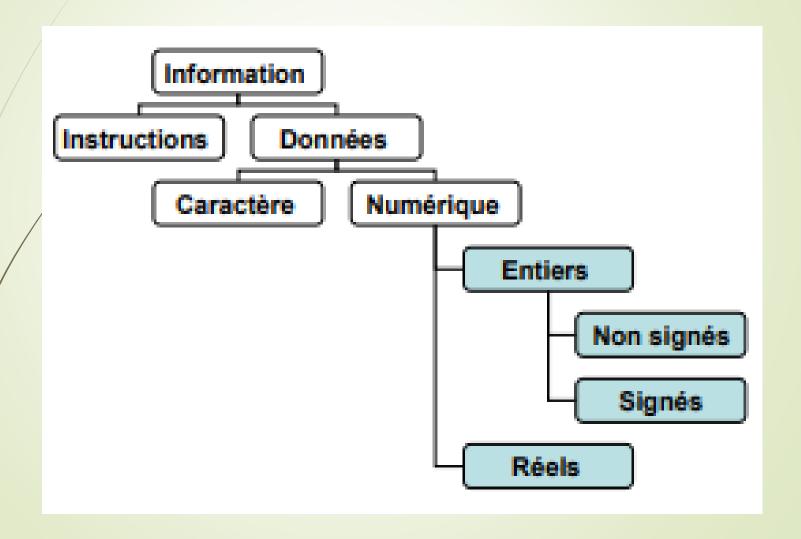
Arithmétique de l'ordinateur : Représentation des réels.

Le matériel

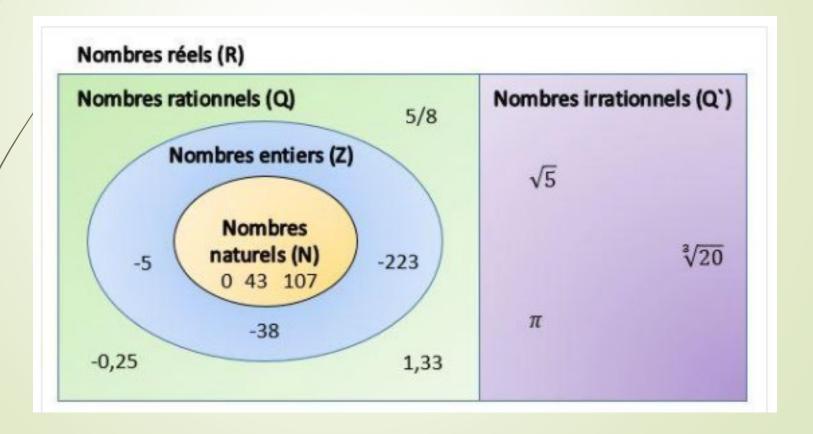
- Synthèse du professeur.
- Site Web: https://www.prodafor.com/informatique
 - Section Notation scientifique
 - Représentation normalisée

Introduction



Représentation des nombres réels

Rappel: un nombre réel :



 Nous allons représenter les nombres réels en version « nombres à virgule » :

Attention: seuls les réels qui peuvent se représenter avec un nombre fini de décimales après la virgule pourront être stockés dans un ordinateur. Pour les autres réels, nous utiliserons une approximation.

Représenter des réels en décimal

Décortiquons 6 431,986...

Position	3	2	1	0	,	-1	-2	-3
Valeur	10 ³	10 ²	10 ¹	100	,	10-1	10-2	10-3
Symbole	6	4	3	1	,	9	8	6
=	6000	400	30	1		0,9	0,08	6

Nombre réels en binaire

- Une possibilité (sur 16 bits):
 - · mettons la virgule au milieu:

```
Valeur 2<sup>7</sup> 2<sup>6</sup> 2<sup>5</sup> 2<sup>4</sup> 2<sup>3</sup> 2<sup>2</sup> 2<sup>1</sup> 2<sup>0</sup> , 2<sup>-1</sup> 2<sup>-2</sup> 2<sup>-3</sup> 2<sup>-4</sup> 2<sup>-5</sup> 2<sup>-6</sup> 2<sup>-7</sup> 2<sup>-8</sup>

Position b<sub>15</sub> b<sub>14</sub> b<sub>13</sub> b<sub>12</sub> b<sub>11</sub> b<sub>10</sub> b<sub>9</sub> b<sub>8</sub> , b<sub>7</sub> b<sub>6</sub> b<sub>5</sub> b<sub>4</sub> b<sub>3</sub> b<sub>2</sub> b<sub>1</sub> b<sub>0</sub>
```

- 8 premiers bits: 2º à 27 (0 à 255)
- 8 derniers bits: 2-1 à 2-8 (1/256 à 255/256, 0.00390625 à 0.99609375)

0 0 1 0 1 1 0 0 1 0 1 0 0 0 0

Problèmes:

- Très limité
 - Quelle est la valeur maximale?

$$255,99609375 (255 \frac{255}{256})$$

Quelle est la précision (ou la plus petite différence entre deux nombre)?

$$\frac{1}{256}$$

Représentation en mode virgule flottante

- On commence par mettre le nombre sous la forme 0,...
- Il faut ensuite mettre :
 - Un bit pour le signe
 - r bits pour la mantisse
 - s+1 bits pour l'exposant (s bits pour l'exposant + 1 bit pour le signe de l'exposant)
- Exemple:

Mettre 5 909 en mode virgule flottante si r = 4 et s = 2

+ 5 9 0 9 + 0 4

Exercices pour r=4 et s=2

0,000547

1489621

-0,050002

La norme IEEE 754

- La norme IEEE 754 a été adoptée en 2008 pour les nombres rationnels sur 32, 64 et 128 bits
- Très similaire à la notation scientifique:

(signe) 1, mantisse x 2^(exposant-127)

- Sur 32 bits:
 - signe: un bit
 - base: 2, donc binaire. Comme cette base est toujours 2, on n'a pas besoin de la stocker (c'est implicite)
 - caractéristique: toujours 1. On n'a pas besoin de la stocker (implicite également)
 - exposant (décalé): 8 bits (donc de 0 à 255), mais on soustrait 127, donc de -127 à +128

1 bit	8 bits	23 bits
signe	exposant	mantisse

Quelques précisions.

- La précision <u>double</u> contenu dans 64 bits possède un bit de signe, 11 bits d'exposant et 52 bits de mantisse.
- Il existe une précision <u>étendue</u> nécessitant 80 bits.
- L'exposant est noté en **excès de 127** en précision simple et en excès de 1023 en précision double.
- La mantisse est tronquée, arrondie ou encore complétée à droite par des zéros.
- Le bit 1 à gauche de la mantisse n'est pas mémorisé.

Ex-1: Écrire 54,625 selon la norme IEEE 754 en simple précision.

54	2
27	0
13	1
6	1
3	0
1	1
0	1

0,625	x2
1,	25
0,	5
1,	0

L'exposant par excès de 127:

10000100

 $110110,101 = 1,10110101 \times 2^5$



signe

exposant

mantisse

Ex-2: 1648,75 (Du décimal vers IEEE754)

Ex-3: (Du IEEE754 vers le décimal)

Convertir 0x411A0000 (écrit en IEEE754) en décimal.

Bit de signe = 0, donc nombre positif

Exposant = 0b10000010 = 130. 130-127 = 3

Mantisse = 0b0011010...

• =
$$1x2^{-3} + 1x2^{-4} + 1x2^{-6} = 0,125 + 0,0625 + 0,015625 = 0,203125$$

$$(+)$$
 1,203125 x 2^3 = 9,625

(signe) 1,mantisse x 2(exposant-127)								
1 bit	8 bits	23 bits						
signe	exposant	mantisse						

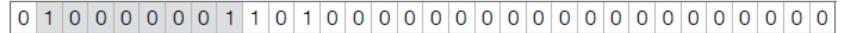
Ex-4: (Du IEEE754 vers le décimal)

Convertir 0x40D00000 (écrit en IEEE754) en décimal

· Tout d'abord, convertissons en binaire

		4	4			(0)			()			()			()			()			()	
(0	1	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

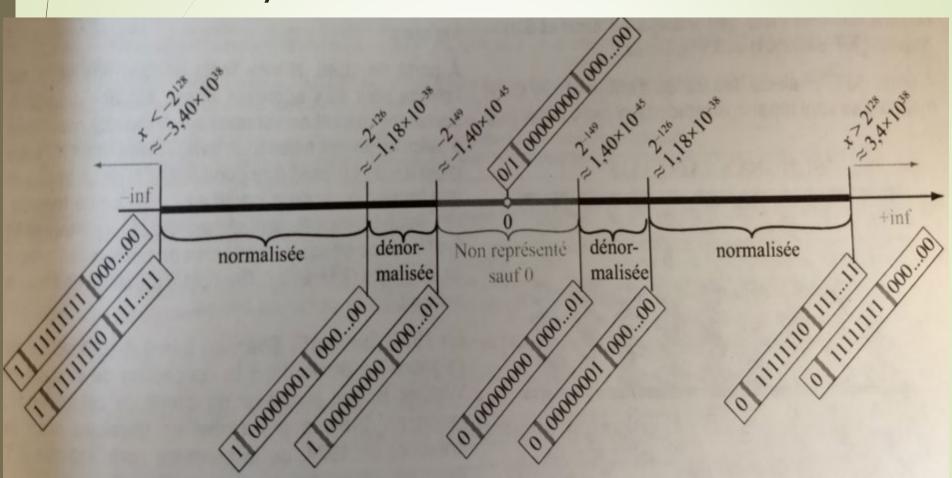
 Séparons la chaîne de bits en sections correspondants aux champs de l'IEEE754 (signe, exposant, mantisse)



Forme dénormalisée (IEEE 754)

- La forme normalisée d'un nombre sous forme IEEE 754 est : $signe \times (1 + mantisse) \times 2^{exposant-127}$
- Lorsque l'exposant = 0 et que la mantisse = 0, c'est la représentation de 0
- Lorsque l'exposant = 0, mais que la mantisse ≠ 0, alors on modifie la notation pour une forme dénormalisée
- Forme dénormalisée : $signe \times mantisse \times 2^{-126}$
- Permet une plus grande précision

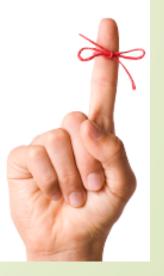
Forme dénormalisée (IEEE 754)



- A priori, nous ne pouvons pas savoir ce qu'une chaîne binaire signifie.
 - Ex: que veut dire 0x416C6C6F (sur 32 bits)?
 - La bonne réponse est: ça dépend!

entier non-signé	1097624687
entier signé	1097624687
réel	14,47764

 Il nous faut donc savoir quel format utiliser pour bien interpréter les données



Devoir

Faire les exercices du fichier (sur Omnivox):

Exercices arithmétique ordinateur réels

- Écouter, si nécessaire, les capsules vidéo suivantes (sur prodafor.com):
 - NotationSC01
 - RepNorme01