Qeios

Peer Review

Review of: "Evaluation of Bfloat16, Posit, and Takum Arithmetics in Sparse Linear Solvers"

Daniel Lemire[1]

1. Data Science Research Center, Télé-université, Québec, Canada

The paper explores the effectiveness of alternative number formats (bfloat16, posit, and takum arithmetic) in solving sparse linear systems, which are crucial in numerous computational fields like structural analysis, circuit simulation, and machine learning. Traditional IEEE 754 floating-point numbers are compared with these emerging formats, particularly in the context of LU, QR, GMRES solvers, and Mixed Precision Iterative Refinement (MPIR).

The article highlights the advantage of tapered precision in posits and takums, which offers better accuracy in direct solvers and reduces iteration counts in iterative solvers. Takum arithmetic, in particular, provides an extensive dynamic range even at low precision, which is crucial for various applications where precision can be traded for computational efficiency.

The study used the SuiteSparse Matrix Collection for a diverse set of real-world matrices.

This paper provides a comprehensive evaluation of bfloat16, posit, and takum arithmetics in the context of sparse linear solvers, showing clear advantages of the latter two formats in terms of accuracy, stability, and efficiency.

It is elegant and clear. The experiments are thorough and supported by good analysis.

Declarations

Potential competing interests: No potential competing interests to declare.

doi.org/10.32388/RKBWUJ