Introduction à la théorie des nombres

Dimitris Koukoulopoulos

Université de Montréal

Dernière mise-à-jour : 10 octobre 2022

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 $f(x) = a_0+a_1x+\cdots+a_dx^d$ 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 $f(x) = a_0+a_1x+\cdots+a_dx^d$, 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

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

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 $n \leqslant x < n+1$. Une autre propriété de la partie entier qu'on utilisera beaucoup est que si $x > 0$, alors

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

La lettre p dénote toujours un nombre premier (définition 1.1). La notation $a|b$ veut dire que a divise b (définition 2.1). 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 2.4), 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 2.12) ou l'intervalle fermé dont les limites sont a et b (c'est-à-dire l'ensemble $\{x \in \mathbb{R} : a \leqslant x \leqslant b\}$).

Première partie

La structure multiplicative des entiers

Chapitre 1

Nombres premiers

La structure additive des nombres naturels est très simple : chaque nombre $n \in \mathbb{N}$ peut s'écrire comme la somme des 1s :

$$n = \underbrace{1 + 1 + \cdots + 1}_{n \text{ fois}}.$$

En plus, on ne peut pas décomposer n dans des morceaux additifs plus petits.

Par contre, la structure multiplicative des nombres naturels est beaucoup plus compliquée et c'est le premier sujet qu'on étudiera à ce cours.

Prenons comme exemple le nombre 420. On veut le décomposer complètement dans des morceaux minimaux. Puisque le dernier chiffre est zéro, on peut écrire $420 = 10 \cdot 42$. Puis, on sait que $10 = 2 \cdot 5$ et que $42 = 6 \cdot 7$, donc $420 = 2 \cdot 5 \cdot 6 \cdot 7$. Mais, on n'a pas fini, car on peut aussi factoriser 6 comme $2 \cdot 3$. On conclut que

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

On ne peut plus décomposer les facteurs $2, 3, 5, 7$, donc notre tache s'est terminée.

Si on prend un autre nombre, par exemple 299, on peut trouver la factorisation

$$299 = 13 \cdot 23.$$

Les facteurs 13 et 23 ne se factorisent plus, donc on a écrit

Afin de formaliser cette discussion, on introduit la notion des *nombres premiers*.

Définition 1.1. Un nombre naturel $n > 1$ est appelé *composé* s'il existe deux nombres naturels a et b tels que $1 < a, b < n$ et $n = ab$.

Un nombre $n > 1$ est appelé *premier* s'il n'est pas composé.

Remarque. Grâce à leur indécomposabilité, les nombres premiers sont souvent appelés les *atomes de la multiplication*.[1].

Or, cette définition et la discussion qui la précède nous amène naturellement au fait important suivant :

Proposition 1.2. *Chaque nombre naturel $n > 1$ peut s'écrire comme produit de certains nombres premiers.*

1. Le mot *atome* provient du mot grec $\alpha\tau o\mu o$ qui veut dire littéralement ce qui ne se coupe pas.

Démonstration. Si n est déjà premier, alors il n'y a rien à montrer. Supposons alors que n est composé. On peut alors l'écrire comme $n = ab$ avec $1 < a, b < n$. Si a et b sont premiers, on a fini. Sinon, on les écrit comme un produit de deux nombres strictement plus petits. En continuant de cette façon, il faut arriver (après un nombre fini d'étapes) à un produit des nombres premiers.

Afin de formaliser cette preuve et la rendre plus rigoureuse, on utilise induction sur n.

L'étape de base est quand $n = 2$: il s'agit d'un nombre premier, donc le résultat est vrai pour lui.

Puis, on considère un nombre $n > 2$ et on suppose que le résultat est vrai pour chaque entier $m \in \{2, \dots, n-1\}$. Si n est déjà premier, alors le résultat est vrai automatiquement. D'autre côté, si n est composé, il s'écrit comme $n = ab$ avec $1 < a, b < n$. En utilisant l'hypothèse d'induction, on voit que a et b sont de produits de certains nombres premiers. Par la suite, n l'est aussi. Ceci conclut l'étape inductive et donc la preuve. □

En voyant ce résultat, il y a des questions naturelles qui se posent :

— *Question 1 :* Combien de façons existe-t-il d'écrire n comme produit de nombres premiers ?
— *Question 2 :* Combien de nombres premiers existe-t-il ?
— *Question 3 :* Est-ce qu'il y a une façon simple de décrire et d'identifier les nombres premiers ?

Comme on va le voir, la réponse à la Question 1 est que chaque entier possède seulement une factorisation en nombre premiers (si on ignore le fait qu'on peut permuter les facteurs). Ceci est une propriété fondamentale des entiers :

Théorème 1.3 (Théorème fondamental de l'arithmétique)**.** *Chaque nombre naturel $n > 1$ peut s'écrire comme produit de nombres premiers. De plus, cette factorisation est unique, sauf pour les permutations possibles des facteurs.*

Remarque. En termes plus formels, l'unicité de la factorisation première veut dire que si $n = p_1 \cdots p_r = q_1 \cdots q_s$ pour quelques premiers $p_1, \dots, p_r$ et $q_1, \dots, q_s$, alors $r = s$ et il existe une permutation $\sigma \in S_r$ telle que $p_i = q_{\sigma(i)}$ pour tout $i \in \{1, \dots, r\}$.

Afin de montrer le théorème 1.3, il faut d'abord développer certains outils importants, ce qui sera l'objectif du prochain chapitre.

En ce qui concerne les deux autres questions, notre compréhension est beaucoup plus limitée. On donne une première réponse à la deuxième avec le résultat suivant.

Théorème 1.4 (Euclide)**.** *Il existe une infinité de nombres premiers.*

Démonstration. Supposons, au contraire, qu'il existe qu'une liste finie de nombres premiers, soit $p_1, \dots, p_k$. Puis, on considère le nombre $n = 1 + p_1 \cdots p_k$. Puisque la liste $p_1, \dots, p_k$ comprend tous les nombres premiers, la proposition 1.2 implique que $n = p_1^{\nu_1} \cdots p_k^{\nu_k}$ pour quelques entiers $\nu_1, \dots, \nu_k \geqslant 0$. Il faut que $\nu_i \geqslant 1$ pour au moins un indice i, car $n > 1$. Mais, dans ce cas-ci on voit que p_i est un facteur commun de n et de $p_1 \cdots p_k$, ce qui implique que $n - p_1 \cdots p_i$ peut s'écrire comme $p_i m$ avec $m \in \mathbb{Z}$. D'autre côté, on a que $n - p_1 \cdots p_k = 1$, donc $1 = p_i m$. Cette relation ne peut pas être vraie, car elle implique aussi que $m \neq 0$, et donc que $|p_i m| \geqslant p_i \geqslant 2$. (On utilise ici que chaque premier est > 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 théorème 1.4 répond partiellement à la Question 2 mais, au même temps, il suscite une nouvelle question plus raffinée :

— *Question 2 raffinée :* Peut-on donner une approximation précise de la fonction de comptage

$$\pi(x) := \#\{p \leqslant x : p \text{ premier}\} \ ?$$

Observons que si p_n dénote le n-ième premier (donc, $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, etc.), alors on a que $\pi(p_n) = n$. Donc, si nous avons une bonne approximation pour $\pi(x)$, nous en trouvons une pour p_n également, ce qui répond partiellement à la Question 3 ci-dessus.

Il y a une façon alternative d'approcher la Question 3 : étant donné un entier n, est-ce qu'il y a un algorithme rapide qui détermine si n est premier ou non ? En aillant même plus loin, on peut aussi se demander s'il existe un algorithme qui détermine toute la factorisation première de n s'il est composé. On va revenir à ces deux questions à la section 4.

1.1 Exercices

EXERCICE 1.1. Montrez que, pour chaque entier $n \geqslant 1$, le nombre $4n^3+6n^2+4n+1$ est composé. [*Indice :* Développez $(x+y)^4$.]

EXERCICE 1.2.

(a) Montrer que si le nombre $2^n - 1$ est premier, alors il faut que n soit premier aussi.

(b) Montrer que si le nombre $2^n + 1$ est premier, alors il faut que $n = 2^k$ pour un entier $k \geqslant 0$.

Chapitre 2

Diviseurs et multiples

Dans cette section, on développe les outils nécessaires afin de montrer le théorème fondamental de l'arithmétique. Plusieurs notions qu'on verra sont très utiles elles-mêmes et jouent un rôle central en théorie des nombres.

Définition 2.1 (Divisibilité)**.** 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 s'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 2.2. *Soit $a, b \in \mathbb{Z}$ avec $a \neq 0$.*

(a) Si $a|b$, alors soit $b = 0$ soit $|b| \geqslant |a|$.

(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 $a|b$.

Démonstration. (a) Supposons que $b \neq 0$. Puisque $a|b$, il existe $k \in \mathbb{Z}$ tel que $b = ka$. Puisque $b \neq 0$, il faut aussi que $k \neq 0$. Par la suite, $|k| \geqslant 1$, d'où on déduit que $|b| \geqslant |a|$.

On laisse les autres parties comme exercices. □

2.1 Division euclidienne

En général, si nous avons deux entiers a, b, il est possible que $a \nmid b$. Dans ce cas-ci, on peut quand même bien approximer b par un multiple de a. Ceci est le contenu du résultat suivant.

Théorème 2.3 (Division euclidienne)**.** *Soit $a, b \in \mathbb{Z}$ avec $a \neq 0$. Il existe $q, r \in \mathbb{Z}$ uniques tels que*

$$b = qa + r \quad \textit{et} \quad 0 \leqslant r < |a|.$$

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 $a > 0$.

On commence en prouvant l'existence de q et de r. Considérons tous les multiples de a, c'est-à-dire les nombres ka avec $k \in \mathbb{Z}$. Définissons $q = \max\{k \in \mathbb{Z} : ka \leqslant b\}$, pour que $qa \leqslant b < (q+1)a$. Si on pose $r := b - qa$, on a alors que $0 \leqslant r < a$ et que $b = qa + r$. Ceci montre l'existence de q et de r.

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

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

ce qui implique que $a|r' - r$. Cependant $|r' - r| < a$ par notre hypothèse que $0 \leqslant r, r' < a$. Donc lemme 2.2(a) implique que $r' = r$. Donc on trouve que $(q - q')a = 0$ et, puisque $a \geqslant 1$, alors $q' = q$ aussi. □

2.2 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 2.4 (Le plus grand commun diviseur)**.** Soient $a, b \in \mathbb{Z}$ qui ne sont pas les deux égaux à 0. On définit leur *plus grand commun diviseur* d'être le nombre

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

Si $(a, b) = 1$, alors on dit que a et b sont *co-premiers*.

Remarque. 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 2.5 (Périodicité du pcgd)**.** *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 $b \geqslant a$. On écrit $b = qa + r$ avec $0 \leqslant r < a$ 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 $0 \leqslant r' < r$ 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

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

Exemples. (a) Calculons $(91, 65)$. On a que

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

(b) 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 2.6 (le lemme de Bezout)**.** *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. 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 deux exemples qu'on a étudié avant le lemme de Bezout. 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

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

La puissance du théorème 2.6 est révélée dans les démonstration des résultats suivants, qui sont très utils.

Lemme 2.7. *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 2.2(b).

Réciproquement, supposons que $d|(a, b)$. Par définition, on a que $(a, b)|a$ et que $(a, b)|b$. Donc lemme 2.2(d) implique que $d|a$ et que $d|b$. □

Lemme 2.8. *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 $b = d\ell$. On affirme que $(k, \ell) = 1$. En effet, si on avait $(k, \ell) > 1$, il existerait $g > 1$ divisant k et ℓ. Donc, on aurait que $dg > d$ est un diviseur commun de a et b, ce qui est impossible de notre hypothèse que d est le *plus grand* commun diviseur de a de b. □

Lemme 2.9 (le lemme d'Euclide)**.** *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 2.10. *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'Euclide nous donne que $a|k$. Alors, $k = a\ell$ pour un $\ell \in \mathbb{N}$. Ceci implique que $c = \ell ab$, ce qui conclut la démonstration. □

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

Définition 2.11 (pgcd de k entiers)**.** 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

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

2.3 Le plus petit commun multiple

On introduit maintenant une notion «duale» à la celle du plus grand commun diviseur :

Définition 2.12 (Le plus grand commun multiple)**.** 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. 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 2.13. *Si $a, b \in \mathbb{N}$, alors $[a, b](a, b) = ab$.*

Démonstration. Soit $m = [a, b]$ et $d = (a, b)$. D'après la proposition 2.8, 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 $m \leqslant dk\ell$. 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 2.10 implique que $k\ell|m/d$. En particulier, $k\ell \leqslant m/d$, ce qui conclut la preuve que $m = dk\ell$. □

Lemme 2.14. *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 2.3 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 2.15. 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

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

2.4 Exercices

EXERCICE 2.1. Soit $a, b \in \mathbb{N}$. Montrez que $a = b$ si et seulement si $a|b$ et $b|a$.

EXERCICE 2.2. Soient $a, b \in \mathbb{N}$ tels que $1/a + 1/b$ est entier. Montrez que soit $a = b = 1$ soit $a = b = 2$.

EXERCICE 2.3. 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 2.4. Soient $a, b, c \in \mathbb{N}$. Montrer que $(ab, ac) = a \cdot (b,c)$.

EXERCICE 2.5. Soit $a, b, c \in \mathbb{N}$ tels que $a|bc$. Montrer que $\frac{a}{(a,b)}|c$.

EXERCICE 2.6. Montrez que si $(a,c) = 1$, alors $(ab,c) = (b,c)$.

EXERCICE 2.7. Soient $a, b, c \in \mathbb{N}$ avec $(a,b) = 1$. Si $(a,c) = (b,c) = 1$, alors montrer que $(ab,c) = 1$.

EXERCICE 2.8. Définissons la suite de Fibonacci $(F_n)_{n=0}^{\infty}$ par les relations suivantes : $F_0 = 1$, $F_1 = 1$ et $F_n = F_{n-1} + F_{n-2}$ pour tout $n \geqslant 2$. Montrer que $(F_n, F_{n-1}) = 1$ pour chaque $n \geqslant 1$.

EXERCICE 2.9. Soient $1 \leqslant k \leqslant n$ et $d = (n,k)$. Montrez que $\frac{n}{d}|\binom{n}{k}$.

EXERCICE 2.10. 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 2.11. Montrez que $((a+b)^2, a^2+ab+b^2) = (a,b)^2$, pour tous $a, b \in \mathbb{N}$.

EXERCICE 2.12. (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 $d|(a_1, \ldots, a_k)$.

EXERCICE 2.13. 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 $[a_1, \ldots, a_k] = a_1 \cdots a_k$.

EXERCICE 2.14. (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 2.15. Soit $b \geqslant 2$. Montrer que chaque entier $n \geqslant 0$ a un *developpement b-adique* : il existe de coefficients uniques $c_0, c_1, \ldots, c_k \in \{0, 1, \ldots, b-1\}$ tels que $n = c_0 + c_1 b + c_2 b^2 + \cdots + c_k b^k$.

EXERCICE 2.16. Pour tous $a, b \in \mathbb{N}$, montrer que $(2^a - 1, 2^b - 1) = 2^{(a,b)} - 1$.

Chapitre 3

Le théorème fondamental de l'arithmétique

Le but de ce chapitre est de montrer le théorème fondamental de l'arithmétique (cf. théorème 1.3). On commence avec un résultat préparatoire.

Lemme 3.1. *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'Euclide 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. On a déjà montré l'existence d'une factorisation première de n à la proposition 1.2. Il reste alors à montrer que cette factorisation est unique. Soient deux factorisations de n, $n = p_1 \cdots p_r = q_1 \cdots q_s$. Bien sûr, on a que $q_1|p_1 \ldots p_r$. Donc, le lemme 3.1(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 3.1(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, de façon inductive, on trouve que il existe s indices $i_1, \ldots, i_s$ distincts appartenant 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

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

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

Corollaire 3.2. *Si $n > 1$, alors il existe nombres premiers $p_1 < \cdots < p_r$ uniques et exposants $v_1, \ldots, v_r \in \mathbb{N}$ uniques tels que $n = p_1^{\nu_1} \cdots p_r^{\nu_r}$.*

Démonstration. Soit $n = q_1 \cdots q_s$ la factorisation de n en nombres premiers. Il est possible que $q_i = q_j$ pour $i \neq j$. Soient $p_1 < p_2 < \cdots < p_r$ les nombres premiers distincts parmi $q_1, \ldots, q_s$. Pour chaque p_i, soit ν_i le nombre de fois qu'il apparaît dans la liste $q_1, \ldots, q_s$. On a alors que

$$n = q_1 \cdots q_s = p_1^{\nu_1} \cdots p_r^{\nu_r}.$$

Finalement, l'unicité des nombres premiers $p_1, \ldots, p_r$ et des exposants $\nu_1, \ldots, \nu_r$ est une conséquence directe de l'unicité de $q_1, \ldots, q_s$ (à part d'une permutation possible). □

La factorisation $n = p_1^{\nu_1} \cdots p_r^{\nu_r}$ avec $p_1 < \cdots < p_r$ est appelée la *factorisation première* de n. Motivés par elle, on définit pour chaque $n \in \mathbb{N}$ et chaque nombre premier p, la *valuation de n au p*

$$v_p(n) := \max\{\nu \in \mathbb{Z}_{\geqslant 0} : p^\nu | n\}.$$

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 3.2 implique tout de suite que

$$n = \prod_p p^{v_p(n)}.$$

Souvent, étant donné un nombre premier p, on utilisera la notation $p^\nu \| n$, qui veut dire que la plus grande puissance de p qui divise n est égale à ν, c'est-à-dire $p^\nu | n$ et $p^{\nu+1} \nmid n$. Bien sûr, on a que

$$p^\nu \| n \quad \text{si et seulement si} \quad \nu = v_p(n).$$

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

(a) On a que $a|b$ si et seulement si $v_p(a) \leqslant v_p(b)$, 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 $v_p(a) \leqslant v_p(b)$.

Réciproquement, supposons que $v_p(a) \leqslant v_p(b)$ 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 $v_p(d) \leqslant v_p(a)$ et $v_p(d) \leqslant v_p(b)$ pour tout p, si et seulement si $v_p(d) \leqslant \min\{v_p(a), v_p(b)\}$ 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 $v_p(a) \leqslant v_p(m)$ et $v_p(b) \leqslant v_p(m)$ pour tout p, si et seulement si $\max\{v_p(a), v_p(b)\} \leqslant v_p(m)$ 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). □

3.1 Exercices

EXERCICE 3.1. (a) Soit $k \geqslant 2$ et $n \in \mathbb{N}$ dont la première factorisation est $n = p_1^{\nu_1} \cdots p_r^{\nu_r}$. Montrez que n est une k-ième puissance parfaite si est seulement si $k|\nu_i$, 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 3.2. Soient $k \geqslant 2$ et $a, b \in \mathbb{N}$. Montrez que $a^k|b^k$ si et seulement si $a|b$.

EXERCICE 3.3. Un nombre n est appelé *sans facteur carré* 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 facteur 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 $n = p_1^{\nu_1} \cdots p_r^{\nu_r}$ pour quelques nombres premiers distincts $p_1, \ldots, p_r$ et quelques exposants $\nu_1, \ldots, \nu_r \geqslant 2$.

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

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

EXERCICE 3.4. Combien de zéros existent à la fin (côté droit) du développement décimal du nombre $1000! = 1 \cdot 2 \cdots 1000$?

EXERCICE 3.5. Si $n = p_1^{\nu_1} \cdots p_r^{\nu_r}$ est la première factorisation de n, alors on pose

$$\omega(n) = r \quad \text{et} \quad \Omega(n) = \nu_1 + \cdots + \nu_r,$$

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

(a) Montrer que $v_p(mn) = v_p(m) + v_p(n)$ et que $\Omega(mn) = \Omega(m) + \Omega(n)$ pour tous $m, n \in \mathbb{N}$.

(b) Montrer que si m et n sont deux nombres naturels co-premiers, alors $\omega(mn) = \omega(m) + \omega(n)$.

EXERCICE 3.6 (Furstenberg). On muni $\mathbb{Z}$ d'une topologie, comme suit :

— $\emptyset$ est ouvert ;

— pour tout $a, b \in \mathbb{Z}$, la progression arithmétique $S(a, b) := \{an + b : n \in \mathbb{Z}\}$ est ouverte ;

— les réunions d'ensembles ouverts sont ouvertes.

Montrer que :

(a) Les progressions arithmétiques sont également fermées.

(b) Il n'existe pas d'ensemble ouvert qui est fini et non-vide.

(c) $\mathbb{Z} \setminus \{-1, 1\}$ n'est pas fermé.

(d) $\mathbb{Z} \setminus \{-1, 1\} = \bigcup_{p \text{ premier}} S(p, 0)$.

Conclure qu'il existe une infinité de nombres premiers.

‘

Chapitre 4

Aspects algorithmiques de la multiplication

4.1 Le crible d'Eratosthène

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'Eratosthène*. Elle se base sur le résultat suivant.

Théorème 4.1. *Si $n > 1$ est composé, alors il existe un nombre premier $p \leqslant \sqrt{n}$ qui divise n.*

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

À partir de ce théorème, on peut construire un algorithme qui détermine 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 $n \in \left(1, \sqrt{x}\right]$ pas encore encerclé ou supprimé et encercler-le. Si un tel n n'existe pas, terminer l'algorithme.
Étape 3 : Supprimer tous les multiples de n qui sont strictement plus grands que n.
Étape 4 : Retourner à la deuxième étape.

Les nombres qui ne sont pas supprimés quand cet algorithme termine sont exactement les nombres premiers appartenant à $[1, x]$.

Exemple. Trouvons tous les nombres premiers jusqu'à $x = 40$. 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. ■

5

4.2 La vitesse de la division euclidienne

Soit $a, b \in \mathbb{N}$, et soit q et r le quotient et le reste, respectivement, de la division euclidienne de b par a, c'est-à-dire $b = qa + r$ avec $0 \leqslant r < b$. On veut être capable de calculer q et r rapidement. Supposons que les nombres a et b sont donnés en base de 10 (l'argument est similaire dans la base de 2 dans laquelle les ordinateurs fonctionnent).

Si $a < b$, alors $q = 0$ et $r = a$, donc il n'y a rien à calculer.

Supposons alors que $a \geqslant b$. On va calculer q en utilisant un argument itératif. Supposons que a et b ont k et ℓ chiffres décimaux, respectivement. Donc,

$$10^{k-1} \leqslant a < 10^k \quad \text{and} \quad 10^{\ell-1} \leqslant b < 10^{\ell}. \tag{4.1}$$

Observons alors que

$$10^{k-\ell-1}b < a < 10^{k-\ell+1}.$$

Calculons les nombres $10^{k-\ell-1}b, 10^{k-\ell}b, 10^{k-\ell+1}b$ car on les aura besoin en dessous. Puisque on travaille dans la base de 10, ce calcul est très facile : il faut tout simplement ajouter $k-\ell-1, k-\ell$ ou $k-\ell+1$ zéros à la fin du développement décimal de b. Or, afin de continuer, on distingue deux cas :

Cas 1 : $10^{k-\ell}b \leqslant a < 10^{k-\ell+1}b$. Dans ce cas-ci, posons $n_1 := k-\ell$ et $a_1 := a - 10^{n_1}b$, et observons que $0 \leqslant a_1 < \frac{9}{10}a$.

Cas 2 : $10^{k-\ell-1}b < a < 10^{k-\ell}b$. Dans ce cas-ci, posons $n_1 = k-\ell-1$ et $a_1 := a - 10^{n_1}b$, et observons que $0 < a_1 < \frac{9}{10}a$.

Si $a_1 < b$, alors notre algorithme a terminé et nous avons que $q = 10^{n_1}$ et que $r = a_1$. Sinon, on répète l'argument ci-dessus afin de trouver $n_2 \in \mathbb{Z}_{\geqslant 0}$ tel que le nombre $a_2 := a_1 - 10^{n_2}b$ se trouve dans $[0, \frac{9}{10}a_1)$. En continuant de cette façon, on construit éventuellement certains nombres $n_1, n_2, \ldots, n_J \in \mathbb{Z}_{\geqslant 0}$ tels que nous avons les propriétés suivantes : si on pose $a_0 = a$ et $a_j = a_{j-1} - 10^{n_j}$ pour $j = 1, \ldots, J$, alors $0 \leqslant a_J < b \leqslant a_{J-1} < \cdots < a$. De plus, nous avons que $a_{j+1} \leqslant \frac{9}{10}a_j$ pour $j = 0, 1, \ldots, J-1$.

Un argument inductif facile implique que $a_j = a - (10^{n_1} + \cdots + 10^{n_j})b$. Puisque $0 \leqslant a_J < b$, il faut alors que $r = a_J$ et que $q = 10^{n_1} + 10^{n_2} + \cdots + 10^{n_J}$.

Pour effectuer chaque étape, on a fait 4 calculs : on a trouvé $10^{k-\ell-1}b, 10^{k-\ell}b, 10^{k-\ell+1}b$ et $a - 10^{n_1}b$. Il reste à borner J, le nombre d'étapes de l'algorithme. Afin de le faire, nous observons que nous avons les inégalités

$$1 \leqslant b \leqslant a_{J-1} < (9/10)^{J-1}a,$$

où la dernière inégalité découle du fait que $a_{j+1} \leqslant \frac{9}{10}a_j$ pour $j = 0, 1, \ldots, J-1$, en utilisant un argument inductif. On en déduit que

$$J < 1 + \frac{\log(a/b)}{\log(10/9)} \leqslant 1 + \frac{\log(10)}{\log(10/9)}(k-\ell+1),$$

d'après (4.1). On voit alors que le nombre d'étapes que cet algorithme prend pour terminer et déterminer q et r est linéaire dans $k-\ell$, le différence du nombre de chiffres de a et de b. De plus, le nombre de calculs exigés est

$$< 4J < 4 + \frac{4\log(10)}{\log(10/9)}\log(b/a) \leqslant 4 + \frac{4\log(10)}{\log(10/9)}(k-\ell+1). \tag{4.2}$$

On appelle un tel algorithme *linéaire*.

4.3 L'algorithme euclidien

En utilisant la discussion de la section précédente, on peut maintenant analyser la vitesse de l'algorithme euclidien. On se rappelle que le but est de calculer le pgcd de deux nombres naturels a et b. Supposons sans perte de généralité que $a \leqslant b$. Posons

$$b_0 = b \quad \text{et} \quad b_1 = a.$$

Puis, faisons la division euclidienne de b_0 par b_1, en trouvant que

$$b_0 = q_1 b_1 + b_2$$

pour quelques $b_2 \in \{0, 1, \dots, b_1 - 1\}$ et $q_1 \in \mathbb{N}$ (le fait que $q_1 \geqslant 1$ est grâce à notre hypothèse que $b \geqslant a$). Si $b_2 = 0$, ce qui veut dire que $b|a$, alors l'algorithme a terminé et on trouve que $(a, b) = b = 0 \cdot a + 1 \cdot b$. D'autre côté, si $1 \leqslant b_2 < b_1$, on utilise le fait que $(a, b) = (b_1, b_2)$ et répète l'argument avec b_1 et b_2 au lieu de b_0 et de b_1, respectivement. En itérant cet argument, on construit une suite

$$b_0 \geqslant b_1 > b_2 > \cdots \geqslant 0.$$

Après un nombre fini d'étapes, soit n, il faut avoir $b_{n+1} = 0$ pour que l'algorithme se termine et on ait que $(a, b) = b_n$. On veut donner une borne supérieure à n afin d'estimer la vitesse de l'algorithme.

Pour chaque $m \leqslant n - 2$, il existe $q_m \in \mathbb{Z}$ tel que

$$b_m = q_{m+1} b_{m+1} + b_{m+2}.$$

Puisque $b_m \geqslant b_{m+1}$, on a que $q_m \geqslant 1$. Donc,

$$b_m \geqslant b_{m+1} + b_{m+2} > 2b_{m+2}$$

pour tous $m \in \{1, \dots, n-2\}$. Un argument itératif montre alors que $b_{2k} \leqslant b_2/2^{k-1} \leqslant 2a/\sqrt{2}^{2k}$ et que $b_{2k-1} \leqslant b_1/2^{k-1} = \sqrt{2}a/\sqrt{2}^{2k-1}$, pour tout $k \geqslant 1$. En particulier, on a que $1 \leqslant b_n \leqslant 2a/\sqrt{2}^n$, ce qui implique que $2^{n/2} \leqslant 2a$. Donc, le nombre total d'étapes que l'algorithme a besoin est

$$n \leqslant 2 + \frac{2 \log a}{\log 2}. \tag{4.3}$$

Cependant, la vraie vitesse de l'algorithme ne dépend pas seulement du nombre d'étapes mais du nombre de calculs exigés au total. D'après la relation (4.2), le calcul de b_{m+2} prend $\leqslant 4 + C \log(b_{m+1}/b_m)$ sous-calculs, où $C = 4 \log(10)/\log(10/9) < 87,5$. Donc le calcul de b_n prend

$$\leqslant \sum_{m=0}^{n-2} \big(4 + C \log(b_{m+1}/b_m)\big) \leqslant 4(n-1) + C \log a \leqslant 4 + \Big(\frac{8}{\log 2} + C\Big) \log a$$

sous-calculs. On voit alors que l'algorithme euclidien est linéaire au nombre de chiffres décimaux du plus petit nombre du paire (a, b).

4.4 Exercices

EXERCICE 4.1. Analyser la vitesse de la division euclidienne dans la base de 2.

EXERCICE 4.2. Dans l'algorithme euclidien, montrer que

$$b_{n-k} \geqslant F_{k+1} b_n \quad \text{pour } k = 0, 1, \dots, n,$$

où F_k dénote le k-ième nombre de Fibonacci (donc $F_1 = F_2 = 1$ et $F_k = F_{k-1} + F_{k-2}$ pour $k \geqslant 2$).

Conclure que $F_n \leqslant a$ et améliorer la borne (4.3).

Chapitre 5

Fonctions arithmétiques

Une fonction $f : \mathbb{N} \to \mathbb{C}$ est appelée une *fonction arithmétique*. On peut l'identifier si on le veut avec la suite de ses valeurs $f(1), f(2), f(3), \ldots$. L'étude de ces fonctions joue un rôle central dans la théorie des nombres, car on peut poser plusieurs questions importantes en termes certaines fonctions arithmétiques.

Voici quelques exemples importants des fonctions arithmétiques :

— La *fonction «nombre de diviseurs»* $\tau(n) := \#\{d \in \mathbb{N} : d|n\}$;
— Les fonctions $\omega(n) = \#\{p \text{ premier} : p|n\}$ et $\Omega(n) = \sum_{p^\nu \| n} \nu$;
— La *fonction «somme de diviseurs»* $\sigma(n) := \sum_{d|n} d$;
— La *fonction ϕ d'Euler*, définie par $\phi(n) := \#\{1 \leqslant a \leqslant n : (a, n) = 1\}$ pour tout $n \in \mathbb{N}$;
— La *fonction μ de Möbius*, supportée sur les entiers n sans facteur carré, où elle vaut $(-1)^{\omega(n)}$;
— La *fonction Λ de von Mangoldt*, supportée sur les puissances de nombres premiers p^ν, où elle vaut $\log p$.

5.1 Fonction multiplicatives

Si les valeurs d'une fonction arithmétique respectent la structure multiplicative de $\mathbb{N}$, on appelle cette fonction *multiplicative* :

Définition 5.1 (Fonction multiplicative)**.** 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

$$f(mn) = f(m)f(n) \quad \text{quand} \quad (m, n) = 1. \tag{5.1}$$

Si la relation (5.1) tient pour tous $m, n \in \mathbb{N}$, on dit que f est *complètement multiplicative*.

Remarque 5.2. Il peut être contre-intuitif au début qu'on exige que $f(mn) = f(m)f(n)$ seulement si $(m, n) = 1$. La raison est que quand $(m, n) = 1$, alors plusieurs phénomènes multiplicatives reliés à la structure multiplicative de m se comportent de façon «indépendante» des phénomènes correspondant à n. On revient à ce sujet à la section 9.2.

Exemples. (a) La fonction-constante 1 est complètement multiplicative.

(b) La fonction constante 0 n'est pas multiplicative.

(c) Si $\alpha \in \mathbb{C}$ est une constante, alors la fonction $\mathbb{N} \ni n \to n^\alpha \in \mathbb{C}$ est complètement multiplicative. (Si α est complexe, on définit $n^\alpha := e^{\alpha \log n}$; donc, la multiplicativité de la fonction découle du fait que $\log(mn) = \log m + \log n$ pour tous $m, n \in \mathbb{N}$.)

(d) La fonction nombre de diviseurs τ est multiplicative. En effet, soient m et n deux nombres naturels co-premiers. On construit une bijection

$$f : \{(d_1, d_2) \in \mathbb{N} \times \mathbb{N} : d_1|m,\ d_2|n\} \to \{d \in \mathbb{N} : d|mn\}, \tag{5.2}$$

définie par $f(d_1, d_2) := d_1 d_2$. Cette fonction est évidemment bien définie, car si $d_1|m$ et $d_2|n$, alors $d_1 d_2|mn$.

Or montrons que f est aussi surjective. Soit d un diviseur de mn. Le fait que m et n sont co-premiers veut dire qu'on peut diviser les facteurs premiers de d dans deux ensembles *disjoints* : ceux qui divisent m et ceux qui divisent n. On peut alors décomposer $d = d_1 d_2$, où $d_1 = \prod_{p^\nu \| d,\ p|m} p^\nu$ et $d_2 = \prod_{p^\nu \| d,\ p|n} p^\nu$. Par la suite, on a que d_1 est le pgcd de d et de m ; en particulier, $d_1|m$. Idem, on a aussi que $d_2 = (d, n)|n$. On a montré alors que $d = f(d_1, d_2)$, et donc que d est dans l'image de f. Ceci montre que f est surjective.

Finalement, montrons que f est injective. En effet, supposons que $d_1 d_2 = e_1 e_2$, où $d_1, e_1|m$ et $d_2, e_2|n$. Donc, $(d_1, e_2) = 1$. Puisque $d_1|e_1 e_2$, il faut que $d_1|e_1$ d'après le lemme 2.9 d'Euclide. En inversant les rôles de d_i et de e_i, on peut ainsi montrer que $e_1|d_1$. Donc, $d_1 = e_1$ (voir l'exercice 2.1). Puisque $d_1 d_2 = e_1 e_2$, il faut aussi que $d_2 = e_2$. Ceci montre que f est injective.

On a montré alors que f est une bijection. En particulier, son domaine doit avoir la même cardinalité que son image, c'est-à-dire on a que $\tau(m)\tau(n) = \tau(mn)$, ce qui est ce qu'on voulait démontrer.

(e) On peut généraliser la construction de la partie (d). Pour chaque $\alpha \in \mathbb{C}$, définissons la fonction arithmétique $\sigma_\alpha : \mathbb{N} \to \mathbb{C}$ par

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

En particulier, $\sigma_0 = \tau$ et $\sigma_1 = \sigma$.

Or, on montre que σ_α est multiplicative pour chaque $\alpha \in \mathbb{C}$. Soient $m, n \in \mathbb{N}$ avec $(m, n) = 1$. En utilisant le fait que la fonction f définie ci-dessus est une bijection, on a que

$$\sigma_\alpha(mn) = \sum_{d|mn} d^\alpha = \sum_{d_1|m,\ d_2|n} \sum d_1^\alpha d_2^\alpha = \Big(\sum_{d_1|m} d_1^\alpha\Big)\Big(\sum_{d_2|n} d_2^\alpha\Big) = \sigma_\alpha(m)\sigma_\alpha(n).$$

Ceci montre que σ_α est multiplicative.

(f) Si $\alpha \in \mathbb{C}$, alors l'exercice 3.5 implique que la fonction α^ω est multiplicative et que la fonction α^Ω est complètement multiplicative.

(g) La fonction μ de Möbius est multiplicative. En effet, soient $m, n \in \mathbb{N}$ avec $(m, n) = 1$.

Supposons d'abord que $\mu(mn) = 0$, ce qui veut dire qu'il existe un nombre premier p tel que $p^2|mn$. Puisque m et n sont co-premiers, soit $p^2|m$ (auquel cas $\mu(m) = 0$), ou $p^2|n$ (auquel cas $\mu(n) = 0$). Dans tous les cas, on trouve que $\mu(m)\mu(n) = 0 = \mu(mn)$.

Or, supposons que $\mu(mn) \neq 0$, ce qui veut dire que mn est sans facteur carré. Donc, m et n ont la même propriété. On trouve alors que $\mu(m) = (-1)^{\omega(m)}$ et $\mu(n) = (-1)^{\omega(n)}$. Puisque $(m, n) = 1$, alors l'exercice 3.5 implique que $\omega(mn) = \omega(m) + \omega(n)$, et donc que $\mu(mn) = \mu(m)\mu(n)$. ■

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

Théorème 5.3. *Soit f une fonction multiplicative. Alors, on a que*

$$f(1) = 1 \quad \textit{et} \quad f(n) = \prod_{p^\nu \| n} f(p^\nu) \quad \textit{pour tout } n \geqslant 2. \tag{5.3}$$

Si, de plus, f est complètement multiplicative, alors

$$f(n) = \prod_{p^\nu \| n} f(p)^\nu \quad \textit{pour tout } n \geqslant 2. \tag{5.4}$$

Remarque. La converse du théorème 5.3 est aussi vraie : si f est une fonction arithmétique satisfaisant (5.3), alors elle est multiplicative. Si, de plus, elle satisfait (5.4), alors elle est complètement multiplicative.

Démonstration. 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, $f(1) = 1$.

Puis, on montre que $f(n) = \prod_{p^\nu \| n} f(p^\nu)$. On utilise induction sur le nombre de facteurs distincts de n, qu'on dénote par $\omega(n)$ (voir l'exercice 3.5). Si $\omega(n) = 1$, alors n est déjà une puissance première, donc il n'y a rien à montrer. Supposons maintenant que la proposition est vraie quand $\omega(n) \leqslant r - 1$, et considérons n avec $\omega(n) = r$, soit $n = p_1^{\nu_1} \cdots p_r^{\nu_r}$. En utilisant la multiplicativité de f, et puis l'hypothèse d'induction, on a que

$$f(n) = f(p_1^{\nu_1} \cdots p_{r-1}^{\nu_{r-1}}) \cdot f(p_r^{\nu_r}) == \left(f(p_1^{v_1}) \cdots f(p_{r-1}^{\nu_{r-1}})\right) \cdot f(p_r^{v_r}).$$

Ceci termine la démonstration de la première partie.

Finalement, si f est complètement multiplicative, alors

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

pour chaque $\nu \in \mathbb{N}$ par induction, ce qui conclut la démonstration. □

Corollaire 5.4. *On a que*

$$\tau(n) = \prod_{p^\nu \| n} (\nu + 1) \quad \textit{et} \quad \sigma(n) = \prod_{p^\nu \| n} \frac{p^{\nu+1} - 1}{p - 1}.$$

Démonstration. Exercice. □

5.1.1 Nombres parfaits

On conclut cette section avec une application classique de la théorie des fonctions multiplicatives qui caractérise les nombres parfaits pairs :

Définition 5.5. 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 Euclide et plusieurs autres personnes pendant les années. En fait, Euclide 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 Euclide et Euler :

Théorème 5.6 (Euclide, Euler)**.** *Un nombre pair n est parfait si et seulement s'il $n = 2^{p-1}(2^p - 1)$ pour quelque nombre premier p tel que $2^p - 1$ est également un nombre premier.*

Remarque. Les nombres premiers p pour lesquels $2^p - 1$ est aussi premier sont appelés *nombres premiers de Mersenne*. Entre autres, ils sont important dans la recherche par ordinateurs de grands nombres premiers.

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 pair. On écrit $n = 2^\nu m$, où $\nu \geqslant 1$ et m est impair. On a que $\sigma(n) = \sigma(2^\nu)\sigma(m) = (2^\nu - 1)\sigma(m)$. D'autre coté, on a que $\sigma(n) = 2n = 2^{\nu+1}m$, c'est-à-dire $(2^{\nu+1} - 1)\sigma(m) = 2^{\nu+1}m$. Puisque $(2^{\nu+1}, 2^{\nu+1} - 1) = 1$ et $2^{\nu+1} - 1$ divise $2^{\nu+1}m$, alors le lemme 2.9 d'Euclide implique que $m = (2^{\nu+1} - 1)\ell$ pour quelque $\ell \in \mathbb{N}$. Donc, on trouve que

$$2^{\nu+1}(2^{\nu+1} - 1)\ell = (2^{\nu+1} - 1)\sigma(m)$$

et, par la suite, que

$$2^{\nu+1}\ell = \sigma(m) = \sigma((2^{\nu+1} - 1)\ell)$$

On affirme que $\ell = 1$. Sinon, les nombres $1, \ell$ et m seraient tous de diviseurs distincts de $m = (2^{\nu+1} - 1)\ell$ et, par conséquent, on aurait que

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

ce qui est impossible. Donc $\ell = 1$, comme affirmé, ce qui implique que

$$\sigma(2^{\nu+1} - 1) = 2^{\nu+1}.$$

La relation au dessus implique tout de suite que $2^{\nu+1}-1$ est un nombre premier ; sinon, il existerait un entier $d|2^{\nu+1}-1$ tel que $1<d<2^{\nu+1}-1$ et, par conséquent,

$$\sigma(2^{\nu+1}-1) \geqslant 1+d+(2^{\nu+1}-1) > 2^{\nu+1},$$

ce qui est une contradiction. Donc, $2^{\nu+1}-1$ est un nombre premier, comme on l'a affirmé. Donc, l'exercice 1.2 (a) implique que $\nu+1$ doit être aussi un nombre premier, soit p. Ceci termine la preuve. □

5.2 La convolution de Dirichlet

La preuve de la multiplicativité de σ_α dans la section précédente se généralise et nous amène naturellement au concept de la *convolution de Dirichlet* :

Définition 5.7 (convolution de Dirichlet)**.** Soient f et g deux fonctions arithmétiques. Leur *convolution de Dirichlet* est la fonction arithmétique $f*g$ définie par

$$(f*g)(n) := \sum_{d|n} f(d)g(n/d)$$

pour tout $n\in\mathbb{N}$, où la somme est prise sur tous les diviseurs $d\in\mathbb{N}$ de n.

En utilisant cette notion, nous avons une généralisation vaste du fait que σ_α est une fonction multiplicative.

Théorème 5.8. *Si f et g sont deux fonctions multiplicatives, alors $f*g$ est aussi multiplicative.*

Démonstration. Soient $m,n\in\mathbb{N}$ avec $(m,n)=1$. En utilisant le fait (cf. (5.2)) qu'il y a une bijection entre les ensembles $\{(d_1,d_2)\in\mathbb{N}^2 : d_1|m,\ d_2|n\}$ et $\{d|mn\}$, on trouve que

$$(f*g)(mn) = \sum_{d|mn} f(d)g(mn/d) = \sum_{d_1|m,\ d_2|n}\sum f(d_1d_2)g\Big(\frac{mn}{d_1d_2}\Big).$$

Puisque $(m,n)=1$, on a aussi que $(d_1,d_2)=(m/d_1,n/d_2)=1$ pour chaque pair (d_1,d_2) où $d_1|m$ et $d_2|n$. En utilisant la multiplicativité de f et de g, on en déduit que

$$\begin{aligned}(f*g)(mn) &= \sum_{d_1|m,\ d_2|n}\sum f(d_1)f(d_2)g\Big(\frac{m}{d_1}\Big)g\Big(\frac{m}{d_1}\Big)\\ &= \Big(\sum_{d_1|m} f(d_1)g\Big(\frac{m}{d_1}\Big)\Big)\Big(\sum_{d_2|n} f(d_2)g\Big(\frac{n}{d_2}\Big)\Big)\\ &= (f*g)(m)\cdot(f*g)(n).\end{aligned}$$

Ceci termine la démonstration. □

La convolution de Dirichlet a plusieurs autres propriétés importantes. On les résume au théorème suivant. Dans la partie (c), l'*addition de fonctions arithmétiques* se réfère à cette opération

qui, étant données deux fonctions arithmétiques f et g, elle produit la nouvelle fonction arithmétique $f+g$ dont ses valeurs sont données par $(f+g)(n) := f(n)+g(n)$ (où le symbole $+$ au côté droit se réfère à l'opération d'addition de nombres complexes).

Théorème 5.9 (Propriétés algébriques de la convolution de Dirichlet)**.**

(a) La convolution de Dirichlet est une opération commutative, c'est-à-dire $f*g = g*f$ pour toutes les fonctions arithmétiques f et g.

(b) La convolution de Dirichlet est une opération associative, c'est-à-dire $(f*g)*h = f*(g*h)$ pour toutes les fonctions arithmétiques f, g et h.

(c) La convolution de Dirichlet est une opération distributive [1] par rapport à l'addition de fonctions arithmétiques, c'est-à-dire $f*(g+h) = f*g + f*h$ pour toutes les fonctions arithmétiques f, g et h.

(d) La convolution de Dirichlet possède un élément neutre δ, défini par la relation $\delta(n) = 1_{n=1}$ pour tout $n \in \mathbb{N}$. C'est-à-dire, on a que [2] $f*\delta = f$ pour toute fonction arithmétique f.

Remarque. Le théorème 5.9 implique que l'ensemble des fonctions arithmétiques, muni des opérations de l'addition et de la convolution de Dirichlet, est un anneau commutatif et unitaire.

Démonstration. (a) Soit $n \in \mathbb{N}$. Il est facile de vérifier que l'ensemble $\{d \in \mathbb{N} : d|n\}$ est en bijection avec l'ensemble $\{(a,b) \in \mathbb{N}\times\mathbb{N} : ab = n\}$; cette bijection est donnée par la fonction $d \to (d, n/d)$. Donc, on a que

$$(f*g)(n) = \sum_{ab=n} f(a)g(b). \tag{5.5}$$

Si on pose $a' = b$ et $b' = a$, on trouve alors que

$$(f*g)(n) = \sum_{a'b'=n} f(b')g(a') = (g*f)(n).$$

(b) En utilisant la formule (5.5) deux fois, on a que

$$[(f*g)*h](n) = \sum_{kc=n} (f*g)(k)h(c) = \sum_{kc=n} \Big(\sum_{ab=k} f(a)g(b)\Big)h(c) = \sum_{abc=n} f(a)g(b)h(c),$$

où on a éliminé la variable k car elle est fixée quand a, b et c sont connus par la relation $k = ab$.

Idem, on a que

$$[f*(g*h)](n) = \sum_{a\ell=n} f(a)*(g*h)(\ell) = \sum_{a\ell=n} f(a)\Big(\sum_{bc=\ell} g(b)h(c)\Big) = \sum_{abc=n} f(a)g(b)h(c).$$

Ceci montre que $(f*g)*h = f*(g*h)$, ce qui est ce qu'il fallait démontrer.

1. On n'a pas besoin de vérifier aussi que $(f+g)*h = f*h+g*h$, car l'opération $*$ est commutative.
2. On n'a pas besoin de vérifier aussi que $\delta*f = f$, car l'opération $*$ est commutative.

(c) Soit $n \in \mathbb{N}$. Nous avons que

$$\begin{aligned}
[f * (g+h)](n) &= \sum_{d|n} f(d) \cdot (g+h)(n/d) \\
&= \sum_{d|n} f(d) \cdot \big(g(n/d) + h(n/d)\big) \\
&= \sum_{d|n} \big(f(d)g(n/d) + f(d)h(n/d)\big) \\
&= \sum_{d|n} f(d)g(n/d) + \sum_{d|n} f(d)h(n/d) \\
&= (f*g)(n) + (f*h)(n) = (f*g + f*h)(n).
\end{aligned}$$

(d) Soit f une fonction arithmétique. Pour tout $n \in \mathbb{N}$, on a que

$$(f*\delta)(n) = f(n)\delta(1) + \sum_{d|n,\, d<1} f(d)\delta(n/d).$$

Par la définition de δ, on a que $\delta(1) = 1$ et que $\delta(n/d) = 0$ quand $d|n$ et $d < n$. Donc, $(f*\delta)(n) = f(n)$, ce qui conclut la démonstration. □

Une question importante qui se pose est quelles fonctions arithmétiques possèdent d'un inverse par rapport à la convolution de Dirichlet.

Définition 5.10. Soit f une fonction arithmétique. S'il existe une autre fonction arithmétique g telle que $f * g = \delta$, alors on dit que f est *inversible (par rapport à la convolution de Dirichlet)* et on appelle g son *inverse de Dirichlet*.

Remarque. Si g_1, g_2 sont deux fonctions arithmétiques telles que $f * g_1 = f * g_2 = \delta$, alors on a que

$$g_2 = g_2 * \delta = g_2 * (f * g_1) = (g_2 * f) * g_1 = (f * g_2) * g_1 = \delta * g_1 = g_1.$$

Donc l'inverse de Dirichlet de f, si elle existe, est uniquement définie.

Théorème 5.11. *Soit $f : \mathbb{N} \to \mathbb{C}$ une fonction arithmétique. Alors, f est inversible si et seulement si $f(1) \neq 0$.*

Démonstration. Supposons que f est inversible. Donc, il existe $g : \mathbb{N} \to \mathbb{C}$ telle que $f * g = \delta$. En particulier, $(f*g)(1) = 1$. Puisque $(f*g)(1) = f(1)g(1)$, on conclut que $f(1) \neq 0$.

Réciproquement, supposons que $f(1) \neq 0$. On va construire l'inverse de Dirichlet de façon inductive. Tout d'abord, on pose $g(1) = 1/f(1)$, pour que $(f*g)(1) = 1$. Puis, fixons $n \geqslant 2$ et supposons qu'on définit les valeurs $g(1), \ldots, g(n-1)$ pour que $(f*g)(m) = \delta(m)$ pour $m = 1, 2, \ldots, n-1$. On va définir aussi $g(n)$ pour que $(f*g)(n) = \delta(n) = 0$. Notons que

$$(f*g)(n) = f(1)g(n) + \sum_{d|n,\, d>1} f(d)g(n/d).$$

Tous les sommés avec $d > 1$ sont déjà connus par l'hypothèse d'induction et par le fait que f est connue. Donc, on pose

$$g(n) = -\frac{1}{f(1)} \sum_{d|n,\, d>1} f(d)g(n/d)$$

pour que $(f * g)(n) = 0 = \delta(n)$. Ceci conclut l'étape inductive et donc la construction de g, qui est l'inverse de Dirichlet de f. □

Théorème 5.12. *Chaque fonction multiplicative est inversible et son inverse de Dirichlet est aussi une fonction multiplicative.*

Démonstration. Soit f une fonction multiplicative. En particulier, $f(1) = 1 \neq 0$ d'après le théorème 5.3. Puis, construisons une fonction arithmétique g telle que : (i) g est multiplicative ; (ii) $f * g = \delta$. Par l'unicité de l'inverse de Dirichlet, ceci montre le théorème.

Afin de construire g, posons d'abord $g(1) = 1$ pour que $(f * g)(1) = 1 = \delta(1)$. Puis, on construit les valeurs $g(p^\nu)$ par induction sur ν en modifiant légèrement l'argument de la preuve du théorème 5.11. Quand $\nu = 1$, on pose $g(p) = -f(p)$ pour que $(f * g)(p) = f(p) + g(p) = 0$. Puis, fixons $\nu \geqslant 2$ et supposons qu'on a défini $g(p^k)$ pour $k = 1, \ldots, \nu - 1$. Posons alors $g(p^\nu) = -\sum_{k=1}^{\nu} f(p^k)g(p^{\nu-k})$ pour que $(f * g)(p^\nu) = 0$. Ceci conclut l'étape inductive de la construction de $g(p^\nu)$.

Finalement, on prolonge g à tous les nombres naturels par la relation

$$g(n) := \prod_{p^\nu \| n} g(p^\nu).$$

Ceci définit une fonction multiplicative (cf. exercice 5.1). De plus, g satisfait les relations $(f * g)(1) = 1$ et $(f * g)(p^\nu) = 0$ pour chaque premier p et chaque entier $\nu \geqslant 1$. Il faut montrer que $(f * g)(n) = 0$ pour chaque $n \geqslant 2$. Puisque f et g sont multiplicatives, leur convolution $f * g$ l'est aussi. Donc, on a que

$$(f * g)(n) = \prod_{p^\nu \| n} (f * g)(p^\nu). \tag{5.6}$$

Si $n \geqslant 2$, alors il existe au moins une puissance première p^ν qui le divise exactement. Puisque $(f * g)(p^\nu) = 0$, tout le produit au côté droit de (5.6) s'annule. Ceci conclut la démonstration. □

Théorème 5.13. *L'inverse de Dirichlet de la fonction constante* 1 *est la fonction* μ *de Möbius définie par*

$$\mu(n) := \begin{cases} 0 & \text{s'il existe un premier } p \text{ tel que } p^2 | n, \\ (-1)^{\omega(n)} & \text{sinon.} \end{cases}$$

Remarque. Une façon équivalente d'énoncer le théorème 5.13 est de dire que la fonction de Möbius satisfait la relation

$$\sum_{d|n} \mu(d) = 1_{n=1}. \tag{5.7}$$

On appelle cette relation la *formule d'inversion de Möbius*.

Démonstration. Puisque 1 et μ sont de fonctions multiplicatives, il suffit de vérifier que $(1 * \mu)(p^\nu) = 0$ pour chaque nombre premier p et chaque entier $\nu \geqslant 1$, ce qui est une conséquence directe de la définition de μ. □

Théorème 5.14. *La fonction ϕ d'Euler, définie par $\phi(n) = \#\{1 \leqslant a \leqslant n : (a,n) = 1\}$, est multiplicative et ses valeurs sont données par*

$$\phi(n) = \prod_{p^\nu | n} (p^\nu - p^{\nu-1}). \tag{5.8}$$

Démonstration. On divise les entiers $1, 2, \ldots, n$ selon leur pgcd avec n. On a alors que

$$n = \sum_{d|n} \#\{1 \leqslant m \leqslant n : (m,n) = d\}.$$

Quand $(m,n) = d$, on sait d'après le lemme 2.8 que $m = da$ avec $(a, n/d) = 1$. De plus, les relations $1 \leqslant m \leqslant n$ impliquent que $1 \leqslant a \leqslant n/d$. Donc, on trouve que

$$\#\{1 \leqslant m \leqslant n : (m,n) = d\} = \#\{1 \leqslant a \leqslant n/d : (a, n/d) = 1\} = \phi(n/d).$$

En conclusion, on a montré que

$$n = \sum_{d|n} \phi(n/d) = (1 * \phi)(n).$$

ce qui veut dire que $\mathrm{id} = 1 * \phi$, où la fonction id dénote la fonction-identité donnée par $\mathrm{id}(n) := n$. En multipliant cette relation à gauche par μ, et en utilisant que $\mu * 1 = \delta$, on conclut que $\phi = \mu * \mathrm{id}$. Les fonctions μ et id sont multiplicatives, donc ϕ l'est aussi. Il reste à calculer ϕ à montrer (5.8). Par multiplicativité et par le théorème 5.3, il suffit de considérer le cas où $n = p^\nu$. Dans ce cas-ci, on a que

$$\phi(p^\nu) = (\mu * \mathrm{id})(p^\nu) = \mu(1)p^\nu + \mu(p)p^{\nu-1} + \sum_{k=2}^{\nu} \mu(p^k)p^{\nu-k} = p^\nu - p^{\nu-1}$$

car $\mu(1) = 1$, $\mu(p) = -1$ et $\mu(p^k) = 0$ pour tout $k \geqslant 2$. Ceci termine la preuve. □

5.3 Séries de Dirichlet

Une autre façon d'étudier une fonction arithmétique f est d'empaqueter toutes ses valeurs dans une série-génératrice. L'exemple le plus classique d'une telle série est la série de puissances, donnée par $\sum_{n \geqslant 1} f(n)x^n$. Par contre, cette série est mieux adopter pour captiver les propriétés *additives* de f, car les fonctions x^n satisfont les relations $x^{m+n} = x^m \cdot x^n$. Afin d'étudier les propriétés *multiplicatives* de f, on introduit sa *série de Dirichlet*

$$F(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s}.$$

On traitera cette série de façon formelle, sans s'inquiétant de questions de convergence. Pour une étude de telles questions analytique, on dirige les lecteurs vers un cours en théorie des nombres analytique (voir aussi [3, 5]).

Notons que si $F(s) = \sum_{n=1}^{\infty} f(n)/n^s$ et $G(s) = \sum_{n=1}^{\infty} f(n)/n^s$, on peut les manipuler formellement et les ajouter

$$F(s) + G(s) = \sum_{n=1}^{\infty} \frac{f(n) + g(n)}{n^s}.$$

On peut aussi multiplier F et G comme suit :

$$\begin{aligned} F(s)G(s) &= \left(\sum_{a=1}^{\infty} \frac{f(a)}{a^s} \right) \left(\sum_{b=1}^{\infty} \frac{g(b)}{b^s} \right) \\ &= \sum_{a,b \geqslant 1} \sum \frac{f(a)g(b)}{(ab)^s} \\ &= \sum_{n=1}^{\infty} \sum_{\substack{a,b \geqslant 1 \\ ab=n}} \sum \frac{f(a)g(b)}{n^s} \\ &= \sum_{n=1}^{\infty} \frac{(f * g)(n)}{n^s}. \end{aligned}$$

On voit donc que le résultat de l'addition formelle de $F + G$ est la série de Dirichlet de la fonction arithmétique $f + g$, et le résultat de la multiplication formelle de $F \cdot G$ est la série de Dirichlet de la fonction arithmétique $f * g$.

En termes algébriques, ceci veut dire que l'ensemble des séries de Dirichlet muni des opérations formelles d'addition et de multiplication est un anneau *isomorphe* à ce des fonctions arithmétiques. On peut donc traiter les fonctions arithmétiques et les séries de Dirichlet comme les deux côtés de la même pièce.

La série de Dirichlet la plus fameuse est la fonction zêta de Riemann, définie par

$$\sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Puisque $1 * \mu = \delta$, on voit tout de suite que

$$\zeta(s)^{-1} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s},$$

où le symbole $\zeta(s)^{-1}$ ici veut l'inverse par rapport à la multiplication formelle de séries de Dirichlet. (Cette relation est aussi vraie dans $\mathbb{C}$ quand $\mathrm{Re}(s) > 1$.)

La fonction ζ et toutes les séries de Dirichlet de fonctions multiplicatives se factorisent. Ceci est une réflexion analytique du théorème 5.3. En effet, soit f une fonction multiplicative et soit F sa série de Dirichlet. De plus, soit $p_1 < p_2 < p_3 < \cdots$ la suite de tous les nombres premiers écrits

en ordre croissante. En réarrangeant les sommés de F, on a que

$$\begin{aligned} F(s) &= \lim_{k\to\infty}\lim_{N\to\infty} \sum_{\substack{n=p_1^{\nu_1}\cdots p_k^{\nu_k}\\ 0\leqslant\nu_1,\dots,\nu_k\leqslant N}} \frac{f(n)}{n^s} \\ &= \lim_{k\to\infty}\lim_{N\to\infty} \sum_{0\leqslant\nu_1,\dots,\nu_k\leqslant N} \frac{f(p_1^{\nu_1})f(p_2^{\nu_2})\cdots f(p_k^{\nu_k})}{p_1^{\nu_1 s}p_2^{\nu_2 s}\cdots p_k^{\nu_k s}} \\ &= \lim_{k\to\infty}\lim_{N\to\infty} \left(\sum_{\nu_1=0}^{N}\frac{f(p_1^{\nu_1})}{p_1^{\nu_1 s}}\right)\left(\sum_{\nu_2=0}^{N}\frac{f(p_2^{\nu_2})}{p_2^{\nu_2 s}}\right)\cdots\left(\sum_{\nu_k=0}^{N}\frac{f(p_k^{\nu_k})}{p_k^{\nu_k s}}\right) \\ &= \lim_{k\to\infty} \left(\sum_{\nu_1=0}^{\infty}\frac{f(p_1^{\nu_1})}{p_1^{\nu_1 s}}\right)\left(\sum_{\nu_2=0}^{\infty}\frac{f(p_2^{\nu_2})}{p_2^{\nu_2 s}}\right)\cdots\left(\sum_{\nu_k=0}^{\infty}\frac{f(p_k^{\nu_k})}{p_k^{\nu_k s}}\right). \end{aligned}$$

En conclusion, on arrive au *produit d'Euler* de F, donné par

$$F(s) = \prod_p \left(1 + \frac{f(p)}{p^s} + \frac{f(p^2)}{p^{2s}} + \cdots\right). \tag{5.9}$$

Dans le cas de la fonction ζ, cette expression se simplifie à

$$\zeta(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}, \tag{5.10}$$

en utilisant le fait que $1 + x + x^2 + \cdots = 1/(1-x)$ avec $x = p^{-s}$.

Bien que toutes les manipulations du paragraphe précédent soient formelles, les relations (5.9) et (5.10) peuvent être justifiées rigoureusement quand les séries de Dirichlet converge absolument. Pour ζ, ceci est vrai donc quand $\mathrm{Re}(s) > 1$.

Observons que la relation (5.10) peuvent être inversée très facilement :

$$\zeta(s)^{-1} = \prod_p \left(1 - \frac{1}{p^s}\right).$$

En comparant le côté gauche avec (5.9), on voit qu'elle est la série de Dirichlet de la fonction multiplicative f unique qui satisfait les relations $f(p) = -1$ et $f(p^\nu) = 0$ pour tout $\nu \geqslant 2$. On voit donc que $f = \mu$, c'est-à-dire que $\zeta(s)^{-1}$ est la série de Dirichlet de μ. Puisque ζ et ζ^{-1} sont inverses l'une de l'autre, on en déduit que les fonctions arithmétiques correspondant les sont aussi (par rapport à la convolution de Dirichlet cette fois-ci). Ceci montre d'une autre façon que $1 * \mu = \delta$.

En général, il est souvent plus de calculer l'inverse de Dirichlet d'une fonction multiplicative donnée en utilisant le produit d'Euler (5.9).

Exemple. On veut calculer l'inverse de la fonction μ^2, qui est la fonction-indicatrice de l'ensemble des entiers sans carré facteur. D'après (5.9) avec $f = \mu^2$, on a que

$$\sum_{n=1}^{\infty}\frac{\mu^2(n)}{n^s} = \prod_p \left(1 + \frac{\mu^2(p)}{p^s} + \frac{\mu^2(p^2)}{p^{2s}} + \cdots\right) = \prod_p \left(1 + \frac{1}{p^s}\right).$$

Puisque $1 + x = (1 - x^2)/(1 - x)$, on en déduit que

$$\sum_{n=1}^{\infty} \frac{\mu^2(n)}{n^s} = \prod_p \frac{1 - 1/p^{2s}}{1 - 1/p^s} = \frac{\zeta(s)}{\zeta(2s)}. \tag{5.11}$$

La fonction $1/\zeta(2s) = \prod_p (1 - 1/p^{2s})$ est la série de Dirichlet de la fonction

$$f(n) = \begin{cases} \mu(m) & \text{si } n = m^2, \\ 0 & \text{sinon.} \end{cases} \tag{5.12}$$

Donc, la relation formelle (5.11) suggère que

$$\mu^2(n) = \sum_{d^2 | n} \mu(d). \tag{5.13}$$

La lectrice attentive est peut être sceptique de cette déduction. Cependant, elle a maintenant une formule concrète à vérifier ; cette tache est facile à accomplir rigoureusement en utilisant les outils des sections précédentes. En effet, on a que μ^2 est multiplicative et que le côté gauche de (5.13) est égale à $1 * f$, où f est donnée par (5.12). Puisque 1 et f sont les deux multiplicatives, leur convolution $1 * f$ l'est également. En conclusion, afin de vérifier (5.13), il suffit de la tester quand n est une puissance première p^ν. Dans ce cas, on trouve que

$$\sum_{d^2 | p^\nu} \mu(d) = \begin{cases} 1 & \text{si } \nu = 1, \\ 1 - 1 = 0 & \text{si } \nu = 2. \end{cases}$$

Le côté droit est égale à $\mu^2(p^\nu)$, ce qui montre la relation (5.13) rigoureusement.

La morale de cette histoire est qu'on peut utiliser les manipulations formelles de cette section pour deviner certaines identités de convolution, et puis d'utiliser les outils des sections 5.1 et 5.2 pour les montrer rigoureusement. ■

Exemple. On sait que $n = \prod_{p^\nu \| n} p^\nu$, et donc que

$$\log n = \sum_{p^\nu \| n} \nu \log p = \sum_{p^k | n} \log p.$$

Pour voir la deuxième égalité, on note que si $p^\nu \| n$, alors $p^k | n$ pour exactement ν exposants $k \in \mathbb{N}$. On peut écrire cette relation de façon plus compacte comme

$$\log = \Lambda * 1, \tag{5.14}$$

où Λ est la fonction de von Mangoldt qu'on a définit au début du chapitre 5 ; on se rappelle que

$$\Lambda(n) = \begin{cases} \log p & \text{si } n = p^\nu, \\ 0 & \text{sinon.} \end{cases}$$

En multipliant (5.14) à gauche par μ, on trouve que

$$\Lambda = \log * \mu = \mu * \log . \tag{5.15}$$

Notons que $\zeta'(s) = -\sum_{n\geqslant 1} \frac{\log n}{n^s}$ en différenciant de façon formelle la série $\sum_{n\geqslant 1} 1/n^s$ terme par terme. Donc, la série de Dirichlet de la fonction de von Mangoldt est la fonction $-\zeta'/\zeta$, c'est-à-dire

$$\sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} = -\frac{\zeta'}{\zeta}(s).$$

Encore une fois, cette dernière relation peut se justifie rigoureusement dans le domaine de convergence absolue de la fonction ζ, c'est-à-dire quand $\mathrm{Re}(s) > 1$. ■

5.4 Exercices

EXERCICE 5.1. Soit f une fonction arithmétique avec $f(1) = 1$.

(a) Montrer que si $f(n) = \prod_{p^\nu \| n} f(p^\nu)$ pour chaque $n \geqslant 2$, alors f est multiplicative. De plus, montrer que si $f(n) = \prod_{p^\nu \| n} f(p)^v$, alors f est fortement multiplicative.

(b) Montrer que si $f(n) = \prod_{p|n} f(p)$ pour chaque $n \geqslant 2$, alors

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

Déduire que f est multiplicative.

EXERCICE 5.2. Est-ce que la fonction $f(n) = (-1)^{n-1}$ multiplicative ou pas ?

EXERCICE 5.3. Prouver le corollaire 5.4.

EXERCICE 5.4. Trouver les trois premiers nombres parfaits.

EXERCICE 5.5. Montrez que n est un nombre parfait si et seulement si $\sum_{d|n} 1/d = 2$.

EXERCICE 5.6. 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.

EXERCICE 5.7. (a) Soient $F, G : \mathbb{R}_{\geqslant 1} \to \mathbb{C}$ deux fonctions. Montrer que les deux relations suivantes sont équivalentes :

— $F(x) = \sum_{n\leqslant x} G(x/n)$ pour tout $x \geqslant 1$;
— $G(x) = \sum_{n\leqslant x} \mu(n) F(x/n)$ pour tout $x \geqslant 1$.

(b) Montrer que $1 = \sum_{n\leqslant x} \mu(n) \lfloor x/n \rfloor$ et en déduire que

$$\left| \sum_{n\leqslant x} \frac{\mu(n)}{n} \right| \leqslant 1.$$

EXERCICE 5.8. Soit $\phi(n) = \#\{1 \leqslant a \leqslant n : (a, n) = 1\}$ la fonction d'Euler.

(a) Montrer que si $n > 2$, alors $\phi(n)$ est un nombre pair.

(b) Montrer que

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

en utilisant le crible d'Eratosthène.

(c) Montrer que pour tous $m, n \in \mathbb{N}$, on a

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

(d) Soit $f(n) = \phi(n)/n$ et soit $(n_k)_{k=1}^{\infty}$ la suite de valeurs n sur lesquelles la fonction f atteint un «record bas», c'est-à-dire $n_1 = 1$ et pour chaque $k \geqslant 2$, on définit n_k d'être le plus petit entier $> n_{k-1}$ tel que $f(n_k) < f(n)$ pour tout $n < n_k$. (Par exemple, $n_2 = 2$ et $n_3 = 6$.) Trouver une formule général pour n_k et pour $f(n_k)$.

EXERCICE 5.9.

(a) Montrer que

$$1_{n \text{ sans facteur carré}} = \mu^2(n) = \sum_{d^2|n} \mu(d).$$

(b) On dit qu'un entier n est *plein de carrés* si pour chaque premier p qui divise n, on a que p^2 divise aussi n. Montrer que

$$1_{n \text{ plein de carrés}} = \sum_{a^2b^3=n} \mu^2(b).$$

(c) Montrer que $2^{\omega} = 1 * \mu^2$.

EXERCICE 5.10. Trouver l'inverse de Dirichlet des fonctions arithmétiques suivantes : (a) μ ; (b) μ^2 ; (c) τ ; (d) 2^{ω} ; (e) ϕ ; (e) σ ; (f) la fonction indicatrice des entiers pleins de carrés.

Chapitre 6

Estimations asymptotiques pour les nombres premiers

Depuis qu'Euclide a prouvé qu'il existe une infinité de nombres premiers, la distribution de ces objets fondamentaux a fasciné les mathématiciens. Contrairement à d'autres suites qui présentent une structure très régulière, comme la suite des carrés, les nombres premiers ne semblent pas suivre de motif. Par conséquent, deviner la location exacte du n-ième premier apparaît être un défi impossible à révéler lorsque $n \to \infty$.

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 comptage des nombres premiers

$$\pi(x) = \#\{p \leqslant x\}.$$

Si p_n dénote le n-ième nombre premier, alors $\pi(p_n) = n$, pour que chaque approximation à $\pi(x)$ se traduise immédiatement à une approximation à p_n, et vice versa.

L'étude de la distributions des nombres premiers a préoccupé le jeune Gauss. Après avoir examiné des tables des 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 intégral, ceci veut dire qu'une bonne approximation pour $\pi(x)$ est donnée par le *logarithme intégral*

$$\mathrm{li}(x) := \int_2^x \frac{\mathrm{d}t}{\log t}.$$

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 plus en détail de cette notation dans la section suivante. De manière équivalente, l'estimation de Gauss

(6.1) constate que $p_n \sim n \log n$ quand $n \to \infty$, une affirmation qu'on laisse comme exercice pour les lecteurs.

Il a fallu plus d'un siècle pour prouver la conjecture de Gauss pour $\pi(x)$. La route vers la preuve a été pavée par le mémoire révolutionnaire de Riemann dans son mémoire *Über die Anzahl der Primzahlen unter einer gegebenen Grösse*, publié en 1859. Dans son travail, Riemann a expliqué comment $\pi(x)$ est intimement lié aux propriétés analytiques de sa fonction ζ. Il a procédé de proposer un programme dont la complétion amènerait à une compréhension profonde des nombres premiers. En particulier, il a postulé la fameuse *Hypothèse de Riemann*, dont la vérité implique l'inégalité [9]

$$|\pi(x) - \mathrm{li}(x)| \leqslant \frac{1}{8\pi}\sqrt{x}\log x \quad \text{pour tout} \quad x \geqslant 2657, \tag{6.2}$$

une forme très forte de l'estimation de Gauss. [1] En 1895, von Mangoldt a prouvé rigoureusement toutes les étapes du plan de Riemann, sauf une : l'hypothèse de Riemann. Néanmoins, en 1896, Hadamard et de la Vallée Poussin ont montré une forme faible de cette hypothèse, qui leur ai permit d'établir l'estimation de Gauss, connue aujourd'hui comme le *théorème des nombres premiers* :

Théorème des nombres premiers. *On a que* $\pi(x) \sim x/\log x$ *lorsque* $x \to \infty$.

La preuve de ce théorème est un de résultats principaux d'un cours d'introduction à la théorie des nombres analytique (p. ex., le cours MAT3634). Dans ces notes, on obtient quelques estimations plus modestes pour $\pi(x)$, ainsi que pour la somme $\sum_{p\leqslant x} 1/p$.

Afin d'énoncer nos résultats, il est convenable d'utiliser la notation asymptotique des «grands-O» qu'on développe à la section suivante.

6.1 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 $x \geqslant 1$. 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 $\lfloor x \rfloor \leqslant x < \lfloor x \rfloor + 1$, 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*

$$\lfloor x \rfloor = x + O(1). \tag{6.3}$$

Généralement, étant données les fonctions complexes f, g and h, et un sous-ensemble I de leurs domaines de définition, on écrit

$$f(x) = g(x) + O(h(x)) \quad (x \in I), \tag{6.4}$$

1. En fait, (6.2) est équivalente à l'hypothèse de Riemann.

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

$$|f(x) - g(x)| \leqslant c \cdot h(x) \quad \text{pour chaque } x \in I.$$

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 (6.3) 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 $\sum_{n \leqslant x} \lfloor x/n \rfloor$. 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 $\sum_{n \leqslant x} \lfloor x/n \rfloor = x \sum_{n \leqslant x} 1/n + O(x)$, car

$$\left| \sum_{n \leqslant x} \lfloor x/n \rfloor - x \sum_{n \leqslant x} 1/n \right| = \left| \sum_{n \leqslant x} (\lfloor x/n \rfloor - x/n) \right| \leqslant \sum_{n \leqslant x} 1 \leqslant x.$$

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 grandeur* de fonctions différentes : si

$$f(x) = O(g(x)) \quad (x \in I),$$

alors on dit que «f a plus petite ordre de grandeur 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 grandeur 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 (Une preuve analytique de l'infinité des nombres premiers)**.** Pour donner une idée du pouvoir des estimations asymptotiques, on les utilise pour donner une démonstration alternative 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,

$$\{n \leqslant x\} = \{2^k \leqslant x : k \geqslant 0\}.$$

Le côté gauche a cardinalité $\lfloor x \rfloor = x + O(1) \sim x$ lorsque $x \to \infty$. Par contre, le côté droit a cardinalité

$$\begin{aligned}\#\{2^k \leqslant x : k \geqslant 0\} &= \#\{0 \leqslant k \leqslant \log x/\log 2\} \\ &= 1 + \lfloor \log x/\log 2 \rfloor \\ &= \frac{\log x}{\log 2} + O(1) \\ &\sim \frac{\log x}{\log 2}\end{aligned}$$

lorsque $x \to \infty$. Mais ceci est absurde 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

$$\{n \leqslant x\} = \{2^k 3^\ell \leqslant x : k, \ell \geqslant 0\}.$$

Mais on a que

$$\begin{aligned}\#\{2^k 3^\ell \leqslant x : k, \ell \geqslant 0\} &\leqslant \#\{(k,\ell) \in \mathbb{Z}^2 : 0 \leqslant k \leqslant \log x/\log 2,\ 0 \leqslant \ell \leqslant \log x/\log 3\} \\ &\leqslant \left(1 + \frac{\log x}{\log 2}\right)\left(1 + \frac{\log x}{\log 3}\right) \\ &\sim \frac{(\log x)^2}{(\log 2)(\log 3)} = o_{x\to\infty}(x),\end{aligned}$$

ce qui est absurd.

Plus généralement, si on suppose que les seuls nombres premiers sont $p_1, \ldots, p_k$, alors

$$\{n \leqslant x\} = \{p_1^{\nu_1} \cdots p_k^{\nu_k} : \nu_1, \ldots, \nu_m \geqslant 0\}.$$

Cependant,

$$\begin{aligned}\#\{p_1^{\nu_1} p_2^{\nu_2} \cdots p_k^{\nu_k} : \nu_1, \ldots, \nu_k \geqslant 0\} &\leqslant \#\{(\nu_1, \ldots, \nu_k) \in \mathbb{Z}^m : 0 \leqslant \nu_j \leqslant \log x/\log p_j\ \forall j\} \\ &\leqslant \prod_{j=1}^{k} \left(1 + \frac{\log x}{\log p_j}\right) \\ &\sim \frac{(\log x)^k}{\prod_{j=1}^k \log p_j} = o_{x\to\infty}(x)\end{aligned}$$

d'après la règle de l'Hôpital. On est arrivé encore à une contradiction. Ceci montre l'infinité des premiers. ■

Remarque. En fait, l'argument ci-dessus peut rendre une estimation quantitative de $\pi(x)$ si on l'utilise avec dextérité.

6.2 La somme des réciproques des nombres premiers

La théorie de la fonction zêta de Riemann nous permet de donner une estimation à la somme

$$\sum_{p\leqslant x}\frac{1}{p}.$$

Afin d'expliquer l'idée de cette estimation, on commence avec un argument d'Euler qui montre que la série infinie $\sum_p 1/p=\infty$; en particulier, il existe un nombre infini de nombres premiers.

D'après le produit d'Euler de ζ (cf. 5.10), nous avons que

$$\log\zeta(\sigma)=\sum_p \log\frac{1}{1-1/p^\sigma}$$

pour $\sigma>1$. On a aussi le développement de Taylor $-\log(1-x)=\sum_{k=1}^\infty x^k/k$ pour $x\in(-1,1)$. Donc,

$$\log\zeta(\sigma)=\sum_p\sum_{k=1}^\infty\frac{1}{kp^{k\sigma}}=\sum_p\frac{1}{p^\sigma}+\sum_p\sum_{k=2}^\infty\frac{1}{kp^{k\sigma}}. \tag{6.5}$$

Si $\sum_p 1/p<\infty$, il faudrait que le côté droit est uniformément borné pour $\sigma>1$, soit par $M>0$. Dans ce cas-ci, on devrait avoir aussi que

$$\sum_{n=1}^N\frac{1}{n^\sigma}\leqslant e^M$$

pour tout $N\geqslant 1$ et pour tout $\sigma>1$. Si on fixe $N\geqslant 1$ et on laisse $\sigma\to 1^+$, on déduit aussi que

$$\sum_{n=1}^N\frac{1}{n}\leqslant e^M$$

pour tout $N\geqslant 1$. Mais ceci contredit le fait que la série harmonique $\sum_{n=1}^\infty 1/n$ diverge.

On a montré donc que la divergence de la série harmonique $\sum_{n=1}^\infty 1/n$ implique la divergence de la série $\sum_p 1/p$. En quantifiant cet argument, on trouvera une estimation précise de la vitesse de divergence de cette dernière somme.

Théorème 6.1. *Pour $x\geqslant 2$, on a l'estimation uniforme*

$$\sum_{p\leqslant x}\frac{1}{p}=\log\log x+O(1).$$

On commence avec une estimation précise de ζ et de sa dérivée proche de 1.

Proposition 6.2. *Pour $\sigma > 1$, on a que*

$$\frac{1}{\sigma - 1} \leqslant \zeta(\sigma) \leqslant \frac{\sigma}{\sigma - 1}$$

et

$$-1 + \frac{1}{(\sigma-1)^2} \leqslant -\zeta'(\sigma) \leqslant 1 + \frac{1}{(\sigma-1)^2}.$$

Démonstration. La fonction $x \to x^{-\sigma}$ est décroissante pour $x > 1$. Donc, on a que

$$\int_n^{n+1} \frac{\mathrm{d}x}{x^\sigma} \frac{1}{n^\sigma} \leqslant \int_{n-1}^{n} \frac{\mathrm{d}x}{x^\sigma}$$

pour tout $n \in \mathbb{N}$. Par la suite,

$$\zeta(\sigma) = 1 + \sum_{n=2}^{\infty} \frac{1}{n^\sigma} \leqslant 1 + \int_1^\infty \frac{\mathrm{d}x}{x^\sigma} = 1 + \frac{1}{\sigma - 1}$$

et

$$\zeta(\sigma) = \sum_{n=1}^{\infty} \frac{1}{n^\sigma} \geqslant \int_1^\infty \frac{\mathrm{d}x}{x^\sigma} = \frac{1}{\sigma - 1}.$$

Ainsi, on a que $-\zeta'(\sigma) = \sum_{n=1}^\infty (\log n)/n^\sigma$.[2] La fonction $x \to (\log x)/x^{1+\sigma}$ est décroissante pour $x \geqslant e$. Donc, on a que

$$\frac{\log n}{n^\sigma} \leqslant \int_{n-1}^{n} \frac{\mathrm{d}x}{x^\sigma} \quad \text{pour } n \geqslant 4 \qquad \text{et} \qquad \frac{\log n}{n^\sigma} \geqslant \int_n^{n+1} \frac{\mathrm{d}x}{x^\sigma} \quad \text{pour } n \geqslant 3.$$

On en déduit que

$$-\zeta'(\sigma) = \frac{\log 2}{2^\sigma} + \frac{\log 3}{3^\sigma} + \sum_{n=4}^{\infty} \frac{1}{n^\sigma} \leqslant \frac{\log 2}{2} + \frac{\log 3}{3} + \int_3^\infty \frac{\log x}{x^\sigma} \mathrm{d}x \leqslant 1 + \frac{1}{(\sigma-1)^2},$$

car $\int_1^\infty \frac{\log x}{x^\sigma} \mathrm{d}x = 1/(\sigma-1)^2$ par intégration par parties, et que

$$-\zeta'(\sigma) = \frac{\log 2}{2^\sigma} + \sum_{n=3}^{\infty} \frac{1}{n^\sigma} \geqslant \int_3^\infty \frac{\log x}{x^\sigma} \geqslant -1 + \frac{1}{(\sigma-1)^2},$$

car $\int_1^3 \frac{\log x}{x^\sigma} \mathrm{d}x \leqslant \int_1^x \frac{\log x}{x} \mathrm{d}x = \frac{(\log 3)^2}{2} < 1$. □

En utilisant le produit d'Euler 5.10, nous pouvons facilement tourner nos inégalités pour ζ et pour ζ' à des estimations pour certaines sommes de nombres premiers.

2. Fixons $\varepsilon > 0$ pour que $\sigma \geqslant 1 + \varepsilon$. Il est facile de voir, par exemple par le critère de Weierstrass, que la séries $\sum_{n=1}^\infty 1/n^w$ et sa série de dérivées $\sum_{n=1}^\infty \frac{\mathrm{d}}{\mathrm{d}w}(1/n^w) = -\sum_{n=1}^\infty (\log n)/n^w$ convergent uniformément pour $w \geqslant 1 + \varepsilon$, donc $\zeta'(w) = \frac{\mathrm{d}}{\mathrm{d}w} \sum_{n=1}^\infty 1/n^w = -\sum_{n=1}^\infty (\log n)/n^w$ pour tout $w \geqslant 1 + \varepsilon$, en particulier pour $w = \sigma$.

Corollaire 6.3. *Pour $1 < \sigma \leqslant 2$, nous avons les estimations uniformes*

$$\sum_p \frac{1}{p^\sigma} = \log \frac{1}{\sigma - 1} + O(1) \quad \textit{et} \quad \sum_p \frac{\log p}{p^\sigma} = \frac{1}{\sigma - 1} + O(1).$$

Démonstration. Notre point de départ est la relation (6.5). On utilise la proposition 6.2 pour borner son côté droit : on a que

$$\log \frac{1}{\sigma - 1} \leqslant \log \zeta(\sigma) \leqslant \log \sigma + \log \frac{1}{\sigma - 1} \leqslant \log 2 + \log \frac{1}{\sigma - 1}$$

pour $\sigma \in (1, 2]$. D'autre côté, nous avons que

$$\sum_p \sum_{k=2}^\infty \frac{1}{kp^{k\sigma}} \leqslant \sum_p \sum_{k=2}^\infty \frac{1}{p^k} = \sum_p \frac{1}{p(p-1)} \leqslant \sum_{n=2}^\infty \frac{1}{n(n-1)} = 1,$$

en utilisant le fait que $1/(n(n-1)) = 1/(n-1) - 1/n$. En insérant les deux relations ci-dessous à (6.5), on déduit que

$$\sum_p \frac{1}{p^\sigma} \leqslant \log \zeta(\sigma) \leqslant \log \frac{1}{\sigma - 1} + \log 2,$$

ainsi que

$$\sum_p \frac{1}{p^\sigma} \geqslant \log \zeta(\sigma) - 1 \geqslant \log \frac{1}{\sigma - 1} - 1$$

pour tout $\sigma \in (1, 2]$. Ceci montre la première partie du théorème.

Puis, on considère la somme $\sum_p (\log p)/p^\sigma$. En différentiant les deux côtés de (6.5), on trouve que [3]

$$-\frac{\zeta'}{\zeta}(\sigma) = \sum_p \sum_{k=1}^\infty \frac{\log p}{p^{k\sigma}} = \sum_p \frac{\log p}{p^\sigma} + \sum_p \sum_{k=2}^\infty \frac{\log p}{p^{k\sigma}}.$$

La proposition 6.2 implique que

$$-\frac{\zeta'}{\zeta}(\sigma) \leqslant \frac{1/(\sigma-1)^2 + 1}{1/(\sigma-1)} = \frac{1}{\sigma - 1} + (\sigma - 1) \leqslant \frac{1}{\sigma - 1} + 1$$

pour $\sigma \in (1, 2]$, et donc on a que

$$\sum_p \frac{\log p}{p^\sigma} \leqslant \frac{1}{\sigma - 1} + 1.$$

De façon similaire, on trouve que

$$-\frac{\zeta'}{\zeta}(\sigma) \geqslant \frac{1/(\sigma-1)^2 - 1}{\sigma/(\sigma-1)} = \frac{1}{\sigma - 1} \cdot \frac{1 - (\sigma-1)^2}{1 + (\sigma-1)} = \frac{1}{\sigma - 1} - 1$$

3. Fixons $\varepsilon > 0$ pour que $\sigma \geqslant 1 + \varepsilon$. La série $\sum_p \sum_{k \geqslant 1} \frac{1}{kp^{kw}}$ et sa série de dérivées $\sum_p \sum_{k \geqslant 1} \frac{\mathrm{d}}{\mathrm{d}w}(\frac{1}{kp^{kw}}) = -\sum_p \sum_{k \geqslant 1} \frac{\log p}{p^{kw}}$ convergent les deux uniformément pour $w \geqslant 1 + \varepsilon$ d'après le critére de Weierstass, donc $\frac{\zeta'}{\zeta}(w) = \frac{\mathrm{d}}{\mathrm{d}w} \log \zeta(w) = \frac{\mathrm{d}}{\mathrm{d}w} \sum_p \sum_{k \geqslant 1} \frac{1}{kp^{kw}} = -\sum_p \sum_{k \geqslant 1} \frac{\log p}{p^{kw}}$ pour $w \geqslant 1 + \varepsilon$ et, en particulier, pour $w = \sigma$.

pour $\sigma \in (1,2]$. D'autre côté, on a que

$$\sum_p \sum_{k=2}^{\infty} \frac{\log p}{p^{k\sigma}} \leqslant \sum_p \sum_{k=2}^{\infty} \frac{\log p}{p^k} = \sum_p \frac{\log p}{p(p-1)} \approx 0.75524 < 1.$$

Donc,

$$\sum_p \frac{\log p}{p^\sigma} = -\frac{\zeta'}{\zeta}(\sigma) - \sum_p \sum_{k=2}^{\infty} \frac{\log p}{p^{k\sigma}} \geqslant \frac{1}{\sigma - 1} - 2.$$

Ceci conclut la démonstration. □

On est enfin prêt de montrer le théorème 6.1.

Démonstration du théorème 6.1. Pour $t \in [0,1]$, nous avons que $e^t \leqslant 1 + et$, et donc que $1 \leqslant e^{-t}(1+et)$. Pour chaque $p \leqslant x$, on utilise cette inégalité avec $t = \frac{\log p}{\log x}$. On en déduit que

$$\begin{aligned}
\sum_{p \leqslant x} \frac{1}{p} &\leqslant \sum_{p \leqslant x} \frac{p^{-1/\log x}(1 + e \log p / \log x)}{p} \\
&= \sum_{p \leqslant x} \frac{1}{p^{1+1/\log x}} + \frac{e}{\log x} \sum_{p \leqslant x} \frac{\log p}{p^{1+1/\log x}} \\
&\leqslant \sum_{p} \frac{1}{p^{1+1/\log x}} + \frac{e}{\log x} \sum_{p} \frac{\log p}{p^{1+1/\log x}} \\
&\leqslant \log\log x + O(1)
\end{aligned}$$

d'après le corollaire 6.3 avec $\sigma = 1 + 1/\log x$.

De façon similaire, nous avons que

$$\begin{aligned}
\sum_{p \leqslant x} \frac{1}{p} &\geqslant \sum_{p \leqslant x} \frac{1}{p^{1+1/\log x}} \\
&= \sum_{p} \frac{1}{p^{1+1/\log x}} - \sum_{p > x} \frac{1}{p^{1+1/\log x}} \\
&\geqslant \sum_{p} \frac{1}{p^{1+1/\log x}} - \sum_{p > x} \frac{\log p / \log x}{p^{1+1/\log x}} \\
&\geqslant \sum_{p} \frac{1}{p^{1+1/\log x}} - \sum_{p} \frac{\log p / \log x}{p^{1+1/\log x}} \\
&= \log\log x + O(1).
\end{aligned}$$

Ceci conclut la démonstration. □

6.3 Le crible d'Eratosthène revisité

Dans la section 4.1, on a vu que le crible d'Eratosthène nous permet de trouver tous les nombres premiers jusqu'à une borne donnée x. Dans cette section-ci, on va examiner le crible d'Eratosthène comme un outil théorique.

On se rappelle que le crible d'Eratosthène se base fondamentalement au théorème 4.1, qui dit que si $n > 1$ est composé, alors il a un facteur premier $p \leqslant \sqrt{n}$. Un corollaire directe de ce fait est la relation suivante : pour tout $x \geqslant 1$, on a que

$$\{n \leqslant x : p \leqslant \sqrt{x} \implies p \nmid n\} = \{1\} \cup \{\sqrt{x} < p \leqslant x\}. \tag{6.6}$$

En effet, le nombre 1 appartient aux ensembles de tous les deux côtés de (6.6). Or, soit $1 < n \leqslant x$. Si n n'a pas de facteurs premiers $\leqslant \sqrt{x}$, il n'a pas de facteurs premiers $\leqslant \sqrt{n}$ non plus. Donc, selon le théorème 4.1, n est un nombre premier. Puisque n n'a pas de facteurs premiers $\leqslant \sqrt{x}$, il doit être un nombre premier $> \sqrt{x}$. Ceci montre que l'ensemble au côté droit de (6.6) est contenu à ce du côté gauche. Il est facile de voir que l'inclusion converse est aussi vraie. Ceci montre (6.6).

Or, on interprète (6.6) d'un point de vu combinatoire qui s'applique facilement dans un contexte plus général. Supposons que nous avons un ensemble fini $\mathcal{A} \subset \mathbb{N}$ et un ensemble fini de nombres premiers $\mathcal{P}$, et que nous voulons comprendre combien d'éléments de $\mathcal{A}$ n'ont pas de facteurs premiers dans $\mathcal{P}$. Pour cette raison, on définit la quantité

$$S(\mathcal{A}, \mathcal{P}) := \{a \in \mathcal{A} : p \in \mathcal{P} \implies p \nmid a\}.$$

Voici certains exemples pour voir le type de questions que cette quantité captive :

(a) Si $\mathcal{A} = \{n \leqslant x\}$ et $\mathcal{P} = \{p \leqslant \sqrt{x}\}$, alors la relation (6.6) implique que

$$S(\mathcal{A}, \mathcal{P}) = 1 + \#\{\sqrt{x} < p \leqslant x\} = 1 + \pi(x) - \pi(\sqrt{x}).$$

(b) Si m est un nombre naturel, $\mathcal{A} = \{1, 2, \dots, m\}$ et $\mathcal{P} = \{p|m\}$, alors

$$S(\mathcal{A}, \mathcal{P}) = \phi(m).$$

(c) Si $\mathcal{A} = \{n(n+2) : 1 < n \leqslant x\}$ et $\mathcal{P} = \{p \leqslant \sqrt{x+2}\}$, alors un argument similaire avec ce menant à (6.6) montre que

$$S(\mathcal{A}, \mathcal{P}) = \#\{\sqrt{x+2} < p \leqslant x : p + 2 \text{ est aussi un nombre premier}\}.$$

On voit donc que la quantité $S(\mathcal{A}, \mathcal{P})$ compte ici paires de nombres premiers jumeaux, c'est paires de nombres premiers (p, q) avec $q = p + 2$. Une conjecture fameuse ouverte en théorie des nombres prédit qu'il existe une infinité de telles paires. Un théorème récent [11] se basant sur les travails de Zhang–Maynard–Tao spectaculaires [12, 6, 10] montre qu'il existe une infinité de paires de nombres premiers (p, q) avec $p < q \leqslant p + 246$.

Afin d'évaluer $S(\mathcal{A}, \mathcal{P})$, on utilise le principe d'inclusion-exclusion. On fait, on veut compter tous les éléments de $\mathcal{A}$ qui n'appartiennent pas à aucun sous-ensemble de la forme $\mathcal{A}_p := \{a \in \mathcal{A} : p|a\}$ avec $p \in \mathcal{P}$. Le principe d'inclusion-exclusion nous dit que

$$S(\mathcal{A}, \mathcal{P}) = |\mathcal{A}| - \sum_{p \in \mathcal{P}} |\mathcal{A}_p| + \sum_{\substack{p_1 < p_2 \\ p_1, p_2 \in \mathcal{P}}} |\mathcal{A}_{p_1} \cap \mathcal{A}_{p_2}| - \sum_{\substack{p_1 < p_2 < p_3 \\ p_1, p_2, p_3 \in \mathcal{P}}} |\mathcal{A}_{p_1} \cap \mathcal{A}_{p_2}| \pm \cdots. \tag{6.7}$$

On observe que si $p_1, \dots, p_k$ sont quelques nombres premiers distincts, alors les k conditions $p_1|a, p_2|a, \dots, p_k|a$ sont équivalentes à la condition $p_1 p_2 \cdots p_k|a$. Donc,

$$\mathcal{A}_{p_1} \cap \mathcal{A}_{p_2} \cap \cdots \cap \mathcal{A}_{p_k} = \mathcal{A}_{p_1 \cdots p_k},$$

où on l'a posé

$$\mathcal{A}_d := \{a \in \mathcal{A} : d|a\}.$$

De plus, le signe du terme $|\mathcal{A}_{p_1} \cap \mathcal{A}_{p_2} \cap \cdots \cap \mathcal{A}_{p_k}|$ dans (6.7) est égal à $(-1)^k = \mu(p_1 \cdots p_k)$. On conclut que

$$S(\mathcal{A}, \mathcal{P}) = \sum_{d|P} \mu(d)|\mathcal{A}_d|, \tag{6.8}$$

où

$$P := \prod_{p \in \mathcal{P}} p.$$

Remarque 6.4. Il y a une façon alternative d'arriver à (6.8). On observe que

$$S(\mathcal{A}, \mathcal{P}) = \#\{a \in \mathcal{A} : (a, P) = 1\} = \sum_{a \in \mathcal{A}} 1_{(a,P)=1}.$$

Puis, on se rappelle que la formule d'inversion de Möbius dit que $\sum_{d|n} \mu(d) = 1_{n=1}$. En appliquant cette formule avec $n = (a, P)$ et en utilisant le lemme 2.7, on trouve que

$$S(\mathcal{A}, \mathcal{P}) = \sum_{a \in \mathcal{A}} \sum_{d|(a,P)} \mu(d) = \sum_{a \in \mathcal{A}} \sum_{d|a,\ d|P} \mu(d).$$

La prochaine étape est de changer l'ordre de sommation des variables a et d, et d'exécuter d'abord la sommation sur d et puis celle sur a. Il faut que $d|P$ et, étant donné un tel d, les seules conditions sur a sont que $a \in \mathcal{A}$ et que $d|a$. Donc, on trouve que

$$S(\mathcal{A}, \mathcal{P}) = \sum_{d|P} \sum_{\substack{a \in \mathcal{A} \\ d|a}} \mu(d) = \sum_{d|P} \mu(d) \sum_{\substack{a \in \mathcal{A} \\ d|a}} 1 = \sum_{d|P} \mu(d)|\mathcal{A}_d|,$$

ce qui donne une montre la formule (6.8).

On voit que la formule d'inversion de Möbius encode le principe d'inclusion-exclusion qui est la clé pour «produire» les nombres premiers. Donc, la fonction de Möbius est liée d'une façon fondamentale avec la répartition des nombres premiers. Par exemple, puisque les valeurs de μ se trouvent toujours dans $\{-1, 0, 1\}$, l'inégalité du triangle implique que

$$\Big|\sum_{n \leqslant x} \mu(n)\Big| \leqslant \sum_{n \leqslant x} 1 \leqslant x.$$

Landau a montré que le théorème des nombres premiers, qui dit que $\pi(x) \sim x/\log x$ lorsque $x \to \infty$, est équivalent à la relation asymptotique $\sum_{n \leqslant x} \mu(n) = o(x)$ lorsque $x \to \infty$. En mots, cette dernière relation dit que si on choisit on entier dans $[1, x]$ sans facteurs carrés, alors la probabilité d'avoir un nombre pair de facteurs premiers est 50% et la probabilité d'avoir un nombre impair de facteurs premiers est aussi 50%. Donc, afin de montrer le théorème des nombres premiers, il faut montrer qu'il n'y a pas de conspiration qui biaise la parité du nombre de facteurs premiers d'un entier «aléatoire».

Or, utilisons la relation (6.8) quand $\mathcal{A} = \{n \leqslant x\}$, avec le but d'obtenir une estimation asymptotique de $\pi(x)$. Dans ce cas spécial, et en abusant un peut la notation, posons

$$S(x, \mathcal{P}) := \#\{n \leqslant x : p|n \implies p \notin \mathcal{P}\} = \#\{n \leqslant x : (n, P) = 1\}.$$

D'après (6.8), on a que

$$S(x, \mathcal{P}) = \sum_{d|P} \mu(d) \#\{n \leqslant x : d|n\}.$$

On peut paramétriser les entiers $n \leqslant x$ qui sont divisibles par d comme $n = md$, où $1 \leqslant m \leqslant x/d$. On a donc $\lfloor x/d \rfloor$ pour le paramètre m, ce qui implique que

$$S(x, \mathcal{P}) = \sum_{d|P} \mu(d) \lfloor x/d \rfloor .$$

Puis, on approxime $\lfloor x/d \rfloor$ par x/d. En effet, on a que $\lfloor x/d \rfloor = x/d - \{x/d\}$, et donc

$$S(x, \mathcal{P}) = x \sum_{d|P} \frac{\mu(d)}{d} - \sum_{d|P} \mu(d)\{x/d\}.$$

On ne peut pas contrôler l'erreur en général, donc on l'estime de façon triviale : on observe que $|\mu(d)| \leqslant 1$ et que $0 \leqslant \{x/d\} < 1$ pour chaque d, donc l'inégalité du triangle implique que

$$\Big| \sum_{d|P} \mu(d)\{x/d\} \Big| \leqslant \sum_{d|P} 1 = \tau(P) = 2^{|\mathcal{P}|},$$

où on a utilisé le corollaire 5.4 et le fait que $P = \prod_{p \in \mathcal{P}} p$. En conclusion, on a que

$$S(x, \mathcal{P}) = x \sum_{d|P} \frac{\mu(d)}{d} + O(2^{|\mathcal{P}|}).$$

Il reste à estimer le terme principal.

La fonction $n \to \sum_{d|n} \mu(d)/d$ est multiplicative car elle est égale à $(\mu/\mathrm{id}) * 1$ (on utilise ici le théorème 5.8). Donc, on a que

$$\sum_{d|P} \frac{\mu(d)}{d} = \prod_{p \in \mathcal{P}} \sum_{d|p} \frac{\mu(d)}{d} = \prod_{p \in \mathcal{P}} \Big(1 - \frac{1}{p}\Big).$$

On a alors montré le théorème suivant :

Théorème 6.5. *Soit $x \geqslant 1$ et soit $\mathcal{P}$ un ensemble de nombres premiers fini. Donc,*

$$S(x, \mathcal{P}) = x \prod_{p \in \mathcal{P}} \Big(1 - \frac{1}{p}\Big) + O(2^{|\mathcal{P}|}).$$

Remarque. Ce théorème a une belle interprétation probabiliste. Le nombre de multiples de p dans $[1, x]$ est $\lfloor x/p \rfloor \approx x/p$. Donc, si on choisit un entier n uniformément au hasard de $[1, x]$, la probabilité qu'il n'est pas un multiple de p est $\approx 1 - 1/p$. Si tous ces événements pour $p \in \mathcal{P}$ étaient indépendants, on aurait alors que la probabilité que n n'est pas divisible par aucun $p \in \mathcal{P}$ est $\approx \prod_{p \in \mathcal{P}}(1 - 1/p)$. Le théorème 6.5 montre que cet argument heuristique est vrai d'une façon approximative si l'ensemble $\mathcal{P}$ n'est pas trop grand (pour que l'erreur $O(2^{|\mathcal{P}|})$ reste sous contrôle).

On utilise maintenant le théorème 6.5 pour estimer $\pi(x)$. Le problème est que si on prend $\mathcal{P} = \{p \leqslant \sqrt{x}\}$, pour que $S(x, \mathcal{P}) = 1 + \#\{\sqrt{x} < p \leqslant x\}$, alors l'erreur devient $2^{\#\{p \leqslant \sqrt{x}\}}$. On s'attend à ce que $\#\{p \leqslant \sqrt{x}\} \sim \sqrt{x}/\log\sqrt{x}$ d'après le théorème des nombres premiers, donc l'«erreur» est beaucoup plus grand que le terme principal présumé, qui a taille $\leqslant x$. [4]

Legendre a trouvé une façon de contourner ce problème, en prenant un ensemble $\mathcal{P}$ plus petit. Soit $y \in [1, \sqrt{x}]$ un paramètre et soit $\mathcal{P} = \{p \leqslant y\}$. Observons que si p est un nombre premier dans $(y, x]$, alors il est compté par $S(x, \mathcal{P})$ car il est dans $[1, x]$ et il n'est pas divisible par aucun nombre premier $\leqslant y$. Donc,

$$\#\{y < p \leqslant x\} \leqslant S(x, \mathcal{P}) \quad \Longrightarrow \quad \pi(x) \leqslant S(x, \mathcal{P}) + y,$$

où on a utilisé la borne triviale $\#\{p \leqslant y\} \leqslant y$. D'autre côté, on a que

$$S(x, \mathcal{P}) = x \prod_{p \leqslant y} \Big(1 - \frac{1}{p}\Big) + O(2^y),$$

d'après le théorème 6.5, où on a utilisé encore une fois la borne triviale $\#\{p \leqslant y\} \leqslant y$. Prenons $y = \log x$, pour que $2^y = e^{(\log 2)(\log x)} = x^{\log 2}$, qui est beaucoup plus petit que x. En conclusion, nous avons que

$$\pi(x) \leqslant x \prod_{p \leqslant \log x} \Big(1 - \frac{1}{p}\Big) + O(x^{\log 2}), \tag{6.9}$$

où on a utilisé que $y = O(2^y)$. Afin de procéder, il faut avoir une estimation pour le produit au côté droit. On peut prendre logarithmes, utiliser la formule de Taylor $\log(1 - x) = -\sum_{k \geqslant 1}^{\infty} x^k/k$ avec $x = 1/p$, et puis appliquer le théorème 6.1. Par contre, il y a une façon plus simple et plus nette de faire cette estimation.

Lemme 6.6. *Pour $y > 1$, nous avons que*

$$\prod_{p \leqslant y} \Big(1 - \frac{1}{p}\Big) \leqslant \frac{1}{\log y}.$$

Démonstration. En inversant les étapes menant à (5.10), nous avons que

$$\prod_{p \leqslant y} \Big(1 - \frac{1}{p}\Big)^{-1} = \prod_{p \leqslant y} \Big(1 + \frac{1}{p} + \frac{1}{p^2} + \cdots\Big) = \sum_{\substack{n \geqslant 1 \\ p \mid n \,\Rightarrow p \leqslant y}} \frac{1}{n}.$$

En particulier, si $n \leqslant y$, alors n est parmi les sommés de cette dernière somme. Donc,

$$\prod_{p \leqslant y} \Big(1 - \frac{1}{p}\Big)^{-1} \geqslant \sum_{n \leqslant y} \frac{1}{n} \geqslant \sum_{n \leqslant y} \int_n^{n+1} \frac{\mathrm{d}t}{t} \geqslant \int_1^y \frac{\mathrm{d}t}{t} = \log y.$$

Ceci termine la preuve. □

4. La façon la plus simple de comparer les fonctions $2^{\sqrt{x}/\log\sqrt{x}}$ et x et de prendre leurs logarithmes, qui sont $\sqrt{x}\log 4/\log x$ et $\log x$, respectivement. Clairement, la première fonction est beaucoup plus grande que la deuxième. Pour le voir, on peut par exemple utiliser la règle de l'Hôpital pour montrer que $\lim_{x\to\infty} \frac{\sqrt{x}/\log x}{\log x} = \infty$.

En combinant le lemme 6.6 avec la relation (6.9), on déduit que

$$\pi(x) \leqslant \frac{x}{\log\log x} + O(x^{\log 2}) \ll \frac{x}{\log\log x} \tag{6.10}$$

pour $x \geqslant 3$. En particulier, on voit que, d'un point de vue asymptotique, 0% d'entiers sont de nombres premiers. Par contre cette estimation est loin de la prédiction de Gauss que $\pi(x) \sim x/\log x$. À la section suivante, on verra une approche alternative et toujours élémentaire qui nous mènera à une estimation beaucoup plus précise de $\pi(x)$.

Remarque. Le travail de Legendre restait pour longtemps une observation intéressante mais très utile. Au début du 20ième siècle, le mathématicien Viggo Brun a transformé complètement le crible d'Eratosthène–Legendre à un outil très flexible et indispensable en théorie des nombres analytique.

6.4 Une estimation de $\pi(x)$

En 1852, avant la preuve du théorème des nombres premiers par Hadamard et de la Vallée-Poussin, Tchebyshev a réussi à montrer l'estimation suivante :

Théorème 6.7 (Tchebyshev)**.** *Pour tout $x \geqslant 3$, on a que*

$$\frac{x}{2\log x} \leqslant \pi(x) \leqslant \frac{6x}{\log x}.$$

Le point de départ de la preuve du théorème 6.7 est l'identité de convolution (5.14), qui dit que

$$\log = \Lambda * 1.$$

En ajoutant cette formula sur tous les entiers $n \in \{1, \dots, N\}$, on trouve que

$$\sum_{n=1}^{N} \log n = \sum_{n=1}^{N} (\Lambda * 1)(n). \tag{6.11}$$

On peut estimer le côté droit assez facilement :

Lemme 6.8. *Pour tout $N \in \mathbb{N}$, on a que*

$$N\log N - N + 1 \leqslant \sum_{n=1}^{N} \log n \leqslant (N+1)\log N - N + 1.$$

Démonstration. Le logarithme est une fonction croissante. Par conséquent, on a les inégalités $\int_{n-1}^{n} \log t \,\mathrm{d}t \leqslant \log n \leqslant \int_{n}^{n+1} \log t \,\mathrm{d}t$ pour chaque entier $n \geqslant 2$. On en déduit que

$$\sum_{n=1}^{N} \log n = \sum_{n=2}^{N} \log n \geqslant \int_{1}^{N} \log t \,\mathrm{d}t = N\log N - N + 1$$

et que

$$\sum_{n=1}^{N} \log n \leqslant \int_{2}^{N+1} \log t \,\mathrm{d}t = (N+1)\log(N+1) - N + 1 - \log 4 \leqslant (N+1)\log N - N + 1,$$

puisque $(N+1)\log(1+1/N) \leqslant \log 4$ pour tout $N \geqslant 1$. Ceci termine la preuve. □

Puis, examinons le côté droit de (6.11). En ouvrant la convolution de Dirichlet, on a que

$$\sum_{n=1}^{N}(\Lambda * 1)(n) = \sum_{n=1}^{N}\sum_{a|n}\Lambda(a) = \sum_{a\leqslant N}\Lambda(a)\sum_{\substack{1\leqslant n\leqslant N\\ a|n}} 1 = \sum_{a\leqslant N}\Lambda(a)\lfloor x/a\rfloor .$$

Nous concluons alors que

$$\sum_{a=1}^{N}\Lambda(a)\lfloor N/a\rfloor = \log(N!) = N\log N - N + O(\log N) \quad \text{pour tout } N\in\mathbb{N}. \tag{6.12}$$

On pense à (6.12) comme un système linéaire infini dont les coefficients sont $\Lambda(a)$. Notre but est d'extraire de ces donnés une estimation pour $\sum_{a=1}^{N}\Lambda(a)$ premièrement et, ensuite, pour $\pi(x)$. Bien sûr, on peut utiliser la formule d'inversion de Möbius (cf. exercice 5.7), mais ceci est d'utilité limitée sans une bonne compréhension de la fonction μ, quelque chose qu'on manque avec les outils théoriques qu'on a vu jusqu'à maintenant. Par contre, comme on va le voir, il y a une façon de manipuler le système d'équations (6.12) de sorte qu'on évite la fonction μ.

La première étape est d'examiner les coefficient de $\Lambda(a)$ dans (6.12) est $\lfloor N/a\rfloor$. On a que

$$\lfloor N/a\rfloor = k \qquad\Longleftrightarrow\qquad N/(k+1) < a \leqslant N/k.$$

Donc, on trouve que

$$\sum_{a=1}^{N}\Lambda(a)\lfloor N/a\rfloor = \sum_{N/2<a\leqslant N}\Lambda(a) + 2\sum_{N/3<a\leqslant N/2}\Lambda(a) + 3\sum_{N/4<a\leqslant N/3}\Lambda(a) + \cdots \tag{6.13}$$

Les coefficients varie trop pour utiliser cette expression afin d'estimer $\sum_{a=1}^{N}\Lambda(a)$. Par contre, on peut réduire les coefficients en comparant (6.12) plus qu'une fois : on a que

$$\begin{aligned}\sum_{a=1}^{2N}\Lambda(a)\lfloor 2N/a\rfloor = {} & \sum_{N<a\leqslant 2N}\Lambda(a) + 2\sum_{2N/3<a\leqslant N}\Lambda(a) + 3\sum_{N/2<a\leqslant 2N/3}\Lambda(a)\\ & + 4\sum_{2N/5<a\leqslant N/2}\Lambda(a) + 5\sum_{N/3<a\leqslant 2N/5}\Lambda(a) + 6\sum_{2N/7<a\leqslant N/3}\Lambda(a) + \cdots\end{aligned} \tag{6.14}$$

Afin de comparer cette expression avec celle de (6.13), on réécrit (6.13) dans une forme plus pratique :

$$\begin{aligned}\sum_{a=1}^{N}\Lambda(a)\lfloor N/a\rfloor = {} & \sum_{2N/3<a\leqslant N}\Lambda(a) + \sum_{N/2<a\leqslant 2N/3}\Lambda(a)\\ & + 2\sum_{2N/5<a\leqslant N/2}\Lambda(a) + 2\sum_{N/3<a\leqslant 2N/5}\Lambda(a) + 3\sum_{2N/7<a\leqslant N/3}\Lambda(a)\\ & + 3\sum_{N/4<a\leqslant 2N/7}\Lambda(a) + \cdots\end{aligned} \tag{6.15}$$

Si on multiplie (6.15) par 2 et la soustraire de (6.14), on trouve que

$$\sum_{a=1}^{2N}\Lambda(a)\,\lfloor 2N/a\rfloor - 2\sum_{n=1}^{N}\Lambda(a)\,\lfloor N/a\rfloor \\ = \sum_{N<a\leqslant 2N}\Lambda(a) + \sum_{N/2<a\leqslant 2N/3}\Lambda(a) + \sum_{N/3<a\leqslant 2N/5}\Lambda(a) + \cdots .$$

En particulier, on a que

$$\sum_{a=N+1}^{2N}\Lambda(a) \leqslant \sum_{a=1}^{2N}\Lambda(a)\,\lfloor 2N/a\rfloor - 2\sum_{n=1}^{N}\Lambda(a)\,\lfloor N/a\rfloor \leqslant \sum_{a=1}^{2N}\Lambda(a).$$

D'autre côté, (6.11) implique

$$\sum_{a=1}^{2N}\Lambda(a)\,\lfloor 2N/a\rfloor - 2\sum_{n=1}^{N}\Lambda(a)\,\lfloor N/a\rfloor = \sum_{n=1}^{2N}\log n - 2\sum_{n=1}^{N}\log n.$$

On conclut alors que

$$\sum_{a=N+1}^{2N}\Lambda(a) \leqslant \sum_{n=1}^{2N}\log n - 2\sum_{n=1}^{N}\log n \leqslant \sum_{a=1}^{2N}\Lambda(a). \tag{6.16}$$

Il reste à estimer l'expression $\sum_{n=1}^{2N}\log n - 2\sum_{n=1}^{N}\log n$. On pourrait tout simplement utiliser le lemme 6.8 deux fois pour voir que cette expression est égale à $2N\log(2N) - 2N - 2(N\log N - N) + O(\log N) = N\log 4 + O(\log N)$. En fait, on peut être un peu plus précis :

Lemme 6.9. *Pour tout $N \in \mathbb{N}$, on a que*

$$N\log 4 - \log(2N+1) \leqslant \sum_{n=1}^{2N}\log n - 2\sum_{n=1}^{N}\log n \leqslant N\log 4.$$

Démonstration. On observe que

$$\sum_{n=1}^{2N}\log n - 2\sum_{n=1}^{N}\log n = \sum_{n=N+1}^{2N}\log n - \sum_{n=1}^{N}\log n = \sum_{n=1}^{N}\big(\log(N+n) - \log n\big).$$

La fonction $x \to \log(N+x) - \log x = \log(1+N/x)$ est décroissante pour $x > 0$, donc

$$\sum_{n=1}^{N}\big(\log(N+n) - \log n\big) \leqslant \sum_{n=1}^{N}\int_{n-1}^{n}\big(\log(N+x) - \log x\big)\mathrm{d}x \\ = \int_{0}^{N}\big(\log(N+x) - \log x\big)\mathrm{d}x = N\log 4.$$

De façon similaire, on trouve aussi que

$$\begin{aligned}\sum_{n=1}^{2N}\log n - 2\sum_{n=1}^{N}\log n &\geqslant \int_1^{N+1}\big(\log(N+x)-\log x\big)\mathrm{d}x\\ &= (2N+1)\log(2N+1) - 2(N+1)\log(N+1)\\ &= 2(N+1)\log\frac{2N+1}{N+1} - \log(2N+1)\\ &\geqslant N\log 4 - \log(2N+1),\end{aligned}$$

car $(1+1/N)\log\frac{2+1/N}{1+1/N} \geqslant \log 2$ pour tout $N \geqslant 1$, du fait que la fonction $(1+x)\log\frac{2+x}{1+x}$ est croissante pour $x \geqslant 0$ et vaut $\log 2$ en $x = 0$. Ceci termine la preuve. □

En combinant ce lemme avec (6.16), on arrive aux inégalités suivantes :

$$\sum_{a\leqslant 2N}\Lambda(a) \geqslant N\log 4 - \log(2N+1) \tag{6.17}$$

et

$$\sum_{N<p\leqslant 2N}\log p \leqslant \sum_{a\leqslant 2N}\Lambda(a) \leqslant N\log 4 \tag{6.18}$$

pour tout entier $N \geqslant 1$. On est maintenant prêt à montrer l'estimation de Tchebyshev.

Démonstration du théorème 6.7. On peut vérifier numériquement que le théorème est vrai quand $x \in [3, 100]$. Or, supposons que $x > 100$.

Pour la borne inférieure, on observe que

$$\begin{aligned}\sum_{a\leqslant 2N}\Lambda(a) = \sum_{\substack{p \text{ premier},\, k\in\mathbb{N}\\ p^k\leqslant x}}\sum \log p &= \sum_{p\leqslant 2N}(\log p)\sum_{1\leqslant k\leqslant \log(2N)/\log p} 1\\ &\leqslant \sum_{p\leqslant 2N}(\log p)\cdot\frac{\log 2N}{\log p}\\ &= \pi(2N)\log(2N).\end{aligned}$$

En combinant cette relation avec (6.17), on trouve que

$$\pi(2N) \geqslant \frac{N\log 4 - \log(2N+1)}{\log(2N)} \geqslant \frac{1.1N}{\log(2N)} \quad \text{pour } N \geqslant 12.$$

On pose $N := \lfloor x/2\rfloor \geqslant 50$ et on observe que $x/2 \geqslant N \geqslant x/2 - 1 \geqslant \frac{49}{100}x$. Donc,

$$\pi(x) \geqslant \pi(2N) \geqslant \frac{1.1N}{\log(2N)} \geqslant \frac{x}{2\log x}.$$

Passons maintenant à la borne supérieure. Fixons $k \in \mathbb{Z}_{\geqslant 0}$ tel que $2^k < x \leqslant 2^{k+1}$; en fait, il faut que $k \geqslant 6$ de notre hypothèse que $x \geqslant 100$. Fixons aussi $\ell \in \{1, \dots, k\}$ tel que $2^\ell \leqslant \sqrt{x} < 2^{\ell+1}$.

Si p est un nombre premier $\leqslant x$, soit $p \leqslant x^{1/2}$ ou $\sqrt{x} < p \leqslant x$. Au deuxième cas, il existe $j \in \{\ell, \ell+1, \dots, k\}$ tel que $2^j < p \leqslant 2^{j+1}$. Par la suite,

$$\pi(x) \leqslant x^{1/2} + \sum_{j=\ell}^{k} \sum_{\substack{2^j < p \leqslant 2^{j+1} \\ p > x^{1/2}}} 1 \leqslant x^{1/2} + \sum_{j=\ell}^{k} \sum_{2^j < p \leqslant 2^{j+1}} \frac{\log p}{\log(x^{1/2})}.$$

Donc, la relation (6.18), appliquée avec $N = 2^j$, implique que

$$\pi(x) \leqslant x^{1/2} + \sum_{j=\ell}^{k} \frac{2^j \log 4}{\log(x^{1/2})}.$$

On a que $\sum_{j=0}^{k} 2^j = 2^{k+1} - 2^{\ell+1} \leqslant 2x - \sqrt{x}$, donc on conclut que

$$\pi(x) \leqslant x^{1/2} + \frac{4(\log 4)x - 2(\log 4)\sqrt{x}}{\log x} \leqslant \frac{6x}{\log x}$$

pour $x \geqslant 100$. Ceci conclut la preuve du théorème 6.7 de Tchebyshev. □

Remarque. Souvent, l'argument ci-dessus est donné en termes plus combinatoires. Considérons le coefficient binomial central

$$\binom{2N}{N} = \frac{(N+1)\cdots(2N)}{N!}.$$

Il s'agit d'un entier qui se trouve dans l'intervalle $[4^N/(2N+1), 4^N]$ (pour le voir, on utilise le théorème du binôme à l'expression $(1+1)^{2N}$, en observant que $\max_{0 \leqslant k \leqslant 2N} \binom{2N}{k} = \binom{2N}{N}$). De plus, on observe que si $p \in (N, 2N]$, alors p divise le numérateur $(N+1)\cdots(2N)$ et il est co-premier avec $N!$. Donc, il faut que $p | \binom{2N}{N}$ pour tout nombre premier $p \in (N, 2N]$. Donc, $\prod_{N < p \leqslant 2N} p$ divise $\binom{2N}{N}$, ce qui implique que

$$\prod_{N < p \leqslant 2N} p \leqslant \binom{2N}{N} \leqslant 4^N.$$

En prenant logarithmes, on déduit (une version plus nette de) la relation (6.18).

On peut aussi déduire (6.17). Tous les nombres premiers qui divisent $\binom{2N}{N}$ sont $\leqslant 2N$. De plus, on a que

$$v_p\left(\binom{2N}{N}\right) = v_p(2N!) - 2v_p(N!).$$

On observe aussi que

$$v_p(m!) = \sum_{k=1}^{\infty} \lfloor m/p^k \rfloor \quad \text{pour tout } m \in \mathbb{N}.$$

Donc

$$\log \binom{2N}{N} = \sum_{p \leqslant 2N} v_p\left(\binom{2N}{N}\right) \log p = \sum_{p \leqslant 2N} \sum_{k=1}^{\infty} \left(\lfloor 2N/p^k \rfloor - 2 \lfloor N/p^k \rfloor \right) \log p.$$

Puisque le côté droit est $\geqslant \log(4^N/(2N+1))$, on déduit que

$$\sum_{p\leqslant 2N}\sum_{k=1}^{\infty}\big(\,\lfloor 2N/p^k\rfloor - 2\,\lfloor N/p^k\rfloor\,\big)\log p \geqslant N\log 4 - \log(2N+1),$$

ce qui montre (6.17) d'une façon alternative.

6.5 Le nombre de facteurs premiers d'un entier aléatoire

La dernière application des méthodes analytiques est une estimation de certaines propriétés statistiques de la fonction ω qui compte le nombre de facteurs premiers.

Théorème 6.10. *Pour $x \geqslant 3$, on a que*

$$\sum_{n\leqslant x}\omega(n) = x\log\log x + O(x) \tag{6.19}$$

et que

$$\sum_{n\leqslant x}\big(\omega(n) - \log\log x\big)^2 = O(x\log\log x). \tag{6.20}$$

Avant montrer le théorème 6.10, on fait quelques remarques importants. On peut interpréter ce résultat d'un point de vue probabiliste : la relation (6.19) montre que la valeur moyenne de $\omega(n)$ est $\log\log x + O(1)$ quand n est choisi uniformément au hasard de $[1,x]\cap\mathbb{Z}$, et la relation (6.20) montre que la variance de ω est $\ll \log\log x$. On sait de la théorie des probabilités qu'une telle variable aléatoire se concentre presque toujours autour de sa moyenne. En effet, nous avons le corollaire suivant du théorème 6.10 qui montre que presque tous les entiers dans $[1,x]$ ont $\sim \log\log x$ facteurs premiers.

Corollaire 6.11. *Soit $\varepsilon > 0$. Alors, on a que*

$$\lim_{x\to\infty}\frac{\#\big\{n\leqslant x : (1-\varepsilon)\log\log x \leqslant \omega(n) \leqslant (1+\varepsilon)\log\log x\big\}}{x} = 1.$$

Démonstration. Fixons $\varepsilon > 0$ et considérons l'ensemble $\mathcal{E}(x) := \{n \leqslant x : |\omega(n) - \log\log x| > \varepsilon\log\log x\}$. On veut montrer que $\frac{\#\mathcal{E}(x)}{x} \to 0$ lorsque $x\to\infty$. Pour chaque $n\in\mathcal{E}(x)$, on a que $(\omega(n)-\log\log x)^2 \geqslant (\varepsilon\log\log x)^2$. De plus, si $n\in[1,x]\setminus\mathcal{E}(x)$ on a que $(\omega(n)-\log\log x)^2 \geqslant 0$. Donc,

$$\#\mathcal{E}(x) \leqslant \sum_{n\leqslant x}\frac{(\omega(n)-\log\log x)^2}{(\varepsilon\log\log x)^2} = O\Big(\frac{x}{\varepsilon^2\log\log x}\Big)$$

d'après le théorème 6.10. On voit alors que $\frac{\#\mathcal{E}(x)}{x}\to 0$, comme désiré. □

Démonstration du théorème 6.10. On a que $\omega(n) = \sum_{p\mid n} 1$, donc

$$\sum_{n\leqslant x}\omega(n) = \sum_{n\leqslant x}\sum_{p\mid n}1 = \sum_{p\leqslant x}\sum_{\substack{n\leqslant x\\ p\mid n}}1 = \sum_{p\leqslant x}\lfloor x/p\rfloor\,.$$

On utilise le fait que $\lfloor x/p \rfloor = x/p + O(1)$ pour voir que

$$\sum_{n \leqslant x} \omega(n) = x \sum_{p \leqslant x} \frac{1}{p} + \sum_{p \leqslant x} O(1).$$

Pour estimer la première somme au côté droit, on utilise le théorème 6.1. Pour la somme de grand-Os, on observe tout simplement que $\sum_{p \leqslant x} O(1) = O(\pi(x)) = O(x)$ de façon triviale. Ceci montre (6.19).

Passons maintenant à la preuve de (6.20). Le côté gauche est toujours $\geqslant 0$, donc il suffit de montrer une borne supérieure seulement. En développant le carré, on trouve que

$$\sum_{n \leqslant x} (\omega(n) - \log\log x)^2 = \sum_{n \leqslant x} \omega(n)^2 - 2(\log\log x) \sum_{n \leqslant x} \omega(n) + (\log\log x)^2 \sum_{n \leqslant x} 1.$$

En utilisant (6.19) et l'estimation triviale $\sum_{n \leqslant x} 1 = x + O(1)$, on déduit que

$$\sum_{n \leqslant x} (\omega(n) - \log\log x)^2 = \sum_{n \leqslant x} \omega(n)^2 - x(\log\log x)^2 + O(x \log\log x).$$

On a donc réduire (6.20) à l'inégalité suivante :

$$\sum_{n \leqslant x} \omega(n)^2 \leqslant x(\log\log x)^2 + O(x \log\log x). \tag{6.21}$$

On observe que

$$\sum_{n \leqslant x} \omega(n)^2 = \sum_{n \leqslant x} \Big(\sum_{p \mid n} 1 \Big)^2 = \sum_{n \leqslant x} \sum_{p_1, p_2 \mid n} \sum 1 = \sum_{p_1, p_2 \leqslant x} \sum_{\substack{n \leqslant x \\ p_1, p_2 \mid n}} 1 \leqslant \sum_{p_1, p_2 \leqslant x} \sum \frac{x}{[p_1, p_2]}, \tag{6.22}$$

puisque les relations $p_1, p_2 | n$ sont équivalentes à la relation $[p_1, p_2] | n$ (cf. lemme 2.14) et on a utilise le fait que $\#\{n \leqslant x : d | n\} = \lfloor x/d \rfloor \leqslant x/d$ pour tout $x \geqslant 1$ et tout $d \in \mathbb{N}$.

La dernière somme de (6.22) se divise naturellement dans deux sous-sommes, dépendant de si $p_1 = p_2$ ou $p_1 \neq p_2$. Pour la première partie, on a que

$$S_1 := \sum_{p \leqslant x} \frac{x}{p} = x \log\log x + O(x)$$

d'après le théorème 6.1. De plus, on a que

$$\begin{aligned} S_2 := \sum_{\substack{p_1, p_2 \leqslant x \\ p_1 \neq p_2}} \sum \frac{x}{[p_1, p_2]} = \sum_{\substack{p_1, p_2 \leqslant x \\ p_1 \neq p_2}} \sum \frac{x}{p_1 p_2} &\leqslant x \Big(\sum_{p \leqslant x} \frac{1}{p} \Big)^2 \\ &= x(\log\log x + O(1))^2 \\ &= x(\log\log x)^2 + O(x \log\log x) \end{aligned}$$

en utilisant le théorème 6.1 encore une fois. En combinant (6.22) avec les estimations de S_1 et de S_2, on déduit la borne (6.21). Ceci termine la preuve du théorème 6.10. □

En 1940, Erdős et Kac [2] ont montré un résultat beaucoup plus précis que le théorème 6.10 : ils ont prouvé que si on normalise $\omega(n)$ en soutrant sa moyenne et en divisant cette différence par son écart type, alors on obtient une variable aléatoire qui converge en distribution vers la loi normale de Gauss.

Théorème 6.12 (Erdős – Kac)**.** *Soient $-\infty < \alpha \leqslant \beta < +\infty$ qu'on considère fixe. Alors, on a que*

$$\#\left\{n \leqslant x : \alpha \leqslant \frac{\omega(n) - \log\log x}{\sqrt{\log\log x}} \leqslant \beta\right\} \sim x \cdot \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-t^2/2}\mathrm{d}t \qquad \textit{lorsque } x \to \infty.$$

La preuve de ce théorème dépasse le contenu de ce cours. On mentionne tout simplement l'idée de base de sa démonstration.

La forme classique du théorème central limite dit que si $X_1, X_2, \dots$ sont quelques variables aléatoires équidistribuées et mutuellement indépendantes, dont la moyenne est égale μ et la variance à $\sigma^2 > 0$, alors la variable aléatoire

$$Y_n := \frac{X_n - \mu n}{\sigma\sqrt{n}}$$

converge en distribution vers la loi normale de Gauss lorsque $n \to \infty$, c'est-à-dire la fonction de répartition $\mathbb{P}(Y_n \leqslant u)$ tend vers $\Phi(u) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-t^2/2}\mathrm{d}t$ lorsque $n \to \infty$.

Il existe des formes beaucoup plus générales du théorème central limite. Par exemple, supposons que $X_1, X_2, \dots$ sont quelques variables aléatoires mutuellement indépendantes telles que X_i a moyenne μ_i et variance σ_i. Définissons

$$s_n := \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$$

et supposons que la *condition de Lyapunov* est vraie, c'est-à-dire qu'il existe $\delta > 0$ tel que

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}\big[|X_i - \mu_i|^{2+\delta}\big] = 0.$$

Alors, la variable aléatoire

$$\frac{X_1 + X_2 + \cdots + X_n - (\mu_1 + \mu_2 + \cdots + \mu_n)}{s_n}$$

converge en distribution vers la loi normale de Gauss.

Dans le contexte du théorème 6.12, on a que

$$\omega(n) = \sum_{p|n} 1 = \sum_{p\leqslant x} B_p(n),$$

où $B_p(n) = 1_{p|n}$ est une variable aléatoire de Bernoulli. On a que

$$\sum_{n\leqslant x} B_p(n) = \#\{n \leqslant x : p|n\} = \lfloor x/p \rfloor = x/p + O(1).$$

Donc, la moyenne de la variable aléatoire B_p est $\sim 1/p$ sur l'ensemble d'entiers $\{n \leqslant x\}$, muni de la distribution uniforme. De plus, on a que

$$\begin{aligned}\sum_{n\leqslant x}\Big(B_p(n)-1/p\Big)^2 &= \sum_{n\leqslant x} B_p(n)^2 - \frac{2}{p}\sum_{n\leqslant x} B_p(n) + \frac{1}{p^2}\sum_{n\leqslant x} 1\\ &= \frac{x}{p} - \frac{2x}{p^2} + \frac{x}{p^2} + O(1) = x\cdot\Big(\frac{1}{p}-\frac{1}{p^2}\Big)+O(1),\end{aligned}$$

où on a utilisé que $B_p^2 = B_p$. Donc, la variance de B_p sur $\{n \leqslant x\}$ est $1/p - 1/p^2$. En utilisant le théorème 6.1, on voit que la somme des moyennes de B_p est égale à

$$\sim \sum_{p\leqslant x}\frac{1}{p} \sim \log\log x$$

et la somme de leurs variances est égale à

$$\sim \sum_{p\leqslant x}\Big(\frac{1}{p}-\frac{1}{p^2}\Big) = \log\log x + O(1) \sim \log\log x,$$

puisque $\sum_{p\leqslant x} 1/p^2 = O(1)$ du fait que la série infinie $\sum_p 1/p^2$ converge. Si les variables aléatoires B_p étaient mutuellement indépendantes, le théorème central limite impliquerait alors que $\omega(n) - \log\log x)/\sqrt{\log\log x}$ converge en distribution vers la loi normale de Gauss, comme le théorème 6.12 le dit.

Le problème ici est que les variables aléatoires B_p ne sont pas indépendantes. Par contre, on peut montrer assez facilement qu'elles sont «quasi-indépendantes» : si $p_1, \ldots, p_m$ sont de nombres premiers distincts, alors

$$\begin{aligned}\#\big\{n\leqslant x : B_{p_1}(n) = \cdots = B_{p_m}(n) = 1\big\} &= \#\big\{n\leqslant x : p_1\cdots p_m | n\big\}\\ &= \lfloor x/(p_1\cdots p_m)\rfloor\\ &= \frac{x}{p_1\cdots p_m} + O(1).\end{aligned} \tag{6.23}$$

On voit donc que si $p_1\cdots p_m$ est petit par rapport à x, alors

$$\frac{\#\big\{n\leqslant x : B_{p_1}(n) = \cdots = B_{p_m}(n) = 1\big\}}{x} \sim \prod_{j=1}^{m}\frac{\#\big\{n\leqslant x : B_{p_j}(n) = 1\big\}}{x}.$$

Cette indépendance approximative des variables B_p suffit pour établir le théorème 6.12. Pour les détails de la preuve, voir le chapitre 15 du livre [5].

6.6 Exercices

EXERCICE 6.1. Soit $m \in \mathbb{N}$.

(a) Montrer que

$$\sum_{n=1}^{\infty}\frac{(\log n)^m}{n^\sigma} = \frac{1}{m!(\sigma-1)^{m+1}} + O_m(1)$$

pour tout $\sigma \in [1,2]$ (i.e., la constante implicite dans le grand O dépend de m, mais pas de σ).

(b) Montrer (de façon heuristique si vous n'avez pas étudié encore la théorie de convergence uniforme des séries de fonctions) que

$$\left(\frac{-\zeta'}{\zeta}\right)^{(m-1)}(\sigma) = \sum_{n=1}^{\infty} \frac{\Lambda(n)(\log n)^{m-1}}{n^{\sigma}}.$$

pour tout $\sigma > 1$.

(c) Montrer que

$$\sum_{p} \frac{(\log p)^m}{p^{\sigma}} = \frac{1}{(m-1)!(\sigma-1)^{m+1}} + O_m(1)$$

pour tout $\sigma \in [1, 2]$.

EXERCICE 6.2. Soit $p_1 < p_2 < \cdots$ la suite des nombres premiers, et soit $P_k = p_1 p_2 \cdots p_k$.

(a) Pour tout $k \geqslant 2$, montrer que $p_k \asymp k \log k$ et que $\log P_k \asymp k \log k$.

(b) En supposant la vérité du théorème des nombres premiers, montrer que $p_k \sim k \log k$ et que $\log P_k \sim k \log k$ lorsque $k \to \infty$.

(c) Montrer que $\omega(n) \ll \log n / \log\log n$ pour tout $n \in \mathbb{Z}_{\geqslant 3}$. [*Indice* : que peut-on dire de la taille de $\omega(n)$ si on sait que $n \leqslant P_k$?]

(d) Montrer que

$$\frac{\phi(n)}{n} \sim \prod_{\substack{p \mid n \\ p \leqslant \log n}} \left(1 - \frac{1}{p}\right) \geqslant \prod_{p \leqslant \log n} \left(1 - \frac{1}{p}\right) \quad (n \to \infty).$$

EXERCICE 6.3. Pour tout $n \in \mathbb{N}$ et tout premier p, montrer que

$$v_p(n!) = \sum_{k=0}^{\infty} \left\lfloor n/p^k \right\rfloor.$$

EXERCICE 6.4 (Nair). Cet exercice fournit une preuve alternative de la borne inférieure au théorème 6.7. Considérer l'intégrale $I_n = \int_0^1 x^n (1-x)^n \mathrm{d}x$ et l'entier $N = \mathrm{ppcm}[n+1, n+2, \ldots, 2n]$. Montrer que :

(a) $I_n \cdot N$ est un entier non-négatif ;

(b) $I_n \leqslant 4^{-n}$;

(c) $N \leqslant (2n)^{\pi(2n)}$.

EXERCICE 6.5 (Mertens). Pour tout $x \geqslant 1$, montrer les estimations suivantes :

(a) $\sum_{n \leqslant x} \log n = x \log x - x + O(\log x)$;

(b) $\sum_{n \leqslant x} \log n = \sum_{a \leqslant x} \Lambda(a) \lfloor x/a \rfloor = x \sum_{a \leqslant x} \frac{\Lambda(a)}{a} + O\big(\sum_{a \leqslant x} \Lambda(a)\big)$;

(c) $\sum_{a \leqslant x} \Lambda(a) \ll x$;

(d) $\sum_{p \leqslant x} \frac{\log p}{p} = \log x + O(1)$.

Deuxième partie

Arithmétique modulaire

Chapitre 7

L'algèbre des résidus

7.1 La division euclidienne revisitée

Soit n un nombre naturel. Une propriété fondamentale de division euclidienne par n est que le reste une fonction n-périodique. En effet, soit a un entier et supposons que

$$a = qn + r \quad \text{avec} \quad 0 \leqslant r < n.$$

Évidemment, ceci veut dire aussi que

$$a + n = (q + 1)n + r.$$

Cette dernière relation implique notre affirmation que le nombre $a+n$ a le même reste que a quand divisé par n.

Grâce à la n-périodicité du reste, on peut penser à la division euclidienne par n comme une opération qui agit sur l'ensemble des progressions arithmétiques d'étape n, c'est-à-dire sur les ensembles de la forme $\{a + kn : k \in \mathbb{Z}\}$ avec $a \in \mathbb{Z}$.

Le lemme suivant clarifie et précise la discussion ci-dessus.

Lemme 7.1. *Soit $n \in \mathbb{N}$, et soient $a, b \in \mathbb{Z}$. Les propriétés suivantes sont équivalentes :*

(a) a et b ont le même reste quand divisé par n ;

(b) $n|(a - b)$;

(c) Il existe une progression arithmétique d'étape n qui contient a et b.

Démonstration. $(a) \Rightarrow (b)$: soit r le reste commun de a et de b quand on les divisent par n. Donc, il existe $k, \ell \in \mathbb{Z}$ tels que $a = kn + r$ et $b = \ell n + r$. On en déduit que $a - b = (k - \ell)n$, ce qui montre que $n|(a - b)$.

$(b) \Rightarrow (c)$: Soit P la progression arithmétique $\{a + kn : k \in \mathbb{Z}\}$. Puisque $n|(a - b)$, il existe $k \in \mathbb{Z}$ tel que $b - a = kn$, ce qui veut dire que $b = a + kn \in P$.

$(c) \Rightarrow (a)$: Soit $P = \{c + kn : k \in \mathbb{Z}\}$ une progression arithmétique qui contient a et b. Donc, il existe $k, \ell \in \mathbb{Z}$ tels que $a = kn + c$ et $b = \ell n + c$. Puis, supposons que les restes de la division euclidienne de a et de b par n sont égaux à r et à s, respectivement. Par la suite, $a = qn + r$ et

$b = q'n + s$ pour quelques entiers q et q'. Donc, on a à notre disponibilité deux façons à calculer $a - b$:

$$a - b = (k - \ell)n \quad \text{et} \quad a - b = (q - q')n + r - s.$$

En comparant ces deux relations, on voit que n divise $r - s$. Mais, on sait que $0 \leqslant r, s < n$, et par conséquent $-n < r - s < n$. Le seul multiple de n qui satisfait ces inégalités et le nombre zéro. Ceci montre que $r = s$, comme désire. La preuve est alors complète. □

Le lemme 7.1 nous amène naturellement à la définition suivante :

Définition 7.2 (Congruence mod n)**.** Soit n un nombre naturel. On dit que deux entiers a et b sont *congruents modulo* n (ou, plus simplement, *congruents mod* n) si $n|(a - b)$. Dans ce cas-ci, on écrit $a \equiv b \pmod{n}$.

D'après le lemme 7.1, la relation de congruence mod n est une relation d'équivalence sur $\mathbb{Z}$ dont les classes sont les ensembles $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_{n-1}$, où $\mathcal{C}_r$ est l'ensemble de tous les entiers dont leur reste quand divisés par n est égal à r, c'est-à-dire les classes d'équivalences sont les n différentes progressions arithmétiques d'étape n. [1]

Définition 7.3 (Résidus)**.** Soit n un nombre naturel.

(a) Les classes d'équivalence de la relation $a \equiv b \pmod{n}$ sont appelées *classes de congruence* mod n ou même *résidus* mod n. Lorsque le module n est clair d'après le contexte, on n'a pas besoin de le préciser et on dit simplement «classe de congruence» ou «résidu».

(b) On écrit $a \pmod{n}$ pour la classe d'équivalence de a mod n.

(c) On dénote l'ensemble de tous les résidus mod n par $\mathbb{Z}/n\mathbb{Z}$.

(d) Un système complet de représentants des classes d'équivalence mod n est appelé un *système complet de résidus* mod n.

De notre discussion avant la définition 7.3, on a tout de suite le lemme suivant :

Théorème 7.4. *Pour tout $n \in \mathbb{N}$, l'ensemble $\{0, 1, \ldots, n-1\}$ est un système complète de résidus mod n, c'est-à-dire*

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

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

À partir du système complet de résidus ci-dessus, on peut en construire d'autres :

Corollaire 7.5. *Soient $n \in \mathbb{N}$ et $a, b \in \mathbb{Z}$. Si $(a, n) = 1$, alors l'ensemble*

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

est un système complet de résidus.

1. On peut penser à une relation d'équivalence $\sim$ sur un ensemble X comme une partition de X dans certains sous-ensembles disjoints. Puis, on a que $x \sim y$ si et seulement si x et y appartiennent au même sous-ensemble de notre partition. Dans notre cas spécifique, le lemme implique 7.1 implique que la relation $a \equiv b \pmod{n}$ partition $\mathbb{Z}$ dans les ensembles $\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_{n-1}$. De façon alternative, on invite les lecteurs de vérifier directement que la relation $a \equiv b \pmod{n}$ satisfait les trois propriétés définissant une relation d'équivalence : réflexivité, symétrie et transitivité.

Démonstration. Puisque $|\mathbb{Z}/n\mathbb{Z}| = n$, il suffit de montrer que les n nombres $aj + b$ avec $j \in \{0, 1, \ldots, n-1\}$ ne sont pas congruents mod n. En effet, si $ai + b \equiv aj + b \pmod{n}$ avec $0 \leqslant i, j < n$, alors n divise $(ai + b) - (aj + b) = a(i - j)$. Puisque $(a, n) = 1$, le lemme 2.9 d'Euclide implique que $n|(i - j)$. Cependant, on a que $-n < i - j < n$. Le seul multiple de n satisfaisant ces inégalités est le nombre 0. Donc $i = j$, comme affirmé. □

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

Corollaire 7.6. *Soit $n \in \mathbb{N}$. Tout ensemble de n nombres entiers consécutifs est un système complet de résidus mod n. En particulier, l'ensemble $\mathbb{Z} \cap (-n/2, n/2]$ est un système complet de résidus mod n.*

Démonstration. La première partie découle du corollaire 7.5, puisque un ensemble de n nombres entiers consécutifs peut s'écrire dans la forme $\{j + b : 0 \leqslant j \leqslant n - 1\}$ 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 (on laisse la vérification de cette affirmation comme exercice). □

7.2 Addition et multiplication mod n

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

Théorème 7.7. *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. Puisque $a \equiv b \pmod{n}$ et $c \equiv d \pmod{n}$, il existe $k, \ell \in \mathbb{Z}$ tels que $a = b + kn$ et $c = d + \ell n$. Donc, $a + c = b + d + (k + \ell)n$ et $ac = bd + (kd + \ell b + k\ell n)n$, ce qui implique que $a + c \equiv b + d \pmod{n}$ et $ac \equiv bd \pmod{n}$, comme affirmé. □

Exemple. Voici une application pratique du théorème 7.7. Soient a, b et n trois nombres naturels avec a et b beaucoup plus grands que n (p. ex. $n = 10$ et a et b ont taille approximative $\approx 10^{10}$). Supposons qu'on a déjà exécute la division euclidienne de a et de b par q, et que nous avons trouvé que

$$a = kn + r \quad \text{et} \quad b = \ell n + s \quad \text{avec } 0 \leqslant r, s < n.$$

On veut maintenant calculer le reste de la division euclidienne de ab par n. Évidemment, on peut calculer le produit ab et après le diviser par n. Par contre, il existe une autre façon qui est beaucoup plus efficace : on a que $ab \equiv rs \pmod{n}$ d'après le théorème 7.7. Par la suite, le lemme 7.1 implique que ab et rs ont le même reste quand divisés par n. Donc il suffit de calculer le reste de rs qui est un nombre beaucoup plus petit que ab. ■

Le théorème 7.7 implique que la relation de congruence mod n est compatible avec les opérations algébriques d'addition et de multiplication. En particulier, il nous permettre de bien définir la somme et le produit de deux résidus $a \pmod{n}$ et $b \pmod{n}$ par

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

D'après le théorème 7.7, cette définition ne dépend pas du choix des représentants des classes de congruence $a \pmod n$ et $b \pmod n$.

Évidemment, les opérations d'addition et de multiplication mod n héritent plusieurs bonnes propriétés des opérations correspondantes de $\mathbb{Z}$: elles sont les deux commutatives et associatives, et la multiplication est distributive par rapport à l'addition. De plus, il existe un élément neutre de l'addition (le résidu $0 \pmod n$) et un élément neutre de la multiplication (le résidu $1 \pmod n$). Finalement, chaque résidu $a \pmod n$ possède un inverse additif, donné par le résidu $-a \pmod n$.

Remarque. La discussion du paragraphe précédent veut dire que l'ensemble $\mathbb{Z}/n\mathbb{Z}$ muni des opérations de l'addition et de la multiplication est un anneau commutatif et unitaire.

D'autre côté, il existe une différence fondamentale entre l'arithmétique de $\mathbb{Z}$ et celle de $\mathbb{Z}/n\mathbb{Z}$. On sait que les seuls nombres entiers qui possèdent un inverse multiplicatif dans $\mathbb{Z}$ sont les nombres -1 et 1. Ce n'est pas le cas pour $\mathbb{Z}/n\mathbb{Z}$, où plusieurs éléments possèdent d'inverses :

Théorème 7.8. *Soit $n \in \mathbb{N}$ et $a \in \mathbb{Z}$. Alors $(a, n) = 1$ si et seulement s'il existe $b \in \mathbb{Z}$ tel que $ab \equiv 1 \pmod n$.*

Démonstration. Supposons d'abord que $(a, n) = 1$. Il y a deux façons à montrer l'existence de b. La première utilise le corollaire 7.5, qui dit que l'ensemble $\{aj : 0 \leqslant j < n\}$ est un système complet de résidus. En particulier, il existe $b \in \{0, 1, \ldots, n-1\}$ tel que $1 \equiv ab \pmod n$.

Il existe aussi une façon plus directe qui fournit aussi une façon de construire n. Puisque $(a, n) = 1$, le lemme de Bezout (cf. théorème 2.6) implique l'existence de quelques entiers b et c tels que $ab + nc = 1$. Donc $ab \equiv 1 \pmod n$. (Afin de construire b pour un choix donné de a et n, voir l'algorithme euclidien dans la section 2.2.)

Vice versa, supposons qu'il existe $b \in \mathbb{Z}$ tel que $ab \equiv 1 \pmod n$. Donc, il y a un entier k tel que $ab = 1 + kn$. Si $d = (a, n)$, alors d divise la combinaison linéaire $b \cdot a - k \cdot n = 1$. On en déduit que $d = 1$ comme voulu. □

Remarque 7.9. (a) Si $(a, n) = 1$, alors $(a + kn, n) = 1$ pour tout $k \in \mathbb{Z}$, d'après le lemme 2.5. Donc, $(a', n) = 1$ pour tout $a' \equiv a \pmod n$.

(b) Si $(a, n) = 1$ et $b, b' \in \mathbb{Z}$ sont deux entiers tels que $ab \equiv ab' \equiv 1 \pmod n$, alors $b \equiv b' \pmod n$. En effet, puisque $ab \equiv ab \pmod n$, le nombre n divise $ab - ab' = a(b - b')$. D'après le lemme d'Euclide, ceci veut dire que $n | (b - b')$, et donc que $b \equiv b' \pmod n$.

Le théorème 7.8 et le remarque 7.9 nous permettent de faire la définition importante suivante :

Définition 7.10 (Inverses multiplicatives mod n)**.**

(a) Soit $(a, n) = 1$. On définit $\overline{a} \pmod n$ d'être la classe de congruence mod n telle que $a \cdot \overline{a} \equiv 1 \pmod n$. On appelle chaque représentant de la classe $\overline{a} \pmod n$ l'*inverse de a mod n*.

(b) Si $a \pmod n$ est tel que $(a, n) = 1$, alors on dit que $a \pmod n$ est un *résidu réduit*. On dénote l'ensemble de tous les résidus réduits par

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

Théorème 7.11. *On a que*

$$\left|(\mathbb{Z}/n\mathbb{Z})^*\right| = \phi(n).$$

Démonstration. D'après le corollaire 7.6, l'ensemble $\{1, 2, \ldots, n\}$ est un système complet de résidus. Puisque on a définit $\phi(n) = \#\{1 \leqslant a \leqslant n : (a, n) = 1\}$, le théorème en découle directement. □

7.3 Un corps fini

Quand un résidu est réduit, on peut diviser par lui. En effet, si $a \in \mathbb{Z}/n\mathbb{Z}$ et $b \in (\mathbb{Z}/n\mathbb{Z})^*$, alors on peut considérer le résidu $a\bar{b} \pmod n$, qui joue le rôle de la fraction a/b dans l'ensemble $\mathbb{Z}/n\mathbb{Z}$. Dans le cas spécial où $n = p$, un nombre premier, tous les résidus non-zéros mod p sont copremiers avec p et donc réduits. Ceci veut dire qu'on peut diviser par tous les résidus non-zéros et rester dans l'ensemble $\mathbb{Z}/p\mathbb{Z}$. Ceci est une propriété que $\mathbb{Z}/p\mathbb{Z}$ partage avec les ensembles $\mathbb{Q}$, $\mathbb{R}$ et $\mathbb{C}$, où on peut aussi inverser tous les éléments non-zéro. Comme on le verra plus bas, cette propriété fondamentale implique que plusieurs résultats concernant $\mathbb{R}$ ont des analogies dans $\mathbb{Z}/p\mathbb{Z}$.

Remarque. Un anneau commutatif et unitaire où tout élément non-zéro est inversible par rapport à la multiplication est appelé un *corps*. Donc, $\mathbb{R}$ et $\mathbb{Z}/p\mathbb{Z}$ sont de corps. Le premier est un corps infini et le deuxième fini.

D'abord, on examinera la résolution d'équations polynomiales $\mathbb{Z}/p\mathbb{Z}$. On commence avec un résultat simple qui illustre l'analogie entre $\mathbb{R}$ et $\mathbb{Z}/p\mathbb{Z}$. (Voir aussi l'exercice 7.3 pou le cas où le modulus est composé.)

Lemme 7.12. *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 3.1(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 théorème 7.7 implique que $a^2 \equiv (\pm b)^2 \equiv b^2 \pmod p$. □

On examine maintenant des équations polynomiales plus générales.

Lemme 7.13. *Soit n un entier et soit $f(x) \in \mathbb{Z}[x]$. Si $a \equiv b \pmod n$, alors $f(a) \equiv f(b) \pmod n$.*

Démonstration. Soit $f(x) = c_d x^d + \cdots + c_1 x + c_0$ avec $c_0, c_1, \ldots, c_d \in \mathbb{Z}$. En utilisant le théorème 7.7 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 $j = 0, 1, 2, \ldots$ (le cas $j = 0$ étant trivial). Donc, on a que $c_j a^j \equiv c_j b^j \pmod n$ pour tout $j \in \{0, 1, \ldots, d\}$ d'après le théorème 7.7. Finalement, une troisième application de ce théorème implique (de façon inductive) que

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

ce qui est ce qu'il fallait démontrer. □

Le lemme précédent 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, l'ensemble

$$\mathcal{R}_f(n) := \{x \pmod n : f(x) \equiv 0 \pmod n\} \tag{7.1}$$

est bien défini grâce au lemme 7.13. Montrons maintenant que $n = p$, alors $|\mathcal{R}_f(p)| \leqslant \deg(f)$, exactement comme dans le cas de polynômes sur $\mathbb{R}$. En fait, nous avons le résultat suivant plus précis :

Théorème 7.14. *Soit p un nombre premier et soit $f(x) = c_0 + c_1x + \cdots + c_dx^d$ un polynôme dont les coefficients c_j sont entiers. On suppose que $p \nmid c_d$ et que*

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

avec $r \in \mathbb{Z}_{\geqslant 0}$. Alors, il existe qulques nombres naturels $m_1, \ldots, m_r$ et un polynôme $g(x)$ à coefficients entiers tels que :

(a) $m_1 + \cdots + m_r \leqslant d$;
(b) $\deg(g) = d - (m_1 + \cdots + m_r)$;
(c) $g(x) \not\equiv 0 \pmod p$ *pour tout* $x \in \mathbb{Z}$;
(d) 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, $|\mathcal{R}_f(p)| = r \leqslant \sum_{j=1}^{r} m_j \leqslant d$.

Démonstration. On utilise induction sur d. Si $d = 0$, nécessairement $f(x) = c_0 \not\equiv 0 \pmod p$ pour chaque $x \in \mathbb{Z}$ de notre hypothèse que $p \nmid c_d = c_0$. 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 $< d$ dont le coefficient en tête n'est pas divisible par p. Si l'ensemble $\mathcal{R}_f(p)$ est vide (auquel cas $r = 0$), on prend $g(x) = f(x)$ pour montrer le théorème. Sinon, on considère le résidu $\alpha_1 \pmod p$, qui est une racine de $f(x)$ mod p. On veut montrer que $f(x)$ mod p a comme facteur le polynôme $x - \alpha_1$. Pour le faire, on observe que $x - \alpha_1$ est un facteur de $f(x) - f(\alpha_1)$: en effet,

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

où

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

De plus, puisque $f(\alpha_1) \equiv 0 \pmod p$, on voit que les polynômes $f(x)$ et $(x - \alpha_1)\tilde{f}(x)$ diffèrent par un multiple de p, donc ils ont les mêmes coefficients mod p. En particulier,

$$f(x) \equiv (x - \alpha_1)\tilde{f}(x) \pmod p \quad \text{pour tout } x \in \mathbb{Z},$$

d'où on trouve que les racines de $\tilde{f}$ mod p sont aussi de racines de f mod p. Plus précisément, nous avons que

$$\mathcal{R}_{\tilde{f}}(p) = \begin{cases} \mathcal{R}_f(p) & \text{si } \tilde{f}(\alpha_1) \not\equiv 0 \,(\mathrm{mod}\, p), \\ \mathcal{R}_f(p) \setminus \{\alpha_1 \,(\mathrm{mod}\, p)\} & \text{sinon.} \end{cases}$$

Dans tous les cas, l'hypothèse inductive implique qu'il existe d'entiers $m_1' \geqslant 0$ et $m_2, \ldots, m_r \geqslant 1$ et un polynôme $g(x)$ à coefficients entiers tels que :

(a) $m_1' + \sum_{i=2}^{r} m_i \leqslant d-1$;
(b) $\deg(\tilde{g}) \leqslant d - 1 - m_1' - \sum_{i=2}^{r} m_i$;
(c) $g(x) \not\equiv 0 \,(\mathrm{mod}\, p)$ pour tout $x \in \mathbb{Z}$;
(d) $\tilde{f}(x)$ et $(x-\alpha_1)^{m_1'}(x-\alpha_2)^{m_2} \cdots (x-\alpha_r)^{m_r} g(x)$.

De plus, on a que $m_1' \geqslant 1$ si et seulement si $\tilde{f}(\alpha_1) \equiv 0 \,(\mathrm{mod}\, p)$. Par conséquent, si on pose $m_1 = m_1' + 1 \geqslant 1$, on trouve que polynômes $f(x)$ et $g(x)$ et les nombres naturels $m_1, \ldots, m_r$ satisfont les propriétés (a)–(d) de l'énoncé du théorème. Ceci conclut l'étape inductive et, par la suite, la démonstration. □

On conclut cette section avec un résultat joli grâce à Wilson qui utilise le fait que les propriétés fondamentales de $\mathbb{Z}/p\mathbb{Z}$: qu'il est un corps et qu'il est un ensemble fini.

Théorème 7.15 (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 $p \geqslant 5$. Pour chaque $j \in \{1, 2, \ldots, p-1\}$, on écrit $\overline{j}$ pour signifier, avec un petit abus de notation, l'élément unique de $\{1, \ldots, p-1\}$ pour lequel $j \cdot \overline{j} \equiv 1 \,(\mathrm{mod}\, p)$. On observe que j et $\overline{j}$ sont parmi les facteurs du produit $(p-1)! = 1 \cdot 2 \cdots (p-1)$. Si $j \neq \overline{j}$, ces deux facteurs s'annulent mod p. Donc, on trouve tout de suite que

$$(p-1)! \equiv \prod_{\substack{1 \leqslant j \leqslant p-1 \\ j = \overline{j}}} j.$$

Il reste à trouver tous les nombres $j \in \{1, \ldots, p-1\}$ avec $\overline{j} = j$. Ceci est équivalent au fait que $j^2 \equiv 1 \,(\mathrm{mod}\, p)$. D'après le lemme 7.12, les seuls solutions à cette équation sont $1 \,(\mathrm{mod}\, p)$ et $-1 \,(\mathrm{mod}\, p)$, qui correspondent aux nombres $j = 1$ et $j = p-1$. On en déduit que

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

ce qui est ce qu'il fallait démontrer. □

Remarque 7.16. L'idée de s'accoupler les éléments de $\{1, \ldots, p-1\}$ de façon appropriée afin de 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 7.17. *Si $n > 4$ est composé, alors*

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

Démonstration. Soit $n = p_1^{\nu_1} \cdots p_k^{\nu_r}$ la factorisation première de n.

Si $k \geqslant 2$, alors $p_j^{\nu_j} < n$ pour chaque j et, par la suite, $p_j^{\nu_j} | (n-1)!$ pour chaque j. On conclut alors que $n|(n-1)!$ dans ce cas-ci.

Supposons, maintenant, que $k = 1$, c'est-à-dire $n = p^\nu$ pour un nombre premier p et un exposant $\nu \geqslant 2$ (car n est composé). Les nombres $p, p^2, \ldots, p^{\nu-1}$ sont tous $< n$, donc leur produit divise $(n-1)!$. On voit aussi que

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

L'exposant est $\geqslant 1 + (\nu - 1) = \nu =$ quand $\nu \geqslant 3$, ce qui conclut la preuve dans ce cas-ci.

Il reste à considérer le cas où $\nu = 2$, c'est-à-dire le cas où $n = p^2$. Puisque on a supposé que $n > 4$, il faut que $p > 2$. Il faut alors montrer que $p^2 | 1 \cdot 2 \cdots (p^2 - 1)$. On sait déjà que p se trouve parmi les nombres $1, 2, \ldots, p^2 - 1$, donc on cherche un autre multiple de p dans cette liste. On observe que $2p < p^2$ quand $p > 2$, et donc $p \cdot 2p$ divise $(p^2 - 1)!$. Ceci termine la démonstration. □

En combinant les théorèmes 7.15 et 7.17, on déduit le résultat suivant qu'on peut voir comme un critère de primalité. Par contre, il faut remarquer que ce critère n'est pas très utile en pratique car il exige trop d'opérations pour calculer le reste de $(n-1)!$ quand divisé par n.

Corollaire 7.18. *Un nombre $n > 1$ est premier si et seulement si*

$$(n-1)! \equiv -1 \pmod{n}.$$

Démonstration. Si n est premier, on a que $(n-1)! \equiv -1 \pmod{n}$, du théorème de Wilson. Si $n = 4$, alors on a que $(4-1)! \equiv 2 \not\equiv -1 \pmod{4}$. Finalement, si n est composé et > 4, alors $(n-1)! \equiv 0 \not\equiv -1 \pmod{n}$ d'après le théorème 7.17. □

7.4 Exercices

EXERCICE 7.1. Soient $n \in \mathbb{N}$ et $a, x, y \in \mathbb{Z}$. Montrer que $ax \equiv ay \pmod{n}$ si et seulement si $x \equiv y \pmod{n/(a,n)}$.

EXERCICE 7.2. Soient $n_1, \ldots, n_k \in \mathbb{N}$ et $x, y \in \mathbb{Z}$. Montrer que

$$x \equiv y \pmod{n_j} \quad \text{pour tout } j = 1, 2, \ldots, k \qquad \Longleftrightarrow \qquad x \equiv y \pmod{[n_1, \ldots, n_k]}.$$

EXERCICE 7.3. 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 7.4. Soient $n \in \mathbb{N}$ et $a \in \mathbb{Z}$ tels que $(a, n) = 1$. Soit l'application $T : \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$, définie par $T(x \pmod{p}) := ax \pmod{p}$. Montrer que :

(a) T est bijective ;

(b) $T((\mathbb{Z}/n\mathbb{Z})^*) = (\mathbb{Z}/n\mathbb{Z})^*$.

EXERCICE 7.5. (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 $\alpha, \beta \geqslant 2$? [*Indice :* Montrez que α doit être impair et factorisez $p^\alpha + 1$.]

EXERCICE 7.6. 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 7.7. Soient p un premier, $f(x) \in \mathbb{Z}[x]$ et $x_0 \in \mathbb{Z}$.

(a) Montrer 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 :* Considérer la fonction $f(x + x_0)$ et écrire-la comme un polynôme de x.]

(b) Montrer que $f(x_0) \equiv 0 \pmod{p}$ si et seulement si $a_0 \equiv 0 \pmod{p}$.

(c) Montrer 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 7.8. 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 7.9. Soit $p > 2$ un nombre premier et soit $p_0 = (p-1)/2$. Montrer que

$$p_0!^2 \equiv (-1)^{p_0+1} \pmod{p}.$$

[*Indice :* $(\mathbb{Z}/p\mathbb{Z})^* = \{a \pmod{p} : 1 \leqslant a \leqslant p_0\} \sqcup \{-a \pmod{p} : 1 \leqslant a \leqslant p_0\}$.]

Chapitre 8

L'ordre multiplicatif mod n

8.1 Critères de divisibilité

À l'école, on apprend quelques critères simples pour déterminer si un nombre est divisible par $5, 10, 3$ ou 9. Comme on le verra dans cette section, tous ces critères sont une conséquence simple de l'arithmétique modulaire et ils se généralisent facilement pour obtenir de critères de divisibilité par n'importe quel nombre. De plus, cette discussion va nous aider naturellement à une notion abstraite très importante : l'ordre multiplicatif d'un entier mod n.

Soit $n = (a_k a_{k-1} \cdots a_1 a_0)_{10}$ le développement décimal du nombre n, c'est-à-dire

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

et les coefficients $a_0, a_1, \ldots, a_k$ se trouvent dans $\{0, 1, \ldots, 9\}$. Soit aussi un nombre naturel d. On veut tester si $d|n$. La réalisation-clé est que ceci est équivalent à la relation $n \equiv 0 \pmod{d}$. D'autre côté, le théorème 7.7 nous permet de déterminer la classe de congruence de $n \pmod{d}$ si on connait la classe de congruence de toutes les puissances de 10 modulo d. On examine quelques cas particuliers pour voir comment cela fonction en pratique.

Exemples. (a) $d \in \{2, 5, 10\}$. Ceux sont les cas les plus simples : nous avons que $10^j \equiv 0 \pmod{d}$ pour tout $j \geqslant 1$. Par la suite, $n \equiv a_0 \pmod{d}$, ce qui implique le critère de divisibilité familier suivant :

$$d|n \quad \Longleftrightarrow \quad d|a_0.$$

(b) $d \in \{3, 9\}$. Dans ces cas, on observe que $10 \equiv 1 \pmod{d}$ et donc $10^j \equiv 1 \pmod{d}$ pour tout $j \geqslant 1$ par un argument inductif. Par la suite, $n \equiv a_0 + a_1 + \cdots + a_k \pmod{d}$, ce qui implique le critère de divisibilité familier suivant :

$$d|n \quad \Longleftrightarrow \quad d|(a_0 + a_1 + \cdots + a_k).$$

En pratique, il faut souvent itérer ce critère plusieurs fois afin de déterminer la divisibilité d'un nombre par 9. Par exemple, on a que

$$\begin{aligned} 9|87106361528299 \quad &\Longleftrightarrow \quad 9|8+7+1+0+6+3+6+1+5+2+8+2+9+9 = 67 \\ &\Longleftrightarrow \quad 9|6+7 = 15. \end{aligned}$$

La dernière relation est évidemment fausse, donc la première l'est aussi.

(c) $d = 7$. On essai maintenant de deviner un critère de divisibilité par 7. On commence en calculant le reste des puissances de 10 mod 7. Ce calcul peut être effectué de façon itérative, comme suit :

$$\begin{aligned}
10^0 &= 1 \equiv 1 \pmod 7\\
10^1 &= 10 \equiv 3 \pmod 7\\
10^2 &\equiv 3^2 \equiv 9 \equiv 2 \pmod 7\\
10^3 &= 10 \cdot 10^2 \equiv 3 \cdot 2 \equiv 6 \equiv -1 \pmod 7\\
10^4 &= 10 \cdot 10^3 \equiv 3 \cdot (-1) \equiv -3 \pmod 7\\
10^5 &= 10 \cdot 10^4 \equiv 3 \cdot (-3) \equiv -9 \equiv -2 \pmod 7\\
10^6 &= 10 \cdot 10^5 \equiv 3 \cdot (-2) \equiv -6 \equiv 1 \pmod 7.
\end{aligned}$$

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

$$\begin{aligned}
n \equiv\ & (a_0 + a_6 + a_{12} + \cdots) + 3(a_1 + a_7 + a_{13} + \cdots) + 2(a_2 + a_8 + a_{14} + \cdots)\\
& - (a_3 + a_9 + a_{15} + \cdots) - 3(a_4 + a_{10} + a_{16} + \cdots) - 2(a_5 + a_{11} + a_{17} + \cdots) \pmod 7,
\end{aligned}$$

ce qui implique que

$$7|n \quad \Longleftrightarrow \quad 7|\sum_{j\geqslant 0} a_{6j} + 3\sum_{j\geqslant 0} a_{6j+1} + 2\sum_{j\geqslant 0} a_{6j+2} - \sum_{j\geqslant 0} a_{6j+3} - 3\sum_{j\geqslant 0} a_{6j+4} - 2\sum_{j\geqslant 0} a_{6j+5}.$$

On conclut la discussion de cet exemple avec une application : on a que

$$\begin{aligned}
9|87106361528299 \quad &\Longleftrightarrow \quad 9|(7+1+9) + 3(8+6+9) + 2(3+2)\\
&\qquad\qquad - (6+8) - 3(0+2) - 2(1+5) = 64\\
&\Longleftrightarrow \quad 9|4 + 3\cdot 6 = 22\\
&\Longleftrightarrow \quad 9|2 + 3\cdot 2 = 8.
\end{aligned}$$

Puisque le dernier nombre n'est pas divisible par 7, on a aussi que $7 \nmid 87106361528299$. ■

Il est possible de généraliser la procédure décrite aux exemples ci-dessus afin d'obtenir un critère de divisibilité par n'importe quel entier d et dans n'importe quelle base.

Soient deux entiers $b, d \geqslant 2$. On veut trouver un critère général qui, étant donné un entier n dans son développement b-adique, détermine si $d|n$. À cette fin, supposons que $n = (c_r c_{r-1} \cdots c_0)_b$ pour quelques nombres $c_0, c_1, \ldots, c_r \in \{0, 1, \ldots, b-1\}$ ou, de façon équivalente, que

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

Le corollaire 7.6 implique que, pour chaque $j \geqslant 0$, 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 grandeur $\approx b^r$, à la divisibilité de $c_0\beta_0 + c_1\beta_1 + \cdots + c_r\beta_r$, un nombre de grandeur $\leqslant (r+1)(b-1)d/2$, qui est considérablement plus petit que b^r (à condition que d ne soit pas énorme).

L'étape cruciale pour déterminer le critère correcte est le calcul de nombres β_j. À première vue, ce calcul prend r étapes : en supposant qu'on a calcul β_{j-1}, on a que $\beta_j \equiv b\beta_{j-1} \pmod d$. Puisque on sait déjà la valeur de β_{j-1}, qui se trouvent dands $(-d/2, d/2]$, le nombre $b\beta_{j-1}$ a valeur absolue $\leqslant (b-1)d/2$, donc sa réduction mod d se fait très rapidement.

Cependant, on affirme que le calcul de nombres β_j prend $\leqslant d$ opérations. En effet, puisque il existe seulement d distinctes classes de congruence mod d, le principe du pigeonnier garantit l'existence d'au moins deux nombres parmi les $d+1$ nombres $b^0, b^1, \ldots, b^d$ qui tombent dans la même classe de congruence mod d. En notation, ceci veut dire qu'il existe deux entiers i_0 et k tels que $0 \leqslant i_0 < i_0 + k \leqslant d$ et $b^{i_0} \equiv b^{i_0+k} \pmod d$. Ceci implique que la suite $\{\beta_i\}_{i\geqslant 0}$ est k-périodique à partir de i_0 : en effet, si $i \geqslant i_0$, alors on a que

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

Puisque $-d/2 < \beta_{i+k}, \beta_i \leqslant d/2$, on déduit que $\beta_{i+k} = \beta_i$ pour tout $i \geqslant i_0$. En conclusion, il suffit de calculer les nombres $\beta_0, \beta_1, \ldots, \beta_{i_0+k-1}$; les autres sont déterminés par périodicité. Puisque $i_0 + k \leqslant d$, ceci montre notre affirmation que le calcul des nombres β_j prend $\leqslant d$ opérations.

Remarque. Sans perte de généralité, on peut supposer que les nombres i_0 et k sont choisis de façon minimale : on peut supposer que i_0 est le plus petit entier $i \geqslant 0$ pour lequel il existe un autre entier $\ell \geqslant 1$ tel que $b^i \equiv b^{i+\ell} \pmod d$. Après avoir choisi i_0, on sélectionne k comme le plus petit nombre naturel tel que $b^{i_0+k} \equiv b^{i_0} \pmod d$.

Notons que Si $(b, d) = 1$, alors la relation $b^{i_0} \equiv b^{i_0+k} \pmod d$ et le fait que b est inversible mod d impliquent que

$$b^k \equiv \overline{b}^{i_0}b^{i_0+k} \equiv \overline{b}^{i_0}b^{i_0} \equiv 1 \pmod d.$$

Le plut petit nombre naturel k qui satisfait l'idéntité $b^k \equiv 1 \pmod d$ est appelé l'*ordre multiplicatif* de b mod d. Il s'agit d'une notion fondamentale qu'on étudiera profondément pour le reste du chapitre.

On conclut cette section avec un dernier exemple d'un critère de divisibilité.

Exemple. 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

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

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

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

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 $n = (10000111100010011 1000)_2$, 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 $6 \nmid n$. ■

8.2 Généralités

On étudie maintenant le concept de l'ordre multiplicatif mod n qu'on a vu brièvement à la dernière section. Voici la définition formelle.

Définition 8.1 (Ordre multiplicatif mod n)**.** Soient $n \in \mathbb{N}$ et $a \in \mathbb{Z}$ avec $(a, n) = 1$. On définit l'*ordre multiplicatif* de a mod n d'être le nombre

$$\mathrm{ord}_n(a) := \min\{k \in \mathbb{N} : a^k \equiv 1 \,(\mathrm{mod}\, n)\}.$$

Remarques. (a) Évidemment, tous les entiers dans le même classe de congruence réduite ont le même ordre multiplicatif, donc on peut aussi parler de l'ordre multiplicatif d'un réside $a \,(\mathrm{mod}\, n)$, qu'on définit d'être $\mathrm{ord}_n(a)$.

(b) On se rappelle que les arguments de la section 8.1 montrent que $\mathrm{ord}_n(a)$ est toujours bien défini et $\leqslant n$. En effet, le principe du pigeonnier implique qu'il existe au moins deux nombres parmi les $n+1$ nombres $a^0, a^1, \ldots, a^n$ qui tombent dans la même classe d'équivalence mod n, soit $a^i \equiv a^j \,(\mathrm{mod}\, n)$ avec $0 \leqslant i < j \leqslant n$. Puisque $(a, n) = 1$, le résidu $a \,(\mathrm{mod}\, n)$ est inversible, d'où on déduit que $a^{j-i} \equiv 1 \,(\mathrm{mod}\, n)$. Puisque $1 \leqslant j - i \leqslant n$, ceci montre notre affirmation. On verra plus tard le résultat plus fort que $\mathrm{ord}_n(a)$ est toujours un diviseur de $\phi(n)$.

Notre premier résultat dans la théorie des ordres multiplicatifs est une preuve que l'ensemble d'entiers k avec $a^k \equiv 1 \,(\mathrm{mod}\, n)$ a une structure très simple :

Lemme 8.2. *Soit n un nombre naturel, soit $a \in \mathbb{Z}$ avec $(a, n) = 1$, et soit $k \in \mathbb{Z}$. Alors,*

$$a^k \equiv 1 \,(\mathrm{mod}\, n) \quad \Longleftrightarrow \quad \mathrm{ord}_n(a) | k.$$

Démonstration. Soit $m = \mathrm{ord}_n(a)$. Si $k = \ell m$ avec $\ell \in \mathbb{Z}$, on a que $a^k \equiv (a^m)^\ell \equiv 1 \,(\mathrm{mod}\, n)$.

Vice versa, supposons que $a^k \equiv 1 \,(\mathrm{mod}\, n)$. Par la division euclidienne (cf. théorème 2.3), on a que $k = qm + r$ avec $q \in \mathbb{Z}$ et $0 \leqslant r < m$. Puisque $a^k \equiv a^m \equiv 1 \,(\mathrm{mod}\, n)$ et a est inversible mod n, on a aussi que $a^r = a^k (a^m)^{-q} \equiv 1 \,(\mathrm{mod}\, n)$. Mais $m = \min\{j \in \mathbb{N} : a^j \equiv 1 \,(\mathrm{mod}\, n)\}$ et $0 \leqslant r < m$. Il faut donc que $r = 0$, ce qui conclut la démonstration. □

Maintenant, on passe à la preuve du théorème central suivant :

Théorème 8.3 (le théorème d'Euler)**.** *Soient $n \in \mathbb{N}$ et $a \in \mathbb{Z}$ avec $(a, n) = 1$. Alors,*

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

En particulier, $\mathrm{ord}_n(a) | \phi(n)$.

Démonstration. D'après le corollaire 7.5, on a que

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

On trouve alors que

$$\prod_{\substack{1\leqslant k\leqslant n\\(k,n)=1}} k \equiv \prod_{\substack{1\leqslant k\leqslant n\\(k,n)=1}} (ak) \pmod n \tag{8.1}$$

Tous les facteurs k avec $(k,n) = 1$ sont inversibles mod n. On peut donc multiplier les deux côtés de (8.1) avec $\overline{k} \pmod n$ pour tout $k \pmod n \in (\mathbb{Z}/n\mathbb{Z})^*$. Donc, on arrive à l'identité

$$\prod_{\substack{1\leqslant k\leqslant n\\(k,n)=1}} 1 \equiv \prod_{\substack{1\leqslant k\leqslant n\\(k,n)=1}} a \pmod n,$$

ce qui montre que $a^{\phi(n)} \equiv 1 \pmod n$.

Finalement, on applique le lemme 8.2 pour déduire que $\mathrm{ord}_n(a)|\phi(n)$. □

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

Corollaire 8.4 (le petit théorème de Fermat)**.** *Soient p un nombre premier et $a \in \mathbb{Z}$ tels que. Alors,*

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

et donc $\mathrm{ord}_p(a)$ est un diviseur de $p-1$.

8.3 Racines primitives

Le théorème d'Euler nous amène naturellement à la question suivante : si n est un nombre naturel donné, est-ce qu'il existe un résidu réduit $a \pmod n$ tel que $\mathrm{ord}_n(a) = \phi(n)$? Évidemment, un tel élément serait d'ordre maximal. Quand il existe, on l'appelle une *racine primitive* :

Définition 8.5 (Racine primitive)**.** Soit $n \in \mathbb{N}$ et soit $a \in \mathbb{Z}$ tel que $(a,n) = 1$. Si $\mathrm{ord}_n(a) = \phi(n)$, alors on dit que a est une *racine primitive* mod n.

Quand il existe de racines primitives mod n, l'ensemble $(\mathbb{Z}/n\mathbb{Z})^*$ admet une description très simple d'un point de vue algébrique :

Théorème 8.6. *Soit $n \in \mathbb{N}$ tel qu'il existe une racine primitive a mod n. Alors,*

$$(\mathbb{Z}/n\mathbb{Z})^* = \{1 \pmod n, a \pmod n, a^2 \pmod n, \ldots, a^{\phi(n)-1} \pmod n\}.$$

Démonstration. Il suffit de montrer que les $\phi(n)$ nombres $1, a, a^2, \ldots, a^{\phi(n)-1}$ sont tous distincts mod n. En effet, supposons au contraire que $a^i \equiv a^j \pmod n$ avec $0 \leqslant i < j < \phi(n)$. Puisque $(a,n) = 1$, il est inversible mod n, on trouve que $a^{j-i} \equiv 1 \pmod n$. Cependant, on a que $1 \leqslant j - i < \phi(n) = \mathrm{ord}_n(a)$, ce qui contredit la définition de l'ordre de a. On est arrivé à une contradiction. Donc, il faut que les nombres $1, a, a^2, \ldots, a^{\phi(n)-1}$ sont tous distincts mod n, comme on l'a affirmé. Ceci termine la preuve du théorème. □

Une fonction qui joue un rôle central dans la théorie des racines primitives est la *fonction de Carmichael* λ, définie par

$$\lambda(n) := \min\{k \in \mathbb{N} : a^k \equiv 1 \pmod{n} \text{ pour tout entier } a \text{ avec } (a,n) = 1\}. \tag{8.2}$$

Le théorème d'Euler implique qu'elle bien définie et qu'elle satisfait l'inégalité

$$\lambda(n) \leqslant \phi(n). \tag{8.3}$$

Le résultat suivant qui donne une expression alternative pour $\lambda(n)$ qui la connecte directement avec l'existence de racines primitives (cf. corollaire 8.8).

Théorème 8.7. *Pour tout nombre naturel n, on a que*

$$\lambda(n) = \max\{\mathrm{ord}_n(a) : 1 \leqslant a \leqslant n,\ (a,n) = 1\}.$$

En particulier, $\lambda(n)$ divise $\phi(n)$.

Avant montrer ce théorème fondamental, on record deux corollaires. Le premier est tout simplement une observation, mais le deuxième s'agit d'un résultat central de l'arithmétique modulaire.

Corollaire 8.8. *Soit $n \in \mathbb{N}$. Il existe de racines primitives mod n si et seulement si $\lambda(n) = \phi(n)$.*

Corollaire 8.9. *Si p est un nombre premier, alors il existe de racines primitives mod p.*

Démonstration. Soit $k = \lambda(p)$. Par la définition de la fonction de Carmichael, on a que $x^k \equiv 1 \pmod{p}$ quand $p \nmid x$. On voit donc que le polynôme $x^k - 1$ a $\geqslant p-1$ racines mod p (car tous les résidus réduits sont de racines). Le théorème 7.14 implique que $k = \deg(f) \geqslant p-1$. D'autre côté, on a que $k = \lambda(p) \leqslant \phi(p) = p-1$ d'après (8.3). On en déduit que $\lambda(p) = \phi(p)$. Donc, le corollaire 8.7 implique qu'il existe de racines primitives mod p. □

On reviendra au sujet des racines primitives pour d'autres moduli à la section 10.3.

On conclut cette section en montrant le théorème 8.7. On commence avec un lemme préparatoire.

Lemme 8.10. *Soient $n \in \mathbb{N}$ et $a, b \in \mathbb{Z}$ tels que $(a,n) = (b,n) = 1$.*

(a) On a que $\mathrm{ord}_n(ab)$ divise $\mathrm{ppcm}[\mathrm{ord}_n(a), \mathrm{ord}_n(b)]$.
(b) Si $\mathrm{pgcd}(\mathrm{ord}_n(a), \mathrm{ord}_n(b)) = 1$, alors $\mathrm{ord}_n(ab) = \mathrm{ord}_n(a) \cdot \mathrm{ord}_n(b)$.
(c) Si d est un diviseur de $\mathrm{ord}_n(a)$, alors $\mathrm{ord}_n(a^d) = \mathrm{ord}_n(a)/d$.

Démonstration. Soit $k = \mathrm{ord}_n(a)$, $\ell = \mathrm{ord}_n(b)$ et $m = \mathrm{ord}_n(ab)$.

(a) Puisque $k | [k,\ell]$, le lemme 8.2 implique $a^{[k,\ell]} \equiv 1 \pmod{n}$. Idem, on a que $b^{[k,\ell]} \equiv 1 \pmod{n}$ car $\ell | [k,\ell]$. On en déduit que $(ab)^{[k,\ell]} = a^{[k,\ell]} b^{[k,\ell]} \equiv 1 \pmod{n}$. En appliquant le lemme 8.2 encore une fois, on déduit alors que $m | [k,\ell]$, comme souhaité.

(b) D'après la partie (a) et de notre hypothèse que $(k,\ell) = 1$, on a que $m | [k,\ell] = k\ell$ (cf. théorème 2.13). Vice versa, montrons que $k\ell | m$. On a que $(ab)^m \equiv 1 \pmod{n}$, et donc $(ab)^{mk} \equiv$

$1 \pmod n$. Puisque $a^k \equiv 1 \pmod n$, on en déduit que $b^{mk} \equiv 1 \pmod n$. Le lemme 8.2 implique alors que $\ell | mk$. Puisque $(k, \ell) = 1$, le lemme 2.9 d'Euclide implique que $\ell | m$. En inversant les rôles de a et de b dans l'argument précédent, on voit aussi que $\ell | m$. Finalement, puisque $(k, \ell) = 1$, les relations $k | m$ et $\ell | m$ impliquent que $k\ell | m$, comme affirmé.

(c) Évidemment, on a que $(a^d)^{k/d} \equiv 1 \pmod n$. De plus, k/d est le plus petit nombre ayant cette propriété par la minimalité de k. Donc $k/d = \mathrm{ord}_n(a)$, comme affirmé. □

Démonstration du théorème 8.7. Soit $k = \max\{\mathrm{ord}_n(a) : 1 \leqslant a \leqslant n,\ (a, n) = 1\}$. Il existe donc $a \in \{1, \ldots, n\}$ tel que $(a, n) = 1$ et $\mathrm{ord}_n(a) = k$. D'après la définition de $\lambda(n)$, on a que $a^{\lambda(n)} \equiv 1 \pmod n$. Donc, il faut que $k | \lambda(n)$ par le lemme 8.2. Pour terminer la preuve, on montre aussi que $k \geqslant \lambda(n)$. Il suffit de prouver que $b^k \equiv 1 \pmod n$ pour tout $b \in \mathbb{Z}$ avec $(b, n) = 1$ ou, de façon équivalente, que $\mathrm{ord}_n(b) | k$ pour chaque tel b. Supposons, au contraire, qu'il existe $b \in \mathbb{Z}$ tel que $(b, n) = 1$ mais $\mathrm{ord}_n(b) \nmid k$. Si on pose $\ell = \mathrm{ord}_n(b)$, on trouve alors 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 $(\mathbb{Z}/n\mathbb{Z})^*$ 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 8.10 : la partie (c) de ce lemme implique que $\mathrm{ord}_n(b^{\ell_1}) = p^w$ et $\mathrm{ord}_n(a^{p^v}) = k_1$. Puisque $p \nmid k_1$, lemme 8.10 (b) implique que $\mathrm{ord}_n(b^{\ell_1} a^{p^v}) = p^w k_1 > k$. Ceci est impossible par le choix du k. On en déduit que $\ell | k$, comme on le voulait. Ceci termine la démonstration. □

8.4 L'algorithme RSA

La théorie des nombres a plusieurs applications pratiques, particulièrement à la cryptographie. La cryptographie existe depuis longtemps, 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 César. L'algorithme est facile : il se base sur une permutation des lettres de l'alphabet. Toutefois, c'est facile de déchiffrer tels algorithmes, spécialement si le message est assez long : on peut utiliser des informations statistiques connues concernant la fréquence des lettres et de mots communs (articles, prépositions, les verbes «avoir» et «être» et ses conjugaisons, etc.) et reconstruire la permutation utilisée pour le cryptage du message.

Avec le temps, les méthodes de cryptage ont évolué 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* pour l'expéditeur du message, *Bob* pour le destinataire et *Ève* pour l'adversaire, correspondant à «personne A», «personne B» et «Evil Eve» (en anglais).

On peut aussi formaliser les procédures de cryptage et de décryptage. Soit X l'ensemble de tous les messages possibles de k lettres. Le cryptage de ces messages est une bijection $f : X \to X$ et leur décryptage est la fonction réciproque $f^{-1} : X \to X$. Donc, si Alice veut transmettre

un message x de N lettres avec $N \geqslant k$, on le divise en paquets de $\leqslant k$ lettres chacun, soit $x = x_1x_2\cdots x_r$, et on envoie au lieu de x le message crypté $x' := f(x_1)f(x_2)\cdots f(x_r)$. Bob possède une clé qui lui permettre de calculer rapidement $f^{-1}(f(x_i)) = x_i$ pour tout i. Sans cette clè, la procédure de décryptage (c'est-à-dire de trouver x_i à partir de $f(x_i)$) doit être difficile, donc Ève aura beaucoup de difficulté de trouver x même si elle voit le message transmis x'. [1]

Les premiers cryptosystèmes développés se basaient sur la *cryptographie de clé privée*, appelée aussi *cryptographie symétrique*. Dans ce schème, les clés 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 établi 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 schème souffre de plusieurs déficiences. Premièrement, si Ève peut découvrir la clé de décryptage seulement une fois, elle pourra lire toutes les communications futures. Par exemple, pendant la Seconde Guerre mondiale, l'armée Allemagne avait développé 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 dû à la distance géographique entre l'Italie et la Chine !

La vraie révolution dans le domaine de la cryptographie est venue avec la naissance de la *cryptographie de clé publique*, aussi appelée *cryptographie asymétrique*. Ce protocole de cryptographie, proposé par Ralph Merkle, utilise un schème de communication asymétrique. Dans ce schème, 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, qui crypte les informations, mais il garde secrète la fonction réciproque f^{-1}. 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 schème de communication enlève l'exigence de l'échange des clés de cryptage et de décryptage entre Alice et Bob. Cependant, il rend la communication moins sécuritaire : Ève 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 schème de la cryptographie de clé publique l'échange 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. Il choisit deux premiers de cette liste d'ordre de grandeur similaire, soit p et q, et il crée le nombre $n := pq$. (Il est important que $p, q > M$.) Après cela, Bob peut facilement

1. La fonction f est une permutation. Donc, en principe, Ève peut trouver f^{-1} si elle possède assez d'informations. Par contre, ceci pourrait prendre beaucoup de temps. Le but d'Alice et de Bob est de s'assurer que le message envoyé reste protégé pour le temps maximal possible.

calculer $\phi(n) = (p-1)(q-1)$. Avec cette information, et en utilisant l'algorithme euclidien, il calcule aussi un nombre $E \in \{1, 2, \ldots, \phi(n)\}$ tel que $\operatorname{pgcd}(E, \phi(n)) = 1$. 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 mod n, soit $C \in \{0, 1, \ldots, n-1\}$. 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 : puisque $p, q > M$, on a que $(M, n) = 1$ et, par la suite, $M^{\phi(n)} \equiv 1 \pmod{n}$. Donc, il suffit que Bob calcule un nombre F tel que $EF \equiv 1 \pmod{\phi(n)}$, car dans ce cas-ci $EF = 1 + k\phi(n)$ pour quelque $k \in \mathbb{Z}$ 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 $M \pmod{n}$. Puisque $M < p, q < n$, ceci détermine M complètement.

Il reste à trouver une façon rapide de calculer le nombre F. Puisque ce nombre est déterminé par la relation $EF \equiv 1 \pmod{\phi(n)}$, il faut simplement inverser $E \pmod{\phi(n)}$. Son inverse F est la clé privée de notre cryptosystème.

L'algorithme euclidien nous permet de calculer rapidement F si on sait E et $\phi(n)$. Tout le monde sait E (cela fait partie de la clé publique), et Bob sait p et q, donc il sait aussi $\phi(n) = (p-1)(q-1)$. En fait, on a que $\phi(n) = n + 1 - p - q$, donc Ève saura aussi $\phi(n)$, si elle apprend $s := p + q$. Évidemment, Ève peut apprendre s en factorisant $n = pq$. Vice versa, si Ève sait s, alors elle peut résoudre facilement l'équation quadratique $x^2 - sx + n = 0$ (rappelons que n est publique) et calculer ses racines, qui sont p et q. En conclusion, Ève peut apprendre $\phi(n)$ si et seulement si elle peut factoriser $n = pq$.

On doit vérifier que ce cryptosystème satisfait les trois principes mentionnés au deuxième paragraphe. Étant donné n et E, le calcul de $M^E \pmod{n}$ est rapide. En effet, soit $E = e_0 + e_1 2 + e_2 2^2 \cdots + e_k 2^k$, où $e_j \in \{0, 1\}$ pour tout j, le développement binaire E. On a donc que

$$M^E = M^{e_0} (M^2)^{e_1} \cdots (M^{2^k})^{e_k}. \tag{8.4}$$

On sait les chiffres $e_0, e_1, \ldots, e_k$: les ordinateurs fonctionnent habituellement dans la base de 2. (Même si E est donné dans une autre base b, il est facile de calculer son développement binaire.) Donc il suffit de calculer les puissances $M, M^2, M^4, \ldots, M^{2^k}$ modulo n, ce qu'on peut faire facilement dans une façon itérative : d'abord, on trouve le reste de M mod n, soit $M_0 \in \{0, 1, \ldots, n-1\}$. Puis, on calcule M_0^2 et on trouve son reste mod n, soit M_1. En général, étant donné $M_i \in \{0, 1, \ldots, N-1\}$ tel que $M_i \equiv M^{2^i} \pmod{N}$, on trouve $M_{i+1} \in \{0, 1, \ldots, N-1\}$ tel que $M_{i+1} \equiv M_i^2 \equiv M^{2^{i+1}} \pmod{n}$. Étant donné M_i, on a besoin seulement de calculer $M_i \cdot M_i$ (une seule opération) et puis de le réduire mod n (ce qui prend $O(\log n)$ opérations selon la section 4.3). Donc, cela prend $O(\log n)$ opérations pour trouver M_{i+1}. Par induction, on conclut alors qu'on a besoin de $O(k \log n)$ opérations pour trouver $M^{2^i} \pmod{N}$, pour tout $i \in \{0, 1, \ldots, k\}$. Puis, on remplace M^{2^i} par M_i dans (8.4), et on calculer $M^E \pmod{N}$ de façon itérative : en supposant qu'on a calculé $P_j := \prod_{i=0}^{j-1} M_i^{e_i}$ mod n, le calcul de P_{j+1} mod n prend $O(\log n)$ opérations : on calcul $P_j \cdot M_j^{e_j}$, et puis on le réduit mod n. En total, on trouve que le calcul de M^E mod n exige $O(k \log n)$ opérations.

De même, le calcul de $C^F \pmod{n}$ exige $O(\ell \log n)$ opérations, où $\ell + 1$ dénote le nombre de chiffres binaire de F. Puisque $1 \leqslant F \leqslant \phi(n)$, on a que $\ell = O(\log n)$, donc le nombre de calcul exigé pour calculer $C^F \pmod{n}$ est $O((\log n)^2)$, c'est-à-dire un multiple constant du carré 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. Comme on l'a discuté, Ève peut lire le message M si elle peut factoriser le nombre n et trouver p et q. La sécurité de 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 d'ordre de grandeur similaire, et qui évitent certaines conditions additionnelles). C'est un algorithme exponentielle (au nombre de chiffres de n), donc il est assez lent, ce qui veut dire que Ève 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 seulement une espérance en ce moment-ci et non un théorème.

Afin de motiver la recherche en théorie des nombres computationnelle, la société *RSA Security* a crée la *compétition de factorisation RSA* [2]. Par exemple, elle a offert un prix de 75 000 dollars américains pour la factorisation du nombre

$$\begin{aligned}&412023436986659543855531365332575948179811699844327982845455626433876445565248426198098870423161841879261420247188869492560931776375033421130982397485150944909106910269861031862704114880866970564902903653658867433731720813104105\\&190864254793282601391257624033946373269391.\end{aligned}$$

Ce problème computationnel reste toujours ouvert.

8.5 Exercices

EXERCICE 8.1. Trouvez un critères de divisibilité d'un nombre par 11, 13 et 37, si le nombre est donné dans la base décimale.

EXERCICE 8.2. Trouvez un critères de divisibilité d'un nombre par 10, si le nombre est donné dans la base binaire.

EXERCICE 8.3. 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).

2. https://fr.wikipedia.org/wiki/Comp%C3%A9tition_de_factorisation_RSA

EXERCICE 8.4. Soit p un nombre premier. Cet exercice donne un troisième argument pour l'existence de racines primitives mod p. À cette fin, posons $\psi(d) = \#\{1 \leqslant a \leqslant p-1 : \mathrm{ord}_p(a) = d\}$. On veut montrer que $\psi(p-1) > 0$.

(a) Soit $d|(p-1)$, $d \geqslant 1$. Montrer que le polynôme $x^d - 1$ a exactement d racines mod p. [*Indice :* On sait que $x^d - 1$ a au plus d racines mod p d'après le théorème 7.14. D'autre côté, montrer que $x^{p-1} - 1 = (x^d - 1)f(x)$ pour un polynôme f à coefficients entiers de degré $p - 1 - d$. et en déduire que $x^d - 1$ a au moins d racines mod p.]

(b) Montrer que $\sum_{d|n} \psi(d) = n$ pour tout $n|(p-1)$.

(c) Montrer que $\psi(n) = \phi(n)$ pour tout $n|(p-1)$. [*Indice :* inversion de Möbius.]

(d) Déduire l'existence de racines primitives mod p.

EXERCICE 8.5. 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 $k \in \mathbb{Z}_{\geqslant 1}$.

(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 $d \geqslant 1$, 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 $\deg(f') = d - 1$.

Chapitre 9

Le théorème des restes chinois

Supposons qu'un empereur possède un armé énorme et il veut savoir le nombre exacte de ses soldats. Il cherche alors une façon de le calculer rapidement. Un théorème fameux d'arithmétique modulaire, appelé le *théorème des restes chinois*, lui offre une solution géniale : soit S le nombre des soldats. L'empereur ordonne à tous ses soldats de s'arranger en paires. À la fin, il reste soit aucun soldat ou un seul soldat sans paire. Cet exercice détermine le réside de S mod 2. Puis, l'empereur demandent aux soldats de former des triplets. À la fin, il reste 0, 1 ou 2 soldats, ce qui détermine le résidu de S mod 3. La procédure continue ainsi : pour tout nombre premier p plus petit qu'une certaine borne z, les soldats s'arrangent en p-tuples afin de déterminer le résidu de S mod p. Selon le théorème de restes chinois (cf. théorème 9.1), ceci détermine uniquement le résidu de S mod $m := \prod_{p \leqslant z} p$. Si on choisi la paramètre z pour que $m > S$ (par exemple, quand $z = 30$, on a $m = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 21 \cdot 23 \cdot 29$, qui est à peu prés égal à 6,5 billions. Certainement, S est plus petit que ce nombre, donc si on connait son résidu mod m, on connait S soi-même !

9.1 Systèmes linéaires de congruences

Le théorème des restes chinois (TRC) est un résultat fondamental de l'arithmétique modulaire qui nous permet de résoudre n'improrte quel système linéaire de congruences, c'est-à-dire de trouver tous les entiers x qui satisfont les équations

$$\begin{cases} r_1 x + s_1 \equiv 0 \pmod{m_1} \\ \vdots \\ r_k x + s_k \equiv 0 \pmod{m_k} \end{cases} \tag{9.1}$$

pour quelques moduli $m_1, \dots, m_k \in \mathbb{N}$ et quelques coefficients $r_1, \dots, r_k \in \mathbb{Z}$ et $s_1, \dots, s_k \in \mathbb{Z}$.

Tout d'abord, considérons le cas de base avec $k = 1$. Dans ce cas-ci, on a une seule équation, qu'on écrit comme

$$rx + s \equiv 0 \pmod{m} \tag{9.2}$$

pour simplifier la notation. Évidemment, ceci est équivalent à la relation $m|(rx + s)$. Si on pose $d = (r, m)$, alors on voit que $d|m$ et donc d doit diviser $rx + s$ pour toute solution x. Mais on sait aussi que $d|a$, donc il faut que $d|s$ s'il existe de solutions x de l'équation (9.2).

Réciproquement, on affirme que $d|s$, alors il existe de solutions x de (9.2). En effet, dans ce cas-ci, on peut écrire $r = dr'$, $s = ds'$ et $m = dn$, où $(r', n) = 1$. Avec cette notation, l'exercice 7.1 implique que l'équation (9.2) est équivalente $r'x \equiv -s' \pmod{n}$. Puisque $(r', n) = 1$, l'inverse de r' existe mod n, soit $\overline{r'} \pmod{n}$. Donc, la congruence $r'x \equiv -s' \pmod{n_j}$ est équivalente à

$$x \equiv a \pmod{n} \tag{9.3}$$

où $a \equiv -\overline{r'}s' \pmod{n}$. En particulier, on voit que (9.2) a de solutions qui forment une seule classe de congruence mod $n = m/(r, m)$.

En généralisant la discussion ci-dessus, on voit qu'une condition nécessaire pour que le système (9.1) ait de solutions est que

$$(r_j, m_j)|s_j \quad \text{pour tout } j = 1, 2, \ldots, k. \tag{9.4}$$

De plus, dans ce cas-ci, on peut transformer le système (9.1) au système simplifié

$$\begin{cases} x \equiv a_1 \pmod{n_1} \\ \vdots \\ x \equiv a_k \pmod{n_k} \end{cases} \tag{9.5}$$

avec $n_j = m_j/(r_j, m_j)$ et $a_j \equiv -\overline{r'_j}s_j \pmod{n_j}$, où $\overline{r'_j}$ dénote l'inverse multiplicatif de $r'_j = r_j/(r_j, m_j)$ mod n_j.

Étudions maintenant le système (9.5) quand $k \geqslant 2$. Évidemment, si x est une solution de (9.5), alors $x \equiv a_i \pmod{(n_i, n_j)}$ et $x \equiv a_j \pmod{(n_i, n_j)}$. Donc, il est nécessaire d'avoir la «condition de compatibilité»

$$(n_i, n_j)|a_i - a_j \quad \text{pour tous } i \neq j \tag{9.6}$$

pour que le système (9.5) ait de solutions. En fait, comme on va le voir, la condition (9.6) est aussi suffisante pour que (9.5) ait de solutions. De plus, ses solutions forment toutes ensemble une seule classe de congruence mod $[n_1, \ldots, n_k]$.

La théorie de la résolution du système (9.5) assume sa forme la plus simple quand les moduli $n_1, \ldots, n_k$ sont deux-à-deux copremiers, c'est-à-dire quand $(n_i, n_j) = 1$ pour tous $i \neq j$. Dans ce cas-ci, la condition (9.6) est automatiquement satisfaite.

Prenons d'abord le cas $k = 2$. On cherche x tel que $x \equiv a_1 \pmod{n_1}$ et $x \equiv a_2 \pmod{n_2}$. La forme générale des x satisfaisant la première relation est $x = qn_1 + a_1$ avec $q \in \mathbb{Z}$. Donc on cherche q tel que

$$qn_1 + a_1 \equiv a_2 \pmod{n_2} \quad \Leftrightarrow \quad qn_1 \equiv a_2 - a_1 \pmod{n_2}.$$

Si $(n_1, n_2) = 1$, alors n_1 est inversible mod n_2, d'où on déduit que $q \equiv \overline{n}_1(a_2 - a_1) \pmod{n_2}$ où $\overline{n}_1$ dénote l'inverse de n_1 mod n_2. Si $b_2 \in \{0, 1, \ldots, n_2 - 1\}$ est un représentant de la classe de congruences $\overline{n}_1(a_2 - a_1) \pmod{n_2}$, on trouve que $q = rn_2 + b_2$ avec $r \in \mathbb{Z}$. Donc, la solution générale est donnée par $x = (rn_2 + b_2)n_1 + a_1 = rn_1n_2 + (b_2n_1 + a_1)$, ce qui montre au même temps

l'existence de solutions x et qu'elles forment toutes ensemble une seule classe de congruence mod $n_1 n_2 = [n_1, n_2]$.

Pour traiter le cas général, on utilise un argument itératif : d'après le cas $k = 2$ qu'on a montré ci-dessus, on peut transformer le système de congruences $x \equiv a_1 \pmod{n_1}$ et $x \equiv a_2 \pmod{n_2}$ à une seule congruence, soit $x \equiv a \pmod{n_1 n_2}$. Donc, on peut passer d'un système avec k congruences dont les moduli sont $n_1, \ldots, n_k$. à un nouveau système avec $k-1$ congruences dont les moduli sont les nombres $n_1 n_2$ et $n_3, \ldots, n_k$. Les nouveaux moduli sont aussi copremiers deux-à-deux car les anciens le sont.

En continuant de cette façon, on arrive au théorème des restes chinois :

Théorème 9.1 (théorème des restes chinois)**.** *Soient $n_1, \ldots, n_k \in \mathbb{N}$ tels que $(n_i, n_j) = 1$ si $i \neq j$. Soient aussi $a_1, \ldots, a_k \in \mathbb{Z}$. Le système de congruences*

$$\begin{cases} x \equiv a_1 \pmod{n_1} \\ \vdots \\ x \equiv a_k \pmod{n_k} \end{cases} \tag{9.7}$$

a une solution unique modulo $n_1 \cdots n_k$.

Il existe une autre preuve de ce théorème qui est un peu plus directe. De plus, elle nous donne une façon alternative de calculer les solutions au système (9.7) qui est plus utile dans les applications théoriques du théorème 9.1.

Deuxième démonstration du TRC. On pose $N = n_1 \cdots n_k$ et, pour chaque $j \in \{1, \ldots, k\}$, on pose

$$N_j = \frac{N}{n_j} = n_1 \cdots n_{j-1} n_{j+1} \cdots n_k.$$

Les deux observations fondamentales sont que $(n_j, N_j) = 1$ et que $N_j \equiv 0 \pmod{n_i}$ pour $i \neq j$. La première relation garanti que N_j est inversible mod n_j. Soit $\overline{N}_j$ un représentant de la classe de congruence inverse de $N_j \pmod{n_j}$, c'est-à-dire $\overline{N}_j N_j \equiv 1 \pmod{n_j}$. Par conséquent,

$$\overline{N}_j N_j \equiv \begin{cases} 1 \pmod{n_j} \\ 0 \pmod{n_i} \quad \text{si } i \neq j. \end{cases}$$

On voit alors que si on pose

$$x_0 = a_1 \overline{N}_1 N_1 + \cdots + a_k \overline{N}_k N_k,$$

on a que $x_0 \equiv a_j \pmod{n_j}$ pour chaque $j \in \{1, \ldots, k\}$, c'est-à-dire que x_0 est une solution du système de congruences (9.7).

Il reste à montrer que la solution x_0 qu'on a trouvé est unique mod N. En effet, soit x_1 une autre solution du système 9.7. On a alors que $x_0 \equiv x_1 \pmod{n_j}$ pour $j = 1, 2, \ldots, k$. Donc, $x_0 \equiv x_1 \pmod{[n_1, \ldots, n_k]}$ d'après l'exercice 7.2. Puisque on a supposé que les nombres $n_1, \ldots, n_k$ sont deux-à-deux co-premiers, on a que $[n_1, \ldots, n_k] = n_1 \cdots n_k = N$ (cf. exercice 2.13 (c)). Par la suite, on conclut que $x_0 \equiv x_1 \pmod{M}$, comme désiré. □

Le cas spécial traité au théorème 9.1 nous permet facilement de traiter aussi le cas général du système 9.7 :

Théorème 9.2 (TRC - le cas général)**.** *Soient $n_1, \dots, n_k \in \mathbb{N}$ et $a_1, \dots, a_k \in \mathbb{Z}$, et soit le système de congruences*

$$\begin{cases} x \equiv a_1 \pmod{n_1} \\ \vdots \\ x \equiv a_k \pmod{n_k} \end{cases} \tag{9.8}$$

Si ce système a de solutions, alors

$$(n_i, n_j) | a_i - a_j \quad \textit{pour tous } i \neq j. \tag{9.9}$$

Réciproquement, si (9.9) est vraie, alors le système (9.8) a de solutions et ses solutions forment toutes ensemble une seule classe de congruence modulo $[n_1, \dots, n_k]$.

Démonstration. On a déjà montré que Pour tout $j \in \{1, \dots, k\}$, soit la factorisation première $n_j = \prod_p p^{v_p(n_j)}$. Ses facteurs sont deux-à-deux copremiers, donc les exercices 7.2 et 2.13 (c) impliquent que

$$x \equiv a_j \pmod{n_j} \quad \Longleftrightarrow \quad x \equiv a_j \pmod{p^{v_p(n_j)}} \quad \text{pour tout } p.$$

Ceci veut dire que le système (9.8) est équivalent au système

$$x \equiv a_j \pmod{p^{v_p(n_j)}} \quad \text{pour tout } p \text{ et tout } j = 1, \dots, k. \tag{9.10}$$

Fixons p et posons

$$\nu_p = \max\{v_p(n_1), \dots, v_p(n_k)\} = v_p([n_1, \dots, n_k]).$$

Soit $j_p \in \{1, \dots, k\}$ tel que $\nu_p = v_p(n_{j_p})$ (si plusieurs tels indices existent, prenons le plus petit), et soit $b_p = a_{j_p}$. Notre hypothèse que $(n_i, n_j) | a_i - a_j$ pour tous $i \neq j$ implique que $p^{\min\{v_p(n_i), v_p(n_j)\}} | a_i - a_j$ pour tous $i \neq j$. En particulier, on voit que

$$x \equiv a_j \pmod{p^{v_p(n_j)}} \quad \text{pour tout } j = 1, 2, \dots, k \qquad \Longleftrightarrow \qquad x \equiv b_p \pmod{p^{\nu_p}}.$$

En conclusion, le système (9.10) est équivalent au système

$$x \equiv b_p \equiv p^{\nu_p} \quad \text{pour tout } p. \tag{9.11}$$

Ici $\nu_p = v_p([n_1, \dots, n_k])$, donc $\nu_p = 0$ pour tous les premiers p qui sont assez grands. Donc, le système (9.11) est fini et ses moduli sont co-premiers. D'après le théorème 9.1, on déduit que le système (9.11) a une solution unique modulo $\prod_p p^{\nu_p} = [n_1, \dots, n_k]$. Ceci conclut la preuve. □

Exemples. (a) On veut résoudre le système de congruences

$$\begin{cases} x \equiv 1 \pmod 5 \\ x \equiv 3 \pmod 4 \\ x \equiv 2 \pmod 3. \end{cases} \tag{9.12}$$

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 (9.12), qu'on peut réécrire comme $x \equiv 11 \pmod{60}$, comme avant.

(b) Soit le système de congruences

$$\begin{cases} x \equiv 1 \pmod 6 \\ x \equiv 4 \pmod{15} \end{cases} \tag{9.13}$$

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 l'exercice 7.1. Donc $y = 5z + 3$, ce qui implique que $x = 6(5z + 3) + 1 = 30z + 19$, c'est-à-dire la solution générale au système (9.13) est $x \equiv 19 \pmod{30}$.

(c) Or, résolvons le système de congruences

$$\begin{cases} x \equiv 1 \pmod 5 \\ x \equiv 2 \pmod{10} \\ x \equiv 3 \pmod 7. \end{cases} \tag{9.14}$$

On observe que $(5, 10) = 5$ mais $5 \nmid 1 - 2$. Donc la condition (9.6) échoue quand $i = 1$ et $j = 2$, ce qui veut dire que le système (9.14) n'a pas de solutions.

(d) Notre dernier exemple est le système de congruences

$$\begin{cases} 3x \equiv 3 \pmod{27} \\ 3x \equiv 2 \pmod 5 \\ x \equiv 7 \pmod{21}. \end{cases} \tag{9.15}$$

Tout d'abord, il faut enlever le coefficient 3 qui multiplie x aux premières deux congruences. On a que

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

selon l'exercice 7.1. De plus,

$$3x \equiv 2 \pmod{5} \quad \Leftrightarrow \quad x \equiv \overline{3} \cdot 2 \pmod{5} \quad \Leftrightarrow \quad x \equiv 4 \pmod{5},$$

puisque $\overline{3} \equiv 2 \pmod{5}$. Donc le système (9.15) est équivalent au système

$$\begin{cases} x \equiv 1 \pmod{9} \\ x \equiv 4 \pmod{5} \\ x \equiv 7 \pmod{21} \end{cases} \tag{9.16}$$

Puis, 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

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

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

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

Donc $z = 7w + 3$, ce qui implique que $x = 45(7z + 3) + 19 = 315w + 154$, c'est-à-dire la solution générale du système (9.16) est $x \equiv 154 \pmod{315}$. ■

9.2 Multiplicativité de fonctions arithmétiques

Le théorème des restes chinois donne plus contexte au remarque 5.2 que quand $(m, n) = 1$, alors plusieurs phénomènes multiplicatives reliés à la structure multiplicative de m se comportent de façon «indépendante» des phénomènes correspondant à n. Pour voir cette affirmation en pratique, on utilise le théorème des restes chinois pour montrer la multiplicativité de certaines fonctions arithmétiques.

On commence avec la multiplicativité de la fonction ϕ d'Euler, qu'on a déjà montré de deux autres façons.

Théorème 9.3. *La fonction ϕ d'Euler est multiplicative.*

Démonstration. Soient $m, n \in \mathbb{N}$ avec $(m, n) = 1$. Considérons la fonction 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 $x \equiv y \pmod{n}$, d'après l'exercice 7.2.

Ces relations montrent que la fonction f est bien définie et 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 $x \equiv b \pmod{n}$. Puisque $(a, m) = 1$ et $x \equiv a \pmod{m}$, on a que $(x, m) = 1$. Idem, on trouve que $(x, n) = 1$. Donc, $(x, mn) = 1$,

ce qui implique que $x \pmod{mn}$ est dans le domaine. Puisque on a aussi que $f(x \pmod{mn}) = (a \pmod{m}, b \pmod{n})$, il en découle que f est surjective.

On a alors montré que f est une bijection. Par la suite,

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

où on a utilisé le théorème 7.11 deux fois. Ceci conclut la démonstration. □

Théorème 9.4. *Soit $f(x)$ un polynôme à coefficients entiers. Soit aussi ρ_f la fonction arithmétique définie par*

$$\rho_f(n) := \#\big\{x \pmod{n} : f(x) \equiv 0 \pmod{n}\big\}$$

pour tout $n \in \mathbb{N}$. Alors, ρ_f est multiplicative.

Démonstration. On rappelle que $\mathcal{R}_f(n) = \{x \pmod{n} : f(x) \equiv \pmod{n}\}$ (voir (7.1)). En particulier, $\rho_f(n) = |\mathcal{R}_f(n)|$ pour tout $n \in \mathbb{N}$.

Soient $m, n \in \mathbb{N}$ avec $(m, n) = 1$. Comme dans la preuve du théorème 9.3, on considère la fonction $g : \mathcal{R}_f(mn) \to \mathcal{R}_f(m) \times \mathcal{R}_f(n)$, définie par $f(x \pmod{mn}) := (x \pmod{m}, x \pmod{n})$. On laisse comme exercice aux lecteurs de vérifier que :

— g est bien définie, ce qui veut que si $x \equiv y \pmod{mn}$, alors $(x \pmod{m}, x \pmod{n}) = (y \pmod{m}, y \pmod{n})$;
— g est injective ;
— g est surjective.

Donc, g est une bijection entre $\mathcal{R}_f(mn)$ et $\mathcal{R}_f(m) \times \mathcal{R}_f(n)$. On en déduit que

$$\rho_f(mn) = |\mathcal{R}_f(mn)| = |\mathcal{R}_f(m)| \cdot |\mathcal{R}_f(n)| = \rho_f(m)\rho_f(n),$$

ce qui termine la preuve. □

Finalement, on donne deux applications du théorème des restes chinois concernant l'ordre multiplicatif.

Lemme 9.5. *Soient $m, n \in \mathbb{N}$ avec $(m, n) = 1$, et soit $a \in \mathbb{Z}$ tel que $(a, mn) = 1$. On a que* $\operatorname{ord}_{mn}(a) = [\operatorname{ord}_m(a), \operatorname{ord}_n(a)]$.

Démonstration. Soit $k = \operatorname{ord}_m(a)$ et $\ell = \operatorname{ord}_n(a)$. On a que

$$a^r \equiv 1 \pmod{mn} \overset{\text{exercice 7.2}}{\iff} \begin{cases} a^r \equiv 1 \pmod{m} \\ a^r \equiv 1 \pmod{n} \end{cases} \overset{\text{lemme 8.2}}{\iff} \begin{cases} k|r \\ \ell|r \end{cases} \overset{\text{lemme 2.14}}{\iff} [k, \ell]|r.$$

Évidemment, ces équivalences impliquent que le plus petit $r \in \mathbb{N}$ tel que $x^r \equiv 1 \pmod{mn}$ est le nombre $r = [k, \ell]$. Donc, $\operatorname{ord}_{mn}(a) = [k, \ell]$, comme affirmé. □

Théorème 9.6. *Si $m, n \in \mathbb{N}$ avec $(m, n) = 1$, alors $\lambda(mn) = [\lambda(m), \lambda(n)]$.*

Démonstration. Posons $k = \lambda(m)$ et $\ell = \lambda(n)$. Soit $x \in \mathbb{Z}$ avec $(x, mn) = 1$. Il faut donc que $(x, m) = (x, n) = 1$. Puisque $[k, \ell]$ est un multiple de k et de ℓ, et on sait que $x^k \equiv 1 \pmod m$ et $x^\ell \equiv 1 \pmod n$, on trouve aussi que $x^{[k,\ell]} \equiv 1 \pmod m$ et que $x^{[k,\ell]} \equiv 1 \pmod n$. Donc, $x^{[k,\ell]} \equiv 1 \pmod{mn}$, d'après l'exercice 7.2. Puisque cette relation est vraie pour n'importe quel résidu réduit $x \pmod{mn}$, la définition de la fonction de Carmichael implique que $\lambda(mn) \leqslant [k, \ell]$.

Afin de montrer l'inégalité inverse, on considère $a \pmod m \in (\mathbb{Z}/m\mathbb{Z})^*$ d'ordre maximale mod m et $b \pmod n \in (\mathbb{Z}/n\mathbb{Z})^*$ d'ordre maximal mod n. Donc, $\mathrm{ord}_m(a) = \lambda(m) = k$ et $\mathrm{ord}_n(b) = \lambda(n) = \ell$ d'après le théorème 8.7. On applique le théorème de restes chinois pour trouver $x \pmod{mn}$ tel que $x \equiv a \pmod m$ et $x \equiv b \pmod n$. Donc, $\mathrm{ord}_m(x) = \mathrm{ord}_m(a) = k$ et $\mathrm{ord}_n(x) = \mathrm{ord}_n(b) = \ell$. Une application du lemme 9.5 nous donne que $\mathrm{ord}_{mn}(x) = [k, \ell]$. D'après le théorème 8.7, il découle que $\lambda(mn)$, qui est l'ordre maximal mod mn, est $\geqslant [k, \ell]$. Ceci conclut la preuve du théorème. □

Corollaire 9.7. *Si $n = p_1^{\nu_1} \cdots p_k^{\nu_k}$, où $p_1, \ldots, p_k$ sont de nombres premiers distincts, alors*

$$\lambda(n) = [\lambda(p_1^{\nu_1}), \ldots, \lambda(p_k^{\nu_k})].$$

Démonstration. On utilise induction sur k et l'exercice 2.13 (a). □

9.3 Nombres premiers de la forme $2^n + c$

Le théorème 5.6 nous amène naturellement à l'étude des nombres premiers $2^p - 1$, où p est également premier, appelés les *nombres premiers de Mersenne*. En fait, on n'a pas besoin de l'exigence que p est premier : si un entier de la forme $2^n - 1$ est premier, alors il faut que n soit premier également (voir la dernière partie de la démonstration du théorème 5.6). Les nombres premiers de Mersenne sont très mystérieux : même s'ils sont très rares (ils sont moins nombreux que les puissances de 2), on croit qu'il en existe un nombre infini. Cependant, leur pénurie rende leur étude très difficile et on sait presque rien pour eux.

Fermat a étudié une question «jumelle» : quels nombres de la forme $2^n + 1$ sont premiers ? Pour que $2^n + 1$ soit un nombre premier, il faut que $n = 2^k$ pour quelque $k \in \mathbb{N}$ (cf. exercice 1.2 (b)). Fermat a ensuite remarqué que les nombres $2^{2^k} + 1$ sont premiers quand $k \in \{0, 1, 2, 3, 4\}$, ce qui correspond aux nombres premiers $2, 5, 17, 65537$, respectivement. À partir de cette observation, Fermat a conjecturé que les nombres $2^{2^k} + 1$ sont toujours premiers, une conclusion qui serait vraiment magnifique, car elle montrerait l'existence d'une suite très structurée parmi les nombres premiers. Cependant, Euler a réfuté la conjecture de Fermat en prouvant que 641 divise le nombre $2^{2^5} + 1 = 4,294,967,297$. En fait, aujourd'hui on croit qu'il existe seulement un nombre fini de $k \in \mathbb{N}$ tels que $2^{2^k} + 1$ est un nombre premier. Certains mathématiciens même pensent que les nombres $2^{2^k} + 1$ sont tous composés dès que $k \geqslant 5$. Par contre, l'étude de la primalité des nombres de la forme $2^{2^k} + 1$ n'est pas accessible par les techniques disponibles modernes.

Comme on l'a vu, notre compréhension de la primalité des nombres de la forme $2^n - 1$ et $2^n + 1$ est très limitée. Néanmoins, Paul Erdős a montré en 1950 [1] qu'il existe un nombre naturel k pour lequel tous les nombres $2^n + k$ avec $n \in \mathbb{Z}_{\geqslant 0}$ sont composés. Sa preuve produit la valeur $k = 9,262,111$.

Voici l'idée d'Erdős. Afin de trouver k tel que $2^n + k$ est toujours composé, il suffit de trouver quelques nombres $d_1, \ldots, d_r$ qui ont la propriété suivante : pour tout $n \in \mathbb{N}$, il existe $j \in \{1, \ldots, r\}$

tel que $d_j | 2^n + k$. On peut réécrire cette dernière égalité comme

$$2^n \equiv -k \pmod{d_j}. \tag{9.17}$$

On impose la condition que les nombres $d_1, \ldots, d_r$ soient impairs, pour que 2 est inversible mod d_j pour tout j. En particulier, ceci implique que la fonction $n \to 2^n \pmod{d_j}$ est m_j-périodique, où $m_j := \mathrm{ord}_{d_j}(2)$. Par la suite, $2^n \equiv 2^{a_j} \pmod{d_j}$ quand $n \equiv a_j \pmod{m_j}$. Donc, si on choisit $k \equiv -2^{a_j} \pmod{d_j}$, on garantit que (9.17) est vrai. On conclut alors qu'il suffit de trouver quelques classes de congruences $a_1 \pmod{m_1}, \ldots, a_r \pmod{m_r}$ telles que

$$\mathbb{Z} = \bigcup_{j=1}^{r} \{n \in \mathbb{Z} : n \equiv a_j \pmod{m_j}\} \tag{9.18}$$

et pour lesquelles il existe des nombres $d_1, \ldots, d_r$ avec $m_j = \mathrm{ord}_{d_j}(2)$. Si on peut faire celui-ci, alors on peut choisir c satisfaisant $k \equiv -2^{a_j} \pmod{d_j}$ pour tout $j \in \{1, \ldots, r\}$, ce qu'on peut faire d'après le théorème des restes à la condition additionnelle que les nombres $d_1, \ldots, d_r$ sont deux à deux co-premiers.

Définition 9.8. Un ensemble de classes de congruences $\{a_1 \pmod{m_1}, \ldots, a_r \pmod{m_r}\}$ est appelé un *système couvrant de congruences* s'il satisfait (9.18).

Un exemple trivial d'un système couvrant est donné par les congruences $0 \pmod 2$ et $1 \pmod 2$. Erdős a construit un exemple beaucoup plus élaboré qui lui a permis de montrer que les nombres $2^n + 9\,262\,111$ sont toujours composés. Il a considéré l'ensemble

$$\{1 \pmod 2, 2 \pmod 4, 1 \pmod 3, 4 \pmod 8, 8 \pmod{12}, 0 \pmod{24}\}.$$

Tous les moduli divisent 24. Donc, on peut vérifier facilement que l'exemple d'Erdős est en effet un système couvrant de congruences en montrant que les nombres $1, 2, \ldots, 24$ sont tous couverts. De plus, on a que

$$\mathrm{ord}_3(2) = 2,\ \mathrm{ord}_5(2) = 4,\ \mathrm{ord}_7(2) = 3,\ \mathrm{ord}_{17}(2) = 8,\ \mathrm{ord}_{13}(2) = 12,\ \mathrm{ord}_{241}(2) = 24.$$

Donc, on choisit k appartenant à l'intersection des classes de congruence

$$-2^1 \pmod 2,\ -2^2 \pmod 5,\ 2^1 \pmod 7,\ -2^4 \pmod 8,\ -2^8 \pmod{12},\ -2^0 \pmod{241}. \tag{9.19}$$

Ceci est possible d'après le théorème des restes chinois. Par la discussion ci-dessous, tout tel choix de k a la propriété que les nombres $2^n + k$ sont tous composés ; en fait, ils sont divisés par un nombre parmi $3, 5, 7, 17, 13$ et 241. On peut montrer facilement que le plus petit positif k qui appartient à l'intersection des classes de congruence (9.19) est $k = 9\,262\,111$.

9.4 Exercices

EXERCICE 9.1. 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 9.2. Montrer que les congruences suivantes forment un système couvrant :

$0 \pmod 3, 0 \pmod 4, 0 \pmod 5, 1 \pmod 6, 6 \pmod 8, 3 \pmod{10}, 5 \pmod{12}, 11 \pmod{15},$
$7 \pmod{20}, 10 \pmod{24}, 2 \pmod{30}, 34 \pmod{40}, 59 \pmod{60}, 98 \pmod{120}$

EXERCICE 9.3. Montrer qu'il existe $k \in \mathbb{N}$ tel que $k \cdot 2^n + 1$ est premier pour tout $n \in \mathbb{Z}_{\geqslant 0}$.

Chapitre 10

Équations polynomiales modulo p^ν

Soit $f(x)$ un polynôme à coefficients 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 $\mathbb{Z}$. On reviendra à ce sujet à la section 13.1.

10.1 Le lemme de Hensel

Soit l'équation polynomiale $f(x) \equiv 0 \pmod n$ où $f(x)$ est un polynôme à coefficients entiers. D'après le théorème des restes chinois (voir la preuve du théorème 9.4 et le théorème 5.3), il suffit de considérer le cas où n est une puissance première p^ν. Dans ce chapitre, on verra comment réduire ensuite ce problème au cas où $\nu = 1$: on montrera que si on connait toutes les racines de $f(x)$ mod p et si p n'appartient pas à un certain ensemble fini de nombres premiers exceptionnels, alors on peut aussi déterminer toutes les racines de $f(x)$ mod p^ν pour tout $\nu \geqslant 2$.

Avant énoncer les résultats, il est utile de visualiser le problème. Considérons l'ensemble $\mathcal{R}_f(p^\nu)$ de toutes les racines mod p^ν du polynôme $f(x)$. Si on réduit tous ses éléments mod $p^{\nu-1}$, on trouve un sous-ensemble $\mathcal{R}_f(p^{\nu-1})$. Il est possible que plusieurs éléments de $\mathcal{R}_f(p^\nu)$ se réduisent à la même classe mod $p^{\nu-1}$, et que certains racines mod $p^{\nu-1}$ ne sont pas obtenues de cette façon. Cette procédure donc crée une forêt [1] qui a $|\mathcal{R}_f(p)|$ arbres, un pour chaque élément de $\mathcal{R}_f(p)$. Les branches des arbres forment de suites $x_\nu \pmod{p^\nu} \in \mathcal{R}_f(p^\nu)$, $\nu = 1, 2, \ldots$, satisfaisant la «condition de compatibilité»

$$x_{\nu+1} \equiv x_\nu \pmod{p^\nu} \quad \text{pour tout } \nu \geqslant 1. \tag{10.1}$$

Exemples. (a) Soit $f(x) = x$. Donc, $\mathcal{R}_f(p^\nu) = \{0 \pmod{p^\nu}\}$ pour tout $\nu \geqslant 1$, ce qui veut dire qu'on a la forêt triviale avec une seule branche, donnée par $0 \pmod p, 0 \pmod{p^2}, 0 \pmod{p^3}, \ldots$ (cf. figure 10.1)

1. Une forêt est un graphe acyclique ; un arbre est un graphe acyclique qui est aussi connexe. Évidemment, on peut écrire chaque forêt comme une réunion d'arbres.

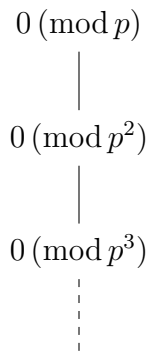

FIGURE 10.1 – La forêt des racines de $f(x) = x$ modulo p^ν

(b) Soit $f(x) = x^2$. Dans ce cas-ci, on a que $x \pmod{p^\nu} \in \mathcal{R}_f(p^\nu)$ si et seulement si $p^\nu | x^2$. On écrit $\nu = 2k - r$ avec $r \in \{0, 1\}$ et $k \in \mathbb{N}$. Donc, $p^\nu | x^2$ si et seulement si $p^k | x$. Par la suite,

$$\mathcal{R}_f(p^\nu) = \begin{cases} \{p^k m \pmod{p^\nu} : 0 \leqslant m < p^{k-1}\} & \text{si } \nu = 2k-1, \\ \{p^k m \pmod{p^\nu} : 0 \leqslant m < p^k\} & \text{si } \nu = 2k. \end{cases}$$

La forêt correspondant aux racines de $f(x)$ modulo les puissances de 2 est décrite dans la figure 10.2.

(c) Soit $f(x) = x^2 - 1$. Si $p > 2$, on affirme que

$$\mathcal{R}_f(p^\nu) = \{1 \pmod{p^\nu}, -1 \pmod{p^\nu}\} \quad \text{pour tout } \nu \geqslant 1. \tag{10.2}$$

En effet, on a que $x \pmod{p^\nu}$ si et seulement si $p^\nu | (x^2 - 1) = (x-1)(x+1)$. Donc, il doit exister $k, \ell \geqslant 0$ tels que $\nu = k + \ell$, $p^k | x - 1$ et $p^\ell | x + 1$. Afin de montrer (10.2), il suffit de prouver que soit $k = \nu$ ou $\ell = \nu$. Tout d'abord, si $k = 0$, on a que $\ell = \nu - 0 = \nu$ comme il faut. Or, supposons que $k \geqslant 1$. En particulier, on a que $p | x - 1$. D'autre côté, $-1 \not\equiv 1 \pmod{p}$, car $p > 2$. Donc, il faut que $p \nmid x + 1$. Par la suite, la seule façon d'avoir que $p^\ell | x + 1$ est si $\ell = 0$, comme désiré. Ceci termine la preuve de (10.2). Par conséquent, le forêt de solutions de $f(x)$ mod puissances de p a deux seules branches distinctes, correspondant aux deux racines mod p (cf. figure 10.3).

Par contre, quand $p = 2$, la situation change. On affirme que

$$\mathcal{R}_f(2^\nu) = \begin{cases} \{1 \pmod{2}\} & \text{si } \nu = 1, \\ \{1 \pmod{4}, -1 \pmod{4}\} & \text{si } \nu = 2, \\ \{1 \pmod{2^\nu}, 1 + 2^{\nu-1} \pmod{2^\nu}, -1 \pmod{2^\nu}, -1 - 2^{\nu-1} \pmod{2^\nu}\} & \text{si } \nu \geqslant 3. \end{cases}$$

On peut vérifier les cas où $\nu = 1, 2, 3$ par un calcul direct. Or, considérons le cas avec $\nu \geqslant 4$. On a que $x \pmod{2^\nu} \in \mathcal{R}_f(2^\nu)$ si et seulement si $2^\nu | (x-1)(x+1)$. Comme avant, ceci veut dire qu'il existe $k, \ell \in \mathbb{Z}_{\geqslant 0}$ tels que $k + \ell = \nu$, $2^k | x - 1$ et $2^\ell | x + 1$. On montrera que soit $k \in \{0, 1\}$ ou $\ell \in \{0, 1\}$. Ceci découle du fait qu'il n'est pas possible d'avoir $2^2 | x - 1$ et $2^2 | x + 1$, car $1 \not\equiv -1 \pmod{2^2}$. Nous avons donc quatre cas pour la paire (k, ℓ) :

— $(k, \ell) = (0, \nu)$, auquel cas $x \equiv -1 \pmod{2^\nu}$;

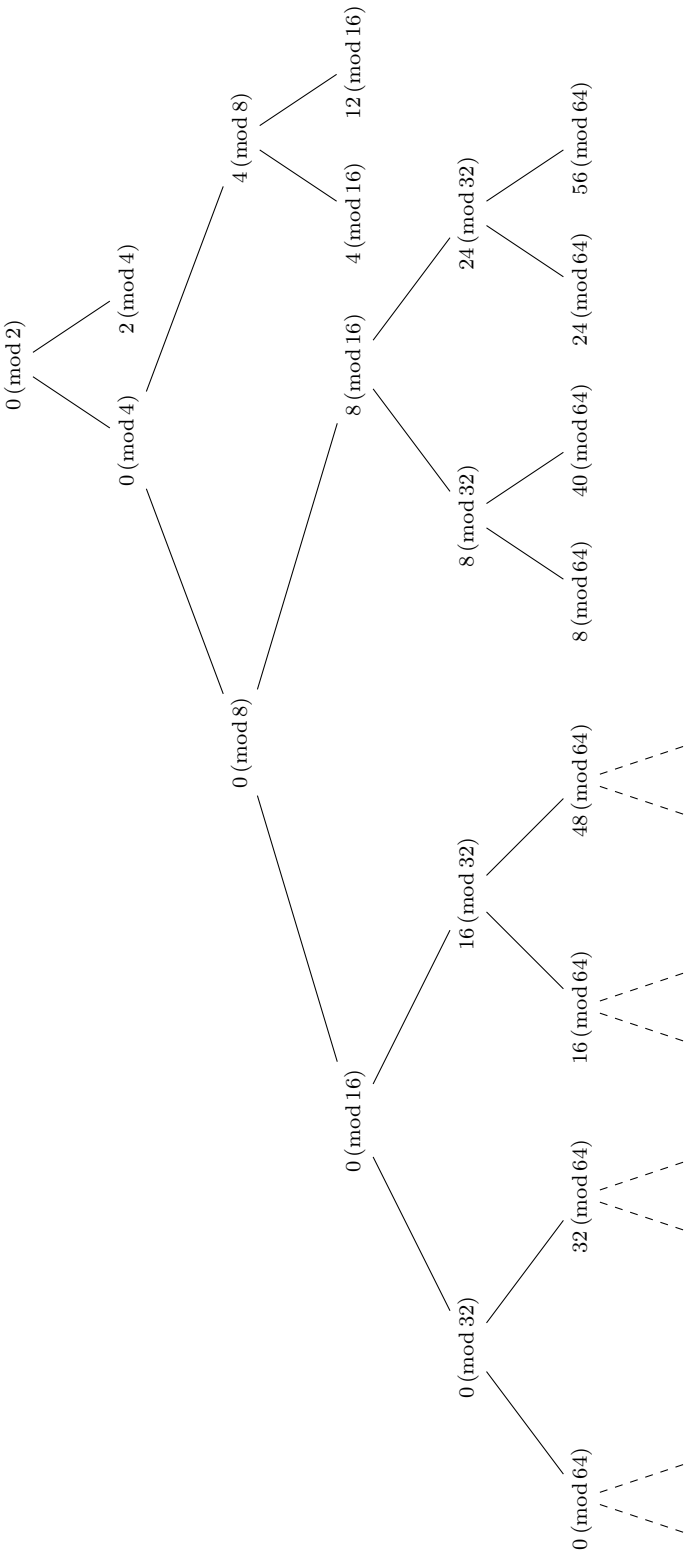

FIGURE 10.2 – La forêt des racines de $f(x) = x^2$ modulo 2^ν

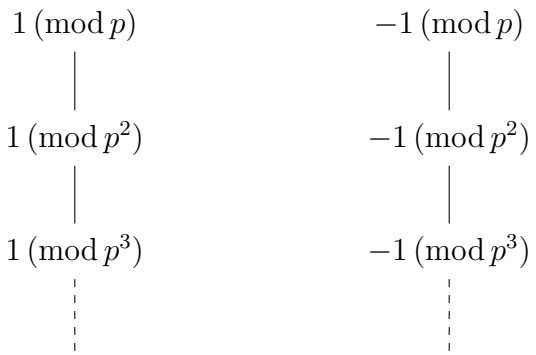

FIGURE 10.3 – Le forêt des racines de $f(x) = x^2 - 1$ modulo p^ν quand $p > 2$

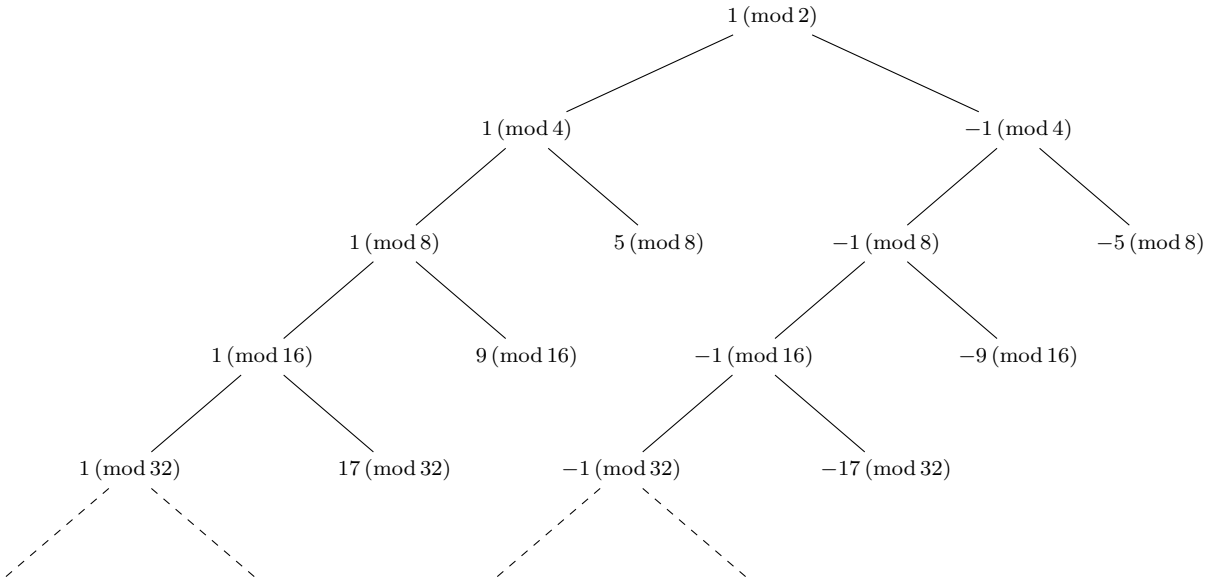

FIGURE 10.4 – La forêt des racines de $f(x) = x^2 - 1$ modulo 2^ν

— $(k, \ell) = (\nu, 0)$, auquel cas $x \equiv 1 \pmod{2^\nu}$;
— $(k, \ell) = (1, \nu - 1)$, auquel cas $x \equiv 1 \pmod 2$ et $x \equiv -1 \pmod{2^{\nu-1}}$. La deuxième relation implique la première, donc $x \equiv -1 \pmod{2^{\nu-1}}$, ce qui veut dire que soit $x \equiv -1 \pmod{2^\nu}$ ou $x \equiv -1 - 2^{\nu-1} \pmod{2^\nu}$.
— $(k, \ell) = (\nu - 1, 1)$, auquel cas $x \equiv 1 \pmod{2^{\nu-1}}$ et $x \equiv -1 \pmod 2$. Comme au cas précédent, ceci veut dire que soit $x \equiv 1 \pmod{2^\nu}$ ou $x \equiv 1 + 2^{\nu-1} \pmod{2^\nu}$.

Ceci conclut notre calcul de $\mathcal{R}_f(2^\nu)$.

On remarque que, dans cet exemple, on a

$$\{x \pmod{2^{\nu-1}} : x \pmod{2^\nu} \in \mathcal{R}_f(2^\nu)\} = \{1 \pmod{2^{\nu-1}}, -1 \pmod{2^{\nu-1}}\} \subsetneq \mathcal{R}_f(2^{\nu-1})$$

pour tout $\nu \geqslant 4$. La forêt correspondant aux racines de $f(x)$ modulo les puissances de 2 est décrite dans la figure 10.4. ■

Parmi tous les exemples qu'on a étudie ci-dessus, les plus simples de loin sont quand $f(x) = x$ et quand $f(x) = x^2 - 1$ et $p > 2$. La propriété commune de ces deux exemples est que les racines de $f(x)$ modulo p sont toutes simples, c'est-à-dire les exposants m_j au théorème 7.14 sont tous égaux à 1. Comme on sait du cas des polynômes réels, une racine α d'un polynôme f est simple si sa dérivée f' ne s'annule pas en α. Le même résultat est vrai aussi mod p. Pour cette raison, étant donné un polynôme $f(x) = c_d x^d + \cdots + c_1 x + c_0$, on définit sa dérivé par

$$f'(x) = dc_d x^{d-1} + (d-1)c_{d-1}x^{d-2} + \cdots 2c_2 x + c_1.$$

Avec cette notation, nous avons le résultat suivant fondamental.

Théorème 10.1 (Hensel)**.** *Soit $f(x) \in \mathbb{Z}[x]$ et soit $a \in \mathbb{Z}$ tel que $f(a) \equiv 0 \pmod p$ et $f'(a) \not\equiv 0 \pmod p$. Alors, il existe une suite unique de résidus $x_\nu \pmod{p^\nu}$, $\nu = 1, 2, \ldots$, telle que :*

(a) $x_1 \equiv a \pmod p$ *;*
(b) $f(x_\nu) \equiv 0 \pmod{p^\nu}$ *pour tout* $\nu \geqslant 1$ *;*
(c) $x_{\nu+1} \equiv x_\nu \pmod{p^\nu}$ *pour tout* $\nu \geqslant 1$.

Le théorème au-dessus est un corollaire direct du *lemme de Hensel* :

Lemme 10.2 (le lemme de Hensel)**.** *Soit p un nombre premier, $\nu \geqslant 1$ et $f(x) \in \mathbb{Z}[x]$. Supposons que $f(x_\nu) \equiv 0 \pmod{p^\nu}$ et écrivons $f(x_\nu) = k \cdot p^\nu$ avec $k \in \mathbb{Z}$. S*

Notons que chaque classe de congruence $x_{\nu+1} \pmod{p^{\nu+1}}$ telle que $x_{\nu+1} \equiv x_\nu \pmod{p^\nu}$ peut s'écrire comme $x_{\nu+1} = x_\nu + tp^\nu$ pour quelque $t \in \mathbb{Z}$. Avec cette notation, on a l'équivalence suivante :

$$f(x_{\nu+1}) \equiv 0 \pmod{p^{\nu+1}} \quad \Longleftrightarrow \quad f'(x_\nu) \cdot t \equiv -k \pmod p. \tag{10.3}$$

En particulier, si $f'(x_v) \not\equiv 0 \pmod p$, alors il existe une solution unique t modulo p à la deuxième équation de (10.3). Par la suite, si $f'(x_v) \not\equiv 0 \pmod p$, alors il existe une classe de congruence unique $x_{\nu+1} \pmod{p^{\nu+1}}$ telle que $x_{\nu+1} \equiv x_\nu \pmod{p^\nu}$ et $f(x_{\nu+1}) \equiv 0 \pmod{p^{\nu+1}}$.

Déduction du théorème 10.1 du lemme 10.2. Soit $x_1 \equiv a \pmod p$. Par induction sur ν, le lemme 10.2 implique qu'il existe une suite $x_2, x_3, x_4, \ldots$ telle que $x_\nu \pmod{p^\nu}$ est la classe de congruence unique mod p^v pour laquelle $x_\nu \equiv x_{\nu-1} \pmod{p^{\nu-1}}$ et $f(x_\nu) \equiv 0 \pmod{p^\nu}$. Le résultat en découle tout de suite. □

Démonstration du lemme 10.2. Soit $f(x) = c_d x^d + \cdots + c_1 x + c_0$. Fixons $m \in \{0, 1, \ldots, d\}$. Si $x_{\nu+1} = x_\nu + tp^\nu$, alors on a que

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

puisque $2\nu \geqslant \nu + 1$ lorsque $\nu \geqslant 1$. On en déduit que

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

On a supposé que $f(x_\nu) = p^\nu k$. Par la suite,

$$\begin{aligned} f(x_\nu + tp^\nu) \equiv 0 \,(\mathrm{mod}\, p^{\nu+1}) \quad &\iff \quad tp^\nu f'(x_v) \equiv -kp^\nu \,(\mathrm{mod}\, p^{\nu+1}) \\ &\iff \quad tf'(x_\nu) \equiv -k \,(\mathrm{mod}\, p), \end{aligned}$$

comme affirmé. Finalement, si $f'(x_\nu) \not\equiv 0 \,(\mathrm{mod}\, p)$, alors l'inverse multiplicatif de $f'(x_v)$ existe modulo p, ce qui implique que $t \equiv -(f'(x_\nu))^{-1}k \,(\mathrm{mod}\, p)$ est la seule solution de l'équation $tf'(x_\nu) \equiv -k \,(\mathrm{mod}\, p)$. Ceci conclut la démonstration. □

Exemple. Résolvons l'équation quadratique

$$x^2 - 4x + 8 \equiv 0 \,(\mathrm{mod}\, 25).$$

Soit $f(x) = x^2 - 4x + 8$. Tout d'abord, on résoudre l'équation modulo 5. On peut vérifier directement que les seules solutions sont $1 \,(\mathrm{mod}\, 5)$ et $3 \,(\mathrm{mod}\, 5)$. Pour le voir, on peut aussi utiliser la méthode familière de la *complétion du carré* : on a que

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

Alors

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

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

$$f'(1) \cdot t \equiv -f(1)/5 \,(\mathrm{mod}\, 5) \quad \Leftrightarrow \quad 2t \equiv 1 \,(\mathrm{mod}\, 5) \quad \Leftrightarrow \quad t \equiv 3 \,(\mathrm{mod}\, 5).$$

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

$$f'(3)s \equiv -f(3)/5 \,(\mathrm{mod}\, 5) \quad \Leftrightarrow \quad 2s \equiv -1 \,(\mathrm{mod}\, 5) \quad \Leftrightarrow \quad s \equiv 2 \,(\mathrm{mod}\, 5).$$

Donc $b \equiv 13 \,(\mathrm{mod}\, 25)$.

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

10.2 Les nombres p-adiques

Soit p un nombre premier. Le théorème 10.1 nous amène naturellement à l'étude de suites de résidus $x_\nu \,(\mathrm{mod}\, p^\nu)$ qui satisfont les «conditions de compatibilité»

$$x_{\nu+1} \equiv x_\nu \,(\mathrm{mod}\, p^\nu) \quad \text{pour tout } \nu \geqslant 1. \tag{10.4}$$

Soit $y_\nu \in \{0, 1, 2, \ldots, p^\nu - 1\}$ un représentant de la classe de congruence $x_\nu \,(\mathrm{mod}\, p^\nu)$. Pour tout ν, on peut écrire y_ν dans la base de p. La condition (10.4) implique que :

Affirmation. *Il certains chiffres* $a_0, a_1, \cdots \in \{0, 1, \ldots, p-1\}$ *tels que*

$$\tilde{x}_\nu = a_0 + a_1 p + \cdots + a_{\nu-1} p^{\nu-1} \quad \textit{pour tout } \nu \geqslant 1.$$

Démonstration. Pour tout $\nu \geqslant 1$, il existe chiffres $a_{\nu,j}$ tels que

$$\tilde{x}_\nu = a_{\nu,0} + a_{\nu,1} p + \cdots + a_{\nu,\nu-1} p^{\nu-1}.$$

Soit $j \geqslant 0$. On pose $a_j := a_{j+1,j}$ et on affirme que $a_{\nu,j} = a_j$ pour tout $\nu \geqslant j+1$. On montre cela par induction sur j.

Tout d'abord, on considère le cas de base $j = 0$. La relation (10.4) implique que $\tilde{x}_\nu \equiv \tilde{x}_1 \pmod p$ pour tout $\nu \geqslant 1$. Mais on a aussi que $\tilde{x}_\nu \equiv a_{\nu,0} \pmod p$ et que $\tilde{x}_1 = a_{1,0} = a_0$. Donc, $a_{\nu,0} \equiv a_0 \pmod p$. Puisque $a_{\nu,0}, a_0 \in \{0, 1, \ldots, p-1\}$, on a alors que $a_{\nu,0} = a_0$ pour tout $\nu \geqslant 1$, comme affirmé.

Or, soit $j \geqslant 1$ et supposons qu'on a établit que $a_{\nu,i} = a_i$ quand $0 \leqslant i < j$ et $\nu \geqslant i+1$. Montrons aussi que $a_{\nu,j} = a_j$ pour tout $\nu \geqslant j+1$. En effet, on a que

$$\tilde{x}_\nu \equiv \tilde{x}_{j+1} \pmod{p^{j+1}} \quad \text{pour tout } \nu \geqslant j+1. \tag{10.5}$$

Mais on a aussi que

$$\tilde{x}_\nu \equiv a_{\nu,0} + a_{\nu,1} p + \cdots + a_{\nu,j} p^j \equiv a_0 + a_1 p + \cdots + a_{j-1} p^{j-1} + a_{\nu,j} p^j \pmod{p^{j+1}}$$

de la condition (10.4) et de l'hypothèse d'induction. De même, et puisque $a_{j+1,j} = a_j$ par définition, on a que

$$\tilde{x}_\nu = a_{j+1,0} + a_{j+1,1} p + \cdots + a_{j+1,j} p^j = a_0 + a_1 p + \cdots + a_{j-1} p^{j-1} + a_j p^j.$$

Par la suite, (10.5) implique que $a_{\nu,j} p^j \equiv a_j p^j \pmod{p^{j+1}}$, ce qui veut dire que $a_{\nu,j} \equiv a_j \pmod p$. Puisque $a_{\nu,j}, a_j \in \{0, 1, \ldots, p-1\}$, on a alors que $a_{\nu,j} = a_j$, comme affirmé. Ceci conclut l'étape inductive et donc la démonstration. □

On voit donc que chaque suite de résidus $x_\nu \pmod{p^\nu}$, $\nu = 1, 2, \ldots$, satisfaisant les conditions de compatibilité (10.4) est en correspondance avec une suite de chiffres $a_0, a_1, \cdots \in \{0, 1, \ldots, p-1\}$ qu'on peut encoder dans l'expression formelle

$$a_0 + a_1 p + a_2 p^2 + \cdots$$

Ceci justifie la définition suivante :

Définition 10.3 (entiers p-adiques)**.** Soit p un nombre premier. On définit $\mathbb{Z}_p$ d'être l'ensemble de toutes les suites de résidus $\left(x_\nu \pmod{p^\nu}\right)_{\nu=1}^{\infty}$, satisfaisant les conditions de compatibilité

$$x_{\nu+1} \equiv x_\nu \pmod{p^\nu} \quad \text{pour tout } \nu \geqslant 1. \tag{10.6}$$

Les membres de l'ensemble $\mathbb{Z}_p$ sont appelés *entiers p-adiques*.

Remarque. Parfois, on écrira tout simplement $(x_\nu)_{\nu=1}^{\infty}$ pour un entier p-adique.

Étant donnés deux entiers p-adiques $x = \big(x_\nu \pmod{p^\nu}\big)_{\nu=1}^{\infty}$ et $y = \big(y_\nu \pmod{p^\nu}\big)_{\nu=1}^{\infty}$, on peut définir leur somme

$$x + y := \big(x_\nu + y_\nu \pmod{p^\nu}\big)_{\nu=1}^{\infty}$$

et leur produit

$$xy := \big(x_\nu y_\nu \pmod{p^\nu}\big)_{\nu=1}^{\infty}.$$

Évidemment, ces opérations héritent plusieurs bonnes propriétés des opérations correspondantes de $\mathbb{Z}/p^\nu\mathbb{Z}$ (héritées originellement de $\mathbb{Z}$) : elles sont les deux commutatives et associatives, et la multiplication est distributive par rapport à l'addition. De plus, il existe un élément neutre de l'addition (l'entier p-adique $(0 \pmod{p^\nu})_{\nu=1}^{\infty}$) et un élément neutre de la multiplication (l'entier p-adique $(1 \pmod{p^\nu})_{\nu=1}^{\infty}$). Finalement, chaque entier p-adique $x = \big(x_\nu \pmod{p^\nu}\big)_{\nu=1}^{\infty}$ possède un inverse additif, donné par $-x := \big(-x_\nu \pmod{p^\nu}\big)_{\nu=1}^{\infty}$.

Remarques. (a) La discussion du paragraphe précédent veut dire que l'ensemble $\mathbb{Z}_p$ muni des opérations de l'addition et de la multiplication est un anneau commutatif et unitaire.

(b) On peut plonger $\mathbb{Z}$ dans $\mathbb{Z}_p$: soit $f : \mathbb{Z} \to \mathbb{Z}_p$ l'application qui associe à l'entier n l'entier p-adique $\big(n \pmod{p^\nu}\big)_{\nu=1}^{\infty}$. Il est facile de voir que f est injective et qu'elle respect les opérations d'addition et de multiplication, dans le sens que $f(m+n) = f(m) + f(n)$ et $f(mn) = f(m)f(n)$ pour tous $m, n \in \mathbb{Z}$, ainsi que $f(0) = 0$ et $f(1) = 1$. On dit alors que f un *monomorphisme d'anneaux*.

La discussion ci-dessous implique que $\mathbb{Z}_p$ contient une copie isomorphe de $\mathbb{Z}_p$. Pour cette raison, on peut abuser la notation $0, 1, 2, \dots$ pour dénoter soit l'élément correspondant de $\mathbb{Z}$ ou ce de $\mathbb{Z}_p$. ■

Il existe des éléments de $\mathbb{Z}_p$ qui ne sont pas inversibles dans $\mathbb{Z}_p$. Par exemple, l'élément $(0, 1, 1, \dots)$ n'est pas inversible car $(0, 1, 1, \dots)(x_1, x_2, \dots) = (0, x_2, x_3, \dots) \neq 1$ pour tout $x = (x_\nu)_{\nu=1}^{\infty} \in \mathbb{Z}_p$. Le lemme suivant caractérise tous les éléments inversibles de $\mathbb{Z}_p$. On remarque qu'il ressemble la théorie des fonctions arithmétiques inversibles (cf. théorèmes 5.11 et 5.12).

Lemme 10.4. *Soit p un nombre premier et soit $x = \big(x_\nu \pmod{p^\nu}\big)_{\nu=1}^{\infty} \in \mathbb{Z}_p$. Alors x est inversible si et seulement si $x_1 \not\equiv 0 \pmod{p}$.*

Démonstration. Si x est inversible, il existe $y = \big(y_\nu \pmod{p^\nu}\big)_{\nu=1}^{\infty} \in \mathbb{Z}_p$ tel que $xy = 1$. En particulier, $x_1 y_1 \equiv 1 \pmod{p}$, ce qui implique que $x_1 \not\equiv 0 \pmod{p}$.

Vice versa, supposons que $x_0 \not\equiv 0 \pmod{p}$. Construisons son inverse y par un argument itératif. Il est plus facile de le construire dans la forme $y = a_0 + a_1 p + a_2 p^2 + \cdots$ avec $a_0, a_1, \cdots \in \{0, 1, 2 \dots, p-1\}$. Il faut que $a_0 x_0 \equiv 1 \pmod{p}$, donc on choisit $a_0 \in \{1, 2, \dots, p-1\}$ d'être un représentant de l'inverse de la classe inverse de congruence $x_0 \pmod{p}$, qui existe car $x_0 \not\equiv 0 \pmod{p}$. Puis, fixons $\nu \geqslant 1$ et supposons qu'on a construit $a_0, \dots, a_{\nu-1}$ tels que $x_j(a_0 + a_1 + \cdots + a_{j-1}p^j) \equiv 0 \pmod{p^j}$ pour $j = 1, 2, \dots, \nu$. On veut trouver $a_\nu \in \{0, 1, \dots, p-1\}$ tel que

$$x_{\nu+1}(a_0 + a_1 + \cdots + a_\nu p^\nu) \pmod{p^{\nu+1}}.$$

On a que $x_{\nu+1} \equiv x_\nu \pmod{p^\nu}$, donc il existe un unique $t \in \{0, 1, \dots, p-1\}$ tel que $x_{\nu+1} \equiv x_\nu + tp^\nu \pmod{p^{\nu+1}}$. On pose aussi $y_\nu = a_0 + a_1 p + \cdots + a_{\nu-1} p^{\nu-1}$. On chercher alors $a_\nu \in \{0, 1, \dots, p-1\}$ tel que

$$(x_\nu + tp^\nu)(y_\nu + a_\nu p^\nu) \equiv 1 \pmod{p^{\nu+1}}.$$

On a que

$$\begin{aligned}(x_\nu + tp^\nu)(y_\nu + a_\nu p^\nu) &\equiv x_\nu y_\nu + (a_\nu x_\nu + ty_\nu)p^\nu + ta_\nu p^{2\nu} \pmod{p^{\nu+1}} \\ &\equiv x_\nu y_\nu + (a_\nu x_\nu + ty_\nu)p^\nu \pmod{p^{\nu+1}},\end{aligned}$$

car $2\nu \geqslant \nu + 1$ quand $\nu \geqslant 1$. De plus, on sait que $x_\nu y_\nu \equiv 1 \,(\mathrm{mod}\, p^\nu)$ par l'hypothèse inductive, donc il existe $s \in \{0, 1, \ldots, p-1\}$ tel que $x_\nu y_\nu \equiv 1 + sp^\nu \,(\mathrm{mod}\, p^{\nu+1})$. On trouve alors que

$$(x_\nu + tp^\nu)(y_\nu + a_\nu p^\nu) \equiv 1 + (s + a_\nu x_\nu + ty_\nu)p^\nu \pmod{p^{\nu+1}}.$$

Pour que cette expression soit $1 \,(\mathrm{mod}\, p^{\nu+1})$, il faut que $s + a_\nu x_\nu + ty_\nu \equiv 0 \,(\mathrm{mod}\, p)$ ou, de façon équivalente, que $a_\nu x_\nu \equiv -s - ty_\nu \,(\mathrm{mod}\, p)$. Ici, t, s, x_ν, y_ν sont tous connus. De plus, $x_\nu \equiv x_1 \,(\mathrm{mod}\, p)$, ce qui veut dire que x_ν est inversible mod p car x_1 l'est. Donc, il existe un unique $a_\nu \in \{0, 1, \ldots, p-1\}$ tel que $a_\nu \equiv -\overline{x}_\nu(s + ty_\nu) \,(\mathrm{mod}\, p)$. Ceci termine l'étape inductive de la construction de l'inverse de x. □

Or, soit $x \in \mathbb{Z}_p$ qu'on réalise dans la forme $a_0 + a_1 p + a_2 p^2 + \cdots$ avec $a_0, a_1, \cdots \in \{0, 1, \ldots, p-1\}$. Si $x \neq 0$, alors il existe $k \in \mathbb{Z}_{\geqslant 0}$ minimal tel que $a_k \neq 0$, c'est-à-dire $x = a_k p^k + a_{k+1} p^{k+1} + \cdots$ avec $a_k \neq 0$. L'entier p-adique $y = a_k + a_{k+1} p + a_{k+2} p^2 + \cdots$ est inversible car $a_k \neq 0$. Soit $b_0 + b_1 p + b_2 p^2 + \cdots$ son inverse avec $b_1, b_2, \cdots \in \{0, 1, \ldots, p-1\}$. Donc, on peut formellement imaginer que $x^{-1} = b_0 p^{-k} + b_1 p^{-k+1} + b_2 p^{-k+2} + \cdots$. Ceci nous amène à la définition suivante :

Définition 10.5 (nombre p-adiques)**.** Soit p un nombre premier. On définit $\mathbb{Q}_p$ d'être l'ensemble de toutes les séries formelles de la forme $\sum_{j \geqslant k} a_j p^j$ pour quelque $k \in \mathbb{Z}$ et quelques coefficients $a_j \in \{0, 1, \ldots, p-1\}$ pour tout j. Les membres de l'ensemble $\mathbb{Q}_p$ sont appelés *nombres p-adiques*.

Étant donnés deux nombres p-adiques $x = \sum_{j \geqslant k} a_j p^j$ et $y = \sum_{j \geqslant \ell} b_j p^j$, on peut les ajouter et les multiplier comme suit :

$$x + y = \sum_{j \geqslant \min\{k,\ell\}} (a_j + b_j) p^j$$

avec la convention que $a_j = 0$ pour $j < k$ et que $b_j = 0$ pour $j < \ell$, et

$$xy = \sum_{j \geqslant k+\ell} c_j p^j,$$

où $\sum_{j \geqslant 0} c_{j+k+\ell} p^j$ est le produit dans $\mathbb{Z}_p$ des entiers p-adiques $\sum_{j \geqslant 0} a_{j+k} p^j$ et $\sum_{j \geqslant 0} b_{j+\ell} p^j$.

On peut vérifier par un calcul simple mais long que ces opérations sont les deux commutatives et associatives, et la multiplication est distributive par rapport à l'addition. De plus, il existe un élément neutre de l'addition, et un élément neutre de la multiplication (l'entier p-adique $(1 \,(\mathrm{mod}\, p^\nu))_{\nu=1}^\infty$). On voit aussi facilement que chaque nombre p-adique possède un inverse additif. Finalement, tout nombre p-adique non-zéro possède un inverse multiplicatif. Donc, $\mathbb{Q}_p$ est ce qu'on appelle un *corps*. Donc, il satisfait l'analogue du théorème 7.14.

Théorème 10.6. *Soit p un nombre premier et soit $f(x)$ un polynôme à coefficients dans $\mathbb{Q}_p$. Soient $\alpha_1, \ldots, \alpha_r$ les racines distinctes de $f(x)$ dans $\mathbb{Q}_p$. Alors, il existe quelques nombres naturels $m_1, \ldots, m_r$ et un polynôme $g(x)$ à coefficients dans$\mathbb{Q}_p$ tels que :*

(a) $m_1 + \cdots + m_r \leqslant d$;
(b) $\deg(g) = d - (m_1 + \cdots + m_r)$;
(c) $g(x) \not\equiv 0 \pmod p$ *pour tout* $x \in \mathbb{Z}$;
(d) $f(x) = (x - \alpha_1)^{m_1} \cdots (x - \alpha_r)^{m_r} g(x)$.

En particulier, $r \leqslant \sum_{j=1}^{r} m_j \leqslant d$.

Démonstration. Exercice. □

Exemple. On a considéré avant les racines des polynômes x^2 et $x^2 - 1$ modulo puissances de 2 et on a vu qu'elles ont structure assez compliquée. Par contre, quand on regarde ces deux polynômes sur l'ensemble $\mathbb{Q}_2$, on voit qu'ils ont une structure de racines très simple : x^2 a une double racine en 0, et $x^2 - 1$ a deux simples racines $+1$ et en -1. Ceci correspond au fait que les seules branches infinies des arbres des figures 10.2 et 10.4 sont les suites $\big(0 \pmod{2^\nu}\big)_{\nu=1}^{\infty}$ pour la première, et $\big(1 \pmod{2^\nu}\big)_{\nu=1}^{\infty}$ et $\big(-1 \pmod{2^\nu}\big)_{\nu=1}^{\infty}$ pour la deuxième. ■

10.3 Racines primitives, encore

Dans cette section, on revient à la question d'existence de racines primitives et on montre le résultat suivant :

Théorème 10.7. *Soit n un nombre naturel. Il existe de racines primitives mod n si et seulement si $n \in \{1, 2, 4\} \cup \{p^\nu : p > 2$ premier, $\nu \geqslant 1\} \cup \{2p^\nu : p > 2$ premier, $\nu \geqslant 1\}$.*

La démonstration de ce théorème se base de façon cruciale au résultat suivant :

Théorème 10.8. *Soit p un nombre premier et soit ν un nombre naturel. On a que*

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

Démonstration. D'abord, on considère le cas où $p > 2$. Dans ce cas-ci, il faut montrer que $\lambda(p^\nu) = \phi(p^\nu)$. D'après le théorème 8.7, ceci est équivalent à l'existence de racines primitives mod p^ν. On distingue trois cas :

Cas 1 : $\nu = 1$. Ici, on peut tout simplement utiliser le corollaire 8.9.

Cas 2 : $\nu = 2$. Soit g une racine primitive mod p. Si $x \pmod{p^2}$ tel que $x \equiv g \pmod p$ et on pose $k = \operatorname{ord}_{p^2}(x)$, 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 $\operatorname{ord}_p(g) = p - 1 | k$. 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 $\operatorname{ord}_{p^2}(x) = p(p-1)$. 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 $p \geqslant 2$ classes d'équivalence modulo $x \pmod{p^2}$

telles que $x \equiv g \pmod p$. Donc il existe exactement $p-1 \geqslant 1$ 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.

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

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

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

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

Finalement, supposons que $p = 2$. Si $\nu \in \{1, 2\}$, c'est facile de vérifier qu'il existe une racine primitive mod 2^ν. Donc on a que $\lambda(2^\nu) = \phi(2^\nu)$ dans ces deux cas.

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

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

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

ce qui prouve notre affirmation.

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

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

puisque $2 \nmid b$. Par conséquent, $\mathrm{ord}_{2^{\nu+1}}(g) = 2^{\nu-1}$, comme affirmé. Ceci implique que $\lambda(2^{\nu+1}) \geqslant 2^{\nu-1}$ selon le théorème 8.7, ce qui conclut la démonstration du théorème. □

Remarque 10.9. La démonstration du théorème 10.8 dans le cas de moduli 2^ν avec $\nu \geqslant 3$ nous permet de déterminer un g explicit d'ordre maximale $2^{\nu-2}$: on a que $\mathrm{ord}_8(5) = 2 = 2^{3-2}$, et que $\mathrm{ord}_{16}(5) = 4 = 2^{4-2}$. Donc il a ordre $2^{\nu-2}$ mod 2^ν, pour tout $\nu \geqslant 3$, d'après la démonstration.

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

Démonstration du théorème 10.7. 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 9.7 et du théorème 10.8. 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 $v_1, v_2 \geqslant 1$ 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 9.6 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 8.7, on trouve que

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

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

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

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

car $2|\lambda(2^v)$ et $2|p-1$ du théorème 10.8. Donc $n \notin \mathcal{N}$.

Finalement, si $n = 2^v$ avec $v \geqslant 3$, alors le théorème 10.8 implique que $\lambda(2^v) = 2^{v-2} < \phi(2^v)$ et donc $n \notin \mathcal{N}$. □

10.4 Exercices

EXERCICE 10.1. Trouver toutes les solutions aux équations polynomiales 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) $x^2 - x + 10 \equiv 0 \pmod{50}$
(e) $x^2 - 3x + 3 \equiv 0 \pmod{343}$.
(f) $x^2 + 7x - 9 \equiv 0 \pmod{63}$.
(g) $x^{12} + 2x^{11} - 3x + 9 \equiv 0 \pmod{121}$. [*Indice :* Diviser le polynôme par $x^{11} - x$ (cf. petit théorème de Fermat).]
(h) $x^3 - 30x^2 + 5x - 2 \equiv 0 \pmod{169}$.
(i) $x^2 - 5x - 4 \equiv 0 \pmod{169}$.

EXERCICE 10.2. Dans le lemme de Hensel, on étudie des racines x_ν de l'équation polynomiale $f(x) \equiv 0 \pmod{p^\nu}$ avec $p \nmid f'(x_\nu)$. Décrire 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.

EXERCICE 10.3. Soit p un nombre premier et $\nu \geqslant 1$. Montrez que si g est une racine primitive mod $p^{\nu+1}$, alors il est une racine primitive mod p^ν.

EXERCICE 10.4. Quand $\nu \geqslant 3$, montrer que

$$(\mathbb{Z}/2^\nu\mathbb{Z})^* = \left\{5^k \pmod{2^\nu} : 0 \leqslant k < 2^{\nu-2}\right\} \cup \left\{-5^k \pmod{2^\nu} : 0 \leqslant k < 2^{\nu-2}\right\}.$$

[*Indice :* voir le remarque 10.9.]

EXERCICE 10.5. Trouver un nombre g qui est une racine primitive mod 5^ν, pour chaque $\nu \geqslant 1$. Faire la même chose mod 7^ν et mod 11^ν.

Chapitre 11

Résidus quadratiques

11.1 Équations quadratiques mod p

Dans le chapitre 10, on a vu comment réduire la résolution d'équations polynomiales mod p^ν au cas où $\nu = 1$. La compréhension de ce cas spécial pour de polynômes généraux est un problème très difficile, qui est partiellement responsable pour le développement du domaine de la *théorie des nombres algébriques* et de la *théorie des corps de classes*. Ici, on se concentre sur le cas de polynômes quadratiques qui est déjà très riche et qui nous amènera à un résultat surprenant et profond de Gauss.

Soit p un nombre premier, et soient $a, b \in \mathbb{Z}$. On veut trouver tous les entiers x qui satisfont l'équation quadratique

$$x^2 + ax + b \equiv 0 \pmod{p}. \tag{11.1}$$

Quand $p = 2$, il est facile de trouver toutes les solutions de (11.1) : si $a \equiv 0 \pmod 2$, alors $x \equiv b \pmod 2$; si $a \equiv 1 \pmod 2$ et $b \equiv 0 \pmod 2$, alors tous les entiers x satisfont (11.1) ; et, finalement, si $a \equiv b \equiv 1 \pmod 2$, il n'y a aucune solution de (11.1).

Or, supposons que $p > 2$. En particulier, $(2, p) = 1$, donc 2 est inversible mod p. Ceci nous permet d'imiter la résolution d'équations quadratiques sur $\mathbb{R}$. La première étape dans cette théorie est de «compléter le carré», qu'on peut effecteur grâce à l'inversibilité de 2 mod p :

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

où $D := a^2 - 4b$ est le discriminant du polynôme $x^2 + ax + b$. Par conséquent, l'équation (11.1) est équivalent à l'équation

$$(x + \overline{2}a)^2 \equiv \overline{4}D \pmod{p}.$$

Il existe alors deux cas :

Cas 1 : il existe $r \pmod p$ tel que $D \equiv r^2 \pmod p$. Dans ce cas-ci, on a que $(x + \overline{2}a)^2 \equiv (\overline{2}r)^2 \pmod p$, et donc le lemme 7.12 implique que $x + \overline{2}a \equiv \pm\overline{2}r \pmod{p}$. On trouve alors que x satisfait (11.1) si et seulement si

$$x \equiv \overline{2} \cdot (-a \pm r) \pmod{p}.$$

Si $D \equiv 0 \pmod p$, il faut que $r \equiv 0 \pmod p$, donc il existe une seule solution mod p à (11.1), donnée par $x \equiv -\overline{2}a \pmod p$. D'autre côté, si $D \not\equiv 0 \pmod p$, alors $r \not\equiv -r \pmod p$, ce qui veut dire qu'il existe deux solutions distinctes mod p.

Cas 2 : il n'existe pas de $r \pmod p$ tel que $D \equiv r^2 \pmod p$. Dans ce cas-ci, l'équation (11.1) n'a aucune solution.

En conclusion, on est arrivé à un résultat analogue du théorème familier concernant la résolution $x^2 + ax + b = 0$ avec $x \in \mathbb{R}$.

11.2 Le symbole de Legendre

La discussion de la section précédente nous amène naturellement à la définition suivante :

Définition 11.1 (symbole de Legendre)**.** Soit p un nombre premier et soit a un entier. On définit *le symbole de Legendre*

$$(a|p) = \left(\frac{a}{p}\right) := \begin{cases} 0 & \text{si } a \equiv 0 \pmod p, \\ 1 & \text{si } a \not\equiv 0 \pmod p \text{ et il existe } r \pmod p \text{ tel que } a \equiv r^2 \pmod p, \\ -1 & \text{sinon.} \end{cases}$$

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

Avec cette notation, la discussion de la section précédente implique le théorème suivant :

Théorème 11.2. *Soit $p > 2$ un nombre premier, soient $a, b \in \mathbb{Z}$, et soit $D = a^2 - 4b$ le discriminant du polynôme $x^2 + ax + b$. Alors, on a que*

$$\#\big\{x \pmod p : x^2 + ax + b \equiv 0 \pmod p\big\} = 1 + \left(\frac{D}{p}\right).$$

Afin de rendre pratique ce théorème, il faut développer une méthode qui nous dit quand un résidu réduit $a \pmod p$ est un résidu ou un non-résidu quadratique. On commence avec le lemme de base suivant.

Lemme 11.3. *Soit $p > 2$ un nombre premier et soit $p_0 = (p-1)/2$.*

(a) On a que

$$\big\{a \pmod p : (a|p) = 1\big\} = \big\{j^2 \pmod p : 1 \leqslant j \leqslant p_0\big\}$$

(b) On a que

$$\#\big\{a \pmod p : (a|p) = 1\big\} = \#\big\{a \pmod p : (a|p) = -1\big\} = p_0.$$

(c) Si $(a|p) = 1$, alors $(\overline{a}|p) = 1$, où $\overline{a}$ dénote un représentant de la classe inverse de $a \pmod p$.

(d) Le multiple de deux résidus quadratiques est un résidu quadratique ; le multiple d'un résidu quadratique avec un non-résidu quadratique est un non-résidu quadratique ; le multiple de deux non-résidus quadratiques est un résidu quadratique.

Démonstration. (a) Si $a \equiv j^2 \pmod p$ avec $1 \leqslant j \leqslant p_0$, alors c'est clair que $(a|p) = 1$. Vice versa, supposons que $(a|p) = 1$. Alors, $p \nmid a$ et il existe $r \in \mathbb{Z}$ tel que $a \equiv r^2 \pmod p$. D'après le corollaire 7.6, on peut supposer que $|r| \leqslant p_0$. De plus, on a que $r \neq 0$. Par la suite, il existe $j \in \{1, \ldots, p_0\}$ tel que $r \in \{j, -j\}$. Dans tout cas, $r^2 \equiv j^2 \pmod p$, et donc $a \equiv j^2 \pmod p$, comme désiré.

(b) On pose $\mathcal{R} = \{a \pmod p : (a|p) = 1\}$ et $\mathcal{N} = \{a \pmod p : (a|p) = -1\}$. Il existe $p - 1 = 2p_0$ résidus réduits $a \pmod p$ et pour chacun on a $(a|p) \in \{-1, +1\}$. Donc, $|\mathcal{R}| + |\mathcal{N}| = 2p_0$. Il suffit alors de montrer que $|\mathcal{R}| = p_0$. Selon la partie (a), ceci sera vrai si on peut montrer que les résidus $j^2 \pmod p$ avec $1 \leqslant j \leqslant p_0$ sont distincts. En effet, si $i^2 \equiv j^2 \pmod p$, alors le lemme 7.12 implique que $i \equiv \pm j \pmod p$. Si $1 \leqslant i, j \leqslant p_0$, alors la condition $i \equiv \pm j \pmod p$ implique que $i = j$. Ceci termine la preuve de la partie (b).

(c) Si $(a|p) = 1$, alors $a \equiv r^2 \pmod p$ pour quelque $r \in \mathbb{Z}$ avec $p \nmid r$. Donc, r est inversible mod p. On peut aussi vérifier que $\overline{r^2} \equiv \overline{r}^2 \pmod p$ (c'est-à-dire, l'inverse de r^2 mod p est égal à l'inverse de r mod p au carré), car $r^2 \cdot (\overline{r}^2) \equiv (r\overline{r})^2 \equiv 1^2 \equiv 1 \pmod p$. Donc, $\overline{a} \equiv \overline{r^2} \equiv \overline{r}^2 \pmod p$, ce qui montre que $(\overline{a}|p) = 1$.

(d) Si $a, b \in \mathcal{R}$, alors $a \equiv r^2 \pmod p$ et $b \equiv s^2 \pmod p$ avec $p \nmid r, s$. Donc, $ab \equiv (rs)^2 \pmod p$ et $p \nmid rs$, ce qui montre que $xy \in \mathcal{R}$, comme voulu.

Puis, on considère le cas où $(a|p) = 1$ et $(b|p) = -1$. On affirme que $(ab|p) = -1$. On sait que $p \nmid a, b$, donc $p \nmid ab$. Alors, si $(ab|p) \neq -1$, il faut que $(ab|p) = 1$. Mais, on sait aussi que $(\overline{a}|p) = 1$ d'après la partie (c) et notre hypothèse que $(a|p) = 1$. Donc, on trouve que $b \equiv \overline{a} \cdot (ab) \pmod p$ est équivalent mod p au produit de deux résidus quadratiques. L'argument du paragraphe précedent montre alors que $(b|p) = 1$, ce qui est impossible. Il faut donc que $(ab|p) = -1$, comme on l'a affirmé.

Finalement, on suppose que $(a|p) = (b|p) = -1$. On remarque que $(a, p) = 1$. Donc, l'application $f : (\mathbb{Z}/p\mathbb{Z})^* \to (\mathbb{Z}/p\mathbb{Z})^*$, définie par $f(x \pmod p) := ax \pmod p$, est une bijection (cf. exercice 7.4).

Or, on rappelle les ensembles $\mathcal{R}$ et $\mathcal{N}$, définis à la partie (b). D'après la deuxième affirmation de la partie (d), qu'on a déjà montré ci-dessus, on a que $f(\mathcal{R}) \subseteq \mathcal{N}$. Puisque f est injective, on a aussi que $|f(\mathcal{R})| = |\mathcal{R}|$. D'autre côté, la partie (b) implique que $|\mathcal{R}| = |\mathcal{N}|$. Par la suite, $|f(\mathcal{R})| = |\mathcal{N}|$. Puisque on a aussi que $f(\mathcal{R}) \subseteq \mathcal{N}$, il faut que

$$f(\mathcal{R}) = \mathcal{N}.$$

On affirme que ceci implique aussi que

$$f(\mathcal{N}) = \mathcal{R}.$$

En effet, puisque f est une bijection, on a que

$$f(\mathcal{R} \sqcup \mathcal{N}) = f(\mathcal{R}) \sqcup f(\mathcal{N}).$$

On a aussi que $\mathcal{R} \sqcup \mathcal{N} = (\mathbb{Z}/p\mathbb{Z})^*$, et donc

$$f(\mathcal{R} \sqcup \mathcal{N}) = f((\mathbb{Z}/p\mathbb{Z})^*) = (\mathbb{Z}/p\mathbb{Z})^* = \mathcal{R} \sqcup \mathcal{N}.$$

On en déduit que $f(\mathcal{R}) \sqcup f(\mathcal{N}) = \mathcal{R} \sqcup \mathcal{N}$. Puisque $f(\mathcal{R}) = \mathcal{N}$, il faut que $f(\mathcal{N}) = \mathcal{R}$, comme affirmé. En particulier, $f(b \pmod p) = ab \pmod p$ est dans $\mathcal{R}$ de notre hypothèse que $(b|p) = -1$. Ceci termine la preuve. □

Un corollaire facile du lemme 11.3 est le résultat suivant.

Corollaire 11.4. *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.

11.3 Calcul du symbole de Legendre : préliminaires

Puisque le symbole de Legendre est une fonction complètement multiplicative, le théorème 11.4 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 valeur de $(-1|p)$ et de $(q|p)$ pour chaque nombre premier q. On calcul d'abord $(-1|p)$ et $(2|p)$. La valeur de $(q|p)$, où q est un nombre premier impair, sera discutée dans la section suivante.

On commence avec le lemme suivant.

Lemme 11.5 (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. Soit $p_0 = (p-1)/2$. Le lemme est trivial quand $p|a$, donc supposons que $p \nmid a$. Le lemme est aussi facile à montrer si $(a|p) = 1$, car dans ce cas-ci on a que $a \equiv r^2 \pmod{p}$ pour quelque $r \not\equiv 0 \pmod{p}$. Donc, le petit théorème de Fermat (cf. corollaire 8.4) implique que $a^{p_0} \equiv r^{2p_0} \equiv 1 \pmod{p}$.

Il reste à considérer le cas où $(a|p) = -1$. On le traite de deux façon différentes.

Première méthode : Soit g une racine primitive mod p, et soit $h = g^{p_0}$. On a que $h^2 \equiv 1 \pmod{p}$, mais que $h \not\equiv 1 \pmod{p}$. Donc $h \equiv -1 \pmod{p}$. Puis, on écrit a en termes de g : il existe $k \in \mathbb{Z}_{\geqslant 0}$ tel que $a \equiv g^k \pmod{p}$. Puisque $(a|p) = -1$, il faut que k soit pair. On peut alors l'écrire comme $k = 2\ell + 1$, d'où

$$a^{p_0} \equiv (g^{2\ell+1})^{p_0} \equiv (g^{p-1})^{\ell} \cdot g^{p_0} \equiv -1 \pmod{p},$$

ce qui termine la démonstration.

Deuxième méthode : Le théorème 7.15 de Wilson dit que $(p-1)! \equiv -1 \pmod{p}$. D'autre côté, le lemme 11.3 implique que les $p-1$ nombres $1^2, 2^2, \ldots, p_0^2, a\cdot 1^2, a\cdot 2^2, \ldots, a\cdot p_0^2$ sont un système complet de tous les résidus réduits mod p. Donc,

$$\left(\prod_{j=1}^{p_0} j^2\right)\left(\prod_{j=1}^{p_0}(aj^2)\right) \equiv (p-1) \equiv -1 \pmod{p}.$$

Le côté droit est égal à $a^{p_0}p_0!^4$, donc on déduit que

$$a^{p_0}p_0!^4 \equiv -1 \pmod{p}.$$

Donc, il reste à montrer que $p_0!^4 \equiv 1$. En effet, les nombres $1, 2, \ldots, p_0, -1, -2, \ldots, -p_0$ sont un système complet de tous les résidus réduits mod p. Donc,

$$\left(\prod_{j=1}^{p_0} j\right)\left(\prod_{j=1}^{p_0}(-j)\right) \equiv (p-1) \equiv -1 \pmod{p}.$$

Le côté droit est égal à $(-1)^{p_0}p_0!^2$, d'où on déduit que

$$p_0!^2 \equiv (-1)^{p_0+1} = (-1)^{\frac{p+1}{2}} \pmod{p}. \tag{11.2}$$

Dans tous les cas, on a que $p_0!^4 \equiv 1 \pmod{p}$ comme voulu. Ceci conclut la preuve. □

En utilisant le critère d'Euler quand $a = -1$, on a le résultat suivant :

Corollaire 11.6. *Si $p > 2$ un nombre premier, alors on a que*

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

Puis, on calcul $(2|p)$ par une variation de la deuxième méthode menant au lemme 11.5.

Théorème 11.7. *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

$$\prod_{\substack{1 \leqslant m \leqslant p-1 \\ 2|m}} m = \prod_{j=1}^{p_0}(2j) = 2^{p_0}p_0!. \tag{11.3}$$

Puis, on réécrit le produit au côté gauche. Les nombres pairs $2j \in [p_0+1, p-1]$ sont en correspondance avec les nombres impairs $2i - 1 \in [1, p_0]$. En effet, si $2j \in [p_0+1, p-1]$, alors $p - 2j$ est un nombre impair qui appartient à $[1, p_0]$, et vice-versa. Il existe $j_0 := \lfloor (p_0+1)/2 \rfloor = \lfloor (p+1)/4 \rfloor$ nombres impairs dans $[1, p_0]$, donc il existe j_0 nombres pairs dans $[p_0+1, p-1]$. Par conséquent,

$$\prod_{\substack{p_0 < m \leqslant p-1 \\ 2|m}} m \equiv (-1)^{j_0} \prod_{\substack{p_0 < m \leqslant p-1 \\ 2|m}} (p-m) \equiv (-1)^{j_0} \prod_{\substack{n \leqslant p_0 \\ 2 \nmid n}} n \pmod{p}.$$

Par la suite,

$$\prod_{\substack{1\leqslant m\leqslant p-1\\ 2|m}} m \equiv \Bigg(\prod_{\substack{m\leqslant p_0\\ 2|m}} m\Bigg)\Bigg(\prod_{\substack{p_0<m\leqslant p-1\\ 2|m}} m\Bigg) \pmod p$$

$$\equiv (-1)^{j_0}\Bigg(\prod_{\substack{m\leqslant p_0\\ 2|m}} m\Bigg)\Bigg(\prod_{\substack{m\leqslant p_0\\ 2\nmid m}} m\Bigg) \pmod p$$

$$\equiv (-1)^{j_0} p_0! \pmod p. \tag{11.4}$$

En comparant les relations (11.3) et (11.4), et puisque $(p_0!, p) = 1$, on trouve que

$$2^{p_0} \equiv (-1)^{j_0} \pmod p.$$

Par conséquent, le critère d'Euler nous donne que $(2|p) \equiv (-1)^{j_0} \pmod p$. Puisque les nombres $(2|p)$ et $(-1)^{j_0}$ prennent que les valeurs ± 1, la dernière congruence est, en fait, une égalité. Le résultat affirmé suit d'une étude du signe de $(-1)^{j_0}$ selon la classe de congruence de p mod 8. □

11.4 Réciprocité quadratique

Selon la section 11.3, il reste à déterminer $(q|p)$ quand q est un nombre premier impair diffèrent que p. Gauss a découvert un résultat très profond et surprenant qui relie deux objets qui sont complètement indépendants à la première vue : les symboles de Legendre $(q|p)$ et $(p|q)$.

Théorème 11.8 (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 étudie comment il peut nous aider de calculer le symbole de Legendre. Par exemple, calculons $(3|p)$. 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 11.9. *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. Gauss en a découvert huit ! On va présenter deux preuves ici.

La première preuve est grâce à Gauss soi-même. On peut voir l'idée principale plus facilement que dans d'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 11.7. Rappelez que dans cette dernière preuve, on a montré que $(2|p) = (-1)^n$, où $n = \#\{j \in \mathbb{N} : (p-1)/2 < 2j \leqslant p-1\}$. On généralise maintenant ce résultat.

Lemme 11.10. *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 $\{x \pmod p : (p-1)/2 < x \leqslant p-1\}$.*

(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

$$N_p(a) = \#\{1 \leqslant j \leqslant p_0 : r_j < 0\}.$$

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, $|i - \varepsilon j| \leqslant i + j \leqslant 2p_0 = p-1$, ce qui implique $i - \varepsilon j = 0$. Puisque $1 \leqslant i,j \leqslant p_0$ 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. Donc, il en existe exactement p_0 tels nombres distincts. Mais, tous ces nombres appartiennent tous à $\{1,2,\dots,p_0\}$. Par conséquent,

$$\{|r_1|, \dots, |r_{p_0}|\} = \{1,\dots,p_0\}. \tag{11.5}$$

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 que 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

$$s_j = \begin{cases} r_j & \text{si } 0 \leqslant r_j \leqslant p_0, \\ p + r_j & \text{si } -p_0 < r_j < 0. \end{cases}$$

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 (11.5). 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 $0 \leqslant s_j < p$, 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 11.8. 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

$$\begin{aligned} \sum_{j=1}^{p_0} \left\lfloor \frac{qj}{p} \right\rfloor = \sum_{j=1}^{p_0} \sum_{1 \leqslant k \leqslant qj/p} 1 &= \#\{(j,k) \in \mathbb{N}^2 : j \leqslant p_0,\ k \leqslant qj/p\} \\ &= \#\{(j,k) \in \mathbb{N}^2 : j \leqslant p_0,\ k \leqslant q_0,\ pk \leqslant qj\}. \end{aligned}$$

De même,

$$\begin{aligned} \sum_{k=1}^{q_0} \left\lfloor \frac{pk}{q} \right\rfloor = \sum_{k=1}^{q_0} \sum_{1 \leqslant j \leqslant pk/q} 1 &= \#\{(j,k) \in \mathbb{N}^2 : k \leqslant q_0,\ j \leqslant pk/q\} \\ &= \#\{(j,k) \in \mathbb{N}^2 : j \leqslant p_0,\ k \leqslant q_0,\ pk \geqslant qj\}. \end{aligned}$$

Puisque il n'y a pas de pair $(j,k) \in \mathbb{N}^2$ avec $1 \leqslant j \leqslant p_0$, $1 \leqslant k \leqslant q_0$ et $pk = qj$ (sinon, on trouve que $p|j$ et donc $j \geqslant p$, une contradiction), alors

$$\sum_{j=1}^{p_0} \left\lfloor \frac{qj}{p} \right\rfloor + \sum_{k=1}^{q_0} \left\lfloor \frac{pk}{q} \right\rfloor = \#\{(j,k) \in \mathbb{N}^2 : j \leqslant p_0,\ k \leqslant q_0\} = p_0 q_0.$$

En combinant cette relation avec le lemme 11.10, 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. □

On conclut ce chapitre en présentant une preuve différente de la loi de réciprocité quadratique grâce à George Rousseau [8]. [1] L'avantage de cette preuve est que son idée de base est beaucoup plus facile à comprendre et à mémoriser.

Démonstration alternative du théorème 11.8. Soit $p_0 = (p-1)/2$ et $q_0 = (q-1)/2$. On considère le produit directe $G := (\mathbb{Z}/p\mathbb{Z})^* \times (\mathbb{Z}/q\mathbb{Z})^*$. On va construire trois ensembles $A, B, C \subset G$ qui contient exactement un moitié de G, i.e. pour tout $g \in G$, ils contient soit g ou $-g$, mais pas les deux. On choisit

$$A = \{i \pmod p : 1 \leqslant i \leqslant p_0\} \times \mathbb{Z}/q\mathbb{Z}^*$$

et

$$B = (\mathbb{Z}/p\mathbb{Z})^* \times \{j \pmod q : 1 \leqslant j \leqslant q_0\}.$$

Finalement, C est l'ensemble de tous les pairs $(a \pmod p, b \pmod q) \in G$ pour lesquels il existe un entier $k \in [1, pq/2)$ tel que $k \equiv a \pmod p$ et $k \equiv b \pmod q$ (cf. théorème des restes chinois).

Par construction, les produits $\prod_{a\in A} a$, $\prod_{b\in B} b$ et $\prod_{c\in C} c$ différent par quelques signes. On les calculent exactement. Avec un petit abus de notation, on écrira (x,y) dans ce calcul pour dénoter le pair $(x \pmod p, y \pmod p)$.

Tout d'abord, on a que

$$P_A := \prod_{a\in A} a \equiv \prod_{1\leqslant i\leqslant p_0} \prod_{1\leqslant j\leqslant q-1} (i,j) \equiv \prod_{1\leqslant i\leqslant p_0} \left(i^{q-1}, (q-1)!\right) \equiv \left(p_0!^{q-1}, (q-1)!^{p_0}\right).$$

On a $q - 1 = 2q_0$ et $p_0!^2 \equiv (-1)^{p_0+1} \pmod p$ par (11.2). De plus, $(q-1)! \equiv -1 \pmod q$ par le théorème 7.15 de Wilson. On en déduit que

$$P_A \equiv \left((-1)^{p_0 q_0 + q_0}, (-1)^{p_0}\right). \tag{11.6}$$

De même, on trouve que

$$P_B := \prod_{b\in B} b \equiv ((p-1)!^{q_0}, q_0!^{p-1}) \equiv \left((-1)^{q_0}, (-1)^{p_0 q_0 + p_0}\right). \tag{11.7}$$

1. La preuve est aussi décrite au lien suivant : https://mathoverflow.net/questions/1420/whats-the-best-proof-of-quadratic-reciprocity.

En particulier,

$$P_A = (-1)^{p_0 q_0} P_B. \tag{11.8}$$

Finalement, calculons $P_C := \prod_{c \in C} c$. Le théorème des restes chinois implique que

$$P_C \equiv \big(K \pmod p, K \pmod q\big) \quad \text{avec} \quad K := \prod_{\substack{1 \leqslant k < pq/2 \\ p,q \nmid k}} k.$$

On calcul d'abord $K \pmod p$. On a que

$$K = \Big(\prod_{\substack{1 \leqslant k < pq/2 \\ p \nmid k}} k \Big) \Big/ \Big(\prod_{\substack{1 \leqslant k < pq/2 \\ q \mid k}} k \Big).$$

De plus,

$$\prod_{\substack{1 \leqslant k < pq/2 \\ q \mid k}} k = \prod_{1 \leqslant m < p/2} (qm) = q^{p_0} p_0! \equiv \left(\frac{q}{p}\right) p_0! \pmod p$$

d'après le critère d'Euler. Puis, on observe que $k < pq/2$ si et seulement si $k \leqslant (pq-1)/2 = q_0 p + p_0$. Donc,

$$\begin{aligned} \prod_{\substack{1 \leqslant k < pq/2 \\ p \nmid k}} k &= \Big[\prod_{j=1}^{q_0} \Big(\prod_{(j-1)p < k < jp} k \Big) \Big] \prod_{q_0 p < k \leqslant q_0 p + p_0} k \\ &\equiv (p-1)!^{q_0} p_0! \pmod p \\ &\equiv (-1)^{q_0} p_0! \pmod p \end{aligned}$$

d'après le théorème de Wilson. En conclusion, on a que $K \equiv (-1)^{p_0}(q|p) \pmod p$. De même, on trouve aussi que $K \equiv (-1)^{q_0}(p|q) \pmod q$. Par conséquent,

$$P_C \equiv \left((-1)^{p_0} \left(\frac{q}{p}\right), (-1)^{q_0} \left(\frac{p}{q}\right) \right). \tag{11.9}$$

Mais on sait que $P_C = \varepsilon P_A$ et $P_C = \varepsilon' P_B$ pour deux signes $\varepsilon, \varepsilon' \in \{-1, 1\}$. En comparant (11.6) et (11.9), on voit que $\varepsilon = (q|p)$. Puis, En comparant (11.7) et (11.9), on voit que $\varepsilon' = (p|q)$. Donc,

$$P_A = \left(\frac{q}{p}\right) P_C = \left(\frac{q}{p}\right)\left(\frac{p}{q}\right) P_B.$$

En combinant cette relation avec (11.8) termine la preuve du théorème 11.8. □

11.5 Exercices

EXERCICE 11.1. Soit $p > 3$ un nombre premier. Calculer $(3|p)$ en utilisant l'idée en arrière de la démonstration du théorème 11.7. [*Indice* : Considérer le produit $P = \prod_{j \leqslant p_0}(3j)$, où $p_0 = (p-1)/2$.]

EXERCICE 11.2 ($*$). 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 11.3. Résolvez l'équation $x^2 - 20x + 139 \equiv 0 \pmod{1583}$.

EXERCICE 11.4. Si $p > 2$ est un nombre premier, montrez que

$$\sum_{a=1}^{p} \left(\frac{a}{p} \right) = 0.$$

Troisième partie

Équations diophantiennes

Chapitre 12

Équations avec solutions paramétriques

12.1 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

$$ax + by = c? \tag{12.1}$$

Tout d'abord, on observe que si l'équation (12.1) 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 (12.1) a de solutions. En effet, théorème 2.6 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 (12.1). 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 (12.1). 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 (12.1). Réciproquement, on montrera que si $(x, y) \in \mathbb{Z}^2$ est une autre solution à (12.1), 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'Euclide 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 12.1. *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 $ax_0 + by_0 = c$.*

12.2 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

$$x^2 + y^2 = z^2. \tag{12.2}$$

Réciproquement, si les nombres x, y et z satisfassent l'équation (12.2), 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 (12.2). 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 (12.2) 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

$$(x, y, z) = (u^2 - v^2, 2uv, u^2 + v^2), \tag{12.3}$$

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 12.2. *On a que*

$$\{(x, y, z) \textit{ triplet pythagoricien primitif} : 2|y\} = \{(u^2 - v^2, 2uv, u^2 + v^2) : 2|uv,\ (u, v) = 1\}.$$

Chapitre 13

Équations diophantiennes insolubles

13.1 Solutions globales et locales

Comme on l'a vu au dernier chapitre, 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

$$x^2 = -1 - y^2 \tag{13.1}$$

n'a pas de solutions réelles, puisque le coté gauche est toujours $\geqslant 0$ et le coté droit est toujours $\leqslant -1$. *A fortiori*, elle n'a pas de solutions entières.

Un autre exemple de nature différente est l'équation

$$x^2 + 3y^2 = 1570. \tag{13.2}$$

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 (13.2) 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 $n \geqslant 3$. 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

$$x^3 + 3y^3 = 9z^3. \tag{13.3}$$

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

$$0 < \max\{|x_1|, |y_1|z_1|\} < \max\{|x_0|, |y_0|, |z_0|\}. \tag{13.4}$$

En itérant cette procedure, on peut construire une infinité de triplets distingues $(x, y, z) \in \mathbb{Z}^3$ dont tous les cordonnés sont $\leqslant \max\{|x_0|, |y_0|, |z_0|\}$ 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 (13.3) 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 (13.4), et la construction est complète.

13.2 Le dernier théorème de Fermat

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 13.1. *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, $|z/d^2| \geqslant |z|$. 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 12.2. 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 12.2 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 $A = \emptyset$. □

Chapitre 14

Représentation des entiers par des polynômes

14.1 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 14.1. *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 $u \neq 0$ $|1/u| \geqslant 1$ et, par la suite, $|u| \leqslant 1$. Les seuls entiers gaussiens $u \neq 0$ qui satisfont cette inégalité sont les nombres $\pm 1, \pm i$.

Définition 14.2. 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 14.3. *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 14.3, 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 11.6). 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 14.4. *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 évident. 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 1.3) est le lemme 2.9 d'Euclide. Puis, la clé dans la démonstration du lemme d'Euclide est la division euclidienne. On pourrait alors montré l'analogue de la division euclidienne dans $\mathbb{Z}[i]$. On a le théorème suivant :

Théorème 14.5. *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 $|x-a|, |y-b| \leqslant 1/2$. On pose $q = a+ib$, pour que $|z/w - q|^2 = |x-a|^2 + |y-b|^2 \leqslant 1/4 + 1/4 = 1/2$. Donc $r := z - qw$ a magnitude $|r| \leqslant |w|/\sqrt{2} < |w|$. □

En utilisant le théorème 14.5, 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'Euclide 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 14.6. *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

$$m = \prod_{\substack{1 \leqslant i \leqslant r \\ p_i = 2 \text{ ou } p_i \equiv 1 \pmod 4}} p_i.$$

Du théorème 14.4, 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 14.1 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 14.4 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 14.4. 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 $\lfloor x \rfloor \leqslant x < \lfloor x \rfloor + 1$, 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

$$X = \{(a, b) \in \mathbb{Z}^2 : 0 \leqslant a, b \leqslant M\}.$$

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 $-M \leqslant a' - a \leqslant M$ et $-M \leqslant b' - b \leqslant M$, ce qui implique que $m \leqslant 2M^2 < 2p$. 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. □

14.2 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 14.7 (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 14.8. *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

$$\begin{aligned} ab = {} & (x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4)^2 + (-x_1y_2 + x_2y_1 - x_3y_4 + x_4y_3)^2 \\ & + (-x_1y_3 + x_3y_1 + x_2y_4 - x_4y_2)^2 + (-x_1y_4 + x_1y_4 - x_2y_3 + x_3y_2)^2, \end{aligned} \tag{14.1}$$

d'où le lemme découle tout de suite. □

Démonstration du théorème 14.7. Le lemme 14.8 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 14.4. 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 14.4). On considère l'ensemble

$$A := \{a = x_1^2 + x_2^2 + x_3^2 + x_4^2 : x_i \in \mathbb{Z}\ (1 \leqslant i \leqslant 4),\ p|a,\ a > 0\}$$

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, $0 \leqslant x \leqslant (p-1)/2$, sont tous distincts modulo p. De même, les nombres $-1-y^2$, $0 \leqslant y \leqslant (p-1)/2$, sont aussi distincts modulo p. Puisque il existe seulement p distincts classes de congruences modulo p, alors il existe $0 \leqslant x, y \leqslant (p-1)/2$ 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 (14.1) : 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. □

14.3 Les quaternions de Hamilton

On conclut ce chapitre section avec une discussion qui donne une explication plus concret du lemme 14.8. 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 (14.1).

14.4 Exercices

EXERCICE 14.1. 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 14.2. 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 14.3.

(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 14.4. L'équation $x^4 + x^2 = y^4 + 5$ possède-t-elle des solutions entières en x et y ?

EXERCICE 14.5. 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 14.6. 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 14.7.

(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 14.8. 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 $\mathbb{R}^4$?

Quatrième partie

Méthodes transcendantales

Chapitre 15

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 a 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 15.1 (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 suit : 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

$\{a_n\}_{n\geqslant 1}$:

$$\begin{aligned}
a_1 &= 0.\textcircled{a_{11}}a_{12}a_{13}\cdots a_{1n}\cdots\\
a_2 &= 0.a_{21}\textcircled{a_{22}}a_{23}\cdots a_{2n}\cdots\\
a_3 &= 0.a_{31}a_{32}\textcircled{a_{33}}\cdots a_{3n}\cdots\\
&\vdots\\
a_n &= 0.a_{n1}a_{n2}a_{n3}\cdots\textcircled{a_{nn}}\cdots\\
&\vdots
\end{aligned}$$

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 15.2. *Le nombre e est irrationnel.*

Démonstration. On a l'expansion de la fonction $x \to e^x$ à sa série de Taylor

$$e^x = \sum_{n\geqslant 0} \frac{x^n}{n!}.$$

Donc

$$e = \sum_{n\geqslant 0} \frac{1}{n!}.$$

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. 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,

$$\begin{aligned} 0 < b! \cdot \left(e - \sum_{n=0}^{b} \frac{1}{n!} \right) = b! \sum_{n=b+1}^{\infty} \frac{1}{n!} = \sum_{n=b+1}^{\infty} \frac{1}{(b+1)(b+2)\cdots n} &< \sum_{n=b+1}^{\infty} \frac{1}{(b+1)^{n-b}} \\ &= \frac{1}{b+1} \cdot \frac{1}{1-1/(b+1)} \\ &= \frac{1}{b} \leqslant 1. \end{aligned}$$

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 $|x - a/b| \leqslant 1/b$; sinon, on peut trouver un autre nombre $a' \in \mathbb{Z}$ tel que $|x - a'/b| \leqslant 1/b < |x - a/b|$. 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 $n < m \leqslant n + C$, 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

$$x - x_n \geqslant \frac{b_m}{10^m} \geqslant \frac{1}{10^{n+C}}.$$

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 16. Pour le moment, on montre le théorème suivant.

Théorème 15.3 (Théorème d'approximation de Dirichlet)**.** *Soient $x \in \mathbb{R}$ et $Q \geqslant 1$. Alors il existe une fraction réduite a/q telle que $1 \leqslant q \leqslant Q$ et*

$$\left| x - \frac{a}{q} \right| \leqslant \frac{1}{qQ} \leqslant \frac{1}{q^2}.$$

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 $[(j-1)/Q, j/Q)$, $1 \leqslant j \leqslant Q$. Le principe des tiroirs implique que au moins deux

nombres entre $\{x\}, \{2x\}, \ldots, \{(Q+1)x\}$, soient $\{kx\}$ et $\{\ell x\}$ avec $1 \leqslant k < \ell \leqslant Q+1$, 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 $1 \leqslant q \leqslant \ell - k \leqslant Q$ 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 15.4. 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 15.5. 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 15.7 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 15.6. *Si α est un nombre réel algébrique de degré $n \geqslant 2$, alors il existe un constant c dépendant seulement de α tel que*

$$\left|\alpha - \frac{a}{b}\right| \geqslant \frac{c}{b^n},$$

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 $\geqslant 2$ 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 $|a/b - \alpha| \geqslant \varepsilon \geqslant \varepsilon/b^n$ et le théorème découle à condition que $c \leqslant \varepsilon$. Finalement, supposons que $a/b \in I$ et posons $M = \sup_{x \in I} |f'(x)|$. Donc

$$\left|\alpha - \frac{a}{b}\right| \geqslant \frac{|f(a/b)|}{M}.$$

De plus, on a que

$$|f(a/b)| = \frac{|a_n a^n + a_{n-1}a^{n-1}b + \cdots + a_1 a b^{n-1} + a_0 b^n|}{b^n} \geqslant \frac{1}{b^n}$$

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 15.7 (Liouville)**.** *Le nombre*

$$x = \sum_{n \geqslant 1} \frac{1}{10^{n!}} = 0.11000100000000000000000001\ldots$$

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 15.6. En effet, soit

$$x_k = \sum_{n=1}^{k} \frac{1}{10^{n!}}$$

pour tout $k \geqslant 1$, 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,

$$0 \leqslant x - x_k = \sum_{n=k+1} \frac{1}{10^{n!}} \leqslant \frac{1}{10^{(k+1)!}} \sum_{n=k+1}^{\infty} \frac{1}{10^{n-k-1}} = \frac{1}{9 \cdot 10^{(k+1)!-1}} = \frac{10}{9 b_k^{k+1}}.$$

Alors x doit être transcendant. Sinon, il serait algébrique, soit de degré d. Selon le théorème 15.6, ceci impliquerait que

$$|x - x_k| \geqslant \frac{c}{b_k^d},$$

pour un $c > 0$ qui est independent de b_k. Alors il faudrait que

$$\frac{c}{b_k^d} \leqslant \frac{10}{9b_k^{k+1}} \quad \Longrightarrow \quad b_k^{k+1-d} \leqslant \frac{10}{9c}.$$

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. □

15.1 Exercices

EXERCICE 15.1. Soient $n, k \in \mathbb{N}$ avec $k \geqslant 2$. Montrez que le nombre $n^{1/k}$ est rationnel si et seulement si n est une k-ième puissance parfaite.

EXERCICE 15.2. Est-ce que le nombre $\sqrt{2} + \sqrt{3}$ est rationnel ou irrationnel ?

Chapitre 16

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 16.1. 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 16.1).

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

$$x = 5 + \frac{15}{17} = 5 + \frac{1}{17/15} = 5 + \cfrac{1}{1 + \cfrac{2}{15}} = 5 + \cfrac{1}{1 + \cfrac{1}{15/2}} = 5 + \cfrac{1}{1 + \cfrac{1}{7 + \cfrac{1}{2}}} = [5, 1, 7, 2].$$

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 16.2. *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

$$a_1 \leqslant \frac{k}{\ell} < a_1 + 1 \qquad \Longrightarrow \qquad a_1\ell \leqslant k < a_1\ell + \ell.$$

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 $0 \leqslant r_2/r_1 < 1$. Par la suite,

$$x = a_1 + \frac{1}{a_2 + \dfrac{r_2}{r_1}} = a_1 + \frac{1}{a_2 + \dfrac{1}{r_1/r_2}}.$$

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 $0 \leqslant r_n < r_{n-1} < \cdots < r_1 < \ell$, 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

$$[a_1, a_2, a_3] = a_1 + \frac{1}{a_2 + \dfrac{1}{a_3}} = a_1 + \frac{a_3}{a_2 a_3 + 1} = \frac{a_1(a_2 a_3 + 1) + a_3}{a_2 a_3 + 1}, \tag{16.1}$$

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

$$[a_1, \ldots, a_n] = a_1 + \frac{1}{[a_2, \ldots, a_n]} \tag{16.2}$$

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 16.3. On définit deux séquences de fonctions $\{F_n : \mathbb{R}^n \to \mathbb{R}\}_{n\geqslant 1}$ et $\{G_n : \mathbb{R}^n \to \mathbb{R}\}_{n\geqslant 1}$ 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

$$\frac{F_n(x_1, \ldots, x_n)}{G_n(x_1, \ldots, x_n)} = x_1 + \frac{1}{\frac{F_{n-1}}{G_{n-1}}(x_2, \ldots, x_n)}.$$

En utilisant cette relation et la relation (16.2), 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 $p_3 = a_3 p_2 + p_1$ et que $q_3 = a_3 q_2 + q_1$. On donne un dernière exemple : on a que

$$\begin{aligned}
[a_1, a_2, a_3, a_4] = a_1 + \frac{1}{[a_2, a_3, a_4]} &= a_1 + \frac{a_4 a_3 + 1}{a_4(a_2 a_3 + 1) + a_2} \\
&= \frac{a_1[a_4(a_2 a_3 + 1) + a_2] + a_4 a_3 + 1}{a_4(a_2 a_3 + 1) + a_2} \\
&= \frac{a_4(a_1(a_2 a_3 + 1) + a_3) + a_1 a_2 + 1}{a_4(a_2 a_3 + 1) + a_2}.
\end{aligned}$$

Par la suite, $p_4 = a_4(a_1(a_2 a_3 + 1) + a_3) + a_1 a_2 + 1 = a_4 p_3 + p_2$ et $q_4 = a_4(a_2 a_3 + 1) + a_2 = a_4 q_3 + q_2$. C'est un phénomène général :

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.

Théorème 16.4. *Pour tout $n \geqslant 1$, 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 16.5. *Pour tout $n \geqslant 1$, 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 16.6. *Soit $x \in \mathbb{R}$ et $\{p_n/q_n\}$ la suite de ses convergents.*

(a) On a que

$$\frac{p_1}{q_1} < \frac{p_3}{q_3} < \cdots < \frac{p_{2n-1}}{q_{2n-1}} < \cdots \leqslant x \leqslant \cdots < \frac{p_{2n}}{q_{2n}} < \cdots < \frac{p_4}{q_4} < \frac{p_2}{q_2}.$$

(b) Pour tout $n \in \mathbb{N}$, on a que

$$\left| x - \frac{p_n}{q_n} \right| \leqslant \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2}.$$

En particulier, $\lim_{n\to\infty} p_n/q_n = x$.

Démonstration. (a) Tout d'abord, on montre que

$$\frac{p_{2n-1}}{q_{2n-1}} \leqslant x \leqslant \frac{p_{2n}}{q_{2n}}, \tag{16.3}$$

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 (16.3) découle pour tout $i \geqslant 1$ 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 $\{p_{2n-1}/q_{2n-1}\}_{n\geqslant 1}$ est strictement croissante et que la suite $\{p_{2n}/q_{2n}\}_{n\geqslant 1}$ est strictement décroissante. En effet, on que

$$\begin{aligned} \frac{p_{n+2}}{q_{n+2}} - \frac{p_n}{q_n} = \left(\frac{p_{n+2}}{q_{n+2}} - \frac{p_{n+1}}{q_{n+1}} \right) + \left(\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} \right) &= \frac{(-1)^n}{q_{n+1}q_{n+2}} + \frac{(-1)^{n-1}}{q_n q_{n+1}} \\ &= \frac{(-1)^{n-1}(q_{n+2} - q_n)}{q_n q_{n+1} q_{n+2}}, \end{aligned}$$

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 (16.3). Donc

$$\left| x - \frac{p_n}{q_n} \right| \leqslant \left| \frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n} \right| = \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2},$$

ce qui conclut la démonstration. □

Remarque 16.7. 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 $\{p_{2n-1}/q_{2n-1}\}_{n\geqslant 1}$ est croissante et bornée au-dessus et, par la suite, elle est convergente. Soit x_1 sa limite. De même, la sous-suite $\{p_{2n}/q_{2n}\}_{n\geqslant 1}$ est décroissante et bornée en-dessous et, par la suite, elle est convergente. Soit x_2 sa limite. Finalement, puisque

$$0 \leqslant \frac{p_{2n}}{q_{2n}} - \frac{p_{2n-1}}{q_{2n-1}} = \frac{1}{q_{2n-1}q_{2n}} \to 0 \quad (n \to \infty),$$

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 16.8. *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 $1 \leqslant b \leqslant q_n$ 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 16.6(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 16.5. 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

$$\frac{q_{n+1} - q_{n-1}}{q_{n-1}q_n q_{n+1}} > \frac{1}{2q_{n-1}q_n} \quad \Leftrightarrow \quad q_{n+1} \geqslant 2q_{n-1}.$$

Mais on a que $q_{n+1} = a_n q_n + q_{n-1} \geqslant q_n + q_{n-1} > 2q_{n-1}$, 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

$$\left|\frac{a}{b} - \frac{p_n}{q_n}\right| = \frac{|aq_n - bp_n|}{bq_n} \geqslant \frac{1}{bq_n} \geqslant \frac{1}{bq_n} \geqslant \frac{1}{q_n^2},$$

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 $b \leqslant q_n$.

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,

$$\frac{p_n}{q_n} - x \geqslant \frac{p_n}{q_n} - \frac{a}{b} \geqslant \frac{1}{bq_n} \geqslant \frac{1}{q_n^2},$$

ce qui contredit théorème 16.6(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 $a/b \leqslant p_{n-1}/q_{n-1}$; sinon, on aurait que $p_{n-1}/q_{n-1} < a/b < x$. Par conséquent, on trouverait que

$$x - \frac{p_{n-1}}{q_{n-1}} \geqslant \frac{a}{b} - \frac{p_{n-1}}{q_{n-1}} = \frac{aq_{n-1} - bp_{n-1}}{bq_{n-1}} \geqslant \frac{1}{bq_{n-1}} \geqslant \frac{1}{q_n q_{n-1}},$$

car $b \leqslant q_n$, 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 $a/b \leqslant p_{n-1}/q_{n-1} < x$, ce qui implique que

$$\left|x - \frac{a}{b}\right| \geqslant \left|\frac{p_{n-1}}{q_{n-1}} - x\right| > \left|x - \frac{p_n}{q_n}\right|,$$

selon le cas où $a/b = p_{n-1}/q_{n-1}$ qu'on a déjà montré. Ceci conclut la démonstration du théorème. □

Remarque 16.9. Le théorème 16.8 implique directement le théorème d'approximation de Dirichlet. En effet, si $p_1/q_1, p_2/q_2, \ldots$ est la suite des convergents de x, alors il y a deux cas. Si $q_n \leqslant Q$ pour tout $n \geqslant 1$, alors $x \in \mathbb{Q}$ et $x = b/r$ avec $r \leqslant Q$, et le résultat découle tout de suite en posant $a/q = b/r$. Finalement, si il existe $n \geqslant 1$ tel que $q_n > Q$, on peut trouver m tel que $q_m \leqslant Q < q_{m+1}$. Donc si on pose $a/q = p_m/q_m$, on a que

$$\left|x - \frac{a}{q}\right| = \left|x - \frac{p_m}{q_m}\right| \leqslant \frac{1}{q_m q_{m+1}} \leqslant \frac{1}{q_m Q} = \frac{1}{qQ},$$

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

$$x = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \ldots}}}.$$

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 16.10. Une fraction continue $[a_1, a_2, \dots]$ est appelée *périodique* si il existe $k \geqslant 1$ tel que $a_{n+k} = a_n$, pour tout $n \geqslant 1$. 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 $k \geqslant 1$ et $\ell \geqslant 1$ tels que $a_{n+k} = a_n$, pour tout $n \geqslant \ell$. 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 16.11. *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 $s \geqslant 0$.*

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 16.2.

Supposons que $x = [a_1, \dots, a_{\ell-1}, \overline{a_\ell, \dots, a_{k+\ell-1}}]$ pour quelques $k, \ell \geqslant 1$. 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, $1 \leqslant i \leqslant k+1$, 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 16.2. Donc

$$q_k y^2 + q_{k-1} y = p_k y + p_{k-1} \qquad \Longrightarrow \qquad q_k y^2 + (q_{k-1} - p_k) y - p_{k-1} = 0.$$

En appliquant la formule quadratique, on conclut que $y = r' + \sqrt{s'}$ pour quelques $r', s' \in \mathbb{Q}$ (nécessairement $s' \geqslant 0$ 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 $s \geqslant 0$. Alors il existe $a, b, c \in \mathbb{Z}$ tels que

$$ax^2 + bx + c = 0. \tag{16.4}$$

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 16.2 implique que

$$x = \frac{x_n p_{n-1} + p_{n-2}}{x_n q_{n-1} + q_{n-2}} \quad (n \geqslant 3). \tag{16.5}$$

(Notez que si p'_i/q'_i, $1 \leqslant i \leqslant n$, sont les convergents de la fraction continue $[a_1, \ldots, a_{n-1}, x_n]$, alors $p'_i = p_i$ et $q_i = q'_i$ pour $i \in \{1, \ldots, n-1\}$.) 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 $\{x_n : n \geqslant 3\}$ et fini (et, par la suite, $x_{n_1} = x_{n_2}$ pour quelques n_1 et n_2 qui sont différents).

Le théorème 16.5 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 (16.4). 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 $n \geqslant 3$, 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 16.6(b) implique que $|h_i| < 1/(q_i q_{i+1}) \leqslant 1$. 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

$$|f(x + h_i)| = |f(x) + h_i f'(t_i)| = |h_i f'(t_i)| \leqslant \frac{1}{q_i^2} \cdot M,$$

où $M = \sup_{x-1 \leqslant t \leqslant x+1} |f'(t)|$. Par la suite,

$$|A| = q_{n-1}^2 |f(x + h_{n-1})| \leqslant M \quad \text{et} \quad |C| = q_{n-2}^2 |f(x + h_{n-2})| \leqslant M,$$

ce qui implique aussi que

$$|B| = \sqrt{B^2} = \sqrt{d + 4AC} \leqslant \sqrt{d} + 2\sqrt{|AC|} \leqslant \sqrt{d} + 2M.$$

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 $\{x_n : n \geqslant 3\}$ est fini, comme affirmé. Ceci conclut la démonstration. □

16.1 Exercices

EXERCICE 16.1. Montrez que la fraction continue d'un nombre est uniquement définie.

Bibliographie

[1] P. Erdős, *On integers of the form $2^k + p$ and some related problems*, Summa Brasil. Math. 2 (1950), 113–123.

[2] P. Erdős, M. Kac, *The Gaussian Law of Errors in the Theory of Additive Number Theoretic Functions.*. American Journal of Mathematics. 62 (1940), 738–742.

[3] A. Granville, *Number theory revealed : a masterclass.* American Mathematical Society, Providence, RI, 2019.

[4] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers.* Sixth edition. Revised by D. R. Heath-Brown and J. H. Silverman. With a foreword by Andrew Wiles. Oxford University Press, Oxford, 2008.

[5] D. Koukoulopoulos, *The distribution of prime numbers.* Graduate Studies in Mathematics, 203. American Mathematical Society, Providence, RI, 2019.

[6] J. Maynard, *Small gaps between primes.* Ann. of Math. (2) 181 (2015), no. 1, 383–413.

[7] I. Niven, H. Zuckerman and H. L. Montgomery, *An introduction to the theory of numbers.* Fifth edition. John Wiley & Sons, Inc., New York, 1991.

[8] G. Rousseau, *On the quadratic reciprocity law.* J. Austral. Math. Soc. Ser. A 51 (1991), no. 3, 423–425.

[9] L. Schoenfeld, *Sharper bounds for the Chebyshev functions $\theta(x)$ and $\psi(x)$. II.* Math. Comp. 30 (1976), no. 134, 337–360.

[10] T. Tao, *Polymath8b : Bounded intervals with many primes, after Maynard*, blog post (2013), https://terrytao.wordpress.com/2013/11/19/polymath8b-bounded-intervals-with-many-primes-after-maynard/

[11] D. H. J. Polymath, *Variants of the Selberg sieve, and bounded intervals containing many primes.* Research in the Mathematical Sciences 1 :12 (2014).

[12] Y. Zhang, *Bounded gaps between primes.* Ann. of Math. (2) 179 (2014), no. 3, 1121–1174.