Communication Hiding Pipelined Krylov Methods

On Parallel Performance and Numerical Stability of Pipelined Conjugate Gradients

Siegfried COOLS [a,1], and Wim VANROOSE [a]

[a] *Applied Mathematics Research Group, Department of Mathematics and Computer Science, University of Antwerp, Middelheimlaan 1, 2020 Antwerp, BE*

Abstract. Pipelined Krylov subspace methods achieve significantly improved parallel scalability compared to standard Krylov methods on present-day HPC hardware. However, this typically comes at the cost of a reduced maximal attainable accuracy. This paper presents and compares several stabilized versions of the communication-hiding pipelined Conjugate Gradients method. The main novel contribution of this work is the reformulation of the multi-term recurrence pipelined CG algorithm by introducing shifts in the recursions for specific auxiliary variables. These shifts reduce the amplification of local rounding errors on the residual. Given a proper choice for the shift parameter, the amplification of local rounding errors is reduced and the resulting algorithm improves the attainable accuracy. Numerical results on a variety of SPD benchmark problems compare different stabilization techniques for the pipelined CG algorithm, showing that the stabilized algorithms are able to attain a high accuracy while displaying excellent parallel performance.

Keywords. Pipelined Krylov subspace methods, Latency hiding, Avoiding global communication, Numerical stability, Maximal attainable accuracy

1. Introduction

Krylov subspace methods [1,2] are well-known as efficient solvers for large scale linear systems $Ax = b$. These iterative algorithms constructs a sequence of approximate solutions $\{x_i\}_i$ with $x_i \in x_0 + \text{span}\{r_0, Ar_0, A^2 r_0, \ldots, A^{i-1} r_0\}$, where $r_0 = b - Ax_0$ is the initial residual. The Conjugate Gradient (CG) method [3], Alg. 1, which allows for the solution of systems with symmetric positive definite (SPD) matrices A, is generally considered as the first Krylov subspace method. Driven by the transition of hardware towards the exascale regime, research on the scalability of Krylov subspace methods on massively parallel architectures has recently (re)gained increasing attention [4,5]. Since for many applications the system matrix is sparse and thus inexpensive to apply, the main bottleneck for efficient parallel execution is typically not the sparse matrix-vector product (SPMV), but the communication overhead caused by global reductions in dot-product computations.

[1]Corresponding Author: S. Cools, Department of Mathematics & Computer Science, University of Antwerp. Address: Middelheimlaan 1, 2020 Antwerp, Belgium. E-mail: siegfried.cools@uantwerp.be.
S.C. acknowledges funding by Research Foundation Flanders (FWO) under grant application no. 12H4617N.