QR Decomposition of Block Low-Rank Matrices

Muhammad Ridwan Apriansyah
School of Computing,
Tokyo Institute of Technology
Tokyo, Japan
apriansyah.m.aa@m.titech.ac.jp

Rio Yokota
Global Scientific Information Center,
Tokyo Institute of Technology
Tokyo, Japan
rioyokota@gsic.titech.ac.jp

1 INTRODUCTION

QR decomposition is a fundamental operation in linear algebra, and is often used in scientific computations. The cost of $O(N^2)$ storage and $O(N^3)$ computations to perform QR decomposition could be reduced if the matrix is approximated using hierarchical low-rank approximation. In this work, we study the QR decomposition of block low-rank matrices (BLR-matrices), which is a simple variant of hierarchical matrices. BLR-matrices allows us to express dense matrix as blocks of full rank and low-rank matrices.

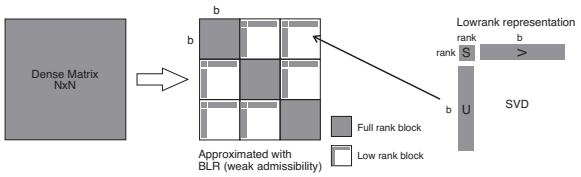


Figure 1: A dense matrix expressed as block low-rank matrix

2 METHOD

2.1 Blockwise QR Decomposition

We extend the tiled QR decomposition algorithm[1] to work on BLR-matrices. The functions are based on LAPACK subroutines.

Algorithm 1: Blockwise QR decomposition

Input: BLR matrix A of size $p \times q$ with block size b
Output: BLR matrix A and block matrix T

for k = *1 to min(p,q)* **do**
  GEQRT($A_{kk}, Y_{kk}, T_{kk}, T_{kk}$);
  for j = *k+1 to q* **do**
    LARFB($A_{kj}, Y_{kk}, T_{kk}, R_{kj}$);
  end
  for i = *k+1 to p* **do**
    TPQRT($R_{kk}, A_{ik}, Y_{ik}, T_{ik}$);
    for j = *k+1 to q* **do**
      TPMQRT($R_{kj}, A_{ij}, Y_{ik}, T_{ik}$);
    end
  end
end

2.2 Parallel Blockwise QR Decomposition

Some of the operations in Algorithm 1 are independent of each other and can be done in parallel. By formulating each function call as a task and considering the dependencies between them, we can generate the dependency graph and execute the tasks in parallel based on that.

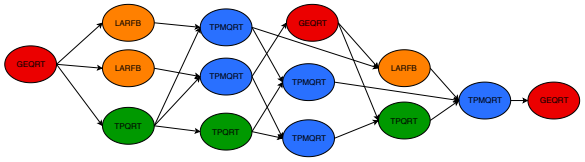
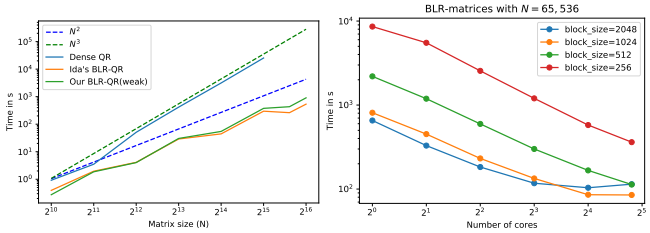


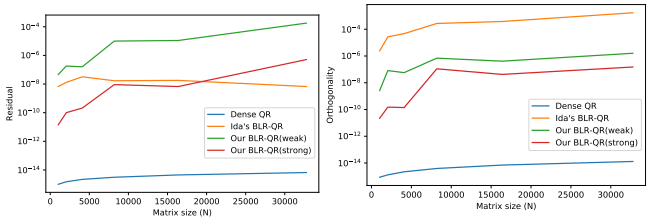
Figure 2: Dependency graph of Algorithm 1 on 3x3 BLR-matrix

3 RESULTS

The experiments were conducted using TSUBAME 3.0 supercomputer. Each node has up to 28 cores. We compare our method with the dense QR decomposition routine of Intel MKL and an existing BLR-QR method[2]. OpenMP 4.5 is used for parallelization.



(a) Execution time **(b) Parallel scalability**



(c) Residual $\frac{\|A-QR\|}{\|A\|}$ **(d) Orthogonality** $\frac{\|I-Q^TQ\|}{\sqrt{N}}$

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Numbers JP18H03248, JP17H01749. This work is supported by "Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures" in Japan (Project ID: jh190043-NAHI).

REFERENCES

[1] Brian C. Gunter and Robert A. Van De Geijn. 2005. Parallel Out-of-core Computation and Updating of the QR Factorization. *ACM Trans. Math. Softw.* 31, 1 (March 2005), 60–78. https://doi.org/10.1145/1055531.1055534

[2] Akihiro Ida, Hiroshi Nakashima, Tasuku Hiraishi, Ichitaro Yamazaki, Rio Yokota, and Takeshi Iwashita. 2019. QR Factorization of Block Low-rank Matrices with Weak Admissibility Condition. *Journal of Information Processing* 27 (2019), 831–839. https://doi.org/10.2197/ipsjjip.27.831