Designing Scalable Solvers for Enlarged Krylov Subspace Methods [a]

Shelby Lockhart
Amanda Bienz
William D. Gropp
Luke Olson

9th JLESC Workshop

[a] Grigori, L., Moufawad, S., & Nataf, F. (2016). Enlarged Krylov paper in References.

Current Solvers State

Current State and Motivation

- Current Solvers
  - AMG, CG and PCG (with AMG preconditioning) solvers
  - Node-aware SpMVs for communication reduction
- Blocked Problems
  - Oil reservoir simulations require solving a pressure block system which has nonzeroes dispersed in a block format, but the blocks are small (2x2 or 3x3)
  - We would like to boost this arithmetic intensity by increasing the block size via Enlarged Krylov Subspace methods - which would then allow us to take advantage of GPUs

Enlarged Krylov Solvers: Short Recurrence Enlarged - Conjugate Gradient

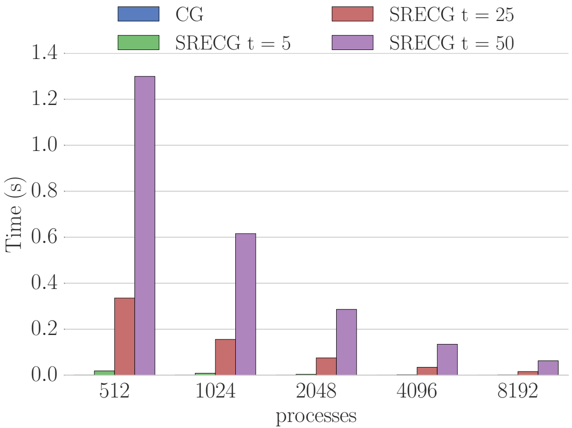
Short Recurrence Enlarged CG

Algorithm 1 SRE-CG

```
t := number of subdomains to split residual
r_0 := b − A x_0
for k = 0, 1, ..., until convg do
    if k == 0 then
        Let W_1 = τ(r_0, t)
        A-orthonormalize vectors of W_1
    else
        Let W_k = A W_{k−1}
        A-orthonormalize vectors of W_k
        against W_{k−1} and W_{k−2} for k > 2
        A-orthonormalize vectors of W_k
    end if
    α̃_k := W_k^T r_k
    x_{k+1} := x_k + W_k α̃_k
    r_{k+1} := r_k − A W_k α̃_k
end for
```

- Only requires storage of two additional iterations' search directions
- Consists of single vector updates
- Does not require a system solve
- Greater risk of suffering from loss of A-orthogonality depending on chosen A-orthonormalization routine

Short Recurrence Enlarged CG Computation per Iteration



Solving a discontinuous Galerkin finite element discretization of the Laplace problem

$$-\triangle u = 1$$

with ≈ 1 million degrees of freedom

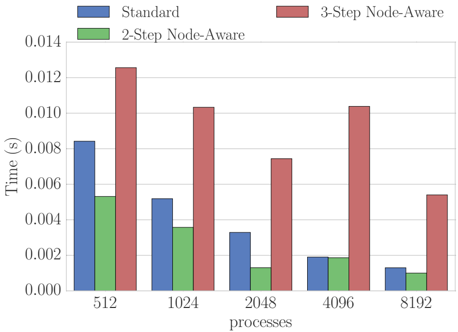
Short Recurrence Enlarged CG: Key Kernel

Algorithm 2 SRE-CG

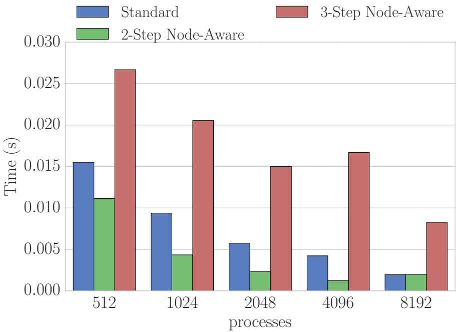
$t :=$ number of subdomains to split residual
$r_0 := b - Ax_0$
for $k = 0, 1, \ldots,$ until convg **do**
    if $k == 0$ **then**
        Let $W_1 = \tau(r_0)$
        A-orthonormalize vectors of W_1
    else
        Let $W_k = AW_{k-1}$
        A-orthonormalize vectors of W_k
        against W_{k-1} and W_{k-2} for $k > 2$
        A-orthonormalize vectors of W_k
    end if
    $\tilde{\alpha}_k := W_k^T r_k$
    $x_{k+1} := x_k + W_k \tilde{\alpha}_k$
    $r_{k+1} := r_k - AW_k \tilde{\alpha}_k$
end for

- Block Vector SpMVs occur throughout the algorithm including the A-orthonormalization routines
- Sparse - Dense Matrix multiplication with a tall and skinny dense matrix

Node-Aware Block Vector SpMV Profiling



$t = 25$



$t = 50$

Solving a discontinuous Galerkin finite element discretization of the Laplace problem

$$-\triangle u = 1$$

with ≈ 1 million degrees of freedom

Next Steps

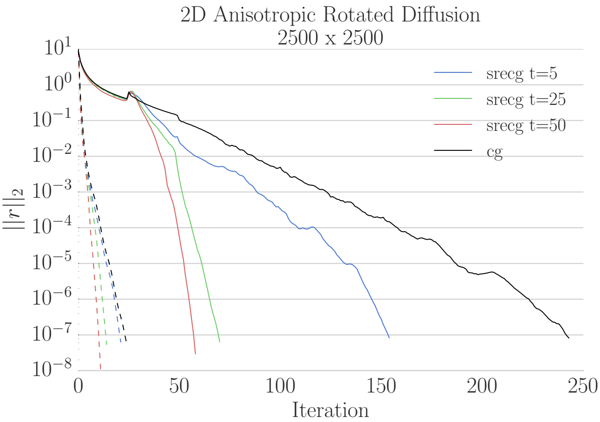
Preconditioned Short Recurrence Enlarged CG

Algorithm 3 Preconditioned SRE-CG

$t :=$ number of subdomains to split residual
$r_0 := b - Ax_0$
$\hat{r_0} := M^{-1} r_0$
for $k = 0, 1, \ldots,$ until convg **do**
    if $k == 0$ **then**
        Let $\hat{W}_1 = \tau(\hat{r_0}, t)$
        A-orthonormalize vectors of $\hat{W}_1$
    else
        Let $\hat{W}_k = M^{-1} A \hat{W}_{k-1}$
        A-orthonormalize vectors of $\hat{W}_k$
        against $\hat{W}_{k-1}$ and $\hat{W}_{k-2}$ for $k > 2$
        A-orthonormalize vectors of $\hat{W}_k$
    end if
    $\hat{\alpha}_k := \hat{W}_k^T \hat{r}_k$
    $\hat{x}_{k+1} := \hat{x}_k + \hat{W}_k \hat{\alpha}_k$
    $r_{k+1} := r_k - A \hat{W}_k \hat{\alpha}_k$
end for

- Finish extending AMG preconditioning to the block vector structure in RAPtor
- Look at designing a new Enlarged AMG algorithm as a preconditioner and standalone solver

Preconditioned Short Recurrence Enlarged CG



Problem Statement:

$$-div(\begin{bmatrix} cos(\frac{\pi}{4}) & -sin(\frac{\pi}{4}) \\ sin(\frac{\pi}{4}) & cos(\frac{\pi}{4}) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0.001 \end{bmatrix} \begin{bmatrix} cos(\frac{\pi}{4}) & sin(\frac{\pi}{4}) \\ -sin(\frac{\pi}{4}) & cos(\frac{\pi}{4}) \end{bmatrix} \nabla u) = f$$

Adding On-Node Parallelism to SRECG

- Currently, finishing adding shared memory parallelism with OpenMP to our SRECG solver for profiling
- Plans to port SRECG and AMG to GPUs in the coming months

Challenges Moving Forward

- Determine optimal A-orthonormalization routine
- Test other kernels of SRECG besides block-SpMV as candidates for node-aware communication
- Determine a model for communication that involves GPU communication, not just communication between CPUs
- Explore compression techniques to reduce message sizes where applicable

This research is part of the Blue Waters sustained petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications.

MFEM: Modular finite element methods library.
mfem.org.

Amanda Bienz, William D. Gropp, and Luke N. Olson.
Tapspmv: Topology-aware parallel sparse matrix vector multiplication.
CoRR, abs/1612.08060, 2016.

Amanda Bienz and Luke N. Olson.
RAPtor: parallel algebraic multigrid v0.1.
`https://github.com/raptor-library/raptor`, 2017.

Acknowledgements & References ii

Moufawad S. Grigori L. and Nataf F.

Enlarged krylov subspace conjugate gradient methods for reducing communication.

SIAM Journal on Matrix Analysis and Applications, 37(2):744–773, 2016.