Communication-avoiding TT-tensor orthogonalization and rounding procedures

Hussam Al Daas*[1], Gray Ballard[2], and Peter Benner[1]

[1]Max Planck Institute for Dynamics of Complex Technical Systems, Department of Computational Methods in Systems and Control Theory, Magdeburg, Germany
[2]Department of Computer Science Wake Forest University Winston-Salem, North Carolina, USA

High dimensional data arise in various areas of scientific computing. Parametric PDEs, molecular simulations, and classification are examples among many others [2–4]. For this reason, low-rank tensor approaches gained intensive attention of researchers in recent years. These methods allow to store the data implicitly and perform arithmetic operations on them with a reasonable complexity avoiding the curse of dimensionality. Most of the research focused on the establishment of representation formats and their corresponding arithmetic operations that reduce the floating point operations complexity [1, 3, 5]. However, few work considered efficient parallel algorithms for tensor computations.

In this work, we present communication-avoiding algorithms for tensors represented in tensor train (TT) format. Left and right orthogonalization procedures play an important role in most computations with TT tensors, e.g., during the projection step in the Alternating Least Squares method, rounding of formal structures, etc. We analyze data distribution and communication cost of the orthogonalization and rounding procedures. Due to the sequential scheme of TT tensors with respect to the modes, the performance of parallel algorithms becomes quickly communication bounded when increasing the number of modes, d. To tackle this issue, we use a mixed representation TT-Tucker to which it was pointed out in the literature [5] and introduce a communication-avoiding orthogonalization and rounding corresponding procedures.

A factor of d in terms of number of messages is saved with respect to the TT variant. Numerical experiments on large number of processes demonstrate the scalability of the proposed methods.

References

[1] W. Hackbusch and S. Kühn. A new scheme for the tensor representation. *J. Fourier Anal. Appl.*, 15(5):706–722, 2009.

[2] Boris N. Khoromskij. *Tensor numerical methods in scientific computing*, volume 19 of *Radon Series on Computational and Applied Mathematics*. De Gruyter, Berlin, 2018.

[3] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Rev.*, 51(3):455–500, 2009.

[4] Daniel Kressner and Christine Tobler. Krylov subspace methods for linear systems with tensor product structure. *SIAM J. Matrix Anal. Appl.*, 31(4):1688–1714, 2009/10.

[5] I. V. Oseledets. Tensor-train decomposition. *SIAM J. Sci. Comput.*, 33(5):2295–2317, 2011.

*aldaas@mpi-magdeburg.mpg.de, https://www.aldaas.net/