ДОКЛАД

профессора Петрова Ю.П. "Точная оценка погрешностей решений систем алгебраических уравнений"на расширенном семинаре кафедры МЭКС факультета ПМ-ПУ СПбГУ, заслушанном 9 сентября 2008 года.

Доклад посвящён проблеме оценки погрешностей решений $x_1; x_2; ...; x_n$ систем уравнений вида

$$Ax = B, \tag{1}$$

точнее - погрешностей происходящих из-за вариаций коэффициентов матрицы A (квадратной матрицы размера nxn) и вектор-столбца правой части B. Эти коэффициенты определяются почти всегда из опыта или измерений, имеющих конечную точность, поэтому обычно известны лишь интервалы, внутри которых заключены их точные значения $\bar{a}_{ij}$, и $\bar{b}_i$, и при расчётах приходится учитывать неравенства:

$$a_{ij}(1 - \varepsilon_{ij} \leq \bar{a}_{ij}) \leq a_{ij}(1 + \varepsilon_{ij}) \tag{2}$$

$$b_i(1 - \delta_i) \leq \bar{b}_i \leq b_i(1 + \delta_i) \tag{3}$$

в которых a_{ij} и b_i - номинальные значения коэффициентов используемые при расчёте, а $\varepsilon_{ij} a_{ij}$ и $b_i \delta_i$ - их вариации. Вариации коэффициентов неизбежно приводят к вариациям решений и поэтому с учётом неравенств (2) и (3) можно лишь утверждать, что каждое из решений x_i системы (1) находится внутри некоторого интервала:

$$x_i - \Delta_1 x_i \leq \bar{x}_1 \leq x_i + \Delta_2 x_i \tag{4}$$

Величины $\Delta_1 x_i$ и $\Delta_2 x_i$ можно рассматривать как погрешности решения x_i, происходящие из-за вариаций (погрешностей) коэффициентов, и очень важной задачей является их оценка. До тех пор пока величина погрешности - или хотя бы её "оценка сверху не известна, не вычислена, любое решение не надёжно. Сложность задачи оценки погрешности заключается в том, что погрешность зависит от сочетания знаков вариаций $\varepsilon_{ij} a_{ij}$, а число возможных сочетаний знаков равно 2^{n^2} и очень быстро растёт с ростом порядка системы (1). Если при $n = 3$ будет $2^{n^2} = 512$, то уже при $n = 5$ их будет $2^{25} > 10^7$, а при $n = 10$ будет $2^{100} > 10^{30}$. Между тем каждая комбинация знаков ε_{ij} приводит к своей погрешности Δx_i, часто отличной от погрешности, возникающей при других комбинациях знаков.

Оценка погрешностей решений систем вида (1) особенно важна потому, что необходимость решать подобные системы возникает в самых различных областях техники и физики.

1. В задачах строительной механики, сопротивления материалов, механики твёрдого тела - при расчёте усилий и нагрузок в элементах самых различных конструкций постоянно приходится решать системы вида (1) самых различных порядков.

2. В электротехнике при расчёте токов и напряжений уравнения Кирхгофа приводят к системам вида (1).

3. При решении дифференциальных уравнений в частных производных одним из этапов вычисления часто оказывается, как известно, решение систем вида (1).

4. То же относится и к очень широкому кругу задач, сводящихся к решению интегральных уравнений. Как известно, при приближённой замене интеграла, входящего в уравнения, конечной суммой возникает система уравнений вида (1), коэффициенты которой приближенны.

Широкая распространённость уравнений вида (1) и важность проблемы оценки влияния вариации коэффициентов на погрешность решения обусловила большое число исследований, посвящённых этой проблеме.

Наиболее распространённой и приводимой во многих учебниках по методам вычислений является оценка по "числу обусловленности", равному произведению нормы матрицы A в системе (1) на норму обратной матрицу A^{-1} (смотри публикации [1], [2], [3]). Если для всех ε_{ij} и δ_i справедливы оценки $|\varepsilon_{ij}| \leq \varepsilon_0$ и $|\delta_i| \leq \delta_0$, то используют формулу:

$$\frac{||\Delta x||}{||x||} \leq ||A|| \cdot ||A^{-1}||(\frac{\varepsilon_0}{||A||} + \frac{\delta_0}{||B||}) \tag{5}$$

(двойными вертикальными чертами обозначены нормы матрицы и вектора).

Оценка по "числу обусловленности", хотя и считается наиболее надёжной и чаще всего используется, обладает целым рядом недостатков:

1. Она даёт лишь осреднённую оценку всех состовляющих вектора погрешности Δx_i, в то время как наиболее важно дать оценку погрешности каждого из решений x_i (если, например, погрешность решения x_1 велика, то уже одно это может быть причиной аварии, невзирая на то, что погрешности решений x_i, $i > 1$, а значит и $||\Delta x||$, малы).

2. В большинстве конкретных примеров оценка по "числу обусловленности"не является точной и оказывается завышенной и грубой.

3. Недавно были обнаружены и описаны в [4] дополнительные недостатки оценки по "числу обусловленности". Поэтому важную задачу оценки погрешности решений систем алгебраических уравнений до последнего времени нельзя было считать решённой.

Большой специалист по интегральным уравнениям и решению систем вида (1) профессор В.С.Сизиков сказал в прошлом году:"тот, кто найдёт метод точной оценки погрешности каждой составляющей x_i решения системы (1) получит Нобелевскую премию".

В книге [4] на основе методики "модульных определителей"была найдена оценка погрешности каждой составляющей x_i решения системы (1), но эта оценка является точной лишь для $n = 2$, а для $n > 2$ она служит только "оценкой сверху". Дальнейшие исследования позволили на основе той же методики, описанной в [4],получить более сложную, но уже точную оценку погрешности и найти величины $\Delta_1 x_i$ и $\Delta_2 x_i$

в неравенствах (4). Слова "точная оценка"означают, что для каждой конкретной системы уравнений (1) существует такая комбинация вариаций коэффициентов a_{ij} и b_i, подчинённых неравенствам (2) и (3), для которой неравенства (4) выполняются со знаком равенства. Численная проверка на ЭВМ, выполненная М.В.Волошиным, подтвердила точность оценки. Оценка погрешностей решений сводится - как ранее было показано в [4] - к предварительной оценке погрешностей определителей, входящих в формулы Крамера, и отысканию наиболее неблагоприятного сочетания знаков погрешностей ε_{ij} и δ_i.

Примеры.

Пример 1. Рассматривается определитель:

$\det = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 1 & 2 \\ 3 & 4 & 5 \end{vmatrix} = 8$, для которого, согласно методике, используемой в [4] наиболее неблагоприятное сочетание знаков вариаций его элементов соответствует таблице: $\begin{vmatrix} - & - & + \\ + & - & + \\ + & + & - \end{vmatrix}$ - то есть, при $|\varepsilon_{ij}| = 0,01$ определитель, после вариаций его элементов, переходит в определитель:

$$\begin{vmatrix} 0,99 & 1,98 & 3,03 \\ 4,04 & 0,99 & 2,02 \\ 3,03 & 4,04 & 4,95 \end{vmatrix} = 9,6604 \tag{6}$$

который в наибольшей степени для данного $|\varepsilon_{ij}| = 0,01$ отличается от исходного опредлелителя.

В ходе "численного эксперимента"были вычислены все $2^9 = 512$ определителей со всеми возможными сочетаниями знаков вариаций элементов $\pm 0,01$ и было подтверждено, что действительно наихудшее сочетание знаков вариаций элементов и наибольшая погрешность $\Delta_2 = 1,6604$ будем реализовываться именно в определителе (6).

Пример 2.

В известном учебнике [5] на странице 205 рассмотрен расчёт усилий и моментов в одной из конструкций. Этот расчёт сводится, как показано в [5], к решению системы уравнений:

$$\begin{cases} 14x_1 + 12x_2 + 15x_3 = 3 \\ 12x_1 + 16x_2 + 12x_3 = 5 \\ 5x_1 + 4x_2 + 6x_3 = 1 \end{cases} \tag{7}$$

(x_1 и x_2 - усилия, x_3 - момент).

Используя формулы Крамера, получим:

$$x_1 = \frac{D_1}{D} = \frac{\begin{vmatrix} 3 & 12 & 15 \\ 5 & 16 & 12 \\ 1 & 4 & 6 \end{vmatrix}}{\begin{vmatrix} 14 & 12 & 15 \\ 12 & 16 & 12 \\ 5 & 4 & 6 \end{vmatrix}} = -0,25; \; x_2 = \frac{D_2}{D} = \frac{\begin{vmatrix} 14 & 3 & 15 \\ 12 & 5 & 12 \\ 5 & 1 & 6 \end{vmatrix}}{\begin{vmatrix} 14 & 12 & 15 \\ 12 & 16 & 12 \\ 5 & 4 & 6 \end{vmatrix}} = 0,4375;$$

$$x_3 = \frac{D_3}{D} = \frac{\begin{vmatrix} 14 & 12 & 3 \\ 12 & 16 & 5 \\ 5 & 4 & 1 \end{vmatrix}}{\begin{vmatrix} 14 & 12 & 15 \\ 12 & 16 & 12 \\ 5 & 4 & 6 \end{vmatrix}} = 0,0733 \tag{8}$$

Исследуя возможные погрешности решений x_1; x_2; x_3 с помощью "числа обусловленности", получаем: евклидова норма матрицы А уравнения (7) равна: $||A|| = 34,438$, та же норма обратной матрицы: $||A^{-1}|| = 2,907$, число обусловленности $||A|| \cdot ||A^{-1}|| = 100,111$, что говорит, казалось бы, о хорошей обусловленности решений (8).

Это обстоятельство, очевидно, и стало причиной того, что система (7) оказалась в качестве примера в учебнике [5], который только к 1979 году выдежрал восемь изданий большими тиражами. Этот пример решали десятки тысяч студентов, их решения проверяли тысячи опытных преподавателей, но никто не заметил, что решение x_3 совершенно ненадёжно.

Действительно, рассматривая определитель D_3:

$$D_3 = \begin{vmatrix} 14 & 12 & 3 \\ 12 & 16 & 5 \\ 5 & 4 & 1 \end{vmatrix} = 4, \tag{9}$$

можно установить, что наиболее неблагоприятное сочетание знаков вариаций его элементов соответствует таблице: $\begin{vmatrix} - & + & - \\ + & - & + \\ + & - & + \end{vmatrix}$ и поэтому уже при $|\varepsilon_{ij}| = 0,0066$ определитель D_3 изменит свой знак, а это означает, что исследуемый момент x_3 уже при $|\varepsilon_{ij}| = 0,6\%$ будет иметь знак, противоположный расчётному, а поэтому вся рассчитываемая конструкция может рухнуть, поскольку все элементы конструкции вряд ли имели точность, лучшую, чем 1%.

Таким образом, традиционные методы оценки погрешности при расчёте не позволили заметить ненадёжность расчёта исследуемой конструкции, а усовершенствованные методы смогли это сделать и тем самым позволили предотвротить возможную аварию (авария произошла бы в том случае, если конструкию, описанную в [5], и кажущуюся безопасной при расчёте по "числу обусловленности", стали бы воплощать в реальное изделие).

Обобщения. Рассмотренная методика позволяет дать точную оценку погрешности любого решения x_i системы (1) не только при относительных вариациях коэффициентов a_{ij} и b_i, описываемых неравенствами (2) и (3), но при "абсолютных"вариациях, когда вместо (2) имеют место неравенства:

$$a_{ij} - \varepsilon_{ij} \leq \bar{a}_{ij} \leq a_{ij} + \varepsilon_{ij}. \tag{10}$$

Точно так же может быть дана точная оценка погрешности решения для тех случаев, когда известно, например, что некоторые $\varepsilon_{ij} = 0$. Дело в том, что главная трудность заключается в нахождении наиболее неблагоприятных (для величины погрешности решения) сочетаний знаков чисел ε_{ij}. Все остальное легко обобщается на самые различные варианты, встречающихся на практике. Проведена и проверка путём численного эксперимента на ЭВМ.

Оценка необходимого объёма вычислений.

Как известно, для вычисления определителя порядка n требуется примерно n^3 умножений (обоснование приведено, например в [1], [2]). Поэтому для получения "оценки сверху"о погрешности решения x_i на основе приведённого в [4] метода "модульных определителей"требуется примерно $2n^3$ умножений, а для получения точной оценки погрешности x_i требуется примерно n^5 умножений. При $n = 5$ будет требоваться $5^5 = 3125$ умножений, при $n = 10$ потребуется 10^5 умножений. Это не очень мало, но для компьютера посильно.

Для сокращения объёма вычислений можно комбинировать два метода: сперва вычислить "оценку сверху"по методике, подробно описанной в [4], что требует $2n^3$ умножений, и только если оценка сверху оказалась пессимистичной, перейти к предлагаемой докладчиком точной оценке, требующей n^5 умножений.

Практические применения.

Доложенные результаты могут быть использованы для существенного улучшения надёжности результатов расчёта во всех тех многочисленных разделах физики и техники, в которых одним из этапов расчёта является решение систем линейных алгебраических уравнений.

Литература.

1. Зализняк В.Е. Основы научных вычислений. Москва-Ижевск, 2006, 264 стр.

2. Шноль Э.Э. Семь лекций по вычислительной математике M.URRS, 3-е издание, 2007, 106 стр.

3. Мышкис А.Д. Прикладная математика для интенеров, специальные курсы, 3-е издание, М. Физматгиз, 2007, 687 стр.

4. Петров Ю.П. Обеспечение достоверности и надёжности компьютерных расчётов. СПб, издательство "БХВ-Петербург", 2008, 160 стр.

5. Феодосьев В.И. Сопротивление материалов. Учебник для ВУЗов, издание восьмое. М. Наука, 1979, 559 стр.