

УДК 519.61

doi 10.26089/NumMet.v16r453

ПАРАЛЛЕЛЬНЫЙ АЛГОРИТМ РАЗРЕЖЕННОГО QR -РАЗЛОЖЕНИЯ ДЛЯ ПРЯМОУГОЛЬНЫХ ВЕРХНИХ КВАЗИТРЕУГОЛЬНЫХ МАТРИЦ СО СТРУКТУРОЙ ТИПА ВЛОЖЕННЫХ СЕЧЕНИЙ

С. А. Харченко¹

Рассматривается параллельный алгоритм вычисления разреженного QR -разложения специальным образом упорядоченной прямоугольной матрицы на основе разреженных блочных преобразований Хаусхолдера. Для построения необходимого упорядочивания можно использовать столбцовое упорядочивание типа вложенных сечений, построенное по структуре матрицы $A^T A$, где A — исходная прямоугольная матрица. Для сеточных задач упорядочивание может быть построено на основе известного объемного разбиения расчетной сетки. В качестве базового алгоритма для организации параллельных вычислений используется QR -разложение для наборов строк матрицы с дополнением в виде нулевого начального блока.

Ключевые слова: разреженная прямоугольная матрица, верхняя квазитреугольная матрица, объемное разбиение, вложенные сечения, QR -разложение, преобразование Хаусхолдера, параллельный алгоритм.

1. Введение. QR -разложение прямоугольной матрицы является одним из базовых вычислительных алгоритмов для многих задач вычислительной математики. В частности, подобные вычисления возникают при решении систем линейных алгебраических уравнений (СЛАУ), при решении задач методом наименьших квадратов, задач на собственные значения матриц [1] и др. Возможность эффективно параллельным образом вычислять QR -разложение разреженной матрицы в некоторых случаях означает возможность использования новых классов вычислительных алгоритмов, поэтому подобные разработки представляют конкретный практический интерес.

В современной вычислительной практике существенное использование разреженности матрицы зачастую является практически единственным способом эффективного решения задач большой размерности. Относительно давно научились эффективно использовать разреженность при параллельном вычислении полных треугольных разложений типа LU -разложения [2] (разложения Холецкого в симметричном случае) и при вычислении их неполных вариантов [3]. Что касается параллельного разреженного QR -разложения, то эта тематика получила особое развитие относительно недавно, прежде всего в работах [4, 5]. В этих работах предложен мультифронтальный алгоритм построения QR -разложения, в котором параллельно вычисляемые разреженные профильные блочно-строчные QR -разложения по дереву вычислений комбинируются группами для формирования объединяющих QR -разложений. В работе [6] рассматривается многоуровневый вариант вычисления неполного разреженного QR -разложения как предобуславливание при итерационном решении задачи метода наименьших квадратов.

В настоящей статье рассматриваются идейно близкие алгоритмы параллельного вычисления полного разреженного QR -разложения матрицы. Основные отличия предлагаемого здесь подхода состоят в следующем:

- за счет введения начального нулевого блока вычисления проводятся по возможности с более разреженными Q -факторами;
- введено строчное упорядочивание и разбиение, которое позволяет дополнительно уменьшить связность вычислений и заполнение Q -факторов;
- предложен алгоритм построения представления матрицы на основе объемного разбиения расчетной сетки, удобного для параллельного вычисления разреженного QR -разложения.

Данная работа является теоретической основой для практической работы [7], содержащей описание реализации алгоритма на гибридной параллельной MPI+threads+SIMD архитектуре и начальные результаты экспериментов на такой архитектуре для тестовых задач. Кроме того, работа является базовой для

¹ ООО «ТЕСИС», ул. Юннатов, 18, офис 705, 127083, Москва; ведущий программист, e-mail: skh@tesis.com.ru

планируемой серии работ автора по новым параллельным итерационным алгоритмам решения СЛАУ и задач метода наименьших квадратов на основе композиции подпространств, порождаемых разреженным базисом.

Статья построена следующим образом. В разделе 2 описывается модификация алгоритма вычисления блочного преобразования Хаусхолдера [8, 9] и его применение для профильной разреженной QR -факторизации прямоугольной матрицы, аналогичная используемой в работах [4, 5]. В разделе 3 описывается алгоритм вычисления расширенного профильного разреженного QR -разложения с более разреженным во многих практических случаях Q -фактором для матрицы, расширенной квадратным нулевым начальным блоком. Проводится сравнение профильного и расширенного профильного разреженного QR -разложения для некоторых матриц. В разделе 4 описывается построение QR -разложения матрицы в случае строчного разбиения матрицы. В разделе 5 вводится понятие верхней квазитреугольной матрицы с разреженностью типа вложенных сечений и приводится параллельный алгоритм вычисления разреженного QR -разложения для такой матрицы. В разделе 6 описывается способ вычисления столбцевого и строчного упорядочиваний и разбиений, которые приводят матрицу, заданную на объемной сетке с известным объемным геометрическим разбиением, к виду верхней квазитреугольной матрицы с разреженностью типа вложенных сечений.

2. Блочное преобразование Хаусхолдера и профильное разреженное QR -разложение.

2.1. Блочное преобразование Хаусхолдера. Классическое преобразование Хаусхолдера — это унитарная матрица вида [1]

$$G = I_M + \tau f f^T, \tag{1}$$

где $f \in \mathbb{R}^M$ — некоторый вектор, $1 \leq \|f\| \leq 2$, τ — некоторое вещественное число, $-2 \leq \tau \leq 0$. Преобразование Хаусхолдера строится, как правило, из условия, чтобы для некоторого вектора $c \in \mathbb{R}^M$ в векторе $c' = G^T c$ все компоненты, кроме заданного набора начальных компонент, были нулевыми.

Рассмотрим прямоугольную матрицу $C \in \mathbb{R}^{M \times s}$. Классический алгоритм [1] вычисления QR -разложения этой матрицы на основе преобразований Хаусхолдера (1) приводит к соотношениям

$$C = \left(\prod_{j=1}^s G_j \right) \begin{bmatrix} R \\ 0 \end{bmatrix}, \tag{2}$$

где $R \in \mathbb{R}^{s \times s}$ — квадратная верхняя треугольная матрица; $G_j \in \mathbb{R}^{M \times M}$ — ортогональные матрицы преобразования Хаусхолдера, имеющие вид $G_j = I_M + \tau_j f_j f_j^T$; τ_j — некоторое число; $f_j \in \mathbb{R}^M$ — некоторый вектор, первые $(j - 1)$ компонент которого нулевые. Расширение QR -разложения для расширенной на один столбец матрицы $[C, c] \in \mathbb{R}^{M \times (s+1)}$, как это хорошо известно, можно вычислить применением в соответствующем порядке транспонированных преобразований Хаусхолдера к столбцу c и построением нового преобразования Хаусхолдера. Для явного нахождения i -го базисного вектора Q -фактора QR -разложения в неявной форме (2) достаточно вычислить результат умножения

$$q_i = \left(\prod_{j=1}^s G_j \right) e_i, \tag{3}$$

где e_i — i -й орт.

Обозначим $\Omega_s = \prod_{j=1}^s G_j$. Рассмотрим модификацию хорошо известного [7] блочного обобщения преобразований Хаусхолдера (1) в виде

$$\Omega_s = I_M + F_s T_s F_s^T, \tag{4}$$

где $\Omega_s \in \mathbb{R}^{M \times M}$, $\Omega_s^T \Omega_s = I_M$, $F_s \in \mathbb{R}^{M \times s}$ и $T_s \in \mathbb{R}^{s \times s}$. Решать задачу построения блочного преобразования Хаусхолдера (4) будем по индукции.

Для $s = 1$ положим $F_1 = f_1$ и $T_1 = \tau_1$. Пусть для некоторого $s \geq 1$ блочное преобразование Хаусхолдера (4) уже построено. Построим это преобразование для $s' = s + 1$. По определению и по предположению индукции имеют место равенства $\Omega_s = I_M + F_s T_s F_s^T$ и $\Omega_{s+1} = \Omega_s G_{s+1}$.

Положим $F_{s+1} = [F_s, f_{s+1}]$, $T_{s+1} = \begin{bmatrix} T_s & T_s F_s^T f_{s+1} \tau_{s+1} \\ 0 & \tau_{s+1} \end{bmatrix}$, тогда, как легко проверить, имеет место

равенство (4) в том числе и для $s' = s + 1$, при этом $\Omega_{s+1} = \prod_{j=1}^{s+1} G_j$ и $\Omega_{s+1}^T \Omega_{s+1} = I_M$.

Таким образом, получается следующая последовательность вычислений блочного преобразования Хаусхолдера. Сначала вычисляется обычное векторное QR -разложение матрицы с одновекторными преобразованиями Хаусхолдера. Из полученных наборов векторов составляется матрица F_s . Далее вычисляется матрица $\Psi_s = F_s^T F_s$ и по ней по рекуррентным формулам вычисляется матрица T_s . Подобная организация вычислений позволяет эффективно использовать векторные вычисления на основе SIMD-инструкций [6].

В последующем изложении мы будем рассматривать вопросы вычисления QR -разложения для прямоугольных так называемых мелкоблочных разреженных матриц. Это означает, что разреженность понимается в смысле блоков малого размера, каждый из которых является плотной в общем случае прямоугольной матрицей. Для простоты последующего изложения мы будем предполагать, что строчные и столбцовые размеры всех мелких блоков совпадают между собой и равны некоторому малому значению s , $s \geq 1$. При этом все представленные в работе алгоритмы могут быть легко обобщены на случай переменного столбцового и строчного мелкоблочного разбиения. В противовес мелким плотным $s \times s$ блокам мы будем говорить также о крупноблочном, или просто блочном разбиении, строчном и столбцовом. Это всегда будет означать, что соответствующие подматрицы составлены из некоторого количества мелкоблочных строк и столбцов. При этом под блочным преобразованием Хаусхолдера мы всегда будем иметь в виду преобразование (4) для одного мелкоблочного столбца.

2.2. Профильное разреженное QR -разложение. Блочные преобразования Хаусхолдера (4) можно применить для вычисления разреженного QR -разложения разреженной матрицы $C \in \mathbb{R}^{M \times Ns}$, составленной из N мелкоблочных столбцов. Для этого можно поступить, например, следующим образом. Для минимизации заполнения R -фактора QR -разложения вычисляем структуру разреженности матрицы $C^T C$ и вычисляем упорядочивание, минимизирующее заполнение матрицы с такой структурой при ее треугольной факторизации. Можно выбрать, например, упорядочивание вложенных сечений ND (Nested Dissection) или упорядочивание RCM (Reverse Cuthill–McKee) [2], что определяет упорядочивание мелкоблочных столбцов матрицы C . Далее упорядочивание мелкоблочных строк можно выбрать последовательно, нумеруя сначала элементы первого столбца матрицы C , потом второго и т.д. Переупорядоченная таким образом матрица будет иметь вид, приведенный на рис. 1 слева. В качестве базового столбцового упорядочивания для этой матрицы было выбрано упорядочивание ND.

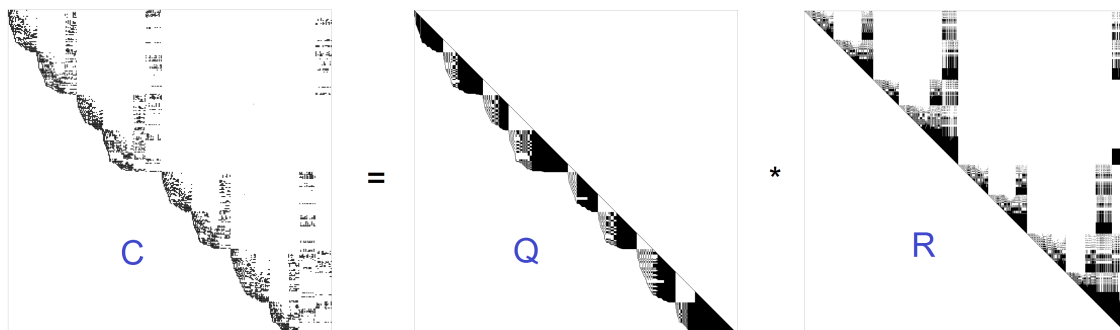


Рис. 1. Структуры разреженности матрицы C и ее Q - и R -факторов для тестовой задачи

При построении так называемого профильного разреженного QR -разложения таким образом упорядоченной мелкоблочной матрицы C будем действовать по аналогии со случаем плотной матрицы. По первому мелкоблочному столбцу матрицы строим разреженное блочное преобразование Хаусхолдера с разреженностью столбца такой, чтобы обнулить мелкие блоки матрицы под первой блочной диагональю. Применяем транспонированное блочное преобразование Хаусхолдера ко второму мелкоблочному столбцу, для полученного результата строим следующее разреженное блочное преобразование Хаусхолдера для обнуления элементов под второй мелкоблочной диагональю и т.д. На рис. 1 справа показана мелкоблочная разреженность Q - и R -факторов QR -разложения для матрицы C слева. Для этого теста заметно довольно сильное заполнение в Q -факторе внутри соответствующего профиля матрицы.

3. Расширенное профильное разреженное QR -разложение. Профильное разреженное QR -разложение матрицы, описанное выше, не является единственным вариантом построения QR -разложения. Существует вариант, для которого более естественным образом выделяется его параллельная структура и который во многих практически важных случаях приводит к меньшему заполнению в факторе Q . Это так называемое расширенное профильное QR -разложение.

Схематически профильное и расширенное профильное QR -разложение изображены на рис. 2. Факти-

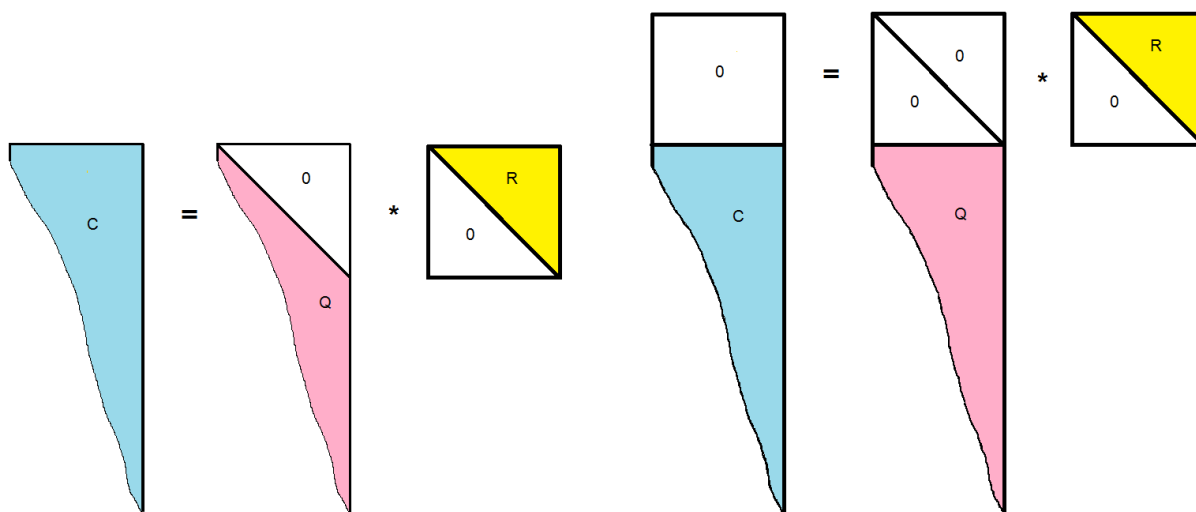


Рис. 2. Профильное (слева) и расширенное профильное (справа) QR -разложения

чески расширенное профильное QR -разложение — это профильное QR -разложение, примененное к матрице, расширенной сверху нулевым квадратным блоком. При этом структура разреженности Q -фактора пополняется возможными дополнительными элементами на месте бывшего фактора R и отсоединенными диагональными элементами, примыкающими к новому R -фактору.

Покажем прежде всего, что в случае матрицы C полного столбцевого ранга профильное и расширенное профильное QR -разложения математически эквивалентны. Обозначим $\tilde{C} = \begin{bmatrix} 0 \\ C \end{bmatrix}$. Для матрицы

полного столбцевого ранга C матрица R в R -факторе QR -разложения не вырождена и является очевидно треугольным фактором матрицы $C^T C$, а значит, матрица R является R -фактором и для матрицы \tilde{C} , поскольку $\tilde{C}^T \tilde{C} = C^T C$. Далее, весь набор $Q = [q_1, \dots, q_{N_s}]$ базисных векторов Q -фактора QR -разложения можно, в соответствии с мелкоблочным аналогом формулы (3) в терминах блочных преобразований Хаусхолдера, вычислить по формуле $Q = \left(\prod_{j=1}^N \Omega_j \right) \begin{bmatrix} I_{N_s} \\ 0 \end{bmatrix}$. Очевидно, имеет место равенство $Q = CR^{-1}$, а

также аналогично $\tilde{Q} = \tilde{C}R^{-1}$, но тогда $Q = CR^{-1} = [0 \ I_M] \tilde{C}R^{-1} = [0 \ I_M] \tilde{Q}$. Таким образом, базисные векторы профильного Q -фактора QR -разложения можно вычислять, построив базисные векторы расширенного Q -фактора QR -разложения и отбросив их первые компоненты. Заметим, что первые компоненты расширенного базисного вектора нулевые, поскольку в противном случае после умножения на невырожденную матрицу R результирующий вектор будет ненулевым, что противоречит тождеству $\tilde{C} = \tilde{Q}R$. В случае неполного столбцевого ранга эквивалентности нет; это связано именно с тем, что в этом случае некоторые начальные компоненты базисных векторов расширенного QR -разложения будут ненулевыми.

Исходя из описания расширенного профильного QR -разложения может сложиться впечатление, что его Q -фактор всегда более заполнен, чем в исходном профильном разложении. В случаях, когда строится QR -разложение матрицы, близкой к квадратной, это действительно может иметь место. Например, для матрицы, изображенной на рис. 1, Q -фактор ее расширенного профильного QR -разложения имеет вид, показанный на рис. 3 (отсоединенная диагональ не показана), т.е. он действительно значительно более заполнен, чем тот, что показан на рис. 1. Тем не менее, даже для матриц, близких к квадратным, имеются примеры, когда заполнение расширенного профильного Q -фактора заметно меньше. Это имеет место, например, для матрицы, изображенной слева на рис. 4. Рассмотрим вопрос о заполнении Q -фактора QR -разложения подробнее.

Имеется достаточно простой способ определения мелкоблочного заполнения Q -фактора в расширенном профильном QR -разложении. Будем предполагать известной мелкоблочную струк-

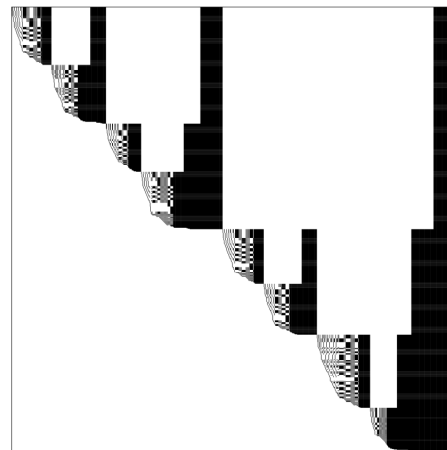


Рис. 3. Заполнение Q -фактора

туру матрицы $C^T C$. По этой структуре можно легко вычислить заполнение R -фактора QR -разложения как результат символической треугольной факторизации мелкоблочной структуры $C^T C$. Но тогда очевидно без учета отсоединенной диагонали, что мелкоблочная структура любого i -го блочного преобразования Хаусхолдера при расширенной профильной факторизации матрицы C есть сумма структуры i -го столбца матрицы C плюс сумма структур предыдущих блочных преобразований Хаусхолдера только для тех предыдущих индексов, для которых имеется соответствующий структурный элемент в наддиагональной части i -го мелкоблочного столбца матрицы R . Это следует из того простого факта, что в силу особенностей структуры расширенного профильного разложения соответствующий элемент в структуре R при разложении может возникнуть только в этом случае, а большего взаимодействия между преобразованиями Хаусхолдера быть также не может, иначе при некоторых числовых значениях будет иметь место дополнительное заполнение в матрице R . В некотором смысле это наименьшее возможное заполнение преобразований Хаусхолдера в поддиагональной части неявного представления Q -фактора. В частности, заполнение блочных преобразований Хаусхолдера в расширенном профильном разложении не зависит от упорядочивания мелкоблочных строк.

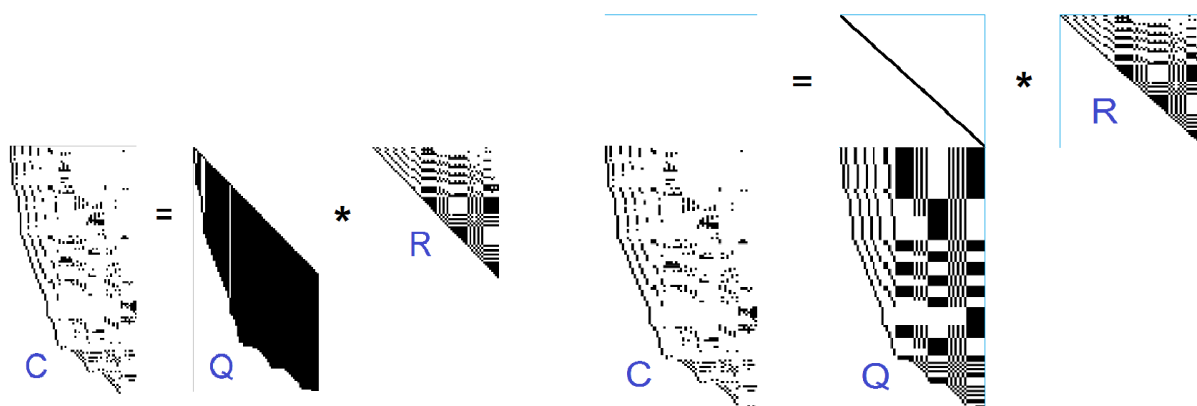


Рис. 4. Профильное (слева) и расширенное профильное (справа) QR -разложения

Более того, для матрицы полного столбцевого ранга, очевидно, структура разреженности любого блочного базисного вектора совпадает со структурой соответствующего блочного преобразования Хаусхолдера для расширенного профильного QR (без учета отсоединенной диагонали). Для матрицы полного ранга имеет место еще более тесная связь с базисными блочными векторами Q -фактора: сам i -й блочный вектор F_s расширенного профильного QR -разложения имеет прямое отношение к i -у базисному блоку направлений Q -фактора. Тем не менее, мы не будем нигде явно использовать эту возможную связь, поскольку в приложениях постоянно приходится иметь дело с матрицами неполного ранга. В случае с обычным профильным разложением заполнение блочных преобразований Хаусхолдера в Q -факторе дополнительно порождается также возникающим мелкоблочным диагональным структурным элементом, из-за чего начинают взаимодействовать те преобразования Хаусхолдера, которые могли бы не взаимодействовать в случае расширенного профильного разложения. Именно это взаимодействие через диагональный элемент и приводит к различию заполнений на рис. 4.

Дополнительное заполнение блочных преобразований Хаусхолдера Q -фактора в наддиагональной части расширенного профильного QR -разложения может быть значительным, как это уже было показано выше. Тем не менее, в частном случае матриц с числом строк, значительно превышающем число столбцов, этим дополнительным заполнением можно пренебречь на фоне возможного избыточного внутривекторного поддиагонального заполнения из-за взаимодействий преобразований Хаусхолдера на диагональном элементе. В контексте интересующих автора приложений число столбцов зачастую суммарно в сотни и тысячи раз меньше числа строк, поэтому в последующей части изложения в качестве базового разложения будет использоваться только расширенное профильное QR -разложение. Расширенное профильное QR -разложение имеет еще одно естественное структурное преимущество. Если число строк в матрице C меньше числа столбцов (что редко, но случается в приложениях из-за, например, дополнительного строчного разбиения матрицы), то в этом случае в профильном разложении нужно делать что-то искусственное, в то время как в расширенном разложении это не приводит к особым ситуациям.

4. QR -разложение матрицы с заданным строчным разбиением. Для прямоугольной матрицы

$C \in \mathbb{R}^{M \times Ns}$ введем ее блочно-строчное разбиение в виде

$$C = \begin{bmatrix} C_1 \\ \vdots \\ C_k \end{bmatrix}, \tag{5}$$

где $C_t \in \mathbb{R}^{M_t \times Ns}$, $t = 1, \dots, k$, и $\sum_{t=1}^k M_t = M$. Пусть для каждой из матриц имеет место QR -разложение типа (2) с блочными преобразованиями Хаусхолдера:

$$C_t = \left(\prod_{j=1}^N \Omega_j^{(t)} \right) \begin{bmatrix} R^{(t)} \\ 0 \end{bmatrix}, \tag{6}$$

$t = 1, \dots, k$. Очевидно, что в мелкоблочной разреженной матрице при введении ее строчного разбиения в строчных мелкоблочных подматрицах может быть много нулевых мелкоблочных столбцов. Для их учета в (6) будем предполагать, что в каждой подматрице C_t имеется ровно N_t ненулевых блочных столбцов, $0 \leq N_t \leq N$, $t = 1, \dots, k$, и что для каждой подматрицы C_t известна такая мелкоблочная перестановка блочных столбцов Ξ_t , $\Xi_t \in \mathbb{R}^{Ns \times Ns}$, что в матрице $[C'_t, 0] = C_t \Xi_t$ ровно первые N_t блочных столбцов ненулевые. Пусть, далее, для матрицы C'_t вычислено ее QR -разложение вида

$$C'_t = \left(\prod_{j=1}^{N_t} \Omega'_j{}^{(t)} \right) \begin{bmatrix} R'^{(t)} \\ 0 \end{bmatrix} \tag{7}$$

с квадратной матрицей $R'^{(t)} \in \mathbb{R}^{N_t s \times N_t s}$; тогда представление (6) получается из (7) неявным добавлением в произведении $(N - N_t)$ блочных множителей с единичной матрицей (частный случай блочного преобразования Хаусхолдера) и с квадратной матрицей $R^{(t)}$ вида

$$R^{(t)} = \begin{bmatrix} R'^{(t)} & 0 \\ 0 & 0 \end{bmatrix} \Xi_t^T. \tag{8}$$

Заметим, что в контексте формулы (8) матрица $R^{(t)}$ в (6) не обязательно верхняя треугольная. Обозначим $\Phi_t = \prod_{j=1}^N \Omega_j^{(t)}$.

Рассмотрим задачу построения QR -разложения всей матрицы из (5) на основе разложений (6). Для этого рассмотрим разреженную матрицу

$$\widehat{C} = \begin{bmatrix} R^{(1)} \\ \vdots \\ R^{(k)} \end{bmatrix} \tag{9}$$

и ее QR -разложение

$$\widehat{C} = \left(\prod_{j=1}^N \widehat{\Omega}_j \right) \begin{bmatrix} \widehat{R} \\ 0 \end{bmatrix}. \tag{10}$$

Обозначим $\widehat{\Phi} = \left(\prod_{j=1}^N \widehat{\Omega}_j \right)$. Имеют место соотношения

$$C = \begin{bmatrix} C^{(1)} \\ \vdots \\ C^{(k)} \end{bmatrix} = \begin{bmatrix} \Phi_1 R^{(1)} \\ \vdots \\ \Phi_k R^{(k)} \end{bmatrix} = \begin{bmatrix} \Phi_1 & 0 \\ \vdots & \\ 0 & \Phi_k \end{bmatrix} \widehat{\Phi} \begin{bmatrix} \widehat{R} \\ 0 \end{bmatrix} = \Phi \begin{bmatrix} \widehat{R} \\ 0 \end{bmatrix}, \tag{11}$$

где

$$\Phi = \begin{bmatrix} \Phi_1 & 0 \\ \vdots & \\ 0 & \Phi_k \end{bmatrix} \widehat{\Phi}, \tag{12}$$

причем $\Phi^T \Phi = I_M$. Отсюда следует, что неявное представление (12) совместно с равенством $C = \Phi \begin{bmatrix} \widehat{R} \\ 0 \end{bmatrix}$ из (11) есть QR -разложение матрицы C , при этом матрица Φ представляет собой Q -фактор QR -разложения, а квадратная верхняя треугольная матрица \widehat{R} есть R -фактор QR -разложения.

Описанная хорошо известная конструкция очевидно позволяет параллельным образом вычислять QR -разложение матрицы за счет введения строчного разбиения. Действительно, строчные QR -разложения можно считать независимо. Синхронизация вычислений происходит только при вычислении объединяющего QR -разложения (10). С другой стороны, понятно, что подобный подход к основному распараллеливанию вычислений может быть эффективен только тогда, когда число столбцов в матрице существенно меньше числа строк, иначе затраты на объединяющее QR -разложение могут доминировать в вычислениях.

Заметим также, что введенное выше промежуточное упорядочивание мелкоблочных столбцов во введенном блочно-строчном разбиении можно использовать не только для того, чтобы избежать обработки нулевых мелкоблочных столбцов, но и для дальнейшей оптимизации затрат памяти и вычислительных затрат при вычислении соответствующих блочно-строчных QR -разложений. В данной работе такая возможность рассматриваться не будет.

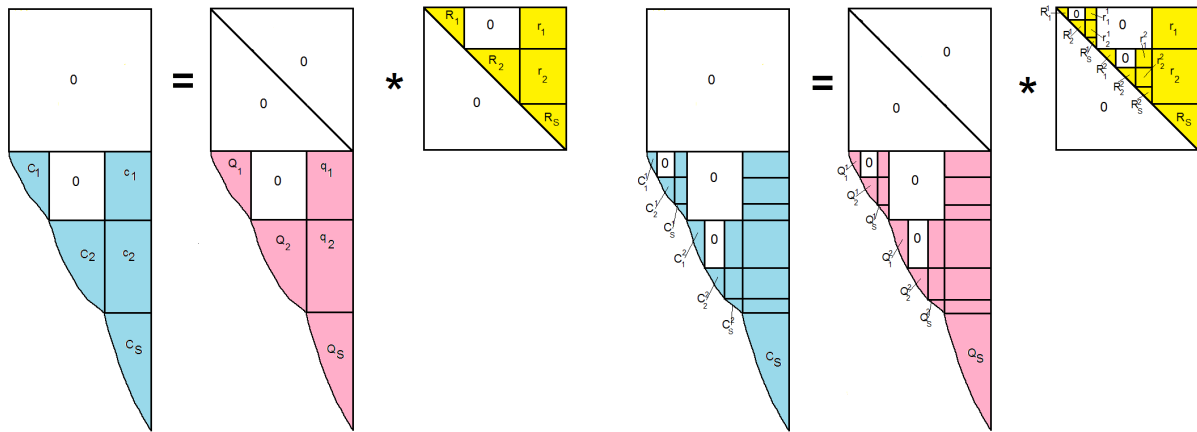


Рис. 5. Двух- (слева) и трех- (справа) уровневая организация параллельного вычисления QR -разложения

5. Верхняя квазитреугольная матрица с разреженностью типа вложенных сечений и параллельный алгоритм построения ее QR -разложения. Описанный в предыдущем разделе параллельный алгоритм вычисления QR -разложения по блочным строкам можно сделать эффективным за счет использования дополнительной столбцевой разреженности матрицы. Для начала рассмотрим двухуровневую организацию вычислений для прямоугольной матрицы со структурой разреженности, изображенной на рис. 5 слева. Пусть число мелкоблочных столбцов в матрицах C_1 , C_2 и C_S равно соответственно N_1 , N_2 и N_S , $N_1 + N_2 + N_S = N$.

Действуя в духе предыдущего раздела, вычислим блочно-строчные QR -разложения:

$$[C_1, c_1] = \left(\prod_{j=1}^{N_1+N_S} \Omega_j^{(1)} \right) \begin{bmatrix} R^{(1)} & r^{(1)} \\ 0 & Z^{(1)} \\ 0 & 0 \end{bmatrix}, \quad [C_2, c_2] = \left(\prod_{j=1}^{N_2+N_S} \Omega_j^{(2)} \right) \begin{bmatrix} R^{(2)} & r^{(2)} \\ 0 & Z^{(2)} \\ 0 & 0 \end{bmatrix}, \quad C_S = \left(\prod_{j=1}^{N_S} \Omega_j^{(S)} \right) \begin{bmatrix} Z_S \\ 0 \end{bmatrix}.$$

В соответствии с (9) объединяющее QR -разложение нужно строить для матрицы

$$\widehat{C} = \begin{bmatrix} R^{(1)} & 0 & r^{(1)} \\ 0 & 0 & Z^{(1)} \\ 0 & R^{(2)} & r^{(2)} \\ 0 & 0 & Z^{(2)} \\ 0 & 0 & Z_S \end{bmatrix}.$$

Если в матрице \hat{C} переставить вторую и третью блочные строки, то получим матрицу \hat{C}' вида

$$\hat{C}' = \begin{bmatrix} R^{(1)} & 0 & r^{(1)} \\ 0 & R^{(2)} & r^{(2)} \\ 0 & 0 & Z^{(1)} \\ 0 & 0 & Z^{(2)} \\ 0 & 0 & Z_S \end{bmatrix}.$$

В первых двух крупноблочных столбцах матрица \hat{C}' уже треугольная, все блочные преобразования Хаусхолдера для них можно считать единичными матрицами, а значит, для построения объединяющего QR -разложения достаточно вычислить QR -разложение матрицы

$$\hat{Z} = \begin{bmatrix} Z^{(1)} \\ Z^{(2)} \\ Z_S \end{bmatrix}. \tag{13}$$

На основе (7) и (8) легко показать, что матрица \hat{Z} в (13) с учетом разреженности имеет мелкоблочный размер $(n_1 + n_2 + N_S) \times N_S$, здесь n_1 и n_2 равны числу ненулевых мелкоблочных столбцов в матрицах c_1 и c_2 соответственно.

Аналогичным образом можно рекурсивно организовать вычисления для трехуровневой матрицы справа на рис. 5. При этом редуцированные объединяющие QR -разложения типа (13) вычисляются только для столбцов соответствующего уровня окаймления.

Обобщая этот алгоритм на общий случай, введем следующее определение.

Определение. Прямоугольную мелкоблочную матрицу с введенным на ней блочным строчным и столбцевым разбиениями будем называть *верхней квазитреугольной L -уровневой матрицей со структурой разреженности типа вложенных сечений*, если в терминах крупных блоков матрица является квадратной верхней треугольной и имеет структуру разреженности типа вложенных сечений, описываемой L -уровневым деревом зависимостей вычислений.

В частности, матрица на рис. 5 слева в терминах определения является двухуровневой верхней квазитреугольной с двухуровневым бинарным деревом зависимостей вычислений. Аналогично, матрица на рис. 5 справа является трехуровневой верхней квазитреугольной с трехуровневым бинарным деревом зависимостей вычислений.

Как следует из предыдущего изложения, для эффективного вычисления QR -разложения верхних квазитреугольных матриц с разреженностью типа вложенных сечений можно использовать следующий параллельный алгоритм:

- 1) параллельно и независимо для каждой блочной строки матрицы строим расширенное профильное разреженное QR -разложение на основе блочных преобразований Хаусхолдера;
- 2) параллельно в порядке, определяемом деревом зависимостей вычислений, достраиваем QR -разложения для объединяющих подматриц типа (13) для вычисления QR -разложений соответствующих мелкоблочных столбцевых окаймлений.

Существует множество способов, при помощи которых разреженную прямоугольную мелкоблочную матрицу C можно привести к виду многоуровневой верхней квазитреугольной матрицы типа вложенных сечений. Простейший из них, например, такой.

По структуре разреженности матрицы $C^T C$ вычисляем упорядочивание вложенных сечений. По структуре переупорядоченной матрицы $C^T C$ строим блочное разбиение и бинарное дерево нужной глубины, которое описывает разреженность этой матрицы в контексте вложенных сечений. Переупорядочиваем мелкоблочные столбцы матрицы C в соответствии с упорядочиванием вложенных сечений. Упорядочивание строк вводим последовательно нумеруя элементы первого мелкоблочного столбца матрицы C , затем второго и т.д. Столбцевое разбиение вводим в соответствии с разбиением вложенных сечений. Строчное разбиение вводим по мере того как заканчиваются столбцевые блоки. В результате получится переупорядоченная по столбцам и строкам матрица с введенным строчным и столбцевым разбиением (рис. 6).

В некоторых случаях удобно ввести дополнительное строчное упорядочивание и разбиение для оптимизации затрат при обработке окаймлений матрицы. Для этого в каждой блочной строке последними

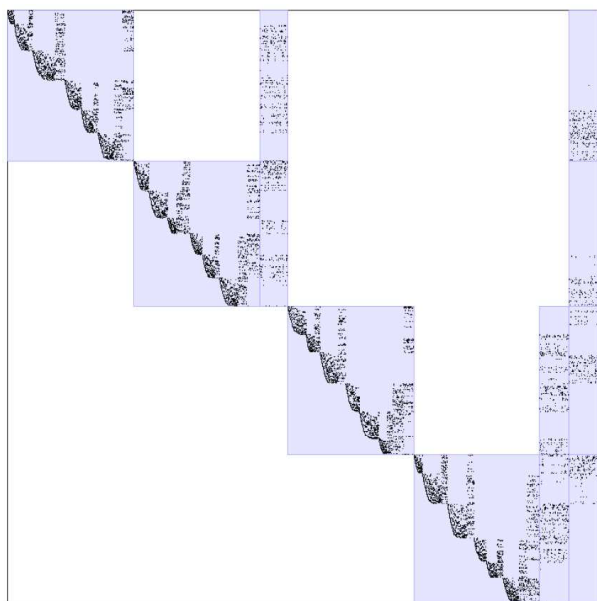


Рис. 6. Упорядочивание и разбиение тестовой матрицы

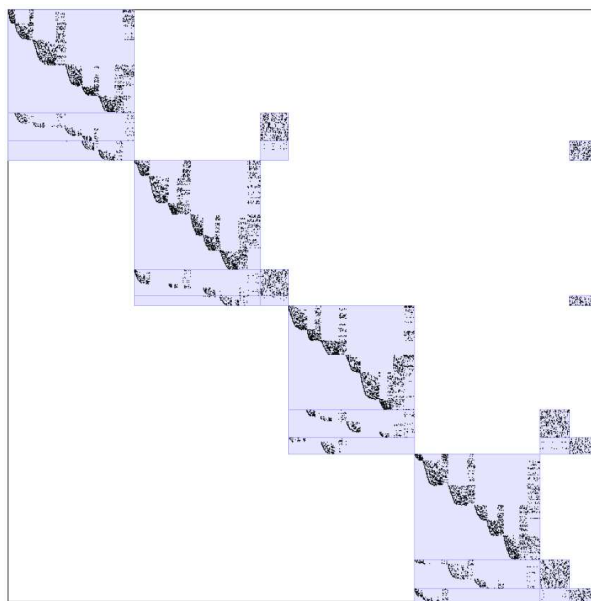


Рис. 7. Дополнительное строчное упорядочивание и разбиение

ставим строки, в которых есть элементы в последнем окаймлении, и вводим новый строчный блок, если такие строки есть; предпоследними ставим строки, которые не были перенесены в конец ранее и которые имеются в предпоследнем окаймлении, и ставим новый строчный блок, если такие строки имеются, и т.д. Для тестовой задачи на рис. 6 получится новое строчное упорядочивание и строчное разбиение, изображенное на рис. 7. На рисунке видно заметно меньшее блочное заполнение в окаймлениях матрицы. Для задачи в таком дополнительном разбиении построить QR -разложение можно параллельным алгоритмом, аналогичным описанному выше. Дополнительным является этап объединяющего строчного QR -разложения в рамках одного узла дерева зависимостей, которое возникает по причине введения дополнительного строчного разбиения в каждой основной подматрице.

6. Построение упорядочиваний и разбиений ND-типа на основе объемного разбиения расчетной сетки на домены. Описанный в предыдущем разделе основанный на явном вычислении упорядочивания ND-способ построения упорядочиваний и разбиений, при которых мелкоблочная матрица становится многоуровневой верхней квазитреугольной типа вложенных сечений, не является практически удобным для множества приложений, в которых матрица возникает как следствие аппроксимации уравнений на расчетной сетке. Для подобных задач необходима специальная методика построения соответствующих упорядочиваний и разбиений. В данном разделе предлагается подобная методика, которая основана на предположении о том, что столбцы и строки матрицы ассоциируются естественным образом с элементами расчетной сетки, а для этой сетки имеется ее геометрически локальное разбиение на достаточно малые объемы — домены.

Таким образом, предположим, что для расчетной области имеется разбиение на домены и что строки и столбцы матрицы ассоциируются с расчетными доменами. Это означает, что можно ввести нумерацию доменов, нумерацию строк и столбцов матрицы и строчное и столбцовое разбиения таким образом, чтобы в общем случае в прямоугольном блоке $(1,1)$ были элементы связей ячеек первого домена между собой, в блоке $(1,2)$ элементы связей ячеек первого домена с ячейками второго и т.д.

Для начала рассмотрим вопрос о построении дерева зависимостей вычислений. Для этого введем в рассмотрение две квадратные разреженные целочисленные матрицы размером $k \times k$, здесь k — полное число доменов. Первую матрицу будем называть матрицей связей. В этой матрице в каждой строке есть элементы только на диагонали, а также в тех позициях (i, j) , в которых есть хоть один структурный элемент в блоке связи между i -м и j -м доменами. Вторую целочисленную матрицу будем называть матрицей весов. Ее структура разреженности совпадает

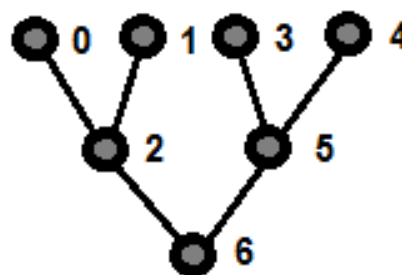


Рис. 8. Нумерация узлов дерева

со структурой разреженности матрицы связей. Диагональные элементы матрицы весов содержат число элементов геометрии в домене (число ячеек), а внедиагональный элемент в позиции (i, j) содержит полное число связей между элементами i -го и j -го доменов. По матрицам связей и весов рекурсивно проводим разбиение всей расчетной области на две большие подобласти, каждый раз содержащие близкое общее количество элементов геометрии (ячеек сетки) и имеющие минимальное количество связей между этими двумя подобластями. Это разбиение продолжается рекурсивно до тех пор, пока в подобласти имеется более чем один домен. Таким образом возникает представление набора доменов в виде квазибинарного дерева.

Пусть L — число уровней этого квазибинарного дерева. Все домены являются листьями некоторого уровня дерева. Построенное квазибинарное дерево и будем считать требуемым L -уровневым деревом зависимостей вычислений. Упорядочим узлы дерева зависимостей таким образом, чтобы каждый узел дерева, имеющий сыновей, имел номер, не меньший чем номера всех узлов поддеревьев этого узла. Такое упорядочивание номеров узлов дерева очевидно существует. На рис. 8 показана нумерация узлов для трехуровневого бинарного дерева.

Для окончательного построения строчного и столбцевого упорядочиваний матрицы, а также ее строчного и столбцевого разбиений рассмотрим подробнее структуру каждого домена (двумерный пример приведен на рис. 9).

В каждом домене выберем такой набор его геометрических элементов, который не имеет связей с внешними геометрическими элементами (подобласть 0 на рис. 9), в терминах трехмерных задач назовем его объемной внутренней частью домена. Из оставшихся элементов домена выделим связанные подобласти геометрических элементов, связанных ровно с одним фиксированным внешним доменом (подобласти 1–4 на рис. 9), в терминах трехмерных задач назовем эти подобласти квазиповерхностными (квазидвумерными) границами домена. Далее из оставшихся геометрических элементов домена выделим связанные подобласти геометрических элементов, связанные ровно с двумя фиксированными внешними доменами (подобласти 5–8 на рис. 9), в терминах трехмерных задач назовем эти подобласти квазиодномерными границами домена. Для не менее чем трехмерных задач далее из оставшихся геометрических элементов выделим связанные подобласти геометрических элементов, связанных ровно более чем с двумя внешними доменами; в терминах трехмерных задач назовем эти подобласти квазинульмерными границами домена.

Упорядочим матрицу в соответствии с введенным разбиением и упорядоченным квазибинарным деревом зависимостей следующим образом. Для этого введем соответствие между узлами дерева и введенными новыми строчным и столбцевым блочными разбиениями. Поместим каждый столбцевой блок в соответствии с его связями следующим образом. Внутренние столбцевые блоки домена поместим в соответствующий домену лист дерева. Столбцевые квазидвумерные границы всех доменов поместим в наименьший по номеру узел упорядоченного дерева зависимостей, в поддереве которого имеются узлы дерева, содержащие соответствующие родительские для этой квазидвумерной границы 2 домена. Аналогично, столбцевые квазиодномерные границы всех доменов поместим в наименьший по номеру узел упорядоченного дерева зависимостей, в поддереве которого имеются узлы дерева, содержащие соответствующие родительские для этой квазиодномерной границы 3 домена. Наконец, остальные столбцевые квазинульмерные границы всех доменов поместим в наименьший по номеру узел упорядоченного дерева зависимостей, в поддереве которого имеются все узлы дерева, содержащие соответствующие домены, родительские для этой квазинульмерной границы с номером, большим 3. Строчные подблоки каждого домена разместим в тех узлах упорядоченного дерева зависимостей, в котором они встречаются в первый раз, если блочные столбцы заранее упорядочить в соответствии с описанным выше алгоритмом. Далее, все блочные столбцы в каждом узле упорядочим по их мерности — чем больше мерность, тем раньше расположен столбцевой блок. После подобных преобразований, очевидно, матрица станет L -уровневой верхней квазитреугольной матрицей типа вложенных сечений, а ее структура разреженности будет иметь вид, подобный изображенному на рис. 7. Для уменьшения заполнения Q - и R -факторов QR -разложений в

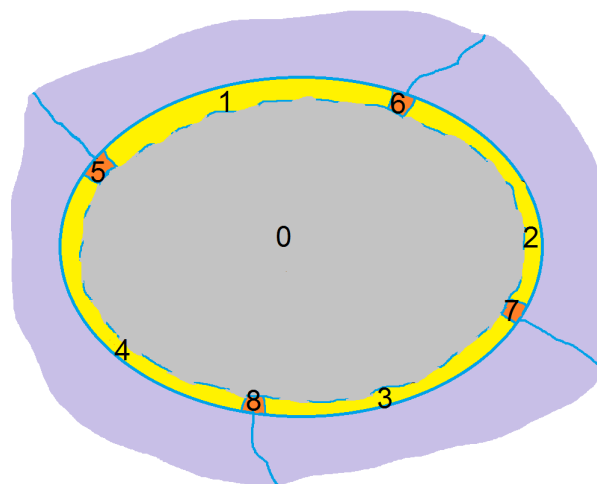


Рис. 9. Геометрическое разбиение двумерного домена на подобласти

столбцевых блоках узлов (листьях дерева) можно дополнительно использовать упорядочивание ND, как это было описано выше.

7. Заключение. В настоящей статье представлен параллельный алгоритм вычисления QR -разложения многоуровневой разреженной верхней квазитреугольной матрицы со структурой разреженности типа вложенных сечений. Базовым вычислительным примитивом алгоритма являются блочные разреженные преобразования Хаусхолдера. Предложены варианты алгоритмов, приводящих заданную матрицу к упомянутому виду, удобному для параллельного вычисления ее разреженного QR -разложения, в том числе для задач аппроксимации уравнений на расчетной сетке. Результаты численных экспериментов с предложенным алгоритмом для тестовых задач на гибридной параллельной MPI+threads+SIMD архитектуре приведены в работе [6].

Статья рекомендована к публикации Программным комитетом Международной научной конференции “Суперкомпьютерные дни в России 2015” (<http://russianscdays.org>).

СПИСОК ЛИТЕРАТУРЫ

1. Тьртъшников Е.Е. Методы численного анализа. М.: Академия, 2007.
2. Джордж А., Лю Дж. Численное решение больших разреженных систем уравнений. М.: Мир, 1984.
3. Kaporin I.E. High quality preconditioning of a general symmetric positive definite matrix based on its $U^tU+U^tR+R^tU$ -decomposition // Numer. Linear Algebra Appl. 1998. **5**, N 6. 483–509.
4. Davis T.A. Algorithm 915: SuiteSparseQR, a multifrontal multithreaded sparse QR factorization package // ACM Trans. Math. Softw. 2011. **38**, N 1. 8:1–8:22.
5. Yeralan S.N., Davis T.A., Ranka S. Algorithm 9xx: sparse QR -factorization on the GPU // ACM Trans. Math. Softw. 2015. **1**, N 1. doi 10.1145/0000000.0000000.
6. Li N., Saad Y. MIQR: a multilevel incomplete QR -preconditioner for large sparse least-squares problems // SIAM J. Matrix Anal. Appl. 2006. **28**, N 2. 524–550.
7. Харченко С.А., Ющенко А.А. Параллельная реализация алгоритма разреженного QR разложения для прямоугольных верхних квазитреугольных матриц со структурой типа вложенных сечений // Вестник ЮУрГУ (в печати).
8. Rotella F., Zambettakis I. Block Householder transformation for parallel QR factorization // Appl. Math. Letters. 1999. **12**, N 4. 29–34.
9. Haidar A., Dong T.T., Tomov S., et al. A framework for batched and GPU-resident factorization algorithms applied to block Householder transformations // Lecture Notes in Computer Science. Vol. 9137. Heidelberg: Springer, 2015. 31–47.

Поступила в редакцию
04.09.2015

A Parallel Algorithm for the Sparse QR Decomposition of a Rectangular Upper Quasi-Triangular Matrix with ND-Type Sparsity

S. A. Kharchenko¹

¹ Limited Liability Company “TESIS”; ulitsa Yunnatov 18, Moscow, 127083, Russia; Leading Programmer, e-mail: skh@tesis.com.ru

Received September 4, 2015

Abstract: An algorithm for computing the sparse QR decomposition of a specially ordered rectangular matrix is proposed. This decomposition is based on the block sparse Householder transformations. For ordering computations, the nested dissection ordering is used for the matrix A^tA , where A is the original rectangular matrix. For mesh based problems, the ordering can be constructed starting from an appropriate volume partitioning of the computational mesh. Parallel computations are based on sparse QR decomposition for sets of rows with an additional initial zero block.

Keywords: sparse rectangular matrix, upper quasi-triangular matrix, volume partitioning, nested dissection, QR decomposition, Householder transformations, parallel algorithm.

References

1. E. E. Tyrtysnikov, *Methods of Numerical Analysis* (Akademiya, Moscow, 2007) [in Russian].
2. A. George and J. W.-H. Liu, *Computer Solution of Large Sparse Positive Definite Systems* (Prentice-Hall, Englewood Cliffs, 1981; Mir, Moscow, 1984).
3. I. E. Kapurin, "High Quality Preconditioning of a General Symmetric Positive Definite Matrix Based on its $U^T U + U^T R + R^T U$ -Decomposition," *Numer. Linear Algebra Appl.* **5** (6), 483–509 (1998).
4. T. A. Davis, "Algorithm 915: SuiteSparseQR, a Multifrontal Multithreaded Sparse QR Factorization Package," *ACM Trans. Math. Softw.* **38** (1), 8:1–8:22 (2011).
5. S. N. Yeralan, T. A. Davis, and S. Ranka, "Algorithm 9xx: Sparse QR Factorization on the GPU," *ACM Trans. Math. Softw.* **1** (1) (2015). doi 10.1145/0000000.0000000
6. N. Li and Y. Saad, "MIQR: A Multilevel Incomplete QR Preconditioner for Large Sparse Least-Squares Problems," *SIAM J. Matrix Anal. Appl.* **28** (2), 524–550 (2006).
7. S. A. Kharchenko and A. A. Yushchenko, "Parallel Implementation of Sparse QR Decomposition of a Rectangular Upper Quasitriangular Matrix with ND-Type Sparsity," *Vestn. South Ural State Univ. Ser. Vychisl. Mat. Inf.*, 2015 (in press).
8. F. Rotella and I. Zambettakis, "Block Householder Transformation for Parallel QR Factorization," *Appl. Math. Lett.* **12** (4), 29–34 (1999).
9. A. Haidar, T. T. Dong, S. Tomov, et al., "A Framework for Batched and GPU-Resident Factorization Algorithms Applied to Block Householder Transformations," in *Lecture Notes in Computer Science* (Springer, Heidelberg, 2015), Vol. 9137, pp. 31–47.