The 8th European Congress on Computational Methods in Applied Sciences and Engineering
ECCOMAS Congress 2022
5-9 June 2022, Oslo, Norway

RACE: Speeding Up Sparse Iterative Solvers Using Level-Based Parallelization and Blocking Techniques

Christie Alappat*, Georg Hager and Gerhard Wellein

Erlangen National High Performance Computing Center,
Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany
Email: {christie.alappat, georg.hager, gerhard.wellein}@fau.de

Keywords: *HPC, Recent Trends, Parallel algorithms, Performance, Sparse Iterative Solvers*

Sparse linear iterative solvers are indispensable for large-scale simulations in the fields of computational fluid dynamics and solid mechanics. In this talk, we present methods to accelerate some of the existing solvers by using the concept of *levels* as developed in the context of our RACE library framework [1]. Levels are constructed using breadth-first search (BFS) on the graph related to the underlying sparse matrix and form the basis of two performance optimization strategies: *parallelization* and *blocking*.

The *parallelization* strategy is designed for sparse kernels that have inherent data dependencies. In this case, the level-based approach enables efficient shared-memory parallel execution of the kernel by resolving the dependencies. In [1], the application of the method to symmetric sparse-matrix vector multiplication (SpMV) has been demonstrated. On modern multicore CPU socket, the method achieved an overall performance speedup of 1.5× compared to other existing algorithms. The approach is not limited to symmetric SpMV but can be applied to other routines, such as preconditioners like Gauss-Seidel and ICCG, or Kaczmarz solvers.

The second and novel performance optimization strategy using our level-based approach is *blocking*, where the matrix elements are blocked in the cache to allow for high spatial and temporal reuse. It finds its use in kernels like sparse-matrix-power vector multiplication and Chebyshev polynomial iterations, which perform repetitive back-to-back iterations without global synchronization in between. The method is highly effective and achieves performance levels of 50–100 GF/s on a single modern Intel or AMD multicore chip, providing speed-ups of typically 2–4× compared to a highly optimized classical SpMV implementation.

After introducing optimization strategies, the talk sheds light on the application of these optimized kernels in iterative solvers. To this end, we discuss the coupling of the RACE library with the Trilinos framework [2] and address the application to commonly found solvers like CG, GMRES, and its s-step variants, which benefit from our blocking optimization.

The work on RACE has been started in the ESSEX project within the DFG priority program SPPEXA and has received multiple international awards including the SC18 Best Student Poster Award and the second place in the 2019 ACM Student Research

Competition Grand Finals [3].

REFERENCES

[1] C. Alappat, A. Basermann, A. R. Bishop, H. Fehske, G. Hager, O. Schenk, J. Thies, and G. Wellein, "A recursive algebraic coloring technique for hardware-efficient symmetric sparse matrix-vector multiplication," *ACM Trans. Parallel Comput.*, vol. 7, June 2020.

[2] T. Trilinos Project Team, *The Trilinos Project Website*, 2020 (acccessed Dec 7, 2021). https://trilinos.github.io.

[3] ACM, "Acm student research competition," 2019. https://www.acm.org/media-center/2019/may/src-2019-grand-finals.