Mixed precision numerical methods for solving large systems of ordinary differential equations

Mouhamad AL SAYED ALI, Inria Lyon - Villeurbanne
Samuel BERNARD, Inria Lyon - Villeurbanne
Arsène MARZORATI, Inria Lyon - Villeurbanne
Jonathan ROUZAUD-CORNABAS, Université de Lyon, INSA-Lyon, Inria, CNRS
Université Claude Bernard Lyon 1, ECL, Université Lumière Lyon 2, LIRIS UMR5205 - Lyon

On modern architectures, the performance of $32-bit$ (single precision) operations is often at least twice as fast as the performance of $64-bit$ (double precision) operations (see *e.g.* [1]). By using a combination of $32-bit$ and $64-bit$ floating point arithmetic (mixed precision), we can design numerical methods that run faster and use less memory while limiting the loss of arithmetic precision due to the use of less precise numerical format. Furthermore, future architectures will add more floating point precisions such as different types of $16-bit$ half precision. These new formats should be used in the future to even improve the performance of numerical methods.
Here we study the use of mixed precision in solving large systems of ordinary differential equations (ODE) by both explicit and implicit schemes. For the explicit schemes, for example (explicit Runge-Kutta (RK2, RK4, RK6), Adams Bashforth,...), lowering the precision of the intermediate stages brings increased speed and reduced communication and energy costs, but it produces results of correspondingly low accuracy.
Implicit integration schemes require, at each integration step, the solution of a large nonlinear system. The nonlinear system is solved by the Newton method that leads to a set of linear systems involving the Jacobian matrix of the ODE and are solved by Krylov subspace methods. The convergence of the whole process relies on the quality of the initial solutions for both the inexact Newton iteration and the linear systems. To improve global convergence, line search and trust region algorithms can be used to improve initial solutions. We explore some approaches in reducing the arithmetic precision in the resolution of the nonlinear system to accelerate the numerical solution of the ODE. These approaches combine the performance of lower precision arithmetic with the accuracy of higher precision arithmetic. We have tested our results on several models (the Neural Field model, the multiscale mathematical model for the regulation of the cell cycle by the circadian clock which is relatively stiff). The numerical results show that parallel schemes running in single precision are up to 2 times faster than thoses running in double precision, but they do not provide sufficient accuracy. However, the numerical results show that the parallel schemes running in Mixed precision are up to 1.5 times faster than those running in double precision with sufficient accuracy. The numerical results also show that the parallel implicit schemes in mixed precision are more accurate and faster than the explicit ones.

[1] N. J. Higham, T. Mary. Mixed precision algorithms in numerical linear algebra. Acta Numerica, **31**, 347–414, 2022.

Contact : `mouhamad.al-sayed-ali@inria.fr`