

УДК 519.6

АНАЛІЗ МАСШТАБОВАНOSTI ПАРАЛЕЛЬНИХ БЛОКОВИХ АЛГОРИТМІВ РЕКУРСИВНОГО МАТРИЧНОГО ДОБУТКУ

А. О. ЧОРНА¹, І. А. НАЗАРОВА²

¹ магістрант кафедри МЗ ЕОМ, ДНУ ім. О. Гончара, Дніпропетровськ, УКРАЇНА

² канд. техн. наук, доцент кафедри ПМІ, ДонНТУ, Красноармійськ, УКРАЇНА

* email: anastasiyachorna@mail.ru

В докладі наведено матеріали досліджень авторів, що присвячені розробці та аналізу ефективності і масштабованості паралельних блокових алгоритмів матричного добутку для щільно заповнених матриць. Робота є продовженням досліджень, викладених у [1, 2].

Масштабованість є однією з основних характеристик паралельного алгоритму у сукупності з архітектурою, на якій його реалізовано. Дослідити масштабованість виключно важливо, тому що використання найпотужніших суперкомп'ютерів виправдано тільки у випадку, якщо воно ефективно. Оцінка масштабованості проводиться з використанням функції ізоефективності. Ця метрика пов'язує розмір задачі з кількістю процесорів, необхідних для збільшення прискорення пропорційно числу процесорів [3].

Рекурсивний алгоритм Штрассена призначений для швидкого множення матриць. Існує модифікація алгоритму Штрассена-Винограда, для якої потрібно 7 добутків і 15 складань. Цей алгоритм, на жаль, не є масштабованим, бо для нього вимагається число процесорів рівне ступеню 7. Для того, щоб зробити алгоритм масштабованим, використовується поліалгоритмічний підхід, тобто даний алгоритм комбінується з алгоритмом Кеннона, який буде на верхньому рівні, далі буде слідувати рекурсія алгоритму Штрассена-Винограда, і після мінімальної розмірності матриць відбуватиметься традиційне матричне множення. Знаючи розмірність перемножуваних матриць, можна обчислити оптимальну глибину рекурсії. Розроблена функція ізоефективності поліалгоритму в парі з алгоритмом Фокса має наступний вигляд:

$$m = \sqrt{\frac{t_s p (2\sqrt{p} - 3)}{t_{op} \left(\sqrt{p} - \left(5 \left(\frac{7}{4} \right)^d - 4 \right) (\sqrt{p} - 1) \right) + t_w \left(\frac{p}{2} - 3\sqrt{p} + 3 \right)},} \quad (1)$$

де p – кількість процесорів паралельної системи; m – розмірність матриці; t_s – латентність, тривалість підготовки повідомлення для передачі; t_w – час передачі одного байта; t_{op} – час виконання операції з плаваючою точкою; d – глибина рекурсії.

Глибина рекурсії d дуже сильно впливає на вигляд функції накладних витрат, а, отже, і функції ізоефективності. Побудуємо функції ізоефективності за формулою (1) для різних значень d (рис. 1, а).

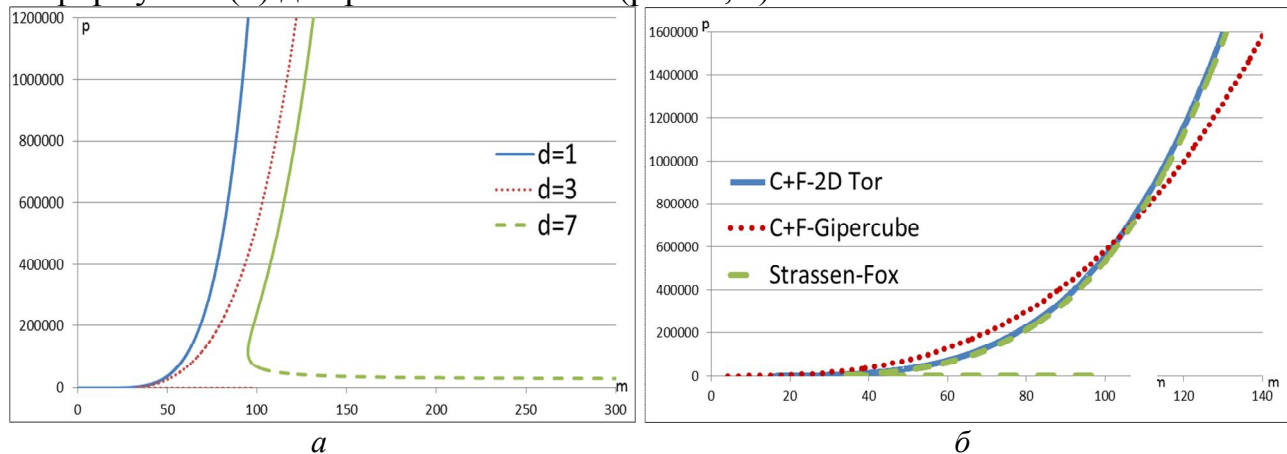


Рис. 1 – Функції ізоефективності: а – поліалгоритму; б – порівняння

Для кожного значення d при певних значеннях p за допомогою поліалгоритму Штрассена можливо обчислювати добуток матриць значної розмірності, але при подальшому збільшенні кількості процесорів можливе значення m зменшується дуже стрімко. Це підтверджує те, що глибину рекурсії треба добирати дуже ретельно. Глибина рекурсії ще й обмежує кількість процесорів: чим більше значення d , тим більшим стає мінімальне значення кількості процесорів.

На рисунку 1б представлено функції ізоефективності алгоритмів Фокса і Штрассена при значенні глибини рекурсії $d = 3$ на топології 2D-тор, Фокса і Кеннона на топологіях 2D-тор і гіперкуб. Вище графіку краще алгоритм Фокса, нижче – поліалгоритм Штрассена. Як бачимо, функція Фокса – Штрассена потребує меншу кількість процесорів для обчислення добутку матриць тієї ж розмірності, що і функції Кеннона – Фокса.

Таким чином, було проаналізовано залежність використовуваних процесорів в поліалгоритмі Кеннона-Штрассена-Винограда від глибини рекурсії, за допомогою функцій ізоефективності було визначено ступень масштабованості алгоритмі

Список літератури

1. Назарова І.А. Дослідження масштабованості паралельних систем на основі функції ізоефективності / І.А. Назарова, А.О. Чорна // Наукові праці Донецького національного технічного університету. Серія: Інформатика, кібернетика та обчислювальна техніка. – Красноармейськ, ДонНТУ, 2015. – Вип. № 2 (21). – С. 56-62.
2. Чорна А.О. Аналіз масштабованості паралельних алгоритмів матричного добутку на топології гіперкуб / А.О. Чорна, І.А. Назарова // VI Міжнародна конференція «Моделювання і комп'ютерна графіка» (м. Красноармійськ, ДонНТУ, 25-29 травня 2015 р.). – Красноармійськ: ДонНТУ, Міністерство освіти та науки України. – 2015. – С. 50-56.
3. Gupta A. Scalability of parallel algorithm for matrix multiplication / A. Gupta, V. Kumar // Technical report TR-91-54, Department of CSU of Minneapolis, 2001. – P. 1-211.