УДК 004.272.2:519.63

О.А. Дмитриева (докт. техн. наук, проф.)
ГВУЗ «Донецкий национальный технический университет», г. Донецк
кафедра прикладной математики и информатики
E-mail: dmitrieva.donntu@gmail.com

ОПТИМИЗАЦИЯ ВЫПОЛНЕНИЯ МАТРИЧНО-ВЕКТОРНЫХ ОПЕРАЦИЙ ПРИ ПАРАЛЛЕЛЬНОМ МОДЕЛИРОВАНИИ ДИНАМИЧЕСКИХ ПРОЦЕССОВ

Рассматриваются вопросы параллельной организации процессов моделирования динамических задач большой размерности, которые описываются системами обыкновенных дифференциальных уравнений. Учет разреженности матриц коэффициентов позволяет значительно сократить трудоемкость численной реализации динамических моделей, так как каждый шаг интегрирования таких задач, по обыкновению, сводится к выполнению матричных или матрично-векторных операций, от оптимальной реализации которых зависит возможность получения эффективного решения с применением параллельных ЭВМ. Особое внимание уделено способам компактного размещения элементов матриц в памяти, выполнению матричных операций, а также распределению ресурсов многопроцессорных систем при работе с разреженными матрицами.

Ключевые слова: *задача Коши, разреженные матрицы, параллельный метод, упаковочный формат, альтернативный формат, ускорение.*

Общая постановка проблемы

Рассматривая особенности моделирования линейных динамических систем (1), нельзя не признать, что в параллельных вычислительных системах с распределенной памятью наиболее трудоёмкими с вычислительной и коммуникационной точек зрения являются операции вычисления матричного и матрично-векторного произведений [1]. Поэтому ускорение параллельного выполнения этих базовых операции линейной алгебры определяет сокращение времени реализации метода решения линейной задачи в целом.

$$x' = Dx(t) + f(t),\ x(t_0) = (x_1(t_0), x_2(t_0),..., x_N(t_0))^T, \quad (1)$$

где D – матрица размерностью $N \times N$ с постоянными коэффициентами.

Необходимо отметить, что моделирование сложных динамических систем в механике, теплотехнике, электродинамике и других отраслях связано с составлением и решением систем дифференциальных и интегро-дифференциальных уравнений, порядок которых может быть очень большим в зависимости от сложности системы и требований к получаемым результатам [2]. Кроме того, отличительной чертой моделей большой размерности, как и вообще больших систем разной природы, является разреженная структура технологических матриц [3-4]. Примером могут служить решение уравнений в частных производных методами конечных разностей или конечных элементов, неявные разностные схемы, которые описывают решение краевых задач или задач Коши [2,5]. При работе с матрицами большой размерности ресурсные затраты на представление одной большой системы как совокупности подсистем, каждая из которых характеризуется активным взаимодействием составных компонентов и имеет сравнительно незначительную плотность внешних связей с другими подсистемами, могут быть неоправданными. Кроме того, несмотря на значительную разреженность, технологические матрицы не всегда структурированные, что не позволяет применять для их расчета эффективные методы анализа [6].

Постановка задач исследования

Использование известных на сегодняшний день высокоэффективных алгоритмов Штрассена [7], Копперсмита и Винограда [8], ориентированных на заполненные матрицы коэффициентов, позволяет значительно сократить трудоемкость численной реализации динамических моделей, которые описываются системами обыкновенных дифференциальных уравнений, так как каждый шаг интегрирования таких задач, по обыкновению, сводится к выполнению матричных или матрично-векторных операций [1, 9]. Но эти алгоритмы не учитывают разреженный характер матриц и особенности параллельной реализации. Существует класс специальных блочных алгоритмов умножения матрицы на матрицу или на вектор, которые являются модификацией классических параллельных алгоритмов, а именно, алгоритмов Кеннона и Фокса [10]. Но и эти алгоритмы не дают преимуществ на разреженных матрицах, в которых число элементов отличных от нуля незначительно в сравнении с количеством нулевых элементов. В то же время, разреженная матричная структура динамических моделей является важной предпосылкой для построения алгоритмов, которые наиболее полно, исходя из критериев вычислительной эффективности, используют разреженность.

Среди множества определений разреженной матрицы в данной работе принято то, которое утверждает, что приписывание матрице свойства разреженности эквивалентно утверждению о существовании алгоритма, использующего ее разреженность и делающего вычисления с ней дешевле по сравнению со стандартными алгоритмами [11]. Поэтому основное внимание в работе уделяется созданию эффективных алгоритмов, учитывающих разреженность используемых матриц, которые представляются в упакованном виде, а также распределению ресурсов многопроцессорных систем при работе с разреженными матрицами.

Генерация альтернативного формата для представления разреженных матриц

Несмотря на то, что вопросы выполнения матричных или матрично-векторных операций сравнительно редко представляют самостоятельный интерес для применений, от эффективной реализации таких задач часто зависит результативность математического моделирования, поскольку значительная часть численных методов включает в себя выполнение операций такого типа как элементарный шаг соответствующего алгоритма [11–12]. Кроме того, отличительной особенностью матричных операций при моделировании реальных объектов, является, как правило, большая размерность и разреженность. В работе для решения проблемы большой размерности и высокой степени разреженности матрицы разработан формат хранения разреженных матриц, преимущество которого состоит в возможности параллельного выполнения любых матричных операций без распаковывания, что значительно сокращает время выполнения операций и объем занимаемой памяти.

$$A = \begin{array}{c c} & \begin{array}{cccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 0 & 1 & 0 & 3 & 0 & 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

Строка 1 $\{1,2\},\{3,4\},\{5,8\}$

Строка 2 $\{2,3\}$

Строка 3 $\{3,6\},\{1,8\}$

Рисунок 1 – Представление матрицы в альтернативном формате

Альтернативой известному и наиболее часто используемому формату хранения разреженных матриц RR(C)O (row — wise representation complete and ordered) [11–12] в работе предлагается формат, схема хранения данных в котором изображена на рис. 1. В этом случае все ненулевые элементы исходной разреженной матрицы сохраняются в одном массиве. Каждый элемент массива представляет, в свою очередь, двухэлементный вектор, первый элемент которого содержит ненулевое значение исходной матрицы, второй - номер столбца исходной матрицы, в котором он находился. Номер строки каждого элемента нового массива соответствует номеру той строки, в которой он находился в исходной матрице.

Согласно алгоритму, приведенному в [11], для упаковки разреженной матрицы в формат RR(C)O нужно осуществить подсчет количества ненулевых элементов, при этом для матрицы $N \times N$ с количеством ненулевых элементов z нужно выполнить $N \times N$ сравнений, $z + N \times N$ операций сложения и непосредственную упаковку матрицы. Это действие требует $N \times N$ операций сравнения, $N + 2z$ операций переприсваивания, $2z + N \times N$ операций сложения. Для представления разреженной матрицы в альтернативном формате нужно создать маску нового массива, для чего выполнить $N \times N$ сравнений, $2z + N \times N$ сложений, $2N$ присваиваний, и осуществить упаковку матрицы, на которую потребуется $N \times N$ сравнений, $2z + N \times N$ сложений, $2N$ присваиваний. Исходя из приведенных данных видно, что второй этап для создания альтернативного упаковочного формата требует меньше операций, что приводит к снижению временных затрат на подготовительный этап. Данные выводы подкрепляются практическими результатами. Сравнительная характеристика затрат времени на упаковку разреженной матрицы разных размеров представлена на рис. 2.

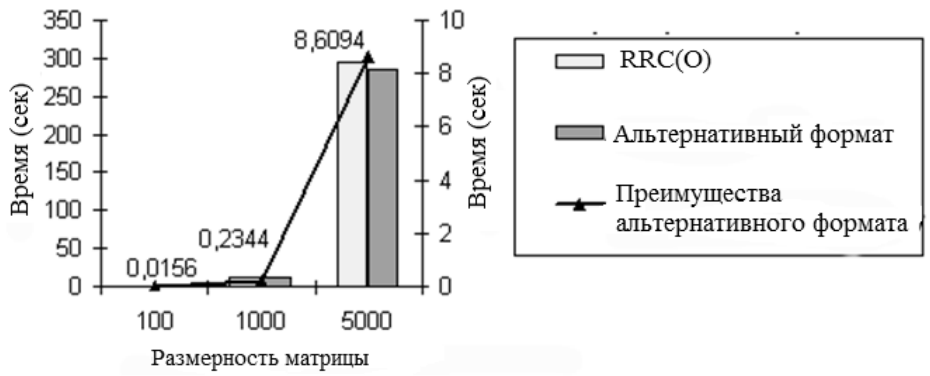

Рисунок 2 – Зависимость затрат времени на упаковку от размерности матрицы

Итак, с ростом мерности матрицы время, необходимое на ее упаковку, растет, однако уже на этом этапе использование альтернативного формата дает преимущество во времени, тем большее, чем больше размерность матрицы. Затраты времени на упаковку матрицы в специальные форматы зависят от степени разреженности матрицы, поэтому при фиксированной размерности данные затраты можно считать постоянными. В дальнейшем затраты на упаковку упоминаться не будут, поскольку целью является минимизация переменных затрат, которые будут увеличиваться с ростом размерности матрицы и интервала поиска решения. Анализ соотношения временных постоянных и переменных затрат проводился на примере реализации стадийного метода для систем дифференциальных уравнений размерностью 100, 1000 и 5000 со степенями разреженности 128, 256 и 512. Работа с матрицей в упаковочном формате осуществлялась в 3 этапа. Проводился подсчет количества ненулевых элементов исходной матрицы коэффициентов, осуществлялась упаковка матрицы и проведение вычислительных операций.

С ростом степени разреженности матрицы увеличивается часть постоянных затрат на реализацию метода. К ним отнесены затраты на подсчет количества ненулевых элементов матрицы и затраты времени на упаковку матрицы. Кроме того, поскольку с ростом степени разреженности матрицы снижается время вычислений (переменные затраты, поскольку прямо зависят от длины интервала поиска решения и шага, с которым осуществляется поиск), то часть постоянных затрат с ростом разреженности матрицы также растет.

Моделирование динамических процессов с представлением разреженных матриц в альтернативном формате

В работе для моделирования динамических процессов используется частный случай системы дифференциальных уравнений с особой правой частью (1), которая представляет собой линейную зависимость от $x(t)$.

Матрица коэффициентов D характеризуется большой размерностью и является разреженной. В таких случаях временные затраты на получение численного решения стадийными методами без предварительной подготовки матрицы можно считать неоправданно высокими [13]. Для их снижения был использованный предложенный в работе упаковочный формат хранения матрицы коэффициентов. Исследование эффективности использования данных форматов проводилось на линейной модели вида (1) со случайно сгенерированными коэффициентами матрицы D различной степени разреженности.

Поскольку целью хранения разреженной матрицы в упаковочном формате было не только сокращение объемов используемой памяти, но и повышение скорости выполнения вычислительных операций, то главным критерием эффективности использования формата считалось соотношение экономии затрат времени и памяти на реализацию явных и неявных стадийных методов, с помощью которых в работе осуществлялся поиск решения системы уравнений (1). Базовыми для сравнения эффективности упаковочных форматов хранения считались ресурсные затраты, необходимые для реализации стадийных методов без использования специальных алгоритмов.

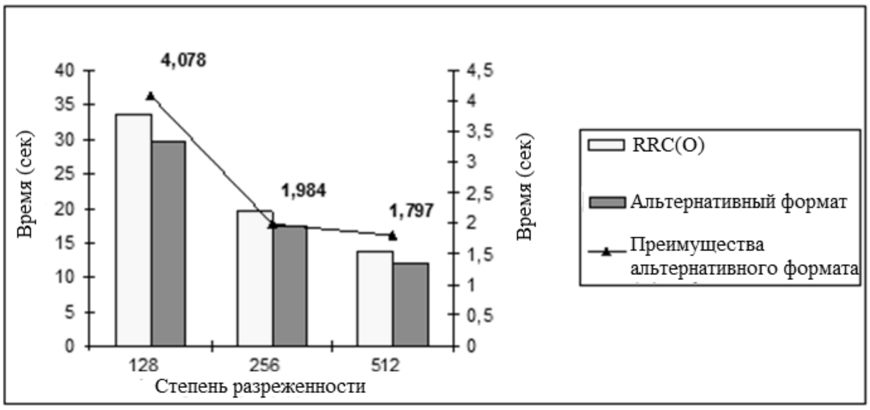

Рисунок 3 – Затраты времени на поиск решения системы мерностью 1000 на интервале 0-2

Задачей первого эксперимента было выяснение зависимости экономии времени и памяти, которая достигается благодаря использованию упаковочных форматов для системы дифференциальных уравнений размерностью 1000 на интервале от 0 до 2 с фиксированным шагом интегрирования, от разной степени разреженности матрицы коэффициентов. На рис. 3 приведено время, которое затрачено на поиск решения (1) диагонально неявным параллельным стадийным методом [13] для матриц с разреженностью 128, 256 и 512 соответственно, которые представлялись в форматах RR(C)O и предложенном в работе альтернативном. Итак, благодаря использованию разработанного формата, удалось сократить время на выполнение задачи в среднем на 10%. Что касается памяти, то для данного эксперимента необходимо было от 60000 до 24000 байт памяти в зависимости от степени разреженности матрицы. Чем она выше, тем меньше были ресурсные затраты, которые можно увидеть на рис. 4. При этом благодаря использованию альтернативного формата размер необходимой памяти удалось снизить, по меньшей мере, на 30% в сравнении с форматом RR(C)O. Для работы с системой без упаковки элементов понадобилось, в среднем, в два раза больше памяти, чем для работы с использованием специальных алгоритмов.

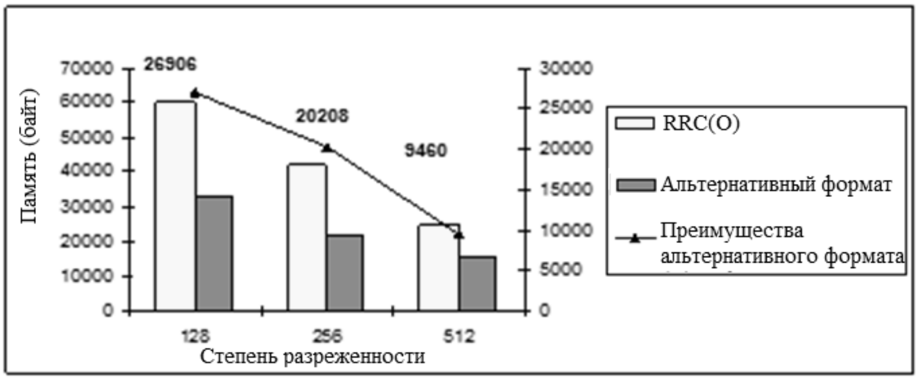

Рисунок 4 – Затраты памяти на поиск решения системы мерностью 1000 на интервале 0-2

С ростом размерности системы дифференциальных уравнений до 5000 время, израсходованное на поиск решения в упаковочном формате, увеличилось в среднем в 20 раз, но благодаря использованию альтернативного формата удалось сэкономить время получения результата приблизительно на 10% в сравнении со стандартным форматом RR(C)O. Экономия памяти составляла в среднем 35% и зависела от степени разреженности матрицы коэффициентов системы. Такие зависимости сохранялись и при увеличении интервала поиска решения.

Выводы

1. Исходя из того факта, что большие линейные системы дифференциальных уравнений, которые описывают поведение динамических объектов, как правило, имеют разреженные матрицы коэффициентов, для их описания и численной реализации в работе предложено использование альтернативного упаковочного формата. При этом выигрыш времени прямо зависит от степени разреженности матрицы исходной системы. Сокращение количества выполняемых операций происходит за счет невыполнения операций с нулевыми элементами, а сокращение объема памяти за счет хранения только отличных от нуля элементов и информации об их расположении. Преимущество данного подхода растет с увеличением интервала поиска решения и со снижением длины шага.

2. Разработанный альтернативный упаковочный формат хранения матриц, который используется в работе в качестве альтернативы стандартному упаковочному формату RR(C)O, позволил сократить время реализации моделирования динамических систем большой размерности и объем используемой памяти. Благодаря использованию альтернативного формата удалось сократить время получения результата примерно на 10% по сравнению со стандартным форматом RR(C)O. Экономия памяти достигала, в среднем, 35% и зависела от степени разреженности матрицы коэффициентов системы. Такая зависимость сохранялась и при расширении интервала поиска решения.

3. Показано, что относительная доля ресурсных затрат на операции, которые выполняются параллельно, в общем количестве израсходованных времени и памяти возрастает с ростом степени разреженности матрицы коэффициентов. Это связано с тем, что общее количество затрат уменьшается за счет сокращения ресурсных затрат на использование многократных операций, а размер затрат, необходимых на упаковку матрицы остается неизменным.

Список использованной литературы

1. Дмитрієва О.А. Організація високопродуктивних обчислень в динамічних задачах з розрідженими матрицями / О. А. Дмитриева, О.М. Григор'єва // Вісник

Східноукраїнського національного університету ім. В.Даля. –2012. – № 2(173). – С. 85–89. – ISSN 1998-7927
2. Dmitrieva O., Firsova A. Parallel Algorithms of Simulation. Increase of simulation of dynamic objects with the lumped parameters into parallel computer systems. - Lambert Academic Publishing. – 2012. – 192 p.
3. Баландин М. Ю. Методы решения СЛАУ большой размерности / М. Ю. Баландин, Э. П. Шурина. – Новосибирск: Изд-во НГТУ, 2000. – 70 с.
4. Yuster R. Fast sparse matrix multiplication [Электронный ресурс] / R. Yuster, U. Zwick.. – 2007.– Режим доступа: http://www.cs.tau.ac.il/~zwick/ papers/sparse.pdf.
5. Raz R. On the complexity of matrix product / R. Raz // SIAM Journal on Computing. – 2003. – Vol.32. – P.1356–1369.
6. Genseberger M. Improving the parallel performance of a domain decomposition preconditioning technique in the Jacobi–Davidson method for large scale eigenvalue problems / M. Genseberger // Applied Numerical Mathematics – 2010. – Vol. 60, № 11. – P. 1083–1099.
7. Strassen V. Gaussian elimination is not optimal / V. Strassen // Numerische Mathematik. – 1999. – Vol.13. – P.354–356.
8. Coppersmith D. Matrix multiplication via arithmetic progressions / D. Coppersmith, S. Winograd // Journal of Symbolic Computation. – 1990. – Vol.9. – P. 251–280.
9. Голуб Д. Матричные вычисления / Д. Голуб, Ч. Ван Лоун: пер. с англ. – М.: Мир, 1999 – 548 с. – ISBN 5-03-002406-9
10. Solving Problems on Concurrent Processors / [G. Fox, M. Johnson, G. Lyzenga, S. Otto, J. Salmon, D. Walker]. – Prentice Hall, 1990. – 350 p. – ISBN 978-0138297145.
11. Писсанецки С. Технология разреженных матриц / С. Писсанецки: пер. с англ. – М. Мир, 1988. – 410 с. – ISBN 5-03-000960-4.
12. Davis T. Direct Methods for Sparse Linear Systems / T. Davis. – Philadelphia: SIAM, 2006. – 217 p. – ISBN 0-89871-613-6.
13. Дмитрієва О.А. Паралельні різницеві методи розв'язання задачі Коші / О. А. Дмитрієва. – Донецьк: ДонНТУ, 2011. – 265 с. – ISBN 978-966-377-104-5.

References

1. Dmitrieva, O. A. and Grigoryeva, O. M., (2012), "Organisation of highly productive calculations in dynamic problems with sparse matrices", *Visnik Shidnoukrainskogo nacionalnogo universitetu im. V. Dalja*, vol. 2(173), pp. 85–89.
2. Dmitrieva and O., Firsova, A., (2012), *Parallel Algorithms of Simulation. Increase of simulation of dynamic objects with the lumped parameters into parallel computer systems*, Lambert Academic Publishing, Saarbrücken, Germany.
3. Balandin, M. Ju. and Shurina, P., (2002), *Metody reshenija SLAU bolshoj razmernosti* [Methods of the decision of SLAE of big dimension], Novosibirsk, NSTU, Rossija.
4. Yuster, R., and Zwick, U., (2007), *Fast sparse matrix multiplication*, available at: http://www.cs.tau.ac.il/~zwick/ papers/sparse.pdf.
5. Raz, R., (2003), "On the complexity of matrix product", *SIAM Journal on Computing*, vol.32, pp.1356–1369.
6. Genseberger, M., (2010), "Improving the parallel performance of a domain decomposition preconditioning technique in the Jacobi–Davidson method for large scale eigenvalue problems", *Applied Numerical Mathematics*, vol. 60, № 11, pp. 1083–1099.
7. Strassen, V., (1999), "Gaussian elimination is not optimal", *Numerische Mathematik*, vol.13, pp.354–356.
8. Coppersmith, D. and Winograd, S., (1990), "Matrix multiplication via arithmetic progressions", *Journal of Symbolic Computation*, vol.9, pp. 251–280.
9. Golub, D. and Loun, Van Ch. (1999), *Matrichnye vychislenija* [Matrix computations],

Translated by Voevodin V. V., Mir, Moscow, Russia.
10. Fox, G., Johnson, M., Lyzenga, G., Otto, S., Salmon, J. and Walker, D., (1990), *Solving Problems on Concurrent Processors*, Prentice Hall.
11. Pissanecki, S., (1988), *Tehnologija razrezhennyh matric* [Sparse Matrix Technology], Translated by Ikramov X. D. and Kaporin I. E., Mir, Moscow, USSR.
12. Davis, T., (2006), *Direct Methods for Sparse Linear Systems*, SIAM, Philadelphia.
13. Dmitrieva, O. A., (2011), *Paralel'ni riznicevi metodi rozv'jazannja zadachi Caushi* [Parallel differential methods of the solution of Cauchy problem], DonNTU, Donetsk, Ukraine.

Надійшла до редакції:
12.05.2014 р.

Рецензент:
докт. техн. наук, проф. Фельдман Л.П.

О.А. Дмитрієва
ДВНЗ «Донецький національний технічний університет»

Оптимізація виконання матрично-векторних операцій при паралельному моделюванні динамічних процесів. *Розглядаються питання паралельної організації процесів моделювання динамічних задач великої розмірності, які описуються системами звичайних диференціальних рівнянь. Урахування розрідженості матриць коефіцієнтів дозволяє значно скоротити трудомісткість чисельної реалізації динамічних моделей, тому що кожний крок інтегрування таких задач, зазвичай, зводиться до виконання матричних або матрично-векторних операцій, від оптимальної реалізації яких залежить можливість одержання ефективного розв’язку із застосуванням паралельних ЕОМ. Особлива увага приділена способам компактного розміщення елементів матриць у пам'яті, виконанню матричних операцій, а також розподілу ресурсів багатопроцесорних систем при роботі з розрідженими матрицями.*

Ключові слова: *задача Коші, розріджені матриці, паралельний метод, пакувальний формат, альтернативний формат, прискорення.*

O.A. Dmitrieva
Donetsk National Technical University

Optimization of performance of matrix and vector operations at parallel simulation of dynamic processes. *Questions of the parallel organization of processes modeling of dynamic problems big dimension which are described by systems of the ordinary differential equations, are considered. The accounting of a sparseness of matrixes of coefficients allows to reduce considerably labor input of numerical realization of dynamic models as each step of integration such tasks, as usual, is reduced to performance matrix or matrix and vector operations on which optimum realization possibility of obtaining the effective decision with use of parallel computers depends. The special attention is paid to ways of compact placement of matrixes elements in memory, to performance of matrix operations, and also distribution of resources of the multiprocessor systems during the work with the sparse matrixes.*

Key words*: Cauchy problem, sparse matrices, parallel method, packing format, alternative format, acceleration.*

Дмитриева Ольга Анатольевна, Украина, докт. техн. наук, профессор кафедры прикладной математики и информатики ГВУЗ «Донецкий национальный технический университет» (ул. Артема, 58, г. Донецк, 83001, Украина). Основное направление научной деятельности – параллельное моделирование сложных динамических объектов с сосредоточенными параметрами.