Le sous-langage sans a du pavage de Goldbach : construction inconditionnelle du sous-shift, du groupoïde de fin-de-mot-équivalence et de sa C^*-algèbre

Denise Vella-Chemla pilotant claude, septembre 2026

Résumé

On développe ici le point 5 de la note précédente, *Pavage apériodique représentant les décompositions de Goldbach : pourquoi il n'apporte rien de plus* https://denisevellachemla.eu/indiscernabilite-locale-pavage-conj-Goldbach-claude-dvc-20260904.pdf , le sous-langage du pavage évitant la lettre a (donc les lettres b, c, d seulement) porte, lui, une construction complète et inconditionnelle de type Matsumoto/Carlsen - sous-shift, groupoïde de fin-de-mot-équivalence, C^*-algèbre. On établit ce résultat en détail, puis on explique précisément pourquoi il ne mène pas plus loin arithmétiquement, sans esquiver la question.

1. Rappel du cadre

Pour n pair, $n \geq 6$, on note $p_1 < \cdots < p_{m(n)}$ les entiers impairs de $[3, n/2]$ et $q_i = n - p_i$. On pose

$$u_i(n) = \mathbf{1}[p_i \text{ premier}], \qquad v_i(n) = \mathbf{1}[q_i \text{ premier}],$$

et $L(n,i) = (u_i(n), v_i(n)) \in \{a, b, c, d\}$ avec $a = (1,1)$, $b = (1,0)$, $c = (0,1)$, $d = (0,0)$.

On note $w_n \in \{b, c, d\}^*$ le mot obtenu en supprimant, dans la ligne $L(n,1) \dots L(n, m(n))$, toutes les occurrences de a. C'est ce mot w_n qui porte toute la construction ci-dessous.

2. Le mot w_n tend vers l'infini : un fait inconditionnel

Lemme 1 (Borne de crible)**.** *Il existe une constante $C > 0$ telle que, pour n pair assez grand, le nombre $r(n)$ d'indices i avec $L(n,i) = a$ vérifie*

$$r(n) \leq C \, \frac{n}{2(\ln n)^2}.$$

Justification. Le crible de Selberg donne, pour le nombre de couples *ordonnés* (p,q) avec $p+q = n$, p, q premiers, p parcourant $[3, n-3]$, la majoration classique et inconditionnelle $\mathrm{O}\left(\dfrac{n}{(\ln n)^2}\right)$ (Halberstam-Richert, *Sieve Methods*, ch. 3). Mais ici i n'indexe que les $p_i \in [3, n/2]$ (convention retenue dès le cadre de la section 1 : une décomposition $n = p+q$ n'est comptée qu'une fois, via son représentant $p \leq q$) - soit exactement la moitié de l'intervalle de crible complet. D'où le facteur $1/2$: $r(n) = \mathrm{O}\left(\dfrac{n}{2(\ln n)^2}\right)$, cohérent avec la convention de comptage utilisée par ailleurs pour la minoration $f(x) = \dfrac{(1+\sqrt{5})x}{2\sqrt{5}(\ln x)^2}$ de la comète de Goldbach. Contrairement à la minoration correspondante (conjecturale, équivalente à Goldbach), cette majoration est un théorème inconditionnel. □

Proposition 2. $|w_n| = m(n) - r(n) \to \infty$ *quand* $n \to \infty$*, et plus précisément* $|w_n|/m(n) \to 1$.

Démonstration. $m(n) \sim n/4$ (comptage des entiers impairs de $[3, n/2]$), tandis que

$$r(n) = \mathrm{O}\left(\frac{n}{(\ln n)^2}\right) = \mathrm{o}(n)$$

par le lemme 1. Donc $|w_n| = m(n) - r(n) \sim n/4$. □

C'est le seul fait arithmétique non trivial utilisé dans toute la suite, et il est entièrement inconditionnel.

3. Construction du sous-shift X_{bcd}

On veut fabriquer, à partir de la suite de mots finis $(w_n)_n$, un véritable sous-shift bi-infini sur $\{b, c, d\}$ - l'objet sur lequel Matsumoto/Carlsen savent construire groupoïde et C^*-algèbre.

Définition 3 (Langage limite)**.** *Pour* $u \in \{b, c, d\}^*$*, on dit que* $u \in L$ *si* u *est facteur (sous-mot consécutif) de* w_n *pour une infinité de* n*.*

Lemme 4 (L est factoriel)**.** *Si* $u \in L$ *et* u' *est facteur de* u*, alors* $u' \in L$*.*

Démonstration. Immédiat : tout n tel que u est facteur de w_n vérifie aussi que u' est facteur de w_n. □

Lemme 5 (L est extensible à droite et à gauche)**.** *Pour tout* $u \in L$*, il existe* $x, y \in \{b, c, d\}$ *tels que* $xu \in L$ *et* $uy \in L$*.*

Démonstration. Soit $k = |u|$ et $S = \{n : u \text{ facteur de } w_n\}$, infini par hypothèse. Par la proposition 2, $|w_n| > k$ pour tout $n \in S$ assez grand ; on ne considère que ceux-là, ensemble encore infini, noté S'.

Fixons $n \in S'$ et une occurrence de u dans w_n, disons commençant à la position j. Deux cas, non exclusifs :

- si $j > 1$: la lettre en position $j - 1$ existe, notons-la x_n ; alors $x_n u$ est facteur de w_n ;
- si $j + k - 1 < |w_n|$: la lettre en position $j + k$ existe, notons-la y_n ; alors uy_n est facteur de w_n.

Comme $|w_n| > k$, on ne peut pas avoir simultanément $j = 1$ *et* $j + k - 1 = |w_n|$ (cela forcerait $|w_n| = k$). Donc au moins un des deux cas ci-dessus a lieu pour *chaque* $n \in S'$. Par le principe des tiroirs, comme il n'y a que $2 \times 3 = 6$ couples (côté, lettre) possibles, une infinité de $n \in S'$ partagent le même couple, disons (droite, y) sans perte de généralité. Alors $uy \in L$. Le même argument, appliqué à l'autre côté sur le complémentaire (encore infini si le premier cas n'épuise pas S', ou directement si c'est le cas dominant), donne l'extensibilité de l'autre côté - au pire on itère la construction séparément pour chaque côté, chacun donnant une infinité de n car S' est infini et on ne le partitionne qu'en 6 classes. □

Théorème 6 (Existence inconditionnelle du sous-shift)**.** *Soit*

$$X_{bcd} = \Big\{ x \in \{b, c, d\}^{\mathbb{Z}} \ : \ \textit{tout facteur fini de } x \textit{ appartient à } L \Big\}.$$

Alors X_{bcd} *est non vide, fermé et invariant par le décalage* σ *- c'est un sous-shift au sens usuel de la dynamique symbolique.*

Démonstration. Fermeture et σ-invariance sont immédiates de la définition (propriété de cylindre). La non-vacuité est le point substantiel ; elle découle des lemmes 4 et 5 par une construction standard de type König :

On construit x coordonnée par coordonnée. Partant d'un mot $u_0 \in L$ non vide (qui existe : par la proposition 2, $|w_n| \geq 1$ pour n grand, donc au moins une lettre de $\{b, c, d\}$ est facteur d'une infinité de w_n, par tiroirs sur les 3 lettres possibles), on applique le lemme 5 pour obtenir $u_1 = x_{-1}u_0y_1 \in L$ de longueur $|u_0| + 2$, puis on répète indéfiniment. On obtient une suite croissante emboîtée (u_k) de mots de L, centrés en position 0, dont la réunion des positions couvre $\mathbb{Z}$. Le point $x \in \{b, c, d\}^{\mathbb{Z}}$ ainsi défini a, par construction, tous ses facteurs finis dans L (chaque facteur de x est facteur d'un u_k pour k assez grand, donc dans L par le lemme 4 appliqué à un $u_k \in L$). Donc $x \in X_{bcd}$. □

C'est exactement l'énoncé visé par le point 5 de la note précédente : *un objet légitime et démontrable dès aujourd'hui*, sans rien de conditionnel - la seule brique arithmétique utilisée est la majoration de crible du lemme 1.

4. Le groupoïde de fin-de-mot-équivalence

Définition 7. *Sur X_{bcd}, on définit la relation de fin-de-mot-équivalence :*

$$x \sim y \iff \exists\, k, l \geq 0,\ \sigma^k(x) = \sigma^l(y).$$

Le groupoïde associé est

$$G_{bcd} = \Big\{(x, k - l, y) : x, y \in X_{bcd},\ k, l \geq 0,\ \sigma^k(x) = \sigma^l(y)\Big\},$$

muni de la topologie engendrée par les cylindres, avec unités X_{bcd}, source/but $(x, \cdot, y) \mapsto y$, x, et composition $(x, m, y)(y, m', z) = (x, m + m', z)$.

Proposition 8. *G_{bcd} est un groupoïde étale, localement compact, à base dénombrable, et moyennable.*

Justification. C'est la construction de Deaconu-Renault pour un sous-shift quelconque (Matsumoto 1997, Carlsen 2004) : elle ne suppose *aucune* propriété du sous-shift (ni type fini, ni sofique, ni minimalité) - seulement son existence en tant que fermé σ-invariant de $\{b, c, d\}^{\mathbb{Z}}$, établie au théorème 6. La moyennabilité de ce type de groupoïde (groupoïde de Deaconu-Renault d'un système dynamique topologiquement libre sur un espace localement compact) est un résultat général (Renault, Anantharaman-Delaroche-Renault). □

5. La C^*-algèbre

Le groupoïde étant étale et moyennable, l'algèbre de convolution $C_c(G_{bcd})$ admet une unique complétion en C^*-algèbre (pleine = réduite) :

$$\mathcal{A}_{bcd} := C^*(G_{bcd}).$$

C'est l'algèbre de Matsumoto du sous-shift X_{bcd} - la généralisation, à un sous-shift quelconque (pas nécessairement de type fini), de la construction de Cuntz-Krieger. C'est *l'analogue exact*, pour le pavage de Goldbach débarrassé de la lettre a, de ce que Connes fait pour Penrose : espace des configurations globales, groupoïde de fin-de-mot-équivalence, C^*-algèbre du groupoïde.

6. Pourquoi cela n'apporte, en l'état, rien de plus

Ce qui manque pour que $\mathcal{A}_{bcd}$ dise quelque chose d'arithmétiquement intéressant sur Goldbach est précisément ce que la construction ci-dessus n'utilise pas :

- **Pas de règle de substitution.** Pour Penrose, $K_0 \cong \mathbb{Z}[\varphi]$ tombe de la matrice de Perron-Frobenius de la règle de subdivision - un mécanisme combinatoire explicite. Ici, la loi

$$(u, v) \cdot (x, y) = (u, y)$$

  n'est *pas* une substitution (elle est dégénérée, cf. note précédente) : X_{bcd} n'est défini que comme le langage-limite *observé* des bits de primalité, sans règle génératrice interne. Il n'y a donc aucune matrice dont on pourrait tirer un analogue de $\mathbb{Z}[\varphi]$.
- L **reste inconnu explicitement.** Le théorème 6 garantit que X_{bcd} existe et n'est pas vide, mais ne dit rien de *quels* mots il contient réellement - cela dépend des corrélations fines entre premiers consécutifs le long de deux progressions arithmétiques opposées (p_i croissant, q_i décroissant), un objet du même ordre de difficulté que les constantes de Hardy-Littlewood pour les configurations de type nombres premiers jumeaux. Sans cette information, L pourrait très bien être le langage plein $\{b, c, d\}^*$ (sous-shift = shift plein, entropie maximale, K_0 sans contenu arithmétique) - on ne sait pas l'exclure.
- **Le calcul de** $K_*(\mathcal{A}_{bcd})$ pour un sous-shift général (non sofique) est lui-même un problème dur en dynamique symbolique pure, indépendamment de Goldbach - il n'existe pas de formule générale comme pour les algèbres de Cuntz-Krieger (où K_0 se lit sur $\mathrm{coker}(I - A^T)$).

Remarque : On a donc, précisément, le tableau suivant :

- *Inconditionnel et acquis dès aujourd'hui* : existence de X_{bcd}, de G_{bcd}, de $\mathcal{A}_{bcd}$ (théorème 6 et ses corollaires).
- *Non acquis, et probablement aussi dur que Goldbach lui-même (ou que les corrélations de type Hardy-Littlewood en général)* : toute description explicite de L, donc tout calcul de $K_*(\mathcal{A}_{bcd})$, donc tout contenu arithmétique de la construction.

Le pont formel (sous-shift $\to$ groupoïde $\to$ C^*-algèbre) est donc bien un résultat réel, pas une esquive - mais il transporte fidèlement, dans le dictionnaire non commutatif, un problème dont la difficulté ne diminue pas en le traduisant.