

АЛГОРИТМ ГЛОБАЛЬНОЙ ОПТИМИЗАЦИИ МЕТОДОМ СЕЛЕКТИВНОГО УСРЕДНЕНИЯ КООРДИНАТ С ИСПОЛЬЗОВАНИЕМ ГИБРИДНЫХ АРХИТЕКТУР ВИДА CPU + GPU

Новиков А.А.

**научный руководитель доктор технических наук, профессор Рубан А.И.
Сибирский федеральный университет**

Рассматривается решение задачи глобальной оптимизации путем применения метода селективного усреднения координат в области дискретно–непрерывных переменных. Параллелизация вычислений осуществляется с помощью гибридной (гетерогенной) архитектуры, на основе технологии NVidiaCUDA.

Основной тенденцией в современных вычислительных системах является повышение степени параллелизма (увеличение количества параллельных блоков) и эффективности параллельных вычислений (расширение команд, увеличение скорости работы и объема памяти и т.п.). Что привело к существенному изменению архитектуры высокопроизводительных вычислительных систем. Появившаяся в 2007 году технология NVidia CUDA, а позднее в 2008 году OpenCL позволили создать на основе существующих графических ускорителей высокопроизводительные вычислительные системы.

Выпуск специализированных графических ускорителей NVidia Tesla в 2008, предназначенных исключительно для высокопроизводительных вычислений, позволил оснащать узлы классических кластеров данными ускорителями, в результате чего широкое распространение также стали получать так называемые гибридные суперкомпьютеры. К примеру, в Топ50 Российской Федерации и СНГ7 производительных суперкомпьютеров оснащены графическими ускорителями, в том числе самый производительный.

Дальнейшие планы производителей микропроцессоров, имеют тенденцию к повышению значимости графических ускорителей в будущем. Новая архитектура процессоров AMD Fusion совмещает на одном кристалле графическое ядро вместе с вычислительными ядрами, используя общую память. Решение под кодовым названием Knight Corner от компании Intel позволит устанавливать платы с 50 дополнительными x86-совместимыми ядрами. Компания NVidia продолжает развивать линейку специализированных графических ускорителей NVidia Tesla, увеличивая их производительность каждый два года, в два-три раза.

В настоящее время гибридные архитектуры набирают популярность при моделировании и инженерных расчетах и применяются для решения следующих задач: разработка лекарств, проектирование (газодинамика), разведка нефти и газа, медицинская визуализация (КТ и МРТ), биржевая торговля, прогнозирование погоды, обработка спутниковых изображений. При этом обладают следующие преимущества по сравнению с традиционными высокопроизводительными системами основанными только на CPU:

- Высокая производительность, обеспечиваемая за счет большого количества параллельных вычислительных устройств;
- Низкое энергопотребление;
- Низкая удельная стоимость, за счет малого потребления электроэнергии и низкого тепловыделения;
- Экономия пространства центров обработки данных.

В тоже время поиск оптимального решения занимает все более значимую роль при решении инженерных задач – это различные задачи оптимального управления, потребления ресурсов, эффективного использования рабочего времени и аппроксимации поведения объекта с помощью математической модели. Словом, такие задачи возникают везде, где необходимо получить наилучший результат целевой функции в допустимой области, задаваемой конечным множеством ограничений – неравенств.

Нахождение оптимального решения связано с высокими вычислительными затратами, связано это с большим количеством параметров оказывающих влияние на систему в реальных инженерных задачах, высокой степенью нелинейности минимизируемой функции и наличия множества ограничений накладываемых как на параметры так и на саму минимизируемую функцию. Большая часть алгоритма глобальной оптимизации методом селективного усреднения координат может быть распараллелина. Поэтому применение гибридных архитектур может позволить решать сложные инженерные задачи в приемлемое время.

Постановка задачи

Рассматривается задача отыскания условного глобального минимум функции многих переменных с учетом ограничений в виде неравенств $\varphi_j(x, y) \leq 0, j = \overline{1, m_1}$.

$$\begin{cases} I(x, y) = \min_{x, y} \\ \varphi_j(x, y) \leq 0, j = \overline{1, m_1}; x \in X; y \in Y, Y \subset Z \end{cases}, \quad (1)$$

где $x = (x_1, \dots, x_k)$ – вектор непрерывные переменные;

$y = (y_1, \dots, y_l)$ – вектор дискретных переменных.

Априорная информация о характере глобального поведения целевой функции, отсутствует, целевая функция может быть многоэкстремальной, разрывной, не дифференцируемой и может быть искажена аддитивной помехой. Ограничения в виде неравенств, так же могут быть нелинейными, а их количество счетно. Далее под решением экстремальной задачи в ходе всей статьи будем понимать поиск минимума функции многих переменных. Для решения задачи (1) при отсутствии априорной информации о характере глобального поведения целевой функции и структуре её локальных экстремумов, неизбежно требуется применение методов, осуществляющих в том или ином виде перебор и сравнение “всех точек” области определения. Для практической оптимизации функций большого числа аргументов целесообразно привлекать вероятностные подходы. Все наиболее успешные методы решения проблемы глобальной оптимизации относятся к методам стохастической оптимизации.

Для поиска глобального экстремума в пространстве дискретно-непрерывных переменных в ходе работы использовался метод селективного усреднения координат. Данный метод изложен в монографии [2]. Исследование свойств метода селективного усреднения координат для пространства непрерывных переменных и его реализация рассмотрены в работе [1].

Для случая дискретно-непрерывных переменных, алгоритм поиска экстремума должен разделять обработку дискретных и непрерывных переменных. На начальном этапе поиска определяет количество возможных значений дискретных переменных в решаемой задаче. Далее, для каждого фиксированного набора значений дискретных переменных формируется функция качества, которая зависит только от непрерывных переменных. В результате исходная задача разбивается на m_2 нелинейных подзадач, где

$m_2 = \sum_{i=1}^l \mu_i$ – количество значений каждой дискретной переменной. Полученные

решения нелинейных подзадач с непрерывными параметрами для каждого из фиксированного значения дискретной переменной являются допустимыми решениями задачи глобальной оптимизации. Решение, при котором целевая функция принимает наименьшее значение, есть точка предполагаемого глобального экстремума.

Разработанная реализация метода селективного усреднения координат в области дискретно-непрерывных переменных состоит из двух частей: расчетная реализована на языке СИ для платформы Nvidia CUDA и управляющая реализована на языке Java. Взаимодействие осуществляется с помощью библиотеки JCuda (www.jcuda.de). Данный подход позволил в разы увеличить скорость вычислений по сравнению с аналогичной реализацией на языке Java и использования потоков для параллелизации вычислений центрального процессора (JavaThreadPool). Увеличение производительности в большей степени было достигнуто за счет большого количества параллельных блоков у графических ускорителей в сравнение с центральным процессором и более быстрой реализации генератора случайных чисел.

Выводы

Полученные в ходе работы результаты свидетельствуют о значительном превосходстве гибридные (гетерогенные) вычислительные системы (на примере Nvidia CUDA) над традиционными высокопроизводительными системами и возможности применения при решении широкого спектра инженерных задач.

Дальнейшее развитие

Основным направлением работ является развитие текущей программной системы для кластерных гибридных архитектур. С такими существующими технологиями параллельного программирования, такими как NVidia CUDA и OpenCL.

Планируется создание специального слоя абстракций, позволяющим взаимодействовать с SDK, а заменить обобщенным. С точки зрения программиста это будет означать появление дополнительных параллельных примитивов, которые будут скрывать за собой особенности работы с той или иной технологией гибридных архитектур.

Библиографические ссылки

1. Кузнецов А.В. Алгоритмы глобальной оптимизации функций в пространстве непрерывных переменных при наличии ограничений–неравенств: дис. канд. техн. наук: 05.13.01 Красноярск: КГТУ. – 2006. – С. 138 : 61 06–5/2857
2. Рубан А.И. Глобальная оптимизация методом усреднения координат. : моногр. Красноярск: ИПЦ КГТУ. – 2004. – 302 с.