Numerical Software: Design, Analysis and Verification
4–6 July 2012, Santander (Spain).

On the accuracy of an algorithm due to Kahan for evaluating ad-bc

Jean-Michel Muller[1]

[1] *INRIA - LIP, ENS de Lyon,*
University of Lyon, France

emails: `Jean-Michel.Muller@ens-lyon.fr`

Abstract

We provide a detailed analysis of Kahan's algorithm for the accurate computation of the determinant of a 2×2 matrix. This algorithm requires the availability of a fused multiply-add instruction. Assuming radix-β, precision-p floating-point arithmetic with β even, $p \geq 2$, and barring overflow or underflow we show that the absolute error of Kahan's algorithm is bounded by $(\beta + 1)/2$ ulps of the exact result and that the relative error is bounded by $2u$ with $u = \frac{1}{2}\beta^{1-p}$ the unit roundoff. Furthermore, we provide input values showing that i) when $\beta/2$ is odd—which holds for 2 and 10, the two radices that matter in practice—, the absolute error bound is optimal; ii) the relative error bound is asymptotically optimal, that is, for such input the ratio (relative error)$/2u$ has the form $1 - O(\beta^{-p})$. We also give relative error bounds parametrized by the relative order of magnitude of the two products in the determinant, and we investigate whether the error bounds can be improved when adding constraints: When the products in the determinant have opposite signs, which covers the computation of a sum of squares, or when Kahan's algorithm is used for computing the discriminant of a quadratic equation. This work is to appear in the journal "Mathematics of Computation"