Параллельная реализация конвейерного метода сопряжённых градиентов с использованием технологий MPI, OpenMP и CUDA

Недожогин Никита Сергеевич

старший преподаватель

Удмуртский Государственный Университет,

ИМИТиФ, кафедра вычислительной механики

Nedozhogin@inbox.ru

 В работе рассматривается подход к распараллеливанию конвейерной схемы метода сопряжённых градиентов [1,2] решения систем линейных алгебраических уравнений для multiGPU кластера. Такой кластер подразумевает узлы содержащие несколько CPU и GPU. Совершенно другая архитектура вычислительных устройств, производительность и модели программирования CPU и GPU представляют уникальные проблемы в гетерогенных системах.

 При решении систем линейных алгебраических уравнений, на каждом шаге метода сопряжённых градиентов существует три неявных точки синхронизации: дважды вычисляется скалярное произведение векторов и вычисление нормы невязки для проверки условия выхода. Особенностью конвейерного алгоритма является сокращение этих неявных точек синхронизации на каждой итерации метода. Это происходит за счёт объединения всех скалярных произведений в одном месте — в конце итерации. Недостатком метода являются дополнительные 5 векторных операций axpy.

 Для сокращения времени обмена сообщениями между узлами, конвейерный алгоритм позволяет перекрыть коммуникации с вычислениями операций axpy. Таким образом, после матрично-векторного произведения, в момент выполнения операций над векторами, происходит сбор результатов. А суммирование скалярных произведений выполняется параллельно операции предобуславливания системы.

 Для оценки масштабируемости полученной реализации использовался набор матриц, генерируемых аналогично тесту HPСG, в котором СЛАУ с сильно разреженной матрицей решается итерационным методом сопряженных градиентов. Полученные результаты показывают, что данный подход обладает сильной масштабируемостью только при достаточно больших размерах системы.

 Так же были проведены сравнения производительности с блочным классическим методом сопряжённых градиентов [3] на матрицах из библиотеки университета Флориды. В ходе экспериментов были сделаны выводы о том, что конвейерный метод обладает лучшей масштабируемостью, которая сохраняется как на маленьких, так и на больших системах.

 За помощь в работе автор выражает благодарность своему научному руководителю — д.ф-м.н. Копысову С.П.. Данная работа выполнена при финансовой поддержке УдГУ в рамках конкурса грантов «Научный потенциал», проект №2020-04-03.

1. Cornelis J., Cools S., Vanroose W. The Communication-Hiding Conjugate Gradient Method with Deep Pipelines // CoRR abs/1801.04728. 2018

2. Ghysels P., Vanroose W. Hiding global synchronization latency in the preconditioned Conjugate Gradient algorithm // Parallel Computing. 2014. V. 40. №7. P. 224-238.

3. Kopysov S., Kuzmin I., Nedozhogin N., Novikov A., Sagdeeva Y. Scalable hybrid implementation of the Schur complement method for multi-GPU systems // The Journal of Supercomputing. 2014. v. 69. P. 81-88