10. URL: http://icybcluster.org.ua.
11. URL: https://sparse.tamu.edu.

USE OF MIXED PRECISION IN MATHEMATICAL MODELING

The work proposes a method by which it is possible to accelerate the time of mathematical modeling of complex systems, using a mixed precision in calculations. The mixed precision allows you to improve the computing performance and save memory. Showing this approach in constructing an algorithm for solving systems of linear algebraic equations with sparse matrices.

Key words: *mathematical modeling, mixed precision, parallel algorithms, sparse matrices.*

Одержано 15.02.2019

УДК 519.6
DOI: 10.32626/2308-5878.2019-19.187-192

О. В. Чистяков, канд. фіз.-мат. наук

Інститут кібернетики імені В. М. Глушкова НАН України, м. Київ

ГІБРИДНИЙ ІТЕРАЦІЙНИЙ АЛГОРИТМ ДЛЯ РОЗВ’ЯЗУВАННЯ ЧАСТКОВОЇ ПРОБЛЕМИ ВЛАСНИХ ЗНАЧЕНЬ

Розглядається паралельний алгоритм поперемінно-трикутного методу для розв’язування на багатоядерному комп’ютері з графічними прискорювачами часткової узагальненої алгебраїчної проблеми власних значень розріджених симетричних матриць на основі їх зведення до блочно-діагонального виду з обрамленням.

Ключові слова: *гібридний комп’ютер, поперемінно-трикутний метод, розріджена матриця, алгебраїчна проблема власних значень.*

Вступ. Багато прикладних задач зводяться до розв’язування часткової узагальненої алгебраїчної проблеми власних значень (АПВЗ) для розріджених матриць великих розмірів. Наприклад, такі задачі виникають в електричних та механічних системах, в яких власні значення відповідають власним частотам коливань, а власні вектори характеризують відповідні форми (моди) коливань [1]. Визначення власних значень та векторів дають можливість аналізувати процеси та управляти ними.

З метою підвищення ефективності розв’язування задач на власні значення розріджених матриць великих розмірів у гібридному алгоритмі поперемінно-трикутного методу, який пропонується, використано ідею попереднього зведення вихідної розрідженої матриці A задачі виду

© О. В. Чистяков, 2019

$Ax = \lambda Bx$ за допомогою методу паралельних перерізів до блочно-діагональної матриці з обрамленням [1]. Таке представлення розрідженої матриці дає можливість більш ефективно виконувати розпаралелення обчислень на багатоядерному комп’ютері з графічними процесами (гібридному комп’ютері) як на CPU, так і на GPU. Слід також зазначити, що виконання ітерацій у методі, що розглядається, здійснюється послідовно, тому доцільно розглядати математичні операції, які можуть бути розпаралелені на кожній ітерації. Причому, найбільш трудомісткими математичними операціями та підзадачами (логічно завершенні частини алгоритму) на кожній ітерації щодо витрат обчислювальних ресурсів і часу виконання є множення розрідженої матриці на вектор та розв’язування систем лінійних алгебраїчних рівнянь з трикутними матрицями. Тому саме ці підзадачі підлягають розпаралелюванню на графічних процесорах, що значно прискорює процес обчислень.

Постановка задачі. Розглядаємо розв’язування задачі на власні значення

$$Ax = \lambda Bx, \tag{1}$$

де A — розріджена додатно визначена матриця порядку n, B — діагональна матриця з додатними елементами, що діють в n-вимірному евклідовому просторі H із скалярним добутком (,), λ та x — відповідно власне значення та власний вектор.

Застосуємо до задачі (1) наступну канонічну ітераційну однокрокову схему знаходження λ_1 та x_1 [2]:

$$Z(y_{k+1} - y_k) + \tau_{k+1} r_k = 0, \text{ для } k = 0, 1, 2, 3, \ldots,$$

де y_0 — довільне початкове значення, $r_k = Ay_k - \mu_k By_k$ — нев’язка; $\mu_k = (Ay_k, y_k)(\mathrm{B}\, y_k, y_k)^{-1}$ — наближення до власного значення; y_k — нормоване наближення до власного вектора; $\tau_к$ — ітераційний параметр, що обчислюється за формулою $\tau_{k+1} = (w_k, r_k) / (Aw_k, w_k)$; Z — матриця (регуляризатор), яка впливає на швидкість збіжності ітераційного процесу.

В залежності від вибраної матриці Z та наборів параметрів τ та α можуть бути отримані різноманітні схеми ітераційних методів, наприклад, для попеременно-трикутного методу:

$$w_k = B^{-1} r_k, \; r_k = Ay_k - \mu_k y_k.$$

Матриця Z має вигляд

$$Z = (E + \omega \tilde{L})(E + \omega \tilde{L}^T).$$

Матриця A представляється у вигляді $A = \tilde{L} + \tilde{L}^T$, де $\tilde{L}$ та $\tilde{L}^T$ — відповідно нижня та верхня трикутні матриці, сформовані з умови, що діагональні елементи $l_{ii} = \frac{1}{2} a_{ii}$ при розв’язуванні задачі

$$(E+\omega_{(k)}\tilde{L})(E+\omega_{(k)}\tilde{L}^{T})\,w_{(k)}=r^{(k)}\,.$$

Зведення матриці до блочно-діагонального вигляду з обрамленням. Застосуємо метод паралельних перерізів для зведення матриці A задачі (1) до блочно-діагонального вигляду з обрамленням [3]:

$$\tilde{A}=P^{T}AP=\begin{pmatrix} D_1 & 0 & 0 & \dots & 0 & C_1 \\ 0 & D_2 & 0 & \dots & 0 & C_2 \\ 0 & 0 & D_3 & & 0 & C_3 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & & D_{p-1} & C_{p-1} \\ C_1 & C_2 & C_3 & \dots & C_{p-1} & D_p \end{pmatrix}, \tag{2}$$

де P — матриця перестановок, а блоки D_i і C_i зберігають розрідженість.

Таким чином, вихідна задача (1) зводиться до $\tilde{A}y=\lambda\tilde{B}y$, де $\tilde{A}$ — блочно-діагональна матриця з обрамленням, $\tilde{B}=P^{T}BP$ — розріджена додатно визначена матриця.

Оскільки тут використано перетворення подібності, то власні значення отриманої матриці $\tilde{A}$ і вихідної A співпадають.

Передобумовлювач (матриця Z) для попеременно-трикутного методу має вигляд

$$Z=(E+\omega\tilde{L})(E+\omega\tilde{L}^{T})\ ,\ A=\tilde{L}+\tilde{L}^{T}\ . \tag{3}$$

При цьому матриця $\tilde{L}$ зберігає блочно-трикутну структуру:

$$\tilde{L}=\begin{pmatrix} \tilde{L}_1 \\ \tilde{L}_2 \\ \tilde{L}_3 \\ \vdots \\ \tilde{L}_{p-1} \\ \tilde{L}p \end{pmatrix}=\begin{pmatrix} \tilde{D}_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \tilde{D}_2 & 0 & \dots & 0 & 0 \\ 0 & 0 & \tilde{D}_3 & \dots & 0 & 0 \\ \vdots & \vdots & & \ddots & & \dots \\ 0 & 0 & 0 & & \tilde{D}_{p-1} & 0 \\ C_1 & C_2 & C_3 & \dots & C_{p-1} & \tilde{D}_p \end{pmatrix}, \tag{4}$$

де блоки $\tilde{D}_i$ — нижні трикутні матриці D_i, C_i — прямокутні матриці, $1\le i\le p$.

Розглядається гібридний комп'ютер, архітектура якого складається з багатоядерного комп'ютера MIMD-архітектури з топологією комунікацій між процесорами — «гіперкуб» та декількох графічних процесорів (GPU) SIMT-архітектури (Single Instruction, Multiple Threads). Тобто задача розв'язується на гібридному комп'ютері при використанні p процесів на CPU, позначимо $p\,\mathrm{CPU}$, та p відповідних до них графічних процесорів — $p\,\mathrm{GPU}$. Під $\mathrm{CPU}(q)$ будемо розуміти паралельний процес, що виконується на ядрі CPU з логічним номером q ($q = 1, 2..., p$) при використанні графічного процесора з таким же номером — $\mathrm{GPU}(q)$.

Для розподілу блоків (підматриць) $\tilde{L}$ між процесами CPU гібридного комп’ютера використовуємо блочну схему. З огляду на структуру $\tilde{L}$ це означає, що процеси з номерами $1 \le i < p$ зберігають блоки $\tilde{D}_i$ та C_i, а процес з номером p зберігає блок $\tilde{D}_p$.

Попеременно-трикутний метод належить до класу методів, ітераційні параметри до яких вибираються з урахуванням апріорної інформації про оператори ітераційної схеми.

Для цього методу такою інформацією є величини δ та Δ. За умов $\|\hat{x}\|_{\hat{A}}^2 \ge \delta \|\hat{x}\|^2$, $\|\hat{L}^{\mathrm{T}} \hat{x}\| \le \Delta \|\hat{x}\|_{\hat{L}}^2 / 4$ параметр $\omega = \dfrac{2}{\sqrt{\delta \Delta}}$ забезпечує збіжність ітераційного процесу зі швидкістю геометричної прогресії [4].

Реалізація гібридного алгоритму попеременно-трикутного методу. Гібридна реалізація попеременно-трикутного методу на багатоядерних комп’ютерах з графічними прискорювачами визначається блочно-трикутною структурою матриць $\tilde{L}_i$ та $\tilde{L}_i^T$ при розв’язуванні системи $Zw = r$, де матриця Z визначається за формулою (3). Отже, розв’язування задачі (3) зводиться до розв’язування на гібридній комп’ютерній архітектурі, з використанням $p\,\mathrm{CPU}$ та відповідних $p\,\mathrm{GPU}$, двох систем:

$$(E + \omega \tilde{L})\, y = r \quad \text{та} \quad (E + \omega \tilde{L}^T)\, w = y$$

за наступною обчислювальною схемою.

Гібридний алгоритм розв’язування системи з нижньою трикутною матрицею

$$(E + \omega \tilde{L})\, y = r \tag{5}$$

зводиться до виконання таких кроків:

- одночасно і незалежно процеси $\mathrm{CPU}(q)$, виконуючи на $\mathrm{GPU}(q)$ багатопоточні матрично-векторні операції, розв’язують трикутні системи $(E + \omega \tilde{D}_q) y_q = r_q$, де $1 \le q < p$, та обчислюють вектори $\tilde{y}_q$ за формулою $\tilde{y}_q = C_q y_q$;
- обчислені вектори $\tilde{y}_q$ одночасно пересилаються від процесів з номерами $1 \le q < p$ в процес $\mathrm{CPU}(p)$, в якому на $\mathrm{GPU}(p)$ розв’язується система $(E + \omega \tilde{D}_p) y_p = r_p - \sum_{q=1}^{p-1} \tilde{y}_q$.

Гібридний алгоритм розв'язування системи з верхньою трикутною матрицею

$$(E + \omega \tilde{L}^T)\, w = y \tag{6}$$

зводиться до виконання таких кроків:

- процес $\mathrm{CPU}(p)$, використовуючи $\mathrm{GPU}(p)$, розв'язує систему $(E + \omega \tilde{D}_p^T) w_p = y_p$ та розсилає отриманий вектор y_p процесам $\mathrm{CPU}(q)$, $1 \le q < p$;
- процеси $\mathrm{CPU}(q)$, використовуючи $\mathrm{GPU}(q)$, одночасно та незалежно розв'язують трикутні системи $(E + \omega \tilde{D}_q^T)\, w_q = y_q - C_q^T w_p$;
- обчислені вектори y_q пересилаються від процесів $\mathrm{CPU}(q)$ ($1 \le q < p$) в процес $\mathrm{CPU}(p)$, в якому, на $\mathrm{GPU}(p)$, розв'язується система $(E + \omega \tilde{D}_p)\, y_p = r_p - \sum_{q=1}^{p-1} \tilde{y}_q$.

Експериментальне дослідження гібридного алгоритму проводилось на інтелектуальній робочій станції гібридної архітектури Інпарком_g з такими технічними характеристиками: CPU — серії Intel(R) Xeon(R) E5606; тактова частота 2.13 GHz; швидкість 4,8 GT/s; кеш-пам'ять 8 Mb; у вузлі — 2 CPU по 4 ядра, Max Memory Size 288 Gb; графічні процесори — Nvidia Tesla M2090; пам'ять 6 Gb.

На графіках (рисунок) показано прискорення розробленого алгоритму при розв'язуванні на різній гібридній архітектурі (CPU + GPU) часткової АПВЗ для розріджених (стрічкових) матриць порядку 250 000 з різною напівшириною стрічки, які отримано шляхом дискретизації методом скінченних елементів змішаної крайової задачі для оператора Лапласа в прямокутному паралелепіпеді.

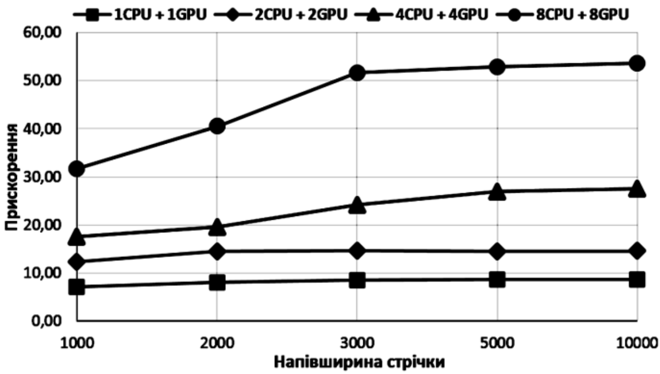

Рисунок. *Прискорення гібридного алгоритму*

З рисунку видно, що створений гібридний алгоритм добре масштабований, прискорення значно зростає при збільшенні кількості CPU та GPU в гібридній архітектурі комп'ютера у порівнянні з прискоренням на архітектурі $1\text{CPU}+1\text{GPU}$ для всіх задач.

Найбільше прискорення отримується для задач з напівшириною стрічки 3000 і більше. Це пояснюється зростанням заповненості GPU-процесорів при виконанні матрично-векторних операцій. Як відомо, GPU здатний миттєво обробляти до 1024 потоків, кожному потоку ставиться у відповідність один елемент матриці або компонента вектора.

Отже, найефективніше можна використати продуктивність графічних процесорів, якщо одна і та ж послідовність математичних операцій виконується над великим обсягом даних та кожен GPU буде максимально заповнений.

Висновки. Розглянуто гібридний алгоритм попереміннотрикутного методу на основі методу паралельних перерізів з передобумовлювачем для розв'язування часткової узагальненої АПВЗ розріджених додатно визначених матриць. Дослідження і апробація створеного гібридного алгоритму показали, що застосування передобумовлювача зводить задачу до розв'язування лінійних систем з блочно-діагональними трикутними матрицями, що визначає високу ступінь паралелізму та збалансованості паралельного процесу. Ефективність алгоритму можна значно покращити за рахунок багатопоточного виконання матрично-векторних операцій з великими обсягами даних на графічних процесорах синхронно з копіюванням даних між CPU та GPU.

Список використаних джерел:

1. Писанецки С. Технология разреженных матриц. М. : Мир, 1988, 410 с.
2. Приказчиков В. Г., Химич А. Н. Итерационные методы решения задач устойчивости и колебания пластин и оболочек. *Прикладная механика.* 1984. Т. 20, № 1. С. 88–94.
3. Джордж А., Лю Дж. Численное решение больших разреженных систем уравнений. М. : Мир, 1984. 334 с.
4. Самарский А. А., Николаев Е. С. Методы решения сеточных уравнений. М. : Наука, 1978. 592 с.

HYBRID ITERATIVE ALGORITHM FOR SOLVING PARTIAL EIGENVALUE PROBLEM

A parallel alternating-triangular method is considered for solving a partial generalized algebraic problem of eigenvalues of sparse symmetric matrices on a multi-core computer with graphic accelerators based on their reduction to block-diagonal form with a frame.

Key words: *hybrid computer, alternating-triangular method, rarefied matrix, algebraic problem of eigenvalues.*

Одержано 17.02.2019