Московский международный институт эконометрики, информатики, финансов и права

И.Н. Мастяева
О.Н. Семенихина

Вычислительная математика

Учебное пособие

Москва 2004

ББК 22.19
УДК 519.6

Мастяева И.Н., Семенихина О.Н. Численные методы: Учебное пособие / Московский международный институт эконометрики, информатики, финансов и права. –М., 2004. –103 стр.

В пособии излагаются численные методы алгебры, анализа и решения дифференциальных уравнений, наиболее часто применяемые при решении практических задач на ЭВМ.

Пособие предназначено для студентов MIFP всех специальностей.

Рецензенты: к.э.н. И.Н. Орлова, к. физ.-мат. н. Э.И. Применко

СОДЕРЖАНИЕ:

1. ПОГРЕШНОСТЬ РЕЗУЛЬТАТА ЧИСЛЕННОГО РЕШЕНИЯ ЗАДАЧИ

1.1. Источники и классификация погрешностей

Погрешности решения задач обусловливаются следующими причинами:

1. Математическое описание задач (математическая модель) большей частью является неточным;

2. Методы решения задач (например, дифференциального уравнения) не являются точными. Во многих случаях получение точного решения требует выполнения неограниченного количества шагов. Обрыв бесконечного процесса приводит к получению приближенного решения;

3. Исходные данные для решения задач, как правило, получаются из эксперимента, а каждый эксперимент может дать результат с ограниченной точностью;

4. При вводе исходных данных в машину, при выполнении арифметических операций, при выводе информации производятся округления;

5. Погрешности приближенных чисел (погрешности исходных данных и погрешности округления) в процессе решения задачи будут последовательно переходить (чаще всего в увеличенном размере) в результаты вычислений и порождать новые погрешности.

В соответствии с указанными источниками погрешностей можно осуществить классификацию последних:

А) Неустранимые погрешности:

1) математического описания задачи [1];

2) исходных данных [3];

Б) погрешности метода [2];

В) вычислительные погрешности [4,5].

1.2. Точные и приближенные числа. Правила округления чисел

В повседневной практической деятельности, а также при решении той или иной задачи используются числа двух видов: точные и приближенные. Например, число 3 является точным, если речь идет о числе сторон треугольника. Если же число 3 – длина стороны треугольника или оно используется вместо числа π в вычислениях, то оно будет числом приближенным.

Определение 1. Приближенным числом a называется число, незначительно отличающееся от точного числа A и заменяющее его в вычислениях.

Приближенные числа обычно получаются в результате либо измерений, либо счета, либо выполнения различных математических

операций (округления, деления, извлечения корня, вычисления синуса, логарифма и т.д.).

При работе с приближенными числами вычислитель должен уметь решать следующие задачи:

1. давать математические характеристики точности приближенных чисел;
2. зная степень точности исходных данных, оценивать степень точности результата (прямая задача теории погрешностей);
3. выбирать исходные данные с той точностью, которая обеспечит заданную точность результата (обратная задача теории погрешностей);
4. оптимальным образом строить вычислительный процесс, чтобы не производить расчетов, не влияющих на точные цифры результата.

Как уже говорилось, одним из источников получения приближенных чисел является округление. Сформулируем правила округления:

1. если отбрасываемые при округлении цифры составляют число, большее половины единицы последнего оставляемого разряда, то последняя оставляемая цифра увеличивается на единицу;
2. если отбрасываемые при округлении цифры составляют число, меньшее половины единицы последнего оставляемого разряда, то оставляемые цифры остаются без изменения;
3. при округлении, когда отбрасываемые цифры составляют число, равное половине единицы последнего оставляемого разряда, то последняя оставляемая цифра увеличивается на единицу, если она нечетная, и остается без изменения, если она четная (правило четной цифры).

Пример 1. Округлить следующие числа:

$A_1 = 271{,}5001$ до целых,
$A_2 = 271{,}15$ до десятых,
$A_3 = 271{,}25$ до десятых,
$A_4 = 0{,}15497$ до сотых.

Решение. Так как при округлении числа A_1 до целых отбрасываемые цифры (5001) составляют число 0,5001, большее половины от единицы (последнего оставляемого разряда), последняя оставляемая цифра увеличивается на единицу (пункт 1). Поэтому после округления A_1 получаем число $a_1 = 272$.

При округлении чисел A_2 и A_3 по правилу четной цифры (пункт 3) получаем $a_2 = 271{,}2$; $a_3 = 271{,}2$.

При округлении числа A_4 (пункт 2) получаем $a_4 = 0{,}15$.

1.3. Математические характеристики точности приближенных чисел

Определение 2. Абсолютной погрешностью приближенного числа a назовем величину Δa, про которую известно, что

$$|A - a| \le \Delta a. \qquad (1.1)$$

Таким образом, точное число заключено в границах

$$a - \Delta a \le A \le a + \Delta a \qquad (1.2)$$

или сокращенно

$$A = a \pm \Delta a. \qquad (1.3)$$

Пример 2. Приближенные числа $a_1 = 2{,}87; a_2 = 300; a_3 = 3 \cdot 10^2$ получены округлением, точные значения чисел неизвестны. Что можно сказать об абсолютной погрешности данных приближенных чисел?

Решение. Пользуясь правилами округления чисел, можно сказать, что абсолютные погрешности приближенных чисел не превосходят половины единицы последнего разряда, т.е.

$$|A_1 - a_1| \le 0{,}005 = \Delta a_1,$$
$$|A_2 - a_2| \le 0{,}5 = \Delta a_2,$$
$$|A_3 - a_3| \le 50 = \Delta a_3.$$

Кроме того, можно записать:

$$A_1 = 2{,}87 \pm 0{,}005,$$
$$A_2 = 300 \pm 0{,}5,$$
$$A_3 = (3 \pm 0{,}5) \cdot 10^2.$$

Пример 3. Округлить числа π = 3,14159265…и e = 2,71828182…до сотых и определить абсолютную погрешность полученных приближенных чисел.

Решение. В силу правил округления имеем

$a_1 = 3{,}14$; $a_2 = 2{,}72$.

По определению абсолютной погрешности

$$|\pi - a_1| = |3{,}14159... - 3{,}14| = 0{,}00159... < 0{,}0016 = \Delta a_1,$$
$$|e - a_2| = |2{,}71828... - 2{,}72| = 0{,}00171... < 0{,}0018 = \Delta a_2.$$

Замечание 1. Абсолютную погрешность принято записывать в виде числа, содержащего не более одной или двух цифр, отличных от нуля (двух значащих цифр).

Замечание 2. В силу определения погрешности абсолютную погрешность округляют до одной или двух значащих цифр только в большую сторону (не придерживаясь сформулированных выше правил округления чисел).

В примере 3 в качестве абсолютной погрешности чисел a_1 и a_2 можно взять значения:

$$\Delta a_1 = 0{,}0016 \quad \textit{либо} \quad \Delta a_1 = 0{,}002.$$
$$\Delta a_2 = 0{,}0018 \quad \textit{либо} \quad \Delta a_2 = 0{,}002.$$

Абсолютная погрешность отражает лишь количественную сторону погрешности, но не качественную, т.е. не показывает, хорошо или плохо проведено измерение или вычисление.

Пример 4. при измерении толщины и длины крышки стола были получены результаты:

$l_1 = 3\,см \pm 1\,см;\quad l_2 = 113\,см \pm 1\,см.$

Определить, в каком случае измерение было сделано более качественно.

Решение. Абсолютная погрешность измерения для l_1 и l_2 одинакова и равна

$$\Delta l_1 = \Delta l_2 = 1\,см\,.$$

Однако очевидно, что измерение l_2 было проведено более качественно, чем l_1 . Для того, чтобы определить качество измерений и вычислений, необходимо выяснить, какую долю составляет абсолютная погрешность от определяемой величины. В связи с этим вводится понятие относительной погрешности.

Определение 3. Относительной погрешностью δa приближенного числа a называется отношение абсолютной погрешности Δa к абсолютной величине приближенного числа a:

$$\delta a = \frac{\Delta a}{|a|};\ a \neq 0 \qquad (1.4)$$

В примере 4 относительные погрешности измерения толщины и длины соответственно равны

$$\frac{\Delta l_1}{|l_1|} = \delta l_1 = \frac{1\,см}{3\,см} = 0{,}34\ или\ 34\%\,,$$

$$\frac{\Delta l_2}{|l_2|} = \delta l_2 = \frac{1\,см}{113\,см} = 0{,}0089\ или\ 0{,}9\%\,.$$

Следовательно, измерение длины l_2 было произведено намного качественнее.

Замечание 3. Относительная погрешность представляет собой безразмерную величину.

При вычислении относительную погрешность округляют в большую сторону и записывают в виде числа, содержащего одну-две значащие цифры.

1.4. Число верных знаков приближенного числа. Связь абсолютной и относительной погрешности с числом верных знаков. Правила подсчета числа верных знаков

Всякое положительное десятичное число a может быть единственным образом представлено в виде конечной или бесконечной десятичной дроби:

$$a = \alpha_1 \cdot 10^m + \alpha_2 \cdot 10^{m-1} + \ldots + \alpha_n \cdot 10^{m-n+1} + \ldots \qquad (1.5)$$

или $$a = \sum_{i=1}^{n} \alpha_i \cdot 10^{m-i+1} \qquad (1.6)$$

где α_i - десятичные цифры ($\alpha_i = 0,1,2,\ldots,9$), причем $\alpha_1 \neq 0$, m – некоторое число (старший разряд числа a). Например, в десятичной системе счисления:

$$28{,}0496 = 2 \cdot 10^1 + 8 \cdot 10^0 + 0{,}10^{-1} + 4 \cdot 10^{-2} + 9 \cdot 10^{-3} + 6 \cdot 10^{-4};$$

$$7{,}54 = 7 \cdot 10^0 + 5 \cdot 10^{-1} + 4 \cdot 10^{-2};$$

$$0{,}006783 = 6 \cdot 10^{-3} + 7 \cdot 10^{-4} + 8 \cdot 10^{-5} + 3 \cdot 10^{-6}.$$

Определение 4. Значащими цифрами числа a называют все цифры в его записи (1.5) начиная с первой слева, отличной от нуля. Например, приводимые ниже числа имеют следующее количество значащих цифр:

5423,47	6 значащих цифр,
0,0000605	3 значащие цифры,
0,060500	5 значащих цифр.

Как видно из приведенных примеров, цифра 0 имеет особое значение при определении числа значащих цифр. Например, в числе 0,00710300 первые три нуля не являются значащими цифрами и служат только для установления старшего десятичного разряда числа. Остальные три являются значащими цифрами, так как первый из них находится между значащими цифрами, а второй и третий, как отражено в записи, указывают, что в приближенном числе сохранены десятичные разряда 10^{-7} и 10^{-8}. Если же в данном числе 0,00710300 последние две цифры не являются значащими цифрами, то это число лучше записать в виде 0,007103. Числа 0,00710300 и 0,007103 не равноценны, так как первое из них имеет 6 значащих цифр, а второе – только 4 значащих цифры. Цифра 0, стоящая в конце числа, может иметь двоякий смысл, как это видно из следующих утверждений:

а) 1 кг = 1000 г;

б) население США по одной из переписей составляло 195530000 человек

В первом случае имеем точное соотношение, поэтому все нули здесь – значащие цифры. Во втором случае нули стоят вместо неизвестных цифр, и число имеет только 5 значащих цифр. Для того чтобы избежать недоразумения, никогда не следует писать нули вместо неизвестных цифр, а лучше применять такую форму записи:

$$19553 \cdot 10^4 \quad \textit{или} \quad 1{,}9553 \cdot 10^8$$

Пример 5. Пусть в результате измерения получено число, имеющее две значащие цифры, l = 72 мм. Если этот результат, не измеряя отрезок с большей точностью, выразить в метрах, километрах или микронах и написать, что l =0,072 м, или l =0,000072 км, или l =72000 мкм, то нули ни в первом, ни во втором, ни в третьем случаях не будут значащими. В дальнейшем условимся различать такие числа, как 7,2; 7,20; 7,200.

Все они выражают одно и то же числовое значение некоторой величины, но определены с разным количеством значащих цифр.

Точность приближенного числа зависит не от количества значащих цифр, а от количества верных значащих цифр. Различают значащие цифры верные в узком и широком смыслах.

Определение 5. Цифры $\alpha_1, \alpha_2, \ldots, \alpha_n$ приближенного числа a называют верными в узком смысле, если абсолютная погрешность Δa приближенного числа a не превосходит половины единицы *(m-n+1)* – го разряда, которому принадлежит цифра α_n, т.е. если

$$\Delta a \le 0.5 \cdot 10^{m-n+1}. \qquad (1.7)$$

Пример 6. Оценить абсолютную погрешность приближенного числа $a = 4{,}483$, если известно, что оно имеет 3 верных знака в узком смысле.

Решение. По определению 5

$$\Delta a \le 0.5 \cdot 10^{m-n+1}.$$

В нашем случае старший разряд числа равен 10^0, т.е. $m = 0$, а $n = 3$. Поэтому получаем

$$\Delta a \le 0{,}5 \cdot 10^{0-3+1} = 0{,}5 \cdot 10^{-2} = 0{,}005.$$

В математических таблицах все числа определены до верных значащих цифр в узком смысле. Так, например, в четырехзначных таблицах Брадиса В.М. гарантировано, что абсолютная погрешность квадратных корней не превосходит $0{,}5*10^{-3}$ (так как там приведены квадратные корни чисел от 1 до 100). В некоторых случаях, например при получении числа путем измерения, удобнее говорить о числе верных знаков в широком смысле.

Определение 6. Цифры $\alpha_1, \alpha_2, \ldots, \alpha_n$ приближенного числа a называют верными в широком смысле, если абсолютная погрешность Δa приближенного числа a не превосходит единицы *(m-n+1)* – го разряда, которому принадлежит цифра α_n, т.е. если

$$\Delta a \le 10^{m-n+1}. \qquad (1.8)$$

Например, если число $a = 4{,}483$ имеет $n = 3$ верных знака в широком смысле, то его абсолютная погрешность не превосходит

$$\Delta a \le 1 \cdot 10^{0-3+1} = 1 \cdot 10^{-2} = 0{,}01 \quad \textit{или} \quad \Delta a \le 0{,}01.$$

Определения 5 и 6 можно обобщить.

Определение 7. Цифры $\alpha_1, \alpha_2, \ldots, \alpha_n$ приближенного числа a называются верными в смысле ω, если абсолютная погрешность числа a не превосходит величины $\omega \cdot 10^{m-n+1}$, т.е.

$$\Delta a \le \omega \cdot 10^{m-n+1}. \qquad (1.9)$$

Определение числа верных значащих цифр позволяет решать и обратную задачу, т.е. определять, какие знаки в приближенном числе верные, а какие нет, если известна его абсолютная погрешность.

Пример 7. Определить, какие значащие цифры приближенного числа $a = 2{,}4483$ будут верными в узком (широком) смысле, если его абсолютная погрешность равна $\Delta a = 0{,}008$.

Решение. Следуя определению числа верных значащих цифр, для того чтобы $\alpha_1, \alpha_2, \ldots, \alpha_n$ были верными значащими цифрами числа a, необходимо потребовать выполнения неравенства:

$$\Delta a \le \omega \cdot 10^{m-n+1}, \text{ где } \omega = \frac{1}{2}(1),$$

которое в нашем примере имеет вид

$$0{,}008 \le \omega \cdot 10^{0-n+1} = \omega \cdot 10^{1-n}.$$

Решая неравенство при $\omega = \frac{1}{2}$, получим

$$0{,}008 \le 0{,}5 \cdot 10^{-n+1},$$
$$10^{-n+1} \ge 0{,}016, \quad n \le 2;$$

а при ω = 1 получим

$$0{,}008 \le 10^{-n+1}, \; m.e.\; n \le 3.$$

Таким образом, у числа a = 2,4483 три верные цифры в широком смысле и две – в узком. Остальные цифры приближенного числа 2,4483 не верны.

Приведенный способ определения числа верных значащих цифр по известной абсолютной погрешности, связанный с решением неравенства, можно заменить более простым правилом: число верных знаков в приближенном числе отсчитывается, начиная с первой значащей цифры числа до первой значащей цифры его абсолютной погрешности.

Пример 8. Определить количество верных значащих цифр в узком и широком смысле для числа a = 0,0076539, если $\Delta a = 0{,}000037$.

Решение. Напишем абсолютную погрешность над числом

$$\begin{array}{r|l} \Delta a = 0{,}0000 & 37 \\ a = 0{,}0076 & 539 \end{array}$$

Очевидно, что все значащие цифры, стоящие слева перед вертикальной чертой, проведенной перед первой значащей цифрой погрешности, будут всегда верными в широком смысле, так как число, стоящее за вертикальной чертой (в погрешности), всегда меньше единицы разряда, стоящего слева от черты, в данном случае

$$0{,}000037 < 0{,}0001$$

В нашем случае значащие цифры 7 и 6, стоящие слева от черты, будут верными и в узком смысле, так как величина погрешности 0.000037<0.00005 -половины единицы разряда десятитысячных, которому принадлежит последняя цифра 6. Если же для числа a = 0,0076539 $\Delta a = 0{,}0000503$, то по этому же правилу

$$\begin{array}{r|l} \Delta a = 0{,}0000 & 503 \\ a = 0{,}0076 & 539 \end{array}$$

число будет иметь две значащие цифры в широком смысле слова и только одну в узком, так как

$$0{,}503 \cdot 10^{-4} > 0.5 \cdot 10^{-4}.$$

На основании обобщенного определения абсолютная погрешность приближенного числа a связана с числом верных знаков соотношением (1.9)

$$\Delta a \le \omega \cdot 10^{m-n+1}.$$

В какой же зависимости от числа верных значащих цифр находится относительная погрешность?

Пусть приближенное число a,

$$|a| = \alpha_1 \cdot 10^m + \alpha_2 \cdot 10^{m-1} + \ldots + \alpha_n \cdot 10^{m-n+1} + \ldots \qquad (1.10)$$

имеет n верных значащих цифр в смысле определения 7.

Разделив обе части неравенства (1.9) на выражение (1.10), получим

$$\delta a = \frac{\Delta a}{|a|} = \frac{\omega \cdot 10^{m-n+1}}{\left|\alpha_1 \cdot 10^m + \ldots + \alpha_n \cdot 10^{m-n+1}\right|} \le \frac{\omega \cdot 10^{m-n+1}}{\left|\alpha_1 \cdot 10^m\right|} \le \frac{\omega}{\alpha_1} \cdot 10^{1-n},$$

т.е.

$$\delta a \le \frac{\omega}{\alpha_1} \cdot 10^{1-n}, \qquad (1.11)$$

где α_1 - первая значащая цифра числа, n – количество верных значащих цифр.

1.5. Общая формула теории погрешностей (погрешность вычисления значения функции)

Основная задача теории погрешностей заключается в следующем: известны погрешности некоторой системы величин, требуется определить погрешность данной функции от этих величин.

Пусть в некоторой области $G \in E_n$ задана дифференцируемая функция

$$y = f(x_1, x_2, \ldots, x_n) \qquad (1.12)$$

и известны абсолютные погрешности аргументов

$$|X_i - x_i| \le \Delta x_i;\ i = \overline{1, n}. \qquad (1.13)$$

Обозначим через

$$\varepsilon_i = X_i - x_i; \quad X_i = x_i + \varepsilon_i, \qquad (1.14)$$

тогда

$$|\varepsilon_i| \le \Delta x_i; \quad i = \overline{1, n}. \qquad (1.15)$$

Абсолютная погрешность функции выражается следующим образом:

$$\Delta y = \left| f(X_1,\ldots,X_n) - f(x_1,\ldots,x_n) \right| = \left| f(x_1+\varepsilon_1,\ldots,x_n+\varepsilon_n) - f(x_1,\ldots,x_n) \right|. \quad (1.16)$$

Согласно формуле Лагранжа

$$\Delta y = \left| \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(x_1+\theta\varepsilon_1,\ldots,x_n+\theta\varepsilon_n)\varepsilon_i \right| \le \sum_{i=1}^{n} \left| \frac{\partial}{\partial x_i} f(x_1+\theta\varepsilon_1,\ldots,x_n+\theta\varepsilon_n) \right| \Delta x_i; \quad (1.17)$$

$0 \le \theta \le 1.$

Отсюда

$$\Delta y \le \sum_{i=1}^{n} B_i \Delta x_i , \quad (1.18)$$

где

$$B_i = \max_{0\le\theta\le1} \left| \frac{\partial}{\partial x_i} f(x_1+\theta\varepsilon_1,\ldots,x_n+\theta\varepsilon_n) \right|, i = \overline{1,n}. \quad (1.19)$$

Когда погрешности аргументов Δx_i малы, величины B_i допустимо заменить на абсолютные значения частных производных функции $f(x_1,\ldots,x_n)$ в точке $(x_1,\ldots,x_n)$.

С учетом этого для абсолютной погрешности функции получится приближенное, но более простое выражение

$$\Delta y \le \sum_{i=1}^{n} \left| \frac{\partial}{\partial x_i} f(x_1,\ldots,x_n) \right| \Delta x_i . \quad (1.20)$$

Данное выражение для определения абсолютной погрешности функции носит название общей (или основной) формулы теории погрешностей.

Разделив обе части выражения (1.20) на $|y|$, получим выражение для относительной погрешности функции:

$$\delta y \le \sum_{i=1}^{n} \left| \frac{1}{y} \frac{\partial}{\partial x_i} f(x_1,\ldots,x_n) \right| \Delta x_i = \sum_{i=1}^{n} \left| \frac{\partial}{\partial x_i} ln\, f(x_1,\ldots,x_n) \right| \Delta x_i . \quad (1.21)$$

В случае функции одного аргумента выражения для погрешностей функции упрощаются. Действительно, если

$$y = f(x),$$

то

$$\Delta y \le \left| f'(x) \right| \Delta x,$$

$$\delta y \le \left| \frac{d}{dx} \ln f(x) \right| \Delta x.$$

В частности, для основных элементарных функций получаем следующие правила:

1. $y = x^{\alpha}, \quad \Delta y = \left|\alpha x^{\alpha-1}\right| \Delta x, \quad \delta y = |\alpha| \delta x;$

2. $y = a^x, \quad \Delta y = a^x \ln a \cdot \Delta x, \quad \delta y = \ln a \, \Delta x; \quad a > 1,$

   $y = e^x, \quad \Delta y = e^x \Delta x, \quad \delta y = \Delta x;$

3. $y = \lg_a x, \quad \Delta y = \dfrac{1}{x|\ln a|} \Delta x, \quad \delta y = \dfrac{\delta x}{|\ln a \cdot \lg_a x|},$

   $y = \ln x, \quad \Delta y = \dfrac{\Delta x}{x} = \delta x, \quad \delta y = \dfrac{\delta x}{|\ln x|};$

4. $y = \sin x, \quad \Delta y = |\cos x| \cdot \Delta x \le \Delta x,$

   $y = \cos x, \quad \Delta y = |\sin x| \cdot \Delta x \le \Delta x,$

   $y = \operatorname{tg} x, \quad \Delta y = (1 + \operatorname{tg}^2 x) \cdot \Delta x \ge \Delta x,$

   $y = \operatorname{ctg} x, \quad \Delta y = (1 + \operatorname{ctg}^2 x) \cdot \Delta x \ge \Delta x.$

1.6. Погрешность арифметических действий

1. Абсолютная погрешность алгебраической суммы нескольких приближенных чисел равна сумме абсолютных погрешностей слагаемых.

Действительно, если

$$y = \sum_{i=1}^{n} x_i \qquad (1.22)$$

то на основании общей формулы теории погрешностей (1.20)

$$\Delta y = \sum_{i=1}^{n} \Delta x_i \, .$$

Из полученной формулы следует: абсолютная погрешность алгебраической суммы не может быть меньше абсолютной погрешности наименее точного из слагаемых, так как увеличение точности за счет остальных слагаемых невозможно. Поэтому, чтобы не производить лишних вычислений, не следует сохранять лишние знаки и в более точных слагаемых.

Пример 9. Найти сумму приближенных чисел

$$y = 5{,}8 + 287{,}649 + 0{,}008064$$

и оценить погрешность результата, считая все знаки слагаемых верными в узком смысле.

Решение. Вычислим указанную сумму тремя способами

5,8	5,8	5,8
287,649	287,6	287,65
0,008064	0,0	0,01
293,457064	293,4	$293,46 \approx 293,5$

Очевидно, что только последний из приведенных способов сложения будет правильным, так как в числе 5,8 отброшенные знаки неизвестны, поэтому нет смысла получать результат с точностью до миллионных, которая ничем не гарантирована.

Второй способ также не верен, так как не использует большую точность двух других слагаемых. Поэтому будет правильным сохранить в остальных слагаемых один лишний десятичный знак, а после сложения результат округлить до десятых согласно с точностью числа, имеющего наибольшую абсолютную погрешность. При большом числе слагаемых вычисления лучше вести с двумя запасными десятичными знаками.

Погрешность полученной суммы будет равна сумме трех слагаемых:

I) сумма погрешностей исходных данных

$\Delta_1 = 0,05 + 0,0005 + 0,0000005 < 0,051$;

2) абсолютная величина суммы ошибок округления слагаемых

$\Delta_2 < |-0,001 - 0,002| = 0,003$;

3) погрешность округления результата

$\Delta_3 = 0,04$.

Следовательно,

$$\Delta y = \Delta_1 + \Delta_2 + \Delta_3 = 0,094,$$
$$y = 293,5 \pm 0,094,$$

т.е. y имеет 4 верных знака в широком смысле и 3 в узком.

2. Относительная погрешность суммы нескольких чисел одного и того же знака заключена между наименьшей и наибольшей из относительных погрешностей слагаемых:

$$\min_{1 \le k \le n} \delta x_k \le \delta y \le \max_{1 \le k \le n} \delta x_k . \qquad (1.23)$$

Действительно, если

$$y = x_1 + x_2 + \ldots + x_n, \quad x_i \ge 0 , \qquad (1.24)$$

то

$$\delta y = \frac{\Delta y}{y} = \frac{\Delta x_1 + \ldots + \Delta x_n}{x_1 + \ldots + x_n} = \frac{x_1 \delta x_1 + \ldots + x_n \delta x_n}{x_1 + \ldots + x_n} . \qquad (1.25)$$

Обозначив

$$\bar{\delta} = \max_k \delta x_k ; \quad \underline{\delta} = \min_k \delta x_k , \qquad (1.26)$$

получим

$$\delta y \geq \frac{x_1\underline{\delta}+\ldots+x_n\underline{\delta}}{x_1+\ldots+x_n}=\underline{\delta};$$
$$\delta y \leq \frac{x_1\overline{\delta}+\ldots+x_n\overline{\delta}}{x_1+\ldots+x_n}=\overline{\delta},\quad \underline{\delta}\leq\delta y\leq\overline{\delta}. \tag{1.27}$$

3. Относительная погрешность разности двух положительных чисел больше относительных погрешностей этих чисел, особенно, если эти числа близки между собой. Это приводит к потере точности при вычитании близких чисел, что следует учитывать при выборе вычислительной схемы.

Действительно, если

$$y = x_1 - x_2\ , \tag{1.28}$$

то

$$\Delta y = \Delta x_1 + \Delta x_2;\quad \delta y = \frac{\Delta x_1 + \Delta x_2}{|x_1 - x_2|}.$$

Пример 10. Найти разность двух чисел $a = 5{,}069$ *и* $в = 5{,}061;\ \ \Delta a = \Delta в = 0{,}0005$. Оценить погрешность результата.

Решение.

$$y = a - b = 0{,}008;\quad \Delta y = \Delta a + \Delta b = 0{,}001;\quad \delta y = \frac{10^{-3}}{8\cdot 10^{-3}} < 13\%.$$

Таким образом, результат имеет один верный знак в широком смысле, хотя сами числа имеют по четыре верных знака. Относительная погрешность разности y более чем в тысячу раз больше относительной погрешности самих чисел $\delta a = \delta b = \frac{0{,}5\cdot 10^{-3}}{5} = 0{,}01\%$.

4. При умножении и делении приближенных чисел складываются их относительные погрешности.

Действительно, если

$$u = \frac{x_1\cdot x_2\cdot\ldots\cdot x_n}{y_1\cdot y_2\cdot\ldots\cdot y_m}, \tag{1.30}$$

то

$$ln\, u = \sum_{i=1}^{n} x_i - \sum_{i=1}^{m} y_i$$

и на основании этого выражения получаем

$$\delta u = \sum_{i=1}^{n}\left|\frac{\partial\, ln\, u}{\partial x_i}\right|\Delta x_i + \sum_{i=1}^{m}\left|\frac{\partial\, ln\, u}{\partial y_i}\right|\Delta y_i = \sum_{i=1}^{n}\frac{\Delta x_i}{|x_i|} + \sum_{i=1}^{m}\frac{\Delta y_i}{|y_i|} = \sum_{i=1}^{n}\delta x_i + \sum_{i=1}^{m}\delta y_i \tag{1.31}$$

Из полученного выражения видно, что относительная погрешность произведения и частного не может быть меньше, чем относительная погрешность наименее точного из сомножителей,

следовательно, число верных знаков произведения не может быть больше наименьшего числа верных знаков сомножителей. Поэтому при перемножении нескольких чисел, имеющих разное число верных значащих цифр, выполняют следующие правила:

1) выделяют число, имеющее наименьшее число верных значащих цифр;

2) округляют оставшиеся сомножители, оставляя в них на одну значащую цифру больше, чем в выделенном сомножителе;

3) сохраняют в произведении столько значащих цифр, сколько верных значащих цифр имеет выделенный сомножитель.

1.7. Обратная задача теории погрешностей

Основная задача теории погрешностей заключалась в том, что по известным погрешностям аргументов находилась погрешность функции. На практике очень важное значение имеет и обратная задача: каковы должны быть погрешности аргументов, чтобы абсолютная погрешность функции не превышала заданной величины?

На основании общей формулы теории погрешностей имеем

$$\Delta y \le \sum_{i=1}^{n} \left| \frac{\partial y}{\partial x_i} \right| \Delta x_i \,. \qquad (1.32)$$

Задача отыскания допустимых значений абсолютной погрешности аргументов по известной абсолютной погрешности функции является математически неопределенной, так как в общем случае для определения n неизвестных Δx_i мы имеем одно уравнение.

Обратная задача теории погрешностей имеет однозначное решение только для функции одного аргумента. Действительно, на основании общей формулы теории погрешностей

$$y = f(x); \quad f'(x) \neq 0 \quad \forall x \in [a,b] \,,$$

следовательно,

$$\Delta x = \frac{\Delta y}{|f'(x)|} \,. \qquad (1.33)$$

1. Принцип равных влиянии. Согласно этому принципу предполагается, что все выражения

$$\left| \frac{\partial}{\partial x_i} f(x_1, \ldots, x_n) \right| \Delta x_i ; \quad i = \overline{1,n} \,,$$

одинаково влияют на образование общей абсолютной погрешности, т.е.

$$\left| \frac{\partial f}{\partial x_1} \right| \Delta x_1 = \ldots = \left| \frac{\partial f}{\partial x_n} \right| \Delta x_n \,.$$

Пусть нам задана абсолютная погрешность $\Delta y \le \varepsilon$. На основании общей формулы теории погрешностей можно написать

$$\sum_{i=1}^{n} \left| \frac{\partial y}{\partial x_i} \right| \Delta x_i \le \varepsilon .$$

Тогда

$$n \cdot \left| \frac{\partial y}{\partial x_i} \right| \Delta x_i \le \varepsilon ,$$

откуда

$$\Delta x_i \le \frac{\varepsilon}{n \cdot \left| \frac{\partial y}{\partial x_i} \right|}; \quad i = \overline{1, n} .$$

2. Принцип равных абсолютных погрешностей. Согласно этому принципу предполагается, что

$$\Delta x_1 = \Delta x_2 = \ldots = \Delta x_n .$$

Тогда из общей формулы теории погрешностей будем иметь

$$\Delta y \le \Delta x_i \sum_{i=1}^{n} \left| \frac{\partial y}{\partial x_i} \right| \le \varepsilon$$

или

$$\Delta x_i \le \frac{\varepsilon}{\sum_{i=1}^{n} \left| \frac{\partial y}{\partial x_i} \right|}; \quad i = \overline{1, n} .$$

3. Принцип равных относительных погрешностей. Согласно этому принципу предполагается, что

$$\delta x_1 = \delta x_2 = \ldots = \delta x_n .$$

По определению $\delta x_i = \frac{\Delta x_i}{|x_i|}$, тогда $\Delta x_i = \delta x_i \cdot |x_i|$.

Подставляя это выражение в общую формулу, получим

$$\Delta y = \sum_{i=1}^{n} \left| \frac{\partial y}{\partial x_i} \right| |x_i| \, \delta x_i \le \varepsilon ,$$

откуда

$$\delta x_i \le \frac{\varepsilon}{\sum_{i=1}^{n} \left| \frac{\partial y}{\partial x_i} \cdot x_i \right|}; \quad \Delta x_i \le \frac{\varepsilon \cdot |x_i|}{\sum_{i=1}^{n} \left| \frac{\partial y}{\partial x_i} \cdot x_i \right|}; \quad i = \overline{1, n} .$$

2. ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

2.1. Отделение корней

Рассмотрим некоторую функцию *f(x)*.

Определение. Всякое число ξ обращающее функцию в нуль, т.е. такое, что $f(\xi)=0$, называется корнем (нулем) функции или корнем уравнения

$$f(x)=0. \qquad (2.1)$$

Приближенное вычисление корня, как правило, распадается на две задачи:

1 отделение корней, т.е. определение интервалов, в каждом из которых содержится только один корень уравнения;

2 уточнение корня, т.е. вычисление его с заданной степенью точности.

При отделении корней уравнения общего вида (2.1) часто используется известная из курса математического анализа теорема Больцано - Коши:

пусть функция *f(x)* непрерывна на отрезке $[a,b]$ и на концах отрезка принимает значения разных знаков, т.е.

$f(a)\cdot f(b)<0$. Тогда существует такая точка ξ, принадлежащая интервалу (a,b), в которой функция обращается в нуль. Заметим, что корень будет единственным, если $f'(x)$ (или $f''(x)$) существует и сохраняет знак на рассматриваемом отрезке.

Остановимся более подробно на алгебраических уравнениях

$$P_n(x)=a_0x^n+a_1x^{n-1}+\ldots+a_{n-1}x+a_n. \qquad (2.2)$$

Верхнюю границу модулей корней уравнения (2.2) дает следующая теорема.

Пусть $A=\max\limits_{1\le i\le n}|a_i|$. Тогда любой корень ξ уравнения (2.2) удовлетворяет условию

$$|\xi|<1+\frac{A}{|a_0|}=R. \qquad (2.3)$$

Допустим, что существует корень α уравнения (2.2), не удовлетворяющий условию (2.3), т.е.

$$|\alpha|\ge 1+\frac{A}{|a_0|}. \qquad (2.4)$$

Из (2.4) следует, что

$|\alpha|>1$.

Тогда

$$|P_n(\alpha)| = |a_0\alpha^n + a_1\alpha^{n-1} + \ldots + \alpha_n| \geq |a_0||\alpha|^n - |a_1\alpha^{n-1} + \ldots + a_n| \geq$$

$$\geq |a_0||\alpha|^n - A\frac{|\alpha|^n - 1}{|\alpha| - 1} > |a_0||\alpha|^n - \frac{A|\alpha|^n}{|\alpha| - 1} = \frac{|a_0||\alpha|^n}{|\alpha| - 1}\left(|\alpha| - 1 - \frac{A}{|a_0|}\right).$$

Согласно (2.4)

$$\left(|\alpha| - 1 - \frac{A}{|a_0|}\right) \geq 0 \text{ и } |P_n(\alpha)| > 0,$$

что противоречит предположению о том, что α - корень уравнения (2.2).

2.2. Метод половинного деления

Пусть функция *f(x)* непрерывна на отрезке *[a,b]* и имеет на его концах разные по знаку значения. Задача состоит в том, чтобы вычислить корень уравнения (2.1.), принадлежащий отрезку *[a,b]* с заданной степенью точности ε , т.е. найти такое приближенное значение корня X_n, (*n* - номер итерации), что

$$|\xi - x_n| \leq \Delta x_n \leq \varepsilon. \qquad (2.5)$$

В методе половинного деления за приближенное значение корня принимается середина отрезка

$$x_1 = \frac{a+b}{2}.$$

При этом очевидно, что

$$|\xi - x_1| \leq \frac{b-a}{2} = \Delta x_1.$$

Затем определяется знак $f(x_1)$ и для дальнейшего деления пополам выбирается тот из двух отрезков $[a, x_1]$ *или* $[x_1, b]$, на концах которого функция *f(x)* имеет разные по знаку значения. Расчет продолжается до тех пор, пока не выполнится условие (2.5) либо условие

$$f(x_n - \varepsilon) \cdot f(x_n + \varepsilon) < 0. \qquad (2.6)$$

2.3. Метод хорд (секущих)

Предполагая опять, что *f(x)* непрерывна на отрезке *[a,b]* и имеет разные знаки на его концах, получим формулы для приближенного вычисления корня уравнения (2.1), учитывающие не только знаки *f(x)*, но и ее значения. Для этого соединим точки $A(a; f(a))$ *и* $B(b; f(b))$ хордой АВ (рис. 2.1). Точку пересечения x_1 этой хорды с осью абсцисс примем за приближенное значение корня.

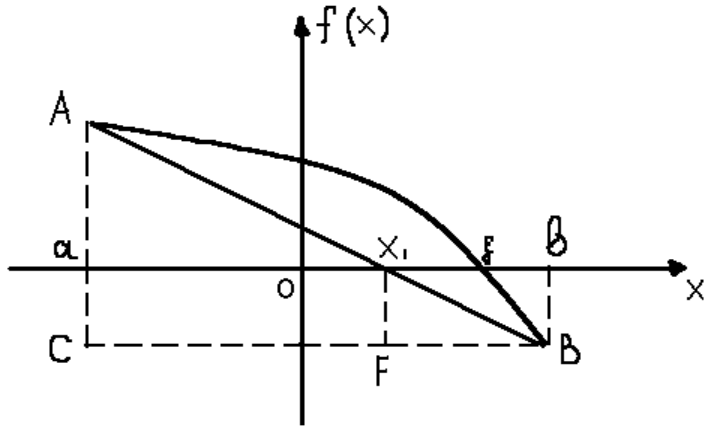

Рис.2.1.

Из подобия треугольников ABC и x_1BF, ABC и Ax_1a следуют соотношения

$$\frac{b-a}{b-x_1}=\frac{f(a)-f(b)}{-f(b)};\quad \frac{b-a}{x_1-a}=\frac{f(a)-f(b)}{f(a)},$$

откуда получим соответственно две формулы метода хорд для x_1:

$$x_1=b-\frac{f(b)}{f(b)-f(a)}(b-a); \qquad (2.7)$$

$$x_1=a-\frac{f(a)}{f(b)-f(a)}(b-a). \qquad (2.8)$$

Выбрав одну из формул, (2.7) или (2.8), вычислим x_1, определим знак $f(x_1)$ и, как в методе половинного деления, для дальнейших вычислений выберем тот из отрезков $[a,x_1]$ *или* $[x_1,b]$, на концах которого функция $f(x)$ имеет разные по знаку значения. Оценкой абсолютной погрешности приближенного значения x_1 здесь может служить величина

$$\Delta x_1=\max(b-x_1,x_1-a),$$

Получим еще одну оценку абсолютной погрешности Δx_1 при дополнительном предположении, что на отрезке *[a, b] f(x)* дифференцируема и

$$|f'(x)|\ge m_1>0.$$

$$|f(x_1)|=|f(\xi)-f(x_1)|=|f'(\alpha)||\xi-x_1|,$$

где α - точка, расположенная между корнем ξ и x_1.

Отсюда

$$|\xi-x_1|=\frac{|f(x_1)|}{|f'(\alpha)|}\le\frac{|f(x_1)|}{m_1}=\Delta x_1. \qquad (2.9)$$

Если уравнение (2.1) имеет на отрезке *[a, b]* несколько корней, то метод хорд, как и метод половинного деления, вычислит с точностью до

ε один из них. Если же функция *f(x)* имеет на отрезке *[a, b]* непрерывную первую и вторую производные, сохраняющие свои знаки, то можно показать [5], что последовательность приближенных значений метода хорд, построенная по формуле

$$x_{n+1} = x_n - \frac{f(x_n)}{f(c) - f(x_n)}(c - x_n), \qquad (2.10)$$

где с - один из концов отрезка *[a, b]*, удовлетворяющий условию

$$f(c) \cdot f''(c) > 0,$$

а x_0 - противоположный конец отрезка, сходится к единственному на этом отрезке корню уравнения (2.1) монотонно.

2.4. Метод касательных (метод Ньютона)

Пусть искомый корень $x = \xi$ уравнения (2.1) принадлежит отрезку *[a, b]*, $x_0 \in [a,b]$. Представим *f(ξ)* с помощью разложения функции *f(x)* в ряд Тейлора в окрестности точки x_0

$$0 \equiv f(\xi) = f(x_0) + f'(x_0)(\xi - x_0) + \frac{f''(\alpha)}{2}(\xi - x_0)^2, \quad (2.11)$$

где α - точка, находящаяся между точками ξ и x_1. Пренебрегая в (2.11) остаточным членом, найдем приближенное значение x_1 корня ξ:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}. \qquad (2.12)$$

Подставив в правую часть (2.12) вместо x_0 полученное значение x_1, получим x_2 и т.д. Докажем, что последовательность

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad (2.13)$$

монотонно сходится к единственному на отрезке корню ξ уравнения (2.1), если:

1) $f(a) \cdot f(b) < 0$;

2) $f'(x), f''(x)$ непрерывны, отличны от нуля и сохраняют свои знаки на *[a, b];*

3) начальное приближение x_0 удовлетворяет условию: $f(x_0) \cdot f''(x_0) > 0$. Существование и единственность корня следуют из условий I и 2.

Докажем сходимость последовательности (2.13) для случая, когда (рис. 2.2)

$$f(a) < 0, f(b) > 0, f'(x) > 0, f''(x) > 0.$$

В остальных случаях доказательство ведется аналогичным образом.

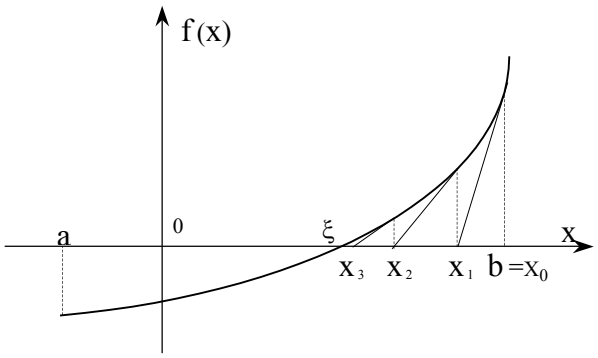

Рис. 2.2.

За начальное приближение удобно взять один из концов отрезка *[a, b],*. В данном случае $x_0 = b$, так как $f(b) \cdot f''(b) > 0$. Методом индукции докажем, что последовательность x_n, построенная по формуле (2.13), ограничена снизу точным корнем ξ .Действительно, $x_0 = b > \xi$. Допустим, что все приближения $x_j > \xi, j = 1,2,\ldots,n$ и докажем, что $x_{n+1} > \xi$. Разложим функцию *f(x)* в ряд Тейлора в окрестности точки x_n:

$$f(x) = f(x_n) + f'(x_n)(x - x_n) + \frac{f''(\alpha)}{2}(x - x_n)^2, \qquad (2.14)$$

где точка α расположена между x и x_n. Подставим в (2.14) x = ξ

$$0 \equiv f(\xi) = f(x_n) + f'(x_n)(\xi - x_n) + \frac{f''(\alpha)}{2}(\xi - x_n)^2.$$

Поскольку $f''(x) > 0$ на *[a, b],*

$$f(x_n) + f'(x_n)(\xi - x_n) < 0,$$

откуда

$$\xi < x_n - \frac{f(x_n)}{f'(x_n)} = x_{n+1},$$

т.е. ограниченность снизу последовательности x_n доказана. Отсюда $f(x_n) > 0, \quad n = 1,2,\ldots$ т.е. $x_{n+1} > x_n$ – последовательность x_n, монотонно убывает и, значит, имеет предел:

$$\lim_{n\to\infty} x_n = A.$$

Перейдем к пределу при $n \to \infty$ в равенстве (2.13)

$$A = A - \frac{f(A)}{f'(A)},$$

$f(A) = 0, \quad m.e. \quad A = \xi$

Абсолютную погрешность приближения x_n, полученного методом касательных, можно оценить формулой (2.9). Преобразуем эту оценку с помощью (2.13). Представим f(x_n)разложением в ряд Тейлора в окрестности точки x_{n-1}:

$$f(x_n) = f(x_{n-1}) + f'(x_{n-1})(x_n - x_{n-1}) + \frac{f''(\alpha)}{2}(x_n - x_{n-1})^2,$$

где α - точка, расположенная между x_{n-1} и x_n.

Согласно (2.13)

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} \text{ или } f(x_{n-1}) + f'(x_{n-1})(x_n - x_{n-1}) = 0,$$

откуда

$$f(x_{n-1}) = \frac{f''(\alpha)}{2}(x_n - x_{n-1})^2.$$

Обозначим $M_2 = \max\limits_{[a,b]} |f''(x)|$.

Получим следующую оценку абсолютной погрешности величины x_n:

$$|\xi - x_n| \le \frac{|f(x_n)|}{m_1} \le \frac{M_2}{2m_1}(x_n - x_{n-1})^2 = \Delta x_n.$$

На свойстве монотонности последовательностей метода хорд (2.10) и метода касательных (2.13) основан комбинированный метод, заключающийся в одновременном использовании этих двух методов:

$$\bar{x}_{n+1} = \bar{x}_n - \frac{f(\bar{x}_n)}{f'(x_n)}; \quad \bar{\bar{x}}_{n+1} = \bar{\bar{x}}_n - \frac{f(\bar{\bar{x}}_n)(\bar{x}_{n+1} - \bar{\bar{x}}_n)}{f(\bar{x}_{n+1}) - f(\bar{\bar{x}}_n)};$$

$$x_{n+1} = \frac{\bar{x}_{n+1} + \bar{\bar{x}}_{n+1}}{2}; \quad \Delta x_{n+1} = \frac{\left|\bar{x}_{n+1} - \bar{\bar{x}}_{n+1}\right|}{2}; \quad n = 0,1,2,\ldots$$

Если в формуле (2.13) положить

$$f'(x_n) \approx \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}},$$

то формула (2.13) примем вид

$$x_{n+1} = x_n - \frac{f(x_n)}{f(x_n) - f(x_{n-1})}(x_n - x_{n-1})$$

Эта модификация метода касательных носит название двухшагового метода хорд.

2.5. Метод итераций

Преобразуем уравнение (2.1) к эквивалентному виду

$$x = \varphi(x). \qquad (2.15)$$

Выбрав в качестве начального приближения точку $x_0 \in [a,b]$, построим последовательность

$$x_{n+1} = \varphi(x_n), \quad n = 0,1,2,\ldots \qquad (2.16)$$

Докажем, что эта последовательность при любом $x_0 \in [a,b]$ сходится к единственному на отрезке *[a, b]* корню уравнения (2.1), если:

1) функция $\varphi(x)$ определена и дифференцируема на отрезке *[a, b];*

2) все ее значения принадлежат этому отрезку при $x \in [a,b]$;

3) существует такое число 0<q<1, что $|\varphi'(x)| \le q$ при $x \in [a,b]$.

Рассмотрим ряд

$$x_0 + (x_1 - x_0) + (x_2 - x_1) + \ldots + (x_n - x_{n-1}) + \ldots, \qquad (2.17)$$

где x_n определено формулой (2.16). Частичная сумма этого ряда

$$S_n = x_n.$$

Оценим по модулю каждый член ряда

$$|x_{i+1} - x_i| = |\varphi(x_i) - \varphi(x_{i-1})| = |\varphi'(\alpha)||x_i - x_{i-1}| \le q|x_i - x_{i-1}|,$$

где точка α - расположена между x_{i-1} и x_i.

Имеем:

$$|x_2 - x_1| \le |x_1 - x_0| q;$$
$$|x_3 - x_2| \le q|x_2 - x_1| \le q^2|x_1 - x_0|;$$
$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$
$$|x_{n+1} - x_n| \le q^n|x_1 - x_0|.$$

Следовательно, ряд (2.17) сходится абсолютно, т.е. существует

$$\lim_{n\to\infty} S_n = \xi,$$

откуда следует сходимость последовательности (2.16)

$$\lim_{n\to\infty} x_n = \xi.$$

Перейдем к пределу в равенстве (2.16):

$$\xi = \varphi(\xi),$$

т.е. ξ - является корнем уравнения (2.15) и эквивалентного ему уравнения (2.1). Докажем единственность ξ. Пусть существуют два корня уравнения (2.15): ξ *и* $\xi_1 \in [a,b]$

$$\xi = \varphi(\xi); \quad \xi_1 = \varphi(\xi_1);$$
$$|\xi - \xi_1| = |\varphi(\xi) - \varphi(\xi_1)| = |\varphi'(\alpha)| \cdot |\xi - \xi_1|,$$

где точка α расположена между ξ и ξ_1, т.е. $\alpha \in [a,b]$ Преобразуем это равенство

$$|\xi - \xi_1|\left(1 - |\varphi'(\alpha)|\right) = 0.$$

Но $|\varphi'(x)| < 1$ на *[a, b]*, значит,

$$\xi - \xi_1 = 0, \quad m.e. \quad \xi = \xi_1.$$

Оценим абсолютную погрешность приближения x_n, полученного методом итераций

$$|\xi - x_n| = |\varphi(\xi) - \varphi(x_{n-1})| = |\varphi'(\alpha)||\xi - x_{n-1}| \le q|\xi - x_n + x_n - x_{n-1}| \le$$
$$\le q(|\xi - x_n| + |x_n - x_{n-1}|),$$

$$|\xi - x_n| \le \frac{q}{1-q}|x_n - x_{n-1}| = \Delta x_n .$$

Укажем теперь достаточно общий прием построения функции $\varphi(x)$, для которой будет обеспечено выполнение условий сходимости итерационного процесса (2.16). Пусть на отрезке *[a, b]* существует *f'(x)* и сохраняет знак так, что

$$0 < m_1 \le f'(x) \le M_1$$

(мы приняли здесь, что $f'(x) > 0$, в противном случае рассматривается функция – *f(x)*). Умножив уравнение (2.1) на число λ и вычтя результат из тождества $x \equiv x$, получим

$$x = x - \lambda f(x) \equiv \varphi(x) .$$

Выберем λ так, чтобы

$$|\varphi'(x)| = |1 - \lambda f'(x)| < 1 .$$

Отсюда

$$-1 < 1 - \lambda f'(x) < 1 .$$

Из правого неравенства получим $\lambda > 0$, а из левого

$$\lambda < \frac{2}{f'(x)}, \quad x \in [a,b] .$$

Обычно полагают $\lambda = \frac{2}{M_1 + m_1}$. Тогда

$$|\varphi'(x)| \le \frac{M_1 - m_1}{M_1 + m_1} = q < 1 .$$

3. ЧИСЛЕННЫЕ МЕТОДЫ ЛИНЕЙНОЙ АЛГЕБРЫ

К численным методам линейной алгебры относятся численные методы решения систем линейных алгебраических уравнений, обращения матриц, вычисления определителей и нахождения собственных векторов матриц.

Методы решения систем линейных алгебраических уравнений делятся на две группы. К первой группе принадлежат прямые (или точные) методы, которые позволяют найти точное решение системы за конечное число арифметических действий. Отметим, что вследствие погрешностей округления при решении задач на ЭВМ прямые методы на самом деле не приводят к точному решению и назвать их точными можно, лишь отвлекаясь от вычислительной погрешности. Наиболее распространенными среди прямых методов являются метод Гаусса и метод прогонки.

Вторую группу составляют итерационные методы (их называют также методами последовательных приближений), которые состоят в том, что точное решение $\bar{x}^*$ системы находится как предел последовательных приближений $\bar{x}^{(n)}$, где n - номер итерации.

3.1. Метод Гаусса

Рассмотрим систему линейных алгебраических уравнений

$$A\bar{x} = \bar{b}. \qquad (3.1)$$

Будем предполагать, что определитель матрицы А отличен от нуля. Метод Гаусса основан на приведении матрицы А к треугольному виду. Это достигается последовательным исключением неизвестных из уравнений системы. Сначала с помощью первого уравнения исключается x_1 из всех последующих уравнений системы. Затем с помощью второго уравнения исключается x_2 из третьего и всех последующих уравнений. Этот процесс, называемый прямым ходом метода Гаусса, продолжается до тех пор, пока в левой части последнего (n-го) уравнения не останется лишь один член с неизвестным $x_п$, т.е. матрица системы будет приведена к треугольному виду.

Обратный ход метода Гаусса состоит в последовательном вычислении искомых неизвестных, т.е. решая последнее уравнение, находим значение $x_п$; далее, используя это значение, из предыдущего уравнения вычисляем $x_{п-1}$ и т.д. Последним найдем x_1 из первого уравнения.

При реализации на ЭВМ прямого хода метода Гаусса нет необходимости действовать с переменными $x_1, x_2, \ldots, x_n$. Достаточно указать алгоритм, согласно которому исходная матрица преобразуется к треугольному виду, и указать соответствующее преобразование правых

частей системы. Пусть осуществлены первые (к-1) шагов, т.е. уже исключены переменные $x_1, x_2, \ldots, x_{\kappa-1}$. Тогда имеем систему

$$\begin{aligned}
&a_{11}x_1 + a_{12}x_2 + \ldots + a_{1k}x_k + \ldots + a_{1n}x_n = b_1;\\
&\qquad a_{22}^{(1)}x_2 + \ldots + a_{2k}^{(1)}x_k + \ldots + a_{2n}^{(1)}x_n = b_2^{(1)};\\
&\ldots \quad \ldots \qquad \ldots \quad \ldots \qquad \ldots \quad \ldots \qquad \ldots \quad \ldots\\
&a_{k-1,k-1}^{(k-2)}x_{k-1} + a_{k-1,k}^{(k-2)}x_k + \ldots + a_{k-1,n}^{(k-2)}x_n = b_{k-1}^{(k-2)};\\
&\ldots \quad \ldots \qquad \ldots \quad \ldots \qquad \ldots \quad \ldots \qquad \ldots \quad \ldots\\
&\qquad a_{nk}^{(k-1)}x_k + \ldots + a_{nn}^{(k-1)}x_n = b_n^{(k-1)},
\end{aligned} \tag{3.2}$$

где $a_{11}, a_{12}, \ldots, a_{1n}$ - коэффициенты первой строки матрицы А; $a_{ij}^{(m)}$ - коэффициент i -го уравнения при j переменной, полученный в результате преобразований системы на m –м шаге. Предположим, что в к - ом уравнении коэффициент $a_{kk}^{(k-1)} \neq 0$. Умножим к -ое уравнение системы на $\dfrac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}$ и вычтем полученное соотношение из i -го уравнения системы (3.2), где $i = k+1, k+2, \ldots, n$.

В результате последняя группа уравнений системы (3.2) примет вид:

$$\begin{aligned}
&a_{kk}^{(k-1)}x_k + a_{k,k+1}^{(k-1)}x_{k+1} + \ldots + a_{kn}^{(k-1)}x_n = b_k^{(k-1)};\\
&\qquad a_{k+1,k+1}^{(k)}x_{k+1} + \ldots + a_{k+1,n}^{(k)}x_n = b_{k+1}^{(k)};\\
&\qquad\quad \ldots \quad \ldots \qquad \ldots \quad \ldots \qquad \ldots \quad \ldots\\
&\qquad\qquad a_{n,k+1}^{(k)}x_{k+1} + \ldots + a_{nn}^{(k)}x_n = b_n^{(k)},
\end{aligned}$$

где

$$\begin{aligned}
&a_{ij}^{(k)} = a_{ij}^{(k-1)} - \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}a_{kj}^{(k-1)};\\
&b_i^{(k)} = b_i^{(k-1)} - \frac{a_{ik}^{(k-1)}}{a_{kk}^{(k-1)}}b_k^{(k-1)}; \quad i = k+1, \ldots, n.
\end{aligned} \tag{3.3}$$

Коэффициенты $a_{kj}^{(k-1)}$, $j = k, k+1, \ldots, n$ и правая часть $b_k^{(k-1)}$ при каждом $k = 1,2, \ldots, n-1$ хранятся в памяти ЭВМ и используются при осуществлении обратного хода.

Обратный ход, как уже указывалось, заключается в вычислении неизвестных $x_n, x_{n-1}, \ldots, x_1$. Последнее уравнение будет иметь вид

$$a_{nn}^{(n-1)}x_n = b_n^{(n-1)}.$$

Откуда

$$x_n = \frac{b_n^{(n-1)}}{a_{nn}^{(n-1)}}.$$

Общая формула обратного хода для вычисления переменной $x_к$, имеет вид

$$x_k = \frac{b_k^{(k-1)} - \sum_{j=k+1}^{n} a_{kj}^{(k-1)} x_j}{a_{kk}^{(k-1)}}, \quad k = n-1,\ldots,1.$$

Основным ограничением метода является предположение о том, что все элементы $a_{kk}^{(k-1)}$, на которые производится деление, отличны от нуля. Число $a_{kk}^{(k-1)}$ называется ведущим (или направляющим) элементом на к - м шаге исключения. Даже если какой-то ведущий элемент не равен нулю, а просто близок к нему, в процессе вычислений может происходить сильное накопление погрешностей. Избежать указанных трудностей позволяет метод Гаусса с выбором главного элемента. Основная идея метода состоит в том, чтобы на очередном шаге исключать не следующую по номеру переменную, а ту, коэффициент при которой является наибольшим по модулю.

Пусть на к - ом шаге имеем систему (3.2). Сначала добиваемся выполнения условия

$$\left|a_{kk}^{(k-1)}\right| \geq \left|a_{ij}^{(k-1)}\right|, \quad i, j = k, k+1,\ldots,n,$$

путем перестановки двух уравнений системы (3.2), а также двух столбцов неизвестных со своими коэффициентами и соответствующей перенумерацией коэффициентов и неизвестных. При этом если переставляются столбцы, то соответствующая перестановка и перенумерация производятся и в уравнениях, которые на к - ом шаге не преобразуются, т.е. при $i = 1,2,\ldots,k-1$.

Найденный максимальный по модулю элемент (если $det\, A \neq 0$, то этот главный элемент всегда отличен от нуля) называется к - м главным элементом. Затем осуществляются преобразования к - го шага по формулам (3.3).

В большинстве существующих стандартных программ одновременно с решением системы линейных алгебраических уравнений (3.1) вычисляется определитель матрицы А. Определитель полученной в результате прямого хода треугольной матрицы равен произведению её диагональных элементов и отличается от определителя $|A|$ лишь знаком, поскольку в процессе преобразований осуществлялась перестановка строк в столбцов. Поэтому

$$det\, A = (-1)^s a_{11} \cdot a_{22}^{(1)} \cdot a_{33}^{(2)} \cdot \ldots \cdot a_{nn}^{(n-1)},$$

где S - суммарное число перестановок строк и столбцов.

Если матрица заданной системы вырожденная, то перед исключением некоторой неизвестной главный элемент окажется равным нулю. Этим самым и обнаружится, что определитель заданной системы равен нулю.

Метод Гаусса используется также для вычисления обратной матрицы. Пусть матрица A^{-1} с элементами x_{ij} является обратной к матрице А. Тогда имеем матричное уравнение

$$A \cdot A^{-I} = E,$$

где Е - единичная матрица. Отсюда каждый j – $й$ столбец $\overline{x_j} = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \dots \\ x_{nj} \end{pmatrix}$ матрицы A^{-1} удовлетворяет уравнению

$$A\bar{x}_j = \overline{e_j}, \qquad (3.4)$$

где $\overline{e_j}$ - вектор-столбец, у которого j - я компонента равна единице, а остальные компоненты равны нулю. Таким образом, вычисление обратной матрицы сводится к решению n систем вида (3.4) для $j = 1,2,\dots,n$, отличающихся только правыми частями.

Как уже указывалось выше, решение системы (3.1), полученное методом Гаусса, может быть искажено вычислительной погрешностью, являющейся следствием округлений при вычислениях. Рассмотрим способ уточнения решения.

Пусть найдено решение $\bar{x}^{(I)}$. Определим для него вектор невязок

$$\bar{b}^{(I)} = \bar{b} - A\bar{x}^{(I)}. \qquad (3.5)$$

Обозначив вектор уточнений $\bar{\Delta}^{(I)} = \bar{x} - \bar{x}^{(I)}$ и вычтя из (3.1) уравнение (3.5), получим, что $\bar{\Delta}^{(I)}$ удовлетворяет системе

$$A\bar{\Delta}^{(I)} = \bar{b}^{(I)}. \qquad (3.6)$$

Решим эту систему и положим $\bar{x}^{(2)} = \bar{x}^{(I)} + \bar{\Delta}^{(I)}$. Если точность нового приближения представляется неудовлетворительной, то повторяем эту операцию. При решении системы (3.6) над компонентами правой части производятся те же линейные операции, что и над компонентами правой части при решении системы (3.1). Поэтому при вычислениях на машине с плавающей запятой естественно ожидать, что относительные погрешности решений этих систем будут одинаковы. Поскольку погрешности округлений обычно малы, $\left|b_j^{(I)}\right| << \left|b_j\right|, \quad j = 1,\dots,n$. Тогда, $\left|\Delta_j^{(I)}\right| << \left|x_j^{(I)}\right|, \quad j = 1,\dots,n$ и,

по-видимому, решение системы (3.6) определяется с существенно меньшей абсолютной погрешностью, чем решение системы (3.1). Таким образом, применение описанного приема приводит к повышению точности приближенного решения.

3.2. Метод прогонки

Метод прогонки применяется для решения систем специального вида, матрица которых является трехдиагональной:

$$\begin{aligned}
-c_0x_0 + b_0x_1 &= f_0;\\
a_1x_0 - c_1x_1 + b_1x_2 &= f_1;\\
a_2x_1 - c_2x_2 + b_2x_3 &= f_2;\\
\dots\quad\dots\quad\dots\quad\dots\quad\dots\quad\dots\quad\dots\quad\dots&\\
a_{n-1}x_{n-2} - c_{n-1}x_{n-1} + b_{n-1}x_n &= f_{n-1};\\
a_nx_{n-1} - c_nx_n &= f_n.
\end{aligned} \qquad (3.7)$$

Такие системы обычно возникают при численном решении краевых задач для дифференциальных уравнений, интерполировании сплайнами и моделировании некоторых процессов.

Выразим из первого уравнения системы (3.7) переменную x_0, а из последнего - переменную x_n, предполагая, что $c_0 \neq 0$, $c_n \neq 0$, и запишем эту систему в следующем виде:

$$a_ix_{i-1} - c_ix_i + b_ix_{i+1} = f_i, \quad i = 1,2,\dots,n-1; \qquad (3.8)$$

$$x_0 = \alpha_0x_1 + \beta_0; \quad x_n = \alpha_nx_{n-1} + \beta_n; \quad \alpha_0 = \frac{b_0}{c_0}; \quad \alpha_n = \frac{a_n}{c_n};$$

$$\beta_0 = -\frac{f_0}{c_0}; \quad \beta_n = -\frac{f_n}{c_n}. \qquad (3.9)$$

Уравнения (3.8) в совокупности обычно называются разностным уравнением второго порядка или точечным разностным уравнением, а уравнения (3.9) - краевыми условиями для разностного уравнения (3.8). Система же (3.8)-(3.9) в целом называется разностной краевой задачей.

Выведем расчетные формулы метода прогонки для решения системы (3.8)-(3.9). Подставим первое краевое условие $x_0 = \alpha_0x_1 + \beta_0$ в первое уравнение (3.8). Получим уравнение

$$a_1(\alpha_0x_1 + \beta_0) - c_1x_1 + b_1x_2 = f_1 \quad \textit{или}$$

$$x_1 = \alpha_1x_2 + \beta_1, \qquad (3.10)$$

$$\textit{где} \quad \alpha_1 = \frac{b_1}{c_1 - a_1\alpha_0}; \quad \beta_1 = \frac{a_1\beta_0 - f_1}{c_1 - a_1\alpha_0}.$$

Найденное выражение (3.10) для x_1 подставим в следующее уравнение (3.8) и получим уравнение, связывающее переменные x_2 и x_3 и т.д.

Допустим, что уже найдено соотношение

$$x_{k-1} = \alpha_{k-1} \cdot x_k + \beta_{k-1}, \quad k = 1,2,\ldots,n-1. \qquad (3.11)$$

Подставим (3.11) в *k*-е уравнение (3.8)

$$a_k(\alpha_{k-1}x_k + \beta_{k-1}) - c_k x_k + b_k x_{k+1} = f_k .$$

Разрешим это уравнение относительно x_k:

$$x_k = \alpha_k \cdot x_{k+1} + \beta_k \qquad (3.12)$$

$$\text{где} \quad \alpha_k = \frac{b_k}{c_k - a_k\alpha_{k-1}}; \quad \beta_k = \frac{a_k\alpha_{k-1} - f_k}{c_k - a_k\alpha_{k-1}}. \qquad (3.13)$$

Таким образом, коэффициенты уравнений (3.12), связывающие соседние переменные x_k *и* x_{k+1}, $k = 1,2,...,n-1$, можно определить из рекуррентных соотношений (3.13), поскольку α_0 *и* β_0 заданы в (3.9).

Подставив во второе краевое условие (3.9) выражение для x_{n-1}, вытекающее из формулы (3.12) при $k = n-1$, получим

$$x_n = \alpha_n(\alpha_{n-1}x_n + \beta_{n+1}) + \beta_n, \qquad (3.14)$$

где α_n *и* β_n - заданные в (3.9) коэффициенты, а α_{n-1} *и* β_{n-1} вычислены по формулам (3.13). Из уравнения (3.14) вычисляем x_n:

$$x_n = \frac{\beta_n + \beta_{n+1}\alpha_n}{1 - \alpha_n\alpha_{n-1}}. \qquad (3.15)$$

Затем по формуле (3.12) в обратном порядке вычисляем остальные неизвестные $x_{n-1}, x_{n-2}, \ldots, x_0$. Формула (3.12) при $k = 0$ совпадает с первым краевым условием (3.9).

Процесс вычисления коэффициентов $\alpha_k, \beta_k, k = 1,2,\ldots,n-1$ по формулам (3.13) называется прямой прогонкой, а вычисление неизвестных x_i, $i = n, n-1, \ldots, 0$ по формулам (3.15), (3.12) - обратной прогонкой.

Метод прогонки можно применять, если знаменатели формул (3.15), (3.12) не обращаются в нуль. Докажем, что для возможности применения метода прогонки достаточно потребовать, чтобы коэффициенты системы (3.8)-(3.9) удовлетворяли условиям

$$a_i \neq 0, b_i \neq 0, |c_i| \geq |a_i| + |b_i|, \quad i = 1,2,\ldots,n-1, \qquad (3.16)$$

$$|\alpha_0| \leq 1, \quad |\alpha_n| < 1. \qquad (3.17)$$

Сначала докажем индукцией, что при условиях (3.16), (3.17)

$$|\alpha_i| \leq 1, \quad i = 1,2,\ldots,n-1.$$

По первому условию (3.17)

$$|\alpha_0| \leq 1.$$

Предположим, что все

$$|\alpha_i| \leq 1, \quad i = 1,2,\ldots,\kappa,$$

и докажем, что

$$|\alpha_{\kappa+1}| \leq 1.$$

Из оценок

$$|c_i - a_i\alpha_{i-1}| \geq |c_i| - |a_i||\alpha_{i-1}| \geq |c_i| - |a_i|$$

и условий (3.16) получаем

$$|c_i - a_i\alpha_{i-1}| \geq |b_i| > 0,$$

т.е. знаменатели выражений (3.13) не обращаются в нуль. Кроме того,

$$|\alpha_{k+1}| = \frac{|b_{k+1}|}{|c_{k+1} - a_{k+1}\alpha_k|} \leq 1,$$

Следовательно,

$$|\alpha_i| \leq 1, \quad i = 1,2,\ldots,n-1.$$

Далее, учитывая второе из условий (3.17) и только что доказанное неравенство $|\alpha_{n-1}| \leq 1$, имеем

$$|1 - \alpha_n \cdot \alpha_{n-1}| \geq 1 - |\alpha_n| \cdot |\alpha_{n-1}| \geq 1 - |\alpha_n| > 0,$$

т.е. не обращается в нуль и знаменатель в выражении для $x_п$.

К аналогичному выводу можно прийти и в том случае, когда условия (3.16), (3.17) заменяются условиями

$$a_i \neq 0, b_i \neq 0, |c_i| > |a_i| + |b_i|, \quad i = 1,2,\ldots,n-1, \qquad (3.18)$$

$$|\alpha_0| \leq 1, \quad |\alpha_n| \leq 1 \qquad (3.19)$$

или условиями

$$a_i \neq 0, b_i \neq 0, |c_i| \geq |a_i| + |b_i|, \quad i = 1,2,\ldots,n-1, \qquad (3.20)$$

$$|\alpha_0| < 1, \quad |\alpha_n| \leq 1. \qquad (3.21)$$

При условиях (3.18), (3.19) из предположения $|\alpha_i| \leq 1$ следует

$$|c_{i+1} - a_{i+1}\alpha_i| \geq |c_{i+1}| - |a_{i+1}||\alpha_i| \geq |c_{i+1}| - |a_{i+1}| > |b_i|; \quad |\alpha_{i+1}| < 1.$$

Т.е. все прогоночные коэффициенты, начиная с первого, по модулю строго меньше единицы. При этом

$$|1 - \alpha_n \cdot \alpha_{n-1}| \geq 1 - |\alpha_n| \cdot |\alpha_{n-1}| \geq 1 - |\alpha_{n-1}| > 0.$$

При условиях (3.20), (3.21) из предположения $|\alpha_i| < 1$ следует

$$|c_{i+1} - a_{i+1}\alpha_i| \geq |c_{i+1}| - |a_{i+1}||\alpha_i| > |c_{i+1}| - |a_{i+1}| \geq |b_i|; \quad |\alpha_{i+1}| < 1,$$

$$|1 - \alpha_n \cdot \alpha_{n-1}| \geq 1 - |\alpha_n| \cdot |\alpha_{n-1}| \geq 1 - |\alpha_{n-1}| > 1.$$

Таким образом, при выполнении условий (3.16), (3.17) (так же как и условий (3.18), (3.19) или условий (3.20), (3.21)) система (3.8)-(3.9) эквивалентна системе (3.12), (3.15), т.е. эти условия гарантируют существование и единственность решения системы (3.8)-(3.9) и возможность нахождения этого решения методом прогонки. Кроме того, доказанные неравенства $|\alpha_i| \leq 1, \quad i = 1,2,\ldots,n-1$ обеспечивают устойчивость счета по рекуррентным формулам (3.12). Последнее означает, что погрешность, внесенная на каком - либо шаге вычислений, не будет возрастать при переходе к следующим шагам. Действительно,

пусть в формуле (3.12) при $k = k_0 + 1$ вместо x_{k_0+1} вычислена величина $\widetilde{x}_{k_0+1} = x_{k_0+1} + \Delta_{k_0+1}$.

Тогда на следующем шаге вычислений, т.е. при $k = k_0$, вместо $x_{k_0} = \alpha_{k_0} x_{k_0} + \beta_{k_0}$ получим величину $\widetilde{x}_{k_0} = \alpha_{k_0}(x_{k_0+1} + \Delta_{k_0+1})$ и погрешность окажется равной $\Delta_{k_0} = \widetilde{x}_{k_0} - x_{k_0} = \alpha_{k_0}\Delta_{k_0+1}$.

Отсюда получим, что

$$\left|\Delta_{k_0}\right| = \left|\alpha_{k_0}\right| \cdot \left|\Delta_{k_0+1}\right| \le \left|\Delta_{k_0+1}\right|,$$

т.е. погрешность не возрастает.

3.3. Норма вектора и норма матрицы

При изучении итерационных процессов нам понадобятся понятия норм вектора и матрицы. Введем в *n* - мерном векторном пространстве P_n норму вектора.

Нормой вектора $\bar{x}$ называется число $\left\|\bar{x}\right\|$, удовлетворяющее следующим аксиомам нормы:

1) $\left\|\bar{x}\right\| > 0$ *для любого* $\bar{x} \in P_n, \bar{x} \neq \bar{0}$ *и* $\left\|\bar{0}\right\| = 0$

2) $\left\|\alpha\bar{x}\right\| = |\alpha| \cdot \left\|\bar{x}\right\|$ *для любого числа* α *и любого* $\bar{x} \in P_n$

3) $\left\|\bar{x} + \bar{y}\right\| \le \left\|\bar{x}\right\| + \left\|\bar{y}\right\|$ *для любых* $\bar{x}, \bar{y} \in P_n$

Наиболее употребительны в пространстве векторов следующие нормы:

$$\left\|\bar{x}\right\|_1 = \max_{1 \le i \le n} |x_i|; \tag{3.22}$$

$$\left\|\bar{x}\right\|_2 = \sum_{i=1}^{n} |x_i|; \tag{3.23}$$

$$\left\|\bar{x}\right\|_3 = \sqrt{\sum_{i=1}^{n} x_i^{\,2}} = \sqrt{(\bar{x}, \bar{x})}. \tag{3.24}$$

Для всех этих норм выполняются аксиомы нормы. Докажем это для нормы $\left\|\bar{x}\right\|_3$. Выполнение первой аксиомы очевидно. Справедливость второй аксиомы следует из равенства:

$$\left\|\alpha\bar{x}\right\|_3 = \sqrt{(\alpha\bar{x}, \alpha\bar{x})} = \sqrt{\alpha^2(\bar{x}, \bar{x})} = |\alpha|\sqrt{(\bar{x}, \bar{x})} = |\alpha| \cdot \left\|\bar{x}\right\|_3$$

Выполнение третьей аксиомы можно доказать, воспользовавшись неравенством Коши - Буняковского [12].

$$\begin{aligned} &\left|(\bar{x}, \bar{y})\right| \le \sqrt{(\bar{x}, \bar{x})} \cdot \sqrt{(\bar{y}, \bar{y})} \quad \textit{или} \\ &\left|(\bar{x}, \bar{y})\right| \le \left\|\bar{x}\right\|_3 \cdot \left\|\bar{y}\right\|_3. \end{aligned} \tag{3.25}$$

Действительно,

$$\left\|\bar{x}+\bar{y}\right\|_3^2 = \left((\bar{x}+\bar{y}),(\bar{x}+\bar{y})\right) = (\bar{x},\bar{x}) + 2(\bar{x},\bar{y}) + (\bar{y},\bar{y}) \le$$
$$\le \left\|\bar{x}\right\|_3^2 + 2\left\|\bar{x}\right\|_3 \cdot \left\|\bar{y}\right\|_3 + \left\|\bar{y}\right\|_3^2 = \left(\left\|\bar{x}\right\|_3 + \left\|\bar{y}\right\|_3\right)^2,$$

откуда

$$\left\|\bar{x}+\bar{y}\right\|_3 \le \left\|\bar{x}\right\|_3 + \left\|\bar{y}\right\|_3.$$

Очевидно, что введенные нормы векторов удовлетворяют следующим соотношениям:

$$\left\|\bar{x}\right\|_1 \le \left\|\bar{x}\right\|_2 \le n\left\|\bar{x}\right\|_1;$$
$$\left\|\bar{x}\right\|_1 \le \left\|\bar{x}\right\|_3 \le \sqrt{n}\left\|\bar{x}\right\|_1.$$

Введем теперь в пространстве матриц понятие нормы матрицы, согласованной с данной нормой вектора (подчиненной данной, норме вектора).

Нормой матрицы A, согласованной с данной нормой вектора, называется число

$$\left\|A\right\| = \sup_{\substack{\bar{x}\in P_n \\ \bar{x}\neq\bar{0}}} \frac{\left\|A\bar{x}\right\|}{\left\|\bar{x}\right\|}. \qquad (3.26)$$

Докажем, что для нормы матрицы $\left\|A\right\|$ выполнены все три аксиомы нормы.

Выполнение первой аксиомы очевидно. Далее имеем

$$\left\|\alpha A\right\| = \sup_{\substack{\bar{x}\in P_n \\ \bar{x}\neq\bar{0}}} \frac{\left\|\alpha A\bar{x}\right\|}{\left\|\bar{x}\right\|} = \sup_{\substack{\bar{x}\in P_n \\ \bar{x}\neq\bar{0}}} \frac{|\alpha|\left\|A\bar{x}\right\|}{\left\|\bar{x}\right\|} = |\alpha|\left\|A\right\|;$$

$$\left\|A+B\right\| = \sup_{\substack{\bar{x}\in P_n \\ \bar{x}\neq\bar{0}}} \frac{\left\|(A+B)\bar{x}\right\|}{\left\|\bar{x}\right\|} = \sup_{\substack{\bar{x}\in P_n \\ \bar{x}\neq\bar{0}}} \frac{\left\|A\bar{x}+B\bar{x}\right\|}{\left\|\bar{x}\right\|} \le \sup_{\substack{\bar{x}\in P_n \\ \bar{x}\neq\bar{0}}} \frac{\left\|A\bar{x}\right\|}{\left\|\bar{x}\right\|} + \sup_{\substack{\bar{x}\in P_n \\ \bar{x}\neq\bar{0}}} \frac{\left\|B\bar{x}\right\|}{\left\|\bar{x}\right\|} = \left\|A\right\| + \left\|B\right\|.$$

Нормами матриц, согласованными с нормами векторов (3.22), (3.23) и (3.24), являются соответственно нормы

$$\left\|A\right\|_1 = \max_{1\le i\le n} \sum_{j=1}^{n} \left|a_{ij}\right|; \qquad (3.27)$$

$$\left\|A\right\|_2 = \max_{1\le j\le n} \sum_{i=1}^{n} \left|a_{ij}\right|; \qquad (3.28)$$

$$\left\|A\right\|_3 = \sqrt{\max_{1\le i\le n} \lambda_i(A'A)}, \qquad (3.29)$$

где A' - транспонированная матрица A, а $\lambda_i(A'A)$ собственные значения матрицы A'A,

$$i = 1,2,\ldots,n.$$

Приведем вывод этих соотношений для вещественного случая.

Согласно (3.22)

$$\left\|A\bar{x}\right\|_1 = \max_i \left|\sum_{j=1}^n a_{ij}x_j\right| \le \max_i \left(\sum_{j=1}^n \left|a_{ij}\right| \cdot \left|x_j\right|\right) \le$$

$$\le \max_i \left(\sum_{j=1}^n \left|a_{ij}\right| \cdot \max_j \left|x_j\right|\right) \le \max_j \left|x_j\right| \cdot \max_i \sum_{j=1}^n \left|a_{ij}\right| = \left\|\bar{x}\right\|_1 \cdot \max_i \sum_{j=1}^n \left|a_{ij}\right|,$$

откуда имеем, что для любого вектора $\bar{x} \neq \bar{0}$ справедливо неравенство

$$\frac{\left\|A\bar{x}\right\|_1}{\left\|\bar{x}\right\|_1} \le \max_i \sum_{j=1}^n \left|a_{ij}\right|. \qquad (3.30)$$

Пусть $\max_i \sum_{j=1}^n \left|a_{ij}\right|$ достигается при $i = e$.

Рассмотрим вектор $\bar{x}^{(0)}$, у которого

$$x_j^{(0)} = \frac{\left|a_{ej}\right|}{a_{ej}} \quad при \quad a_{ej} \neq 0 \quad и$$

$$x_j^{(0)} = 0 \quad при \quad a_{ej} = 0$$

Очевидно, что $\left\|\bar{x}^{(0)}\right\|_1 = 1$.

$$\left\|A\bar{x}^{(0)}\right\|_1 = \max_i \left|\sum_{j=1}^n a_{ij}x_j^{(0)}\right| \ge \left|\sum_{j=1}^n a_{ej}x_j^{(0)}\right| = \sum_{j=1}^n \left|a_{ej}\right|,$$

откуда

$$\frac{\left\|A\bar{x}^{(0)}\right\|_1}{\left\|\bar{x}^{(0)}\right\|_1} \ge \sum_{j=1}^n \left|a_{ej}\right| = \max_i \sum_{j=1}^n \left|a_{ij}\right| \qquad (3.31)$$

Поскольку для всякого вектора и для $\bar{x}^{(0)}$, в частности, справедливо противоположное неравенство (3.30), заключаем, что

$$\sup_{\substack{\bar{x}\in P_n \\ \bar{x}\neq\bar{0}}} \frac{\left\|A\bar{x}\right\|_1}{\left\|\bar{x}\right\|_1} = \max_i \sum_{j=1}^n \left|a_{ij}\right| = \left\|A\right\|_1.$$

Согласно (3.23)

$$\left\|A\bar{x}\right\|_2 = \sum_{i=1}^n \left|\sum_{j=1}^n a_{ij}x_j\right| \le \sum_{i=1}^n \sum_{j=1}^n \left|a_{ij}\right| \cdot \left|x_j\right| = \sum_{j=1}^n \left(x_j \cdot \sum_{i=1}^n \left|a_{ij}\right|\right) \le$$

$$\le \sum_{j=1}^n \left(x_j \cdot \max_j \sum_{i=1}^n \left|a_{ij}\right|\right) = \left\|\bar{x}\right\|_2 \cdot \max_j \sum_{i=1}^n \left|a_{ij}\right|,$$

откуда заключаем, что для любого вектора $\bar{x} \neq \bar{0}$ справедливо неравенство

$$\frac{\left\|A\bar{x}\right\|_2}{\left\|\bar{x}\right\|_2} \leq \max_j \sum_{i=1}^{n} \left|a_{ij}\right|. \qquad (3.32)$$

Пусть $\max_j \sum_{i=1}^{n} \left|a_{ij}\right| = \sum_{i=1}^{n} \left|a_{il}\right|$.

Рассмотрим вектор $\bar{x}^{(I)}$, у которого l - я координата равна $x_l^{(1)} \neq 0$, а остальные координаты - нули.

Для этого вектора $\left\|\bar{x}^{(1)}\right\|_2 = \sum_{j=1}^{n} \left|x_j^{(1)}\right| = \left|x_l^{(1)}\right|$ и

$$\left\|A\bar{x}^{(1)}\right\|_2 = \sum_{i=1}^{n} \left|\sum_{j=1}^{n} a_{ij} x_j^{(1)}\right| = \sum_{i=1}^{n} \left|a_{il}\right| \cdot \left|x_l^{(1)}\right| = \left\|x_l^{(1)}\right\| \cdot \sum_{i=1}^{n} \left|a_{il}\right| =$$

$$= \left\|\bar{x}^{(1)}\right\|_2 \cdot \max_j \sum_{i=1}^{n} \left|a_{ij}\right|,$$

откуда

$$\frac{\left\|A\bar{x}^{(1)}\right\|_2}{\left\|\bar{x}^{(1)}\right\|_2} = \max_j \sum_{i=1}^{n} \left|a_{ij}\right|. \qquad (3.33.)$$

Из (3.32) и (3.33) следует, что

$$\sup_{\substack{\bar{x} \in P_n \\ \bar{x} \neq \bar{0}}} \frac{\left\|A\bar{x}\right\|_2}{\left\|\bar{x}\right\|_2} = \max_j \sum_{i=1}^{n} \left|a_{ij}\right| = \left\|A\right\|_2 .$$

Согласно (3.26) и (3.24)

$$\left\|A\right\|_3 = \sup_{\substack{\bar{x} \in P_n \\ \bar{x} \neq \bar{0}}} \frac{\left\|A\bar{x}\right\|_3}{\left\|\bar{x}\right\|_3} = \sup_{\substack{\bar{x} \in P_n \\ \bar{x} \neq \bar{0}}} \sqrt{\frac{(A\bar{x}, A\bar{x})}{(\bar{x}, \bar{x})}} = \sqrt{\sup_{\substack{\bar{x} \in P_n \\ \bar{x} \neq \bar{0}}} \frac{(A\bar{x}, A\bar{x})}{(\bar{x}, \bar{x})}}.$$

Матрица A' A -симметрическая, поскольку

$$(A'A)' = A' \cdot (A')' = A'A.$$

Известно, что для всякой вещественной симметрической матрицы B существует базис, составленный из ее собственных векторов [8,11].

Пусть $\bar{e_1}, \bar{e_2}, \ldots, \bar{e_n}$ - ортонормированный базис собственных векторов, а $\lambda_1, \lambda_2, \ldots, \lambda_n$ – соответствующие собственные значения. Всякий вектор $\bar{x}$ представим в виде

$$\bar{x} = \sum_{i=1}^{n} c_i \bar{e_i} .$$

Имеем

$$(B\bar{x}, \bar{x}) = \left(\left(\sum_{i=1}^{n} \lambda_i c_i \bar{e_i} \right), \left(\sum_{i=1}^{n} c_i \bar{e_i} \right) \right) = \sum_{i=1}^{n} \lambda_i c_i^2 ,$$

поэтому

$$(B\bar{x}, \bar{x}) \le \max_i \lambda_i \cdot \sum_{i=1}^{n} c_i^2 = \max_i \lambda_i \cdot (\bar{x}, \bar{x}) \qquad (3.34)$$

и

$$(B\bar{x}, \bar{x}) \ge \min_i \lambda_i \cdot (\bar{x}, \bar{x}) \qquad (3.35)$$

В то же время

$$\frac{(B\bar{e_i}, \bar{e_i})}{(\bar{e_i}, \bar{e_i})} = \frac{(\lambda_i \bar{e_i}, \bar{e_i})}{(\bar{e_i}, \bar{e_i})} = \lambda_i .$$

Из этих соотношений следует, что

$$\sup_{\substack{\bar{x} \in P_n \\ \bar{x} \neq \bar{0}}} \frac{(B\bar{x}, \bar{x})}{(\bar{x}, \bar{x})} = \max_i \lambda_i . \qquad (3.36)$$

Поскольку $(A' A\bar{x}, \bar{x}) = (A\bar{x}, A\bar{x}) \ge 0$, то все $\lambda_i (A' A) \ge 0$ [1]. Полагая в (3.36) $B = A'A$, получим

$$\sup_{\substack{\bar{x} \in P_n \\ \bar{x} \neq \bar{0}}} \frac{(A' A\bar{x}, \bar{x})}{(\bar{x}, \bar{x})} = \max_i \lambda_i (A' A),$$

откуда следует (3.29).

Отметим важный частный случай. Если A - симметрическая матрица, то

$$\lambda_i (A' A) = \lambda_i (A^2) = \lambda_i^2 (A).$$

Поэтому для неё

$$\|A\|_3 = \max_i |\lambda_i (A)| .$$

Рассмотрим некоторые свойства нормы матрицы.

I $\left\| A\bar{x} \right\| \le \|A\| \cdot \left\| \bar{x} \right\|$. (3.37)

Из определения нормы матрицы следует, что для любого $\bar{x} \neq \bar{0}$

$$\|A\| \ge \frac{\left\| A\bar{x} \right\|}{\left\| \bar{x} \right\|} \quad \textit{или} \quad \left\| A\bar{x} \right\| \le \|A\| \cdot \left\| \bar{x} \right\| .$$

Для $\bar{x} = \bar{0}$ имеем

$$A\bar{x} = \bar{0},$$

поэтому (3.37) заполняется как строгое равенство.

II. $\left\|A^k \bar{x}\right\| \le \|A\|^k \cdot \left\|\bar{x}\right\|$. (3.38)

На основании (3.37)

$$\left\|AB\bar{x}\right\| = \left\|A \cdot (B\bar{x})\right\| \le \|A\| \cdot \left\|B\bar{x}\right\| \le \|A\| \cdot \|B\| \cdot \left\|\bar{x}\right\|$$

и, следовательно, имеет место неравенство (3.38).

III. $\|A\| \ge \left|\bar{\lambda}\right|$ (3.39)

где $\bar{\lambda}$ - наибольшее по модулю собственное значение матрицы А.

Пусть $\bar{y}$ - собственный вектор матрицы А, соответствующий $\bar{\lambda}$. Имеем

$$A\bar{y} = \bar{\lambda}\,\bar{y},$$

$$\left\|A\bar{y}\right\| = \left\|\bar{\lambda}\,\bar{y}\right\| = \left|\bar{\lambda}\right| \left\|\bar{y}\right\| \le \|A\| \cdot \left\|\bar{y}\right\|,$$

откуда следует (3.39).

3.4. Метод простой итерации

Простейшим итерационным методом решения систем линейных уравнений является метод простой итерации. Система уравнений (3.1) преобразуется к эквивалентному виду

$$\bar{x} = B\bar{x} + \bar{c}. \quad (3.40)$$

Метод простой итерации состоит в следующем. Выбирается произвольный вектор $\bar{x}^{(0)} \in P_n$ (начальное приближение) и строится итерационная последовательность векторов по формуле

$$\bar{x}^{(k)} = B\bar{x}^{(k-1)} + \bar{c}, \quad k = 1,2,\ldots \quad (3.41)$$

Приведем теорему о достаточном условии сходимости метода простой итерации.

Если $\|B\| < 1$, то система уравнений (3.40) имеет единственное решение $\bar{x}^*$ и итерационный процесс (3.41) сходится к решению со скоростью геометрической прогрессии.

Допустим, что $\bar{x}^*$ - одно из решений системы (3.40), т.е. выполняется равенство

$$\bar{x}^* = B\bar{x}^* + \bar{c}. \quad (3.42)$$

Отсюда, используя третью аксиому нормы и неравенство (3.37), получим

$$\left\|\bar{x}^*\right\| \le \left\|B\bar{x}^*\right\| + \left\|\bar{c}\right\| \le \|B\| \cdot \left\|\bar{x}^*\right\| + \left\|\bar{c}\right\|$$

и

$$(1-\|B\|)\left\|\bar{x}^*\right\| \le \|\bar{c}\|$$

или, поскольку $1-\|B\|>0$,

$$\left\|\bar{x}^*\right\| \le \frac{\|\bar{c}\|}{1-\|B\|}.$$

Из этого неравенства следует единственность решения однородной системы $\bar{x}=B\bar{x}$, т.е. при $\bar{c}=\bar{0}$, а следовательно, существование и единственность решения системы (3.41) при любом свободном члене $\bar{c}$.

Вычтем из равенства (3.42) равенство (3.41). Получим

$$\bar{x}^* - \bar{x}^{(k)} = B\left(\bar{x}^* - \bar{x}^{(k-1)}\right), \quad k=1,2,... \qquad (3.43)$$

и, следовательно,

$$\bar{x}^* - \bar{x}^{(k)} = B^k\left(\bar{x}^* - \bar{x}^{(0)}\right).$$

Отсюда на основании (3.37) имеем

$$\left\|\bar{x}^* - \bar{x}^{(k)}\right\| \le \|B\|^k \left\|\bar{x}^* - \bar{x}^{(0)}\right\|,$$

т.е. норма разности между точным решением и k-м приближением стремится к нулю при $k\to\infty$ не медленнее геометрической прогрессии со знаменателем $q=\|B\|<1$.

Оценим погрешность k-го приближения. Преобразуем равенство (3.43) к виду

$$\bar{x}^* - \bar{x}^{(k-1)} = \bar{x}^k - \bar{x}^{(k-1)} + B\left(\bar{x}^* - \bar{x}^{(k-1)}\right).$$

Согласно третьей аксиоме нормы и равенству (3.37)

$$\left\|\bar{x}^* - \bar{x}^{(k-1)}\right\| \le \left\|\bar{x}^k - \bar{x}^{(k-1)}\right\| + \|B\| \cdot \left\|\bar{x}^* - \bar{x}^{(k-1)}\right\|,$$

откуда

$$\left\|\bar{x}^* - \bar{x}^{(k-1)}\right\| \le \frac{\left\|\bar{x}^{(k)} - \bar{x}^{(k-1)}\right\|}{1-\|B\|}. \qquad (3.44)$$

Кроме того, в силу (3.43) имеем

$$\left\|\bar{x}^* - \bar{x}^{(k)}\right\| \le \|B\| \cdot \left\|\bar{x}^* - \bar{x}^{(k-1)}\right\|. \qquad (3.45)$$

Из (3.44) и (3.45) окончательно получаем

$$\left\|\bar{x}^* - \bar{x}^{(k)}\right\| \le \frac{\|B\|}{1-\|B\|}\left\|\bar{x}^{(k)} - \bar{x}^{(k-1)}\right\|.$$

Приведем без доказательства теорему о необходимом и достаточном условии сходимости метода простой итерации.

Пусть система (3.40) имеет единственное решение. Итерационный процесс (3.41) сходится к решению системы (3.40) при любом начальном приближении тогда и только тогда, когда все собственные значения матрицы B по модулю меньше единицы.

Эта теорема дает более общие условия сходимости метода простой итерации, однако воспользоваться ею в общем случае непросто. В частном случае, когда матрица B симметрическая, можно воспользоваться изложенным в разделе 3.5 методом отыскания максимального по модулю собственного значения, чтобы проверить условия этой теоремы.

Некоторую модификацию метода простой итерации представляет собой метод Зейделя. Основная его идея заключается в том, что при вычислении k - го приближения неизвестной $x_i^{(k)}$ используются уже вычисленные ранее k - е приближения неизвестных $x_1^{(k)}, x_2^{(k)}, \ldots, x_{i-1}^{(k)}$:

$$x_i^{(k)} = \sum_{j=1}^{i-1} b_{ij} x_j^{(k)} + \sum_{j=i}^{n} b_{ij} x_j^{(k-1)} + c_i, \quad i = 1,2,\ldots,n.$$

Условия сходимости методов простой итерации и Зейделя не совпадают, но пересекаются. Обычно метод Зейделя сходится быстрее, чем метод простой итерации [4,5].

3.5. Частичная проблема собственных значений

Задача определения собственных значений и собственных векторов важна и как самостоятельная задача, и как вспомогательная задача. Её можно разбить на три естественных этапа:

построение характеристического многочлена

$$P_n(\lambda) = det(A - \lambda E);$$

решение алгебраического уравнения $P_n(\lambda) = 0$,

т.е. отыскание собственных значений $\lambda_1, \lambda_2, \ldots, \lambda_n$ матрицы;

отыскание ненулевых решений однородной системы

$$(A - \lambda_i E)\bar{x} = \bar{0}, \quad i = 1,2,\ldots,n,$$

т.е. нахождение собственных векторов матрицы A. Каждый из трех отмеченных этапов представляет собой достаточно сложную задачу. Однако иногда можно вычислить собственные значения и соответствующие им собственные векторы, минуя этап построения характеристического многочлена и не прибегая к решению указанных выше систем однородных алгебраических уравнений. Этого удается достичь при помощи различных косвенных соображений, используя те

или иные свойства собственных значений и собственных векторов матрицы.

Мы рассмотрим приближенный метод решения частичной проблемы собственных значений, т.е. задачи нахождения не всех собственных значений и соответствующих им собственных векторов матрицы, а только некоторых из них - метод отыскания максимального по модулю собственного значения матрицы.

Предположим, что квадратная матрица A порядка n, имеет n собственных линейно независимых нормированных векторов, т.е. эти векторы образуют базис n -мерного векторного пространства (как известно, это всегда имеет место, если A - симметрическая матрица).

$$A\overline{y_i} = \lambda_i \overline{y_i}, \quad i = 1,2,\ldots,n. \tag{3.46}$$

$$\left\|\overline{y_i}\right\|_3 = \sqrt{(\overline{y_i}, \overline{y_i})} = \sqrt{y_{i1}^2 + y_{i2}^2 + \ldots + y_{in}^2} = 1, \quad i = 1,2,\ldots,n. \tag{3.47}$$

Допустим, что

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n|. \tag{3.48}$$

Возьмем произвольный вектор $\overline{x}^{(0)} \neq \overline{0}$. Имеем

$$\overline{x}^{(0)} = c_1\overline{y}_1 + c_2\overline{y}_2 + \ldots + c_n\overline{y}_n,$$

где $c_1, c_2, \ldots, c_n$ - координаты вектора $\overline{x}^{(0)}$ в базисе собственных векторов $\overline{y_1}, \overline{y_2}, \ldots, \overline{y_n}$. Предположим, что

$$c_1 \neq 0. \tag{3.49}$$

Последовательно находим векторы

$$\overline{x}^{(k)} = A\overline{x}^{(k-1)}, \quad k = 1,2,\ldots \tag{3.50}$$

Тогда согласно (3.46)

$$\overline{x}^{(1)} = A\overline{x}^{(0)} = A\left(c_1\overline{y_1} + c_2\overline{y_2} + \ldots + c_n\overline{y_n}\right) = c_1\lambda_1\overline{y_1} + c_2\lambda_2\overline{y_2} + \ldots + c_n\lambda_n\overline{y_n},$$

$$\overline{x}^{(2)} = A\overline{x}^{(1)} = c_1\lambda_1^2\overline{y_1} + c_2\lambda_2^2\overline{y_2} + \ldots + c_n\lambda_n^2\overline{y_n}$$

и вообще

$$\overline{x}^{(k)} = c_1\lambda_1^k\overline{y_1} + c_2\lambda_2^k\overline{y_2} + \ldots + c_n\lambda_n^k\overline{y_n} = \lambda_1^k\left(c_1\overline{y_1} + \overline{z}^{(k)}\right), \tag{3.51}$$

где $\overline{z}^{(k)} = c_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\overline{y_2} + c_3\left(\frac{\lambda_3}{\lambda_1}\right)^k\overline{y_3} + \ldots + c_n\left(\frac{\lambda_n}{\lambda_1}\right)^k\overline{y_n}$.

В силу (3.48) $\left\|\overline{z}^{(k)}\right\|_3 \to 0$ при $k \to \infty$ и

$$\left\|\overline{z}^{(k)}\right\|_3 = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^k\right). \tag{3.52}$$

Значит, вектор $\bar{x}^{(k)}$ при больших k близок к собственному вектору матрицы А, соответствующему собственному значению λ_1.

Используя (3.51), найдем скалярное произведение

$$\begin{aligned}&\left(\bar{x}^{(k)},\bar{x}^{(k-1)}\right)=\lambda_1^k\left(c_1\overline{y_1}+\bar{z}^{(k)}\right)\cdot\lambda_1^{k-1}\left(c_1\overline{y_1}+\bar{z}^{(k-1)}\right)=\\&=\lambda_1^{2k-1}\left[c_1^2\left(\overline{y_1},\overline{y_1}\right)+c_1\left(\overline{y_1},\bar{z}^{(k-1)}\right)+c_1\left(\overline{y_1},\bar{z}^{(k)}\right)+\left(\bar{z}^{(k)},\bar{z}^{(k-1)}\right)\right]\end{aligned}\qquad(3.53)$$

Согласно (3.47) $(\overline{y_1},\overline{y_1})=\left\|\overline{y_1}\right\|_3^2=1$. Для каждого из остальных скалярных произведений в (3.53) воспользуемся неравенством Коши - Буняковского (3.25):

$$\left|\left(\overline{y_1},\bar{z}^{(k-1)}\right)\right|\le\left\|\overline{y_1}\right\|_3\cdot\left\|\bar{z}^{(k-1)}\right\|_3=\left\|\bar{z}^{(k-1)}\right\|_3;$$

$$\left|\left(\overline{y_1},\bar{z}^{(k)}\right)\right|\le\left\|\overline{y_1}\right\|_3\cdot\left\|\bar{z}^{(k)}\right\|_3=\left\|\bar{z}^{(k)}\right\|_3;$$

$$\left|\left(\bar{z}^{(k)},\bar{z}^{(k-1)}\right)\right|\le\left\|\bar{z}^{(k)}\right\|_3\cdot\left\|\bar{z}^{(k-1)}\right\|_3.$$

Теперь из (3.53) с учётом (3.52) получим

$$\left(\bar{x}^{(k)},\bar{x}^{(k-1)}\right)=\lambda_1^{2k-1}\left[c_1^2+O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{k-1}\right)\right].$$

Аналогично можем получить

$$\left(\bar{x}^{(k-1)},\bar{x}^{(k-1)}\right)=\lambda_1^{2k-2}\left[c_1^2+O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{k-1}\right)\right].$$

Последние две соотношения дадут

$$\frac{\left(\bar{x}^{(k)},\bar{x}^{(k-1)}\right)}{\left(\bar{x}^{(k-1)},\bar{x}^{(k-1)}\right)}=\lambda_1+O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{k-1}\right),$$

$$\lambda_{1k}=\frac{\left(\bar{x}^{(k)},\bar{x}^{(k-1)}\right)}{\left(\bar{x}^{(k-1)},\bar{x}^{(k-1)}\right)}\to\lambda_1\quad при\quad k\to\infty.\qquad(3.54)$$

Таким образом, при условии (3.48) итерационный процесс (3.50) позволяет найти с любой точностью максимальное по модулю собственное значение λ_1 и соответствующий ему собственный вектор.

Следует заметить, что если $|\lambda_1|>1,\quad то\quad\left\|\bar{x}^{(k)}\right\|\to\infty\ \ при\ k\to\infty.$

Если же $|\lambda_1|<1,\quad то\quad\left\|\bar{x}^{(k)}\right\|\to 0\quad при\ k\to\infty$. То и другое явление при счете на ЭВМ нежелательно. В первом случае может наступить переполнение (выход за допустимый диапазон чисел). Во втором случае

$\left\|\bar{x}^{(k)}\right\|$ может стать машинным нулем (слишком малой величиной), и информация теряется. Поэтому целесообразно на каждой итерации нормировать собственный вектор $\bar{x}^{(k)}$, т.е. итерации вести по формулам:

$$\bar{y}_1^{(0)} = \frac{\bar{x}^{(0)}}{\left\|\bar{x}^{(0)}\right\|_3};$$

$$\bar{x}^{(k)} = A\bar{y}_1^{(k-1)}; \quad \lambda_{1\kappa} = \left(\bar{x}^{(\kappa)}, \bar{x}^{(\kappa-1)}\right);$$

$$\bar{y}_1^{(k)} = \frac{\bar{x}^{(k)}}{\left\|\bar{x}^{(k)}\right\|_3}.$$

Подтверждением того, что λ_1 не является кратным собственным значением и что нет собственного значения, равного - λ_1, служит сходимость итерационного процесса при выборе различных $\bar{x}^{(0)}$ к одному и тому же собственному вектору (с точностью до противоположного вектора).

Рассмотрим теперь, как, используя метод отыскания максимального по модулю собственного значения матрицы, определить максимальное и минимальное собственные значения симметрической матрицы. Как известно, все собственные значения вещественной симметрической матрицы A действительны [8, II] и существует ортонормированный базис $\bar{e}_1, \bar{e}_2, \ldots, \bar{e}_n$, составленный из собственных векторов матрицы A.

Пусть $P_1(t) = a_0 t + a_1$ -

некоторый алгебраический многочлен от t первой степени с действительными коэффициентами. Обозначим через B следующую матрицу

$$B = a_0 A + a_1 E,$$

где E - единичная матрица. Докажем, что собственные значения матриц A и B связаны соотношением

$$\lambda_i(B) = P_1(\lambda_i(A)), \qquad (3.55)$$

а собственный вектор матрицы A, соответствующий собственному значению $\lambda_i(A)$, является собственным вектором матрицы B, соответствующим собственному значению $\lambda_i(B)$.

Пусть $\bar{e}_i$ - собственный вектор матрицы A, соответствующий собственному значению $\lambda_i(A)$:

$$A\bar{e}_i = \lambda_i(A) \cdot \bar{e}_i.$$

Тогда

$$B\bar{e}_i = (a_0 A + a_1 E)\bar{e}_i = a_0 \lambda_i(A)\bar{e}_i + a_1 \bar{e}_i =$$
$$= (a_0 \cdot \lambda_i(A) + a_1)\bar{e}_i = \lambda_i(B)\bar{e}_i.$$

Допустим, что максимальное по модулю собственное значение $\bar{\lambda}(A)$ симметрической матрицы A известно. Постоим матрицу

$$B = A - \bar{\lambda}(A)E \qquad (3.56)$$

и определим для нее максимальное по модулю собственное значение $\bar{\lambda}(B)$.

Если $\bar{\lambda}(A) > 0$, то очевидно, что

$$\max_{1 \le i \le n} \lambda_i(A) = \bar{\lambda}(A).$$

Кроме того, согласно (3.55) и (3.56)

$$\lambda_i(B) = \lambda_i(A) - \bar{\lambda}(A) \le 0, \quad i = 1,2,\ldots,n.$$

Поэтому

$$\bar{\lambda}(B) = \min_{1 \le i \le n} \left[\lambda_i(A) - \bar{\lambda}(A)\right] = \min_{1 \le i \le n} \lambda_i(A) - \bar{\lambda}(A),$$

т.е.
$$\min_{1 \le i \le n} \lambda_i(A) = \bar{\lambda}(A) + \bar{\lambda}(B). \qquad (3.57)$$

Если $\bar{\lambda}(A) < 0$, то

$$\min_{1 \le i \le n} \lambda_i(A) = \bar{\lambda}(A)$$

и

$$\lambda_i(B) = \lambda_i(A) - \bar{\lambda}(A) \ge 0, \quad i = 1,2,\ldots,n.$$

Поэтому

$$\bar{\lambda}(B) = \max_{1 \le i \le n} \lambda_i(A) - \bar{\lambda}(A),$$

откуда

$$\max_{1 \le i \le n} \lambda_i(A) = \bar{\lambda}(A) + \bar{\lambda}(B).$$

4. ИНТЕРПОЛИРОВАНИЕ

Задача приближения (аппроксимации) функций возникает и как самостоятельная, и при решении многих других задач. Простейшая ситуация, приводящая к приближению функций, заключается в следующем. При некоторых значениях аргумента $x_0, x_1, \ldots x_n$, называемых узлами, заданы значения функции $y_i=f(x_i), \ i=0,1\ldots,n.$ Требуется восстановить значения функции при других x. Подобная же задача возникает при многократном вычислении на ЭВМ одной и той же сложной функции в различных точках. Вместо этого часто бывает целесообразно вычислять значения этой функции в небольшом числе характерных точек x_i, а в остальных точках вычислять ее значения по некоторому более простому правилу, используя информацию об уже известных значениях y_i.

Другими распространенными примерами приближения функций являются задачи определения производной $f'(x)$ и интеграла $\int_a^b f(x)dx$ по заданным значениям y_i.

Классический подход к решению подобных задач заключается в том, чтобы, используя имеющуюся информацию о функции $f(x)$, рассмотреть другую функцию $\varphi(x)$, близкую к $f(x)$, позволяющую выполнить над ней соответствующую операцию и получить оценку погрешности такой «аналитической замены».

При выборе класса, к которому принадлежит аппроксимирующая функция $\varphi(x)$, следует руководствоваться тем, что $\varphi(x)$, с одной стороны, должна отражать характерные особенности аппроксимируемой функции $f(x)$, с другой стороны, быть достаточно удобной в обращении.

Вопрос о близости аппроксимируемой и аппроксимирующей функций решается по-разному. Если параметры, от которых зависит функция $\varphi(x)$, определяются из условия совпадения значений функций $f(x)$ и $\varphi(x)$ в узлах, то такой способ аппроксимации называется интерполированием (интерполяцией).

Наличие большого количества различных способов приближения объясняется многообразием различных постановок задачи. Далее мы рассмотрим лишь один раздел теории приближения – интерполирование многочленами. Аппарат интерполирования многочленами является важнейшим аппаратом численного анализа. На его основе строится большинство численных методов решения других задач.

4.1. Интерполяционный полином, его существование и единственность. Остаточный член

Будем строить аппроксимирующую функцию в виде

$$\varphi(x) = P_n(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_n . \qquad (4.1)$$

Коэффициенты $a_0, a_1, \ldots, a_n$ определим из условий

$$P_n(x_i) = y_i, \quad i = 0,1,\ldots n . \qquad (4.2)$$

Распишем подробно эти условия:

$$P_n(x_0) = a_0 x_0^n + a_1 x_0^{n-1} + \ldots + a_n = y_0 .$$

$$\ldots\ldots\ldots\ldots\ldots$$

$$P_n(x_n) = a_0 x_n^n + a_1 x_n^{n-1} + \ldots + a_n = y_n .$$

Определитель этой системы

$$\begin{vmatrix} x_0^n & x_0^{n-1} & \ldots & x_0 & 1 \\ x_1^n & x_1^{n-1} & \ldots & x_1 & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_n^n & x_n^{n-1} & \ldots & x_n & 1 \end{vmatrix}$$

может быть получен из определителя Вандермонда

$$\begin{vmatrix} 1 & 1 & \ldots & 1 \\ x_0 & x_1 & \ldots & x_n \\ x_0^2 & x_1^2 & \ldots & x_n^2 \\ \ldots & \ldots & \ldots & \ldots \\ x_0^n & x_1^n & \ldots & x_n^n \end{vmatrix}$$

транспонированием матрицы и последующей перестановкой ее строк, т.е. будет отличаться от определителя Вандермонда лишь знаком.

Последний, как известно, равен $\Pi(x_i - x_j)$ [8], т.е. отличен от $1 \le x_j < x_i \le n$ нуля, если узлы интерполирования x_i различны.

Следовательно, коэффициенты $a_0, a_1, \ldots, a_n$ интерполяционного полинома (4.1) всегда могут быть определены, и при том единственным образом. Таким образом, доказано существование и единственность интерполяционного полинома (4.1).

Оценим остаточный член интерполирования

$$R_n(x^*) = f(x^*) - P_n(x^*), \qquad (4.3)$$

где x* – точка, в которой значение функции вычисляется с помощью интерполяционного полинома.

Предположим, что узлы упорядочены: $a \le x_0 < x_1 < \ldots < x_n \le b$ и $f^{n+1}(x)$ непрерывна на $[a, b]$, $x^* \in [a,b]$.

Введем вспомогательную функцию

$$F(x) = f(x) - P_n(x) - k(x - x_0)(x - x_1)\ldots(x - x_n), \qquad (4.4)$$

где константа k выбирается так, чтобы

$$F(x^*) = 0,$$

отсюда

$$k = \frac{f(x^*) - P_n(x^*)}{(x^* - x_0)(x^* - x_1)\ldots(x^* - x_n)}. \qquad (4.5)$$

При таком выборе k функция *f(x)* обращается в нуль в *(n+2)* точках $x_0, x_1, \ldots, x_n, x^*$. На основании теоремы Ролля ее производная *F'(x)* обращается в нуль, по крайней мере, в *(n+1)*-й точке. Применяя теорему Ролля к *F'(x)*, получаем, что ее производная *F''(x)* обращается в нуль по крайней мере в *n* точках. Продолжая эти рассуждения дальше, получаем, что $F^{(n+1)}(x)$ обращается в нуль по крайней мере в одной точке ξ, принадлежащей отрезку *[a, b]*. Поскольку

$$F^{(n+1)}(x) = f^{(n+1)}(x) - k(n+1)!,$$

из условия $F^{(n+1)}(\xi) = 0$ будем иметь

$$k = \frac{f^{(n+1)}(\xi)}{(n+1)!}. \qquad (4.6)$$

Приравнивая правые части (4.5) и (4.6), получим представление остаточного члена в точке x*

$$R_n(x^*) = f(x^*) - P_n(x^*) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x^* - x_0)(x^* - x_1)\ldots(x^* - x_n), \qquad (4.7)$$

где $\xi \in [a,b]$.

Остаточная абсолютная погрешность интерполирования Δ_1 в точке x^* может быть оценена как

$$\left|R_n(x^*)\right| \le \Delta_1 = \frac{M_{n+1}}{(n+1)!}\left|(x^* - x_0)(x^* - x_1)\ldots(x^* - x_n)\right|, \qquad (4.8)$$

где $M_{n+1} = \max\limits_{[a,b]} \left|f^{(n+1)}(x)\right|$.

Так как точка x^* – произвольная точка отрезка *[a, b]*, выражение (4.7) остаточного члена справедливо для любой точки $x \in [a,b]$. Найдем оценку остаточной погрешности интерполирования на всем отрезке *[a, b]:*

$$\left|R_n(x)\right| \le \Delta_1 = \frac{M_{n+1}}{(n+1)!} \cdot \max_{[a,b]} \left|\omega_n(x)\right|,$$

где $\omega_n(x) = \left|(x - x_0)(x - x_1)\ldots(x - x_n)\right|$.

Оценить $\omega_n(x)$ при произвольном расположении узлов интерполяции сложно. Если же узлы расположены на одинаковом расстоянии *h* друг от друга., то $\omega_n(x)$ имеет примерно такой вид, как показано на рисунке 4.1. для *n = 5* [3].

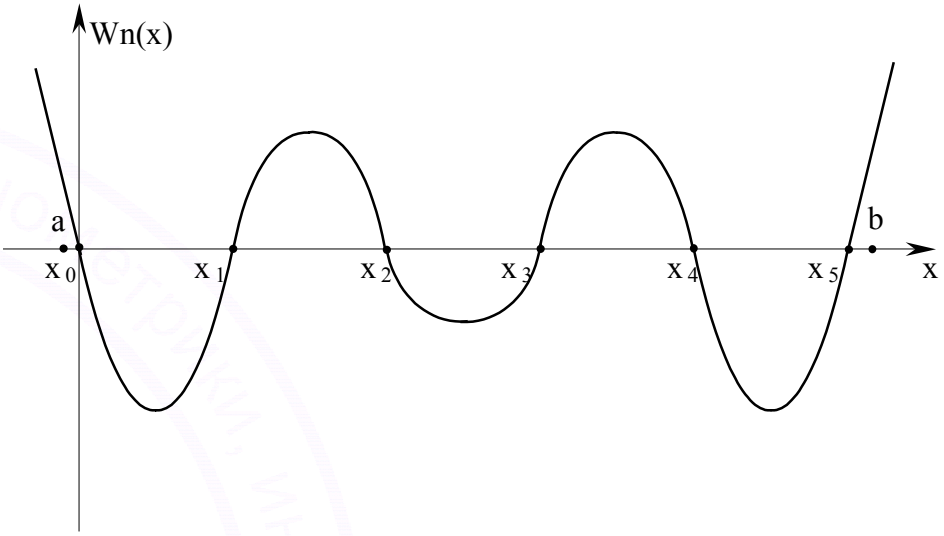

Рис. 4.1.

Вблизи центрального узла интерполяции экстремумы невелики, вблизи крайних узлов – несколько больше, а если X выходит за крайние узлы интерполяции, то $\omega_n(x)$ быстро возрастает. Термин «интерполяция» в узком смысле употребляют, если x заключен между крайними узлами; если же он выходит из этих пределов, то говорят об экстраполяции. Очевидно, что при экстраполяции далеко за крайним узлом ошибка может быть велика, поэтому экстраполяция малонадежна.

4.2. Интерполяционный полином Лагранжа

Будем строить интерполяционный полином в виде

$$P_n(x) = \sum_{i=0}^{n} l_i^{(n)}(x) y_i \,, \qquad (4.9)$$

где $l_i^{(n)}(x)$ – многочлены степени не выше *n,* обладающие следующим свойством:

$$l_i^{(n)}(x) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}, \quad i, j = 0,1,\ldots,n \,.$$

Действительно, в этом случае полином (4.9) в каждом узле x_j, $j=0,1,\ldots n$, равен соответствующему значению функции y_j, т.е. является интерполяционным.

Построим такие многочлены. Поскольку $l_i^{(n)}(x)=0$ при $x=x_0,x_1,\ldots x_{i-1},x_{i+1},\ldots x_n$, $l_i^{(n)}(x)$ можно следующим образом разложить на множители

$$l_i^{(n)}(x)=c(x-x_0)(x-x_1)\ldots(x-x_{i-1})(x-x_{i+1})\ldots(x-x_n)=$$

$$=c\prod_{\substack{j=0\\ j\neq i}}^{n}(x-x_j),$$

где c – постоянная. Из условия $l_i^{(n)}(x_i)=1$ получим, что

$$c=\frac{1}{\prod\limits_{\substack{j=0\\ j\neq i}}^{n}(x-x_j)} \quad \text{и} \quad l_i^{(n)}(x)=\prod_{\substack{j=0\\ j\neq i}}^{n}\frac{(x-x_j)}{(x_i-x_j)}.$$

Интерполяционный полином (4.1), записанный в форме

$$P_n(x)\equiv L_n(x)=\sum_{i=0}^{n}y_i\prod_{\substack{j=0\\ j\neq i}}^{n}\frac{(x-x_j)}{(x_i-x_j)}, \qquad (4.10)$$

называют интерполяционным полиномом Лагранжа.

Приближенное значение функции в точке x^*, вычисленное с помощью полинома Лагранжа, будет иметь остаточную погрешность (4.8). Если значения функции y_i в узлах интерполирования x_i заданы приближенно с одинаковой абсолютной погрешностью Δ^*, то вместо точного значения $L_n(x^*)$ будет вычислено приближенное значение $\overline{L_n}(x^*)$, причем

$$\left|L_n(x^*)-\overline{L_n}(x^*)\right|\le\Delta_2=\Delta^*\sum_{i=0}^{n}\prod_{\substack{j=0\\ j\neq i}}^{n}\left|\frac{(x^*-x_j)}{(x_i-x_j)}\right|,$$

где Δ_2 – вычислительная абсолютная погрешность интерполяционного полинома Лагранжа. Окончательно имеем следующую оценку полной погрешности приближенного значения $\overline{L_n}(x^*)$.

$$\left|f(x^*)-\overline{L_n}(x^*)\right|=\left|f(x^*)-L_n(x^*)+L_n(x^*)-\overline{L_n}(x^*)\right|\le\Delta_1+\Delta_2=\Delta_{полн}.$$

В частности, полиномы Лагранжа первой и второй степени будут иметь вид

$$L_1(x)=\frac{x-x_1}{x_0-x_1}y_0+\frac{x-x_0}{x_1-x_0}y_1;$$

$$L_2(x)=\frac{(x-x_1)(x-x_2)}{(x_0-x_{1)(x_0-x_2)}}y_0+\frac{(x-x_0)(x-x_2)}{(x_1-x_{0)(x_1-x_2)}}y_1+\frac{(x-x_0)(x-x_1)}{(x_2-x_{0)(x_2-x_1)}}y_2,$$

а их полные погрешности в точке x^*

$$\Delta_{полн.}(L_1)=\frac{M_2}{2!}\left|(x^*-x_0)(x^*-x_1)\right|+\Delta^*\left(\left|\frac{x^*-x_1}{x_0-x_1}\right|+\left|\frac{x^*-x_0}{x_1-x_0}\right|\right),$$

$$\Delta_{полн.}(L_2)=\frac{M_3}{3!}\left|(x^*-x_0)(x^*-x_1)(x^*-x_2)\right|+$$

$$+\Delta^*\left(\left|\frac{(x^*-x_1)(x^*-x_2)}{(x_0-x_1)(x_0-x_2)}\right|+\left|\frac{(x^*-x_0)(x^*-x_2)}{(x_1-x_0)(x_1-x_2)}\right|+\left|\frac{(x^*-x_0)(x^*-x_2)}{(x_2-x_0)(x_2-x_1)}\right|\right),$$

где $M_2=\max\limits_{[a,b]}\left|f''(x)\right|$; $M_3=\max\limits_{[a,b]}\left|f'''(x)\right|$.

Существуют другие формы записи того же интерполяционного полинома (4.1), например, рассматриваемая далее интерполяционная формула Ньютона с разделенными разностями и ее варианты. При точных вычислениях значения $Pn(x^*)$, получаемые по различным интерполяционным формулам, построенным по одним и тем же узлам, совпадают. Наличие же вычислительной погрешности приводит к различию получаемых по этим формулам значений. Запись многочлена в форме Лагранжа приводит, как правило, к меньшей вычислительной погрешности [1-3].

Использование формул для оценки погрешностей, возникающих при интерполировании, зависит от постановки задачи. Например, если известно количество узлов, а функция задана с достаточно большим количеством верных знаков, то можно поставить задачу вычисления $f(x^*)$ с максимально возможной точностью. Если, наоборот, количество верных знаков небольшое, а количество узлов велико, то можно поставить задачу вычисления $f(x^*)$ с точностью, которую допускает табличное значение функции, причем для решения этой задачи может потребоваться как разрежение, так и уплотнение таблицы.

4.3. Разделенные разности и их свойства

Понятие разделенной разности является обобщенным понятием производной. Пусть в точках x_0, x_1,…x_n заданы значения функций $f(x_0)$, $f(x_1),\ldots,f(x_n)$. Разделенные разности первого порядка определяются равенствами

$$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0};$$

$$f[x_1, x_2] = \frac{f(x_2) - f(x_1)}{x_2 - x_1};$$

.....

$$f[x_i, x_{i+1}] = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i};$$

разделенные разности второго порядка – равенствами,

$$f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0};$$

$$f[x_1, x_2, x_3] = \frac{f[x_2, x_3] - f[x_1, x_2]}{x_3 - x_1};$$

.....

$$f[x_i, x_{i+1}, x_{i+2}] = \frac{f[x_{i+1}, x_{i+2}] - f[x_i, x_{i+1}]}{x_{i+2} - x_i},$$

а разделенные разности *k*-го порядка определяются следующей рекуррентной формулой:

$$f[x_i, x_{i+1}, \ldots, x_{i+k}] = \frac{f[x_{i+1}, \ldots, x_{i+k}] - f[x_i, \ldots, x_{i+k-1}]}{x_{i+k} - x_i} \quad (4.11)$$

Разделенные разности обычно помещаются в таблицу следующего вида:

x_i	$f(x_i)$	Разделение разности			
		I порядка	II порядка	III порядка	IV порядка
x_0	y_0				
		$f[x_0,x_1]$			
x_1	y_1		$f[x_0,x_1,x_2]$		
		$f[x_1,x_2]$		$f[x_0,x_1,x_2,x_3]$	
x_2	y_2		$f[x_1,x_2,x_3]$		$f[x_0,x_1,x_2,x_3,x_4]$
		$f[x_2,x_3]$		$f[x_1,x_2,x_3,x_4]$	
x_3	y_3		$f[x_2,x_3,x_4]$		
		$f[x_3,x_4]$			
x_4	y_4				

Рассмотрим следующие свойства разделенных разностей.

1. Разделенные разности всех порядков являются линейными комбинациями значений $f(x_i)$, т.е. имеет место следующая формула:

$$f[x_0, x_1, \ldots, x_k] = \sum_{i=0}^{k} \frac{f(x_i)}{\prod\limits_{\substack{j=0 \\ j \neq i}}^{k} (x_i - x_j)} \quad . \quad (4.12)$$

Докажем справедливость этой формулы индукцией по порядку разностей. Для разностей первого порядка

$$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{f(x_1)}{x_1 - x_0} + \frac{f(x_0)}{x_0 - x_1}.$$

Формула (4.12) справедлива. Предположим теперь, что она справедлива для всех разностей порядка $k \le n$.

Тогда, согласно (4.11) и (4.12) для разностей порядка $k=n+1$ имеем

$$f[x_0, x_1, \ldots, x_{n+1}] = \frac{f[x_1, \ldots, x_{n+1}] - f[x_0, \ldots, x_n]}{x_{n+1} - x_0} =$$

$$\frac{1}{x_{n+1} - x_0}\left[\sum_{i=1}^{n} \frac{f(x_i)}{\prod\limits_{\substack{j=1 \\ j\neq i}}^{n+1}(x_i - x_j)} - \sum_{i=1}^{n} \frac{f(x_i)}{\prod\limits_{\substack{j=0 \\ j\neq i}}^{n}(x_i - x_j)}\right] + \frac{f(x_{n+1})}{\prod\limits_{j=0}^{n}(x_{n+1} - x_j)} -$$

$$- \frac{f(x_0)}{\prod\limits_{j=1}^{n+1}(x_0 - x_j)}.$$

Слагаемые, содержащие $f(x_0)$ и $f(x_{n+1})$, имеют требуемый вид. Рассмотрим слагаемые, содержащие $f(x_i)$, $i=1, 2, \ldots, n$. Таких слагаемых два - из первой и второй сумм:

$$\frac{1}{x_{n+1} - x_0}\left[\frac{f(x_i)}{\prod\limits_{\substack{j=1 \\ j\neq i}}^{n+1}(x_i - x_j)} - \frac{f(x_i)}{\prod\limits_{\substack{j=0 \\ j\neq i}}^{n}(x_i - x_j)}\right] = \frac{(x_i - x_0)f(x_i) - (x_i - x_{n+1})f(x_i)}{(x_{n+1} - x_0)\prod\limits_{\substack{j=0 \\ j\neq i}}^{n+1}(x_i - x_j)} =$$

$$= \frac{f(x_i)}{\prod\limits_{\substack{j=0 \\ j\neq i}}^{n+1}(x_i - x_j)},$$

т.е. формула (4.12) справедлива для разности порядка $k=n+1$, доказательство закончено.

2. Разделенная разность есть симметрическая функция своих аргументов $x_0, x_1, \ldots x_n$ (т.е. не меняется при любой их перестановке):

$$f[x_0, x_1, \ldots, x_n] = f[x_1, x_0, \ldots, x_n] = \ldots = f[x_n, x_{n-1}, \ldots, x_1, x_0].$$

Это свойство непосредственно следует из равенства (4.12).

3. Простую связь разделенной разности $f[x_0, x_1, \ldots, x_n]$ и производной $f^{(n)}(x)$ дает следующая теорема.

Пусть узлы $x_0, x_1, \ldots x_n$ принадлежат отрезку $[a, b]$ и функция $f(x)$ имеет на этом отрезке непрерывную производную порядка n. Тогда существует такая точка $\xi \in [a, b]$, что

$$f[x_0, x_1, \ldots, x_n] = \frac{f^{(n)}(\xi)}{n!} \quad . \qquad (4.13)$$

Докажем сначала справедливость соотношения

$$f(x) - L_n(x) = f[x_0, x_1, \ldots, x_n, x](x - x_0)(x - x_1)\ldots(x - x_n), \qquad (4.14)$$

где $x \in [a,b], \quad x \neq x_i, \quad i = 0,1,\ldots,x_n$.

$$f(x) - L_n(x) = f(x) - \sum_{i=0}^{n} f(x_i) \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{(x - x_j)}{(x_i - x_j)} =$$

$$\prod_{j=0}^{n}(x - x_j) \cdot \left[\frac{f(x)}{\prod_{j=0}^{n}(x - x_j)} - \sum_{i=0}^{n} \frac{f(x_i)}{(x - x_i) \prod_{\substack{j=0 \\ j \neq i}}^{n}(x_i - x_j)} \right] =$$

$$= \prod_{j=0}^{n}(x - x_j) \cdot \left[\frac{f(x)}{\prod_{j=0}^{n}(x - x_j)} + \sum_{i=0}^{n} \frac{f(x_i)}{(x_i - x) \prod_{\substack{j=0 \\ j \neq i}}^{n}(x_i - x_j)} \right].$$

Согласно (4.12) выражение в квадратных скобках есть

$$f[x_0, x_1, \ldots, x_n, x].$$

Из сравнения (4.14) с выражением (4.7) для остаточного члена $R_n(x)=f(x)-L_n(x)$ получим (4.13), теорема доказана.

Из этой теоремы вытекает простое следствие. Для полинома n-ой степени

$$f(x) = a_0 x^n + a_1 x^{n-1} + \ldots a_n$$

производная порядка n, очевидно, есть

$$f^{(n)}(x) = a_0 n!$$

и соотношение (4.13) дает для разделенной разности значение

$$f[x_0, x_1, \ldots, x_n] = \frac{a_0 n!}{n!} = a_0 .$$

Итак, у всякого многочлена степени n разделенные разности порядка n равны постоянной величине – коэффициенту при старшей степени многочлена. Разделенные разности высших порядков (больше n), очевидно, равны нулю. Однако этот вывод справедлив лишь в случае отсутствия вычислительной погрешности у разделенных разностей.

4.3. Интерполяционный полином Ньютона с разделенными разностями

Запишем интерполяционный полином Лагранжа в следующем виде:

$$L_n(x) \equiv L_0(x) + [L_1(x) - L_0(x)] + \ldots + [L_k(x) - L_{k-1}(x)] + \ldots$$
$$\ldots + [L_n(x) - L_{n-1}(x)], \qquad (4.15)$$

где $L_0(x) = f(x_0) = y_0$, а $L_k(x)$ – интерполяционный полином Лагранжа степени k, построенный по узлам $x_0, x_1, \ldots, x_k$. Тогда $[L_k(x) - L_{k-1}(x)]$

есть полином степени k, корнями которого являются точки $x_0, x_1, \ldots, x_{k-1}$. Следовательно, его можно разложить на множители

$$L_k(x) - L_{k-1}(x) = A_k(x - x_0)(x - x_1)\ldots(x - x_{k-1}) \quad (4.16)$$

где A_k – постоянная.

В соответствии с (4.14) получим

$$L_k(x) - L_{k-1}(x) = f(x_k) - L_{k-1}(x) = f[x_0, \ldots, x_k](x_k - x_0)\ldots(x_k - x_{k-1}) \quad (4.17)$$

Сравнивая (4.16) и (4.17) получим, что $A_k = f[x_0, \ldots, x_k]$ и (4.15) примет вид

$$\begin{aligned} L_n(x) = N_n(x) = y_0 + f[x_0, x_1](x - x_0) + \ldots + \\ + f[x_0, x_1, \ldots, x_n](x - x_0)\ldots(x - x_{n-1}), \end{aligned} \quad (4.18)$$

который носит название интерполяционного полинома Ньютона с разделенными разностями.

Этот вид записи интерполяционного полинома более нагляден (добавлению одного узла соответствует появление одного слагаемого) и позволяет лучше проследить аналогию проводимых построений с основными построениями математического анализа.

Остаточная погрешность интерполяционного полинома Ньютона выражается формулой (4.8), но ее, с учетом (4.13), можно записать и в другой форме

$$\Delta_1 \approx \max_{[a,b]} \left| f[x_i, x_{i+1}, \ldots, x_{i+n+1}] \right| \cdot \left| (x^* - x_0)\ldots(x^* - x_n) \right|,$$

т.е. остаточная погрешность может быть оценена модулем первого отброшенного слагаемого в полиноме $N_n(x^*)$.

Вычислительная погрешность $N_n(x^*)$ определится погрешностями разделенных разностей. Узлы интерполяции, лежащие ближе всего к интерполируемому значению x^*, окажут большее влияние на интерполяционный полином, лежащие дальше – меньшее. Поэтому целесообразно, если это возможно, за x_0 и x_1 взять ближайшие к x^* узлы интерполирования и произвести сначала линейную интерполяцию по этим узлам. Затем постепенно привлекать следующие узлы так, чтобы они возможно симметричнее располагались относительно x^*, пока очередной член по модулю не будет меньше абсолютной погрешности входящей в него разделенной разности.

4.4. Конечные разности и их свойства

Пусть узлы x_i, в которых заданы значения функции $f(x_i)=y_i$, являются равноотстающими, т.е. $x_1 = x_0 + h$, $x_2 = x_1 + h = x_0 + 2h \ldots, x_i = x_0 + ih, \ldots x_n = x_0 + nh$, где h – шаг таблицы.

Назовем конечными разностями первого порядка разности

$$\Delta y_0 = y_1 - y_0,$$
$$\Delta y_1 = y_2 - y_1,$$
$$\ldots\ldots\ldots \quad \ldots\ldots\ldots \quad \ldots\ldots$$
$$\Delta y_{n-1} = y_n - y_{n-1}.$$

конечными разностями второго порядка

$$\Delta^2 y_0 = \Delta y_1 - \Delta y_0,$$
$$\Delta^2 y_1 = \Delta y_2 - \Delta y_1,$$
....
$$\Delta^2 y_{n-2} = \Delta y_{n-1} - \Delta y_{n-2}$$

и т.д. Конечные разности *(к+1)*-го порядка вычисляются по формуле

$$\Delta^{k+1} y_i = \Delta^k y_{i+1} - \Delta^k y_i . \qquad (4.19)$$

Конечные разности, как и разделенные, располагаются в таблице.

x_i	$f(x_i)$	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$
x_0	y_0				
		Δy_0			
x_1	y_1		$\Delta^2 y_0$		
		Δy_1		$\Delta^3 y_0$	
x_2	y_2		$\Delta^2 y_1$		$\Delta^4 y_0$
		Δy_2		$\Delta^3 y_1$	
x_3	y_3		$\Delta^2 y_2$		
		Δy_3			
x_4	y_4				

Рассмотрим некоторые свойства конечных разностей.

1. Конечная разность связана с соответствующей разделенной разностью следующим соотношением:

$$\Delta^k y_i = h^k k!\, f\left[x_i, x_{i+1}, \ldots, x_{i+k}\right] . \qquad (4.20)$$

Докажем справедливость этого соотношения методом математической индукции. Для конечных разностей первого порядка имеем

$$\Delta y_i = y_{i+1} - y_i = h\frac{y_{i+1} - y_i}{x_{i+1} - x_i} = h1!\, f\left[x_i, x_{i+1}\right] .$$

Допустим, что соотношение верно для некоторого $k \le m$. Тогда,

$$\Delta^{m+1} y_i = \Delta^m y_{i+1} - \Delta^m y_i = h^m m!\, f[x_{i+1}, \ldots, x_{i+m+1}] -$$
$$- h^m m!\, f[x_i, \ldots, x_{i+m}] = h^m m! \frac{f[x_{i+1}, \ldots, x_{i+m+1}] - f[x_i, \ldots, x_{i+m}]}{x_{i+m+1} - x_i}$$
$$\cdot (m+1)h = h^{m+1}(m+1)!\, f[x_i, \ldots, x_{i+m+1}],$$

что и требовалось доказать.

2. Конечная разность связана с соответствующей производной соотношением

$$\Delta^k y_i = h^k f^{(k)}(\xi), \quad \xi \in [x_i, x_{i+k}] . \qquad (4.21)$$

Это равенство непосредственно следует из только что доказанного соотношения (4.20) и ранее доказанного равенства (4.13).

Как следствие (4.21) получим, что конечные разности порядка n от полинома степени n постоянны и равны $h^n n! a_0$, а конечные разности любого более высокого порядка равны нулю. Однако, этот вывод справедлив лишь для случая, когда исходные значения функции y_i являются точными и конечные разности любого порядка подсчитаны без вычислительных погрешностей.

Поскольку числа y_i, как правило, задаются с некоторой абсолютной погрешностью Δ^*, конечные разности первого порядка будут иметь абсолютную погрешность $2\Delta^*$, конечные разности второго порядка - $4\Delta^*$ и т.д., т.е. конечные разности порядка k будут иметь абсолютную погрешность $2^k\Delta^*$.

Если у функции $f(x)$ производные достаточно высоких порядков остаются ограниченными, то согласно (4.21) соответствующие конечные разности $\Delta^k y_i$ будут убывать с ростом k. Поэтому, естественно, наступит такой момент, когда погрешности конечных разностей станут сравнимы или даже больше абсолютных величин самих конечных разностей. Следовательно, информация, содержащаяся в таблице этих разностей, станет информацией о погрешностях, а не функции, и использование ее станет нецелесообразным. При этом говорят, что порядок последних конечных разностей, которые еще целесообразно использовать в вычислениях, есть порядок правильности таблицы конечных разностей.

4.5. Интерполяционные формулы Ньютона

Рассмотрим интерполяционный полином Ньютона с разделенными разностями (4.18), взяв в качестве узлов интерполирования равноотстоящие точки x_0, $x_1 = x_0 + h$, ..., $x_i = x_0 + ih$, ..., $x_n = x_0 + nh$. Заменяя разделенные разности их выражениями через конечные разности согласно (4.20)

$$f[x_0, x_1, \ldots x_k] = \frac{\Delta^k y_0}{h^k k!},$$

получим

$$N_n(x) = y_0 + \Delta y_0 \frac{x - x_0}{h} + \frac{\Delta^2 y_0}{h^2 2!}(x - x_0)(x - x_1) + \ldots + \frac{\Delta^n y_0}{h^n n!}(x - x_0)\ldots(x - x_{n-1})$$

Введем переменную $t = \dfrac{x - x_0}{h}$. Тогда формула примет вид

$$N_n(x) = N_n^{\,I}(x_0 + ht) = y_0 + \Delta y_0 t + \frac{\Delta^2 y_0}{2!} t(t-1) + \ldots + \frac{\Delta^n y_0}{n!} t(t-1)\ldots[t-(n-1)]. \qquad (4.22)$$

Полученную формулу называют первым интерполяционным полиномом Ньютона или полиномом Ньютона для интерполирования вперед.

Остаточная погрешность значения $N_n^{\,I}(x_0 + ht^*)$ выражается формулой (4.8). Если заменить $x^* = x_0 + ht^*$, то она примет следующий вид:

$$\Delta_1 = \frac{M_{n+1} h^{n+1}}{(n+1)!} \cdot \left| t^*(t^* - 1)\ldots(t^* - n) \right|.$$

На практике величина $M_{n+1} = \max\limits_{[a,b]} \left| f^{(n+1)}(x) \right|$ оценивается согласно (4.21) с помощью конечных разностей *(n+1)*-го порядка

$$M_{n+1} \approx \frac{max \left| \Delta^{n+1} y_i \right|}{h^{n+1}}$$

или Δ_1 определяется абсолютной величиной первого отброшенного слагаемого.

Введем еще одну интерполяционную формулу Ньютона. Для этого запишем полином Ньютона с разделенными разностями (4.18), присоединяя узлы в следующем порядке: $x_n, x_{n+1}, \ldots, x_1, x_0$:

$$N_n(x) = y_n + f[x_n, x_{n+1}](x - x_n) + \ldots + f[x_n, \ldots, x_1, x_0](x - x_n)\ldots(x - x_1)$$

Введем переменную $q = \frac{x - x_n}{h}$. и выразим разделенные разности через конечные.

$$N_n(x) = N_n^{\,II}(x_n + hq) = y_n + \Delta y_{n-1} q + \frac{\Delta^2 y_{n-2}}{2!} q(q+1) + \ldots + \frac{\Delta^n y_0}{n!} q(q+1)\ldots[q+n-1]. \quad (4.23)$$

Эта формула называется вторым интерполяционным полиномом Ньютона, или полиномом Ньютона для интерполирования назад.

Оценка (4.8) остаточной погрешности приближенного значения $N_n^{\,II}(x_n + q^* h)$ представится в виде

$$\Delta_1 = \frac{M_{n+1} h^{n+1}}{(n+1)!} \cdot \left| q^*(q^* + 1)\ldots(q^* + n) \right|, \quad q^* = \frac{x^* - x_n}{h},$$

$$M_{n+1} = \max_{[a,b]} \left| f^{(n+1)}(x) \right| .$$

Итак, получены две новые формулы интерполирования, и далее будут получен еще ряд таких формул. Однако следует заметить, что каждая из них является лишь другой формой записи интерполяционного полинома Лагранжа. Поэтому, если отвлечься от различия в обозначениях и в форме записи, то все эти формулы тождественны, когда они построены по одним и тем же узлам интерполирования. Однако в практике вычислений применяются в различных случаях

разные формулы. Как уже отмечалось, во-первых, дело связано с тем, что обычно бывает удобнее вести вычисления, если при интерполировании сначала используются ближайшие к x^* узлы, а затем подключаются все более удаленные. При этом первые члены интерполяционных формул дадут основной вклад в искомую величину, а остальные будут давать лишь уменьшающиеся (по модулю) добавки. В этом случае легко установить, на какой разности следует закончить вычисления.

Во-вторых, как было отмечено в разделе 4.1, максимальные значения $|\omega_n(x)| = |(x-x_0)\ldots(x-x_n)|$ убывают к середине отрезка, содержащего все узлы, и возрастают к концам его. Поэтому, если имеется возможность при вычислениях для различных x строить интерполяционный полином по различным узлам, то их следует выбирать так, чтобы точка x находилась вблизи середины отрезка, содержащего все узлы интерполирования. В этом смысле мы можем сравнивать по точности различные интерполяционные формулы

4.6. Интерполяционные полиномы с центральными разностями

Возьмем в качестве узлов интерполирования точки $x_0, x_1, x_{-1}, \ldots, x_k, x_{-k}$, где $x_i = x_0 + ih,\ i = 0, \pm 1, \ldots, \pm k$. Построим интерполяционный полином Ньютона с разделенными разностями

$$N_{2k}(x) = y_0 + f[x_0, x_1](x-x_0) + f[x_0, x_1, x_{-1}](x-x_0)(x-x_1) + \ldots$$
$$+ f[x_0, x_1, x_{-1}, x_2, x_{-2}, \ldots, x_k](x-x_0)(x-x_1)(x-x_{-1})\ldots(x-x_{-(k-1)}) +$$
$$+ f[x_0, x_1, x_{-1}, \ldots, x_k, x_{-k}](x-x_0)(x-x_1)(x-x_{-1})\ldots(x-x_k)$$

Используя симметричность разделенных разностей относительно своих аргументов и связь их с конечными разностями (4.20), получим

$$f[x_0, x_1, x_{-1}, \ldots, x_i, x_{-i}] = \frac{\Delta^{2i} y_{-i}}{(2i)!\, h^{2i}},$$

$$f[x_0, x_1, x_{-1}, \ldots, x_{-i}, x_{i+1}] = \frac{\Delta^{2i+1} y_{-i}}{(2i+1)!\, h^{2i+1}}.$$

Отсюда

$$N_{2k}(x) = y_0 + \frac{\Delta y_0}{n}(x - x_0) + \frac{\Delta^2 y_{-1}}{2!h^2}(x - x_0)(x - x_1) + \dots$$
$$+ \frac{\Delta^{2k-1} y_{-(k-1)}}{(2k-1)!h^{2k-1}}(x - x_0)(x - x_1)\dots(x - x_{-(k-1)}) +$$
$$+ \frac{\Delta^{2k} y_{-k}}{(2k)!h^{2k}}(x - x_0)(x - x_1)\dots(x - x_{-(k-1)})(x - x_k).$$

Введя переменную $t = \frac{x - x_0}{h}$, получим первый интерполяционный полином Гаусса, или полином Гаусса для интерполирования вперед,

$$N_{2k}(x) = G_{2k}{}^{I}(x_0 + ht) = y_0 + \Delta y_0 t + \frac{\Delta^2 y_{-1}}{2!}t(t-1) + \frac{\Delta^3 y_{-1}}{3!}t(t^2-1) + \dots$$
$$+ \frac{\Delta^{2k-1} y_{-(k-1)}}{(2k-1)!}t(t^2-1)(t^2-2^2)\dots\left[t^2-(k-1)^2\right] + \qquad (4.24)$$
$$+ \frac{\Delta^{2k} y_{-k}}{(2k)!}t(t^2-1)(t^2-2^2)\dots\left[t^2-(k-1)^2\right](t-k).$$

В этой формуле используются следующие конечные разности (подчеркнуты):

x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$
x_{-3}	y_{-3}				
		Δy_{-3}			
x_{-2}	y_{-2}		$\Delta^2 y_{-3}$		
		Δy_{-2}		$\Delta^3 y_{-3}$	
x_{-1}	y_{-1}		$\Delta^2 y_{-2}$		$\Delta^4 y_{-3}$
		Δy_{-1}		$\Delta^3 y_{-2}$	
x_0	$\underline{y_0}$		$\underline{\Delta^2 y_{-1}}$		$\underline{\Delta^4 y_{-2}}$
		$\underline{\Delta y_0}$		$\underline{\Delta^3 y_{-1}}$	
x_1	y_1		$\Delta^2 y_0$		$\Delta^4 y_{-1}$
		Δy_1		$\Delta^3 y_0$	
x_2	y_2		$\Delta^2 y_1$		
		Δy_2			
x_3	y_3				

Если взять узлы интерполирования в другом порядке, а именно $x_0, x_{-1}, x_1, \ldots, x_{-k}, x_k$, то совершенно аналогично можно получить второй интерполяционный полином Гаусса, или интерполяционный полином Гаусса для интерполирования назад,

$$G_{2k}^{\ II}(x_0+ht) = y_0 + \Delta y_{-1} t + \frac{\Delta^2 y_{-1}}{2!} t(t+1) + \frac{\Delta^3 y_{-2}}{3!} t(t^2-1) +$$
$$\frac{\Delta^4 y_{-2}}{4!} t(t^2-1)(t+2) + \ldots + \frac{\Delta^{2k-1} y_{-k}}{(2k-1)!} t(t^2-1)(t^2-2^2)\ldots\left[t^2-(k-1)^2\right] + \qquad (4.25)$$
$$+ \frac{\Delta^{2k} y_{-k}}{(2k)!} t(t^2-1)(t^2-2^2)\ldots\left[t^2-(k-1)^2\right](t+k).$$

Вторая интерполяционная формула Гаусса использует следующие конечные разности:

x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$
x_{-3}	y_{-3}				
		Δy_{-3}			
x_{-2}	y_{-2}		$\Delta^2 y_{-3}$		
		Δy_{-2}		$\Delta^3 y_{-3}$	
x_{-1}	y_{-1}		$\Delta^2 y_{-2}$		$\Delta^4 y_{-3}$
		$\underline{\Delta y_{-1}}$		$\underline{\Delta^3 y_{-2}}$	
x_0	$\underline{y_0}$		$\underline{\Delta^2 y_{-1}}$		$\underline{\Delta^4 y_{-2}}$
		Δy_0		$\Delta^3 y_{-1}$	
x_1	y_1		$\Delta^2 y_0$		$\Delta^4 y_{-1}$
		Δy_1		$\Delta^3 y_0$	
x_2	y_2		$\Delta^2 y_1$		
		Δy_2			
x_3	y_3				

Взяв полусумму интерполяционных формул Гаусса, получим интерполяционный полином Стирлинга в виде формулы:

$$S_{2k}(x_0+ht)=y_0+\frac{\Delta y_0+\Delta y_{-1}}{2}t+\frac{\Delta^2 y_{-1}}{2!}t^2+\frac{\Delta^3 y_{-1}+\Delta^2 y_{-2}}{2}\frac{t(t^2-1)}{3!}+$$
$$\ldots+\frac{\Delta^{2k-1}y_{-(k-1)}+\Delta^{2k-1}y_{-k}}{2}\frac{t(t^2-1)(t^2-2^2)\ldots\left[t^2-(k-1)^2\right]}{(2k-1)!}+ \qquad (4.26)$$
$$\frac{\Delta^{2k}y_{-k}}{(2k)!}t^2(t^2-1)(t^2-2^2)\ldots\left[t^2-(k-1)^2\right]$$

Интерполяционный полином Стирлинга использует следующие конечные разности:

x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$
x_{-3}	y_{-3}	Δy_{-3}			
x_{-2}	y_{-2}	Δy_{-2}	$\Delta^2 y_{-3}$	$\Delta^3 y_{-3}$	
x_{-1}	y_{-1}	$\underline{\Delta y_{-1}}$	$\Delta^2 y_{-2}$	$\underline{\Delta^3 y_{-2}}$	$\Delta^4 y_{-3}$
x_0	$\underline{y_0}$	$\underline{\Delta y_0}$ 1/2	$\underline{\Delta^2 y_{-1}}$	$\underline{\Delta^3 y_{-1}}$ 1/2	$\underline{\Delta^4 y_{-2}}$
x_1	y_1	Δy_1	$\Delta^2 y_0$	$\Delta^3 y_0$	$\Delta^4 y_{-1}$
x_2	y_2	Δy_2	$\Delta^2 y_1$		
x_3	y_3				

Остаточный член интерполяционных формул (4.24), (4.25) и (4.26) имеет следующий вид

$$R_{2k}(x)=\frac{f^{(2k+1)}(\xi)}{(2k+1)!}(x-x_0)\ldots(x-x_{-k})(x-x_k)=$$
$$=\frac{f^{(2k+1)}(\xi)}{(2k+1)!}h^{2k+1}t(t^2-1)(t^2-2^2)\ldots(t^2-k^2),\ \xi\in[x_{-k},x_k]. \qquad (4.27)$$

Например, для полинома Стирлинга второй степени

$$S_2(x_0+ht)=y_0+\frac{\Delta y_0+\Delta y_{-1}}{2}t+\frac{\Delta^2 y_{-1}}{2!}t^2,$$

остаточный член

$$R_2(x_0+ht)=\frac{f'''(\xi)}{3!}h^3t(t^2-1),\quad \xi\in[x_{-1},x_1].$$

Получим еще одну форму интерполяционного полинома. Для этого применим вторую интерполяционную формулу Гаусса к точке x_1, используя для ее построения узлы $x_1,x_0,x_2,x_{-1},\ldots,x_{-k},x_{k+1}$. Тогда

$$G_{2k+1}^{\ II}(x_1+ht')=y_1+\Delta y_0t'+\frac{\Delta^2y_0}{2!}t'(t'+1)+\frac{\Delta^3y_{-1}}{3!}t'[(t')^2-1]+$$
$$+\frac{\Delta^4y_{-1}}{4!}t'[(t')^2-1](t'+2)+\ldots+\frac{\Delta^{2k}y_{-(k-1)}}{(2k)!}t'[(t')^2-1]\ldots(t'+k)+$$
$$+\frac{\Delta^{2k+1}y_{-k}}{(2k+1)!}t'[(t')^2-1]\ldots[(t')^2-k^2),$$

где $t'=\dfrac{x-x_1}{h}$. Легко видеть, что $t'=t-1$, где $t=\dfrac{x-x_0}{h}$. Выразим в $G_{2k+1}^{\ II}$ t' через t. Получим

$$G_{2k+1}^{\ II}(x_0+ht)=y_1+\Delta y_0(t-1)+\frac{\Delta^2y_0}{2!}t(t-1)+\frac{\Delta^3y_{-1}}{3!}t(t-1)(t-2)+$$
$$+\frac{\Delta^4y_{-1}}{4!}t(t^2-1)(t-2)+\ldots+\frac{\Delta^{2k}y_{-(k-1)}}{(2k)!}t(t^2-1)(t^2-2^2)\ldots$$
$$\ldots\left[t^2-(k-2)^2\right]\left[t-(k-1)\right](t-k)+\frac{\Delta^{2k+1}y_{-k}}{(2k+1)!}t(t^2-1)(t^2-2^2)\ldots$$
$$\ldots\left[t^2-(k-1)^2\right](t-k).$$

Полусумма этой формулы и первой формулы Гаусса (4.24), построенной по узлам $x_0,x_1,x_{-1},\ldots,x_k,x_{-k},x_{k+1}$, даст интерполяционный полином Бесселя:

$$B_{2k+1}(x_0+ht)=\frac{y_0+y_1}{2}+\Delta y_0\left(t-\frac{1}{2}\right)+\frac{\Delta^2y_0+\Delta^2y_{-1}}{2}\frac{t(t-1)}{2!}+$$
$$+\frac{\Delta^3y_{-1}}{3!}t(t-1)\left(t-\frac{1}{2}\right)+\frac{\Delta^4y_{-1}+\Delta^4y_{-2}}{2}\frac{t(t^2-1)(t-2)}{4!}+\ldots+$$
$$+\frac{\Delta^{2k}y_{-(k-1)}+\Delta^{2k}y_{-k}}{2}\frac{t(t^2-1)\ldots\left[t^2-(k-1)^2\right](t-k)}{(2k)!}+ \qquad (4.28)$$
$$+\frac{\Delta^{2k+1}y_{-k}}{(2k+1)!}t(t^2-1)\ldots\left[t^2-(k-1)^2\right](t-k)\left(t-\frac{1}{2}\right).$$

Полином Бесселя особенно удобен для интерполирования на середину, т.е. для $t=\dfrac{1}{2}$. Действительно, в этом случае члены, содержащие разности нечетного порядка, обращаются в нуль. В формуле Бесселя используются следующие разности:

x_i	y_i	Δy_i	$\Delta^2 y_i$	$\Delta^3 y_i$	$\Delta^4 y_i$	$\Delta^5 y_i$
x_{-3}	y_{-3}					
		Δy_{-3}				
x_{-2}	y_{-2}		$\Delta^2 y_{-3}$			
		Δy_{-2}		$\Delta^3 y_{-3}$		
x_{-1}	y_{-1}		$\Delta^2 y_{-2}$		$\Delta^4 y_{-3}$	
		Δy_{-1}		$\Delta^3 y_{-2}$		$\Delta^5 y_{-3}$
x_0	$\underline{y_0}$ 1/2		$\underline{\Delta^2 y_{-1}}$ 1/2		$\underline{\Delta^4 y_{-2}}$ 1/2	
		$\underline{\Delta y_0}$		$\underline{\Delta^3 y_{-1}}$		$\underline{\Delta^5 y_{-2}}$
x_1	$\underline{y_1}$		$\underline{\Delta^2 y_0}$		$\underline{\Delta^4 y_{-1}}$	
		Δy_1		$\Delta^3 y_0$		$\Delta^5 y_{-1}$
x_2	y_2		$\Delta^2 y_1$		$\Delta^4 y_0$	
		Δy_2		$\Delta^3 y_1$		
x_3	y_3		$\Delta^2 y_2$			
		Δy_3				
x_4	y_4					

Остаточный член интерполяционного полинома Бесселя имеет вид

$$R_{2k+1}(x) = \frac{f^{(2k+2)}(\xi)}{(2k+2)!}(x-x_0)(x-x_1)(x-x_{-1})\ldots(x-x_{-k})(x-x_{k+1}) =$$

$$= \frac{f^{(2k+2)}(\xi)}{(2k+2)!}h^{2k+2}t(t^2-1)(t^2-2^2)\ldots(t^2-k^2)(t-k-1), \qquad (4.29)$$

$$\xi \in [x_{-k}, x_{k+1}].$$

В частности, для полинома Бесселя первой и третьей степени

$$B_1(x_0+ht) = \frac{y_0+y_1}{2} + \Delta y_0\left(t-\frac{1}{2}\right),$$

$$B_3(x_0+ht) = \frac{y_0+y_1}{2} + \Delta y_0\left(t-\frac{1}{2}\right) + \frac{\Delta^2 y_0 + \Delta^2 y_{-1}}{2}\frac{t(t-1)}{2!} +$$

$$+ \frac{\Delta^3 y_{-1}}{3!}t(t-1)\left(t-\frac{1}{2}\right)$$

остаточные члены имеют вид

$$R_1(x_0+ht)=h^2\frac{f''(\xi)}{2!}t(t-1),\quad \xi\in[x_0,x_1];$$

$$R_3(x_0+ht)=h^4\frac{f^{IV}(\xi)}{4!}t(t^2-1)(t-2),\quad \xi\in[x_{-1},x_2].$$

4.8.Обратное интерполирование

В практике вычислений часто встречается следующая задача. Функция *f(x)* задана своими значениями y_i в точках $x_i\in[a,b], i=0,1,\ldots,n$. Требуется определить значение аргумента $x^*\in[x_i,x_{i+1}]$, соответствующее заданному значению $y^*\in[y_i,y_{i+1}]$ функции *f(x)*, т.е. найти корень уравнения

$$f(x)=y^*, \qquad (4.30)$$

принадлежащий интервалу (x_i,x_{i+1}). Предполагается, что интервал (x_i,x_{i+1}) настолько мал, что $\sqrt{x^*}$ – единственный.

Поставленная задача называется задачей обратного интерполирования. Один из возможных путей решения этой задачи заключается в следующем. Функцию *f(x)* аппроксимируем ее интерполяционным полиномом $P_n(x)$, а уравнение (4.30) заменяем уравнением

$$P_n(x)=y^*. \qquad (4.31)$$

Находим действительный корень $\bar{x}$ уравнения (4.31), принадлежащий интервалу (x_i,x_{i+1}). Практически мы получаем лишь приближенное решение уравнения (4.31) – точку $\bar{\bar{x}}$. И теперь полагаем, что $x^*\approx\bar{\bar{x}}$.

Оценим погрешность такого решения. Пусть суммарная погрешность интерполирования есть Δ, т.е.

$$\left|f(\bar{x})-P_n(\bar{x})\right|\le\Delta, \qquad (4.32)$$

а погрешность решения уравнения (4.31) есть ε, т.е.

$$\left|\bar{x}-\bar{\bar{x}}\right|\le\varepsilon. \qquad (4.33)$$

Тогда приращение функции в точке $\bar{x}$ можно представить как

$$y^*-f(\bar{x})=f(x^*)-f(\bar{x})=f'(\xi)(x^*-\bar{x}),$$

$$\xi=\bar{x}+\theta(x^*-\bar{x}),\quad \theta\in(0,1).$$

Отсюда, принимая во внимание, что

$$P_n(\bar{x})=y^*,$$

имеем

$$P_n(\bar{x}) - f(\bar{x}) = f'(\xi)(x^* - \bar{x}).$$

Предположив теперь, что

$$\min_{[x_i, x_{i+1}]} |f'(x)| \geq m_1 > 0,$$

и используя оценку (4.32), получим

$$\left|x^* - \bar{x}\right| \leq \frac{\Delta}{m_1}. \qquad (4.34)$$

Далее,

$$\left|x^* - \bar{\bar{x}}\right| = \left|x^* - \bar{x} + \bar{x} - \bar{\bar{x}}\right| \leq \left|x^* - \bar{x}\right| + \left|\bar{x} - \bar{\bar{x}}\right|.$$

Следуя оценкам (4.33) и (4.34), окончательно находим

$$\left|x^* - \bar{\bar{x}}\right| \leq \frac{\Delta}{m_1} + \varepsilon. \qquad (4.35)$$

Таким образом, как решение задачи обратного интерполирования, так и погрешность (4.35) определяются двумя процессами: построением интерполяционного полинома и решением уравнения (4.31), т.е. нахождением корней интерполяционного полинома.

Может показаться, что эти два момента ничем не связаны между собой. Однако это совсем не так. Следует иметь в виду, что увеличение степени полинома, с одной стороны, уменьшает погрешность Δ, с другой – увеличивает трудоемкость решения уравнения (4.31). Поэтому степень интерполяционного полинома должна быть наименьшей при условии достижения требуемой точности [1-3].

При практическом решении задачи обратного интерполирования на равномерной сетке узлов в качестве интерполяционных полиномов обычно используются полиномы Стирлинга и Бесселя. В этом случае уравнение (4.31), записанное с переменной $t = \frac{x - x_0}{h}$, приводится к виду $t = \varphi(t)$ и решается методом итераций.

Например, при использовании полинома Стирлинга имеем

$$t = \frac{2}{\Delta y_0 + \Delta y_{-1}} \left[\begin{array}{l} y^* - y_0 - \frac{\Delta^2 y_{-1}}{2!} t^2 - \frac{\Delta^3 y_{-1} + \Delta^3 y_{-2}}{2} \cdot \frac{t(t^2-1)}{3!} - \\ - \frac{\Delta^4 y_{-2}}{4!} t^2 (t^2 - 1) \end{array} \right]. \qquad (4.36)$$

В качестве начального приближения t_0 обычно принимается $t_0=0$. После того, как t^* – решение уравнения (4.36) - получено, x^* определяется по формуле

$$x^* = x_0 + ht^*.$$

Аналогичным образом можно получить решение задачи обратного интерполирования при помощи полинома Бесселя или первого и второго интерполяционных полиномов Ньютона.

Рассмотрим еще один подход к решению задачи обратного интерполирования, основанный на существовании гладкой функции *g(y)*, обратной к *f(x)*.

Пусть функция *g(y)* непрерывна вместе с достаточным количеством своих производных на минимальном интервале, содержащем значения $y_i = f(x_i),\ \ i = 0, \pm 1, \ldots\ \ y^* = f(x^*)$. В этом случае определение x^* эквивалентно вычислению обратной функции *g(y)*, заданной своими значениями x_i в узлах y_i, в точке $y = y^*$, так как

$$x^* = g(y^*).$$

Таким образом, задача обратного интерполирования сведена к задаче интерполирования обратной функции *g(y)*.

Например, если обратную функцию *g(y)* приближать интерполяционным полиномом Лагранжа, то решение поставленной задачи в этом случае будет иметь вид

$$x^* \approx L_n(y^*) = \sum_{i=0}^{n} x_i \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{(y^* - y_j)}{(y_i - y_j)}.$$

Оценка остаточного члена будет такая же, как и при прямом интерполировании:

$$x^* - L_n(y^*) = \frac{g^{(n+1)}(\xi)}{(n+1)!}(y^* - y_0)(y^* - y_1)\ldots(y^* - y_n),$$

где $g^{(n+1)}(\xi)$ – производная *(n+1)*-го порядка обратной функции в точке ξ, принадлежащей интервалу между минимальным и максимальным $y_i,\ i=0,1,\ldots,n$. Оценка же вычислительной погрешности усложнится, поскольку теперь выражение $L_n(y^*)$ нелинейно зависит от приближенных величин $y_i,\ i=0,1,\ldots,n$.

Приведенный способ решения задачи обратного интерполирования является более эффективным, нежели прием, содержащий одним из этапов решение уравнения. Особенно он удобен, если решение задачи обратного интерполирования требуется найти для достаточно большого числа значений y^* или когда требуется получить явное выражение для корня уравнения (4.30). Недостатком рассмотренного метода является требование гладкой обратной функции, что далеко не всегда выполнимо (например, это требование не выполняется для немонотонных функций).

4.9. Численное дифференцирование

К численному дифференцированию приходится прибегать в том случае, когда функция *f(x)*, для которой нужно найти производную, задана таблично или же имеет сложное аналитическое выражение. В первом случае методы дифференциального исчисления просто неприменимы, а во втором случае их использование вызывает значительные трудности.

Одним из способов построения формул численного дифференцирования является дифференцирование интерполяционных полиномов. Пусть известны значения функции *f(x)* в точках $x_0, x_1, \ldots, x_n$. Требуется вычислить $f^{(m)}(x), \quad m \le n$. Построим интерполяционный полином $L_n(x)$ и положим

$$f^{(m)}(x) \approx L^{(m)}{}_n(x). \qquad (4.37)$$

Точно так же мы можем заменять значения производных функций значениями производных других интерполяционных полиномов: Стирлинга, Бесселя и т.д. Можно показать [3], что остаточный член формул численного дифференцирования (4.37) имеет следующий вид:

$$R_n{}^{(m)}(x) = f^{(m)}(x) - L^{(m)}{}_n(x) = \\ = \sum_{j=0}^{m} \frac{m!}{(m-j)!(n+j+1)!} f^{(n+j+1)}(\xi_i)\omega_n{}^{(m-j)}(x), \qquad (4.38)$$

где

$$\omega_n(x) = (x - x_0)(x - x_1)\ldots(x - x_n), \quad \omega_n{}^{(m-j)} = \frac{d^{(m-j)}}{dx^{m-j}}\omega_n(x),$$

а ξ_i – некоторые точки из интервала между наименьшим и наибольшим из чисел $x, x_0, x_1, \ldots, x_n$.

Пусть функция задана на равномерной сетке узлов с шагом *h*. Взяв интерполяционный полином Стирлинга, построенный по точкам $x_0, x_i = x_0 + ih, \quad i = \pm 1, \pm 2, \ldots, \pm k$, продифференцируем его один раз. Получим следующую формулу для первой производной:

$$f'(x) \approx \frac{dS_{2k}(x_0 + ht)}{dx} = \frac{dt}{dx} \cdot \frac{d}{dt}\left[y_0 + \frac{\Delta y_0 + \Delta y_{-1}}{2} t + \frac{\Delta^2 y_{-1}}{2!} t^2 + \frac{\Delta^3 y_{-1} + \Delta^3 y_{-2}}{2} \cdot \right.$$

$$\left. \cdot \frac{t(t^2-1)}{3!} + \frac{\Delta^4 y_{-2}}{4!} t^2(t^2-1) + \ldots \right] = \frac{1}{h}\left[\frac{\Delta y_0 + \Delta y_{-1}}{2} + \Delta^2 y_{-1} t + \frac{\Delta^3 y_{-1} + \Delta^3 y_{-2}}{12}(3t^2 - 1) + \right. \qquad (4.39)$$

$$\left. + \frac{\Delta^4 y_{-2}}{12}(2t^3 - t) + \ldots \right]$$

где $t = \dfrac{x - x_0}{h}$.

Для второй производной, дифференцируя по *x* (4.39), получим

$$f''(x) \approx \frac{1}{h^2}\left[\Delta^2 y_{-1} + \frac{\Delta^3 y_{-1} + \Delta^3 y_{-2}}{2} t + \frac{\Delta^4 y_{-2}}{12}(6t^2 - 1) + \ldots\right] \quad (4.40)$$

В частности, при $x=x_0$ $(t=0)$ будем иметь

$$f'(x_0) \approx \frac{1}{h}\left[\frac{\Delta y_0 + \Delta y_{-1}}{2} - \frac{\Delta^3 y_{-1} + \Delta^3 y_{-2}}{12} + \ldots\right] \quad (4.41)$$

$$f''(x_0) \approx \frac{1}{h^2}\left[\Delta^2 y_{-1} - \frac{\Delta^4 y_{-2}}{12} + \ldots\right] \quad (4.42)$$

В некоторых случаях выгоднее выражать производные в узловых точках не через конечные разности, а непосредственно через значения функции. Преобразуем к такому виду формулы (4.41) и (4.42).

Если в формулах (4.41) и (4.42) ограничиться одним слагаемым, что соответствует полиному Стирлинга второй степени, то получим соответственно

$$f'(x_0) \approx \frac{1}{h} \cdot \frac{\Delta y_0 + \Delta y_{-1}}{2} = \frac{y_1 - y_{-1}}{2h} \quad ; \quad (4.43)$$

$$f''(x_0) \approx \frac{\Delta^2 y_{-1}}{h^2} = \frac{y_{-1} - 2y_0 + y_1}{h^2}. \quad (4.44)$$

Взяв в формулах (4.41) и (4.42) по два слагаемых (полином Стирлинга четвертой степени), будем соответственно иметь

$$f'(x_0) \approx \frac{y_{-2} - 8y_{-1} + 8y_1 - y_2}{12h}; \quad (4.45)$$

$$f''(x_0) \approx \frac{-y_2 + 16y_{-1} - 30y_0 + 16y_1 - y_2}{12h^2}. \quad (4.46)$$

Получим остаточный член формулы численного дифференцирования (4.41). Для этого продифференцируем по x остаточный член полинома Стирлинга степени $2k$ и подставим $x=x_0$:

$$\frac{d}{dx} R_{2k}(x)\Big|_{x=x_0} = \frac{dt}{dx} \cdot \frac{d}{dx}\left[\frac{f^{(2k+1)}(\xi(x))}{(2k+1)!} h^{2k+1} t(t^2 - 1)\ldots(t^2 - k^2)\right]_{t=0} =$$

$$= \frac{f^{(2k+1)}(\xi)}{(2k+1)!}(-1)^k [k!]^2 h^{2k}, \quad \xi \in [a,b]. \quad (4.47)$$

Для формул (4.43) и (4.45) остаточный член (4.47) будет соответственно иметь вид

$$-\frac{h^2 f'''(\xi_1)}{6} \quad и \quad \frac{h^4 f^V(\xi_2)}{30}.$$

Исследуем полную погрешность формул численного дифференцирования, например, для формулы (4.43)

$$\Delta полн.(h) = \frac{h^2 M_3}{6} + \frac{\Delta^*}{h}, \quad (4.48)$$

где $M_3 = \max\limits_{[a,b]} |f'''(x)|$,

Δ^* – абсолютная погрешность каждого из чисел y_i.

В (4.48) первое слагаемое (остаточная погрешность) убывает с уменьшением h, а второе (вычислительная погрешность) возрастает с уменьшением h. Возникает вопрос о подборе для данной формулы численного дифференцирования оптимального шага h^*, для которого полная погрешность имела бы минимальное значение. Найдем такой шаг

$$\Delta' полн.(h) = \frac{2hM_3}{6} - \frac{\Delta^*}{h^2} = 0,$$

откуда

$$h^* = \sqrt[3]{\frac{3\Delta^*}{M_3}}.$$

В точке $h = h^*$ функция $\Delta полн.(h)$ имеет действительно минимальное значение, поскольку

$$\Delta'' полн.(h) = \frac{M_3}{3} + \frac{\Delta^*}{h^2} > 0.$$

При вычислении второй производной или производных более высокого порядка, когда в знаменатель соответствующей формулы численного дифференцирования входит h^2 или h^k и $k>2$, вопрос о выборе оптимального шага является еще более актуальным.

5. ИНТЕРПОЛИРОВАНИЕ С КРАТНЫМИ УЗЛАМИ И СПЛАЙНЫ

Рассмотрим теперь более общую постановку задачи интерполирования полиномами.

В узлах $x_i \in [a,b], \quad i = 0,1,\ldots,n$, среди которых нет совпадающих, известны значения функции $f(x_i)$ и ее производных $f^{(j)}(x_i)$ до порядка k_i-1 включительно, $j = 0,1,\ldots,k_i - 1$. Таким образом, информация о функции $f(x)$ задается следующим образом:

$$\begin{array}{cccc} f(x_0) & f(x_1) & \ldots & f(x_n); \\ f'(x_0) & f'(x_1) & \ldots & f'(x_n); \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ f_{(x_0)}^{(k_0-1)} & f_{(x_1)}^{(k_1-1)} & \ldots & f_{(x_n)}^{(k_n-1)}. \end{array} \qquad (5.1)$$

Здесь значения $k_i \geq 1$ для различных i, вообще говоря, различны, но допустим случай, когда $k_0 = k_1 = \ldots = k_n$. Следовательно, всего задано $k_0 + k_1 + \ldots + k_n$ величин. Требуется построить алгебраический многочлен $H_m(x)$ степени $m = k_0 + k_1 + \ldots + k_n - 1$, для которого выполняются условия

$$H_m{}^{(j)}(x_i) = f^{(j)}(x_i), \quad j = 0,1,\ldots,k_i - 1; \quad i = 0,1,\ldots,n. \qquad (5.2)$$

Многочлен $H_m(x)$, удовлетворяющий условиям (5.2), называется интерполяционным полиномом Эрмита для функции $f(x)$ или интерполяционным полиномом с кратными узлами. Числа $k_0, k_1, \ldots, k_n$ называются кратностями узлов $x_0, x_1, \ldots, x_n$ соответственно.

Интерполяционный полином $H_m(x)$ определяется единственным образом. В самом деле, предположив противное, будем иметь два полинома степени m, удовлетворяющих условию (5.2). Тогда их разность $Q_m(x)$ удовлетворяет соотношениям

$$Q_m(x_i) = Q'_m(x_i) = \ldots = Q_m{}^{(k_i-1)}(x_i) = 0, \quad i = 0,1,\ldots,n$$

т.е. точки $x_0, x_1, \ldots, x_n$ являются корнями полинома $Q_m(x)$ кратности $k_0, k_1, \ldots, k_n$ соответственно. Мы получили, что многочлен $Q_m(x)$ степени m имеет $m+1$ корней. Следовательно, $Q_m(x) \equiv 0$.

Существование интерполяционного полинома Эрмита $H_m(x)$ докажем, получив для него явное выражение. Далее предположим, что функция $f(x)$ непрерывно дифференцируема $(m+1)$ раз.

5.1. Разделенные разности с повторяющимися (кратными) узлами

Зададимся последовательностью совокупностей точек

$$\begin{array}{llll} x_0 & x_0^{(1)} & x_0^{(2)} & \dots \quad x_0^{(k_0-1)}; \\ x_1 & x_1^{(1)} & x_1^{(2)} & \dots \quad x_1^{(k_1-1)}; \\ \dots & \dots & \dots & \dots \quad \dots \\ x_n & x_n^{(1)} & x_n^{(2)} & \dots \quad x_n^{(k_n-1)}, \end{array}$$

удовлетворяющих условию: все точки $x_i^{(j)}$ – различны. В частности, можно положить

$$x_i^{(j)} = x_i + (j-1)\varepsilon, \quad i = 0,1,\dots,n, \quad j = 1,2,\dots,k_i - 1,$$

где $\varepsilon > 0$ – малая величина. Построим по всем этим точкам разделенную разность порядка $m = k_0 + k_1 + \dots + k_n - 1$. Определим

$$f\left[\underbrace{x_0, x_0, \dots, x_0}_{k_0 \text{ раз}}; \underbrace{x_1, x_1, \dots, x_1}_{k_1 \text{ раз}}; \dots; \underbrace{x_n, x_n, \dots, x_n}_{k_n \text{ раз}}\right] = \tag{5.3}$$

$$= \lim_{x_i^{(j)} \to x_i (\varepsilon \to 0)} f\left[x_0, x_0^{(1)}, \dots, x_0^{(k_0-1)}; \dots; x_n, x_n^{(1)}, \dots, x_n^{(k_n-1)}\right].$$

Рассмотрим сначала случай, когда под знаком разделенной разности левой части (5.3) повторяется только один узел x_i и разделенная разность порядка k_i-1 вычисляется только по этому повторяющемуся узлу. Согласно определению (5.3)

$$f\left[\underbrace{x_i, x_i, \dots, x_i}_{k_i \text{ раз}}\right] = \lim_{x_i^{(j)} \to x_i} f\left[x_i, x_i^{(1)}, \dots, x_i^{(k_i-1)}\right]$$

По формуле связи (4.13) между разделенной разностью и производной имеем

$$f\left[x_i, x_i^{(1)}, \dots, x_i^{(k_i-1)}\right] = \frac{f^{(k_i-1)}(\xi)}{(k_i-1)!}, \tag{5.4}$$

где ξ – точка, принадлежащая наименьшему отрезку, содержащему все точки $x_i, x_i^{(1)}, \dots, x_i^{(k_i-1)}$. Перейдя в равенстве (5.4) к пределу при $x_i^{(j)} \to x_i$, получим

$$f\left[\underbrace{x_i, x_i, \dots, x_i}_{k_i \text{ раз}}\right] = \frac{f^{(k_i-1)}(x_i)}{(k_i-1)!}. \tag{5.5}$$

Итак, если при $i = 0,1,\dots,n$ производная $f^{(k_i-1)}(x)$ непрерывна, то существуют разделенные разности

$$f\left[\underbrace{x_i, x_i, \dots, x_i}_{k_i \text{ раз}}\right], \qquad i = 1,2,\dots,n.$$

Но это обеспечивает также существование разделенной разности с кратными узлами левой части (5.3), т.к. все остальные разделенные разности, необходимые для ее вычисления, находятся путем последовательного применения рекуррентных формул

$$f[x_0, x_1, \ldots, x_k] = \frac{f[x_1, x_2, \ldots, x_k] - f[x_0, x_1, \ldots, x_{k-1}]}{x_k - x_0}$$

и их обобщений. Чтобы не проводить громоздкого вывода для общего случая формулы (5.3), рассмотрим иллюстративную таблицу. Приведенные в этой таблице вычисления переносятся на общий случай без всяких принципиальных затруднений.

Требуется найти $f[x_0, x_0; x_1, x_1; x_2, x_2, x_2]$, если заданы $f(x_0), f'(x_0); f(x_1), f'(x_1); f(x_2), f'(x_2), f''(x_2)$.

№ строк	1 x_i	2 $f(x_i)$	3 $f[x_i, x_j]$	4 Разности II порядка	5 III пор.	6 IV пор	7 V пор	8 VI пор
0	x_0	$f(x_0)$						
1			$\frac{f'(x_0)}{1!}$					
2	x_0	$f(x_0)$		$f[x_0; x_0, x_1]$				
3			$f[x_0, x_1]$		$f[x_0, x_0; x_1, x_1]$			
4	x_1	$f(x_1)$		$f[x_0; x_1, x_1]$		$f[x_0, x_0; x_1, x_1; x_2]$		
5			$f'(x_1)$		$f[x_0, x_1; x_1; x_2]$		$f[x_0, x_0; x_1; x_1; x_2, x_2]$	
6	x_1	$f(x_1)$		$f[x_1, x_1; x_2]$		$f[x_0; x_1, x_1; x_2; x_2]$		$f[x_0, x_0; x_1; x_1; x_2, x_2, x_2]$
7			$f[x_1, x_2]$		$f[x_1, x_1; x_2, x_2]$		$f[x_0; x_1, x_1; x_2, x_2, x_2]$	
8	x_2	$f(x_2)$		$f[x_1; x_2, x_2]$		$f[x_1, x_1; x_2, x_2, x_2]$		
9			$f'(x_2)$		$f[x_1; x_2; x_2, x_2]$			
10	x_2	$f(x_2)$		$\frac{f''(x_2)}{2!}$				
11			$f'(x_2)$					
12	x_2	$f(x_2)$						

Левый столбец таблицы – для нумерации строк, верхняя строка – для нумерации столбцов. В первом столбце в строках с четным номером приведены аргументы искомой разделенной разности. Во втором столбце в тех строках, что и аргументы, помещены соответствующие значения функции. Третий столбец предназначен для разделенных разностей первого порядка. Они размещаются в строках с нечетными номерами между строк, в которых находятся соответствующие узлы (аргументы) и значения функции. Если узлы повторяются, как это имеет место для строк 1, 5, 9, 11, то сюда помещают значение первой производной. В строках 3, 7 помещены обычные разделенные разности

первого порядка. Столбец 4 предназначен для разделенных разностей второго порядка. За исключением последней из них (строка 10), где

$$f[x_2, x_2, x_2] = \frac{f''(x_2)}{2!},$$

они находятся обычным способом по рекуррентной формуле. Так,

$$f[x_0, x_0, x_1] = \frac{f[x_0, x_1] - f[x_0, x_0]}{x_1 - x_0} = \frac{f[x_0, x_1] - f'(x_0)}{x_1 - x_0}.$$

Аналогично и для остальных разностей. В пятом, шестом, седьмом и восьмом столбцах находятся, соответственно, разделенные разности третьего, четвертого, пятого и шестого порядков. Они вычисляются по обычным рекуррентным формулам. Например,

$$f[x_0, x_0; x_1, x_1; x_2, x_2, x_2] = \frac{f[x_0; x_1, x_1; x_2, x_2, x_2] - f[x_0, x_0; x_1, x_1; x_2, x_2]}{x_2 - x_0}.$$

5.2. Интерполяционный полином Эрмита

Перейдем теперь к задаче построения полинома Эрмита. Для этого, как и при определении разделенных разностей с кратными узлами, наряду с данными точками $x_0, x_1, \ldots, x_n$ выберем на отрезке *[a,b]* точки $x_i^{(j)}$, $j = 1,2,\ldots,k_{i-1}$; $i = 0,1,\ldots,n$. Все эти узлы различны. Построим по совокупности $m+1 = k_0 + k_1 + \ldots + k_n$ точек интерполяционный полином Ньютона с разделенными разностями

$$N_m(x) = f(x_0) + (x - x_0) f\left[x_0, x_0^{(1)}\right] + (x - x_0)(x - x_0^{(1)}) f\left[x_0, x_0^{(1)}, x_0^{(2)}\right] +$$
$$\ldots + (x - x_0)(x - x_0^{(1)})\ldots(x - x_0^{(k_0-2)}) f\left[x_0, x_0^{(1)}, \ldots, x_0^{(k_0-1)}\right] + \ldots$$
$$+ (x - x_0)\ldots(x - x_0^{(k_0-1)})(x - x_1)\ldots(x - x_1^{(k_1-1)})\ldots(x - x_n^{(k_n-2)}) \cdot$$
$$\cdot f\left[x_0, \ldots, x_n^{(k_n-1)}\right]$$

Перейдем в обеих частях этого равенства к пределу при $x_i^{(j)} \to x_i$. Получим

$$H_m(x) = f(x_0) + (x - x_0) f'(x_0) + (x - x_0)^2 \frac{f''(x_0)}{2!} + \ldots$$
$$+ (x - x_0)^{k_0-1} \frac{f^{k_0-1}(x_0)}{(k_0 - 1)!} + (x - x_0)^{k_0} f\left[\underbrace{x_0, x_0, \ldots, x_0}_{k_0 \text{ раз}}; x_1\right] + \ldots \qquad (5.6)$$
$$+ (x - x_0)^{k_0} (x - x_1)^{k_1} \ldots (x - x_n)^{k_n - 1} f\left[\underbrace{x_0, x_0, \ldots, x_0}_{k_0 \text{ раз}}; \ldots; \underbrace{x_n, x_n, \ldots, x_n}_{k_n \text{ раз}}\right]$$

Покажем, что полученный таким образом полином $H_m(x)$ решает поставленную задачу, т.е. удовлетворяет условиям (5.2). Первые k_0 членов правой части (5.6) являются первыми k_0 членами разложения

функции *f(x)* в ряд Тейлора. Остальные же члены содержат множитель $(x-x_0)^{k_0}$. Поэтому выполняются условия (5.2), относящиеся к узлу x_0. Но мы могли бы записать $N_m(x)$, взяв за начальный узел не x_0, а любую из точек $x_1, x_2, \ldots, x_n$. При этом ни сам многочлен, ни его предел не изменятся, изменится только форма записи этих многочленов. Таким образом, условия (5.2) будут выполнены и для остальных узлов.

Остаточный член полинома Эрмита получится из остаточного члена полинома $N_m(x)$ переходом к пределу при $x_i^{(j)} \to x_i, i = 0,1,\ldots,n$:

$$\lim_{x_i^{(j)} \to x_i} R_m(x) = \lim_{x_i^{(j)} \to x_i} \left[\frac{f^{(m+1)}(\xi)}{(m+1)!}(x-x_0)\ldots(x-x_0^{(k_0-1)})\ldots \right.$$
$$\left. \ldots(x-x_n)\ldots(x-x_n^{(k_n-1)}) \right] =$$
$$= \frac{f^{(m+1)}(\alpha)}{(m+1)!}(x-x_0)^{k_0}(x-x_1)^{k_1}\ldots(x-x_n)^{k_n},$$

и остаточная погрешность определится как

$$\Delta_1 = \frac{M_{m+1}}{(m+1)!}\left|(x-x_0)^{k_0}(x-x_1)^{k_1}\ldots(x-x_n)^{k_n}\right|, \; M_{m+1} = \max_{[a,b]}\left|f^{(m+1)}(x)\right|.$$

Интерполяционный полином Эрмита можно получить другим способом. Наряду с $H_m(x)$ рассмотрим интерполяционный полином Лагранжа $L_n(x)$, принимающий в точках $x_0, x_1, \ldots, x_n$ значения $f(x_0), f(x_1), \ldots, f(x_n)$. Разность $H_m(x) - L_n(x)$ должна быть многочленом степени не выше *m*, обращающимся в нуль в точках $x_0, x_1, \ldots, x_n$. Следовательно,

$$H_m(x) - L_n(x) = \omega_{n+1}(x)H_{m-n-1}(x),$$

Где $\omega_{n+1}(x) = (x-x_0)(x-x_1)\ldots(x-x_n)$, а $H_{m-n-1}(x)$ – многочлен степени *(m-n-1)*. При любом $H_{m-n-1}(x)$ функция

$$H_m(x) = L_n(x) + \omega_n(x)H_{m-n-1}(x)$$

принимает в узлах интерполирования x_i значения $f(x_i)$. Подберем теперь $H_{m-n-1}(x)$ так, чтобы были выполнены и остальные условия (5.2). Дифференцируя последнее равенство, получим

$$H_m'(x) = L_n'(x) + \omega_{n+1}'(x)H_{m-n-1}(x) + \omega_{n+1}(x)H'_{m-n-1}(x).$$

Полагая здесь $x = x_i$, будем иметь

$$H_m'(x_i) = L_n'(x_i) + \omega_{n+1}'(x_i)H_{m-n-1}(x_i).$$

Так как $\omega'_{n+1}(x_i) \neq 0$, в каждой точке, в которой задана величина $H'_m(x_i) = f'(x_i)$, мы найдем $H_{m-n-1}(x_i)$.

Дифференцируя еще раз, получим

$$H_m''(x) = L_n''(x) + \omega_{n+1}''(x)H_{m-n-1}(x) + 2\omega'_{n+1}(x)H'_{m-n-1}(x) +$$
$$+ \omega_{n+1}(x)H''_{m-n-1}(x).$$

Полагая снова $x = x_i$, найдем

$$H_m''(x_i) = L_n''(x_i) + \omega_n''(x_i) H_{m-n-1}(x_i) + 2\omega'_n(x_i) H'_{m-n-1}(x)$$

Из этого равенства мы сумеем найти $H'_{m-n-1}(x)$ в тех точках, в которых заданы $H''_m(x_i) = f''(x_i)$. Продолжим этот процесс далее. Каждый раз коэффициентом при старшей производной от $H_{m-n-1}(x)$ в точках x_i будет $\omega'_{n+1}(x_i)$. Таким образом, мы сведем нашу задачу отыскания $H_m(x)$ к задаче отыскания $H_{m-n-1}(x)$, удовлетворяющего условиям:

$$\begin{array}{llll}
H_{m-n-1}(x_0) = z_0; & H_{m-n-1}(x_1) = z_1; & \cdots; & H_{m-n-1}(x_n) = z_n; \\
H'_{m-n-1}(x_0) = z'_0; & H'_{m-n-1}(x_1) = z'_1; & \cdots; & H'_{m-n-1}(x_n) = z_n; \\
\cdots\cdots\cdots & \cdots\cdots\cdots & \cdots & \cdots\cdots\cdots \\
H^{(k_0-2)}_{m-n-1}(x_0) = z_0^{(k_0-2)}; & H^{(k_1-2)}_{m-n-1}(x_1) = z_1^{(k_1-2)}; & \cdots; & H^{(k_n-2)}_{m-n-1}(x_n) = z_n^{(k_n-2)},
\end{array}$$

где z_i^j, $j = 0,1,\ldots,k_i - 2$, $i = 0,1,\ldots,n$, - известные числа. Для построения $H_{m-n-1}(x)$ применим точно такой же прием. Получим некоторые условия, наложенные на $H_{m-n-1}(x)$, где $m_1 = k_0 + k_1 + \ldots + k_n - n - 1$. В конце концов, нам потребуется построить интерполяционный полином Лагранжа по его значениям в некоторых из точек x_i.

На практике полином Эрмита часто записывают в различных формах, которые определяются количеством заданных узлов и их кратностью. Например, полином Эрмита третьей степени, построенный по точкам x_0, x_1, в которых заданы еще значения первой производной функции, можно записать в виде

$$H_3(x) = F_1(x) \cdot f(x_0) + F_2(x) \cdot f(x_1) + F_3(x) \cdot f'(x_0) + F_4(x) \cdot f'(x_1), \qquad (5.7)$$

где $F_1(x), F_2(x), F_3(x), F_4(x)$ – полиномы третьей степени, удовлетворяющие условиям:

$$\begin{array}{llll}
F_1(x_0) = 1; & F_2(x_0) = 0; & F_3(x_0) = 0; & F_4(x_0) = 0; \\
F_1(x_1) = 0; & F_2(x_1) = 1; & F_3(x_1) = 0; & F_4(x_1) = 0; \\
F_1'(x_0) = 0; & F_2'(x_0) = 0; & F_3'(x_0) = 1; & F_4'(x_0) = 0; \\
F_1'(x_1) = 0; & F_2'(x_1) = 0; & F_3'(x_1) = 0; & F_4'(x_1) = 1.
\end{array} \qquad (5.8)$$

Очевидно, что $H_3(x)$, определяемый формулой (5.7), удовлетворяет (5.2):

$$H_3(x_0) = f(x_0); \quad H_3(x_1) = f(x_1);$$
$$H_3'(x_0) = f'(x_0); \quad H_3'(x_1) = f'(x_1).$$

Иногда интерполяционный многочлен Эрмита строится методом неопределенных коэффициентов, т.е. рассматривается многочлен

$$H_m(x) = C_0 x^m + C_1 x^{m-1} + \ldots + C_m$$

и коэффициенты $C_0, C_1, \ldots, C_m$ определяются из условий (5.2).

Вычислительная погрешность интерполяционного полинома Эрмита в точке x для каждой из его форм определяется так же, как и для интерполяционных полиномов Лагранжа, Ньютона и т.д. Например, для (5.7) вычислительная погрешность

$$\Delta_2 = |F_1(x)|\Delta^* + |F_2(x)|\Delta^{**} + |F_3(x)|\overline{\Delta^*} + |F_4(x)|\overline{\Delta^{**}}$$

Где $\Delta^*, \Delta^{**}, \overline{\Delta^*}, \overline{\Delta^{**}}$ – абсолютные погрешности величин $f(x_0), f(x_1), f'(x_0), f'(x_1)$ соответственно.

5.3. Интерполирование сплайнами

Интерполирование многочленом Лагранжа, Ньютона или их модификациями на всем отрезке *[a, b]* с использованием большого числа узлов часто приводит к плохому приближению из-за большой вычислительной погрешности, чувствительности таких многочленов к ошибкам при вычислении их коэффициентов. Для того, чтобы избежать этих затруднений, отрезок *[a, b]* разбивают на частичные отрезки и на каждом из них приближенно заменяют функцию *f(x)* многочленом невысокой степени (кусочно-полиномиальная интерполяция).

Одним из способов интерполирования на всем отрезке является интерполирование с помощью сплайн-функций. Сплайн-функцией или просто сплайном называют кусочно-полиномиальную функцию, определенную на отрезке *[a, b]* и имеющую на этом отрезке некоторое число непрерывных производных. Слово «сплайн» (английское spline) означает гибкую линейку, используемую для проведения гладких кривых через заданные точки плоскости.

Максимальная по всем частичным отрезкам степень полиномов называется степенью сплайна, а разность между степенью сплайна и порядком наивысшей непрерывной на *[a, b]* производной – дефектом сплайна.

Мы рассмотрим частный, но распространенный в вычислительной практике случай, когда сплайн определяется с помощью многочленов третьей степени (кубический сплайн).

Пусть в точках

$$a = x_0 < x_1 < \ldots < x_n = b$$

заданы значения функции *f(x)* $y_i = f(x_i), \quad i = 0,1,\ldots,n$. Кубическим сплайном, соответствующим данной функции *f(x)* и данным узлам x_i, называется функция *P(x)*, удовлетворяющая следующим условиям:

а) на каждом отрезке $[x_i, x_{i+1}]$, $i = 0,1,\ldots,n-1$, функция *P(x)* является многочленом третьей степени;

б) функция *P(x)*, а также ее первая (и вторая) производные непрерывны на *[a, b]*;

в) $P(x_i) = y_i \quad i = 0,1,\ldots,n$.

Последнее условие называется условием интерполирования, а сплайн, определяемый условиями а-в, называется также интерполяционным кубическим сплайном.

Предположим, что нам известны значения производной функции *f(x)* в каждой из узловых точек: $f'(x_i) = y'_i \quad i = 0,1,\ldots,n$. По заданным значениям y_i и y'_i, $i = 0,1,\ldots,n$, построим для функции *f(x)* на каждом отрезке $[x_i, x_{i+1}]$, $i = 0,1,\ldots,n-1$, интерполяционный полином Эрмита третьей степени в виде (5.7). Для этого ведем вспомогательные функции:

$$
\begin{aligned}
F_{1i}(x) &= 3\left(\frac{x_{i+1}-x}{h_i}\right)^2 - 2\left(\frac{x_{i+1}-x}{h_i}\right)^3;\\
F_{2i}(x) &= 3\left(\frac{x-x_i}{h_i}\right)^2 - 2\left(\frac{x-x_i}{h_i}\right)^3;\\
F_{3i}(x) &= h_i\left[\left(\frac{x_{i+1}-x}{h_i}\right)^2 - \left(\frac{x_{i+1}-x}{h_i}\right)^3\right];\\
F_{4i}(x) &= -h_i\left[\left(\frac{x-x_i}{h_i}\right)^2 - \left(\frac{x-x_i}{h_i}\right)^3\right],
\end{aligned}
\qquad (5.9)
$$

где $h_i = x_{i+1} - x_i$, $i = 0,1,\ldots,n-1$. Для функций (5.9) и их производных

$$
\begin{aligned}
F'_{1i} &= -6\frac{x_{i+1}-x}{h_i^2} + 6\frac{(x_{i+1}-x)^2}{h_i^3};\\
F'_{2i} &= 6\frac{x-x_i}{h_i^2} - 6\frac{(x-x_i)^2}{h_i^3};\\
F'_{3i} &= -2\frac{x_{i+1}-x}{h_i} + 3\frac{(x_{i+1}-x)^2}{h_i^2};\\
F'_{4i} &= -2\frac{x-x_i}{h_i} + 3\frac{(x-x_i)^2}{h_i^2}
\end{aligned}
$$

определены условия (5.8), $i = 0,1,\ldots,n-1$, поэтому полином

$$P_i(x) = F_{1i}(x)y_i + F_{2i}(x)y_{i+1} + F_{3i}(x)y'_i + F_{4i}(x)y'_{i+1} \qquad (5.10)$$

является полиномом Эрмита для функции *f(x)* на каждом отрезке $[x_i, x_{i+1}]$, а функция $P(x) = P_i(x)$, $i = 0,1,\ldots,n-1$, – интерполяционным кубическим сплайном с непрерывной первой производной, т.е. с дефектом, равным 2. Если значения первой производной функции не заданы, их определяют с помощью формул численного дифференцирования. Построенный таким образом сплайн называется локальным, т.к. он строится отдельно на каждом отрезке $[x_i, x_{i+1}]$ непосредственно по формуле (5.10). Этот способ построения сплайна

удобен, если в процессе работы со сплайном (а она часто происходит в диалоговом режиме с визуализацией результатов на экране) требуется исправить какое-то одно значение функции или ее производной.

Построим теперь интерполяционный кубический сплайн с непрерывной второй производной, т.е. с дефектом, равным 1, по y_i, $i = 0,1,\ldots,n$.

Потребуем непрерывности $P''(x)$ в узлах, т.е.

$$P''_{i-1}(x_i) = P''_i(x_i) \quad i = 1,2,\ldots,n-1.$$

Выразим с помощью (5.10) обе части этого соотношения:

$$y_{i-1}F''_{1i-1}(x_i) + y_i F''_{2i-1}(x_i) + y'_{i-1}F''_{3i-1}(x_i) + y'_i F''_{4i-1}(x_i) =$$
$$= y_i F''_{1i}(x_i) + y_{i+1}F''_{1i}(x_i) + y'_i F''_{3i}(x_i) + y'_{i+1}F''_{4i}(x_i).$$

Вторые производные от функций (5.9) имеют вид:

$$F''_{1i-1}(x) = \frac{6}{h_{i-1}^2} - 12\frac{x_i - x}{h_{i-1}^3}; \quad F''_{2i-1}(x) = \frac{6}{h_{i-1}^2} - 12\frac{x - x_{i-1}}{h_{i-1}^3};$$

$$F''_{3i-1}(x) = \frac{2}{h_{i-1}} - 6\frac{x_i - x}{h_{i-1}^2}; \quad F''_{4i-1}(x) = -\frac{2}{h_{i-1}} + 6\frac{x - x_{i-1}}{h_{i-1}^2};$$

$$F''_{1i}(x) = \frac{6}{h_i^2} - 12\frac{x_{i+1} - x}{h_i^3}; \quad F''_{2i}(x) = \frac{6}{h_i^2} - 12\frac{x - x_i}{h_i^3};$$

$$F''_{3i}(x) = \frac{2}{h_i} - 6\frac{x_{i+1} - x}{h_i^2}; \quad F''_{4i}(x) = -\frac{2}{h_i} + 6\frac{x - x_i}{h_i^2}$$

и в точке x_i соответственно равны

$$\left.\begin{aligned} F''_{1i-1}(x_i) &= \frac{6}{h_{i-1}^2}; \quad F''_{2i-1}(x_i) = -\frac{6}{h_{i-1}^2}; \\ F''_{3i-1}(x_i) &= \frac{2}{h_{i-1}}; \quad F''_{4i-1}(x_i) = \frac{4}{h_{i-1}}; \end{aligned}\right\} \tag{5.11}$$

$$\left.\begin{aligned} F''_{1i}(x_i) &= -\frac{6}{h_i^2}; \quad F''_{2i}(x_i) = \frac{6}{h_i^2}; \\ F''_{3i}(x_i) &= -\frac{4}{h_i}; \quad F''_{4i}(x_i) = -\frac{2}{h_i}. \end{aligned}\right\} \tag{5.12}$$

Отсюда получаем

$$\frac{6}{h_{i-1}^2}y_{i-1} - \frac{6}{h_{i-1}^2}y_i + \frac{2}{h_{i-1}}y'_{i-1} + \frac{4}{h_{i-1}}y'_i = -\frac{6}{h_i^2}y_{i+} \frac{6}{h_i^2}y_{i+1} - \frac{4}{h_i}y'_i - \frac{2}{h_i}y'_{i+1}$$

или

$$\frac{1}{h_{i-1}}y'_{i-1}+2\frac{h_i+h_{i-1}}{h_i\cdot h_{i-1}}y'_i+\frac{1}{h_i}y'_{i+1}=3\left[-\frac{1}{h_{i-1}^2}y_{i-1}+\frac{h_i^2-h^2{}_{i-1}}{h_i^2\cdot h^2{}_{i-1}}y_i+\frac{1}{h_i^2}y_{i+1}\right], \qquad (5.13)$$

$$i=1,2,\ldots,n-1.$$

Введем обозначения

$$\alpha_i=\frac{h_{i-1}}{h_{i-1}+h_i};\quad i=1,2,\ldots,n-1.$$

Очевидно, что $0<\alpha_i<1$ и

$$1-\alpha_i=\frac{h_i}{h_{i-1}+h_i}.$$

Умножим равенство (5.13) на $\frac{h_{i-1}\cdot h_i}{h_{i-1}+h_i}$. Получим

$$(1-\alpha_i)y'_{i-1}+2y'_i+\alpha_i y'_{i+1}=3\left[-\frac{(1-\alpha_i)}{h_{i-1}}y_{i-1}+\frac{h_i-h_{i-1}}{h_{i-1}\cdot h_i}y_i+\frac{\alpha_i}{h_i}y_{i+1}\right],$$

$$i=1,2,\ldots,n-1,$$

или

$$(1-\alpha_i)y'_{i-1}+2y'_i+\alpha_i y'_{i+1}=3\left[(1-\alpha_i)\frac{y_i-y_{i-1}}{h_{i-1}}+\alpha_i\frac{y_{i+1}-y_i}{h_i}\right], \qquad (5.14)$$

$$i=1,2,\ldots,n-1.$$

Итак, получена система из *(n-1)* уравнений с *(n+1)* неизвестными $y'_0, y'_1,\ldots,y'_n$. Нужно задать еще два условия, которые называются краевыми, т.к. они обычно связаны с «крайними» значениями y'_0 *и* y'_n. Рассмотрим два варианта задания краевых условий.

1. Известны значения первой производной на концах отрезка

$$y'_0=f'(a);\quad y'_n=f'(b). \qquad (5.15)$$

2. Известны значения второй производной на концах отрезка *f''(a)* и *f''(b)*. Потребуем, чтобы *P''(a) = f''(a)* и *P''(b) = f''(b)*. Тогда, согласно (5.12) при *i=0* и согласно (5.11) при *i=n* соответственно имеем

$$\begin{aligned}&-\frac{6}{h_0^2}y_0+\frac{6}{h_0^2}y_1-\frac{4}{h_0}y'_0-\frac{2}{h_0}y'_1=f''(a);\\&\frac{6}{h_{n-1}^2}y_{n-1}-\frac{6}{h_{n-1}^2}y_n+\frac{2}{h_{n-1}}y'_{n-1}+\frac{4}{h_{n-1}}y'_n=f''(b),\end{aligned} \qquad (5.16)$$

где *f''(a)* и *f''(b)* – заданные величины.

Из соотношений (5.16) получим

$$y'_0 = -\frac{1}{2}y'_1 - \frac{3}{2h_0}y_0 + \frac{3}{2h_0}y_1 - \frac{h_0}{4}f''(a);$$

$$y'_n = -\frac{1}{2}y'_{n-1} - \frac{3}{2h_{n-1}}y_{n-1} + \frac{3}{2h_{n-1}}y_n + \frac{h_{n-1}}{4}f''(b).$$

Таким образом, оба варианта задания краевых условий дают нам два соотношения вида

$$\begin{aligned} y'_0 &= c_0 y'_1 + d_0, \quad |c_1| < 1 \\ y'_n &= c_n y'_{n-1} + d_n, \quad |c_n| < 1 \end{aligned} \tag{5.17}$$

Если постановка задачи не несет в себе задания краевых условий, то обычно полагают $P''(x_0) = P''(x_n) = 0$. Такой сплайн называется свободным кубическим сплайном и обладает свойствами минимальной кривизны [1,9].

В силу диагонального преобладания существует единственное решение системы (5.13) с краевыми условиями (5.17), оно может быть получено методом прогонки, который является устойчивым.

Оценим остаточную и вычислительную погрешности в точке x для локального сплайна.

Предполагается, что аппроксимируемая функция имеет на отрезке *[a, b]* непрерывную производную четвертого порядка. Значения $f(x_i)$ известны с одинаковой абсолютной погрешностью Δ^*. тогда для любого $x \in [x_i, x_{i+1}]$, $i = 0,1,\ldots,n-1$, остаточная погрешность

$$\Delta_1 = \frac{M_4^i (x - x_i)^2 (x - x_{i+1})^2}{4!}; \quad M_4^i = \max_{[x_i, x_{i+1}]} \left| f^{IV}(x) \right|,$$

а вычислительная погрешность

$$\Delta_2 = |F_{1i}(x)|\Delta^* + |F_{2i}(x)|\Delta^* + |F_{3i}(x)|\overline{\Delta^*} + |F_{4i}(x)|\overline{\Delta^{**}},$$

где $\overline{\Delta^*}$ u $\overline{\Delta^{**}}$ – полные абсолютные погрешности величин $f'(x_i)$, $f'(x_{i+1})$ соответственно, заданных или вычисленных по формулам численного дифференцирования.

Для глобального сплайна на всем отрезке *[a, b]* справедливы следующие оценки остаточных погрешностей в случае равноотстоящих узлов $x_i = 0,1,\ldots,n$ [9]:

$$\max_{[a,b]} |f(x) - P(x)| \le M_4 h^4;$$

$$\max_{[a,b]} |f'(x) - P'(x)| \le M_4 h^3;$$

$$\max_{[a,b]} |f''(x) - P''(x)| \le M_4 h^2;$$

$$M_4 = \max_{[a,b]} |f^{IV}(x)|.$$

6. ЧИСЛЕННОЕ ИНТЕГРИРОВАНИЕ

Пусть требуется вычислить интеграл

$$J = \int_a^b f(x)dx. \qquad (6.1)$$

Если функция *f(x)* является непрерывной на отрезке *[a, b]*, то интеграл (6.1) существует и может быть вычислен по формуле Ньютона-Лейбница

$$J = \int_a^b f(x)dx = F(b) - F(a). \qquad (6.2)$$

Однако для большинства функций *f(x)* первообразную *F(x)* не удается выразить через элементарные функции. Кроме того, функция *f(x)* часто задается в виде таблицы ее значений для определенных значений аргумента. Все это порождает потребность в построении формул численного интегрирования или квадратурных формул.

Приближенное равенство

$$J = \int_a^b f(x)dx \approx (b-a)\sum_{i=1}^{N} A_i \cdot f(x_i) = J_N \qquad (6.3)$$

называется квадратурной формулой, определяемой узлами $x_i \in [a,b]$ и коэффициентами A_i.

Величина

$$R_N(f) = J - J_N \qquad (6.4)$$

называется остаточным членом квадратурной формулы.

В зависимости от способа задания подынтегральной функции *f(x)* будем рассматривать два различных в смысле реализации случая численного интегрирования.

Задача 1. На отрезке *[a, b]* в узлах x_i заданы значения f_i некоторой функции *f*, принадлежащей определенному классу *F*. Требуется приближенно вычислить интеграл (6.1) и оценить погрешность полученного значения.

Так обычно ставится задача численного интегрирования в том случае, когда подынтегральная функция задана в виде таблицы.

Задача 2. На отрезке *[a, b]* функция *f(x)* задана в виде аналитического выражения. Требуется вычислить интеграл (6.1) с заданной предельно допустимой погрешностью ε.

Рассмотрим алгоритм решения задач 1 и 2.

Алгоритм решения задачи 1.

1. Выбирают конкретную квадратурную формулу (6.3) и вычисляют J_N. Если значения функции *f(x)* заданы приближенно, то фактически вычисляют лишь приближенное значение $\overline{J_N}$ для точного J_N.

2. Приближенно принимают, что $J \approx \overline{J_N}$.

3. Пользуясь конкретным выражением для остаточного члена $R_N(f)$ или оценкой его для выбранной квадратурной формулы, вычисляют погрешность метода

$$\Delta_1 \geq \left|J - J_N\right| = \left|R_N(f)\right|.$$

4. Определяют погрешность вычисления $\overline{J_N}$

$$\Delta_2 \geq \left|J_N - \overline{J_N}\right|$$

по погрешностям приближенных значений $f(x_i)$.

5. Находят полную абсолютную погрешность приближенного значения $\overline{J_N}$:

$$\left|J - \overline{J_N}\right| \leq \Delta_1 + \Delta_2 = \Delta.$$

6. Получают решение задачи в виде

$$J = \overline{J_N} \pm \Delta.$$

Алгоритм решения задачи 2.

1. Представляют ε в виде суммы трех неотрицательных слагаемых:

$$\varepsilon = \varepsilon_1 + \varepsilon_2 + \varepsilon_3,$$

где ε_1 – предельно допустимая погрешность метода: ε_2 – предельно допустимая погрешность вычисления $\overline{J_N}$; ε_3 – предельно допустимая погрешность округления результата.

2. Выбирают N в квадратурной формуле так, чтобы выполнялось неравенство

$$\Delta_1 = \left|J - J_N\right| = \left|R_N(f)\right| \leq \varepsilon_1.$$

3. Вычисляют $f(x_i)$ с такой точностью, чтобы при подсчете $\overline{J_N}$ по формуле (6.3) обеспечить выполнение неравенства

$$\Delta_2 = \left|J - \overline{J_N}\right| \leq \varepsilon_2.$$

Для этого, очевидно, достаточно вычислить все $f(x_i)$ с абсолютной погрешностью

$$\Delta^* \leq \frac{\varepsilon_2}{(b-a)\sum_{i=1}^{N}\left|A_i\right|}.$$

4. Найденную в п.3. величину $\overline{J_N}$ округляют (если $\varepsilon_3 \neq 0$) с предельно допустимой погрешностью ε_3 до величины $\overline{\overline{J_N}}$.

5. Получают решение задачи в виде

$$J = \overline{\overline{J_N}} \pm \varepsilon.$$

6.1. Формула прямоугольников

Допустим, что $f(x) \in C_2[a,b]$.

Отрезок *[a, b]* разделим на *N* равных частичных отрезков $[x_{i-1}, x_i]$, где $x_i = a + ih, \quad i = \overline{0, N-1}; \quad x_n = b; h = \frac{(b-a)}{N}$.

Тогда

$$\int_a^b f(x)dx = \sum_{i=1}^{N} \int_{x_{i-1}}^{x_i} f(x)dx. \qquad (6.5)$$

Обозначим среднюю точку отрезка $[x_{i-1}, x_i]$ через

$$\xi_i = \frac{x_{i-1} + x_i}{2}. \qquad (6.6)$$

Запишем для функции *f(x)* на каждом их отрезков $[x_{i-1}, x_i]$ формулу Тейлора с остаточным членом в форме Лагранжа

$$f(x) = f(\xi_i) + f'(\xi_i)(x - \xi_i) + \frac{f''(\eta_i)}{2!}(x - \xi_i)^2, \qquad (6.7)$$

$\eta_i \in (x_{i-1}, x_i)$.

Подставим в правую часть соотношения (6.5) вместо *f(x)* ее представление (6.7) и получим (6.8):

$$\int_a^b f(x)dx = \sum_{i=1}^{N} \int_{x_{i-1}}^{x_i} \left[f(\xi_i) + f'(\xi_i)(x - \xi_i) + \frac{f''(\eta_i)}{2!}(x - \xi_i)^2 \right] dx =$$

$$= \sum_{i=1}^{N} \left[f(\xi_i) \int_{x_{i-1}}^{x_i} dx + f'(\xi_i) \int_{x_{i-1}}^{x_i} (x - \xi_i)dx + \int_{x_{i-1}}^{x_i} \frac{f''(\eta_i)(x - \xi_i)^2}{2} dx \right].$$

Используя для вычисления $\int_{x_{i-1}}^{x_i} \frac{f''(\eta_i)(x - \xi_i)^2}{2} dx$ теорему о среднем значении интеграла и учитывая, что $\int_{x_{i-1}}^{x_i} (x - \xi_i)dx = 0$, получим

$$\int_a^b f(x)dx = h \sum_{i=1}^{N} f(\xi_i) + \frac{h^3}{24} \sum_{i=1}^{N} f''(\overline{\eta_i}), \quad \overline{\eta_i} \in (x_{i-1}, x_i). \qquad (6.9)$$

В силу непрерывности $f''(x)$ существует такая точка $\eta \in (a,b)$, что

$$\sum_{i=1}^{N} f''(\overline{\eta_i}) = Nf''(\eta). \qquad (6.10)$$

Используя (6.10), получаем

$$\int_a^b f(x)dx = h \sum_{i=1}^{N} f(\xi_i) + \frac{h^3}{24} Nf''(\eta).$$

или, так как $h = \frac{(b-a)}{N}$,

$$\int_a^b f(x)dx = (b-a)\sum_{i=1}^{N}\frac{1}{N}f(\xi_i) + \frac{(b-a)}{24}h^2 f''(\eta). \qquad (6.11)$$

Приближенное равенство

$$\int_a^b f(x)dx \approx (b-a)\sum_{i=1}^{N}\frac{1}{N}f(\xi_i) = J_N^{np} \qquad (6.12)$$

называется квадратурной формулой прямоугольников, определяемой узлами $\xi_i \in [a,b]$ и коэффициентами $A_i = \frac{1}{N}$. Величина

$$R_N(f) = \int_a^b f(x)dx - J_N^{np} = \frac{(b-a)}{24}h^2 f''(\eta) \qquad (6.13)$$

является остаточным членом формулы прямоугольников.

Оценка остаточной погрешности формулы прямоугольников может быть записана в виде

$$|R_N(f)| \le \frac{(b-a)}{24}h^2 M_2 = \Delta_1 , \qquad (6.14)$$

где

$$M_2 = \max_{[a,b]} |f''(x)|.$$

Выражения для остаточного члена (6.13) и остаточной погрешности (6.14) показывают, что формула прямоугольников (6.12) является точной для любой линейной функции, т.к. вторая производная такой функции равна нулю и, следовательно, $\Delta_1 = 0$.

Оценим вычислительную погрешность Δ_2 формулы прямоугольников, которая возникает за счет приближенного вычисления значений функции *f(x)* в узлах ξ_i.

Пусть, например, значения $f(\xi_i)$ в формуле (6.12) вычислены с одинаковой абсолютной погрешностью Δ^*, тогда

$$\Delta_2 = \left|J_N - \overline{J_N}\right| = (b-a)\sum_{i=1}^{N}\frac{1}{N}\Delta^* = (b-a)\Delta^* . \qquad (6.15)$$

6.2. Формула трапеций

Предположим, что $f(x) \in C_2[a,b]$. Разделим отрезок *[a, b]* на *N* равных частей, тогда

$$\int_a^b f(x)dx = \sum_{i=1}^{N}\int_{x_{i-1}}^{x_i} f(x)dx , \qquad (6.16)$$

где $x_i = a + ih, \quad i = \overline{0, N-1}; \quad x_N = b, \quad h = \dfrac{b-a}{N}$.

Заменим функцию *f(x)* на каждом из отрезков $[x_{i-1}, x_i]$ первой интерполяционной формулой Ньютона первой степени

$$f(x) = f(x_{i-1}) + \frac{x - x_{i-1}}{h}\left(f(x_i) - f(x_{i-1}) + \frac{f''(\eta_i)}{2}(x - x_{i-1})(x - x_i) \right), \quad (6.17)$$

$\eta_i \in (x_{i-1}, x_i)$.

Подставляя формулу (6.17) в правую часть (6.16), интегрируя и используя теорему о среднем значении интеграла, получим

$$\int_a^b f(x)dx = \sum_{i=1}^{N} h\frac{f(x_{i-1}) + f(x_i)}{2} - \frac{h^3}{12}\sum_{i=1}^{N} f''(\eta_i), \quad (6.18)$$

$$\eta_i \in (x_{i-1}, x_i).$$

В силу (6.10) получаем

$$\int_a^b f(x)dx = h\left(\frac{f(x_0) + f(x_N)}{2} + \sum_{i=1}^{N-1} f(x_i) \right) - \frac{h^2}{12}(b-a)f''(\eta), \quad (6.19)$$

$\eta \in (a, b)$.

Приближенное равенство

$$J = \int_a^b f(x)dx \approx \frac{b-a}{N}\left(\frac{f(x_0) + f(x_N)}{2} + \sum_{i=1}^{N-1} f(x_i) \right) = J_N^{mp} \quad (6.20)$$

называется формулой трапеций. Величина

$$R_N(f) = J - J_N^{mp} = -\frac{h^2}{12}(b-a)f''(\eta) \quad (6.21)$$

является остаточным членом формулы трапеций. Оценка остаточной погрешности формулы трапеций может быть записана в виде

$$|R_N(f)| \le \frac{b-a}{12}h^2 M_2 = \Delta_1. \quad (6.22)$$

Формула трапеций, как и формула прямоугольников, является точной для любой линейной функции. Вычислительная погрешность формулы трапеций также равна

$$\Delta_2 = (b-a)\Delta^*. \quad (6.23)$$

Так как остаточные члены формул прямоугольников и трапеций (6.13) и (6.21) имеют противоположные знаки, формулы (6.12) и (6.20) дают двустороннее приближение для интеграла (6.1), т.е.

$$J_N^{np} < J < J_N^{mp}, \quad \text{если } f''(x) > 0;$$

$$J_N^{mp} < J < J_N^{np}, \quad \text{если } f''(x) < 0.$$

В таком случае можно принять, что

$$J \approx \frac{J_N^{mp} + J_N^{np}}{2} = \widetilde{J}, \quad (6.24)$$

тогда

$$\left|J-\tilde{J}\right|<\frac{\left|J_N^{mp}-J_N^{np}\right|}{2}, \qquad (6.25)$$

т.е. погрешность выражается через приближенные значения интегралов.

6.3. Формула Симпсона

Предположим, что $f(x)\in C_4[a,b]$. Разделим отрезок *[a, b]* на четное число равных частей *N=2k*, тогда

$$\int_a^b f(x)dx=\sum_{i=0}^{k-1}\int_{x_{2i}}^{x_{2i+2}} f(x)dx, \qquad (6.26)$$

где $x_i=a+ih,\quad i=0,1,\ldots,2k-1;\quad x_{2k}=b,\quad h=\frac{b-a}{N}=\frac{b-a}{2k}$.

Заменим функцию *f(x)* на каждом отрезке $[x_{2i},x_{2i+2}]$ длиной *2h* интерполяционным полиномом Лагранжа второй степени и положим

$$\int_{x_{2i}}^{x_{2i+2}} f(x)dx\approx\int_{x_{2i}}^{x_{2i+2}} L_2(x)dx. \qquad (6.27)$$

Возьмем интеграл в правой части (6.27). Получим:

$$\int_{x_{2i}}^{x_{2i+2}} L_2(x)dx=\int_{x_{2i}}^{x_{2i+2}}\left[\begin{array}{l}\frac{(x-x_{2i+1})(x-x_{2i+2})}{(x_{2i}-x_{2i+1})(x_{2i}-x_{2i+2})}f(x_{2i})+\frac{(x-x_{2i})(x-x_{2i+2})}{(x_{2i+1}-x_{2i})(x_{2i+1}-x_{2i+2})}f(x_{2i+1})+\\ +\frac{(x-x_{2i})(x-x_{2i+1})}{(x_{2i+2}-x_{2i})(x_{2i+2}-x_{2i+1})}f(x_{2i+2})\end{array}\right]dx= \qquad (6.28)$$

$$=\frac{h}{3}\left[f(x_{2i})+4f(x_{2i+1})+f(x_{2i+2})\right].$$

Подставив (6.28) в (6.26), получим квадратурную формулу Симпсона

$$\int_a^b f(x)dx\approx\frac{h}{3}\left[f(x_0)+f(x_{2k})+4\sum_{i=0}^{k-1}f(x_{2i+1})+2\sum_{i=0}^{k-2}f(x_{2i+2})\right].$$

Остаточный член интерполяционного полинома Лагранжа второй степени, построенного на каждом отрезке $[x_{2i},x_{2i+2}]$, равный

$$\frac{f'''(\xi(x))}{3!}(x-x_{2i})(x-x_{2i+1})(x-x_{2i+2}),$$

обращается в нуль, если *f(x)* – полином второй степени. Следовательно, формула Симпсона является точной для полинома второй степени.

Докажем, что формула Симпсона является точной и для полинома третьей степени. Действительно, для $f(x)=x^3$ имеем по формуле Симпсона

$$\int_{x_{2i}}^{x_{2i+2}} f(x)dx \approx \frac{h}{3}\left[x_{2i}^3 + 4x_{2i+1}^3 + x_{2i+2}^3\right] =$$

$$= \frac{x_{2i+2} - x_{2i}}{6} \cdot \left[x_{2i}^3 + 4\left(\frac{x_{2i} + x_{2i+2}}{2}\right)^3 + x_{2i+2}^3 \right] =$$

$$= \frac{x_{2i+2} - x_{2i}}{6} \cdot 3\frac{x_{2i}^3 + x_{2i}^2 \cdot x_{2i+2} + x_{2i} \cdot x_{2i+2}^2 + x_{2i+2}^3}{2} = \frac{x_{2i+2}^4 - x_{2i}^4}{4},$$

что равно точному значению этого интеграла, полученному по формуле Ньютона-Лейбница

$$\int_{x_{2i}}^{x_{2i+2}} x^3 dx = \frac{x^4}{4}\Bigg|_{x_{2i}}^{x_{2i+2}} = \frac{x_{2i+2}^4 - x_{2i}^4}{4}.$$

Таким образом, формула Симпсона является точной для полинома второй степени и для функции $f(x)=x^3$, а значит, и для произвольного полинома третьей степени.

Получим остаточный член формулы Симпсона. Для этого представим подынтегральную функцию $f(x)$ на каждом отрезке $[x_{2i}, x_{2i+2}]$ интерполяционным полиномом Эрмита третьей степени с двукратным узлом x_{2i+1}:

$$J = \int_a^b f(x)dx = \sum_{i=0}^{k-1} \int_{x_{2i}}^{x_{2i+2}} H_3^{(i)}(x)dx +$$

$$+ \sum_{i=0}^{k-1} \int_{x_{2i}}^{x_{2i+2}} \frac{f^{IV}(\xi_i(x))}{4!}(x - x_{2i})(x - x_{2i+1})2(x - x_{2i+2})dx. \qquad (6.29)$$

Заменим первую сумму правой части (6.29) формулой Симпсона, которая дает точное значение каждого интеграла $\int_{x_{2i}}^{x_{2i+2}} H_3^{(i)}(x)dx$.

Вторую сумму преобразуем, интегрируя с помощью теоремы о среднем для определенного интеграла и применяя затем теорему о среднем значении непрерывной функции. Получим

$$J = \frac{h}{3}\left[f(x_0) + f(x_{2k}) + 4\sum_{i=0}^{k-1} f(x_{2i+1}) + 2\sum_{i=0}^{k-2} f(x_{2i+2}) \right] -$$

$$- \frac{f^{IV}(\eta)(b-a)}{180} h^4, \qquad \eta \in [a,b].$$

Величина

$$R_N(f) = -\frac{f^{IV}(\eta)(b-a)}{180} h^4$$

является остаточным членом формулы Симпсона.

6.4. Правило Рунге практической оценки погрешности квадратурных формул. Уточнение приближенного значения интеграла по Ричардсону

Пусть функция $f(x) \in C_4[a,b]$ и интеграл (6.1) вычисляется по формуле прямоугольников. Получим следующее соотношение:

$$J = \int_a^b f(x)dx = J_N^{np} + ch^2 + O(h^4), \tag{6.30}$$

где с – постоянная, не зависящая от h.

Введем вспомогательную функцию

$$F(x) = \int_{x_i - \frac{h}{2}}^{x} f(t)dt, \quad x \in [x_{i-1}, x_i].$$

Очевидно, что

$$F\left(x_i - \frac{h}{2}\right) = 0; \quad F'(x) = f(x); \quad F''(x) = f'(x);$$
$$F'''(x) = f''(x); \quad F^{IV}(x) = f'''(x); \quad F^{V}(x) = f^{IV}(x). \tag{6.31}$$

Разложим функцию $F(x)$ в ряд Тейлора в окрестности точки $x_i - \frac{h}{2}$.

$$F(x) = F(x_i - \frac{h}{2}) + F'(x_i - \frac{h}{2})(x - x_i + \frac{h}{2}) +$$
$$+ \frac{F''(x_i - \frac{h}{2})}{2!}(x - x_i + \frac{h}{2})^2 + \frac{F'''(x_i - \frac{h}{2})}{3!}(x - x_i + \frac{h}{2})^3 + \tag{6.32}$$
$$+ \frac{F^{IV}(x_i - \frac{h}{2})}{4!}(x - x_i + \frac{h}{2})^4 + \frac{F^{V}(\xi_i)}{5!}(x - x_i + \frac{h}{2})^5,$$
$$\xi_i \in [x_{i-1}, x_i].$$

С помощью (6.31) и (6.32) имеем

$$F(x_i) = f\left(x_i - \frac{h}{2}\right) \cdot \frac{h}{2} + \frac{f'\left(x_i - \frac{h}{2}\right)}{8}h^2 + \frac{f''\left(x_i - \frac{h}{2}\right)}{48}h^3 +$$
$$+ \frac{f'''\left(x_i - \frac{h}{2}\right)}{24} \cdot \frac{h^4}{16} + \frac{f^{IV}(\xi_i)}{120} \cdot \frac{h^5}{32}, \quad \xi_i \in [x_{i-1}, x_i];$$

$$F(x_{i-1}) = -f\left(x_i - \frac{h}{2}\right)\cdot\frac{h}{2} + \frac{f'\left(x_i - \frac{h}{2}\right)}{8}h^2 - \frac{f''\left(x_i - \frac{h}{2}\right)}{48}h^3 +$$

$$+ \frac{f'''\left(x_i - \frac{h}{2}\right)}{24}\cdot\frac{h^4}{16} - \frac{f^{IV}(\overline{\xi_i})}{120}\cdot\frac{h^5}{32}, \quad \overline{\xi_i} \in [x_{i-1}, x_i].$$

Вычитая из верхнего равенства нижнее, получим

$$\int_{x_{i-1}}^{x_i} f(t)dt = F(x_i) - F(x_{i-1}) = f\left(x_i - \frac{h}{2}\right)h + \frac{f''\left(x_i - \frac{h}{2}\right)}{24}h^3 +$$

$$+ \frac{h^5}{120\cdot 16}\cdot\frac{f^{IV}(\xi) + f^{IV}(\overline{\xi})}{2} = f\left(x_i - \frac{h}{2}\right)h + \frac{f''\left(x_i - \frac{h}{2}\right)}{24}h^3 + \qquad (6.33)$$

$$+ \frac{f^{IV}(\eta_i)}{1920}h^5, \qquad \eta_i \in [x_{i-1}, x_i],$$

откуда

$$\int_a^b f(t)dt = h\sum_{i=1}^{N} f\left(x_i - \frac{h}{2}\right) + \frac{h^3}{24}\sum_{i=1}^{N} f''\left(x_i - \frac{h}{2}\right) + \frac{h^4(b-a)}{1920}f^{IV}(\eta), \qquad (6.34)$$

$\eta \in [a,b]$.

На основании (6.11)

$$\int_a^b f''(t)dt = h\sum_{i=1}^{N} f''\left(x_i - \frac{h}{2}\right) + \frac{h^2(b-a)}{24}f^{IV}(\eta),$$

откуда

$$\frac{h^3}{24}\sum_{i=1}^{N} f''\left(x_i - \frac{h}{2}\right) = \frac{h^2}{24}\left[\int_a^b f''(t)dt - \frac{h^2}{24}(b-a)f^{IV}(\varsigma)\right], \qquad (6.35)$$

$\varsigma \in [a,b]$

Подставим (6.35) в (6.34):

$$\int_a^b f(t)dt = h\sum_{i=1}^{N} f\left(x_i - \frac{h}{2}\right) + \frac{h^2}{24}\left[\int_a^b f''(t)dt - \frac{h^2}{24}(b-a)f^{IV}(\varsigma)\right] +$$

$$+ \frac{(b-a)h^4 f^{IV}(\eta)}{1920} = h\sum_{i=1}^{N} f\left(x_i - \frac{h}{2}\right) + h^2 c + O(h^4),$$

где $c = \frac{1}{24}\int_a^b f''(t)dt$ не зависит от h. Соотношение (6.30) получено.

Величина ch^2 называется главной частью погрешности формулы прямоугольников.

Если $f(x) \in C_4[a,b]$, то справедливо аналогичное соотношение и для формулы трапеций

$$J = J_N^{mp} + c_1 h^2 + O(h^4), \qquad (6.36)$$

где $c_1 = -\frac{1}{12}\int_a^b f''(t)dt$

не зависит от h.

При условии $f(x) \in C_6[a,b]$ можно получить аналогичное соотношение для формулы Симпсона

$$J = J_N^C + \bar{c}h_4 + O(h^6), \qquad (6.37)$$

где $\bar{c}$ – не зависящая от h постоянная.

Обозначим через J_h приближенное значение интеграла (6.1), найденное по одной из трех формул: прямоугольников, трапеций, Симпсона, и объединим соотношения (6.30), (6.36), (6.37) в одно

$$J = J_h + ch^k + O(h^{k+2}), \qquad (6.38)$$

где с не зависит от h, $k = 2$ для формул прямоугольников и трапеций, $k = 4$ для формулы Симпсона. Предполагается, что $f(x) \in C_{k+2}[a,b]$. Запишем соотношение (6.38) для $h_1 = 2h$:

$$J = J_{2h} + c(2h)^k + O(h^{k+2}), \qquad (6.39)$$

вычтем из (6.39) (6.38) и получим

$$0 = -J_h + J_{2h} + ch^k(2^k - 1) + O(h^{k+2}),$$

$$J_h - J_{2h} = ch^k(2^k - 1) + O(h^{k+2}) \quad \textit{или}$$

$$J - J_h = ch^k + O(h^{k+2}) = \frac{J_h - J_{2h}}{2^k - 1} + O(h^{k+2}),$$

следовательно, с точностью до $O(h^{k+2})$ имеем

$$|J - J_h| \approx \frac{|J_h - J_{2h}|}{2^k - 1}. \qquad (6.40)$$

Вычисление приближенной оценки погрешности квадратурной формулы по формуле (6.40) называется правилом Рунге.

Вычитая из умноженного на 2^k равенства (6.38) равенство (6.39), получим

$$J(2^k - 1) = 2^k J_h - J_{2h} + O(h^{k+2}), \qquad (6.41)$$

откуда

$$J = \frac{2^k J_h - J_{2h}}{2^k - 1} + O(h^{k+2}). \qquad (6.42)$$

Число $J_h^R = \frac{2^k J_h - J_{2h}}{2^k - 1}$ называется уточненным по Ричардсону приближенным значением интеграла J.

Согласно (6.42)

$$J - J_h^R = O(h^{k+2}).$$

Таким образом, с помощью приближенных значений интегралов J_h, J_{2h}, найденных по соответствующим квадратурным формулам с шагом h и $2h$, можно, во-первых, оценить погрешность более точного значения интеграла J_h по правилу Рунге и, во-вторых, вычислить уточненное по Ричардсону приближенное значение интеграла J_h^R, имеющее погрешность более высокого порядка относительно h, чем J_h.

7. ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Дифференциальным уравнением называется уравнение вида

$$F(x, y, y', \ldots, y^{(n)}) = 0,$$

которое кроме независимых переменных и неизвестных функций от них содержит еще и производные неизвестных функций или их дифференциалы. Наивысший порядок входящих в уравнение производных неизвестных функций называется порядком дифференциального уравнения. Если искомые функции, входящие в дифференциальное уравнение, зависят от одной независимой переменной, то уравнение называется обыкновенным дифференциальным уравнением.

Задачи решения дифференциальных уравнений возникают при математическом моделировании многих реальных явлений. При этом, как правило, точное решение не удается выразить через элементарные функции. Доля задач, решаемых в явном виде, ничтожна мала. Поэтому возникает необходимость применять приближенные методы решения дифференциальных уравнений. В зависимости от того, ищется ли приближенное решение в аналитическом виде или в виде таблицы чисел, приближенные методы делятся соответственно на аналитические и численные. Например, при доказательстве существования решения дифференциального уравнения

$$y' = f(x, y) \qquad (7.1)$$

с начальным условием

$$y(x_0) = y_0 \qquad (7.2)$$

-задачи Коши – используют метод последовательных приближений Пикара. При этом точное решение получается как предел последовательности

$$y_0(x), y_1(x), y_2(x), \ldots, y_n(x), \ldots,$$

где

$$y_n(x) = y_0 + \int_{x_0}^{x} f(x, y_{n-1}(x))dx, \quad n = 1,2,\ldots$$

Эта последовательность равномерно сходится к решению y*(x)* задачи (7.1) и (7.2) на отрезке $[x_0 - h, x_0 + h], h = min(a, \frac{b}{M})$, $M = \sup_R |f(x, y)|$, если выполнены следующие условия:

1. Функция *f(x, y)* непрерывна в области

$$R = \{|x - x_0| \le a; \quad |y - y_0| \le b\}.$$

2. Функция *f(x, y)* удовлетворяет в *R* условию Липшица по y:

$$\left|f(x,\bar{y})-f(x,\bar{\bar{y}})\right| \le L\left|\bar{y}-\bar{\bar{y}}\right|,$$

где L – постоянная, не зависящая от $x,\bar{y},\bar{\bar{y}}$;

Точки $(x,\bar{y}),(x,\bar{\bar{y}})$ – произвольные точки области R.

Погрешность приближенного решения $y_n(x)$ в любой точке $x \in [x_0-h, x_0+h]$ оценивается следующей формулой:

$$\left|y(x)-y_n(x)\right| \le \frac{ML^n}{(n+1)!}\left|x-x_0\right|^{n+1}.$$

Наиболее распространенным численным методом решения задачи Коши является метод Рунге-Кутта.

7.1. Метод Рунге-Кутта

Пусть нам требуется найти решение задачи Коши (7.1)-(7.2) в точке $x_1 = x_0 + h$. Предположим, что в рассматриваемой области *f(x,y)* имеет непрерывные частные производные до некоторого порядка *n+1*. Тогда искомое решение будет иметь непрерывные производные до порядка *n*. Приближенное значение y_1 для решения $y(x_1)$ будет вычисляться следующим образом:

$$y(x_1) \approx y_1 = y_0 + \sum_{i=1}^{r} p_i k_i(h), \qquad (7.3)$$

где p_i – постоянные,

$$k_i(h) = hf(\xi_i, \eta_i), \qquad (7.4)$$

$$\left.\begin{aligned} \xi_i &= x_0 + \alpha_i h, \\ \eta_i &= y_0 + \beta_{i1}k_1(h) + \beta_{i2}k_2(h) + \ldots + \beta_{i,i-1}k_{i-1}(h), \end{aligned}\right\} \qquad (7.5)$$

α_i, β_{ij} – постоянные, $j = 1,2,\ldots,i-1;\quad i = 1,2,\ldots,r;\quad \alpha_1 = 0$.

Распишем последовательно формулы (7.4) и (7.5):

$$\begin{aligned}
&k_1(h) = hf(\xi_1,\eta_1) = hf(x_0,y_0);\\
&k_2(h) = hf(\xi_2,\eta_2) = hf(x_0+\alpha_2 h, y_0+\beta_{21}k_1);\\
&k_3(h) = hf(\xi_3,\eta_3) = hf(x_0+\alpha_3 h, y_0+\beta_{31}k_1+\beta_{32}k_2);\\
&\quad \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\\
&k_r(h) = hf(\xi_r,\eta_r) = hf(x_0+\alpha_r h, y_0+\beta_{r1}k_1+\beta_{r2}k_2+\ldots+\beta_{r,r-1}k_{r-1}).
\end{aligned} \qquad (7.6)$$

Рассмотрим вопрос о выборе параметров $p_i, \alpha_i, \beta_{ij}$. Обозначим через $R(h)$ разность между точным и приближенным значениями решения в точке x_1:

$$R(h) = y(x_1) - y_1 = y(x_0+h) - y_1.$$

В соответствии с (7.3) будем иметь

$$R(h) = y(x_0+h) - y_0 - \sum_{i=1}^{r} p_i k_i(h). \qquad (7.7)$$

Разложим $R(h)$ в ряд Маклорена:

$$R(h) = \sum_{i=0}^{s} \frac{R^{(i)}(0)}{i!} h^i + \frac{R^{(s+1)}(\theta h)}{(s+1)!} h^{s+1}, \quad 0 \le \theta \le 1. \qquad (7.8)$$

Будем подбирать параметры $p_i, \alpha_i, \beta_{ij}$ так, чтобы

$$R(0) = R'(0) = \ldots = R^{(s)}(0) = 0,$$
$$R^{(s+1)}(0) \neq 0,$$

причем s было бы как можно больше при произвольной *f(x, y)*. Величина $R(h)$ называется погрешностью метода Рунге-Кутта на одном шаге, а *(s+1)* – порядком погрешности. Таким образом, погрешность на шаге при таком выборе параметров согласно (7.8) будет равна:

$$y(x_1) - y_1 = R(h) = \frac{h^{s+1} R^{(s+1)}(\theta h)}{(s+1)!}, \quad 0 \le \theta \le 1. \qquad (7.9)$$

Очевидно, что условие $R(0) = 0$ будет выполнено всегда, так как $k_i(0) = 0$.

Теперь рассмотрим частные случаи формулы (7.3).

1 случай: *r = 1*. При этом сама формула (7.3) приобретает вид

$$y(x_0 + h) \approx y_1 = y_0 + p_1 k_1(h) = y_0 + p_1 h f(x_0, y_0),$$

a $R(h) = y(x_0 + h) - y_0 - p_1 h f(x_0, y_0).$

Найдем производные функции $R(h)$:

$$R'(h) = y'(x_0 + h) - p_1 f(x_0, y_0).$$

Условие

$$R'(0) = y'(x_0) - p_1 f(x_0, y_0) = 0$$

удовлетворяется при $p_1 = 1$.

Далее имеем

$$R''(h) = y''(x_0 + h).$$

Значение $R''(0) = y''(x_0)$ не зависит от констант и в общем случае не может быть равным нулю. Таким образом, приближенная формула

$$y(x_0 + h) \approx y_1 - y_0 + h f(x_0, y_0) \qquad (7.10)$$

имеет ошибку на одном шаге, равную

$$R(h) = \frac{h^2 y''(\xi)}{2}, \quad x_0 \le \xi \le x_0 + h. \qquad (7.11)$$

Говорят, что в этом случае погрешность метода на одном шаге имеет порядок h^2. Формула (7.10) называется еще методом Эйлера.

2 случай: *r = 2*. Формула (7.3) при этом имеет вид

$$y(x_0 + h) \approx y_1 = y_0 + p_1 k_1(h) + p_2 k_2(h);$$
$$k_1(h) = h f(x_0, y_0);$$
$$k_2(h) = h f(x_0 + \alpha_2 h, y_0 + \beta_{21} k_1).$$

Погрешность на шаге

$$R(h) = y(x_0 + h) - y_0 - p_1 h f(x_0, y_0) - p_2 h f(x_0 + \alpha_2 h, y_0 + \beta_{21} h f(x_0, y_0)).$$

Ее первая производная

$$R'(h) = y'(x_0 + h) - p_1 f(x_0, y_0) - p_2 f(x_0 + \alpha_2 h, y_0 + \beta_{21} h f(x_0, y_0)) -$$

$$- p_2 h \left[\alpha_2 \frac{\partial f}{\partial x}\left(x_0 + \alpha_2 h, y_0 + \beta_{21} h f(x_0, y_0)\right) + \right.$$

$$\left. + \beta_{21} f(x_0, y_0) \frac{\partial f}{\partial y}\left(x_0 + \alpha_2 h, y_0 + \beta_{21} h f(x_0, y_0)\right) \right],$$

$$R'(0) = y'(x_0) - (p_1 + p_2) f(x_0, y_0).$$

Таким образом, $R(0) = 0$ в том и только в том случае, если

$$p_1 + p_2 = 1.$$

Далее,

$$R''(h) = y''(x_0 + h) - 2p_2 \left[\alpha_2 \frac{\partial f}{\partial x}(x_0 + \alpha_2 h, y_0 + \beta_{21} h f(x_0, y_0)) + \right.$$

$$\left. + \beta_{21} f(x_0, y_0) \cdot \frac{\partial f}{\partial y}(x_0 + \alpha_2 h, y_0 + \beta_{21} h f(x_0, y_0)) \right] -$$

$$- p_2 h \left\{ \alpha_2^2 \frac{\partial^2 f}{\partial x_2}(x_0 + \alpha_2 h, y_0 + \beta_{21} h f(x_0, y_0)) + \right.$$

$$+ 2\alpha_2 \beta_{21} f(x_0, y_0) \frac{\partial^2 f}{\partial x \partial y}(x_0 + \alpha_2 h, y_0 + \beta_{21} h f(x_0, y_0)) +$$

$$\left. + \beta_{21}^2 (f(x_0, y_0))^2 \cdot \frac{\partial^2 f}{\partial y^2}(x_0 + \alpha_2 h, y_0 + \beta_{21} h f(x_0, y_0)) \right\}.$$

Выражение для $y''(x_0 + h)$ найдем, дифференцируя уравнение (7.1):

$$y''(x) = \frac{\partial f}{\partial x}(x, y) + y'(x) \frac{\partial f}{\partial y}(x, y) = \frac{\partial f}{\partial x}(x, y) + f(x, y) \frac{\partial f}{\partial y}(x, y), \qquad (7.12)$$

$$R''(0) = \frac{\partial f}{\partial x}(x_0, y_0)[1 - 2p_2\alpha_2] + \frac{\partial f}{\partial y}(x_0, y_0) f(x_0, y_0)[1 - 2p_2\beta_{21}].$$

Необходимым и достаточным условием обращения $R''(0)$ в нуль будет

$$1 - 2p_2\alpha_2 = 0;$$

$$1 - 2p_2\beta_{21} = 0.$$

Третья производная $R'''(h)$ будет равна:

$$\lambda'''(h) = y'''(x_0 + h) - 3p_2 \left[\alpha_2^2 \frac{\partial^2 f}{\partial x^2}(x_0 + \alpha_2 h, y_0 + \beta_{21} h f(x_0, y_0)) + \right.$$

$$+ 2\alpha_2 \beta_{21} f(x_0, y_0) \frac{\partial^2 f}{\partial x \partial y}(x_0 + \alpha_2 h, y_0 + \beta_{21} h f(x_0, y_0)) +$$

$$\left. + \beta_{21}^2 (f(x_0, y_0))^2 \cdot \frac{\partial^2 f}{\partial y^2}(x_0 + \alpha_2 h, y_0 + \beta_{21} h f(x_0, y_0)) \right] - p_2 h A(h),$$

A(h) – производная по *h* от выражения в фигурной скобке предыдущего равенства для *R''(h)*.

Дифференцируя (7.12) получим

$$y'''(x) = \frac{\partial^2 f}{\partial x^2}(x,y) + 2y'(x)\frac{\partial^2 f}{\partial x \partial y}(x,y) + (y'(x))^2 \frac{\partial^2 f}{\partial y^2}(x,y) + y''(x)\frac{\partial f}{\partial y}(x,y).$$

$$R''(0) = \frac{\partial^2 f}{\partial x^2}(x,y)\left[1 - 3p_2\alpha_2^2\right] + 2f(x_0,y_0)\frac{\partial^2 f}{\partial x \partial y}(x_0,y_0)\left[1 - 3p_2\alpha_2\beta_{21}\right] +$$

$$+ (f(x_0,y_0))^2 \frac{\partial^2 f}{\partial y^2}(x_0,y_0)\left[1 - 3p_2\beta_{21}^2\right] + \frac{\partial f}{\partial y}(x_0,y_0)y''(x_0).$$

Очевидно, что последнее слагаемое, а следовательно, и все выражение для $R'''(0)$, вообще говоря, не обращается в нуль. Таким образом, беря $p_1, p_2, \alpha_2, \beta_{21}$, удовлетворяющие условиям

$$\begin{aligned} &p_1 + p_2 = 1, \\ &1 - 2p_2\alpha_2 = 0, \\ &1 - 2p_2\beta_{21} = 0, \end{aligned} \qquad (7.13)$$

мы получим формулы, имеющие порядок ошибки на шаге h^3. Из (7.13) следует, что $p_2 \neq 0;\ \ \alpha_2 \neq 0;\ \ \alpha_2 = \beta_{21}$.

Равенства (7.13) являются системой трех уравнений относительно четырех неизвестных. Эта система имеет бесчисленное множество решений. Каждое решение дает формулу, имеющую порядок ошибки h^3.

Можно, например, взять $p_1 = \frac{1}{2}$. Тогда $p_3 = \frac{1}{2}$, $\alpha_2 = \beta_{21} = 1$. Формула (7.3) примет вид

$$y(x_0 + h) \approx y_1 = y_0 + \frac{h}{2}\left[f(x_0,y_0) + f(x_0 + h, y_0 + hf(x_0,y_0)\right].$$

Если обозначить

$$\tilde{y}_1 = y_0 + hf(x_0,y_0), \qquad (7.14)$$

то $y_1 = y_0 + \frac{h}{2}[f(x_0,y_0) + f(x_1,\tilde{y}_1)]$. (7.15)

Формулы (7.14)-(7.15) носят название метода Эйлера-Коши.

Если взять $p_1 = 0$, то $p_2 = 1, \alpha_2 = \beta_{21} = \frac{1}{2}$ и будем иметь следующую формулу:

$$y(x_0 + h) \approx y_1 = y_0 + hf(x_0 + \frac{h}{2}, y_0 + \frac{h}{2}f(x_0,y_0)), \qquad (7.16)$$

которая называется уточненным методом Эйлера.

На практике из формул, имеющих погрешность на шаге порядка h^3, используются именно эти две формулы: метод Эйлера-Коши и уточненный метод Эйлера, т.к. они имеют простой, удобный для вычислений вид.

3 случай: *r = 3*. Тогда, согласно (7.3) и (7.6)

$$y(x_0+h) = p_1k_1(h) + p_2k_2(h) + p_3k_3(h);$$
$$k_1(h) = hf(x_0, y_0);$$
$$k_2(h) = hf(x_0 + \alpha_2 h, y_0 + \beta_{21}k_1;$$
$$k_3(h) = hf(x_0 + \alpha_3 h, y_0 + \beta_{31}k_1 + \beta_{32}k_2)$$

и погрешность на шаге

$$R(h) = y(x_0+h) - y_0 - p_1k_1(h) - p_2k_2(h) - p_3k_3(h).$$

Для того, чтобы получить систему уравнений относительно неизвестных параметров $p_1, p_2, p_3, \alpha_2, \alpha_3, \beta_{21}, \beta_{31}, \beta_{32}$, нужно, как и в предыдущих случаях, выписать $R'(0), R''(0), R'''(0), R^{IV}(0)$ и потребовать их обращения в нуль. Оказывается [3], что в этом случае для произвольной *f(x, y)* можно обратить в нуль только $R'(0), R''(0)$ *и* $R'''(0)$ и порядок погрешности будет равен 4:

$$R(h) = \frac{h^4\lambda^{IV}(\xi)}{4!}, \quad 0 \le \xi \le h.$$

Чтобы выполнялось требование

$$R'(0) = R''(0) = R'''(0) = 0,$$

необходимо и достаточно выполнение следующих соотношений:

$$\begin{aligned}
&1 - p_1 - p_2 - p_3 = 0;\\
&\alpha_2 - \beta_{21} = 0;\\
&\alpha_3 - \beta_{31} - \beta_{32} = 0;\\
&\alpha_3(\alpha_3 - \alpha_2) - \beta_{32}\alpha_2(2 - 3\alpha_2) = 0;\\
&1 - 6p_3\beta_{32}\alpha_2 = 0;\\
&1 - 2p_2\alpha_2 - 2p_3\alpha_3 = 0.
\end{aligned} \qquad (7.17)$$

Система (7.17), содержащая 8 неизвестных и 6 уравнений, имеет бесчисленное множество решений, каждое из которых определяет формулу метода Рунге-Кутта с погрешностью на шаге четвертого порядка. Одна из широко употребляемых на практике формул соответствует решению

$$\alpha_2 = \beta_{21} = \frac{1}{2}; \quad \alpha_3 = 1; \quad \beta_{31} = -1; \quad \beta_{32} = 2;$$
$$p_1 = \frac{1}{6}; \quad p_2 = \frac{2}{3}; \quad p_3 = \frac{1}{6}$$

и имеет следующий вид:

$$y(x_0+h) \approx y_1 = y_0 + \frac{1}{6}[k_1 + 4k_2 + k_3], \qquad (7.18)$$

где

$$k_1(h) = hf(x_0, y_0);$$
$$k_2(h) = hf(x_0 + \frac{h}{2}, y_0 + \frac{k_1}{2}; \qquad (7.19)$$
$$k_3(h) = hf(x_0 + h, y_0 - k_1 + 2k_2).$$

4 случай: *r = 4*. Формулы (7.3) и (7.6) примут вид

$$y_1 = y_0 + p_1k_1(h) + p_2k_2(h) + p_3k_3(h) + p_4k_4(h);$$
$$k_1(h) = hf(x_0, y_0);$$
$$k_2(h) = hf(x_0 + \alpha_2 h, y_0 + \beta_{21}k_1;$$
$$k_3(h) = hf(x_0 + \alpha_3 h, y_0 + \beta_{31}k_1 + \beta_{32}k_2);$$
$$k_4(h) = hf(x_0 + \alpha_4 h, y_0 + \beta_{41}k_1 + \beta_{42}k_2 + \beta_{43}k_3).$$

В этом случае удается построить формулы с погрешностью на шаге пятого порядка

$$R(h) = \frac{h^5 \lambda^{IV}(\xi)}{5!}, \quad 0 \le \xi \le R,$$

из которых самой распространенной является следующая:

$$y(x_0+h) \approx y_1 = y_0 + \frac{1}{6}[k_1 + 2k_2 + 2k_3 + k_4], \qquad (7.20)$$

где

$$k_1(h) = hf(x_0, y_0);$$
$$k_2(h) = hf(x_0 + \frac{h}{2}, y_0 + \frac{k_1}{2});$$
$$k_3(h) = hf(x_0 + \frac{h}{2}, y_0 + \frac{k_2}{2}); \qquad (7.21)$$
$$k_4(h) = hf(x_0 + h, y_0 + k_3).$$

Дальнейшие исследования показывают, что в случае *r = 5* не удается достигнуть увеличения порядка точности на шаге, поэтому эти формулы применения не находят. При *r = 6* можно получить формулы, имеющие порядок ошибки h^6, но они очень громоздкие и практического применения также не находят [3].

Применяя ту или иную формулу Рунге-Кутта, мы находим $y(x_1) \approx y_1$. Затем, взяв за начальное значение y_1, можно продвинуться еще на один шаг такой же или другой длины. Повторяя этот процесс, мы получим таблицу значений искомого решения в некоторых точках.

Найдем приближенную оценку погрешности решения y_i^h, полученного после *2n* шагов с помощью одной из формул метода Рунге-Кутта.

Предположим, что $R^{(s+1)}(\xi)$ – мало меняющаяся функция на отрезке от 0 до *h*, т.е. на каждом шаге допущена одинаковая погрешность

$$\frac{R^{(s+1)}(\xi)}{(s+1)!}h^{s+1}, \quad 0 \le \xi \le h.$$

Тогда

$$y(x_i) = y_i^h + 2n\frac{R^{(s+1)}(\xi)}{(s+1)!}h^{s+1}. \qquad (7.22)$$

Если провести расчет по той же формуле с шагом *2h*, то получим другое приближенное решение y_i^{2h} в точках x_i:

$$y(x_i) = y_i^{2h} + n\frac{R^{(s+1)}(\xi)(2h)^{s+1}}{(s+1)!} = y_i^{2h} + 2n\frac{R^{(s+1)}(\xi)h^{s+1}2^s}{(s+1)!} \qquad (7.23)$$

Для того, чтобы оценить погрешность за *2n* шагов

$$\left|y(x_i) - y_i^h\right| \approx \left|2n\frac{R^{(s+1)}(\xi)h^{s+1}}{(s+1)!}\right|$$

вычтем (7.22) из (7.23). Получим

$$2n\frac{R^{(s+1)}(\xi)h^{s+1}}{(s+1)!}(2^s - 1) + y_i^{2h} - y_i^h = 0$$

и $$\left|2n\frac{R^{(s+1)}(\xi)h^{s+1}}{(s+1)!}\right| = \frac{\left|y_i^{2h} - y_i^h\right|}{2^s - 1}. \qquad (7.24)$$

Из (7.24) следует, что для метода Эйлера (7.10) погрешность приближенного решения y_i^h будет оцениваться формулой $\left|y_i^{2h} - y_i^h\right|$, для методов Эйлера-Коши (7.14)-(7.15) и уточненного метода Эйлера (7.16) – формулой

$$\frac{\left|y_i^h - y_i^{2h}\right|}{3},$$

а для методов Рунге-Кутта с погрешностями на шаге четвертого и пятого порядков – формулами

$$\frac{\left|y_i^h - y_i^{2h}\right|}{7} \quad и \quad \frac{\left|y_i^h - y_i^{2h}\right|}{15}$$

соответственно.

7.2. Разностный метод решения краевой задачи

Рассмотрим краевую задачу для дифференциального уравнения второго порядка следующего вида:

$$Ly = y'' - P(x)y = f(x), \quad x \in [0, H]; \qquad (7.25)$$

$$y(0) = a; \quad y(H) = b \qquad (7.26)$$

зададим шаг $h = \frac{H}{n}$, n – целое. Точки $x_j = jh, \quad j = 0,1,\ldots,n$, примем за узлы сетки, $y(x_j)$ – неизвестные значения искомого решения в узлах. Выразим производную $y''(x_j)$ в узлах сетки по формуле численного дифференцирования

$$y''(x_j) \approx \frac{y(x_{j+1}) - 2y(x_j) + y(x_{j-1})}{h^2}, \quad j = 1,2,...,n-1.$$

Пусть

$$P_j = p(x_j); \quad f_j = f(x_j); \quad j = 0,1,...,n.$$

Вместо дифференциальной краевой задачи (7.25)-(7.26) будем иметь разностную краевую задачу

$$l(y_j) = \frac{y_{j+1} - 2y_j + y_{j-1}}{h^2} - p_j y_j = f_j; \qquad (7.27)$$

$$j = 1,2,...,n-1; \quad y_0 = a; \quad y_n = b. \qquad (7.28)$$

где y_j – приближенное значение точного решения $y(x_j)$ в узлах $x_j, \quad j = 1,2,...,n-1$.

Перепишем систему линейных алгебраических уравнений (7.27)-(7.28) в виде

$$y_{j-1} - (2 + h^2 p_j) y_j + y_{j+1} = f_j h^2, \quad j = 1,2,...,n-1.$$

Эта система с трех диагональной матрицей при $P(x) \geq 0$ на *[0, H]* имеет решение, причем единственное, которое может быть получено методом прогонки, при этом условие $P(x) \geq 0$ гарантирует устойчивость прогонки. Дадим оценку этому решению.

Лемма 1. Пусть $P(x) \geq 0$ и числа $z_0, z_1, ..., z_n$ таковы, что $l(z_j) \leq 0, \quad z_0, z_n \geq 0$. Тогда $z_j \geq 0$ для всех j. Пусть $d = \min\limits_{0 \leq j \leq n} z_j$. Предположим, что $\alpha < 0$. Следовательно, $d \neq z_0, z_n$. Пусть q – наименьшее целое, для которого $z_q = d$. Из определения d и q имеем: $z_{q-1} > d; \quad z_{q+1} \geq d$.

Тогда

$$l(z_q) = \frac{z_{q+1} - 2z_q + z_{q-1}}{h^2} - p_q z_q = \frac{(z_{q+1} - z_q) + (z_{q-1} - z_q)}{h^2} - p_q z_q \geq$$

$$\geq \frac{(z_{q-1} - z_q)}{h^2} > 0$$

- противоречие с $l(q_j) \leq 0$.

Лемма 1 доказана.

Лемма 2. Если $P(x) \geq 0$, то для любой системы чисел z_j выполняется неравенство

$$\max_{0 \leq j \leq n} |z_j| \leq max(|z_0|, |z_n|) + Z \frac{H^2}{8},$$

где $Z = \max\limits_{0 \leq j \leq n-1} |l(z_j)|$.

Введем в рассмотрение функцию

$$\omega(x) = |z_0|\left(1 - \frac{x}{H}\right) + |z_n|\frac{x}{H} + Z\frac{x(H-x)}{2}$$

Через ω_j обозначим

$$\omega_j = \omega(x_j) = |z_0|\left(1 - \frac{jh}{H}\right) + |z_n|\frac{jh}{H} + Z\frac{jh(H-jh)}{2}.$$

Очевидно, что $\omega_j \geq 0$.

Это многочлен второй степени. Для него конечная разность второго порядка $\Delta^2\omega_j = h^2 2!\frac{(-z)}{2}$, следовательно,

$$\frac{\Delta^2\omega_j}{h^2} = \frac{\omega_{j+1} - 2\omega_j + \omega_{j-1}}{h^2} = -Z.$$

Отсюда следует, что

$$l(\omega_j) = \frac{\omega_{j+1} - 2\omega_j + \omega_{j-1}}{h^2} - p_j\omega_j \leq Z;$$
$$l(\omega_j \pm z_j) \leq Z \pm l(z_j) \leq 0.$$

Очевидно, что

$$\omega_0 \pm z_0 = |z_0| \pm z_0 \geq 0; \quad \omega_n \pm z_n = |z_n| \pm z_n \geq 0.$$

Числа $\omega_j \pm z_j$ удовлетворяют условиям леммы 1. Поэтому $\omega_j \pm z_j \geq 0$. Отсюда следует оценка

$$|z_j| \leq |\omega_j| \leq \max_{0\leq j\leq n} |\omega_j|.$$

Имеем неравенство

$$|z_0|\left(1 - \frac{jh}{H}\right) + |z_n|\frac{jh}{H} \leq max(|z_0|,|z_n|)\left(1 - \frac{jh}{H} + \frac{jh}{H}\right) = max(|z_0|,|z_n|).$$

Кроме того,

$$\max_{0\leq x\leq H} |x(H-x)| = \frac{H^2}{4}; \quad |jh(H-jh)| \leq \frac{H^2}{4}.$$

Поэтому имеем

$$\max_{0\leq j\leq n} |\omega_j| \leq \left(|z_0|,|z_n|\right) + \frac{ZH^2}{8}.$$

Лемма 2 доказана.

Рассмотрим случай, когда функции *P(x)* и *f(x)* дважды непрерывно дифференцируемы. В курсе дифференциальных уравнений доказывается, что когда краевая задача (7.25)-(7.26) имеет единственное решение *y(x)*, которое четырежды непрерывно дифференцируемо. Наша

задача – оценить разность $R_j = y(x_j) - y_j$ для $j = 0,1,...n$. $R_0 = R_n = 0$ – это краевые условия.

Рассмотрим

$$l(y(x_j)) - f_j = \frac{y(x_{j+1}) - 2y(x_j) + y(x_{j-1})}{h^2} - p_j y(x_j) - f_j.$$

Согласно дифференциальному уравнению (7.25) для любого j

$$y''(x_j) - p_j y(x_j) = f_j, \text{ т.е.}$$

$$y''(x_j) = p_j y(x_j) + f_j$$

Следовательно,

$$l(y(x_j)) - f_j = \frac{y(x_{j+1}) - 2y(x_j) + y(x_{j-1})}{h^2} - y''(x_j).$$

Левая часть этого равенства есть разность между приближенным значением второй производной в точке x_j, полученным по формуле численного дифференцирования, и точным значением этой производной $y(x_j)$ и равна остаточному члену этой формулы

$$l(y(x_j)) - f_j = \frac{h^2 y^{IV}(\xi_j)}{12}; \qquad (7.29)$$

$$\xi_i \in [x_{j-1}, x_j], \quad j = 1,2,...,n-1.$$

Согласно (7.27) имеем

$$l(y_j) - f_j = 0, \quad j = 1,...,n-1. \qquad (7.30)$$

Вычтем (7.30) из (7.29)

$$l(y(x_j) - y_j) = \frac{h^2 y^{IV}(\xi_j)}{12}; \text{ т.е.}$$

$$\left|l(R_j)\right| \le \frac{h^2 M_4}{12}; \quad M_4 = \max_{[0,H]} \left|y^{IV}(x)\right|.$$

Воспользуемся леммой 2 для чисел R_j, $j = 0,1,...,n$. Имеем

$$\left|R_j\right| \le max(\left|R_0\right|, \left|R_n\right|) + \frac{H^2}{8}\frac{h^2 M_4}{12} = \frac{H^2 M_4 h^2}{96}.$$

Таким образом, при $h \to 0$, т.е. неограниченном сгущении сетки, решение разностной задачи приближается к решению дифференциальной.

Разностный метод решения краевой задачи (7.25)-(7.26) используется также и при $P(x) < 0$, хотя успешный результат заранее предвидеть трудно. Для оценки получаемого решения в этом случае нужно провести расчеты для различных значений шага h (не менее трех) и убедиться в том, что полученные значения функции в одних и тех же узлах близки между собой и их разность уменьшается, что говорит о стремлении решения к некоторому пределу при $h \to 0$.

Литература

1. Бахвалов Н.С. Численные методы. –М: Наука, 1975.
2. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. – М.: Лаборатория Базовых Знаний, 2001.
3. Березин И.С., Жидков Н.П. Методы вычислений. – Т.1. – М.: Наука, 1966; - Т.2. – М.: Физматгиз, 1962.
4. Волков Е.А. Численные методы. – М.: Наука, 1987.
5. Демидович Б.П., Марон И.А. Основы вычислительной математики. – М.: Наука, 1970.
6. Демидович Б.П., Марон И.А., Шувалова Э.З. численные методы анализа. – М: Наука, 1967.
7. Калиткин Н.Н., Численные методы. - М.: Наука, 1978.
8. Курош А.Г. Курс высшей алгебры. - М.: Наука, 1968.
9. Самарский А.А., Гулин А.В. Численные методы. – М: Наука, 1989.
10. Турчак Л.И. Основы численных методов. – М: Наука, 1987.
11. Фаддеев Д.К., Фаддеева В.Н. вычислительные методы линейной алгебры. – М: Физматгиз, 1963.
12. Фихтенгольц Г.М. Математический анализ. –Т.1, 2. – М: Гостехиздат, 1957.
13. Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений. – М: Мир, 1980.