

ВЫЧИСЛИТЕЛЬНЫЕ ТЕНЗОРНЫЕ МЕТОДЫ ДЛЯ СВЕРХБОЛЬШИХ ЗАДАЧ ФИЗИКИ, ХИМИИ, МАТЕМАТИКИ

И.В. Оселедец

Экзафлопные вычисления накладывают ряд существенных ограничений на используемые алгоритмы. Пожалуй, только методы типа Монте-Карло или задачи, в которых требуется перебрать варианты, можно тривиальным образом перенести на такие системы. Однако, такие методы обладают плохой масштабируемостью по числу степеней свободы: например, в задачах перебора экспоненциальная зависимость не является чем-то удивительным. Такая же экспоненциальная зависимость возникает при попытке решения различных многомерных задач химии, физики и математики: число параметров растет экспоненциально. Это не означает, что такие задачи невозможно решать. Например, в вычислительной химии широкое распространение получили базисы специального вида, и число базисных функций уже относительно небольшое. Однако, при этом теряется «идеальная масштабируемость» с точки зрения параллельных вычислений. Поэтому вопрос о создании класса методов которые как хорошо параллелизуются, так и не содержат экспоненциального по числу степеней свободы числа параметров, стоит особенно актуально.

Вычислительные тензорные методы --- пример таких методов, которые имеют большие перспективы. Данные представляются как многомерный массив, однако элементы этого массива не хранятся целиком, а представлены с помощью некоторого малопараметрического представления. В качестве такого представления используется ТТ-формат. Как при этом возможно распараллеливать соответствующие алгоритмы? Для примера, опишем процесс ТТ-интерполяции. В этом процесс вычисляется небольшое число элементов массива, потом по этим элементам адаптивно вычисляются, если это необходимо, те новые элементы, которые необходимо вычислить. Это похоже на метод Монте-Карло, только точки вычисляются с помощью детерминистического адаптивного алгоритма, что часто дает гораздо более высокую точность. Примеры задач, где применимы вычислительные тензорные методы включают задачи вычислительной химии, физики твердого тела, математической физики, финансовой математике.

В 2009 году в ряде работ Оселедцем И.В. и Тыртышниковым Е.Е. были предложены абсолютно новые представления для многомерных массивов, основанные на так называемом ТТ-формате (tensor train format). Он позволяет в ряде случаев существенно сократить размер памяти, требуемой для хранения многомерного массива. В отличие от известных подходов, основанных на каноническом разложении и разложении Таккера, ТТ-разложение не имеет «встроенной» экспоненциальной зависимости (как в разложении Таккера) и может быть вычислено устойчивым образом (в отличие от канонического разложения). Это делает его очень перспективным для решения многомерных задач. Были предложены все базовые алгоритмы линейной алгебры для работы с тензорами в таком формате: сложение, умножение матрицы на вектор, а также «тензорное округление» --- приближение заданного массива другим с сохранением требуемой точности. Поэтому, можно провести аналогию между ТТ-форматом и представлением вещественных чисел с плавающей точкой, только вместо цифр используются специальные параметрические представления. Арифметика ТТ-формата, представления некоторых стандартных операторов и функций, а также приближенные методы нахождения собственных значений и собственных векторов в таких форматах реализованы в рамках программного пакета ТТ-Toolbox 2.1 (<http://spring.inm.ras.ru/osel>) С его помощью уже решен ряд

Какие практические задачи можно решать с использованием ТТ-формата? Выделим три группы задач, где они уже нашли достаточно успешное применение. Первое --- это задачи вычислительной химии. Вычислительная химия представляет собой огромный пласт задач, связанный с компьютерным расчетом различных свойств молекул. Различают расчет электронной структуры, когда положение атомов зафиксировано, и задачи квантовой молекулярной динамики, когда требуется либо вычислить энергию системы при различных положениях атомов, или вычислить колебательные спектры. Расчет электронной структуры «сводится» к решению электронного уравнения Шредингера. Однако, несмотря на внешнюю простоту этого уравнения, нахождение приближенного решения этого уравнения представляет огромную сложность и по сей день, не в последнюю очередь благодаря тому, что на решение накладывается дополнительное условие антисимметрии. Возможно, ТТ-формат можно применить для ускорения вычислений электронной структуры, но пока не очень понятно, как это можно сделать, именно в силу условий антисимметрии. В задачах квантовой молекулярной динамики удалось получить очень интересные результаты. При каждом фиксированном положении атомов R_1, \dots, R_N , энергия системы определяется как решение электронного уравнения Шредингера, например с помощью методов DFT (Density Functional Theory). Такие методы реализованы в известных квантово-химических пакетах, таких как GAMESS, NWChem, и многих других. Даже вычисление функции $E=E(R_1, \dots, R_N)$ (она называется Potential Energy Surface, PES) в одной точке достаточно трудоемко, и для больших молекул (здесь N – число атомов в молекуле) может потребовать использования многоядерных архитектур и систем с распределенной памятью. Для анализа свойств молекулы требуется знание всей

структуры PES, что формально приводит к огромному числу вычислений. Для сокращения вычислений требуется некоторое малопараметрическое представление PES. В качестве такого представления естественно применять TT-формат. Нам удалось применить TT-формат для эффективного представления многомерной PES для молекул формальдегида (6 степеней свободы), метилового спирта и этилена (10 степеней свободы). Расчеты электронной структуры проводились с помощью пакета GAMESS (US). С помощью полученных PES был проведен расчет колебательных уровней энергии, и результаты при фиксированном базисе для электронных расчетов оказались лучше, чем для подходов, реализованных в GAMESS. Это можно рассматривать как первое успешное применение TT-формата и TT-Toolbox к задачам квантовой химии. Отметим, что несмотря на то, что основные вычислительные процедуры реализованы на языке MATLAB, решение даже таких многомерных задач удалось получить за небольшое время на ноутбуке. Переход к большим молекулам, конечно, потребует больших вычислительных ресурсов. Более того, на этапе расчета аппроксимации PES присутствует «встроенный параллелизм», однако такой расчет необходимо провести один раз. Это очень перспективно в плане использования кластерных систем, в том числе и эксафлопного класса. Один большой расчет сводит задачу к малопараметрической, которую можно уже потом решать многократно на персональных компьютерах или даже смартфонах(!)

К слову о смартфонах. Опишем очень интересный класс задач, где применимы тензорные методы. Это так называемые многопараметрические задачи и редукция моделей большой размерности к моделям с небольшим числом степеней свободы. Предположим, у нас есть некоторая система, которая описывается большим числом степеней свободы. Такие системы возникают при численном решении практически любой задачи математической физики, например в аэрокосмическом моделировании. Более того, такая система может зависеть от параметров (числа Маха, и пр.). Идея состоит в том, чтобы заменить такую систему меньшей, на основе малоранговой аппроксимации и TT-формата, и изучать динамику уже меньшей системы. Удивительный пример приведен в работе D. Amsallem, C. Farhat. Исходная система описывает самолет F-16 и содержит миллионы степеней свободы. Один расчет требует использования небольшого кластера. Редуцированная система имеет порядка 100 степеней свободы и расчет взлетных характеристик самолета можно проводить на смартфоне (в статье приведен пример использования Iphone). Естественно, такие подходы крайне актуальны и укладываются в базовую схему:

1. Выбор малопараметрического формата
2. Расчет тестовых характеристик «большой» системы, с использованием высокопроизводительных систем
3. На основе вычислений формулировка редуцированной модели.

Третья группа задач, где уже применялся TT-формат – задачи сжатия информации с потерями. В 2010-2011 году нам удалось показать связь TT-формата с вейвлет-разложениями, что дает возможность применять его для сжатия неструктурированных данных (изображения, массивы изображений, видео, массивы данных расчета координаты-время). Несмотря на то, что существуют различные методы, которые крайне эффективны для сжатия изображений и видео, они «настроены» на определенный тип задач. В нашем случае, одна и та же программа применялась без модификаций для различных типов данных, и были получены результаты сравнимые со стандартными подходами. Для задачи сжатия поля температуры, полученной с помощью модели океана, удалось получить сжатия в 22 раза при требуемой точности 0.1 К.

В заключение отметим, что TT-разложение и TT-формат не является «универсальным методом» решения любой многомерной задачи. Существует достаточно интересных примеров, когда прямое и наивное применение не дает небольшого числа параметров. В некоторых случаях эту проблему удается решить за счет «трюков», однако остается огромное количество проблем, как теоретических и практических. Требуется разработка новых эффективных вычислительных алгоритмов, создание связи со стандартными пакетами, внедрение процедур в образовательную и вычислительную практику. Однако с уверенностью можно сказать, что нам удалось создать принципиально новый инструмент и технологию работы с многомерными массивами, который по надежности и эффективности можно сравнить разве только лишь с сингулярным разложением для матриц (SVD), но только в многомерном случае.