

Оценка быстродействия параллельного алгоритма пакетного решения линейных систем с трехдиагональными матрицами различной размерности на графических процессорах

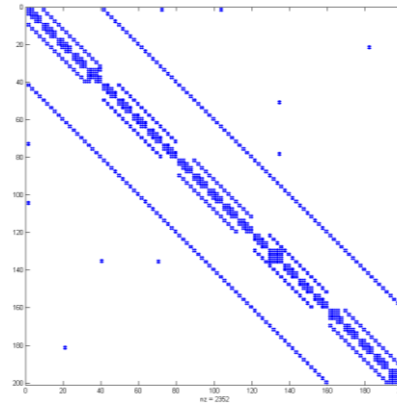
А.С. Добровольцев, М.А. Сохатский, А.В. Юлдашев

Уфимский университет науки и технологий

Решение систем линейных алгебраических уравнений (СЛАУ) с разреженной матрицей является вычислительным ядром множества задач математического моделирования, в том числе задачи моделирования фильтрационных течений углеводородов в пористой среде. Время решения СЛАУ при численном решении задачи фильтрации углеводородов может достигать от 60 до 90 % от всего времени расчета.

Особенности матриц:

- размерность $10^6 - 10^8$,
- заполненность менее 0,01 %,
- плохая обусловленность,
- несимметричность (по значениям),
- мелкозернистая блочность,
- наличие нерегулярной части,
- блочная трехдиагональная часть.



Название матрицы	Размерность	Количество ненулевых элементов	Ср. кол-во ненулевых элементов в строке
imsh	1 500 000	55 815 624	37,2
immn	2 304 102	42 859 314	18,6
krrv	4 320 921	85 471 137	19,8
mmnt	5 637 747	109 595 799	19,4
fdrv	6 610 263	118 221 633	17,9
kmms	8 630 895	167 332 329	19,4
lkms	13 665 705	254 102 823	18,6

Использование метода AIPS в качестве предобусловливателя на GPU*

Если представить матрицу в виде $A_p = P + R$, где P – трехдиагональная часть A_p , то на основе разложения в ряд Неймана можно аппроксимировать обратную к ней матрицу следующим образом:

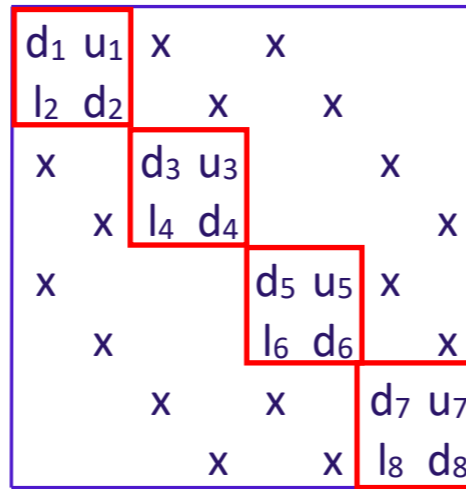
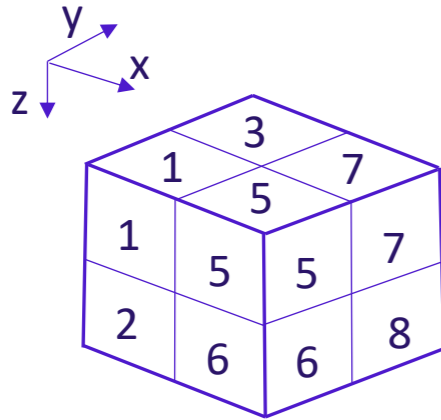
$$A_p^{-1} \approx M_N^{-1} = \left(\sum_{k=0}^N (-1)^k (P^{-1}R)^k \right) P^{-1}, \text{ при условии } \rho(P^{-1}R) < 1.$$

$$A_p \delta x_p = r_p \quad \longrightarrow \quad \delta x_p \approx M_N^{-1} r_p = P^{-1} r_p + \sum_{k=1}^N (-1)^k (P^{-1}R)^k P^{-1} r_p$$

1) Умножение разреженной матрицы R на вектор

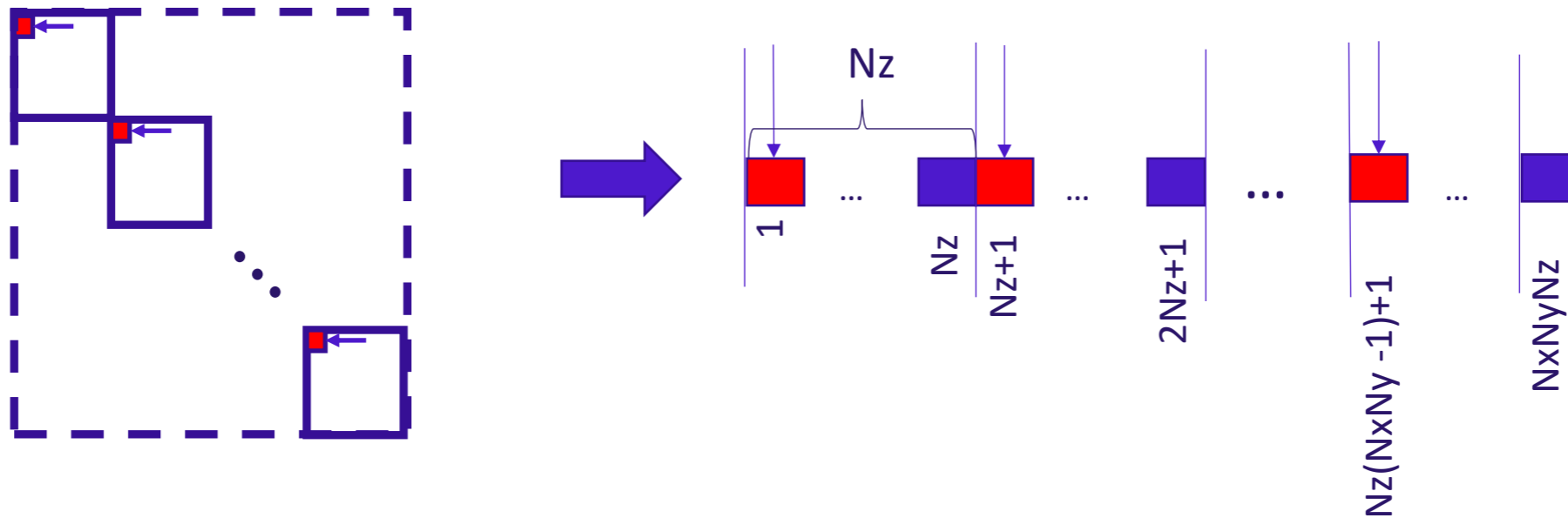
2) Решение СЛАУ с трехдиагональной матрицей P

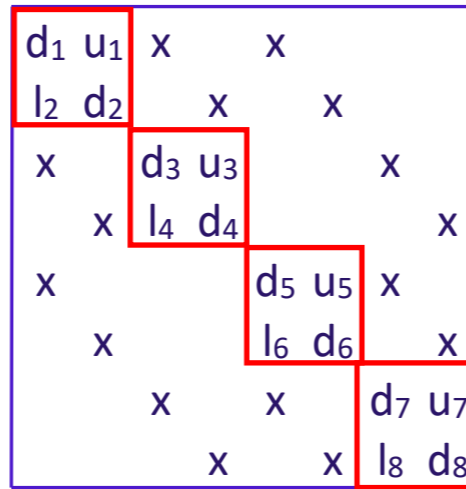
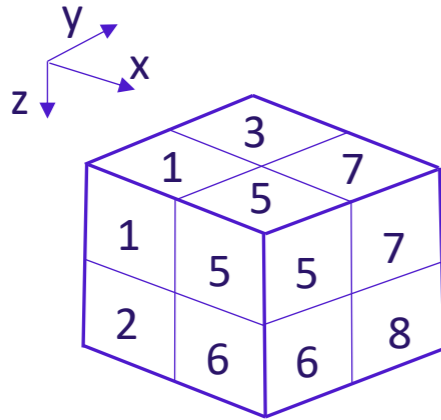
* Юлдашев А.В., Репин Н.В., Спеле В.В. Параллельный предобусловливатель на основе степенного разложения обратной матрицы для решения разреженных линейных систем на графических процессорах // Вычислительные методы и программирование. 2019. Т. 20. С. 444-456. DOI: 10.26089/NumMet.v20r439.



d – элементы главной диагонали
 u – элементы верхней диагонали
 l – элементы нижней диагонали
 x – элементы, вне трехдиагональной части

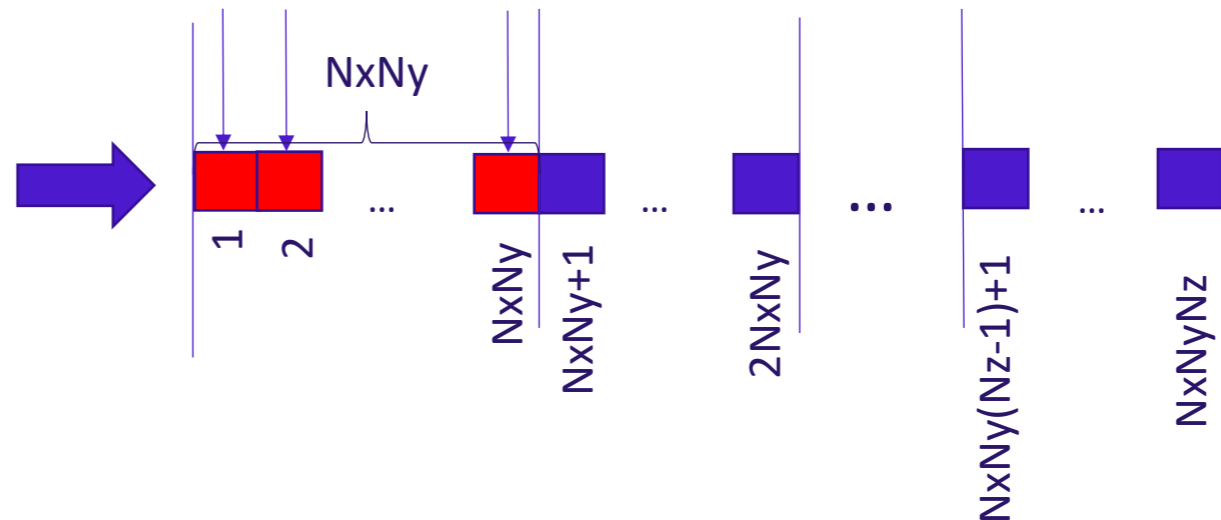
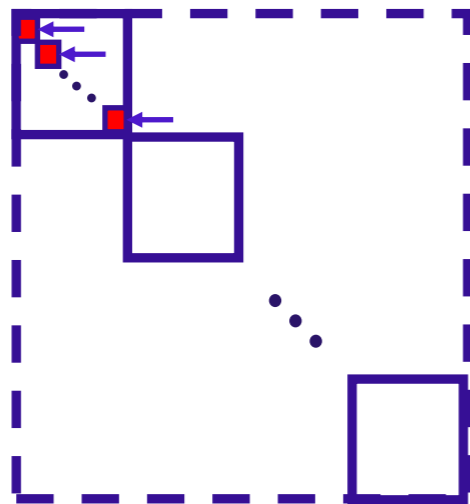
Для повышения производительности доступа к памяти элементы трехдиагональной матрицы переставляются таким образом, чтобы обращения выполнялись к последовательно расположенным адресам памяти.

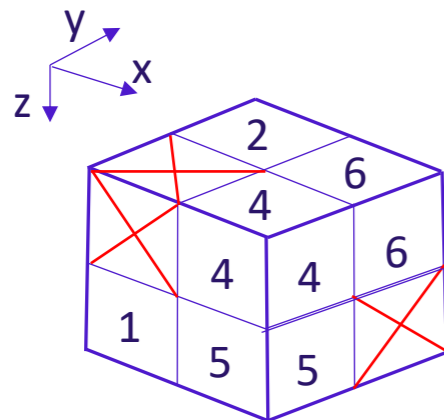




d – элементы главной диагонали
 u – элементы верхней диагонали
 l – элементы нижней диагонали
 x – элементы, вне трехдиагональной части

Для повышения производительности доступа к памяти элементы трехдиагональной матрицы переставляются таким образом, чтобы обращения выполнялись к последовательно расположенным адресам памяти.





d ₁	x	x		
x	d ₂ u ₂			x
x	l ₃ d ₃			
		d ₄ u ₄	x	
		l ₅ d ₅		
	x	x		d ₆

d – элементы главной диагонали
 u – элементы верхней диагонали
 l – элементы нижней диагонали
 x – элементы, вне трехдиагональной части

В целях балансировки нагрузки блоки переупорядочиваются согласно их размерности, к каждой группе блоков одинаковой размерности применяется предыдущая оптимизация.

l	d	u
x	d ₁	x
x	d ₂	u ₂
l ₃	d ₃	x
x	d ₄	u ₄
l ₅	d ₅	x
x	d ₆	x



l	d	u
x	d ₁	x
x	d ₆	x
x	d ₂	u ₂
l ₃	d ₃	x
x	d ₄	u ₄
l ₅	d ₅	x

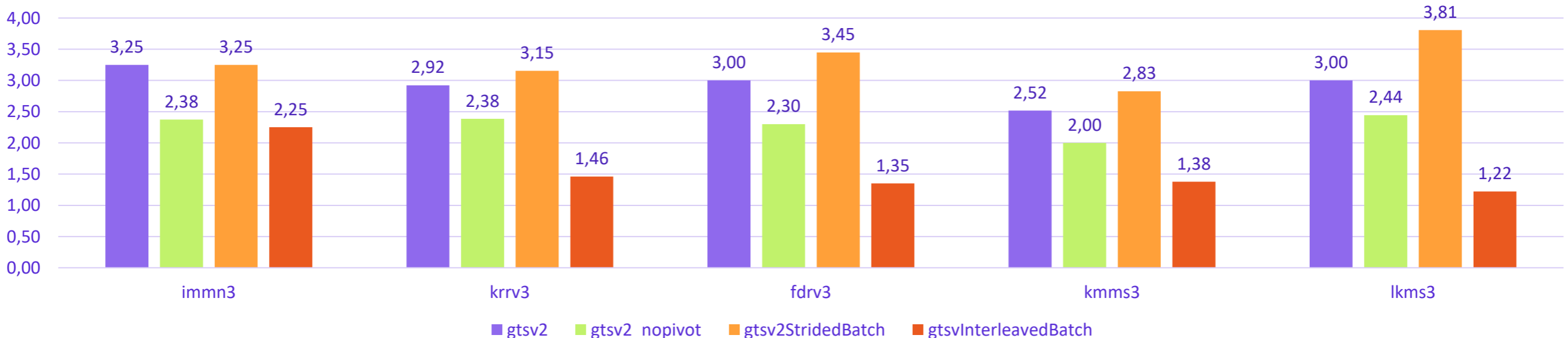


l	d	u
x	d ₁	x
x	d ₆	x
x	d ₂	u ₂
x	d ₄	u ₄
l ₃	d ₃	x
l ₅	d ₅	x

Проведено сравнение на GPU A100 с алгоритмами, реализованными в библиотеке cuSPARSE: `cusparselt>gtsv2` и `cusparselt>gtsv2_nopivot` – функции, предназначенные для решения трехдиагональных СЛАУ общего вида, а также `cusparselt>gtsv2StridedBatch` и `cusparselt>gtsvInterleavedBatch` – функции для пакетного решения множества трехдиагональных СЛАУ одинаковой размерности.

Название матрицы	Размерность матрицы	Минимальный размер блока	Максимальный размер блока	Количество независимых блоков
immn3	768 034	1	34	221 402
krrv3	1 440 307	1	39	233 866
fdrv3	2 203 421	1	55	633 202
kmms3	2 876 965	1	121	389 199
lkms3	4 555 235	1	27	1 167 013

Ускорение предложенного алгоритма относительно cuSPARSE



Среднее ускорение относительно cuSPARSE: 2,94 (gtsv2) / 2,3 (gtsv2_nopivot) / 3,3 (gtsvStridedBatch) / 1,53 (gtsvInterleavedBatch).