

BUTLLETÍ

DE LA SOCIETAT CATALANA
DE MATEMÀTIQUES
Institut d'Estudis Catalans

— í n d e x —

- ☐ *Jaume Agudé*
Un repàs a les rotacions en baixa dimensió 5
- ☐ *Jaume Coll Guerrero*
Poliedres i papiroflèxia modular 37
- ☐ *Xavier Guitart i Marc Masdeu*
Punts racionals en corbes el·líptiques 73
- ☐ *Jaume de Haro*
Les equacions de Friedmann: mecànica newtoniana *versus* relativitat
general 103
- ☐ *Luis Sierra, Marta Casanellas i Piotr Zwiernik*
Tensors en estadística algebraica 131
- ☐ English summaries 171



Volum 39 • Número 1-2 • Any 2024



Institut
d'Estudis
Catalans

BUTLLETÍ

DE LA SOCIETAT CATALANA

DE MATEMÀTIQUES

Institut d'Estudis Catalans

Volum 39 • Número 1-2 • Desembre 2024

BARCELONA

2024

© dels autors dels articles

Editat per la Societat Catalana de Matemàtiques
filial de l'Institut d'Estudis Catalans
Carrer del Carme, 47
08001 Barcelona

Text original revisat lingüísticament
per la Unitat d'Edició del Servei Editorial de l'IEC

Compost per Rosa M. Rodríguez
Imprès a Ediciones Gráficas Rey, SL

ISSN: 0214-316-X
Dipòsit Legal: B 19272-1987



Els continguts del BUTLLETÍ DE LA SOCIETAT CATALANA DE MATEMÀTIQUES estan subjectes —llevat que s'indiqui el contrari en el text o en el material gràfic— a una llicència Reconeixement - No comercial - Sense obres derivades 3.0 Espanya de Creative Commons, el text complet de la qual es pot consultar a <https://creativecommons.org/licenses/by-nc-nd/3.0/es/deed.ca>. Així, doncs, s'autoritza el públic en general a reproduir, distribuir i comunicar l'obra sempre que se'n reconegui l'autoria i l'entitat que la publica i no se'n faci un ús comercial ni cap obra derivada.

Índex

JAUME AGUADÉ
Un repàs a les rotacions en baixa dimensió 5

JAUME COLL GUERRERO
Políedres i papiroflèxia modular37

XAVIER GUITART I MARC MASDEU
Punts racionals en corbes el·líptiques 73

JAUME DE HARO
Les equacions de Friedmann: mecànica newtoniana *versus* relativitat general..103

LUIS SIERRA, MARTA CASANELLAS I PIOTR ZWIERNIK
Tensors en estadística algebraica 131

English summaries171

Un repàs a les rotacions en baixa dimensió

JAUME AGUADÉ

Resum: Comencem amb les rotacions de la geometria elemental i les matrius ortogonals i acabem construint de manera explícita, elemental i autocontinguda els isomorfismes clàssics entre els grups d'espínors en dimensió < 7 i certs grups unitaris o simplèctics. La intenció del treball és essencialment didàctica i es pot llegir com una introducció a la teoria dels grups de Lie compactes.

Paraules clau: rotació, espínor, grup de Lie, quaternions.

Classificació MSC2020: 22-01, 20G20, 22E15.

1 Presentació

En aquest treball de caràcter elemental —i intenció clarament didàctica—, estudiarem amb detall les rotacions i els espínors en dimensions menors que set, i els grups que formen. Abans de començar, pot ser útil incloure uns comentaris adreçats als lectors que ja coneguin la teoria bàsica dels grups de Lie compactes. Aquests lectors ja saben que, per a valors molt petits del rang n hi ha coincidències entre els diagrames de Dynkin de les famílies A_n , B_n , C_n , D_n . Aquestes coincidències, juntament amb el teorema que relaciona els grups de Lie compactes connexos simples amb els diagrames de Dynkin a través dels seus sistemes d'arrels, impliquen que hi ha isomorfismes entre els representants simplement connexos d'alguns grups clàssics de rang petit. Aquests isomorfismes s'anomenen *isogènies esporàdiques*. Amb aquestes isogènies, els grups d'espínors **Spin**(3), **Spin**(4), **Spin**(5), **Spin**(6) —i només aquests— són isomorfs a certs grups simplèctics o unitaris.

L'existència d'aquestes isogènies esporàdiques és una conseqüència d'un teorema de grandíssima envergadura com és el que ens dona la classificació dels grups de Lie compactes connexos simples, però aquestes isogènies també es poden construir explícitament per mètodes que podríem qualificar

d'*elementals*.¹ Són unes construccions que duen sovint l'eufemístic qualificatiu de «ben conegudes» però, més enllà del cas de **Spin**(3), no és fàcil trobar-ne una exposició elemental i completa.

El primer objectiu d'aquest treball va ser construir explícitament aquestes isogènies esporàdiques i, al mateix temps, repassar certs resultats clàssics sobre la geometria de les rotacions en baixa dimensió, però ràpidament ens vam adonar que, si bé aquestes isogènies eren la nostra *Ítaca* —«sense ella no hauríem partit»—, tots els paisatges que anàvem trobant fent camí eren potser més interessants que l'objectiu final. El text va anar adquirint un fort component didàctic, i tal com el presentem ara potser es pot llegir com una *introducció a la teoria dels grups de Lie compactes*, basada en l'estudi concret de les rotacions en baixa dimensió i la seva relació amb els espinors.²

2 Preliminars: parlem de rotacions

Euclides, probablement per evitar les famoses *apories del moviment* (Zenó, Aquil·les...), intenta desenvolupar tota la seva geometria sense utilitzar mai el *moviment*. Tanmateix, no pot deixar de fer-ho (dissimuladament) en algun punt concret, el més escandalós dels quals és ni més ni menys que la proposició quarta del llibre primer: el *criteri CAC* de congruència de triangles. De fet, aquest criteri és indemostrable a partir dels axiomes d'Euclides: cal incloure'l a la llista d'axiomes —com va fer Hilbert— o bé disposar d'un grup prou gran de moviments, és a dir, de les rotacions i les translacions. En qualsevol cas, d'una manera o d'una altra, el concepte de *rotació* és inherent a la geometria euclidiana i el podem definir com *qualsevol transformació del pla que deixi un punt fix, conservi la distància i conservi l'orientació*.

Una conseqüència significativa dels axiomes de la geometria euclidiana és que podem identificar els punts i les rectes d'Euclides amb els punts i les rectes de l'espai afí de dimensió 2 o 3 sobre un cos ordenat arquimedià pitagòric. Si, a més, volem disposar de tots els angles, de tots els polígons regulars, del nombre π i, en general, si volem poder utilitzar conceptes com *continuitat*, *connexió* i *completesa*, aleshores aquest cos pitagòric ha de ser el cos real i hem d'identificar la geometria d'Euclides plana a la geometria afí euclidiana ordinària de $\mathbb{R}^2$. En aquest context, no és difícil demostrar que una rotació no és altra cosa que una aplicació lineal $R: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ que conserva el producte escalar ordinari i conserva l'orientació. Si M_R és la matriu de R en la base canònica, R és una rotació si i només si $M_R M_R^t = I$ (és a dir, si les columnes de M_R (i les files) són vectors ortonormals) i $\det(M_R) = 1$.

Amb això que hem dit, el concepte de *rotació* admet una generalització òbvia a l'espai euclidià $\mathbb{R}^n$ en qualsevol dimensió n . Les matrius reals M que compleixen $MM^t = I$ s'anomenen *matrius ortogonals* i, necessàriament, tenen determi-

¹ Sense àlgebres de Clifford, àlgebres de Lie, sistemes d'arrels, teoria de representacions, geometria diferencial, grups de Coxeter, grups de Weyl ni, evidentment, el teorema de classificació.

² Quan he impartit cursos sobre grups de Lie he dedicat més temps a treballar «a mà» amb els grups clàssics que no pas a fer teoria general. Aquest treball segueix les mateixes idees.

nant ± 1 . Formen un grup que es designa $\mathbf{O}(n)$: el *grup ortogonal* en dimensió n . Les rotacions s'identificaran als elements de $\mathbf{O}(n)$ de determinant 1 i formen un subgrup normal de $\mathbf{O}(n)$ que es denota amb $\mathbf{SO}(n)$: el grup *especial* ortogonal. Si incloem $\mathbb{R}^n \subset \mathbb{R}^{1+n}$ com l'hiperplà $x_1 = 0$, tindrem una cadena de subgrups

$$\{I\} = \mathbf{SO}(1) \subset \mathbf{SO}(2) \subset \mathbf{SO}(3) \subset \mathbf{SO}(4) \subset \mathbf{SO}(5) \subset \mathbf{SO}(6) \subset \dots$$

L'objectiu d'aquest treball és estudiar d'una manera minuciosa els sis primers grups d'aquesta cadena i relacionar-los amb els *espinors*, un concepte que ja explicarem quan arribi el moment.

La geometria afí sobre els espais vectorials complexos $\mathbb{C}^n$ és força diferent de la geometria euclidiana, però també és extraordinàriament important. Quan passem de $\mathbb{R}^n$ a $\mathbb{C}^n$ el producte escalar estàndard —que és l'eina bàsica de la geometria d'Euclides— s'ha de reemplaçar pel *producte hermític*:

$$\langle (u_1, \dots, u_n), (v_1, \dots, v_n) \rangle := u_1 \bar{v}_1 + \dots + u_n \bar{v}_n \in \mathbb{C}.$$

En aquest nou context, les matrius que conserven el producte hermític de $\mathbb{C}^n$ són les que compleixen $MM^* = I$ on $M^* := \bar{M}^t$. S'anomenen *matrius unitàries* i formen un grup: el *grup unitari* $\mathbf{U}(n)$. Igual que abans, les matrius unitàries de determinant 1 formen un subgrup normal $\mathbf{SU}(n)$ de $\mathbf{U}(n)$ i tenim una cadena de subgrups

$$\{I\} = \mathbf{SU}(1) \subset \mathbf{SU}(2) \subset \mathbf{SU}(3) \subset \mathbf{SU}(4) \subset \mathbf{SU}(5) \subset \mathbf{SU}(6) \subset \dots$$

Rotacions del pla

La teoria de les rotacions del pla és tan coneguda que potser no ens adonem de totes les matemàtiques que hi ha al seu darrere, de la quantitat de matemàtica que necessitem per fonamentar-la rigorosament. Caldria parlar, com a mínim, de la teoria geomètrica dels angles, de la quadratura de la circumferència —i, per tant, del nombre π —, de les funcions trigonomètriques, del cos dels nombres complexos, de la funció exponencial complexa i de la fórmula d'Euler.³ Aquí no toca repetir tota aquesta teoria, però sí que en farem un resum ràpid:

- Podem identificar els punts de la circumferència unitat amb els *angles*, que són els elements de $\mathbb{R}/2\pi\mathbb{Z}$. Els angles formen un grup abelià amb la suma.
- Si $A \in \mathbf{SO}(2)$, existeix un angle θ únic tal que $A = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$. El producte de rotacions planes es correspon amb la *suma* d'angles.
- Els vectors del pla $\mathbb{R}^2$ es poden identificar amb els nombres complexos $\mathbb{C}$, de manera que la circumferència unitat S^1 de $\mathbb{R}^2$ coincideix amb el conjunt dels nombres complexos de mòdul 1: $S^1 = \{z \in \mathbb{C} : z\bar{z} = 1\}$.

³ Podeu llegir, per exemple, el darrer capítol del llibre *Matemàtiques: comenceu per aquí* (2024, accés lliure amb llicència Creative Commons), del mateix autor que aquest treball.

- Els nombres complexos de mòdul 1 formen un grup amb la multiplicació: és el grup $U(1)$. Per tant, la circumferència S^1 adquireix una estructura de grup (abelià) de dues maneres (que són la mateixa): si identifiquem la circumferència als angles, podem considerar la suma d'angles; si identifiquem la circumferència als complexos de mòdul 1, tenim el producte de complexos:

$$S^1 \cong \mathbf{SO}(2) \cong U(1) \cong \mathbb{R}/2\pi\mathbb{Z} \cong \{z \in \mathbb{C} : |z| = 1\}.$$

- Si z és un nombre complex de mòdul 1, la transformació $w \mapsto zw$ és una rotació del pla $\mathbb{R}^2$ (identificat a $\mathbb{C}$). L'angle d'aquesta rotació és l'únic angle θ tal que $z = (\cos \theta, \sin \theta) = e^{i\theta}$. Els nombres complexos ens permeten donar una forma molt més compacta a l'expressió d'una rotació: una rotació del pla és la multiplicació per un nombre complex unitari.

Rotacions en qualsevol dimensió

A la vista de tot això, podríem pensar que una rotació a $\mathbb{R}^n$ deu ser una cosa ben complicada, i encara més a mesura que n es fa gran. No és així: tota rotació a $\mathbb{R}^n$ és simplement una composició de rotacions planes en plans ortogonals.⁴ Més acuradament, es compleix aquest resultat notable:

Si R és una rotació de $\mathbb{R}^n$, existeix una descomposició ortogonal

$$\mathbb{R}^n = P_0 \perp P_1 \perp \cdots \perp P_k, \quad \dim(P_0) \leq 1, \dim P_i = 2, i = 1, \dots, k,$$

tal que R és la identitat sobre P_0 i R és una rotació plana sobre cada P_i , $i = 1, \dots, k$. (Observem que $P_0 = 0$ si n és parell i $P_0 \cong \mathbb{R}$ si n és senar.)

Podem demostrar això d'aquesta manera. Definim $S := R + R^t$, que és una matriu simètrica i, per tant, serà diagonalitzable. Sigui w un vector propi de S de valor propi λ i considerem el subespai $W := \langle w, R(w) \rangle \subseteq \mathbb{R}^n$. Si W té dimensió 1, aleshores $R(w) \in W$ i W és estable per R . Si W té dimensió 2, aleshores $\lambda w = S(w) = R(w) + R^t(w)$ i, per tant, $R^2(w) = \lambda R(w) - w$, amb la qual cosa W és estable per R . Hem demostrat que R té un subespai estable W de dimensió 1 o 2.

Escrivim ara $\mathbb{R}^n = W \perp W^\perp$ i procedim per inducció: arribem a la conclusió que hi ha una base ortonormal de $\mathbb{R}^n$ en la qual la matriu de R s'expressa en forma diagonal $R = \text{Diag}(\lambda_1, \dots, \lambda_k, B_1, \dots, B_r)$, on cada B_i és una matriu 2×2 . Com que R és ortogonal, $\lambda_i = \pm 1$ i cada B_i és també ortogonal. Fent un canvi de base ortonormal, podem aconseguir que $\det(B_i) = 1$. Agrupant per parelles els -1 també obtenim blocs que són rotacions de 180 graus. Quedarà, com a màxim, un λ_i desaparellat i, com que R té determinant 1, serà $\lambda_i = 1$. $\square$

⁴ Estem dient que cada rotació de $\mathbb{R}^n$ és, geomètricament, un objecte ben senzill perquè es redueix a un producte de rotacions planes; però no estem dient, ni molt menys, que els grups $\mathbf{SO}(n)$ es redueixin a un producte de grups $\mathbf{SO}(2)$!

Una mica de topologia

L'estudi de la topologia dels grups $\mathbf{U}(n)$ i $\mathbf{O}(n)$ té un interès immens, però aquí només podem esbossar breument algunes propietats elementals. En primer lloc, tots aquests grups són, per definició, grups de matrius (reals o complexes) i, com que podem identificar l'espai vectorial real $\mathcal{M}_{n \times m}$ a $\mathbb{R}^{nm}$ i l'espai vectorial complex $\mathcal{M}_{n \times m}$ a $\mathbb{R}^{2nm}$, tots els grups de matrius hereten una topologia Hausdorff. A més, com que les operacions de multiplicació i pas a l'invers es poden escriure a partir de funcions algebraiques i, per tant, contínues, podem parlar de *grups topològics*.⁵

També és immediat observar que els grups $\mathbf{O}(n)$, $\mathbf{SO}(n)$, $\mathbf{U}(n)$ i $\mathbf{SU}(n)$ són *compactes*:⁶ són clarament tancats perquè estan definits per l'anul·lació de certes funcions algebraiques, i són acotats perquè les columnes de les matrius d'aquests grups són vectors unitaris.

Observem que, com a espai topològic, $\mathbf{U}(n)$ es pot descompondre com a $S^1 \times \mathbf{SU}(n)$, utilitzant la funció que transforma cada parella (z, A) en la matriu que s'obté multiplicant per z la primera fila de la matriu A . És una funció contínua i bijectiva que, com que els espais involucrats són compactes Hausdorff, serà un homeomorfisme. Cal remarcar que aquesta descomposició *no* és compatible amb l'estructura de grup de $\mathbf{U}(n)$: els grups $\mathbf{U}(n)$ i $S^1 \times \mathbf{SU}(n)$ són diferents, però els espais topològics $\mathbf{U}(n)$ i $S^1 \times \mathbf{SU}(n)$ són homeomorfs.

Una eina útil per a l'estudi d'aquests grups és la identificació dels espais de classes laterals $\mathbf{SO}(n)/\mathbf{SO}(n-1)$ i $\mathbf{SU}(n)/\mathbf{SU}(n-1)$ amb esferes. Concretament, si assignem a cada matriu ortogonal/unitària el vector unitari format per la seva primera columna, tenim aplicacions contínues exhaustives $\mathbf{SO}(n) \rightarrow S^{n-1}$ i $\mathbf{SU}(n) \rightarrow S^{2n-1}$ que factoritzen per homeomorfismes

$$\mathbf{SO}(n)/\mathbf{SO}(n-1) \cong S^{n-1}, \quad \mathbf{SU}(n)/\mathbf{SU}(n-1) \cong S^{2n-1}.$$

Aquests homeomorfismes es poden explotar per demostrar inductivament propietats topològiques dels grups $\mathbf{SO}(n)$ i $\mathbf{SU}(n)$. Per exemple:

Per a $n \geq 1$ es compleix que $\mathbf{SO}(n)$, $\mathbf{U}(n)$, $\mathbf{SU}(n)$ són connexos, $\mathbf{O}(n)$ té dos components connexos i $\mathbf{SU}(n)$ és simplement connex.

La primera part d'aquest teorema es demostra per inducció sobre n , començant pel fet que $\mathbf{SO}(1) = \mathbf{SU}(1) = \{I\}$ i $\mathbf{U}(1) = S^1$. Fem-ho amb detall per al cas del grup de rotacions $\mathbf{SO}(n)$, $n > 1$. Com hem dit abans, tenim una aplicació contínua exhaustiva

$$f: \mathbf{SO}(n) \rightarrow S^{n-1}$$

que, com que els espais involucrats són compactes Hausdorff, serà també tancada. Si poguéssim escriure $\mathbf{SO}(n)$ com una unió disjunta $A \cup B$ de tancats

⁵ De fet, els grups $\mathbf{O}(n)$, $\mathbf{SO}(n)$, $\mathbf{U}(n)$ i $\mathbf{SU}(n)$ són molt més que grups topològics: són varietats diferenciables i les operacions de multiplicació i pas a l'invers són diferenciables. Direm que són *grups de Lie*. Però en aquest text elemental no volem aprofundir en aquesta direcció.

⁶ Com que tenim una estimació especial pels grups compactes, els distingim usant lletra **negreta**.

no buits, tindriem $S^{n-1} = f(A) \cup f(B)$ i, com que S^{n-1} és un espai connex, aquesta unió no pot ser disjunta. Si $x \in f(A) \cap f(B)$, considerem $X := f^{-1}(x)$ i escrivim $X = (X \cap A) \cup (X \cap B)$, que és una unió disjunta de tancats de X . Observem ara que X està format per totes les matrius ortogonals que tenen com a primera columna el vector x i aquest subespai és clarament homeomorf a $\mathbf{SO}(n-1)$, que, per hipòtesi d'inducció, és connex. Arribem, doncs, a una contradicció i $\mathbf{SO}(n)$ també ha de ser connex. Pel que fa a $\mathbf{O}(n)$, el resultat és clar perquè $\mathbf{O}(n)$ és unió disjunta de dos tancats: $\mathbf{SO}(n)$ i el subespai de les matrius ortogonals de $\det = -1$. Com que aquest subespai és homeomorf a $\mathbf{SO}(n)$, tenim ja els dos components connextos de $\mathbf{O}(n)$. Finalment, una inducció utilitzant les identifications $\mathbf{SU}(n)/\mathbf{SU}(n-1) \cong S^{2n-1}$ i uns teoremes (senzills) de topologia que ara no podem entrar a discutir ens demostren que tots els grups $\mathbf{SU}(n)$ són simplement connextos. $\square$

3 Rotacions en dimensió 3 i espinors

La descripció general de les rotacions en qualsevol dimensió que hem vist abans ens diu que tota rotació no trivial en tres dimensions tindrà un eix de rotació ben definit i es comportarà com una rotació en el pla perpendicular a aquest eix. És a dir, per determinar una rotació R a $\mathbb{R}^3$ prendrem un vector unitari $\vec{u} \in \mathbb{R}^3$ —que ens indicarà l'eix— i un angle θ —que serà l'angle de rotació en el pla ortogonal a l'eix. Aleshores, la rotació R , que podem designar com a $R_{\theta, \vec{u}}$, deixarà fixos tots els punts de la recta $\langle \vec{u} \rangle$ i rotarà un angle θ els punts del pla $H := \langle \vec{u} \rangle^\perp$, entenent que una base positiva de H serà qualsevol base ortonormal $\vec{e}_1, \vec{e}_2$ tal que $\langle \vec{e}_1 \times \vec{e}_2, \vec{u} \rangle > 0$. Amb una mica de trigonometria elemental podem trobar fàcilment una expressió de $R_{\theta, \vec{u}}(\vec{v})$ que es coneix com a *fórmula d'Olinde Rodrigues*:

$$R_{\theta, \vec{u}}(\vec{v}) = \cos \theta \vec{v} + \sin \theta (\vec{u} \times \vec{v}) + (1 - \cos \theta) \langle \vec{u}, \vec{v} \rangle \vec{u}.$$

Tenim, doncs, una bona descripció de cada rotació —tinguem en compte que $\vec{u}$ i θ no estan unívocament determinats—, però encara estem lluny d'entendre el conjunt de *totes* les rotacions de l'espai, és a dir, la topologia de $\mathbf{SO}(3)$ i la seva estructura com a grup —en particular, el producte de dues rotacions $R_{\theta, \vec{u}} R_{\theta', \vec{u}'}$. Per donar resposta a aquestes qüestions hem de parlar d'uns objectes anomenats *espinors*.

Espinors

Un espinor és una parella de nombres complexos (ζ_1, ζ_2) tals que

$$\zeta_1 \bar{\zeta}_1 + \zeta_2 \bar{\zeta}_2 = 1.$$

Si escrivim aquesta equació en coordenades reals, obtenim $a^2 + b^2 + c^2 + d^2 = 1$, que defineix l'esfera unitat de $\mathbb{R}^4$, és a dir, l'esfera S^3 . En conclusió, els espinors estan en correspondència bijectiva amb els punts de l'esfera S^3 .

Els espinors es poden multiplicar d'una manera curiosa:

$$(\zeta_1, \zeta_2)(\eta_1, \eta_2) = (\zeta_1 \eta_1 - \bar{\zeta}_2 \eta_2, \zeta_2 \eta_1 + \bar{\zeta}_1 \eta_2).$$

Aquesta estranya multiplicació i les seves propietats s'entenen molt millor si escrivim els espinors en aquesta forma matricial:

$$(\zeta_1, \zeta_2) \rightsquigarrow \begin{pmatrix} \zeta_1 & -\bar{\zeta}_2 \\ \zeta_2 & \bar{\zeta}_1 \end{pmatrix}.$$

Aleshores, la multiplicació dels espinors és la multiplicació de matrius i podem comprovar que els espinors formen un grup: el designarem **Spin**(3).

Què podem dir de les matrius complexes 2×2 anteriors? No és difícil veure que són matrius unitàries de determinant 1, és a dir, pertanyen al grup **SU**(2). A més, tota matriu de **SU**(2) és de la forma $\begin{pmatrix} \zeta_1 & -\bar{\zeta}_2 \\ \zeta_2 & \bar{\zeta}_1 \end{pmatrix}$ per algun espinor (ζ_1, ζ_2) . Això és conseqüència de la identificació $\mathbf{SU}(2) = \mathbf{SU}(2)/\mathbf{SU}(1) \cong S^3$ que hem esmentat abans, però també admet una demostració elemental directa. En conclusió:

*Un espinor és la primera columna d'una matriu especial unitària 2×2 . Els espinors formen un grup designat **Spin**(3) que és isomorf al grup **SU**(2). Aquests dos grups isomorfs s'identifiquen, topològicament, a l'esfera S^3 .⁷*

Espinors i rotacions de l'espai

Hem introduït els espinors i el grup **Spin**(3) amb la intenció d'entendre millor les rotacions de $\mathbb{R}^3$, és a dir, el grup **SO**(3). Per comprendre la relació entre els espinors i les rotacions de l'espai, considerem les *matrius antihermítiques* de traça zero:

$$\mathcal{H}'_0 := \{H \in \mathcal{M}_{2 \times 2}(\mathbb{C}) : H + H^* = 0, \text{tr}(H) = 0\}.$$

$\mathcal{H}'_0$ és un $\mathbb{R}$ -espai vectorial i podem prendre aquesta base ortonormal:

$$\sigma_x = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}$$

de manera que cada vector de $\mathbb{R}^3$ s'identifica a la matriu hermítica

$$\vec{u} = (x, y, z) \mapsto H_{\vec{u}} := \begin{pmatrix} xi & -y - zi \\ y - zi & -xi \end{pmatrix}.$$

Aquestes propietats de les matrius $H_{\vec{u}}$ són senzilles de demostrar:

- **SU**(2) normalitza $\mathcal{H}'_0$, en el sentit que, si $U \in \mathbf{SU}(2)$, aleshores $U \mathcal{H}'_0 U^* = \mathcal{H}'_0$.
- $H_{\vec{u}} H_{\vec{v}} = -\langle \vec{u}, \vec{v} \rangle I + H_{\vec{u} \times \vec{v}}$ i $H_{\vec{v}} H_{\vec{u}} - H_{\vec{u}} H_{\vec{v}} = -2H_{\vec{u} \times \vec{v}}$.
- $H_{\vec{u}} H_{\vec{v}} H_{\vec{u}} = -\langle \vec{u}, \vec{v} \rangle H_{\vec{u}} + H_{(\vec{u} \times \vec{v}) \times \vec{u}}$. Si recordem la identitat $(\vec{u} \times \vec{v}) \times \vec{w} = \langle \vec{u}, \vec{w} \rangle \vec{v} - \langle \vec{u}, \vec{v} \rangle \vec{w}$, obtenim que, si $\vec{u}$ és unitari,

$$H_{\vec{u}} H_{\vec{v}} H_{\vec{u}} = -2\langle \vec{u}, \vec{v} \rangle H_{\vec{u}} + H_{\vec{v}}.$$

⁷ Pot semblar estrany que donem un nom com **Spin**(3) a un grup que no és altra cosa que el grup **SU**(2). Quan passem a dimensions superiors entendrem millor el motiu de tot això.

- Si $U \in \mathbf{SU}(2)$, existeix un vector unitari $\vec{u} \in \mathbb{R}^3$ i un angle $\alpha \in [0, \pi]$ tals que $U = \cos \alpha I + \sin \alpha H_{\vec{u}}$. Per comprovar això, observem que la primera columna de U és un espinor $(a + bi, c + di)$ amb $a^2 + b^2 + c^2 + d^2 = 1$. Aleshores, la identitat anterior és certa amb $\cos \alpha = a$ i $\vec{u} = (\sin \alpha)^{-1}(b, c, -d)$.

La primera propietat de la llista anterior ens diu que cada matriu de $\mathbf{SU}(2)$ i, per tant, cada espinor de $\mathbf{Spin}(3)$ dona lloc, per conjugació, a una aplicació lineal $\mathbb{R}^3 \rightarrow \mathbb{R}^3$. Calculem aquesta aplicació. Sigui $U \in \mathbf{SU}(2)$, i expressem U en la forma $U = \cos \alpha I + \sin \alpha H_{\vec{u}}$, on $\vec{u}$ és unitari. Volem calcular

$$(\cos \alpha I + \sin \alpha H_{\vec{u}})H_{\vec{v}}(\cos \alpha I + \sin \alpha \overline{H_{\vec{u}}})^t.$$

Tenim, aplicant les propietats que acabem d'explicar:

$$\begin{aligned} U H_{\vec{v}} U^{-1} &= (\cos \alpha I + \sin \alpha H_{\vec{u}})H_{\vec{v}}(\cos \alpha I + \sin \alpha \overline{H_{\vec{u}}})^t = \\ &= \cos^2 \alpha H_{\vec{v}} + \sin 2\alpha H_{\vec{u} \times \vec{v}} - \sin^2 \alpha H_{\vec{u}} H_{\vec{v}} H_{\vec{u}} = \\ &= \cos 2\alpha H_{\vec{v}} + \sin 2\alpha H_{\vec{u} \times \vec{v}} + 2 \sin^2 \alpha \langle \vec{u}, \vec{v} \rangle H_{\vec{u}} = H_{\vec{w}}. \end{aligned}$$

Comparant amb la fórmula d'Olinde Rodrigues, obtenim $\vec{w} = R_{2\alpha, \vec{u}}(\vec{v})$. En conclusió:

Hi ha un epimorfisme $\pi_3: \mathbf{Spin}(3) \rightarrow \mathbf{SO}(3)$ segons el qual l'espinor

$$[\cos \alpha + ix \sin \alpha, y \sin \alpha - iz \sin \alpha]$$

dona la rotació d'angle 2α amb vector director $\vec{u} = (x, y, z)$.

Clarament, el nucli de π_3 té ordre 2 i està format per les matrius $\pm I \in \mathbf{SU}(2)$. Aquest homomorfisme ens dona una excel·lent descripció topològica de $\mathbf{SO}(3)$. Recordem que $\mathbf{Spin}(3) \cong \mathbf{SU}(2) \cong S^3$. Per tant, hi ha un homeomorfisme $\mathbf{SO}(3) \cong S^3 / \{\pm 1\} = \mathbb{R}P(3)$ entre el grup de les rotacions de l'espai de dimensió 3 i l'espai projectiu real de dimensió 3. Deduïm que $\mathbf{SO}(3)$ no és *simplement connex*: el seu grup fonamental té dos elements i el seu recobriment universal és l'esfera S^3 . A més, ara que coneixem el grup fonamental de $\mathbf{SO}(3)$, podem demostrar per inducció —utilitzant la identificació $\mathbf{SO}(n) / \mathbf{SO}(n-1) \cong S^{n-1}$ — que el grup fonamental de *tots* els grups $\mathbf{SO}(n)$ per $n > 2$ és el grup de dos elements.

No és difícil demostrar que el recobriment universal d'un grup topològic admet una estructura canònica de grup topològic compatible amb l'aplicació de recobriment. Això ens permet *definir* el grup dels espinors $\mathbf{Spin}(n)$ com el recobriment universal del grup de les rotacions $\mathbf{SO}(n)$. Ara bé, aquesta definició ens diu ben poc sobre l'estructura d'aquests nous grups: convé seguir llegint aquest treball.

Després de tot això que hem vist, podem pensar els espinors —d'una manera molt informal— com si fossin «rotacions amb signe». Per cada rotació hi ha dos espinors que la defineixen, i aquests dos espinors difereixen en un signe.

Digressió # 1: el cinturó de Conway

Com que el grup fonamental de $\text{SO}(3)$ té dos elements, ha d'existir algun camí ω de $\text{SO}(3)$ que comenci i acabi a I tal que $2\omega \sim *$ però $\omega \not\sim *$.⁸ L'existència d'aquest camí es pot visualitzar molt bé amb el que es coneix com a *cinturó de Conway* (vegeu la figura 1):

Imaginem una cinta llarga i estreta —com un cinturó— de manera que un dels seus extrems està fixat a l'origen de $\mathbb{R}^3$ i l'altre extrem pot estar situat a qualsevol punt de $\mathbb{R}^3$. Suposem, per exemple, que la cinta és com el rectangle $[0, 2] \times [0, 100]$. En cada instant de temps $t \in [0, 100]$, considerem una base ortonormal positiva de $\mathbb{R}^3$ situada al punt $(1, t)$ de la cinta tal que el primer vector apunta segons l'eix de la cinta, en la direcció de l'extrem lliure, el segon vector apunta cap al punt $(0, t)$ i el tercer vector és normal a la cinta. Cadascuna d'aquestes bases ortonormals defineix unívocament una rotació de $\mathbb{R}^3$ i la família uniparamètrica de bases al llarg de la cinta defineix un camí de $\text{SO}(3)$ que comença a la identitat.

Comencem amb la cinta plana sobre el pla horitzontal i girem l'extrem lliure 360 graus en sentit positiu respecte de l'eix central de la cinta. La cinta quedarà entortolligada. Això definirà un camí ω de I a I a $\text{SO}(3)$. Si ara, sense canviar l'orientació de l'extrem lliure, intentem desentortolligar la cinta, no podrem fer-ho: el camí que hem construït no és trivial, no es pot deformar cap al camí constant. A continuació, tornem a girar l'extrem lliure 360 graus en el mateix sentit d'abans: hauré construït el camí 2ω . Ara la cinta estarà molt més entortolligada, però només aparentment: fàcilment podrem aconseguir, sense canviar l'orientació de l'extrem lliure de la cinta, retornar la cinta a la seva posició original. El camí 2ω és trivial.

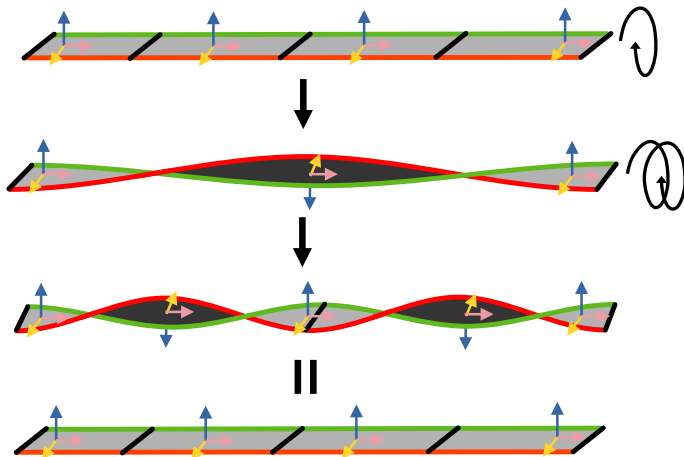


FIGURA 1: El cinturó de Conway.

⁸ Amb el símbol $\sim$ indiquem una homotopia, és a dir, $\alpha_0 \sim \alpha_1$ vol dir que hi ha una família de camins α_t , $t \in [0, 1]$, que depèn contínuament del paràmetre t .

Digressió # 2: els espinors i l'electró

L'electró té *moment angular* com un petit planeta en rotació al voltant d'un eix. En mòdul, el moment angular de tots els electrons té sempre el mateix valor: $\hbar/2$. Però el moment angular és una magnitud vectorial —com les rotacions de l'espai— i, per a cada electró concret, en un instant concret, vindrà donat per un vector $\vec{m}$ en tres dimensions: $\vec{m} = m_x \vec{x} + m_y \vec{y} + m_z \vec{z}$. Pel principi d'incertesa, no podem mesurar simultàniament els tres valors m_x , m_y , m_z . Suposem, doncs, que mesurem m_x . Pel fenomen de la quantització, quan mesurem m_x obtindrem un d'aquests dos valors $\pm \hbar/2$. Segons el valor obtingut, diem que l'electró té *spin* $\pm 1/2$.

Segons el formalisme estàndard de la física quàntica, l'estat inicial del paràmetre m_x es descriu per una funció d'ona Ψ que és combinació lineal dels dos possibles resultats de la mesura: $\Psi = \zeta_1 \Psi_+ + \zeta_2 \Psi_-$, on ζ_1 , ζ_2 són nombres complexos i $||\zeta_1||^2$, $||\zeta_2||^2$ són les probabilitats de cadascun dels dos resultats. En particular, $||\zeta_1||^2 + ||\zeta_2||^2 = 1$ i, per tant, (ζ_1, ζ_2) és un *espinor*. És en aquest sentit que diem que el moment angular de l'electró no és un vector (una rotació), sinó que és un espinor (una «rotació amb signe»).

4 Rotacions en dimensió 3 i quaternions

Recordem que l'existència dels nombres complexos ens permet donar una versió més compacta de les rotacions del pla, *sense sinus ni cosinus*: els nombres complexos ens van permetre passar de la descripció real de les rotacions com a

$$R_\theta(a, b) = (\cos \theta a - \sin \theta b, \sin \theta a + \cos \theta b)$$

a la descripció de les rotacions com a *multiplicar per un nombre complex de mòdul 1*. És possible que hi hagi una estructura *més enllà dels complexos* que tingui un rol similar a les rotacions de l'espai? La resposta és sí i l'estructura és l'anell de divisió⁹ dels *quaternions*, que discutirem ara.

Probablement, la millor manera de definir el cos dels nombres complexos és

$$\mathbb{C} := \left\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} : a, b \in \mathbb{R} \right\} \subset \mathcal{M}_{2 \times 2}(\mathbb{R}),$$

de manera que $i \in \mathbb{C}$ s'identifica a la matriu $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, cada nombre real $a \in \mathbb{R}$ s'identifica a la matriu $\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$ i, aleshores, $a + ib \in \mathbb{C}$ és la matriu $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. Amb aquesta definició de $\mathbb{C}$, les propietats de cos són immediates. En particular, el fet que tot element $z \neq 0$ tingui invers es dedueix del fet que el determinant de la matriu corresponent a z és $a^2 + b^2 = ||z||^2$. Aleshores, $||zz'|| = ||z|| ||z'||$.

⁹ Un anell de divisió és un anell amb unitat $1 \neq 0$ en el qual tot element no nul té invers bilateral. Dit informalment, un anell de divisió és un *cos no commutatiu*.

Si ara repetim aquesta estratagema amb $a, b \in \mathbb{C}$, observem que l'existència d'inversos es perd. Si volem obtenir un cos que amplii el cos dels complexos, cal fer una petita modificació utilitzant la conjugació de $\mathbb{C}$. Amb la definició següent sí que obtenim una $\mathbb{R}$ -àlgebra de divisió de dimensió 4:

$$\mathbb{H} := \left\{ \begin{pmatrix} z & -\bar{w} \\ w & \bar{z} \end{pmatrix} : z, w \in \mathbb{C} \right\} \subset \mathcal{M}_{2 \times 2}(\mathbb{C}).$$

Els elements de $\mathbb{H}$ s'anomenen *quaternions* i compleixen aquestes propietats senzilles de demostrar:

- Si identifiquem $z \in \mathbb{C}$ amb $\begin{pmatrix} z & 0 \\ 0 & \bar{z} \end{pmatrix}$, tenim que $\mathbb{C}$ és un subcòs de $\mathbb{H}$. Si definim $j := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathbb{H}$, veiem que $j^2 = -1$ i que $k := ij = -ji$, amb la qual cosa en el pas de $\mathbb{C}$ a $\mathbb{H}$ hem perdut la commutativitat de la multiplicació. Cada $q \in \mathbb{H}$ s'escriu com a $q = z + jw$ amb $z, w \in \mathbb{C}$ i també com a $q = a + ib + jc + kd$ amb $a, b, c, d \in \mathbb{R}$. $\mathbb{H}$ és, doncs, una $\mathbb{R}$ -àlgebra associativa amb unitat, de dimensió 4, no commutativa.
- $\mathbb{H}$ també està dotat d'una conjugació $q \mapsto \bar{q}$ que és $\mathbb{R}$ -lineal, generalitza la conjugació de $\mathbb{C}$ i compleix $\overline{pq} = \bar{q} \bar{p}$ —és a dir, és un *antiautomorfisme* de $\mathbb{H}$. El conjugat de $a + ib + jc + kd$ és $a - ib - jc - kd$. Interpretant —com hem fet— els quaternions com a matrius complexes 2×2 , si q és la matriu H , aleshores $\bar{q}$ és la matriu H^* .
- L'existència d'inversos per a cada $q \neq 0$ es garanteix pel fet que la matriu corresponent a q té determinant $\|q\|^2$, i aquesta interpretació ens dona la multiplicativitat de la norma a $\mathbb{H}$: $\|pq\| = \|p\| \|q\|$. Tenim, doncs, que $\mathbb{H}$ és un anell de divisió i una $\mathbb{R}$ -àlgebra normada.
- Una altra manera d'escriure un quaternió q és en la forma $q = \lambda + \vec{u}$, on $\lambda = (q + \bar{q})/2 \in \mathbb{R}$ s'anomena *part real* o *part escalar* de q i $\vec{u} = (q - \bar{q})/2 \in \langle i, j, k \rangle_{\mathbb{R}}$ es denomina *part vectorial* de q . Els quaternions amb part escalar zero direm que són *quaternions purs*. Si $\|q\| = 1$, podem escriure

$$q = \cos \alpha + \sin \alpha \vec{u},$$

on $\alpha \in [0, \pi]$ i $\vec{u} \in \mathbb{H}$ és pur de norma 1, unívocament determinat si $q \notin \mathbb{R}$.

- Si u, v són quaternions purs, es compleix

$$uv = -\langle u, v \rangle + u \times v,$$

en què $u \times v$ s'entén mirant u, v com a vectors en la base ortonormal i, j, k .

- **Fórmula d'anticommutació.** Una identitat trivial molt útil és

$$p \bar{q} + q \bar{p} = 2\langle p, q \rangle,$$

que té com a conseqüència que dos quaternions p, q són ortogonals (com a vectors de $\mathbb{R}^4$) si compleixen $p \bar{q} = -q \bar{p}$. D'altra banda, si $q \in \mathbb{H}^*$ i

considerem la conjugació $c_q: \mathbb{H} \rightarrow \mathbb{H}$ com un isomorfisme $\mathbb{R}$ -lineal de $\mathbb{R}^4$, la fórmula anterior ens permet comprovar que c_q conserva el producte escalar de $\mathbb{R}^4$ i és, per tant, una aplicació *ortogonal*.

- Si $\lambda \in \mathbb{R}$ és la part escalar de q , es compleix aquesta identitat:

$$q^2 - 2\lambda q + ||q||^2 = 0.$$

En particular, les solucions de $x^2 = -1$ són tots els quaternions unitaris purs, és a dir, els punts de l'esfera unitat de $\langle i, j, k \rangle_{\mathbb{R}} \cong \mathbb{R}^3$.¹⁰

- Trivialment, el centre de l'anell de divisió $\mathbb{H}$ és $\mathbb{R}$. Això ens permetrà determinar fàcilment el grup d'automorfismes de $\mathbb{H}$ (pàgina 21).

Àlgebra lineal sobre els quaternions

La teoria bàsica dels espais vectorials s'estén sense problemes al cas que els escalars no formin un cos, sinó un anell de divisió. Per tant, podem fer àlgebra lineal amb els quaternions com a escalars i, per abús de llenguatge, parlarem de $\mathbb{H}$ -*espais vectorials*. Cal només tenir presents alguns petits detalls:

- La no commutativitat de $\mathbb{H}$ dona importància a la distinció entre $\mathbb{H}$ -espais vectorials per la dreta i $\mathbb{H}$ -espais vectorials per l'esquerra, segons que el producte d'un escalar $q \in \mathbb{H}$ i un vector $\vec{v}$ s'escriu $\vec{v}q$ o $q\vec{v}$. En el nostre cas, considerarem que tots els $\mathbb{H}$ -espais vectorials són per l'esquerra.
- Si $\varphi: \mathbb{H}^n \rightarrow \mathbb{H}^n$ és una aplicació $\mathbb{H}$ -lineal bijectiva, l'aplicació φ^{-1} també és $\mathbb{H}$ -lineal i podem parlar, per tant, del grup lineal $GL(n, \mathbb{H})$.
- Si A és una matriu $n \times n$ amb coeficients a $\mathbb{H}$, l'aplicació $\vec{v} \mapsto A\vec{v}$ amb $\vec{v} \in \mathbb{H}^n$ no és $\mathbb{H}$ -lineal, però $\vec{v} \mapsto \vec{v}A$ sí que ho és. En conseqüència, en un $\mathbb{H}$ -espai vectorial per l'esquerra, les matrius actuen per la dreta sobre vectors fila. Si ho fem així, tenim isomorfisme entre les matrius $n \times n$ i les aplicacions $\mathbb{H}$ -lineals de $\mathbb{H}^n$ a $\mathbb{H}^n$ (que formen un $\mathbb{H}$ -espai vectorial *per la dreta*) i els isomorfismes lineals $\mathbb{H}^n \rightarrow \mathbb{H}^n$ s'identifiquen a les matrius invertibles $n \times n$.
- Com que la multiplicació de $\mathbb{H}$ no és commutativa, la teoria clàssica del determinant deixa de ser vàlida. Per exemple, $M = \begin{pmatrix} i & j \\ j & i \end{pmatrix}$ és clarament invertible ($M^2 = -2I$) però, si intentem calcular-ne el determinant (multiplicant els factors en qualsevol ordre), sempre obtenim zero.¹¹
- Quan treballem amb matrius quaterniòniques, ni la transposició ni la conjugació no respecten el producte de matrius. En canvi, la conjugació seguida de la transposició sí que és un antiautomorfisme: $(AB)^* = B^*A^*$.

¹⁰ Veiem, doncs, que el conegut fet que el nombre d'arrels d'un polinomi no pot superar el grau del polinomi deixa de ser cert si la multiplicació no és commutativa.

¹¹ Per a una discussió més a fons sobre l'extensió del determinant complex a un determinant quaterniònic, vegeu N. Cohen, S. de Leo, «The quaternionic determinant», *Electron. J. Linear Algebra*, 7 (2000), 100-111.

- Quan els coeficients de les matrius pertanyen a un domini no commutatiu pot passar que una matriu tingui inversa per un costat i no per l'altre, o que en tingui pels dos costats, però siguin diferents. Aquests fets patològics no passen en el cas de les matrius quaterniòniques, precisament perquè podem identificar les matrius amb les aplicacions lineals i una aplicació lineal $\mathbb{H}^n \rightarrow \mathbb{H}^n$ exhaustiva o injectiva és automàticament bijectiva.
- Per decidir si una matriu quaterniònica és invertible i per calcular-ne la inversa, podem mirar l'àlgebra $\mathcal{M}_{n \times n}(\mathbb{H})$ com a subàlgebra de $\mathcal{M}_{2n \times 2n}(\mathbb{C})$ de la manera següent: si $A \in \mathcal{M}_{n \times n}(\mathbb{H})$, A defineix una aplicació $\mathbb{H}$ -lineal $A: \mathbb{H}^n \rightarrow \mathbb{H}^n$, que ens podem mirar com a aplicació $\mathbb{C}$ -lineal $A_{\mathbb{C}}: \mathbb{C}^{2n} \rightarrow \mathbb{C}^{2n}$. Aleshores, A és invertible si i només si $\det A_{\mathbb{C}} \neq 0$. Evidentment, si A^{-1} existeix, també existeix $A_{\mathbb{C}}^{-1}$ i, per tant, $\det(A_{\mathbb{C}}) \neq 0$. Recíprocament, si $\det A_{\mathbb{C}} \neq 0$, aleshores $A_{\mathbb{C}}$ és bijectiva i A també. Per tant, A té inversa.

De cara al que farem més endavant, ens convé explicitar clarament la inclusió de les matrius quaterniòniques en les matrius complexes de mida doble. Denotem amb $\vec{e}_1, \dots, \vec{e}_n$ la base canònica de $\mathbb{H}^n$ i considerem aquesta $\mathbb{C}$ -base de $\mathbb{H}^n$

$$\vec{e}_1, \dots, \vec{e}_n, j\vec{e}_1, \dots, j\vec{e}_n.$$

Recordem que mirem $\mathbb{H}^n$ com a $\mathbb{H}$ -mòdul per l'esquerra i, per tant, les matrius actuen per la dreta. Per coherència, també les matrius complexes actuaran per la dreta. Sigui $A \in \mathcal{M}_{n \times n}(\mathbb{H})$ i escrivim $A = A_1 + A_2 j$ amb $A_1, A_2 \in \mathcal{M}_{n \times n}(\mathbb{C})$. Aleshores, un càlcul senzill mostra que l'aplicació $\mathbb{H}$ -lineal donada per A és la mateixa que l'aplicació $\mathbb{C}$ -lineal donada per la matriu

$$A_{\mathbb{C}} := \begin{pmatrix} A_1 & A_2 \\ -\bar{A}_2 & \bar{A}_1 \end{pmatrix}.$$

És fàcil comprovar que aquesta aplicació $\mathcal{M}_{n \times n}(\mathbb{H}) \rightarrow \mathcal{M}_{2n \times 2n}(\mathbb{C})$ és un homomorfisme d'anells i respecta l'operació $A \mapsto A^*$.

El producte hermític de $\mathbb{C}^n$ es pot generalitzar als vectors de $\mathbb{H}^n$, però ara cal fer atenció a l'ordre dels factors i al fet que les matrius actuen per la dreta:¹²

- $\langle \vec{u} + \vec{u}', \vec{v} \rangle = \langle \vec{u}, \vec{v} \rangle + \langle \vec{u}', \vec{v} \rangle$, $\langle \vec{u}, \vec{v} + \vec{v}' \rangle = \langle \vec{u}, \vec{v} \rangle + \langle \vec{u}, \vec{v}' \rangle$, $\langle p\vec{u}, q\vec{v} \rangle = p \langle \vec{u}, \vec{v} \rangle \bar{q}$;
- $\langle \vec{u}, \vec{v} \rangle = \overline{\langle \vec{v}, \vec{u} \rangle}$;
- $\langle \vec{u}, \vec{u} \rangle = ||\vec{u}||^2$ on $||\vec{u}||$ és el mòdul de $\vec{u}$ com a vector de $\mathbb{R}^{4n}$;
- (no degenerat) $\langle \vec{u}, \vec{v} \rangle = 0$ per a tot $\vec{v}$ si i només si $\vec{u} = 0$;
- $\langle \vec{u}A, \vec{v} \rangle = \langle \vec{u}, \vec{v}A^* \rangle$.

¹² En el cas $n = 1$, la notació $\langle \vec{u}, \vec{v} \rangle$ és ambigua perquè pot fer referència al producte escalar dels quaternions $\vec{u}, \vec{v}$ com a vectors de $\mathbb{R}^4$ o al producte hermític de $\vec{u}, \vec{v}$ com a vectors de $\mathbb{H}^1$.

Els grups simplèctics

Lamentablement, hi ha dues famílies de grups que reben el mateix nom de *grups simplèctics*, com a conseqüència del fet que hi ha dos conceptes diferents de *matriu simplèctica*. Haurem de deduir del context de quina de les dues famílies estem parlant. Per distingir-los, utilitzarem notacions diferents i, a més, com que els grups d'una de les famílies són compactes i els de l'altra no ho són, en cas de dubte direm que parlem del grup simplèctic compacte o del grup simplèctic no compacte.

Els grups simplèctics no compactes. El pas de la geometria euclidiana a la *geometria simplèctica* —important a la mecànica hamiltoniana, l'òptica geomètrica i la mecànica quàntica, entre altres camps— comença substituint el producte escalar (simètric) per un producte antisimètric. Més concretament, suposem que V és un espai vectorial de dimensió finita sobre un cos k de característica zero. Una estructura simplèctica sobre V ve donada per una 2-forma bilineal ω que compleixi

1. $\omega(\vec{v}, \vec{v}) = 0$ per a tot $\vec{v} \in V$.
2. Si $\omega(\vec{u}, \vec{v}) = 0$ per a tot $\vec{v} \in V$, aleshores $\vec{u} = 0$.

La matriu de ω és antisimètrica i invertible i això implica immediatament que la dimensió de V ha de ser parella $2n$. A més, podem trobar una base de V en la qual la matriu de ω tingui la forma $\Omega = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$, on I és la matriu identitat $n \times n$.

De la mateixa manera que a la geometria euclidiana ens interessen els automorfismes que conserven la forma quadràtica estàndard de $\mathbb{R}^n$ —les transformacions ortogonals—, en la geometria simplèctica ens interessen els automorfismes que conserven la forma simplèctica estàndard Ω —en direm, naturalment, *transformacions simplèctiques*.

És senzill de veure que una matriu M és simplèctica si compleix $M^t \Omega M = \Omega$. Aquestes matrius formen un subgrup de $\text{GL}_{2n}(k)$ que denotarem amb $\text{Sp}_{2n}(k)$. Aquests grups s'anomenen *grups simplèctics* i obtenim una cadena de subgrups

$$\text{Sp}_2(k) \subset \text{Sp}_4(k) \subset \text{Sp}_6(k) \subset \dots$$

completant cada matriu simplèctica amb una matriu identitat. Com que les matrius $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix}$ són simplèctiques per a tot $\lambda \neq 0$, veiem que els grups $\text{Sp}_{2n}(\mathbb{R})$, $\text{Sp}_{2n}(\mathbb{C})$ no són compactes.

En el cas particular $n = 1$ és immediat que $\text{Sp}_2(k) = \text{SL}_2(k)$. La igualtat deixa de ser certa en dimensions superiors, però sí que és cert en general que $\text{Sp}_{2n}(k)$ és un subgrup de $\text{SL}_{2n}(k)$. Aquest fet que el determinant de tota matriu simplèctica és igual a 1 no és evident a partir de la condició $M^t \Omega M = \Omega$, però es pot demostrar amb aquest raonament:

Comencem recordant que tota aplicació lineal $\phi: V \rightarrow V$ indueix aplicacions lineals $\phi_i: \bigwedge^i V \rightarrow \bigwedge^i V$ en els productes exteriors. En particular, si V és de dimensió m , l'espai vectorial $\bigwedge^m V$ és de dimensió 1 i l'aplicació $\phi_m: \bigwedge^m V \rightarrow \bigwedge^m V$ és la multiplicació pel determinant de ϕ . Sigui ara V un espai vectorial

simplèctic amb 2-forma $\omega \in V^* \wedge V^*$. Prenem una base $\vec{u}_1, \dots, \vec{u}_n, \vec{v}_1, \dots, \vec{v}_n$ de V en la qual la matriu de ω sigui Ω , és a dir,

$$\omega = \vec{u}_1^* \wedge \vec{v}_1^* + \dots + \vec{u}_n^* \wedge \vec{v}_n^*,$$

on $\vec{u}_1^*, \dots, \vec{u}_n^*, \vec{v}_1^*, \dots, \vec{v}_n^*$ és la base dual de V^* . Sigui M una matriu simplèctica i sigui $\phi: V \rightarrow V$ l'aplicació lineal induïda per M en la base anterior. El fet que M sigui simplèctica es traduirà en el fet que $\phi_2^*(\omega) = \omega$, on $\phi^*: V^* \rightarrow V^*$ és l'aplicació dual. Considerem $\omega^n := \omega \wedge \dots \wedge \omega \in \wedge^{2n}(V^*)$. Aleshores, $\det(\phi) \omega^n = \phi_{2n}^*(\omega^n) = \omega^n$ i tot es redueix a demostrar que $\omega^n \neq 0$. Si observem que, per a tot i, j , els elements $\vec{u}_i^* \wedge \vec{v}_i^*, \vec{u}_j^* \wedge \vec{v}_j^* \in \wedge^2(V^*)$ commuten, veurem immediatament que

$$\omega^n = \vec{u}_1^* \wedge \vec{v}_1^* \wedge \dots \wedge \vec{u}_n^* \wedge \vec{v}_n^* \neq 0. \quad \square$$

Els grups simplèctics compactes. El producte hermític de $\mathbb{H}^n$ ens permet estendre els conceptes de matriu ortogonal (real) i matriu unitària (complexa) al cas quaterniònic. En direm *matrius simplèctiques* (però no les hem de confondre amb les matrius simplèctiques de l'apartat anterior): una matriu $A \in \mathcal{M}_{n \times n}(\mathbb{H})$ és simplèctica (en aquest nou context) si conserva el producte hermític:

$$\langle \vec{u}A, \vec{v}A \rangle = \langle \vec{u}, \vec{v} \rangle \text{ per tot } \vec{u}, \vec{v} \in \mathbb{H}^n.$$

Vegem algunes primeres propietats de les matrius simplèctiques.

- Clarament una matriu simplèctica representa una aplicació bijectiva i, per tant, és una matriu invertible, i la inversa també és simplèctica. Per tant, les matrius simplèctiques formen un grup que designarem $\mathbf{Sp}(n)$.
- El grup $\mathbf{Sp}(1)$ s'identifica al grup multiplicatiu dels quaternions de mòdul 1 i, per la mateixa definició de $\mathbb{H}$, aquest grup és $\mathbf{SU}(2)$. Topològicament, es tracta de S^3 , l'esfera de dimensió 3.
- Si $A \in \mathbf{Sp}(n)$, A conserva la norma de $\mathbb{H}^n$, i recíprocament.
- Igual que en el cas del grup unitari $\mathbf{U}(n)$, les matrius simplèctiques estan caracteritzades per la propietat $AA^* = I$, que és equivalent a $A^* = A^{-1}$ i és equivalent a $A^*A = I$.
- L'aplicació $\mathcal{M}_{n \times n}(\mathbb{H}) \rightarrow \mathcal{M}_{2n \times 2n}(\mathbb{C})$ que hem construït anteriorment dona una inclusió de grups $\mathbf{Sp}(n) \subset \mathbf{U}(2n)$. En particular, els grups $\mathbf{Sp}(n)$ són compactes: ja sabem que $\mathbf{U}(2n)$ és compacte i és clar que $\mathbf{Sp}(n)$ és tancat (es pot definir a partir d'equacions algebraïques).
- Els grups $\mathbf{Sp}(n)$ són connexos i simplement connexos. La demostració és la mateixa que hem indicat a la pàgina 9 i utilitza $\mathbf{Sp}(n) / \mathbf{Sp}(n-1) \cong S^{4n-1}$.

En resum:

El mateix concepte de matriu que conserva el producte intern, aplicat a les tres àlgebres de divisió $\mathbb{R}$, $\mathbb{C}$ i $\mathbb{H}$, ens proporciona tres famílies infinites de grups compactes: $\mathbf{O}(n)$, $\mathbf{U}(n)$, $\mathbf{Sp}(n)$, $n \geq 1$.

Relació entre les dues famílies de grups simplèctics. Entre el grup simplèctic no compacte $\mathrm{Sp}_{2n}(\mathbb{C})$ de les matrius $M \in \mathcal{M}_{2n \times 2n}(\mathbb{C})$ tals que $M^t \Omega M = \Omega$ i el grup simplèctic compacte $\mathbf{Sp}(n) \subset \mathrm{U}(2n)$ de les matrius $M \in \mathcal{M}_{n \times n}(\mathbb{H})$ tals que $MM^* = I$ hi ha aquesta relació:

$$\mathbf{Sp}(n) = \mathrm{Sp}_{2n}(\mathbb{C}) \cap \mathrm{U}(2n) = \mathrm{Sp}_{2n}(\mathbb{C}) \cap \mathbf{SU}(2n).$$

Demostrem aquesta identitat. Si tenim la matriu $A = A_1 + A_2 j \in \mathbf{Sp}(n)$, considerem la matriu $A_{\mathbb{C}} \in \mathrm{U}(2n)$ que hem definit a la pàgina 17 Aleshores, $\overline{A_{\mathbb{C}}}^t A_{\mathbb{C}} = I$ ens dona

$$\overline{A_1}^t A_1 + A_2^t \overline{A_2} = \overline{A_2}^t A_2 + A_1^t \overline{A_1} = I, \quad \overline{A_1}^t A_2 - A_2^t \overline{A_1} = \overline{A_2}^t A_1 - A_1^t \overline{A_2} = 0.$$

D'altra banda,

$$A_{\mathbb{C}}^t \Omega A_{\mathbb{C}} = \begin{pmatrix} \overline{A_2}^t A_1 & -A_1^t \overline{A_2} & \overline{A_2}^t A_2 + A_1^t \overline{A_1} \\ -\overline{A_1}^t A_1 & -A_2^t \overline{A_2} & -\overline{A_1}^t A_2 + A_2^t \overline{A_1} \end{pmatrix} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

i hem demostrat que $A_{\mathbb{C}} \in \mathrm{Sp}_{2n}(\mathbb{C})$. Recíprocament, si $M \in \mathrm{Sp}_{2n}(\mathbb{C}) \cap \mathrm{U}(2n)$, tenim $M^t \Omega M = \Omega$ i, al mateix temps, $M^{-1} = M^*$. Deduïm que $\Omega^t M = \overline{M} \Omega^t$ i això implica que M és de la forma $A_{\mathbb{C}}$ per a alguna matriu $A \in \mathbf{Sp}(n)$. $\square$

En conclusió:

Si volem prescindir dels quaternions, podem definir el grup simplèctic $\mathbf{Sp}(n)$ com el grup format per les matrius complexes $2n \times 2n$ que conserven, simultàniament, la forma hermitica i la forma simplèctica.

Els quaternions i les rotacions de $\mathbb{R}^3$

Abans hem vist que els espinors —que hem identificat amb els elements del grup $\mathbf{SU}(2)$ — ens donen les rotacions de l'espai de dimensió 3. Ara hem pogut identificar els espinors amb els quaternions de norma 1. En aquest apartat integrarem aquests dos fets i veurem com el grup multiplicatiu dels quaternions $\mathbb{H}^*$ ens dona una excel·lent descripció del grup de les rotacions $\mathbf{SO}(3)$.

A l'apartat *Espinors i rotacions de l'espai*, de la pàgina 11, hem vist, d'una manera molt explícita, com els elements de $\mathbf{SU}(2)$ actuen per conjugació sobre $\mathcal{H}'_0$, un $\mathbb{R}$ -espai vectorial de dimensió 3. D'altra banda, com que $\mathbb{R}$ és el centre de $\mathbb{H}$, i sabem (pàgina 16) que la conjugació per un quaternió $q \neq 0$ conserva el producte escalar de $\mathbb{R}^4$, tindrem que c_q dona un automorfisme ortogonal de $\langle i, j, k \rangle_{\mathbb{R}} \subset \mathbb{H}$, el subespai (real) dels quaternions purs. Observem que aquest subespai de $\mathbb{H}$, si el mirem dins de $\mathcal{M}_{2 \times 2}(\mathbb{C})$, com hem fet quan hem definit els quaternions, coincideix amb $\mathcal{H}'_0$. D'aquesta manera, l'acció dels quaternions de norma 1 per conjugació sobre $\langle i, j, k \rangle_{\mathbb{R}}$ és exactament la mateixa que l'acció de $\mathbf{SU}(2)$ sobre $\mathcal{H}'_0$ que hem estudiat detalladament a l'apartat 3.

En definitiva, de la mateixa manera que els nombres complexos ens van proporcionar una manera molt més simple de descriure les rotacions del pla

$$z \mapsto zz_0 \quad \text{vs.} \quad \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix},$$

també els quaternions ens donen una descripció molt més simple de les rotacions de l'espai de tres dimensions:

$$h \mapsto p_0 h p_0^{-1} \quad \text{vs.} \quad R_{\theta, \vec{u}}(\vec{v}) = \cos \theta \vec{v} + \sin \theta (\vec{u} \times \vec{v}) + (1 - \cos \theta) \langle \vec{u}, \vec{v} \rangle \vec{u}.$$

Fem ara explícita la relació entre un quaternió unitari $q \in \mathbb{H}$ i una rotació descrita geomètricament en la forma $R_{\theta, \vec{u}}$:

$$q = \cos \alpha + \sin \alpha (x i + y j + z k) \rightsquigarrow R_{2\alpha, \overline{(x, y, z)}},$$

entenent, però, que a cada rotació de $\text{SO}(3)$ li corresponen dos quaternions (o dos espinors) $\pm q$. Observem també que l'eix de la rotació associada a un quaternió p és simplement el vector $p - \bar{p} \in \langle i, j, k \rangle_{\mathbb{R}} \subset \mathbb{H}$ perquè $p(p - \bar{p})p^{-1} = p - \bar{p}$.

Digressió # 3: automorfismes de $\mathbb{H}$

Ara podem descriure fàcilment els automorfismes de $\mathbb{H}$. Curiosament, el comportament patològic de $\mathbb{C}$ respecte dels automorfismes —recordem que l'axioma de l'elecció ens permet demostrar l'existència d'una infinitat no numerable d'automorfismes de $\mathbb{C}$ — no es generalitza al cas dels quaternions. El motiu rau en el fet que $\mathbb{R}$ és el *centre* de $\mathbb{H}$ i qualsevol automorfisme (i també qualsevol antiautomorfisme) ha de respectar el centre. Per tant, si $\phi: \mathbb{H} \rightarrow \mathbb{H}$ és un (anti)automorfisme, aleshores ϕ és $\mathbb{R}$ -lineal i, en conseqüència, és continu i ve donat per una matriu real 4×4 . Per estudiar els automorfismes de $\mathbb{H}$ ens convé considerar simultàniament els automorfismes i els antiautomorfismes. Observem que, en tot cas, ϕ ha de conservar les solucions de l'equació $q^2 = -1$ i aquestes solucions (pàgina 16) formen l'esfera unitat dins de $\langle i, j, k \rangle_{\mathbb{R}} \subset \mathbb{H} = \mathbb{R}^4$. Per tant, ϕ donarà un isomorfisme lineal de $\mathbb{R}^3$ que conservarà la norma i, per tant, vindrà donat per una matriu ortogonal. Si ϕ té determinant 1, serà una rotació i, segons hem vist, ϕ serà la conjugació per un quaternió de norma 1. Si ϕ té determinant -1 , tindrem que $c\phi$ serà la conjugació per un quaternió de norma 1, on c és l'antiautomorfisme de $\mathbb{H}$ donat per la conjugació. En conclusió, tots els automorfismes de $\mathbb{H}$ són interns i tots els antiautomorfismes seran producte d'un automorfisme intern per la conjugació c . Pel que hem vist abans, deduïm que els automorfismes de $\mathbb{H}$ formen el grup $\text{SO}(3)$.

Rotacions en dimensió 3: conclusió

L'estudi de les rotacions de $\mathbb{R}^3$ ens ha dut a tres descripcions d'un mateix objecte

$$\text{SU}(2) \cong \text{Spin}(3) \cong \text{Sp}(1),$$

el recobriment universal de $\mathbf{SO}(3)$, un grup amb un subgrup normal $\{\pm I\}$ amb quocient $\mathbf{SO}(3)$, de manera que, topològicament, obtenim el recobriment universal

$$S^3 \rightarrow \mathbb{R}P^3.$$

Com hem dit abans, donar tres noms diferents a un mateix grup quedarà justificat quan vegem, més endavant, que les tres famílies $\mathbf{SU}(n)$, $\mathbf{Spin}(n)$ —que hem definit com el recobriment universal de $\mathbf{SO}(n)$ — i $\mathbf{Sp}(n)$ divergeixen quan n creix.

Digressió # 4: l'aplicació de Hopf

El grup $\mathbf{Sp}(1)$ està format pels quaternions de mòdul 1 i, per tant, admet com a subgrup el grup $\mathbf{U}(1)$, format pels complexos de mòdul 1. Això ens permet considerar l'espai de les classes laterals $\mathbf{Sp}(1)/\mathbf{U}(1)$. Topològicament, $\mathbf{Sp}(1)$ s'identifica a l'esfera S^3 i $\mathbf{U}(1)$ s'identifica a la circumferència S^1 , que actua per la dreta sobre S^3 . Aleshores, $\mathbf{Sp}(1)/\mathbf{U}(1)$ és un espai topològic que podem identificar fàcilment de la manera següent. Pensem S^2 com l'esfera unitat dins dels quaternions purs i escollim com a punt base el punt $i \in S^2 \subset \langle i, j, k \rangle_{\mathbb{R}}$. Aleshores, la representació $S^3 = \mathbf{Spin}(3) \rightarrow \mathbf{SO}(3)$ ens dona una acció ortogonal de S^3 sobre $\mathbb{R}^3$ i, en conseqüència, una acció de S^3 sobre l'esfera $S^2 \subset \mathbb{R}^3$. Aleshores, si $q \in S^3$, podem considerar $qi\bar{q} \in S^2$. En conclusió, tenim una aplicació contínua

$$\eta: S^3 \rightarrow S^2,$$

que es coneix com a *aplicació de Hopf* i té propietats remarcables com aquestes:

- Ens permet identificar $\mathbf{Sp}(1)/\mathbf{U}(1)$ amb l'esfera S^2 . En particular, l'aplicació η és un *fibrat principal* i una *fibració* i, localment, es comporta com un producte de S^2 per S^1 , que és la *fibra*. Geomètricament, podem descompondre l'esfera tridimensional com a unió disjunta de circumferències, una per a cada punt de l'esfera 2-dimensional. A Internet¹³ hi podeu trobar moltes representacions gràfiques magnífiques d'aquesta descomposició.
- η va ser el primer exemple d'una aplicació homotòpicament no trivial entre esferes de dimensions diferents. Heinz Hopf va poder demostrar que tota aplicació contínua de S^3 a S^2 és homotòpicament equivalent a un *múltiple* enter de η , en un sentit que ara no explicitem, i, d'aquesta manera, va demostrar que $\pi_3(S^2) = \mathbb{Z}\eta$, el primer càlcul d'un grup d'homotopia d'esferes $\pi_k(S^n)$ amb $k > n$. S'ha afirmat que aquest resultat de Hopf representa el moment fundacional de la *teoria d'homotopia*.

Hi ha una senzilla relació entre l'aplicació de Hopf i la *geometria projectiva*. Recordem que, si $V \neq 0$ és un k -espai vectorial —encara que k sigui només un anell de divisió—, podem definir el seu espai projectiu $P(V)$ com el *conjunt*

¹³ Vegeu, per exemple, aquesta figura de l'entrada «Hopf fibration» de la Wikipedia.

dels subespais de dimensió 1 de V , i aquest espai $P(V)$ és l'àmbit natural de la geometria lineal. Si $k = \mathbb{R}, \mathbb{C}$, obtenim els espais projectius reals i complexos

$$\begin{aligned} * &= \mathbb{R}P^0 \subset \mathbb{R}P^1 \subset \mathbb{R}P^2 \subset \mathbb{R}P^3 \subset \mathbb{R}P^4 \subset \dots, \\ * &= \mathbb{C}P^0 \subset \mathbb{C}P^1 \subset \mathbb{C}P^2 \subset \mathbb{C}P^3 \subset \mathbb{C}P^4 \subset \dots \end{aligned}$$

i, si $k = \mathbb{H}$, obtenim els espais projectius quaterniònics

$$* = \mathbb{H}P^0 \subset \mathbb{H}P^1 \subset \mathbb{H}P^2 \subset \mathbb{H}P^3 \subset \mathbb{H}P^4 \subset \dots,$$

en els quals la geometria lineal es comporta pràcticament igual que en els casos clàssics reals i complexos llevat que, degut a la no commutativitat de la multiplicació a $\mathbb{H}$, no es compleix el *teorema de Pappos*. En els tres casos anteriors, podem identificar la recta projectiva amb la compactificació per un punt de la recta afi, és a dir, amb l'esfera de la dimensió corresponent: $\mathbb{R}P^1 \cong S^1$, $\mathbb{C}P^1 \cong S^2$, $\mathbb{H}P^1 \cong S^4$.

En els tres casos real, complex i quaterniònic, cada subespai de dimensió 1 de k^{n+1} està determinat per un vector unitari, i dos vectors unitaris defineixen el mateix subespai si un s'obté de l'altre multiplicant per un escalar de mòdul 1. Això ens permet identificar kP^n amb el quocient (topològic) de l'esfera unitat de k^{n+1} per l'acció de l'esfera unitat de k :

$$\mathbb{R}P^n = S^n / \pm 1, \quad \mathbb{C}P^n = S^{2n+1} / S^1, \quad \mathbb{H}P^n = S^{4n+3} / S^3.$$

Per a $n = 1$ obtenim

$$S^1 \cong \mathbb{R}P^1 = S^1 / \pm 1, \quad S^2 \cong \mathbb{C}P^1 = S^3 / S^1, \quad S^4 \cong \mathbb{H}P^1 = S^7 / S^3.$$

En el cas real obtenim $S^1 \rightarrow S^1$, que és l'aplicació $z \mapsto z^2$ que es denota per 2ι . En el cas complex obtenim l'aplicació de Hopf η i en el cas quaterniònic obtenim una generalització de l'aplicació de Hopf

$$\nu: S^7 \rightarrow S^4,$$

que també és una aplicació no homotòpicament trivial significativa.

Precisament la relació que hem comentat entre les aplicacions de Hopf 2ι , η i ν i la geometria projectiva ens dona la clau de la seva no trivialitat homotòpica. Considerem, per exemple, el cas de $\eta: S^3 \rightarrow S^2$. No és difícil veure que el pla projectiu $\mathbb{C}P^2$ s'obté adjuntant a la recta projectiva $\mathbb{C}P^1 = S^2$ una bola de dimensió 4 a través de l'aplicació η . Aleshores, si η fos homotòpica a l'aplicació constant, tindriem que $\mathbb{C}P^2$ seria homotòpicament equivalent a $S^2 \vee S^4$. Això és impossible perquè la cohomologia de $\mathbb{C}P^2$ i la cohomologia de $S^2 \vee S^4$ són diferents.

La importància que l'aplicació de Hopf va tenir en el desenvolupament de la teoria d'homotopia va fer que es busquessin generalitzacions a dimensions superiors, del tipus $S^{2n+1} \rightarrow S^{n+1}$ amb fibra S^n . Per a $n = 0, 1, 3$ els plans projectius real, complex i quaterniònic ens donen aplicacions com aquesta i

la pregunta va ser si es podia anar més enllà. Resulta que per construir una aplicació de tipus Hopf $S^{2n+1} \rightarrow S^{n+1}$ n'hi ha prou que la presumpta fibra S^n tingui una *multiplicació contínua amb unitat* i la construcció és tan senzilla com això:

- Considerem aquestes dues construccions topològiques elementals:

- La *suspensió*. Si X és un espai, la seva suspensió és l'espai

$$\Sigma X := (X \times [0, 1]) / \{(x, 0) \sim (x', 0), (x, 1) \sim (x', 1) : x, x' \in X\}.$$

- La *juntura*. Si X, Y són espais, la seva juntura és l'espai

$$X * Y := (X \times Y \times [0, 1]) / \mathcal{R};$$

$$\mathcal{R} := \{(x, y, 0) \sim (x, y', 0), (x, y, 1) \sim (x', y, 1) : x, x' \in X, y, y' \in Y\}.$$

Si Y és l'espai amb un únic punt, aleshores $X * Y$ es coneix com el *con* de X , i es denota CX .

- La *cofibra homotòpica*. Si $f: X \rightarrow Y$ és una aplicació contínua, la cofibra homotòpica de f és l'espai quocient

$$(Y \cup_f CX) := (X \cup CY) / \{(x, 0) \sim f(x) : x \in X\}.$$

El cas més interessant és quan X és una esfera S^n i CX és una bola tancada e^{n+1} . En aquest cas, diem que la cofibra homotòpica de $f: S^n \rightarrow Y$ és l'espai que s'obté *adjuntant* a Y una *cel·la* de dimensió $n + 1$.

- No és difícil comprovar que les dues operacions anteriors transformen esferes en esferes (llevat d'homeomorfisme): $\Sigma S^n \cong S^{n+1}$, $S^n * S^m \cong S^{n+m+1}$.
- Si S^n admet una multiplicació amb unitat, podem considerar l'aplicació

$$f: S^{2n+1} \cong S^n * S^n \rightarrow \Sigma S^n \cong S^{n+1}, \quad f([x, y, t]) := [xy, t],$$

que, per a $n = 0, 1, 3$, és, llevat d'homotopia, l'aplicació de Hopf. A més, si utilitzem f per adjuntar a S^{n+1} una cel·la e^{2n+2} , obtenim un espai $S^{n+1} \cup_f e^{2n+2}$, que podem interpretar com el *pla projectiu* associat a S^n .

En particular, si S^n és un grup —com ho són $S^0 = \{\pm 1\}$, $S^1 = \mathbf{U}(1)$ i $S^3 = \mathbf{Sp}(1)$ —, ja tenim l'aplicació de Hopf $S^{2n+1} \rightarrow S^{n+1}$ que volíem, però no cal que S^n compleixi tots els axiomes de grup, només el que afirma l'existència de l'element neutre! El cas és que un teorema clàssic de Frank Adams —molt difícil, si bé s'han trobat demostracions força més senzilles que l'original d'Adams— afirma que les úniques esferes que poden tenir una multiplicació contínua amb unitat són S^0 , S^1 , S^3 i S^7 i, en conseqüència, les úniques aplicacions de tipus Hopf que poden existir són les tres que ja coneixem: 2ι , η , ν i una hipotètica $\sigma: S^{15} \rightarrow S^8$ amb fibra S^7 , si és que realment S^7 —que no sembla que puguem identificar amb cap grup conegut— admet com a mínim una multiplicació contínua amb unitat.¹⁴

¹⁴ S^7 sí que admet aquesta multiplicació, però aquest fet —de conseqüències immenses— el discutirem en algun altre lloc.

5 Rotacions en dimensió 4

Volem estudiar ara les rotacions de l'espai euclidià $\mathbb{R}^4$ amb la intenció de trobar, potser, alguna relació entre el grup **Spin**(4) —que hem definit de manera abstracta com el recobriment universal de **SO**(4)— i algun grup conegut. Com veurem, l'estudi del grup **SO**(4) és molt bonic, i també és ben singular perquè **Spin**(4) és l'únic grup d'espinors que té una descomposició no trivial com a producte cartesià.

Considerem l'aplicació

$$\pi_4: \mathbf{Sp}(1) \times \mathbf{Sp}(1) \longrightarrow \mathbf{SO}(4)$$

definida d'aquesta manera: identifiquem $\mathbb{R}^4$ amb $\mathbb{H}$ i definim $\pi_4(p, q)$ com l'aplicació $\mathbb{R}$ -lineal $\mathbb{H} \rightarrow \mathbb{H}$ donada per

$$h \mapsto p h \bar{q}.$$

Al llarg d'aquest apartat, per simplificar, escriurem π en lloc de π_4 . La fórmula d'anticommutació ens diu que $\pi(p, q)$ conserva el producte escalar, és a dir, $\pi(p, q) \in \mathbf{O}(3)$ i, com que $\mathbf{Sp}(1) \times \mathbf{Sp}(1)$ és un espai connex, tenim que $\pi(p, q) \in \mathbf{SO}(3)$. Que π és un homomorfisme de grups és evident. Les rotacions $\pi(p, 1)$ i $\pi(1, p)$ per $p \in \mathbf{Sp}(1)$ reben el nom d'*isoclíniques*.

Vegem primer que π és una aplicació exhaustiva. Sigui $A \in \mathbf{SO}(4)$ i sigui (a, b, c, d) la primera columna de A . Aleshores, si $p := a + bi + cj + dk \in \mathbb{H}$, observem que $B := \pi(\bar{p}, 1)A$ deixa fix el primer vector de la base canònica de $\mathbb{R}^4 = \mathbb{H}$, amb la qual cosa $B \in \mathbf{SO}(3)$ —entès com el subgrup de $\mathbf{SO}(4)$ de les matrius amb primera fila i primera columna igual a $(1, 0, \dots, 0)$ — i, com hem vist en l'estudi de les rotacions de $\mathbb{R}^3$, existeix un quaternió unitari q tal que $\phi = \pi(q, q)$. En conclusió, $A = \pi(pq, q)$ i hem comprovat que π és exhaustiva. D'altra banda, és clar que el nucli de π és el subgrup diagonal $\{(1, 1), (-1, -1)\}$.

Per tot això, podem afirmar que el recobridor universal de **SO**(4) és **Sp**(1) $\times$ **Sp**(1) i això ens duu a identificar

$$\mathbf{Spin}(4) = \mathbf{Sp}(1) \times \mathbf{Sp}(1),$$

de manera que cada rotació de l'espai euclidià de quatre dimensions ve determinada per una parella de quaternions unitaris, única llevat del signe.

Observem que, com ja havíem anunciat, si bé **Spin**(3) $\cong$ **SU**(2) $\cong$ **Sp**(1), aquests isomorfismes ja no es mantenen si augmentem la dimensió en una unitat. Tanmateix, **Spin**(4) no ens aporta cap grup nou més enllà de les famílies **SU**(n) i **Sp**(n).¹⁵ Amb les definicions que hem donat, la compatibilitat entre

¹⁵ Espòiler: Cal arribar a la dimensió 7 perquè els grups **Spin**(n) es mostrin com una família independent de les famílies **SU**(n) i **Sp**(n).

les descripcions que tenim de les rotacions de $\mathbb{R}^3$ i les de $\mathbb{R}^4$ és immediata: tenim un diagrama commutatiu

$$\begin{array}{ccc} \mathbf{Sp}(1) \cong \mathbf{Spin}(3) & \xrightarrow{\Delta} & \mathbf{Spin}(4) \cong \mathbf{Sp}(1) \times \mathbf{Sp}(1) \\ \downarrow \pi_3 & & \downarrow \pi_4 \\ \mathbf{SO}(3) & \xrightarrow{\iota} & \mathbf{SO}(4) \end{array}$$

on ι ve donada per la inclusió de $\mathbb{R}^3 = \langle i, j, k \rangle_{\mathbb{R}}$ en $\mathbb{R}^4 = \mathbb{H}$, Δ és l'aplicació diagonal i π_3, π_4 són les representacions que hem explicat en aquest text.

L'epimorfisme π també ens permet obtenir una bona descripció de l'estructura de grup de $\mathbf{SO}(4)$. Considerem aquests tres subgrups de $\mathbf{SO}(4)$

$$H_l = \{\pi(p, 1) : p \in \mathbf{Sp}(1)\} \cong \mathbf{Sp}(1),$$

$$H_r = \{\pi(1, p) : p \in \mathbf{Sp}(1)\} \cong \mathbf{Sp}(1),$$

$$H_{\Delta} = \{\pi(p, p) : p \in \mathbf{Sp}(1)\} \cong \mathbf{SO}(3).$$

Observem que H_l i H_r són normals a $\mathbf{SO}(4)$, commuten entre ells i tenen intersecció $\{\pm I\}$. En tots dos casos, el grup quocient és $\mathbf{SO}(3)$. El grup H_{Δ} no és normal i les classes laterals s'identifiquen a $\mathbf{Sp}(1)$ com a espais topològics.

L'epimorfisme π també ens permet obtenir una bona descripció de la *topologia* de l'espai $\mathbf{SO}(4)$. D'entrada, $\mathbf{SO}(4)$ s'identifica a l'espai quocient de $S^3 \times S^3$ per l'acció diagonal de l'aplicació antipodal τ , però, d'altra banda, l'aplicació $f(p, q) = (p, pq)$ de $S^3 \times S^3$ ens dona un homeomorfisme

$$(S^3 \times S^3)/\tau \cong (S^3/\tau) \times S^3,$$

on l'acció de τ sobre l'espai de l'esquerra és diagonal. En conclusió, topològicament —no com a grup!— $\mathbf{SO}(4)$ s'identifica al producte $\mathbb{R}P^3 \times S^3 = \mathbf{SO}(3) \times \mathbf{Sp}(1)$.

Finalment, voldríem relacionar la descripció de $\mathbf{SO}(4)$ anterior amb la *geometria* de les rotacions en quatre dimensions. Recordem que tota matriu de $\mathbf{SO}(4)$ es pot escriure, prenent una base ortonormal adient, en la forma

$$\begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \oplus \begin{pmatrix} \cos \beta & -\sin \beta \\ \sin \beta & \cos \beta \end{pmatrix}.$$

És a dir, hi ha dos plans invariants ortogonals i sobre cada pla tenim una rotació d'angle α i β , respectivament. Com podem llegir els plans invariants i els angles de rotació de la rotació $\pi(p, q)$? D'ençà que Cayley va descriure per primera vegada (1855) les rotacions en quatre dimensions com a productes de quaternions $x \mapsto px\bar{q}$, diversos autors com Klein, Hurwitz o Coxeter van completar aquesta teoria que desenvoluparem a continuació.

Rotacions no isoclíniques. Considerem dos quaternions unitaris $p, q \neq \pm 1$ i escrivim-los en la forma

$$\begin{aligned} p &= \cos \alpha + \sin \alpha u, \\ q &= \cos \beta + \sin \beta v, \end{aligned}$$

on $\alpha, \beta \in (0, \pi)$ i u, v són quaternions unitaris purs. Suposem, en primer lloc, que $u \neq \pm v$ i definim

$$\begin{aligned} x_1^+ &= 1 + uv, & x_2^+ &= u - v, \\ x_1^- &= 1 - uv, & x_2^- &= u + v. \end{aligned}$$

Uns càlculs elementals demostren aquestes propietats de $x_1^+, x_2^+, x_1^-, x_2^-$ (tingueu en compte les fórmules de la pàgina 16):

- $||x_1^+||^2 = ||x_2^+||^2 = 2 - 2\langle u, v \rangle$; $||x_1^-||^2 = ||x_2^-||^2 = 2 + 2\langle u, v \rangle$; $x_1^+, x_2^+, x_1^-, x_2^-$ són ortogonals dos a dos.
- Com que estem suposant $u \neq \pm v$, normalitzant els quaternions $x_1^+, x_2^+, x_1^-, x_2^-$, obtenim una base ortonormal de $\mathbb{H}$ com a $\mathbb{R}$ -espai vectorial. A més, podem afirmar que aquesta base és *positiva*. Això es dedueix d'aquest resultat:

Si u, v són quaternions purs, aleshores, per la desigualtat de Cauchy-Schwarz,

$$\det(1, u, v, uv) = ||u||^2 ||v||^2 - \langle u, v \rangle^2 \geq 0.$$

Per comprovar-ho, recordem la fórmula del producte vectorial triple $\vec{a} \times (\vec{b} \times \vec{c}) = \langle \vec{a}, \vec{c} \rangle \vec{b} - \langle \vec{a}, \vec{b} \rangle \vec{c}$ i recordem també que en dimensió 3 tenim aquesta relació entre el determinant i el producte vectorial: $\det(\vec{e}_1, \vec{e}_2, \vec{e}_3) = \langle \vec{e}_1, \vec{e}_2 \times \vec{e}_3 \rangle$. Aleshores:

$$\begin{aligned} \det(1, u, v, uv) &= \det(1, u, v, -\langle u, v \rangle + u \times v) = \\ &= \det(u, v, u \times v) = \langle u, v \times (u \times v) \rangle = \\ &= \langle u, ||v||^2 u - \langle u, v \rangle v \rangle = ||u||^2 ||v||^2 - \langle u, v \rangle^2 \geq 0. \quad \square \end{aligned}$$

En el cas de la base $x_1^+, x_2^+, x_1^-, x_2^-$ és fàcil veure que el seu determinant és igual a $4 \det(1, u, v, uv) > 0$.

- L'acció de $\pi(p, q)$ compleix aquestes fórmules:

$$\begin{aligned} p x_1^+ \bar{q} &= \cos(\alpha + \beta) x_1^+ + \sin(\alpha + \beta) x_2^+, \\ p x_2^+ \bar{q} &= -\sin(\alpha + \beta) x_1^+ + \cos(\alpha + \beta) x_2^+, \\ p x_1^- \bar{q} &= -\cos(\alpha - \beta) x_1^- + \sin(\alpha - \beta) x_2^-, \\ p x_2^- \bar{q} &= -\sin(\alpha - \beta) x_1^- + \cos(\alpha - \beta) x_2^-. \end{aligned}$$

- Les fórmules anteriors demostren que $H^+ := \langle x_1^+, x_2^+ \rangle$ i $H^- := \langle x_1^-, x_2^- \rangle$ són plans ortogonals invariants per $\pi(p, q)$ i que $\pi(p, q)$ és la rotació plana d'angle $\alpha + \beta$ en el pla H^+ i la rotació plana d'angle $\alpha - \beta$ en el pla H^- .

El cas $u = \pm v$ es comporta pràcticament igual. Per exemple, si $u = v$, seguim tenint el pla invariant $H^- = \langle 1, u \rangle$ sobre el qual $\pi(p, q)$ és la rotació plana d'angle $\alpha - \beta$. Escollim ara un quaternió pur unitari w ortogonal a u i definim $H^+ := \langle w, uw \rangle$. Aleshores, H^+ i H^- són dos plans invariants ortogonals i uns càlculs com els anteriors ens mostren que $\pi(p, q)$ actua sobre H^+ com la rotació plana d'angle $\alpha + \beta$. Observem que, pel resultat anterior sobre el determinant, la base que hem pres és positiva. El cas $u = -v$ és exactament igual, excepte que ara $H^- := \langle w, uw \rangle$ i $H^+ = \langle 1, u \rangle$.

Estudiem ara l'existència d'altres *plans invariants* en la rotació $\pi(p, q)$ amb $p, q \neq \pm 1$. Si descartem el cas $\alpha = \beta = \pi/2$ —en el qual $\pi(p, q)$ té, evidentment, infinits plans invariants—, demostrarem que els plans H^+ i H^- anteriors són els únics plans invariants de $\rho := \pi(p, q)$.

Suposem que $H \neq H^+, H^-$ és un pla invariant i sigui $h \in H$ tal que $h = h^+ + h^-$ amb $h^+ \in H^+$, $h^- \in H^-$ i $h^+, h^- \neq 0$. Les hipòtesis que hem fet sobre α, β impliquen que $\rho(h) \neq \pm h$ i que com a mínim una de les dues parelles de vectors $(h^+, \rho(h^+))$ i $(h^-, \rho(h^-))$ és linealment independent. En particular, h i $\rho(h)$ formen una base de H . Pel que hem vist més amunt sobre la rotació ρ , el polinomi anul·lador de ρ sobre H^+ és $x^2 - 2 \cos(\alpha + \beta)x + 1$ i el polinomi anul·lador de ρ sobre H^- és $x^2 - 2 \cos(\alpha - \beta)x + 1$. Calculem ara $\rho^2(h)$:

$$\rho^2(h) = 2 \cos(\alpha + \beta) \rho(h^+) - h^+ + 2 \cos(\alpha - \beta) \rho(h^-) - h^-.$$

Com que estem suposant que H és un pla invariant, tindrem $\rho^2(h) - h = \lambda h + \mu \rho(h)$ i, com que $H^+ \cap H^- = \{0\}$, obtenim aquestes dues igualtats:

$$(2 \cos(\alpha + \beta) - \mu) \rho(h^+) = \lambda h^+,$$

$$(2 \cos(\alpha - \beta) - \mu) \rho(h^-) = \lambda h^-,$$

que impliquen (recordem que com a mínim una de les dues parelles $(h^+, \rho(h^+))$ i $(h^-, \rho(h^-))$ és linealment independent) $\lambda = 0$ i $\cos(\alpha + \beta) = \cos(\alpha - \beta)$, que es contradiu amb la hipòtesi $p, q \neq \pm 1$.

Rotacions isoclíniques. Resta només estudiar el cas de les rotacions *isoclíniques*, és a dir, el cas en què $p = 1$ o $q = 1$. En aquest cas, hi ha infinits plans invariants i cada quaternió $h \neq 0$ pertany a un d'aquests plans. Concretament, sigui p un quaternió unitari $p = \cos \alpha + \sin \alpha u \neq \pm 1$ i sigui $h \in \mathbb{H} - \{0\}$. Si apliquem la identitat $p^2 = 2 \cos \alpha p - 1$, veiem que $p^2 h = 2 \cos \alpha p h - h \in \langle h, p h \rangle$ i $\langle h, p h \rangle$ és un pla invariant per $\pi(p, 1)$ i, anàlogament, $\langle h, h \bar{p} \rangle$ és un pla invariant per $\pi(1, p)$.

Si w és un quaternió unitari pur ortogonal a u , podem prendre la base ortonormal positiva (pàgina 27) de $\mathbb{H}$ formada per $1, u, w, uw$ i observem

que $\pi(p, 1)$ és la rotació d'angle α en cadascun dels plans invariants ortogonals $\langle 1, u \rangle$ i $\langle w, uw \rangle$. En canvi, $\pi(1, p)$ és la rotació d'angle $-\alpha$ a $\langle 1, u \rangle$ i la rotació d'angle α a $\langle w, uw \rangle$. Veiem, doncs, que les rotacions isoclíniques es classifiquen en dos tipus: les que giren angles iguals en dos plans invariants ortogonals i les que giren angles oposats en dos plans invariants ortogonals ($\pm I$ pertany als dos tipus). Les primeres s'obtenen per multiplicació per l'esquerreta per un quaternió unitari i, les segones, per multiplicació per la dreta per un quaternió unitari. Aquesta distinció es correspon als dos subgrups normals de $\mathbf{SO}(4)$ —no conjugats, és clar— H_l i H_r que hem considerat abans.

Digressió # 5: rotacions de l'espai temps

L'espai temps de la teoria de la relativitat (especial, és a dir, sense gravetat) no és l'espai euclidià $\mathbb{R}^4$ perquè la coordenada *temps* (t) es comporta de manera diferent de les tres coordenades *espacials* (x, y, z) i, per tant, el concepte de *rotació* requerirà una definició diferent de la que hem utilitzat en la geometria euclidiana de $\mathbb{R}^n$, és a dir, en els altres apartats d'aquest text. Expliquem de manera molt succinta —que vol dir que ho explicarem necessàriament malament—¹⁶ quina és la *distància* que hem d'utilitzar a $\mathbb{R}^4$ quan l'identifiquem a l'espai temps de la teoria de la relativitat.

Considerem un sistema de referència inercial S amb un sistema de coordenades ortonormal x, y, z i un rellotge que mesura el temps t ,¹⁷ i un altre sistema inercial S' amb coordenades x', y', z' i un rellotge que mesura el temps t' . A cada esdeveniment el sistema S li assignarà unes coordenades $(x, y, z, t) \in \mathbb{R}^4$ i el sistema S' li assignarà unes altres coordenades $(x', y', z', t') \in \mathbb{R}^4$. Suposem que, quan $t = 0$, també $t' = 0$ i els orígens de S i S' coincideixen. Suposem que en l'instant $t = t' = 0$ un fotó surt de l'origen i tenim un observador a S i un altre a S' que volen calcular la velocitat del fotó. Suposem que aquest fotó passa per un punt P . Aleshores, S detectarà el fotó en el punt P amb coordenades (x, y, z) en l'instant t i S' detectarà el fotó en el mateix punt P amb coordenades (x', y', z') en l'instant t' . És a dir, S observarà que el fotó ha recorregut una distància $\sqrt{x^2 + y^2 + z^2}$ en un temps t i S' observarà que el fotó ha recorregut una distància $\sqrt{x'^2 + y'^2 + z'^2}$ en un temps t' . Per l'axioma fonamental de la teoria de la relativitat, S i S' han d'arribar al mateix valor per a la velocitat del fotó. Per tant,

$$\frac{\sqrt{x^2 + y^2 + z^2}}{t} = c = \frac{\sqrt{x'^2 + y'^2 + z'^2}}{t'}.$$

Aquest raonament ens diu que el punts en què s'anulla la forma quadràtica

$$Q(X, Y, Z, T) = X^2 + Y^2 + Z^2 - c^2 T^2$$

¹⁶ Si voleu més detalls sobre tot el que explicarem ara, llegiu el llibre imprescindible *Geometria diferencial i relativitat*, de Joan Girbau, en el qual trobareu molt ben explicat tot el que aquí hem de resumir excessivament.

¹⁷ De fet, un rellotge a cada punt de l'espai, tots sincronitzats.

segons l'observador de S són els mateixos punts en què s'anulla segons l'observador de S' . Dit amb més exactitud: sigui A el canvi de coordenades (que no coneixem) entre (x, y, z, t) i (x', y', z', t') , i sigui Q^A la quàdrica transformada de Q per A ; aleshores,

$$Q(\vec{e}) = 0 \iff Q^A(\vec{e}) = 0.$$

Això implica¹⁸ que les dues quàdriques difereixen en una constant no nul·la. A partir d'aquí, fent uns càlculs més o menys llargs,¹⁹ podem deduir:

- Les clàssiques fórmules de Lorentz que expressen el canvi de coordenades $(x, y, z, t) \mapsto (x', y', z', t')$ en funció de la velocitat v de S' respecte de S (sempre que $|v| < c$).
- $\lambda = 1$ i el canvi de coordenades *conserva la forma quadràtica* Q .

El que ens interessa de tot això és aquest darrer punt: a l'espaitemps, els moviments han de respectar la forma Q , que, per tant, fa el paper que fa la forma quadràtica $X^2 + Y^2 + Z^2 + T^2$ a l'espai euclidià. En conclusió, deixant de banda les translacions, el grup dels moviments rígids de l'espaitemps estarà format per les matrius reals invertibles 4×4 que conservin la forma quadràtica $Q = X^2 + Y^2 + Z^2 - c^2 T^2$, una forma quadràtica no singular de rang 4 i índex 1 equivalent a la forma $X^2 + Y^2 + Z^2 - T^2$. Aquestes matrius formen el que es coneix com a *grup de Lorentz*. El denotarem amb $O(3, 1)$ i és el grup del qual volem dir alguna cosa en aquesta digressió.

Si Γ és la matriu diagonal $(1, 1, 1, -1)$ —és a dir, la matriu de la forma quadràtica Q anterior—, aleshores

$$O(3, 1) = \{A \in \mathcal{M}_{4 \times 4}(\mathbb{R}) : A^t \Gamma A = \Gamma\}$$

és clarament un subgrup de $GL_4(\mathbb{R})$. Podem considerar el subgrup $O(3) \times O(1) \subset O(3, 1)$ tenint en compte les transformacions ortogonals de les tres primeres coordenades i el canvi de signe de l'última coordenada.

$O(3, 1)$ no és compacte: per a cada $\lambda \in \mathbb{R}$ podem considerar la matriu

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \oplus \begin{pmatrix} \mu & \lambda \\ \lambda & \mu \end{pmatrix} \in O(3, 2)$$

¹⁸ És curiós que molts llibres de teoria de la relativitat ometen aquest punt (una excepció és el llibre de Bernard Schutz), és a dir, de la igualtat dels conjunts de punts en què les quàdriques s'anul·len passen a la igualtat de les formes, sense cap comentari. Matemàticament, el problema es planteja en aquests termes: siguin f i g dos polinomis irreductibles en n variables tals que el conjunt de zeros de f sigui el mateix que el conjunt de zeros de g , és a dir, $V(\langle f \rangle) = V(\langle g \rangle)$, en el llenguatge de la geometria algebraica; podem deduir que $\langle f \rangle = \langle g \rangle$? Si el cos base és algebraicament tancat, la resposta és sí i es dedueix immediatament del Nullstellensatz, però, si el cos és $\mathbb{R}$, com en el nostre cas, la resposta és no: $x^2 + y^2$ i $x^2 + 2y^2$ tenen els mateixos zeros a $\mathbb{R}^2$ i no difereixen en una constant. Cal alguna hipòtesi addicional. Afortunadament, el llibre *Afinitats, moviments i quàdriques*, d'Agustí Reventós, sí que tracta el tema (apèndix D) i demostra que, si f i g són de grau dos sobre $\mathbb{R}$, amb els mateixos conjunts de zeros, i un d'ells compleix que té un zero *regular* (és a dir, amb gradient no nul), aleshores f i g difereixen en una constant no nul·la. Com que en el cas de la forma quadràtica Q la condició de regularitat es compleix, podem afirmar que $Q = \lambda Q^A$.

¹⁹ Ho podeu trobar molt ben explicat al llibre citat de Joan Girbau.

amb $\mu = \sqrt{1 + \lambda^2}$. Usualment, aquestes matrius s'escriuen utilitzant les funcions trigonomètriques hiperbòliques:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \oplus \begin{pmatrix} \cosh x & \sinh x \\ \sinh x & \cosh x \end{pmatrix}, \quad x \in \mathbb{R},$$

que recorda l'expressió de les rotacions amb sinus i cosinus. Es pot demostrar que el subgrup $O(3) \times O(1)$ és un subgrup compacte maximal de $O(3, 1)$.

És clar que les matrius de $O(3, 1)$ tenen determinant ± 1 i tenim, per tant, un homomorfisme de grups $\det: O(3, 1) \rightarrow \mathbb{Z}/2$ que és exhaustiu i, en particular, ens diu que el grup de Lorentz no és connex. Tal com fem amb el grup ortogonal, definim $SO(3, 1)$ com el subgrup normal de $O(3, 1)$ de les matrius de determinant 1.

Ara apareix un fenomen que no tenim en el cas dels grups ortogonals clàssics. Hi ha encara un altre homomorfisme exhaustiu $\theta: O(3, 1) \rightarrow \mathbb{Z}/2$ que es defineix d'aquesta manera: si $A \in O(3, 1)$, escrivim $A = \begin{pmatrix} M & v \\ w^t & \mu \end{pmatrix}$ amb $\mu \in \mathbb{R}$. De la condició $A^t \Gamma A = \Gamma$ es dedueix $\mu^2 = v^t v + 1 = ||v||^2 + 1 \geq 1$. Per tant, $\mu \neq 0$ i podem definir $\theta(A) := \mu/|\mu| \in \{\pm 1\} = \mathbb{Z}/2$. Resulta que θ és un homomorfisme, però no conec cap demostració elemental d'aquest fet. Resulta que θ s'identifica amb un concepte general de formes quadràtiques que s'anomena *norma espinorial* i que, per la seva mateixa definició, és un homomorfisme. En definitiva, tenim un epimorfisme

$$O(3, 1) \rightarrow \mathbb{Z}/2 \times \mathbb{Z}/2$$

que ens diu que $O(3, 1)$ té, com a mínim, quatre components connexes —són exactament quatre— i ens permet definir aquest subgrup normal *connex*

$$SO^+(3, 1) := \{A \in O(3, 1) : \det(A) = \theta(A) = 1\}.$$

Ara podríem començar a estudiar un grup $\text{Spin}(3, 1)$ que fos un recobriment doble de $SO^+(3, 1)$, identificar-lo a algun altre grup conegut (coincideix amb el grup simplèctic no compacte $\text{Sp}_2(\mathbb{C})$), etc., però serà millor que ho deixem aquí.

6 Rotacions en dimensió 5

En aquest apartat estudiarem les rotacions de l'espai euclidià $\mathbb{R}^5$, amb la mateixa intenció dels casos anteriors: veure si podem identificar els espinors en dimensió 5 amb algun grup conegut. Avancem que la resposta que trobarem serà que **Spin**(5) s'identifica a **Sp**(2). Per veure-ho, considerem aquest $\mathbb{R}$ -espai vectorial de dimensió 5:

$$V := \left\{ \begin{pmatrix} \lambda & q \\ \bar{q} & -\lambda \end{pmatrix} \in \mathcal{M}_{2 \times 2}(\mathbb{H}) : \lambda \in \mathbb{R}, q \in \mathbb{H} \right\}.$$

Observem:

- V està format per les matrius hermítiques quaternioniques de traça zero: $V = \{M \in \mathcal{M}_{2 \times 2}(\mathbb{H}) : M^* = M, \text{tr}(M) = 0\}$. Escriurem els vectors de V en la forma $[\lambda, q]$ i prendrem com a base ortonormal de V la formada per les matrius $[1, 0]$, $[0, 1]$, $[0, i]$, $[0, j]$, $[0, k]$.
- $\mathbf{Sp}(2)$ actua sobre V per conjugació. Això és evident pel punt anterior i pel fet que, per definició, les matrius de $\mathbf{Sp}(2)$ compleixen $Q^{-1} = Q^*$.
- Un càlcul immediat mostra que, si $M \in V$, aleshores $M^2 = ||M||^2 I$, on la norma l'entendem amb el producte escalar euclidià de V com a $\mathbb{R}$ -espai vectorial. Com a conseqüència d'això, la conjugació per una matriu de $\mathbf{Sp}(2)$ conserva la norma de $V = \mathbb{R}^5$ i, per tant, és una transformació ortogonal.
- Tenim, per tant, un homomorfisme $\mathbf{Sp}(2) \rightarrow \mathbf{O}(5)$ i, com que $\mathbf{Sp}(2)$ és connex, la imatge estarà continguda en el component de la identitat:

$$\pi_5: \mathbf{Sp}(2) \rightarrow \mathbf{SO}(5).$$

Recordem l'epimorfisme $\mathbf{Sp}(1) \times \mathbf{Sp}(1) \rightarrow \mathbf{SO}(4)$ de la secció anterior. És evident que, si identifiquem $\mathbb{R}^4 = \mathbb{H}$ amb l'hiperplà $\lambda = 0$ de V i identifiquem $\mathbf{Sp}(1) \times \mathbf{Sp}(1)$ amb les matrius diagonals de $\mathbf{Sp}(2)$, aleshores l'homomorfisme $\pi_5: \mathbf{Sp}(2) \rightarrow \mathbf{SO}(5)$ és una extensió de la representació de les rotacions en dimensió 4 i tenim un diagrama commutatiu

$$\begin{array}{ccc} \mathbf{Sp}(1) \times \mathbf{Sp}(1) & \longrightarrow & \mathbf{Sp}(2) \\ \downarrow \pi_4 & & \downarrow \pi_5 \\ \mathbf{SO}(4) & \xrightarrow{\iota} & \mathbf{SO}(5) \end{array}$$

Calculem el nucli de π_5 , és a dir, les matrius $Q \in \mathbf{Sp}(2)$ que commuten amb totes les matrius de V . La commutativitat amb $[1, 0]$ implica que Q és diagonal. La commutativitat amb $[0, 1]$ implica que Q és de la forma pI amb $||p|| = 1$. Finalment, la commutativitat amb totes les matrius $[0, q]$ per a tot q implica que p és central a $\mathbb{H}$ i, per tant, $Q = \pm I$.

Finalment, demostrem que π_5 és exhaustiva. Sigui $A \in \mathbf{SO}(5)$ i sigui $[\lambda, q] \in V$ un eix de A . No és restrictiu suposar que $||[\lambda, q]|| = 1$ i $\lambda \geq 0$. Sigui ara

$$\mu := \sqrt{\frac{1+\lambda}{2}}, \quad p := \frac{q}{2\mu}, \quad Q := \begin{pmatrix} \mu & -p \\ \bar{p} & \mu \end{pmatrix}.$$

Es compleix que $Q \in \mathbf{Sp}(2)$ i que $Q[1, 0]Q^* = [\lambda, q] \in V$. Aleshores $B := \pi_5(Q)^{-1}A\pi_5(Q) \in \mathbf{SO}(4)$ perquè fixa l'eix $[1, 0]$. Per tant, B pertany a la imatge de π_5 i A també.

Observem que els dos subgrups normals H_l, H_r de les rotacions isoclíniques de $\mathbf{SO}(4)$ passen a ser conjugats a $\mathbf{SO}(5)$. No tinc constància que s'hagi fet un estudi dels eixos i els plans invariants com el que hem explicat a l'apartat de les rotacions en dimensió 4.

En conclusió, $\mathbf{Sp}(2)$ és el recobriment universal de $\mathbf{SO}(2)$ i podem identificar

$$\mathbf{Spin}(5) = \mathbf{Sp}(2).$$

Una vegada més, tampoc no ens ha calgut descobrir cap grup nou per descriure els espinors en dimensió 5.

7 Rotacions en dimensió 6

Estem arribant al límit del que considerem *baixa dimensió* en aquest treball: estudiarem les rotacions en dimensió 6 i veurem com també les podem descriure utilitzant uns espinors apropiats que, novament, identificarem amb algun dels grups que ja coneixem. Recordem que hem identificat $\mathbf{Spin}(5)$ amb $\mathbf{Sp}(2)$ i recordem també que hem vist (pàgina 20) que $\mathbf{Sp}(2)$ és un subgrup de $\mathbf{SU}(4)$, concretament, $\mathbf{Sp}(2) = \mathbf{SU}(4) \cap \mathbf{Sp}_4(\mathbb{C})$. Això ens pot dur a intuir que potser podrem identificar el grup $\mathbf{Spin}(6)$ amb $\mathbf{SU}(4)$ i, efectivament, aquesta intuïció és correcta, però alguns detalls seran més complicats que en les dimensions inferiors.²⁰ Per començar, ens caldrà repassar alguns conceptes d'àlgebra multilineal.

Sigui $V = \mathbb{C}^n$ amb el producte hermític ordinari i la base canònica $e_1, \dots, e_n$. Ens interessa considerar aquestes representacions de $\mathbf{GL}_n(\mathbb{C})$: l'acció natural sobre V , l'acció sobre V^* donada per

$$(A \cdot \omega)(e) = \omega(\overline{A}^t e) = \omega(A^* e)$$

i les accions naturals que es desprenen d'aquestes dues sobre els productes exteriors $\bigwedge^k V$ i $\bigwedge^k V^*$, per $k = 1, \dots, n$. A partir d'aquí, observem:

- Hi ha un isomorfisme $\mathbf{GL}_n(\mathbb{C})$ -equivariant $\alpha: \bigwedge^k V^* \rightarrow (\bigwedge^k V)^*$ donat pel determinant: $\alpha(\omega_1 \wedge \dots \wedge \omega_k)(u_1 \wedge \dots \wedge u_k) := \det(\omega_i(u_j))$.
- Tenim una *dualitat* $d: V \rightarrow V^*$ definida així:

$$d(e)(u) := \langle u, e \rangle.$$

d és bijectiva i *semilineal* en el sentit que $d(\lambda e) = \overline{\lambda} d(e)$. Direm que d és un c-isomorfisme. Observem que d és $\mathbf{U}(n)$ -equivariant. També podem considerar el c-isomorfisme $\bigwedge^k d: \bigwedge^k V \rightarrow \bigwedge^k V^*$.

- Si $n = 2k$, podem considerar aquesta *forma bilineal simètrica* B sobre $\bigwedge^k V$:

$$B(u_1 \wedge \dots \wedge u_k, v_1 \wedge \dots \wedge v_k) := \det(u_1, \dots, u_k, v_1, \dots, v_k).$$

Aquesta forma bilineal dona lloc, de manera natural, a un isomorfisme $b: \bigwedge^k V \rightarrow (\bigwedge^k V)^*$ i és senzill comprovar que b és $\mathbf{SU}(n)$ -equivariant.

²⁰ Seguirem el camí indicat per Paul Garret a *Sporadic isogenies to orthogonal groups* (2015).

- Combinant les aplicacions anteriors

$$\bigwedge^k V \xrightarrow{\wedge^k d} \bigwedge^k V^* \xrightarrow{\alpha} (\bigwedge^k V)^* \xleftarrow{b} \bigwedge^k V$$

obtenim un c-isomorfisme $J: \bigwedge^k V \rightarrow \bigwedge^k V$ que és $\mathbf{SU}(n)$ -equivariant.

- Sigui $W := \text{Fix}(J)$ el conjunt dels $\rho \in \bigwedge^k V$ fixos per J . Observem que W no és un subespai vectorial de $\bigwedge^k V$ —perquè J no és lineal—, però sí que és un $\mathbb{R}$ -subespai vectorial. D'altra banda, és clar que W és tancat per l'acció de $\mathbf{SU}(n)$ —perquè J és equivariant.

Per a l'estudi del grup $\mathbf{Spin}(6)$ ens interessa l'aplicació J en el cas $n = 4$

$$J: \bigwedge^2 \mathbb{C}^4 \rightarrow \bigwedge^2 \mathbb{C}^4,$$

que no és difícil de calcular de manera explícita. Considerem la base canònica $e_{ij} := e_i \wedge e_j$, $i < j$, de $\bigwedge^2 \mathbb{C}^4$. Un càlcul senzill ens dona

$$\begin{aligned} J(e_{12}) &= e_{34}, & J(e_{34}) &= e_{12}, & J(e_{13}) &= -e_{24}, \\ J(e_{24}) &= -e_{13}, & J(e_{23}) &= e_{14}, & J(e_{14}) &= e_{23}. \end{aligned}$$

És a dir, la matriu de J com a aplicació $\mathbb{R}$ -lineal en la base

$$\{e_{12}, e_{34}, ie_{12}, ie_{34}, e_{13}, e_{24}, ie_{13}, ie_{24}, e_{23}, e_{14}, ie_{23}, ie_{14}\}$$

és

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \oplus \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}.$$

A partir d'aquí obtenim aquesta base de W com a $\mathbb{R}$ -espai vectorial:

$$\begin{aligned} w_0 &:= ie_{13} + ie_{24}, & w_1 &:= e_{24} - e_{13}, & w_2 &:= -e_{14} - e_{23}, \\ w_3 &:= ie_{23} - ie_{14}, & w_4 &:= e_{12} + e_{34}, & w_5 &:= ie_{12} - ie_{34}. \end{aligned}$$

Recordem que sobre $\bigwedge^2 \mathbb{C}^4$ —i, per tant, sobre W — hi tenim una forma bilineal simètrica B . Si calculem els valors que pren B sobre la $\mathbb{R}$ -base de W , observem que obtenim una matriu diagonal amb 2 a la diagonal. És a dir, B és un producte escalar (real) definit positiu sobre W i $\{w_i/\sqrt{2} : i = 0, \dots, 5\}$ és una base ortonormal respecte de B . D'altra banda, l'acció de $\mathbf{SU}(4)$ conserva la forma bilineal B . En conclusió, hem obtingut un homomorfisme de grups

$$\mathbf{SU}(4) \rightarrow \mathbf{O}(6).$$

Com que els grups $\mathbf{SU}(n)$ són connexos, tenim

$$\pi_6: \mathbf{SU}(4) \rightarrow \mathbf{SO}(6).$$

A continuació, demostrem que el nucli de π_6 és $\{\pm I\}$. Suposem que $A = (a_{ij}) \in \mathbf{SU}(4)$ és tal que deixa fixos els elements de la base de W . A partir d'aquí és fàcil veure que A actua trivialment sobre $\wedge^2 \mathbb{C}^4$: per a cada $i < j$ tindrem $Ae_i \wedge Ae_j = e_i \wedge e_j$; prenem $k \neq i, j$, considerant els coeficients de e_{ik} , e_{jk} i e_{ij} a la igualtat anterior, arribem a

$$a_{ii}a_{kj} = a_{ki}a_{ij}, \quad a_{ji}a_{kj} = a_{ki}a_{jj}, \quad a_{ii}a_{jj} - a_{ji}a_{ij} = 1.$$

Deduïm que A ha de ser diagonal i el producte de dos elements qualssevol de la diagonal ha de ser igual a 1, que només és possible si $A = \pm I$.

Relacionem ara $\pi_6: \mathbf{SU}(4) \rightarrow \mathbf{SO}(6)$ amb l'epimorfisme $\pi_5: \mathbf{Sp}(2) \rightarrow \mathbf{SO}(5)$ de l'apartat anterior. Recordem que havíem vist que $\mathbf{Sp}(2)$ s'identifica al subgrup de $\mathbf{SU}(4)$ format per les matrius que conserven la forma alternada no degenerada sobre V donada per $\Omega = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$ en la base canònica e_1, e_2, e_3, e_4 (pàgina 20). Aquesta forma dona una aplicació lineal $\omega: \wedge^2 V \rightarrow \mathbb{C}$ que compleix

$$\omega(e_{12}) = \omega(e_{14}) = \omega(e_{23}) = \omega(e_{34}) = 0, \quad \omega(e_{13}) = \omega(e_{24}) = 1$$

i, per tant, $\omega(w_0) = 2i$ i $\omega(w_i) = 0$ per $i \neq 0$. Això ens diu que un element de $\mathbf{SU}(4)$ conserva ω si i només si deixa fix w_0 . En conclusió, $\pi_6: \mathbf{SU}(4) \rightarrow \mathbf{SO}(6)$ envia $\mathbf{Spin}(5) = \mathbf{Sp}(2)$ a $\mathbf{SO}(\langle w_0 \rangle^\perp) = \mathbf{SO}(5)$. Ara caldrà relacionar aquest epimorfisme de $\mathbf{Spin}(5)$ a $\mathbf{SO}(5)$ amb π_5 : es compleix que aquest diagrama és commutatiu:

$$\begin{array}{ccc} \mathbf{Sp}(2) & \longrightarrow & \mathbf{SU}(4) \\ \downarrow \pi_5 & & \downarrow \pi_6 \\ \mathbf{SO}(5) & \xrightarrow{\iota} & \mathbf{SO}(6) \end{array}$$

La manera elemental de demostrar la commutativitat d'aquest diagrama és fent un càlcul directe: prenem una matriu arbitrària $A \in \mathbf{Sp}(2)$, calculem explícitament la matriu $\pi_6(A_{\mathbb{C}})$ i la comparem amb la matriu $\pi_5(A)$.²¹

Finalment, demostrarem que π_6 és exhaustiva. Considerem, doncs, una matriu $R \in \mathbf{SO}(6)$ per a la qual volem trobar $P \in \mathbf{SU}(4)$ tal que $\pi_6(P) = R$. Escrivim

$$R(w_0) = \lambda w_0 + \mu w, \quad w \in \langle w_0 \rangle^\perp, \quad ||w||^2 = ||w_0||^2 = 2, \quad \lambda^2 + \mu^2 = 1.$$

Segui $S \in \mathbf{SO}(\langle w_0 \rangle^\perp)$ una rotació qualsevol tal que $S(w) = w_1$, de manera que $SR(w_0) = \lambda w_0 + \mu w_1$. Ja sabem que existirà $A \in \mathbf{Sp}(2) \subset \mathbf{SU}(4)$ tal que $S = \pi_6(A)$. Considerem la matriu complexa diagonal

$$M := \text{Diag}(\lambda i - \mu, \lambda i + \mu, -i, -i).$$

²¹ Per fer aquest càlcul i trobar la base $w_0, \dots, w_5$ que fa que el diagrama sigui commutatiu vaig utilitzar *sagemath*, en particular, aquest *script* (accessible des de la versió electrònica de l'article). Seria bonic trobar una demostració breu i no computacional de la commutativitat del diagrama, però la veritat és que fer aquest càlcul amb *sagemath* també té el seu encant.

Tenim que $M \in \mathbf{SU}(4)$ i $\pi_6(M)(w_0) = \lambda w_0 + \mu w_1$. Aleshores $\pi_6(M^{-1})\pi_6(A)R(w_0) = w_0$, amb la qual cosa existirà $B \in \mathbf{Sp}(2)$ tal que $\pi_6(M^{-1})\pi_6(A)R = \pi_6(B)$ i deduïm que R és a la imatge de π_6 .

8 Conclusió

Hem completat el nostre objectiu de donar descripcions explícites, detallades i coherents de les *isogènies esporàdiques* entre els grups d'espinors $\mathbf{Spin}(k)$ per $k < 7$ i grups de les famílies $\mathbf{SU}(n)$ i $\mathbf{Sp}(n)$, i hem obtingut aquest diagrama commutatiu:

$$\begin{array}{ccccccc}
 \mathbf{Spin}(3) & \longrightarrow & \mathbf{Spin}(4) & \longrightarrow & \mathbf{Spin}(5) & \longrightarrow & \mathbf{Spin}(6) \\
 \parallel & & \parallel & & \parallel & & \parallel \\
 \mathbf{Sp}(1) & \xrightarrow{\Delta} & \mathbf{Sp}(1) \times \mathbf{Sp}(1) & \longrightarrow & \mathbf{Sp}(2) & \longrightarrow & \mathbf{SU}(4) \\
 \downarrow \pi_3 & & \downarrow \pi_4 & & \downarrow \pi_5 & & \downarrow \pi_6 \\
 \mathbf{SO}(3) & \xrightarrow{\iota} & \mathbf{SO}(4) & \xrightarrow{\iota} & \mathbf{SO}(5) & \xrightarrow{\iota} & \mathbf{SO}(6)
 \end{array}$$

Esperem que, més enllà de l'interès que aquestes isogènies clàssiques puguin tenir, el camí que hem seguit —i les digressions que hem fet— puguin ser útils per iniciar-se en l'estudi dels *grups de Lie compactes* i les seves apassionants propietats geomètriques i topològiques.²²

CATEDRÀTIC DE TOPOLOGIA JUBILAT
 UNIVERSITAT AUTÒNOMA DE BARCELONA
 Jaume.Aguade@uab.cat

²² Si us ha agradat aquest estudi de les rotacions en dimensió < 7 , no us perdeu els fets extraordinaris que succeeixen en dimensió 8.

Políedres i papiroflèxia modular

JAUME COLL GUERRERO

La papiroflèxia és la manera més fàcil
de conèixer la geometria.

Joan Bascuñana, artista de la vida olotí

Resum: En aquest treball s'expliquen tècniques de papiroflèxia modular que ens permeten fer amb paper una gran varietat de políedres: els cinc sòlids platònics, els tretze políedres arquimedians, els vuit deltaedres convexos i les seves estelacions, prismes, antiprismes, piràmides i cúpules geodèsiques fullerianes, entre d'altres.

Paraules clau: papiroflèxia modular, políedres, sòlids platònics, políedres arquimedians, tècniques de papiroflèxia, tessellacions.

Classificació MSC2020: 00A08, 00A66, 05B45, 51M20, 52B05, 52B10.

Introducció

Ara fa cinquanta anys es va publicar amb el nom de *caixa de colors* un cub de paper que actualment és molt popular i que es coneix com a *cub de Sonobe*. Aquesta figura de paper es construeix a partir de sis mòduls bàsics que s'uneixen mitjançant pestanyes i butxaques, sense utilitzar cola, per formar un cub. El cub de Sonobe va donar lloc al naixement de tota una nova branca de la papiroflèxia: la papiroflèxia modular. Des d'aleshores s'han creat altres mòduls per fer políedres modulars en paper, com el tetràedre i l'octàedre de Kuni-hiko Kasahara (Omnibus Origami, 1988), l'icosàedre de Tomoko Fuse (BOS Convention, 1987), el dodecàedre de Lewis Simon (*Pajarita* núm. 26, 1989), el cubooctàedre de Lewis Simon (*Pajarita Extra*, 1988), el cub truncat d'Oriol Bach (*Pajarita* núm. 29, 1989), el rombicuboctàedre de Natale Fietta (CDO, 1986) o el dodecàedre ròmbic de David Brill (BOS Convention, 1986).

Tots aquests mòduls s'uneixen de la mateixa manera que el cub de Sonobe i l'essència de la majoria d'aquests mòduls és aconseguir un angle molt concret per fer els polígons que formaran el políedre.

Cal destacar un mòdul de tipus aresta molt versàtil, flexible i resistent, el Kantenmodul (2002), de l'alemanya Carmen Sprung; amb aquest mòdul es poden construir una gran varietat de políedres, com per exemple quatre dels cinc sòlids platònics, la família dels vuit deltaedres convexos, molts deltaedres còncaus, prismes, antiprismes, piràmides, cúpules fullerianes, triangulacions de sòlids platònics i arquimedians, etc.

També cal destacar els mòduls aresta de 120 graus (1987), 135 graus (1987) i 144 graus (2013) de l'origamista de Singapur Francis Ow (1949–2018), que permeten fer els tretze políedres arquimedians. La col·lecció dels tretze políedres arquimedians fets amb origami modular fou exposada per Dirk Eisner l'any 2011 a Friburg de Brisgòvia en la 6 Internationale Tagunz Zur Faltikdidaktik, convenció de papiroflèxia organitzada per Nick Robinson i el badaloní Joan Sallas. Els mòduls utilitzats en aquestes figures i dissenyats l'any 2007 conjuntament per Tomoko Fuse, Dirk Eisner i el mateix Francis Ow eren variants dels mòduls de Francis Ow.

Acabem la introducció explicant l'estructura d'aquest article. A la primera part del treball es fan algunes observacions sobre l'origen dels políedres i es demostra que tan sols existeixen cinc sòlids platònics o regulars i tretze políedres arquimedians o semiregulars.

A la segona part del treball es descriuen els mòduls aresta de Francis Ow i el mòdul aresta Kantenmodul de Carmen Sprung i es veuen algunes de les seves possibilitats. En aquesta part del treball descriurem una generalització dels mòduls aresta de 120 i 135 graus de Francis Ow. Aquesta generalització descriu un procediment que ens permet canviar l'angle general per un altre de concret, obtenir casos particulars de mòduls aresta i construir altres tipus de políedres. Els mòduls aresta de 120 i 135 graus de Francis Ow són casos particulars d'aquest mètode més general.

A la tercera part del treball es descriu el mòdul de Sonobe i també es veuen algunes de les seves possibilitats, en particular la construcció de fractals, com l'esponja de Menger. En aquest apartat es calcula i es demostra quin és el gènere de la superfície que s'obté en l'enèsima iteració per definir aquest fractal.

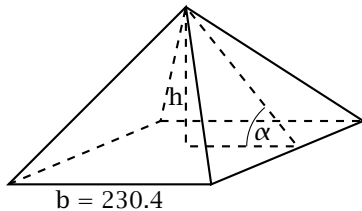
La darrera part del treball està dedicada a la construcció amb papiroflèxia modular de mosaics plans. Veurem que alguns motius de l'Alhambra de Granada també poden tessel·lar un cub i en farem efectius dos: l'avió nassarita i l'os nassarita. El treball finalitza amb la construcció d'una figura no modular multiforme inspirada en un tercer motiu dels mosaics de l'Alhambra de Granada: un tetraèdre no modular construït a partir d'un ocellet nassarita tridimensional.

1 Políedres

1.1 Els humans i els políedres

Fa milers d'anys que els humans coneixen els políedres. Un dels exemples més antics, enigmàtic i emotiu és la gran piràmide de Kheops, segon faraó de la IV dinastia egípcia. Aquesta construcció faraònica està situada a la planura de Guiza, avui dia al costat de la ciutat del Caire o Al-Qahira (la Victòria). Essent

la més gran de totes les piràmides egípcies, té una alçària de $h = 146.6$ m i el costat de la base fa $b = 230.4$ m de longitud, de la qual cosa podem deduir que la seva inclinació té un valor numèric que es pot aproximar per l'arrel quadrada del nombre auri $m = \tan(\alpha) \approx \sqrt{\phi}$, amb un error relatiu del 0.04 %, o també un valor aproximat de $m = \tan(\alpha) \approx \sqrt[12]{18}$, amb un error relatiu del 0.01 %. En la resta de les piràmides de l'antic Egipte no es donen aquestes aproximacions numèriques i, per tant, cal esperar que aquestes i d'altres aproximacions expressades amb funcions senzilles siguin casualitats numèriques.



Altres políedres força enigmàtics es poden veure en el Museu Galoromà Lugdunum de la ciutat francesa de Lió. A més de daus i compassos de l'època romana, aquest museu té dos dodecaèdres de bronze datats al voltant dels segles II i III de la nostra era. En les seves dotze cares foradades per cercles de diferents mides es poden veure algunes marques circulars, i en cadascun dels seus vint vèrtexs que connecten les trenta arestes hi ha unes petites esferes. Avui dia es coneixen més de cent d'aquests dodecaèdres, la majoria trobats a França, i un únic icosaèdre al LVR-Landesmuseum de Bonn. Encara no es coneix quina era la utilitat d'aquests objectes perquè no se n'han trobat referències històriques documentades.

Com han arribat els humans a conèixer figures geomètriques tan regulars? Una part de la resposta és que observessin aquestes formes geomètriques a la natura. Molts minerals cristallitzen macroscòpicament en diferents tipus de políedres força regulars. Aquest procés natural es produeix a l'escorça de la terra des de milions d'anys abans de l'aparició dels humans. Així, tenim que la pirita cristallitza en cubs, octàedres i dodecaèdres, i la fluorita i la magnetita, en octàedres gairebé perfectes per a l'ull humà. Aquests són tres exemples de políedres en què totes les cares són polígons regulars iguals i en tots els vèrtexs conflueix el mateix nombre de cares. Aquests políedres s'anomenen *políedres regulars* o *sòlids platònics* en honor de Plató, que els menciona en el *Timeu* [10]: «El foc està format per tetràedres; l'aire, d'octàedres; l'aigua, d'icosaèdres; la terra, de cubs; com que encara és possible una cinquena forma, Déu ha utilitzat aquesta, el dodecaèdre pentagonal, perquè serveixi de límit al món».

La pirita també pot cristallitzar en forma de cubooctàedre, que és un políedre semiregular format per sis quadrats i vuit hexàgons regulars. Un darrer exemple és el granat, que pot cristallitzar com a dodecaèdre ròmbic, que és un dels tretze políedres de Catalan.

Els políedres semiregulars també s'anomenen *políedres* o *sòlids arquimedians* en honor d'Arquimedes de Siracusa (segle III aC). Tot i que no es conserven traduccions de l'obra original d'Arquimedes que parlin d'aquests políedres, hi ha referències d'autors grecs, com Papos d'Alexandria (segle IV dC), que en el llibre V secció XIX de la seva *Collecció matemàtica*, dedicada als cossos sòlids, descriu els cinc políedres platònics i tretze políedres formats per diferents polígons equilàters i equiangles que atribueix a Arquimedes [19].

1.2 Políedres platònics i arquimedians

Els políedres estan formats per elements de dimensió 2 (les cares C), de dimensió 1 (les arestes A) i de dimensió zero (els vèrtexs V). Per a qualsevol políedre homeomorf a l'esfera, aquests elements estan relacionats per la fórmula d'Euler:

$$C + V = A + 2. \quad (1)$$

Leonhard Euler (Basilea, 1707 - Sant Petersburg, 1783) va proposar aquesta igualtat en una carta escrita a Berlín amb data 14 de novembre de 1750 i dirigida al seu amic Christian Goldbach, que vivia a Sant Petersburg. La primera demostració rigorosa de la fórmula d'Euler no fou donada fins al 1847 per Karl von Staudt [16]. Aquesta equació diofàntica lineal (1) té infinites solucions, però no totes les solucions es corresponen a un políedre; per exemple, no existeix cap políedre amb els valors $C = V = A = 2$, tot i que aquests valors compleixen la relació d'Euler $2 + 2 = 2 + 2$.

A continuació ens centrarem en els políedres convexos, que tenen certa regularitat en les cares i en els vèrtexs.

DEFINICIÓ 1. Direm que un políedre convex és regular o platònic si totes les seves cares són polígons regulars d'un únic tipus i en cada vèrtex conflueix el mateix nombre de cares.

Altres políedres força regulars són els prismes regulars i els antiprismes regulars. Aquests políedres estan formats per dos polígons regulars paral·lels de n costats units per n quadrats i per $2n$ triangles equilàters, respectivament. Existeixen infinits prismes regulars i infinits antiprismes regulars, un per cada polígon regular. El prisma regular amb base un quadrat és el cub i l'antiprisma regular amb base un triangle equilàter és l'octàedre.

DEFINICIÓ 2. Direm que un políedre convex, que no sigui un prisma o antiprisma, és semiregular o arquimedià si totes les seves cares són polígons regulars de dos o més tipus i en cada vèrtex conflueix el mateix tipus de cares i en el mateix ordre.

Per demostrar que tan sols existeixen cinc políedres platònics i tretze políedres arquimedians, imposarem les respectives condicions de regularitat a la fórmula d'Euler.

TEOREMA 3. *Tan sols existeixen cinc políedres regulars o platònics.*

PROVA. Cada aresta està limitada per dos vèrtexs. A cada aresta conflueixen dues cares, i dues cares comparteixen com a màxim una única aresta.

Com que el poliedre és regular, totes les cares que el formen són polígons regulars amb n arestes i n vèrtexs; a més, a cada vèrtex del poliedre conflueixen m cares i m arestes. Evidentment, $n \geq 3$ i $m \geq 3$. Aquesta informació ens permet obtenir relacions entre el nombre de cares C , el nombre d'arestes A i el nombre de vèrtexs V :

$$A = \frac{nC}{2}, \quad V = \frac{nC}{m}, \quad V = \frac{2A}{m}.$$

Substituint aquestes expressions en la fórmula d'Euler (1), obtenim els valors $C, A, V \in \mathbb{N}$ en funció dels paràmetres naturals $n, m \in \mathbb{N}$:

$$\begin{aligned} C + \frac{nC}{m} &= \frac{nC}{2} + 2 \Rightarrow C = \frac{4m}{2n + 2m - nm}, \\ \frac{2A}{n} + \frac{2A}{m} &= A + 2 \Rightarrow A = \frac{2nm}{2n + 2m - nm}, \\ \frac{mV}{n} + V &= \frac{mV}{2} + 2 \Rightarrow V = \frac{4n}{2n + 2m - nm}. \end{aligned} \quad (2)$$

Ara cal resoldre aquestes tres noves equacions diofàntiques algebraiques. Primer de tot veurem que un dels paràmetres val 3: $n = 3$ o bé $m = 3$. En efecte, és evident que el denominador d'aquestes expressions és positiu:

$$2n + 2m - nm > 0 \Rightarrow \frac{1}{n} + \frac{1}{m} > \frac{1}{2}. \quad (3)$$

Si els dos paràmetres són diferents de 3, aleshores $n \geq 4$ i $m \geq 4$, i obtenim la contradicció següent:

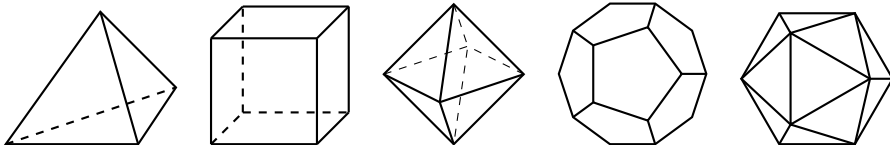
$$\frac{1}{2} = \frac{1}{4} + \frac{1}{4} \geq \frac{1}{n} + \frac{1}{m} > \frac{1}{2}.$$

Si $n = 3$, de la desigualtat (3) deduïm que $m \leq 5$, i per simetria de les variables n, m en la mateixa desigualtat (3), tenim que $n \leq 5$ si $m = 3$. Així doncs, tenim cinc possibilitats, i amb les expressions (2) podem deduir el nombre de cares, arestes i vèrtexs.

n	m	C	A	V	
3	5	20	30	12	Icosàedre
3	4	8	12	6	Octàedre
3	3	4	6	4	Tetràedre
4	3	6	12	8	Cub
5	3	12	30	20	Dodecàedre

□

Observem que el nombre d'arestes sempre és un múltiple de 6 i que el nombre de cares i de vèrtexs és parell.



Canviant la condició de convex per còncav, s'obté una família de quatre políedres regulars còncavs, coneguts amb el nom de *políedres de Kepler-Poinsot*.

Seguint els passos de la demostració anterior i amb un argument més elaborat, es pot demostrar que tan sols existeixen tretze políedres arquimedians o semiregulars [2].

Prèviament veurem tres lemes que ens permetran descartar molts possibles casos. Definim s com el nombre de tipus de polígons regulars diferents que té el políedre arquimièdia.

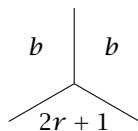
LEMA 4. *En cada vèrtex d'un políedre semiregular o arquimièdia no poden confluïr quatre o més tipus de polígons diferents. És a dir, $s \leq 3$.*

PROVA. Suposem el contrari; aleshores els angles interiors dels polígons regulars que formin el políedre arquimièdia seran superiors o iguals a $\frac{\pi}{3}$ rad (60 graus, triangle equilàter), $\frac{\pi}{2}$ rad (90 graus, quadrat), $\frac{3\pi}{5}$ rad (108 graus, pentàgon regular) i $\frac{2\pi}{3}$ rad (120 graus, hexàgon regular). D'aquí obtenim la contradicció següent:

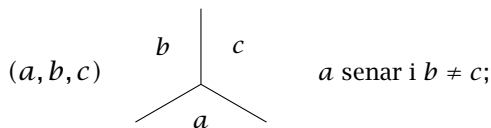
$$\frac{\pi}{3} + \frac{\pi}{2} + \frac{3\pi}{5} + \frac{2\pi}{3} = \frac{21}{10}\pi > 2\pi,$$

perquè la suma dels angles interiors dels polígons regulars que conflueixen en un mateix vèrtex ha de ser inferior a 2π rad, ja que el políedre és convex. $\square$

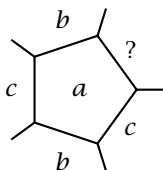
LEMA 5. *Si a cada vèrtex d'un políedre semiregular o arquimièdia hi conflueixen tres polígons i un d'ells té un nombre senar de cares, aleshores els altres dos polígons han de tenir el mateix nombre de cares.*



PROVA. Suposem el contrari, és a dir, existeix una configuració del tipus

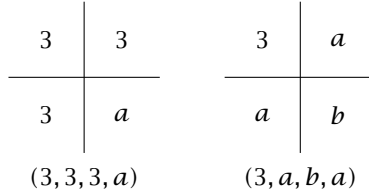


aleshores un dels vèrtexs del polígon amb un nombre senar de costats no tindrà la mateixa configuració, degut a l'alternança dels valors b, c al voltant del polígon amb un nombre senar de costats. Ho il·lustrem amb un pentàgon.



$\square$

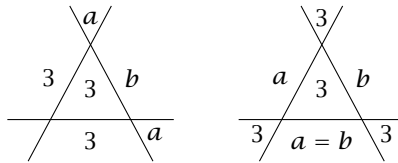
LEMA 6. Si a cada vèrtex d'un políedre semiregular o arquimedià hi conflueixen quatre polígons i com a mínim un d'ells és un triangle equilàter, aleshores les úniques configuracions de vèrtexs són $(3, 3, 3, a)$ i $(3, a, b, a)$ amb $a \neq 3$, és a dir, podem aconseguir una configuració de vèrtexs on un dels triangles equilàters té els dos polígons adjacents amb el mateix nombre de cares.



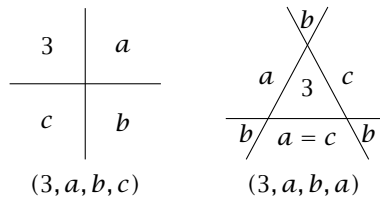
PROVA. Si a cada vèrtex conflueixen quatre triangles equilàters, aleshores tenim un políedre regular.

Si a cada vèrtex conflueixen exactament tres triangles equilàters, aleshores obtenim la configuració de vèrtexs dels antiprismes $(3, 3, 3, n)$.

Si a cada vèrtex conflueixen exactament dos triangles equilàters, aleshores tenim dues possibilitats $(3, 3, a, b)$ (triangles equilàters adjacents) i $(3, a, 3, b)$ (triangles equilàters oposats). Els dos casos no són possibles com s'il·lustra en el dibuix següent:



Si a cada vèrtex conflueix exactament un triangle equilàter, aleshores en la configuració $(3, a, b, c)$ la cara de b costats sempre està oposada al triangle equilàter, de la qual cosa es dedueix que $a = c$.



□

TEOREMA 7. Tan sols existeixen tretze políedres semiregulars o arquimedians.

PROVA. La demostració d'aquest teorema es divideix en l'anàlisi de diferents casos. Observem que en un mateix vèrtex no poden confluir més de cinc polígons, perquè aleshores la suma dels angles seria superior als 2π rad. Utilitzem la notació següent: $n(c)$ vol dir n polígons de c costats, $\{\geq n\}(c)$ vol dir un nombre superior o igual a n polígons de c costats, $n(2m \geq c)$ vol dir n polígons amb c o més costats que compleixen la condició $m \geq c/2$.

1. Poliedres formats tan sols per triangles equilàters i quadrats.

(4, 3, 3)	en contradicció amb el lema 5
(4, 3, 3, 3)	antiprisma quadrangular
(4, 3, 3, 3, 3)	$s = 2, m_3 = 4, m_4 = 1, m = 5$ (I) cub xato
$1(4) + \{\geq 5\}(3)$	La suma d'angles supera 2π
(4, 4, 3)	prisma triangular
(4, 4, 3, 3)	en contradicció amb el lema 6
(4, 3, 4, 3)	$s = 2, m_3 = 2, m_4 = 2, m = 4$ (II) cubooctàedre
$2(4) + \{\geq 3\}(3)$	La suma d'angles val $\geq 2\pi$
(4, 4, 4, 3)	$s = 2, m_3 = 1, m_4 = 3, m = 5$ (III) rombicuboctàedre
$3(4) + \{\geq 2\}(3)$	La suma d'angles supera 2π
$\{\geq 4\}(4) + \{\geq 1\}(3)$	La suma d'angles supera 2π

2. Poliedres formats tan sols per triangles equilàters i pentàgons.

(5, 3, 3)	en contradicció amb el lema 5
(5, 3, 3, 3)	antiprisma pentagonal
(5, 3, 3, 3, 3)	$s = 2, m_3 = 4, m_5 = 1, m = 5$ (IV) dodecàedre xato
$1(5) + \{\geq 5\}(3)$	La suma d'angles supera 2π
(5, 5, 3)	en contradicció amb el lema 5
(5, 5, 3, 3)	en contradicció amb el lema 6
(5, 3, 5, 3)	$s = 2, m_3 = 2, m_5 = 2, m = 4$ (V) icosidodecàedre
$2(5) + \{\geq 3\}(3)$	La suma d'angles supera 2π
$\{\geq 3\}(5) + \{\geq 1\}(3)$	La suma d'angles supera 2π

3. Poliedres formats tan sols per triangles equilàters i hexàgons.

(6, 3, 3)	en contradicció amb el lema 5
(6, 3, 3, 3)	antiprisma hexagonal
$1(6) + \{\geq 4\}(3)$	La suma d'angles val $\geq 2\pi$
(6, 6, 3)	$s = 2, m_3 = 1, m_6 = 2, m = 3$ (VI) tetràedre truncat
$2(6) + \{\geq 2\}(3)$	La suma d'angles val $\geq 2\pi$
$\{\geq 3\}(6) + \{\geq 1\}(3)$	La suma d'angles val $\geq 2\pi$

4. Políedres formats tan sols per triangles equilàters i n -àgons, $n \geq 7$.

$(n, 3, 3)$	en contradicció amb el lema 5
$(n, 3, 3, 3)$	antiprisma n -gonal
$1(n) + \{\geq 4\}(3)$	La suma d'angles supera 2π
$(8, 8, 3)$	$s = 2, m_3 = 1, m_8 = 2, m = 3$ (VII) cub truncat
$(10, 10, 3)$	$s = 2, m_3 = 1, m_{10} = 2, m = 3$ (VIII) dodecàedre truncat
$(2n + 1, 2n + 1, 3)$	en contradicció amb el lema 5
$2(2m \geq 12) + \{\geq 1\}(3)$	La suma d'angles supera 2π
$2(n) + \{\geq 2\}(3)$	La suma d'angles supera 2π
$\{\geq 3\}(n) + \{\geq 1\}(3)$	La suma d'angles supera 2π

5. Políedres formats tan sols per quadrats i n -àgons, $n \geq 5$.

$(n, 4, 4)$	prisma n -gonal
$1(n) + \{\geq 3\}(4)$	La suma d'angles supera 2π
$(2n + 1, 2n + 1, 4)$	en contradicció amb el lema 5
$(6, 6, 4)$	$s = 2, m_4 = 1, m_6 = 2, m = 3$ (IX) octàedre truncat
$(6, 6, 4, 4)$	La suma d'angles supera 2π
$2(2m \geq 8) + \{\geq 1\}(4)$	La suma d'angles supera 2π
$\{\geq 3\}(n) + \{\geq 1\}(4)$	La suma d'angles supera 2π

6. Políedres formats tan sols per pentàgons i n -àgons, $n \geq 6$.

$(n, 5, 5)$	en contradicció amb el lema 5
$1(n) + \{\geq 3\}(5)$	La suma d'angles supera 2π
$(6, 6, 5)$	$s = 2, m_5 = 1, m_6 = 2, m = 3$ (X) icosaèdre truncat
$2(2m \geq 7) + \{\geq 1\}(5)$	La suma d'angles supera 2π
$\{\geq 3\}(n) + \{\geq 1\}(5)$	La suma d'angles supera 2π

7. Políedres formats tan sols per dos tipus diferents de n -àgons, $n \geq 6$.

Aquest cas és impossible perquè la suma dels angles interns sempre és superior a 2π .

8. Políedres formats tan sols per triangles equilàters, quadrats i n -àgons, $n \geq 5$.

$(n, 4, 3)$	en contradicció amb el lema 5
$(n, 4, 3, 3)$	en contradicció amb el lema 6
$(n, 3, 4, 3)$	en contradicció amb el lema 6
$(n, 3, 3, 4)$	en contradicció amb el lema 6
$(5, 4, 4, 3)$	en contradicció amb el lema 6

$(5, 4, 3, 4)$	$s = 3, m_3 = 1, m_4 = 2, m_5 = 1, m = 4$
	(XI) petit rombicododecàedre
$(5, 3, 4, 4)$	en contradicció amb el lema 6

$1(n) + 1(4) + \{\geq 3\}(3)$	La suma d'angles supera 2π
$1(n \geq 6) + \{\geq 1\}(4) + \{\geq 1\}(3)$	La suma d'angles supera 2π
$\{\geq 2\}(n) + \{\geq 1\}(4) + \{\geq 1\}(3)$	La suma d'angles supera 2π

9. Poliedres formats tan sols per tres tipus de polígons que no siguin triangles.

En aquest darrer cas la suma de quatre angles interns diferents sempre és superior a 2π :

$$\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 > 4 \cdot \pi / 2 = 2\pi.$$

Per tant, en cada vèrtex conflueixen tres polígons regulars diferents que, degut al lema 5, tindran un nombre parell de costats; en aquesta situació les úniques combinacions que no superen els 2π radians són:

$(8, 6, 4)$	$s = 3, m_4 = 1, m_6 = 1, m_8 = 1, m = 3$
	(XII) gran rombicuboctàedre
$(10, 6, 4)$	$s = 3, m_4 = 1, m_6 = 1, m_{10} = 1, m = 3$
	(XIII) gran rombicododecàedre

□

Els arguments anteriors són de tipus local i no hem utilitzat la característica d'Euler-Poincaré del poliedre. Per calcular el nombre exacte de cares, vèrtexs i arestes d'aquests poliedres, definim C_n com el nombre de n -àgons de n arestes i n vèrtexs d'un poliedre arquimedià, m com el nombre de cares o arestes que conflueixen en un vèrtex i m_i com el nombre de cares n_i -gonals que conflueixen en un vèrtex.

TEOREMA 8. *El nombre de vèrtexs, cares i arestes dels poliedres arquimedians ve donat per la taula de valors següent.*

	Poliedre	Vèrtexs	Arestes	Cares
(I)	<i>cub xato</i>	24	60	38
(II)	<i>cuboctàedre</i>	12	24	14
(III)	<i>petit rombicuboctàedre</i>	24	48	26
(IV)	<i>dodecàedre xato</i>	60	150	92
(V)	<i>icosidodecàedre</i>	30	60	32
(VI)	<i>tetràedre truncat</i>	12	18	8
(VII)	<i>cub truncat</i>	24	36	14
(VIII)	<i>dodecàedre truncat</i>	60	90	32
(IX)	<i>octàedre truncat</i>	24	36	14
(X)	<i>icosàedre truncat</i>	60	90	32
(XI)	<i>petit rombicododecàedre</i>	60	120	62
(XII)	<i>gran rombicuboctàedre</i>	48	72	26
(XIII)	<i>gran rombicododecàedre</i>	120	180	62

PROVA. Utilitzant el mateix argument que en el cas platònic deduïm:

$$C = \sum_{i=1}^s C_{n_i}, \quad A = \frac{1}{2} \sum_{i=1}^s n_i C_{n_i}, \quad V = \frac{1}{m} \sum_{i=1}^s n_i C_{n_i}. \quad (4)$$

En aquest cas el nombre de vèrtexs també compleix les igualtats següents:

$$V = \frac{n_i C_{n_i}}{m_i}, \quad \forall i = 1, \dots, s.$$

Aleshores podem deduir els valors de C_{n_i} resolent el sistema de s equacions lineals en C_{n_i} format per la relació donada per la fórmula d'Euler i $s - 1$ de les equacions anteriors:

$$\sum_{i=1}^s \left(1 - \frac{n_i}{2} + \frac{n_i}{m}\right) C_{n_i} = 2,$$

$$\frac{n_1 C_{n_1}}{m_1} = \frac{n_i C_{n_i}}{m_i}, \quad \forall i = 2, \dots, s - 1.$$

Concretament, si a cada vèrtex conflueixen dos tipus de polígons ($s = 2$), cal resoldre un sistema de dues equacions i dues incògnites:

$$\begin{pmatrix} 1 - \frac{n_1}{2} + \frac{n_1}{m} & 1 - \frac{n_2}{2} + \frac{n_2}{m} \\ -\frac{n_1}{m_1} & \frac{n_2}{m_2} \end{pmatrix} \begin{pmatrix} C_{n_1} \\ C_{n_2} \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

i, en el cas que en cada vèrtex conflueixin tres tipus de polígons ($s = 3$), cal resoldre un sistema de tres equacions i tres incògnites:

$$\begin{pmatrix} 1 - \frac{n_1}{2} + \frac{n_1}{m} & 1 - \frac{n_2}{2} + \frac{n_2}{m} & 1 - \frac{n_3}{2} + \frac{n_3}{m} \\ -\frac{n_1}{m_1} & \frac{n_2}{m_2} & 0 \\ -\frac{n_1}{m_1} & 0 & \frac{n_3}{m_3} \end{pmatrix} \begin{pmatrix} C_{n_1} \\ C_{n_2} \\ C_{n_3} \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}.$$

Els valors m i n_i , m_i , $\forall i = 1, \dots, s$, els hem determinat en la demostració del teorema 7, cosa que permet calcular els valors C_{n_i} , $\forall i = 1, \dots, s$, i també la terna de valors (C, V, A) per a cada tipus de políedre mitjançant les igualtats (4). $\square$

El cub xato i el dodecàedre xato existeixen en dues variants que són isomètriques: la forma levogira i la forma dextrogira. El rombicuboctàedre està format per dues cúpules i un prisma octogonal central; girant 45 graus una d'aquestes cúpules, s'obté l'anomenat *pseudorombicuboctàedre*, també anomenat *sòlid de Sommerville* [14] o *sòlid de Miller* [7]. Aquest políedre té les mateixes propietats de regularitat en els vèrtexs que el rombicuboctàedre, però té menys simetries axials i especulars. El rombicuboctàedre i el pseudorombicuboctàedre no són figures isomètriques.

2 Papiroflèxia d'alguns mòduls aresta

2.1 Què és la papiroflèxia?

Aquesta mateixa pregunta es devien fer el novembre de 1936 els lectors del diari La Prensa de Buenos Aires, en el qual, en una entrevista de Carlos A. Leumann a l'odontòleg burgalès Vicente Solórzano Sagredo, apareix per primer cop la paraula *papiroflèxia*. De fet, en aquest article surten dues paraules creades pel docor Solórzano (vegeu també [12, 13]):

- «papirola: todas las producciones científico-artísticas realizadas con papel.»
- «papiroflexia: El arte de doblar geométricamente el papel.»

Aquestes dues paraules, així com la paraula *papirofléxico*, no van ser incloses al diccionari de la RAE fins l'any 1986, tot i que sí que apareixien en altres diccionaris. El doctor Solórzano també va fundar l'any 1954 el Museu de Papiroflèxia a la ciutat de Buenos Aires. Posteriorment, l'any 1961, aquesta collecció es va traslladar a Valladolid, on un dels seus companys d'estudis, Nemesio Montero Pérez, la va dipositar en el Museu de Ciències Naturals d'aquesta ciutat. Amb els anys, i després d'algunes gestions amb les germanes del doctor Montero, les figures d'aquesta collecció van ser cedides a l'Asociación Española de Papiroflexia (AEP), on es conserven actualment en dos grans baguls, que estan pendants d'estudi i catalogació.

Trenta-sis anys abans de la publicació del llibre del doctor Solórzano, quan encara no existia la paraula *papiroflèxia*, l'escriptor i filòsof Miguel de Unamuno va publicar un apèndix al llibre *Amor y pedagogía* amb el títol *Apuntes para un tratado de cocotología*; en aquest apèndix podem llegir el següent [18]:

La palabra *cocotología* se compone de dos, de la francesa *cocotte*, pajarita de papel, y de la griega *logia*, de *logos*, tratado. La palabra francesa *cocotte* es una palabra infantil y que se aplica en su sentido primitivo y recto a los pollos y por extensión a las aves. En sentido traslaticio, a las pajaritas de papel y a las mozas de vida alegre.

Poques llengües del món tenen una paraula pròpia equivalent a *papiroflèxia*. Trobem la paraula *paper folding* en anglès, *Papierfalten* en alemany, *zhézhǐ* en xinès i *origami* en japonès (de *oru*, «plegar», i *kami*, «paper»). Aquesta darrera paraula és la que s'utilitza en la majoria de llengües del món per designar aquest art de plegar el paper.

Avui dia no està clar que la papiroflèxia s'originés al Japó. D'aquest país, a més d'imposar-se la paraula *origami* en les altres llengües del món, es coneix el llibre més antic dedicat íntegrament a la papiroflèxia. És un anònim de 1797 que porta per títol *Hiden Senbazuru Orikata*, i que podríem traduir per *El secret del plegatge de les 1000 grues*, on es descriuen les construccions de diferents tipus de grues de paper. Del Japó també es coneix documentalment un cub de paper anomenat *Tamatebako* o cofre del tresor, publicat l'any 1734 per Hayato Ohoka en el llibre *Ranma Zushiki*.

A Occident la figura d'origami documentada més antiga que es coneix és la caixa del *Llibre d'hores*, de Caterina de Clèveris (segle xv). Uns 500 anys més

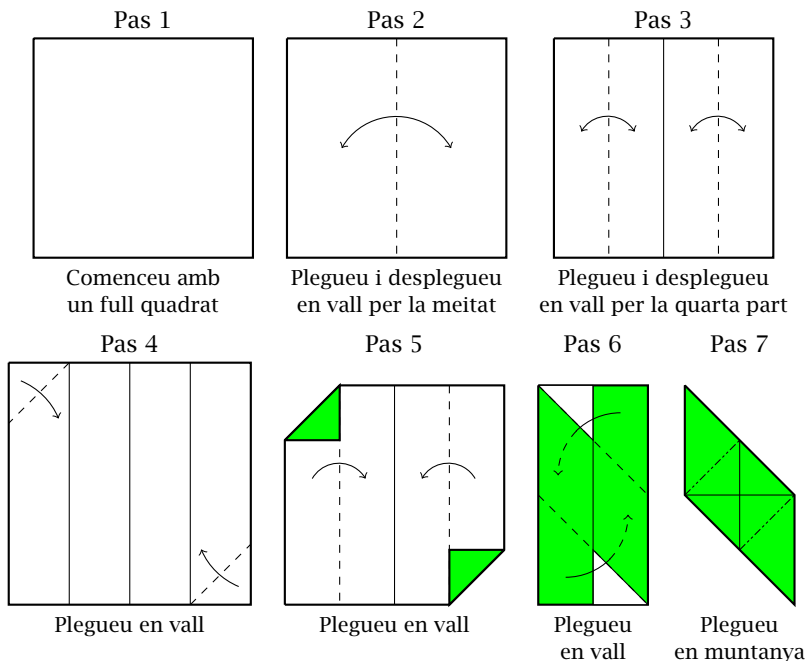
tard podem trobar la descripció de la seva construcció en la figura número 6 del llibre anònim *Una hoja de papel*, publicat per l'Editorial Salvatella [1]. Vicente Palacios Garrido, expresident de l'AEP, va esbrinar que aquest llibre fou escrit per Lorenzo Herrero Sáiz. Molt probablement Lorenzo Herrero no coneixia el *Llibre d'hores*, de Caterina de Clèveris i, per tant, cal esperar que aquesta caixa i d'altres figures tradicionals i molt populars que apareixen en el mateix llibre, com l'*ocellet* (*pajarita* en castellà, *cocotte* en francès), s'hagin transmès per tradició oral al llarg dels segles.

A més de la papiroflèxia clàssica, en què es construeix una figura a partir d'un únic paper, que normalment és un full dinA4 o un paper quadrat, avui dia existeixen diferents branques de la papiroflèxia, com la papiroflèxia modular, en què es construeixen figures de paper unint diverses còpies d'un mateix mòdul (o de diferents mòduls, papiroflèxia multimodular), la papiroflèxia cinètica, en què es construeixen figures de paper amb moviment, o la papiroflèxia multiforme, on apareixen diferents figures en una successió de plects.

En el material suplementari¹ veurem com podem construir amb paper els cinc sòlids platònics i els tretze políedres arquimediàns amb tècniques de papiroflèxia modular.

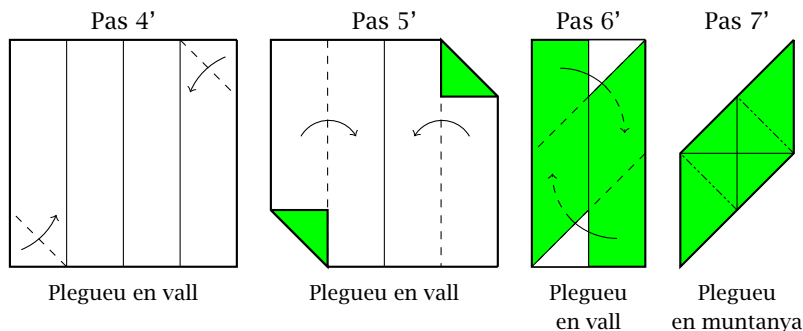
3 El cub de Sonobe

El cub de Sonobe és el cub de paper més popular. Aquest cub és modular i està format per sis mòduls iguals, un per cada cara del cub. El mòdul de Sonobe es construeix fent uns determinats plects en un full quadrat de paper.

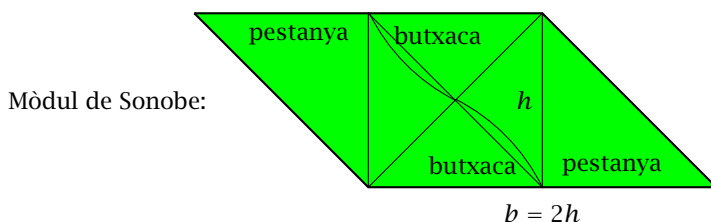


¹ Vegeu el material suplementari a l'edició digital d'aquest volum, *Butlletí de la Societat Catalana de Matemàtiques* (en línia) (2024), vol. 39 (1-2), <https://revistes.iec.cat/index.php/BSCM>.

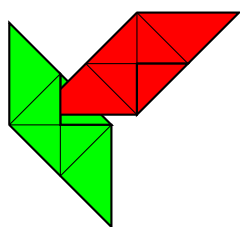
Observem que la construcció del mòdul de Sonobe és asimètrica; en efecte, en el pas 4 hem fet els plecs en dos vèrtexs oposats del quadrat. També es poden fer els plecs triant els altres dos vèrtexs oposats del quadrat i repetir els passos 4 a 7; d'aquesta manera obtenim un mòdul simètric.



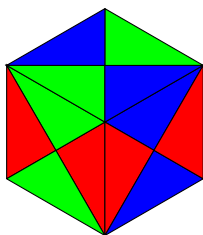
El mòdul de Sonobe té forma de romboide amb angles de 45 i 135 graus, i la seva base és el doble de la seva alçària. Podeu comprovar que cada punt del romboide té quatre capes de paper i, per tant, la seva superfície és la quarta part del quadrat inicial. Per unir els sis mòduls romboidals i formar el cub no cal utilitzar cola; s'uneixen perquè cada mòdul té dues butxaques i dues pestanyes. Les dues butxaques estan situades a la diagonal menor del romboide, que també és una diagonal del quadrat central, i les pestanyes són els dos triangles rectangles isòscels laterals.



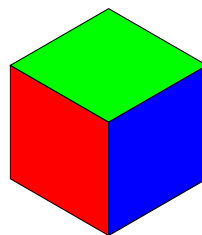
Per unir dos mòduls, en girem un 90 graus i introduïm una pestanya del primer mòdul en una butxaca del segon. La resta dels mòduls s'uneixen de la mateixa manera. El mòdul obtingut en el pas 7 no es pot encaixar de manera natural amb el mòdul del pas 7'. Per construir un cub cal fer sis mòduls idèntics com en el pas 7 o sis mòduls idèntics com el pas 7'; en els dos casos l'efecte visual és el mateix. Si en el pas 7 o 7' fem un plec en vall en lloc de muntanya, aleshores, quan unim sis mòduls, obtenim un cub amb les cares llises, no dividides en triangles.



Unió de dos mòduls



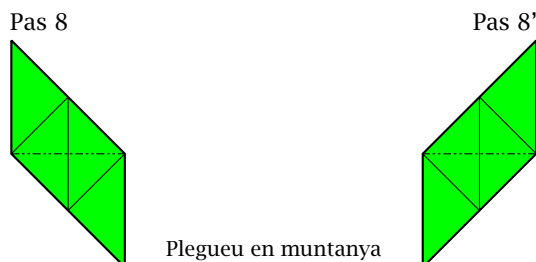
Unió de sis mòduls del pas 7 o 7' en muntanya



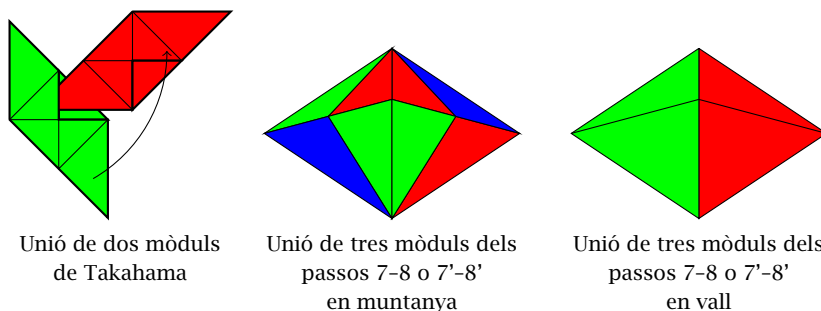
Unió de sis mòduls del pas 7 o 7' en vall

3.1 La joia de Takahama

Fent un nou plec en el mòdul de Sonobe s'obté una variant que en aquesta secció anomenaré *mòdul de Takahama*.



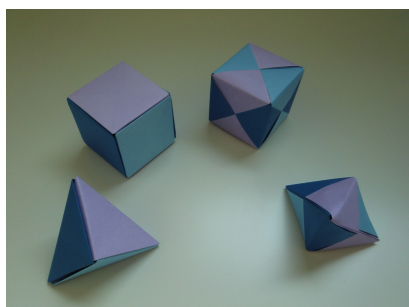
Unint tres mòduls de Takahama obtenim un altre hexàedre: la bipiràmide triangular. La unió de dos mòduls de Takahama es fa com en el cas del mòdul de Sonobe. Les altres propietats del mòdul de Takahama són semblants a les del mòdul de Sonobe. El mòdul obtingut en el pas 8 no es pot encaixar de manera natural amb el mòdul del pas 8'. Per construir una bipiràmide triangular cal fer tres mòduls idèntics com en el pas 8, o tres mòduls idèntics com en el pas 8'; en els dos casos l'efecte visual és el mateix. Si en els passos 7-8 o 7'-8' fem plects en vall en lloc de muntanya, aleshores, quan unim tres mòduls, obtenim una bipiràmide triangular amb les cares llises no dividides en triangles.



Nota històrica: segons l'historiador de l'origami David Lister (Grimsby, 1930–2013), la joia de Takahama fou publicada per primer cop l'any 1974 en el llibre *Vida creativa amb origami creatiu* [17], escrit en japonès per l'origamista nipona Toshie Takahama (Tòquio, 1910–1999). El nom de *joia* és degut al fet que Toshie Takahama feia collarets unint uns quants d'aquests hexaedres amb una corda. A finals de la dècada dels anys seixanta, Toshie Takahama va fundar el Sosaku Origami Group 67 (grup de creació o d'estudi de l'origami), del qual formaven part, entre d'altres, Mitsunobu Sonobe i Kunihiro Kasahara. Aquest petit grup de plegadors es reunia periòdicament a l'escola Tamagawa Gakuen, a l'oest de Tòquio.

David Lister també afirma que el cub de Sonobe fou publicat per primer cop l'any 1968 a les pàgines 10 i 11 de la segona de les set revistes editades pel Sosaku Origami Group 67.

Amb les publicacions de Toshie Takahama es va donar a conèixer el mòdul de Sonobe. Així, el cub de Sonobe i la joia de Takahama van ser les figures modulars que van donar lloc a l'actual especialitat de l'origami modular. El cub de Sonobe es va publicar posteriorment en diverses revistes i llibres de papiroflèxia amb el nom de *caixa de colors*, com per exemple en la pàgina 19 del número 0, volum 3, del butlletí de la NOA (Nippon Origami Association), de l'any 1974, del qual era directora Toshie Takahama.



Cubs de Sonobe i joies de Takahama

Toshie Takahama i Mitsunobu Sonobe

Mitsunobu Sonobe (園部 光伸) va néixer l'any 1937. En el llibre *Atarashii Origami Nyumon* [15], editat pel Sosaku Origami Group 67 l'any 1970, en el qual apareix publicada la caixa de colors, podem llegir que tenia sis anys d'experiència en el plegatge de paper i que vivia al barri d'Ikebukuro, en el districte de Toshima-ku de la ciutat de Tòquio. Treballava en un grup de teatre i el que més li agradava era la creació de tot tipus de flors de paper. En el mateix llibre es fan petits apunts biogràfics dels altres membres del Sosaku Origami Group 67.

Als anys setanta, Toshie Takahama era més coneguda que Mitsunobu Sonobe i durant algunes dècades el mòdul de Sonobe va ser conegut a Occident com a *mòdul de Takahama*, però és clar que l'autor d'aquest mòdul és Mitsunobu Sonobe, com apareix en el llibre *Papiroflexia para expertos* [6], dels mestres d'origami Kunihiro Kasahara (笠原 邦彦) i Toshie Takahama (高濱 利恵), contemporanis de Mitsunobu Sonobe.

Cal destacar la semblança entre el cub de Sonobe i el cub de paper anomenat *tamatebako* o *cofre del tresor*, que apareix en una il·lustració del llibre *Ranma Zushiki*, publicat l'any 1734 per Hayato Ohoka.

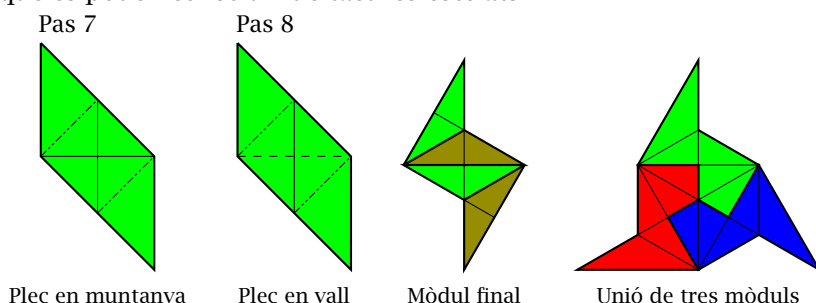
Podeu trobar un altre cub de paper força diferent del de Sonobe en el llibre *Guía práctica del arte manual* [3], publicat l'any 1977, tot i que les instruccions de la seva construcció no són gaire clares.

3.2 Deltaedres estelats

Els deltaedres són poliedres totes les cares dels quals són triangles equilàters. Existeixen infinits deltaedres còncavs i tan sols vuit deltaedres convexos de 4, 6, 8, 10, 12, 14, 16 i 20 cares, com van demostrar l'any 1947 Freudenthal i Van der Waerden [4]. Els valors mínim i màxim del nombre de cares d'un deltaedre

convex és 4 i 20, respectivament, que es corresponen als sòlids platònics o regulars amb els valors mínim i màxim de cares: el tetràedre i l'icosàedre. Existeix un tercer políedre, l'octàedre, que és deltaedre i sòlid platònic alhora. L'octàedre també es pot interpretar com una bipiràmide quadrangular. Els altres cinc deltaedres (les bipiràmides triangular i pentagonal, el dodecadeltaedre, el tetracaicadeltaedre i l'hexacaicadeltaedre) no són regulars en els seus vèrtexs. Cal destacar que no existeixen deltaedres convexos de 18 cares o d'un nombre de cares superior a 20.

El mòdul de Sonobe té moltes més possibilitats. Si en el pas 7 fem el plec en muntanya i en el pas 8 fem el plec en vall, aleshores obtenim un nou mòdul amb què es poden construir deltaedres estelats.

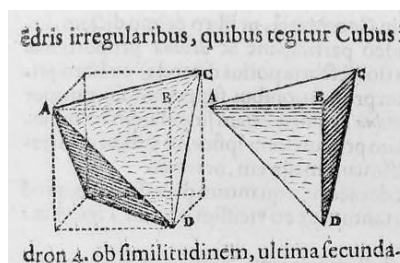


Unint tres d'aquests mòduls obtenim una piràmide triangular oberta en què la base és un triangle equilàter i els tres costats laterals són triangles rectangles isòsceles. Amb aquestes noves configuracions de segon nivell podem fer estelacions de deltaedres. Amb vuit piràmides obtenim un octàedre estelat, i amb vint, un icosàedre estelat. Com que cada mòdul forma part de dues piràmides adjuntes, observem que es necessiten dotze mòduls per fer l'octàedre estelat i trenta mòduls per fer l'icosàedre estelat. En general, si tenim un deltaedre de C cares i volem fer un deltaedre estelat, ens calen

$$\frac{3C}{2}$$

mòduls; d'aquest raonament origàmic deduïm que qualsevol deltaedre té sempre un nombre parell de cares.

Observem que, si fem l'estelació del tetràedre amb aquest tipus de configuració de segon nivell, obtenim el cub de Sonobe. Això és així perquè es pot inscriure un tetràedre dins d'un cub; les sis arestes del tetràedre coincideixen amb sis de les dotze diagonals del cub. Aquest fet ja fou observat per Kepler en l'*Harmonices Mundi* (1619).



Kepler. *Harmonices Mundi*, llibre v



Segell amb un octàedre estelat

De la il·lustració del llibre de Kepler deduïm que la longitud de l'aresta del cub l coincideix amb la distància entre dues arestes oposades del tetràedre, i la longitud de l'aresta del tetràedre regular inscrit val $\sqrt{2}l$. A més, la descomposició del cub en cinc tetràedres ens permet calcular el volum del tetràedre regular.

$$l^3 = V_{\text{cub}} = V_{\text{tetràedre regular}} + 4 \cdot V_{\text{tetràedres no regulars}},$$

$$l^3 = V_{\text{tetràedre regular}} + 4 \cdot \frac{1}{6}l^3,$$

$$V_{\text{tetràedre regular}} = \frac{l^3}{3}.$$

Observem que el volum del tetràedre regular és el doble que el volum del tetràedre rectangle.

L'estelació de l'octàedre apareix en un segell que es va emetre al Japó per commemorar la celebració a Kyoto, l'any 1990, del XXI Congrés Internacional de Matemàtiques.

Els deltaedres estelats construïts amb aquest mòdul tenen les cares triangulades. Si invertim el tipus de plecs en el pas 7 (vall en lloc de muntanya) i el pas 8 (muntanya en lloc de vall), aleshores obtindrem els deltaedres estelats amb les cares llises.

Pas 7	Pas 8	Políedre
Plec en muntanya		Cub de Sonobe amb cares triangulades
Plec en vall		Cub de Sonobe amb cares llises
Plec en muntanya	Plec en muntanya	Joia de Takahama amb cares triangulades
Plec en vall	Plec en vall	Joia de Takahama amb cares llises
Plec en muntanya	Plec en vall	Deltaedres estelats amb cares triangulades
Plec en vall	Plec en muntanya	Deltaedres estelats amb cares llises

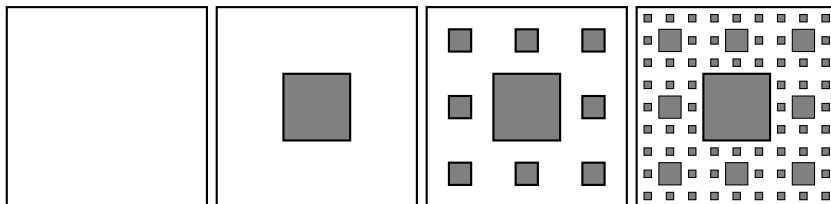
3.3 El cub de Sonobe i els fractals

L'esponja de Menger és un fractal descrit per primer cop pel matemàtic austríac Karl Menger (1902–1985). La descripció d'aquest fractal, que denotarem amb M_∞ , comença amb un cub M_0 de costat L . Dividim aquest cub en vint-i-set cubs iguals de costat $L/3$ i eliminem el cub central i els sis cubs centrals de cada cara. Aquesta nova figura M_1 té vint dels vint-i-set cubs inicials. Ara iterem infinitament el procés per cadascun dels cubs. Observem que en cada iteració cada cub es transforma en vint cubs i, per tant, l'enèsima iteració M_n està formada per la unió connexa de $K_n = 20^n$ cubs de costat $L/3^n$. D'aquesta manera tenim una successió de varietats topològiques. Definim l'esponja de Menger, que denotarem amb M_∞ , com la intersecció de totes aquestes varietats topològiques.

$$M_\infty = \bigcap_{k \in \mathbb{N}} M_k \subsetneq \cdots \subsetneq M_n \subsetneq M_{n-1} \subsetneq \cdots \subsetneq M_1 \subsetneq M_0.$$

De fet, l'esponja de Menger és una generalització a $\mathbb{R}^3$ de la catifa de Sierpinski ($\mathbb{R}^2$)

$$S_\infty = \bigcap_{k \in \mathbb{N}} S_k \subsetneq \cdots \subsetneq S_n \subsetneq S_{n-1} \subsetneq \cdots \subsetneq S_1 \subsetneq S_0$$



i del conjunt o pols de Cantor ($\mathbb{R}^1$),

$$C_\infty = \bigcap_{k \in \mathbb{N}} C_k \subsetneq \cdots \subsetneq C_n \subsetneq C_{n-1} \subsetneq \cdots \subsetneq C_1 \subsetneq C_0$$



Aquests i d'altres fractals van ser estudiats amb profunditat al segle xx; per exemple, les seves dimensions de Hausdorff són, respectivament:

$$\log 20 / \log 3 \approx 2.727, \quad \log 8 / \log 3 \approx 1.8927, \quad \log 2 / \log 3 \approx 0.631.$$

El mòdul de Sonobe ens permet construir i entendre algunes propietats d'aquesta successió de superfícies. El cub inicial M_0 es pot fer amb $C_0 = 6$ mòduls de Sonobe; aquest cub té $V_0 = 8$ vèrtexs i $A_0 = 12$ arestes, la seva característica d'Euler-Poincaré val $\chi_0 = C_0 - V_0 + A_0 = 6 - 8 + 12 = 2$ i el seu gènere és $g_0 = (2 - \chi_0)/2 = 0$.

A continuació veurem quants mòduls de Sonobe són necessaris per construir la iteració enèsima de l'esponja de Menger M_n .

Inicialment tenim $K_0 = 1$ cub de costat L . En fer la primera iteració de l'esponja de Menger dividim cada aresta en tres parts iguals, així cada aresta aportarà dos nous vèrtexs.

En la primera iteració de l'esponja de Menger observem que cada cara quadrada de costat L del cub inicial es transforma en vuit noves cares quadrades de longitud $L/3$; a aquestes noves cares cal afegir quatre cares centrals interiors per cada cara del cub inicial. Així doncs, la primera iteració de l'esponja de Menger M_1 es pot fer amb $C_1 = 24 + 8C_0 = 72$ mòduls de Sonobe. El nombre d'arestes és el doble del nombre de cares $A_1 = 4C_1/2 = 144$. Per calcular el nombre de vèrtexs V_1 observem que cada aresta A_0 es divideix en tres parts iguals i, per tant, cada aresta aportarà dos vèrtexs en la iteració, als quals cal afegir els vèrtexs V_0 del cub inicial, així com els nous vèrtexs que apareixen a la part central del cub, quatre per cada cara inicial i vuit més al centre del cub $V_1 = V_0 + 2A_0 + 4C_0 + 8 = 64$; així doncs, $\chi_1 = C_1 - A_1 + V_1 = 72 - 144 + 64 = -8$ i, per tant, el seu gènere és $g_1 = (2 - \chi)/2 = 5$.

De fet, aquest argument és vàlid per a qualsevol iteració i ens proporciona fórmules recursives per calcular els valors del nombre de cares, arestes, vèrtexs, característica d'Euler i gènere de la superfície de cada iteració M_n .

Nombre de cares	$C_n = 8C_{n-1} + 24K_{n-1}$
Nombre d'arestes	$A_n = 2C_n$
Nombre de vèrtexs	$V_n = V_{n-1} + 2A_{n-1} + 2^2C_{n-1} + 2^3K_{n-1} + 8(20^{n-1} - 8^{n-1})$
Característica d'Euler	$\chi_n = V_n - A_n + C_n$
Gènere de la superfície	$g_n = (2 - \chi_n)/2$

Utilitzant les recurrències anteriors, i avaluant polinomis ciclotòmics en dues variables, es podem demostrar per inducció les següents fórmules que donen els valors exactes en funció de n :

Nombre de cares	$C_n = 4 \cdot 8^n + 2 \cdot 20^n$
Nombre d'arestes	$A_n = 8 \cdot 8^n + 4 \cdot 20^n$
Nombre de vèrtexs	$V_n = 8 + 24 \cdot \frac{8^n - 1}{7} + 32 \cdot \frac{20^n - 1}{19}$
Característica d'Euler	$\chi_n = 8 + 24 \cdot \frac{8^n - 1}{7} + 32 \cdot \frac{20^n - 1}{19} - 4 \cdot 8^n - 2 \cdot 20^n$
Gènere de la superfície	$g_n = 2 \cdot \frac{8^n - 1}{7} + 3 \cdot \frac{20^n - 1}{19}$

Per tant, els valors del nombre de mòduls de Sonobe necessaris per construir l'enèsima iteració de l'esponja de Menger i el gènere de la superfície venen donats per les successions

$$\{C_n\}_{n \in \mathbb{N}} = \{6, 72, 1056, 18048, 336384, 6531072, \dots, \infty\},$$

$$\{g_n\}_{n \in \mathbb{N}} = \{0, 5, 81, 1409, 26433, 514625, \dots, \infty\}.$$

Prenent límit quan n tendeix a ∞ , deduïm que l'esponja de Menger M_∞ té una superfície d'àrea infinita amb volum 0.

$$\text{Àrea}(M_\infty) = \lim_{n \rightarrow \infty} \frac{C_n}{3^{2n}} = \lim_{n \rightarrow \infty} 2 \cdot \left(\frac{20}{9}\right)^n + 4 \cdot \left(\frac{8}{9}\right)^n = \infty,$$

$$\text{Volum}(M_\infty) = \lim_{n \rightarrow \infty} \frac{K_n}{3^{3n}} = \lim_{n \rightarrow \infty} \left(\frac{20}{27}\right)^n = 0.$$

Aplicant el teorema de Ringel-Youngs [11], veiem que el nombre mínim de colors (nombre cromàtic) necessaris per acolorir la superfície donada per l'enèsima iteració de l'esponja de Menger sense que dues cares adjacents tinguin el

mateix color ve donat per la fórmula següent, conjecturada per Heawood el 1891, que tan sols depèn del gènere de la superfície

$$H(g_n) = \left\lceil \frac{7 + \sqrt{1 + 48g_n}}{2} \right\rceil = \left\lceil \frac{7 + \sqrt{1 + 48 \left(2 \cdot \frac{8^n - 1}{7} + 3 \cdot \frac{20^n - 1}{19} \right)}}{2} \right\rceil.$$

Els primers valors d'aquesta successió de nombres cromàtics són:

$$\{H(g_n)\}_{n \in \mathbb{N}} = \{4, 11, 34, 133, 566, 2488, 11056, 49323, 220373, 985176, \dots, \infty\}.$$

3.4 Cubs de paper i cubs aritmètics

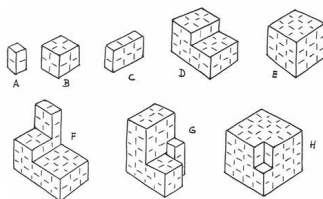
Juan Manuel Gimeno Viguera, erudit de la papiroflèxia, va proposar l'any 1992 un trencaclosques geomètric-papiroflèctic inspirat en la igualtat

$$6^3 = 5^3 + 4^3 + 3^3.$$

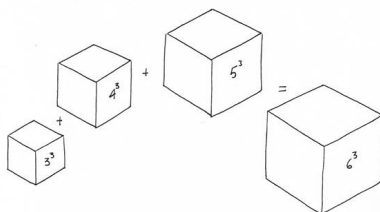
Aquest joc consistia en la construcció de vuit peces fetes amb el mòdul de Sonobe-Takahama, que calia reordenar per aconseguir els quatre cubs.

Los cubos disjuntivos

- Usando los módulos de 90° de Takahama, construir las siguientes piezas:



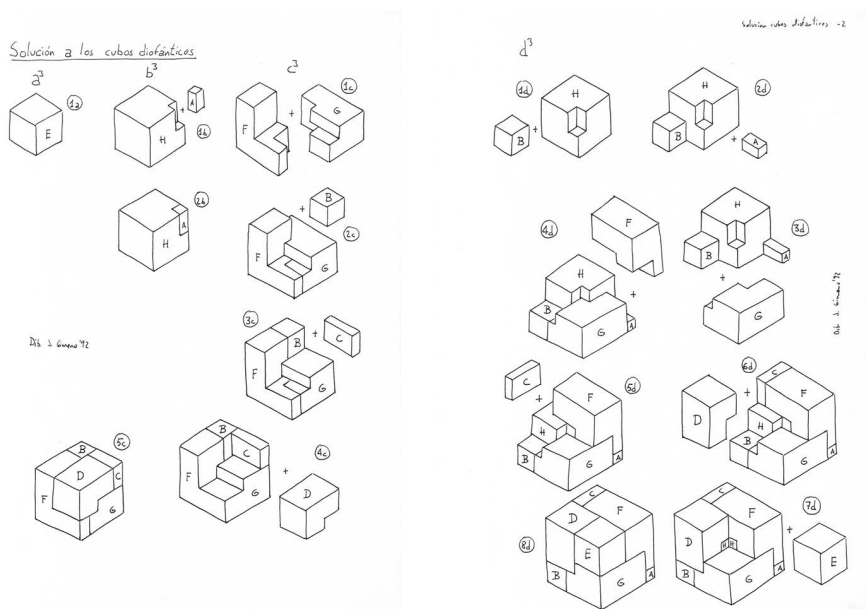
- Hacer con todas las piezas tres cubos (3^3 , 4^3 , 5^3) de tal manera que todas juntas formen otro cubo (6^3) que demuestren que $3^3 + 4^3 + 5^3 = 6^3$



Una segona igualtat amb cubs ben coneguda és l'exemple de Srinivasa Ramanujan, que expressa el nombre 1729 de dues formes diferents com a la suma de dos cubs

$$1729 = 1^3 + 12^3 = 9^3 + 10^3.$$

Aquesta igualtat i el Taxi Cab que la va inspirar apareixen en la pel·lícula biogràfica *L'home que coneixia l'infinit*, 2015.



Però com es poden deduir aquestes o altres igualtats amb nombres naturals? Observem que les igualtats numèriques anteriors són solucions enteres de la superfície cúbica

$$F : x^3 + y^3 + z^3 + w^3 = 0. \quad (5)$$

L'estudi profund de les superfícies cúbiques comença l'any 1849, quan Arthur Cayley demostra que tota superfície cúbica llisa sobre el cos $\mathbb{C}$ té un nombre finit de rectes i el mateix any George Salmon precisa que el nombre de rectes és exactament 27. Cal destacar també els treballs d'Alfred Clebsch, que el 1871 descobreix una superfície cúbica que conté 27 rectes reals. Aquesta superfície cúbica es pot expressar amb l'equació

$$(x + y + z + w)^3 = x^3 + y^3 + z^3 + w^3.$$

Observem que les 27 rectes de la superfície (5) venen donades per les equacions

$$\begin{aligned} L_{i,j} &= \{x + \xi^i y = 0, z + \xi^j w = 0\}, \\ L'_{i,j} &= \{x + \xi^i z = 0, y + \xi^j w = 0\}, \\ L''_{i,j} &= \{x + \xi^i w = 0, y + \xi^j z = 0\}, \end{aligned} \quad (6)$$

on ξ és una arrel primitiva cúbica de la unitat i els índexs $\{i, j\}$ prenen els valors $\{0, 1, 2\}$. Tres d'aquestes rectes ($i = j = 0$) tenen coeficients en el cos dels nombres racionals $\mathbb{Q}$ i les altres 24 rectes ($i \neq 0$ o $j \neq 0$) tenen coeficients en el cos $\mathbb{Q}(\sqrt{-3})$. Es poden obtenir parametritzacions de la superfície (5) triant dues rectes que no es tallin d'aquesta col·lecció de 27 rectes (6).

Noam Elkies i Irene Polo-Blanco van trobar, respectivament, les parametritzacions següents:

$$\begin{aligned} \Phi: \quad \mathbb{P}^2 &\longrightarrow F \\ (r:s:t) &\longrightarrow \begin{aligned} &-(s+r)t^2 + (s^2 + 2r^2)t - s^3 + rs^2 - 2r^2s - r^3 : \\ &t^3 - (s+r)t^2 + (s^2 + 2r^2)t + rs^2 - 2r^2s + r^3 : \\ &-t^3 + (s+r)t^2 - (s^2 + 2r^2)t + 2rs^2 - r^2s + 2r^3 : \\ &(s-2r)t^2 + (r^2 - s^2)t + s^3 - rs^2 + 2r^2s - 2r^3, \end{aligned} \end{aligned} \quad (7)$$

$$\begin{aligned} \Psi: \quad \mathbb{P}^2 &\longrightarrow F \\ (r:s:t) &\longrightarrow \begin{aligned} &(-r^3 - 2r^2t + 3r^2s + 12rst - 3rs^2 - 4rt^2 + 6s^2t + 12st^2 + 9s^3 : \\ &r^3 + 2r^2t + 3r^2s + 12rst + 3rs^2 + 4rt^2 - 6s^2t + 12st^2 + 9s^3 : \\ &-8t^3 - 8rt^2 - 9s^3 - r^3 - 3r^2s - 3rs^2 - 4r^2t - 12s^2t : \\ &8t^3 + 8rt^2 - 9s^3 + r^3 - 3r^2s + 3rs^2 + 4r^2t + 12s^2t). \end{aligned} \end{aligned}$$

De fet, existeixen transformacions lineals f, g que fan que aquestes dues parametritzacions siguin equivalents:

$$\begin{array}{ccc} \mathbb{P}^2 & \xrightarrow{\Phi} & F \\ f \downarrow & & \uparrow g \\ \mathbb{P}^2 & \xrightarrow{\Psi} & F \end{array} \quad g \circ \Psi \circ f = \Phi,$$

on

$$f: \quad \mathbb{P}^2 \longrightarrow \mathbb{P}^2 \quad (r:s:t) \longrightarrow (2s-r:-r:t-s) \quad \text{i} \quad g: \quad F \longrightarrow F \quad (w:x:y:z) \longrightarrow (w:z:y:x).$$

Donant valors als paràmetres $(r:s:t)$ en (7), podem obtenir noves relacions numèriques entre cubs de nombres naturals:

$$\begin{aligned} r=1 \quad s=1 \quad t=1: & \quad 6^3 = 5^3 + 4^3 + 3^3, \\ r=1 \quad s=-1 \quad t=1: & \quad 9^3 = 8^3 + 6^3 + 1^3, \\ r=1 \quad s=2 \quad t=3: & \quad 19^3 = 18^3 + 10^3 + 3^3, \\ r=2 \quad s=1 \quad t=2: & \quad 16^3 + 2^3 = 15^3 + 9^3, \\ r=3 \quad s=1 \quad t=3: & \quad 60^3 + 3^3 = 59^3 + 22^3, \\ r=3 \quad s=2 \quad t=3: & \quad 51^3 + 12^3 = 43^3 + 38^3. \end{aligned}$$

L'exemple de Ramanujan es pot obtenir amb els següents valors dels paràmetres

$$r = -\frac{2}{\sqrt[3]{21}} \quad s = \frac{1}{\sqrt[3]{21}} \quad t = -\frac{6}{\sqrt[3]{21}}: \quad 1^3 + 12^3 = 9^3 + 10^3.$$

En el jardí matemàtic, ubicat entre els edificis Torretamarit i Torrepinet del campus de la Universitat Miguel Hernández d'Elx, podeu veure una escultura de 3 metres d'altura feta amb pedra natural de la superfície cúbica $w(x+y+z)^2 + xyz = 0$. Aquesta superfície té quinze rectes i un únic punt

singular $(0 : 0 : 0 : 1)$. L'escultura fou realitzada l'any 2009 per l'escultor murcià Cayetano Ramírez López i és coneguda popularment com a *Damatemàtica* per la seva semblança amb la Dama d'Elx. Aquest escultor va restaurar l'any 2005 la col·lecció de superfícies cúbiques de la Universitat de Groningen; en particular, va fer amb escaiola i amb resina la superfície cúbica de Cayley definida per l'equació $xyz + yzw + zwx + wxy = 0$. La cúbica de Cayley té nou rectes i quatre punts singulars, que és el màxim nombre de punts singulars que pot tenir una superfície cúbica. També va fer uns models en polièster de la superfície de Clebsch, on es poden veure clarament les 27 rectes reals, i d'altres superfícies cúbiques que van ser exposades en el XXV Congrés Internacional de Matemàtiques, celebrat a Madrid l'any 2006, el mateix congrés en què el matemàtic rus Grigori Perelman va rebutjar la medalla Fields. Posteriorment es va fer una exposició itinerant sota el títol *Art i matemàtiques* amb aquestes escultures sobre superfícies cúbiques i altres escultures, com una ampolla de Klein, un nàutil basat en la proporció àuria, alguns cargols amb formes espirals i un romanesc per il·lustrar el concepte de fractal.



Damatemàtica



Cúbica de Clebsch

Dues escultures de Cayetano Ramírez

Actualment les escultures de les superfícies cúbiques de Cayetano Ramírez es troben en el Museu Didàctic i Interactiu de Ciències del Baix Segura (MUDIC, <http://www.mudic.es>) ubicat a l'Escola Politècnica Superior d'Oriola de la Universitat Miguel Hernández (Elx) i dirigit per la matemàtica Carmen Perea.

Fins ara hem vist que existeixen infinits cubs naturals que es poden descompondre com a suma de tres cubs naturals no nuls. Aquesta afirmació s'il·lustrava amb un trencaclosques que utilitzava el mòdul de Sonobe. En canvi, no podem fer un trencaclosques semblant amb aquest mòdul per descompondre un cub com a suma de dos cubs. En efecte, com va dir Pierre de Fermat el 1635:

Cubum autem in duos cubos, aut quadrato-quadratum in duos quadrato-quadratos, et generaliter nullam in infinitum ultra quadratum potestatem in duos eiusdem nominis fas est dividere cuius rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.

Aquesta és possiblement la frase més famosa de les matemàtiques i actualment es coneix com a *darrer teorema de Fermat*.

Durant 359 anys els matemàtics van intentar demostrar sense èxit aquesta afirmació que Pierre de Fermat (Bèumont de Lomanha, 1601 - Castres, 1665) va fer el 1635. El cas $n = 4$ es pot demostrar amb la tècnica del descens infinit; així doncs, per demostrar el darrer teorema de Fermat és suficient estudiar el cas $n = p$ nombre primer i, a més, podem suposar que x, y, z són nombres naturals, no nuls i dos a dos coprims entre si.

$$\forall x, y, z \in \mathbb{N} - \{0\}, \quad x^p + y^p \neq z^p, \quad p \geq 3.$$

Dels nombrosos articles dedicats a aquest teorema, cal destacar el treball de la matemàtica francesa Sophie Germain, que en una carta dirigida a Gauss i amb data 12 de maig de 1819 intenta fer una demostració general per a tots els exponents del darrer teorema de Fermat. Les tècniques de Sophie Germain van permetre demostrar el teorema per a tots els nombres primers $p < 100$ i dividir el darrer teorema de Fermat en dos casos: $p \nmid xyz$ (cas 1) i $p \mid xyz$ (cas 2).

A principis del segle xx els treballs de Dmitri Mirimànov (1905) i d'Arthur Wieferich (1909) redueixen el cas 1 a trobar les solucions de la congruència:

$$2^p \equiv 2 \pmod{p^2}.$$

De fet, tan sols es coneixen dues solucions $p = 1093$ (Meisner, 1913) i $p = 3511$ (Beeger, 1921) de la congruència de Wieferich per a $p < 4 \cdot 10^{12}$. Per tant, el cas 1 del darrer teorema de Fermat és cert per a gairebé tot primer $p < 4 \cdot 10^{12}$.

Però la demostració definitiva del darrer teorema de Fermat comença l'any 1986 amb el treball de Gerhard Frey, que associa a una possible solució $a^p + b^p = c^p$ amb nombres enters no nuls de l'equació de Fermat, la corba el·líptica

$$y^2 = x(x - a^p)(x + b^p).$$

Estudiant les seves propietats, Frey demostra que aquesta corba el·líptica no pot ser modular, en contra de la conjectura de Shimura-Taniyama-Weil: tota corba el·líptica és modular. El cercle es va tancar el dia de la revetlla de Sant Joan de l'any 1993 en el Sir Isaac Newton Institute de la Cambridge University, quan Andrew Wiles va finalitzar tres conferències en dies consecutius sota el títol *Modular forms, elliptic curves and Galois representations* afirmant haver demostrat la conjectura de Shimura-Taniyama-Weil i, com a corollari, el darrer teorema de Fermat. Al cap de poc temps els experts van trobar un forat en la demostració, que va ser resolta pel mateix Wiles i el seu deixeble Richard Taylor. Finalment, la demostració rigorosa fou publicada pels dos autors el maig de 1995 en els *Annals of Mathematics*. El treball de Wiles es redueix a estudiar els punts de 3-torsió i de 5-torsió d'una corba el·líptica. Posteriorment, Noam Elkies va fer una simplificació de la demostració i va provar que pel darrer teorema de Fermat no calia estudiar els punts de 5-torsió.

A continuació veurem una demostració elemental del darrer teorema de Fermat per al cas 1 de l'exponent $p = 3$.

PROPOSICIÓ 9. *Siguin $x, y, z \in \mathbb{N}$ no nuls tals que $x^3 + y^3 = z^3$. Aleshores, $3 \mid xyz$.*

PROVA. Siguin x, y, z tres nombres naturals, no nuls, 2 a 2 coprimers i no divisibles per 3 que compleixen l'equació de Fermat per a $p = 3$.

$$x^3 + y^3 = z^3, \quad x, y, z \in \mathbb{N} - \{0\}, \quad 3 \nmid xyz. \quad (8)$$

La suma de dos cubs $x^3 + y^3$ es pot descompondre com el producte $(x + y)(x^2 - xy + y^2)$ de dos factors. Aquests factors no són divisibles per 3 perquè el seu producte val z^3 , que no és divisible per 3. A més, també són coprimers, en efecte, $\text{mcd}(x + y, x^2 - xy + y^2) = \text{mcd}(x + y, 3xy) = 1$; per tant, es poden expressar com a potències cúbiques. Un raonament anàleg es pot aplicar per a les diferències de dos cubs $z^3 - x^3$ i $z^3 - y^3$.

$$\begin{aligned} x + y &= z_1^3, & x^2 - xy + y^2 &= z_2^3, & z_1 z_2 &= z, \\ z - x &= y_1^3, & z^2 + zx + x^2 &= y_2^3, & y_1 y_2 &= y, \\ z - y &= x_1^3, & z^2 + zy + y^2 &= x_2^3, & x_1 x_2 &= x. \end{aligned}$$

Ara tenim sis nous valors no nuls $(x_1, y_1, z_1, x_2, y_2, z_2)$ que són 2 a 2 coprimers i no divisibles per 3. De fet, els valors (x_2, y_2, z_2) no intervenen en la resta de la demostració.

De les tres primeres equacions lineals podem aïllar els valors x, y, z en funció de x_1, y_1, z_1

$$x = \frac{z_1^3 - y_1^3 + x_1^3}{2}, \quad y = \frac{z_1^3 - x_1^3 + y_1^3}{2}, \quad z = \frac{z_1^3 + x_1^3 + y_1^3}{2}.$$

Substituïm aquestes combinacions lineals de cubs en l'expressió (8) i multipliquem per 8,

$$(z_1^3 - y_1^3 + x_1^3)^3 + (z_1^3 - x_1^3 + y_1^3)^3 = (z_1^3 + x_1^3 + y_1^3)^3.$$

Desenvolupant els cubs obtenim:

$$(z_1^3 - x_1^3 - y_1^3)^3 = 24x_1^3 y_1^3 z_1^3.$$

D'aquesta darrera igualtat es dedueix la contradicció que 3 és un cub perfecte al conjunt dels nombres racionals:

$$3 = \left(\frac{z_1^3 - x_1^3 - y_1^3}{2x_1 y_1 z_1} \right)^3 \in \mathbb{Q}.$$

□

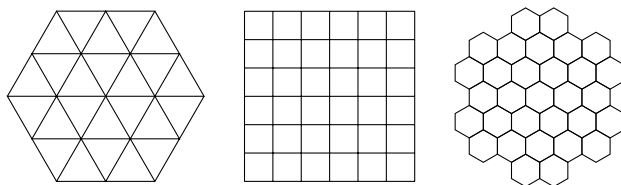
A la localitat occitana de Bèumont de Lomanha podeu visitar la casa natal de Fermat, avui convertida en museu, així com un monument de l'illustre matemàtic al costat d'un preciós mercat cobert de fusta del segle xv. Aquesta petita població està situada a uns 60 km al nord-oest de Tolosa de Llenguadoc, on Fermat treballava com a advocat.

4 Mosaics de paper

Els diferents motius dels mosaics de l'Alhambra de Granada són des de fa temps una font d'inspiració per a la creació artística i matemàtica. En aquest palau de la dinastia nassarita, construït entre els segles XIII i XV, es poden contemplar diferents motius geomètrics. Alguns d'aquests motius es coneixen amb un nom propi: ocellet nassarita, os nassarita, avió nassarita o ratpenat nassarita.

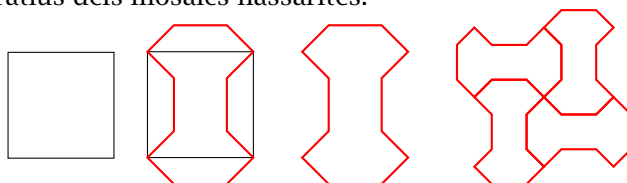
Però què és un mosaic? La seva etimologia ens diu que aquesta paraula prové del llatí *mosaicum opus*, és a dir, una obra de les muses. Antigues civilitzacions com la grega o la romana ja feien aquestes composicions decoratives amb petites peces de diferents colors i materials anomenades *tessel·les*. Molts d'aquests antics mosaics i altres de més moderns, com els que es poden veure en les obres modernistes d'Antoni Gaudí, tot i la seva bellesa, no tenen cap tipus de simetria. També existeixen mosaics que mostren simetries, com els mosaics de les voreres del passeig de Gràcia de Barcelona, fets amb rajoles hexagonals, o els mosaics del pati d'entrada de l'Institut d'Estudis Catalans, elaborats amb rajoles catalanes. Des d'un punt de vista matemàtic, definirem un mosaic o tessellació com una divisió del pla sense encavalcaments ni buits mitjançant una o més formes geomètriques, també anomenades *tessel·les*. Per estudiar els mosaics respondrem les preguntes següents: Quins són els mosaics més simples? Com podem crear nous mosaics? Quants mosaics essencialment diferents existeixen? Com podem classificar els mosaics?

Els mosaics més simples, que anomenarem *regulars*, s'aconsegueixen prenent com a tessella bàsica tres tipus de polígons regulars: el triangle equilàter, el quadrat i l'hexàgon regular. Els mosaics regulars tenen simetries centrals o rotacionals de 60, 90 i 120 graus, respectivament, així com diferents tipus de simetries axials, en què els eixos de simetria formen entre si angles de 30, 45 i 60 graus, respectivament.



Mosaics regulars

Es poden fer nous mosaics modificant els mosaics regulars. Aquesta observació va ser desenvolupada en les obres de l'artista neerlandès Maurits Cornelius Escher en quedar impressionat en la seva visita a l'Alhambra, l'any 1936, pels motius decoratius dels mosaics nassarites.



Quadrat

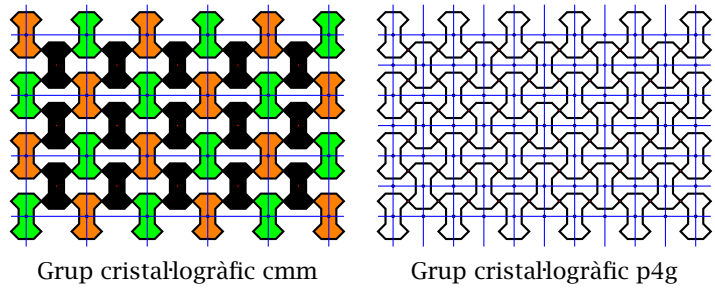
Os nassarita

La tessella os nassarita es pot aconseguir a partir d'un quadrat. En dos costats oposats d'un quadrat seccionem un trapezi que tingui per base major el costat del quadrat, per alçària la quarta part de la longitud del costat del quadrat i angles interns de 45, 135, 135 i 45 graus. Unim per la part exterior del quadrat la base major d'aquests trapezis als altres dos costats del quadrat, com es mostra en la figura anterior.

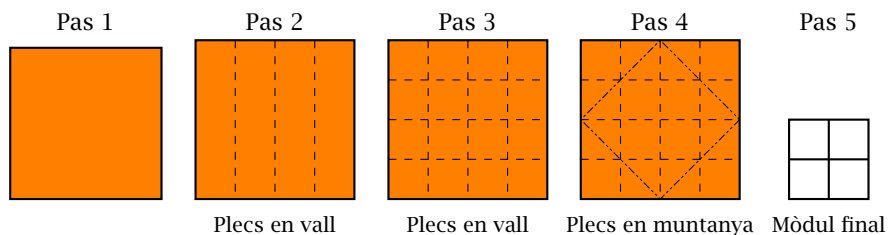
El teorema de Fedorov (1891) ens diu que tan sols existeixen disset tipus de mosaics essencialment diferents [5]. Els diferents tipus de mosaics venen descrits per disset grups de simetries. Aquests grups també es coneixen amb el nom de *grups cristal·logràfics plans*, perquè Fedorov els va trobar estudiant la forma en què cristal·litzen a la natura els minerals. Existeixen diferents notacions equivalents per classificar els grups cristal·logràfics plans: Pólya, Conway, Fejes Thót-Cadwell, Niggli, Speiser, Shubinov-Koptsik, Wells-Bell, internacional llarga, etc. En la taula següent s'utilitza la notació internacional curta.

Angle mínim de rotació	Existeix alguna reflexió?				
	Sí		No		
$360^\circ/6=60^\circ$	p6m		p6		
$360^\circ/4=90^0$	Tots els centres de rotació estan sobre els eixos de reflexió?		p4		
	Sí: p4m	No: p4g			
$360^\circ/3=120^\circ$	Tots els centres de rotació estan sobre els eixos de reflexió?		p3		
	Sí: p31m	No: p3m1			
$360^\circ/2=180^\circ$	Existeixen reflexions amb eixos perpendiculars?		Existeix alguna reflexió amb lliscament?		
	Sí		No	Sí: pgg	No: p2
	Tots els centres de rotació estan sobre els eixos de reflexió?		pmg		
	Sí: cmm	No: pmm			
$360^\circ/1=0^\circ$	Existeix alguna reflexió amb lliscament amb eix diferent d'un eix de reflexió?		Existeix alguna reflexió amb lliscament?		
	Sí: cm	No: pm	Sí: pg	No: p1	

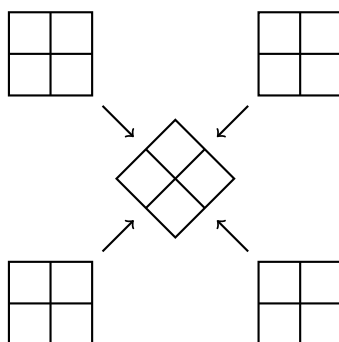
A l'Alhambra de Granada hi ha catorze dels disset tipus de mosaics si es tenen en compte els colors [9] i tots disset si no es tenen en compte els colors [8].



A continuació es descriu un mòdul de paper que permet fer mosaics. Aquest mòdul i les seves variants es coneixen amb el nom de *variacions Froebel*, pel fet que va ser utilitzat al segle XIX passat pel pedagog alemany Frederic Fröbel en les escoles Kindergarten (jardins d'infància).

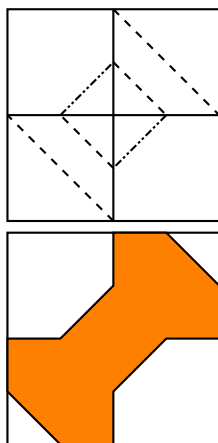


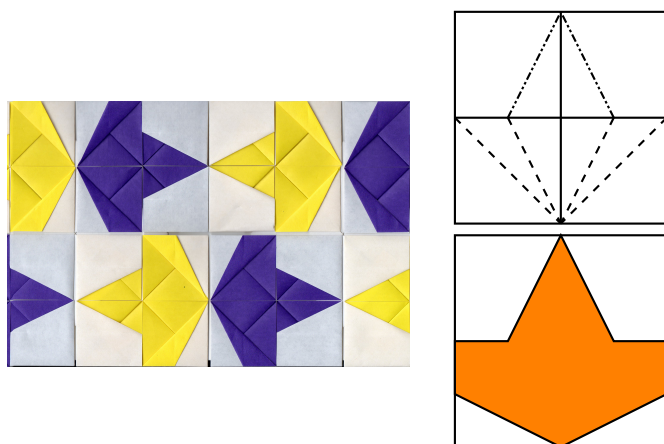
Per aconseguir diferents tipus de motius és necessari utilitzar un paper bi-color, com per exemple el paper de tipus kami (blanc/color). En el pas 1 comencem amb la cara de color al davant i la de color blanc al darrere; en el pas 2 pleguem verticalment en vall el paper per la meitat i per la quarta part; en el pas 3 fem el mateix que en el pas 2 però horitzontalment; en el pas 4 pleguem en muntanya els quatre vèrtexs del quadrat cap al centre, i en el pas 5 obtenim el mòdul final col·lapsant els quatre costats del quadrat cap a l'interior.



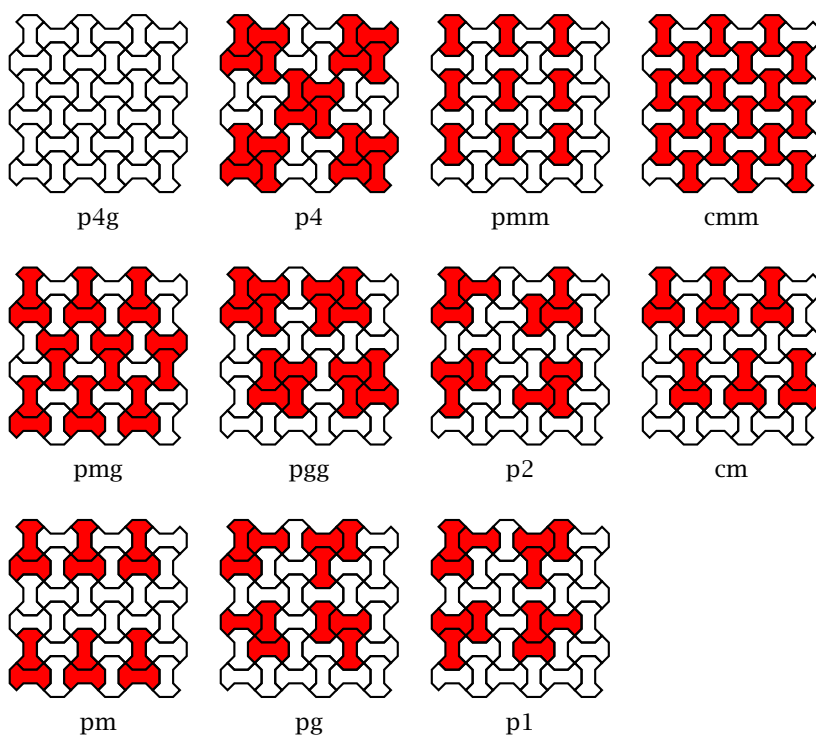
Unió de cinc mòduls

Per aconseguir els motius os nassarita i avió nassarita cal fer els següents plecs en el mòdul final.



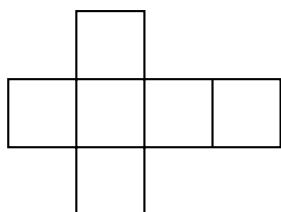


Combinant els colors en els mosaics fets amb la tessella de l'os nassarita es poden aconseguir onze de les disset configuracions dels grups cristal·logràfics.

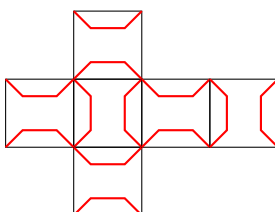


És clar que el desenvolupament pla del cub té sis quadrats. Vist a l'inrevés es pot pensar que el cub és un mosaic tridimensional format per sis tesselles quadrades. D'aquesta manera podem construir diferents tipus de mosaics

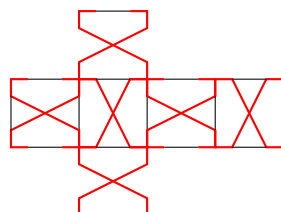
cúbics fent modificacions en el quadrat llis inicial. Per exemple, si considerem sis tesselles de l'os nassarita o de l'avió nassarita, es pot fer un mosaic cúbic tridimensional amb cadascun d'aquests motius.



Desenvolupament del cub

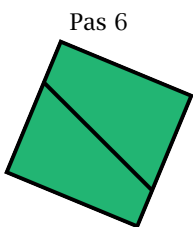
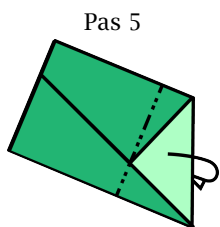
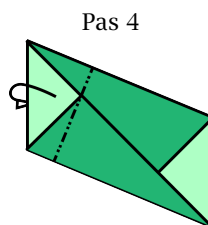
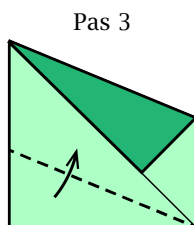
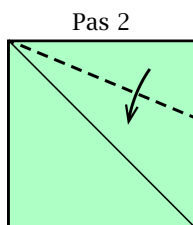
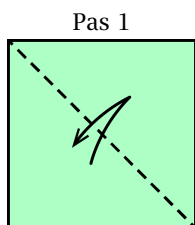


Desenvolupament del cub os nassarita

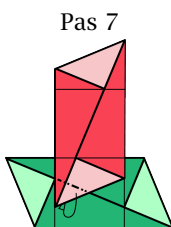


Desenvolupament del cub avió nassarita

Es poden construir cubs amb els tres motius anteriors amb sis mòduls idèntics fent determinats plecs. El primer motiu, un quadrat llis, es correspon al cub de Sonobe. El cub avió nassarita es pot fer amb els plecs següents:



Mòdul final

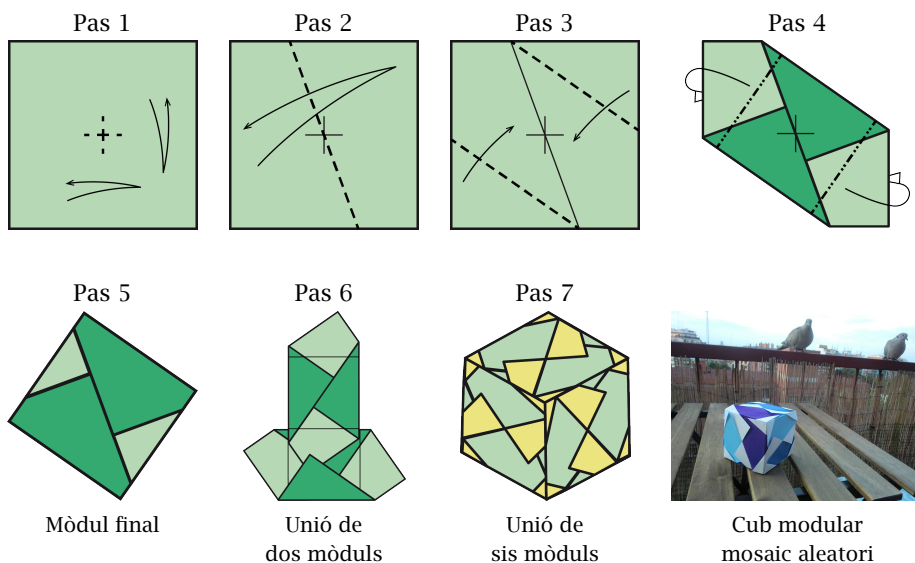


Unió de dos mòduls

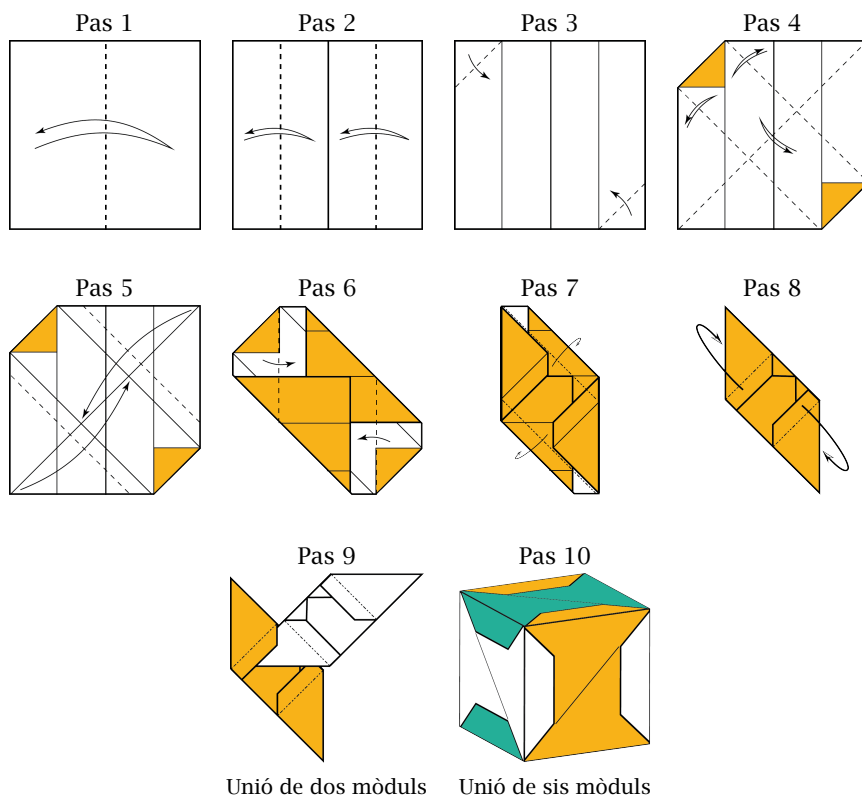


Unió de sis mòduls

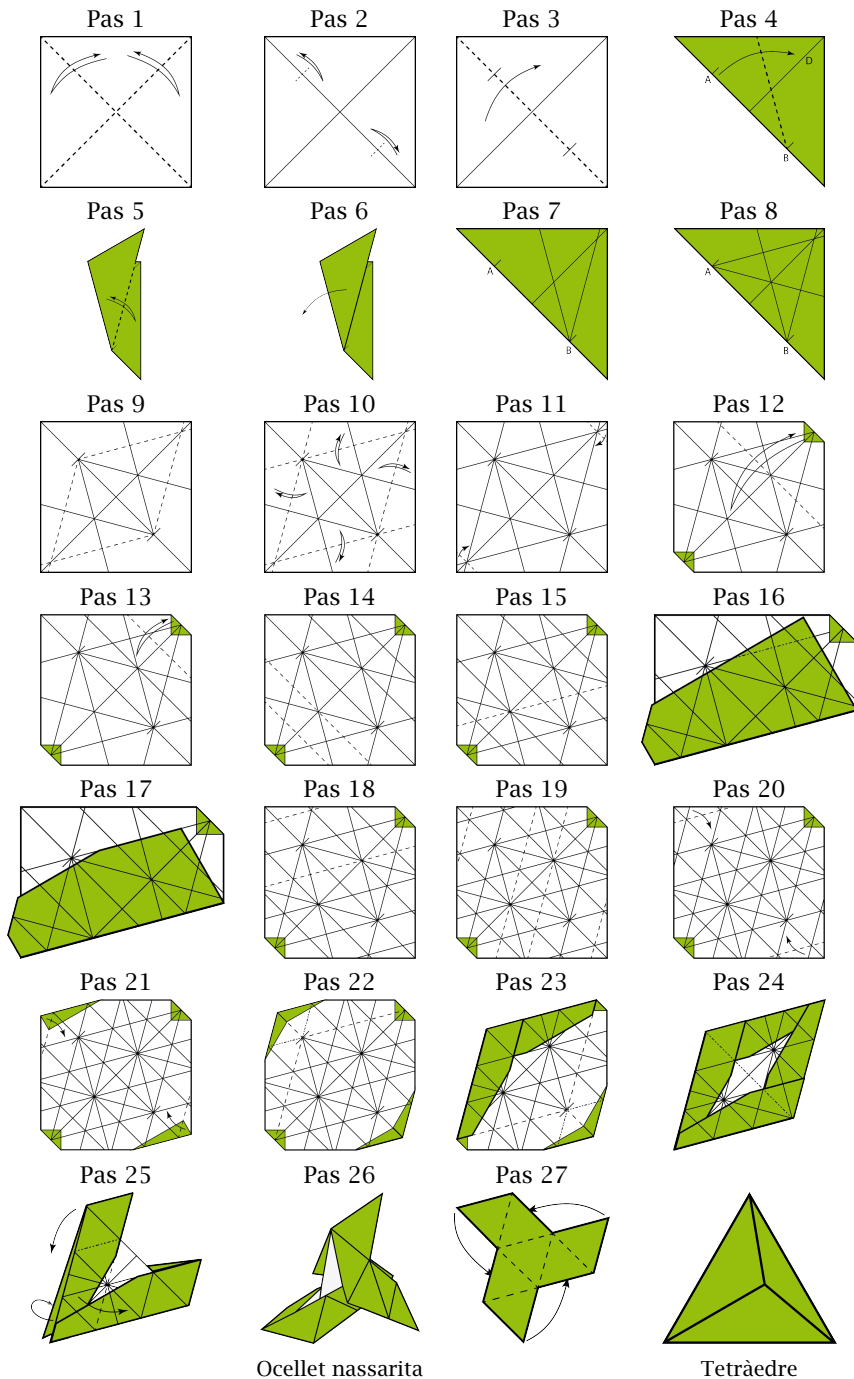
Si a la construcció anterior, en lloc de la diagonal, prenem un segment que passi pel centre del quadrat amb pendent aleatori $m \in \mathbb{R}$, aleshores obtenim una infinitud no numerable de motius.



La tessellació del cub amb el motiu os nassarita es pot fer amb els plects següents:



Finalitzem aquesta dissertació amb una figura no modular multiforme: la construcció d'un tetràedre obtingut mitjançant el col·lapse d'un ocellet nassarita tridimensional.



Agraïments

A Minerva Ciruela per introduir-me en el món de la papiroflèxia; a Juan Gimeno per les seves converses sobre el tema; a Joan Sallas per convidar-me a experimentar i desenvolupar aquestes idees en les jornades FALTIK a Friburg de Brisgòvia i en les jornades APLEC a Badalona; a Carmen Sprung per deixar-me publicar els seus diagrames, i a David Brill per facilitar-me la imatge d'una sessió del Sosaku Origami Group 67 en què apareixen Toshie Takahama i Mitsunobu Sonobe.

Referències

- [1] ANÒNIM. *Una hoja de papel*. 2a ed. Barcelona: Miguel A. Salvatella, 1964.
- [2] CROMWELL, P. R. *Polyhedra*. Cambridge: Cambridge University Press, 1997.
- [3] DIFERENTS AUTORS. *Guía práctica del arte manual*. Barcelona: Kairós, 1977.
- [4] FREUDENTHAL, H.; VAN DER WAERDEN, B. L. «Over een bewering van Euclides». *Simon Stevin*, 25 (1947), 115-128.
- [5] FYODOROV, Y. «Simmetrija na ploskosti». *Zapiski Imperatorskogo Sant-Petersburgskogo Mineralogicheskogo Obschestva*, s. 2, 28 (1891), 245-291.
- [6] KASAHARA, K.; TAKAHAMA, T. *Papiroflexia para expertos*. Madrid: EDAF, 2011.
- [7] MILLER J. C. P. *Polyhedron*. 11a ed. Encyclopædia Britannica. Nova York: University Press, 1910.
- [8] MONTESINOS, J. M. *Classical Tessellations and Three-Manifolds*. Berlín: Springer-Verlag, 1987. (Universitext)
- [9] MÜLLER, E. «El estudio de ornamentos como aplicación de la teoría de grupos de orden finito». *Euclides (Madrid)*, 6 (1946), 42-52.
- [10] PLATÓ. *Timeu Crítias*. Barcelona: Edicions 62, 2010.
- [11] RINGEL, G.; YOUNGS, J. W. T. «Solution of the Heawood map-coloring problem». *Proc. Nat. Acad. Sci. U.S.A.*, 60 (1968), 438-445.
- [12] SOLÓRZANO, V. *Papirolas. Tratado de papiroflexia. Figuras geoméricas de papel doblado*. Buenos Aires: Librería de la Salud, 1939.
- [13] SOLÓRZANO, V. *Tratado de papiroflexia superior*. Buenos Aires, 1945.
- [14] SOMMERVILLE, D. M. Y. «Semi-regular networks of the plane in absolute geometry». *Transactions of the Royal Society of Edinburgh*, 41, part III, 29, 725-747.
- [15] SOSAKU ORIGAMI GROUP '67. *Atarashii Origami Nyumon*. Tòquio: Nihon Bungeisha, 1970.
- [16] VON STAUDT, K. *Geometrie der lage*. Nuremberg: Verlag von Bauer und Raspe, 1847.
- [17] TAKAHAMA, T. *Creative Life with Creative Origami 1*. 1974. [En japonès]

- [18] DE UNAMUNO, M. *Amor y pedagogía. Apuntes para un tratado de cocotología*. Barcelona: Espasa Calpe, 1964. (Austral; 141)
- [19] VER EECKE, P. *Pappus d'Alexandrie. La collection mathématique*. París: Blanchard, 1982.

CENTRE DE FORMACIÓ DE PERSONES ADULTES RAFAEL FARRÉ (MOLINS DE REI)
DEPARTAMENT DE MATEMÀTIQUES
UNIVERSITAT AUTÒNOMA DE BARCELONA
jaume.coll.guerrero@uab.cat

Punts racionals en corbes el·líptiques

XAVIER GUITART I MARC MASDEU

Resum: Les corbes el·líptiques són un dels objectes més estudiats en teoria de nombres actualment. Es descriuen per equacions cúbiques i allò que les distingeix de la resta de corbes—i les fa tan fascinants—és l'estructura algebraica tan rica que presenta el conjunt de les seves solucions. Aquest article té com a objectiu explicar què són les corbes el·líptiques i explorar algunes de les seves propietats més importants, que sovint es troben entre els resultats més rellevants de les matemàtiques dels segles XX i XXI, així com algunes de les conjectures encara obertes i que guien gran part de la recerca actual. Per situar-les en un context històric i conceptual més ampli, les presentem com un cas particular d'equacions diofàntiques, un tema fonamental i transversal dins la teoria de nombres.

Paraules clau: equacions diofàntiques, corbes el·líptiques, punts de Heegner, integració p -àdica, punts de Stark-Heegner.

Classificació MSC2020: 14H52, 11G05.

1 Equacions diofàntiques

Un dels temes centrals en teoria de nombres és l'estudi de les *equacions diofàntiques*, en què hom està interessat en les solucions enteres o racionals d'equacions polinòmiques amb coeficients enters. De manera més precisa, una equació diofàntica és una equació de la forma

$$F(x_1, \dots, x_n) = 0, \quad (1.1)$$

on F és un polinomi amb coeficients enters, i volem determinar les solucions en què totes les components viuen a $\mathbb{Z}$ o, alternativament, a $\mathbb{Q}$. Per exemple, $(-1, 0)$ és una solució entera de l'equació diofàntica

$$x^2 + y^2 - 1 = 0, \quad (1.2)$$

i $(-3/5, 4/5)$ n'és una solució racional. De fet, i això ho veurem més endavant, aquesta equació diofàntica té infinites solucions racionals.

El terme *diofàntic* fa referència al matemàtic Diofant d'Alexandria, que visqué durant el segle III i feu, entre altres, un tractat d'aritmètica en què estudià problemes que es resolten amb aquestes equacions. Veiem, doncs, que l'interès per les equacions diofàntiques ve de molt lluny. Al llarg dels segles, matemàtics il·lustres, com ara Fermat, Euler o Gauss, n'han tractat casos concrets i, de fet, sovint han estat la motivació d'algunes nocions algebraïques que han acabat formant part del cos teòric de la matemàtica actual. Un dels casos més famosos és el de l'equació de Fermat

$$F_n : x^n + y^n - z^n = 0, \quad (1.3)$$

on n és un natural fixat. Fermat deixà anotat en un marge d'un dels seus llibres d'estudi, justament un volum de l'*Aritmètica* de Diofant (vegeu figura 1), que, si $n \geq 3$, aleshores l'equació (1.3) no té solucions enteres no trivial; és a dir, no té solucions (x, y, z) amb $x, y, z \in \mathbb{Z}$ tots tres no nuls. Aquest enunciat, d'aparença innocent, romangué sense demostrar durant més de 300 anys, i moltes de les nocions i tècniques matemàtiques actuals provenen dels intents d'aquells que van succeir Fermat en la tasca de provar l'enunciat.¹

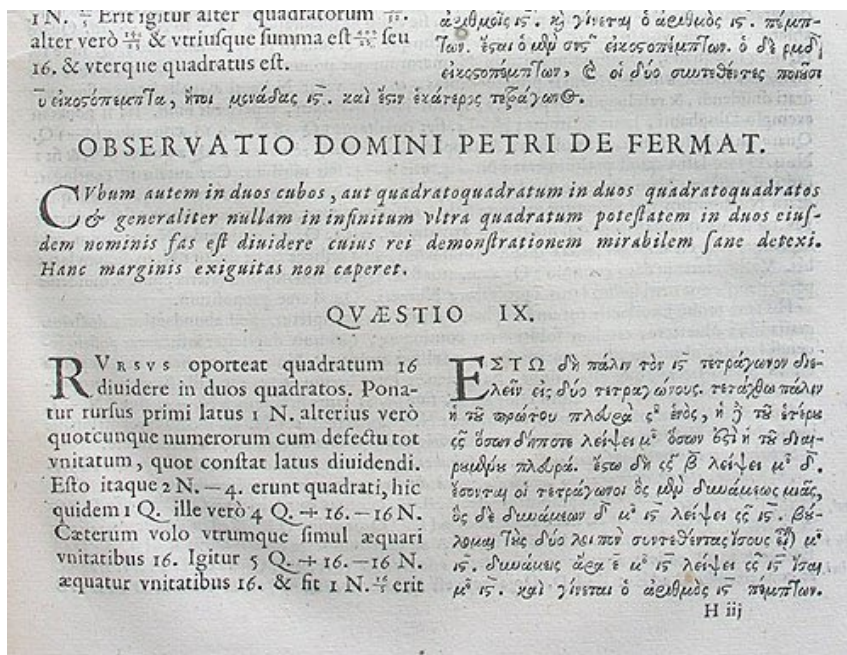


FIGURA 1: Edició posterior de l'*Aritmètica* de Diofant, amb l'observació de Fermat. (Font: https://en.wikipedia.org/wiki/Fermat%27s_Last_Theorem.)

¹ Fermat afirmà que n'havia trobat una demostració, però que el marge on escrivia era massa petit per a encabir-la-hi. Aquesta suposada prova no es va trobar mai, i aquest és, en part, el motiu pel qual tants matemàtics provaren de demostrar l'anomenat *darrer teorema de Fermat*.

Per exemple, la teoria general d'ideals en anells sorgí de l'estratègia ideada per Kummer per a atacar el problema; o, més recentment, els treballs de Wiles [37] i Taylor i Wiles [35] sobre la conjectura de modularitat, de la qual parlarem més endavant, foren el pas definitiu que resolgué el problema l'any 1994.

No es pot dir que hi hagi hagut sempre al llarg de la història un interès per una teoria general de les equacions diofàntiques, sinó que més aviat es tractaven de manera fragmentada, cas per cas. Un dels primers a plantejar la necessitat de sistematitzar-ne l'estudi fou Hilbert, que a la seva llista de 23 problemes que presentà al congrés internacional de matemàtiques de París l'any 1900, inclogué com a problema número 10 el següent:

Donada una equació diofàntica amb qualsevol nombre de quantitats desconegudes i amb coeficients numèrics enters: Idear un procés segons el qual es pugui determinar, en un nombre finit d'operacions, si l'equació és resoluble en nombres enters.

Aquest procés en un nombre finit de passos de què parlava Hilbert és el que avui en dia coneixem com a *algoritme*. El problema, tal i com el va plantejar Hilbert amb solucions enteres, resulta que no té solució. El 1970 Matiyasevich ([25]) donà una família d'equacions diofàntiques per a les quals no pot existir cap algoritme que pugui decidir si tenen solucions enteres o no. Això no obstant, el problema anàleg per a solucions racionals segueix obert. De fet, a partir d'ara ens centrarem en solucions racionals, de manera que, quan parlem d'equacions diofàntiques, ens hi referirem en el sentit de solucions racionals.

Malgrat ser més generals que les enteres, les solucions racionals solen presentar més estructura, la qual cosa n'acostuma a facilitar l'estudi. Un primer exemple d'això és el fet que una equació com (1.1) dona lloc a una varietat algebraica, i podem emprar les eines provinents de la geometria algebraica. Una conseqüència terminològica d'aquest punt de vista és que pensem les solucions $(x_1, \dots, x_n)$ com a punts de l'espai afí (o de l'espai projectiu, si el polinomi que defineix l'equació és homogeni). Així doncs, sovint ens referirem a una solució de (1.1) com a un punt, entenent que ens referim a un punt de la varietat definida per l'equació. Per exemple, l'equació (1.2) defineix una circumferència al pla afí, i les solucions racionals es corresponen als punts de la circumferència amb les dues components racionals.

Seguint amb l'esperit del problema de Hilbert, algunes de les preguntes més generals que podem plantejar per a equacions diofàntiques són:

1. Existeix algun algoritme que, donada una equació diofàntica, decideixi si aquesta té alguna solució racional?
2. Si l'equació té solucions, hi ha algun algoritme per a calcular-les totes? O, en qualsevol cas, per a calcular-ne alguna?
3. Podem dir alguna cosa sobre el nombre de solucions? Per exemple, decidir si és finit o infinit i donar-ne una fita en el cas finit?

Aquests problemes segueixen oberts, no només per al cas general, sinó, fins i tot, per al cas particular, molt restringit però, tot i això, molt interessant, de corbes planes, que serà en el que ens centrarem a partir d'ara.

Una *corba plana*² ve donada per un polinomi en dues variables $f(x, y) = 0$ en el cas afí, o per un polinomi homogeni en tres variables $F(X, Y, Z) = 0$ en el cas projectiu.³ De fet, l'exemple (1.2) és una corba plana afí i (1.3) és una corba plana projectiva.

L'invariant més important d'una corba és el seu *gènere*, que podem pensar en primera instància com una noció topològica. Considerem una corba C donada per una equació de la forma

$$C : f(x, y) = 0.$$

Com a equació diofàntica, estem interessats en les solucions racionals o, dit d'una altra manera, en el conjunt de punts racionals de C , que denotarem amb $C(\mathbb{Q})$. Més concretament,

$$C(\mathbb{Q}) = \{(x, y) \in \mathbb{Q} \times \mathbb{Q} : f(x, y) = 0\}.$$

Però també té sentit considerar el conjunt de punts de C amb components complexes; és a dir,

$$C(\mathbb{C}) = \{(x, y) \in \mathbb{C} \times \mathbb{C} : f(x, y) = 0\}.$$

Clarament $C(\mathbb{Q}) \subseteq C(\mathbb{C})$ i $C(\mathbb{C})$ admet una estructura topològica que governa determinats aspectes aritmètics de $C(\mathbb{Q})$. Per començar, vist com a subespai de $\mathbb{C} \times \mathbb{C}$, resulta que $C(\mathbb{C})$ és una superfície. Una manera de pensar-ho és escrivint les variables complexes x i y com $x = a + bi$ i $y = c + di$ amb $a, b, c, d \in \mathbb{R}$, i observant que l'equació $f(x, y) = 0$ dona lloc a dues equacions de nombres reals (una per a la part real i una per a la part imaginària) en les quatre variables reals a, b, c, d ; dues equacions en quatre variables reals defineixen una superfície. La compactificació d'aquesta superfície és una superfície orientable compacta que és, doncs, homeomorfa a un tor amb g forats per a algun $g \geq 0$; aquest g és el *gènere* de la corba. Per exemple, si la corba $C(\mathbb{C})$ és homeomorfa a l'esfera, aleshores $g = 0$; i si és homeomorfa al tor estàndard, $g = 1$.

En el cas de corbes planes, el gènere està relacionat amb un altre invariant important, que és el grau: si la corba és no singular⁴ de grau d , el gènere de la

² Ens restringirem al cas de corbes planes per simplicitat, però tot el que direm en aquesta secció també és cert per a corbes algebraïques més generals, no necessàriament planes (és a dir, que no estiguin contingudes necessàriament al pla afí o al pla projectiu, sinó en espais de dimensió superior).

³ En el cas projectiu, com que el polinomi $F(X, Y, Z)$ és homogeni, si $(X_0, Y_0, Z_0) \neq (0, 0, 0)$ és solució, aleshores també ho és $(\lambda X_0, \lambda Y_0, \lambda Z_0)$ per a tot $\lambda \in \mathbb{Q}$ no nul; totes aquestes solucions s'identifiquen amb un únic punt de l'espai projectiu.

⁴ Una corba és no singular si tot punt té una única recta tangent. La fórmula per al gènere que donem només val per a corbes planes que segueixen essent no singulars quan les pensem al pla projectiu.

corba és $(d-1)(d-2)/2$. Per exemple, la corba de Fermat F_n té gènere $(n-1)(n-2)/2$; en particular, F_2 té gènere 0 i F_3 té gènere 1.

Doncs bé, aquest invariant topològic permet distingir tres comportaments aritmètics molt diferenciats en les corbes, i és que dona informació molt important sobre el conjunt de punts racionals.

En primer lloc, les corbes de gènere 0 són còniques planes; és a dir, venen donades per $f(x, y) = 0$, on f és un polinomi de grau 2. Per a aquestes corbes, existeix un algorisme que permet decidir si hi ha algun punt racional. En cas que n'hi hagi algun, aleshores n'hi ha infinits i es poden trobar tots a partir d'un d'ells. Per exemple, en el cas de l'equació (1.2), que té gènere 0, podem partir del punt racional $(-1, 0)$. Aleshores, la recta que passa per aquest punt i té pendent t ve descrita per l'equació

$$y = t(x + 1). \quad (1.4)$$

Aquesta recta talla la corba en un altre punt P_t (vegi's la figura 2). Substituint (1.4) a (1.2), obtenim que

$$P_t = \left(\frac{1-t^2}{1+t^2}, \frac{2t}{1+t^2} \right).$$

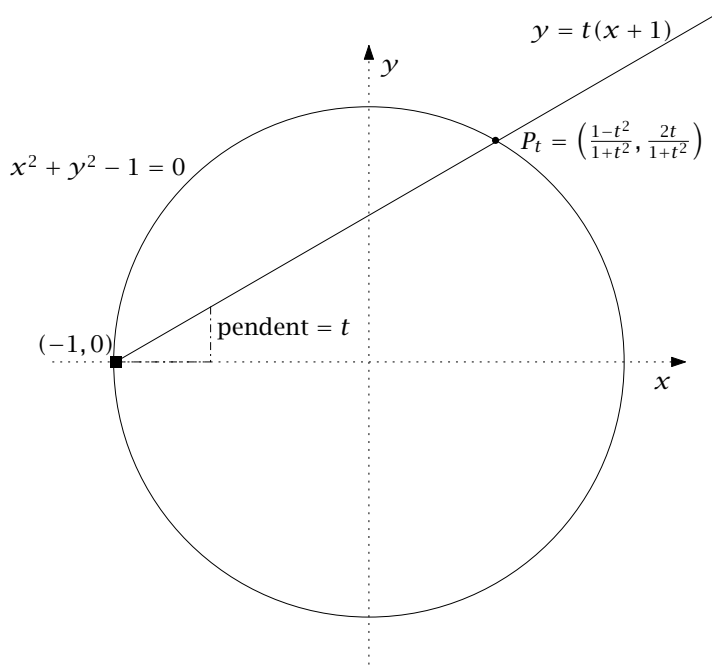


FIGURA 2: A partir del punt racional $(-1, 0)$ del cercle, podem trobar tots els altres amb rectes de pendent racional.

D'aquí es pot veure que P_t té coordenades racionals si i només si t és racional, i que si P' és un punt racional de la corba, aleshores es correspon a un punt de la forma P_t per algun valor $t \in \mathbb{Q}$. Hem vist, doncs, que la corba té infinits punts racionals i que els podem calcular explícitament si en coneixem un. El mateix procediment funciona per a qualsevol corba de gènere 0.

El cas següent serien les corbes de gènere 1, però aquest el deixarem per al final i ara parlarem de què succeeix amb les corbes de gènere més gran o igual que 2. Mordell va conjecturar l'any 1922 que, si una corba té gènere $g \geq 2$, aleshores té un nombre finit de punts racionals. Aquesta conjectura va ser demostrada per Faltings el 1983 [12], en un resultat que li va valer una medalla Fields l'any 1986. També hi ha resultats que donen una fita pel nombre de punts. Un dels més recents, de Dimitrov-Gao-Habegger [11], ens diu que

$$\#C(\mathbb{Q}) \leq c_g^{1+r(C)},$$

on c_g és una constant que només depèn de g i $r(C)$ és un invariant de C (el rang del grup de punts racionals de la seva jacobiana). Tot i que aquests resultats ens donen molta informació sobre el conjunt de punts d'una corba de gènere ≥ 2 , avui dia encara no es coneix cap algorisme per a determinar si una corba de gènere ≥ 2 té algun punt racional o no, ni per a calcular-los tots en cas que en tingui.

Ens queda, doncs, veure quina és la situació en el cas de corbes de gènere 1. Hem vist que una corba de gènere 0 o bé no té punts o bé en té infinits, i que una corba de gènere ≥ 2 té un nombre finit de punts. Les corbes de gènere 1 exhibeixen un comportament encara diferent. En primer lloc, de manera semblant al cas de gènere ≥ 2 , no es coneix cap algorisme per a determinar si una corba de gènere 1 té punts racionals o no. Però, a diferència dels casos de gènere 0 o gènere ≥ 2 , si en té, tant en pot tenir un nombre finit com un nombre infinit. De fet, una corba de gènere 1 amb algun punt racional és, justament, el que es coneix amb el nom de *corba el·líptica*, i el conjunt dels seus punts racionals admet una estructura molt especial que li confereix un seguit de propietats molt notables. Aquestes són les corbes o, si es vol, les equacions diofàntiques, a què dedicarem la resta de l'article.

2 Corbes el·líptiques

Una *corba el·líptica sobre* $\mathbb{Q}$ és una corba projectiva no singular de gènere 1 amb un punt racional distingit. Aquesta definició pot semblar una mica abstracta, però es pot veure que tota corba el·líptica es pot descriure, de fet, per una equació afí de la forma⁵

$$E: y^2 = x^3 + ax + b, \text{ amb } a, b \in \mathbb{Q} \text{ i } 4a^3 + 27b^2 \neq 0. \quad (2.1)$$

⁵ Tot i que el nom pugui dur a confusió, les corbes el·líptiques no són el·lipses, com es veu d'aquesta descripció: l'equació d'una el·lipse és de grau 2, mentre que la d'una corba el·líptica és de grau 3. Des del punt de vista històric sí que hi ha, però, una certa relació, a través de les funcions el·líptiques, que apareixen en el càlcul de la longitud de l'arc de les el·lipses i que es troben a l'origen del terme *corba el·líptica*.

La corba projectiva associada ve donada per l'equació

$$Y^2Z = X^3 + aXZ^2 + bZ^3.$$

La condició $4a^3 + 27b^2 \neq 0$ és equivalent al fet que la corba sigui no singular, i la quantitat $\Delta_E = -16(4a^3 + 27b^2)$ s'anomena *discriminant* de la corba. De l'equació projectiva veiem que el punt $O = [0 : 1 : 0]$ és un punt racional de E ; s'anomena *punt de l'infinit*, ja que és l'únic punt de la corba a la recta de l'infinit en el pla projectiu (és a dir, és l'únic punt de la corba projectiva que no veiem al model afí (2.1)).

L'equació (2.1) s'anomena *equació de Weierstrass* de la corba el·líptica. Sovint és convenient treballar amb equacions una mica més generals, de la forma

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \quad (2.2)$$

que s'anomenen *equacions generalitzades de Weierstrass*. Si treballem sobre $\mathbb{Q}$, com és el nostre cas, tota equació de la forma (2.2) es pot transformar mitjançant un canvi de variables en una de la forma (2.1). Però en cossos de característica 2 o 3 no sempre és possible, per això cal admetre també equacions generalitzades.

Seguint amb el plantejament de la secció 1, estem interessats en E com a equació diofàntica; és a dir, ens interessen els punts racionals de E :

$$E(\mathbb{Q}) = \{(x, y) \in \mathbb{Q} \times \mathbb{Q} : y^2 = x^3 + ax + b\} \cup \{O\}.$$

Com que hem definit els punts racionals fent servir l'equació afí, ens cal afegir-hi el punt O , que no veiem al model afí. Alternativament, podríem haver definit els punts racionals directament com els punts de l'espai projectiu que satisfan l'equació projectiva de la corba.

Ja ens hem trobat amb un exemple de corba el·líptica. Recordem que a la secció 1 hem vist que la cúbica de Fermat

$$F_3 : X^3 + Y^3 = Z^3$$

té gènere 1. Com que té punts racionals (per exemple, el punt $[1 : -1 : 0]$), és una corba el·líptica. De fet, el canvi de variables

$$x = \frac{12Z}{X+Y}, \quad y = \frac{36(X-Y)}{X+Y}$$

transforma l'equació F_3 en

$$E_3 : y^2 = x^3 - 432,$$

que és de la forma (2.1). Fixem-nos que el canvi de variables involucra funcions racionals amb coeficients a $\mathbb{Q}$; per tant, transforma punts racionals de E_3 en punts racionals de F_3 i estableix, de fet, una bijecció entre els punts racionals de les dues corbes.

La propietat més important de les corbes el·líptiques és que, a partir de dos punts racionals, podem fabricar-ne un altre amb un procediment geomètric. Ho explicarem amb un exemple. Suposem que tenim la corba el·líptica

$$E: y^2 = x^3 - 16x + 16. \quad (2.3)$$

Per ajudar-nos a visualitzar el procés, és útil dibuixar els punts reals de la corba, de manera semblant a com hem fet amb l'exemple del cercle anteriorment. Podem veure el resultat a la figura 3. Remarquem que d'aquests punts amb coordenades reals, les solucions racionals de l'equació diofàntica $y^2 = x^3 - 16x - 16$ es corresponen amb els punts de la gràfica amb les dues components racionals. Per exemple, els dos punts

$$P = (0, 4), \quad Q = (4, 4)$$

són punts racionals de la corba, és a dir, $P, Q \in E(\mathbb{Q})$, ja que tenen components racionals i satisfan l'equació.

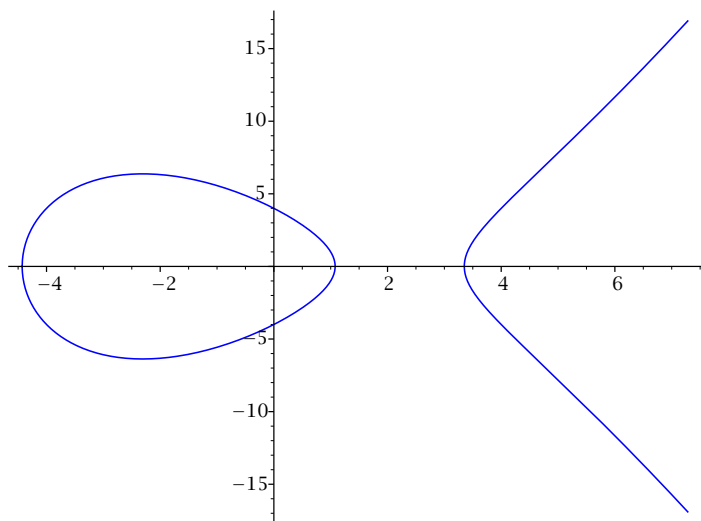


FIGURA 3: Els punts reals de la corba $y^2 = x^3 - 16x - 16$.

Resulta que la recta que passa per P i Q talla la corba E en un tercer punt. Escrivint l'equació de la recta $\overline{PQ}$ i substituint-la a l'equació de E , podem calcular que aquest tercer punt (vegeu la figura 4) és el $(-4, 4)$, que també té coordenades racionals i que anomenarem $P \star Q$.

Aquest procediment funciona per a tota corba el·líptica E i per a tot parell de punts racionals P i Q : la recta $\overline{PQ}$ ve donada per una equació de grau 1, i E per una de grau 3, de manera que la recta i la corba es tallen en tres punts. A més, un càlcul explícit permet veure que, si P i Q tenen coordenades racionals, el nou punt $P \star Q$ també.

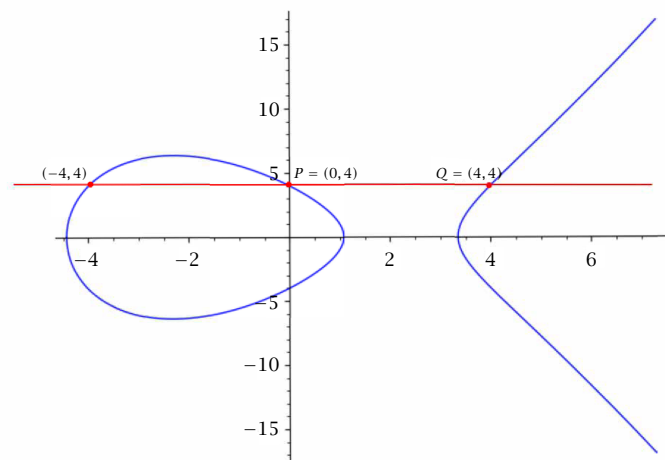


FIGURA 4: A partir dels punts racionals $P = (0, 4)$ i $Q = (4, 4)$ podem trobar el punt $(-4, 4)$, prenent la intersecció de la corba E amb la recta que passa per P i Q .

Per tant, donats dos elements de $E(\mathbb{Q})$, podem fabricar un altre element de $E(\mathbb{Q})$ amb aquest procediment geomètric, i això dona lloc a una llei de composició; és a dir, a una aplicació

$$\begin{aligned} E(\mathbb{Q}) \times E(\mathbb{Q}) &\longrightarrow E(\mathbb{Q}) \\ (P, Q) &\longmapsto P \star Q. \end{aligned}$$

Bé, hi ha un parell de detalls pendents que no hem tractat. El primer és què passa si prenem $P = Q$, ja que en aquest cas no té sentit parlar de la recta que uneix P i Q . Però, com que la corba és no singular, podem prendre la corba tangent a E que passa per P (si la corba fos singular en P , hi hauria més d'una recta tangent i això no tindria sentit). Podem repetir l'argument anterior amb aquesta recta tangent: la seva intersecció amb E ens dona un punt $P \star P$, que també és un punt racional.

El segon detall pendent és què passa si un dels punts que sumem és el punt O , que no veiem al model afí. La resposta és que aquest punt el podem pensar com «la direcció vertical»; per exemple, si volem calcular $P \star O$, prendrem com a recta $\overline{PO}$ la recta vertical que passa per P .

A aquest procediment per a construir nous punts a partir de punts coneguts se'l coneix com a *mètode de la secant i la tangent*, però la llei de composició que hem descrit cal modificar-la una mica per tal que tingui millors propietats

algebraiques. En comptes de $P \star Q$, prenem un altre punt definit de la manera següent: fem la recta vertical que passa per $P \star Q$ i considerem el punt d'intersecció d'aquesta recta amb E (vegeu la figura 5). Es pot veure que aquest punt, que anomenarem $P + Q$, també té coordenades racionals. Obtenim, doncs, una nova llei de composició

$$\begin{aligned} + : E(\mathbb{Q}) \times E(\mathbb{Q}) &\longrightarrow E(\mathbb{Q}) \\ (P, Q) &\longmapsto P + Q. \end{aligned}$$

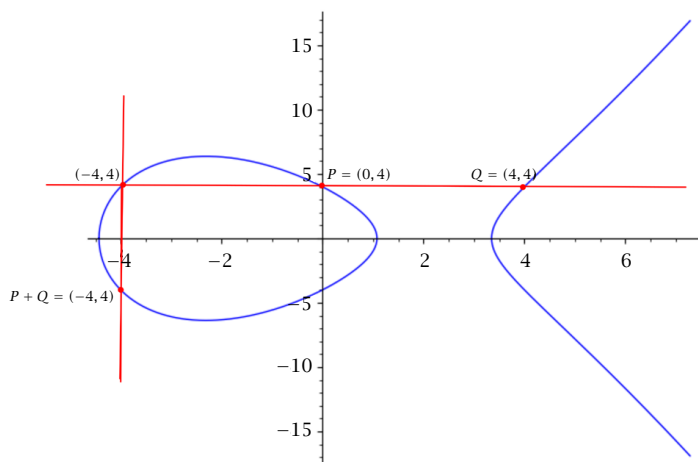


FIGURA 5: A partir del punt $P \star Q = (-4, 4)$, trobem un nou punt racional fent la recta vertical que passa per $P \star Q$ i tallant amb E . Aquest nou punt, en aquest cas el $(-4, -4)$, és el que anomenem $P + Q$.

Poincaré va demostrar que aquesta operació $+$ satisfà un seguit de propietats molt rellevants:

1. Per a tot $P \in E(\mathbb{Q})$ se satisfà que $P + O = O$. És a dir, O és un element neutre per a l'operació.
2. Tot element té un invers: per a tot $P \in E(\mathbb{Q})$, existeix un únic punt Q tal que $P + Q = O$. Aquest punt l'anomenarem $-P$.
3. L'operació és associativa: $(P + Q) + R = P + (Q + R)$.
4. L'operació és commutativa: $P + Q = Q + P$.

Veiem, doncs, que $E(\mathbb{Q})$ amb l'operació $+$ satisfà els axiomes de grup abelià.⁶ Aquest és el primer resultat clau de la teoria de corbes el·líptiques [30].

TEOREMA 2.1 (POINCARÉ, 1901). *L'operació $P, Q \mapsto P + Q$ dona una estructura de grup abelià al conjunt $E(\mathbb{Q})$.*

⁶ En canvi, l'operació $P, Q \mapsto P \star Q$ no compleix els axiomes d'una llei de grup, ja que no és associativa.

Poincaré també va conjecturar al mateix article⁷ que sempre podem trobar un conjunt finit de punts racionals a partir dels quals es pot construir qualsevol altre punt racional aplicant repetidament el mètode de la secant i la tangent. Dit amb terminologia moderna de teoria de grups, Poincaré conjecturà que $E(\mathbb{Q})$ és un grup finitament generat. Això fou demostrat per Mordell [28] un quart de segle després.

TEOREMA 2.2 (MORDELL, 1922). $E(\mathbb{Q})$ és un grup finitament generat.

Aquest resultat ja ens proporciona molta informació sobre l'estructura de $E(\mathbb{Q})$. El teorema de classificació de grups abelians finitament generats ens diu que es té un isomorfisme de la forma

$$E(\mathbb{Q}) \simeq \mathbb{Z}^r \oplus T,$$

on T és un grup finit i $r \in \mathbb{Z}_{\geq 0}$. El grup T es correspon amb els punts racionals de E d'ordre finit, o punts de torsió; és a dir, punts $P \in E(\mathbb{Q})$ tals que per a algun $n \in \mathbb{Z}_{>0}$ es té que⁸ $nP = O$. L'enter r denota el nombre màxim de punts d'ordre infinit linealment independents, i es coneix com a *rang* de E .

EXEMPLE 2.3. Per la corba E_3 corresponent a la cúbica de Fermat, resulta que

$$E(\mathbb{Q}) \simeq \mathbb{Z}/3\mathbb{Z}.$$

És a dir, $T = \mathbb{Z}/3\mathbb{Z}$ (hi ha tres punts racionals) i el rang és 0 (no hi ha cap punt d'ordre infinit). De fet,

$$E_3(\mathbb{Q}) = \{(12, 36), (12, -36)\} \cup \{O\}.$$

El punt $(12, 36)$ té ordre 3, i es compleix que $2(12, 36) = (12, -36)$ i $3(12, 36) = O$.

EXEMPLE 2.4. La corba E de (2.3) satisfà que

$$E(\mathbb{Q}) \simeq \mathbb{Z}.$$

És a dir, l'únic punt de torsió és O , i hi ha un punt d'ordre infinit a partir del qual es poden obtenir totes els altres punts racionals. El punt $P = (0, 4)$ és un generador; alguns dels seus múltiples són:

$$2P = (4, 4), \quad 3P = (-4, -4), \quad 4P = (8, -20), \quad 5P = (1, -1), \quad 6P = (24, 116).$$

Podria semblar que sempre obtindrem punts amb coordenades enteres. Però això no és cert; ja al següent múltiple de P veiem que $7P = (-20/9, 172/27)$ i a partir d'aquí tots els altres múltiples tenen coordenades racionals no enteres.

⁷ Per a una visió més completa de les contribucions de Poincaré a l'aritmètica, podeu consultar [1].

⁸ Com és habitual en els grups abelians, nP denota $P + P + \dots + P$.

EXEMPLE 2.5. Podem considerar ara la corba

$$E : y^2 = x^3 - 15\,058\,251x + 21\,601\,366\,470,$$

que satisfà que $E(\mathbb{Q}) \simeq \mathbb{Z}^2 \oplus \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z}$. Els generadors són

$$(51, 27), \quad (45, 153), \quad (-124, 62), \quad (9, 594).$$

Els dos primers punts tenen ordre infinit, el tercer té ordre 2 i el darrer té ordre 4.

Veiem, doncs, una propietat que havíem avançat a la secció 1: una corba el·líptica tant pot tenir un nombre finit de punts racionals com un nombre infinit. Però, gràcies a l'estructura algebraica del conjunt de punts, ara podem precisar molt més aquesta qüestió: podem preguntar-nos quina és l'estructura de grup de $E(\mathbb{Q})$. De manera natural, el problema es pot desglossar en dos, corresponents a la part de torsió i a la part lliure:

1. Quina és l'estructura del grup de punts de torsió?
2. Quin és el rang de E ?

També ens podem preguntar, és clar, sobre com calcular els generadors de $E(\mathbb{Q})$. Novament, en aquest problema podem distingir els generadors d'ordre finit i els d'ordre infinit.

Relacionat amb el punt 1 hi ha la qüestió sobre quins possibles grups apareixen com a grups de torsió de corbes el·líptiques. Això va quedar completament resolt gràcies a un resultat de Mazur [26] que havia estat conjecturat el 1908 per Levi.

TEOREMA 2.6 (MAZUR, 1977). *Si E és una corba el·líptica sobre $\mathbb{Q}$. Aleshores $E(\mathbb{Q})$ és isomorf a algun dels grups següents:*

- $\mathbb{Z}/N\mathbb{Z}$ amb $1 \leq N \leq 10$ o $N = 12$.
- $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2N\mathbb{Z}$ amb $1 \leq N \leq 4$.

A més, se sap que cadascun d'aquests 15 grups apareix com a grup de torsió d'infinites corbes el·líptiques.

En canvi, el rang és un invariant de les corbes el·líptiques del qual sabem molt menys. Per exemple, no se sap si hi ha corbes el·líptiques de rang arbitràriament gran. Abans hem mostrat exemples de corbes de rang 1 i 2. La corba el·líptica de rang més gran que es coneix⁹ va ser trobada per Elkies-Klagsbrun el 2024 i té rang ≥ 29 . El rècord anterior era una corba de rang ≥ 28 , i havia estat trobada per Elkies el 2006.

⁹ És la corba d'equació:

$$y^2 = x^3 - 35000013481153675195091309180305500105059398789899919909339650987x \\ + 2578119979771429916851971456123315357665330364495541948135763833876310204081495487106234897116966.$$

No se sap si hi ha corbes de rang més gran. Un argument heurístic de Park-Poonen-Voight-Wood [29] suggereix que els possibles rangs haurien d'estar fitats i, de manera més precisa, que només hi ha un nombre finit de corbes el·líptiques de rang més gran que 21.

Un dels problemes oberts més importants en teoria de nombres és, doncs, el de determinar el rang de les corbes el·líptiques, ja sigui des d'un punt de vista algorítmic o mitjançant alguna fórmula que permeti entendre millor la naturalesa d'aquest invariant. Aquest segon punt de vista és, precisament, el que persegueix la conjectura de Birch i Swinnerton-Dyer, que forma part de la llista de set problemes del mil·lenni presentats per la Fundació Clay l'any 2000. Abans de parlar d'aquesta conjectura, però, hem de fer un petit parèntesi en el nostre estudi de les solucions a $\mathbb{Q}$ i parlar, breument, de les solucions en cossos finits.

3 Punts sobre cossos finits

Suposem que tenim una corba el·líptica com ara, per exemple,

$$E: y^2 = x^3 - x + 1. \quad (3.1)$$

Si p és un nombre primer, podem pensar a trobar solucions de l'equació (3.1) mòdul p . Per exemple, per a $p = 3$ resulta que el punt $(2, 1)$ és una solució mòdul 3. En efecte, si substituïm $x = 2$ i $y = 1$ a l'equació de la corba, no obtenim una igualtat de nombres enters, però sí una congruència mòdul 3:

$$1^2 \equiv 2^3 - 2 + 1 \pmod{3}. \quad (3.2)$$

El costat esquerre de (3.2) val 1 i el costat dret val 7, i $7 \equiv 1 \pmod{3}$ o, el que és el mateix, $7 - 1$ és múltiple de 3. Fixem-nos que $(2, 1)$ no és un punt racional de la corba, ja que, com hem dit, quan substituïm $x = 2$ i $y = 1$ a l'equació de la corba no obtenim una igualtat de nombres racionals. Però sí que obtenim una congruència mòdul 3; equivalentment, obtenim una igualtat de classes a $\mathbb{Z}/3\mathbb{Z}$. Diem que $(2, 1)$ és un punt de E a $\mathbb{Z}/3\mathbb{Z}$. De fet, és fàcil trobar tots els punts de E a $\mathbb{Z}/3\mathbb{Z}$, ja que les úniques classes de residus mòdul 3 que cal considerar per a x i y són 0, 1 i 2, i podem provar totes les possibilitats i quedar-nos amb aquelles que proporcionin punts mòdul 3. Fent una cerca exhaustiva veiem que

$$E(\mathbb{Z}/3\mathbb{Z}) = \{(0, 1), (0, 2), (1, 1), (2, 1), (2, 2)\} \cup \{O\}.$$

El cos $\mathbb{Z}/p\mathbb{Z}$ es denota habitualment amb $\mathbb{F}_p$, i és l'únic cos finit de cardinal p , llevat d'isomorfisme. Per a cada nombre primer p , podem considerar la quantitat N_p de punts mòdul p ; és a dir, $N_p = \#E(\mathbb{F}_p)$. Per a l'exemple (3.1) hem calculat N_p per als primers valors de p :

p	2	3	5	7	11	13	17	19	23	29	31	37
N_p	3	7	8	12	10	19	14	22	23	37	35	36

Podem observar que N_p creix aproximadament al mateix ritme que p . De fet, un argument heurístic ens mostra que podem esperar que el valor de N_p estigui al voltant de $p + 1$. Per veure si $(x_0, y_0) \in \mathbb{F}_p \times \mathbb{F}_p$ pertany a $E(\mathbb{F}_p)$, substituïm x_0 a la part de la dreta de l'equació (3.1). Si el valor $x_0^3 - x_0 + 1$ és 0 a $\mathbb{F}_p$, aleshores $(x_0, 0)$ serà un punt; si és diferent de 0, aleshores trobarem dos punts si aquest valor és un quadrat a $\mathbb{F}_p$ i no trobarem cap punt si no és un quadrat. Com que la meitat dels valors no nuls de $\mathbb{F}_p$ són quadrats i la meitat són no quadrats, per a cadascun dels p valors possibles de x_0 esperem trobar un punt, en mitjana. Això ens dona p punts en total, que, si hi sumem el punt O , ens dona $p + 1$ com a valor esperat de N_p . Aquest argument és heurístic perquè no sabem quina és la distribució dels valors de $x_0^3 - x_0 + 1$ quan x_0 recorre $\mathbb{F}_p$. En qualsevol cas, podem esperar que la quantitat $a_p = a_p(E) = p + 1 - N_p$, que mesura la diferència entre el nombre de punts real i l'esperat, serà petita. A la figura 6 veiem representada la quantitat $p + 1 - \#E(\mathbb{F}_p)$ respecte als primers p en l'interval $2 \leq p \leq 10^5$, i veiem que no només aquest valor és petit respecte a p , sinó que sembla sempre confinat a una paràbola. Això és cert, i és el contingut d'aquest resultat conjecturat per Artin a la seva tesi i demostrat per Hasse.

TEOREMA 3.1 (HASSE, 1936). *Sigui E una corba el·líptica sobre $\mathbb{Q}$, p un primer i posem $a_p = p + 1 - \#E(\mathbb{F}_p)$. Aleshores, $|a_p| \leq 2\sqrt{p}$.*

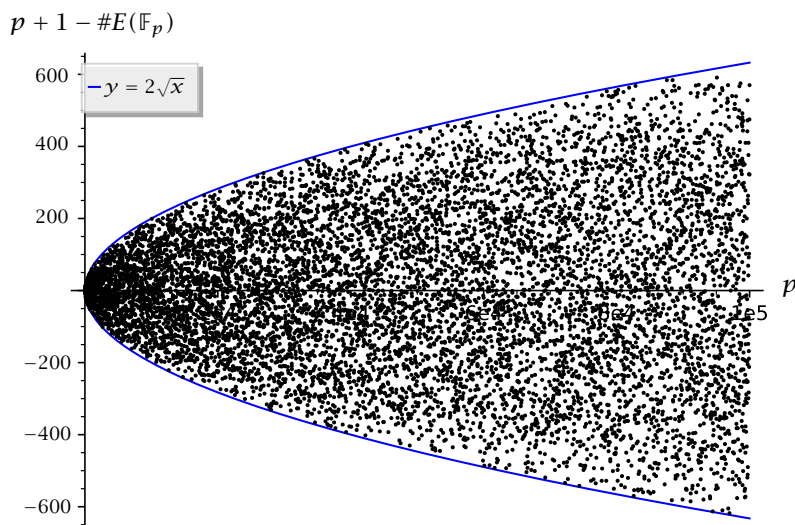


FIGURA 6: Il·lustració del teorema de Hasse.

No es coneix una fórmula tancada que, donada una corba E , calculi la quantitat a_p . En certa manera, la funció $p \mapsto a_p$ es comporta de manera aleatòria, i el que sí que coneixem és la distribució de probabilitat que segueix. Per

formalitzar aquest fet, comencem observant que per la fita de Hasse del teorema 3.1 la quantitat $\bar{a}_p = a_p / \sqrt{p}$ és un nombre real que cau a l'interval $[-2, 2]$. Podem preguntar-nos per la «distribució de probabilitat» de la quantitat $\bar{a}_p$, quan pensem p com una quantitat aleatòria. Dit d'una altra manera, donat un subinterval $[\alpha, \beta]$ de l'interval $[-2, 2]$, quina és la probabilitat que, si prenem un primer p a l'atzar, el corresponent $\bar{a}_p$ caigui a l'interval $[\alpha, \beta]$?

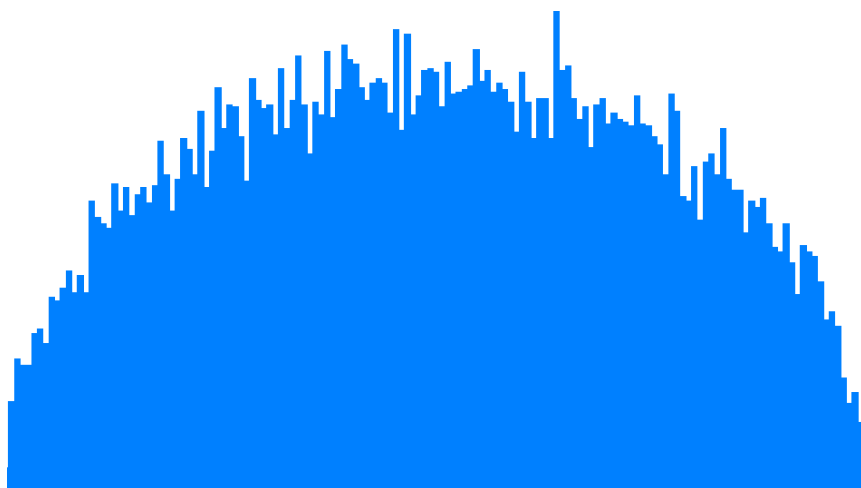


FIGURA 7: Il·lustració de la conjectura de Sato-Tate. Font: [33].

Observem a la figura 7 un histograma (per la corba $y^2 = x^3 + x + 1$) on s'han repartit els valors $\bar{a}_p$ entre 151 subintervalls de $[-2, 2]$ per a $2 \leq p \leq 2^{18}$. L'any 1960 Sato i Tate, de manera independent, van conjecturar que la distribució de probabilitat venia donada per un semicercle d'entre -2 i 2 ; és a dir, era la funció $\frac{4}{\pi} \sqrt{4 - x^2}$. La conjectura de Sato-Tate fou finalment demostrada a finals de la primera dècada del segle XXI [34, 20].

TEOREMA 3.2 (CLOZEL-HARRIS-SHEPHERD-BARRON-TAYLOR, 2008). *Si E és «genèrica», aleshores*

$$\lim_{x \rightarrow \infty} \frac{\#\{p \leq x : \bar{a}_p \in [\alpha, \beta]\}}{\#\{p \leq x\}} = \frac{4}{\pi} \int_{\alpha}^{\beta} \sqrt{4 - x^2} dx.$$

El terme *genèrica* fa referència al fet que la corba no tingui multiplicació complexa. Per no allargar-nos excessivament no en donarem la definició precisa (el lector interessat pot consultar [8, § 3.1]), però, en el cas en què la corba no sigui genèrica, també se sap quina és la distribució de probabilitat i, de fet, el resultat és anterior al del cas genèric.

4 La conjectura de Birch i Swinnerton-Dyer

La intuïció darrere la conjectura de Birch i Swinnerton-Dyer (d'ara en endavant, BSD) és que un rang gran de E hauria de donar lloc a molts punts mòdul p ; és a dir, esperaríem que N_p fos sistemàticament més gran que el nombre de punts esperat en general, que és $p + 1$. Per tal de quantificar aquesta idea, a finals de la dècada dels 1950 Birch i Swinnerton-Dyer van estudiar experimentalment la funció

$$C_E(x) = \prod_{p \leq x} \frac{N_p(E)}{p}, \quad (4.1)$$

on per a cada valor real de x el producte recorre tots els nombres primers p menors o iguals que x .

Emprant un dels primers ordinadors electrònics de la història, l'EDSAC, de la Universitat de Cambridge, van comprovar que per a corbes de rang 0 la funció $C_E(x)$ semblava comportar-se asimptòticament com una constant. En canvi, per a corbes de rang $r > 0$ la funció semblava tendir a infinit quan $x \rightarrow \infty$, i l'observació clau fou que ho feia proporcionalment a la funció $(\log x)^r$, com podem veure a la figura 8, on s'il·lustra amb corbes de rang fins a tres.¹⁰ És a dir, semblava que el rang de E es manifestava en l'ordre de creixement de la funció $C_E(x)$. Això va donar lloc a una primera versió de la conjectura [4].

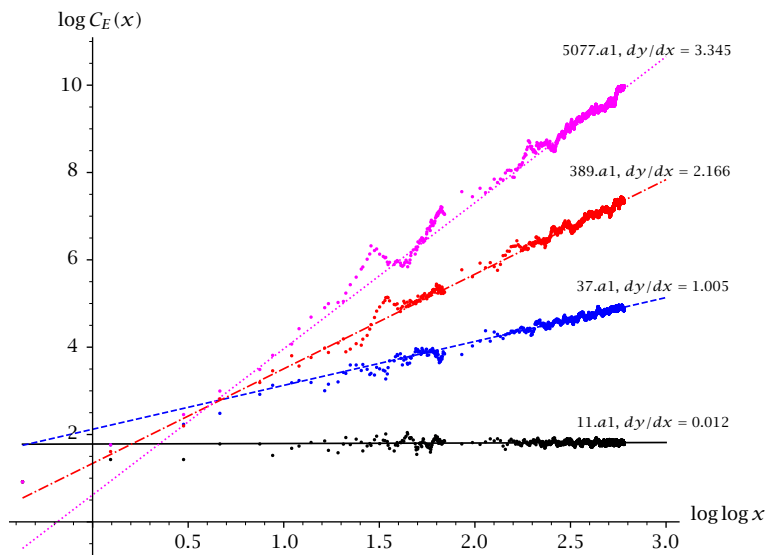


FIGURA 8: Il·lustració de la conjectura BSD, per les corbes 11.a1, 37.a1, 389.a1 i 5077.a1, de rangs 0, 1, 2 i 3, respectivament.

¹⁰ Les corbes anomenades 11.a1, 37.a1, 389.a1 i 5077.a1 es poden trobar a <https://www.lmfdb.org>, i són les «primeres» que tenen rangs 0, 1, 2 i 3, respectivament.

A l'enunciat següent, el símbol $\sim$ denota que el quocient de les dues funcions tendeix a 1 quan $x \rightarrow \infty$.

CONJECTURA 4.1 (BIRCH I SWINNERTON-DYER). *Si E és una corba el·líptica sobre $\mathbb{Q}$ de rang r , aleshores existeix una constant α_E tal que*

$$C_E(x) \sim \alpha_E (\log x)^r.$$

La formulació d'aquesta conjectura és un exemple de com les matemàtiques també poden tenir una component experimental essencial, i és que el càlcul computacional i l'observació de les dades obtingudes fou el que revelà la relació precisa entre el nombre de punts sobre cossos finits i el rang. De fet, una eina fonamental en l'estudi de les corbes el·líptiques són les taules que recullen exemples concrets de corbes i alguns dels seus invariants més notables (rang, punts de torsió, etc.), com ara les taules d'Anvers [3], les taules de Cremona [6] o la base de dades online LMFDB [22], accessible a <https://www.lmfdb.org/>.

La funció $C_E(x)$ resulta ser complicada de controlar des del punt de vista analític. Per això, habitualment la conjectura BSD s'enuncia en termes d'una altra funció amb millors propietats analítiques, que s'anomena *funció L de E* i és, de fet, una funció de variable complexa.¹¹ Aquesta funció es defineix com un producte infinit de funcions analítiques de variable complexa, indexades pels nombres primers i en què la contribució corresponent a un primer p incorpora la informació sobre el nombre de punts mòdul p , via la quantitat $a_p(E)$. De manera més precisa, si E ve donada per una equació

$$E: y^2 = x^3 + ax + b,$$

la funció L incompleta de E es defineix com a

$$L(E, s) = \prod_p \frac{1}{1 - a_p(E)p^{-s} + p^{1-2s}}, \quad (4.2)$$

on $s \in \mathbb{C}$ és la variable de la funció i p recorre tots els primers senars que no divideixen $4a^3 + 27b^2$. Es pot veure que la fita de Hasse $|a_p(E)| \leq 2\sqrt{p}$ implica que aquest producte infinit convergeix a una funció analítica en el semiplà del pla complex format pels nombres de part real més gran que $3/2$. De fet, Hasse també va conjecturar que $L(E, s)$ admet continuació analítica a tot el pla complex. Això està demostrat avui en dia, com a conseqüència del teorema de modularitat, demostrat per Breuil-Conrad-Diamond-Taylor el 2001 [5] amb les tècniques introduïdes per Wiles i Taylor i Wiles en la seva demostració del darrer teorema de Fermat.

En particular, té sentit avaluar $L(E, s)$ en $s = 1$. No podem fer-ho directament a la fórmula (4.2), ja que el producte infinit només convergeix per a part real

¹¹ Una altra funció de variable complexa que apareix de manera notable en aritmètica és la funció ζ de Riemann [31]. De fet, hi ha una construcció general que permet associar funcions L a objectes geomètrics i obtenir $\zeta(s)$ i $L(E, s)$ com a casos particulars, així que podem pensar en $L(E, s)$ com una determinada generalització de $\zeta(s)$.

de s més gran que $3/2$. Però, si ens oblidem per un moment del problema amb la convergència i substituïm formalment $s = 1$ al producte infinit, multiplicant i dividint el terme p -èsim per p obtenim que

$$L(E, 1) \ll \gg \prod_p \frac{1}{1 - a_p(E)p^{-1} + p^{-1}} \ll \gg \prod_p \frac{p}{N_p(E)}.$$

Les cometes als signes d'igualtat indiquen que l'argument és heurístic i no són igualtats rigoroses. Però, tolerant aquest inconvenient, si comparem amb (4.1), veiem que això hauria de correspondre a l'invers del valor $\lim_{x \rightarrow \infty} c_E(x)$. Per tant, sembla que el comportament de $L(E, s)$ en $s = 1$ hauria de reflectir el comportament asimptòtic de $c_E(x)$, i, com que un és l'invers de l'altre, com més gran sigui el rang r , més ràpid hauria de tendir a 0 la funció en $s = 1$. Això és, justament, el que diu la versió més habitual de la conjectura BSD; per exemple, la que apareix en la descripció oficial del problema del mil·lenni de la Fundació Clay [38].

CONJECTURA 4.2 (BSD). *Si E té rang r , aleshores $L(E, s)$ té un zero d'ordre r en $s = 1$. És a dir,*

$$L(E, s) = c(s - 1)^r + \text{termes d'ordre superior},$$

per a determinada constant $c \neq 0$.

Malgrat que a primer cop d'ull pugui semblar el contrari, aquesta conjectura no ens ajuda gaire en el problema algorítmic del càlcul del rang, ja que no es coneixen algorismes per a calcular l'ordre d'anul·lació de $L(E, s)$ en $s = 1$ quan aquest ordre és més gran que 3. De fet, avui dia no coneixem cap corba el·líptica E per la qual puguem demostrar que l'ordre d'anul·lació de $L(E, s)$ és 4.

Hi ha una «versió forta» de la conjectura, que involucra la funció L completa (en què s'incorporen també termes corresponents a $p \mid 2(4a^3 + 27b^2)$, que són una mica més complicats de definir). En aquest cas la conjectura prediu el valor de la constant c en el primer terme del desenvolupament de Taylor en funció de certs invariants aritmètics i geomètrics de E . Si aquesta versió forta fos certa, això sí que donaria lloc a un algorisme per a calcular el rang i un conjunt de generadors de $E(\mathbb{Q})$.

La conjectura BSD, fins i tot en la seva «versió dèbil» que hem enunciat, encara està oberta en completa generalitat. Això no obstant, se'n coneixen alguns resultats parcials. El més important des del punt de vista del present article és, sens dubte, el que es deriva dels treballs de Gross, Zagier i Kolyagin de finals de la dècada del 1980 [16, 21].

TEOREMA 4.3 (GROSS-ZAGIER 1986, KOLYVAGIN 1988). *Si l'ordre d'anul·lació de la funció $L(E, s)$ en $s = 1$ és 0 o 1, aleshores la conjectura BSD per a E és certa.*

Dit d'una altra manera: si $L(E, 1) \neq 0$, el rang és 0; i si $L(E, 1) = 0$ i $L'(E, 1) \neq 0$, el rang és 1. En particular, si $L(E, 1) = 0$ i $L'(E, 1) \neq 0$, aleshores hi ha un punt d'ordre infinit a E . Aquesta és la part del resultat que van provar Gross i Zagier, i per a la qual van resoldre el problema clau: donada una corba el·líptica amb $L(E, 1) = 0$ i $L'(E, 1) \neq 0$, com construir un punt racional d'ordre infinit? L'eina fonamental que van emprar són els anomenats *punts de Heegner*, introduïts uns anys abans per Birch i que, en certa manera, ja havien estat descoberts per Heegner en un context lleugerament diferent.

5 Punts de Heegner i exemples

Com ja hem avançat més amunt, els punts de Heegner són punts definits a la corba el·líptica E que són d'ordre infinit quan l'ordre d'anul·lació de la funció $L(E, s)$ és exactament 1, i, per tant, en aquesta secció assumirem que aquest és el cas. A més, ens cal utilitzar el fet que E és *modular*, així que primer veurem què vol dir que una corba el·líptica sigui modular. De fet, una manera ràpida de dir-ho és dir que la corba E està associada a una forma modular. Però aleshores ens cal explicar què és una forma modular, i què vol dir que E hi estigui associada.

Considerem, doncs, el semiplà superior de Poincaré $\mathbb{H}$, que simplement és conjunt de complexos amb part imaginària positiva:

$$\mathbb{H} = \{z \in \mathbb{C} \mid \text{Im}(z) > 0\}.$$

Aquest espai, que també s'anomena *semiplà hiperbòlic*, té una mètrica associada donada per

$$ds = \frac{dx dy}{y^2}, \quad z = x + iy.$$

Les isometries de $\mathbb{H}$ venen donades justament pel grup

$$\text{PSL}_2(\mathbb{R}) = \{\gamma \in \text{M}_2(\mathbb{R}) \mid \det \gamma = 1\} / \{\pm 1\},$$

on l'acció d'una matriu $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ és per una transformació fraccionària lineal

$$z \mapsto \frac{az + b}{cz + d}.$$

En particular, el grup $\text{SL}_2(\mathbb{R})$ actua a $\mathbb{H}$, i ens interessa estudiar funcions que es comporten de manera molt simètrica respecte a determinats subgrups. Concretament, en el cas de les corbes el·líptiques fixarem un enter positiu N i considerarem el grup

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{M}_2(\mathbb{Z}) \mid ad - bc = 1, N \mid c \right\},$$

és a dir, en el grup de matrius dos per dos a coeficients enters, determinant 1 i amb l'entrada inferior esquerra divisible per N .

DEFINICIÓ 5.1. Una forma modular de nivell N (cuspidal i de pes 2) és una funció holomorfa $f: \mathbb{H} \rightarrow \mathbb{C}$ tal que:

1. $f(yz) = (cz + d)^2 f(z)$ per a tota $y = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$, i
2. $\lim_{z \rightarrow i\infty} (cz + d)^{-2} f(yz) = 0$ per a tota $y = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$.

Com que la matriu $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ pertany a $\Gamma_0(N)$ sigui quin sigui N , tota forma modular f és 1-periòdica, i, per tant, té una sèrie de Fourier. De fet, com que f és holomorfa, podem escriure

$$f(z) = \sum_{n=1}^{\infty} a_n(f) e^{2\pi i n z}, \quad a_n(f) \in \mathbb{C}.$$

Sovint s'abreuja $q = e^{2\pi i z}$ i s'anomena $\sum_{n=1}^{\infty} a_n(f) q^n$ la q -expansió de f . Evidentment és temptador considerar la sèrie de Dirichlet associada a aquests coeficients, que no és més que la funció

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

Un teorema de Hecke ens assegura que per a tot primer p es té $|a_p(f)| = O(p)$, i, com a conseqüència dels resultats de Deligne sobre les conjectures de Weil, sabem que, de fet, $|a_p(f)| \leq 2\sqrt{p}$. En tot cas, la sèrie $L(f, s)$ convergeix per $\mathrm{Re}(s)$ suficientment gran. A més, la simetria de f ens permet continuar analíticament $L(f, s)$ a una funció holomorfa a tot $\mathbb{C}$.

TEOREMA 5.2 (TAYLOR-WILES, WILES, BREUIL-CONRAD-DIAMOND-TAYLOR). *Donada una corba el·líptica E , hi ha un enter positiu N_E i una forma modular f_E de nivell N_E , de manera que per a tot primer p es té*

$$a_p(f_E) = p + 1 - \#E(\mathbb{F}_p).$$

La quantitat N_E s'anomena *conductor* de la corba, i hi ha algorismes que el calculen de manera senzilla a partir de l'equació de Weierstrass. A més, els primers que divideixen N_E són essencialment els mateixos que divideixen el discriminant de E . És més, la funció $L(E, s)$ es pot definir de manera que també incorpori els primers que divideixen el discriminant, i aleshores el teorema de modularitat ens diu que hi ha una forma modular f_E tal que $L(f_E, s) = L(E, s)$.

Tal com l'hem enunciat, no està clar com es pot aprofitar aquest teorema per a construir punts a la corba el·líptica, que recordem que és el nostre objectiu.

Per a entendre com l'existència de la forma modular f_E ens permet trobar punts a E ens caldria entendre l'estructura algebraica que hi ha al darrere del quocient analític $Y_0(N) = \Gamma_0(N) \backslash \mathbb{H}$. En aquesta nota, però, ens conformarem amb la següent conseqüència.

Tal com hem vist a la secció 1, els punts complexos, que denotem amb $E(\mathbb{C})$, formen una superfície a l'espai de dimensió (real) 4 que és $\mathbb{C} \times \mathbb{C}$, que s'anomena

superfície de Riemann, ja que fou Riemann qui primer estudià aquest tipus de superfícies. En el cas que ens ocupa, es pot calcular un reticle

$$\Lambda_E = \{m + n\tau_E \mid m, n \in \mathbb{Z}\} \subset \mathbb{C}$$

per a un determinat element $\tau_E \in \mathbb{H}$, i es té un isomorfisme analític¹²

$$\mathbb{C}/\Lambda_E \xrightarrow{\cong} E(\mathbb{C}).$$

L'aplicació

$$\tau \mapsto \int_{i\infty}^{\tau} 2\pi i f_E(z) dz$$

envia $Y_0(N)$ a $\mathbb{C}/\Lambda$, on Λ és homotètic a Λ_E . Això fa que, reescalant la integral anterior, puguem considerar una aplicació analítica

$$\Phi_N: Y_0(N) \rightarrow E(\mathbb{C}),$$

que anomenarem *parametrització modular*. Val a dir que aquesta funció (o una aproximació arbitràriament bona) es pot calcular a partir de la q -expansió de f_E , i que s'ha fet molta recerca per a aconseguir que aquest càlcul sigui molt eficient.

L'últim ingredient que ens falta, doncs, per a obtenir punts a $E(\mathbb{Q}) \subset E(\mathbb{C})$ és trobar punts adequats a $Y_0(N)$. La funció Φ_N és altament transcendent, i, per tant, en principi no esperariem que punts algebraics a $\mathbb{H}$ donin lloc a punts algebraics a $E(\mathbb{C})$, cosa que gairebé no passa mai. Tot i així, la teoria de multiplicació complexa garanteix que, si $\tau \in \mathbb{H}$ és un quadràtic imaginari (és a dir, satisfà una equació de la forma $A\tau^2 + B\tau + C = 0$ amb $A, B, C \in \mathbb{Z}$ i $B^2 - 4AC < 0$), aleshores $\Phi_N(\tau) \in E(\mathbb{C})$ té coordenades definides¹³ en una extensió algebraica de $\mathbb{Q}(\tau)$. La situació és semblant al que passa amb les funcions trigonomètriques: la funció $x \mapsto \sin(2\pi x)$ és transcendent, però resulta que, si x és racional, aleshores $\sin(2\pi x)$ és un nombre algebraic.¹⁴

Vegem un exemple que ens ajudarà a clarificar la construcció que hem fet. Considerem la corba el·líptica¹⁵

$$E: y^2 + y = x^3 - x,$$

que és la corba el·líptica «més simple» de rang 1. La forma modular f_E associada a E per modularitat té q -expansió

$$f_E(q) = q - 2q^2 - 3q^3 + 2q^4 - 2q^5 + 6q^6 - q^7 + 6q^9 + \dots$$

¹² Un isomorfisme analític és aquell que ve donat per funcions holomorfes amb inversa holomorfa.

¹³ Per tal que això sigui cert ens cal assumir una condició tècnica, coneguda com la «hipòtesi de Heegner»: els primers divisors de N no poden ser *inerts* a $\mathbb{Q}(\tau)$.

¹⁴ De fet, la teoria de Galois ens permet predir en quina extensió de $\mathbb{Q}$ està definit $\sin(2\pi x)$ en aquest cas.

¹⁵ <https://www.lmfdb.org/EllipticCurve/Q/37/a/1>.

Considerem ara

$$\tau = \frac{-21 + \sqrt{-3}}{74} \in \mathbb{H}.$$

Calculem de manera aproximada $\Phi_N(\tau)$ i obtenim

$$\Phi_N(\tau) \simeq (-1.000\,000, 1.11 \cdot 10^{-16}) \in E(\mathbb{C}).$$

Podem comprovar que $(-1, 0) = 3(0, -1) + (0, 1)$, i que $E(\mathbb{Q}) = \{O, (0, 1)\} \times \langle (0, -1) \rangle$. Així doncs, la construcció dels punts de Heegner ens ha donat lloc al punt d'ordre infinit l'existència del qual hauríem pogut predir (gràcies a la conjectura BSD) del fet que la L-sèrie $L(E, s)$ tingués un zero (simple) a $s = 1$.

Observem que en aquest cas el punt que hem obtingut té coordenades racionals, però això prové del fet que $\text{rang } E(\mathbb{Q}) = \text{rang } E(\mathbb{Q}(\sqrt{-3})) = 1$.

Vegem per acabar un altre exemple que mostra com els punts de Heegner tenen coordenades a extensions (potser trivials) de cossos quadràtics imaginaris. En aquest cas, considerem la corba¹⁶ donada per l'equació

$$E: y^2 + y = x^3 - x^2 - 10x - 20,$$

que té conductor 11 i forma modular associada

$$f_E = q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 - 2q^9 + \dots$$

En aquest cas prenem

$$\tau = \frac{-9 + \sqrt{-7}}{22},$$

que genera el cos quadràtic imaginari $\mathbb{Q}(\sqrt{-7})$, i calculem

$$\Phi_N(\tau) \simeq (0.500\,000 - 1.322\,876\sqrt{-1}, -2.000\,00 - 5.291\,502\,6\sqrt{-1}).$$

El punt obtingut és molt proper a

$$P_{\sqrt{-7}} = \left(\frac{1 + \sqrt{-7}}{2}, -2 + 2\sqrt{-7} \right) \in E(\mathbb{Q}(\sqrt{-7})),$$

que es pot escriure com a

$$P_{\sqrt{-7}} = \left(-6, \frac{-1 - 11\sqrt{-7}}{2} \right) + (5, 5).$$

Es pot veure que $E(\mathbb{Q}(\sqrt{-7}))$ és isomorf (com a grup abelià) a $\mathbb{Z}/5\mathbb{Z} \oplus \mathbb{Z}$, amb generador de $\mathbb{Z}/5\mathbb{Z}$ donat per $(5, 5)$ i generador de la part lliure donat per $(-6, \frac{-1 - 11\sqrt{-7}}{2})$.

¹⁶ <https://www.lmfdb.org/EllipticCurve/Q/11/a/2>.

6 Corbes el·líptiques sobre els p -àdics

A finals del segle passat, Darmon proposà a [7] una construcció anàloga a la dels punts de Heegner, en el cas que l'extensió quadràtica que considerem sigui real. En aquesta secció i la propera intentarem donar una idea d'aquesta construcció, que, a diferència de la construcció de punts de Heegner, dona lloc a punts que—avui dia—encara no sabem demostrar que siguin algebraics.

Ja d'entrada ens adonem fàcilment que la construcció de Heegner no funcionarà en aquesta situació, ja que, si τ és real, aleshores segur que no pertany a $\mathbb{H}$ (recordem que $\mathbb{H}$ està format per complexos amb part imaginària *estrictament* positiva). La idea brillant de Darmon va ser considerar un altre tipus de semiplà que sí que contindria aquests punts reals. Així, substituï $\mathbb{H}$ per un pla p -àdic, on p és un primer divisor del conductor de la corba. Per tal de simplificar l'exposició suposarem que el conductor de la corba el·líptica que considerem és primer.

Els nombres p -àdics, que habitualment s'escriuen com a $\mathbb{Q}_p$, s'obtenen a partir dels racionals de manera anàloga a com se n'obtenen els nombres reals, és a dir, prenent una compleció dels racionals respecte d'un valor absolut. Si prenem el valor absolut usual, obtenim els reals $\mathbb{R}$, però, si prenem el valor absolut p -àdic (que veurem seguidament), aleshores obtindrem els nombres p -àdics.

El valor absolut p -àdic, que denotarem amb $|\cdot|_p$, es defineix en els enters com a

$$|a|_p = p^{-v_p(a)}, \quad v_p(a) = \max\{k \mid p^k \text{ divideix } a\}.$$

Per exemple, com que 3^2 divideix 18 i 3^3 no, tenim $|18|_3 = 1/9$. Així, si pensem 3-àdicament, el nombre 18 és *més petit que* el nombre 1 (per a qualsevol p , es té $|1|_p = 1$). De manera semblant, els nombres $3, 9, 27, \dots, 3^n, \dots$ tendeixen a zero 3-àdicament, encara que amb el valor absolut usual ens sembli que es fan grans.

El valor absolut p -àdic es pot estendre de manera directa als racionals, definint

$$\left| \frac{a}{b} \right|_p = \frac{|a|_p}{|b|_p},$$

i així obtenim un valor absolut que respecta la multiplicació: per a tot $x, y \in \mathbb{Q}$, tenim $|xy|_p = |x|_p |y|_p$.

De la mateixa manera que podem aproximar qualsevol real no nul mitjançant un enter multiplicat per una potència de 10, també podem aproximar qualsevol p -àdic no nul mitjançant un enter multiplicat per una potència de p . En aquest cas, dos enters seran molt propers (p -àdicament) si la seva diferència és divisible per una potència molt gran de p . Per exemple, els nombres 14 i 500 són bastant propers a $\mathbb{Q}_3$, ja que la seva diferència és $486 = 2 \cdot 3^5$, que té valor absolut 3-àdic igual a $3^{-5} \approx 0.004115$. D'altra banda, 14 està a distància 1 de 2, ja que $|14 - 2|_3 = |13|_3 = 1$.

De la mateixa manera que per a definir els punts de Heegner no n'hem tingut prou amb els reals i hem hagut de considerar els nombres complexos, en

el cas p -àdic també ens caldrà considerar un cos més gran que $\mathbb{Q}_p$. Recordem que per construir els complexos adjuntem una arrel del polinomi $x^2 + 1$, que és irreductible si el pensem en els reals. Així, considerarem una extensió $\mathbb{C}_p$ de grau 2 de $\mathbb{Q}_p$, que es pot obtenir, per exemple, adjuntant l'arrel d'un polinomi mònic de grau 2 amb coeficients a $\mathbb{Z}$ que sigui irreductible mòdul p .¹⁷

Hi ha una teoria de corbes el·líptiques sobre $\mathbb{Q}_p$, i resulta que també donen lloc a tors (en aquest cas, tors p -àdics), però en aquesta situació ens cal una versió multiplicativa. Fixem-nos que, en el cas complex, la funció exponencial $z \mapsto e^{2\pi iz}$ dona lloc a un isomorfisme $\mathbb{C}/\Lambda_E \cong \mathbb{C}^\times / q_E^\mathbb{Z}$, on $q_E = e^{2\pi i\tau_E}$ i

$$q_E^\mathbb{Z} = \{q_E^n \mid n \in \mathbb{Z}\}.$$

En el cas p -àdic, es té un isomorfisme analític (donat per sèries de potències amb coeficients p -àdics)

$$\eta_{\text{Tate}}: \mathbb{C}_p^\times / q_E^\mathbb{Z} \longrightarrow E(\mathbb{C}_p),$$

on $q_E \in \mathbb{Q}_p^\times$, que s'anomena *període de Tate*, s'obté també de manera senzilla a partir de l'equació de E .

Fins aquí hem intentat convèncer el lector de substituir els punts complexos de la corba el·líptica $E(\mathbb{C})$ pels punts p -àdics $E(\mathbb{C}_p)$. Ens hem de preguntar, doncs, per quins objectes anàlegs substituïrem el semiplà de Poincaré $\mathbb{H}$ i el grup $\Gamma_0(p)$. D'entrada, podem pensar que $\mathbb{H} = (\mathbb{C} \setminus \mathbb{R})^+$, on el símbol $+$ fa referència a la component connexa formada pels complexos amb part imaginària positiva. Així, no ens sorprendrà que definim el *semiplà p -àdic* $\mathbb{H}_p$ com a $\mathbb{C}_p \setminus \mathbb{Q}_p$ (en aquest cas, no té sentit considerar una component «positiva»). El grup $\text{PSL}_2(\mathbb{Q}_p)$ actua a $\mathbb{H}_p$ per transformacions fraccionàries lineals, i el grup anàleg a $\Gamma_0(p)$ que considerà Darmon, seguint Ihara, és el grup

$$\Gamma = \text{SL}_2(\mathbb{Z}[1/p]) = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{M}_2(\mathbb{Z}[1/p]) \mid \det \gamma = 1 \right\}.$$

La notació $\mathbb{Z}[1/p]$ significa l'anell format pels racionals a/b , on b és una potència de p . Ara ja tenim definits objectes p -àdics $\Gamma \backslash \mathbb{H}_p$ i $E(\mathbb{C}_p)$, i ens resta per veure:

1. Quin és l'anàleg p -àdic de l'aplicació Φ_N de la secció anterior, i
2. Quins punts interessants es poden definir a $\Gamma \backslash \mathbb{H}_p$ que donin lloc a punts algebraics a $E(\mathbb{C}_p)$?

En la secció següent veurem que el camí no és tan directe, i ens veurem obligats a combinar la teoria p -àdica amb la complexa per definir els punts que cerquem a $E(\mathbb{C}_p)$.

¹⁷ A la literatura sovint s'escriu $\mathbb{C}_p$ per a referir-se a la completació de la clausura algebraica de $\mathbb{Q}_p$, que aquí no farem servir.

7 Punts de Stark-Heegner i exemples

Recordem que el teorema de modularitat ens permet associar a una corba el·líptica E una certa forma modular f_E . De fet, el que acabem integrant és una 1-forma diferencial $\omega_E = 2\pi i f'(z) dz$. La propietat de simetria per al grup $\Gamma_0(p)$ (recordem que estem assumint que el conductor $N_E = p$ és primer) fa que ω_E sigui invariant per a aquest grup i, per tant, la forma diferencial ω_E és una forma a $Y_0(p)$.

Un dels punts clau de la construcció de Darmon consisteix a considerar el producte $\mathbb{H}_p \times \mathbb{H}$, on actua el grup Γ de manera diagonal, i a associar a E una $(1, 1)$ -forma diferencial Ω_E a $X_\Gamma = \Gamma \backslash (\mathbb{H}_p \times \mathbb{H})$. Val a dir que el llenguatge que estem fent servir no és gaire precís, ja que $\mathbb{H}_p$ és totalment disconnex amb la topologia p -àdica i la diferenciabilitat no funciona com amb els complexos. Per fer-ho precís hauríem de recórrer a la geometria rigidanalítica, però en aquest article l'analogia ens servirà.

De manera anàloga a com en la secció anterior hem integrat una 1-forma diferencial al llarg d'un camí complex $i\infty \rightsquigarrow \tau$, aquesta nova $(1, 1)$ -forma diferencial l'hauréem d'integrar sobre una 2-cadena. Fixem-nos que el camí complex $i\infty \rightsquigarrow \tau$ té com a vora la 0-cadena tancada $(\tau) - (i\infty)$. Així, si volem seguir l'analogia, hauríem de trobar un 1-cicle exacte, és a dir, que sigui la vora d'una 2-cadena. Heus ací com podem procedir: donat un punt $\tau \in \mathbb{H}_p$ que satisfaci una equació $A\tau^2 + B\tau + C = 0$ amb $B^2 - 4AC > 0$, es pot considerar el subgrup estabilitzador

$$\Gamma_\tau = \{\alpha \in \Gamma \mid \alpha\tau = \tau\},$$

que resulta estar generat per una matriu γ . També es pot veure que una certa potència¹⁸ de γ es pot escriure com a producte de commutadors:

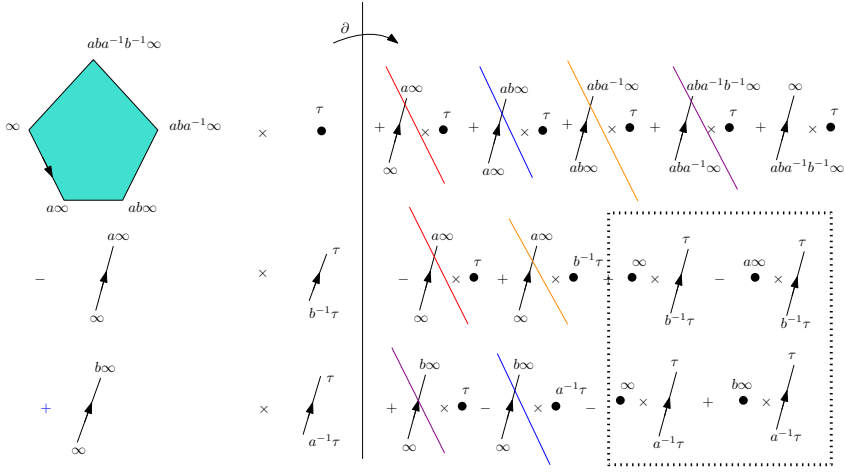
$$\gamma^e = [a_1, b_1][a_2, b_2] \cdots [a_t, b_t].$$

El cicle que considera Darmon és

$$\Theta = (\gamma^e \infty \rightsquigarrow \infty) \times \{\tau\} \in \Gamma \backslash (\mathbb{H} \times \mathbb{H}_p).$$

Aquest cicle és tancat, i, de fet, es pot veure que és exacte, és a dir, que hi ha una 2-cadena Δ tal que $\partial\Delta = \Theta$. Per fer-ho suposem, per tal de simplificar la notació, que només apareix un commutador en la descomposició de γ^e , així que $\gamma^e = aba^{-1}b^{-1}$. El diagrama següent mostra la 2-cadena corresponent:

¹⁸ En general, cal modificar γ fent servir uns operadors determinats coneguts com a *operadors de Hecke*, però podem obviar aquest detall en aquesta primera aproximació.



Observem que molts dels termes es cancel·len. D'altra banda, la part que apareix enquadrada es pot reescriure com a

$$\bullet \times \begin{array}{c} \tau \\ \nearrow \\ b^{-1}\tau \end{array} + \bullet \times \begin{array}{c} b^{-1}\tau \\ \nearrow \\ \tau \end{array} + \bullet \times \begin{array}{c} a^{-1}\tau \\ \nearrow \\ \tau \end{array} + \bullet \times \begin{array}{c} \tau \\ \nearrow \\ a^{-1}\tau \end{array} = \infty \times \left(\begin{array}{c} \tau \\ \nearrow \\ b^{-1}\tau \end{array} + \begin{array}{c} a^{-1}b^{-1}\tau \\ \nearrow \\ a^{-1}\tau \end{array} + \begin{array}{c} a^{-1}\tau \\ \nearrow \\ \tau \end{array} + \begin{array}{c} b^{-1}\tau \\ \nearrow \\ b^{-1}a^{-1}\tau \end{array} \right)$$

Però, com que

$$a^{-1}b^{-1}\tau = b^{-1}a^{-1}aba^{-1}b^{-1}\tau = b^{-1}a^{-1}\gamma\tau = b^{-1}a^{-1}\tau,$$

aquesta 1-cadena és nul·la. En conclusió, tots els termes excepte un es cancel·len i, per tant,

$$\Theta = \partial \left(\diamond \times \{\tau\} - (\infty \rightarrow a\infty) \times (b^{-1}\tau \rightarrow \tau) + (\infty \rightarrow b\infty) \times (a^{-1}\tau \rightarrow \tau) \right).$$

Podem definir la integral

$$J_\tau = \int_\infty^{a\infty} \int_\tau^{b^{-1}\tau} \Omega_E - \int_\infty^{b\infty} \int_\tau^{a^{-1}\tau} \Omega_E \in \mathbb{C}_p.$$

Per exemple, considerem la corba el·líptica de l'apartat anterior

$$E : y^2 + y = x^3 - x^2 - 10x - 20,$$

de conductor $p = 11$. Com que $x^2 - 13$ no té arrels mòdul 11, podem escollir

$$\tau = \frac{3 - \sqrt{13}}{2} \in \mathbb{H}_{11}.$$

Calculem una aproximació de la integral anterior, que dona lloc al punt

$$J_\tau = 275\,514\,415\,488\,849\,148\,207\tau + 302\,729\,526\,018\,381\,608\,293 \pmod{11^{20}}.$$

Aplicant l'isomorfisme a $E(\mathbb{C}_{11})$, obtenim

$$\eta_{\text{Tate}}(J_{\tau}) \simeq \left(645\,077\,502\,996 + 732\,103\,938\,720\tau + O(11^{20}), \right. \\ \left. 4\,172\,141\,459\,976\,479 + 4\,172\,141\,459\,976\,479\tau + O(11^{20}) \right),$$

que és proper al punt algebraic d'ordre infinit

$$2 \cdot \left(-8\,071\,272/139\,129\tau + 52\,783\,800/139\,129, \right. \\ \left. 109\,494\,875\,952/51\,895\,117\tau - 329\,723\,275\,860/51\,895\,117 \right) \in E(\mathbb{Q}(\sqrt{13})).$$

Com hem remarcat a la secció anterior, una diferència important entre la teoria de punts de Heegner de la secció 5 i la teoria de punts de Stark-Heegner que hem vist en aquesta secció és que aquesta darrera és una teoria encara majoritàriament conjectural. Això es deu al fet que la construcció que hem explicat de punts de Heegner és la vessant analítica d'una teoria geomètrica de punts de Heegner, que permet demostrar que els punts obtinguts no només són punts amb components complexes, sinó que, de fet, són punts amb components algebraiques. La construcció de punts de Stark-Heegner es pot veure com un anàleg analític p -àdic de la construcció complexa de punts de Heegner, però en el cas de Stark-Heegner no es coneix l'anàleg de la construcció geomètrica.

Tot i així, al llarg dels anys han aparegut molts resultats que donen evidència de la validesa d'aquesta conjectura, tant de caire computacional i experimental ([9, 17]), en què es comprova en nombrosos exemples que els punts de Stark-Heegner són molt propers a punts racionals d'ordre infinit, com de caire teòric ([27, 23, 2, 24, 10]), en què es demostren casos particulars de la conjectura.

8 Epíleg

La construcció original de Darmon s'ha generalitzat en múltiples direccions en les més de dues dècades que han passat des de l'article fundacional [7]. La hipòtesi de Heegner (sobre el comportament en el cos quadràtic dels divisors del conductor) es pot relaxar a canvi de considerar subgrups de $\text{SL}_2(\mathbb{R})$ o $\text{SL}_2(\mathbb{Q}_p)$ donats per àlgebres de quaternions en comptes dels grups $\Gamma_0(N)$ ([15]), i es poden considerar corbes el·líptiques sobre cossos de nombres diferents de $\mathbb{Q}$, com ara cossos quadràtics imaginaris ([36]). Avui en dia, disposem de construccions conjecturals de punts algebraics sempre que el signe de l'equació funcional de $L(E/K, s)$ sigui senar, i les conjectures prediuen que aquestes construccions donen lloc a punts algebraics d'ordre infinit sempre que l'ordre d'anul·lació de $L(E/K, s)$ a $s = 1$ sigui exactament 1. Per a més detalls, convidem el lector a consultar [19] i [18], en què la construcció es presenta des del punt de vista més general.

També cal mencionar l'existència de generalitzacions que van més enllà de la situació de corbes el·líptiques de rang 1 tractada en aquest article. Per

exemple, els cicles de Stark-Heegner ([32]) o, més recentment, els anomenats *punts plèctics*, que, inspirats per les construccions de Darmon, proporcionen invariants p -àdics en certes situacions de rang > 1 ([13, 14]).

En aquest article hem volgut fer un tast d'algunes conjectures sobre corbes el·líptiques i les corresponents respostes parcials que sabem donar. Concretament, la conjectura de Birch i Swinnerton-Dyer segueix essent una font d'inspiració per a la teoria de nombres del segle XXI, i les construccions explícites que va introduir Heegner i ha generalitzat Darmon i altres contemporanis ens proveeixen d'una petita ajuda a l'hora d'entendre aquesta misteriosa relació entre l'anàlisi de les L -sèries i l'aritmètica dels punts racionals. Segurament caldran altres idees fonamentalment noves per a poder demostrar la conjectura en casos de rangs superiors a 1, on fins avui tenim ben pocs resultats teòrics.

Referències

- [1] BAYER, P. «Les contribucions de Poincaré a l'aritmètica». *Butlletí de la Societat Catalana de Matemàtiques*, 21 (1) (2006), 5–38.
- [2] BERTOLINI, M.; DARMON, H. «The rationality of Stark-Heegner points over genus fields of real quadratic fields». *Ann. of Math. (2)*, 170 (1) (2009), 343–370.
- [3] BIRCH, B. J.; KUYK, W. (ED.). *Modular Functions of One Variable. IV*. Proceedings of the International Summer School, University of Antwerp, July 17–August 3, 1972. Berlin; Nova York: Springer-Verlag, 1975. (Lecture Notes in Math.; 476)
- [4] BIRCH, B. J.; SWINNERTON-DYER, H. P. F. «Notes on elliptic curves. II». *J. Reine Angew. Math.*, 218 (1965), 79–108.
- [5] BREUIL, C.; CONRAD, B.; DIAMOND, F.; TAYLOR, R. «On the modularity of elliptic curves over $\mathbf{Q}$: wild 3-adic exercises». *J. Amer. Math. Soc.*, 14 (4) (2001), 843–939.
- [6] CREMONA, J. E. *Algorithms for Modular Elliptic Curves*. 2a ed. Cambridge: Cambridge University Press, 1997.
- [7] DARMON, H. «Integration on $\mathcal{H}_p \times \mathcal{H}$ and arithmetic applications». *Ann. of Math. (2)*, 154 (3) (2001), 589–639.
- [8] DARMON, H. *Rational Points on Modular Elliptic Curves*. Washington, DC: Conference Board of the Mathematical Sciences; Providence, RI: American Mathematical Society, 2004. (CBMS Reg. Conf. Ser. Math.; 101)
- [9] DARMON, H.; POLLACK, R. «Efficient calculation of Stark-Heegner points via overconvergent modular symbols». *Israel J. Math.*, 153 (2006), 319–354.
- [10] DARMON, H.; ROTGER, V. «Stark-Heegner points and diagonal classes». A: *Heegner Points, Stark-Heegner Points, and Diagonal Classes*. *Astérisque*, 434 (2022), 1–28.
- [11] DIMITROV, V.; GAO, Z.; HABEGGER, P. «Uniformity in Mordell-Lang for curves». *Ann. of Math. (2)*, 194 (1) (2021), 237–298.

- [12] FALTINGS, G. «Endlichkeitssätze für abelsche Varietäten über Zahlkörpern». *Invent. Math.*, 73 (3) (1983), 349–366.
- [13] FORNEA, M.; GEHRMANN, L. «Plectic Stark-Heegner points». *Adv. Math.*, 414 (2023), article núm. 108861, 42 p.
- [14] FORNEA, M.; GUITART, X.; MASDEU, M. «Plectic p -adic invariants». *Adv. Math.*, 406 (2022), article núm. 108484, 26 p.
- [15] GREENBERG, M. «Stark-Heegner points and the cohomology of quaternionic Shimura varieties». *Duke Math. J.*, 147 (3) (2009), 541–575.
- [16] GROSS, B. H.; ZAGIER, D. B. «Heegner points and derivatives of L -series». *Invent. Math.*, 84 (2) (1986), 225–320.
- [17] GUITART, X.; MASDEU, M. «Elementary matrix decomposition and the computation of Darmon points with higher conductor». *Math. Comp.*, 84 (292) (2015), 875–893.
- [18] GUITART, X.; MASDEU, M.; MOLINA, S. «An automorphic approach to Darmon points». *Indiana Univ. Math. J.*, 69 (4) (2020), 1251–1274.
- [19] GUITART, X.; MASDEU, M.; ŞENGÜN, M. H. «Darmon points on elliptic curves over number fields of arbitrary signature». *Proc. Lond. Math. Soc. (3)*, 111 (2) (2015), 484–518.
- [20] HARRIS, M.; SHEPHERD-BARRON, N.; TAYLOR, R. «A family of Calabi-Yau varieties and potential automorphy». *Ann. of Math. (2)*, 171 (2) (2010), 779–813.
- [21] KOLYVAGIN, V. A. «Euler systems». A: *The Grothendieck Festschrift, Vol. II*. Boston, MA: Birkhäuser Boston, Inc., 1990, 435–483. (Progr. Math.; 87)
- [22] LMFDB, «The L-functions and modular forms database (LMFDB)». Publicat en línia (2024). <https://www.lmfdb.org>.
- [23] LONGO, M.; MARTIN, K.; HU, Y. «Rationality of Darmon points over genus fields of non-maximal orders». *Ann. Math. Qué.*, 44 (1) (2020), 173–195.
- [24] LONGO, M.; VIGNI, S. «The rationality of quaternionic Darmon points over genus fields of real quadratic fields». *Int. Math. Res. Not. IMRN*, 2014 (13) (2014), 3632–3691.
- [25] MATIYASEVICH, JU. V. «The Diophantineness of enumerable sets». *Dokl. Akad. Nauk SSSR*, 191 (2) (1970), 279–282. [En rus]
- [26] MAZUR, B. «Modular curves and the Eisenstein ideal». Amb un apèndix de B. Mazur i M. Rapoport. *Inst. Hautes Études Sci. Publ. Math.*, 47 (1977), 33–186.
- [27] MOK, C. P. «On a theorem of Bertolini-Darmon on the rationality of Stark-Heegner points over genus fields of real quadratic fields». *Trans. Amer. Math. Soc.*, 374 (2) (2021), 1391–1419.
- [28] MORDELL, L. J. «On the rational solutions of the indeterminate equations of the third and fourth degrees». *Proc. Cambridge Philos. Soc.*, 21 (1922/23), 179–192.

- [29] PARK, J.; POONEN, B.; VOIGHT, J.; WOOD, M. M. «A heuristic for boundedness of ranks of elliptic curves». *J. Eur. Math. Soc. (JEMS)*, 21 (9) (2019), 2859–2903.
- [30] POINCARÉ, H. «Sur les propriétés arithmétiques des courbes algébriques». *J. Math. Pures Appl.*, 7 (3) (1901), 161–233.
- [31] QUER, J. «La funció ζ de Riemann». *Butlletí de la Societat Catalana de Matemàtiques*, 22 (2) (2007), 197–228.
- [32] ROTGER, V.; SEVESO, M. A. « $\mathcal{L}$ -invariants and Darmon cycles attached to modular forms». *J. Eur. Math. Soc. (JEMS)*, 14 (6) (2012), 1955–1999.
- [33] SUTHERLAND, A. V. «Sato-Tate distributions in genus 1». <https://math.mit.edu/~drew/g1SatoTateDistributions.html>. [Consulta: 10 setembre 2024].
- [34] TAYLOR, R. «Automorphy for some l -adic lifts of automorphic mod l Galois representations. II». *Publ. Math. Inst. Hautes Études Sci.*, 108 (2008), 183–239.
- [35] TAYLOR, R.; WILES, A. «Ring-theoretic properties of certain Hecke algebras». *Ann. of Math. (2)*, 141 (3) (1995), 553–572.
- [36] TRIFKOVIĆ, M. «Stark-Heegner points on elliptic curves defined over imaginary quadratic fields». *Duke Math. J.*, 135 (3) (2006), 415–453.
- [37] WILES, A. «Modular elliptic curves and Fermat’s last theorem». *Ann. of Math. (2)*, 141 (3) (1995), 443–551.
- [38] WILES, A. «The Birch and Swinnerton-Dyer conjecture». A: *The Millennium Prize Problems*. Cambridge, MA; Clay Mathematics Institute, 2006, 31–41.

XAVIER GUITART
 DEPARTAMENT DE MATEMÀTIQUES I INFORMÀTICA
 UNIVERSITAT DE BARCELONA
 xevi.guitart@gmail.com

MARC MASDEU
 DEPARTAMENT DE MATEMÀTIQUES
 UNIVERSITAT AUTÒNOMA DE BARCELONA
 marc.masdeu@uab.cat

Les equacions de Friedmann: mecànica newtoniana *versus* relativitat general

JAUME DE HARO

Com pot ser que les matemàtiques, essent al cap i a la fi un producte del pensament humà independent de l'experiència, siguin tan admirablement adequades als objectes de la realitat? És que la raó humana, aleshores, sense experiència, només amb el pensament, pot comprendre les propietats de les coses reals? En la meua opinió: la resposta a aquesta pregunta és, breument, aquesta: en la mesura que les proposicions de les matemàtiques es refereixen a la realitat, no són certes; i en la mesura que són certes, no es refereixen a la realitat.

Albert Einstein (1879–1955)

Resum: En aquest assaig, presentem una derivació heurística de les equacions de Friedmann pertorbades basada en la mecànica newtoniana, oferint una perspectiva intuïtiva sobre aquestes relacions cosmològiques. Aquest enfocament ressalta l'acord notable entre les descripcions newtonianes i relativistes dins de certs límits, actuant tant com un nexa conceptual i fins i tot filosòfic i com una eina pedagògica per aprofundir en els principis físics que governen l'evolució de l'univers.

Simplificant el complex marc matemàtic de la teoria de la relativitat general, aquest mètode proporciona una via accessible per explorar la dinàmica de l'univers. Subratlla com la mecànica clàssica, quan s'aplica amb cura, pot oferir perspectives que s'alineen amb els resultats relativistes en l'aproximació lineal. Aquesta connexió no només aprofundeix la nostra comprensió de la universalitat de les lleis físiques, sinó que també reforça la idea que els fenòmens relativistes sovint es poden entendre a través d'anàlegs clàssics familiars.

Paraules clau: equacions de Friedmann, perturbacions cosmològiques, mecànica newtoniana, relativitat general.

Classificació MSC2020: 83C05, 83C99, 70F99, 83F05.

1 Introducció

En el present estudi, ens centrem en les equacions de Friedmann, formulades per primera vegada per Aleksandr Friedmann a principis de la dècada dels anys vint del segle passat. Aquestes equacions, sota certes hipòtesis, representen la forma que adopten les equacions de camp d'Einstein en el marc de la relativitat general (RG) quan s'apliquen a la cosmologia. Constitueixen la pedra angular de la cosmologia moderna, ja que descriuen la dinàmica a gran escala de l'univers sota les hipòtesis d'homogeneïtat i isotropia, tenint en compte la curvatura i el contingut d'energia i matèria de l'univers. Proporcionen un marc matemàtic per explorar com l'univers evoluciona al llarg del temps, abastant tant escenaris d'expansió com de contracció.

L'innovador treball de Friedmann [10, 11] inicialment va tenir dificultats per obtenir reconeixement. Durant anys, les solucions de les seves equacions, que revelaven la possibilitat d'universos en expansió i contracció, van ser en gran part ignorades per la comunitat científica [20]. Fins i tot Einstein, malgrat el seu paper fonamental en el desenvolupament de la RG, va desestimar inicialment els resultats de Friedmann, considerant-los inconsistents amb les seves pròpies equacions. Ben aviat, però, a instàncies del mateix Friedmann, Einstein va reconèixer-ne la validesa després d'un examen més detallat, tot i que durant deu anys va seguir sent escèptic sobre la idea d'un univers en expansió. No va ser fins a la formulació de la llei de Hubble-Lemaître, que va establir una relació directa entre el desplaçament cap al vermell de la llum de les galàxies i les seves distàncies, que el treball de Friedmann va ser plenament apreciat. Aquest avenç va connectar la cosmologia observacional amb la física teòrica, i va consolidar la rellevància de les solucions de les equacions de Friedmann per entendre l'evolució de l'univers.

Partint de les contribucions pioneres de Friedmann, l'estudi que presentem deriva les equacions de Friedmann pertorbades a través d'un enfocament alternatiu. Inspirant-nos en treballs anteriors, com els de McCrea i Milne [25] i posteriorment els de Callan, Dicke i Peebles [4], utilitzarem un mètode heurístic basat en la mecànica newtoniana. Aquesta perspectiva alternativa no només serveix com a eina pedagògica per comprendre les equacions de Friedmann, sinó que també subratlla la coherència entre els marcs newtonians i relativistes sota condicions específiques. A través d'aquest enfocament, es pretén il·luminar els principis físics subjacents a l'evolució còsmica i oferir perspectives que complementin la visió relativista.

Les implicacions d'aquest treball van més enllà de l'aspecte pedagògic. La derivació posa de manifest el paper de les inhomogeneïtats del factor d'escala i la taxa d'expansió de Hubble (la mesura de la velocitat en què s'expandeix l'univers), un fet amb profundes conseqüències observacionals. En particular, suggereix que les variacions locals en la taxa d'expansió de Hubble podrien sorgir a causa de diferències en els entorns físics dels objectes còsmics utilitzats per mesurar-la. Aquesta perspectiva podria oferir una nova visió sobre la *tensió de Hubble*, la persistent discrepància entre el valor actual de la taxa d'expansió de Hubble mesurada localment i el seu valor inferit del fons de microones còsmic, i proporcionaria així un marc per interpretar aquestes variacions.

A més a més, aquest enfocament posa èmfasi en la importància de l'anàlisi pertorbativa per connectar la teoria amb l'observació. Obre la porta a una exploració més profunda de com les estructures i dinàmiques locals influeixen en els paràmetres cosmològics a gran escala, connectant la física a petita escala amb el relat més ampli de l'evolució còsmica. Aquest enfocament dual —que ofereix tant una introducció pedagògica com perspectives sobre problemes avançats— subratlla la versatilitat del treball que presentem.

El treball està estructurat de la manera següent: A la secció 2, i partint de la consideració que l'univers presenta homogeneïtat i isotropia a gran escala, asumirem que està ple d'un fluid perfecte homogeni, els elements del qual es poden identificar amb les galàxies que el componen. En aquest context, derivarem les equacions de Friedmann en la seva forma homogènia, les solucions de les quals descriuen tant l'evolució d'aquest flux com l'expansió de l'univers, basant-nos en el marc teòric de la relativitat general. A la secció 3 s'exposa com aquestes equacions es poden derivar novament fent ús exclusivament de conceptes propis de la mecànica newtoniana. La secció 4, que constitueix l'apartat més rellevant d'aquest estudi, està dedicada a la determinació, de manera heurística, de les equacions de Friedmann pertorbades en primera aproximació. Som conscients que l'univers no és completament homogeni a gran escala, i aquestes inhomogeneïtats ens proporcionen informació essencial sobre la seva estructura. Per aquest motiu, és fonamental determinar les equacions pertorbades, ja que permeten descriure i entendre l'evolució d'aquestes inhomogeneïtats. A la secció 5, deduirem les equacions de conservació i d'Euler relativistes, les quals, juntament amb les equacions pertorbades de Friedmann, constitueixen el sistema que descriu de manera completa l'evolució de les inhomogeneïtats. A la secció 6 establim la relació entre la mètrica de Schwarzschild —una de les primeres solucions de les equacions d'Einstein— i la mecànica newtoniana. També analitzarem la seva connexió amb els forats negres i discutirem com, a partir d'aquesta mètrica, es pot reproduir correctament la precessió del periheli de Mercuri. Finalment, a la secció 7, debatrem el significat físic de les singularitats matemàtiques que, sovint, apareixen en cosmologia.

2 Les equacions de Friedmann a partir de la relativitat general

En aquesta secció obtindrem les equacions homogènies de Friedmann a partir de la relativitat general. Per poder fer-ho, primer de tot hem d'assumir l'homogeneïtat i la isotropia de l'univers a gran escala, cosa que ens porta a descriure'l geomètricament com una varietat de Lorentz (una varietat pseudo-riemanniana amb signatura $(1, 3)$, que és la generalització amb curvatura de l'espai-temps de Minkowski que apareix en la relativitat especial) amb la mètrica de Friedmann-Lemaître-Robertson-Walker (FLRW)

$$ds^2 = -N^2(t) dt^2 + a^2(t) \left(\frac{dr^2}{1 - kr^2} + r^2 d\Omega \right),$$

on $d\Omega = d\theta^2 + \sin^2 \theta d\phi^2$, $N(t)$ denota la funció lapse, que representa la llibertat de reparametritzar el temps, k representa la curvatura de l'espai, i $a(t)$ indica el factor d'escala: per entendre'ns, una mena de «radi de l'univers», almenys en el cas que l'espai tingui curvatura positiva.

Per a aquesta mètrica, l'invariant geomètric més important és l'escalar de Ricci o curvatura escalar, que s'expressa com a

$$R = 6 \left(\frac{1}{aN} \frac{d}{dt} \left(\frac{\dot{a}}{N} \right) + \frac{H^2}{N} + \frac{k}{a^2} \right),$$

on, en aquesta secció, el «punt» indica la derivada respecte al *temps còsmic* t , i $H = \frac{\dot{a}}{a}$ representa la taxa d'expansió de Hubble, també anomenada *paràmetre de Hubble*, ja que depèn del temps còsmic. Per tant, no sembla gaire apropiat anomenar-la *constant de Hubble*, com es fa habitualment.

Quan considerem un univers homogeni i isòtrop a gran escala, ple d'un fluid perfecte en què la pressió depèn únicament de la densitat d'energia, el model de FLRW ens permet descriure diferents etapes evolutives. Inicialment, el fluid estaria compost per radiació (dominant en la seva etapa primitiva), seguit d'una distribució homogènia dominada per matèria bariònica i, sobretot, matèria fosca. En l'època actual, el component predominant seria el que s'anomena *energia fosca*, en què la pressió és aproximadament igual a la densitat d'energia canviada de signe.

En aquest context, el lagrangiana d'Einstein-Hilbert ([18]) es pot expressar com una funció de l'escalar de Ricci, la mesura espai-temps i el contingut material, i pren la forma següent:

$$L_{EH} = \frac{1}{2} R \sqrt{-g} - 8\pi G \rho_0 \sqrt{-g} = \frac{1}{2} R a^3 N - 8\pi G \rho_0 a^3 N,$$

on ρ_0 representa la densitat d'energia homogènia, G és la constant de la gravitació universal, i $\sqrt{-g} = Na^3$ actua com a mesura quadridimensional per a la mètrica de FLRW.

Cal destacar que aquest lagrangiana es pot reformular com a

$$L_{EH} = \frac{d}{dt} \left(\frac{\dot{a}a^2}{N} \right) - \frac{3\dot{a}^2 a}{N} + 3kaN - 8\pi G \rho_0 a^3 N.$$

Per tant, tenint en compte que el primer terme és una derivada total, aquest lagrangiana és equivalent al següent:

$$\bar{L}_{EH} = -\frac{3\dot{a}^2 a}{N} + 3kaN - 8\pi G \rho_0 a^3 N.$$

Noti's que la variació respecte de la funció lapse condueix a la coneguda *restricció hamiltoniana*

$$0 = \frac{\partial \bar{L}_{EH}}{\partial N} = \frac{3\dot{a}^2 a}{N^2} + 3ka - 8\pi G \rho_0 a^3,$$

on, després d'escollir $N(t) = 1$ (recordem que la funció lapse només representa la llibertat de reparametritzar el temps. En triar-la igual a 1, el que estem fent és seleccionar el temps propi d'un observador que es troba en repòs respecte al fluid que omple l'univers), s'arriba a la primera equació de Friedmann:

$$H^2 + \frac{kc^2}{a^2} = \frac{8\pi G}{3}\rho_0, \quad (1)$$

que no és altra cosa que una restricció, ja que només conté la derivada primera del factor d'escala.

Per derivar l'equació dinàmica, realitzem una variació respecte del factor d'escala. Després de fixar $N = 1$, l'equació d'Euler-Lagrange

$$\frac{d}{dt} \left(\frac{\partial \tilde{L}_{EH}}{\partial \dot{a}} \right) = \frac{\partial \tilde{L}_{EH}}{\partial a}$$

ens porta a

$$6\ddot{a}a + 3\dot{a}^2 = 8\pi G \frac{d}{da}(\rho_0 a^3) - 3ka,$$

on, degut a la dependència de la densitat d'energia únicament respecte del factor d'escala, substituïm $\frac{\partial}{\partial a}$ per $\frac{d}{da}$.

A continuació, assumim una evolució adiabàtica en què l'entropia total es conserva, i utilitzem la primera llei de la termodinàmica $dE = -p dV$ (la variació de l'energia és igual a menys la pressió per la variació del volum). En el nostre cas, $E = a^3 \rho_0$, $V = a^3$ i $p = p_0$, essent p_0 la pressió homogènia, obtenim

$$d(\rho_0 a^3) = -p_0 da^3 \Rightarrow \frac{d}{da}(\rho_0 a^3) = -3p_0 a^2,$$

i, per tant:

$$\frac{\ddot{a}}{a} = -4\pi G p_0 - \frac{1}{2} \left(H^2 + \frac{k}{a^2} \right).$$

Finalment, a partir de la primera equació de Friedmann (1), arribem a l'equació de l'acceleració, o segona equació de Friedmann

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(3p_0 + \rho_0) = -\frac{4\pi G}{3}(1 + 3w_0)\rho_0, \quad (2)$$

on hem introduït el *paràmetre de l'equació d'estat* $w_0 \equiv \frac{p_0}{\rho_0}$.

Observem que aquesta última equació ens diu que, quan $w_0 > -1/3$, l'univers s'expandeix de manera desaccelerada, i, quan $w_0 < -1/3$, ho fa de manera accelerada.

Aquest fet és important, ja que, segons el model actual més utilitzat, l'anomenat Λ CDM-model (model Λ de matèria fosca freda, on Λ és la *constant cosmològica* d'Einstein), l'univers va experimentar inicialment, com explicarem a la secció 7, una fase d'expansió accelerada molt ràpida coneguda com a *inflació* [13]. Actualment, com sabem gràcies a les observacions de supernoves de

tipus la fetes pels grups liderats pels astrònoms guardonats amb el Premi Nobel Perlmutter, Riess i Schmidt a finals del segle passat, l'univers es troba de nou en una fase d'expansió accelerada. Per explicar aquesta segona fase, els cosmòlegs postulen que el component dominant de l'univers responsable d'aquesta acceleració s'anomena *energia fosca*. Curiosament, com ja hem comentat, el candidat més utilitzat per modelar-la és la famosa constant cosmològica, la qual dona lloc a una força repulsiva, i que Einstein ja havia introduït l'any 1917 per compensar l'atracció gravitatòria en el seu model estàtic.

Posteriorment, Einstein va haver de renunciar-hi un cop es va reconèixer que l'univers s'estava expandint, gràcies a les observacions de Hubble i altres astrònoms. Tot i així, ja abans d'aquest descobriment, cosmòlegs de prestigi, com ara Willem de Sitter, havien identificat que el model estàtic d'Einstein era dinàmicament inestable i que petites pertorbacions podrien desencadenar el col·lapse o l'expansió de l'univers. Aquesta inestabilitat inherent va ser un dels primers indicis que el model estàtic no era sostenible.

Com es diu amb certa ironia, «si s'esternuda en el model d'Einstein, l'univers col·lapsa», una expressió que, en el fons, subratlla que fins i tot una petita pertorbació seria suficient per alterar l'equilibri precari que plantejava. Avui dia, si ho analitzem des del punt de vista dels sistemes dinàmics, diríem que, qualitativament, el punt d'equilibri que representa matemàticament la solució d'Einstein és inestable; de fet, correspon a un punt de sella (vegeu, per exemple, [16]). Tanmateix, cal recordar que, en aquella època, la teoria qualitativa dels sistemes dinàmics encara no estava prou desenvolupada per proporcionar una anàlisi formal i detallada d'aquesta inestabilitat.

3 Les equacions de Friedmann a partir de la mecànica newtoniana

Ara tornarem a derivar les dues equacions de Friedmann homogènies, però des d'una perspectiva newtoniana. Considerem, en coordenades comòbils (són les coordenades adoptades per un observador que es mou amb el fluid homogeni que omple l'univers), una bola homogènia de mida gran amb un radi $\bar{R}$ en l'espai euclidià (també podem considerar $\bar{R} = \infty$, però, per a un radi finit, la massa total dins de la bola és finita: $\bar{M} = \frac{4\pi}{3} \varrho_0 \bar{R}^3$, on ϱ_0 és la densitat de massa homogènia).

Suposem que la bola s'expandeix radialment. Això vol dir que, si O és el centre de la bola, un punt P dins de la bola a l'instant t_0 es transforma en el punt P_t a l'instant t , i la distància des de O fins a P_t ve donada per $d_{OP_t} \equiv d_{OP}(t) = a(t) d_{OP} = a(t) |\vec{OP}|$, on $a(t)$, amb $a(t_0) = 1$, és el factor d'escala que ens ha aparegut a la secció anterior.

A més a més, a l'instant t_0 , considerem el triangle $\widehat{POQ}$, que es transforma en el triangle equivalent $\widehat{P_tOQ_t}$ a l'instant t . Per tant, com que $d_{OP_t} = a(t) d_{OP}$ i $d_{OQ_t} = a(t) d_{OQ}$, utilitzant el teorema de Tales, trobem que, per a qualsevol parell de punts P i Q dins de la bola, $d_{P_tQ_t} = a(t) d_{PQ}$.

La relació mostra que qualsevol bola, a l'instant t_0 , centrada en un punt P amb radi $R \ll \bar{R}$ s'expandeix radialment amb la mateixa taxa que la bola gran original. A més a més, la velocitat relativa entre P_t i Q_t segueix la llei de Hubble-Lemaître

$$\frac{d}{dt}(d_{P_t Q_t}) = \dot{a}(t) d_{PQ} = H(t) d_{P_t Q_t}.$$

L'equació de moviment per al factor d'escala en la mecànica newtoniana es deriva considerant una bola centrada en un punt donat P i amb un radi inicial R a l'instant t_0 . A l'instant t , la força radial en un punt donat Q_t , situat a la vora de la bola, es calcula com a $\mathbf{F}(Q_t) = f(a(t)R) \frac{\overrightarrow{P_t Q_t}}{a(t)R}$. Per determinar la funció f , es calcula el flux que entra a la bola

$$\Psi = \int \mathbf{F} \cdot \mathbf{n} dS = 4\pi a^2(t) R^2 f(a(t)R),$$

on $\mathbf{n}$ és la normal exterior a l'esfera que envolta la bola i dS és la mesura de l'àrea de l'esfera.

D'altra banda, a partir de l'equació clàssica de Poisson $\nabla \cdot \mathbf{F} = -4\pi G \varrho_0$, i el teorema de la divergència de Gauss, el flux també ve donat per

$$\Psi = \int \nabla \cdot \mathbf{F} dV = -\frac{16\pi^2 G}{3} \varrho_0 a^3(t) R^3,$$

la qual cosa condueix a la següent expressió de f :

$$f(a(t)R) = -\frac{4\pi G}{3} \varrho_0 a(t)R \Rightarrow \mathbf{F}(Q_t) = -\frac{4\pi G}{3} \varrho_0 \overrightarrow{P_t Q_t}.$$

Per tant, l'acceleració que experimenta una partícula de prova de massa m al punt Q a causa de la bola es determina mitjançant la segona llei de Newton com a

$$m \frac{d^2}{dt^2}(\overrightarrow{P_t Q_t}) = -\frac{4\pi G m}{3} \varrho_0 \overrightarrow{P_t Q_t} \Rightarrow \frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \varrho_0,$$

on hem usat que $\overrightarrow{P_t Q_t} = a(t) \overrightarrow{PQ}$, i ens proporciona la segona equació de Friedmann per a un fluid de pols, és a dir, amb pressió nul·la.

Aquesta equació es pot derivar a partir del lagrangià

$$L_N = \frac{R^2 \dot{a}^2}{2} + \frac{GM}{aR} = \frac{R^2 \dot{a}^2}{2} + \frac{4\pi G}{3} R^2 a^2 \varrho_0,$$

on $M = \frac{4\pi}{3} a^3 R^3 \varrho_0$ representa la massa dins de la bola.

De fet, utilitzant l'equació d'Euler-Lagrange s'obté

$$\frac{d}{dt} \left(\frac{\partial L_N}{\partial \dot{a}} \right) = \frac{\partial L_N}{\partial a} \Rightarrow \ddot{a} = \frac{4\pi G}{3} \frac{\partial}{\partial a} (a^2 \varrho_0) = -\frac{4\pi G}{3} \varrho_0 a,$$

on hem usat la conservació de la massa

$$\frac{\partial}{\partial a} (a^3 \varrho_0) = 0 \Rightarrow \frac{\partial}{\partial a} (a^2 \varrho_0) = -a \varrho_0.$$

Observem que el radi R de la bola triada no afecta les equacions dinàmiques. Per tant, fixem $R = 1$.

Per derivar la segona equació de Friedmann per a un fluid qualsevol, ara hem de tenir en compte l'energia del moviment microscòpic de les partícules que constitueixen el fluid, així com la interacció entre elles; per tant, substituïm la densitat de massa per la densitat d'energia, diem-n'hi ρ_0 , en el lagrangiana newtoniana amb $R = 1$, de la qual cosa resulta

$$\bar{L}_N = \frac{\dot{a}^2}{2} + \frac{4\pi G}{3} a^2 \rho_0.$$

Utilitzant l'equació d'Euler-Lagrange i la primera llei de la termodinàmica,

$$\begin{aligned} d(\rho_0 a^3) &= ad(\rho_0 a^2) + \rho_0 a^2 da \Rightarrow -3p_0 a^2 da = \\ &= ad(\rho_0 a^2) + \rho_0 a^2 da \Rightarrow d(\rho_0 a^2) = -(3p_0 + \rho_0)ada, \end{aligned}$$

derivem fàcilment (2).

El pas següent és obtenir la primera equació de Friedmann. Això es pot aconseguir combinant la segona equació amb la primera llei de la termodinàmica, expressada de la manera següent:

$$\dot{\rho}_0 = -3H(\rho_0 + p_0).$$

Primer, reescrivim (2) com a

$$\frac{\ddot{a}}{a} = -4\pi G(p_0 + \rho_0) + \frac{8\pi G}{3}\rho_0, \quad (3)$$

i calculem

$$\frac{dH^2}{dt} = 2H\frac{\ddot{a}}{a} - 2H^3.$$

Inserint (3) en aquesta última equació, obtenim

$$\frac{d}{dt} \left(H^2 - \frac{8\pi G}{3} \rho_0 \right) = -2H \left(H^2 - \frac{8\pi G}{3} \rho_0 \right) \Rightarrow \frac{d(H^2 - \frac{8\pi G}{3} \rho_0)}{H^2 - \frac{8\pi G}{3} \rho_0} = -2 \frac{da}{a}.$$

La solució d'aquesta equació ve donada per

$$H^2 - \frac{8\pi G}{3} \rho_0 = \frac{C}{a^2},$$

i establint la constant d'integració C igual a $-k$, obtenim la primera equació de Friedmann.

I, finalment, considerant que l'energia d'una bola homogènia de radi a és $E = \frac{4\pi}{3} a^3 \rho_0$, el lagrangiana newtoniana apareix com a $\bar{L}_N = E_{\text{cin}} - V$, on $E_{\text{cin}} = \frac{\dot{a}^2}{2}$ és l'energia cinètica per unitat de massa i $V = -\frac{GE}{a}$ és el potencial gravitatori generat per la bola amb massa en repòs E .

4 Les equacions de Friedmann pertorbades

Aquesta secció està basada en els recents resultats obtinguts a [17], en què tindrem en compte les petites inhomogeneïtats que té l'univers a gran escala, i que considerarem pertorbacions. Comencem amb la mètrica següent, anomenada $\mathbf{g}$, en l'aproximació del camp feble per un potencial newtonià estàtic $|\Phi_N| \ll 1$, que és la pertorbació de la mètrica de FLRW

$$ds^2 = (1 + 2\Phi_N(\mathbf{x})) dt^2 - (1 - 2\Phi_N(\mathbf{x})) d\mathbf{x}^2. \quad (4)$$

Aquesta expressió és una solució no trivial de les equacions pertorbades d'Einstein quan es considera un camp newtonià feble i estàtic. Es pot trobar a la fórmula (106.3) de [21] i també se'n fa referència en el llibre d'Einstein *The Meaning of Relativity* [7]. En la terminologia moderna, s'anomena habitualment *calibració newtoniana*. Aquí, $\mathbf{x}$ representa les coordenades físiques. Tenint en compte l'expansió de l'univers, descrita pel factor d'escala, i la variabilitat del potencial newtonià, la mètrica (4) en coordenades comòbils $\mathbf{q}$, on $\mathbf{x} = a\mathbf{q}$, adopta la forma dinàmica d'una mètrica *conformestàtica* ([31]), és a dir, d'una mètrica amb part espacial conformement plana i estàtica

$$ds^2 = \frac{a^2}{a_N^2(\mathbf{q}, t)} dt^2 - a_N^2(\mathbf{q}, t) d\mathbf{q}^2, \quad (5)$$

on $a_N(\mathbf{q}, t) \cong a(t)(1 - \Phi_N(\mathbf{q}, t))$ representa el factor d'escala pertorbat, que té un paper vital en la comprensió de la dinàmica de l'univers.

Ara expressem la primera equació de Friedmann en un univers FLRW pla espacialment i utilitzant el temps conforme $dt = a d\eta$, com el lligam següent:

$$\mathcal{H}((a^2)' - \mathcal{H}a^2) = \frac{8\pi G a^4}{3} \rho_0,$$

on usem la derivada respecte al temps conforme i $\mathcal{H} = a'/a$ és el paràmetre de Hubble conforme.

D'altra banda, escrivint $\rho = \rho_0 + \delta\rho$, $p = p_0 + \delta p$, l'equació clàssica de Poisson ve donada per

$$\Delta_{\mathbf{x}}\Phi_N = 4\pi G\delta\rho,$$

que, utilitzant les coordenades comòbils, té la forma

$$-\frac{1}{3}\Delta_{\mathbf{q}}(-2a^2\Phi_N) = \frac{8\pi G a^4}{3}\delta\rho.$$

Per reconciliar ambdues equacions, i considerant la linealitat, s'assumeix que la primera equació de Friedmann pertorbada és

$$\mathcal{H}(\partial_{\eta}(a_N^2) - \mathcal{H}a_N^2) - \frac{1}{3}\Delta_{\mathbf{q}}a_N^2 = \frac{8\pi G a^4}{3}\rho.$$

A continuació, ens ocupem de la segona equació de Friedmann no pertorbada, és a dir, la dinàmica, que es pot escriure com a

$$(a^2)'' - 2\mathcal{H}^2 a^2 = \frac{8\pi G a^4}{3} T_0, \quad (6)$$

on $T_0 = \rho_0 - 3p_0$ és la traça no pertorbada del tensor d'energia-impuls $\mathcal{T} = (\rho + p)\mathbf{u} \otimes \mathbf{u} - p\mathbf{g}$, essent $\mathbf{u}$ la forma dual de la velocitat del fluid que omple l'univers, i $\otimes$ el producte de Kronecker.

En l'esperit de les primeres teories escalars de la gravetat [1, 2, 27, 28], la generalització natural de l'equació clàssica de Poisson és

$$\partial_{t^2}^2 \Phi_N - \frac{1}{3} \Delta_{\mathbf{x}} \Phi_N = -\frac{4\pi G}{3} \delta T, \quad (7)$$

on podem veure que, en aquest marc, el potencial newtonià Φ_N és com una ona sonora que viatja amb velocitat $1/\sqrt{3}$, impulsada per la pertorbació de la font δT .

Una vegada tenim aquest, les equacions (6) i (7), l'equació dinàmica més senzilla que les conté és

$$\partial_{\eta^2}^2 a_N^2 - \frac{1}{3} \Delta_{\mathbf{q}} a_N^2 - 2\mathcal{H}^2 a_N^2 = \frac{8\pi G a^4}{3} T.$$

Per tant, en aquest enfocament, les equacions de Friedmann pertorbades són

$$\mathcal{H}(\partial_{\eta}(a_N^2) - \mathcal{H}a_N^2) - \frac{1}{3} \Delta_{\mathbf{q}} a_N^2 = \frac{8\pi G a^4}{3} \rho, \quad (8)$$

$$\partial_{\eta^2}^2 a_N^2 - \frac{1}{3} \Delta_{\mathbf{q}} a_N^2 - 2\mathcal{H}^2 a_N^2 = \frac{8\pi G a^4}{3} T. \quad (9)$$

És important destacar que aquestes equacions corresponen a les equacions linealitzades de la relativitat general. Concretament, les equacions pertorbades de la RG són ([26])

$$\begin{aligned} \Delta_{\mathbf{q}} \Phi_N - 3\mathcal{H}(\partial_{\eta} \Phi_N + \mathcal{H} \Phi_N) &= 4\pi G a^2 \delta \rho, \\ \partial_{\eta^2}^2 \Phi_N + 3\mathcal{H} \partial_{\eta} \Phi_N + (2\mathcal{H}' + \mathcal{H}^2) \Phi_N &= 4\pi G a^2 \delta p, \end{aligned}$$

les quals, després de combinar-se, condueixen, en l'aproximació lineal, a les equacions de Friedmann pertorbades (8) i (9).

Per trobar una forma més familiar de les equacions de Friedmann pertorbades, és important reconèixer que, en aquest enfocament, igual que en l'espai-temps de FLRW, emergeix un sistema de referència privilegiat: el sistema de referència comòbil, que es desplaça amb el fluid que omple l'univers. De fet, com que estem tractant amb un problema d'evolució, procedim com en el formalisme d'Arnótt-Deser-Misner (ADM) [3], en què la varietat de Lorentz té

una foliació privilegiada per a superfícies espacials Σ_t , i adoptem coordenades comòbils. En aquest marc, definim per ∂_t el vector normal de tipus temps, la taxa d'expansió de Hubble escalar en «temps còsmic» $H_N \equiv \frac{\partial_t a_N}{a_N}$ i el vector tridimensional de Hubble $\mathbf{H}_N = \frac{1}{a_N} \nabla_{\mathbf{q}} a_N$. Llavors, utilitzant aquestes definicions, les equacions de Friedmann pertorbades adquireixen la forma habitual de l'espai temps homogeni

$$H_N^2 - \frac{2}{3a_N^2} \nabla_{\mathbf{q}} \cdot \mathbf{H}_N = \frac{8\pi G}{3} \rho, \quad (10)$$

$$\frac{\partial_t^2 a_N}{a_N} + \frac{1}{3a_N^2} \nabla_{\mathbf{q}} \cdot \mathbf{H}_N = -\frac{4\pi G}{3} (\rho + 3p), \quad (11)$$

amb la peculiaritat que ens apareix la divergència del vector de Hubble.

També considerem el vector gradient quadridimensional d'una funció f , definit en geometria diferencial com a $\text{grad}_{\mathbf{g}} f = g^{\mu\nu} \partial_\mu f \partial_\nu$. Llavors, en la mètrica (5), tenim

$$\frac{1}{a_N} \text{grad}_{\mathbf{g}} a_N = H_N \partial_t - \frac{1}{a_N^2} \mathbf{H}_N.$$

De la mateixa manera, la divergència quadridimensional d'un quadrivector és $\text{div}_{\mathbf{g}} \mathbf{w} = \frac{1}{\sqrt{-g}} \partial_\mu (\sqrt{-g} w^\mu)$. Llavors, l'equació dinàmica de Friedmann s'expressa com la relació entre la geometria de la varietat de Lorentz i la matèria que omple l'univers

$$\text{div}_{\mathbf{g}} \mathbf{H}_{\mathbf{g},N} = 4\pi G(\rho - p), \quad (12)$$

on hem introduït el quadrivector de Hubble $\mathbf{H}_{\mathbf{g},N} \equiv \frac{1}{a_N} \text{grad}_{\mathbf{g}} a_N$.

Les altres equacions dinàmiques provenen de la conservació del tensor d'energia-impuls, és a dir, de $\text{div}_{\mathbf{g}} \mathbf{T} = 0$, que, en l'aproximació lineal, condueix a

$$\begin{aligned} \partial_t \rho + (\rho + p)[3H_N + \nabla_{\mathbf{q}} \cdot \mathbf{v}] &= 0, \\ \partial_t (a_N^2 (\rho + p) \mathbf{v}) + (\rho + p)[3H_N a_N^2 \mathbf{v} - \mathbf{H}_N] + \nabla_{\mathbf{q}} p &= 0, \end{aligned} \quad (13)$$

on $\mathbf{v} \equiv \frac{d\mathbf{q}}{ds}$ és la velocitat espacial.

En conclusió, les equacions dinàmiques que governen l'evolució de l'univers prenen una forma geomètrica força senzilla

$$\text{div}_{\mathbf{g}} \mathbf{H}_{\mathbf{g},N} = 4\pi G(\rho - p), \quad \text{div}_{\mathbf{g}} \mathbf{T} = 0.$$

Per acabar la secció, farem quatre observacions rellevants:

1. Podem escriure la primera equació de Friedmann en una forma més geomètrica si es considera la curvatura intrínseca, diem-n'hi $\mathcal{R}$, de la varietat de Lorentz. Com que al primer ordre de pertorbacions tenim $\mathcal{R} = \frac{4}{a_N^2} \nabla_{\mathbf{q}} \cdot \mathbf{H}_N$ i $H_N^2 = \mathbf{g}(\mathbf{H}_{\mathbf{g},N}, \mathbf{H}_{\mathbf{g},N})$, podem escriure

$$\mathbf{g}(\mathbf{H}_{\mathbf{g},N}, \mathbf{H}_{\mathbf{g},N}) - \frac{1}{6} \mathcal{R} = \frac{8\pi G}{3} \rho.$$

2. La combinació de les dues equacions de Friedmann (10) i (11) juntament amb l'equació de conservació (13) condueix a la restricció

$$\Delta_{\mathbf{q}} \partial_t a_N = 4\pi G a_N^3 (\rho + p) \nabla_{\mathbf{q}} \cdot \mathbf{v}, \quad (14)$$

i, combinant les dues equacions de Friedmann, obtenim

$$\partial_t H_N + \frac{1}{a_N^2} \nabla_{\mathbf{q}} \cdot \mathbf{H}_N = -4\pi G (\rho + p), \quad (15)$$

i també

$$2\partial_t H_N + 3H_N^2 = -8\pi G p.$$

3. Multiplicant (15) per H_N , combinant amb l'equació de conservació i la restricció (14), s'obté la primera equació de Friedmann.
4. Definint les *coordenades espacials físiques* com a $\mathbf{x}_N = a_N \mathbf{q}$, prenent la derivada respecte al temps propi s i utilitzant que $\partial_t = \frac{a_N}{a} \partial_t$, s'obté

$$\frac{d\mathbf{x}_N}{ds} = \left(\frac{dt}{ds} \partial_t a_N + \mathbf{v} \cdot \nabla a_N \right) \mathbf{q} + a_N \mathbf{v},$$

i, tenint en compte que $\frac{ds^2}{ds^2} = 1 = \left(\frac{a}{a_N} \right)^2 \left(\frac{dt}{ds} \right)^2 - a_N^2 |\mathbf{v}|^2$, s'arriba a

$$\frac{d\mathbf{x}_N}{ds} = \left(\sqrt{1 + a_N^2 |\mathbf{v}|^2} H_N + \mathbf{v} \cdot \mathbf{H}_N \right) \mathbf{x}_N + a_N \mathbf{v},$$

que, en l'aproximació lineal, condueix a la llei de Hubble-Lemaître pertorbada

$$\frac{d\mathbf{x}_N}{ds} = H_N \mathbf{x}_N + a_N \mathbf{v}.$$

Així, hem aconseguit reobtenir, de manera heurística, les equacions que descriuen l'evolució de les inhomogeneïtats de l'univers, les quals ens permeten comprendre amb més profunditat la formació d'estructures complexes com les galàxies i els cúmuls de galàxies. A més a més, aquestes equacions també ofereixen la possibilitat de confrontar els resultats teòrics amb les observacions actuals, cosa que facilita una validació empírica del model cosmològic. En particular, ens permeten veure com el model actual de la inflació explica adequadament la fase més primitiva de l'univers, i contribuir així a la nostra comprensió del seu origen i de la seva evolució. Aquestes connexions entre teoria i dades observacionals són fonamentals per consolidar les nostres idees sobre l'univers i les seves característiques a gran escala.

5 Les equacions de conservació i d'Euler relativistes

Un cop obtingudes les equacions de Friedmann pertorbades, podem trobar les equacions d'evolució per a la densitat d'energia i la velocitat de les pertorbacions, a partir de l'equació de conservació $\text{div}_g \mathcal{T} = 0$. No obstant això,

volem trobar-les a partir del primer principi de la termodinàmica i de l'equació clàssica d'Euler dels fluids.

Primer de tot, volem obtenir l'equació de conservació de la relativitat general. Per fer-ho, necessitem el resultat següent: Sigui $\varphi_\eta: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ el flux d'un fluid perfecte, amb $\varphi_{\bar{\eta}}$ sent la transformació identitat. Definim la velocitat tridimensional en temps conforme $\bar{\mathbf{u}}(\varphi_\eta(\mathbf{q}), \eta) = \frac{d\varphi_\eta(\mathbf{q})}{d\eta}$. Aleshores, arribem al resultat fonamental (trobareu més detalls en el molt recomanable llibre del professor Girbau [12])

$$\left[\frac{d}{d\eta} \int_{\varphi_\eta(V)} f(\mathbf{q}, \eta) dq_1 dq_2 dq_3 \right]_{\eta=\bar{\eta}} = \int_V (\partial_\eta f + \nabla_{\mathbf{q}} \cdot (f \bar{\mathbf{u}}))_{\eta=\bar{\eta}} dq_1 dq_2 dq_3.$$

Ara, considerem un volum comòbil infinitesimal δV , on suposarem que, dins d'aquest element de volum, la pressió és la mateixa en tots els punts. Aleshores, aplicant aquest resultat a la primera llei de la termodinàmica, obtenim ([14])

$$\left[\frac{d}{d\eta} \int_{\varphi_\eta(\delta V)} \rho(\mathbf{q}, \eta) dV \right]_{\eta=\bar{\eta}} = -p(\mathbf{q}, \eta) \left[\frac{d}{d\eta} \int_{\varphi_\eta(\delta V)} 1 dV \right]_{\eta=\bar{\eta}},$$

on l'element de volum espacial és $dV = a_N^3 dq_1 dq_2 dq_3$. Aquesta formulació integral és equivalent a l'equació diferencial

$$D_\eta \rho + (\rho + p)[3\mathcal{H}_N + \nabla_{\mathbf{q}} \cdot \bar{\mathbf{u}}] = 0,$$

on hem utilitzat la derivada total $D_\eta \rho = \frac{\partial \rho}{\partial \eta} + \bar{\mathbf{u}} \cdot \nabla_{\mathbf{q}} \rho$ i hem introduït la taxa d'expansió de Hubble $\mathcal{H}_N \equiv \frac{\partial_\eta a_N}{a_N}$ en temps conforme.

Mantenint només els termes lineals, s'obté

$$\partial_\eta \delta \rho + 3\mathcal{H}(\delta \rho + \delta p) + (\rho_0 + p_0)[\nabla_{\mathbf{q}} \cdot \bar{\mathbf{u}} - 3\partial_\eta \Phi_N] = 0,$$

el qual coincideix, fins al primer ordre, amb les equacions de la relativitat general, i a l'ordre zero s'obté òbviament l'equació de conservació en l'espaitemps de FLRW pla, és a dir, $\dot{\rho}_0 = -3H(\rho_0 + p_0)$.

Finalment, necessitem l'equació d'evolució de la velocitat $\bar{\mathbf{u}}$. En primer lloc, recordem que per a un fluid de pols, és a dir, amb $|p| \ll \rho$, prenent l'element de línia $ds^2 = dt^2 - d\mathbf{q}^2$ en l'espaitemps de Minkowski, l'equació d'Euler clàssica es pot escriure com a

$$\left[\frac{d}{dt} \int_{\varphi_t(V)} \varrho \bar{\mathbf{v}} dV \right]_{t=\bar{t}} = \int_{\partial V} T(\mathbf{n}) dS - \int_V \varrho \nabla_{\mathbf{q}} \Phi_N dV,$$

on ϱ és la densitat de massa, $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ és el *tensor d'esforç*, $\bar{\mathbf{v}} = \frac{d\mathbf{q}}{dt}$, dS és l'element d'àrea, i $\mathbf{n}$ és la normal exterior a la frontera. Tenint en compte que

per a un fluid perfecte es té que $T(\mathbf{n}) = -p\mathbf{n}$, i, a partir del teorema de la divergència de Gauss, l'equació d'Euler es pot escriure com a

$$\left[\frac{d}{dt} \int_{\varphi_t(V)} \varrho \tilde{\mathbf{v}} dV \right]_{t=\bar{t}} = - \int_V (\nabla_{\mathbf{q}} p + \varrho \nabla_{\mathbf{q}} \Phi_N) dV, \quad (16)$$

la forma diferencial de la qual és

$$\partial_t(\varrho \tilde{\mathbf{v}}) + \nabla_{\mathbf{q}} \cdot (\varrho \tilde{\mathbf{v}}) \mathbf{v} + \varrho \tilde{\mathbf{v}} \cdot \nabla_{\mathbf{q}} \tilde{\mathbf{v}} + \nabla_{\mathbf{q}} p + \varrho \nabla_{\mathbf{q}} \Phi_N = 0, \quad (17)$$

o bé

$$\partial_t \tilde{\mathbf{v}} + \tilde{\mathbf{v}} \cdot \nabla_{\mathbf{q}} \tilde{\mathbf{v}} + \frac{1}{\varrho} \nabla_{\mathbf{q}} p + \nabla_{\mathbf{q}} \Phi_N = 0,$$

on hem utilitzat la primera llei de la termodinàmica per a un fluid de pols, és a dir, l'equació de continuïtat $\partial_t \varrho + \nabla_{\mathbf{q}} \cdot (\varrho \tilde{\mathbf{v}}) = 0$.

Cal assenyalar que l'equació (17) és incompatible amb la relativitat especial. Per aquest motiu, la compararem amb la llei de conservació del tensor energia-impuls, $\text{div}_{\mathfrak{g}} \mathfrak{T} = 0$, en l'espaitemps de FLRW pla.

Hem de tenir en compte que, usant coordenades, l'equació de conservació s'escriu $\nabla_{\mu} T_{\nu}^{\mu} = 0$, on amb ∇_{μ} es denota la derivada covariant. Apliquem la fórmula (86.11) de [21]

$$\nabla_{\mu} T_{\nu}^{\mu} = \frac{1}{\sqrt{-g}} \partial_{\mu} (\sqrt{-g} T_{\nu}^{\mu}) - \frac{1}{2} \partial_{\nu} (g_{\mu\alpha}) T^{\mu\alpha},$$

a la mètrica de FLRW $ds^2 = a^2(d\eta^2 - d\mathbf{q}^2)$, on

$$u^{\mu} = \left(\frac{d\eta}{ds}, \frac{d\mathbf{q}}{ds} \right) = \frac{1}{a\sqrt{1 - |\tilde{\mathbf{u}}|^2}} (1, \tilde{\mathbf{u}}) \cong \frac{1}{a} (1, \tilde{\mathbf{u}}).$$

Atès que suposem que la velocitat del fluid és molt menor que la velocitat de la llum, aleshores, $\nabla_{\mu} T_k^{\mu} = 0$, que és equivalent a $\partial_{\mu} (a^4 T_k^{\mu}) = 0$, perquè, pel fet d'estar en un espaitemps homogeni, $\partial_k g_{\mu\nu} = 0$ condueix a

$$\partial_{\eta}((\rho + p)\tilde{\mathbf{u}}) + 4\mathcal{H}(\rho + p)\tilde{\mathbf{u}} + \nabla_{\mathbf{q}} \cdot ((\rho + p)\tilde{\mathbf{u}})\tilde{\mathbf{u}} + (\rho + p)\tilde{\mathbf{u}} \cdot \nabla_{\mathbf{q}} \tilde{\mathbf{u}} + \nabla_{\mathbf{q}} p = 0.$$

Per tant, l'equació clàssica d'Euler (16) en un univers en expansió és compatible amb la relativitat especial i l'expansió de l'univers quan substituïm la densitat de massa ϱ per la funció de *calor per unitat de volum* $(\rho + p)$ i $\tilde{\mathbf{v}}$ per $a\tilde{\mathbf{u}}$.

Així, tenint en compte aquesta substitució, quan s'inclou la gravetat, la forma integral de l'equació d'Euler es converteix en ([14])

$$\left[\frac{d}{a_N d\eta} \int_{\varphi_{\eta}(V)} a_N (\rho + p) \tilde{\mathbf{u}} dV \right]_{\eta=\bar{\eta}} = - \int_V \left(\nabla_{\mathbf{q}} p - \frac{\rho + p}{a_N} \nabla_{\mathbf{q}} a_N \right) dV,$$

amb $dV = a_N^3 dq_1 dq_2 dq_3$ (l'element de volum tridimensional) i on hem fet la substitució $\nabla_{\mathbf{q}}\Phi_N \rightarrow -\frac{1}{a_N}\nabla_{\mathbf{q}}a_N$. I, en forma diferencial, l'equació d'Euler relativista esdevé

$$\begin{aligned} \frac{1}{a_N}\partial_\eta[a_N^4(\rho+p)\bar{\mathbf{u}}] + \bar{\mathbf{u}} \cdot \nabla_{\mathbf{q}}[a_N^3(\rho+p)\bar{\mathbf{u}}] + \\ + a_N^3 \left[(\rho+p)(\nabla_{\mathbf{q}} \cdot \bar{\mathbf{u}})\bar{\mathbf{u}} + \nabla_{\mathbf{q}}p - \frac{\rho+p}{a_N}\nabla_{\mathbf{q}}a_N \right] = 0, \end{aligned}$$

que, utilitzant la derivada total i desconsiderant alguns termes de segon ordre, es pot escriure com a

$$D_\eta((\rho+p)\bar{\mathbf{u}}) + (\rho+p) \left[(4\mathcal{H}_N + \nabla_{\mathbf{q}} \cdot \bar{\mathbf{u}})\bar{\mathbf{u}} - \frac{1}{a_N}\nabla_{\mathbf{q}}a_N \right] + \nabla_{\mathbf{q}}p = 0.$$

Finalment, tinguem en compte que, al primer ordre de pertorbacions, l'equació esdevé

$$\partial_\eta((\rho_0+p_0)\bar{\mathbf{u}}) + 4\mathcal{H}(\rho_0+p_0)\bar{\mathbf{u}} + (\rho_0+p_0)\nabla_{\mathbf{q}}\Phi_N + \nabla_{\mathbf{q}}\delta p = 0,$$

que coincideix amb la linearització de l'equació de conservació del tensor energia-impuls, $\text{div}_{\mathbf{g}}\mathbb{T} = 0$.

6 La mètrica de Schwarzschild a partir de la mecànica newtoniana

Una altra aplicació d'aquesta analogia entre la mecànica newtoniana i la relativitat general és l'obtenció, sense necessitat de resoldre les equacions de camp de la RG, de la famosa mètrica de Schwarzschild, una de les primeres solucions de les equacions d'Einstein, que descriu la mètrica produïda per una partícula puntual de massa M situada a l'origen de coordenades.¹ En la seva forma més coneguda, la mètrica de Schwarzschild és la que van obtenir de manera independent els matemàtics David Hilbert i Johannes Droste [18, 6]:

$$ds^2 = (1 + 2\Phi_N(r)) dt^2 - (1 + 2\Phi_N(r))^{-1} dr^2 - r^2 d\Omega, \quad (18)$$

on $d\Omega = d\theta^2 + \sin^2\theta d\varphi^2$, r representa la distància euclidiana, i $\Phi_N(r) = -\frac{MG}{r}$ és el potencial newtonià creat per una partícula puntual de massa M .

Notem que aquesta mètrica és singular a l'anomenat *radi de Schwarzschild* $r = 2MG$, i, a més a més, no és estàtica per a $r < 2MG$ (el vector ∂_r es torna temporal), tot i que el físic Karl Schwarzschild va obtenir una família de solucions que depenien d'un paràmetre, i va triar intencionadament aquest

¹ Recordem la famosa frase atribuïda a John Archibald Wheeler: «La matèria li diu a l'espai temps com s'ha de corbar, i l'espai temps li diu a les partícules de prova com s'han de moure», que, d'alguna manera, podríem dir que és la realització, a segon ordre (un ordre més de la mecànica newtoniana), del pensament d'Ernst Mach sobre la inèrcia.

paràmetre de manera que la mètrica només fos singular a l'origen de coordenades, on estava situada la partícula ([30]). Det fet, la solució que ell va trobar va ser

$$ds^2 = \left(1 - \frac{\alpha}{(r^3 + \sigma)^{1/3}}\right) dt^2 - \frac{r^4}{(r^3 + \sigma)^{4/3}} \left(1 - \frac{\alpha}{(r^3 + \sigma)^{1/3}}\right)^{-1} dr^2 + (r^3 + \sigma)^{2/3} d\Omega, \quad (19)$$

amb $\alpha = 2MG$ i on σ era un paràmetre lliure, que, per traslladar la singularitat a l'origen (mireu el coeficient que multiplica a dt^2 en l'equació (19)), Schwarzschild va triar igual a $\sigma = \alpha^3$. Noteu que, si triem $\sigma = 0$, obtenim la solució trobada per Hilbert i Droste.

De fet, Schwarzschild, usant coordenades esfèriques, escriu la mètrica, estacionària amb simetria esfèrica, més general possible com a

$$ds^2 = F(r) dt^2 - H(r) dr^2 - G(r) d\Omega,$$

i, resolent les equacions d'Einstein al buit, troba els valors de F , H i R en la forma que es veu a (19).

La diferència amb els treballs de Hilbert i Droste és que, degut a la covariància de les equacions del camp gravitatori, van triar la nova variable $r^* = \sqrt{G(r)}$, amb la qual cosa la mètrica queda de la forma

$$ds^2 = W(r^*) dt^2 - Z(r^*) dr^{*2} - r^{*2} d\Omega$$

en les noves coordenades. I, com va escriure Hilbert en el seu article [18]: «estem igualment justificats a interpretar r^* , θ i φ com a coordenades esfèriques».

Desafortunadament, Schwarzschild morí l'any 1916, durant la Primera Guerra Mundial, en contraure pèmfing mentre era soldat (va ser en el front de guerra on va obtenir la seva solució), i mai no va conèixer la solució obtinguda per Hilbert i Droste. De totes maneres, veurem que la mecànica newtoniana ens condueix a la versió de Hilbert i Droste. A més a més, en mecànica newtoniana, la velocitat d'escapament ve donada per $v_e = \sqrt{\frac{2MG}{r}}$, la qual cosa implica que la llum, amb velocitat igual a 1 en unitats naturals, no pot escapar de l'interior del radi de Schwarzschild. Això, en relativitat, només es compleix si es tria $\sigma = 0$.

Cal destacar que, basant-se en la versió de Hilbert i Droste, l'any 1939 Oppenheimer i Snyder van publicar el seu treball sobre el col·lapse gravitatori, en el qual van considerar un model simplificat, específicament una estrella sense rotació, sense pressió i amb simetria esfèrica [29]. En aquest treball, els autors van establir la relació entre les coordenades d'un observador estàtic i les d'un observador que es mou conjuntament amb la matèria. El fet crucial d'aquest treball és que, per a un observador extern, el temps que triga la superfície de l'estrella a entrar dins del radi de Schwarzschild és infinit, en contrast amb el temps experimentat per un observador que es mou conjuntament amb la matèria. Aquest fenomen es va entendre inicialment com la formació d'un «estel congelat» (cada vegada alienta més la seva contracció i mai no arriba a entrar dins del radi de Schwarzschild) fins a mitjans dels anys seixanta.

Posteriorment va ser rebatejat com la formació d'un «forat negre», primer per J. A. Wheeler, però des d'una perspectiva diferent: la d'un observador que es mou conjuntament amb la matèria que es contreu. Des d'aquest punt de vista, un forat negre es forma en un temps finit, cosa que ha donat lloc a una àmplia literatura dedicada a explorar-ne l'estructura.

Arribats a aquest punt, un es podria preguntar: què és el que observen els astrònoms, estels congelats o forats negres? De fet, els astrònoms no tenen accés directe a les observacions des del punt de vista d'un observador comòbil. El que realment detecten són fenòmens associats a un camp gravitatori extrem, com ara la radiació emesa per matèria que orbita a prop del radi de Schwarzschild. Així, és temptador dir que el que s'observa són els efectes de la formació d'un forat negre, tal com es percep des d'una perspectiva externa. Un aspecte clau en aquesta discussió és la formació del que es coneix com a *horitzó d'esdeveniments*. Tanmateix, una anàlisi detallada d'aquest fenomen implicaria endinsar-nos en consideracions de diversa índole que ens allunyarien de l'objectiu principal d'aquesta secció.

Fet aquest breu resum històric, vegem com obtenir la mètrica (18) utilitzant només conceptes newtonians i una mica de relativitat especial, i, per tant, sense haver de resoldre les complicades equacions de la RG d'Einstein. Comencem per la força d'inèrcia introduïda per Newton als seus *Principia*, és a dir, aquella resistència que tenen els cossos a romandre en repòs o en moviment rectilini respecte d'un suposat «espai absolut». Aquesta s'expressa com a $F_i = -m_i a_c$, on m_i és la massa inercial del cos i a_c la seva acceleració.

D'altra banda, la força d'atracció de la Terra sobre un cos és $F_g = -gm_g$, on m_g és la massa gravitatòria del cos. Arribats a aquest punt, cal destacar que tots els experiments, des de Galileu fins a l'actualitat, ens indiquen que el valor numèric de la massa inercial i la massa gravitatòria és el mateix. Aquesta igualtat es coneix com a *principi d'equivalència feble*, el qual podem expressar a través de la segona llei de Newton com $a_c = -g$ o, equivalentment, com a

$$F_i + F_g = 0 \quad \text{amb} \quad m_i = m_g, \quad (20)$$

que ens diu que la suma de forces —aquí no comptem amb la fricció— que actuen sobre una partícula en caiguda lliure és zero; la força d'inèrcia compensa la gravetat. Dit «a l'estil d'Einstein», una persona en caiguda lliure, i qualsevol objecte que caigui amb ell, no senten el seu pes. Aquest fet, que ens sembla tan natural, és el que va guiar Einstein en la formulació de la seva teoria relativista de la gravetat, i al qual ell mateix va referir-se com «la idea més feliç de la meua vida».

Seguint aquesta idea, considerem la caiguda lliure, en direcció radial, d'un objecte puntual dins un camp amb simetria radial $\Phi_N(r)$, on r torna a ser la distància euclidiana a l'origen. Llavors, la llei (20) s'escriu com a

$$a_c = -\partial_r \Phi_N,$$

on a_c és l'acceleració de l'objecte en caiguda lliure. Ara bé, com calculem l'acceleració a_c ?

Recordem que la relativitat especial ja ens ensenya que el temps no és absolut. Per tant, en quin sistema de referència hem d'aplicar la segona llei de Newton? Per fer-nos-en una idea, tornem a Einstein: l'objecte no sent el seu pes, és a dir, les forces que actuen sobre ell, la gravetat i la inercial, s'anul·len.

Però quina ha de ser la força d'inèrcia que sent l'objecte i que compensa la gravetat? Semblaria que la més natural ha de ser aquella que es calcula a partir del temps propi de l'objecte $F_i = -m \frac{d^2 r}{ds^2}$, ja que és l'objecte el que la sent, i, per tant, l'acceleració que hem d'agafar en la segona llei de Newton és l'acceleració pròpia, l'obtinguda en el sistema de referència de l'objecte en caiguda lliure. Una vegada feta aquesta hipòtesi, la segona llei de Newton s'escriu com a

$$\frac{d^2 r}{ds^2} = -\partial_r \Phi_N. \quad (21)$$

El pas següent és ara trobar la mètrica, les geodèsiques radials de la qual (recordem que l'objecte es mou lliurement, ja que no actua cap força sobre ell) ens donen la llei de Newton (21). Per simetria, la busquem, en coordenades esfèriques, de la forma

$$ds^2 = A(r) dt^2 - B(r) dr^2 - r^2 d\Omega,$$

i, en fer la variació de l'acció relativista $S = \int ds$ respecte de r , s'obté l'equació de la geodèsica radial

$$-\frac{d^2 r}{ds^2} - \frac{1}{2} \partial_r A = 0 \quad \text{amb la condició} \quad A(r)B(r) = 1,$$

i, imposant que la mètrica sigui la de Minkowski a l'infinit, ja que el potencial gravitatori s'anul·la a l'infinit, juntament amb el fet que es compleixi la segona llei de Newton (21), obtenim

$$A(r) = 1 + 2\Phi_N(r) \quad \text{i} \quad B(r) = (1 + 2\Phi_N(r))^{-1}.$$

Veiem, doncs, que, a partir de la mecànica newtoniana i imposant que la força d'inèrcia ve donada per l'acceleració pròpia del cos, es troba la versió de Hilbert i Droste de la mètrica de Schwarzschild, en què les coordenades que s'utilitzen s'han d'interpretar, segons ens acaba de dir la mecànica newtoniana, com a esfèriques, tal com ja va suggerir Hilbert. Observem, doncs, que és la física la que ens proporciona la interpretació adequada de les coordenades. Dit d'una altra manera: les coordenades no tenen un significat independent, desvinculat de la solució de la mètrica. Cada una de les múltiples solucions de la mètrica porta associada la seva pròpia interpretació física de les seves coordenades locals, com ens il·lustra i ens ajuda a comprendre l'exemple de la mètrica de Schwarzschild (vegi's la discussió que es fa a [19]).

6.1 Sobre la precessió del periheli de Mercuri

La llei de la gravitació universal de Newton havia estat una eina extraordinàriament precisa per descriure el moviment dels planetes durant més de dos

segles. No obstant això, a finals del segle XIX, els astrònoms van identificar una discrepància persistent en el moviment del periheli (el punt de màxima proximitat al Sol) de l'òrbita de Mercuri, que no podia ser explicada completament amb la teoria de Newton.

Segons els càlculs basats en la mecànica newtoniana, el periheli de Mercuri havia de precessionar en part a causa de les interaccions gravitacionals amb els altres planetes del sistema solar. Tot i això, les observacions mostraven un excés d'aproximadament 43 segons d'arc per segle que no es podia justificar amb les lleis conegudes.

Durant dècades es van proposar diverses hipòtesis per explicar aquesta discrepància, incloent-hi l'existència d'un nou planeta (anomenat hipotèticament *Vulcà*) o correccions a la llei de Newton. Cap d'aquestes teories, però, no es va demostrar que fos concloent.

Afortunadament, poc després de trobar les seves equacions, i usant perturbacions de segon ordre de les seves equacions, Einstein va trobar que el terme addicional associat a la curvatura de l'espai-temps explicava exactament els 43 segons d'arc per segle observats experimentalment. Això va ser un èxit espectacular, que va convèncer Einstein que la seva teoria era correcta (anys abans del tan famós eclipsi), i va marcar una fita en la història de la ciència. De fet, aquest càlcul basat en aproximacions de segon ordre de les equacions de la RG, i que aportava noves correccions a la mecànica newtoniana, va ser el que va motivar Schwarzschild a buscar una solució exacta de les equacions gravitatòries de la RG, la qual permetia descriure correctament el moviment d'un planeta al voltant del Sol.

Aquí només farem una breu menció de com es pot trobar la precessió del periheli de Mercuri a través de la calibració newtoniana. Comencem amb les anomenades *coordenades harmòniques* ([32])

$$\begin{cases} X_1 = R \sin \theta \cos \varphi, & X_2 = R \sin \theta \sin \varphi, \\ X_3 = R \cos \theta, & t = t, \end{cases}$$

on $R = r - MG$. En aquestes coordenades, la mètrica de Schwarzschild esdevé

$$ds^2 = \left(\frac{1 - MG/R}{1 + MG/R} \right) dt^2 - \left(1 + \frac{MG}{R} \right)^2 d\mathbf{X}^2 - \left(\frac{1 + MG/R}{1 - MG/R} \right) \frac{M^2 G^2}{R^4} (\mathbf{X} \cdot d\mathbf{X})^2,$$

la qual, en l'aproximació $MG \ll R$, i retenint només termes lineals en MG/R , s'escriu com a

$$\begin{aligned} ds^2 &\cong \left(1 - \frac{2MG}{R} \right) dt^2 - \left(1 + \frac{2MG}{R} \right) d\mathbf{X}^2 \\ \Rightarrow ds^2 &\cong \left(1 - \frac{2MG}{r} \right) dt^2 - \left(1 + \frac{2MG}{r} \right) dr^2 - r^2 d\Omega, \end{aligned}$$

que és la calibració newtoniana per a una partícula puntual, i que també dona la mateixa precessió del periheli de Mercuri que la mètrica de Schwarzschild, ja

que ambdues mètriques, en l'aproximació $MG/r \ll 1$, donen lloc a la mateixa equació clau ([32, 24])

$$\left(\frac{d\varphi}{du}\right)^2 = \frac{1 + 2MG(u + u_+ + u_-)}{(u - u_+)(u_- - u)}, \quad (22)$$

on hem introduït la notació $u = 1/r$, $u_+ = 1/r_+$ i $u_- = 1/r_-$, essent r_- la distància del periheli i r_+ la de l'afeli. La integració de l'equació (22) permet calcular la precessió del periheli de Mercuri. (Podeu trobar tots els càlculs detallats en el llibre de Weinberg [32].)

El mateix ocorre si fem servir les anomenades *coordenades isotròpiques*, en les quals la mètrica de Schwarzschild s'expressa de la manera següent:

$$ds^2 = \left(\frac{1 - \frac{MG}{2\tilde{r}}}{1 + \frac{MG}{2\tilde{r}}}\right)^2 dt^2 - \left(1 + \frac{MG}{2\tilde{r}}\right)^4 (d\tilde{r}^2 + \tilde{r}^2 d\Omega),$$

amb $r = \tilde{r}(1 + \frac{MG}{2\tilde{r}})^2$, i, per tant, per $MG/r \ll 1$, tornem a obtenir la relació $\tilde{r} = r - MG$, ara entre coordenades esfèriques i isotròpiques.

Per concloure, convé destacar una observació rellevant: Si considerem la primera equació de Friedmann (10) al buit i busquem solucions estacionàries amb simetria esfèrica, obtindrem

$$\nabla_{\mathbf{q}} \cdot \mathbf{H}_N = 0 \Rightarrow \partial_q \left(\frac{q^2}{a_N} \partial_q a_N \right) = 0,$$

on $q = |\mathbf{q}|$ és el mòdul de $\mathbf{q}$. La solució d'aquesta equació, que, alineant-se amb la mètrica de Minkowski, a l'infinit val 1, és: $a_N = e^{C/q}$. Per a valors grans de q , tindrem $a_N \cong 1 + \frac{C}{q}$, i, per tant, si volem obtenir la mètrica creada per una partícula puntual de massa M , hem de triar $C = MG$.

Finalment, per trobar la relació amb les coordenades esfèriques que apareixen en la mètrica de Hilbert i Droste, és a dir, entre q i la distància euclidiana r , si comparem la part angular —la que multiplica a $d\Omega$ — per $q \geq MG$, s'obté:

$$r = qe^{MG/q} \Rightarrow q \cong r - MG$$

per $MG/q \ll 1$. Així doncs, per a valors grans de r , la relació és la mateixa que hi ha entre les coordenades esfèriques i harmòniques o isotròpiques, i, per tant, amb la mètrica que s'obté a partir de la primera equació de Friedmann, també es troba la precessió del periheli de Mercuri.

Efectivament, considerem un moviment en el pla $\theta = \pi/2$, llavors la mètrica de Hilbert-Droste s'escriu com a

$$ds^2 = a_N^{-2}(r) dt^2 - a_N^2(r) dr^2 - r^2 d\varphi^2,$$

on $a_N^2(r) = \frac{1}{1+2\Phi_N(r)}$. D'altra banda, l'acció relativista es pot escriure com a $S = \int ds = \int \frac{ds^2}{ds^2} ds$, i, per tant, el lagrangà és

$$\mathcal{L} = a_N^{-2}(r) \dot{t}^2 - a_N^2(r) \dot{r}^2 - r^2 \dot{\varphi}^2, \quad \text{amb} \quad \mathcal{L} = 1.$$

Com que el lagrangiana no depèn ni de t ni de φ , tenim dues quantitats que es conserven

$$a_N^{-2}(r)\dot{t} = \sqrt{2\bar{E} + 1} \quad \text{i} \quad r^2\dot{\varphi} = \bar{L},$$

amb la qual cosa la condició $\mathcal{L} = 1$ ens porta a

$$\dot{r}^2 = 2\bar{E} - 2\Phi_N(r) - (1 + 2\Phi_N(r))\frac{\bar{L}^2}{r^2}. \quad (23)$$

Si ara considerem la mètrica conformestàtica

$$ds^2 = a_N^{-2}(q) dt^2 - a_N^2(q) d\mathbf{q}^2, \quad \text{amb} \quad a_N^2(q) = e^{2MG/q} \cong 1 - 2\Phi_N(q),$$

obtindrem

$$\dot{q}^2 = 2E + 1 - a_N^{-2}(q) - a_N^4\frac{L^2}{q^2},$$

i, usant la relació $r^2 = a_N^2 q^2$, s'arriba, en l'aproximació $MG/q \ll 1$, a (23).

7 Algunes reflexions sobre el significat físic de les singularitats matemàtiques que emergeixen en cosmologia

El big-bang, tot i ser una singularitat matemàtica de les equacions d'Einstein (només cal resoldre les equacions homogènies de Friedmann amb una equació d'estat lineal: $p_0 = w\rho_0$ on $w > -1$), sovint s'ha representat de manera inexacta com una explosió literal al començament de l'univers. Per exemple, en el reconegut programa televisiu *Cosmos*, emès als anys vuitanta i presentat per l'astrònom Carl Sagan, l'inici de l'univers es presentava sovint com una colossal explosió. Tanmateix, aquesta representació manca de fonament científic quan s'examina des de la perspectiva de la relativitat general. Recordem que el mateix Albert Einstein va remarcar que, per a densitats d'energia elevades, les seves equacions perden el seu significat [7]. En les seves paraules:

Els dubtes teòrics es basen en el fet que, en el moment inicial de l'expansió, la mètrica esdevé singular i la densitat d'energia, ρ , es torna infinita. En aquest sentit, cal tenir en compte el següent: la teoria actual de la relativitat es basa en una divisió de la realitat física en un camp mètric (gravetat), d'una banda, i en un camp electromagnètic i matèria, de l'altra. En realitat, l'espai probablement tindrà un caràcter uniforme, i la teoria actual només serà vàlida com un cas límit. Per a densitats elevades del camp i de la matèria, les equacions de camp, i fins i tot les variables de camp que hi intervenen, no tindran cap significat real. Per tant, no es pot assumir la validesa de les equacions per a densitats molt altes de camp i matèria, ni concloure que el començament de l'expansió implica necessàriament una singularitat en el sentit matemàtic. El que cal entendre és que les equacions no es poden continuar en aquestes regions. Aquesta consideració, però, no altera el fet que el principi del món constitueix realment un inici des del punt de vista del desenvolupament dels estels i sistemes d'estels actuals, moment en què aquests estels i sistemes encara no existien com a entitats individuals.

Per tant, el que aquesta singularitat matemàtica indica és que, per a un temps còsmic molt proper a $t = 0$, on hem traslladat la singularitat a l'instant $t = 0$, les equacions de la relativitat general no són vàlides per descriure clàssicament l'univers. En aquest sentit, s'assumeix generalment que la relativitat general descriu l'univers només per a escales d'energia o temperatura per sota de la de Planck. Aquesta idea l'expressà clarament Alan Guth, l'inventor del paradigma inflacionari, a [13] quan va escriure:

El model estàndard presenta una singularitat que convencionalment es considera en el temps $t = 0$. Quan $t \rightarrow 0$, la temperatura de l'univers tendeix a l'infinit. Això implica que no es pot definir un problema de valors inicials en $t = 0$. No obstant això, quan la temperatura de l'univers és de l'ordre de la massa de Planck ($m_{\text{pl}} \sim 1.22 \times 10^{19}$ GeV) o superior, les equacions del model estàndard són indubtablement inoperants, ja que s'espera que els efectes gravitacionals quàntics esdevinguin essencials.

Per tant, dins dels límits del nostre coneixement, és raonable iniciar l'escenari del big-bang calent a una temperatura que sigui còmodament inferior a la massa de Planck. En aquest moment, es pot prendre la descripció de l'univers com un conjunt de condicions inicials, i les equacions de moviment descriuen llavors l'evolució posterior. Per descomptat, l'equació d'estat de la matèria a aquestes temperatures no es coneix realment, però es poden fer diverses hipòtesis i explorar-ne les conseqüències.

Ampliant aquesta concepció profundament arrelada d'una creació de l'univers —possiblement heretada, almenys a Occident, d'una tradició paulina—, la representació del big-bang com una explosió dramàtica prové més del desig de transmetre una narrativa captivadora que d'una representació precisa dels principis que regeixen la natura. De fet, el terme *big-bang* va ser encunyat inicialment durant una emissió de ràdio de la BBC l'any 1949 com una etiqueta una mica despectiva per l'astrònom Fred Hoyle, que preferia el seu model d'estat estacionari davant de l'univers en evolució proposat pel físic teòric i sacerdot belga Georges Lemaître. Lemaître va suggerir que l'univers es va expandir des d'un *àtom primigeni* [22]. De fet, com Dirac va explicar a [5], la proposta de Lemaître era que l'univers es va originar a partir d'un únic àtom superradioactiu amb una massa extremadament gran, que es va desintegrar de manera molt abrupta, i es va trencar en fragments, i aquests fragments en peces encara més petites. En paraules del mateix Lemaître:

Necessitem una teoria de l'evolució en forma de focs artificials. Els darrers dos mil milions d'anys són una evolució lenta: són cendres i fum d'uns focs artificials brillants i molt ràpids.

És important destacar que, en aquesta frase, l'ús figuratiu del terme *focs artificials* ha generat malentesos persistents, perpetuats fins avui per narratives populars que presenten l'inici de l'univers com una explosió de matèria localitzada en l'espai exterior. Tanmateix, d'una manera més aclaridora, Lemaître va escriure a [23]:

Si el món ha començat amb un sol quàntum, les nocions d'espai i temps no tindrien cap sentit al principi i només començarien a adquirir algun significat quan el quàntum original s'hagués dividit en un nombre suficient de quàntums. Si aquesta proposta és correcta, l'inici del món va succeir una mica abans de l'inici de l'espai i el temps. Un inici del món d'aquest tipus està prou allunyat de l'ordre actual de la natura, i no és en absolut repugnant.

És evident que la comprensió actual del *big-bang* ha evolucionat significativament, especialment amb la introducció del paradigma inflacionari [13]. La majoria dels cosmòlegs accepten avui que una descripció clàssica de l'univers només és factible a les escales de la gran teoria unificada (GTU), moment en què una expansió accelerada és impulsada per un camp escalar conegut com a *inflató*. Després de la inflació, l'univers experimenta un procés de reescal·lament, ja sigui a través de l'energia alliberada per l'inflató o mitjançant la producció gravitacional (aquí entra la mecànica quàntica, no pas en la seva totalitat quantitzant la gravetat, però sí, de manera semiclàssica, introduint un camp quàntic interaccionant amb la gravetat, la qual li cedeix part de la seva energia per produir partícules supermassives que finalment decauran en les del *model estàndard* [15]) de partícules, per tal d'alinear-se amb l'univers calent de Friedmann.

De fet, no tenim cap experiment a escales properes a la de Planck, en què probablement els efectes quàntics tinguin una importància fonamental, o potser a aquestes escales l'acció d'Einstein-Hilbert presenta no-linealitats en l'escalar de Ricci, i, per tant, les equacions d'Einstein presentaran fortes correccions o simplement s'hauran de canviar en aquesta escala. Tot això són especulacions i van més enllà de la física que coneixem, i és clar que podem fer teories a aquesta escala, com per exemple la teoria quàntica de llaços, la de supercordes i tantes més, i obtenir resultats sorprenents, però això va més enllà del que entenem per física, i, de fet, es podria qualificar com una mena de metafísica que fa ús d'un immens aparell matemàtic. Sovint, aquest aparell resulta tan complex que la rica i intricada estructura de les matemàtiques, amb totes les seves múltiples ramificacions, dificulta, especialment en aquestes teories alternatives que aspiren a transcendir els límits de la relativitat general, l'obtenció d'una visió clara i nítida de la realitat física. Aquesta perspectiva està àmpliament desenvolupada al llibre [8].

Així, adoptant una actitud positivista, sempre es pot argumentar que no cal preocupar-se per les singularitats matemàtiques que apareixen en física, car la mateixa física les resoldrà. Un exemple clar el trobem en l'electrodinàmica quàntica, en què els treballs de Feynman, Schwinger i Tomonaga, guardonats amb el Premi Nobel l'any 1965, van demostrar com eliminar els infinits: primer regularitzant la teoria i després renormalitzant-la, la qual cosa permet obtenir resultats finits i d'acord amb les observacions experimentals.

Aquest precedent ens convida a pensar que les singularitats matemàtiques que poden aparèixer en altres àmbits, com ara en la cosmologia, no són més que indicadors de les limitacions de les teories actuals. Per exemple, les singularitats

associades al big-bang o als forats negres reflecteixen el fet que la relativitat general no és aplicable a escales extremes, on els efectes de la gravetat quàntica haurien de tenir un paper predominant. Així doncs, aquestes singularitats no s'han d'interpretar com una fallida fonamental de la física, sinó com un senyal de la necessitat d'una teoria més completa.

En altres paraules, les singularitats matemàtiques no són més que un símptoma de la inaplicabilitat de les teories físiques actuals a certes escales o condicions (vegi's [9], per a una discussió més detallada). Aquestes situacions, com les associades al big-bang o l'interior dels forats negres, plantegen desafiaments per als físics teòrics, ja que els experiments directes en aquestes escales no són accessibles. Tanmateix, la història de la física ens mostra que les limitacions d'una teoria sovint condueixen al naixement de nous paradigmes científics, obrint la porta a una comprensió més profunda i objectiva de la realitat.

8 Conclusions i una reflexió final

En aquest assaig, s'ha presentat una derivació heurística de les equacions de Friedmann relativistes al primer ordre de pertorbacions, partint dels principis fonamentals de la mecànica newtoniana. Aquest enfocament connecta la mecànica clàssica amb la cosmologia relativista, i ofereix una perspectiva intuïtiva sobre aquestes equacions fonamentals.

Un resultat especialment sorprenent d'aquesta formulació és que el factor d'escala, i en conseqüència el paràmetre de Hubble, no són estrictament homogenis. Aquesta inhomogeneïtat suggereix que les desviacions en els valors mesurats de la taxa de Hubble poden sorgir en funció dels objectes còsmics utilitzats com a referència per a aquestes mesures. Diferents tipus d'objectes, com supernoves, galàxies o quàsars, poden sondejar regions de l'univers amb propietats locals diferents, cosa que condueix a discrepàncies en la taxa d'expansió inferida.

Aquest fet pot tenir implicacions significatives en l'anomenada *tensió de Hubble*, el desacord observat entre el valor actual del paràmetre de Hubble mesurat localment i el seu valor inferit del fons de microones còsmic. Les inhomogeneïtats predites per aquest marc podrien proporcionar una explicació parcial per a aquestes variacions, tot motivant una exploració més profunda del paper de les estructures còsmiques locals en la configuració de la nostra comprensió de l'expansió de l'univers.

D'altra banda, aquest enfocament permet alleugerar la bretxa entre la mecànica clàssica i la relativitat general, vinculant els fonaments intuïtius de la física newtoniana amb els profunds coneixements de la cosmologia moderna mitjançant la senzilla relació entre geometria i matèria representada en l'equació (12). Aquesta tasca no només enriqueix la nostra comprensió teòrica de l'evolució dinàmica de l'univers, sinó que també ofereix una perspectiva unificada sobre com les descripcions clàssiques i relativistes convergeixen en l'estudi dels fenòmens cosmològics.

Voldria acabar amb una reflexió que em sembla adient. A primer cop d'ull, el lector podria estar temptat a pensar que en aquest assaig estic prioritant la mecànica newtoniana sobre la relativitat general. Ans al contrari, la meua guia sempre han estat les equacions linealitzades de la relativitat, i només buscant unes equacions que, en l'aproximació lineal, coincidissin amb les d'Einstein és com s'han pogut trobar les equacions de Friedmann pertorbades. És així com funcionen les ciències naturals: primer es busquen unes equacions que, en una certa aproximació, coincideixin amb les conegudes, i després, si cal, ja es trobarà un principi variacional que les justifiqui. Aquest va ser el punt de vista d'Einstein per trobar les seves equacions per al camp gravitatori. Volia aconseguir unes equacions tensorials, i, per tant, covariants per difeomorfismes passius (canvis de coordenades), que, en primera aproximació, s'alineessin amb les de Newton i Poisson, i va ser només pocs dies —de fet, setmanes— després de presentar-les a l'Acadèmia Prussiana de Ciències, el novembre de 1915, quan Hilbert les va reobtenir a partir d'un principi variacional, ja coneixent-les prèviament.

Agraïments

Aquest treball compta amb el suport de la subvenció espanyola PID2021-123903NB-I00 finançada per MCIN/AEI/10.13039/501100011033 i per «FEDER Una manera de hacer Europa».

Referències

- [1] ABRAHAM, M. «Zur Theorie der Gravitation». *Physik. Zs.*, 13 (1912), 1-5.
- [2] ABRAHAM, M. «Neuere Gravitationstheorien». *Jahrbuch der Radioaktivität und Elektronik* (1915), 470-520.
- [3] ARNOWITT, R.; DESER, S.; MISNER, C. W. «Republication of: The dynamics of general relativity». *Gen. Relativity Gravitation*, 40 (2008), 1997-2027.
- [4] CALLAN, C.; DICKE, R. H.; PEEBLES, P. J. E. «Cosmology and Newtonian Mechanics». *Amer. J. Phys.*, 33 (2) (1965), 105-108.
- [5] DIRAC, P. A. M. «The scientific work of Georges Lemaître». *Comment. Pontificia Acad. Sci.*, 2 (11) (1968), 1-20.
- [6] DROSTE, J. «The field of a single centre in Einstein's theory of gravitation, and the motion of a particle in that field». *Nederl. Akad. Wetensch. Proc.*, 19 (1917), 197-215.
- [7] EINSTEIN, A. *The Meaning of Relativity*. 4a ed. Princeton: Princeton University Press, 1953.
- [8] ELIZALDE, E. *The True Story of Modern Cosmology: Origins, Protagonists and Breakthroughs*. Berlín: Springer, 2021.

- [9] ELIZALDE, E. «Mathematical singularities in the farthest confines of the Universe—and a brief report on its evolutionary history». *Universe*, 9, 33 (2023), 1–24.
- [10] FRIEDMANN, A. «On the curvature of space». *Z. Physik*, 10 (1922), 377–386.
- [11] FRIEDMANN, A. «Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes». *Z. Physik*, 21 (1924), 326–332. [Traduït a: «On the possibility of a world with constant negative curvature of space». *Gen. Relativity Gravitation*, 31 (12) (1999), 2001–2008]
- [12] GIRBAU, J. *Geometria Diferencial i Relativitat*. Bellaterra: Manuals de la Universitat Autònoma de Barcelona, 1993.
- [13] GUTH, A. H. «Inflationary universe: a possible solution to the horizon and flatness problems». *Phys. Rev. D* (3), 23 (2) (1981), 347–356.
- [14] DE HARO, J. «Relations between Newtonian and Relativistic Cosmology». *Universe*, 10, 263 (2024).
- [15] DE HARO, J.; ARESTÉ SALÓ, L.; PAN, S. «Gravitational reheating formulas and bounds in oscillating backgrounds». *Phys. Rev. D*, 110 (12) (2024), 123504.
- [16] DE HARO, J.; ELIZALDE, E. «Topics in cosmology—Clearly explained by means of simple examples». *Universe*, 8 (3), 166 (2022).
- [17] DE HARO, J.; ELIZALDE, E.; PAN, S. «On the perturbed Friedmann equations in Newtonian Gauge». Preprint (2024). [Disponible en línia a: <https://arxiv.org/abs/2412.15139>]
- [18] HILBERT, D. «Die Grundlagen der Physik. (Zweite Mitteilung)». *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1917 (1917), 53–76.
- [19] JOHNS, O. D. «Validity of the Einstein hole argument». *Stud. Hist. Philos. Sci. B Stud. Hist. Philos. Modern Phys.*, 68 (2019), 62–70.
- [20] KLIMCHITSKAYA, G. L.; MOSTEPANENKO, V. M. «Centenary of Alexander Friedmann’s prediction of the universe expansion and the quantum vacuum». *Physics*, 4 (3) (2022), 981–994.
- [21] LANDAU, L. D.; LIFSHITZ, E. M. *The Classical Theory of Fields*. 3a edició en anglès revisada. Oxford: Pergamon Press, 1971. (Course of Theoretical Physics; 2)
- [22] LEMAÎTRE, A. «Contributions to a British Association Discussion on the Evolution of the Universe». *Nature*, 128 (1931), 704–706.
- [23] LEMAÎTRE, G. «The beginning of the world from the point of view of quantum theory». *Nature*, 127 (1) (1931), 706. [Reproduït com a Golden Oldie Editorial a: LUMINET, J.-P. «Editorial note to: Georges Lemaître, The beginning of the world from the point of view of quantum theory». *Gen. Relativity Gravitation*, 43 (10) (2011), 2911–2928]

- [24] MAGNAN, C. «Complete calculations of the perihelion precession of Mercury and the deflection of light by the Sun in General Relativity». Preprint (2007). [Disponible en línia a: <https://arxiv.org/abs/0712.3709>]
- [25] MCCREA, W. H.; MILNE, E. A. «Newtonian universes and the curvature of space». *Q. J. Math.*, 5 (1) (1934), 73–80.
- [26] MUKHANOV, V. *Physical Foundations of Cosmology*. Amb un pròleg d'Andrei Linde. Nova York: Cambridge University Press, 2005.
- [27] NORDSTRÖM, G. «The principle of relativity and gravitation». *Phys. Zeit.*, 13 (1912), 1126–1129.
- [28] NORDSTRÖM, G. «Zur Theorie der Gravitation vom Standpunkt des Relativitätsprinzips». *Ann. Phys.*, 347 (13) (1913), 533–554.
- [29] OPPENHEIMER, J. R.; SNYDER, H. «On continued gravitational contraction». *Phys. Rev. (2)*, 56 (5) (1939), 455–459.
- [30] SCHWARZSCHILD, K. «On the gravitational field of a mass point according to Einstein's theory». *Gen. Relativity Gravitation*, 35 (5) (2003), 951–959. [Traduït de l'article original alemany «Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie». *Sitzungsber. Königl. Preussisch. Akad. Wiss. Berlin Phys. Math. Kl.* (1916), 189–196 per S. Antoci i A. Loinger]
- [31] SYNGE, J. L. *Relativity: The General Theory*. Amsterdam: North-Holland Publishing Co.; Nova York: Interscience Publishers, Inc., 1960. (Series in Physics)
- [32] WEINBERG, S. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*. Nova York: John Wiley & Sons, 1972.

DEPARTAMENT DE MATEMÀTIQUES
 UNIVERSITAT POLITÈCNICA DE CATALUNYA
 C. COLOM, 11, 08222 TERRASSA
jaime.haro@upc.edu

Tensors en estadística algebraica

LUIS SIERRA, MARTA CASANELLAS I PIOTR ZWIERNIK

Resum: En estadística i anàlisi de dades trobem tensors contínuament. Els objectes centrals que vinculen la ciència de dades amb la teoria de tensors i l'àlgebra són els models amb variables latents. En aquest article proporcionem una visió general de la teoria de tensors, posant èmfasi especial en les seves aplicacions en estadística algebraica. Reforcem aquest enfocament amb nombrosos exemples amb l'objectiu d'il·lustrar els conceptes clau. A més, s'inclou una revisió extensa de la literatura per guiar els lectors cap a estudis més detallats sobre la matèria.

Paraules clau: tensors, estadística algebraica, models gràfics, descomposició tensorial.

Classificació MSC2020: 14M99, 15A69, 62H22, 62R01, 62R07.

1 Introducció

Els tensors, com a concepte que generalitza les matrius a dimensions superiors, van ser introduïts a finals del segle XIX per autors com William Rowan Hamilton, Woldemar Voigt o Gregorio Ricci-Curbastro, i es van fer populars amb el treball [69]. Tot i que els tensors van ser desenvolupats per al càlcul diferencial i la geometria, avui en dia tenen un paper crucial en estadística i ciència de dades, ja que ofereixen un marc matemàtic potent per entendre fenòmens complexos, capturar relacions intrincades i extreure intuïcions valuoses de conjunts de dades d'alta dimensió (vegeu el resum [46], per exemple). En aquest article volem mostrar la importància dels tensors en estadística i ciència de dades, centrant-nos en els seus fonaments matemàtics i les connexions amb l'estadística algebraica.

Aplicacions bàsiques. Els tensors tenen aplicacions extenses en diversos àmbits, incloent-hi l'aprenentatge automàtic, la neurociència i la genòmica. En l'aprenentatge automàtic, els tensors faciliten l'anàlisi de dades multidimensionals i permeten tasques com el reconeixement d'imatges, el resum automàtic de vídeos i el processament del llenguatge natural. Per exemple, en

el reconeixement d'imatges s'utilitzen tensors per representar imatges d'alta resolució, capturar relacions espacials i permetre el desenvolupament de models sofisticats d'aprenentatge profund (vegeu [42, 9]).

En neurociència, els tensors tenen un paper crucial per desxifrar els patrons de connectivitat cerebral i comprendre l'organització funcional del cervell (vegeu, per exemple, [28]). Les xarxes cerebrals, representades com a tensors d'ordre superior, capturen les interaccions entre diverses regions cerebrals i proporcionen informació sobre processos cognitius, trastorns neurològics i plasticitat cerebral. Els tensors permeten l'extracció de característiques significatives de les dades d'imatges per ressonància magnètica funcional, cosa que ajuda els investigadors a identificar les regions cerebrals associades a tasques o condicions específiques; vegeu [16].

La genòmica és un altre àmbit en què els tensors ofereixen un marc potent per analitzar dades biològiques d'alta dimensió. En el camp de les dades òmiques, s'empren tensors per integrar diverses característiques moleculars com ara l'expressió gènica, la metilació de l'ADN i les modificacions d'histones (vegeu, per exemple, [94, 60, 74]). Mitjançant la modelització d'aquestes relacions complexes amb tensors, els investigadors són capaços d'identificar signatures moleculars associades a malalties, predir resultats de pacients i orientar estratègies de tractament personalitzades.

Modelització tensorial. La complexitat geomètrica dels tensors planteja reptes en la seva anàlisi i interpretació. Per superar-los s'han desenvolupat mètodes que aprofiten estructures de tensors més simples. Molts dels models aplicats es basen en altres de més bàsics: tensors diagonals i tensors de rang u o de baix rang. Aquesta simplificació millora la interpretabilitat i facilita la computació eficient i la inferència en altes dimensions.

Com veurem, tant els tensors diagonals com els de rang u apareixen de manera natural en el context d'independència de variables aleatòries. En general, molts dels models estadístics que fan servir independència de variables (per exemple, l'anàlisi de components independents, o la independència condicionada) es basen a trobar una descomposició tensorial en termes de tensors de rang inferior.

El camp de l'*estadística algebraica* s'erigeix com el marc adequat per estudiar els tensors dins dels models estadístics (la bibliografia sobre estadística algebraica inclou els llibres [66, 61, 34, 85, 99]). Aprofitant conceptes algebraics, aquesta nova disciplina proporciona nous punts de vista sobre les propietats i el comportament dels tensors en contextos estadístics diversos. Les eines algebraiques s'han aplicat amb èxit per abordar reptes relacionats amb la identificabilitat de models de variables latents, l'estimació de paràmetres, l'ajust de models, el disseny d'experiments i altres problemes estadístics fonamentals.

Variables latents i models gràfics. En ciència de dades, la inclusió de variables *latents* (o *ocultes*, no observades) és una metodologia habitual per a millorar l'expressivitat d'un model sense incrementar dràsticament la seva complexitat

computacional. Malgrat el seu immens potencial, la presència de variables latents introdueix reptes importants en l'estimació i interpretació del model. Per abordar aquests reptes sovint es requereixen algorismes sofisticats i tècniques estadístiques complexes. Això ha constituït una motivació important per al desenvolupament de la teoria d'aprenentatge singular (vegeu [92]) i una gran part dels avenços recents en estadística algebraica se centra en problemes relacionats.

Aprofitant l'estructura algebraica dels tensors, s'ha aconseguit una comprensió més profunda dels models estadístics amb variables latents i s'han desenvolupat algorismes d'estimació eficients. La modelització de la dependència entre les variables aleatòries (tant les observades com les latents) se sol fer mitjançant un *model gràfic*. En un model gràfic, els vèrtexs d'un graf representen variables aleatòries i les seves arestes codifiquen les relacions de dependència condicionada. Quan aquesta representació utilitza arestes dirigides en un graf acíclic, el model també és conegut com a *xarxa bayesiana* o *graf acíclic dirigit*.

Els arbres latents formen la família més manejable de xarxes bayesianes amb variables latents i poden utilitzar-se per modelitzar estructures de dependència quan s'esperen variables de confusió (*confounders*) no observades (vegeu, per exemple, [62, secció 2]). No obstant això, hi ha diverses altres raons per les quals els models d'arbres latents han esdevingut populars en ciències. En primer lloc, els models d'arbres latents engloben un rang més ampli de distribucions de probabilitat que els models d'arbres totalment observats, però mantenen avantatges computacionals, especialment en escenaris d'alta dimensió. A més, l'algorisme *max-product* permet una inferència eficient d'estats no observats. En segon lloc, els arbres poden representar processos evolutius i s'utilitzen com a tals en anàlisi filogenètica (vegeu, per exemple, [80] i la secció 6.1), en lingüística, modelitzant l'evolució del llenguatge (vegeu [70, 76]), i en tomografia de xarxes per inferir estructures d'Internet ([19, 36]). Els models d'arbres latents gaussians (per exemple, els models d'arbres de moviment brownià) capturen correlacions de forma natural per a la tomografia de xarxes gràcies a la disminució de correlacions amb la distància en l'arbre; vegeu la secció 4.3. En tercer lloc, els models d'arbres latents capturen l'estructura jeràrquica en conjunts de dades complexes i estan estretament relacionats amb mètodes de clusterització jeràrquics ([10, 50, 96, 59]). Aquest marc també troba aplicació en visió per computador; vegeu [93] i [24].

Els models d'arbres latents generalitzen els models de Markov ocults (HMM), coincidint amb aquests en arbres eruga ([68]). Els models d'arbres de Markov ocults (HMTM; vegeu [29]) relaxen les restriccions dels HMM i permeten qualsevol estructura d'arbre latent. Les aplicacions inclouen el processament del senyal ([23, 72]), la biomedicina ([52, 65]) i la lingüística ([95]). Aquests models inclouen models filogenètics, models de Bayes primaris, models d'arbres de moviment brownià i un model d'anàlisi factorial, fet que condueix a un marc unificat per a diverses classes de models ([90]).

Un altre exemple important de model de variables latents és l'*anàlisi factorial*, que es fa servir popularment en l'estudi de dades multivariades per

identificar factors latents que influeixen en les variables observades. Aquest camp té aplicacions en finances, psicologia i ciències socials, entre d'altres. Una classe de models relacionats és l'*anàlisi de components principals* (PCA), una tècnica àmpliament utilitzada per a la reducció de la dimensionalitat i la compressió de dades. La versió probabilística, coneguda com a *PCA probabilístic*, introdueix variables latents per tenir en compte el soroll de les dades, i el fa més flexible en aplicacions reals (vegeu la secció 6.3).

Una altra classe important de models gràfics és la de les *màquines de Boltzmann restringides* (RBM; vegeu la secció 6.2), que són components essencials en l'aprenentatge profund, particularment en l'entrenament de xarxes de creença profundes (*deep belief networks*) i altres models generatius. Les RBM involucren variables latents que capturen patrons ocults en les dades i es poden estendre a models gràfics probabilístics amb múltiples capes de variables latents. Han revolucionat camps com la visió per computador i el reconeixement de veu i, des de la perspectiva algebraica, aquesta classe de models ha estat àmpliament estudiada per Montúfar i col·laboradors en [58, 57].

En aquest article expositiu expliquem el paper que tenen els tensors en l'estadística algebraica i en mostrem diversos exemples. A la secció 2 presentem tensors que apareixen de manera natural en estadística i proporcionem els primers resultats que donen resposta a preguntes estadístiques en termes de propietats algebraiques dels tensors. A la secció 3 aprofundim en els models gràfics i presentem exemples de models gràfics dirigits i no dirigits, amb variables discretes o contínues, fent èmfasi especial en els models d'arbres latents. La secció 4 la dediquem al cas de variables gaussianes, que requereix un tractament específic. Basant-nos en els exemples anteriors, a la secció 5 il·lustrem els conceptes de rang tensorial i descomposició tensorial tot explorant el mètode dels moments en aquest context. Finalment, la secció 6 està dedicada a aplicacions addicionals dels tensors en estadística algebraica, en què ens centrem en l'anàlisi filogenètica, les màquines de Boltzmann restringides, i l'anàlisi de components independents.

2 Primers exemples i definicions

En aquesta secció introduïm breument els principals objectes d'estudi: tensors, tensors simètrics, tensors diagonals, tensors de rang u, i els seus noms corresponents en estadística. Ens centrem sobretot en el cas de tensors discrets per a aquesta secció, però alguns resultats també valen per al cas general. Les referències bàsiques per a aquests conceptes inclouen [48, 99, 13, 56].

Una matriu real $r_1 \times r_2$ és una taula dos dimensional, $A = (a_{i_1, i_2})$, on $i_1 \in \{0, \dots, r_1 - 1\}$, $i_2 \in \{0, \dots, r_2 - 1\}$, i $a_{i_1, i_2} \in \mathbb{R}$. Donats $m \in \mathbb{N}$ i enters $r_1, \dots, r_m \geq 1$, una *matriu* $r_1 \times \dots \times r_m$ -dimensional és una col·lecció de nombres reals $a_{i_1, \dots, i_m}$ on $i_j \in \{0, \dots, r_j - 1\}$ per a $j = 1, \dots, m$. Com explicarem a continuació, aquestes matrius multidimensionals s'identifiquen de manera natural amb els *tensors*.

Començant pel cas de les matrius usals 2-dimensionals, sigui $\{e_i^j\}_{i=0, \dots, r_j-1}$ la base canònica a $\mathbb{R}^{r_j}$, $j = 1, \dots, m$. Una matriu A es pot identificar amb una

aplicació bilinear de $\mathbb{R}^{r_1} \times \mathbb{R}^{r_2}$ en $\mathbb{R}$, $(x, y) \mapsto x^T A y$ (on l'element (e_i^1, e_j^2) s'envia a $a_{i,j}$). Si $E^{i,j}$ denota la matriu formada per zeros excepte un 1 en l'entrada en la posició (i, j) , la seva aplicació bilinear corresponent es denota amb el tensor $e_i^1 \otimes e_j^2$. D'aquesta manera, una matriu A com la d'abans es correspon amb el tensor $\sum a_{i,j} e_i^1 \otimes e_j^2$ en l'espai vectorial de tensors $\mathbb{R}^{r_1} \otimes \mathbb{R}^{r_2}$ (pensat com a l'espai de formes bilineals en $\mathbb{R}^{r_1} \times \mathbb{R}^{r_2}$).

Donats dos vectors $v^1 = (v_0^1, \dots, v_{r_1-1}^1) \in \mathbb{R}^{r_1}$ i $v^2 = (v_0^2, \dots, v_{r_2-1}^2) \in \mathbb{R}^{r_2}$, el seu *producte tensorial* es defineix com el tensor $v^1 \otimes v^2 = \sum_{i_1, i_2} v_{i_1}^1 v_{i_2}^2 e_{i_1}^1 \otimes e_{i_2}^2 \in \mathbb{R}^{r_1} \otimes \mathbb{R}^{r_2}$. Per exemple, si $v^1 = (2, 0) \in \mathbb{R}^2$ i $v^2 = (1, 1) \in \mathbb{R}^2$, aleshores $v^1 \otimes v^2 = 2e_0^1 \otimes e_0^2 + 2e_0^1 \otimes e_1^2$ és un tensor diferent de $v^2 \otimes v^1 = 2e_0^1 \otimes e_0^2 + 2e_1^1 \otimes e_0^2$.

Més generalment, un *tensor* $r_1 \times \dots \times r_m$ sovint es defineix com l'aplicació multilinear $\mathbb{R}^{r_1} \times \dots \times \mathbb{R}^{r_m} \rightarrow \mathbb{R}$: el tensor $e_{i_1}^1 \otimes \dots \otimes e_{i_m}^m$ es correspon amb l'aplicació multilinear

$$\begin{aligned} \mathbb{R}^{r_1} \times \dots \times \mathbb{R}^{r_m} &\rightarrow \mathbb{R} \\ (e_{j_1}^1, \dots, e_{j_m}^m) &\mapsto \begin{cases} 1 & \text{si } (j_1, \dots, j_m) = (i_1, \dots, i_m), \\ 0 & \text{altrament} \end{cases} \end{aligned}$$

i aquests tensors formen una base natural de l'espai vectorial $\mathbb{R}^{r_1} \otimes \dots \otimes \mathbb{R}^{r_m}$ dels tensors $r_1 \times \dots \times r_m$. En aquest sentit, el tensor $t = \sum_{i_1, \dots, i_m} a_{i_1, \dots, i_m} e_{i_1}^1 \otimes \dots \otimes e_{i_m}^m$ en aquest espai es correspon amb la matriu multidimensional $A = [a_{i_1, \dots, i_m}]_{i_j=0, \dots, r_j-1}$ de dimensions $r_1 \times \dots \times r_m$. Quan parlem de coordenades d'un tensor, ens referim a les seves coordenades en la base natural.

Com abans, donats m vectors

$$v^1 = (v_0^1, \dots, v_{r_1-1}^1) \in \mathbb{R}^{r_1}, \dots, v^m = (v_0^m, \dots, v_{r_m-1}^m) \in \mathbb{R}^{r_m},$$

el seu *producte tensorial* és el tensor $v^1 \otimes \dots \otimes v^m = \sum_{i_1, \dots, i_m} v_{i_1}^1 v_{i_2}^2 \dots v_{i_m}^m e_{i_1}^1 \otimes \dots \otimes e_{i_m}^m$.

Diem que un tensor T de $\otimes^r \mathbb{R}^n = \mathbb{R}^n \otimes \dots \otimes \mathbb{R}^n$ és *simètric* si $T_{i_1 \dots i_r} = T_{j_1 \dots j_r}$ per a tota permutació $(j_1, \dots, j_r)$ de $(i_1, \dots, i_r)$. L'espai de tensors simètrics de $\otimes^r \mathbb{R}^n$ es denota per $S^r(\mathbb{R}^n)$. La dimensió de $S^r(\mathbb{R}^n)$ és $\binom{n+r-1}{r}$, donat que un tensor $T \in S^r(\mathbb{R}^n)$ es pot codificar amb les entrades $T_{i_1 \dots i_r}$, on $1 \leq i_1 \leq \dots \leq i_r \leq n$.

2.1 Mesures discretes com a tensors

Considerem una col·lecció de m variables aleatòries discretes, cadascuna amb r_i possibles estats, $X_i \in \mathcal{X}_i = \{0, \dots, r_i - 1\}$ amb $i = 1, \dots, m$. El vector aleatori $X = (X_1, \dots, X_m)$ pren valors a $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_m$ i una distribució de probabilitat $p = (p(x))_{x \in \mathcal{X}}$ de X es pot identificar amb un punt a

$$\mathbb{R}^{\mathcal{X}} := \mathbb{R}^{\mathcal{X}_1} \otimes \dots \otimes \mathbb{R}^{\mathcal{X}_m},$$

on $\mathbb{R}^{\mathcal{X}_i}$ és una còpia de $\mathbb{R}^{r_i}$ amb coordenades indexades per $\mathcal{X}_i$. Les coordenades de l'espai tensorial $\mathbb{R}^{\mathcal{X}}$ es denoten amb p_x , $x \in \mathcal{X}$. Per exemple, si $m = 2$,

$r_1 = r_2 = 2$, aleshores cada X_i és binària i les coordenades de $p \in \mathbb{R}^X$ són p_{00} , p_{01} , p_{10} , p_{11} , $p = p_{00}e_0 \otimes e_0 + p_{01}e_0 \otimes e_1 + p_{10}e_1 \otimes e_0 + p_{11}e_1 \otimes e_1$ (ometem els superíndexs perquè en aquest cas els dos espais són el mateix).

Per definició, totes les distribucions de probabilitat discretes es poden representar al simplex $\Delta^X \subseteq \mathbb{R}^X$, on

$$\Delta^X := \left\{ p \in \mathbb{R}^X : p_x \geq 0, \sum_{x \in X} p_x = 1 \right\}.$$

Qualsevol model estadístic per a X és, per definició, una família de distribucions de probabilitat i, per tant, una família de punts de Δ^X . Això dona una identificació bàsica dels models estadístics discrets amb subconjunts de l'espai tensorial $\mathbb{R}^X$.

Quan es fan consideracions geomètriques referents als tensors, sol ser més convenient treballar sobre l'espai projectiu. Denotem amb $\text{Proj}(V)$ la projectivització d'un espai vectorial V i sigui $\mathbb{P}_{\mathbb{R}}^X := \text{Proj}(\mathbb{R}^X)$. Per definició, aquest és el conjunt de punts $\mathbb{R}^X \setminus \{0\}$, identificant aquells que es troben sobre la mateixa recta que passa per l'origen. Observem que el simplex de probabilitat Δ^X està en correspondència amb la part no negativa de $\mathbb{P}_{\mathbb{R}}^X$. En efecte, la bijecció envia un punt $q \in \mathbb{P}_{\mathbb{R}}^X$ amb coordenades q_x no negatives a $p = (p_x)_{x \in X} \in \Delta^X$, on $p_x = q_x / \sum_{y \in X} q_y$.

La immersió en el projectiu és sovint convenient perquè molts models de dependències complexes sovint estan definits mòdul una constant de normalització (vegeu, per exemple, els models definits a (2)).

2.2 Model d'independència i tensors de rang u

Donada una col·lecció de variables discretes $X = (X_1, \dots, X_m) \in \mathcal{X}$, diem que les components de X són independents (o que X pertany al *model d'independència total*) si el tensor $p \in \Delta^X$ corresponent a la distribució de X es pot escriure com el producte tensorial de les distribucions individuals $p^i \in \mathbb{R}^{X_i}$, $p = p^1 \otimes \dots \otimes p^m$. En altres paraules,

$$p_{i_1 i_2 \dots i_m} = p_{i_1}^1 p_{i_2}^2 \dots p_{i_m}^m, \quad \text{per tot } x = (i_1, \dots, i_m) \in X. \quad (1)$$

Per tenir un equivalent geomètric d'aquesta construcció, diem que $p \in \mathbb{R}^X$ és un *tensor de rang u* (o un *tensor pur o descomponible*) si és el producte tensorial d'una m -tupla de vectors $p^i \in \mathbb{R}^{X_i}$ per a $i = 1, \dots, m$. El model d'independència per al vector X està contingut en el conjunt de tensors de rang u. Quan $m = 2$, els tensors de rang u es corresponen amb les matrius de rang u. En efecte, $p = p^1 \otimes p^2$ és el tensor $\sum_{i,j} p_i^1 p_j^2 e_i^1 \otimes e_j^2$, que, amb la correspondència introduïda a dalt, es pot identificar amb la matriu de rang u

$$\begin{pmatrix} p_0^1 \\ \vdots \\ p_{r_1-1}^1 \end{pmatrix} \begin{pmatrix} p_0^2 & \dots & p_{r_2-1}^2 \end{pmatrix}.$$

En el context de la geometria algebraica projectiva, el conjunt de tensors de rang u es correspon amb la varietat de Segre real

$$\text{Seg}(\mathbb{P}_{\mathbb{R}}^{X_1} \times \cdots \times \mathbb{P}_{\mathbb{R}}^{X_m}) \subset \mathbb{P}_{\mathbb{R}}^X = \text{Proj}(\mathbb{R}^{X_1} \otimes \cdots \otimes \mathbb{R}^{X_m});$$

vegeu [13, 10.5.12]. Per exemple, $\text{Seg}(\mathbb{P}_{\mathbb{R}}^1 \times \mathbb{P}_{\mathbb{R}}^1)$ és la immersió de $\mathbb{P}_{\mathbb{R}}^1 \times \mathbb{P}_{\mathbb{R}}^1$ a $\mathbb{P}_{\mathbb{R}}^3 = \text{Proj}(\mathbb{R}^2 \otimes \mathbb{R}^2)$ donada per l'aplicació

$$\begin{aligned} \mathbb{P}_{\mathbb{R}}^1 \times \mathbb{P}_{\mathbb{R}}^1 &\longrightarrow \mathbb{P}_{\mathbb{R}}^3 \\ ([p_0^1 : p_1^1], [p_0^2 : p_1^2]) &\longmapsto [p_0^1 p_0^2 : p_0^1 p_1^2 : p_1^1 p_0^2 : p_1^1 p_1^2]. \end{aligned}$$

Aquesta varietat està formada pels punts $[p_{00} : p_{01} : p_{10} : p_{11}]$ a $\mathbb{P}_{\mathbb{R}}^3$ que satisfan l'equació $p_{00}p_{11} - p_{01}p_{10} = 0$ i es correspon amb els tensors de rang u , $p^1 \otimes p^2$, o matrius

$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$$

de rang u . Anàlogament, si pensem els elements de $\mathbb{R}^{X_1} \otimes \mathbb{R}^{X_2}$ com a matrius, $\text{Seg}(\mathbb{P}_{\mathbb{R}}^{X_1} \otimes \mathbb{P}_{\mathbb{R}}^{X_2})$ és el conjunt de matrius $r_1 \times r_2$ de rang u (definides pels menors 2×2 nuls). De fet, es pot provar fàcilment que la part no negativa de $\text{Seg}(\mathbb{P}_{\mathbb{R}}^{X_1} \times \cdots \times \mathbb{P}_{\mathbb{R}}^{X_m})$ és isomorfa al model d'independència total.

2.3 Tensors de moments - cas discret

Considerem les mesures discretes de la secció 2.1. Donats $u = (u_1, \dots, u_m) \in \mathbb{N}^m$ i un vector $x = (x_1, \dots, x_m) \in X$, definim el monomi

$$x^u := x_1^{u_1} \cdots x_m^{u_m},$$

on fem servir la convenció $0^0 = 1$. Si X té distribució p , aleshores el *moment* corresponent és

$$\mu_u = \mathbb{E}[X^u] = \sum_{x \in X} p_x x^u$$

i es diu que és un *moment d'ordre k* si $k = u_1 + \cdots + u_m$. En particular, $\mu_{0 \dots 0} = 1$.

EXEMPLE 1. Considerant el cas binari $(X_1, X_2) \in \{0, 1\}^2$, es té $\mu_{00} = 1$, $\mu_{10} = p_{10} + p_{11}$, $\mu_{01} = p_{01} + p_{11}$, $\mu_{11} = p_{11}$.

PROPOSICIÓ 1. L'aplicació $\mathbb{R}^X \rightarrow \mathbb{R}^X$ definida com a

$$\mu_u = \mathbb{E}[X^u], \quad \text{per a tot } u \in X,$$

és una bijecció lineal que envia el conjunt $\{\sum_{x \in X} p_x = 1\}$ al conjunt $\{\mu_{0 \dots 0} = 1\}$.

Hi ha altres maneres de codificar una distribució de probabilitat discreta que no sigui només amb els seus moments. Un exemple important és el dels *cumulants*. Des del punt de vista geomètric, tractar amb cumulants correspon a fer un canvi de variables (no lineal) que pot ser més adient per estudiar la geometria de segons quins models; vegeu [83, 98, 25]. Si considerem l'exemple 1, podem emprar la transformació en cumulants: $\kappa_{00} = 0$, $\kappa_{10} = \mu_{10}$, $\kappa_{01} = \mu_{01}$, i $\kappa_{11} = \mu_{11} - \mu_{10}\mu_{01}$. En aquest cas la imatge del model d'independència per a aquesta aplicació és el subespai lineal donat per $\kappa_{11} = 0$.

3 Models gràfics

Molts models en estadística i aprenentatge automàtic involucren modelitzar la dependència entre diverses components d'un vector aleatori. Entre els múltiples exemples trobem l'anàlisi factorial, les màquines de Boltzmann restringides, o el model de Bayes primari. Els models gràfics constitueixen el llenguatge natural per a aquestes tasques de modelització multivariant, ja que representen les relacions de dependència mitjançant grafs. En aquest context, els vèrtexs d'un graf representen variables aleatòries i les arestes codifiquen les relacions de dependència condicionada entre elles. Segons el tipus d'arestes, podem distingir dos tipus principals de models gràfics: els que tenen arestes dirigides i els que no, cadascun adient per a diferents tasques.

Per a ambdós tipus de models gràfics és convenient començar per la noció de *distribució factoritzada*. Donat un vector aleatori $X = (X_1, \dots, X_m)$ que pren valors a $\mathcal{X}$ com abans i donat $\mathcal{F}$ un conjunt de subconjunts de $[m]$, podem considerar el conjunt de distribucions sobre $\mathcal{X}$ que satisfan

$$p(x) = \frac{1}{Z} \prod_{F \in \mathcal{F}} \phi_F(x_F), \quad \text{per tot } x \in \mathcal{X}, \quad (2)$$

on ϕ_F són funcions no negatives sobre $\mathbb{R}^{X_F}$ (anomenades *potencials*) i Z és la constant normalitzadora.

Un exemple trivial és el de (1), on la distribució factoritza en components individuals. En aquest cas, els factors es corresponen amb distribucions marginals, però aquesta propietat no és certa en general.

EXEMPLE (MODEL D'ISING). El model d'Ising considera distribucions a $\mathcal{X} = \{0, 1\}^m$ definides per

$$p(x) = \frac{1}{Z} \prod_{i=1}^m e^{h_i x_i} \prod_{i < j} e^{J_{ij} x_i x_j} = \frac{1}{Z} e^{h^\top x + \frac{1}{2} x^\top J x},$$

on $h = [h_i] \in \mathbb{R}^m$, $J = [J_{ij}] \in \mathbb{R}^{m \times m}$, és simètrica amb zeros a la diagonal, i Z és la constant normalitzadora

$$Z = \sum_{x \in \{0,1\}^n} e^{h^\top x + \frac{1}{2} x^\top J x}.$$

Es tracta, per tant, d'una distribució factoritzada. En moltes aplicacions d'aquest model, J té suport en un graf G (i en aquest cas, $J_{ij} = 0$, sempre que no hi hagi una aresta connectant i amb j a G). Des del punt de vista algebraic, aquest model es coneix com a *model gràfic binari* i ha estat estudiat a [31].

Si $\mathcal{X}$ és finit com a la secció 2.1, està clar que (2) imposa restriccions algebraiques sobre el tensor de probabilitat $p \in \mathbb{R}^{\mathcal{X}}$. Més endavant estudiarem algunes de les formes que poden prendre aquestes restriccions. En aquest punt només val la pena assenyalar que aquestes propietats de factorització són equivalents als models de xarxa de tensors, àmpliament considerats en aplicacions industrials diverses. L'article [71] explora aquesta connexió amb més detall.

3.1 Grafs acíclics dirigits

Sigui $\mathcal{G} = (V, E)$ un graf dirigit. Si $(a, b) \in E$ és una aresta del graf, també designada per $\{a \rightarrow b\}$, diem que a és un *predecessor* de b , i b un *successor* de a . Un graf dirigit $\mathcal{G}$ és acíclic (breument DAG) si no conté cicles dirigits. Si $v \in V$ és un vèrtex d'un DAG $\mathcal{G}$, es denota amb $\mathbf{pa}(v)$ el conjunt dels *predecessors* de v i amb $\mathbf{de}(v)$ el conjunt de *descendents* de v format pels vèrtexs w tals que existeix un camí dirigit $\{v \rightarrow w\} \subseteq \mathcal{G}$. Contràriament, també podem definir $\mathbf{nd}(v)$ com el conjunt de *no descendants* $V \setminus (\{v\} \cup \mathbf{de}(v))$.

Donat un DAG $\mathcal{G} = (V, E)$ amb $V = \{1, \dots, m\}$, assignem una variable aleatòria X_v a cada node $v \in V$. El model probabilístic associat al vector aleatori $X = (X_1, \dots, X_m)$ i al DAG $\mathcal{G}$, també anomenat *xarxa bayesiana*, pren valors a $\mathcal{X} = \prod_{i=1}^m \mathcal{X}_i$ i conté totes les distribucions sobre $\mathcal{X}$ que satisfan la factorització recursiva següent:

$$p(x) = \prod_{i=1}^m p(x_i | \mathbf{pa}(x_i)), \quad (3)$$

on, per a $X_C = \mathbf{pa}(X_i)$, tenim que $p(x_i | x_C)$ denota la funció de densitat de la probabilitat condicionada de X_i donat $X_C = x_C$, especificada per

$$p(x_i | \mathbf{pa}(x_i)) = \frac{p(x_i, \mathbf{pa}(x_i))}{p(\mathbf{pa}(x_i))},$$

si $p(\mathbf{pa}(x_i)) > 0$.

NOTA 1. Per simplificar la notació, utilitzem la mateixa lletra per denotar totes les densitats condicionades. Per evitar confusions, un estat denotat en minúscula amb el mateix nom que una variable aleatòria (en majúscula) es refereix a tal observació de la variable aleatòria.

En aquest article ens centrem en la situació en què el vector aleatori X és exclusivament o bé discret, o bé continu, però també són d'interès les situacions mixtes; vegi's [49].

Donades tres variables aleatòries X, Y, Z , diem que X és independent de Y donada Z (denotat amb $X \perp\!\!\!\perp Y|Z$), si la distribució condicionada satisfà $p(x, y|z) = p(x|z)p(y|z)$ per a tot x, y, z (o, equivalentment, $p(x|y, z) = p(x|z)$). La factorització (3) permet una interpretació important en termes d'independència condicionada coneguda com a *propietat de Markov dirigida local*:

$$X_v \perp\!\!\!\perp X_{\text{nd}(v) \setminus \text{pa}(v)} \mid X_{\text{pa}(v)}, \quad v \in V. \quad (4)$$

EXEMPLE 2. Considerem l'estructura d'una *cadena de Markov*, en què les variables aleatòries $X_1, \dots, X_m$ satisfan la propietat $p(x_k|x_{k-1}, \dots, x_1) = p(x_k|x_{k-1})$ per a tot $k = 2, \dots, m$ (colloquialment l'interpretem com que el futur només es veu afectat pel present immediat, però no pel passat). Degut a aquesta propietat d'independència condicionada, una cadena de Markov es pot representar com un model probabilístic sobre un graf línia en què cada node representa un estat (modelitzat com una variable aleatòria) X_k i X_{k-1} és el seu únic predecessor directe, donat que X_k és independent del «passat» quan coneixem X_{k-1} . Si $m = 3$, el graf associat és el que il·lustrem a la figura 1(a).

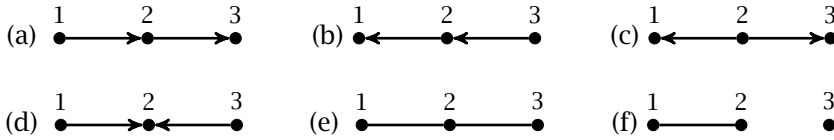


FIGURA 1: Diversos grafs sobre tres nodes.

L'anàlisi de models DAG pot portar a observacions inesperades. Per exemple, el graf (a) de la figura 1 defineix exactament el mateix model que els grafs (b) i (c). En particular, sense hipòtesis addicionals, no existeix una bijecció entre models en DAG i els seus grafs corresponents [89, 21].

EXEMPLE (COLLISIONADORS). Considerem un vector aleatori format per tres variables aleatòries discretes X_1, X_2, X_3 , en què el graf de dependència està donat per la figura 1(d). En aquest model, la funció de densitat es pot factoritzar com a $p(x_1, x_2, x_3) = p(x_2|x_1, x_3)p(x_1)p(x_3)$. Observeu que la distribució marginal de (X_1, X_3) es factoritza com a $p(x_1, x_3) = p(x_1)p(x_3)$ i, per tant, aquestes dues variables són independents. No obstant això, en condicionar X_1, X_3 respecte de $X_2 = x_2$, podem observar que es trenca la propietat d'independència, ja que

$$p(x_1, x_3|x_2) = \frac{p(x_2|x_1, x_3)p(x_1)p(x_3)}{p(x_2)}.$$

Així doncs, tot i que $X_1 \perp\!\!\!\perp X_3$, tenim que $X_1 \not\perp\!\!\!\perp X_3|X_2 = x_2$. Intuïtivament, el fenomen es correspon amb la idea que, malgrat que dos progenitors siguin (genèticament) independents, si observem un fill en comú, tot el que no vingui explicat per un dels progenitors haurà de ser explicat per l'altre, i, per tant,

deixen de ser independents. Aquesta configuració s'anomena *col·lisionador* (*collider* en anglès) i és una coneguda font de confusió en inferència causal, també coneguda com a *paradoxa de Berkson*.

EXEMPLE (DAG ESTRELLA). Considerem un DAG senzill sobre quatre nodes amb arestes dirigides $(0, 1)$, $(0, 2)$ i $(0, 3)$ com el de la figura 2. El model probabilístic associat correspon a la factorització

$$p(x) = p(x_0)p(x_1|x_0)p(x_2|x_0)p(x_3|x_0), \quad \text{per a tot } x \in \mathcal{X}.$$

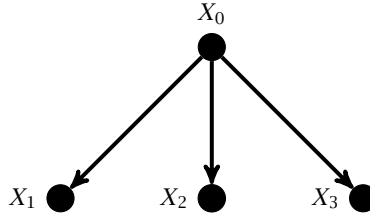


FIGURA 2: Model gràfic sobre quatre nodes amb un predecessor i tres descendents.

Suposem que $\mathcal{X} = \{0, 1\}^4$. Cada distribució condicionada $p(x_i|x_0)$, $i = 1, 2, 3$, es pot representar per una *matriu estocàstica* 2×2 (una matriu no negativa en què totes les files sumen 1)

$$A^i = \begin{pmatrix} p(x_i = 0|x_0 = 0) & p(x_i = 1|x_0 = 0) \\ p(x_i = 0|x_0 = 1) & p(x_i = 1|x_0 = 1) \end{pmatrix}.$$

Denotant amb $\pi_x = p(X_0 = x)$ i $a_{x,y}^i$ l'entrada (x, y) -èsima de la matriu A_i , veiem que el model està parametritzat per

$$p_{x_0x_1x_2x_3} = \pi_{x_0} a_{x_0,x_1}^1 a_{x_0,x_2}^2 a_{x_0,x_3}^3. \quad (5)$$

Observem que la dimensió de l'espai dels catorze paràmetres $\pi_x, a_{x,y}^i$ és set perquè les sumes de les files de A^i és u i $\pi_0 + \pi_1 = 1$. De l'equació (5) és senzill veure que, fixant l'estat a X_0 igual a i , les matrius

$$\begin{pmatrix} p_{i000} & p_{i001} \\ p_{i010} & p_{i011} \\ p_{i100} & p_{i101} \\ p_{i110} & p_{i111} \end{pmatrix}, \quad \begin{pmatrix} p_{i000} & p_{i010} \\ p_{i001} & p_{i011} \\ p_{i100} & p_{i110} \\ p_{i101} & p_{i111} \end{pmatrix}, \quad \begin{pmatrix} p_{i000} & p_{i100} \\ p_{i001} & p_{i101} \\ p_{i010} & p_{i110} \\ p_{i011} & p_{i111} \end{pmatrix}$$

són de rang u. D'aquí, podem verificar fàcilment les condicions

$$p_{ijkl}p_{ij'k'l'} = p_{ijk'l}p_{ij'k'l'} = p_{ijk'l}p_{ij'k'l'} = p_{ij'kl}p_{ijk'l'}, \quad \text{per a } i, j, j', k, k' \in \{0, 1\}.$$

D'aquesta manera, les condicions d'independència es tradueixen en relacions polinomials entre les entrades del vector de probabilitats p . En termes de geometria algebraica, podem pensar en la factorització (5) com l'aplicació polinomial

$$\varphi: \mathbb{R}^7 \rightarrow \otimes^4 \mathbb{R}^2,$$

que envia cada conjunt de set paràmetres lliures a la distribució p . Tant des del punt de vista estadístic com de geometria algebraica, és d'interès conèixer un conjunt generador minimal d'equacions polinomials que defineixen el model (és a dir, que s'anul·lin precisament en la imatge de φ). En aquest cas, és conegut que l'ideal de polinomis que s'anul·len a $\text{Im}(\varphi)$ és generat minimalment per divuit equacions quadràtiques de les que hem llistat a dalt; vegeu [40].

Aquests són el tipus de preguntes que es troben en el centre de l'estadística algebraica: per models d'estadística algebraica, utilitzar tècniques de geometria algebraica i àlgebra commutativa per trobar equacions en polinomis i restriccions semialgebraiques que defineixin el model. Conèixer-les pot ser útil per esbrinar si els paràmetres són identificables o no, per estimar els paràmetres, per inferència de versemblança i per ajustar el model.

3.2 Models gràfics no dirigits

Sigui $\mathcal{G}$ un graf no dirigit i sigui C el conjunt de cliques (subgrafs complets) maximals de $\mathcal{G}$. Un *model gràfic no dirigit* és un conjunt de distribucions de probabilitat de tipus

$$f(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C), \quad \text{on } Z = \int_{\mathcal{X}} \prod_{C \in \mathcal{C}} \psi_C(x_C) dx, \quad (6)$$

i $\psi_C: \mathcal{X}_C \rightarrow [0, \infty)$ es coneixen com a *funcions potencials* degut a la seva interpretació en el camp de la física estadística. Aquest és un cas particular de l'expressió a (2) prenent com a $\mathcal{F}$ el conjunt C de cliques de $\mathcal{G}$.

Aquesta definició pot semblar una mica allunyada de les nocions d'independència condicionada que hem descrit per als grafs dirigits, però el 1971, en un article no publicat, John Hammersley i Peter Clifford varen demostrar que, sota la condició $p(x) > 0$ per a tot $x \in \mathcal{X}$, la factorització (6) en un graf no dirigit $\mathcal{G}$ és equivalent a les *propietats globals de Markov*: per a qualssevol subconjunts $A, B, C \subset V$ es té $X_A \perp\!\!\!\perp X_B | X_C$ sempre que C separi A, B en $\mathcal{G}$ (és a dir, que tots els camins entre A i B passin per C).

Reprenem el cas de variables discretes de la secció 2.1. Sigui $i_C := (i_j)_{j \in C} \in \mathcal{X}_C$ un estat del vector X_C per una clica $C \subseteq V$. Llavors, la funció potencial $\psi_C: \mathcal{X}_C \rightarrow [0, \infty)$ es pot veure com un vector $\theta^{(C)}$ amb entrades $\theta_{i_C}^{(C)} = \psi_C(i_C) \in [0, \infty)$. Tal com vam veure a l'exemple DAG estrella, això ens dona la parametrització

$$p_{i_1 i_2 \dots i_n} = \frac{1}{Z(\theta)} \prod_{C \in \mathcal{C}} \theta_{i_C}^{(C)}. \quad (7)$$

D'aquesta manera, aconseguim de nou un model paramètric amb una estructura semialgebraica. Podríem representar una distribució que satisfés (7) com un punt a $\mathbb{P}_{\mathbb{R}}^X$ (ignorant la constant normalitzadora). Veiem que la família de tots aquests punts admet una parametrització monomial i com a tal forma una varietat tòrica en terminologia de geometria algebraica; vegi's [85] per a més detalls.

3.3 Models gràfics latents

Els models gràfics latents estan dissenyats per gestionar situacions en què determinades variables Z no es poden observar de manera directa però tenen un paper crucial a l'hora d'explicar patrons i relacions en les dades observades X . De fet, l'ús de variables latents permet, fent servir un menor nombre de variables, assolir una major expressibilitat del model, és a dir, capturar una major complexitat i variabilitat de les dades. És per això que aquestes variables *latents*, també conegudes com a variables *ocultes*, estan presents en moltes àrees i són essencials per modelitzar fenòmens complexos. Malgrat això, com ja hem esmentat anteriorment, aquests models presenten dificultats a l'hora de fer inferència.

EXEMPLE (MODEL BAYES PRIMARI). Considerem un problema de classificació sobre unes dades de variables discretes $X = (X_1, X_2, \dots, X_m)$ amb una altra variable discreta Z codificant a quina de les k classes $C = \{c_1, c_2, \dots, c_k\}$ pertany una observació. Si la classe és desconeguda per a noves observacions, la podem modelitzar com una variable latent. La hipòtesi que es fa servir en el model Bayes primari (també anomenat *classificador bayesià naïf*) és que els atributs són condicionalment independents si es coneix la classe, és a dir, $X_i \perp\!\!\!\perp X_j \mid Z$. Aquesta propietat es pot representar com un model gràfic acíclic dirigit (DAG) en el graf següent:

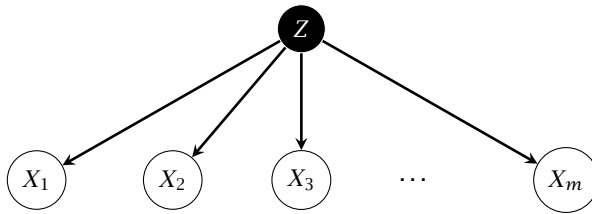


FIGURA 3: Model gràfic amb una variable latent Z i variables observables $X_1, \dots, X_m$ condicionalment independents donat $Z = z$.

Per tal de predir la classe $z \in C$ per una observació concreta de les dades $\mathbf{x} = (x_1, \dots, x_m)$, podem fer servir la regla de Bayes per calcular la probabilitat *a posteriori* de cada classe:

$$p(z|\mathbf{x}) \propto p(z) \prod_{i=1}^m p(x_i|z),$$

i, segons aquesta probabilitat, la classe $z \in C$ s'escull segons la regla de decisió bayesiana, la qual maximitza aquesta probabilitat *a posteriori*:

$$z = \arg \max_{z \in C} \left\{ p(z) \prod_{i=1}^m p(x_i|z) \right\}.$$

Aquesta estratègia s'utilitza àmpliament en el processament de llenguatge natural i per a tasques de classificació de texts. Tot i ser un model tan simple, el model Bayes primari obté resultats sorprenentment bons a la pràctica, especialment per a tasques de detecció de correu brossa, anàlisi d'opinions i classificació temàtica.

EXEMPLE (MODEL EVOLUTIU EN UN TRÍPODE). Com s'ha mencionat anteriorment, els models d'arbres latents formen un exemple important de models de variables latents i han estat aplicats amb èxit a la biologia. A la secció 5, considerem processos de Markov en arbres filogenètics, però aquí considerem un cas senzill mitjançant el model Bayes primari amb $m = 3$, en què totes les variables aleatòries X_1, X_2, X_3 i Z són discretes i prenen valors a $S = \{0, 1, 2, 3\}$. Aquest conjunt d'estats representa els quatre nucleòtids (adenina, citosina, timina i guanina) que es troben en les seqüències d'ADN. En aquest cas, les variables aleatòries X_1, X_2, X_3 corresponen a observacions de nucleòtids en seqüències d'ADN associades a espècies biològiques actuals (per exemple, humana, ximpanzé i goril·la), mentre que la variable latent Z codifica nucleòtids en una seqüència d'ADN d'un ancestre comú a aquestes espècies. La propietat de Markov local en aquest DAG dona

$$p(x_1, x_2, x_3, z) = p(z)p(x_1|z)p(x_2|z)p(x_3|z)$$

(cf. exemple DAG estrella), i, marginalitzant sobre la variable latent, obtenim que

$$p(x_1, x_2, x_3) = \sum_{z \in S} p(z)p(x_1|z)p(x_2|z)p(x_3|z),$$

per a qualssevol $(x_1, x_2, x_3) \in S^3$. Aquestes probabilitats $p(x_1, x_2, x_3)$ representen la probabilitat d'observar els nucleòtids x_1, x_2, x_3 a les fulles de l'arbre i es poden estimar a partir de les observacions de les dades a les espècies actuals, mentre que $p(z)$ i $p(x_i|z)$ són paràmetres (desconeguts) del model. En filogenètica, l'estimació d'aquests paràmetres és rellevant perquè mesuren la quantitat de substitucions de nucleòtids transcorregudes des de l'ancestre comú fins a cada espècie actual, que dona una noció de *distància evolutiva* entre una espècie ancestral i els seus descendents. A la secció 5 explorem mètodes basats en moments per estimar aquests paràmetres, i a la secció 6 comentem aquests models amb més detall.

EXEMPLE (BOSSA DE PARAULES). Considerem ara una versió restringida del model Bayes primari en què les X_i condicionades respecte de la variable latent Z són idènticament distribuïdes. Com a exemple pràctic, considerem el model

de la bossa de paraules (*bag-of-words* en anglès), en què un document format per m paraules ha de ser classificat en un dels k temes ocults. Suposem que el vocabulari té una mida d i que cada tema $c_h \in C$ té el seu propi vector de probabilitat $\mu_h \in \Delta^d$ per emetre paraules del vocabulari: primer s'escull un tema $c_h \in C$ amb probabilitat $p(z = c_h)$, i després es mostregen m paraules $(x_1, \dots, x_m)$ de manera independent amb vector de probabilitat μ_h .

En aquest exemple, l'ordre de les paraules en cada document és irrellevant, només es té en compte si hi són presents o no, i, per tant, els mots es consideren com a variables aleatòries *intercanviables*. Això significa que la seva distribució conjunta $p(x_1, \dots, x_m)$ és invariant per permutacions dels índexs i, en conseqüència, el tensor p corresponent és un tensor simètric.

L'estimació de paràmetres en un model latent no és una tasca fàcil. A l'exemple *mètode dels moments per a la bossa de paraules* explicarem un mètode específic per al model de bossa de paraules basat en el mètode dels moments. Una altra aproximació és optimitzar la versemblança.

NOTA 2. Com hem apuntat en seccions anteriors, una distribució $p(x, z | \theta)$ amb variables latents on z denota un estat latent, la inferència de paràmetres pot ser complicada donat que z no és conegut. L'aproximació més clàssica és la de mirar de maximitzar la versemblança marginal sobre les variables observades

$$p(x|\theta) = \sum_z p(x, z | \theta),$$

on podem substituir la suma per una integral si Z també inclou variables contínues. Cal remarcar que tota la informació coneguda sobre les variables latents Z es troba a través de la seva distribució *a posteriori* $p(\cdot | x, \theta)$. En no disposar de la versemblança per al conjunt de dades complet, com a alternativa es procedeix a estimar-la prenent el valor esperat de la distribució *a posteriori* (pas E), i seguidament aquesta versemblança estimada es maximitza respecte als paràmetres θ (pas M). Aquest procediment rellevant per tractar les dades incompletes que provenen de variables latents es coneix com a *algorisme EM* (*expectation maximization* en anglès); vegeu [30].

4 El cas gaussià

En aquesta secció posem el focus en variables aleatòries contínues i, en particular, en variables normals multivariants. Generalitzem primer els tensors de moments introduïts abans en el cas discret.

4.1 Tensors de moments - cas general

Els tensors simètrics apareixen de manera natural com a derivades d'ordre superior de funcions diferenciables $f: \mathbb{R}^n \rightarrow \mathbb{R}$: les derivades parcials d'ordre r de f es poden organitzar com un tensor $T \in S^r(\mathbb{R}^n)$ amb coordenades

$$T_{i_1 \dots i_r} = \frac{\partial^r}{\partial x_{i_1} \dots \partial x_{i_r}} f(x), \quad \text{per } 1 \leq i_1, \dots, i_r \leq n.$$

En relació amb l'estadística i l'aprenentatge automàtic, hi ha dos tipus concrets de funcions i les seves derivades parcials que són particularment rellevants. Considerem un vector aleatori $X = (X_1, \dots, X_m)$ en què ara permetem que els espais d'estats X_i siguin subconjunts arbitraris de $\mathbb{R}$. Siguin $M_X(\mathbf{s}) = \mathbb{E}[e^{\mathbf{s}^\top X}]$ i $K_X(\mathbf{s}) = \log \mathbb{E}[e^{\mathbf{s}^\top X}]$ les funcions generatrius de moments i de cumulants, respectivament (vegeu [54]). El *tensor de moment* d'ordre r es denota amb $\mu_r(X)$ i és la multitaula de dimensions $m \times \dots \times m$, l'entrada $(i_1, \dots, i_r)$ de la qual és

$$\mu_{i_1 \dots i_r}(X) = \mathbb{E}[X_{i_1} \dots X_{i_r}] = \frac{\partial^r}{\partial s_{i_1} \dots \partial s_{i_r}} M_X(\mathbf{s}) \Big|_{\mathbf{s}=0}.$$

De manera similar, el *tensor cumulant* $\kappa_r(X)$ es defineix en coordenades com a

$$\kappa_{i_1 \dots i_r}(X) = \text{cum}(X_{i_1}, \dots, X_{i_r}) = \frac{\partial^r}{\partial s_{i_1} \dots \partial s_{i_r}} K_X(\mathbf{s}) \Big|_{\mathbf{s}=0}.$$

La relació entre $\mu_r(X)$ i $\kappa_r(X)$ per a una r qualsevol és rebuscada però està molt ben entesa; vegi's [79, 54]. Directament, per construcció, $\mu_r(X)$ i $\kappa_r(X)$ són tensors simètrics de l'espai $S^r(\mathbb{R}^m)$.

NOTA 3. Aquí estem fent servir una notació estàndard per a tensor de moments i cumulants. En aquesta convenció tenim, per exemple, que $\mu_1(X) = \mathbb{E}[X_1]$, $\mu_{13}(X) = \mathbb{E}[X_1 X_3]$, $\kappa_{12}(X) = \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]$. Això contrasta amb la notació que emprem en la secció 2.3, que també es fa servir regularment en l'estudi algebraic de models discrets.

Com a exemple, noteu que, si $X = (X_1, X_2)$ són dues variables aleatòries independents, aleshores $\mu_{12} = \mu_1 \mu_2$ i $\kappa_{12} = 0$.

No obstant això, els moments i cumulants d'ordre baix no són suficients per a l'estudi d'un model estadístic, en general. Una excepció important és el cas de distribucions normals multivariants (vegi's la subsecció següent), que es parametritzen pel vector de mitjanes μ i per la matriu de variàncies i covariàncies Σ , que són moments/cumulants d'ordre 1 i 2, respectivament.

4.2 Distribució normal multivariant

La distribució normal o gaussiana multivariada és la distribució més important en l'estadística multivariant i en ciència de dades en general. També té un paper important en moltes tècniques de reducció de dimensionalitat per a tensors que esmentarem en aquest document.

Denotant amb $S_+^2(\mathbb{R}^n)$ el conjunt de matrius simètriques $n \times n$ definides positives, la distribució gaussiana n -variada amb mitjana $\mu \in \mathbb{R}^n$ i matriu de covariància $\Sigma \in S_+^2(\mathbb{R}^n)$ és una distribució sobre $\mathbb{R}^n$ amb funció de densitat

$$f(x) = \det(2\pi\Sigma)^{-1/2} \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right). \quad (8)$$

Les funcions generadores de moments i cumulants d'un vector aleatori normal X són

$$M_X(\mathbf{s}) = e^{\mu^\top \mathbf{s} + \frac{1}{2} \mathbf{s}^\top \Sigma \mathbf{s}}, \quad K_X(\mathbf{s}) = \mu^\top \mathbf{s} + \frac{1}{2} \mathbf{s}^\top \Sigma \mathbf{s}.$$

Observem que la funció generatriu de cumulants és un polinomi quadràtic i, per tant, tots els cumulants d'ordre tres o superior, com ara l'asimetria o la curtosi, que són els cumulants d'ordre tres i quatre, respectivament, són nuls. Per un resultat clàssic de Józef Marcinkiewicz, la distribució normal és l'única distribució tal que la seva funció generatriu de cumulants és polinomial [53, 51]. El fet que aquesta funció sigui un polinomi té conseqüències importants en probabilitat i estadística (vegeu, per exemple, [44]) i el fet que cap altra distribució no tingui aquesta propietat també és rellevant per justificar el resultat d'identificabilitat de [26] que apareixerà en la secció 6.3.

Una de les propietats fonamentals de la distribució normal és que és tancada per la consideració de marginals i condicionades. De manera més precisa, per a un vector aleatori X amb distribució normal n -variant de mitjana μ i matriu de covariància Σ , considerant una divisió arbitrària en dos blocs (X_A, X_B) , la distribució del subvector X_A és també normal amb mitjana $\mu_A = (\mu_i)_{i \in A}$ i matriu de covariància $\Sigma_{A,A} = [\Sigma_{ij}]_{i,j \in A}$. Així mateix, la distribució condicionada de X_A donat $X_B = x_B$ és normal amb mitjana $\mu_A + \Sigma_{A,B} \Sigma_{B,B}^{-1} (x_B - \mu_B)$ i covariància $\Sigma_{A,A} + \Sigma_{A,B} \Sigma_{B,B}^{-1} \Sigma_{B,A}$. Per tant, treballar amb distribucions marginals i condicionades d'una normal multivariant implica una transformació purament algebraica en termes dels paràmetres.

4.3 Models gràfics gaussians

Les restriccions dels models gràfics tant per al cas dirigit com no dirigit permeten una descripció algebraica senzilla quan el vector X segueix una distribució normal multivariant.

Suposem que $G = (V, E)$ és un graf no dirigit amb vèrtexs $V = \{1, \dots, m\}$ i que $X = (X_1, \dots, X_m)$ segueix una distribució normal multivariant. Aleshores, desenvolupant (8), és fàcil veure que la distribució factoritza com un model (6) en G sempre que la matriu de covariància Σ satisfaci

$$(\Sigma^{-1})_{ij} = 0, \quad \text{per tot } (i, j) \notin E, \quad i \neq j.$$

Per tant, parametritzar el model en termes de la inversa de la matriu de covariància $K = \Sigma^{-1}$ ens permet definir el model a partir de senzilles restriccions lineals. Aquesta observació ha motivat una direcció de recerca en estadística algebraica per entendre els subespais lineals de matrius simètriques, per exemple, [87, 81, 82, 100].

Les distribucions gaussianes definides per DAG G també tenen una descripció simple, la qual és equivalent al model d'equacions lineals estructurals (vegeu la secció 6.3) a base de considerar el sistema d'equacions lineals estocàstiques

$$X_i = \sum_{j \in \text{pa}(X_i)} \lambda_{ij} X_j + \varepsilon_i, \quad \text{per } i \in V, \quad (9)$$

on els ε_i per a $i \in V$ se suposen independents entre si i independents dels $\text{pa}(X_i)$ corresponents. Suposem que els ε_i són variables normals de mitjana

zero, i els coeficients λ_{ij} prenen valors reals arbitraris. En forma matricial podem escriure-ho com a $X = \Lambda X + \varepsilon$, on Λ és triangular superior (reordenant X , si cal). Com que G és acíclic, $I_m - \Lambda$ és invertible i tenim que $X = (I_m - \Lambda)^{-1} \varepsilon$. D'aquí segueix que la inversa de la matriu de covariància $K = \Sigma^{-1}$ satisfà

$$K = (I_m - \Lambda)^\top \Omega^{-1} (I_m - \Lambda),$$

on Ω és la matriu diagonal amb les variàncies dels ε_i com a entrades. Per tant, aquest model es pot descriure de manera equivalent amb condicions d'anul·lació sobre la factorització de Cholesky de K . Aquests models han estat estudiats des d'una perspectiva algebraica a [84].

EXEMPLE (ANÀLISI FACTORIAL). Per a un conjunt de m variables observables $X = (X_1, \dots, X_m)$ amb distribució conjunta normal multivariant $X \sim \mathcal{N}_m(\mu, \Sigma)$, ens pot interessar reduir aquestes variables en un espai de $d < m$ dimensions mitjançant una projecció lineal. Equivalentment, busquem modelitzar m variables observades gaussianes que depenen linealment de d variables gaussianes latents independents $Z = (Z_1, \dots, Z_d)$. Podem escriure, doncs, que $X = \mu + WZ + \varepsilon$ on $\mu \in \mathbb{R}^d$, W és una matriu $m \times d$, i el terme de soroll ε podem suposar que té covariància $\Omega := \text{Cov}(\varepsilon)$ diagonal. La matriu de covariància Σ en aquest cas ve donada per

$$\mathbb{E}[(WZ + \varepsilon)(WZ + \varepsilon)^\top] = WW^\top + \text{Cov}(\varepsilon).$$

Així doncs, Σ viu al conjunt

$$F_{m,d} = \{\Omega + WW^\top \in \mathbb{R}^{m \times m} : \Omega \succ 0 \text{ diagonal}, W \in \mathbb{R}^{m \times d}\},$$

on fem servir $\succ 0$ per dir que la matriu és definida positiva. Observi's que $F_{m,d}$ és un conjunt semialgebraic, ja que es pot veure com la imatge de $\mathbb{R}_{>0}^m \times \mathbb{R}^{m \times d}$ sota una aplicació racional. Per tant, el model d'anàlisi factorial format per les distribucions normals multivariants X amb matriu de covariància $\Sigma \in F_{m,d}$ és un exemple de model estadístic semialgebraic. Entendre quines són les equacions algebraiques i semialgebraiques que defineixen el conjunt $F_{m,d}$ és clau per saber quines distribucions poden encaixar en aquest model.

5 Descomposició tensorial en estadística algebraica

De manera anàloga a la descomposició de matrius, la descomposició de tensors té com a objectiu representar un tensor d'altres dimensions com una sèrie d'operacions elementals en tensors més simples. Les descomposicions de tensors s'han utilitzat en diferents camps, com ara en la visió per computador, la neurociència computacional, la filogenètica o la psicometria. S'utilitzen per obtenir separació cega de fonts (*blind source separation*), l'anàlisi de components independents i per proporcionar estimadors en models de variables latents. Remetem el lector a [46] i [45], on trobarà excel·lents compilacions sobre la descomposició de tensors i les seves aplicacions en aprenentatge automàtic.

5.1 Rang d'un tensor

Un exercici senzill d'àlgebra lineal (usant la descomposició en valors singulars, per exemple) consisteix a veure que una matriu M de dimensions $n_1 \times n_2$ té rang k si i només si es pot escriure com a

$$M = u^1 \otimes v^1 + \dots + u^k \otimes v^k$$

per a determinats vectors $u^i \in \mathbb{R}^{n_1}$, $v^i \in \mathbb{R}^{n_2}$, i no és possible escriure aquesta descomposició amb menys sumands (aquí $u_i \otimes v_i$ s'ha de pensar com a la matriu de rang u corresponent).

Una manera de generalitzar aquest concepte del rang per tensors és la següent.

DEFINICIÓ 2. El *rang* d'un tensor $T \in \mathbb{R}^{r_1} \otimes \dots \otimes \mathbb{R}^{r_m}$ és el menor enter k tal que T es pot escriure com a

$$\sum_{j=1}^k u^{1,j} \otimes u^{2,j} \otimes \dots \otimes u^{m,j} \quad (10)$$

per a determinats $u^{i,j} \in \mathbb{R}^{r_i}$, $i \in [m]$, $j \in [k]$. Dit d'una altra manera, un tensor té rang k si es pot escriure com una combinació lineal de k tensors de rang u (purs), però no com a suma d'un nombre menor de k tensors purs.

El *rang simètric* d'un tensor simètric $T \in S^r(\mathbb{R}^m)$ es defineix com l'enter k més petit tal que $T = \sum_{j=1}^k u^j \otimes u^j \otimes \dots \otimes u^j$ per a determinats $u^j \in \mathbb{R}^m$, $j \in [k]$.

Ja hem vist que els tensors que apareixen en el model d'independència total són tensors de rang u. A continuació donem exemples de tensors de rang superior. En termes de geometria algebraica, donat que la varietat $\text{Seg}(\mathbb{P}_{\mathbb{R}}^{X_1} \times \dots \times \mathbb{P}_{\mathbb{R}}^{X_m})$ està formada pels tensors de rang u, el conjunt de tensors de rang k es troba a la seva k -èsima varietat secant (de fet, la varietat secant és la clausura en la topologia de Zariski d'aquest conjunt de tensors).

EXEMPLE (RANG TENSORIAL). En els tensors que codifiquen la distribució de X en el model de l'exemple *model Bayes primari* amb una variable latent Z que pren valors en un conjunt C de cardinal k , el rang és com a màxim k . En efecte, la propietat de Markov dona la factorització de la distribució conjunta, i, marginalitzant sobre la variable latent Z , obtenim

$$p(x_1, \dots, x_m) = \sum_{z \in C} p(z) \prod_{i=1}^m p(x_i | z)$$

per a qualsevol conjunt d'observacions $x = (x_1, \dots, x_m) \in X$. Llavors, considerant els vectors

$$u^{i,z} = (p(x_i = 0 | z), \dots, p(x_i = r_i - 1 | z)) \in \mathbb{R}^{X_i}, \quad \text{on } z \in C, i = 1, \dots, m,$$

podem escriure la distribució $p = (p(x))_{x \in X} \in \mathbb{R}^X$ de $X = (X_1, \dots, X_m)$ com a

$$p = \sum_{z \in C} p(z) u^{1,z} \otimes \dots \otimes u^{m,z}. \quad (11)$$

Per tant, p té rang tensorial menor o igual que k . En particular, la distribució conjunta p a les fulles del model evolutiu sobre el tripode de l'exemple *model evolutiu en un tripode* és un tensor de rang menor o igual a quatre.

EXEMPLE 3. Considerem el model introduït a la bossa de paraules. En aquest cas, els vectors $u_{i,z}$ coincideixen amb μ_h si z és el tema c_h (i no depèn de i). Llavors, l'expressió (11) esdevé

$$p = \sum_{h=1}^k p(z = c_h) \mu_h \otimes \cdots \otimes \mu_h$$

i obtenim que p és un tensor simètric de rang com a màxim k .

5.2 Descomposició tensorial

Una descomposició com la de (10) ha estat introduïda en diversos àmbits. Es coneix com a *descomposició de rang tensorial* o com a descomposició *poliàdica canònica* (CP en anglès) o també *factor paral·lel* (PARAFAC); vegeu [43, 41].

Aquesta descomposició no és única en general: es poden reescalar els vectors $u_{i,j}$ o permutar els sumands per obtenir una altra descomposició. Aquestes són ambigüitats inherents que solen poder resoldre's per a cada aplicació particular en què treballem. Restringint-nos a descomposicions especials, podem obtenir la unicitat mòdul aquestes ambigüitats. Com en el cas del teorema espectral per a matrius simètriques, l'ortogonalitat és un requisit comú per descompondre tensors simètrics, però no tots els tensors simètrics tenen una descomposició ortogonal. La unicitat en la descomposició CP (mòdul permutacions i reescalatge) es pot obtenir per a tensors de baix rang, com veurem a continuació.

EXEMPLE 4. Considerem el model Bayes primari de l'exemple *model Bayes primari* amb $m = 3$ i deixem que k sigui el cardinal del conjunt d'estats de la variable latent. A [47], l'autor va donar condicions explícites per a la identificabilitat dels sumands (mòdul permutacions i escalars) en la descomposició de rang del tensor distribució conjunta. Aquesta condició explícita es compleix per a tensors *genèrics* del model si k és menor que el nombre r_i d'estats de cada variable observada, i contrasta fortament amb la no unicitat de descomposicions de matrius de rang k com a sumes de matrius de rang u . A [1], els autors mostren com aquest resultat bàsic es pot usar de manera potent per establir la identificabilitat d'una gran classe de models de variables latents.

En el context dels arbres filogenètics com a l'exemple *model evolutiu en un tripode*, aquest resultat es va redescobrir a [20]. En aquest cas, el resultat diu que els paràmetres del model (la distribució de nucleòtids en la variable latent i les probabilitats condicionades) es poden recuperar de manera única (mòdul permutacions) a partir de la distribució conjunta de les variables observades i es dona una construcció explícita per a la seva recuperació utilitzant descomposicions espectrals de determinades matrius extrems de la distribució conjunta.

5.3 Rang tensorial no negatiu

Els tensors que provenen de distribucions de probabilitat, ja sigui discreta o contínua, formen un subconjunt especial dins de la classe de tots els tensors. Per aquesta raó, i especialment per al cas discret, sovint és útil parlar no del seu rang com en (10), sinó d'una variant coneguda com a *rang no negatiu*: diem que un tensor T té *rang no negatiu* com a màxim k si es pot escriure com la suma de k tensors de rang u com abans

$$T = \sum_{j=1}^k u^{1,j} \otimes \dots \otimes u^{m,j}, \quad u^{i,j} \in \mathbb{R}_{\geq 0}^{r_i}, \quad i \in [m], j \in [k], \quad (12)$$

on ara requerim que tots els vectors $u^{i,j}$ tinguin entrades no negatives. El *rang no negatiu* denotat amb rank_+ és el menor enter k tal que existeix una expressió com la de (12). El rang no negatiu no ha de coincidir necessàriament amb el rang usual d'un tensor, i per definició es té $\text{rank } T \leq \text{rank}_+ T$. Per a més generalitzacions i variants del rang de tensors, vegeu [48].

NOTA 4. És un exercici senzill comprovar que la matriu següent és de rang tres, però el seu rang no negatiu és quatre.

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Aquest és un exemple senzill sobre com el rang clàssic i el rang no negatiu poden diferir.

Determinar si els rangs clàssic i no negatiu coincideixen és un problema NP-difícil. De fet, la factorització no negativa de matrius és per si mateixa un problema NP-difícil, tot i que mitjançant heurístiques de cerca local [88] es pot aconseguir un rendiment polinomial. Afortunadament, hi ha condicions sota les quals aquestes descomposicions de rang no negatiu són úniques, i aquestes estan estretament lligades la identificabilitat del model subjacent [67].

La descomposició no negativa de tensors ha estat estudiada en el context PARAFAC incloent-hi restriccions de no negativitat en quimiometria [14, 15], i es pot interpretar també com una descomposició de Bayes primari de distribucions de probabilitat respecte a condicionals com es fa a [40].

Tenint en compte l'expressió a (1), el model d'independència total és un exemple d'un tensor de rang no negatiu ≤ 1 . Per exemple, el subconjunt de matrius $r_1 \times r_2$ amb $\text{rank}_+ \leq 1$ és la part no negativa de la varietat de Segre $\mathbb{P}^{X_1} \times \mathbb{P}^{X_2}$. Les distribucions del model d'independència de dues variables discretes corresponen a matrius en la intersecció de Δ^X amb el conjunt de matrius de $\text{rank}_+ \leq 1$.

Els models de mixtura o de barreja (*mixture model* en anglès) fan ús implícit del rang tensorial no negatiu. Per a un model estadístic $\mathcal{P} \subset \Delta^{\mathcal{X}}$, el k -èsim *model de mixtura* $\text{Mixt}^k(\mathcal{P})$ ve donat per

$$\text{Mixt}^k(\mathcal{P}) = \left\{ \sum_{i=1}^k \pi_i p^i : \pi \in \Delta^{k-1}, p^i \in \mathcal{P} \right\}.$$

Està format per totes les distribucions de probabilitat de rang no negatiu menor o igual que k i és també un model gràfic amb variables ocultes. Un tractament complet sobre aquesta relació es pot trobar a [34].

Un cas important és el dels tensors de rang no negatiu ≤ 2 , que són clau en l'estudi de models de Markov en arbres filogenètics amb estats binaris. El model Bayes primari en 4 amb variables observades binàries i dues classes latents es pot representar per tensors de dimensions $2 \times 2 \times \dots \times 2$ de rang no negatiu com a màxim 2 i corresponen a mixtures de dues distribucions. Vegeu [3] per a una caracterització d'aquests tensors amb $\text{rank}_+ \leq 2$ i la seva geometria.

EXEMPLE (MODELS TEMÀTICS). El model descrit a bossa de paraules és un classificador basat en el contingut del document, però no és un model generatiu i el seu potencial representatiu és limitat. Els models temàtics (*topic models* en anglès), [11], s'utilitzen quan els documents poden exhibir més d'una temàtica i la classificació del model *bossa de paraules* seria una aproximació massa restrictiva. Un exemple d'aquest tipus de model és el de l'assignació latent de Dirichlet (LDA) [12], que és un model generatiu que funciona com una xarxa bayesiana (secció 3.1), on es té una col·lecció de n documents i k temes, amb la propietat que els elements poden pertànyer a diverses classes (és a dir, un document pot tractar sobre més d'un tema).

Cada *topic* o tema es representen com a distribucions de probabilitat sobre les paraules, i els documents es representen, al seu torn, com a distribucions sobre temes latents. Per a cada document, s'assigna un escalar $\theta_{ij} \in [0, 1]$ a cada tema $j \in \{1, 2, \dots, k\}$ segons la seva adequació al document, és a dir, la probabilitat que el tema j aparegui en el document i , com un model de mixtura en què assumim que el nombre total de temes k és conegut. Els pesos associats als temes del document segueixen una distribució de Dirichlet $\text{Dir}(\alpha)$, amb una funció de densitat definida com a

$$p(\theta_i; \alpha) = f(\theta_{i1}, \dots, \theta_{ik}; \alpha_1, \dots, \alpha_k) = \frac{1}{\mathbf{B}(\alpha)} \prod_{j=1}^k \theta_{ij}^{\alpha_j - 1},$$

on $\mathbf{B}$ és la funció Beta que normalitza la distribució. Els pesos $\theta_i = (\theta_{i1}, \dots, \theta_{ik}) \in \Delta^{k-1}$ sumen u i, per tant, formen una mesura discreta com teníem a la secció 2.1. D'altra banda, una paraula i té una probabilitat φ_{ij} de pertànyer al tema j , on φ_{ij} segueix una distribució de Dirichlet $\text{Dir}(\beta)$. Els tres conjunts de variables aleatòries que representen paraules, temes i documents es poden interpretar com un model gràfic. Concatenant aquests tres conjunts de variables, obtenim de manera natural una xarxa bayesiana amb capes de variables en l'ordre esmentat de predecessors a descendents.

5.4 Mètode dels moments

Fer aprenentatge en els tipus de models discutits a la secció 3.3 genera molts problemes. Els mètodes de versemblança, que són una opció popular per a molts models simples, ràpidament es tornen intractables en presència de variables latents. Amb poques alternatives bones, el mètode dels moments [64] ara s'utilitza rutinàriament en mixtures gaussianes [5] i en models d'arbres latents [7], [6], [73]. Com que aquest mètode és intrínsecament algebraic, el descrivim amb més detall en aquest manuscrit.

La idea del mètode és senzilla. Suposem que la parametrització del model induïx una parametrització explícita d'alguns moments $\mu = f(\theta)$ de la distribució subjacent. Un estimador pel mètode dels moments procedeix en dues etapes: primer, es troba un bon estimador $\hat{\mu}$ dels moments a partir de la mostra i després es defineix $\hat{\theta}$ com una solució de $\hat{\mu} = f(\theta)$.

Un problema evident d'aquest mètode és que, per a un $\hat{\mu}$ donat, no tenim garanties que l'equació $\hat{\mu} = f(\theta)$ tingui solució. En certa mesura, això es pot solucionar escollint adequadament uns moments determinats per a l'anàlisi. Però, fins i tot si existeix una solució, pot ser que no compleixi algunes de les restriccions necessàries. Per aquest motiu, el mètode dels moments normalment s'aplica cas per cas i requereix una bona comprensió de la classe de models que estem tenint en compte.

A tall d'exemple, en el treball clàssic [63], es va utilitzar una mixtura de dues variables gaussianes per modelitzar les mides de les parts del cos d'una població de crancs. Per fer-ho, Pearson va proposar un sistema d'equacions que involucrava els primers cinc moments de la distribució, i seguidament va aconseguir eliminar les variables i va obtenir un polinomi de grau nou en una única variable. Resolent el polinomi de grau nou resultant, va poder obtenir els paràmetres de la distribució de mixtura. Per a més detalls d'aquest primer exemple bàsic del mètode dels moments, consulteu [4].

EXEMPLE (MÈTODE DELS MOMENTS PER A LA BOSSA DE PARAULES). Considerem la bossa de paraules de l'exemple *bossa de paraules* i suposem que el nombre m de paraules en el document és almenys 3. Anandkumar *et al.* ([7]) van mostrar que, fent servir moments d'ordre com a màxim tres, hom pot recuperar els paràmetres del model (és a dir, els vectors de les distribucions condicionades μ_h i les probabilitats $w_h = p(z = c_h)$ per a $h = 1, \dots, k$). Aquí expliquem breument aquest mètode dels moments seguint [6, apèndix D].

Podem representar les paraules del vocabulari amb els vectors de la base canònica $e_1, \dots, e_d \in \mathbb{R}^d$, i considerem X tal que l'espai d'estats de X_i sigui $\mathcal{D} = \{e_1, \dots, e_d\}$. Considerem les variables X_i com a indicadors de la manera següent:

$$X_i = e_j \iff \text{la } i\text{-èsima paraula del document és } j.$$

Aleshores els moments creuats d'aquests vectors aleatoris es corresponen amb les probabilitats conjuntes:

$$\mathbb{E}[X_1 \otimes X_2] = \sum_{i,j} p(x_1 = e_i, x_2 = e_j) e_i \otimes e_j$$

i obtenim

$$\mathbb{E}[X_1 \otimes X_2] = \sum_{h=1}^k \sum_{i,j} p(z = c_h) p(x_1 = e_i | z = c_h) p(x_2 = e_j | z = c_h) e_i \otimes e_j.$$

Amb la notació introduïda es pot escriure com a

$$\mathbb{E}[X_1 \otimes X_2] = \sum_{h=1}^k w_h \mu_h \otimes \mu_h. \quad (13)$$

De manera anàloga obtenim

$$\mathbb{E}[X_1 \otimes X_2 \otimes X_3] = \sum_{h=1}^k w_h \mu_h \otimes \mu_h \otimes \mu_h.$$

Si prenem $V = (\mu_1 | \dots | \mu_k)$ i $D = \text{diag}(w_1, \dots, w_k)$, aleshores, considerant $M_2 = \mathbb{E}[X_1 \otimes X_2]$ com una matriu, podem expressar (13) com a

$$M_2 = V D V^\top. \quad (14)$$

No es tracta d'una diagonalització de $\mathbb{E}[X_1 \otimes X_2]$ (llevat de quan V és ortogonal), però podem emprar tècniques espectrals incorporant els moments de tercer ordre. De fet, considerem $M_3 = \mathbb{E}[X_1 \otimes X_2 \otimes X_3]$ i per a qualsevol $\eta \in \mathbb{R}^d$ definim el 2-tensor $M_3 \bullet \eta$ com a

$$(M_3 \bullet \eta)_{i,j} = \sum_l (M_3)_{i,j,l} \eta_l.$$

Llavors, com a matrius tenim $M_3 \bullet \eta = V D d(\eta) V^\top$, on $d(\eta) = \text{diag}(\mu_1^\top \eta, \dots, \mu_k^\top \eta)$, i si $M = (M_3 \bullet \eta) M_2^{-1}$, obtenim $M = V d(\eta) V^{-1}$. Prenent η amb entrades no repetides, les columnes de V estan determinades pels vectors propis de M (llevat de permutacions i multiplicació per escalars). Un cop es recupera V , $w_1, \dots, w_k$ són fàcilment recuperables mitjançant l'expressió (14).

Existeixen enfocaments més robusts per a l'estimació de paràmetres utilitzant moments, però van més enllà de l'abast d'aquest estudi. Val la pena assenyalar que els paràmetres estimats pel mètode dels moments també es poden utilitzar com a pas inicial d'un algorisme EM (vegeu la nota 2); en aquest cas, l'EM sovint convergeix a un màxim local després d'un sol pas; vegeu [97].

6 Aplicacions de l'estadística algebraica

6.1 Anàlisi filogenètica

Un *arbre filogenètic* T sobre un conjunt L d'espècies (o altres entitats biològiques) és un arbre amb les fulles etiquetades pels elements de L . La figura 4 mostra tal arbre amb fulles $L = \{\text{humà, ximpanzé, goril·la, orangutan}\}$.

L'objectiu principal de la filogenètica és reconstruir la història evolutiva del conjunt L d'espècies actuals a partir de la informació proporcionada per una col·lecció de molècules d'ADN associades a elles. Degut a l'estructura de doble hèlix de l'ADN, aquestes molècules es poden interpretar com a paraules o seqüències en l'alfabet que representen els quatre nucleòtids, $\{A, C, G, T\}$. Sota certes hipòtesis biològiques es pot suposar que cada posició d'aquestes seqüències evoluciona independentment de les altres posicions i que les posicions estan idènticament distribuïdes.

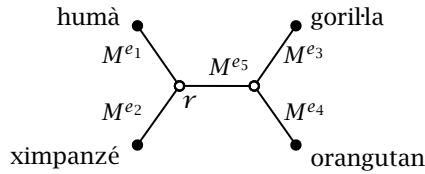


FIGURA 4: Un arbre filogenètic amb conjunt de fulles donat per $L = \{\text{humà, ximpanzé, goril·la, orangutan}\}$, node arrel r , i matrius estocàstiques per files associades al procés de Markov.

Si T és un arbre filogenètic, modelem la substitució de nucleòtids en T seleccionant primer un node interior r que faci el paper d'arrel de l'arbre (que podria representar l'ancestre comú a totes les espècies de L) i dirigint totes les arestes des de r per veure $T = (V, E)$ com un DAG. A cada $v \in V$ li associem una variable aleatòria X_v prenent valors a $S = \{0, 1, 2, 3\}$ (en representació dels quatre nucleòtids) i considerem la propietat de Markov local sobre el DAG. Aquesta propietat és equivalent al fet que cada variable aleatòria sigui independent de les seves variables no descendents donades les observacions al seu node pare immediat, com s'expressa a (4). És conegut que, per obtenir la identificabilitat dels paràmetres, cal que l'arbre no tingui nodes de grau 2. Per a un arbre de tres fulles, aquest model és el que es dona a l'exemple *model evolutiu en un trípede*.

Per escriure la factorització de la distribució conjunta (1) sota aquest model de Markov, codifiquem les distribucions condicionades en matrius estocàstiques per files: a cada aresta $e = \{u \rightarrow v\} \in E$ associem M^e , una matriu 4×4 amb entrades $M^e_{x,y} = p(X_v = y | X_u = x)$, la probabilitat condicionada que un estat x en el node ancestral u es vegi substituït per l'estat y en el descendent v .

Aleshores, la probabilitat conjunta d'observar estats (nucleòtids) x_v a les variables X_v , per a $v \in V$, és

$$p(\{x_v\}_{v \in V}) = p(x_r) \prod_{e=\{u-u'\} \in E} M_{x_{u'}, x_u}^e. \quad (15)$$

En filogenètica només tenim observacions per a les variables aleatòries trobades a les fulles de l'arbre, de manera que les variables aleatòries als nodes interiors són latents (ja que representen espècies extintes). Així, si el conjunt de fulles L és $[n]$, la probabilitat $p_{i_1 \dots i_n}$ d'observar el nucleòtid i_j a la fulla j per $j \in [n]$ es pot obtenir marginalitzant (15) sobre els nodes interiors:

$$p_{i_1 \dots i_n} = \sum_{\{x_v\}_{v \in V} | x_j = i_j, j \in [n]} p(\{x_v\}_v).$$

Per cada arbre filogenètic T sobre $L = [n]$ definim com a $p^\top \in \mathbb{R}^X = \mathbb{R}^4 \otimes \dots \otimes \mathbb{R}^4$ el tensor de la distribució conjunta $(p_{0 \dots 0}, p_{0 \dots 1}, \dots, p_{3 \dots 3})$ obtingut d'aquesta manera. Aquest model s'anomena *procés de Markov general sobre l'arbre filogenètic*.

Si l'arbre filogenètic que explica la història evolutiva del conjunt L és conegut i tenim unes seqüències d'ADN observades sobre el conjunt L , els paràmetres del procés de Markov corresponent se solen estimar mitjançant un mètode de màxima versemblança. No obstant això, un dels problemes principals en filogenètica és obtenir l'arbre T que s'ajusti millor a les dades: com que el nombre d'arbres filogenètics creix superexponencialment en el nombre de fulles, no és possible buscar exhaustivament per tot l'espai d'arbres. Aquí l'estadística algebraica té un paper important, tal com expliquem a continuació (consulteu també [17]).

A finals dels anys vuitanta, els biòlegs Cavender, Felsenstein i Lake es van adonar que algunes equacions polinomials satisfetes per les coordenades de $p^\top$ permeten distingir entre els diferents arbres que donen lloc a la distribució. És a dir, hi ha equacions polinomials sobre les coordenades d'un tensor $p \in \mathbb{R}^4 \otimes \dots \otimes \mathbb{R}^4$ que es compleixen si $p = p^\top$ per a alguns arbres T però no per a altres. Un exemple d'aquestes equacions prové de l'*aplanament* (*flattening* en anglès) del tensor, com detallem a continuació.

Considerem una bipartició $A|B$ de $L = [n]$: un subconjunt A i el seu complement $B = L \setminus A$. Com teníem a la secció 2.2, aquesta divisió indueix un isomorfisme de l'espai de tensors a l'espai de matrius

$$\begin{aligned} \mathbb{R}^X &\cong \mathbb{R}^{X_A} \otimes \mathbb{R}^{X_B} \longrightarrow \mathcal{M}_{|X_A| \times |X_B|}(\mathbb{R}) \\ p &= (p_{x_1 \dots x_n}) \longmapsto \text{flatt}_{A|B}(p), \end{aligned} \quad (16)$$

on l'entrada (x_A, x_B) de $\text{flatt}_{A|B}(p)$ és la coordenada de p que es correspon amb la probabilitat d'observar estats $x_A = \{x_l\}_{l \in A}$ a les fulles de A i $x_B = \{x_l\}_{l \in B}$ a les fulles de B . Per exemple, a l'arbre filogenètic de la figura 4, si les fulles

humà, ximpanzé, gorilla i orangutan es denoten amb 1, 2, 3, 4, respectivament, tenim que

$$\text{flatt}_{12|34}(p) = \begin{pmatrix} p_{0000} & p_{0001} & p_{0002} & \dots & p_{0033} \\ p_{0100} & p_{0101} & p_{0102} & \dots & p_{0133} \\ p_{0200} & p_{0201} & p_{0202} & \dots & p_{0233} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{3300} & p_{3301} & p_{3302} & \dots & p_{3333} \end{pmatrix}.$$

Suposem que $p = p^\top$ per a algun arbre filogenètic T , i considerem una bipartició $A|B$ del conjunt de fulles induïda per la supressió d'una aresta e de T , que anomenarem *bipartició d'aresta*. Llavors $\text{flatt}_{A|B}(p^\top)$ es pot considerar com una taula de probabilitat conjunta de les variables aleatòries $X_A = (X_l)_{l \in A}$ i $X_B = (X_l)_{l \in B}$. Per la propietat de Markov en T sabem que $X_A \perp\!\!\!\perp X_B | X_u$, on X_u és la variable oculta corresponent a un dels vèrtexs de e . Aquest és un model de Bayes primari amb dues variables aleatòries observades X_A, X_B ($m = 2$) i una de latent. Per tant, $p^\top$ és un tensor de rang com a màxim quatre segons l'exemple rang tensorial. En particular, $\text{flatt}_{A|B}(p^\top)$ és una matriu de rang com a màxim quatre perquè (16) envia una suma de tensors de rang u a una suma de matrius de rang u . Així doncs, hem obtingut la primera afirmació del següent teorema:

TEOREMA 3 ([2]). *Sigui T un arbre filogenètic i $A|B$ una divisió del seu conjunt de fulles L . Sigui $p^\top$ un tensor de distribució obtingut per un procés de Markov en T . Aleshores, si $A|B$ és una bipartició d'aresta en T , $\text{flatt}_{A|B}(p^\top)$ té rang menor o igual a quatre. A més, si $A|B$ no és una bipartició d'aresta en T , el rang de $\text{flatt}_{A|B}(p^\top)$ és més gran que quatre (sempre que $p^\top$ compleixi determinades condicions genèriques).*

Així doncs, l'anul·lació dels menors 5×5 de $\text{flatt}_{A|B}(p)$ defineix equacions polinomials que es compleixen per a distribucions que provenen d'arbres que tenen $A|B$ com a bipartició d'aresta, però que no es compleixen per a distribucions en altres arbres. En lloc d'utilitzar directament aquestes equacions algebraïques, és més natural calcular la distància entre matrius de rang quatre directament utilitzant valors singulars. Aquesta estratègia s'ha aprofitat per proporcionar diversos mètodes per a la reconstrucció filogenètica; consulteu, per exemple, [22, 37, 18].

6.2 Màquines de Boltzmann restringides

Considerem un conjunt de variables aleatòries binàries $V = \{X_1, \dots, X_n, Y_1, \dots, Y_m\}$ en els vèrtexs d'un model gràfic no dirigit, on $X = (X_1, \dots, X_n)$ són les variables observades i $Y = (Y_1, \dots, Y_m)$ són les variables ocultes. Aquests models gràfics, coneguts com a *màquines de Boltzmann restringides* (*restricted Boltzmann machines* o RBM en anglès) venen dotats d'una estructura de graf bipartit de manera que únicament hi ha arestes entre variables observades i

variables latents; vegeu la figura 5. L'energia d'una configuració donada de variables aleatòries amb els seus paràmetres en aquest graf es defineix com la funció

$$E(x, y; \theta) = - \sum_{i=1}^n b_i x_i - \sum_{j=1}^m c_j y_j - \sum_{\{i,j\}} W_{ij} x_i y_j,$$

on $\theta = (W, b, c)$ són els paràmetres del model. La notació que usem per als paràmetres està influenciada per la notació en la literatura de perceptrons i xarxes neuronals. A partir de l'energia, podem definir la probabilitat d'una realització del vector aleatori com a

$$p(x, y; \theta) = \frac{1}{Z(\theta)} \exp(-E(x, y; \theta)), \quad (x, y) \in \{0, 1\}^{n+m}, \quad (17)$$

on $Z(\theta)$ és la constant normalitzadora. Com podem veure per la connectivitat del graf, les variables observades són independents entre elles condicionades sobre les latents, i viceversa. La distribució marginal de X es pot obtenir de l'equació (17) sumant sobre els estats de les variables latents $y \in \{0, 1\}^m$, i després, mitjançant les manipulacions adients, obtindrem el producte següent (vegeu [57] per a més detalls sobre la derivació i les màquines de Boltzmann en general):

$$\begin{aligned} p(x; \theta) &= \frac{1}{Z(\theta)} \exp(b^\top x) \prod_{j=1}^m (1 + \exp(w_j x + c_j)) \\ &= \frac{1}{Z(\theta)} \prod_{j=1}^m \left(\lambda_j \prod_{i=1}^n p'_{ji}(x_i) + (1 - \lambda_j) \prod_{i=1}^n p''_{ji}(x_i) \right) \end{aligned} \quad (18)$$

per a λ_j entre zero i u, on w_j és la fila j -èsima de W i p'_{ji}, p''_{ji} són distribucions que depenen exclusivament de x_i . D'aquesta manera trobem una expressió que ens recorda una distribució producte $q(x_1, \dots, x_n) = \prod q_i(x_i)$ que lliga amb un producte tensorial. Concretament, l'expressió de l'equació (18) mostra una estructura de producte tensorial de rang no negatiu dos, del tipus que hem esmentat a la secció 5.3. Ho veiem en l'equació següent, que fa explícita aquesta connexió fent servir notació tensorial:

$$p = \frac{1}{Z(\theta)} \prod_{j=1}^m (q'_{j1} \otimes \dots \otimes q'_{jn} + q''_{j1} \otimes \dots \otimes q''_{jn}),$$

per a determinats q'_{ji}, q''_{ji} . Així doncs, les RBM són un exemple d'un producte de m factors tensorials de rang no negatiu com a màxim dos.

Aquests models han estat explorats i utilitzats en processament de la informació a [78] i a [39] com a precursor de les xarxes neuronals de dues capes, de manera que les seves propietats com a aproximadors de funcions són especialment rellevants en aplicacions d'aprenentatge automàtic. Els tensors de rang no negatiu dos han estat descrits en termes de restriccions de submodularitat a [3].

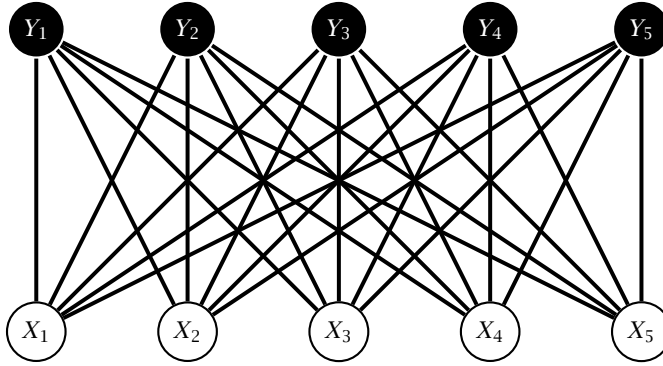


FIGURA 5: Diagrama que mostra una màquina de Boltzmann restrictiva com un model gràfic amb cinc variables latents i cinc variables observades.

NOTA 5. L'anàlisi de les distribucions que sorgeixen d'aquests tensors de rang dos és remarcadament difícil. No obstant això, per al cas en què el model és la suma de dos tensors $2 \times 2 \times 2$, Seigal i Montúfar ([75]) han demostrat que el model de distribucions no degenerades (és a dir, les que cauen a l'interior del simplex) són equivalents a un model gràfic en un arbre amb una única variable oculta de tres estats.

6.3 Components independents i models d'equacions estructurals

Considerem el model següent per a un vector aleatori X ,

$$X = A\varepsilon, \quad (19)$$

on $A \in \mathbb{R}^{m \times m}$ i ε és un vector aleatori (latent) amb una estructura de dependència senzilla. Els models d'aquesta forma s'utilitzen extensament en aplicacions. Depenent del context, podem imposar restriccions addicionals a la matriu A o al vector aleatori (latent) ε . Per exemple, en separació cega de fonts és habitual suposar que A és invertible i que ε té components independents, [27]. La proposició ens ajudarà en aquesta discussió:

PROPOSICIÓ 4. *Les components de X són independents si i només si el tensor cumulants $\kappa_r(X)$ és diagonal per a tot $r \geq 2$.*

El resultat anterior té conseqüències importants en ciència de dades. En l'anàlisi de components independents (ICA), es postula que el vector aleatori observat $X = (X_1, \dots, X_m)$ es pot escriure com una transformació lineal $A\varepsilon$, on $\varepsilon = (\varepsilon_1, \dots, \varepsilon_m)$ té components independents i $A \in \mathbb{R}^{m \times m}$ és invertible. Sota condicions febles sobre la distribució del vector no observat ε , podem recuperar la matriu A a partir de les observacions X (llevat de permutacions de files i canvis en els seus signes), [26]. A la pràctica aquesta recuperació es realitza amb els cumulants d'ordre baix $\kappa_3(Y)$ i $\kappa_4(Y)$, aprofitant el fet que aquests són tensors diagonals.

Donat que el vector ε no s'observa directament, la pregunta principal és si la matriu A es pot recuperar únicament a partir de les observacions de X . Sense pèrdua de generalitat, suposem que ε té mitjana zero i matriu de covariància identitat. Sota aquesta hipòtesi, la covariància de X satisfà $\text{Cov}(X) = AA^\top$, cosa que ens permet identificar A mòdul l'acció del grup ortogonal sobre l'espai de les columnes de A ($AA^\top = (AQ)(AQ)^\top$ per a qualsevol matriu ortogonal Q). Aquest resultat d'identificabilitat es pot millorar si ε és no gaussià. Concretament, en un treball clau, Comon ([26]) va demostrar que A es pot recuperar llevat de permutacions i reescalaments de les columnes per a ± 1 , sempre que com a màxim una component de ε sigui normal. Sota condicions genèriques febles, A es pot recuperar a partir dels moments de tercer o quart ordre de X . A la pràctica, molts mètodes utilitzen aquests moments d'ordre inferior.

Per connectar el model (19) a la nostra discussió sobre tensors, considerem $\mu_r(X), \mu_r(\varepsilon) \in S^r(\mathbb{R}^m)$ com els moments tensorials de X i ε , respectivament, tal com hem introduït a la secció 4.1. Suposem que ε té components independents i, per tant, per la proposició 4, $\mu_r(\varepsilon)$ és un tensor diagonal. Així mateix, per la propietat multilinear dels moments, tenim que

$$\mu_r(X) = \mu_r(A\varepsilon) = A \bullet \mu_r(\varepsilon),$$

on la notació $A \bullet \mu_r(\varepsilon) \in S^r(\mathbb{R}^m)$ és l'acció multilinear de $A \in \mathbb{R}^{m \times m}$ sobre $S^r(\mathbb{R}^m)$:

$$(A \bullet \mu_r(\varepsilon))_{i_1, \dots, i_r} = \sum_{j_1, \dots, j_r} A_{i_1 j_1} \cdots A_{i_r j_r} (\mu_r(\varepsilon))_{j_1, \dots, j_r}.$$

En altres paraules, l'equació (19) estableix que, per a tot $r \geq 2$, el tensor moment r -èsim de X és de la forma $A \bullet D$, per a $A \in \mathbb{R}^{m \times m}$ i un tensor diagonal $D \in S^r(\mathbb{R}^m)$. La pregunta es converteix, doncs, en: en quins casos permet aquesta condició, per a r fixada, identificar la matriu A ? Com que la matriu de covariància ja ens permetia identificar A mòdul l'acció del grup ortogonal, podem reformular-ho de la manera següent. Suposem $A \bullet D = (AQ) \bullet D'$ per a A invertible i Q ortogonal, on D, D' són tensors diagonals; equivalentment, $Q \bullet D'$ és diagonal. Llavors, podem concloure que Q ha de ser una matriu de permutació llevat de signes? Per a més detalls sobre aquest enfocament, vegeu [55].

EXEMPLE 5. Considerem el cas $r = 3, m = 2$. La condició que $Q \bullet D'$ sigui diagonal per a un tensor diagonal D' es pot expressar pel sistema d'equacions cúbic en les entrades de Q :

$$\begin{aligned} (Q \bullet D')_{112} &= Q_{11}^2 Q_{21} D'_{111} + Q_{12}^2 Q_{22} D'_{222} = 0, \\ (Q \bullet D')_{122} &= Q_{11} Q_{21}^2 D'_{111} + Q_{12} Q_{22}^2 D'_{222} = 0. \end{aligned}$$

En forma matricial això és

$$Q \cdot \begin{bmatrix} Q_{11} & 0 \\ 0 & Q_{22} \end{bmatrix} \cdot \begin{bmatrix} Q_{21} & 0 \\ 0 & Q_{12} \end{bmatrix} \cdot \begin{bmatrix} D'_{111} \\ D'_{222} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Com que Q és ortogonal, cadascuna de les matrius diagonals anteriors han de ser o bé la matriu de zeros, o bé invertibles. Si alguna de les dues és idènticament zero, aleshores Q ha de ser una matriu de permutació llevat de signes, i l'equació se satisfà trivialment. Si ambdues matrius són invertibles, l'equació només es pot satisfer si $D'_{111} = D'_{222} = 0$, i en aquest cas D' es correspon amb el tensor idènticament zero. Això ens mostra que, mentre D' sigui un tensor diagonal diferent de zero, $Q \bullet D'$ és diagonal només si Q és una matriu de permutació signada.

En la majoria d'aplicacions suposem que ε té components independents, però també és possible relaxar aquestes condicions (per exemple, en el model de variància comuna). Per a més motivació i resultats bàsics, vegeu [55]. Addicionalment, el cas sobredeterminat on A és una matriu rectangular és altament rellevant a la pràctica; vegeu [91] per als detalls.

Els models de la forma (19) també s'han estudiat extensivament en el context d'anàlisi de la causalitat, i, en aquest àmbit, la matriu A pren una estructura especial. Pearl ([62]) considera el model següent d'*equacions estructurals* definit sobre un graf dirigit:

$$X_i = f_i(\mathbf{pa}(i), \varepsilon_i), \quad i = 1, \dots, m,$$

és a dir que cada variable del sistema es pot expressar com una funció dels seus nodes predecessors i un terme d'error ε_i . Aquest sistema és no lineal i no paramètric, però generalitza la versió lineal àmpliament utilitzada en els camps de l'economia i les ciències socials que hem presentat a (9). Per al cas del model lineal, es pot escriure de forma compacta com a

$$X = \Lambda X + \varepsilon,$$

on $\Lambda_{ij} = 0$ tret que $j \rightarrow i$ sigui una aresta de $\mathcal{G}$. Prenent $A = (I_m - \Lambda)^{-1}$, recuperem el model que teníem a (19).

Donat que, en aquest context, la matriu A té una estructura particular, podríem aspirar a identificar-la directament amb la matriu de covariàncies de X . Per a resultats en aquesta direcció, vegeu [33, 35, 38]. L'exemple següent ha estat adaptat de [32].

EXEMPLE 6. Fumar durant l'embaràs afecta el pes en néixer del nadó? Per respondre aquesta pregunta, suposem que un estudi registra el nivell de tabaquisme matern durant l'embaràs (X_1) i el pes en néixer d'un nadó (X_2). Suposant que existeix un efecte causal del tabaquisme sobre el pes en néixer, pretenem quantificar aquest efecte. A la pràctica, se centren les dades i es fa servir un model lineal:

$$X_1 = \varepsilon_1, \quad X_2 = \lambda_{12}X_1 + \varepsilon_2;$$

l'objectiu és inferir l'efecte λ_{12} . Si $\varepsilon_1, \varepsilon_2$ són variables no correlacionades de mitjana 0, podem fer servir que $\text{Cov}(X_1, X_2) = \lambda_{12} \text{Var}(X_1)$ per recuperar λ_{12} a partir de la distribució observada.

Una altra generalització important és la d'incloure variables latents a (9) per modelitzar potencials variables de confusió. Aquestes poden distorsionar les relacions de causalitat observades entre les components de X ; vegeu [8, 86]. L'exemple a continuació ha estat adaptat també de [32].

EXEMPLE 7. L'efecte λ_{12} de l'exemple anterior pot no ser directe i pot veure's distorsionat per altres variables de confusió latents. Per exemple, factors de predisposició genètica o socioeconòmics poden afectar tant l'hàbit de fumar com el pes en néixer dels nadons. En aquesta situació, la inferència sobre el quocient $\text{Cov}(X_1, X_2) / \text{Var}(X_1)$ no seria vàlida. Una possible solució és la d'introduir una «variable instrumental» que influènciï directament el nivell de tabaquisme però no el pes en néixer, i que no estigui influenciada per la variable latent. Seguint la bibliografia, es pot emprar una variable instrumental X_3 que mesuri els impostos sobre el tabac. Les equacions lineals que defineixen el model amb la variable latent H vindrien donades per:

$$X_1 = \lambda_{01} + \lambda_{31}X_3 + \lambda_{H1}H + \varepsilon_1,$$

$$X_2 = \lambda_{02} + \lambda_{12}X_1 + \lambda_{H2}H + \varepsilon_2,$$

$$X_3 = \lambda_{03} + \varepsilon_3,$$

$$H = \lambda_{0H} + \varepsilon_H,$$

on els errors $\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_H\}$ són independents dos a dos. D'aquí se segueix que l'equació $\text{Cov}(X_2, X_3) = \lambda_{12} \text{Cov}(X_1, X_3)$ se satisfà per a aquest model i, sempre que $\text{Cov}(X_1, X_3) \neq 0$, podrem deduir propietats de λ_{12} a partir del quocient $\text{Cov}(X_2, X_3) / \text{Cov}(X_1, X_3)$.

Un altre problema conegut en l'anàlisi de causalitat és si l'estructura del graf es pot recuperar a partir de les observacions de X . Un altre cop, això requereix moments d'ordre superior de les dades, la no gaussianitat, i els resultats d'identificabilitat com els que hem presentat per ICA; vegeu [77].

Referències

- [1] ALLMAN, E. S.; MATIAS, C.; RHODES, J. A. «Identifiability of parameters in latent structure models with many observed variables». *Ann. Statist.*, 37 (6A) (2009), 3099–3132.
- [2] ALLMAN, E. S.; RHODES, J. A. «Phylogenetic ideals and varieties for the general Markov model». *Adv. in Appl. Math.*, 40 (2) (2008), 127–148.
- [3] ALLMAN, E. S.; RHODES, J. A.; STURMFELS, B.; ZWIERNIK, P. «Tensors of nonnegative rank two». *Linear Algebra Appl.*, 473 (2015), 37–53.
- [4] AMÉNDOLA, C.; CASANELLAS, M.; GARCÍA PUENTE, L. D. «Tapas of algebraic statistics». *Notices Amer. Math. Soc.*, 65 (8) (2018), 936–938.
- [5] AMÉNDOLA, C.; FAUGÈRE, J.-C.; STURMFELS, B. «Moment varieties of Gaussian mixtures». *J. Algebr. Stat.*, 7 (1) (2016), 14–28.

- [6] ANANDKUMAR, A.; GE, R.; HSU, D.; KAKADE, S. M.; TELGARSKY, M. «Tensor decompositions for learning latent variable models». *J. Mach. Learn. Res.*, 15 (2014), 2773–2832.
- [7] ANANDKUMAR, A.; HSU, D.; KAKADE, S. M. «A method of moments for mixture models and hidden Markov models». A: *Proceedings of the 25th Annual Conference on Learning Theory*. Vol. 23 (2012), 33.1–33.34.
- [8] BARBER, R. F.; DRTON, M.; STURMA, N.; WEIHS, L. «Half-trek criterion for identifiability of latent variable models». *Ann. Statist.*, 50 (6) (2022), 3174–3196.
- [9] BERTOLINI, M.; TURRINI, C. «Problems and related results in algebraic vision and multiview geometry». *Rend. Circ. Mat. Palermo* (2), 73 (6) (2024), 2205–2231.
- [10] BISHOP, C. M.; TIPPING, M. E. «A hierarchical latent variable model for data visualization». *IEEE Trans. Pattern Anal. Mach. Intell.*, 20 (3) (1998), 281–293.
- [11] BLEI, D. M. «Probabilistic topic models». *Comm. ACM*, 55 (4) (2012), 77–84.
- [12] BLEI, D. M.; NG, A. Y.; JORDAN, M. I. «Latent Dirichlet allocation». *J. Mach. Learn. Res.*, 3 (2003), 993–1022.
- [13] BOCCI, C.; CHIANTINI, L. *An Introduction to Algebraic Statistics with Tensors*. La Matematica per il 3+2. Cham: Springer, 2019. (Unitext; 118)
- [14] BRO, R.; DE JONG, S. «A fast non-negativity-constrained least squares algorithm». *Journal of Chemometrics*, 11 (5) (1997), 393–3401.
- [15] BRO, R.; SIDIROPOULOS, N. D. «Least squares algorithms under unimodality and non-negativity constraints». *Journal of Chemometrics*, 12 (4) (1998), 223–247.
- [16] CALAMANTE, F.; MØRUP, M.; HANSEN, L. K. «Defining a local arterial input function for perfusion MRI using independent component analysis». *Magnetic Resonance in Medicine*, 52 (4) (2004), 789–797.
- [17] CASANELLAS, M. «Phylogenetic reconstruction based on Algebra». A: *Cutting-Edge Mathematics*. Cham: Springer, 2024, 26–44. (RSME Springer Ser.; 13)
- [18] CASANELLAS, M.; FERNÁNDEZ-SÁNCHEZ, J.; GARROTE-LÓPEZ, M. «SAQ: semi-algebraic quartet reconstruction». *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 18 (6) (2021), 2855–2861.
- [19] CASTRO, R.; COATES, M.; LIANG, G.; NOWAK, R.; YU, B. «Network tomography: recent developments». *Statist. Sci.*, 19 (3) (2004), 499–517.
- [20] CHANG, J. T. «Full reconstruction of Markov models on evolutionary trees: identifiability and consistency». *Math. Biosci.*, 137 (1) (1996), 51–73.
- [21] CHICKERING, D. M. «A transformational characterization of equivalent Bayesian network structures». A: *Uncertainty in Artificial Intelligence* (Montreal, PQ, 1995). San Francisco, CA: Morgan Kaufmann, 1995, 87–98.

- [22] CHIFMAN, J.; KUBATKO, L. «Quartet inference from SNP data under the coalescent model». *Bioinformatics*, 30 (23) (2014), 3317–3324.
- [23] CHOI, H.; BARANIUK, R. G. «Multiscale image segmentation using wavelet-domain hidden Markov models». *IEEE Trans. Image Process.*, 10 (9) (2001), 1309–1321.
- [24] CHOI, M. J.; LIM, J. J.; TORRALBA, A.; WILLSKY, A. S. «Exploiting hierarchical context on a large database of object categories». A: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Francisco, CA: IEEE, 2010, 129–136.
- [25] CILIBERTO, C.; CUETO, M. A.; MELLA, M.; RANESTAD, K.; ZWIERNIK, P. «Cremona linearizations of some classical varieties». A: *From Classical to Modern Algebraic Geometry*. Cham: Birkhäuser/Springer, 2016, 375–407. (Trends Hist. Sci.)
- [26] COMON, P. «Independent component analysis, a new concept?». *Signal Process.*, 36 (3) (1994), 287–314.
- [27] COMON, P.; JUTTEN, C. (ED.). *Handbook of Blind Source Separation. Independent Component Analysis and Applications*. Academic Press, 2010.
- [28] CONG, F.; LIN, Q.-H.; KUANG, L.-D.; GONG, X.-F.; ASTIKAINEN, P.; RISTANIEMI, T. «Tensor decomposition of EEG signals: A brief review». *Journal of Neuroscience Methods*, 248 (2015), 59–69.
- [29] CROUSE, M. S.; NOWAK, R. D.; BARANIUK, R. G. «Wavelet-based statistical signal processing using hidden Markov models». *IEEE Trans. Signal Process.*, 46 (4) (1998), 886–902.
- [30] DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. «Maximum likelihood from incomplete data via the EM algorithm». *Amb discussió. J. Roy. Statist. Soc. Ser. B*, 39 (1) (1977), 1–38.
- [31] DEVELIN, M.; SULLIVANT, S. «Markov bases of binary graph models». *Ann. Comb.*, 7 (4) (2003), 441–466.
- [32] DRTON, M. «Algebraic problems in structural equation modeling». A: *The 50th Anniversary of Gröbner Bases*. Tòquio: Mathematical Society of Japan, 2018, 35–86. (Adv. Stud. Pure Math.; 77)
- [33] DRTON, M.; FOYGE, R.; SULLIVANT, S. «Global identifiability of linear structural equation models». *Ann. Statist.*, 39 (2) (2011), 865–886.
- [34] DRTON, M.; STURMFELS, B.; SULLIVANT, S. *Lectures on Algebraic Statistics*. Basilea: Birkhäuser Verlag, 2009. (Oberwolfach Semin.; 39)
- [35] DRTON, M.; WEIHS, L. «Generic identifiability of linear structural equation models by ancestor decomposition». *Scand. J. Stat.*, 43 (4) (2016), 1035–1045.
- [36] ERIKSSON, B.; DASARATHY, G.; BARFORD, P.; NOWAK, R. «Toward the practical use of network tomography for internet topology discovery». A: *Proceedings IEEE INFOCOM*. San Diego, CA: IEEE, 2010, 1–9.

- [37] FERNÁNDEZ-SÁNCHEZ, J.; CASANELLAS, M. «Invariant versus classical quartet inference when evolution is heterogeneous across sites and lineages». *Systematic Biology*, 65 (2) (2016), 280–291.
- [38] FOYGEL, R.; DRAISMA, J.; DRTON, M. «Half-trek criterion for generic identifiability of linear structural equation models». *Ann. Statist.*, 40 (3) (2012), 1682–1713.
- [39] FREUND, Y.; HAUSSLER, D. «Unsupervised learning of distributions on binary vectors using two layer networks». A: *Advances in Neural Information Processing Systems*. Vol. 4. Morgan-Kaufmann, 1991, 912–919.
- [40] GARCIA, L. D.; STILLMAN, M.; STURMFELS, B. «Algebraic geometry of Bayesian networks». *J. Symbolic Comput.*, 39 (3-4) (2005), 331–355.
- [41] HARSHMAN, R. A.; LUNDY, M. E. «PARAFAC: Parallel factor analysis». *Comput. Statist. Data Anal.*, 18 (1) (1994), 39–72.
- [42] HARTLEY, R.; ZISSERMAN, A. *Multiple View Geometry in Computer Vision*. 2a ed. Amb un pròleg d'Olivier Faugeras. Cambridge: Cambridge University Press, 2003.
- [43] HITCHCOCK, F. L. «The expression of a tensor or a polyadic as a sum of products». *J. Math. and Phys.*, 6 (1-4) (1927), 164–189.
- [44] JANSON, S. «Normal convergence by higher semi-invariants with applications to sums of dependent random variables and random graphs». *Ann. Probab.*, 16 (1) (1988), 305–312.
- [45] JANZAMIN, M.; GE, R.; KOSSAIFI, J.; ANANDKUMAR, A. «Spectral learning on matrices and tensors». *Foundations and Trends® in Machine Learning*, 12 (5-6) (2019), 393–536.
- [46] JI, Y.; WANG, Q.; LI, X.; LIU, J. «A survey on tensor techniques and applications in machine learning». *IEEE Access*, 7 (2019), 162950–162990.
- [47] KRUSKAL, J. B. «Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics». *Linear Algebra Appl.*, 18 (2) (1977), 95–138.
- [48] LANDSBERG, J. M. *Tensors: Geometry and Applications*. Providence, RI: American Mathematical Society, 2012. (Grad. Stud. Math.; 128)
- [49] LAURITZEN, S. L. *Graphical Models*. Oxford Sci. Publ. Nova York: The Clarendon Press, Oxford University Press, 1996. (Oxford Statist. Sci. Ser.; 17)
- [50] LAWRENCE, N. D. «Gaussian process latent variable models for visualisation of high dimensional data». A: *NIPS'03: Proceedings of the 16th International Conference on Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2004, 329–336.
- [51] LUKACS, E. «Some extensions of a theorem of Marcinkiewicz». *Pacific J. Math.*, 8 (1958), 487–501.

- [52] MAKHIJANI, M. K.; BALU, N.; YAMADA, K.; YUAN, C.; NAYAK, K. S. «Accelerated 3D MERGE carotid imaging using compressed sensing with a hidden Markov tree model». *Journal of Magnetic Resonance Imaging*, 36 (5) (2012), 1194–1202.
- [53] MARCINKIEWICZ, J. «Sur une propriété de la loi de Gauß». *Math. Z.*, 44 (1) (1939), 612–618.
- [54] MCCULLAGH, P. *Tensor Methods in Statistics. Monographs on Statistics and Applied Probability*. Chapman and Hall/CRC, 2018.
- [55] MESTERS, G.; ZWIERNIK, P. «Nonindependent components analysis». *Ann. Statist.*, 52 (6) (2024), 2506–2528.
- [56] MICHAŁEK, M.; STURMFELS, B. *Invitation to Nonlinear Algebra*. Providence, RI: American Mathematical Society, 2021. (Grad. Stud. Math.; 211)
- [57] MONTÚFAR, G. «Restricted Boltzmann machines: introduction and review». A: *Information Geometry and its Applications*. Cham: Springer, 2018, 75–115. (Springer Proc. Math. Stat.; 252)
- [58] MONTÚFAR, G.; MORTON, J. «Discrete restricted Boltzmann machines». *J. Mach. Learn. Res.*, 16 (2015), 653–672.
- [59] MOURAD, R.; SINOQUET, C.; ZHANG, N. L.; LIU, T.; LERAY, P. «A survey on latent tree models and applications». *J. Artificial Intelligence Res.*, 47 (2013), 157–203.
- [60] NAFEEES, S.; RICE, S. H.; WAKEMAN, C. A. «Analyzing genomic data using tensor-based orthogonal polynomials with application to synthetic RNAs». *NAR Genomics and Bioinformatics*, 2 (4) (2020), lqaa101.
- [61] PACHTER, L.; STURMFELS, B. (ED.). *Algebraic Statistics for Computational Biology*. Nova York: Cambridge University Press, 2005.
- [62] PEARL, J. *Causality. Models, Reasoning, and Inference*. Cambridge: Cambridge University Press, 2000.
- [63] PEARSON, K. «Contributions to the mathematical theory of evolution». *Philos. Trans. Roy. Soc. London Ser. A*, 185 (1894), 71–110.
- [64] PEARSON, K. «Method of moments and method of maximum likelihood». *Biometrika*, 28 (1/2) (1936), 34–59.
- [65] PFEIFFER, M.; BETIZEAU, M.; WALTISPURGER, J.; PFISTER, S. S.; DOUGLAS, R. J.; KENNEDY, H.; DEHAY, C. «Unsupervised lineage-based characterization of primate precursors reveals high proliferative and morphological diversity in the OSVZ». *Journal of Comparative Neurology*, 524 (3) (2016), 535–563.
- [66] PISTONE, G.; RICCOMAGNO, E.; WYNN, H. P. *Algebraic Statistics. Computational Commutative Algebra in Statistics*. Boca Raton, FL: Chapman & Hall/CRC, 2001. (Monogr. Statist. Appl. Probab.; 89)
- [67] QI, Y.; COMON, P.; LIM, L.-H. «Semialgebraic geometry of nonnegative tensor rank». *SIAM J. Matrix Anal. Appl.*, 37 (4) (2016), 1556–1580.

- [68] RABINER, L. R. «A tutorial on hidden Markov models and selected applications in speech recognition». A: *Proceedings of the IEEE*, 77 (2) (1989), 257–286.
- [69] RICCI, M. M. G.; LEVI-CIVITA, T. «Méthodes de calcul différentiel absolu et leurs applications». *Math. Ann.*, 54 (1-2) (1900), 125–201.
- [70] RINGE, D.; WARNOW, T.; TAYLOR, A. «Indo-European and computational cladistics». *Transactions of the Philological Society*, 100 (1) (2002), 59–129.
- [71] ROBEVA, E.; SEIGAL, A. «Duality of graphical models and tensor networks». *Inf. Inference*, 8 (2) (2019), 273–288.
- [72] ROMBERG, J. K.; CHOI, H.; BARANIUK, R. G. «Bayesian tree-structured image modeling using wavelet-domain hidden Markov models». *IEEE Trans. Image Process.*, 10 (7) (2001), 1056–1068.
- [73] RUFFINI, M.; CASANELLAS, M.; GAVALDÀ, R. «A new method of moments for latent variable models». *Mach. Learn.*, 107 (8-10) (2018), 1431–1455.
- [74] SCHREIBER, J.; DURHAM, T.; BILMES, J.; NOBLE, W. S. «Avocado: a multi-scale deep tensor factorization method learns a latent representation of the human epigenome». *Genome Biology*, 21, article núm. 81 (2020).
- [75] SEIGAL, A.; MONTÚFAR, G. «Mixtures and products in two graphical models». *J. Algebr. Stat.*, 9 (1) (2018), 1–20.
- [76] SHIERS, N.; ASTON, J. A. D.; SMITH, J. Q.; COLEMAN, J. S. «Gaussian tree constraints applied to acoustic linguistic functional data». *J. Multivariate Anal.*, 154 (2017), 199–215.
- [77] SHIMIZU, S.; INAZUMI, T.; SOGAWA, Y.; HYVÄRINEN, A.; KAWAHARA, Y.; WASHIO, T.; HOYER, P. O.; BOLLEN, K. «DirectLiNGAM: a direct method for learning a linear non-Gaussian structural equation model». *J. Mach. Learn. Res.*, 12 (2011), 1225–1248.
- [78] SMOLENSKY, P. «Information processing in dynamical systems: Foundations of harmony theory». *Parallel Distributed Process*, 1 (1986).
- [79] SPEED, T. P. «Cumulants and partition lattices». *Austral. J. Statist.*, 25 (2) (1983), 378–388.
- [80] STEEL, M. *Phylogeny—Discrete and Random Processes in Evolution*. Filadelfia, PA: Society for Industrial and Applied Mathematics (SIAM), 2016. (CBMS-NSF Regional Conf. Ser. in Appl. Math.; 89)
- [81] STURMFELS, B.; TIMME, S.; ZWIERNIK, P. «Estimating linear covariance models with numerical nonlinear algebra». *Algebr. Stat.*, 11 (1) (2020), 31–52.
- [82] STURMFELS, B.; UHLER, C. «Multivariate Gaussian, semidefinite matrix completion, and convex algebraic geometry». *Ann. Inst. Statist. Math.*, 62 (4) (2010), 603–638.
- [83] STURMFELS, B.; ZWIERNIK, P. «Binary cumulant varieties». *Ann. Comb.*, 17 (1) (2013), 229–250.

- [84] SULLIVANT, S. «Algebraic geometry of Gaussian Bayesian networks». *Adv. in Appl. Math.*, 40 (4) (2008), 482–513.
- [85] SULLIVANT, S. *Algebraic Statistics*. Providence, RI: American Mathematical Society, 2018. (Grad. Stud. Math.; 194)
- [86] TRAMONTANO, D.; DRTON, M.; ETESAMI, J. «Parameter identification in linear non-Gaussian causal models under general confounding». Preprint (2024). [Disponible en línia a: <https://arxiv.org/abs/2405.20856>]
- [87] UHLER, C. «Geometry of maximum likelihood estimation in Gaussian graphical models». *Ann. Statist.*, 40 (1) (2012), 238–261.
- [88] VAVASIS, S. A. «On the complexity of nonnegative matrix factorization». *SIAM J. Optim.*, 20 (3) (2009), 1364–1377.
- [89] VERMA, T.; PEARL, J. «Equivalence and synthesis of causal models». A: *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*. Nova York: Elsevier Science Inc., 1990, 220–227.
- [90] WAINWRIGHT, M. J.; JORDAN, M. I. «Graphical models, exponential families, and variational inference». *Foundations and Trends® in Machine Learning*, 1 (1-2) (2008), 1–305.
- [91] WANG, K.; SEIGAL, A. «Identifiability of overcomplete independent component analysis». Preprint (2024). [Disponible en línia a: <https://arxiv.org/abs/2401.14709>]
- [92] WATANABE, S. *Algebraic Geometry and Statistical Learning Theory*. Cambridge: Cambridge University Press, 2009. (Cambridge Monogr. Appl. Comput. Math.; 25)
- [93] WILLSKY, A. S. «Multiresolution Markov models for signal and image processing». A: *Proceedings of the IEEE*, 90 (8) (2002), 1396–1458.
- [94] YANG, K. D.; KATCOFF, A.; UHLER, C. «Characterizing and learning equivalence classes of causal DAGs under interventions». *Proceedings of Machine Learning Research*, 80 (2018), 5537–5546.
- [95] ŽABOKRTSKÝ, Z.; POPEL, M. «Hidden Markov tree model in dependency-based machine translation». A: *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, 2009, 145–148.
- [96] ZHANG, N. L. «Hierarchical latent class models for cluster analysis». *J. Mach. Learn. Res.*, 5 (2003/04), 697–723.
- [97] ZHANG, Y.; CHEN, X.; ZHOU, D.; JORDAN, M. I. «Spectral methods meet EM: a provably optimal algorithm for crowdsourcing». *J. Mach. Learn. Res.*, 17 (2016), article núm. 102, 44 p.
- [98] ZWIERNIK, P. « L -cumulants, L -cumulant embeddings and algebraic statistics». *J. Algebr. Stat.*, 3 (1) (2012), 11–43.
- [99] ZWIERNIK, P. *Semialgebraic Statistics and Latent Tree Models*. Boca Raton, FL: Chapman & Hall/CRC, 2016. (Monogr. Statist. Appl. Probab.; 146)

- [100] ZWIERNIK, P.; UHLER, C.; RICHARDS, D. «Maximum likelihood estimation for linear Gaussian covariance models». *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79 (4) (2017), 1269-1292.

LUIS SIERRA
DEPARTMENT OF STATISTICAL SCIENCES
UNIVERSITY OF TORONTO
luis.sierra@mail.utoronto.ca

MARTA CASANELLAS
DEPARTAMENT DE MATEMÀTIQUES
UNIVERSITAT POLITÈCNICA DE CATALUNYA
marta.casanellas@upc.edu

PIOTR ZWIERNIK
DEPARTMENT OF ECONOMICS AND BUSINESS
UNIVERSITAT POMPEU FABRA
piotr.zwiernik@upf.edu

English summaries

Jaume Agudé

A review of rotations in low dimensions

We begin by discussing rotations in linear geometry and orthogonal matrices, and go on to present explicit, elementary and self-contained constructions of the classical isomorphisms between spinor groups in dimensions < 7 and certain unitary or symplectic groups. The paper is intended to serve as lecture notes for an introductory course on compact Lie groups.

Keywords: rotation, spinor, Lie group, quaternions.

MSC2020 Subject Classification: 22-01, 20G20, 22E15.

Jaume Coll Guerrero

Polyhedra and modular origami

This paper presents modular origami techniques that allow us to create a wide variety of polyhedra from paper: the five Platonic solids, the thirteen Archimedean polyhedra, the eight convex deltahedra and their stellations, prisms, antiprisms, pyramids and Fullerian geodesic domes, among others.

Keywords: modular origami, polyhedra, Platonic solids, Archimedean polyhedra, origami techniques, tessellations.

MSC2020 Subject Classification: 00A08, 00A66, 05B45, 51M20, 52B05, 52B10.

Xavier Guitart and Marc Masdeu

Rational points on elliptic curves

Elliptic curves are currently among the most extensively studied objects in number theory. They can be described by cubic equations in two variables, but what sets them apart – and makes them fascinating – is the rich algebraic structure exhibited by their solutions. The aim of this article is to explain what elliptic curves are and to explore their most significant properties, some of which rank among the most important results in 20th- and 21st-century mathematics. We will also discuss several open conjectures that continue to shape current research. To place elliptic curves within a broader historical and conceptual context, we present them as a specific class of Diophantine equations – a recurrent and central theme in number theory.

Keywords: Diophantine equations, elliptic curves, Heegner points, p -adic integration, Stark-Heegner points.

MSC2020 Subject Classification: 14H52, 11G05.

Jaume de Haro

The Friedmann equations: Newtonian mechanics vs. general relativity

In this paper, we present a heuristic derivation of the perturbed Friedmann equations based on Newtonian mechanics, offering an intuitive perspective on these cosmological relations. This approach highlights the remarkable agreement between Newtonian and relativistic descriptions within certain limits, serving both as a conceptual – and even philosophical – bridge, and as a pedagogical tool to deepen understanding of the physical principles governing the evolution of the universe. By simplifying the complex mathematical framework of general relativity, this method provides an accessible means of exploring the dynamics of the universe. It underscores how classical mechanics, when applied carefully, can offer insights that align with relativistic results in the linear approximation. This connection not only enhances our understanding of the universality of physical laws but also reinforces the idea that relativistic phenomena can often be interpreted through familiar classical analogues.

Keywords: Friedmann equations, cosmological perturbations, Newtonian mechanics, general relativity.

MSC2020 Subject Classification: 83C05, 83C99, 70F99, 83F05.

Luis Sierra, Marta Casanellas and Piotr Zwiernik

Tensors in algebraic statistics

Tensors are ubiquitous in statistics and data analysis. The central object that links data science to tensor theory and algebra is that of a model with latent variables. We provide an overview of tensor theory, with particular emphasis on its applications in algebraic statistics. This high-level overview is accompanied by numerous examples to illustrate key concepts. Additionally, an extensive literature review is included to guide readers towards more detailed studies on the subject.

Keywords: tensors, algebraic statistics, graphical models, tensor decomposition.

MSC2020 Subject Classification: 14M99, 15A69, 62H22, 62R01, 62R07.

Instruccions per als autors

Els articles sotmesos a publicació s'han d'enviar als editors o a qualsevol membre del comitè editorial per correu electrònic, preferentment en format PDF. Els originals han de contenir la versió anglesa del títol, un resum breu en català i en anglès, paraules clau en català i en anglès i els codis de la classificació per matèries MSC2020.

Les versions definitives dels articles acceptats s'han de presentar en codi $\text{\TeX}$, preferentment en l'estil $\text{\LaTeX}$ propi del BUTLLETÍ. Aquest estil es pot obtenir a les pàgines web de la Societat Catalana de Matemàtiques (SCM). Es recomana una extensió d'entre 15 i 35 pàgines. Fem notar que en aquesta publicació s'utilitza preferentment el punt per a separar decimals, en lloc de la coma recomanada per l'IEC, per a poder facilitar la comprensió de les expressions matemàtiques. Per tal d'accelerar el procés de producció, es prega als autors que segueixin les indicacions contingudes en el document d'exemple.

La versió en paper del BUTLLETÍ s'imprimeix en blanc i negre. Quan un article contingui figures en color i es consideri convenient, l'autor proporcionarà una versió dels gràfics substituint el color per tons de grisos i línies de gruix variable. Així mateix, modificarà els comentaris que facin referència al color de les figures. En qualsevol cas, el BUTLLETÍ publicarà l'original en color en el seu format electrònic.

La propietat intel·lectual dels articles és dels respectius autors.

Els autors, en el moment de lliurar els articles al BUTLLETÍ per a sol·licitar-ne la publicació, accepten els termes següents:

- Els autors cedeixen a la SCM (filial de l'Institut d'Estudis Catalans) els drets de reproducció, comunicació pública i distribució dels articles presentats per a ser publicats al BUTLLETÍ.

- Els autors responen davant la SCM de l'autoria i l'originalitat dels articles presentats.

- És responsabilitat dels autors l'obtenció dels permisos per a la reproducció de tot el material gràfic inclòs en els articles.

- La SCM està exempta de tota responsabilitat derivada de l'eventual vulneració de drets de propietat intel·lectual per part dels autors.

Cada autor rebrà una còpia en PDF d'alta qualitat de la versió digital del seu article i un exemplar imprès del número del BUTLLETÍ en el qual es publiqui.

La correspondència administrativa relacionada amb el BUTLLETÍ s'ha d'adreçar a la SCM.

Pel que fa a la protecció de dades personals, l'Institut d'Estudis Catalans (IEC) compleix el que estableix el Reglament general de protecció de dades de la Unió Europea (Reglament 2016/679, del 27 d'abril de 2016). De conformitat amb aquesta norma, s'informa que, amb l'acceptació de les normes de publicació, els autors autoritzen que les seves dades personals (nom i cognoms, dades de contacte i dades de filiació) puguin ser publicades en el corresponent volum del BUTLLETÍ.

Aquestes dades seran incorporades a un tractament que és responsabilitat de l'IEC amb la finalitat de gestionar aquesta publicació. Únicament s'utilitzaran les dades dels autors per a gestionar la publicació del BUTLLETÍ i no seran cedides a tercers, ni es produiran transferències a tercers països o organitzacions internacionals. Un cop publicat el BUTLLETÍ, aquestes dades es conservaran com a part del registre històric d'autors. Els autors poden exercir els drets d'accés, rectificació, supressió, oposició, limitació en el tractament i portabilitat, adreçant-se per escrit a l'Institut d'Estudis Catalans (carrer del Carme, 47, 08001 Barcelona), o bé enviant un correu electrònic a l'adreça dades.personals@iec.cat, en què s'especifiqui de quina publicació es tracta.

Comitè editorial

Antoni Guillamon (editor en cap)
Departament de Matemàtiques
Universitat Politècnica de Catalunya
antoni.guillamon@upc.edu

Carme Cascante
Dep. de Matemàtiques i Informàtica
Universitat de Barcelona
cascante@ub.edu

Bartomeu Coll
Dep. de Matemàtiques i Informàtica
Universitat de les Illes Balears
tomeu.coll@uib.cat

Núria Fagella
Dep. de Matemàtiques i Informàtica
Universitat de Barcelona
fagella@maia.ub.es

Armengol Gasull
Departament de Matemàtiques
Universitat Autònoma de Barcelona
armengol.gasull@uab.cat

Gábor Lugosi
ICREA i Departament d'Economia
Universitat Pompeu Fabra
gabor.lugosi@upf.edu

Rosa Camps (editora adjunta)
Departament de Matemàtiques
Universitat Autònoma de Barcelona
rosa.camps@uab.cat

Marc Noy
Departament de Matemàtiques
Universitat Politècnica de Catalunya
marc.noy@upc.edu

Francesc Planas
Departament de Matemàtiques
Universitat Politècnica de Catalunya
francesc.planas@upc.edu

Joan Saldaña
Dep. d'Informàtica, Mat. Aplicada i Estadística
Universitat de Girona
joan.saldana@udg.edu

Marta Sanz-Solé
Dep. de Matemàtiques i Informàtica
Universitat de Barcelona
marta.sanz@ub.edu

Gil Solanes
Departament de Matemàtiques
Universitat Autònoma de Barcelona
gil.solanes@uab.cat



Societat Catalana de Matemàtiques

La **Societat Catalana de Matemàtiques (SCM)** és una societat filial de l'Institut d'Estudis Catalans que continua les activitats de la Secció de Matemàtiques de la Societat Catalana de Ciències, que fou fundada per l'Institut l'any 1931. Les finalitats de la **SCM** són: el conreu de les ciències matemàtiques, l'extensió del seu coneixement en la societat catalana, el foment del seu ensenyament i de la seva investigació teòrica i aplicada, així com la publicació de tota mena de treballs que s'adeqüin a aquests objectius. La **SCM** desenvolupa les seves activitats en les terres de llengua i cultura catalanes. El català és, doncs, la llengua pròpia de la **SCM** i la que és usada normalment en tots els seus actes i publicacions.

La **SCM** edita les publicacions periòdiques *SCM/Notícies*, *Butlletí de la Societat Catalana de Matemàtiques*, *NouBiaix* (en col·laboració amb la FEEMCAT) i *Reports@SCM*. Els socis de la **SCM** reben, gratuïtament, aquestes publicacions.

La **SCM** té convenis de reciprocitat amb diverses societats matemàtiques d'arreu del món, mitjançant els quals els socis de la **SCM** obtenen una reducció en la quota de soci d'aquestes societats. Així mateix, els socis de la **SCM** poden fer-se socis de la Societat Matemàtica Europea pagant una quota complementària.

La Junta Directiva de la **SCM** està constituïda per les persones següents:

PRESIDENTA: Montserrat Alsina i Aubach

VICEPRESIDENT: Josep Vives i Santa-Eulàlia

SECRETÀRIA: Margarida Mitjana i Riera

TRESORER: Albert Granados i Corsellas

VOCALS: Clara Mateo Campo, Ainoa Murillo López, Albert Ruiz Cirera,
David Virgili Correas

DELEGAT DE L'IEC: Joaquim Bruna i Floris

L'adreça de la **SCM** és carrer del Carme, 47, 08001 Barcelona. Telèfon: 933 248 583. Fax: 932 701 180. Correu electrònic: scm@iec.cat. Adreça web: <https://scm.iec.cat>.

El BUTLLETÍ DE LA SOCIETAT CATALANA DE MATEMÀTIQUES publica, en llengua catalana, exposicions matemàtiques de qualitat, que puguin interessar a un nombre elevat de lectors. Es donarà prioritat a aquells treballs en què destaquin la claredat d'exposició i l'interès general del tema. El BUTLLETÍ està obert a tots els camps de la matemàtica i també als aspectes matemàtics de les ciències experimentals, la tecnologia, l'economia, etc., així com a altres àrees, com la història, la didàctica i la filosofia, sempre que els treballs tinguin un component matemàtic important. També tenen cabuda al BUTLLETÍ aquells articles que desenvolupin un aspecte significatiu de la problemàtica de la professió matemàtica al nostre país.

El BUTLLETÍ publica un volum a l'any, dividit en dos números, que es trameten gratuïtament a tots els socis. El BUTLLETÍ es publica també en format electrònic. L'edició electrònica del BUTLLETÍ pot obtenir-se des del portal de revistes científiques en línia de l'IEC o al servidor <https://scm.iec.cat>.

La correspondència administrativa s'ha d'adreçar a la Societat Catalana de Matemàtiques.

Editor en cap _____

Antoni Guillamon

Departament de Matemàtiques

Universitat Politècnica de Catalunya

Editora adjunta _____

Rosa Camps

Departament de Matemàtiques

Universitat Autònoma de Barcelona

Comitè Editorial _____

Carne Cascante

Dep. de Matemàtiques i Informàtica

Universitat de Barcelona

Bartomeu Coll

Dep. de Matemàtiques i Informàtica

Universitat de les Illes Balears

Núria Fagella

Dep. de Matemàtiques i Informàtica

Universitat de Barcelona

Armengol Gasull

Departament de Matemàtiques

Universitat Autònoma de Barcelona

Gábor Lugosi

ICREA i Departament d'Economia

Universitat Pompeu Fabra

Marc Noy

Departament de Matemàtiques

Universitat Politècnica de Catalunya

Francesc Planas

Departament de Matemàtiques

Universitat Politècnica de Catalunya

Joan Saldaña

Dep. d'Informàtica, Mat. Aplicada i Estadística

Universitat de Girona

Marta Sanz-Solé

Dep. de Matemàtiques i Informàtica

Universitat de Barcelona

Gil Solanes

Departament de Matemàtiques

Universitat Autònoma de Barcelona

Societat Catalana de Matemàtiques

Carrer del Carme, 47 - 08001 Barcelona

tel. 933 248 583 - fax 932 701 180

scm@iec.cat - <https://scm.iec.cat>