V. V. Voevodin, A. Yu. Kuznetsov, E. E. Tyrtyshnikov, A tribute to goodness, intelligence and talent, *Algebra i Analiz*, 1990, Volume 2, Issue 6, 10–33

В.В.Воеводин, А.Ю.Кузнецов, Е.Е.Тыртышников

ПОКЛОНЕНИЕ ДОБРОТЕ, УМУ И ТАЛАНТУ

Эта статья не является обзором трудов двух выдающихся людей нашего времени, лауреатов Государственной премии СССР Веры Николаевны и Дмитрия Константиновича Фаддеевых. Не является она и их жизнеописанием. Это всего лишь три не связанных между собой отблеска того яркого пламени, которое, теперь уже в далекое время, разожгли эти люди в вычислительных науках, особенно в вычислительной алгебре.

По-разному жизнь сводила нас с Верой Николаевной и Дмитрием Константиновичем. Формально мы не являемся их учениками. Но каждому Вера Николаевна и Дмитрий Константинович давали свое благословение при вступлении на путь самостоятельных научных исследований. Получить это благословение было совсем не просто. Они были нетерпимы к какой-либо фальши в науке, да и не только в науке. Предельная работоспособность, одержимость в достижении цели, высочайший профессионализм, энциклопедичность в знании предмета исследований, - это только некоторые из требований, которые они предъявляли как к себе, так и ко всем тем, кто работал или хотел работать с ними.

И еще одна характерная деталь. Вера Николаевна и Дмитрий Константинович никогда подолгу не обсуждали планы своей будущей деятельности. Они просто начинали их осуществлять. И мы, тогда совсем еще молодые люди, нередко уступали им в инициативе, причем на самых современных направлениях. Так было, например, с модными теперь параллельными вычислениями. Именно Вера Николаевна и Дмитрий Константинович стали в нашей стране пионерами в этом современном направлении научных исследований. „Меньше говорить, больше делать" - этот их принцип с предельной четкостью описал В.В.Вересаев в своем коротком, но ярком рассказе „Мимоходом", посвященном Вере Николаевне.

Вокруг них была атмосфера удивительной доброжелательности. И это несмотря на то, что Вера Николаевна могла выразить свое несогласие с чем-либо в довольно резкой форме. Присутствие же Дмитрия Константиновича всегда направляло дискуссию в спокойное русло. Дух требовательности и доброты создавал особенно благоприятную почву для воспитания молодежи. Нам повезло, что в период своего научного становления удалось соприкоснуться с этим духом.

Ключевые слова: вычислительные методы линейной алгебры, прямые методы решения линейных систем, матричные итерационные методы в подпространствах, параллельные вычисления.

В смутное сегодняшнее время, когда нередко без разбору разрушается прошлое и перечеркиваются идеалы, очень нужна точка опоры. Многим из тех, кому посчастливилось знать Веру Николаевну и Дмитрия Константиновича, память об этих людях помогает не потерять себя сегодня.

§ 1. Вычислительные методы линейной алгебры

В области вычислительной линейной алгебры имена Веры Николаевны и Дмитрия Константиновича Фаддеевых имеют мировую известность и воспринимаются специалистами как имена классиков.

Прежде всего они, как и „положено" классикам, находились у истоков этой науки и в период интенсивного ее формирования создали уникальный в то время обобщающий труд, вобравший в себя, по-видимому, все сколько-нибудь ценные факты, идеи и алгоритмы тогда еще молодой вычислительной науки, вызванной к жизни бурным развитием компьютеров. Еще в 1950 г. В.Н.Фаддеева выпустила в свет книгу, называвшуюся „Вычислительные методы линейной алгебры", а большой совместный труд с тем же названием, ставший настольной книгой алгебраистов-вычислителей того времени, появился (в первом издании) в 1960 г. Теперь эта книга может служить надежной точкой отсчета, обращаясь к которой можно судить, как сильно и чем именно обогатилась эта область науки за прошедшие 30 лет. Благодаря энциклопедическому охвату материала книга Фаддеевых до сих пор сохраняет значение как основательный справочник, в котором можно найти факты и алгоритмы, упоминаемые вскользь или отсутствующие в современных книгах. О том, насколько важно иметь книгу такого рода, свидетельствует большое и продолжающее расти число цитирующих ее авторов как в нашей стране, так и за рубежом. В области вычислительных методов наследие Фаддеевых, конечно, не исчерпывается названной книгой. Помимо ряда оригинальных научных работ, им принадлежат исключительно полезные обзоры [4,5]. Под руководством В.Н.Фаддеевой была выполнена поистине титаническая работа по созданию серии библиографических указателей по вычислительным методам линейной алгебры и ее программному обеспечению [8], собрана замечательная коллекция тестовых матриц [7].

В этом разделе мы представим некоторые результаты по вычислительной линейной алгебре, появившиеся после выхода в свет книги Фаддеевых. Между собой они почти не связаны, но тем не менее есть нечто их объединяющее - каждый из них по-своему показывает, каким образом продолжает жить наследие Фаддеевых, обеспечивая почву новым исследованиям.

1.1. „Принципиально новые" алгоритмы. „Принципиально новые" алгоритмы нередко оказываются реминисценцией методов давно известных, но по тем или иным причинам отодвинутых на второй план. Яркой иллюстрацией к сказанному является потенциально самый быстрый алгоритм решения систем линейных алгебраических уравнений с невырожденной матрицей коэффициентов в условиях использования неограниченного числа процессоров. Пусть система имеет вид

$$Ax = b, \qquad (1.1.1)$$

где A - невырожденная матрица порядка n. Классический подход заключается в выполнении некоторого левостороннего преобразования N матрицы A, превращающего ее в верхнюю треугольную матрицу U:

$$NA = U. \qquad (1.1.2)$$

Предположим, что все ведущие миноры в A отличны от нуля. Тогда к (1.1.2) можно прийти, используя метод исключения Гаусса, без выбора ведущего элемента:

$$a_{ij}^{(0)} = a_{ij}, \quad 1 \le i, j \le n;$$

$k=1,\dots,n-1$:

$$a_{ij}^{(k)} = a_{ij}^{(k-1)} - (a_{ik}^{(k-1)}/a_{kk}^{(k-1)})a_{kj}^{(k-1)}, \quad k+1 \le i, j \le n. \qquad (1.1.3)$$

Очевидно, $U=[a_{ij}^{(n-1)}]$.

Формулы (1.1.3) таковы, что при каждом k операции по вычислению $a_{ij}^{(k)}$ для всех i,j можно выполнять одновременно. Если считать, что любая операция выполняется за одну единицу времени (будем говорить, за один такт), то формулы (1.1.3) можно реализовать за $3n+O(1)$ тактов. В общем случае, когда некоторые из ведущих миноров могут быть и нулевыми, требуется выбор ведущего элемента. Независимо от того, осуществляется он по столбцу или же по всей „активной" подматрице, при параллельном счете один ведущий элемент можно выбрать, затратив $O(\log_2 n)$ тактов. Таким образом, все операции, связанные с выбором, довольно „дорогие" - они требуют $O(n\log_2 n)$ тактов.

Чтобы понять, можно ли получить универсальный алгоритм с числом тактов меньше, чем $O(n\log_2 n)$, попробуем рассмотреть еще один классический подход, в котором исключение элементов проводится с помощью плоских вращений (метод Гивенса):

$$a_{ij}^{(0)} = a_{ij}, \quad 1 \le i, j \le n;$$

$k = 1,\dots,n-1$:

$i = k + 1,\dots,n$:

$$c_{ik} = \frac{a_{kk}^{(k-1)}}{\sqrt{|a_{kk}^{(k-1)}|^2 + |a_{ik}^{(k-1)}|^2}}, \quad s_{ik} = \frac{a_{ik}^{(k-1)}}{\sqrt{|a_{kk}^{(k-1)}|^2 + |a_{ik}^{(k-1)}|^2}}, \qquad (1.1.4)$$

$$a_{ij}^{(k)} = c_{ik}a_{ij}^{(k-1)} - s_{ik}a_{kj}^{(k-1)},$$

$$a_{kj}^{(k)} = \bar{s}_{ik}a_{ij}^{(k-1)} + \bar{c}_{ik}a_{kj}^{(k-1)}, \quad k + 1 \le j \le n.$$

В итоге мы получаем (1.1,2), где $U=[a_{ij}^{(n-1)}]$, а матрица N теперь является произведением матриц плоских вращений.

На первый взгляд, с точки зрения распараллеливания алгоритм (1.1.4) хуже метода Гаусса, потому что в отличие от последнего теперь мы не можем исключать элементы одного столбца одновременно. В действительности же, делая такой вывод, мы рассматриваем изолированно каждый фрагмент, связанный с исключением элементов одного столбца. В данном случае необходимо рассмотреть алгоритм в целом – и тогда обнаружится, что возможность одновременного исключения элементов разных столбцов здесь имеется. Мы можем исключать одновременно элементы в позициях, указанных в каждой из нижеследующих строк:

$$
\begin{array}{llll}
(2,1); & & & \\
(3,1); & & & \\
(4,1), & (3,2); & & \\
(5,1), & (4,2); & & \\
(6,1), & (5,2), & (4,3); & \\
(7,1), & (6,2), & (5,3); & \\
\ldots\ldots\ldots & & & \\
(n,1), & (n-1,2),\ldots\ldots; & & \\
 & (n,2),\ldots\ldots\ldots; & & \\
 & & & (n,n-1).
\end{array}
\qquad (1.1.5)
$$

Легко подсчитать, что общее число строк здесь равно $2n-3$. Следовательно, для параллельной реализации метода вращений достаточно $O(n)$ тактов. Обратим внимание на то, что порядок выполнения операций изменен таким образом, что в действительности все операции будут выполняться в параллельном алгоритме с теми же самыми данными, что и в исходном алгоритме. Поэтому и реально вычисленные (с учетом округлений) величины в обоих алгоритмах будут одинаковыми. „Новый" параллельный алгоритм Гивенса оказался математически эквивалентным исходному алгоритму.

Итак, у нас есть универсальный алгоритм с числом тактов $O(n)$. В то же время очевидная (и до сих пор не улучшенная) нижняя оценка для числа тактов имеет вид $\mathrm{const}\cdot\log_2 n$. Просматривая алгоритмы, обычно приводимые в современных книгах по вычислительной линейной алгебре, мы обнаружим, что линейная зависимость от n для числа тактов – это лучшее, что удается из них „выжать". Но в книге Фаддеевых мы можем найти то, что нужно! Раскроем ее на с. 311 (2-е издание); это – § 47 с заголовком „Метод Леверье и видоизменение Д. К. Фаддеева". Пусть характеристический многочлен матрицы A имеет вид

$$\det(A-tI) = (-1)^n(t^n-p_1t^{n-1}-p_2t^{n-2} - \ldots - p_n). \qquad (1.1.6)$$

Тогда по теореме Кэли-Гамильтона

$$A^{-1} = \frac{1}{p_n}\,(A^{n-1} - p_1A^{n-2} - p_2A^{n-3} - \ldots - p_{n-1}I). \qquad (1.1.7)$$

Матрицы $A, A^2,\ldots,A^{n-1}$ можно вычислить за $O(\log_2^2 n)$ тактов, используя схему сдваивания [5]. Если известны коэффициенты $p_1,\ldots,p_n$, то A^{-1} можно найти по формуле (1.1.7), вновь используя схему сдваивания, т.е. за $O(\log_2 n)$ тактов.

Чтобы вычислить коэффициенты характеристического многочлена, предлагается использовать формулы Ньютона:

$$kp_k = s_k - p_1s_{k-1} - \ldots - p_{k-1}s_1, \quad 1 \le k \le n, \qquad (1.1.8)$$

$$s_k = \operatorname{sp} A^k, \quad 1 \le k \le n. \tag{1.1.9}$$

Следы матриц A^k, очевидно, вычисляются по схеме сдваивания - за $O(\log_2 n)$ тактов. Далее, соотношения (1.1.8) в матрично-векторной записи имеют вид

$$\begin{bmatrix} 1 & & & & \\ s_1 & 2 & & & \\ s_2 & s_1 & 3 & & \\ \cdots & \cdots & \cdots & & \\ s_{n-1} & s_{n-2} & s_{n-3} & \cdots & s_1 n \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ \cdots \\ p_n \end{bmatrix} = \begin{bmatrix} s_1 \\ s_2 \\ s_3 \\ \cdots \\ s_n \end{bmatrix} \tag{1.1.10}$$

Следовательно, остается найти способ решения треугольной линейной системы, требующей $O(\log_2^2 n)$ тактов, тогда мы получим универсальный алгоритм для систем общего вида, дающий решение за $O(\log_2^2 n)$ тактов. Параллельное решение треугольной системы требует, конечно, некоторой изобретательности, но эта задача, тем не менее является относительно простой; в обзоре [5] можно найти даже несколько различных алгоритмов с числом тактов $O(\log_2^2 n)$.

Описанный выше подход к параллельному решению линейных алгебраических систем предложен Ксанки [5]. По-видимому, он лишен какого-либо практического интереса ввиду почти очевидной численной неустойчивости процесса. В то же время это очень важный результат, так как он показывает потенциальные возможности распараллеливания. Пока еще открытым остается вопрос: можно ли получить численно устойчивый алгоритм с числом тактов $O(\log_2^\gamma n)$? До сих пор мы ничего не знаем о существовании алгоритма с такого же вида оценкой, в которой $\gamma<2$.

1.2. Быстрые реализации метода окаймления. Обычно о методе окаймления говорят в контексте обращения последовательности ведущих подматриц некоторой матрицы. Однако он представляет интерес и с точки зрения решения линейных алгебраических систем. В книге Фаддеевых этим вопросам посвящен § 25, с. 203.

Пусть исходная система имеет вид

$$Az = b,$$
$$A = [a_{ij}]_{i\,j=0}^{n-1}, \quad z = [z_0 \ldots z_{n-1}]^T, \quad b = [b_0 \ldots b_{n-1}]^T. \tag{1.2.1}$$

В методе окаймления эта система рассматривается как одна (последняя) из семейства усеченных систем

$$A_k z^{(k)} = b^{(k)},$$
$$A_k = [a_{ij}]_{i\,j=0}^{k}, \quad z^{(k)} = [z_0^{(k)} \ldots z_k^{(k)}]^T, \quad b^{(k)} = [b_0 \ldots b_k]^T. \tag{1.2.2}$$

Легко проверить, что вектор

$$z^{(k)} - \begin{bmatrix} z^{(k-1)} \\ \\ 0 \end{bmatrix}$$

пропорционален последнему столбцу в A_k^{-1}; соответствующий коэффициент пропорциональности легко вычисляется, причем для этого нужно знать лишь вектор $z^{(k-1)}$.

Более точно, предположим, что для некоторого $k\ge1$ матрица A невырожденная.

Обозначим $y^{(k)}$ последний столбец в A_k^{-1}. Тогда если $(k-1)$-я усеченная система совместна и $z^{(k-1)}$ есть ее решение, то

$$z^{(k)} = \begin{bmatrix} z^{(k-1)} \\ 0 \end{bmatrix} + y^{(k)}\alpha_k, \tag{1.2.3}$$

$$\alpha_k = b_k - \sum_{j=0}^{k-1} a_{kj} z_j^{(k-1)}. \tag{1.2.4}$$

Для записи алгоритмов, построенных на базе метода окаймления, удобно использовать следующие операторы $\underline{\uparrow}$ и $\overline{\downarrow}$, переводящие m-вектор $z=[z_0 \dots z_{m-1}]^T$ в m+1-векторы:

$$\underline{\uparrow}\, z \equiv [z_0 \dots z_{m-1} 0]^T, \tag{1.2.5}$$

$$\overline{\downarrow}\, z \equiv [0 z_0 \dots z_{m-1}]^T. \tag{1.2.6}$$

Тогда (1.2.3) принимает вид

$$z^{(k)} = \underline{\uparrow} z^{(k-1)} + y^{(k)}\alpha_k.$$

Соотношения (1.2.3) и (1.2.4) по существу можно найти в книге Фаддеевых (§ 25, с. 203). В ряде работ [9,17,20] показано, каким образом они позволяют получать весьма экономичные алгоритмы, учитывающие специфику матрицы A. Ясно, что реализация соотношений (1.2.3) и (1.2.4) требует выполнения $O(k)$ арифметических операций. При изменении k от 0 до $n-1$ это составит $O(n^2)$ операций. Следовательно, для того чтобы построить $O(n^2)$-алгоритм решения линейной алгебраической системы (1.2.1), остается найти ускоренный способ рекуррентного вычисления векторов $y^{(k)}$.

Вообще говоря, мы можем получить нужные нам векторы $y^{(k)}$, рекуррентно вычисляя матрицы $A_0^{-1}, A_1^{-1}, \dots, A_{n-1}^{-1}$. Однако во многих случаях, особенно для матриц специального вида, есть смысл отказаться от вычисления всех элементов обратных матриц, а вместо этого перейти к получению их „в операторном виде". Последнее означает, что имеются некоторые представления обратных матриц, допускающие быстрое умножение их на вектор, а задача обращения матрицы понимается как задача вычисления входящих в эти представления параметров.

Предположим, что матрица A теплицева, т. е.

$$a_{ij} = a_{i-j}, \quad 0 \le i, j \le n-1. \tag{1.2.7}$$

Чтобы организовать рекуррентное вычисление векторов $y^{(k)}$, в данном случае мы можем прибегнуть к конструкции того же метода окаймления (см. [9,17]). Системы

$$A_k y^{(k)} = [0 \dots 01]^T \tag{1.2.8}$$

нельзя считать усеченными; однако от них можно перейти к системам

$$(JA_kJ)(Jy^{(k)}) = [10 \dots 0]^T, \tag{1.2.9}$$

которые являются усеченными для системы

$$(JAJ)(Jy) = [10 \dots 0]^T. \tag{1.2.10}$$

Это означает, что матрицы JA_kJ суть ведущие подматрицы в матрице JAJ. Здесь

$$J = \begin{bmatrix} 0 & & 1 \\ & \cdots & \\ 1 & & 0 \end{bmatrix}, \qquad (1.2.11)$$

причем порядок определяется в зависимости от контекста.

Заметим, что последний столбец матрицы $(JA_kJ)^{-1}$ можно записать в виде $Jx^{(k)}$, где $x^{(k)}$ - первый столбец в A_k^{-1}. Поэтому, согласно (1.2.3), (1.2.4) для соответствующего β_k, находим

$$Jy^{(k)} = \uparrow Jy^{(k-1)} + Jx^{(k)}\beta_k$$

или, после применения J к обеим частям,

$$y^{(k)} = \downarrow y^{(k-1)} + x^{(k)}\beta_k. \qquad (1.2.12)$$

Поскольку вектор x^k удовлетворяет усеченной системе

$$A_kx^{(k)} = [10\ldots0]^T,$$

для соответствующего α_k имеем

$$x^{(k)} = \uparrow x^{(k-1)} + y^{(k)}\alpha_k. \qquad (1.2.13)$$

Объединяя (1.2.12) и (1.2.13), получаем

$$x^{(k)}(1-\alpha_k\beta_k) = \uparrow x^{(k-1)} + \downarrow y^{(k-1)}\alpha_k,$$

$$y^{(k)}(1-\beta_k\alpha_k) = \uparrow x^{(k-1)}\beta_k + \downarrow y^{(k-1)}, \qquad (1.2.14)$$

$$\alpha_k = -\sum_{j=0}^{k-1} a_{k-j}x_j^{(k-1)}, \qquad \beta_k = -\sum_{j=0}^{k-1} a_{-1-j}y_j^{(k-1)}.$$

В силу теплицевости подматрица, дополнительная к A_0 в A_k, есть не что иное, как A_{k-1}; поэтому $x_0^{(k-1)}$ и $y_{k-1}^{(k-1)}$ отличны от нуля, если при всех k подматрицы A_k невырожденные. Тогда $1-\alpha_k\beta_k\neq0$ при $1\le k\le n-1$.

Чтобы сократить число операций, целесообразно ввести нормировку вычисляемых векторов (см. [17,20]). Будем считать, что первый и последний столбцы в A_k^{-1} имеют вид $x^{(k)}p_k$ $y^{(k)}q_k$, где p_k, q_k - ненулевые нормировочные коэффициенты. В этих обозначениях алгоритм вычисления параметров для определения A^{-1} можно сформулировать следующим образом:

$$k=0:\ p_0= q_0=a_0^{-1},\ x^{(0)}=y^{(0)}=I;$$

$k = 1,\ldots, n-1$:

$$F_k = \sum_{j=0}^{k-1} a_{k-j}x_j^{(k-1)},\ G_k = \sum_{j=0}^{k-1} a_{-1-j}\ y_j^{(k-1)},$$

$$s_k = -q_{k-1}F_k,\ t_k = -p_{k-1}G_k, \qquad (1.2.15)$$

$$p_k = (I - t_ks_k)^{-1}p_{k-1},\ q_k = (I - s_kt_k)^{-1}q_{k-1},$$

$$x^{(k)} = \uparrow x^{(k-1)} + \downarrow y^{(k-1)}s_k,\ y^{(k)} = \uparrow x^{(k-1)}t_k + \downarrow y^{(k-1)}.$$

Алгоритм (1.2.15) является некоторой модификацией хорошо известного алгоритма Левинсона-Дурбина [13]; он требует выполнения (в главном члене) $2n^2$ умножений и

$2n^2$ сложений-вычитаний. Однако он записан таким образом, что без изменений может быть использован в случае блочно-теплицевой матрицы. Пусть a_{i-j} - блоки порядка p. В этом случае p_k, q_k суть невырожденные блоки порядка p; число умножений будет равно $2p^3n^2$, число сложений-вычитаний такое же.

В работах [17,18] показано, каким образом алгоритм (1.2.15) можно преобразовать в алгоритм, требующий лишь $(8/3)n\log_2^2 n$ умножений и $8n\log_2^2 n$ сложений-вычитаний; это делается с использованием разработанной в [17] теории векторизации.

Быстрые реализации метода окаймления удается получить для различных матриц специального вида (см. [17]). При этом исключительно важны как способ представления исходной матрицы, так и способ представления обратной матрицы. Например, матрицу A можно задавать ее теплицевым разложением

$$A - ZAZ^T = \sum_{i=1}^{t} u_i v_i^T, \tag{1.2.16}$$

где u, v суть вектор-столбцы размерности n;

$$Z = \begin{bmatrix} 0 & & & 0 \\ I & 0 & & \\ & I & 0 & \\ 0 & & \ddots \; I & 0 \end{bmatrix}_{n\times n} \tag{1.2.17}$$

Пусть компоненты матрицы Z и векторов u, v являются блоками порядка p. В этом случае метод окаймления приводит нас к алгоритму, в котором $\frac{1}{2}(5tp^3+2p^2)n^2$ умножений и столько же сложений-вычитаний. Заметим, что в общем случае можно утверждать, что $t\le n$. Однако часто бывает известно, что t много меньше, чем n; более того, часто само теплицево разложение возникает непосредственно, т.е. без дополнительных вычислений. Нередко естественным образом возникают такие теплицевы разложения, в которых

$$v_2 = u_2 = [I_2 0 \; \dots \; 0]^T. \tag{1.2.18}$$

Для них удается получить несколько модифицированный алгоритм, содержащий $\frac{1}{2}(5t-6)p^3n^2+p^2n^2$ умножений и столько же сложений [17].

Наряду с теплицевым разложением (1.2.16) могут представлять интерес и некоторые другие разложения. Предположим, что блочная матрица A удовлетворяет соотношению

$$\begin{bmatrix} x_0 & & 0 \\ & \ddots & \\ 0 & & x_{n-1} \end{bmatrix} A - A \begin{bmatrix} y_0 & & 0 \\ & \ddots & \\ 0 & & y_{n-1} \end{bmatrix} = \sum_{\ell=1}^{r} \begin{bmatrix} \varphi_{\ell 0} \\ \dots \\ \varphi_{\ell, n-1} \end{bmatrix} [\psi_{\ell 0} \dots \psi_{\ell, n-1}], \tag{1.2.19}$$

где x_i, y_j, $\varphi_{\ell i}$, $\psi_{\ell j}$ - некоторые заданные блоки. Будем считать, что никакие два из блоков $x_0,\dots,x_{n-1},\dots,y_0,\dots,y_{n-1}$ не имеют общих собственных значений. Тогда из теории матричных уравнений вытекает, что A из (1.2.19) определяется однозначно (см. [17]). Указанный способ задания матрицы будем называть разложением Коши.

Если разложение Коши невырожденной матрицы A является r-членным, то разложение Коши матрицы A^{-1} также будет r-членным. Действительно, после умножения обеих частей (1.2.19) слева и справа на A^{-1} получаем

$$A^{-1}\begin{bmatrix} x_0 & & 0 \\ & \ddots & \\ 0 & & x_{n-1}\end{bmatrix} - \begin{bmatrix} y_0 & & 0 \\ & \ddots & \\ 0 & & y_{n-1}\end{bmatrix}A^{-1} = \sum_{\ell=1}^{r}\begin{bmatrix} u_0 \\ \dots \\ u_{\ell,n-1}\end{bmatrix}[v_{\ell 0}\dots v_{\ell,n-1}], \qquad (1.2.20)$$

$$A\begin{bmatrix} u_0 \\ \dots \\ u_{\ell,n-1}\end{bmatrix} = \begin{bmatrix} \varphi_{\ell 0} \\ \dots \\ \varphi_{\ell,n-1}\end{bmatrix}, \quad \ell = 1,\dots,r; \qquad (1.2.21)$$

$$[v_{\ell 0}\dots v_{\ell,n-1}]A = [\psi_{\ell 0}\dots\psi_{\ell,n-1}], \quad \ell=1,\dots,r. \qquad (1.2.22)$$

Теперь предположим, что все блочные подматрицы A_k невырождены. Следуя методу окаймления, введем в рассмотрение также столбцы (блочные) $c^{(k)}$ и $s^{(k)}$, составленные соответственно из (блочных) компонент последних столбца и строки матрицы A_k^{-1}. Тогда [17]

$$u_\ell^{(k)} = \triangleq u_\ell^{(k-1)} + c^{(k)}\alpha_{\ell k},$$

$$\alpha_{\ell k} = \varphi_{\ell k} - \sum_{j=0}^{k-1} a_{kj}\, u_{\ell j}^{(k-1)}, \quad \ell = 1,\dots,r; \qquad (1.2.23)$$

$$v_\ell^{(k)} = \triangleq v_\ell^{(k-1)} + \beta_{\ell k}\, s^{(k)},$$

$$\beta_{\ell k} = \psi_{\ell k} - \sum_{j=0}^{k-1} v_{\ell j}^{(k-1)}\, a_{jk}, \quad \ell = 1,\dots,r. \qquad (1.2.24)$$

Принимая во внимание эти соотношения, равенства $a_{kk}^{(-1)}=c_k^{(k)}=s_k^{(k)}$ и (1.2.20), получаем

$$a_{kk}^{(-1)}x_k - y_k a_{kk}^{(-1)} = a_{kk}^{(-1)} f_k a_{kk}^{(-1)}, \qquad (1.2.25)$$

где

$$f_k = \sum_{\ell=1}^{r} \alpha_{\ell k}\beta_{\ell k}. \qquad (1.2.26)$$

В силу условий на собственные значения блоков x_k и y_k это матричное уравнение однозначно разрешимо относительно $(a_{kk}^{(-1)})^{-1}$.

Договоримся о следующем обозначении. Пусть задано матричное уравнение

$$xa - ay = f, \qquad (1.2.27)$$

где матрицы x и y не имеют общих собственных значений. Тогда a однозначно выражается через x, y, f, и мы будем писать

$$a = \mathrm{sol}(x,y,f). \qquad (1.2.28)$$

Таким образом, согласно (1.2.20),

$$c_k^{(k)} = s_k^{(k)} = (\mathrm{sol}(x_k, y_k, f_k))^{-1}. \tag{1.2.29}$$

Далее, при $0 \le i \le k-1$ находим

$$a_{ik}^{(-1)} x_k - y_i a_{ik}^{(-1)} = \sum_{\ell=1}^{r} u_{\ell i}^{(k)} \ v_{\ell k}^{(k)} =$$

$$= \sum_{\ell=1}^{r} u_{\ell i}^{(k-1)} \beta_{\ell k} c_k^{(k)} + a_{ik}^{(-1)} \sum_{\ell=1}^{r} \alpha_{\ell k} \beta_{\ell k} c_k^{(k)}.$$

Последнее слагаемое преобразуем с учетом (1.2.25), (1.2.26):

$$a_{ik}^{(-1)} (c_k^{(k)})^{-1} (c_k^{(k)} f_k c_k^{(k)}) = a_{ik}^{(-1)} (c_k^{(k)})^{-1} (c_k^{(k)} x_k - y_k c_k^{(k)}).$$

Следовательно,

$$a_{ik}^{(-1)} ((c_k^{(k)})^{-1} y_k c_k^{(k)}) - y_i a_{ik}^{(-1)} = \sum_{\ell=1}^{r} u_{\ell i}^{(k-1)} \beta_{\ell k} c_k^{(k)}.$$

В силу условий на собственные значения матриц y_k и y_i полученное матричное уравнение однозначно разрешимо:

$$a_{ik}^{(-1)} = \mathrm{sol}(-y_i, (c_k^{(k)})^{-1} y_k c_k^{(k)}, \ \sum_{\ell=1}^{r} u_{\ell i}^{(k-1)} \beta_{\ell k} c_k^{(k)}), \quad 0 \le i \le k-1. \tag{1.2.30}$$

Для блоков $a_{kj}^{(-1)}$ при $0 \le j \le k-1$ получаем такие соотношения:

$$a_{kj}^{(-1)} x_j - y_k a_{kj}^{(-1)} = \sum_{\ell=1}^{r} u_{\ell k}^{(k)} \ v_{\ell j}^{(k)} =$$

$$c_k^{(k)} \sum_{\ell=1}^{r} \alpha_{\ell k} v_{\ell j}^{(k-1)} + c_k^{(k)} \sum_{\ell=1}^{r} \alpha_{\ell k} \ \beta_{\ell k} a_{kj}^{(-1)}.$$

С учетом (1.2.24)-(1.2.26) выводим

$$a_{kj}^{(-1)} x_j - (c_k^{(k)} x_k (c_k^{(k)})^{-1}) \ a_{kj}^{(-1)} = c_k^{(k)} \sum_{\ell=1}^{r} \alpha_{\ell k} v_{\ell j}^{(k-1)},$$

откуда

$$a_{kj}^{(-1)} = \mathrm{sol}(-c_k^{(k)} x_k (c_k^{(k)})^{-1}, \ x_j, \ \sum_{\ell=1}^{r} \alpha_{\ell k} v_{\ell j}^{(k-1)}), \quad 0 \le j \le k-1.$$

Описанный алгоритм предложен в [17]. Он, очевидно, имеет более простой вид, если блоки x_i, y_j суть скалярные матрицы. При $p=1$ общее число умножений составляет $(4r+1)n^2$ (в главном члене); число сложений-вычитаний такое же.

1.3. Формулы Фробениуса и ленточные матрицы. Имеется очень много работ, посвященных ленточным матрицам и специфике обратных к ним матриц. На наш взгляд, наиболее ясный и простой путь к изучению обратных матриц связан с использованием подхода, предложенного Д.К.Фаддеевым [2].

Этот подход опирается на так называемые формулы Фробениуса (см. [3,10,20]).

Пусть матрицы A и A^{-1} представлены в блочном виде:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad A^{-1} = \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix}, \tag{1.3.1}$$

где A_{11} - квадратная невырожденная матрица. Тогда, действуя в духе метода Гаусса, имеем

$$\begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix} =$$

$$\begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22}-A_{21}A_{11}^{-1}A_{12} \end{bmatrix} \begin{bmatrix} \hat{A}_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ -A_{21}A_{11}^{-1} & I \end{bmatrix}.$$

Отсюда очевидным образом получаем

$$\begin{aligned} \hat{A}_{22} &= (A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1}, \quad \hat{A}_{21} = -\hat{A}_{22}A_{21}A_{11}^{-1}, \\ \hat{A}_{11} &= A_{11}^{-1}-A_{11}^{-1}A_{12}\hat{A}_{21}, \quad \hat{A}_{12} = -A_{11}^{-1}A_{12}\hat{A}_{22}, \end{aligned} \tag{1.3.2}$$

Это и есть формулы Фробениуса. Непосредственно из них вытекает важное матричное соотношение

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -A_{11}^{-1}A_{12} \\ I \end{bmatrix} [\hat{A}_{21}\hat{A}_{22}]. \tag{1.3.3}$$

Теперь рассмотрим матрицу $G=[g_{ij}]_{i\,j=1}^{n}$ и предположим, что для некоторых неотрицательных целых чисел ℓ и r, не превосходящих n, элемент g_{ij} может быть отличен от нуля только в том случае, когда $\ell-1 \ge i-j \ge r-1$. Число ℓ называется левой, а r - правой шириной ленты матрицы G. Число $\ell+r-1$ называют шириной ленты, а саму матрицу G при этом называют $\ell+r-1$-диагональной. При $r=2$ матрицу называют левой хессенберговой, а при $\ell=2$ - правой хессенберговой.

Предположим, что $g_{ij} \ne 0$ при $i-j=-r+1$. В этом случае запишем G и G^{-1} в блочном виде

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix}, \quad G^{-1} = \begin{bmatrix} \hat{G}_{11} & \hat{G}_{12} \\ \hat{G}_{21} & \hat{G}_{22} \end{bmatrix}, \tag{1.3.4}$$

где G_{12} и $\hat{G}_{21}$ - квадратные матрицы порядка $n-r+1$; G_{21} и $\hat{G}_{12}$, G_{11} и $\hat{G}_{22}$, G_{22} и $\hat{G}_{11}$ имеют размеры соответственно $(r-1)\times(r-1)$, $(n-r+1)\times(r-1)$, $(r-1)\times(n-r+1)$. Очевидно, квадратная матрица G_{12} треугольная и невырожденная. Поэтому мы имеем право записать формулы Фробениуса для матрицы

$$A = \begin{bmatrix} G_{12} & G_{11} \\ G_{22} & G_{21} \end{bmatrix}; \tag{1.3.5}$$

тогда

$$A^{-1} = \begin{bmatrix} \hat{G}_{21} & \hat{G}_{22} \\ \hat{G}_{11} & \hat{G}_{12} \end{bmatrix}. \qquad (1.3.6)$$

В силу (1.3.3)

$$G^{-1} = \begin{bmatrix} 0 & 0 \\ G_{12}^{-1} & 0 \end{bmatrix} + \begin{bmatrix} I \\ -G_{12}^{-1}\, G_{11} \end{bmatrix} [\hat{G}_{11}\hat{G}_{12}]. \qquad (1.3.7)$$

Матрица G_{12}^{-1} нижняя треугольная. Поэтому элементы $g_{ij}^{(-1)}$ матрицы G^{-1} при $i-j<r-1$ совпадают с соответствующими элементами некоторой матрицы порядка n ранга $r-1$. Более точную формулировку полученного дает следующая

Теорема. *Для того чтобы невырожденная матрица имела правую (левую) ширину ленты* $r(\ell)$ *и отличные от нуля элементы на диагонали* $i-j=-r+1$ $(i-j=\ell-1)$, *необходимо и достаточно, чтобы элементы обратной к ней матрицы при* $i-j<r-1$ $(i-j>-\ell+1)$ *совпадали с соответствующими элементами некоторой матрицы такого же порядка, имеющей ранг не выше* $r-1$ $(\ell-1)$.

Эта теорема легко обобщается на случай, когда g_{ij} суть квадратные блоки одного порядка для всех i, j. Условие $g_{ij}\neq 0$ в этом случае заменяется требованием невырожденности блока g_{ij}. Однако теорема может быть обобщена и на случай разбиения матрицы на прямоугольные блоки; такое обобщение получено в [16].

Пусть G составлена из блоков g_{ij}, и при этом g_{ij} имеет размер $k_i\times\ell_j$. Если G – квадратная матрица порядка N, то имеем

$$\sum_{i=1}^{n} k_i = \sum_{j=1}^{n} \ell_j = N.$$

Будем считать для определенности, что матрица G - левая блочно-хессенбергова, т.е. $g_{ij}=0$ при $i-j<-1$. Представим G и G^{-1} в блочном виде (1.3.4), где G_{12} и $\hat{G}_{21}^{T}$, G_{21} и $\hat{G}_{12}^{T}$, G_{11} и $\hat{G}_{12}^{T}$, G_{22} и $\hat{G}_{11}^{T}$ имеют соответственно размеры

$$\sum_{i=1}^{n-1} k_i \times \sum_{j=2}^{n} \ell_j, \quad k_n \times \ell_1,$$

$$\sum_{i=1}^{n-1} k_i \times \ell_1, \quad k_n \times \sum_{j=2}^{n} \ell_j.$$

В рассмотренной ситуации формулы Фробениуса „не работают". Однако соотношение типа (1.3.3) получить все-таки можно. Если матрица A_{11} прямоугольная, но имеет полный столбцовый ранг, то для некоторой матрицы $\tilde{A}_{11}$ имеем

$$\tilde{A}_{11}A_{11} = I. \qquad (1.3.4)$$

Легко проверить, что (1.3.3) остается в силе, если A_{11}^{-1} заменить на $\tilde{A}_{11}$. На этом пути получается следующий достаточно общий результат [16].

Пусть невырожденная матрица G и обратная к ней разбиты на блоки g_{ij} и $g_{ij}^{(-1)}$ соответственно размеров $k_i\times\ell_j$ и $\ell_i\times k_j$ $1\le i,j\le n$. Пусть G является левой (правой) блочно-хессенберговой и для наддиагональных (поддиагональных) блоков имеет место одно из двух - или ранг каждого блока равен числу его столбцов, или ранг каждого блока равен числу его строк. Тогда при $i-j\le 0$ $(i-j\ge 0)$ блоки $g_{ij}^{(-1)}$ совпадают с блоками аналогичного блочного разбиения матрицы такого же порядка и ранга не выше $\ell_1(\ell_n)$ в первом случае и $k_n(k_1)$ во втором случае.

§ 2. Матричные итерационные методы

Замечательная книга Веры Николаевны Фаддеевой [6], пожалуй, впервые акцентировала внимание на вычислительной линейной алгебре как самостоятельной области вычислительной математики, богатой своими идеями и методами.

Одной из таких идей было особое внимание к теории норм и анализу в целом в связи с конечномерными пространствами и операторами. Эта идея была развита и дополнена новыми результатами в фундаментальной монографии [3]. В данном параграфе очень кратко мы проиллюстрируем использование и дальнейшее развитие этих идей применительно к теории итерационных методов решения линейных систем, ограничиваясь для простоты вещественным случаем.

2.1. Сходимость стационарных итерационных методов. Применим для решения совместной системы линейных алгебраических уравнений

$$Au = f \qquad (2.1.1)$$

с матрицей A порядка n и вектором $f \in \operatorname{im} A$ стационарный итерационный метод

$$u^k = u^{k-1} - H(Au^{k-1} - f), \quad k = 1,2,\dots, \qquad (2.1.2)$$

где H - некоторая заданная матрица порядка n. Этот итерационный метод называется сходящимся, если для любого начального приближения $u^\circ \in \mathbb{R}^n$ последовательность векторов u^k сходится к некоторому решению u^* системы (2.1.1).

В случае, когда матрица A неособенная, ответ на вопрос о необходимом и достаточном условии сходимости дает следующее утверждение [3]:

метод (2.1.2) *сходится тогда и только тогда, когда*

$$\rho(T) < 1, \qquad (2.1.3)$$

где $\rho(T)$ *- спектральный радиус матрицы* $T=I-HA$ *и* I *- единичная матрица.*

Из этого утверждения, в частности, следует, что матрица H из (2.1.2) должна быть неособенной.

В случае произвольной матрицы A утверждение о необходимом и достаточном условии сходимости стационарного итерационного метода имеет другую формулировку [14] -

метод (2.1.2) *сходится тогда и только тогда, когда одновременно выполняются следующие три условия*:

1) $\ker H \cap \operatorname{im} A = \{\theta\}$;

2) $r(T) \equiv \max\limits_{\lambda\in Sp(T)} |\lambda| \le 1$;

3) *кратность собственного числа* $\lambda=1$ *матрицы* T *равна размерности* $\ker A$.

Здесь θ - нулевой вектор и через $\mathrm{Sp}(T)$ обозначается множество собственных чисел матрицы T.

Последнее утверждение имеет другую формулировку, более компактную и понятную с точки зрения теории линейных операторов в конечномерных пространствах [11] -

метод (2.1.2) *сходится тогда и только тогда, когда одновременно выполняются следующие два условия*:

1) *матрица* $P = \lim\limits_{k\to\infty} T^k$ *существует и является проектором*;

2) $\mathrm{im}\ P = \ker A$.

Легко проверить, что в случае выполнения этих двух условий, пространство $\mathbb{R}^n$ является прямой суммой подпространств $\ker A$ и $\mathrm{im}\ HA$. Более того, матрицы

$$P^* = \lim_{k\to\infty} (I-A^TH^T)^k,$$

$$P_1 = \lim_{k\to\infty} (I-AH)^k, \qquad (2.1.4)$$

$$P_1^* = \lim_{k\to\infty} (I-H^TA^T)^k,$$

также будут существовать и также являться проекторами. В частности,

$$\ker P^* = \mathrm{im}\ A^T,$$

$$\ker P_1 = \mathrm{im}\ A, \qquad (2.1.5)$$

$$\mathrm{im}\ P_1^* = \ker A^T,$$

и, следовательно, имеют место ортогональные прямые суммы

$$\mathbb{R}^n = \mathrm{im}\ P \dot{\oplus} \ker P^*,$$

$$\mathbb{R}^n = \mathrm{im}\ P_1^* \dot{\oplus} \ker P_1. \qquad (2.1.6)$$

Важнейшим вопросом теории сходимости стационарных итерационных методов решения систем с особенными матрицами становится проверка сформулированных выше условий для конкретных матриц. Для этих целей весьма удобным аппаратом является следующий вариационный принцип [14]:

пусть в подпространстве $\mathrm{im}\ A$ *задана некоторая норма такая, что для любого вектора* $w\in\mathbb{R}^n$, $w \bar{\in} \ker A$ выполняется неравенство

$$\|AT_w\|_* < \|Aw\|_*. \qquad (2.1.7)$$

Тогда метод (2.1.2) сходится.

Таким образом, чтобы установить сходимость конкретного стационарного итерационного метода, нам достаточно найти подходящую норму в подпространстве $\mathrm{im}\ A$ и установить выполнение неравенства (2.1.7). Можно показать, что существование такой нормы является также необходимым условием сходимости любого стационарного итерационного метода. Из сказанного фактически следует, что необходимым и достаточным условием сходимости стационарного итерационного метода, т.е. сходимости для любого начального приближения $u^\circ\in\mathbb{R}^n$, является существование в подпространстве $\mathrm{im}\ A$ такой нормы $\|\cdot\|_*$, что

$$\|I - AH\|_* < 1. \qquad (2.1.8)$$

Подробнее все эти вопросы, включая структуру предельного вектора $u^\infty = \lim_{k\to\infty} u^k$ в зависимости от матрицы HA, изучаются в [11,14].

Мы рассмотрим только один практически важный пример использования сформулированного вариационного принципа - достаточного условия сходимости стационарного итерационного метода. Пусть A - симметричная положительно полуопределенная матрица и имеет место представление

$$A = \Lambda - F - F^T, \tag{2.1.9}$$

где Λ - симметричная положительно определенная матрица. Выберем

$$H \equiv H_\omega = (\tfrac{1}{\omega}\Lambda - F)^{-1}, \tag{2.1.10}$$

где ω - вещественный параметр. Соответствующий метод (2.1.2) называется методом последовательной верхней релаксации и широко применяется при решении систем сеточных уравнений [3].

Определим в подпространстве $\mathrm{im}\, A$ норму $\|\cdot\|_*$ с помощью соотношения

$$\|\xi\|_* = (A^+\xi,\xi)^{1/2}, \quad \xi \in \mathrm{im}\, A, \tag{2.1.11}$$

где A^+ - обобщенная обратная матрица. С помощью несложных преобразований получим

$$\|(I-AH)\xi\|_*^2 = \|\xi\|_*^2 - 2(H\xi,\xi) + (AH\xi,H\xi) = \|\xi\|_*^2 - (\tfrac{2}{\omega} - 1)(\Lambda H\xi,H\xi),$$

где ξ - произвольный ненулевой вектор из $\mathrm{im}\, A$. Отсюда сразу следует, что для любого вектора $w\in\mathbb{R}^n$, $w\bar{\in}\ker A$, т.е. для любого ненулевого вектора $\xi=Aw$ и для любого значения $\omega\in(0;2)$ выполняется строгое неравенство

$$\|A(I-HA)w\|_* < \|Aw\|_*. \tag{2.1.12}$$

Таким образом, достаточное условие выполнено и, следовательно, метод последовательной верхней релаксации сходится для систем с положительно полуопределенными матрицами для любого значения параметра ω из интервала $(0;2)$.

2.2. Методы в подпространствах. Итерационные методы решения систем с особенными матрицами являются простейшим случаем итерационных методов в подпространствах. В последние годы интерес к итерационным методам в подпространствах сильно возрос в связи с развитием новых подходов к решению сеточных систем, в первую очередь методов фиктивных компонент и методов разбиения областей [15].

Пусть заданы некоторая матрица H, удовлетворяющая условию $\ker H\cap \mathrm{im} A=\{\theta\}$, и некоторое подпространство U, инвариантное относительно матрицы AH. Тогда при условии, что начальные векторы невязок $\xi^\circ=Au^\circ-f$ метода (2.1.2) принадлежат U, этот итерационный метод может рассматриваться в данном подпространстве, так как все последующие векторы невязок $\xi^k=Au^k-f$ также будут принадлежать U. Более того, поскольку для любого набора чисел $\gamma_1,\dots,\gamma_s$ и любых $\xi\in U$ векторы

$$\eta = \xi - \sum_{i=1}^{s} \gamma_i (AH)^i \xi \tag{2.2.1}$$

также принадлежат U, то методы типа обобщенных методов сопряженных градиентов и минимальных итераций, а также чебышевские итерационные методы могут рассматриваться в этом же подпространстве.

Для начала мы рассмотрим блочный метод Гаусса-Зейделя как итерационный метод в подпространстве. Представим матрицу A системы (2.1.1) в виде блочной 2×2-матрицы

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{2.2.2}$$

с квадратными неособенными диагональными подматрицами A_{11} и A_{22} порядка n_1 и n_2 соответственно. Выберем

$$H = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}^{-1}. \tag{2.2.3}$$

Тогда легко видеть, что

$$(I - AH)\xi = -\begin{bmatrix} 0 & A_{12} \\ 0 & 0 \end{bmatrix} H\xi, \tag{2.2.4}$$

т.е. для любого $\xi \in \mathbb{R}^n$ вектор $(I-AH)\xi$ принадлежит подпространству

$$U = \operatorname{im} \begin{bmatrix} 0 & A_{12} \\ 0 & 0 \end{bmatrix}. \tag{2.2.5}$$

Поэтому если вектор ξ^0 принадлежит этому подпространству, то все последующие векторы невязок ξ^k метода (2.1.2) и других методов, использующих полиномиальные процедуры типа (2.2.1), также будут принадлежать этому подпространству.

Очевидно, что подпространство из (2.2.5) не единственное, в котором мы можем рассматривать итерационный метод (2.1.2) с матрицей H из (2.2.3). Например, мы можем определить подпространство

$$\overset{*}{U} = \{\xi : \xi = \begin{bmatrix} \xi_1 \\ 0 \end{bmatrix}, \ \xi \in \mathbb{R}^n, \ \xi_1 \in \mathbb{R}^{n_1}\}, \tag{2.2.6}$$

которое также инвариантно относительно матрицы AH и которое содержит в себе подпространство (2.2.5). С другой стороны, в случае, когда матрица A особенная и вектор $\xi \in \operatorname{im} A$, то вектор

$$\eta = \xi - AH\xi \tag{2.2.7}$$

принадлежит одновременно подпространству $\operatorname{im} A$ и подпространству из (2.2.5). Поэтому подпространство

$$\overset{**}{U} = \operatorname{im} A \cap \operatorname{im} \begin{bmatrix} 0 & A_{12} \\ 0 & 0 \end{bmatrix} \tag{2.2.8}$$

одновременно будет инвариантным относительно матрицы AH и содержаться в подпространстве U из (2.2.5). Выбор того или иного подпространства зависит от теоретических или практических целей, которые преследует исследователь, рассматривая данный вариант блочного метода Гаусса-Зейделя.

Важное обобщение метода Гаусса-Зейделя достигается выбором

$$H = \begin{bmatrix} B_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}^{-1}, \qquad (2.2.9)$$

где B_{11} - некоторая неособенная матрица порядка n_1. Именно такие матрицы H возникают в методах декомпозиции области с альтернирующими краевыми условиями Неймана-Дирихле [15]. Остановимся на этих методах подробнее, предполагая далее для простоты рассуждений, что исходная матрица A и матрица B_{11} симметричны и положительно определены.

Легко проверить, что в рассматриваемом случае подпространство $\overset{*}{U}$ из (2.2.6) инвариантно относительно матрицы AH. Иначе говоря, для векторов невязок метод (2.1.2) может быть переписан в виде

$$\begin{bmatrix} \xi_1^k \\ 0 \end{bmatrix} = \begin{bmatrix} \xi_1^{k-1} \\ 0 \end{bmatrix} - AH \begin{bmatrix} \xi_1^{k-1} \\ 0 \end{bmatrix} \qquad (2.2.10)$$

или, эквивалентно,

$$\xi_1^k = \xi_1^{k-1} - S_{11}\, B_{11}^{-1}\, \xi_1^{k-1}, \quad k = 1,2,\ldots, \qquad (2.2.11)$$

где матрица

$$S_{11} = A_{11} - A_{12}A_{22}^{-1}A_{21} \qquad (2.2.12)$$

называется дополнением Шура матрицы A. Таким образом, рассматриваемый метод в подпространстве $\overset{*}{U}$ может быть представлен как итерационный метод в пространстве $\mathbb{R}^{n_1}$:

$$u_1^k = u_1^{k-1} - B_{11}^{-1}(S_{11}u_1^{k-1} - g_1), \quad k = 1,2,\ldots, \qquad (2.2.13)$$

применяемый для решения редуцированной системы

$$S_{11}u_1 = g_1, \qquad (2.2.14)$$

где $g_1=f_1-A_{12}A_{22}^{-1}f_2$. Здесь $f=\begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$, $f_1\in\mathbb{R}^{n_1}$, $f_2\in\mathbb{R}^{n_2}$.

Так как матрицы B_{11} и S_{11} симметричны и положительно определены, то метод (2.2.13) может рассматриваться как базовая итерационная процедура для обобщенного метода сопряженных градиентов или чебышевских итерационных методов. Отсюда следует вывод, что рассмотрение метода (2.1.2) с несиметричной матрицей H из (2.2.9), если его рассматривать в подпространстве $\overset{*}{U}$, приводит к классическому симметризуемому итерационному методу (2.2.13). Иначе говоря, метод (2.1.2), (2.2.9) является симметризуемым по отношению к подпространству $\overset{*}{U}$.

Нетрудно убедиться, что точно такими же свойствами по отношению к подпространству $\overset{*}{U}$ (в смысле сведения к системе (2.2.14)) обладают итерационные методы (2.1.2) с матрицами

$$H = \begin{bmatrix} B_{11}+A_{12}A_{22}A_{21} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1}. \qquad (2.2.15)$$

Эти матрицы уже являются симметричными и положительно определенными. Такие матрицы H явились основой для развития многоуровневых методов декомпозиции области.

В определенном смысле двойственными по отношению к рассмотренным обобщениям блочного метода Гаусса-Зейделя являются методы фиктивных компонент [15]. Основная идея этого подхода заключается в следующем. Исходная система

$$B_{11}V_1 = g_1 \qquad (2.2.16)$$

с симметрично положительно определенной матрицей B_{11} порядка n_1 заменяется, например, системой

$$\begin{bmatrix} B_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} g_1 \\ 0 \end{bmatrix}, \qquad (2.2.17)$$

которую обозначим

$$BV = g. \qquad (2.2.18)$$

Предположим, что нами выбрана симметричная и положительно определенная матрица A_{11} такая, что симметричная матрица

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \qquad (2.2.19)$$

где $A_{12}=A_{21}^{\mathrm{T}}$, является положительно определенной. Рассмотрим стационарный итерационный метод

$$V^{k} = V^{k-1} - A^{-1}(BV^{k-1} - g), \quad k = 1,2,\ldots \qquad (2.2.20)$$

Нетрудно видеть, что подпространство $\overset{*}{U}$ из (2.2.6) инвариантно относительно матрицы BA^{-1}, а редуцированный итерационный метод выглядит следующим образом:

$$V_1^{k} = V_1^{k-1} - S_{11}^{-1}\,(B_{11}V_1^{k-1} - g_1), \quad k = 1,2,\ldots \qquad (2.2.21)$$

Опять мы приходим к выводу, что по отношению к подпространству $\overset{*}{U}$ матрица BA^{-1} является симметризуемым оператором. Использование свойств конструктивной симметризуемости по отношению к инвариантным подпространствам будет кратко рассмотрено в следующем разделе. Естественно, что все сказанное выше переносится на случай особенных матриц, но уже по отношению к подпространству $\overset{**}{U}$.

2.3. Обобщенный метод сопряженных градиентов. Пусть U - некоторое подпространство $\mathbb{R}^n$, инвариантное относительно матрицы AH, и матрица D удовлетворяет следующим условиям

$$(D\psi,\varphi) = (\psi,D\varphi) \quad \forall\psi,\varphi\in U,$$

$$(D\varphi,\varphi) > 0 \quad \forall\varphi\in U, \quad \varphi\neq\theta.$$

Матрицу AH мы будем называть D-самосопряженной в U, если

$$(DAH\psi,\varphi) = (D\psi, AH\varphi) \quad \forall\psi,\varphi\in U$$

и D-положительно определенной в U, если

$$(DAH\varphi,\varphi) > 0 \quad \forall\varphi\in U, \quad \varphi \neq \theta.$$

Предположим, что матрица AH является D-самосопряженной и D-положительно определенной в подпространстве U. Тогда для решения системы (2.1.1) при условии $\xi^0 \in U$ применим обобщенный метод сопряженных градиентов в подпространстве U. Двучленные формулы этого метода имеют вид [14]

$$p_k = \begin{cases} H\xi^0 & , k=1, \\ H\xi^{k-1} - \alpha_k \xi p_{k-1}, & k=1, \end{cases} \tag{2.3.1}$$
$$u^k = u^{k-1} - \beta_k p_k,$$

где

$$\alpha_k = \frac{(DAH\xi^{k-1}, Ap_{k-1})}{(DAp_{k-1},\ Ap_{k-1})}, \quad \beta_k = \frac{(D\xi^{k-1}, Ap_k)}{(DAp_k, Ap_k)}, \tag{2.3.2}$$
$$\xi^k = Au^k - f, \quad k = 1, 2, \ldots,$$

а трехчленные имеют вид [12], $e_{-1}=0$,

$$u^{k+1} = u^k - \frac{1}{q_k}\,[H\xi^k - e_{k-1}(u^k - u^{k-1})],$$
$$q_k = \frac{(DAH\xi^k, AH\xi^k)}{(DAH\xi^k, \xi^k)} - e_{k-1}, \quad e_k = q_k \frac{(DAH\xi^{k+1}, \xi^{k+1})}{(DAH\xi^k, \xi^k)}, \quad k=1,2,\ldots \tag{2.3.3}$$

Аналогичным образом могут быть рассмотрены обобщенный метод минимальных итераций Ланцоша и другие варианты методов сопряженных напряжений [3]. Некоторые вопросы реализации этих методов с учетом структуры подпространств U, которые базируются на алгоритмах частичного решения линейных систем, рассмотрены в [12].

§ 3. Параллельные вычисления

Как-то в середине 70-х годов на одной из встреч Вера Николаевна задала неожиданный вопрос: „А Вы еще не начали изучать векторные компьютеры? Мы с Дмитрием Константиновичем уже начали". Этим вопросом-ответом было сказано очень много - и выражена почти уверенность, что мы еще не начали, и констатирован факт, что они уже начали, и не оставлено место для сомнения, что именно это направление является наиболее важным.

Конечно, тогда мы еще не начали. Задавленное проблемами использования отечественной вычислительной техники наше сообщество математиков-вычислителей практически никак не реагировало на изменение путей развития архитектуры вычислительных систем на Западе. Да и сейчас, спустя полтора десятка лет, ситуация радикально не изменилась. Тем более удивительно, что Вера Николаевна и Дмитрий Константинович, имевшие в то время за плечами солидный груз лет, смогли увидеть не только новое перспективное направление, но и начали активно его развивать и пропагандировать. Их первые публикации [5] в области параллельных вычислений стали тем фундаментом, отталкиваясь от которого начала развиваться в нашей стране новая научная область.

3.1. Концепция неограниченного параллелизма. В работах [5] по существу была предложена концепция исследования параллельных свойств алгоритмов. Отдельные ее аспекты появились в конце 50-х - начале 60-х годов, когда о параллельных вычислительных системах только начали говорить. В то время относительно структуры этих систем было известно очень мало, разве лишь то, что в них одновременно может работать большое число устройств. Пока имели место разрозненные исследования, о концепции не было речи. С появлением обобщающего труда В.Н. и Д.К.Фаддеевых [5] стало ясно, что концепция уже сформировалась. Так или иначе, но на нее должна была последовать реакция со стороны математиков математиков-вычислителей.

Концепция получила название концепции неограниченного параллелизма. В ее основе явно или неявно лежит предположение, что алгоритм реализуется на параллельной вычислительной системе, не накладывающей на его реализации никаких ограничений, кроме числа используемых процессоров. В этих условиях основными объектами в изучении параллельных свойств алгоритма становятся число необходимых процессоров и число параллельных шагов алгоритма.

Концепция неограниченного параллелизма стала моделью, в рамках которой математики получили возможность приобщиться к параллельным компьютерам. В этом, безусловно, была большая ценность концепции, особенно на раннем этапе развития параллельных вычислений. Как и любая другая модель, концепция неограниченного параллелизма является идеализированной. Ее применение будет целесообразным лишь постольку, поскольку сделанные выводы не вступают в противоречие с практикой. Однако простота концепции оказалась столь привлекательной, что сама концепция неожиданно стала очень популярной. И сейчас на ее основе проводится немало исследований, хотя уже давно ясно, что ее основополагающие предположения не очень жизненные. Возможно, что каждый человек, интересующийся параллельными компьютерами, должен пройти через концепцию неограниченного параллелизма. А насколько долго он на ней задержится, зависит от того, что в действительности он хочет и каковы его цели.

Как бы ни относиться к концепции неограниченного параллелизма, в 70-годах она была прогрессивной моделью вычислений, во всяком случае, значительно более прогрессивной, чем концепция последовательных вычислений. Появление работ В.Н. и Д.К.Фаддеевых послужило толчком, вызвавшим огромный поток исследований в области параллельных вычислений. В этом легко убедиться, проанализировав, например, библиографию тех лет [19].

Исследование параллельных свойств методов обнаружило удивительные факты. Так, например, одно лишь изменение порядка суммирования может сделать метод обратной подстановки либо обладающим очень хорошими параллельными свойствами, либо просто последовательным. Выяснилось далее, что итерационные методы с хорошими свойствами скорости сходимости по числу операций не имеют достаточно большого параллелизма.

Стали устанавливаться факты связи параллельных свойств методов и возможности использования многоуровневой памяти. Но, пожалуй, самым неожиданным оказалось то, что практически все известные методы решения проблем собственных значений и сингулярных чисел для плотных матриц имеют очень плохие свойства с точки зрения использования параллельных компьютеров. Снова пришлось вернуться к интенсивному исследованию спектральных методов. Теперь на первый план опять выходят методы типа Якоби. А ведь они, казалось, были окончательно списаны со счета еще 25 лет назад.

Эти и многие другие факты стали известны далеко не сразу. На первых же порах проходило лишь осмысление отдельных результатов. Почти одновременно началась и критическая оценка самой концепции неограниченного параллелизма. Уточнялась модель вычислений. Вводились в рассмотрение новые характеристики такие, как загруженность процессоров, число конфликтов в памяти, требования к коммуникационной сети и многое другое.

3.2. Систематизация и узкие места. Нужно подчеркнуть очень важную особенность стиля исследований, проводимых Верой Николаевной и Дмитрием Константиновичем. Это - тщательная систематизация излагаемого материала. Их работы по параллельным вычислениям также содержат систематизированные сведения.

В наше время, когда поток публикаций исчисляется тысячами единиц, нет ничего более желанного в каждой предметной области, чем хорошо систематизированное ее описание. Наличие систематизированного материала не только указывает связь отдельных объектов между собой. Появляется возможность сравнивать объекты по тем или иным характеристикам, изучать истоки появления различий в характеристиках, исследовать еще не рассмотренные особенности и т.п. Систематизация - это верный ключ к поиску новых направлений и обнаружению узких мест в сформировавшихся направлениях. Но, наверное, нет труда более тяжелого, чем создание систематизированного описания. Остается только удивляться мужеству людей, которые все-таки берутся за подобные дела. Плохо написанный материал редко подвергается критике. Чаще всего его просто не замечают. Систематизированный материал подвергается критике почти всегда. Причем чем глубже систематизация, тем охотнее она критикуется. Ведь в этом случае в представленном материале всегда можно что-нибудь не найти. Не стали исключением и работы В.Н. и Д.К.Фаддеевых по параллельным вычислениям.

Конечно, концепция неограниченного параллелизма как модель вычислений плохо согласуется с реальностью. Несмотря на успехи микроэлектроники, трудно ожидать, что в ближайшее будущее для решения задач с матрицами порядка n будет предоставлена вычислительная система, имеющая в своем составе n^3 или n^4 параллельно работающих процессоров. Трудно также представить, что при большом числе процессоров каждый из них будет иметь быстрый доступ к любому участку памяти, причем независимо от других процессоров. Не будут процессоры выполнять операции точно.

Если требования на число процессоров и их связь с памятью сделать более жесткими, то многие из методов, рассматриваемых в рамках концепции неограниченного параллелизма, сразу оказываются несостоятельными. При малом числе процессоров приходится иначе организовывать вычислительные процессы. При большом числе процессоров с ограниченной связью с памятью трудности начинают появляться даже при реализации таких простых операций, как вычисление сумм чисел или скалярных произведений векторов по схеме сдваивания. Если число слагаемых сравнимо с числом процессоров или больше, то операции становятся весьма чувствительными к различным особенностям параллельных вычислительных систем.

Несмотря на то что исследование устойчивости к влиянию ошибок округления является обязательным при решении вопроса о целесообразности практического применения того или иного метода, концепция неограниченного параллелизма не акцентирует внимание на необходимости проведения такого рода исследований. Выполненные исследования параллельных методов решения задач алгебры показали исключительно сильную неустойчивость новых параллельных алгоритмов, имеющих высоты порядка $\log_2^2 n$ для задач с матрицами порядка n. По-видимому, без радикального изменения они не имеют сколько-нибудь серьезной перспективы для использования в больших параллельных вычислительных системах. Относительно быстрых устойчивых методов сейчас не известно даже, существуют ли они. Что же касается параллельных методов, полученных путем формального распараллеливания соответствующих последовательных алгоритмов, то, как выяснилось, влияние ошибок округления в них точно такое, как в исходных методах.

С точки зрения математика единственное достоинство параллельных вычислительных систем заключается в том, что они дают принципиальную возможность реализовывать алгоритмы быстрее, чем на последовательных ЭВМ, за счет использования большого числа процессоров. Но за достижение этой возможности приходится платить. Эта плата определяется не только числом используемых процессоров. С точки зрения математика ухудшаются свойства вычислительных систем. Точнее, у вычислительных систем появляются такие свойства, которые, с одной стороны, требуют обязательного своего учета для достижения высокой скорости решения задач, но, с другой стороны, они не привычны и не всегда легко понимаемы для математика. Отсюда идет стремление математика отгородиться от этих неудобных для него свойств вычислительной системы созданием некоторой подходящей математической модели и заменить исследование процессов решения задач на реальных вычислительных системах их модельными исследованиями. Критика концепции неограниченного параллелизма потому и возникла, что сама концепция не отражает многих особенностей реальных вычислительных процессов на параллельных системах.

Вокруг концепции неограниченного параллелизма появилось немало других исследований. Возникли даже новые направления. Одно из них связано со сложностью параллельных алгоритмов. Основными изучаемыми объектами этого направления являются верхние и нижние оценки высот алгоритмов для решения классов задач. Собственно говоря, основной предмет исследований представляют нижние оценки. Интерес к сложности параллельных алгоритмов вызывается во многом тем, что анализ этой проблемы может дать результаты, отличающиеся от результатов для последовательных вычислений.

Известно, например, что сложности задач нахождения произведения двух квадратных матриц и обратной матрицы в случае последовательных вычислений совпадают. Доказано, что, с некоторыми оговорками, алгоритм одной задачи порождает алгоритм той же сложности для другой задачи. Однако для параллельных вычислений подобное утверждение не известно. Для задачи нахождения произведения матриц существует параллельный алгоритм, высота которого по порядку зависимости от n совпадает с нижней оценкой высоты и равна $\log_2 n$. Для задачи вычисления обратной матрицы самый быстрый из известных алгоритмов имеет высоту порядка $\log_2^2 n$.

В рамках концепции неограниченного параллелизма стали развиваться новые подходы конструирования параллельных алгоритмов. Основные из них следующие:

расщепление задачи на большие независимые подзадачи;

выделение в задаче большого потока независимых однотипных подзадач;

применение принципа „разделяй и властвуй" с рекурсивным расщеплением большой задачи на две или несколько меньших задач.

Заметим, что новые параллельные алгоритмы для решения типовых задач, как правило, имеют более сложную структуру связей между операциями, чем традиционные последовательные алгоритмы, и не очень хорошо отображаются на архитектуру существующих параллельных систем. К тому же они обладают несходными вычислительными характеристиками в первую очередь в отношении устойчивости. И наконец, их реальное преимущество становится заметным только при достаточно большом числе процессоров.

В настоящее время основное внимание в деле создания параллельных методов и соответствующего программного обеспечения направлено на выявление параллельных структур в традиционных последовательных алгоритмах. Эта задача исключительно сложна и требует усилий многих коллективов. Требует она также опробирования различных подходов. Один из таких подходов и был заложен в работах Веры Николаевны и Дмитрия Константиновича. Его осмысление и критика позволили нащупать узкие места проблематики, устранить эти узкие места и начать новый цикл исследований.

Список литературы

[1] Фаддеев Д.К. Лекции по алгебре. М.: Наука, 1984.

[2] Фаддеев Д.К. О свойствах матрицы, обратной к хессенберговой // Зап.науч.семинаров ЛОМИ. 1981. Т.3. С.177-179.

[3] Фаддеев Д.К., Фаддеева В.Н. Вычислительные методы линейной алгебры. М., Л.: Физматгиз, 1963.

[4] Фаддеев Д.К., Фаддеева В.Н. Вычислительные методы линейной алгебры // Зап.науч.семинаров ЛОМИ. 1975. Т.54.

[5] Фаддеев Д.К., Фаддеева В.Н. Параллельные вычисления в линейной алгебре. Препринт ЛОМИ, Р-6-77, Р-6-81.

[6] Фаддеева В.Н. Вычислительные методы линейной алгебры. Гостехиздат, 1950.

[7] Фаддеева В.Н., Колотилина Л.Ю. Вычислительные методы линейной алгебры. Набор матриц для тестирования. Ч.1,2,3. Л., 1982.

[8] Фаддеева В.Н., Кузнецов Ю.А. и др. Вычислительные методы линейной алгебры. Библиографический указатель. Новосибирск, 1976.

[9] Gohberg I., Kailath T., Koltracht I. Efficient solution of linear systems of equations with recursive structure // Linear Algebra Appl. 1986. N 80. P.80-113.

[10] Колотилина Л.Ю. О некоторых структурных свойствах обратных матриц // Зап.науч.семинаров ЛОМИ. 1984. Т.139. С.61-73.

[11] Kuznetsov Yu.A. Matrix iterative methods in subspaces // Proc. Int. Congress. Math., Warsaw, 1983. North-Holland, Amsterdam, 1984. P.1509-1521.

[12] Кузнецов Ю.А. Вычислительные методы в подпространствах // Вычисл. процессы и системы. М.: Наука, 1985. С.265-350.

[13] Levinson N. The Wiener RMS essor critesion in filter besign and prediction // J. Math. Phys. 1947. N 25. P.261-278.

[14] Марчук Г.И., Кузнецов Ю.А. Итерационные методы и квадратичные функционалы. Новосибирск: ВЦ СО АН СССР, 1972. 205 с.

[15] Marchuk G.I., Kuznetsov Yu.A., Matsokin A.M. Fictitious domain and domain decomposition methods // Sov. J.Numer. Anal. Math. Modelling, 1. 1986. P.3-35.

[16] Нечепуренко Ю.М. Об одной факторизации элементов обратной матрицы // ЖВМ и МФ. 1984. N 4. С.601-605.

[17] Тыртышников Е.Е. Теплицевы матрицы, некоторые их аналоги и приложения. ОВМ АН СССР, 1989.

[18] Тыртышников Е.Е. Новые быстрые алгоритмы для систем с генкелевой и теплицевой матрицами // ЖВМ и МФ. 1989. Т.29, N 5. С.645-652.

[19] Воеводин В.В. Математические модели и методы в параллельных процессах. М.: Наука, 1986.

[20] Воеводин В.В., Тыртышников Е.Е. Вычислительные процессы с теплицевыми матрицами. М.: Наука, 1987.

Отдел вычислительной математики АН СССР
119034, Москва, ул.Рылеева, д.29

Поступило 14 июля 1990 г.