TD 2: Représentation des flottants

Suivant la norme IEEE 754, la représentation d'un nombre flottant en simple (respectivement double) précision nécessite 32 (respectivement 64) bits (voir cours).

Exercice 1 : Donner une représentation flottante normalisée IEEE 754 en simple précision des nombres suivants :

(a) $\left(\frac{5}{32}\right)_{10}$
(b) -6.125_{10}
(c) 3.5_{10}
(d) -0.75_{10}
(e) -8_{10}
(f) 1_{10}
(g) $-\left(\frac{1}{16}\right)_{10}$
(h) 0
(i) 0.1_{10}
(j) 12.875_{10}

Exercice 2 : Quel est le nombre en base 10 représenté par le mot de 32 bits suivant :

(a) 1 | 1000 0001 | 0100 0000 0000 0000 0000 000
(b) 0 | 1000 0101 | 1001 0001 0000 0000 0000 000
(c) 1 | 0111 1101 | 1100 0000 0000 0000 0000 000

Exercice 3 : Quels sont, en valeur absolue, le plus grand et le plus petit nombres représentables avec la simple précision IEEE 754 ? On rappelle que le cadre de la normalisation impose que l'exposant décalé d'un nombre non nul et non infini varie entre 1 et 254 pour un flottant.
Hors normalisation, il est possible de représenter des flottants avec l'exposant décalé nul. Dans ce cas, il n'y a plus de bit caché dans la mantisse

Exercice 4 : Les concepteurs de cette norme souhaitaient définir une représentation qui puisse aussi être facilement soumise à des opérations entières (par exemple comparaison de deux nombres).

(a) Pourquoi la norme IEEE 754 place le bit signe en premier ? l'exposant avant la mantisse ?
(b) Pourquoi l'exposant est-il codé par excédent 127 et non par le complément à 2 ?

Exercice 5 : Effectuer les opérations sur les flottants suivants, représentés selon la norme IEEE754 simple précision :

(a) $x+y$ avec $x = 0$ | 0111 1100 | 01000000000000000000000 et $y = 0$ | 1000 0001 | 10001000000000000000000
(b) $x+y$ avec $x = 1$ | 0111 1100 | 01000000000000000000000 et $y = 1$ | 1000 0000 | 11000000000000000000000
(c) $x+y$ avec $x = 0$ | 0111 1101 | 10000000000000000000000 et $y = 0$ | 0111 1100 | 00000000000000000000000
(d) $x \times y$ avec $x = 1$ | 0111 1101 | 10011000000000000000000 et $y = 0$ | 0111 1100 | 00000000000000000000000