

Ю. Ю. Громов, С. И. Татаренко

**ВВЕДЕНИЕ
В МЕТОДЫ
ЧИСЛЕННОГО
АНАЛИЗА**



• ИЗДАТЕЛЬСТВО ТГТУ •

Министерство образования Российской Федерации
Тамбовский государственный технический университет

Ю. Ю. Громов, С. И. Татаренко

**ВВЕДЕНИЕ В МЕТОДЫ
ЧИСЛЕННОГО АНАЛИЗА**

*Утверждено Ученым советом университета
в качестве лекций к курсу*

ВВЕДЕНИЕ

Большинство математических описаний объектов и явлений разной природы приводят к задачам решения уравнений и систем уравнений (линейных, нелинейных, дифференциальных, интегральных) и необходимости решать такие уравнения. В качестве примера можно привести уравнения колебаний маятника, уравнения теплопроводности, уравнения баланса спроса и предложения.

Основополагающими методами изучения математических моделей являются аналитические методы: получение точного решения в частных случаях (например табличные интегралы, квадратные уравнения), разложения в ряды. Приближенные методы решения издавна играли лишь вспомогательную роль, но с появлением ЭВМ положение изменилось. Применение компьютеров радикально расширило возможности приложения приближенных математических методов в традиционных областях (механике, физике, технике) и вызвало их бурное проникновение в нетрадиционные области (управление, экономику, биологию и т.п.).

В настоящем учебном пособии приводятся основные понятия, идеи и методы вычислительной математики, овладение которыми необходимо для успешного решения конкретных задач разных областей. Книга адресована студентам и специалистам различных технических и экономических специальностей, для которых вычислительные методы не являются профилирующим предметом. Поэтому строгость и обоснованность численных методов не являлись при изложении самоцелью и большое внимание в пособии уделено анализу точности и достоверности, получаемых с помощью приближенных методов результатов.

Учитывая учебный характер издания, многие заимствования даны без ссылок. Список литературы не претендует на полноту; к этим и другим книгам могут обратиться читатели, желающие углубить свои знания по численным методам.

Тема 1 ЧИСЛЕННОЕ РЕШЕНИЕ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

Пусть задана функция $f(x)$ и требуется найти корни уравнения

$$f(x) = 0. \quad (1)$$

Эта задача разделяется на два этапа:

- локализация корней, т.е. нахождение таких отрезков на оси x , на каждом из которых находится только один корень;
- вычисление корня, принадлежащего заданному отрезку с заданной степенью точности.

Для решения задачи локализации корней нет каких-либо рекомендаций, кроме утверждения о том, что если на концах некоторого отрезка $[a, b]$ функция $f(x)$ имеет разные знаки, т.е. $f(a)f(b) < 0$ и монотонна, т.е. $f'(x)$ не меняет знак при $x \in [a, b]$, то на отрезке $[a, b]$ имеется единственный корень уравнения (1).

При решении практических задач очень часто отрезок $[a, b]$, содержащий требуемый корень, известен заранее.

Рассмотрим методы решения задачи о нахождении корня уравнения (1), принадлежащего заданному отрезку $[a, b]$ с заданной степенью точности.

Метод простых итераций (последовательных приближений) - является наиболее общим и многие другие методы можно представить как некоторую вариацию метода простых итераций.

Представим уравнение (1) в виде

$$x = s(x). \quad (2)$$

Это можно сделать, например, прибавив x к обеим частям уравнения (1). Очевидно, что при подстановке искомого корня x^* в уравнение (2), оно превращается в тождество $x^* \equiv s(x^*)$.

Рассмотрим последовательность чисел x_i , которая определяется следующим образом: $x_0 \in [a, b]$ - задано, $x_{k+1} = s(x_k)$.

При некоторых определенных свойствах функции $s(x)$ эта последовательность сходится к искомому корню, т.е. $x_k \rightarrow x^*$ при $k \rightarrow \infty$.

Метод простых итераций имеет наглядную геометрическую интерпретацию (см. рис. 1). Решением уравнения (2) будет абсцисса точки пересечения прямой $y = x$ с кривой $y = s(x)$. При выполнении итераций значение функции $s(x)$ в точке x_i необходимо отложить по оси абсцисс. Это можно сделать, если провести горизонталь до пересечения с прямой $y = x$ и из точки их пересечения опустить перпендикуляр на ось абсцисс. На рис. 1 показаны разные ситуации: *a* - сходимость к корню односторонняя; *b* - сходимость с разных сторон; *в*, *г* - расходящиеся итерационные процессы.

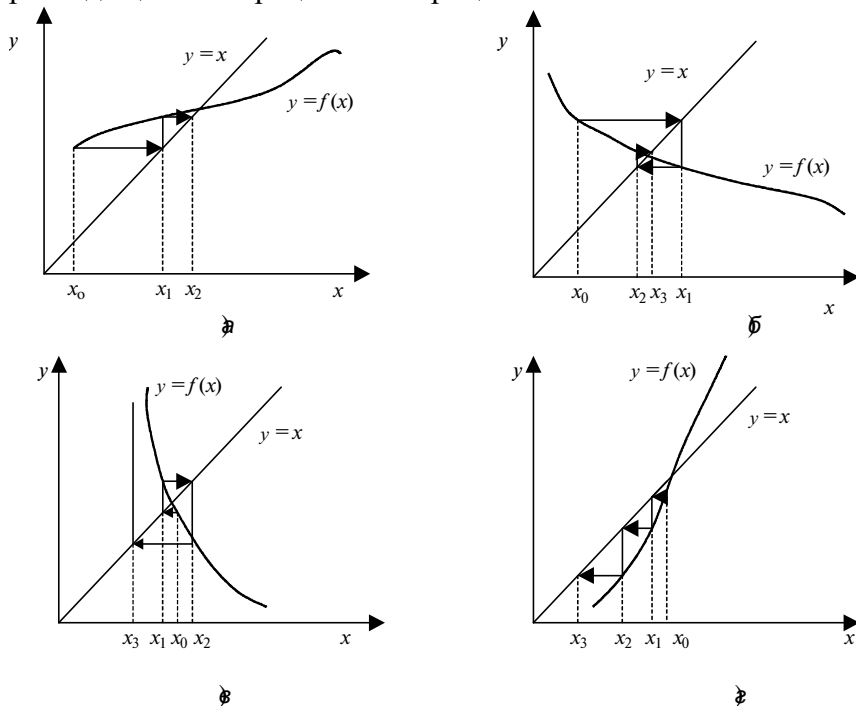


Рис. 1

Достаточные условия сходимости формулируются в теореме.

Теорема. Пусть в шаровой окрестности радиусом r некоторой точки ε $U_r(\varepsilon) = |x - \varepsilon| \leq r$ функция $s(x)$ удовлетворяет условиям:

1) Коши-Липшица: для любых $x_1, x_2 \in U_r(\varepsilon)$

$$|s(x_1) - s(x_2)| \leq q |x_1 - x_2|, \text{ причем } q = \text{const} < 1; \quad (4)$$

$$2) \quad |s(\varepsilon) - \varepsilon| \leq (1 - q)r. \quad (5)$$

Тогда уравнение (1) имеет на $U_r(\varepsilon)$ единственный корень x^* , последовательность (3) сходится к x^* при любом $x_0 \in U_r(\varepsilon)$, а для погрешности справедлива оценка

$$|x_k - x^*| \leq q^k |x_0 - x^*|. \quad (6)$$

На практике в качестве окрестности $U_r(\varepsilon)$ рассматривают интервал $[a, b]$, а вместо условия (4) обычно требуют выполнения менее сильного условия:

$$|s'(x)| \leq q < 1. \quad (7)$$

Условие (5) гарантирует, что все вычисляемые итерации x_k будут принадлежать окрестности $U_r(\varepsilon)$, но при решении практических задач доказать выполнение этого условия невозможно, тем более, что интервал поиска $[a, b]$ не совпадает с интервалом $U_r(\varepsilon)$. В итоге можно столкнуться с ситуацией, когда на отрезке $[a, b]$ выполнено условие (7), но уже x_1 выходит за границы интервала $[a, b]$ и метод итераций расходится. Для таких случаев можно рекомендовать каким-либо образом сузить интервал локализации, из которого выбирается начальное приближение. Такой подход обосновывается следующим следствием.

Следствие. Пусть в окрестности некоторой точки $U_r(\varepsilon)$ функция $s(x)$ удовлетворяет условию Коши-Липшица (4) и известно, что $x^* \in U_r(\varepsilon)$; тогда существует такая окрестность $U \in U_r(\varepsilon)$, содержащая искомый корень $x^* \in U$, что если выбрать $x_0 \in U$, то последовательность (3) сходится к корню x^* и имеет место оценка (6).

Скорость сходимости последовательности (3) такова, что на каждом шаге погрешность убывает в q раз:

$$|x_{k+1} - x^*| \leq q |x_k - x^*|;$$

ТАКАЯ СХОДИМОСТЬ НАЗЫВАЕТСЯ ЛИНЕЙНОЙ. ИСПОЛЬЗУЯ ЭТО НЕРАВЕНСТВО
МОЖНО ПОЛУЧИТЬ ОЦЕНКУ ЧИСЛА ИТЕРАЦИЙ, НЕОБХОДИМЫХ ДЛЯ
ДОСТИЖЕНИЯ ЗАДАННОЙ ТОЧНОСТИ, $Q^N |X_0 - X^*| \leq \varepsilon$.

Мажорируя неизвестную величину $|x_0 - x^*|$ размером интервала $[a, b]$, получим

$$n \geq \frac{\lg\left(\frac{\varepsilon}{b-a}\right)}{\lg q}.$$

ПРИ ПЕРЕХОДЕ ОТ УРАВНЕНИЯ (1) К УРАВНЕНИЮ (2) МОЖНО ИСПОЛЬЗОВАТЬ
РАЗЛИЧНЫЕ ПРИЕМЫ, НЕОБХОДИМО ТОЛЬКО, ЧТОБЫ ВСЕ КОРНИ УРАВНЕНИЙ (1)
И (2) СОВПАДАЛИ, А ИТЕРИРУЮЩАЯ ФУНКЦИЯ $S(X)$ УДОВЛЕТВОРЯЛА
УСЛОВИЯМ ТЕОРЕМЫ О СХОДИМОСТИ. СОБСТВЕННО МЕТОДОМ ПРОСТЫХ
ИТЕРАЦИЙ НАЗЫВАЕТСЯ МЕТОД, ПРИ КОТОРОМ ИТЕРИРУЮЩАЯ ФУНКЦИЯ
ЗАДАЕТСЯ В ВИДЕ

$$S(X) = F(X) + X.$$

РАССМОТРИМ ДРУГИЕ ВАРИАНТЫ ПОЛУЧЕНИЯ ИТЕРИРУЮЩЕЙ ФУНКЦИИ.
МЕТОД РЕЛАКСАЦИЙ: УМНОЖИМ УРАВНЕНИЕ (1) НА НЕКОТОРОЕ ЧИСЛО τ
 $= \text{CONST} \neq 0$ И ПРИМЕНИМ К НЕМУ МЕТОД ПРОСТЫХ ИТЕРАЦИЙ

$$\tau F(X) = 0; \quad X = \tau F(X) + X = S(X),$$

получим последовательность, $x_0 \in [a, b]$ задано, $x_{k+1} = \tau f(x_k) + x_k$.

ПОДБЕРЕМ ПАРАМЕТР τ ТАК, ЧТОБЫ ВЫПОЛНЯЛОСЬ УСЛОВИЕ СХОДИМОСТИ
 $|S'(X)| = |\tau F'(X) + 1| \leq 1$, ОТКУДА ПОЛУЧАЕМ $\tau F'(X) < 0$, $\tau > -2/F'(X)$.

Метод Ньютона (касательных). Зададим некоторое начальное приближение $x_0 \in [a, b]$ и линеаризуем функцию $f(x)$ в окрестности x_0 с помощью отрезка ряда Тейлора $f(x) = f(x_0) + f'(x_0)(x - x_0)$.

Вместо уравнения (1) решим линеаризованное уравнение

$$f(x_0) + f'(x_0)(x - x_0) = 0, \quad (8)$$

трактуя его решение x как первое приближение к корню:

$$x_1 = x_0 - f(x_0)/f'(x_0).$$

Продолжая этот процесс, приходим к формуле Ньютона

$$x_{k+1} = x_k - f(x_k)/f'(x_k), \quad (9)$$

которую можно считать итерационным процессом с итерирующей функцией $s(x) = x - f(x)/f'(x)$.

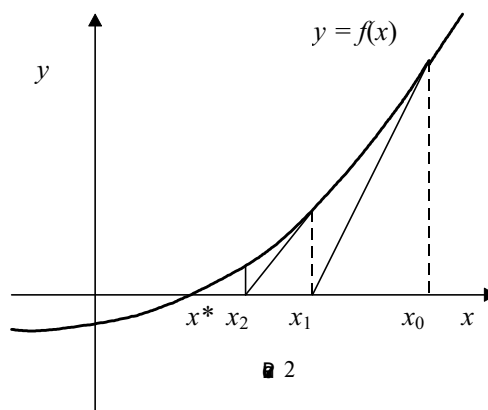
Геометрическая интерпретация этого процесса показана на рис. 2. Уравнение (8) является уравнением линии касательной к кривой $f(x)$ в точке x_0 , поэтому этот метод называют методом касательных.

Сходимость метода Ньютона оценивается неравенством

$$|x_{k+1} - x^*| \leq |x_k - x^*|^2 \frac{M_2}{m_1},$$

где $M_2 = \max |f''(x)|$,

$$m_1 = 2 \min |f'(x)|, \quad x \in [a, b].$$



Такая сходимость называется квадратичной, так как на каждом шаге погрешность

пропорциональна
квадрату предыдущей.

Следует отметить,
что улучшение
сходимости этого

метода по сравнению с предыдущими достигается увеличением затрат на выполнение каждого шага, так как в каждом шаге требуется вычислять не только значение функции $f(x)$, но и ее производной $f'(x)$.

Чтобы избавиться от необходимости вычислять производную, на каждом шаге используют так называемый модифицированный метод Ньютона, в котором производная вычисляется только однажды:

$$x_{k+1} = x_k - f(x_k)/f'(x_0).$$

ЭТОТ МЕТОД ТАКЖЕ МОЖНО СЧИТАТЬ МЕТОДОМ РЕЛАКСАЦИЙ С ПАРАМЕТРОМ

$$\tau = -1/f'(x_0).$$

Сходимость его как и метода релаксаций линейная.

Метод секущих - можно получить из метода Ньютона, заменив производную $f'(x)$ отношением разности функции к разности аргумента в окрестности рассматриваемой точки

$$f'(x) \approx (f(x+h) - f(x))/h.$$

Подставляя это выражение в (9), получим

$$x_{i+1} = x_i - f(x_i)h/(f(x_i+h) - f(x_i)). \quad (10)$$

Геометрически это означает, что приближенным значением корня считается точка пересечения секущей, проходящей через две точки функции $f(x_i)$ и $f(x_i+h)$, с осью абсцисс (рис. 3).

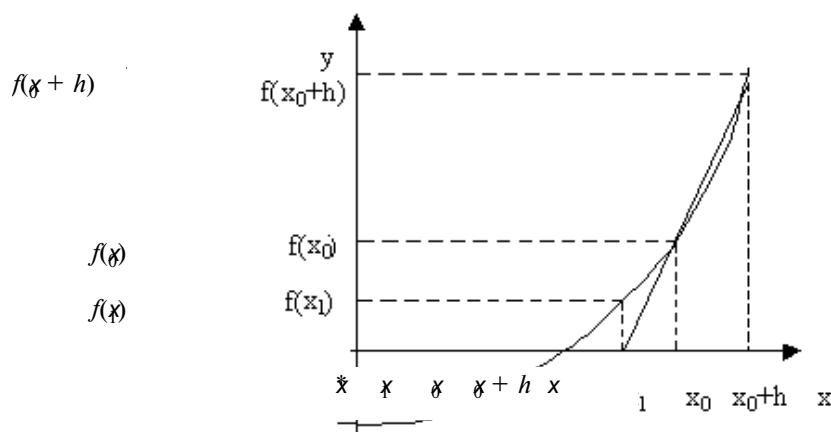


Рис. 3

При использовании этого метода следует уменьшать величину h по мере приближения к корню.

ВАРИАЦИЕЙ МЕТОДА СЕКУЩИХ ЯВЛЯЕТСЯ МЕТОД ЛОЖНЫХ ПОЛОЖЕНИЙ,

КОГДА ИХ ДЛЯ ПРОВЕДЕНИЯ СЕКУЩЕЙ ИСПОЛЬЗУЮТСЯ ТЕКУЩАЯ И ПРЕДЫДУЩАЯ ТОЧКИ. ПЕРВОЕ ПРИБЛИЖЕНИЕ ВЫЧИСЛЯЕТСЯ ПО ФОРМУЛЕ (10), А ОСТАЛЬНЫЕ - ПО ФОРМУЛЕ

$$x_{i+1} = x_i - f(x_i)(x_{i-1} - x_i)/(f(x_{i-1}) - f(x_i)).$$

Метод хорд - на отрезке $[a, b]$ строится хорда, стягивающая концы дуги графика функции $y = f(x)$, а точка пересечения хорды с осью абсцисс считается приближенным значением корня (рис. 4):

$$c = a - (f(a)(a - b))/(f(a) - f(b)),$$

$$c = b - (f(b)(a - b))/(f(a) - f(b)).$$

Следующее приближение ищется на интервале $[a, c]$ или $[c, b]$ в зависимости от знаков значений функции в точках a, b, c :

$$x^* \in [c, b], \text{ если } f(c)f(a) > 0;$$

$$x^* \in [a, c], \text{ если } f(c)f(b) < 0.$$

Если $f'(x)$ не меняет знак на $[a, b]$, то обозначая $c = x_1$ и считая начальным приближением a или b , получим итерационные формулы метода хорд с закрепленной правой или левой точкой:

$$x_0 = a, \quad x_{i+1} = x_i - f(x_i)(b - x_i)/(f(b) - f(x_i)) \text{ при } f'(x)f''(x) > 0;$$

$$x_0 = b, \quad x_{i+1} = x_i - f(x_i)(x_i - a)/(f(x_i) - f(a)) \text{ при } f'(x)f''(x) < 0.$$

Сходимость метода хорд линейная.

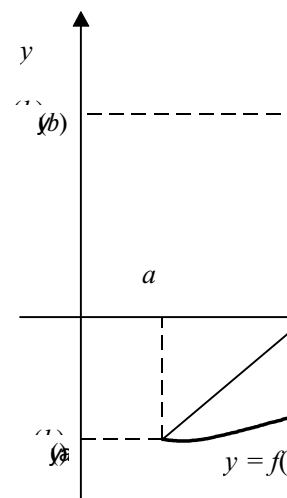
Рис. 4

Метод половинного деления - простой метод вычисления корня уравнения. Разделим исходный отрезок $[a, b]$ пополам: $c = (a + b)/2$.

Проверяя знаки $f(a), f(b), f(c)$, выясним, в каком из отрезков $[a, c]$ или $[c, b]$ содержится корень:

$$x^* \in [a, c], \text{ если } f(a)f(c) < 0; \quad x^* \in [c, b], \text{ если } f(c)f(b) < 0.$$

Выбранный отрезок принимаем за $[a, b]$ и повторяем это до тех пор пока получаемый отрезок не сожмется до заданной степени точности. При n



итерациях получим соотношение $(b - a)/2^n \leq \varepsilon$, из которого можно вычислить число итераций, необходимое для достижения заданной степени точности $n \geq \ln_2(b - a)/\varepsilon$.

Ввиду медленной сходимости этот метод редко используется для нахождения значения корня, обычно его применяют для локализации корня с дальнейшим уточнением значения корня каким-либо другим методом.

Т е м а 2 ЧИСЛЕННОЕ РЕШЕНИЕ СИСТЕМ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ

Необходимость решать системы линейных алгебраических уравнений (СЛАУ) возникает при реализации очень многих методов вычислительной математики. Для решения этой задачи разработано много различных методов, которые можно разделить на прямые и итерационные. Прямые методы позволяют теоретически получить точное решение задачи (т.е. при отсутствии ошибок округления при вычислениях) за конечное число операций. Итерационные методы позволяют вычислять некоторую последовательность векторов, сходящуюся к решению задачи.

2.1 Прямые методы решения СЛАУ

Задача ставится следующим образом: пусть требуется найти решение системы линейных алгебраических уравнений

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + a_{1,3}x_3 + \dots + a_{1,n}x_n &= b_1; \\ a_{2,1}x_1 + a_{2,2}x_2 + a_{2,3}x_3 + \dots + a_{2,n}x_n &= b_2; \\ &\vdots \\ &\vdots \\ &\vdots \\ &\vdots \\ a_{n,1}x_1 + a_{n,2}x_2 + a_{n,3}x_3 + \dots + a_{n,n}x_n &= b_n, \end{aligned} \quad (1)$$

или в векторной форме $AX = B$, где A - матрица коэффициентов; X - вектор неизвестных; B - вектор правых частей.

Будем считать, что $\Delta = \det A \neq 0$, т.е. решение существует и единственно.

Рассмотрим вначале прямые методы. В явном виде решение системы (1) записывается в виде формул Крамера $x_i = \Delta_i/\Delta$, где Δ_i - определитель матрицы, которая получается из матрицы A путем замены i -го столбца на столбец правых частей.

Этот метод очень неэкономичен, так как для его применения требуется $(n + 1)!$ операций, поэтому на практике используются различные варианты метода исключения переменных (Гаусса), который состоит из двух этапов: прямого хода, заключающегося в преобразовании исходной системы к системе с треугольной матрицей коэффициентов, и обратного хода, т.е. решения системы с треугольной матрицей.

Схема единственного деления. Пусть $a_{1,1} \neq 0$, если это не так, то всегда можно переставить уравнения таким образом, чтобы это требование выполнилось. Умножим первое уравнение системы (1) на множитель $l_{i1}^{(1)} = -a_{i,1}/a_{1,1}$ и складывая его с i -м уравнением исключим неизвестную x_1 из всех уравнений, кроме первого ($i = 2, 3, \dots, n$). Получим систему

$$a_{1,1}x_1 + a_{1,2}x_2 + a_{1,3}x_3 + \dots + a_{1,n}x_n = b_1;$$

$$\begin{array}{ccccccc}
0 & + & a_{2,2}^{(1)} x_2 & + & a_{2,3}^{(1)} x_3 & + & \dots + a_{2,n}^{(1)} x_n & = & b_2^{(1)}; \\
: & & : & & : & & : & & : \\
0 & + & a_{n,2}^{(1)} x_2 & + & a_{n,3}^{(1)} x_3 & + & \dots + a_{n,n}^{(1)} x_n & = & b_n^{(1)},
\end{array}$$

где $a_{ij}^{(1)} = a_{i,j} + l_i^{(1)} a_{1,j}$; $b_i^{(1)} = b_i + l_i^{(1)} b_1$, $i, j = 2, 3, 4, \dots, n$.

Предполагая, что $a_{2,2} \neq 0$ и используя множители $l_i^{(2)} = -a_{i,2}^{(1)}/a_{2,2}$, исключим x_2 из всех уравнений кроме, первого и второго. Аналогично из оставшихся уравнений исключим x_3 , затем x_4 и т.д. Выполнив $n - 1$ шагов, придем к системе с треугольной матрицей:

$$\begin{array}{ccccccc}
a_{1,1}x_1 & + & a_{1,2}x_2 & + & a_{1,3}x_3 & + & \dots + a_{1,n}x_n & = & b_1; \\
0 & + & a_{2,2}^{(1)}x_2 & + & a_{2,3}^{(1)}x_3 & + & \dots + a_{2,n}^{(1)}x_n & = & b_2; \\
0 & + & 0 & + & a_{3,3}^{(2)}x_3 & + & \dots + a_{3,n}^{(2)}x_n & = & b_3; \\
: & & : & & : & & : & & : \\
0 & + & 0 & + & 0 & + & \dots + a_{n,n}^{(n-1)}x_n & = & b_n.
\end{array} \quad (2)$$

Решение полученной системы (2) выполняется снизу вверх. Из последнего уравнения найдем x_n :

$$x_n = b_n^{(n-1)} / a_{n,n}^{(n-1)};$$

подставив его в предпоследнее уравнение найдем x_{n-1} :

$$x_{n-1} = (b_n^{(n-1)} - a_{n-1,n}^{(n-2)} x_n) / a_{n-1,n-1}^{(n-2)};$$

затем найдем x_{n-2} , x_{n-3} , x_{n-4} и последним x_1 по формулам

$$x_m = (b_m^{(m-1)} - a_{m,n}^{(m-1)} x_n - a_{m,n-1}^{(m-1)} x_{n-1} - \dots - a_{m,m+1}^{(m-1)} x_{m+1}) / a_{m,m}^{(m-1)}.$$

Этот метод требует $[n(n - 1)(4n - 7) + n^2]/6$ операций, что при больших n примерно равно $2n^3/3$.

При выполнении прямого хода из-за малости диагональных элементов, т.е. при $a_{i,i}$ близких к нулю могут возникать большие арифметические погрешности. Для уменьшения таких погрешностей используют схемы Гаусса с выбором главных элементов.

Выбор главного элемента по столбцам. Перед исключением x_1 найдем номер максимального по модулю элемента первого столбца

$$i_0 = \arg \{ \max |a_{i,1}| \}, \quad 1 \leq i \leq n$$

и поменяем местами первое и i_0 уравнения, после чего осуществим первый шаг исключения. Перед исключением x_2 отыскиваем следующий максимальный элемент во втором столбце

$$i_1 = \arg \{ \max |a_{i,2}^{(1)}| \}, \quad 2 \leq i \leq n$$

и осуществляем перестановку второго и i_1 -го уравнений. Продолжая таким образом, приведем систему к треугольному виду. Обратный ход выполняем так же как и в предыдущем случае.

Выбор главного элемента по строкам. Перед исключением x_1 найдем номер максимального по модулю элемента первой строки

$$j_0 = \arg \{ \max |a_{1,j}| \}, \quad 1 \leq j \leq n$$

и взаимно поменяем номера неизвестных x_1 и x_{j_0} (т.е. поменяем местами столбцы 1 и j_0), затем исключим переменную x_1 . На следующем шаге найдем следующий главный элемент

$$j_1 = \arg \{ \max |a_{2,j}^{(1)}| \}, \quad 2 \leq j \leq n$$

и, поменяв столбцы j_1 и 2, исключим переменную x_2 . Продолжая таким образом, приведем систему к треугольному виду и затем выполним обратный ход. Для восстановления исходной нумерации неизвестных выполним взаимные перестановки найденных значений в порядке, обратном их выполнению при прямом ходе.

Выбор главного элемента по всей матрице. Определим максимальный по модулю элемент матрицы $\max \{ |a_{i,j}| \}$ ($1 \leq i \leq n, 1 \leq j \leq n$). Пусть, например, это элемент с номером $i_0 j_0$. Уравнение (1) поменяем местами с уравнением i_0 , а переменную x_1 поменяем с переменной x_{j_0} и исключим x_1 из всех уравнений, кроме первого. Далее найдем новый главный элемент, не рассматривая первую строку и первый столбец $\max \{ |a_{i,j}^{(1)}| \}$ ($2 \leq i \leq n, 2 \leq j \leq n$), и выполним второй шаг исключения неизвестных. После приведения системы к треугольному виду и выполнения обратного хода переставим местами найденные значения неизвестных в порядке, обратном их перестановкам при прямом ходе.

ВАРИАНТЫ МЕТОДА ГАУССА С ВЫБОРОМ ГЛАВНЫХ ЭЛЕМЕНТОВ ПОЗВОЛЯЮТ СУЩЕСТВЕННО УМЕНЬШИТЬ НЕБЛАГОПРИЯТНОЕ ВЛИЯНИЕ ПОГРЕШНОСТЕЙ ОКРУГЛЕНИЯ НА РЕЗУЛЬТАТЫ РАСЧЕТА. ОДНАКО, ДЛЯ СИСТЕМ С ТАК НАЗЫВАЕМЫМ ДИАГОНАЛЬНЫМ ПРЕОБЛАДАНИЕМ МОЖНО НЕ ЗАБОТИТЬСЯ О ВЛИЯНИИ ПОГРЕШНОСТЕЙ ОКРУГЛЕНИЯ И ПРИМЕНЯТЬ СХЕМУ ЕДИНСТВЕННОГО ДЕЛЕНИЯ. УСЛОВИЕ ДИАГОНАЛЬНОГО ПРЕОБЛАДАНИЯ ЗАПИСЫВАЕТСЯ В ВИДЕ

$$|a_{i,i}| > \sum_{i=1, j=1}^n |a_{i,j}| \quad \text{при } i = 1, 2, 3, \dots, n.$$

ПРИ ВЫПОЛНЕНИИ ЭТОГО УСЛОВИЯ ПЕРЕД КАЖДЫМ ШАГОМ ИСКЛЮЧЕНИЯ НЕИЗВЕСТНЫХ ГЛАВНЫЙ ЭЛЕМЕНТ ВСЕГДА БУДЕТ НАХОДИТЬСЯ В НУЖНОЙ ПОЗИЦИИ.

ПРИ РЕАЛИЗАЦИИ ЧИСЛЕННЫХ МЕТОДОВ РЕШЕНИЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ ЧАСТО ВОЗНИКАЮТ СЛАУ СПЕЦИАЛЬНОГО ВИДА - ТРЕХДИАГОНАЛЬНЫЕ. У ТАКИХ СИСТЕМ МАТРИЦА КОЭФФИЦИЕНТОВ ИМЕЕТ ОТЛИЧНЫЕ ОТ НУЛЯ ЭЛЕМЕНТЫ ТОЛЬКО НА ГЛАВНОЙ ДИАГОНАЛИ И ДВУХ ДИАГОНАЛЯХ, ПРИЛЕЖАЩИХ К ГЛАВНОЙ. ЗАПИШЕМ ТАКУЮ СИСТЕМУ В СЛЕДУЮЩЕМ ВИДЕ:

$$1x_1 - d_1x_2 = f_1;$$

$$\begin{array}{rcll}
a_2x_1 - b_1x_2 + c_2x_3 & & & = f_2; \\
a_3x_2 - b_3x_3 - c_3x_4 & & & = f_3; \\
: & : & : & : \\
a_{n-1}x_{n-2} - b_{n-1}x_{n-1} + c_{n-1}x_n & & & = f_{n-1}; \\
-d_nx_{n-1} + 1x_n & & & = f_n. \quad (3)
\end{array}$$

КАК ПРАВИЛО, ТАКИЕ СИСТЕМЫ УДОВЛЕТВОРЯЮТ УСЛОВИЮ ДИАГОНАЛЬНОГО ПРЕОБЛАДАНИЯ, ЗАПИСАННОМУ В ВИДЕ

$$|d_1| < 1, \quad |d_n| < 1, \quad |b_i| > |a_i| + |c_i|, \quad i = 2, 3, \dots, n - 1.$$

Для таких систем используется специальный вариант метода исключения, называемый методом прогонки. Выразим x_1 из первого уравнения системы (3)

$$x_1 = d_1x_2 + f_1,$$

и, подставим его во второе, получим уравнение относительно x_2 и x_3 :

$$a_2(d_1x_2 + f_1) - b_2x_2 + c_2x_3 = f_2; \quad (4)$$

введем обозначения

$$d_2 = c_2/(b_2 - a_2d_1), \quad g_1 = f_1, \quad g_2 = (a_2g_1 - f_2)/(b_2 - a_2d_1)$$

и получим уравнение аналогичное (4):

$$x_2 = d_2x_3 + g_2.$$

Подставив x_2 в третье уравнение, получим уравнение относительно x_3 и x_4 . Поступая аналогично для k -го, уравнения, получим соотношения аналогичные (4):

$$x_k = d_kx_{k+1} + g_k \quad (5)$$

$$d_k = c_k/(b_k - a_k d_{k-1}), \quad g_k = (a_k g_{k-1} - f_k)/(b_k - a_k d_{k-1}); \quad (6)$$

дойдя до последнего уравнения, найдем из него последнюю переменную

$$x_n = (f_n + d_n g_{n-1})/(1 - d_n d_{n-1}), \quad (7)$$

а затем последовательно найдем все остальные переменные по уравнению (5).

Таким образом метод прогонки состоит из прямой прогонки, т.е. нахождения промежуточных коэффициентов $d_k, g_k, k = 1, 2, \dots, n - 1$ по формулам (6) и обратной прогонки, т.е. нахождения неизвестных по формулам (7) и (5).

Метод прогонки относится к экономичным методам и требует порядка n операций; это достигается за счет того, что над нулевыми элементами исходной матрицы никаких действий не выполняется.

Для оценки неустранимой погрешности при решении линейных систем необходимы некоторые сведения о нормах векторов и матриц.

Нормой вектора X называется положительное число, обозначаемое $\|X\|$ и удовлетворяющее условиям:

- 1) $\|X\| \geq 0, \|X\| = 0 \Leftrightarrow X = 0$;
- 2) $\|\alpha X\| = |\alpha| \|X\|, \alpha$ - скаляр;
- 3) $\|X + Y\| \leq \|X\| + \|Y\|$ (неравенство треугольника).

Способ введения нормы может быть различен, например $\|X\|_1 = \sum |x_i|$,

$$\|X\|_2 = \sqrt{\sum x_i^2} \text{ (Евклидова норма), } \|X\|_3 = \max_{i=1, n} |x_i| \text{ (равномерная норма).}$$

Пространство с введенной в нем нормой называется нормированным и метрическим, поскольку норма определяет метрику $\rho(X, Y)$ - расстояние между элементами:

$$\rho(X, Y) = \|X - Y\|.$$

Нормой квадратной матрицы A называется положительное число, обозначаемое $\|A\|$ и удовлетворяющее условиям:

- 1) $\|A\| \geq 0, \|A\| = 0 \Leftrightarrow A = 0$;
- 2) $\|\alpha A\| = |\alpha| \|A\|, \alpha$ - скаляр;
- 3) $\|A + B\| \leq \|A\| + \|B\|$, (неравенство треугольника);
- 4) $\|A B\| \leq \|A\| \|B\|$.

Норма матрицы вводится с помощью нормы вектора. Норма матрицы $\|A\|$ называется согласованной с нормой вектора $\|X\|$, если

$$\|A X\| \leq \|A\| \|X\|$$

Норма матрицы $\|A\|$ называется подчиненной норме вектора $\|X\|$, если норма матрицы $\|A\|$ вводится с помощью нормы вектора $\|X\|$ следующим образом:

$$\|A\| = \sup_{X \neq 0} \frac{\|A X\|}{\|X\|} = \sup_{\|X\|=1} \|A X\|.$$

Подчиненная форма является согласованной с соответствующей метрикой векторного пространства. В частности, используя нормы векторов $\|X\|_1, \|X\|_2$ получим нормы матриц

$$\|A\|_1 = \max_j \sum_{i=1}^n |a_{i,j}|;$$

$$\|A\|_2 = \sqrt{\rho(A^T A)};$$

$$\|A\|_3 = \max_i \sum_{j=1}^n |a_{i,j}|,$$

где $\rho(A^T A) = \max(\lambda_{A^T A})$ - спектральный радиус матрицы $A^T A$ (A^T - транспонированная матрица A); $\lambda_{A^T A}$ - собственные значения матрицы $A^T A$.

Для симметричной матрицы $A^T = A$, имеем $A^T A = A^2$,

$$\rho(A^T A) = \sqrt{\max(\lambda_{A^T A})^2}, \quad \|A\|_2 = \max |\lambda_A|.$$

Оценим сначала влияние на точность получаемого решения ошибок, содержащихся в исходных данных, предполагая, что арифметические действия выполняются без округлений.

Пусть вместо системы $AX = B$ имеем систему

$$(A + \delta A)(X + \delta X) = B + \delta B.$$

Здесь δA - матрица возмущений, задающая ошибки коэффициентов матрицы A ; δB - возмущения правых частей системы; δX - вектор ошибок, обусловленный влиянием возмущений δA , δB .

Преобразуем систему к виду

$$A\delta X + \delta A\delta X = \delta B - \delta AX$$

и, отбрасывая члены второго порядка малости $\delta A \delta X$, найдем ошибку решения

$$\delta X = A^{-1}(\delta B - \delta AX).$$

Используя рассмотренные нормы векторов и матриц, получим оценку величины погрешности

$$\|\delta X\| = \|A^{-1}(\delta B - \delta AX)\| \leq \|A^{-1}\| \left(\|B\| \frac{\|\delta B\|}{\|B\|} + \|A\| \frac{\|\delta A\|}{\|A\|} \|X\| \right).$$

Учитывая, что $\|B\| = \|AX\| \leq \|A\| \|X\|$, получим

$$\|\delta X\| \leq \|A^{-1}\| \|A\| \|X\| \left(\frac{\|\delta B\|}{\|B\|} + \frac{\|\delta A\|}{\|A\|} \right),$$

откуда оценка относительной погрешности решения равна

$$\frac{\|\delta X\|}{\|X\|} \leq \mu_A \left(\frac{\|\delta B\|}{\|B\|} + \frac{\|\delta A\|}{\|A\|} \right).$$

Здесь $\mu_A = \|A^{-1}\| \cdot \|A\|$ - число обусловленности матрицы A , именно эта величина определяет насколько сильно погрешности исходных данных влияют на решение системы. При умеренных значениях числа обусловленности ($1 < \mu_A < 10$) система считается хорошо обусловленной. Если μ_A велико ($\mu_A > 1000$), то система считается плохо обусловленной и ее решение сильно зависит от точности входных данных.

Оценку влияния на решение системы погрешностей округления при выполнении арифметических операций произведем следующим образом. Машинное решение $X^* = X + \delta X$ удовлетворяет системе с некоторой возмущенной матрицей $A + \delta A$:

$$(A + \delta A)X^* = B.$$

Для нормы матрицы эквивалентных возмущений справедлива оценка

$$\|\delta A\| \approx ng(A) \|A\| p^{-t},$$

где n - порядок системы; p - основание машинной арифметики (обычно $p = 2$); t - число цифр мантиссы;

$$g(A) = \left(\frac{\max_{i,j} \|a_{i,j}^k\|}{\max_{i,j} \|a_{i,j}\|} \right).$$

Величина $g(A)$ называется коэффициентом роста и показывает насколько могут возрасти пересчитываемые элементы матрицы A при выполнении прямого хода. Используя эту оценку, получим

$$\frac{\|\delta X\|}{\|X\|} \approx \mu_A n g(A) p^{-t}.$$

Если, например, решается система из 1000 уравнений с хорошей обусловленностью и диагональным преобладанием $\mu_A \approx 1$, $g(A) \approx 1$ на 32-разрядной машине ($p = 2$, $t = 26$), то нельзя получить точность выше

$$\frac{\|\delta X\|}{\|X\|} \approx 1000 \cdot 2^{-26} \approx 10^{-4},$$

а если при этом система плохо обусловлена $\mu_A \approx 10\,000$, то может произойти полная потеря точности.

2.2 Итерационные методы решения СЛАУ

При итерационном решении линейная система

$$AX = B. \quad (1)$$

преобразуется каким-либо образом к эквивалентной системе

$$X = PX + G, \quad (2)$$

а затем строится итерационный процесс

$$X^{(k)} = PX^{(k-1)} + G. \quad (3)$$

Сходимость итерационного процесса обеспечивается при выполнении условий следующей теоремы.

Теорема 1. Для сходимости итераций (3) к решению X^* системы (2) достаточно, чтобы в какой-либо норме выполнялось условие

$$\|P\| \leq q < 1, \quad (4)$$

при этом независимо от выбора $X^{(0)}$ справедлива оценка точности

$$\|X^{(k)} - X^*\| \leq q^k \|X^{(0)} - X^*\|. \quad (5)$$

Приведенная в теореме оценка точности показывает, что итерационный процесс имеет линейную скорость сходимости и позволяет оценить число итераций гарантирующих, что вычисленное решение будет отличаться от точного не более, чем на малую величину ε

$$\|X^{(k)} - X^*\| \leq q^k \|X^{(0)} - X^*\| \leq \varepsilon,$$

откуда получаем оценку k

$$k \geq \frac{\lg \frac{\varepsilon}{\|X^{(0)} - X^*\|}}{\lg q},$$

которой можно пользоваться, если мажорировать каким-то способом начальную погрешность $\|X^{(0)} - X^*\|$. Практически итерации прекращают, когда в какой-либо норме выполняется условие

$$\|X^{(k)} - X^{(k-1)}\| \leq \varepsilon.$$

Проще всего использовать равномерную норму

$$\|X\| = \max_{i=1,n} |x_i|.$$

ПРИ ПРИВЕДЕНИИ СИСТЕМЫ (1) К СИСТЕМЕ (2) ИСПОЛЬЗУЮТСЯ РАЗНЫЕ ПОДХОДЫ. РАССМОТРИМ НЕКОТОРЫЕ ИЗ НИХ.

Метод Якоби (простых итераций). Разрешим каждое из уравнений (1) относительно неизвестного в главной диагонали, т.е. из первого уравнения выразим x_1 , из второго x_2 и т.д.:

$$x_m = \frac{1}{a_{m,m}} (b_m - a_{m,1}x_1 - a_{m,2}x_2 - \dots - a_{m,m-1}x_{m-1} - a_{m,m+1}x_{m+1} - \dots - a_{m,n}x_n),$$

$$m = 1, 2, 3, \dots, n.$$

Получим систему (2) с матрицей вида

$$P = \begin{pmatrix} 0 & a_{1,2} & a_{1,3} & \dots & a_{1,n} \\ a_{2,1} & 0 & a_{2,3} & \dots & a_{2,n} \\ a_{2,2} & a_{2,2} & a_{2,2} & \dots & a_{2,2} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n,1} & a_{n,2} & a_{n,3} & \dots & 0 \\ a_{n,n} & a_{n,n} & a_{n,n} & \dots & \dots \end{pmatrix}.$$

Вычисление итераций в этом случае производится по формулам

$$x_m^{(k)} = \frac{1}{a_{m,m}} \left(b_m - \sum_{i=1}^{m-1} a_{m,i}x_i^{(k-1)} - \sum_{i=m+1}^n a_{m,i}x_i^{(k-1)} \right).$$

Этот метод сходится при выполнении условий диагонального преобладания (6), но при прочих равных условиях он имеет скорость сходимости примерно в два раза выше, чем метод Гаусса.

Метод релаксаций (однопараметрический метод итераций): умножим обе части уравнения (1) на некоторое число τ :

$$\tau XA = \tau B,$$

и приведем полученное уравнение к виду $X = (E - \tau A)X + \tau B$.

Фактически мы получили уравнение (2), где $P = (E - \tau A)$, $G = \tau B$.

Чтобы обеспечить сходимость итераций (3) параметр τ должен быть таким, чтобы в одной из норм выполнялось условие

$$\|P\| = \|E - \tau A\| \leq q < 1.$$

Для определения значения параметра τ будем использовать следующую теорему.

Теорема 2. Для сходимости итерационного процесса (3) к решению системы (1) необходимо и достаточно, чтобы все собственные значения ν_i матрицы P по абсолютной величине были меньше единицы:

$$|\nu_i| < 1.$$

Если матрица исходной системы симметрична и положительна ($A^T = A$, $A > 0$), то собственные значения λ_i матрицы A связаны с собственными значениями матрицы P соотношением $\nu_i = 1 - \tau \lambda_i$, из чего можно получить условие $\tau \leq \frac{2}{\max_i |\lambda_i|}$.

ВЫПОЛНИМ ОЦЕНКУ ВЛИЯНИЯ ОКРУГЛЕНИЙ ПРИ ВЫЧИСЛЕНИЯХ ИТЕРАЦИОННЫМИ МЕТОДАМИ НА ТОЧНОСТЬ ПОЛУЧАЕМЫХ РЕЗУЛЬТАТОВ. ЧТО КАСАЕТСЯ ВЛИЯНИЯ ОШИБОК ВХОДНЫХ ДАННЫХ, ТО ИХ ВЛИЯНИЕ БЫЛО РАССМОТРЕНО В ПРЕДЫДУЩЕМ РАЗДЕЛЕ НЕЗАВИСИМО ОТ МЕТОДА.

Будем использовать тот факт, что суммарный эффект ошибок округлений при выполнении одного шага итерационного процесса (3) можно трактовать как возмущение правой части уравнения, т.е. вектора G . Реальный процесс запишем в виде

$$\tilde{X}^{(n)} = P X^{(n-1)} + G + \delta G.$$

Тогда для разности между реально вычисляемыми приближениями $\tilde{X}^{(n)}$ и идеальными $X^{(n)}$, которые получались бы, если бы вычисления производились без округлений, получим

$$\tilde{X}^{(n)} - X^{(n)} = P(\tilde{X}^{(n-1)} - X^{(n-1)}) + \delta G$$

сходимости $\|P\| \leq q < 1$, и, оценивая разности векторов, получим

$$\begin{aligned} \|\tilde{X}^{(n)} - X^{(n)}\| &\leq q \|\tilde{X}^{(n-1)} - X^{(n-1)}\| + \|\delta G^{(n)}\| \leq \\ &\leq q^2 \|\tilde{X}^{(n-2)} - X^{(n-2)}\| + q \|\delta G^{(n-1)}\| + \|\delta G^{(n)}\| \leq \\ &\leq q^n \|\tilde{X}^{(0)} - X^{(0)}\| + (\max_k \|\delta G^{(k)}\|)(1 + q + q^2 + \dots + q^{n-1}). \end{aligned}$$

Поскольку начальное приближение не вычисляется, а задается, будем считать, что

$$\|\tilde{X}^{(0)} - X^{(0)}\| = 0.$$

Обозначая максимально возможную суммарную (в пределах одной итерации) неустранимую погрешность $\delta = \max_k \|\delta G^{(k)}\|$ и заменяя сумму геометрической прогрессии, получим окончательную оценку погрешности

$$\|\tilde{X}^{(n)} - X^{(n)}\| \leq \delta \frac{q^n}{q-1} \leq \frac{\delta}{1-q}. \quad (7)$$

Эта мера является мерой максимально возможной точности, поэтому бессмысленно задавать допустимый уровень погрешности ε меньше чем $\delta/(1-q)$.

Представление о порядке δ можно получить, используя статистические данные анализа погрешностей [19]. Выполнение N однотипных арифметических операций приводит к суммарной ошибке округления порядка $\Delta_0 \sqrt{N}$, где Δ_0 - погрешность выполнения одной операции. При использовании итерационных методов для вычисления одной компоненты результата требуется порядка n умножений и делений, где n - порядок системы. Следовательно, $\delta \approx \Delta_0 \sqrt{n}$.

Пусть, как и в примере, приведенном в предыдущем разделе, решается большая, плохо обусловленная система $n \approx 1000$, $\mu_A \approx 10\,000$.

Значение знаменателя в выражении (7) оценим с помощью числа обусловленности $1 - q \approx 1/\mu_A$, тогда точность будет оцениваться выражением

$$\frac{\delta}{1-q} \approx \frac{\Delta_0 \sqrt{1000}}{10^{-4}} \approx 10^5 \Delta_0.$$

При стандартной точности вычислительной машины $\Delta_0 \approx 10^{-8}$ можно получить точность порядка 10^{-2} , в то время как метод Гаусса при таких условиях приводит к полной потере точности.

Заметим, что если матрица системы сильно разрежена (т.е. число нулевых элементов порядка n), как это довольно часто бывает на практике, то оценка неустранимой погрешности при использовании итерационных методов уменьшится еще на два порядка, а оценка погрешности метода Гаусса никак не изменится.

Тема 3 ЧИСЛЕННОЕ РЕШЕНИЕ СИСТЕМ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

Рассмотрим систему нелинейных уравнений вида

$$f_i(x_1, x_2, \dots, x_n) = 0, \quad i = 1, 2, \dots, n;$$

или в векторной записи

$$F(X) = 0,$$

где X - вектор неизвестных величин, а F - вектор-функция:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad F(X) = \begin{bmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ \vdots \\ f_n(x_1, x_2, \dots, x_n) \end{bmatrix}.$$

Будем считать, что система (1) имеет хотя бы одно решение и что интересующее нас решение X^* принадлежит некоторой области U выявленной на стадии локализации решения.

Метод простых итераций. Приведем каждое из уравнений к виду

$$x_i = \varphi_i(x_1, x_2, \dots, x_n), \quad i = 1, 2, \dots, n;$$

или в векторной форме $\mathbf{X} = \Phi(\mathbf{X})$ с вектор функцией

$$\Phi = \begin{bmatrix} \varphi_1(x_1, x_2, \dots, x_n) \\ \varphi_2(x_1, x_2, \dots, x_n) \\ \vdots \\ \varphi_n(x_1, x_2, \dots, x_n) \end{bmatrix}.$$

Используя некоторое начальное приближение $\mathbf{X}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})$, построим итерационный процесс

$$\mathbf{X}^{(k+1)} = \Phi(\mathbf{X}^{(k)}).$$

Условие сходимости этого процесса к решению системы (1) формулируются в следующей теореме.

Теорема. Пусть \mathbf{X}^* решение системы (1). Если в окрестности $U_r(\mathbf{X}^*) = \{\|\mathbf{X} - \mathbf{X}^*\| < r\}$ вектор функция $\Phi(\mathbf{X})$ удовлетворяет условию Коши-Липшица, т.е. для любых $\mathbf{X}', \mathbf{X}'' \in U_r(\mathbf{X}^*)$ имеет место

$$\|\Phi(\mathbf{X}') - \Phi(\mathbf{X}'')\| \leq q \|\mathbf{X}' - \mathbf{X}''\|$$

с постоянной $q < 1$, то последовательность \mathbf{X}^k при $\mathbf{X}^{(0)} \in U_r(\mathbf{X}^*)$ сходится к решению \mathbf{X}^* и имеет место оценка сходимости

$$\|\mathbf{X}^{(k)} - \mathbf{X}^*\| \leq q \|\mathbf{X}^{(0)} - \mathbf{X}^*\|.$$

Приведенное в теореме условие Коши-Липшица в большинстве случаев достаточно сложно доказать, поэтому на практике в качестве достаточного критерия сходимости используется другое условие, аналогичное условию $|\varphi'(x)| < 1$ в случае одного уравнения.

Введем в рассмотрение матрицу Якоби от вектор-функции Φ :

$$M_\Phi = \begin{vmatrix} \frac{\partial \varphi_1}{\partial k_1} & \frac{\partial \varphi_1}{\partial k_2} & \dots & \frac{\partial \varphi_1}{\partial k_n} \\ \frac{\partial \varphi_2}{\partial k_1} & \frac{\partial \varphi_2}{\partial k_2} & \dots & \frac{\partial \varphi_2}{\partial k_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial \varphi_n}{\partial k_1} & \frac{\partial \varphi_n}{\partial k_2} & \dots & \frac{\partial \varphi_n}{\partial k_n} \end{vmatrix}.$$

Разность между компонентами искомого корня системы \mathbf{X}^* и найденным приближенным значением, используя теорему о среднем для функции нескольких переменных, можно записать в виде

$$x_i^{(k)} - x_i^* = \varphi(x_1^{(k-1)}, x_2^{(k-1)}, \dots, x_n^{(k-1)}) - \varphi(x_1^*, x_2^*, \dots, x_n^*) =$$

$$= \sum_{j=1}^n \frac{\partial \varphi_i}{\partial x_j} (x_j^{(k-1)} - x_j^*), \quad i = 1, 2, \dots, n;$$

или в векторной форме $\mathbf{X}^{(k)} - \mathbf{X}^* = \mathbf{M}_\Phi (\mathbf{X}^{(k-1)} - \mathbf{X}^*)$.

Пусть \mathbf{M} - мажорирующая матрица с элементами $m_{i,j} = \max_U \left| \frac{\partial \varphi_i}{\partial x_j} \right|$,

т.е. $\|\mathbf{M}_\Phi\| \leq \|\mathbf{M}\|$, тогда $\|\mathbf{X}^{(k)} - \mathbf{X}^*\| \leq \|\mathbf{M}_\Phi\| \|\mathbf{X}^{(k-1)} - \mathbf{X}^*\| \leq \|\mathbf{M}\| \|\mathbf{X}^{(k-1)} - \mathbf{X}^*\|$

и если $\|\mathbf{M}\| \leq q < 1$, то итерационный процесс сходится к \mathbf{X}^* . Таким образом, для сходимости итераций достаточно, чтобы в какой-либо норме выполнялось условие $\|\mathbf{M}\| < 1$.

Для увеличения скорости сходимости в методе простых итераций применяют схему Зейделя, которая предусматривает использование найденных компонент вектора $\mathbf{X}^{(k+1)}$ при вычислении других компонент этого вектора

$$x_1^{(k+1)} = \varphi_1(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)});$$

$$x_2^{(k+1)} = \varphi_2(x_1^{(k+1)}, x_2^{(k)}, \dots, x_n^{(k)});$$

$$\vdots$$

$$x_n^{(k+1)} = \varphi_n(x_1^{(k+1)}, x_2^{(k+1)}, \dots, x_n^{(k)}).$$

Метод Пикара. Пусть в уравнениях системы (1) можно выделить линейную $l(\mathbf{X})$ и нелинейную $g(\mathbf{X})$ части функций $f_i(\mathbf{X}) = l_i(\mathbf{X}) + g_i(\mathbf{X})$. Тогда систему уравнений (1) можно записать в виде $l_i(\mathbf{X}) = -g_i(\mathbf{X})$, $i = 1, 2, 3, \dots, n$; или в векторной форме $A\mathbf{X} = -\mathbf{G}(\mathbf{X})$, где A - матрица коэффициентов линейных частей уравнений; $\mathbf{G}(\mathbf{X})$ - вектор-функция нелинейных частей уравнений.

Выберем некоторый начальный вектор $\mathbf{X}^{(0)}$ и построим итерационный процесс в виде

$$A\mathbf{X}^{(k+1)} = -\mathbf{G}(\mathbf{X}^{(k)}).$$

Для выполнения одной итерации таким методом необходимо решать систему линейных уравнений, у которой вектором свободных членов будут нелинейные части функций $f_i(\mathbf{X})$. Причем поскольку матрица A остается неизменной при всех итерациях, то для решения СЛАУ можно использовать специальные алгоритмы, предусматривающие возможность преобразования только столбца свободных членов.

Метод Ньютона - можно рассматривать как линейризацию исходной задачи. Выберем в окрестности корня некоторое начальное приближение $\mathbf{X}^0 \in U_r(\mathbf{X}^*)$ и линейризуем уравнение (1) с помощью ряда Тейлора, ограничившись только величинами первого порядка малости

$$f_i(x_1^0, x_2^0, \dots, x_n^0) + \sum_{j=1}^n \frac{\partial f_i}{\partial x_j} (x_1^0, x_2^0, \dots, x_n^0) (x_j - x_j^0) = 0, \quad i = 1, 2, \dots, n.$$

Обозначим решение этого уравнения $\mathbf{X}^{(1)}$, используя его, получим $\mathbf{X}^{(2)}$. Итерационный процесс запишется в виде

$$\sum_{j=1}^n \frac{\partial f_i^{(k-1)}}{\partial x_j} (x_j^{(k)} - x_j^{(k-1)}) = -f_i^{(k-1)};$$

где $f_i^{(k)} = f_i(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})$, $\frac{\partial f_i^{(k)}}{\partial x_j} = \frac{\partial f_i(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})}{\partial x_j}$.

Введем в рассмотрение матрицу Якоби для вектор-функции $F(\mathbf{X})$

$$W = \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_n} \end{vmatrix}$$

Систему (1) можно записать в виде

$$W(\mathbf{X}^{(k-1)}) (\mathbf{X}^{(k)} - \mathbf{X}^{(k-1)}) = F(\mathbf{X}^{(k-1)}),$$

решив это уравнение относительно $\mathbf{X}^{(k)}$, получим

$$\mathbf{X}^{(k)} = \mathbf{X}^{(k-1)} - W^{-1}(\mathbf{X}^{(k-1)}) F(\mathbf{X}^{(k-1)}).$$

Как и для случая с одной переменной метод Ньютона можно модифицировать так, чтобы вычислять Якобиан только в одной точке:

$$\mathbf{X}^{(k)} = \mathbf{X}^{(k-1)} - W^{-1}(\mathbf{X}^{(0)}) F(\mathbf{X}^{(k-1)}).$$

Скорость сходимости модифицированного метода Ньютона значительно ниже, чем метода Ньютона.

Метод релаксаций. Перепишем систему (1) в виде $\mathbf{X} = \mathbf{X} + \tau F(\mathbf{X})$, где τ - некоторая константа, и построим итерационный процесс по схеме $\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} + \tau F(\mathbf{X}^{(k)})$.

Параметр τ должен быть таким, чтобы в окрестности решения выполнялось достаточное условие сходимости $\|E + \tau W\| < 1$, где E - единичная матрица.

На практике выполнение этого условия достаточно сложно проверить, поэтому значение параметра τ выбирают пробным путем, проверяя выполнение необходимого условия сходимости после выполнения каждой итерации

$$\|\mathbf{X}^{(k)} - \mathbf{X}^{(k-1)}\| < \|\mathbf{X}^{(k-1)} - \mathbf{X}^{(k-2)}\|.$$

Если окажется, что на какой-либо итерации это условие не выполняется, то необходимо изменить значение параметра τ .

Тема 4 ПРИБЛИЖЕНИЕ ФУНКЦИЙ ИНТЕРПОЛЯЦИОННЫМИ ПОЛИНОМАМИ

Пусть функция $f(x)$ задана для дискретного набора точек x_i множеством своих значений $f_i = f(x_i)$, т.е. таблицей, и пусть требуется найти приближенное значение $f(x)$ при $x \neq x_i$. При достаточно подробной таблице, т.е. когда $|x_i - x|$ мало и $x \approx x_i$, можно положить $f(x) \approx f(x_i)$ или более точно для $x \in [x_i, x_{i+1}]$

$$f(x) \approx f_i + \frac{f_{i+1} - f_i}{x_{i+1} - x_i} (x - x_i),$$

где $f_i = f(x_i)$.

Последнее выражение получено путем замены функции $f(x)$ отрезком прямой, проходящей через точки (x_i, f_i) , (x_{i+1}, f_{i+1}) .

В общем случае будем заменять функцию $f(x)$ полиномом степени n

$$P_n(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n,$$

совпадающим с функцией $f(x)$ в $n + 1$ табличных точках x_i , называемых узлами интерполяции

$$a_0 + a_1x_i + a_2x_i^2 + \dots + a_nx_i^n = f(x_i), \quad i = 0, 1, 2, \dots, n. \quad (1)$$

Система уравнений (1) может быть решена, если среди узлов x_i нет совпадающих. Используя решение этой системы, интерполяционный полином можно записать в виде полинома Лагранжа

$$\begin{aligned} P_n(x) = & f_0 \frac{(x-x_1)(x-x_2)\dots(x-x_n)}{(x_0-x_1)(x_0-x_2)\dots(x_0-x_n)} + \dots \\ & \dots + f_k \frac{(x-x_0)\dots(x-x_{k-1})(x-x_{k+1})\dots(x-x_n)}{(x_k-x_0)\dots(x_k-x_{k-1})(x_k-x_{k+1})\dots(x_k-x_n)} + \dots \\ & \dots + f_n \frac{(x-x_0)(x-x_1)\dots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\dots(x_n-x_{n-1})}; \end{aligned}$$

Полином Лагранжа удобно использовать, если требуется находить приближения различных функций заданных в одних и тех же табличных точках. В таких случаях можно предварительно вычислить коэффициенты Лагранжа

$$\begin{aligned} L_i^{(n)}(x) = & \frac{(x-x_0)(x-x_1)\dots(x-x_{k-1})(x-x_{k+1})\dots(x-x_n)}{(x_i-x_0)(x_i-x_1)\dots(x_i-x_{k-1})(x_i-x_{k+1})\dots(x_i-x_n)} = \\ = & \prod_{i=0, i \neq j}^n \frac{(x-x_j)}{(x_i-x_j)}, \end{aligned}$$

а для вычисления значений функций использовать формулу

$$f(x) \approx P_n(x) = \sum_{i=0}^n f_i L_i^{(n)}(x).$$

Заметим, что значения коэффициентов Лагранжа инвариантны относительно линейной подстановки, т.е. если

$$x = at + b, \quad x_i = at_j + b, \quad j = 0, 1, 2, \dots, n, \quad \text{то} \quad L_i^{(n)}(x) = L_j^{(n)}(t).$$

Если узлы интерполяции равноотстоящие и упорядочены по величине, так что $x_{i+1} - x_i = h = \text{const}$, т.е. $x_i = x_0 + ih$, то интерполяционный полином можно записать в форме Ньютона. Первая интерполяционная формула Ньютона имеет вид

$$P_n(x) \approx y_0 + q\Delta y_0 + \frac{q(q-1)}{2!} \Delta^2 y_0 + \frac{q(q-1)(q-2)\dots(q-n+1)}{n!} \Delta^n y_0,$$

где $q = \frac{(x-x_0)}{h}$, $y_i = f_i$, а выражения вида $\Delta^k y_i$ - конечные разности:

$\Delta y_i = y_{i+1} - y_i$ - разность первого порядка; $\Delta^2 y_i = \Delta y_{i+1} - \Delta y_i$ - разность второго порядка;
 $\Delta^k y_i = \Delta^{k-1} y_{i+1} - \Delta^{k-1} y_i$ - разность k -го порядка.

Вторая интерполяционная формула Ньютона записывается в виде

$$P_n(x) = y_n + q\Delta y_{n-1} + \frac{q(q+1)}{2!} \Delta^2 y_{n-2} + \frac{q(q+1)(q+2)\dots(q+n-1)}{n!} \Delta^n y_0,$$

где $q = \frac{x-x_n}{h}$.

Интерполяционные полиномы в форме Ньютона удобно использовать, если точка интерполирования находится вблизи начала (первая формула Ньютона) или конца таблицы (вторая формула Ньютона). Эти полиномы интересны тем, что каждая частичная сумма первых m слагаемых есть интерполяционный полином $(m - 1)$ степени, построенный по m первым (для первой формулы) или последним (для второй формулы) табличным точкам. Поэтому интерполяционные полиномы Ньютона удобно использовать при последовательном увеличении степени интерполяционного многочлена. Пусть, например, для вычисления значения $f(x)$ некоторой точке x вычислен полином $P_n(x)$ по $(n + 1)$ точкам x_0, x_1, \dots, x_n , а затем для уточнения этого значения привлечен еще один узел x_{n+1} , тогда для вычисления $P_{n+1}(x)$ по $(n + 2)$ узлам с помощью формулы Лагранжа нужно заново пересчитывать все слагаемые, а в первой формуле Ньютона требуется вычислить только одно слагаемое

$$\frac{q(q-1)(q-2)\dots(q-n-1)(q-n-2)}{(n+1)!} \Delta^{n+1} y_0,$$

по величине которого можно, кстати, определить и целесообразность дальнейшего увеличения степени интерполяции.

Если точка интерполирования x лежит вблизи какого-либо внутреннего узла, то обозначая этот узел x_0 , а узлы слева от него отрицательными индексами $x_{-1} = x_0 - h$, $x_{-2} = x_0 - 2h$ и т.д. можно записать интерполяционный полином в виде первой формулы Гаусса (для $x > x_0$)

$$P_n(x) = y_0 + q\Delta y_0 + \frac{q(q-1)}{2!} \Delta^2 y_{-1} + \frac{q(q^2-1)}{3!} \Delta^3 y_{-2} + \frac{q(q^2-1)\dots(q^2-(n-1)^2)(q-n)}{(2n)!} \Delta^{2n} y_{-n},$$

где $q = (x - x_0)/h$; или в виде второй формулы Гаусса (для $x < x_0$)

$$P_n(x) = y_0 + q\Delta y_{-1} + \frac{q(q+1)}{2!} \Delta^2 y_{-1} + \frac{q(q^2-1)}{3!} \Delta^3 y_{-2} + \frac{q(q^2-1)\dots(q^2-(n-1)^2)(q+n)}{(2n)!} \Delta^{2n} y_{-n}.$$

При интерполировании вблизи середины таблицы, т.е. при $x \approx x_0$, используется интерполяционная формула Ньютона-Бесселя

$$P_n(x) = \frac{y_0 + y_1}{2} + (q-0,5)\Delta y_0 + \frac{q(q-1)}{2!} \frac{\Delta^2 y_0 + \Delta^2 y_{-1}}{2} +$$

$$+ \frac{(q-0,5)q(q-1)}{3!} \Delta^3 y_{-1} +$$

$$+ \frac{(q+n-1)\dots(q-n+1)}{(2n-2)!} \frac{X\Delta^{2n-2} y_{1-n} + \Delta^{2n-2} y_{2-n}}{2}.$$

Если точка интерполирования x совпадает с серединой отрезка $[x_0, x_1]$ (интерполирование на середину), то формула Ньютона-Бесселя значительно упрощается

$$P\left(x_0 + \frac{h}{2}\right) = \frac{y_0 + y_1}{2} + \frac{\Delta^2 y_0 + \Delta^2 y_{-1}}{16} + \frac{3}{16} (\Delta^4 y_{-1} + \Delta^4 y_{-2}) + \dots$$

Ошибка приближения функции интерполяционным полиномом это разность $R_n(x) = f(x) - P_n(x)$. Используя теорему Ролля, можно получить следующие выражения для погрешности интерполяции:

- для формулы Лагранжа

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(N+1)!} (x-x_0)(x-x_1)\dots(x-x_n);$$

- для первой формулы Ньютона

$$R_n(x) = h^{n+1} \frac{q(q-1)(q-2)\dots(q-n)}{(n+1)!} f^{(n+1)}(\xi);$$

- для второй формулы Ньютона

$$R_n(x) = h^{n+1} \frac{q(q+1)(q+2)\dots(q+n)}{(n+1)!} f^{(n+1)}(\xi);$$

- для первой формулы Гаусса

$$R_n(x) = h^{2n+1} \frac{q(q^2-1^2)\dots(q^2-n^2)}{(2n+1)!} f^{(2n+1)}(\xi);$$

- для формулы Ньютона-Бесселя

$$R_n(x) = h^{2n} \frac{(q+n-1)\dots(q-n)}{(2n)!} f^{(2n)}(\xi),$$

где ξ - некоторая точка интервала $[x_0, x_n]$ или $[x_{-n}, x_n]$.

Конкретная величина погрешности $R_n(x)$ зависит от x и от положения точки ξ , но для оценки погрешности для формул Лагранжа и Ньютона справедливо соотношение

$$|R_n(x)| \leq \frac{\xi \max_{[x_0, x_n]} |f^{(n+1)}(\xi)|}{(n+1)!} |x-x_0| |x-x_1| \dots |x-x_n|,$$

а для формул Гаусса и Бесселя соотношение

$$|R_n(x)| \leq \frac{\xi \max_{[x_0, x_1]} |f^{(2n+1)}(\xi)|}{(2n+1)!} |x-x_{-n}| |x-x_{-n+1}| \dots |x-x_{n-1}| |x-x_n|.$$

Заметим, что за пределами отрезка интерполирования, когда $x \notin [x_0, x_n]$ (или $x \notin [x_{-n}, x_n]$) для формул Гаусса и Бесселя, т.е. при экстраполяции, оценка погрешности сильно возрастает.

Заметим также, что для некоторых функций $f(x)$, определенных на отрезке $x \in [a, b]$, интерполяционный многочлен $P_n(x)$ построенный по значениям $f(x_i)$ в равноотстоящих узлах $x_i = a + ih, x_0 = a, x_n = b, h = (b - a)/n$ с ростом числа узлов n , вовсе не будет иметь убывающую погрешность интерполирования. Например для функции $f(x) = 1/(x^2 + 0,5^2)$ на участке $[-1, 1]$ погрешность интерполяции при равноотстоящих узлах не стремится к нулю при увеличении n . Это обусловлено тем, что равноотстоящие узлы не являются лучшими с точки зрения уменьшения погрешности интерполирования. Если имеется возможность выбора узлов интерполирования, то их следует выбирать так, чтобы обеспечить минимум погрешности интерполяции $\min |R_n(x)|$. Такому требованию на отрезке $[-1, 1]$ удовлетворяют узлы полиномов Чебышева

$$T_{n+1}(x) = \frac{1}{2^n} \cos((n+1) \arccos x),$$

вычисляемые по формуле $x_k = \cos \frac{2k+1}{2(n+1)} \pi, k = 0, 1, 2, \dots, n$.

В случае произвольного отрезка $[a, b]$ узлы интерполяции вычисляются по формуле $x_k = \frac{a+b}{2} + \frac{b-a}{2} \cos \frac{2k+1}{2(n+1)} \pi, k = 0, 1, 2, \dots, n$.

Интерполяционный полином, построенный по таким узлам, является алгебраическим полиномом наилучшего приближения.

Кусочно-многочленная интерполяция. Поскольку увеличение степени интерполанта не всегда ведет к уменьшению погрешности интерполирования, для восстановления значений функции в произвольных точках можно пользоваться функцией, которая между двумя соседними узлами является полиномом заданной невысокой степени, например первой, второй и т.д. Соответствующая интерполяция называется кусочно-линейной, кусочно-квадратичной и т.д.

В случае кусочно-линейной интерполяции на каждом отрезке $x_i < x < x_{i+1}$ для приближения функции используется полином $f_i + \frac{f_{i+1} - f_i}{x_{i+1} - x_i} (x - x_i)$, а интерполант представляет собой ломаную линию, состоящую из отрезков прямых. При кусочно-квадратичной интерполяции интерполант состоит из отрезков парабол, проходящих через три соседних узла.

Такой подход к интерполированию приводит к тому, что интерполант в узлах интерполяции может не иметь производной даже первого порядка.

Кусочно-многочленная гладкая интерполяция (сплайны) - это кусочно-многочленная интерполяционная функция, обладающая заданным числом производных всюду, включая узлы интерполирования.

Пусть функция $f(x)$ определена на отрезке $[a, b]$ и известны ее значения в системе узлов $a = x_0 < x_1 < x_2 \dots < x_n = b$. Интерполяционным сплайном порядка m называется функция $S_m(x)$, удовлетворяющая следующим условиям:

- на каждом отрезке $[x_i, x_{i+1}], i = 1, 2, \dots, n$ $S_m(x)$ является многочленом степени m ;
- на всем отрезке $[a, b]$ $S_m(x)$ имеет непрерывные производные до порядка $(m - 1)$;
- $S_m(x_i) = f(x_i), i = 0, 1, 2, \dots, n$.

Если $m \geq 2$, то для единственности $S_m(x)$ необходимо задать еще $(m - 1)$ дополнительных условий, которые обычно задаются на концах отрезка $[a, b]$ либо произвольно, либо из дополнительной информации о поведении $f(x)$.

Рассмотрим построение кубического сплайна $S_3(x)$. По определению $S_3(x)$ на i -м отрезке $[x_i, x_{i+1}]$ является кубическим полиномом. Обозначим его $P_i(x)$ и будем искать в виде

$$P_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3, \quad i = 0, 1, 2, \dots, n. \quad (2)$$

По определению $S_m(x)$ должны выполняться условия

$$\begin{aligned} P_i(x_i) &= f(x_i), & i &= 0, 1, 2, \dots, n-1; \\ P_i(x_{i+1}) &= f(x_{i+1}), & i &= 0, 1, 2, \dots, n-1; \\ P'_i(x_{i+1}) &= P'_{i+1}(x_{i+1}), & i &= 0, 1, 2, \dots, n-2; \\ P''_i(x_{i+1}) &= P''_{i+1}(x_{i+1}), & i &= 0, 1, 2, \dots, n-2. \end{aligned} \quad (3)$$

Зададим дополнительные условия в виде

$$P''_1(x_0) = 0, \quad P''_{n-1}(x_n) = 0. \quad (4)$$

Введем обозначения

$$h_i = x_{i+1} - x_i, \quad g_i = \frac{f(x_{i+1}) - f(x_i)}{h_i}, \quad i = 0, 1, 2, \dots, n-1. \quad (5)$$

Подставляя в (3) - (4) выражение (2) и используя (5), получаем следующие соотношения:

$$a_i = f(x_i), \quad i = 0, 1, 2, \dots, n-1; \quad (6)$$

$$b_i + c_i h_i + d_i h_i^2 = a_{i+1}, \quad i = 0, 1, 2, \dots, n-1; \quad (7)$$

$$b_i + 2c_i h_i + 3d_i h_i^2 = b_{i+1}, \quad i = 0, 1, 2, \dots, n-2; \quad (8)$$

$$d_i = \frac{c_{i+1} - c_i}{3h_i}, \quad i = 0, 1, 2, \dots, n-2 \quad (9)$$

$$c_0 = 0, \quad c_{n-1} = 0. \quad (10)$$

Подставив (9) в (7) получим формулу

$$b_i = g_i - h_i(c_{i+1} - 2c_i)/3, \quad i = 0, 1, 2, \dots, n-1. \quad (11)$$

Используя (9) и (11) из (8) получим уравнение

$$h_i c_i + 2(h_i + h_{i+1})c_{i+1} + h_{i+1}c_{i+2} = 3(g_{i+1} - g_i), \quad i = 0, 1, 2, \dots, n-2. \quad (12)$$

Система (10), (12) является трехдиагональной системой линейных уравнений относительно коэффициентов c_i , решив ее методом прогонки из уравнений (9) можно найти значения d_i , а затем из уравнений (11) значения b_i . Значения коэффициентов a_i определяются из уравнений (6).

Тема 5 АППРОКСИМАЦИЯ ФУНКЦИЙ МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ

Рассмотрим принципиально другой, нежели интерполирование, подход к приближению функции заданной таблицей своих значений $f(x_i) = y_i$ в точках $x_i, i = 0, 1, 2, \dots, n$. Будем искать приближение в виде полинома степени m

$$P_m(x) = a_0 + a_1x + a_2x^2 + \dots + a_mx^m,$$

причем такого, чтобы значения этого полинома в табличных точках x_i не слишком сильно отличались от значений функции y_i , т.е. разности $|P_m(x_i) - y_i|$ должны быть малы. Для оценки степени приближения полинома $P_m(x)$ к точкам y_i используется метод наименьших квадратов, согласно которому наилучшее приближение достигается при минимуме суммы квадратов отклонений полинома от значений функции

$$\Delta(a_0, a_1, a_2, \dots, a_m) = \sum_{i=0}^m (P_m(x_i) - y_i)^2 \rightarrow \min.$$

Очевидно, что при $m = n$ решением такой задачи является интерполяционный многочлен, так как в табличных точках x_i его значения $P_m(x)$ равны значениям функции y_i и минимум квадратов разностей будет равен нулю $\Delta(a_0, a_1, a_2, \dots, a_m) = 0$.

При $m < n$ поставленная задача будет иметь единственное решение, которое можно найти, приравняв частные производные $\frac{\partial \Delta(a_0, a_1, \dots, a_m)}{\partial a_i}$ нулю. При $m > n$ поставленная задача будет иметь множество решений, для поиска одного из них можно задать произвольно $(m - n - 1)$ коэффициентов a_i , а остальные найти, решив задачу интерполяции.

Остановимся подробнее на случае $m < n$, т.е. когда число табличных точек больше числа параметров аппроксимационной функции. Приравняв нулю частные производные функции $\Delta(a_0, a_1, a_2, \dots, a_m)$, получим систему уравнений

$$\frac{\partial \sum_{i=0}^n (a_0 + a_1 x_i + a_2 x_i^2 + \dots + a_m x_i^m - y_i)^2}{\partial a_k} = 0, \quad k = 0, 1, 2, \dots, m;$$

или после преобразований

$$a_0 \sum_{i=0}^n x_i^k + a_1 \sum_{i=0}^n x_i^{k+1} + \dots + a_m \sum_{i=0}^n x_i^{k+m} = \sum_{i=0}^n y_i x_i^k, \quad k = 0, 1, 2, \dots, m.$$

Таким образом, для определения коэффициентов a_k получаем систему линейных уравнений с симметрической матрицей коэффициентов

$$\begin{bmatrix} n+1 & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \dots & \sum_{i=1}^n x_i^m \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \dots & \sum_{i=1}^n x_i^{m+1} \\ \dots & \dots & \dots & \dots & \dots \\ \sum_{i=1}^n x_i^m & \sum_{i=1}^n x_i^{m+1} & \sum_{i=1}^n x_i^{m+2} & \dots & \sum_{i=1}^n x_i^{m+m} \end{bmatrix}$$

и столбцом свободных членов

$$\begin{pmatrix} \sum_{i=0}^n y_i \\ \sum_{i=0}^n y_i x_i \\ \vdots \\ \sum_{i=0}^n y_i x_i^k \end{pmatrix},$$

элементы которых вычисляются по табличным данным. Найденный из решения этой системы полином называется среднеквадратичным приближением функции заданной таблицей.

Основная сфера применения таких полиномов - обработка экспериментальных данных (построение эмпирических формул). Дело в том, что интерполяционный полином, построенный по значениям функции, полученным с помощью эксперимента, будет

испытывать сильное влияние "экспериментального шума", к тому же при интерполировании узлы интерполяции не могут повторяться, т.е. нельзя использовать результаты повторных экспериментов при одинаковых условиях. Среднеквадратичный же полином сглаживает шумы и позволяет использовать результаты многократных экспериментов.

В ряде случаев аппроксимирующую функцию ищут не в виде полинома, а в виде некоторой другой зависимости, например показательной $y(x) = ab^x$. Если применить к этой зависимости метод наименьших квадратов, то получим систему нелинейных уравнений, решить которую весьма затруднительно. Чтобы облегчить решение можно произвести замену переменных таким образом, чтобы линеаризовать функцию $y(x)$. В частности, для показательной функции можно применить логарифмирование

$$\lg y = \lg a + x \lg b,$$

после замены переменных $q = \lg y$, $z = x$, $p = \lg a$, $s = \lg b$, получим линейную форму $q = p + zs$.

Для нахождения значений p и s методом наименьших квадратов необходимо решить систему линейных уравнений:

$$\begin{aligned} p(n+1) + s \sum_{i=0}^n z_i &= \sum_{i=0}^n q_i; \\ p \sum_{i=0}^n z_i + s \sum_{i=0}^n z_i^2 &= \sum_{i=0}^n q_i z_i. \end{aligned}$$

После нахождения значений p и s параметры a и b находятся по формулам $a=10^p$, $b=10^s$.

Далее приведены замены переменных для некоторых часто используемых нелинейных функций

Функция	Замена переменных
$y = ab^x$	$q = \lg y$, $z = \lg x$, $p = \lg a$, $s = \lg b$
$y = ax^b$	$q = \lg y$, $z = \lg x$, $p = a$, $s = \lg b$
$y = a + \frac{b}{x}$	$q = y$, $z = \frac{1}{x}$, $p = a$, $s = b$
$y = \frac{1}{ax+b}$	$q = \frac{1}{y}$, $z = x$, $p = a$, $s = b$
$y = \frac{x}{ax+b}$	$q = \frac{x}{y}$, $z = x$, $p = a$, $s = b$
$y = a \lg x + b$	$q = y$, $z = \lg x$, $p = a$, $s = b$

Если значения x или y отрицательны или нулевые, то логарифмирование и деление на ноль выполнить невозможно, поэтому следует сдвинуть шкалу x или y в положительную область. Например, для показательной функции $y = ab^x$ можно использовать замену переменных $q = \lg(y + w)$, $z = \lg(x + v)$, $p = \lg a$, $s = \lg b$, при этом получим аппроксимирующую функцию вида $y - w = ab^{x-v}$, где v и w сдвиги координат x и y . Для

обеспечения возможности вычислений эти значения должны удовлетворять условиям $v > \max_{x_i < 0} |x_i|$, $w > \max_{y_i < 0} |y_i|$, $i = 0, 1, 2, \dots, n$.

Тема 6 ПРИБЛИЖЕННОЕ ВЫЧИСЛЕНИЕ ПРОИЗВОДНЫХ

Численное дифференцирование используется для приближенного вычисления производных функции, заданной таблицей, и для функций, которые по разным причинам неудобно или невозможно дифференцировать аналитически. В последнем случае вычисляется таблица функции в окрестности исследуемой точки и по этим значениям вычисляется приближенное значение производной.

Итак, пусть в точках x_i , $i = 0, 1, 2, \dots, n$ известны значения функции $y_i = f(x_i)$. Способ построения формул численного дифференцирования состоит в том, что по табличным точкам строится интерполянт $P_n(x)$, который дифференцируется нужное число раз и делается допущение о том, что производная от функции приблизительно равна производной от интерполианта

$$\frac{d^k f(x)}{dx^k} \approx \frac{d^k P_n(x)}{dx^k}.$$

Погрешность такой формулы характеризуется k -ой производной от ошибки интерполяции $\frac{d^k R(x)}{dx^k}$.

При кусочно-линейной интерполяции по равноотстоящим узлам $x_i = x_0 + ih$

$$f(x) \approx P_1(x) = y_i + (y_{i+1} - y_i) \frac{x - x_i}{h},$$

получим очевидную формулу для $x \in [x_i, x_{i+1}]$

$$f'(x) \approx P_1'(x) = \frac{y_{i+1} - y_i}{h}, \quad (1)$$

либо, если использовать интерполяцию вида

$$f(x) \approx P_1(x) = y_i + (y_i - y_{i-1}) \frac{x - x_i}{h},$$

для $x \in [x_{i-1}, x_i]$ получим

$$f'(x) \approx P_1'(x) = \frac{y_i - y_{i-1}}{h}. \quad (2)$$

Формулы (1) и (2) называются правым разностным соотношением и левым разностным соотношением, их можно было бы записать сразу, используя определение производной и не привлекая интерполяцию как промежуточный этап. Однако для получения более точных формул и для вычисления производных высших порядков без интерполирования не обойтись.

Дифференцируя интерполяционный многочлен в виде первой формулы Ньютона, получим

$$f'(x) \approx P_n'(x) = \frac{1}{h} \left(\Delta y_0 + \frac{2q+1}{2!} \Delta^2 y_0 + \frac{3q^2 - 6q + 2}{3!} \Delta^3 y_0 + \dots \right), \quad (3)$$

где $q = (x - x_0)/h$.

Такая форма записи позволяет последовательно получать формулы дифференцирования различной степени точности, но для практических вычислений более удобна форма содержащая не конечные разности, а значения функции в табличных точках. Так, при квадратичной интерполяции для $x \in [x_i, x_{i+2}]$ получим

$$f'(x) \approx \frac{1}{h} [(q-1/2)y_i + 2qy_{i+1} + (q+1/2)y_{i+2}]. \quad (4)$$

При кубической интерполяции для

$$f'(x) \approx \frac{1}{h} \left(-\frac{3q^2 - 6q + 2}{6} y_i + \frac{3q^2 - 4q - 1}{2} y_{i+1} - \frac{3q^2 - 2q - 2}{2} y_{i+2} + \frac{3q^2 - 1}{6} y_{i+3} \right). \quad (5)$$

Если требуется вычислить производную в одной из табличных точек, то формулы упрощаются. В частности, для $x = x_i$ по формуле (4) получим

$$f'(x_i) \approx \frac{y_{i+1} - y_{i-1}}{2h}.$$

Это так называемое центральное разностное отношение.

При $x = x_0$ из (4) получим выражение

$$f'(x_0) \approx \frac{1}{h} (-1,5y_0 + 2y_1 - 0,5y_2),$$

а из (5) получаем

$$f'(x_0) \approx \frac{-11y_0 + 2y_1 - 9y_2 + 2y_3}{6h}.$$

При интерполировании по пяти точкам можно получить

$$f'(x_i) \approx \frac{y_{i-2} - 8y_{i-1} + 8y_{i+1} - y_{i+2}}{12h}.$$

Формулы для вычисления производных второго порядка получаются путем повторного дифференцирования интерполянта. Например, дифференцируя (3) еще раз и подставляя $x = x_0$, получим

$$f''(x_0) \approx \frac{1}{h^2} (\Delta^2 y_0 - \Delta^3 y_0 + \frac{11}{12} \Delta^4 y_0 - \frac{5}{6} \Delta^5 y_0 \dots).$$

При использовании значений функции вместо конечных разностей для квадратичной интерполяции получим формулу

$$f''(x_i) \approx \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2}.$$

При численном дифференцировании возникает методическая ошибка, вызванная заменой функции $f(x)$ интерполяционным многочленом $P_n(x)$, и ошибка округления, вызванная неточным заданием значений функции в табличных точках. Методическая погрешность оценивается производной от ошибки интерполирования $\frac{d^k P_n(x)}{dx}$ и при уменьшении шага, как правило, уменьшается. В частности, для формулы (1) величина методической погрешности оценивается соотношением

$$\left| f'(x) - \frac{dP_n(x)}{dx} \right| \leq \frac{M_2}{2} h,$$

где $M_2 = \max_{x \in [x_i, x_{i+1}]} |f''(x)|$.

Погрешность округления для всех случаев оценивается величиной $\frac{2\delta}{h}$, где δ - абсолютная погрешность исходных значений $f(x_i)$.

Таким образом, методическая погрешность при уменьшении шага интерполирования уменьшается, а погрешность округления обратно пропорциональна шагу и при уменьшении шага может как угодно сильно исказить получаемый результат, т.е. численное дифференцирование неустойчиво. Исходя из этого, шаг следует выбирать таким, чтобы суммарная погрешность была минимальна:

$$\Delta = \frac{M_2}{2} h + \frac{2\delta}{h} = \Phi(h) \rightarrow \min.$$

Приравнявая нулю производные $\frac{d\Phi}{dh}$, получим $\frac{M_2}{2} - \frac{2\delta}{h^2} = 0$, отсюда $h_{\text{опт}} = 2\sqrt{\frac{\delta}{M_2}}$. При этом $\Phi(h_{\text{опт}}) = 2\sqrt{M_2\delta}$.

Таким образом, в лучшем случае (при оптимальном шаге) можно лишь гарантировать, что производная $f'(x)$, вычисленная по формуле (1), будет отличаться от истинного значения не более чем на $2\sqrt{M_2\delta}$. Например, при $\delta \leq 0,01$ и $M_2 \leq 1$ шаг должен быть примерно 0,2 и при этом получим точность не выше 0,2.

Тема 7 ПРИБЛИЖЕННОЕ ВЫЧИСЛЕНИЕ ОПРЕДЕЛЕННЫХ ИНТЕГРАЛОВ

Вычисление определенных интегралов вида $\int_a^b f(x)dx$ основано на замене подынтегральной функции интерполяционным полиномом $P_n(x)$ по некоторой системе узлов и вычислении интеграла от интерполанта

$$\int_a^b f(x)dx \approx \int_a^b P_n(x)dx.$$

Получаемые таким образом формулы численного интегрирования называются квадратурными и имеют вид

$$\int_a^b f(x)dx \approx \sum_{i=0}^n C_i f(x_i),$$

где x_i - узлы квадратурных формул $x_i \in [a, b]$, $i = 0, 1, 2, \dots, n$; C_i - коэффициенты $\sum_{i=0}^n C_i = b - a$.

Конкретные квадратурные формулы различаются выбором узлов и значениями коэффициентов.

Погрешность такой формулы характеризуется k -ой производной от ошибки интерполяции $\frac{d^k R(x)}{dx}$.

При кусочно-линейной интерполяции по равноотстоящим узлам $x_i = x_0 + ih$

$$f(x) \approx P_1(x) = y_i + (y_{i+1} - y_i) \frac{x - x_i}{h},$$

получим очевидную формулу для $x \in [x_i, x_{i+1}]$

$$f'(x) \approx P_1'(x) = \frac{y_{i+1} - y_i}{h}, \quad (1)$$

либо, если использовать интерполяцию вида

$$f(x) \approx P_1(x) = y_i + (y_i - y_{i-1}) \frac{x - x_i}{h},$$

для $x \in [x_{i-1}, x_i]$ получим

$$f'(x) \approx P_1'(x) = \frac{y_i - y_{i-1}}{h}. \quad (2)$$

Формулы (1) и (2) называются правым разностным соотношением и левым разностным соотношением, их можно было бы записать сразу, используя определение производной и не привлекая интерполяцию как промежуточный этап. Однако для получения более точных формул и для вычисления производных высших порядков без интерполирования не обойтись.

Дифференцируя интерполяционный многочлен в виде первой формулы Ньютона, получим

$$f'(x) \approx P'_n(x) = \frac{1}{h} \left(\Delta y_0 + \frac{2q+1}{2!} \Delta^2 y_0 + \frac{3q^2-6q+2}{3!} \Delta^3 y_0 + \dots \right), \quad (3)$$

где $q = (x - x_0)/h$.

Такая форма записи позволяет последовательно получать формулы дифференцирования различной степени точности, но для практических вычислений более удобна форма содержащая не конечные разности, а значения функции в табличных точках. Так, при квадратичной интерполяции для $x \in [x_i, x_{i+2}]$ получим

$$f'(x) \approx \frac{1}{h} [(q-1/2)y_i + 2qy_{i+1} + (q+1/2)y_{i+2}]. \quad (4)$$

При кубической интерполяции для

$$f'(x) \approx \frac{1}{h} \left(-\frac{3q^2-6q+2}{6} y_i + \frac{3q^2-4q-1}{2} y_{i+1} - \frac{3q^2-2q-2}{2} y_{i+2} + \frac{3q^2-1}{6} y_{i+3} \right). \quad (5)$$

Если требуется вычислить производную в одной из табличных точек, то формулы упрощаются. В частности, для $x = x_i$ по формуле (4) получим

$$f'(x_i) \approx \frac{y_{i+1} - y_{i-1}}{2h}.$$

Это так называемое центральное разностное отношение.

При $x = x_0$ из (4) получим выражение

$$f'(x_0) \approx \frac{1}{h} (-1,5y_0 + 2y_1 - 0,5y_2),$$

а из (5) получаем

$$f'(x_0) \approx \frac{-11y_0 + 2y_1 - 9y_2 + 2y_3}{6h}.$$

При интерполировании по пяти точкам можно получить

$$f'(x_i) \approx \frac{y_{i-2} - 8y_{i-1} + 8y_{i+1} - y_{i+2}}{12h}.$$

Формулы для вычисления производных второго порядка получаются путем повторного дифференцирования интерполянта. Например, дифференцируя (3) еще раз и подставляя $x = x_0$, получим

$$f''(x_0) \approx \frac{1}{h^2} (\Delta^2 y_0 - \Delta^3 y_0 + \frac{11}{12} \Delta^4 y_0 - \frac{5}{6} \Delta^5 y_0 \dots).$$

При использовании значений функции вместо конечных разностей для квадратичной интерполяции получим формулу

$$f''(x_i) \approx \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2}.$$

При численном дифференцировании возникает методическая ошибка, вызванная заменой функции $f(x)$ интерполяционным многочленом $P_n(x)$, и ошибка округления, вызванная неточным заданием значений функции в табличных точках. Методическая погрешность оценивается производной от ошибки интерполирования $\frac{d^k P_n(x)}{dx}$ и при уменьшении шага, как правило, уменьшается. В частности, для формулы (1) величина методической погрешности оценивается соотношением

$$\left| f'(x) - \frac{dP_n(x)}{dx} \right| \leq \frac{M_2}{2} h,$$

где $M_2 = \max_{x \in [x_i, x_{i+1}]} |f''(x)|$.

Погрешность округления для всех случаев оценивается величиной $\frac{2\delta}{h}$, где δ - абсолютная погрешность исходных значений $f(x_i)$.

Таким образом, методическая погрешность при уменьшении шага интерполирования уменьшается, а погрешность округления обратно пропорциональна шагу и при уменьшении шага может как угодно сильно исказить получаемый результат, т.е. численное дифференцирование неустойчиво. Исходя из этого, шаг следует выбирать таким, чтобы суммарная погрешность была минимальна:

$$\Delta = \frac{M_2}{2}h + \frac{2\delta}{h} = \Phi(h) \rightarrow \min.$$

Приравняв нулю производные $\frac{d\Phi}{dh}$, получим $\frac{M_2}{2} - \frac{2\delta}{h^2} = 0$, откуда $h_{\text{опт}} = 2\sqrt{\frac{\delta}{M_2}}$. При этом $\Phi(h_{\text{опт}}) = 2\sqrt{M_2\delta}$.

Таким образом, в лучшем случае (при оптимальном шаге) можно лишь гарантировать, что производная $f'(x)$, вычисленная по формуле (1), будет отличаться от истинного значения не более чем на $2\sqrt{M_2\delta}$. Например, при $\delta \leq 0,01$ и $M_2 \leq 1$ шаг должен быть примерно 0,2 и при этом получим точность не выше 0,2.

Т е м а 7 П Р И Б Л И Ж Е Н Н О Е В Ы Ч И С Л Е Н И Е О П Р Е Д Е Л Е Н Н Ы Х И Н Т Е Г Р А Л О В

Вычисление определенных интегралов вида $\int_a^b f(x)dx$ основано на замене подынтегральной функции интерполяционным полиномом $P_n(x)$ по некоторой системе узлов и вычислении интеграла от интерполянта

$$\int_a^b f(x)dx \approx \int_a^b P_n(x)dx.$$

Получаемые таким образом формулы численного интегрирования называются квадратурными и имеют вид

$$\int_a^b f(x)dx \approx \sum_{i=0}^n C_i f(x_i),$$

где x_i - узлы квадратурных формул $x_i \in [a, b]$, $i = 0, 1, 2, \dots, n$; C_i - коэффициенты $\sum_{i=0}^n C_i = b - a$.

Конкретные квадратурные формулы различаются выбором узлов и значениями коэффициентов.

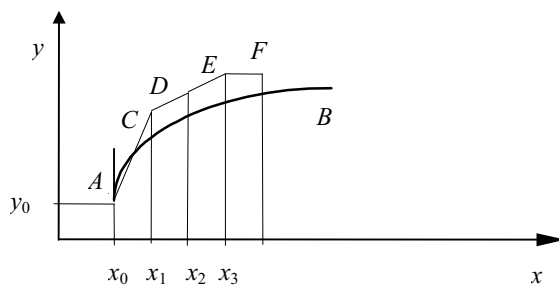


Рис. 1 Решение задачи Коши методом Эйлера

Метод Эйлера очень прост в реализации, но имеет невысокую точность. Он является методом первого порядка точности относительно шага h , а его глобальная погрешность оценивается выражением

$$|y_i - y_i^*| < \text{const}h = O(h),$$

где y_i^* - точное значение задачи (1), (2) в точке x_i ; $O(h)$ - погрешность на одном шаге h .

Усовершенствованный метод Эйлера (метод Эйлера второго порядка). В этом методе для вычисления функции $y(x)$ в одной точке требуется дважды вычислить функцию $f(x, y)$:

$$y_{i+1} = y_i + hf(x_i + h/2, y_i + hf(x_i, y_i)).$$

Погрешность этого метода пропорциональна h^2 , т.е. $|y_i - y_i^*| < O(h^2)$.

Метод Эйлера-Коши также относится к методам второго порядка и тоже требует двукратного вычисления функции $f(x, y)$:

$$y_{i+1}^0 = y_i + hf(x_i, y_i); \quad (4)$$

$$y_{i+1} = y_i + (f(x_i, y_i) + f(x_{i+1}, y_{i+1}^0)) h/2. \quad (5)$$

Методы Эйлера относятся к группе с общим названием метода Рунге-Кутты, к этой же группе принадлежит и метод, называемый методом Рунге-Кутты четвертого порядка. Согласно этому методу для вычисления одного значения функции $y(x)$ необходимо вычислить функцию $f(x, y)$ в четырех точках:

$$K1_i = f(x_i, y_i);$$

$$K2_i = f(x_i + h/2, y_i + K1_i/2);$$

$$K3_i = f(x_i + h/2, y_i + K2_i/2);$$

$$K4_i = f(x_i + h, y_i + K3_i);$$

$$y_{i+1} = y_i + h (K1_i + 2K2_i + 2K3_i + K4_i)/6.$$

Погрешность этого метода пропорциональна h^4 , т.е. $|y_i - y_i^*| < O(h^4)$.

Методы Рунге-Кутты относятся к так называемым одношаговым методам, поскольку для вычисления значения функции $y(x)$ в точке x_{i+1} требуется знать только значение функции $y(x)$ в одной предыдущей точке x_i .

Многошаговые методы, называемые также методами Адамса, построены путем интерполирования по нескольким соседним точкам; для их использования необходимо знать значение функции $y(x)$ в нескольких предыдущих точках. Достоинство многошаговых методов состоит в том, что независимо от порядка метода для вычисления значения функции $y(x)$ в одной точке требуется один раз вычислить функцию $f(x, y)$.

Метод Адамса второго порядка записывается следующим образом

$$y_{i+1} = y_i + h(3y'_i - y'_{i-1})/2,$$

где $y'_i = f(x_i, y_i)$.

Методы Адамса третьего и четвертого порядков имеют вид

$$y_{i+1} = y_i + h (23y'_i - 16y'_{i-1} - 5y'_{i-2})/12;$$

$$y_{i+1} = y_i + h (55y'_i - 59y'_{i-1} + 37y'_{i-2} - 9y'_{i-3})/24.$$

Погрешность решения, найденного многошаговым методом, оценивается равенством $|y_i - y_i^*| < O(h^m)$, где m - порядок метода.

Таким образом, метод Рунге-Кутты четвертого порядка и метод Адамса четвертого порядка имеют одинаковую оценку погрешности, но метод Адамса требует примерно вчетверо меньшего объема вычислений.

При получении формулы Эйлера (3) была использована замена производной правой конечной разностью $y'_i = (y_{i+1} - y_i)/h$, если использовать левую конечную разность $y'_i = (y_i - y_{i-1})/h$, получим уравнение

$$y_i - y_{i-1} - h f(x_i, y_i) = 0,$$

в которое искомая величина y_i входит нелинейным образом. Для решения такого уравнения можно использовать итерационные методы. Задавая начальное приближение $y_i^{(0)} = y_{i-1}$ и последовательно уточняя значение y_i по формуле

$$y_i^{(k+1)} = y_{i-1} + h f(x_i, y_i^{(k)}),$$

получаем неявный метод Эйлера (или метод Эйлера с итерационной обработкой). Итерации выполняют до тех пор, пока не выполнится условие $|y_{i+1}^{(k)} - y_{i+1}^{(k-1)}| < \varepsilon$.

Если после трех-четырех итераций требуемая точность ε не достигается, то производится уменьшение шага.

Неявную схему можно применить и для методов высших порядков, при этом получаются, так называемые методы прогноза и коррекции.

Метод Эйлера-Коши (4)-(5) можно рассматривать как один из простейших методов типа прогноза и коррекции, согласно которому вначале вычисляется грубое приближение $y_{i+1}^{(0)}$ по методу Эйлера, а затем оно используется для уточнения по формуле (5), причем коррекция может быть использована многократно:

$$y_{i+1}^{(k+1)} = y_i + (f(x_i, y_i) + f(x_{i+1}, y_{i+1}^{(k)})) h/2.$$

Методы прогноза и коррекции можно получить на основе использования многошаговых методов:

- второго порядка

$$y_{i+1}^{(0)} = y_i + h(3y'_i - y'_{i-1})/2;$$

$$y_{i+1}^{(k+1)} = y_i + h(y'_{i+1}^{(k)} + y'_i)/2,$$

- третьего порядка

$$y_{i+1}^{(0)} = y_i + h(23y'_i - 16y'_{i-1} + 5y'_{i-2})/12;$$

$$y_{i+1}^{(k+1)} = y_i + h(5y'_{i+1}^{(k)} + 8y'_i - y'_{i-1})/12,$$

- четвертого порядка

$$y_{i+1}^{(0)} = y_i + h(55y'_i - 59y'_{i-1} + 37y'_{i-2} - 9y'_{i-3})/24;$$

$$y_{i+1}^{(k+1)} = y_i + h(9y'_{i+1}^{(k)} + 19y'_i - 5y'_{i-1} + y'_{i-2})/24.$$

Во всех этих формулах для вычисления $y'_{i+1}^{(k)}$ используется значение y_{i+1} полученное в текущем приближении

$$y'_{i+1}^{(k)} = f(x_{i+1}, y_{i+1}^{(k)}).$$

В методе Милна используются формулы прогноза и коррекции построенные по разным точкам

$$y_i^{(0)} = y_{i-4} + 4h(2y'_{i-3} - y'_{i-2} + 5y'_{i-1})/3;$$

$$y_{i+1}^{(k+1)} = y_{i-2} + h(y'_{i-2} + 4y'_{i-1} - y'^{(k)}_i)/3.$$

Задача Коши для обыкновенного дифференциального уравнения высшего порядка ставится следующим образом: найти решение уравнения

$$y^{(n)} = f(x, y^{(n-1)}, y^{(n-2)}, \dots, y', y), \quad (6)$$

удовлетворяющее начальным условиям

$$y^{(n-1)}(x_0) = y_0^{(n-1)};$$

$$y^{(n-2)}(x_0) = y_0^{(n-2)};$$

⋮

$$y'(x_0) = y_0';$$

$$y(x_0) = y_0.$$

Путем введения дополнительных переменных

$$z = y, \quad z_1 = y', \quad z_2 = y^{(2)}, \quad z_3 = y^{(3)}, \quad \dots, \quad z_{n-1} = y^{(n-1)}$$

можно преобразовать уравнение (6) в систему уравнений первого порядка:

$$z_{n-1}' = f(x, z_{n-1}, z_{n-2}, z_{n-3}, \dots, z', z);$$

$$z_{n-2}' = z_{n-1};$$

$$z_{n-3}' = z_{n-2}; \quad (8)$$

⋮

$$z_1' = z_2;$$

$$z' = z_1$$

с начальными условиями

$$z_{n-1}(x_0) = y_0^{(n-1)};$$

$$z_{n-2}(x_0) = y_0^{(n-2)}; \quad (9)$$

⋮

$$z_1(x_0) = y_0';$$

$$z(x_0) = y_0.$$

Таким образом, задача Коши (6), (7) для дифференциального уравнения степени n эквивалентна задаче Коши (8), (9) для системы n дифференциальных уравнений первого порядка. Поэтому далее будем рассматривать только методы решения систем дифференциальных уравнений.

Для решения задачи Коши для системы дифференциальных уравнений можно использовать те же методы, что и для одного уравнения, но все этапы решения нужно выполнять по всем переменным одновременно. Проще всего представить формулы, реализующие соответствующий метод для одного уравнения в векторной форме, а затем заменить векторную форму алгебраической. Рассмотрим это на примере задачи Коши для системы двух уравнений

$$\begin{aligned} t' &= v(x, t, s), & t(x_0) &= t_0; \\ s' &= w(x, t, s), & s(x_0) &= s_0. \end{aligned} \quad (10)$$

Метод Рунге-Кутты второго порядка для одного уравнения можно записать в виде

$$K_i = hf(x_i, y_i); \quad y_{i+1} = y_i + hf(x_i + h/2, y_i + K_i/2).$$

Запишем это в векторной форме:

$$K_i = hf(x_i, y_i); \quad y_{i+1} = y_i + hf(x_i + h/2, y_i + K_i/2).$$

В соответствие с задачей Коши (10) векторы будут состоять из компонент

$$y_i = \{t_i, s_i\}; \quad K_i = \{kt_i, ks_i\}; \quad f = \{v, w\}.$$

Заменяя векторную форму алгебраической, получим формулы

$$\begin{aligned} kt_i &= hv(x_i, t_i, s_i); & ks_i &= hw(x_i, t_i, s_i); \\ t_{i+1} &= t_i + hv(x_i + h/2, t_i + kt_i/2, s_i + ks_i/2); \end{aligned} \quad (11)$$

$$s_{i+1} = s_i + hw(x_i + h/2, t_i + kt_i/2, s_i + ks_i/2).$$

Метод Адамса второго порядка для задачи (10) реализуется следующим образом:

$$t_{i+1} = t_i + h(v(x_i, t_i, s_i) + v(x_{i-1}, t_{i-1}, s_{i-1}))/2;$$

$$s_{i+1} = s_i + h(w(x_i, t_i, s_i) + w(x_{i-1}, t_{i-1}, s_{i-1}))/2.$$

Еще раз заметим, что все этапы вычислений нужно полностью произвести для всех переменных, и только после этого переходить к следующему этапу. Например, при вычислении по формулам (11) можно менять последовательность вычислений первой и второй строк и третьей и четвертой, но нельзя после первой строки вычислять третью, а затем вторую и четвертую (т.е. выполнять все этапы сначала для одной переменной, а затем для другой).

Тема 9 КРАЕВЫЕ ЗАДАЧИ ДЛЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Рассмотрим случай, когда дифференциальное уравнение и краевые условия линейны. Такая краевая задача называется линейной краевой задачей. Линейное дифференциальное уравнение можно записать в виде

$$L[y] = f(x), \quad (1)$$

где $L[y]$ - линейный дифференциальный оператор:

$$L[y] = p_0(x)y^n + p_1(x)y^{n-1} + \dots + p_n(x)y,$$

причем обычно предполагается, что $p_i(x)$, $i = \overline{0, n}$ и $f(x)$ известные непрерывные функции на отрезке $[a, b]$. Для простоты будем предполагать, что в краевые условия входят две абсциссы $x_1 = a$, $x_2 = b$ ($a < b$) концы отрезка $[a, b]$. Такие условия называют двухточечными.

Краевые условия

$$R_v[y] = j_v, \quad (v = \overline{1, n}) \quad (2)$$

называют линейными, если они имеют вид

$$R_n[y] = \sum_{k=0}^{n-1} [\alpha_k^{(n)} y^{(k)}(a) + \beta_k^{(n)} y^{(k)}(b)],$$

$\alpha_k^{(v)}$, $\beta_k^{(v)}$, j - заданные постоянные и выполняются условия

$$\sum_{k=0}^{n-1} (|\alpha_k^{(v)}| + |\beta_k^{(v)}|) \neq 0, \quad v = \overline{1, n}.$$

Линейная краевая задача состоит в нахождении функции $y(x)$, удовлетворяющей дифференциальному уравнению (1) и краевым условиям (2), причем последние - линейно независимые.

Линейная краевая задача называется однородной, если:

$$1) f(x) \equiv 0 \text{ в (1);} \quad (1)$$

$$2) j_v = 0; \quad (2)$$

в противном случае задача неоднородная.

Метод конечных разностей. Рассмотрим линейное дифференциальное уравнение

$$y'' + p(x)y' + g(x)y = f(x) \quad (3)$$

с двухточечными краевыми условиями

$$\begin{aligned} & \alpha_0 y(a) + \alpha_1 y'(a) = A; \\ & \beta_0 y(b) + \beta_1 y'(b) = B; \\ & (|\alpha_0| + |\alpha_1| \neq 0; \quad |\beta_0| + |\beta_1| \neq 0), \end{aligned} \quad (4)$$

где $p(x)$, $g(x)$, $f(x)$ - непрерывны на отрезке $[a, b]$.

Одним из самых простых способов решения задачи (3), (4) является сведение ее к системе конечно-разностных уравнений. Для этого необходимо разбить отрезок $[a, b]$ на n равных частей длины h (шаг), где $h = (b - a)/n$. Точками разбиения абсциссы будут $x_i = x_0 + ih$, где $x_0 = a$, $x_n = b$.

Значение функции $y(x)$ в точках x_i и ее производных обозначим соответственно $y(x_i) = y_i$, $y'(x_i) = y'_i$, $y''(x_i) = y''_i$.

Введем следующие обозначения: $P_i = p(x_i)$, $g_i = g(x_i)$, $f_i = f(x_i)$.

Заменим производные конечно-разностными отношениями. Для внутренних точек отрезка $[a, b]$, будем иметь:

$$y'_i = (y_{(i+1)} - y_{(i-1)})/2h; \quad y''_i = (y_{(i+1)} - 2y_i + y_{(i-1)})/h, \quad i = 1, 2, \dots, n - 1. \quad (5)$$

Для крайних точек $x_0 = a, x_n = b$, чтобы не выходить за концы отрезка $[a, b]$, можно положить:

$$y'_0 = (y_1 - y_0)/h; \quad y'_n = (y_{n-1} - y_n)/(-h). \quad (6)$$

Однако, если функция $y(x)$ достаточно гладкая, то более точные значения дают формулы

$$y'_0 = (-y_2 + 4y_1 - 3y_0)/2h; \quad y'_n = (3y_n - 4y_{n-1} + y_{n-2})/2h; \quad n \geq 2. \quad (7)$$

Используя (5) и (6), можно представить краевую задачу (3), (4) в виде:

$$(y_{i+1} - 2y_i + y_{i-1})/h + p_i(y_{i+1} - y_{i-1})/2h + g_i y_i = f_i, \quad i = 1, \dots, n - 1;$$

$$a_0 y_0 + a_1 (y_1 - y_0)/h = A;$$

$$b_0 y_n + b_1 (y_n - y_{n-1})/h = B. \quad (8)$$

Таким образом, получили систему $n + 1$ линейных уравнений с $n + 1$ неизвестными y_0, y_1, \dots, y_n , решение которой позволяет найти значение искомой функции $y(x)$ в точках x_0, x_1, \dots, x_n .

СИСТЕМА (8) ЯВЛЯЕТСЯ ТРЕХДИАГОНАЛЬНОЙ И ДЛЯ ЕЕ РЕШЕНИЯ ИСПОЛЬЗУЕТСЯ МЕТОД ПРОГОНКИ, ПРИВЕДЕННЫЙ В ЛЕКЦИИ 2.

ПРИ ИСПОЛЬЗОВАНИИ ПОДСТАНОВОК (7) ПОЛУЧАЕМ СИСТЕМУ ЛИНЕЙНЫХ УРАВНЕНИЙ:

$$(y_{i+1} - 2y_i + y_{i-1})/h + p_i(y_{i+1} - y_{i-1})/2h + g_i y_i = f_i, \quad i = 1, \dots, n - 1;$$

$$a_0 y_0 + a_1 (-y_2 + y_1 - 3y_0)/(2h) = A;$$

$$b_0 y_n + b_1 (3y_n - 4y_{n-1} + y_{n-2})/(2h) = B.$$

ДЛЯ РЕШЕНИЯ ТАКОЙ СИСТЕМЫ МОЖНО ИСПОЛЬЗОВАТЬ КАКОЙ-ЛИБО ИТЕРАЦИОННЫЙ МЕТОД.

Метод Галеркина. Основан на одной теореме из теории общих рядов Фурье.

Теорема. Пусть $\{U_m(x)\}$ полная система функций ортогональных на интервале $[a, b]$ с ненулевой нормой. Если непрерывная функция $f(x)$ ортогональна на $[a, b]$ ко всем функциям $U_n(x)$, то есть

$$\int_a^b f(x) U_n(x) dx = 0 \quad \text{при } n = 0, 1, 2, \dots, m,$$

то $f(x) \equiv 0$ при $x \in [a, b]$.

Пусть имеем двухточечную линейную краевую задачу

$$L[y] = f(x), \quad (9)$$

где $L[y] = y'' + p(x)y' + g(x)y$ с линейными краевыми условиями

$$\Gamma_a[y] = \alpha_0 y(a) + \alpha_1 y'(a) = A; \quad (10)$$

$$\Gamma_b[y] = \beta_0 y(b) + \beta_1 y'(b) = B;$$

$$(|\alpha_0| + |\alpha_1| = 0, \quad |\beta_0| + |\beta_1| = 0).$$

Выберем конечную систему базисных функций $\{U_i(x)\}$ $i = 0, n$, составляющих часть некоторой полной системы, причем потребуем, чтобы $U_0(x)$ удовлетворяла неоднородным краевым условиям

$$\Gamma_a[U_0] = A, \quad \Gamma_b[U_0] = B,$$

а остальные функции $U_i(x)$ ($i = 1, n$) удовлетворяли однородным краевым условиям

$$\Gamma_a[U_i] = \Gamma_b[U_i] = 0, \quad i = 1, n.$$

Решение краевой задачи (9) будем искать в виде

$$y = U_0(x) + \sum_{i=1}^n C_i U_i(x). \quad (11)$$

При таком выборе базисных функций $U_i(x)$ функция $y(x)$, определяемая (11), очевидно удовлетворяет краевым условиям при любом выборе C_i .

Если подставить в левую часть дифференциального уравнения (9) выражение (10) и вычесть правую часть, то получим функцию невязки:

$$R(x, C_1, C_2, \dots, C_n) = L[U_0] + \sum_{i=1}^n C_i L[U_i] - f(x). \quad (12)$$

Эта функция характеризует отклонение формы (11) от действительного решения краевой задачи (9), (10). Если форма (11) при некоторых значениях коэффициентов C_i совпадает с решением краевой задачи, то невязка будет равна нулю $R(x) \equiv 0$ при $x \in [a, b]$, поэтому для получения приближенного решения, близкого к точному, нужно подобрать C_i так, чтобы функция невязки была в каком-то смысле мала.

Согласно приведенной выше теореме для того, чтобы функция невязки была равна нулю, необходимо, чтобы невязка была ортогональна ко всем базисным функциям:

$$\int_a^b U_i(x) R(x, C_1, C_2, \dots, C_n) dx = 0, \quad i = 1, 2, \dots, n.$$

При достаточно большом n это обеспечивает малость невязки в среднем, насколько приближенное решение будет при этом близко к точному вопрос остается открытым.

Используя выражение для невязки (12), после преобразований получим систему линейных уравнений вида

$$\sum_{k=1}^n C_k \int_a^b U_i(x) L[U_k] dx = \int_a^b U_i(x) (f(x) - L[U_0]) dx, \quad i = 1, 2, \dots, n,$$

решив которую можно найти значения коэффициентов C_i .

Согласно условиям теоремы используемая система функций U_i должна быть полной и ортогональной, но на практике подобрать такую систему не всегда представляется возможным, поэтому при использовании этого метода требуют, чтобы система функций была линейно-независимой на $[a, b]$.

Метод коллокаций. В этом методе выбирают n точек принадлежащих рассматриваемому участку $x_i \in [a, b]$, $i = 1, 2, \dots, n$, называемых точками коллокации, и требуют, чтобы невязка в этих точках была равна нулю,

$$R(x_i, C_1, C_2, \dots, C_n) = 0, \quad i = 1, 2, \dots, n.$$

После преобразований получим систему линейных уравнений

$$\sum_{k=1}^n C_k L[U_k(x_i)] dx = f(x_i) - L[U_0(x_i)], \quad i = 1, 2, \dots, n.$$

Следует иметь в виду, что получаемая система будет хорошо обусловленной, если точки коллокации являются корнями ортогональных многочленов (например многочленов Чебышева); при равноотстоящих точках получаемая система, особенно при большой размерности, может быть плохо обусловленной.

Метод экспоненциальной прогонки. Во многих приложениях, особенно в механике жидкости, при моделировании электрических цепей, химических реакций и в теории управления возникают задачи с начальными данными, которым присуще свойство, называемое "жесткостью". Жесткие уравнения - это такие уравнения, решения которых имеют одновременно как быстрое, так и медленно изменяющиеся компоненты. Первые из них часто не представляют интереса, однако если используются классические методы интегрирования, то эти быстро затухающие компоненты могут привести к возникновению трудных вычислительных задач. Следовательно, важно создать подходящие численные методы для решения таких задач. С этой целью был предложен ряд схем.

Некоторые методы предназначены для систем средней жесткости и требуют использования очень малых шагов в области быстрого изменения решения. Фактически это означает, что эффективность этих способов уменьшается с увеличением жесткости.

Для того чтобы показать, как пограничный слой может возникнуть в случае простых задач, мы приводим пример линейной двухточечной краевой задачи с таким поведением:

$$\begin{aligned} \varepsilon u''(x) + (1+x^2)u'(x) - \left(x - \frac{1}{2}\right)^2 u(x) &= -(x^2 - \exp(x)), \quad 0 < x < 1, \\ u(0) &= -1, \quad u(1) = 0, \end{aligned}$$

Приведенные ниже восемь графиков (рис. 1) иллюстрируют возникновение и развитие пограничного слоя по мере изменения коэффициента при второй производной от средних до малых значений, точные решения для восьми значений параметра ε .

В этой части мы рассмотрим двухточечные краевые задачи, являющиеся математическими моделями диффузионно-конвективных процессов или родственных физических явлений. Диффузионным членом является член, включающий производные второго порядка, а конвективный член включает производные первого порядка.

Во многих практических задачах коэффициент при производной первого порядка мал. Примерами могут служить задачи о переносе тепла с большими числами Пекле, о течениях

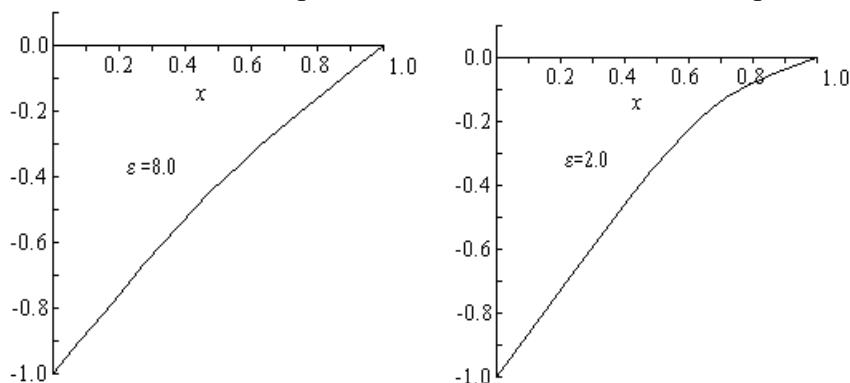
Навье-Стокса с большими числами Рейнольдса и задачи магнитной гидродинамики с большими числами Хартмана.

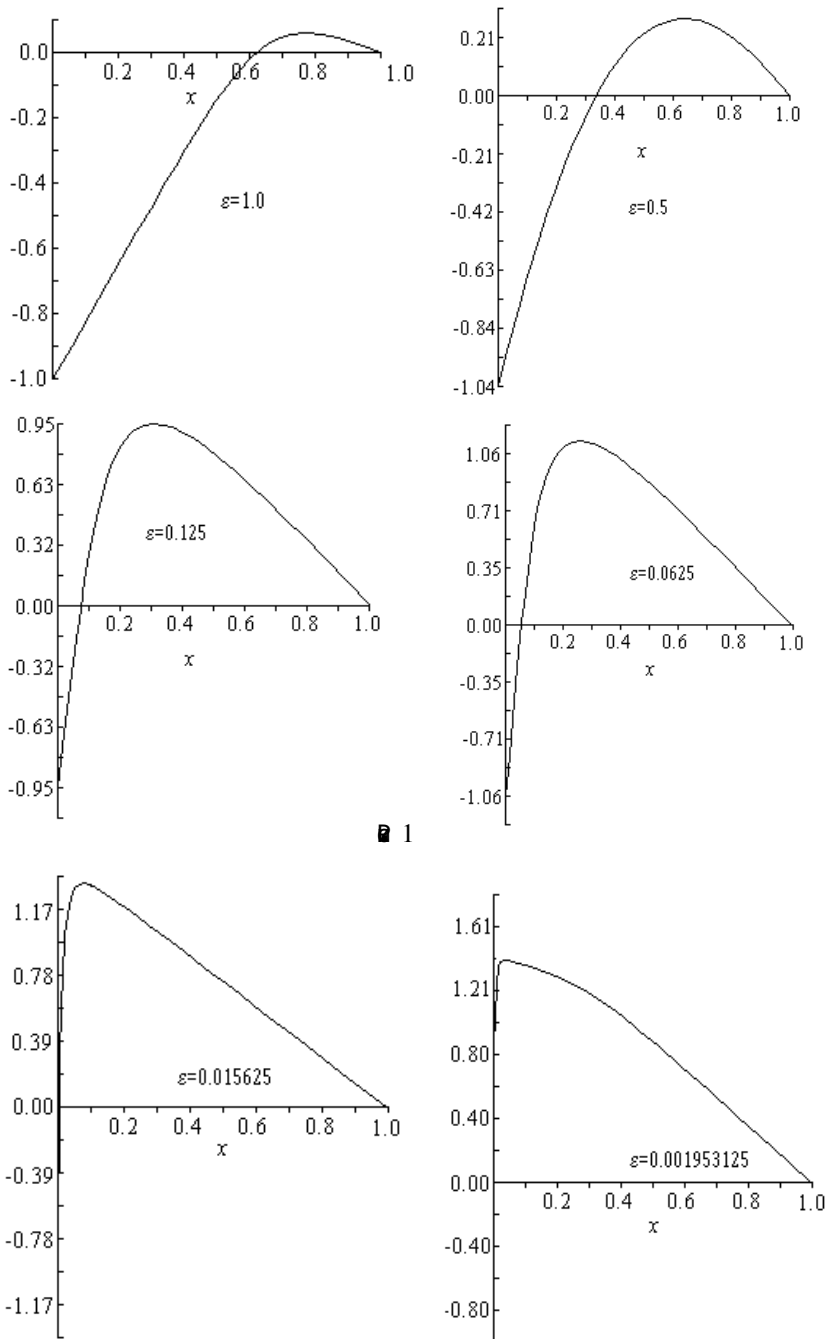
Решения указанного выше типа задач обычно не могут быть представлены в виде рядов по возрастающим степеням коэффициента диффузии. Таким образом, эти задачи не являются регулярными возмущениями соответствующих задач более низкого порядка. Они называются сингулярно возмущенными задачами. Так как в предельном случае при нулевом коэффициенте диффузии дифференциальные уравнения имеют более низкий порядок, некоторые из граничных условий станут излишними. Решение задачи может быстро изменяться вблизи граничных точек, что соответствует наличию пограничного слоя. Описание решений в таком пограничном слое требует введения другой независимой переменной (см. например, Смит [16]).

При решении указанных задач стандартными численными методами возникают большие трудности, причина которых заключается в неустойчивости численного процесса. Чтобы избавиться от нее, используем модифицированную дискретизацию диффузионного члена. Если нужно, то результирующие схемы могут быть представлены в консервативной форме.

Далее исследуются двухточечные краевые задачи для обыкновенных дифференциальных уравнений второго порядка, у которых коэффициент при второй производной является малым параметром. Если задача имеет член с первой производной, возможен один пограничный слой. С другой стороны, если нет члена с первой производной, то пограничные слои могут быть в окрестности обеих граничных точек. Также кратко рассматриваются задачи с малыми коэффициентами при первой и второй производных и задачи с точками возврата. С аналитической теорией задач этого типа можно подробно ознакомиться в работах Бендера и Орсага [17] и О'Малли [18].

Нельзя рассчитывать на то, что классические методы будут работать одинаково хорошо во всей области изменения величин h и ε . Их использование аналогично аппроксимации экспоненциальной функции отрезком ряда Тейлора или рациональной аппроксимации Паде. Эти классические методы также требуют строгих ограничений на величину шага для обеспечения устойчивости при малых ε , за исключением специальных случаев. Например, хорошо известный метод направленных разностей приводит к схемам, не требующим наложения ограничений на величину шага для обеспечения устойчивости. Ошибка схемы с направленной разностью всегда содержит член порядка $O(\varepsilon)$. Это приемлемо, если ε достаточно мало по сравнению с h , но может быть неприемлемым в других случаях.





Продолжение рис. 1

Рассмотрим задачу вида

$$Lu(x) = \varepsilon u''(x) + a(x)u'(x) - b(x)u(x) = f(x), \quad 0 \leq x \leq 1, \quad (13)$$

$$u(0) = A, \quad u(1) = B,$$

$$a(x) \geq \alpha > 0, \quad b(x) \geq 0, \quad \forall x \in [0, 1],$$

где ε - малый параметр.

Случай, когда $a(x) \leq \alpha < 0$ приводится к (13) заменой независимой переменной x на $1 - x$.

В случае $b \equiv 0$ схема решения задачи имеет следующий вид:

$$\frac{\varepsilon}{h^2}(\gamma_i^- u_{i-1} + \gamma_i^c u_i + \gamma_i^+ u_{i+1}) = q_i^- f(x_{i-1}) + q_i^c f(x_i) + q_i^+ f(x_{i+1});$$

$$1 \leq i \leq N-1; \quad (14)$$

$$u_0 = A, \quad u_N = B;$$

$$\text{- слева} \quad \gamma_i^- = \frac{p_i^- e^{-p_i^-}}{1 - e^{-p_i^-}};$$

$$\text{- справа} \quad r_i^+ = \frac{p_i^+}{1 - e^{-p_i^+}};$$

$$\gamma_i^0 = -(\gamma_i^- + \gamma_i^+);$$

$$q_i^+ = \frac{\gamma_i^+ - 1}{2p_i^+}; \quad q_i^c = q_i^- + q_i^+; \quad q_i^- = \frac{1 - \gamma_i^-}{2p_i^-},$$

$$\text{где} \quad p_i^- = \frac{h(a(x_i) + a(x_{i-1}))}{2\varepsilon}; \quad p_i^+ = \frac{h(a(x_{i+1}) + a(x_i))}{2\varepsilon}; \quad p = h/\varepsilon.$$

УРАВНЕНИЯ (14) СВЯЗЫВАЮТ ТРИ СОСЕДНИЕ ТОЧКИ ИСКОМОГО РЕШЕНИЯ И РЕШАТЬ ЭТУ СИСТЕМУ МОЖНО МЕТОДОМ ПРОГОНКИ.

В случае $b \neq 0$ схема решения задачи имеет вид

$$\frac{\varepsilon}{h^2}(\gamma_i^- u_{i-1} + \gamma_i^c u_i + \gamma_i^+ u_{i+1}) - q_i^- b(x_{i-1})u_{i-1} - q_i^c b(x_i)u_i - q_i^+ b(x_{i+1})u_{i+1} =$$

$$= q_i^- f(x_{i-1}) + q_i^c f(x_i) + q_i^+ f(x_{i+1}),$$

$$1 \leq i \leq N-1,$$

где $r_i^-, r_i^c, r_i^+, q_i^-, q_i^c, q_i^+$ – выражаются зависимостями (14).

САМОСОПРЯЖЕННЫЕ ЗАДАЧИ В КОНСЕРВАТИВНОЙ ФОРМЕ. В ПРИЛОЖЕНИЯХ ЧАСТО ВСТРЕЧАЮТСЯ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ, ЗАПИСАННЫЕ В КОНСЕРВАТИВНОЙ ФОРМЕ. СКОНСТРУИРУЕМ СХЕМЫ ЭКСПОНЕНЦИАЛЬНОЙ ПОДГОНКИ ДЛЯ РЕШЕНИЯ ТАКИХ ЗАДАЧ И ПОКАЖЕМ, ЧТО РЕШЕНИЯ ЭТИХ СХЕМ СХОДЯТСЯ РАВНОМЕРНО С ПЕРВЫМ ПОРЯДКОМ К ТОЧНЫМ РЕШЕНИЯМ.

Рассмотрим несамосопряженную задачу в консервативном виде

$$Lu(x) \equiv -\varepsilon(p(x)u'(x))' + q(x)u(x) = f(x), \quad 0 < x \leq 1, \quad (15)$$

$$u(0) = A, \quad u(1) = B,$$

где $p(0), g(0), r(0)$ - гладкие и удовлетворяют условиям

$$p(x) \geq \alpha > 0; \quad q(x) \geq \beta > 0; \quad p'(x) \geq 0.$$

Используем конечно-разностную схему экспоненциальной подгонки для задачи (15), получим

$$L^k u_i = -\varepsilon \sigma_i \delta(p(x_i) \delta u(x)) + q(x_i) u(x_i) = f(x_i), \quad 1 \leq i \leq N-1,$$

$$u_0 = A, \quad u_N = B,$$

где $\sigma_i = \frac{1}{4}\gamma_i p^2 s h^{-2} \frac{1}{2} p \sqrt{\gamma_i}; \quad \gamma_i = q(x_i)/p(x_i); \quad p = h/\sqrt{\varepsilon}.$

ИСПОЛЬЗУЯ ТЕХНИКУ, ОПИСАННУЮ ВЫШЕ, МОЖНО ТАКЖЕ ПОКАЗАТЬ, ЧТО ДРУГИЕ СИСТЕМЫ ДАЮТ РЕШЕНИЯ, СХОДЯЩИЕСЯ РАВНОМЕРНО С ПЕРВЫМ ПОРЯДКОМ.

Пример 1.

$$\varepsilon p(x)u''(x) + q(x)u'(x) + \gamma(x)u(x) = f(x);$$

$$u(0) = A, \quad u(1) = B,$$

где функции p , g и r гладкие и удовлетворяют условиям

$$p(x) \geq \alpha > 0, \quad q(x) \geq \beta > 0, \quad \gamma(x) \leq 0.$$

Для этой задачи разностный аналог имеет вид

$$\varepsilon \sigma_i p(x_i) \delta^2 u_i + q(x_i) \delta u_i + \gamma(x_i) u_i = f(x_i), \quad 1 \leq i \leq N-1;$$

$$u_0 = A, \quad u_N = B, \quad \sigma_i = \frac{pq(x_i)}{2p(x_i)} \operatorname{cth} \frac{pq(x_i)}{2p(x_i)}.$$

Решение этой задачи сходится равномерно по ε с первым порядком к решению непрерывной задачи.

Пример 2.

$$\varepsilon(p(x)u'(x))' + q(x)u'(x) + \gamma(x)u(x) = f(x), \quad \varepsilon \in (0,1);$$

$$u(0) = A, \quad u(1) = B,$$

$$p(x) \geq \alpha > 0, \quad p'(x) \geq 0, \quad q(x) \geq \beta > 0, \quad \gamma(x) \leq 0.$$

Разностный аналог имеет вид

$$\varepsilon \sigma_i \delta(p(x_i) \delta u_i) + q(x_i) \delta u_i + \gamma(x_i) u_i = f(x_i), \quad 1 \leq i \leq N-1;$$

$$u_0 = A, \quad u_N = B; \quad \sigma_i = \frac{pq(x_i)}{2p\left(x_i - \frac{1}{2}h\right)} \operatorname{cth} \frac{pq(x_i)}{2p\left(x_i - \frac{1}{2}h\right)}.$$

Пример 3.

$$\varepsilon(p(x)u'(x))' + (q(x)u(x))' + \gamma(x)u(x) = f(x), \quad \varepsilon \in (0,1);$$

$$u(0) = A, \quad u(1) = B;$$

$$p(x) \geq \alpha > 0, \quad p'(x) \geq 0, \quad q(x) \geq \beta > 0, \quad q'(x) \geq 0, \quad \gamma(x) \leq 0.$$

Разностный аналог имеет вид

$$\varepsilon \delta(\sigma_i p(x_i) \delta u_i) + \delta(q(x_i) u_i) + \gamma(x_i) u_i = f(x_i), \quad 1 \leq i \leq N-1;$$

$$u_0 = A, \quad u_N = B, \quad \sigma_i = \frac{pq(x_i)}{2p(x_i)} \operatorname{cth} \frac{pq(x_i)}{2p(x_i)}.$$

Задачи со смешанными граничными условиями возникают во многих практических случаях. В качестве примера можно назвать задачи, связанные с гидродинамикой, электролизом, моделированием полупроводниковых приборов и химическими реакциями.

Рассмотрим задачу вида

$$Lu(x) \equiv \varepsilon u''(x) + a(x)u'(x) - b(x)u(x) = f(x); \quad (16)$$

$$B_0 u(0) \equiv \alpha u(0) - \beta u'(0) = A;$$

$$B_1 u(1) \equiv \gamma u(1) + \delta u'(1) = B,$$

а a, b, f - гладкие функции, причем a и b ограничены;

$$A^* \geq a(x) \geq a^* > 0, \quad B^* \geq b(x) \geq b^* \geq 0;$$

$$\alpha, \beta, \gamma, \delta \geq 0, \quad \alpha + \beta > 0, \quad \gamma + \delta > 0, \quad \gamma + b^* > 0.$$

Численный аналог задачи (16) имеет вид

$$L^k u_i \equiv \varepsilon \sigma_i D_+ D_- u_i + a(x_i) D_0 u_i - b(x_i) u_i = f(x_i), \quad 1 \leq i \leq N-1;$$

$$B_0^k u_0 \equiv \alpha u_0 - \beta(u_N - u_{N-1})/h = A;$$

$$B_1^k u_N \equiv \gamma u_N + \delta(u_N - u_{N-1})/h = B,$$

где

$$\sigma_i = \frac{1}{2} p a(x_i) \operatorname{cth} \frac{1}{2} p a(x_i);$$

$$D_+ u_i = (u_{i+1} - u_i)/2h;$$

$$D_- u_i = (u_i - u_{i-1})/2h;$$

$$D_0 u_i = (D_+ + D_-) u_i / 2 = \frac{1}{2} D_+ u_i + \frac{1}{2} D_- u_i.$$

СХЕМЫ ВЫСОКОГО ПОРЯДКА. РАНЕЕ БЫЛИ ПОКАЗАНЫ РАЗНОСТНЫЕ СХЕМЫ, СХОДЯЩИЕСЯ РАВНОМЕРНО С ПЕРВЫМ ПОРЯДКОМ К РЕШЕНИЯМ НЕПРЕРЫВНЫХ ЗАДАЧ. РАССМОТРИМ КОНЕЧНО-РАЗНОСТНЫЕ СХЕМЫ ЭКСПОНЕНЦИАЛЬНОЙ ПОДГОНКИ, ДАЮЩЕЕ БОЛЕЕ ТОЧНЫЕ АППРОКСИМАЦИИ НЕПРЕРЫВНЫХ ЗАДАЧ. НАЧНЕМ С ЗАДАЧИ

$$Lu(x) \equiv \varepsilon u''(x) + au'(x) - bu(x) = f(x), \quad 0 \leq x \leq 1; \quad (17)$$

$$u(0) = A, \quad u(1) = B,$$

где f - гладкая функция и a, b - постоянные, удовлетворяющие условиям

$$b \geq 0 \text{ и } |a| + b \neq 0.$$

Разностный аналог задачи (17)

$$L^k u_i \equiv \varepsilon \sigma_i D_+ D_- u_i + a \sigma_2 D_0 u_i - b u_i = Q_i(f), \quad 1 \leq i \leq N-1;$$

$$u_0 = A, \quad u_N = B,$$

где $\sigma_1 = -\frac{h^2 b}{4\varepsilon} \left(1 + \operatorname{cth} \frac{h\lambda_1}{2} + \operatorname{cth} \frac{h\lambda_2}{2} \right); \quad \sigma_2 = \frac{hb}{2a} \left(\operatorname{cth} \frac{h\lambda_1}{2} + \operatorname{cth} \frac{h\lambda_2}{2} \right); \quad \lambda_1, \lambda_2 \text{ — корни}$

характеристического уравнения $\varepsilon \lambda^2 + a\lambda - b = 0; \quad Q_i(f)$ - линейная комбинация значений функций f в x_i ,

$$Q_i(f) = \alpha_i f(x_{i-1}) + \beta_i f(x_i) + \gamma_i f(x_{i+1}), \quad (18)$$

где

$$\alpha_i = \frac{\tilde{a}_i}{a(x_{i-1})} \left(\frac{1}{2p\tilde{a}_{i-1}} - \frac{\tilde{\sigma}_i - 1}{p^2 \tilde{a}_{i-1} \tilde{a}_i} \right);$$

$$\gamma_i = \frac{\tilde{a}_i}{a(x_{i+1})} \left(\frac{\tilde{\sigma}_i - 1}{p\tilde{a}_i} + \frac{\alpha_i a(x_{i+1}) \tilde{a}_{i-1}}{(\tilde{a}_i)^2} \right);$$

$$\beta_i = \frac{\tilde{a}_i}{a(x_i)} \left(1 - \gamma_i \frac{a(x_{i+1})}{p\tilde{a}_i} - \alpha_i \frac{a(x_{i-1})}{p\tilde{a}_i} \right);$$

$$\tilde{a}_i = \frac{1}{12} [5a(x_{i+1}) + 8a(x_i) - a(x_{i-1})],$$

В некоторых случаях зависимость вида (18) может несколько упрощаться. В частности, при $b = 0$ получим

$$Q_i(f) = \alpha_1 f(x_{i-1}) + \alpha_2 f(x_i) + \alpha_3 f(x_{i+1}),$$

где $\alpha_1 = \frac{1}{6} - \frac{\varepsilon\sigma - \varepsilon}{2ah} + \frac{\varepsilon(\varepsilon - \varepsilon\sigma)}{a^2 h^2}; \quad \alpha_2 = \frac{2}{3} - \frac{2\varepsilon(\varepsilon - \varepsilon\sigma)}{a^2 h^2};$

$$\alpha_3 = \frac{1}{6} + \frac{\varepsilon\sigma - \varepsilon}{2ah} + \frac{\varepsilon(\varepsilon - \varepsilon\sigma)}{a^2 h^2}; \quad \sigma = \frac{ha}{2\varepsilon} \operatorname{cth} \frac{ha}{2\varepsilon}.$$

В случае $a = 0$ зависимость (18) можно записать в виде

$$Q_i(f) = \alpha_1 f(x_{i-1}) + \alpha_2 f(x_i) + \alpha_3 f(x_{i+1}),$$

где $\alpha_1 = \alpha_3 = \frac{\varepsilon\sigma - \varepsilon}{ah^2}; \quad \alpha_2 = 1 - \frac{2\varepsilon(\varepsilon - \varepsilon\sigma)}{ah^2}; \quad \sigma = \frac{h^2 b}{4\varepsilon} \operatorname{sh}^{-2} \frac{h\sqrt{b}}{2\sqrt{\varepsilon}}.$

ИМЕЕМ ЧЕТВЕРТЫЙ ПОРЯДОК РАВНОМЕРНОЙ СХОДИМОСТИ.

КВАЗИЛИНЕАРИЗАЦИЯ (МЕТОД НЬЮТОНА-КАНТОРОВИЧА). РАССМОТРИМ КРАЕВУЮ ЗАДАЧУ ВИДА

$$u'' = f(u); \quad u(0) = u(b) = 0. \quad (19)$$

ПУСТЬ $u_0(x)$ - НЕКОТОРОЕ НАЧАЛЬНОЕ ПРИБЛИЖЕНИЕ, А $\{u_n\}$ - ПОСЛЕДОВАТЕЛЬНОСТЬ, ОПРЕДЕЛЯЕМАЯ РЕКУРЕНТНЫМ СООТНОШЕНИЕМ

$$\begin{aligned} u_n'' &= f(u_{n-1}) + (u_n - u_{n-1})f'(u_{n-1}); \\ u_n(0) &= u_n(b) = 0. \end{aligned} \quad (20)$$

КАЖДАЯ ФУНКЦИЯ $u_n(x)$ ЕСТЬ РЕШЕНИЯ ЛИНЕЙНОГО УРАВНЕНИЯ, ЧТО ЯВЛЯЕТСЯ ВЕСЬМА ВАЖНОЙ ХАРАКТЕРНОЙ ОСОБЕННОСТЬЮ ЭТОГО АЛГОРИТМА. АЛГОРИТМ СЛЕДУЕТ ИЗ МЕТОДА АППРОКСИМАЦИИ НЬЮТОНА-РАФСОНА-КАНТОРОВИЧА В ФУНКЦИОНАЛЬНОМ ПРОСТРАНСТВЕ И ОТ НЕГО МОЖНО ОЖИДАТЬ КВАДРАТИЧНУЮ СХОДИМОСТЬ, ЕСЛИ СХОДИМОСТЬ ВООБЩЕ ИМЕЕТ МЕСТО.

ПОСКОЛЬКУ КАЖДОЕ РЕШЕНИЕ ОПРЕДЕЛЯЕТСЯ ДВУХТОЧЕЧНЫМИ КРАЕВЫМИ УСЛОВИЯМИ, ЗАРАНЕЕ ОТНЮДЬ НЕ ОЧЕВИДНО, ЧТО ПОСЛЕДОВАТЕЛЬНОСТЬ $\{u_n\}$, ЗАДАВАЕМАЯ (20), ДЕЙСТВИТЕЛЬНО СУЩЕСТВУЕТ. ПОЭТОМУ ВНАЧАЛЕ ПОКАЖЕМ, ЧТО УПОМЯНУТАЯ ПОСЛЕДОВАТЕЛЬНОСТЬ ПОЛНОСТЬЮ ОПРЕДЕЛЕНА, ЕСЛИ ОБЛАСТЬ ИЗМЕНЕНИЯ ОГРАНИЧЕНА МАЛЫМ ОТРЕЗКОМ $[A, B]$, А ЗАТЕМ ДОКАЖЕМ КВАДРАТИЧНУЮ СХОДИМОСТЬ ПОСЛЕДОВАТЕЛЬНОСТИ.

ПРЕЖДЕ ВСЕГО ДОКАЖЕМ СУЩЕСТВОВАНИЕ И РАВНОМЕРНУЮ ОГРАНИЧЕННОСТЬ ПОСЛЕДОВАТЕЛЬНОСТИ $\{u_n(x)\}$ ПРИ ДОСТАТОЧНО МАЛОМ ЗНАЧЕНИИ B . ИЗМЕНИВ ИНДЕКСАЦИЮ В (20), ЗАПИШЕМ

$$\begin{aligned} u_{n+1}'' &= f(u_n) + (u_{n+1} - u_n)f'(u_n); \\ u_{n+1}(0) &= u_{n+1}(b) = 0 \end{aligned} \quad (21)$$

и получим линейное интегральное уравнение

$$u_{n+1} = \int_0^b K(x, y) [f(u_n) + (u_{n+1} - u_n)f'(u_n)] dy,$$

где $K(x, y)$ - функция Грина:

$$K(x, y) = \begin{cases} \frac{x(y-b)}{b}, & 0 \leq x < y \leq b; \\ \frac{(x-b)y}{b}, & b \geq x > y \geq 0. \end{cases}$$

ДЛЯ ТОГО, ЧТОБЫ ПОЛУЧИТЬ (22), ПРЕДСТАВИМ (21) В ВИДЕ УРАВНЕНИЯ

$$u_n = \gamma(t), \quad u(0) = u(b) = 0$$

и затем используем результат

$$u = \int_0^b K(t, t_1, b) \gamma(t_1) dt_1;$$

$$K(t, t_1, b) = \begin{cases} \frac{t(t_1-b)}{b}, & t \leq t_1 \leq b; \\ \frac{(t-b)t_1}{b}, & 0 \leq x \leq t. \end{cases}$$

Отметим, что $\max_{x,y} |K(x,y)| = \frac{b}{4}$, где максимизация проведена по области $0 \leq x, y \leq b$.

Положим $\max_{|u| \geq 1} (|f(u)|, |f'(u)|) = m$, предполагая $m < \infty$, выберем $u_0(x)$ так, чтобы $|u_0(x)| \leq 1, 0 \leq x \leq b$. Возвращаясь к (22), имеем

$$|u_{n+1}| \leq \int_0^b |K(x,y)| [|f(u_n)| + |u_n| |f'(u_n)| + |u_{n+1}| |f'(u_n)|] dy.$$

Отсюда, обозначая $m_1 = \max_{0 \leq x \leq b} |u_1(x)|$, получаем при $n = 0$

$$m_1 \leq \frac{b}{4} \int_0^b [2m + mm_1] dy \leq \frac{b^2 m}{2} + \frac{b^2 m}{4} m_1.$$

Следовательно, при условии $b^2 m / 4 < 1$ находим границу

$$m_1 \leq \frac{b^2 m / 2}{1 - b^2 m / 4}.$$

Эта верхняя граница m_1 будет превосходить 1, если выполнено условие $b^2 \leq 4/3m$, поэтому следует выбирать отрезок $[0, b]$ достаточно малым, чтобы выполнялось условие $b^2 \leq 4/3m$. Итак, доказано, что принятое индуктивное определение последовательности $\{u_n(x)\}$ имеет смысл.

Доказательство сходимости проведем стандартным путем; единственное отличие в том, что сходимость окажется квадратичной.

Обращаясь к рекуррентному соотношению (21), вычтем n -е уравнение из $(n-1)$ -го:

$$(u_{n+1} - u_n)'' = f(u_n) - (u_n - u_{n-1})f'(u_{n-1}) + f'(u_n)(u_{n+1} - u_n).$$

РАССМАТРИВАЯ (23) КАК УРАВНЕНИЕ ОТНОСИТЕЛЬНО $u_{n+1} - u_n$ И, ПРЕОБРАЗУЯ ЕГО В ИНТЕГРАЛЬНОЕ УРАВНЕНИЕ, КАК ЭТО УЖЕ ДЕЛАЛОСЬ, ПОЛУЧАЕМ

$$(u_{n+1} - u_n) = \int_0^b K(x, y) [(u_n) - f(u_{n-1}) - (u_n - u_{n-1})f'(u_{n-1}) + f'(u_n)(u_{n+1} - u_n)] dy.$$

Из теоремы о среднем следует

$$f(u_n) - f(u_{n-1}) - (u_n - u_{n-1})f'(u_{n-1}) = \frac{(u_n - u_{n-1})^2}{2} f'(\theta),$$

где θ лежит между u_{n-1} и u_n .

Тогда, полагая $k = \max_{|u| \leq 1} |f'(u)|$, имеем

$$|u_{n+1} - u_n| \leq \frac{b}{4} \int_0^b \left[\frac{k}{2} (u_n - u_{n-1})^2 + m |u_{n+1} - u_n| \right] dy.$$

Отсюда, проводя рассуждения, аналогичные тем, которые использовались при доказательстве предыдущего метода, получаем

$$u(x) = -\ln 2 + 2 \ln \left[c \sec \frac{c(x-1/2)}{2} \right], \quad (24)$$

где c - корень уравнения $\sqrt{2} = c \sec(c/4)$, лежащий между 0 и $\pi/2$, а именно $c = 1,3360557$.

Напомним, как получено решение уравнения (24). Умножая на u' обе части уравнения и интегрируя, получаем u' как функцию от u . Дальнейшее интегрирование приводит к искомому явному решению уравнения (24).

Используя методы, изложенные выше, получаем аппроксимирующую последовательность

$$u''_{n+1} = e^{u_n} + e^{u_n} (u_{n+1} - u_n), \quad u_{n+1}(0) = u_{n+1}(b) = 0.$$

Выбирая наиболее естественное начальное приближение $u_0(x) = 0$, вычислим функции $u_1(x)$ и $u_2(x)$, используя метод интегрирования Рунге-Кутты.

Таблица иллюстрирует скорость сходимости. Значащие цифры величины $u_n(x)$, совпадающие со знаками $u(x)$, выделены подчеркиванием. Приближение $u_n(x)$ к $u(x)$.

x	$u_0(x)$	$u_1(x)$	$u_2(x)$	$u(x)$
0,0	0,000000	<u>0,000000</u>	<u>0,000000</u>	0,000000
0,1	0,000000	<u>-0,041285</u>	<u>-0,041436</u>	-0,041436
0,2	0,000000	<u>-0,072974</u>	<u>-0,0732683</u>	-0,0732686
0,3	0,000000	<u>-0,095386</u>	<u>-0,095800</u>	-0,095800
0,4	0,000000	<u>-0,108743</u>	<u>-0,109238</u>	-0,109238
0,5	0,000000	<u>-0,113181</u>	<u>-0,113704</u>	-0,113704
0,6	0,000000	<u>-0,108743</u>	<u>-0,109238</u>	-0,109238
0,7	0,000000	<u>-0,095386</u>	<u>-0,095800</u>	-0,095800
0,8	0,000000	<u>-0,072974</u>	<u>-0,0732683</u>	-0,0732686

0,9	0,000000	<u>-0,041285</u>	<u>-0,041436</u>	-0,041436
1,0	0,000000	<u>0,000000</u>	<u>0,000000</u>	0,000000

Нелинейные дифференциальные уравнения второго порядка общего вида. Нетрудно применить тот же метод для исследования уравнения

$$u'' = f(u', u, x) \quad (25)$$

с нелинейными краевыми условиями вида

$$g_1(u(0), u'(0)) = 0; \quad g_2(u(b), u'(b)) = 0 \quad (26)$$

или даже более общего вида

$$g_1(u(0), u'(0), u(b), u'(b)) = 0;$$

$$g_2(u(0), u'(0), u(b), u'(b)) = 0.$$

Теперь применим квазилинеаризацию как к уравнению, так и к граничным условиям. Тогда в случае уравнения (25) и условий (26) получим последовательность $\{u_n(x)\}$ при помощи уравнения

$$u''_{n+1} = f(u'_n, u_n, x) + f_{u'}(u'_n, u_n, x)(u'_{n+1} - u'_n) + f_u(u'_n, u_n, x)(u_{n+1} - u_n)$$

с линеаризованными краевыми условиями

$$g_1(u_n(0), u'_n(0)) + g_{1u}(u_n(0), u'_n(0))(u_{n+1}(0) - u_n(0)) + \\ + g_{1u'}(u_n(0), u'_n(0))(u'_{n+1}(0) - u'_n(0)) = 0.$$

Аналогично соотношения выписываются для g_2 в точке $x = b$.

Легко доказать существование, единственность и квадратичную сходимость при достаточно малом b и очевидных ограничениях на величины $f_u, f_{u'}, g_{iu}, g_{iu'}$, $i = 1, 2$.

В качестве примера рассмотрим уравнение

$$-u'' = 1 + a^2(u')^2, \quad u(0) = u(b) = 0,$$

возникающее при изучении конечных деформаций упругой струны под действием поперечной нагрузки. Точное аналитическое решение при $a^2 = 0,49$ и $b = 1$ имеет вид

$$u(x) = \frac{1}{a^2} \ln \left[\frac{\cos a(x-1/2)}{\cos(a/2)} \right].$$

Используя рекуррентное соотношение

$$-u''_{n+1} = 1 - 0,49u_n'^2 + 2 \cdot 0,49u_n' u_{n+1}';$$

$$u_{n+1}(0) = u_{n+1}(b) = 0$$

с начальным приближением $u_0(x) = 0$, получим результаты, которые приведены в таблице. Приближение $u_n(x)$ к $u(x)$

x	$u_0(x)$	$u_1(x)$	$u_2(x)$	$u(x)$
0,0	0,000000	0,000000	0,000000	0,000000
0,1	0,000000	<u>0,045000</u>	<u>0,046570</u>	0,046571
0,2	0,000000	<u>0,080000</u>	<u>0,082302</u>	0,082304
0,3	0,000000	<u>0,105000</u>	<u>0,107571</u>	0,107573
0,4	0,000000	<u>0,120000</u>	<u>0,122632</u>	0,122635
0,5	0,000000	<u>0,125000</u>	<u>0,127636</u>	0,127639
0,6	0,000000	<u>0,120000</u>	<u>0,122632</u>	0,122635
0,7	0,000000	<u>0,105000</u>	<u>0,107571</u>	0,107573
0,8	0,000000	<u>0,080000</u>	<u>0,082302</u>	0,082304
0,9	0,000000	<u>0,045000</u>	<u>0,046570</u>	0,046571
1,0	0,000000	0,000000	0,000000	0,000000

Тема 10 РЕШЕНИЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ

При решении задач с дифференциальными уравнениями в частных производных используется метод сеток, который заключается в следующем. На область решения накладывается некоторая сетка и путем замены производных в дифференциальном уравнении и в краевых условиях конечными разностями различного вида, получают соотношения, связывающие значения искомой функции в узлах сетки. Полученная таким образом система уравнений решается каким-либо численным методом. Рассмотрим этот подход подробнее применительно к некоторым уравнениям математической физики.

Уравнение Лапласа (эллиптическое уравнение). Это уравнение описывает некоторое установившееся стационарное состояние в пространстве (x, y) в некоторой области G :

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = 0, \quad (x, y) \in G. \quad (1) \quad (1)$$

Если область G является замкнутой и на границе этой области Γ задается значение искомой функции

$$U(x, y) = \varphi(x, y), \quad (x, y) \in \Gamma, \quad (2) \quad (2)$$

то такая задача называется задачей Дирихле.

Для замены этой задачи разностной задачей наложим на область G сетку из прямых, параллельных осям координат с шагом $h > 0$. Точки пересечения этих прямых будут образовывать узлы $(x_i = ih; y_j = jh; i = \overline{1, n}; j = \overline{1, m})$. Узлы, которые вместе со своими четырьмя ближайшими соседями будут принадлежать области G , называются внутренними узлами. Граничные узлы - это совокупность точек, у которых хотя бы одна из ближайших точек лежит вне области G .

Внутренние узлы будут образовывать некоторую сеточную область G_c , а граничные - некоторую сеточную область Γ_c .

Для замены дифференциального уравнения заменим производные центральными конечными разностями:

$$\frac{\partial^2 U(x, y)}{\partial x^2} \approx \frac{U(x+h, y) - 2U(x, y) + U(x-h, y)}{h^2};$$

$$\frac{\partial^2 U(x, y)}{\partial y^2} \approx \frac{U(x, y+h) - 2U(x, y) + U(x, y-h)}{h^2}.$$

Тогда вместо уравнения (1) будем иметь

$$\frac{U(x+h, y) - 2U(x, y) + U(x-h, y)}{h^2} + \frac{U(x, y+h) - 2U(x, y) + U(x, y-h)}{h^2} = 0;$$

следовательно, $U(x, y)$ можно выразить в следующем виде

$$U(x, y) = \frac{1}{4}(U(x+h, y) + U(x-h, y) + U(x, y+h) + U(x, y-h)).$$

Однако, чтобы иметь возможность оценить точность такой замены, следует идти по несколько иному пути, используя для получения конечно-разностного уравнения формулу Тейлора:

$$U(x+h, y+s) = U(x, y) + \left(\frac{\partial}{\partial x}h + \frac{\partial}{\partial y}s\right)U(x, y) + \frac{1}{2}\left(\frac{\partial}{\partial x}h + \frac{\partial}{\partial y}s\right)^2 U(x, y) + \dots + \frac{1}{k!}\left(\frac{\partial}{\partial x}h + \frac{\partial}{\partial y}s\right)^k U(x, y) + \dots + \frac{\theta^k}{k!}\left(\frac{\partial}{\partial x}h + \frac{\partial}{\partial y}s\right)^k U(x+\theta h, y+\theta s),$$

$$0 \leq \theta \leq 1;$$

при этом можно воспользоваться разными схемами.

Первая схема. Рассмотрим точку A и ближайшие к ней соседние точки B, C, D, E с координатами $A = (x, y), B = (x-h, y), C = (x+h, y), D = (x, y+h), E = (x, y-h)$ (рис. 1, а).

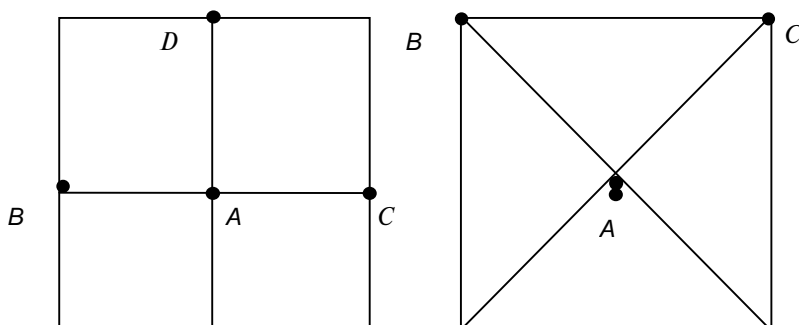




Рис. 1

Согласно формуле (4) при $k = 4, s = h$ получим

$$\begin{aligned}
 U(x-h, y) &= U(x, y) - hU_x + \frac{1}{2}h^2U_{xx} - \frac{1}{3!}h^3U_{xxx} + \frac{1}{4!}h^4\bar{U}_{xxxx}; \\
 U(x+h, y) &= U(x, y) + hU_x + \frac{1}{2}h^2U_{xx} + \frac{1}{3!}h^3U_{xxx} + \frac{1}{4!}h^4\tilde{U}_{xxxx}; \\
 U(x, y-h) &= U(x, y) - hU_y + \frac{1}{2}h^2U_{yy} - \frac{1}{3!}h^3U_{yyy} + \frac{1}{4!}h^4\hat{U}_{yyyy}; \\
 U(x, y+h) &= U(x, y) + hU_y + \frac{1}{2}h^2U_{yy} + \frac{1}{3!}h^3U_{yyy} + \frac{1}{4!}h^4\check{U}_{yyyy}.
 \end{aligned}$$

Здесь $U_x, U_{xx}, U_{xxx}, U_y, U_{yy}, U_{yyy}$ - значения производных функции $U(x, y)$ в точке (x, y) , а $\bar{U}_{xxxx}, \tilde{U}_{xxxx}, \hat{U}_{yyyy}, \check{U}_{yyyy}$ - значения производных в некоторых промежуточных точках.

Складывая эти значения, получаем

$$\begin{aligned}
 U(x-h, y) + U(x+h, y) + U(x, y-h) + U(x, y+h) &= \\
 = 4U(x, y) + h^2(U_{xx} + U_{yy}) + 12R(x, y) & \quad (5)
 \end{aligned}$$

где остаточный член R определяется равенством

$$R(x, y) = \frac{h^4}{4!}(\bar{U}_{xxxx} + \tilde{U}_{xxxx} + \hat{U}_{yyyy} + \check{U}_{yyyy}),$$

т.е. имеет четвертый порядок малости $R(x, y) = O(h^4)$.

Заметим, что выражение $(U_{xx} + U_{yy})$ является оператором Лапласа

$$\Delta U = \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = U_{xx} + U_{yy},$$

поэтому из (5) можно получить

$$\Delta U = O(h^4) + \frac{1}{h^2}(U(x-h, y) + U(x+h, y) + U(x, y+h) - 4U(x, y)),$$

которое называется первой основной конечно-разностной формой оператора Лапласа. Путем приравнивания этого выражения нулю и отбрасывания члена $O(h^4)$ можно получить соотношение (3).

Вторая схема. Рассмотрим дальние соседние точки около точки $A = (x, y)$ (рис. 1, б). Это будут точки $B = (x - h, y + h)$, $C = (x + h, y + h)$, $D = (x + h, y - h)$, $E = (x - h, y - h)$.

Как и в первой схеме воспользуемся выражением (3) для разложения функции в ряд Тейлора в точках B, C, D, E :

$$U(x+h, y-h) = U(x, y) + h(U_x - U_y) + \frac{h^2}{2!}(U_{xx} - 2U_{xy} + U_{yy}) + \frac{h^3}{3!}(U_{xxx} - 3U_{xxy} + 3U_{xyy} - U_{yyy}) + \frac{h^4}{4!}\left(\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right)^4 U(\xi_1, \zeta_1);$$

$$U(x-h, y-h) = U(x, y) + h(-U_x - U_y) + \frac{h^2}{2!}(U_{xx} + 2U_{xy} - U_{yy}) + \frac{h^3}{3!}(-U_{xxx} - 3U_{xxy} - 3U_{xyy} - U_{yyy}) + \frac{h^4}{4!}\left(-\frac{\partial}{\partial x} - \frac{\partial}{\partial y}\right)^4 U(\xi_2, \zeta_2);$$

$$U(x-h, y+h) = U(x, y) + h(-U_x + U_y) + \frac{h^2}{2!}(U_{xx} - 2U_{xy} - U_{yy}) + \frac{h^3}{3!}(-U_{xxx} + 3U_{xxy} - 3U_{xyy} + U_{yyy}) + \frac{h^4}{4!}\left(-\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)^4 U(\xi_3, \zeta_3);$$

$$U(x+h, y+h) = U(x, y) + h(U_x + U_y) + \frac{h^2}{2!}(U_{xx} + 2U_{xy} + U_{yy}) + \frac{h^3}{3!}(U_{xxx} + 3U_{xxy} + 3U_{xyy} + U_{yyy}) + \frac{h^4}{4!}\left(\frac{\partial}{\partial x} + \frac{\partial}{\partial y}\right)^4 U(\xi_4, \zeta_4).$$

Здесь $(\xi_1, \zeta_1), (\xi_2, \zeta_2), (\xi_3, \zeta_3), (\xi_4, \zeta_4)$ - некоторые промежуточные точки. Сложив эти выражения, будем иметь

$$U(x+h, y+h) + U(x-h, y+h) + U(x-h, y-h) + U(x+h, y-h) = 4U(x, y) + 2h^2 \Delta U + O(h^4),$$

откуда можно получить вторую конечно-разностную форму оператора Лапласа:

$$\Delta U = \frac{1}{2h^2} [U(x+h, y+h) + U(x-h, y+h) + U(x+h, y-h) + U(x-h, y-h) - 4U(x, y)] + O(h^4).$$

Отбрасывая остаточный член $O(h^4)$, получаем уравнение вида

$$U(x, y) = \frac{1}{4}(U(x+h, y+h) + U(x+h, y-h) + U(x-h, y+h) + U(x-h, y-h)). \quad (6)$$

Таким образом, в результате применения двух различных схем установлено, что любая внутренняя точка может быть вычислена как среднееарифметическое своих дальних или ближних соседних точек.

Соотношение (3) справедливо для каждой точки области G , поэтому оно будет справедливо для любой внутренней точки $(x_i, y_j) \in G_c$. Введя обозначение $U(x_i, y_j) = U_{i,j}$ и используя (3), получим систему уравнений

$$U_{i,j} = \frac{1}{4}(U_{i+1,j} + U_{i-1,j} + U_{i,j+1} + U_{i,j-1}); \quad (x_i, y_j) \in G_c.$$

Дополним эту систему уравнениями для граничных точек

$$U_{i,j} = \varphi_{i,j}; \quad (x_i, y_j) \in \Gamma_c.$$

Полученная система линейных уравнений (7), (8) может быть решена любым из методов, приведенных в лекции 2, но поскольку эта система имеет достаточно большую размерность (при сетке 10×10 имеем 100 уравнений) и очень разрежена (в каждом из уравнений участвуют не более пяти неизвестных), для решения таких систем рекомендуется использовать итерационные методы, а алгоритм построить так, чтобы не выполнять действий над нулевыми элементами матрицы коэффициентов. Простейшим способом будет задание некоторого начального приближения для внутренних точек $U_{i,j}^{(0)}, (x_i, y_j) \in G_c$ и вычисление следующего приближения методом простых итераций:

$$U_{i,j}^{(k+1)} = \frac{1}{4}(U_{i+1,j}^{(k)} + U_{i-1,j}^{(k)} + U_{i,j+1}^{(k)} + U_{i,j-1}^{(k)})$$

При использовании итерационных методов необходимо задавать начальное приближение для значений всех внутренних узлов. От близости этого приближения к истинному решению задачи зависит время расчета. Для уменьшения этого времени используют подход, предложенный Либманом и заключающийся в следующем.

Строится сетка с достаточно большим шагом и решается система уравнений небольшой размерности. В результате находится некоторое грубое приближение к решению задачи в узлах сетки. Для получения более точного решения шаг сетки уменьшается вдвое и решается система вчетверо большей размерности. Для сокращения времени расчета в качестве начального приближения для системы с мелкой сеткой используются значения, полученные в результате решения системы с крупной сеткой, а в тех узлах мелкой сетки, которые не совпадают с узлами крупной сетки, используется аппроксимация по уравнению (3) или (6). Такой процесс деления шага сетки можно выполнить несколько раз, получая все более точное решение поставленной задачи.

Уравнение теплопроводности (параболическое уравнение). Типичная задача, описывающая теплоперенос вдоль одномерного стержня единичной длины $0 \leq x \leq 1$ при заданном начальном распределении температуры $\varphi(x)$ и заданных температурных режимах на концах стержня $\psi(t), \phi(t)$ ставится следующим образом:

$$\frac{\partial U}{\partial t} - a^2 \frac{\partial^2 U}{\partial x^2} = f(x,t), \quad t > 0; \quad 0 \leq x \leq 1;$$

$$U(0, x) = \varphi(x); \quad (1)$$

$$U(t, 0) = \psi(t); \quad U(t, 1) = \phi(t).$$

Эта модельная задача является смешанной, так как содержит начальные (при $t = 0$) и краевые условия (при $x = 0$ и $x = 1$). Она описывает различные диссипативные процессы (типа теплопроводности или диффузии) в твердых телах, газах, плазме, в магнитной гидродинамике, в биологии и т.д.

Областью решения задачи (1) будет являться полуполоса $t > 0$, $0 \leq x \leq 1$. Путем замены переменных $\tau = a^2 t$ дифференциальное уравнение приводится к виду

$$\frac{\partial U}{\partial \tau} - \frac{\partial^2 U}{\partial x^2} = f(x, \tau),$$

поэтому избавимся от множителя a^2 , считая его равным единице. Тогда задача запишется в виде

$$\frac{\partial U}{\partial t} - \frac{\partial^2 U}{\partial x^2} = f(x, t), \quad 0 \leq t \leq T, \quad 0 \leq x \leq 1;$$

$$U(0, x) = \varphi(x); \quad (2)$$

$$U(t, 0) = \psi(t); \quad U(t, 1) = \phi(t).$$

Наложим на область $0 \leq t \leq T$, $0 \leq x \leq 1$ сетку, образованную прямыми параллельными осям координат с шагом по x равным, и шагом по t равным, т.е. $x_i = ih$; $t_j = j\tau$, $i = \overline{0, n}$, $j = \overline{0, m}$ где n и m количество линий сетки. Введем обозначения $U_{i,j} = U(x_i, t_j)$, $f_{i,j} = f(x_i, t_j)$, $\varphi_i = \varphi(x_i)$, $\psi_j = \psi(t_j)$, $\phi_j = \phi(t_j)$ и заменим производную по длине центральным разностным отношением

$$\frac{\partial^2 U(x_i, t_j)}{\partial x^2} \approx \frac{U_{i-1,j} - 2U_{i,j} + U_{i+1,j}}{h^2},$$

а производную по времени - правым конечно-разностным отношением

$$\frac{\partial U(x_i, t_j)}{\partial t} \approx \frac{U_{i,j+1} - U_{i,j}}{\tau}.$$

Подставив эти выражения в дифференциальное уравнение, получим конечно-разностное уравнение, справедливое для всех внутренних узлов сетки:

$$\frac{U_{i,j+1} - U_{i,j}}{\tau} - \frac{U_{i-1,j} - 2U_{i,j} + U_{i+1,j}}{h^2} = f_{i,j}, \quad i = \overline{1, n-1}; \quad j = \overline{0, m-1}. \quad (3)$$

Начальные и краевые условия дадут уравнения

$$U_{i,0} = \varphi_i, \quad i = \overline{1, n-1}; \quad (4)$$

$$U_{0,j} = \psi_j; \quad U_{n,j} = \phi_j, \quad j = \overline{0, m}.$$

ТАКИМ ОБРАЗОМ, ЗАДАЧА (2) СВОДИТСЯ К СИСТЕМЕ ЛИНЕЙНЫХ УРАВНЕНИЙ (3), (4), РЕШИВ КОТОРУЮ, МОЖНО ПОЛУЧИТЬ ЗНАЧЕНИЯ ИСКОМОЙ ФУНКЦИИ В УЗЛАХ СЕТКИ.

КОНФИГУРАЦИЯ УЗЛОВ, ЗНАЧЕНИЯ СЕТОЧНОЙ ФУНКЦИИ В КОТОРЫХ ОПРЕДЕЛЯЮТ ВИД РАЗНОСТНЫХ УРАВНЕНИЙ, ЗАПИСЫВАЕМЫХ ДЛЯ ВНУТРЕННИХ ТОЧЕК СЕТКИ, НАЗЫВАЕТСЯ ШАБЛОНОМ РАЗНОСТНОЙ СХЕМЫ. ДЛЯ СХЕМЫ (3) ШАБЛОН ИМЕЕТ ВИД, ПОКАЗАННЫЙ НА РИС 2, А; ОН СВЯЗЫВАЕТ ОДНУ ТОЧКУ СЛОЯ (ВРЕМЕННОГО) С ТРЕМЯ ТОЧКАМИ СЛОЯ.

Введем обозначение $\sigma = \frac{\tau}{h^2}$ и решим уравнение (3) относительно $U_{i,j}$:

$$U_{i,j+1} = U_{i,j} + \sigma(U_{i-1,j} - 1U_{i,j} + U_{i+1,j}) + \tau f_{i,j}.$$

Пользуясь этим уравнением и значениями в начальных и граничных точках будем последовательно восстанавливать значения искомой функции в узлах сетки от слоя к слою. Такая схема называется явной.

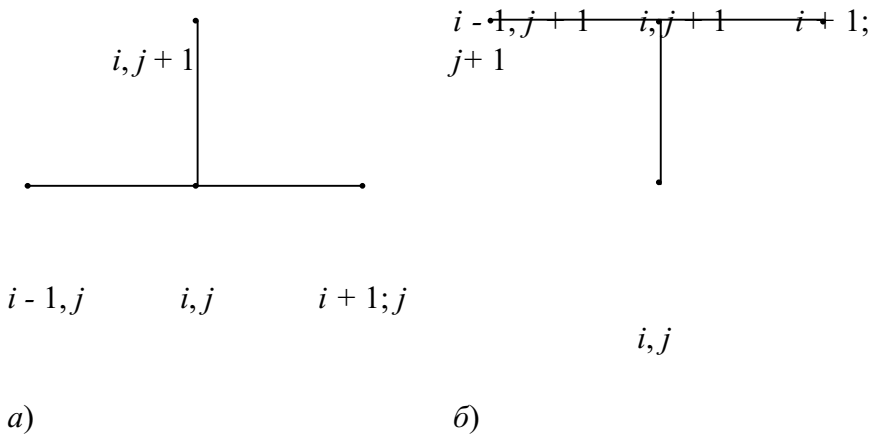


Рис. 2

Если заменить производную по времени левым конечно-разностным уравнением

$$\frac{\partial U(x_i, t_j)}{\partial t} \approx \frac{U_{i,j} - U_{i,j-1}}{\tau},$$

то получим неявную схему, связывающую три точки слоя j с одной точкой предыдущего слоя $j-1$:

$$\frac{U_{i,j} - U_{i,j-1}}{\tau} - \frac{U_{i-1,j} - 2U_{i,j} + U_{i+1,j}}{h^2} = f_{i,j}, \quad i = \overline{1, n-1}, \quad j = \overline{1, m};$$

$$U_{i,0} = \varphi_i, \quad i = \overline{1, n-1};$$

$$U_{0,j} = \psi_j; \quad U_{n,j} = \phi_j, \quad j = \overline{0, m}. \quad (5)$$

Получаемая система линейных уравнений не эквивалентна системе с треугольной матрицей как при использовании явной схемы и для ее решения используется специальный прием, позволяющий разделить систему на некоторое число систем меньшей размерности.

Рассмотрим точки первого слоя с номерами $i = \overline{0, n}$, $j = 1$. Поскольку значения искомой функции на предыдущем слое известны из начальных условий для каждой точки этого слоя получим уравнение, связывающее три неизвестных, причем в крайних точках слоя значения функции известны из граничных условий, т.е. получаем систему уравнений с трехдиагональной матрицей, решить которую можно методом прогонки. Затем, используя это решение, можно найти значения функции на следующем слое и т.д.

При решении задач с частными производными особенно актуальным становится вопрос об устойчивости конечно-разностных схем. Пусть конечно-разностная схема с некоторым шагом

h (скалярным или векторным) записывается в виде некоторого конечно-разностного оператора

$$L_h U^{(h)} = F^{(h)},$$

где $U^{(h)}$ - совокупность значений искомой функции в узлах сетки; $F^{(h)}$ - совокупность значений правой части дифференциального уравнения в узлах сетки.

Определение. Решение, полученное с помощью схемы $L_h U^{(h)} = F^{(h)}$, называется *устойчивым*, если для любого $\varepsilon > 0$ найдется ограничение шага $h_0 > 0$ такое, что решение возмущенной задачи с ограниченным возмущением $L_h Z^{(h)} = F^{(h)} + \delta f^{(h)}$, $\|\delta f^{(h)}\| \leq \varepsilon$ будет не сильно отклоняться от решения исходной задачи

$$\|U^{(h)} - Z^{(h)}\| \leq C \|\delta f^{(h)}\| \leq C\varepsilon,$$

при любых значениях шага $0 \leq h \leq h_0$ и константе C , независимой от шага h .

Устойчивость явной схемы (3) зависит от соотношения шагов h и τ ; при выполнении условия $0 < \sigma < 0,5$ эта схема будет устойчивой, в противном случае, т.е. при $\sigma > 0,5$ явная схема перестает быть устойчивой. Неявная схема (5) устойчива при любом выборе шагов.

Уравнение колебаний (гиперболическое уравнение). Поперечные колебания струны единичной длины, имеющей начальную форму $f(x)$, начальную скорость $F(x)$ и подвергающуюся воздействию на концах $\varphi(t)$, $\psi(t)$ описываются краевой задачей

$$\begin{aligned} \frac{\partial^2 U}{\partial t^2} &= a^2 \frac{\partial^2 U}{\partial x^2}, \quad 0 \leq x \leq 1, \quad 0 \leq t \leq \infty; \\ U(x, 0) &= f(x); \quad \frac{\partial U(x, 0)}{\partial t} = F(x); \\ U(0, t) &= \varphi(t); \quad U(1, t) = \psi(t). \end{aligned}$$

Как и для параболического уравнения ограничимся рассмотрением решения в прямоугольной области $0 \leq x \leq 1$, $0 \leq t \leq T$ и избавимся от параметра a^2 , считая его равным единице. Тогда задача примет вид

$$\begin{aligned} \frac{\partial^2 U}{\partial t^2} &= \frac{\partial^2 U}{\partial x^2}, \quad 0 \leq x \leq 1, \quad 0 \leq t \leq T; \\ U(x, 0) &= f(x); \quad \frac{\partial U(x, 0)}{\partial t} = F(x); \\ U(0, t) &= \varphi(t); \quad U(1, t) = \psi(t). \end{aligned}$$

Наложим на область решения сетку с шагом по длине h и шагом по времени τ : $x_i = ih$; $t_j = j\tau$, $i = \overline{0, n}$; $j = \overline{0, m}$. Как и для эллиптического уравнения заменим производные центральными разностными отношениями и, введя обозначение $\sigma = \tau/h$, получим разностную схему вида

$$U_{i,j+1} - 2U_{i,j} + U_{i,j-1} + \sigma^2 (U_{i+1,j} - 2U_{i,j} + U_{i-1,j}). \quad () \quad (6)$$

Эта схема является явной и, если известны значения функции $U_{i,j}$ на двух слоях j , $(j - 1)$, то можно вычислить значения в слое $(j + 1)$.

Устойчивость этой схемы обеспечивается при $\sigma \leq 1$.

Для нахождения значений функции в первых двух слоях ($j = 0$, $j = 1$) будем использовать начальные условия. Из условия $U(x, 0) = f(x)$ получим уравнения $U_{i,0} = f_i$, ($i = \overline{1, n-1}$). Из

второго начального условия путем замены производной

$$\frac{\partial U_{i,j}}{\partial t} \approx \frac{U_{i,j} - U_{i,j-1}}{\tau}$$

при $j = 1$ получим уравнения

$$U_{i,1} = f_i + \tau F_i, \quad i = \overline{1, n-1}.$$

Далее можно использовать уравнение (6), а для определения значений функции в крайних точках - воспользоваться уравнениями, полученными из краевых условий $U_{0,j} = \varphi_j, \quad U_{n,j} = \psi_j, \quad j = \overline{0, m}$.

Точность получаемого таким образом решения существенно зависит от точности определения значений функции в слое $j = 1$. Для увеличения точности будем использовать разложение в ряд Тейлора:

$$U_{i,1} \approx U_{i,0} + \tau \frac{\partial U_{i,0}}{\partial t} + \frac{\tau^2}{2} \frac{\partial^2 U_{i,0}}{\partial t^2}.$$

Используя исходное дифференциальное уравнение, заменим вторую производную по времени производной по длине

$$U_{i,1} \approx U_{i,0} + \tau \frac{\partial U_{i,0}}{\partial t} + \frac{\tau^2}{2} \frac{\partial^2 U_{i,0}}{\partial x^2}.$$

Тогда, замечая, что

$$U_{i,0} = f_i; \quad \frac{\partial U_{i,0}}{\partial t} = F_i; \quad \frac{\partial^2 U_{i,0}}{\partial x^2} = f''_i,$$

получим более точное значение для слоя $j = 1$:

$$U_{i,1} = f_i + \tau F_i + \frac{\tau^2}{2} f''_i.$$

Заметим, что если функция f имеет вторую производную, равную нулю, то этот метод совпадает с первым.

Тема 11 МЕТОДЫ РЕШЕНИЯ ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ

Под интегральными уравнениями понимается уравнение, содержащее неизвестную функцию $y(x)$ под знаком определенного интеграла. Рассмотрим наиболее часто встречающиеся уравнения:

$$\int_a^b k(x,s)y(s)ds = f(x); \quad (1)$$

$$y(x) + \lambda \int_a^b k(x,s)y(s)ds = f(x). \quad (2)$$

Уравнения (1) и (2) называются уравнениями Фредгольма, соответственно, первого и второго рода. Функция $k(x, s)$ называется ядром интегрального уравнения. Те значения λ , при которых уравнение (2) имеет нетривиальные решения, называются собственными решениями

(собственными числами) ядра $k(x, s)$, а отвечающие им ненулевые решения $y(x)$ - собственными функциями.

Интегральное уравнение может быть получено из линейной краевой задачи. Рассмотрим одну из них: (3)

$$L[y] = p(x)y''(x) + p'(x)y'(x) + g(x)y(x) = f(x); \quad p(x) > 0, \quad x \in [a, b]; \quad (3)$$

$$\Gamma_a[y] \equiv \alpha_0 y(a) + \alpha_1 y'(a) = 0; \quad (4)$$

$$\Gamma_b[y] \equiv \beta_0 y(b) + \beta_1 y'(b) = 0;$$

$$(|\alpha_0| + |\alpha_1| \neq 0, \quad |\beta_0| + |\beta_1| \neq 0), \quad (4)$$

где $p(x)$, $p'(x)$, $g(x)$ непрерывные на $[a, b]$ функции.

Определение. Функция $G(x, s)$ называется функцией Грина задачи (3), (4), если выполняются следующие условия:

- 1) $G(x, s)$ - определена и непрерывна в области $a \leq x \leq b; a \leq s \leq b$.
- 2) $L_x [G(x, s)] = 0$ при $x \neq s$.
- 3) $\Gamma_a[G(a, s)] = 0, \quad \Gamma_b[G(b, s)] = 0$.
- 4) $G'_s(s+0, s) - G'_s(s-0, s) = -\frac{1}{p(s)}$.

Если функция Грина найдена, то решение краевой задачи (3), (4) дается формулой

$$y(x) = -\int_a^b G(x, s) f(s) ds.$$

5) Функция Грина симметрическая, если $G(x, s) = G(s, x)$.

Метод последовательных приближений. Запишем уравнение Фредгольма в виде

$$y(x) = f(x) + \lambda \int_a^b k(x, s) f(s) ds. \quad (5)$$

Пусть функции ядра $k(x, s)$ и свободного члена $f(x)$ непрерывны. Будем искать решение в форме степенного ряда

$$y(x) = \sum_{n=0}^{\infty} \lambda^n \varphi_n(x). \quad (6)$$

Подставим (6) в (5), приравнявая коэффициенты при одинаковых степенях параметра получим

$$\begin{aligned} \varphi_0(x) &= f(x); \\ \varphi_n(x) &= \int_a^b k(x, s) \varphi_{n-1}(s) ds, \quad (n=1, 2, \dots) \end{aligned}$$

Пусть функции $k(x, s)$ и $f(x)$ ограничены:

$$|k(x, s)| \leq M; \quad |f(x)| \leq N, \quad x, s \in [a, b],$$

тогда из (7) получаем

$$|\varphi_n(x)| \leq M^n N (b-a)^n.$$

Сходимость (6) обеспечивается, если $|x| < \frac{1}{M(b-a)}$.

Приняв $y(x) \approx y_n(x) = \sum_{k=0}^n \lambda^k \varphi^k(x)$, получим решение (5) с погрешностью

$$E_n = \frac{N[M(b-a)|\lambda|]^{n+1}}{1 - M(b-a)|\lambda|}. \quad (8)$$

Из формулы (7) вытекает, что решение (6) можно представить в виде

$$y(x) = f(x) + \lambda \sum_{n=1}^{\infty} \lambda^{n-1} \int_a^b k_n(x,s) f(s) ds \quad (9)$$

Интегрированные ядра $k_n(x, s)$ могут быть последовательно найдены

$$k_n(x,s) = \int_a^b k(x,t) k_{n-1}(t,s) dt, \quad n = 2, 3, \dots$$

Функция

$$R(x,s,\lambda) = \sum_{n=1}^{\infty} \lambda^{n-1} k_n(x,s) \quad (10)$$

называется резольвентой уравнения (5) и при малых λ определяется степенью ряда. Резольвенту $R(x, s, \lambda)$ можно продолжить на всю комплексную плоскость параметра λ за исключением собственных значений $\lambda_1, \dots, \lambda_n$ (особые точки), которые являются корнями резольвенты. Тогда (10) дает решение интегрального уравнения (5) при любом $\lambda \neq \lambda_i$ ($i = \overline{1, n}$).

Решение интегрального уравнения методом конечных сумм. Метод основывается на приближенном вычислении определенного интеграла в интегральном уравнении с помощью некоторой квадратурной формулы

$$\int_a^b f(x) dx = \sum_{i=1}^n A_i f(x_i) + R.$$

Применим этот подход к уравнению (2), получим

$$y(x) - \lambda \sum_{i=1}^n A_i k(x, s_i) y(s_i) + R = f(x).$$

Эта формула справедлива для любых значений x из интервала $[a, b]$, в том числе и для значений x , совпадающих с узлами квадратурных формул. Записав эти уравнения для $x_i = s_i$ и отбрасывая остаточный член, получим систему уравнений

$$y_i - \lambda \sum_{i=1}^n A_i k_{i,j} y_i = f_i, \quad i = \overline{1, n}, \quad (11)$$

где $y_i = y(x_i)$; $k_{i,j} = k(x_i, x_j)$; $f_i = f(x_i)$, $i, j = \overline{1, n}$.

Используя символ Кронекера $\delta_{i,j} = \begin{cases} 0, & i \neq j; \\ 1, & i = j \end{cases}$ и учитывая, что $y_i = \sum_{j=1}^n \delta_{i,j} y_j$, систему (11)

запишем в виде

$$\sum_{j=1}^n (\delta_{i,j} - \lambda A_j k_{i,j}) y_j = f_i, \quad i = \overline{1, n} \quad (12)$$

если $\Delta(\lambda) = \det(\delta_{i,j} - \lambda A_j k_{i,j}) \neq 0$, то система (12) имеет единственное решение y_i , которое можно найти любым методом решения системы линейных уравнений. Найдя y_i , получаем приближенное решение (2) в виде

$$y(x) = f(x) + \lambda \sum_{j=1}^n A_j k(x, x_j) y_j.$$

Метод вырожденного ядра. Введем следующее.

О п р е д е л е н и е. Ядро $k(x, s)$ называется вырожденным, если его можно представить в виде конечной суммы парных произведений

$$k(x, s) = \sum_{i=1}^n \alpha_i(x) \beta_i(s),$$

где $\alpha_i(x), \beta_i(x)$ - линейно независимые функции.

Для таких ядер уравнение Фредгольма второго рода можно записать в виде

$$y(x) = f(x) + \lambda \sum_{i=1}^n c_i \alpha_i(x), \quad (13)$$

ãäå

$$c_i = \int_a^b \beta_i(s) y(s) ds, \quad i = \overline{1, n}. \quad (14)$$

Если (14) подставить в (13), то для определения c_i получим систему линейных алгебраических уравнений

$$c_i = \int_a^b \beta_i(s) f(s) ds + \lambda \int_a^b \beta_i(s) \sum_{j=1}^n c_j \alpha_j(s) ds, \quad i = \overline{1, n}.$$

Запишем ее в виде

$$c_i - \lambda \sum_{j=1}^n c_j j_{i,j} = f_i, \quad i = \overline{1, n}, \quad (15)$$

где $f_i = \int_a^b \beta_i(s) f(s) ds; \quad j_{i,j} = \int_a^b \alpha_i(s) \beta_j(s) ds.$

Используя символ Кронекера, представим (15) в виде

$$\sum_{j=1}^n (\delta_{i,j} - \lambda j_{i,j}) c_j = f_i, \quad i = \overline{1, n}.$$

$$\Delta(\lambda) = \det(\delta_{i,j} - \lambda j_{i,j}) = \begin{vmatrix} 1 - \lambda j_{1,1} & -\lambda j_{2,1} & \dots & -\lambda j_{n,1} \\ -\lambda j_{1,2} & 1 - \lambda j_{2,2} & \dots & -\lambda j_{n,2} \\ \dots & \dots & \dots & \dots \\ -\lambda j_{1,n} & -\lambda j_{2,n} & \dots & 1 - \lambda j_{n,n} \end{vmatrix}$$

Если $\Delta(\lambda) \neq 0$, то неизвестные коэффициенты c_i определяются по формулам Крамера

$$c_i = \frac{\sum_{j=1}^n \Delta_{i,j}(\lambda) f_j}{\Delta(\lambda)}, \quad i = \overline{1, n}$$

и уравнение (2) имеет единственное решение:

$$y(x) = f(x) + \lambda \sum_{i=1}^n \sum_{j=1}^n \frac{\Delta_{j,i}(\lambda)}{\Delta(\lambda)} f_j \alpha_i(x).$$

В том случае, если $\Delta(\lambda) = 0$, т.е. когда λ собственное значение, имеем либо множество решений, либо вообще таких не имеем.

Метод коллокаций. Рассмотрим интегральное уравнение Фредгольма второго рода в виде

$$y(x) - \lambda \int_a^b k(x,s) y(s) ds - f(x) = 0. \quad (16)$$

Будем искать приближенное решение уравнения (16) в виде некоторой функции определенного вида $y_n = \Phi(x, c_1, c_2, \dots, c_n)$ со свободными параметрами c_1, c_2, \dots, c_n . Подставляя ее в (16), получим невязку

$$R[y_n] = y_n(x) - \lambda \int_a^b k(x,s) y_n(s) ds - f(x) \quad (17)$$

Если y_n точное решение, то невязка $R[y] \equiv 0$. Поэтому нужно подобрать параметры $c_i (i = \overline{1, n})$ так, чтобы невязка $R[y_n]$ в определенном смысле была возможно малой.

Как и в методах решения двухточечной краевой задачи, изложенных в лекции 9, минимизировать $R[y_n]$ можно различными способами. Рассмотрим некоторые из них.

Пусть

$$y_n = \varphi_0(x) + \sum_{i=1}^n c_i \varphi_i(x), \quad (18)$$

где $\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x)$ - заданная система линейно независимых функций (координатные функции); $c_i, (i = \overline{1, n})$ - неопределенные коэффициенты. Заметим, что можно положить $\varphi_0(x) \equiv 0$. Подставим (18) в (16), получим невязку

$$R[y_n] = \varphi_0(x) + \sum_{i=1}^n c_i \varphi_i(x) - f(x) - \lambda \int_a^b k(x,s) \left[\varphi_0(s) + \sum_{i=1}^n c_i \varphi_i(s) \right] ds$$

или
$$R[y_n(x)] = \psi_0(x, \lambda) + \sum_{i=1}^n c_i \psi_i(x, \lambda), \quad (19)$$

где

$$\psi_0(x, \lambda) = \varphi_0(x) - f(x) - \lambda \int_a^b k(x, s) \varphi_0(s) ds;$$

$$\psi_i(x, \lambda) = \varphi_i(x) - \lambda \int_a^b k(x, s) \varphi_i(s) ds, \quad i = \overline{1, n}.$$

Согласно методу коллокаций потребуем, чтобы невязка $R[y_n(x)]$ обращалась в ноль в заданной системе точек $x_j \in [a, b]$, $j = \overline{1, n}$ (точки коллокаций), т.е. $R[y_n(x_j)] = 0$, $j = \overline{1, n}$; $a \leq x_1 < x_2 < \dots < x_{n-1} < x_n \leq b$.

Отсюда, используя (19), получаем систему алгебраических уравнений для определения c_i :

$$\sum_{i=1}^n c_i \psi_i(x_j, \lambda) = -\psi_0(x_j, \lambda), \quad j \in \overline{1, n}. \quad (20)$$

Если $\Delta(x) = \det[\psi_i(x_j, \lambda)] \neq 0$, то из (20) однозначно определяются коэффициенты c_i , $i = \overline{1, n}$.

Метод наименьших квадратов. В этом методе, как и в предыдущем, решение уравнения (16) представляется в виде некоторой функции (18) зависящей от n коэффициентов c_i . Потребуем, чтобы интеграл квадрата невязки был минимален

$$I = \int_a^b \{ R [T_n(x)] \}^2 dx = \int_a^b \left[\psi_0(x, \lambda) + \sum_{i=1}^n c_i \psi_i(x, \lambda) \right]^2 dx \rightarrow \min,$$

для чего приравняем нулю частные производные $\frac{\partial I}{\partial c_j} = 0$, $j = \overline{1, n}$.

Получим

$$\frac{1}{2} \frac{\partial I}{\partial c_j} = \int_a^b \psi_j(x, \lambda) \left(\psi_0(x, \lambda) + \sum_{i=1}^n c_i \psi_i(x, \lambda) \right) dx = 0, \quad i = \overline{1, n}. \quad (22)$$

Используя скалярные произведения, систему (22) запишем в виде

$$(\psi_i, \psi_j) = \int_a^b \psi_i(x, \lambda) \psi_j(x, \lambda) dx;$$

$$c_1(\psi_1, \psi_1) + c_2(\psi_1, \psi_2) + \dots + c_n(\psi_1, \psi_n) = -(\psi_0, \psi_1);$$

$$c_1(\psi_2, \psi_1) + c_2(\psi_2, \psi_2) + \dots + c_n(\psi_2, \psi_n) = -(\psi_2, \psi_0);$$

.....

$$c_1(\psi_n, \psi_1) + c_2(\psi_n, \psi_2) + \dots + c_n(\psi_n, \psi_n) = -(\psi_n, \psi_0).$$

Поскольку $\varphi_0 = 0$, то $\psi_0 = -f(x)$ и, следовательно, $-(\psi_i, \varphi_0) = (\psi_i, f)$, $i = \overline{1, n}$, т.е. матрица системы (23) симметричная и решить такую систему можно методом квадратного корня.

Метод моментов: Коэффициенты формы (18) определяются из условия ортогональности невязки ко всем координатным функциям $\varphi_0(x), \varphi_1(x), \dots, \varphi_n(x)$, что дает систему уравнений

$$\int_a^b R[y_n] \varphi_i(x) dx = 0, \quad i = \overline{1, n};$$

или после преобразований

$$\sum_{j=1}^n c_j (a_{i,j} - \lambda \beta_{i,j}) = \lambda j_i, \quad i = \overline{1, n}, \quad (24)$$

где

$$\begin{aligned} \alpha_{i,j} &= \int_a^b \varphi_i(x) \varphi_j(x) dx; \\ \beta_{i,j} &= \int_a^b dx \int_a^b k(x,s) \varphi_i(x) \varphi_j(s) ds; \\ j_i &= \int_a^b dx \int_a^b k(x,s) \varphi_i(x) f(s) ds. \end{aligned}$$

Если определитель системы (24) $\Delta(\lambda) = \det(\alpha_{i,j} - \lambda \beta_{i,j})$ отличен от нуля, то из этой системы можно однозначно определить коэффициенты c_1, c_2, \dots, c_n . Тогда (2) дает приближенное решение (18). Из уравнения $\Delta(\lambda) = 0$ приближенно находятся $\lambda_1, \lambda_2, \dots, \lambda_n$, собственные значения ядра $k(x, s)$.

Тема 12 ОСНОВЫ ТЕОРИИ ЦЕПНЫХ ДРОБЕЙ

Цепной или непрерывной дробью называется выражение вида

$$b_0 + \frac{a_1}{b_1 + \frac{a_2}{b_2 + \dots}}; \quad (1)$$

:

$$+ \frac{a_k}{b_k + \dots},$$

где $a_i b_j, 1=1, \dots; j=0, \dots$ - переменные; в частности они могут быть числами, матрицами, операторами, элементами абстрактных пространств, в которых введены или соответствующим образом определены операции сложения и деления. Компоненты a_i называются частными числителя,

b_j - частными знаменателя, b_0 - свободный член, дроби вида $\frac{a_i}{b_i}$ называются i -ми звеньями.

Далее будут рассматриваться цепные дроби, элементами которых являются действительные или комплексные числа:

$$\begin{aligned} \frac{b_0}{1} = \frac{P_0}{Q_0}; \quad b_0 + \frac{a_1}{b_1} = \frac{P_1}{Q_1}; \quad b_0 + \frac{a_1}{b_1 + \frac{a_2}{b_2}} = \frac{P_2}{Q_2} \dots \\ \dots b_0 + \frac{a_1}{b_1 + \frac{a_2}{b_2 + \dots}} = \frac{P_k}{Q_k}; \\ \vdots \\ \frac{a_k}{b_k}. \end{aligned} \quad (2)$$

Дроби вида (2) называются подходящими для цепной дроби (1). Для вычисления P_k, Q_k , соответственно числителя и знаменателя k -й подходящей дроби для (1), существуют непосредственно проверяемые рекуррентные соотношения.

Уже более 2000 лет математики занимаются изучением цепных дробей. Еще и теперь в каждом реферативном журнале "Математика" появляются рефераты работ, посвященных цепным дробям и их приложениям. Понятно, что такой огромный интервал времени может выдержать только теория, имеющая фундаментальное значение. С помощью цепных дробей решается целый ряд важных задач теории дифференциальных уравнений, вычислительной математики, теории устойчивости и оптимального управления.

В виде цепных дробей представляются решения обыкновенных дифференциальных уравнений и решения конечно-разностных уравнений. В частности, метод цепочки эквивалентен вычислению конечной дроби.

В настоящих лекциях рассмотрены возможности применения цепных дробей для решения задач интерполяции и аппроксимации функций одной и нескольких действительных переменных, решения дифференциальных уравнений, а также уравнений в частных производных и решения ряда задач теории устойчивости. Кроме этого, рассматриваются основы теории цепных дробей, которая может быть использована для решения задач теории графов и принятия решения.

$$P_k = b_k P_{k-1} + a_k P_{k-2}; \quad Q_k = b_k Q_{k-1} + a_k Q_{k-2}, \quad k = 1, 2, \dots \quad (3)$$

при начальных условиях

$$P_0 = b_0, \quad P_1 = 1, \quad Q_0 = 1, \quad Q_{-1} = 0; \quad (4)$$

используя (3) и (4), можно получить формулу

$$\frac{P_k}{Q_k} - \frac{P_{k-1}}{Q_{k-1}} = (-1)^{k+1} \frac{a_1 a_2 \dots a_k}{Q_k Q_{k-1}}, \quad k = 1, 2, \dots, \quad (5)$$

$$\frac{P_k}{Q_k} - \frac{P_{k-2}}{Q_{k-2}} = (-1)^k \frac{a_1 a_2 \dots a_k + b_k}{Q_k Q_{k-2}}, \quad k = 2, 3, \dots \quad (6)$$

Цепные дроби с комплексными компонентами удобно рассматривать как произведение дробно-линейных отображений. Пусть

$$T_0(z) = b_0 + z; \quad T_k(z) = \frac{a_k}{b_k + z}, \quad k = 1, 2, \dots \quad (7)$$

В этом случае

$$\frac{P_0}{Q_0} = T_0(0); \quad \frac{P_1}{Q_1} = T_0 T_1(0), \dots, \quad \frac{P_k}{Q_k} = T_0 T_1 \dots T_k(0). \quad (8)$$

В дальнейшем для сокращения записи дроби вида (1) будут записываться в виде

$$b_0 + \frac{a_1}{|b_1|} + \frac{a_2}{|b_2|} + \dots + \frac{a_k}{|b_k|} = b_0 + \sum_{k=1}^{\infty} \frac{a_k}{|b_k|} \quad (9)$$

или

$$b_0 + \frac{a_1}{b_1} + \frac{a_2}{b_2} + \dots + \frac{a_k}{b_k} = b_0 + \sum_{k=1}^{\infty} \frac{a_k}{b_k}. \quad (10)$$

Наиболее хорошо изучены цепные дроби с положительными компонентами. Для них характерным свойством является свойство «вилки», непосредственно следующее из формул (5), (6). Приведем без доказательства ряд теорем, условия, полученные в которых, будут использоваться ниже.

Теорема 1. Если у цепной дроби (1) все компоненты положительные, то для подходящих дробей (2) выполняются неравенства

$$\begin{aligned} \frac{P_1}{Q_1} &> \frac{P_3}{Q_3} > \dots > \frac{P_{2i-1}}{Q_{2i-1}} > \dots \\ \frac{P_0}{Q_0} &< \frac{P_2}{Q_2} < \dots < \frac{P_{2k}}{Q_{2k}} < \dots \\ \frac{P_{2i-1}}{Q_{2i-1}} &> \frac{P_{2k}}{Q_{2k}}, \quad i = 1, 2, \dots; \quad k = 0, 1, \dots \end{aligned} \quad (11)$$

Неравенство (11) называется свойством "вилки". Рассмотрим важный вопрос о сходимости цепной дроби (1). Цепная дробь (1) с произвольными компонентами называется сходящейся, если существует и конечен предел последовательности подходящих дробей $\{P_n/Q_n\}$. Величину этого предела в случае его существования будем называть значением бесконечной дроби (1).

Пусть члены дроби (1) положительны, тогда, используя свойство "вилки" можно сформулировать следующее предположение.

Предположение 1. Необходимыми и достаточными условиями сходимости цепной дроби (1) с положительными компонентами является стремление к нулю разности [13]:

$$\frac{P_{2p+1}}{Q_{2p+1}} - \frac{P_{2p}}{Q_{2p}} \text{ при } p \rightarrow \infty.$$

Вопрос сходимости цепных дробей с положительными компонентами полностью решается теоремой Зейделя.

Теорема 2 (Зейдель). Для того, чтобы цепная дробь $b_0 + \sum_{i=1}^{\infty} \frac{1}{|b_i|}$ с положительными компонентами сходилась, необходимо, чтобы ряд $\sum_{n=1}^{\infty} b_n$ расходился.

Достаточный признак сходимости цепных дробей с комплексными компонентами сформулирован в следующей теореме.

Теорема 3. Цепная дробь (1) с комплексными компонентами сходится, если выполняется неравенство [13]:

$$|b_i| - |a_i| \geq 1, \quad i = 1, 2, \dots \quad (12)$$

Арифметические операции над цепными дробями

А. Я. Хинчин в работе [3], анализируя преимущества и недостатки цепных дробей как математического аппарата, отмечал, что эффективное использование всякого аппарата требует достаточно простых формул для выполнения арифметических операций над его элементами, без чего аппарат не может служить средством вычислений.

Рассмотрим арифметические операции над двумя цепными дробями общего вида с неодинаковым количеством звеньев:

$$S' = b_0 + \frac{a_1}{|b_1|} + \frac{a_2}{|b_2|} + \dots + \frac{a_{k_1}}{|b_{k_1}|}, \quad (13)$$

$$S'' = d_0 + \frac{c_1}{|d_1|} + \frac{c_2}{|d_2|} + \dots + \frac{c_{k_2}}{|d_{k_2}|}. \quad (14)$$

Предположим, что число звеньев дроби (14) больше числа звеньев дроби (13), т.е. $k_2 > k_1$. Частный случай равенства $k_1 = k_2$ рассмотрим ниже. Для простоты изложения переобозначим элементы цепных дробей (13) и (14) следующим образом:

$$\begin{aligned} \alpha_{2i-1} = a_i; \quad \beta_{2i-1} = b_i; \quad \alpha_{2i} = c_i; \quad \beta_{2i} = d_i, \quad i = \overline{1, k}; \\ \alpha_{k_1+i} = c_i; \quad \beta_{k_1+i} = b_i, \quad i = \overline{k_1+1, k_2}. \end{aligned} \quad (15)$$

Потребуем, чтобы $\alpha_i \neq 0$, $\beta_i \neq 0$, $i = \overline{1, k_1 + k_2}$. В [13] предполагается следующая методика решения задачи.

• Обозначим через A_n/B_n , $i = \overline{0, k_1}$ подходящие цепные дроби (13), через C_n/D_n , $n = \overline{0, k_2}$ – подходящие цепные дроби дроби (14) из элементов последовательностей подходящих дробей $\{A_n/B_n\}_0^{k_1}, \{C_n/D_n\}_0^{k_2}$ образовывается последовательность $\{P_n/Q_n\}_0^{k_1+k_2}$, у которых i -ый элемент P_i/Q_i содержит звенья $b_0 \otimes d_0, \alpha_1/\beta_1, \alpha_2/\beta_2, \dots, \alpha_i/\beta_i$ исходных цепных дробей (13), (14). Эта последовательность называется последовательностью непрерывных подходящих дробей результата арифметической операции. Элементы последовательности $\{P_n/Q_n\}_0^{k_1+k_2}$ выражается через элементы последовательностей $\{A_n/B_n\}_0^{k_1}$ и $\{C_n/D_n\}_0^{k_2}$ следующим образом:

$$\frac{P_n}{Q_n} = \frac{A_{n_1}}{B_{n_1}} \otimes \frac{C_{n_2}}{D_{n_2}}, \quad n_1 = \overline{0, 2k_1}, \quad (16)$$

где $n_1 = [(n+1)/2]$; $n_2 = [n/2]$, квадратные скобки означают деление нацело.

$$\frac{P_{k_1+n}}{Q_{k_1+n}} = \frac{A_{k_1}}{B_{k_1}} \otimes \frac{C_n}{D_n}, \quad n = \overline{k_1+1, k_1+k_2}, \quad (17)$$

\otimes - определяет одну из четырех арифметических операций.

• Представим подходящие цепные дроби из последовательностей (16), (17) в виде отношения двух определителей специального вида, составленных из элементов цепных дробей (13), (14), обладающих следующими свойством: определитель n -ой подходящей дроби получается из определителей $(n+1)$ -й подходящей дроби, если в последних определителях вычеркнуть последнюю строку и последний столбец.

• Разложим определители числителя и знаменателя по элементам последнего столбца (или строки). Полученные разложения с учетом свойства (2) будут играть роль рекуррентных соотношений для числителя и знаменателя подходящих дробей последовательности (16), (17).

• Численное значение результата арифметической операции над цепными дробями (13), (14) будем вычислять по полученным в (3) рекуррентным формулам, при соответствующих данной операции начальных условиях.

Проведем конкретные выкладки по предложенной методике. Из элементов цепных дробей (13), (14) составим определитель $\Delta_{k_1+k_2, l, m} = |\omega_{i,j}|_1^{k_1+k_2+2}$ размера k_1+k_2+2 следующего вида:

где $M_{l,m}$ - матрица размера 5×5 , элементы которой будут определены ниже в зависимости от рассматриваемой арифметической операции.

Остальные элементы $\omega_{i,j}$ определителя $\Delta_{k_1+k_2, l, m}$ определяются через элементы цепных дробей (13), (14) в обозначениях (15) соотношениями

$$\begin{aligned} \omega_{i+2, i+2} &= \beta_i, \quad i = \overline{4, k_1+k_2}; \\ \omega_{i, i+2} &= \alpha_i, \quad i = \overline{4, 2k_1}; \\ \omega_{i+1, i+2} &= \alpha_i, \quad i = \overline{2k_1+1, k_1+k_2}; \\ \omega_{i+2, i} &= -1, \quad i = \overline{4, 2k_1}; \\ \omega_{i+2, i+1} &= -1, \quad i = \overline{2k_1+1, k_1+k_2}; \end{aligned}$$

$$\begin{aligned} \omega_{i,j} &= 0; \quad |i-j| > 0, \quad i = \overline{1,5}, j = \overline{6, k_1 + k_2 + 2}; \quad i = \overline{6, k_1 + k_2 + 2}, \quad j = \overline{1,5}; \\ \omega_{i,j} &= 0; \quad |i-j| = 1, \quad i, j = \overline{5, 2k_1 + 2}; \\ \omega_{i,j} &= 0; \quad |i-j| = 2, \quad i, j = \overline{2k_1 + 2, k_1 + k_2 + 2}. \end{aligned}$$

Значение индекса l определяет арифметическую операцию: $l = 1$ - сложения, $l = 2$ - вычитания, $l = 3$ - умножения, $l = 4$ - деления. Индекс m принимает значения 1, 2 в зависимости от того, где находится определитель - в числителе ($m = 1$) или в знаменателе ($m = 2$). Обозначим через $\Delta_{n,l,m}$, $n = \overline{0, k_1 + k_2}$ главный минор определителя $\Delta_{k_1+k_2, l, m}$, используя выше приведенные обозначения, рассмотрим теорему.

Теорема 1. Для каждой из арифметических операций ($l = \overline{1,4}$) n -я подходящая дробь из последовательности (16), (17) представляются в виде отношения следующих определителей [13]:

сложение $l = 1$ ($S' + S''$):

$$\frac{P_n}{Q_n} = \frac{\Delta_{n,1,1}}{\Delta_{n,1,2}}, \quad n = \overline{0, k_1 + k_2},$$

$$\text{где } M_{1,1} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & b_0 + d_0 & \alpha_1 & \alpha_2 & 0 \\ 0 & -1 & \beta_1 & 0 & \alpha_3 \\ 0 & -1 & 0 & \beta_2 & 0 \\ 0 & 0 & -1 & 0 & \beta_3 \end{vmatrix}; \quad M_{1,2} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \beta_1 & 0 & \alpha_3 \\ 0 & 0 & 0 & \beta_2 & 0 \\ 0 & 0 & -1 & 0 & \beta_3 \end{vmatrix}. \quad (18)$$

вычитание $l = 2$ ($S' - S''$):

$$\frac{P_n}{Q_n} = \frac{\Delta_{n,2,1}}{\Delta_{n,2,2}}, \quad n = \overline{0, k_1 + k_2}; \quad (19)$$

$$M_{2,1} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & b_0 - d_0 & \alpha_1 & \alpha_2 & 0 \\ 0 & -1 & \beta_1 & 0 & \alpha_3 \\ 0 & 1 & 0 & \beta_2 & 0 \\ 0 & 0 & -1 & 0 & \beta_3 \end{vmatrix}; \quad M_{2,2} = M_{1,2};$$

умножение $l = 3$ ($S' * S''$):

$$\frac{P_n}{Q_n} = \frac{\Delta_{n,3,1}}{\Delta_{n,3,2}}, \quad n = \overline{0, k_1 + k_2}; \quad (20)$$

$$M_{3,1} = \begin{vmatrix} b_0 & 0 & \alpha_1 & 0 & 0 \\ 0 & d_0 & 0 & \alpha_2 & 0 \\ -1 & 0 & \beta_1 & 0 & \alpha_3 \\ 0 & -1 & 0 & \beta_2 & 0 \\ 0 & 0 & -1 & 0 & \beta_3 \end{vmatrix}; \quad M_{3,2} = M_{1,2};$$

деление $l = 4$ ($S' : S''$):

$$\frac{P_n}{Q_n} = \frac{\Delta_{n,4,1}}{\Delta_{n,4,2}}, \quad n = \overline{0, k_1 + k_2}; \quad (21)$$

$$M_{4,1} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & b_0 & \alpha_1 & 0 & 0 \\ 0 & -1 & \beta_1 & 0 & \alpha_3 \\ 0 & 0 & 0 & \beta_2 & 0 \\ 0 & 0 & -1 & 0 & \beta_3 \end{vmatrix}; \quad M_{4,2} = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & d_0 & 0 & \alpha_2 & 0 \\ 0 & 0 & \beta_1 & 0 & \alpha_3 \\ 0 & -1 & 0 & \beta_2 & 0 \\ 0 & 0 & -1 & 0 & \beta_3 \end{vmatrix}.$$

Из теоремы 2 следует, что определители $\Delta_{n,l,m}$, $l = \overline{1,4}$; $m = \overline{1,2}$; $n = \overline{4, k_1 + k_2}$ при фиксированном n отличаются между собой лишь значениями элементов $\omega_{i,j}$, $i, j = \overline{1,6}$, значения же остальных элементов у всех определителей одинаковые: поэтому разложение этих определителей по элементам последнего столбца (или строки) дают единое для всех значений l, m соотношения

$$\begin{aligned} \Delta_{n,l,m} &= \beta_n \Delta_{n-1,l,m} + \alpha_n (\beta_{n-1} \Delta_{n-3,l,m} + \alpha_{n-1} \Delta_{n-4,l,m}), \quad n = \overline{4, 2k_1}; \\ \Delta_{n,l,m} &= \beta_n \Delta_{n-1,l,m} + \alpha_n \Delta_{n-2,l,m}, \quad n = \overline{2k_1 + 1, k_1 + k_2}. \end{aligned} \quad (22)$$

Учитывая (18), (21) и (22), можно получить единое для всех арифметических операций рекуррентное соотношение для числителя и знаменателя подходящих дробей последовательности (16), (17):

$$\begin{aligned} P_n &= \beta_n P_{n-1} + \alpha_n (\beta_{n-1} P_{n-3} + \alpha_{n-1} P_{n-4}); \\ Q_n &= \beta_n Q_{n-1} + \alpha_n (\beta_{n-1} Q_{n-3} + \alpha_{n-1} Q_{n-4}), \quad n = \overline{4, 2k_1}; \\ P_n &= \beta_n P_{n-1} + \alpha_n P_{n-2}; \\ Q_n &= \beta_n Q_{n-1} + \alpha_n Q_{n-2}, \quad n = \overline{2k_1 + 1, k_1 + k_2} \end{aligned} \quad (23)$$

с начальными условиями

$$\begin{aligned} P_0 &= \Delta_{0,l,1}; \quad P_1 = \Delta_{1,l,1}; \quad P_2 = \Delta_{2,l,1}; \quad P_3 = \Delta_{3,l,1}; \\ Q_0 &= \Delta_{0,l,2}; \quad Q_1 = \Delta_{1,l,2}; \quad Q_2 = \Delta_{2,l,2}; \quad Q_3 = \Delta_{3,l,2}. \end{aligned} \quad (24)$$

Выполнение арифметических операций

$$(S' + S''), (S' - S''), (S' * S''), (S' : S'')$$

над исходными цепными дробями (13), (14) производится по рекуррентным формулам (23) с начальными условиями (24) при соответствующем данной операции значении индекса l . Операция $S'' - S'$ выполняется по рекуррентным формулам:

$$\begin{aligned} P_n &= -\beta_n P_{n-1} - \alpha_n (\beta_{n-1} P_{n-3} + \alpha_{n-1} P_{n-4}); \\ Q_n &= \beta_n Q_{n-1} + \alpha_n (\beta_{n-1} Q_{n-3} + \alpha_{n-1} Q_{n-4}), \quad n = \overline{4, 2k_1}; \\ P_n &= -\beta_n P_{n-1} - \alpha_n P_{n-2}; \\ Q_n &= \beta_n Q_{n-1} + \alpha_n Q_{n-2}, \quad n = \overline{2k_1 + 1, k_1 + k_2}, \end{aligned} \quad (25)$$

с начальными условиями (24) при $l = 2$. Операция $S'' : S'$ выполняется при $l = 4$ по рекуррентным формулам (23) с начальным условием:

$$\begin{aligned} P_0 &= \Delta_{0,4,2}; \quad P_1 = \Delta_{1,4,2}; \quad P_2 = \Delta_{2,4,2}; \quad P_3 = \Delta_{3,4,2}; \\ Q_0 &= \Delta_{0,4,1}; \quad Q_1 = \Delta_{1,4,1}; \quad Q_2 = \Delta_{2,4,1}; \quad Q_3 = \Delta_{3,4,1}. \end{aligned} \quad (26)$$

В случае одинакового количества звеньев $k_1 = k_2$ в исходных цепных дробях (13), (14) включение операций

$$(S' + S''), (S'' + S'), (S' - S''), (S' * S''), (S'' * S'), (S' : S'')$$

производится по рекуррентным формулам (23) с начальными условиями (24) при соответствующем данной операции значении индекса l . Операция $(S'' - S')$ с начальными условиями (24) при $l = 2$. Операция $(S'' : S')$ по формулам (23) с начальными условиями (26) при $l = 4$.

Тема 13 АНАЛОГ ФОРМУЛЫ ТЕЙЛОРА В ТЕОРИИ ЦЕПНЫХ ДРОБЕЙ

Цепные дроби соответствующие степенному ряду. Функцию можно разложить и в ряд, и в цепную дробь. Естественно, возникает вопрос о том, какое же разложение лучше? Для ответа на этот вопрос в теории цепных дробей исследуется связь цепных дробей и степенных рядов.

Опишем численный метод типа Рунге-Кутты третьего порядка. Выберем на отрезке $[x_0, X]$ некоторую сетку $\{x_n, 0 \leq n \leq N\}$ значений аргумента так, чтобы выполнялись соотношения $x_0 < x_1 < x_2 < \dots < x_N = X$.

Разложение решения $y(x)$ задачи (1), (2) на отрезке $x_n \leq x \leq x_{n+1}$ по формуле Тиле, ограничим дробью с тремя звеньями, хотя в принципе можно выбрать дробь с большим числом звеньев.

$$y_{n+1} = y_n + \frac{h}{\sqrt{[y(x)]_{x=x_n} + \frac{h}{2\sqrt{[y(x)]_{x=x_n} + \frac{h}{3\sqrt{[y(x)]_{x=x_n}}}}}}$$

где $y_{n+1} \approx y(x_{n+1})$, $h = x_{n+1} - x_n$; $y = 1/y'$;

$$2\sqrt{[y]} = -\frac{2(y')^2}{y''}; \quad 3\sqrt{[y]} = \frac{3(y'')^2}{2(y')^2 y''' - 3y'(y'')^2}.$$

После замены обратных производных получаем

$$y_{n+1} = y_n + h y'_n + \frac{3h^2 (y''_n)^2}{6y''_n - 2h y'''_n}.$$

Исходя из приведенной формулы, будем искать желаемое решение в виде

$$y_{n+1} = y_n + h k_1 + h \frac{(k_2 - k_1)^2}{p_1 k_1 + p_2 k_2 + p_2 k_2}, \quad (3)$$

где

$$\begin{aligned} k_1 &= f(x_n, y_n); \\ k_2 &= f(x_n + \alpha_2 h, y_n + \beta_{21} h k_1); \\ k_3 &= f(x_n + \alpha_3 h, y_n + \beta_{31} h k_1 + \beta_{32} h k_2); \\ \alpha_2 &= \beta_{21}; \quad \alpha_3 = \beta_{31} + \beta_{32}. \end{aligned}$$

Неизвестные константы могут быть найдены из решения системы уравнений

$$\begin{aligned}
p_1 + p_2 + p_3 &= 0; \\
p_2\alpha_2 + p_3\alpha_3 &= 2\alpha_2^2; \\
p_2\alpha_2^2 + p_3\alpha_3^3 &= -4\alpha_2^2(\alpha_2 - 1/3); \\
p_3\beta_{32} &= -2\alpha_2/3; \\
\alpha_2 &= \beta_{21}; \alpha_3 = \beta_{31} + \beta_{32}.
\end{aligned}$$

Эта система имеет бесчисленное множество решений. С помощью двух параметров, которые являются произвольными, можно минимизировать точность аппроксимации. Следовательно, мы имеем двухпараметрическое (по числу свободных параметров) семейство нелинейных методов типа Рунге-Кутты.

Одним из решений приведенной системы нелинейных уравнений будет

$$p_1 = -7/6; p_2 = 4/3; p_3 = -1/6; \alpha_2 = \beta_{21} = 1/2; \alpha_3 = 1; \beta_{31} = -1; \beta_{32} = 2.$$

Тогда формула (3) будет иметь вид

$$\begin{aligned}
y_{n+1} &= y_n + hk_1 - h \frac{6(k_2 - k_1)^2}{7k_1 - 8k_2 + k_3}; \\
k_1 &= f(x_n, y_n); \\
k_2 &= f(x_n + 1/2h, y_n + 1/2hk_1); \\
k_3 &= f(x_{n+h}, y_n - hk_1 + 2hk_2).
\end{aligned}$$

Получим аналогично формулу для нелинейного аналога метода Рунге-Кутты. Для этого будем искать численное решение в виде

$$y_{n+1} = \frac{v_0}{1 - \frac{v_1 h}{1 - \frac{v_2 h}{1 - \frac{v_3 h}{1 - v_4 h}}}}, \quad (4)$$

где

$$\begin{aligned}
v_0 &= y_n; v_1 = \frac{k_1}{y_n}; v_2 = \frac{y_n q_1 - k_1^2}{y_n k_1}; v_3 = \frac{y_n(k_1 q_3 - q_2^2)}{k_1(y_n q_1 - k_1^2)}; \\
v_4 &= \frac{k_1 \{q_1(k_1 q_2 - q_1^2) + q_2(k_1 q_1 - y_n q_2) + q_3(y_n q_1 - k_1^2)\}}{(y_n q_1 - k_1^2)(k_1 q_2 - q_1^2)}; \\
q_1 &= a_1 k_1 + a_2 k_2; q_2 = b_1 k_1 + b_2 k_2 + b_3 k_3; \\
q_3 &= c_1 k_1 + c_2 k_2 + c_3 k_3 + c_4 k_4; \\
k_1 &= f(x_n, y_n); k_2 = f(x_n + \alpha_2 h, y_n + \beta_{21} h k_1); \\
k_3 &= f(x_n + \alpha_3 h, y_n + \beta_{31} h k_1 + \beta_{32} h k_2); \\
k_4 &= f(x_n + \alpha_4 h, y_n + \beta_{41} h k_1 + \beta_{42} h k_2 + \beta_{43} h k_3).
\end{aligned}$$

Для того, чтобы метод (4) имел четвертый порядок точности, необходимо, чтобы неизвестные параметры удовлетворяли следующей системе уравнений:

$$\begin{aligned}
a_1 + a_2 + 0; a_2 \alpha_2 &= 1/2; \\
b_1 + b_2 + b_3 &= 0; \\
b_2 \alpha_2 + b_3 \alpha_3 &= 0; \\
(a_2 + b_2 + c_2) \alpha_2^2 + (b_3 + c_3) \alpha_3^2 + c_4 \alpha_4^2 &= 1/3; \\
c_2 \alpha_2 + c_3 \alpha_3 + c_4 \alpha_4 &= 0; \\
c_1 + c_2 + c_3 &= 0; \\
b_3 \beta_{32} \alpha_2 &= 1/6; \\
c_3 \beta_{32} \alpha_2 + c_4 (\beta_{42} \alpha_2 + \beta_{43} \alpha_3) &= 0; \\
(a_2 + b_2 + c_2) \alpha_2^2 + (b_3 + c_3) + c_4 \alpha_4 &= 1/4; \\
(b_3 + c_3) \alpha_2^2 \beta_{32} + c_4 (\beta_{42} \alpha_2^2 + \beta_{43} \alpha_3^2) &= 1/12; \\
(\beta_3 + c_3) \alpha_2 \beta_{32} \alpha_3 + c_4 \alpha_4 (\beta_{42} \alpha_2 + \beta_{43} \alpha_3) &= 1/8; \\
c_4 \beta_{43} \beta_{32} \alpha_2 &= 1/24.
\end{aligned}$$

В итоге получаем систему 13 уравнений с 15 неизвестными. Выразим все параметры через два свободных параметра α_2, α_3 , в итоге получаем

$$\begin{aligned}
a_1 &= -\frac{1}{2\alpha_2}; \quad a_2 = \frac{1}{2\alpha_2}; \\
b_1 &= \frac{1-2\alpha_2}{3\alpha_2\alpha_3}; \quad b_2 = \frac{2\alpha_2-1}{3\alpha_2(\alpha_3-\alpha_2)}; \quad b_3 = \frac{1-2\alpha_2}{3\alpha_3(\alpha_3-\alpha_2)}; \\
c_1 &= \frac{\alpha_3(3(1+2\alpha_2\alpha_3)-4(\alpha_2+\alpha_3))-(1-2\alpha_2)(3-4\alpha_3)}{12\alpha_2\alpha_3(1-\alpha_3)}; \\
c_2 &= \frac{(1-\alpha_2)(1-2\alpha_2)(3-4\alpha_3)-(\alpha_3-\alpha_2)(3(1+2\alpha_2\alpha_3)-4(\alpha_2+\alpha_3))}{12\alpha_2(1-\alpha_2)(1-\alpha_3)(\alpha_3-\alpha_2)}; \\
c_3 &= \frac{(2\alpha_2-1)(3-4\alpha_3)}{12\alpha_3(\alpha_3-\alpha_2)(1-\alpha_3)}; \quad c_4 = \frac{3(1+2\alpha_2\alpha_3)-4(\alpha_2+\alpha_3)}{12(1-\alpha_2)(1-\alpha_3)}; \\
\beta_{31} &= \frac{2\alpha_2\alpha_3(1-\alpha_2)-\alpha_3(\alpha_3-\alpha_2)}{2\alpha_2(1-2\alpha_2)}; \quad \beta_{32} = \frac{\alpha_3(\alpha_3-\alpha_2)}{2\alpha_2(1-2\alpha_2)}; \\
\beta_{42} &= \frac{2(1-\alpha_3)(1-\alpha_2)(2\alpha_3-1)-(1-\alpha_2)(\alpha_3-\alpha_2)}{2\alpha_2(\alpha_3-\alpha_2)(3(1+2\alpha_2\alpha_3)-4(\alpha_2+\alpha_3))}; \\
\beta_{43} &= \frac{(1-\alpha_3)(1-\alpha_2)(1-2\alpha_2)}{\alpha_3(\alpha_3-\alpha_2)(3(1+2\alpha_2\alpha_3)-4(\alpha_2+\alpha_3))}; \\
\alpha_4 &= 1; \quad \beta_{41} = \alpha_4 - \beta_{42} - \beta_{43}.
\end{aligned}$$

Понятно, что в случае использования предлагаемой методики можно получить метод любой заданной точности. Трудности при построении таких методов, как видно, будут вычислительного характера.

Выбирая соответствующим образом α_2, α_3 , получим двухпараметрическое семейство методов для решения задач Коши.

В настоящее время известны работы, в которых на основе изложенного подхода сделана попытка построения численных методов для решения жестких систем обыкновенных дифференциальных уравнений.

Решение дифференциальных уравнений с частными производными. В первую очередь рассмотрим линейные дифференциальные уравнения с частными производными. Известно, что такое уравнение на основе использования метода конечных разностей можно привести к системе линейных алгебраических уравнений

при условии, что граничные условия также будут линейными. Решения этой системы уравнений дают приближительные решения уравнения в частных производных.

Использование метода ветвящихся целых дробей для решения хорошо приближающихся систем конечноразностных уравнений дают хорошие результаты и для приближенного нахождения решений уравнений с частными производными. Оказалось, что алгоритмы Гаусса, окаймления, ... и некоторые другие могут быть интерпретированы как алгоритмы ветвящихся цепных дробей специального вида. Это открывает новые возможности для построения новых алгоритмов в виде ветвящихся цепных дробей иных типов. Для построения приближенных методов решения нелинейных дифференциальных уравнений с частными производными можно использовать формулу разложения функций в ветвящуюся цепную дробь. Приведем некоторые результаты, основанные на формулах разложения функций в ветвящиеся цепные дроби.

А. А. Самарский в 1980 г. высказал идею применения цепных и ветвящихся цепных дробей для нелинейного приближения производных, имеющих в дифференциальном уравнении. Покажем это на примере.

Как известно, производную $y'(x)$, $y''(x)$ можно линейно заменить, используя формулу Тейлора. Пусть шаг сетки h , обозначим $y_k = y(kh)$, тогда $y_{n+1} = y_n + hy'(nh)$. Поэтому

$$\begin{aligned} y'(nh) &= (y_{n+1} - y_n) / h y'; \\ y_{n+1} &= y_n + hy'(nh) + 1/2 h^2 y''(nh); \\ y_{n-1} &= y_n - hy'(nh) + 1/2 h^2 y''(nh). \end{aligned}$$

Отсюда

$$y''(nh) = \frac{\begin{vmatrix} h & y_{n+1} - y_n \\ -h & y_{n-1} - y_n \end{vmatrix}}{\begin{vmatrix} h & \frac{1}{2}h^2 \\ -h & \frac{1}{2}h^2 \end{vmatrix}} = \frac{y_{n+1} - y_n - y_n + y_{n-1}}{2 \frac{h^2}{2}} = \frac{y_{n+1} - 2y_n + y_{n-1}}{h^2}.$$

Однако можно заменить эти же две производные, используя формулу Тиле:

$$y_{n+1} = y_n + \frac{h}{\frac{1}{y'(nh)}}; \quad y_{n+1} = y_n + \frac{h}{\alpha_1 + \frac{h}{\alpha_2}}; \quad y_{n-1} = y_n + \frac{-h}{\alpha_1 + \frac{-h}{\alpha_2}}.$$

Из последних двух соотношений получаем нелинейное выражение для y'' . Понятно, что на этом пути можно получить приближенное значение производных сколь угодно высокого порядка. Нужно только увеличивать количество звеньев в формулах Тиле и количество уравнений для определения производных.

Рассмотрим численное решение рядом с начальными условиями для уравнений теплопроводности с помощью явной нелинейной сеточной модели

$$L(u) = f,$$

где

$$\begin{aligned} L(u) &= \begin{cases} \frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2}; \\ u(x, 0). \end{cases} \quad \bar{D} = \{-\infty < x < \infty; 0 \leq t \leq T \leq \infty\}; \\ f &= \begin{cases} \varphi(x, t); \\ \psi(x) \end{cases} \quad \bar{D} = \{-\infty < x < \infty; 0 \leq t \leq T \leq \infty\}. \end{aligned}$$

Предполагается, что все известные функции - гладкие. Выберем прямоугольную сетку и заменим область \bar{D} сеточной областью \bar{D}_h , причем координаты узловых точек сетки (x_m, t_n) определяются по формулам

$$\begin{aligned} x_m = mh; t_n = n\tau; m = 0, \pm 1, \pm 2, \dots, h > 0; \\ t_n = n\tau, n = \overline{0, N}; \tau > 0; N\tau \leq T \leq (N+1)\tau, \end{aligned} \quad x_m = mh; t_n = n\tau.$$

Вместо задачи (1) рассмотрим ее разностную модель, когда производная определяется приблизительно не на основе разложения в ряд Тейлора, а на основе разложения функции в цепные дроби.

$$\left. \frac{\partial u}{\partial t} \right|_{(x_m, t_n)} = \frac{u_m^n (u_m^{n+1} - u_m^n)}{u_m^{n+1} \tau} - \frac{\tau}{2} \left. \frac{\partial^2 u}{\partial t^2} \right|_{(x_m, t_n)}; \quad (5)$$

$$\left. \frac{\partial^2 u}{\partial x^2} \right|_{(x_m, t_n)} = \frac{4(u_{m+1}^n - 2u_m^n + u_{m-1}^n)(u_{m+1}^n - u_m^n)(u_m^n - u_{m-1}^n)}{(u_{m+1}^n - u_{m-1}^n)^2 h^2} - \frac{h^2}{12} \left. \frac{\partial^4 u}{\partial x^4} \right|_{(x_m, t_n)}, \quad (6)$$

где $u_m^n = u(x_m, t_n)$.

Формула (5) получается, если воспользоваться разложением функции $u(x, t)$ в цепную дробь без нулевого члена, а формула (6) - если воспользоваться разложением той же функции $u(x, t)$ по формуле Тиле. При такой замене вместо линейной задачи (1) нужно решать систему линейных алгебраических уравнений:

$$\begin{aligned} Lh(u^h) &= f^{(h)}; \\ f^h &= \begin{cases} \Phi_m^n, \\ \Psi_m; \end{cases} \\ u^h &= u(x_m, t_n). \end{aligned}$$

Эффективность предлагаемых схем можно увидеть на текстовых примерах с использованием ЭВМ.

Рассмотрим еще один подход, основанный на использовании теории ветвящихся цепных дробей для решения дифференциальных уравнений в частных производных. Продемонстрируем сущность предлагаемого подхода на решении следующей задачи:

$$\frac{\partial u}{\partial t} = a^2(x) \frac{\partial^2 u}{\partial x^2} + f(x, t), \quad -a \leq x \leq 1$$

- с граничным условием первого рода:

$$\begin{aligned} u(x, t)|_{x=0} &= F_{\text{ГР1}}(t); \\ u(x, t)|_{x=1} &= F_{\text{ГР2}}(t), \end{aligned} \quad (7)$$

- с начальным условием вида $u(x, t)|_{t=0} = E(x)$.

Предположим, что решение $u(x, t)$ задачи (7) может быть приближено аппроксимирующим рядом $u(x, t) = \sum_n V_n(t) \Phi_n(x)$, в котором $\Phi_n(x)$ - известная функция,

например, x^n , x^{2n} , $\cos nx$ и др., имеющие соответствующую степень гладкости по x .

Производные $\frac{\partial u(x,t)}{\partial x}$, $\frac{\partial^2 u(x,t)}{\partial x^2}$ с использованием граничных и начальных условий могут быть разложены в ряды по функциям $u(x, t)$. Например,

$$\frac{\partial^2 u(x,t)}{\partial x^2} = \sum_n V_n(t) \frac{\partial^2 \Phi}{\partial x^2} = \sum_i D_i(x, x_i) u(x,t),$$

где координаты $D_i(x, x_i)$ выражаются через известную функцию $\Phi_n(x)$, их производные - через граничные условия. Явный вид $D_i(x, x_i)$ можно определить разными способами.

Введем ряд опорных точек в область определения функции $u(x, t)$ по x , получим x_0, x_1, \dots, x_n . Далее построим ветвящуюся цепную дробь, интерполирующую $u(x, t)$ по опорным точкам. Для этого используем представление в виде формулы Тиле. Далее величины $u(-1, t)$ и $u(1, t)$ определяются из граничных условий, т.е. $u(0, t) = F_{cp1}(t)$, $u(1, t) = F_{cp2}(t)$.

Частные производные $\frac{\partial u}{\partial t}$, $\frac{\partial^2 u}{\partial x^2}$ определяются непосредственным дифференцированием ветвящейся дроби, получаемой по опорным точкам функции $u(x, t)$ по x .

В результате будем иметь систему обыкновенных дифференциальных уравнений, каждое может быть решено известными методами. Отметим, что, как правило, полученная система уравнений обладает свойствами жесткости.

Решение задач устойчивости систем автоматического управления. Рассмотрим решение задач об устойчивости линейной системы автоматического управления, характеристическое уравнение которой имеет следующий вид:

$$f(S) = S^n + a_{n-1}S^{n-1} + a_{n-2}S^{n-2} + \dots + a_1S + a_0. \quad (8)$$

Как правило, для решения задач об устойчивости линейной системы автоматического управления, характеристическое уравнение которой имеет вид (8), использовались критерии Гаусса или Гурвица, однако их использование связано со значительными вычислительными затратами. Ниже без доказательства приведен ряд утверждений, использование которых позволяет решить задачу об устойчивости с меньшими вычислительными затратами.

Утверждение 1. Пусть $\hat{f}(S) = S^n + a_{n-1}S^{n-1} + a_{n-2}S^{n-2} + \dots + a_0$ - полином с вещественными коэффициентами и пусть

$$g(S) = a_{n-1}S^{n-1} + a_{n-3}S^{n-3} + a_{n-5}S^{n-5} + \dots$$

Тогда $\hat{f}(S)$ является устойчивым полиномом в том и только в том случае, когда проверяемую дробь $t = g/\hat{f}$ можно представить как неправильную дробь вида

$$t(S) = \frac{g(S)}{\hat{f}(S)} = \frac{1}{1+d_1S} + \frac{1}{d_2S} + \frac{1}{d_3S} + \dots + \frac{1}{d_nS},$$

где $d_i > 0$, $i = \overline{1, n}$

Пример 1.

$$\begin{aligned}\hat{f}(S) &= S^4 + 5S^3 + 10S^2 + 10S + 4; \\ g(S) &= 5S^3 + 10S; \\ t(S) &= \frac{1}{\hat{f}(S)/g(S)} = \frac{1}{1+h_1(S)}; \\ h_1(S) &= (1/5)S + (8S^2 + 4)/(5S^3 + 10S); \\ h_1(S) &= (1/5)S + 1/h_2(S),\end{aligned}$$

где после деления $5S^3 + 10S$ на $8S^2 + 4$ получаем

$$h_2(S) = (5/8)S + (15/2)S / (8S^2 + 4),$$

следовательно, $h_2(S) = (5/8)S + 1/h_3(S)$, где после деления $8S^2 + 4$ на $15S/2$ получаем

$$h_3(S) = (16/15)/S + 1/[(15/8)S].$$

Комбинируя предыдущие результаты, получаем представление в виде непрерывной дроби

$$t(S) = \frac{1}{1+(1/5)S} + \frac{1}{(5/8)S} + \frac{1}{(16/5)S} + \frac{1}{(15/8)S}.$$

Из утверждения вытекает, что $\hat{f}(S)$ устойчивый полином.

Пример 2.

При каком значении вещественной константы

$$\hat{f}(S) = S^4 + 5S^3 + 10S^2 + 10S + c$$

будет устойчивым. Следуя процедуре, использованной в примере 1, получаем

$$t(S) = \frac{1}{1+(1/5)S} + \frac{1}{(5/8)S} + \frac{1}{d_3S} + \frac{1}{d_4S},$$

где $d_3 = 8/[10 - (5/8)c]$; $d_4 = [10 - (5/8)c]$, следовательно, $d_3 > 0$ тогда, когда $c > 16$, а d_4 больше нуля, когда $c > 0$. Таким образом, можно сделать вывод, что $\hat{f}(S)$ будет устойчивым полиномом в том случае, если $0 < c < 16$. Очевидно, что для устойчивости полинома $\hat{f}(S)$ необходимо существование, с одной стороны, его представления в виде:

$$\frac{1}{1+d_1S} + \frac{1}{d_2S} + \frac{1}{d_3S} + \dots + \frac{1}{d_mS},$$

а с другой стороны, обязательно выполнение условия $m = n$, где n - степень полинома $\hat{f}(S)$. Приведем еще два утверждения, так же без доказательства.

Утверждение 2. Пусть $f(S) = a_0 + a_1S + a_2S^2 + \dots + a_nS^n$, $a_0 \neq 0$, вещественный полином степени n и пусть

$$D_1 = a_1; \quad D_2 = \begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix}; \quad D_3 = \begin{vmatrix} a_1 & a_3 & a_5 \\ a_0 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{vmatrix},$$

где D_k - определитель ($k \times k$), $a_m = 0$ при $m > n$ и $D_0 = 1$; $D_1 = 1/a_0$; $D_2 = 1/a_0^2$.

Тогда f является устойчивым полиномом в том и только в том случае, когда определители удовлетворяют условиям:

$$D_{2k} > 0, \quad k = \overline{1, n/2};$$

$$\text{sign} D_{2k+1} = \text{sign} a_0 k = 0, \quad [(n-1)/2].$$

Отметим, что с вычислительной точки зрения значительно проще установить устойчивость полинома $f(S)$ методом деления (Утверждение 1) или путем вычисления определителей. Утверждение 2 носит название теоремы Гурвица.

Тема 14 ЭЛЕМЕНТЫ ТЕОРИИ ВЕТВЯЩИХСЯ ДРОБЕЙ

Определение ветвящейся цепной дроби. Ветвящаяся цепная дробь является новым математическим понятием, возникшим на основе геометризации понятия обычной цепной дроби. Так, цепную дробь

$$\frac{a_1}{b_1 + \frac{a_2}{b_2 + \frac{a_3}{b_3 + \dots + \frac{a_n}{b_n + \dots}}}} \quad (1)$$

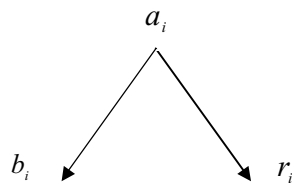


Рис. 1

состоящую из звеньев

$$\frac{a_i}{b_i + r_i}, \quad i = \overline{1, \infty},$$

можно изобразить и в геометрической форме, если каждому такому звену поставить в соответствие фигуру, изображенную на рис. 1, где число a_i делится на сумму двух чисел $(b_i + r_i)$; на основе этого легко получить геометрическое изображение цепной дроби (1), которая представляет собой ориентированный граф типа дерева (рис. 2).

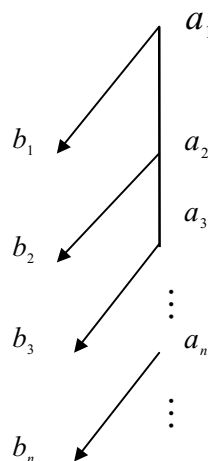


Рис. 2

Возникает вопрос, какой аналитический математический объект отвечает в предлагаемой интерпретации дереву общего вида.

Рассмотрим геометрически ориентированное дерево, представленное на рис. 3. Поставим

в соответствии каждой концевой вершине $B_{ij}^{(k)}$ дерева (Д) оператор $b_{ij}^{(k)}$, а каждой концевой вершине $A_i^{(k)}$ дерева (Д) оператор $a_i^{(k)}$. В частности, можно поставить в соответствие вершинам числа, функции, независимые переменные или другие математические объекты.

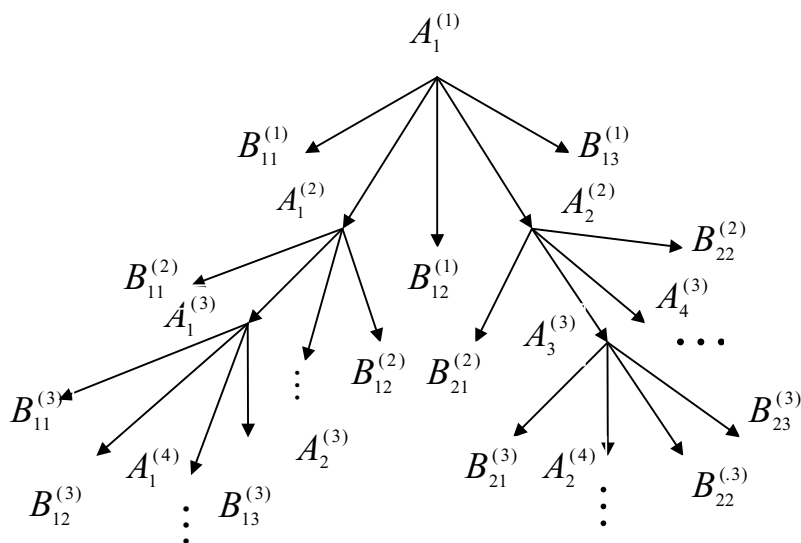


Рис. 3

Далее ориентируем дерево (Д) при помощи стрелок, которые выходят из каждой неконцевой вершины. Вершину $A_1^{(1)}$ принимаем за корень дерева. Тогда в каждую вершину $A_j^{(k)}, B_{ij}^{(k)}$ входит только одна стрелка, а из каждой вершины $A_i^{(k)}$ выходит несколько стрелок, из которых m_{ik} штук идут к концевым вершинам $B_{ij}^{(k)}$, n_{ik} штук к следующим неконцевым вершинам дерева (Д). Поэтому каждая из вершин $A_i^{(k)}$ является точкой ветвления дерева (Д) с числом ветвлений n_{ik} . Корень дерева (Д) и совокупность стрелок, подходящих к вершинам $B_{ij}^{(k)}$, образуют первый этаж дерева (Д). Второй этаж дерева (Д) образуют вершины $A_j^{(2)}$ и стрелки, которые выходят из них в направлении вершин $B_{ij}^{(2)}$ и т.д. Поставим в соответствие вершине $A_i^{(k)}$ и совокупности стрелок, которые из нее выходят к вершине $B_{ij}^{(k)}$, дробь

$$\frac{a_i^{(k)}}{b_{i1}^{(k)} + b_{i2}^{(k)} + b_{i3}^{(k)} + \dots + b_{imik}^{(k)}}; \tag{2}$$

тогда (если положив для простоты $n_{ik} = n, m_{ik} = m$) дереву указанного вида соответствует ветвящаяся цепная дробь

$$\begin{aligned} & \frac{a_1^{(1)}}{b_{11}^{(1)} + b_{12}^{(1)} + \dots + b_{1m}^{(1)} + \frac{a_1^{(2)}}{b_{11}^{(2)} + b_{12}^{(2)} + \dots + b_{1m}^{(2)} + \frac{a_1^{(3)}}{\vdots} + \dots + \frac{a_4^{(3)}}{\vdots}} + \dots \\ & \rightarrow \frac{a_4^{(2)}}{b_{41}^{(2)} + \dots + b_{4m}^{(2)} + \frac{a_{n(n-1)+1}}{\dots} + \dots + \frac{\dots}{\dots}} \end{aligned} \quad (3)$$

с числом ветвей n .

Обратно, каждой ветвящейся цепной дроби соответствует в предлагаемой геометрической интерпретации единственное дерево.

Пример. Построим ветвящуюся дробь для дерева, представленного на рис. 4:

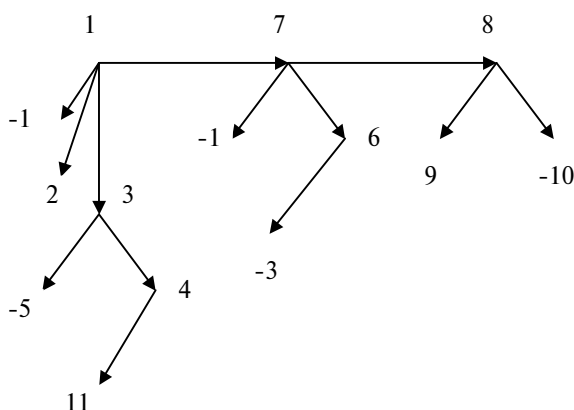


рис. 4

$$\frac{1}{-1+2+\frac{3}{-5+\frac{4}{11}}+\frac{7}{-1+\frac{6}{-3}+\frac{8}{9-10}}}$$

Обычная цепная дробь является частным случаем ветвящейся цепной дроби, когда число ветвей равно единице.

Ветвящуюся цепную дробь можно определить и по другому. Известно, что при изучении цепных дробей с комплексными компонентами очень удобно их рассматривать как композицию дробно-линейных отображений. Ветвящаяся цепная дробь с N ветками ветвления, которая является естественным обобщением обычной цепной дроби, может быть определена как композиция N -мерных дробно-линейных отображений в поле комплексных чисел.

Пусть $b_0, b_{i_1 i_2 \dots i_k}, a_{i_1 i_2 \dots i_k}$ - комплексные числа, $z_{i_1 i_2 \dots i_k}$ - комплексные переменные и

$$T_0(z_1, z_2, \dots, z_n) = b_0 + z_1 + z_2 + \dots + z_n;$$

$$T_{i_1}(z_{i_1 1}, z_{i_1 2}, \dots, z_{i_1 N}) = \frac{a_{i_1}}{b_{i_1} + z_{i_1 1} + z_{i_1 2} + \dots + z_{i_1 N}};$$

$$T_{i_1 i_2 \dots i_k}(z_{i_1 i_2 \dots i_k 1}, \dots, z_{i_1 i_2 \dots i_k N}) = \frac{a_{i_1 i_2 \dots i_k}}{b_{i_1 i_2 \dots i_k} + z_{i_1 i_2 \dots i_k}}, \quad (4)$$

$(i_k = \overline{1, N}, k = \overline{1, \infty})$ - некоторая совокупность N -мерных дробно-линейных отображений, взятых, вообще говоря, в разных точках. Тогда рассматриваемые композиции этих отображений приводят к следующей записи ветвящейся цепной дроби с N витками:

$$b_0 + \sum_{i_1=1}^n \frac{a_{i_1}}{b_{i_1} + \sum_{i_2=1}^n \frac{a_{i_1 i_2}}{b_{i_1 i_2} + \dots + \sum_{i_k=1}^n \frac{a_{i_1 i_2 \dots i_k}}{b_{i_1 i_2 \dots i_k} + \dots}}} \quad (5)$$

где $a_{i_1 i_2 \dots i_k}, b_0, b_{i_1 i_2 \dots i_k}, i_k = \overline{1, N}, k = \overline{1, \infty}$ переменные; в частности, они могут быть числами, матрицами, операторами, элементами абстрактных пространств, в которых введены операции сложения и деления. Компоненты ветвящейся цепной дроби $a_{i_1 i_2 \dots i_k}$ называются частными делителями, $b_{i_1 i_2 \dots i_k}$ - частными знаменателями.

Если число компонентов конечно, то дробь называется конечной, в противном случае бесконечной. Впредь, если это не будет оговорено, будем рассматривать ветвящиеся цепные дроби, компонентами которых являются действительные или комплексные числа.

Дробь $a_{i_1 i_2 \dots i_k} / b_{i_1 i_2 \dots i_k}$ - называют звеном k -го порядка ветвящейся цепной дроби с индексом $i_1 i_2 \dots i_k$. Звено $a_{i_1 i_2 \dots i_k} / b_{i_1 i_2 \dots i_k}$ называется оценочным звеном k -го порядка с индексом $i_1 i_2 \dots i_k$, если все следующие части числителя $a_{i_1 i_2 \dots i_k}, i_{k+1} = \overline{1, N}$ равны нулю при фиксированном наборе чисел $i_1 i_2 \dots i_k$, т.е. ветвящаяся цепная дробь на k -м элементе на ветке с индексами $i_1 i_2 \dots i_k$ обрывается.

Ветвящуюся цепную дробь с N ветками ветвления можно компактно записать в виде

$$b_0 + \sum_{i_1=1}^N \frac{a_{i_1}}{b_{i_1}} + \sum_{i_2=1}^N \frac{a_{i_1 i_2}}{b_{i_1 i_2}} + \dots + \sum_{i_k=1}^N \frac{a_{i_1 i_2 \dots i_k}}{b_{i_1 i_2 \dots i_k}} + \dots \quad (6)$$

или в виде

$$b_0 + \sum_{i_k=1}^n \frac{a_{i_1 i_2 \dots i_k}}{b_{i_1 i_2 \dots i_k}}. \quad (7)$$

Формула разности подходящих дробей. Свойство вилки. При изучении вопросов сходимости и устойчивости ветвящихся цепных дробей важное значение имеет формула разности двух подходящих дробей. В теории цепных дробей для решения аналогичных вопросов широко используются линейные рекуррентные формулы второго порядка для числителей и знаменателей подходящих дробей. Однако для ветвящихся цепных дробей такие формулы, вообще говоря, не имеют места. Простых и удобных для исследования рекуррентных формул для числителей и знаменателей подходящих дробей пока построить не удастся. Введем рекуррентные соотношения

$$a_{i_1 i_2 \dots i_p}^{(p)} = b_{i_1 i_2 \dots i_p};$$

$$a_{i_1 i_2 \dots i_k}^{(p)} = b_{i_1 i_2 \dots i_k} + \sum_{i_{k+1}}^N \frac{a_{i_1 i_2 \dots i_{k+1}}^{(p)}}{Q_{i_1 i_2 \dots i_{k+1}}^{(p)}}, \quad i_k = \overline{1, N}, \quad k = \overline{p-1, 1}, \quad p = \overline{1, \infty},$$

на основе которых можно получить

$$\frac{P_n}{Q_n} - \frac{P_m}{Q_m} = (-1)^m \sum_{i_1 i_2 \dots i_{m+1}=1}^n \frac{\prod_{k=1}^{m+1} a_{i_1 i_2 \dots i_k}}{\prod_{k=1}^{m+1} Q_{i_1 i_2 \dots i_k}^{(m)} \prod_{k=1}^m Q_{i_1 i_2 \dots i_k}^{(m)}}, \quad (8)$$

$Q_{i_1 i_2 \dots i_k}^{(s)}$ - рациональная функция от компонент ветвящейся цепной дроби.

Свойствовилки сформулировано в виде следующей теоремы.

Теорема. Если у ветвящейся цепной дроби (7) все компоненты положительны, то для подходящих дробей (7) выполняются свойствавилки:

- 1) $\frac{P_1}{Q_1} > \frac{P_3}{Q_3} > \dots > \frac{P_{2i-1}}{Q_{2i-1}} > \dots;$
- 2) $\frac{P_0}{Q_0} < \frac{P_2}{Q_2} < \dots < \frac{P_{2k}}{Q_{2k}} < \dots;$
- 3) для произвольных i, k $\frac{P_{2i-1}}{Q_{2i-1}} > \frac{P_{2k}}{Q_{2k}}.$

На этом мы заканчиваем рассмотрение основ теории непрерывных дробей, далее будет показана возможность использования теории непрерывных дробей для решения ряда задач вычислительной математики.

Тема 15 ПРИБЛИЖЕНИЯ ФУНКЦИЙ МНОГИХ ПЕРЕМЕННЫХ ВЕТВЯЩИМИСЯ ЦЕПНЫМИ ДРОБЯМИ

Частные обратные разности. Интерполяционная формула с линейными частными числителями.

Пусть для действительной функции N переменных $f(x)$, где $x = (x_1, x_2, \dots, x_N)$ заданы ее значения в точках

$$x_{k_n}^{n_i} = (x_1^n, x_2^n, \dots, x_{k_n}^n, x_{k_{n+1}}^n, \dots, x_N^n) \in R^N, \quad k_n = \overline{1, N}, \quad n, i = \overline{0, m},$$

не лежащие одновременно в подпространстве $R^K (K < N)$, применим $x_{k_n}^{n_i} = x^n$, если $i = n$ или $k_n = N$, $x = (x_1, x_2, \dots, x_N)$ $x_{k_n}^{n_i} = x^n$ - произвольная, но фиксированная точка из R^N . Все точки $x_{k_n}^{n_i}$ разобьем на две группы: точки $x^i = (x_1^i, \dots, x_N^i)$, $i = \overline{1, m}$ назовем основными, а все остальное - промежуточными точками.

Частными разностями первого порядка $x_{k_1} v_1(x^0, x)$ по аргументу $x_{k_1} (k_1 = \overline{1, N})$ называют соотношения вида

$$x_{k_1} v_1(x^0, x^1) = (x_{k_1}^0 - x_{k_1}^1) \{k_1 f(x^0, x^1)\}^{-1}, \quad (1)$$

где

$$k_1 f(x^0, x^1) = \frac{1}{2} \{ [v_0(x_{k_1-1}^{10}) - v_0(x_{k_1}^{10})] - [v_0(x_{k_1-1}^{01}) - v_0(x_{k_1}^{01})] \}, \quad (2)$$

$$x_{k_0} v_0(x) = v_0(x) = f(x).$$

Поскольку $k_1 f(x^0, x^1) = -k_1(x^1, x^0)$ при произвольном $k_1 (k_1 = \overline{1, N})$, то частные обратные разности первого порядка $x_{k_1} v_1(x^0, x^1)$ функции $f(x)$ являются симметричные функции точек x^0, x^1 .

Например, для функции $f(x, y)$ частные обратные разности первого порядка примут вид

$$\begin{aligned} x v_1(x^0, x^1) &= \frac{x_0 - x_1}{\frac{1}{2} \{f(x_0, y_0) - f(x_1, y_0) - f(x_1, y_1) + f(x_0, y_1)\}}; \\ y v_1(x^0, x^1) &= \frac{y_0 - y_1}{\frac{1}{2} \{f(x_1, y_0) - f(x_1, y_1) - f(x_0, y_1) + f(x_0, y_0)\}}. \end{aligned} \quad (3)$$

Полезно отметить, что

$$\sum_{k_1=1}^n k_1 f(x^0, x^1) \equiv f(x^0) - f(x^1) \quad (4)$$

Частные обратные разности второго порядка $x_{k_2} x_{k_1} v_2(x^0, x^2, x^1)$ по аргументам $x_{k_2}, x_{k_1}, k_i = \overline{1, N}, i = 1, 2$ называется соотношением вида:

$$x_{k_2} x_{k_1} v_2(x^0, x^2, x^1) = (x_{k_2}^0 - x_{k_2}^2) \{k_2 k_1 f(x^0, x^2, x^1)\}^{-1}, \quad (5)$$

где

$$\begin{aligned} k_2 k_1 f(x^0, x^2, x^1) &= \frac{1}{2} \{x_{k_1} v_1(x_{k_2}^{20}, x^1) - x_{k_1} v_1(x_{k_2}^{20}, x^1) - \\ &- x_{k_1} v_1(x_{k_2}^{02}, x^1) + x_{k_1} v_1(x_{k_2}^{02}, x^1)\} \end{aligned} \quad (6)$$

так как $k_2 k_1 f(x^0, x^2, x^1) = -k_2 k_1 f(x^2, x^0, x^1)$ при произвольных k_2, k_1 , то частные обратные разности второго порядка удовлетворяют соотношению

$$x_{k_2} x_{k_1} v_2(x^0, x^2, x^1) = x_{k_2} x_{k_1} v_2(x^2, x^0, x^1). \quad (7)$$

Для функции двух аргументов $f(x, y)$ частные обратные разности второго порядка определяются из соотношений

$$\begin{aligned} x x v_2(x^0, x^2, x^1) &= (x^0 - x^2) 2 \{x v_1(x^0, x^1) - x v_1(x_1^{20}, x^1) - \\ &- x v_1(x^2, x^1) + x v_1(x_1^{02}, x^1)\}^{-1} \quad (8) \\ x y v_2(x^0, x^2, x^1) &= (x_0 - x_2) 2 \{y v_1(x^0, x^1) - y v_1(x_1^{20}, x^1) - \\ &- y v_1(x^2, x^1) + y v_1(x_1^{02}, x^1)\}^{-1}; \\ y y v_2(x^0, x^2, x^1) &= (y_0 - y_2) 2 \{y v_1(x_1^{02}, x^1) - \\ &- y v_1(x^2, x^1) - y v_1(x_1^{02}, x^1) + y v_1(x^0, x^1)\}^{-1} \quad (9); \end{aligned}$$

$$\begin{aligned} x x v_2(x^0, x^2, x^1) &= (y_0 - y_2) 2 \{x v_1(x_1^{20}, x^1) - x v_1(x_1^{20}, x^1) - \\ &- x v_1(x_1^{02}, x^1) + x v_1(x^0, x^1)\}^{-1}, \end{aligned}$$

где $x_1^{20} = (x_2, y_0); x_1^{02} = (x_0, y_2)$.

Как и в случае обычных разностей первого порядка справедлива формула

$$\sum_{k_2=1}^N k_2 k_1 f(x^0, x^2, x^1) \equiv x_{k_1} v_1(x^0, x^1) - x_{k_1} v_1(x^2, x^1). \quad (10)$$

Такие разности n -го порядка ($n \geq 2$) называются соотношениями вида:

$$x_{k_n} \dots x_{k_1} v_n(x^0, x^n, x^{n-1}, \dots, x^1) = (x_{k_n}^0 - x_{k_n}^n) \{k_n \dots k_1 f(x^0, x^n, x^{n-1}, \dots, x^1)\}^{-1}; \quad (11)$$

где

$$\begin{aligned}
 k_n \dots k_1 f(x^0, x^n, \dots, x^1) = & \frac{1}{2} \left\{ x_{k_{n-1}} \dots x_{k_1} v_{n-1} \left(x_{k_{n-1}}^{n0}, x^{n-1}, \dots, x^1 \right) - \right. \\
 & - x_{k_{n-1}} \dots x_{k_1} v_{n-1} \left(x_{k_n}^{n0}, x^{n-1}, \dots, x^1 \right) - x_{k_{n-1}} \dots x_{k_1} v_{n-1} \left(x_{k_{n-1}}^{0n}, x^{n-1}, \dots, x^1 \right) + \\
 & \left. + x_{k_{n-1}} \dots x_{k_1} v_{n-1} \left(x_{k_n}^{0n}, x^{n-1}, \dots, x^1 \right) \right\}; \quad (12)
 \end{aligned}$$

Отметим также, что

$$k_n \dots k_1 f(x^0, x^n, x^{n-1}, \dots, x^1) = -k_n \dots k_1 f(x^n, x^0, x^{n-1}, \dots, x^1)$$

Приведем в качестве примеров формулы для разложения

$$\begin{aligned}
 f(x) = f(x^1) + \sum_{k_1=1}^N \frac{x_{k_1} - x_{k_1}^1}{x_{k_1} v_1(x^2, x^1) + \sum_{k_2=1}^N \frac{x_{k_2} - x_{k_2}^2}{x_{k_2} x_{k_1} v_2(x^3, x^2, x^1) + \dots} \quad (13) \\
 \vdots \\
 \sum_{k_m=1}^N \frac{x_{k_m} - x_{k_m}^m}{x_{k_m} \dots x_{k_1} v_m(x, x^m, x^{m-1}, \dots, x^1)}
 \end{aligned}$$

Для функций двух переменных $f(x, y)$ разложение имеет вид

$$\begin{aligned}
 f(x, y) = f(x_1, y_1) + \frac{x - x_1}{x v_1(x^2, x^1) + \frac{x - x_2}{x x v_2(x^3, x^2, x^1) + \dots} + \frac{y - y_2}{y x v_2(x^3, x^2, x^1) + \dots} \\
 + \frac{y - y_1}{y x_1(x^2, x^1) + \frac{x - x_2}{x y v_2(x^3, x^2, x^1) + \dots} + \frac{y - y_2}{y y v_2(x^3, x^2, x^1) + \dots}
 \end{aligned}$$

Пример. Пусть заданы значения функции $f(x, y) = \frac{1}{x^2 + y^2 - 1,5}$ в основных точках (0, 0), (2, 1), (1, 2) и в промежуточных (0,1), (0,2), (2,0), (2,2), (1,0), (1,1). Используя только основные точки, получим интерполяционную дробь

$$-\frac{2}{3} + \frac{x}{\frac{105}{88} + \frac{x-2}{1672/5601} + \frac{y-1}{209/162}} + \frac{y}{-\frac{105}{76} + \frac{x-2}{209/162} + \frac{y-1}{1672/5601}}$$

Вместо значения функции $f(x, y) = \frac{1}{x^2 + y^2 - 1,5}$ в точке (112, 112) с точностью до 10^{-5} по интерполяционной дроби, получаем -0,90368, при приближении полином Лагранжа 4-й

степени (построение по тем же основным точкам и промежуточным) $f(112, 112) \approx -0,18242$.

Точное значение функции $f(x, y) = \frac{1}{x^2 + y^2 - 1,5}$ в точке $(112, 112)$ равно -1 .

Интерполяционная формула с полиномиальными компонентами. Введем в рассмотрение следующие обозначения:

$$\rho_{00}(x, y) = f(x, y);$$

$$\rho_{10}(x_1, x, y_1) = \frac{x - x_1}{f(x, y_1) - f(x_1, y_1)};$$

$$\rho_{01}(x_1, y_1, y) = \frac{y - y_1}{f(x_1, y) - f(x_1, y_1)};$$

$$\rho_{11}(x_1, x, y_1, y) = (x - x_1)(y - y_1) \{f(x, y) - f(x_1, y) + f(x_1, y_1) - f(x, y_1)\}^{-1};$$

$$\rho_{mk}(x_1, \dots, x_m, x, y_1, \dots, y_k, y) = \begin{cases} (x - x_m) \{ \rho_{m-1k}(x_1, \dots, x_{m-1}, x, y_1, \dots, y_k, y) - \\ - \rho_{m-1k}(x_1, \dots, x_m, y_1, \dots, y_k, y) \}^{-1}, & m > k; \\ (y - y_k) \{ \rho_{mk-1}(x_1, \dots, x_m, x, y_1, \dots, y_{k-1}, y) - \\ - \rho_{mk-1}(x_1, \dots, x_m, x, y_1, \dots, y_k) \}^{-1}, & k > m; \\ (x - x_m)(y - y_m) \{ \rho_{m-1m-1}(x_1, \dots, x_{m-1}, y_1, \dots, y_{m-1}, y) - \\ - \rho_{m-1m-1}(x_1, \dots, x_m, y_1, \dots, y_{m-1}, y) + \\ + \rho_{m-1m-1}(x_1, \dots, x_m, y_1, \dots, y_m) - \\ - \rho_{m-1m-1}(x_1, \dots, x_{m-1}, x, y_1, \dots, y_m) \}^{-1}, & m = k = \overline{0, \infty}; \\ \rho_{-1k} = \rho_{m-1} = \rho_{-1-1} = 0. \end{cases}$$

Рассмотрим пример на основе приведенных соотношений:

$$f(x, y_1) = f(x_1, y_1) + \frac{x - x_1}{\rho_{10}(x_1, x, y_1)};$$

$$f(x_1, y) = f(x_1, y_1) + \frac{y - y_1}{\rho_{01}(x_1, y_1, y)};$$

$$f(x, y) - f(x_1, y) - f(x, y_1) = -f(x_1, y_1) + \frac{(x - x_1)(y - y_1)}{\rho_{11}(x_1, x, y_1, y)}.$$

Сложив равенства, получим

$$f(x, y) = f(x_1, y_1) + \frac{x - x_1}{\rho_{10}(x_1, x, y_1)} + \frac{y - y_1}{\rho_{01}(x_1, y_1, y)} + \frac{(x - x_1)(y - y_1)}{\rho_{11}(x_1, x, y_1, y)};$$

$$\rho_{10}(x_1, x, y_1) = \rho_{10}(x_1, x_2, y_1) + \frac{x - x_2}{\rho_{20}(x_1, x_2, x, y_1)};$$

$$\rho_{01}(x_1, y_1, y) = \rho_{01}(x_1, y_1, y_2) + \frac{y - y_2}{\rho_{02}(x_1, y_1, y_2, y)};$$

$$\begin{aligned} \rho_{11}(x_1, x, y_1, y) - \rho_{11}(x_1, x, y_1, y_2) - \rho_{11}(x_1, x_2, y_1, y) = \\ = -\rho_{11}(x_1, x_2, y_1, y_2) + \frac{(x - x_2)(y - y_2)}{\rho_{22}(x_1, x_2, x, y_1, y_2, y)}; \end{aligned}$$

$$\rho_{11}(x_1, x, y_1, y_2) = \rho_{11}(x_1, x_2, y_1, y_2) + \frac{x - x_2}{\rho_{21}(x_1, x_2, x, y_1, y_2)};$$

$$\rho_{11}(x_1, x_2, y_1, y) = \rho_{11}(x_1, x_2, y_1, y_2) + \frac{y - y_2}{\rho_{12}(x_1, x_2, y_1, y_2, y)}.$$

Пусть $f(x, y) = \frac{1+x}{1-y}$ задана таблицей

$x \backslash y$	0	1/2	3/2	2
0	1	2	-2	-1
1/2	3/2	3	-3	-3/2
3/2	5/2	5	-5	-5/2
2	3	6	-6	-3

Тогда интерполяционная дробь, построенная на основе приведенных соотношений, имеет вид

$$\frac{1+x}{1-y} \approx 1+x + \frac{y}{1-y} + \frac{xy}{1-y}.$$

Разложение функции многих переменных в ветвящиеся целые дроби. При определении частных обратных разностей предполагалось, что все точки $x^i, i = \overline{1, m}$ различны. Если предположить, что две или три, а может и больше точек совпадают, то получится последовательность частных обратных разностей с повторяющимися значениями аргумента.

Особый интерес представляет случай, когда все точки x^i совпадают, в таком случае для частных обратных разностей с повторяющимися значениями аргумента введем следующее определение.

Частной обратной производной m -ого порядка функции $f(x)$ называется выражение вида $x_{k_m}, \dots, x_{k_1}^{(m)} f(x) = \lim_{x^i \rightarrow x} \{x_{k_m} \dots x_{k_1} v_m(x, x^m, \dots, x^1) + x_{k_{m-2}} \dots x_{k_1} v_{m-2}(x^{m-1}, \dots, x^1) + \dots\}$ где $k_i = \overline{1, N}; i = \overline{1, m}, x_{k_0} v_0(x) = f(x), x_{k_{m-2}} \dots x_{k_1} v_{m-2}(x^{m-1}, \dots, x^1) = 0; m = 1.$

Частные обратные производные первого порядка в окрестности точки x , получим

$$x_{k_1} \setminus f(x) = \lim_{x^1 \rightarrow x} x_{k_1} v_1(x, x^1) = \lim_{x^1 \rightarrow x} \frac{1}{\frac{k_1 f(x, x^1) - f(x)}{x_0 - x_{k_1}}} = \frac{1}{f'_{x_{k_1}}(x)};$$

$$x_{k_1} \setminus f(x) = \frac{1}{f'_{x_{k_1}}(x)}, \quad k_1 = \overline{1, N}.$$

В случае функции двух переменных можно записать

$$x \setminus f(x, y) = \lim_{\substack{x_1 \rightarrow x \\ y_1 \rightarrow y}} \frac{2}{\frac{f(x, y) - f(x_1, y)}{x - x_1} + \frac{f(x, y_1) - f(x_1, y_1)}{x - x_1}} = \frac{1}{f'_x(x, y)};$$

$$f(x) = f(x^0) + \sum_{k_1=1}^N \frac{x_{k_1} - x_{k_1}^0}{x_{k_1} \setminus f(x)_{x=x_0}} + \dots$$

Для функции двух переменных в окрестности точки (x_0, y_0) разложение имеет вид

$$f(x, y) = f(x_0, y_0) + \frac{x - x_0}{\sqrt{x f(x_0, y_0) + \frac{x - x_0}{2x \sqrt{[x f(x, y)]_{x=x_0, y=y_0}} + \dots}} + \frac{y - y_0}{2y \sqrt{[x \sqrt{f(x, y)]_{x=x_0, y=y_0}} + \dots}} + \frac{y - y_0}{\sqrt{y f(x_0, y_0) + \frac{x - x_0}{2x \sqrt{[y f(x, y)]_{x=x_0, y=y_0}} + \dots}} + \frac{y - y_0}{2y \sqrt{[x \sqrt{f(x, y)]_{x=x_0, y=y_0}} + \dots}}$$

Рассмотрим пример разложения функции $\sqrt{x^2 + y}$ в окрестности точки (1,3) в ветвящуюся цепную дробь

$$\sqrt{x^2 + y} = 2 + \frac{x - 1}{2 + \frac{x - 1}{-\frac{4}{3} + \frac{x - 1}{6 + \dots} + \frac{y - 3}{19} + \dots}} + \frac{y - 3}{8 + \frac{x - 1}{2} + \frac{y - 3}{12} + \dots} + \frac{y - 3}{4 + \frac{x - 1}{2 + \frac{x - 1}{-3 + \dots} + \frac{y - 3}{5 + \dots}} + \frac{y - 3}{4 + \frac{x - 1}{2 + \dots} + \frac{y - 3}{4 + \dots}}$$

Аппроксимация Паде. Пусть задан ряд

$$f(z) = \sum_{i=0}^{\infty} c_i z^i, \quad (14)$$

представляющий функцию $f(z)$. Разложение (14) является исходным пунктом любого анализа, использующего аппроксимацию Паде. Всюду в дальнейшем через $c_i, i = 0, \infty$ определяются коэффициенты ряда, а через $f(z)$ - соответствующая функция. Аппроксимация Паде - это рациональная функция вида

$$[L/M] = \frac{a_0 + a_1 z + \dots + a_L z^L}{b_0 + b_1 z + \dots + b_M z^M}, \quad (15)$$

разложения которой в ряд Тейлора в узле совпадает с разложением (14) до тех пор, пока это возможно. Отметим, что функция (15) имеет $(L + 1)$ коэффициентов в числителе и $(M + 1)$ - в знаменателе. Весь выбор коэффициентов определяется с точностью до общего множителя, для определенности можно предположить, что $b_0 = 1$, таким образом, мы имеем $(L + 1)$ свободных параметров в числителе и M - в знаменателе формулы (15) вместо $(L + M + 1)$ свободных параметров. Это означает, что в общем случае коэффициенты Тейлоровского разложения функции $[L/M]$ при системах $1, z, z^2, \dots, z^{L+M}$ должны совпадать с соответствующими коэффициентами ряда (14), другими словами,

$$\sum_{i=0}^{\infty} c_i z^i = \frac{a_0 + a_1 z + \dots + a_L z^L}{b_0 + b_1 z + \dots + b_M z^M} + 0(z^{L+M+1}) \quad (16)$$

Рассмотрим примеры:

$$\begin{aligned}
f(z) &= 1 - \frac{1}{2}z + \frac{1}{3}z^2 + \dots + \frac{(-1)^n}{n+1}z^n; \\
[1/0] &= 1 - \frac{1}{2}z = f(z) + 0(z^2); \\
[0/1] &= \frac{1}{1 + \frac{1}{2}z} = f(z) + 0(z^2); \\
[1/1] &= \frac{1 + \frac{1}{6}z}{1 + \frac{2}{3}z} = f(z) + 0(z^3)
\end{aligned}$$

Умножим (16) на знаменатель дроби; находим, что

$$(b_0 + b_1z + \dots + b_mz^m)(c_0 + c_1z + \dots) = a_0 + a_1z + \dots + a_Lz^L + 0(z^{L+M+1}) \quad (17)$$

Сравнивая коэффициенты при $z^{L+1}, z^{L+2}, \dots, z^{L+M}$, получим равенства:

$$\begin{aligned}
b_Mc_{L+M+1} + b_{M-1}c_{L-M+2} + \dots + b_0c_{L+1} &= 0; \\
b_Mc_{L+M+2} + b_{M-1}c_{L-M+3} + \dots + b_0c_{L+2} &= 0; \\
\vdots & \\
b_Mc_L + b_{M-1}c_{L+1} + \dots + b_0c_{L+M} &= 0.
\end{aligned} \quad (18)$$

Для полноты предположим $c_j = 0$ при $j < 0$. С учетом соотношения $b_0 = 1$ равенства (18) можно представить в виде системы M линейных уравнений с M неизвестными коэффициентами знаменателя:

$$\begin{pmatrix} C_{L+M+1} & C_{L-M+2} & C_{L-M+3} & \dots & C_L \\ C_{L-M+2} & C_{L-M+3} & C_{L-M+4} & \dots & C_{L+1} \\ C_{L-M+3} & C_{L-M+4} & C_{L-M+5} & \dots & C_{L+2} \\ \dots & \dots & \dots & \dots & \dots \\ C_L & C_{L+1} & C_{L+3} & \dots & C_{L+M-1} \end{pmatrix} \begin{pmatrix} b_M \\ b_{M-1} \\ b_{M-2} \\ \vdots \\ b_1 \end{pmatrix} = \begin{pmatrix} C_{L+1} \\ C_{L+2} \\ C_{L+3} \\ \vdots \\ C_{L+M} \end{pmatrix}. \quad (19)$$

Отсюда могут быть найдены b_i , коэффициенты числителя a_0, a_1, \dots, a_L находим теперь из (17) сравнением коэффициентов при $1, z, z^2, \dots, z^L$:

$$\begin{aligned}
a_0 &= c_0; \\
a_1 &= c_1 + b_1c_0; \\
a_2 &= c_2 + b_1c_1 + b_2c_0; \\
\vdots & \\
a_L &= c_L + \sum_{i=1}^{\min(L,M)} b_i c_{L-i}.
\end{aligned} \quad (20)$$

Уравнения (19) и (20) называются аппроксимациями Паде: в случае, когда система (19) разрешилась, они определяют коэффициенты числителя и знаменателя аппроксимации Паде $[L/M]$. Коэффициенты числителя Тейлоровского разложения этой функции при $1, z, z^2, \dots, z^{L+M}$ совпадают с соответствующими коэффициентами ряда (14).

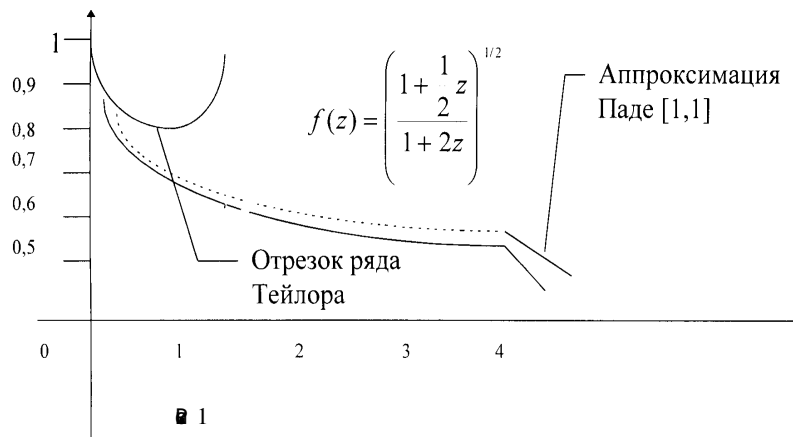
Поскольку искомой точкой разделения являются коэффициенты ряда $\sum_{i=0}^{\infty} c_i z^i$, то нет необходимости заранее считать эти коэффициенты коэффициентами ряда Тейлора какой-либо функции. Чтобы продемонстрировать, насколько хорошо срабатывает аппроксимация Паде, рассмотрим следующий пример:

$$f(z) = \sqrt{\frac{1 + \frac{1}{2}z}{1 + 2z}} = 1 - \frac{3}{4}z + \frac{39}{32}z^2 - \dots$$

Выпишем аппроксимацию Паде $[1/1]$. Используя уравнения (19), (20), получим $b_0 = 1$, $b_1 = 13/8$, $a_0 = 1$, $a_1 = 7/8$. Правильность вычислений можно проверить выполнением равенства

$$(1 + 13/8z)\left(1 - 3/4z + 39/32z^2\right) = 1 + 7/8z + O(z^3)$$

Таким образом, мы получаем, что $[1/1] = \frac{1 + \frac{7}{8}z}{1 + \frac{13}{8}z}$.



На рис. 1 приведены для сравнения графики функций $f(z)$ и $[1/1](z)$ при $z \geq 0$.

В частности, имеем $f(\infty) = 0,5$; $[1/1](\infty) = 7/13 = 0,54$, значения в бесконечности аппроксимация Паде дается с точностью 80 %. Этот пример показывает замечательную точность, достигаемую аппроксимацией Паде, которая использует всего три числа ряда.

Однако необходимо отметить одну особенность, связанную с приближенными вычислениями аппроксимации Паде, эти вычисления требуют большей точности, чем можно было бы заранее предположить. Аппроксимация Паде экстраполирует всю последовательность коэффициентов по их конечному числу, поэтому исходные коэффициенты должны быть вычислены достаточно точно. Приведем ниже таблицу аппроксимаций Паде функции $f(z) = e^z$. Отметим, что аппроксимация Паде этой функции находит широкое применение в теории автоматического управления при исследовании систем с запаздыванием.

L/M	0	1	2
0	$\frac{1}{1}$	$\frac{1+z}{1}$	$\frac{2+2z+z^2}{2}$
1	$\frac{1}{1-z}$	$\frac{2+z}{-z+2}$	$\frac{6+4z+z^2}{6-2z}$
2	$\frac{2}{2+2z+z^2}$	$\frac{6+2z}{6-4z+z^2}$	$\frac{12+6z+z^2}{12-6z+z^2}$

Используя различия аппроксимации, представленной в таблице, можно получить аппроксимацию e^z с различной степенью точности, что имеет важное значение при решении задачи синтеза и анализа систем автоматического управления.

Тема 16 ИСПОЛЬЗОВАНИЕ ТЕОРИИ ВЕТВЯЩИХСЯ ДРОБЕЙ ДЛЯ РЕШЕНИЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Решение обыкновенных дифференциальных уравнений. Для решения обыкновенных дифференциальных уравнений хорошо разработаны линейные методы численного решения Рунге-Кутта, Адамса и др. Рассмотрим нелинейные численные методы, основанные на дробно-рациональном представлении искомых решений.

Обыкновенное дифференциальное уравнение

$$y' = f(x, y), \quad x \in [x_0, X] \quad (1)$$

с начальным условием

$$y(x_0) = y_0. \quad (2)$$

Как известно, основным источником линейных численных методов решения задач (1), (2) является разложение функции $y(x)$ в ряд Тейлора. Аналогично использование формулы Тиле, приведенной в предыдущих темах, может служить новым источником численных методов решения задач (1), (2). Методы решения обыкновенных дифференциальных уравнений, основанные на дроби Тиле, будут нелинейными. Можно предвидеть, что нелинейные численные методы будут устойчивыми относительно малых отклонений значений обоих аргументов, поскольку цепные дроби устойчивы относительно малых отклонений своих аргументов. Функции можно приближать дробно-рациональными выражениями. Используя эти методы, можно получить дробно-рациональное приближенное решение дифференциального уравнения (1).

Опишем численный метод типа Рунге-Кутта третьего порядка. Выберем на отрезке $[x_0, X]$ некоторую сетку $\{x_n, 0 \leq n \leq N\}$ значений аргумента так, чтобы выполнялись соотношения $x_0 < x_1 < x_2 < \dots < x_N = X$.

Разложение решения $y(x)$ задачи (1), (2) на отрезке $x_n \leq x \leq x_{n+1}$ по формуле Тиле, ограничим дробью с тремя звеньями, хотя в принципе можно выбрать дробь с большим числом звеньев.

$$y_{n+1} = y_n + \frac{h}{\sqrt{[y(x)]_{x=x_n} + \frac{h}{2\sqrt{[y(x)]_{x=x_n} + \frac{h}{3\sqrt{[y(x)]_{x=x_n}}}}}},$$

где $y_{n+1} \approx y(x_{n+1})$, $h = x_{n+1} - x_n$; $y = 1/y'$;

$$2\sqrt{[y]} = -\frac{2(y')^2}{y''}; \quad 3\sqrt{[y]} = \frac{3(y'')^2}{2(y')^2 y''' - 3y'(y'')^2}.$$

После замены обратных производных получаем

$$y_{n+1} = y_n + hy'_n + \frac{3h^2 (y''_n)^2}{6y''_n - 2hy'''_n}.$$

Исходя из приведенной формулы, будем искать желаемое решение в виде

$$y_{n+1} = y_n + hk_1 + h \frac{(k_2 - k_1)^2}{p_1 k_1 + p_2 k_2 + p_3 k_3}, \quad (3)$$

где

$$\begin{aligned} k_1 &= f(x_n, y_n); \\ k_2 &= f(x_n + \alpha_2 h, y_n + \beta_{21} h k_1); \\ k_3 &= f(x_n + \alpha_3 h, y_n + \beta_{31} h k_1 + \beta_{32} h k_2); \\ \alpha_2 &= \beta_{21}; \quad \alpha_3 = \beta_{31} + \beta_{32}. \end{aligned}$$

Неизвестные константы могут быть найдены из решения системы уравнений

$$\begin{aligned} p_1 + p_2 + p_3 &= 0; \\ p_2 \alpha_2 + p_3 \alpha_3 &= 2\alpha_2^2; \\ p_2 \alpha_2^2 + p_3 \alpha_3^2 &= -4\alpha_2^2 (\alpha_2 - 1/3); \\ p_3 \beta_{32} &= -2\alpha_2/3; \\ \alpha_2 &= \beta_{21}; \quad \alpha_3 = \beta_{31} + \beta_{32}. \end{aligned}$$

Эта система имеет бесчисленное множество решений. С помощью двух параметров, которые являются произвольными, можно минимизировать точность аппроксимации. Следовательно, мы имеем двухпараметрическое (по числу свободных параметров) семейство нелинейных методов типа Рунге-Кутты.

Одним из решений приведенной системы нелинейных уравнений будет

$$p_1 = -7/6; \quad p_2 = 4/3; \quad p_3 = -1/6; \quad \alpha_2 = \beta_{21} = 1/2; \quad \alpha_3 = 1; \quad \beta_{31} = -1; \quad \beta_{32} = 2.$$

Тогда формула (3) будет иметь вид

$$\begin{aligned} y_{n+1} &= y_n + hk_1 - h \frac{6(k_2 - k_1)^2}{7k_1 - 8k_2 + k_3}; \\ k_1 &= f(x_n, y_n); \\ k_2 &= f(x_n + 1/2h, y_n + 1/2hk_1); \\ k_3 &= f(x_n + h, y_n - hk_1 + 2hk_2). \end{aligned}$$

Получим аналогично формулу для нелинейного аналога метода Рунге-Кутты. Для этого будем искать численное решение в виде

$$y_{n+1} = \frac{v_0}{1 - \frac{v_1 h}{1 - \frac{v_2 h}{1 - \frac{v_3 h}{1 - v_4 h}}}}, \quad (4)$$

где

$$\begin{aligned}
v_0 &= y_n; v_1 = \frac{k_1}{y_n}; v_2 = \frac{y_n q_1 - k_1^2}{y_n k_1}; v_3 = \frac{y_n(k_1 q_3 - q_2^2)}{k_1(y_n q_1 - k_1^2)}; \\
v_4 &= \frac{k_1\{q_1(k_1 q_2 - q_1^2) + q_2(k_1 q_1 - y_n q_2) + q_3(y_n q_1 - k_1^2)\}}{(y_n q_1 - k_1^2)(k_1 q_2 - q_1^2)}; \\
q_1 &= a_1 k_1 + a_2 k_2; q_2 = b_1 k_1 + b_2 k_2 + b_3 k_3; \\
q_3 &= c_1 k_1 + c_2 k_2 + c_3 k_3 + c_4 k_4; \\
k_1 &= f(x_n, y_n); k_2 = f(x_n + \alpha_2 h, y_n + \beta_{21} h k_1); \\
k_3 &= f(x_n + \alpha_3 h, y_n + \beta_{31} h k_1 + \beta_{32} h k_2); \\
k_4 &= f(x_n + \alpha_4 h, y_n + \beta_{41} h k_1 + \beta_{42} h k_2 + \beta_{43} h k_3).
\end{aligned}$$

Для того, чтобы метод (4) имел четвертый порядок точности, необходимо, чтобы неизвестные параметры удовлетворяли следующей системе уравнений:

$$\begin{aligned}
a_1 + a_2 + 0; a_2 \alpha_2 &= 1/2; \\
b_1 + b_2 + b_3 &= 0; \\
b_2 \alpha_2 + b_3 \alpha_3 &= 0; \\
(a_2 + b_2 + c_2) \alpha_2^2 + (b_3 + c_3) \alpha_3^2 + c_4 \alpha_4^2 &= 1/3; \\
c_2 \alpha_2 + c_3 \alpha_3 + c_4 \alpha_4 &= 0; \\
c_1 + c_2 + c_3 &= 0; \\
b_3 \beta_{32} \alpha_2 &= 1/6; \\
c_3 \beta_{32} \alpha_2 + c_4 (\beta_{42} \alpha_2 + \beta_{43} \alpha_3) &= 0; \\
(a_2 + b_2 + c_2) \alpha_2^2 + (b_3 + c_3) \alpha_3^2 + c_4 \alpha_4^2 &= 1/4; \\
(b_3 + c_3) \alpha_2^2 \beta_{32} + c_4 (\beta_{42} \alpha_2^2 + \beta_{43} \alpha_3^2) &= 1/12; \\
(\beta_3 + c_3) \alpha_2 \beta_{32} \alpha_3 + c_4 \alpha_4 (\beta_{42} \alpha_2 + \beta_{43} \alpha_3) &= 1/8; \\
c_4 \beta_{43} \beta_{32} \alpha_2 &= 1/24.
\end{aligned}$$

В итоге получаем систему 13 уравнений с 15 неизвестными. Выразим все параметры через два свободных параметра α_2, α_3 , в итоге получаем

$$\begin{aligned}
a_1 &= -\frac{1}{2\alpha_2}; a_2 = \frac{1}{2\alpha_2}; \\
b_1 &= \frac{1-2\alpha_2}{3\alpha_2\alpha_3}; b_2 = \frac{2\alpha_2-1}{3\alpha_2(\alpha_3-\alpha_2)}; b_3 = \frac{1-2\alpha_2}{3\alpha_3(\alpha_3-\alpha_2)}; \\
c_1 &= \frac{\alpha_3(3(1+2\alpha_2\alpha_3)-4(\alpha_2+\alpha_3))-(1-2\alpha_2)(3-4\alpha_3)}{12\alpha_2\alpha_3(1-\alpha_3)}; \\
c_2 &= \frac{(1-\alpha_2)(1-2\alpha_2)(3-4\alpha_3)-(\alpha_3-\alpha_2)(3(1+2\alpha_2\alpha_3)-4(\alpha_2+\alpha_3))}{12\alpha_2(1-\alpha_2)(1-\alpha_3)(\alpha_3-\alpha_2)}; \\
c_3 &= \frac{(2\alpha_2-1)(3-4\alpha_3)}{12\alpha_3(\alpha_3-\alpha_2)(1-\alpha_3)}; c_4 = \frac{3(1+2\alpha_2\alpha_3)-4(\alpha_2+\alpha_3)}{12(1-\alpha_2)(1-\alpha_3)}; \\
\beta_{31} &= \frac{2\alpha_2\alpha_3(1-\alpha_2)-\alpha_3(\alpha_3-\alpha_2)}{2\alpha_2(1-2\alpha_2)}; \beta_{32} = \frac{\alpha_3(\alpha_3-\alpha_2)}{2\alpha_2(1-2\alpha_2)}; \\
\beta_{42} &= \frac{2(1-\alpha_3)(1-\alpha_2)(2\alpha_3-1)-(1-\alpha_2)(\alpha_3-\alpha_2)}{2\alpha_2(\alpha_3-\alpha_2)(3(1+2\alpha_2\alpha_3)-4(\alpha_2+\alpha_3))}; \\
\beta_{43} &= \frac{(1-\alpha_3)(1-\alpha_2)(1-2\alpha_2)}{\alpha_3(\alpha_3-\alpha_2)(3(1+2\alpha_2\alpha_3)-4(\alpha_2+\alpha_3))}; \\
\alpha_4 &= 1; \beta_{41} = \alpha_4 - \beta_{42} - \beta_{43}.
\end{aligned}$$

Понятно, что в случае использования предлагаемой методики можно получить метод любой заданной точности. Трудности при построении таких методов, как видно, будут вычислительного характера.

Выбирая соответствующим образом α_2, α_3 , получим двухпараметрическое семейство методов для решения задач Коши.

В настоящее время известны работы, в которых на основе изложенного подхода сделана попытка построения численных методов для решения жестких систем обыкновенных дифференциальных уравнений.

Решение дифференциальных уравнений с частными производными. В первую очередь рассмотрим линейные дифференциальные уравнения с частными производными. Известно, что такое уравнение на основе использования метода конечных разностей можно привести к системе линейных алгебраических уравнений при условии, что граничные условия также будут линейными. Решения этой системы уравнений дают приближительные решения уравнения в частных производных.

Использование метода ветвящихся целых дробей для решения хорошо приближающихся систем конечноразностных уравнений дают хорошие результаты и для приближенного нахождения решений уравнений с частными производными. Оказалось, что алгоритмы Гаусса, окаймления, ... и некоторые другие могут быть интерпретированы как алгоритмы ветвящихся цепных дробей специального вида. Это открывает новые возможности для построения новых алгоритмов в виде ветвящихся цепных дробей иных типов. Для построения приближенных методов решения нелинейных дифференциальных уравнений с частными производными можно использовать формулу разложения функций в ветвящуюся цепную дробь. Приведем некоторые результаты, основанные на формулах разложения функций в ветвящиеся цепные дроби.

А. А. Самарский в 1980 г. высказал идею применения цепных и ветвящихся цепных дробей для нелинейного приближения производных, имеющих в дифференциальном уравнении. Покажем это на примере.

Как известно, производную $y'(x)$, $y''(x)$ можно линейно заменить, используя формулу Тейлора. Пусть шаг сетки h , обозначим $y_k = y(kh)$, тогда $y_{n+1} = y_n + hy'(nh)$. Поэтому

$$\begin{aligned} y'(nh) &= (y_{n+1} - y_n) / h y'; \\ y_{n+1} &= y_n + hy'(nh) + 1/2 h^2 y''(nh); \\ y_{n-1} &= y_n - hy'(nh) + 1/2 h^2 y''(nh). \end{aligned}$$

Отсюда

$$y''(nh) = \frac{\begin{vmatrix} h & y_{n+1} - y_n \\ -h & y_{n-1} - y_n \end{vmatrix}}{\begin{vmatrix} h & \frac{1}{2}h^2 \\ -h & \frac{1}{2}h^2 \end{vmatrix}} = \frac{y_{n+1} - y_n - y_n + y_{n-1}}{2 \frac{h^2}{2}} = \frac{y_{n+1} - 2y_n + y_{n-1}}{h^2}.$$

Однако можно заменить эти же две производные, используя формулу Тиле:

$$y_{n+1} = y_n + \frac{h}{\frac{1}{y'(nh)}}; \quad y_{n+1} = y_n + \frac{h}{\alpha_1 + \frac{h}{\alpha_2}}; \quad y_{n-1} = y_n + \frac{-h}{\alpha_1 + \frac{-h}{\alpha_2}}.$$

Из последних двух соотношений получаем нелинейное выражение для y'' . Понятно, что

на этом пути можно получить приближенное значение производных сколь угодно высокого порядка. Нужно только увеличивать количество звеньев в формулах Тиле и количество уравнений для определения производных.

Рассмотрим численное решение рядом с начальными условиями для уравнений теплопроводности с помощью явной нелинейной сеточной модели

$$L(u) = f,$$

где

$$L(u) = \begin{cases} \frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2}; \\ u(x, 0). \end{cases} \quad \bar{D} = \{-\infty < x < \infty; 0 \leq t \leq T \leq \infty\};$$

$$f = \begin{cases} \varphi(x, t); \\ \psi(x) \end{cases} \quad \bar{D} = \{-\infty < x < \infty; 0 \leq t \leq T \leq \infty\}.$$

Предполагается, что все известные функции - гладкие. Выберем прямоугольную сетку и заменим область \bar{D} сеточной областью \bar{D}_h , причем координаты узловых точек сетки (x_m, t_n) определяются по формулам

$$x_m = mh; \quad t_n = n\tau; \quad m = 0, \pm 1, \pm 2, \dots, \quad h > 0; \quad x_m = mh; \quad t_n = n\tau.$$

$$t_n = n\tau, \quad n = 0, N; \quad \tau > 0; \quad N\tau \leq T \leq (N+1)\tau,$$

Вместо задачи (1) рассмотрим ее разностную модель, когда производная определяется приблизительно не на основе разложения в ряд Тейлора, а на основе разложения функции в цепные дроби.

$$\frac{\partial u}{\partial t} \Big|_{(x_m, t_n)} = \frac{u_m^n (u_m^{n+1} - u_m^n)}{u_m^{n+1} \tau} - \frac{\tau}{2} \frac{\partial^2 u}{\partial t^2} \Big|_{(x_m, t_n)}; \quad (5)$$

$$\frac{\partial^2 u}{\partial x^2} \Big|_{(x_m, t_n)} = \frac{4(u_{m+1}^n - 2u_m^n + u_{m-1}^n)(u_{m+1}^n - u_m^n)(u_m^n - u_{m-1}^n)}{(u_{m+1}^n - u_{m-1}^n)^2 h^2} - \frac{h^2 \partial^4 u}{12 \partial x^4} \Big|_{(x_m, t_n)}, \quad (6)$$

где $u_m^n = u(x_m, t_n)$.

Формула (5) получается, если воспользоваться разложением функции $u(x, t)$ в цепную дробь без нулевого члена, а формула (6) - если воспользоваться разложением той же функции $u(x, t)$ по формуле Тиле. При такой замене вместо линейной задачи (1) нужно решать систему линейных алгебраических уравнений:

$$Lh(u^h) = f^{(h)};$$

$$f^h = \begin{cases} \Phi_m^n; \\ \Psi_m; \end{cases}$$

$$u^h = u(x_m, t_n).$$

Эффективность предлагаемых схем можно увидеть на текстовых примерах с использованием ЭВМ.

Рассмотрим еще один подход, основанный на использовании теории ветвящихся цепных дробей для решения дифференциальных уравнений в частных производных. Продемонстрируем сущность предлагаемого подхода на решении следующей задачи:

$$\frac{\partial u}{\partial t} = a^2(x) \frac{\partial^2 u}{\partial x^2} + f(x, t), \quad -a \leq x \leq 1$$

- с граничным условием первого рода:

$$\begin{aligned} u(x, t)|_{x=0} &= F_{\text{Гр1}}(t); \\ u(x, t)|_{x=1} &= F_{\text{Гр2}}(t), \end{aligned} \quad (7)$$

- с начальным условием вида $u(x, t)|_{t=0} = E(x)$.

Предположим, что решение $u(x, t)$ задачи (7) может быть приближено аппроксимирующим рядом $u(x, t) = \sum_n V_n(t) \Phi_n(x)$, в котором $\Phi_n(x)$ - известная функция, например, x^n , x^{2n} , $\cos nx$ и др., имеющие соответствующую степень гладкости по x .

Производные $\frac{\partial u(x, t)}{\partial x}$, $\frac{\partial^2 u(x, t)}{\partial x^2}$ с использованием граничных и начальных условий могут быть разложены в ряды по функциям $u(x, t)$. Например,

$$\frac{\partial^2 u(x, t)}{\partial x^2} = \sum_n V_n(t) \frac{\partial^2 \Phi}{\partial x^2} = \sum_i D_i(x, x_i) u(x, t),$$

где координаты $D_i(x, x_i)$ выражаются через известную функцию $\Phi_n(x)$, их производные - через граничные условия. Явный вид $D_i(x, x_i)$ можно определить разными способами.

Введем ряд опорных точек в область определения функции $u(x, t)$ по x , получим x_0, x_1, \dots, x_n . Далее построим ветвящуюся цепную дробь, интерполирующую $u(x, t)$ по опорным точкам. Для этого используем представление в виде формулы Тиле. Далее величины $u(-1, t)$ и $u(1, t)$ определяются из граничных условий, т.е. $u(0, t) = F_{\text{Гр1}}(t)$, $u(1, t) = F_{\text{Гр2}}(t)$.

Частные производные $\frac{\partial u}{\partial t}$, $\frac{\partial^2 u}{\partial x^2}$ определяются непосредственным дифференцированием ветвящейся дроби, получаемой по опорным точкам функции $u(x, t)$ по x .

В результате будем иметь систему обыкновенных дифференциальных уравнений, каждое может быть решено известными методами. Отметим, что, как правило, полученная система уравнений обладает свойствами жесткости.

Решение задач устойчивости систем автоматического управления. Рассмотрим решение задач об устойчивости линейной системы автоматического управления, характеристическое уравнение которой имеет следующий вид:

$$f(S) = S^n + a_{n-1}S^{n-1} + a_{n-2}S^{n-2} + \dots + a_1S + a_0. \quad (8)$$

Как правило, для решения задач об устойчивости линейной системы автоматического управления, характеристическое уравнение которой имеет вид (8), использовались критерии Гаусса или Гурвица, однако их использование связано со значительными вычислительными затратами. Ниже без доказательства приведен ряд утверждений, использование которых позволяет решить задачу об устойчивости с меньшими вычислительными затратами.

Утверждение 1. Пусть $\hat{f}(S) = S^n + a_{n-1}S^{n-1} + a_{n-2}S^{n-2} + \dots + a_0$ - полином с вещественными коэффициентами и пусть

$$g(S) = a_{n-1}S^{n-1} + a_{n-3}S^{n-3} + a_{n-5}S^{n-5} + \dots$$

Тогда $\hat{f}(S)$ является устойчивым полиномом в том и только в том случае, когда проверяемую дробь $t = g/\hat{f}$ можно представить как неправильную дробь вида

$$t(S) = \frac{g(S)}{\hat{f}(S)} = \frac{1}{1+d_1S} + \frac{1}{d_2S} + \frac{1}{d_3S} + \dots + \frac{1}{d_nS},$$

где $d_i > 0$, $i = \overline{1, n}$

Пример 1.

$$\hat{f}(S) = S^4 + 5S^3 + 10S^2 + 10S + 4;$$

$$g(S) = 5S^3 + 10S;$$

$$t(S) = \frac{1}{\hat{f}(S)/g(S)} = \frac{1}{1+h_1(S)};$$

$$h_1(S) = (1/5)S + (8S^2 + 4)/(5S^3 + 10S);$$

$$h_1(S) = (1/5)S + 1/h_2(S),$$

где после деления $5S^3 + 10S$ на $8S^2 + 4$ получаем

$$h_2(S) = (5/8)S + (15/2)S/(8S^2 + 4),$$

следовательно, $h_2(S) = (5/8)S + 1/h_3(S)$, где после деления $8S^2 + 4$ на $15S/2$ получаем

$$h_3(S) = (16/15)/S + 1/[(15/8)S].$$

Комбинируя предыдущие результаты, получаем представление в виде непрерывной дроби

$$t(S) = \frac{1}{1+(1/5)S} + \frac{1}{(5/8)S} + \frac{1}{(16/5)S} + \frac{1}{(15/8)S}.$$

Из утверждения вытекает, что $\hat{f}(S)$ устойчивый полином.

Пример 2.

При каком значении вещественной константы

$$\hat{f}(S) = S^4 + 5S^3 + 10S^2 + 10S + c$$

будет устойчивым. Следуя процедуре, использованной в примере 1, получаем

$$t(S) = \frac{1}{1+(1/5)S} + \frac{1}{(5/8)S} + \frac{1}{d_3S} + \frac{1}{d_4S},$$

где $d_3 = 8/[10 - (5/8)c]$; $d_4 = [10 - (5/8)c]$, следовательно, $d_3 > 0$ тогда, когда $c > 16$, а d_4 больше нуля, когда $c > 0$. Таким образом, можно сделать вывод, что $\hat{f}(S)$ будет устойчивым полиномом в том случае, если $0 < c < 16$. Очевидно, что для устойчивости полинома $\hat{f}(S)$ необходимо существование, с одной стороны, его представления в виде:

$$\frac{1}{1+d_1S} + \frac{1}{d_2S} + \frac{1}{d_3S} + \dots + \frac{1}{d_mS},$$

а с другой стороны, обязательно выполнение условия $m = n$, где n - степень полинома $\hat{f}(S)$. Приведем еще два утверждения, так же без доказательства.

Утверждение 2. Пусть $f(S) = a_0 + a_1S + a_2S^2 + \dots + a_nS^n$, $a_0 \neq 0$, вещественный полином степени n и пусть

$$D_1 = a_1; \quad D_2 = \begin{vmatrix} a_1 & a_3 \\ a_0 & a_2 \end{vmatrix}; \quad D_3 = \begin{vmatrix} a_1 & a_3 & a_5 \\ a_0 & a_2 & a_4 \\ 0 & a_1 & a_3 \end{vmatrix},$$

где D_k - определитель $(k \times k)$, $a_m = 0$ при $m > n$ и $D_0 = 1$; $D_{-1} = 1/a_0$; $D_{-2} = 1/a_0^2$.

Тогда f является устойчивым полиномом в том и только в том случае, когда определители удовлетворяют условиям:

$$D_{2k} > 0, \quad k = 1, n/2;$$

$$\text{sign} D_{2k+1} = \text{sign} a_0 k = 0, [(n-1)/2]$$

Отметим, что с вычислительной точки зрения значительно проще установить устойчивость полинома $f(S)$ методом деления (Утверждение 1) или путем вычисления определителей. Утверждение 2 носит название теоремы Гурвица.

СПИСОК ЛИТЕРАТУРЫ

- 1 *Иванов В. В.* Методы вычислений на ЭВМ: Справочное пособие. Киев: Наукова думка, 1986. 584 с.
- 2 *Демидович Б. П., Марон И. А.* Основы вычислительной математики. М.: Гос. изд-во физ.-мат. литературы, 1963. 659 с.
- 3 *Рябенский В. С.* Введение в вычислительную математику: Учеб. пособие для вузов. М.: Физматлит, 1994. 336 с.
- 4 *Ланцон К.* Практические методы прикладного анализа: Справочное руководство. М.: Гос. изд-во физ.-мат. литературы, 1961. 524 с.
- 5 *Хемминг Р. В.* Численные методы для научных работников и инженеров. М.: Наука, 1972. 400 с.
- 6 *Коллатц Л.* Функциональный анализ и вычислительная математика. М.: Мир, 1969. 447 с.
- 7 *Демидович Б. П., Марон И. А., Шувалова Э. З.* Численные методы анализа. Приближение функций, дифференциальные и интегральные уравнения. М.: Физматлит, 1967. 368 с.
- 8 *Бахвалов И. С., Жидков Н. П., Кобельков Г. М.* Численные методы. М.: Наука, 1987. 600 с.
- 9 *Воробьева С. Н., Данилова А. Н.* Практикум по вычислительной математике. М.: Высшая школа, 1990. 208 с.
- 10 *Березин И. С., Жидков Н. П.* Методы вычислений. Т. 1. М.: Физматлит, 1959. 464 с.
- 11 *Березин И. С., Жидков Н. П.* Методы вычислений. Т. 2. М.: Физматлит, 1959. 620 с.
- 12 *Самарский А. А.* Введение в численные методы. М.: Наука, 1982. 272 с.

- 13 Хинчин А. Я. Цепные дроби. М.: Физматгиз, 1961. 112 с.
- 14 Дрейкер Д. М., Грейвс-Моррис П. Аппроксимации Паде. М.: Мир, 1986. 502 с.
- 15 Джоунс У., Трон В. Непрерывные дроби. Аналитическая теория и приложения. М.: Мир, 1987. 414с.
- 16 D. R. Smith. The multivariable method in singular perturbation, SIAM Rev.,17, 1975. P. 221 - 273.
- 17 C. M. Bender, S. A. Orszag. Advanced mathematical methods for scientists and engineers. McGrawHill, 1978.
- 18 R. E. O'Malley. Jr Introduction to singular perturbation. Academic Press, New York, 1974.
- 19 Êàëèðèè Ì. Ì. ×ñëãííå ìàðíå. М.: Наука, 1978.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ		3
.....		
.....		
Тема 1	ЧИСЛЕННОЕ РЕШЕНИЕ НЕЛИНЕЙНЫХ УРАВНЕНИЙ	4
	
	
Тема 2	ЧИСЛЕННОЕ РЕШЕНИЕ СИСТЕМ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ	10
Тема 3	ЧИСЛЕННОЕ РЕШЕНИЕ СИСТЕМ НЕЛИНЕЙНЫХ УРАВНЕНИЙ	20
	
Тема 4	ПРИБЛИЖЕНИЕ ФУНКЦИЙ ИНТЕРПОЛЯЦИОННЫМИ ПОЛИНОМАМИ	24
	
Тема 5	АППРОКСИМАЦИЯ ФУНКЦИЙ МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ	30
	
Тема 6	ПРИБЛИЖЕННОЕ ВЫЧИСЛЕНИЕ ПРОИЗВОДНЫХ	33
Тема 7	ПРИБЛИЖЕННОЕ ВЫЧИСЛЕНИЕ ОПРЕДЕЛЕННЫХ ИНТЕГРАЛОВ	36
	
Тема 8	МЕТОДЫ РЕШЕНИЯ ЗАДАЧИ КОШИ	43
	

Тема 9	КРАЕВЫЕ ЗАДАЧИ ДЛЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ	48
Тема 10	РЕШЕНИЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ	65
Тема 11	МЕТОДЫ РЕШЕНИЯ ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ	74
Тема 12	ОСНОВЫ ТЕОРИИ ЦЕПНЫХ ДРОБЕЙ	81
Тема 13	АНАЛОГ ФОРМУЛЫ ТЕЙЛОРА В ТЕОРИИ ЦЕПНЫХ ДРОБЕЙ	89
Тема 14	ЭЛЕМЕНТЫ ТЕОРИИ ВЕТВЯЩИХСЯ ДРОБЕЙ	101
Тема 15	ПРИБЛИЖЕНИЯ ФУНКЦИЙ МНОГИХ ПЕРЕМЕННЫХ ВЕТВЯЩИМИСЯ ЦЕПНЫМИ ДРОБЯМИ	106
Тема 16	ИСПОЛЬЗОВАНИЕ ТЕОРИИ ВЕТВЯЩИХСЯ ДРОБЕЙ ДЛЯ РЕШЕНИЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ	116
СПИСОК	ЛИТЕРАТУРЫ	126
.....		