Devoir sur la représentation des nombres en IEEE 754

EBANDA MBARGA Eric Jordan
Matricule : 21T2462

1. Identification des nombres limites en simple et double précision

En Simple Précision

- **Plus petit nombre positif représentable (dénormalisé)** : Environ 1.4×10^{-45}. Ce nombre est obtenu avec un exposant nul (indiquant un nombre dénormalisé) et la plus petite mantisse possible, soit 2^{-149}.

- **Plus petit nombre négatif représentable (dénormalisé)** : Environ -1.4×10^{-45}. Ce nombre est similaire au cas positif, avec le signe inversé.

- **Plus grand nombre positif représentable** : Environ 3.4×10^{38}. Ce nombre est obtenu avec un exposant de 254 et une mantisse composée uniquement de 1, soit calculé par $(2 - 2^{-23}) \times 2^{127}$.

- **Plus grand nombre négatif représentable** : Environ -3.4×10^{38}, avec une valeur similaire au cas positif mais un signe opposé.

En Double Précision

- **Plus petit nombre positif représentable (dénormalisé)** : Environ 4.9×10^{-324}, avec un exposant nul et la plus petite mantisse possible, soit 2^{-1074}.

- **Plus petit nombre négatif représentable (dénormalisé)** : Environ -4.9×10^{-324}, similaire au cas positif avec un signe inversé.

- **Plus grand nombre positif représentable** : Environ 1.8×10^{308}, calculé avec un exposant de 1023 et la mantisse maximale, soit $(2 - 2^{-52}) \times 2^{1023}$.

- **Plus grand nombre négatif représentable** : Environ -1.8×10^{308}.

2. Calcul des écarts entre nombres consécutifs

En Simple Précision

- **Plus petit écart entre deux nombres consécutifs** : Environ 1.4×10^{-45}, correspondant à la différence entre le plus petit nombre positif représentable, 2^{-149}, et zéro.

- **Plus grand écart entre deux nombres consécutifs** : Environ 2×10^{31}, entre les deux plus grands nombres consécutifs proches de 3.4×10^{38} et $3.4 \times 10^{38} - 2^{104}$.

En Double Précision

- **Plus petit écart entre deux nombres consécutifs** : Environ 4.9×10^{-324}, soit la différence entre 2^{-1074} et zéro.
- **Plus grand écart entre deux nombres consécutifs** : Environ 2×10^{292}, entre 1.8×10^{308} et $1.8 \times 10^{308} - 2^{971}$.

3. Gestion des cas exceptionnels selon la norme IEEE 754

- **Overflow (dépassement de capacité)** : Se produit lorsqu'un calcul dépasse le nombre maximal représentable. Il est alors représenté par $+\infty$ ou $-\infty$.
- **Underflow (sous-capacité)** : Lorsque le calcul produit un résultat inférieur au minimum normalisé, on utilise des nombres dénormalisés pour représenter ce résultat.
- **Exposant nul** : Si la mantisse est non nulle, cela indique un nombre dénormalisé. Si la mantisse est également nulle, cela représente zéro.
- **Mantisse nulle** : Si l'exposant est différent de zéro, cela correspond à une puissance de 2. Si l'exposant est nul, cela représente zéro.

4. Exposants valides et cas d'exception

- **Nombre d'exposants valides** :
  - **Simple précision** : 254, allant de 1 à 254.
  - **Double précision** : 2046, allant de 1 à 2046.
- **Exposants réservés** :
  - L'exposant 0 est réservé pour les nombres dénormalisés ou zéro.
  - L'exposant maximum (255 en simple précision et 2047 en double précision) est réservé pour représenter $+\infty$, $-\infty$, ou NaN.

5. Exercice 4 du support de cours

Question : Quelle est l'erreur commise lors du stockage d'une puissance de 2 dans l'intervalle $[minF, maxF]$?

Réponse : Lorsqu'une puissance de 2 est stockée dans cet intervalle, l'erreur de stockage est nulle. Cependant, en dehors de cet intervalle, l'erreur peut devenir infinie (en cas d'overflow) ou être limitée (en cas d'underflow).