TD4 - Représentation des entiers relatifs et des réels

1 Représentation des entiers relatifs en complément à 2

① Donner la représentation en complément à 2 sur 7 bits, quand cela est possible, des entiers relatifs qui suivent : -53, -7, 0, 59 et 88.

② En considérant la même représentation, donner la valeur codée par les séquences binaires 10110101 et 11101001.

Correction

① Sur k bits on peut représenter les entiers N tels que $-\left(2^{k-1}\right) \leq N \leq +\left(2^{k-1}-1\right)$.
Par conséquent, sur 7 bits, on a $-64 \leq N \leq +63$. Mis à part 88, les autres entiers relatifs sont donc représentables.

- Zéro et les entiers positifs ont la même représentation que les entiers naturels donc $0 = 0000000$ et $59 = 0111011$.
- Pour les entiers négatifs on utilise le binaire de leur valeur absolue, on calcule le complément à 1, puis on ajoute 1. On obtient ainsi :
  - $53_{10} = 0110101_2$, puis 1001010 et $1001011 = -53$;
  - $7_{10} = 0000111_2$, puis 1111000 et $1111001 = -7$.

② Déterminer l'entier relatif représenté par une séquence binaire

(a) Valeur codée par 10110101
- Bit de signe à 1 $\rightarrow$ nombre négatif ;
- pour obtenir la valeur absolue du nombre codé, il suffit de calculer le complément à 2 du nombre.
  1. Inversion des bits (complément à 1)
     $10110101 \rightarrow 01001010$
  2. Ajout de 1

$$\begin{array}{rcccccccc} & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ + & & & & & & & & 1 \\ \hline & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \end{array}$$

Or

$$\begin{aligned} 1001011_2 &= 1 \times 2^6 + 1 \times 2^3 + 1 \times 2^1 + 1 \times 2^0 \\ &= 64 + 8 + 2 + 1 \\ &= 75 \end{aligned}$$

Donc, c'est -75 qui est codé par 10110101 sur 8 bits.

(b) Valeur codée par 11101001
— Bit de signe à 1 $\rightarrow$ nombre négatif ;
— pour obtenir la valeur absolue du nombre codé, il suffit de calculer le complément à 2 du nombre.
1. Inversion des bits (complément à 1)
11101001 $\rightarrow$ 00010110
2. Ajout de 1

$$\begin{array}{r} 0\;1\;1\;0\;0\;0\;0\;1 \\ +\qquad\qquad\qquad 1 \\ \hline 0\;1\;1\;0\;0\;0\;1\;0 \end{array}$$

Or

$$\begin{aligned} 10111_2 &= 1 \times 2^4 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0 \\ &= 16 + 4 + 2 + 1 \\ &= 23 \end{aligned}$$

Donc, c'est -23 qui est codé par 11101001 sur 8 bits.

2 Représentation des réels

2.1 Représentation en virgule fixe

Correction

① Conversion en base 10
— $10110,1101_2$

$$\begin{aligned} 10110,1101_2 &= 1 \times 2^4 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-4} \\ &= 2^4 + 2^2 + 2^1 + \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^4} \\ &= 16 + 4 + 2 + \frac{1}{2} + \frac{1}{4} + \frac{1}{16} \\ &= 22,8125 \end{aligned} \tag{1}$$

On obtient donc $10110,1101_2 = 22,8125_{10}$.
— $7,361_8$

$$\begin{aligned} 7,361_8 &= 7 \times 8^0 + 3 \times 8^{-1} + 6 \times 8^{-2} + 1 \times 8^{-3} \\ &= 7 \times 1 + 3 \times \frac{1}{8} + 6 \times \frac{1}{64} + 1 \times \frac{1}{512} \\ &= 7 + \frac{3}{8} + \frac{3}{32} + \frac{1}{512} \\ &= 7,470703125_{10} \end{aligned}$$

On obtient donc $7,361_8 = 7,470703125_{10}$.

Autre approche possible : utiliser la notation scientifique
- $10110,1101_2 = 101101101_2 \times 2^{-4} = 365 \times \frac{1}{16} = 22,8125_{10}$;
- $7,361_8 = 7361_8 \times 8^{-3} = 3825 \times \frac{1}{512} = 7,470703125_{10}$.

② Conversion en base 2

— $106,6875_{10}$
  - Partie entière : $106_{10} = 01101010_2$ sur 8 bits.
  - Partie fractionnaire :

$$\begin{aligned}
0,6875 \times 2 &= 1,375 \rightarrow 1 \\
0,375 \times 2 &= 0,75 \rightarrow 0 \\
0,75 \times 2 &= 1,5 \rightarrow 1 \\
0,5 \times 2 &= 1,0 \rightarrow 1
\end{aligned}$$

  d'où $0,6875_{10} = 0,10110000_2$ sur 8 bits.

  On obtient donc $106,6875_{10} = 01101010,10110000_2$.

— $54,4_6$

  Il faut tout d'abord convertir le nombre en base 10 :

$$54,4_6 = 5 \times 6^1 + 4 \times 6^0 + 4 \times 6^{-1} = 34,6666..._{10}$$

  - Partie entière : $54_6 = 34_{10} = 100010_2$.
  - Partie fractionnaire :

$$\begin{aligned}
0,6666 \times 2 &= 1,3332 \rightarrow 1 \\
0,3332 \times 2 &= 0,6664 \rightarrow 0 \\
0,6664 \times 2 &= 1,3328 \rightarrow 1 \\
0,3328 \times 2 &= 0,6656 \rightarrow 0 \\
0,6656 \times 2 &= 1,3312 \rightarrow 1 \\
\text{etc.} &= \rightarrow 0 \\
&= \rightarrow 1 \\
&= \rightarrow 0
\end{aligned}$$

  d'où $0,6666_{10} = 0,10101010_2$ sur 8 bits.

  On obtient donc $54,4_6 = 00100010,10101010$, soit une valeur approchée.

2.2 Représentation en virgule flottante

Correction

① Représention IEEE 754 simple précision d'un réel.

(a) Conversion en base 2
  - Partie entière : $281_{10} = 100011001_2$.
  - Partie fractionnaire :

$$\begin{aligned}
0,34375 \times 2 &= 0,6875 \rightarrow 0 \\
0,6875 \times 2 &= 1,375 \rightarrow 1 \\
0,375 \times 2 &= 0,75 \rightarrow 0 \\
0,75 \times 2 &= 1,5 \rightarrow 1 \\
0,5 \times 2 &= 1,0 \rightarrow 1
\end{aligned}$$

  d'où $0,34375_{10} = 0,01011_2$.

On obtient donc $281,34375_{10} = 100011001,01011_2$.

(b) Norme IEEE 754

On réécrit le nombre avec la représentation en virgule flottante, en normalisant la mantisse :

$$281,34375_{10} = 1,0001100101011_2 \times 2^8.$$

D'où :

— Signe = + $\rightarrow S = 0$.

— Exposant

$$E_{\text{réel}} = 8 \rightarrow E_{\text{biaisé}} = 8 + 127 = 135,$$

et $135_{10} = 10000111_2 = E$ sur 8 bits.

— Mantisse = $1,0001100101011_2 \rightarrow M = 00011001010110\ldots0$.

et donc on abouti au code suivant :

4	3	8	C	A	C	0 0
0 1 0 0	0 0 1 1	1 0 0 0	1 1 0 0	1 0 1 0	1 1 0 0	0 ... 0

S (1 bit) — E (8 bits) — M (23 bits)

Ainsi en IEEE 754 simple précision $438CAC00_{16}$ code le réel $281,34375_{10}$.

② Valeur d'un réel représenté en IEEE 754 simple précision.

(a) Norme IEEE 754

On identifie le Signe, l'Exposant et la Mantisse.

C	1	6	5	0	0 0 0	
1 1 0 0	0 0 0 1	0 1 1 0	0 1 0 1	0 0 0 0	0 ... 0	

S — E — M

D'où :

— $S = 1 \rightarrow$ Signe = -.

— $E = 10000010_2 = 130_{10} = E_{\text{biaisé}}$

$$E_{\text{biaisé}} = 130 \rightarrow E_{\text{réel}} = 130 - 127 = 3$$

— $M = 110010100\ldots0 \rightarrow$ Mantisse = $1,1100101_2$.

et donc on abouti à la représentation en virgule flottante suivante :

$$-1,1100101_2 \times 2^3 = -1110,0101_2 = -11100101 \times 2^{-4} = -14,3125$$