Fast and Accurate Midpoint-Radius Interval Operations in Rounding to Nearest

Naoya Yamanaka[1), **Shin'ichi Oishi**[2)]

[1)]Research Institute for Science and Engineering, Waseda University, Tokyo, 169-8555, Japan,
naoya_yamanaka@suou.waseda.jp

[2)]Faculty of Science and Engineering, Waseda University, Tokyo, 169-8555 Japan,
oishi@waseda.jp

Keywords: *interval arithmetic; accurate computation; rounding to nearest.*

Abstract

Numerical computations on a computer are plagued by the problem of roundoff error and its accumulation. Very often we trust results obtained using floating point computations, whose truth we are not certain about. Verifying the correctness of such results may be a very difficult task, to solve which various mathematical or programming tools may be used.

Multi-precision computation has a variety of application areas. Because of this, many arbitrary precision algorithms and libraries have been developed. Bailey and Briggs have developed algorithms and software for "doubled double" precision, twice the double precision. They used the multiple-component format, where a doubled double number is represented as an unevaluated sum of a leading double and a trailing double. The doubled double number represents the exact sum $a = a_0 + a_1$, where a_0 is the most significant component. We have designed and implemented algorithms for basic arithmetic operations.

In this paper we will describe fast and efficient algorithms to compute the interval based on doubled double numbers containing the true interval. The interval arithmetic is a method of finding the possible values of a result by performing a computation on a manner which preserves these bounds, and thus developing numerical method that yield reliable results. Proposed algorithms are working in rounding to nearest, so that they don't need to take time for changing rounding mode. Instead of changing rounding modes, they evaluate the rounding error of the approximate value in rounding to nearest mode, and find an interval represented by doubled double numbers including the true interval. By numerical experiments it is shown that the algorithms are faster compared to the computation time of the existing algorithm.

References

[1] D. E. Knuth: The Art of Computer Programming: Seminumerical Algorithms, vol. 2, Addison-Wesley, Reading, Massachusetts, 1969.

[2] T. J. Dekker: A floating-point technique for extending the available precision, Numer. Math., 18 (1971), pp. 224–242.

[3] T. Ogita, S. M. Rump, S. Oishi: Accurate sum and dot product, SIAM J. Sci. Comput., 26:6 (2005), pp. 1955–1988.

[4] Y. Hida, X. S. Li, D. Bailey: Library for Double-Double and Quad-Double Arithmetic http://crd.lbl.gov/~dhbailey/mpdist/