V. V. Voevodin, On an order of elimination of unknowns, *Zh. Vychisl. Mat. Mat. Fiz.*, 1966, Volume 6, Number 4, 758–760

ЖУРНАЛ
ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ и МАТЕМАТИЧЕСКОЙ ФИЗИКИ

Том 6 Июль — Август № 4

УДК 518:512.25

НАУЧНЫЕ СООБЩЕНИЯ

ОБ ОДНОМ ПОРЯДКЕ ИСКЛЮЧЕНИЯ НЕИЗВЕСТНЫХ

В. В. ВОЕВОДИН

(Москва)

Пусть решается система линейных алгебраических уравнений. Будем считать, что применяется один из вариантов метода исключения, основанный на использовании ортогональных преобразований (см. [1]).

Обозначим через A_0 расширенную матрицу системы. Реальный вычислительный процесс приводит к построению матриц $A_1, A_2, \ldots, A_N$, связанных между собой следующими рекуррентными соотношениями:

$$\begin{aligned} A_1 &= R_1A_0 + F_1, \\ A_2 &= R_2A_1 + F_2, \\ &\ldots\ldots\ldots \\ A_N &= R_NA_{N-1} + F_N. \end{aligned} \tag{1}$$

Здесь R_k — точная ортогональная матрица, вычисляемая по матрице A_{k-1}, F_k — матрица ошибок, вносимых на k-м шаге процесса за счет неточного вычисления R_k и неточного перемножения матриц, A_N — матрица с нулевыми поддиагональными элементами.

Для оценки погрешности решения исходной задачи необходимо оценить какую-нибудь норму (например, евклидову) матрицы

$$\Delta_N = A_N - R_NR_{N-1}\ldots R_1A_0.$$

В [2] получена практически неулучшаемая мажорантная оценка для циклического порядка исключения неизвестных. В настоящей заметке будут даны оценки, соответствующие другому порядку исключения. Эти оценки лучше оценок [2].

Предположим, что исключение поддиагональных элементов матрицы A_0 осуществляется с помощью матриц вращения T_{ij}. Обычно при умножении на T_{ij} исключается элемент, стоящий в позиции (i, j). Мы откажемся от этого требования и будем считать, что может исключаться любой элемент из i-й или j-й строки. Исключение элементов разобьем на циклы, которые определим следующим образом:

1) цикл состоит из конечного числа последовательных умножений на матрицы вращения;

2) при каждом умножении на матрицу вращения сохраняются все ранее исключенные элементы и дополнительно исключается еще один элемент, находящийся в столбце с наименьшим возможным номером;

3) никакая строка матрицы не преобразуется в течение цикла более одного раза;

4) в течение каждого цикла исключается максимально возможное количество элементов.

Поясним сказанное на примере матрицы 6-го порядка. Будем последовательно умножать на матрицы (циклы отделены точкой с запятой)

$$T_{21},\ T_{43},\ T_{65};\ T_{31},\ T_{64};\ T_{51},\ T_{42};\ T_{32},\ T_{64};\ T_{52},\ T_{43};\ T_{53},\ T_{64};\ T_{54};\ T_{65}$$

и при этом исключать такие элементы:

$$(2,1),\ (4,1),\ (6,1);\ (3,1),\ (6,2);\ (5,1),\ (4,2);\ (3,2),\ (6,3);\ (5,2),\ (4,3);\ (5,3),\ (6,4);\ (5,4);\ (6,5).$$

Методом математической индукции можно доказать, что число циклов, необходимое для исключения всех поддиагональных элементов матрицы, имеющей n строк, не превосходит $2(n-1)$. Интересно отметить, что некоторые матрицы вращения могут быть использованы несколько раз, а некоторые — ни разу.

Основной операцией этого метода является умножение вектора второго порядка слева на матрицу T_{ij}. Результатом операции будет

$$b = T_{ij}a + f, \tag{2}$$

где f — вектор ошибок, аналогичный матрице ошибок F_k из (1). Мы не будем подробно останавливаться на оценке f, так как она в значительной мере зависит как от способа вычисления матрицы T_{ij}, так и от способа выполнения арифметических операций в вычислительной машине. Заметим только, что для вычислений с плавающей запятой

$$\|f\|_E \leqslant c2^{-t}\|a\|_E, \tag{3}$$

а для фиксированной

$$\|f\|_E \leqslant d2^{-t}, \tag{4}$$

если $\|a\|_E \leqslant 1$. Здесь t — разрядность машины, c и d — константы, не зависящие ни от t, ни от a. По данным [2] имеем $c \leqslant 6$ при выполнении всех операций с t-разрядной точностью, $d \leqslant 2.5$ при использовании вычисления скалярных произведений с удвоенной точностью.

Пусть R_k — матрица преобразования, соответствующая k-му циклу, тогда из (1) получаем

$$\|\Delta_N\|_E \leqslant \|\Delta_{2(n-1)}\|_E \leqslant \|F_1\|_E + \|F_2\|_E + \ldots + \|F_{2(n-1)}\|_E. \tag{5}$$

Как следует из (2), (3),

$$\|b\|_E \leqslant (1 + c2^{-t})\|a\|_E,$$

поэтому

$$\|F_k\|_E \leqslant c2^{-t}\|A_{k-1}\|_E, \qquad \|A_k\|_E \leqslant (1 + c2^{-t})\|A_{k-1}\|_E.$$

Окончательно

$$\|\Delta_N\|_E \leqslant c2^{-t}[1 + (1 + c2^{-t}) + (1 + c2^{-t})^2 + \ldots + (1 + c2^{-t})^{2n-3}]\|A_0\|_E \leqslant \tag{6}$$
$$\leqslant 2c2^{-t}(n-1)(1 + c2^{-t})^{2n-3}\|A_0\|_E.$$

Эта оценка в $\sqrt{n/2}$ раз лучше соответствующей оценки (24.1) из [2], гл. III.

Аналогично получается оценка для фиксированной запятой. Обозначим через l_k количество элементов, исключенных в течение k-го цикла, и пусть v_k — фиксированный столбец матрицы A_k. Имеем

$$\|v_k\|_E \leqslant \|v_{k-1}\|_E + d2^{-t}\sqrt{l_k} \leqslant \|v_{k-1}\|_E + d2^{-t}\sqrt{n/2},$$

если только $\|v_{k-1}\|_E \leqslant 1$. Для выполнения условия $\|v_k\|_E \leqslant 1$ для всех $k = 0, 1, \ldots, 2(n-1)$ потребуем, чтобы

$$\|v_0\|_E \leqslant 1 - 2d2^{-t}(n-1)\sqrt{n/2}.$$

Предположим, что матрица A_0 имеет m столбцов, тогда

$$\|F_k\|_E \leqslant d2^{-t}\sqrt{ml_k},$$

и далее, учитывая (5),

$$\|\Delta_N\|_E \leqslant d2^{-t}\sqrt{m}(\sqrt{l_1} + \sqrt{l_2} + \ldots + \sqrt{l_{2(n-1)}}) \leqslant \tag{7}$$
$$\leqslant d2^{-t}\sqrt{m}\sqrt{2(n-1)}\sqrt{l_1 + l_2 + \ldots + l_{2(n-1)}} = d2^{-t}(n-1)\sqrt{mn}.$$

Эта оценка лучше соответствующей оценки (33.3) из [2] примерно в n раз, если исходные матрицы привести к одной норме.

З а м е ч а н и я. 1. Полученные формулы можно использовать для оценки погрешности преобразования каждого отдельного столбца g матрицы A_0, если в (6) заменить $\|A_0\|_E$ на $\|g\|_E$, а в (7) положить $m = 1$.

2. Описанный процесс позволяет получать более точное разложение произвольной матрицы в произведение ортогональной и треугольной матрицы, а также осуществлять более точное умножение произвольной матрицы на последовательность матриц вращения. Это дает возможность точнее реализовать некоторые методы решения полной проблемы собственных значений (метод односторонних вращений, ортогональный степенной метод, метод Якоби и т. п.).

3. Рассмотрим любой метод исключения, сводящийся к левостороннему умножению исходной матрицы на последовательность матриц вращения или матриц отражения. Пусть пересчет столбцов при каждом таком умножении осуществляется сколь угодно точно. В условиях t-разрядного хранения коэффициентов преобразуемой матрицы неизбежно по крайней мере одно последнее округление во всех пересчетах. Можно показать, что неулучшаемые мажорантные оценки лишь этих ошибок совпадают по порядку величины на классе матриц, соответственно, с оценками (6) и (7). Таким образом, полученные здесь оценки не могут быть существенно улучшены. В связи с этим заметим, что оценка (6) совпадает с оценкой из [2], полученной для метода отражений с использованием вычисления скалярных произведений с удвоенной точностью. При выводе оценки (6) применение подобной операции не предполагалось.

Поступила в редакцию
22.01.1966

Цитированная литература

1. Д. К. Фаддеев, В. Н. Фаддеева. Вычислительные методы линейной алгебры. М.— Л., Физматгиз, 1963.
2. J. H. Wilkinson. The algebraic eigenvalue problem. Oxford, Clarendon Press, 1965.

УДК 518:517.949.12

О ФОРМУЛАХ ЧИСЛЕННОГО ДИФФЕРЕНЦИРОВАНИЯ

Н. Н. ГУДОВИЧ

(Воронеж)

В [1] показано, что разностная схема

$$\frac{1}{h^n}[a_{-s}u(t-sh)+\ldots+a_{-1}u(t-h)+a_0u(t)+a_1u(t+h)+\ldots+a_lu(t+lh)] \tag{1}$$

тогда и только тогда аппроксимирует n-ю производную функции $u(t)$ с порядком $p+1$, когда ее характеристическая функция

$$\varphi(z)=\frac{a_{-s}}{z^s}+\ldots+\frac{a_{-1}}{z}+a_0+a_1z+\ldots+a_lz^l \tag{2}$$

имеет вид

$$\varphi(z)=z^{-s}(z-1)^n(1+C_1(z-1)+C_2(z-1)^2+\ldots), \tag{2'}$$

где коэффициенты $C_1, \ldots, C_p$ при заданном s определяются однозначно; там же даны формулы для этих коэффициентов. В настоящей заметке приводятся другие формулы для нахождения этих коэффициентов, преимуществом которых является возможность использования таблиц чисел Стирлинга. Изложим вывод этих формул.

Как легко устанавливается с помощью формулы Тейлора, для того чтобы выражение (1) аппроксимировало n-ю производную с порядком $p+1$, необходимо и достаточно, чтобы коэффициенты a_k схемы (1) удовлетворяли следующей линейной системе:

$$\sum_{k=-s}^{l}k^ia_k=\begin{cases}0, & i\neq n;\\ n!, & i=n;\end{cases}\qquad i=0,1,\ldots,n,n+1,\ldots,n+p. \tag{3}$$