Параллельное решение систем РМГ на графических процессорах

Копысов С.П., Новиков А.К., Сагдеева Ю.А.

Институт прикладной механики УрО РАН

Семинар: "Решение инженерных и научных задач на гибридных вычислительных системах, графические процессоры и архитектура CUDA"
Институт математики и механики УрО РАН, Екатеринбург,
24 марта 2011 г.

Содержание

Достоинства метода

- использование несогласованных сеток (составные конструкции,композиционные материалы, сопряженные задачи);
- использование комбинированных сеток (разные типы элементов);
- адаптивные схемы (локальное повышение порядка аппроксимации и/или престроение сетки);
- большой потенциал распараллеливания (декомпозиция области, параллельные адаптивные схемы).

Недостатки

- Увеличение размера конечно-элементной СЛАУ
- Необходимость выбора штрафного параметра
- Зависимость обусловленности СЛАУ от штрафного параметра

Предполагаемая область применения

- Моделирование напряженно-деформированного состояния в сложных (составных) конструкциях.
- Исследование напряженно-деформированного состояния конструкций в условиях длительной эксплуатации с накоплением повреждения материала.
- Получение эффективных характеристик композиционных материалов.
- Сопряженные задачи механики деформируемого твердого тела и газодинамики.

Формулировка РМГ по методу внутреннего штрафа для задачи теории упругости

Пусть $\tilde{\Omega}$ — разделение области Ω на M элементов K_i, таких что $\cup_{i=1}^{M} K_i = \tilde{\Omega}$, $K_i \cap K_j = \emptyset$. $\tilde{\Gamma}$ — совокупность внутренних ребер между элементами. Γ — внешняя граница области Ω.
Найти функцию $u^h \in U^h$ такую, что для всех $\omega^h \in W^h$

$$\int\limits_{\tilde{\Omega}} \nabla^s \omega^h : C : \nabla^s u^h d\Omega + \theta_{DG} \int\limits_{\tilde{\Gamma}} \langle \nabla^s \omega^h : C \rangle n_K \cdot [\![u^h]\!] d\Gamma -$$
$$- \int\limits_{\tilde{\Gamma}} [\![\omega^h]\!] \cdot \langle C : \nabla^s u^h \rangle n_K d\Gamma + \int\limits_{\tilde{\Gamma}} [\![\omega^h]\!] \eta [\![u^h]\!] d\Gamma = \int\limits_{\tilde{\Omega}} \omega^h \cdot g d\Omega + \int\limits_{\Gamma^N} \omega^h \cdot f^N d\Gamma, \quad (1)$$

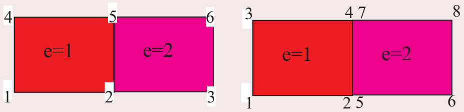
здесь u^h — пробная функция; $\eta \geqslant C_1 h^{-\beta}$ — штраф в скачке перемещений $[\![u^h]\!]$ между элементами; $\beta \geqslant 1/(d-1)$ — параметр, зависящий от размерности задачи; θ_{DG} — параметр, определяющий метод штрафа: $\theta_{DG} = -1$ симметричный РМГ с внутренним штрафом (СРМГ); $\theta_{DG} = 1$ несимметричный РМГ с внутренним штрафом (НРМГ).

Оператор среднего $\langle u^h \rangle := \frac{1}{2}\left(u_1^h + u_2^h \right)$ на $\tilde{\Gamma}$, $\langle u^h \rangle := u^h$ на Γ.

Оператор скачка $[\![u^h]\!] := u_1^h - u_2^h$ на $\tilde{\Gamma}$, $[\![u^h]\!] := u^h$ на Γ.
Граничные условия заданы на Γ^D и Γ^N, так что $\Gamma^D \cup \Gamma^N = \Gamma$, $\Gamma^D \cap \Gamma^N = \emptyset$.

Пример формирования СЛАУ симметричного РМГ

Нумерация узловых неизвестных в стандартном и разрывном методе Галёркина



Матрица жесткости в симметричном разрывном методе Галёркина (СРМГ)

$$A = \begin{bmatrix} \int_{K_1} B^T_{e=1} C B_{e=1} d\Omega \\ \int_{K_2} B^T_{e=2} C B_{e=2} d\Omega \end{bmatrix} - \int\limits_{\tilde{\Gamma}} (B_I^{<\ >})^T \phi_I^{[\![\ ]\!]} d\Gamma - \int\limits_{\tilde{\Gamma}} (\phi_I^{[\![\ ]\!]})^T B_I^{<\ >} d\Gamma + \qquad (2)$$

$$+ \int\limits_{\tilde{\Gamma}} (\phi_I^{[\![\ ]\!]})^T \eta \phi_I^{[\![\ ]\!]} d\Gamma = A_0 - A_d - A_s + A_r,$$

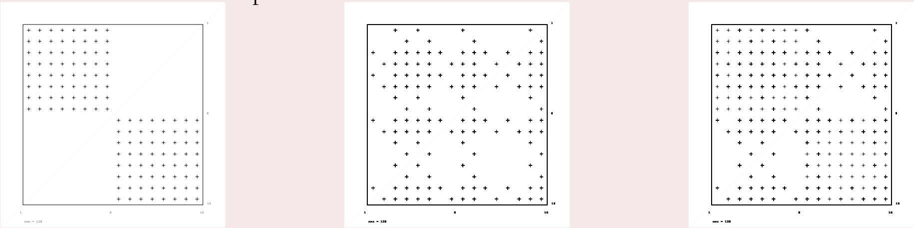


Рис. 1: Вклад элементных матриц A_0

Рис. 2: Вклад интегралов по границе $-A_d - A_s + A_r$

Рис. 3: Глобальная матрица СРМГ $A = A_0 - A_d - A_s + A_r$ $(N = 16,\ Nnz = 192)$

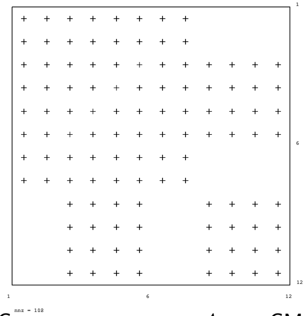


Рис. 4: Структура матрицы A для СМГ ($N = 12$, $Nnz = 112$)

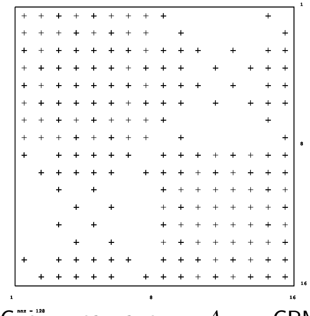


Рис. 5: Структура матрицы A для СРМГ ($N = 16$, $Nnz = 192$)

Система уравнений

$$Au = f. \tag{3}$$

Свойства матрицы системы РМГ

- $u^T A u \geq 0, \ \forall u$
- $A_{N\times N} = A^T_{N\times N}$ для СРМГ и $A_{N\times N} \neq A^T_{N\times N}$ для НРМГ.
- $Nnz \ll N^2$
- $N_{\text{РМГ}} > N_{\text{СМГ}}$
- $k(A) \sim \mathcal{O}(h^{-2})$

Тестовая задача

Область радиуса $b = 200$,
с включением радиуса $a = 5$.
Характеристики материала:
$E_1 = 1,\ E_2 = 1,\ \nu_1 = 0.3,\ \nu_2 = 0.25.$
Граничные условия:

$$u_r\big|_{r=a} = a,\ u_\theta\big|_{r=a} = 0.$$

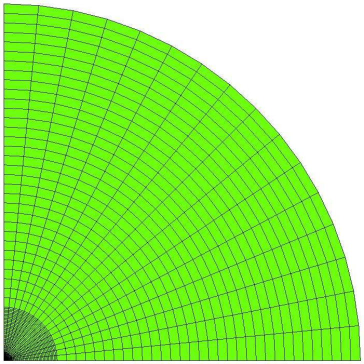
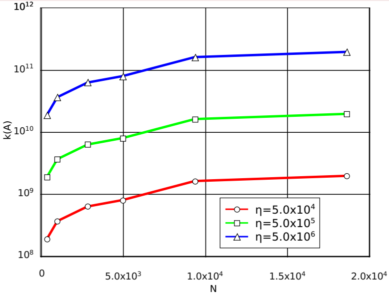


Рис. 6: Расчётная сетка
(1216 конечных элементов, 1305 узлов)

Обусловленность систем РМГ в зависимости от N и параметра η



Экономичные (гибридные) методы решения

- Прямые методы (на основе LU, LL^T, LDL^T разложений матрицы A);
- Итерационные методы крыловского типа (CG, GMRES, BiCGSTAB, ...);
- Гибридные методы:
  - Прямые методы с итерационным уточнением;
  - Итерационные методы крыловского типа с предобуславливанием на основе неполного разложения A (ILU/CG, IC(0)/CG, ...);
  - Методы декомпозиции — метод дополнения Шура (прямой метод — в подобластях, итерационный метод — на границе подобластей).

Сравнение прямого и итерационного методов решения на GPU

Таблица 1: Решение систем СМГ

Размер системы	Время, сек			$\|u - u_{LU}\|/\|u_{LU}\|$	
N	LU	CG	DIAG/CG	CG	DIAG/CG
952	0.039	0.077	0.065	2.61×10^{-8}	2.76×10^{-8}
2568	0.260	0.147	0.114	2.82×10^{-8}	3.58×10^{-8}
4960	1.063	0.236	0.188	6.08×10^{-8}	4.9×10^{-8}
7640	3.380	0.306	0.243	6.44×10^{-8}	6.1×10^{-8}
100660	—	2.384	1.807	—	—

Таблица 2: Решение систем СРМГ

Размер системы	Время, сек			$\|u - u_{LU}\|/\|u_{LU}\|$	
N	LU	CG	DIAG/CG	CG	DIAG/CG
1042	0.042	4.72	2.71	2.23×10^{-9}	1.01×10^{-8}
2426	0.217	5.83	3.29	3.46×10^{-9}	1.80×10^{-8}
4758	0.922	8.56	4.80	6.28×10^{-9}	2.54×10^{-8}
7390	2.881	12.73	7.21	1.06×10^{-8}	2.85×10^{-8}
100866	—	147.80	80.99	—	—

Для симметричных систем РМГ (СРМГ):

- метод сопряженных градиентов (CG) с предобуславливанием;
- прямые и гибридные методы на основе LL^T и LDL^T разложения A.

Для несимметричных систем (НРМГ):

- метод бисопряженных градиентов (BiCGSTAB)
- метод обобщенных минимальных невязок (GMRES)
- прямые и гибридные методы на основе LU разложения A.

Параллельный алгоритм CG с предобусловливанием

$u, r, p, q, z \in \mathbb{R}^{N_1}$, $A, M^{-1} \in \mathbb{R}^{N_1 \times N_1}$, $N_1 = N/\sqrt{P}$
$i \leftarrow 0, u_i \leftarrow 0$, $r_i \leftarrow f$
$z_i \leftarrow M^{-1} r_i$ или решение $M z_i = r_i$
$p_i \leftarrow z_i$
$\rho_i \leftarrow (r_i, z_i)$ {здесь и далее $(\cdot,\cdot) = \sum_P (\cdot,\cdot)^{(P)}$}
while $||r_i||_2/||f||_2 > \varepsilon$ **do**
  $q_i \leftarrow A p_i$ {предполагает, что $q_i \leftarrow \sum_{P_j=1}^{N_1} A^{(P_j)} p_i$}
  $\alpha_i \leftarrow (r_i, z_i)/(q_i, p_i)$
  $u_{i+1} \leftarrow u_i + \alpha_i p_i$
  $r_{i+1} \leftarrow r_i - \alpha_i q_i$
  $z_{i+1} \leftarrow M^{-1} r_{i+1}$ или решение $M z_{i+1} = r_{i+1}$
  $\rho_{i+1} \leftarrow (r_{i+1}, z_{i+1})$
  $\beta_{i+1} \leftarrow \rho_{i+1}/\rho_i$
  $p_{i+1} \leftarrow z_{i+1} + \beta_{i+1} p_i$
  $i \leftarrow i + 1$
end while

Явные предобусловливатели $M^{-1} \approx A^{-1}$

AINV — основан на неполной факторизации обратной матрицы в виде $M^{-1} = ZD^{-1}Z^T \approx A^{-1}$, где $Z \approx L^{-T}$ из $A = LDL^T$.

SPAI — получается в процессе минимизации нормы Фробениуса $min\|I - M^{-1}A\|_F < \varepsilon$, здесь ε — заданная точность.

DIAG — диагональный предобусловливатель $M^{-1} = diag(A)^{-1}$.

Неявные предобусловливатели $M \approx A$

ILU — неполная LU(p, τ)-факторизация, где p — максимальное число внедиагональных элементов в неполном LU-разложении, τ — порог отбрасывания внедиагональных элементов в предобусловливателе.

AMG — сглаживающий агрегирующий алгебраический многосеточный метод

Распараллеливание метода решения

- Параллельная реализация матрично-векторного произведения $q_i \leftarrow Ap_i$ при компактном хранении матрицы (форматы COO, CRS, ELL, BCRS2, ...).
- Использование библиотек линейной алгебры (MKL, ACML, ATLAS, GOTO, CUBLAS).
- Построение предобуславливателя.
  - Последовательное построение предобуславливателя на CPU (M_{CPU}).
  - Параллельное построение предобуславливателя на CPU (M_{MCPU}).
  - Параллельное построение предобуславливателя на GPU (M_{GPU}).
- Операции предобуславливания
  - Явное предобуславливание $z_i \leftarrow M^{-1} r_i$.
    - последовательное на CPU (M_{CPU}).
    - параллельное на CPU (M_{MCPU}).
    - параллельное предобуславливание на GPU (M_{GPU}).
  - Неявное предобуславливание $M z_i = r_i$.
    - последовательное на CPU.
    - параллельное на CPU.
    - параллельное предобуславливание на GPU.

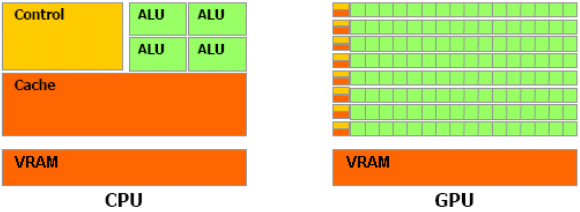


Рис. 7: Архитектуры CPU и GPU

Рис. 8: Видеокарта GeForce 470GTX

Таблица 3: Характеристики аппаратного обеспечения

Характеристика	GPU	CPU
Процессор	NVIDIA GF100 rev. A3	AMD Athlon 64 X2 5600+
Число процессоров	1	1
Число ядер	448	2
Тактовая частота,ГГц	1.22	1.00
Память индив., Мбайт	1280	1
Память ОЗУ, Мбайт	2048	2048

Решение систем стандартного метода Галёркина и симметричного РМГ

Таблица 4: Время построения предобуславливателя/решение системы и число итераций

Метод	СМГ		СРМГ		
	$N = 2416$		$N = 9346$		$N = 100866$
	CPU	GPU	CPU	GPU	GPU
CG	0.36 (169)	0.04 (171)	130.08 (35901)	10.97 (35901)	147.8 (74053)
DIAG/CG	0.12 (149)	0.03 (150)	50.29 (13416)	4.31 (13416)	80.9911 (39697)
AINV/CG	0.16/0.144 (81)		6.02/82.5 (28583)		4.54/93.33 (26225)
SPAI/CG	0.01/0.06 (99)		0.06/— ($> 10^5$)		
AMG/CG	0.02/0.09 (30)	0.02/0.03 (30)	0.05/57.72 (5529)	0.05/7.84 (5529)	0.24/80.8 (22057)
ILU(p,τ)/CG	$(20, 10^{-6})$ 0.59/0.4 (46)	$(20, 10^{-6})$ 0.56/0.56 (46)	$(80, 10^{-8})$ 15.6/3.18 (17)	$(80, 10^{-8})$ 16.45/2.18 (17)	

Форматы сжатого хранения матриц

- COO (Координатный формат) — хранятся упорядоченные по строкам значения $a_{ij} \neq 0$ и их строчные i и столбцовые индексы j;
- CRS (Сжатый строчный формат) — вместо строчных индексов хранятся указатели на начальные позиции каждой строки в списках $a_{ij} \neq 0$ и j);
- BCRS2 (Блочный вариант CRS) — используются блоки 2×2;
- ELL — формат с постоянным числом ненулевых элементов в строке;
- HYB (Гибридный) — комбинация ELL и формата COO.

Таблица 5: Время решения при различных форматах хранения матриц

Формат матрицы	CG		DIAG/CG	
	CPU	GPU	CPU	GPU
CRS	130.08	28.27	50.29	11.08
BCRS2		15.82		6.51
COO		10.97		4.31
HYB		6.99		3.13
ELL		6.41		2.93

Таблица 6: Решение несимметричной системы РМГ, $N = 9346$

Метод	CPU	GPU
ILU$(45, 10^{-8})$/ GMRES	7.62/5.48 (50)	8.24/3.95 (50)
ILU$(45, 10^{-8})$/ BICGSTAB	7.62/4.65 (22)	8.24/3.31 (22)

- Для эффективного решения систем РМГ на графических ускорителях процесс построения предобусловливателя должен быть распараллелен.
- В итерационном процессе, выполняемом на GPU предпочтительно применение явных предобусловливателей.
- Необходимым условием для получения максимального ускорения вычислений на GPU является соответствие формата хранения матриц архитектуре графического ускорителя.
- При увеличении размерности решаемых систем уравнений предполагается рассмотрение вариантов: multiGPU, multiGPU/multiCPU.