

А. А. АБРАМОВ

ОБ ОШИБКЕ ОКРУГЛЕНИЙ ПРИ РЕШЕНИИ СИСТЕМ ЛИНЕЙНЫХ УРАВНЕНИЙ

(Представлено академиком М. А. Лаврентьевым 12 IV 1954)

В работе приводится оценка невязок, возникающих вследствие ошибок округления при решении системы линейных алгебраических уравнений методом исключения. В ряде вопросов (например, при уравнивании по методу наименьших квадратов) непосредственное значение имеют именно невязки, а не погрешности в неизвестных.

Часто системы обладают тем свойством, что многие коэффициенты в левых частях равны нулю, тем самым сокращается число действий, проводимых приближенно; оценка, приведенная в работе, учитывает это обстоятельство. Хотя сам способ подсчета погрешности чрезвычайно прост и не раз был использован ^(1, 2), однако окончательный результат предлагаемой работы не вытекает из выводов ^(1, 2) даже для систем полных уравнений.

Формулы (1), (2), (3) и (4) могут быть использованы для оценки накопления ошибок и для изменения точности на отдельных этапах вычислений.

1°. Рассмотрим влияние ошибок прямого хода при исключении. Пусть система

$$\sum_{j=1}^{n+1} a_{ij} x_j = 0 \quad *, \quad i = 1, 2, \dots, n,$$

приведена известным методом исключения Гаусса к виду

$$\sum_{j=i}^{n+1} b_{ij} x_j = 0, \quad i = 1, 2, \dots, n.$$

Здесь

$$b_{ij} = b_{ij}^{(i-1)}, \quad \text{где } b_{ij}^{(p)} = b_{ij}^{(p-1)} - \alpha_{pi} b_{pj}^{(p-1)}, \quad \alpha_{pi} = \frac{b_{ip}^{(p-1)}}{b_{pp}^{(p-1)}}, \quad i > p, \quad b_{ij}^{(0)} = a_{ij}.$$

При реализации этого приведения вносятся погрешности вычислений и фактически имеет место

$$\bar{b}_{ij}^{(p)} = \bar{b}_{ij}^{(p-1)} - \bar{\alpha}_{pi} \bar{b}_{pj}^{(p-1)} + \varepsilon_{ij}^{(p)}, \quad \bar{\alpha}_{pi} = \frac{\bar{b}_{ip}^{(p-1)}}{\bar{b}_{pp}^{(p-1)}} + \eta_{pi} \quad \text{при } i > p, j > p;$$

$$\bar{b}_{ij}^{(p)} = 0 \quad \text{при } i > p, j \leq p; \quad \bar{b}_{ij}^{(0)} = a_{ij}, \quad \bar{b}_{ij}^{(i-1)} = \bar{b}_{ij}.$$

* Здесь для простоты свободные члены включены в общую сумму, т. е. $x_{n+1} = 1$.

Пусть $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$ — решение системы

$$\sum_{j=1}^{n+1} \bar{b}_{ij} \bar{x}_j = 0, \quad i = 1, 2, \dots, n.$$

Найдем величины

$$\varphi_i = \sum_{j=1}^{n+1} a_{ij} \bar{x}_j.$$

Обозначим

$$\varphi_i^{(p)} = \sum_{j=p+1}^{n+1} \bar{b}_{ij}^{(p)} \bar{x}_j, \quad i \geq p.$$

Тогда

$$\varphi_i^{(p)} = \sum_{j=p+1}^{n+1} \bar{b}_{ij}^{(p)} \bar{x}_j, \quad \varphi_i^{(p-1)} = \sum_{j=p}^{n+1} \bar{b}_{ij}^{(p-1)} \bar{x}_j, \quad 0 = \sum_{j=p}^{n+1} \bar{b}_{pj}^{(p-1)} \bar{x}_j.$$

Изменим у обеих частей второго равенства знаки на обратные, умножим обе части третьего равенства на α_{pi} и сложим получившиеся три равенства. Получим

$$\varphi_i^{(p)} - \varphi_i^{(p-1)} = (\alpha_{pi} \bar{b}_{pp}^{(p-1)} - \bar{b}_{ip}^{(p-1)}) \bar{x}_p + \sum_{j=p+1}^{n+1} (\bar{b}_{ij}^{(p)} - \bar{b}_{ij}^{(p-1)} + \alpha_{pi} \bar{b}_{pj}^{(p-1)}) \bar{x}_j,$$

т. е.

$$\varphi_i^{(p)} - \varphi_i^{(p-1)} = \bar{b}_{pp}^{(p-1)} \eta_{pi} \bar{x}_p + \sum_{j=p+1}^{n+1} \varepsilon_{ij}^{(p)} \bar{x}_j.$$

Поэтому

$$\varphi_i = \varphi_i^{(0)} = - \sum_{p=1}^{i-1} (\varphi_i^{(p)} - \varphi_i^{(p-1)}) = - \sum_{p=1}^{i-1} \bar{b}_{pp}^{(p-1)} \eta_{pi} \bar{x}_p - \sum_{j=1}^{n+1} \left(\sum_{p=1}^{i-1} \varepsilon_{ij}^{(p)} \right) \bar{x}_j.$$

Будем считать, что действия проводятся с фиксированным числом знаков после запятой. Сложение и вычитание чисел при этом ошибок не вызывают. В тех случаях, когда происходит деление нуля на какое-либо число или умножение на ноль, действие проводится, очевидно, точно и вычислительной ошибки не возникает. В тех случаях, когда делится ненулевое число или производится умножение отличных от нуля чисел, может возникнуть ошибка. Эти ошибки мы будем считать независимыми случайными величинами

$$M\eta_{pi} = 0, \quad D\eta_{pi} = \sigma_1^2, \quad M\varepsilon_{ij}^{(p)} = 0, \quad D\varepsilon_{ij}^{(p)} = \sigma_2^2 \text{ при } j < n+1,$$

$$M\varepsilon_{in+1}^{(p)} = 0, \quad D\varepsilon_{in+1}^{(p)} = \sigma_3^2.$$

(Мы считаем, что получение величин $\bar{\alpha}_{pi}$ может вестись с другой точностью, чем $\bar{\alpha}_{pi} \bar{b}_{pj}$, $(n+1)$ -е неизвестное выделяется, так как оно играет роль правой части и в операциях над ним тоже возможна другая, чем над остальными, точность.)

При этих предположениях получаем

$$M\varphi_i = 0, \quad D\varphi_i = \sigma_1^2 \sum_{p=1}^{i-1} \bar{b}_{pp}^2 \bar{x}_p^2 \Delta_{ip} + \sigma_2^2 \sum_{j=1}^n A_{ij} \bar{x}_j^2 + \sigma_3^2 A_{in+1}. \quad (1)$$

Здесь $\Delta_{ip} = \text{sgn}^2 \bar{b}_{ip}^{(p-1)}$, $A_{ij} = \sum_{p=1}^{i-1} \text{sgn}^2 (\bar{b}_{ip}^{(p-1)} \bar{b}_{pj}^{(p-1)})$ — число таких преобразований (фактических) i -й строки, при которых коэффициент при x_j действительно преобразуется.

Если $\bar{x}_p^2 \leq \bar{x}^2$, то получаем оценку:

$$D\varphi_i \leq \sigma_1^2 \bar{x}^2 \sum_{p=1}^{i-1} \bar{b}_{pp}^2 \Delta_{pi} + \sigma_2^2 \bar{x}^2 A_i + \sigma_3^2 A_{in+1}. \quad (2)$$

Здесь A_i — полное число всех фактически проводимых действий в i -й строке (кроме преобразований $b_{in+1}^{(p)}$).

Если получение величин α_{ip} ведется с различной точностью, то первое слагаемое в (1) изменяется; так, если $D\eta_{pi} = \frac{\tilde{\sigma}_1^2}{\bar{b}_{pp}^2}$, то

$$\begin{aligned} D\varphi_i &= \tilde{\sigma}_1^2 \sum_{p=1}^{i-1} \Delta_{pi} \bar{x}_p^2 + \sigma_2^2 \sum_{j=1}^n A_{ij} \bar{x}_j^2 + \sigma_3^2 A_{in+1} \leq \\ &\leq \tilde{\sigma}_1^2 \bar{x}^2 \sum_{p=1}^{i-1} \Delta_{pi} + \sigma_2^2 \bar{x}^2 A_i + \sigma_3^2 A_{in+1}, \end{aligned} \quad (3)$$

$\sum_{p=1}^{i-1} \Delta_{pi}$ — число преобразований i -й строки при помощи других.

2°. Для ошибок обратного хода при решении методом исключения имеем

$$\sum_{j=1}^{n+1} b_{ij} \bar{x}_j = \theta_i b_{ii},$$

где θ_i — ошибка деления $\sum_{j=i+1}^{n+1} b_{ij} \bar{x}_j$ на b_{ii} .

Так как

$$\sum_{j=1}^{n+1} a_{ij} \bar{x}_j = \sum_{j=1}^{n+1} b_{ij}^{(i-1)} \bar{x}_j + \sum_{p=1}^{i-1} \alpha_{pi} \sum_{j=1}^{n+1} b_{pj} \bar{x}_j,$$

то получаем

$$\sum_{j=1}^{n+1} a_{ij} \bar{x}_j = \theta_i b_{ii} + \sum_{p=1}^{i-1} \alpha_{pi} b_{pp} \theta_p = \theta_i b_{ii} + \sum_{p=1}^{i-1} b_{ip}^{(p-1)} \theta_p = \sum_{p=1}^i b_{ip}^{(p-1)} \theta_p.$$

Попрежнему будем считать $\theta_1, \theta_2, \dots, \theta_n$ независимыми случайными величинами, $M\theta_i = 0$, $D\theta_i = \sigma_i'^2$. Тогда для $\psi_i = \sum_{j=1}^{n+1} a_{ij} \bar{x}_j$ получим $M\psi_i = 0$ и

$$D\psi_i = \sum_{p=1}^{i-1} (b_{ip}^{(p-1)})^2 \sigma_p'^2. \quad (4)$$

Поступило
5 IV 1954

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

¹ А. Тюринг, Усп. матем. наук, 6, в. 6 (1951). ² J. v. Neumann, H. H. Goldstein, Bull. Am. Math. Soc., 53, 1021 (1947).