Recent advances in mixed-precision low-rank tensor approximations

Eda Oktay*[1]

[1]Max Planck Institute for Dynamics of Complex Technical Systems – Germany

Abstract

Low-rank tensor decompositions are widely used in scientific computing, data analysis, and engineering applications due to their advantages in memory efficiency and scalability when handling high-dimensional data. These benefits have motivated the development of numerous decomposition techniques and their application across various fields, such as quantum physics and machine learning. However, constructing such methods and performing operations within these structures can be computationally intensive, often due to the high cost of tensor construction and the complexity of numerical computations.
To overcome these challenges, high-performance computing (HPC) strategies such as parallelism, randomized algorithms, and mixed-precision arithmetic have become essential for improving computational performance and reducing memory usage and communication overhead in (multi)linear algebra operations. In this talk, we present recent advances in high-performance low-rank tensor decompositions, focusing on mixed-precision techniques. We explore how combining low- and high-precision arithmetic enables significant acceleration without sacrificing accuracy and discuss their potential for developing large-scale applications in scientific computing.

*Speaker