

Параллельный алгоритм разреженного QR разложения для прямоугольных верхних квази-треугольных матриц со структурой типа вложенных сечений

С.А. Харченко

ООО "ТЕСИС", Вычислительный Центр РАН

В работе рассматривается параллельный алгоритм вычисления разреженного QR разложения специальным образом упорядоченной прямоугольной матрицы на основе разреженных блочных преобразования 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 биения расчетной сетки. В качестве базового алгоритма для организации параллельных вычислений используется QR разложение для наборов строк матрицы с дополнением в виде нулевого начального блока. Алгоритм предполагается использовать в качестве вычислительного ядра разрабатываемых автором параллельных итерационных алгоритмов решения СЛАУ и задач наименьших квадратов на основе композиции подпространств, порождаемых разреженными базисами.

1. Введение

QR разложение прямоугольной матрицы является одним из базовых вычислительных алгоритмов для многих задач вычислительной математики. В частности, подобные вычисления возникают при решении СЛАУ, при решении задач наименьших квадратов и задач на собственные значения [1], и т.д. Возможность эффективно параллельным образом вычислять QR разложение разреженной матрицы в некоторых случаях означает возможность использования новых классов вычислительных алгоритмов, и 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 вычислений при параллельном вычислении традиционных 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] Тима Дэвиса с соавторами. В этих работах предложен мульти- фронтальный 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 данной работе рассматриваются идейно близкие алгоритмы параллельного вычисления разреженного QR разложения матрицы. Основные отличия предлагаемого в работе подхода состоят в следующем:

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

Данная работа является теоретической основой для практической работы [6], содержащей описание реализации алгоритма на гибридной параллельной MPI+threads+SIMD архитектуре и начальные результаты экспериментов на такой архитектуре для тестовых задач. Также работа является базовой для планируемой серии работ автора по новым параллельным итерационным

алгоритмам решения СЛАУ и задач наименьших квадратов на основе композиции подпространств, порождаемых разреженными базисами.

Работа построена следующим образом. В Разделе 2 описывается блочное преобразование Хаусхолдера [7] и его применение для профильной разреженной QR факторизации прямоугольной матрицы, аналогичной используемым 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 биения матрицы. В Разделе 5 вводится понятие верхней квази-треугольной 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 биений, которые приводят матрицу, заданную на объемной сетке с известным объемным геометрических биением, к виду верхней квази-треугольной 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]:

$$G = I_M + \tau f f^T, \quad (2.1)$$

где $f \in \mathbb{R}^M$ - некоторый вектор, τ - некоторое вещественное число. При этом строится преобразование Хаусхолдера как правило из условия, чтобы для некоторого вектора $c \in \mathbb{R}^M$ в векторе $\hat{c} = G^T c$ 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 (2.1) приводит к соотношениям:

$$C = (\prod_{j=1}^s G_j) \begin{bmatrix} R \\ 0 \end{bmatrix}, \quad (2.2)$$

где матрица $R \in \mathbb{R}^{s \times s}$ - квадратная верхняя треугольная, а матрицы $G_j \in \mathbb{R}^{M \times M}$ - ортогональные матрицы преобразования Хаусхолдера, имеющие вид $G_j = I_M + \tau_j f_j f_j^T$, где 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 (2.2) достаточно вычислить результат умножения:

$$q_i = (\prod_{j=1}^s G_j) e_i, \quad (2.3)$$

где e_i - i -й единичный орт.

Обозначим $\Omega_s = \prod_{j=1}^s G_j$. Рассмотрим хорошо известное [7] блочное обобщение преобразований Хаусхолдера (2.1) в виде

$$\Omega_s = I_M + F_s T_s F_s^T, \quad (2.4)$$

где $\Omega_s \in \mathbb{R}^{M \times M}$, $\Omega_s^T \Omega_s = I_M$, $F_s \in \mathbb{R}^{M \times s}$, $T_s \in \mathbb{R}^{s \times s}$. Решать задачу построения блочного преобразования Хаусхолдера (2.4) будем по индукции.

Для $s = 1$ положим $F_1 = f_1$, $T_1 = \tau_1$. Пусть для некоторого $s \geq 1$ блочное преобразование Хаусхолдера (2.4) уже построено. Построим это преобразование для $s' = s + 1$. По определению и по предположению индукции имеют место равенства:

$$\begin{cases} \Omega_s = I_M + F_s T_s F_s^T, \\ \Omega_{s+1} = \Omega_s G_{s+1}. \end{cases}$$

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 (2.4) 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 одно- векторными 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 мелко блочно 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 столбцевые 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 столбцевого и строчного мелко блочного биения. В противовес мелким плотным $s \times s$ блокам мы будем говорить также о крупно- блочном или просто блочном биении, строчном и столбцевом. Это всегда будет означать, что соответствующие подматрицы составлены из некоторого количества мелко блочных строк и столбцов. При этом под блочным преобразованием Хаусхолдера мы всегда будем иметь в виду преобразование (2.4) для одного мелко блочного столбца.

2.2 Профильное разреженное QR разложение

Блочные преобразования Хаусхолдера (2.4) можно применить для вычисления разреженного QR разложения разреженной матрицы $C \in \mathbb{R}^{M \times N_s}$, составленной из N мелко блочных 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 или упорядочивание RCM [2]. Это определяет упорядочивание мелко блочных столбцов матрицы C . Далее упорядочивание мелко блочных строк можно выбрать последовательно нумеруя сначала элементы первого столбца C , потом второго, и т.д. Переупорядоченная таким образом матрица будет иметь вид типа приведенного на Рисунке 1 слева. В качестве базового столбцевого упорядочивания для этой матрицы было выбрано упорядочивание ND.

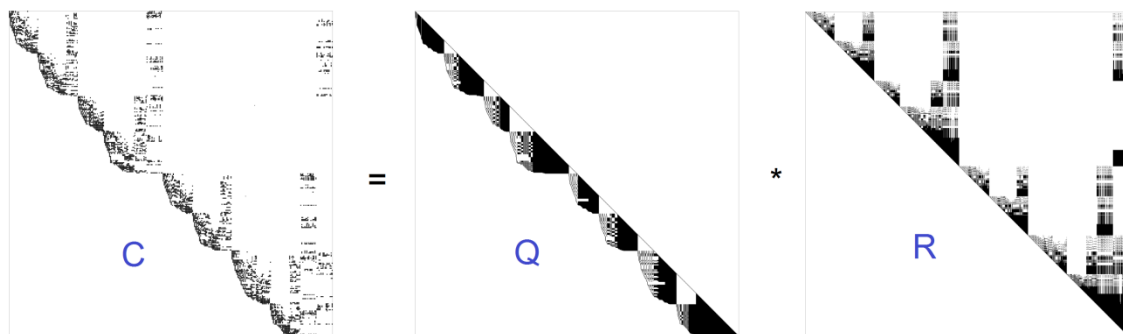


Рис. 1. Структуры разреженности матрицы C и ее Q- и R- факторов для тестовой задачи

При построении так называемого профильного разреженного QR разложения таким образом упорядоченной мелко блочной матрицы C будем действовать по аналогии со случаем плотной матрицы. По первому мелко блочному 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 мелко блочному столбцу, для полученного результата строим следующее разреженное блочное преобразование Хаусхолдера для обнуления элементов под второй мелко блочной диагональю, и т.д. На Рис. 1 справа показана мелко блочная 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 изображены на Рисунке 2. Фактически 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.

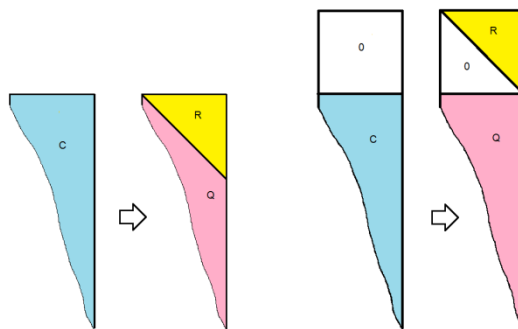


Рис. 2. Профильное (слева) и расширенное профильное (справа) QR разложения

Покажем прежде всего, что в случае матрицы C полного столбцевого ранга профильное и расширенное профильное QR разложения математически эквивалентны. Обозначим $\check{C} = \begin{bmatrix} 0 \\ C \end{bmatrix}$. Для матрицы полного столбцевого ранга C матрица R невырождена и является очевидно треугольным фактором матрицы $C^T C$, а значит матрица R является R- фактором и для матрицы \check{C} , поскольку $\check{C}^T \check{C} = C^T C$. Далее, весь набор $Q = [q_1, \dots, q_{N_S}]$ базисных векторов Q- фактора QR разложения можно, в соответствии с мелко блочным аналогом формулы (2.3) 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 $\check{Q} = \check{C}R^{-1}$, но тогда очевидно $Q = CR^{-1} = \begin{bmatrix} 0 & I_M \end{bmatrix} \check{C}R^{-1} = \begin{bmatrix} 0 & I_M \end{bmatrix} \check{Q}$. 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 $\check{C} =$

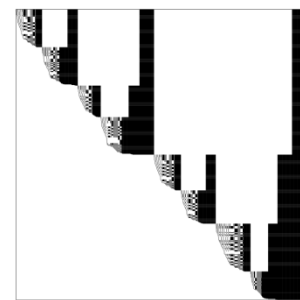


Рис. 3. Заполнение Q фактора

$\tilde{Q}R$. В случае неполного столбцевого ранга эквивалентности нет, и это связано именно с тем, что в этом случае некоторые начальные компоненты базисных векторов расширенного QR 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 зачастую действительно может быть так. Например, для матрицы, изображенной на Рисунке 1, Q- фактор ее расширенного профильного QR имеет вид, показанный на Рисунке 3 (отсоединенная диагональ не показана), т.е., действительно значительно более заполнен чем тот, что показан на Рисунке 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 Рисунке 4. Рассмотрим вопрос о заполнении Q- фактора QR разложения подробнее.

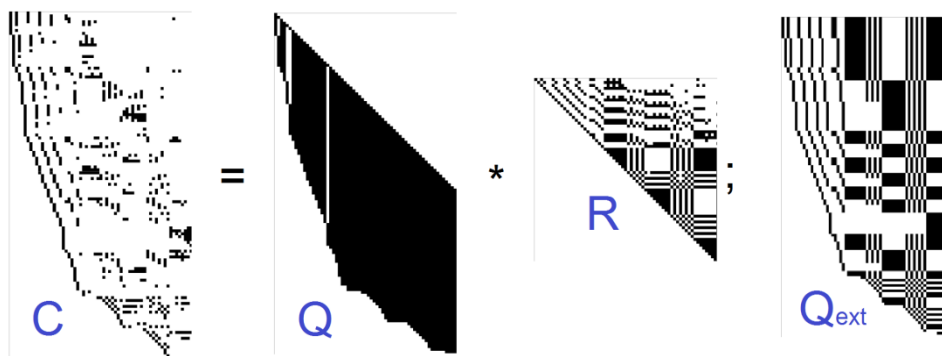


Рис. 4. Профильное (слева) и Q фактор расширенного профильного (справа) QR разложения

Имеется достаточно простой способ определения мелко блочного заполнения Q фактора в расширенном профильном QR. Будем предполагать известной мелко блочную структуру матрицы $C^T C$. По этой структуре можно легко вычислить заполнение R- фактора QR разложения как результат символьной треугольной факторизации мелко блочной структуры $C^T C$. Но тогда очевидно без учета отсоединенной диагонали мелко блочная структура любого i -го блочного преобразования Хаусхолдера при расширенной профильной факторизации C есть сумма структуры i -го столбца C плюс сумма структур предыдущих блочных преобразований Хаусхолдера только для тех предыдущих индексов, для которых имеется соответствующий структурный элемент в над-диагональной части i -го мелко блочного 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 R . В некотором смысле это наименьшее возможное заполнение преобразований Хаусхолдера в под-диагональной части неявного представления Q- фактора. В частности, заполнение блочных преобразований Хаусхолдера в расширенном профильном разложении не зависит от упорядочивания мелко блочных строк. Более того, для матрицы полного 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 имеет прямое отношение к i -му 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 мелко блочно диагональным структурным элементом, из-за чего начинают взаимодействовать те преобразования Хаусхолдера, которые могли бы не взаи-

действовать в случае расширенного профильного разложения. Именно это взаимодействие через диагональный элемент и приводит к разнице заполнений на Рисунке 4.

Дополнительное заполнение блочных преобразований Хаусхолдера Q- фактора в над- диагональной 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 внутри профильного под- диагонального заполнения из-за взаимодействий преобразований Хаусхолдера на диагональном элементе. В контексте интересующих автора приложений число столбцов всегда будет 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 биения 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 биением

Для прямоугольной матрицы $C \in \mathbb{R}^{M \times N_s}$ введем ее блочно строчное биение в виде:

$$C = \begin{bmatrix} C_1 \\ \vdots \\ C_k \end{bmatrix}, \quad (4.1)$$

где $C_t \in \mathbb{R}^{M_t \times N_s}$, $t = 1, \dots, k$, и $\sum_{t=1}^k M_t = M$. Пусть для каждой из матриц имеет место QR разложение типа (2.2) с блочными преобразованиями Хаусхолдера:

$$C_t = \left(\prod_{j=1}^{N_t} \Omega_j^{(t)} \right) \begin{bmatrix} R^{(t)} \\ 0 \end{bmatrix}, \quad (4.2)$$

$t = 1, \dots, k$. Очевидно, что в мелко блочной разреженной матрице при введении ее строчного биения в строчных мелко блочных подматрицах может быть много нулевых мелко блочных столбцов. Для их учета в (4.2) 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 мелко блочная перестановка блочных столбцов Π_t , $\Pi_t \in \mathbb{R}^{N_s \times N_s}$, что в матрице $\hat{C}_t = C_t \Pi_t$ ровно первые N_t блочных столбцов ненулевые. Пусть далее для матрицы \hat{C}_t вычислено ее QR разложение вида

$$\hat{C}_t = \left(\prod_{j=1}^{N_t} \hat{\Omega}_j^{(t)} \right) \begin{bmatrix} \hat{R}^{(t)} \\ 0 \end{bmatrix} \quad (4.3)$$

с квадратной матрицей $\hat{R}^{(t)} \in \mathbb{R}^{N_t \times N_t}$, тогда представление (4.2) получается из (4.3) неявным добавлением в произведении $(N - N_t)$ блочных множителей с единичной матрицей (частный случай блочного преобразования Хаусхолдера) и с квадратной матрицей $R^{(t)}$ вида:

$$R^{(t)} = \begin{bmatrix} \hat{R}^{(t)} & 0 \\ 0 & 0 \end{bmatrix} \Pi_t^T. \quad (4.4)$$

Заметим, что в контексте формулы (4.4) матрица $R^{(t)}$ в (4.2) не обязательно верхняя треугольная. Обозначим $\Phi_t = \prod_{j=1}^{N_t} \Omega_j^{(t)}$.

Рассмотрим задачу построения QR разложения всей матрицы из (4.1) на основе разложений (4.2). Для этого рассмотрим разреженную матрицу

$$\hat{C} = \begin{bmatrix} R^{(1)} \\ \vdots \\ R^{(k)} \end{bmatrix} \quad (4.5)$$

и ее QR разложение

$$\hat{C} = \left(\prod_{j=1}^N \hat{\Omega}_j \right) \begin{bmatrix} \hat{R} \\ 0 \end{bmatrix}. \quad (4.6)$$

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

$$C = \begin{bmatrix} C_1 \\ \vdots \\ C_k \end{bmatrix} = \begin{bmatrix} \Phi_1 R^{(1)} \\ \vdots \\ \Phi_k R^{(k)} \end{bmatrix} = \begin{bmatrix} \Phi_1 & & 0 \\ & \ddots & \\ 0 & & \Phi_k \end{bmatrix} \widehat{\Phi} \begin{bmatrix} \widehat{R} \\ 0 \end{bmatrix} = \Phi \begin{bmatrix} \widehat{R} \\ 0 \end{bmatrix}, \quad (4.7)$$

где

$$\Phi = \begin{bmatrix} \Phi_1 & & 0 \\ & \ddots & \\ 0 & & \Phi_k \end{bmatrix} \widehat{\Phi}, \quad (4.8)$$

причем $\Phi^T \Phi = I_M$. Отсюда следует, что неявное представление (4.8) совместно с равенством $C = \Phi \begin{bmatrix} \widehat{R} \\ 0 \end{bmatrix}$ из (4.7) 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 конструкция очевидно позволяет параллельным образом вычислять QR разложение матрицы за счет введения строчного биения. Действительно, строчные QR разложения можно считать независимо. Синхронизация вычислений происходит только при вычислении объединяющего QR разложения (4.6). С другой стороны понятно, что подобный подход к основному распараллеливанию вычислений может быть эффективен только если число столбцов в матрице существенно меньше числа строк, иначе затраты на объединяющее QR разложение могут доминировать в вычислениях.

5. Верхняя квази-треугольная 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 столбцовой разреженности матрицы.

Для начала рассмотрим двух-уровневую организацию вычислений для прямоугольной матрицы со структурой разреженности, изображенной на Рисунке 5 слева. Пусть число мелко блочных столбцов в матрицах C_1, C_2 и C_S равны соответственно N_1, N_2 и N_S соответственно, $N_1 + N_2 + N_S = N$.

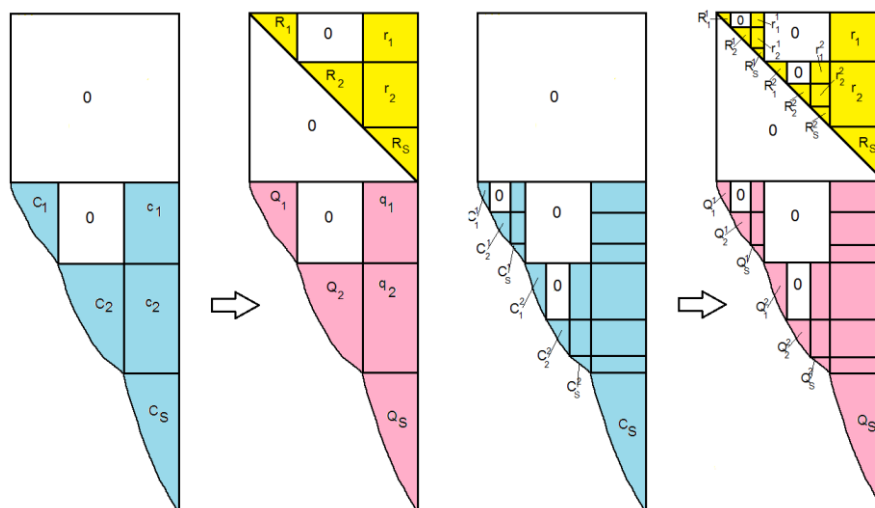


Рис. 5. Двух- (слева) и трех- (справа) уровневая организация параллельного вычисления QR разложения

Действуя в духе предыдущего раздела вычислим блочно строчные QR разложения:

$$\begin{cases} [C_1, c_1] = \left(\prod_{j=1}^{N_1+N_S} \Omega_j^{(1)} \right) \begin{bmatrix} R^{(1)} & r^{(1)} \\ 0 & Z^{(1)} \\ 0 & 0 \end{bmatrix}, \\ [C_2, c_2] = \left(\prod_{j=1}^{N_2+N_S} \Omega_j^{(2)} \right) \begin{bmatrix} R^{(2)} & r^{(2)} \\ 0 & Z^{(2)} \\ 0 & 0 \end{bmatrix}, \\ c_S = \left(\prod_{j=1}^{N_S} \Omega_j^{(S)} \right) \begin{bmatrix} Z_S \\ 0 \end{bmatrix}. \end{cases} \quad (5.1)$$

В соответствии с (4.5) объединяющий QR нужно строить для матрицы

$$\hat{C} = \begin{bmatrix} R^{(1)} & 0 & r^{(1)} \\ 0 & 0 & Z^{(1)} \\ 0 & R^{(2)} & r^{(2)} \\ 0 & 0 & Z^{(2)} \\ 0 & 0 & Z_S \end{bmatrix}. \quad (5.2)$$

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

$$\hat{\hat{C}} = \begin{bmatrix} R^{(1)} & 0 & r^{(1)} \\ 0 & R^{(2)} & r^{(2)} \\ 0 & 0 & Z^{(1)} \\ 0 & 0 & Z^{(2)} \\ 0 & 0 & Z_S \end{bmatrix}. \quad (5.3)$$

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

$$\hat{Z} = \begin{bmatrix} Z^{(1)} \\ Z^{(2)} \\ Z_S \end{bmatrix}. \quad (5.4)$$

С учетом формул (4.3) и (4.4) легко показать, что матрица \hat{Z} в (5.4) с учетом разреженности имеет мелко блочный размер $(n_1 + n_2 + N_S) \times N_S$, здесь n_1 и n_2 соответственно число ненулевых мелко блочных столбцов в матрицах c_1 и c_2 соответственно.

Аналогичным образом можно рекурсивно организовать вычисления для трех- уровневой матрицы справа на Рисунке 5. При этом редуцированные объединяющие QR разложения типа (5.4) вычисляются только для столбцов соответствующего уровня окаймления.

Обобщая этот алгоритм на общий случай введем следующее определение.

Определение 1. Прямоугольную мелко блочную матрицу с введенным на ней блочным строчным и столбцевым биениями будем называть *верхней квази- треугольной 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 Рисунке 5 слева в терминах Определения 1 является двух уровневой верхней квази- треугольной с двух уровневом бинарным деревом зависимостей вычислений. Аналогично, матрица на Рисунке 5 справа является трех уровневой верхней квази- треугольной с трех уровневом бинарным деревом зависимостей вычислений.

Как следует из предыдущего изложения, для эффективного вычисления QR разложения верхних квази- треугольных 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 (5.4) для вычисления QR разложений соответствующих мелко блочных столбцевых окаймлений.

Существует множество способов, при помощи которых разреженную прямоугольную мелко блочную матрицу C можно привести к виду многоуровневой верхней квази-треугольной 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 биение и бинарное дерево нужной глубины, которое описывает разреженность этой матрицы в контексте вложенных сечений. Переупорядочиваем мелко блочные столбцы матрицы C в соответствии с упорядочиванием вложенных сечений. Упорядочивание строк вводим последовательно нумеруя элементы первого мелко блочного столбца C , затем второго и т.д. Столбцовое биение вводим в соответствии с биением вложенных сечений. Строчное биение вводим по мере того как заканчиваются столбцовые блоки. В результате получится переупорядоченная по столбцам и строкам матрица с введенным строчным и столбцовым биением типа изображенной на Рисунке 6.

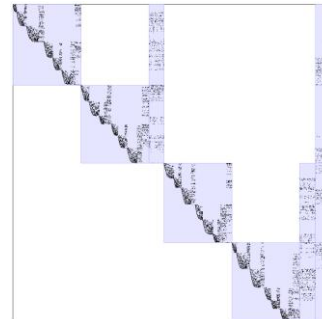


Рис. 6. Упорядочивание и биение тестовой матрицы

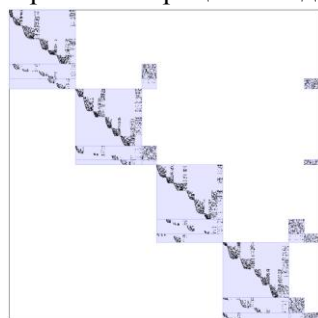


Рис. 7. Дополнительное строчное упорядочивание и биение

В некоторых случаях удобно ввести дополнительное строчное упорядочивание и биение для оптимизации затрат при обработке окаймлений матрицы. Для этого в каждой блочной строке последними 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 Рисунке 6 получится новое строчное упорядочивание и строчное биение, изображенное на Рисунке 7. На рисунке видно заметно меньшее блочное заполнение в окаймлениях матрицы. Для задачи в таком дополнительном биении 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 биения в каждой основной подматрице.

6. Построение упорядочиваний и биений ND-типа на основе объемного биения расчетной сетки на домены

Описанный в предыдущем разделе основанный на явном вычислении упорядочивании ND способ построения упорядочиваний и биений, при которых мелко блочная матрица становится многоуровневой верхней квази-треугольной 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 биений. В данном разделе предлагается подобная методика, которая основана на предположении о том, что столбцы и строки матрицы ассоциируются естественным образом с элементами расчетной сетки, и 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 возможно ввести нумерацию доменов, нумерацию строк и столбцов матрицы и строчное и столбцовое биения таким образом, чтобы в общем случае в прямоугольном блоке (1,1) были элементы связей первого домена между собой, в блоке (1,2) элементы связей первого домена со вторым, и т.д.

Для начала рассмотрим вопрос о построении дерева зависимостей вычислений. Для этого введем в рассмотрение две квадратных разреженных целочисленных 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 -тым и j -тым доменами. Вторую целочисленную матрицу будем называть *матрицей весов*. Ее структура разреженности совпадает со структурой разреженности матрицы связей. Диагональные элементы матрицы весов содержат число элементов геометрии в домене (число ячеек), а вне- диагональный элемент

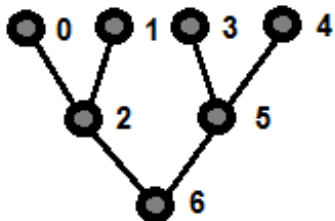


Рис. 8. Нумерация узлов дерева

никает представление набора доменов в виде квази- бинарного дерева. Пусть L - число уровней этого квази- бинарного дерева. Все домены являются листьями некоторого уровня дерева. Построенное квази- бинарное дерево и будем считать требуемым L - уровнем 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 Рисунке 8 показана нумерация узлов для трех-уровневого бинарного дерева.

Для окончательного построения упорядочиваний матрицы, ее строчного и столбцевого биений рассмотрим подробнее структуру каждого домена (двумерный пример приведен на Рисунке 9).

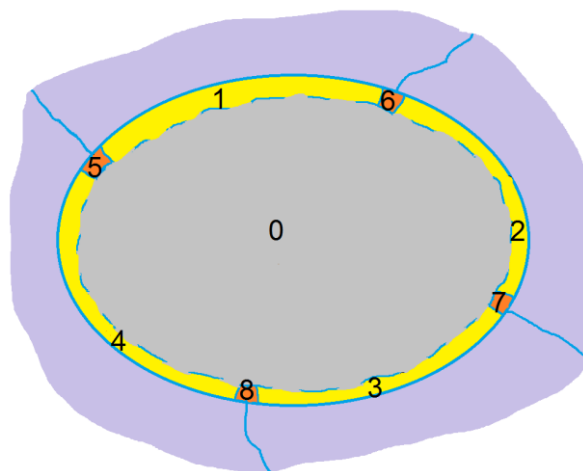


Рис. 9. Геометрическое разбиение двумерного домена на подобласти

В каждом домене выберем такой набор его геометрических элементов, который не имеет связей с внешними геометрическими элементами (подобласть 0 на Рисунке 9), в терминах 3-х мерных задач назовем его объемной внутренней частью домена. Из оставшихся элементов домена выделим связанные подобласти геометрических элементов, связанных ровно с одним внешним доменом (подобласти 1-4 на Рисунке 9), в терминах 3-х мерных задач назовем эти подобласти квази- поверхностными (квази- двумерными) границами домена. Далее из оставшихся геометрических элементов домена выделим связанные подобласти геометрических элементов, связанных ровно с двумя внешними доменами (подобласти 5-8 на Рисунке 9), в терминах 3-х мерных задач назовем эти подобласти квази- одномерными 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 мерных задач назовем эти подобласти квази нуль- мерными границами домена.

Упорядочим матрицу в соответствии с введенным биением и упорядоченным квази-бинарным деревом зависимостей следующим образом. Для этого введем соответствие между узлами дерева и введенными новыми строчным и столбцевым блочными биениями. Поместим каждый столбцевой блок в соответствии с его связями следующим образом. Внутренние столбцевые блоки домена поместим в соответствующий домену лист дерева. Столбцевые квази-двухмерные границы всех доменов поместим в наименьший по номеру узел упорядоченного дерева зависимостей, в сыновьях которого имеются узлы дерева, содержащие соответствующие родительские для этой квази-двухмерной границы 2 домена. Аналогично, столбцевые квази-одномерные границы всех доменов поместим в наименьший по номеру узел упорядоченного дерева зависимостей, в сыновьях которого имеются узлы дерева, содержащие соответствующие родительские для этой квази-одномерной границы 3 домена. Наконец, остальные столбцевые квази-нульмерные границы всех доменов поместим в наименьший по номеру узел упорядоченного дерева зависимостей, в сыновьях которого имеются все узлы дерева, содержащие соответствующие родительские для этой квази-нульмерной границы >3 доменов. 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 квази-треугольной матрицей типа вложенных сечений, и ее структура разреженности будет иметь вид, подобный изображенному на Рисунке 7. Для уменьшения заполнения Q- и R- факторов QR разложений в столбцевых блоках узлов можно дополнительно использовать упорядочивание ND как было описано выше.

Заключение

В работе представлен параллельный алгоритм вычисления QR разложения многоуровневой разреженной верхней квази-треугольной 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], также представленной в качестве доклада на конференцию.

Литература

1. Тыртышников Е.Е. Методы численного анализа : учеб. пособие для студ. вузов / — М. : Издательский центр «Академия», 2007. — 320 с. — (Университетский учебник. Сер. Прикладная математика и информатика), ISBN 978-5-7695-3925-1.
2. George A., Liu J.W., "Computer Solution of Large Sparse Positive Definite Systems", Prentice Hall, 1981.
3. Kaporin I.E. "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, 483-509 (1998).
4. Davis T. A. 2011. Algorithm 915: SuiteSparseQR, a multifrontal multithreaded sparse QR factorization package. ACM Trans. Math. Softw. 38, 1 (2011), 8:1–8:22.
5. Yeralan S.N., Davis T.A., Ranka S. Algorithm 9xx: Sparse QR Factorization on the GPU. ACM Transactions on Mathematical Software, Vol. 1, No. 1, Article 1, Publication date: January 2015.
6. Харченко С.А., Ющенко А.А. Параллельная реализация алгоритма разреженного QR разложения для прямоугольных верхних квази-треугольных матриц со структурой типа вло-

женных сечений // Представлена в качестве доклада на первую объединенную международную конференцию "Суперкомпьютерные дни в России", Москва, 28-29 сентября 2015 г.

7. Haidar A., Dong T., Tomov S., Luszczek P., Dongarra J. Framework for Batched and GPU-resident Factorization Algorithms Applied to Block Householder Transformations.

Parallel algorithm for sparse QR decomposition of a rectangular upper quasi triangular matrix with ND-type sparsity

Sergey Kharchenko

Keywords: sparse rectangular matrix, upper quasi triangular matrix, nested dissection, volume partitioning, QR decomposition, Householder transformations, parallel algorithm, SLAE, least squares problem

The paper considers algorithm for computing sparse QR decomposition of a specially ordered rectangular matrix. Decomposition is based on block sparse Householder transformations. For ordering computations the ND-type ordering for sparsity of ATA matrix can be used, here A - original rectangular matrix. For mesh based problems the ordering can be constructed starting from appropriate volume partitioning of the computational mesh. Parallel computations are based on sparse QR decomposition for sets of rows with additional zero block at the beginning. The suggested algorithm is planned to be used as main computational kernel in the developed by the author parallel iterative algorithms for solving SLAEs and least squares problems. The corresponding algorithms will be based on composition of the subspaces represented by sparse bases.