Nombres premiers: mystères et enjeux

Jean-Marie De Koninck

Un peu d'histoire

La suite des nombres premiers commence ainsi:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, \ldots$$

- 300 A.C.: Euclide:
  - *Il existe une infinité de nombres premiers*:
    Preuve: considérer le nombre $p_1 p_2 \ldots p_k + 1$
  - *La factorisation de tout entier $n \geq 2$ est unique:* $n = q_1^{\alpha_1} q_2^{\alpha_2} \ldots q_r^{\alpha_r}$.
- Soit $\pi(x) := \#\{p \text{ premier} : p \leq x\}$, alors
  - 1737: Euler: $\displaystyle\sum_{p \leq x} \frac{1}{p} \sim \log\log x$
  - 1762: Euler: $\pi(x) \approx x/\log x$
  - 1791: Gauss: $\pi(x) \approx x/\log x$
  - 1798: Legendre: $\pi(x) \approx \dfrac{x}{A\log x + B}$
  - 1801: Legendre: $\pi(x) \approx \dfrac{x}{\log x - 1.08366}$
  - 1810: Bessel: $\pi(x) \approx \mathrm{li}(x) := \lim\limits_{\varepsilon \to 0^+} \left(\displaystyle\int_1^{1-\varepsilon} + \int_{1+\varepsilon}^{x} \right) \frac{dt}{\log t} = \int_2^x \frac{dt}{\log t} - 1.04\ldots$
  - 1849: Gauss: $\pi(x) \approx x/\log x$, mais mieux encore, $\pi(x) \approx \mathrm{li}(x)$
  - 1838: Dirichlet croit avoir une preuve de la formule de Legendre
  - 1848: Chebyshev: Il existe des constantes $a < 1 < b$ telles

$$\frac{ax}{\log x} < \pi(x) < \frac{bx}{\log x} \qquad (x \geq x_0)$$

    et de plus, si $\lim\limits_{x\to\infty} \dfrac{\pi(x)\log x}{x}$ existe, alors cette limite est égale à 1.

- La preuve du TNP (*théorème des nombres premiers*): $\pi(x) \sim \dfrac{x}{\log x}$:

– 1737: Euler: Pour tout nombre réel $s > 1$,

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \ldots\right) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}.$$

– Weierstrass (1842) et Riemann (1851): l'invention du prolongement analytique des fonctions holomorphes

– 1859: Riemann prolonge la fonction zêta à tout le plan complexe. Ainsi

$$\zeta(s) = \frac{1}{s-1} + 1 - s\int_1^{\infty} \frac{t - [t]}{t^{s+1}}\, dt \qquad (\Re(s) > 0,\ s \neq 1).$$

Riemann établit son *équation fonctionnelle*

$$\zeta(1-s) = \frac{2}{(2\pi)^s} \cos\left(\frac{\pi s}{2}\right) \Gamma(s)\zeta(s) \qquad (s \neq 0, s \neq 1),$$

de laquelle il découle que $\zeta(-2) = \zeta(-4) = \ldots = 0$. Riemann énonce ce que l'on appelle aujourd'hui l'*hypothèse de Riemann*:

$$\zeta(\sigma + it) = 0 \text{ et } 0 < \sigma < 1 \Longrightarrow \sigma = \frac{1}{2}.$$

Riemann donne des résultats heuristiques menant à TNP.

– 1896: Jacques Hadamard et Charles Jean de la Vallée Poussin: Première preuve rigoureuse du TNP. En particulier

$$\zeta(1 + it) \neq 0 \Longleftrightarrow \pi(x) \sim \frac{x}{\log x}.$$

– 1948: Erdős et Selberg: Première preuve élémentaire du TNP.

Quelques équivalences de l'hypothèse de Riemann (HR)

$$HR \qquad \Longleftrightarrow \qquad \pi(x) - \mathrm{li}(x) = O(\sqrt{x}\log x)$$

$$HR \qquad \Longleftrightarrow \qquad \sum_{d|n} d < e^{\gamma} n \log\log n, \quad \forall n \geq 5041$$

La conjecture *abc* (1985)

Soit $\varepsilon > 0$. Alors il existe une constante $M = M(\varepsilon) > 0$ telle que pour tout triplet d'entiers positifs a, b, c relativement premiers deux à deux et satisfaisant $a + b = c$, on a

$$c < M \left(\prod_{p|abc} p \right)^{1+\varepsilon}.$$

Exemple: L'équation $x^4 + y^4 = z^4$ ne peut avoir qu'un nombre fini de solutions en entiers positifs x, y, z.

En effet, soit x, y, z une solution primitive (avec $x < y$) et soit $\varepsilon > 0$ un petit nombre. Selon la conjecture *abc*, on a

$$z^4 < M \cdot \left(\prod_{p|x^4y^4z^4} p \right)^{1+\varepsilon} \leq M \cdot (xyz)^{1+\varepsilon} < M \cdot z^{3(1+\varepsilon)},$$

de sorte que

$$z^{1-3\varepsilon} < M,$$

ce qui est faux si z est assez grand.

- -

Quel est donc le lien entre ces 3 notions ?

Nombres puissants
$p|n \Rightarrow p^2|n$

Équation de Fermat
$x^p + y^p = z^p$

Nombres premiers
de Wieferich
$2^{p-1} \equiv 1 \pmod{p^2}$

Nombres puissants, nombres premiers de Wieferich et théorème de Fermat

Nombres puissants

Un entier $n > 1$ est dit *puissant* si $p|n \Rightarrow p^2|n$. La suite des nombres puissants commence comme suit:

$$4, 8, 9, 16, 25, 27, 32, 36, 49, 64, 72, 81, 100, 108, 121, 125, 128, 144, 169, 196,$$

$$200, 216, 225, 243, 256, 288, 289, 324, 343, 361, 392, 400, 432, 441, 484, 500 \ldots$$

Existe-t-il une infinité de nombres puissants consécutifs comme 8,9 et 288,289 ?

OUI, cela résulte du fait que l'équation de Pell $x^2 - 2y^2 = 1$ admet une infinité de solutions.

Existe-t-il une infinité de triplets de nombres puissants $n, n+1, n+2$?

On peut démontrer que si la conjecture *abc* est vraie, alors il ne peut exister qu'un nombre fini de tels triplets.

En effet, supposons que $n-1, n, n+1$ sont puissants.

Observons que si m est puissant, alors $\prod_{p|m} p \leq \sqrt{m}$.

On applique donc la conjecture *abc* à l'équation $n^2 = (n^2-1)+1$, auquel cas

$$n^2 < M \cdot \left(\prod_{p|(n-1)n(n+1)} p \right)^{1+\varepsilon} \leq M \cdot \left(\sqrt{n(n^2-1)} \right)^{1+\varepsilon} < M n^{\frac{3}{2}(1+\varepsilon)},$$

de sorte que

$$n^{\frac{1}{2} - \frac{3\varepsilon}{2}} < M,$$

ce qui n'est pas possible si n est assez grand.

DÉFINITION: p est un *nombre premier de Wieferich* si $2^{p-1} \equiv 1 \pmod{p^2}$.

Rappelons que Fermat a démontré que $2^{p-1} \equiv 1 \pmod{p}$ pour tout nombre premier p.

Soit W l'ensemble des nombres premiers de Wieferich.

- 1909: Arthur Wieferich démontre:

  > *Si p est un nombre premier tel que $x^p + y^p = z^p$ pour des entiers positifs x, y, z tels que $p \nmid xyz$, alors $2^{p-1} \equiv 1 \pmod{p^2}$.*

- 1913: Meissner: $1093 \in W$
- 1922: Beeger: $3511 \in W$.
- 2003: Si $p \in W$ et $p \neq 1093, 3511$, alors $p > 1.25 \cdot 10^{15}$.
- Par un argument heuristique, on a $W(x) \sim \log\log x$.
- On ne sait pas démontrer que $|W| = +\infty$, ni que $|W^c| = +\infty$.
- 1988: Silverman: Si la conjecture *abc* est vraie, alors $W^c(x) \gg \log x$.
- On peut démontrer que s'il n'existe pas de triplets puissants $n, n+1, n+2$, alors $|W^c| = +\infty$.
- Si $q^2 | 2^p - 1$, alors q est un nombre premier de Wieferich.

Voici la factorisation de $2^{1092} - 1$:

$$3^2 \cdot 5 \cdot 7^2 \cdot 13^2 \cdot 29 \cdot 43 \cdot 53 \cdot 79 \cdot 113 \cdot 127 \cdot 157 \cdot 313 \cdot 337 \cdot 547 \cdot 911 \cdot 1093^2 \cdot 1249$$
$$\cdot 1429 \cdot 1613 \cdot 2731 \cdot 3121 \cdot 4733 \cdot 5419 \cdot 8191 \cdot 14449 \cdot 21841121369 \cdot 224771$$
$$\cdot 503413 \cdot 1210483 \cdot 1948129 \cdot 22366891 \cdot 108749551 \cdot 112901153 \cdot 23140471537$$
$$\cdot 25829691707 \cdot 105310750819 \cdot 467811806281 \cdot 4093204977277417 \cdot 8861085190774909$$
$$\cdot 556338525912325157 \cdot 2757007179515465669468544 97 \cdot 86977595801949844993$$
$$\cdot 2926531131471572057791275268 27 \cdot 3194753987813988499397428643895659569.$$

D'une fausse preuve d'une conjecture
à la preuve d'une autre conjecture

- Il découle du TNP qu'en moyenne $p_{k+1} - p_k$ est $\log p_k$, ce qui veut dire que l'écart entre deux nombres premiers consécutifs de taille x est en moyenne $\log x$.
- Il est facile de démontrer que $\limsup_{k\to\infty}(p_{k+1} - p_k) = +\infty$:

  il suffit de considérer la suite de nombres composés $(k+1)! + 2, (k+1)! + 3, \ldots, (k+1)! + (k+1)$.
- On ne sait pas démontrer que $\liminf_{k\to\infty}(p_{k+1} - p_k)$ existe, mais on croit que $\liminf_{k\to\infty}(p_{k+1} - p_k) = 2$ (conjecture des nombres premiers jumeaux).
- 1940: Paul Erdős: Il existe une constante positive $C < 1$ telle que $p_{k+1} - p_k < C(\log p_k)$ pour une infinité de k.
- 1965: $C \leq 0.90625$ (trois chinois).
- 1966: $C \leq 0.467$ (Davenport & Bombieri).
- 1988: $C \leq 0.248$ (Maier).
- Mars 2003: D. Goldston et C.Y. Yildirim affichent sur le WEB une preuve qu'il existe une constante positive $C < 1$ telle que

$$p_{k+1} - p_k < (\log p_k)^C.$$

- Juillet 2003: A.Granville et Soundarajan: En utilisant l'argumentation de Goldston & Yildirim, on peut obtenir que $\liminf_{k\to\infty}(p_{k+1} - p_k) = 12$.

  Ils trouvent une ERREUR dans le calcul d'une intégrale multiple.

D É C E P T I O N

\- -

Les nombres premiers en progression arithmétique

La suite 199, 409, 619, 829, 1039, 1249, 1459, 1669, 1879, 2089 est une suite de $k = 10$ termes dont chacun est un nombre premier (la différence entre les termes consécutifs est égale à 210).

- ≈ 1770: Lagrange et Waring: combien grand peut-on choisir k ?
- 1923: Hardy et Littlewood conjecturent que la suite des nombres premiers contient des progressions arithmétiques arbitrairement longues. De plus, ils donnent un argument heuristique à l'effet que le nombre de k-tuplets $(p_1, \ldots, p_k)$ appartenant à une même progression arithmétique (avec $p_i \leq x$) est

$$\sim C_k \frac{x^2}{\log^k x} \quad \text{pour une certaine constante } C_k.$$

- 1939: van der Corput démontre le cas $k = 3$, i.e. qu'il existe une infinité de triplets de nombres premiers en progression arithmétique.
- 1981: Heath-Brown: $k = 4$, mais avec 3 termes premiers et un 4^e terme premier ou le produit de deux nombres premiers.
- 1993: Moran, Pritchard, Thyssen: Les nombres

$$11\,410\,337\,850\,553 + 4\,609\,098\,694\,200 \cdot k, \quad \text{pour } k = 0, 1, 2, \ldots, 21$$

  sont tous premiers.

- 2003: Frind: Les nombres

$$376\,859\,931\,192\,959 + 18\,549\,279\,769\,020 \cdot k, \quad \text{pour } k = 0, 1, 2, \ldots, 21$$

  sont tous premiers.

- 2004: Ben Green et Terence Tao: *La suite des nombres premiers contient des progressions arithmétiques de longueur k pour tout entier positif k donné.*

  La preuve de Green & Tao utilise trois outils:

  1. Un théorème de Szemeredi: *Tout sous-ensemble d'entiers de densité positive contient des progressions arithmétiques de longueur arbitraire.*
  2. Un ensemble pseudo-aléatoire de densité positive contient des progressions arithmétiques de longueur arbitraire.
  3. Une technique utilisée par Goldston & Yildirim.

  La preuve est disponible à `http://arxiv.org/abs/math.NT/0404188`

La recherche des nombres premiers

Les tests de primalité

- La division par de petits nombres premiers: si n n'est pas premier, il est "dénoncé" par un diviseur premier $\leq \sqrt{n}$.
- Le théorème de Wilson: Si $n \geq 2$, alors

$$(n-1)! \equiv -1 \pmod{n} \iff n \text{ est premier.}$$

- La réciproque du petit théorème de Fermat. En effet, selon le petit théorème de Fermat,

$$p \text{ premier}, (a,p) = 1 \qquad \Longrightarrow \qquad a^{p-1} \equiv 1 \pmod{p}.$$

  Même si la réciproque est fausse, elle est presque toujours vérifiée.

- Le test de Lucas:

$$\begin{array}{l} a^{n-1} \equiv 1 \pmod{n} \\ \\ a^{(n-1)/p} \not\equiv 1 \pmod{n} \quad \forall p | n-1 \end{array} \qquad \Longrightarrow \quad n \text{ est premier}$$

- Le test de Lucas-Lehmer pour tester la primalité des nombres de Mersenne:

  Soit $M_p = 2^p - 1$, où p est un nombre premier impair. Soit $s_1 = 4$ et, pour $k \geq 2$, soit $s_k \equiv s_{k-1}^2 - 2 \pmod{M_p}$. Alors

$$M_p \text{ est premier } \iff M_p | s_{p-1}.$$

  C'est ainsi qu'on sait que

$$2^{24\,036\,583} - 1 \text{ est premier.}$$

- Le test de Pépin pour tester la primalité des nombres de Fermat:

  Soit k un entier positif et $F_k = 2^{2^k} + 1$. Alors

$$3^{\frac{F_k - 1}{2}} \equiv -1 \pmod{F_k} \iff F_k \text{ est premier.}$$

  On sait que F_0, F_1, F_2, F_3, F_4 sont premiers et que F_n est composé pour $n = 5, 6, \ldots, 32$.

- Le test probabiliste de Miller-Rabin:

  Si un nombre impair n est déclaré composé, il l'est.

  S'il est déclaré premier, alors la probabilité d'une erreur est d'environ $\frac{1}{4^k}$, où k est le nombre de fois qu'on a effectué le test.

- Le test de probabilité utilisant les courbes elliptiques.

Comment mesure-t-on la rapidité d'un algorithme ?

Soit n un nombre de r chiffres ($r \approx \log n$).

- Temps polynomial: r^c
- Temps exponentiel: $c^{f(r)}$
- Temps sous-exponentiel: $\exp\{c(\log n)^a(\log\log n)^{1-a}\}$
  pour un certain $0 < a < 1$.

Rapidité des tests de primalité:

Le test	Nombre d'opérations	Temps d'exécution
Division par les petits nombres premiers	$\sqrt{n} = e^{(\log n)/2}$	exponentiel
Le test de Wilson	$n = e^{\log n}$	exponentiel
Courbes elliptiques	$\log^6 n$ en moyenne	polynomial en moyenne
Miller-Rabin	$\log^4 n$ (sous GRH)	polynomial (GRH)
Adleman, Pomerance, Rumely	$(\log n)^{\log\log\log n}$	presque polynomial

Enfin un test de primalité en temps polynomial !

- Août 2002
- Trois chercheurs de l'*Indian Institute of Technology*:
  Manindra Agrawal et ses deux étudiants Neeraj Kayal et Nitin Saxena
- Découverte d'un test de primalité dont le nombre d'étapes est borné approximativement par d^{12}, où d est le nombre de chiffres du nombre testé.
- En 2004, le nombre d'étapes est réduit à d^6.
- Outils:
  - le petit théorème de Fermat, i.e. $a^p \equiv a \pmod{p}$ pour tout entier positif a et tout nombre premier p.
  - $(x+y)^p \equiv x^p + y^p \pmod{p}$ pour tout nombre premier p.
- Disponible sur le WEB à l'adresse `www.cse.iitk.ac.in/primality.pdf`.

Le temps de factorisation d'un nombre

La méthode	le nombre d'opérations	année	auteur	le nombre de chiffres
La division par de petits nombres premiers	$\frac{2\sqrt{n}}{\log n}$	l'an 0		20
Test de Fermat $n = x^2 - y^2$	rapide pour $n = pq$, avec p et q proches	1650	Fermat	20
$x^2 \equiv y^2 \pmod{n}$ (amélioration du test de Fermat)	$e^{\sqrt{c \log n \log\log n}}$	1920	Kraitchik	30
Méthode des fractions continues	$e^{\sqrt{2 \log n \log\log n}}$	1979	Brillhart Morisson	50
Crible quadratique	$e^{\sqrt{\log n \log\log n}}$	1990	Pomerance	116
Number Field Sieve (pour les nombres n de la forme $n = p^a + b$)	$e^{(\log n)^{1/3}(\log\log n)^{2/3}}$	1996	H. Lenstra	130
Courbes elliptiques (trouve les facteurs de moins de 40 chiffres)	$e^{\sqrt{\log n \log\log n}}$	1985	H. Lenstra	150

La méthode RSA et la méthode des courbes elliptiques – un aperçu

La méthode RSA, initiée par Rivest, Shamir et Adleman en 1978, est une méthode de cryptographie à clés publiques qui exploite le fait qu'il est facile d'identifier des nombres premiers de 150 chiffres, alors qu'il est pratiquement impossible de factoriser des entiers n de 300 chiffres de la forme $n = pq$.

Toutefois, son grand inconvénient est qu'elle nécessite beaucoup d'espace-mémoire pour sauvegarder la clé privée.

Or ce n'est pas le cas avec la méthode des courbes elliptiques de H. Lenstra, laquelle assure pratiquement le même niveau de sécurité avec cette fois des clés beaucoup plus petites.

Ainsi, comparons la méthode RSA avec celle de Lenstra.

La méthode RSA repose sur la fonction de multiplication (ici $\mathcal{P}$ désigne l'ensemble des nombres premiers impairs)

$$\begin{array}{rcl} f \ : \ \mathcal{P} \times \mathcal{P} & \to & \mathbf{N} \\ (p, q) & \mapsto & n = pq, \end{array}$$

qui, à deux nombres premiers p, q associe leur produit $n = pq$. Or le calcul de f^{-1} est en général très difficile, voire même quasi-impossible pour de grands nombres n qui sont produits de deux nombres premiers également très grands.

La méthode des courbes elliptiques considère un point rationnel P pris sur une courbe elliptique E, $y^2 = x^3 + ax + b$, définie sur un corps fini (en l'occurence $\mathbf{Z}_p = \{0, 1, 2, \ldots, p-1\}$), et applique alors la fonction

$$\begin{array}{rcl} g \ : \ \mathbf{N} & \to & E \\ k & \mapsto & kP = \underbrace{P + P + \ldots + P}_{k}, \end{array} \tag{1}$$

où kP est tout simplement le point P additionné k fois, jusqu'à ce que l'addition mène à une impasse. Or, tout comme dans le cas de RSA, la fonction inverse de (**??**) est très difficile à calculer: c'est ce qu'on appelle le *problème du logarithme discret*.

Alors que le récent "défi RSA", qui consistait à factoriser un nombre de 155 chiffres, a été relevé avec succès, le défi pour calculer l'inverse de (**??**) concernait un nombre k de 33 chiffres. La raison est que l'on dispose d'une foule d'algorithmes relativement efficaces pour factoriser un nombre n donné (disons, de moins de 100

chiffres), alors qu'on en connaît très peu pour résoudre le problème du logarithme discret.

Le problème P versus NP

D'abord, quelques problèmes:

1. Problème de tri: Étant donné un ensemble d'entiers, les ranger en ordre croissant.
2. Problème de l'emploi du temps: Étant donnée une liste de cours devant être offerts, la grille de disponibilité des étudiants et celle des professeurs, construire un emploi du temps qui ne contient aucun conflit d'horaire.
3. Problème du commis voyageur: Étant donné un certain nombre de villes et les coûts de déplacement d'une ville à l'autre, quel est l'itinéraire le moins coûteux permettant de traverser chacune d'elle, une et une seule fois ? Le problème consiste donc à trouver l'option la plus économique parmi un nombre fini, mais astronomiquement élevé, de possibilités.

Il s'agit là de problèmes d'optimisation combinatoire.

On désigne par **P** la classe des problèmes que l'on peut résoudre au moyen d'un algorithme dans un temps polynomial.

On désigne par **NP** la classe des problèmes que l'on peut résoudre au moyen d'un algorithme non déterministe et vérifier sa solution en un temps polynomial.

Deux exemples:

1. Le problème de la factorisation d'un entier positif donné n.
2. La recherche d'une personne dans un party.

$$\mathbf{P} \subset \mathbf{NP}, \qquad \text{mais} \quad \mathbf{NP} \subset \mathbf{P} \quad ??.$$