

Logarithme avec arrondi correct (stage de M2)

Lieu. INRIA/LORIA, Nancy, www.loria.fr.

Encadrant. Paul Zimmermann, directeur de recherche, équipe Caramba, Paul.Zimmermann@inria.fr.

Contexte. Le standard IEEE 754 définit ce qu'on appelle l'« arrondi correct ». Étant donné une fonction mathématique f et un nombre flottant x , l'arrondi correct de $f(x)$ est le nombre flottant y le plus proche de $f(x)$ selon le mode d'arrondi donné (au plus proche, vers zéro, vers $-\infty$ ou vers $+\infty$). Si le standard IEEE 754 impose l'arrondi correct pour les quatre opérations arithmétiques de base (addition, soustraction, multiplication, division) et la racine carrée, il ne fait que le recommander pour les fonctions mathématiques (exp, sin, pow, ...) Les bibliothèques mathématiques couramment utilisées ne garantissent pas l'arrondi correct [1]. Le projet CORE-MATH vise à produire des implantations avec arrondi correct, en vue d'une intégration dans ces bibliothèques mathématiques.

Objectif du stage. L'objectif du stage est d'implanter un logarithme avec arrondi correct, en utilisant pour les calculs internes des opérations sur des entiers de 64 bits. En effet, il a été démontré dans [2] que l'utilisation d'opérations sur des entiers est très compétitive par rapport à une implantation utilisant uniquement des nombres flottants. On commencera par reproduire les résultats de [2] pour la double précision (format `binary64` de IEEE 754), puis on les étendra à la double précision étendue (type `long double` sur processeur x86) et à la quadruple précision (format `binary128`). Le code produit au cours de ce stage sera diffusé sous une licence open-source, et pourra être intégré dans le cadre du projet CORE-MATH aux bibliothèques mathématiques existantes.

Références

- [1] INNOCENTE, V., AND ZIMMERMANN, P. Accuracy of mathematical functions in single, double, extended double and quadruple precision. <https://members.loria.fr/PZimmermann/papers/accuracy.pdf>, 2021. Version of September 7, 22 pages.
- [2] LE MAIRE, J., BRUNIE, N., DE DINECHIN, F., AND MULLER, J.-M. Computing floating-point logarithms with fixed-point operations. In *23rd IEEE Symposium on Computer Arithmetic* (Santa Clara, United States, 2016), P. Montuschi, M. J. Schulte, J. Hormigo, S. F. Oberman, and N. Revol, Eds., IEEE, pp. 156–163.