

ОПТИМИЗАЦИЯ СЕТОЧНОГО РАЗБИЕНИЯ ДЛЯ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ В МОДЕЛИ SIBCIOM

Вячеслав Сергеевич Градов

Новосибирский государственный университет, 630090, Россия, г. Новосибирск, ул. Пирогова, 2, аспирант, тел. (383)330-64-60, e-mail: gradov.v.s@gmail.com

Геннадий Алексеевич Платов

Институт вычислительной математики и математической геофизики СО РАН, 630090, Россия, г. Новосибирск, пр. Академика Лаврентьева, 6, доктор физико-математических наук, главный научный сотрудник, тел. (383)330-64-50, e-mail: plat@ommfao.sccc.ru

С развитием вычислительной техники развиваются и численные модели, и это развитие может заключаться либо в создании новых моделей, либо в модификации, улучшении уже имеющихся. Модификации могут касаться как математической части, например, смена численных схем, параметризации различных физических процессов, так и технической части, например, адаптация алгоритмов для использования на других вычислительных системах или распараллеливание алгоритмов. В работе рассматривается вопрос оптимизации параллельных вычислений для модели льда и океана SibCIOM. Оптимизация заключается в построении сеточного разбиения, для которого узлы распределены равномерно по вычислительным ядрам, а время, затрачиваемое на обмены, минимально. Для реализации такого сеточного разбиения используется пакет METIS. Излагается концептуальное описание реализации обменных процессов для такого разбиения.

Ключевые слова: параллельные вычисления, разбиение графа, сетка

OPTIMIZATION OF MESH PARTITIONING FOR PARALLEL COMPUTATIONS IN THE SIBCIOM MODEL

Viacheslav S. Gradov

Novosibirsk State University, 2, Pirogova St., Novosibirsk, 630090, Russia, Ph. D. Student, phone: (383)330-64-60, e-mail: gradov.v.s@gmail.com

Gennady A. Platov

Institute of the Computational Mathematics and Mathematical Geophysics SB RAS, 6, Prospect Akademik Lavrentiev St., Novosibirsk, 630090, Russia, D. Sc., Chief Researcher, phone: (383)330-64-50, e-mail: plat@ommfao.sccc.ru

Numerical models develop with development of computational technique, and this development can include either creation of new models, or modification, improvement of existing ones. Modifications can concern both mathematical part, for example, change of numerical schemes, parameterization of various physical processes, and technical part, for example, adaptation of algorithms for use on other computer systems or parallelization of algorithms. The paper considers the issue of optimization of parallel computations for the ice and ocean model SibCIOM. Optimization consists in constructing a mesh partitioning for which the nodes are evenly distributed across the computational cores and the time spent on exchanges is minimal. The METIS package is used to implement such a mesh partitioning. A conceptual description of the implementation of exchange processes for such partitioning is presented.

Keywords: parallel computations, graph partition, mesh

Введение

Модель SibCIOM (Siberian Coupled Ice-Ocean Model) [1, 2] – численная модель океана и морского льда, разработанная в ИВМиМГ СО РАН. Она используется для расчета крупномасштабного распределения гидродинамических полей в Северном Ледовитом океане и Северной Атлантике.

По мере развития вычислительной техники, численные модели тоже развиваются, и это развитие может заключаться либо в создании новых моделей, либо в модификации, улучшении уже имеющихся. Модификации могут касаться как математической части, например, применение новых численных схем или математических методов, оптимизация алгоритмов, так и технической части, например, смена вычислительной среды или парадигмы программирования (переход к параллельным вычислениям).

Модель SibCIOM не является исключением – над ней постоянно ведется работа. В текущей параллельной версии данной модели используются протоколы MPI, которые необходимы для организации параллельных вычислений. Ядра вычислительной машины получают части расчетной области модели, проводят на ней необходимые вычисления, например, расчет какой-либо динамической величины на одном шаге по времени, и синхронизируют результаты вычислений между собой, чтобы корректно продолжить счет на следующем шаге. Правильная синхронизация между вычислительными ядрами состоит в том, что текущее ядро не может продолжить вычисления, пока не получит сигнала от связанных с ним ядер о том, что вычисления могут быть продолжены без возникновения критических ситуаций, например, взаимной блокировки или несвоевременной перезаписи данных.

Расчетная область модели SibCIOM охватывает Северный Ледовитый океан и Атлантический океан к северу от 20°S (рис. 1, 2).

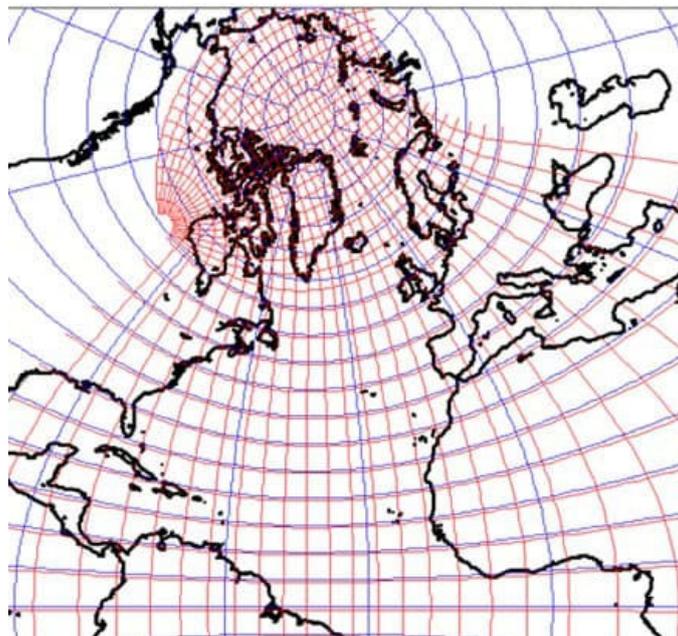


Рис. 1. Расчетная область модели SibCIOM



Рис. 2. Пример разбиения расчетной области

Текущее разбиение устроено следующим образом: прямоугольная расчетная область, разбивается сеткой на прямоугольники одинакового размера (рис. 2). Каждый прямоугольник ассоциируется с вычислительным ядром и содержит некоторые узлы: расчетные (океан) или не расчетные (суша). Прямоугольники, в которых нет вычислительных узлов, не учитываются. При таком разбиении возникает ситуация, когда в некоторых прямоугольниках находится малое количество расчетных узлов или они вовсе отсутствуют, а в некоторых гораздо большее количество, что приводит к простаиванию вычислительных ядер и, как следствие, к увеличению времени вычислений. Помимо этого может возникать проблема обмена информацией об узлах между вычислительными ядрами, например, передача слишком большого количества информации. Как видно из рис. 1, текущая концепция разбиения области позволяет достаточно просто определить массивы граничных узлов и синхронизировать их между собой. Но количество узлов в таких массивах может очень сильно различаться, что тоже может увеличить время вычисления.

Полностью избавиться от этих эффектов нельзя, но можно модифицировать разбиение области так, чтобы распределить между вычислительными ядрами примерно одинаковое количество вычислительных узлов и минимизировать обмена между вычислительными ядрами. Так целью данной работы является разработка концепции оптимального разбиения двумерной расчетной сетки и организации параллельных вычислений на такой сетке для текущей параллельной версии модели SibCIOM.

Методы и материалы

Модель SibCIOM имеет разрешение 0.5° в горизонтальных направлениях в районе средних широт и экватора. В приполярных районах разрешение составляет примерно 10-25 км. В вертикальном направлении имеется 38 уровней, расположенных неравномерно с минимальным разрешением 5 м у поверхности. Модель так же учитывает приток 52 рек, а состояние ледяного покрова рассчитывается с помощью модели льда CICE3 [3].

Для оптимального разбиения расчётной области используется пакет METIS, разработанный в Миннесотском университете [4]. Данный пакет позволяет производить разбиение графов и расчетных сеток для использования в параллельных вычислениях. Пакет позволяет задавать критерии разбиения, такие как минимизация обменов между подобластями, равномерное распределение количества узлов для каждой подобласти, минимизация или максимизация суммарного веса вершин в каждой подобласти и т.д. Разбиение сетки сводится к разбиению графа, так как узлы можно ассоциировать с вершинами графа, а геометрия сетки (прямоугольная, треугольная и т. д.) задается ребрами между вершинами графа. Для разбиения графа используется алгоритм многоуровневого k -ступенчатого разбиения графа [5]. Это базовый алгоритм данного пакета.

После разбиения настраивается синхронизация данных между вычислительными ядрами с помощью протоколов MPI. Так как геометрия границ подобластей может быть слишком сложной, требуется более тонкая настройка синхронизации, по сравнению с предыдущей концепцией обмена.

Результаты и обсуждение

На рис. 3 показан пример разбиения расчетной области, изображенной на рис. 2, на 7 частей с помощью пакета METIS.

Разбиение графа осуществляется с помощью алгоритма многоуровневого k -ступенчатого разбиения. Этот алгоритм позволяет использовать различные параметры разбиения. Один из них – целевая функция разбиения, которая принимает два значения: *cut* – минимизация разрезов по ребрам графа, и *vol* – минимизация общего числа обменов. Так же есть дополнительные параметры: *-minconn* и *-contig*. Первый параметр указывает, что процедуры разделения должны пытаться минимизировать количество связей между ядрами. Вторым параметром указывается, что процедуры разделения должны пытаться создать связные подграфы.

Для сложных сеток чаще всего невозможно достичь ситуации, когда одновременно узлы равномерно распределены между вычислительными ядрами и количество обменов минимально и одинаково для всех ядер. Но описанные выше параметры разбиения позволяют минимизировать возможные потери. Далее будет обоснован выбор параметров METIS для разбиения.

Для разного количества ядер NP рассмотрим 4 конфигурации разбиения и следующие выходные данные: NN – среднее количество узлов на вычислительное ядро, D – максимальное отклонение от среднего, $MnNC$, $MxNC$ – минимальное и максимальное количество обменов на каком-либо ядре и отношение $R = \frac{MxNC}{MnNC}$.



Рис. 3. Пример разбиения расчетной области модели SibCIOM на 7 частей с помощью пакета METIS

Подбор конфигурации осуществляется по следующему критерию: требуется минимизировать среднее отклонение D и отношение R . Чем меньше данное отношение, тем меньше максимальное время ожидания во время обменных процессов.

Из табл. 1 и 2 видно, что разбиение сетки $NP > 24$ в основном не оптимально, так как либо R достаточно велико, либо D относительно среднего числа узлов NN велика. Работа с большим числом ядер может быть оптимальнее, если сетка будет более подробной.

Таблица 1

Параметры при различных конфигурациях

NP	NN	Конфигурация METIS							
		Vol, -minconn, -contig				Cut, -minconn, -contig			
		D	MnNC	MxNC	R	D	MnNC	MxNC	R
4	76241,75	8,25	560	1568	2,80	81,75	574	1730	3,01
8	38120,88	463,13	710	1575	2,22	83,88	908	1799	1,98
10	30496,70	361,70	570	1608	2,82	9,30	782	1806	2,31
12	25413,92	560,92	234	1557	6,65	16,92	533	1567	2,94
14	21783,36	635,36	204	1362	6,68	29,36	234	1492	6,38
24	19060,44	569,56	306	1357	4,43	196,56	391	1876	4,80
32	9530,22	278,22	42	1212	28,86	96,78	42	1347	32,07
48	6353,48	185,48	114	930	8,16	3190,48	188	1141	6,07
64	4765,11	1415,11	57	893	15,67	125,11	64	978	15,28

Параметры при различных конфигурациях

NP	NN	Конфигурация METIS							
		Vol, -minconn				Cut, -minconn			
		D	MnNC	MxNC	R	D	MnNC	MxNC	R
4	76241,75	133,75	637	1424	2,24	9,75	580	1728	2,98
8	38120,88	396,88	585	1480	2,53	27,88	516	1540	2,98
10	30496,70	283,70	529	1298	2,45	21,70	620	1825	2,94
12	25413,92	234,08	358	1451	4,05	115,92	357	1796	5,03
14	21783,36	371,36	207	1569	7,58	52,36	246	1831	7,44
24	19060,44	186,44	195	1468	7,53	16,44	399	1735	4,35
32	9530,22	138,22	133	1199	9,02	263,78	70	1302	18,60
48	6353,48	183,48	55	1074	19,53	121,52	140	1096	7,83
64	4765,11	142,89	56	892	15,93	139,11	62	963	15,53

В целом масштаб величины D при $NP \leq 24$ в основном мал в сравнении с NN , поэтому более значимым параметром можно считать величину R . Так же можно заметить, что связность подграфов не играет большой роли. Хотя при отсутствии связности компоненты какого-либо подграфа могут распределиться по разным частям расчетной области, что может увеличить количество связей с другими вычислительными ядрами.

После организации разбиения необходимо настроить синхронизацию между вычислительными ядрами, учитывая сложную геометрию разбиения. В текущей реализации разбиения явно строится функция соответствия между индексами глобальной сетки и некоторой её прямоугольной подобласти. Это позволяет без хранения лишних массивов осуществлять синхронизацию. В новой реализации предлагается следующее: каждое вычислительное ядро будет хранить маску своей подобласти сетки. Маска является прямоугольным массивом, который охватывает самые крайние элементы подобласти, и поэтому имеет разный размер для каждого ядра. Далее маска расширяется расчетными узлами соседних ядер, т. е. граница окаймляется граничными узлами соседних ядер. Это окаймление может состоять из 1 и более узлов, в зависимости от реализуемой численной схемы. Далее для каждого ядра создается одномерный массив NA , размерности количества ядер NP , k -ый элемент которого ($1 \leq k \leq NP$) содержит информацию о количестве смежных с границей k -го ядра узлов. Суммированием элементов NA считается суммарное число обменов NC для рассматриваемого ядра. Следующий массив $coord$ размера $4 \times NC$ будет содержать координаты граничных узлов, обрабатываемых ядром, и граничных узлов соседних ядер, при этом ненулевые элементы NA будут служить разделителями для массива $coord$. Построенные массивы помогут установить граничную взаимосвязь между ядрами с помощью средств MPI.

Плюсом данной концепции является то, что она позволяет установить взаимосвязь между границами подобластей сетки вне зависимости от геометрии гра-

ницы. Минусом является повышенный, по сравнению с текущим подходом, объем хранимой информации, для осуществления обменов.

Так же стоит отметить, что разбиение трехмерной сетки может быть сведено к разбиению двумерной сетки, добавлением в граф исходной двумерной сетки весового коэффициента для каждого его узла, соответствующего количеству вертикальных уровней в данном узле. Предложенная выше концепция обменов будет требовать модификации, так как в данном случае появляется связь между вертикальными слоями.

В будущем предполагается построение разбиения для трехмерной сетки, реализация концепции организации обменов для двумерных и трехмерных сеток и внедрение в текущую версию модели SibCIOM.

Заключение

Для двумерной расчетной сетки модели SibCIOM осуществлено разбиение с помощью пакета METIS. Обоснован выбор параметров разбиения сетки, а также предложена концепция организации обменов при параллельных вычислениях. Стоит отметить, что данный подход может быть обобщен для разбиения трехмерной сетки.

Работа выполнена в рамках государственного задания ИВМиМГ СО РАН (проект 0215-2021-0003).

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Голубева Е. Н. Численное моделирование динамики атлантических вод в Арктическом бассейне с использованием схемы QUICKEST // Вычислительные технологии. – 2008. – Т. 13. – 5. – С. 11–24.
2. Голубева Е. Н., Платов Г. А. Численное моделирование отклика арктической системы океан – лед на вариации атмосферной циркуляции 1948-2007 гг. // Известия Российской академии наук. Физика атмосферы и океана. – 2009. – Т. 45. – 1. – С. 145–160.
3. Hunke E. C., Dukowicz J. K. An Elastic–Viscous–Plastic Model for Sea Ice Dynamics // Journal of Physical Oceanography. – 1997. – V. 27. – 9. – P. 1849–1867. doi.org/10.1175/1520-0485(1997)027<1849:AEVPMF>2.0.CO;2.
4. Karypis G., METIS. A Software Package for Partitioning Unstructured Graphs, Partitioning Meshes, and Computing Fill-Reducing Orderings of Sparse Matrices, pp. 35. [Electronic resource]. – Mode of access : <http://glaros.dtc.umn.edu/gkhome/fetch/sw/metis/manual.pdf> (дата обращения: 15.05.2021).
5. Karypis G. and Kumar V. Multilevel k-way partitioning scheme for irregular graphs. // Journal of Parallel and Distributed Computing. – 1998. – V. 48. – 1. – P. 96–129.

© В. С. Градов, Г. А. Платов, 2021