Project Milestone Report: Symmetry-Exploiting Compiler Optimizations for MLIR Tensor Programs

GAURAV ARYA and APURVA GANDHI

1 Project Web Page

https://github.com/gaurav-arya/compilers-symmetry-final-project

2 Major Changes

While there have not been any major changes in our goals, as we discuss below in Section 5, there were some concurrent efforts from the EnzymeAD group at MIT. We are working with them to merge our changes with theirs and deduplicate some work.

3 What We Have We Accomplished So Far

(1) We wrote an initial analysis pass to detect symmetry generation and propogation for 2D tensors (matrices): https://github.com/EnzymeAD/Enzyme-JAX/compare/main...gaurav-arya:Enzyme-JAX:symmetric-result-analysis

(2) We later created a PR in the Enzyme-JAX repo to deduplicate our analysis pass with one they wrote. In particular, we added the following additional analyses:
   - We annotate that the result of transposing a symmetric matrix is still symmetric: https://github.com/EnzymeAD/Enzyme-JAX/pull/1626
   - We annotate that the result of applying StableHLO's `dot_general` op for $A * A$ will still be symmetric if A is symmetric: https://github.com/EnzymeAD/Enzyme-JAX/commit/1659f787d960ef5ae97b20d9afe72d5403921b64

(3) We implemented an optimization that takes advantage of the symmetry analysis. In particular, when performing a StableHLO `dot_general` op where at least one of the input matrices A is symmetric, we swap the reduction axes to row-major ones for that matrix (if not already in this form) to better exploit spatial locality. We can do this sort of 'implicit transpose' for an input matrix without changing the result only if we know that it is symmetric: https://github.com/EnzymeAD/Enzyme-JAX/commit/1a2a98a8b10efa18fd6a4c133c06154be430268d

(4) We verified that our symmetry analysis and optimizations work on example StableHLO MLIR: https://github.com/EnzymeAD/Enzyme-JAX/blob/173903fe8d8fd9883a01c3d4dacc13779301a3b4/test/lit_tests/structured_tensors/propagate_symmetric.mlir and https://github.com/EnzymeAD/Enzyme-JAX/blob/1a2a98a8b10efa18fd6a4c133c06154be430268d/test/lit_tests/structured_tensors/dot_general_symmetric.mlir

4 Meeting Our Milestone

Yes, as described in Section 3, we have met our milestone goals of (1) detecting and annotate/propogate symmetry generated in at least one tensor operation and (2) leveraging this annotation to perform at least one symmetry-exploiting optimization.

5 Surprises

After we wrote our initial pass for detecting symmetry for matrices, we learned of a similar PR contributed by the EnzymeAD group: https://github.com/EnzymeAD/Enzyme-JAX/pull/1549. We are now attending their weekly meetings and working more closely to try avoiding duplication of

Authors' Contact Information: Gaurav Arya, gauravar@cs.cmu.edu; Apurva Gandhi, apurvag@cs.cmu.edu.

work, and as described in Section 3, we have made a PR to add some additional scenarios to the analysis pass on top of their original version. Also, since the work at MIT mainly focuses on the matrix (2-dim tensor) case, we plan to contribute a generalization of the analysis that works for the N-dim tensor case, also adding a notion of partial symmetry.

6 Revised Schedule

- **Nov 20 - 26:** Implement at least one more symmetry-exploiting optimization pass and benchmark existing ones. Start working on generalizing the analysis pass for N-dim tensor case.
- **Nov 27 - Dec 3:** Finish up the more general N-dim tensor pass and use it to perform at least one optimization. Write final report and prepare poster.

We plan to pair code and work together on these items.

7 Resources Needed

We were able to build and test our changes to the Enzyme-JAX repo on our local machines. We do not anticipate the need of any additional resources.