Conjecture de Goldbach, réécriture, contradiction

Denise Vella-Chemla

30/3/14

1 16 règles de réécriture

On rappelle qu'on a choisi de représenter le fait qu'un entier est premier par le booléen 0 et le fait qu'il est composé par le booléen 1.

On a également pris comme conventions que ($3 \leqslant p \leqslant n/2$) :

- la lettre a symbolise une décomposition de n de la forme $p+q$ avec p et q premiers ;
- la lettre b symbolise une décomposition de n de la forme $p+q$ avec p composé et q premier ;
- la lettre c symbolise une décomposition de n de la forme $p+q$ avec p premier et q composé ;
- la lettre d symbolise une décomposition de n de la forme $p+q$ avec p et q composés.

La lettre a code la matrice $\begin{pmatrix}0\\0\end{pmatrix}$, et respectivement b $\begin{pmatrix}0\\1\end{pmatrix}$, c $\begin{pmatrix}1\\0\end{pmatrix}$ et enfin d $\begin{pmatrix}1\\1\end{pmatrix}$

Exemple : Ci-dessous le mot $m_{abcd}(40)$.

40	37	35	33	31	29	27	25	23	21
	0	1	1	0	0	1	1	0	1
	0	0	0	1	0	0	1	0	0
	3	5	7	9	11	13	15	17	19
$m_{abcd}(40)$	a	c	c	b	a	c	d	a	c

Dans la suite, on utilise l'opération ainsi définie sur les matrices :

$$\begin{pmatrix}x_1\\x_2\end{pmatrix}.\begin{pmatrix}y_1\\y_2\end{pmatrix}=\begin{pmatrix}x_1\\y_2\end{pmatrix}$$

L'opération ci-dessus fournit 16 règles de réécriture de couples de lettres, qui semblent pertinentes pour l'étude de la conjecture de Goldbach :

1) $aa \to a$	5) $ba \to a$	9) $ca \to c$	13) $da \to c$
2) $ab \to b$	6) $bb \to b$	10) $cb \to d$	14) $db \to d$
3) $ac \to a$	7) $bc \to a$	11) $cc \to c$	15) $dc \to c$
4) $ad \to b$	8) $bd \to b$	12) $cd \to d$	16) $dd \to d$

2 Rappels de théorie des langages

Un alphabet est un ensemble fini de symboles.

Les alphabets utilisés ci-après sont : $A = \{a, b, c, d\}$, $A_{ab} = \{a, b\}$, $A_{cd} = \{c, d\}$, $A_{ac} = \{a, c\}$ et $A_{bd} = \{b, d\}$.

Un mot sur l'alphabet X est une séquence finie et ordonnée, éventuellement vide, d'éléments de l'alphabet. C'est une concaténation de lettres. On note X^* l'ensemble des mots sur l'alphabet X.

Un mot est préfixe d'un autre s'il contient, sur toute sa longueur, les mêmes lettres aux mêmes positions (Soient un alphabet X et $w, u \in X^*$. u est préfixe de w si et seulement si $\exists v \in X^*$ tel que $w = u.v$).

3 Observer les mots

Observons les mots associés aux nombres pairs compris entre 6 et 36.

$$
\begin{array}{r|llllllll}
6: & a \\
8: & a \\
10: & a & a \\
12: & c & a \\
14: & a & c & a \\
16: & a & a & c \\
18: & c & a & a & d \\
20: & a & c & a & b \\
22: & a & a & c & b & a \\
24: & c & a & a & d & a \\
26: & a & c & a & b & c & a \\
28: & c & a & c & b & a & c \\
30: & c & c & a & d & a & a & d \\
32: & a & c & c & b & c & a & b \\
34: & a & a & c & d & a & c & b & a \\
36: & c & a & a & d & c & a & d & a
\end{array}
$$

FIGURE 1 : mots des nombres pairs de 6 à 36

4 Propriétés des mots

Les mots diagonaux (diagonales) ont leurs lettres soit dans l'alphabet A_{ab} soit dans l'alphabet A_{cd}.

Toute diagonale est préfixe de la diagonale suivante définie sur le même alphabet.

Une diagonale code en effet des décompositions de même second sommant et de premier sommant un nombre impair de la liste des impairs successifs à partir de 3.

Par exemple, la diagonale *aaaba*, qui commence au a première lettre du mot de 26 sur la figure 1 code les décompositions $3 + 23, 5 + 23, 7 + 23; 9 + 23, 11 + 23$ et $13 + 23$.

Ainsi, les diagonales sur l'alphabet A_{ab} "codent" les décompositions dont le second sommant est premier ; les lettres de telles diagonales codent soit par des a correspondant aux nombres premiers, soit par des b correspondant aux nombres composés la séquence des caractères de primalité des entiers impairs, à partir de 3.

Les diagonales sur l'alphabet A_{cd} "codent" quant à elles des décompositions dont le second sommant est composé ; les lettres de telles diagonales codent soit par des c correspondant aux nombres premiers, soit par des d correspondant aux nombres composés la séquence des caractères de primalité des entiers impairs, à partir de 3.

Les mots verticaux ont leurs lettres soit dans l'alphabet A_{ac} soit dans l'alphabet A_{bd}. Un mot vertical code des décompositions successives en somme de deux impairs de même premier sommant. Tout mot vertical est contenu dans un mot vertical qui se trouve à sa gauche et qui est défini sur le même alphabet.

5 Quelques régularités

On observe quelques régularités facilement explicables, qui lient entre eux les nombres de lettres de chaque sorte apparaissant dans un mot et dans une portion de la colonne des premières lettres. La figure 2 ci-dessous présente schématiquement les noms des variables qui seront utiles au raisonnement :

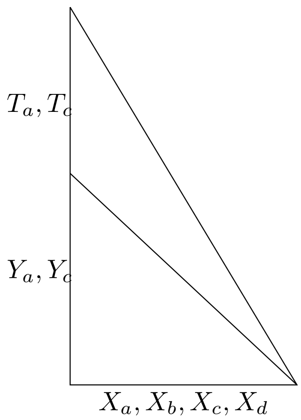

FIGURE 2 : variables liées

Le triangle global contient les mots associés aux nombres pairs de 6 à n.

X_a, X_b, X_c et X_d comptent le nombre de a, b, c ou d du mot associé à n.

T_a et T_c comptent le nombre de lettres a ou c qui sont premières lettres des mots associés aux nombres pairs compris entre 6 et $2\left\lceil \dfrac{n+2}{4} \right\rceil$.

T_a correspond ainsi aux décompositions de la forme $n' = 3 + p_i$, p_i premier, $n' \leqslant 2\left\lceil \dfrac{n+2}{4} \right\rceil$. Par exemple, pour $n = 34$, $T_a = \#\{3{+}3, 3{+}5, 3{+}7, 3{+}11, 3{+}13\}$.

T_c correspond aux décompositions de la forme $n' = 3 + c_i$, c_i composé pour $n' \leqslant 2\left\lceil \dfrac{n+2}{4} \right\rceil$. Par exemple, pour $n = 34$, $T_c = \#\{3+9, 3+15\}$.

Y_a et Y_c comptent le nombre de lettres a ou c qui sont premières lettres de mots associés aux nombres pairs compris entre $2\left\lceil \dfrac{n+2}{4} \right\rceil + 2$ et n.

La bijection triviale sur le deuxième sommant des décompositions permet d'expliquer aisément pourquoi $Y_a = X_a + X_b$ ou bien pourquoi $Y_c = X_c + X_d$. La simple présentation des ensembles en extension suffit à s'en convaincre.

$$\begin{array}{ll}
Y_a & = \#\{3+17, 3+19, 3+23, 3+29, 3+31\} \\
X_a & = \#\{3+31, 5+29, 11+23, 17+17\} \\
X_b & = \#\{15+19\} \\
 & \\
Y_c & = \#\{3+21, 3+25, 3+27\} \\
X_c & = \#\{7+27, 13+21\} \\
X_d & = \#\{9+25\}
\end{array}$$

Ci-après deux figures permettant de "fixer les idées" pour les nombres pairs $n = 32$ ou $n = 34$.

	6 :	a							
	8 :	a							
	10 :	a	a						
$T_a = 5$	12 :	c	a						
$T_c = 2$	14 :	a	c	a					
	16 :	a	a	c					
	18 :	c	a	a	d				
	20 :	a	c	a	b				
	22 :	a	a	c	b	a			
$Y_a = 4$	24 :	c	a	a	d	a			
$Y_c = 3$	26 :	a	c	a	b	c	a		
	28 :	c	a	c	b	a	c		
	30 :	c	c	a	d	a	a	d	
	32 :	a	c	c	b	c	a	b	$\leftarrow X_a = 2, X_b = 2, X_c = 3, X_d = 0$

FIGURE 3 : premier exemple : $n = 32$

	6 :	a								
	8 :	a								
	10 :	a	a							
$T_a = 5$	12 :	c	a							
$T_c = 2$	14 :	a	c	a						
	16 :	a	a	c						
	18 :	c	a	a	d					
	20 :	a	c	a	b					
	22 :	a	a	c	b	a				
	24 :	c	a	a	d	a				
	26 :	a	c	a	b	c	a			
$Y_a = 5$	28 :	c	a	c	b	a	c			
$Y_c = 3$	30 :	c	c	a	d	a	a	d		
	32 :	a	c	c	b	c	a	b		
	34 :	a	a	c	d	a	c	b	a	$\leftarrow X_a = 4, X_b = 1, X_c = 2, X_d = 1$

FIGURE 4 : second exemple : $n = 34$

Les contraintes suivantes sont toujours vérifiées :

$$\begin{array}{l} Y_a = X_a + X_b \\ Y_c = X_c + X_d \\ T_a + T_c + Y_a + Y_c + \epsilon = 2(X_a + X_b + X_c + X_d) \end{array}$$

$\epsilon = 1$ si n est un double d'impair, $\epsilon = 0$ sinon.

On a vu que les contraintes ci-dessus peuvent être aisément comprises si l'on revient aux nombres de décompositions qu'elles représentent et si l'on utilise des bijections “à la Cantor”.

Les lettres des différents mots sont ainsi très *intriquées* et ces intrications ont pour conséquence que tout mot contient un a. On va prouver cela en menant un raisonnement par l'absurde dans la section 7.

6 Visualisation des bijections de Cantor

Ci-dessous, on visualise les bijections de Cantor pour les cas $n = 32$ et $n = 34$.

La bijection f qui permet de passer de la ligne 2 du tableau à la ligne 1 est telle que $f(a) = f(c) = a$ et $f(b) = f(d) = c$.

La bijection g qui permet de passer de la ligne 2 du tableau à la ligne 3 est telle que $g(a) = g(b) = a$ et $g(c) = g(d) = c$.

C'est la duplication de la décomposition $3 + n/2$ dans le cas des doubles d'impairs qui nécessite l'introduction de la variable ϵ qui vaut alors 1 et 0 sinon.

– Bijections pour $n = 32$

	3	3	3	3	3	3	3
1	a	a	a	c	a	a	c
	3	5	7	9	11	13	15
	3	5	7	9	11	13	15
2	a	c	c	b	c	a	b
	29	27	25	23	21	19	17
	29	27	25	23	21	19	17
3	a	c	c	a	c	a	a
	3	3	3	3	3	3	3

– Bijections pour $n = 34$

	3	3	3	3	3	3	3	3
1	a	a	a	c	a	a	c	a
	3	5	7	9	11	13	15	17
	3	5	7	9	11	13	15	17
2	a	a	c	d	a	c	b	a
	31	29	27	25	23	21	19	17
	31	29	27	25	23	21	19	17
3	a	a	c	c	a	c	a	a
	3	3	3	3	3	3	3	3

7 Rechercher une contradiction

Imaginons que le mot m_n est associé à un nombre n qui contredit la conjecture de Goldbach, i.e. m_n ne contient aucune lettre a, la lettre a symbolisant on le rappelle la somme de deux nombres premiers.

m_n ne contenant aucune lettre a, on a $X_a = 0$. Mais puisque $Y_a = X_a + X_b$, on a alors $Y_a = X_b$. En identifiant Y_a à X_b et Y_c à $X_c + X_d$ dans la dernière contrainte toujours respectée fournie au paragraphe précédent, on obtient la suite d'égalités suivante :

$$\begin{aligned} T_a + T_c + Y_a + Y_c + \epsilon &= 2(X_a + X_b + X_c + X_d) \\ T_a + T_c + X_b + X_c + X_d + \epsilon &= 2X_a + 2X_b + 2X_c + 2X_d \\ T_a + T_c + \epsilon &= X_b + X_c + X_d \\ T_a + T_c + \epsilon &= X_b + Y_c \end{aligned}$$

Il s'agit maintenant de se souvenir de ce que représentent ces différentes variables :

- $T_a + T_c = \left\lfloor \dfrac{n-4}{4} \right\rfloor$;
- X_b compte le nombre de décompositions de n sous la forme d'une somme de deux nombres impairs $p + q$ avec $p \leqslant n/2$ composé et q premier ;
- Y_c compte le nombre de nombres composés impairs compris entre $n/2$ et $n - 3$.

Le nombre X_b de décompositions de n sous la forme d'une somme de deux nombres impairs $p + q$ avec $p \leqslant n/2$ composé et q premier étant forcément inférieur au nombre de nombres premiers compris entre $n/2$ et $n-3$, on a $X_b < Y_a$ (on a utilisé ici une sorte de "principe des tiroirs" inversé : si on met 0 ou 1 objet dans k tiroirs, on ne peut avoir plus d'objets que de tiroirs, i.e. plus de k objets). Mais le nombre de nombres premiers contenus dans un intervalle $[2k+3, 4k+1]$ est toujours inférieur au nombre de nombres composés impairs contenus dans cet intervalle pour $k > 25$. Dans ces cas, $Y_a < Y_c$ et $X_b + Y_c < Y_a + Y_c < 2Y_c$.

Or $\left\lfloor \dfrac{n-4}{4} \right\rfloor$ est pour tout entier supérieur à un certain entier assez petit (tel que 100) supérieur à $2Y_c$. Cela assure qu'on a jamais l'égalité $T_a + T_c + \epsilon = X_b + Y_c$ qui découlerait de l'absence de lettre a dans un mot.

On a ainsi abouti à une contradiction qui serait conséquence de l'absence de a dans un mot. Cela entraîne l'impossibilité qu'un nombre pair contredise la conjecture de Goldbach. Les règles de réécriture intriquent totalement les lettres des mots de telle manière que les nombres de lettres de chaque sorte doivent respecter impérativement certaines contraintes. On se situe dans une théorie lexicale des nombres, selon laquelle les nombres sont des mots. Cette théorie exploite le fait que l'ordre des lettres dans les mots de Goldbach est primordial.