TD4 - Représentation des entiers relatifs et des réels

Représentation en virgule flottante

Correction

Représentation IEEE 754 simple précision d'un réel.

Il faut tout d'abord convertir le nombre en base 2, puis le réécrire en normalisant la mantisse sous la forme attendue. Il reste ensuite, pour définir la séquence binaire sur 32 bits représentant le nombre en IEEE 754, à identifier :

— *le **S**igne → occupe 1 bit (le bit de poids fort) ;*
— *l'**E**xposant (biaisé) → occupe 8 bits ;*
— *la **M**antisse → occupe les derniers 23 bits.*

① Conversion en base 2
— Partie entière : $187_{10} = 10111011_2$
— Partie fractionnaire :

$$\begin{aligned} 0,4375 \times 2 &= 0,875 \to 0 \\ 0,875 \times 2 &= 1,75 \to 1 \\ 0,75 \times 2 &= 1,5 \to 1 \\ 0,5 \times 2 &= 1,0 \to 1 \end{aligned}$$

d'où $0,4375_{10} = 0,0111_2$.

On obtient donc $187,4375_{10} = 10111011,0111_2$.

② Norme IEEE 754
On réécrit le nombre avec la représentation en virgule flottante, en normalisant la mantisse (*mise sous la forme* $1,\ldots$) :

$$\begin{aligned} 187,4375_{10} &= 10111011,0111_2 \times 2^0 \\ &= 1,01110110111_2 \times 2^7. \end{aligned}$$

D'où :
— Signe = + → $S = 0$.
— Exposant (*la valeur du biais en simple précision est de 127*)

$$E_{\text{réel}} = 7 \to E_{\text{biaisé}} = 7 + 127 = 134,$$

et $134_{10} = 10000110_2 = E$ sur 8 bits.
— Mantisse normalisée = $1,01110110111_2 \to M = 011101101110\ldots0$

Remarques :

— le bit à gauche de la virgule $(1, \ldots)$ n'est pas stocké dans M, d'où le fait qu'il soit appelé *bit caché* ;

— M doit faire 23 bits, il faut donc compléter à droite avec des zéros si nécessaire.

et donc on aboutit au code suivant :

$$\overbrace{0\,1\,0\,0}^{4}\,\overbrace{0\,0\,1\,1}^{3}\,\overbrace{0\,0\,1\,1}^{3}\,\overbrace{1\,0\,1\,1}^{B}\,\overbrace{0\,1\,1\,1}^{7}\,\overbrace{0\ldots\,0}^{0\,0\,0}$$

$$\underbrace{0}_{S}\,\underbrace{1\,0\,0\,0\,0\,1\,1\,0}_{E}\,\underbrace{0\,1\,1\,1\,0\,1\,1\,0\,1\,1\,1\,0\ldots\,0}_{M}$$

Ainsi en IEEE 754 simple précision $433B7000_{16}$ code le réel $187,4375_{10}$.

Le lien suivant : `http://www.h-schmidt.net/FloatConverter/IEEE754.html` vous permettra de vérifier que le résultat est correct. Vous pourrez également, en calculant la représentation de $29,3333_{10}$, constater que ce nombre peut pas être représenté de manière exacte.