

BUTLLETÍ

DE LA SOCIETAT CATALANA
DE MATEMÀTIQUES
Institut d'Estudis Catalans

_ í n d e x _

- ☐ *Jordi Delgado i Enric Ventura*
Autòmats de Stallings, un camí d'anada i tornada 5
- ☐ *Gábor Lugosi*
Estimació de la mitjana: una revisió d'avenços recents 61
- ☐ English summaries 95



Volum 37 • Número 1 • Any 2022



Institut
d'Estudis
Catalans

BUTLLETÍ

DE LA SOCIETAT CATALANA

DE MATEMÀTIQUES

Institut d'Estudis Catalans

Volum 37 • Número 1 • Juny 2022

BARCELONA

2022

© 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, en les fotografies o en altres il·lustracions— 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 <http://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

JORDI DELGADO I ENRIC VENTURA	
Autòmats de Stallings, un camí d'anada i tornada	5
GÁBOR LUGOSI	
Estimació de la mitjana: una revisió d'avenços recents.....	61
English summaries	95

Autòmats de Stallings, un camí d'anada i tornada

JORDI DELGADO I ENRIC VENTURA

L'algèbre n'est qu'une géometrie écrite;
la géométrie n'est qu'une algèbre figurée.

Sophie Germain (1776–1831)

Resum: En aquest article revisem algunes de les propietats fonamentals del grup lliure i fem una exposició detallada de la teoria dels autòmats de Stallings, una interpretació geomètrica dels seus subgrups que ha estat (i segueix essent) immensament fructífera, tant com a mitjà per a entendre resultats clàssics com com a font de nous resultats. N'expliquem alguns dels més rellevants.

Paraules clau: grup lliure, subgrup, autòmat, Stallings, problema algorísmic, problema de decisió.

Classificació MSC2010: 20-02, 20E05, 20F05, 20F10, 20F65, 05C25.

1 Introducció

La relació entre l'àlgebra i la geometria és, probablement, una de les més fructíferes entre les diferents àrees matemàtiques. Val a dir que es tracta d'una relació bidireccional. Si bé inicialment l'àlgebra es va constituir com un potent recurs per a resoldre (entre d'altres) problemes geomètrics (per exemple, amb el treball d'Euclides al Llibre II dels *Elements*, molts segles més tard, de forma sistemàtica, amb el descobriment de Descartes de la geometria analítica, o, més modernament, amb el desenvolupament de la topologia algebraica), a mesura que aquesta guanyava en abstracció, i es constituïa en una disciplina matemàtica *per se*, s'ha anat configurant també una relació fructífera en sentit contrari: els arguments geomètrics i topològics han esdevingut un instrument essencial per a resoldre problemes d'origen i motivació algebraics.

Aquesta influència ha estat especialment acusada durant el darrer segle i escaig en l'àmbit de la teoria de grups infinits, on l'enfocament combinatori

(basat en la idea de presentació d'un grup) ha donat pas a una explosió de mètodes i arguments geomètrics que han acabat conformant l'àrea anomenada *teoria geomètrica de grups*.

El concepte de presentació d'un grup (vegeu la definició 20) es basa en el de *grup lliure*, una mena de llenç¹ on tots els grups poden ser dibuixats. No és, per tant, gens sorprenent la rellevància que ha tingut l'estudi del grup lliure en el desenvolupament de la teoria de grups, especialment dels grups infinits. Per als no avesats, convé recordar que el grup lliure admet una interpretació molt senzilla i intuïtiva com a «conjunt de paraules» usant un conjunt predefinit de lletres (vegeu la secció 2).

La concisió de la descripció combinatòria donada per les presentacions és especialment propícia per a la formulació de qüestions algorísmiques, que, durant el darrer segle, han esdevingut un tercer ingredient estretament lligat a l'àlgebra i la geometria dels grups.

Malgrat la seva aparent simplicitat, reforçada pel fet que els seus subgrups són també lliures (teorema de Nielsen-Schreier), les relacions entre els subgrups del grup lliure són intrincades i amaguen sorpreses interessants. Per exemple, veurem que —en contrast amb el que succeeix en els ambients abelians clàssics— el grup lliure de rang 2 té subgrups de qualsevol rang finit, i fins i tot de rang infinit. I estudiarem el comportament de les seves interseccions, que va donar lloc a la coneguda conjectura de Hanna Neumann, un problema obert durant més de mig segle, i resolt fa uns anys independentment per Friedman a [18] i per Mineyev a [36] (vegeu també les simplificacions remarcables de Dicks a [18, apèndix B] i [17]).

Tot i la innegable complexitat del seu reticle de subgrups, veurem que, des del punt de vista algorísmic, els grups lliures, en general, es comporten molt bé. Multitud de problemes algorísmics naturals (incloent-hi els problemes clàssics de Dehn, així com els de la pertinença,² la intersecció de subgrups o la finitud de l'índex) són decidibles al grup lliure $\mathbb{F}_n$; molts d'ells gràcies, precisament, a tècniques basades en els autòmats de Stallings. Vegeu [12] per a un inventari (no exhaustiu) de problemes decidibles usant autòmats de Stallings, els protagonistes d'aquest article. Cal assenyalar, però, que els comportaments difícils tampoc no estan lluny: és suficient fer un producte directe de dos grups lliures per començar a trobar problemes molt naturals que són algorísmicament indecidibles (vegeu [35]).

En aquest article començarem fent una breu introducció al grup lliure i les seves propietats principals. Tot seguit, exposem els fonaments de la teoria dels autòmats de Stallings, una interpretació geomètrica molt aclaridora dels subgrups del grup lliure que, a més, ha resultat immensament fructífera en termes d'aplicacions. A la secció 4 en revisarem algunes de les més importants.

¹ Com veurem a la secció 2, un full d'origami, del qual sorgeixen tots els altres grups després de «plegar-lo» adequadament, és potser una metàfora més acurada.

² En anglès, *Membership Problem* (MP).

1.1 Notació, terminologia i convenis

La major part de la notació i la terminologia utilitzades en aquest article són estàndard; tanmateix, a continuació aclarim alguns aspectes per evitar possibles confusions.

El conjunt dels números naturals, designat per $\mathbb{N}$, inclou el zero, i especifiquem condicions sobre ell mitjançant subíndexs; per exemple, designem per $\mathbb{N}_{\geq 1}$ el conjunt de números naturals estrictament positius. El cardinal d'un conjunt S es designa per $\#S$, i la notació $|\cdot|$ es reserva per a indicar longitud (en diferents contextos). Escrivim com a $[m, n] = \{k \in \mathbb{N} : m \leq k \leq n\}$ el conjunt de números naturals entre m i n (ambdós inclosos) i $[0, \aleph_0] = \mathbb{N} \cup \{\aleph_0\}$.

Designem per $\mathbb{F}$ un grup lliure genèric, mentre que les notacions $\mathbb{F}_A$ i $\mathbb{F}_\kappa$ (resp., $\mathbb{F}_n$) s'usen per a emfatitzar una base A i el rang κ (resp., n , si és finit) de $\mathbb{F}$, respectivament. Les primeres lletres de l'alfabet llatí ($a, b, c, \dots$) s'usen normalment per a designar els símbols dels alfabetos formals, mentre que les últimes ($u, v, w, \dots$) solen designar paraules formals o elements del grup lliure.

Les funcions actuen per la dreta. És a dir, designem per $(x)\varphi$ (o simplement per $x\varphi$) la imatge de l'element x per la funció φ , i designem per $\phi\psi$ la composició $A \xrightarrow{\phi} B \xrightarrow{\psi} C$. En conseqüència, escrivim, per exemple, $g^h = h^{-1}gh$ (el conjugat de g per h) i $[g, h] = g^{-1}h^{-1}gh$ (el commutador de g i h).

Designem la conjugació (d'elements o subgrups) per $\sim$, i escrivim $H \leq G$, $H \trianglelefteq G$, $H \trianglelefteq_{\text{fg}} G$, $H \trianglelefteq_{\text{ff}} G$, $H \trianglelefteq_{\text{fi}} G$, per designar que H és subgrup, subgrup normal, subgrup finitament generat, factor lliure, i subgrup d'índex finit (de G), respectivament.

2 Grup lliure

Comencem recordant les definicions de monoide i de grup. Un *monoide* és un parell $(G, \cdot)$ on G és un conjunt arbitrari, $\cdot : G \times G \rightarrow G$, $(g_1, g_2) \mapsto g_1 \cdot g_2$ és una operació associativa amb element neutre (que representarem per 1). Si, a més, tot element $g \in G$ té un (únic) invers g^{-1} (i. e., tal que $g \cdot g^{-1} = g^{-1} \cdot g = 1$) diem que $(G, \cdot)$ és un *grup*. Si $A \subseteq G$, designarem per A^{-1} el conjunt d'inversos d'elements de A , i. e., $A^{-1} = \{a^{-1} : a \in A\}$. Usualment ometrem el símbol per a l'operació i escriurem $g \cdot h = gh$. Si, a més, $gh = hg$ per a tot $g, h \in G$, direm que el monoide (o el grup) és *commutatiu* o *abelià*; en aquest cas l'operació se sol designar per «+» i l'element neutre per 0 (*notació additiva*).

Hi ha moltes maneres de definir els grups lliures i les seves bases: geomètriques, algebraiques, combinatòries, categòriques... La més concreta és, potser, la que adapta la noció estàndard de base de l'àlgebra lineal al context de grups, mentre que la més abstracta és probablement la categòrica.³ Al llarg d'aquesta secció les desenvoluparem totes dues i veurem com estan relacionades. Abans, però, emfasitzem el polimorfisme del grup lliure amb dues caracteritzacions addicionals, de caràcter geomètric, que enunciem sense demostració.

³ Vegeu https://en.wikipedia.org/wiki/Abstract_nonsense.

TEOREMA 1. *Sigui F un grup. Aleshores, els enunciats següents són equivalents:*

- (a) *F és un grup lliure;*
- (b) *F és el grup fonamental d'un graf connex;⁴*
- (c) *F actua lliurement i sense inversions sobre les arestes d'un arbre.⁵*

Tot i la indubtable rellevància d'aquest resultat, hem preferit fer una presentació «algebraica» del grup lliure, amb l'esperança que el lector la trobi més natural. Comencem notant que és immediat de la definició de grup que els factors trivials 1 i els productes gg^{-1} (anomenats *cancel·lacions*) són superflus en qualsevol producte $g_1 \cdot \dots \cdot g_n$ d'elements d'un grup G (per exemple, $h1k = hk$ i $hgg^{-1}k = hk$). Un producte en el qual no apareixen factors trivials ni cancel·lacions s'anomena *producte reduït* (incloent-hi el producte buit, que representa l'element neutre —també anomenat *element trivial* en aquest context). És clar que, eliminant repetidament factors trivials i cancel·lacions, tot producte es pot convertir en un producte reduït, amb el mateix resultat dins de G .

DEFINICIÓ 2. Sigui G un grup i $A \subseteq G$. Diem que A és un subconjunt *lliure* (o *independent*) en G si dos productes reduïts diferents d'elements de $A^\pm = A \cup A^{-1}$ sempre donen resultats diferents de G (o, equivalentment, si el buit és l'únic producte reduït que dona l'element trivial de G). Diem que A *genera* (o que és un *conjunt de generadors* per a) G si tot element de G és igual a un producte (que podem suposar reduït) d'elements de $A^\pm$. Finalment, diem que A és una *base* de G i que G és un grup lliure (sobre A) si A és lliure en G i genera G . Farem servir la notació $\mathbb{F}$ per a referir-nos als grups lliures.

DEFINICIÓ 3. S'anomena *rang* d'un grup G , designat per $\text{rk}(G)$, la mínima cardinalitat d'un conjunt de generadors per a G . Si un grup admet un conjunt de generadors finit, direm que és *finitament generat*.

EXEMPLE 4. Considerem el grup dels enters $\mathbb{Z}$ amb notació additiva.⁶ Els conjunts $\{1\}$ i $\{-1\}$ en són bases ja que, clarament, generen $\mathbb{Z}$ i cap expressió de la forma

$$1 + 1 + \dots + 1 \quad \text{o bé} \quad (-1) + (-1) + \dots + (-1)$$

amb $n \neq 0$ dona igual a zero. En particular, $\{1\}$ és un conjunt de generadors (òbviament mínim) de $\mathbb{Z}$, i per tant $\text{rk}(\mathbb{Z}) = 1$.

EXEMPLE 5. En canvi, $\{\bar{1}\} = \{1 + n\mathbb{Z}\}$ no és una base de $\mathbb{Z}/n\mathbb{Z}$, per a $n \geq 2$, ja que, tot i ser-ne generador, no és un subconjunt lliure, degut a la *relació* $\bar{1} + \dots + \bar{1} = \bar{0}$. De fet, cap subconjunt $A \subseteq \mathbb{Z}/n\mathbb{Z}$ no és base ja que el subconjunt buit no genera, i qualsevol element $\bar{a} \in \mathbb{Z}/n\mathbb{Z}$ compleix la relació: $\bar{a} + \dots + \bar{a} = \bar{0}$; per tant, el grup $\mathbb{Z}/n\mathbb{Z}$ no és lliure.

⁴ Vegeu la definició 33 per a la definició de grup fonamental, i l'observació 50 i la proposició 54(ii) per a una reformulació en termes d'autòmats.

⁵ Un arbre és un graf connex sense cicles. Vegeu [6, capítol 3].

⁶ En notació additiva, 1 és el generador natural del grup $(\mathbb{Z}, +)$ i no pas el neutre, designat per 0.

Noteu que el grup trivial és lliure (amb base $\emptyset$ i, per tant, amb rang 0) i, com hem vist a l'exemple anterior, el grup dels enters $\mathbb{Z}$ també és lliure (i té dues bases, $\{1\}$ i $\{-1\}$). És immediat de la definició que aquests són els únics grups lliures abelians. D'altra banda, hem vist que $\mathbb{Z}/n\mathbb{Z}$ no és lliure per a cap $n \geq 2$. Amb el mateix argument es pot demostrar que cap grup finit no és lliure, excepte el trivial.

OBSERVACIÓ 6. Si $A \subseteq \mathbb{F}$ és base d'un grup (lliure) $\mathbb{F}$, aleshores A és un sistema de generadors *minimal* de $\mathbb{F}$; és a dir, $\langle A \rangle = \mathbb{F}$, però $\langle A \setminus S \rangle \neq \mathbb{F}$ per a tot $\emptyset \neq S \subseteq A$.

PROVA. Si existís un $S \neq \emptyset$ tal que $A \setminus S$ generés tot $\mathbb{F}$, aleshores, per qualsevol $s \in S$, tindríem $s \in \mathbb{F} = \langle A \setminus S \rangle$, i l'element $s \in G$ el podríem obtenir com una lletra de A i, alhora, com un producte reduït en $A \setminus S$, contradient la hipòtesi que A és una família lliure. $\square$

Tot seguit construirem grups lliures amb bases de cardinal arbitrari.

DEFINICIÓ 7. Sigui A un conjunt (finit o infinit), que anomenarem *alfabet*, o conjunt de *lletres* elementals. Una *paraula* sobre A és una seqüència ordenada i finita de lletres, $a_1 a_2 \cdots a_n$, on $n \geq 0$, i $a_i \in A$, amb possibles repeticions. El nombre total de lletres n d'una paraula s'anomena *longitud*, i escrivim $|a_1 a_2 \cdots a_n| = n$. Com a conveni, designarem per 1 l'única paraula de longitud zero, o *paraula buida*. El conjunt de totes les paraules sobre A el designem per A^* .

OBSERVACIÓ 8. Per a tot alfabet A , el conjunt A^* és un monoide amb l'operació de concatenació, $u \cdot v = uv$, $u, v \in A^*$. A més, $|uv| = |u| + |v|$ i, per tant, l'únic element invertible és el neutre, 1.

Aprofitant la notació exponencial per a productes successius d'un element amb si mateix, $u^n = u \cdot \overset{n}{\cdot} u$, podem abreujar les paraules de A^* tot escrivint la potència corresponent cada vegada que una lletra apareix diverses vegades consecutivament; per exemple, $aabaaabbab = a^2 b a^3 b^2 a b$. Si fem això agrupant al màxim, cadascuna de les potències de lletres que apareixen s'anomena una *síl·laba*; per exemple, la paraula anterior té longitud $2 + 1 + 3 + 2 + 1 + 1 = 10$, i 6 síl·labes.

Observem que, sobre un alfabet d'una sola lletra, $A = \{a\}$, el monoide A^* és isomorf al dels números naturals (en efecte, $A^* = \{1, a, a^2, a^3, \dots\}$ i $a^n \cdot a^m = a^{n+m}$, per $n, m \geq 0$). En canvi, sobre un alfabet A de dues o més lletres, el monoide A^* és força més complicat; en particular, com que $ab \neq ba$, no és un monoide commutatiu.

Anem en la bona direcció per aconseguir que A sigui una base de A^* : efectivament A genera A^* , i cap producte de longitud $n \neq 0$ d'elements de A no dona mai igual a 1. El problema és que A^* només és un monoide (de fet, un monoide lliure); però és molt lluny de ser un grup ja que cap element, llevat del neutre 1, no té invers. Per a convertir-lo en un grup, haurem d'introduir els

inversos de tots els elements. Com veurem, serà suficient d'introduir «inversos formals» per a les lletres elementals, $A^{-1} = \{a^{-1} : a \in A\}$. És a dir, doblem l'alfabet A amb una lletra nova per cadascuna de les antigues, i el designem per $A^{\pm} = A \sqcup A^{-1} = \{a, a^{-1} \mid a \in A\}$. La notació usada suggereix que cada a^{-1} és l'invers del corresponent a , i així serà en acabar la construcció, però de moment és només una qüestió notacional: a i a^{-1} són simplement dues lletres diferents del nou alfabet $A^{\pm}$ (inverses formals, si voleu).

Considerem el monoide $(A^{\pm})^*$, és a dir, el conjunt de totes les paraules sobre l'alfabet $A^{\pm}$ amb l'operació de concatenació. Per exemple, si $A = \{a, b\}$ tindrem $A^{\pm} = \{a, a^{-1}, b, b^{-1}\}$ i $(A^{\pm})^*$ conté exactament $4^2 = 16$ paraules de longitud 2, a saber $a^2, aa^{-1}, ab, ab^{-1}, a^{-1}a, a^{-2}, a^{-1}b, a^{-1}b^{-1}, ba, ba^{-1}, b^2, bb^{-1}, b^{-1}a, b^{-1}a^{-1}, b^{-1}b$, i b^{-2} (on $x^{-n} = (x^{-1})^n$, per $n \geq 0$). Per aconseguir que, per cada lletra $a \in A$, l'element a^{-1} sigui realment l'invers de a , cal fer alguna cosa per a forçar que $aa^{-1} = 1 = a^{-1}a$. Una manera d'aconseguir-ho és fent el conjunt quocient per la relació d'equivalència apropiada.

DEFINICIÓ 9. Anomenem *reducció elemental*, designada per $\rightsquigarrow$, la transformació consistent a eliminar una cancel·lació dins d'una paraula; és a dir, si $a \in A$ i $u, v \in (A^{\pm})^*$, aleshores

$$uaa^{-1}v \rightsquigarrow uv \quad \text{i} \quad ua^{-1}av \rightsquigarrow uv. \quad (1)$$

La transformació inversa l'anomenem *inserció elemental*, i la seva clausura simètrica (designada per $\rightsquigarrow$), *transformació elemental*. És a dir,

$$w \rightsquigarrow w' \Leftrightarrow w \rightsquigarrow w' \text{ ó } w' \rightsquigarrow w. \quad (2)$$

Òbviament, si $w \rightsquigarrow w'$ llavors $|w| = |w'| \pm 2$. Finalment, definim $\sim$ com la *clausura reflexotransitiva* de $\rightsquigarrow$; és a dir, per a tot $w, w' \in (A^{\pm})^*$

(i) $w \sim w$ i

(ii) $w \sim w' \Leftrightarrow \exists w_0 = w, w_1, \dots, w_n = w' \text{ t.q. } w_0 \rightsquigarrow w_1 \rightsquigarrow \dots \rightsquigarrow w_n$,

és a dir, $w \sim w'$ si es pot passar de w a w' fent una quantitat finita de reduccions i/o insercions elementals, entenent que això inclou el cas amb $n = 0$ passes. Per exemple, $a^2a^{-1}bb^{-1}a^{-2}b \sim a^2c^{-1}ca^{-3}b$ ja que

$$a^2a^{-1}bb^{-1}a^{-2}b \rightsquigarrow abb^{-1}a^{-2}b \rightsquigarrow aa^{-2}b \rightsquigarrow a^2a^{-3}b \rightsquigarrow a^2c^{-1}ca^{-3}b.$$

És immediat de la definició que $\sim$ és una relació d'equivalència al conjunt $(A^{\pm})^*$. Podem considerar, doncs, el conjunt quocient, que designem per $\mathbb{F}_A = (A^{\pm})^* / \sim$. És clar que a la classe $[w] \in \mathbb{F}_A$ hi ha precisament totes les paraules «iguals» a $w \in (A^{\pm})^*$ mòdul les igualtats elementals desitjades $aa^{-1} = 1 = a^{-1}a$, $a \in A$. Finalment, definim a $\mathbb{F}_A = (A^{\pm})^* / \sim$ una operació binària (clarament ben definida) adaptant de manera natural la concatenació de $(A^{\pm})^*$: per a tot $u, v \in (A^{\pm})^*$,

$$[u] \cdot [v] = [uv]. \quad (3)$$

PROPOSICIÓ 10. *El conjunt $\mathbb{F}_A$ amb l'operació (3) és un grup.*

PROVA. La propietat associativa és conseqüència immediata de l'associativitat de la concatenació en $(A^\pm)^*$; l'element neutre és $[1]$; i l'invers d'una classe $[a_{i_1}^{\epsilon_1} \cdots a_{i_n}^{\epsilon_n}] \in \mathbb{F}_A$ (on $\epsilon_j = \pm 1$) és la classe $[a_{i_n}^{-\epsilon_n} \cdots a_{i_1}^{-\epsilon_1}]$. Per tant, $\mathbb{F}_A$ amb l'operació (3) té estructura de grup. $\square$

Ja tenim el grup $\mathbb{F}_A$ construït. Vegem ara que $\mathbb{F}_A$ és lliure amb base A . La manera natural de veure A dins de $\mathbb{F}_A$ és com el conjunt de classes de les paraules positives de longitud 1, és a dir, $\{[a] : a \in A\}$. Però hi ha un petit problema tècnic aquí: estem segurs que dues lletres positives diferents estan sempre en classes diferents? En altres paraules, estem segurs que l'aplicació $\iota_A: A \rightarrow \mathbb{F}_A$, $a \mapsto [a]$, és injectiva? Intuïtivament sembla clar que, si $a, b \in A$ són dues lletres diferents, llavors $a \not\sim b$ ja que no sembla possible transformar la lletra a en una de diferent b , simplement usant reduccions i insercions elementals. Per a demostrar-ho rigorosament usarem la proposició següent, que és important també per a altres qüestions.

DEFINICIÓ 11. Una paraula $w \in (A^\pm)^*$ és *reduïda* si no conté cap parell de lletres consecutives mútuament inverses formals. És a dir, $w = a_{i_1}^{\epsilon_1} a_{i_2}^{\epsilon_2} \cdots a_{i_n}^{\epsilon_n}$ és reduïda si, sempre que dues lletres consecutives coincideixin, $a_{i_j} = a_{i_{j+1}}$, els seus signes també, $\epsilon_j = \epsilon_{j+1}$. El conjunt de paraules reduïdes en A el designarem per $R(A) \subseteq (A^\pm)^*$.

PROPOSICIÓ 12. *Tota classe d'equivalència $[w] \in \mathbb{F}_A$ conté una i només una paraula reduïda (designada per $\overline{w}$).*

PROVA. És clar que tota classe conté alguna paraula reduïda, ja que podem prendre un representant qualsevol $w \in [w]$ i (si no és ja reduït) reduir-lo aplicant reduccions elementals successives fins que no en quedi cap de disponible. Aquest procés sempre acaba en un nombre finit de passos, ja que $|w|$ és finit i a cada pas la longitud disminueix en dues unitats.

Per veure la unicitat, suposem que dues paraules reduïdes diferents $w, w' \in R(A)$ pertanyen a la mateixa classe i busquem una contradicció. Sigui $w = w_0 \rightsquigarrow w_1 \rightsquigarrow \cdots \rightsquigarrow w_{n-1} \rightsquigarrow w_n = w'$ una successió de transformacions elementals minimitzant $N = \sum_{i=0}^n |w_i|$. Com que $w \neq w'$ i ambdues són reduïdes, no pot ser ni $n = 0$, ni $n = 1$, ni $n = 2$; per tant, $n \geq 3$. A més, $|w_0| < |w_1|$ i $|w_{n-1}| > |w_n|$ i, per tant, hi ha un $j \in [1, n-1]$ tal que $|w_{j-1}| < |w_j| > |w_{j+1}|$. Fixem-nos ara en les reduccions elementals $w_{j-1} \rightsquigarrow w_j \rightsquigarrow w_{j+1}$, i en les dues lletres de w_j afectades per cadascuna d'aquestes reduccions. Si n'hi ha una (o dues) en comú, aleshores $w_{j-1} = w_{j+1}$ cosa que contradiu la minimalitat de N . I si els parells de lletres eliminades en ambdues reduccions són disjunts, aleshores $w_j = x a^\epsilon a^{-\epsilon} y b^\delta b^{-\delta} z$, per a certes paraules $x, y, z \in (A^\pm)^*$, certes lletres $a, b \in A$, i certs signes $\epsilon, \delta = \pm 1$; canviant $w_{j-1} \rightsquigarrow w_j \rightsquigarrow w_{j+1}$ per $w_{j-1} = x y b^\delta b^{-\delta} z \rightsquigarrow x y z \rightsquigarrow x a^\epsilon a^{-\epsilon} y z = w_{j+1}$ ó $w_{j-1} = x a^\epsilon a^{-\epsilon} y z \rightsquigarrow x y z \rightsquigarrow x y b^\delta b^{-\delta} z = w_{j+1}$ segons convingui, reduïm el valor de N , en contradicció també amb la seva minimalitat. $\square$

OBSERVACIÓ 13. Donat que a tota classe $[w] \in \mathbb{F}_A$ hi ha una única paraula reduïda $\bar{w} \in R(A)$ (tal que $[w] = [\bar{w}]$), podem pensar el grup $\mathbb{F}_A$, alternativament, com el conjunt de paraules reduïdes $R(A)$ amb l'operació $u \cdot v = \overline{uv}$. Sovint aplicarem aquesta interpretació i escriurem simplement $w \in \mathbb{F}_A$ per referir-nos a l'element $[w] = [\bar{w}] \in \mathbb{F}_A$.

Com que les lletres (les paraules de llargada 1, per ser precisos) són clarament reduïdes, el resultat següent és immediat de la proposició 12.

COROLLARI 14. *L'aplicació $\iota_A: A \rightarrow \mathbb{F}_A, a \mapsto [a]$ és injectiva.* $\square$

Ara, pensant $A \subseteq \mathbb{F}_A$ via ι_A , és clar que tota paraula reduïda és producte d'elements de $A^\pm$ (és a dir, A genera $\mathbb{F}_A$), i que el buit és l'únic producte reduït d'elements de $A^\pm$ que dona $1 \in \mathbb{F}_A$ (és a dir, A és un subconjunt lliure de $\mathbb{F}_A$).

COROLLARI 15. *Per a tot alfabet A , el grup $\mathbb{F}_A$ és lliure amb base A .* $\square$

Hem provat l'existència de grups lliures amb bases de qualsevol cardinal. La pregunta natural ara és quan dos d'aquests grups, $\mathbb{F}_A$ i $\mathbb{F}_{A'}$, són isomorfs (com a grups). I és força raonable pensar que serà, precisament, quan A i A' tinguin el mateix cardinal. Efectivament, així és; i per a demostrar-ho ens serà de molta utilitat la caracterització següent de «grup lliure», que, en realitat, n'és la definició estàndard en termes categòrics.

PROPOSICIÓ 16. *Siguin F un grup i $A \subseteq F$. Designem per $\iota_A: A \hookrightarrow F$ la inclusió natural. Aleshores, F és lliure amb base A si i només si, per a tot grup G , i tota aplicació (de conjunts) $\varphi: A \rightarrow G$, existeix un únic morfisme de grups $\tilde{\varphi}: F \rightarrow G$ tal que $\iota_A \tilde{\varphi} = \varphi$.*

$$\begin{array}{ccc} A & \xrightarrow{\varphi} & G \\ \downarrow \iota_A & \nearrow \tilde{\varphi} & \\ F & & \end{array}$$

FIGURA 1: Definició categòrica de grup lliure F .

PROVA. Suposem que F és lliure amb base A . Donats G i $\varphi: A \rightarrow G$, el morfisme $\tilde{\varphi}: F \rightarrow G$ que busquem ha de complir $a\tilde{\varphi} = a\varphi$ per a tot element $a \in A$ i, per tant, $(a_{i_1}^{\epsilon_1} \cdots a_{i_n}^{\epsilon_n})\tilde{\varphi} = (a_{i_1}\tilde{\varphi})^{\epsilon_1} \cdots (a_{i_n}\tilde{\varphi})^{\epsilon_n} = (a_{i_1}\varphi)^{\epsilon_1} \cdots (a_{i_n}\varphi)^{\epsilon_n}$, per a tot producte reduït d'elements de A , $a_{i_1}^{\epsilon_1} \cdots a_{i_n}^{\epsilon_n}$. Com que A genera F , el possible morfisme $\tilde{\varphi}$ queda completament determinat per φ , la qual cosa ens dona la unicitat. D'altra banda, com que A és un subconjunt lliure de F (i. e., cap element de F no admet dues expressions diferents com a producte reduït d'elements de A) és clar que la igualtat anterior per a $\tilde{\varphi}$ ens dona una aplicació $F \rightarrow G$ ben definida. Finalment, veiem que $\tilde{\varphi}$ és morfisme de grups: donats $x, y \in F$, considerem les seves (úniques) expressions reduïdes,

$$x = a_{i_n}^{\epsilon_n} \cdots a_{i_2}^{\epsilon_2} a_{i_1}^{\epsilon_1} \quad \text{i} \quad y = b_{j_1}^{\delta_1} b_{j_2}^{\delta_2} \cdots b_{j_m}^{\delta_m},$$

amb $a_{i_1}, \dots, a_{i_n}, b_{j_1}, \dots, b_{j_m} \in A$ i $\epsilon_1, \dots, \epsilon_n, \delta_1, \dots, \delta_m = \pm 1$. Fixant-nos en les $0 \leq r \leq \min\{n, m\}$ cancel·lacions que apareixen en el producte xy , tenim que $a_{i_1}^{\epsilon_1} = b_{j_1}^{-\delta_1}, \dots, a_{i_r}^{\epsilon_r} = b_{j_r}^{-\delta_r}$ i $a_{i_{r+1}}^{\epsilon_{r+1}} \neq b_{j_{r+1}}^{-\delta_{r+1}}$. En aquesta situació, l'única expressió reduïda per a l'element xy en termes de A és $xy = a_{i_n}^{\epsilon_n} \dots a_{i_{r+1}}^{\epsilon_{r+1}} b_{j_{r+1}}^{\delta_{r+1}} \dots b_{j_m}^{\delta_m}$ i tenim

$$\begin{aligned}
 (xy)\tilde{\varphi} &= (a_{i_n}^{\epsilon_n} \dots a_{i_{r+1}}^{\epsilon_{r+1}} b_{j_{r+1}}^{\delta_{r+1}} \dots b_{j_m}^{\delta_m})\tilde{\varphi} = \\
 &= (a_{i_n}\varphi)^{\epsilon_n} \dots (a_{i_{r+1}}\varphi)^{\epsilon_{r+1}} \cdot (b_{j_{r+1}}\varphi)^{\delta_{r+1}} \dots (b_{j_m}\varphi)^{\delta_m} = \\
 &= [(a_{i_n}\varphi)^{\epsilon_n} \dots (a_{i_{r+1}}\varphi)^{\epsilon_{r+1}}] \cdot [(a_{i_r}\varphi)^{\epsilon_r} \dots (a_{i_1}\varphi)^{\epsilon_1}] \cdot \\
 &\quad \cdot [(b_{j_1}\varphi)^{\delta_1} \dots (b_{j_r}\varphi)^{\delta_r}] \cdot [(b_{j_{r+1}}\varphi)^{\delta_{r+1}} \dots (b_{j_m}\varphi)^{\delta_m}] = \\
 &= (a_{i_n}^{\epsilon_n} \dots a_{i_{r+1}}^{\epsilon_{r+1}} \cdot a_{i_r}^{\epsilon_r} \dots a_{i_1}^{\epsilon_1})\tilde{\varphi} \cdot (b_{j_1}^{\delta_1} \dots b_{j_r}^{\delta_r} \cdot b_{j_{r+1}}^{\delta_{r+1}} \dots b_{j_m}^{\delta_m})\tilde{\varphi} = \\
 &= (x)\tilde{\varphi} \cdot (y)\tilde{\varphi}.
 \end{aligned}$$

Per veure l'altra implicació, suposem certa la propietat universal per a F .

Considerem $H = \langle A \rangle \leq F$, el subgrup de F generat per A , i designem per $i: H \hookrightarrow F$ la inclusió. Aplicant la propietat universal a la inclusió $\varphi: A \hookrightarrow H$, obtenim un morfisme $\tilde{\varphi}: F \rightarrow H$ que compleix $\iota_A \tilde{\varphi} = \varphi$. Però ara tornem a aplicar la propietat universal a la composició $\iota_A = \varphi i: A \hookrightarrow H \hookrightarrow F$: el morfisme $\tilde{\varphi} i$ compleix $\iota_A(\tilde{\varphi} i) = (\iota_A \tilde{\varphi})i = \varphi i$, i la identitat $\text{id}_F: F \rightarrow F$ també ho compleix $\iota_A \text{id}_F = \iota_A = \varphi i$. Per tant, per la unicitat, $\text{id}_F = \tilde{\varphi} i$, d'on deduïm que $F = \text{im}(\text{id}_F) = \text{im}(\tilde{\varphi} i) = \text{im } \tilde{\varphi} \leq H = \langle A \rangle$ i, per tant, que A genera F .

$$\begin{array}{ccc}
 A & \xrightarrow{\varphi} & H \\
 \downarrow \iota_A & \nearrow \tilde{\varphi} & \\
 F & &
 \end{array}
 \qquad
 \begin{array}{ccccc}
 A & \xrightarrow{\varphi} & H & \xrightarrow{i} & F \\
 \downarrow \iota_A & \nearrow \tilde{\varphi} = \text{id}_F & & & \\
 F & & & &
 \end{array}$$

Per altra banda, podem pensar en el conjunt A com un alfabet formal i construir el grup $G = \mathbb{F}_A$; aplicant la propietat universal a la inclusió $\varphi: A \hookrightarrow \mathbb{F}_A$, $a \mapsto [a]$, existeix un morfisme $\tilde{\varphi}: F \rightarrow \mathbb{F}_A$ que compleix $\iota_A \tilde{\varphi} = \varphi$, és a dir, que compleix $(a_{i_1}^{\epsilon_1} \dots a_{i_n}^{\epsilon_n})\tilde{\varphi} = (a_{i_1}\varphi)^{\epsilon_1} \dots (a_{i_n}\varphi)^{\epsilon_n} = [a_{i_1}]^{\epsilon_1} \dots [a_{i_n}]^{\epsilon_n} = [a_{i_1}^{\epsilon_1} \dots a_{i_n}^{\epsilon_n}]$, per a cada producte reduït $a_{i_1}^{\epsilon_1} \dots a_{i_n}^{\epsilon_n}$ d'elements de A . Per tant, si dos productes reduïts d'elements de A , $a_{i_1}^{\epsilon_1} \dots a_{i_n}^{\epsilon_n}$ i $b_{j_1}^{\delta_1} \dots b_{j_m}^{\delta_m}$ donen el mateix resultat a F , $a_{i_1}^{\epsilon_1} \dots a_{i_n}^{\epsilon_n} =_F b_{j_1}^{\delta_1} \dots b_{j_m}^{\delta_m}$, llavors $[a_{i_1}^{\epsilon_1} \dots a_{i_n}^{\epsilon_n}] =_{\mathbb{F}_A} [b_{j_1}^{\delta_1} \dots b_{j_m}^{\delta_m}]$, i la proposició 12 ens assegura que són la mateixa expressió reduïda: $n = m$, $a_{i_1} = b_{j_1}, \dots, a_{i_n} = b_{j_n}$, i $\epsilon_1 = \delta_1, \dots, \epsilon_n = \delta_n$. Per tant, A és un subconjunt lliure de F . $\square$

PROPOSICIÓ 17. *Siguin $\mathbb{F}$ un grup lliure amb base $A \subseteq \mathbb{F}$, $\mathbb{F}'$ un grup lliure amb base $A' \subseteq \mathbb{F}'$, i designem per ι_A i $\iota_{A'}$ les respectives inclusions. Llavors, $\mathbb{F}$ i $\mathbb{F}'$ són grups isomorfs si i només si A i A' tenen el mateix cardinal: $\mathbb{F} \simeq \mathbb{F}' \Leftrightarrow \#A = \#A'$.*

PROVA. Per la implicació cap a l'esquerra, suposem $\#A = \#A'$ i prenem una aplicació bijectiva $\eta: A \rightarrow A'$. Aplicant la propietat universal de $\mathbb{F}$ a $\eta\iota_{A'}$ obtenim un morfisme $\alpha: \mathbb{F} \rightarrow \mathbb{F}'$ que compleix $\iota_A \alpha = \eta\iota_{A'}$; anàlogament, aplicant la propietat universal de $\mathbb{F}'$ a $\eta^{-1}\iota_A$, obtenim un morfisme $\beta: \mathbb{F}' \rightarrow \mathbb{F}$ que compleix $\iota_{A'} \beta = \eta^{-1}\iota_A$; vegeu la part superior de (4). Però ara apliquem la propietat universal de $\mathbb{F}$ a la inclusió ι_A : òbviament la identitat compleix $\iota_A \text{id}_{\mathbb{F}} = \iota_A$; però el morfisme $\alpha\beta: \mathbb{F} \rightarrow \mathbb{F}' \rightarrow \mathbb{F}$ també: $\iota_A \alpha\beta = \eta\iota_{A'}\beta = \eta\eta^{-1}\iota_A = \iota_A$. Per unicitat, deduïm que $\alpha\beta = \text{id}_{\mathbb{F}}$. I amb un argument simètric, $\beta\alpha = \text{id}_{\mathbb{F}'}$. Per tant, $\mathbb{F} \simeq \mathbb{F}'$.

$$\begin{array}{ccc}
 A & \xrightarrow{\eta} & A' \xrightarrow{\iota_{A'}} \mathbb{F}' \\
 \downarrow \iota_A & \nearrow \alpha & \\
 \mathbb{F} & &
 \end{array}
 \qquad
 \begin{array}{ccc}
 A' & \xrightarrow{\eta^{-1}} & A \xrightarrow{\iota_A} \mathbb{F} \\
 \downarrow \iota_{A'} & \nearrow \beta & \\
 \mathbb{F}' & &
 \end{array}
 \tag{4}$$

$$\begin{array}{ccc}
 A & \xrightarrow{\iota_A} & \mathbb{F} \\
 \downarrow \iota_A & \nearrow \alpha\beta=\text{id}_{\mathbb{F}} & \\
 \mathbb{F} & &
 \end{array}
 \qquad
 \begin{array}{ccc}
 A' & \xrightarrow{\iota_{A'}} & \mathbb{F}' \\
 \downarrow \iota_{A'} & \nearrow \beta\alpha=\text{id}_{\mathbb{F}'} & \\
 \mathbb{F}' & &
 \end{array}$$

Per a la implicació contrària farem un argument de naturalesa diferent. És fàcil veure que, si $\#A \geq \aleph_0$, llavors $\#\mathbb{F}_A = \#A$. Per tant, si $\#A, \#A' \geq \aleph_0$ el resultat és evident. Restringim-nos, doncs, al cas $\#A < \infty$.

Observeu que, per cada grup G , la propietat universal de $\mathbb{F}$ ens dona una bijecció, $\text{Map}(A, G) \rightarrow \text{Hom}(\mathbb{F}, G)$, $\varphi \mapsto \tilde{\varphi}$, entre el conjunt d'aplicacions de A a G , i el conjunt de morfismes de grups de $\mathbb{F}$ a G (la inversa és la restricció, $\phi|_A \mapsto \phi$). Per tant, aquests dos conjunts tenen el mateix cardinal, $\#\text{Map}(A, G) = \#\text{Hom}(\mathbb{F}, G)$. Si $\mathbb{F}$ i $\mathbb{F}'$ són grups isomorfs, llavors, mirant aplicacions i morfismes a $\mathbb{Z}/2\mathbb{Z}$, tenim

$$\begin{aligned}
 2^{\#A} &= \#\text{Map}(A, \mathbb{Z}/2\mathbb{Z}) = \#\text{Hom}(\mathbb{F}, \mathbb{Z}/2\mathbb{Z}) = \\
 &= \#\text{Hom}(\mathbb{F}', \mathbb{Z}/2\mathbb{Z}) = \#\text{Map}(A', \mathbb{Z}/2\mathbb{Z}) = 2^{\#A'},
 \end{aligned}$$

d'on es dedueix que $\#A = \#A'$. □

OBSERVACIÓ 18. Un grup lliure $\mathbb{F}$, doncs, pot ser-ho sobre diversos subconjunts seus $A, A' \subseteq \mathbb{F}$; però, en aquest cas, tots hauran de tenir el mateix cardinal $\#A = \#A'$. Donat que les bases de $\mathbb{F}$ són sistemes de generadors minimal, veiem ara que, de fet, són sistemes de generadors de cardinal mínim: si A és base de $\mathbb{F}$, aleshores $\text{rk}(\mathbb{F}) = \#A$. Escrivem $\mathbb{F}_\kappa$ per designar el grup lliure de rang κ (si no volem fer referència a cap base concreta).

Aquesta situació és anàloga a la de l'àlgebra lineal: un K -espai vectorial E sobre un cos K pot tenir diverses bases, però totes han de tenir sempre el mateix cardinal (la K -dimensió de E). Com ja hem vist, hi ha grups lliures de

qualsevol rang, de la mateixa manera que hi ha K -espais vectorials de qualsevol dimensió. Una diferència important, però, és que en àlgebra lineal tot K -espai vectorial té bases (i, per tant, dimensió), mentre que hi ha molts grups que no tenen cap base, és a dir que *no són lliures* sobre cap subconjunt (els grups finits no trivials, per exemple).

Algebraicament, la família dels grups lliures és d'una gran rellevància. Com que, llevat d'isomorfia, n'hi ha un de sol per a cada rang κ , el designarem per $\mathbb{F}_\kappa$ si volem obviar la referència a l'alfabet concret A usat per a la seva construcció. Així tindrem:

$$\mathbb{F}_0 \simeq 1, \quad \mathbb{F}_1 \simeq \mathbb{Z}, \quad \mathbb{F}_2, \quad \mathbb{F}_3, \dots, \mathbb{F}_{\aleph_0}, \dots^7$$

Clarament, els únics grups commutatius són els dos primers, força especials dins d'aquesta família. La rellevància d'aquests grups rau en el fet que, d'alguna manera, contenen *tota* la informació possible sobre *tots* els grups possibles. Això es concreta a través del resultat clàssic següent, força senzill de demostrar, i alhora de fonamental importància per a la teoria de grups.

TEOREMA 19. *Tot grup G és quocient d'un grup lliure. És a dir, per a tot grup G existeixen un cardinal κ i un subgrup normal $N \trianglelefteq \mathbb{F}_\kappa$ tals que $G \simeq \mathbb{F}_\kappa / N$.*

PROVA. Sigui $A \subseteq G$ un conjunt de generadors de G (sempre es pot prendre $A = G$), i sigui $\kappa = \#A$ el seu cardinal. Pensem A com un conjunt abstracte, i considerem el grup lliure $\mathbb{F}_A$. Per la propietat universal aplicada a la inclusió $\varphi: A \rightarrow G$, existeix un (únic) morfisme de grups $\tilde{\varphi}: \mathbb{F}_A \rightarrow G$ tal que $[a]\tilde{\varphi} = a$, per a tot $a \in A$. Com que A genera G , $\tilde{\varphi}$ és exhaustiu i, pel primer teorema d'isomorfia, $N = \ker \tilde{\varphi}$ és un subgrup normal de $\mathbb{F}_A$ que compleix $\mathbb{F}_A / N \simeq \text{Im}(\tilde{\varphi}) = G$. $\square$

Si al teorema 19 fixem una base per a $\mathbb{F}_\kappa$, i una família de generadors de N com a subgrup normal de $\mathbb{F}_\kappa$, aleshores obtenim el concepte de presentació, essencial en teoria de grups.

DEFINICIÓ 20. Sigui G un grup. Una *presentació* per a G és un parell (A, R) on A és un conjunt de símbols, R és un subconjunt de $\mathbb{F}_A$, i $G \simeq \mathbb{F}_A / \langle\langle R \rangle\rangle$.⁸ Abusant lleugerament del llenguatge, se sol escriure $G = \langle A \mid R \rangle$, i es diu que els elements de A (resp., R) són els *generadors*⁹ (resp., *relators*) de G donats per la presentació $\langle A \mid R \rangle$, i que $w \in (A^\pm)^*$ és una *paraula en els generadors* de G .

COROLLARI 21. *El grup lliure amb base A admet la presentació $\mathbb{F}_A = \langle A \mid - \rangle$.*

⁷ Qui vulgui que siguin... Vegeu https://ca.wikipedia.org/wiki/Hipòtesi_del_continu.

⁸ La clausura normal $\langle\langle R \rangle\rangle$ és el subgrup (normal) generat per tots els conjugats d'elements de R .

⁹ Formalment, els generadors de G donats per la presentació $G = \langle A \mid X \rangle$ són les imatges $[a]\tilde{\varphi}$ dels elements $a \in A$ per la seqüència d'homomorfismes exhaustius $(A^\pm)^* \twoheadrightarrow \mathbb{F}_A \twoheadrightarrow G$.

Noteu que el teorema 19 ens diu que tot grup admet una presentació —de fet, infinites presentacions. Diem que una presentació $\langle A \mid R \rangle$ és *finita* si tant A com R són conjunts finits. Es diu que un grup és *finitament presentat* si admet una presentació finita.

El teorema 19 admet dues interpretacions contraposades. Per un costat, implica que, col·lectivament, els reticles de subgrups (normals) dels grups lliures contenen tota la informació algebraica possible sobre qualsevol grup concebible; això els converteix en uns candidats naturals i molt concrets on buscar informació sobre qualsevol grup. Però, per altra banda, per aquesta mateixa raó, aquests reticles han de ser extraordinàriament complicats, i és utòpic pretendre entendre'ls completament.

Les presentacions constitueixen la manera natural de descriure els grups des del punt de vista de la teoria combinatòria i algorísmica de grups. La contrapartida geomètrica més elemental és l'anomenat *digraf de Cayley*, definit a continuació.

DEFINICIÓ 22. Sigui G un grup i $S \subseteq G$ un conjunt de generadors per G . Aleshores, el *digraf de Cayley de G respecte de S* , designat per $\text{Cay}(G, S)$, és el graf dirigit (digraf) amb conjunt de vèrtexs G , i un arc $g \xrightarrow{s} gs$ per a cada element $g \in G$ i cada generador $s \in S^\pm$.

Observeu que els digrafs de Cayley (respecte a conjunts generadors) són connexos: com que $\langle S \rangle = G$, qualsevol $g \in G$ es pot escriure en la forma $g = s_{i_1}^{\epsilon_1} \cdots s_{i_k}^{\epsilon_k}$, per a certs $k \geq 0$, $s_{i_j} \in S$, i $\epsilon_j = \pm 1$ i, llavors, el camí

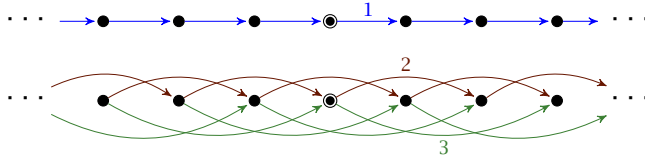
$$1 \xrightarrow{s_{i_1}^{\epsilon_1}} s_{i_1}^{\epsilon_1} \xrightarrow{s_{i_2}^{\epsilon_2}} s_{i_1}^{\epsilon_1} s_{i_2}^{\epsilon_2} \xrightarrow{s_{i_3}^{\epsilon_3}} \cdots \xrightarrow{s_{i_k}^{\epsilon_k}} g \quad (5)$$

connecta el vèrtex $1 \in G$ amb el vèrtex genèric $g \in G$. Per altra banda, els digrafs de Cayley són també $(2\#S)$ -regulars:¹⁰ per definició, de cada vèrtex $g \in G$ en surt exactament una aresta per cada $s \in S^\pm$ (que va al vèrtex $gs \in G$) i n'arriba també exactament una per cada $s \in S^\pm$ (que ve de $gs^{-1} \in G$); en general, algunes d'aquestes arestes poden ser paral·leles, o llaços.

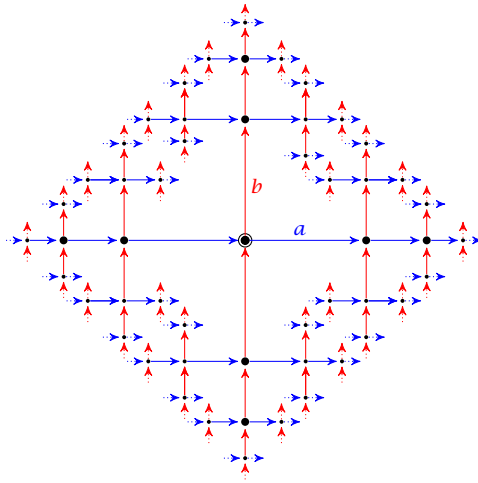
Recalquem que $\text{Cay}(G, S)$ depèn fortament del conjunt de generadors escollit S (vegeu figura 2). Per tant, no és un objecte geomètric canònicament associat al grup G , sinó al grup G i a un conjunt de generadors concret S .

EXEMPLE 23. El digraf de Cayley del grup lliure $\mathbb{F}_{\{a\}} \simeq \mathbb{Z}$ respecte a la base $\{a\}$ (és a dir, $\{1\}$ en notació additiva) és un camí dirigit amb infinits vèrtexs consecutius, connectats cadascun al següent per una a -aresta (i a l'anterior per una (a^{-1}) -aresta). A la figura 2 podeu veure dibuixats els digrafs de Cayley $\text{Cay}(\mathbb{Z}, \{1\})$ i $\text{Cay}(\mathbb{Z}, \{2, 3\})$ on, com és habitual amb els digrafs involutius (vegeu la definició 32), dibuixem només els arcs positius, i assenyallem amb $\bullet$ el vèrtex corresponent a l'element neutre.

¹⁰ Un digraf Γ és *regular* si el nombre d'arcs que entren en un vèrtex i el nombre d'arcs que en surten són independents del vèrtex $p \in V\Gamma$.

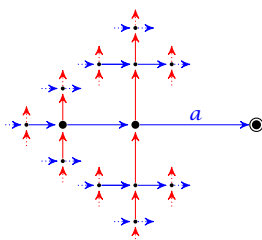
FIGURA 2: $\text{Cay}(\mathbb{Z}, \{1\})$ (a dalt) i $\text{Cay}(\mathbb{Z}, \{2, 3\})$ (a sota).

EXEMPLE 24. És fàcil veure que el digraf de Cayley del grup lliure $\mathbb{F}_{\{a,b\}}$ respecte de la base $\{a, b\}$ té l'aspecte de la figura 3. Per a convèncer-nos-en, només cal tenir en compte les dues observacions anteriors (es tracta d'un digraf connex, i 4-regular), juntament amb els fets d'ésser infinit ($\#\mathbb{F}_{\{a,b\}} = \aleph_0$) i d'ésser un arbre: un hipotètic camí tancat ens donaria dues expressions diferents d'un mateix element g com a producte reduït de les lletres a, a^{-1}, b, b^{-1} , contradient el fet que $\mathbb{F}_{\{a,b\}}$ és lliure en $\{a, b\}$. Anàlogament, el digraf de Cayley del grup lliure $\mathbb{F}_\kappa$ de rang κ respecte a una base és el corresponent arbre infinit 2κ -regular.

FIGURA 3: El digraf de Cayley de $\mathbb{F}_{\{a,b\}}$.

Com veurem a la secció següent, el digraf de Cayley dels grups lliures respecte a una base tindrà protagonisme en la teoria d'autòmats de Stallings. Concretament, el subgraf definit a continuació.

DEFINICIÓ 25. Sigui A un alfabet, i $a \in A^\pm$. Anomenem *a-branca de Cayley* de $\mathbb{F}_A$ la component connexa de $\text{Cay}(\mathbb{F}_A, A) \setminus F$ que conté $\bullet$, on F és el conjunt d'arcs incidents a $\bullet$ excepte $\bullet \xrightarrow{a} \bullet$ (i el seu invers); vegeu la figura 4.

FIGURA 4: La (a^{-1}) -branca de Cayley de $\mathbb{F}_{\{a,b\}}$.

Dedicarem la resta d'aquest article a fer una primera aproximació a l'estudi del reticle de subgrups de $\mathbb{F}_n$, amb $n \geq 2$ finit. I, tal com intuïem més amunt, ens adonarem de seguida que, efectivament, és extremadament complicat. D'entrada, presenta certs comportaments que recorden patrons de l'àlgebra lineal i suggereixen que, potser, alguns aspectes no seran tan intricats; però gairebé sempre acaba apareixent algun comportament distintiu, sovint de tipus fractal, que el distancia radicalment d'ambients algebraics més benignes.

Acabem aquesta secció proposant un parell de problemes sobre el reticle de subgrups de $\mathbb{F}_2 = \mathbb{F}_{\{a,b\}}$ indicatius de la complexitat a la qual ens referim. Es tracta de problemes molt naturals, anàlegs a problemes típics de l'àlgebra lineal, que en aquell context es resolten fàcilment plantejant i resolent sistemes d'equacions. Només cal endinsar-se una mica en ambdós problemes, com farem a continuació, per a copsar la dificultat extra que suposa el fet d'estar tractant amb grups fortament no commutatius. Convidem el lector a intentar resoldre'ls abans de continuar llegint, per tal de percebre la complexitat a què ens referim. Tot i que en un primer apropament poden semblar difícils de resoldre, veurem a la secció 4 que hi ha una manera elegant, eficient i sorprenentment senzilla de tractar-los: usar la teoria dels autòmats de Stallings, que expliquem en detall a la secció 3.

EXEMPLE 26. Sigui $\mathbb{F}_2$ el grup lliure sobre $A = \{a, b\}$ i considerem el subgrup $H = \langle v_1, v_2, v_3 \rangle \leq \mathbb{F}_2$ generat pels elements $v_1 = baba^{-1}$, $v_2 = aba^{-1}$, i $v_3 = aba^2$. És cert que l'element $u = b^2aba^{-1}b^7a^{-2}b^{-1}a^2 \in \mathbb{F}_2$ pertany a H ? I, en cas afirmatiu, sabem expressar-lo com a producte de v_1, v_2, v_3 i els seus inversos?

Per a certes paraules, decidir la pertinença a H és fàcil. Per exemple, per a veure que a^3 pertany a H , és suficient observar que $v_2^{-1}v_3 = (ab^{-1}a^{-1})(aba^2) = a^3$. Un altre argument senzill ens permet deduir que l'element a no pertany a H : tant se val com multipliquem v_1, v_2, v_3 o els seus inversos, el a -exponent (és a dir, el nombre total de a obtingut sumant les d'exponent positiu i restant les d'exponent negatiu) al resultat sempre serà múltiple de 3. En efecte, els a -exponents dels generadors són $|v_1|_a = 1 - 1 = 0$, $|v_2|_a = 1 - 1 = 0$, i $|v_3|_a = 1 + 2 = 3$; com que $|w_1w_2|_a = |w_1|_a + |w_2|_a$ (ja que la possible cancel·lació entre w_1 i w_2 elimina la mateixa quantitat de a positives que de a negatives), tots els elements del subgrup $H = \langle v_1, v_2, v_3 \rangle$ han de tenir forçosament a -exponent múltiple de 3. I, com que $|a|_a = 1$ no és múltiple de 3, deduïm que $a \notin H$.

Però en el cas de l'element $u = b^2aba^{-1}b^7a^{-2}b^{-1}a^2$, el a -exponent total és $|u|_a = 0$, per tant, no podem descartar-lo usant l'argument anterior. Òbviament, això tampoc no ens assegura que hi hagi de pertànyer: no tenim, *a priori*, cap indicació de com podem multiplicar entre si les paraules v_1, v_2, v_3 i els seus inversos, per tal d'aconseguir u ; ni de si això és realment possible, o no. Si els elements v_1, v_2, v_3 commutessin, el problema es reduiria a decidir si u és de la forma $v_1^\alpha v_2^\beta v_3^\gamma$ amb $\alpha, \beta, \gamma \in \mathbb{Z}$, és a dir, si u és una *combinació lineal* de v_1, v_2, v_3 amb coeficients enters. Com es pot percebre, però, la no commutativitat dels elements v_1, v_2, v_3 complica molt el plantejament i la resolució d'aquest problema.

La resposta en aquest cas és afirmativa, però trobar un producte adequat (que doni u) ja no és tan fàcil a simple vista:

$$\begin{aligned}
 v_1 v_2^{-1} v_1 (v_1 v_2^{-1})^7 v_3^{-1} v_2^{-1} v_3 &= \\
 &= baba^{-1}(aba^{-1})^{-1}baba^{-1}(baba^{-1}(aba^{-1})^{-1})^7(aba^2)^{-1}(aba^{-1})^{-1}aba^2 = \\
 &= baba^{-1}ab^{-1}a^{-1}baba^{-1}(baba^{-1}ab^{-1}a^{-1})^7a^{-2}b^{-1}a^{-1}ab^{-1}a^{-1}aba^2 = \\
 &= b baba^{-1}b^7a^{-2}b^{-1}a^2 = b^2aba^{-1}b^7a^{-2}b^{-1}a^2 = u.
 \end{aligned}$$

Evidentment, fer els càlculs per comprovar si un determinat producte dels generadors dona u és fàcil. Però trobar, o més encara, decidir l'existència d'un producte de $v_1^{\pm 1}, v_2^{\pm 1}, v_3^{\pm 1}$ que doni la paraula candidata u , no ho sembla tant.¹¹ És fàcil adonar-se que la complicació deriva de com de recargolada sigui l'eventual expressió (*de llargada finita però arbitràriament gran*) del candidat u en termes dels generadors. A la subsecció 4.2 explicarem una manera sistemàtica de resoldre aquest problema.

Preguntes afins amb una evident importància algebraica són: és aquesta l'única manera d'obtenir u a partir de v_1, v_2, v_3 ? Si n'hi ha d'altres, podem descriure-les totes? Podem calcular la paraula més curta (potser no única) que doni u ?...

Observeu que el problema anàleg en àlgebra lineal seria del tipus següent: «pertany el vector $(1, 2, 3)$ al subespai vectorial $\langle (-1, 2, 1), (0, 1, -3) \rangle$ de $\mathbb{R}^3$?». Per a respondre, només hem de mirar si el sistema d'equacions $(1, 2, 3) = \alpha(-1, 2, 1) + \beta(0, 1, -3)$ té o no solucions a $\mathbb{R}$. En l'ambient no commutatiu en què ens trobem no coneixem, de moment, cap eina que pugui fer un paper semblant al que fan els sistemes d'equacions en l'àlgebra lineal. Com veurem a continuació, en el cas dels grups lliures, els autòmats de Stallings faran aquest paper.

Per acabar la secció, us proposem un altre problema que involucra interseccions de subgrups del grup lliure.

¹¹ Almenys a la majoria dels humans (vegeu https://ca.wikipedia.org/wiki/P_versus_NP).

EXEMPLE 27. Sigui $\mathbb{F}_2$ el grup lliure sobre $A = \{a, b\}$, i considerem els subgrups $H = \langle u_1, u_2, u_3 \rangle \leq \mathbb{F}_2$ i $K = \langle v_1, v_2, v_3 \rangle \leq \mathbb{F}_2$ donats per les paraules $u_3 = a^{-1}bab^{-1}a$, i $v_1 = ab$, $v_2 = a^3$, i $v_3 = a^{-1}ba$, $u_1 = b$, $u_2 = a^3$. Com veurem al corol·lari 66, els subgrups del grup lliure no tenen per què ser finitament generats. Ho és la intersecció $H \cap K$? I, en tal cas, podeu calcular-ne un conjunt de generadors?

A simple vista ja es veu que l'element a^3 és present a ambdós subgrups: $a^3 = v_2 \in H$, i $a^3 = u_2 \in K$. També se'n veu fàcilment un altre:

$$H \ni u_1^{-1}u_2u_1 = b^{-1}a^3b = v_1^{-1}v_2v_1 \in K.$$

I amb una mica més d'imaginació (o de treball de càlcul) aquí tenim un tercer element de la intersecció:

$$H \ni u_3^3 = a^{-1}ba^3b^{-1}a = v_3v_2v_3^{-1} \in K.$$

Com que $H \cap K$ és un subgrup de $\mathbb{F}_2$, deduïm que $\langle a^3, b^{-1}a^3b, a^{-1}ba^3b^{-1}a \rangle$ està contingut a $H \cap K$. Seguir trobant nous elements a simple vista no sembla fàcil; però la pregunta contrària encara sembla més complicada: és cert que $H \cap K = \langle a^3, b^{-1}a^3b, a^{-1}ba^3b^{-1}a \rangle$? O cal trobar i afegir-hi més generadors?

El lector interessat pot provar de respondre totes aquestes preguntes per si sol. Nosaltres, de moment, ho deixarem aquí, i tornarem a tractar aquests exemples a la secció 4, on ja disposarem d'eines adequades per a respondre totes aquestes preguntes... i més!

3 Autòmats de Stallings

DEFINICIÓ 28. Sigui A un alfabet. Un A -autòmat Γ (o simplement *autòmat*, si l'alfabet és clar pel context) és un graf dirigit amb els arcs etiquetats en A i un vèrtex distingit; més formalment, és una tupla $\Gamma = (V, E, \iota, \tau, \ell, \bullet)$, on V i E són conjunts, $\iota, \tau: E \rightarrow V$ i $\ell: E \rightarrow A$ són funcions, i $\bullet \in V$. Els conjunts $V = V\Gamma$ i $E = E\Gamma$ s'anomenen conjunt de *vèrtexs* i d'*arcs* (o *arestes dirigides*) de Γ , respectivament, i les funcions ι , τ i ℓ assignen a cada arc $e \in E$ el seu *origen* $\iota(e) \in V$, el seu *final* $\tau(e) \in V$, i la seva *etiqueta* $\ell(e) \in A$, respectivament. Si $\iota e = p$, $\tau e = q$ i $\ell(e) = a$, aleshores diem que e és un a -arc de p a q , i escrivim $e \equiv p \xrightarrow{a} q$. El vèrtex distingit $\bullet$ s'anomena *vèrtex base* o *punt base* del A -autòmat. El *digraf subjacent* d'un A -autòmat Γ és el digraf resultat d'ignorar-ne la funció ℓ i el punt base $\bullet$.

Observeu que, segons la nostra definició, un autòmat no és necessàriament connex, i admet la possibilitat tant d'arcs amb el mateix origen i final (*llaços*), com també de múltiples arcs amb el mateix origen i el mateix final (*arcs paral·lels*).

DEFINICIÓ 29. Sigui Γ un A -autòmat. Un vèrtex p de Γ s'anomena *saturat* si per a cada lletra $a \in A$ hi ha (almenys) un a -arc amb origen p . En cas contrari, diem que p és *insaturat* (o *a-deficient*, si volem referir-nos a l'etiqueta que

falta). El a -dèficit de Γ , designat per $\text{def}_a(\Gamma)$, és el nombre (cardinal) de vèrtexs a -deficients a Γ . Un A -autòmat Γ s'anomena *saturat* si tots els seus vèrtexs ho són, i. e., si $\text{def}_a(\Gamma) = 0$ per a tot $a \in A$.

DEFINICIÓ 30. Un *camí (dirigit)* en un A -autòmat Γ és una seqüència alternada finita $\gamma = p_0 e_1 p_1 \cdots e_l p_l$, on $p_i \in V\Gamma$, $e_i \in E\Gamma$, $\iota e_i = p_{i-1}$, i $\tau e_i = p_i$ per a $i = 1, \dots, l$. Aleshores, p_0 i p_l s'anomenen *origen* i *final* de γ respectivament, i escrivim $p_0 = \iota(\gamma)$ i $p_l = \tau(\gamma)$. Diem que γ és un camí de p_0 a p_l , i ho designem per $\gamma: p_0 \rightsquigarrow p_l$. Si l'origen i el final de γ coincideixen, diem que γ és un *camí tancat*. Un camí tancat de p a p s'anomena *p-camí*. La *longitud* d'un camí γ , designada per $|\gamma|$, és el nombre d'arcs a γ (comptant possibles repeticions). Els camins de longitud 0 s'anomenen *camins trivials* i corresponen als vèrtexs de Γ .

Recordem que $A^\pm = A \sqcup A^{-1}$, on A^{-1} és el conjunt d'inversos formals de A .

DEFINICIÓ 31. Un A -autòmat *involutiu* és un $A^\pm$ -autòmat amb una involució $e \mapsto e^{-1}$ als seus arcs tal que si $e \equiv p \xrightarrow{a} q$ aleshores $e^{-1} \equiv p \xleftarrow{a^{-1}} q$. Aleshores, diem que (l'arc etiquetat) e^{-1} és l'*invers* de e . Observeu que $(e^{-1})^{-1} = e$; és a dir, en un autòmat involutiu, els arcs etiquetats apareixen aparellats amb els seus respectius inversos (vegeu la figura 5).

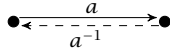


FIGURA 5: Un arc en un autòmat involutiu amb el seu invers puntejat.

DEFINICIÓ 32. Un arc en un A -autòmat involutiu Γ es diu *positiu* (resp., *negatiu*) si la seva etiqueta pertany a A (resp., A^{-1}). S'anomena *part positiva* de Γ , designada per Γ^+ , l'autòmat obtingut després d'eliminar de Γ tots els arcs negatius. És clar que tot autòmat involutiu queda determinat per la seva part positiva, que habitualment farem servir per a descriure l'autòmat, amb el conveni tàcit que els arcs positius $p \xrightarrow{a} q$ es poden travessar cap enrere (i. e., de q a p) llegint la lletra inversa a^{-1} . Observeu que, identificant els parells d'arcs mútuament inversos (en una *aresta*, no dirigida), recuperem la noció estàndard de graf *no dirigit*, que anomenarem *graf (no dirigit) subjacent* de Γ . Farem servir aquesta correspondència per a traslladar a l'àmbit dels autòmats algunes nocions bàsiques de teoria de grafs. Per exemple, direm que un autòmat Γ és *connex* si ho és el seu graf subjacent, direm que T és un *arbre generador*¹² de Γ si ho és del seu graf subjacent, i definirem el *grau* d'un vèrtex p , designat per $\deg(p)$, com el nombre d'arestes incidents a p en el graf subjacent, i el *rang* de Γ , designat per $\text{rk}(\Gamma)$, com el rang del seu graf subjacent (vegeu la definició 33). En particular, si Γ és un autòmat involutiu, connex i finit, aleshores $\text{rk}(\Gamma) = 1 - \#V\Gamma + \#E\Gamma^+$.

¹² En anglès, *spanning tree*.

Si Γ és un A -autòmat involutiu, aleshores el $A^\pm$ -etiquetatge de Γ s'estén de forma natural a un $(A^\pm)^*$ -etiquetatge dels camins, tot concatenant les corresponents etiquetes: si $\gamma = p_0 e_1 p_1 \cdots e_l p_l$ és un camí, definim la seva *etiqueta*, i la seva *etiqueta reduïda* com $\ell(\gamma) = \ell(e_1) \cdots \ell(e_l) \in (A^\pm)^*$, i $\bar{\ell}(\gamma) = \bar{\ell}(\gamma) \in \mathbb{F}_A$, respectivament. En ambdós casos, s'entén que l'etiqueta d'un camí trivial és l'element trivial 1. Si $\ell(\gamma) = w \in (A^\pm)^*$ diem que el camí γ *llegeix* w , o que la paraula w *etiqueta* el camí γ , i escrivim $\gamma: p_0 \xrightarrow{w} p_l$.

Es diu que un camí presenta *retrocés*¹³ si té dos arcs successius inversos l'un de l'altre. Un camí sense retrocés es diu *camí reduït*. És obvi que, si $\ell(\gamma)$ és una paraula reduïda, aleshores el camí γ ha de ser reduït, però el recíproc no és cert en general, ja que es poden donar les situacions (no deterministes) de la figura 6. Observeu també que, si T és un arbre generador d'un A -autòmat connex Γ , aleshores per a cada parell de vèrtexs p, q a Γ existeix un únic camí *reduït* de p a q usant només arcs de T ; l'anomenem T -geodèsica de p a q , i el dissenyem per $\gamma_T[p, q]$, o simplement per $p \xrightarrow{T} q$.

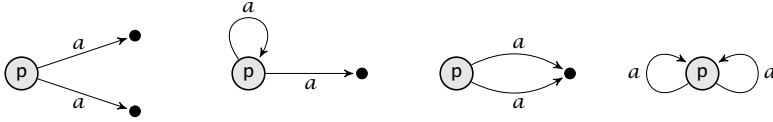


FIGURA 6: Situacions no deterministes al vèrtex p .

Si $\gamma = p_0 e_1 p_1 \cdots e_l p_l$ és un camí en un A -autòmat involutiu Γ (llegint $\ell(\gamma) = \ell(e_1) \cdots \ell(e_l)$), aleshores s'anomena *camí invers* de γ a $\gamma^{-1} = p_l e_l^{-1} p_{l-1} \cdots e_1^{-1} p_0$ (llegint $\ell(\gamma^{-1}) = \ell(e_l)^{-1} \cdots \ell(e_1)^{-1}$). És clar que, també, $\bar{\ell}(\gamma^{-1}) = \bar{\ell}(\gamma)^{-1}$.

És el moment de començar a relacionar els A -autòmats amb els grups. Potser la manera més natural de fer-ho és a través de la idea topològica de grup fonamental d'un graf. Com veurem, si admetem etiquetes a les arestes, aquesta idea desemboca en la de subgrup reconegut, que, com es mostra a continuació, és un subgrup concret de $\mathbb{F}_A$, en lloc d'un grup abstracte.

DEFINICIÓ 33. Sigui Γ un graf no dirigit i connex, i $p \in V\Gamma$. El *grup fonamental de Γ al vèrtex p* , designat per $\pi_p(\Gamma)$, és el conjunt de classes de p -camins a Γ mòdul retrocés, o mòdul reducció de camins (que, en el cas dels grafs vistos com a espais topològics, equival a dir mòdul homotopia), juntament amb l'operació de concatenació de p -camins. És ben sabut que, independentment del vèrtex p triat, $\pi_p(\Gamma)$ és un grup lliure de rang igual al nombre (cardinal) d'arestes fora d'un arbre generador T de Γ ; a aquest valor, que no depèn de l'arbre T , l'anomenem el *rang (gràfic)* de Γ , designat per $\text{rk}(\Gamma)$; és fàcil veure que, si Γ és finit, aleshores

$$\text{rk}(\Gamma) - 1 = \#E\Gamma - \#V\Gamma = \frac{1}{2} \sum_{q \in V\Gamma} (\deg(q) - 2), \quad (6)$$

on $\#V\Gamma$ és el nombre de vèrtexs i $\#E\Gamma$ el nombre d'arestes (no dirigides) de Γ .

¹³ En anglès, *backtracking*.

DEFINICIÓ 34. Sigui Γ un A -autòmat involutiu i connex, i $p \in V\Gamma$. És fàcil veure que el conjunt d'etiquetes reduïdes de p -camins a Γ ,

$$\langle \Gamma \rangle_p = \{ \bar{\ell}(\gamma) : \gamma : p \rightsquigarrow p \}, \quad (7)$$

és un subgrup de $\mathbb{F}_A$; l'anomenem *subgrup reconegut per Γ al vèrtex p* . El subgrup $\langle \Gamma \rangle_\bullet$ el designem simplement per $\langle \Gamma \rangle$.

OBSERVACIÓ 35. En general, el subgrup reconegut per un A -autòmat Γ i el grup fonamental de (el subgraf subjacent de) Γ en un vèrtex p són conceptes diferents, però estretament lligats per l'aplicació *llegir etiquetes reduïdes*:

$$\begin{aligned} \tilde{\ell}_\Gamma : \pi_p(\Gamma) &\twoheadrightarrow \langle \Gamma \rangle_p \leq \mathbb{F}_A \\ [\gamma] &\mapsto \bar{\ell}(\gamma). \end{aligned} \quad (8)$$

És clar que $\tilde{\ell}_\Gamma$ està ben definit i és un homomorfisme de grups exhaustiu i *no injectiu, en general*. Més endavant veurem una condició sobre Γ , sota la qual $\tilde{\ell}_\Gamma$ serà injectiu.

La definició 34 (prenent $p = \bullet$) determina l'aplicació següent, que relaciona els A -autòmats amb el reticle de subgrups de $\mathbb{F}_A$:

$$\begin{aligned} \{(\text{classes d'isomorfia de}) A\text{-autòmats}\} &\rightarrow \{\text{subgrups de } \mathbb{F}_A\} \\ \Gamma &\mapsto \langle \Gamma \rangle. \end{aligned} \quad (9)$$

És fàcil veure que aquesta aplicació és exhaustiva; és a dir, tot subgrup $H \leq \mathbb{F}_A$ és el subgrup reconegut per algun A -autòmat. En efecte, si $S = \{w_i\}_{i \in I} \subseteq \mathbb{F}_A$ és un conjunt de paraules reduïdes que generen H , aleshores considerem, per a cada $w_i = a_{i_1} a_{i_2} \cdots a_{i_l} \in S$ ($a_j \in A^\pm$), el cicle orientat llegint w_i (ó w_i^{-1} en direcció contrària); anomenat *pètal* associat a w_i , i designat per $\text{Fl}(w_i)$ (vegeu la figura 7).

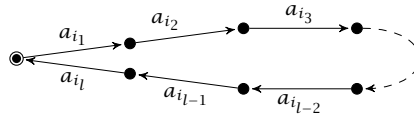
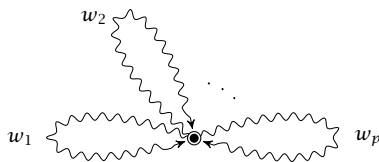


FIGURA 7: El pètal $\text{Fl}(w_i)$.

A continuació definim l'*autòmat flor* $\text{Fl}(S)$ com l'autòmat obtingut després d'identificar els punts base dels pètals associats als elements de S (vegeu la figura 8).

FIGURA 8: L'autòmat flor $\text{Fl}(w_1, w_2, \dots, w_p)$.

És clar que les etiquetes reduïdes dels $\bullet$ -camins de $\text{Fl}(S)$ descriuen precisament el subgrup H ; i.e., $\langle \text{Fl}(S) \rangle = \langle S \rangle = H \leq \mathbb{F}_A$. Per tant, l'aplicació (9) és exhaustiva. Observeu també que (9) està lluny de ser injectiva ja que, per exemple, diferents famílies de generadors per a H proporcionen diferents autòmats flor, tots preimatge de H per (9).

El resultat clau de Stallings a [50] ens diu que es pot eliminar tota la redundància de la representació donada per (9) (i assolir un únic representant Γ per a cada subgrup $H \leq \mathbb{F}_A$) simplement imposant un parell de condicions força naturals sobre els autòmats involucrats: ésser determinista, i ésser cor.

DEFINICIÓ 36. Un A -autòmat Γ es diu *determinista* si de cap dels seus vèrtexs no surten dos arcs diferents amb la mateixa etiqueta; és a dir, si per a cada vèrtex $p \in V\Gamma$, i per a cada parell d'arcs e, e' sortint de p , $\ell(e) = \ell(e')$ implica $e = e'$. (Noteu que un autòmat involutiu és determinista si i només si de cap vèrtex no surten i a cap vèrtex no arriben arcs diferents amb la mateixa etiqueta positiva.)

DEFINICIÓ 37. Un autòmat involutiu es diu *cor*¹⁴ si tot vèrtex (i, per tant, tot arc) apareix en algun $\bullet$ -camí reduït. El *cor* d'un autòmat involutiu Γ , designat per $\text{core}(\Gamma)$, és el màxim subautòmat cor de Γ que conté el punt base. És clar que $\text{core}(\Gamma)$ és un subautòmat induït,¹⁵ de Γ (de fet, de la component connexa de Γ que conté $\bullet$), connex, i que reconeix el mateix subgrup, i.e., $\langle \text{core}(\Gamma) \rangle = \langle \Gamma \rangle$.

Observeu que $\text{core}(\Gamma)$ pot tenir encara un vèrtex de grau 1 que, en cas d'existir, ha de ser necessàriament el punt base a l'*extrem d'un «pèl» penjant de $\text{core}(\Gamma)$* . Eliminant aquest eventual pèl obtenim el que anomenem *cor restringit* de Γ , designat per $\text{core}^*(\Gamma)$ (formalment, el dígraf etiquetat obtingut després d'eliminar successivament tots els vèrtexs de grau 1 de $\text{core}(\Gamma)$ i ignorar el vèrtex base).

OBSERVACIÓ 38. És fàcil veure que $\text{rk}(\text{core}^*(\Gamma)) = \text{rk}(\text{core}(\Gamma)) = \text{rk}(\Gamma)$ (vegeu la subsecció 4.6), i que $\langle \text{core}^*(\Gamma) \rangle_p$ és conjugat a $\langle \text{core}(\Gamma) \rangle = \langle \Gamma \rangle$, per a tot $p \in V(\text{core}^*(\Gamma))$ (vegeu la subsecció 4.4).

DEFINICIÓ 39. Un A -autòmat involutiu es diu *reduït* si és determinista i cor.

¹⁴ En anglès, *core*.

¹⁵ Un subautòmat és *induït* si conté tots els arcs (de l'autòmat original) incidents als seus vèrtexs.

EXEMPLE 40. Principals variants d'autòmat involutiu i determinista:

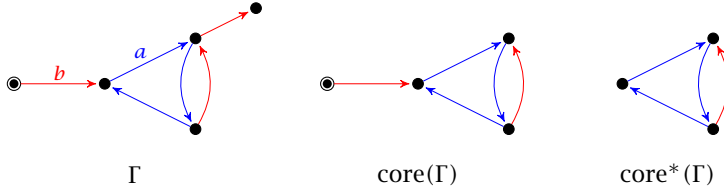


FIGURA 9: D'esquerra a dreta, un autòmat determinista no reduït Γ (amb dos pès), el seu cor $\text{core}(\Gamma)$, i el seu cor restringit $\text{core}^*(\Gamma)$.

Per veure que cada fibra de (9) admet un únic representant reduït adaptarem un resultat clàssic de teoria de llenguatges.

DEFINICIÓ 41. Siguin $\Gamma = (V, E, \iota, \tau, \ell, \odot)$ i $\Gamma' = (V', E', \iota', \tau', \ell', \odot')$ A -autòmats. Un *homomorfisme (de A -autòmats)* de Γ a Γ' és una funció $\theta: V \rightarrow V'$ que envia el punt base al punt base ($\odot\theta = \odot'$) i tal que, per a cada $p, q \in V$ i cada $a \in A$, si $p \xrightarrow{a} q$ aleshores $p\theta \xrightarrow{a} q\theta$. Habitualment, abusem lleugerament del llenguatge i el dissenyem per $\theta: \Gamma \rightarrow \Gamma'$.

PROPOSICIÓ 42. Siguin Γ i Γ' A -autòmats reduïts. Aleshores, $\langle \Gamma \rangle \leq \langle \Gamma' \rangle$ si i només si existeix un homomorfisme $\Gamma \rightarrow \Gamma'$ i, en aquest cas, l'homomorfisme és únic.

PROVA. La unicitat és conseqüència del determinisme de Γ' . En efecte, suposem que $\theta_1, \theta_2: \Gamma \rightarrow \Gamma'$ són homomorfismes, i $p \in V\Gamma$ un vèrtex tal que $p\theta_1 \neq p\theta_2$. Aleshores les imatges per θ_1 i θ_2 de qualsevol camí $\gamma: \odot \rightsquigarrow p$ a Γ seran camins a Γ' amb el mateix origen, $\odot'$, i la mateixa etiqueta, $\ell(\gamma)$, però diferent final, $p\theta_1 \neq p\theta_2$, en contradicció amb el determinisme de Γ' .

[$\Leftarrow$] la implicació cap a l'esquerra és certa sense cap hipòtesi addicional sobre Γ, Γ' : si $\theta: \Gamma \rightarrow \Gamma'$ és un homomorfisme d'autòmats, aleshores la imatge de qualsevol $\odot$ -camí a Γ és un $\odot'$ -camí a Γ' amb la mateixa etiqueta, i per tant, la mateixa etiqueta reduïda; per tant, $\langle \Gamma \rangle \leq \langle \Gamma' \rangle$.

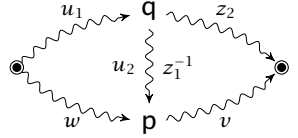
[$\Rightarrow$] Suposem ara que Γ i Γ' són A -autòmats reduïts amb $\langle \Gamma \rangle \leq \langle \Gamma' \rangle$ i construïm un homomorfisme $\theta: \Gamma \rightarrow \Gamma'$. Òbviament, prenem $\odot\theta = \odot'$. Per a tot vèrtex $p \neq \odot$, considerem un $\odot$ -camí reduït γ a Γ que passi per p :

$$\gamma: \odot \rightsquigarrow^u p \rightsquigarrow^v \odot$$

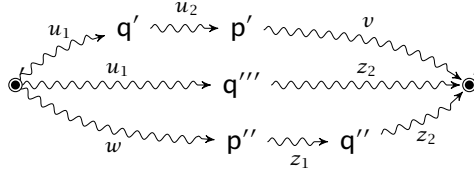
(sabem que n'existeix almenys un perquè Γ és cor). A més, com que Γ és també determinista, tenim que tant u com v són paraules reduïdes, i sense cancel·lació entre elles. En particular, $uv \in \langle \Gamma \rangle$ i per tant $uv \in \langle \Gamma' \rangle$. Això vol dir que uv és l'etiqueta reduïda d'un $\odot'$ -camí a Γ' i, com que Γ' és determinista, uv és també l'etiqueta d'un $\odot'$ -camí reduït γ' a Γ' , que podem descompondre com a

$$\gamma': \odot' \rightsquigarrow^u p' \rightsquigarrow^v \odot'.$$

Definim $p\theta = p'$. Per veure que està ben definit, sigui $\bullet \xrightarrow{w} p \xrightarrow{z} \bullet$ un altre $\bullet$ -camí reduït de Γ que passa per p , considerem $\bullet' \xrightarrow{w} p'' \xrightarrow{z} \bullet'$ el corresponent $\bullet'$ -camí reduït a Γ' , i vegem que $p' = p''$. Observeu que $\bullet \xrightarrow{u} p \xrightarrow{z} \bullet$ és un $\bullet$ -camí a Γ amb possible retrocés a p ; si anomenem q el vèrtex on acaba aquest retrocés, aleshores tenim



on $u = u_1 u_2$, $z = z_1 z_2$, $u_2 = z_1^{-1}$, i el camí $\bullet \xrightarrow{u_1} q \xrightarrow{z_2} \bullet$ és reduït. Per tant, $uv = u_1 u_2 v$, $wz = w z_1 z_2$ i $u_1 z_2$ han de ser llegibles (sense cancel·lació entre síl·labes) com a etiquetes de $\bullet'$ -camins reduïts a Γ' :



Finalment, pel determinisme de Γ' , és clar, primer que $q' = q''' = q''$, i després (tenint en compte que $u_2 = z_1^{-1}$) que $p' = p''$, tal com reclamàvem. Per tant, $\theta: V\Gamma \rightarrow V\Gamma'$ està ben definit.

Per acabar, vegem que θ és un homomorfisme d'autòmats entre Γ i Γ' . En efecte, com que Γ és cor, donat un arc $e \equiv p \xrightarrow{a} q$ a Γ , sabem que existeix un $\bullet$ -camí reduït a Γ que usa l'arc e :

$$\bullet \xrightarrow{u} p \xrightarrow{a} q \xrightarrow{v} \bullet.$$

Per tant, $uav \in \langle \Gamma \rangle \leq \langle \Gamma' \rangle$ (sense cancel·lació entre síl·labes). I, per la definició de θ , a Γ' hi tenim el $\bullet'$ -camí reduït

$$\bullet' \xrightarrow{u} p\theta \xrightarrow{a} q\theta \xrightarrow{v} \bullet'.$$

En particular, tenim un arc $p\theta \xrightarrow{a} q\theta$ a Γ' , i per tant, θ és un homomorfisme d'autòmats de Γ a Γ' , tal com volíem demostrar. $\square$

Els corollaris següents són immediats, i capturen exactament la propietat que perseguim.

COROLLARI 43. Si Γ és un autòmat reduït, aleshores l'únic homomorfisme $\Gamma \rightarrow \Gamma$ és la identitat. $\square$

COROLLARI 44. Si dos A -autòmats reduïts reconeixen el mateix subgrup, aleshores són isomorfs. És a dir, si Γ i Γ' són A -autòmats reduïts, aleshores

$$\langle \Gamma \rangle = \langle \Gamma' \rangle \Leftrightarrow \Gamma \simeq \Gamma'.$$

Per tant, d'entre tots els A -autòmats que reconeixen un cert subgrup $H \leq \mathbb{F}_A$, serà suficient triar-ne un de reduït per tal d'obtenir un representant únic, i restringir (9) a una bijecció. Vegem a continuació que aquest representant reduït sempre existeix, i com es relaciona amb un altre autòmat naturalment associat al subgrup H corresponent.

DEFINICIÓ 45. Siguin G un grup, H un subgrup de G , i $S \subseteq G$ un conjunt de generadors per a G . Aleshores, l'autòmat de Schreier (per la dreta) de H respecte a S , designat per $\text{Sch}_G(H, S)$, és el S -autòmat amb conjunt de vèrtexs $H \backslash G$ (el conjunt de classes laterals per la dreta de G mòdul H), un arc $Hg \xrightarrow{s} Hgs$ per a cada classe lateral $Hg \in H \backslash G$ i cada element $s \in S^\pm$, i la classe lateral H com a vèrtex base.

OBSERVACIÓ 46. Noteu que, si $H \trianglelefteq G$, llavors $\text{Sch}_G(H, S) = \text{Cay}(G/H, S)$. En particular, $\text{Sch}_G(1, S) = \text{Cay}(G, S)$.

Donat que d'ara endavant sempre considerarem $G = \mathbb{F}_A$, prendrem habitualment la base A com a conjunt de generadors i escriurem simplement $\text{Sch}_{\mathbb{F}_A}(H, A) = \text{Sch}(H, A)$. El A -autòmat $\text{Sch}(H, A)$ té les propietats següents.

PROPOSICIÓ 47. *Sigui H un subgrup de $\mathbb{F}_A$. Aleshores,*

- (i) $\text{Sch}(H, A)$ és un A -autòmat involutiu, determinista i saturat;
- (ii) $\text{Sch}(H, A)$ és connex;
- (iii) $\langle \text{Sch}(H, A) \rangle = H$;
- (iv) cada arc e no pertanyent a $\text{core}(\text{Sch}(H, A))$ és un pont de $\text{Sch}(H, A)$; és a dir, $\text{Sch}(H, A) \setminus \{e\}$ té dues components connexes i, a més, la que no conté $\bullet$ és un arbre infinit i saturat excepte al vèrtex incident a e ;
- (v) $\text{Sch}(H, A)$ no és necessàriament cor.

PROVA. L'apartat (i) és immediat de la definició d'autòmat de Schreier (definició 45). La connexió de $\text{Sch}(H, A)$ és conseqüència directa del fet que tot vèrtex Hw està connectat amb el punt base: en efecte, posant $w = a_{i_1}^{\epsilon_1} a_{i_2}^{\epsilon_2} \cdots a_{i_l}^{\epsilon_l}$ com a paraula en $A^\pm$, tenim el camí

$$\gamma: \bullet \xrightarrow{a_{i_1}^{\epsilon_1}} Ha_{i_1}^{\epsilon_1} \xrightarrow{a_{i_2}^{\epsilon_2}} Ha_{i_1}^{\epsilon_1} a_{i_2}^{\epsilon_2} \xrightarrow{a_{i_3}^{\epsilon_3}} \cdots \xrightarrow{a_{i_l}^{\epsilon_l}} Hw. \quad (10)$$

Això demostra (ii). L'apartat (iii) se segueix del fet que el camí (10) és tancat si i només si $Hw = \bullet = H$, és a dir, si i només si $\ell(\gamma) = w \in H$.

Per veure (iv) considerem un arc e de $\text{Sch}(H, A)$ no pertanyent a $\text{core}(\text{Sch}(H, A))$. Si hi hagués camins, i per tant camins reduïts, γ i η a $\text{Sch}(H, A) \setminus \{e\}$, de $\bullet$ a ιe , i de $\bullet$ a τe , respectivament, llavors $\gamma e \eta^{-1}$ seria un $\bullet$ -camí reduït, cosa que contradiu que e no forma part del cor. Per tant, e és un pont a $\text{Sch}(H, A)$. Canviant e per e^{-1} si cal, podem suposar que hi ha un camí γ de $\bullet$ a ιe , que no passa per l'arc e . Llavors, si la component connexa de $\text{Sch}(H, A) \setminus \{e\}$ que conté τe no fos arbre, hi hauria un τe -camí reduït no trivial ξ , i $\gamma e \xi e^{-1} \gamma^{-1}$

seria un $\bullet$ -camí reduït que passa per e i contradiu un altre cop el fet que e no pertany al cor. Per tant, la component connexa de $\text{Sch}(H, A) \setminus \{e\}$ que conté τe és un arbre; i haurà de ser forçosament infinit ja que si no, tindria vèrtexs de grau 1, cosa que contradiria el fet que $\text{Sch}(H, A)$ és saturat. Com que cada vèrtex és saturat, aquest arbre és necessàriament una branca de Cayley (vegeu la figura 10).

Finalment, per a veure (v) només cal observar l'autòmat $\text{Sch}(\langle b \rangle, \{a, b\})$, dibuixat a la figura 10: un b -llaç al punt base $\bullet$, i dues branques de Cayley naixent dels dos arcs disponibles al punt base. $\square$

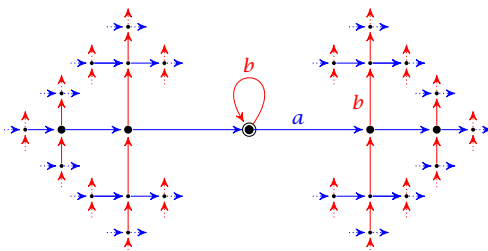


FIGURA 10: L'autòmat de Schreier de $\langle b \rangle \leq \mathbb{F}_{\{a,b\}}$.

De la proposició 47 és clar que la propietat d'ésser cor és l'única que li falta a $\text{Sch}(H, A)$ per a ser un autòmat reduït que reconeix H .

DEFINICIÓ 48. Sigui H un subgrup de $\mathbb{F}_A$. Anomenem *autòmat de Stallings* de H respecte a A , designat per $\text{St}(H, A)$, el cor de $\text{Sch}(H, A)$; és a dir $\text{St}(H, A) = \text{core}(\text{Sch}(H, A))$. També definim l'*autòmat restringit de Stallings* de H respecte a A com el digraf $\text{St}^*(H, A) = \text{core}^*(\text{St}(H, A))$. Si la base de referència A és clara pel context, habitualment l'ometrem i escriurem simplement $\text{Sch}(H)$, $\text{St}(H)$, i $\text{St}^*(H)$.

EXEMPLE 49. (La part positiva de) l'autòmat de Stallings $\text{St}(\mathbb{F}_A, A)$ té un únic vèrtex i un llaç etiquetat per cada $a \in A$. Aquest autòmat s'anomena *rosa* (de $\#A$ pètals), i es designa per $\mathcal{B}_A$ o $\mathcal{B}_n$, on $n = \#A$:

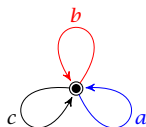


FIGURA 11: $\mathcal{B}_3$, una rosa de tres pètals.

OBSERVACIÓ 50. És clar que, per a qualsevol cardinal n , el grup fonamental de la rosa de n pètals és el grup lliure de rang n ; és a dir, $\pi_{\bullet}(\mathcal{B}_n) \simeq \mathbb{F}_n$.

A continuació resumim les propietats principals de l'autòmat de Stallings i la seva relació amb el de Schreier.

PROPOSICIÓ 51. *Segui H un subgrup de $\mathbb{F}_A$. Aleshores,*

- (i) $\text{St}(H, A)$ és un A -autòmat involutiu, determinista, i cor (per tant, reduït);
- (ii) $\langle \text{St}(H, A) \rangle = H$;
- (iii) $\text{St}(H, A)$ és saturat $\Leftrightarrow \text{St}(H, A) = \text{Sch}(H, A) \Leftrightarrow \text{Sch}(H, A)$ és cor;
- (iv) $\text{St}(H, A)$ no és necessàriament saturat;
- (v) $\text{Sch}(H, A)$ consisteix en $\text{St}(H, A)$ amb una a -branca de Stallings adjacent a cada vèrtex a -deficient de $\text{St}(H, A)$, on $a \in A^\pm$;
- (vi) l'índex $|\mathbb{F}_A : H|$ és finit si i només si $\text{St}(H, A)$ és saturat i $\#V\text{St}(H, A) < \infty$.

PROVA. L'apartat (i) és immediat de la definició d'autòmat de Stallings (definició 48), mentre que (ii) se segueix de la mateixa definició i l'observació 38.

Per a veure (iii) només cal recordar que $\text{Sch}(H)$ és sempre saturat, $\text{St}(H)$ és sempre cor, i $\text{St}(H) = \text{core}(\text{Sch}(H))$. Per tant, (iv) és conseqüència de la proposició 47(v).

(v) Si p és un vèrtex a -deficient de $\text{St}(H)$, l'arc de $\text{Sch}(H)$ llegint a des de p no pertany a $\text{St}(H)$ i aleshores el resultat reclamat no és més que una reinterpretació de la proposició 47(iv).

(vi) Recordem que $|\mathbb{F}_A : H| = \#V\text{Sch}(H, A)$. Si $\text{St}(H)$ és saturat i té un nombre finit de vèrtexs, aleshores $|\mathbb{F}_A : H| = \#V\text{Sch}(H, A) = \#V\text{St}(H, A) < \infty$, on hem usat l'apartat (iii) a la segona igualtat. Recíprocament, suposem $|\mathbb{F}_A : H| < \infty$: per una banda $\#V\text{St}(H, A) \leq \#V\text{Sch}(H, A) = |\mathbb{F}_A : H| < \infty$; i per una altra, si $\text{St}(H, A)$ fos insaturat, (v) implicaria que $\text{Sch}(H, A)$ té arbres infinits penjant de tots els vèrtexs insaturats, cosa que contradiu que $\#V\text{Sch}(H, A) = |\mathbb{F}_A : H| < \infty$. $\square$

És convenient tenir present que, com a conseqüència de la teoria desenvolupada, les propietats (i) i (ii) de la proposició 47, i la propietat (i) de la proposició 51, caracteritzen, de fet, els respectius tipus d'autòmats.

COROLLARI 52. *Segui Γ un A -autòmat involutiu i determinista. Aleshores:*

- (i) si Γ és saturat i connex, aleshores $\Gamma \simeq \text{Sch}(\langle \Gamma \rangle, A)$;
- (ii) si Γ és cor, aleshores $\Gamma \simeq \text{St}(\langle \Gamma \rangle, A)$. $\square$

Disposem ja de tots els ingredients necessaris per a establir la bijecció que buscàvem entre A -autòmats i subgrups de $\mathbb{F}_A$, tot restringint (9) al domini adequat.

TEOREMA 53 (J. R. STALLINGS [50]). *La funció*

$$\{(classes\ d'isomorfia\ de)\ A\text{-autòmats reduïts}\} \rightarrow \{subgrups\ de\ \mathbb{F}_A\} \quad (11)$$

$$\Gamma \mapsto \langle \Gamma \rangle$$

és una bijecció, amb inversa $\text{St}(H, A) \mapsto H$.

PROVA. Les dues aplicacions estan ben definides: en efecte, per a tot Γ , $\langle \Gamma \rangle$ és un subgrup de $\mathbb{F}_A$; i, gràcies a la proposició 51(i), per a tot subgrup $H \leq \mathbb{F}_A$, $\text{St}(H, A)$ és un A -autòmat reduït. I són una inversa de l'altra: per la proposició 51(ii), $\langle \text{St}(H, A) \rangle = H$ i, com que Γ és reduït, $\text{St}(\langle \Gamma \rangle, A) = \Gamma$, a conseqüència del corollari 44. $\square$

El pas següent és garantir el caràcter computacional de la bijecció (11), que serà fonamental per als nostres interessos. Veurem que els subgrups *finitament generats* corresponen precisament als A -autòmats reduïts *finitos* i que la conversió entre ambdós és algorísmica. Trobarem mètodes eficients per a computar les dues direccions de la bijecció (11): (a) donat un A -autòmat reduït i finit Γ , calcularem explícitament un sistema de generadors finit de $\langle \Gamma \rangle$ (millor encara, en calcularem una base, i demostrarem de passada que $\langle \Gamma \rangle$ sempre torna a ser un grup lliure); i (b) donat un sistema de generadors finit de $\langle \Gamma \rangle$, construirem el seu autòmat (finit) de Stallings $\text{St}(H, A)$.

PROPOSICIÓ 54. *Sigui Γ un A -autòmat involutiu i connex, sigui T un arbre generador de Γ , i considerem*

$$S_T = \{w_e : e \in E^+\Gamma \setminus E\Gamma\} \subseteq \langle \Gamma \rangle \leq \mathbb{F}_A, \quad (12)$$

on $w_e = \bar{\ell}(\odot \xrightarrow{T} \bullet \xrightarrow{e} \bullet \xrightarrow{T} \odot)$. Llavors,

- (i) S_T és un sistema de generadors per $\langle \Gamma \rangle$;
- (ii) si Γ és determinista, aleshores $\langle \Gamma \rangle$ és un grup lliure i S_T n'és una base; en particular, el rang algebraic $\text{rk}(\langle \Gamma \rangle)$ i el rang gràfic $\text{rk}(\Gamma)$ coincideixen;
- (iii) si Γ és reduït, aleshores $\langle \Gamma \rangle$ és finitament generat si i només si Γ és finit; i en aquest cas, $\text{rk}(\langle \Gamma \rangle) = 1 - \#\Gamma + \#E^+\Gamma$.

PROVA. (i) Sigui $w \in \langle \Gamma \rangle$, i expressem-lo com a $w = \bar{\ell}(y)$, per a cert $\odot$ -camí reduït y a Γ . Si distingim les visites de y a arcs fora de T , podem escriure

$$y: \odot \xrightarrow{T} \bullet \xrightarrow{e_1^{\epsilon_1}} \bullet \xrightarrow{T} \bullet \xrightarrow{e_2^{\epsilon_2}} \bullet \xrightarrow{T} \bullet \xrightarrow{\dots} \bullet \xrightarrow{T} \bullet \xrightarrow{e_l^{\epsilon_l}} \bullet \xrightarrow{T} \odot,$$

on $e_1, \dots, e_l \in E^+\Gamma \setminus E\Gamma$ (amb possibles repeticions) i $\epsilon_j = \pm 1$. Observem ara que y té la mateixa etiqueta reduïda que el camí ampliat

$$y': \odot \xrightarrow{T} \bullet \xrightarrow{e_1^{\epsilon_1}} \bullet \xrightarrow{T} \odot \xrightarrow{T} \bullet \xrightarrow{e_2^{\epsilon_2}} \bullet \xrightarrow{T} \odot \dots \odot \xrightarrow{T} \bullet \xrightarrow{e_l^{\epsilon_l}} \bullet \xrightarrow{T} \odot. \quad (13)$$

Per tant, $w = \bar{\ell}(y) = \bar{\ell}(y') = w_{e_1^{\epsilon_1}} w_{e_2^{\epsilon_2}} \dots w_{e_l^{\epsilon_l}} \in \langle S_T \rangle$.

(ii) Suposem ara, a més, que Γ és determinista. Demostrarem que S_T és, de fet, una base de $\langle \Gamma \rangle$ i que, per tant, $\langle \Gamma \rangle$ és un grup lliure. És suficient veure que qualsevol paraula reduïda no buida en S_T representa un element no trivial de $\mathbb{F}_A$. En efecte, sigui $w = w_{e_1^{\epsilon_1}} w_{e_2^{\epsilon_2}} \dots w_{e_l^{\epsilon_l}}$, amb $l \geq 1$, $e_1, \dots, e_l \in E^+\Gamma \setminus E\Gamma$ (amb possibles repeticions), i tal que no hi ha dos factors successius inversos l'un de l'altre, i. e., $e_i = e_{i+1}$ implica $\epsilon_i = \epsilon_{i+1}$. Aleshores,

$$\begin{aligned} \bar{w} &= \overline{w_{e_1^{\epsilon_1}} w_{e_2^{\epsilon_2}} \dots w_{e_l^{\epsilon_l}}} \\ &= \bar{\ell}(\odot \xrightarrow{T} \bullet \xrightarrow{e_1^{\epsilon_1}} \bullet \xrightarrow{T} \odot \xrightarrow{T} \bullet \xrightarrow{e_2^{\epsilon_2}} \bullet \xrightarrow{T} \odot \dots \odot \xrightarrow{T} \bullet \xrightarrow{e_l^{\epsilon_l}} \bullet \xrightarrow{T} \odot) \\ &= \bar{\ell}(\odot \xrightarrow{T} \bullet \xrightarrow{e_1^{\epsilon_1}} \bullet \xrightarrow{T} \bullet \xrightarrow{e_2^{\epsilon_2}} \bullet \xrightarrow{T} \bullet \dots \bullet \xrightarrow{T} \bullet \xrightarrow{e_l^{\epsilon_l}} \bullet \xrightarrow{T} \odot). \end{aligned}$$

Però aquest últim $\odot$ -camí és no buit ($l \geq 1$), i reduït (donat que els arcs $e_i^{\epsilon_i}$ estan tots fora de T , i dos de consecutius mai no presenten retrocés). Com que Γ és determinista, això significa que $\overline{w} \neq 1$, com volíem demostrar. Per tant, S_T és una base de $\langle \Gamma \rangle$ i el rang algebraic i el rang gràfic coincideixen, $\text{rk}(\langle \Gamma \rangle) = \#S_T = \text{rk}(\Gamma)$.

(iii) Finalment, suposem que Γ és reduït. Si Γ és finit, llavors $\#(E^+\Gamma \setminus ET) < \infty$ i $\langle \Gamma \rangle$ és finitament generat. Recíprocament, si $\Gamma = \text{core}(\Gamma)$ té rang finit, aleshores és finit (ja que, afegint un nombre finit d'arcs a un arbre infinit, no resulta mai un cor). I en aquest cas, $\text{rk}(\langle \Gamma \rangle) = \text{rk}(\Gamma) = 1 - \#\text{V}\Gamma + \#E^+\Gamma$. La demostració està completa. $\square$

EXEMPLE 55. Considereu l'autòmat determinista següent, i l'arbre generador T format pels arcs horitzontals (representats amb traç més gruixut).



D'acord amb la proposició 54(ii), el subgrup $\langle \Gamma \rangle$ reconegut per Γ és lliure amb base el conjunt $S_T = \{b, a^3, a^2bab^{-1}a^{-2}\}$ i, per tant, de rang 3.

El següent resultat fonamental, demostrat primer per J. Nielsen, per a subgrups finitament generats, i uns anys més tard per O. Schreier, amb total generalitat, se segueix ara de manera transparent de la teoria de Stallings.

TEOREMA 56 (TEOREMA DE NIELSEN-SCHREIER, [40, 45]). *Tot subgrup d'un grup lliure és lliure.*

PROVA. Pel teorema 53, tot subgrup $H \leq \mathbb{F}_A$ és de la forma $H = \langle \text{St}(H, A) \rangle$; i, per la proposició 54, és lliure. $\square$

Finalment, situem-nos en el cas finit, i busquem un algorisme eficient per computar els dos sentits de la bijecció (11). El càlcul d'una base del subgrup reconegut per un A -autòmat reduït i finit Γ és directe de la proposició anterior: calculem un arbre generador T de Γ i, per a cada arc $e \in E^+\Gamma \setminus ET$ (n'hi ha un nombre finit), calculem la paraula reduïda $w_e = \overline{\ell}(\odot \xrightarrow{T} \bullet \xrightarrow{e} \bullet \xrightarrow{T} \odot)$. Per la proposició 54(ii), $S_T = \{w_e : e \in E^+\Gamma \setminus ET\}$ és una base de $\langle \Gamma \rangle \leq \mathbb{F}_A$.

Per a computar la direcció contrària, donat un subgrup $H \leq \mathbb{F}_A$ (mitjançant una família finita S de generadors, que podem suposar paraules reduïdes), hem de trobar una manera efectiva de calcular $\text{St}(H, A)$; o, el que és el mateix, una manera de calcular un A -autòmat Γ que sigui involutiu, reduït, finit, i que reconegui $\langle \Gamma \rangle = H$. Ja coneixem un A -autòmat, molt fàcil de construir, i que gairebé compleix totes aquestes condicions: l'autòmat flor $\text{Fl}(S)$ és involutiu, cor, finit, reconeix $\langle \text{Fl}(S) \rangle = \langle S \rangle = H$, i és determinista *excepte potser al vèrtex base* $\odot$ (ja que els diferents elements de S poden començar o acabar amb la mateixa lletra de $A^\pm$). Per a «reparar» l'eventual no determinisme de $\text{Fl}(S)$, necessitem un procediment que converteixi un A -autòmat finit en determinista (sense perdre les altres propietats); el desenvolupem a continuació.

DEFINICIÓ 57. Sigui Γ un A -autòmat involutiu; i siguin e_1 i e_2 dos arcs de Γ que violen el determinisme, i. e., que compleixen $\iota e_1 = \iota e_2$, $\ell(e_1) = \ell(e_2)$, però $e_1 \neq e_2$. Anomenem *plegament (elemental) de Stallings*,¹⁶ i el designem per $\Gamma \rightsquigarrow \Gamma'$, la transformació consistent a identificar els arcs e_1 i e_2 (i els seus corresponents inversos) a Γ . Si els arcs e_1 i e_2 no són paral·lels (i. e., si $\tau e_1 \neq \tau e_2$), direm que és un *plegament obert*; i en cas contrari, direm que és un *plegament tancat* (vegeu la figura 12).



FIGURA 12: Un plegament obert (esquerra) i un de tancat (dreta).

Un senzill recompte de vèrtexs i arcs proporciona el resultat següent, que serà rellevant més endavant.

OBSERVACIÓ 58. Si Γ és finit i $\Gamma \rightsquigarrow \Gamma'$ és un plegament elemental de Stallings, aleshores:

$$\text{rk}(\Gamma') = \begin{cases} \text{rk}(\Gamma) & \text{si } \Gamma \rightsquigarrow \Gamma' \text{ és obert,} \\ \text{rk}(\Gamma) - 1 & \text{si } \Gamma \rightsquigarrow \Gamma' \text{ és tancat.} \end{cases} \quad (14)$$

És clar que els plegaments de Stallings no modifiquen el subgrup reconegut.

LEMA 59. Si $\Gamma \rightsquigarrow \Gamma'$ és un plegament de Stallings, aleshores $\langle \Gamma' \rangle = \langle \Gamma \rangle$. $\square$

Llavors, si Γ és un A -autòmat involutiu, finit, reconeixent H , el *procés de Stallings* per a convertir Γ en determinista és com segueix: si Γ ja és determinista, no cal fer res; en cas contrari, detectem successivament possibles plegaments (oberts o tancats) i els realitzem fins que no en quedi cap de disponible, és a dir, fins a obtenir un autòmat determinista. Observeu que realitzant un plegament podem provocar-ne de nous i fins i tot *augmentar* el nombre de plegaments visibles a fer. Malgrat això, està garantit que arribarem a una situació determinista amb un nombre finit de passes ja que l'autòmat inicial Γ és finit, i en cada plegament el nombre d'arcs decreix exactament en una unitat. Un cop arribem a un A -autòmat determinista, prenem el cor i designem per Γ' el A -autòmat reduït resultant. Observeu que, pel lema 59 i l'observació 38, $\langle \Gamma' \rangle = \langle \Gamma \rangle$. I pel corollari 44, Γ' ha de ser (isomorf a) l'autòmat de Stallings $\text{St}(H, A)$. El procés descrit s'anomena (una) *seqüència de plegaments de Stallings* per a Γ :

$$\Gamma = \Gamma_0 \xrightarrow{\phi_1} \Gamma_1 \xrightarrow{\phi_2} \dots \xrightarrow{\phi_p} \Gamma_p \xrightarrow{\text{core}} \text{core}(\Gamma_p) \simeq \text{St}(H, A).$$

¹⁶ En anglès, *stallings folding*.

Observeu que hi pot haver, en principi, diferents plegaments —i per tant, diferents seqüències de plegaments— disponibles a Γ . Ara bé, com que l'autòmat final és reduït, el corollari 44 ens garanteix que l'objecte final obtingut serà sempre el mateix (mòdul isomorfisme), *independentment de la seqüència de plegaments seguida*.

Construint una seqüència de plegaments de Stallings per a l'autòmat flor $\text{Fl}(S)$, obtenim una manera efectiva de calcular $\text{St}(H, A)$ a partir de qualsevol conjunt de generadors finit S per a H . No és difícil demostrar que, en aplicar una seqüència de plegaments de Stallings a l'autòmat flor $\text{Fl}(S)$ d'un conjunt S de paraules *reduïdes*, cap dels plegaments de la seqüència no trenca el caràcter cor de l'autòmat inicial i, per tant, no és necessari realitzar el pas final de prendre el cor:

$$\text{Fl}(S) = \Gamma_0 \xrightarrow{\phi_1} \Gamma_1 \xrightarrow{\phi_2} \cdots \xrightarrow{\phi_p} \Gamma_p = \text{St}(H, A).$$

Observeu finalment que el resultat de la seqüència de plegaments, $\text{St}(H, A)$, depèn canònicament del subgrup H i, per tant, *també és independent del sistema de generadors finit S de H amb què haguem començat el procés* (mentre que $\text{Fl}(S)$ en depèn fortament). Això conclou la computació de la bijecció (11).

Incorporant a l'enunciat del teorema 53 els resultats que acabem de demostrar, obtenim el teorema principal d'aquesta secció.

TEOREMA 60 (J. R. STALLINGS [50]). *La funció*

$$\begin{aligned} \text{St}: \{\text{subgrups de } \mathbb{F}_A\} &\rightarrow \{(classes d'isomorfia de) A\text{-autòmats reduïts}\} \\ H &\mapsto \text{St}(H, A) \end{aligned} \quad (15)$$

és una bijecció, amb inversa $\langle \Gamma \rangle \mapsto \Gamma$. A més, els subgrups finitament generats corresponen precisament als autòmats reduïts finits i, en aquest cas, la bijecció és computable en ambdós sentits. $\square$

EXEMPLE 61. Sigui $\mathbb{F}_2$ el grup lliure sobre $A = \{a, b\}$, $S = \{u_1, u_2, u_3\} \subseteq \mathbb{F}_2$ i $H = \langle S \rangle \leq \mathbb{F}_2$ el subgrup generat pels elements $u_1 = a^3$, $u_2 = abab^{-1}$, i $u_3 = a^{-1}bab^{-1}$. Per obtenir una base (i, per tant, el rang) de H , calcularem primer $\text{St}(H, A)$. Comencem construint l'autòmat flor $\Gamma_0 = \text{Fl}(S)$, i anem fent plegaments fins a obtenir l'autòmat de Stallings $\text{St}(H, A)$: a la figura 13 teniu una possible seqüència de Stallings $\Gamma_0 \leadsto \Gamma_1 \leadsto \Gamma_2 \leadsto \Gamma_3 \leadsto \Gamma_4 \leadsto \Gamma_5 \leadsto \Gamma_6 = \text{St}(H)$, tot indicant, en cada pas, amb traç gruixut, els arcs que es pleguen al pas següent (observeu que el primer pas són en realitat tres plegaments fets simultàniament):

Finalment, prenent $\bullet \xrightarrow{b} \bullet$ com a arbre generador de $\text{St}(H)$, la proposició 54(ii) ens diu que $\{a, bab^{-1}\}$ és una base del subgrup H ; en particular, $\text{rk}(H) = 2$.

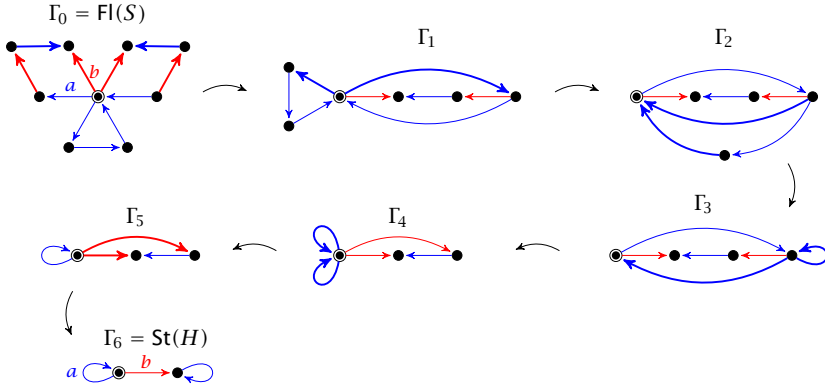


FIGURA 13: Una seqüència de plegaments de Stallings pel subgrup $H = \langle a^3, abab^{-1}, a^{-1}bab^{-1} \rangle \leq \mathbb{F}_{\{a,b\}}$.

Tornem per un moment al morfisme $\tilde{\ell}_\Gamma$ definit a (8). Dèiem més amunt que no era un morfisme injectiu, en general. I, efectivament, pel A -autòmat $\Gamma_0 = \text{Fl}(S)$ de l'exemple anterior (al vèrtex base $\bullet$) no ho és ja que $\pi_\bullet(\Gamma_0)$ és un grup lliure de rang $\text{rk}(\Gamma_0) = 1 - \#\text{V}\Gamma_0 + \#\text{E}^+\Gamma_0 = 1 - 9 + 11 = 3 = \#S$, mentre que $\langle \Gamma_0 \rangle = \langle S \rangle = H$ és un (sub)grup (de $\mathbb{F}_A$) lliure de rang 2. Això vol dir que els tres generadors inicials u_1, u_2, u_3 de H no són base de H , és a dir, no són independents; per tant, hi ha d'haver un producte reduït i no buit d'ells i els seus inversos que doni l'element trivial, $w(u_1, u_2, u_3) = 1$. Això és, exactament, la contrapartida algebraica del fet gràfic que $\tilde{\ell}_\Gamma$ no sigui injectiu: hi ha un element $1 \neq [\gamma] \in \pi_\bullet(\Gamma_0)$ (i. e., un $\bullet$ -camí reduït i no buit al A -autòmat flor Γ_0) tal que $\bar{\ell}(\gamma) = 1$ (i. e., que llegeixi l'etiqueta reduïda trivial). Fent càlculs veiem, per exemple, $u_2 u_3^{-1} u_1^{-1} u_2 u_3^{-1} u_1^{-1} u_2 u_3^{-1} = 1$. Podem anar encara una mica més enllà en aquest paral·lelisme entre àlgebra i geometria, i quantificar la no injectivitat del morfisme $\tilde{\ell}_\Gamma: \pi_\bullet(\Gamma) \twoheadrightarrow \langle \Gamma \rangle \leq \mathbb{F}_A$, $[\gamma] \mapsto \bar{\ell}(\gamma)$, des dels dos punts de vista: una manera algebraica de fer-ho és mirant el nombre mínim de generadors de $\ker(\tilde{\ell}_\Gamma)$ com a subgrup normal de $\pi_\bullet(\Gamma)$; com veurem a continuació, la contrapartida gràfica és el nombre de plegaments tancats que apareixen en una (i, per tant, en qualsevol) seqüència de plegaments de Stallings de Γ a $\text{St}(H, A)$.

OBSERVACIÓ 62. Si Γ és determinista, llavors $\tilde{\ell}_\Gamma$ és injectiu. $\square$

És clar que la pèrdua de rang gràfic en qualsevol seqüència de Stallings per un A -autòmat finit Γ és precisament el nombre de plegaments tancats, ja que els plegaments oberts no alteren el rang, i els tancats el redueixen en una unitat (vegeu l'observació 58). Es pot usar un argument geomètric-algebraic per a demostrar que aquest nombre també coincideix amb la quantitat mínima de generadors de $\ker(\tilde{\ell}_\Gamma)$ com a subgrup normal de $\pi_\bullet(\Gamma)$ (per raons d'espai, ometem els detalls tècnics d'aquesta demostració).

PROPOSICIÓ 63. *Sigui Γ un A -autòmat involutiu, connex, i finit, que reconeix el subgrup $H \leq \mathbb{F}_A$, i sigui*

$$\Gamma \xrightarrow{\phi_1} \Gamma_1 \xrightarrow{\phi_2} \dots \xrightarrow{\phi_p} \Gamma_p \xrightarrow{\text{core}} \text{core}(\Gamma_p) = \text{St}(H, A) \quad (16)$$

una seqüència de plegaments de Stallings per a Γ . Aleshores, els tres nombres següents coincideixen:

- (a) *el nombre de plegaments tancats a (16);*
- (b) $\text{rk}(\Gamma) - \text{rk}(\text{St}(H, A))$;
- (c) *el nombre mínim de generadors de $\ker(\tilde{\ell}_\Gamma)$ com a subgrup normal de $\pi_\bullet(\Gamma)$.*

Per tant, aquest nombre és independent de la seqüència concreta de plegaments i només depèn de Γ .

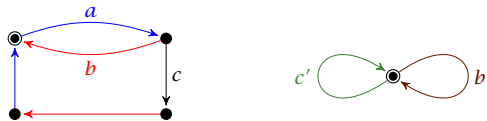
Aquest valor —independent de la seqüència concreta de plegaments i només dependent de Γ — s'anomena la *pèrdua* de Γ , designada per $\text{loss}(\Gamma)$. Podem ampliar aquesta definició per a cobrir A -autòmats infinits de la manera següent:

DEFINICIÓ 64. *Sigui Γ un A -autòmat involutiu i connex. Definim la *pèrdua* de Γ , designada per $\text{loss}(\Gamma)$, com el suprem (que pot ser ∞ , si Γ és infinit) del nombre de plegaments tancats, pres sobre el conjunt de totes les seqüències (finites) de plegaments de Stallings començant a Γ .*

Observeu que si Γ és finit les tals seqüències de llargada maximal són, precisament, les que acaben en un Γ_p determinista; i com que $\text{core}(\Gamma_p) = \text{St}(\langle \Gamma \rangle)$, totes elles tenen el mateix nombre de plegaments tancats, per la proposició 63. Per tant, en el cas finit, el suprem a la definició de $\text{loss}(\Gamma)$ és un màxim, i es pren sobre totes les seqüències de plegaments de Stallings començant a Γ .

Acabem la secció amb un exemple per a il·lustrar que el canvi de la base de referència pot canviar dràsticament l'aspecte de l'autòmat de Stallings. De fet, entendre com es comporta la funció $\text{St}(H, A) \mapsto \text{St}(H, B)$ (on A, B són bases de $\mathbb{F}$) és un dels grans reptes encara per aclarir de la teoria de Stallings.

EXEMPLE 65 (EXEMPLE 2.2 A [34]). *Sigui $\mathbb{F}_3$ un grup lliure amb base $A = \{a, b, c\}$ i sigui $H = \langle ab, acba \rangle$. És fàcil veure que $B = \{a', b', c'\}$ és també una base de $\mathbb{F}_3$, on $a' = a$, $b' = ab$, i $c' = acba$. Els autòmats de Stallings de H respecte de A i B estan representats a la figura següent.*



4 Aplicacions

En aquesta secció presentem algunes de les aplicacions més rellevants de la teoria dels autòmats de Stallings, desenvolupada a la secció 3. Començarem amb alguns teoremes clàssics sobre els subgrups del grup lliure que emergeixen de forma natural —de vegades com a simples reinterpretacions— dels resultats de la teoria de Stallings, i anirem progressant cap a aplicacions més elaborades, com ara el problema de la intersecció, font d'una de les qüestions més famoses del darrer segle en teoria geomètrica de grups.

4.1 Estructura dels subgrups

Tal com hem vist a la secció 3, una de les primeres aplicacions del teorema de Stallings (teorema 60) és el teorema de Nielsen-Schreier (teorema 56). Fixeu-vos que la bijecció (15) ens permet no només garantir que tots els subgrups d'un grup lliure són, de nou, lliures, sinó també descriure exactament (les classes d'isomorfia de) els subgrups d'un grup lliure donat: com que d'autòmats de Stallings sobre un alfabet de dues lletres n'hi ha de qualsevol rang fins a numerable (inclòs), la descripció següent se segueix immediatament.

COROLLARI 66. *Sigui $\mathbb{F}_\kappa$ el grup lliure de rang $\kappa \in [2, \aleph_0]$. Aleshores, per a tot cardinal $\mu \in [0, \aleph_0]$ existeix un subgrup $H \leq \mathbb{F}_\kappa$ tal que $H \simeq \mathbb{F}_\mu$. $\square$*

En particular, subgrups de qualsevol rang finit i, fins i tot, de rang infinit numerable poden aparèixer com a subgrups de $\mathbb{F}_2$; a continuació presentem dos exemples clàssics d'aquest fet. Aquest és un comportament radicalment diferent del de l'àlgebra lineal (un K -espai vectorial de dimensió n només conté subespais de dimensions $0, 1, \dots, n$), que confereix al reticle de subgrups d'un grup lliure un cert caràcter fractal.

EXEMPLE 67. Sigui $\mathbb{F}_{\{a,b\}} = \langle a, b \mid - \rangle$ un grup lliure de rang 2. Aleshores, la clausura normal de b , designada per $\langle\langle b \rangle\rangle$, és un subgrup de $\mathbb{F}_{\{a,b\}}$ de rang infinit amb autòmat de Stallings (de rang infinit) $\text{St}(\langle\langle b \rangle\rangle)$ representat a la figura 14. Prenent com a arbre generador el marcat amb traç gruixut (de fet, l'únic possible) obtenim la base $\{a^k b a^{-k} : k \in \mathbb{Z}\}$.

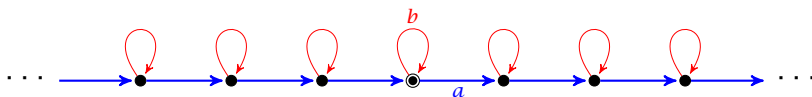


FIGURA 14: L'autòmat de Stallings de $\langle\langle b \rangle\rangle \leq \mathbb{F}_{\{a,b\}}$.

EXEMPLE 68. Un altre exemple típic és el commutador $H = [\mathbb{F}_{\{a,b\}}, \mathbb{F}_{\{a,b\}}] \trianglelefteq \mathbb{F}_{\{a,b\}}$, i.e., el subgrup normal de $\mathbb{F}_{\{a,b\}}$ generat per tots els elements de la forma $[v, w] = v^{-1}w^{-1}vw$, amb $v, w \in \mathbb{F}_{\{a,b\}}$. El subgrup H té rang infinit ja que, com tots els subgrups normals, el seu autòmat de Stallings $\text{St}(H) = \text{Sch}(H)$ és el dígraf de Cayley del grup quocient $\mathbb{F}_{\{a,b\}}/H \simeq \mathbb{Z}^2$, respecte dels generadors $\{[a], [b]\}$, i.e., respecte de la base canònica de $\mathbb{Z}^2$: la graella bidimensional

infinita, amb els arcs horitzontals etiquetats per a , i els arcs verticals etiquetats per b ; vegeu la figura 15. Proposem al lector descriure una base del commutador H a partir del vostre arbre generador preferit.

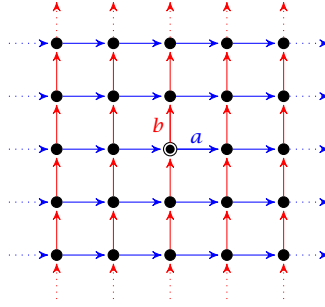


FIGURA 15: L'autòmat de Stallings del commutador $[F_{\{a,b\}}, F_{\{a,b\}}]$.

4.2 El problema de la pertinença

En aquesta secció estudiarem el problema de la pertinença en grups lliures i hi donarem una solució general. A continuació l'enunciem en general per a un grup finitament presentat $G = \langle A \mid R \rangle$.

PROBLEMA DE LA PERTINENÇA A $G = \langle A \mid R \rangle$, $\text{MP}(G)$. *Donada una família finita $u, v_1, \dots, v_k$ de paraules en els generadors de G , decidir si (l'element $g \in G$ representat per) u pertany al subgrup $\langle v_1, \dots, v_k \rangle_G \leq G$; i, en cas afirmatiu, trobar una expressió per a u com a producte dels $v_i^{\pm}$.*

Situem-nos al grup lliure $\mathbb{F}_A$; els elements $u, v_1, \dots, v_k$ són paraules (que podem suposar reduïdes, ja que la reducció és trivialment algorísmica) en l'alfabet A , i hem de decidir si $u \in H = \langle v_1, \dots, v_k \rangle$.

A tal efecte, primer calculem l'autòmat de Stallings $\text{St}(H)$ tal com hem vist a la secció anterior: construïm l'autòmat flor, $\text{Fl}(S)$, de pètals $S = \{v_1, \dots, v_k\}$, i anem fent plegaments successius (en qualsevol ordre) fins a obtenir $\text{St}(H)$,

$$\text{Fl}(S) = \Gamma_0 \curvearrowright \Gamma_1 \curvearrowright \cdots \curvearrowright \Gamma_{p-1} \curvearrowright \Gamma_p = \text{St}(H). \quad (17)$$

Per les propietats (i) i (ii) a la proposició 51, sabem que $\text{St}(H)$ és un autòmat determinista que reconeix el subgrup H . Per tant, $u \in \langle H \rangle$ si i només si podem llegir la paraula reduïda u com a etiqueta d'algun $\odot$ -camí a $\text{St}(H)$. Però, gràcies al determinisme de $\text{St}(H)$, això és molt fàcil de comprovar: començant a $\odot$, anem resseguint l'únic possible camí etiquetat amb les lletres consecutives de $u = a_{i_1}^{e_1} a_{i_2}^{e_2} \cdots a_{i_l}^{e_l}$; si en un determinat punt no podem llegir la lletra següent, deduïm que $u \notin H$; si podem completar la lectura de u amb un camí que no acaba a $\odot$, deduïm igualment que $u \notin H$; i finalment, si completem la lectura de u mitjançant un camí tancat que retorna a $\odot$ (i. e., un $\odot$ -camí), concloem que $u \in H$. Donat que aquest procediment cobreix tots els desenllaços possibles, això ens permet decidir algorísmicament si $u \in H$.

La segona part de l'algorisme consisteix a suposar que la resposta és afirmativa (i. e., que $u \in H$ i tenim aquesta paraula realitzada com l'etiqueta d'un $\bullet$ -camí reduït γ a l'autòmat $\text{St}(H)$), i buscar una manera efectiva d'expressar u en termes dels generadors originals $v_1, \dots, v_k$. De passada, entendrem quan aquesta expressió és única i quan no. La idea és la següent: anomenem $\gamma_p = \gamma$ i l'anem *elevant enrere*, plegament a plegament, al llarg de la seqüència de plegaments usada per a calcular $\text{St}(H)$, fins a obtenir un $\bullet$ -camí reduït γ_0 a l'autòmat flor $\text{Fl}(S) = \Gamma_0$:

$$\begin{aligned} \text{Fl}(S) = \Gamma_0 &\curvearrowright \Gamma_1 \curvearrowright \dots \curvearrowright \Gamma_{p-1} \curvearrowright \Gamma_p = \text{St}(H) \\ \gamma_0 &\leftarrow \gamma_1 \leftarrow \dots \leftarrow \gamma_{p-1} \leftarrow \gamma_p = \gamma. \end{aligned} \quad (18)$$

Per una banda, gràcies a la forma de l'autòmat $\text{Fl}(S)$, aquest últim camí γ_0 no és res més que una successió de pètals o pètals inversos, és a dir, un camí reduït amb etiqueta no necessàriament reduïda a $(A^\pm)^*$, però igual a una paraula reduïda en $v_1^{\pm 1}, \dots, v_k^{\pm 1}$. Per altra banda, farem l'elevació $\gamma_{i+1} \mapsto \gamma_i$ a través de cada plegament elemental de manera que el resultat sigui un camí reduït amb etiqueta $\ell(\gamma_i) = u_i$ no necessàriament reduïda, però tal que $\overline{u_i} = \overline{u}$; és a dir, representant el mateix element que u a $\mathbb{F}_A$. Per tant, el procés d'elevació de γ produirà una seqüència de paraules $u = u_p, u_{p-1}, \dots, u_1, u_0 \in (A^\pm)^*$, totes amb etiqueta reduïda igual a $\overline{u} \in \mathbb{F}_A$, introduint una o diverses cancel·lacions a cada pas, de manera que u_0 sigui l'etiqueta d'un $\bullet$ -camí reduït de $\text{Fl}(S)$, és a dir, un cert producte dels $v_i^{\pm 1}$, que és el que estem buscant.

Fixem-nos, doncs, en un dels plegaments elementals $\Gamma_i \curvearrowright \Gamma_{i+1}$ de la seqüència (17), considerem un $\bullet$ -camí reduït γ_{i+1} a Γ_{i+1} (amb $u_{i+1} = \ell(\gamma_{i+1}) \in (A^\pm)^*$), i estudiem com fer-ne l'elevació a Γ_i . Permutant i invertint lletres de A si cal, podem suposar que el plegament elemental en qüestió consisteix en la identificació de dos arcs diferents, diguem-ne e_1 i e_2 , tals que $\iota e_1 = \iota e_2$ i $\ell(e_1) = \ell(e_2) = a$; vegeu la figura 12.

Observeu que els autòmats Γ_i i Γ_{i+1} són idèntics excepte en els dos arcs e_1 i e_2 (que són diferents a Γ_i i estan identificats a Γ_{i+1}), i excepte potser en els vèrtexs $q_1 = \tau e_1$ i $q_2 = \tau e_2$ (que són iguals a Γ_{i+1} , i potser diferents de Γ_i). Si $q_1 \neq_{\Gamma_i} q_2$ el plegament és obert, i en cas contrari tancat. Això és independent del fet que el vèrtex $p = \iota e_1 = \iota e_2$ coincideixi o no amb q_1 i/o amb q_2 (és a dir, que e_1 i/o e_2 siguin o no llaços); aquesta possibilitat complica la visualització del procediment que farem, però no afecta els arguments. Distingim dos casos.

- *Cas 1:* $\Gamma_i \curvearrowright \Gamma_{i+1}$ és obert (i. e., $q_1 \neq_{\Gamma_i} q_2$). Sigui $k \geq 0$ el nombre de visites del camí reduït γ_{i+1} al vèrtex identificat $q_1 = q_2$; cadascuna d'aquestes visites s'anomena *crítica* si γ_{i+1} arriba a $q_1 = q_2$ per un arc incident a q_1 en Γ_i i en surt per un d'incident a q_2 (o viceversa). És evident que cada segment γ de γ_{i+1} entre visites crítiques consecutives a $q_1 = q_2$ s'eleva al camí idèntic γ de Γ_i (entenent que cada visita a l'arc plegat $e_1 = e_2$ cal substituir-la per e_1 o e_2 segons quins siguin els arcs adjacents a γ_{i+1} ; observeu que només hi ha una possibilitat). Això ens dona una elevació de γ_{i+1} a un « $\bullet$ -camí» de Γ_i amb la mateixa etiqueta u_{i+1} , però amb

una discontinuïtat (saltant de q_1 a q_2 , o viceversa) per cada visita crítica de y_{i+1} al vèrtex $q_1 = q_2$. Aquest problema el podem resoldre fàcilment intercalant el segment $e_1^{-1}e_2$ (ó $e_2^{-1}e_1$) per cadascuna de les visites crítiques, i obtenim així el $\bullet$ -camí reduït y_i a Γ_i . A més, com que l'etiqueta d'aquests segments afegits és sempre $a^{-1}a$, l'etiqueta $u_i = \ell(y_i)$ serà $\ell(y_{i+1}) = u_{i+1}$ amb k' segments $a^{-1}a$ intercalats en les posicions de les $k' \leq k$ visites crítiques a $q_1 = q_2$; en particular, $\overline{u_i} = \overline{u_{i+1}} \in \mathbb{F}_A$.

Cal esmentar una possibilitat degenerada que pot ocórrer quan $\bullet = q_1 = q_2$ a Γ_{i+1} (i, posem per cas, $\bullet = q_1$ a Γ_i). En aquesta situació, el $\bullet$ -camí y_{i+1} pot començar (resp., acabar) amb un arc e incident a q_2 : en aquests casos, cal afegir el segment $e_1^{-1}e_2$ (resp., $e_2^{-1}e_1$) al començament (resp., final) de y_i per tal que l'elevació realment comenci (resp., acabi) a $\bullet \in \Gamma_i$.

- *Cas 2:* $\Gamma_i \curvearrowright \Gamma_{i+1}$ és tancat (i. e., $q_1 =_{\Gamma_i} q_2$). La noció de visita crítica al vèrtex $q_1 = q_2$ no té sentit ja que ara $q_1 = q_2$ també a Γ_i , i l'elevació de y_{i+1} es pot fer d'una tirada (com si tot y_{i+1} fos un sol segment entre visites crítiques): només cal considerar el mateix $\bullet$ -camí a Γ_i utilitzant només un qualsevol dels arcs plegats. És més, hi ha infinites possibles elevacions de y_{i+1} (a $\bullet$ -camins reduïts de Γ_i llegint $\bar{\ell}(y_{i+1})$): per cada factorització $y_{i+1} = y'ry''$, i per cada camí reduït y''' de r a $p = te_1 = te_2$ (resp., a $q = \tau e_1 = \tau e_2$), podem elevar y_{i+1} com a $y'ry'''p(e_1qe_2^{-1})^k p(y''')^{-1}ry''$ (resp., $y'ry'''q(e_1^{-1}pe_2)^k q(y''')^{-1}ry''$), on $k \in \mathbb{Z} \setminus \{0\}$. Es pot veure que aquesta és la família de totes les elevacions possibles de y_{i+1} a $\bullet$ -camins reduïts de Γ_i llegint $\overline{u_{i+1}} \in \mathbb{F}_A$.

Tot aquest procediment demostra el resultat següent.

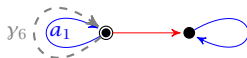
TEOREMA 69. *El problema de la pertinença per a grups lliures, $MP(\mathbb{F}_A)$, és resoluble.* $\square$

Podeu veure il·lustrat aquest mètode gràfic amb la resolució de l'exemple següent. Proposem al lector el seguiment d'aquest mateix mètode per deduir la paraula $v_1 v_2^{-1} v_1 (v_1 v_2^{-1})^7 v_3^{-1} v_2^{-1} v_3$, que a l'exemple 26 ens ha permès d'expressar u en termes de v_1, v_2, v_3 (sense explicar-ne l'origen).

EXEMPLE 70. Sigui $\mathbb{F}_2$ el grup lliure sobre $A = \{a, b\}$ i considerem el subgrup $H = \langle v_1, v_2, v_3 \rangle \leq \mathbb{F}_2$ generat pels elements $v_1 = a^3$, $v_2 = abab^{-1}$, i $v_3 = a^{-1}bab^{-1}$. Decidiu si l'element $u = a \in \mathbb{F}_2$ pertany a H i, en cas afirmatiu, expresseu-lo en termes de v_1, v_2, v_3 .

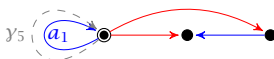
Per començar, construïm l'autòmat $St(H)$, a partir del conjunt de generadors $\{v_1, v_2, v_3\}$; això està fet a l'exemple 61; vegeu la figura 13. Observeu que el primer pas $\mathcal{F} \curvearrowright \Gamma_1$ són en realitat tres plegaments elementals simultanis (no ens caldrà distinguir-los entre ells). Noteu també que tots els plegaments d'aquesta seqüència són oberts excepte un, $\Gamma_4 \curvearrowright \Gamma_5$, que és tancat perquè identifica dos a -llaços en un de sol. D'aquest procés deduïm que $H = \langle v_1, v_2, v_3 \rangle = \langle a, bab^{-1} \rangle$; o, millor encara, que $\{a, bab^{-1}\}$ és una base de H (corresponent a l'arbre maximal $\bullet \rightarrow \bullet$ de Γ_6).

D'aquí veiem clarament que $a \in H$, ja que és l'etiqueta del $\odot$ -camí $\gamma_6 = a_1$ de $\Gamma_6 = \text{St}(H)$, assenyalat al dibuix amb una línia discontinua de color gris:

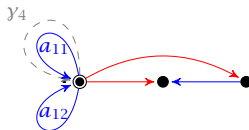


Al a -llaç incident a $\odot$ (i. e., l'únic arc travessat per γ_6) l'hem anomenat a_1 ; ens referirem als diversos arcs dels autòmats Γ_i de la seqüència de plegaments amb la lletra de les seves respectives etiquetes (a per a les blaves, i b per a les vermelles) i uns certs subíndexs que anirem definint dinàmicament; als arcs no usats no els assignarem cap nom.

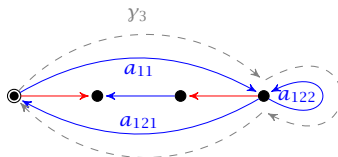
Com a primer pas, elevem el camí γ_6 a través de l'últim plegament $\Gamma_5 \curvearrowright \Gamma_6$: com que γ_6 no visita el vèrtex identificat, s'eleva al $\odot$ -camí $\gamma_5 = a_1$ de Γ_5 (on, amb un petit abús de llenguatge, tornem a dir a_1 a l'arc corresponent de Γ_5):



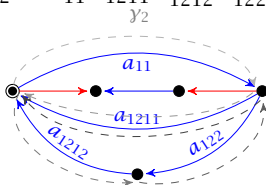
Observeu ara el plegament $\Gamma_4 \curvearrowright \Gamma_5$, que és tancat; per mantenir una notació coherent, anomenem a_{11} i a_{12} els dos a -llaços de Γ_4 que es pleguen en a_1 . Per elevar γ_5 a l'autòmat Γ_4 , la manera més senzilla és usant un dels dos a -llaços plegats, per exemple, $\gamma_4 = a_{11}$:



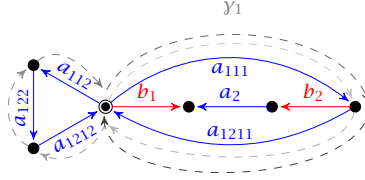
(Hi ha, però, altres maneres de fer-ho, $\gamma_4 = a_{12}$, $\gamma_4 = a_{11}a_{12}a_{11}^{-1}$, $\gamma_4 = a_{11}a_{12}^{-1}a_{11}$, etc). El pas següent és l'elevació a Γ_3 : designem per a_{121} i a_{122} els dos a -arcs de Γ_3 que es pleguen sobre a_{12} , i elevem el camí γ_4 a $\gamma_3 = a_{11}a_{122}^{-1}a_{121}$:



Anomenant a_{1211} i a_{1212} els dos arcs de Γ_2 que es pleguen sobre a_{121} , el camí γ_3 s'eleva a Γ_2 com a $\gamma_2 = a_{11}a_{1211}a_{1212}^{-1}a_{122}^{-1}a_{1211}$:

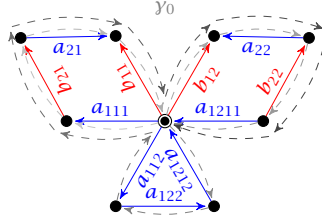


El pas següent és elevar γ_2 a Γ_1 com a $\gamma_1 = a_{111}a_{1211}a_{1212}^{-1}a_{122}^{-1}a_{112}^{-1}a_{111}a_{1211}$:



Finalment, si designem per a_2 , b_1 , i b_2 els arcs de Γ_1 encara no usats, i mantenim el conveni de notació per als arcs corresponents de $\Gamma_0 = \mathcal{F}$, podem elevar γ_1 a

$$\gamma_0 = a_{111}b_{21}a_{21}b_{11}^{-1}b_{12}a_{22}^{-1}b_{22}^{-1}a_{1211}a_{1212}^{-1}a_{122}^{-1}a_{112}^{-1}a_{111}b_{21} \cdot a_{21}b_{11}^{-1}b_{12}a_{22}^{-1}b_{22}^{-1}a_{1211}.$$



Assenyalant amb parèntesis les visites completes a cada pètal o pètal invers (marcades per les visites de γ_0 a $\bullet$) obtenim l'expressió desitjada de $u = a$ en termes de v_1 , v_2 , v_3 :

$$\begin{aligned} a &= (abab^{-1})(ba^{-1}b^{-1}a)(a^{-1}a^{-1}a^{-1})(abab^{-1})(ba^{-1}b^{-1}a) = \\ &= v_2v_3^{-1}v_1^{-1}v_2v_3^{-1}. \end{aligned}$$

Com hem comentat més amunt, les elevacions a través de plegaments oberts són úniques, però l'elevació de γ_5 a través del plegament tancat $\Gamma_4 \leadsto \Gamma_5$ té diverses possibilitats (que acabaran en resultats diferents, *i. e.*, en expressions alternatives de $u = a$ en termes de $v_1^{\pm 1}$, $v_2^{\pm 1}$, $v_3^{\pm 1}$). Convidem el lector a prendre $\gamma_4 = a_{12}$ en lloc de $\gamma_4 = a_{11}$, i a calcular la seva elevació fins a Γ_0 per obtenir la nova expressió $a = (a^{-1}bab^{-1})(ba^{-1}b^{-1}a^{-1})(aaa) = v_3v_2^{-1}v_1$.

El fet de poder obtenir a en termes de $v_1^{\pm 1}$, $v_2^{\pm 1}$, $v_3^{\pm 1}$ de *dues maneres diferents*, $v_2v_3^{-1}v_1^{-1}v_2v_3^{-1} = a = v_3v_2^{-1}v_1$, ens confirma el fet que $\{v_1, v_2, v_3\}$ generen H , però *no són un conjunt lliure* (vegeu la definició 2). Dit d'una altra manera, $v_2v_3^{-1}v_1^{-1}v_2v_3^{-1}v_1^{-1}v_2v_3^{-1} = 1$ és una relació no trivial entre els tres generadors v_1 , v_2 , v_3 de H .

4.3 Generadors, bases i rang

Comencem recordant que la proposició 54 garanteix la computabilitat d'una base per qualsevol subgrup de $\mathbb{F}_A$ a partir d'un conjunt *finit* de generadors.

COROLLARI 71. *Si S és un conjunt finit d'elements d'un grup lliure $\mathbb{F}_A$, aleshores una base (i, per tant, el rang) de $\langle S \rangle \leq \mathbb{F}_A$ és computable. Més concretament, $\text{rk}(\langle S \rangle) = 1 - \#\Gamma + \#E^+\Gamma$, on $\Gamma = \text{St}(\langle S \rangle)$.*

Molts altres resultats fonamentals sobre bases de (subgrups de) grups lliures es deriven fàcilment de la construcció de Stallings. L'observació següent ens portarà a la important propietat de hopfianitat dels grups lliures.

Sigui Γ un A -autòmat involutiu, finit, i connex, i sigui $\Gamma \xrightarrow{\phi} \Gamma'$ un plegament elemental. El corresponent morfisme d'autòmats $\phi: \Gamma \rightarrow \Gamma'$ indueix un morfisme exhaustiu de grups $\phi: \pi_\bullet(\Gamma) \twoheadrightarrow \pi_\bullet(\Gamma')$, $[\gamma] \mapsto [\gamma\phi]$ tal que $\phi\tilde{\ell}_{\Gamma'} = \tilde{\ell}_\Gamma$. A més, és fàcil veure que, si el plegament és obert, llavors ϕ és bijectiu, mentre que si és tancat (posem per cas, amb $e_1, e_2 \in E\Gamma$ els dos arcs que s'identifiquen a Γ' , $p = \iota e_1 = \iota e_2$, $q = \tau e_1 = \tau e_2$, i $\ell(e_1) = \ell(e_2) \in A^\pm$), llavors $\ker \phi$ està generat, com a subgrup normal de $\pi_\bullet(\Gamma)$, per l'element $[\eta p e_1 q e_2^{-1} p \eta^{-1}]$, on η és un camí qualsevol a Γ de $\bullet$ a p .

L'enunciat següent mostra com la teoria desenvolupada captura geomètricament els dos ingredients (algebraics) de la definició elemental de base (definició 2).

PROPOSICIÓ 72. *Sigui $S \subseteq \mathbb{F}_A$. Aleshores,*

- (i) S genera $\mathbb{F}_A$ si i només si $\text{St}(\langle S \rangle) = \mathcal{B}_A$;
- (ii) S és una família lliure de $\mathbb{F}_A$ si i només si $\text{loss}(\text{Fl}(S)) = 0$; i, en aquest cas, $\text{rk}(\langle S \rangle) = \#S$.

Ambdues condicions són algorísmicament decidibles si S és finit.

PROVA. L'apartat (i) és clar ja que $\text{St}(\mathbb{F}_A) = \mathcal{B}_A$.

Per veure l'apartat (ii) distingim dos casos. Si S és finit, considerem l'autòmat flor $\text{Fl}(S)$, i una seqüència qualsevol de plegaments elementals fins a obtenir $\text{St}(\langle S \rangle)$,

$$\text{Fl}(S) = \Gamma_0 \xrightarrow{\phi_1} \Gamma_1 \xrightarrow{\phi_2} \dots \xrightarrow{\phi_{p-1}} \Gamma_{p-1} \xrightarrow{\phi_p} \Gamma_p = \text{St}(\langle S \rangle).$$

Component els corresponents morfismes obtenim el diagrama commutatiu

$$\begin{array}{ccccccc} \pi_\bullet(\Gamma_0) & \xrightarrow{\phi_1} & \pi_\bullet(\Gamma_1) & \xrightarrow{\phi_2} & \dots & \xrightarrow{\phi_{p-1}} & \pi_\bullet(\Gamma_{p-1}) & \xrightarrow{\phi_p} & \pi_\bullet(\Gamma_p) \\ & \searrow & \downarrow \tilde{\ell}_1 & & & \downarrow \tilde{\ell}_{p-1} & & \downarrow \tilde{\ell}_p & \searrow \\ & & & \tilde{\ell}_0 & & & & & \langle S \rangle \end{array} \quad (19)$$

on $\text{rk}(\pi_\bullet(\Gamma_0)) = \#S$, amb (les classes de) els pètals com a base; i on $\tilde{\ell}_p$ és bijectiu ja que Γ_p és determinista. Si $\text{loss}(\text{Fl}(S)) = 0$, tots els plegaments són oberts, cadascun dels morfismes $\phi_1, \dots, \phi_p$ és bijectiu i concloem que $\tilde{\ell}_0 = \phi_1 \dots \phi_p \tilde{\ell}_p$ és també bijectiu; per tant, $\langle S \rangle$ està lliurement generat per

les etiquetes dels pètals de $\text{Fl}(S)$, és a dir, per S . Recíprocament, si S és una família lliure d'elements de $\mathbb{F}_A$, llavors S és base de $\langle S \rangle \leq \mathbb{F}_A$, i el morfisme $\tilde{\ell}_0$ és bijectiu; la commutativitat del diagrama anterior implica aleshores que $\phi_1, \dots, \phi_p$ són tots bijectius i, per tant, tots els plegaments de la seqüència són oberts; en conseqüència, $\text{loss}(\text{Fl}(S)) = 0$.

Finalment, suposem que S és infinit. Si S no és una família lliure, llavors existeix un subconjunt finit $S_0 \subset S$ que tampoc no és lliure i, aplicant el cas finit, deduïm que $\text{loss}(\text{Fl}(S)) \geq \text{loss}(\text{Fl}(S_0)) \geq 1$. Recíprocament, si S és una família lliure, tota subfamília finita $S_0 \subseteq S$ també és lliure i, aplicant el cas finit, $\text{loss}(\text{Fl}(S_0)) = 0$; per tant, cap seqüència finita de plegaments elementals començant a $\text{Fl}(S)$ (que involucrarà forçosament un nombre finit de pètals) no pot contenir un plegament tancat; és a dir, $\text{loss}(\text{Fl}(S)) = 0$. $\square$

Noteu que d'aquest resultat se'n segueix una prova alternativa per a la proposició 17: si B és una base de $\mathbb{F}_A$, aleshores $\text{loss}(\text{Fl}(B)) = 0$ i $\text{St}(\langle B \rangle, A) = \mathcal{B}_A$; per tant, $\#B = \text{rk}(\text{Fl}(B)) = \text{rk}(\text{St}(\langle B \rangle, A)) = \text{rk}(\mathcal{B}_A) = \#A$. També en podem deduir la important propietat de hopfianitat dels grups lliures finitament generats:

TEOREMA 73. *El grup lliure $\mathbb{F}_n$ és hopfià: tot endomorfisme $\varphi: \mathbb{F}_n \rightarrow \mathbb{F}_n$ exhaustiu és automàticament injectiu. (Equivalentment, $S \subseteq \mathbb{F}_n$ és base de $\mathbb{F}_n$ si i només si $\langle S \rangle = \mathbb{F}_n$ i $\#S = n$.)*

PROVA. De nou, la demostració és immediata de la teoria de Stallings: si $\langle S \rangle = \mathbb{F}_n$ i $\#S = n$, aleshores $\text{St}(\langle S \rangle) = \mathcal{B}_n$, i $\text{rk}(\text{Fl}(S)) = n$. Per tant,

$$\text{loss}(\text{Fl}(S)) = \text{rk}(\text{Fl}(S)) - \text{rk}(\text{St}(\langle S \rangle)) = n - n = 0 \quad (20)$$

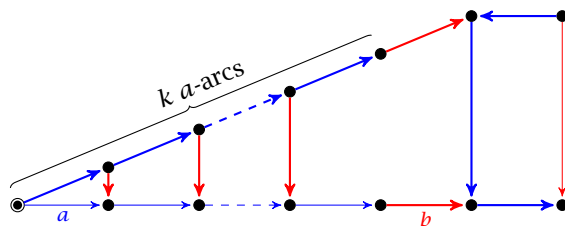
i, per la proposició 72(ii), tenim que S és una família lliure; per tant, S és base de $\mathbb{F}_n$. $\square$

D'acord amb l'observació 6, i tal com succeeix als K -espais vectorials, les bases de grups lliures $\mathbb{F}$ són sempre conjunts minimalis de generadors. El recíproc —també cert als K -espais vectorials— és fals als grups lliures, fins a la màxima degeneració possible.

LEMA 74. *Existeixen conjunts de generadors minimalis de $\mathbb{F}_n$, $n \geq 1$, de qualsevol cardinalitat finita $m \geq n$.*

PROVA. Aquest comportament ja es pot observar a $\mathbb{Z} = \mathbb{F}_1$ per raons purament aritmètiques: considerem un conjunt de números primers $\{p_1, \dots, p_m\}$ diferents dos a dos, i prenem $S = \{\prod_{i \neq j} p_i : j = 1, \dots, m\}$; clarament $\text{mcd}(S) = 1$ i, per tant, $\langle S \rangle = \mathbb{Z}$. Però $\text{mcd}(S \setminus \{\prod_{i \neq j} p_i\}) = p_j$ i, per tant, $\langle S \setminus \{\prod_{i \neq j} p_i\} \rangle = p_j \mathbb{Z} \neq \mathbb{Z}$.

Usant la teoria de Stallings podem construir exemples més sofisticats. Per exemple, al grup $\mathbb{F}_{\{a,b\}}$ considereu la família d'autòmats involutius Γ_k , $k \geq 2$, representada a continuació:



Prenent com a arbre generador el donat pels arcs representats en traç gruixut obtenim el conjunt de generadors:

$$S_k = \{aba^{-1}, a^k bab^{-1}a^{-1}b^{-1}a^{k-1}, a^k ba^{-1}ba^{-2}b^{-1}a^{-k}\} \cup \{a^i ba^{-1}b^{-1}a^{-(i-1)} : i \in [2, k-1]\}$$

(de cardinal $k+1$) per a $\langle \Gamma_k \rangle$. Observem, però, que Γ_k no és determinista (ho seria sense un dels dos a -arcs sortint de $\bullet$) i que, fent-li plegaments de Stallings, col·lapsa fins a la rosa $\mathcal{B}_{\{a,b\}}$. Per tant, $\langle S_k \rangle = \langle \Gamma_k \rangle = \mathbb{F}_{\{a,b\}}$ i, per a tot $k \geq 2$, S_k és una família de $k+1$ generadors de $\mathbb{F}_{\{a,b\}}$. Ara bé, si eliminem un qualsevol dels generadors de S_k (equivalentment, si eliminem de Γ_k l'arc corresponent fora de l'arbre generador), el procés de plegaments queda bloquejat justament a l'arc eliminat i s'obté un A -autòmat determinista diferent de $\mathcal{B}_{\{a,b\}}$. Per tant, S_k és un sistema de generadors *minimal* de $\mathbb{F}_{\{a,b\}}$, de cardinal $k+1$. $\square$

Fixeu-vos, finalment, que la família S_k també serveix com a contraexemple al grup lliure de la propietat fonamental en àlgebra lineal que tota família de generadors conté una base.

4.4 Conjugació i normalitat

Un dels conceptes fonamentals en teoria de grups és el de *conjugació*: si G és un grup, es diu que dos elements $g, g' \in G$ (resp., subgrups $H, H' \leq G$) són conjugats, i escrivim $g \sim g'$ (resp., $H \sim H'$) si existeix un element $z \in G$ tal que $z^{-1}gz = g'$ (resp., $z^{-1}Hz = H'$). Aleshores diem que g' (resp., H') és g (resp., H) *conjugat per* z , i escrivim $g' = g^z = z^{-1}gz$ (resp., $H' = H^z = z^{-1}Hz$).

La teoria de Stallings permet caracteritzar la conjugació de subgrups de forma molt transparent.

LEMA 75. *Siguin H un subgrup de $\mathbb{F}_A$, i $w \in \mathbb{F}_A$. Aleshores, $\text{St}(H^w, A)$ és el cor de l'autòmat $\text{Sch}_{H^w}(H, A)$ obtingut canviant el punt base de $\text{Sch}(H, A)$ per Hw .*

PROVA. És clar de la proposició 47 i la definició 34 que $\text{Sch}_{H^w}(H, A)$ és un A -autòmat involutiu, determinista, saturat i connex que reconeix el subgrup H^w . El resultat se segueix immediatament del corollari 52(ii). $\square$

Com a conseqüència del lema anterior obtenim la següent caracterització gràfica de la conjugació de subgrups al grup lliure, que és òbviament decidible en cas que els subgrups siguin finitament generats.

PROPOSICIÓ 76. *Siguin H i K dos subgrups de $\mathbb{F}_A$. Aleshores, H i K són conjugats si i només si $\text{St}^*(H) = \text{St}^*(K)$.*

PROVA. Observeu que, per construcció, $\text{St}(H)$ coincideix amb $\text{St}^*(H)$ reintroduint el pèl (eventualment buit) fins a $\bullet_H$.

Suposem que $\text{St}^*(H) = \text{St}^*(K)$. Reintroduïm a $\text{St}^*(H) = \text{St}^*(K)$ els dos pèls fins als vèrtexs $\bullet_H$ i $\bullet_K$ i sigui w l'etiqueta d'un camí reduït qualsevol de $\bullet_H$ a $\bullet_K$. És fàcil veure que $H^w = K$; per tant, H i K són conjugats.

Recíprocament, suposem que $H^w = K$ per alguna paraula (reduïda) $w \in \mathbb{F}_A$. Siguí γ l'únic camí possible a $\text{St}(H)$ començant a $\bullet_H$ i llegint w : la unicitat és clara pel determinisme de $\text{St}(H)$; i si no n'existeix cap, afegim a $\text{St}(H)$ el pèl necessari per a poder completar γ . Declarem $\tau\gamma$ com a nou punt base, i designem per Γ l'autòmat obtingut. Per construcció, $\text{core}^*(\Gamma) = \text{core}^*(\text{St}(H)) = \text{St}^*(H)$ és determinista, i $\langle \Gamma \rangle = w^{-1} \langle \text{St}(H) \rangle w = H^w = K$. Per tant, $\text{core}^*(\Gamma) = \text{core}^*(\text{St}(K)) = \text{St}^*(K)$, d'on deduïm $\text{St}^*(H) = \text{St}^*(K)$. $\square$

Observeu que, en cas que H i K siguin finitament generats i conjugats, la demostració anterior proporciona, a més, un mètode per a calcular efectivament un conjugador w . Això demostra que

COROLLARI 77. *El problema de la conjugació de subgrups $\text{SCP}(\mathbb{F})$ és decidable per a grups lliures: existeix un algorisme que, donats dos subconjunts finits $S_1, S_2 \subseteq \mathbb{F}$, decideix si els corresponents subgrups generats són conjugats i, en cas afirmatiu, retorna un element conjugador.* $\square$

Recordem que un subgrup H d'un grup G es diu *normal* (a G), designat $H \trianglelefteq G$, si i només si $H^g = H$, per a tot $g \in G$. Del lema 75 també se segueix fàcilment una caracterització gràfica de normalitat en el grup lliure (i la seva decidibilitat en el cas finitament generat).

PROPOSICIÓ 78. *Sigui $H \neq 1$ un subgrup de $\mathbb{F}_A$. Aleshores, els enunciats següents són equivalents:*

- (a) H és normal a $\mathbb{F}_A$;
- (b) $\text{Sch}(H)$ és vèrtex-transitiu;¹⁷
- (c) $\text{Sch}(H)$ és vèrtex-transitiu i cor;
- (d) $\text{St}(H)$ és vèrtex-transitiu i saturat.

PROVA. L'equivalència entre (a) i (b) se segueix clarament de les definicions a través del lema 75. Concretament, per a tot $w \in \mathbb{F}_A$:

$$H^w = H \Leftrightarrow \text{Sch}(H^w) = \text{Sch}(H) \Leftrightarrow \text{Sch}_{H^w}(H) = \text{Sch}(H).$$

La interpretació de l'esquerra correspon a la normalitat de H , mentre que la de la dreta correspon a la vèrtex-transitivitat de $\text{Sch}(H)$.

¹⁷ Un A -autòmat Γ és *vèrtex-transitiu* si per a qualsevol parell de vèrtexs, p, q de Γ , existeix un automorfisme de A -dígrafs etiquetats (és a dir, respectant les etiquetes de les arestes i ignorant el punt base) de Γ que envia p a q . (Informalment, si « Γ té el mateix aspecte vist des de qualsevol vèrtex»).

Per veure l'equivalència entre (b) i (c) és suficient veure que (b) implica (c). En efecte, si $\text{Sch}(H)$ és vèrtex-transitiu, aleshores ha de ser cor ja que, clarament, no existeix cap automorfisme de A -autòmats Γ enviant un vèrtex del cor de Γ (i, per tant, pertanyent a un $\bullet$ -camí reduït de Γ) a un vèrtex fora del cor de Γ (i, per tant, no pertanyent a cap $\bullet$ -camí reduït de Γ).

Finalment, l'equivalència entre (c) i (d) és conseqüència immediata de la proposició 51(iii). $\square$

OBSERVACIÓ 79. Tot i que per decidir si un subgrup finitament generat $H \leq \mathbb{F}_A$ és normal, n'hi ha prou de comprovar si els conjugats dels generadors per les lletres de $A^\pm$ tornen a pertànyer a H (un nombre finit d'instàncies del problema de la pertinença), la proposició anterior ens dona una caracterització gràfica molt natural per a la normalitat, que també es pot usar per a donar un algorisme alternatiu per decidir la normalitat de H .

4.5 El problema de l'índex

Recordem que, si H és un subgrup d'un grup G , aleshores G sempre admet una partició $G = \bigsqcup_{i \in I} Hg_i$, on $g_i \in G$. Els conjunts Hg_i s'anomenen *classes laterals (per la dreta)* de H a G ; el cardinal del conjunt de classes laterals $H \backslash G = \{Hg_i : i \in I\}$ s'anomena *índex* de H en G , i es designa per $|G : H|$; es diu també que $\{g_i\}_{i \in I}$ és una *família de representants* (o *transversal*) de les classes laterals (per la dreta) de H en G .

En aquesta secció usarem els autòmats de Stallings per a estudiar qüestions relatives a l'índex dels subgrups dels grups lliures, i resoldrem el problema de la finitud de l'índex, que enunciem en general a continuació.

PROBLEMA DE LA FINITUD DE L'ÍNDEX A $G = \langle A \mid R \rangle$, $\text{FIP}(G)$. *Donada una família finita $v_1, \dots, v_k$ de paraules en els generadors de G , decidir si el subgrup $\langle v_1, \dots, v_k \rangle \leq G$ és d'índex finit a G i, en cas afirmatiu, calcular aquest índex i una família de representants de les corresponents classes laterals.*

Recordem (definició 45) que els vèrtexs de l'autòmat de Schreier $\text{Sch}(H)$ són precisament les classes laterals per la dreta de H a $\mathbb{F}$, és a dir, $|\mathbb{F} : H| = \#\text{V}\text{Sch}(H)$; i observem que l'apartat (vi) de la proposició 51 pren la forma següent per a subgrups finitament generats.

COROLLARI 80. *Sigui H un subgrup finitament generat d'un grup lliure $\mathbb{F}$. Aleshores, l'índex $|\mathbb{F} : H|$ és finit si i només si $\text{St}(H)$ és saturat; en aquest cas, $|\mathbb{F} : H| = \#\text{V}\text{St}(H)$.* $\square$

Com a primera conseqüència podem deduir que $\mathbb{F}_n$ té una quantitat finita de subgrups d'un índex finit k donat (vegeu [21] per a una fórmula recurrent que compta aquest nombre de subgrups). D'aquí es dedueix fàcilment el mateix resultat per a qualsevol grup G finitament generat.

COROLLARI 81. *Sigui G un grup finitament generat. Per a tot $k \geq 1$, G conté una quantitat finita de subgrups d'índex k .*

EXEMPLE 82. A continuació representem els autòmats de Stallings dels 13 subgrups d'índex 3 de $\mathbb{F}_2 = \langle a, b \mid - \rangle$ (amb les corresponents classes de conjugació encerclades en gris). Òbviament, els subgrups normals corresponen a les classes amb un únic representant.

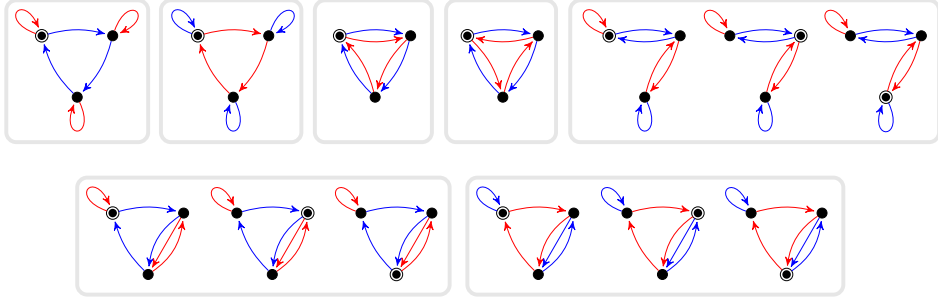


FIGURA 16: Els 13 subgrups d'índex 3 de $\mathbb{F}_2 = \langle a, b \mid - \rangle$.

Del corollari 80 se segueix immediatament el resultat següent.

TEOREMA 83. *El problema de la finitud de l'índex per a grups lliures, $\text{FIP}(\mathbb{F})$, és computable.*

PROVA. La decidibilitat de $\text{FIP}(\mathbb{F})$ és conseqüència del corollari 80 i de la computabilitat de $\text{St}(H)$ en el cas finitament generat.

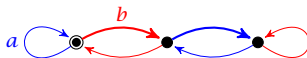
Suposem ara que $\text{St}(H)$ és saturat i finit. Per a calcular una família de representants de les classes per la dreta mòdul H , només hem d'observar que, a l'autòmat $\text{St}(H) = \text{core}(\text{Sch}(H)) = \text{Sch}(H)$, qualsevol camí γ amb inici $\bullet$ té la forma (13) i, per tant, compleix $\tau\gamma = Hw$, on $w = \ell(\gamma) \in \mathbb{F}_A$. Per tant, podem formar una família de representants de les classes laterals per la dreta mòdul H simplement prenent un camí qualsevol del punt base $\bullet$ a cadascun dels vèrtexs de $\text{St}(H)$. Una manera sistemàtica de fer-ho és calcular un arbre generador T de $\text{St}(H)$, i prendre les etiquetes de les T -geodèsiques de $\bullet$ a cada vèrtex: $\{\ell(T[\bullet, p]) \mid p \in \text{VSt}(H)\}$ és una família de representants de les classes laterals per la dreta mòdul H , és a dir,

$$H \backslash \mathbb{F}_A = \bigsqcup_{p \in \text{VSt}(H)} H \ell(T[\bullet, p]). \quad (21)$$

Òbviament, l'índex $|\mathbb{F} : H|$ és el cardinal (finit) d'aquest conjunt. Això resol completament el problema de la finitud de l'índex per a grups lliures. (Si volem representants de les classes per l'esquerra, només hem d'invertir les paraules obtingudes, ja que $(Hw)^{-1} = w^{-1}H^{-1} = w^{-1}H$.) $\square$

EXEMPLE 84. Sigui $\mathbb{F}_2$ el grup lliure sobre $A = \{a, b\}$ i considerem el subgrup $H = \langle v_1, v_2, v_3, v_4 \rangle \leq \mathbb{F}_2$ generat pels elements $v_1 = a$, $v_2 = b^2$, $v_3 = ba^2b^{-1}$,

i $v_4 = baba^{-1}b^{-1}$. Per decidir si H és d'índex finit a $\mathbb{F}_2$ calculem $\text{St}(H)$ a partir de l'autòmat flor $\text{Fl}(\{v_1, v_2, v_3, v_4\})$, tal com hem explicat al teorema 53. El resultat és l'autòmat següent:



Com que els tres vèrtexs de $\text{St}(H)$ tenen un a -arc entrant i un de sortint, i un b -arc entrant i un de sortint, l'autòmat $\text{St}(H)$ és saturat i, per tant, H és d'índex finit a $\mathbb{F}_2$, concretament d'índex $\#\text{VSt}(H) = 3$. Prenent com a arbre generador el donat pels dos arcs dibuixats amb traç gruixut, obtenim $\{1, b, ba\}$ com a conjunt de representants de les classes per la dreta mòdul H , és a dir, $\mathbb{F}_2 = H \sqcup Hb \sqcup H(ba)$.

Com veiem a continuació, la coneguda fórmula de Schreier, que relaciona l'índex d'un subgrup d'índex finit amb el seu rang, també es dedueix de forma transparent de la teoria de Stallings.

TEOREMA 85 (FÓRMULA DE SCHREIER). *Sigui $\mathbb{F}_\kappa$ el grup lliure de rang κ , i sigui $H \leq \mathbb{F}_\kappa$ un subgrup d'índex finit. Aleshores, $\text{rk}(H) - 1 = |\mathbb{F}_\kappa : H|(\kappa - 1)$. En particular, un subgrup d'índex finit és finitament generat si i només si el rang ambient κ és finit.*

PROVA. Per la proposició 51(vi), si H és d'índex finit, llavors $\text{St}(H)$ és 2κ -regular i $\#\text{VSt}(H) < \infty$. Per tant, $\text{rk}(\text{St}(H)) < \infty$ (i. e., hi ha un nombre finit d'arcs fora de qualsevol arbre generador) si i només si $\kappa < \infty$. En particular, si $\kappa = \infty$ llavors $\text{rk}(H) = \infty$ i ja hem acabat.

Suposem ara $\kappa < \infty$. Sabem que H és finitament generat i, per tant, $\text{St}(H)$ és finit. Sigui T un arbre generador de $\text{St}(H)$; tal com s'ha demostrat al final de la prova del teorema 53, H té una base de $\#(\text{E}^+\text{St}(H) \setminus \text{ET})$ elements; per tant,

$$\begin{aligned} \text{rk}(H) - 1 &= \#(\text{E}^+\text{St}(H) \setminus \text{ET}) - 1 = \#\text{E}^+\text{St}(H) - \#\text{ET} - 1 = \\ &= \#\text{E}^+\text{St}(H) - \#\text{VT} = \kappa\#\text{VSt}(H) - \#\text{VSt}(H) = \\ &= |\mathbb{F}_\kappa : H|(\kappa - 1), \end{aligned}$$

on la penúltima igualtat prové de $2\kappa\#\text{VSt}(H) = 2\#\text{E}^+\text{St}(H)$, obtingut sumant els graus de tots els vèrtexs. $\square$

Combinant les caracteritzacions gràfiques d'índex finit (corol·lari 80) i normalitat (proposició 78), arribem a una mena de recíproc del lema de Schreier per a subgrups normals no trivials del grup lliure.

COROL·LARI 86. *Un subgrup normal no trivial H de $\mathbb{F}_n$ és finitament generat si i només si té índex finit.* $\square$

COROLLARI 87. *Un subgrup normal no trivial H de $\mathbb{F}_{\aleph_0}$ té sempre rang infinit.* $\square$

Acabem aquesta secció amb una demostració gràfica particularment neta d'un resultat clàssic de Marshall Hall, Jr., que mostra especialment bé la potència de l'enfocament geomètric per obtenir resultats algebraics.

Recordem que, donats H i K subgrups d'un grup lliure $\mathbb{F}$, es diu que H és *factor lliure* de K , i ho designem per $H \leq_{\text{ff}} K$, si una base (i, per tant, totes les bases) de H es pot ampliar a una base de K . Observeu també que, si Δ és un subautòmat de Γ , aleshores $\langle \Delta \rangle$ és un factor lliure $\langle \Gamma \rangle$ (ampliant un arbre generador de Δ a un arbre generador de Γ).

Un resultat ben sabut d'àlgebra lineal diu que, donat un espai vectorial E i un subespai seu $F \leq E$, tota base de F es pot ampliar a una base de E (tècnicament, això correspon al fet que F és un sumand directe de E , noció de la qual els factors lliures són la generalització a ambients no commutatius). El resultat corresponent en un grup lliure és fals: sabem que $\mathbb{F}_n$ té subgrups H de rang superior a n (fins i tot de rang infinit) que, clarament, no en seran factors lliures, $H \not\leq_{\text{ff}} \mathbb{F}_n$. El clàssic teorema de Marshall Hall, Jr. ens diu que un resultat d'aquesta naturalesa sí que és cert, si ens restringim a subgrups finitament generats, i canviem $\mathbb{F}_n$ per un subgrup d'índex finit adequat.

TEOREMA 88 (M. HALL, JR. [22]). *Si H és un subgrup finitament generat de $\mathbb{F}_A$, aleshores H és un factor lliure d'un subgrup d'índex finit de $\mathbb{F}_A$; és a dir:*

$$H \leq_{\text{fg}} \mathbb{F} \Rightarrow \exists K : H \leq_{\text{ff}} K \leq_{\text{fi}} \mathbb{F}. \quad (22)$$

PROVA. Si H ja és d'índex finit a $\mathbb{F}_A$, el resultat és obvi. En cas contrari, sabem que $\text{St}(H)$ és un A -autòmat finit i insaturat. Ara és suficient adonar-se del següent fet purament geomètric:¹⁸ en un A -autòmat involutiu i finit, per a tot $a \in A$, $\text{def}_a(\Gamma) = \text{def}_{a^{-1}}(\Gamma)$. Per tant, només caldrà saturar $\text{St}(H)$ aparellant, per cada $a \in A$, els vèrtexs a -deficients amb els (a^{-1}) -deficients, i afegint un a -arc dels primers als segons. L'autòmat Γ obtingut és finit i saturat (per tant, reconeix un subgrup d'índex finit) i conté $\text{St}(H)$ com a subautòmat. Per tant, $H = \langle \text{St}(H) \rangle \leq_{\text{ff}} \langle \Gamma \rangle$, tal com volíem demostrar. Noteu que l'aparellament de vèrtexs deficients de $\text{St}(H)$ es pot fer, en general, de diverses maneres i, per tant, el subgrup obtingut K no és únic, en general. $\square$

EXEMPLE 89. Sigui $H = \langle a^{-1}b, ab^2, a^{-2}b^4, a^{-2}baba^{-1}b^{-1}a^2 \rangle \leq \mathbb{F}_{\{a,b\}}$. Per a obtenir un subgrup d'índex finit de $\mathbb{F}_{\{a,b\}}$ del qual H sigui factor lliure, és suficient calcular $\text{St}(H)$ (en traç continu a la figura 17) i completar-lo a un autòmat saturat (per exemple, tal com s'indica amb traç discontinu). L'autòmat obtingut reconeix un subgrup K d'índex 6 a $\mathbb{F}_{\{a,b\}}$ que conté H com a factor lliure.

¹⁸ Es tracta d'una adaptació natural del ben conegut *handshaking lemma*.

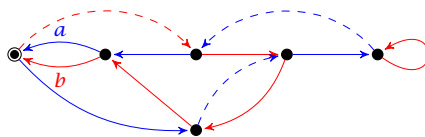


FIGURA 17: Exemple del teorema 88 (Marshall Hall, Jr.).

4.6 El problema de la intersecció

En aquesta darrera secció usarem els autòmats de Stallings per a resoldre el problema proposat a l'exemple 27: el problema de la intersecció de subgrups. Recordem que existeixen grups G que contenen parelles de subgrups $H, K \leq G$ finitament generats amb intersecció $H \cap K$ no finitament generada (a [9] i [14] trobareu exemples relativament senzills de grups d'aquest tipus; analitzats, a més, usant generalitzacions de les tècniques gràfiques que expliquem en aquesta secció). A. G. Howson va estudiar aquest fenomen en el grup lliure i el seu nom s'usa per a referir-se als grups que no tenen aquest comportament «estrany»: un grup G es diu que té la *propietat de Howson* (o que és *Howson*) si la intersecció de qualssevol dos subgrups finitament generats $H, K \leq G$ és sempre finitament generada (per inducció, el mateix és cert llavors per a interseccions d'una quantitat finita de subgrups). Per a un grup qualsevol $G = \langle A \mid R \rangle$ té sentit, doncs, plantejar-se el problema de decisió següent:

PROBLEMA DE LA INTERSECCIÓ DE SUBGRUPS A $G = \langle A \mid R \rangle, \text{SIP}(G)$. *Donades dues famílies finites de paraules en els generadors de G , $u_1, \dots, u_k; v_1, \dots, v_l \in (A^\pm)^*$, decidir si la intersecció dels corresponents subgrups generats $\langle u_1, \dots, u_k \rangle \cap \langle v_1, \dots, v_l \rangle$ (a G) és finitament generada i, en cas afirmatiu, calcular-ne un sistema de generadors.*

A l'article [23], Howson va demostrar que els grups lliures satisfan la propietat que porta el seu nom. A continuació, veurem com els autòmats de Stallings aporten una demostració gràfica molt neta i transparent (i, a més, constructiva!) d'aquest fet remarcable: de la mateixa demostració en sortirà una manera efectiva de calcular una base per a la intersecció, i una fita superior per al rang de $H \cap K$ en termes dels rangs de H i de K : la famosa conjectura de Hanna Neumann, resolta recentment, de la qual parlarem al final de la secció.

DEFINICIÓ 90. Siguin $\Gamma_1 = (V_1, E_1, \iota_1, \tau_1, \ell_1, \odot_1)$ i $\Gamma_2 = (V_2, E_2, \iota_2, \tau_2, \ell_2, \odot_2)$ dos A -autòmats. El *producte*¹⁹ de Γ_1 i Γ_2 , designat per $\Gamma_1 \times \Gamma_2$, és el A -autòmat amb:

- conjunt de vèrtexs el producte cartesià $V_1 \times V_2$;
- un arc $(p_1, p_2) \xrightarrow{a} (q_1, q_2)$ per a cada parell d'arcs $p_1 \xrightarrow{a} q_1$ a Γ_1 , i $p_2 \xrightarrow{a} q_2$ a Γ_2 amb la mateixa etiqueta $a \in A$;

¹⁹ O *pull-back*, en terminologia categòrica.

- les funcions naturals ι , τ , ℓ com a funcions d'incidència i d'etiquetatge;
- vèrtex base $\odot = (\odot_1, \odot_2)$;

vegeu la figura 18.

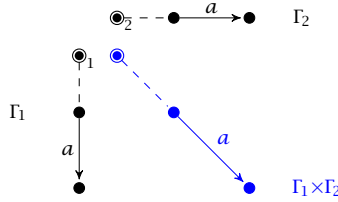


FIGURA 18: Esquema del producte (en blau) de dos A -autòmats (en negre).

Per tant, $\Gamma_1 \times \Gamma_2$ no és res més que el producte tensorial o categòric²⁰ de Γ_1 i Γ_2 , *respectant les etiquetes dels arcs* (i. e., no aparellant mai arcs amb etiquetes diferents). Amb un exemple s'entén clarament el funcionament d'aquesta operació ben natural entre A -autòmats.

EXEMPLE 91. Considerem els subgrups $H = \langle b, a^3, a^{-1}bab^{-1}a \rangle$ i $K = \langle ab, a^3, a^{-1}ba \rangle$ de $\mathbb{F}_{\{a,b\}}$. A la figura 19 podem veure l'autòmat $\text{St}(H)$ dibuixat a la part esquerra (en format vertical), l'autòmat $\text{St}(K)$ dibuixat a la part superior (en format horitzontal), i el producte $\text{St}(H) \times \text{St}(K)$ a la part central (i reorganitzat, a la dreta).

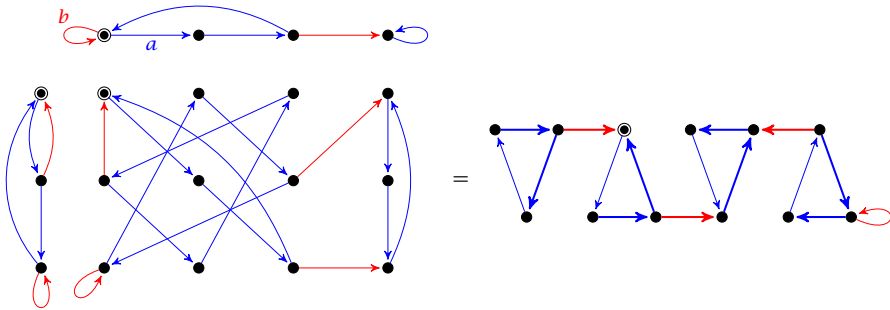


FIGURA 19: Producte dels autòmats $\text{St}(H)$ i $\text{St}(K)$.

La proposició següent recull les propietats principals d'aquest producte d'autòmats, que el relacionen molt de prop amb la intersecció de subgrups.

PROPOSICIÓ 92. *Siguin $\Gamma_1 = (V_1, E_1, \iota_1, \tau_1, \ell_1, \odot_1)$ i $\Gamma_2 = (V_2, E_2, \iota_2, \tau_2, \ell_2, \odot_2)$ dos A -autòmats. Aleshores,*

²⁰ Vegeu https://en.wikipedia.org/wiki/Graph_product.

- (i) per a tot $(p, q) \in \Gamma_1 \times \Gamma_2$, tenim $0 \leq \deg(p, q) \leq \min\{\deg(p), \deg(q)\}$;
- (ii) si Γ_1 i Γ_2 són deterministes, llavors $\Gamma_1 \times \Gamma_2$ també és determinista, i, en tal cas, $\langle \Gamma_1 \times \Gamma_2 \rangle = \langle \Gamma_1 \rangle \cap \langle \Gamma_2 \rangle$;
- (iii) encara que Γ_1 i Γ_2 siguin connexos, $\Gamma_1 \times \Gamma_2$ pot no ser connex;
- (iv) encara que Γ_1 i Γ_2 siguin cor, $\Gamma_1 \times \Gamma_2$ pot no ser cor.

PROVA. (i). És clar de la definició 90, ja que tot a -arc sortint de $(p, q) \in V(\Gamma_1 \times \Gamma_2)$ correspon a un a -arc sortint tant de $p \in V\Gamma_1$ com de $q \in V\Gamma_2$.

(ii). Que el producte d'autòmats deterministes és, de nou, determinista és clar de la definició. Suposem que Γ_1 i Γ_2 són deterministes; aleshores, si $w \in \langle \Gamma_1 \times \Gamma_2 \rangle$, hi ha un $\bullet$ -camí reduït a $\Gamma_1 \times \Gamma_2$ amb etiqueta w ; projectant-lo a la primera i a la segona coordenades, obtenim $\bullet$ -camins reduïts a Γ_1 i a Γ_2 , respectivament, llegint igualment w ; per tant, $w \in \langle \Gamma_1 \rangle \cap \langle \Gamma_2 \rangle$. I, recíprocament, si $w \in \langle \Gamma_1 \rangle \cap \langle \Gamma_2 \rangle$, aleshores hi ha $\bullet$ -camins reduïts a Γ_1 i a Γ_2 , respectivament, ambdós llegint la paraula w ; «cosint-los» arc per arc, obtenim un $\bullet$ -camí reduït a $\Gamma_1 \times \Gamma_2$ que llegeix w i, per tant, $w \in \langle \Gamma_1 \times \Gamma_2 \rangle$.

(iii) i (iv). Els subgrups $H = \langle a^2, b \rangle$ i $K = \langle b, aba^{-1} \rangle$ de $\mathbb{F}_{\{a,b\}}$ serveixen de contraexemple (deixem al lector el càlcul dels respectius autòmats de Stallings, i del seu producte). $\square$

COROLLARI 93. Per $H, K \leq F_A$, tenim $\text{St}(H \cap K) = \text{core}(\text{St}(H) \times \text{St}(K))$. $\square$

Combinat amb la proposició 54, això ens demostra la propietat de Howson, la fita de Hanna Neumann, i ens resol el problema de la intersecció per a grups lliures. Definim el *rang reduït* d'un grup lliure $\mathbb{F}_n$ com $\widetilde{\text{rk}}(\mathbb{F}_n) = \max\{0, n - 1\}$, és a dir, $\widetilde{\text{rk}}(\mathbb{F}_n) = n - 1$ excepte per al grup trivial, per la qual cosa posem $\widetilde{\text{rk}}(1) = 0$ (en lloc de -1).

TEOREMA 94 (HOWSON, [23]; H. NEUMANN, [38]). Els grups lliures $\mathbb{F}$ (de qualsevol rang) són Howson. A més, si $H, K \leq_{\text{fg}} \mathbb{F}$, aleshores $\widetilde{\text{rk}}(H \cap K) \leq 2\widetilde{\text{rk}}(H)\widetilde{\text{rk}}(K)$.

PROVA. Observeu que, com que la propietat de Howson només involucra subgrups finitament generats, és suficient demostrar-la quan l'ambient $\mathbb{F}$ és de rang finit, cosa que assumirem durant la demostració. Si $H, K \leq \mathbb{F}_n$ són finitament generats, aleshores $\text{St}(H)$ i $\text{St}(K)$ són finits, i per tant, $\text{St}(H) \times \text{St}(K)$ també serà finit, d'on (per la proposició 92(ii)) es dedueix que $H \cap K$ torna a ser finitament generada. En conseqüència, $\mathbb{F}_n$ (i, per tant, tot grup lliure) és Howson.

Per veure la desigualtat, observem primer que, si Γ és un graf connex i finit, i Δ és el graf obtingut després d'eliminar successivament vèrtexs de grau 1 (i les corresponents arestes incidents), aleshores $\text{rk}(\Delta) = \text{rk}(\Gamma)$. D'altra banda, per la proposició 92(i), si $\deg(p) \geq 2$ i $\deg(q) \geq 2$, aleshores

$$\deg(p, q) - 2 \leq (\deg(p) - 2)(\deg(q) - 2). \quad (23)$$

És clar que ens podem restringir al cas $H, K, H \cap K \neq \{1\}$. Siguin ara $\Gamma_H = \text{St}(H)$, $\Gamma_K = \text{St}(K)$, Γ la component connexa de $\Gamma_H \times \Gamma_K$ que conté el punt base, i Γ_H^* , Γ_K^* i $\Gamma^* \subseteq \Gamma_H^* \times \Gamma_K^*$ els respectius cors reduïts. Com que Γ^* té tots els vèrtexs (p, q) de grau 2 o superior (i, per tant, p i q també), usant les observacions anteriors, l'equació 6 i les propietats del cor reduït obtenim:

$$\begin{aligned}
 \text{rk}(H \cap K) - 1 &= \text{rk}(\Gamma^*) - 1 = \\
 &= \frac{1}{2} \sum_{(p,q) \in \Gamma^*} (\deg(p, q) - 2) \leq \\
 &\leq \frac{1}{2} \sum_{(p,q) \in \Gamma^*} (\deg(p) - 2)(\deg(q) - 2) \leq \\
 &\leq \frac{1}{2} \sum_{(p,q) \in V(\Gamma_H^* \times \Gamma_K^*)} (\deg(p) - 2)(\deg(q) - 2) = \\
 &= \frac{1}{2} (\sum_{p \in V\Gamma_H^*} (\deg(p) - 2)) (\sum_{q \in V\Gamma_K^*} (\deg(q) - 2)) = \\
 &= 2(\text{rk}(H) - 1)(\text{rk}(K) - 1). \quad \square
 \end{aligned}$$

Amb una demostració una mica més tècnica es pot veure una desigualtat més forta, que va demostrar Walter Neumann a [39]:

$$\sum_{w \in H \setminus \mathbb{F}_A / K} \widetilde{\text{rk}}(H^w \cap K) \leq 2\widetilde{\text{rk}}(H)\widetilde{\text{rk}}(K), \quad (24)$$

on $H^w = w^{-1}Hw$, i el sumatori recorre tots els w d'un transversal qualsevol del conjunt de classes laterals dobles $H \setminus \mathbb{F}_A / K = \{HwK \mid w \in \mathbb{F}_A\}$ (es pot veure que els sumands no nuls es corresponen bijectivament amb les components connexes no arbre i no cíclics del producte $\text{St}(H) \times \text{St}(K)$).

Arribats a aquest punt, és obligat mencionar que Hanna Neumann primer (sobre la desigualtat al teorema 94) i Walter Neumann més tard (sobre la desigualtat (24)) van conjecturar que el factor «2» es pot eliminar de les fites respectives. Aquestes són les famoses «conjectura de Hanna Neumann» i «conjectura forta de Hanna Neumann»,²¹ demostrades (independentment i quasi simultània) fa pocs anys per J. Friedman i I. Mineyev; vegeu [18], [36], i les remarcables simplificacions obtingudes per W. Dicks, a [18, apèndix B] i a [17], respectivament.

TEOREMA 95 (J. FRIEDMAN [18] i I. MINEYEV [36]). *Sigui $\mathbb{F}$ un grup lliure i H, K subgrups finitament generats de $\mathbb{F}$. Aleshores,*

$$\sum_{w \in H \setminus \mathbb{F}_A / K} \widetilde{\text{rk}}(H^w \cap K) \leq \widetilde{\text{rk}}(H)\widetilde{\text{rk}}(K).$$

Com ja hem justificat, la tècnica del producte d'autòmats també ens permet resoldre el problema de la intersecció.

TEOREMA 96. *El problema de la intersecció de subgrups per a grups lliures, $\text{SIP}(\mathbb{F})$, és resoluble.* $\square$

²¹ En anglès, *Hanna Neumann conjecture* i *Strengthened Hanna Neumann conjecture*.

EXEMPLE 97. Com a exemple, acabem el càlcul començat a l'exemple 27. Es tractava de calcular la intersecció de $H = \langle u_1, u_2, u_3 \rangle \leq K = \langle v_1, v_2, v_3 \rangle$ donats per les paraules $u_1 = b$, $u_2 = a^3$, $u_3 = a^2bab^{-1}a$, $v_1 = ab$, $v_2 = a^3$, i $v_3 = a^{-1}ba$, com a subgrups de $\mathbb{F}_A$, $A = \{a, b\}$. A simple vista ja hem trobat els elements $a^3, b^{-1}a^3b, a^{-1}ba^3b^{-1}a \in H \cap K$; però no quedava gens clar si aquests tres elements generen $H \cap K$, o en falten de més complicats per descobrir.

A l'exemple 91 hem calculat $\text{St}(H)$, $\text{St}(K)$ i $\text{St}(H) \times \text{St}(K)$. Com que aquest últim ja és connex i cor, el corollari 93 ens diu que $\text{St}(H \cap K) = \text{St}(H) \times \text{St}(K)$. Prenent com a arbre maximal T l'indicat amb arcs en traç gruixut a la figura 19, obtenim la base

$$S_T = \{b^{-1}a^3b, a^3, a^{-1}ba^3b^{-1}a, a^{-1}bab^{-1}a^3ba^{-1}b^{-1}a, \\ a^{-1}bab^{-1}aba^{-1}ba^{-1}b^{-1}a\}$$

per a la intersecció $H \cap K$.

A més, com que cadascun dels cinc generadors a S_T és l'etiqueta d'un $\bullet$ -camí reduït a $\text{St}(H) \times \text{St}(K)$, el podem projectar a la primera i a la segona coordenades, respectivament, i, usant el mètode de la secció 4.2, podem reescriure'l com a paraula en $\{u_1, u_2, u_3\}$ i en $\{v_1, v_2, v_3\}$:

$$\begin{aligned} H \ni u_1^{-1}u_2u_1 &= b^{-1}a^3b &= v_1^{-1}v_2v_1 \in K, \\ H \ni u_2 &= a^3 &= v_2 \in K, \\ H \ni u_3^3 &= a^{-1}ba^3b^{-1}a &= v_3v_2v_3^{-1} \in K, \\ H \ni u_3u_2u_3^{-1} &= a^{-1}bab^{-1}a^3ba^{-1}b^{-1}a &= v_3v_1^{-1}v_2v_1v_3^{-1} \in K, \\ H \ni u_3u_1u_3^{-1} &= a^{-1}bab^{-1}aba^{-1}ba^{-1}b^{-1}a &= v_3v_1^{-1}v_2v_3v_2^{-1}v_1v_3^{-1} \in K. \end{aligned}$$

OBSERVACIÓ 98. Tal com succeeix amb el problema de la pertinença (vegeu [35]), no és difícil demostrar que el problema de la intersecció és també indecidible per a productes directes de grups lliures (vegeu [14, corollari 2.3]), fet que confirma la indocilitat algorísmica d'aquesta família, en cert sentit propera al grup lliure.

Per acabar, vegem que la mateixa idea és útil també per a resoldre el problema de la intersecció de classes laterals en un grup lliure. Recordem que, si G és un grup qualsevol, $u, v \in G$ dos elements, i $H, K \leq G$ dos subgrups, llavors les classes laterals Hu i Kv o bé són disjunes, o bé s'intersequen exactament en una classe lateral de $H \cap K$, i. e., $Hu \cap Kv = (H \cap K)w$.

PROBLEMA DE LA INTERSECCIÓ DE CLASSES LATERALS A $G = \langle A \mid R \rangle$, $\text{CIP}(G)$. Donades dues famílies finites de paraules $u, u_1, \dots, u_k; v, u_1, \dots, u_l \in (A^\pm)^*$, decidir si la intersecció $\langle u_1, \dots, u_k \rangle u \cap \langle v_1, \dots, v_l \rangle v \subseteq G$ és buida, i, en cas negatiu, calcular-ne un representant.

TEOREMA 99. El problema de la intersecció de classes laterals per a grups lliures, $\text{CIP}(\mathbb{F})$, és resoluble.

PROVA. Donades paraules reduïdes $u, u_1, \dots, u_k; v, v_1, \dots, v_l \in \mathbb{F}_A$, calculem els autòmats de Stallings de $H = \langle u_1, \dots, u_k \rangle$ i $K = \langle v_1, \dots, v_l \rangle$. Ara intentem seguir un camí reduït a $\text{St}(H)$ començant a $\odot_H$ i llegint $u = a_{i_1} \cdots a_{i_n}$ (on $a_{i_1}, \dots, a_{i_n} \in A^\pm$). Si no és possible, perquè després de llegir a_{i_j} arribem a un vèrtex $(a_{i_{j+1}})$ -deficient r , ampliïm l'autòmat $\text{St}(H)$ tot afegint el pèl

$$r \xrightarrow{a_{i_{j+1}}} \bullet \xrightarrow{a_{i_{j+2}}} \bullet \xrightarrow{a_{i_{j+3}}} \dots \xrightarrow{a_{i_n}} p \quad (25)$$

(per poder completar la lectura de u com a etiqueta d'un camí de $\odot_H$ a, diguem-ne p) i dissenyem per $\text{St}_u(H)$ l'autòmat resultant. Anàlogament, construïm $\text{St}_v(K)$ (assegurant-nos que conté un camí llegint v , de $\odot_K$ a, diguem-ne q) i calculem el producte $\Gamma = \text{St}_u(H) \times \text{St}_v(K)$.

Ara, si $(\odot_H, \odot_K)$ i (p, q) estan a la mateixa component connexa de Γ , prenem un camí γ de $(\odot_H, \odot_K)$ a (p, q) , i el projectem a $\text{St}_u(H)$ i a $\text{St}_v(K)$, respectivament. Com que ambdues projeccions tenen etiqueta $w = \bar{\ell}(\gamma)$, i van de $\odot_H$ a p , i de $\odot_K$ a q , respectivament, concloem que $Hu = Hw$ i $Kv = Kw$; per tant, $w \in Hu \cap Hv$.

Recíprocament, si $(\odot_H, \odot_K)$ a (p, q) són a components connexes diferents, no hi ha cap parella de camins reduïts, de $\odot_H$ a p (a $\text{St}_u(H)$) i de $\odot_K$ a q (a $\text{St}_v(K)$), compartint la mateixa etiqueta. Per tant, $Hu \cap Kv \neq \emptyset$ si i només si $(\odot_H, \odot_K)$ i (p, q) són a la mateixa component connexa de $\text{St}_u(H) \times \text{St}_v(K)$. Això completa la demostració. $\square$

Per acabar d'il·lustrar la utilitat d'aquesta tècnica del producte de A -autòmats, donem un argument força curt i elegant per demostrar, en l'àmbit dels grups lliures, un resultat sobre subgrups força conegut i vàlid en general (la prova del cas general requereix eines algebraïques més sofisticades, com pot ser el teorema del subgrup de Kurosh).

PROPOSICIÓ 100. *Si G un grup i siguin H, K, H', K' subgrups de G . Si $H \leq_{\text{ff}} K$ i $H' \leq_{\text{ff}} K'$, llavors $H \cap H' \leq_{\text{ff}} K \cap K'$.*

PROVA. (Fem la demostració només en el cas que G és un grup lliure, $G = \mathbb{F}$.)

Veurem primer que, si $H \leq_{\text{ff}} K \leq \mathbb{F}$, i $L \leq \mathbb{F}$, llavors $H \cap L \leq_{\text{ff}} K \cap L$. Considerem una base A de K que estengui una base de H , i observem que $\text{St}(H, A)$ és simplement una rosa d'uns quants pètals (precisament, els etiquetats pels elements de A pertanyents a H). Considerem $\text{St}(K \cap L, A)$ i calculem $H \cap L = H \cap (K \cap L)$ fent el producte d'aquests dos A -autòmats: clarament, $\text{St}(H, A) \times \text{St}(K \cap L, A)$ és el A -subautòmat de $\text{St}(K \cap L, A)$ determinat pels arcs amb etiqueta a H . Per tant, $H \cap L \leq_{\text{ff}} K \cap L$, tal com volíem.

Aplicant aquest fet dues vegades, obtenim $H \cap H' \leq_{\text{ff}} K \cap H' \leq_{\text{ff}} K \cap K'$ i, per tant, $H \cap H' \leq_{\text{ff}} K \cap K'$, tal com volíem demostrar. $\square$

5 Per saber-ne més

La teoria dels autòmats de Stallings és un tema molt conegut i ben representat a la literatura. A més de l'article original de Stallings [50] (d'orientació més aviat

topològica), existeixen diverses revisions amb un punt de vista més proper al que s'ha usat en aquest article (vegeu *e. g.* [2, 12, 25]). Els articles que fan servir la teoria de Stallings per a resoldre altres problemes específics són innombrables i continuen apareixent amb regularitat. Podeu trobar-ne alguns exemples a [1, 3, 4, 10, 15, 31, 33, 34, 43, 49, 52].

Arran de l'èxit immens de la teoria de Stallings per al grup lliure, també hi ha hagut múltiples extensions d'aquesta idea a àmbits més generals, per exemple: grafs de grups [19, 20, 26, 44, 46, 47, 51]; monoides i semigrups [16, 30, 31, 53]; grups que satisfan certes propietats de petita cancel·lació²² [1]; grups totalment residualment lliures [28, 37, 41]; productes lliures [24]; amalgames de grups finits [32]; grups que actuen lliurement a $\mathbb{Z}^n$ -arbres [42]; grups virtualment lliures [48]; subgrups quasiconvexos [27]; productes directes i semidirectes de grups lliures amb grups lliure-abelians [8, 11, 13]; complexos cúbics CAT(0) [5]; grups relativament hiperbòlics [29]; i grups de Coxeter d'angle recte²³ [7], entre d'altres.

Agraïments

Els autors agraïm el suport parcial rebut de l'Agència Estatal de Investigación, a través del projecte de recerca MTM2017-82740-P (AEI/FEDER, UE). El primer autor va realitzar la primera part d'aquest treball a la Universitat del País Basc (UPV/EHU) amb suport parcial del MINECO a través del projecte PID2019-107444GA-I00, i del Govern basc amb el projecte IT974-16.

Agraïm als revisors i editors l'acurada lectura d'aquest text i els pertinents comentaris i suggeriments realitzats.

Referències

- [1] ARZHANTSEVA, G. N.; OL'SHANSKII, A. YU. «The class of groups all of whose subgroups with lesser number of generators are free is generic». *Math. Notes*, 59 (4) (1996), 350–355.
- [2] BARTHOLDI, L.; SILVA, P. V. «Rational subsets of groups». A: *Handbook of Automata Theory. Vol. II. Automata in Mathematics and Selected Applications*. Berlín: EMS Press, 2021, 841–869.
- [3] BASSINO, F.; MARTINO, A.; NICAUD, C.; VENTURA, E.; WEIL, P. «Statistical properties of subgroups of free groups». *Random Structures Algorithms*, 42 (3) (2013), 349–373.
- [4] BASSINO, F.; NICAUD, C.; WEIL, P. «Random generation of finitely generated subgroups of a free group». *Internat. J. Algebra Comput.*, 18 (2) (2008), 375–405.

²² En anglès, *small cancellation*.

²³ En anglès, *right-angled Coxeter groups*.

- [5] BEEKER, B.; LAZAROVICH, N. «Stallings' folds for cube complexes». *Israel J. Math.*, 227 (1) (2018), 331–363.
- [6] CLAY, M.; MARGALIT, D. (ed.). *Office Hours with a Geometric Group Theorist*. Princeton, Nova Jersey: Princeton University Press, 2017.
- [7] DANI, P.; LEVCOVITZ, I. «Subgroups of right-angled Coxeter groups via Stallings-like techniques». *J. Comb. Algebra*, 5 (3) (2021), 237–295.
- [8] DELGADO, J. «Extensions of free groups: algebraic, geometric, and algorithmic aspects». Tesi doctoral. Universitat Politècnica de Catalunya, 2017.
- [9] DELGADO, J.; ROY, M.; VENTURA, E. «Intersection configurations in free and free times free-abelian groups». Preprint (2022). [Disponible en línia a: arXiv:2107.12426]
- [10] DELGADO, J.; SILVA, P. V. «On the lattice of subgroups of a free group: complements and rank». *J. Groups Complex. Cryptol.*, 12 (1) (2020), article núm. 1, 24 p.
- [11] DELGADO, J.; VENTURA, E. «Algorithmic problems for free-abelian times free groups». *J. Algebra*, 391 (2013), 256–283.
- [12] DELGADO, J.; VENTURA, E. «A list of applications of Stallings automata». *Trans. Comb.*, 11 (2022), 181–235.
- [13] DELGADO, J.; VENTURA, E. «Stallings automata for free-times-abelian groups: intersections and index». *Publ. Mat.*, 66 (2) (2022), 789–830. DOI: 10.5565/PUBLMAT6622209.
- [14] DELGADO, J.; VENTURA, E.; ZAKHAROV, A. «Intersection problem for Droms RAAGs». *Internat. J. Algebra Comput.*, 28 (7) (2018), 1129–1162.
- [15] DELGADO, J.; VENTURA, E.; ZAKHAROV, A. «Relative order and spectrum in free and related groups». *Commun. Contemp. Math.* (per aparèixer). [Disponible en línia a: arXiv:2105.03798]
- [16] DELGADO, M.; MARGOLIS, S.; STEINBERG, B. «Combinatorial group theory, inverse monoids, automata, and global semigroup theory». *Internat. J. Algebra Comput.*, 12 (1-2) (2002), 179–211.
- [17] DICKS, W. «Simplified Mineyev's proof of Hanna Neumann conjecture». Preprint (2011).
- [18] FRIEDMAN, J. «Sheaves on graphs, their homological invariants, and a proof of the Hanna Neumann conjecture: with an appendix by Warren Dicks». *Mem. Amer. Math. Soc.*, 233 (1100) (2015), 106 p.
- [19] GUIRADEL, V. «Approximations of stable actions on $\mathbf{R}$ -trees». *Comment. Math. Helv.*, 73 (1) (1998), 89–121.
- [20] GUIRADEL, V. «Reading small actions of a one-ended hyperbolic group on $\mathbf{R}$ -trees from its JSJ splitting». *Amer. J. Math.*, 122 (4) (2000), 667–688.
- [21] HALL, M., JR. «Subgroups of finite index in free groups». *Canad. J. Math.*, 1 (1949), 187–190.
- [22] HALL, M., JR. «A topology for free groups and related groups». *Ann. of Math.* (2), 52 (1950), 127–139.

- [23] HOWSON, A. G. «On the intersection of finitely generated free groups». *J. London Math. Soc.*, 29 (1954), 428–434.
- [24] IVANOV, S. V. «On the intersection of finitely generated subgroups in free products of groups». *Internat. J. Algebra Comput.*, 9 (5) (1999), 521–528.
- [25] KAPOVICH, I.; MYASNIKOV, A. «Stallings foldings and subgroups of free groups». *J. Algebra*, 248 (2) (2002), 608–668.
- [26] KAPOVICH, I.; WEIDMANN, R.; MIASNIKOV, A. «Foldings, graphs of groups and the membership problem». *Internat. J. Algebra Comput.*, 15 (1) (2005), 95–128.
- [27] KHARLAMPOVICH, O.; MIASNIKOV, A.; WEIL, P. «Stallings graphs for quasi-convex subgroups». *J. Algebra*, 488 (2017), 442–483.
- [28] KHARLAMPOVICH, O. G.; MYASNIKOV, A. G.; REMESLENNIKOV, V. N.; SERBIN, D. E. «Subgroups of fully residually free groups: algorithmic problems». A: *Group Theory, Statistics, and Cryptography*. Providence, RI: Amer. Math. Soc., 2004, 63–101. (Contemp. Math.; 360)
- [29] KHARLAMPOVICH, O.; WEIL, P. «On the generalized membership problem in relatively hyperbolic groups». A: *Fields of Logic and Computation. III*. Cham: Springer, 2020, 147–155. (Lecture Notes in Comput. Sci.; 12180)
- [30] MARGOLIS, S. W.; MEAKIN, J. C. «Free inverse monoids and graph immersions». *Internat. J. Algebra Comput.*, 3 (1) (1993), 79–99.
- [31] MARGOLIS, S.; SAPIR, M.; WEIL, P. «Closed subgroups in pro- V topologies and the extension problem for inverse automata». *Internat. J. Algebra Comput.*, 11 (4) (2001), 405–445.
- [32] MARKUS-EPSTEIN, L. «Stallings foldings and subgroups of amalgams of finite groups». *Internat. J. Algebra Comput.*, 17 (8) (2007), 1493–1535.
- [33] MARTINO, A.; VENTURA, E. «On automorphism-fixed subgroups of a free group». *J. Algebra*, 230 (2) (2000), 596–607.
- [34] MIASNIKOV, A.; VENTURA, E.; WEIL, P. «Algebraic extensions in free groups». A: *Geometric Group Theory*. Basel: Birkhäuser, 2007, 225–253. (Trends Math.)
- [35] MIHAĬLOVA, K. A. «The occurrence problem for direct products of groups». *Dokl. Akad. Nauk SSSR*, 119 (1958), 1103–1105. [En rus]
- [36] MINEYEV, I. «Submultiplicativity and the Hanna Neumann conjecture». *Ann. of Math. (2)*, 175 (1) (2012), 393–414.
- [37] MYASNIKOV, A. G.; REMESLENNIKOV, V. N.; SERBIN, D. E. «Fully residually free groups and graphs labeled by infinite words». *Internat. J. Algebra Comput.*, 16 (4) (2006), 689–737.
- [38] NEUMANN, H. «On the intersection of finitely generated free groups». *Publ. Math. Debrecen*, 4 (1956), 186–189.
- [39] NEUMANN, W. D. «On intersections of finitely generated subgroups of free groups». A: *Groups—Canberra 1989*. Berlin: Springer, 1990, 161–170. (Lecture Notes in Math.; 1456)

- [40] NIELSEN, J. «Om Regning med ikke-kommutative Faktorer og dens Anvendelse i Gruppeteorien». *Mat. Tidsskr. B* (1921), 77–94.
- [41] NIKOLAEV, A. V.; SERBIN, D. E. «Finite index subgroups of fully residually free groups». *Internat. J. Algebra Comput.*, 21 (4) (2011), 651–673.
- [42] NIKOLAEV, A.; SERBIN, D. «Membership problem in groups acting freely on $\mathbb{Z}^n$ -trees». *J. Algebra*, 370 (2012), 410–444.
- [43] PUDER, D. «Primitive words, free factors and measure preservation». *Israel J. Math.*, 201 (1) (2014), 25–73.
- [44] RIPS, E.; SELA, Z. «Cyclic splittings of finitely presented groups and the canonical JSJ decomposition». *Ann. of Math. (2)*, 146 (1) (1997), 53–109.
- [45] SCHREIER, O. «Die Untergruppen der freien Gruppen». *Abh. Math. Sem. Univ. Hamburg*, 5 (1) (1927), 161–183. [En alemany]
- [46] SELA, Z. «Acyindrical accessibility for groups». *Invent. Math.*, 129 (3) (1997), 527–565.
- [47] SELA, Z. «Diophantine geometry over groups. I. Makanin-Razborov diagrams». *Publ. Math. Inst. Hautes Études Sci.*, 93 (2001), 31–105.
- [48] SILVA, P. V.; SOLER-ESCRIVÀ, X.; VENTURA, E. «Finite automata for Schreier graphs of virtually free groups». *J. Group Theory*, 19 (1) (2016), 25–54.
- [49] SILVA, P. V.; WEIL, P. «On an algorithm to decide whether a free group is a free factor of another». *Theor. Inform. Appl.*, 42 (2) (2008), 395–414.
- [50] STALLINGS, J. R. «Topology of finite graphs». *Invent. Math.*, 71 (3) (1983), 551–565.
- [51] STALLINGS, J. R. «Foldings of G -trees». A: *Arboreal Group Theory*. Nova York: Springer, 1991, 355–368. (Math. Sci. Res. Inst. Publ.; 19)
- [52] STALLINGS, J. R.; WOLF, A. R. «The Todd-Coxeter process, using graphs». A: *Combinatorial Group Theory and Topology*. Princeton, NJ: Princeton Univ. Press, 1987, 157–161. (Ann. of Math. Stud.; 111)
- [53] STEINBERG, B. «Inverse automata and profinite topologies on a free group». *J. Pure Appl. Algebra*, 167 (2-3) (2002), 341–359.

DEPARTAMENT DE MATEMÀTIQUES

UNIVERSITAT POLITÈCNICA DE CATALUNYA (UPC)

ESCOLA POLITÈCNICA SUPERIOR D'ENGINYERIA DE MANRESA (EPSEM)

AV. DE LES BASES DE MANRESA, 61, 73, 08242 MANRESA (BARCELONA)

{jorge.delgado,enric.ventura}@upc.edu

Estimació de la mitjana: una revisió d'avenços recents

GÁBOR LUGOSI

Resum: Segurament el problema estadístic més bàsic és estimar el valor esperat d'una variable aleatòria basant-ho en observacions independents de la mateixa distribució. Aquesta pregunta aparentment senzilla planteja dificultats sorprenents tant des del punt de vista estadístic com computacional. En aquest article analitzem alguns dels avenços recents en l'estimació de la mitjana. En particular, es descriuen estimadors de mitjanes subgaussians per a dades que poden tenir cues pesades, tant en configuracions univariants com multivariants. Ens centrem en estimadors basats en tècniques de medianes de mitjanes però també es revisen altres mètodes com la mitjana retallada i l'estimador de Catoni. Presentem demostracions detallades dels resultats bàsics.

Paraules clau: estimació de la mitjana, distribucions de cua pesada, robustesa, aprenentatge estadístic.

Classificació MSC2010: 62G05, 62G15, 62G35.

1 Introducció

Sens dubte, el problema més fonamental de l'estadística és estimar el valor esperat μ d'una variable aleatòria X basant-se en una mostra de n observacions independents i idènticament distribuïdes extretes d'una distribució de X . L'elecció òbvia d'estimador és, per descomptat, la mitjana empírica. Les seves propietats són ben conegudes a partir de resultats clàssics de la teoria de la probabilitat. Tanmateix, des dels inicis, els experts en estadística s'han preocupat per la qualitat de la mitjana empírica, especialment quan la distribució pot ser de *cua pesada* o hi pot haver *dades atípiques*. Aquesta preocupació va donar lloc a l'àrea de l'*estadística robusta*, que aborda el problema d'estimació de la mitjana (i altres problemes estadístics) per a aquest tipus de dades. Les referències clàssiques inclouen Huber [31], Huber i Ronchetti [32], Hampel, Ronchetti, Rousseeuw i Stahel [23] i Tukey [59].

Motivat per aplicacions en aprenentatge automàtic i en ciència de dades, en els darrers anys hi ha hagut un interès creixent en la construcció d'estimacions

de la mitjana i de funcions de regressió amb el requisit que els estimadors hagin d'aconseguir *alta precisió* amb una gran *confiança*. Avui dia, s'entén molt millor quin és el millor balanç precisió/confiança que es pot assolir i l'objectiu d'aquest article és analitzar-ne alguns dels avenços recents. Ens centrarem en el problema d'estimació de la mitjana, tant en configuracions univariants com multivariants. Presentem una discussió detallada sobre quin és el millor rendiment que hom pot esperar, sobre com descriure una varietat d'estimadors i sobre com analitzar-ne el rendiment. Parem una atenció especial a una metodologia senzilla però potent basada en tècniques de *medianes de mitjanes*.

A la secció 2, abordem el problema més simple, el d'estimació de la mitjana univariant. Ens centrem en estimadors *subgaussians* i n'explorem les possibilitats i limitacions. La secció 3 es dedica al problema multivariant, que és substancialment més difícil. Ampliem la noció d'estimador subgaussià en el context multivariant i analitzem diversos estimadors.

Per a una revisió més detallada, remetem el lector al recent article de Lugosi i Mendelson [45].

2 Estimació de la mitjana d'una variable aleatòria real

Comencem examinant el problema clàssic d'estimar la mitjana d'una variable aleatòria. Siguin $X_1, \dots, X_n$ variables aleatòries independents, distribuïdes de manera idèntica, amb mitjana $\mu = \mathbb{E}X_1$. Es pretén estimar μ a partir d'observacions d'aquestes variables aleatòries. Un estimador $\hat{\mu}_n = \hat{\mu}_n(X_1, \dots, X_n)$ és simplement una funció (mesurable) de $X_1, \dots, X_n$.

Per descomptat, l'opció més natural d'un estimador de la mitjana és la mitjana empírica estàndard

$$\bar{\mu}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Tanmateix, tal com argumentem a continuació, hi pot haver altres estimadors més adients, segons quin sigui l'objectiu.

La qualitat d'un estimador es pot mesurar de diverses maneres. Els primers treballs estadístics es van centrar, en la seva majoria, en mesures del risc esperat, com ara l'*error quadràtic mitjà*

$$\mathbb{E}[(\hat{\mu}_n - \mu)^2].$$

Si hom se centra en l'error quadràtic mitjà, aleshores la mitjana empírica té un rendiment excel·lent. Suposem que els X_i tenen un segon moment finit i $\sigma^2 = \mathbb{E}[(X_i - \mu)^2]$ denota la seva variància. Aleshores, és un exercici fàcil veure que l'error quadràtic mitjà de $\bar{\mu}_n$ és igual a σ^2/n . Al mateix temps, això és òptim fins i tot si suposem que la distribució de X_i és gaussiana. De fet, del teorema clàssic de Rao-Blackwell ([55, 4, 36]) es desprèn que no es pot esperar obtenir un error quadràtic mitjà més petit quan s'estima la mitjana d'una variable aleatòria gaussiana.

No obstant això, les mesures de risc com l'error quadràtic mitjà poden induir a error. De fet, si la diferència $|\hat{\mu}_n - \mu|$ no és prou concentrada, aleshores el valor esperat no necessàriament reflecteix el comportament «típic» de l'error. Per motius com aquest, sovint preferim els estimadors $\hat{\mu}_n$, que són propers a μ amb alta probabilitat. Així, el nostre objectiu és entendre, per a qualsevol mida de mostra n i el paràmetre de confiança $\delta \in (0, 1)$, el valor més petit possible $\epsilon = \epsilon(n, \delta)$ tal que

$$\mathbb{P}\{|\hat{\mu}_n - \mu| > \epsilon\} \leq \delta. \quad (1)$$

És important destacar que (1) és un criteri no asimptòtic: hom voldria obtenir estimacions quantitatives sobre com la precisió ϵ s'escala amb el paràmetre de confiança δ i la mida de la mostra n . Aquest tipus d'estimació recorda el PAC (acrònim de *probablement aproximadament correcte*), un marc adoptat habitualment en la teoria de l'aprenentatge estadístic; vegeu Valiant [61], Vapnik i Chervonenkis [63] i Blumer, Ehrenfeucht, Haussler i Warmuth [5].

Les propietats de la mitjana empírica s'entenen molt bé, ja que les mitjanes de variables aleatòries independents s'han estudiat des dels inicis de la teoria de la probabilitat. Una pedra angular d'aquesta teoria és el teorema del límit central. Aquest teorema garanteix que la mitjana empírica té cues gaussianes, asimptòticament, quan n tendeix a infinit. Més precisament,

$$\mathbb{P}\left\{|\bar{\mu}_n - \mu| > \frac{\sigma \Phi^{-1}(1 - \delta/2)}{\sqrt{n}}\right\} \rightarrow \delta,$$

on $\Phi(x) = \mathbb{P}\{G \leq x\}$ és la funció de distribució acumulada d'una variable aleatòria normal estàndard G . Es pot veure fàcilment (per exemple, utilitzant el fet que per a $t \geq 1$, $\exp(-t^2/2) \leq t \exp(-t^2/2)$), que per a tots els $x \geq 0$,

$$1 - \Phi(x) \leq e^{-x^2/2}.$$

Això implica que $\Phi^{-1}(1 - \delta/2) \leq \sqrt{2 \log(2/\delta)}$, i el teorema del límit central afirma que

$$\lim_{n \rightarrow \infty} \mathbb{P}\left\{|\bar{\mu}_n - \mu| > \frac{\sigma \sqrt{2 \log(2/\delta)}}{\sqrt{n}}\right\} \leq \delta.$$

No obstant això, aquesta és una estimació asimptòtica i no la quantitativa que esperàvem. Tot i així, el nostre objectiu és obtenir fites de rendiment no asimptòtiques de la mateixa forma. En particular, diem que un estimador de la mitjana $\hat{\mu}_n$ és *L-subgaussiana* si existeix una constant $L > 0$ de manera que per a totes les mides de mostra n i amb probabilitat almenys $1 - \delta$, se satisfà

$$|\hat{\mu}_n - \mu| \leq \frac{L\sigma \sqrt{\log(2/\delta)}}{\sqrt{n}}. \quad (2)$$

Val la pena esmentar, en aquest punt, el fet conegut que, si tot el que se sap és que la distribució desconeguda és gaussiana, llavors la mitjana empírica

és òptima per a totes les mides de mostra n i tots els nivells de confiança δ (vegeu Catoni [8, proposició 6.1] per a un enunciat precís). A més, la següent observació, establerta per Devroye, Lerasle, Lugosi i Oliveira [17], mostra que (2) és essencialment el millor que es pot esperar en general, fins i tot si hom està interessat en un nivell de confiança fixat:

TEOREMA 1. *Sigui $n > 5$ un número enter positiu. Siguin $\mu \in \mathbb{R}$, $\sigma > 0$ i $\delta \in (2e^{-n/4}, 1/2)$. Aleshores, per a qualsevol estimador de la mitjana $\hat{\mu}_n$, existeix una distribució amb mitjana μ i variància σ^2 tal que*

$$\mathbb{P}\left\{|\hat{\mu}_n - \mu| > \sigma \sqrt{\frac{\log(1/\delta)}{n}}\right\} \geq \delta.$$

PROVA. Per trobar la fita inferior, n'hi ha prou de considerar dues distribucions, P_+ , P_- , ambdues concentrades en dos punts, definits per

$$P_+(\{0\}) = P_-(\{0\}) = 1 - p, \quad P_+(\{c\}) = P_-(\{-c\}) = p,$$

on $p \in [0, 1]$ i $c > 0$. Observem que les mitjanes de les dues distribucions són $\mu_{P_+} = pc$ i $\mu_{P_-} = -pc$ i que totes dues tenen variància $\sigma^2 = c^2p(1 - p)$.

Per a $i = 1, \dots, n$, considerem parells independents de variables aleatòries de valors reals (X_i, Y_i) tals que

$$\mathbb{P}\{X_i = Y_i = 0\} = 1 - p \quad \text{i} \quad \mathbb{P}\{X_i = c, Y_i = -c\} = p.$$

Fixem-nos que X_i es distribueix com P_+ i Y_i es distribueix com P_- . Sigui $\delta \in (0, 1/2)$. Si $\delta \geq 2e^{-n/4}$ i $p = (1/(2n)) \log(2/\delta)$, aleshores (utilitzant $1 - p \geq \exp(-p/(1 - p))$),

$$\mathbb{P}\{(X_1, \dots, X_n) = (Y_1, \dots, Y_n)\} = (1 - p)^n \geq 2\delta.$$

Sigui $\hat{\mu}_n$ un estimador de la mitjana qualsevol, que possiblement depèn de δ . Aleshores,

$$\begin{aligned} & \max(\mathbb{P}\{|\hat{\mu}_n(X_1^n) - \mu_{P_+}| > cp\}, \mathbb{P}\{|\hat{\mu}_n(Y_1^n) - \mu_{P_-}| > cp\}) \geq \\ & \geq \frac{1}{2} \mathbb{P}\{|\hat{\mu}_n(X_1, \dots, X_n) - \mu_{P_+}| > cp \quad \text{o} \quad |\hat{\mu}_n(Y_1, \dots, Y_n) - \mu_{P_-}| > cp\} \geq \\ & \geq \frac{1}{2} \mathbb{P}\{\hat{\mu}_n(X_1, \dots, X_n) = \hat{\mu}_n(Y_1, \dots, Y_n)\} \geq \\ & \geq \frac{1}{2} \mathbb{P}\{(X_1, \dots, X_n) = (Y_1, \dots, Y_n)\} \geq \delta. \end{aligned}$$

A partir de $\sigma^2 = c^2 p(1-p)$ i $p \leq 1/2$, tenim que $cp \geq \sigma\sqrt{p/2}$, i per tant,

$$\max \left(\mathbb{P} \left\{ |\hat{\mu}_n(X_1, \dots, X_n) - \mu_{P_+}| > \sigma \sqrt{\frac{\log \frac{2}{\delta}}{n}} \right\}, \right. \\ \left. \mathbb{P} \left\{ |\hat{\mu}_n(Y_1, \dots, Y_n) - \mu_{P_-}| > \sigma \sqrt{\frac{\log \frac{2}{\delta}}{n}} \right\} \right) \geq \delta.$$

D'aquí se'n desprèn el teorema 1. $\square$

Considerant el teorema 1, el nostre objectiu és dissenyar estimadors que tinguin un rendiment amb una taxa d'error subgaussiana.

Naturalment, la primera tasca a fer és comprovar si l'elecció més òbvia d'estimador de la mitjana —la mitjana empírica— és L -subgaussiana per a algun L . D'una banda, és fàcil veure que, sota determinades condicions sobre la distribució de X_i , la mitjana empírica mostra un rendiment subgaussiana. De fet, si els X_i són tals que existeix $L > 0$ tal que per a tot $\lambda > 0$

$$\mathbb{E} e^{\lambda(X_i - \mu)} \leq e^{\sigma^2 \lambda^2 / L^2},$$

aleshores la mitjana empírica $\bar{\mu}_n$ és L -subgaussiana per a tota $\delta \in (0, 1)$, tal com es veu fàcilment amb la fita de Chernoff.

D'altra banda, els supòsits d'aquest tipus són força restrictius i imposen condicions fortes sobre el decaïment de la probabilitat de la cua de les X_i . Quan les X_i no presenten aquest decaïment de la cua, la mitjana empírica no té per què ser subgaussiana.

Per exemple, si assumim només que σ existeix (és a dir, que la variància de X_i és finita), llavors es pot utilitzar la desigualtat de Txebychev. Aquesta desigualtat implica que, amb probabilitat almenys $1 - \delta$,

$$|\bar{\mu}_n - \mu| \leq \sigma \sqrt{\frac{1}{n\delta}}. \quad (3)$$

No obstant això, sota el supòsit de la finitud del segon moment, això és essencialment el millor que es pot esperar. Tot i que la fita de (3) decau amb la mida de la mostra a la velocitat òptima de $O(n^{-1/2})$, la dependència del paràmetre de confiança δ és exponencialment pitjor que a (2). Remetem a Catoni [8, proposició 6.2] per a una formulació precisa i per a un exemple senzill que (gairebé) satura la desigualtat de Txebychev.

Tot plegat condueix a una conclusió inevitable: si hom està buscant un estimador de la mitjana que sigui subgaussiana per a qualsevol variable aleatòria que tingui una mitjana i una variància ben definides, llavors cal trobar alternatives a la mitjana mostral. Cal dir que, potser sorprenentment, existeixen estimadors de la mitjana que aconsegueixen un rendiment subgaussiana per a totes les distribucions amb una variància finita. En les dues seccions següents es presenten i s'analitzen dos estimadors força diferents.

2.1 L'estimador mediana de les mitjanes

L'estimador mediana de les mitjanes que es presenta a continuació s'ha proposat de maneres diferents en diversos articles; vegeu Nemirovsky i Yudin [53], Hsu [28], Jerrum, Valiant i Vazirani [33] i Alon, Matias i Szegedy [1].

L'estimador mediana de les mitjanes divideix primer les dades en k grups de mida aproximadament igual, calcula la mitjana empírica de cada grup, i pren la mediana dels valors obtinguts.

Formalment, recordem que la mediana de k números reals $x_1, \dots, x_k \in \mathbb{R}$ es defineix com a $M(x_1, \dots, x_k) = x_i$, on x_i és tal que

$$|\{j \in [k] : x_j \leq x_i\}| \geq \frac{k}{2} \quad \text{i} \quad |\{j \in [k] : x_j \geq x_i\}| \geq \frac{k}{2}.$$

(Si hi ha diversos índexs i que s'ajusten a la descripció anterior, prenem el més petit.)

Considerem ara $1 \leq k \leq n$ i dividim el conjunt $[n] = \{1, \dots, n\}$ en k blocs $B_1, \dots, B_k$, cadascun d'ells de mida $|B_i| \geq \lfloor n/k \rfloor \geq 2$.

Donat $X_1, \dots, X_n$, calculem la mitjana mostral de cada bloc

$$Z_j = \frac{1}{|B_j|} \sum_{i \in B_j} X_i$$

i definim l'estimador mediana de les mitjanes com a $\hat{\mu}_n = M(Z_1, \dots, Z_k)$.

Per copsar de manera intuïtiva per què funciona aquest estimador, observem que, per a cada bloc, la mitjana empírica és un estimador no esbiaixat de la mitjana, amb desviació estàndard $\sigma/\sqrt{n/k}$. Per tant, la mediana de la distribució de les mitjanes empíriques per blocs difereix com a màxim $\sigma/\sqrt{n/k}$ de l'esperança. Així, doncs, la mediana empírica és un estimador altament concentrat d'aquesta mediana.

A continuació establim una fita del rendiment de l'estimador. Per simplicitat, suposem que n és divisible per k , de manera que cada bloc té $m = n/k$ elements.

TEOREMA 2. *Siguin $X_1, \dots, X_n$ variables aleatòries independents i distribuïdes de manera idèntica amb mitjana μ i variància σ^2 . Siguin m, k enters positius i suposem que $n = mk$. Aleshores, l'estimador mediana de mitjanes $\hat{\mu}_n$ amb k blocs satisfà*

$$\mathbb{P}\{|\hat{\mu}_n - \mu| > \sigma\sqrt{4/m}\} \leq e^{-k/8}.$$

En particular, per a qualsevol $\delta \in (0, 1)$, si $k = \lceil 8 \log(1/\delta) \rceil$, aleshores, amb probabilitat almenys $1 - \delta$,

$$|\hat{\mu}_n - \mu| \leq \sigma \sqrt{\frac{32 \log(1/\delta)}{n}}.$$

PROVA. Per la desigualtat de Txebeixev, per a cada $j = 1, \dots, k$, amb probabilitat almenys $3/4$,

$$|Z_j - \mu| \leq \sigma \sqrt{\frac{4}{m}}.$$

Així, $|\hat{\mu}_n - \mu| > \sigma \sqrt{4/m}$ implica que almenys $k/2$ dels mitjanes Z_j són tals que $|Z_j - \mu| > \sigma \sqrt{4/m}$. Per tant, si $\text{Bin}(k, 1/4)$ és una variable aleatòria binomial $(k, 1/4)$,

$$\begin{aligned} \mathbb{P}\{|\hat{\mu}_n - \mu| > \sigma \sqrt{4/m}\} &\leq \mathbb{P}\{\text{Bin}(k, 1/4) \geq \frac{k}{2}\} = \\ &= \mathbb{P}\left\{\text{Bin}(k, 1/4) - \mathbb{E} \text{Bin}(k, 1/4) \geq \frac{k}{4}\right\} \leq \\ &\leq e^{-k/8}, \end{aligned}$$

on en l'últim pas hem utilitzat una fita exponencial estàndard per a la cua d'una variable aleatòria binomial (per exemple, la desigualtat de Hoeffding [24]). $\square$

El teorema 2 mostra que l'estimador mediana de les mitjanes té un rendiment subgaussià amb $L = 8$ per a totes les distribucions amb una variància finita. Tanmateix, és important assenyalar que l'estimador $\hat{\mu}_n$ depèn del nivell de confiança δ ja que el nombre de blocs k es tria en funció de δ . Es tracta d'una propietat poc desitjable, ja que per a diferents valors del paràmetre de confiança δ s'obté un estimador puntual diferent. Malgrat això, sense informació addicional sobre la distribució, és impossible construir un sol estimador que sigui subgaussià per a un rang no trivial de valors del paràmetre de confiança δ ; vegeu Devroye, Lerasle, Lugosi i Oliveira [17] per a més detalls. No obstant, en el teorema 4, que enunciem més endavant, es presenta un resultat més positiu.

Els resultats de Bubeck, Cesa-Bianchi i Lugosi [7] i Devroye, Lerasle, Lugosi i Oliveira [17] mostren que l'estimador mediana de les mitjanes es pot utilitzar fins i tot si la distribució de X_i té una variància infinita però té un moment finit d'ordre $1 + \alpha$ per a alguna $\alpha \in (0, 1)$. En concret, tenim el resultat següent:

TEOREMA 3. *Sigui $\alpha \in (0, 1]$ i suposem que $X_1, \dots, X_n$ són variables aleatòries independents i distribuïdes idènticament amb mitjana μ i moment central $(1 + \alpha)$ -èsim $M = \mathbb{E}[|X_i - \mu|^{1+\alpha}]$. Siguin m, k números enters positius i suposem que $n = mk$. Aleshores, l'estimador mediana de mitjanes amb $k = \lceil 8 \log(2/\delta) \rceil$ blocs satisfà*

$$\mathbb{P}\left\{|\hat{\mu}_n - \mu| > 8 \left(\frac{12M^{1/\alpha} \log(1/\delta)}{n} \right)^{\alpha/(1+\alpha)}\right\} \leq \delta.$$

A més, per a qualsevol estimador de la mitjana $\hat{\mu}_n$, existeix una distribució amb mitjana μ i moment central $(1 + \alpha)$ -èsim M tal que

$$\mathbb{P}\left\{|\hat{\mu}_n - \mu| > \left(\frac{M^{1/\alpha} \log(2/\delta)}{n} \right)^{\alpha/(1+\alpha)}\right\} \geq \delta.$$

La prova de la primera part s'obté demostrant que, si $c(\alpha)$ és una constant adient que només depèn de α i

$$\eta \geq c(\alpha) (\mathbb{E}|X_i - \mu|^{1+\alpha})^{1/(1+\alpha)} \left(\frac{1}{m}\right)^{\alpha/(1+\alpha)},$$

aleshores

$$\mathbb{P}\left(\left|\frac{1}{m} \sum_{i=1}^m X_i - \mu\right| \geq \eta\right) \leq 0.2.$$

La prova de la segona afirmació va en la línia del teorema 1.

Acabem aquesta secció demostrant que, si la distribució de X té un moment finit d'ordre $2 + \alpha$ per a algun $\alpha > 0$, aleshores l'estimador mediana de mitjanes té un rendiment subgaussià sota un rang més ampli d'eleccions del paràmetre k que compta el nombre de blocs. La fita següent es deu a Minsker i Strawn [52]. Per simplificar l'exposició, només considerem el cas $\alpha = 1$.

TEOREMA 4. *Siguin $X_1, \dots, X_n$ variables aleatòries independents distribuïdes idènticament amb mitjana μ , variància σ^2 , i tercer moment central $\rho = \mathbb{E}|X - \mu|^3$. Siguin m, k números enters positius tals que $n = mk$. Suposem que*

$$\sqrt{\frac{\log(2/\delta)}{2k}} + \frac{\rho}{2\sigma^3\sqrt{m}} \leq 1/4. \quad (4)$$

Aleshores, l'estimador mediana de mitjanes $\hat{\mu}_n$ amb k blocs satisfà que, amb probabilitat almenys $1 - \delta$,

$$|\hat{\mu}_n - \mu| \leq \frac{1}{c} \left(\sigma \sqrt{\frac{\log(2/\delta)}{2n}} + \frac{\rho k}{2\sigma^2 n} \right),$$

on $c = \phi(\Phi^{-1}(3/4))$ és una constant, mentre que ϕ i Φ denoten, respectivament, les funcions de densitat i de distribució normals estàndard.

Observem que el primer terme a la part dreta de la fita és de forma subgausiana. El segon terme és menor que el primer sempre que el nombre k de blocs satisfaci

$$k \leq \frac{2\sigma^3}{\rho} \sqrt{n \log(2/\delta)}.$$

En particular, $k \leq \frac{2\sigma^3}{\rho} \sqrt{n}$ és suficient per tal d'obtenir un rendiment subgaussià. Això és bo, ja que amb aquesta opció l'estimador no depèn del valor del paràmetre de confiança δ i, simultàniament, l'estimador és subgaussià per a tot el rang de valors de δ permesos per la condició (4). A més, observem que es pot triar el nombre de blocs molt més gran que l'elecció suggerida pel teorema 2. En particular, k pot ser tan gran com un múltiple constant de $\sqrt{n}$. En aquest cas, l'estimador mediana de mitjanes és subgaussià simultàniament

per a tota $\delta \geq e^{-c_0\sqrt{n}}$ per a una constant c_0 adequada. El preu a pagar és la hipòtesi addicional de l'existència del tercer moment. Minsker i Strawn [52] també demostren que, quan $k = o(\sqrt{n})$, aleshores, sota les hipòtesis del teorema 4, $\sqrt{n}(\hat{\mu}_n - \mu)$ és asimptòticament normal amb mitjana zero i variància $\sigma^2\pi/2$.

PROVA. Observem que $\hat{\mu}_n \in [\mu - a, \mu + a]$ si $a > 0$ és tal que

$$\frac{1}{k} \sum_{j=1}^k \mathbb{1}_{Z_j - \mu \leq a} \geq \frac{1}{2} \quad \text{i} \quad \frac{1}{k} \sum_{j=1}^k \mathbb{1}_{Z_j - \mu \geq -a} \geq \frac{1}{2}.$$

Demostrem que, amb probabilitat almenys $1 - \delta$, podem prendre

$$a = \frac{1}{c} \left(\sigma \sqrt{\frac{\log(2/\delta)}{2n}} + \frac{\rho k}{2\sigma^2 n} \right).$$

A tal efecte, notem que

$$\begin{aligned} \frac{1}{k} \sum_{j=1}^k \mathbb{1}_{Z_j - \mu \leq a} &= \frac{1}{k} \sum_{j=1}^k (\mathbb{1}_{Z_j - \mu \leq a} - \mathbb{P}\{Z_j - \mu \leq a\}) + \\ &\quad + \left(\mathbb{P}\{Z_1 - \mu \leq a\} - \mathbb{P}\left\{G \frac{\sigma}{\sqrt{m}} \leq a\right\} \right) + \\ &\quad + \mathbb{P}\left\{G \frac{\sigma}{\sqrt{m}} \leq a\right\} \end{aligned}$$

(on G és una variable aleatòria normal estàndard).

Notem primer que, per la desigualtat de Hoeffding, amb probabilitat almenys $1 - \delta/2$,

$$\frac{1}{k} \sum_{j=1}^k (\mathbb{1}_{Z_j - \mu \leq a} - \mathbb{P}\{Z_j - \mu \leq a\}) \geq -\sqrt{\frac{\log(2/\delta)}{2k}}.$$

Per al segon terme del costat dret, podem utilitzar el teorema de Berry-Esseen (vegeu Shevtsova [56]), que implica que

$$\mathbb{P}\{Z_1 - \mu \leq a\} - \mathbb{P}\left\{G \frac{\sigma}{\sqrt{m}} \leq a\right\} \geq -\frac{\rho}{2\sigma^3\sqrt{m}}.$$

Per tant, tenim que, amb probabilitat almenys $1 - \delta/2$,

$$\frac{1}{k} \sum_{j=1}^k \mathbb{1}_{Z_j - \mu \leq a} \geq \mathbb{P}\left\{G \frac{\sigma}{\sqrt{m}} \leq a\right\} - \sqrt{\frac{\log(2/\delta)}{2k}} - \frac{\rho}{2\sigma^3\sqrt{m}}.$$

Així, $(1/k) \sum_{j=1}^k \mathbb{1}_{Z_j - \mu \leq a} \geq \frac{1}{2}$ amb probabilitat almenys $1 - \delta/2$, sempre que a sigui tal que

$$\mathbb{P}\left\{G \leq a \frac{\sqrt{m}}{\sigma}\right\} \geq \frac{1}{2} + \sqrt{\frac{\log(2/\delta)}{2k}} + \frac{\rho}{2\sigma^3\sqrt{m}}.$$

Si $\sqrt{\frac{\log(2/\delta)}{2k}} + \frac{\rho}{2\sigma^3\sqrt{m}} \leq 1/4$, aleshores n'hi ha prou de considerar valors de a tals que $a\sqrt{m}/\sigma \leq \Phi^{-1}(3/4)$. Llavors,

$$\mathbb{P}\left\{G \leq a \frac{\sqrt{m}}{\sigma}\right\} \geq \frac{1}{2} + c \frac{a\sqrt{m}}{\sigma}$$

amb $c = \phi(\Phi^{-1}(3/4))$. Per tant, podem prendre

$$a = \frac{\sigma}{c\sqrt{m}} \left(\sqrt{\frac{\log(2/\delta)}{2k}} + \frac{\rho}{2\sigma^3\sqrt{m}} \right) = \frac{1}{c} \left(\sigma \sqrt{\frac{\log(2/\delta)}{2n}} + \frac{\rho k}{2\sigma^2 n} \right).$$

El mateix argument mostra que, amb probabilitat almenys $1 - \delta/2$,

$$\frac{1}{k} \sum_{j=1}^k \mathbb{1}_{Z_j - \mu \geq -a} \geq \frac{1}{2}$$

per l'elecció de a que hem fet més amunt. □

2.2 L'estimador de Catoni

A continuació, presentem un enfocament completament diferent per construir un estimador de mitjanes, introduït i analitzat per Catoni [8]. Per introduir la idea de Catoni, observem primer que la mitjana empírica $\bar{\mu}_n$ és precisament la solució $\gamma \in \mathbb{R}$ de l'equació

$$\sum_{i=1}^n (X_i - \gamma) = 0.$$

Catoni va proposar substituir el costat esquerre de l'equació anterior per una altra funció estrictament decreixent de γ de la forma

$$R_{n,\alpha}(\gamma) = \sum_{i=1}^n \psi(\alpha(X_i - \gamma)),$$

on $\psi: \mathbb{R} \rightarrow \mathbb{R}$ és una funció creixent antisimètrica i $\alpha \in \mathbb{R}$ és un paràmetre. La idea és que, si $\psi(x)$ augmenta molt més lentament que x , llavors es redueix l'efecte de les dades atípiques presents a causa de les cues pesades. Catoni ofereix tot un ventall de *funcions d'influència* ψ . Per facilitar l'exposició, triem una funció concreta, com ara

$$\psi(x) = \begin{cases} \log(1 + x + x^2/2) & \text{if } x \geq 0, \\ -\log(1 - x + x^2/2) & \text{if } x < 0. \end{cases}$$

Definim l'estimador de la mitjana de Catoni $\hat{\mu}_{\alpha,n}$ com el valor únic y tal que $R_{n,\alpha}(y) = 0$ amb aquesta elecció de ψ . Com que $\psi(x) \leq \log(1 + x + x^2/2)$ per a tot $x \in \mathbb{R}$, tenim que, per a tot $y \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}[e^{R_{n,\alpha}(y)}] &\leq \left(\mathbb{E} \left[1 + \alpha(X - y) + \frac{\alpha^2(X - y)^2}{2} \right] \right)^n = \\ &= \left(1 + \alpha(\mu - y) + \frac{\alpha^2(\sigma^2 + (\mu - y)^2)}{2} \right)^n \leq \\ &\leq \exp \left(n\alpha(\mu - y) + \frac{n\alpha^2(\sigma^2 + (\mu - y)^2)}{2} \right), \end{aligned}$$

sempre que els X_i tinguin una variància finita σ^2 . Així, per la desigualtat de Markov, tenim que, per a qualsevol $y \in \mathbb{R}$ fix i $\delta \in (0, 1)$,

$$\mathbb{P} \left\{ R_{n,\alpha}(y) \geq n\alpha(\mu - y) + \frac{n\alpha^2(\sigma^2 + (\mu - y)^2)}{2} + \log(1/\delta) \right\} \leq \delta.$$

Suposem que el paràmetre α és tal que $\alpha^2\sigma^2 + 2\log(1/\delta)/n \leq 1$. Aleshores el polinomi quadràtic en y

$$n\alpha(\mu - y) + \frac{n\alpha^2(\sigma^2 + (\mu - y)^2)}{2} + \log(1/\delta)$$

té almenys una arrel. En particular, prenent l'arrel menor

$$y_+ = \mu + \frac{\frac{\alpha\sigma^2}{2} + \frac{\log(1/\delta)}{n\alpha}}{\frac{1}{2} + \frac{1}{2}\sqrt{1 - \alpha^2\sigma^2 - \frac{2\log(1/\delta)}{n}}},$$

tenim que $R_{n,\alpha}(y_+) < 0$ amb probabilitat almenys $1 - \delta$. Com que $R_{n,\alpha}(y)$ és estrictament decreixent, això implica que $\hat{\mu}_{\alpha,n} < y_+$ amb probabilitat almenys $1 - \delta$. Un argument simètric serveix per a demostrar que $\hat{\mu}_{\alpha,n} > y_-$ amb probabilitat almenys $1 - \delta$, on

$$y_- = \mu - \frac{\frac{\alpha\sigma^2}{2} + \frac{\log(1/\delta)}{n\alpha}}{\frac{1}{2} + \frac{1}{2}\sqrt{1 - \alpha^2\sigma^2 - \frac{2\log(1/\delta)}{n}}}.$$

Ara, mitjançant fites senzilles i escollint el paràmetre α per tal d'optimitzar les fites, obtenim la següent estimació del rendiment.

TEOREMA 5. *Siguin $X_1, \dots, X_n$ variables aleatòries independents i distribuïdes de manera idèntica amb mitjana μ i variància σ^2 . Sigui $\delta \in (0, 1)$ tal que $n > 2 \log(1/\delta)$. L'estimador de la mitjana de Catoni $\hat{\mu}_{n,\alpha}$ amb paràmetre*

$$\alpha = \sqrt{\frac{2 \log(1/\delta)}{n \sigma^2 (1 + \frac{2 \log(1/\delta)}{n - 2 \log(1/\delta)})}}$$

satisfà, amb probabilitat almenys $1 - 2\delta$,

$$|\hat{\mu}_{n,\alpha} - \mu| < \sqrt{\frac{2 \sigma^2 \log(1/\delta)}{n - 2 \log(1/\delta)}}. \quad (5)$$

El teorema destaca que, amb un paràmetre α triat de manera escaient, l'estimador de Catoni té un rendiment subgaussià. Cal remarcar que la constant $\sqrt{2}$ és la millor possible. Un desavantatge de l'estimador de Catoni respecte de la mediana de les mitjanes és que —almenys en la forma donada al teorema— depèn de la variància σ^2 . En general, no és realista suposar el coneixement de σ^2 . Si hom substitueix σ^2 a la fórmula de α per una fita superior ν , llavors la fita (5) encara es manté si substituïm σ^2 per ν . En cas que no existeixi cap fita superior bona per a σ^2 , Catoni ([8]) mostra com seleccionar α a partir de les dades que tenen un rendiment gairebé òptim. Huber ([30]) combina l'estimador mediana de les mitjanes amb l'estimador de Catoni en un procediment de dos passos i obté un estimador amb la constant del terme dominant òptima en la fita subgaussiana quan $|\sigma/\mu|$ està fitada per una constant coneguda.

Un altre problema —compartit amb l'estimador mediana de les mitjanes— és que l'estimador de Catoni també depèn del nivell de confiança requerit δ . Aquesta dependència, tal com s'ha esmentat anteriorment, és necessària. Una solució ràpida és utilitzar l'estimador amb un paràmetre independent de δ , tot i que l'estimació resultant, naturalment, no podrà ser subgaussiana. Una opció raonable és $\alpha = \sqrt{2/(n\sigma^2)}$. En aquest cas, és fàcil veure que, sempre que $n > 2(1 + \log(1/\delta))$, l'estimador de Catoni satisfà, amb probabilitat almenys $1 - 2\delta$,

$$|\hat{\mu}_{n,\alpha} - \mu| < \sqrt{\frac{\sigma^2}{2n}} \cdot \frac{1 + \log(1/\delta)}{1 - \frac{1 + \log(1/\delta)}{n}}.$$

A causa d'un factor addicional $\sqrt{\log(1/\delta)}$, aquesta fita no és subgaussiana, però les probabilitats «subexponencials» de les cues són encara no trivials i útils.

2.3 Mitjana retallada

L'intent més natural de millorar el rendiment de la mitjana empírica és potser el d'eliminar possibles valors atípics mitjançant un truncament de X . De fet, l'estimador anomenat *mitjana retallada* (o *mitjana truncada*) es defineix eliminant una fracció de la mostra, que consisteix en els ϵn punts més grans i més petits per a algun paràmetre $\epsilon \in (0, 1)$, i després fer la mitjana sobre la resta.

Aquesta idea és una de les eines més clàssiques en estadística robusta i remetem a Tukey i McLaughlin [60], Huber i Ronchetti [32], Bickel [3] i Stigler [58] per consultar els primers treballs sobre les propietats teòriques de l'estimador mitjana retallada. No obstant això, fa ben poc que es va introduir la propietat subgaussiana no asimptòtica de la mitjana retallada. En efecte, Oliveira i Orensstein ([54]) van demostrar que, si s'escull ϵ proporcionalment a $\log(1/\delta)/n$, aleshores l'estimador mitjana retallada té un rendiment subgaussià per a totes les distribucions amb una variància finita.

Per mostrar de la manera més senzilla com funciona aquesta idea, a continuació analitzem una variant senzilla de l'estimador mitjana retallada.

L'estimador divideix les dades en dues parts iguals. Una meitat s'utilitza per a determinar el nivell de truncament correcte. Amb els punts de l'altra meitat se'n calcula la mitjana, sense tenir en compte les dades que queden fora de la regió de truncament. Per comoditat de notació, suposem que les dades consisteixen en $2n$ còpies independents de la variable aleatòria X , denotades per $X_1, \dots, X_n, Y_1, \dots, Y_n$.

Per a $\alpha \leq \beta$, definim la funció de truncament

$$\phi_{\alpha,\beta}(x) = \begin{cases} \beta & \text{si } x > \beta, \\ x & \text{si } x \in [\alpha, \beta], \\ \alpha & \text{si } x < \alpha, \end{cases}$$

i per a $x_1, \dots, x_m \in \mathbb{R}$ sigui $x_1^* \leq x_2^* \leq \dots \leq x_m^*$ la seva reordenació no decreixent.

Un cop establerta aquesta notació, la definició de l'estimador és la següent:

- (1) Donat el nivell de confiança $\delta \geq 8e^{-3n/16}$, definim

$$\epsilon = \frac{16 \log(8/\delta)}{3n}.$$

- (2) Siguin $\alpha = Y_{\epsilon n}^*$ i $\beta = Y_{(1-\epsilon)n}^*$ (assumint, per simplificar, que ϵn és un enter) i definim

$$\hat{\mu}_{2n} = \frac{1}{n} \sum_{i=1}^n \phi_{\alpha,\beta}(X_i).$$

TEOREMA 6. *Siguin $X_1, \dots, X_n, Y_1, \dots, Y_n$ variables aleatòries independents i distribuïdes de manera idèntica amb mitjana μ i variància σ^2 . Sigi $\delta \in (0, 1)$ tal que $n > (16/3) \log(8/\delta)$. Aleshores, amb probabilitat almenys $1 - \delta$,*

$$|\hat{\mu}_{2n} - \mu| \leq 9\sigma \sqrt{\frac{\log(8/\delta)}{n}}.$$

PROVA. Comencem demostrant que el nivell de truncament és proper als quantils adequats de la distribució. Amb aquesta finalitat, per a $p \in (0, 1)$, introduïm els quantils

$$Q_p = \sup\{M \in \mathbb{R} : \mathbb{P}\{X \geq M\} \geq 1 - p\}.$$

Per facilitar l'exposició, suposem que X té una distribució no atòmica. (Aquesta suposició no és necessària, però simplifica la notació.) En aquest cas, $\mathbb{P}\{X > Q_p\} = \mathbb{P}\{X \geq Q_p\} = 1 - p$.

Mitjançant una aplicació senzilla d'una desigualtat de cua binomial (en particular, la desigualtat de Bernstein), amb probabilitat almenys $1 - 2 \exp(-(3/16)\varepsilon n)$, tenim tant que

$$|\{i \in [n] : Y_i \geq Q_{1-2\varepsilon}\}| \geq \varepsilon n$$

com que

$$|\{i \in [n] : Y_i \leq Q_{1-\varepsilon/2}\}| \geq (1 - \varepsilon)n.$$

Això implica que, amb probabilitat almenys $1 - 2 \exp(-(3/16)\varepsilon n)$,

$$Q_{1-2\varepsilon} \leq Y_{(1-\varepsilon)n}^* \leq Q_{1-\varepsilon/2}. \quad (6)$$

Seguint el mateix argument, amb probabilitat almenys $1 - 2 \exp(-(3/16)\varepsilon n)$,

$$Q_{\varepsilon/2} \leq Y_{\varepsilon n}^* \leq Q_{2\varepsilon}. \quad (7)$$

A partir d'aquí, simplement necessitem provar que $|\mathbb{E}\phi_{\alpha,\beta}(X) - \mu|$ és petit i que $(1/n) \sum_{i=1}^n \phi_{\alpha,\beta}(X_i)$ es concentra al voltant de la seva mitjana.

Per al primer pas, considerem l'esdeveniment E consistent en el fet que tant (6) com (7) són certes. Aquest esdeveniment té una probabilitat d'almenys $1 - 4 \exp(-(3/16)\varepsilon n) = 1 - \delta/2$. En l'esdeveniment E ,

$$\begin{aligned} |\mathbb{E}[\phi_{\alpha,\beta}(X) | Y_1, \dots, Y_n] - \mu| &\leq \\ &\leq |\mathbb{E}[(X - \alpha) \mathbb{1}_{X \leq \alpha} | Y_1, \dots, Y_n]| + |\mathbb{E}[(X - \beta) \mathbb{1}_{X \geq \beta} | Y_1, \dots, Y_n]| \leq \\ &\leq |\mathbb{E}(X - Q_{2\varepsilon}) \mathbb{1}_{X \leq Q_{2\varepsilon}}| + |\mathbb{E}(X - Q_{1-2\varepsilon}) \mathbb{1}_{X \geq Q_{1-2\varepsilon}}|. \end{aligned}$$

Per a relacionar aquests dos termes, cal tenir en compte que, per la desigualtat de Txebychev,

$$2\varepsilon = \mathbb{P}\{X \geq Q_{1-2\varepsilon}\} \leq \frac{\sigma_X^2}{(Q_{1-2\varepsilon} - \mu)^2},$$

i, en particular,

$$Q_{1-2\varepsilon} \leq \mu + \frac{\sigma}{\sqrt{2\varepsilon}}.$$

Per tant, per la desigualtat de Cauchy-Schwarz,

$$\begin{aligned} |\mathbb{E}(X - Q_{1-2\varepsilon}) \mathbb{1}_{X \geq Q_{1-2\varepsilon}}| &= |\mathbb{E}(X - \mu) - (Q_{1-2\varepsilon} - \mu) \mathbb{1}_{X \geq Q_{1-2\varepsilon}}| \leq \\ &\leq \mathbb{E}|(X - \mu) \mathbb{1}_{X \geq Q_{1-2\varepsilon}}| + (Q_{1-2\varepsilon} - \mu) \mathbb{P}\{X \geq Q_{1-2\varepsilon}\} \leq \\ &\leq \sigma \sqrt{\mathbb{P}\{X \geq Q_{1-2\varepsilon}\}} + 2\varepsilon(Q_{1-2\varepsilon} - \mu) \leq \\ &\leq \sigma \sqrt{8\varepsilon}. \end{aligned}$$

Un argument simètric prova que $|\mathbb{E}(X - Q_{2\varepsilon}) \mathbb{1}_{X \leq Q_{2\varepsilon}}| \leq \sigma \sqrt{8\varepsilon}$, i per tant, en l'esdeveniment E , tenim que

$$|\mathbb{E}[\phi_{\alpha,\beta}(X)|Y_1, \dots, Y_n] - \mu| \leq \sigma \sqrt{32\varepsilon} \leq 6\sigma \sqrt{\frac{\log(8/\delta)}{n}}$$

per a la nostra elecció de ε . A continuació, considerem

$$Z = \frac{1}{n} \sum_{i=1}^n \phi_{\alpha,\beta}(X_i) - \mathbb{E}[\phi_{\alpha,\beta}(X)|Y_1, \dots, Y_n]$$

i observem que

$$Z = \frac{1}{n} \sum_{i=1}^n \phi_{\alpha-\mu, \beta-\mu}(X_i - \mu) - \mathbb{E}[\phi_{\alpha-\mu, \beta-\mu}(X - \mu)|Y_1, \dots, Y_n].$$

Per tant, en l'esdeveniment E (que només depèn de $Y_1, \dots, Y_n$), Z és una mitjana de variables aleatòries centrades que està fitada punt a punt per $M = \max\{|Q_{\varepsilon/2} - \mu|, |Q_{1-\varepsilon/2} - \mu|\} \leq \sigma \sqrt{2/\varepsilon}$ i la variància de la qual és com a màxim σ^2 . Per tant, per la desigualtat de Bernstein, amb probabilitat almenys $1 - \delta/2$,

$$Z \leq \sigma \sqrt{\frac{2 \log(2/\delta)}{n}} + \frac{\log(2/\delta) \sigma \sqrt{2/\varepsilon}}{n} \leq 3\sigma \sqrt{\frac{\log(2/\delta)}{n}}.$$

Ajuntant totes les peces, obtenim la fita anunciada. $\square$

A més de la seva senzillesa conceptual, un avantatge important de la mitjana retallada en comparació amb altres estimadors amb rendiment subgaussià és que és robusta a la contaminació antagonista de les dades. Aquesta propietat es va formalitzar i demostrar a [47], on també se'n presenta i analitza una extensió multivariant.

Tanquem aquesta secció observant que, en un treball recent, Lee i Valiant ([41]) construeixen un estimador subgaussià amb la constant (gairebé) òptima $L = \sqrt{2} + o(1)$. El seu estimador es basa en una combinació intel·ligent de mediana de les mitjanes, mitjana retallada i estimador de Catoni. Per una altra banda, Minsker i Ndaoud ([51]) van proposar un enfocament diferent. Igual que la mediana de mitjanes, el seu estimador de la mitjana també comença calculant mitjanes empíriques en blocs disjunts de les dades. A continuació, tornen a ponderar les mitjanes del bloc en funció de la seva desviació estàndard empírica. Utilitzant propietats no trivials de sumes autonormalitzades, obtenen un estimador que no només és subgaussià sinó que també és asimptòticament eficient, en el sentit que l'estimador és asimptòticament normal amb una variància asimptòtica tan petita com sigui possible en el sentit minimax.

3 Estimació de la mitjana d'un vector aleatori

A continuació, discutim les extensions del problema d'estimació de la mitjana al cas multivariant. Per tal d'establir el problema, considerem un vector aleatori X que pren valors a $\mathbb{R}^d$. Suposem que el vector de mitjanes $\mu = \mathbb{E}X$ i la matriu de covariància $\Sigma = \mathbb{E}(X - \mu)(X - \mu)^T$ existeixen. Donades n mostres independents i distribuïdes de manera idèntica $X_1, \dots, X_n$ extreptes de la distribució de X , volem estimar el vector de mitjanes.

Igual que en el cas univariant, una elecció natural és la mitjana mostral $\bar{\mu}_n = (1/n) \sum_{i=1}^n X_i$, que té un comportament gairebé òptim sempre que la distribució tingui una cua prou lleugera. Tanmateix, com passa en el cas univariant, sempre que ens preocupin les cues pesades, cal evitar la mitjana mostral ja que pot tenir un rendiment subòptim.

3.1 Rendiment subgaussia

Hem vist diversos exemples d'estimadors de mitjanes amb un rendiment subgaussia. Per tal d'establir correctament el nostre objectiu per al cas d -dimensional, primer hem d'entendre què vol dir «rendiment subgaussia». Com en el cas univariant, hom voldria construir estimadors que fossin «a prop» de la veritable mitjana μ amb «alta probabilitat». A partir d'ara, $\|\cdot\|$ denotarà la norma euclidiana habitual a $\mathbb{R}^d$.

Si X té una distribució normal multivariant amb vector de mitjanes μ i matriu de covariància Σ , llavors la mitjana mostral $\bar{\mu}_n$ també és normal multivariant amb mitjana μ i matriu de covariància $(1/n)\Sigma$. Així, per a tot $t > 0$,

$$\mathbb{P}\{\|\bar{\mu}_n - \mu\| \geq \mathbb{E}\|\bar{\mu}_n - \mu\| + t\} = \mathbb{P}\{\|\bar{X}\| - \mathbb{E}\|\bar{X}\| \geq t\sqrt{n}\},$$

on $\bar{X}$ és un vector gaussià en $\mathbb{R}^d$ amb mitjana zero i matriu de covariància Σ . Una propietat clau dels vectors gaussians és que $\bar{X}$ té la mateixa distribució que $\Sigma^{1/2}Y$, on Y és un vector normal estàndard (és a dir, amb una matriu de covariància igual a la identitat i amb mitjana zero) i $\Sigma^{1/2}$ és l'arrel quadrada semidefinida positiva de Σ . Observem també que per a tots els $\mathcal{Y}, \mathcal{Y}' \in \mathbb{R}^d$,

$$\left| \|\Sigma^{1/2}\mathcal{Y}\| - \|\Sigma^{1/2}\mathcal{Y}'\| \right| \leq \|\Sigma^{1/2}(\mathcal{Y} - \mathcal{Y}')\| \leq \|\Sigma^{1/2}\|_{2 \rightarrow 2} \cdot \|\mathcal{Y} - \mathcal{Y}'\|,$$

on $\|\Sigma^{1/2}\|_{2 \rightarrow 2}$ és la norma espectral de $\Sigma^{1/2}$. Així, $\Sigma^{1/2}\mathcal{Y}$ és una funció de Lipschitz de $\mathcal{Y} \in \mathbb{R}^d$ amb una constant de Lipschitz $\|\Sigma^{1/2}\|_{2 \rightarrow 2} = \sqrt{\lambda_{\max}}$, on $\lambda_{\max} = \lambda_{\max}(\Sigma)$ denota el valor propi més gran de la matriu de covariància Σ . Ara, de la desigualtat de concentració gaussiana de Tsirelson, Ibragimov i Sudakov [14] (vegeu també Ledoux [39] i Boucheron, Lugosi i Massart [6] per a més informació), es dedueix que

$$\mathbb{P}\{\|\bar{X}\| - \mathbb{E}\|\bar{X}\| \geq t\sqrt{n}\} \leq e^{-nt^2/(2\lambda_{\max})}.$$

Observant que

$$\mathbb{E}\|\bar{X}\| \leq \sqrt{\mathbb{E}\|\bar{X}\|^2} = \sqrt{\text{Tr}(\Sigma)},$$

on $\text{Tr}(\Sigma)$ és la traça de la matriu de covariància Σ , tenim que, per a $\delta \in (0, 1)$, amb probabilitat almenys $1 - \delta$,

$$\|\bar{\mu}_n - \mu\| \leq \sqrt{\frac{\text{Tr}(\Sigma)}{n}} + \sqrt{\frac{2\lambda_{\max} \log(1/\delta)}{n}}. \quad (8)$$

Així, en el cas multivariant, direm que un estimador de la mitjana és *subgaussiana* si, amb probabilitat almenys $1 - \delta$, satisfà una desigualtat de la forma (8) (amb factors constants possiblement diferents). Observem que, per a qualsevol distribució amb mitjana μ i matriu de covariància Σ , l'error quadràtic mitjà de la mitjana empírica és igual a

$$\mathbb{E}\|\bar{\mu}_n - \mu\|^2 = \frac{\text{Tr}(\Sigma)}{n}.$$

En particular, $\mathbb{E}\|\bar{\mu}_n - \mu\| \leq \sqrt{\frac{\text{Tr}(\Sigma)}{n}}$. Una característica important de la propietat subgaussiana (8) és que les fluctuacions aleatòries estan controlades per la norma espectral $\lambda_{\max}$ de la matriu de covariància, que possiblement és molt més petita que $\text{Tr}(\Sigma)$, la suma de tots els valors propis de Σ .

3.2 Mediana de mitjanes multivariant

Per a distribucions no gaussianes i possiblement de cua pesada, no es pot esperar un comportament subgaussiana de la mitjana mostral semblant a (8).

Com a alternativa, es pot intentar estendre l'estimador mediana de mitjanes al cas multivariant. Una idea òbvia és dividir les dades en blocs disjunts, calcular la mitjana empírica dins de cada bloc i calcular una mediana multivariant de les mitjanes empíriques obtingudes. Tanmateix, no hi ha una noció estàndard de mediana per a dades multivariants, i no està del tot clar quina definició de mediana multivariant funciona millor per als estimadors medianes de mitjanes. Entre les nombroses possibilitats, esmentem la *mediana per coordenades*, la *mediana geomètrica (o espacial)*, la *mediana de Tukey (o de semiespai)*, la *mediana d'Oja* i la *mediana de Liu*; vegeu la revisió de Small [57], que conté, a més, referències rellevants.

Independentment de quina noció de mediana multivariant decidim adoptar, comencem dividint $[n] = \{1, \dots, n\}$ en k blocs $B_1, \dots, B_k$, cadascun dels quals de mida $|B_i| \geq \lfloor n/k \rfloor \geq 2$. Aquí, k és un paràmetre de l'estimador que es triarà més endavant. Per simplificar, suposem que $km = n$ per a algun enter positiu m . Com abans, calculem la mitjana mostral dels vectors aleatoris dins de cada bloc: per a $j = 1, \dots, k$, considerem

$$Z_j = \frac{1}{m} \sum_{i \in B_j} X_i.$$

El primer intent més natural és, possiblement, definir $\hat{\mu}_n$ com el vector de les medianes coordinada a coordinada de les Z_j (és a dir, la ℓ -èsima component del vector $\hat{\mu}_n$ és la mediana de les ℓ -èsimes components de $Z_1, \dots, Z_k$, per a $\ell \in [d]$). Aleshores, el teorema 2 i la fita de la unió (desigualtat de Boole) impliquen que, per a qualsevol $\delta \in (0, 1)$, prenent $k = \lceil 8 \log(1/\delta) \rceil$, amb probabilitat almenys $1 - \delta$,

$$\|\hat{\mu}_n - \mu\| \leq \sqrt{\frac{32 \operatorname{Tr}(\Sigma) \log(d/\delta)}{n}},$$

on hem utilitzat el fet que $\operatorname{Tr}(\Sigma) = \mathbb{E}\|X - \mathbb{E}X\|^2$ és la suma de les variàncies de les d components de X . Clarament, aquesta fita queda lluny de la desigualtat subgaussiana (8) per diverses raons. En primer lloc, no és independent de la dimensió ja que d hi apareix explícitament. Segurament, però, és més important encara que $\log(d/\delta)$ ve multiplicat per $\operatorname{Tr}(\Sigma)$ en lloc de $\lambda_{\max}(\Sigma)$, i això pot ser una diferència rellevant en problemes d'alta dimensió, especialment quan hom està interessat en probabilitats d'error petites. Un exemple il·lustratiu és quan tots els valors propis de Σ són idèntics i iguals a $\lambda_{\max}$. Si la dimensió d és gran, l'expressió (8) és de l'ordre de $\sqrt{(\lambda_{\max}/n)(d + \log(1/\delta))}$ mentre que la fita anterior només dona l'ordre $\sqrt{(\lambda_{\max}/n)(d \log(d/\delta))}$.

De tota manera, això es pot millorar amb força facilitat amb una noció diferent (no estàndard) de mediana a la definició de l'estimació: triem $\hat{\mu}_n$ com el punt de $\mathbb{R}^d$ que té la propietat que la bola euclidiana centrada en $\hat{\mu}_n$ que conté més de $k/2$ dels punts Z_j té un radi mínim. Com que $\mathbb{E}\|Z_j - \mu\|^2 = \operatorname{Tr}(\Sigma)/m$, per la desigualtat de Txebeixev, $\|Z_j - \mu\| \leq r := 2\sqrt{\operatorname{Tr}(\Sigma)/m}$ amb una probabilitat d'almenys $3/4$. Així, escollint $k = \lceil 8 \log(1/\delta) \rceil$, tenim que, amb probabilitat almenys $1 - \delta$, més de la meitat dels punts Z_j compleixen

$$\|Z_j - \mu\| \leq r.$$

Denotem ara aquest esdeveniment amb E (així, $\mathbb{P}\{E\} \geq 1 - \delta$). En l'esdeveniment E , aquest radi és com a màxim r . Per tant, almenys un dels Z_j es troba a la distància r tant de μ com de $\hat{\mu}_n$ i, per la desigualtat triangular, $\|\hat{\mu}_n - \mu\| \leq 2r$. Hem obtingut, doncs, la proposició següent.

PROPOSICIÓ 7. *Siguin $X_1, \dots, X_n$ vectors aleatoris i.i.d. en $\mathbb{R}^d$ amb mitjana μ i matriu de covariància Σ . Sigui $\delta \in (0, 1)$ i sigui $\hat{\mu}_n$ l'estimador definit anteriorment amb $k = \lceil 8 \log(1/\delta) \rceil$. Aleshores, amb probabilitat almenys $1 - \delta$,*

$$\|\hat{\mu}_n - \mu\| \leq 4\sqrt{\frac{\operatorname{Tr}(\Sigma)(8 \log(1/\delta) + 1)}{n}}.$$

La fita de la proposició 7 és remarcable ja que no depèn de la dimensió i no es fa cap hipòtesi que no sigui l'existència de la matriu de covariància. Tanmateix, encara no assoleix una fita de rendiment subgaussiana que s'assembli a (8). A més, la noció de mediana utilitzada aquí (és a dir, el centre de la bola més petita que conté almenys la meitat dels punts) és problemàtica des del

punt de vista computacional, ja que calcular aquesta mediana és un problema no trivial.

Una versió calculable de manera eficient d'una mediana multivariant és l'anomenada *mediana geomètrica*, definida com a

$$\hat{\mu}_n = \operatorname{argmin}_{m \in \mathbb{R}^d} \sum_{i=1}^k \|Z_i - m\|.$$

Aquest estimador va ser proposat per Minsker [50] i independentment per Hsu i Sabato [29] (vegeu també Lerasle i Oliveira [43]). En particular, Minsker ([50]) va demostrar que aquesta versió multivariant de l'estimador mediana de les mitjanes assoleix un rendiment semblant a la proposició 7. A més, calcular la mediana geomètrica i, per tant, l'estimador mediana de les mitjanes multivariant, implica resoldre un problema d'optimització convex. Així, la mediana geomètrica es pot aproximar de manera eficient; vegeu Cohen, Lee, Miller, Pachocki i Sidford [15] per al resultat més recent i per a la rica història del problema. Remetem el lector a Aloupis [2] per a una revisió dels aspectes computacionals d'altres nocions diverses de medianes multivariants.

Per a un estimador de la mitjana molt diferent, basat en la idea de la mediana de les mitjanes amb garanties «gairebé» subgaussianes però amb un cost computacional important, vegeu Joly, Lugosi i Oliveira [34].

Per tal d'aconseguir un rendiment realment subgaussià, hem de definir un nou estimador. A continuació, en definim un que aconsegueix el rendiment desitjat: aquest estimador es va introduir a [46], basat en la idea de torneigs de mediana de mitjanes. També esmentem un altre estimador, proposat a [47], que es defineix a partir d'una extensió multivariant de l'estimador mitjana retallada. Tanmateix, abans de presentar aquestes estimacions, recordem un estimador molt diferent introduït per Catoni i Giulini [10].

3.3 Llindar de la norma: l'estimador de Catoni-Giulini

En aquesta secció discutim breument un estimador notablement senzill, suggerit i analitzat per Catoni i Giulini [10]. L'estimador de Catoni-Giulini és

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n X_i \min \left(1, \frac{1}{\alpha \|X_i\|} \right), \quad (9)$$

on $\alpha > 0$ és un paràmetre (petit). Fixem-nos que $\hat{\mu}_n$ és simplement una mitjana empírica dels X_i , on els punts que tenen una norma gran es fan tendir a zero. Aquesta estimació és trivial de calcular, a diferència dels estimadors més complexos que discutirem a la subsecció 3.4. D'altra banda, val a dir que fer tendir a zero és una mica arbitrari i antinatural. De fet, l'estimador no és invariant per translacions de les dades en el sentit que $\hat{\mu}_n(X_1 + a, \dots, X_n + a)$ no és necessàriament igual a $\hat{\mu}_n(X_1, \dots, X_n) + a$ quan $a \neq 0$.

Catoni i Giulini demostren que, si es tria el paràmetre com a

$$\alpha = \sqrt{\frac{c \log(1/\delta)}{vn}},$$

on $v \geq \lambda_{\max}$ i $c > 0$ és una constant numèrica, aleshores l'estimador (9) compleix, amb probabilitat almenys $1 - \delta$,

$$\|\hat{\mu}_n - \mu\| \leq C \sqrt{\frac{(\text{Tr}(\Sigma) + v + \|\mu\|^2) \log(1/\delta)}{n}},$$

on C és una constant que només depèn de c . Aquesta fita és similar però més feble que la de la proposició 7, principalment a causa de dos fets. En primer lloc, l'estimador requereix un coneixement previ (una bona fita superior) de $\lambda_{\max}$, mentre que l'estimador mediana geomètrica de les mitjanes no assumeix aquesta informació prèvia. En segon lloc, $\|\mu\|^2$ apareix a la fita superior que, *a priori*, pot ser arbitràriament gran en comparació amb $\text{Tr}(\Sigma)$. La presència d'aquest terme es deu a la manca d'invariància per translacions de l'estimador. Aquest segon problema es pot solucionar mitjançant la definició d'un estimador de dues etapes: primer es pot utilitzar un estimador invariant per translacions com la mediana geomètrica de mitjanes definit a la secció anterior per tal d'obtenir una estimació aproximada de la mitjana. Aleshores, utilitzant un nou conjunt de dades independents, es poden centrar les dades a la mitjana estimada i després utilitzar l'estimador de Catoni-Giulini per a les dades centrades. Aquest nou estimador és invariant per translacions, i el terme $\|\mu\|^2$ es pot substituir per l'error quadràtic de l'estimador del primer pas, és a dir, per $\text{Tr}(\Sigma) \log(1/\delta)/n$. Però, fins i tot amb aquesta modificació, la fita no és subgaussiana en el sentit de (8).

Tanmateix, és remarcable que el rendiment de l'estimador de Catoni-Giulini s'acosti a ser subgaussiana en el sentit desitjat amb només una petita hipòtesi addicional. En particular, si $\mathbb{E}\|X\|^\beta < \infty$ per a alguns $\beta > 2$, llavors, amb la mateixa elecció de α que l'anterior, tenim que

$$\|\hat{\mu}_n - \mu\| \leq C \left(\sqrt{\frac{v \log(1/\delta)}{n}} + \sqrt{\frac{(\text{Tr}(\Sigma) + v)}{n}} + \frac{\kappa_\beta}{n^{(\beta-1)/2}} \right),$$

on κ_β és una constant que depèn de β i del β -èsim moment de $\|X\|$. Així, si el paràmetre anterior v és una bona estimació de $\lambda_{\max}$ en el sentit que està fitat per un múltiple constant, aleshores els dos primers termes de la fita són de la forma subgaussiana desitjada. El tercer terme és d'ordre més petit, encara que, novament, pot ser arbitràriament gran si la mitjana està lluny de l'origen, cosa que es pot solucionar fent l'estimador més complex. Remetem a Catoni i Giulini [9] per a altres estimacions amb un esperit similar i per a més discussió sobre la qüestió. Les tècniques de demostració de [9, 10] es basen en les anomenades desigualtats *PAC-bayesianes*, els detalls de les quals sobrepassen l'abast d'aquesta revisió.

3.4 Torneigs de mediana de mitjanes

En aquesta secció introduïm un estimador de la mitjana, proposat per Lugosi i Mendelson [46], amb un rendiment subgaussià per a totes les distribucions la matriu de covariància de les quals existeixi. L'estimador que es presenta a continuació és l'exemple primari i més senzill del que anomenem *torneigs de mediana de mitjanes*.

Recordem que partim d'una mostra $X_1, \dots, X_n$ de vectors aleatoris i.i.d. a $\mathbb{R}^d$. Com en el cas de l'estimador mediana de les mitjanes, comencem dividint el conjunt $\{1, \dots, n\}$ en k blocs $B_1, \dots, B_k$, cadascun dels quals de mida $|B_j| \geq m = \lfloor n/k \rfloor$, on k és un paràmetre de l'estimador el valor del qual depèn del nivell de confiança desitjat, tal com s'especifica a continuació. Per simplificar la presentació, suposem que n és divisible per k i, per tant, $|B_j| = m$ per a tot $j = 1, \dots, k$.

Definim la mitjana mostral dins de cada bloc com a

$$Z_j = \frac{1}{m} \sum_{i \in B_j} X_i.$$

Per a cada $a \in \mathbb{R}^d$, sigui

$$T_a = \{x \in \mathbb{R}^d : \exists J \subset [k] \text{ tal que } |J| \geq k/2 \text{ i, per a tot } j \in J, \|Z_j - x\| \leq \|Z_j - a\|\} \quad (10)$$

i definim l'estimador de la mitjana com a

$$\hat{\mu}_n \in \operatorname{argmin}_{a \in \mathbb{R}^d} \operatorname{radi}(T_a),$$

on $\operatorname{radi}(T_a) = \sup_{x \in T_a} \|x - a\|$. Així, $\hat{\mu}_n$ es tria per a minimitzar, sobre tot $a \in \mathbb{R}^d$, el radi del conjunt T_a definit com el conjunt de punts $x \in \mathbb{R}^d$ per als quals $\|Z_j - x\| \leq \|Z_j - a\|$ per a la majoria dels blocs. Si hi ha diversos minimitzadors, se'n pot triar un qualsevol.

El conjunt T_a es pot veure com el conjunt de punts de $\mathbb{R}^d$ que estan almenys tan a prop del núvol de punts $\{Z_1, \dots, Z_k\}$ com el punt a . L'estimador $\hat{\mu}_n$ s'obté minimitzant el radi de T_a .

Observem que sempre s'aconsegueix el mínim. Això es deu al fet que $\operatorname{radi}(T_a)$ és una funció contínua de a (ja que, per a cada a , T_a és la intersecció d'una unió finita de boles tancades, i els centres i radis de les boles tancades són contínues en a).

Es pot interpretar $\operatorname{argmin}_{a \in \mathbb{R}^d} \operatorname{radi}(T_a)$ com una altra noció multivariant de la mediana de $Z_1, \dots, Z_k$. De fet, quan $d = 1$, és una elecció particular de la mediana i l'estimador coincideix amb l'estimador mediana de les mitjanes.

La fita de rendiment següent mostra que l'estimador té el rendiment subgaussià desitjat.

TEOREMA 8 (LUGOSI I MENDELSON [46]). *Segui $\delta \in (0, 1)$ i considerem l'estimador de la mitjana $\hat{\mu}_n$ amb el paràmetre $k = \lceil 200 \log(2/\delta) \rceil$. Si $X_1, \dots, X_n$ són vectors aleatoris i.i.d. a $\mathbb{R}^d$ amb mitjana $\mu \in \mathbb{R}^d$ i matriu de covariància Σ , aleshores per a tot n , amb probabilitat almenys $1 - \delta$,*

$$\|\hat{\mu}_n - \mu\| \leq \max\left(960\sqrt{\frac{\text{Tr}(\Sigma)}{n}}, 240\sqrt{\frac{\lambda_{\max} \log(2/\delta)}{n}}\right).$$

Igual que la fita de rendiment de la proposició 7, el teorema 8 és «infinito-dimensional» en el sentit que la fita no depèn explícitament de la dimensió d . De fet, el mateix estimador es pot definir per a vectors aleatoris que prenguin valors en espais de Hilbert i el teorema 8 continua sent vàlid mentre $\text{Tr}(\Sigma) = \mathbb{E}\|X - \mu\|^2$ sigui finit.

El teorema 8 és després del resultat següent.

TEOREMA 9. *Utilitzant la notació prèvia i fixant*

$$r = \max\left(960\sqrt{\frac{\text{Tr}(\Sigma)}{n}}, 240\sqrt{\frac{\lambda_{\max} \log(2/\delta)}{n}}\right),$$

amb probabilitat almenys $1 - \delta$, per a qualsevol $a \in \mathbb{R}^d$ tal que $\|a - \mu\| \geq r$, es té $\|Z_j - a\| > \|Z_j - \mu\|$ per a més de $k/2$ índexs j . En altres paraules, $\|a - \mu\| \geq r$ implica que $a \notin T_\mu$.

El teorema 9 implica que per a una col·lecció «típica» $X_1, \dots, X_n$, μ és més proper a la majoria dels Z_j en comparació amb qualsevol $a \in \mathbb{R}^d$ que estigui prou lluny de μ . Òbviament, per a una col·lecció arbitrària $x_1, \dots, x_n \subset \mathbb{R}^d$, ni tan sols cal que existeixi aquest punt, i és sorprenent que per a una configuració típica i.i.d. μ satisfaci aquesta propietat.

El fet que el teorema 9 impliqui el teorema 8 és senzill. De fet, la definició de $\hat{\mu}_n$ i el teorema 9 impliquen que, amb probabilitat almenys $1 - \delta$, $\text{radi}(T_{\hat{\mu}_n}) \leq \text{radi}(T_\mu) \leq r$. Com que, o bé $\mu \in T_{\hat{\mu}_n}$ o bé $\hat{\mu} \in T_\mu$, llavors seguir que $\|\hat{\mu}_n - \mu\| \leq r$, tal com es demana.

Certament, les constants que apareixen al teorema 8 no són òptimes; es van obtenir amb l'objectiu de fer la demostració més transparent.

La demostració del teorema 9 es basa en la idea següent. La mitjana μ és el minimitzador de la funció $f(x) = \mathbb{E}\|X - x\|^2$. Un enfocament possible és utilitzar les dades disponibles per a endevinar, per a qualsevol parell $a, b \in \mathbb{R}^d$, si $f(a) < f(b)$. Una opció natural és fer servir un estimador mediana de mitjanes per tal de decidir quin dels dos és millor. El «torneig» és simplement una manera de comparar cada parell, tal com es descriu a continuació.¹

¹ Tal com expliquem a continuació, n'hi ha prou d'assegurar-se que la comparació és correcta entre μ i qualsevol punt que no estigui massa a prop de μ .

Per definir el torneig, recordem que $[n]$ està dividit en k blocs disjunts $B_1, \dots, B_k$ de mida $m = n/k$. Per a $a, b \in \mathbb{R}^d$, diem que a *derrota* b si

$$\frac{1}{m} \sum_{i \in B_j} (\|X_i - b\|^2 - \|X_i - a\|^2) > 0 \quad (11)$$

en més de $k/2$ blocs B_j . El principal lema tècnic és el següent.

LEMA 10. *Sigui $\delta \in (0, 1)$, $k = \lceil 200 \log(2/\delta) \rceil$, i definim*

$$r = \max \left(960 \sqrt{\frac{\text{Tr}(\Sigma)}{n}}, 240 \sqrt{\frac{\lambda_{\max} \log(2/\delta)}{n}} \right).$$

Amb una probabilitat almenys $1 - \delta$, μ derrota tots els $b \in \mathbb{R}^d$ de manera que $\|b - \mu\| \geq r$.

El resultat del lema 10 és raonable: si $\|b - \mu\|$ és prou gran, això es reflectirà en els «valors típics» de $(\|X_i - \mu\|)_{i=1}^n$ i $(\|X_i - b\|)_{i=1}^n$. La comparació dels valors mitjançant (11) assegura «estabilitat», i el fet que b estigui lluny de μ es mostra amb alta probabilitat. Subratllem que l'estimació de probabilitat ha de ser *uniforme* en b . Aquestes estimacions uniformes esdevindran un tema recurrent en el que resta d'article.

PROVA. Notem que

$$\|X_i - b\|^2 - \|X_i - \mu\|^2 = \|X_i - \mu + \mu - b\|^2 - \|X_i - \mu\|^2 = -2\langle X_i - \mu, b - \mu \rangle + \|b - \mu\|^2,$$

definim $\bar{X} = X - \mu$ i denotem $v = b - \mu$. Així, per a un b fix que satisfaci $\|b - \mu\| \geq r$, μ derrota b si

$$-\frac{2}{m} \sum_{i \in B_j} \langle \bar{X}_i, v \rangle + \|v\|^2 > 0$$

a la majoria dels blocs B_j .

Per tant, per a demostrar la nostra afirmació necessitem que, amb probabilitat almenys $1 - \delta$, per a cada $v \in \mathbb{R}^d$ amb $\|v\| \geq r$,

$$-\frac{2}{m} \sum_{i \in B_j} \langle \bar{X}_i, v \rangle + \|v\|^2 > 0 \quad (12)$$

en més de $k/2$ blocs B_j . Clarament, n'hi ha prou de demostrar que (12) es compleix quan $\|v\| = r$.

Considerem un $v \in \mathbb{R}^d$ fix amb $\|v\| = r$. Per la desigualtat de Txeixev, amb una probabilitat d'almenys $9/10$,

$$\left| \frac{1}{m} \sum_{i \in B_j} \langle \bar{X}_i, v \rangle \right| \leq \sqrt{10} \sqrt{\frac{\mathbb{E} \langle \bar{X}, v \rangle^2}{m}} \leq \sqrt{10} \|v\| \sqrt{\frac{\lambda_{\max}}{m}},$$

on recordem que $\lambda_{\max}$ és el valor propi més gran de la matriu de covariància Σ de X . Així, si

$$r = \|v\| \geq 4\sqrt{10}\sqrt{\frac{\lambda_{\max}}{m}}, \quad (13)$$

aleshores amb probabilitat almenys $9/10$,

$$-\frac{2}{m} \sum_{i \in B_j} \langle \bar{X}_i, v \rangle \geq \frac{-r^2}{2}. \quad (14)$$

Aplicant la desigualtat de Hoeffding ([24]), obtenim que (14) és certa per a un sol v amb probabilitat almenys $1 - \exp(-k/50)$ en almenys $8/10$ dels blocs B_j .

Ara ho hem d'estendre a tots els vectors amb norma r a partir d'un vector fix v . Per tal de demostrar que (14) és certa simultàniament per a tots els $v \in r \cdot S^{d-1}$ en almenys $7/10$ dels blocs B_j , primer considerem un conjunt $V_1 \subset r \cdot S^{d-1}$ maximal i ϵ -separat respecte a la norma $L_2(X)$. En altres paraules, V_1 és un subconjunt de $r \cdot S^{d-1}$ de cardinalitat maximal tal que per a tots els $v_1, v_2 \in V_1$, $\|v_1 - v_2\|_{L_2(X)} = \langle v_1 - v_2, \Sigma(v_1 - v_2) \rangle^{1/2} \geq \epsilon$. Podem estimar aquesta cardinalitat per la desigualtat «dual de Sudakov» (vegeu Ledoux i Talagrand [40] i també Vershynin [64] per a una versió amb la constant específica utilitzada aquí): la cardinalitat de V_1 està fitada per

$$\log(|V_1|/2) \leq \frac{1}{32} \left(\frac{\mathbb{E}[\langle G, \Sigma G \rangle^{1/2}]}{\epsilon/r} \right)^2,$$

on G és un vector normal estàndard a $\mathbb{R}^d$. Observem que per a qualsevol $a \in \mathbb{R}^d$, $\mathbb{E}_X \langle a, X \rangle^2 = \langle a, \Sigma a \rangle$ i, per tant,

$$\begin{aligned} \mathbb{E}[\langle G, \Sigma G \rangle^{1/2}] &= \mathbb{E}_G[(\mathbb{E}_X[\langle G, \bar{X} \rangle^2])^{1/2}] \leq (\mathbb{E}_X \mathbb{E}_G[\langle G, \bar{X} \rangle^2])^{1/2} = \\ &= (\mathbb{E}[\|\bar{X}\|^2])^{1/2} = \sqrt{\text{Tr}(\Sigma)}. \end{aligned}$$

Per tant, fixant

$$\epsilon = 2r \left(\frac{1}{k} \text{Tr}(\Sigma) \right)^{1/2}, \quad (15)$$

tenim $|V_1| \leq 2e^{k/100}$ i per la fita de la unió, amb probabilitat almenys $1 - 2e^{-k/100} \geq 1 - \delta/2$, (14) és vàlida per a tots els $v \in V_1$ en almenys $8/10$ dels blocs B_j .

A continuació, comprovem que la propietat (12) es compleix simultàniament per a tots els x amb $\|x\| = r$ en almenys $7/10$ dels blocs B_j .

Per a cada $x \in r \cdot S^{d-1}$, sigui v_x l'element més proper a x en V_1 respecte a la norma $L_2(X)$. N'hi ha prou de demostrar que, amb probabilitat almenys $1 - \exp(-k/200) \geq 1 - \delta/2$,

$$\sup_{x \in r \cdot S^{d-1}} \frac{1}{k} \sum_{j=1}^k \mathbb{1}_{\{|m^{-1} \sum_{i \in B_j} \langle \bar{X}_i, x - v_x \rangle| \geq r^2/4\}} \leq \frac{1}{10}. \quad (16)$$

De fet, en aquest esdeveniment es dedueix que per a cada $x \in r \cdot S^{d-1}$, almenys en 7/10 dels blocs B_j , tant

$$-\frac{2}{m} \sum_{i \in B_j} \langle \bar{X}_i, v_x \rangle \geq \frac{-r^2}{2} \quad \text{com} \quad 2 \left| \frac{1}{m} \sum_{i \in B_j} \langle \bar{X}_i, x \rangle - \frac{1}{m} \sum_{i \in B_j} \langle \bar{X}_i, v_x \rangle \right| < \frac{r^2}{2}$$

se satisfan i, per tant, en aquests blocs, $-\frac{2}{m} \sum_{i \in B_j} \langle \bar{X}_i, x \rangle + r^2 > 0$, tal com volíem.

Queda per demostrar (16). Observeu que

$$\frac{1}{k} \sum_{j=1}^k \mathbb{1}_{\{|m^{-1} \sum_{i \in B_j} \langle \bar{X}_i, x - v_x \rangle| \geq r^2/4\}} \leq \frac{4}{r^2} \frac{1}{k} \sum_{j=1}^k \left| \frac{1}{m} \sum_{i \in B_j} \langle \bar{X}_i, x - v_x \rangle \right|.$$

Com que $\|x - v_x\|_{L_2(X)} = \sqrt{\mathbb{E} \langle \bar{X}, x - v_x \rangle^2} \leq \epsilon$, es dedueix que, per a cada j ,

$$\mathbb{E} \left| \frac{1}{m} \sum_{i \in B_j} \langle \bar{X}_i, x - v_x \rangle \right| \leq \sqrt{\frac{\mathbb{E}[\langle \bar{X}, x - v_x \rangle^2]}{m}} \leq \frac{\epsilon}{\sqrt{m}},$$

i, per tant,

$$\begin{aligned} & \mathbb{E} \sup_{x \in r \cdot S^{d-1}} \frac{1}{k} \sum_{j=1}^k \mathbb{1}_{\{|m^{-1} \sum_{i \in B_j} \langle \bar{X}_i, x - v_x \rangle| \geq r^2/4\}} \leq \\ & \leq \underbrace{\frac{4}{r^2} \mathbb{E} \sup_{x \in r \cdot S^{d-1}} \frac{1}{k} \sum_{j=1}^k \left(\left| \frac{1}{m} \sum_{i \in B_j} \langle \bar{X}_i, x - v_x \rangle \right| - \mathbb{E} \left| \frac{1}{m} \sum_{i \in B_j} \langle \bar{X}_i, x - v_x \rangle \right| \right)}_{(A)} + \underbrace{\frac{4\epsilon}{r^2 \sqrt{m}}}_{(B)}. \end{aligned}$$

Per fitar (B), notem que, usant (15),

$$\frac{4\epsilon}{r^2 \sqrt{m}} = 8 \left(\frac{\text{Tr}(\Sigma)}{n} \right)^{1/2} \cdot \frac{1}{r} \leq \frac{1}{60}$$

sempre que

$$r \geq 480 \left(\frac{\text{Tr}(\Sigma)}{n} \right)^{1/2}.$$

Podem fitar (A) mitjançant tècniques estàndard de processos empírics com ara la simetrització, la contracció per a mitjanes de Rademacher i la dessimetrització. En efecte, si $\sigma_1, \dots, \sigma_n$ són variables aleatòries de Rademacher

independents (és a dir, $\mathbb{P}\{\sigma_i = 1\} = \mathbb{P}\{\sigma_i = -1\} = 1/2$), independentment de tots els X_i , aleshores

$$(A) \leq \frac{8}{r^2} \mathbb{E} \sup_{x \in r \cdot S^{d-1}} \frac{1}{k} \sum_{j=1}^k \sigma_j \left| \frac{1}{m} \sum_{i \in B_j} \langle \bar{X}_i, x - v_x \rangle \right| \leq$$

(per una desigualtat de simetrització estàndard;

vegeu, per exemple, [62, Lemma 2.3.6])

$$\leq \frac{8}{r^2} \mathbb{E} \sup_{x \in r \cdot S^{d-1}} \left| \frac{1}{k} \sum_{j=1}^k \sigma_j \frac{1}{m} \sum_{i \in B_j} \langle \bar{X}_i, x - v_x \rangle \right| \leq$$

(per un lema de contracció per a mitjanes de Rademacher; vegeu [40])

$$\leq \frac{16}{r^2} \mathbb{E} \sup_{x \in r \cdot S^{d-1}} \left| \frac{1}{n} \sum_{i=1}^n \langle \bar{X}_i, x - v_x \rangle \right| \leq$$

(vegeu una altra vegada [62, Lemma 2.3.6])

$$\leq \frac{32}{r} \mathbb{E} \sup_{\{t: \|t\| \leq 1\}} \left| \frac{1}{n} \sum_{i=1}^n \langle \bar{X}_i, t \rangle \right| \leq$$

(observem que $\|x - v_x\| \leq 2r$)

$$\leq \frac{32}{r} \cdot \frac{\sqrt{\mathbb{E} \|\bar{X}\|^2}}{\sqrt{n}} = \frac{32}{r} \left(\frac{\text{Tr}(\Sigma)}{n} \right)^{1/2} \leq \frac{1}{30}$$

sempre que $r \geq 960 \left(\frac{\text{Tr}(\Sigma)}{n} \right)^{1/2}$.

Així, per a

$$Y = \sup_{x \in r \cdot S^{d-1}} \frac{1}{k} \sum_{j=1}^k \mathbb{1}_{\{|m^{-1} \sum_{i \in B_j} \langle \bar{X}_i, x - v_x \rangle| \geq r^2/4\}},$$

hem demostrat que $\mathbb{E}Y \leq 1/60 + 1/30 = 1/20$. Finalment, per a obtenir (16), n'hi ha prou de demostrar que $\mathbb{P}\{Y > \mathbb{E}Y + 1/20\} \leq e^{-k/200}$, que es desprèn de la desigualtat de diferències fitades (vegeu, per exemple, [6, teorema 6.2]). $\square$

Demostració del teorema 9 El teorema 9 es dedueix fàcilment del lema 10. Fixem un bloc B_j i recordem que $Z_j = \frac{1}{m} \sum_{i \in B_j} X_i$. Siguin $a, b \in \mathbb{R}^d$. Aleshores,

$$\begin{aligned} \frac{1}{m} \sum_{i \in B_j} (\|X_i - a\|^2 - \|X_i - b\|^2) &= \frac{1}{m} \sum_{i \in B_j} (\|X_i - b - (a - b)\|^2 - \|X_i - b\|^2) = \\ &= -\frac{2}{m} \sum_{i \in B_j} \langle X_i - b, a - b \rangle + \|a - b\|^2 = (*). \end{aligned}$$

Observem que $-\frac{2}{m} \sum_{i \in B_j} \langle X_i - b, a - b \rangle = -2 \langle \frac{1}{m} \sum_{i \in B_j} X_i - b, a - b \rangle = -2 \langle Z_j - b, a - b \rangle$, i, per tant,

$$\begin{aligned} (*) &= -2 \langle Z_j - b, a - b \rangle + \|a - b\|^2 = \\ &= -2 \langle Z_j - b, a - b \rangle + \|a - b\|^2 + \|Z_j - b\|^2 - \|Z_j - b\|^2 = \\ &= \|Z_j - b - (a - b)\|^2 - \|Z_j - b\|^2 = \|Z_j - a\|^2 - \|Z_j - b\|^2. \end{aligned}$$

En conseqüència, $(*) > 0$ (és a dir, b derrota a al bloc B_j) si i només si $\|Z_j - a\| > \|Z_j - b\|$.

Recordem que el lema 10 estableix que, amb probabilitat almenys $1 - \delta$, si $\|a - \mu\| \geq r$, llavors en més de $k/2$ blocs B_j , $\frac{1}{m} \sum_{i \in B_j} (\|X_i - a\|^2 - \|X_i - \mu\|^2) > 0$, que, segons l'argument anterior, és el mateix que dir que almenys per a $k/2$ índexs j , es compleix que $\|Z_j - a\| > \|Z_j - \mu\|$.

Reflexionant-hi, queda clar que les idees utilitzades en la demostració del teorema 8 són força generals. De fet, són al cor del *mètode de la bola petita* introduït per Mendelson [48] (vegeu també [49] per a mètodes de caire similar). El mètode de la bola petita s'aplica en situacions molt més generals que el teorema 8. Per tal d'explicar com se'n pot estendre l'argument, tornem a esbossar els tres passos que ens han permès comparar cada b i μ :

- (1) Per a qualsevol b fixat de $\mathbb{R}^d$, obtenim una fita que es compleix amb una alta probabilitat.
- (2) Aleshores, gràcies a l'estimació amb alta probabilitat de (1), podem invocar la fita de la unió i controlar una col·lecció gran (però encara finita) de punts.

Tenim total llibertat per a triar aquesta col·lecció, i la seleccionem com una xarxa de radi ϵ al conjunt en qüestió.

- (3) La part crucial de l'argument és passar del control que tenim en cada punt de la xarxa al control uniforme desitjat a tota la classe; concretament, mostrem que si un «centre», és a dir, un element de la xarxa, té un bon comportament,² llavors el mateix és cert per a qualsevol punt prou proper al centre. Amb aquesta finalitat, mostrem que les «oscil·lacions aleatòries» no destrueixen el bon comportament d'un centre en gaires blocs.

3.5 Consideracions computacionals

Un tema important que hem passat de llarg fins a aquest punt és la viabilitat computacional dels estimadors de mitjanes. Tot i que la mitjana empírica és trivial de calcular, molts dels estimadors més sofisticats que hem discutit en aquest article estan lluny de ser-ho. En particular, un requisit bàsic perquè qualsevol estimador multivariant de la mitjana sigui útil a la pràctica és que es pugui calcular en temps polinomial (és a dir, en un temps que sigui un polinomi de la mida de la mostra n i de la dimensió d). Com ja hem assenyalat,

² A la demostració del teorema 8, «bon comportament» significa que (12) és vàlida per a la majoria dels blocs.

alguns dels estimadors descrits anteriorment entren en aquesta categoria. Per exemple, l'estimador mediana geomètrica de les mitjanes o l'estimador de Catoni-Giulini són calculables de manera eficient en aquest sentit. No obstant això, aquests estimadors estan lluny de ser subgaussians. L'estimador de torneig de medianes de mitjanes és subgaussà, però el seu càlcul suposa un repte gens trivial. De fet, tal com es defineix l'estimador, és probable que sigui computacionalment intractable (és a dir, NP-complex). No obstant això, en un article recent, Hopkins ([25]) defineix una relaxació semidefinida enginyosa de l'estimador de torneig de medianes de mitjanes que es pot calcular en temps $O(nd + d \log(1/\delta)^c)$ per a una constant independent de la dimensió i , al mateix temps, garanteix la condició subgaussiana desitjada sota l'únic supòsit que existeixi la matriu de covariància. Aquest és el primer estimador de la mitjana multivariant subgaussà calculable de manera eficient. Encara més recentment, Cherapanamjeri, Flammario i Bartlett ([13]) van millorar el temps d'execució a $O(nd + d \log(1/\delta)^2 + \log(1/\delta)^4)$ combinant les idees de Hopkins amb una optimització astuta de gradient descendent. Basant-se en idees de «reponderació espectral» de Cheng, Diakonikolas i Ge [12], Depersin i Lecué [16] i Lei, Luh, Venkat i Zhang [42] milloren encara més el temps d'execució. Hopkins, Li i Zhang ([27]) mostren com la reponderació espectral és essencialment equivalent a la noció de mediana introduïda anteriorment. Remetem a aquests articles per a una revisió exhaustiva de la literatura sobre els aspectes computacionals de l'estimació de la mitjana robusta, que està creixent ràpidament.

A la comunitat de la informàtica teòrica hi ha hagut recentment un augment important de resultats que aborden el problema de l'estimació de la mitjana *robusta* i computacionalment eficient. En aquest context, un estimador es defineix com a robust si funciona bé en presència d'una petita fracció constant de valors atípics (possiblement antagonistes). S'han proposat diversos models; vegeu Charikar, Steinhardt i Valiant [11], Diakonikolas, Kamath, Kane, Li, Moitra i Stewart [18, 19, 20], Diakonikolas, Kane i Stewart [21], Diakonikolas, Kong i Stewart [22], Hopkins i Li [26], Klivans, Kothari i Meka [35], Kothari, Steinhardt i Steurer [37], Lai, Rao i Vempala [38], Loh i Tan [44], per a tenir una mostra d'aquest important gruix de literatura en plena expansió. La revisió d'aquesta àrea, però, va més enllà de l'abast d'aquest article.

Referències

- [1] ALON, N.; MATIAS, Y.; SZEGEDY, M. «The space complexity of approximating the frequency moments». Twenty-eighth Annual ACM Symposium on the Theory of Computing (Philadelphia, PA, 1996). *J. Comput. System Sci.*, 58 (1), part 2 (1999), 137-147.
- [2] ALOUPIS, G. «Geometric measures of data depth». A: *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*. Provi-

- dence, RI: Amer. Math. Soc., 2006, 147–158. (DIMACS Ser. Discrete Math. Theoret. Comput. Sci.; 72)
- [3] BICKEL, P. J. «On some robust estimates of location». *Ann. Math. Statist.*, 36 (1965), 847–858.
- [4] BLACKWELL, D. «Conditional expectation and unbiased sequential estimation». *Ann. Math. Statistics*, 18 (1947), 105–110.
- [5] BLUMER, A.; EHRENFUCHT, A.; HAUSSLER, D.; WARMUTH, M. K. «Learnability and the Vapnik-Chervonenkis dimension». *J. Assoc. Comput. Mach.*, 36 (4) (1989), 929–965.
- [6] BOUCHERON, S.; LUGOSI, G.; MASSART, P. *Concentration Inequalities. A Nonasymptotic Theory of Independence*. Prefaci de Michel Ledoux. Oxford: Oxford University Press, 2013.
- [7] BUBECK, S.; CESA-BIANCHI, N.; LUGOSI, G. «Bandits with heavy tail». *IEEE Trans. Inform. Theory*, 59 (11) (2013), 7711–7717.
- [8] CATONI, O. «Challenging the empirical mean and empirical variance: a deviation study». *Ann. Inst. Henri Poincaré Probab. Stat.*, 48 (4) (2012), 1148–1185.
- [9] CATONI, O.; GIULINI, I. «Dimension-free PAC-Bayesian bounds for matrices, vectors, and linear least squares regression». Preprint (2017). [Disponible en línia a: [arXiv:1712.02747](https://arxiv.org/abs/1712.02747)]
- [10] CATONI, O.; GIULINI, I. «Dimension-free PAC-Bayesian bounds for the estimation of the mean of a random vector». Preprint (2018). [Disponible en línia a: [arXiv:1802.04308](https://arxiv.org/abs/1802.04308)]
- [11] CHARIKAR, M.; STEINHARDT, J.; VALIANT, G. «Learning from untrusted data». A: *STOC'17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*. Nova York: ACM, 2017, 47–60.
- [12] CHENG, Y.; DIAKONIKOLAS, I.; GE, R. «High-dimensional robust mean estimation in nearly-linear time». A: *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. Filadèlfia, PA: SIAM, 2019, 2755–2771.
- [13] CHERAPANAMJERI, Y.; FLAMMARION, N.; BARTLETT, P. L. «Fast mean estimation with sub-Gaussian rates». Preprint (2019). [Disponible en línia a: [arXiv:1902.01998](https://arxiv.org/abs/1902.01998)]
- [14] CIREL'SON, B. S.; IBRAGIMOV, I. A.; SUDAKOV, V. N. «Norms of Gaussian sample functions». A: *Proceedings of the Third Japan-USSR Symposium on Probability Theory*. Berlín: Springer, 1976, 20–41. (Lecture Notes in Math.; 550)
- [15] COHEN, M. B.; LEE, Y. T.; MILLER, G.; PACHOCKI, J.; SIDFORD, A. «Geometric median in nearly linear time». A: *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*. Nova York: ACM, 2016, 9–21.

- [16] DEPERSIN, J.; LECUÉ, G. «Robust subgaussian estimation of a mean vector in nearly linear time». Preprint (2019). [Disponible en línia a: arXiv:1906.03058]
- [17] DEVROYE, L.; LERASLE, M.; LUGOSI, G.; OLIVEIRA, R. I. «Sub-Gaussian mean estimators». *Ann. Statist.*, 44 (6) (2016), 2695–2725.
- [18] DIAKONIKOLAS, I.; KAMATH, G.; KANE, D. M.; LI, J.; MOITRA, A.; STEWART, A. «Robust estimators in high dimensions without the computational intrac-tability». A: *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2016*. Los Alamitos, CA: IEEE Computer Soc., 2016, 655–664.
- [19] DIAKONIKOLAS, I.; KAMATH, G.; KANE, D. M.; LI, J.; MOITRA, A.; STEWART, A. «Being robust (in high dimensions) can be practical». A: *ICML’17: Procee-dings of the 34th International Conference on Machine Learning*. Vol. 70. Sydney, NSW, Austràlia: JMLR.org, 2017, 999–1008.
- [20] DIAKONIKOLAS, I.; KAMATH, G.; KANE, D. M.; LI, J.; MOITRA, A.; STEWART, A. «Robustly learning a Gaussian: getting optimal error, efficiently». A: *Pro-ceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*. Filadèlfia, PA: SIAM, 2018, 2683–2702.
- [21] DIAKONIKOLAS, I.; KANE, D. M.; STEWART, A. «Efficient robust proper learn-ing of log-concave distributions». Preprint (2016). [Disponible en línia a: arXiv:1606.03077]
- [22] DIAKONIKOLAS, I.; KONG, W.; STEWART, A. «Efficient algorithms and lower bounds for robust linear regression». A: *Proceedings of the Thirtieth An-nual ACM-SIAM Symposium on Discrete Algorithms*. Filadèlfia, PA: SIAM, 2019, 2745–2754.
- [23] HAMPEL, F. R.; RONCHETTI, E. M.; ROUSSEEUW, P. J.; STAHEL, W. A. *Robust Statistics. The Approach Based on Influence Functions*. Nova York: John Wiley & Sons, Inc., 1986. (Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics)
- [24] Hoeffding, W. «Probability inequalities for sums of bounded random variables». *J. Amer. Statist. Assoc.*, 58 (1963), 13–30.
- [25] HOPKINS, S. B. «Mean estimation with sub-Gaussian rates in polynomial time». *Ann. Statist.*, 48 (2) (2020), 1193–1213.
- [26] HOPKINS, S. B.; LI, J. «Mixture models, robustness, and sum of squares proofs». A: *STOC’18—Proceedings of the 50th Annual ACM SIGACT Sym-posium on Theory of Computing*. Nova York: ACM, 2018, 1021–1034.
- [27] HOPKINS, S. B.; LI, J.; ZHANG, F. «Robust and heavy-tailed mean estimation made simple, via regret minimization». Preprint (2020). [Disponible en línia a: arXiv:2007.15839]
- [28] HSU, D. «Robust statistics» (2010). [No publicat]
- [29] HSU, D.; SABATO, S. «Loss minimization and parameter estimation with heavy tails». *J. Mach. Learn. Res.*, 17, treball núm. 18 (2016), 40 p.

- [30] HUBER, M. «An optimal (ϵ, δ) -randomized approximation scheme for the mean of random variables with bounded relative variance». *Random Structures Algorithms*, 55 (2) (2019), 356–370.
- [31] HUBER, P. J. «Robust estimation of a location parameter». *Ann. Math. Statist.*, 35 (1964), 73–101.
- [32] HUBER, P. J.; RONCHETTI, E. M. *Robust Statistics*. 2a ed. Hoboken, NJ: John Wiley & Sons, Inc., 2009. (Wiley Series in Probability and Statistics)
- [33] JERRUM, M. R.; VALIANT, L. G.; VAZIRANI, V. V. «Random generation of combinatorial structures from a uniform distribution». *Theoret. Comput. Sci.*, 43 (2-3) (1986), 169–188.
- [34] JOLY, E.; LUGOSI, G.; OLIVEIRA, R. I. «On the estimation of the mean of a random vector». *Electron. J. Stat.*, 11 (1) (2017), 440–451.
- [35] KLIVANS, A.; KOTHARI, P. K.; MEKA, R. «Efficient algorithms for outlier-robust regression». A: *Proceedings of Machine Learning Research*. Vol. 75. PMLR, 2018, 1420–1430.
- [36] KOLMOGOROV, A. N. «Unbiased estimates». *Izvestiya Akad. Nauk SSSR. Ser. Mat.*, 14 (1950), 303–326. [En rus]
- [37] KOTHARI, P. K.; STEINHARDT, J.; STEURER, D. «Robust moment estimation and improved clustering via sum of squares». A: *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. Nova York: ACM, 2018, 1035–1046.
- [38] LAI, K. A.; RAO, A. B.; VEMPALA, S. «Agnostic estimation of mean and covariance». A: *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2016*. Los Alamitos, CA: IEEE Computer Soc., 2016, 665–674.
- [39] LEDOUX, M. *The Concentration of Measure Phenomenon*. Providence, RI: American Mathematical Society, 2001. (Mathematical Surveys and Monographs; 89)
- [40] LEDOUX, M.; TALAGRAND, M. *Probability in Banach Spaces. Isoperimetry and Processes*. Berlín: Springer-Verlag, 1991. (Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3); 23]
- [41] LEE, J. C. H.; VALIANT, P. «Optimal sub-Gaussian mean estimation in $\mathbb{R}$ ». A: *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science—FOCS 2021*. Los Alamitos, CA: IEEE Computer Soc., 2022, 672–683.
- [42] LEI, Z.; LUH, K.; VENKAT, P.; ZHANG, F. «A fast spectral algorithm for mean estimation with sub-Gaussian rates». A: *Proceedings of Machine Learning Research*. Vol. 125, PMLR, 2020, 2598–2612.
- [43] LERASLE, M.; OLIVEIRA, R. I. «Robust empirical mean estimators». Preprint (2011). [Disponible en línia a: [arXiv:1112.3914](https://arxiv.org/abs/1112.3914)]
- [44] LOH, P.-L.; TAN, X. L. «High-dimensional robust precision matrix estimation: cellwise corruption under ϵ -contamination». *Electron. J. Stat.*, 12 (1) (2018), 1429–1467.

- [45] LUGOSI, G.; MENDELSON, S. «Mean estimation and regression under heavy-tailed distributions: a survey». *Found. Comput. Math.*, 19 (5) (2019), 1145–1190.
- [46] LUGOSI, G.; MENDELSON, S. «Sub-Gaussian estimators of the mean of a random vector». *Ann. Statist.*, 47 (2) (2019), 783–794.
- [47] LUGOSI, G.; MENDELSON, S. «Robust multivariate mean estimation: the optimality of trimmed mean». *Ann. Statist.*, 49 (1) (2021), 393–410.
- [48] MENDELSON, S. «Learning without concentration». *J. ACM*, 62 (3) (2015), Art. 21, 25 p.
- [49] MENDELSON, S. «Learning without concentration for general loss functions». *Probab. Theory Related Fields*, 171 (1-2) (2018), 459–502.
- [50] MINSKER, S. «Geometric median and robust estimation in Banach spaces». *Bernoulli*, 21 (4) (2015), 2308–2335.
- [51] MINSKER, S.; NDAOUD, M. «Robust and efficient mean estimation: an approach based on the properties of self-normalized sums». *Electron. J. Stat.*, 15 (2) (2021), 6036–6070.
- [52] MINSKER, S.; STRAWN, N. «Distributed statistical estimation and rates of convergence in normal approximation». Preprint (2017). [Disponible en línia a: arXiv:1704.02658]
- [53] NEMIROVSKY, A. S.; YUDIN, D. B. *Problem Complexity and Method Efficiency in Optimization*. A Wiley-Interscience Publication. Nova York: John Wiley & Sons, Inc., 1983. (Wiley-Interscience Series in Discrete Mathematics). [Traduït del rus i amb prefaci d'E. R. Dawson]
- [54] OLIVEIRA, R. I.; ORENSTEIN, P. «The sub-gaussian property of trimmed means estimators». Informe tècnic, IMPA (2019).
- [55] RADHAKRISHNA RAO, C. «Information and the accuracy attainable in the estimation of statistical parameters». *Bull. Calcutta Math. Soc.*, 37 (1945), 81–91.
- [56] SHEVTSOVA, I. G. «On absolute constants in inequalities of Berry-Esseen type». *Dokl. Akad. Nauk*, 456 (6) (2014), 650–654. [En rus; traducció a: *Dokl. Math.*, 89 (3) (2014), 378–381]
- [57] SMALL, C. G. «A survey of multidimensional medians». *Int. Stat. Rev.*, 58 (3) (1990), 263–277.
- [58] STIGLER, S. M. «The asymptotic distribution of the trimmed mean». *Ann. Statist.*, 1 (1973), 472–477.
- [59] TUKEY, J. W. «Mathematics and the picturing of data». A: *Proceedings of the International Congress of Mathematicians*. Vol. 2. Montreal, Que.: Canad. Math. Congress, 1975, 523–531.
- [60] TUKEY, J. W.; MCLAUGHLIN, D. H. «Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Winsorization. I». *Sankhyā Ser. A*, 25 (1963), 331–352.

- [61] VALIANT, L. G. «A theory of the learnable». *Communications of the ACM*, 27 (11) (1984), 1134-1142.
- [62] VAN DER VAART, A. W.; WELLNER, J. A. *Weak Convergence and Empirical Processes. With Applications to Statistics*. Nova York: Springer-Verlag, 1996. (Springer Series in Statistics)
- [63] VAPNIK, V. N.; CHERVONENKIS, A. YA. *Theory of Pattern Recognition*. Moscow: Statistical problems of learning Izdat. «Nauka», 1974. [En rus]
- [64] VERSHYNIN, R. «Lectures in Geometric Functional Analysis» (2009). [No publicat]

DEPARTAMENT D'ECONOMIA I EMPRESA, UNIVERSITAT POMPEU FABRA
ICREA, PG. LLUÍS COMPANYS 23, 08010 BARCELONA
BARCELONA GRADUATE SCHOOL OF ECONOMICS
gabor.lugosi@upf.edu

English summaries

Jordi Delgado and Enric Ventura

Stallings automata: A roundtrip between algebra and geometry

In this paper, we review the fundamental properties of the free group and give a detailed account of Stallings's automata theory, a geometric interpretation of its subgroups. Stallings automata theory has been (and remains) immensely fruitful, both as a means of understanding classical results and as a source of new ones. We describe some of the most important applications of this theory.

Keywords: free group, subgroup, automaton, Stallings, algorithmic problem, decision problem.

MSC2010 Subject Classification: 20-02, 20E05, 20F05, 20F10, 20F65, 05C25.

Gábor Lugosi

Mean estimation: A survey of recent advances

Perhaps the most basic statistical problem is to estimate the expected value of a random variable based on independent observations of the same distribution. This seemingly simple question raises surprising difficulties both from a statistical and a computational point of view. In this paper we survey some of the recent advances in mean estimation. In particular, we describe sub-Gaussian mean estimators for possibly heavy-tailed data in both the univariate and multivariate settings. We focus on estimators based on median-of-means techniques, although other methods such as the trimmed mean and Catoni's estimator are also reviewed. We offer detailed evidence for our core findings.

Keywords: mean estimation, heavy-tailed distributions, robustness, statistical learning.

MSC2010 Subject Classification: 62G05, 62G15, 62G35.

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 MSC2010.

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). 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
gasull@mat.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
rcamps@mat.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
solanes@mat.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* i *Butlletí de la Societat Catalana de Matemàtiques*. Els socis de la **SCM** reben, gratuïtament, aquestes dues 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: Dolors Herbera i Espinal

VICEPRESIDENT: Josep Vives i Santa-Eulàlia

ADJUNT A LA VICEPRESIDÈNCIA: Abraham de la Fuente Pérez

SECRETÀRIA: Immaculada Baldomà i Barraca

TRESORER: Albert Granados i Corsellas

VOCAL EMPRESA: Aleix Ruiz de Villa

VOCAL PUBLICACIONS: Montserrat Alsina i Aubach

VOCALS: Josep Grané i Manlleu, Mireia Lopez i Beltran, Albert Ruiz i Cirera, Manel Udina i Abelló

DELEGADA DE L'IEC: Pilar Bayer i Isant

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: <http://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 <http://scm.iec.cat>.

La correspondència administrativa s'ha d'adreçar a la Societat Catalana de Matemàtiques.

Editor en cap _____

Antoni Guillaumon

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 Carne, 47 - 08001 Barcelona

tel. 933 248 583 - fax 932 701 180

scm@iec.cat - <http://scm.iec.cat>