yz)$.
    - **commutativa**: $\forall x, y,\ xy = yx$.
    - **element neutre**: $\exists 1\, \forall x,\ x1 = x$.
  - La multiplicació compleix la propietat **distributiva** respecte de la suma:

$$\forall x, y, z,\ x(y + z) = xy + xz.$$

- **Domini** (també anomenat *domini d'integritat*). Les mateixes propietats d'anell i a més

  - **No hi ha divisors de zero**: Si $xy = 0$, aleshores $x = 0$ o $y = 0$.

- **Cos**: Les mateixes propietats d'anell i a més
  - **No trivialitat**: $1 \neq 0$.
  - **inversos multiplicatius**: Per tot $x \neq 0$ existeix y tal que $xy = 1$. Aquesta propietat implica que no hi ha divisors de zero. Per tant, tot cos és un domini.

Amb les operacions de suma i producte que ja coneixem, $\mathbb{Z}$ i $\mathbb{Q}$ són dominis, $\mathbb{Q}$ és un cos i $\mathbb{Z}$ no ho és. La suma i el producte de nombres naturals compleixen algunes de les propietats anteriors, però $\mathbb{N}$ no és un anell. Si afeblim o eliminem algunes de les propietats anteriors obtenim algunes estructures algebraiques importants que no estudiarem en aquest curs. Per exemple, són importants els *anells no commutatius*, en els que la multiplicació pot no ser commutativa, i els *anells de divisió* en els que es compleixen tots els axiomes de cos excepte la commutativitat del producte.[5]

[5] Si no es compleix la propietat commutativa del producte cal ser més exigents en la propietat distributiva, l'element neutre i els inversos. L'element neutre 1 ha de complir que $1x = x1 = x$ per tot x, a la propietat distributiva, a més de $x(y + z) = xy + xz$ cal també exigir $(x + y)z = xz + yz$ i, en el cas dels anells de divisió, cal exigir que per cada x ha d'existir un y tal $xy = yx = 1$.

17 | Divisibilitat

La gran diferència entre els enters $\mathbb{Z}$ i els racionals $\mathbb{Q}$ es troba en la **divisibilitat**: com que $\mathbb{Q}$ és un cos, podem dividir per qualsevol nombre diferent de zero; a $\mathbb{Z}$, en canvi, la divisió només és possible en alguns casos i apareix el problema de la divisibilitat i tot el que això comporta: divisors i múltiples, unitats, mcd i mcm, elements irreductibles —també anomenats *primers*—, congruències, etc. A més, els mètodes elementals de resolució d'equacions lineals deixen de ser vàlids. Per exemple, una equació com $3x+2y=5$ que es resol de manera trivial a $\mathbb{Q}$, quan la volem resoldre a $\mathbb{Z}$ necessitem uns altres mètodes. L'estudi d'aquests problemes lligats a la divisibilitat formen el que es coneix com a **aritmètica**.

Comencem introduint diversos conceptes bàsics relacionats amb la divisibilitat. Són conceptes que tenen sentit en qualsevol anell, però en aquest moment ens interessen en el cas de l'anell dels enters.

- **Divisors**. Direm que d *divideix* a si existeix c tal que $a = dc$. Una notació que s'utilitza de vegades és $d \mid a$.
- **Múltiples**. Direm que a *és múltiple de* b si existeix c tal que $a = bc$, és a dir, si b divideix a.
- **Unitats**. Direm que u és una *unitat* o un element *invertible* si existeix u' tal que $uu' = 1$. Equivalentment, si $u \mid b$ per tot b. Si A és un anell, els elements invertibles de A formen un grup (amb la multiplicació) que es denota A^*.
- **Primers i irreductibles**. Direm que p és *primer* o *irreductible* si no és una unitat i els seus únics divisors són les unitats i els productes up on u és una unitat. Observem que, explícitament, considerem que les unitats no són elements primers.[1]

Tots aquests conceptes ens interessen ara a l'anell dels nombres enters $\mathbb{Z}$, però els hem definit de manera que les definicions tinguin sentit a qualsevol domini.

[1] Aquí estem considerant com a sinònims els termes *primer* i *irreductible* que, en general, són dos conceptes diferents que, tanmateix, coincideixen en els cas dels anells que estudiarem en aquest curs. Estrictament parlant, el que acabem de definir són els elements *irreductibles*, però en els exemples que estudiarem aquests elements coincideixen amb els elements *primers*, que es defineixen de la següent manera: p és primer si $p \neq 0$, p no és unitat i si p divideix ab, aleshores p divideix a o p divideix b. Vegeu l'exercici IV.13.

És clar que, sobre un cos, tots aquests conceptes són trivials: tot element diferent de zero divideix tot altre element, tot element diferent de zero és una unitat i no hi ha cap element primer.

En el cas de $\mathbb{Z}$, només hi ha dues unitats: $\mathbb{Z}^* = \{-1, 1\}$. Observem també que ± 1 divideix qualsevol enter i qualsevol enter divideix 0.

Els múltiples d'un element

Quina estructura té el conjunt de tots els múltiples d'un element fixat? Estudiem-ho.

Sigui A un domini —el cas que ens interessa ara és $A = \mathbb{Z}$— i sigui $a \in A$. Denotarem per $(a) \subseteq A$ el conjunt format per tots els múltiples de a.[2] Observem:

- Casos trivials: $(0) = \{0\}$, $(1) = A$. Si u és una unitat, aleshores és clar que $(u) = A$ i que $(a) = (ua)$ per tot $a \in A$. Recíprocament, si $(a) = (b)$, existeix una unitat $u \in A$ tal que $a = ub$. En el cas $A = \mathbb{Z}$ això vol dir que $(a) = (b)$ si i només si $a = \pm b$, i també que $\pm a$ són els dos elements de (a) amb valor absolut mínim entre tots els elements de (a).
- (a) és tancat respecte de la suma: si $b, c \in (a)$ aleshores $b + c \in (a)$. Es compleix que si $b \in (a)$, també $-b \in (a)$. Finalment, $0 \in (a)$. Per tot això, (a) és un *subgrup additiu* de A.
- (a) és tancat respecte de productes amb elements de A: si $b \in (a)$ i $c \in A$, aleshores $bc \in (a)$.

Totes aquestes propietats són molt senzilles de comprovar. Les dues últimes són tan importants que mereixen una definició:

> ***Definició.*** *Un subconjunt $I \subseteq A$ diem que és un* ***ideal*** *si és un subgrup additiu i compleix*
>
> $$a \in I,\ b \in A \ \Rightarrow\ ab \in I.$$

En conclusió, el conjunt de múltiples d'un element forma un *ideal* de l'anell. D'altra banda, en principi, hi pot haver ideals que no siguin de la forma (a). Els ideals de la forma (a) direm que són **ideals principals**.

La divisió amb residu

Centrem-nos ara en l'anell dels enters $\mathbb{Z}$ i tornem a considerar el problema de la divisió. Observem que no podem dividir, per exemple, 173 entre 8, però sí que

[2] Una altra notació que també s'utilitza per denotar el conjunt de tots els múltiples de a és $a\mathbb{Z}$, que potser és més explícita perquè, efectivament, els múltiples de a s'obtenen multiplicant a per tots els elements de $\mathbb{Z}$.

podem fer una **divisió amb residu** i obtenir

$$173 = 21 \times 8 + 5$$

on diem que 173 és el dividend, 8 és el divisor, 21 és el quocient i 5 és el **residu**. Aquest fet, l'existència —a l'anell $\mathbb{Z}$ i potser a algun altre anell, però no a tots— de la divisió amb residu, és una eina fonamental en l'estudi de l'aritmètica. Demostrem rigorosament l'existència de la divisió amb residu.

> ***Divisió amb residu***. *Siguin $D, d \in \mathbb{Z}$, $d \neq 0$. Existeixen $q, r \in \mathbb{Z}$ únics tals que*
>
> 1. $0 \leq r < |d|$.
> 2. $D = qd + r$.

Demostració. Ho demostrarem en el cas $D, d > 0$ perquè les altres possibilitats es dedueixen fàcilment d'aquesta (exercici **III**.5). Considerem aquest subconjunt (no buit) de $\mathbb{N}$

$$X := \{D - kd : k \in \mathbb{N}\} \cap \mathbb{N}$$

que, per una propietat fonamental dels nombre naturals, tindrà un mínim $r \geq 0$ que serà de la forma $r = D - qd$ per algun q. Ara cal demostrar que $r < d$. Si no fos així, escriuríem $r = d + s$ amb $s \geq 0$ i aleshores

$$r = D - qd > D - qd - d = D - (q + 1)d = s \geq 0$$

en contradicció amb que r sigui el mínim de X.

Demostrem ara la unicitat de q, r: Suposem $qd + r = q'd + r'$ amb $0 \leq r, r' < d$. Tindríem

$$|q - q'|d = |r' - r| < d$$

d'on obtenim $q = q'$ i $r = r'$. Això acaba la demostració del teorema sobre la divisió amb residu.

Aquesta divisió amb residu —també coneguda com a *divisió euclidiana*— és la que utilitzarem majoritàriament, però hi ha una segona versió que pot ser útil en algun cas: la divisió *per excés* en la qual expressem

$$D = q'd - r', \qquad 0 \leq r' < |d|.$$

Per exemple, la divisió per excés de 173 entre 8 és $173 = 22 \times 8 - 3$.

18 | $\mathbb{Z}$ és un DIP

Sabem què és un **domini** —un anell on el producte de dos elements diferents de zero no pot donar zero—, sabem què és un **ideal** —un subgrup tancat per productes amb elements de l'anell— i sabem què és un **ideal principal** —el format pels múltiples d'un element. Aleshores, un **domini d'ideals principals**, un DIP, és un domini en el que tots els ideals són principals. Aquest fet —que tots els ideals siguin principals— té, com veurem, conseqüències aritmètiques molt importants. Resulta que $\mathbb{Z}$ és un DIP:

Teorema. *$\mathbb{Z}$ és un DIP.*

Demostració. Sigui I un ideal de $\mathbb{Z}$. Hem de veure que I és principal, és a dir, I està format per tots els múltiples d'un cert element. Si $I = \{0\}$, ja hem acabat. En cas contrari, podem escollir un element $a \in I$ tal que a sigui mínim entre els elements positius de I. Evidentment, $(a) \subseteq I$ i tot es redueix a demostrar que $I \subseteq (a)$. L'eina clau de la demostració és la **divisió amb residu** que sabem que podem dur a terme a $\mathbb{Z}$. Sigui $b \in I$ i fem la divisió amb residu de b per a:

$$b = qa + r, \quad 0 \leq r < a.$$

Aleshores, és clar que $r = b - qa \in I$ i, com que a és mínim entre els elements positius de I, ha de ser $r = 0$. En conclusió, $b = qa \in (a)$ i hem acabat la demostració.

mcd i mcm

La primera conseqüència del fet que $\mathbb{Z}$ sigui un DIP és l'existència del **màxim comú divisor** i el **mínim comú múltiple** de dos nombres enters —o de qualsevol conjunt de nombres enters. En aquest apartat definirem amb precisió aquests dos conceptes i estudiarem les seves propietats més bàsiques.

Comencem observant que podem fer aquestes dues operacions senzilles amb ideals:

- **Intersecció d'ideals**. És trivial veure que si I i J són ideals d'un anell, aleshores $I \cap J$ també és un ideal de l'anell.

- **Suma d'ideals**. Si tenim dos ideals I, J d'un anell, podem considerar aquest conjunt
$$I + J := \{a + b : a \in I,\ b \in J\}$$
format per les sumes d'elements dels dos ideals. És un exercici fàcil comprovar que $I + J$ és també un ideal.

Si $a, b \in \mathbb{Z}$ podem considerar els ideals principals (a) i (b) i podem aplicar, a aquests dos ideals, les operacions d'intersecció i suma que acabem de veure. Per evitar casos trivials, suposarem que $a, b \neq 0$.

- $(a) \cap (b)$ serà també un ideal de $\mathbb{Z}$ i, **com que $\mathbb{Z}$ és un DIP**, serà un ideal principal:
$$(a) \cap (b) = (m)$$
per un cert $m \in \mathbb{Z}$, únic llevat del signe. Prenem $m \geq 0$. Aquest element m compleix aquestes propietats:
  - m és múltiple de a i múltiple de b.
  - $m \neq 0$ perquè $ab \in (m)$ i $ab \neq 0$.
  - m és el més petit entre els múltiples comuns positius de a i b.
  - m també és *mínim* en aquest segon sentit: si m' és múltiple de a i múltiple de b, aleshores m' és múltiple de m.
  - Qualsevol de les dues propietats anteriors caracteritza m entre els enters positius.

  En definitiva, podem dir que m és el **mínim comú múltiple** de a i b. Escriurem $m = \operatorname{mcm}(a, b)$.

- $(a) + (b)$ serà també un ideal de $\mathbb{Z}$ i, **com que $\mathbb{Z}$ és un DIP**, serà un ideal principal:
$$(a) + (b) = (d)$$
per un cert $d \in \mathbb{Z}$, únic llevat del signe. Prenem $d \geq 0$. Aquest element d compleix aquestes propietats:
  - d divideix a i divideix b. En particular, $d \neq 0$.
  - **Identitat de Bézout**. Existeixen $n, m \in \mathbb{Z}$ tals que
$$d = na + mb.$$
    Aquesta conseqüència immediata de la definició de d és extremadament important i útil, com anirem veient.
  - d és *màxim* en aquest sentit: si d' també divideix a i divideix b, aleshores d' divideix d. Aquesta propietat es demostra immediatament a partir de la identitat de Bézout.

- d és el més gran dels divisors comuns de a i b, entenent *gran* en l'ordre habitual dels nombres enters.
- Qualsevol de les dues propietats anteriors caracteritza d entre els enters positius.

En definitiva, podem dir que d és el **màxim comú divisor** de a i b. Escriurem $d = \text{mcd}(a, b)$. Si $d = 1$, direm que a i b són **coprimers**.

També té sentit parlar de mcm i mcd quan algun dels dos enters és zero. Tenim $\text{mcm}(0, b) = 0$ i $\text{mcd}(0, b) = b$.

L'algorisme d'Euclides

Euclides ens ensenya un algorisme —a les dues primeres proposicions del llibre setè dels *Elements*— que ens permet calcular fàcilment el màxim comú divisor de dos nombres enters. Podem dir que és l'algorisme més antic que coneixem que encara s'utilitza avui dia. Es fonamenta en la divisió amb residu i, específicament, en aquest lema fàcil de demostrar:

Lema. *Si* $D = qd + r$, *aleshores* $\text{mcd}(D, d) = \text{mcd}(d, r)$.

A partir d'aquest lema, l'algorisme d'Euclides funciona d'aquesta manera. Suposem que tenim dos enters $a, b > 0$ amb $a > b$.

- Fem la divisió amb residu de a per b. Obtenim $a = q_0 b + r_0$ amb $r_0 < b$.
- Fem la divisió amb residu de b per r_0. Obtenim $b = q_1 r_0 + r_1$ amb $r_1 < r_0$.
- Fem la divisió amb residu de r_0 per r_1. Obtenim $r_0 = q_2 r_1 + r_2$ amb $r_2 < r_1$.
- repetim el procés. Com que els residus (que no poden ser negatius) són cada vegada més petits, arribarem a un residu $r_n = 0$ i ens aturem aquí.

Aleshores, segons el lema,

$$\text{mcd}(a, b) = \text{mcd}(b, r_0) = \text{mcd}(r_0, r_1) = \cdots = \text{mcd}(r_{n-2}, r_{n-1}) = \text{mcd}(r_{n-1}, 0) = r_{n-1}.$$

Hem dit que la *identitat de Bézout* és una eina molt important a l'aritmètica. L'algorisme d'Euclides que ens ha permès calcular ràpidament el màxim comú divisor de dos nombres també en permet calcular ràpidament els enters n i m de la identitat de Bézout. En efecte, cada pas de l'algorisme d'Euclides és una divisió amb residu de la forma

$$r_i = q_{i+2} r_{i+1} + r_{i+2}$$

i aquesta igualtat ens permet escriure r_{i+2} com a combinació lineal de r_i, r_{i+1}. Aleshores, treballant recursivament podrem obtenir $r_{n-1} = \mathrm{mcd}(a, b)$ com a combinació lineal de a i b.

Exemple. Prenem $a = 177$, $b = 39$. Els passos de l'algorisme d'Euclides són aquests:

$$\begin{aligned} 177 &= 4 \times 39 + 21 \\ 39 &= 1 \times 21 + 18 \\ 21 &= 1 \times 18 + 3 \\ 18 &= 6 \times 3 + 0 \end{aligned}$$

Per tant, $\mathrm{mcd}(177, 39) = 3$. També:

$$\begin{aligned} 3 = 21 - 18 &= 21 - (39 - 21) = 2 \times 21 - 39 \\ &= 2\,(177 - 4 \times 39) - 39 = 2 \times 177 - 9 \times 39 \end{aligned}$$

I això ens dóna la identitat de Bézout entre 177 i 39:

$$2 \times 177 - 9 \times 39 = 3.$$

Més endavant entendrem fins a quin punt tota aquesta teoria mil·lenària té una importància immensa en la revolució digital que estem vivint.

Finalment, quan hem calculat el màxim comú divisor de dos nombres, el càlcul del mínim comú múltiple és immediat (excloem el cas $a = b = 0$):

$$\mathrm{mcm}(a, b) = \frac{ab}{\mathrm{mcd}(a, b)}.$$

Demostració. Escrivim $d := \mathrm{mcd}(a, b)$ i $m := ab/d$. En primer lloc, és evident que m és múltiple de a i és múltiple de b. Segons la caracterització del mcm que hem vist abans, n'hi ha prou amb demostrar que si m' és també múltiple de a i de b, aleshores m' divideix m. Escrivim la identitat de Bézout entre a i b

$$d = ra + sb.$$

Aleshores, si $m' = ak = bl$, tenim

$$dl = ral + sbl = ral + sak \Rightarrow rl + sk = \frac{dl}{a} \Rightarrow \frac{dl}{a} \in \mathbb{Z}.$$

Per tant,

$$m' = bl = \frac{ab}{d}\,\frac{dl}{a} = m\frac{dl}{a} \Rightarrow m | m'$$

i això acaba la demostració.

19 | Els nombres primers

En aquest capítol començarem a parlar dels **nombres primers**. No caldria insistir en la importància immensa que tenen els nombres primers, però és apropiat fer esment d'aquests punts:

- Hi ha bons motius per considerar-los com **els àtoms** dels que estan fets tots els nombres. Com demostrarem aviat, tot nombre té una expressió única com a producte de primers i els primers són *indivisibles*, en un sentit concret que veurem.
- La importància dels nombres primers transcendeix la teoria de nombres i s'escampa per totes les matemàtiques. A moltes teories matemàtiques és possible fer una mena de *descomposició aritmètica* dels problemes que ens permet atacar-los *per a cada primer per separat*.
- Amb la *revolució digital*, els nombres primers s'han convertit en un **producte industrial de primera necessitat** fins el punt que la recerca sobre nombres primers ha adquirit un gran valor estratègic. La seguretat digital —com veurem més endavant— es fonamenta en la teoria dels nombres primers i un avenç significatiu en aquest camp pot millorar —o destruir!— la seguretat digital mundial, amb unes conseqüències inimaginables i, potser, catastròfiques.

Recordem que hem definit el concepte d'element primer (o irreductible) en qualsevol domini, però ara ens centrarem en els primers del domini $\mathbb{Z}$. En aquest cas, la definició d'element primer que vam donar al capítol 17 es concreta així:

> *Diem que $p \in \mathbb{Z}$ és **primer** si $p \neq \pm 1$ i els únics divisors de p són ± 1 i $\pm p$.*

Dèiem abans que els primers són com els àtoms. Aquestes primeres propietats bàsiques dels nombres primers ho justifiquen:

- *Si p és primer i p divideix ab, aleshores p divideix a o p divideix b.* Ho podem demostrar d'aquesta manera. Suposem que p no divideix a. En

aquest cas, per la definició de nombre primer, tindrem que $\mathrm{mcd}(p, a) = 1$ i la identitat de Bézout entre p i a

$$1 = np + ma$$

dóna, multiplicant per b als dos costats de la igualtat, $b = npb + mab$. Aleshores, p ha de dividir b.

- *Si c divideix ab i a, c són coprimers, aleshores c divideix b.* És una generalització de la propietat anterior que es demostra de la mateixa manera.
- *Si a i b divideixen k i a i b són coprimers, aleshores ab també divideix k.* Per veure-ho, suposem $k = ar = bs$, escrivim la identitat de Bézout $1 = na + mb$ i, multiplicant per k, observem que

$$k = nak + mbk = nabs + mbar.$$

Passem ara a un resultat fonamental —que reforça encara més l'analogia entre nombres primers i àtoms: tot enter es pot expressar com a producte de nombres primers i, a més, aquesta *descomposició en primers* és essencialment única.

Teorema de factorització en primers. *Tot $n \in \mathbb{Z}$, $n \neq 0, \pm 1$, s'expressa de manera única en la forma*

$$n = \pm p_1 p_2 \cdots p_k$$

on $p_1 \leq p_2, \cdots \leq p_k$ són nombres primers positius.

Evidentment, n'hi ha prou amb demostrar aquest teorema quan $n > 1$. La demostració té dues parts: existència de la factorització i unicitat de la factorització.

Existència de la factorització. Ho demostrarem pel mètode del contraexemple minimal. Sigui $n > 1$ el menor enter que no es pot expressar com a producte de primers. Això vol dir que n no és primer i, per tant, podem escriure $n = mn'$ amb $m, n' > 1$. Com que n és un contraexemple minimal i com que m i n' són $< n$, m i n' sí que es poden escriure com a producte de primers i, per tant, n també.

Unicitat de la factorització. També ho demostrarem pel mètode del contraexemple minimal. Suposem que n és el mínim enter > 1 que té dues descomposicions diferents:

$$n = p_1 \cdots p_k = q_1 \cdots q_r$$

amb els primers positius i ordenats en ordre creixent. Observem ara que p_1 divideix $q_1 \cdots q_r$ i, per una propietat dels primers que hem vist anteriorment, p_1 dividirà q_s per algun s. Però q_s és primer i, per tant, $p_1 = q_s$. Això ens diu que podem simplificar la igualtat anterior dividint per p_1 i obtenir dues descomposicions de $n/p_1 < n$. Com que n era el contraexemple minimal, aquestes dues

descomposicions de n/p_1 seran iguals i ara és fàcil veure que això només pot passar si les dues descomposicions de n són també iguals.

Aquesta propietat de $\mathbb{Z}$ consistent en que tot element té una descomposició única en producte d'elements irreductibles té una gran importància en l'estudi dels nombres enters. No és estrany pensar que també ha de tenir una gran importància en altres àmbits on es compleixi. Això justifica la introducció d'aquesta definició:

> *Un domini que compleixi el teorema anterior*[1] *direm que és un* ***domini de factorització única*** *(DFU). Per exemple, $\mathbb{Z}$ és un DFU.*[2]

Quants primers hi ha?

Ara que comencem a entendre la gran importància dels nombres primers, és lògic que ens preguntem quants n'hi ha. La resposta a aquesta pregunta la coneixem, com a mínim, des dels temps d'Euclides, que a la proposició 20 del llibre IX afirma i demostra que *hi ha més nombres primers que qualsevol multitud donada de nombres primers*. En el nostre context actual diríem això mateix d'aquesta manera:

> **Teorema**. *Hi ha infinits nombres primers.*

Demostració. Suposem que hi hagués una quantitat finita de nombres primers (positius) $p_1, \ldots, p_k$. Considerem aquest nombre enter:

$$n = p_1 \cdots p_k + 1.$$

És evident que $n \neq p_1, \ldots, p_k$ i, per tant, no pot ser primer. Sabem que n s'expressarà com a producte de primers i, en particular, hi haurà un $1 \leq i \leq k$ tal que p_i divideix n. Però això és impossible perquè, per la manera com hem definit n, tindríem que p_i divideix 1, que és absurd.

Per tant, la resposta a la pregunta de quants primers hi ha és: n'hi ha infinits. Però els matemàtics no ens podem donar per satisfets amb aquesta resposta perquè ens agradaria anar més enllà i preguntar-nos, per exemple, per la **densitat** dels nombres primers entre tots els nombres. Per exemple, sabem que hi ha infinits nombres parells però també sabem que un de cada dos nombres és parell, una densitat del 50%. Podem arribar a saber quina és la densitat dels primers? Cada quants nombres n'hi ha un de primer? Quants primers hi ha, per exemple, entre 10^{50} i 10^{51}?

[1] Convenientment reformulat per evitar l'ús de la relació d'ordre $\leq$ que no té sentit en un anell general. Caldrà dir que *la descomposició és única llevat de l'ordre dels factors*.

[2] Hi ha un teorema que afirma que tot DIP és un DFU. La demostració utilitza l'axioma de l'elecció.

Si fem alguns càlculs amb nombres petits, observem que els nombres primers es van fent cada vegada més escassos a mida que treballem amb nombres més i més grans. Per exemple, hi ha 25 primers ≤ 100, és a dir, tenim aquí una densitat del 25%, però només hi ha 168 primers ≤ 1000, una densitat del 16.8%. Fen càlculs més extensos obtenim aquests resultats:

interval	densitat
[2,100]	0.25
[2,1000]	0.168
[2,10000]	0.1229
[2,100000]	0.09592
[2,1000000]	0.078498
[2,10000000]	0.0664579
[2,100000000]	0.05761455

Si designem per $\pi(x)$ el nombre de primers a l'interval $[2, x]$, la **densitat** dels nombres primers la podem definir com la funció

$$\rho(x) := \frac{\pi(x)}{x}$$

i ens podem preguntar si aquesta funció $\rho(x)$ es pot calcular o aproximar per alguna funció matemàtica coneguda. Aquesta pregunta va preocupar Legendre i Gauss (quan tenia quinze anys!) a finals del segle XVIII. Els dos matemàtics van conjecturar que aquesta densitat es comportava com la funció $1/\log(x)$.[3] Concretament, això és el que van conjecturar:

> **El teorema dels nombres primers**. *Quan $x \to \infty$, la funció $\rho(x)$ és asimptòtica a la funció*
>
> $$\frac{1}{\log x}.$$

Van ser diversos els grans matemàtics que van treballar en la demostració d'aquest importantíssim teorema. Finalment, després dels decisius avenços que havia fet Bernhard Riemann, van ser Jacques Hadamard i Charles-Jean de la Vallée-Poussin els que van completar —independentment un de l'altre— la demostració el 1896, cent anys després que Legendre i Gauss conjecturessin el resultat.

[3] De fet, la funció que conjecturava Legendre era $1/(\log(x) - A)$ on A era una constant que, segons Legendre, tenia un valor aproximat de 1.08366.

20 | Aritmètica modular

La idea de les **congruències**, la idea de **«reduir un problema aritmètic mòdul k»** és una idea antiga extraordinàriament poderosa. A més, en el moment actual, té una importància tècnica i econòmica que era impensable no fa gaires anys. En aquest capítol definirem el concepte de congruència i estudiarem les seves propietats bàsiques. També donarem dues aplicacions immediates —però molt interessants— que ens mostraran com les congruències poden ser útils.

L'àmbit on treballem és el dels nombres enters $\mathbb{Z}$. Comencem fixant un **mòdul**, que és un enter $k > 1$ qualsevol.

> **Definició**. *Si $a, b \in \mathbb{Z}$, direm que a i b són **congruents** mòdul k si $a - b$ és múltiple de k. Escriurem $a \equiv b \mod k$ i, si k es sobreentén, escriurem simplement $a \equiv b$.*

Observem aquestes propietats senzilles de les congruències:

- La congruència és una relació d'equivalència en el conjunt $\mathbb{Z}$.
- La congruència és compatible amb les operacions de suma i producte. És a dir, si $a \equiv a'$ i $b \equiv b'$, també $a + b \equiv a' + b'$ i $ab \equiv a'b'$.
- Cada nombre enter és congruent a exactament un d'aquests enters

  $$0, 1, \ldots, k - 1.$$

  Per demostrar això, si tenim $a \in \mathbb{Z}$ fem la divisió amb residu $a = qk + r$ amb $0 \leq r < k$, que ens mostra que $a \equiv r \in \{0, 1, \ldots, k - 1\}$. Pel que fa a la unicitat, si $i, j \in \{0, 1, \ldots, k - 1\}$ i $i \equiv j$, tindrem $i = j + sk$ per algun s, però això només és possible si $s = 0$ i $i = j$.

  Si $a \equiv r$ per $0 \leq r < k$, direm que r és la *reducció de a mòdul k*.
- Si a, k són coprimers, l'equació $aX \equiv 1$ té solució i dues solucions són congruents. En efecte, la identitat de Bézout ens donarà enters r, s tals que $ra + sk = 1$ i, per tant, r serà una solució de l'equació $aX \equiv 1$. Si r, r' fossin solucions, tindríem $ar \equiv ar' \equiv 1$. Per tant, $r \equiv rar' \equiv r'$. Observem que, gràcies a l'algorisme d'Euclides, podem calcular explícitament la identitat

de Bézout de dos enters i, en conseqüència, podem trobar explícitament la solució de l'equació $aX \equiv 1$. La condició que a sigui coprimer amb el mòdul k és necessària per tal que l'equació tingui solució, perquè si $ax \equiv 1$ aleshores $ax = rk + 1$ i a, k són coprimers.

- Si a, k són coprimers, també podem resoldre explícitament qualsevol equació de la forma $aX \equiv c$. La identitat de Bézout ens diu que existeixen enters n, m tals que $na + mk = 1$. Aleshores, $x = nc$ és solució de l'equació i totes les solucions són de la forma $nc + hk$ amb $h \in \mathbb{Z}$.
- Si k és primer, aleshores $ab \equiv 0$ implica $a \equiv 0$ o $b \equiv 0$. Això és una conseqüència immediata del fet que si un primer divideix un producte ha de dividir algun dels factors (pàgina 136).

Equacions diofàntiques lineals

El terme *equació diofàntica*[1] fa referència a qualsevol equació que només involucri nombres enters i les operacions de suma i producte i de la qual volem trobar les solucions en que les incògnites tenen valors que siguin nombres enters. N'hem parlat breument al capítol 6.

Una primera aplicació de les congruències és que permeten trobar les solucions enteres d'una equació diofàntica lineal en dues variables

$$aX + bY = c.$$

Vegem com podem resoldre aquesta equació. Observem en primer lloc que sense l'exigència que les solucions siguin enteres —és a dir, si no tenim en compte el caràcter *diofàntic* del problema— el problema és trivial: si $b \neq 0$, X pot ser qualsevol nombre racional i Y ha de ser $(c - aX)/b$.

Descartem els casos trivials $a, b = 0, \pm 1$ i comencem trobant el màxim comú divisor de a i b. Sigui $d := \mathrm{mcd}(a, b)$. Aleshores, és trivial adonar-se que

Si d no divideix c, l'equació no té solució.

En cas contrari, podem dividir tota l'equació per d i obtenim una equació equivalent (vol dir: amb les mateixes solucions) en la qual a i b són coprimers. Ara hem de resoldre aquesta equació. La idea crucial és

Idea: *reduïm mòdul b i passem a l'equació d'una variable $aX \equiv c$* mòd b.

[1] El nom fa referència al matemàtic del segle III Diofant l'Alexandria.

Aquesta nova equació, com que a, b són coprimers, ja hem vist abans com la podem resoldre explícitament: escrivim la identitat de Bézout entre a i b, $a'a + b'b = 1$ i totes les solucions de $aX \equiv c$ són

$$X = a'c + \lambda b, \quad \lambda \in \mathbb{Z}.$$

Ara només ens cal trobar els valors de Y. Ho fem així: a partir de $aX + bY = c$ obtenim

$$aa'c + \lambda ab + bY = c.$$

Com que $aa' = 1 - b'b$ arribem a la conclusió que les solucions de l'equació diofàntica són

$$\left.\begin{aligned} Y &= b'c - \lambda a \\ X &= a'c + \lambda b \end{aligned}\right\}$$

on λ varia sobre els nombres enters. Observem que el mètode de resolució, gràcies a l'algorisme d'Euclides, és totalment efectiu i es pot programar fàcilment perquè el resolgui un ordinador de manera força ràpida.[2]

El teorema xinès del residu

El que es coneix com el *teorema xinès del residu* és un mètode per trobar un enter n sabent les seves reduccions mòdul diversos enters coprimers $n_1, \ldots, n_r$. Se li dóna aquest nom perquè el resultat apareix per primera vegada en una obra del segle III del matemàtic xinès Sunzi.

Suposem que tenim enters > 1 $n_1, \ldots, n_r$, i suposem que són coprimers dos a dos. Suposem que tenim enters $a_1, \ldots, a_r$ i volem resoldre el sistema d'equacions

$$X \equiv a_i \mod n_i,\ 1 \leq i \leq r.$$

El teorema xinès del residu afirma que, en aquestes condicions,

- El sistema té solució i es pot calcular explícitament.
- Si n és una solució, totes les solucions són $n + \lambda n_1 \cdots n_r$, variant $\lambda \in \mathbb{Z}$.

La solució es troba de manera inductiva sobre el nombre r d'equacions.

Cas de dues equacions. Hem de resoldre

$$\begin{aligned} X &\equiv a_1 \mod n_1 \\ X &\equiv a_2 \mod n_2 \end{aligned}$$

[2] Un cop resoltes les equacions lineals en dues variables, per un mètode inductiu podrem resoldre les equacions diofàntiques lineals en diverses variables. Vegeu l'exercici **IV**.26.

amb $\text{mcd}(n_1, n_2) = 1$. Comencem escrivint la identitat de Bézout $m_1n_1 + m_2n_2 = 1$. Aleshores,

$$x := a_1 + (a_2 - a_1)m_1n_1 = a_1 + (a_2 - a_1)(1 - m_2n_2)$$

és clarament una solució del sistema. D'altra banda, si y és una altra solució, haurà de ser $x \equiv y \mod n_1$ i també $x \equiv y \mod n_2$. Per tant, n_1 i n_2 divideixen $x - y$ i, com que n_1, n_2 són coprimers, per una propietat dels nombres primers vista a la pàgina 137, tenim que n_1n_2 també divideix $x - y$, amb la qual cosa $x \equiv y \mod n_1n_2$.

Cas de $r > 2$ equacions. Resolem les dues primeres equacions pel mètode anterior. Trobarem b tal que les solucions de les dues primeres equacions són les mateixes de $X \equiv b \mod n_1n_2$. Això ens redueix el problema a $r-1$ equacions. Inductivament, arribarem a la solució del sistema inicial de r equacions.

21 Els anells $\mathbb{Z}/(m)$ i els cossos finits

Recordeu el *conjunt quocient*? Recordeu que dèiem que era una de les operacions més poderoses de les matemàtiques perquè ens permet considerar com iguals coses que no ho són? En aquest capítol utilitzarem aquella idea per convertir la **congruència** d'enters en igualtat i, molt més important que això, construir unes noves **estructures** que juguen un paper crucial en l'aritmètica.[1]

Recordem, doncs, que a $\mathbb{Z}$, un cop hem escollit un *mòdul* $m > 1$, tenim una relació anomenada de congruència que *és una relació d'equivalència*. Per tant, podem fer el **pas al quocient**

$$\mathbb{Z} \longrightarrow \mathbb{Z}/\sim$$

que associa a cada enter a la seva classe d'equivalència $[a] \in \mathbb{Z}/\sim$. Pel que hem estudiat abans, hi ha exactament m classes d'equivalència

$$\mathbb{Z}/\sim = \{[0], [1], \ldots, [m-1]\}.$$

Si recordem el concepte de *quocient d'un grup per un subgrup normal* (pàgina 108) i tenim en compte que, en el cas dels grups abelians, tots els subgrups

[1] Aquí tindrem un exemple paradigmàtic d'una de les idees més importants de les matemàtiques: la idea d'**estructura**. És una idea difícil d'explicar, que només es pot arribar a entendre en tota la seva magnitud quan l'aprenent de matemàtic vagi avançant en la seva carrera. Donem ara alguns exemples. No és el mateix conèixer molt bé les trajectòries aparents dels planetes vistos per un observador situat a la Terra (com va fer Tycho Brahe), que elaborar un *sistema del món* com va fer Johannes Kleper. No és el mateix saber que en el pòquer és més fàcil tenir una doble parella que un trio, que formular axiomàticament el concepte d'espai de probabilitat com va fer Kolmogorov. No és el mateix ser un geni de la simetria com els artistes de l'Alhambra que imaginar l'estructura de grup que regeix i unifica totes les idees de simetria i permet plantejar-se la seva classificació. No és el mateix posar taronges en una pila ben feta que entendre que al darrera d'aquesta acció hi ha l'estructura fonamental i extraordinàriament multidisciplinar de *reticle*. No és el mateix intuir que *les rectes paral·leles es tallen a l'infinit* que descobrir l'*espai projectiu*, l'àmbit natural dels fenòmens geomètrics. No és el mateix veure que els nombres negatius són molt pràctics per indicar deutes o temperatures per sota del nivell de congelació de l'aigua, que imaginar l'estructura de DIP de $\mathbb{Z}$, l'àmbit natural de l'aritmètica. Ni és el mateix utilitzar coordenades cartesianes per situar un punt al pla o a l'espai que concebre l'estructura d'espai vectorial, una de les més fonamentals de les matemàtiques. I, tornant al tema d'aquest capítol, no és el mateix aprendre a treballar amb **congruències** que adonar-se de l'existència d'unes **estructures** noves i potents com les que estudiarem en aquest capítol.

són automàticament normals, ens adonarem que el conjunt quocient $\mathbb{Z}/\sim$ és exactament el mateix que el grup quocient del grup $\mathbb{Z}$ pel subgrup (m) format pels múltiples de m. En particular, $\mathbb{Z}/(m)$ és un grup abelià i el pas al quocient

$$\mathbb{Z} \longrightarrow \mathbb{Z}/(m)$$

és un homomorfisme de grups. L'operació de suma a $\mathbb{Z}/(m)$ ve donada per

$$[a] + [b] := [a + b].$$

Encara més: a $\mathbb{Z}/(m)$ podem definir una *multiplicació*

$$[a][b] := [ab]$$

i és un exercici fàcil veure que aquesta multiplicació està ben definida i dota $\mathbb{Z}/(m)$ d'estructura l'**anell**.[2] En conclusió:

> *Si fixem un mòdul $m > 1$ obtenim un anell $\mathbb{Z}/(m)$ amb m elements, i un homomorfisme d'anells*[3]
>
> $$\pi : \mathbb{Z} \longrightarrow \mathbb{Z}/(m).$$

És a dir, tenim uns nous objectes matemàtics —els anells finits $\mathbb{Z}/(m)$— i uns homomorfismes $\pi : \mathbb{Z} \to \mathbb{Z}/(m)$ que ens permeten *reduir mòdul m* qualsevol problema aritmètic sobre nombres enters i obtenir altre un problema aritmètic —probablement més fàcil d'atacar— sobre un anell finit.

Què hi guanyem —i què hi perdem— quan passem de $\mathbb{Z}$ a $\mathbb{Z}/(m)$?

- $\mathbb{Z}$ és infinit, mentre que $\mathbb{Z}/(m)$ és finit i té exactament m elements, cadascun dels quals és una classe d'equivalència d'enters:

  $$[k] = \{k + \lambda m : \lambda \in \mathbb{Z}\}.$$

  Normalment, si no hi ha perill de confusió, l'element $[k] \in \mathbb{Z}/(m)$ es denota simplement k. És a dir:

  $$\mathbb{Z}/(m) = \{0, 1, 2, \ldots, m-1\}.$$

- $\mathbb{Z}$ és un domini, és a dir, no té *divisors de zero*:

  $$a, b \in \mathbb{Z},\ ab = 0 \quad \Rightarrow \quad a = 0 \text{ o } b = 0.$$

  En canvi, $\mathbb{Z}/(m)$ pot tenir-ne o no tenir-ne. De què depèn? La resposta ens la dóna aquest resultat important:

[2] Com hem indicat a la pàgina 130, aquest anell també es denota $\mathbb{Z}/m\mathbb{Z}$.

[3] Un *homomorfisme d'anells* és una aplicació $\phi : A \to B$ entre dos anells que és homomorfisme de grups (respecte de la suma) i, a més, compleix $\phi(1) = 1$ i, per tot $x, y \in A$, $\phi(xy) = \phi(x)\phi(y)$.

Teorema. *$\mathbb{Z}/(m)$ és un **domini** si i només si m és **primer**.*

La demostració és molt senzilla. Si m no és primer, podem escriure $m = uv$ amb $u, v \neq \pm m$. Aleshores, a $\mathbb{Z}/(m)$ tenim $uv = m = 0$, però $u, v \neq 0$. Recíprocament, si $uv = 0$ a $\mathbb{Z}/(m)$ això vol dir que uv és un múltiple de m o, equivalentment, que m divideix uv. Si m és primer, això implica que m divideix u o m divideix v que, a $\mathbb{Z}/(m)$, vol dir $u = 0$ o $v = 0$.

- Des del punt de vista de la divisibilitat, $\mathbb{Q}$ i $\mathbb{Z}$ se situen en dos extrems oposats: $\mathbb{Q}$ és *«dòcil»* perquè tot element diferent de zero és invertible i $\mathbb{Z}$ és *«molt esquerp»* perquè només té dos elements invertibles: $1, -1$. En canvi, $\mathbb{Z}/(m)$ pot tenir molts més elements invertibles que $\mathbb{Z}$. Quants? Quins són els elements invertibles de $\mathbb{Z}/(m)$? Veurem que aquesta és una pregunta important, encara que la resposta sigui ben senzilla.

  Teorema. *a és invertible a $\mathbb{Z}/(m)$ si i només si a, m són coprimers.*

  La demostració és fàcil amb el que hem estudiat fins ara. Si a, m són coprimers, podem escriure la identitat de Bézout $na + sm = 1$ i deduir que, a $\mathbb{Z}/(m)$, tenim $na = 1$. Recíprocament, si a és invertible tindrem $aa' \equiv 1$ mòd m, és a dir, $aa' = 1 + \lambda m$ i d'aquí es dedueix immediatament (lema de la pàgina 134) que $\mathrm{mcd}(a, m) = 1$.

- Si m és primer —i només si m és primer— aleshores $\mathbb{Z}/(m)$ **és un cos!** Això es dedueix del que acabem de dir. Aquest fet és transcendental! Si considerem congruències mòdul un primer fugim de l'esquerp planeta dels enters i aterrem en una estructura tan potent com és **un cos**! Un cos finit! De fet, una família infinita de cossos finits, un per a cada primer:

  $$\mathbb{Z}/(2),\ \mathbb{Z}/(3),\ \mathbb{Z}/(5),\ \mathbb{Z}/(7),\ \mathbb{Z}/(11),\ \mathbb{Z}/(13),\ \ldots$$

  Observem, doncs, que el pas al quocient $\mathbb{Z} \to \mathbb{Z}/(p)$ on p és qualsevol primer ens permet traslladar qualsevol problema aritmètic sobre nombres enters a un problema sobre un cos finit. Es-pec-ta-cu-lar!!!!

Els cossos finits

Acabem de descobrir una família infinita de cossos finits. Els cossos finits —els que acabem de descobrir i d'altres que existeixen— són estructures extraordinàriament importants. Durant segles, han tingut una importància teòrica molt gran; ara, amb la revolució digital, també tenen una importància pràctica no gens menor. L'estudi dels cossos finits formarà part del programa d'una assignatura que l'estudiant cursarà més endavant, i ara no és el moment de dur-lo a terme perquè no tenim ni el temps ni les eines per fer-ho. Però sí que pot ser interessant fer un breu resum de quins són els cossos finits, més enllà dels que acabem de descobrir.

Comencem introduint el concepte de **característica** d'un cos.

> *Sigui k un cos. Definim la seva **característica** com el mínim enter $n > 0$ tal que $1 + \cdots + 1 = 0$, on a la suma hi ha exactament n termes 1. Si sumant 1 mai no obtenim zero —com passa, per exemple, en els cossos $\mathbb{Q}$ i $\mathbb{R}$— direm que la característica és 0.*

Com a exemples, tenim cossos de característica zero —que ja coneixíem— com els racionals $\mathbb{Q}$ i els reals $\mathbb{R}$ i cossos amb característica p per a qualsevol primer p, com els cossos $\mathbb{Z}/(p)$ que acabem de descobrir. És fàcil veure que un cos finit no pot tenir mai característica zero.

Els fets bàsics sobre els cossos finits —que ara no podrem demostrar— són aquests:

- Per cada primer $p \geq 2$ i cada $r > 0$ existeix un cos finit amb p^r elements, denotat $\mathbb{F}_{p^r}$. Aquest cos té característica p.
- Tot cos finit té p^r elements, per algun primer $p \geq 2$ i algun enter $r > 0$.
- Dos cossos finits amb el mateix nombre d'elements són *isomorfs*.

És a dir, per cada p^r hi ha, essencialment, un únic cos amb p^r elements. Evidentment, quan $r = 1$ tenim els cossos $\mathbb{F}_p = \mathbb{Z}/(p)$ que ja coneixem. Com són la resta de cossos finits que encara no coneixem? No ho podem explicar —però alguna cosa en direm més endavant. El que és clar és que, si $r > 1$, el cos $\mathbb{F}_{p^r}$ **no** és $\mathbb{Z}/(p^r)$ perquè aquest anell, com hem vist, té divisors de zero.

Per tant, els cossos més petits que hi ha són aquests:

$$\mathbb{F}_2,\ \mathbb{F}_3,\ \mathbb{F}_4,\ \mathbb{F}_5,\ \mathbb{F}_7,\ \mathbb{F}_8,\ \mathbb{F}_9,\ \mathbb{F}_{11},\ \mathbb{F}_{13},\ \mathbb{F}_{16},\ldots$$

Els que tenen un nombre primer d'elements ja els coneixem

$$\mathbb{F}_2 = \mathbb{Z}/(2) = \{0,1\},\quad \mathbb{F}_3 = \mathbb{Z}/(3) = \{0,1,2\} = \{0,1,-1\},\quad \ldots$$

Per tant, el cos més petit que encara no coneixem és el cos que té 4 elements. Aquest cos és tan petit que simplement per assaig–error podem escriure les seves taules de sumar i multiplicar. Serà $\mathbb{F}_4 = \{0,1,a,b\}$ i els elements no-nuls $\{1,a,b\}$ formaran un grup de tres elements. A l'exercici **III**.18 ja vam veure que hi ha una única possibilitat, donada per aquesta taula de multiplicar:

$\cdot$	1	a	b
1	1	a	b
a	a	b	1
b	b	1	a

En canvi, el grup additiu té 4 elements i en aquell mateix exercici vam veure que hi havia dues possibilitats i n'hem de descartar una. Demostrarem que $1+1=0$. Suposem $1+1=a$ (el cas $1+1=b$ és equivalent). Aleshores, $1=ab=(1+1)b=b+b$ i la taula de sumar tindria aquesta forma

$+$	0	1	a	b
0	0	1	a	b
1	1	a	?	?
a	a	?	?	?
b	b	?	?	1

Si ara recordem que a cada fila i a cada columna no hi pot haver repeticions, veiem que no és possible completar la taula. Per tant, $1+1=0$ i la taula de sumar ha de ser necessàriament aquesta:

$+$	0	1	a	b
0	0	1	a	b
1	1	0	b	a
a	a	b	0	1
b	b	a	1	0

que correspon al grup que a l'exercici **III**.18 vam denotar K_4. Observem, doncs, que el cos de 4 element $\mathbb{F}_4$ té característica 2 ($x+x=0$ per tot $x\in\mathbb{F}_4$) i està format per quatre elements 0, 1, a, $1+a$ que compleixen $a^2+a+1=0$.

22 La funció φ d'Euler

El 1763 Leonhard Euler va considerar per primera vegada una certa funció definida sobre els enters positius que ha tingut, des d'aquell moment, un paper important en el desenvolupament de l'aritmètica. Actualment, a més, aquesta funció és molt rellevant per al desenvolupament de la seguretat digital, com veurem més endavant. La funció en qüestió s'acostuma a conèixer com *la funció φ d'Euler*[1] i en aquest capítol l'estudiarem des d'un punt de vista teòric.

La **definició** de la funció φ és molt senzilla. Donarem dues definicions equivalents:

1. Si $n > 0$, definim $\varphi(n)$ com la quantitat d'enters entre 1 i n (incloent 1 i n) que són *coprimers* amb n. Observem que $\varphi(1) = 1$. És a dir

$$\varphi(n) := \big|\{k : 1 \leq k \leq n,\ \mathrm{mcd}(k, n) = 1\}\big|.$$

2. Definim $\varphi(1) := 1$ i si $n > 1$, definim $\varphi(n)$ com la quantitat d'elements invertibles a l'anell $\mathbb{Z}/(n)$. És a dir,

$$\varphi(n) := \big|\big(\mathbb{Z}/(n)\big)^*\big|.$$

Que les dues definicions són equivalents es desprèn de l'estudi que vam fer sobre els invertibles de l'anell $\mathbb{Z}/(n)$ (pagina 146). Recordem que havíem demostrat que $a \in \mathbb{Z}/(n)$ és invertible si i només si a i n són coprimers. Els elements invertibles d'un anell A es denoten A^* (pàgina 129) i és important tenir en compte aquesta observació senzilla:

> *Si A és un anell, aleshores A^* és un grup amb l'operació de multiplicació.*

Per tant, si $n > 1$, la funció d'Euler $\varphi(n)$ calcula l'ordre (pàgina 97) del grup $(\mathbb{Z}/(n))^*$.

[1] En anglès aquesta funció es coneix com *Euler's totient function*. La paraula *totient* és un neologisme inventat el 1879 pel matemàtic anglès James Joseph Sylvester.

Propietats i càlcul de la funció φ

- Si p és primer, aleshores $\varphi(p) = p - 1$. Això és evident.
- Si p és primer i $r \geq 1$, aleshores

$$\varphi(p^r) = p^r - p^{r-1} = p^r \left(1 - \frac{1}{p}\right).$$

  Això també és evident.

- Si n, m són coprimers, aleshores $\varphi(nm) = \varphi(n)\varphi(m)$. Per demostrar aquesta propietat crucial —que ja no és evident— utilitzarem la definició 2 de la funció d'Euler. Veiem que es tracta de ser capaços de relacionar aquests tres grups

$$\big(\mathbb{Z}/(nm)\big)^*, \quad \big(\mathbb{Z}/(n)\big)^*, \quad \big(\mathbb{Z}/(m)\big)^*.$$

  Comencem observant que aquestes aplicacions

$$\mathbb{Z}/(nm) \longrightarrow \mathbb{Z}/n, \quad \mathbb{Z}/(nm) \longrightarrow \mathbb{Z}/(m)$$

  donades per $a \mapsto a$ estan ben definides i són *homomorfismes d'anells*. Això ens permet definir un homomorfisme d'anells

$$\Phi : \mathbb{Z}/(nm) \longrightarrow \mathbb{Z}/(n) \times \mathbb{Z}/(m)$$

  com $a \mapsto (a, a)$. Vist això, el punt clau és que Φ és una aplicació bijectiva. Per què? **Pel teorema xinès del residu!** Per tant, l'anell de l'esquerra i el de la dreta, com que són isomorfs, tenen el mateix nombre d'elements invertibles. D'altra banda, és senzill veure que $(a, b) \in \mathbb{Z}/(n) \times \mathbb{Z}/(m)$ és invertible si i només si a és invertible a $\mathbb{Z}/(n)$ i b és invertible a $\mathbb{Z}/(m)$. En conclusió, hem demostrat que $\varphi(nm) = \varphi(n)\varphi(m)$.

- **Càlcul de $\varphi(n)$ per qualsevol n.** Com que tot enter es pot descompondre en producte de potències de primers diferents, les dues propietats anteriors de la funció φ ja ens permeten calcular trivialment $\varphi(n)$ quan tenim la descomposició en primers $n = p_1^{r_1} \cdots p_k^{r_k}$, amb $p_1 < p_2 \cdots < p_k$:

$$\varphi(n) = p_1^{r_1} \left(1 - \frac{1}{p_1}\right) \cdots p_k^{r_k} \left(1 - \frac{1}{p_k}\right) = n \prod_{p|n} \left(1 - \frac{1}{p}\right).$$

- **Fórmula de Gauss.**

$$\sum_{d|n} \varphi(d) = n.$$

  La demostració d'aquesta fórmula és curiosa —i, de fet, elemental— i és un bon exemple d'un mètode de demostració d'igualtats d'aquesta mena consistent en *comptar els elements d'un conjunt de dues maneres diferents*. El conjunt que comptarem serà el conjunt $\{1, \ldots, n\}$ que, evidentment, té

n elements. Per comptar els elements d'aquest conjunt el subdividirem en subconjunts A_i, un per a cada divisor de n. Si $d_1, \ldots, d_r$ són els divisors de n, definim

$$A_i := \left\{ k : 1 \leq k \leq n,\ \mathrm{mcd}(k, n) = \frac{n}{d_i} \right\}, \quad i = 1, \ldots, r.$$

Aleshores, si sumem el nombre d'elements de cada subconjunt A_i hem d'obtenir n. Quants elements té A_i? Considerem aquests altres conjunts

$$B_i := \{ m : 1 \leq m \leq d_i,\ \mathrm{mcd}(m, d_i) = 1 \}, \quad i = 1, \ldots, r$$

que, per definició de la funció d'Euler, tenen exactament $\varphi(d_i)$ elements cadascun. Observem ara que

$$k \mapsto \frac{k d_i}{n} \in \mathbb{Z}$$

dóna una bijecció de A_i a B_i, per tot $i = 1, \ldots, r$. Amb tota aquesta informació ja tenim demostrada la fórmula de Gauss.

- **Fórmula de Moebius**.

$$\varphi(n) = n \sum_{d|n} \mu(d) \frac{1}{d}.$$

Hem d'explicar el significat de $\mu(d)$, que es coneix com la **funció μ de Moebius**. És una funció definida sobre els enters positius que només pren tres valors: $0, 1, -1$:

  - $\mu(1) = 0$.
  - Si hi ha un primer p tal que p^2 divideix n, aleshores $\mu(n) = 0$.
  - Si n s'expressa com a producte de k primers diferents, aleshores $\mu(n) = (-1)^k$.

Ara podem veure que la fórmula de Moebius és una conseqüència immediata del càlcul de $\varphi(n)$ a partir de la descomposició en primers $n = p_1^{r_1} \cdots p_k^{r_k}$, amb $p_1 < p_2 \cdots < p_k$:

$$\begin{aligned} \varphi(n) &= n \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_k}\right) \\ &= n \sum_{i_1 < \cdots < i_j} (-1)^j \frac{1}{p_{i_1} \cdots p_{i_j}} = n \sum_{d|n} \mu(d) \frac{1}{d}. \end{aligned}$$

Una excursió: polígons construïbles amb regle i compàs

Si anem a Saldes (el Berguedà) i el dia és clar, seria imperdonable no continuar una mica més enllà i arribar al mirador de la paret nord del Pedraforca —encara

que no estiguem en disposició d'escalar-la, en aquell moment. Igualment, si hem seguit fins aquí aquest camí de la funció d'Euler, tenint en compte que aquest mateix camí passa molt a prop d'un paisatge matemàtic grandiós, seria lamentable no acostar-se fins al mirador i contemplar aquest paisatge —encara que no puguem ni vulguem «escalar-lo» en aquesta ocasió.

El «paisatge» del que parlem és la teoria del dibuix, amb regle i compàs, dels polígons regulars: triangle equilàter, quadrat, pentàgon regular, etc.

És molt senzill inscriure un triangle equilàter o un quadrat a una circumferència. Als *Elements* d'Euclides hi ha les construccions del pentàgon regular i del pentadecàgon regular, però a la matemàtica clàssica no es coneixia cap altra construcció d'un polígon regular amb un nombre senar n de costats, més enllà dels casos $n = 3, 5, 15$. Observem que, com que sempre podem traçar la bisectriu de qualsevol angle, si hem construït un polígon regular de n costats, també podem construir-ne un de $2n$ costats i, en general, un de $2^r n$ costats, per tot r. Recíprocament, si hem construït un polígon regular de $2^r n$ costats, també en tenim un de n costats, clarament. Per tant, el problema rau en *construir amb regle i compàs polígons regulars amb un nombre senar de costats* $n \geq 7$. Com és que no es coneixia cap construcció a banda de les tres que hem esmentat? Són *massa difícils* de trobar? O potser són *impossibles*?

En aquestes preguntes, les paraules que convé no oblidar són *«amb regle i compàs»*. Els polígons regulars amb un nombre qualsevol de costats *existeixen* al pla real, això no ho discutim pas —tot i que aquesta existència no és pas una qüestió trivial!. Que es puguin dibuixar amb uns mitjans concrets com són el regle i el compàs, això és el que estem dilucidant ara.

El famós problema de la construcció amb regle i compàs dels polígons regulars va restar obert fins l'any 1800 quan Gauss va trobar una *condició suficient* —que té a veure amb la funció φ d'Euler— perquè el polígon regular de n costats es pugui construir amb regle i compàs. Més endavant, el 1837 Pierre Wantzel va demostrar que aquella condició també és necessària, de manera que el problema va quedar totalment dilucidat. El teorema és aquest:

> *El polígon regular de n costats és construïble amb regle i compàs si i només si $\varphi(n)$ és una potència de 2.*

En particular, ni el polígon regular de 7 costats, ni el de 9, ni el de 11, ni el de 13 es poden construir, però el de 17 sí que es pot construir perquè $\varphi(17) = 16 = 2^4$. Ara ens hem de preguntar

> *Per a quins valors de n es compleix que $\varphi(n)$ és una potència de dos?*

Escrivim la descomposició de n en factors primers $n = p_1^{r_1} \cdots p_k^{r_k}$ i recordem el càlcul que hem fet de la funció φ:

$$\varphi(n) = p_1^{r_1-1} \cdots p_k^{r_k-1}(p_1 - 1) \cdots (p_k - 1).$$

Quan és aquesta expressió una potència de 2? La resposta no és difícil:

> *$\varphi(n)$ és una potència de 2 si i només si es compleixen:*
> *a)* $n = 2^r(2^{k_1}+1)\cdots(2^{k_m}+1)$.
> *b)* $2^{k_1}+1,\ldots,2^{k_m}+1$ *són primers.*

Quins són els enters n que tenen aquestes dues propietats? Fem una definició:

> **Primers de Fermat**. *Un primer de Fermat és un nombre primer p tal que $p-1$ és una potència de dos. És a dir, són els primers de la forma 2^r+1 per algun r, els primers que estan immediatament després d'una potència de dos.*

Per exemple, 3, 5 i 17 són primers de Fermat. N'hi ha algun més?

> *Només es coneixen 5 primers de Fermat:* 3, 5, 17, 257, 65537. *Es conjectura que no n'hi ha cap més.*

Tornant a la construcció dels polígons regulars, el resultat anterior ens diu que hi ha molts pocs polígons regulars *amb un nombre primer de costats* que sapiguem que es poden construir amb regle i compàs: el triangle equilàter, el pentàgon regular i els polígons regulars de 17, 257 i 65537 costats.[2]

[2] És comprensible que l'estudiant, a la vista del teorema anterior de Gauss i Wantzel, trobi inversemblant que hi hagi una relació profunda entre dos temes tant aparentment diversos com les construccions amb regle i compàs i una funció aritmètica com la φ d'Euler. No podem demostrar aquest teorema —necessitaríem eines que l'estudiant aprendrà més endavant en la seva carrera— però, en canvi, no és gens difícil d'adquirir una certa idea sobre quin és el lligam entre els dos temes. La primera observació que cal fer és pensar quines operacions algebraiques apareixen quan traduïm una construcció amb regle i compàs a càlculs amb coordenades cartesianes. Mentre treballem només amb rectes —unir dos punts amb una recta, trobar el punt de tall de dues rectes— les coordenades cartesianes dels nous punts que anem obtenint es relacionen amb les coordenades dels punts inicials amb les operacions de suma, resta, multiplicació i divisió. En canvi, quan utilitzem circumferències —trobar la intersecció d'una recta i una circumferència, trobar la intersecció de dues circumferències— les coordenades dels nous punts s'obtenen amb les operacions anteriors i, a més, l'operació *arrel quadrada*. La conclusió és que un punt es pot arribar a construir amb regle i compàs si i només si les seves coordenades no contenen cap més operació que suma, resta, multiplicació, divisió i arrel quadrada.

Què podem dir, doncs, del polígon regular de n costats? Per construir-lo hem de dibuixar un angle de $2\pi/n$ radiants o, equivalentment, hem de construir el punt de coordenades $(\cos 2\pi/n, \sin 2\pi/n)$. Això redueix el problema a aquesta pregunta: és $\cos 2\pi/n$ expressable només amb operacions racionals i arrels quadrades? Tots sabem que la resposta és sí per als angles de 30, 45, 60 graus perquè el seu cosinus val $\sqrt{3}/2$, $\sqrt{2}/2$, $1/2$, respectivament. És menys conegut que el cosinus de $2\pi/5$ (l'angle del pentàgon regular) val $(\sqrt{5}-1)/4$ o que

$$\cos\left(\frac{2\pi}{15}\right) = \frac{\sqrt{30-6\sqrt{5}}+\sqrt{5}+1}{8}$$

i aquest és el motiu que explica per què el pentàgon i el pentadecàgon regulars es poden construir

amb regle i compàs. En definitiva, la feina que van fer Gauss i Wantzel va consistir en relacionar $\varphi(n)$ amb el fet que $\cos 2\pi/n$ es pugui expressar amb les operacions de suma, resta, multiplicació, divisió i arrel quadrada.

Com que 17, 257 i 65537 són primers de Fermat, sabem que $\cos(2\pi/17)$, $\cos(2\pi/257)$ i $\cos(2\pi/65537)$ s'han de poder expressar utilitzant les operacions de l'aritmètica més l'arrel quadrada. L'expressió de $\cos(2\pi/17)$ —i, per tant, la possibilitat de construir amb regle i compàs el polígon regular de 17 costats— va ser trobada per Gauss quan tenia divuit anys. És aquesta:

$$16\cos\left(\frac{2\pi}{17}\right) = \sqrt{34-2\sqrt{17}}+\sqrt{17}-1+2\sqrt{17+3\sqrt{17}-\sqrt{34-2\sqrt{17}}-2\sqrt{34+2\sqrt{17}}}.$$

A la mateixa breu nota on anunciava el seu descobriment —presentada a una important revista literària per un dels seus professors al *Collegium Carolinum*— Gauss revela que ja tenia en ment la solució del problema general: *«aquest descobriment és, de fet, només un corol·lari d'una teoria d'un abast més gran, encara no completada»*..

23 El teorema de Fermat-Euler

En una carta que va escriure a un amic seu el 1640, Pierre de Fermat va enunciar —sense demostració, com era habitual en ell— aquesta propietat dels nombres primers:

(Fermat) *p primer* $\Rightarrow\ a^{p-1} \equiv 1 \mod p$ *per tot a coprimer amb p.*

L'any 1736 Euler va demostrar aquesta afirmació de Fermat —que ara es coneix amb el nom de *petit teorema de Fermat*— i, més endavant, va generalitzar aquest resultat utilitzant la seva funció φ que hem estudiat en el tema anterior.

(Euler) $a^{\varphi(n)} \equiv 1 \mod n$ *per tot a coprimer amb n.*

Aquest teorema d'Euler és un resultat de gran importància a la teoria de nombres i, com veurem en el capítol següent, és una peça clau de la seguretat digital actual.

Abans de parlar de la demostració d'aquest teorema d'Euler fem una observació important: el petit teorema de Fermat és una *condició necessària* perquè p sigui primer. Resulta que comprovar que un nombre p és primer és extraordinàriament laboriós. En canvi, comprovar si p satisfà la condició $a^{p-1} \equiv 1 \mod p$ per algun valor de a (coprimer amb p) és molt més ràpid de fer. Aleshores, si trobem un valor de a tal que p no compleix la congruència de Fermat, ja podem descartar que p sigui primer. El petit teorema de Fermat ens dóna, doncs, un **test de primalitat** per a cada valor de a: si p no passa algun d'aquests testos, podem afirmar que no és primer. Si els passa, pot ser primer o pot no ser-ho.

Direm que p és un **pseudoprimer en base** a si $a^{p-1} \equiv 1 \mod p$. Un pseudoprimer pot no ser primer. Per exemple, 341 és un pseudoprimer en base 2, però $341 = 11 \times 31$.

També hi ha enters p que passen el test *per tot a* (coprimer amb p). Es diu que són **pseudoprimers forts**. Curiosament, passar tots els testos no garanteix primalitat: 561 és un pseudoprimer fort que no és primer.

Demostració del teorema de Fermat-Euler

La demostració del teorema de Fermat-Euler és extraordinàriament senzilla *quan es situa en un context apropiat*. Per aquest motiu, té un alt valor didàctic i és pertinent discutir-la en aquest text de fonaments.

Resulta que, encara que no ho sembli, el teorema

$$a^{\varphi(n)} \equiv 1 \quad \text{mòd } n$$

no té res a veure ni amb les congruències, ni amb els nombres enters, ni amb l'aritmètica. A la demostració no utilitzarem res de tot això i obtindrem un magnífic exemple de la importància de ser capaços de situar una propietat matemàtica en el seu context més general possible per aconseguir, d'aquesta manera, despullar-la de tot allò que no juga cap paper en la demostració i poder anar a l'arrel d'allò que volem demostrar. El premi, molt sovint, és que quan plantegem la propietat en el seu context més essencial, la demostració pot esdevenir evident.

Fem aquestes observacions sobre el teorema de Fermat-Euler:

- El teorema parla de congruències mòdul n. Per tant, el seu àmbit natural és l'anell $\mathbb{Z}/(n)$, en el que la congruència és *igualtat*.
- L'element a que surt a l'enunciat és coprimer amb n. Per tant, és una unitat a l'anell $\mathbb{Z}/(n)$.
- Les unitats a $\mathbb{Z}/(n)$ formen un grup que hem denotat $(\mathbb{Z}/(n))^*$. Per tant, $a \in (\mathbb{Z}/(n))^*$.
- Sabem que $(\mathbb{Z}/(n))^*$ té exactament $\varphi(n)$ elements.

Intentem abstreure les idees principals de tot això.

- Tenim un grup multiplicatiu G. En el nostre cas, és el grup $(\mathbb{Z}/(n))^*$.
- El grup G té m elements. En el nostre cas, $m = \varphi(n)$.
- Tenim un element $a \in G$.
- Volem demostrar que al grup G es compleix $a^m = 1$.

És a dir, *podria ser* que aquesta gran generalització del teorema de Fermat-Euler —que no parla ni de congruències ni de nombres coprimers ni d'aritmètica— fos certa:

Si G és un grup (multiplicatiu) d'ordre m i $a \in G$, aleshores $a^m = 1$.

És clar que aquest teorema general podria ser fals i que, nogensmenys, el teorema de Fermat-Euler fos cert. Però si el teorema general és cert, aleshores el teorema de Fermat-Euler ja està demostrat perquè és un cas particular del teorema general. La situació és que el teorema anterior és cert i la seva demostració és força senzilla.

Demostració: En primer lloc, com que G és un grup finit, existirà $r > 0$ mínim tal que $a^r = 1$: és el que n'havíem dit l'*ordre* de l'element a al grup G i té la propietat que

$$H := \{1, a, a^2, \ldots, a^{r-1}\}$$

és un subgrups i té exactament r elements. Apliquem ara el teorema de Lagrange (pàgina 108): r ha de ser un divisor de m i tindrem $m = rs$ per algun enter s. Aleshores:

$$a^m = a^{rk} = (a^r)^k = 1^k = 1$$

i el teorema està demostrat.

24 | Criptografia de clau pública

Ja hem anat dient que els teoremes clàssics d'aritmètica que hem estudiat als capítols anteriors, després de ser considerats, durant segles o, fins i tot, millenis, com el paradigma de la matemàtica «pura» —la que tanta i tanta gent (incloent-hi matemàtics eminents[1]) afirmava fins fa ben poc que *no servia per a res*— amb la revolució digital han quedat situats al centre mateix de les tecnologies imprescindibles per al món digital que ara ens envolta —seguretat digital, criptomonedes, etc.

En aquest capítol volem explicar les bases matemàtiques del primer sistema criptogràfic *de clau pública* que van publicitar Ron Rivest, Adi Shamir i Leonard Adleman el 1977 i es coneix amb l'acrònim **RSA**. Abans d'estudiar com funciona el sistema RSA hem de repassar quin era la idea bàsica de l'encriptació abans de 1977.

Si A vol enviar un missatge m a B de manera que cap altre agent C pugui llegir-lo, cal que A i B es posin d'acord en una *clau secreta compartida*, és a dir una funció f i la seva inversa f^{-1} que només A i B coneixen. Aleshores, A envia a B el missatge encriptat $x := f(m)$ i B el desencripta en la forma $m = f^{-1}(x)$. El problema evident que té aquest sistema és que C en té prou amb interceptar la clau secreta compartida (f, f^{-1}) en el moment que A i B se la intercanviïn.

Més enllà que aquest sistema que hem descrit és completament vulnerable, és inimaginable com podria funcionar de manera segura en les circumstàncies actuals on una immensitat d'agents s'han d'intercanviar constantment una immensitat de missatges segurs per canals vulnerables.

El sistemes de clau pública funcionen igual que el sistema anterior *però amb una diferència essencial*. En aquests sistemes B fa pública la seva clau f. D'aquesta manera, qualsevol agent pot enviar un missatge encriptat a B enviant $f(m)$. Però B manté secreta —només B la coneix— la funció de desencriptació f^{-1}. És evident que aquest mètode només pot funcionar si **el coneixement de f no permet deduir el valor de la funció f^{-1}**. Això sembla absurd perquè la inversa d'una funció bijectiva està unívocament determinada, però del que es tracta és de trobar una funció f fàcil de calcular però amb la propietat que f^{-1} sigui

[1] *«I have never done anything 'useful'. No discovery of mine has made, or is likely to make, directly or indirectly, for good or ill, the least difference to the amenity of the world.»* G.H. Hardy, *A Mathematicians Apology*.

extraordinàriament difícil de calcular. I aquí és on intervé l'aritmètica i la funció φ d'Euler.

Els fonaments teòrics del sistema RSA

Suposem, doncs, que B vol establir un sistema criptogràfic de clau pública, de manera que qualsevol agent li pugui enviar missatges encriptats que només B, aplicant la seva clau privada, pugui llegir. Aquesta situació es produeix cada moment que, per exemple, entrem en una url que utilitza el protocol segur HTTPS. Imaginem que nosaltres som B. Aquests són els passos que hem de seguir:

1. Escollim dos nombres primers grans p, q.
2. Multipliquem p i q i el resultat $n := pq$ l'anomenem *el mòdul*.
3. Calculem $k := (p-1)(q-1) = \varphi(n)$.
4. Escollim un nombre natural e qualsevol que sigui coprimer amb k. En diem *l'exponent*.
5. Calculem la identitat de Bézout entre k i e i trobem d tal que $ed \equiv 1$ mòd k. Diem que d és la nostra *clau privada*.
6. Fem públics el *mòdul* n i *l'exponent* e.
7. Guardem en secret al nostre ordinador p, q, k i la *clau privada* d.

Havent fet això, la manera de procedir és aquesta:

- Qualsevol agent que ens vulgui enviar un missatge secret m ($0 < m < n$) utilitza la nostra clau pública n i el nostre exponent e (també públic) per calcular
$$X := m^e \mod n.$$
Aleshores, l'agent ens envia el missatge encriptat X.
- Quan nosaltres rebem un missatge encriptat X, utilitzem la nostra clau privada d per calcular $Y := X^d \mod n$.
- Finalment, el teorema de Fermat-Euler ens diu que
$$Y \equiv X^d \equiv m^{ed} \equiv m^{1+\lambda k} \equiv m\, m^{\lambda\varphi(n)} \equiv m \mod n$$
i, d'aquesta manera, el missatge que hem obtingut desencriptant el missatge que hem rebut és el mateix missatge que ens volien enviar de manera segura.

Què hauria de fer un hacker per llegir el missatge m? Observem que la informació que pot tenir el hacker consisteix en n, e i X. Per recuperar m necessita conèixer d i per conèixer d ha de conèixer $k = \varphi(n)$. En té prou amb saber calcular la funció d'Euler d'un nombre n. On és el problema, doncs? El problema del hacker és que —fins on sabem ara (abril de 2024)— no hi ha cap mètode prou ràpid per calcular la funció φ ni per descompondre n en factors primers.

És a dir, la seguretat del sistema RSA es fonamenta en que quan multipliquem dos primers p, q per obtenir un nombre $n = pq$, no es coneix cap mètode ràpid per *desfer* aquesta multiplicació i recuperar p, q a partir de n.

Alguns exemples

Tots els ordinadors tenen la capacitat de treballar amb encriptació RSA i, per tant, de generar les claus pública i privada necessàries. Si demanem a un ordinador que generi aquestes claus, obtindrem alguna cosa com aquesta:

```
RSA Private-Key (2048 bit, 2 primes)
modulus
    22811093846362383056605010433288438613172319843087653670721435098666
    34277486834786592881276316796266554028427897886380276718538401102555
    94975183633565671839330773492965538500160756991769216925374033398644
    84833604623628942374332439816731196701794182953856428972015989710344
    43226568268038287456495815531269529853880847326031630341695588321455
    97902283112174872331310120638817309645815192271229209605924317201766
    45873733844990692899443031273306842959696437392026905097028030219066
    79025666587855019253931111541616214410893356012557249483800595777177
    33135054568533446217389968615714006698146621906478458304242436208677
    84137468980121
publicExponent 65537 (0x10001)
privateExponent
    18266069456387254534500730007912232326482295085911728901014544342015
    68458248494785626559294746923894479284341176025441362007906629102055
    25790052365929061331099068650340090253596233672140267864362175545533
    41831068511580073345783177581247790892373116365342196561066363007170
    23926744833275573911369820151448703132606839790356456201683234102390
    07049226117102338625754162752187335614631413725655490533840568436591
    88448276934113863334305729229438043977926836950606943080857868870075
    69511121358749575790437871200563381055982085212419152239279555768111
    33038027890563354737817516613414393609121184627966839836761037931340
    01030584366557
prime1
    15233382076298582734651154849939420250899333046082240882703274478915
    53199540412757043917235362010139801846310926622234596833378675589985
    27067029898138030261692438663107337562635375561490575237696403048745
    66472685348792278305769271116983346426091045520489641573022852121175
    36902171757863261511366192374105987171787
prime2
    14974411940900413459763021012539452887066122823099208015225271772475
    35501437798300291968149039755383034964989713849753746345458358718985
    88481030135918239246462390550733200488312100702331601970737783772455
    75393893030733885437318533421351631276193985531741947761581775321255
    20704502725307333684934592510922478061483
```

Veiem que el mòdul (la clau pública) n és un enter de 2048 bits, que vol dir un nombre de 617 xifres (en base 10), obtingut multiplicant dos primers p_1, p_2, cadascun dels quals és un enter de 1024 bits. També veiem que l'exponent (públic)

e és el primer 65537 —el més gran dels primers de Fermat coneguts (coincidència?). Finalment, tenim la clau privada d que compleix $ed \equiv 1 \mod k$ amb $k = (p_1 - 1)(p_2 - 1)$.

Si ara, per exemple, entrem a la pàgina `uab.cat`, veurem que és una pàgina segura amb protocol HTTPS. Evidentment, li podem demanar quina és la seva clau pública —perquè és pública!— i si ho fem obtenim aquesta resposta:

```
Algorisme: RSA
Mida de la clau: 2048
Exponent: 65537
Mòdul:
   D4:D0:93:61:42:C0:EA:B5:0B:DE:37:65:CF:A2:27:37:38:66:30:E8:00:B7:B8:
   3A:75:03:CB:83:04:D5:6A:4A:D1:42:64:10:C6:C2:68:09:51:EE:4E:E5:2A:8D:
   B2:31:15:42:13:4B:03:FB:A0:D8:FF:CC:33:A0:82:83:C1:F7:31:10:B4:A7:8C:
   F3:AE:6A:91:51:4E:A3:0A:F9:26:E4:E7:9A:F4:6F:64:1B:62:33:93:B5:7E:C2:
   4F:37:67:66:5A:05:81:79:19:38:59:27:ED:9A:DE:A2:62:9C:1E:71:AB:E2:EC:
   B3:1C:49:9F:8F:09:85:9E:2C:2A:F7:0D:0A:DC:C6:F0:ED:CE:AB:8F:0A:F1:49:
   13:98:18:F3:E7:AE:C5:49:4C:8C:68:5F:69:18:AC:5A:9F:3A:5E:DA:34:60:5B:
   E2:FA:69:B2:91:24:DD:FD:E8:B8:F3:E7:08:69:F1:83:CB:A5:E3:EE:3F:04:97:
   9D:03:EC:EA:89:38:85:9C:21:FA:D3:EF:C5:4C:61:CF:41:90:F9:A7:F2:AE:28:
   AA:7D:BE:E5:0E:F0:30:86:78:2B:0B:9D:F3:52:99:A0:74:12:40:60:BA:A7:3B:
   E4:B9:ED:1F:1C:48:3E:60:A0:BF:00:55:A6:C1:66:D5:BD:61:DC:C2:59:5F:93:
   0F:73:A1
```

Observem que l'exponent públic torna a ser el famós 65537 —no pot tornar a ser una coincidència!—, la mida és el valor típic 2048 —com més gran és la mida, més segura és la clau— i el mòdul o clau pública ara ens ve donada en notació hexadecimal. Si aconseguíssim descompondre aquest mòdul com a producte dels dos primers (secrets) que el divideixen, hauríem trencat la clau RSA de la UAB i podríem desxifrar els missatges que s'envien a la seva pàgina web, per exemple, les contrasenyes dels usuaris. El temps necessari perquè un ordinador (no quàntic) pugui trobar aquests dos primers és de l'ordre de 10^{14} anys —si ningú troba alguna manera més ràpida de fer-ho.

Algunes consideracions tècniques

Hem vist que la teoria que hi ha al darrere de RSA és ben senzilla i forma part del coneixement matemàtic clàssic des de fa prop de tres-cents anys. Però més enllà del fonament teòric, hi ha una sèrie de consideracions pràctiques que també juguen un paper important. Repassem-ne algunes:

- **D'on surten els dos primers?**

  La tecnologia digital necessita grans quantitats de nombres primers molt grans, constantment, fins el punt que els nombres primers són un producte industrial de consum massiu, de primera necessitat. Existeixen, aquests primers? N'hi ha prou per cobrir les necessitats?

  En aquest punt entra en joc el **teorema dels nombres primers** que vam discutir al capítol 19, el gran teorema que van intuir Gauss i Legendre i que ens diu quina és la densitat dels nombres primers en el conjunt de tots

els nombres naturals, és a dir, quina és, aproximadament, la probabilitat de trobar un nombre primer de, diguem, 1024 bits: serà com buscar una agulla en un paller?

Recordem que el teorema dels nombres primers ens diu que, si x és gran, hi ha aproximadament $x/\log(x)$ primers $\leq x$. En particular, hi ha aproximadament 2.53×10^{305} primers $\leq 2^{1024}$ i 1.27×10^{305} primers $\leq 2^{1023}$. Per tant, la probabilitat que un nombre de 1024 bits sigui primer és aproximadament del 0.14%. Sembla poc, però pensem que si eliminem els parells, els múltiples de 3, els múltiples de 5, etc. aquesta probabilitat ja comença a ser suficientment gran per no haver de patir perquè se'ns acabin els primers o no siguem capaços de trobar-los.

En resum, el que es fa per trobar primers de 1024 bits és:

1. Prenem un nombre q de 1024 bits a l'atzar.
2. Mirem si q és divisible per algun dels 100 nombres primers més petits $2, 3, 5, \ldots, 89, 97$. Si ho és, el descartem i tornem a començar.
3. Apliquem a q diversos tests de primalitat, per exemple el test basat en el petit teorema de Fermat. Si no passa algun d'aquests tests, descartem q i tornem a començar.
4. Al final, obtenim un pseudoprimer q que *«amb probabilitat molt alta»* és primer.
5. Observem que no podem tenir la seguretat absoluta que el pseudoprimer q que hem obtingut sigui primer. Determinar-ho amb seguretat seria del mateix ordre de dificultat que voler trencar una clau RSA de 1024 bits.

- **Exponencial modular**

  Recordem que per encriptar i desencriptar pel mètode RSA cal calcular potències de nombres molt grans amb exponents molt grans, sempre mòdul n. I aquestes operacions s'han de poder fer molt de pressa. Clarament, és impensable calcular primer les potències a $\mathbb{Z}$ i després reduir el resultat mòdul n. Ha calgut, doncs, desenvolupar tècniques molt eficients per calcular potències a $\mathbb{Z}/(n)$.

  Hem trobat dues vegades que l'exponent públic de la clau RSA és precisament el primer de Fermat 65537. Gairebé sempre s'utilitza aquest exponent i els motius són fàcils d'entendre. En primer lloc, és bo que e sigui un nombre primer perquè recordem que e ha de ser coprimer amb $(p_1 - 1)(p_2 - 1)$. D'altra banda,

  $$65537 = 2^{2^4} + 1$$

  i és molt ràpid calcular m^e mòd n elevant repetidament al quadrat.

Algunes idees (molt bàsiques) sobre el problema $P = NP$

Hem vist que la seguretat de l'encriptació del tipus RSA es basa en una funció f que és molt ràpida de calcular però que, en canvi, tots els mètodes que es coneixen per calcular f^{-1} són extraordinàriament lents. En el cas concret que hem estudiat, el càlcul «impossible» és descompondre un nombre de 617 xifres en producte de dos primers.

Comparem aquests dos problemes:

- Decidiu si 7487 divideix 25702871.
- Trobeu un divisor de 25702871.

El primer es resol en molt poc temps, mentre que el segon requereix molt més temps: és ben habitual que sigui molt més senzill *comprovar* la solució d'un problema que *trobar* la solució d'aquest mateix problema. La pregunta és: podem donar una justificació teòrica d'aquesta comprovació empírica? Es considera que respondre aquesta pregunta és un dels problemes oberts més importants de les matemàtiques del segle XXI, un problema la solució del qual sembla molt llunyana. Se'n diu el problema $P = NP$. Expliquem-ho amb una mica més de detall.

Es diu que un problema (de càlcul amb nombres naturals) pertany a la classe NP quan el temps necessari per **comprovar** que un nombre és solució del problema creix segons una **funció polinòmica** de la mida del nombre. Per exemple, trobar un divisor de n és clarament un problema de classe NP: quan tenim un candidat a divisor de n, comprovar si ho és o no es pot fer amb una simple divisió, i dividir dos nombres de k xifres requereix, en el pitjor dels casos, de l'ordre de k^2 operacions elementals.

En canvi, direm que un problema (de càlcul amb nombres naturals) pertany a la classe P quan el temps necessari per **trobar** una solució del problema creix segons una **funció polinòmica** de la mida del nombre. Evidentment, tot problema de classe P és també de classe NP, però no se sap si hi ha algun problema de classe NP —per exemple, factoritzar un nombre natural— que **no** sigui de classe P.

Demostrar $P = NP$ implicaria, conceptualment, que no hi pot haver cap sistema criptogràfic essencialment segur. Demostrar que $P \neq NP$ voldria dir que, en principi, poden existir sistemes criptogràfics que són essencialment segurs.

Exercicis d'aritmètica

1. Demostreu les propietats següents, vàlides a tot anell A:
   (a) L'element neutre respecte del producte és únic.
   (b) $0a = 0$ per a tot $a \in A$.
   (c) $(-1)\,a = -a$ per a tot $a \in A$.
2. A partir de la relació d'ordre ordinària a $\mathbb{Z}$, definiu una relació d'ordre total a $\mathbb{Q}$ que compleixi aquestes propietats:[2]
   (a) Si $a < b$, aleshores $a + c < b + c$ per tot $c \in \mathbb{Q}$.
   (b) Si $0 < a$ i $0 < b$, aleshores $0 < ab$.
3. Sigui K un cos ordenat (és a dir, un cos amb una relació d'ordre total que compleixi les propietats (a) i (b) de l'exercici anterior). Demostreu:
   (a) $1 > 0$.
   (b) $a > 0 \Leftrightarrow -a < 0$.
   (c) $a > 0$, $b < 0 \Rightarrow ab < 0$.
   (d) $a < 0$, $b < 0 \Rightarrow ab > 0$.
4. Fixem un primer $p \in \mathbb{Z}$. Si $n \in \mathbb{Z}$, $n \neq 0$, definim la *valoració p-àdica* de n com el màxim r tal que p^r divideix n. Escrivim $v_p(n) := r$. Definim $v_p(0) := \infty$. Demostreu (sense utilitzar el teorema de factorització única) aquestes propietats de la valoració p-àdica:
   (a) $v_p(mn) = v_p(n) + v_p(m)$ per tot $n, m \in \mathbb{Z}$.
   (b) $v_p(n + m) \geq \min\{v_p(n), v_p(m)\}$ per tot $n, m \in \mathbb{Z}$.
   (c) Si $v_p(n) \neq v_p(m)$, aleshores $v_p(n + m) = \min\{v_p(n), v_p(m)\}$.
5. (a) A la pàgina 131 es demostra l'existència de divisió amb residu en el cas que dividend i divisor siguin positius. Completeu la demostració en els altres casos.
   (b) Trobeu el quocient i el residu en aquestes divisions enteres:
   (a) 19 entre 7 (b) -111 entre 11 (c) 0 entre 19
   (d) -1 entre 3 (e) -107 entre 101 (f) -23 entre -17.

[2] Un cos amb una relació d'ordre total que compleix aquestes dues propietats es diu que és un *cos ordenat*.

(c) Escriviu la identitat de Bézout en els casos anteriors.

(d) Escriviu un programa per fer divisions enteres i calcular la identitat de Bézout.

6. Sigui p un nombre primer. Definim $\mathbb{Q}_{(p)}$ com el subconjunt de $\mathbb{Q}$ format per tots els nombres racionals que es poden expressar per una fracció amb denominador no divisible per p. Demostreu que $\mathbb{Q}_{(p)}$ és un anell amb la suma i el producte de $\mathbb{Q}$. Determineu quines són les unitats de $\mathbb{Q}_{(p)}$ i quins són, llevat de producte per unitats, els elements primers de $\mathbb{Q}_{(p)}$. Què podeu dir de l'anell $\mathbb{Q}_{(p)}/(p)$?

7. Suposem que tenim una *cadena d'ideals* de l'anell $\mathbb{Z}$

$$I_0 \subseteq I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots$$

Demostreu que és *estacionària*, és a dir, a partir d'un cert n, tots els ideals I_i amb $i \geq n$ són iguals. Podem afirmar que tota cadena d'ideals

$$I_0 \supseteq I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$$

és estacionària?

8. Proveu que no hi ha cap parella d'enters a, b tals que $\mathrm{mcd}(a, b) = 7$ i $a + b = 100$. En canvi, hi ha infinites parelles de nombres que compleixen $\mathrm{mcd}(a, b) = 5$ i $a + b = 100$.

9. (a) Demostreu que la fracció $\dfrac{4n+5}{2n+3}$ és irreductible per tot $n \in \mathbb{N}$.

(b) Calculeu $\mathrm{mcd}(28n + 5, 35n + 2)$ per a tot $n \geq 1$.

10. Calculeu, per tot nombre natural n, el màxim comú divisor de $6n + 12$ i $4n + 7$.

11. Per simplificar la notació, en aquest exercici escrivim $(a, b) := \mathrm{mcd}(a, b)$ i suposem també que tots els nombres enters que hi apareixen són positius. Demostreu, sense utilitzar la descomposició en primers, aquestes propietats del mcd:

(a) (Propietat associativa) $(a, (b, c)) = ((a, b), c)$. Això ens permet considerar, sense ambigüitat, el mcd de més de dos enters, com (a, b, c, d).

(b) (Propietat distributiva) $c(a, b) = (ca, cb)$.

(c) $(a, b)\,(a^2, b^2) = (a^3, ab^2, ba^2, b^3) = (a, b)^3$.

(d) $(a^2, b^2) = (a, b)^2$.

12. Sigui $C = \{1, x, x^2, \ldots, x^{n-1}\}$ un grup cíclic d'ordre n (vegeu l'exercici **III**.22). Demostreu que $x^i \in C$ té ordre n si i només si r, n són coprimers.

13. En un domini A sigui $p \in A$, $p \neq 0$, que tingui la propietat que si $p|ab$, aleshores $p|a$ o $p|b$. Demostreu que p no és producte de dues no-unitats.[3]

14. Sigui p un nombre primer. Demostreu que $(x + y)^p \equiv x^p + y^p$ (mòd p).

15. Sigui p un primer. Trobeu explícitament les solucions a $\mathbb{Z}/(p^2)$ de l'equació $px - p = x + 1$.

[3] En el llenguatge de la nota de la pàgina 129, en aquest exercici es demana demostrar que, en un domini, tot element primer és irreductible. El recíproc no és cert.

16. El conjunt $\{a + bi \in \mathbb{C} : a, b \in \mathbb{Z}\}$ té una estructura d'anell amb la suma i la multiplicació de $\mathbb{C}$ (recordem que $i^2 = -1$).[4] S'anomena *l'anell dels enters de Gauss* i es denota $\mathbb{Z}[i]$. Definim $N(a + bi) = a^2 + b^2$. Demostreu:

    (a) $N(uv) = N(u)N(v)$ per tot $u, v \in \mathbb{Z}[i]$.

    (b) Demostreu que $\mathbb{Z}[i]$ és un domini i que les seves unitats són ± 1, $\pm i$.

    (c) Demostreu que $2 \in \mathbb{Z}[i]$ no és primer, però $3 \in \mathbb{Z}[i]$ sí que ho és.[5]

17. La successió de Fibonacci $0, 1, 1, 2, 3, 5, \ldots, F_n, \ldots$ es defineix recursivament per $F_0 = 0$, $F_1 = 1$, $F_n + F_{n+1} = F_{n+2}$. Demostreu que dos termes consecutius de la successió de Fibonacci són sempre primers entre ells.

18. Sigui a un enter congruent amb 3 mòdul 4. Demostreu que a no és suma de dos quadrats.

19. Decidiu (sense fer la divisió) si 2590245817021027 és divisible per 11. És a dir, enuncieu i demostreu un *criteri de divisibilitat* per 11, en base 10.

20. Els matemàtics egipcis a partir del segle XIX aC s'esforçaven en escriure els nombres racionals (positius) com a suma de fraccions diferents de numerador 1 —el que avui es coneix com *fraccions egípcies*. Fibonacci va descobrir un mètode per expressar qualsevol fracció en forma de fracció egípcia, basat en aquesta identitat

$$\frac{x}{y} = \frac{1}{q} + \frac{r}{yq}$$

    on q, r són, respectivament, el quocient i el residu de la divisió euclidiana *per excés* (pàgina 131) de y per x. Utilitzant aquesta fórmula, demostreu (de manera constructiva) que tot racional positiu es pot expressar com a suma d'un nombre natural i una fracció egípcia (no oblideu comprovar que els denominadors de les fraccions egípcies són diferents). Escriviu un programa per calcular aquesta expressió per a qualsevol fracció. Apliqueu-ho a resoldre el problema 31 del papir Rhind: *Quin és el nombre que sumat amb els seus dos terços, la seva meitat i la seva setena part dóna 33?*

21. Utilitzeu congruències per trobar una relació entre un nombre n i el nombre que s'obté sumant les xifres de n.[6] Apliqueu-ho a aquest problema: Sigui $x = 2095^{1777}$; sigui a la suma de les seves xifres; sigui b la suma de les xifres de a; calculeu (sense utilitzar cap calculadora ni ordinador) la suma de les xifres de b. (Indicació: calculeu a mà 2.1^3.)

22. Demaneu a una persona que multipliqui el dia del mes en que va néixer per 12, que multipliqui el número del mes per 31 i que us digui només el resultat de sumar aquests dos valors. Determineu quan l'heu de felicitar pel seu aniversari.

---

[4] Aquest exercici utilitza alguns coneixements sobre el cos dels nombres complexos que s'estudiaran a la part **VI** d'aquest text.

[5] En aquest anell, igual com passa a $\mathbb{Z}$, els dos conceptes de *primer* i *irreductible* coincideixen. Vegeu la nota de la pàgina 129.

[6] La resposta a aquesta pregunta ens dóna la justificació teòrica de l'antiquíssima *prova del nou* per comprovar els resultats de les operacions aritmètiques.

23. Els cometes 2P/Encke, 4P/Faye i 8P/Tuttle tenen períodes de 3, 8 i 13 anys, respectivament i van passar pels seus perihelis el 2017, 2014 i 2008, respectivament.[7] Apliqueu el teorema xinès del residu per saber quin any passaran tots tres pels seus perihelis?

24. Suposem que volem calcular la reducció de a mòdul mn, amb n i m coprimers. Utilitzeu el teorema xinès del residu per reduir el problema a calcular a mòdul m i a mòdul n. Apliqueu-ho a determinar les dues últimes xifres de 49^{127}.

25. Resoleu les equacions diofàntiques següents:

$$(a)\ 45x + 21y = 3, \qquad (b)\ 9x + 12y = 2, \qquad (c)\ 7x - 5y = 1,$$
$$(d)\ -16x + 12y = 20, \quad (e)\ 111x + 36y = 15, \quad (f)\ 10x + 26y = 1224.$$

26. Suposem que volem resoldre l'equació diofàntica en tres variables $ax + by + cz = t$. Procedim d'aquesta manera:

    (a) Resolem l'equació diofàntica en dues variables $\mathrm{mcd}(a, b)\, w + cz = t$.

    (b) Resolem l'equació diofàntica en dues variables $ax + by = \mathrm{mcd}(a, b)\, w$.

    Demostreu que aquest mètode ens dóna exactament totes les solucions de l'equació inicial. Apliqueu el mètode a l'equació $6x + 10y + 15z = 7$ i trobeu totes les seves solucions enteres.

27. Sigui $a > 0$ un enter i sigui $b = a + 2$.

    (a) Escriviu la identitat de Bézout entre a i b.

    (b) Decidiu en quin casos l'equació diofàntica $ax + by = 5$ té solució i, en els casos en què en té, trobeu-les totes.

28. Utilitzeu congruències per demostrar que l'equació $3x^2 + 5y^3 = 11$ no té cap solució en els nombres enters.

29. Si A és un anell, podem definir la característica de A de la mateixa manera que ho hem fet en el cas que A sigui un cos (pagina 147). Demostreu que si A és un domini, la característica de A és 0 o un nombre primer.

30. Un ideal d'un anell A diem que és *propi* si és diferent de $\{0\}$ i de A.

    (a) Demostreu que A és un cos si i només si no té cap ideal propi.

    (b) Mostreu un ideal propi a l'anell $\mathbb{Z}/(15)$.

    (c) Demostreu que si $\phi : A \to B$ és un homomorfisme d'anells, aleshores $\ker \phi$ és un ideal de A.

    (d) Demostreu que si $\phi : A \to B$ és un homomorfisme d'anells i A és un cos, aleshores ϕ és injectiu.

31. Considereu l'anell $A = \mathbb{Z}/(6)$.

    (a) Escriviu tots els divisors de zero de A, si és que n'hi ha algun.

[7] Aquestes dades són una gran simplificació. Els períodes reals són 3.30, 7.48 i 13.6 anys, respectivament, i varien amb el temps.

(b) Escriviu tots els elements invertibles de A i trobeu els seus inversos.

(c) Mostreu un ideal de A diferent de $\{0\}$ i de A.

(d) Mostreu un subconjunt no buit de A que contingui 0 i no sigui un ideal.

32. Considereu l'aplicació $f : \mathbb{Z}/(2) \times \mathbb{Z}/(7) \to \mathbb{Z}/(14)$ donada per $f([a],[b]) = [7a + 2b]$. Està ben definida? En cas afirmatiu, es injectiva?

33. Resoleu l'equació de segon grau $x^2 + x + 1 = 0$ als cossos $\mathbb{F}_2$, $\mathbb{F}_3$, $\mathbb{F}_4$, $\mathbb{F}_5$ i $\mathbb{F}_{139}$. (En l'últim cas, utilitzeu la igualtat $54^2 = 21 \times 139 - 3$.)

34. Resoleu aquest sistema d'equacions lineals al cos $\mathbb{F}_5$

$$\left.\begin{aligned} 2x - y + z &= 1 \\ x - 3y + 2z &= -3 \\ 2x + y + z &= 1 \end{aligned}\right\}.$$

35. Sigui p un primer de la forma $p = 4k + 3$ i considereu el cos de p elements, $\mathbb{F}_p$.

(a) Demostreu que si $z \in \mathbb{F}_p$, $z \neq 0$, la condició necessària i suficient perquè z sigui un quadrat és que es compleixi $z^{2k+1} = 1$.

(b) Resoleu l'equació $7x^2 + 2x - 1 = 0$ al cos de 71 elements.

36. Observeu aquesta igualtat, vàlida per a n senar:

$$(x^n + 1) = (x + 1)\,(x^{n-1} - x^{n-2} + \cdots + 1).$$

Recordeu que un *primer de Fermat* és un nombre primer de la forma 2^n+1. Demostreu que tot primer de Fermat és de la forma $2^{2^k} + 1$ per algun enter $k \geq 0$.

37. Un enter a es diu que és un *residu quadràtic mòdul n* si a és un quadrat a l'anell $\mathbb{Z}/(n)$. Demostreu que si $p > 2$ és primer i a és coprimer amb p i és un un residu quadràtic mòdul p, aleshores $a^{(p-1)/2} \equiv 1 \mod p$. (El recíproc també és cert, segons un teorema d'Euler.)

38. Sigui $a = 7^{7583185}$. Calculeu a mà a mòdul 17 i a mòdul 18.

39. Els *nombres de Fermat* són els enters $F_n := 2^{2^n} + 1$ per $n \geq 0$. Demostreu per inducció aquesta fórmula, vàlida per tot $n \geq 1$:

$$F_n = F_0 \cdots F_{n-1} + 2.$$

Deduïu aquest fet (conegut com a *teorema de Goldbach*): dos nombres de Fermat diferents són coprimers.

40. Feu un programa per aplicar el test de primalitat del petit teorema de Fermat amb $a = 5$ als 1000 primers enters senars > 1. Mireu quants nombres no primers passen el test.

41. L'enter

$$\begin{aligned} n = {} & 1340780792994259709957402499820584612751876782678878785 6 \\ & 9358404838219076439420807809743890199269686208790385534 6 \\ & 1676276924914696884102863881983907197952819 \end{aligned}$$

s'ha obtingut multiplicant dos primers diferents $n = pq$. Es tracta de trobar p i q. Comenceu observant que els mètodes elementals de factorització (`factor(n)`, `qsieve(n)`) no són capaços de resoldre el problema. Observeu que

$$x := \frac{p+q}{2},\ y = \frac{p-q}{2} \quad \Longrightarrow \quad n = x^2 - y^2 = (x+y)\,(x-y)$$

i aquesta ha de ser la factorització de n que busquem. Si no hem sigut prou curosos i hem triat els primers p i q massa propers, tindrem que y serà petit i potser podem trobar el valor de y. Pel teorema de les mitjanes aritmètica i geomètrica sabem que $x \geq \sqrt{n}$ i podem començar a buscar x a partir de $\sqrt{n}$. Feu un programa que implementi aquest mètode (conegut com a factorització de Fermat) i trobeu p i q en menys d'un mil·lisegon.

42. Feu un programa per crear unes claus RSA de 2048 bits i un altre programa per encriptar i desencriptar. Comproveu el seu funcionament enviant-vos missatges encriptats entre companys.

43. Recordeu que en la descripció del sistema criptogràfic RSA (pàgina 159) preníem $k := (p-1)(q-1)$. En canvi, en molts casos s'utilitza $k = \text{mcm}(p-1, q-1)$ (el que es coneix com a mètode de Carmichael). Demostreu que amb aquest nou valor de k tot el mètode RSA funciona sense cap variació.

44. Habitualment, quan un usuari U vol entrar a un servei S ho fa utilitzant una contrasenya x que U tramet a S i S valida. El problema d'aquest sistema és que requereix que S conegui i tingui arxivada la contrasenya x de U i podria fer-ne un mal ús o revelar-la, voluntària o involuntàriament, a terceres persones. La mateixa idea de la criptografia RSA que hem estudiat ens dóna un mètode perquè U demostri a S que posseeix la contrasenya necessària per entrar al servei *sense que S conegui la contrasenya*. Es parla d'un mètode de *coneixement-zero*. Funciona d'aquesta manera:

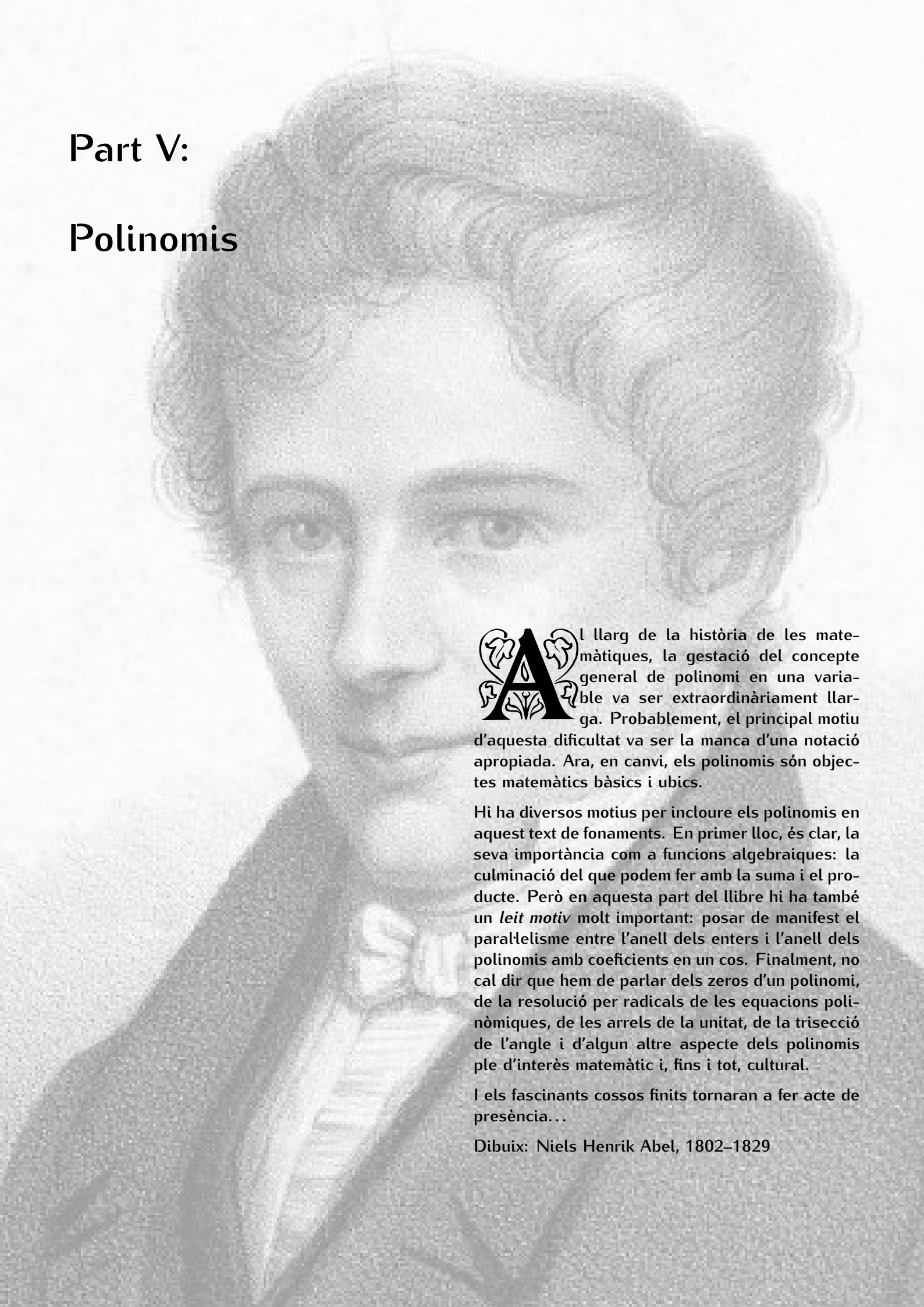
    - Quan U es registra per primera vegada a S:
        - U i S comparteixen un primer gran p i un element $g \in \mathbb{F}_p^*$ d'ordre molt gran.
        - U tria una contrasenya x (un nombre natural) i la manté en secret.
        - U envia a S el valor $y := g^x$.
    - Cada vegada que U es vol registrar a S:
        - U tria un $r \in [0, p-2]$ a l'atzar, calcula $c := g^r \mod p$ i envia c a S.
        - S decideix aleatòriament entre aquestes dues accions:
            (a) Demanar a U el valor de r i comprovar que $c = g^r \mod p$.
            (b) Demanar a U el valor $u := x + r \mod p-1$ i comprovar que $cy = g^u \mod p$.
        - Es repeteix el procés anterior un cert nombre de vegades, i si U supera sempre el test, S considera que, amb una probabilitat ≈ 1, U posseeix realment la paraula de pas x.

    Completeu els detalls d'aquest mètode i observeu que compleix aquestes propietats: (a) S no adquireix cap informació sobre el valor de x; (b) Ningú que no conegui x pot passar aquest test (excepte amb una probabilitat ≈ 0); (c) S no pot demostrar a cap tercera persona que U coneix la paraula de pas x.

pàgina (gairebé) en blanc

Part V:

Polinomis



Al llarg de la història de les matemàtiques, la gestació del concepte general de polinomi en una variable va ser extraordinàriament llarga. Probablement, el principal motiu d'aquesta dificultat va ser la manca d'una notació apropiada. Ara, en canvi, els polinomis són objectes matemàtics bàsics i ubics.

Hi ha diversos motius per incloure els polinomis en aquest text de fonaments. En primer lloc, és clar, la seva importància com a funcions algebraiques: la culminació del que podem fer amb la suma i el producte. Però en aquesta part del llibre hi ha també un *leit motiv* molt important: posar de manifest el paral·lelisme entre l'anell dels enters i l'anell dels polinomis amb coeficients en un cos. Finalment, no cal dir que hem de parlar dels zeros d'un polinomi, de la resolució per radicals de les equacions polinòmiques, de les arrels de la unitat, de la trisecció de l'angle i d'algun altre aspecte dels polinomis ple d'interès matemàtic i, fins i tot, cultural.

I els fascinants cossos finits tornaran a fer acte de presència...

Dibuix: Niels Henrik Abel, 1802–1829

25 | Polinomis: conceptes bàsics

Què és un polinomi

Si A és un anell (commutatiu amb unitat, com sempre durant aquest text, i amb $1 \neq 0$) i x és una *variable* —és a dir, una lletra que no ha estat assignada a cap valor— un **polinomi** en x amb coeficients a A és una expressió de la forma

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

Aquestes *expressions* es poden sumar i multiplicar de la manera òbvia, respectant les propietats d'un anell —associativitat, commutativitat, distributivitat, etc. Aleshores, tots aquests polinomis formen un *anell* que es denota $A[x]$.

Amb tot això que acabem de dir, el concepte de polinomi i les operacions amb polinomis són clars i senzills, i perfectament vàlids per a la pràctica matemàtica quotidiana, però en un text de *fonaments de les matemàtiques* no podem considerar la descripció anterior com una *definició* formalment vàlida. Dir que un polinomi és una *expressió de la forma tal* o que *x és una variable* no ho podem admetre com a definicions en una fonamentació de les matemàtiques basada en la teoria de conjunts.

Donem, doncs, una definició formal —que, de fet, és la que utilitzen, per exemple, els ordinadors— de **polinomi amb coeficients en un anell** A. La idea consisteix en identificar el polinomi $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ amb la successió dels seus coeficients: $a_0, a_1, \ldots, a_n$.

- Un polinomi és una *successió* $(a_0, a_1, \ldots)$ d'elements de A en la qual tots els a_i són zero excepte una quantitat finita.
- Definim la suma de dos polinomis $(a_i) + (b_i)$ com el polinomi (c_i) amb $c_i = a_i + b_i$ per tot $i \leq 0$.
- Definim el producte de dos polinomis $(a_i) \cdot (b_i)$ com el polinomi (c_i) amb

$$c_i := \sum_{j=0}^{i} a_j b_{i-j}.$$

- Definim $x := (0,1,0,0,0,\ldots)$, que és, per tant, un polinomi perfectament ben definit —i no és cap *incògnita* ni cap *variable* ni cap dels altres termes ambigus dels que utilitzem usualment per referir-nos a la x d'un polinomi.

- Amb la definició anterior del polinomi x, és immediat comprovar que tot polinomi $p \neq 0$ s'expressa[1] de manera única com a $p = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, per algun $n \geq 0$, amb $a_n \neq 0$.

- Una comprovació llarga i avorrida demostra que el conjunt de tots els polinomis amb coeficients a A té, amb les operacions anteriors, una estructura d'anell commutatiu amb unitat que es denota $A[x]$. Evidentment, si el polinomi $(0,1,0,0,0,\ldots)$, en lloc de denotar-lo amb la lletra x l'haguéssim denotat amb una altra lletra, per exemple t, escriuríem l'anell dels polinomis com $A[t]$.

- Si A és un cos, l'anell $A[x]$ té també una estructura natural d'**espai vectorial** —una estructura que no forma part del contingut d'aquestes notes però que, molt probablement, el lector ja coneix. Es tracta d'un espai vectorial de *dimensió infinita* amb una base canònica formada pels polinomis $1, x, x^2, x^3, \ldots$

- A l'anell $A[x]$ podem considerar també una operació molt interessant que podem anomenar *substitució*, *composició* o *canvi de variable*. Si tenim dos polinomis $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ i $q(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0$, podem considerar el polinomi $p \circ q$ que obtenim substituint la x del polinomi p per $q(x)$:

$$(p \circ q)(x) = a_n(b_m x^m + \cdots + b_0) + \cdots + a_1(b_m x^m + \cdots + b_0) + a_0.$$

Podem parlar, doncs, de polinomis sobre un anell qualsevol, però a partir d'ara considerarem que l'anell A al que pertanyen els coeficients dels polinomis no té divisors de zero, és a a dir, A és un *domini*.

Grau d'un polinomi

Un dels conceptes més bàsics en l'estudi dels polinomis és el concepte de **grau**. Ja hem dit que tot polinomi $p \neq 0$ s'expressa de manera única com a $p = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$, per algun $n \geq 0$, amb $a_n \neq 0$. En aquest cas, direm que p té grau n. Aquesta definició no ens serveix per al polinomi 0 que, per tant,

[1] Potser és pertinent fer un breu comentari sobre la *notació* que utilitzarem per designar un polinomi genèric. Com en tots els objectes matemàtics, utilitzarem lletres —majúscules, minúscules, gregues, etc.— i escriurem, per exemple, *el polinomi p*, *el polinomi F* o *el polinomi* Ψ. Malgrat això, és cert que hi ha un cert costum d'indicar els polinomis amb una notació que inclou la variable que estiguem utilitzant. Escrivim, per exemple, *el polinomi* $p(x)$, *el polinomi* $F(z)$ o *el polinomi* $\Psi(u)$ per parlar de polinomis genèrics de $A[x]$, $A[z]$ o $A[u]$. Aquesta notació pot ser útil en algunes circumstàncies, però té l'inconvenient d'afavorir la confusió entre polinomis i funcions polinòmiques. En aquest text utilitzarem indistintament les dues notacions.

no té grau. Tanmateix, és útil definir el grau del polinomi zero com $-\infty$. Les propietats bàsiques del grau d'un polinomi són (recordem que estem suposant que A és un domini):

- $\text{grau}(pq) = \text{grau}(p) + \text{grau}(q)$.
- $\text{grau}(p+q) \leq \text{màx}(\text{grau}(p), \text{grau}(q))$.
- Si $\text{grau}(p) \neq \text{grau}(q)$, aleshores a la fórmula anterior val la igualtat.

Fem aquestes observacions senzilles:

- El fet de definir $\text{grau}(0) := -\infty$ fa que les fórmules anteriors sobre el grau siguin vàlides per a tots els polinomis, siguin o no zero, degut a que entenem que $-\infty$ compleix que, per tot n, $-\infty + n = n - \infty = -\infty$ i $-\infty < n$.
- Els polinomis $1, x, x^2, x^3, x^4, \ldots$ són tots diferents i, per tant, l'anell $A[x]$ té sempre infinits elements.
- Els polinomis de $A[x]$ de grau ≤ 0 formen un anell que podem identificar a l'anell A. Direm que A és l'*anell de coeficients* de $A[x]$.
- $A[x]$ és també un domini.
- Les unitats de l'anell $A[x]$ són les unitats de A. Si A és un cos, les unitats de $A[x]$ són, doncs, els polinomis de grau zero.

Funcions polinòmiques i zeros

Si tenim un polinomi $p \in A[x]$ i tenim un element $a \in A$, podem substituir a p la variable x per l'element a i obtenim un element de A que, naturalment, denotarem $p(a)$. És a dir, cada polinomi de $A[x]$ ens defineix una aplicació $A \to A$ per la fórmula $a \mapsto p(a)$. Direm que tenim una **funció polinòmica** sobre A que, per abús de llenguatge, designarem també per p o $p(x)$.

És important no confondre el polinomi $p \in A[x]$ amb la funció polinòmica $p : A \to A$. Són dos conceptes íntimament relacionats, però són *coses diferents*. Per exemple, sigui qui sigui A, hi ha infinits polinomis amb coeficients a A, però si, per exemple, A és un anell finit, només hi pot haver una quantitat finita d'aplicacions $A \to A$ i, en particular, una quantitat finita de funcions polinòmiques sobre A. Això ens demostra que és perfectament possible que **polinomis diferents donin lloc a funcions polinòmiques iguals**.

Un cas extrem seria quan $A = \mathbb{F}_2$. En aquest cas, només hi ha dues funcions $A \to A$: la identitat i l'aplicació zero. En canvi, hi ha infinits polinomis a $A[x]$. Un exemple més interessant és el cas dels polinomis de $\mathbb{F}_p[x]$. Pel petit teorema de Fermat, el polinomi no nul $x^p - x$ dóna lloc a la funció polinòmica zero.

Si considerem un polinomi $p \in A[x]$ i $a \in A$ és tal que $p(a) = 0$, direm que a és un **zero** de p, o també que a és una **arrel** de p. Les preguntes de si un polinomi p té algun zero, de quants en té i de com els podem obtenir són preguntes molt importants que discutirem més endavant.

El concepte de zero d'un polinomi ens duu al concepte d'**element algebraic** que ja ha aparegut anteriorment en aquestes notes (exercici **II**.43). Suposem que tenim dos anells $A \subseteq B$. Un element $b \in B$ direm que és *algebraic sobre* A si existeix un polinomi $p \in A[x]$ tal que b és un zero de p. El cas més interessant és quan els dos anells són $\mathbb{Q} \subset \mathbb{R}$. En aquest cas, direm que un nombre real a és algebraic quan és algebraic sobre $\mathbb{Q}$, segons la definició anterior. Per exemple $\sqrt{2} \in \mathbb{R}$ és un nombre algebraic perquè és un zero del polinomi $x^2 - 2 \in \mathbb{Q}[x]$.

Divisió euclidiana de polinomis

Recordem la importància que va tenir, en l'estudi de l'aritmètica dels nombres enters, la *divisió amb residu*. De la mateixa manera, un fet molt transcendent en l'estudi de les propietats dels polinomis **sobre un cos** és l'existència d'una divisió amb residu pràcticament anàloga a la dels nombres enters.

> *Siguin $D, d \in k[x]$, on k és un cos i $d \neq 0$. Aleshores, existeixen $q, r \in k[x]$ únics, tals que*
>
> 1. $D = qd + r$.
> 2. $\text{grau}(r) < \text{grau}(d)$

Com que el teorema és evidentment cert si $D \in k$, el demostrarem en general per inducció sobre el grau de D. Suposem

$$D = a_n x^n + \cdots + a_0, \quad a_n \neq 0$$
$$d = b_m a^m + \cdots + b_0, \quad b_m \neq 0.$$

Si $n < m$, prenem $q = 0$, $r = D$. En cas contrari, considerem

$$D - \frac{a_n}{b_m} x^{n-m} d$$

que és un polinomi de grau $< n$ sobre el qual podem aplicar la hipòtesi d'inducció. Tindrem

$$D - \frac{a_n}{b_m} x^{n-m} d = qd + r$$

i això ens demostra que el teorema és cert per a D.

Finalment, cal demostrar la unicitat de q i r. Suposem $qd + r = q'd + r'$. Tindrem

$$(q - q')d = r' - r$$

i, igualant els graus dels dos termes, tindrem

$$\text{grau}(q - q') + \text{grau}(d) \leq \text{màx}\{\text{grau}(r), \text{grau}(r')\} < \text{grau}(d).$$

Aquesta desigualtat només és possible si $\text{grau}(q-q') = -\infty$, és a dir, si $q-q' = 0$, i aleshores, també $r - r' = 0$.

La divisió amb residu de polinomis **sobre un cos** és completament efectiva, en el sentit que hi ha un algorisme senzill per calcular explícitament el quocient q i el residu r, un algorisme que apareix implícit a la demostració per inducció anterior. També, igual com va passar en el cas dels nombres enters, aquesta divisió amb residu té conseqüències molt importants en l'estudi dels polinomis, com veurem més endavant. De moment, presentem només un primer corol·lari fonamental que ens **relaciona els zeros amb la divisibilitat**.

Sigui $p \in k[x]$, k un cos. Aleshores, són equivalents:

1. *$a \in k$ és un zero de p.*
2. *p és divisible per $x - a$.*

La demostració és molt senzilla. Si p és divisible per $x - a$, escrivim $p = (x - a)q$ i és evident que $p(a) = 0$. Recíprocament, suposem que $p(a) = 0$, fem la divisió de p per $x - a$

$$p = q(x - a) + r$$

i ara substituïm x per a als dos costats de la igualtat anterior. Obtenim que $r(a) = 0$. Però $\text{grau}(r) < \text{grau}(x - a) = 1$ i $r \in k$ és una constant. Com que $r(a) = 0$, tenim que $r = 0$ i $p = q(x - a)$.

Polinomis en diverses variables

El procés de construir, a partir d'un anell A, l'anell de polinomis $A[x]$ es pot iterar i podem construir **anells de polinomis en diverses variables**:

$$\begin{aligned} A[x, y] &:= A[x][y] \\ A[x, y, z] &:= A[x, y][z] \\ A[x, y, z, t] &:= A[x, y, z][t] \\ &\dots \end{aligned}$$

Podem parlar, doncs, dels anells $A[x_1, \dots, x_n]$. Observem, però, que encara que comencem amb un cos k, l'anell $k[x]$ ja no és un cos i, per tant, ja no podem assegurar que l'anell $k[x, y]$ tingui totes les bones propietats de l'anell $k[x]$. Per exemple, els anells $k[x_1, \dots, x_n]$ amb $n > 1$ ja no són DIP's, però sí que segueixen sent DFU's.

Les disciplines que estudien els anells de polinomis en diverses variables són l'àlgebra commutativa i la geometria algebraica i depassen el que és raonable d'incloure en un text de fonaments de les matemàtiques com aquest.

26 | $k[x]$ s'assembla molt a $\mathbb{Z}$

En tot aquest capítol k denota un **cos** qualsevol.

Si repassem amb atenció l'aritmètica dels enters que hem estudiat en una part anterior del text, veurem que tot el que vam fer amb els enters —$\mathbb{Z}$ és un DIP, mcd i mcm, algorisme d'Euclides, $\mathbb{Z}$ és un DFU, congruències, els anells quocients de $\mathbb{Z}$, etc— només utilitzava aquestes dues propietats dels enters:

- El valor absolut $a \mapsto |a| \in \mathbb{N}$ que ens va permetre fer demostracions per inducció, entre altres coses.
- L'existència de la *divisió amb residu*.

Resulta que l'anell de polinomis $k[x]$ també disposa d'aquestes dues eines essencials. El paper del valor absolut el pot jugar el concepte de *grau d'un polinomi* i ja hem vist que existeix la divisió amb residu de dos polinomis. Això implica que l'anell $k[x]$ tindrà unes propietats aritmètiques molt similars a les que tenia l'anell $\mathbb{Z}$. Aquest fet és transcendental i no requereix cap més demostració que anar repassant els capítols pertinents dedicats a l'aritmètica de $\mathbb{Z}$ i adaptar-los a l'anell $k[x]$, sense necessitat de fer, pràcticament, cap canvi, més enllà de substituir el valor absolut pel grau. Fem un repàs ràpid:

- $k[x]$ és un DIP i la demostració que vam fer per a l'anell $\mathbb{Z}$ és totalment vàlida en l'anell $k[x]$, precisament perquè existeix la divisió amb residu que era l'eina fonamental de la demostració. Aleshores, també podrem parlar del *màxim comú divisor* i el *mínim comú múltiple* de dos (o més) polinomis. Recordem com es feia: si $p, q \in k[x]$, considerem els ideals $(p) \cap (q)$ i $(p)+(q)$ que, com que $k[x]$ és un DIP, seran ideals **principals** $(p)\cap(q) = (m)$ i $(p) + (q) = (d)$. Aleshores, $d = \text{mcd}(p, q)$ i $m = \text{mcm}(p, q)$.
- Recordem que el mcd i el mcm estan definits llevat d'**unitats** de l'anell. En el cas de $\mathbb{Z}$, les úniques unitats són ± 1 i escollíem d i m que fossin positius. A l'anell $k[x]$ hi ha més unitats que a $\mathbb{Z}$: les unitats de $k[x]$ són tots els polinomis de grau zero, és a dir, les constants diferents de zero, que havíem denotat k^*. Per tant, el mcd i el mcm estan determinats llevat del producte per una constant diferent de zero i podem desfer l'ambigüitat prenent d i m que siguin **polinomis mònics**: un polinomi mònic és aquell en el que el coeficient del monomi de grau màxim és igual a 1.

- Podem parlar de polinomis *primers* o *irreductibles*, que són els polinomis p de grau > 0 que no tenen altres divisors que u i up on u és qualsevol constant diferent de zero. Normalment, en el cas dels polinomis, es prefereix parlar d'*irreductibles*. Els primers positius de $\mathbb{Z}$ són $2, 3, 5, 7, 11, \ldots$ i ens podem preguntar quins són els polinomis irreductibles de $k[x]$. És una pregunta important i, en molts casos, difícil de respondre. Podem fer aquestes consideracions trivials:
  - Si un polinomi de grau > 1 té un zero ja no pot ser irreductible perquè, si el zero és $a \in k$, ja sabem que podrem dividir el polinomi per $x-a$. És important recordar que el recíproc **no és cert**: per exemple, el polinomi $x^4 + 2x^2 + 1 \in \mathbb{Q}[x]$ no té cap zero, però no és irreductible perquè és igual a $(x^2+1)(x^2+1)$.
  - Quins siguin els polinomis irreductibles de $k[x]$ dependrà del cos k. Per exemple, el polinomi $x^2 + x - 1$ és clarament irreductible si $k = \mathbb{Q}$, però si $k = \mathbb{R}$ ja no és irreductible perquè
$$x^2 + x - 1 = \left(x - \frac{-1+\sqrt{5}}{2}\right)\left(x - \frac{-1-\sqrt{5}}{2}\right).$$
  - Els polinomis de grau 1 són clarament irreductibles i els polinomis de grau dos o tres són irreductibles si i només si no tenen cap zero. Això és evident perquè qualsevol descomposició d'un polinomi de grau dos o tres ha d'incloure un polinomi de grau 1, i els polinomis de grau 1 sempre tenen un zero.
- El mcd i el mcm es poden calcular amb l'algorisme d'Euclides, igual que en el cas dels nombres enters, i també tenim **identitat de Bézout** que es pot determinar amb el mateix algorisme d'Euclides.
- $k[x]$ és un DFU. És a dir, tot polinomi $p \in k[x]$ s'expressa de manera única (llevat de l'ordre) com
$$p = uq_1 \cdots q_r$$
on els polinomis q_i són mònics i irreductibles.

Tornem a insistir que estem considerant que k és un cos. Si considerem, per exemple, l'anell $\mathbb{Z}[x]$ dels polinomis amb coeficients enters, ja no tenim divisió amb residu i no podem garantir que les propietats anteriors siguin vàlides. Per exemple, considerem aquest subconjunt de $\mathbb{Z}[x]$:
$$I := \{p \in \mathbb{Z}[x] : p(0) \text{ és parell}\}.$$
És fàcil veure que I és un ideal de $\mathbb{Z}[x]$ i que **no** és *principal*. Amb això ja veiem que $\mathbb{Z}[x]$ no és un DIP i (en principi) no podem parlar ni de mcd ni de mcm a $\mathbb{Z}[x]$.[1]

[1] De fet, sí que podem definir a $\mathbb{Z}[x]$ els conceptes de mcd i mcm com el divisor comú de grau màxim i el múltiple comú de grau mínim, respectivament, però no tenim, en general, identitat de Bézout. També es compleix que $\mathbb{Z}[x]$ és un DFU perquè hi ha un teorema que afirma que si A és un DFU, també ho és $A[x]$. Vegeu l'exercici V.7.

Una altra conseqüència important de la teoria de la divisibilitat a $k[x]$ i de la relació entre divisibilitat i zeros que hem vist a la pàgina 176 és aquesta:

> *$a \in k$ és un zero comú de dos polinomis $p, q \in k[x]$ si i només si a és un zero de* $\operatorname{mcd}(p, q)$.

La demostració d'aquest fet és senzilla: sigui $d := \operatorname{mcd}(p, q)$, escrivim $p = hd$, $q = td$ i $d = rp + sq$ (identitat de Bézout). Aleshores, si $d(a) = 0$, és clar que $p(a) = q(a) = 0$ i, recíprocament, si $p(a) = q(a) = 0$, la identitat de Bézout ens dóna $d(a) = 0$.

Exemple

Suposem que volem resoldre aquest sistema d'equacions sobre un cert cos k de característica $\neq 2, 3$:

$$\begin{cases} x^4 - 5x^2 - 2x + 3 = 0 \\ x^4 + 2x^3 + x^2 - 1 = 0 \end{cases}$$

Comencem calculant el mcd dels dos polinomis $p = x^4 - 5x^2 - 2x + 3$, $q = x^4 + 2x^3 + x^2 - 1$ anteriors, utilitzant l'algorisme d'Euclides:

$$\begin{aligned} p &= q + r_0, & r_0 &= -2x^3 - 6x^2 - 2x + 4 \\ q &= \frac{1 - x}{2} r_0 + r_1, & r_1 &= 3x^2 + 3x - 3 \\ r_0 &= -\frac{2x + 4}{3} r_1 + 0 \end{aligned}$$

Per tant, el mcd és $x^2 + x - 1$ i les solucions del sistema d'equacions seran els zeros d'aquest polinomi. Com que estem suposant que k és un cos de característica diferent de 2, podem utilitzar la fórmula universal per als zeros dels polinomis de grau dos i observem que si 5 no és un quadrat a k, el sistema no té solució, i si 5 és un quadrat a k, diguem $5 = \zeta^2$, les solucions són $(-1 \pm \zeta)/2$.

27 | Multiplicitat d'un zero. Polinomis irreductibles

Recordem que la primera conseqüència que vam deduir de l'existència de la divisió euclidiana als anells de polinomis sobre un cos —i en aquest capítol seguirem suposant que l'anell de coeficients és un **cos**— va ser aquesta relació entre els zeros d'un polinomi i la divisibilitat:

> *Sigui $p \in k[x]$ un polinomi $\neq 0$, k un cos. Aleshores, són equivalents:*
>
> 1. *$a \in k$ és un zero de p.*
> 2. *p és divisible per $x - a$.*

Ara volem anar una mica més enllà en aquesta línia i introduirem el concepte de **multiplicitat** d'un zero d'un polinomi:

> *Sigui $p \in k[x]$, $p \neq 0$, i sigui $a \in k$ un zero de p. Definim la **multiplicitat de** a **com a zero de** p com el màxim $r > 0$ tal que $(x - a)^r$ divideix p.*

Un zero de multiplicitat > 1 direm que és un *zero múltiple*. Ara podem contestar aquesta pregunta important: *quants zeros pot tenir, com a màxim, un polinomi de grau n?*

> *El nombre de zeros d'un polinomi de grau $n \geq 0$, cadascun comptat tantes vegades com indiqui la seva multiplicitat, és $\leq n$.*

La demostració és immediata perquè cada zero de p de multiplicitat r dóna lloc a un factor de p de grau r. Per tant, si els zeros (diferents) de p són $a_1, \ldots, a_k$ i cadascun té una multiplicitat $r_1, \ldots, r_k$, aleshores podem descompondre

$$p = (x - a_1)^{r_1}(x - a_2)^{r_2} \cdots (x - a_k)^{r_k} q$$

i, en conseqüència, $r_1 + \cdots + r_k \leq \text{grau}(p)$.

Derivada d'un polinomi

El concepte de *derivada d'una funció* pertany al càlcul infinitesimal perquè involucra el concepte de límit, però quan es tracta d'un polinomi, podem donar una definició completament algebraica de la derivada, vàlida per a polinomis sobre un cos qualsevol:

Definim *la derivada d'un polinomi* $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ *com*

$$p'(x) = na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + a_1.$$

Aleshores, les derivades que acabem de definir tenen un comportament similar al de les derivades de funcions reals:[1]

- $(p+q)' = p' + q'$. La demostració és trivial.
- $(pq)' = pq' + p'q$. Vegeu l'exercici V.8.
- També es compleix la *regla de la cadena*: $[p^r]' = rp^{r-1}p'$. La demostració és senzilla per inducció sobre r. Vegeu l'exercici V.8.
- La derivada d'un polinomi constant és zero i, **si la característica del cos** k **és zero**, també és cert el recíproc: si $p' = 0$ aleshores p és constant.

Observem que quan la característica del cos és > 0 la derivada té un comportament una mica estrany. Per exemple, sigui p un nombre primer i considerem $x^p \in \mathbb{F}_p[x]$. La derivada d'aquest polinomi és zero, però no és un polinomi constant.

Sobre cossos de característica zero hi ha una relació important entre la derivada i les arrels múltiples:

Sigui $p \neq 0$ un polinomi sobre un cos de característica zero.

- *Si a és un zero de p de multiplicitat $r > 1$, aleshores a és un zero de p' de multiplicitat $r-1$.*
- *Els zeros múltiples de p són els zeros de* $\operatorname{mcd}(p, p')$.

Demostració. Si a és un zero de p de multiplicitat $r > 1$, podrem escriure $p = (x-a)^r q$ amb $q(a) \neq 0$. Calculem ara la derivada de p:

$$p' = [(x-a)^r q]' = (x-a)^r q' + r(x-a)^{r-1}q = (x-a)^{r-1}((x-a)q' + rq).$$

Sigui $h := (x-a)q' + rq$ i observem que $h(a) = rq(a) \neq 0$ i, per tant, a és un zero de p' de multiplicitat $r-1$.

[1] Recordem que no podem utilitzar les demostracions del càlcul infinitesimal perquè estem considerant polinomis sobre un cos arbitrari k en el que el concepte de *límit* pot no tenir sentit.

Si a és un zero múltiple de p, hem vist que és un zero de p' i, per tant, $x-a$ dividirà p i dividirà p'. Per tant, $x-a$ dividirà $d := mcd(p,p')$ i a serà un zero de d. Recíprocament, suposem que a és un zero de d, amb la qual cosa $x-a$ divideix d. Com que, per definició, d divideix p i divideix p', obtenim que $p(a) = p'(a) = 0$. Utilitzem ara el càlcul de p' que hem fet abans i suposem que $r = 1$. Obtenim $0 = p'(a) = rq(a) \neq 0$, una contradicció. Per tant, $r > 1$ i a és zero múltiple de p.[2]

Polinomis irreductibles sobre $\mathbb{Z}$ i sobre $\mathbb{Q}$

Sabem que $k[x]$ és un DFU per tot cos k i també és cert (però no ho hem demostrat) que $\mathbb{Z}[x]$ és un DFU. Una pregunta natural és com podem determinar si un polinomi és irreductible i, més en general, com podem factoritzar un polinomi en els seus factors irreductibles. El cas dels polinomis a coeficients enters o racionals és especialment important de cara a les aplicacions pràctiques —que inclouen, entre altres coses, la *criptografia de clau pública*— i s'han desenvolupat algorismes prou eficients per factoritzar aquests polinomis. El cas dels polinomis sobre cossos finits també és molt important perquè forma part dels algorismes de factorització sobre $\mathbb{Z}$ i $\mathbb{Q}$. Tots aquests problemes constitueixen un capítol important de l'*àlgebra computacional* i aquí no podem tractar-los amb profunditat. Però, com hem fet en altres casos al llarg d'aquestes notes, sí que considerarem alguns principis elementals sobre aquest tema.

Suposem, doncs, que tenim un polinomi de grau $n > 0$ amb coeficients enters

$$Q(x) = a_n x^n + a_{n-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$$

i suposem que es tracta d'un polinomi *primitiu*, que vol dir que els seus coeficients no tenen cap divisor comú ($\neq \pm 1$), és a dir, $\mathrm{mcd}(a_0, \ldots, a_n) = 1$. Ens preguntem si Q és irreductible sobre $\mathbb{Q}$ i si ho és sobre $\mathbb{Z}$.

- **Casos trivials a $k[x]$.** Com hen dit més amunt, tot polinomi de grau 1 és irreductible i un polinomi de grau 2 o 3 és irreductible si i només si no té cap zero.
- **Factoritzar Q sobre $\mathbb{Z}$ és el mateix que factoritzar Q sobre $\mathbb{Q}$.** Aquest fet es coneix com a **lema de Gauss**:

  > **Lema de Gauss**. *Sigui Q un polinomi de grau > 0, amb coeficients enters. Si Q es pot descompondre com a producte de dos polinomis de $\mathbb{Q}[x]$ de grau > 0, també es pot descompondre com a producte de dos polinomis de $\mathbb{Z}[x]$ de grau > 0.*

[2] Observem que no afirmem el recíproc de la primera part del teorema: que a sigui un zero de p' de multiplicitat $r-1$ **no implica** que a sigui un zero de p. Un contraexemple senzill seria $p = x^2 + 1$.

Demostració. Suposem que

$$Q = A_1 A_2$$

on $A_1, A_2 \in \mathbb{Q}[x]$ són polinomis de grau > 0. L'objectiu serà demostrar que $Q = B_1 B_2$ on $B_1, B_2 \in \mathbb{Z}[x]$ són polinomis de grau > 0.

Siguin d_1, d_2 els mcm dels denominadors dels coeficients de A_1, A_2, respectivament, de manera que

$$A_1' := d_1 A_1, \quad A_2' := d_2 A_2$$

són polinomis amb coeficients enters. Sigui $d = d_1 d_2$, de manera que

$$dQ = A_1' A_2'. \qquad (*)$$

Si $d = 1$, la descomposició inicial ja és una descomposició sobre $\mathbb{Z}$ i hem acabat. Si $d > 1$, sigui p un primer que divideixi d i *reduïm ara la igualtat (*) mòdul p*. Obtenim

$$0 = A_1' A_2' \text{ a } \mathbb{F}_p[x].$$

Però $\mathbb{F}_p[x]$ és un *domini* i, per tant, un dels dos polinomis A_1', A_2' serà zero a $\mathbb{F}_p[x]$. Suposem, sense pèrdua de generalitat, que $A_1' = 0 \in \mathbb{F}_p[x]$. Això vol dir que tots els coeficients de A_1' són divisibles per p i, per tant, $\frac{1}{p} A_1'$ és un polinomi a coeficients enters i tenim una descomposició a $\mathbb{Z}[x]$

$$\frac{d}{p} Q = \left(\frac{1}{p} A_1 \right) A_2$$

en la que podem observar que hem reduït el factor que multiplica Q de d a d/p. Si ara anem repetint aquest procés arribarem a $d = 1$ i haurem acabat.

- **El Criteri d'Eisenstein** ens permet concloure **en alguns casos** que un polinomi és irreductible sobre $\mathbb{Z}$.

  > **Criteri d'Eisenstein.** *Sigui $Q = a_n x^n + a_{n-1} x^{n-1} + a_1 x + a_0 \in \mathbb{Z}[x]$ un polinomi primitiu i sigui p un primer que divideixi els coeficients $a_0, \ldots, a_{n-1}$ i tal que p^2 no divideixi a_0. Aleshores, Q és irreductible a $\mathbb{Z}[x]$ (i, pel lema de Gauss, també ho és a $\mathbb{Q}[x]$).*

  La **demostració** és una aplicació interessant de les coses que hem anat estudiant fins ara. Suposem que Q es pogués expressar com a producte de dos polinomis de $\mathbb{Z}[x]$ de grau > 0: $Q = GH$. Considerem ara la *reducció mòdul p* que ens passa de $\mathbb{Z}[x]$ a $\mathbb{F}_p[x]$. Tindrem una igualtat de polinomis a $\mathbb{F}_p[x]$ $\overline{Q} = \overline{G}\,\overline{H}$ i les hipòtesis del criteri ens diuen que $\overline{Q} = a_n x^n \in \mathbb{F}_p[x]$. Observem que, com que Q és un polinomi primitiu, p no pot dividir a_n i $\overline{Q} \neq 0$.

  Ara apliquem el fet important que $\mathbb{F}_p[x]$ és un DFU. Deduïm que $\overline{G} = b x^r$ i $\overline{H} = c x^{n-r}$. Observem que $0 < r < n$ perquè $\overline{G}, \overline{H}$ són polinomis de grau $< n$. Tenim, doncs, $\overline{G}(0) = \overline{H}(0) = 0$ i això vol dir que $G(0) \equiv 0$ mòdul p i $H(0) \equiv 0$ mòdul p. Per tant, $a_0 = Q(0) = G(0)H(0) \equiv 0$ mòdul p^2 i això contradiu les hipòtesis del teorema.

Arrels de la unitat

Considerem el polinomi $x^r - 1$ amb $r > 1$. De manera natural, direm que els zeros d'aquest polinomi són **les arrels r-èsimes de la unitat** i, per les propietats dels zeros d'un polinomi que hem estudiat, d'aquestes arrels n'hi ha un màxim de r i un mínim d'una ($x = 1$). Depenent de r i del cos base k, hi haurà més o menys arrels r-èsimes de la unitat. Per exemple

- Si el cos base és $\mathbb{Q}$ o $\mathbb{R}$, les arrels r-èsimes de la unitat són ± 1 si r és parell i 1 si r és senar.
- En característica zero, no hi pot haver arrels de la unitat *múltiples* perquè, clarament, $x^r - 1$ i la seva derivada rx^{r-1} són coprimers.
- Si el cos base és $\mathbb{F}_p$, sabem pel (petit) teorema de Fermat $a^{p-1} = 1$ per tot $a \in \mathbb{F}_p - \{0\}$. Per tant, sobre el cos $\mathbb{F}_p$ hi ha exactament $p - 1$ arrels $(p-1)$-èsimes de la unitat, el màxim possible.

Considerem a partir d'ara el cas important en que cos base és $\mathbb{Q}$. Si r és senar, el polinomi $x^r - 1$ sempre es pot dividir per $x - 1$:

$$x^r - 1 = (x-1)\,(x^{r-1} + x^{r-2} + \cdots + x + 1)$$

i si r és parell, $x^r - 1$ es pot dividir per $x^2 - 1$:

$$x^r - 1 = (x^2-1)\,(x^{r-2} + x^{r-4} + \cdots + x^2 + 1).$$

però la descomposició completa de $x^r - 1$ en producte de polinomis irreductibles és un tema complex de l'aritmètica. Un cas interessant i relativament senzill que podem dilucidar amb el que hem après fins ara és aquest:

Si p és un primer, $x^{p-1} + x^{p-2} + \cdots + x + 1$ és irreductible.

La demostració és una aplicació del criteri d'Eisenstein. Observem que

$$x^{p-1} + x^{p-2} + \cdots + x + 1 = \frac{x^p - 1}{x - 1}$$

i observem que **no** es compleixen les hipòtesis del criteri d'Eisenstein, però sí que es compleixen si fem el canvi de variable $x \mapsto x + 1$ i obtenim

$$\frac{(x+1)^p - 1}{x} = x^{p-1} + \binom{p}{1} x^{p-2} + \cdots + \binom{p}{2} x + p.$$

Per una propietat ben coneguda (exercici **IV**.14) dels nombres combinatoris

$$\binom{p}{i} \equiv 0 \quad \text{mòd } p \text{ per tot } 0 < i < p.$$

Per tant, podem aplicar el criteri d'Eisenstein al polinomi anterior i obtenim que és irreductible sobre $\mathbb{Q}$. Això implica que el polinomi inicial (abans de fer el canvi de variable $x \mapsto x + 1$) també és irreductible perquè, de manera trivial, una factorització $Q(x) = G(x)\,H(x)$ dóna una factorització $Q(x+1) = G(x+1)H(x+1)$.

28 Resolució d'equacions polinòmiques

En aquest capítol parlarem de la resolució d'equacions de la forma $p(x) = 0$ on p és un polinomi de l'anell $k[x]$, per un cert cos k.[1] Plantegem-nos, doncs, què podem dir sobre les solucions de

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0, \quad a_n \neq 0,\ n > 1,$$

més enllà del que ja hem estudiat en els capítols precedents. Recordem que les solucions de l'equació $p(x) = 0$ també reben els noms de *zeros* o *arrels* de p.

Solucions enteres o racionals: un cas trivial

Podem afirmar que trobar les solucions enteres o racionals d'una equació polinòmica qualsevol

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

en la que els coeficients són *enters* és un problema essencialment trivial. Observem que si els coeficients fossin racionals, podem multiplicar l'equació pel mcm dels denominadors i obtenir una equació amb les mateixes solucions i els coeficients enters. Que aquest cas sigui trivial es deu a aquest senzill i ben conegut teorema:

> **Teorema de les arrels racionals.** *Si el nombre racional u/v (amb u, v enters coprimers) és una arrel del polinomi amb coeficients enters $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$, aleshores u divideix a_0 i v divideix a_n. En particular, qualsevol arrel entera divideix a_0.*

Com que a_n i a_0 tenen un nombre finit de divisors, obtenim un conjunt finit de nombres racionals que conté totes les arrels racionals de $p(x)$. Avaluant $p(x)$

[1] L'àmbit natural d'aquest tema de la resolució de les equacions algebraiques és una assignatura de teoria de cossos —el nom de l'assignatura pot ser també *teoria de Galois*— que, probablement, l'estudiant trobarà més endavant en els seus estudis, però és apropiat que en un text de fonaments de les matemàtiques hi hagi una introducció elemental a aquest tema important.

en cadascun dels elements d'aquest conjunt podem trobar, en principi, totes les arrels racionals de qualsevol polinomi enter.

La **demostració** del teorema anterior és ben senzilla. Si u/v és un zero del polinomi $p(x)$, tindrem $p(u/v) = 0$ i

$$a_n \frac{u^n}{v^n} + a_{n-1} \frac{u^{n-1}}{v^{n-1}} + \cdots + a_1 \frac{u}{v} = -a_0.$$

Multipliquem ara els dos membres per v^n i obtenim aquesta expressió sense denominadors:

$$u(a_n u^{n-1} + a_{n-1} v u^{n-2} + \cdots + a_1 v^{n-1}) = -a_0 v^n$$

que, com que u i v són coprimers, implica que u divideix a_0. Si ara escrivim la igualtat anterior en la forma

$$v(a_{n-1} u^{n-1} + \cdots + a_0 v^{n-1}) = -a_n u^n$$

obtenim que v divideix a_n.

Primeres consideracions sobre el cas general

Suposem ara que tenim una equació polinòmica general

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

sobre un cos qualsevol k i volem trobar les seves solucions. Com ho podem fer? Comencem amb aquestes observacions elementals:

- No és restrictiu suposar $a_n = 1$ perquè si dividim el polinomi per a_n obtenim un altre polinomi que té exactament els mateixos zeros que el polinomi inicial.
- Si en el cos k tenim $n \neq 0$, no és restrictiu suposar $a_{n-1} = 0$. En efecte, el canvi de variable
$$y := x + \frac{a_{n-1}}{n}$$
ens converteix l'equació inicial en una equació de la forma
$$y^n + b_{n-2} y^{n-2} + \cdots + b_0 = 0$$
i les solucions de l'equació original es corresponen amb les solucions de la segona equació a través del canvi de variable que hem utilitzat.
- Les observacions anteriors ens permeten resoldre les equacions de segon grau en qualsevol cos de característica diferent de 2. Efectivament, considerem l'equació ($a \neq 0$)
$$ax^2 + bx + c = 0.$$

En primer lloc, dividim per a i obtenim

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0.$$

Ara fem el canvi de variable $x = y - \frac{b}{2a}$ i obtenim l'equació

$$y^2 = \frac{b^2 - 4ac}{4a^2}.$$

Ara, si convenim en *denotar* $\pm\sqrt{u}$ els elements de k tals que el seu quadrat és igual a u (en el cas que existeixin), i desfem el canvi de variable, obtenim la conegudíssima *fórmula* de les arrels de l'equació de segon grau, vàlida sobre qualsevol cos de característica diferent de 2:

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

Per tant, ja sabem resoldre les equacions de segon grau sobre un cos qualsevol (de característica $\neq 2$). N'estem segurs? Parlem-ne!

Què volem dir quan parlem de resoldre una equació?

Què volem dir –exactament— quan diem que volem trobar les solucions d'una equació algebraica? Quan el cos base és $\mathbb{Q}$ o un cos finit, la resposta a aquesta pregunta és molt clara. No hi ha cap ambigüitat en voler trobar les solucions de $x^3 + 2x^2 + 2x + 1 = 0$ sobre el cos $\mathbb{Q}$ (només en té una: $x = -1$) o sobre el cos $\mathbb{F}_7$ (en té tres: $x = 2, 4, 6$) però si el cos on volem trobar les solucions és el cos real $\mathbb{R}$, on els elements poden tenir una infinitat de decimals, com volem, realment, escriure les solucions? Quan podrem dir que les hem *trobat*?

Per exemple, què volem dir quan parlem de *trobar* les solucions reals de $x^2 = 2$? Dir que la solució és $x = \pm\sqrt{2}$ és entrar en un *cercle viciós* perquè $\pm\sqrt{2}$ vol dir —*per definició!*— les solucions de l'equació $x^2 = 2$.

Què volem dir quan diem que la fórmula de l'apartat anterior *resol l'equació de segon grau*? Observem que el que realment fa la fórmula és reduir el problema de resoldre qualsevol equació de segon grau a la resolució d'una equació del tipus $x^2 = d$. És a dir, quan diem que sabem resoldre les equacions de segon grau, estem dient que sabem expressar totes les solucions utilitzant les operacions del cos —suma, resta, multiplicació i divisió— i, a més, una nova operació que és l'arrel quadrada. Direm que la clàssica fórmula de l'equació de segon gran ens permet **resoldre l'equació per radicals**.

En el cas d'una equació de grau n, també direm que la sabem resoldre per radicals si som capaços d'escriure les solucions utilitzant les operacions del cos i, a més, arrels quadrades, cúbiques, etc, fins a grau n.

D'altra banda, si el que volem, en el cas real, són valors aproximats —amb qualsevol nivell d'aproximació que desitgem— dels nombres reals que compleixen l'equació, hi ha molts mètodes —coneguts com a *mètodes numèrics*— que ens ho permeten fer. Considerem, per exemple, l'equació

$$x^3 - 3x + 1 = 0.$$

Si demanem a algun programa de càlcul matemàtic que ens resolgui aquesta equació, això és el que obtenim:

```
> x = polygen(RR)
> (x^3-3*x+1).roots()
  [(-1.87938524157182, 1), (0.347296355333861, 1), (1.53208888623796, 1)]
```

i ja hem trobat les tres solucions de l'equació, en el sentit que tenim les seves expressions decimals, amb una precisió de 14 xifres decimals. En els casos pràctics, és això el que voldrem, però des d'un punt de vista teòric potser ens agradaria tenir una expressió *exacta* de les solucions, per exemple, utilitzant arrels quadrades i cúbiques, o alguna altra funció. Per exemple, les tres solucions de l'equació anterior són *«exactament»* aquestes:

$$x_1 = 2\cos\left(\frac{2\pi}{9}\right), \quad x_2 = 2\cos\left(\frac{8\pi}{9}\right), \quad x_3 = 2\cos\left(\frac{14\pi}{9}\right).$$

Per comprovar-ho, observem que si θ és qualsevol dels tres angles anteriors, aleshores $3\theta = 2\pi/3$ i $\cos(3\theta) = -1/2$. Apliquem ara la fórmula trigonomètrica del cosinus de l'angle triple

$$\cos 3\theta = 4\cos^3\theta - 3\cos\theta$$

i observem que, efectivament, x_1, x_2, x_3 són tres solucions diferents de l'equació de tercer grau inicial i són, per tant, totes les solucions de l'equació, expressades d'una manera *exacta* però ni *algebraica* ni *per radicals*. És això el que volem? Potser sí. Havíem demanat solucions *exactes* i aquestes ho són, encara que, potser hauríem preferit utilitzar, en lloc de la funció cosinus, les funcions arrel quadrada o arrel cúbica?

Resolució per radicals de les equacions polinòmiques

Aquest problema clàssic demana trobar fórmules que expressin les solucions de qualsevol equació polinòmica sobre els reals utilitzant només les operacions del cos (suma, resta, producte, divisió) i arrels (de qualsevol grau menor o igual al de l'equació) —de la mateixa manera que sabem fer-ho per a l'equació de segon grau, com a mínim des del segle III AC. El primer cas que calia resoldre era, evidentment, el de l'equació de tercer grau, la *cúbica*.

La història del descobriment al segle XVI de la solució de la cúbica *per radicals* és apassionant però no la podem incloure en aquest text. El que sí que és

important que fem aquí és, com a mínim, sembrar el dubte sobre si és assenyat esperar que totes les equacions es puguin resoldre per radicals. Plantegem la situació d'aquesta manera:

- Hi ha dos tipus de nombres reals irracionals: els que són solució d'alguna equació polinòmica i els que no són solució de cap equació polinòmica. Els primers s'anomenen nombres algebraics, els segons s'anomenen nombres transcendents. Entre els nombres algebraics podem fer esment de

$$\sqrt{2}, \quad \varphi = \frac{1+\sqrt{5}}{2} \text{ (la raó àuria)}, \quad \cos\left(\frac{2\pi}{9}\right), \ldots$$

  Entre els transcendents podem esmentar

$$e, \quad \pi, \quad e^{\pi}, \quad \cos(1), \ldots$$

  Designem per $\mathcal{A} \subsetneq \mathbb{R}$ el subconjunt dels nombres algebraics, que es pot demostrar que formen un cos.

- També podem considerar tots els nombres reals que es poden expressar a partir dels nombres racionals amb les operacions de suma, resta, multiplicació, divisió i arrel n-èsima, per tot n. Designem $\mathcal{R}$ el conjunt de tots aquests nombres. Es pot demostrar que tots aquests nombres són algebraics i, evidentment, formen un cos. Tenim

$$\mathcal{R} \subseteq \mathcal{A} \subsetneq \mathbb{R}$$

  i no tenim cap motiu per pensar que $\mathcal{R} = \mathcal{A}$. En canvi, si *totes* les equacions polinòmiques es poguessin resoldre *per radicals*, això implicaria $\mathcal{R} = \mathcal{A}$.

El problema de decidir si totes les equacions polinòmiques es poden resoldre per radicals o no va ser un problema molt important que va restar obert durant molts anys. La resposta a aquest problema es va veure que era negativa. En particular, $\mathcal{R} \neq \mathcal{A}$ i un exemple de nombre algebraic que **no** es pot expressar per radicals és, precisament, $\cos(2\pi/9)$. O potser sí que es pot? Ho estudiem a continuació.

Resolució per radicals de la cúbica

Tres matemàtics italians del segle XVI —Scipione del Ferro, Niccolò Tartaglia, Gerolamo Cardano— van trobar un mètode per resoldre *per radicals* l'equació de tercer grau (sobre el cos dels nombres reals). Amb la notació actual i, sobre tot, amb la utilització dels nombres negatius, la solució d'aquests matemàtics és relativament senzilla d'explicar.

Com hem vist abans, n'hi ha prou amb considerar l'equació $x^3 + px + q = 0$. També n'hi ha prou amb trobar *una* solució, perquè quan tenim una solució a podem dividir per $x - a$ i el problema de trobar les dues altres solucions (si existeixen) queda reduït a resoldre una equació de segon grau. La solució de la cúbica s'obté en dos passos.

- **Primer pas.** En aquest primer pas resolem, utilitzant una equació de segon grau, aquest sistema de dues equacions amb dues incògnites:

$$\begin{cases} u^3 + v^3 = a \\ uv = b \end{cases}$$

Aquest sistema és molt senzill de resoldre perquè

$$(X - u^3)(X - v^3) = X^2 - aX + b^3$$

i, aleshores, u^3, v^3 són les solucions d'aquesta equació de segon gran. Si apliquem ara la fórmula per a les arrels d'una equació de segon grau, obtenim

$$u = \sqrt[3]{\frac{a + \sqrt{a^2 - 4b^3}}{2}}, \quad v = \frac{b}{u} = \sqrt[3]{\frac{a - \sqrt{a^2 - 4b^3}}{2}}.$$

- **Segon pas.** Siguin ara u, v solucions del sistema del primer pas amb $a = -q$ i $b = -p/3$. Una comprovació trivial demostra que $x := u + v$ és solució de la cúbica inicial. Tenim, doncs, una solució de la cúbica que ve donada per la que es coneix com a *fórmula de Cardano*,[2] que ens dóna la solució *per radicals* de l'equació de tercer grau:

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

Apliquem ara la fórmula de Cardano per resoldre un parell d'exemples:

- L'equació $x^3 + x + 1 = 0$. Aplicant directament la fórmula de Cardano, una solució és

$$x = \sqrt[3]{\frac{-9 + \sqrt{93}}{18}} + \sqrt[3]{\frac{-9 - \sqrt{93}}{18}} \approx -0.682327803828019.$$

No té cap més solució real.

- L'equació $x^3 - 3x + 1 = 0$. Les tres solucions d'aquesta equació ja les havíem trobat anteriorment:

$$x_1 = 2\cos\left(\frac{2\pi}{9}\right) \approx 1.532,\ x_2 = 2\cos\left(\frac{8\pi}{9}\right) \approx -1.879,\ x_3 = 2\cos\left(\frac{14\pi}{9}\right) \approx 0.347$$

i la fórmula de Cardano ens hauria de trobar una expressió amb radicals d'una d'aquestes solucions. Curiosament, la fórmula de Cardano dóna aquest resultat:

$$x = \sqrt[3]{\frac{-1 + \sqrt{-3}}{2}} + \sqrt[3]{\frac{-1 - \sqrt{-3}}{2}}$$

[2] Observem com, una vegada més, presentem la solució d'un problema —i demostrem que és efectivament una solució— sense donar cap indici sobre com els descobridors de la solució van arribar a trobar-la. Això seria una altra història.

que, com que involucra nombres *impossibles* com l'arrel quadrada de -3, sembla indicar que l'equació no té cap solució —cosa que ja sabem que no és certa— o que la fórmula de Cardano conté un error —però la comprovació de la fórmula que hem fet és impecable. Aquesta situació va confondre profundament el mateix Cardano i altres matemàtics de l'època. De fet, si coneixem la teoria dels *nombres complexos*, aquesta situació aparentment paradoxal queda immediatament resolta:

- La fórmula de Cardano és correcta i ens dóna una solució expressada amb radicals, en els nombres complexos.
- El valor x que dóna la fórmula de Cardano és suma de dos nombres complexos no reals, però x és **real** perquè els dos nombres complexos són *conjugats*. De fet, $x \approx 1.532$ i coincideix amb la solució $2\cos(2\pi/9)$ que havíem trobat abans.[3]
- Tota aquesta situació posa de manifest que l'àmbit naturals per estudiar les solucions d'una equació polinòmica és el dels *nombres complexos*, **encara que només ens interessin les arrels reals**.
- Efectivament, la fórmula de Cardano ens demostra que totes les cúbiques es poden resoldre per radicals *sobre els nombres complexos*.
- L'exemple que acabem d'estudiar suggereix que, en canvi, no totes les cúbiques es poden resoldre per radicals si exigim de romandre sempre dins dels nombres reals. Es pot demostrar que és impossible expressar les solucions de l'equació $x^3 - 3x + 1 = 0$ utilitzant només nombres reals, arrels cúbiques i arrels quadrades de nombres positius.[4]

Resolució per radicals de les equacions de graus superiors

El pas següent a la resolució de la cúbica per radicals seria la resolució per radicals de l'equació de quart grau —la *quàrtica*— per a la qual el mateix Cardano (any 1545) va trobar una solució que involucrava, entre altres coses, resoldre una

[3] Aquí estem obviant el fet que, quan treballem sobre els complexos, els nombres diferents de zero tenen *tres* arrels cúbiques diferents i, aleshores, la fórmula de Cardano no ens dóna una solució de l'equació, sinó que ens les dóna totes tres. Vegeu l'exercici **VI**.14.

[4] Ara que *passem per aquí*, seria imperdonable no fer esment de la importància d'aquesta equació concreta en la resolució d'un dels grans problemes no resolts de la geometria d'Euclides: la **trisecció de l'angle**. Aquest problema demana trobar una construcció *amb regle i compàs* que divideixi qualsevol angle donat en tres angles iguals. La solució al problema és que aquesta construcció amb regle i compàs *no existeix*. Per demostrar-ho, observem que la construcció clàssica amb regle i compàs del triangle equilàter ens permet dibuixar un angle de $2\pi/3$. Si la trisecció fos possible —amb regle i compàs!— podríem dibuixar un angle de $2\pi/9$ i, a partir d'aquí, podríem dibuixar un segment de longitud $x_1 = 2\cos(2\pi/9)$. Recordem (pàgina 152) que quan vam parlar dels polígons construïbles amb regle i compàs vam observar que en una construcció amb regle i compàs només podem obtenir punts les coordenades dels quals s'expressin utilitzant sumes, restes, multiplicacions, divisions i arrels quadrades. Com que hem dit que es pot demostrar que el nombre real x_1 no es pot expressar d'aquesta manera, la conclusió és que la trisecció de l'angle és, en general, impossible —si ens restringim, és clar, a les construccions amb regle i compàs.

cúbica. La fórmula final és força complicada i té molt poc interès pràctic. Els problemes que presenta, pel que fa a les arrels *impossibles*, són com els que hem discutit abans. En conclusió, la quàrtica general es pot resoldre per radicals (en els nombres complexos).

La situació, però, canvia «radicalment» a partir del cinquè grau: els esforços per trobar una solució per radicals de l'equació de cinquè grau —la *quíntica*— van ser debades. Finalment, l'any 1824, —gairebé tres-cents anys després de la resolució de la cúbica i la quàrtica!— Niels Henrik Abel va resoldre negativament el problema: va demostrar que l'equació *general*[5] de cinquè grau **no** és resoluble per radicals (ni en els nombres complexos).

Finalment, Évariste Galois va crear una *teoria general de les equacions polinòmiques* i va demostrar que, en grau $n \geq 5$, la resolució per radicals de l'equació general és impossible. Molt probablement, els lectors d'aquest text de fonaments de les matemàtiques estudiaran aquesta teoria més endavant en la seva carrera i, per tant, no és ara el moment d'anar més enllà en aquest tema.

[5] Quan diem que l'equació general de grau n no es pot resoldre per radicals volem dir que si considerem una equació amb coeficients indeterminats $a_n x^n + \cdots + a_0 = 0$, no hi ha cap funció de $a_n, \ldots, a_0$, formada només per les operacions del cos i per arrels, que determini, en tots els casos, una solució de l'equació. Una altra cosa diferent és que per una equació concreta la solució es pugui expressar per radicals o no. Per exemple, les solucions de $x^n - a = 0$ sempre es poden expressar per radicals. Evidentment, si tenim una solució per radicals de l'equació general de grau n (com passa en els casos de grau 2, 3 i 4), necessàriament *totes* les equacions de grau n són resolubles per radicals. Però podria passar que no hi hagués cap solució per radicals de l'equació general i en canvi totes les equacions es poguessin resoldre per radicals. En aquest sentit, Abel també va demostrar que hi ha equacions concretes de cinquè grau que no són resolubles per radicals. Finalment, la teoria de Galois resol completament els dos problemes: no hi ha solució per radicals de les equacions generals de grau > 4 i, per a cada equació concreta, hi ha un criteri per decidir si es pot resoldre per radicals o no.

29 Com fer que qualsevol equació tingui solució i crear nous cossos

Hem vist que el comportament de l'anell de polinomis sobre un cos té unes propietats molt similars a les de l'anell dels enters. Aquesta semblança prové de que en tots dos casos disposem de la **divisió amb residu** i és d'aquesta simple operació que es desprenen les propietats de ser un DIP i un DFU, la identitat de Bézout, l'algorisme d'Euclides, el bon comportament de les congruències o l'existència dels cossos $\mathbb{F}_p$.[1]

En aquest capítol volem traslladar a $k[x]$ la teoria dels anells $\mathbb{Z}/(m)$ que vam estudiar en un capítol anterior. Recordem aquestes propietats de l'anell dels enters:

- Per cada $a \in \mathbb{Z}$ podem considerar la relació d'equivalència *«congruència mòdul a»*. Aleshores, designem per $\mathbb{Z}/(a)$ el conjunt quocient per aquesta relació.
- Els conjunts $\mathbb{Z}/(a)$ tenen una estructura d'anell, de manera que la projecció canònica $\pi : \mathbb{Z} \to \mathbb{Z}/(a)$ és un homomorfisme d'anells (la *reducció mòdul a*).
- Si $a \in \mathbb{Z}$ és primer, l'anell $\mathbb{Z}/(a)$ és, de fet, un cos. La identitat de Bézout ens permet trobar efectivament l'invers de qualsevol element no nul $x \in \mathbb{Z}/(a)$.
- D'aquesta manera, per a cada primer p hem construït un cos amb p elements designat per $\mathbb{F}_p$. Aquests cossos són molt útils: per exemple, si tenim un problema sobre nombres enters, el podem convertir en un problema en cadascun dels cossos finits $\mathbb{F}_p$ a través dels homomorfismes $\pi : \mathbb{Z} \to \mathbb{F}_p$.

Tot això mateix, sense canviar pràcticament ni una coma, ho tenim en l'anell $k[x]$ quan k és un cos. En aquest capítol considerarem que el cos base és el cos dels nombres racionals, és a dir, estudiarem l'anell $\mathbb{Q}[x]$.

[1] La utilitat del paral·lisme entre $\mathbb{Z}$ i $k[x]$ no s'acaba aquí. Probablement, l'estudiant aprendrà més endavant teoremes com la classificació dels grups abelians finitament generats i la classificació dels endomorfismes d'un espai vectorial de dimensió finita: dues versions paral·leles d'un teorema que és vàlid en tots els anells que tenen divisió amb residu.

- Per cada polinomi $p(x) \in \mathbb{Q}[x]$ podem considerar la relació d'equivalència

$$q_1(x) \equiv q_2(x) \;\Leftrightarrow\; q_1(x) = q_2(x) + r(x)p(x) \text{ per algun } r(x) \in \mathbb{Q}[x]$$

  i podem fer el conjunt quocient, que denotarem

$$\mathbb{Q}[x]/(p(x)).$$

- Aquests conjunts quocients tenen una estructura d'anell, de manera que la projecció

$$\mathbb{Q}[x] \to \mathbb{Q}[x]/(p(x))$$

  és un homomorfisme d'anells.

- Si $p(x) \in \mathbb{Q}[x]$ és un polinomi **irreductible**, l'anell $\mathbb{Q}[x]/(p(x))$ és, de fet, un **cos**. En parlarem més endavant.

- La identitat de Bézout ens permet trobar efectivament l'invers de qualsevol element no nul del cos $\mathbb{Q}[x]/(p(x))$.

Estudiem una mica més a fons els nous cossos que podem construir d'aquesta manera. Ho farem amb un exemple.

Un exemple

Considerem el polinomi $Q := x^2 - x - 1 \in \mathbb{Q}[x]$. Clarament, es tracta d'un polinomi irreductible perquè és de segon grau i no té zeros a $\mathbb{Q}$. Per tant, $\mathbb{K} := \mathbb{Q}[x]/(Q)$ és un **cos**. Quines propietats té aquest cos?

- $\mathbb{Q}$ és un *subcòs* de $\mathbb{K}$. És a dir, hi ha un homomorfisme de cossos injectiu[2] $i : \mathbb{Q} \to \mathbb{K}$ donat per $i(q) := [q]$. Podem identificar els nombres racionals amb les seves imatges a $\mathbb{K}$. Direm que $\mathbb{Q} \subseteq \mathbb{K}$ és una *extensió de cossos*. Hem passat del cos $\mathbb{Q}$ a un cos més gran $\mathbb{K}$.

- $\mathbb{K}$ és un subcòs de $\mathbb{R}$. En efecte, considerem l'homomorfisme d'anells $j : \mathbb{Q}[x] \to \mathbb{R}$ definit per

$$j(a_n x^n + \cdots + a_1 x + a_0) := a_n \phi^n + \cdots + a_1 \phi + a_0$$

  on

$$\phi := \frac{1 + \sqrt{5}}{2} \in \mathbb{R}.$$

  Observem que j compleix la propietat necessària i suficient per definir un homomorfisme $j : \mathbb{K} \to \mathbb{R}$ perquè

$$j(x^2 - x - 1) = \phi^2 - \phi - 1 = 0.$$

[2] No ens ha de sorprendre que sigui injectiu perquè, de fet, tot homomorfisme d'un cos en un anell ho és (exercici IV.30).

Tenim, doncs, un homomorfisme[3] injectiu

$$\bar{j} : \mathbb{K} \to \mathbb{R}.$$

És a dir, $\mathbb{K}$ és un cos *intermedi* entre els racionals i els reals

$$\mathbb{Q} \subset \mathbb{K} \subset \mathbb{R}.$$

- Observem que l'equació $x^2 - x - 1 = 0$ no té cap solució al cos $\mathbb{Q}$, **però passa a tenir-ne al cos més gran $\mathbb{K}$!** En efecte, si denotem $\alpha := [x] \in \mathbb{K}$, és evident que
$$\alpha^2 - \alpha - 1 = [x^2 - x - 1] = 0 \in \mathbb{K}.$$
és a dir, hem agafat un polinomi (irreductible) que no tenia cap zero a $\mathbb{Q}$ i hem construït un cos més gran $\mathbb{K} \supset \mathbb{Q}$ en el qual el polinomi sí que té una arrel. Això mateix ho podem fer amb **qualsevol** polinomi irreductible (de grau > 1) de manera que, tal com diu el títol del capítol, podem fer que qualsevol equació (polinòmica, és clar) tingui solució. Magnífic!

- Podem pensar que el que li «faltava» a $\mathbb{Q}$ perquè $x^2 - x - 1 = 0$ tingués solució era simplement el nombre real $\sqrt{5}$ i, efectivament, la diferència entre $\mathbb{Q}$ i el cos més gran $\mathbb{K}$ que hem construït és que $\sqrt{5} \in \mathbb{K}$:
$$(2\alpha - 1)^2 = 4\alpha^2 - 4\alpha + 1 = 4[x^2 - x - 1] + 4 + 1 = 5 \in \mathbb{K}.$$
Per aquest motiu, de vegades s'escriu $\mathbb{K}$ com $\mathbb{Q}(\sqrt{5})$, el cos més petit que conté els nombres racionals i el nombre real $\sqrt{5}$, i diem, de manera informal que $\mathbb{K}$ *«s'ha obtingut a partir de* $\mathbb{Q}$ *adjuntant* $\sqrt{5}$*»*. És un exemple del que es coneixen com a **cossos de nombres**.

El que hem fet amb aquest exemple té un caràcter completament general, amb una excepció: si l'equació amb la que hem començat tampoc no hagués tingut cap arrel a $\mathbb{R}$, aleshores no hauríem pogut incloure el nou cos $\mathbb{K}$ en el cos $\mathbb{R}$.

Un exemple finit

Si fem el mateix procés de l'exemple anterior començant amb un cos finit $\mathbb{F}_p$ en lloc de començar amb el cos $\mathbb{Q}$, també obtindrem nous cossos que ara seran **finits**. De fet, es pot demostrar que tots els cossos finits es poden obtenir per aquest procediment. Estudiem, per exemple, el cas del cos de 4 elements.

Considerem el polinomi $x^2 + x + 1 \in \mathbb{F}_2[x]$ que és clarament un polinomi irreductible (és de segon grau i no té cap arrel). Per tant,

$$K := \mathbb{F}_2[x]/(x^2 + x + 1)$$

[3] Observem el fet remarcable que també podríem definir un segon homomorfisme j' substituint ϕ per $\psi := (1 - \sqrt{5})/2$. Això és molt interessant, però no podem entrar ara a discutir-ho més a fons.

serà un cos. Mirem quants elements té. Tot $\alpha \in K$ prové d'un polinomi, de manera que $\alpha = [x^n + a_{n-1}x^{n-1} + \cdots + a_0]$. Però $[x^2 + x + 1] = 0$ a K, és a dir, $[x^2] = [x + 1] \in K$. D'aquí deduïm que els únics elements de K són

$$0,\ 1,\ [x],\ [x+1].$$

Si ara escrivim la taula de multiplicar de K veurem que es tracta del cos de 4 elements que ja vam estudiar en un capítol anterior (pàgina 147) i que, de fet, és l'únic cos de 4 elements (llevat d'isomorfisme): $K = \mathbb{F}_4$.

La resta de cossos $\mathbb{F}_{p^r}$ de què hem parlat al capítol 21 es construeixen amb aquest mateix mètode.

* * *

Resumim el que hem fet per tal d'adonar-nos millor de la seva importància.

- Hem trobat un mètode per construir nous cossos. En particular, hem trobat el mètode que ens permet construir **tots** els cossos finits.
- Hem demostrat que tot polinomi de $k[x]$ de grau $n > 0$ té n zeros. És possible que alguns d'aquests zeros no siguin a k, però ara estem segurs que sí que són a algun cos una mica més gran que k.
- En particular, podem treballar amb les arrels d'un polinomi amb tota naturalitat perquè ja sabem que totes aquestes arrels són a un cos convenient. La frase *«aquest polinomi no té arrels»* és essencialment incorrecta. Hauríem de dir: *«les arrels d'aquest polinomi, existeixen, però no pertanyen al cos tal»*.
- En certa manera, això que hem fet és una monumental generalització del que havíem fet a la pàgina 124. Quan encara no coneixíem el cos dels racionals, dèiem que l'equació $2x = 3$ no tenia solució. Gràcies a $\mathbb{Q}$, ara sabem que sí que en té: no a $\mathbb{Z}$ però sí a $\mathbb{Q}$. De la mateixa manera, ara ja no podem dir que l'equació $x^2 = -1$ no tingui solució: no en té a $\mathbb{Q}$, però sí que en té a $\mathbb{Q}(\sqrt{-1}) := \mathbb{Q}[x]/(x^2 + 1)$.

Exercicis de polinomis

1. Sigui k un cos finit. Trobeu un polinomi diferent de zero $p(x) \in k[x]$ que doni lloc a la funció polinòmica idènticament zero sobre k.

2. Trobeu el mcd, el mcm i els coeficients de la identitat de Bézout entre aquests dos polinomis de $\mathbb{Q}[x]$:
$$x^3 + x - 1, \quad x^5 + 2x^3 + 1.$$

3. Trobeu tots els polinomis $p \in k[x]$ (k un cos) que compleixen $p(x^2) = p(x)^2$. Tracteu per separat el cas que k tingui característica 2.

4. Sigui $\alpha = \sqrt{2}$, $\beta = (1 + \sqrt{5})/2$, $\alpha, \beta \in \mathbb{R}$. Demostreu que $\alpha + \beta, \alpha\beta \in \mathbb{R}$ són nombres algebraics.

5. Sigui $\alpha \in \mathbb{R}$ un nombre algebraic i considereu $\mathbb{R}$ com a $\mathbb{Q}$-espai vectorial. Considereu el conjunt de totes les combinacions lineals de les potències de α amb coeficients racionals. És a dir.
$$V_\alpha := \{\lambda_0 + \lambda_1\alpha + \cdots + \lambda_n\alpha^n \in \mathbb{R} : \lambda_i \in \mathbb{Q},\ n \geq 0\} \subset \mathbb{R}.$$
Demostreu que V_α és un subespai vectorial de dimensió finita de $\mathbb{R}$. Utilitzeu aquesta idea per demostrar que si $\alpha, \beta \in \mathbb{R}$ són nombres algebraics, aleshores $\alpha + \beta \in \mathbb{R}$ també és un nombre algebraic.

6. Sigui $p(x) \in \mathbb{Q}[x]$ un polinomi mònic de grau $n > 0$. Demostreu que existeix una matriu $n \times n$ amb coeficients a $\mathbb{Q}$ tal que el seu polinomi característic és $(-1)^n p(x)$. (Feu-ho primer per a un polinomi $p(x)$ de grau dos.)

7. Trobeu dos polinomis $p, q \in \mathbb{Z}[x]$ que siguin coprimers però tals que no es pugui escriure la identitat de Bézout entre ells, és a dir, no existeixin polinomis $r, s \in \mathbb{Z}[x]$ tals que $rp + sq = 1$.

8. Demostreu aquestes propietats de la derivació de polinomis:
   (a) *regla del producte*: $(fg)' = f'g + fg'$.
   (b) *regla de la cadena*: $\frac{d}{dx} f(g(x)) = f'(g(x))\, g'(x)$.
   (c) *Primitives*: si el cos de coeficients té característica zero, tot polinomi té alguna primitiva i dues primitives difereixen en una constant.

9. Trobeu totes les solucions a $\mathbb{R}$ d'aquest sistema d'equacions polinòmiques:
$$x^4 + 2x^3 + 4x^2 + 10x - 5 = 0$$
$$x^5 + 2x^4 + 3x^2 + x - 1 = 0$$

10. Demostreu (per inducció sobre $n+m$) que si $n, m > 0$, aleshores

$$\mathrm{mcd}(x^n - 1, x^m - 1) = x^d - 1 \text{ amb } d = \mathrm{mcd}(n, m).$$

11. *Interpolació de Lagrange.* Sigui k un cos i siguin $a_0, \ldots, a_n \in k$, tots diferents. Trobeu un polinomi $p(x)$ de grau n tal que $p(a_i) = 0$ per $i = 1, \ldots, n$ i $p(a_0) = 1$. Siguin $b_0, \ldots, b_n \in k$. Trobeu un polinomi $P(x)$ de grau $\leq n$ tal que $P(a_i) = b_i$ per $i = 0, \ldots, n$.

12. Trobeu un polinomi de $\mathbb{Q}[x]$ que sigui congruent amb $-x^2+2x+1$ mòdul x^3+x^2+x+1 i sigui congruent amb $3x^3 - 4x + 2$ mòdul $x^4 - x^2 + 1$.

13. Siguin $p_1, p_2, p_3 \in \mathbb{R}[x]$ tals que $xp_1^2 = p_2^2 - xp_3^2$. Demostreu que $p_1, p_2, p_3 = 0$. Trobeu un contraexemple sobre un cos finit.

14. Direm que $p \in \mathbb{Q}[x]$ és *lliure de quadrats* si no hi ha cap polinomi no constant q tal que q^2 divideix p. Demostreu:

    (a) p és lliure de quadrats si i només si p es pot escriure com a producte de polinomis irreductibles diferents.

    (b) Si p, p' (la derivada de p) són coprimers, aleshores p és lliure de quadrats.

    (c) Recíprocament, si p és lliure de quadrats, aleshores p, p' són coprimers.

    (d) Tot polinomi p es pot descompondre com

$$p = q_1 q_2^2 q_3^3 \cdots q_n^n$$

    on els polinomis q_i, $1 \leq i \leq n$ són lliures de quadrats.

15. Trobeu tots els zeros del polinomi $x^2 - 1$ a $\mathbb{Z}/(16)$. Trobeu alguna contradicció amb els resultats que hem demostrat sobre $k[x]$?

16. Demostreu aquesta igualtat vàlida a $\mathbb{F}_p[x]$:

$$x^p - x = (x-1)\cdots(x-p).$$

    Apliqueu aquest resultat a demostrar la *congruència de Wilson*: Si p és primer, $(p-1)! \equiv -1$ mòdul p.

17. Trobeu tots els polinomis mònics irreductibles de grau 2 i 3 de $\mathbb{F}_3[x]$.

18. Els *polinomis de Txebixov* són polinomis $T_n \in \mathbb{Z}[x]$, $n \geq 0$ que es defineixen de manera recursiva amb aquestes fórmules

$$T_0 = 1, \quad T_1 = x, \quad T_{n+1} = 2xT_n - T_{n-1}.$$

    (a) Demostreu, aplicant les fórmules trigonomètriques, que $T_n(\cos\theta) = \cos(n\theta)$ per tot $n \geq 0$.

    (b) Demostreu que T_n és solució de l'equació diferencial $(1-x^2)y'' - xy' + n^2 y = 0$. (Feu el canvi de variable $x = \cos t$.)

    (c) Demostreu que $T_n(T_m(x)) = T_{nm}(x)$.

19. Considereu polinomis sobre un cos k de característica zero.

    (a) Demostreu que si p, q són polinomis tals que $p' = q'$ i existeix $a \in k$ tal que $p(a) = q(a)$, aleshores $p = q$.

    (b) Demostreu (per inducció) la *fórmula de Taylor*: si p és un polinomi de grau n, aleshores per tot $a \in k$ es compleix

    $$p = p(a) + \frac{p'(a)}{1!}(x-a) + \cdots + \frac{p^{(n)}(a)}{n!}(x-a)^n.$$

20. Sigui $P \in \mathbb{F}_p[x]$ un polinomi de grau $n > 0$ tal que $P' = 0$. Demostreu:

    (a) El grau de P és divisible per p.

    (b) Si a és un zero de P, la multiplicitat de a és un múltiple de p.

    (c) El nombre màxim de zeros de P és n/p.

21. Demostreu que a $\mathbb{Q}[x]$ hi ha polinomis irreductibles de tots els graus > 0.

22. Escriviu aquests polinomis de $\mathbb{Q}[x]$ com a producte de polinomis irreductibles:

    (a) $x^9 + 15x^5 + 30x^2 + 45$.

    (b) $x^5 + x^4 + 4x^2 + 6x + 2$.

    (c) $x^5 - 2x^4 - x^2 + 4$.

23. Demostreu que el polinomi $x^4 + 4x + 1$ és irreductible sobre $\mathbb{Q}$. (Indicació: feu un canvi de variable $y = x + a$.)

24. Considereu els polinomis $P_n = 1 + x + x^2 + \cdots + x^n \in \mathbb{Q}[x]$ per $n > 0$. Trobeu tots els seus zeros a $\mathbb{Q}$ i determineu també la multiplicitat d'aquests zeros.

25. (a) Sigui $P(x) \in \mathbb{Z}[x]$ i sigui p un primer que no divideixi el coeficient del terme de grau màxim de $P(x)$. Demostreu que si $P(x)$ és irreductible a $\mathbb{F}_p[x]$, aleshores $P(x)$ és irreductible a $\mathbb{Z}[x]$.

    (b) Utilitzeu la reducció mòdul 2 per demostrar que el polinomi $x^4 - x^3 + x^2 - x + 1$ és irreductible a $\mathbb{Z}[x]$. Què podem dir a $\mathbb{Q}[x]$?

    (c) Trobeu la descomposició de $x^{10} - 1 \in \mathbb{Q}[x]$ com a producte de polinomis irreductibles. (Utilitzeu l'apartat anterior.)

26. Considereu aquest polinomi de $\mathbb{Q}[x]$:

    $$P = x^6 + 6x^5 + 15x^4 + 20x^3 + 3x^2 - 18x + 5.$$

    (a) Mireu si P té algun zero real doble.

    (b) Trobeu tots els zeros reals de P.

    (c) Trobeu la descomposició de P en polinomis irreductibles a $\mathbb{Q}[x]$.

27. Apliqueu el mètode de Cardano per trobar una arrel real de $x^3 + 3x^2 + 4x + 3$. Comproveu que les altres dues arrels no són reals.

28. El *discriminant* de l'equació de tercer grau $Q_3 := x^3 + px + q$ es defineix com $\Delta := -4p^3 - 27q^2$. Suposem que α, β, γ són les tres arrels de Q_3 en algun cos de característica zero. Demostreu que

$$\Delta = (\alpha - \beta)^2(\alpha - \gamma)^2(\beta - \gamma)^2.$$

Suposeu ara que el cos és el cos dels reals i que les tres arrels són diferents. Demostreu que la fórmula de Cardano que troba un zero de Q_3 involucra arrels quadrades que no es poden resoldre al cos real. (És a dir, la fórmula de Cardano només expressa una arrel real completament dins les operacions reals quan la cúbica té una única arrel real.)

29. Suposem que volem trobar els zeros de l'equació de quart grau $P_4 = x^4 + cx^2 + dx + e$. Aquest és un mètode:

(a) Escriviu P_4 com a producte de dues equacions de segon grau amb coeficients indeterminats

$$P_4 = x^4 + cx^2 + dx + e = (x^2 + px + q)(x^2 - px + r).$$

N'hi ha prou amb trobar els valors de p, q, r per tenir resolt el problema. Expresseu c, d, e en funció de p, q, r.

(b) Demostreu $(c + p^2)^2 - (d/p)^2 = 4e$.

(c) escriviu $y := p^2$ i comproveu que y compleix aquesta equació de tercer grau:

$$y^3 + 2cy^2 + (c^2 - 4e)y - d^2 = 0.$$

(d) Resolent aquesta cúbica podem determinar p, q, r i resoldre el problema.

(e) Apliqueu-ho a resoldre $x^4 - 5x^2 - 2x + 3 = 0$.

30. Considereu aquests anells

$$A := \mathbb{Q}[x]/(x^4 + 4x + 1),\ B := \mathbb{Q}[x]/(x^3 + 3x^2 + 4x + 3),\ C := \mathbb{Q}[x]/(x^4 - 5x^2 - 2x + 3).$$

Demostreu que A i B són cossos, però C no ho és. Demostreu que hi ha un únic homomorfisme de cossos $j : B \to \mathbb{R}$. (Utilitzeu els exercicis 23, 27, 29.)

31. Designem α l'arrel real positiva de $x^2 + x - 1$. Demostreu:

(a) Si $n \geq 0$, es compleix $\alpha^n = u_n + v_n\alpha$ on els enters u_n, v_n estan determinats per la fórmula recursiva

$$u_{n+1} = v_n, \quad v_{n+1} = u_n - v_n, \quad u_0 = 1, \quad v_0 = 0.$$

(b) Sigui $\varphi := \alpha^{-1}$. Demostreu $\varphi = 1 + \alpha$ i comproveu que φ és l'arrel real positiva de $x^2 - x - 1$.[4]

(c) Trobeu una fórmula recursiva per expressar les potències negatives de α en la forma $\alpha^{-n} = u'_n + v'_n\alpha$.

(d) Demostreu les identitats $(1 + \alpha)(1 - \alpha) = \alpha$ i $\alpha^3(3 + 2\alpha) = 1$.

[4] És a dir, φ és la *raó àuria*.

(e) Utilitzeu la teoria de les progressions geomètriques per demostrar aquesta igualtat:

$$\sum_{i\in\mathbb{Z}} \alpha^{|i|} = 3 + 2\alpha.$$

32. (a) Sigui $\alpha \in \mathbb{R}$ com a l'exercici 31. Si A és un subconjunt finit de $\mathbb{Z}^2$, podem associar a A aquest nombre real:

$$f(A) := \sum_{(i,j)\in A} \alpha^{|i|+|j|} \in \mathbb{R}.$$

Demostreu aquestes propietats de la funció f:

i. Sigui F_j la fila j de $\mathbb{Z}^2$, és a dir, $F_j = \{(i,j) : i \in \mathbb{Z}\}$. Calculeu $f(A_j)$.

ii. Sigui $j_0 \geq 0$ i sigui B_{j_0} el subconjunt de $\mathbb{Z}^2$ format per totes les files j amb $j \geq j_0$. Demostreu

$$f(B_{j_0}) = \alpha^{j_0-1}(5 + 3\alpha).$$

Demostreu $f(B_5) = 1$.

iii. Donats $i, j \in \mathbb{Z}$, considereu aquests subconjunts de $\mathbb{Z}^2$:

$$A = \{(i,j),(i+1,j)\},\ A' = \{(i+2,j)\};$$

$$B = \{(i+1,j),(i+2,j)\},\ B' = \{(i,j)\}.$$

Demostreu $f(A) \geq f(A')$, $f(B) \geq f(B')$.

(b) (**Les Dames de Conway**) Imaginem un tauler com el de jugar a escacs o a les dames, de mida arbitràriament gran. A cada casella del tauler hi pot haver una fitxa, o no haver-n'hi cap. Les fitxes es poden desplaçar horitzontalment o verticalment, cap a la dreta o cap a l'esquerra, amunt o avall, sempre que ho facin *«saltant»* sobre una altra fitxa contigua, que desapareix del tauler. Hi pot haver una quantitat il·limitada de fitxes.

En la posició inicial les fitxes estan totes per sota d'una línia horitzontal —la *«frontera»*— i l'objectiu del joc és aconseguir que alguna fitxa, seguint les normes del joc, arribi a una fila el més allunyada possible de la frontera.

És trivial aconseguir que una fitxa avanci fins la primera o la segona fila més enllà de la frontera. És més difícil aconseguir dur una fitxa a la tercera o quarta fila més enllà de la frontera. El **teorema de Conway** afirma que és impossible que cap fitxa arribi fins la cinquena fila.

Es tracta de demostrar aquest resultat de Conway. Per fer-ho, suposeu que ha arribat una fitxa a la cinquena fila i situeu l'origen de coordenades a la casella on hi hagi aquesta fitxa. Utilitzeu l'apartat anterior per associar a cada posició del joc un nombre real —diguem-ne el *valor de la posició*. Observeu que el valor de la posició a l'inici del joc és < 1 i que, en cada moviment del joc, el valor de la posició no pot créixer. Observeu, finalment, que una fitxa a la casella $(0,0)$ al final del joc implica que el valor de la posició final és ≥ 1, una contradicció.

33. Sigui K un cos i sigui G un subgrup finit de K^*, és a dir, G està format per elements $\neq 0$ de K i és tancat per productes i pas a l'invers. Es tracta de demostrar que G és un grup cíclic (vegeu l'exercici **III**.22). Seguiu aquests passos

(a) Sigui n l'ordre de G i sigui d un divisor de n. Demostreu que el nombre d'elements de G d'ordre d és 0 o $\varphi(d)$. Utilitzeu l'exercici **IV**.12 i el fet que sobre un cos un polinomi d'ordre d no pot tenir més de d zeros.

(b) Apliqueu la fórmula de Gauss de la pàgina 150 i acabeu la demostració del teorema.

(c) Observeu que, con a corol·lari, el grup multiplicatiu de qualsevol cos finit $\mathbb{F}_p$ és un grup cíclic d'ordre $p-1$.

Part VI:

Els nombres complexos

En les parts I i II vam explicar quins eren els (complicats) fonaments dels nombres naturals. Més endavant, vam poder definir amb exactitud i amb una relativa facilitat els nombres enters i els nombres racionals. El pas següent en l'ampliació del concepte de nombre consisteix en construir —o definir axiomàticament— el cos dels nombres reals. Però aquesta construcció torna a ser difícil i problemàtica —com ho va ser la dels nombres naturals— i hem preferit deixar-la fora d'aquest text sobre els fonaments de les matemàtiques.

Encara més enllà dels reals hi ha el cos dels nombres complexos i, com que el pas dels reals als complexos torna a ser, com els dels naturals als enters o el dels enters als racionals, relativament senzill, és escaient incloure en aquest text de fonaments una brevíssima introducció —més conceptual que no pas tècnica— sobre el cos $\mathbb{C}$.

Tancarem aquest llibre amb una demostració —necessàriament incompleta, però molt acurada— de la famosíssima fórmula

$$e^{i\pi} + 1 = 0.$$

Foto: Bernhardt Riemann, 1826–1866.

30 | Tres definicions dels nombres complexos

Primera: una multiplicació a l'espai vectorial $\mathbb{R}^2$

La primera definició dels nombres complexos que donarem és la més elemental i la més incomprensible. No és incomprensible en el sentit que sigui difícil d'entendre sinó en el sentit que no veiem d'on surt ni com és possible. L'hem de considerar *elemental* perquè pràcticament no requereix cap coneixement previ ni utilitza cap estructura sofisticada.

Considerem l'espai vectorial $\mathbb{C}_1 := \mathbb{R}^2$ que, en particular, és un grup abelià amb l'operació de suma de vectors. Els vectors de $\mathbb{C}_1$ són parelles de nombres reals: $\overrightarrow{v} = (a, b)$, $a, b \in \mathbb{R}$. En general, no hi ha cap concepte de multiplicació de vectors en un espai vectorial $\mathbb{R}^n$. És cert que hi ha l'anomenat *producte escalar* de vectors, però el resultat d'un producte escalar no és un vector, sinó que és un escalar. També és cert que a $\mathbb{R}^3$ (i només en dimensió 3) tenim el curiós *producte vectorial* que, efectivament, ens permet multiplicar dos vectors i obtenir un tercer vector, però les propietats algebraiques d'aquesta multiplicació de vectors en dimensió 3 no són gaire satisfactòries: hi ha divisors de zero i la multiplicació no és associativa.

Malgrat tot això, en dimensió dos sí que podem definir una —sorprenent— multiplicació de vectors que té bones —excel·lents!— propietats algebraiques. De fet, com veurem, té propietats molt millors del que podríem imaginar.

Aquesta multiplicació es defineix així:

$$(a, b)\,(c, d) := (ac - bd, ad + bc).$$

Per què la definim així? No ho sabem —de moment!— i és per això que hem dit que aquesta definició dels complexos és *incomprensible*. Estudiem ara les propietats d'aquesta multiplicació.

- És **commutativa**. La comprovació és molt senzilla.
- És **associativa**. Això es pot comprovar fent un càlcul llarg i avorrit que no farem.

- És **distributiva** respecte de la suma. Novament, un càlcul avorrit i trivial ens permet comprovar aquesta propietat.
- $(1,0) \in \mathbb{C}_1$ actua com a **element neutre**.
- Per tot això, $\mathbb{C}_1$ és un **anell** commutatiu amb unitat.
- L'aplicació $a \mapsto (a,0) \in \mathbb{C}_1$ és un homomorfisme d'anells injectiu que ens permet identificar el cos $\mathbb{R}$ com a subanell de l'anell $\mathbb{C}_1$. Per abús de llenguatge escriurem $a \in \mathbb{C}_1$ quan ens referim a $(a,0) \in \mathbb{C}_1$.
- Definim $i \in \mathbb{C}_1$ com l'element $i := (0,1)$. Aleshores, els elements $1, i \in \mathbb{C}_1$ formen una **base** de l'espai vectorial $\mathbb{C}_1$: tot $v \in V$ s'expressa de manera única com combinació lineal $v = a + bi$ amb $a, b \in \mathbb{R}$. Diem que a és la *part real* de v i b és la *part imaginària* de v.
- Per la manera com hem definit la multiplicació tenim la igualtat $i^2 = -1$.
- Com que $\mathbb{C}_1 = \mathbb{R}^2$, podem considerar el **mòdul** dels elements de $\mathbb{C}_1$

  $$||a + bi|| = ||(a,b)|| := \sqrt{a^2 + b^2} \in \mathbb{R}.$$

  Aquest mòdul té un bon comportament respecte de la multiplicació que hem definit:

  $$||uv|| = ||u||\,||v|| \text{ per tot } u, v \in \mathbb{C}_1.$$

  El mòdul de $\mathbb{C}_1$ generalitza el valor absolut de $\mathbb{R}$ i, si ho preferim, el podem escriure amb una única barra vertical: $|v|$.
- L'aplicació $\mathbb{C}_1 \to \mathbb{C}_1$ donada per $(a,b) \mapsto (a,-b)$ és un isomorfisme d'anells. S'anomena la **conjugació** i es denota $u \mapsto \overline{u}$. És fàcil veure que la conjugació compleix aquestes propietats: per tot $u, v \in \mathbb{C}_1$
  - $\overline{\overline{u}} = u$.
  - $\overline{uv} = \overline{u}\,\overline{v}$.
  - $|u| = |\overline{u}|$.
  - $u\,\overline{u} = |u|^2 \in \mathbb{R}$.
  - $u + \overline{u} \in \mathbb{R}$.
  - $u = \overline{u}$ si i només si $u \in \mathbb{R}$.
- Tot $u \in \mathbb{C}_1$, $u \neq 0$, té **invers** multiplicatiu. Efectivament,[1]

  $$u \left(\frac{\overline{u}}{|u|^2} \right) = 1.$$

[1] Podria semblar que aquí estem caient en una *petitio principii*: per demostrar que hi ha inversos multiplicatius utilitzem una divisió, és a dir, una multiplicació per un invers multiplicatiu. Observem, però, que estem dividint per $|u|^2$, que sabem que és un nombres real. Com que $\mathbb{C}_1$, per construcció, és l'espai vectorial $\mathbb{R}^2$, sempre podem dividir qualsevol vector per qualsevol *escalar* no nul.

És a dir: $\mathbb{C}_1$ és un **cos** que conté el cos dels nombres reals. Tenim

$$\mathbb{R} \subsetneq \mathbb{C}_1$$

i aquest nou cos $\mathbb{C}_1$ que hem construït té la curiosa propietat que els nombres reals negatius tenen arrels quadrades. De fet, com veurem, té moltes altres propietats encara més extraordinàries!

Segona: adjuntar $\sqrt{-1}$ als reals

La segona definició del nombres complexos és la més avançada, la més natural i la més breu. És *avançada* perquè utilitza estructures algebraiques i conceptes importants que hem estudiat al llarg d'aquest curs; és *natural* perquè es veu clarament què volem fer, què fem i per què ho fem; és *breu* perquè, com veurem, el fet que els nombres complexos formin un cos serà una conseqüència trivial de la definició.

Al capítol 29 hem vist com podem aconseguir que una equació que no tingui solució en un cos passi a tenir-ne en un cos més gran. Què succeeix si apliquem aquell mètode al cos $\mathbb{R}$ i a l'equació $x^2 + 1 = 0$? Doncs que obtenim un cos $\mathbb{C}_2$. Ras i curt!

$$\mathbb{C}_2 := \mathbb{R}(\sqrt{-1}) := \mathbb{R}[x]/(x^2 + 1).$$

Pràcticament no hi res a demostrar: hem definit un cos més gran que $\mathbb{R}$ on hi ha un element $i := [x]$ que té la propietat que $i^2 = -1$.

En tot cas, hauríem de demostrar que aquesta segona definició —tan senzilla, natural i lògica, si la comparem amb la primera— defineix realment el mateix cos de l'apartat anterior. L'isomorfisme entre el cos $\mathbb{C}_1$ que hem definit abans i aquest cos $\mathbb{C}_2$ que acabem de definir ara és clar:

$$(a, b) \mapsto [a + bx] \in \mathbb{R}[x]/(x^2 + 1).$$

Tercera: unes matrius que formen un cos

En la primera definició dels nombres complexos hem hagut d'**inventar** una multiplicació a $\mathbb{R}^2$. En la segona definició, hem **creat** $\mathbb{C}$ afegint a $\mathbb{R}$ un nou element. En aquesta tercera definició, en canvi, **trobarem** $\mathbb{C}$ en un lloc ja conegut, sense que ens calgui inventar res nou: les matrius quadrades 2×2.[2]

[2] Aquesta tercera definició ens fa pensar en la trama de la coneguda narració *The Purloined Letter* d'Edgar Allan Poe: la policia regira minuciosament i infructuosament la residència del suposat lladre buscant un document valuosíssim que ha estat robat; el detectiu Auguste Dupin troba el document a l'interior d'una carta que ha estat tota l'estona a la vista de tothom.

Les matrius quadrades es poden sumar i multiplicar, de manera que tenen una estructura d'anell. Però aquest anell està molt lluny de ser un cos. Hi ha divisors de zero

$$\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

la multiplicació no és commutativa

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

i moltes matrius no tenen inversa

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ és impossible.}$$

Però, curiosament, hi ha un subconjunt de matrius quadrades 2×2 que no té cap d'aquests problemes. Definim

$$\mathbb{C}_3 := \left\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} : a, b \in \mathbb{R} \right\}.$$

És molt senzill veure que aquest conjunt de matrius té aquestes propietats

- És un subanell de l'anell de totes les matrius: la resta i el producte de dues matrius de $\mathbb{C}_3$ és una matriu de $\mathbb{C}_3$ i la matriu identitat és a $\mathbb{C}_3$.
- Totes les matrius de $\mathbb{C}_3$ diferents de zero tenen determinant diferent de zero i, per tant, són invertibles.
- Per tant, $\mathbb{C}_3$ és un cos. A més, l'aplicació

  $$a \mapsto \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$$

  és un homomorfisme de cossos de $\mathbb{R}$ a $\mathbb{C}_3$.

Només ens caldria veure que aquest nou cos $\mathbb{C}_3$ que hem definit ara és isomorf als que hem definit als apartats anteriors. Això és trivial perquè l'aplicació

$$(a, b) \mapsto \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

ens dóna un isomorfisme de cossos entre el cos $\mathbb{C}_1$ que hem definit al primer apartat i el cos $\mathbb{C}_3$ que acabem de definir ara. En particular, el *«misteriós»* o *«imaginari»* nombre i és aquesta matriu

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

que no té res de misteriosa ni imaginària i que, clarament, el seu quadrat és -1 (és a dir, la matriu $-I$).[3]

* * *

Tenim, doncs, tres cossos isomorfs i cadascun d'ells podem dir que és **el cos dels nombres complexos**, denotat per la lletra $\mathbb{C}$. Podem utilitzar la definició que ens sigui més útil en cada moment, però la definició $\mathbb{C}_1$ és la més habitual.

[3] Algú no pot imaginar una cosa que el seu quadrat sigui -1? Que pensi en una rotació de 90 graus!

31 El teorema fonamental de l'àlgebra

Quan vam afegir als nombre reals l'arrel quadrada de -1 per tal que totes les equacions de segon grau tinguessin solució difícilment podíem imaginar que aquesta maniobra tindria uns efectes immensament superiors als esperats: aconseguir que **totes** les equacions polinòmiques tinguin solució! Aquest fet tan important es coneix com a **teorema fonamental de l'àlgebra**:[1]

Tot polinomi no constant de $\mathbb{R}[x]$ *té un zero a* $\mathbb{C}$.

En aquestes notes no podem demostrar aquest teorema —del que s'han trobat moltíssimes demostracions diferents, cap d'elles completament elemental. Observem, en primer lloc, que en aquest teorema els nombres reals hi juguen un paper decisiu: no és cert, de cap manera, que si en cos qualsevol totes les equacions de segon grau tenen solució, també n'hagin de tenir totes les altres equacions polinòmiques. Per aquest motiu, és clar que en les demostracions del teorema caldrà utilitzar alguna propietat dels nombres reals basada en la continuïtat — per exemple, el *teorema de Bolzano* que implica, per exemple, que tot polinomi real de grau senar ha de tenir alguna arrel real— però també caldrà utilitzar eines pròpies de l'anàlisi complexa o de la topologia que estan més enllà dels temes elementals d'aquest text de Fonaments. Admetem, doncs, aquest teorema sense demostració i passem a explicar algunes de les seves conseqüències.

- *Tot polinomi no constant de* $\mathbb{C}[x]$ *té un zero a* $\mathbb{C}$. Considerem un polinomi no constant $p(x) \in \mathbb{C}[x]$ i considerem el polinomi $q(x) := p(x)\,\overline{p}(x)$ on $\overline{p}(x)$ indica el polinomi que hem obtingut a partir de $p(x)$ substituint cada coeficient pel seu conjugat. Les propietats de la conjugació ens asseguren que $q(x)$ és un polinomi no constant de $\mathbb{R}[x]$ (exercici **VI**.7) i, segons el teorema fonamental, tindrà un zero $z \in \mathbb{C}$. Aleshores, o bé $p(z) = 0$ o bé $\overline{p}(z) = 0$. En el primer cas ja hem trobat un zero de $p(x)$ i en el segon cas també perquè $\overline{p(\overline{z})} = \overline{p}(z) = 0$ i, per tant, $p(\overline{z}) = 0$.

[1] És un nom excessivament ampul·lós —no és tan *fonamental* i, de fet, ni tan sols és un un teorema estrictament d'àlgebra (perquè a la construcció de $\mathbb{C}$ hi intervenen els nombres reals)— però l'expressió *«teorema fonamental de l'àlgebra»* va fer fortuna en algun moment de la història de les matemàtiques i així és com el seguim anomenant ara.

- *Tot polinomi mònic $p(x) \in \mathbb{C}[x]$ de grau $n > 0$ es pot expressar com a producte de n polinomis de primer grau*

$$p(x) = (x - z_1)(x - z_2) \cdots (x - z_n).$$

És a dir, els únics polinomis irreductibles de $\mathbb{C}[x]$ són els de primer grau i tot polinomi de grau $n > 0$ de $\mathbb{C}[x]$ té exactament n zeros, comptats amb les seves multiplicitats. Quan un cos té aquesta propietat es diu que és un **cos algebraicament tancat**.

- *Els únics polinomis irreductibles mònics de grau > 1 de $\mathbb{R}[x]$ són els de la forma*

$$x^2 + bx + c \text{ amb } b^2 - 4c < 0.$$

És a dir, tot polinomi mònic no constant $p(x) \in \mathbb{R}[x]$ s'escriu de manera única (llevat de l'ordre) com

$$p(x) = (x - a_i) \cdots (x - a_r)(x^2 + b_1 x + c_1) \cdots (x^2 + b_s x + c_s),$$

amb $a_i, b_i, c_i \in \mathbb{R}$, $b_i^2 - 4c_i < 0$. Per demostrar això, sigui $p(x) \in \mathbb{R}[x]$ un polinomi mònic no constant. Si el considerem com a polinomi de $\mathbb{C}[x]$ el podrem descompondre en monomis de primer grau

$$p(x) = (x - z_1)(x - z_2) \cdots (x - z_n).$$

Observem que, com que $p(x) \in \mathbb{R}[x]$, tenim $\overline{p}(x) = p(x)$. Aleshores

$$p(x) = \overline{p}(x) = (x - \overline{z}_1)(x - \overline{z}_2) \cdots (x - \overline{z}_n).$$

Per tant, per cada $i = 1, \ldots, n$ tindrem $\overline{z}_i = z_j$ per algun j. Distingim dos casos. Si $\overline{z}_i = z_i$, aleshores $z_i \in \mathbb{R}$. En cas contrari, si $\overline{z}_i = z_j \neq z_i$, aleshores $z_i, z_i \notin \mathbb{R}$ i

$$(x - z_i)(x - z_j) = (x - z_i)(x - \overline{z}_i) = x - (z_i + \overline{z}_i)x + z_i\overline{z}_i.$$

Recordem ara que, per les propietats de la conjugació, $z_i + \overline{z}_i$ i $z_i\overline{z}_i$ són nombres reals i, per tant, el polinomi anterior és un polinomi de segon grau amb coeficients reals, sense zeros reals. Això demostra que, efectivament, $p(x)$ es pot descompondre a $\mathbb{R}[x]$ de la manera indicada.

- *El «Nullstellensatz» de Hilbert.* L'existència de solucions per a totes les equacions polinòmiques (no constants) sobre el cos $\mathbb{C}$ va ser generalitzada per David Hilbert als *sistemes d'equacions polinòmiques en diverses variables*

$$\begin{cases} p_1(x_1, \ldots, x_n) = 0 \\ \cdots\cdots\cdots \\ p_k(x_1, \ldots, x_n) = 0 \end{cases}$$

en un famós teorema que sovint s'anomena amb el seu nom en alemany: *Nullstellensatz* —literalment, el «teorema de col·locació dels zeros».

Considerem, doncs, una família (també pot ser infinita) de polinomis complexos no constants en diverses variables. El teorema afirma que, sota una condició necessària evident, existeix un punt $(x_1, \ldots, x_n) \in \mathbb{C}^n$ on tots aquests polinomis s'anul·len. La condició necessària és aquesta: és impossible escriure

$$1 = h_1 p_1 + \cdots + h_k p_k$$

on $h_1, \ldots, h_k \in \mathbb{C}[x_1, \ldots, x_n]$.

- *El teorema de Bézout sobre els punts d'intersecció de dues corbes*. Un polinomi de grau n en dues variables $p(x, y)$ defineix el que es coneix com una *corba algebraica plana* de grau n. Per exemple, $x^2 + y^2 - 1$ és la circumferència unitat del pla, $(x^2 + y^2)^2 - 2(x^2 - y^2)$ és la *lemniscata de Bernoulli* i $x^3 + y^3 - 3xy$ és el *foli de Descartes*. Aleshores, el teorema afirma que, en el pla $\mathbb{C}^2$, una corba algebraica $p(x, y)$ de grau n i una corba algebraica $q(x, y)$ de grau m es tallen exactament en nm punts, si admetem unes condicions que ara explicarem.

  En primer lloc, el teorema no pot ser cert si els dos polinomis tenen un divisor comú de grau > 1, perquè en aquest cas, tots els punts de la corba definida per aquest divisor serien punts d'intersecció, i n'hi ha infinits.

  En segon lloc, cada punt d'intersecció s'ha de comptar amb la seva *multiplicitat*, que és un concepte similar al de multiplicitat d'un zero. Per exemple, és clar que la recta $x = 1$ i la circumferència $x^2 + y^2 = 1$ tenen un únic punt en comú: el punt $(1, 0)$, però aquest punt, en un sentit que es pot formalitzar matemàticament, *«compta»* com dos.

  Finalment, si apliquem aquest teorema a dues rectes del pla tindrem que s'han de tallar en un punt, però si les rectes són paral·leles, no es tallen. El que succeeix aquí és que el teorema ha d'incloure també els *punts de l'infinit* on es tallen les rectes paral·leles, uns punts que estan presents en el concepte d'*espai projectiu*.[2]

En conclusió, les propietats geomètriques del cos $\mathbb{C}$ són molt més senzilles que les del cos real —i no diguem les del cos racional $\mathbb{Q}$!— i això fa que $\mathbb{C}$ sigui l'àmbit idoni on estudiar els fenòmens de la *geometria algebraica* —fins i tot en el cas que només ens interessin els punts amb coordenades reals![3]

[2] Pot semblar difícil de creure que, per exemple, les circumferències $x^2 + y^2 = 1$ i $x^2 + y^2 = 2$ es tallin en quatre punts, tal com afirma el teorema de Bézout. De fet, aquestes dues circumferències tenen en comú dos punts dobles a l'infinit: $\{1, i, 0\}$ i $\{-1, i, 0\}$.

[3] Aquí tenim un exemple excel·lent d'una de les grans idees de les les matemàtiques: sovint, una situació complicada en un cert àmbit esdevé comprensible i natural quan la miren en un àmbit més ampli. El cas del teorema de Bézout és especialment paradigmàtic. Si ens interessen els punts d'intersecció d'una corba de grau 3 amb una corba de grau 4 (irreductibles), el teorema ens diu que n'hi ha exactament 12, sempre que ens mirem el problema a l'espai projectiu complex i comptem cada punt segons la seva multiplicitat. Si, dit això, ens interessen només els punts reals que no estiguin a l'infinit, és clar que d'aquells 12 punts només en «veurem» uns quants però és molt més simple pensar que els 12 punts hi són, encara que alguns estiguin en un lloc que, pel que sigui, ara no ens interessa, que no pas obstinar-se en voler fer una teoria en la que alguns punts de tall puguin «no existir».

32 La fórmula més bella: $e^{i\pi} + 1 = 0$

Es diu que la *fórmula d'Euler*

$$e^{i\pi} + 1 = 0$$

és la més bonica de les fórmules matemàtiques: hi apareixen els que en podríem dir els cinc nombres més importants (0, 1, e, π, i) i les tres operacions fonamentals (suma, producte i exponenciació), relacionats d'una manera gens trivial. Tanmateix, aquesta fórmula és un cas particular de la fórmula general

$$e^{a+ib} = e^a(\cos(b) + i\sin(b)).$$

En aquest capítol final volem demostrar aquesta fórmula, però ha de quedar clar, abans de res, que el context en què ens movem en aquest text és insuficient per demostrar la fórmula d'Euler —fins i tot és insuficient per enunciar-la. Efectivament, les definicions de e, de π i de la funció exponencial requereixen el concepte de pas al límit i el seu àmbit natural és el del *càlcul infinitesimal*. Per tant, només amb les eines del càlcul podem plantejar-nos demostrar la fórmula d'Euler d'una manera completa.

Malgrat tot això que diem, donarem una demostració de la fórmula d'Euler el més completa possible i, especialment, evitarem qualsevol mistificació o qualsevol argument heurístic.[1]

La demostració de la fórmula d'Euler —en la seva versió general— que explicarem a continuació conté una exquisida barreja de geometria, àlgebra i càlcul diferencial i integral. Comencem amb la part més geomètrica de la demostració.

[1] La fórmula d'Euler i la trigonometria –en la qual es fonamenta la fórmula— són eines matemàtiques molt importants. En conseqüència, conceptes com *sinus, cosinus* o *radiant*, juntament amb les seves propietats més bàsiques, són temes amb els que tots els alumnes de secundària hi estan familiaritzats. Tanmateix, la fonamentació rigorosa d'aquests conceptes no és gens trivial i, en conseqüència, el tractament que se'n fa fora dels estudis de matemàtiques és heurístic, de manera que no és apropiat per a un text com aquest. Donem un parell d'exemples: quan es defineix l'angle d'un grau com el resultat de dividir la circumferència en 360 parts iguals, *què volem dir quan diem parts iguals?*; si el que es defineix és el sinus d'un angle, *com és que després es parla del sinus d'un nombre?* La conclusió de tot això és que hem de demanar ara al lector d'aquest text que faci un exercici de *dubte cartesià* i deixi de banda tot el que sap d'angles i funcions trigonomètriques, com a pas previ per començar a fonamentar aquests conceptes sobre uns principis sòlids.

Angles

Hem de començar definint clarament què entenem per *angle*. Hi ha diverses definicions vàlides i ens decantem, en el context actual, per aquesta:

> *Un **angle** és una parella ordenada de rectes que es tallen en un únic punt.*

És a dir, entenem que dues rectes que es tallen en un punt formen un angle (observem que amb aquesta definició no tenim ni l'angle «pla» ni l'angle «nul»). També podríem dir que en formen dos o quatre però, entre les diverses definicions possibles d'angle, escollim aquesta. Observem que es tracta d'una definició purament *geomètrica*: un angle son dues rectes, no pas un nombre real! Com que dues rectes concurrents sempre estan sobre un pla, a partir d'ara considerarem només rectes del pla $\mathbb{R}^2$.

Ara que ja sabem què és un angle, hem d'introduir una *relació d'equivalència* entre els angles que ens permeti parlar d'*angles congruents* o, per abús de llenguatge, «angles iguals».

> *Dos angles (L_1, L_2), (L_1', L_2') direm que són **congruents** si existeix un* moviment rígid *de $\mathbb{R}^2$ que conserva l'orientació i transforma $L_1 \mapsto L_1'$ i $L_2 \mapsto L_2'$.*

El concepte de **moviment rígid** forma part de la geometria euclidiana i no és aquest el lloc per fer-ne un estudi a fons. Diguem només que un moviment rígid del pla euclidià és una aplicació $f : \mathbb{R}^2 \to \mathbb{R}^2$ que conserva les distàncies entre punts, és a dir: $d(P, Q) = d(f(P), f(Q))$ per tota parella de punts $P, Q \in \mathbb{R}^2$. Tot moviment rígid es pot descompondre en una *translació* seguida d'una aplicació *ortogonal* $\phi : \mathbb{R}^2 \to \mathbb{R}^2$, és a dir, un isomorfisme lineal que conserva el *producte escalar*

$$\overrightarrow{u} \bullet \overrightarrow{v} = \phi(\overrightarrow{u}) \bullet \phi(\overrightarrow{v}) \text{ per tot } \overrightarrow{u}, \overrightarrow{v} \in \mathbb{R}^2.$$

ϕ vindrà donat per una matriu A tal que $AA^t = I$. Com que $\det(\phi) \neq 0$, només hi ha dues possibilitats: $\det(\phi) > 0$ o $\det(\phi) < 0$; en el primer cas direm que el moviment rígid *conserva l'orientació* i en el segon cas direm que *inverteix l'orientació*.[2]

Si designem per $\mathcal{A}$ el conjunt de tots els angles del pla $\mathbb{R}^2$, podem considerar el conjunt quocient de $\mathcal{A}$ per la relació d'equivalència anterior, $\mathcal{A}/\sim$.

Funcions trigonomètriques: la versió geomètrica

A continuació definirem el **cosinus d'un angle** en la seva versió geomètrica. Per tal de no confondre aquest concepte amb el de la *funció cosinus* que definirem

[2] Per un tractament rigorós i complet d'aquests conceptes de la geometria euclidiana, vegeu el llibre *Un curs de geometria lineal*, del mateix autor d'aquestes notes.

més endavant, la notació que utilitzarem serà Cos(−).

Sigui (L_1, L_2) *un angle. Definim*

$$\mathrm{Cos}(L_1, L_2) := \overrightarrow{u_1} \bullet \overrightarrow{u_2} \in \mathbb{R}$$

on cada $\overrightarrow{u_i}$ *és un vector director unitari de* L_i *per* $i = 1, 2$*, de manera que la parella ordenada* $(\overrightarrow{u_1}, \overrightarrow{u_2})$ *és una* base positiva *de* $\mathbb{R}^2$*.*

Una base de $\mathbb{R}^2$ diem que és **positiva** si la matriu de canvi de base respecte de la base canònica té determinant positiu.[3] Observem que $\mathrm{Cos}(L_1, L_2)$ està ben definit, malgrat que cadascuna de les rectes L_i té dos vectors directors unitaris. Les propietats del producte escalar —en particular, la important desigualtat de Cauchy-Schwarz— ens diuen que

$$\mathrm{Cos}(L_1, L_2) \in (-1, 1)$$

i això ens permet definir el **sinus** (geomètric!) d'un angle con

$$\mathrm{Sin}(L_1, L_2) := \sqrt{1 - \mathrm{Cos}(L_1, L_2)^2} \in (0, 1]$$

de manera que es compleix —per definició!— la identitat fonamental

$$\mathrm{Sin}(L_1, L_2)^2 + \mathrm{Cos}(L_1, L_2)^2 = 1.$$

Algunes conseqüències d'aquestes definicions geomètriques de Sin i Cos:

- Reprodueixen els conceptes clàssics de *«catet contigu dividit per hipotenusa»* i *«catet oposat dividit per hipotenusa»*, coneguts des de fa diversos mil·lenis. En efecte, considerem el triangle rectangle de $\mathbb{R}^2$ de vèrtex $(0, 0)$, $(a, 0)$ i (a, b) amb $a, b > 0$ i sigui $h = \sqrt{a^2 + b^2}$ la longitud de la hipotenusa del triangle. Considerem l'angle (L_1, L_2) on L_1 és la recta $y = 0$ i L_2 és la recta $bx = ay$. Aleshores, $\mathrm{Cos}(L_1, L_2) = a/h$ i $\mathrm{Sin}(L_1, L_2) = b/h$.
- És evident que dos angles congruents tenen els mateixos Sinus i Cosinus.
- Recíprocament, *si dos angles tenen el mateix Cosinus, aleshores són congruents*. Efectivament, suposem que tenim dos angles (L_1, L_2) i (L_1', L_2') tals que $\mathrm{Cos}(L_1, L_2) = \mathrm{Cos}(L_1', L_2')$. En primer lloc, fent dues translacions, podem suposar que els vèrtex dels dos angles són l'origen de $\mathbb{R}^2$. Siguin $\overrightarrow{e_1} = (a_1, b_1)$, $\overrightarrow{e_2} = (a_2, b_2)$, $\overrightarrow{e_1}' = (a_1', b_1')$, $\overrightarrow{e_2}' = (a_2', b_2')$ vectors com els de la definició del cosinus, de manera que

$$\overrightarrow{e_1} \bullet \overrightarrow{e_2} = \mathrm{Cos}(L_1, L_2) = \mathrm{Cos}(L_1', L_2') = \overrightarrow{e_1}' \bullet \overrightarrow{e_2}'.$$

[3] De manera informal, una base $\overrightarrow{u_1}, \overrightarrow{u_2}$ de $\mathbb{R}^2$ és positiva si el camí més curt de $\overrightarrow{u_1}$ a $\overrightarrow{u_2}$ és en el sentit contrari a les agulles del rellotge.

Sigui $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ l'isomorfisme lineal que transforma la base (positiva) $\overrightarrow{e_1}, \overrightarrow{e_2}$ en la base (positiva) $\overrightarrow{e_1}', \overrightarrow{e_2}'$. La matriu de ϕ serà

$$\phi = \begin{pmatrix} a_1' & a_2' \\ b_1' & b_2' \end{pmatrix} \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}^{-1}.$$

Aleshores, un càlcul senzill (que utilitza la igualtat $\overrightarrow{e_1} \bullet \overrightarrow{e_2} = \overrightarrow{e_1}' \bullet \overrightarrow{e_2}'$) mostra que $\phi^t \phi = I$ i ϕ és, per tant, un moviment rígid que conserva l'orientació, de manera que hem trobat un moviment rígid que transforma l'angle (L_1, L_2) en l'angle (L_1', L_2').

- Tot nombre real entre -1 i 1 és Cosinus d'algun angle. Si $r \in (-1,1)$, $r \neq 0$, podem considerar les rectes $L_1 : \{y = 0\}$ i $L_2 : \{y = mx\}$ amb $m := \sqrt{1-r^2}/r$ i comprovar que $\mathrm{Cos}(L_1, L_2) = r$.

Tenim, doncs, una funció bijectiva

$$\mathrm{Cos} : \mathcal{A}/\sim \longrightarrow (-1,1).$$

Podem adjuntar al conjunt $\mathcal{A}/\sim$ dos nous elements, anomenats *«angle nul»* i *«angle pla»* —que, segons la definició que hem donat, **no** són angles— i assignar a aquests nous elements els valors de Cos $= 1, -1$, respectivament. Si designem per $\overline{\mathcal{A}/\sim}$ aquest conjunt ampliat, tenim una bijecció

$$\mathrm{Cos} : \overline{\mathcal{A}/\sim} \longrightarrow [-1,1].$$

En definitiva, podem podem utilitzar l'aplicació Cos —el cosinus *geomètric*— com una *mesura d'angles*: per saber si dos angles són congruents o no ho són, n'hi ha prou amb calcular els seus Cosinus. De tota manera, aquesta mesura d'angles té l'inconvenient de no ser *additiva*: en general, una bona mesura compleix que la mesura de l'objecte que obtenim afegint (en un determinat sentit) dos objectes és la suma de les mesures de cadascun dels objectes. L'aplicació Cos no compleix aquesta propietat, però ens interessa molt estudiar com es comporta respecte de la suma d'angles —un concepte que ens caldrà definir.

No podem definir una suma d'angles en general, però sí que podem definir una operació de suma en un cas particular: suposem que tenim tres rectes que passen per l'origen:

$$L_1 := \langle \overrightarrow{e_1} \rangle,\ L_2 := \langle \overrightarrow{e_2} \rangle,\ L_3 := \langle \overrightarrow{e_3} \rangle$$

i suposem que $\overrightarrow{e_1}$, $\overrightarrow{e_2}$, $\overrightarrow{e_3}$ són vectors unitaris tals que

$$(\overrightarrow{e_1}, \overrightarrow{e_2}),\ (\overrightarrow{e_1}, \overrightarrow{e_3}),\ (\overrightarrow{e_2}, \overrightarrow{e_3})$$

són bases positives de $\mathbb{R}^2$. En aquest cas, podem entendre que l'angle (L_1, L_3) és la suma $(L_1, L_2) + (L_2, L_3)$. Intentem ara relacionar $\mathrm{Cos}(L_1, L_3)$ amb $\mathrm{Cos}(L_1, L_2)$ i $\mathrm{Cos}(L_2, L_3)$. Tenim, per definició:

$$\begin{aligned} C_{1,2} &:= \mathrm{Cos}(L_1, L_2) = \overrightarrow{e_1} \bullet \overrightarrow{e_2}, \\ C_{1,3} &:= \mathrm{Cos}(L_1, L_3) = \overrightarrow{e_1} \bullet \overrightarrow{e_3}, \\ C_{2,3} &:= \mathrm{Cos}(L_2, L_3) = \overrightarrow{e_2} \bullet \overrightarrow{e_3}. \end{aligned}$$

i denotem $S_{1,2}, S_{1,3}, S_{2,3}$ els sinus corresponents. Escrivim

$$\overrightarrow{e_3} = \lambda\overrightarrow{e_1} + \mu\overrightarrow{e_2}$$

i observem que, com que $(\overrightarrow{e_1}, \overrightarrow{e_2})$ i $(\overrightarrow{e_2}, \overrightarrow{e_3})$ són bases positives, la matriu de canvi de base ha de tenir determinant positiu i deduïm que $\lambda < 0$. D'altra banda, $||\overrightarrow{e_3}|| = 1$ ens dóna l'equació

$$\lambda^2 + \mu^2 + 2\lambda\mu\, C_{1,2} = 1. \qquad (*)$$

Observem que

$$C_{2,3} = \overrightarrow{e_2} \bullet \overrightarrow{e_3} = \lambda C_{1,2} + \mu$$

i, substituint μ a l'equació $(*)$, obtenim $\lambda = \pm S_{2,3}/S_{1,2}$ que, com que $\lambda < 0$, ens determina unívocament el valor de $\lambda = -S_{2,3}/S_{1,2}$. D'altra banda,

$$C_{1,3} = \overrightarrow{e_1} \bullet \overrightarrow{e_3} = \lambda + \mu C_{1,2}$$

i tenim demostrada la fórmula trigonomètrica ben coneguda

$$\mathrm{Cos}(L_1, L_3) = \mathrm{Cos}(L_1, L_2)\,\mathrm{Cos}(L_2, L_3) - \mathrm{Sin}(L_1, L_2)\,\mathrm{Sin}(L_2, L_3).$$

Funcions trigonomètriques: la versió analítica

Tot el que hem fet a l'apartat anterior ho hem pogut fonamentar perfectament i és ben conegut des dels inicis de les matemàtiques. Les aplicacions Sinus i Cosinus que hem introduït estan definides sobre els *angles*, és a dir, sobre el *conjunt* $\mathcal{A}$ i no són, a diferència de les funcions sinus i cosinus que s'utilitzen a la ciència i la tecnologia, funcions de variable real. Si volem funcions

$$\sin, \cos : \mathbb{R} \longrightarrow \mathbb{R}$$

ens cal parlar de la **longitud d'un arc de circumferència** i, per fer-ho, necessitem la teoria dels nombres reals i el càlcul diferencial i integral.

Les eines del càlcul infinitesimal —en particular, la teoria de la *integració*— permeten, en el cas de les corbes anomenades *rectificables*, definir la **longitud d'una corba parametritzada** $\sigma : [a, b] \to \mathbb{R}^2$ com el límit de les longituds d'aproximacions de la corba per secants. Aquest concepte, per la seva pròpia definició, és invariant per moviments rígids, perquè els moviments rígids conserven les distàncies entre punts. La semi-circumferència unitat del semiplà superior és la corba $\sigma : [-1, 1] \to \mathbb{R}^2$ definida per $\sigma(t) := (-t, \sqrt{1-t^2})$ i es pot demostrar que és una corba rectificable. D'aquesta manera, tenim una funció contínua i estrictament creixent

$$s : [-1, 1] \longrightarrow \mathbb{R}$$

tal que $s(t)$ és la longitud de l'arc de circumferència entre els punts $\sigma(-1) = (1, 0)$ i $\sigma(t)$. Ara podem definir el nombre $\pi \in \mathbb{R}$ com

$$\pi := s(1).$$

És a dir, $\pi \in \mathbb{R}$ és la longitud de l'arc de circumferència unitat entre els punts $(1,0)$ i $(-1,0)$. Aleshores, $s : [-1,1] \to [0,\pi]$ és una funció contínua, creixent i bijectiva i podem definir la funció

$$\cos : [0,\pi] \to [-1,1]$$

com $\cos(u) := -s^{-1}(u)$ i la funció

$$\sin : [0,\pi] \to [0,1]$$

com $\sin(u) := \sqrt{1-\cos^2(u)}$.

El concepte de *longitud d'un arc de circumferència* i la funció s ens permeten *mesurar* els angles d'una manera additiva definint la mesura d'un angle (L_1, L_2) com la longitud de l'arc de circumferència unitat entre dos vectors directors unitaris de L_1 i L_2 que formin una base positiva. D'aquesta manera tenim una relació entre les versions geomètrica i analítica de les funcions trigonomètriques. Si $a \in [-1,1]$ i L_a és la recta que passa per l'origen i pel punt $(a, \sqrt{1-a^2})$ aleshores, clarament,

$$\mathrm{Cos}(y=0, L_a) = \cos(s(-a)) = a.$$

Tenim, doncs, les funcions trigonomètriques sin i cos —expressades en *radiants*!— que tothom coneix, i que mantenen el significat geomètric de les funcions $\mathrm{Sin}, \mathrm{Cos}$ que hem definit abans. Ara és senzill estendre els dominis d'aquestes dues funcions a tot el cos real definint $\cos(x+\pi) = -\cos(x)$ i $\sin(x+\pi) = -\sin(x)$ i tenir funcions periòdiques $\sin, \cos : \mathbb{R} \to \mathbb{R}$.

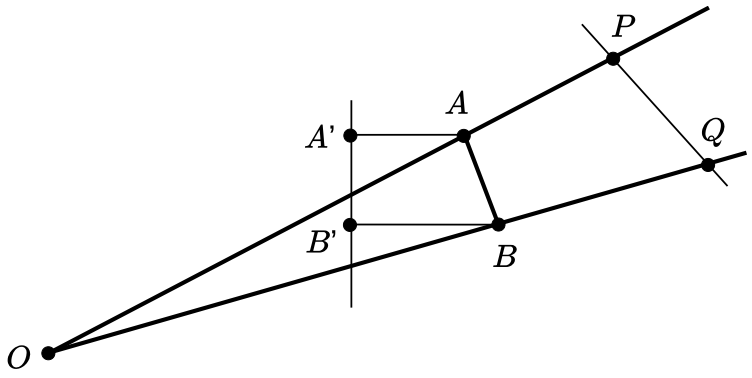
El pas següent seria demostrar les propietats bàsiques d'aquestes dues funcions, en particular, demostrar que són diferenciables i calcular les seves derivades i les seves sèries de Taylor. També els valors per $x = 2\pi/n$ per diversos valors de n. Per fer-ho en tenim prou amb les fórmules —que ja hem vist que són vàlides— de $\cos(x+y)$ i $\sin(x+y)$ i una propietat fonamental que no hem demostrat encara:[4]

$$\lim_{x\to 0} \frac{\sin(x)}{x} = 1.$$

Per demostrar aquesta propietat[5] necessitem aquest lema geomètric:

[4] Aquesta propietat requereix una reflexió sobre tot el que hem fet fins ara. D'una banda, hem definit les funcions trigonomètriques sobre els *angles*. Per convertir aquestes funcions en funcions reals, n'hem tingut prou amb definir una *mesura d'angles* i, com que era desitjable que aquesta mesura fos additiva, hem escollit com a mesura d'un angle la longitud de l'arc de circumferència unitat que determina. És a dir, hem mesurat els angles en *radiants*. Nogensmenys, és evident que qualsevol múltiple d'aquesta mesura —per exemple, la mesura en *graus*— ens hauria servit igual per definir les funcions trigonomètriques sin i cos. Però l'avantatge immens que representa mesurar els angles en radiants ens el descriu perfectament aquest límit igual a 1 —que només val 1 si a la funció $\sin(x)$ la variable x està expressada en radiants. Com que aquest límit, en particular, és el que ens garanteix que la derivada de la funció sin sigui la funció cos, veiem que la mesura d'angles en radiants simplifica dràsticament una infinitat de fórmules matemàtiques que usem diàriament.

[5] Observem que no podem utilitzar la famosa *regla de l'Hôpital* per calcular aquest límit perquè la demostració que la derivada de la funció sinus sigui la funció cosinus utilitza que aquest límit sigui 1: tindríem un cercle viciós.



En el dibuix anterior (situat en el pla $\mathbb{R}^2$), suposem que els angles a A' i B' són rectes i que els segments OA i OB tenen la mateixa longitud, és a dir, $|OA| = |OB|$. Aleshores

$$|A'B'| \leq |AB| \leq |PQ|.$$

Demostració. La primera desigualtat és senzilla: $A'B'$ és la projecció ortogonal d'un segment sobre una recta i, pel teorema de Pitàgores, $|A'B'| \leq |AB|$. La segona desigualtat, en canvi, és una mica més complicada. Denotem $\overrightarrow{e_1} := \overrightarrow{OA}$, $\overrightarrow{e_2} := \overrightarrow{OB}$ i suposem, sense pèrdua de generalitat, $||\overrightarrow{e_1}|| = ||\overrightarrow{e_2}|| = 1$, amb la qual cosa $c := ||\overrightarrow{e_1} \bullet \overrightarrow{e_2}|| < 1$. Tindrem $\overrightarrow{OP} = \lambda\overrightarrow{e_1}$, $\overrightarrow{OQ} = \mu\overrightarrow{e_2}$, per uns certs $\lambda, \mu > 1$. Un càlcul senzill ens dóna

$$||\overrightarrow{PQ}||^2 - ||\overrightarrow{AB}||^2 = \lambda^2 + \mu^2 - 2 - 2c(\lambda\mu - 1) > \lambda^2 + \mu^2 - 2\lambda\mu = (\lambda - \mu)^2 \geq 0$$

i això és el que volíem demostrar.[6]

Ara ja és molt senzill demostrar que el límit de $\sin(x)/x$ quan $x \to 0$ és 1. Considerem un arc XY de longitud x a la circumferència unitat. Per definició de la longitud d'una corba rectificable, x serà el límit de les sumes de les longituds de segments secants com el segment AB. Pel lema anterior tenim

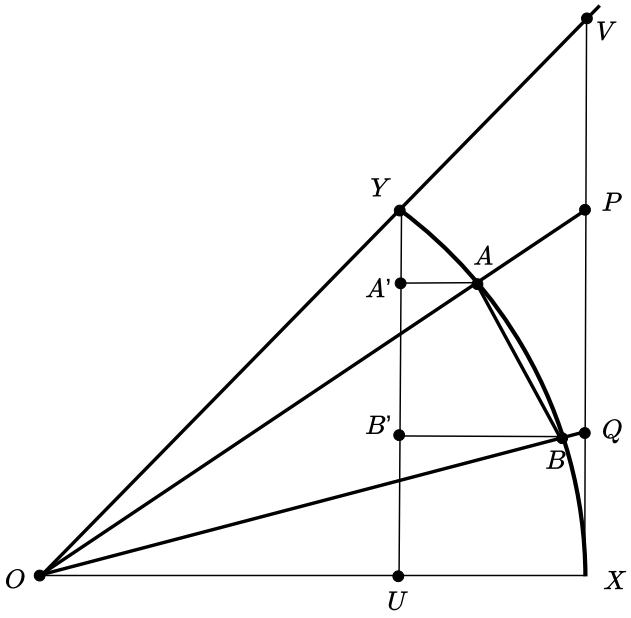
$$|A'B'| \leq |AB| \leq |PQ|$$

i, per tant, $|UY| \leq x \leq |XV|$. Però, per la definició geomètrica de les funcions trigonomètriques, $|UY| = \sin x$ i $|XV| = \sin x/\cos x$. Això ens dóna la desigualtat

$$1 \leq \frac{x}{\sin x} \leq \frac{1}{\cos x}$$

i, per continuïtat, com que els termes a l'esquerra i la dreta tendeixen a 1, el terme del mig també ha de tendir a 1.

[6] També podríem haver demostrat aquest mateix resultat utilitzant raonaments geomètrics propis de la geometria d'Euclides.



Sèries de potències formals

El concepte de polinomi amb coeficients a un cos k es pot generalitzar a la situació

$$F(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots$$

en la que (a_i) sigui una successió qualsevol d'elements de k. Direm que F és una *sèrie (de potències) formal* en la variable x i escriurem

$$F(x) = \sum_{n=0}^{\infty} a_n x^n.$$

Denotarem $k[[x]]$ el conjunt de totes les sèries formals i, amb les operacions òbvies de suma i producte —que generalitzen les dels polinomis— resulta que $k[[x]]$ és un *anell* (que conté l'anell de polinomis $k[x]$).

A diferència del que passava amb els polinomis, una sèrie formal F no es pot avaluar en un element $a \in k$, és a dir, una sèrie formal no dóna una funció $F : k \to k$ i no té sentit escriure $F(a)$. Sí que és cert que quan k és un cos com $\mathbb{R}$ o $\mathbb{C}$ en el qual existeix un concepte de *límit*, es pot donar un significat a l'avaluació d'alguna sèrie formal $F(x)$ en alguns elements a del cos.

Tal com passava amb els polinomis, podem definir la *derivada* d'una sèrie formal sense necessitat d'invocar cap pas al límit:

$$F'(x) := \sum_{n=0}^{\infty} (n+1)a_{n+1}x^n.$$

El concepte de sèrie formal es pot estendre al cas de *diverses variables*. Per exemple, podem parlar de l'anell $k[[x,y]]$ format pels elements

$$F(x,y)=a_{00}+a_{01}y+a_{10}x+a_{02}y^2+a_{11}xy+a_{20}x^2+\cdots$$

que escriurem en la forma

$$F(x,y)=\sum_{n=0}^{\infty}\sum_{i=0}^{n}a_{i,n-i}\,x^i y^{n-i}.$$

Una sèrie formal molt especial

Aquesta sèrie formal té propietats molt interessants:

$$E(x):=\sum_{n=0}^{\infty}\frac{1}{n!}\,x^n\in\mathbb{Q}[[x]].$$

Efectivament:

- Coincideix amb la seva derivada: $E'(x)=E(x)$. La comprovació és trivial.
- Transforma sumes en productes, és a dir:

  $$E(x+y)=E(x)E(y).$$

  Aquesta fórmula es compleix a l'anell $\mathbb{Q}[[x,y]]$ i la comprovació és senzilla, aplicant la fórmula de les potències d'un binomi:

  $$\begin{aligned}E(x+y)=&\sum_{n=0}^{\infty}\frac{1}{n!}(x+y)^n=\sum_{n=0}^{\infty}\frac{1}{n!}\sum_{r=0}^{n}\binom{n}{r}x^r y^{n-r}\\=&\sum_{n=0}^{\infty}\frac{1}{n!}\sum_{r=0}^{n}\frac{n!}{r!(n-r)!}\,x^r y^{n-r}\\=&\sum_{n=0}^{\infty}\sum_{r=0}^{n}\frac{x^r}{r!}\,\frac{y^{n-r}}{(n-r)!}\\=&E(x)E(y).\end{aligned}$$

- Afegim ara al cos base $\mathbb{Q}$ un nou element i que sigui un zero del polinomi x^2+1, és a dir, tal que $i^2=-1$ (com hem après a fer al capítol 29) i considerem la sèrie formal $E(ix)$. Observem que

  $$i^n=\begin{cases}1 & n\equiv 0 \mod 4\\ i & n\equiv 1 \mod 4\\ -1 & n\equiv 2 \mod 4\\ -i & n\equiv 3 \mod 4\end{cases}$$

D'aquí deduïm aquesta igualtat:

$$E(ix) = C(x) + iS(x)$$

on $C(x)$ i $S(x)$ són aquestes sèries formals

$$C(x) = \sum_{n=0}^{\infty} (-1)^n \frac{1}{(2n)!}\, x^{2n}, \qquad S(x) = \sum_{n=0}^{\infty} (-1)^n \frac{1}{(2n+1)!}\, x^{2n+1}.$$

Quan el cos base és $\mathbb{C}$...

Quan el cos base és $\mathbb{C}$, el càlcul infinitesimal ens diu que la sèrie formal $E(x)$ es pot avaluar a qualsevol nombre complex $z \in \mathbb{C}$, de manera que tenim una funció $\mathbb{C} \to \mathbb{C}$ definida per $z \mapsto E(z)$ que té les propietats que abans hem vist que té la sèrie formal $E(x)$. D'aquesta funció importantíssima en diem la **funció exponencial**

$$\exp : \mathbb{C} \to \mathbb{C}.$$

Això ens permet definir $e := \exp(1) \in \mathbb{R}$ i escriure $e^z := \exp(z)$. A més, les sèries formals $C(x)$ i $S(x)$ també es poden avaluar a qualsevol nombre real $a \in \mathbb{R}$ i obtenim dues funcions reals. Comparant les sèries formals $C(x)$ i $S(x)$ amb les sèries de Taylor de les funcions cos i sin que hem discutit abans, veiem que $C(x)$ i $S(x)$, quan les avaluem en els nombres reals, donen lloc a les funcions cos i sin.

En conclusió, La fórmula $E(ix) = C(x) + iS(x)$ de l'apartat anterior, que era una igualtat algebraica entre sèries formals, es converteix en una igualtat a $\mathbb{C}$:

$$e^{a+ib} = e^a e^{ib} = e^a\,(\cos(b) + i\sin(b))$$

vàlida per a tot $a, b \in \mathbb{R}$. Aquesta és la **fórmula d'Euler** i, quan l'apliquem a $a = 0$, $b = \pi$, obtenim *la fórmula més bonica de les matemàtiques*

$$e^{i\pi} + 1 = 0.$$

Algunes conseqüències de la fórmula d'Euler

- **Expressió polar d'un nombre complex**. Si $z = a + bi \in \mathbb{C}$ té norma 1, el punt (a, b) està sobre la circumferència unitat i, per tant, si $\theta \in [0, 2\pi)$ és la longitud de l'arc entre $(1, 0)$ i (a, b), tindrem $a = \cos\theta$, $b = \sin\theta$ i podrem escriure

  $$z = a + bi = \cos\theta + i\sin\theta = e^{i\theta}.$$

  Direm que θ és l'*argument* de z. En general, si $z \neq 0$, podrem escriure

  $$z = ||z||\,\frac{z}{||z||} = ||z||\,e^{i\theta}.$$

  Aleshores, la *forma polar* de z és $||z||_\theta$, de manera que tot nombre complex diferent de zero s'expressa, de manera única, com r_θ amb $r > 0$ i $\theta \in [0, 2\pi)$ o $\theta \in \mathbb{R}/2\pi\mathbb{Z}$.

- **Producte en forma polar**. Si recordem que l'expressió polar r_θ representa el nombre complex
$$r_\theta = re^{i\theta},$$
observarem que la multiplicació dels nombres complexos s'expressa, en forma polar, d'aquesta manera:
$$r_\theta\, r'_{\theta'} = re^{i\theta}r'e^{i\theta'} = rr'e^{i(\theta+\theta')} = (rr')_{\theta+\theta'}.$$

- **Nombres complexos i rotacions**. El fet que quan multipliquem un nombre complex r_θ per un nombre complex 1_α obtinguem el nombre complex $r_{\theta+\alpha}$ ens diu que podem identificar cada rotació del pla (de centre l'origen) com *multiplicar per un nombre complex de norma 1*. És a dir, el grup de les rotacions del pla s'identifica al conjunt dels nombres complexos de norma 1 que, d'altra banda, s'identifica a la circumferència unitat del pla.

- **Arrels n-èsimes**. Si n és un enter positiu, tot nombre complex diferent de zero té exactament n arrels n-èsimes diferents que s'expressen en forma polar d'aquesta manera: sigui $z = r_\theta \in \mathbb{C} - \{0\}$ l'expressió polar de z, aleshores, les arrels n-èsimes de z són
$$(\sqrt[n]{r})_{\alpha_i}, \quad \alpha_i = \frac{\theta + 2k\pi}{n},\ k = 0, 1, \ldots, n-1.$$

- **Arrels de la unitat**. El teorema fonamental de l'àlgebra ens diu que, per cada $n > 0$, l'equació $x^n - 1$ té exactament n solucions a $\mathbb{C}$ que, d'altra banda, són totes diferents. Les solucions d'aquesta equació són les *arrels n-èsimes de la unitat*. Ara les podem descriure exactament: són els nombres complexos de mòdul 1 i argument $2k\pi/n$ per $0 \leq k < n$ i coincideixen amb els vèrtex del polígon regular de n costats centrat a l'origen que té un vèrtex al punt $(1, 0)$.

Exercicis de nombres complexos

1. (a) Demostreu aquesta propietat dels nombres complexos: $||zw|| = ||z||\,||w||$.
   (b) Demostreu que si dos nombres enters $n, m > 1$ són suma de dos quadrats, aleshores nm també és suma de dos quadrats. (Utilitzeu l'apartat anterior.)

2. Hem vist que les matrius $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ amb $a, b \in \mathbb{R}$ formen un cos isomorf al cos $\mathbb{C}$. Intentem ara repetir el mateix procés amb $a, b \in \mathbb{C}$. Demostreu que el que s'obté no és un cos. En canvi, si considerem les matrius $\begin{pmatrix} a & -b \\ \overline{b} & \overline{a} \end{pmatrix}$, sí que obtenim un cos que conté el cos $\mathbb{C}$, excepte que la multiplicació no és commutativa. Se'n diu l'anell de divisió dels *quaternions*.

3. Considereu el polinomi $x^2 + 1$ sobre els quaternions (que hem definit a l'exercici anterior). Demostreu que té infinits zeros (que formen una esfera de dimensió 2 a $\mathbb{R}^4$). Deduïu que la propietat commutativa de la multiplicació és essencial en la teoria dels polinomis que hem estudiat.

4. Sigui $t \in \mathbb{R}$ i considerem la matriu

$$R_t := \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}.$$

   Demostreu que no és diagonalitzable com a matriu real, però sí que ho és com a matriu complexa. Diagonalitzeu-la sobre $\mathbb{C}$.

5. Dibuixeu les següents regions del pla complex:

$$\Omega_1 = \{z \in \mathbb{C} \,:\, |z+3| < 2\}, \qquad \Omega_2 = \{z \in \mathbb{C} \,:\, |\mathrm{Re}(z)| < 1\},$$
$$\Omega_3 = \{z \in \mathbb{C} \,:\, |z| - z = i\}, \qquad \Omega_4 = \{z \in \mathbb{C} \,:\, |z| \leq \mathrm{Re}(z+2)\}$$

6. (a) D'entre els cossos $\mathbb{Q}$, $\mathbb{C}$, $\mathbb{F}_p$ només n'hi ha un que admeti un automorfisme diferent de la identitat. Demostreu-ho.
   (b) Demostreu que $\mathbb{R}$ tampoc no admet cap automorfisme diferent de la identitat. (Indicació: demostreu que un automorfisme de $\mathbb{R}$ ha de conservar l'ordre.)

7. Si $p \in \mathbb{C}[x]$, denotem $\overline{p} \in \mathbb{C}[x]$ el polinomi que s'obté conjugant tots els coeficients de p. Demostreu que per tot polinomi p es compleix $p\,\overline{p} \in \mathbb{R}[x]$.

8. Siguin: $z_1 = 2 - 3i$, $z_2 = -1 - 4i$, $z_3 = \frac{1}{3} - i$, $z_4 = -3i$. Calculeu:

$$z_1 z_2 - z_3 z_4, \quad |\overline{z_1} - \overline{z_2}^2|, \quad \frac{z_2}{z_4 - \overline{z_3}}.$$

9. Sigui $P(x) = a_n x^n + \cdots + a_1 x + a_0 \in \mathbb{R}[x]$ i suposem que $z \in \mathbb{C} - \mathbb{R}$ és un zero de $P(x)$. Demostreu que $\overline{z}$ també és un zero de $P(x)$. Trobeu un polinomi de grau 2 a $\mathbb{R}[x]$ que divideixi P.

10. Trobeu (en la forma $z = a + bi$) una solució a $\mathbb{C}$ de l'equació $z^6 + 2z^3 + 2 = 0$.

11. Calculeu la part real, la part imaginària, la norma, el conjugat i l'invers dels següents nombres complexos:

$$(1+i)^6,\ i^{17},\ i^7+i+1,\ (3+3i)^4,\ (-i)^{-1},\ \frac{i-4}{2i-3}.$$

12. Calculeu la part real, la part imaginària, la norma, el conjugat i l'invers dels següents nombres complexos:

$$e^{i\pi},\quad 2e^{-2\pi i/3},\quad 12e^{\pi i/6},\quad e^{2\pi i/3}+e^{4\pi i/3},\quad e^{5+\frac{i\pi}{2}}.$$

13. Calculeu $(1-i)^{23}$ i $(\sqrt{2}-\sqrt{2}\,i)^{17}$.

14. El polinomi x^3-2x+1 té tres zeros reals, un d'ells enter. Trobeu-los. Apliqueu ara la fórmula de Cardano. Observeu que apareixen arrels quadrades de nombres negatius però, interpretant apropiadament la fórmula en els nombres complexos, s'obtenen realment els tres zeros reals.[7] (Indicació: apliqueu la fórmula del cosinus de l'angle triple $\cos 3\alpha = 4\cos^3\alpha - 3\cos\alpha$.)

15. Sigui $F(x)$ una sèrie formal amb coeficients a $\mathbb{Q}$. Demostreu que $F(x)$ té un invers multiplicatiu si i només si $F(x)$ no és divisible per x. (Indicació: definiu per recursió els coeficients de F^{-1}.)

16. Sigui $F(x)$ la sèrie formal que té com a coeficients els nombres de Fibonacci (pàgina 166). Demostreu que $F(x) = x + xF(x) + x^2F(x)$.

17. Demostreu la *fórmula de de Moivre*: per tot $t \in \mathbb{R}$ i tot $n > 0$

$$(\cos t + i\sin t)^n = \cos(nt) + i\sin(nt).$$

18. Sigui $z \in \mathbb{C}$ tal que $|z| = 1$. Calculeu $|1+z|^2 + |1-z|^2$.

19. Expresseu aquests nombres complexos en forma polar:

$$-5i,\quad 1+\sqrt{3}\,i,\quad \frac{1}{3+3i},\quad 5\sqrt{3}+5i,\quad -\pi,$$

$$-7+7i,\quad \frac{1-\sqrt{3}\,i}{2},\quad \cos\left(\frac{\pi}{5}\right) - i\sin\left(\frac{\pi}{5}\right).$$

Calculeu $(3+3i)^{829}$, $(5\sqrt{3}+5i)^{135}$, $(-7+7i)^{1017}$, $((1-\sqrt{3}\,i)/2)^{4002}$.

20. Trobeu tots els zeros d'aquests polinomis

a) $z^2 - i$ b) $z^4 + 1$ c) $z^3 - 1 - i\sqrt{3}$
d) $z^2 + z + 1$ e) $z^6 + 1 - i$ f) $z^4 + z^2 + 1$

[7] Aquest exercici permet veure que la fórmula de Cardano té un interès pràctic ben escàs.

21. Factoritzeu completament a $\mathbb{C}[z]$ els polinomis següents:

    (a) $z^4 - 1$

    (b) $z^4 + 16$

    (c) $z^3 + 27$

    (d) $z^3 + (i-1)z^2 + (1-i)z - 1$

    En els tres primers casos, trobeu també la factorització completa a $\mathbb{R}[z]$.

22. Calculeu les arrels quadrades de i, $-i$, $-\pi$, $(1-\sqrt{3}\,i)/2$ i $e^{-1+i\pi/4}$ en forma polar i en forma cartesiana.

23. Expresseu en forma cartesiana el nombre complex

$$z := \frac{1+i}{3e^{i\pi/3}}.$$

    Calculeu z^{2022}.

24. Demostreu que és impossible definir una relació d'ordre total a $\mathbb{C}$ que compleixi aquestes dues propietats (per tot $z_1, z_2, z \in \mathbb{C}$): $z_1 < z_2$ implica $z_1 + z < z_2 + z$; si $z_1 > 0$ i $z_2 > 0$ aleshores $z_1 z_2 > 0$. És a dir, $\mathbb{C}$ no és un cos ordenat (exercici **IV**.3).

25. Uns estudiants de matemàtiques van dissenyar una samarreta que contenia aquesta paradoxa:

$$1 = \sqrt{1} = \sqrt{(-1)(-1)} = \sqrt{-1}\sqrt{-1} = (\sqrt{-1}\,)^2 = -1.$$

    Expliqueu exactament on és la fal·làcia i quines conseqüències en podem treure.

26. Proveu que per a tot nombre complex z es té $e^z \neq 0$. Proveu que per a tota parella z, w de nombres complexos, la igualtat $e^z = e^w$ equival a $z - w \in 2\pi i\mathbb{Z}$.

27. Resoleu les següents equacions, on la indeterminada z pren valors complexos:

$$e^z = 1, \quad e^z = e, \quad e^{3z-1} = i, \quad e^{\log(4)+i\pi} = ie^{z+2}, \quad e^z \in \mathbb{R}.$$

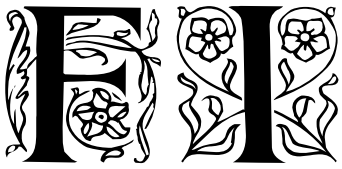
28. Definiu una funció $\log : \mathbb{C}^* \to \mathbb{C}$ que compleixi $e^{\log z} = z$ per tot $z \neq 0$. Observeu que aquesta funció no és única. Demostreu que, en general, no es compleixen aquestes identitats: $\log(e^z) = z$, $\log(z_1 z_2) = \log z_1 + \log z_2$.

29. Demostreu que si z és un nombre complex de mòdul 1, aleshores

$$z = \left(\frac{1+z}{|1+z|}\right)^2.$$

    Deduïu, sense utilitzar el teorema fonamental de l'àlgebra ni la forma polar, que tot nombre complex té arrel quadrada. Trobeu, en forma cartesiana, les solucions d'aquesta equació sobre el nombres complexos:

$$i\,x^2 + (3-5i)\,x - (7-4i) = 0.$$



però això només és el principi...

Índex alfabètic

pàgina (gairebé) en blanc

Jaume Aguadé Bover (Barcelona 1953) ha sigut catedràtic al Departament de Matemàtiques de la Universitat Autònoma de Barcelona durant 34 anys. La seva àrea de treball ha estat la topologia algebraica i la teoria homotòpica de grups de Lie. També s'ha implicat profundament en la docència i ha publicat diversos llibres que reflecteixen la seva manera personal d'entendre l'ensenyament de les matemàtiques: *«Apunts d'un curs de topologia elemental»*, «*Matemàtiques i modelització per a les ciències ambientals»*, «*Un curs de geometria lineal»* i aquest que ara teniu a les mans. En l'actualitat, ja jubilat, treballa en un llibre sobre grups de Lie compactes amb el títol provisional *«Grups que has de conèixer»*. Més enllà de les matemàtiques, ha practicat amb passió l'excursionisme, l'alpinisme i l'esquí de muntanya que, en els darrers anys, ha substituït pel ciclisme de carretera.