Nombres : ensembles, bases et types

1 Ensembles de nombres

1.1 Entiers

Entiers naturels $\mathbb{N}$ désigne l'ensemble des *entiers naturels*, c'est à dire les nombres utilisés pour compter ou ordonner des objets : 0, 1, 2, etc.

Division euclidienne Étant donnés deux entiers naturels n et d, il existe une seule manière d'écrire n sous la forme $d \times q + r$ avec q et r deux entiers naturels et $r < d$. Cette décomposition s'appelle la *division euclidienne* de n par d, n est appelé le dividende, d le diviseur, q le quotient et r le reste. Si $r = 0$, on dit que d divise n, ou que d est un diviseur de n, ou que n est un multiple de d. Un nombre est dit *premier* s'il possède exactement deux diviseurs.

L'équation $x + 3 = 7$ possède une solution dans $\mathbb{N}$ (il s'agit de 4), mais l'équation $x + 5 = 2$ n'a pas de solution.

Entiers relatifs $\mathbb{Z}$ désigne l'ensemble des *entiers relatifs*, c'est à dire les nombres utilisés pour mesurer des différences de quantités : ils incluent les nombres négatifs. Quand j'utilise le mot « entier » tout seul, je parle généralement des entiers relatifs.

Propriété Tout entier relatif possède un opposé :
$\forall n \in \mathbb{Z}, \exists p \in \mathbb{Z}, n + p = 0$. Le nombre p est alors noté $-n$.

La division euclidienne peut être étendue aux entiers relatifs : Si n et d sont deux entiers relatifs, il existe une seule manière d'écrire n sous la forme $d \times q + r$ avec q et r deux entiers relatifs et $0 \leq r < d$.

Les entiers relatifs sont l'ensemble des solutions de toutes les équations de la forme $x + a = b$ avec $a \in \mathbb{N}$ et $b \in \mathbb{N}$. Ainsi, l'équation $x + 5 = 2$ a une solution dans $\mathbb{Z}$. En revanche, l'équation $3x = 2$ n'a pas de solution.

1.2 Rationnels

$\mathbb{Q}$ désigne l'ensemble des *rationnels*, c'est à dire les nombres utilisés pour calculer des rapports entre des quantités.

Représentation On peut représenter un rationnel à l'aide d'une fraction $\frac{n}{p}$. Si $p > 0$ et que n et p sont premiers entre eux (c'est à dire qu'ils n'ont pas de diviseur commun à part 1), on dit que la fraction est *irréductible*.
On peut également le représenter à l'aide d'un nombre à virgule. Dans ce cas, à partir d'un certain rang, les décimales sont périodiques : elles se répètent selon un cycle de taille finie. On note parfois ce cycle à l'aide d'une barre.
Exemples :

$$\begin{aligned}\frac{5}{14} &= 0.3\overline{571428} = 0.3571428571428571428\dots\\ \frac{1}{300} &= 0.00\overline{3} = 0.00333333\dots\\ 1 &= 1.\overline{0} = 1.000000\dots\\ &= 0.\overline{9} = 0.999999\dots\end{aligned}$$

Démonstration de $1 = 0.\overline{9}$

$$\begin{aligned}0.\overline{9} + 9 &= 9.\overline{9}\\ \Leftrightarrow \frac{0.\overline{9} + 9}{10} &= \frac{9.\overline{9}}{10}\\ &= 0.\overline{9}\\ \Leftrightarrow 0.\overline{9} + 9 &= 10 \times 0.\overline{9}\\ \Leftrightarrow 9 &= 9 \times 0.\overline{9}\\ \Leftrightarrow 1 &= 0.\overline{9}\end{aligned}$$

Propriété Tout nombre rationnel non nul possède un inverse :
$\forall q \in \mathbb{Q}^*, \exists p \in \mathbb{Q}^*, q \times p = 1$. Le nombre p est alors noté $1/q$.

Les rationnels sont l'ensemble des solutions de toutes les équations de la forme $px = n$ où n et p sont deux entiers et $p \neq 0$. En revanche, l'équation $x^2 = 2$ n'a pas de solution.

Démonstration de l'irrationnalité de $\sqrt{2}$

— Montrons d'abord que pour tout entier n, si n^2 est pair, alors n est pair. Pour cela, on va montrer la contraposée, c'est à dire que si n est impair, alors n^2 est impair. Soit n un entier pair. On peut écrire $n = 2k + 1$ avec $k \in \mathbb{Z}$. Développons son carré :

$$\begin{aligned}n^2 &= (2k + 1)^2\\ &= 4k^2 + 4k + 1\\ &= 2(2k^2 + 2k) + 1\end{aligned}$$

On constate que n^2 s'écrit sous la forme $2k' + 1$ avec $k' \in \mathbb{Z}$, autrement dit n^2 est impair.

— Montrons maintenant l'irrationnalité de $\sqrt{2}$ à l'aide d'un raisonnement par l'absurde : Supposons que $\sqrt{2} \in \mathbb{Q}$ et trouvons une contradiction. On note $\frac{p}{q}$ la fraction irréductible représentant $\sqrt{2}$, avec $p \in \mathbb{Z}$, $q \in \mathbb{Z}^*$ et $\text{pgcd}(p, q) = 1$. On a :

$$\left(\frac{p}{q}\right)^2 = 2 = \frac{p^2}{q^2}$$
$$\Leftrightarrow p^2 = 2q^2$$

On constate que p^2 est pair. On en déduit que p est pair et peut donc s'écrire $p = 2k$ avec $k \in \mathbb{Z}$. Alors $p^2 = 4k^2$ et :

$$2q^2 = 4k^2$$
$$\Leftrightarrow q^2 = 2k^2$$

On constate que q^2 est pair. On en déduit que q est pair. p et q étant tous les deux pairs, ils sont tous deux divisibles par 2, autrmeent dit $\text{pgcd}(p, q) \geq 2$. Or on a fait l'hypthèse que la fraction est irréductible, soit $\text{pgcd}(p, q) = 1$, ce qui est une contradiction.

— En conclusion, $\sqrt{(}2) \notin \mathbb{Q}$.

1.3 Réels

$\mathbb{R}$ désigne l'ensemble des *réels*, c'est à dire les nombres utilisés pour représenter des phénomènes continus. On les représente parfois à l'aide de la droite réelle : chaque point de cette droite représente un nombre réel. Les réels sont beaucoup utilisés car ils permettent de définir des fonctions continues et des dérivées, et donc d'établir et de résoudre des équations différentielles.

Continuité d'une fonction Une fonction $f : \mathbb{R} \to \mathbb{R}$ est continue en a si et seulement si :

$$\lim_{x \to a^-} f(x) = \lim_{x \to a^+} f(x) = f(a)$$

Une fonction est continue sur un intervalle I si elle est continue en chaque point de I.

Exemple La fonction sinus cardinal est utilisée dans plusierus domaines de la physique pour représenter certaines figures de diffraction ou bien la distribution d'un électron dans un atome. Elle est définie sur $\mathbb{R}$ par :

$$\text{sinc}(x) = \begin{cases} \frac{\sin x}{x} & \text{si } x \neq 0 \\ 1 & \text{si } x = 0 \end{cases}$$

Pour montrer que sinc est continue en 0, on utilise le fait que :

$$\lim_{x \to 0^-} \frac{\sin x}{x} = \lim_{x \to 0^+} \frac{\sin x}{x} = 1$$

Limitations L'équation $x^2 - 5x + 6 = 0$ possède deux solutions réelles, en revanche l'équation $x^2 + 1 = 0$ n'a pas de solution dans $\mathbb{R}$.

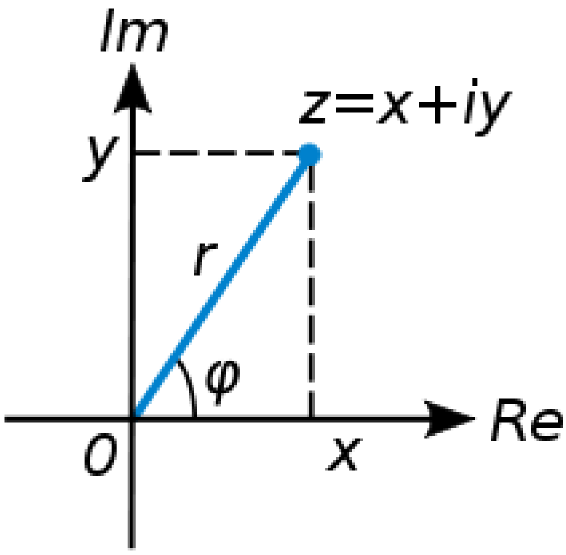
1.4 Complexes

$\mathbb{C}$ désigne l'ensemble des *complexes*, qui sont utilisés entre autres pour représenter des points et des transformations dans le plan.
On postule l'existence de deux solutions à l'équation $x^2 = -1$, que l'on note i et $-i$. Les nombres complexes sont alors l'ensemble des nombres z de la forme $z = a + ib$ où a et b sont deux réels. a est la partie réelle de z, et b est sa partie imaginaire. La forme $a + ib$ est appelée la forme algébrique de z.

Propriétés Si $z = a + ib$ et $z' = c + id$ sont deux nombres complexes, alors :

$$\begin{aligned} z + z' &= a + c + i(b + d) \\ z \times z' &= ac - bd + i(ad + bc) \\ \frac{1}{z} &= \frac{a}{a^2 + b^2} - i\frac{b}{a^2 + b^2} \end{aligned}$$

On appelle *conjugué* de z le nombre $\bar{z} = a - ib$ et *module* de z le réel $|z| = \sqrt{a^2 + b^2}$. On appelle *argument* de z le réel $\theta \in [0; 2\pi[$ tel que $a = |z|\cos\theta$ et $b = |z|\sin\theta$. On représente souvent $\mathbb{C}$ par le plan complexe, construit en utilisant la droite réelle comme axe des abscisses et la droite imaginaire comme axe des ordonnées.



Ici le module de z est noté r et l'argument φ. Source : Wikipédia.

Notation exponentielle Un nombre complexe z de module r et d'argument θ peut aussi s'écrire $z = re^{i\theta}$. Par exemple, $i = e^{i\frac{\pi}{2}}$. Si $z' = se^{i\varphi}$, alors $z \times z' = rse^{i(\theta+\varphi)}$.

La formule d'Euler est parfois décrite comme « la plus belle formule des mathématiques ». Elle s'écrit :

$$e^{i\pi} + 1 = 0$$

Propriété Toute équation du second degré possède une ou deux solutions dans $\mathbb{C}$. Quand le discriminant Δ est strictement négatif, l'équation $ax^2 + bx + c = 0$ possède deux solutions complexes :

$$z_1 = \frac{-b - i\sqrt{-\Delta}}{2a}$$
$$z_2 = \frac{-b + i\sqrt{-\Delta}}{2a}$$

2 Bases de numération

Pour éviter certaines ambiguités, je note parfois un nombre xyz_b, ce qui signifie que le nombre est exprimé en base b (exprimée en base 10) et que ses chiffres dans cette base sont xyz.

2.1 Décimale

La manière usuelle de représenter les nombres entiers est *positionnelle* : la valeur d'un chiffre dépend de sa position dans le nombre. Ainsi le nombre 237_{10} correspond à l'expression $7 \times 10^0 + 3 \times 10^1 + 2 \times 10^2$. Les constituents d'un nombre sont regroupés en paquets de 10, et ces paquets peuvent eux-mêmes être regroupés en paquets de paquets (les centaines), en paquets de paquets de paquets (milliers), etc. On a besoin de 10 chiffres pour représenter les nombres en base 10.

Au cours d'une addition, si la somme de deux chiffres dépasse 10, on forme un paquet qui constitue la retenue, et on note le reste à l'emplacement actuel.

2.2 Binaire

En base deux, on forme des paquets de deux. On n'a donc besoin que de deux chiffres, 0 et 1. Le nombre 1101_2 correspond à l'expression $1 \times 2^0 + 0 \times 2^1 + 1 \times 2^2 + 1 \times 2^3$, ce qui équivaut à 13_{10}.

Table d'addition

+	0	1
0	0	1
1	1	10

Table de muliplication

×	0	1
0	0	0
1	0	1

Exemple $6_{10} \times 5_{10} = 30_{10}$. En binaire : $6_{10} = 110_2$ et $5_{10} = 101_2$. On a alors :

$$
\begin{array}{ccccc}
 & & 1 & 1 & 0 \\
\times & & 1 & 0 & 1 \\
\hline
 & & 1 & 1 & 0 \\
 & 0 & 0 & 0 & \cdot \\
1 & 1 & 0 & \cdot & \cdot \\
\hline
1 & 1 & 1 & 1 & 0
\end{array}
$$

On trouve bien $11110_2 = 30_{10}$. Il existe une méthode alternative pour faire des multiplications plus rapidement, qui s'applique uniquement à la base 2. On pose les deux nombres à multiplier côte à côte, puis, ligne par ligne, on décale tous les chiffres su premier nombre vers la gauche (en ajoutant des zéros) et tous ceux du deuxième nombre vers la droite, jusqu'à ce qu'il ne reste plus que le nombre 1 à droite. On ajoute alors toutes les lignes de la colonne de gauche en ignorant celles pour lesquelles la colonne de droite se termine par zéro.

Exemple $6_{10} \times 5_{10} = 30_{10}$.

$$
\begin{array}{ccccccccc}
 & & 1 & 1 & 0 & \times & 1 & 0 & 1 \\
 & 1 & 1 & 0 & 0 & & & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & & & & 1 \\
\hline
1 & 1 & 1 & 1 & 0 & & & &
\end{array}
$$

Conversions Pour convertir un nombrede l'n écriture décimale vers le binaire, on divise successivement par 2 jusqu'à obtenir un quotient de 0. La liste des restes lues en sens inverse correspond à l'écriture binaire du nombre.

Exemple

$$
\begin{aligned}
35 &= 2 \times 17 + 1 \\
17 &= 2 \times 8 + 1 \\
8 &= 2 \times 4 + 0 \\
4 &= 2 \times 2 + 0 \\
2 &= 2 \times 1 + 0 \\
1 &= 2 \times 0 + 1 \\
35_{10} &= 100011_2
\end{aligned}
$$

Pour convertir un nombre du binaire vers le décimal, on additionne les puissances de 2 correspondant aux chiffres égaux à 1 dans l'écriture binaire.

Exemple

1	0	0	0	1	1	1
	1	0	0	0	1	2
		1	0	0	0	4
			1	0	0	8
				1	0	16
					1	32
						35

2.3 Autres

Hexadéciamal En base 16, on a besoin de 6 symboles supplémentaires pour représenter les chiffres de 10 à 15. On utilise généralement les lettres A à F. Ainsi, $10_{11} = B_{16}$, $16_{10} = 10_{16}$ et $42_{10} = 2A_{16}$. L'hexadécimal est notamment utilisé pour représenter des niveaux de couleurs entre 0 et $255_{10} = FF_{16}$. On rencontre parfois la base 60 (pour compter le temps en heures, minutes et secondes), la base 12 ou 24 (pour les heures dans une journée ou les douzaines d'œufs), etc.

3 Types fondamentaux

3.1 Integer

Dans la plupart des langages de programmation, on réserve un nombre précis de bits en mémoire pour stocker les nombres. Il existe deux conventions pour représenter les nombres entiers, suivant que l'on veut représenter seulement des entiers positifs ou bien des entiers positifs et négatifs.

Entiers non signés On désigne ces types par uintN, où N est le nombre de bits utilisés (généralement 32 ou 64). Les nombres sont simplement représentés sous leur forme binaire. Par exemple, le nombre 2 peut être représenté par l'uint3 010 ou par l'uint6 000010.

Overflow Un uintN ne peut pas représenter plus de 2^N nombres différents, on va donc représenter les entiers de 0 à $2^N - 1$. Le plus grand nombre représentable par un uint3 est donc $2^3 - 1 = 7$. Lors d'une addition, si une retenue dépasse la capacité de stockage du type, elle est ignorée. Cela conduit au phénomène d'overflow : dépasser la limite remet le compteur à zéro. Par exemple, en uint3, 111+001=000 (on ignore la retenue qui devrait donner 1000 car on n'a que 3 bits pour stocker le résultat).

Entiers signés On désigne ces types par intN. Là encore, on ne peut pas représenter plus de 2^N nombres différents, on choisit de représenter les nombres de -2^{N-1} à $2^{N-1} - 1$ en utilisant la représentation par complément à 2^N (souvent appelée complément à 2). Les nombres positifs sont représentés normalement. Un nombre négatif $-x$ est représenté par $2^N - x$, soit le nombre qu'il faut ajouter à x pour obtenir 2^N, c'est-à-dire produire un overflow et obtenir 0. Une manière de calculer cette représentation est d'écrire d'abord le nombre x (positif) puis d'inverser tous ses bits (les 0 deviennent des 1 et inversement) et enfin d'ajouter 1.

Exemple Avec le type int3, le nombre $2_{10} = 10_2$ est représenté par 010, tandis que le nombre $-3_{10} = -011_2$ est représenté par 101.

Propriété Avec la convention du complément à 2, les nombres positifs commencent tous par 0 et les nombres strictement négatifs commencent tous par 1.
L'intérêt de cette méthode est qu'elle permet d'utiliser l'algorithme d'addition des entiers non signés pour additionner des entiers signés sans avoir besoin de le modifier ou de tester le signe.
Attention toutefois, le type intN est tout autant susceptible de produire des overflows que le type uintN.

Integer en python En python, le problème d'overflow ne se pose pas car les entiers n'ont pas de limite de taille. Le langage utilise une structure spéciale qui repose sur des int64 tant que c'est possible, mais qui peut allouer des blocs mémoire supplémentaires pour représenter des entiers plus longs en cas de besoin.

3.2 Float

La plupart des langages de programmation, y compris python, utilisent la norme IEEE-754 pour définir les nombres à virgule flottante. Certains langages distinguent les nombre à simple précision des nombres à double précision, là où python utilise seulement des nombres à double précision. Les nombres à double précision sont stockés sur 64 bits : le premier bit sert au signe s, les onze suivants représentent l'exposant e et les 52 bits restants représentent la mantisse m.

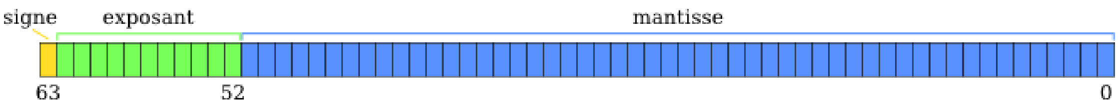


Schéma de la représentation d'un float à double précision. Source : Wikipédia.

Le nombre est donc représenté sous la forme $\pm m \times 2^e$, avec $1 \leq m < 2$, d'une manière analogue à la notation scientifique qui utilise les puissances de 10.
Les 11 bits d'exposant permettent de représenter des exposants positifs et négatifs entre $-2^1 0 et 2^1 0 - 1$ (mais d'une manière différente des compléments à 2). Cette représentation permet également de définir des valeurs « spéciales » pour représenter l'infini ou Nan (Not a Number).
En python, la variable `sys.float_info` du module `sys` permet d'avoir accès à certaines propriétés des float, notamment le plus grand nombre représentable (max), ou encore l'écart entre 1 et le plus petit nombre strictement plus grand que 1 représentable (epsilon), qui donne une idée du niveau de précision maximal atteignable.

Limites Le type float présente trois limites :

- Il est soumis au même risque d'overflow que le type int si l'exposant devient trop grand.
- La précision est également limitée. Pour cette raison, il est déconseillé de comparer deux float avec les opérateurs `==` ou `!=`. Par exemple, l'expression :
  `print(0.1+0.1+0.1 == 0.3)`
  affiche `False`, ce qui peut causer quelques confusions.
- Certains nombres qui ont une écriture à virgule finie en base 10 ont une écriture infinie en base 2. Par exemple, $0.1_{10} = 0.0\overline{0011}_2$

3.3 Complex

En python, i est représenté par la lettre `j`. Tout nombre suivi directement par la lettre `j` ou `J` représente un imaginaire pur. Le nombre $3 + 2i$ s'écrit alors `3+2j` et le nombre i s'écrit `1j`.