

УДК 519.6

## ДРІБНО-ПЛИТКОВИЙ ГІБРИДНИЙ АЛГОРИТМ ФАКТОРИЗАЦІЇ РОЗРІДЖЕНОЇ МАТРИЦІ

**О. М. Хіміч**, член-кор. НАНУ, професор

Інститут кібернетики імені В.М. Глушкова НАН України  
khimich505@gmail.com

**В.А. Сидорук**, молодший науковий співробітник

Інститут кібернетики імені В.М. Глушкова НАН України  
wolodymyr.sydoruk@gmail.com

*В роботі розглядається новий гібридний плитковий алгоритм факторизації розрідженої матриці. Описується його реалізація на гібридному комп'ютері з багатоядерними процесорами та графічними прискорювачами.*

*Khimich A. N., Sydoruk V. A. Finely-tiled hybrid algorithm for the factorization of sparse matrix. The paper discusses a new hybrid tiled algorithm for the factorization of sparse matrix. Describes its implementation on a personal computer with graphics accelerators and multi-core processors.*

**Ключові слова:** РОЗРІДЖЕНА МАТРИЦЯ, ГРАФІЧНІ ПРИСКОРЮВАЧІ, ПЛИТКОВИЙ АЛГОРИТМ, CUDA, ГІБРИДНА АРХІТЕКТУРА.

**Keywords:** SPARSE MATRIX, GRAPHICS ACCELERATORS, TILE ALGORITHM, CUDA, HYBRID ARCHITECTURE.

Розглянемо задачу розв'язання системи лінійних алгебраїчних рівнянь з розрідженою додатно-визначеною симетричною матрицею блочно-діагонального виду з обрамленням

$$Ax=b. \quad (1)$$

Представимо  $A$  наступним чином

$$A = (A_1, A_2, \dots, A_i, \dots, A_p)^T,$$
$$A_i = (D_{11}, 0, \dots, 0, C_{1p}),$$

$$A_i = (0, 0, \dots, D_{ii}, \dots, 0, C_{ip}),$$

$$A_p = (C_{p1}, C_{p2}, \dots, C_{pi}, \dots, C_{p,p-1}, D_{pp}).$$

Варто відзначити, що блоки  $D_{ii}$ ,  $C_{ip}$ ,  $C_{pi}$ ,  $D_{pp}$  зберігають розріджену структуру,  $p$  - кількість діагональних блоків у матриці  $A$ . Отримати таку структуру можна застосувавши до довільної розрідженої матриці метод паралельних перерізів [1].

В роботі [2] описано паралельний гібридний алгоритм розв'язання системи (1). В даній роботі представлено алгоритм, який дозволяє максимально врахувати профільну або розріджену структуру блоків  $D_{ii}$ ,  $C_{ip}$ ,  $C_{pi}$ ,  $D_{pp}$ .

Розіб'ємо матрицю  $A$  на блоки розмірністю  $bl*bl$ . Далі для факторизації блочно-діагональної матриці застосуємо алгоритм запропонований в [3] для щільних матриць.

Для факторизації матриці на  $\kappa$ -ому кроці використаємо наступне співвідношення

$$A^\kappa = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} L_{11}^T & L_{21}^T \\ 0 & L_{22}^T \end{pmatrix} \quad (2)$$

де розмірності блоків  $A_{11}$  -  $bl*bl$ ,  $A_{12}$  -  $(n-k*bl)*bl$ ,  $A_{22}$  -  $(n-k*bl)*(n-k*bl)$ , блоки  $A_{12}$  та  $A_{22}$  враховують структуру блоків  $A_i$ ,  $A_p$ .

Звідси отримаємо алгоритм, за яким проводиться розвинення на  $\kappa$  кроці:

$$A_{11} = L_{11} * L_{11}^T; \quad (3)$$

$$L_{21} = A_{21} * (L_{11}^T)^{-1}; \quad (4)$$

$$\tilde{A}_{22} = A_{22} - L_{21} * L_{21}^T. \quad (5)$$

Зазначимо, що реалізація (3)-(5) для блоку  $A_i$ ,  $i = \overline{0, p-1}$ , враховуючи його структуру, модифікує тільки блоки  $D_{ii}$ ,  $C_{pi}$ ,  $D_{pp}$ .

Нехай на комп'ютері гібридної архітектури маємо  $p$  CPU і  $p$  GPU. Будемо використовувати наступний розподіл даних: на CPU, що відповідають процесам з номерами  $i$ ,  $i = \overline{0, p-1}$  зберігаються відповідні блоки  $D_{ii}$ ,  $C_{pi}$ . На CPU з номером  $p$  зберігається блок  $D_{pp}$ .

Паралелізація обчислень трикутної факторизації полягає в тому, що реалізація (3)-(4) з врахуванням структури  $A_i$ ,  $i = \overline{0, p-1}$  може здійснюватись незалежно у всіх процесорах: для реалізації (3)-(4) в першому GPU необхідні блоки  $D_{11}$ ,  $C_{p1}$ ; в другому GPU - блоки  $D_{22}$ ,  $C_{p2}$ , і т. д.

Гібридний алгоритм факторизації реалізується наступним чином:

1. Копіюємо блоки  $A_i$  на відповідні GPU.

Далі на кожному кроці виконуємо:

2. У всіх GPU з номерами  $i$ ,  $i = \overline{0, p-1}$  одночасно і незалежно факторизуємо  $A_{i1}$  із  $D_i$ ; формуємо  $A_{pp}^{(i)} = L_{21} * L_{21}^T$  і результат копіюємо на  $p$ -й GPU.

3. Виконуємо модифікацію  $A_{pp}$  через мультізбирання в процесі з номером  $p$ :  $\tilde{A}_{pp} = A_{pp} - \sum_{i=0}^{p-1} A_{pp}^{(i)}$ .

4. Після цього факторизуємо блок  $\tilde{A}_{pp}$  згідно (3)-(5), тим самим завершуючи процес факторизації матриці  $A$ .

Основними перевагами даного алгоритму є: можливість вибирати розмірність блоку та узгодити його з розміром кеш пам'яті при факторизації, а також можливість врахувати розріджену або профільну структуру діагональних блоків та блоків обрамлення.

### Література

1. Джордж А., Лю Дж. Численное решение больших разреженных систем уравнений. - М.: Мир, 1984. - 334 с.

2. Хіміч О. М., Сидорук В. А. Гібридний алгоритм розв'язування лінійних систем з розрідженими матрицями на основі блочного  $LL^T$  методу // Комп'ютерна математика. – 2015. – Вип. 1. – С. 67-74.

3. Alfredo Buttari, Julien Langou, Jakub Kurzak, and Jack Dongarra: A Class of Parallel Tiled Linear Algebra Algorithms for Multicore Architectures. *Parallel Computing*, Volume 35, Issue 1, P. 38-53, 2009, ISSN: 0167-8191.