

УДК 519.63 : 519.683

Застосування паралельних обчислень та технології NVIDIA CUDA до розв'язання задач математичної фізики варіаційними методами

М. В. Сидоров, А. В. Шишка

Харківський національний університет радіоелектроніки, Україна

В статті розглядається метод скорочення часу, необхідного для розв'язання крайових задач математичної фізики варіаційними методами, на прикладі задачі розрахунку течій в'язкої рідини в однозв'язних областях шляхом застосування паралельних обчислень на кластері графічних процесорів. Також наведено результати обчислювального експерименту і показано перевагу запропонованого підходу над традиційним.

Ключові слова: паралельні обчислення, чисельний розв'язок, метод Рітца, метод R-функцій, чисельне інтегрування, центральний процесор, графічний процесор, кубічний сплайн.

В статье рассматривается метод сокращения времени, необходимого для решения краевых задач математической физики вариационными методами, на примере задачи расчета течений вязкой жидкости в односвязных областях путем применения параллельных вычислений на кластере графических процессоров. Также приведены результаты вычислительного эксперимента и показано преимущество предложенного подхода над традиционным.

Ключевые слова: параллельные вычисления, численное решение, метод Ритца, метод R-функций, численное интегрирование, центральный процессор, графический процессор, кубический сплайн.

The article proposes a method to reduce the time required for solving boundary problems of mathematical physics by variational methods on the example of a calculation of viscous flow in simply connected domains by the use of parallel computing on cluster of GPUs. Also it brings on results of computational experiment and shows an advantage of this approach over the traditional.

Key words: parallel computing, numeric solution, Ritz method, R-function method, numerical integration, CPU, GPU, cubic spline.

1. Опис проблеми, що вирішується, та її актуальність

Доволі значна частина математичних моделей, що розглядаються при вирішенні проблем, породжених промисловістю, наукою та технікою, зводяться до задач математичної фізики. В наші часи чи не єдиним способом їх розв'язання є чисельні методи та використання ЕОМ. Проте вимоги до їх реалізації постійно зростають: вона має бути швидшою, точнішою і здатною дати лад новим, більш складним задачам. Наразі основною матеріальною базою для проведення подібних обчислень є суперкомп'ютери та кластери зі значною кількістю потужних процесорів. Проте такий підхід є доволі дорогим як з точки зору вартості комплектуючих, так і з урахуванням вартості електроенергії, необхідної для функціонування подібного обчислювального комплексу.

Поява GPU (від англ. graphics processor unit – графічний процесор) загального призначення надала альтернативу для організації суперобчислень. Вони можуть бути використані не тільки для вирішення їх основної задачі (якою безперечно є розрахунок графіки), а і для довільних обчислень. Цього можна досягти викори-

станням спеціалізованих бібліотек: CUDA (від англ. compute unified device architecture – уніфікована обчислювальна архітектура) – для відеоприскорювачів виробництва NVIDIA, OpenCL або DirectCompute. Всі вони базуються на архітектурі SIMD (від англ. single instruction multiple data – одна операція над великою кількістю даних), яка дозволяє запускати велику кількість потоків одночасно і може використовуватись для численних цілей, в тому числі і для наукових розрахунків.

Дослідження присвячено вирішенню актуальної проблеми скорочення часу обчислень при реалізації чисельного розв'язання задач математичної фізики варіаційними методами. Метою роботи є дослідження можливості та доцільності застосування технології NVIDIA CUDA до паралельної імплементації варіаційного методу Рітца розв'язування крайових задач математичної фізики.

2. Витоки дослідження авторів

Ця робота спирається на метод R -функцій акад. В.Л. Рвачова [5] та його застосування до розрахунку плоских стаціонарних течій в'язкої рідини в однов'язних областях складної геометрії [1, 3, 6, 7].

Чисельні експерименти, пов'язані з описаними вище дослідженнями, показали, що досягнення прийнятної точності потребує доволі значної кількості часу, що негативно впливає на застосування розроблених методів для проведення аналізу промислових процесів та прийняття відповідних управлінських рішень в реальному часі. Рішення цієї проблеми було вирішено шукати в застосування паралельних обчислень, а також у використанні оригінальної ідеї проведення неграфічних обчислень на GPU.

3. Нерозв'язані проблеми та цілі роботи

Попри те, що застосування паралельних обчислень на графічних процесорах для проведення наукових розрахунків має низку переваг над традиційним підходом та значні перспективи, досі клас відповідно імplementованих методів залишається доволі вузьким. Наразі існує три класи алгоритмів, що отримали реалізацію на GPU [10]:

- чисельні алгоритми лінійної алгебри;
- методи Монте-Карло;
- сіткові методи.

Третій клас знайшов достойне застосування при моделюванні фізичних явищ, яке не потребує великої точності (в основному мова йде про різноманітні ефекти в 3D-графіці та 3D-іграх). Причиною цього є те, що, по-перше, даний клас задач практично співпадає зі сферою застосування відеоприскорювачів, по-друге, сіткові методи ідеально співвідносяться з архітектурою сучасних GPU.

Проте для вирішення проблем, що потребують надзвичайно точного розв'язку (взяти хоча б до уваги ті задачі математичної фізики, що вимагають не лише апроксимації розв'язку, а і його похідних), в порівнянні із сітковими методами, перевагу надають варіаційним.

Цілями даної роботи є:

- скорочення часу, необхідного для виконання чисельної реалізації варіаційного методу Рітца;

- розширення класу GPU-акселерованих алгоритмів;
- аналіз ефективності та порівняння результатів з традиційним підходом до паралельних обчислень.

4. Постановка задачі та наближений метод розрахунку плоских стаціонарних течій Стокса в однозв'язних областях

Розглянемо плоску стаціонарну течію в'язкої нестисливої рідини в скінченній однозв'язній області Ω з кусково-гладкою межею $\partial\Omega$. У наближенні Стокса ця течія може бути описана крайовою задачею для функції течії ψ [1, 6]:

$$\Delta^2 \psi = \operatorname{Re} F \quad \text{в } \Omega, \quad (4.1)$$

$$\psi|_{\partial\Omega} = \tilde{f}(s), \quad \left. \frac{\partial \psi}{\partial \mathbf{n}} \right|_{\partial\Omega} = \tilde{g}(s), \quad s \in \partial\Omega. \quad (4.2)$$

Тут $\Delta^2 = \frac{\partial^4}{\partial x^4} + 2 \frac{\partial^4}{\partial x^2 \partial y^2} + \frac{\partial^4}{\partial y^4}$ – бігармонічний оператор; Re – число Рейнольдса; $F = \frac{\partial f_y}{\partial x} - \frac{\partial f_x}{\partial y}$ – ротор вектору зовнішніх сил (f_x, f_y) , $\partial\Omega$ – межа області

Ω , \mathbf{n} – зовнішня нормаль до $\partial\Omega$, $\frac{df}{ds}$, \tilde{g} – розподіли нормальної та дотичної складових швидкостей течії відповідно, $\tilde{f} \in W_2^1(\partial\Omega)$, $\tilde{g} \in L_2(\partial\Omega)$. Якщо $v = (v_x, v_y)$ – вектор швидкості в'язкого потоку, то функція течії ψ задається співвідношеннями

$$v_x = \frac{\partial \psi}{\partial y}, \quad v_y = -\frac{\partial \psi}{\partial x}.$$

Побудована відповідно до методу R-функцій [5] загальна структура розв'язку цієї крайової задачі має вигляд [7]

$$\psi = f - \omega(D_1 f + g) + \omega^2 \Phi, \quad (4.3)$$

де f і g – відповідно продовження функцій \tilde{f} і \tilde{g} у Ω ; Φ – невизначена компонента структури; функція $\omega(x, y)$ така, що:

- 1) $\omega(x, y) = 0$ на $\partial\Omega$;
- 2) $\omega(x, y) > 0$ в Ω ;
- 3) $\frac{\partial \omega}{\partial \mathbf{n}} = -1$ на $\partial\Omega$.

У задачі (4.1), (4.2) зробимо заміну $\psi = \varphi + u$, де $\varphi = f - \omega(D_1 f + g)$. Тоді для функції u матимемо задачу з однорідними крайовими умовами

$$\Delta^2 u = -\Delta^2 \varphi + \operatorname{Re} F \quad \text{у } \Omega, \quad (4.4)$$

$$u|_{\partial\Omega} = 0, \quad \left. \frac{\partial u}{\partial \mathbf{n}} \right|_{\partial\Omega} = 0. \quad (4.5)$$

З крайовою задачею (4.4), (4.5) пов'яжемо оператор A цієї крайової задачі, який діє за правилом

$$Au \equiv \Delta^2 u$$

на області визначення

$$D(A) = \left\{ u \mid u \in C^4(\Omega) \cap C^1(\bar{\Omega}), u|_{\partial\Omega} = \frac{\partial u}{\partial \mathbf{n}}|_{\partial\Omega} = 0 \right\}.$$

Як відомо [4], такий оператор є додатно-означеним.

Замкнувши область визначення $D(A)$ у нормі, породженій скалярним добутком

$$[u, v] = \iint_{\Omega} \Delta u \Delta v dx dy,$$

отримаємо енергетичний простір H_A .

Має місце така теорема [1].

Теорема. Якщо $\varphi = f - \omega(D_1 f + g) \in W_2^2(\Omega)$, $F \in L_2(\Omega)$, то послідовність $\psi_N = \varphi + u_N$ збігається при $N \rightarrow \infty$ у нормі простору $W_2^2(\Omega)$ до єдиного узагальненого розв'язку задачі (4.1), (4.2), де $\{u_N\}$ – мінімізуюча послідовність для функціоналу

$$I[u] = \iint_{\Omega} \left[(\Delta u)^2 + 2\Delta u \Delta \varphi - 2\operatorname{Re} u F \right] dx dy. \quad (4.6)$$

Для побудови мінімізуючої послідовності функціонала (4.6) використаємо метод Рітца.

Відповідно до методу Рітца функцію u_N шукатимемо у вигляді

$$u_N = \sum_{k=1}^N c_k \varphi_k, \quad (4.7)$$

де $\{\varphi_k\}$ – координатна послідовність, тобто система функцій, яка задовольняє такі умови:

- 1) $\varphi_k \in H_A$, $k = 1, 2, \dots$;
- 2) $\{\varphi_k\}$ повна в H_A ;
- 3) для будь-якого N функції $\varphi_1, \varphi_2, \dots, \varphi_N$ лінійно незалежні.

З огляду на вигляд (4.3) структури розв'язку, координатну послідовність для нашої задачі можна обрати як $\varphi_k = \omega^2 \tau_k$, де $\{\tau_k\}$ – будь-яка повна у $W_2^2(\Omega)$ система функцій (алгебраїчних поліномів, тригонометричних поліномів, сплайнів тощо).

Для визначення невідомих сталих c_i , $i = 1, 2, \dots, N$, у (4.7) приходимо до системи лінійних алгебраїчних рівнянь (системи Рітца)

$$\sum_{j=1}^N [\varphi_i, \varphi_j] c_j = \operatorname{Re}(F, \varphi_i) - (\Delta \varphi, \varphi_i), \quad i = 1, 2, \dots, N, \quad (4.8)$$

де

$$[\varphi_i, \varphi_j] = \iint_{\Omega} \Delta \varphi_i \Delta \varphi_j dx dy, \quad i, j = 1, 2, \dots, N;$$

$$(F, \varphi_i) = \iint_{\Omega} F \varphi_i dx dy, \quad (\Delta \varphi, \varphi_i) = \iint_{\Omega} \Delta \varphi \cdot \varphi_i dx dy, \quad i = 1, 2, \dots, N.$$

Зауважимо, що при реалізації цього методу на ЕОМ значна частина машинного часу витрачається на формування системи (4.8). Причиною цього є те, що знаходження коефіцієнтів системи пов'язане з обчисленням двовимірних інтегралів (в загальному випадку багатовимірних). При цьому геометрія області може бути досить складною, а тому досягнення прийнятної точності потребує значних часових ресурсів. Вирішення даної проблеми було прийнято шукати у паралельному обчисленні коефіцієнтів системи взагалі та відповідних інтегралів зокрема.

5. Результати обчислювального експерименту

Обчислювальний експеримент був проведений для модельної задачі про розрахунок течії у прямокутній камері $\Omega = \{(x, y) | 0 < x < a, 0 < y < b\}$. Вважалось, що масові сили \mathbf{f} потенціальні, тобто $\text{rot} \mathbf{f} = 0$, а тому, $F(x, y) \equiv 0$. В якості базисних функцій було обрано кубічні сплайни Шенберга [2], які мають низку переваг над поліноміальними координатними послідовностями, зокрема кращу обчислювальну стійкість та апроксимаційні властивості. Інтеграли в системі (4.8) обчислювались наближено за формулою Гауса з 32 вузлами за кожною змінною на кожній ділянці перетину носіїв елементів координатної послідовності. Вважалось, що стінки області тверді, непроникні, нерухомі, окрім верхньої, що рухається вліво зі швидкістю 1. Краєві умови мають вигляд

$$\Psi|_{\partial\Omega} = 0, \quad \frac{\partial \Psi}{\partial \mathbf{n}}|_{\partial\Omega} = \begin{cases} -1, & y = b; \\ 0, & \partial\Omega \setminus \{y = b\}. \end{cases}$$

За цих умов

$$\Psi(x, y) = \frac{\omega(x, y) \omega_2(x, y)}{\omega_1(x, y) + \omega_2(x, y)} + \omega^2(x, y) \Phi(x, y),$$

де

$$\omega(x, y) = \left[\frac{1}{a} x(a-x) \right] \wedge_{\alpha} \left[\frac{1}{b} y(b-y) \right]; \quad \omega_1(x, y) = b - y;$$

$$\omega_2(x, y) = \left[\frac{1}{a} x(a-x) \right] \wedge_{\alpha} y,$$

\wedge_{α} – знак R -кон'юнкції [5].

Невизначену компоненту Φ апроксимуємо виразом

$$\Phi(x, y) = \sum_{i=-1}^{N_x+1} \sum_{j=-1}^{N_y+1} c_{ij} B_3 \left(\frac{N_x x}{a} - i \right) B_3 \left(\frac{N_y y}{b} - j \right) = \sum_{i=-1}^{N_x+1} \sum_{j=-1}^{N_y+1} c_{ij} \varphi_{ij}(x, y),$$

де $B_3(t)$ – кубічний сплайн Шенберга, N_x, N_y – кількість вузлів сітки сплайнів за змінними x та y відповідно.

На рис. 1 та 2 наведені графіки поверхні функції течії $\psi(x, y)$ та її ліній рівня ($a = b = 1$, $N_x = N_y = 10$).

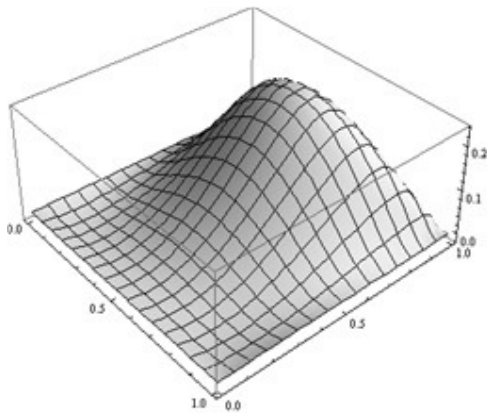


Рис. 1. Поверхня функції течії $\psi(x, y)$

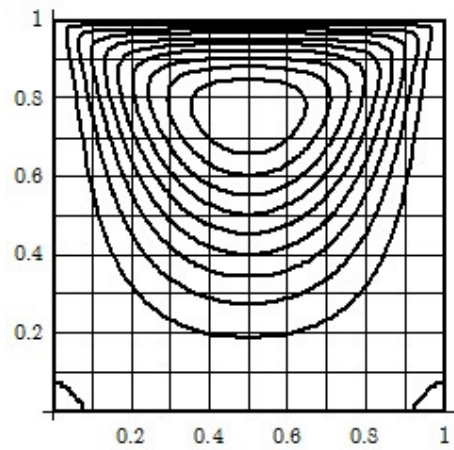


Рис. 2. Лінії рівня функції течії $\psi(x, y)$

6. Порівняння швидкої різних реалізацій методу

Для аналізу ефективності використання технології CUDA в методі Рітца було проведено ряд тестів, які порівнюють час виконання обчислень для різних реалізацій методу:

- однопоточкова програма для CPU, написана на C++ ;
- багатопоточкова програма для CPU, написана на C#;
- багатопоточкова програма для GPU, написана на CUDA C.

Конфігурація тестової системи наступна:

- Процесор: Intel® Core™ i7-860 (4 ядра/8 потоків) @ 2,93 GHz;
- ОЗП: 8 Gb DDR3-1600Hz;
- Відеокарта: NVIDIA® GeForce® GTX 460 1Gb, 336 ядер @ 1,45GHz;
- ОС: Windows 7 Ultimate 64-bit.

Обрана методика оцінювання передбачає порівняння часу, що витрачається на обчислення коефіцієнтів СЛАР, оскільки, по-перше, це займає основну частину виконання алгоритму, по-друге, решта дій, що здійснюються програмою (виділення та звільнення пам'яті, виведення результатів у файл тощо), не залежать від реалізації.

Обчислення проводились для $N_x = N_y = N$, $a = b = 1$, $h = \frac{1}{N}$.

Залежності часу [с] від вимірності апроксимаційного простору (тобто кількості координатних функцій, яка дорівнює $n = (N + 3)^2$) для різних реалізацій наведені у табл. 1.

Таблиця 1 – Залежність часу обчислень від реалізації методу та вимірності апроксимаційного простору

n	64	256	576	1024	1600
h	0,2	0,0769	0,0476	0,0345	0,0270
1 потік на CPU	2,08	31,54	154,65	491,99	1164,64
Багато потоків на CPU	0,77	11,42	54,73	182,9	421
Багато потоків на GPU	0,11	0,7	3,2	9,9	23,99

Графік залежності часу роботи різних реалізацій алгоритму наведено на рис. 3. Особливу увагу варто приділити аналізу рис. 4, оскільки він показує відношення швидкодій GPU та CPU. Можна помітити, що при достатній складності задачі графічний прискорювач працює приблизно до 17-18 разів швидше. Менший коефіцієнт при меншій кількості координатних функцій можна пояснити тим, що загальний час, витрачений GPU саме на обчислення, стає меншим у порівнянні з часом, витраченим на обмін даними з відеокартою через шину PCI Express 2.0 [8, 9].

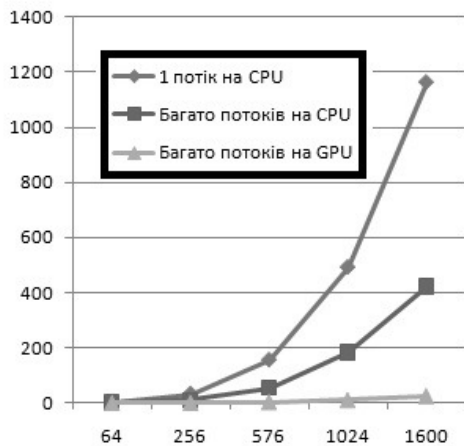


Рис. 3. Час обчислень в залежності від складності задачі

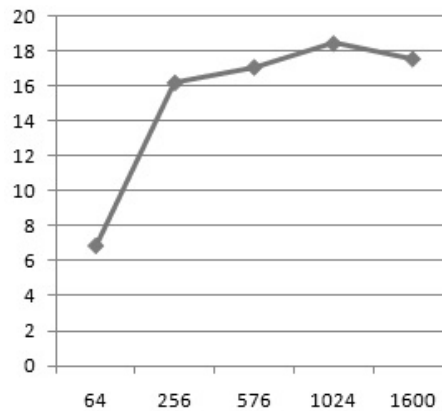


Рис. 4. Відношення швидкодій GPU та CPU в залежності від складності задачі

7. Висновки за результатами та напрямки подальших досліджень

У роботі вперше вирішено актуальну проблему скорочення часу обчислень при реалізації чисельного розв'язання задач математичної фізики варіаційними методами на прикладі методу Рітца. Координатна послідовність для реалізації методу Рітца була побудована за допомогою методу R-функцій на основі кубічних сплайнів Шенберга. Для формування системи Рітца було застосовано паралельні обчислення, що використовують технологію NVIDIA CUDA, та проведено порівняння швидкодії з традиційним підходами (однопотоковою та багатопотоковою реалізацією для CPU).

Результати наукового дослідження можуть бути використані для розробки GPU-акселерованих бібліотек існуючих методів розв'язання задач математичної

фізики, а також для розробки нових методів, орієнтованих на останні GPU архітектури.

ЛІТЕРАТУРА

1. Сидоров М.В. Математичне моделювання та чисельний аналіз течій в'язкої рідини у однозв'язних та багатозв'язних областях методом R-функцій. Автореф. дис. ... канд. фіз.-мат. наук: 01.05.02 – Математичне моделювання та обчислювальні методи. – Харків, 2006. – 18 с.
2. Вержбицкий В.М. Основы численных методов. – М.: Высш. шк., 2002. – 840 с.
3. Колосова С.В., Сидоров М.В. Применение метода R-функций к расчету плоских течений вязкой жидкости // Вісн. ХНУ. Сер. Прикл. матем. і мех. – № 602. – 2003. – С. 61 – 67.
4. Михлин С.Г. Численная реализация вариационных методов. – М.: Наука, 1966. – 432 с.
5. Рвачев В.Л. Теория R-функций и некоторые её приложения. – К.: Наук. думка, 1982. – 552 с.
6. Сидоров М.В. Применение метода R-функций к расчету течений Стокса в квадратичной каверне при малом числе Рейнольдса // Радиоэлектроника и информатика. – № 4. – 2002. – С. 77 – 78.
7. Сидоров М.В. О построении структур решений задачи Стокса // Радиоэлектроника и информатика. – № 3. – 2002. – С. 52 – 54.
8. CUDA C Best Practices Guide Version 3.2., 2010. – 65 с.
9. CUDA C Programming Guide Version 3.2., 2010. – 163 с.
10. CUDA-Accelerated Applications [Електронний ресурс]. – Режим доступу: http://www.nvidia.com/object/cuda_app_tesla.html.