Introduction à la théorie des nombres

Dimitris Koukoulopoulos

Université de Montréal

Dernière mise-à-jour : 18 décembre 2017

Table des matières

Notation

Les symboles $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ et $\mathbb{C}$ dénotent les ensembles des nombres naturels, entiers, rationnels, réels et complexes, respectivement. On n'inclut pas le nombre 0 à l'ensemble de nombres naturels.

Le symbole $\log x$ dénote toujours le logarithme naturel de x (qui est souvent noté par $\ln x$ dans la bibliographie).

Étant donné un ensemble $A \subset \mathbb{C}$, on dénote par $A[x]$ l'ensemble des polynômes $f(x) = a_0 + a_1 x + \cdots + a_d x^d$ dont les coefficients $a_0, a_1, \ldots, a_d$ appartiennent à A. Étant donné un polynôme $f(x) \in \mathbb{C}[x]$ qui n'est pas égal à 0, on peut toujours l'écrire uniquement comme $f(x) = a_0 + a_1 x + \cdots + a_d x^d$, où $a_d \neq 0$. Le nombre d est appelé le degré de f.

Étant donné un nombre réel x, on utilise la notation $\lfloor x \rfloor$ pour dénoter sa partie entier, qui est défini d'être le plus grand entier n qui est plus petit que x, c'est-à-dire

$$\lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \leq x\}.$$

De plus, on utilise la notation $\{x\}$ pour dénoter la partie fractionnel de x, qui est défini par

$$\{x\} = x - \lfloor x \rfloor .$$

On remarque ici que la partie entier de x est l'entier unique n satisfaisant les inégalités $n \leq x < n + 1$. Une autre propriété de la partie entier qu'on utilisera beaucoup est que si $x > 0$, alors

$$\lfloor x \rfloor = \#\{n \in \mathbb{Z} : 1 \leq n \leq x\}.$$

La lettre p dénote toujours un nombre premier (définition 1.25). La notation $a|b$ veut dire que a divise b (définition 1.1). Si p est un nombre premier et $v \in \mathbb{N}$, alors on écrit $p^v \| n$ si $p^v | n$ et $p^{v+1} \nmid n$.

La notation (a, b) pourrait signifier le plus grand commun diviseur de a et b (définition 1.4), l'intervalle ouvert dont les limites sont a et b (c'est-à-dire l'ensemble $\{x \in \mathbb{R} : a < x < b\}$), ou le pair de nombres a et b. Le contexte déterminera sa signification.

De même, le symbole $[a, b]$ pourrait signifier le plus petit commun multiple de a et b (définition 1.20) ou l'intervalle fermé dont les limites sont a et b (c'est-à-dire l'ensemble $\{x \in \mathbb{R} : a \leq x \leq b\}$).

Première partie

Méthodes algébriques

Chapitre 1

La structure multiplicative des entiers

1.1 Divisibilité

Définition 1.1. Soit $a, b \in \mathbb{Z}$ avec $a \neq 0$. On dit que a divise b et on écrit $a|b$ si $b/a \in \mathbb{Z}$, c'est-à-dire si il existe $k \in \mathbb{Z}$ tel que $b = ka$.

Dans ce cas, on dit aussi que b est divisible par a, ou que a est un diviseur de b ou que b est un multiple de a.

Le lemme suivant établit quelques propriétés de base de la notion de la divisibilité :

Lemme 1.2. *Soit $a, b \in \mathbb{Z}$ avec $a \neq 0$.*

(a) Si $a|b$, alors soit $b = 0$ soit $|b| \geq a$.
(b) Si $a|b$ et $a|c$, alors $a|(bx + cy)$ pour tout $x, y \in \mathbb{Z}$.
(c) Si $a|b$ et $c|d$, alors $ac|bd$.
(d) Si $a|b$ et $b|c$, alors $a|c$.
(e) Si $c \neq 0$ et $ac|bc$, alors $a|b$.

On nombre ne divise pas toujours un autre. Par exemple, $3|6$ mais $3 \nmid 7$. En général, on a le théorème suivant.

Théorème 1.3 (Division euclidienne)**.** *Soit $a, b \in \mathbb{Z}$ avec $a \neq 0$. Il existe $q, r \in \mathbb{Z}$ uniques tels que*

$$b = qa + r \quad et \quad 0 \leq r < |a|.$$

Les nombres q et r sont appelés le quotient *et le* reste *de la division de b par a, respectivement.*

Démonstration. Puisque $qa = (-q)(-a)$ et $|-a| = |a|$, on peut supposer, sans perte de généralité, que $a \geq 1$.

On commence en prouvant l'existence de q et de r. On considère les nombres $n_j := b - ja$, $j \in \mathbb{Z}$. Observons que $n_{j+1} - n_j = -a$, c'est-à-dire les nombres n_j sont une progression arithmétique d'étape a. Donc il existe $j_0 \in \mathbb{Z}$ tel que $n_{j_0} \in [0, a)$. (Sinon, on pourrait trouver j tel que $n_{j+1} < 0$ et $n_j \geq a$ et, par la suite, $n_j - n_{j+1} > a$, ce qui est impossible.) On pose $r = n_{j_0} \in [0, a)$ et on observe que $b = j_0 a + n_{j_0}$. Donc l'existence de q et r découle en posant $q = j_0$ et $r = n_{j_0}$.

Finalement, on montre l'unicité de q et r. Supposons que $b = qa + r = q'a + r'$ avec $0 \leq r, r' < a$. Donc

$$(q - q')a = r' - r,$$

ce qui implique que $a|r' - r$. Cependant $|r' - r| < a$ par notre hypothèse que $0 \leq r, r' < a$. Donc lemme 1.2(a) implique que $r' = r$. Donc on trouve que $(q - q')a = 0$ et, puisque $a \geq 1$, alors $q' = q$ aussi. □

Exercices

EXERCICE 1.1. Soit $a, b \in \mathbb{N}$. Montrez que $a = b$ si et seulement si $a|b$ et $b|a$.

EXERCICE 1.2. Soient $a, b \in \mathbb{N}$ tels que $1/a + 1/b$ est entier. Montrez que soit $a = b = 1$ soit $a = b = 2$.

1.2 Le plus grand commun diviseur

Afin d'étudier la relation multiplicative de deux nombres, on introduit la notion de leur *plus grand commun diviseur* :

Définition 1.4. Soient $a, b \in \mathbb{Z}$ qui ne sont pas les deux égaux à 0. Le plus grand commun diviseur de a et b, dénoté par (a, b), est défini d'être le plus grand nombre naturel qui divise a et b. C'est-à-dire

$$(a, b) = \max\{d \in \mathbb{N} : d|a \text{ et } d|b\}.$$

Remarque 1.5. Notre hypothèse que soit $a \neq 0$ soit $b \neq 0$ implique que l'ensemble $\{d \in \mathbb{N} : d|a \text{ et } d|b\}$ est non-vide et fini. Donc (a, b) est bien défini.

On peut utiliser un algorithme rapide pour calculer le plus grand commun diviseur de deux nombres qui se base sur la division euclidienne. L'observation-clé est donnée au lemme suivant :

Lemme 1.6. *Pour tout $a, b, q \in \mathbb{Z}$ avec $a \neq 0$, on a que $(a, b) = (a, b - qa)$. En particulier, si r est le reste dans la division de b par a, on a que $(a, b) = (a, r)$.*

Démonstration. Si $d|a$ et $d|b$, alors $d|b - qa$ aussi. Réciproquement, si $d|a$ et $d|b - qa$, alors $d|qa + (b - qa) = b$. Donc on trouve que

$$\{d \in \mathbb{N} : d|a \text{ et } d|b\} = \{d \in \mathbb{N} : d|a \text{ et } d|b - qa\},$$

ce qui termine la démonstration. □

Ce lemme nous amène à l'algorithme euclidien pour calculer le plus grand commun diviseur de deux nombres. Soient $a, b \in \mathbb{N}$. Sans perde de généralité, on suppose que $b \geq a$. On écrit $b = qa + r$ avec $0 \leq r < a$ pour que $(a, b) = (a, r)$, ce qui nous permet de remplacer le pair (a, b) avec un nouveau par, (a, r), dont le plus petit élément est strictement plus petit qu'on

avait avant. Bien sûr, on peut répéter cette procedure : on a que $a = q'r + r'$ avec $0 \leq r' < r$ et que $(a, r) = (r', r)$, ce qui nous permet de remplace le pair (a, r) avec le nouveau pair (r', r) pour que $\min\{r, r'\} = r' < r = \min\{a, r\}$. En continuant dans cette façon, on doit arriver à un pair dont un de deux membres est égal à 0. Le terme non-zéro du premier tel pair serait le plus grand commun diviseur de a et b. Plus formellement, il existe un n tel que

$$\begin{aligned}
b_0 := b, \quad b_1 := a, \quad & b_0 = q_1 b_1 + r_1, \quad 0 < r_1 < b_1 \\
& \rightsquigarrow \quad (a, b) = (b_0, b_1) = (b_1, r_1) \\
b_2 := r_1 < b_1, \quad & b_1 = q_2 b_2 + r_2, \quad 0 < r_2 < b_2 \\
& \rightsquigarrow \quad (a, b) = (b_1, b_2) = (b_2, r_2) \\
b_3 := r_2 < b_2, \quad & b_2 = q_3 b_3 + r_3, \quad 0 < r_3 < b_3 \\
& \rightsquigarrow \quad (a, b) = (b_2, b_3) = (b_3, r_3) \\
& \vdots \\
b_{n-1} := r_{n-2} < b_{n-2}, \quad & b_{n-2} = q_{n-1} b_{n-1} + r_{n-1}, \quad 0 < r_{n-1} < b_{n-1} \\
& \rightsquigarrow \quad (a, b) = (b_{n-2}, b_{n-1}) = (b_{n-1}, r_{n-1}) \\
b_n := r_{n-1} < b_{n-1}, \quad & b_{n-1} = q_n b_n + r_n, \quad r_n = 0 \\
& \rightsquigarrow \quad (a, b) = (b_{n-1}, b_n) = (b_n, 0) = b_n.
\end{aligned}$$

Exemple 1.7. Calculons $(91, 65)$. On a que

$$\begin{aligned}
91 = 1 \cdot 65 + 26 \quad & \rightsquigarrow \quad (91, 65) = (65, 26) \\
65 = 2 \cdot 26 + 13 \quad & \rightsquigarrow \quad (91, 65) = (65, 26) = (26, 13) \\
26 = 2 \cdot 13 \quad & \rightsquigarrow \quad (91, 65) = (26, 13) = (13, 0) = 13.
\end{aligned}$$

Exemple 1.8. Calculons $(1568, 686)$. On a que

$$\begin{aligned}
1568 = 2 \cdot 686 + 196 \quad & \rightsquigarrow \quad (1586, 686) = (686, 196) \\
686 = 3 \cdot 196 + 98 \quad & \rightsquigarrow \quad (1586, 686) = (686, 196) = (196, 98) \\
196 = 2 \cdot 98 \quad & \rightsquigarrow \quad (1586, 686) = (196, 98) = (98, 0) = 98.
\end{aligned}$$

L'algorithme euclidien nous permet de déduire le théorème suivant qui est très utile.

Théorème 1.9. *Soient $a, b \in \mathbb{Z}$ qui ne sont pas les deux égaux à zéro. Alors, le plus grand commun diviseur de a et b est une combinaison linéaire de a et b, c'est-à-dire il existe deux entiers x et y tels que*

$$(a, b) = ax + by.$$

Démonstration. Le cas où $a = 0$ ou $b = 0$ est facile. Supposons maintenant que a et b ne sont pas zéro. Sans perde de généralité, on peut supposer que $a, b \in \mathbb{N}$. Donc, en utilisant la notation au-dessus, on a que $(a, b) = b_n$, où

$$b_0 = q_1 b_1 + b_2, \quad b_1 = q_2 b_2 + b_3, \quad b_2 = q_3 b_3 + b_4, \quad \ldots \quad b_{n-2} = q_{n-1} b_{n-1} + b_n,$$

Avec $b_0 = b$ et $b_1 = a$. $(a, b) = b_n = b_{n-2} - q_{n-1} b_{n-1}$ est une combinaison linéaire de b_{n-1} et de b_{n-2}. Puisque $b_{n-1} = b_{n-3} - q_{n-2} b_{n-2}$ est une combinaison linéaire de b_{n-2} et de b_{n-3}, on

trouve que (a, b) a la même propriété. On continue en remplaçant b_{n-2} par $b_{n-4} - q_{n-3}b_{n-3}$ pour trouver que (a, b) est une combinaison linéaire de b_{n-3} et de b_{n-4}. De façon inductive, on conclut que (a, b) est une combinaison linéaire de b_0 et de b_1, comme affirmé. □

Remarque 1.10. La dernière démonstration nous permet de trouver les nombres x et y, si on sait les restes $q_1, q_2, \ldots, q_{n-1}$ qui apparait dans l'algorithme euclidien. En effet, on a que

$$\begin{aligned}(a, b) = b_n &= b_{n-2} - q_{n-1}b_{n-1} \\ &= b_{n-2} - q_{n-1}(b_{n-3} - q_{n-2}b_{n-2}) \\ &= (1 + q_{n-1}q_{n-2})b_{n-2} - q_{n-1}b_{n-3} \\ &= (1 + q_{n-1}q_{n-2})(b_{n-4} - q_{n-3}b_{n-3}) - q_{n-1}b_{n-3} \\ &= (1 + q_{n-1}q_{n-2})b_{n-4} - [(1 + q_{n-1}q_{n-2})q_{n-3} + q_{n-1}]b_{n-3} \\ &= \cdots = ax + by.\end{aligned}$$

On applique cet algorithme aux paires des exemples 1.7 et 1.8 au-dessous. On a que

$$\begin{aligned}(91, 65) = 13 &= 65 - 2 \cdot 26 \\ &= 65 - 2 \cdot (91 - 1 \cdot 65) = -2 \cdot 91 + 3 \cdot 65.\end{aligned}$$

De même, on a que

$$\begin{aligned}98 = (1586, 686) &= 686 - 3 \cdot 196 \\ &= 686 - 3 \cdot (1586 - 2 \cdot 686) = -3 \cdot 1586 + 7 \cdot 686.\end{aligned}$$

Remarque 1.11. L'algorithme euclidien est un algorithme très rapide. Supposons que $1 \leq a \leq b$ et que $a \nmid b$, pour que $n \geq 2$. Pour tout $m \in \{0, 1, \ldots, n-2\}$, on a que $b_m = q_{m+1}b_{m+1} + b_{m+2}$. Aussi, on sait que $b_{m+1} \geq b_{m+2}$ et que $q_{m+1} \geq 1$. Donc $b_m \geq 2b_{m+2}$. Un argument itératif montre que $b_{2k} \leq b_2/2^k \leq a/\sqrt{2}^{2k}$ et que $b_{2k-1} \leq b_1/2^{k-1} = \sqrt{2}a/\sqrt{2}^{2k-1}$, pour tout $k \geq 1$. En particulier, on a que $1 \leq b_{n-1} \leq \sqrt{2}a/\sqrt{2}^{n-1}$, ce qui implique que $2^{n/2} \leq 2a$. Donc $n \leq 2\log(2a)/\log 2$, c'est-à-dire l'algorithme termine dans $\leq 2\log(2a)/\log 2$ étapes.

Remarque 1.12. Étant donné un nombre a, on peut l'écrire dans la base binaire, soit $a = 2^k + d_{k-1}2^{k-1} + \cdots + d_1 2 + d_0$, avec $d_0, d_1, \ldots, d_{k-1} \in \{0, 1\}$. On a que

$$2^k \leq a \leq 2^k + 2^{k-1} + \cdots + 2 + 1 = \frac{2^{k+1} - 1}{2 - 1} < 2^{k+1}.$$

Donc

$$k \leq \frac{\log a}{\log 2} < k + 1,$$

c'est-à-dire $k = \lfloor \log a/\log 2 \rfloor$. C'implique que $k \approx \log a/\log 2$ ou que $\log a \approx 2\log 2$. Alors l'algorithme euclidien a besoin d'à peu près $2k$ étapes, qui est une fonction linéaire dans le nombre de chiffres binaires de a (qui est égal à $k + 1$). On dit qu'un tel algorithme est un algorithme linéaire. Si l'algorithme a besoin de k^d étapes, pour un $d \in \mathbb{N}$, on dit que

l'algorithme est polynomial (comme une fonction du nombre des chiffres binaires du donné a). Finalement, si le nombre d'étapes dont l'algorithme a besoin pour terminer est une fonction qui augment plus rapidement que chaque polynôme (par exemple e^k ou $e^{\sqrt{k}}$ ou 2^{k^2}), alors on dit que l'algorithme est exponentiel (toujours comme une fonction du nombre des chiffres binaires du donné a).

La puissance du théorème 1.9 est révélée dans les démonstration des résultats suivants, qui sont très utils.

Lemme 1.13. *Soient $d, a, b \in \mathbb{N}$. Alors $d|a$ et $d|b$ si et seulement si $d|(a, b)$. C'est-à-dire, chaque commun diviseur de a et b divise leur plus grand commun diviseur.*

Démonstration. Supposons que $d|a$ et que $d|b$. On a que $(a, b) = ax + by$ pour quelques $x, y \in \mathbb{Z}$. Donc $d|ax + by = (a, b)$, du lemme 1.2(b).

Réciproquement, supposons que $d|(a, b)$. Par définition, on a que $(a, b)|a$ et que $(a, b)|b$. Donc lemme 1.2(d) implique que $d|a$ et que $d|b$. □

Lemme 1.14. $(ab, ac) = a \cdot (b, c)$

Démonstration. Soit $d = (ab, ac)$ et $e = (b, c)$. On veut montrer que $d = ae$. Il suffit de montrer que $d|ae$ et que $ae|d$.

On a que $e|b$ et que $e|c$. Donc $ae|ab$ et $ae|ac$. Alors lemme 1.13 implique que $ae|(ab, ac) = d$.

Réciproquement, on a que $a|ab$ et que $a|ac$. Par la suite, $a|(ab, ac) = d$. On écrit $d = ad'$ et on observe que $ad'|ab$ et que $ad'|ac$. Alors on trouve que $d'|b$ et que $d'|c$, ce qui implique que $d'|(b, c) = e$. Donc $d = ad'|ae$, ce qui conclut la démonstration. □

Proposition 1.15. *Soit $a, b \in \mathbb{Z}$. Si $d = (a, b)$, alors il existe $k, \ell \in \mathbb{N}$ tels que $a = dk$, $b = d\ell$ et $(k, \ell) = 1$.*

Démonstration. On sait que $d|a$ et $d|b$, donc il existe k, ℓ tels que $a = dk$ et $b = d\ell$. De plus, on a que

$$d = (a, b) = (dk, d\ell) = d \cdot (k, \ell),$$

d'après le lemme 1.14. Donc $(k, \ell) = 1$, comme affirmé, ce qui conclut la démonstration. □

Proposition 1.16 (Lemme d'Euclid)**.** *Soient $a, b, c \in \mathbb{N}$ avec $(a, b) = 1$. Si $a|bc$, alors $a|c$.*

Démonstration. Puisque $(a, b) = 1$, alors il existe $x, y \in \mathbb{Z}$ tels que $1 = ax + by$. Donc $c = acx + bcy$. On a que $a|a$ et que $a|bc$. Par conséquent, on trouve que a divise leur combinaison linéaire $acx + bcy = c$. Ceci termine la démonstration. □

Lemme 1.17. *Soient $a, b, c \in \mathbb{N}$ avec $(a, b) = 1$. Si $a|c$ et $b|c$, alors $ab|c$.*

Démonstration. Soit $c = ka$. On a que $b|c = ka$ et que $(a, b) = 1$. Donc le lemme d'Euclid nous donne que $a|k$. Alors, $k = a\ell$ pour un $\ell \in \mathbb{N}$. C'implique que $c = \ell ab$, ce qui conclut la démonstration. □

Proposition 1.18. *Soient $a, b, c \in \mathbb{N}$ avec $(a, b) = 1$. Si $(a, c) = (b, c) = 1$, alors $(ab, c) = 1$.*

Démonstration. Soit $d = (ab, c)$. L'idée est que d ne peut pas avoir une partie commune ni avec a ni avec b car $d|c$. En effet, soit $d_1 = (d, a)$. On a que $d_1|d|c$ et donc $d_1|c$. Alors $d_1|(a, c) = 1$, ce qui implique que $d_1 = 1$. Puisque $(d, a) = 1$ et $d|ab$, alors le lemme d'Euclid implique que $d|b$. Mais dans ce cas on a que $d|(b, c) = 1$, d'où on déduit que $d = 1$, comme affirmé. □

On conclut cette section avec une généralisation du plus grand commun diviseur.

Définition 1.19. Soient $a_1, \ldots, a_k \in \mathbb{Z}$ qui ne sont pas tous zéros. Le plus grand commun diviseur de $a_1, \ldots, a_k$, dénoté par $(a_1, \ldots, a_k)$, est le plus grand nombre naturel qui divise chacun des nombres $a_1, a_2, \ldots, a_k$. C'est-à-dire

$$(a_1, \ldots, a_k) = \max\{d \in \mathbb{N} : d|a_j \ (1 \leq j \leq k)\}.$$

Exercices

EXERCICE 1.3. Calculez (a, b) et trouvez x et y tels que $ax + by = (a, b)$ quand

(a) $a = 1287$ et $b = 4004$,
(b) $a = 3185$ et $b = 1232$,
(c) $a = 5021$ et $b = 1728$.

EXERCICE 1.4. Soient $1 \leq k \leq n$ et $d = (n, k)$. Montrez que $\frac{n}{d} | \binom{n}{k}$.

EXERCICE 1.5. Montrez que si $(a, c) = 1$, alors $(ab, c) = (b, c)$.

EXERCICE 1.6. Soit $a, b \in \mathbb{Z}$ tels que $(a, b) = 1$. Montrez que

(a) $(a + b, a - b) = 1$ ou 2.
(b) $(2a + b, a + 2b) = 1$ ou 3.
(c) $(a + b, a^2 - 3ab + b^2) = 1$ ou 5.

Classifiez quand chaque cas arrive.

EXERCICE 1.7. Montrez que $((a + b)^2, a^2 + ab + b^2) = (a, b)^2$, pour tous $a, b \in \mathbb{N}$.

EXERCICE 1.8. (a) Montrez que $(a_1, \ldots, a_k) = ((a_1, \ldots, a_{k-1}), a_k)$.
(b) Montrez que $d|a_j$ pour chaque $j \in \{1, \ldots, k\}$ si et seulement si $d|(a_1, \ldots, a_k)$.

1.3 Le plus petit commun multiple

On peut définir une notion qui est duale à la notion du plus grand commun diviseur : c'est le plus petit commun multiple de deux nombres.

Définition 1.20. Soient $a, b \in \mathbb{Z} \setminus \{0\}$. Le plus petit commun multiple de a et b, dénoté par $[a, b]$, est défini d'être le plus petit nombre nombre naturel qui est divisible par a et par b. C'est-à-dire

$$[a, b] = \min\{m \in \mathbb{N} : a|m \text{ et } b|m\}.$$

Remarque 1.21. Notre hypothèse que $a, b \neq 0$ implique que l'ensemble $\{m \in \mathbb{N} : a|m \text{ et } b|m\}$ est non-vide, puisque $|ab|$ y appartient toujours. Donc $[a, b]$ est bien défini.

Le théorème suivant établit une relation simple entre le plus grand commun diviseur et le plus petit commun multiple de deux nombres.

Théorème 1.22. *Si $a, b \in \mathbb{N}$, alors $[a, b](a, b) = ab$.*

Démonstration. Soit $m = [a, b]$ et $d = (a, b)$. Il existe $k, \ell \in \mathbb{N}$ tels que $a = dk$, $b = d\ell$ et $(k, \ell) = 1$. Avec cette notation, il suffit de montrer que $m = [dk, d\ell] = dk\ell$. Bien sur, $dk\ell$ est un commun multiple de dk et de $d\ell$. Donc $m \leq dk\ell$. De plus, on a que $m = dkt$ et que $m = dks$, pour quelques $t, s \in \mathbb{N}$. En particulier, $dk|m$, $k|m/d$ et $\ell|m/d$. Puisque $(k, \ell) = 1$, le lemme 1.17 implique que $k\ell|m/d$. En particulier, $k\ell \leq m/d$, ce qui conclut la preuve que $m = dk\ell$. □

Lemme 1.23. *Si $a, b, m \in \mathbb{N}$, alors $a|m$ et $b|m$ si et seulement si $[a, b]|m$. C'est-à-dire, chaque commun multiple de a et de b est divisible par le plus petit commun multiple de a et de b.*

Démonstration. Soit $(a, b) = d$, $a = dk$ et $b = d\ell$, pour que $(k, \ell) = 1$. Théorème 1.3 implies that $[a, b] = dk\ell$. So it suffices to show that $dk\ell|m$. On a que $a = dk|m$. En particulier, $d|m$ et $k|m/d$. De même, on trouve que $\ell|m/d$. Puisque $(k, \ell) = 1$, alors $k\ell|m/d$, ce qui implique $dk\ell|m$, comme affirmé. La direction réciproque est triviale. □

Finalement, on remarque que, comme pour le plus grand commun diviseur, on peut définir le plus petit commun multiple de plusieurs nombres.

Définition 1.24. Soient $a_1, \ldots, a_k \in \mathbb{Z} \setminus \{0\}$. Le plus petit commun multiple de $a_1, \ldots, a_k$, dénoté par $[a_1, \ldots, a_k]$, est le plus petit nombre naturel qui est un multiple de chacun des nombres $a_1, a_2, \ldots, a_k$. C'est-à-dire

$$[a_1, \ldots, a_k] = \min\{m \in \mathbb{N} : a_j|m \ (1 \leq j \leq k)\}.$$

Exercices

EXERCICE 1.9. Soient $a_1, \ldots, a_k \in \mathbb{Z} \setminus \{0\}$.

(a) Montrez que $[a_1, \ldots, a_k] = [[a_1, \ldots, a_{k-1}], a_k]$.
(b) Montrez que $a_j|m$ pour chaque $j \in \{1, \ldots, k\}$ si et seulement si $[a_1, \ldots, a_k]|m$.
(c) Si $(a_i, a_j) = 1$ pour tout $i \neq j$, montrez que $[a_1, \ldots, a_k] = a_1 \cdots a_k$.

1.4 Nombres premiers

Étant donner un nombre, on peut souvent le factoriser. Par exemple, le nombre 420 est évidemment divisible par 2, puisque $420 = 2 \cdot 210$. On ne peut pas factoriser plus le nombre 2 mais on a que $210 = 2 \cdot 105$. Puis, on trouve que $105 = 5 \cdot 21$ et, finalement, que $21 = 3 \cdot 7$. Alors, on conclut que

$$420 = 2 \cdot 2 \cdot 3 \cdot 5 \cdot 7.$$

On ne peut pas decomposer plus 420 car les nombres 2,3,5 et 7 n'ont pas de diviseurs non-triviaux (c'est-à-dire diviseurs différents de 1 et eux-mêmes). Il existe plus tels nombres. Par exemple, 13 et 23 ont aussi cette propriété. L'indecomposibilité de ces nombres les rends les *atoms* de l'operation de multiplication. Pour cette raison, ils sont très importants et on les donne un nom spécial : ils sont les nombres premiers.

Définition 1.25. Un nombre $n > 1$ est appelé composé si il existe deux nombres a et b tels que $1 < a, b < n$ et $n = ab$.

Un nombre $n > 1$ est appelé premier si il n'est pas composé. De façon équivalente, un nombre $n > 1$ est premier si ses seuls diviseurs positifs sont 1 et n.

Comme l'exemple du nombre 420 pourrait laisser entendre, on peut écrire chaque nombre n comme un produit de certains nombres premiers. De plus, ces facteurs premiers de n sont définis de façon unique. C'est le context du théorème suivant, qui est un résultat d'importance fondamentale, comme son nom l'indique.

Théorème 1.26 (Théorème fondamental de l'arithmétique)**.** *Si $n > 1$, alors il existe des nombres premiers $p_1, \ldots, p_r$ tels que $n = p_1 \cdots p_r$. De plus, les nombres $p_1, \ldots, p_r$ sont définis de façon unique, dans le sens que si $n = q_1 \cdots q_s$ pour quelques nombres premiers $q_1, \ldots, q_s$, alors $s = r$ et les nombres $q_1, \ldots, q_s$ sont une permutation des nombres $p_1, \ldots, p_r$.*

Avant de montrer le théorème fondamental de l'arithmétique, on a besoin d'un résultat préparatoire.

Lemme 1.27. *Soit p un nombre premier.*

(a) Si $a \in \mathbb{Z}$, alors soit $p|a$ soit $(a, p) = 1$.

(b) Si $p|a_1 \cdots a_k$ pour quelques entiers $a_1, \ldots, a_k$, alors $p|a_i$ pour un $i \in \{1, \ldots, k\}$.

Démonstration. (a) Soit $d = (a, p)$. On a que $d|p$. Donc soit $d = 1$ soit $d = p$. Dans le deuxième cas, on déduit que $d = p|a$.

(b) Par induction sur k. Si $k = 1$, le résultat est trivial. Puis, supposons qu'il est vrai pour $k - 1$. Si $p|a_1 \cdots a_k$, alors soit $p|a_k$ soit $p \nmid a_k$. Au premier cas, le résultat découle. Au deuxième cas, on a que $(a_k, p) = 1$ selon la partie (a). Donc le lemme d'Euclid implique que $p|a_1 \cdots a_{k-1}$ et l'hypothèse inductive nous dit que $p|a_i$ pour un $i \in \{1, \ldots, k-1\}$. Ceci termine l'étape inductive et, par la suite, la démonstration. □

Démonstration du théorème fondamental de l'arithmétique. Tout d'abord, on montre l'existence de la factorisation de tous les nombres en nombres premiers. On le ferra de façon inductive. Le théorème est vrai pour $n = 2$, puisque 2 est premier et, par conséquent, il le produit

de nombres premiers. Puis, on considère un nombre $n \geq 3$ et on suppose que le théorème est vrai pour tout les nombres $m \in \{2, \ldots, n-1\}$. Si n est premier, alors il est déjà factorisé comme le produit de nombres premiers et le théorème est vrai pour n aussi. Finalement, si n est composé, alors il existe $a, b \in \mathbb{N}$ tels que $n = ab$ et $1 < a, b < n$. Alors l'hypothèse inductive implique que on peut écrire a et b comme le produit de nombres premiers, soient $a = p_1 \cdots p_r$ et $b = p_{r+1} \cdots p_{r+s}$. Evidemment, c'implique que $n = ab = p_1 p_2 \cdots p_{r+s}$, ce qui conclut l'étape inductive et, par la suite, la première partie de la démonstration.

Finalement, on montre l'unicité de la factorisation de n en nombres premiers. Soient deux factorisations de n, $n = p_1 \cdots p_r = q_1 \cdots q_s$. Bien sur, on a que $q_1 | p_1 \ldots p_r$. Donc, le lemme 1.27(b) implique que $q_1 | p_{i_1}$ pour un $i_1 \in \{1, \ldots, r\}$. Par la suite, $(p_{i_1}, q_1) > 1$ et le lemme 1.27(a) nous donne que $p_i | q_1$ aussi. Alors, on trouve que $q_1 = p_{i_1}$. Puisque $q_1 \cdots q_s = p_1 \cdots p_r$, alors $q_2 \cdots q_s = \prod_{i \neq i_1} p_i$. On répète le même argument avec q_2 : on a que $q_2 | \prod_{i \neq i_1} p_i$ et donc il existe $i_2 \in \{1, \ldots, r\} \setminus \{i_1\}$ tel que $q_2 = p_{i_2}$. On continue de cette façon et, inductivement, on trouve que il existe s indices $i_1, \ldots, i_s$ distincts appartenants à $\{1, \ldots, r\}$ tels que $q_j = p_{i_j}$, pour tout $j \in \{1, \ldots, s\}$. Donc la relation $q_1 \cdots q_s = p_1 \cdots p_r$ devient

$$1 = \prod_{\substack{1 \leq i \leq r \\ i \notin \{i_1, \ldots, i_s\}}} p_i.$$

Ceci montre que $r = s$ aussi et le théorème découle. □

Corollaire 1.28. *Si $n > 1$, alors il existe nombres premiers $p_1 < \cdots < p_r$ et exposants $v_1, \ldots, v_r \in \mathbb{N}$ uniques tels quen $= p_1^{v_1} \cdots p_r^{v_r}$.*

Démonstration. C'est un corollaire directe du théorème fondamental de l'arithmétique : soit $n = q_1 \cdots q_s$ la factorisation de n en nombres premiers. Même si les indices de $q_1, q_2, \ldots, q_s$ sont différents, ils peuvent être le même nombre premier parfois. Soient $p_1, \ldots, p_r$ les facteurs premiers distincts de n. Chaque tel facteur appairait avec une certaine multiplicité dans la factorisation $n = q_1 \cdots q_s$. Si v_i est la multiplicité d'occurrence de p_i, alors on trouve que

$$n = q_1 \cdots q_s = p_1^{v_1} \cdots p_r^{v_r}.$$

Sans perde de généralité, on peut supposer que les nombres $p_1, \ldots, p_r$ sont en order croissant de magnitude ; sinon, on les permute pour garantir ceci. Finalement, l'unicité des nombres $p_1, \ldots, p_r$ et $v_1, \ldots, v_r$ est une conséquence directe de l'unicité de $q_1, \ldots, q_s$ (à part d'une permutation possible). □

La factorisation $n = p_1^{v_1} \cdots p_r^{v_r}$ avec $p_1 < \cdots < p_r$ est appelée la factorisation première de n. Son existence et unicité nous permet de définir

$$v_p(n) := \max\{v \geq 1 : p^v | n\}$$

pour chaque nombre premier p et chaque $n \in \mathbb{N}$. On appel $v_p(n)$ la valuation de n au nombre premier p. Bien sur, si n est fixé, alors $v_p(n) = 0$ pour presque tous les nombres premiers, c'est-à-dire la relation $v_p(n) = 0$ est vraie pour tous les nombres premiers dehors un ensemble fini (qui pourrait dépendre de n). Donc on peut considérer le produit $\prod_p p^{v_p(n)}$ sans problèmes. En fait, le corollaire 1.28 implique tout de suite que $n = \prod_p p^{v_p(n)}$.

Souvent, étant donné un nombre premier p, on utilisera la notation $p^v \| n$, qui veut dire que la plus grande puissance de p qui divise n est égale à v, c'est-à-dire $p^v | n$ et $p^{v+1} \nmid n$. Bien sur, on a que $p^v \| n$ si et seulement si $v = v_p(n)$.

En utilisant la notion de la valuation, on peut écrire le plus grand commun diviseur et le plus petit commun multiple de deux nombres en termes de leurs factorisations premières.

Proposition 1.29. *Soient $a, b \in \mathbb{N}$.*

(a) On a que $a|b$ si et seulement si $v_p(a) \leq v_p(b)$, pour tout nombre premier p.
(b) $(a, b) = \prod_p p^{\min\{v_p(a), v_p(b)\}}$.
(c) $[a, b] = \prod_p p^{\max\{v_p(a), v_p(b)\}}$.

Démonstration. (a) Supposons que $a|b$ et soit p un nombre premier. Puisque $p^{v_p(a)}|a$, alors $p^{v_p(a)}|b$, ce qui implique que $v_p(a) \leq v_p(b)$.

Réciproquement, supposons que $v_p(a) \leq v_p(b)$ pour tout nombre premier p. Donc on peut considérer le nombre entier $k = \prod_p p^{v_p(b)-v_p(a)}$. Evidemment, on a que $b = ka$, cequi implique que $a|b$.

(b) On a que $d|a$ et $d|b$ si et seulement si $v_p(d) \leq v_p(a)$ et $v_p(d) \leq v_p(b)$ pour tout p, si et seulement si $v_p(d) \leq \min\{v_p(a), v_p(b)\}$ pour tout p, si et seulement si $d| \prod_p p^{\min\{v_p(a), v_p(b)\}}$. Ceci conclut la démonstration de la partie (b).

(c) On a que $a|m$ et $b|m$ si et seulement si $v_p(a) \leq v_p(m)$ et $v_p(b) \leq v_p(m)$ pour tout p, si et seulement si $\max\{v_p(a), v_p(b)\} \leq v_p(m)$ pour tout p, si et seulement si $\prod_p p^{\max\{v_p(a), v_p(b)\}}|m$. Ceci conclut la démonstration de la partie (c). □

L'infinité des nombres premiers. Les nombres premiers sont si fondamentaux que la question de leur repartition entre les nombres entiers est inévitable. Euclide a montré qu'il y a un nombre infini de premiers. On étudiera leur repartition de façon precise au chapitre 7.

Théorème 1.30 (Euclide)**.** *Il existe une infinité de nombres premiers.*

Démonstration. Supposons, au contraire, que l'ensemble des nombres premiers est fini, soit $\{p_1, \cdots, p_k\}$. On considère le nombre $n = 1 + p_1 \cdots p_k$. Ce nombre est plus grand que tous les p_i et, par la suite, il est composé. Donc il existe un nombre premier p qui divise n. Nécessairement, $p = p_{i_0}$ pour un certain indice $i_0 \in \{1, \ldots, k\}$. Mais c'est impossible car les relations $p_{i_0}|n$ et $p_{i_0}|p_1 \cdots p_k$ impliquent que $p|(n - p_1 \cdots p_k) = 1$. On est arrivé à une contradiction. Alors l'hypothèse que l'ensemble des nombres premiers est fini ne peut pas être vrai, ce qui est ce qu'il fallait démontrer. □

Le crible d'Eratosthenes. Étant donné le caractère fondamental des nombres premiers, c'est important d'être capable de les calculer. Une méthode pour le faire est appelé le crible d'Eratosthenes. Elle se base au résultat suivant.

Théorème 1.31. *Si $n > 1$ est composé, alors il existe un nombre premier $p \leq \sqrt{n}$ qui divise n.*

Démonstration. Si n est composé, alors $n = ab$ pour quelques nombres $a, b \in \{2, \ldots, n-1\}$. Sans perde de généralité, on peut supposer que $a \leq b$. En particulier, $a^2 \leq ab = n$, ce qui implique que $a \leq \sqrt{n}$. Alors, si p est un facteur premier de a, qui existe du théorème fondamental de l'arithmétique, on trouve que $p|n$ et $p \leq a \leq \sqrt{n}$. Ceci conclut la démonstration. □

À partir de ce théorème, on peut construire un algorithme qui déterminera tous les nombres premiers dans l'intervalle $[1, x]$, pour un x donné. L'algorithme a les étapes suivantes :

Étape 1 : Énumérer tous les entiers dans $(1, x]$.
Étape 2 : Trouver le plus petit entier $n \in (1, \sqrt{x}]$ qui n'est pas déjà ni encerclé ni supprimé et encercler-le. Si un tel n n'existe pas, terminer l'algorithme.
Étape 3 : Supprimer tous les multiples de n qui sont plus grands que n.
Étape 4 : Retourner à la deuxième étape.

Les nombres que ne sont pas supprimés quand cet algorithme termine sont exactement les nombres premiers contenants à $[1, x]$. On ferra un exemple concret pour expliquer le mécanisme de l'algorithme d'Eratosthenes mieux. Supposons qu'on veut trouver tous les nombres premiers jusqu'a $x = 40$. On a que $\sqrt{40} \approx 6.32$, donc on doit répéter le troisième étape de l'algorithme au plus 5 fois (en fait, comme on verra, il suffit de le faire 3 fois). On commence par énumérer tous les nombres jusqu'à 40.

	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19	20	21	22	23	24
25	26	27	28	29	30	31	32
33	34	35	36	37	38	39	40

Le plus petit nombre de notre liste est 2. On l'encercle et on supprime tous ses plus grands multiples :

~~1~~	②	3	~~4~~	5	~~6~~	7	~~8~~
9	~~10~~	11	~~12~~	13	~~14~~	15	~~16~~
17	~~18~~	19	~~20~~	21	~~22~~	23	~~24~~
25	~~26~~	27	~~28~~	29	~~30~~	31	~~32~~
33	~~34~~	35	~~36~~	37	~~38~~	39	~~40~~

Le plus petite nombre qui n'est pas encerclé ni supprimé est 3. On l'encercle et on supprime tous ses plus grands multiples (on a besoin de supprimer seulement les nombres qui n'étaient pas déjà éliminés à la dernière étape ; par exemple, 6 était déjà enlevé comme un multiple de 2) :

~~1~~	②	③	~~4~~	5	~~6~~	7	~~8~~
~~9~~	~~10~~	11	~~12~~	13	~~14~~	~~15~~	~~16~~
17	~~18~~	19	~~20~~	~~21~~	~~22~~	23	~~24~~
25	~~26~~	~~27~~	~~28~~	29	~~30~~	31	~~32~~
~~33~~	~~34~~	35	~~36~~	37	~~38~~	~~39~~	~~40~~

Le seul nombre au-dessous $\sqrt{40} \approx 6.32$ qui reste et qu'il n'est pas ni encerclé ni supprimé est le nombre 5. On l'encercle et on supprime tous ses plus grands multiples :

~~1~~	②	③	~~4~~	⑤	~~6~~	7	~~8~~
~~9~~	~~10~~	11	~~12~~	13	~~14~~	~~15~~	~~16~~
17	~~18~~	19	~~20~~	~~21~~	~~22~~	23	~~24~~
~~25~~	~~26~~	~~27~~	~~28~~	29	~~30~~	31	~~32~~
~~33~~	~~34~~	35	~~36~~	37	~~38~~	~~39~~	~~40~~

Il ne reste plus de nombres dans $[1, \sqrt{40}]$ qui ne s'étaient pas traités, donc l'algorithme termine ici. On conclut que les nombres premiers entre 1 et 40 sont les nombres 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31 et 37.

Exercices

EXERCICE 1.10. (a) Montrez que $(a, b) > 1$ si et seulement si il existe un nombre premier p tel que $p|a$ et $p|b$.

(b) Montrez que $(x^m, x^n + 1) = 1$, pour chaque $m, n \in \mathbb{N}$.

(c) Si $a \equiv 3 \pmod 7$, alors montrez que $(a^3 - a, a^3 - a + 7) = 1$. Quand est-ce que $(a^3 - a, a^3 - a + 7) > 1$?

(d) Montrez que $(ab, a + b^2, a + b + 1) = 1$ pour chaque $a, b \in \mathbb{N}$.

EXERCICE 1.11. (a) Soit $k \geq 2$ et $n \in \mathbb{N}$ dont la première factorisation est $n = p_1^{v_1} \cdots p_r^{v_r}$. Montrez que n est une k-ième puissance parfaite si est seulement si $k|v_i$, pour tout $i \in \{1, \ldots, r\}$.

(b) Est-ce qu'il existe un nombre naturel n tel que sa moitié est un carré parfait, son tiers est un cube est son cinquième est une cinquième puissance ?

EXERCICE 1.12. Soient $k \geq 2$ et $a, b \in \mathbb{N}$. Montrez que $a^k|b^k$ si et seulement si $a|b$.

EXERCICE 1.13. Montrez que, pour chaque entier $n \geq 1$, le nombre $4n^3 + 6n^2 + 4n + 1$ est composé. [*Indice :* Développez $(x + y)^4$.]

EXERCICE 1.14. Un nombre n est appelé sans carré facteur si il n'existe pas de $a > 1$ dont le carré divise n. Un nombre n est appelé plein de carrés si $p^2|n$ pour tout les nombres premiers p qui divisent n.

(a) Montrez que n est sans carré si et seulement si $n = p_1 \cdots p_r$ pour quelques nombres premiers distincts $p_1, \ldots, p_r$.

(b) Montrez que n est plein de carrés si et seulement si $n = p_1^{v_1} \cdots p_r^{v_r}$ pour quelques nombres premiers distincts $p_1, \ldots, p_r$ et quelques exposants $v_1, \ldots, v_r \geq 2$.

(c) Montrez que chaque nombre entier peut s'écrire comme $n = ab$, où a est sans carré facteur, b est plein de carrés et $(a, b) = 1$.

(d) Montrez que chaque nombre entier peut s'écrire comme $n = ab^2$, où a est sans carré facteur.

(e) Montrez qu'un nombre plein de carrés n peut s'écrire comme $n = a^2b^3$, pour quelques entiers a et b.

EXERCICE 1.15. Si $n = p_1^{v_1} \cdots p_r^{v_r}$ est la première factorisation de n, alors on pose

$$\omega(n) = r \quad \text{et} \quad \Omega(n) = v_1 + \cdots + v_r,$$

avec la convention que $\omega(1) = \Omega(1) = 0$.

(a) Montrez que $v_p(mn) = v_p(m) + v_p(n)$ et que $\Omega(mn) = \Omega(m) + \Omega(n)$.
(b) Montrez que si $(m, n) = 1$, alors $\omega(mn) = \omega(m) + \omega(n)$.

EXERCICE 1.16 (Furstenberg). On muni l'ensemble des nombres entiers d'une topologie, comme suivant :

— $\emptyset$ et $\mathbb{Z}$ sont ouverts,
— chaque progression arithmétique $S(a, b) := \{an + n : n \in \mathbb{Z}\}$ est ouvert,
— les unions des ensembles ouverts sont également ouverts.

Montrez que :

(a) Les progressions arithmétiques sont également fermés.
(b) Chaque ensemble fini n'est pas ouvert.
(c) $\mathbb{Z} \setminus \{-1, 1\}$ n'est pas fermé.
(d) $\mathbb{Z} \setminus \{-1, 1\} = \bigcup_{p \text{ premier}} S(p, 0)$.

Concluez que il existe une infinité de nombres premiers.

1.5 Fonctions multiplicatives

Une classe importante de fonctions dans la théorie des nombres sont les fonctions multiplicatives. Elles sont des fonctions qui respectent la structure multiplicative des entiers. La définition précise suit.

Définition 1.32. Une fonction $f : \mathbb{N} \to \mathbb{C}$ est appelée multiplicative si $f \neq 0$ (c'est-à-dire, s'il existe $n \in \mathbb{N}$ tel que $f(n) \neq 0$) et si

$$f(mn) = f(m)f(n) \quad \text{quand} \quad (m, n) = 1. \tag{1.1}$$

Si la relation (1.1) tient pour tous $m, n \in \mathbb{N}$, on dit que f est complètement multiplicative.

Par exemple, les fonctions $n \to n^\alpha$, où $\alpha \in \mathbb{C}$, sont toutes complètement multiplicatives. Le théorème suivant nous fournit d'autres exemples de fonctions multiplicatives.

Théorème 1.33. *Si $\alpha \in \mathbb{C}$, la fonction σ_α, définie par*

$$\sigma_\alpha(n) = \sum_{d|n} d^\alpha,$$

est multiplicative.

Démonstration. Supposons que $(m,n)=1$. Alors, chaque diviseur d de mn peut s'écrire de façon unique comme $d=d_1d_2$, où $d_1|m$ et $d_2|n$. En effet, cette décomposition est donnée par les nombres $d_1=\prod_{p^v\|d,\,p|m}p^v$ et $d_2=\prod_{p^v\|d,\,p|n}p^v$. Par la suite,

$$\sigma_\alpha(mn)=\sum_{d|mn}d^\alpha=\sum_{\substack{d_1|m\\ d_2|n}}d_1^\alpha d_2^\alpha=\left(\sum_{d_1|m}d_1^\alpha\right)\left(\sum_{d_2|n}d_2^\alpha\right)=\sigma_\alpha(m)\sigma_\alpha(n),$$

comme affirmé. □

Remarque 1.34. Les cas avec $\alpha=0$ et $\alpha=1$ du théorème 1.33 sont d'importance spéciale. Le cas $\alpha=0$ correspond à la fonction diviseur

$$\tau(n):=\#\{d|n\}$$

et le cas $\alpha=1$ correspond à la fonction "sommes de diviseurs"

$$\sigma(n):=\sum_{d|n}d.$$

Si on sait qu'une fonction est multiplicative, on peut la calculer à condition qu'on connaisse ses valuers aux puissances de nombres premiers :

Proposition 1.35. *Soit f une fonction multiplicative.*

(a) On a que $f(1)=1$.

(b) Si $n=p_1^{v_1}p_2^{v_2}\cdots p_r^{v_r}$ est la factorisation de n à ses facteurs premiers distincts, alors

$$f(n)=f(p_1^{v_1})f(p_2^{v_2})\cdots f(p_r^{v_r}).$$

Si, de plus, f est complètement multiplicative, alors

$$f(n)=f(p_1)^{v_1}f(p_2)^{v_2}\cdots f(p_r)^{v_r}.$$

Démonstration. (a) Par définition, il existe $n\in\mathbb{N}$ tel que $f(n)\neq 0$. Mais on a que $f(n)=f(n\cdot 1)=f(n)f(1)$ car $(n,1)=1$ et, par la suite, $f(1)=1$.

(b) On utilise induction sur r. Si $r=1$, il n'y a rien à montrer. Supposons maintenant que la proposition est vraie quand n a $r-1$ facteurs premiers distincts. Puisque $(p_1^{v_1},p_2^{v_2}\cdots p_r^{v_r})=1$, on trouve que

$$f(n)=f(p_1^{v_1})f(p_2^{v_2}\cdots p_r^{v_r})=f(p_1^{v_1})f(p_2^{v_2})\cdots f(p_r^{v_r}),$$

selon l'hypothèse inductive. Ceci termine la démonstration de la première partie. Finalement, si f est complètement multiplicative, alors

$$f(p^v)=f(p\cdot p^{v-1})=f(p)f(p^{v-1})=\cdots=\underbrace{f(p)\cdots f(p)}_{v\text{ fois}}=f(p)^v,$$

ce qui conclut la démonstration. □

Corollaire 1.36. *Soit $n = p_1^{v_1} p_2^{v_2} \cdots p_r^{v_r}$ la factorisation première du nombre n. Alors*

$$\tau(n) = (v_1 + 1)(v_2 + 1) \cdots (v_r + 1)$$

et

$$\sigma(n) = \prod_{j=1}^{r} \frac{p_j^{v_j+1} - 1}{p_j - 1}.$$

Démonstration. Exercice. □

On conclut cette section avec une application classique du théorème 1.33 qui caractérise les nombres parfaits :

Définition 1.37. Un nombre n est appelé parfait si il est la somme de ses propres diviseurs, c'est-à-dire si

$$n = \sum_{d|n,\ d<n} d.$$

De façon équivalent, n est parfait si et seulement si $\sigma(n) = 2n$.

Par exemple, 6 est parfait : on a que $6 = 1 + 2 + 3$. D'autres exemples sont les nombres 28 et 496. Les nombres parfaits ont fasciné Euclid et plusieurs autres personnes pendant les années. En fait, Euclid a trouvé une façon générale d'engendrer de nombres parfaits. Cependant, il ne pouvait pas trouver de nombres parfait impairs, ni les autres personnes qu'ont essayés. Aujourd'hui, c'est une conjecture ouverte fameuse de prouver qu'il n'existe pas de nombres parfaits impairs. D'autre coté, les nombres parfaits pairs sont complètement classifiés, grâce à Euclid et Euler :

Théorème 1.38 (Euclid, Euler)**.** *Un nombre pair n est parfait si et seulement si il existe un nombre premier p tel que $2^p - 1$ est également premier et que $n = 2^{p-1}(2^p - 1)$.*

Démonstration. Tout d'abord, si $n = 2^{p-1}(2^p - 1)$, où p est un nombre premier pour lequel $2^p - 1$ est aussi premier, alors on a que

$$\sigma(n) = \sigma(2^{p-1}(2^p - 1)) = \sigma(2^{p-1})\sigma(2^p - 1) = \frac{2^p - 1}{2 - 1} \cdot 2^p = 2n,$$

comme souhaité.

Réciproquement, supposons que n est un nombre parfait impair. On écrit $n = 2^v m$, où $v \geq 1$ et m est impair. On a que $\sigma(n) = \sigma(2^v)\sigma(m) = (2^v - 1)\sigma(m)$. D'autre coté, on a que $\sigma(n) = 2n = 2^{v+1}m$, c'est-à-dire $(2^{v+1} - 1)\sigma(m) = 2^{v+1}m$. Puisque $(2^{v+1}, 2^{v+1} - 1) = 1$ et $2^{v+1} - 1$ divise $2^{v+1}m$, alors $m = (2^{v+1} - 1)\ell$ pour un $\ell \in \mathbb{N}$. Donc on trouve que

$$2^{v+1}(2^{v+1} - 1)\ell = (2^{v+1} - 1)\sigma(m) \quad \Longrightarrow \quad 2^{v+1}\ell = \sigma(m).$$

On affirme que $\ell = 1$. Sinon, on aurait que $1, \ell$ et m sont tous diviseurs distincts de $m = (2^{v+1} - 1)\ell$ et, par la suite,

$$\sigma(m) \geq 1 + m + \ell = 2^{v+1}\ell + 1 > 2^{v+1}\ell,$$

ce qui est impossible. Donc $\ell = 1$, comme affirmé, ce qui implique que $\sigma(2^{v+1}-1) = 2^{v+1}$. De cette dernière relation, on déduit tout de suite que $2^{v+1}-1$ est premier ; sinon, il existait $d \in (1, 2^{v+1}-1)$ qui divisait $2^{v+1}-1$ et, conséquemment,

$$\sigma(2^{v+1}-1) \geq 1 + d + (2^{v+1}-1) > 2^{v+1},$$

ce qui est une contradiction. Donc $2^{v+1}-1$ est en effet premier. Il reste de montrer que $v+1$ est premier. Sinon, on aurait que $v+1 = ab$ pour quelques $a, b \in \mathbb{N}$ avec $a, b > 1$. Mais dans ce cas on aurait que

$$2^{v+1}-1 = (2^a)^b - 1 = (2^a-1)((2^a)^{b-1} + (2^a)^{b-2} + \cdots + 2^a + 1),$$

ce qui contredit le fait que $2^{v+1}-1$ est premier. Alors $v+1$ est premier et le résultat désiré découle. □

Exercices

EXERCICE 1.17. (a) Montrez que si f est une fonction arithmétique telle que $f(n) = \prod_{p^v \| n} f(p^v)$ pour chaque $n \in \mathbb{N}$, alors f est multiplicative. De plus, montrez que si $f(n) = \prod_{p^v \| n} f(p)^v$, alors f est fortement multiplicative.

(b) Montrez que si f est une fonction arithmétique telle que $f(n) = \prod_{p|n} f(p)$ pour chaque $n \in \mathbb{N}$, alors

$$f(mn)f((m,n)) = f(m)f(n) \quad (m, n \in \mathbb{N}).$$

Déduisez que f est multiplicative.

EXERCICE 1.18. Est-ce que la fonction $f(n) = (-1)^{n-1}$ multiplicative ou pas ?

EXERCICE 1.19. Prouvez le corollaire 1.36.

EXERCICE 1.20. Trouvez les trois premiers nombres parfaits.

EXERCICE 1.21. Montrez que n est un nombre parfait si et seulement si $\sum_{d|n} 1/d = 2$.

EXERCICE 1.22. Soit

$$f(n) = \#\{(n_1, n_2) \in \mathbb{N}^2 : [n_1, n_2] = n\}.$$

Montrez que f est multiplicative and évaluez-la aux puissances des nombres premiers.

Chapitre 2

Arithmétique modulaire

L'arithmétique modulaire est une façon d'étudier des fonctions périodiques, ainsi que la divisibilité d'un entier par un autre. Un exemple d'une fonction périodique est le temps : les heures se répètent chaque jour, les jours se répètent chaque semaine, et les mois se répètent chaque an. Un autre exemple plus musical est l'octave : elle octave se compose de douze semitones et se répète périodiquement quand on change la fréquence du son continûment. Cependant, ici on s'intéresse principalement aux phénomènes arithmétiques. Un tel exemple est la propriété d'être divisible par 5. C'est une propriété 5-périodique (c'est-à-dire, de période 5) : soit tous les membres de la progression arithmétiquen $\{a+5k : k \in \mathbb{Z}\}$ sont de multiples de 5, ou aucun membre ne l'est. Un autre exemple, plus élaboré, est la propriété que 7 divise le nombre x^2+1. Comme on le verra, la période ici est 7.

2.1 Congruences

Définition 2.1. Soit $n \in \mathbb{N}$. On dit que deux nombres $a, b \in \mathbb{Z}$ sont congruents modulo n si $n|(a-b)$. Dans ce cas, on écrit $a \equiv b \pmod{n}$.

À partir de cette définition, on voit tout-de-suite que tous les nombres dans la progression arithmétique $\{a+kn : k \in \mathbb{Z}\}$ sont congruents à a. En fait, c'est facile de vérifier que la congruence modulo n est une relation d'équivalence dont les classes d'équivalences sont exactement les différentes progressions arithmétiques d'étape n.

Définition 2.2. Une classe d'équivalence modulo n est appelée une classe de congruence modulo n ou un résidu modulo n. L'ensemble de tous les résidus modulo n est dénoté par $\mathbb{Z}/n\mathbb{Z}$. Finalement, un système complet de représentants des classes d'équivalence modulo n est appelé un système complet de résidus modulo n.

Remarque 2.3. On utilisera la notation $a \pmod{n}$ pour dénoter la classe d'équivalence de a modulo n.

Lemme 2.4. *Pour tout $n \in \mathbb{N}$, l'ensemble $\{0, 1, \ldots, n-1\}$ est un système complète de résidus modulo n, c'est-à-dire*

$$\mathbb{Z}/n\mathbb{Z} = \{0 \pmod{n}, 1 \pmod{n}, \ldots, n-1 \pmod{n}\}.$$

En particulier, $|\mathbb{Z}/n\mathbb{Z}| = n$.

Démonstration. La division euclidienne nous dit que pour chaque $a \in \mathbb{Z}$ il existe $q \in \mathbb{Z}$ et $r \in \{0, 1, \ldots, n-1\}$ tels que $a = qn + r$. En particulier, $a \equiv r \pmod{n}$. Il reste de prouver que si $0 \leq i < j \leq n-1$, alors $i \not\equiv j \pmod{n}$. En effet, on a que $0 < j - i < n$ et il n'y a pas de multiples de n dans l'intervalle $(0, n)$. Donc $i \not\equiv j \pmod{n}$, comme affirmé. □

À partir du système complet de résidus $\{0, 1, \ldots, n-1\}$, on peut construire d'autres systèmes complets :

Lemme 2.5. *Soient* $n \in \mathbb{N}$ *et* $a, b \in \mathbb{Z}$. *Si* $(a, n) = 1$, *alors l'ensemble*

$$\{aj + b : 0 \leq j \leq n-1\}$$

est un système complet de résidus.

Démonstration. Il suffit de montrer que si $aj + b \equiv aj' + b \pmod{n}$ et $0 \leq j, j' \leq n-1$, alors $j = j'$. En effet, on a que $n|(aj + b) - (aj' + b) = a(j - j')$. Puisque $(a, n) = 1$, alors le lemme d'Euclid implique que $n|(j - j')$. Mais $|j - j'| \leq n - 1$ et le seul multiple de n dans l'intervalle $[-(n-1), n-1]$ est 0. Donc $j = j'$, comme affirmé. □

On déduit tout-de-suite le résultat suivant qui est très utile.

Corollaire 2.6. *Soit* $n \in \mathbb{N}$. *Un ensemble de* n *nombres entiers consécutifs est un système complet de résidus modulo* n. *En particulier, l'ensemble* $\mathbb{Z} \cap (-n/2, n/2]$ *est un système complet de résidus modulo* n.

Démonstration. La première partie découle du lemme 2.5, puisque un ensemble de n nombres entiers consécutifs peut s'écrire dans la forme $\{j + b : 0 \leq j \leq n-1\}$ pour un $b \in \mathbb{Z}$. Pour la deuxième partie, on observe que $\mathbb{Z} \cap (-n/2, n/2]$ est toujours un ensemble de n nombres entiers consécutifs. (Vérifiez-le comme exercice.) □

Le concept des congruences révèle sa puissance quand on considère son comportement sous les operations de l'addition et de la multiplication.

Lemme 2.7. *Soient* $n \in \mathbb{N}$ *et* $a, b, c, d \in \mathbb{Z}$. *Si* $a \equiv b \pmod{n}$ *et* $c \equiv d \pmod{n}$, *alors* $a + c \equiv b + d \pmod{n}$ *et* $ac \equiv bd \pmod{n}$.

Démonstration. (a) On a que $n|(a-b)$ et que $n|(c-d)$. Donc $n|(a-b)+(c-d) = (a+c-b-d)$, ce qui implique que $a + c \equiv b + d \pmod{d}$, et que $n|c(a-b) - b(c-d) = ac - bd$, d'où on déduit que $ac \equiv bd \pmod{n}$.

(b) On a que $n|(ab - ac) = a(b - c)$. Puisque $(a, n) = 1$, le lemme d'Euclid implique que $a|b - c$, c'est-à-dire $b \equiv c \pmod{n}$. □

Donc on voit que les congruences respectent les operations algébriques de base. En particulier, on peut définir la somme et le produit de deux résidus $a \pmod{n}$ et $b \pmod{n}$ par

$$a \pmod{n} + b \pmod{n} := a + b \pmod{n} \quad \text{et} \quad a \pmod{n} \cdot b \pmod{n} := ab \pmod{n}.$$

Ces opérations sont bien définies d'après le lemme 2.7. Il existe une difference fondamentale entre l'arithmétique de $\mathbb{Z}$ et l'arithmétique de $\mathbb{Z}/n\mathbb{Z}$. On sait que les seuls nombres entiers qui possèdent un inverse qui est également entier sont -1 et 1. L'inverse des autres entiers n'appartient pas à $\mathbb{Z}$ mais à $\mathbb{Q}$. Ce n'est pas le cas pour $\mathbb{Z}/n\mathbb{Z}$, où plusieurs éléments possèdent d'inverses.

Lemme 2.8. *Soient $n \in \mathbb{N}$ et $a, x, y \in \mathbb{Z}$.*

(a) On a que $ax \equiv ay \pmod{n}$ si et seulement si $x \equiv y \pmod{n/(a,n)}$. En particulier, si $(a,n) = 1$, alors on a que $ax \equiv ay \pmod{n}$ si et seulement si $x \equiv y \pmod{n}$.

(b) Si $(a,n) = 1$, alors il existe un résidu unique $b \pmod{n}$ tel que $ab \equiv 1 \pmod{n}$.

Démonstration. (a) On pose $d = (a,n)$ et on écrit $n = dn_1$ et $a = da_1$ pour que $(n_1, a_1) = 1$.

Si $ax \equiv ay \pmod{n}$, alors $n|(ax - ay) = a(x-y)$, ce qui implique que $n_1|a_1(x-y)$. Puisque $(a_1, n_1) = 1$, le lemme d'Euclid implique que $n_1|(x-y)$, c'est-à-dire $x \equiv y \pmod{n_1}$, comme affirmé.

Réciproquement, si $x \equiv y \pmod{n_1}$, alors $n_1|(x-y)$ et, par la suite, $dn_1|d(x-y)$. En particulier, $n = dn_1|da_1(x-y) = a(x-y)$, ce qui implique que $ax \equiv ay \pmod{n}$.

(b) Puisque $(a,n) = 1$, le lemme 1.9 implique l'existence de $b, c \in \mathbb{Z}$ tels que $ab + nc = 1$. Donc $ab \equiv 1 \pmod{n}$. Finalement, on montre que b est défini uniquement modulo n. En effet, si $ab' \equiv 1 \pmod{n}$ pour un autre $b' \in \mathbb{Z}$, alors $ab \equiv ab' \pmod{n}$ et la partie (a) du lemme implique que $b \equiv b' \pmod{n}$. □

Si $(a,n) = 1$, alors on dénote la classe d'équivalence de $b \pmod{n}$ pour laquelle $ab \equiv 1 \pmod{n}$ par $\overline{a} \pmod{n}$ ou, si ceci ne cause pas de confusion, par $\overline{a}$. On appelle $\overline{a}$ l'inverse de a modulo n. Parfois on écrira aussi $1/a \pmod{n}$ ou $a^{-1} \pmod{n}$. Le fait que les résidus qui sont copremiers à n possèdent un inverse mod n les offrent un rôle spécial. Donc on les appellent résidus réduits. L'ensemble de tous les résidus est dénoté par $(\mathbb{Z}/n\mathbb{Z})^*$, c'est-à-dire

$$(\mathbb{Z}/n\mathbb{Z})^* := \{a \pmod{n} : (a,n) = 1\}.$$

Sa cardinalité est appelée la fonction d'Euler et elle dénotée par

$$\phi(n) := \#(\mathbb{Z}/n\mathbb{Z})^*.$$

Remarque 2.9. La démonstration du lemme 2.8 donne une façon de calculer d'inverse d'un résidu quadratique. Par exemple, si $a = 100$ et $n = 271$, alors on applique l'algorithme d'Euclid afin de trouver que

$$271 = 2 \cdot 100 + 71, \quad 100 = 71 + 29, \quad 71 = 2 \cdot 29 + 13, \quad 29 = 2 \cdot 13 + 3, \quad 13 = 4 \cdot 3 + 1,$$

donc

$$\begin{aligned} 1 &= 13 - 4 \cdot 3 \\ &= 13 - 4 \cdot (29 - 2 \cdot 13) = 9 \cdot 13 - 4 \cdot 29 \\ &= 9 \cdot (71 - 2 \cdot 29) - 4 \cdot 29 = 9 \cdot 71 - 22 \cdot 29 \\ &= 9 \cdot 71 - 22 \cdot (100 - 71) = 31 \cdot 71 - 22 \cdot 100 \\ &= 31 \cdot (271 - 2 \cdot 100) - 22 \cdot 100 \\ &= 31 \cdot 271 - 84 \cdot 100. \end{aligned}$$

Alors, on conclut que $\overline{100} \equiv -84 \pmod{271}$.

L'inversibilité des résidus quadratiques veut dire qu'on peut diviser par eux. En effet, si $a \in \mathbb{Z}/n\mathbb{Z}$ et $b \in (\mathbb{Z}/n\mathbb{Z})^*$, alors $a\bar{b} \pmod{n}$ joue le rôle de la fraction a/b. Dans le cas spécial où $n = p$ est premier, alors tous les résidus modulo p, sauf $0 \pmod{p}$ sont copremiers à p. Donc on peut diviser par tous les résidus non-zéros et on prend un nouveau résidu modulo p, exactement comme on peut diviser un nombre réel par un autre nombre réel qui est non-zéro, et le résultat de cette division est un nouveau nombre réel. C'est une propriété très importante des nombres premiers qu'on exploitera plusieurs fois dans ce chapitre et les suivants. Pour l'instance, on établit un lemme qui montre que certains équations sur $\mathbb{Z}/p\mathbb{Z}$ se comportent de la même façon que sur $\mathbb{R}$.

Lemme 2.10. *Soit p un nombre premier. On a que $a^2 \equiv b^2 \pmod{p}$ si et seulement si $a \equiv b \pmod{p}$ ou $a \equiv -b \pmod{p}$.*

Démonstration. Supposons que $p|(a^2 - b^2) = (a - b)(a + b)$. Puisque p est premier, le lemma 1.27(b) implique que soit $p|(a - b)$ soit $p|(a + b)$, c'est-à-dire soit $a \equiv b \pmod{p}$ soit $a \equiv -b \pmod{p}$. Réciproquement, on a que $(-b)^2 \equiv a^2 \pmod{p}$. Donc si $a \equiv \pm b \pmod{n}$, alors le lemme 2.7 implique que $a^2 \equiv b^2 \pmod{p}$. □

On termine cette section avec une discussion d'équations polynomiales modulo un nombre premier n. On commence avec le lemme suivant.

Lemme 2.11. *Soit $f(x) \in \mathbb{Z}[x]$. Si $a \equiv b \pmod{n}$, alors $f(a) \equiv f(b) \pmod{n}$.*

Démonstration. Soit $f(x) = c_d x^d + \cdots + c_1 x + c_0$. En utilisant le lemme 2.7(a) et induction sur j, on peut montrer que la relation $a \equiv b \pmod{n}$ implique que $a^j \equiv b^j \pmod{n}$, pour tout $j \in \mathbb{N}$. Donc on a que $c_j a^j \equiv c_j b^j \pmod{n}$ pour tout $j \in \{0, 1, \ldots, d\}$. En appliquant encore une fois le lemme 2.7(a) (de façon inductive), on trouve que

$$f(a) = c_d a^d + \cdots + c_1 a + c_0 \equiv c_d b^d + \cdots + c_1 b + c_0 \pmod{n} \equiv f(b) \pmod{n},$$

ce qui est ce qu'il fallait démontrer. □

Le lemme precedent nous permet de parler de solutions d'une équation polynomiale modulo n, ou de racines d'un polynôme $f(x) \in \mathbb{Z}[x]$ modulo un nombre naturel n. En effet, du lemme 2.11, l'ensemble $\{x \pmod{n} : f(x) \equiv 0 \pmod{n}\}$ est bien défini. De plus, si on choisi un système complet de résidus modulo n, par exemple $\{0, 1, \ldots, n-1\}$, on peut écrire cet ensemble comme $\{0 \leq x \leq n - 1 : f(x) \equiv 0 \pmod{n}\}$.

2.2 Critères de divisibilité

On apprend à l'école quelques critères simples pour déterminer si un nombre est divisible par 5, 10, 3 ou 9. Ici on explique l'origin de ces critères et on voit une façon générale d'obtenir de critères de divisibilité par n'importe quel nombre. On commence avec l'explication des critères familiers.

Soit $n = (a_k a_{k-1} \cdots a_1 a_0)_{10}$, où $a_0, a_1, \ldots, a_k \in \{0, 1, \ldots, 9\}$, l'expansion décimale du nombre n, c'est-à-dire

$$n = a_0 + a_1 10 + \cdots + a_{k-1} 10^{k-1} + a_k 10^k.$$

Soit aussi un nombre d, qu'on prendra d'être 3,5,9 ou 10. La réalisation-clé est que $d|n$ si et seulement si $n \equiv 0 \pmod{d}$. Donc si on sait la classe d'équivalence des puissances de 10 modulo d, alors on peut réécrire $n \pmod{d}$ dans une façon qui, avec un peu de chance, sera plus simple.

Par exemple, on a que $10^i \equiv 0 \pmod{10}$, pour tout $i \in \{1, \ldots, k\}$, et donc

$$n \equiv a_0 + a_1 \cdot 0 + \cdots + a_k \cdot 0 \pmod{10} \equiv a_0 \pmod{10}.$$

Par la suite, $10|n$ si et seulement si $10|a_0$. Mais a_0 appartient à $\{0, 1, \ldots, 9\}$ et le seul multiple de 10 dans cet ensemble et le nombre 0. Donc $10|n$ si et seulement si $a_0 = 0$. C'est le critère familier de divisibilité par 10.

De même, puisque $10^i \equiv 0 \pmod{10}$, on a que $n \equiv a_0 \pmod{5}$. Donc $5|n$ si et seulement si $5|a_0$. Puisque $a_0 \in \{0, 1, \ldots, 9\}$, on trouve que $5|a_0$ si et seulement si $a_0 \in \{0, 5\}$.

Les critères de visibilité par 3 et par 9 sont plus compliqués mais leurs démonstrations sont aussi faciles. On a que $3|n$ si et seulement si 3 divise $a_0 + a_1 + \cdots + a_k$, la somme des chiffres décimales de n. Le critère demeure pareil pour 9 : on a que $9|n$ si et seulement si 9 divise $a_0 + a_1 + \cdots + a_k$. On montrera le critère pour 3 ; la lectrice peut vérifier facilement que le même argument marche pour 9 également. Comme avant, la première étape est de calculer la classe d'équivalence des puissances de 10 modulo 3. Bien sur on a que $10 \equiv 1 \pmod{3}$. Donc $10^j \equiv 1^j \pmod{3} \equiv 1 \pmod{3}$. Donc

$$n \equiv a_0 + a_1 \cdot 1 + \cdots + a_k \cdot 1 \pmod{3} \equiv a_0 + a_1 \cdots + a_k \pmod{3}.$$

Alors, on conclut que $3|n$ si et seulement si 3 divise $a_0 + a_1 \cdots + a_k$.

Les critères de divisibilité modulo 3,5,9 et 10 sont particulièrement simple. Est-ce qu'on peut trouver des critères assez simples pour d'autres nombres ? Par exemple, supposons qu'on cherche un critère de divisibilité de n par 7. On commence avec l'étude des puissances de 10. On a que

$$\begin{aligned}
10^0 &= 1 \equiv 1 \pmod{7} \\
10^1 &= 10 \equiv 3 \pmod{7} \\
10^2 &\equiv 3^2 \pmod{7} \equiv 2 \pmod{7} \\
10^3 &= 10 \cdot 10^2 \equiv 3 \cdot 2 \pmod{7} \equiv 6 \pmod{7} \equiv -1 \pmod{7} \\
10^4 &= 10 \cdot 10^3 \equiv 3 \cdot (-1) \pmod{7} \equiv -3 \pmod{7} \\
10^5 &= 10 \cdot 10^4 \equiv 3 \cdot (-3) \pmod{7} \equiv -2 \pmod{7} \\
10^6 &= 10 \cdot 10^5 \equiv 3 \cdot (-2) \pmod{7} \equiv 1 \pmod{7}.
\end{aligned}$$

On arrête ici parce que $10^6 \equiv 10^0 \pmod{7}$. Ceci veut dire que les puissances de 10 sont 6-périodiques modulo 7. En effet, on a que $10^{k+6} = 10^k 10^6 \equiv 10^k \pmod{7}$. Donc on trouve

que

$$\begin{aligned} n &\equiv (a_0 + 3a_1 + 2a_2 - a_3 - 3a_4 - 2a_5) + (a_6 + 3a_7 + 2a_8 - a_9 - 3a_{10} - 2a_{11}) + \cdots \pmod 7 \\ &\equiv (a_0 + 3a_1 + 2a_2) - (a_3 + 3a_4 + 2a_5) + (a_6 + 3a_7 + 2a_8) - (a_9 + 3a_{10} + 2a_{11}) \pm \cdots \pmod 7 \end{aligned}$$

ce qui implique que 7 divise n si et seulement si 7 divise $(a_0+3a_1+2a_2)-(a_3+3a_4+2a_5)\pm\cdots$.

On peut généraliser cette procedure. Supposons que le nombre n est donné dans la base de b, c'est-à-dire $n = (c_r c_{r-1} \cdots c_0)_b$ pour quelques nombres $c_0, c_1, \ldots, c_r \in \{0, 1, \ldots, b-1\}$ ou, de façon équivalente,

$$n = c_0 + c_1 b + \cdots + c_{r-1} b^{r-1} + c_r b^r.$$

Supposons aussi qu'on veut déterminer si n est divisé par un nombre donné d. Le corollaire 2.6 implique que pour chaque $j \geq 0$, on peut trouver $\beta_j \in \mathbb{Z} \cap (-d/2, d/2]$ tel que $b^j \equiv \beta_j \pmod d$. Donc

$$n \equiv c_0\beta_0 + c_1\beta_1 + \cdots + c_r\beta_r \pmod d.$$

C'est une relation très utile parce qu'elle nous permet de réduire la divisibilité de n, un nombre de magnitude $\approx b^r$, à la divisibilité de $c_0\beta_0 + c_1\beta_1 + \cdots + c_r\beta_r$, un nombre de magnitude $\leq r \cdot (b-1) \cdot d/2$ qui est considérablement plus petit que b^r (à condition que d ne soit pas énorme). De plus, on a un avantage additionnel : on peut calculer la séquence β_j avec $\leq 2d$ operations. En effet, les $d+1$ nombres $b^0, b^1, \ldots, b^d$ ne peuvent pas être tous distincts modulo d car il existe seulement d distincts classes d'équivalences modulo d. Donc il existe deux nombres i_0 et k tels que $0 \leq i_0 < i_0 + k \leq d$ et que $b^{i_0} \equiv b^{i_0+k} \pmod d$. C'implique que la suite $\{\beta_i\}_{i \geq 1}$ est k-périodique pour $i \geq i_0$: si $i \geq i_0$, on a que

$$\beta_{i+k} \equiv b^{i+k} \pmod d \equiv b^{i-i_0} b^{i_0+k} \pmod d \equiv b^{i-i_0} b^{i_0} \pmod d \equiv b^i \pmod d \equiv \beta_i \pmod d.$$

Puisque $-d/2 < \beta_{i+k}, \beta_i \leq d/2$, alors on déduit que $\beta_{i+k} = \beta_i$, pour tout $i \geq i_0$. C'implique qu'il suffit de calculer les nombres $\beta_0, \beta_1, \ldots, \beta_{i_0+k-1}$; les autres sont déterminés par périodicité. Finalement, on remarque qu'on peut choisir les nombres i_0 et k dans une façon minimale : on peut supposer que i_0 est le minimum nombre $i \geq 0$ pour lequel il existe un $\ell \geq 1$ tel que $b^i \equiv b^{i+\ell} \pmod d$. Ayant choisi i_0, on choisit k d'être le minimum nombre tel que $b^{i_0+k} \equiv b^{i_0} \pmod d$.

Remarque 2.12. Si $(b, d) = 1$, alors la relation $b^i \equiv b^{i+\ell} \pmod d$ et le lemme 2.8(b) implique que $b^\ell \equiv 1 \pmod d$. On déduit que dans ce cas on peut toujours choisir $i_0 = 0$ et $k \leq d-1$. Comme on verra au chapitre **??**, le nombre minimum $k \geq 1$ pour lequel $b^k \equiv 1 \pmod d$ est appelé l'ordre multiplicative de b modulo d.

On donne un autre exemple concret pour démontrer la procedure décrite ci-dessus. Soit $n = (c_r \cdots c_1 c_0)_2$ un nombre dans la base binaire, c'est-à-dire

$$n = c_0 + c_1 2 + \cdots + c_{r-1} 2^{r-1} + c_r 2^r.$$

On veut déterminer si c'est nombre est divisible par 6. On applique la procedure au-dessus : on a que

$$\begin{aligned}
2^0 &= 1 \equiv 1 \,(\mathrm{mod}\, 6)\\
2^1 &= 2 \equiv 2 \,(\mathrm{mod}\, 6)\\
2^2 &= 4 \equiv -2 \,(\mathrm{mod}\, 6)\\
2^3 &= 2 \cdot 2^2 \equiv 2 \cdot (-2) \,(\mathrm{mod}\, 6) \equiv -4 \,(\mathrm{mod}\, 6) \equiv 2 \,(\mathrm{mod}\, 6).
\end{aligned}$$

Donc on voit que la suite $\{2^j\}_{j\geq 0}$ est 2-périodique pour $j \geq 1$. Alors

$$\begin{aligned}
n &\equiv c_0 + (2c_1 - 2c_2) + (2c_3 - 2c_4) + (2c_5 - 2c_6) + \cdots \,(\mathrm{mod}\, 6)\\
&\equiv c_0 + 2c_1 - 2c_2 + 2c_3 - 2c_4 \pm \cdots \,(\mathrm{mod}\, 6).
\end{aligned}$$

Alors on conclut que $6|n$ si et seulement si 6 divise $c_0 + 2c_1 - 2c_2 + 2c_3 - 2c_4\pm$. Par exemple, si $n = (100001111000100111000)_2$, alors 6 divise n si et seulement si 6 divise

$$0 + 0 - 0 + 2 - 2 + 2 - 0 + 0 - 2 + 0 - 0 + 0 - 2 + 2 - 2 + 2 - 0 + 0 - 0 - 0 + 2 = 4,$$

ce qui n'est pas vrai. Donc $6 \nmid n$.

2.3 Le théorème de Wilson

Le premier résultat important de l'arithmétique modulaire qu'on montrera est un théorème jolie grâce à Wilson qui, entre autres, donne une caractérisation des nombres premiers.

Théorème 2.13 (Wilson)**.** *Si p est un nombre premier, alors*

$$(p-1)! \equiv -1 \,(\mathrm{mod}\, p).$$

Démonstration. Si $p = 2$ ou $p = 3$, le résultat est évident. Supposons maintenant que $p \geq 5$. On couple chaque nombre $j \in \{1, \ldots, p-1\}$ avec son inverse mod p, qu'on dénote par $\overline{j}$ (on suppose qu'on a choisi $\overline{j}$ parmi les nombres $\{1, \ldots, p-1\}$). On observe que j et $\overline{j}$ apparaissent au produit $(p-1)! = 1 \cdot 2 \cdots (p-1)$ et, donc, il vont s'annuler. Presque... on a menti un peu. Afin que s'arrive, il faut que $\overline{j} \neq j$. Ceci nous amène naturellement à comprendre quand $\overline{j} = j$ ou, de façon équivalente, quand $j \cdot j \equiv 1 \,(\mathrm{mod}\, p)$. Alors $p|j^2 - 1 = (j-1)(j+1)$ et, vu que p est premier, ceci implique que $p|j-1$ ou $p|j+1$. Donc, on a montré que $\overline{j} = j$ si, et seulement si, $j \equiv \pm 1 \,(\mathrm{mod}\, p)$. Ceci implique que les éléments de l'ensemble $\{2, 3, \ldots, p-2\}$ peut se diviser en pairs de la forme $(j, \overline{j})$ avec $j \neq \overline{j}$ et, par la suite, $2 \cdot 3 \cdots (p-2) \equiv 1 \,(\mathrm{mod}\, p)$. On conclut que

$$(p-1)! \equiv 1 \cdot (p-1) \equiv -1 \,(\mathrm{mod}\, p),$$

ce qui est ce qu'il fallait démontrer. □

Remarque 2.14. L'idée de s'accoupler les éléments de $\{1, \ldots, p-1\}$ dans de façons appropriées pour calculer un produit est très utile. On la reverra quand on étudie les résidus quadratiques.

Dans le cas où n est composé (et plus grand que 4), on a le résultat suivant complémentaire :

Théorème 2.15. *Si $n > 4$ est composé, alors*

$$(n-1)! \equiv 0 \,(\mathrm{mod}\, n).$$

Démonstration. Soit $n = p_1^{v_1} \cdots p_r^{v_r}$ la factorisation première de n.

Si $r \geq 2$, alors $p_j^{v_j} < n$ et, donc, $p_j^{v_j} | (n-1)!$ pour chaque j. On conclut alors que $n|(n-1)!$ dans ce cas-ci.

Supposons, maintenant, que $r = 1$, c'est-à-dire $n = p^v$ pour un nombre premier p et un exposant $v \geq 2$ (car n est composé). On a que $p, p^2, \ldots, p^{v-1}$ sont tous $< n$, donc leur produit divise $(n-1)!$. On a que

$$p \cdot p^2 \cdots p^{v-1} = p^{1+2+\cdots+(v-1)} = p^{\frac{v(v-1)}{2}}.$$

On observe que $v(v-1)/2 \geq v$ quand $v \geq 3$, ce qui conclut la preuve dans ce cas.

Il reste à considérer le cas où $v = 2$, c'est-à-dire $n = p^2$. On a que $p > 2$ car $n > 4$. Il faut montrer que $p^2 | 1 \cdot 2 \cdots (p^2 - 1)$. On sait déjà que p se trouve entre les nombres $1, 2, \ldots, p^2 - 1$, donc on cherche un autre multiple de p entre eux. On observe que $2p < p^2$ quand $p > 2$, d'où $p \cdot 2p | (p^2 - 1)!$. Ceci termine la démonstration. □

Corollaire 2.16. *Un nombre $n > 1$ est premier si et seulement si*

$$(n-1)! \equiv -1 \,(\mathrm{mod}\, n).$$

Démonstration. Si n est premier, on a que $(n-1)! \equiv -1 \,(\mathrm{mod}\, n)$, du théorème de Wilson. Si $n = 4$, alors on a que $(4-1)! \equiv 2 \not\equiv -1 \,(\mathrm{mod}\, 4)$. Finalement, si n est composé et plus grand que 4, alors $(n-1)! \equiv 0 \not\equiv -1 \,(\mathrm{mod}\, n)$, du théorème 2.15. □

2.4 Le théorème des restes chinois

Supposez que vous êtes un empereur possédant un armé énorme, mais vous ne connaissez pas exactement le nombre de vos soldats. Comment est-ce que vous pouvez le calculez rapidement ? Le théorèmes des restes chinois vous offre une solution géniale : soit N le nombre de vos soldats, qui forment une ligne longue. Comme ça, c'est facile d'arranger les soldats en pairs. À la fin, il reste soit aucun soldat soit un soldat, ce qui implique que N est pair ou impair, respectivement. Puis, les soldats forment des triplets. À la fin, il reste 0, 1 ou 2 soldats, ce qui détermine le reste de N quand divisé par 3. On continue de cette façon : pour tout nombre premier p plus petit qu'un paramètre z, on arrange les soldats en p-tuples pour trouver le reste de N quand divisé par p. Le théorème de restes chinois (cf. théorème 2.18) dit que ceci détermine le reste de N modulo $m := \prod_{p \leq z} p$. Si on choisi la paramètre z pour que $m > N$ (par exemple, si $z = 30$, on a que m est à peu prés égal à 6.5 billions, un nombre qui est certainement plus grand que N), alors le reste de N modulo m est, en fait, égal à N, c'est-à-dire on a déterminé N soi-même !

Avant de montrer le théorème des restes chinois, on établit un lemme préparatoire :

Lemme 2.17 (théorème des restes chinois - version bébé)**.** *Soient $m_1, \dots, m_k \in \mathbb{N}$ et $x, y \in \mathbb{Z}$. On a que*

$$\begin{cases} x \equiv y \,(\mathrm{mod}\, m_1) \\ \vdots \\ x \equiv y \,(\mathrm{mod}\, m_k) \end{cases} \quad \Leftrightarrow \quad x \equiv y \,(\mathrm{mod}\, [m_1, \dots, m_k])$$

Démonstration. On a que $x \equiv y \,(\mathrm{mod}\, n_j)$ pour tout $j \in \{1, \dots, k\}$ si et seulement si $n_j | (x-y)$ pour tout $j \in \{1, \dots, k\}$. D'après l'exercice 1.9, cette relation est équivalente à la relation $[m_1, \dots, m_k] | (x-y)$, c'est-à-dire $x \equiv y \,(\mathrm{mod}\, [m_1, \dots, m_k])$. □

On montre maintenant le théorème des restes chinois :

Théorème 2.18 (théorème des restes chinois)**.** *Soient $m_1, \dots, m_k \in \mathbb{N}$ tels que $(m_i, m_j) = 1$ si $i \neq j$. Soient aussi $a_1, \dots, a_k \in \mathbb{Z}$. Le système de congruences*

$$\begin{cases} x \equiv a_1 \,(\mathrm{mod}\, m_1) \\ \vdots \\ x \equiv a_k \,(\mathrm{mod}\, m_k) \end{cases} \tag{2.1}$$

a une solution unique modulo $m_1 \cdots m_k$.

Démonstration. On donne deux arguments :

Première preuve. On pose $M = m_1 \cdots m_k$ et, pour chaque $j \in \{1, \dots, k\}$, on pose

$$M_j = \frac{M}{m_j} = m_1 \cdots m_{j-1} m_{j+1} \cdots m_k.$$

L'observation fondamentale est que $(m_j, M_j) = 1$ et que $M_j \equiv 0 \,(\mathrm{mod}\, m_i)$ pour $i \neq j$. La première relation garanti que M_j est inversible mod m_j. Soit N_j son inverse. Alors on a que

$$M_j N_j \equiv \mathbf{1}_{j=i} \,(\mathrm{mod}\, m_i) = \begin{cases} 1 \,(\mathrm{mod}\, m_j) \\ 0 \,(\mathrm{mod}\, m_i) \end{cases} \quad \text{si } i \neq j.$$

Alors le nombre $x_0 = a_1 M_1 N_1 + \cdots + a_k M_k N_k$ est une solution au système (2.1).

Il reste de montrer que la solution x_0 qu'on a trouvé est unique mod $M = m_1 \cdots m_k$. En effet, si $x_0 \equiv x_1 \,(\mathrm{mod}\, m_j)$, pour tout $j \in \{1, \dots, k\}$, alors le lemme 2.17 implique que $x_0 \equiv x_1 \,(\mathrm{mod}\, [m_1, \dots, m_k])$. Finalement, puisque $(m_i, m_j) = 1$, en appliquant l'exercice 1.9(c), on déduit que $[m_1, \dots, m_k] = M$, c'est-à-dire $x_0 \equiv x_1 \,(\mathrm{mod}\, M)$, comme affirmé.

Deuxième preuve. On considère d'abord le cas $k = 2$. On cherche x tel que $x \equiv a_1 \,(\mathrm{mod}\, m_1)$ et $x \equiv a_2 \,(\mathrm{mod}\, m_2)$. La forme générale des x satisfaisant la première relation est $x = km_1 + a_1$. Donc on cherche k tel que

$$km_1 + a_1 \equiv a_2 \,(\mathrm{mod}\, m_2) \quad \Leftrightarrow \quad km_1 \equiv a_2 - a_1 \,(\mathrm{mod}\, m_2) \quad \Leftrightarrow \quad k \equiv \overline{m}_1 (a_2 - a_1) \,(\mathrm{mod}\, m_2),$$

où $\overline{m}_1$ est l'inverse de m_1 modulo m_2, qui existe à cause de la co-primalité de m_1 et de m_2. Si $b_2 \in \{0, 1, \dots, m_2 - 1\}$ est un representant de la classe de congruences $\overline{m}_1 (a_2 - a_1) \,(\mathrm{mod}\, m_2)$,

on trouve que $k = \ell m_2 + b_2$. Donc, la solution générale est donnée par $x = (\ell m_2 + b_2)m_1 + a_1 = \ell m_1 m_2 + (b_2 m_1 + a_1)$, ce qui montre au même temps l'existence et l'unicité mod $m_1 m_2$ de x.

Le cas général suit par induction : on a transformé le système de congruences $x \equiv a_1 \pmod{m_1}$ et $x \equiv a_2 \pmod{m_2}$ à une seule congruence, soit $x \equiv a \pmod{m_1 m_2}$. Donc, on a passé d'un système avec k congruences à un système avec $k-1$ congruences. □

Remarque 2.19. La démonstration du théorème des restes chinois nous donne deux algorithmes pour résoudre le système de congruences (2.1).

Dans le premier algorithme, on calcule $M = m_1 \cdots m_k$, $M_j = M/m_j$ et N_j tels que $M_j N_j \equiv 1 \pmod{m_j}$ en utilisant l'algorithme euclidien. La solution générale est donnée par

$$x \equiv a_1 M_1 N_1 + \cdots a_k M_k N_k \pmod{M}.$$

Dans le deuxième algorithme, on calcule la solution générale en commençant par la solution de la première équation, $x = k_1 m_1 + a_1$, où k_1 est une variable libre. Puis, on remplace x par cette expression à la deuxième équation et on trouve la solution générale du système formés par les deux premières équations, soit $x = k_2 m_1 m_2 + a_2'$, où k_2 est une variable libre et a_2' est un nombre déterminé. En continuant de cette façon, on trouve la solution général au système de congruences $x \equiv a_j \pmod{m_j}$, $1 \leq j \leq k$.

Exemple 2.20. Résolvons le système de congruences

$$\begin{cases} x \equiv 1 \pmod{5} \\ x \equiv 3 \pmod{4} \\ x \equiv 2 \pmod{3}. \end{cases} \tag{2.2}$$

Solution. On a que $(5,4) = (4,3) = (5,3) = 1$. Donc on peut appliquer tout de suite l'algorithme mentionné au remarque 2.19. On a $M = 5 \cdot 4 \cdot 3 = 60$, $M_1 = 4 \cdot 3 = 12$, $M_2 = 5 \cdot 3 = 15$ et $M_3 = 5 \cdot 4 = 20$. Puis on calcule N_1, N_2 et N_3. On a que

$$\begin{aligned} M_1 N_1 \equiv 1 \pmod{m_1} \quad &\Leftrightarrow \quad 12 N_1 \equiv 1 \pmod{5} \quad \Leftrightarrow \quad 2 N_1 \equiv 1 \pmod{5} \\ &\Leftrightarrow \quad N_1 \equiv 3 \pmod{5}, \end{aligned}$$

$$\begin{aligned} M_2 N_2 \equiv 1 \pmod{m_2} \quad &\Leftrightarrow \quad 15 N_2 \equiv 1 \pmod{4} \quad \Leftrightarrow \quad -N_2 \equiv 1 \pmod{4} \\ &\Leftrightarrow \quad N_2 \equiv -1 \pmod{4} \end{aligned}$$

et

$$\begin{aligned} M_3 N_3 \equiv 1 \pmod{m_3} \quad &\Leftrightarrow \quad 20 N_3 \equiv 1 \pmod{3} \quad \Leftrightarrow \quad 2 N_3 \equiv 1 \pmod{3} \\ &\Leftrightarrow \quad N_3 \equiv -1 \pmod{3}. \end{aligned}$$

Donc la solution unique au système (2.2) est

$$x \equiv 1 \cdot 12 \cdot 3 + 3 \cdot 15 \cdot (-1) + 2 \cdot 20 \cdot (-1) \pmod{60} \equiv 11 \pmod{60}.$$

De façon alternative, afin de résoudre le système (2.2), on observe que la congruence $x \equiv 1 \pmod 5$ veut dire que $x = 5y+1$, pour un $y \in \mathbb{Z}$. Avec cette substitution, la congruence $x \equiv 3 \pmod 4$ devient

$$5y + 1 \equiv 3 \pmod 4 \quad \Leftrightarrow \quad y \equiv 2 \pmod 4,$$

c'est-à-dire $y = 4z + 2$. Donc $x = 5(4z + 2) + 1 = 20z + 11$ et la congruence $x \equiv 2 \pmod 3$ devient

$$20z + 11 \equiv 2 \pmod 3 \quad \Leftrightarrow \quad 2z \equiv 0 \pmod 3 \quad \Leftrightarrow \quad z \equiv 0 \pmod 3.$$

Par la suite, $z = 3w$, ce qui implique que $x = 60w+11$. C'est la solution générale au système de congruences (2.2), qu'on peut réécrire comme $x \equiv 11 \pmod{60}$, comme avant. □

Exemple 2.21. Résolvons le système de congruences

$$\begin{cases} x \equiv 1 \pmod 6 \\ x \equiv 4 \pmod{15} \end{cases} \tag{2.3}$$

Solution. Ici on a que $(6, 15) = 3 > 1$, donc on ne peut appliquer directement l'algorithme mentionné au remarque 2.19. On écrit factorise 6 et 15 en termes de leurs plus grand commun facteur : $6 = 2 \cdot 3$ et $15 = 3 \cdot 5$. Puis, on observe que le lemme 2.17 implique que

$$\begin{cases} x \equiv 1 \pmod 6 \\ x \equiv 4 \pmod{15} \end{cases} \quad \Leftrightarrow \quad \begin{cases} \begin{cases} x \equiv 1 \pmod 2 \\ x \equiv 1 \pmod 3 \end{cases} \\ \begin{cases} x \equiv 4 \pmod 3 \\ x \equiv 4 \pmod 5 \end{cases} \end{cases} \quad \Leftrightarrow \quad \begin{cases} x \equiv 1 \pmod 2 \\ x \equiv 1 \pmod 3 \\ x \equiv 4 \pmod 5 \end{cases}$$

$$\Leftrightarrow \quad \begin{cases} x \equiv 1 \pmod 6 \\ x \equiv 4 \pmod 5 \end{cases}$$

puisque les congruences $x \equiv 4 \pmod 3$ et $x \equiv \pmod 3$ sont, en fait, la même congruence. Maintenant on est en position d'appliquer l'algorithme du remarque 2.19 au dernier système, puisque $(5, 6) = 1$. On a $M = 5 \cdot 6 = 30$, $M_1 = 5$ et $M_2 = 6$. Puis on calcule N_1 et N_2. On a que

$$M_1N_1 \equiv 1 \pmod{m_1} \quad \Leftrightarrow \quad 5N_1 \equiv 1 \pmod 6 \quad \Leftrightarrow \quad N_1 \equiv -1 \pmod 6$$

et

$$M_2N_2 \equiv 1 \pmod{m_2} \quad \Leftrightarrow \quad 6N_2 \equiv 1 \pmod 5 \quad \Leftrightarrow \quad N_2 \equiv 1 \pmod 5.$$

Donc la solution unique au système (2.3) est

$$x \equiv 1 \cdot 5 \cdot (-1) + 4 \cdot 6 \cdot 1 \pmod{30} \equiv 19 \pmod{30}.$$

Alternativement, afin de résoudre le système (2.2), on observe que $x \equiv 1 \pmod 6$ si et seulement si $x = 6y + 1$. Donc la congruence $x \equiv 4 \pmod{15}$ devient

$$6y + 1 \equiv 4 \pmod{15} \quad \Leftrightarrow \quad 6y \equiv 3 \pmod{15} \quad \Leftrightarrow \quad 2y \equiv 1 \pmod 5 \quad \Leftrightarrow \quad y \equiv 3 \pmod 5,$$

où on a utilisé le lemme 2.8(a). Donc $y = 5z + 3$, ce qui implique que $x = 6(5z + 3) + 1 = 30z + 19$, c'est-à-dire la solution générale au système (2.3) est $x \equiv 19 \pmod{30}$. □

Exemple 2.22. Résolvons le système de congruences

$$\begin{cases} x \equiv 1 \pmod 5 \\ x \equiv 2 \pmod{10} \\ x \equiv 3 \pmod 7. \end{cases} \tag{2.4}$$

Solution. On a que $(5, 10) = 5 > 1$. Donc on écrit $10 = 5 \cdot 2$ et on observe que

$$\begin{cases} x \equiv 1 \pmod 5 \\ x \equiv 2 \pmod{10} \\ x \equiv 3 \pmod 7 \end{cases} \quad \Leftrightarrow \quad \begin{cases} x \equiv 1 \pmod 5 \\ \begin{cases} x \equiv 2 \pmod 2 \\ x \equiv 2 \pmod 5 \end{cases} \\ x \equiv 3 \pmod 7 \end{cases} \quad \Leftrightarrow \quad \begin{cases} x \equiv 1 \pmod 5 \text{ et } x \equiv 2 \pmod 5 \\ x \equiv 0 \pmod 2 \\ x \equiv 3 \pmod 7 \end{cases}$$

Le dernière système de congruences n'a pas de solutions parce que les relations $x \equiv 1 \pmod 5$ et $x \equiv 2 \pmod 5$ sont mutuellement exclusives. Donc le système (2.4) ne possède pas de solutions. □

Exemple 2.23. Résolvons le système de congruences

$$\begin{cases} 3x \equiv 3 \pmod{27} \\ 3x \equiv 2 \pmod 5 \\ x \equiv 7 \pmod{21}. \end{cases} \tag{2.5}$$

Solution. Ici on a deux problèmes. D'abord, on a que x est multiplié par 3 en deux instances. Cependant, on a que

$$3x \equiv 3 \pmod{27} \quad \Leftrightarrow \quad x \equiv 1 \pmod 9$$

selon le lemme 2.8(a) et que

$$3x \equiv 2 \pmod 5 \quad \Leftrightarrow \quad x \equiv 3^{-1} \cdot 2 \pmod 5 \quad \Leftrightarrow \quad x \equiv 4 \pmod 5,$$

puisque $3^{-1} \equiv 2 \pmod 5$. Donc le système (2.5) est équivalent au système

$$\begin{cases} x \equiv 1 \pmod 9 \\ x \equiv 4 \pmod 5 \\ x \equiv 7 \pmod{21} \end{cases} \tag{2.6}$$

Mais même maintenant on ne peut pas appliquer l'algorithme du remarque 2.19 parce que $(9, 21) = 3 > 1$. Donc on écrit 9 et 21 en termes de leur plus grand diviseur, c'est-à-dire $9 = 3 \cdot 3$ et $21 = 3 \cdot 7$. Cependant, ici

$$x \equiv 1 \pmod 9 \quad \not\Leftrightarrow \quad \begin{cases} x \equiv 1 \pmod 3 \\ x \equiv 1 \pmod 3 \end{cases}$$

car $(3,3) = 3 > 1$. En fait, ce qu'il faut faire est de trouver les factorisations premières de 9 et 21, qui sont $9 = 3^2$ et $21 = 3 \cdot 7$. Puis on peut dire que

$$\begin{cases} x \equiv 1 \pmod 9 \\ x \equiv 4 \pmod 5 \\ x \equiv 7 \pmod{21} \end{cases} \quad \Leftrightarrow \quad \begin{cases} x \equiv 1 \pmod{3^2} \\ x \equiv 4 \pmod 5 \\ \begin{cases} x \equiv 7 \pmod 3 \\ x \equiv 7 \pmod 7 \end{cases} \end{cases} \quad \Leftrightarrow \quad \begin{cases} x \equiv 1 \pmod{3^2} \text{ et } x \equiv 1 \pmod 3 \\ x \equiv 4 \pmod 5 \\ x \equiv 0 \pmod 7 \end{cases}$$

Finalement, on observe que la relation $x \equiv 1 \pmod{3^2}$ implique la relation $x \equiv 1 \pmod 3$. Par conséquent, on trouve que

$$\begin{cases} x \equiv 1 \pmod 9 \\ x \equiv 4 \pmod 5 \\ x \equiv 7 \pmod{21} \end{cases} \quad \Leftrightarrow \quad \begin{cases} x \equiv 1 \pmod 9 \\ x \equiv 4 \pmod 5 \\ x \equiv 0 \pmod 7 \end{cases}$$

Ce dernière système peut se traiter selon l'algorithme du remarque 2.19 : on a que $M = 9 \cdot 5 \cdot 7 = 315$, $M_1 = 5 \cdot 7 = 35$, $M_2 = 9 \cdot 7 = 63$ et $M_3 = 9 \cdot 5 = 45$. Puis on trouve N_1, N_2 et N_3 : on a que

$$M_1 N_1 \equiv 1 \pmod{m_1} \quad \Leftrightarrow \quad 35N_1 \equiv 1 \pmod 9 \quad \Leftrightarrow \quad N_1 \equiv -1 \pmod 9,$$

$$\begin{aligned} M_2 N_2 \equiv 1 \pmod{m_2} \quad \Leftrightarrow \quad 63N_2 \equiv 1 \pmod 5 \quad &\Leftrightarrow \quad 3N_2 \equiv 1 \pmod 5 \\ &\Leftrightarrow \quad N_2 \equiv 2 \pmod 5 \end{aligned}$$

et

$$\begin{aligned} M_3 N_3 \equiv 1 \pmod{m_3} \quad \Leftrightarrow \quad 45N_3 \equiv 1 \pmod 7 \quad &\Leftrightarrow \quad 3N_3 \equiv 1 \pmod 7 \\ &\Leftrightarrow \quad N_3 \equiv -2 \pmod 7. \end{aligned}$$

Donc la solution unique au système (2.2) est

$$x \equiv 1 \cdot 35 \cdot (-1) + 4 \cdot 63 \cdot 2 + 0 \cdot 45 \cdot (-2) \pmod{315} \equiv 154 \pmod{315}.$$

Alternativement, afin de résoudre le système (2.2), on observe que $x \equiv 1 \pmod 9$ si et seulement si $x = 9y + 1$ pour un $y \in \mathbb{Z}$. Donc on a que

$$x \equiv 4 \pmod 5 \quad \Leftrightarrow 9y + 1 \equiv 4 \pmod 5 \quad \Leftrightarrow y \equiv 2 \pmod 5,$$

c'est-à-dire $y = 5z + 2$. Par la suite, $x = 9(5z + 2) + 1 = 45z + 19$, ce qui implique que

$$x \equiv 7 \pmod{21} \quad \Leftrightarrow 45z + 19 \equiv 7 \pmod{21} \quad \Leftrightarrow 3z \equiv 9 \pmod{21} \quad \Leftrightarrow z \equiv 3 \pmod 7.$$

Donc $z = 7w + 3$, ce qui implique que $x = 45(7z + 3) + 19 = 315w + 154$, c'est-à-dire la solution générale du système (2.6) est $x \equiv 154 \pmod{315}$. □

La fonction d'Euler

Rappelez que la fonction d'Euler $\phi(n)$ est définie d'être la cardinalité de l'ensemble des résidus réduits, c'est-à-dire

$$\phi(n) = \#(\mathbb{Z}/n\mathbb{Z})^* = \#\{1 \leq a \leq n : (a,n) = 1\}.$$

On montre maintenant quelques propriétés-clés de ϕ.

On commence par le calcul de quelques valeurs. Si p est premiers, on a que $(a,p) = 1$ si et seulement si $p \nmid a$. Donc

$$\phi(p) = \#\{1 \leq a \leq p : p \nmid a\} = p - 1,$$

puisque chaque nombre entre 1 et $p-1$ n'est pas divisible par p. En généralisant cet argument, on observe que si k est n'importe quel nombre naturel, on a que $(a,p^k) = 1$ si et seulement si $p \nmid a$. Par conséquent,

$$\phi(p^k) = p^k - \#\{1 \leq a \leq p^k : p|a\} = p^k - p^{k-1}$$

car les multiples de p entre 1 et p^k sont les nombres $1 \cdot p, 2p, 3p, \ldots, p^{k-1} \cdot p$.

On calcule maintenant $\phi(pq)$, où p et q sont deux nombres premiers distincts. On a que

$$\phi(pq) = \#\{1 \leq n \leq pq : p, q \nmid n\} = pq - \#\{1 \leq n \leq pq : p|n \text{ ou } q|n\}$$

Les nombres $n \in \{1, \ldots, pq\}$ qui sont divisibles par p sont $p, 2p, 3p, \ldots, qp$, et les n divisibles par q sont $q, 2q, 3q, \ldots, pq$. La seule coincidence entre ces deux listes est le nombre pq : en effet, si $p, q|n$, alors $pq|n$ car p et q sont premiers distincts. En particulier, il faut que $n \geq pq$. On en déduit que

$$\phi(pq) = pq - (p + q - 1) = pq - p - q + 1 = (p-1)(q-1).$$

Cette factorisation de $\phi(pq)$ n'est pas un accident :

Théorème 2.24. *Si $(m,n) = 1$, alors $\phi(mn) = \phi(m)\phi(n)$, c'est-à-dire la fonction ϕ est multiplicative.*

Démonstration. Soit $f : (\mathbb{Z}/mn\mathbb{Z})^* \to (\mathbb{Z}/mZ)^* \times (\mathbb{Z}/n\mathbb{Z})^*$ définie par $f(x \pmod{mn}) := (x \pmod{m}, x \pmod{n})$. On a que :

— $(x, mn) = 1$ si et seulement si $(x,m) = (x,n) = 1$;

— $x \equiv y \pmod{mn}$, si et seulement si $x \equiv y \pmod{m}$ et $x \equiv y \pmod{n}$, par le lemme 2.17.

Ces relations montrent que la fonction f est bien définie, ainsi qu'elle est injective. Finalement, on montre que f est surjective : soit $(a,m) = 1$ et $(b,n) = 1$. Par le théorème des restes chinois, il existe $x \pmod{mn}$ tel que $x \equiv a \pmod{m}$ et $x \equiv b \pmod{n}$. Donc $f(x \pmod{mn}) = (a \pmod{m}, b \pmod{n})$, comme voulu. Ceci montre que f est une bijection. Par la suite

$$\phi(mn) = \#(\mathbb{Z}/mn\mathbb{Z})^* = \#\left((\mathbb{Z}/mZ)^* \times (\mathbb{Z}/n\mathbb{Z})^*\right) = \#(\mathbb{Z}/mZ)^* \cdot \#(\mathbb{Z}/n\mathbb{Z})^* = \phi(m)\phi(n),$$

ce qui conclut la démonstration. □

Corollaire 2.25. *On a que*

$$\phi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right).$$

Démonstration. Ceci découle directement de la multiplicativité de n : si la factorisation première de n est donnée par $n = p_1^{v_1} \cdots p_k^{v_k}$, alors les nombres $p_1^{v_1}, \ldots, p_k^{v_k}$ sont deux à deux copremiers, d'où

$$\phi(n) = \prod_{j=1}^{k} \phi(p_j^{v_j}).$$

Puisque $\phi(p_j^{v_j}) = p_j^{v_j} - p_j^{v_j-1} = p_j^{v_j}(1 - 1/p_j)$, comme on l'a vu avant, la formule affirmée pour $\phi(n)$ en suit.

Démonstration combinatoire. Soit $n = p_1^{v_1} \cdots p_k^{v_k}$ comme ci-dessus. On observe que

$$\phi(n) = \#\{1 \le a \le n : p_j \nmid n \ \forall j\}.$$

Soit $X = \{1 \le a \le n\}$ et $E_j = \{1 \le a \le n : p_j | n\}$. Alors, on cherche la cardinalité de l'ensemble $X \setminus \cup_{j=1}^{k} E_j$. D'après la principe d'inclusion-exclusion, on a que

$$\begin{aligned}\phi(n) &= |X| - \sum_{j=1}^{k} |E_j| + \sum_{1 \le j_1 < j_2 \le k} |E_{j_1} \cap E_{j_2}| - \sum_{1 \le j_1 < j_2 < j_3 \le k} |E_{j_1} \cap E_{j_2} \cap E_{j_3}| \pm \cdots \\ &= \sum_{J \subset \{1,2,\ldots,k\}} (-1)^{|J|} \Big| \bigcap_{j \in J} E_j \Big|,\end{aligned}$$

où $\cap_{j \in \emptyset} E_j = X$ par convention. Si $J = \{j_1, \ldots, j_\ell\}$, alors

$$\bigcap_{j \in J} E_j = \{1 \le a \le n : p_{j_1}, \ldots, p_{j_\ell} | a\} = \{1 \le a \le n : p_{j_1} \cdots p_{j_\ell} | a\}$$

puisque les nombres premiers $p_{j_1}, \ldots, p_{j_\ell}$ sont distincts. En écrivant $a = b p_{j_1} \cdots p_{j_\ell}$, on trouve alors que

$$\Big| \bigcap_{j \in J} E_j \Big| = \#\{b \le n / p_{j_1} \cdots p_{j_\ell}\} = \frac{n}{p_{j_1} \cdots p_{j_\ell}}, \tag{2.7}$$

d'où

$$\frac{\phi(n)}{n} = \sum_{J \subset \{1,2,\ldots,k\}} (-1)^{|J|} \prod_{j \in J} \frac{1}{p_j} = \sum_{J \subset \{1,2,\ldots,k\}} \prod_{j \in J} \frac{-1}{p_j}.$$

On affirme que cette expression est égale à $(1 - 1/p_1) \cdots (1 - 1/p_k)$. En effet, quand on développe ce produit, on trouve une somme de 2^k termes de la forme $(-1/p_{j_1}) \cdots (-1/p_{j_\ell})$, déterminés par les sous-ensembles $J = \{j_1, \ldots, j_\ell\}$ de $\{1, \ldots, k\}$, où la correspondence est qu'on choisit soit $-1/p_j$ soit 1 dans la parenthèse $(1 - 1/p_j)$, dépendant de si $j \in J$ ou non, respectivement. Ceci conclut la démonstration combinatoire du théorème.

Démonstration probabiliste. On réécrit (2.7) comme

$$\frac{|\bigcap_{j\in J} E_j|}{|X|} = \prod_{j\in J}\frac{1}{p_j} = \prod_{j\in J}\frac{|E_j|}{|X|}. \tag{2.8}$$

Si on interprète le quotient $\mathbb{P}(E) := |E|/|X|$ comme la probabilité de l'ensemble $E \subset X$, la relation (2.8) implique que les événements E_j, $1 \le j \le k$, sont mutuellement indépendants. Donc

$$\frac{\phi(n)}{n} = \mathbb{P}\left(\bigcap_{j=1}^{k} E_j^c\right) = \prod_{j=1}^{k}\mathbb{P}(E_j^c) = \prod_{j=1}^{k}(1-\mathbb{P}(E_j)) = \prod_{j=1}^{k}\left(1-\frac{1}{p_j}\right).$$

□

Remarque 2.26. On peut essayer d'utiliser l'approche probabiliste pour donner une approximation à la quantité de nombres premiers $\le x$. En effet, les nombres premiers p sont des objets définis par des *exclusions*; les exclusions de la divisibilité par tous les nombres dans l'intervalle ouvert $(1, p)$. En fait, le crible d'Eratosthenes nous dit qu'il suffit d'exclure la divisibilité de p par les premiers dans l'intervalle $[1, \sqrt{p}]$. Ceci implique la formule

$$\#\{\sqrt{x} < p \le x\} = \#\{1 < n \le x : p|n \;\Rightarrow\; p > \sqrt{x}\}.$$

Motivés par l'argument probabiliste qu'on présenté ci-dessus, on pose $E_p = \{1 < n \le x : p|n\}$ et on suppose que les événements E_p sont quasi-indépendants de probabilité $\approx 1/p$. On peut donc conjecturer que

$$\#\{\sqrt{x} < p \le x\} \approx x \prod_{p\le\sqrt{x}}\left(1-\frac{1}{p}\right).$$

Cependant, comme on le verra au chapitre 7, le côté gauche est asymptotique à $x/\log x$, et le côté droit à $cx/\log x$, pour une constante $c > 1$. On voit alors que notre hypothèse d'indépendance n'est pas correcte ici. Le problème est que les événements E_p sont très corrélés pour de grands nombres premiers. Par exemple, si $n \in (1, x]$ et $p_1, p_2, p_3 > x^{1/3}$, alors on sait qu'au moins un p_j ne divise pas n; sinon, on aurait que $n \ge p_1p_2p_3 > x$, ce qui est absurd. On voit alors que $E_{p_1} \cap E_{p_2} \cap E_{p_3} = \emptyset$.

Théorème 2.27 (Théorème d'Euler)**.** *Si* $(a, n) = 1$*, alors*

$$a^{\phi(n)} \equiv 1 \pmod{n}.$$

Démonstration. On sait que $(\mathbb{Z}/n\mathbb{Z})^* = \{k \pmod{n} : 1 \le k \le n,\ (k, n) = 1\}$. Puisque $(a, n) = 1$, on sait aussi que les nombres $a, 2a, \ldots, na$ forment un système complet de résidus mod n. De plus, $(ak, n) = 1$ si et seulement si $(k, n) = 1$. Donc $(\mathbb{Z}/n\mathbb{Z})^* = \{ak \pmod{n} : 1 \le k \le n,\ (k, n) = 1\}$. On trouve alors que

$$\prod_{\substack{1\le k\le n\\(k,n)=1}} k \equiv \prod_{\substack{1\le k\le n\\(k,n)=1}} ak \pmod{n}$$

Si on le pose $m = \prod_{\substack{1\le k\le n\\(k,n)=1}} k$, alors le côté gauche se factorise comme

$$ma^{\#\{1\le k\le n:(k,n)=1\}} = ma^{\phi(n)}.$$

Donc

$$ma^{\phi(n)} \equiv m \pmod n. \tag{2.9}$$

Puisque m est le produit de quelques résidus réduits, il est copremier avec n. Donc $m^{-1} \pmod n$ existe, et le théorème découle d'éliminant m dans la relation (2.9). □

En prenant $n = p$, un premier, on arrive tout de suite au résultat suivant :

Corollaire 2.28 (Petit théorème de Fermat)**.** *Si $p \nmid a$, alors*

$$a^{p-1} \equiv 1 \pmod p.$$

L'algorithme RSA

La théorie des nombres a plusieurs applications pratiques, particulièrement sur la cryptographie. La cryptographie existe depuis long temps, parce que le besoin de transmettre de messages d'une façon sécuritaire est aussi vieux que la civilisation humaine. Par exemple, il existe un algorithme de cryptage inventé par Jules Cesar. L'algorithme est facile : il se base sur une permutation donné des lettres de l'alphabet. Toutefois, c'est facile de décrypter tels algorithmes, spécialement si le message est assez long : on peut utiliser une combinaison des statistiques sur les lettres qui sont les plus fréquentes ou des mots fréquentes (articles, prepositions, les verbes "avoir" et "être" et ses conjugaisons, etc.) et deviner quels sont quelques lettres.

Avec le temps, les méthodes de cryptage ont évolues et, aujourd'hui, elles se basent sur des méthodes arithmétiques. De façon générale, quand on veut transmettre un message de façon sécuritaire, il y a trois choses qu'il faut assurer :

— l'expéditeur du message doit être capable de le crypter rapidement ;
— un adversaire ne doit pas pouvoir décrypter le message transmis rapidement et lire le message original ;
— il faut que le destinataire ait accès à une clé de décryptage qui lui permettra de décrypter le message rapidement.

Habituellement, on utilise les noms Alice (Personne A) pour l'expéditeur du message, Bob (Personne B) pour le destinataire et Eve (Evil Eve, de l'anglais) pour l'adversaire.

On peut aussi formaliser les procedures de cryptage et de décryptage. Soit X l'ensemble de tous les messages possibles de k lettres. Le cryptage est une bijection $f : X \to X$ et le décryptage est la fonction réciproque $f^{-1} : X \to X$. Donc si on veut transmettre un message de x lettres, on le divise en paquets de k lettres ou mois chaque, soit $x = x_1x_2\cdots x_n$, où chaque x_i a $\le k$ lettres, et on envoie le message $f(x_1)f(x_2)\cdots f(x_n)$. Bob possède une clé qui lui permettre de calculer rapidement $f^{-1}(f(x_i)) = x_i$ pour tout i. Puisque Eve n'a pas

accès à cette clé, la procedure de décryptage (c'est-à-dire de trouver x_i à partir de $f(x_i)$) doit être difficile. [1]

Les premiers cryptosystèmes développés se basaient sur la *cryptographie de clé privée*, appelée aussi *cryptographie symétrique*. Dans ce scheme, les clé de cryptage et de décryptage sont les deux privées et échangées entre Alice et Bob avant l'échange des messages. Certainement, l'échange des clés entre Alice et Bob devait être fait d'une façon sécuritaire aussi. Ceci est facile dès que Alice et Bob ont établit une communication sécuritaire : Alice peut créer un nouveau pair de clés et l'envoyer à Bob en utilisant les vieilles clés, déjà établies. Cependant, ce scheme souffre de plusieurs déficiences. Premièrement, si Eve peut découvrir la clé de décryptage seulement une fois, elle pourra lire toutes les communications futures. Par exemple, pendant la seconde guerre mondiale, les Allemagnes avaient développés un cryptosystème appelé *Enigma*. Les clés de cryptage et de décryptage étaient changées périodiquement en utilisant les vieilles clés. Une équipe de mathématiciens britanniques, amenée par Alan Turing, a réussi d'exploiter ce trou et de décrypter Enigma, un fait qui était d'importance cruciale pour le résultat de la guerre. Une autre déficience de la cryptographie de clé privée est que, avant leur première communication, Alice et Bob doivent se rencontrer au mois une fois pour échanger les clés de cryptage et de décryptage, ou de trouver une façon sécuritaire différente pour le faire. Bien sûr, ceci pourrait être très compliqué si, par exemple, on suppose que Alice est une personne en Italie qui veut envoyer quelques informations confidentielles à Bob, qui est une banque à la Chine. Il sera très difficile d'échanger les clés sans contact physique, qui est également difficile grâce à la distance géographique entre l'Italie et la Chine !

La vraie revolution dans le domaine de la cryptographie est venue avec la naissance de la *cryptographie de clé publique*, aussi appelée *cryptographie asymétrique*. Ce protocol de cryptographie, proposé par Ralph Merkle, utilise un scheme de communication asymétrique. Dans ce scheme, Bob, le récipient crée deux clés : une clé de décryptage, qui est privée et laquelle Bob garde pour lui-même, et une clé de cryptage, qui est publique. C'est-à-dire, Bob publie la fonction $f : X \to X$, qui crypte les informations, mais il garde secrete la fonction réciproque $f^{-1} : X \to X$. Puis, Alice utilise la fonction f pour crypter son message et l'envoyer à Bob, qui peut maintenant le décrypter avec la fonction f^{-1}. Ce scheme de communication enlève l'exigence de l'exchange des clés de cryptage et de décryptage entre Alice et Bob. Cependant, il rende la communication moins sécuritaire : Eve peut, en théorie, calculer f^{-1} à partir de f. L'hypothèse-clé de Merkle est que calculer f^{-1} est, en pratique, difficile si le seul donné est la fonction f. Il a développé un tel algorithme, et un autre exemple a été trouvé plus tard par Diffie et Hellman. Aujourd'hui, on appelle souvent le scheme de la cryptographie de clé publique l'*echange de clés de Diffie-Hellman-Merkle*. On décrit ici un exemple très important d'un tel algorithme, appelé l'*algorithme RSA*. Il a été publié en 1978 par Rivest, Shamir et Adleman.

Supposons que Alice veut envoyer un message M à Bob, qui a accès à une liste de grands nombres premiers. En utilisant cette liste, Bob forme un nombre n de la forme pq, où p et q sont deux nombres premiers de taille similaire. Bob, qui sait les nombres premiers premiers

1. La fonction f est une permutation. Donc, en principe, on peut trouver son inverse si on possède assez informations, mais ceci pourrait prendre beaucoup de temps. Le but d'Alice et de faire sûr que le message envoyé reste protégé pour le temps maximal possible.

p et q, peut aussi calculer $\phi(n) = (p-1)(q-1) = n+1-p-q$. Avec cette information, il calcule aussi un exposant E tel que $(E, \phi(n)) = 1$. Bob publie le pair (E, n) ; c'est la clé publique que Alice utilise pour crypter son message. Pour le faire, elle calcule le résidu de $M^E \bmod n$, soit $C \in \{0, 1, \dots, n-1\}$. Le résultat est le message crypté qu'elle envoie à Bob.

La question est maintenant, comment va Bob décrypter le message ? La clé est le théorème d'Euler : si $(M, n) = 1$, alors $M^{\phi(n)} \equiv 1 \pmod n$. Donc, il suffit que Bob calcule un nombre F tel que $EF \equiv 1 \pmod{\phi(n)}$, car dans ce cas $EF = 1 + k\phi(n)$ et, par la suite, $C^F = M^{EF} = M \cdot (M^{\phi(n)})^k \equiv M \pmod n$. Donc, en calculant $C^F \pmod n$, Bob peut déterminer $M \pmod n$. Pour déterminer la valuer exacte de M, il faut s'assurer que le nombre $n > M$, pour que le reste de $M \bmod n$ est égal à M. Alternativement (et plus efficacement pour les messages longues), Bob peut produire une liste de clés publiques (n_j, E_j), $j \le J$, où $n_1, \dots, n_k$ sont mutuellement co-premiers, et calculer $M \pmod{n_j}$ pour tout j. En appliquant le théorème des restes chinois, il peut alors calculer $M \pmod{n_1 \cdots n_k}$. Si $M < n_1 \cdots n_k$, il a réussi à calculer M soi-même.

Il reste trouver une façon rapide de calculer le nombre F. Puisque ce nombre est determiner par la relation $EF \equiv 1 \pmod{\phi(n)}$, il faut simplement inverser $E \pmod{\phi(n)}$. Ceci est facile par l'algorithme euclidien, pourvu qu'on connait E et $\phi(n)$. Tout le monde connait E (cela fait partie de la clé publique), et Bob connait p et q, donc il connait $\phi(n) = (p-1)(q-1) = n+1-p-q$. Les nombres (E, n) forment alors la clé publique et le nombre F la clé privée (qui peut être calculée rapidement si on connait $\phi(n)$).

On doit vérifier que les trois principes mentionnées au deuxième paragraphe sont satisfaites. Tout d'abord, est-ce que c'est possible de calculer rapidement $M^E \pmod n$ et $C^F \pmod n$? La réponse est affirmative et on l'explique dans le premier cas. Soit $E = e_0 + e_1 2 + e_2 2^2 \cdots + e_k 2^k$ l'expansion binaire de E. Alors

$$M^E = M^{e_0} (M^2)^{e_1} \cdots (M^{2^k})^{e_k}. \tag{2.10}$$

On sait les chiffres $e_0, e_1, \dots, e_k$ (habituellement, on travail avec l'expansion binaire dans un ordinateur ; en tout cas, c'est facile de calculer l'expansion de M à la base b, pour n'importe quel b). Donc il suffit de calculer les puissances $M, M^2, M^4, \dots, M^{2^k}$ modulo n, ce qu'on peut faire facilement dans une façon iterative : d'abord, on trouve le reste de $M \bmod n$, soit $M_0 \in \{0, 1, \dots, n-1\}$. Puis, on calcule M_0^2 et on trouve son reste mod n, soit M_1. En général, étant donné $M_i \in \{0, 1, \dots, N-1\}$ tel que $M_i \equiv M^{2^i} \pmod N$, on trouve $M_{i+1} \in \{0, 1, \dots, N-1\}$ tel que $M_{i+1} \equiv M_i^2 \equiv M^{2^{i+1}} \pmod n$. Étant donné M_i, on a besoin seulement d'une operation pour trouver M_{i+1}. Donc, par induction, on a besoin de $k+1$ operations pour trouver $M^{2^i} \pmod N$, pour tout $i \in \{0, 1, \dots, k\}$. Puis, on remplace M^{2^i} par M_i dans (2.10), et on calculer $M^E \pmod N$ avec $\le 2(k+1)$ operations. Puisque $k = \lfloor \log E / \log 2 \rfloor$, la calculation de $M^E \pmod n$ exige $\lesssim \log E$ operations. De même, la calculation de $C^F \pmod n$ exige $\approx \log F$ operations. On peut supposer que $1 \le E, F \le \phi(n) \le n$, donc le calcul de $M^E \pmod n$ et de $C^F \pmod n$ nécessite $\lesssim \log n$ approximations, une fonction linéaire du nombre de chiffres binaires de n.

On a vu que, étant données les clés publiques et privées, le cryptage et le décryptage du message M est très rapide. Cependant, on est aussi intéressé de la sécurité de l'algorithme. Eve a accès à la clé publique (n, E). Supposons aussi qu'elle a aussi réussi à lire le message

crypté $C \equiv M^E \pmod n$. Est-ce qu'elle peut deviner M à partir de ces informations ? La façon évidente pour le faire est d'essayer de calculer l'exposant F, c-est-à-dire la clé privée. Afin de l'accomplir, elle a besoin de connaître la valuer de $\phi(n) = n+1-p-q$. Puisque elle connait déjà $n = pq$, il suffit de trouver la valeur de $p+q$. Mais c'est la même chose que de connaître les premiers p et q, c'est-à-dire la factorisation première de n. Donc si Eve pouvait factoriser n rapidement, elle aurait accès au message crypté M. L'algorithme RSA s'appuie sur l'hypothèse qu'il est difficile de factoriser n, c'est-à-dire qu'il n'existe pas d'algorithme rapide pour le faire. Le meilleur algorithme connu présentement pour la factorisation de n a besoin de

$$\approx \exp\left\{ \sqrt[3]{\frac{64}{9}} (\log n)^{1/3} (\log\log n)^{2/3} \right\}$$

opérations au pire cas (et le pire cas consiste à un nombre n qui est le produit de deux nombres premiers p et q qui sont de taille similaire). C'est un algorithme exponentielle (au nombre de chiffres de n), donc il est assez lent, ce qui veut dire que Eve ne peut pas trouver F facilement. Le problème de factoriser un nombre donné est central à la cryptographie. En général, il est cru qu'un algorithme très rapide ('polynomial' au nombre de digits de n) n'existe pas, mais ceci est juste une espérance et non un théorème.

2.5 Équations polynomiales

Supposons que $f(x) = a_d x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0$ est un polynôme dont les coefficients $a_d, \ldots, a_0$ sont entiers. On veut étudier ses racines mod n. Une motivation pour le faire est pour étudier ses racines entières : si $f(\alpha) = 0$ pour un $\alpha \in \mathbb{Z}$, alors $f(\alpha) \equiv 0 \pmod n$, pour tout n. Par exemple, on peut utiliser cette observation pour montrer que le polynôme x^3+x+1 n'a pas de racines sur $\mathbb{Z}$: on observe que $x^3+x+1 \equiv 1 \pmod 2$ pour chaque x, donc x^3+x+1 n'a pas de racines dans $\mathbb{Z}/2\mathbb{Z}$ et *a fortiori* dans $\mathbb{Z}$. Si on veut étudier les racines rationnelles de $\mathbb{Z}$, on suppose que $x = a/b$ avec $(a,b) = 1$ est une telle racine et on multiplie l'équation $f(a/b) = 0$ par b^3 pour trouve que $a^3 + ab^2 + b^3 = 0$. On a alors une équation polynomiale en deux variables sur $\mathbb{Z}$. Afin de l'étudier, on peut la réduire mod n, pour un nombre convenant n. Par exemple, si on la réduit mod a, on trouve que $b^3 \equiv 0 \pmod a$. Mais $(a,b) = 1$, donc ceci est impossible. On a montré alors que l'équation $x^3 + x + 1$ n'a pas de solutions rationnelles.

Supposons maintenant qu'on veut étudier l'équation polynomiale générale $f(x) \equiv 0 \pmod n$. On observe tout d'abord que si $x \equiv y \pmod n$, alors $x^j \equiv y^j \pmod n$ pour chaque j. Puisque $f(x)$ est une combinaison linéaire de puissances de x, on trouve que $f(x) \equiv f(y) \pmod n$. Ceci veut dire que la valeur de $f(x) \pmod n$ dépende juste de la classe de $x \pmod n$. En particulier, on peut définir

$$\mathcal{R}_f(n) := \{x \in \mathbb{Z}/n\mathbb{Z} : f(x) \equiv 0 \pmod n\} \quad \text{et} \quad \rho_f(n) := \#\mathcal{R}_f(n).$$

Considérons maintenant la factorisation première de n, soit $n = p_1^{v_1} \cdots p_k^{v_k}$. La version bébé

du théorème de restes chinois implique que

$$x \,(\mathrm{mod}\, n) \in \mathcal{R}_f(n) \quad \Leftrightarrow \quad f(x) \equiv 0 \,(\mathrm{mod}\, n) \quad \Leftrightarrow \quad \begin{cases} f(x) \equiv 0 \,(\mathrm{mod}\, p_1^{v_1}) \\ \vdots \\ f(x) \equiv 0 \,(\mathrm{mod}\, p_k^{v_k}) \end{cases}$$
$$\Leftrightarrow \quad \begin{cases} x \,(\mathrm{mod}\, p_j^{v_j}) \in \mathcal{R}_f(p_1^{v_1}) \\ \vdots \\ x \,(\mathrm{mod}\, p)_k^{v_j} \in \mathcal{R}_f(p_j^{v_j}). \end{cases}$$

D'après le théorème de restes chinois, chaque k-tuple $(x_1, \ldots, x_k) \in \mathcal{R}_f(p_1^{v_1}) \times \mathcal{R}_f(p_k^{v_k})$ définit un unique résidu $x \,(\mathrm{mod}\, n)$. En particulier, on a montré que la fonction ρ_f est **multiplicative**.

On voit alors que l'étude des équations polynomials mod n se réduit à l'étude de ces équations quand n est une puissance d'un premier, soit p^v. Le cas difficile est $v = 1$, parce que dans cette section on verra une façon de trouver toutes les solutions mod p^v, $v \geq 1$, si on connait déjà les solutions mod p. Avant de le faire, on montre un résultat qui nous permet de controller le nombre de solutions mod p. Il est l'analogie d'un résultat bien connu en concernant les racines des polynômes dont les coefficients sont de nombres réels.

Théorème 2.29. *Soit $f(x) = a_d x^d + a_{d-1} x^{d-1} + \cdots + a_1 x + a_0 \in \mathbb{Z}[x]$ et soit p un nombre premier ne divisant pas tous les coefficients a_j de f. Si*

$$\mathcal{R}_f(p) = \{\alpha_1 \,(\mathrm{mod}\, p), \ldots, \alpha_r \,(\mathrm{mod}\, p)\},$$

où $r \geq 0$, alors il existe de nombres $m_1, \ldots, m_r \in \mathbb{N}$ et un polynôme $g(x) \in \mathbb{Z}[x]$ de degré $d - \sum_{j=1}^{r} m_j$ tels que :

(a) $g(x) \not\equiv 0 \,(\mathrm{mod}\, p)$, pour tout x ;

(b) les polynômes $f(x)$ et $(x - \alpha_1)^{m_1} \cdots (x - \alpha_r)^{m_r} g(x)$ ont les mêmes coefficients mod p.

En particulier, $\rho_f(p) \leq \sum_{j=1}^{r} m_j \leq d$.

Démonstration. Sans perte de généralité, on peut supposer que $p \nmid d$; sinon, on trouve $m = \max\{0 \leq j \leq d : p \nmid a_j\}$ et on montre le théorème pour le polynôme $a_m x^m + \cdots + a_1 x + a_0$, qui a les mêmes racines avec f mod p.

On utilise induction sur d. Si $d = 0$, nécessairement $f(x) = a_0 \not\equiv 0 \,(\mathrm{mod}\, p)$ pour chaque $x \in \mathbb{Z}$ de notre hypothèse que au moins un coefficient de f (le coefficient a_0 dans ce cas-ci) n'est pas divisible par p. Alors on peut prendre $g(x) = f(x)$ et le résultat en découle. Supposons maintenant que le résultat est vrai pour tous les polynômes de degré $< d$. Si l'ensemble $\mathcal{R}_f(p)$ est vide, on peut prendre $g(x) = f(x)$ pour déduire le théorème. Sinon,

soit $\alpha \pmod p \in \mathcal{R}_f(p)$. On observe que $x - \alpha$ est un facteur de $f(x) - f(\alpha)$: en effet,

$$\begin{aligned} f(x) - f(\alpha) &= \sum_{j=0}^{d} a_j x^j - \sum_{j=0}^{d} a_j \alpha^j = \sum_{j=0}^{d} a_j (x^j - \alpha^j) \\ &= \sum_{j=0}^{d} a_j (x-\alpha)(x^{j-1} + x^{j-2}\alpha_1 + x^{j-3}\alpha^2 + \cdots + x\alpha^{j-2} + \alpha_1^{j-1}) \\ &= (x-\alpha)\tilde{f}(x), \end{aligned}$$

où

$$\begin{aligned} \tilde{f}(x) &:= \sum_{j=0}^{d} a_j (x^{j-1} + x^{j-2}\alpha_1 + x^{j-3}\alpha^2 + \cdots + x\alpha^{j-2} + \alpha_1^{j-1}) \\ &= a_d x^{d-1} + \text{ plus petites puissances de } x. \end{aligned}$$

De plus, puisque $f(\alpha) \equiv 0 \pmod p$, on a que $f(x)$ et $(x-\alpha)\tilde{f}(\alpha)$ ont les mêmes coefficients mod p. En particulier,

$$f(x) \equiv (x-\alpha)\tilde{f}(x) \pmod p \quad \text{pour tout } x \in \mathbb{Z},$$

d'où on trouve que les racines de $\tilde{f}$ mod p sont aussi de racines de f mod p. Alors, soit $\mathcal{R}_{\tilde{f}}(p) = \mathcal{R}_f(p)$, soit $\mathcal{R}_{\tilde{f}}(p) = \mathcal{R}_f(p) \setminus \{\alpha \pmod p\}$ (on est dans le deuxième cas s-si $\tilde{f}(\alpha) \not\equiv 0 \pmod p$). Dans tous les cas, l'hypothèse inductive implique qu'il existe $m_1' \geq 0$, $m_2, \ldots, m_r \geq 1$, et un polynôme $g(x) \in \mathbb{Z}[x]$ de degré $d - 1 - m_1' - \sum_{j=2}^{r} m_j$ tels que

$$g(x) \equiv (x-\alpha_1)^{m_1'}(x-\alpha_2)^{m_2} \cdots (x-\alpha_r)^{m_r} g(x) \quad (x \in \mathbb{Z}).$$

Par conséquent, si on pose $m_1 = m_1' + 1 \geq 1$, on trouve que

$$f(x) = (x-\alpha_1)^{m_1}(x-\alpha_2)^{m_2} \cdots (x-\alpha_r)^{m_r} g(x) \quad (x \in \mathbb{Z}),$$

ce qui conclut l'étape inductive et, par la suite, la démonstration. □

Finalement, on montre comment on peut déterminer les racines d'un polynômes mod p^v en commençant avec ses racines mod p. Le résultat-clé est appelé le *lemme de Hensel*. Dans son énoncé, $f'(x)$ dénote la dérivée de $f(x) = a_d x^d + \cdots + a_1 x + a_0$, définie par

$$f'(x) = d a_d x^{d-1} + (d-1) a_{d-1} x^{d-2} + \cdots 2a_2 x + a_1.$$

Théorème 2.30 (Hensel)**.** *Soit $f(x) \in \mathbb{Z}[x]$. Si $f(x_1) \equiv 0 \pmod p$ et $f'(x_1) \not\equiv 0 \pmod p$, alors pour chaque $v \geq 2$ il existe une classe d'équivalence $x_v \pmod{p^v}$ unique pour laquelle $x_v \equiv x_1 \pmod p$ et $f(x_v) \equiv 0 \pmod{p^v}$.*

Le théorème au-dessus est un corollaire direct du lemme suivant.

Lemme 2.31 (Hensel)**.** *Soit p un nombre premier, $v \geq 1$ et $f(x) \in \mathbb{Z}[x]$. Supposons que $f(x_v) \equiv 0 \pmod{p^v}$ et considérons une classe de congruence $x_{v+1} \pmod{p^{v+1}}$ telle que $x_{v+1} \equiv x_v \pmod{p^v}$, c'est-à-dire $x_{v+1} = x_v + tp^v$ pour un $t \in \mathbb{Z}$. Alors on a que*

$$f(x_{v+1}) \equiv 0 \pmod{p^{v+1}} \quad \Leftrightarrow \quad f'(x_v) \cdot t \equiv -f(x_v)/p^v \pmod{p}.$$

En particulier, si $f'(x_v) \not\equiv 0 \pmod{p}$, alors il existe un unique $t \pmod{p}$ qui satisfait cette congruence et, par la suite, une unique classe de congruence $x_{v+1} \pmod{p^{v+1}}$ telle que $x_{v+1} \equiv x_v \pmod{p^v}$ et $f'(x_{v+1}) \equiv 0 \pmod{p^{v+1}}$.

Démonstration. Si $x_{v+1} = x_v + tp^v$, alors on a que

$$\begin{aligned} x_{v+1}^m = (x_v + tp^v)^m &= x_v^m + \binom{m}{1} x_v^{m-1} tp^v + \binom{m}{2} x_v^{m-2} (tp^v)^2 + \cdots \\ &\equiv x_v^m + m x_v^{m-1} tp^v \pmod{p^{v+1}}, \end{aligned}$$

puisque $2v \geq v + 1$ pour $v \geq 1$. Donc

$$\begin{aligned} f(x_v + tp^v) = \sum_{m=0}^{d} a_m (x_v + tp^v)^m &\equiv \sum_{m=0}^{d} a_m (x_v^m + m x_v^{m-1} tp^v) \pmod{p^{v+1}} \\ &\equiv f(x_v) + tp^v f'(x_v) \pmod{p^{v+1}}. \end{aligned}$$

On a que $p^v | f(x_v)$ par hypothèse. Par la suite

$$\begin{aligned} f(x_v + tp^v) \equiv 0 \pmod{p^{v+1}} \quad &\Leftrightarrow \quad tp^v f'(x_v) \equiv -f(x_v) \pmod{p^{v+1}} \\ &\Leftrightarrow \quad t f'(x_v) \equiv -f(x_v)/p^v \pmod{p}, \end{aligned}$$

comme affirmé. Finalement, si $f'(x_v) \not\equiv 0 \pmod{p}$, alors l'inverse de $f'(x_v)$ existe modulo p, ce qui implique que $t \equiv -(f'(x_v))^{-1} f(x_v)/p^v \pmod{p}$. Ceci conclut la démonstration. □

Démonstration du théorème 2.30. D'après le lemme 2.31 et induction sur v, on trouve qu'il existe une suite $x_2, x_3, x_4, \ldots$ telle que $x_v \pmod{p^v}$ est l'unique classe d'équivalence mod p^v pour laquelle $x_v \equiv x_{v-1} \pmod{p^{v-1}}$ et $f(x_v) \equiv 0 \pmod{p^v}$. Alors le résultat désiré suit tout de suite. □

Exemple 2.32. Résolvons l'équation quadratique

$$x^2 - 4x + 8 \equiv 0 \pmod{25}.$$

Solution. Soit $f(x) = x^2 - 4x + 8$. Tout d'abord, on résoudre l'équation modulo 5. Á fin de faire cela, on compléter le carré :

$$f(x) = (x-2)^2 - 2^2 + 8 = (x-2)^2 + 4 \equiv (x-2)^2 - 1^2 \pmod{5} \equiv (x-3)(x-1) \equiv \pmod{5}.$$

Alors

$$f(x) \equiv 0 \pmod{5} \quad \Leftrightarrow \quad x \equiv 3 \pmod{5} \quad \text{ou} \quad x \equiv 1 \pmod{5}.$$

Puis, on observe que $f'(x) = 2x-4$ et donc $f'(1) = -2 \not\equiv 0 \pmod 5$ et $f'(3) = 2 \not\equiv 0 \pmod 5$. Alors le lemme d'Hensel implique qu'il existe exactement une racine de f modulo 25, soit $a \pmod{25}$, telle que $a \equiv 1 \pmod 5$ et $f(a) \equiv 0 \pmod{25}$. De plus, si on pose $a = 1+5t$, alors

$$f'(1) \cdot t \equiv -f(1)/5 \pmod 5 \quad \Leftrightarrow \quad 2t \equiv 1 \pmod 5 \quad \Leftrightarrow \quad t \equiv 3 \pmod 5.$$

Donc $a \equiv 16 \pmod{25}$. De même, il existe exactement une racine de f modulo 25, soit $\overline{b}$, telle que $b \equiv 3 \pmod 5$ et $f(b) \equiv 0 \pmod{25}$. On pose $b = 3 + 5s$ pour que

$$f'(3)s \equiv -f(3)/5 \pmod 5 \quad \Leftrightarrow \quad 2s \equiv -1 \pmod 5 \quad \Leftrightarrow \quad s \equiv 2 \pmod 5.$$

Donc $b \equiv 13 \pmod{25}$.

Par conséquent, les solutions à l'équation $f(x) \equiv 0 \pmod{25}$ sont $x \equiv 13 \pmod{25}$ et $x \equiv 16 \pmod{25}$. □

Exercices

EXERCICE 2.1. Soit $n \in \mathbb{N}$ et $a, b \in \mathbb{Z}$. Montrez que $a^2 \equiv b^2 \pmod n$ si et seulement si il existe deux nombres naturels k et ℓ tels que $n = k\ell$, $a \equiv b \pmod k$ et $a \equiv -b \pmod \ell$.

EXERCICE 2.2. (a) Montrez que si n est impair, alors $n^2 \equiv 1 \pmod 8$.

(b) Est-ce qu'il y a de solutions à l'équation $p^\alpha + 1 = 2^\beta$, où p est premier et $\alpha, \beta \geq 2$? [*Indice :* Montrez que α doit être impair et factorisez $p^\alpha + 1$.]

EXERCICE 2.3. Montrez que pour tout entier positif n, le nombre $1^n + 2^n + 3^n + 4^n + 5^n + 6^n$ est divisible par 7 si et seulement si n n'est pas un multiple de 6.

EXERCICE 2.4. Trouvez un critères de divisibilité d'un nombre par 11, 13 et 37, si le nombre est donné dans la base décimale.

EXERCICE 2.5. Trouvez un critères de divisibilité d'un nombre par 10, si le nombre est donné dans la base binaire.

EXERCICE 2.6. Dans la base hexadécimale les chiffres sont $0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A = 10, B = 11, C = 12, D = 13, E = 14, F = 15$. Par exemple,

$$(10AB4)_{16} = 4+B\cdot 16+A\cdot 16^2+0\cdot 16^3+1\cdot 16^4 = 4+11\cdot 16+10\cdot 256+0+65536 = (68276)_{10}.$$

Déterminez si le nombre $(5A3BF204BC376F4A)_{16}$ est divisible par 5 (sans faire la conversion à la base de 10).

EXERCICE 2.7. Résolvez les suivants systèmes de congruences :

$$(a)\ \begin{cases} x \equiv 4 \pmod 7 \\ 3x \equiv 2 \pmod{11} \\ 7x \equiv 1 \pmod{13} \end{cases} \qquad (b)\ \begin{cases} x \equiv 2 \pmod{28} \\ 3x \equiv 8 \pmod{10} \end{cases} \qquad (c)\ \begin{cases} 2x \equiv 4 \pmod{10} \\ 3x \equiv 8 \pmod 7 \end{cases}$$

EXERCICE 2.8. Montrez que pour tous $m, n \in \mathbb{N}$, on a

$$\phi(mn) = \phi(m)\phi(n)\frac{(m,n)}{\phi((m,n))}.$$

EXERCICE 2.9. Montrez que si $n > 2$, alors $\phi(n)$ est un nombre pair.

EXERCICE 2.10. Soit p un nombre premier.

(a) Montrez que $x^p \equiv x \pmod{p}$ pour tous $x \in \mathbb{Z}$.
(b) Montrez que $x^{p^k} \equiv x \pmod{p}$ pour tous $x \in \mathbb{Z}$ et tous $k \in \mathbb{Z}_{\geq 1}$.
(c) Montrez que si $f(x) \in \mathbb{Z}[x]$, alors il existe deux polynômes $q(x), g(x) \in \mathbb{Z}[x]$ tels que : (a) $f(x) = q(x)(x^p - x) + g(x)$; (b) $\deg(g) < p$. [*Indice :* utilisez induction sur le degré de $f(x)$.]
(d) Si $f(x)$ et $g(x)$ sont comme ci-dessus, alors montrez que $\mathcal{R}_f(p) = \mathcal{R}_g(p)$.
(e) Si $f(x) \in \mathbb{Z}[x]$ a degré $d \geq 1$, alors on sait que $f'(x)$ est un polynôme sur $\mathbb{Z}$ de degré $d-1$. Cependant, ceci n'est pas toujours vrai sur $\mathbb{Z}/p\mathbb{Z}$: par exemple, le polynôme $x^p - x$ a comme dérivée $px^{p-1} - 1$ qui est égale à -1 mod p. Montrez, quand même, que si $d < p$, alors $\deg(f') = d - 1$.

EXERCICE 2.11. Soient p un premier, $f(x) \in \mathbb{Z}[x]$ et $x_0 \in \mathbb{Z}$.

(a) Montrez qu'il existe quelques nombres $a_j \in \mathbb{Z}$ uniques tels que $f(x) = \sum_{j=0}^{d} a_j(x - x_0)^j$, où $d = \deg(f)$. [*Indice :* Considérez le polynôme $f(x - x_0)$.]
(b) Montrez que $f(x_0) \equiv 0 \pmod{p}$ si et seulement si $a_0 \equiv 0 \pmod{p}$.
(c) Montrez que $f(x_0) \equiv 0 \pmod{p}$ et $f'(x_0) \equiv 0 \pmod{p}$ si et seulement si $a_0 \equiv a_1 \equiv 0 \pmod{p}$.

EXERCICE 2.12. Trouvez toutes les solutions aux équations polynomials suivantes :

(a) $x^2 \equiv 2 \pmod{49}$
(b) $x^3 + x + 1 \equiv 0 \pmod{27}$
(c) $x^2 - 2x + 3 \equiv 0 \pmod{16}$
(d) $x^2 - x + 10 \equiv 0 \pmod{50}$

EXERCICE 2.13. Dans le lemme de Hensel, on étudie des racines x_0 de l'équation $f(x) \equiv 0 \pmod{p^\nu}$ avec $p \nmid f'(x_0)$. Décrivez un algorithme pour déterminer les racines de l'équation $f(x) \equiv 0 \pmod{p^\nu}$ dans le cas où $p | f'(x_0)$. Appliquez votre algorithme pour résoudre l'équation $x^2 \equiv a \pmod{p^\nu}$, où a, p et ν sont donnés. Déterminez quand il existe de solutions et calculez leur nombres.

Chapitre 3

Éléments de la théorie des groupes

Dans ce chapitre, on présente quelques définitions et résultats venant de la théorie des groupes qui vont nous permettre d'apprécier l'arithmétique modulaire d'un point de vue plus abstrait, ainsi que de faciliter quelques démonstrations.

3.1 Le concept d'un groupe

Soit un ensemble G muni d'une opération $*$. Souvent, l'opération $*$ est la multiplication $\cdot$ ou l'addition $+$. Par exemple, on pourrait avoir :

Ex. 1 : l'ensemble $\mathbb{Z}$ muni de l'opération $+$;

Ex. 2 : l'ensemble $\mathbb{Z}_{\geq 0}$ muni de l'opération $+$;

Ex. 3 : l'ensemble $\mathbb{Z}_{>0} = \{1, 2, \ldots\}$ muni de l'opération $\cdot$;

Ex. 4 : l'ensemble $\mathbb{Z}_{<0}$ muni de l'opération $\cdot$;

Ex. 5 : l'ensemble $\mathbb{Q}_{>0}$ muni de l'opération $\cdot$;

Ex. 6 : l'ensemble $\mathbb{R}_{\geq 0}$ muni de l'opération $a * b := |a - b|$.

Pour l'économie de la discussion, on utilise toujours le symbole $*$ dénotant une opération abstraite (qui pourrait être la multiplication ou l'addition familière). Le pair $(G, *)$ est appelé un groupe s'il satisfait quatre propriétés :

- (loi de composition interne) Si $g, h \in G$, alors $g * h \in G$, i.e. le résultat de la multiplication de deux éléments de G est toujours un nouveau élément de G. Ceci veut dire que l'exemple 3 ci-dessus n'est pas un groupe : on a, par exemple, $-1 \cdot (-1) = 1 \notin \mathbb{Z}_{<0}$.
- (associativité) Pour tous $g, h, j \in G$, on a que $(g * h) * j = g * (h * j)$, c'est-à-dire l'ordre dans laquelle on execute les opérations n'est pas importante. Ceci veut dire que l'exemple 6 ci-dessus n'est pas un groupe car $2 * (1 * 1) = 2 * 0 = 2$, mais $(2 * 1) * 1 = 1 * 1 = 0$.
- (élément neutre) il existe un élément neutre dénoté par e ayant la propriété que $e * g = g$ et $g * e = g$ pour tout $g \in G$. Si l'opération de G est dénoté par $\cdot$, alors on dénote l'élément neutre de G par 1 ; si l'opération est dénotée par $+$, alors on le dénote par 0.

  L'élément neutre est unique s'il existe : si $u \in G$ est un autre neutre élément, alors $e * u = e$ et $u * e = u$ par la neutralité de e et de u par la gauche et par la droite, respectivement.

L'élément neutre dans les exemples 1, 2 et 6 est le nombre 0 ; l'élément neutre dans les exemples 3 et 5 est le nombre 1 ; il n'y a pas d'élément neutre dans l'exemple 3.

— (élément inverse/opposé) pour tout $g \in G$, il existe un élément inverse, c'est-à-dire un $h \in G$ tel que $g * h = e = h * g$. On peut montrer qu'il est unique et on le symbolise par g^{-1}. Une exception est quand l'opération de G est l'addition, dans quel cas on appelle h l'élément opposé de g et on le symbolise par $-g$; il satisfait les relations $g + h = 0 = h + g$.

Afin de voir l'unicité de h, supposons que j est un autre inverse de g. On considère l'élément $x := (j*g)*h$ et on le calcule de deux façons. D'un côté, on a que $x = e*h = h$ et, d'autre côté, $x = j * (g * h) = j * e = j$. On en déduit que $j = h$.

L'axiom de l'existence d'inverses implique que l'exemple 2 ci-dessus n'est pas un groupe car l'opposé additif d'un nombre positif est negatif. Donc, l'opposé de $m > 0$ n'appartient pas à $\mathbb{Z}_{\geq 0}$. De façon similaire, dans l'exemple 3, l'inverse du nombre 2 n'existe pas dans l'ensemble $\mathbb{Z}_{>0}$.

C'est facile de vérifier que les seuls groupes entre les exemples 1-6 se donnent par les pairs $(\mathbb{Z}, +)$ et $(\mathbb{Q}_{>0}, \cdot)$. Des autres exemples des groupes sont : $(\mathbb{R}, +)$, $(\mathbb{R} \setminus \{0\}, \cdot)$, $(\mathbb{Q}, +)$, les matrices réels 2×2 munis de l'addition de matrices, ... Cependant, pour le but de ce livre, les exemples les plus importants sont :

$$(\mathbb{Z}/n\mathbb{Z}, +) \quad \text{et} \quad ((\mathbb{Z}/n\mathbb{Z})^*, \cdot).$$

C'est facile de vérifier qu'ils satisfassent les quatre axioms ci-dessus - on le laisse aux lecteurs comme un exercice.

Tous les exemples des groupes qu'on a donné satisfassent un cinquième axiom :

— (commutativité) Pour tous $g, h \in G$, on a que $g * h = h * g$.

Si le paire $(G, *)$ satisfait cet axiom additionnel, on dit qu'il forme un groupe abélien ou commutatif. L'étude de ces groupes est assez plus simple. Il y a des exemples des groupes qui ne sont pas abéliens. Un tel exemple est le groupe des matrices 2 qui sont inversibles : l'ensemble

$$\mathrm{GL}_2(\mathbb{R}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc \neq 0 \right\},$$

muni de la multiplication de matrices. On a que

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \neq \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

3.2 Sous-groupes

Un concept important est d'un sous-groupe d'un groupe $(G, *)$. On dit que H est un sous-groupe de G et on écrit $H < G$ si $e \in H$ et $(H, *)$ est aussi un groupe, c'est-à-dire si : (a) $e \in H$; (b) $xy \in H$ chaque fois que $x, y \in H$; (c) $x^{-1} \in H$ chaque fois que $x \in H$.

Une caractérisation simple d'être sous-groupe suit :

Lemme 3.1. *Soit $(G, *)$ et $H \subset G$. On a que $H < G$ si et seulement si :*

— $H \neq \emptyset$;
— *pout tous* $a, b \in H$, *on a que* $a * b^{-1} \in H$.

Démonstration. Si $H < G$, alors la conclusion du lemme est claire : on a que $e \in H$, donc $H \neq \emptyset$. De plus, si $a, b \in H$, alors $b^{-1} \in H$ (existence d'élément inverse) et donc $a * b^{-1} \in H$ (loi de composition interne).

Réciproquement, supposons que H satisfait la conclusion du lemme. Soit $h \in H$ (il existe au moins un tel élément par l'hypothèse que $H \neq \emptyset$). Donc $e = h * h^{-1} \in H$. De plus, $h^{-1} = e * h^{-1} \in H$ (existence d'élément inverse). Finalement, si h' est un autre élément de H, on va montrer que $h' * h \in H$. Tout d'abord, on note que $(h^{-1})^{-1} = h$. En effet, $h^{-1} * h = e = h * h^{-1}$, donc h satisfait les conditions pour être l'inverse de h^{-1} (donc il est son inverse unique). Vu que $h', h^{-1} \in H$, on trouve que $h' * (h^{-1})^{-1} \in H$ et, par la suite, $h' * h \in H$, comme affirmé. Ceci montre que $(H, *)$ est un groupe (l'associativité découle automatiquement du fait que H est un sous-ensemble de G). □

En utilisant le lemme 3.1, les lecteurs peuvent vérifier facilement que :

(a) L'ensemble $2\mathbb{Z}$ des entiers pairs est un sous-groupe de $(\mathbb{Z}, +)$;
(b) L'ensemble $\mathbb{Z}$ est un sous-groupe de $(\mathbb{R}, +)$;
(c) L'ensemble $\{-1, 1\}$ est un sous-groupe de $(\mathbb{Q}, \cdot)$;
(d) L'ensemble $\{x \pmod n : d|n\}$ est un sous-groupe de $(\mathbb{Z}/n\mathbb{Z}, +)$;
(e) l'ensemble des matrices diagonales non-nuls $\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$, $a \neq 0$, est un sous-groupe de $(\mathrm{GL}_2(\mathbb{R}), \cdot)$.

3.3 Génération de groupes

Soit $(G, *)$ et $X \subset G$. Peut-être X n'est pas un sous-groupe de G et se demande quelle est la façon la plus 'économique' de le rendre un. Il faut ajouter quelques éléments à X et arriver à un sur-ensemble de X qui est un sous-groupe de G. On dénote cet ensemble hypothétique par $\langle X \rangle$. Par exemple, s'il existe $x \in X$ tel que $x^{-1} \notin X$, alors il faut que x^{-1} est ajouté à $\langle X \rangle$. De façon similaire, s'il existe $x, y \in X$ tels que $xy \notin X$, il faut ajouter xy à $\langle X \rangle$. Continuant comme cela jusqu'à l'infini, on peut construire $\langle X \rangle$. Cependant, ceci n'est pas une démonstration. De façon rigoureuse, on définite

$$\langle X \rangle := \bigcap_{\substack{H < G \\ H \supset X}} H.$$

L'exercice 3.7 montre que, en effet, le côté droit est un sous-groupe de G comme l'intersection de quelques sous-groupes de G. Évidemment, il est le *plus petit* sous-groupe de G contenant X.

On montre maintenant que la définition formelle qu'on a donné est d'accord avec la définition plus intuitive dont on a parlé avant :

Théorème 3.2. *Soit $(G, *)$ un groupe et $X \subset G$. On a que*

$$\langle X \rangle = \{x_1^{\varepsilon_1} * \cdots * x_k^{\varepsilon_k} : x_1, \ldots, x_k \in X,\ \varepsilon_1, \ldots, \varepsilon_k \in \{-1, +1\},\ k \in \mathbb{Z}_{\geq 0}\},$$

avec la convention qu'un produit vide est égal à e. (Par exemple, $\langle \emptyset \rangle = \{e\}$.)

Démonstration. Soit

$$\tilde{X} = \{x_1^{\varepsilon_1} * \cdots * x_k^{\varepsilon_k} : x_1, \ldots, x_k \in X,\ \varepsilon_1, \ldots, \varepsilon_k \in \{-1, +1\},\ k \in \mathbb{Z}_{\geq 0}\}.$$

Il est assez évident que $\langle X \rangle \subset \tilde{X}$. En effet, soit $\tilde{x} = x_1^{\varepsilon_1} * \cdots * x_k^{\varepsilon_k}$ dans $\tilde{X}$, où $x_1, \ldots, x_k \in X$ et $\varepsilon_1, \ldots, \varepsilon_k \in \{-1, +1\}$. Si H est un sous-groupe de G contentant X, alors H contient $x_1, \ldots, x_k$. Mais H est un groupe, donc il contient $x_1^{\varepsilon_1}, \ldots, x_k^{\varepsilon_k}$ pour n'importe quel choix de $\varepsilon_1, \ldots, \varepsilon_k \in \{-1, +1\}$. Par la suite, H doit contenir leur produit, c'est-à-dire $\overline{x}$. Ceci montre que $\tilde{X} \subset H$, pour chaque $H < G$ avec $H \subset X$. Par la définition de $\langle X \rangle$, on déduit que $\tilde{X} \subset \langle X \rangle$.

Finalement, on montre que $\langle X \rangle \subset \tilde{X}$. Vu que $\tilde{X}$ contient X (en prenant $k = 1$ et $\varepsilon_1 = 1$ à la définition de $\tilde{X}$), et vu que $\langle X \rangle$ est le plus petit sous-groupe de G contenant X, il suffit de montrer que $\tilde{X}$ est un sous-groupe de G. Pour le faire, on observe que $e \in \tilde{X}$ (par convention : on a supposé que le produit vide est égal à e). De plus, si $\tilde{x} = x_1^{\varepsilon_1} * \cdots * x_k^{\varepsilon_k}$ et $\tilde{y} = y_1^{\delta_1} \cdots y_\ell^{\delta_\ell}$, où $x_1, \ldots, x_k, y_1, \ldots, y_\ell \in X$ et $\varepsilon_1, \ldots, \varepsilon_k, \delta_1, \ldots, \delta_\ell \in \{-1, +1\}$, alors on a que

$$\tilde{x} * \tilde{y}^{-1} = x_1^{\varepsilon_1} * \cdots * x_k^{\varepsilon_k} * y_\ell^{-\delta_\ell} * \cdots * y_1^{-\delta_1},$$

ce qui est évidement un nouvel élément de $\tilde{X}$. En appliquant le lemme 3.1, on voit que $\tilde{X}$ est un sous-groupe de X. Ceci termine la démonstration. □

Un cas important du théorème 3.2 est donné par le sous-groupe de G engendré par un seul élément g : on a que

$$\langle g \rangle = \{g^k : k \in \mathbb{Z}\},$$

où $g^0 := e$, $g^k := \underbrace{g * \cdots * g}_{k \text{ fois}}$ pour $k \in \mathbb{Z}_{>0}$, et $g^k := \underbrace{g^{-1} * \cdots * g^{-1}}_{k \text{ fois}}$ pour $k \in \mathbb{Z}_{<0}$. On peut vérifier que

$$g^{k+\ell} = g^k * g^\ell$$

en examinant des cas différents (selon le signe de k et de ℓ). Donc l'ensemble $\{g^k : k \in \mathbb{Z}\}$ est un sous-groupe de G contenant g, et il est certainement le plus petit sous-groupe de G ayant cette propriété. Donc $\langle g \rangle = \{g^k : k \in \mathbb{Z}\}$, comme affirmé.

Dans le cas où l'opération du groupe est dénotée par $+$, on a que

$$\langle g \rangle = \{k \cdot g : k \in \mathbb{Z}\},$$

où $0 \cdot g := 0$, $k \cdot g := \underbrace{g + \cdots + g}_{k \text{ fois}}$ pour $k \in \mathbb{Z}_{>0}$, et $k \cdot g := \underbrace{(-g) + \cdots + (-g)}_{k \text{ fois}}$ pour $k \in \mathbb{Z}_{<0}$. Par exemple, si $G = \mathbb{Z}$ et l'opération est l'addition des entiers, alors

$$\langle 2 \rangle = 2\mathbb{Z},$$

l'ensemble des entiers pairs.

Si le groupe G est abélien et X est fini, l'énoncé du théorème 3.2 est simplifiée : on a que

$$\langle g_1, \ldots, g_m \rangle = \{g_1^{k_1} * \cdots * g_m^{k_m} : k_1, \ldots, k_m \in \mathbb{Z}\}.$$

On dit que le groupe G est engendré par les éléments g_i, $i \in I$, si $G = \langle \{g_i : i \in I\} \rangle$. Les éléments g_i sont appelés les générateurs de G. En général, les générateurs ne sont pas définis uniquement. Si G est engendré par un seul élément, c'est-à-dire il existe $g \in G$ tel que $G = \langle g \rangle$, alors on dit que G est cyclique. Par exemple, le groupe $(\mathbb{Z}, +)$ est cyclique, engendré par l'élément 1 (ainsi que par l'élément -1). De façon similaire, le groupe $(\mathbb{Z}/n\mathbb{Z}, +)$ est cyclique, engendré par l'élément 1. En général, chaque résidu réduit $a \pmod n$ est un générateur du groupe additif $\mathbb{Z}/n\mathbb{Z}$. Les groupes multiplicatifs $(\mathbb{Z}/n\mathbb{Z})^*$ sont beaucoup plus compliqués, comme on va le voir. On va les étudier au prochain chapitre. On remarque, par exemple, que $(\mathbb{Z}/6\mathbb{Z})^* = \langle -1 \rangle$, que $(\mathbb{Z}/7\mathbb{Z})^* = \langle 3 \rangle$, et que $(\mathbb{Z}/15\mathbb{Z})^* = \langle 7, 11 \rangle$.

3.4 L'ordre d'un élément

Soit $(G, *)$ un groupe et $g \in G$. On considère $\langle g \rangle$, le sous-groupe de G engendré par g. On définit l'ordre de g comme suite : soit l'ensemble $K = \{k \in \mathbb{N} : g^k = e\}$. Alors, l'ordre de g est définie par

$$\operatorname{ord}(g) := \begin{cases} +\infty & \text{si } K = \emptyset, \\ \min K & \text{sinon.} \end{cases}$$

Si G est fini, l'ensemble K est toujours non-vide, donc k est un nombre naturel fini. En effet, si $n = |G|$, les puissances $g^0, g^1, \ldots, g^n$ ne peuvent être toutes distinctes. Soit $g^i = g^j$ pour $0 \leq i < j \leq n$. En multipliant à gauche par g^{-i}, on trouve que $g^{j-i} = e$, donc $j - i \in K$.

Dans le cas où $K \neq \emptyset$, on a que

$$\langle g \rangle = \{e, g, g^2, \ldots, g^{k-1}\}. \tag{3.1}$$

En effet, c'est clair que le côté droit est un sous-ensemble du côté gauche. Vice-versa, si $m \in \mathbb{Z}$, alors $m = qk + r$ pour quelques $q \in \mathbb{Z}$ et $r \in \{0, 1, \ldots, k-1\}$. Donc $g^m = g^{qk+r} = (g^k)^q * g^r = g^r$, ce qui montre (3.1). Cet argument nous amène au lemme suivant :

Lemme 3.3. *Soit $(G, *)$ un groupe et $g \in G$ d'ordre $k < \infty$. On a que $g^m = e$ si et seulement si $k|m$.*

Démonstration. C'est clair que si $k|m$, alors $g^m = e$ (puisque $g^k = e$ et $m = qk$). Vice versa, si $g^m = 1$ et on écrit $m = qk + r$, alors $e = g^m = g^r$. Mais $k = \min\{j \in \mathbb{N} : g^j = e\}$ et $0 \leq r < k$, d'où on trouve que $r = 0$. Ceci conclut la démonstration. □

En utilisant ce lemme, on montre un théorème très utile :

Théorème 3.4. *Soit $(G, *)$ un groupe* **abélien fini**. *Si $n = |G|$, alors l'ordre de chaque élément de G est un diviseur de G. En particulier, $g^n = e$ pour tout $g \in G$.*

Démonstration. Soit $g \in G$. Considérons l'application $f : G \to G$, définie par $f(h) := g * h$. Elle est une bijection : $f(h) = f(h')$ veut dire que $gh = gh'$. En multipliant à gauche par g^{-1}, on trouve que $h = h'$, ce qui montre l'injectivité de f. Puisque G est fini, alors f est bijective. Par la suite, les éléments $g * h$, $h \in G$, sont une permutation des éléments de G, ce qui implique que

$$\prod_{h\in G}(g * h) = \prod_{h\in G} h,$$

c'est-à-dire

$$g^n * \prod_{h\in G} h = \prod_{h\in G} h.$$

En multipliant à droit par l'inverse de $\prod_{h\in G} h$, on trouve que $g^n = e$. Donc $\mathrm{ord}(g)|n$ d'après le lemme 3.3. □

En appliquant le théorème precedent au groupe $G = (\mathbb{Z}/n\mathbb{Z})^*$ des résidus réduits mod n, on arrive au théorème d'Euler, dont un cas particulier est le petit théorème de Fermat :

Corollaire 3.5. *(a) (théorème d'Euler) Si $(a, n) = 1$, alors $a^{\phi(n)} \equiv 1 \pmod n$.*

(b) (petit théorème de Fermat) Si p est premier et a est un entier non-divisible par p, alors $a^{p-1} \equiv 1 \pmod p$.

Soit $(G, *)$ un groupe abélien. Si $k = \mathrm{ord}(g)$ et $\ell = \mathrm{ord}(h)$, alors c'est clair que $(gh)^{[k,\ell]} = e$, donc $\mathrm{ord}(gh)|[k, \ell]$. On peut montrer une relation plus précise dans un cas particulier :

Lemme 3.6. *Soit $(G, *)$ un groupe.*

(a) Supposons que G est abélien. Si $k = \mathrm{ord}(g)$ et $\ell = \mathrm{ord}(h)$ sont **finis et copremiers**, *alors $\mathrm{ord}(gh) = k\ell$.*

(b) $k = \mathrm{ord}(g) < \infty$ et $d|k$, alors $\mathrm{ord}(g^d) = k/d$.

Démonstration. (a) Soit $m = \mathrm{ord}(gh)$. Comme on l'a déjà discuté, on a que $m|[k, \ell] = k\ell$. Vice versa, on observe que $(gh)^m = e$. Donc $(gh)^{mk} = e$ et, puisque $g^k = e$, alors $h^{mk} = e$. En appliquant le lemme 3.3, on trouve que $\ell|mk$ et, en utilisant la coprimalité de k et de ℓ, on trouve que $\ell|m$. De façon similaire, en échangeant les rôles de g et de h, on trouve aussi que $k|m$. En utilisant la coprimalité de k et de ℓ une dernière fois, on conclut que $k\ell|m$. Ceci conclut la démonstration.

(b) Evidemment, on a que $(g^d)^{k/d} = e$. De plus, k/d est le plus petit nombre ayant cette propriété par la minimalité de k. Donc $k/d = \mathrm{ord}(g^d)$, comme affirmé. □

3.5 L'exposant d'un groupe abélien

Le théorème 3.4 nous dit que si $(G, *)$ est un groupe abélien d'ordre n, alors $g^n = e$. Est-ce que n est le plus petit nombre ayant cette propriété ? En général, la réponse est non : par exemple, si $G = (\mathbb{Z}/8\mathbb{Z})^*$, alors $x^2 \equiv 1 \pmod 8$ pour chaque x impair. On définit alors l'exposant de G par

$$\exp(G) := \min\{k \geq 1 : g^k = e \ \forall g \in G\},.$$

Dans le cas spécial où $G = (\mathbb{Z}/n\mathbb{Z})^*$, on écrit

$$\lambda(n) := \exp((\mathbb{Z}/n\mathbb{Z})^*).$$

La fonction λ est appelée la fonction de Carmichael.

On a le résultat suivant qui donne une définition alternative pour $\exp(G)$:

Théorème 3.7. *Si $(G, *)$ est un groupe abélien fini, alors*

$$\exp(G) = \max\{\mathrm{ord}(g) : g \in G\}.$$

En particulier, $\exp(G)$ *divise* $|G|$.

Démonstration. Soit $r = \exp(G)$ et $k = \max\{\mathrm{ord}(g) : g \in G\}$. Puisque G est fini, il existe $g_0 \in G$ tel que $k = \mathrm{ord}(g_0)$. D'après la définition de r, on a que $g_0^r = 1$ et le lemme 3.3 implique que $k|r$. Il reste de montrer que $k \geq r$. Pour le montrer, il suffit de prouver que $g^k = e$ pour chaque $g \in G$ ou, de façon équivalente, que $\mathrm{ord}(g)|k$ pour chaque $g \in G$. Supposons, au contraire, qu'il existe $g_1 \in G$ tel que $\mathrm{ord}(g_1) \nmid k$. Donc, si on pose $\ell = \mathrm{ord}(g_1)$, on trouve qu'il existe un nombre premier p pour lequel $v_p(\ell) > v_p(k)$. On écrit $\ell = p^w\ell_1$ et $k = p^v k_1$, où $w = v_p(\ell) > v_p(k) = v$. En particulier, $p \nmid \ell_1 k_1$. On construira un élément de G d'ordre $p^w k_1 > k$, ce qui est une contradiction à notre hypothèse que g_0 a ordre maximale k. Pour le faire, on utilise le lemme 3.6 : la partie (b) de ce lemme implique que l'élément $g_1^{\ell_1}$ a ordre p^w, et que l'élément $g_0^{p^v}$ a ordre k_1. Donc la partie (a) du lemme 3.6 implique que l'élément $g_1^{\ell_1} g_0^{p^v}$ a ordre $p^w k_1 > k$. Ceci est impossible par le choix du k. On en déduit que $\ell|k$, comme on le voulait. Ceci termine la démonstration. □

Proposition 3.8. *Si $(m, n) = 1$, alors $\lambda(mn) = [\lambda(m), \lambda(n)]$.*

Démonstration. Soit $k = \lambda(m)$ et $\ell = \lambda(n)$. Si $(x, mn) = 1$, alors $(x, m) = (x, n) = 1$ et donc

$$x^k \equiv 1 \,(\mathrm{mod}\, m) \quad \text{et} \quad x^\ell \equiv 1 \,(\mathrm{mod}\, n),$$

de la définition de λ. Puisque $[k, \ell]$ est un multiple de k et de ℓ, on trouve que $x^{[k,\ell]} \equiv 1 \,(\mathrm{mod}\, m)$ et que $x^{[k,\ell]} \equiv 1 \,(\mathrm{mod}\, n)$. Donc, $x^{[k,\ell]} \equiv 1 \,(\mathrm{mod}\, mn)$, d'après le lemme 2.17. Puisque cette relation est vraie pour n'importe quel $x \in (\mathbb{Z}/mn\mathbb{Z})^*$, on déduit que $[k, \ell] \geq \lambda(mn)$.

Afin de montrer l'inégalité inverse, on considère $a \in (\mathbb{Z}/m\mathbb{Z})^*$ et $b \in (\mathbb{Z}/n\mathbb{Z})^*$ d'ordre maximale, soit k et ℓ, respectivement. En particulier, la Proposition 3.7 implique que $k = \lambda(m)$ et que $\ell = \lambda(n)$. On applique le théorème de restes chinois pour trouver un $x \,(\mathrm{mod}\, mn)$ tel que $x \equiv a \,(\mathrm{mod}\, m)$ et $x \equiv b \,(\mathrm{mod}\, n)$. On affirme que $\mathrm{ord}_{mn}(x) = [k, \ell]$. En effet, on a que

$$\begin{aligned} x^r \equiv 1 \,(\mathrm{mod}\, mn) \quad &\Leftrightarrow \quad \begin{cases} x^r \equiv 1 \,(\mathrm{mod}\, m) \\ x^r \equiv 1 \,(\mathrm{mod}\, n) \end{cases} \quad \Leftrightarrow \quad \begin{cases} a^r \equiv 1 \,(\mathrm{mod}\, m) \\ b^r \equiv 1 \,(\mathrm{mod}\, n) \end{cases} \\ &\Leftrightarrow \quad \begin{cases} k = \mathrm{ord}_m(a)|r \\ \ell = \mathrm{ord}_n(b)|r \end{cases} \quad \Leftrightarrow \quad [k, \ell]|r. \end{aligned}$$

Ceci implique que l'ordre de x dans $(\mathbb{Z}/mn\mathbb{Z})^*$ est égale à $[k, \ell]$, comme affirmé. Par conséquent, la proposition 3.7 implique que $\lambda(mn) \geq [k, \ell]$, ce qui conclut la preuve. □

Corollaire 3.9. *Si $n = p_1^{v_1} \cdots p_r^{v_r}$ est la factorisation du nombre n à ses facteurs premiers, alors*

$$\lambda(n) = [\lambda(p_1^{v_1}), \ldots, \lambda(p_r^{v_r})].$$

Démonstration. Par induction sur r. □

On calculera $\lambda(p^v)$ pour chaque nombre premier p et chaque $v \geq 1$ au prochaine chapitre.

3.6 Groupes isomorphes

Deux groupes $(G, *)$ et $(G', *)$ peuvent être secrètement le même groupe. Par exemple, on considère les groupes $(\{-1, 1\}, \cdot)$ et $(\mathbb{Z}/2\mathbb{Z}, +)$. On affirme que ces deux groupes sont pratiquement 'égaux'. En effet, les tables de ce deux groupes sont :

$\cdot$	1	-1
1	1	-1
-1	-1	1

$+$	0	1
0	0	1
1	1	0

Les tables sont identiques : le rôle de 1 au premier table est joué par 0 au deuxième table (ces sont les éléments neutres), et le rôle de -1 au premier table est joué par 1 au deuxième table. On dit donc que les groupes $\{-1, 1\}$ et $\mathbb{Z}/2\mathbb{Z}$ sont isomorphes.

De façon plus générale, on dit que les groupes $(G, *)$ et $(G', *)$ (les opérations peuvent être différentes même si on utilise le même symbole) sont isomorphes s'il existe une bijection $f : G \to G'$ telle que

$$f(a * b) = f(a) * f(b)$$

pour tous $a, b \in G$. Dans ce cas, on écrit $G \cong G'$. La fonction f est appelée un isomorphisme.

La relation d'isomorphisme des groupes est une relation d'équivalence. Les groupes appartenant à la même classe d'équivalence sont 'égaux' du point de vue de la théorie des groupes (les seules choses qui sont différentes sont les noms/symbols qu'on attribue à leurs éléments et à leur opération).

Exemple 3.10. Si $(G, *)$ et $(G', *)$ sont deux groupes, alors l'ensemble $G \times G' = \{(g, g') : g \in G, g' \in G'\}$ est un groupe par rapport à l'opération $(g, g') * (h, h') := (g * h, g' * h')$. L'élément neutre est le pair (e, e'), où e l'élément neutre de G et e' de G'. L'inverse de (g, g') est donné par $(g^{-1}, (g')^{-1})$.

Si $(m, n) = 1$, alors on affirme que

$$\mathbb{Z}/mn\mathbb{Z} \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}, \tag{3.2}$$

où l'opération est l'addition. De plus, on affirme que

$$(\mathbb{Z}/mn\mathbb{Z})^* \cong (\mathbb{Z}/m\mathbb{Z})^* \times (\mathbb{Z}/n\mathbb{Z})^*. \tag{3.3}$$

Ces relations sont une conséquence du théorème des restes chinois. On montre la première et laisse la deuxième comme exercice. On définit $f : \mathbb{Z}/mn\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ par $f(x \pmod{mn}) := (x \pmod{m}, x \pmod{n})$. Comme dans la démonstration du théorème 2.24, on a que f est une bijection. De plus, c'est facile de vérifier que $f(x+y) = f(x) + f(y)$, d'où la relation (3.2) suit.

3.7 Le groupe quotient

Soit $(G, *)$ un groupe et soit $H < G$. À partir de H, on peut définir une relation d'équivalence sur G : on dit que deux éléments g_1, g_2 de G sont équivalents, et on écrit $g_1 \equiv g_1 \pmod H$, si $g_1 * g_2^{-1} \in H$. C'est facile de vérifier que c'est un relation d'équivalence [1].

On peut vérifier facilement que la classe d'équivalence de g et l'ensemble $H * g := \{h * g : h \in H\}$. En effet, si $g' \equiv g \pmod H$, alors $g' * g^{-1} \in H$, c'est-à-dire il existe $h \in H$ tel que $g' * g^{-1} = h$. En multipliant à gauche par g, on trouve que $g' = h * g \in H * g$. On montre de même que si $g' = h * g$ pour un $h \in H$, alors $g \equiv g' \pmod H$. Ceci montre notre affirmation que la classe d'équivalence de g est l'ensemble $H * g$.

Quand G est commutative, alors $H * g = g * H = \{g * h : h \in H\}$. De plus, l'ensemble des classes d'équivalence $H * g$, $g \in G$, dénoté par G/H, devient un groupe. Pour cette raison on l'appelle le groupe quotient de G sur H. Son opération est définie par $(H * g_1) * (Hg_2) := H * (g_1 * g_2)$. Elle est bien définie (i.e. elle ne dépende pas du choix des représentants g_1 et g_2) : si $g_1 \equiv g_3 \pmod H$ et $g_2 \equiv g_4 \pmod H$, alors $g_3 = h_1 * g_1$ et $g_4 = h_2 * g_2$ pour quelques $h_1, h_2 \in H$. Par la suite, $g_3 * g_4 = h_1 * g_1 * h_2 * g_2 = h_1 * h_2 * g_1 * g_2 \equiv g_1 * g_2 \pmod H$, puisque $h_1 * h_2 \in H$, où on a utilisé la commutativité de G. Puisque les classes d'équivalences de deux éléments équivalents sont identiques, on trouve que $H * (g_1 * g_2) = H * (g_3 * g_4)$. C'est facile maintenant de vérifier que l'opération définie ci-dessus rende G/H un groupe : les axioms sont des conséquences directes des axioms de G. L'élément neutre est la classe d'équivalence $H * e = H$.

Remarque 3.11. C'est clair que si G est fini, toutes les classes d'équivalences $H * g$ contiennent le même nombre d'éléments, qui est égal à $|H|$. Puisque il y a $|G/H|$ classes d'équivalences, on trouve que $|H|$ divise $|G|$ et, plus précisément, que

$$|G| = |H| \cdot |G/H|.$$

En prenant $H = \langle g \rangle$, ceci offre une autre démonstration du théorème 3.4 (en fait, cet argument n'a pas besoin de la commutativité de G).

Exemple 3.12. Si on considère le groupe $(\mathbb{Z}, +)$ et son sous-groupe $n\mathbb{Z} = \{m \in \mathbb{Z} : n|m\}$, alors on le groupe quotient $\mathbb{Z}/n\mathbb{Z}$ est le groupe des résidus mod n qu'on a vu au dernier chapitre. Donc, on voit que le groupe quotient généralise le concept de l'arithmétique modulaire.

Remarque 3.13. Le groupe multiplicatif $(\mathbb{Z}/n\mathbb{Z})^*$ n'est pas un groupe quotient. C'est possible de le réaliser d'un point de vu général comme le groupe des éléments inversibles de l'anneau $\mathbb{Z}/n\mathbb{Z}$ (veuillez ignorer ce remarque si vous n'avez pas suive le cours de l'algèbre 2).

Exemple 3.14. Si $G = (\mathbb{Z}/n\mathbb{Z})^*$ est le groupe multiplicatif mod n, alors $H = \{a^2 \pmod n : (a, n) = 1\}$ est un sous-groupe de G. Le quotient G/H sera le sujet de la section 4.2.

1. Réflexivité : $g \equiv g \pmod H$ car $g * g^{-1} = e \in H$; symétrie : si $g_1 \equiv g_2 \pmod H$, alors $g_1 * g_2^{-1} \in H$ et donc son inverse $(g_1 * g_2^{-1})^{-1} = g_2 * g_1^{-1} \in H$ (justifiez pourquoi l'inverse de $g_1 * g_2^{-1}$ est égal à $g_2 * g_1^{-1}$). On trouve alors que $g_2 \equiv g_1 \pmod H$. Transitivité : si $g_1 \equiv g_2 \pmod H$ et $g_2 \equiv g_3 \pmod H$, alors $g_1 * g_2^{-1}, g_2 * g_3^{-1} \in H$ et, par la suite, $g_1 * g_2^{-1} * g_2 * g_3^{-1} = g_1 * g_3^{-1} \in H$, ce qui implique que $g_1 \equiv g_3 \pmod H$.

Supposons, maintenant, que on a deux groupes $(G, *)$ et $(G', *)$. Une fonction $f : G \to G'$ est appelée un morphisme de groupes si elle respecte les lois des deux groupes, c'est-à-dire si

$$f(g_1 * g_2) = f(g_1) * f(g_2) \quad \text{pour tous } g_1, g_2 \in G.$$

Si f est un morphisme bijectif, on l'appelle un isomorphisme. En particulier, les groupes G et G' sont isomorphes si et seulement il existe un isomorphisme $f : G \to G'$. Le noyau de f est l'ensemble

$$\ker(f) := \{g \in G : f(g) = e'\},$$

où e' dénote l'élément neutre de G.

Par exemple, la fonction $f : \mathbb{Z} \to \mathbb{Z}$ définie par $f(x) = nx$ est un morphisme de groupes. Son noyau est trivial, c'est-à-dire $\ker(f) = \{0\}$. (Comme on va le voir tout de suite, l'élément neutre de G appartient toujours à $\ker(f)$.)

Lemme 3.15. *Si $f : G \to G'$ est un morphisme de groupes, alors* $\ker(f) < G$.

Démonstration. On montre d'abord que $\ker(f) \neq \emptyset$ en montrant qu'il contient e. Puisque $e * e = e$, on trouve que $f(e) = f(e * e)$. Mais $f(e * e) = f(e) * f(e)$, d'où $f(e) = f(e) * f(e)$. En multipliant par $f(e)^{-1}$, on trouve que $f(e) = e'$.

Soit, maintenant, $g \in \ker(f)$. Alors $e' = f(e) = f(g * g^{-1}) = f(g) * f(g^{-1})$. Puisque $f(g) = e$, on en déduit que $f(g^{-1}) = e$ et donc $g^{-1} \in \ker(f)$.

Finalement, si $g_1, g_2 \in \ker(f)$, alors $f(g_1) = f(g_2^{-1}) = e$. Donc

$$f(g_1 * g_2^{-1}) = f(g_1) * f(g_2^{-1}) = e' * e' = e'.$$

On trouve alors que $g_1 * g_2^{-1} \in \ker(f)$. D'après le lemme 3.1, ceci montre que $\ker(f)$ est un sous-groupe de G. □

Un théorème très important à l'étude des groupes est le premier théorème d'isomorphisme de groupes qu'on montre au cas special de groupes abéliens. On a besoin de définir le concept d'un épimorphisme, qui est simplement un morphisme de groupes $f : G \to G'$ qui est surjectif (i.e. son image est égale à G').

Lemme 3.16. *Si $f : G \to G'$ est un épimorphisme de groupes et G est abélien, alors*

$$G/\ker(f) \cong G'.$$

Démonstration. On pose $H = \ker(f)$ et on définit la fonction $\phi : G/H \to G'$ par

$$\phi(H * g) := f(g).$$

Cette fonction est bien définie : si $g_1 \equiv g_2 \pmod H$, alors $g_1 * g_2^{-1} \in H = \ker(f)$, c'est-à-dire $f(g_1 * g_2^{-1}) = e'$. En multipliant par $f(g_2)$, on trouve alors que

$$f(g_2) = f(g_1 * g_2^{-1}) * f(g_2) = f(g_1 * g_2^{-1} * g_2) = f(g_1).$$

Donc la définition de ϕ ne dépende pas du choix du représentant de la classe d'équivalence gH.

C'est clair que ϕ est surjectif, car f l'est. De façon similaire, ϕ est un morphisme de groupes car f l'est. Donc, ϕ est un isomorphisme de groupes, ce qui montre le théorème. □

Exemple 3.17. Si $G = \langle g \rangle$ est un groupe cyclique et $n = \mathrm{ord}(g) \in \mathbb{N} \cup \{+\infty\}$, alors on définit $f : \mathbb{Z} \to G$ par $f(m) := g^m$. C'est un épimorphisme de groupes. On a que

$$f(m) = e \quad \Leftrightarrow \quad g^m = e \quad \Leftrightarrow \quad n|m \quad \Leftrightarrow \quad \begin{cases} m = 0 & \text{si } n = +\infty, \\ m \in n\mathbb{Z} & \text{si } 1 \leq n < \infty. \end{cases}$$

Donc le théorème 3.16 implique que

$$G \cong \begin{cases} \mathbb{Z} & \text{si } n = +\infty, \\ \mathbb{Z}/n\mathbb{Z} & \text{si } 1 \leq n < \infty. \end{cases}$$

Alors on voit que, modulo isomorphisme, les seuls groupes cycliques sont les groupes $\mathbb{Z}$ et $\mathbb{Z}/n\mathbb{Z}$.

Exercices

EXERCICE 3.1. Pour quels entiers positifs n l'expression $3^n + 1$ est-elle un multiple de 10 ?

EXERCICE 3.2. Soient $a \in \mathbb{N}$ et p un nombre premier.

(a) Si $(a, p) = 1$, montrez que soit $a^{\frac{p-1}{2}} \equiv 1 \pmod{p}$ soit $a^{\frac{p-1}{2}} \equiv -1 \pmod{p}$.

(b) Montrez que $\frac{n^5}{5} + \frac{n^3}{3} + \frac{7n}{15}$ est un nombre entier pour chaque $n \in \mathbb{N}$.

EXERCICE 3.3. Montrez que

$$\sum_{d|n} \phi(d) = n$$

pour chaque $n \in \mathbb{N}$. [*Indice :* observez que $n = \sum_{d|n} \#\{1 \leq x \leq n : (x, n) = d\}$.]

EXERCICE 3.4. Le petit théorème de Fermat implique que $x^p \equiv x \,(\mathrm{mod}\, p)$, pour chaque $x \in \mathbb{Z}$. Cet exercice donne une nouvelle démonstration de ce fait.

(a) Montrez que $p|\binom{p}{k}$, pour chaque $k \in \{1, \ldots, p-1\}$.

(b) Montrez que $(a + b)^p \equiv a + b \,(\mathrm{mod}\, p)$, pour chaque $a, b \in \mathbb{Z}$.

(c) Montrez que la fonction $f(x) = x^p - x$ est constante mod p. Déduisez que $x^p \equiv x \,(\mathrm{mod}\, p)$, pour chaque $x \in \mathbb{Z}$.

EXERCICE 3.5. Montrez que $x^{48} \equiv 1 \pmod{224}$, pour tout $x \in \mathbb{Z}$ qui est copremier à 14.

EXERCICE 3.6. Montrez que si $n = pq$, où p et q sont deux nombres premiers distincts, alors

$$a^{\phi(n)/2} \equiv 1 \,(\mathrm{mod}\, n),$$

pour tout $a \in \mathbb{Z}$ qui est copremier à n.

EXERCICE 3.7. Soit $(G, *)$ un groupe, et soit $(G_i)_{i \in I}$ une collection non-vide de sous-groupes de G. Montrez que $\cap_{i \in I} G_i$ est un sous-groupe de G.

EXERCICE 3.8. Soit $A = (\mathbb{Z}/p\mathbb{Z})[x]$, l'ensemble de polynômes ayant coefficients dans $\mathbb{Z}/p\mathbb{Z}$. Si $f(x) = \sum_{j=0}^{m} a_j x^j$ et $g(x) = \sum_{j=0}^{n} b_j x^j$ sont deux éléments de A, alors on définit $f(x) + g(x) := \sum_{j=0}^{\max\{m,n\}} (a_j + b_j) x^j$, où on l'a posé $a_j = 0$ si $j > m$ et $b_j = 0$ si $j > n$, et l'addition $a_j + b_j$ est faite mod p.

(a) Montrez que $(A, +)$ est un groupe abelien.

(b) Montrez que $(f(x)+g(x))^p = f(x)+g(x)$ pour tous $f(x), g(x) \in A$. [*Indice :* Consultez l'exercice 3.4.]

(c) Montrez que l'application $\phi : A \to A$, définie par $\phi(f(x)) = f(x)^p$ est un morphisme injectif de groupes. (Il est appelé le morphisme de *Frobenius*.)

EXERCICE 3.9. Soit $(G, +)$ un groupe abelien. Étant donné $n \in \mathbb{N}$ et $g \in G$, définissons $n \cdot g := \underbrace{g + \cdots + g}_{n \text{ fois}}$. Si $n = 0$, alors on pose $0 \cdot g = 0$, et si $n \in \mathbb{Z}_{<0}$, on pose $n \cdot g := (-n) \cdot (-g)$.

(a) Montrez que $(m + n) \cdot g = m \cdot g + n \cdot g$, ainsi que $(mn) \cdot g = m \cdot (n \cdot g)$, pour tous $m, n \in \mathbb{Z}$ et $g \in G$.

(b) Montrez que $m \cdot (g + h) = m \cdot g + m \cdot h$, pour tous $m \in \mathbb{Z}$ et $g, h \in G$.

(c) Soit $g \in G$. Montrez que l'application $\phi_g : \mathbb{Z} \to G$, définie par $\phi_g(n) := n \cdot g$ est un morphisme de groupes. Si ϕ_g n'est pas injectif, alors prouvez qu'il existe un nombre premier $p = p(g)$ tel que $\ker(\phi_g) = p\mathbb{Z}$.

(d) Soit $n \in \mathbb{N}$ et définissons $G_n = \{g \in G : n \cdot g = 0\}$. Montrez que G_n est un sous-groupe de G.

(e) Supposez que G est également muni d'une opération multiplicative $*$ qui est commutative et compatible avec $+$, c'est-à-dire $g*(h+j) = g*h+g*j$. Supposez, de plus, qu'il existe un nombre premier p tel que $p \cdot g = 0$ pour tous $g \in G$. (Un tel G est l'ensemble A de l'exercice précedent.) Montrez que l'application $G \ni g \to g^p \in G$ (où g^p veut dire $\underbrace{g * \cdots * g}_{p \text{ fois}}$, comme habituellement) est un morphisme de groupes.

Chapitre 4

Le groupe $(\mathbb{Z}/n\mathbb{Z})^*$

4.1 Racines primitives

Le but de cette section est de déterminer la structure du groupe multiplicatif $(\mathbb{Z}/n\mathbb{Z})^*$. D'après la relation 3.3, si $n = p_1^{v_1} \cdots p_r^{v_r}$ est la factorisation de n en facteurs premiers, alors

$$(\mathbb{Z}/n\mathbb{Z})^* \cong (\mathbb{Z}/p_1^{v_1}\mathbb{Z})^* \times \cdots \times (\mathbb{Z}/p_r^{v_r}\mathbb{Z})^*. \tag{4.1}$$

Donc, notre but se réduit à l'étude des groupes $(\mathbb{Z}/p^v\mathbb{Z})^*$, où p est premier. On montrera que la majorité de ces groupes sont cycliques, c'est-à-dire il existe un résidu réduit $a \pmod n$ tel que $(\mathbb{Z}/p^v\mathbb{Z})^* = \langle a \pmod{p^v} \rangle$. Un tel a est appelé une racine primitive mod p^v. De façon analogue, on peut définir la notion d'une racine primitive pour n'importe que modulus n.

Les deux théorèmes principaux qu'on montrera dans cette sections sont les suivants :

Théorème 4.1. *Soit p un premier et $v \in \mathbb{N}$. Si $p > 2$ ou si $v \in \{1, 2\}$, alors*

$$(\mathbb{Z}/p^v\mathbb{Z})^* \cong \mathbb{Z}/(p^v - p^{v-1})\mathbb{Z},$$

c'est-à-dire il existe de racines primitives mod p^v. Finalement, si $p = 2$ et $v \geq 3$, alors

$$(\mathbb{Z}/2^v\mathbb{Z})^* \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2^{v-2}\mathbb{Z}.$$

Théorème 4.2. *Le nombre n possède de racines primitives si et seulement si $n \in \{1, 2, 4\} \cup \{p^v : p > 2$ premier, $v \geq 1\} \cup \{2p^v : p > 2$ premier, $v \geq 1\}$.*

On commence en étudiant le cas de moduli premiers :

Théorème 4.3. *Si p est un nombre premier, alors il existe de racines primitives mod p.*

Démonstration. D'après le théorème 3.7, il suffit de montrer que $\lambda(p) = p - 1$. Si $k = \lambda(p)$, on a que $x^k \equiv 1 \pmod p$, pour tout $x \in \{1, 2, \ldots, p-1\}$. Donc le polynôme $f(x) = x^k - 1$ a $\geq p - 1$ racines mod p. D'autre côte, $f(x)$ peut avoir $\leq k = \deg(f)$ racines mod p selon le théorème 2.29. Donc $k \geq p - 1$. Puisque on a toujours que $k | \phi(p) = p - 1$, on en déduit que $k = p - 1$, ce qui conclut la démonstration. □

Démonstration alternative du théorème 4.3. On peut éviter l'utilisation du théorème 2.29 dans la démonstration du théorème 4.3. Cet argument alternatif se base à l'identité

$$\sum_{a=1}^{p-1} a^k \equiv \begin{cases} -1 \pmod p & \text{si } (p-1)|k, \\ 0 \pmod p & \text{sinon,} \end{cases} \tag{4.2}$$

pour chaque $k \in \mathbb{Z}_{k\geq 0}$. Avant de la montrer, on en déduit l'existence de racines primitives mod p.

On pose $k = \lambda(p)$. Nécessairement, $1 \leq k \leq \phi(p) = p-1$. De plus, la définition de $\lambda(p)$ implique que

$$\sum_{a=1}^{p-1} a^k \equiv \sum_{a=1}^{p-1} 1 \equiv p-1 \equiv -1 \pmod p.$$

Donc la relation (4.2) implique que $(p-1)|k$ et, par la suite, $k \geq p-1$. Ceci montre que $k = p-1$ et l'existence d'une racine primitive mod p découle du lemme **??**.

Finalement, on montre (4.2). Puisque $a^{k+p-1} \equiv a^k \pmod p$, pour chaque $a \in \{1, \ldots, p-1\}$, une consequence du petit théorème de Fermat, il suffit de montrer (4.2) quand $0 \leq k \leq p-2$. On pose

$$S_k = \sum_{a=1}^{p-1} a^k \quad \text{et} \quad \delta_k = \begin{cases} 1 & \text{si } (p-1)|k, \\ 0 & \text{sinon.} \end{cases}$$

On montrera que $S_k \equiv -\delta_k \pmod p$ pour $k \in \{0, 1, \ldots, p-2\}$ par induction sur k. Si $k = 0$, alors

$$S_0 = \sum_{a=1}^{p-1} 1 = p-1 \equiv -1 \pmod p \equiv -\delta_0 \pmod p.$$

Puis, supposons que $S_j \equiv -\delta_j \pmod p$ pour chaque $j \in \{0, 1, \ldots, k-1\}$, où $1 \leq k \leq p-2$. On montrera que $S_k \equiv -\delta_k \equiv 0 \pmod p$ également. On a que

$$\sum_{a=1}^{p-1}((a+1)^{k+1} - a^{k+1}) = \sum_{a=1}^{p-1}(a+1)^{k+1} - \sum_{a=1}^{p-1} a^{k+1} = -1 + p^{k+1} + S_{k+1} - S_{k+1} \equiv -1 \pmod p.$$

D'autre côté,

$$\sum_{a=1}^{p-1}((a+1)^{k+1} - a^{k+1}) = \sum_{a=1}^{p-1}\sum_{j=0}^{k} \binom{k+1}{j} a^j = \sum_{j=0}^{k} \binom{k+1}{j} S_j \equiv (k+1)S_k - 1 \pmod p,$$

par l'hypothèse inductive. En comparant les deux relations au-dessus, on conclut tout-de-suite que

$$(k+1)S_k \equiv 0 \pmod p \quad \Longrightarrow \quad S_k \equiv 0 \pmod p,$$

puisque $p \nmid k+1$ par notre hypothèse que $1 \leq k \leq p-2$. Ceci conclut la démonstration du théorème.

□

À partir du théorème précédent, on calcule la valuer de la fonction de Carmichael sur les arguments qui sont une puissance d'un nombre premier. Il est important de noter ici que le théorème 3.7 implique que le modulo n possède de racines primitives si et seulement si $\lambda(n) = \phi(n)$. Donc, le théorème ci-dessus établit, entre autres, l'existence de racines primitives mod p^v quand p est un nombre premier impair.

Théorème 4.4. *Soit p un nombre premier et $v \geq 1$. On a que*

$$\lambda(p^v) = \begin{cases} \phi(p^v) & \text{si } p > 2, \\ \phi(p^v) & \text{si } p = 2 \text{ et } v \leq 2, \\ \phi(p^v)/2 & \text{si } p = 2 \text{ et } v \geq 3. \end{cases}$$

Démonstration. **Cas 1 :** p impair. L'égalité $\lambda(p^v) = \phi(p^v)$ est équivalent à l'existence de racines primitives mod p^v. On distingue trois sous-cas :

Sous-cas 1a : $v = 1$. Le résultat désiré est le théorème 4.3.

Sous-cas 1b : $v = 2$. Soit g une racine primitive mod p. Si $x \pmod{p^2}$ tel que $x \equiv g \pmod{p}$ et on pose $k = \mathrm{ord}_{p^2}(x)$, on a que $k|\phi(p^2) = p(p-1)$ et que $x^k \equiv 1 \pmod{p^2}$. Donc $g^k \equiv x^k \pmod{p} \equiv 1 \pmod{p}$, ce qui implique que $\mathrm{ord}_p(g) = p - 1|k$. Puisque $k|p(p-1)$, les seules possibilités sont $k = p - 1$ ou $k = p(p-1)$. On doit montrer qu'il existe x tel que $\mathrm{ord}_{p^2}(x) = p(p-1)$. De façon équivalente, on doit trouver x tel que $x \equiv g \pmod{p}$ et que $x^{p-1} \not\equiv 1 \pmod{p^2}$. On considère le polynôme $f(x) = x^{p-1} - 1$. On a que $f(g) \equiv 0 \pmod{p}$ et que $f'(g) = (p-1)g^{p-2} \not\equiv 0 \pmod{p}$. Donc le lemme de Hensel implique qu'il existe $x_0 \pmod{p^2}$ *unique* tel que $x_0 \equiv g \pmod{p}$ et $f(x_0) \equiv 0 \pmod{p^2}$. Mais il existe exactement $p \geq 2$ classes d'équivalence modulo $x \pmod{p^2}$ telles que $x \equiv g \pmod{p}$. Donc il existe exactement $p - 1 \geq 1$ classes d'équivalence modulo $x \pmod{p^2}$ telles que $x \equiv g \pmod{p}$ et $f(x) \not\equiv 0 \pmod{p^2}$. Pour chaque telle classe d'équivalence $x \pmod{p^2}$, on a que $\mathrm{ord}_{p^2}(x) = p(p-1)$, c'est-à-dire, x est une racine primitive mod p^2.

Sous-cas 1c : $v \geq 3$. Soit g une racine primitive mod p^2, qui existe du sous-cas 1b. On montrera que g est une racine primitive mod p^v, pour chaque $v \geq 3$. Il suffit de montrer que $g^{p^{v-2}(p-1)} \not\equiv 1 \pmod{p^v}$, pour chaque $v \geq 3$. En effet, si $k = \mathrm{ord}_{p^v}(g)$, on a que $g^k \equiv 1 \pmod{p^v}$ et, par conséquent, $g^k \equiv 1 \pmod{p^2}$. Donc on a que $p(p-1) = \mathrm{ord}_{p^2}(g)|k$. Aussi, on a que $k|\phi(p^v) = p^{v-1}(p-1)$, du théorème d'Euler. Alors $k = p^j(p-1)$ pour un nombre $j \in \{1, \ldots, v-1\}$. Par conséquent, g est une racine primitive si et seulement si $j = v - 1$, si et seulement si $g^{p^{v-2}(p-1)} \not\equiv 1 \pmod{p^v}$, comme clamé.

On montrera que $g^{p^{v-2}(p-1)} \not\equiv 1 \pmod{p^e}$, pour chaque $v \geq 2$, de façon inductive. Si $v = 2$, c'est vrai de notre hypothèse que g est une racine primitive mod p^2. Supposons maintenant que le résultat tient pour un $v \geq 2$. On a que $g^{p^{v-2}(p-1)} = g^{\phi(p^{v-1})} \equiv 1 \pmod{p^{v-1}}$, du théorème d'Euler. Donc $g^{p^{v-2}(p-1)} = 1 + bp^{v-1}$ pour un $b \in \mathbb{Z}$, où $p \nmid b$ car $g^{p^{v-2}(p-1)} \not\equiv 1 \pmod{p^v}$. Par

la suite,

$$\begin{aligned} g^{p^{v-1}(p-1)} = (1+bp^{v-1})^p &= 1 + \binom{p}{1} bp^{v-1} + \binom{p}{2}(bp^{v-1})^2 + \binom{p}{3}(bp^{v-1})^3 + \cdots \\ &\equiv 1 + \binom{p}{1} bp^{v-1} + \binom{p}{2}(bp^{v-1})^2 \pmod{p^{v+1}} \\ &\equiv 1 + bp^v + \frac{p-1}{2} bp^{2v-1} \pmod{p^{v+1}} \\ &\equiv 1 + bp^v \pmod{p^{v+1}}, \end{aligned}$$

parce que $3(v-1) \geq v+1$ et $2v-1 \geq v+1$ pour $v \geq 2$. Puisque $p \nmid b$, on déduit que $g^{p^{v-1}(p-1)} \not\equiv 1 \pmod{p^{v+1}}$, ce qui conclut l'étape inductive et, par conséquent, la démonstration que p^v possède des racines primitives.

Cas 2 : $p = 2$. Si $v \in \{1, 2\}$, c'est facile de verifier qu'il existe une racine primitive mod 2^v. Donc on a que $\lambda(2^v) = \phi(2^v)$ dans ces cas.

Finalement, on considère le cas où $p = 2$ et $v \geq 3$. On peut verifier directement que $\lambda(8) = 2$ et $\lambda(16) = 4$, comme clamé. On montre les autres cas par induction. On suppose que $\lambda(2^w) = \phi(2^w)/2 = 2^{w-2}$ pour $w \in \{3, \ldots, v\}$, où $v \geq 4$, et on prouve que $\lambda(2^{v+1}) = 2^{v-1}$.

D'abord, on montre que $x^{2^{v-1}} \equiv 1 \pmod{2^{v+1}}$, pour chaque x impair, ce qui implique tout de suite que $\lambda(2^{v+1}) \leq 2^{v-1}$. En effet, l'hypothèse inductive implique que $\lambda(2^{v-1}) = 2^{v-3}$ et, par la suite, $x^{2^{v-3}} \equiv 1 \pmod{2^{v-1}}$. Donc $x^{2^{v-3}} = 1 + 2^{v-1}b$ pour un $b \in \mathbb{Z}$. Alors on déduit que

$$x^{2^{v-1}} = (1 + 2^{v-1}b)^4 = 1 + 2^{v+1}b + 6 \cdot 2^{2v-2}b^2 + 4 \cdot (2^{v-1}b)^3 + (2^{v-1}b)^4 \equiv 1 \pmod{2^{v+1}},$$

ce qui prouve notre affirmation.

Finalement, on montre que $\lambda(2^{v+1}) \geq 2^{v-1}$. Soit g impair tel que $\mathrm{ord}_{2^v}(g) = 2^{v-2}$. Il suffit de prouver que $\mathrm{ord}_{2^{v+1}}(g) = 2^{v-1}$. En effet, soit $b \in \mathbb{Z}$ tel que $g^{2^{v-3}} = 1 + 2^{v-1}b$, comme ci-dessus. Nécessairement b est impair : sinon, on aurait que $g^{2^{v-3}} \equiv 1 \pmod{2^v}$, c'est-à-dire $\mathrm{ord}_{2^v}(g) \leq 2^{v-3} < 2^{v-2}$. Donc

$$g^{2^{v-2}} = (1 + 2^{v-1}b)^2 = 1 + 2^v b + 2^{2v-2}b^2 \equiv 1 + 2^v b \pmod{2^{v+1}} \not\equiv 1 \pmod{2^{v+1}},$$

puisque $2 \nmid b$. Par conséquent, $\mathrm{ord}_{2^{v+1}}(g) = 2^{v-1}$, comme affirmé. C'implique que $\lambda(2^{v+1}) \geq 2^{v-1}$ selon la Proposition 3.7, ce qui conclut la démonstration du théorème. □

Remarque 4.5. La démonstration du théorème 4.4 dans le cas de moduli 2^v avec $v \geq 3$ nous permet de déterminer un g explicit d'ordre maximale 2^{v-2} : on a que 5 a ordre $2 = 2^{3-2}$ mod 8, et ordre $4 = 2^{4-2}$ mod 16. Donc il a ordre 2^{v-2} mod 2^v, pour tout $v \geq 3$, d'après la démonstration.

On est maintenant près de montrer les résultats principaux de cette section :

Preuve du théorème 4.1. Le premier cas du corollaire suit directement du théorème 4.4 et de l'exemple 3.17. Considérons maintenant le cas ou le modulus est 2^v avec $v \geq 3$. D'après le remarque 4.5, on sait que 5 a ordre maximale égale à 2^{v-2} modulo 2^v. On affirme que

chaque nombre impair n a une representation unique mod 2^v de la forme $n \equiv \pm 5^k \pmod{2^v}$, où $k \in \{0, 1, \ldots, 2^{v-2}-1\}$. En effet, tous les nombres de la forme 5^k sont $1 \pmod 4$. De plus, il existe 2^{v-2} puissances distinctes de 5 mod 2^v, et aussi il existe 2^{v-2} résidus réduits mod 2^v qui sont $1 \pmod 4$. Par la suite, on trouve que si x est impair, alors $x \equiv 1 \pmod 4$ si et seulement si $x \equiv 5^k \pmod{2^v}$ pour un $k \in \{0, 1, \ldots, 2^{v-2}-1\}$. D'autre côté, si $x \equiv 3 \pmod 4$, alors $-x \equiv 1 \pmod 4$, donc $-x \equiv 5^k \pmod{2^v}$ pour un $k \in \{0, 1, \ldots, 2^{v-2}-1\}$. Ceci montre notre affirmation que chaque x impair a une representation de la forme $\pm 5^k$ modulo 2^v.

On construit maintenant un morphisme $f : \mathbb{Z}/2^{v-2}\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \times G$ par

$$f(k, \ell) := 5^k \cdot (-1)^\ell \pmod{2^v}.$$

Le fait 5 a ordre 2^{v-2} et -1 a ordre 2 implique que f est bien définie, ne dependant pas du choix de $k \pmod{2^v}$ et de $\ell \pmod 2$. Cette fonction est un morphisme, comme on peut facilement vérifier. Elle est surjective par la discussion ci-dessus. Puisque les groupes $\mathbb{Z}/2^{v-2}\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ et G ont la même cardinalité, la fonction f doit être aussi injective. On voit donc que les groupes $\mathbb{Z}/2^{v-2}\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ et G sont isomorphes. Ceci conclut la démonstration du théorème 4.1. □

Démonstration du théorème 4.2. Soient

$$\mathcal{M} = \{1, 2, 4\} \cup \{p^e : p \text{ nombre premier impair}\} \cup \{2p^e : p \text{ nombre premier impair}\}$$

et

$$\mathcal{N} = \{n \in \mathbb{N} : \text{il existe de racines primitives mod } n\}.$$

" $\Rightarrow$ " : Si $n \in \mathcal{M}$, alors $\lambda(n) = \phi(n)$ du corollaire 3.9 et du théorème 4.4. Donc $n \in \mathcal{N}$.

" $\Leftarrow$ " : On montrera que si $n \notin \mathcal{M}$, alors $n \notin \mathcal{N}$.

Si n possède deux facteurs premiers pairs distincts, soient p_1 et p_2, alors on peut écrit $n = p_1^{v_1} p_2^{v_2} m$, où $v_1, v_2 \geq 1$ et $p_1, p_2 \nmid m$. De plus, $\phi(p_i^{v_i}) = (p_i - 1)p_i^{v_i - 1}$ et pair pour $i \in \{1, 2\}$. Donc Proposition 3.8 implique que $\lambda(n) = [\lambda(p_1^{v_1}), \lambda(p_2^{v_2}), \lambda(m)]$. Puisque $\lambda(a) | \phi(a)$ pour chaque $a \in \mathbb{N}$, une conséquence de la Proposition 3.7, on trouve que

$$\lambda(n) \mid [\phi(p_1^{v_1}), \phi(p_2^{v_2}), \lambda(m)] \leq \frac{\phi(p_1^{v_1})\phi(p_2^{v_2})\phi(m)}{2} = \frac{\phi(n)}{2}.$$

Par la suite, $n \notin \mathcal{N}$.

De même, on montre que si $n = 2^v p^e$, avec $v \geq 2$ et $e \geq 1$, alors

$$\lambda(n) = [\lambda(2^v), \lambda(p^e)] \mid [\lambda(2^v), (p-1)p^{e-1}] \leq \frac{\phi(2^v)(p-1)p^{e-1}}{2} = \frac{\phi(n)}{2},$$

car $2 | \lambda(2^v)$ et $2 | p - 1$ du théorème 4.4. Donc $n \notin \mathcal{N}$.

Finalement, si $n = 2^v$ avec $v \geq 3$, alors le théorème 4.4 implique que $\lambda(2^v) = 2^{v-2} < \phi(2^v)$ et donc $n \notin \mathcal{N}$. □

4.2 Résidus quadratiques

Supposons qu'on veut trouver tous les x qui sont de solutions à l'équation quadratique

$$x^2 + ax + b \equiv 0 \pmod{p}, \tag{4.3}$$

où p est un nombre premier impair. Le nombre 2 est inversible mod un nombre impair, ce qui nous permet de completer la carré :

$$\begin{aligned} x^2 + ax + b \equiv x^2 + 2(\overline{2}ax) + b \pmod{p} &\equiv (x + \overline{2}a)^2 - (\overline{2}a)^2 + b \pmod{p} \\ &\equiv (x + \overline{2}a)^2 - \overline{4}d \pmod{p}, \end{aligned}$$

où $d := a^2 - 4b$ est le discriminant du polynôme $x^2 + ax + b$. Donc on arrive à l'équation

$$(x + 2^{-1}a)^2 \equiv 4^{-1}d \pmod{p} \quad \Leftrightarrow \quad (2x + a)^2 \equiv d \pmod{p}.$$

Par la suite,

$$\#\{x \,(\mathrm{mod}\, p) : x^2 + ax + b \equiv 0 \,(\mathrm{mod}\, p)\} = \begin{cases} 0 & \text{si } d \not\equiv r^2 \,(\mathrm{mod}\, p) \text{ pour tout } r \in \mathbb{Z}, \\ 1 & \text{si } d \equiv 0 \,(\mathrm{mod}\, p), \\ 2 & \text{si } p \nmid d \text{ et } d \equiv r^2 \,(\mathrm{mod}\, p) \text{ pour un } r \in \mathbb{Z}. \end{cases}$$

Dans le deuxième cas, la seule solution à (4.3) est $x \equiv \overline{2}a \,(\mathrm{mod}\, p)$, et dans le troisième cas les deux solutions sont $\overline{2} \cdot (-a \pm r) \pmod{p}$, où r est une racine quadratique de d modulo p. On arrive alors au résultat analogue du théorème familier concernant la résolution $x^2+ax+b = 0$ avec $x \in \mathbb{R}$.

Ceci nous amène à la définition suivante :

Définition 4.6. Pour chaque $a \in \mathbb{Z}$ et chaque nombre premier p, on définit le symbole de Legendre

$$(a|p) = \left(\frac{a}{p}\right) := \begin{cases} 0 & \text{si } p|a, \\ 1 & \text{si } a \equiv x^2 \text{ pour un } x \not\equiv 0 \,(\mathrm{mod}\, p), \\ -1 & \text{sinon.} \end{cases}$$

Si $(a|p) = 1$, alors on dit que a est un résidu quadratique modulo p. Si $(a|p) = -1$, alors on dit que a est un non-résidu quadratique modulo p.

Avec cette définition, on a que

$$\#\{x \,(\mathrm{mod}\, p) : x^2 + ax + b \equiv 0 \,(\mathrm{mod}\, p)\} = 1 + \left(\frac{a^2 - 4b}{p}\right).$$

Notre but dans cette section est de comprendre quand un nombre donné a est un résidu quadratique ou un non-résidu quadratique. On commence avec le lemme de base suivant.

Lemme 4.7. *Soit p un nombre premier.*

(a) *Si $p > 2$, alors les résidus quadratiques modulo p occupent les $(p-1)/2$ distinctes classes d'équivalence $1^2 \pmod p$, $2^2 \pmod p$, ..., $((p-1)/2)^2 \pmod p$. Par consequent, il existe $(p-1)/2$ résidus quadratiques modulo p et $(p-1)/2$ non-résidus quadratiques modulo p.*

(b) *Le multiple de deux résidus quadratiques mod p est un résidu quadratique mod p ; le multiple d'un résidu quadratique mod p et un non-résidu quadratique mod p est un non-résidu quadratique mod p ; le multiple de deux non-résidus quadratiques mod p et un résidu quadratique mod p.*

Démonstration. Soient $G = (\mathbb{Z}/p\mathbb{Z})^*$ et $H = \{x^2 \pmod p : p \nmid x\}$. C'est clair que $H < G$. On affirme que $|H| = (p-1)/2 =: p_0$. En effet, on a que $\mathbb{Z}/p\mathbb{Z} = \mathbb{Z} \cap [-p_0, p_0]$ d'après le corollaire 2.6. Donc $G = \{\pm j \pmod p : 1 \leq j \leq p_0\}$ et $H = \{j^2 \pmod p : 1 \leq j \leq p_0\}$. On affirme que les carrés j^2 sont distincts mod p quand $j \in [1, p_0]$. En effet, si $j^2 \equiv i^2 \pmod p$, alors $p|(j^2 - i^2) = (j-i)(j+i)$. Puisque $2 \leq j+i \leq 2p_0 < p$, on a que $p \nmid j+i$. Par la suite, $p|j-i$ et, puisque $1 \leq i, j \leq p_0 < p$, on trouve que $i = j$. Ceci montre que la partie (a) du lemme. En particulier, il montre que $|H| = p_0$.

Maintenant, on passe à la partie (b). La seule affirmation qui n'est pas directe est que le produit de deux résidus non-quadratiques est un résidu quadratique. Puisque $|G| = |G/H| \cdot |H|$, le groupe quotient G/H a cardinalité 2, donc $G = H \cup nH$, où $n \notin H$. Mais $n^2 \in H$ par définition, donc $n^2 H = H$. La classe d'équivalence nH est l'ensemble de tous les résidus non-quadratiques. Si $n_1, n_2 \in nH$, alors $n_1 H = n_2 H = nH$, donc

$$n_1 n_2 H = (n_1 H) \cdot (n_2 H) = (nH) \cdot (nH) = n^2 H = H,$$

ce qui implique que $n_1 n_2 \in H$. Ceci conclut la démonstration. □

Lemme 4.8 (Critère d'Euler)**.** *Si $p > 2$ est un nombre premier et $a \in \mathbb{Z}$, alors*

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod p.$$

Démonstration. Le lemme est trivial quand $p|a$, donc supposons que $p \nmid a$. Soit g une racine primitive mod p, et soit $h = g^{(p-1)/2}$. On a que $h^2 \equiv 1 \pmod p$, mais que $h \not\equiv 1 \pmod p$. Donc $h \equiv -1 \pmod p$. On écrit, maintenant, a en termes de g : on a que $a \equiv g^k \pmod p$ pour un $k \in \mathbb{Z}$. Les résidus quadratiques sont exactement les puissances paires de g. Alors, si $k = 2\ell + r$ avec $r \in \{0, 1\}$, on a que $(a|p) = (-1)^r$. D'autre côté, on a que

$$a^{(p-1)/2} = (g^{p-1})^\ell \cdot h^r \equiv (-1)^r \pmod p,$$

ce qui termine la démonstration. □

Corollaire 4.9. *Soit p un nombre premier. Pour tous $a, b \in \mathbb{Z}$, on a que*

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

En particulier, la fonction $n \to \left(\frac{n}{p}\right)$ est complètement multiplicative.

Puisque le symbole de Legendre est une fonction complètement multiplicative, le théorème 4.9 implique que si $n = \pm q_1^{v_1} q_2^{v_2} \cdots q_r^{v_r} \in \mathbb{Z}$, où $q_1, \ldots, q_r$ sont nombres premiers positifs distincts, alors

$$\left(\frac{n}{p}\right) = \left(\frac{\pm 1}{p}\right)\left(\frac{q_1}{p}\right)^{v_1}\left(\frac{q_2}{p}\right)^{v_2}\cdots\left(\frac{q_r}{p}\right)^{v_r}.$$

Donc, afin de calculer $(n|p)$, il suffit de savoir la valuer de $(-1|p)$ et de $(q|p)$ pour chaque nombre premier q. On calcul d'abord $(-1|p)$ et $(2|p)$. La valuer de $(q|p)$, où q est un nombre premier impair, sera discutée après.

Théorème 4.10. *Si $p > 2$ un nombre premier, alors on a que*

$$\left(\frac{-1}{p}\right) = \begin{cases} 1 & \text{si } p \equiv 1 \pmod 4, \\ -1 & \text{si } p \equiv 3 \pmod 4. \end{cases}$$

Démonstration. Le résultat est un corollaire direct du critère d'Euler. □

Théorème 4.11. *Si $p > 2$ un nombre premier, alors on a que*

$$\left(\frac{2}{p}\right) = \begin{cases} 1 & \text{si } p \equiv \pm 1 \pmod 8, \\ -1 & \text{si } p \equiv \pm 3 \pmod 8, \end{cases}$$

Démonstration. On utilise le critère d'Euler. Soit $p_0 = (p-1)/2$. On observe que

$$\prod_{\substack{1 \le m \le p-1 \\ 2 \mid m}} m = \prod_{j=1}^{p_0} (2j) = 2^{p_0} p_0!. \tag{4.4}$$

Puis, on réécrit le produit au côté droit. Soit $j_0 = \lfloor p_0/2 \rfloor$, pour que $2j \le p-1$ si et seulement si $j \le j_0$. Les nombres pairs $2j \in (p_0, p-1]$ sont en correspondence avec les nombres impairs $2i - 1 \in [1, p_0]$. En effet, si $2j \in (p_0, p-1]$, alors $p - 2j$ est un nombre impair qui appartient à $[1, p_0]$, et vice-versa. Puisque il existe $p_0 - j_0$ nombres pairs dans $(p_0, p-1]$, on trouve que

$$\prod_{\substack{p_0 < m \le p-1 \\ 2 \mid m}} m \equiv (-1)^{p_0 - j_0} \prod_{\substack{p_0 < m \le p-1 \\ 2 \mid m}} (p - m) \equiv \prod_{\substack{m \le p_0 \\ 2 \nmid m}} m \pmod p.$$

Donc

$$\begin{aligned} \prod_{\substack{1 \le m \le p-1 \\ 2 \mid m}} m &\equiv \prod_{\substack{m \le p_0 \\ 2 \mid m}} m \prod_{\substack{p_0 < m \le p-1 \\ 2 \mid m}} m \pmod p \\ &\equiv (-1)^{p_0 - j_0} \prod_{\substack{m \le p_0 \\ 2 \mid m}} m \prod_{\substack{m \le p_0 \\ 2 \nmid m}} m \pmod p \\ &\equiv (-1)^{p_0 - j_0} p_0! \pmod p. \end{aligned} \tag{4.5}$$

En comparant les relations (4.4) et (4.5), et puisque $(p_0!, p) = 1$, on trouve que $2^{p_0} \equiv (-1)^{p_0 - j_0} \pmod p$. Par conséquent, le critère d'Euler nous donne que $(2|p) \equiv (-1)^{p_0 - j_0} \pmod p$.

Puisque les nombres $(2|p)$ et $(-1)^{p_0-j_0}$ prennent les valeurs ± 1, la dernière congruence est, en fait, une égalité. Le résultat affirmé suit d'une examination du signe de $(-1)^{p_0-j_0}$ selon la classe d'équivalence du p modulo 8. □

Le résultat fondamental pour calculer $(q|p)$ quand q est un premier impair est le théorème suivant.

Théorème 4.12 (la loi de réciprocité quadratique)**.** *Si p et q sont deux nombres premiers impairs et districts, alors*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}.$$

Avant de montrer ce théorème, on verra comment il peut nous aider de calculer le symbole de Legendre. Comme nous l'avons vu, il suffit de calculer $(a|p)$ quand $a = -1$ ou a est un nombre premier. On a déjà calculé $(a|p)$ quand $a \in \{-1, 2\}$. Supposons, maintenant, qu'on veut calculer $(3|p)$. Si $p = 2$, alors $(3|p) = 1$, et si $p = 3$, alors $(3|p) = 0$. Supposons maintenant que $p > 3$. La loi de réciprocité quadratique donc implique que

$$\left(\frac{3}{p}\right) = (-1)^{\frac{p-1}{2}\frac{3-1}{2}}\left(\frac{p}{3}\right) = (-1)^{\frac{p-1}{2}}\left(\frac{p}{3}\right).$$

On a que

$$(-1)^{\frac{p-1}{2}} = \begin{cases} 1 & \text{si } p \equiv 1 \pmod 4, \\ -1 & \text{si } p \equiv 3 \pmod 4, \end{cases} \quad \text{et} \quad \left(\frac{p}{3}\right) = \begin{cases} 1 & \text{si } p \equiv 1 \pmod 3, \\ -1 & \text{si } p \equiv 2 \pmod 3. \end{cases}$$

On est arrivé donc au résultat suivant :

Théorème 4.13. *Si $p > 3$ un nombre premier, alors on a que*

$$\left(\frac{3}{p}\right) = \begin{cases} 1 & \textit{si } p \equiv \pm 1 \pmod{12}, \\ -1 & \textit{si } p \equiv \pm 5 \pmod{12}. \end{cases}$$

Il y a plusieurs démonstrations de la loi de réciprocité quadratique. On va présenter deux preuves ici.

La première est grâce à Gauss. On peut voir l'idée principale plus facilement qu'en autres démonstrations, mais on paie le prix d'avoir un argument plus long. L'étape-clé de la démonstration est le lemme suivant. L'argument s'appuie sur une généralisation de la preuve du théorème 4.11. Rappelez que dans cette dernière preuve, on a montré que $(2|p) = (-1)^n$, où $n = \#\{j \in \mathbb{N} : (p-1)/2 < 2j \leq p-1\}$. On généralise maintenant ce résultat.

Lemme 4.14. *Soient $p > 2$ un nombre premier et $a \in \mathbb{Z}$ tels que $(a, p) = 1$. Dénotons par $N_p(a)$ le nombre des entiers $j \in [1, (p-1)/2]$ pour lesquels $aj \pmod p$ appartient à $\{x \pmod p : (p-1)/2 < x \leq p-1\}$.*

(a) (lemme de Gauss) On a que

$$\left(\frac{a}{p}\right) = (-1)^{N_p(a)}.$$

(b) Si a est impair, alors on a que

$$N_p(a) \equiv \sum_{j=1}^{\frac{p-1}{2}} \left\lfloor \frac{aj}{p} \right\rfloor \pmod 2.$$

Démonstration. (a) Soit $p_0 = (p-1)/2$. Pour chaque $j \in \{1, \dots, p_0\}$, on considère $r_j \in \mathbb{Z} \cap (-p_0, p_0]$ tel que $aj \equiv r_j \pmod p$. On observe que $r_j < 0$ si et seulement si le reste de aj quand on le divise par p se trouve dans $(p_0, p-1]$. Donc

$$N_p(a) = \#\{1 \le j \le p_0 : r_j < 0\}.$$

On affirme que les nombres $|r_1|, |r_2|, \dots, |r_{p_0}|$ sont distincts modulo p. En effet, si $|r_i| \equiv |r_j| \pmod p$ pour quelques $i, j \in \{1, \dots, p_0\}$, alors $r_i \equiv \varepsilon r_j \pmod p$ pour un $\varepsilon \in \{-1, 1\}$. Donc $ai \equiv \varepsilon aj \pmod p$. Puisque $(a, p) = 1$, on trouve que $i \equiv \varepsilon j \pmod p$, c'est-à-dire $p|(i - \varepsilon j)$. Cependant, $|i - \varepsilon j| \le i + j \le 2p_0 = p - 1$, ce qui implique $i - \varepsilon j = 0$. Puisque $1 \le i, j \le p_0$ et $\varepsilon \in \{-1, 1\}$, il faut que $\varepsilon = 1$ et $i = j$. Ceci montre notre affirmation que les nombres $|r_1|, |r_2|, \dots, |r_{p_0}|$ sont distincts modulo p. Mais ces sont p_0 nombres qui appartiennent tous à $\{1, 2, \dots, p_0\}$. Par conséquent,

$$\{|r_1|, \dots, |r_{p_0}|\} = \{1, \dots, p_0\}. \tag{4.6}$$

Alors, on trouve que

$$\begin{aligned} p_0! = \prod_{j=1}^{p_0} |r_j| = (-1)^{N_p(a)} \prod_{j=1}^{p_0} r_j &\equiv (-1)^{N_p(a)} \prod_{j=1}^{p_0} (aj) \pmod p \\ &\equiv (-1)^{N_p(a)} a^{\frac{p-1}{2}} \prod_{j=1}^{p_0} j \pmod p \\ &\equiv (-1)^{N_p(a)} \left(\frac{a}{p}\right) p_0! \pmod p, \end{aligned}$$

d'après le critère d'Euler. Puisque $(p_0!, p) = 1$, on trouve que $(a|p) \equiv (-1)^{N_p(a)} \pmod p$. Puisque les nombres $(a|p)$ et $(-1)^{N_p(a)}$ prennent les valeurs ± 1, la dernière congruence est, en fait, une égalité. Ceci conclut la démonstration de la première partie du lemme.

(b) Soit s_j le reste de aj quand divisé par p. On observe que

$$s_j = \begin{cases} r_j & \text{si } 0 \le r_j \le p_0, \\ p + r_j & \text{si } -p_0 < r_j < 0. \end{cases}$$

Donc

$$\sum_{j=1}^{p_0} s_j = \sum_{j=1}^{p_0} r_j + N_p(a)p \equiv N_p(a) + \sum_{j=1}^{p_0} r_j \pmod 2,$$

parce que p est impair. De plus, on observe que $x \equiv |x| \pmod 2$, pour tout $x \in \mathbb{Z}$. Donc

$$\sum_{j=1}^{p_0} r_j \equiv \sum_{j=1}^{p_0} |r_j| \equiv \sum_{j=1}^{p_0} j \pmod 2$$

où on a utilisé la relation (4.6). Par conséquent,

$$N_p(a) \equiv \sum_{j=1}^{p_0} s_j - \sum_{j=1}^{p_0} j \pmod 2.$$

Finalement, puisque $aj = kp + s_j$, pour un $k \in \mathbb{Z}$ et $0 \leq s_j < p$, on a que $\lfloor aj/p \rfloor = \lfloor k + s_j/p \rfloor = k$. Donc $s_j = aj - p\lfloor aj/p \rfloor$, ce qui implique que

$$\sum_{j=1}^{p_0} s_j = a\sum_{j=1}^{p_0} j - \sum_{j=1}^{p_0} \left\lfloor \frac{aj}{p} \right\rfloor \equiv \sum_{j=1}^{p_0} j + \sum_{j=1}^{p_0} \left\lfloor \frac{aj}{p} \right\rfloor \pmod 2,$$

d'après notre hypothèse que a est impair. Ceci conclut la démonstration du lemme. □

Démonstration du théorème 4.12. On pose $p_0 = (p-1)/2$ et $q_0 = (q-1)/2$. On examinera les sommes

$$\sum_{j=1}^{p_0} \left\lfloor \frac{qj}{p} \right\rfloor \quad \text{et} \quad \sum_{k=1}^{q_0} \left\lfloor \frac{pk}{q} \right\rfloor.$$

On observe que

$$\begin{aligned}\sum_{j=1}^{p_0} \left\lfloor \frac{qj}{p} \right\rfloor &= \sum_{j=1}^{p_0} \sum_{1 \leq k \leq qj/p} 1 = \#\{(j,k) \in \mathbb{N}^2 : j \leq p_0,\ k \leq qj/p\} \\ &= \#\{(j,k) \in \mathbb{N}^2 : j \leq p_0,\ k \leq q_0,\ pk \leq qj\}.\end{aligned}$$

De même,

$$\begin{aligned}\sum_{k=1}^{q_0} \left\lfloor \frac{pk}{q} \right\rfloor &= \sum_{k=1}^{q_0} \sum_{1 \leq j \leq pk/q} 1 = \#\{(j,k) \in \mathbb{N}^2 : k \leq q_0,\ j \leq pk/q\} \\ &= \#\{(j,k) \in \mathbb{N}^2 : j \leq p_0,\ k \leq q_0,\ pk \geq qj\}.\end{aligned}$$

Puisque il n'y a pas de pair $(j,k) \in \mathbb{N}^2$ avec $1 \leq j \leq p_0$, $1 \leq k \leq q_0$ et $pk = qj$ (sinon, on trouve que $p|j$ et donc $j \geq p$, une contradiction), alors

$$\sum_{j=1}^{p_0} \left\lfloor \frac{qj}{p} \right\rfloor + \sum_{k=1}^{q_0} \left\lfloor \frac{pk}{q} \right\rfloor = \#\{(j,k) \in \mathbb{N}^2 : j \leq p_0,\ k \leq q_0\} = p_0 q_0.$$

En combinant cette relation avec le lemme 4.14, on en déduit que $N_p(q) + N_q(p) \equiv p_0 q_0 \pmod 2$ et que

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{p_0 q_0},$$

ce qui est ce qu'il fallait montrer. □

Démonstration alternative du théorème 4.12. On présente une preuve différente de la loi de réciprocité quadratique grâce à George Rousseau (*On the quadratic reciprocity law.* J. Austral. Math. Soc. Ser. A 51 (1991), no. 3, 423–425), présentée aussi ici `https://mathoverflow.net/questions/1420/whats-the-best-proof-of-quadratic-reciprocity`.

On considère le groupe $G = (\mathbb{Z}/p\mathbb{Z})^* \times (\mathbb{Z}/q\mathbb{Z})^*$, et trois ensembles $A, B, C \subset G$ qui contient un moitié de G, i.e. pour tout $g \in G$, ils contient soit g ou $-g$. En identifiant G avec $(\mathbb{Z}/pq\mathbb{Z})^*$ (ils sont de groupes isomorphes), on choisit

$$A = \{i \,(\mathrm{mod}\, p) : 1 \le i \le (p-1)/2\} \times \mathbb{Z}/q\mathbb{Z}^*,$$

$$B = (\mathbb{Z}/p\mathbb{Z})^* \times \{j \,(\mathrm{mod}\, q) : 1 \le j \le (q-1)/2\},$$

$$C = \{k \,(\mathrm{mod}\, qp) : 1 \le k < pq/2,\ (k, pq) = 1\}.$$

Par construction, les produits $\prod_{a\in A} a$, $\prod_{b\in B} b$ et $\prod_{c\in C} c$ différent par quelques signes.

Si $p_0 = (p-1)/2$ et $q_0 = (q-1)/2$, alors

$$\prod_{a\in A} a = \prod_{1\le i\le p_0} \prod_{1\le j\le q-1} (i, j) = \prod_{1\le i\le p_0} (i^{q-1}, (q-1)!) = (p_0!^{q-1}, (q-1)!^{p_0})$$

et, de même,

$$\prod_{b\in B} b = ((p-1)!^{q-1}, q_0!^{p-1}).$$

Finalement, on a que □

4.3 Caractères

Le corollaire 4.9 implique que le symbole de Legendre définit un morphisme du groupe $(\mathbb{Z}/p\mathbb{Z})^*$ au groupe multiplicatif $\mathbb{C}^* := \mathbb{C} \setminus \{0\}$. En général, un morphisme

$$\chi : (\mathbb{Z}/q\mathbb{Z})^* \to \mathbb{C}^*$$

est appelé un caractère de Dirichlet mod q. Le nombre q est parfois appelé le modulus du caractère. On peut prolonger χ sur $\mathbb{Z}$ en mettant

$$\chi(n) := \begin{cases} \chi(n \,(\mathrm{mod}\, q)) & \text{si } (n, q) = 1, \\ 0 & \text{sinon.} \end{cases}$$

C'est facile de vérifier que χ est une fonction q-périodique et complètement multiplicative dans le sens que

$$\chi(mn) = \chi(m)\chi(n) \quad \forall m, n \in \mathbb{Z}.$$

Les deux points de vue sont équivalents. Donc, parfois, on va parler de caractères de Dirichlet comme étant de fonctions définies sur $\mathbb{Z}$.

Le concept des caractères de Dirichlet se généralise sur n'import quel groupe. Ici on se concentre au cas des **groupes abéliens finis** qui est plus simple. Soit un $(G, *)$ un tel groupe. Un morphisme $\chi : G \to \mathbb{C}^*$ est appelée un caractère de G. L'ensemble de tous les caractères de G est dénoté par $\hat{G}$.

Remarque 4.15. Si $n = |G|$ et χ est un caractère de G, alors $\chi(g)$ est une n-ième racine de l'unité. En effet, comme on l'a vu à la démonstration du lemme 3.15, on a que $\chi(e) = 1$. De plus, le théorème 3.4 implique que $g^n = e$. Donc, $1 = \chi(g^n) = \chi(g)^n$, ce qui montre notre affirmation.

Exemple 4.16. Soit le groupe additif $G = \mathbb{Z}/q\mathbb{Z}$. C'est un groupe cyclique engendré par $1 \pmod q$. De plus, $\chi(n \pmod q) = \chi(1 \pmod q)^n$ car χ est un morphisme. Alors la valuer de $\chi(1 \pmod q)$, soit ζ, détermine le caractère χ. Selon le remarque 4.15, ζ est une q-ième racine de l'unité, c'est-à-dire $\zeta = e^{2\pi ia/q}$ pour un $a \in \{0, 1, \ldots, q\}$.

Vice versa, si $\zeta = e^{2\pi ia/q}$ est une q-ième racine de l'unité, on peut définir le caractère $\chi(n \pmod q) := e^{2\pi ian/q}$. C'est bien défini, ne dépendant pas du choix de représentant de la classe d'équivalence $n \pmod q$.

On voit donc que $\hat{G}$ est en correspondence avec les q-ièmes racines de l'unité. En particulier, $|\hat{G}| = q$.

Remarque 4.17. On peut munir l'ensemble $\hat{G}$ d'une opération de multiplication comme suivant. Si $\chi, \psi : G \to \mathbb{C}^*$ sont deux caractères, alors c'est immédiat que leur produit $\chi\psi : G \to \mathbb{C}^*$, défini simplement par $(\chi\psi)(g) := \chi(g) \cdot \psi(g)$, est aussi un caractère de G. Le pair $(\hat{G}, \cdot)$ est un groupe abélien. L'inverse de χ est simplement sont réciproque : $\chi^{-1}(g) := 1/\chi(g)$. On note que, puisque $\chi(g)$ est une racine de l'unité, alors $1/\chi(g) = \overline{\chi}(g)$, son conjugué comme nombre complexe. L'élément neutre de $\hat{G}$ est le caractère trivial $1 : G \to \mathbb{C}^*$, défini par $1(g) = 1$. Il est appelé le caractère principal de G.

Théorème 4.18. *Si $(G, *)$ est un groupe abélien fini, alors $|\hat{G}| = |G|$.*

Démonstration. Ce résultat se base sur un théorème profond de la théorie des groupe qu'on donne sans démonstration : il existe nombres $q_1, \ldots, q_k$ tels que

$$G \cong \mathbb{Z}/q_1\mathbb{Z} \times \cdots \times \mathbb{Z}/q_k\mathbb{Z}.$$

En mots, tous les groupes abéliens finis sont de produits de groupes cycliques ! Il suffit alors de monter le résultat quand $G = \mathbb{Z}/q_1\mathbb{Z} \times \cdots \times \mathbb{Z}/q_k\mathbb{Z}$. Dans ce cas, $|G| = q_1 \cdots q_k$.

Pour chaque $j \in \{1, \ldots, k\}$, on définit $\chi_j : \mathbb{Z}/q_j\mathbb{Z} \to \mathbb{C}^*$ en posant $\chi_j(n_j \pmod{q_j})$ d'être la valuer de χ quand evalué au vecteur dont tous les cordonnés valent 0, sauf le j-ième cordonné qui vaut $n_j \pmod{q_j}$. C'est facile de vérifier que χ_j est un caractère du groupe $\mathbb{Z}/q_j\mathbb{Z}$. En particulier, selon l'exemple 4.16, il existe $a_j \in \{0, 1, \ldots, q_j - 1\}$ tel que $\chi_j(n_j \pmod{q_j}) = e^{2\pi ia_jn_j/q_j}$. On trouve alors que

$$\chi(n_1 \pmod{q_1}, \ldots, n_k \pmod{q_k}) = \prod_{j=1}^{k} \chi_j(n_j \pmod{q_j}) = e^{2\pi i(a_1n_1/q_1 + \cdots + a_kn_k/q_k)}. \tag{4.7}$$

Vice versa, une fonction définie par (4.7) est un caractère de G. On a montré alors que $|\hat{G}| = q_1 \cdots q_k = |G|$, comme affirmé. □

L'importance des caractères se trouve, entre autres, à leur orthogonalité. Plus précisément, soit

$$L^2(G) := \{\alpha : G \to \mathbb{C}\},$$

l'espace des fonctions complexes définies sur G. Evidemment, c'est un espace linéaire sur $\mathbb{C}$ et $\hat{G} \subset L^2(G)$. Sa dimension sur $\mathbb{C}$ est égale à $|G|$ (chaque fonction $\alpha : G \to \mathbb{C}$ est définie

par $|G|$ valeurs ; les fonctions $\varepsilon_h : G \to \mathbb{C}$, définies pour chaque $h \in G$ par $\varepsilon_g(g) := \mathbf{1}_{g=h}$ sont une base de $L^2(G)$). De plus, on peut munir $L^2(G)$ d'un produit interne défini par

$$\langle \alpha, \beta \rangle_G := \frac{1}{|G|} \sum_{g \in G} \alpha(g)\overline{\beta}(g).$$

Le théorème 4.19 plus bas montre que $\hat{G}$ est un ensemble orthonormal de $L^2(G)$. En particulier, c'est un ensemble linéairement indépendant. Puisque $|\hat{G}| = |G| = \dim_{\mathbb{C}}(L^2(G))$, alors $\hat{G}$ est une **base orthonormale** de $L^2(G)$, c'est-à-dire chaque fonction $f : G \to \mathbb{C}$ peut s'écrire comme

$$f = \sum_{\chi \in \hat{G}} \langle f, \chi \rangle_G \cdot \chi. \tag{4.8}$$

On définit $\hat{f} : \hat{G} \to \mathbb{C}$, la transformation de Fourier de f, par

$$\hat{f}(\chi) := \langle f, \chi \rangle_G.$$

Alors la formule (4.8) devient

$$f = \sum_{\chi \in \hat{G}} \hat{f}(\chi)\chi, \tag{4.9}$$

le développement de Fourier de f.[1]

Théorème 4.19. *Si $\chi, \psi \in \hat{G}$, alors*

$$\langle \chi, \psi \rangle_G = \mathbf{1}_{\chi=\psi}.$$

Démonstration. Soit le caractère $\xi = \chi\overline{\psi}$, pour que $\langle \chi, \psi \rangle_G = |G|^{-1} \sum_{g \in G} \xi(g)$. Si $\chi = \psi$, alors $\xi = 1$ et le théorème suit directement. Si $\chi \neq \psi$, alors il existe $h \in G$ avec $\xi(h) \neq 1$. Puisque $G * h = G$, on trouve que

$$\sum_{g \in G} \xi(g) = \sum_{g \in G} \xi(g * h) = \sum_{g \in G} \xi(g)\xi(h) = \xi(h) \sum_{g \in G} \xi(g).$$

Mais $\xi(h) \neq 1$, donc $\sum_{g \in G} \xi(g) = 0$. Ceci conclut la démonstration. □

Exemple 4.20. Si $G = (\mathbb{Z}/p\mathbb{Z})^*$, $\chi(a) = (a|p)$ est le symbole de Legendre, et $\psi(a) = 1$ est le caractère principale, alors le théorème 4.19 implique que

$$\sum_{a=1}^{p-1} \left(\frac{a}{p}\right) = 0.$$

Ceci est une autre démonstration du fait qu'il existe aussi beaucoup résidus quadratiques que non-résidus quadratiques.

1. Comparez cette formule avec le développement en séries de Fourier d'une 'bonne' fonction 1-périodique $f : \mathbb{R} \to \mathbb{C}$, donnée par $f(x) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x}$ avec $a_n = \int_0^1 f(x)e^{-2\pi i n x} \mathrm{d}x$. Le groupe ici est le tore $G = \mathbb{R}/\mathbb{Z}$ (car on travaille avec des fonctions 1-périodiques), et son groupe de caractères est infini car G l'est : ils sont les fonctions $x \to e^{2\pi i n x}$ avec $n \in \mathbb{Z}$. Finalement, le produit interne ici est donné par l'intégrale $\langle \alpha, \beta \rangle_G := \int_0^1 \alpha(x)\overline{\beta}(x)\mathrm{d}x$.

Remarque 4.21. En appliquant la formule (4.9) avec la fonction $f(g) = \mathbf{1}_{g=e}$, pour laquelle $\hat{f}(\chi) = \chi(e)/|G| = 1/|G|$, on trouve que

$$\mathbf{1}_{g=e} = \frac{1}{|G|} \sum_{g \in G} \chi(g). \tag{4.10}$$

En remplacant g par $g * h^{-1}$, où h est un autre élément de G, on trouve que

$$\mathbf{1}_{g=h} = \frac{1}{|G|} \sum_{g \in G} \chi(g)\overline{\chi}(h). \tag{4.11}$$

Si $G = (\mathbb{Z}/q\mathbb{Z})^*$, $(a, q) = 1$ et $n \in \mathbb{Z}$, alors cette relation devient

$$\mathbf{1}_{n \equiv a \,(\mathrm{mod}\, q)} = \frac{1}{\phi(q)} \sum_{\chi \,(\mathrm{mod}\, q)} \chi(n)\overline{\chi}(a),$$

où $\chi \,(\mathrm{mod}\, q)$ veut dire que χ est un caractère de Dirichlet de modulus q vu comme une fonction $\chi : \mathbb{Z} \to \mathbb{C}$. Cette relation joue un rôle très important dans l'étude de nombres premiers en progressions arithmétiques.

Remarque 4.22. On a la formule de Parseval

$$\sum_{\chi \in \hat{G}} |\hat{f}(\chi)|^2 = \frac{1}{|G|} \sum_{g \in G} |f(g)|^2. \tag{4.12}$$

En effet, puisque $|z|^2 = z \cdot \overline{z}$, alors on trouve que

$$\sum_{\chi \in \hat{G}} |\hat{f}(\chi)|^2 = \frac{1}{|G|^2} \sum_{\chi \in \hat{G}} \sum_{g \in G} f(g)\overline{\chi}(g) \sum_{h \in G} \overline{f}(h)\chi(h) = \frac{1}{|G|^2} \sum_{g \in G} f(g)\overline{f}(h) \sum_{\chi \in \hat{G}} \overline{\chi}(g)\chi(h),$$

et (4.12) découle de (4.11).

Exercices

EXERCICE 4.1. Trouvez toutes les solutions aux équations suivantes :

(a) $x^2 - 3x + 3 \equiv 0 \pmod{343}$.
(b) $x^2 + 7x - 9 \equiv 0 \pmod{63}$.
(c) $x^{12} + 2x^{11} - 3x + 9 \equiv 0 \pmod{121}$. [*Indice :* Divisez le polynôme par $x^{11} - x$ (cf. petit théorème de Fermat).]
(d) $x^3 - 30x^2 + 5x - 2 \equiv 0 \pmod{169}$.
(e) $x^2 - 5x - 4 \equiv 0 \pmod{169}$.

EXERCICE 4.2. Soient p un nombre premier et d un diviseur de $p - 1$. Montrez que le polynôme $x^d - 1$ a exactement d racines modulo p. [*Indice :* Montrez que $x^{p-1} - 1 = (x^d - 1)f(x)$ pour un polynôme f de degré $p - 1 - d$ et utilisez le théorème 2.29.]

EXERCICE 4.3. Soient p un nombre premier,

$$K_p = \#\{a \in \{0, 1, \ldots, p-1\} : a^3 \equiv 1 \pmod{p}\}$$

et

$$L_p = \#\{(x, y) \in \{0, 1, \ldots, p-1\}^2 : x^3 \equiv y^3 \pmod{p}\}.$$

(a) Trouvez une formule générale pour K_p.

(b) Montrez que $L_p = 1 + (p-1)K_p$.

EXERCICE 4.4. Cette exercice donne un troisième argument pour l'existence de racines primitives modulo p. Posons $\psi(d) = \{1 \leq a \leq p-1 : \mathrm{ord}_p(a) = d\}$.

(a) Montrez que $\sum_{d|n} \psi(d) = n$, pour chaque $n|(p-1)$. [*Indice* : exercice 4.2.]

(b) Montrez que $\psi(d) = \phi(d)$, pour chaque $d|(p-1)$, par induction sur d. [*Indice* : théorème 3.3]

(c) Déduisez l'existence de racines primitives modulo p et le théorème **??** dans le cas où n est premier.

EXERCICE 4.5. Soit p un nombre premier et $v \geq 1$. Montrez que si g est une racine primitive modulo p^{v+1}, alors il est une racine primitive modulo p^v.

EXERCICE 4.6. (a) Trouvez un nombre g qui est une racine primitive mod 5^v, pour chaque $v \geq 1$. Faites la même chose modulo 7^v et modulo 11^v.

(b) Trouvez g tel que $\mathrm{ord}_{2^v}(g) = \lambda(2^v) = 2^{v-2}$, pour chaque $v \geq 3$.

EXERCICE 4.7. Soit $p > 3$ un nombre premier. Calculer $(3|p)$ en utilisant l'idée en arrière de la démonstration du théorème 4.11. [*Indice* : Considérer le produit $P = \prod_{j \leq p_0}(3j)$, où $p_0 = (p-1)/2$.]

EXERCICE 4.8 $(*)$. Soit $p > 2$ un nombre premier.

(a) Si $a \in \mathbb{Z}$ est non divisible par p, alors montrez que

$$\sum_{x=1}^{p} \left(\frac{x^2 + ax}{p}\right) = -1.$$

(b) Si $\left(\frac{d}{p}\right) = -1$, alors montrez que

$$\sum_{x=1}^{p-1} \left(\frac{x^2 - 1}{p}\right) + \sum_{x=1}^{p-1} \left(\frac{dx^2 - 1}{p}\right) = 2\sum_{a=1}^{p-1} \left(\frac{a-1}{p}\right) = -2\left(\frac{-1}{p}\right).$$

Déduisez que

$$\sum_{x=1}^{p} \left(\frac{x^2 - d}{p}\right) = -1.$$

(c) Pour tout $a, b \in \mathbb{Z}$, montrez que

$$\sum_{x=1}^{p} \left(\frac{x^2 + ax + b}{p} \right) = \begin{cases} -1 & \text{si } p \nmid a^2 - 4b, \\ p - 1 & \text{si } p | a^2 - 4b. \end{cases}$$

EXERCICE 4.9. Résolvez l'équation $x^2 - 20x + 139 \equiv 0 \pmod{1583}$.

Chapitre 5

Équations diophantiennes

Dans ce chapitre, on étudie des équations diophantiennes : ces sont d'équations algébriques, données par polynômes, pour lesquelles on cherche de solutions entiers ou rationnels [1] Des exemples sont :

- on cherche de trouver les solutions entières de l'équation linéaire $ax + by = c$;
- on cherche de trouver de triples pythagoriciens, c'est-à-dire de nombres naturels x, y, z tels que $x^2 + y^2 = z^2$ (on peut construire un triangle rectangle pour chaque tel triplet).
- on cherche de trouver les solutions entières x, y, n de l'équation $x^2 + y^2 = n$; c'est-à-dire on cherche de déterminer quels nombres naturels n peuvent s'écrire comme la somme de deux carrés $x^2 + y^2$.

5.1 Une équation diophantienne linéaire

Soient trois nombres entiers fixés, a, b et c avec $a, b \neq 0$. Est-ce qu'il existe $x, y \in \mathbb{Z}$ tels que

$$ax + by = c? \tag{5.1}$$

Tout d'abord, on observe que si l'équation (5.1) a de solutions, nécessairement le plus grand commun diviseur de a et b, soit d, divise c. En effet, on a que $d|a$ et $d|b$ et, par la suite, $d|(ax + by) = c$. Réciproquement, on affirme que si $d = (a, b)$ divise c, alors (5.1) a de solutions. En effet, théorème 1.9 implique qu'il existe $x_0, y_0 \in \mathbb{Z}$ tels que

$$ax_0 + by_0 = d.$$

En multipliant cette équation par c/d, on trouve que le pair $(cx_0/d, cy_0/d)$ est une solution à (5.1). De plus, on peut calculer cette solution en utilisant l'algorithme euclidien.

À partir d'une solution donnée, on peut trouver toutes les solutions à (5.1). Soient x_0 et y_0 comme au-dessus et soit $(x_1, y_1) := (cx_0/d, cy_0/d)$. Observons que tous les pairs de la forme $(x_1 + tb/d, y_1 - ta/d)$, $t \in \mathbb{Z}$, sont de solutions à (5.1). Réciproquement, on montrera

1. La théorie moderne des équations diophantiennes s'occupe de l'étude d'équations algèbriques dans des ensembles plus générales que les entiers et les rationnels.

que si $(x, y) \in \mathbb{Z}^2$ est une autre solution à (5.1), alors on peut l'écrire dans la forme $(x_1 + tb/d, y_1 - ta/d)$, pour un certain $t \in \mathbb{Z}$. En effet, on a que

$$ax + by = d = ax_1 + by_1 \quad \Longrightarrow \quad a(x - x_1) = b(y - y_1).$$

On écrit $a = dk$ et $b = d\ell$, pour que $(k, \ell) = 1$. Donc

$$k(x - x_1) = \ell(y_1 - y).$$

Alors $\ell | k(x - x_1)$ et, puisque $(k, \ell) = 1$, le lemme d'Euclid implique que $\ell | x - x_1$, c'est-à-dire $x = x_1 + \ell t$ pour un $t \in \mathbb{Z}$. Par la suite, $\ell(y_1 - y) = k\ell t$, ce qui implique que $y = y_1 - \ell t$, comme affirmé.

Pour conclure, on a montré le résultat suivant :

Théorème 5.1. *Soient $a, b \in \mathbb{Z} \setminus \{0\}$ et $c \in \mathbb{Z}$. L'équation diophantienne linéaire possède de solutions si et seulement si $d := (a, b) | c$. Dans ce cas, il possède une infinité de solutions : ils sont les éléments de l'ensemble $\{(cx_0/d + tb/d, cy_0/d - ta/d) : t \in \mathbb{Z}\}$, où x_0 et y_0 sont tels que $ax_0 + by_0 = c$.*

5.2 Triplets pythagoriciens

Tout le monde connait le théorème de Pythagore : étant donné un triangle droit, le carré de son hypoténuse est égal à la somme des carrés de ses deux cotés perpendiculaires. Algébriquement, si z est le longueur de l'hypoténuse et x et y sont les longueurs des deux cotés perpendiculaires, alors

$$x^2 + y^2 = z^2. \tag{5.2}$$

Réciproquement, si les nombres x, y et z satisfassent l'équation (5.2), alors on peut construire un triangle droit dont les longueurs des cotés sont x, y et z. Une question naturelle est si il existe de triangles droits dont tous les cotés ont de longueur qui est un nombre entier. De façon équivalente, est ce qu'il y a de triplets $(x, y, z) \in \mathbb{Z}^3$ qui satisfassent l'équation (5.2). Un tel triplet est appelé un triplet pythagoricien. La réponse et qu'oui, il existe de triplets pythagoriciens. Par exemple, $(3, 4, 5)$ en est un et $(5, 12, 25)$ en est un autre. Le but de cette section est de décrire tous les triplets pythagoriciens. L'observation-clé est que l'équation (5.2) peut s'écrire comme

$$y^2 = z^2 - x^2 = (z - x)(z + x).$$

Donc, on déduira que, sous certains conditions, on peut factoriser y en deux facteurs dont les carrés sont égaux à $z - x$ et à $z + x$, respectivement. Afin de faire ceci, il faut faire quelques réductions préparatoires au problèmes.

Tout d'abord, si (x, y, z) est un triplet pythagoricien, alors (mx, my, mz) en est un aussi. De même, si d est un commun diviseur diviseur de x, y et z, alors le triplet $(x/d, y/d, z/d)$ est un triplet pythagoricien. Donc dans notre cherche pour de triplets pythagoriciens, on peut supposer sans perte de généralité que $(x, y, z) = 1$. Un tel triplet est appelé primitif. Les membres d'un triplet pythagoricien sont deux par deux copremiers. En effet, si il existait un

nombre premier p qui divisait x et y, alors p diviserait aussi $z^2 = x^2 + y^2$. Mais dans ce cas $p|z$ et, par la suite, $p|(x, y, z) = 1$, qui est impossible. Alors on déduit que $(x, y) = 1$, comme affirmé. De même façon, on peut montrer que $(x, z) = (y, z) = 1$ aussi.

On peut faire d'autres commentaires faciles : puisque $a^2 \equiv 1 \pmod 4$ pour tout nombre impair a, alors au moins un entre le x et le y doit être pair ; sinon, on aurait que $z^2 \equiv 2 \pmod 4$, ce qui est impossible. Sans perte de généralité, on suppose que $2|y$; sinon, on peut permuter x et y et les renommer. Si $2|y$, alors il faut que x et z soient impairs. Ceci implique que $(x - z, x + z) = 2$. En effet, si $d = (x - z, x + z)$, alors on a que $d|(x - z) + (x + z) = 2x$ et que $d|(x + z) - (x - z) = 2z$, c'est-à-dire $d|(2x, 2z) = 2(x, z) = 2$. Donc soit $d = 1$ soit $d = 2$. Puisque x et z sont les deux impairs, alors $2|(x - z)$ et $2|(x + z)$ et, par la suite, il faut avoir que $d = 2$, comme affirmé. En écrivant $y = 2y_1$, on trouve que

$$4y_1^2 = y^2 = z^2 - x^2 = (z - x)(z + x) \quad \Longrightarrow \quad y_1^2 = \frac{z - x}{2} \cdot \frac{z + x}{2}.$$

Evidemment, si ab est un carré et $(a, b) = 1$, alors a et b sont également de carrés. Donc il existe $u, v \in \mathbb{Z}$ tels que $\frac{z-x}{2} = v^2$ et $\frac{z+x}{2} = u^2$. En particulier, $(u, v) = 1$, $z = u^2 + v^2$ et $x = u^2 - v^2$. Aussi, puisque x est impair, il faut que $2|uv$. Finalement, on a que $y_1^2 = (uv)^2$ et, par la suite, $y_1 = \pm uv$. Sans perte de généralité, on peut supposer que $y_1 = uv$; sinon, on peut remplacer u par $-u$. Donc on conclut que un triplet pythagoricien primitif peut s'écrire comme

$$(x, y, z) = (u^2 - v^2, 2uv, u^2 + v^2), \tag{5.3}$$

où $(u, v) = 1$ et $2|uv$. Réciproquement, si (x, y, z) est comme avant, donc il est clairement primitif et, de plus,

$$x^2 + y^2 = (u^2 - v^2) + (2uv)^2 = u^4 - 2u^2v^2 + v^4 + 4u^2v^2 = (u^2 + v^2) = z^2,$$

c'est-à-dire (x, y, z) est un triplet pythagoricien primitif. Pour conclure, on a montré le résultat suivant.

Théorème 5.2. *On a que*

$$\{(x, y, z) \text{ triplet pythagoricien primitif} : 2|y\} = \{(u^2 - v^2, 2uv, u^2 + v^2) : 2|uv,\ (u, v) = 1\}.$$

5.3 Équations diophantiennes insolubles

Comme on l'a vu à la dernière section, l'équation $x^2 + y^2 = z^2$ a plusieurs solutions. Cependant, ce n'est pas toujours le cas que une équation diophantienne a de solutions entiers. Par exemple, l'équation

$$x^2 = -1 - y^2 \tag{5.4}$$

n'a pas de solutions réelles, puisque le coté gauche est toujours ≥ 0 et le coté droit est toujours ≤ -1. *A fortiori*En particulier, elle n'a pas de solutions entières.

Un autre exemple de nature différente est l'équation

$$x^2 + 3y^2 = 1570. \tag{5.5}$$

Cette équation a plusieurs solutions réelles qui forment une ellipse. Cependant, on affirme qu'elle n'a pas de solutions entières. En effet, pour tout $a \in \mathbb{Z}$, on a que $a^2 \equiv 0, 1 \pmod 4$. Donc $x^2 + 3y^2 \equiv 0, 1, 3 \pmod 4$ mais $1570 \equiv 2 \pmod 4$. On en déduit que (5.5) n'a pas de solutions entières, car chaque telle solution "globale" impliquerait une solution "locale" mod 4.

La discussion ci-dessus nous donne un critère pour le non-existence de solutions à une équation diophantienne : une équation diophantienne qui possède de solutions entières, nécessairement a de solutions réelles et de solutions modulo n, pour n'importe quel nombre n. En considérant la proposition contraposée, si on peut refuter soit l'existence de solutions réelles ou l'existence de solutions « locales »(c'est-à-dire modulo un nombre n) d'une équation, alors cette équation n'a pas de solutions entières.

Cependant, c'est possible qu'une équation diophantienne possède de solutions réelles et de solutions locales pour tout n, mais qu'elle n'a pas de solutions entières. Dans ce cas, on peut utiliser d'autres techniques pour prouver la non existence de solutions. Une telle technique a été développée par Fermat afin d'étudier sa fameuse équation $x^n + y^n = z^n$ quand $n \geq 3$. La méthode de Fermat est appelée la descente infinie et son idée est simple et élégante : on commence avec une solution hypothétique à une équation et, à partir d'elle, on construit une nouvelle solution qui est 'plus petite' (dans un certain sens - habituellement, on mesure la « magnitude » d'une solution en termes de la taille de ses coefficients).

On commence avec un exemple simple pour démontrer la méthode de la descente infinie. Considérons l'équation

$$x^3 + 3y^3 = 9z^3. \tag{5.6}$$

On montrera qu'elle n'a pas de solutions sauf la solution triviale $(0, 0, 0)$. Supposons au contraire que le triplet $(x_0, y_0, z_0) \in \mathbb{Z}^3$ est une solution. On construira un nouveau triplet $(x_1, y_1, z_1) \in \mathbb{Z}^3$ telle que

$$0 < \max\{|x_1|, |y_1|z_1|\} < \max\{|x_0|, |y_0|, |z_0|\}. \tag{5.7}$$

En itérant cette procedure, on peut construire une infinité de triplets distingues $(x, y, z) \in \mathbb{Z}^3$ dont tous les cordonnés sont $\leq \max\{|x_0|, |y_0|, |z_0|\}$ en valuer absolue. C'est clairement absurd, donc l'hypothèse initiale qu'il existe une solution $(x, y, z) \neq (0, 0, 0)$ à l'équation (5.6) est fausse.

En effet, si $x_0^3 + 3y_0^3 = 9z_0^3$, alors $3|x_0$. Si on écrit $x_0 = 3x_1$, alors $9x_1^3 + y_0^3 = 3z_0^3$. Par la suite, $3|y_0$. On écrit $y_0 = 3y_1$ pour que $3x_1^3 + 9y_1^3 = z_0^3$. Donc $3|z_0$ et, si on écrit $z_0 = 3z_1$, alors on trouve que $x_1 3 + 3y_1^3 = z_1^3$. Ceci construit la nouvelle solution promise $(x_1, y_1, z_1) = (x_0/3, y_0/3, z_0/3)$. Evidemment, elle satisfait (5.7), et la construction est complète.

Pierre de Fermat a étudié le livre de Diophantus d'Alexandre décrivant la détermination de tous les triples pythagoriciens. qui contenait la solution de l'équation $x^2 + y^2 = z^2$ avec

$x, y, z \in \mathbb{Z}$. Il s'est arrivé à la question naturelle suivante : est-ce qu'il existe de solutions si on remplaces les carrés par de puissances plus grandes ? L'équation diophantienne est donc $x^n + y^n = z^n$. On cherche de solutions entières et non-triviales, c'est-à-dire avec $xyz \neq 0$ (les triples $(t, 0, t)$ et $(0, t, t)$, $t \in \mathbb{Z}$, sont de solutions triviales). Fermat a conjecturé en 1637 qu'il n'existe pas de solutions non-triviales et il affirmé qu'il avait une "solution magnifique" mais que la marge du livre de Diophantus ne suffisait par pour l'écrire. Il a défie les autres mathématiciens de trouver la preuve de son résultat, qui est devenu connu comme le dernier théorème de Fermat. Cela a pris plus que 350 ans pour battre le défi de Fermat : en 1994, Andrew Wiles, avec la collaboration de Richard Taylor, a enfin publié la démonstration du dernier théorème de Fermat. Les méthodes que Wiles a introduit sont vraiment révolutionnaires (c'est peu probable que Fermat a montré son théorème de cette façon - en fait, c'est un débat ouvert si Fermat possédait d'une preuve correcte), mais ils sont dehors les buts de ces notes. Ici on présent la démonstration du cas $n = 4$ du dernier théorème de Fermat qui est élémentaire. En fait, on montre un résultat plus fort :

Théorème 5.3. *Soient $x, y, z \in \mathbb{Z}$. Si $x^4 + y^4 = z^2$, alors $xyz = 0$.*

Démonstration. On utilise la méthode infinie de Fermat : supposons que

$$A := \{(x, y, z) \in \mathbb{Z}^3 : xyz \neq 0,\ x^4 + y^4 = z^2\} \neq \emptyset.$$

On peut choisir $(x, y, z) \in A$ tels que $|z|$ est minimal, c'est-à-dire

$$|z| = \min\{|z'| : (x', y', z') \in A\}.$$

On construira un triplet $(x', y', z') \in A$ tel que $|z'| < |z|$, ce qui est clairement une contradiction.

Tout d'abord, puisque $(-t)^2 = t^2$, sans perte de généralité, on peut supposer que $x, y, z \in \mathbb{N}$; sinon, on peut les remplacer par $-x$ par $-y$ ou par $-z$, respectivement. On observe que (x^2, y^2, z) est un triplet pythagoricien, qui est primitif. Afin de voir que il est primitif, il suffit de montrer que $d := (x, y, z)$ est égal à 1. En effet, on a que $d^4 | x^4 + y^4 = z^2$ et donc $d^2 | z$. Ceci implique que $(x/d, y/d, z/d^2) \in A$ et, par la suite, $|z/d^2| \geq |z|$. Donc $d = 1$, comme affirmé.

Puisque (x^2, y^2, z) est un triplet pythagoricien primitive, alors soit $2|x^2$ soit $2|y^2$, selon la discussion de la section 5.2. Sans perte de généralité, on peut supposer que $2|y^2$ (sinon, on permute x et y). Par conséquent, x et z sont impairs et y est pair. Alors, le théorème 5.2 implique que

$$x^2 = u^2 - v^2, \quad y^2 = 2uv, \quad z = u^2 + v^2,$$

pour quelques $u, v \in \mathbb{N}$ avec $(u, v) = 1$ et $2|uv$. En fait, puisque $x^2 \equiv 1 \pmod 4$ pour x pair, alors on doit avoir que $2|v$ et que $u \equiv 1 \pmod 2$. On écrit $y = 2y_1$ et on observe que $2y_1^2 = uv$. Donc $u = a^2$ et $v = 2b^2$, où $(a, 2b) = 1$ et $ab = y_1$. D'autre côté, les relations $(u - v, u + v) = 1$ et $x^2 = (u - v)(u + v)$ impliquent que $u - v = k^2$ et $u + v = \ell^2$, pour quelques $k, \ell \in \mathbb{N}$ avec $(k, \ell) = 1$, $2 \nmid k\ell$ et $k\ell = x$. Donc $u = (k^2 + \ell^2)/2$ et $v = (\ell^2 - k^2)/2$ et, par conséquent, $k^2 + \ell^2 = 2a^2$ et $\ell^2 - k^2 = 4b^2$. On a que $(\ell - k, \ell + k) = 2$, d'où on déduit que $\ell - k = 2b_1^2$ et que $\ell + k = 2b_2^2$, où $b_1 b_2 = b$ et $(b_1, b_2) = 1$. En remplaçant ces relations à l'équation $k^2 + \ell^2 = 2a^2$, alors on trouve que

$$2a^2 = (b_1^2 + b_2^2)^2 + (b_2^2 - b_1^2) = 2(b_1^4 + b_2^4) \quad \Longrightarrow \quad b_1^4 + b_2^4 = a^2,$$

c'est-à-dire $(b_1, b_2, a) \in A$. Mais $a < z$, ce qui est une contradiction à la minimalité de z. Ceci conclut la démonstration que $A = \emptyset$. □

5.4 Sommes de deux carrés

Quels sont les nombres entiers qu'on peut écrire comme la somme de deux carrés ? C'est aussi un problème diophantien un peu différent des autres qu'on a étudié aux sections précédentes. Ici on demande pour quels $n \in \mathbb{N}$ il y a de solutions à l'équation $x^2 + y^2 = n$, avec $x, y \in \mathbb{Z}$. Clairement, ce n'est pas le cas toujours : on a que $3 \neq x^2 + y^2$, pour tous $x, y \in \mathbb{Z}$.

On laisse pour l'instant cette question, et on étudie la question semblable de quels nombres peuvent s'écrire comme la *différence de deux carrés*. En notation algébrique, pour quel n existent-ils $x, y \in \mathbb{N}$ tels que $n = x^2 - y^2$. Si c'est le cas, alors $n = (x-y)(x+y)$, c'est-à-dire n a une factorisation de la forme ab avec $a = x - y$ et $b = x + y$. Cette factorisation est un peu spéciale : on a que $2|a-b$. Vice versa, si $n = ab$ avec $2|a-b$, on peut résoudre le système $x - y = a$ et $x + y = b$ pour trouver deux solutions entières $x = (a+b)/2$ et $y = (b-a)/2$.

Alors, on voit que n peut s'écrire comme la différence de deux carrés si et seulement si il a une factorisation $n = ab$ avec $2|a-b$. Ceci est le cas toujours si n est impair : on peut prendre $a = 1$ et $b = n$. Ceci est aussi le cas si $4|n$: on peut prendre $a = 2$ et $b = n/2 \equiv 0 \pmod 2$. Mais ceci n'est pas possible si $2\|n$, car si $n = ab$, alors soit a soit b est pair, mais pas les deux. Donc $a - b$ est toujours impair.

On a classifié donc les nombres n qui peuvent s'écrire comme la différence de deux carrés : ils sont les nombres $n \not\equiv 2 \pmod 4$. On voit, alors, que notre réponse a une périodicité mod 4. Retournons maintenant à la question de nombres n qui peuvent s'écrire comme la somme de deux carrés. Est-ce que une réponse si semble est aussi possible ? Voici les premiers membres de la suite de nombres représentables comme la somme de deux carrés :

$$1, 2, 4, 5, 8, 9, 10, 13, 16, 17, 18, 20, 25, 26, 29, 32, 34, 36, 37, 40, 41, 45, 49, 50, 52, 53, 58, 61, 64,$$
$$65, 68, 72, 73, 74, 80, 81, 82, 85, 89, 90, 97, 98, 100, 101, 104, 106, 109, 113, 116, 117, 121, 122,$$
$$125, 128, 130, 136, 137, 144, 145, 146, 148, 149, 153, 157, 160, \ldots$$

La structure de cette suite semble beaucoup plus compliquée. Cependant, si on se concentre à ses membres premiers, la structure devient beaucoup plus simple :

$$2, 5, 13, 17, 29, 37, 41, 53, 61, 73, 89, 97, 101, 109, 113, 137, 149, 157, \ldots$$

On ose alors à conjecturer que les nombres premiers représentables comme la somme de deux carrés sont exactement le nombre 2 et les nombres premiers congruents à $1 \pmod 4$.

Une partie de notre conjecture est facile à montrer : on a que $2 = 1^2 + 1^2$. De plus, si $p \equiv 3 \pmod 4$, alors p ne peut pas s'écrire comme la somme de deux carrés. En effet, $x^2 \equiv 0, 1 \pmod 4$ pour tous $x \in \mathbb{Z}$, donc $x^2 + y^2 \equiv 0, 1, 2 \pmod 4$ pour tous $x, y \in \mathbb{Z}$. On trouve, alors que $p \neq x^2 + y^2$ pour tous $x, y \in \mathbb{Z}$.

La partie difficile de la conjecture concerne les nombres premiers $p \equiv 1 \pmod 4$ et la preuve qu'ils sont représentables comme la somme de deux carrés. On donne une démonstration algébrique. Le point de départ est la solution du problème de différences de carrés ci-dessus, qu'on essaie à imiter. On veut, alors, factoriser $x^2 + y^2$. Pour le faire, on passe aux

nombres complexes : on a que $x^2 + y^2 = (x + iy)(x - iy)$, où i est l'unité imaginaire pour laquelle $i^2 = -1$. Il apparait, alors, que la clé se trouve dans l'arithmétique des nombres de la forme $x + iy$ avec $x, y \in \mathbb{Z}$. Ces nombres s'appellent entiers gaussiens et on dénote leur ensemble par

$$\mathbb{Z}[i] := \{x + iy \,|\, x, y \in \mathbb{Z}\}.$$

Avant d'étudier les entiers gaussiens, on observe que le passage au plan complexe nous permettre de démontrer une propriété fondamentale de la suite des nombres qui sont la somme de deux carrés :

Lemme 5.4. *Si m et n sont la somme de deux carrés, alors mn l'est aussi.*

Démonstration. On a que $m = a^2 + b^2 = |a + ib|^2$ pour quelques $a, b \in \mathbb{Z}$. De même, $n = c^2 + d^2 = |c + id|^2$ pour quelques $c, d \in \mathbb{Z}$. Donc

$$mn = |a+ib|^2 \cdot |c+id|^2 = |(a+ib)(c+id)|^2 = |(ac-bd)+i(ad+bc)|^2 = (ac-bd)^2+(ad+bc)^2.$$

□

Le lemme ci-dessus nous montre que l'étude des premiers est la clé pour comprendre quels nombres sont la somme de deux carrés.

On revient maintenant à l'étude des entiers gaussiens. On peut imaginer que cette extension des entiers réguliers a des propriétés similaires. Tout d'abord, si on ajoute ou on multiplie deux entiers gaussiens, on obtient un nouvel élément de $\mathbb{Z}[i]$. On peut alors parler des nombres gaussiens premiers : on dit que $z = x + iy$ est premier dans $\mathbb{Z}[i]$ s'il n'a pas une factorisation 'non-triviale'. Mais il faut comprendre c'est quoi une telle factorisation. Par exemple, on peut écrire de façon triviale

$$z = 1 \cdot z \quad \text{et} \quad z = -1 \cdot (-z).$$

Mais, on peut aussi écrire

$$z = i \cdot (-iz) \quad \text{et} \quad z = -i \cdot iz$$

où $iz = -y + ix$ est un autre entier gaussien. Ces sont aussi de factorisations triviales, existantes toujours.

En général, les factorisations triviales sont obtenues par des éléments $u \in \mathbb{Z}[i] \setminus \{0\}$ tels que $1/u \in \mathbb{Z}[i]$, parce que dans ce cas on peut factoriser z dans $\mathbb{Z}[i]$ comme $z = u \cdot (u^{-1}z)$. Si $1/u \in \mathbb{Z}[i]$, alors $u \neq 0$ $|1/u| \geq 1$ et, par la suite, $|u| \leq 1$. Les seuls entiers gaussiens $u \neq 0$ qui satisfont cette inégalité sont les nombres $\pm 1, \pm i$.

Définition 5.5. On dit que l'entier gaussien z est un composé gaussien si on peut l'écrire comme $z = \alpha\beta$ avec $\alpha, \beta \notin \{\pm 1, \pm i\}$. Si z n'est pas un composé gaussien, on l'appelle un composé premier.

Observez que chaque nombre entier $n > 1$ qui est la somme de deux carrés est un composé gaussien quand on le voit comme un élément de $\mathbb{Z}[i]$. En effet, on a que $n = x^2 + y^2 = (x + iy)(x - iy)$ et, puisque $n > 1$, les facteurs $x \pm iy$ ne sont pas triviaux. En particulier, 2 et 5 sont de composés gaussiens.

Un réciproque partiel existe aussi :

Lemme 5.6. *Si p est un premier dans $\mathbb{Z}$ qui est composé dans $\mathbb{Z}[i]$, alors p est la somme de deux carrés.*

Démonstration. On a que $p = (a+ib)(c+id)$ pour quelques $a+ib, c+id \notin \{\pm 1, \pm i\}$. Donc

$$p^2 = |(a+ib)(c+id)|^2 = |a+ib|^2|c+id|^2 = (a^2+b^2)(c^2+d^2).$$

Puisque $a+ib, c+id \notin \{\pm 1, \pm i\}$, on a que $a^2+b^2, c^2+d^2 > 1$. La primalité de p alors implique que $p = a^2+b^2 = c^2+d^2$. □

Soit maintenant un nombre premier $p \equiv 1 \pmod 4$. On veut montrer qu'il est la somme de deux carrés. D'après le lemme 5.6, il suffit de montrer qu'il est composé dans $\mathbb{Z}[i]$. Supposons, au contraire, que p est premier dans $\mathbb{Z}[i]$. L'observation-clé pour obtenir une contradiction est que $(-1|p) = 1$ (voir théorème 4.10).Ceci veut dire qu'il existe un $m \in \mathbb{Z}$ tel que $p|m^2+1 = (m+i)(m-i)$. Si p était premier dans $\mathbb{Z}[i]$, ceci voudrait sûrement dire que soit $p|m+i$ ou $p|m-i$. Mais c'est une contradiction : si, par exemple, $p|m+i$, alors $m+i = p\cdot(a+bi)$ pour quelques $a, b \in \mathbb{Z}$. En particulier, $1 = pb$, ce qui est impossible. De même, on voit que la relation $p|m-i$ est absurde. On a montré alors le théorème suivant :

Théorème 5.7. *Un nombre premier $p > 2$ peut s'écrire comme la somme de deux carrés si et seulement si $p \equiv 1 \pmod 4$.*

Ou, peut-être, non ? On a utilisé dans notre argument que si p est un gaussien premier et $p|(m+i)(m-i)$, alors $p|m+i$ ou $p|m-i$ comme un résultat evident. Mais ce résultat assume que le théorème fondamental de l'arithmétique reste vrai dans $\mathbb{Z}[i]$, ce qui n'est pas évident ! En fait, le théorème fondamental de l'arithmétique peut être violé quand on passe à d'extensions de $\mathbb{Z}$. Considérons, par exemple, l'ensemble

$$\mathbb{Z}[\sqrt{-5}] := \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\}.$$

On a que $6 = 2\cdot 3 = (1+\sqrt{-5})(1-i\sqrt{5})$. On peut aussi montrer que les nombres $2, 3, 1\pm\sqrt{-5}$ ne se factorisent pas de façon non-trivial dans $\mathbb{Z}[\sqrt{-5}]$. Donc on voit que la factorisation unique est violée dans cet ensemble !

Comment peut-on montré que la factorisation unique en facteurs premiers est aussi vraie dans $\mathbb{Z}[i]$? La clé dans la démonstration du théorème fondamental de l'arithmétique (cf. théorème 1.26) est le lemme d'Euclid (cf. lemme). Et la clé dans la démonstration du lemme d'Euclid est la division euclidienne. On pourrait alors montré l'analogue de la division euclidienne dans $\mathbb{Z}[i]$. On a le théorème suivant :

Théorème 5.8. *Si $z, w \in \mathbb{Z}[i]$ avec $w \neq 0$, alors il existe $q, r \in \mathbb{Z}[i]$ tels que $|r| < |w|$ et $z = qw + r$.*

Démonstration. On considère le quotient $z/w = x + iy$ pour quelques $x, y \in \mathbb{Q}$. Il existe des entiers a, b tels que $|x-a|, |y-b| \leq 1/2$. On pose $q = a+ib$, pour que $|z/w - q|^2 = |x-a|^2 + |y-b|^2 \leq 1/4 + 1/4 = 1/2$. Donc $r := z - qw$ a magnitude $|r| \leq |w|/\sqrt{2} < |w|$. □

En utilisant le théorème 5.8, on peut montrer que le pgcd de deux nombres gaussiens z, w est une combinaison linéaire de z et w. On peut alors démontrer l'analogue du lemme d'Euclid sur $\mathbb{Z}[i]$. On laisse les détails aux lecteurs.

On peut enfin répondre à notre question et classifier les entiers qui sont la somme de deux carrés :

Théorème 5.9. *Considérons $n \in \mathbb{N}$ et sa factorisation première $n = p_1^{v_1} \cdots p_r^{v_r}$. Le nombre n peut être écrit comme la somme de deux carrés si et seulement si $2|v_i$ quand $p_i \equiv 3 \pmod 4$.*

Démonstration. Si $n = p_1^{v_1} \cdots p_r^{v_r}$ possède la propriété que $2|v_i$ quand $p_i \equiv 3 \pmod 4$, alors on peut écrire $n = d^2 m$, où

$$m = \prod_{\substack{1 \le i \le r \\ p_i = 2 \text{ ou } p_i \equiv 1 \pmod 4}} p_i.$$

Du théorème 5.7, on trouve que $p_i = x_i^2 + y_i^2$ quand $p_i \equiv 1 \pmod 4$. Aussi, on a trivialement que $2 = 1^2 + 1^2$. Donc le lemme 5.4 implique que m est aussi la somme de deux carrés, soit $m = x^2 + y^2$. Par la suite, $n = (dx) + (dy)^2$, ce qui est ce qu'il fallait démontrer.

Réciproquement, supposons que $n = x^2 + y^2$. On pose $d = (x, y)$ et on écrit $x = da$ et $y = db$, où $(a, b) = 1$, pour que $n = d^2(a^2 + b^2)$. Il suffit de montrer que $a^2 + b^2$ n'est pas divisible par de nombres premiers $p \equiv 3 \pmod 4$. En effet, soit $p|a^2 + b^2$, $p > 2$. Puisque $(a, b) = 1$, alors $(ab, p) = 1$. Donc

$$\begin{aligned} a^2 + b^2 \equiv 0 \pmod p \quad &\Longrightarrow \quad (ab^{-1})^2 \equiv -1 \pmod p \quad \Longrightarrow \quad \left(\frac{-1}{p}\right) = 1 \\ &\Longrightarrow \quad p \equiv 1 \pmod 4, \end{aligned}$$

ce qui termine la démonstration. □

On donne ci-dessus une démonstration alternative et plus directe du théorème 5.7 en évitant la théorie des entiers gaussiens. Cependant, l'idée principale vient de cette théorie et une présentation de cet argument sans l'étude de $\mathbb{Z}[i]$ pourrait le faire apparaitre comme de la 'magique.

Démonstration alternative du théorème 5.7. On a toujours que $x^2 \equiv 0, 1 \pmod 4$. Donc $x^2 + y^2 \equiv 0, 1, 2 \pmod 4$, ce qui implique que si $p > 2$ est représentable comme la somme de deux carrés, alors nécessairement $p \equiv 1 \pmod 4$.

Réciproquement, supposons que $p \equiv 1 \pmod 4$. Donc $\left(\frac{-1}{p}\right) = 1$, c'est-à-dire il existe $r \in \{1, \ldots, p-1\}$ tel que $r^2 \equiv -1 \pmod p$. On pose $M = \lfloor \sqrt{p} \rfloor$, pour que

$$M < \sqrt{p} < M + 1$$

(en général, on a que $\lfloor x \rfloor \le x < \lfloor x \rfloor + 1$, mais dans ce cas on ne peut pas avoir que $M = \sqrt{p}$ parce que un nombre premier n'est pas un carré parfait). Soit

$$X = \{(a, b) \in \mathbb{Z}^2 : 0 \le a, b \le M\}.$$

Pour chaque $(a, b) \in X$, on considère le nombre $a + br$. Puisque $|X| = (M + 1)^2 > p$, les nombres $a + br$ ne peuvent pas être tous différents modulo p. Donc il existe deux éléments de X (a, b) et (a', b') qui sont distincts et pour lesquels $a + br \equiv a' + b'r \pmod p$. C'implique que $(a - a') \equiv r(b' - b) \pmod p$ et, par la suite $(a - a')^2 \equiv r^2(b' - b)^2 \equiv -(b' - b)^2 \pmod p$. Donc le nombre

$$m := (a - a')^2 + (b - b')^2$$

est un multiple de p qui est positif car $(a, b) \neq (a', b')$. De plus, on a que $-M \leq a' - a \leq M$ et $-M \leq b' - b \leq M$, ce qui implique que $m \leq 2M^2 < 2p$. Mais le seul multiple de p qui est dans l'intervalle $(0, 2p)$ est p. Donc $m = p = (a - a')^2 + (b - b')^2$, ce qui est ce qu'il fallait montrer. □

5.5 Sommes de quatre carrés

On a vu déjà que il y a de nombres entiers qui ne peuvent s'exprimer comme la somme de deux carrés. La même chose est vraie pour les sommes de trois carrés : on a que $23 \neq x^2 + y^2 + z^2$, pour tous $x, y, z \in \mathbb{Z}^3$. Cependant, Lagrange a montré le théorème suivant :

Théorème 5.10 (Lagrange)**.** *Chaque nombre naturel peut s'écrire comme la somme de quatre carrés.*

Comme dans le cas de deux carrés, on commence avec un lemme préparatoire qui réduire le théorème aux nombres premiers :

Lemme 5.11. *Si a et b sont de sommes de quatre carrés, alors la même chose est vraie pour ab.*

Démonstration. Si $a = x_1^2 + x_2^2 + x_3^2 + x_4^2$ et $b = y_1^2 + y_2^2 + y_3^2 + y_4^2$, alors on peut vérifier facilement que

$$\begin{aligned} ab = {} & (x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4)^2 + (-x_1y_2 + x_2y_1 - x_3y_4 + x_4y_3)^2 \\ & + (-x_1y_3 + x_3y_1 + x_2y_4 - x_4y_2)^2 + (-x_1y_4 + x_1y_4 - x_2y_3 + x_3y_2)^2, \end{aligned} \tag{5.8}$$

d'où le lemme découle tout de suite. □

Démonstration du théorème 5.10. Le lemme 5.11 nous permet de considérer seulement le cas d'un nombre premier. Alors, soit p un nombre premier. Si $p = 2$, on a que $2 = 1^2+1^2+0^2+0^2$, comme voulu. Supposons maintenant que $p > 2$. Si $p \equiv 1 \pmod 4$, alors on peut utiliser le théorème 5.7. Toutefois, on montre que p peut s'écrire comme la somme de quatre carrés avec un argument unifié pour touts les $p > 2$. Ici on ne peut pas « linéariser » le problème, comme on l'a fait au cas de deux carrés (voir démonstration alternative du théorème 5.7). On considère l'ensemble

$$A := \{a = x_1^2 + x_2^2 + x_3^2 + x_4^2 : x_i \in \mathbb{Z}\ (1 \leq i \leq 4),\ p|a,\ a > 0\}$$

et on montre que $\min A = p$. Tout d'abord, il faut montrer que $A \neq \emptyset$, pour que son minimum soit bien défini. Les nombres x^2, $0 \leq x \leq (p - 1)/2$, sont tous distincts modulo p. De même,

les nombres $-1-y^2$, $0 \leq y \leq (p-1)/2$, sont aussi distincts modulo p. Puisque il existe seulement p distincts classes de congruences modulo p, alors il existe $0 \leq x, y \leq (p-1)/2$ tels que $x^2 \equiv -1-y^2 \pmod{p}$, ce qui implique que $x^2+y^2+1^2+0^2 \in A$. Alors $A \neq \emptyset$, comme voulu. Soit

$$a = x_1^2 + x_2^2 + x_3^2 + x_4^2 = \min A.$$

On a que $a = mp$, pour un $m \in \mathbb{N}$. Il suffit de montrer que $m = 1$. D'abord, on montre que m est impair. En effet, si m était pair, alors $x_1^2 + x_2^2 + x_3^2 + x_4^2 \equiv 0 \pmod{2}$, ce qui implique que soit tous les x_i sont impairs, soit ils sont tous pairs, soit il y en a deux qui sont pairs et deux qui sont impairs. En tout cas, on peut écrire $\{x_1, x_2, x_3, x_4\} = \{x_{i_1}, x_{i_2}\} \cup \{x_{i_3}, x_{i_4}\}$ pour que $x_{i_1} \equiv x_{i_2} \pmod{2}$ et $x_{i_3} \equiv x_{i_4} \pmod{2}$. Donc

$$\frac{a}{2} = \frac{x_{i_1}^2 + x_{i_2}^2 + x_{i_3}^2 + x_{i_4}^2}{2} = \left(\frac{x_{i_1}+x_{i_2}}{2}\right)^2 + \left(\frac{x_{i_1}-x_{i_2}}{2}\right)^2 + \left(\frac{x_{i_3}+x_{i_4}}{2}\right)^2 + \left(\frac{x_{i_3}-x_{i_4}}{2}\right)^2,$$

ce qui implique que $a/2 \in A$. C'est impossible car $a = \min A$ et $a/2 < a$. Donc m est impair, comme affirmé. L'idée maintenant est d'utiliser la relation (5.8) : si $b = y_1^2+y_2^2+y_3^2+y_4^2 > 0$, alors

$$\begin{aligned} mpb = ab &= (x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4)^2 + (-x_1y_2 + x_2y_1 - x_3y_4 + x_4y_3)^2 \\ &\quad + (-x_1y_3 + x_3y_1 + x_2y_4 - x_4y_2)^2 + (-x_1y_4 + x_1y_4 - x_2y_3 + x_3y_2)^2 \\ &=: z_1^2 + z_2^2 + z_3^2 + z_4^2 \in A. \end{aligned}$$

Le but est de trouver $b \equiv 0 \pmod{m}$ tel que $z_i \equiv 0 \pmod{m}$, pour tout $i \in \{1,2,3,4\}$. Dans ce cas, on aura que $pb/m = (z_1/m)^2+(z_2/m)^2+(z_3/m)^2+(z_4/m)^2 \in A$. Si $m > 1$, on affirme qu'on peut choisir un tel b qui satisfait l'inégalité $b < m^2$, ce qui implique que $pb/m < pm = a = \min A$, une contradiction. En effet, pour tout i, on peut choisir $y_i \in \mathbb{Z}/\cap(-m/2, m/2)$ tel que $x_i \equiv y_i \pmod{m}$. Si $b = y_1^2+y_2^2+y_3^2+y_4^2 = 0$, alors $y_i = 0$ pour tout $i \in \{1,\ldots,4\}$ et, par la suite, $a/m^2 = (x_1/m)^2 + (x_2/m)^2 + (x_3/m)^2 + (x_4/m)^2 \in A$. C'est impossible car $a/m^2 < a = \min A$. Donc $b > 0$. De plus, on a que

$$\begin{aligned} b &\equiv x_1^2 + x_2^2 + x_3^2 + x_4^2 \pmod{m} \\ &\equiv 0 \pmod{m} \end{aligned}$$

et $b < 4 \cdot (m/2)^2 = m^2$, comme voulu. Finalement, on a que

$$\begin{aligned} z_1 = x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 &\equiv x_1^2 + x_2^2 + x_3^2 + x_4^2 \pmod{m} \\ &\equiv 0 \pmod{m} \end{aligned}$$

et, de même, $z_2 \equiv z_3 \equiv z_4 \equiv 0 \pmod{m}$. Comme on a vu au-dessous, ceci nous amène à une contradiction car c'implique que $bp/m \in A$ bien que $bp/m < mp = a = \min A$. Par la suite, il faut que $m = 1$ et le théorème découle. □

Les quaternions de Hamilton

On conclut cette section avec une discussion qui donne une explication plus concret du lemme 5.11. Cette explication passe par les *quaternions de Hamilton*.

On sait qu'on peut identifier $\mathbb{C}$ avec le plan $\mathbb{R}^2$. Cependant, la structure multiplicative de $\mathbb{C}$ devient beaucoup plus claire et intuitive quand on utilise la notation $a+bi$ au lieu de (a,b). En effet, avec la règle $i^2=-1$, on trouve tout de suite que $(a+ib)(c+id) = (ac-db)+i(ad+bc)$, la multiplication familière de deux nombres complexes. Comme on l'a vu, cette propriété est très importante dans l'étude des nombres qui peuvent s'écrire comme la somme de deux carrés.

De même façon, une autre structure algébrique devient important dans l'étude des nombres qui sont la somme de quatre carrés. Cette structure est l'ensemble des quaternions de Hamilton, c'est-à-dire l'ensemble

$$\mathbb{H} := \{a+bi+cj+dk : a,b,c,d \in \mathbb{R}\},$$

où i,j,k sont de 'nombres' linéairement indépendants sur $\mathbb{R}$. Alors, $\mathbb{H}$ est juste une autre façon d'écrire $\mathbb{R}^4$ (on identifie $a+bi+cj+dk$ avec le vecteur (a,b,c,d)), exactement de même façon que $\mathbb{C}$ est une autre façon d'écrire $\mathbb{R}^2$.

Comme un espace linéaire, $\mathbb{H}$ possède d'une addition de ses éléments. On peut aussi les multiplier en introduisant les règles $i^2=j^2=k^2=-1$, $ij=k$, $jk=i$ et $ki=j$. À partir de ces équations, on peut déduire que $ji=j(jk)=j^2k=-k$, $kj=k(ki)=k^2i=-i$ et $ik=i(ij)=i^2j=-j$. En particulier, la multiplication dans $\mathbb{H}$ n'est pas commutative. En suivant les règles précédentes, on s'amène à la relation suivante :

$$\begin{aligned}(a+bi+cj+dk)(a'+b'i+c'j+d'k) &= aa'+ab'i+ac'j+ad'k\\ &\quad +ba'i+bb'i^2+bc'ij+bd'ik\\ &\quad +ca'j+cb'ji+cc'j^2+cd'jk\\ &\quad +da'k+db'ki+dc'kj+dd'k^2\\ &= aa'+ab'i+ac'j+ad'k\\ &\quad +ba'i-bb'+bc'k-bd'j\\ &\quad +ca'j-cb'k-cc'+cd'i\\ &\quad +da'k+db'j-dc'i-dd'\\ &= (aa'-bb'-cc'-dd')+(ab'+ba'+cd'-c'd)i\\ &\quad +(ac'+ca'-bd'+db')j+(ad'+da'+bc'-cb')k.\end{aligned}$$

Puis, étant donné un quaternion $\alpha = a+bi+cj+dk$, on défini sa norme $N(\alpha) := a^2+b^2+c^2+d^2$. On peut verifier que

$$N(\alpha) = \alpha\overline{\alpha},$$

où $\overline{\alpha} := a-bi-cj-dk$ est le conjugué de α. Aussi, on peut verifier que

$$\overline{\alpha\beta} = \alpha\beta.$$

Donc

$$N(\alpha\beta) = N(\alpha)N(\beta).$$

En appliquant cette formule avec $\alpha = x_1+x_2i+x_3j+x_4k$ et $\beta = y_1-y_2i-y_3j-y_4k$, on arrive à (5.8).

Exercices

EXERCICE 5.1. Trouvez toutes les solutions composées de nombres entiers aux équations suivantes :

$$(a)\ 10x + 2y = 9 \qquad (b)\ 7x + 11y = 20 \qquad (c)\ 10x + 35y = 100.$$

EXERCICE 5.2. Supposez que vous avez 20 timbres de 7$ chaque et 10 timbres de 11$ chaque. Vous voulez envoyez à Toronto un colis qui coûte 151$. Est-ce que c'est possible de le payer sans acheter d'autres timbres ? Dans combien de façons différentes vous pouvez le payer ? Est-ce que ce serait possible de payer le colis avec les timbres disponibles s'il coûtait 244$?

EXERCICE 5.3.

(a) Trouvez tous les triplets pythagoriciens qui forment une progression arithmétique.
(b) Trouvez toutes les solutions entières à l'équation $x^2 + y^2 = z^4$ sachant que $(x, y, z) = 1$.

EXERCICE 5.4. L'équation $x^4 + x^2 = y^4 + 5$ possède-t-elle des solutions entières en x et y ?

EXERCICE 5.5. Résolvez le système suivant :

$$\begin{aligned} 2x(1 + y + y^2) &= 3(1 + y^4) \\ 2y(1 + z + z^2) &= 3(1 + z^4) \\ 2z(1 + x + x^2) &= 3(1 + x^4) \end{aligned}$$

EXERCICE 5.6. Utilisez la méthode de la descente infinie pour montrer que l'équation $x^2 = 2y^2$ n'a pas de solutions entiers (et, par la suite, $\sqrt{2} \notin \mathbb{Q}$).

EXERCICE 5.7.

(a) Montrez que un nombre premier $p > 2$ peut être écrit comme $x^2 + 2y^2$, où $x, y \in \mathbb{N}$, si et seulement si $p \equiv 1, 3 \pmod 8$.
(b) Montrez que si $p | a^2 + 2b^2$ avec $(a, b) = 1$, alors soit $p = 2$ soit $p \equiv 1, 3 \pmod 8$.
(c) Déterminez quels sont les nombres naturels n qui peuvent s'écrire dans la forme x^2+2y^2.

EXERCICE 5.8. Est-ce que c'est possible de donner une structure multiplicative à $\mathbb{R}^3$ comme on a fait pour $\mathbb{R}^2$ et pour $\mathbb{R}^4$?

Deuxième partie

Méthodes analytiques

Chapitre 6

Et il en exista infiniment beaucoup

Depuis la preuve d'Euclide qu'il existe un nombre infini de nombres premiers, la répartition de ces objets fondamentaux a fasciné les mathématiciens. Différent d'autres suites qui ont une structure très régulière, comme la suite des carrés, les nombres premiers ne semblent pas suivre un motif visible. Par conséquent, deviner la location exacte du n-ième premier apparaît être un défi impossible quand n devient de plus en plus grand.

Puisque la suite des premiers semble être si chaotique, on peut fixer l'objectif plus modeste de comprendre la location *approximative* du n-ième premier. De façon équivalent, on cherche une bonne approximation à la fonction de dénombrement des nombres premiers

$$\pi(x) := \#\{p \leq x\}.$$

Si p_n dénote le n-ième nombre premier, alors $\pi(p_n) = n$, pour que chaque approximation à $\pi(x)$ peut se traduire immédiatement à une approximation à p_n, et vice versa.

L'étude de la repartition de la repartition des premiers a préoccupé le jeune Gauss. Après avoir examiné des tables de nombres premiers, il a observé que la densité des premiers autour de x est à peu prés $1/\log x$. En le traduisant à la langue du calcul, ceci veut dire qu'une bonne approximation pour $\pi(x)$ est donnée par l'intégrale logarithmique

$$\mathrm{li}(x) := \int_2^x \frac{\mathrm{d}t}{\log t}.$$

En appliquant la règle de l'Hôpital, on trouve que

$$\lim_{x\to\infty} \frac{\mathrm{li}(x)}{x/\log x} = 1.$$

Donc l'estimation de Gauss implique que $\pi(x)$ est approximativement égal à $x/\log x$ quand $x \to \infty$. De façon symbolique, on écrit

$$\pi(x) \sim \frac{x}{\log x} \quad (x \to \infty), \tag{6.1}$$

qui veut dire que le rapport de ces deux fonctions tend vers 1 quand $x \to \infty$. On discutera cette notation plus profondément dans la section 7.1. De manière équivalente, l'estimation de Gauss (6.1) constate que $p_n \sim n \log n$ quand $n \to \infty$ (exercice).

Cela a pris plus qu'une siècle de démontrer la conjecture de Gauss pour $\pi(x)$. Le chemin de la preuve a été décrite par Riemann dans son mémoire *Über die Anzahl der Primzahlen unter einer gegebenen Grösse* qui a été publié en 1859 et qui a causé une révolution au sujet. Dans son travail, Riemann a expliqué comment $\pi(x)$ est intimement lié aux propriétés analytiques la fonction zeta de Riemann

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s},$$

définie pour l'instant quand $\Re(s) > 1$. Il est procédé de proposer un programme dont la complétion amènerait à une comprehension profonde des nombres premiers et qui établirait qu'il existe une constant $c > 0$ telle que

$$|\pi(x) - \mathrm{li}(x)| \leq c\sqrt{x}\log x \quad \text{pour tout} \quad x \geq 2, \tag{6.2}$$

une forme très forte de l'estimation de Gauss. En 1895, von Mangoldt a prouvé rigoureusement toutes les étapes du plan de Riemann, sauf une. Cette dernière étape dans son plan reste insaisissable. C'est la fameuse hypothèse de Riemann qu'on discutera dans la section 9. Quand même, en 1896, Hadamard et de la Vallée Poussin ont montré une forme faible de l'hypothèse de Riemann qui a été assez forte pour leur permettre d'établir l'estimation de Gauss, connue aujourd'hui comme le théorème des nombres premiers :

Théorème des nombres premiers. *Quand $x \to \infty$, on a que $\pi(x) \sim x/\log x$.*

On parlera de la démonstration de ce résultat fondamental à la section 9. Avant de le faire, on développe plusieurs outils et on démontre des autres résultats intermédiaires qui sont intéressantes eux-mêmes.

Chapitre 7

Estimations asymptotiques

7.1 La notation asymptotique

Les fonctions qu'on rencontre dans la théorie des nombres sont souvent irrégulières. Alors, on veut les remplacer par d'autres fonctions qui s'en rapprochent et qui sont plus faciles à analyser. Comme un exemple, considérons la fonction $f(x)$ qui compte le nombre d'entiers dans l'intervalle $[1, x]$ avec $x \geq 1$. On peut facilement voir que f est une fonction en escalier qui a de sauts de longueur 1 à tous les entiers. On peut écrire f en termes d'une fonction plus familière, la partie entière de x qu'on dénote par $\lfloor x \rfloor$. Puisque $\lfloor x \rfloor \leq x < x + 1$, c'est clair que $f(x) = \lfloor x \rfloor$, d'où $f(x) = x + E(x)$ pour une fonction $E(x)$ bornée par 1 en valeur absolue. On a donc remplacé la fonction en escalier $f(x) = \lfloor x \rfloor$ par son approximation lisse x, et le terme restant de cette approximation est une fonction bornée. On exprime ce fait par la formule asymptotique

$$\lfloor x \rfloor = x + O(1). \tag{7.1}$$

Généralement, étant données les fonctions complexes f, g and h, et un sous-ensemble I de leurs domaines de définition, on écrit

$$f(x) = g(x) + O(h(x)) \quad (x \in I), \tag{7.2}$$

et on lit '$f(x)$ est égale à $g(x)$ plus grand-O de $h(x)$', s'il existe une constante $c = c(f, g, I)$ telle que

$$|f(x) - g(x)| \leq c \cdot h(x) \quad \text{pour chaque } x \in I.$$

On appelle souvent la constante c absolue pour signifier qu'elle ne dépend pas de l'argument des fonctions f, g et h, ni d'autres paramètres qui peuvent être présents.

On remarque que la différence $x - \lfloor x \rfloor$ dans (7.1) est la partie fractionnelle de x, dénotée par $\{x\}$. Cependant, c'est souvent plus simple d'ignorer la valeur exacte du terme restant et de juste garder en tête que il est une fonction bornée. Supposons, par exemple, qu'on veut trouver une approximation de l'expression $\sum_{n \leq x} \lfloor x/n \rfloor$. Le pouvoir de la notation asymptotique se révèle dans l'évaluation approximative de telles expressions compliquées, car elle transforme des inégalités à des égalités : on a que $\sum_{n \leq x} \lfloor x/n \rfloor = x \sum_{n \leq x} 1/n + O(x)$, car

$$\left| \sum_{n \leq x} \lfloor x/n \rfloor - x \sum_{n \leq x} 1/n \right| = \left| \sum_{n \leq x} (\lfloor x/n \rfloor - x/n) \right| \leq \sum_{n \leq x} 1 \leq x.$$

Mais il faut faire attention : les règles usuelles de l'addition et de la multiplication changent. Par exemple, puisque la somme de deux fonctions bornées est aussi bornée, on a que $O(1) + O(1) = O(1)$. De manière similaire, on a que $O(1) \cdot O(1) = O(1)$ and $O(1) - O(1) = O(1)$.

La notation asymptotique nous permet aussi de comparer l'ordre de magnitude de fonctions différentes : si

$$f(x) = O(g(x)) \quad (x \in I),$$

alors on dit que "f a plus petite ordre de magnitude que g sur I". On peut aussi exprimer la relation asymptotique ci-dessus en utilisant la notation de Vinogradov :

$$f(x) \ll g(x) \quad (x \in I).$$

Si $f(x) \ll g(x)$ et $g(x) \ll f(x)$ pour tout $x \in I$, alors on écrit

$$f(x) \asymp g(x) \quad (x \in I)$$

et on dit que "f et g ont la même ordre de magnitude sur I".

Il existe deux autres notations asymptotiques importantes. On écrit

$$f(x) = o(g(x)) \quad (x \to x_0) \quad \Leftrightarrow \quad \lim_{x \to x_0} \frac{f(x)}{g(x)} = 0$$

et

$$f(x) \sim g(x) \quad (x \to x_0) \quad \Leftrightarrow \quad \lim_{x \to x_0} \frac{f(x)}{g(x)} = 1,$$

où dans les deux définitions $x_0 \in \hat{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ et g est non-zéro dans un voisinage de x_0.

On donne ci-dessous quelques exemples afin d'illustrer l'utilisation des notations asymptotiques qu'on a introduit.

Exemple 7.1. Souvent on a une expression composée qu'on veut évaluer, comme par exemple $\log \lfloor x \rfloor$. Puisque $\lfloor x \rfloor = x + O(1)$, alors le théorème des accroissements finis implique que

$$\log \lfloor x \rfloor = \log x + O(1) \cdot \frac{1}{c},$$

pour un c entre $\lfloor x \rfloor$ and x. Par la suite,

$$\log \lfloor x \rfloor = \log x + O(1/x) \quad (x \geq 1).$$

Exemple 7.2. Une application simple du théorème des accroissements finis est souvent insuffisant car on a besoin de précision extra dans notre approximation. Dans ce cas, on peut utiliser le théorème de Taylor. Par exemple, on a

$$\sqrt{x + \log x} = \sqrt{x} + \frac{\log x}{2\sqrt{x}} - \frac{\log^2 x}{4x^{3/2}} + O\left(\frac{\log^3 x}{x^{5/2}}\right) \quad (x \geq 1).$$

Exemple 7.3. La notation asymptotique peut être aussi utilisée pour obtenir des expansions asymptotiques d'intégrales qu'on ne peut pas calculer exactement en termes de fonctions élémentaires. Comme un exemple, on analyse l'intégrale logarithmique. Par intégration par parties, on a que

$$\begin{aligned}
\mathrm{li}(x) &= \int_2^x x' \frac{\mathrm{d}y}{\log y} \\
&= \frac{x}{\log x} + O(1) + \int_2^x \frac{\mathrm{d}y}{\log^2 y} \\
&= \frac{x}{\log x} + \frac{x}{\log^2 x} + O(1) + 2\int_2^x \frac{\mathrm{d}y}{\log^3 y} \\
&\vdots \\
&= \frac{x}{\log x} + \frac{x}{\log^2 x} + \frac{2!x}{\log^3 x} + \cdots + \frac{(N-1)!x}{\log^N x} + O_N(1) + N!\int_2^x \frac{\mathrm{d}y}{\log^{N+1} y}.
\end{aligned}$$

La dernière intégrale est $\sim x/\log^{N+1} x$ quand $x \to \infty$ d'après la règle de l'Hôpital's. On arrive donc à la formule asymptotique

$$\mathrm{li}(x) = \frac{x}{\log x} + \frac{x}{\log^2 x} + \frac{2!x}{\log^3 x} + \cdots + \frac{(N-1)!x}{\log^N x} + O_N\left(\frac{x}{(\log x)^N}\right) \quad (x \geq 2).$$

7.2 Une preuve analytique de l'infinité des premiers

Pour donner une idée du pouvoir de la méthode analytique, on donne une démonstration analytique de l'infinité des nombres premiers. Pour décrire l'idée, on considère la possibilité absurde que le nombre 2 est le seul nombre premier. Puisque chaque entier positif peut se factoriser en premiers, il suit que chaque entier positif est une puissance de 2. En particulier,

$$\{n \leq x\} = \{2^k \leq x : k \geq 0\}.$$

Le côté gauche a cardinalité $\lfloor x \rfloor = x + O(1) \sim x$ quand $x \to \infty$. Par contre, le côté droit a cardinalité

$$\#\{2^k \leq x : k \geq 0\} = \#\{0 \leq k \leq \log x/\log 2\} = 1 + \lfloor \log x/\log 2 \rfloor = \frac{\log x}{\log 2} + O(1) \sim \frac{\log x}{\log 2}$$

quand $x \to \infty$. Mais ceci est absurd car $\lim_{x\to\infty} \frac{\log x}{x} = 0$. On a montré alors que ce n'est pas possible que 2 est le seul nombre premier pour des raisons analytiques : il n'y a pas assez de puissances de 2 pour créer tous les nombres !

On peut généraliser cette idée : supposons que les seuls nombres premiers étaient 2 et 3. Donc

$$\{n \leq x\} = \{2^k 3^\ell \leq x : k, \ell \geq 0\}.$$

Mais on a que

$$\begin{aligned}\#\{2^k3^\ell \le x : k,\ell \ge 0\} &\le \#\{(k,\ell)\in\mathbb{Z}^2 : 0\le k\le \log x/\log 2,\ 0\le \ell\le \log x/\log 3\}\\ &\le \left(1+\frac{\log x}{\log 2}\right)\left(1+\frac{\log x}{\log 3}\right)\\ &\sim \frac{(\log x)^2}{(\log 2)(\log 3)} = o_{x\to\infty}(x),\end{aligned}$$

ce qui est absurd.

Plus généralement, si on suppose que les seuls nombres premiers sont $p_1,\dots,p_m$, alors

$$\{n\le x\} = \{p_1^{k_1}\cdots p_m^{k_m} : k_1,\dots,k_m\ge 0\}.$$

Cependant,

$$\begin{aligned}\#\{p_1^{k_1}p_2^{k_2}\cdots p_m^{k_m} : k_1,\dots,k_m\ge 0\} &\le \#\{(k_1,\dots,k_m)\in\mathbb{Z}^m : 0\le k_j\le \log x/\log p_j\ \forall j\}\\ &\le \prod_{j=1}^m\left(1+\frac{\log x}{\log p_j}\right)\\ &\sim \frac{(\log x)^m}{\prod_{j=1}^m \log p_j} = o_{x\to\infty}(x)\end{aligned}$$

d'après la règle de l'Hôpital. On est arrivé encore à une contradiction. Ceci montre l'infinité des premiers. En fait, l'argument ci-dessus a la capacité d'estimer $\pi(x)$ si utilisé avec dextérité. On voit alors la puissance potentielle des méthodes analytiques.

7.3 Sommation par parties

Un des outils principaux de la théorie analytique des nombres est une analogie discrète de l'intégration par parties. Elle est appelée sommation par parties ou sommation partielle, et sa forme la plus simple est la formule de sommation d'Abel :

$$\sum_{n=M+1}^{N} a_n b_n = A_n b_n\Big|_{n=M}^{N} - \sum_{n=M}^{N-1} A_n(b_{n+1}-b_n), \tag{7.3}$$

où $A_n = a_1+\cdots+a_n$. La formule d'Abel est prouvée en observant que $a_n = A_n - A_{n-1}$ et, puis, en re-arrangeant la sommation :

$$\sum_{n=M+1}^{N} a_n b_n = \sum_{n=M+1}^{N} A_n b_n - \sum_{n=M+1}^{N} A_{n-1} b_n = \sum_{n=M+1}^{N} A_n b_n - \sum_{n=M}^{N-1} A_n b_{n+1},$$

où on a fait le changement de variables $n\to n+1$ à la deuxième somme. Puisque

$$\sum_{n=M+1}^{N} A_n b_n = A_N b_N - A_M b_M + \sum_{n=M}^{N-1} A_n b_n,$$

la relation (7.3) suit.

La formule de sommation d'Abel nous permet de comprendre les sommes partielles de a_nb_n, à condition qu'on sache déjà le comportement de A_n et de b_n. Dans le cas spécial où $b_n = f(n)$ pour une fonction f qui est continûment differentiable, la formule d'Abel peut être écrite en termes de l'intégrale de Riemann : on a que

$$\sum_{n=M}^{N-1} A_n(b_{n+1} - b_n) = \sum_{n=M}^{N-1} A_n(f(n+1) - f(n)) = \sum_{n=M}^{N-1} A_n \int_n^{n+1} f'(t)\mathrm{d}t.$$

En observant que la fonction en escalier $A(x) := \sum_{1\le n\le x} a_n$ est constante pour $x \in (n, n+1)$, on a que

$$\sum_{n=M}^{N-1} A_n(b_{n+1} - b_n) = \sum_{n=M}^{N-1} \int_n^{n+1} A(t)f'(t)\mathrm{d}t = \int_M^N A(t)f'(t)\mathrm{d}t.$$

On déduit alors que

$$\sum_{n=M+1}^{N} a_nf(n) = A(x)f(x)\Big|_{n=M}^{N} - \int_M^N A(t)f'(t)\mathrm{d}t.$$

Plus généralement, on peut montrer de manière similaire que

$$\sum_{y<n\le z} a_nf(n) = A(t)f(t)\Big|_{t=y}^{z} - \int_y^z A(t)f'(t)\mathrm{d}t \tag{7.4}$$

pour $y, z \in \mathbb{R}_{\ge 0}$.

Souvent, l'information qu'on a concernant $A(x)$ est dans une forme asymptotique : $A(x) = M(x) + R(x)$, où $M(x)$ est le terme principal, donné par une fonction continûment differentiable, et $R(x)$ est le terme restant à l'approximation de $A(x)$ par $M(x)$. Dans ce cas, on a

$$\begin{aligned}\sum_{y<n\le z} a_nf(n) = M(t)f(t)\Big|_{t=y}^{z} - \int_y^z M(t)f'(t)\mathrm{d}t \\ + R(t)f(t)\Big|_{t=y}^{z} - \int_y^z R(t)f'(t)\mathrm{d}t.\end{aligned}$$

On peut réécrire la première ligne en utilisant intégration par parties. On trouve alors que

$$\sum_{y<n\le z} a_nf(n) = \int_y^z f(t)M'(t)\mathrm{d}t + R(t)f(t)\Big|_{t=y}^{z} - \int_y^z R(t)f'(t)\mathrm{d}t. \tag{7.5}$$

Dans le cas spécial où $a_n = 1$, on a que $A(x) = \lfloor x \rfloor$, et on l'écrit comme $x - \{x\}$. Ceci nous amène à la formule de sommation d'Euler-McLaurin :

Théorème 7.4. *Si f est continûment differentiable sur $[y, z]$, alors*

$$\sum_{y<n\le z} f(n) = \int_y^z f(t)\mathrm{d}t - \{t\}f(t)\Big|_{t=y}^{z} + \int_y^z \{t\}f'(t)\mathrm{d}t.$$

En particulier, si f est continûment differentiable sur $[1, +\infty)$, alors

$$\sum_{n\leq x} f(n) = \int_1^x f(t)\mathrm{d}t + f(1) - \{x\}f(x) + \int_1^x \{t\}f'(t)\mathrm{d}t.$$

Remarque 7.5. On peut considérer $\sum_{y<n\leq z} f(n)$ comme une somme de Riemann pour l'intégrale $\int_y^z f(t)\mathrm{d}t$. Bien sûr, ici on fait la sommation sur une (quasi-)partition de $[y, z]$ de diamètre 1, alors la différence $\sum_{y\leq n\leq z} f(n) - \int_y^z f(t)\mathrm{d}t$ pourrait être assez grande. Cependant, si f n'oscille pas trop (c'est-à-dire si sa dérivée est habituellement petite), alors cette différence devrait être assez petite. La formule d'Euler-McLaurin est une quantification de cette idée.

On donne maintenant deux applications fameuses de la formule d'Euler-McLaurin. On commence par une estimation de la croissance de la série harmonique. Le symbole γ dans son énoncé dénote la constante d'Euler-Mascheroni, définie par

$$\gamma = 1 - \int_1^\infty \frac{\{t\}}{t^2}\mathrm{d}t = 0.57721\ldots$$

Théorème 7.6. *Pour tout $x \geq 1$, on a que*

$$\sum_{n\leq x} \frac{1}{n} = \log x + \gamma + O\left(\frac{1}{x}\right).$$

Démonstration. D'après le théorème 7.4, on a que

$$\begin{aligned}\sum_{n\leq x} \frac{1}{n} &= \int_1^x \frac{\mathrm{d}t}{t} + 1 - \frac{\{x\}}{x} - \int_1^x \frac{\{t\}}{t^2}\mathrm{d}t \\ &= \log x + 1 - \int_1^x \frac{\{t\}}{t^2}\mathrm{d}t + O\left(\frac{1}{x}\right).\end{aligned}$$

L'intégrale $\int_1^\infty \{t\}t^{-2}\mathrm{d}t$ converge absolument. De plus, on a l'estimation suivante pour sa queue :

$$\left|\int_x^\infty \frac{\{t\}}{t^2}\mathrm{d}t\right| \leq \int_x^\infty \frac{\mathrm{d}t}{t^2} = \frac{1}{x},$$

ce qui complète la preuve du théorème. □

Une application plus compliquée du théorème 7.4 est donnée par l'approximation de Stirling de la fonction factorielle :

Théorème 7.7 (Stirling's formula)**.** *Pour chaque entier $n \geq 1$, on a que*

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n}\left(1 + O\left(\frac{1}{n}\right)\right).$$

Démonstration. On considére le logarithme de $n!$ et on utilise la formule d'Euler-McLaurin pour le réécrire :

$$\log(n!) = \sum_{j=1}^{n} \log j = \int_1^n \log x \,\mathrm{d}x + \log 1 - \{n\}\log n + \int_1^n \frac{\{t\}}{t}\mathrm{d}t$$
$$= n\log n - n + 1 + \int_1^n \frac{\{t\}}{t}\mathrm{d}t,$$

puisque ici $n \in \mathbb{N}$ et, par conséquent, $\{n\} = 0$. Puis, on pose

$$F(x) = \int_0^x (\{t\} - 1/2)\mathrm{d}t.$$

Le fait que $\{t\} - 1/2$ est une fonction 1-périodique de moyenne 0 sur une période complète (i.e. $\int_0^1 (\{t\} - 1/2)\mathrm{d}t = 0$) implique que F est aussi 1-périodique. En particular, $F(n) = 0$ pour chaque $n \in \mathbb{N}$, et $F(x) = O(1)$ pour tout $x \geq 1$. En intégrant par parties, on trouve que

$$\int_1^n \frac{\{t\}}{t}\mathrm{d}t = \frac{\log n}{2} + \int_1^n \frac{\{t\} - 1/2}{t}\mathrm{d}t = \frac{\log n}{2} + \frac{F(t)}{t}\Big|_{t=1}^{n} + \int_1^n \frac{F(t)}{t^2}\mathrm{d}t$$
$$= \frac{\log n}{2} + \int_1^n \frac{F(t)}{t^2}\mathrm{d}t.$$

(Justifiez pourquoi on peut intégrer par parties, même si F n'est pas differentiable partout.) L'intégrale $\int_1^\infty F(t)t^{-2}\mathrm{d}t$ converge absolument car $F(t) = O(1)$. De plus, sa queue satisfait la borne

$$\int_n^\infty \frac{F(t)}{t^2}\mathrm{d}t \ll \int_n^\infty \frac{\mathrm{d}t}{t^2} = \frac{1}{n}.$$

Ceci montre que $\log(n!) = (n - 1/2)\log n - n + c + O(1/n)$ pour une certaine constante c. Puisque $e^{O(1/n)} = 1 + O(1/n)$ par le théorème de Taylor, la preuve sera complète à condition qu'on puisse montrer que $e^c = \sqrt{2\pi}$. L'argument est expliqué à l'exercise 7.8 ci-dessous. □

Exercices

EXERCICE 7.1. Considérez les fonctions suivantes :

$$f_1(x) = x^{1/\log\log x}, \quad f_2(x) = e^{\sqrt{\log x}}, \quad f_3(x) = x, \quad f_4(x) = (\log x)^A, \quad f_5(x) = \sqrt{x},$$
$$f_6(x) = e^x, \quad f_7(x) = \frac{x}{(\log x)^A}, \quad f_8(x) = \frac{x}{e^{\sqrt{\log x}}}, \quad f_9(x) = \log\log x,$$

où $A > 0$ est fixé mais arbitrairement grand. Ordonnez ces fonctions en termes de leur ordre de magnitude quand $x \to \infty$.

EXERCICE 7.2. Montrez les résultats asymptotiques suivants :

(a) $\log(1 + \delta) = \delta + O(\delta^2)$ pour $\delta \in [-1/2, 1/2]$;

(b) $\sqrt{x+1} = \sqrt{x} + O(1/\sqrt{x})$ pour $x \geq 1$;
(c) $e^{\delta} = 1 + O(\delta)$ pour$|\delta| \leq 1$;
(d) Si $p > 1$, alors $\sum_{n>x} 1/n^p \ll_p x^{1-p}$ pour $x \geq 1$;
(e) Soit $\lambda \in (0,1)$ et considérez une $\{a_n\}_{n=1}^{\infty} \subset \mathbb{R}_{\geq 0}$ telle que

$$a_{n+1} \leq \lambda a_n \quad (n \geq 1).$$

Donc

$$\sum_{n \geq N} a_n \asymp_{\lambda} a_N.$$

EXERCICE 7.3. (a) Prouvez que

$$\sum_{n \leq x} \sqrt{n} = \frac{2}{3} x^{3/2} + O(\sqrt{x}) \quad (x \geq 1)$$

(b) Prouvez qu'il existe une constante c telle que

$$\sum_{n \leq x} \frac{1}{\sqrt{n}} = 2\sqrt{x} + c + O\left(\frac{1}{\sqrt{x}}\right) \quad (x \geq 1).$$

EXERCICE 7.4. (a) Observez que si $f : [1, +\infty) \to [0, \infty)$ est une fonction décroissante, alors

$$\int_n^{n+1} f(t)\mathrm{d}t \leq f(n) \leq \int_{n-1}^{n} f(t)\mathrm{d}t,$$

et déduisez que

$$\int_{\lfloor x \rfloor + 1}^{\infty} f(t)\mathrm{d}t \leq \sum_{n>x} f(n) \leq \int_{\lfloor x \rfloor}^{\infty} f(t)\mathrm{d}t.$$

(b) Prouvez que

$$\frac{1}{\delta x^{\delta}} \leq \sum_{n>x} \frac{1}{n^{1+\delta}} \leq \frac{1}{\delta(x-1)^{\delta}} \quad (\delta > 0,\ x \geq 2),$$

et

$$\log x \leq \sum_{n \leq x} \frac{1}{n} \leq 1 + \log x \quad (x \geq 1).$$

EXERCICE 7.5. Si $(a_n)_{n \geq 1}$ est une suite de nombres complexes et $A(x) = \sum_{n \leq x} a_n$, alors montrez que

$$\sum_{n \leq x} a_n \log n = A(x) \log x - \int_1^x \frac{A(t)}{t}\mathrm{d}t.$$

Déduisez que

$$\sum_{n \leq x} \log n = x \log x + O(\log x) \quad (x \geq 1).$$

EXERCICE 7.6. Soit $(a_n)_{n=1}^{\infty}$ une suite de nombres complexes.

(a) Si

$$\lim_{x\to\infty}\frac{1}{x}\sum_{n\le x}a_n=\ell, \tag{7.6}$$

alors montrez que

$$\lim_{x\to\infty}\frac{1}{\log x}\sum_{n\le x}\frac{a_n}{n}=\ell. \tag{7.7}$$

(b) Construisez un exemple d'une suite de nombres $a_n\in[0,1]$ pour laquelle la limite dans (7.6) n'existe pas, mais la limite dans (7.7) existe.

EXERCICE 7.7. Cet exercise développe plus loin l'argument dans la démonstration du théorème 7.7.

On définit la suite des polynômes de Bernoulli $B_n(x)$ et des nombres de Bernoulli B_n comme suivant : on met $B_0(x)=B_0=1$, $B_1=-1/2$ et $B_1(x)=x+B_1$. Puis, on pose $B_2(x)=B_2+2\int_0^x B_1(x)\mathrm{d}x$, où B_2 est tel que $\int_0^1 B_2(x)\mathrm{d}x=0$, c'est-à-direh $B_2(x)=x^2-x+1/6$. Généralement, en supposant qu'on a défini $B_n(x)$, on pose $B_{n+1}(x)=B_{n+1}+(n+1)\int_0^x B_n(t)\mathrm{d}t$, où B_{n+1} est tel que $\int_0^1 B_{n+1}(x)\mathrm{d}x=0$.

(a) Pour $n\neq 1$, montrez que $B_n(1)=B_n(0)=B_n$. Concluez que la fonction $x\to B_n(\{x\})$ est 1-périodique et continue, et que $\int_0^x B_n(\{t\})\mathrm{d}t=(B_{n+1}(\{x\})-B_{n+1})/(n+1)$.

(b) Pour $n,k\in\mathbb{Z}_{\ge 1}$, montrez que

$$\log n!=(n+1/2)\log n-n+1-\sum_{\ell=2}^{k}\frac{B_\ell}{\ell(\ell-1)}(1-n^{1-\ell})+\int_1^n\frac{B_k(\{x\})}{kx^k}\mathrm{d}x.$$

(c) Généralement, pour si $a<b$ sont deux entiers, et f est une fonction lisse, alors montrez que

$$\sum_{a<n\le b}f(n)=\int_a^b f(x)\mathrm{d}x+\sum_{\ell=1}^{k}\frac{(-1)^\ell B_\ell}{\ell!}(f^{(\ell-1)}(b)-f^{(\ell-1)}(a))+(-1)^{k+1}\int_a^b\frac{B_k(\{x\})f^{(k)}(x)}{k!}\mathrm{d}x.$$

(d) Montrez que

$$\int_0^1 B_k(\{x\})e^{-2\pi imx}\mathrm{d}x=-\mathbf{1}_{m\neq 0}\frac{k!}{(2\pi im)^k}$$

et concluez que pour $k\ge 2$ on a

$$B_k(\{x\})=-\frac{k!}{(2\pi i)^k}\sum_{m\neq 0}\frac{e^{2\pi imx}}{m^k}.$$

(e) Pour $k\ge 1$, montrez que $B_{2k+1}=0$ et que

$$B_{2k}=\frac{(2k)!}{2^{2k-1}\pi^{2k}}\sum_{m=1}^{\infty}\frac{1}{m^{2k}}=\frac{(2k)!\zeta(2k)}{2^{2k-1}\pi^{2k}}.$$

(f) Prouvez que $B_n(x) = \sum_{k=0}^{n} \binom{n}{k} B_{n-k} x^k$ et déduisez que $B_n(x+1) = B_n(x) + nx^{n-1}$.
(g) Prouvez la formule récursive $B_n = -(n+1)^{-1} \sum_{k=2}^{n+1} \binom{n+1}{k} B_{n+1-k}$ pour $n \geq 1$ et en déduisez que $|B_n| \leq (4/5)^n n!$.
(h) Considérez la séries génératrice $F(z,x) = \sum_{n=0}^{\infty} B_n(x) z^n/n!$. Prouvez que $\partial F/\partial x = zF$, ainsi que $F(z,1) - F(z,0) = z$. Déduisez que $F(z,x) = e^{zx} z/(e^z - 1)$.
(i) Montrez que $z/(e^z - 1) + z/2$ est une fonction paire (i.e. ses valuers en z et $-z$ sont égales) et donnez une preuve différente du fait que $B_{2n+1} = 0$ pour $n \geq 1$.
(j) En notant que $z/(e^z - 1) = 1/(1 + \sum_{n=1}^{\infty} z^n/(n+1)!)$, donnez une formule explicite pour B_n.

EXERCICE 7.8. Complétez la démonstration du théorème 7.7 comme suivant :

(a) Montrez par induction que, pour chaque $n \in \mathbb{Z}_{\geq 0}$,

$$I_n := \int_0^{\pi/2} (\cos x)^n \mathrm{d}x = \begin{cases} \dfrac{\pi}{2} \cdot \dfrac{1 \cdot 3 \cdots (2k-1)}{2 \cdot 4 \cdots (2k)} = \dfrac{\pi}{2} \cdot \dfrac{\binom{2k}{k}}{4^k} & \text{si } n = 2k, \\ \dfrac{2 \cdot 4 \cdots (2k)}{1 \cdot 3 \cdots (2k+1)} = \dfrac{4^k}{(2k+1)\binom{2k}{k}} & \text{si } n = 2k+1, \end{cases}$$

(b) Pour chaque $\delta \in (0, \pi/2]$, posons

$$I_n(\delta) := \int_0^{\delta} (\cos x)^n \mathrm{d}x.$$

Montrez que

$$I_n - \frac{\pi}{2}(\cos\delta)^n \leq I_n(\delta) \leq I_n.$$

(c) Pour chaque $\delta \in (0, \pi/2]$, montrez que

$$(\cos\delta) \cdot \frac{I_n - \frac{\pi}{2}(\cos\delta)^n}{I_n} \leq \frac{I_{n+1}}{I_n} \leq 1.$$

Déduisez que

$$\lim_{n\to\infty} \frac{I_{n+1}}{I_n} = 1$$

et complétez la démonstration du théorème 7.7.

EXERCICE 7.9. Trouvez la valeur moyenne asymptotique du plus grand commun diviseur de a et b, quand a et b sont choisis de façon uniforme parmi tous les entiers $\leq x$. C'est-à-dire, calculez

$$\lim_{x\to\infty} \frac{1}{x^2} \sum_{a,b \leq x} (a,b).$$

Chapitre 8

La convolution de Dirichlet

8.1 La méthode de l'hyperbole

Une fonction arithmétique est une fonction de la forme $f : \mathbb{N} \to \mathbb{C}$ (c'est-à-dire une suite de nombres complexes). Par exemple, f pour être la fonction indicatrice des nombres premiers. Un autre exemple classique d'importance arithmétique est la fonction diviseur

$$\tau(n) := \#\{d|n\}.$$

On observe qu'on peut l'écrire comme

$$\tau(n) = \sum_{d|n} 1 = \sum_{ab=n} 1,$$

où tous les variables sont supposées d'être de nombres naturels. C'est un exemple d'une convolution de Dirichlet. Généralement, étant données deux fonctions arithmétiques f et g, leur convolution de Dirichlet est une nouvelle fonction arithmétique dénotée par $f * g$ et définie par

$$(f * g)(n) := \sum_{d|n} f(d)g(n/d) = \sum_{ab=n} f(a)g(b).$$

Donc, on a que

$$\tau = 1 * 1.$$

Un autre exemple important est donné en analysant la fonction logarithmique. Si $n = p_1^{v_1} \cdots p_r^{v_r}$ est la factorisation de n en facteurs premiers, alors

$$\begin{aligned} \log n &= v_1 \log p_1 + \cdots + v_r \log p_r \\ &= \underbrace{\log p_1 + \cdots + \log p_1}_{v_1 \text{ fois}} + \cdots + \underbrace{\log p_r + \cdots + \log p_r}_{v_r \text{ fois}} \\ &= \sum_{p^j | n} \log p. \end{aligned}$$

Donc, si on définit la fonction de von Mangoldt par

$$\Lambda(n) = \begin{cases} \log p & \text{si } n = p^j \text{ pour un premier } p \text{ et un entier } j \geq 1, \\ 0 & \text{sinon,} \end{cases}$$

alors

$$\log = \Lambda * 1. \tag{8.1}$$

Quand une fonction arithmétique f est la convolution de Dirichlet de deux autres fonctions g et h qui sont plus simples, on peut trouver une approximation de ses sommes partielles en utilisant ce qu'on sait concernant les sommes partielles de g et de h. On décrit l'idée en étudiant l'exemple concrète de la fonction diviseur. Puisque $\tau = 1 * 1$, alors

$$\sum_{n\le x}\tau(n) = \sum_{n\le x}(1*1)(n) = \sum_{n\le x}\sum_{ab=n} 1 = \sum_{ab\le x} 1.$$

Une manière de réarranger cette somme est de fixer a et sommer sur b (théorème de Fubini discrète). Ceci nous amène à la formule

$$\begin{aligned}\sum_{n\le x}\tau(n) = \sum_{a\le x}\sum_{b\le x/a} 1 = \sum_{a\le x}\Big(\frac{x}{a}+O(1)\Big) &= x\sum_{a\le x}\frac{1}{a} + O(x)\\ &= x\log x + O(x),\end{aligned} \tag{8.2}$$

où on a utilisé le théorème 7.6. Ceci est une formule asymptotique véritable, mais le terme d'erreur est juste un plus petit que le terme principal. On voudrait avoir une meilleure approximation. En réexaminant notre argument, on remarque que l'approximation $\sum_{b\le x/a} 1 = x/a + O(1)$ n'est par très bonne quand a est grand. Plutôt, quand a est grand, cela serait beaucoup mieux de changer les rôles de a et de b, en fixant b et en sommant premièrement sur a. De façon plus formelle, étant donnés deux paramètres $A, B \ge 1$ avec $AB = x$, on peut réarranger la sommation comme suivant :

$$\sum_{n\le x}\tau(n) = \sum_{ab\le x} 1 = \sum_{\substack{ab\le x\\ a\le A}} 1 + \sum_{\substack{ab\le x\\ a>A}} 1 = \sum_{a\le A}\sum_{b\le x/a} 1 + \sum_{b\le B}\sum_{A<a\le x/b} 1.$$

On observe que

$$\sum_{A<a\le x/b} 1 = \sum_{a\le x/b} 1 - \sum_{a\le A} 1.$$

Par la suite,

$$\begin{aligned}\sum_{n\le x}\tau(n) &= \sum_{a\le A}\sum_{b\le x/a} 1 + \sum_{b\le B}\sum_{a\le x/b} 1 - \Big(\sum_{a\le A} 1\Big)\Big(\sum_{b\le B} 1\Big)\\ &= \sum_{a\le A}\Big(\frac{x}{a}+O(1)\Big) + \sum_{b\le B}\Big(\frac{x}{b}+O(1)\Big) - (A+O(1))(B+O(1)).\end{aligned} \tag{8.3}$$

En appliquant deux fois le théorème 7.6 et en se rappelant que $AB = x$, on déduit que

$$\begin{aligned}\sum_{n\le x}\tau(n) &= x\left(\log(AB) + 2\gamma + O\left(\frac{1}{A}+\frac{1}{B}\right)\right) - AB + O(A+B)\\ &= x\log x + (2\gamma-1)x + O(A+B).\end{aligned}$$

Le choix optimal est $A = B = \sqrt{x}$, et il nous amène au résultat suivant :

Théorème 8.1. *Pour $x \geq 1$, on a que*

$$\sum_{n \leq x} \tau(n) = x \log x + (2\gamma - 1)x + O(\sqrt{x}).$$

Ce théorème est grâce à Dirichlet. La méthode de sa preuve est souvent appelée la méthode de l'hyperbole de Dirichlet. Son nom est justifié par une réévaluation géométrique de la démonstration : la somme $\sum_{n \leq x} \tau(n) = \sum_{ab \leq x} 1$ compte le nombre des points du réseau $\mathbb{N} \times \mathbb{N}$ sous l'hyperbole $ab = x$, c'est-à-dire la cardinalité de l'ensemble $\mathcal{N} = \{(a, b) \in \mathbb{N}^2 : ab \leq x\}$. La manière dont on a réarrangé la sommation correspond à la décomposition de $\mathcal{N}$ comme $X \cup Y$, où $X = \{(a, b) \in \mathbb{N}^2 : a \leq A\}$ and $Y = \{(a, b) \in \mathbb{N}^2 : b \leq B\}$. Par la principe de l'inclusion-exclusion, on a que

$$\sum_{n \leq x} \tau(n) = |X \cup Y| = |X| + |Y| - |X \cap Y|,$$

qui est la relation (8.3). La méthode de l'hyperbole est très utile et on va la réutiliser dans le prochain chapitre.

Exercices

EXERCICE 8.1. Soit $\mathcal{A}$ l'ensemble de toutes les fonctions arithmétiques, et soit $\mathcal{M}$ l'ensemble de toutes les fonctions multiplicatives.

(a) Montrez que la convolution de Dirichlet est une opération commutative : $f * g = g * f$ pour tous $f, g \in \mathcal{A}$.

(b) Montrez que la convolution de Dirichlet est une opération associative : $(f * g) * h = f * (g * h)$ pour tous $f, g, h \in \mathcal{A}$.

(c) Soit $\mathbf{e} \in \mathcal{A}$, définie par $\mathbf{e}(n) = \mathbf{1}_{n=1}$. Montrez que $\mathbf{e}$ est l'élément neutre de la convolution de Dirichlet, c'est-à-dire $f * \mathbf{e} = f$ pour chaque $f \in \mathcal{A}$.

(d) Soit $f \in \mathcal{A}$. On dit que la fonction arithmétique g est l'inverse de Dirichlet de g si $f * g = \mathbf{e}$. Montrez que g existe si et seulement si $f(1) \neq 0$. [*Indice :* construisez g de façon inductive, en mettant $g(n) = f(1)^{-1} \sum_{d|n,\, d>1} f(d)g(n/d)$.]

(e) Montrez que $(\mathcal{M}, *)$ est un groupe abélien.

(f) Soit $f \in \mathcal{M}$, et soit g son inverse de Dirichlet. Calculez $g(p)$ et $g(p^2)$ pour p premier en termes des valeurs de f.

EXERCICE 8.2. Soit μ la fonction de Möbius, définie par

$$\mu(n) := \begin{cases} (-1)^r & \text{si } n \text{ est sans-carré } [1] \text{ et a } r \text{ facteurs premiers,} \\ 0 & \text{sinon.} \end{cases}$$

(a) Montrez que μ est l'inverse de Dirichlet de 1.

(b) Prouvez que si $f = 1 * g$, alors $g = \mu * f$.

(c) Montrez que $\Lambda = \mu * \log$, ainsi que $\Lambda = -1 * \mu \log$.

(d) Montrez que si f est complètement multiplicative et g est l'inverse de Dirichlet de f, alors $g = \mu f$.
(e) Montrez que $\phi = I * \mu$, où $I(n) := n$. [*Indice* : utilisez la multiplicativité de ϕ.]
(f) Montrez que $\mu^2 = 1 * f$, où $f(n) = 0$ si n n'est pas un carré parfait, et $f(n) = \mu(m)$ si $n = m^2$. [*Indice* : utilisez la multiplicativité de μ^2.]
(g) Montrez que $2^\omega = 1 * \mu^2$, où $\omega(n) = \#\{p|n\}$. [*Indice* : multiplicativité.]

EXERCICE 8.3. Montrez que

$$\sum_{n\le x} \omega(n) = x \log\log x + O(x).$$

pour tout $x \ge 3$. [*Indice :* $\omega(n) = \sum_{p|n} 1$.]

EXERCICE 8.4. Définissons la k-ième fonction diviseur $\tau_k(n) := \#\{(d_1, \dots, d_k) \in \mathbb{N}^k : d_1 \cdots d_k = n\}$.
(a) Montrez que τ_k est multiplicative.
(b) Montrez qu'il existe un polynôme P_k de degrè $k-1$ tel que

$$\sum_{n\le x} \tau_k(n) = x \cdot P_k(\log x) + O_k(x^{1-1/k}) \quad (x \ge 1).$$

Calculez le coefficient Calculate the leading coefficient de tête de P_k.

EXERCICE 8.5. Pour chaque $k \in \mathbb{N}$, on définit la k-ième fonction généralisée de von Mangoldt function par $\Lambda^{(k)} = \mu * \log^k$. Prouvez que :
(a) $\Lambda^{(k+1)} = \Lambda^{(k)} \log + \Lambda^{(k)} * \Lambda$;
(b) Λ_k est supporté sur les entiers avec $\le k$ facteurs premiers distincts ;
(c) Si $n = p_1 \cdots p_k$, où $p_1, \dots, p_k$ sont de nombres premiers distincts, alors $\Lambda^{(k)}(n) = (\log p_1) \cdots (\log p_k)$.
(d) $0 \le \Lambda^{(k)}(n) \le (\log n)^k$ pour chaque $n \in \mathbb{N}$;

EXERCICE 8.6. Montrez qu'il existe une constante $c > 0$ telle que

$$\#\{n \le x : n \text{ is square-free}\} = cx + O(\sqrt{x}) \quad (x \ge 1).$$

[*Indice :* Utilisez l'exercice 8.2(e) et la version triviale de la méthode de l'hyperbole (cf. relation (8.2))]

EXERCICE 8.7. Un nombre n est appelé plein de carrés si $p^2|n$ pour chaque facteur premier p de n.
(a) Montrez que n est plein de carrés si et seulement s'il peut s'écrire comme $n = a^2b^3$ pour quelques $a, b \in \mathbb{N}$.
(b) Prouvez que

$$\#\{n \le x : n \text{ est plein de carrés}\} \asymp \sqrt{x} \quad (x \ge 1).$$

EXERCICE 8.8. (a) Si f dénote la fonction indicatrice de la suite des nombres pleins de carrés, alors montrez que $f(n) = \sum_{a^2b^3=n} \mu^2(b)$.

(b) Montrez qu'il existe deux constantes $c_1, c_2 \in \mathbb{R}$ telles que

$$\#\{n \leq x : n \text{ est plein de carrés}\} = c_1 x^{1/2} + c_2 x^{1/3} + O(x^{1/5}).$$

8.2 Les théorèmes de Chebyshev et de Mertens

On utilise maintenant la méthode de l'hyperbole pour étudier la répartition des nombres premiers. Notre point de départ est la relation (8.1). Donc, au lieu de donner une approximation à la fonction $\pi(x)$, on introduit les fonctions de Chebyshev

$$\psi(x) := \sum_{n \leq x} \Lambda(n) = \sum_{p^k \leq x} \log p$$

et

$$\theta(x) := \sum_{p \leq x} \log p.$$

Comme on va montrer tout de suite, une approximation pour $\pi(x)$ peut se transformer facilement à une estimation pour $\theta(x)$ et $\psi(x)$. Le converse est aussi vrai.

Pour voir notre affirmation, on commence en observant que $\theta(x) = \sum_{n \leq x} a_n \log n$, où $a_n = \mathbf{1}_{n \text{ est premier}}$, et on a enlevé le terme avec $n = 1$ car $a_1 = 0$. En appliquant la formule (7.4) avec $f(n) = \log n$, on trouve alors que

$$\theta(x) = \pi(x) \log x - \int_1^x \frac{\pi(t)}{t} \mathrm{d}t. \tag{8.4}$$

Réciproquement, on a que

$$\pi(x) = \sum_{2-\varepsilon < n \leq x} \frac{\tilde{a}_n}{\log n}$$

pour chaque $\varepsilon > 0$, où $\tilde{a}_n = \mathbf{1}_{n \text{ est premier}} \log n$. En appliquant 7.4 avec $f(n) = 1/\log n$, on obtient la relation

$$\pi(x) = \frac{\theta(x)}{\log x} + \int_{2-\varepsilon}^x \frac{\theta(t)}{t \log^2 t} \mathrm{d}t$$

pour chaque $\varepsilon > 0$. On laisse $\varepsilon \to 0^+$ afin de déduire que

$$\pi(x) = \frac{\theta(x)}{\log x} + \int_2^x \frac{\theta(t)}{t \log^2 t} \mathrm{d}t. \tag{8.5}$$

Donc, on voit qu'un estimation approximative pour $\pi(x)$ peut se transformer tout de suite à une approximation pour $\theta(x)$, et vice versa.

De plus, on peut lier $\theta(x)$ et $\psi(x)$: on a que

$$\begin{aligned} 0 \le \psi(x) - \theta(x) = \sum_{p^k \le x,\, k \ge 2} \log p &= \sum_{p \le \sqrt{x}} \log p \sum_{2 \le k \le \log x / \log p} \\ &\le \sum_{p \le \sqrt{x}} \log p \cdot \frac{\log x}{\log p} \\ &= \sum_{p \le \sqrt{x}} \log x \le \sqrt{x} \log x. \end{aligned} \tag{8.6}$$

On attend à ce que $\psi(x)$ soit $\sim x$, et la fonction $\sqrt{x}\log x$ est minuscule en comparaison avec la fonction x.

Comme une application des relations (8.4), (8.5) et (8.6), on invite le lecteur de vérifier que les relations suivantes sont équivalentes quand $x \to \infty$:

$$\pi(x) \sim \frac{x}{\log x}, \quad \theta(x) \sim x, \quad \psi(x) \sim x.$$

Le premier théorème qu'on montrera concernant la repartition des premiers est que $\pi(x) \asymp x/\log x$, c'est-à-dire l'ordre de magnitude de $\pi(x)$ est celle prévue par la conjecture de Gauss :

Théorème 8.2 (Chebyshev)**.** *Pour $x \ge 2$, on a que*

$$\pi(x) \asymp \frac{x}{\log x}, \quad \theta(x) \asymp x, \quad \textit{et} \quad \psi(x) \asymp x.$$

Démonstration. Selon la discussion précédente, il suffit de montrer que $\psi(x) \asymp x$; les autres résultats suivent tout de suite par (8.6) et (8.5).

Vu que $\log = \Lambda * 1$, on a que

$$S(x) := \sum_{n \le x} \log n = \sum_{n \le x} (\Lambda * 1)(n) = \sum_{n \le x} \sum_{ab = n} \Lambda(a) = \sum_{ab \le x} \Lambda(a).$$

On réarrange la sommation en fixant b et en sommant d'abord sur a. On trouve alors que

$$\sum_{n \le x} \log n = \sum_{b \le x} \sum_{a \le x/b} \Lambda(b) = \sum_{b \le x} \psi(x/b).$$

On peut donner une estimation asymptotique du côté gauche par la formule d'Euler-McLaurin. Par exemple, l'exercice 7.5 implique que

$$S(x) = \sum_{n \le x} \log n = x \log x - x + O(\log x).$$

On considère maintenant l'expression $S(x) - 2S(x/2)$. D'un côté, on a que

$$S(x) - 2S(x/2) = x \log x - x \log(x/2) + O(\log x) = x \log 2 + O(\log x).$$

D'autre côté, on a que

$$S(x) - 2S(x/2) = \sum_{b \le x} \psi(x/b) - \sum_{2b \le x} \psi(x/2b) = \sum_{n \le x} (-1)^{n-1} \psi(x/n).$$

La suite $n \to \psi(x/n)$ est décroissante, donc par la théorie des séries alternées, on a que

$$\psi(x) - \psi(x/2) \le S(x) - 2S(x/2) \le \psi(x).$$

On trouve alors que $\psi(x) \ge x \log 2 - O(\log x) \gg x$ pour x grand. Pour x petit, on observe simplement que $\psi(x) \ge \psi(2) = \log 2$.

Finalement, puisque

$$\psi(x) - \psi(x/2) \le x \log 2 + O(\log x) \le cx \quad (x \ge 1)$$

pour une constante absolue $c > \log 2$, on déduit que

$$\psi(x) = \sum_{1 \le 2^j \le x} [\psi(x/2^j) - \psi(x/2^{j+1})] \le \sum_{1 \le 2^j \le x} \frac{cx}{2^j} \le 2cx.$$

Ceci conclut la démonstration. □

Même si la méthode de démonstration du théorème 8.2 ne suffit pas pour obtenir la formule asymptotique $\pi(x) \sim x/\log x$, on peut quand même donner des approximations pour quelques sommes de la forme $\sum_{p \le x} f(p)$, où f une fonction lisse que décroit rapidement comme $f(p) = \log p/p$ et $f(p) = 1/p$. L'estimation de telles sommes est, en général, plus facile car la location exacte d'un nombre premier donné est beaucoup moins importante (son poids est très petit). Voir aussi l'exercice 7.6.

Théorème 8.3. *(a) Pour chaque $x \ge 1$, on a que*

$$\sum_{p \le x} \frac{\log p}{p} = \log x + O(1).$$

(b) Il existe une constante $c \in \mathbb{R}$ telle que

$$\sum_{p \le x} \frac{1}{p} = \log\log x + c + O\left(\frac{1}{\log x}\right) \quad (x \ge 2).$$

(c) Pour chaque $x \ge 2$, on a que

$$\prod_{p \le x} \left(1 - \frac{1}{p}\right) = \frac{e^{-\gamma}}{\log x} \left(1 + O\left(\frac{1}{\log x}\right)\right).$$

Démonstration. (a) Le point de départ est encore l'identité $\log = \Lambda * 1$, d'où on trouve que

$$\sum_{n \le x} \log n = \sum_{ab \le x} \Lambda(a).$$

Contrairement à la démonstration du théorème 8.2, maintenant on arrange la sommation comme

$$\sum_{n\le x}\log n=\sum_{a\le x}\Lambda(a)\sum_{b\le x/a}1.$$

La somme interne est égale à $x/a+O(1)$. Par conséquent,

$$\sum_{n\le x}\log n=\sum_{a\le x}\Lambda(a)\cdot(x/a+O(1))=x\sum_{a\le x}\frac{\Lambda(a)}{a}+O(\psi(x))=x\sum_{a\le x}\frac{\Lambda(a)}{a}+O(x),$$

où on a utilisé le théorème 8.2.

D'autre côté, comme on l'a vu ci-dessus, on a que $\sum_{n\le x}\log n=x\log x-x+O(\log x)$. Donc,

$$\sum_{a\le x}\frac{\Lambda(a)}{a}=\log x+O(1).$$

Puisque $\sum_{a=p^\nu,\,\nu\ge 2}\Lambda(a)/a=O(1)$, la partie (a) du théorème découle tout de suite.

(b) Le résultat suit de la partie (a) et de sommation partielle. Plus précisément, si on écrit

$$\sum_{p\le x}\frac{\log p}{p}=\log x+R(x),$$

où $R(x)=O(1)$, alors la relation (7.5) avec $a_n=\mathbf{1}_{n\text{ est premier}}\frac{\log n}{n}$, $f(n)=1/\log n$, et $M(x)=\log x$ implique que

$$\begin{aligned}\sum_{p\le x}\frac{1}{p}=\lim_{\varepsilon\to 0^+}\sum_{2-\varepsilon<n\le x}a_nf(n)&=\sum_{2^-<n\le x}a_nf(n)\\&=\int_2^x\frac{1}{t\log t}\mathrm{d}t+\frac{R(x)}{\log x}-\frac{R(2^-)}{\log 2}+\int_2^x\frac{R(t)}{t\log^2 t}\mathrm{d}t\\&=\log\log x-\log\log 2+\frac{R(x)}{\log x}-\frac{R(2^-)}{\log 2}+\int_2^x\frac{R(t)}{t\log^2 t}\mathrm{d}t.\end{aligned}$$

On observe que $R(2^-)=\log 2$ et que l'intégrale $\int_2^\infty R(t)/(t\log^2 t)\mathrm{d}t$ converge absolument car $R(t)\ll 1$. Donc, si on pose $c=-\log\log 2-1+\int_2^\infty R(t)/(t\log^2 t)\mathrm{d}t$, on trouve que

$$\sum_{p\le x}\frac{1}{p}=\log\log x+c+O\left(\frac{1}{\log x}+\int_x^\infty\frac{\mathrm{d}t}{t\log^2 t}\right)=\log\log x+c+O\left(\frac{1}{\log x}\right),$$

comme affirmé.

(c) On a que

$$\varepsilon_p:=\log\left(1-\frac{1}{p}\right)+\frac{1}{p}=-\sum_{m=2}^\infty\frac{1}{mp^m}$$

d'après le développement de Taylor autour de 1 de la fonction logarithmique. Puisque

$$\sum_{m=2}^\infty\frac{1}{mp^2}\le\sum_{m=2}^\infty\frac{1}{2p^m}=\frac{1}{2p(p-1)}\le\frac{1}{p^2},$$

on déduit que $\varepsilon_p = O(1/p^2)$. Par la suite,

$$\log \prod_{p\le x}\left(1-\frac{1}{p}\right) = -\sum_{p\le x}\frac{1}{p} + \sum_{p\le x}\varepsilon_p = -\log\log x + c + \sum_p \varepsilon_p + O\left(\frac{1}{\log x} + \sum_{p>x}|\varepsilon_p|\right)$$

$$= -\log\log x + \kappa + O\left(\frac{1}{\log x}\right), \tag{8.7}$$

où $\kappa := c + \sum_p \varepsilon_p$. Il reste à montrer que $\kappa = -\gamma$. Ceci est prouvé en utilisant des informations concernant le comportement analytique de la fonction zeta de Riemann autour de 1 (voir exercice **??**). □

Exercices

EXERCICE 8.9. Complete the proof of Theorem 8.2 by establishing the lower bound as follows :

(a) Show that

$$\nu_p(n!) = \sum_{j\ge 0}\left\lfloor \frac{n}{p^j}\right\rfloor,$$

where $\nu_p(m)$ is the p-adic valuation of m, that is to say the highest power of p that divides m.

(b) Show that $\lfloor 2x\rfloor - 2\lfloor x\rfloor$ is a 1-periodic function taking only the values 0 and 1.

(c) Conclude that

$$\binom{2n}{n} \le (2n)^{\pi(2n)}$$

and complete the proof of Theorem 8.2.

EXERCICE 8.10. This exercise provides an alternative way to establish the lower bound in Theorem 8.2, due to Nair. Consider the integral $I_n = \int_0^1 x^n(1-x)^n \mathrm{d}x$ and the integer $N = \mathrm{lcm}[n+1, n+2, \dots, 2n]$.

(a) Prove that $I_n \cdot N$ is a non-negative integer.

(b) Prove that $I_n \le 4^{-n}$.

(c) Prove that $N \le (2n)^{\pi(2n)}$ and complete the proof of Theorem 8.2.

EXERCICE 8.11. Prove asymptotics for the sums

$$\sum_{p\le x}\frac{\log^k p}{p} \quad \text{and} \quad \sum_{p>x}\frac{1}{p^2}$$

EXERCICE 8.12. Prove that the asymptotic $\pi(x) \sim x/\log x$ is equivalent to having that

$$\sum_{p\le x}\frac{\log p}{p} = \log x + c + o(1) \quad (x\to\infty),$$

for some appropriate constant c.

EXERCICE 8.13. Let $p_1 < p_2 < \cdots$ denote the sequence of primes, and let $P_k = p_1 p_2 \cdots p_k$.

(a) For $k \geq 2$, show that $p_k \asymp k \log k$ and that $\log P_k \asymp k \log k$.

(b) Assuming the Prime Number Theorem, show that $p_k \sim k \log k$ and $\log P_k \sim k \log k$ as $k \to \infty$.

(c) Show that $\omega(n) \ll \log n / \log\log n$ for all n. [*Hint* : what can you say about $\omega(n)$ if $n \leq P_k$?]

(d) Show that

$$\frac{\phi(n)}{n} \sim \prod_{\substack{p \mid n \\ p \leq \log n}} \left(1 - \frac{1}{p}\right) \geq \prod_{p \leq \log n} \left(1 - \frac{1}{p}\right) \quad (n \to \infty).$$

Chapitre 9

La fonction zeta de Riemann

Les approches qu'on a vu dans le dernier chapitre pour compter les nombres premiers s'appellent 'élémentaires' car ils utilisent des méthodes simples venant de la théorie des nombres et de l'analyse réelle. Cependant, la clé pour débarrer les sécrètes des nombres premiers se trouve au plan complexe.

Souvent, afin de comprendre une suite a_n, on forme une série génératrice. L'exemple le plus familier est la série de puissances de a_n, définie par

$$S(z) = \sum_{n=1}^{\infty} a_n z^n.$$

Si $|a_n| \leq 1$, alors cette série converge vers une fonction holomorphe dans le disque $|z| < 1$. De plus, si on peut évaluer $S(z)$ (de façon exacte ou approximative), on peut aussi évaluer les coefficients a_n grâce à la formule d'inversion de Fourier

$$a_n = \frac{r^{-n}}{2\pi} \int_0^{2\pi} S(re^{ix}) e^{-inx} \mathrm{d}x \tag{9.1}$$

pour chaque $r \in (0, 1)$. En effet, on a que

$$\begin{aligned} \int_0^{2\pi} S(re^{2\pi ix}) e^{-2\pi inx} \mathrm{d}x &= \int_0^{2\pi} \sum_{m=1}^{\infty} a_m r^m e^{i(m-n)x} \mathrm{d}x \\ &= \sum_{m=1}^{\infty} a_m r^m \int_0^{2\pi} e^{i(m-n)x} \mathrm{d}x \\ &= 2\pi a_n r^n, \end{aligned}$$

car l'intégrale $\int_0^{2\pi} e^{i(m-n)x} \mathrm{d}x$ vaut zéro, sauf si $m = n$, dans quel cas il vaut 2π.

On va essayer d'appliquer l'approche décrite ci-dessus pour étudier les nombres premiers. Soit a_n la fonction indicatrice des premiers, dont la série de puissances est

$$Q(z) := \sum_p z^p.$$

En ajoutant (9.1) quand $S(z) = Q(z)$, on déduit la formule d'inversion

$$\pi(N) = \sum_{n=1}^{N} \frac{r^{-n}}{2\pi} \int_0^{2\pi} Q(re^{ix}) e^{-inx} \mathrm{d}x = \frac{1}{2\pi} \int_0^{2\pi} S(re^{ix}) \frac{1 - (re^{ix})^{-N}}{re^{ix} - 1} \mathrm{d}x \tag{9.2}$$

pour chaque $N \in \mathbb{Z}_{\geq 1}$. On vois donc qu'une compréhension des propriétés analytiques de $Q(z)$ peut nous amener à une estimation précise de $\pi(N)$.

Cette approche arrive rapidement à un cul-de-sac, car ce n'est pas clair comment on peut étudier la fonction $Q(z)$ sans une connaissance profonde des premiers. En fait, la même objection peut être soulevée pour n'importe quelle fonction génératrice de la suite des premiers : comment est-ce que c'est possible de déterminer son comportement asymptotique sans avoir déjà une bonne compréhension de la répartition des premiers ? Afin de répondre à cette question, on revient à la fonction $Q(z)$ on l'analyse. Cette fonction est naturellement liée à la structure additive des premiers. Par exemple, on note que

$$Q(z)^k = \sum_{p_1, \ldots, p_k} z^{p_1 + \cdots + p_k} = \sum_{n=0}^{\infty} g_k(n) z^n,$$

où $g_k(n)$ est le nombre de façon d'écrire n comme la somme de k nombres premiers. Cependant, les premiers sont d'objets multiplicatif, donc c'est plus naturel de les étudier d'un tel point de vue. Pour le faire, on observe que la fonction logarithmique est un isomorphisme de groupes de $(\mathbb{R}_{>0}, \times)$ à $(\mathbb{R}, +)$. Ceci nous amène naturellement à la considération de la fonction génératrice

$$\sum_p z^{\log p} = \sum_p p^{\log z}.$$

Ceci n'est plus une série de puissances, car les exposants ne sont pas d'entiers. En fait, c'est convenable de faire le changement de variables $s = -\log z$, pour que notre fonction génératrice devient

$$P(s) := \sum_p \frac{1}{p^s}.$$

En suivant la notation de Riemann, on écrit toujours

$$s = \sigma + it.$$

Cette nouvelle fonction génératrice est bien définie pour $\sigma > 1$: on a que

$$|n^s| = |n^\sigma| \cdot |n^{it}| = n^\sigma \cdot |e^{it \log n}| = n^\sigma$$

et on sait que $\sum_n 1/n^\sigma$ converge si $\sigma > 1$. Donc $P(s)$ converge absolument si $\sigma > 1$, ce qui montre notre affirmation.

Pour la k-ième puissance de P, on a que

$$P(s)^k = \sum_{p_1} \frac{1}{p_1^s} \cdots \sum_{p_k} \frac{1}{p_k^s} = \sum_{p_1, \ldots, p_k} \frac{1}{(p_1 \cdots p_k)^s} = \sum_{n=1}^{\infty} \frac{h_k(n)}{n^s},$$

où $h_k(n)$ dénote le nombre de manières d'écrire n comme le produit de k premiers. En particulier, h_k est supportée sur les entiers ayant $\leq k$ facteurs premiers. En comparaison, avant on n'avait pas de contrôle sur le support de g_k. On voit alors tout de suite que $P(s)$ se comporte de meilleure façon que $Q(z)$.

En poussant l'argument précédent, Euler a prouvé que la fonction P peut s'écrire en termes de la fonction zeta de Riemann

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s},$$

convergeant absolument aussi pour $\sigma > 1$. On observe que la fonction ζ est pour la suite des nombres naturels ce qui la fonction P est pour la suite des nombres premiers. L'importance de ζ dans l'étude des premiers se manifeste à l'identité suivante, connue comme le produit d'Euler de ζ :

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}. \tag{9.3}$$

Pour la montrer, on observe que

$$\begin{aligned}\prod_p \left(1 - \frac{1}{p^s}\right)^{-1} &= \prod_p \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots\right) \\ &= \left(\frac{1}{2^{0s}} + \frac{1}{2^s} + \frac{1}{2^{2s}} + \cdots\right)\left(\frac{1}{3^{0s}} + \frac{1}{3^s} + \frac{1}{3^{2s}} + \cdots\right)\left(\frac{1}{5^{0s}} + \frac{1}{5^s} + \frac{1}{5^{2s}} + \cdots\right)\cdots\end{aligned}$$

Si on développe le produit infini, on trouve une série dont les sommés sont de la forme $1/n^s$, où n est chaque produit possible de la forme $2^{v_2}3^{v_3}5^{v_5}7^{v_7}\cdots$ avec $v_p \in \mathbb{Z}_{\geq 0}$, et où les exposants sont presque tous 0. Le théorème fondamental de l'arithmétique implique que les n construits sont exactement tous les nombres naturels, d'où l'identité (9.3) suit.

En fait, l'argument ci-dessus a quelques trous, parce qu'on a ignoré plusieurs problèmes de convergence. Plus rigoureusement, si $\sigma > 1$ on sait que la série $\sum_{n\geq 1} 1/n^s$ converge absolument, donc on peut réarranger ses termes comme on veut. On a alors que

$$\zeta(s) = \lim_{M\to\infty}\lim_{N\to\infty} \sum_{p^\nu \| n \,\Rightarrow\, p\leq M, \nu \leq N} \frac{1}{n^s} = \lim_{M\to\infty}\lim_{N\to\infty} \prod_{p\leq M}\left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots + \frac{1}{p^{Ns}}\right).$$

L'identité (9.3) est donc établie.

En prenant le logarithme de chaque côté de (9.3), on déduit que

$$\log\zeta(s) = \sum_p \sum_{\nu=1}^{\infty} \frac{1}{\nu p^{\nu s}} = \sum_{\nu=1}^{\infty} \frac{P(\nu s)}{\nu} = P(s) + \sum_{\nu=2}^{\infty} \frac{P(\nu s)}{\nu}, \tag{9.4}$$

qui fournit le lien entre P et ζ. Cette formule est le point de départ de la théorie analytique des nombres, puisque il connecte la fonction P, pour laquelle on ne savait rien, à la fonction ζ. Cette dernière est significativement plus simple, puisque elle est définie comme une

sommation sur tous les nombres naturels, qui sont un ensemble très régulier. Il semble alors possible qu'on puisse obtenir des bonnes estimations de $P(s)$.

Comme dans le cas de $Q(z)$, on peut passer par $P(s)$ à $\pi(x)$. L'inversion se faite en utilisant la transformation de Fourier sur $\mathbb{R}$ et elle est donnée par l'identité

$$\frac{\pi(x^+)+\pi(x^-)}{2} = \frac{1}{2\pi}\int_{-\infty}^{\infty} P(\sigma+it)\frac{x^{\sigma+it}}{\sigma+it}\mathrm{d}t, \tag{9.5}$$

valide pour chaque $\sigma > 1$ et chaque $x \geq 1$. (Comme la valuer de l'intégrale impropre, on prend sa *valeur principale*, c'est-à-dire

$$\int_{-\infty}^{\infty} f(t)\mathrm{d}t = \lim_{T\to\infty}\int_{-T}^{T} f(t)\mathrm{d}t.$$

Un calcul simple implique que le côté gauche de (9.5) est égal à $\pi(x) - \mathbf{1}_{x \text{ is prime}}/2$.

En pratique, c'est un peu difficile de travailler avec $P(s)$. On considère alors la série

$$\sum_{n=1}^{\infty}\frac{\Lambda(n)}{n^s} = \sum_{p}\sum_{m=1}^{\infty}\frac{\log p}{p^{ms}}.$$

En dérivant la première égalité de (9.4), on trouve que

$$\sum_{n=1}^{\infty}\frac{\Lambda(n)}{n^s} = -\frac{\zeta'}{\zeta}(s).$$

La formule analogue de (9.5) correspondant à cette série est

$$\frac{\psi(x^+)+\psi(x^-)}{2} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(-\frac{\zeta'}{\zeta}\right)(\sigma+it)\frac{x^{\sigma+it}}{\sigma+it}\mathrm{d}t, \tag{9.6}$$

valide pour chaque $\sigma > 1$ et chaque $x \geq 1$.

Motivée par l'importance de ζ dans l'étude des nombres premiers, on l'étudie de plus. En utilisant la formule d'Euler-McLaurin avec $f(n) = 1/n^s$, on trouve que

$$\begin{aligned}
\zeta(s) &= 1 + \lim_{N\to\infty}\sum_{1<n\leq N}\frac{1}{n^s} \\
&= 1 + \lim_{N\to\infty}\left(\int_1^N \frac{\mathrm{d}y}{y^s} - \frac{\{y\}}{y^s}\Big|_{y=1}^{N} - s\int_1^N \frac{\{y\}}{y^{s+1}}\mathrm{d}y\right) \\
&= 1 + \lim_{N\to\infty}\left(\frac{1-N^{-s+1}}{s-1}\int_1^N \frac{\mathrm{d}y}{y^s} - s\int_1^N \frac{\{y\}}{y^{s+1}}\mathrm{d}y\right) \\
&= \frac{s}{s-1} - s\int_1^{\infty}\frac{\{y\}}{y^{s+1}}\mathrm{d}y. \qquad (9.7)
\end{aligned}$$

L'expression dans (9.7) est bien définie quand $\sigma > 0$: on a que $|\{y\}/y^{s+1}| \leq 1/|y^{s+1}| = 1/y^{\sigma+1}$, et on sait que l'intégrale $\int_1^{\infty} 1/y^p \mathrm{d}y$ converge quand $p > 1$. Donc la formule (9.7)

nous donne une façon de prolonger la définition de ζ pour tous les nombres complexes avec $\sigma > 0$. En fait, on voit qu'on ne peut pas la définir en 1 (i.e. elle 'explode' vers l'infini à ce point), car il faut diviser par $1 - 1 = 0$ à ce point. La fonction ζ a une singularité en 1, ce qui est une conséquence de la divergence de la série harmonique $\sum_{n=1}^{\infty} 1/n$. Cette singularité est très importante à l'étude de $\psi(x)$ et de $\pi(x)$ et elle en accord avec la conjecture de Gauss (pour les connaisseurs d'analyse complexe, l'intégrante $(-\zeta'/\zeta)(s)x^s/s$ dans (9.6) a une singularité de résidu x en 1).

Comme notre dernier remarque sur ζ, on observe que si $\zeta(\rho) = 0$ dans un certain point ρ, la fonction ζ'/ζ a une singularité en ρ. Par sa représentation (9.3), il suit que $\zeta(s) \neq 0$ quand $\sigma > 1$. On peut aussi montrer que les seuls zéros de ζ avec $\sigma < 0$ sont aux points $-2, -4, -6, \ldots$. Tous les autres zéros se trouvent dans la bande $\{s \in \mathbb{C} : 0 \leq \sigma \leq 1\}$, qu'on appelle la bande criticale. Riemann a conjecturé que tous les zéros de ζ dans cette bande se trouve sur la ligne centrale $\sigma = 1/2$. Ceci est la fameuse hypothèse de Riemann qui est encore ouverte et un des plus importants problèmes en mathématiques. Pour montrer le théorème des nombres premiers, de la Vallée-Poussin et Hadamard on montré que $\zeta(s) \neq 0$ quand $\sigma = 1$. Après, de la Vallée-Poussin a amélioré ce résultat en prouvant qu'il existe une constante absolue $c_1 > 0$ telle que

$$\zeta(s) \neq 0 \quad \text{quand} \quad \sigma > 1 - \frac{c_1}{\log(2 + |t|)}. \tag{9.8}$$

Ceci a amené à une forme quantitative du théorème des nombres premiers :

$$\pi(x) = \mathrm{li}(x) + O(xe^{-c_2\sqrt{\log x}}) \quad (x \geq 2), \tag{9.9}$$

où $c_2 > 0$ est une autre constante absolue.

On peut améliorer la qualité de l'estimation (9.11) si on agrandit la région san zéros (9.10). Le record à ce moment-ci, montré par Korobov et Vinogradov en 1958, est que

$$\zeta(s) \neq 0 \quad \text{quand} \quad \sigma > 1 - \frac{c_3}{\log^{2/3}(2 + |t|)(\log\log(3 + |t|))^{1/3}}. \tag{9.10}$$

Ceci a amené à une forme quantitative du théorème des nombres premiers :

$$\pi(x) = \mathrm{li}(x) + O(xe^{-c_4(\log x)^{3/5}/(\log\log x)^{1/5}}) \quad (x \geq 2), \tag{9.11}$$

où c_3 et c_4 sont deux autres constantes absolues.

Exercices

Chapitre 10

Le théorème d'Erdős-Kac

Plusieurs problèmes dans la théorie analytique des nombres peuvent être rapprochés par un perspectif probabiliste. Entre les plus beaux tels résultats se trouvent le théorème d'Erdős-Kac qui concerne la distribution des valeurs de la fonction

$$\omega(n) = \#\{p|n\}.$$

On interprète ω comme une variable aléatoire sur l'ensemble $\{n \leq x\}$, muni du mesure de probabilité uniforme, c'est-à-dire

$$\mathbb{P}_{n\leq x}(A) := \frac{|A|}{\lfloor x \rfloor}.$$

Notre but est de donner une estimation asymptotique à la probabilité

$$\mathbb{P}_{n\leq x}(\omega(n) \in I)$$

quand $x \to \infty$, où I est un intervalle (qui peut dépendre de x).

Afin d'approcher ce problème, on observe que si $n \leq x$, alors

$$\omega(n) = \sum_{p\leq x} \mathbf{1}_{p|n}.$$

Les fonctions $B_p(n) := \mathbf{1}_{p|n}$ sont des variables aléatoires de Bernoulli prenant les valeurs 0 et 1. On a que

$$\mathbb{P}_{n\leq x}(B_p = 1) = \frac{\#\{n \leq x : p|n\}}{\lfloor x \rfloor}.$$

En posant $n = mp$, on trouve que $\#\{n \leq x : p|n\} = \#\{m \leq x/p\} = \lfloor x/p \rfloor$. Donc

$$\mathbb{P}_{n\leq x}(B_p = 1) = \frac{\lfloor x/p \rfloor}{\lfloor x \rfloor} = \frac{x/p + O(1)}{x + O(1)} = \frac{1}{p} + O\left(\frac{1}{x}\right).$$

On a alors que

$$\mathbb{E}_{n\leq x}[B_p(n)] = 0 \cdot \mathbb{P}_{n\leq x}(B_p = 0) + 1 \cdot \mathbb{P}_{n\leq x}(B_p = 1) = \frac{1}{p} + O\left(\frac{1}{x}\right) \sim_{x\to\infty} \frac{1}{p},$$

et que

$$\mathrm{Var}_{n\le x}[B_p(n)] = \mathbb{E}_{n\le x}[B_p(n)^2] - \mathbb{E}_{n\le x}[B_p(n)]^2 = \frac{1}{p} - \frac{1}{p^2} + O\left(\frac{1}{x}\right) \sim_{x\to\infty} \frac{p-1}{p^2}.$$

On étudie maintenant les correlations entre les différentes variables B_p : si $p_1, \dots, p_m$ sont de nombres premiers distincts, alors

(10.1)

$$\begin{aligned}\mathbb{P}_{n\le x}(B_{p_1}(n) = \cdots = B_{p_m}(n) = 1) = \frac{\#\{n \le x : p_1, \dots, p_m p_m | n\}}{\lfloor x \rfloor} &= \frac{\#\{n \le x : p_1 \cdots p_m | n\}}{\lfloor x \rfloor} \\ &= \frac{\lfloor x/(p_1 \cdots p_m)\rfloor}{\lfloor x \rfloor} \\ &= \frac{1}{p_1 \cdots p_m} + O\left(\frac{1}{x}\right),\end{aligned}$$

d'où

$$\mathbb{P}_{n\le x}(B_{p_1}(n) = \cdots = B_{p_m}(n) = 1) \sim \prod_{j=1}^{m} \mathbb{P}_{n\le x}(B_{p_j}(n) = 1) \quad (x \to \infty).$$

On est amené alors à la conclusion que les variables aléatoires sont approximativement indépendantes l'une à l'autre. Sil elles étaient vraiment indépendantes, le théorème central limite impliquerait que leur somme $\omega(n) = \sum_{p\le x} B_p(n)$ tend vers la répartition normale avec moyenne

$$\sum_{p\le x} \mathbb{E}_{n\le x}[B_p(n)] = \sum_{p\le x} \left(\frac{1}{p} + O\left(\frac{1}{x}\right)\right) = \log\log x + O(1)$$

et variation

$$\sum_{p\le x} \mathrm{Var}_{n\le x}[B_p(n)] = \sum_{p\le x} \left(\frac{1}{p} - \frac{1}{p^2} + O\left(\frac{1}{x}\right)\right) = \log\log x + O(1),$$

où on a utilisé le théorème de Mertens (voir théorème 8.3(b)). Erdős et Kac ont montré en 1940 que la quasi-indépendance des variables B_p est suffisamment forte pour obtenir la convergence de la répartition de ω à une variable normale :

Théorème 10.1 (Erdős-Kac)**.** *Soient $\alpha < \beta$ fixés. On a que*

$$\lim_{x\to\infty} \mathbb{P}_{n\le x}\left(\log\log x + \alpha\sqrt{\log\log x} \le \omega(n) \le \log\log x + \beta\sqrt{\log\log x}\right) = \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-t^2/2} \mathrm{d}t.$$

Démonstration. On suit un argument de Billingsley [1]. Soit

$$\mu_x = \sum_{p\le x} \frac{1}{p}$$

1. Voir Section 30 du livre de P. Billingsley, *Probability and measure*, Third edition, Wiley Series in Probability and Mathematical Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1995.

et soit σ_x défini par

$$\sigma_x^2 = \sum_{p\le x}\left(\frac{1}{p}-\frac{1}{p^2}\right).$$

Puisque $\mu_x = \log\log x + O(1)$ et $\sigma_x \sim \sqrt{\log\log x}$, il suffit de montrer que

$$\mathbb{P}_{n\le x}\left(\alpha < \frac{\omega(n)-\mu_x}{\sigma_x} \le \beta\right) \sim \frac{1}{\sqrt{2\pi}}\int_\alpha^\beta e^{-t^2/2}\mathrm{d}t \quad (x\to\infty). \tag{10.2}$$

pour tous $\alpha < \beta$. Un théorème probabiliste implique qu'il suffit de montrer que les moments des variables aléatoires normalisées $(\omega(n)-\mu_x)/\sigma_x$ tendent vers les moments de la distribution normale standard, c'est-à-dire que

$$\mathbb{E}_{n\le x}\left[\left(\frac{\omega(n)-\mu_x}{\sigma_x}\right)^k\right] \sim \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} t^k e^{-t^2/2}\mathrm{d}t \quad (x\to\infty) \tag{10.3}$$

pour chaque $k\in\mathbb{N}$ (considéré fixé quand $x\to\infty$).

On considère maintenant les variables alèatoires de Bernoulli $(K_p)_{p \text{ premier}}$. On construit les K_p de sorte qu'elles soient indépendantes l'une à l'autre, et qu'elles assument juste les valeurs 0 et 1 avec $\mathbb{P}(K_p = 1) = 1/p$. On peut imaginer que les variables K_p sont des 'cousins' des variables B_p qui la bonne propriété additionnelle de leur indépendance de l'une à l'autre. Elles vivent dans un espace de probabilité ambient (dont la définition exacte n'est pas importante). Pour la somme de K_p alors le théorème central limite est un fait. En particulier, puisque

$$\sum_{p\le x}\mathbb{E}[K_p] = \sum_{p\le x}\frac{1}{p} = \mu_x$$

et

$$\sum_{p\le x}\mathrm{Var}[K_p] = \sum_{p\le x}\left(\frac{1}{p}-\frac{1}{p^2}\right) = \sigma_x^2,$$

on a que la suite des variables $(\sum_{p\le x}K_p-\mu_x)/\sigma_x$ converge en loi vers la distribution normale standard quand $x\to\infty$. En fait, la forme spécifique de cette suite nous permet d'établir que leurs moments aussi tendent vers les moments de la distribution normale standard, c'est-à-dire

$$\mathbb{E}\left[\left(\frac{\sum_{p\le x}K_p-\mu_x}{\sigma_x}\right)^k\right] \sim \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} t^k e^{-t^2/2}\mathrm{d}t \quad (x\to\infty).$$

La discussion ci-dessus réduit (10.3) à démontrer que

$$\mathbb{E}_{n\le x}\left[\left(\frac{\omega(n)-\mu_x}{\sigma_x}\right)^k\right] \sim \mathbb{E}\left[\left(\frac{\sum_{p\le x}K_p-\mu_x}{\sigma_x}\right)^k\right] \quad (x\to\infty).$$

En multipliant les deux côtés par σ_x^k, on voit qu'il suffit de prouver que

$$\Delta_k := \mathbb{E}_{n\le x}[(\omega(n)-\mu_x)^k] - \mathbb{E}\left[\left(\sum_{p\le x}K_p-\mu_x\right)^k\right] = o_{x\to\infty}((\log\log x)^k). \tag{10.4}$$

On développe les k-ième puissances en utilisant le théorème du binôme : on a que

$$(\omega(n)-\mu_x)^k = \sum_{j=0}^{k} \binom{k}{j} \omega(n)^j (-\mu_x)^{k-j}.$$

La linéarité de l'espérance implique que

$$\mathbb{E}_{n\leq x}[(\omega(n)-\mu_x)^k] = \sum_{j=0}^{k} \binom{k}{j} (-\mu_x)^{k-j} \mathbb{E}_{n\leq x}[\omega(n)^j].$$

De façon similaire, on a que

$$\mathbb{E}\left[\left(\sum_{p\leq x} K_p - \mu_x\right)^k\right] = \sum_{j=0}^{k} \binom{k}{j} (-\mu_x)^{k-j} \mathbb{E}\left[\left(\sum_{p\leq x} K_p\right)^j\right],$$

d'où

$$\Delta_k = \sum_{j=0}^{k} \binom{k}{j} (-\mu_x)^{k-j} \left(\mathbb{E}_{n\leq x}[\omega(n)^j] - \mathbb{E}\left[\left(\sum_{p\leq x} K_p\right)^j\right]\right).$$

On a réduit donc (10.4) à démontrer que

$$\delta_j := \mathbb{E}_{n\leq x}[\omega(n)^j] - \mathbb{E}\left[\left(\sum_{p\leq x} K_p\right)^j\right] = o_{x\to\infty}((\log\log x)^j) \tag{10.5}$$

pour chaque nombre naturel fixé j.

Afin de montrer (10.5), on développe les j-ièmes puissances : puisque $\omega = \sum_{p\leq x} B_p$, on a que

$$\omega^j = \sum_{p_1,\dots,p_j\leq x} B_{p_1}\cdots B_{p_j}.$$

Donc

$$\mathbb{E}_{n\leq x}[\omega(n)^j] = \sum_{p_1,\dots,p_j\leq x} \mathbb{E}_{n\leq x}[B_{p_1}(n)\cdots B_{p_j}(n)].$$

Puisque les variables aléatoires prennent juste les valeurs 0 et 1, on trouve que

$$\mathbb{E}_{n\leq x}[B_{p_1}(n)\cdots B_{p_j}(n)] = \mathbb{P}_{n\leq x}[B_{p_1}(n) = \cdots = B_{p_j}(n) = 1].$$

On peut éliminer les premiers qui se répètent dans la liste $p_1,\dots,p_j$. Donc (10.1) implique que

$$\mathbb{E}_{n\leq x}[B_{p_1}(n)\cdots B_{p_j}(n)] = \frac{1}{[p_1,\dots,p_j]} + O\left(\frac{1}{x}\right),$$

d'où

$$\mathbb{E}_{n\leq x}[\omega(n)^j] = \sum_{p_1,\dots,p_j\leq x} \left(\frac{1}{[p_1,\dots,p_j]} + O\left(\frac{1}{x}\right)\right) = \sum_{p_1,\dots,p_j\leq x} \frac{1}{[p_1,\dots,p_j]} + O\left(\frac{\pi(x)^j}{x}\right).$$

Le même argument implique que

$$\mathbb{E}\left[\left(\sum_{p\le x} K_p\right)^j\right] = \sum_{p_1,\dots,p_j\le x} \mathbb{E}[K_{p_1}\cdots K_{p_j}] = \sum_{p_1,\dots,p_j\le x} \frac{1}{[p_1,\dots,p_j]}.$$

Par la suite,

$$\delta_j \ll \frac{\pi(x)^j}{x}.$$

Le côté droit est trop grand en comparaison avec $(\log\log x)^j$ si $j \ge 2$.

On est arrivé alors à un impasse. Cependant, on peut utiliser une astuce pour le circumventer. On observe que si $n \le x$, alors $\#\{y < p \le x : p|n\} < \log x/\log y$. En effet, si $p_1,\dots,p_r$ sont les facteurs premiers $> y$ de n, alors $p_1\cdots p_r | n$. En particulier, $p_1\cdots p_r \le n \le x$. D'autre côté, $p_1\cdots p_r > y^r$ car $p_j > y$ pour chaque j. On conclut alors que $y^r < x$, ce qui montre notre affirmation que $r \le \log x/\log y$.

On applique l'observation ci-dessus en prenant $y = y(x) := x^{1/\log\log\log x}$. Si on pose

$$\omega(n;y) = \#\{p|n : p \le y\},$$

on a que

$$0 \le \omega(n) - \omega(n;y) \le \log\log\log x,$$

ainsi que

$$0 \le \mu_x - \mu_y = \sum_{y<p\le x} \frac{1}{p} = \log\frac{\log x}{\log y} + O(1) = \log\log\log\log x + O(1).$$

Alors, au lieu de monter (10.2), il suffit de montrer que

$$\mathbb{P}_{n\le x}\left(\alpha < \frac{\omega(n;y)-\mu_y}{\sigma_y}\right) \sim \frac{1}{\sqrt{2\pi}}\int_\alpha^\beta e^{-t^2/2}\mathrm{d}t \quad (x\to\infty).$$

En fait, comme on l'a discuté avant, il suffit de montrer que

$$\mathbb{E}_{n\le x}\left[\left(\frac{\omega(n;y)-\mu_x}{\sigma_x}\right)^k\right] \sim \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty t^k e^{-t^2/2}\mathrm{d}t \quad (x\to\infty)$$

ou, de façon équivalent, on peut plutôt prouver que

$$\mathbb{E}_{n\le x}\left[\left(\frac{\omega(n;y)-\mu_x}{\sigma_x}\right)^k\right] \sim \mathbb{E}\left[\left(\frac{\sum_{p\le y}K_p - \mu_y}{\sigma_y}\right)^k\right] \quad (x\to\infty).$$

Comme avant, on réduit cette relation à l'estimation

$$\mathbb{E}_{n\le x}[\omega(n;y)^j] - \mathbb{E}\left[\left(\sum_{p\le y} K_p\right)^j\right] = o_{x\to\infty}((\log\log x)^j) \tag{10.6}$$

pour chaque nombre naturel fixé j. On a vu que

$$\mathbb{E}\left[\left(\sum_{p\le y} K_p\right)^j\right] = \sum_{p_1,\dots,p_j\le y} \frac{1}{[p_1,\dots,p_j]}.$$

De plus, on a que

$$\begin{aligned}
\mathbb{E}_{n\le x}[\omega(n;y)^j] &= \mathbb{E}_{n\le x}\left[\left(\sum_{p\le y} B_p(n)\right)^j\right] \\
&= \sum_{p_1,\dots,p_j\le y} \mathbb{E}_{n\le x}[B_{p_1}(n)=\cdots=B_{p_j}(n)=1] \\
&= \sum_{p_1,\dots,p_j\le y} \left(\frac{1}{[p_1,\dots,p_j]} + O\left(\frac{1}{x}\right)\right) \\
&= \mathbb{E}\left[\left(\sum_{p\le y} K_p\right)^j\right] + O\left(\frac{\pi(y)^j}{x}\right).
\end{aligned}$$

Mais ici $\pi(y) \le y = x^{1/\log\log\log x}$, donc l'erreur est miniscule. Ceci établit (10.6) et complète la démonstration du théorème. □

Troisième partie

Méthodes transcendantales

Chapitre 11

Nombres irrationnels et transcendantaux

Les anciens grecques considéraient les nombres d'une façon très géométrique, comme longueurs de lignes. Pythagore de Samos a fondé une association sécrète, appelée « les pythagoriciens », qui est allée même plus loin. Les pythagoriciens ont développé une théorie du cosmos qui se basait sur le *principe d'analogies*. En langue moderne, ils pensaient que tous les nombres constructibles à la règle et au compas sont analogues l'un de l'autre, c'est-à-dire ils sont de nombres rationnels. Cependant un membre des pythagoriciens, Hippase de Métaponte, a découvert un trou dans cette théorie. Considérons le triangle droite dont les cotés perpendiculaires ont longueur 1. Alors le théorème de Pythagore implique que son hypoténuse a longueur $\sqrt{2}$. Hippase a observé que $\sqrt{2}$ est, contrairement à la croyance des pythagoriciens, irrationnel ! L'argument est simple : si $\sqrt{2}$ était rationnel, alors il existerait $a, b \in \mathbb{N}$ tels que $(a, b) = 1$ et $\sqrt{2} = a/b$. Donc $b^2 = 2a^2$, ce qui implique que $2|b^2$ et, par la suite, que $2|b$. Alors $4|b^2 = 2a^2$, d'où on déduit que $2|a^2$, c'est-à-dire $2|a$ aussi. On est arrivé à une conclusion absurde : on a supposé que $(a, b) = 1$, et on montré que $2|a$ et que $2|b$. Par conséquent, notre hypothèse initiale, que $\sqrt{2} \in \mathbb{Q}$, doit être fausse. Ceci conclut la démonstration du fait que $\sqrt{2}$ est un nombre irrationnel.

La découverte hérétique d'Hipasse lui a couté sa propre vie. Aujourd'hui, on sait que les nombres irrationnels sont la grande majorité des nombres réels :

Théorème 11.1 (Cantor)**.** *L'ensemble des nombres réels est indénombrable. En particulier, les nombres irrationnels forment un ensemble indénombrable.*

Démonstration. La démonstration, grâce à Cantor, utilise son fameux argument diagonal. On montrera que $(0, 1]$ est indénombrable, qui est une déclaration plus forte. On considère une suite des nombres $a_1, a_2, \ldots, a_n, \ldots$ appartenants à $(0, 1)$. On construira un nouveau élément de $(0, 1]$ qui est diffèrent des $a_1, a_2, \ldots$. Ceci suffit pour déduire notre affirmation. On construit ce nouveaux nombre comme suivant : pour chacun nombre, on considère son expansion décimale, soit $a_i = 0.a_{i1}a_{i2}\cdots a_{in}\cdots$, où $a_{ij} \in \{0, 1, \ldots, 9\}$. On exige que ce soit une expansion infinie. (Pour tout nombre, il existe une unique telle expansion. Par exemple, si 0.5 appartient à notre suite, on considère son expansion infinie $0.4999\ldots$) On met dans

une liste tous les nombres de la suite $\{a_n\}_{n\geq 1}$:

$$\begin{aligned}
a_1 &= 0.\textcircled{a_{11}}a_{12}a_{13}\cdots a_{1n}\cdots \\
a_2 &= 0.a_{21}\textcircled{a_{22}}a_{23}\cdots a_{2n}\cdots \\
a_3 &= 0.a_{31}a_{32}\textcircled{a_{33}}\cdots a_{3n}\cdots \\
&\vdots \\
a_n &= 0.a_{n1}a_{n2}a_{n3}\cdots\textcircled{a_{nn}}\cdots \\
&\vdots
\end{aligned}$$

On considère les chiffres de la diagonale, $a_{11}, a_{22}, a_{33}, \ldots$, avec lesquels on construit le nombre $b \in (0, 1]$ dont l'expansion décimale $b = 0.b_1b_2b_3\cdots$ est donnée par

$$b_i = \begin{cases} 1 & \text{si } a_{ii} \neq 1, \\ 2 & \text{si } a_{ii} = 1. \end{cases}$$

Donc l'expansion $b = 0.b_1b_2\cdots$ est infinie. De plus, le i-ième chiffre décimal de cette expansion est different que le i-ième chiffre décimale du nombre a_i. Alors $b \neq a_i$, pour tout $i \in \mathbb{N}$, comme voulu. Ceci conclut la démonstration. □

Bien que les nombres irrationnels soient abondants, il est souvent assez difficile de montrer qu'un nombre donné est irrationnel. Ici on montre l'irrationalité d'un constant fameux, le constant d'Euler $e = 2.718\ldots$, défini par $e = \lim_{n\to\infty}(1 + 1/n)^n$.

Théorème 11.2. *Le nombre e est irrationnel.*

Démonstration. On a l'expansion de la fonction $x \to e^x$ à sa série de Taylor

$$e^x = \sum_{n\geq 0} \frac{x^n}{n!}.$$

Donc

$$e = \sum_{n\geq 0} \frac{1}{n!}.$$

C'est la clé pour finir la démonstration : on a exprimé e comme une série des nombres rationnels qui converge très rapidement. Cette idée est centrale à plusieurs démonstrations qu'un nombre est irrationnel (ou, comme on verra à la section **??**, qu'un nombre est transcendant). Supposons que $e = a/b$ pour quelques $a, b \in \mathbb{N}$. Donc le nombre

$$b! \cdot \left(e - \sum_{n=0}^{b} \frac{1}{n!} \right)$$

est entier. En effet, $b!e = (b-1)!a \in \mathbb{Z}$ et $b!\sum_{n=1}^{b} 1/n! = \sum_{n=0}^{n} n!/b! \in \mathbb{Z}$ car $b!/n! = b(b-1)(b-2)\cdots(n+1) \in \mathbb{Z}$ pour tout $n \in \{0, 1, \ldots, b\}$. Cependant,

$$\begin{aligned} 0 < b! \cdot \left(e - \sum_{n=0}^{b} \frac{1}{n!} \right) = b! \sum_{n=b+1}^{\infty} \frac{1}{n!} = \sum_{n=b+1}^{\infty} \frac{1}{(b+1)(b+2)\cdots n} &< \sum_{n=b+1}^{\infty} \frac{1}{(b+1)^{n-b}} \\ &= \frac{1}{b+1} \cdot \frac{1}{1-1/(b+1)} \\ &= \frac{1}{b} \leq 1. \end{aligned}$$

C'est une contradiction : on a trouvé un nombre entier dans l'intervalle $(0, 1)$, ce qui est impossible. □

Bien que la majorité de nombres réels soient irrationnels, pour nous c'est beaucoup plus simple de comprendre les nombres rationnels. Par exemple, les nombres rationnels ont une expansion décimale finie ou périodique. Pour cette raison, les mathématiciens on a essayé de trouver de bonnes approximations rationales des nombres irrationnels. Une façon de le faire est de considérer l'expansion décimale d'un nombre irrationnel x, soit $x = a_k a_{k-1} \cdots a_0.b_1 b_2 \cdots$. Puis on peut considérer les nombres rationnels $x_n = a_k a_{k-1} \cdots a_0.b_1 b_2 \cdots b_n$, qui sont d'approximations de x. Mais comment est-ce qu'on peut mesurer la qualité d'une approximation rationale ? Supposons que a/b est une approximation rationale du nombre réel x. On voudrait que $|x - a/b|$ est petit. En particulier, on voulait que $|x - a/b| \leq 1/b$; sinon, on peut trouver un autre nombre $a' \in \mathbb{Z}$ tel que $|x - a'/b| \leq 1/b < |x - a/b|$. En général, on mesure la qualité de l'approximation rationale a/b en termes de la taille de son dénominateur b (en supposant que la fraction a/b est réduite, c'est-à-dire que $(a, b) = 1$). Les approximations x_n du nombre x, construites au-dessus, ne sont pas si bonnes en général : si il existe un chiffre b_m avec $n < m \leq n + C$, pour un constant C petit (c'est-à-dire si le nombre x n'a pas un très longue lacune environ son n-ième chiffre, un événement rare), alors on a que

$$x - x_n \geq \frac{b_m}{10^m} \geq \frac{1}{10^{n+C}}.$$

Mais si $b_n \neq 0$ (ou si $b_m \neq 0$ pour un $m \in (n - C, n]$), alors le dénominateur de la fraction x_n est $\approx 10^n$, c'est-à-dire si on écrit $x_n = a/b$, alors $x - x_n = x - a/b \approx 1/b$. C'implique que, en général, x_n est une approximation faible à x.

Cependant, on peut construire d'approximations beaucoup mieux pour un nombre donné x. Ils sont les fractions continus, dont la théorie on développera à la section 12. Pour le moment, on montre le théorème suivant.

Théorème 11.3 (Théorème d'approximation de Dirichlet)**.** *Soient $x \in \mathbb{R}$ et $Q \geq 1$. Alors il existe une fraction réduite a/q telle que $1 \leq q \leq Q$ et*

$$\left| x - \frac{a}{q} \right| \leq \frac{1}{qQ} \leq \frac{1}{q^2}.$$

Démonstration. Sans perde de généralité, on peut supposer que $Q \in \mathbb{N}$. On considéré les $Q + 1$ nombres $\{x\}, \{2\alpha\}, \ldots, \{(Q+1)\alpha\}$ qui sont situés dans $[0, 1)$. On partage $[0, 1)$ dans

les Q intervalles $[(j-1)/Q, j/Q)$, $1 \leq j \leq Q$. Le principe des tiroirs implique que au moins deux nombres entre $\{x\}, \{2x\}, \ldots, \{(Q+1)x\}$, soient $\{kx\}$ et $\{\ell x\}$ avec $1 \leq k < \ell \leq Q+1$, appartient au même intervalle entre les moins deux de ces nombres sont situés au même intervalle $[(j-1)/Q, j/Q)$, pour un $j \in \{1, \ldots, Q\}$. En particulier, on a que

$$|\{\ell x\} - \{kx\}| < \frac{1}{Q}.$$

En posant $m = \lfloor kx \rfloor$ et $n = \lfloor \ell x \rfloor$, on trouve que

$$|(\ell - k)x - (n-m)| < \frac{1}{Q} \quad \Rightarrow \quad \left| x - \frac{n-m}{\ell - k} \right| < \frac{1}{(\ell - k)Q}$$

Donc, si a/q est la fraction réduite de $(m-n)/(\ell - k)$, on trouve que $1 \leq q \leq \ell - k \leq Q$ et le théorème découle. □

Jusqu'à ce point, on a examiné les nombres irrationnels et rationnels. Cependant, pas tous les nombres irrationnels ont la même complexité. Par exemple, le nombre $\sqrt{2}$ est une des racines de l'équation $x^2 - 2 = 0$ et le nombre $5^{1/3}$ satisfait l'équation algébrique $x^3 - 5 = 0$. Il y a des exemples plus compliqués : par exemple, le nombre $\sqrt{2} + \sqrt{3}$ est une des racines du polynôme

$$\begin{aligned}
&(x - \sqrt{2} - \sqrt{3})(x - \sqrt{2} + \sqrt{3})(x + \sqrt{2} - \sqrt{3})(x + \sqrt{2} + \sqrt{3}) \\
&\quad = ((x - \sqrt{2})^2 - 3)((x + \sqrt{2})^2 - 3) \\
&\quad = (x^2 - 1 - 2\sqrt{2}x)(x^2 - 1 + 2\sqrt{2}x) \\
&\quad = (x^2 - 1)^2 - 8x^2 = x^4 - 10x^2 + 1.
\end{aligned}$$

Les nombres qui ont la même propriété, d'être une racine d'un polynôme dont les coefficients sont rationnels ont un nom spécial :

Définition 11.4. Un nombre complexe α est appelé algébrique si il existe un polynôme $f(x) \in \mathbb{Q}[x]$ tel que $f(\alpha) = 0$. Un nombre complexe qui n'est pas algébrique est appelé transcendant.

On peut donner un mesure de la complexité d'un nombre algébrique en utilisant le concept du degré :

Définition 11.5. Si α est un nombre algébrique, alors il existe un polynôme $f(x) \in \mathbb{Q}[x]$ de degré **minimal** tel que $f(\alpha) = 0$. Le degré de ce polynôme est appelé le degré α et dénoté par $\deg(\alpha)$.

La discussion au-dessous implique que les nombres $\sqrt{2}$, $5^{1/3}$ et $\sqrt{2} + \sqrt{3}$ sont algébriques, de degré 2, 3 et 4, respectivement [1] D'autre côté, on sait que les nombres e et π sont de nombres transcendants. Au théorème 11.7 on donnera un autre exemple concret d'un nombre transcendant. L'étape-clé est le résultat suivant, qui démontre que les nombres algébriques ne peuvent être très proche d'un nombre rationnel.

1. On peut montrer que si $f(\sqrt{2} + \sqrt{3}) = 0$ pour un polynôme $f(x) \in \mathbb{Q}[x]$, alors nécessairement $f(\sqrt{2} - \sqrt{3}) = f(-\sqrt{2} + \sqrt{3}) = f(-\sqrt{2} - \sqrt{3}) = 0$. Donc $\deg(\sqrt{2} + \sqrt{3}) = 4$.

Théorème 11.6. *Si α est un nombre réel algébrique de degré $n \geq 2$, alors il existe un constant c dépendant seulement de α tel que*

$$\left|\alpha - \frac{a}{b}\right| \geq \frac{c}{b^n},$$

pour tous $a \in \mathbb{Z}$ et $b \in \mathbb{N}$.

Démonstration. Soit $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in \mathbb{Q}[x]$ tel que $f(\alpha) = 0$. Soient a et b comme dans la déclaration du théorème. Le théorème de la valuer moyenne implique que

$$-f\left(\frac{a}{b}\right) = f(\alpha) - f\left(\frac{a}{b}\right) = \left(\alpha - \frac{a}{b}\right) f'(t),$$

pour un t entre α et a/b. Donc si on donne un borne supérieure à $|f'(t)|$ et un borne inférieure à $|f(a/b)|$, on prendra un borne inférieure à $|\alpha - a/b|$. Tout d'abord, on a que $\alpha \neq a/b$ car le degré de α est ≥ 2 et, par la suite, $\alpha \notin \mathbb{Q}$. Aussi, il existe $\varepsilon > 0$ tel que l'intervallee $I := [\alpha - \varepsilon, \alpha + \varepsilon]$ n'a pas de racines de $f(x)$ sauf α. Si $a/b \notin I$, alors $|a/b - \alpha| \geq \varepsilon \geq \varepsilon/b^n$ et le théorème découle à condition que $c \leq \varepsilon$. Finalement, supposons que $a/b \in I$ et posons $M = \sup_{x \in I} |f'(x)|$. Donc

$$\left|\alpha - \frac{a}{b}\right| \geq \frac{|f(a/b)|}{M}.$$

De plus, on a que

$$|f(a/b)| = \frac{|a_n a^n + a_{n-1}a^{n-1}b + \cdots + a_1 a b^{n-1} + a_0 b^n|}{b^n} \geq \frac{1}{b^n}$$

parce que le numérateur est un entier différent de 0 (rappelez que $f(a/b) \neq 0$ car la seule racine de f appartenante à I est α et $a/b \neq \alpha$). Le théorème découle en prennant $c = \min\{\varepsilon, 1/M\}$. □

Théorème 11.7 (Liouville)**.** *Le nombre*

$$x = \sum_{n \geq 1} \frac{1}{10^{n!}} = 0.11000100000000000000000001\ldots$$

est transcendant.

Démonstration. On montrera que, à cause des grandes ensembles de zéros consécutifs, le nombre x est très proche à quelques nombres rationnels, trop proche pour être algébrique, selon le théorème 11.6. En effet, soit

$$x_k = \sum_{n=1}^{k} \frac{1}{10^{n!}}$$

pour tout $k \geq 1$, une suite des nombres rationnels qui approchent x lorsque $k \to \infty$. On a que $x_k = a_k/b_k$ avec $a_k \in \mathbb{N}$ et $b_k = 10^{k!}$. De plus,

$$0 \leq x - x_k = \sum_{n=k+1} \frac{1}{10^{n!}} \leq \frac{1}{10^{(k+1)!}} \sum_{n=k+1}^{\infty} \frac{1}{10^{n-k-1}} = \frac{1}{9 \cdot 10^{(k+1)!-1}} = \frac{10}{9b_k^{k+1}}.$$

Alors x doit être transcendant. Sinon, il serait algébrique, soit de degré d. Selon le théorème 11.6, ceci impliquerait que

$$|x - x_k| \geq \frac{c}{b_k^d},$$

pour un $c > 0$ qui est independent de b_k. Alors il faudrait que

$$\frac{c}{b_k^d} \leq \frac{10}{9b_k^{k+1}} \quad \Longrightarrow \quad b_k^{k+1-d} \leq \frac{10}{9c}.$$

C'est impossible car le côté droit de la dernière relation est non borné lorsque $k \to \infty$. Alors on conclut que x est transcendant, comme affirmé. □

Exercices

EXERCICE 11.1. Soient $n, k \in \mathbb{N}$ avec $k \geq 2$. Montrez que le nombre $n^{1/k}$ est rationnel si et seulement si n est une k-ième puissance parfaite.

EXERCICE 11.2. Est-ce que le nombre $\sqrt{2} + \sqrt{3}$ est rationnel ou irrationnel ?

Chapitre 12

Fractions continues

On développe ici la théorie des fractions continues qui donnent de très bonnes approximations rationales aux nombres irrationnels. De plus, cette théorie se caractérise d'une grande élégance, comme on le verra.

On commence avec l'approximation rationale d'un nombre la plus simple : la partie entier. En effet, si $x \in \mathbb{R}$, alors $x = \lfloor x \rfloor + \{x\}$, où $\lfloor x \rfloor \in \mathbb{Z}$ et $\{x\} \in [0, 1)$. Donc, si on pose $a_1 = \lfloor x \rfloor$ et $\theta_1 = \{x\}$, alors a_1 est une approximation rationale (en fait, entier !) de x et l'erreur de cette approximation est égale à $\theta_1 \in [0, 1)$. Puis, on observe que

$$x = a_1 + \theta_1 = a_1 + \frac{1}{1/\theta_1}.$$

Alors si on approxime $1/\theta_1$ par sa partie entier, soit a_2, et on met $\theta_2 := \{1/\theta_1\}$, donc on trouve que

$$x = a_1 + \frac{1}{a_2 + \theta_2}.$$

Le nombre rationnel $a_1 + 1/a_2$ serve comme une approximation de x. On continue comme avant : on inverse θ_2 et on approxime son inverse, $1/\theta_2$, par $a_2 := \lfloor 1/\theta_2 \rfloor$. Si $\theta_3 = \{1/\theta_2\}$, alors ceci nous amène à la relation

$$x = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \theta_3}} \approx a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3}}.$$

Le nombre rationnel au coté droit de cette relation est une autre approximation de x (meilleure que a_1 et que $a_1 + 1/a_2$, comme on le verra).

Bien sûr, on peut continuer la procedure précédente indéfiniment. Alors on trouve des nombres $a_1 \in \mathbb{Z}$, $a_2, a_3, \cdots \in \mathbb{N}$ et $\theta_1, \theta_2, \theta_3, \cdots \in [0, 1)$ tels que $a_1 = \lfloor x \rfloor$, $\theta_1 = \{x\}$ et $a_{n+1} = \lfloor 1/\theta_n \rfloor$ et $\theta_{n+1} = \lfloor 1/\theta_n \rfloor$, pour tout $n \in \mathbb{N}$. (Si $\theta_N = 0$ pour un $N \in \mathbb{N}$, alors cette procedure termine après N étapes ; donc, les suites des θ_i et des a_i sont, en fait, finies.) De

plus, on a que

$$x = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\cdots + a_{n-1} + \cfrac{1}{a_n + \theta_n}}}} \approx a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\cdots + a_{n-1} + \cfrac{1}{a_n}}}},$$

pour tout $n \in \mathbb{N}$ pour lequel les nombres $a_1, \ldots, a_n$ et $\theta_1, \ldots, \theta_n$ existent. Le nombre rationnel au coté droit de la relation au-dessus est appelé le n-ième convergent de x et il est dénoté par p_n/q_n (pour le moment, on ne précis pas la définition précise nombres p_n et q_n ; on les examinera plus soigneusement plus tard). Aussi, on a la définition suivante.

Définition 12.1. Étant donnés n nombres réels non-zéros $a_1, a_2 \ldots, a_n$, on appelle la fraction continue de $a_1, \ldots, a_n$ le nombre

$$[a_1, \ldots, a_n] := a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{\cdots + \cfrac{1}{a_n}}}}.$$

De plus, étant donnée une suite infinie de nombres réels non-zéros $a_1, a_2 \ldots$, on appelle la fraction continue de $a_1, a_2, \ldots$ le nombre

$$[a_1, a_2, \ldots] := \lim_{n\to\infty} [a_1, \ldots, a_n],$$

si le limite existe, au quel cas on dénote aussi par

$$[a_1, a_2, \ldots] = a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}.$$

Si $x \in \mathbb{R}$ et on définit les suites $a_1, a_2, \ldots$ et $\theta_1, \theta_2, \ldots$ comme au-dessus, alors on trouve que

$$x = [a_1, a_2, \ldots, a_{n-1}, a_n + \theta_n] \quad \text{et que} \quad \frac{p_n}{q_n} = [a_1, a_2, \ldots, a_n].$$

La fraction continue (finie ou infinie) $[a_1, a_2, \ldots]$, obtenue de cette procedure, est appelée la fraction continue de x. Ce n'est pas très difficile de montrer que la fraction continue d'un nombre est défini uniquement (voyez exercice 12.1).

On calcule les fractions continues de deux nombres concrets. Si $x = \frac{41}{13}$, alors on a que

$$x = 3 + \frac{2}{13} = 3 + \frac{1}{13/2} = 3 + \cfrac{1}{6 + \cfrac{1}{2}} = [3, 6, 2].$$

De même, si $x = \frac{100}{17}$, alors on a que

$$x = 5 + \frac{15}{17} = 5 + \frac{1}{17/15} = 5 + \frac{1}{1 + \frac{2}{15}} = 5 + \frac{1}{1 + \frac{1}{15/2}} = 5 + \frac{1}{1 + \frac{1}{7 + \frac{1}{2}}} = [5, 1, 7, 2].$$

On a vu dans les des exemples au-dessus que les fractions continues de $\frac{41}{13}$ et de $\frac{100}{17}$ sont finies. Ce n'a pas été un accident :

Théorème 12.2. *La fraction continue d'un nombre réel x est finie si et seulement si x est rationnel.*

Démonstration. Si la fraction continue de x est finie, c'est evident que x est rationnel car il existe $n \in \mathbb{N}$ tel que $x = [a_1, \ldots, a_n] \in \mathbb{Q}$.

Réciproquement, supposons que $x \in \mathbb{Q}$. On écrit $x = k/\ell$ avec $k \in \mathbb{Z}$ et $\ell \in \mathbb{N}$ et $(k, \ell) = 1$. On observe que si $a_1 = \lfloor k/\ell \rfloor$, alors

$$a_1 \leq \frac{k}{\ell} < a_1 + 1 \quad \Longrightarrow \quad a_1\ell \leq k < a_1\ell + \ell.$$

Donc le nombre $k - a_1\ell$ est un des nombres $0, 1, \ldots, \ell - 1$, c'est-à-dire $k - a_1\ell$ est le reste dans la division de k par ℓ (et, par conséquent, a_1 est le quotient de cette division). Si on pose $r_1 = k - a_1\ell$, alors

$$x = \frac{k}{\ell} = \frac{\ell a_1 + r_1}{\ell} = a_1 + \frac{r_1}{\ell} = a_1 + \frac{1}{\ell/r_1}.$$

Si $r_1 = 0$, on a finit. Sinon, on continue comme au-dessus : afin de trouver a_2, on fait la division euclidienne de ℓ par r_1 et on trouve $a_2 \in \mathbb{N}$ et $r_2 \in \{0, 1, \ldots, r_1 - 1\}$ tels que $\ell = a_2 r_1 + r_2$. Donc $\ell/r_1 = a_2 + r_2/r_1$ et $0 \leq r_2/r_1 < 1$. Par la suite,

$$x = a_1 + \frac{1}{a_2 + \frac{r_2}{r_1}} = a_1 + \frac{1}{a_2 + \frac{1}{r_1/r_2}}.$$

Si $r_2 = 0$, alors on a finit la démonstration. Sinon, on trouve a_3 et r_3 tels que $r_1 = a_3 r_2 + r_3$ et $r_3 \in \{0, 1, \ldots, r_2 - 1\}$. En continuant dans cette façon, après n étapes, on aura construit une suite $r_1, r_2, \ldots, r_n, \ldots$ telle que $0 \leq r_n < r_{n-1} < \cdots < r_1 < \ell$, c'est-à-dire on a construit n nombres distincts appartenants à $\{0, 1, \ldots, \ell\}$. Ceci montre que cette procedure ne peut pas continuer infiniment. Plutôt, il existe un $n \in \{1, \ldots, \ell\}$ tel que $r_n = 0$, ce qui implique que $x = [a_1, \ldots, a_n]$, comme voulu. □

Notre but maintenant est de comprendre la relation entre x et ses convergents p_n/q_n. Tout d'abord, on donne une définition précise de p_n et de q_n. Avant de donner la définition générale, on définit quelques cas spéciaux pour des raisons pédagogiques : on a que la fraction

continue d'un élément a_1 est égale à $[a_1] = a_1 = a_1/1$. Donc on définit $p_1 = a_1$ et $q_1 = 1$. Puis, la fraction continue de a_1 et de a_2 est égale à

$$[a_1, a_2] = a_1 + \frac{1}{a_2} = \frac{a_1 a_2 + 1}{a_2}.$$

Donc on définit $p_2 = a_1 a_2 + 1$ et $q_2 = a_2$. De même, on a que

$$[a_1, a_2, a_3] = a_1 + \frac{1}{a_2 + \dfrac{1}{a_3}} = a_1 + \frac{a_3}{a_2 a_3 + 1} = \frac{a_1(a_2 a_3 + 1) + a_3}{a_2 a_3 + 1}, \tag{12.1}$$

ce qui nous amène à poser $p_3 = a_1(a_2 a_3 + 1) + a_2$ et $q_3 = a_2 a_3 + 1$. En général, ce n'est pas difficile de voir que

$$[a_1, \ldots, a_n] = \frac{F_n(a_1, \ldots, a_n)}{G_n(a_1, \ldots, a_n)}$$

pour quelques polynômes F_n et G_n. De plus, la relation

$$[a_1, \ldots, a_n] = a_1 + \frac{1}{[a_2, \ldots, a_n]} \tag{12.2}$$

implique que

$$\frac{F_n(a_1, \ldots, a_n)}{G_n(a_1, \ldots, a_n)} = a_1 + \frac{1}{\frac{F_{n-1}}{G_{n-1}}(a_2, \ldots, a_n)} = \frac{a_1 F_{n-1}(a_2, \ldots, a_n) + G_{n-1}(a_2, \ldots, a_n)}{F_{n-1}(a_2, \ldots, a_n)}.$$

Cette relation nous amène à la définition rigoureuse suivante.

Définition 12.3. On définit deux séquences de fonctions $\{F_n : \mathbb{R}^n \to \mathbb{R}\}_{n\geq 1}$ et $\{G_n : \mathbb{R}^n \to \mathbb{R}\}_{n\geq 1}$ inductivement par les relations $F_1(x_1) = x_1$, $G_1(x_1) = 1$,

$$F_n(x_1, \ldots, x_n) = x_1 F_{n-1}(x_2, \ldots, x_n) + G_{n-1}(x_2, \ldots, x_n)$$

et

$$G_n(x_1, \ldots, x_n) = F_{n-1}(x_2, \ldots, x_n).$$

Puis, étant donné une fraction continue $[a_1, \ldots, a_n]$, on met $p_i = F_i(a_1, \ldots, a_i)$ et $q_i = G_i(a_1, \ldots, a_i)$, pour tout $n \in \{1, \ldots, n\}$.[1]

La définition de F_n et de G_n implique tout de suite que

$$\frac{F_n(x_1, \ldots, x_n)}{G_n(x_1, \ldots, x_n)} = x_1 + \frac{1}{\frac{F_{n-1}}{G_{n-1}}(x_2, \ldots, x_n)}.$$

1. On pourrait avoir donné une définition moins rigoureuse des polynômes F_n et G_n : ils sont les polynômes qu'on obtient comme numérateur et dénominateur de la fraction [a1, . . . , an] après avoir fait toutes les simplifications qu'on peut (sans diviser). Cependant, cette définition intuitive est plus difficile d'utiliser en pratique.

En utilisant cette relation et la relation (12.2), on peut montrer inductivement que

$$\frac{p_n}{q_n} = \frac{F_n(a_1, \ldots, a_n)}{G_n(a_1, \ldots, a_n)} = [a_1, \ldots, a_n],$$

comme désiré.

Les nombres p_n et q_n satisfassent de relations itératives simples. En effet, on observe que $p_3 = a_3p_2 + p_1$ et que $q_3 = a_3q_2 + q_1$. On donne un dernière exemple : on a que

$$\begin{aligned}[a_1, a_2, a_3, a_4] = a_1 + \frac{1}{[a_2, a_3, a_4]} &= a_1 + \frac{a_4a_3 + 1}{a_4(a_2a_3 + 1) + a_2} \\ &= \frac{a_1[a_4(a_2a_3 + 1) + a_2] + a_4a_3 + 1}{a_4(a_2a_3 + 1) + a_2} \\ &= \frac{a_4(a_1(a_2a_3 + 1) + a_3) + a_1a_2 + 1}{a_4(a_2a_3 + 1) + a_2}.\end{aligned}$$

Par la suite, $p_4 = a_4(a_1(a_2a_3+1)+a_3)+a_1a_2+1 = a_4p_3+p_2$ et $q_4 = a_4(a_2a_3+1)+a_2 = a_4q_3+q_2$. C'est un phénomène général :

Théorème 12.4. *Pour tout $n \geq 1$, on a que $p_{n+2} = a_{n+2}p_{n+1} + p_n$ et que $q_{n+2} = a_{n+2}q_{n+2} + q_n$.*

Démonstration. On utilise induction sur n. Quand n = 1, on a déjà vu que $p_3 = a_3p_2 + p_1$ et que $q_3 = a_3q_2 + q_1$. Puis, on suppose que la conclusion du théorème est valide quand $n \in \{1, \ldots, N-1\}$ et on la montre pour $n = N$ aussi. L'hypothèse inductive et la définition de F_n impliquent que

$$\begin{aligned}F_{N+2}(a_1, \ldots, a_{N+2}) &= a_1F_{N+1}(a_2, \ldots, a_{N+2}) + G_{N+1}(a_2, \ldots, a_{N+2}) \\ &= a_1[a_{N+2}F_N(a_2, \ldots, a_{N+1}) + F_{N-1}(a_2, \ldots, a_N)] \\ &\quad + [a_{N+2}G_N(a_2, \ldots, a_{N+1}) + G_{N-1}(a_2, \ldots, a_N)] \\ &= a_{N+2}(a_1F_N(a_2, \ldots, a_{N+1}) + G_N(a_2, \ldots, a_{N+1})) \\ &\quad + a_1F_{N-1}(a_2, \ldots, a_N) + G_{N-1}(a_2, \ldots, a_N) \\ &= a_{N+2}F_{N+1}(a_1, \ldots, a_{N+1}) + G_{N+1}(a_1, \ldots, a_{N+1}).\end{aligned}$$

Donc $p_{N+2} = a_{N+2}p_{N+1} + p_N$, comme voulu. De même, on a que

$$\begin{aligned}G_{N+2}(a1, \ldots, a_{N+2}) = F_{N+1}(a_2, ..., a_{N+2}) &= a_{N+2}F_N(a_2, \ldots, a_{N+1}) + F_{N-1}(a_2, \ldots, a_N) \\ &= a_{N+2}G_{N+1}(a_1, \ldots, a_{N+1}) + G_N(a_1, \ldots, a_N).\end{aligned}$$

Par la suite, on trouve aussi que $q_{N+2} = a_{N+2}q_{N+1}+q_N$, ce qui conclut la démonstration. □

Un corollaire directe du théorème précèdent est le résultat suivant.

Théorème 12.5. *Pour tout $n \geq 1$, on a que $p_{n+1}q_n - p_nq_{n+1} = (-1)^{n-1}$. En particulier, si $a_1, \ldots, a_n \in \mathbb{Z}$, alors p_n et q_n sont d'entiers copremiers.*

Démonstration. On utilise induction sur n. Soit $D_n = p_{n+1}q_n - p_nq_{n+1}$. Si $n = 1$, alors on a que $D_1 = p_2q_1 - p_1q_2 = (a_1a_2 + 1) \cdot 1 - a_1a_2 = 1$. Puis, on assume que le résultat est vrai pour n. On observe que

$$\begin{aligned} D_{n+1} = \det \begin{pmatrix} p_{n+2} & p_{n+1} \\ q_{n+2} & q_{n+1} \end{pmatrix} &= \det \begin{pmatrix} a_{n+2}p_{n+1} + p_n & p_{n+1} \\ a_{n+2}q_{n+1} + q_n & q_{n+1} \end{pmatrix} \\ (C_1 \to C_1 - a_{n+2}C_2) \qquad &= \det \begin{pmatrix} p_n & p_{n+1} \\ q_n & q_{n+1} \end{pmatrix} \\ (C_1 \leftrightarrow C_2) \qquad &= \det \begin{pmatrix} p_{n+1} & p_n \\ q_{n+1} & q_n \end{pmatrix} \\ &= D_n, \end{aligned}$$

ce qui conclut l'induction. □

Armés avec ce théorème, on peut déduire plus d'informations pour la suite des convergents d'un nombre.

Théorème 12.6. *Soit $x \in \mathbb{R}$ et $\{p_n/q_n\}$ la suite de ses convergents.*

(a) On a que

$$\frac{p_1}{q_1} < \frac{p_3}{q_3} < \cdots < \frac{p_{2n-1}}{q_{2n-1}} < \cdots \leq x \leq \cdots < \frac{p_{2n}}{q_{2n}} < \cdots < \frac{p_4}{q_4} < \frac{p_2}{q_2}.$$

(b) Pour tout $n \in \mathbb{N}$, on a que

$$\left| x - \frac{p_n}{q_n} \right| \leq \frac{1}{q_nq_{n+1}} < \frac{1}{q_n^2}.$$

En particulier, $\lim_{n\to\infty} p_n/q_n = x$.

Démonstration. (a) Tout d'abord, on montre que

$$\frac{p_{2n-1}}{q_{2n-1}} \leq x \leq \frac{p_{2n}}{q_{2n}}, \tag{12.3}$$

pour tout $n \in \mathbb{N}$. Puisque

$$\frac{p_i}{q_i} = [a_1, \ldots, a_i] \quad \text{et} \quad x = [a_1, \ldots, a_{i-1}, a_i + \theta_i]$$

pour un $\theta_i \in [0, 1)$, alors il suffit de montrer que la fonction $t \to f_i(t) := [a_1, \ldots, a_{i-1}, t]$ est croissante quand i est impair et décroissante quand i est pair. En effet, quand $i = 1$, on a que $f_1(t) = t$, une fonction qui est évidemment croissante. La relation (12.3) découle pour tout $i \geq 1$ de la formule

$$f_i(t) = a_1 + \frac{1}{[a_2, \ldots, a_{i-1}, t]}$$

et d'induction sur i.

Il reste de montrer que la suite $\{p_{2n-1}/q_{2n-1}\}_{n\geq 1}$ est strictement croissante et que la suite $\{p_{2n}/q_{2n}\}_{n\geq 1}$ est strictement décroissante. En effet, on que

$$\frac{p_{n+2}}{q_{n+2}} - \frac{p_n}{q_n} = \left(\frac{p_{n+2}}{q_{n+2}} - \frac{p_{n+1}}{q_{n+1}}\right) + \left(\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}\right) = \frac{(-1)^n}{q_{n+1}q_{n+2}} + \frac{(-1)^{n-1}}{q_n q_{n+1}}$$
$$= \frac{(-1)^{n-1}(q_{n+2} - q_n)}{q_n q_{n+1} q_{n+2}},$$

ce qui termine la démonstration de la partie (a).

(b) Pour tout n, alors x est entre p_n/q_n et p_{n+1}/q_{n+1}, selon la relation (12.3). Donc

$$\left|x - \frac{p_n}{q_n}\right| \leq \left|\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}\right| = \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2},$$

ce qui conclut la démonstration. □

Remarque 12.7. En général, si $a_1 \in \mathbb{Z}$ et $a_2, a_3, \cdots \in \mathbb{N}$, alors la suite $\frac{p_n}{q_n} = [a_1, \ldots, a_n]$ converge : sa sous-suite $\{p_{2n-1}/q_{2n-1}\}_{n\geq 1}$ est croissante et bornée au-dessus et, par la suite, elle est convergente. Soit x_1 sa limite. De même, la sous-suite $\{p_{2n}/q_{2n}\}_{n\geq 1}$ est décroissante et bornée en-dessous et, par la suite, elle est convergente. Soit x_2 sa limite. Finalement, puisque

$$0 \leq \frac{p_{2n}}{q_{2n}} - \frac{p_{2n-1}}{q_{2n-1}} = \frac{1}{q_{2n-1}q_{2n}} \to 0 \quad (n \to \infty),$$

alors $x_1 = x_2$. Donc le limite $\lim_{n\to\infty} p_n/q_n$ existe.

Puis on montre que les fractions continues sont les meilleurs approximations rationales des nombres irrationnels.

Théorème 12.8. *Si $x \in \mathbb{R} \setminus \mathbb{Q}$, alors*

$$\left|x - \frac{a}{b}\right| > \left|x - \frac{p_n}{q_n}\right|$$

pour tout $a, b \in \mathbb{Z}$ tels que $1 \leq b \leq q_n$ et $a/b \neq p_n/q_n$.

Démonstration. Supposons que n est pair ; le cas où il est impair est similaire.

Tout d'abord, on montre le théorème dans le cas spécial où $a/b = p_{n-1}/q_{n-1}$. Puisque

$$\frac{p_{n-1}}{q_{n-1}} < x < \frac{p_n}{q_n},$$

une conséquence du théorème 12.6(a) et de notre hypothèse que n est pair, alors on a que $|x - p_{n-1}/q_{n-1}| > |x - p_n/q_n|$ si et seulement la distance entre x et p_{n-1}/q_{n-1} et plus grand que le moyen de la distance entre p_n/q_n et p_{n-1}/q_{n-1}. C'est-à-dire, il suffit de montrer que

$$x - \frac{p_{n-1}}{q_{n-1}} > \frac{1}{2}\left|\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}}\right| = \frac{1}{2q_{n-1}q_n},$$

où on a appliqué le théorème 12.5. On a que $x > p_{n+1}/q_{n+1} > p_{n-1}/q_{n-1}$, par notre hypothèse que n est pair. Donc

$$\begin{aligned} x - \frac{p_{n-1}}{q_{n-1}} > \frac{p_{n+1}}{q_{n+1}} - \frac{p_{n-1}}{q_{n-1}} &= \left(\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}\right) + \left(\frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}}\right) \\ &= -\frac{1}{q_n q_{n+1}} + \frac{1}{q_{n-1}q_n} = \frac{q_{n+1} - q_{n-1}}{q_{n-1}q_n q_{n+1}}. \end{aligned}$$

Alors, il suffit de montrer que

$$\frac{q_{n+1} - q_{n-1}}{q_{n-1}q_n q_{n+1}} > \frac{1}{2q_{n-1}q_n} \quad \Leftrightarrow \quad q_{n+1} \geq 2q_{n-1}.$$

Mais on a que $q_{n+1} = a_n q_n + q_{n-1} \geq q_n + q_{n-1} > 2q_{n-1}$, ce qui montre le théorème dans le cas spécial où $a/b = p_n/q_n$.

Finalement, on considère le cas général. On a que

$$\left|\frac{a}{b} - \frac{p_n}{q_n}\right| = \frac{|aq_n - bp_n|}{bq_n} \geq \frac{1}{bq_n} \geq \frac{1}{bq_n} \geq \frac{1}{q_n^2},$$

car le numérateur est un entier non-zéro de notre hypothèse que $a/b \neq p_n/q_n$ et, aussi, on a supposé que $b \leq q_n$.

On distingue deux sous-cas. Si $a/b > x$, on affirme que $a/b > p_n/q_n$; sinon, on aurait que $x < a/b < p_n/q_n$ et, par la suite,

$$\frac{p_n}{q_n} - x \geq \frac{p_n}{q_n} - \frac{a}{b} \geq \frac{1}{bq_n} \geq \frac{1}{q_n^2},$$

ce qui contredit théorème 12.6(b). Donc $a/b > p_n/q_n$, ce qui implique que

$$\left|x - \frac{a}{b}\right| = \frac{a}{b} - x > \frac{p_n}{q_n} - x = \left|x - \frac{p_n}{q_n}\right|.$$

Puis, on considère le cas où $a/b < x$. D

Donc si $a/b > x$. Dans ce cas, on affirme que $a/b \leq p_{n-1}/q_{n-1}$; sinon, on aurait que $p_{n-1}/q_{n-1} < a/b < x$. Par conséquent, on trouverait que

$$x - \frac{p_{n-1}}{q_{n-1}} \geq \frac{a}{b} - \frac{p_{n-1}}{q_{n-1}} = \frac{aq_{n-1} - bp_{n-1}}{bq_{n-1}} \geq \frac{1}{bq_{n-1}} \geq \frac{1}{q_n q_{n-1}},$$

car $b \leq q_n$, ce qui est impossible car

$$x - \frac{p_{n-1}}{q_{n-1}} < \frac{p_n}{q_n} - \frac{p_{n-1}}{q_{n-1}} = \frac{1}{q_n q_{n-1}}.$$

Donc on a que $a/b \leq p_{n-1}/q_{n-1} < x$, ce qui implique que

$$\left|x - \frac{a}{b}\right| \geq \left|\frac{p_{n-1}}{q_{n-1}} - x\right| > \left|x - \frac{p_n}{q_n}\right|,$$

selon le cas où $a/b = p_{n-1}/q_{n-1}$ qu'on a déjà montré. Ceci conclut la démonstration du théorème. □

Remarque 12.9. Le théorème 12.8 implique directement le théorème d'approximation de Dirichlet. En effet, si $p_1/q_1, p_2/q_2, \dots$ est la suite des convergents de x, alors il y a deux cas. Si $q_n \leq Q$ pour tout $n \geq 1$, alors $x \in \mathbb{Q}$ et $x = b/r$ avec $r \leq Q$, et le résultat découle tout de suite en posant $a/q = b/r$. Finalement, si il existe $n \geq 1$ tel que $q_n > Q$, on peut trouver m tel que $q_m \leq Q < q_{m+1}$. Donc si on pose $a/q = p_m/q_m$, on a que

$$\left| x - \frac{a}{q} \right| = \left| x - \frac{p_m}{q_m} \right| \leq \frac{1}{q_m q_{m+1}} \leq \frac{1}{q_m Q} = \frac{1}{qQ},$$

ce qui est ce qu'il fallait montrer.

On conclut notre discussion des fractions continues avec une étude des fraction continues périodiques. On commence avec le plus simple exemple : considérons le nombre

$$x = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \dots}}}.$$

On observe que x est un auto-similarité : on a que

$$x = 1 + \frac{1}{x}.$$

Donc $x^2 - x - 1 = 0$, ce qui implique que $x = (1 \pm \sqrt{5})/2$. Puisque $x > 0$, alors on conclut que $x = (1 + \sqrt{5})/2$.

Puis, on considère un autre exemple, un plus compliqué. Soit

$$x = 4 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{3 + \cfrac{1}{2 + \cfrac{1}{3 + \cdots}}}}}.$$

On observe que

$$x = 4 + \cfrac{1}{1 + \cfrac{1}{y}},$$

où

$$y = 2 + \cfrac{1}{3 + \cfrac{1}{2 + \cfrac{1}{3 + \cdots}}}.$$

Maintenant on peu utiliser la même astuce : on a que

$$y = 2 + \cfrac{1}{3 + \cfrac{1}{y}} = 2 + \frac{y}{3y+1} = \frac{7y+2}{3y+1}.$$

Donc $3y^2 + y = 7y + 2$ ou, de façon équivalente, $3y^2 - 6y - 2 = 0$. Alors on conclut que $y = (3 \pm \sqrt{15})/3$. Puisque $y > 0$, on doit avoir que $y = (3 + \sqrt{15})/3$, d'où on déduit que

$$x = 4 + \cfrac{1}{1 + \cfrac{3}{3 + \sqrt{15}}} = 4 + \frac{3+\sqrt{15}}{6+\sqrt{15}} = 4 + \frac{1-\sqrt{15}}{7} = \frac{29-\sqrt{15}}{7}.$$

C'est un phénomène général :

Définition 12.10. Une fraction continue $[a_1, a_2, \dots]$ est appelée périodique si il existe $k \geq 1$ tel que $a_{n+k} = a_n$, pour tout $n \geq 1$. Dans ce cas, on écrite

$$[a_1, a_2, \dots] = [\overline{a_1, \dots, a_k}].$$

Une fraction continue $[a_1, a_2, \dots]$ est appelée finalement périodique si il existe $k \geq 1$ et $\ell \geq 1$ tels que $a_{n+k} = a_n$, pour tout $n \geq \ell$. Dans ce cas, on écrite

$$[a_1, a_2, \dots] = [a_1, \dots, a_{\ell-1}, \overline{a_\ell, \dots, a_{k+\ell-1}}].$$

Théorème 12.11. *La fraction continue d'un nombre x est périodique si et seulement si $x = r + \sqrt{s}$ pour quelques $r, s \in \mathbb{Q}$ avec $s \geq 0$.*

Démonstration. Sans perte de généralité, on suppose que $x \in \mathbb{R} \setminus \mathbb{Q}$; sinon, on a que $x = r + \sqrt{s}$ avec $r = x$ et $s = 0$ et que la fraction de x est finie selon le théorème 12.2.

Supposons que $x = [a_1, \dots, a_{\ell-1}, \overline{a_\ell, \dots, a_{k+\ell-1}}]$ pour quelques $k, \ell \geq 1$. Donc

$$x = a_1 + \cfrac{1}{a_2 + \cfrac{1}{\cdots + \cfrac{1}{a_{\ell-1} + \cfrac{1}{y}}}},$$

où $y = [\overline{a_\ell, \dots, a_{k+\ell-1}}]$. Alors on a que

$$y = a_\ell + \cfrac{1}{a_{\ell+1} + \cfrac{1}{\cdots + \cfrac{1}{a_{k+\ell-1} + \cfrac{1}{y}}}} = [a_\ell, a_{\ell+1}, \dots, a_{k+\ell-1}, y].$$

Soient p_i/q_i, $1 \le i \le k+1$, les convergents de la fraction continue $[a_\ell, a_{\ell+1}, \dots, a_{k+\ell-1}, y]$. On a que

$$y = \frac{p_{k+1}}{q_{k+1}} = \frac{yp_k + p_{k-1}}{yq_k + q_{k-1}}$$

selon le théorème 12.2. Donc

$$q_k y^2 + q_{k-1} y = p_k y + p_{k-1} \quad \Longrightarrow \quad q_k y^2 + (q_{k-1} - p_k) y - p_{k-1} = 0.$$

En appliquant la formule quadratique, on conclut que $y = r' + \sqrt{s'}$ pour quelques $r', s' \in \mathbb{Q}$ (nécessairement $s' \ge 0$ car on sait déjà que l'équation quadratique $q_k y^2 + (q_{k-1} - p_k) y - p_{k-1} = 0$ a une solution réelle). Par conséquent, x doit être également de cette forme.

Réciproquement, supposons que $x = r + \sqrt{s}$ pour quelques $r, s \in \mathbb{Q}$ avec $s \ge 0$. Alors il existe $a, b, c \in \mathbb{Z}$ tels que

$$ax^2 + bx + c = 0. \tag{12.4}$$

On a que $a \neq 0$; sinon, on aurait que $x \in \mathbb{Q}$. Mais on déjà traité le cas où $x \in \mathbb{Q}$. Donc $a \neq 0$, comme affirmé. Soit $x = [a_1, a_2, \dots]$. On veut montrer k et ℓ tels que $x = [a_1, \dots, a_{\ell-1}, \overline{a_\ell, \dots, a_{k+\ell-1}}]$. On pose

$$x_n = [a_n, a_{n+1}, \dots]$$

et on observe qu'il suffit de trouver deux nombres différents n_1 et n_2 tels que $x_{n_1} = x_{n_2}$. (Si $n_1 < n_2$, ceci nous permettra de prendre $\ell = n_1$ et $k = n_2 - n_1$.) Soient p_n/q_n les convergents de x. On a que

$$x = [a_1, \dots, a_{n-1}, x_n]$$

et donc le théorème 12.2 implique que

$$x = \frac{x_n p_{n-1} + p_{n-2}}{x_n q_{n-1} + q_{n-2}} \quad (n \ge 3). \tag{12.5}$$

(Notez que si p'_i/q'_i, $1 \le i \le n$, sont les convergents de la fraction continue $[a_1, \dots, a_{n-1}, x_n]$, alors $p'_i = p_i$ et $q_i = q'_i$ pour $i \in \{1, \dots, n-1\}$.) On utilisera cette relation pour montrer que x_n satisfait une équation quadratique $a_n x^2 + b_n x + c_n = 0$ également. Finalement, on montrera que l'ensemble de ces équations quadratiques est, en fait fini. Puisque chaque cette équation a au plus deux racines, ceci nous permettra de montrer que l'ensemble $\{x_n : n \ge 3\}$ et fini (et, par la suite, $x_{n_1} = x_{n_2}$ pour quelques n_1 et n_2 qui sont différents).

Le théorème 12.5 implique que le determinant de la matrice $\begin{pmatrix} p_{n-1} & p_{n-2} \\ q_{n-1} & q_{n-2} \end{pmatrix}$ est ± 1. En général, soient $\alpha, \beta, \gamma, \delta \in \mathbb{Z}$ tels que

$$\det \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \varepsilon = \pm 1.$$

Si

$$x = \frac{\alpha y + \beta}{\gamma y + \delta},$$

alors le nombre y est une racine de l'équation

$$a\left(\frac{\alpha y+\beta}{\gamma y+\delta}\right)^2+b\left(\frac{\alpha y+\beta}{\gamma y+\delta}\right)+c=0,$$

puisque x est une racine de l'équation (12.4). En multipliant par $(\gamma y+\delta)^2$ l'équation au-dessus, on peut l'écrire comme

$$Ay^2+By+C=0,$$

où

$$A=a\alpha^2+b\alpha\gamma+c\gamma^2,\quad B=2a\alpha\beta+b(\alpha\delta+\beta\gamma)+2c\gamma\delta,\quad C=a\beta^2+b\beta\delta+c\delta^2.$$

Quand

$$\begin{pmatrix}\alpha & \beta\\ \gamma & \delta\end{pmatrix}=\begin{pmatrix}p_{n-1} & p_{n-2}\\ q_{n-1} & q_{n-2}\end{pmatrix}$$

pour un $n\geq 3$, comme dans notre cas, on montrera que les nombres A, B et C sont bornés en termes de x, ce qui nous permettra de déduire qu'il y a seulement un nombre fini de possibilités pour les polynômes At^2+Bt+C. Tout d'abord, on note que si $d=b^2-4ac$ et le discriminant du polynôme ax^2+bx+c et $D=B^2-4AC$ est le discriminant du polynôme Ax^2+Bx+C, alors on a que $D=\varepsilon^2 d=d$. Cette relation déjà impose de restrictions sur A, B et C. De plus, si $f(t)=at^2+bt+c$, alors on observe que

$$A=\gamma^2 f\left(\frac{\alpha}{\gamma}\right)=q_{n-1}^2 f\left(\frac{p_{n-1}}{q_{n-1}}\right)\quad\text{et}\quad C=\delta^2 f\left(\frac{\beta}{\delta}\right)=q_{n-2}^2 f\left(\frac{p_{n-2}}{q_{n-2}}\right).$$

Mais $f(x)=0$ par l'hypothèse et $p_i/q_i\to x$ lorsque $i\to\infty$. Alors A et C ne peuvent être très grands. En effet, si on écrit $p_i/q_i=x+h_i$, alors le théorème 12.6(b) implique que $|h_i|<1/(q_iq_{i+1})\leq 1$. Donc, selon le théorème de la valuer moyenne, on trouve qu'il existe un $t_i\in(x-|h_i|,x+|h_i|)\subset[x-1,x+1]$ tel que

$$|f(x+h_i)|=|f(x)+h_if'(t_i)|=|h_if'(t_i)|\leq\frac{1}{q_i^2}\cdot M,$$

où $M=\sup_{x-1\leq t\leq x+1}|f'(t)|$. Par la suite,

$$|A|=q_{n-1}^2|f(x+h_{n-1})|\leq M\quad\text{et}\quad|C|=q_{n-2}^2|f(x+h_{n-2})|\leq M,$$

ce qui implique aussi que

$$|B|=\sqrt{B^2}=\sqrt{d+4AC}\leq\sqrt{d}+2\sqrt{|AC|}\leq\sqrt{d}+2M.$$

Par consequent, on trouve que, étant donné x, il y a seulement un nombre fini de possibilités pour les nombres entiers A, B et C. Alors on déduit que l'ensemble $\{x_n : n\geq 3\}$ est fini, comme affirmé. Ceci conclut la démonstration. □

Exercices

EXERCICE 12.1. Montrez que la fraction continue d'un nombre est uniquement définie.