Mots de nombres premiers

Denise Vella-Chemla

4.3.17

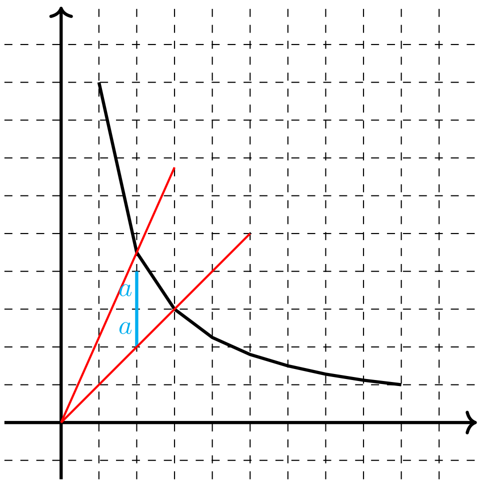
On voudrait ici présenter lentement une découverte qu'on a faite récemment qui permet de caractériser les nombres premiers en étudiant des mots sur un alphabet à 2 lettres a et b.

Cette découverte une fois effectuée est simple, elle utilise des opérations élémentaires : ajouter 1 à un nombre, multiplier deux nombres, comparer deux nombres, coder le sens d'inégalités par des lettres a ou b et comparer des mots.

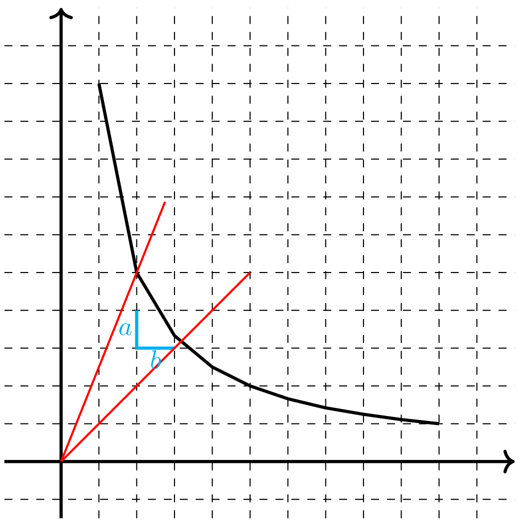
On est parti du fait que l'hyperbole d'équation $xy = n$ avec n un nombre premier ne passe que par des points à coordonnées non-entières du plan, hormis les points $(1, n)$ et $(n, 1)$. Par contre, si n est composé, l'hyperbole passe par autant de points à coordonnées entières du plan que n a de diviseurs. Par exemple, l'hyperbole d'équation $xy = 12$ passe par les points à coordonnées entières $(1, 12), (2, 6), (3, 4), (4, 3), (6, 2), (12, 1)$. Ces considérations (sur les points, dont les 2 coordonnées dans le plan peuvent être entières ou non) sont des considérations géométriques.

On a souhaité comprendre ce qui se passait lors du passage d'une hyperbole à la suivante (i.e. de l'hyperbole d'équation $xy = n$ à l'hyperbole d'équation $xy = n + 1$). Dans ce but, on a choisi d'associer à chaque hyperbole son mot de Christoffel.

Un mot de Christoffel [1] est un mot sur un alphabet de deux lettres qui "colle au plus près" à une courbe par segments liant des points discrets (par le dessus ou par le dessous). Voyons un exemple. Pour que la caractérisation des nombres premiers soit la plus simple possible, on ne conserve qu'une portion de chaque mot de Christoffel appartenant à une zone délimitée par deux droites, dans le quadrant $x > 0, y > 0$ (les droites d'équation $y = 2x$ et $y = x$).



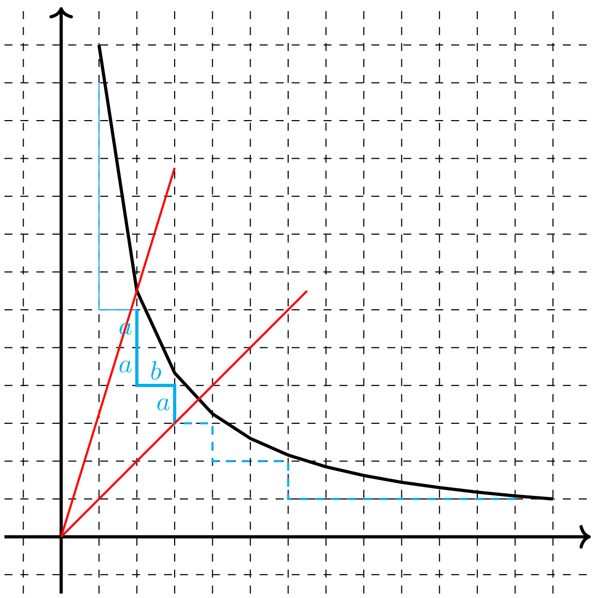
$n = 9$



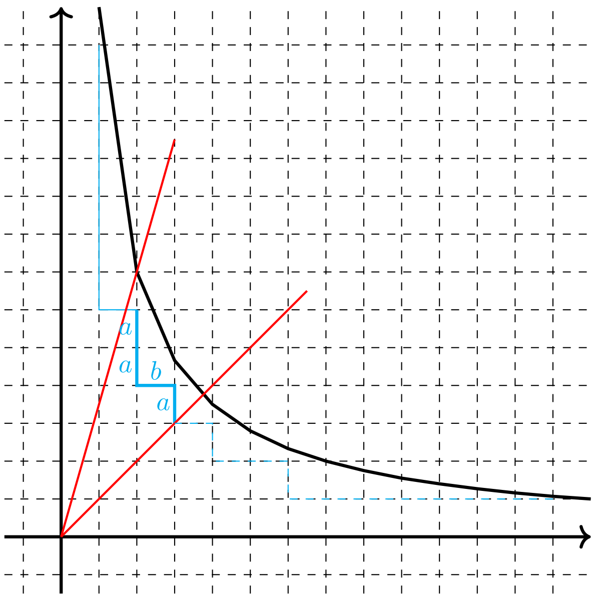
$n = 10$

On constate et c'est simple à comprendre qu'un nombre premier n a le même mot de Christoffel qu'$n + 1$ tandis qu'un nombre composé n a un mot de Christoffel différent de celui d'$n + 1$.

En choisissant la modélisation ci-dessus, on relie trois théories : la théorie des langages (qui contient les concepts d'alphabet, lettre, mot, concaténation, égalité et différence), l'arithmétique (là, les concepts sont ceux de nombre, diviseur, premier, composé) et la géométrie (et ses concepts de points, réseau de points entiers, hyperbole et équation).



$n = 13$



$n = 14$

Qu'est-ce qui se cache derrière cette modélisation informatique ?

Chaque lettre a ou b code une assertion logique, qui correspond ici à une inégalité entre deux nombres. On a un segment de longueur 1, étiqueté b, d'origine le point de coordonnées (x, y) partant horizontalement dans la direction de l'hyperbole[1] associée à n si $(x+1)y \leqslant n$. Dans le cas contraire, descend verticalement de ce point un segment étiqueté a de longueur 1.

Quand on dit qu'*un mot est égal à un autre mot*, une phrase si élémentaire que sa véracité peut être appréhendée par un élève de maternelle, on lie entre elles toute une série d'assertions logiques, les unes associées à l'hyperbole d'équation $xy = n$, les autres associées à l'hyperbole d'équation $xy = n + 1$, et on étudie si toutes les assertions sont deux à deux simultanément vérifiées ou simultanément invérifiées, ce qui est assez complexe à exprimer algébriquement.

Considérons un exemple simple, celui du passage du mot *aaaba* du nombre composé 15 au mot *aabaa* de son successeur.

Les lettres *aaaba* "de" 15 codent les inégalités :

$$\left\{\begin{array}{l} a : (2+1) * 7 \geqslant 15 \\ a : (2+1) * 6 \geqslant 15 \\ a : (2+1) * 5 \geqslant 15 \\ b : (2+1) * 4 < 15 \\ a : (3+1) * 4 \geqslant 15 \end{array}\right.$$

tandis que les lettres *aabaa* "de" 16 codent les inégalités :

$$\left\{\begin{array}{l} a : (2+1) * 7 \geqslant 16 \\ a : (2+1) * 6 \geqslant 16 \\ b : (2+1) * 5 < 16 \\ a : (3+1) * 5 \geqslant 16 \\ a : (3+1) * 4 \geqslant 16 \end{array}\right.$$

Lorsqu'on dit que ces mots sont différents, on regarde chaque doublon de lettres et on étudie si le sens des inégalités pour l'un et l'autre nombre sont égaux. En l'occurrence, on trouve une différence entre les troisièmes lettres des mots $a : (2+1) * 5 \geqslant 15$ et $b : (2+1) * 5 < 16$. A la première différence de sens entre les inégalités associées à n et celles associées à $n + 1$, il y a bifurcation des chemins au niveau des mots de Christoffel.

L'interprétation qui vient d'être donnée est une interprétation algébrique. On peut se placer sur une algèbre de Boole au lieu de se placer dans la théorie des langages : la lettre a correspond à 0, la lettre b à

[1] i.e. on a "la place" pour mettre un b sans toucher à l'hyperbole.

1. Regarder si deux lettres sont identiques consiste à étudier le résultat d'une fonction qui à deux booléens en associe un troisième (une fonction de $\mathbb{B} \times \mathbb{B}$ dans $\mathbb{B}$) qui code le fait que les deux booléens sont égaux ou pas (tester l'égalité de deux booléens $b_1 = b_2$ consiste à calculer $((1 - b_1) + b_2).((1 - b_2) + b_1))$. En annexe, on rappelle les valeurs des connecteurs logiques.

D'un point de vue géométrique, à quoi cela correspond-il ?

Il faut imaginer les hyperboles d'un nombre et de son successeur ainsi que leur mot de Christoffel respectifs comme se situant dans des plans différents. Il faut aussi imaginer qu'à chaque hyperbole est associée sur la gauche une bande de largeur horizontale 1 dans laquelle se développe le mot de Christoffel. Tant que celui-ci peut se voir concaténer un a sans sortir de la bande, c'est ce qui se passe. Dès que le mot "sort de la bande hyperbolique", il faut lui concaténer une lettre b pour le ramener dans la bande (i.e. pour qu'il reste toujours à distance la plus faible possible de l'hyperbole tout en ne s'y collant pas et il s'agit d'une distance de Manhattan horizontale, je crois, ou distance de norme 1). Dès qu'on a ramené le mot dans la bande, on peut se remettre à concaténer des a, c'est pour cette raison que les lettres b n'apparaissent qu'une par une dans les mots, et jouent le rôle de séparateurs (ou ponctuations).

On peut noter qu'une fois établie la modélisation, on peut oublier son interprétation géométrique et n'en conserver que le versant algébrique (les mots de lettres codant des inégalités) en se rappelant toutefois que ce sont les images géométriques (la perception visuelle des objets géométriques) qui ont permis d'aboutir à la modélisation.

Bibliographie :

[1] Jean Berstel, Aaron Lauve, Christophe Reutenauer, Franco Saliola, *Combinatorics on Words : Christoffel Words and Repetitions in Words*, 2008.

[2] B.A. Trahtenbrot, *Algorithmes et machines à calculer*

[3] Alain Connes, *Un extrait d'un cours donné en 1998 à l'Ohio State University, à Colombus aux Etats-Unis, et dans lequel il est question d'hyperboles et points discrets ; il n'y a pas de lien avec ce qui est présenté dans cette note, on ne retiendra que l'expression utilisée à la fin de la présentation de cet exemple "c'est marrant !", qui exprime toute la jubilation que provoque l'étude et la découverte :*
http://denise.vella.chemla.free.fr/ac-hr-osu-1998-8-hyperb-9mn30.mp4

[4] Alain Connes, *Un extrait d'un Colloquium à l'ICTP de Trieste en Italie en mars 2017, dans lequel sont présentées des anagrammes exprimant la non-commutativité du langage :*
http://denise.vella.chemla.free.fr/ac-ictp-mars2017-anagrammes.mp4

Annexe : rappel des valeurs des connecteurs logiques

La notation P de la dernière colonne correspond au calcul de $(b_1 = b_2)$ aussi égal à $((1 - b_1) + b_2).((1 - b_2) + b_1)$ (en logique, on se situe dans un domaine complètement abélien ; c'est uniquement l'importance de l'ordre des lettres dans les mots (i.e. la non-commutativité de la concaténation en théorie des langages) qui rend notre modélisation non-commutative).

b_1	b_2	$1 - b_1$	$1 - b_2$	$(1 - b_1) + b_2$	$(1 - b_2) + b_1$	P
0	0	1	1	1	1	1
0	1	1	0	1	0	0
1	0	0	1	0	1	0
1	1	0	0	1	1	1

Il faut avoir à l'esprit que la formule fournie pour exprimer l'égalité entre 2 booléens fait intervenir l'addition, la soustraction et la multiplication parce que c'est pratique. En fait, on peut générer les connecteurs $\vee$ (ou $+$), $\wedge$ (ou $.$), $\neg b$ (ou $1 - b$) et $=$ (parfois noté $\odot$) à l'aide de la seule implication $\rightarrow$ comme montré dans les tables de vérité ci-dessous (pour le $=$, on extrapolera une fois l'idée comprise à cause de la lourdeur des formules à écrire).

b	$\neg b$	$b \to 0$
0	1	1
1	0	0

b_1	b_2	$b_1 \vee b_2$	$b_1 \to 0$	$(b_1 \to 0) \to b_2$
0	0	0	1	0
0	1	1	1	1
1	0	1	0	1
1	1	1	0	1

b_1	b_2	$b_1 \wedge b_2$	$b_1 \to 0$	$(b_1 \to 0) \to 0$	$b_2 \to 0$	$((b_1 \to 0) \to 0) \to (b_2 \to 0)$	$(((b_1 \to 0) \to 0) \to (b_2 \to 0)) \to 0$
0	0	0	1	0	1	1	0
0	1	0	1	0	0	1	0
1	0	0	0	1	1	1	0
1	1	1	0	1	0	0	1

On a : $\neg b = (b \to 0)$, $b_1 \vee b_2 = (b_1 \to 0) \to b_2$ et $b_1 \wedge b_2 = (((b_1 \to 0) \to 0) \to (b_2 \to 0)) \to 0$.