ЧИСЛЕННЫЕ МЕТОДЫ

курс лекций

Кандоба И. Н.

2023-2024, ИЕНиМ УрФУ, Екатеринбург

Контакты

Кандоба Игорь Николаевич

kandoba@imm.uran.ru

▲1 Учебный план

I семестр

1. 2 контрольные работы
2. 3 лабораторные работы
3. Коллоквиум
4. Зачет

II семестр

1. 2 контрольные работы
2. 3 лабораторные работы
3. Экзамен

Рекомендуемая литература

Классические учебники

1. Березин И. С., Жидков Н. П. Методы вычислений. 1962.
2. Бахвалов Н. С. Численные методы. 1973.
3. Калиткин Н. Н. Численные методы. 1978.
4. Хемминг Р. В. Численные методы. 1977.
5. Волков Е. А. Численные методы. 1982.

Учебные пособия

1. Пименов В. Г. Численные методы : в 2 ч. Ч.1: М-во образования и науки Рос. Федерации, УрФУ— Екатеринбург: Изд-во Урал. ун-та, 2013.— 112 с.
2. Пименов В. Г., Ложников А. Б. Численные методы : в 2 ч. Ч.2: М-во образования и науки Рос. Федерации, УрФУ— Екатеринбург: Изд-во Урал. ун-та, 2014.— 106 с.

ТЕМА 1. Теория погрешностей

1.1. Последовательность моделей и структура погрешности

1. Реальный объект
2. Математическая модель
3. Алгоритмическая модель
4. Вычислительная модель

$$1 \mapsto 2 \mapsto 3 \mapsto 4 \Rightarrow \text{ Результат (Число)}$$

$1 \mapsto 2 \Rightarrow$ Неустранимая погрешность
$2 \mapsto 3 \Rightarrow$ Погрешность метода (погрешность усечения)
$3 \mapsto 4 \Rightarrow$ Вычислительная погрешность[a]

[a] Классификация погрешностей по Колмогорову

Пример 1.

1. Реальный объект — колебательный процесс
2. Математическая модель — $y = \sin(x)$
3. Алгоритмическая модель — $\widetilde{y}(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$
4. Вычислительная модель $\Rightarrow \widetilde{y}(25,7) \approx y^\star = 24,25401855...$

Пример 2.

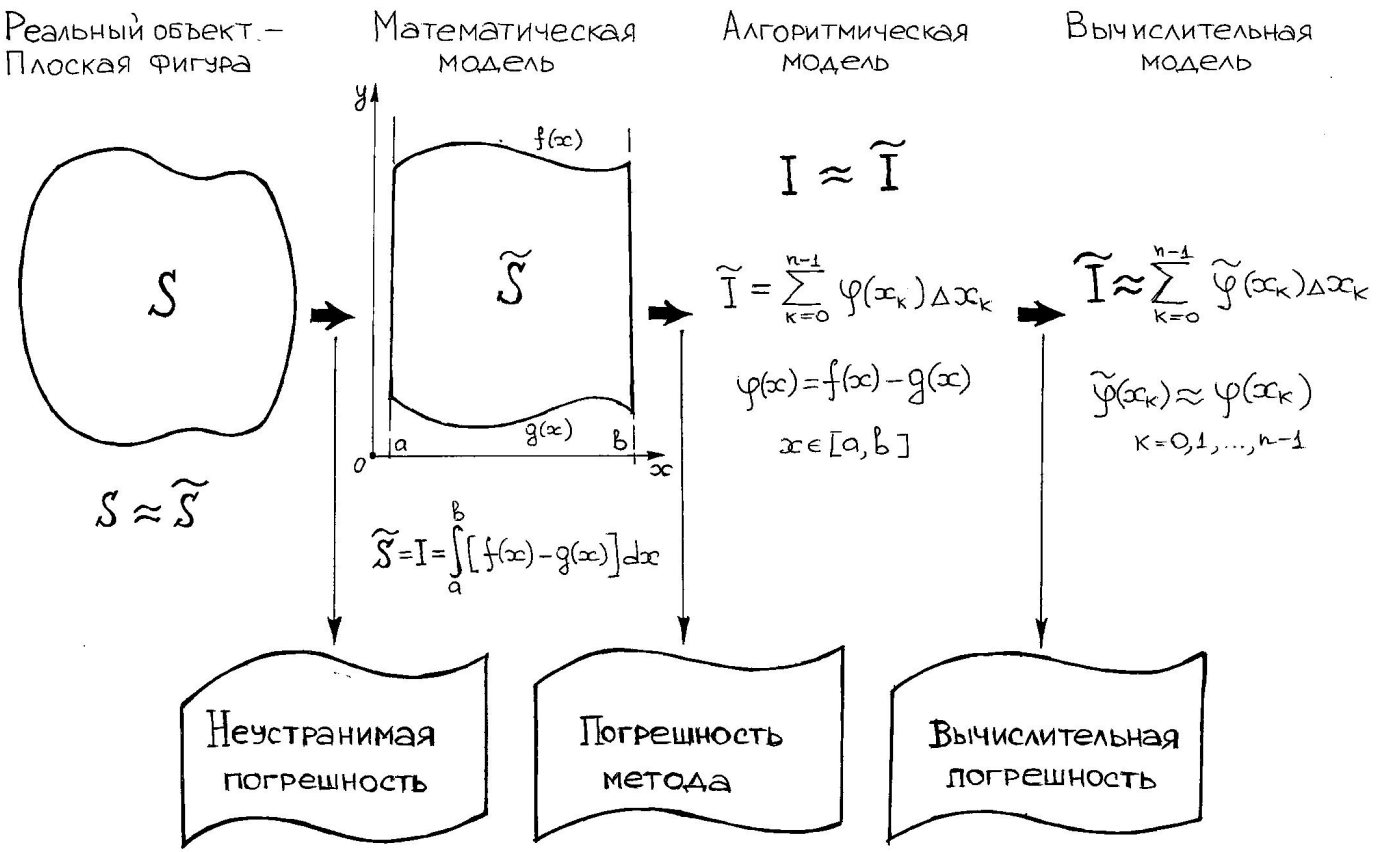


Рис. 1: Вычисление приближенного значения площади плоской фигуры.

1.2. Абсолютная и относительная погрешности

x — точное число, $x^\star$ — приближенное число (приближение)

Определение 1.1.

Абсолютной погрешностью приближения $x^\star$ называется такая величина $A_{x^\star}$, которая удовлетворяет неравенству

$$|x - x^\star| \le A_{x^\star} \tag{1.1}$$

Определение 1.2.

Относительной погрешностью приближения $x^\star$ $(x^\star \neq 0)$ называется такая величина $\Delta_{x^\star}$, которая удовлетворяет неравенству

$$\left|\frac{x - x^\star}{x^\star}\right| \le \Delta_{x^\star} \tag{1.2}$$

$A_{x^\star}$ — величина размерная: $x = x^\star \pm A_{x^\star}$
$\Delta_{x^\star}$ — величина безразмерная: $x = x^\star\,(1 \pm \Delta_{x^\star})$

$$\Delta_{x^\star} = \frac{A_{x^\star}}{|x^\star|} \tag{1.3}$$

1.3. Погрешность и позиционная запись числа

$3,14159265358979323846264 3$ — позиционная запись числа π

Определение 1.3.
Значащими цифрами числа называются все цифры в его позиционной записи, следующие слева направо, начиная с первой ненулевой

Примеры: 1) $0,00\underline{120}$; 2) $\underline{1},\underline{20}$.

Определение 1.4.
Цифра в позиционной записи числа называется верной, если абсолютная погрешность числа не превосходит половины единицы разряда, соответствующего этой цифре[a]

[a] В противном случае цифра называется сомнительной

Примеры:
1) $x^\star = \underline{0},\underline{00120}$, где $A_{x^\star} = 0,00004$;
2) $x^\star = \underline{1},20$, где $A_{x^\star} = 0,07$.

1.4. Определение погрешности значения функции по погрешностям ее аргументов

$y = f(x_1, x_2, \ldots, x_n)$, $\boldsymbol{x} \in D[f] \subseteq \mathbb{R}^n$, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top \in \mathbb{R}^n$.

Пусть $x_i^\star : A_{x_i^\star}, \Delta_{x_i^\star}$ $(i = 1, 2, \ldots, n)$, $\boldsymbol{x}^\star = (x_1^\star, x_2^\star, \ldots, x_n^\star)^\top \in \mathbb{R}^n$.

Требуется построить оценку сверху для

$$|y - y^\star| = |f(x_1, x_2, \ldots, x_n) - f(x_1^\star, x_2^\star, \ldots, x_n^\star)|.$$

Вспомогательная функция

$F(t) = f(\boldsymbol{x}^\star + t\boldsymbol{h})$,
где $\boldsymbol{h} = \boldsymbol{x} - \boldsymbol{x}^\star = (h_1, h_2, \ldots, h_n)^\top$:

$$|h_i| \leq A_{x_i^\star} \quad (i = 1, 2, \ldots, n). \tag{1.4}$$

Определение погрешности значения функции по погрешностям ее аргументов (продолжение)

Пусть $f \in C^1$. Тогда

$$
\begin{aligned}
|y - y^\star| = |F(1) - F(0)| = |F'(\theta)|(1-0) = \\
= |\sum_{i=1}^n \frac{\partial f}{\partial x_i}(\boldsymbol{x}^\star + \theta \boldsymbol{h}) h_i| \le \\
\le \sum_{i=1}^n |\frac{\partial f}{\partial x_i}(\boldsymbol{x}^\star + \theta \boldsymbol{h})| A_{x_i^\star} \le \\
\le \sum_{i=1}^n M_i A_{x_i^\star} = A_{y^\star},
\end{aligned}
\tag{1.5}
$$

где $\theta \in [0,1]$, $M_i = \max_{\theta \in [0,1]} |\frac{\partial f}{\partial x_i}(\boldsymbol{x}^\star + \theta \boldsymbol{h})|$ $(i = 1, 2, \ldots, n)$.

Если $A_{x_i^\star} \ll 1$, то $\frac{\partial f}{\partial x_i}(\boldsymbol{x}^\star + \theta \boldsymbol{h}) \approx \frac{\partial f}{\partial x_i}(\boldsymbol{x}^\star)$ $(i = 1, 2, \ldots, n)$.

$$
A_{y^\star} = \sum_{i=1}^n |\frac{\partial f}{\partial x_i}(\boldsymbol{x}^\star)| A_{x_i^\star} \tag{1.6}
$$

1.5. Погрешности арифметических операций

Пусть $x_i, x_i^\star : A_{x_i^\star}, \Delta_{x_i^\star}\ (i = 1, 2, \ldots, n)$

I. СЛОЖЕНИЕ

Абсолютная погрешность

Поскольку $|\sum_{i=1}^n x_i - \sum_{i=1}^n x_i^\star| \le \sum_{i=1}^n |x_i - x_i^\star| \le \sum_{i=1}^n A_{x_i^\star}$, то

$$A_{\sum_{i=1}^n x_i^\star} = \sum_{i=1}^n A_{x_i^\star} \tag{1.7}$$

Относительная погрешность. Случай 1: $x_i^\star > (<) 0\ (i = 1, 2, \ldots, n)$

В силу (1.3), (1.7)

$$\Delta_{\sum_{i=1}^n x_i^\star} = \frac{A_{\sum_{i=1}^n x_i^\star}}{|\sum_{i=1}^n x_i^\star|} = \frac{\sum_{i=1}^n A_{x_i^\star}}{\sum_{i=1}^n |x_i^\star|} = \frac{\sum_{i=1}^n \Delta_{x_i^\star} |x_i^\star|}{\sum_{i=1}^n |x_i^\star|} \le \max_i \Delta_{x_i^\star}.$$

$$\min_i \Delta_{x_i^\star} \le \Delta_{\sum_{i=1}^n x_i^\star} \le \max_i \Delta_{x_i^\star} \tag{1.8}$$

Погрешности арифметических операций

Относительная погрешность. Случай 2: $x_i^\star$ разных знаков $(i = 1, 2, \ldots, n)$

Частный случай: $y^\star = x_1^\star - x_2^\star$, где $x_1^\star > 0, x_2^\star > 0$.

В силу (1.3), (1.7) $\Delta_{y^\star} = \frac{A_{x_1^\star - x_2^\star}}{|x_1^\star - x_2^\star|} = \frac{A_{x_1^\star} + A_{x_2^\star}}{|x_1^\star - x_2^\star|} = \frac{x_1^\star \Delta_{x_1^\star} + x_2^\star \Delta_{x_2^\star}}{|x_1^\star - x_2^\star|}$.

При вычитании двух положительных близких чисел может произойти значительное увеличение относительной погрешности.

Пример. $x_1 = \sqrt{11}$, $x_1^\star = \underline{3,32}$; $x_2 = \sqrt{10}$, $x_2^\star = \underline{3,16}$

Пусть $A_{x_1^\star} = A_{x_2^\star} = 0,005$. Тогда $A_{x_1^\star - x_2^\star} = A_{x_1^\star} + A_{x_2^\star} = 0,01$.

$$y^\star = x_1^\star - x_2^\star = \underline{0,16} \tag{1.9}$$

$$y^\star = \sqrt{11} - \sqrt{10} = \frac{(\sqrt{11} - \sqrt{10})(\sqrt{11} + \sqrt{10})}{\sqrt{11} + \sqrt{10}} = \tag{1.10}$$

$$= \frac{11 - 10}{\sqrt{11} + \sqrt{10}} = \frac{1}{3,32 + 3,16} = \frac{1}{6,48} = \underline{0,154}$$

[1] Самостоятельно ответить на вопрос: Почему в результате, полученном с помощью формулы (1.10), все цифры верные?

Относительная погрешность операции умножения

II. УМНОЖЕНИЕ

$y = \prod_{i=1}^{n} x_i = f(x_1, x_2, \ldots, x_n)$.
Пусть $x_i^\star \neq 0, A_{x_i^\star}, \Delta_{x_i^\star}$ $(i = 1, 2, \ldots, n)$ и
$y^\star = \prod_{i=1}^{n} x_i^\star = f(x_1^\star, x_2^\star, \ldots, x_n^\star)$.

Нетрудно заметить, что

$$\frac{\partial f}{\partial x_i}(x_1, x_2, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n)}{x_i} \quad \forall i = 1, 2, \ldots, n \qquad (1.11)$$

Тогда, в силу (1.3), (1.6) и (1.11), имеем

$$\Delta_{y^\star} = \frac{A_{y^\star}}{|\prod_{i=1}^{n} x_i^\star|} = \frac{\sum_{i=1}^{n} \frac{|f(x_1^\star, x_2^\star, \ldots, x_n^\star)|}{|x_i^\star|} A_{x_i^\star}}{|f(x_1^\star, x_2^\star, \ldots, x_n^\star)|} =$$

$$= \frac{|f(x_1^\star, x_2^\star, \ldots, x_n^\star)| \sum_{i=1}^{n} \frac{A_{x_i^\star}}{|x_i^\star|}}{|f(x_1^\star, x_2^\star, \ldots, x_n^\star)|} = \sum_{i=1}^{n} \frac{A_{x_i^\star}}{|x_i^\star|} = \qquad (1.12)$$

$$= \sum_{i=1}^{n} \Delta_{x_i^\star}$$

Относительная погрешность операции деления

III. ДЕЛЕНИЕ

$y = \frac{x_1}{x_2} = f(x_1, x_2)$.

Пусть $x_i^\star \neq 0, A_{x_i^\star}, \Delta_{x_i^\star}$ $(i = 1, 2)$ и $y^\star = \frac{x_1^\star}{x_2^\star}$.

$$\frac{\partial f}{\partial x_1}(x_1, x_2) = \frac{1}{x_2}, \quad \frac{\partial f}{\partial x_2}(x_1, x_2) = -\frac{x_1}{x_2^2} \tag{1.13}$$

Тогда, в силу (1.3), (1.6) и (1.13), имеем

$$\begin{aligned} \Delta_{y^\star} &= \frac{A_{y^\star}}{|\frac{x_1^\star}{x_2^\star}|} = \frac{\frac{1}{|x_2^\star|} A_{x_1^\star} + \frac{|x_1^\star|}{(x_2^\star)^2} A_{x_2^\star}}{|\frac{x_1^\star}{x_2^\star}|} = \\ &= \frac{A_{x_1^\star}}{|x_1^\star|} + \frac{A_{x_2^\star}}{|x_2^\star|} = \\ &= \Delta_{x_1^\star} + \Delta_{x_2^\star} \end{aligned} \tag{1.14}$$

$\boxtimes$

▲2 ТЕМА 2. Ускорение сходимости числовых рядов

2.1. Постановка задачи

$$\sum_{n=1}^{\infty} a_n \ : \quad \exists S = \lim_{N\to\infty} S_N, \quad S_N = \sum_{n=1}^{N} a_n \tag{2.1}$$

Определить $\widetilde{S} \approx S$: $|S - \widetilde{S}| \le \varepsilon$, где $\varepsilon > 0$ — заданная точность.

Пусть $\varepsilon = \varepsilon_1 + \varepsilon_2$, где ε_1 — погрешность метода:

$$\widetilde{S} = \sum_{n=1}^{\widetilde{N}} a_n \tag{2.2}$$

$$\widetilde{N} : |S - \widetilde{S}| = |\sum_{n=1}^{\infty} a_n - \sum_{n=1}^{\widetilde{N}} a_n| = |\sum_{n=\widetilde{N}+1}^{\infty} a_n| \le \varepsilon_1 \tag{2.3}$$

ε_2 — вычислительная погрешность:

$$A_{a_n^\star} = \frac{\varepsilon_2}{\widetilde{N}} \quad n = 1, 2, \ldots, \widetilde{N} \tag{2.4}$$

Тогда $|S - \widetilde{S}^\star| = |S - \widetilde{S} + \widetilde{S} - \widetilde{S}^\star| \le |S - \widetilde{S}| + |\widetilde{S} - \widetilde{S}^\star| \le \varepsilon_1 + \varepsilon_2 = \varepsilon$

2.2. Пример

$$\sum_{n=1}^{\infty} a_n, \quad a_n = \frac{1}{n^2+1} \quad \forall n \in \mathbb{N} \tag{2.5}$$

Пусть $\varepsilon = 10^{-5}$ и, например, $\varepsilon = \varepsilon_1 + \varepsilon_2$, где $\varepsilon_1 = \varepsilon_2 = 0,5 \cdot 10^{-5}$. Согласно (2.3), оценим остаток ряда (2.5):

$$\left| \sum_{n=N+1}^{\infty} \frac{1}{n^2+1} \right| = \sum_{n=N+1}^{\infty} \frac{1}{n^2+1} \leq \int_N^{\infty} \frac{1}{x^2+1} dx \tag{2.6}$$

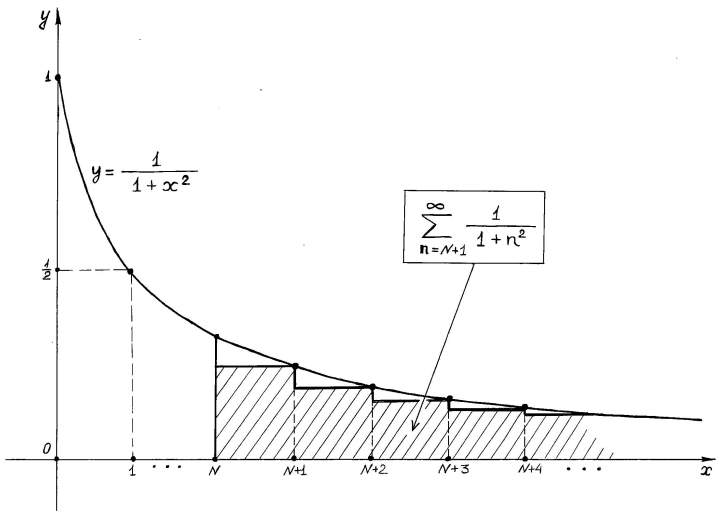


Рис. 2: Геометрическая интерпретация суммы ряда $\sum_{n=N+1}^{\infty} \frac{1}{n^2+1}$.

Пример (продолжение)

Следовательно,

$$\left|\sum_{n=N+1}^{\infty}\frac{1}{n^2+1}\right| \le \int_N^{\infty}\frac{1}{x^2+1}dx \le \\ \le \int_N^{\infty}\frac{1}{x^2}dx = -\frac{1}{x}\Big|_N^{\infty} = \frac{1}{N} \le \varepsilon_1 = \frac{1}{2}\cdot 10^{-5} \tag{2.7}$$

Из (2.7) следует, что $N \ge 2\cdot 10^5$. Откуда

$\widetilde{N} = 2\cdot 10^5$ и $A_{a_n^\star} = \frac{\varepsilon_2}{\widetilde{N}} = \frac{0{,}5\cdot 10^{-5}}{2\cdot 10^5} = 0{,}25\cdot 10^{-10}$.

Вывод

Для того, чтобы вычислить сумму ряда (2.5) с точностью $\varepsilon = 10^{-5}$, достаточно в его частичной сумме просуммировать $2\cdot 10^5$ слагаемых. При этом приближенное значение каждого из слагаемых должно содержать не менее 10 верных цифр после запятой в его позиционной записи.

2.3. Метод Куммера

Увеличить скорость сходимости числового ряда значит преобразовать общий член ряда так, чтобы для вычисления приближенного значения суммы исходного ряда с заданной точностью потребовалось меньшее количество слагаемых в его частичной сумме.

Определение 2.1.

Числовой ряд $\sum_{n=1}^{\infty} b_n$ называется эталонным для ряда (2.1), если

1. $\exists \sum_{n=1}^{\infty} b_n = B < \infty$;
2. $\exists \lim_{n\to\infty} \frac{a_n}{b_n} = \lambda$: $0 < |\lambda| < \infty$.[a]

[a] Общие члены a_n и b_n соответствующих рядов асимптотически эквивалентны.

Метод Куммера

Преобразование Куммера

Пусть ряд $\sum_{n=1}^{\infty} b_n$ — эталонный для ряда (2.1). Тогда

$$\begin{aligned} S = \sum_{n=1}^{\infty} a_n &= \sum_{n=1}^{\infty} (a_n - \lambda b_n + \lambda b_n) = \\ &= \sum_{n=1}^{\infty} (a_n - \lambda b_n) + \lambda \sum_{n=1}^{\infty} b_n = \\ &= \sum_{n=1}^{\infty} b_n \left(\frac{a_n}{b_n} - \lambda \right) + \lambda B \end{aligned} \tag{2.8}$$

Вспомогательный ряд

$$\sum_{n=1}^{\infty} c_n = C, \quad c_n = b_n \left(\frac{a_n}{b_n} - \lambda \right) \quad \forall n \in \mathbb{N} \tag{2.9}$$

$$S = \sum_{n=1}^{\infty} a_n = C + \lambda B \tag{2.10}$$

При $n \to \infty$ $c_n \to 0$ “быстрее”, чем a_n и b_n.

2.4. Примеры эталонных рядов

Класс рядов:

$a_n = \frac{\alpha_0}{n^m} + \frac{\alpha_1}{n^{m+1}+\ldots}$, $\alpha_i \in \mathbb{R}\ \forall i$, $m \geq 2$;

$a_n = \frac{\beta_0 n^p + \beta_1 n^{p-1} + \ldots + \beta_{p-1} n + \beta_p}{\gamma_0 n^q + \gamma_1 n^{q-1} + \ldots + \gamma_{q-1} n + \gamma_q}$, $\beta_i, \gamma_j \in \mathbb{R}\ \forall i, j$, $q \geq p+2$.

Эталонные ряды

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

$$\sum_{n=1}^{\infty} \frac{1}{n^3} = 1,2020569032\ldots$$

$$\sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90} \tag{2.11}$$

$$\sum_{n=1}^{\infty} \frac{1}{n(n+m)} = \frac{1}{m}\left(1 + \frac{1}{2} + \ldots + \frac{1}{m}\right), \quad m \in \mathbb{N}$$

$$\sum_{n=1}^{\infty} \frac{1}{(n+k)(n+k+1)\ldots(n+k+l)} = \frac{1}{l(k+1)\ldots(k+l)}, \quad k, l \in \mathbb{N}$$

2.5. Пример ускорения сходимости числового ряда

Рассмотрим ряд (2.5): $S = \sum_{n=1}^{\infty} a_n$, $a_n = \frac{1}{n^2+1}$.
Пусть $\varepsilon = 10^{-5}$ и $\varepsilon = \varepsilon_1 + \varepsilon_2$, где $\varepsilon_1 = \varepsilon_2 = 0,5 \cdot 10^{-5}$.
Ранее уже установлено, что $S \approx \widetilde{S} = \sum_{n=1}^{\widetilde{N}} a_n$, где $\widetilde{N} = 2 \cdot 10^5$, и $A_{a_n^\star} = 0,25 \cdot 10^{-10}$. Требуется ускорить сходимость ряда (2.5) с помощью метода Куммера.

Очевидно, что эталонным рядом для ряда (2.5) является $\sum_{n=1}^{\infty} \frac{1}{n^2}$.
Для ряда (2.5) построим более "эффективный" эталонный ряд.
Для этого, используя известное разложение в ряд Тейлора в окрестности нуля функции $\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots$, можно преобразовать общий член ряда (2.5) следующим образом

$$a_n = \frac{1}{n^2+1} = \frac{1}{n^2} \cdot \frac{1}{1+\frac{1}{n^2}} =$$

$$= \frac{1}{n^2}\left(1 - \frac{1}{n^2} + \frac{1}{n^4} - \frac{1}{n^6} + \ldots\right) = \qquad (2.12)$$

$$= \left(\frac{1}{n^2} - \frac{1}{n^4}\right) + \frac{1}{n^6} - \ldots$$

Пример ускорения сходимости числового ряда (продолжение)

Из (2.12) следует, что в качестве эталонного для ряда (2.5) может быть использован ряд $\sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} \left(\frac{1}{n^2} - \frac{1}{n^4}\right) = B = \frac{\pi^2}{6} - \frac{\pi^4}{90}$.

Здесь $\lim_{n\to\infty} \frac{a_n}{b_n} = \lim_{n\to\infty} \frac{\frac{1}{n^2+1}}{\frac{1}{n^2}-\frac{1}{n^4}} = \lim_{n\to\infty} \frac{n^4}{n^4-1} = 1 = \lambda.$

Преобразование Куммера

$$S = \sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} (a_n - \lambda b_n + \lambda b_n) =$$

$$= \sum_{n=1}^{\infty} \left(\frac{1}{n^2+1} - \frac{1}{n^2} + \frac{1}{n^4}\right) + \left(\frac{\pi^2}{6} - \frac{\pi^4}{90}\right) = \qquad (2.13)$$

$$= \sum_{n=1}^{\infty} \frac{1}{n^2(n^4+1)} + \left(\frac{\pi^2}{6} - \frac{\pi^4}{90}\right)$$

Вспомогательный ряд

$$\sum_{n=1}^{\infty} c_n = C, \quad c_n = \frac{1}{n^2(n^4+1)} \quad \forall n \in \mathbb{N} \qquad (2.14)$$

Пример ускорения сходимости числового ряда (продолжение)

Аналогично (2.6) оценим модуль остатка вспомогательного ряда (2.14):

$$\left|\sum_{n=N+1}^{\infty}\frac{1}{n^2(n^4+1)}\right| = \sum_{n=N+1}^{\infty}\frac{1}{n^2(n^4+1)} \leq$$

$$\leq \int_N^{\infty}\frac{1}{x^2(x^4+1)}dx \leq \int_N^{\infty}\frac{1}{x^6}dx = \qquad (2.15)$$

$$= -\frac{1}{5x^5}\Big|_N^{\infty} = \frac{1}{5N^5} \leq \varepsilon_1 = \frac{1}{2}\cdot 10^{-5}$$

Из (2.15) следует, что $N \geq \left(\frac{2}{5}\right)^{\frac{1}{5}} \cdot 10$. Откуда $\widetilde{N} = 10$ и $A_{c_n^\star} = \frac{\varepsilon_2}{\widetilde{N}} = \frac{0{,}5\cdot 10^{-5}}{10} = 0,5\cdot 10^{-6}$.

Результат

$S \approx \widetilde{S}^\star = \widetilde{C}^\star + \left(\frac{\pi^2}{6} - \frac{\pi^4}{90}\right)$, $\widetilde{N} = 10$, $A_{c_n^\star} = 0,5\cdot 10^{-6}$,

где $\widetilde{C}^\star = \sum_{n=1}^{\widetilde{N}} c_n^\star$, $c_n^\star \approx \frac{1}{n^2(n^4+1)}$ $\forall n \in \mathbb{N}$.

ТЕМА 3. Численное решение уравнений

3.1. Постановка задачи

$f \in C\left(\mathrm{D}[f]\right), \mathrm{D}[f] \subseteq \mathbb{R}$

$$f(x) = 0 \tag{3.1}$$

$$\xi \in \mathrm{D}[f]: \quad f(\xi) = 0 \tag{3.2}$$

ξ — корень уравнения (3.1)

Проблемы

1. Локализация корней уравнения (3.1) (отделение корней) — $[a, b] \subseteq \mathrm{D}[f]: \quad \xi \in [a, b]$;
2. Выбор метода приближенного решения уравнения (3.1) — построение последовательности $\{x_n\}_{n=0}^{\infty}$: $\lim_{n\to\infty} x_n = \xi$;
3. Оценка погрешности метода — построение оценки вида $|\xi - x_n| \leq C(x_0, n, f)$;
4. Оценка скорости сходимости метода — построение оценки вида $|\xi - x_n| \leq C(\xi, n, f)$

3.2. Локализация корней уравнения $f(x)=0$

Теорема

Если функция непрерывна на некотором отрезке и на концах этого отрезка принимает значения противоположных знаков, то существует точка, в которой значение функции равно нулю[a].

[a] Следствие теоремы Больцано-Коши о промежуточном значении

Теорема

Пусть $f \in C^1(\mathrm{D}[f]), [a,b] \subseteq \mathrm{D}[f]$:
a) $f(a)f(b) \le 0$; b) $f'(x) > (<)\ 0 \quad \forall x \in [a,b]$.

Тогда $\exists!\ \xi \in [a,b] : \ f(\xi) = 0$.

Утверждение

Пусть $P_n(x)$ — многочлен степени n: $P_n^{(k)}(\bar{c}) > 0\ (k = 0, 1, \ldots, n)$.
Тогда $P_n(c) > 0 \quad \forall c > \bar{c}$.

Доказательство: $P_n(x) = \sum_{k=0}^{n} \frac{1}{k!} P_n^{(k)}(\bar{c})(x-\bar{c})^k$. Отсюда $\forall c > \bar{c}$

$$P_n(c)=P_n(\bar{c})+P_n^{(1)}(\bar{c})(c-\bar{c})+\tfrac{1}{2!}P_n^{(2)}(\bar{c})(c-\bar{c})^2+\ldots+\tfrac{1}{n!}P_n^{(n)}(\bar{c})(c-\bar{c})^n>0.$$

3.3. Методы приближенного решения уравнения $f(x)=0$

I. Метод дихотомии (метод деления отрезка пополам)

$f \in C([a,b])$, где $[a,b] \subseteq \mathrm{D}[f]$: $f(a)f(b) < 0$

Численная процедура:

- $[a_0, b_0] = [a, b]$, $\quad x_0 = \frac{a_0+b_0}{2}$;
- $[a_1, b_1] = \begin{cases} [a_0, x_0], & f(a_0)f(x_0)<0 \\ [x_0, b_0], & f(x_0)f(b_0)<0 \end{cases}$, $\quad x_1 = \frac{a_1+b_1}{2}$;
- $\ldots$
- $[a_n, b_n] = \begin{cases} [a_{n-1}, x_{n-1}], & f(a_{n-1})f(x_{n-1})<0 \\ [x_{n-1}, b_{n-1}], & f(x_{n-1})f(b_{n-1})<0 \end{cases}$, $\quad x_n = \frac{a_n+b_n}{2}$;[a]
- $\ldots$

Последовательность $\{x_n\}_{n=0}^{\infty}$ сходится, т.к. она фундаментальная и пространство $\mathbb{R}$ — полное.

[a] если $f(x_n) = 0$, то $\xi = x_n$ и процедура прерывается

Метод дихотомии

Оценка погрешности и скорости сходимости метода

$$|\xi - x_n| \leq \frac{b-a}{2^n} \tag{3.3}$$

Метод дихотомии сходится как геометрическая прогрессия со знаменателем $\frac{1}{2}$.

Оценка числа итераций достаточного для достижения заданной точности ε

Требуется определить корень уравнения (3.1) с заданной точностью, т.е. $|\xi - \widetilde{\xi}| \leq \varepsilon$. Пусть $\widetilde{\xi} = x_n$. Тогда из (3.3) следует

$$n \geq \log_2 \frac{b-a}{\varepsilon} \tag{3.4}$$

В методе дихотомии используется только знак значений функции f, непрерывной на отрезке локализации $[a, b]$ корня ξ уравнения (3.1).

II.a. Метод неподвижных хорд

$f \in C([a,b])$, где $[a,b] \subseteq \mathrm{D}[f]$: $f(a)f(b) < 0$

Пусть $x_0 = a$,[2] $x_1 = b$, ..., $x_n \in [a,b]$ — текущее приближение ξ. Уравнение прямой в отрезках

$$\frac{y - f(x_n)}{x - x_n} = \frac{f(x_0) - f(x_n)}{x_0 - x_n} \tag{3.5}$$

геометрически задает хорду, стягивающую текущую ("плавающую") точку $(x_n, f(x_n))$ и зафиксированную точку $(x_0, f(x_0))$ на графике функции $y = f(x)$. Точка пересечения этой хорды с осью абсцисс геометрически определяет следующее $(n+1$-ое) приближение x_{n+1} корня ξ. Аналитически значение x_{n+1} определяется как решение следующего из (3.5) уравнения

$$y = \frac{f(x_0) - f(x_n)}{x_0 - x_n}(x - x_n) + f(x_n) = 0. \tag{3.6}$$

[2] Ответ на вопрос какой из концов отрезка $[a,b]$ следует взять в качестве начального приближения x_0 зависит от свойств функции f.

Метод неподвижных хорд (продолжение)

Решением уравнения (3.6) является $x = x_n - \frac{f(x_n)}{f(x_n)-f(x_0)}(x_n - x_0)$.

Численная процедура метода неподвижных хорд:

$$x_{n+1} = x_n - \frac{f(x_n)}{f(x_n) - f(x_0)}(x_n - x_0), \quad n = 0, 1, 2, \ldots \tag{3.7}$$

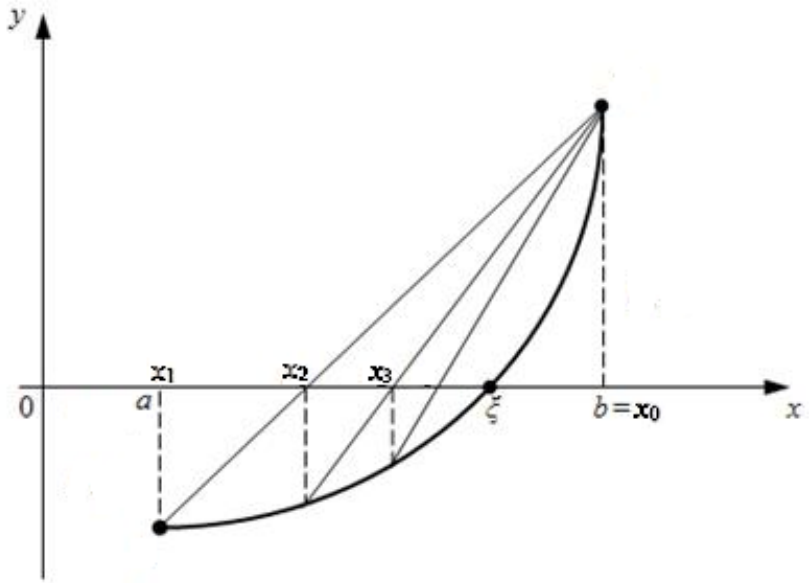


Рис. 3: Геометрическая интерпретация метода неподвижных хорд.

Метод неподвижных хорд (продолжение)

Пусть $f \in C^1([a,b])$. Тогда, используя формулу Лагранжа,

$$f(\xi) - f(x_n) = f'(\theta)(\xi - x_n), \quad \theta \in [\xi, x_n] \tag{3.8}$$

Из (3.8) с учетом того, что $f(\xi) = 0$, следует

$$\xi - x_n = -\frac{f(x_n)}{f'(\theta)}.$$

Откуда

Оценка погрешности метода

Пусть (a) $f \in C^1([a,b])$, (b) $m = \min_{x\in[a,b]} |f'(x)| > 0$. Тогда

$$|\xi - x_n| \le \frac{|f(x_n)|}{m} \tag{3.9}$$

3

[3] Следует ли из того, что величина $|f(x_n)|$ достаточно мала, что приближение x_n уже достаточно близко к корню ξ?

II.6. Метод подвижных хорд

В методе неподвижных хорд (3.7) неподвижная точка $(x_0, f(x_0))$ “размораживается” и очередное приближение x_{n+1} корня ξ уравнения (3.1) строится на основе текущего x_n и ему предшествовавшего x_{n-1} приближений

Численная процедура метода подвижных хорд:

$$x_{n+1} = x_n - \frac{f(x_n)}{f(x_n) - f(x_{n-1})}\,(x_n - x_{n-1}), \quad n = 0, 1, 2, \ldots \qquad (3.10)$$

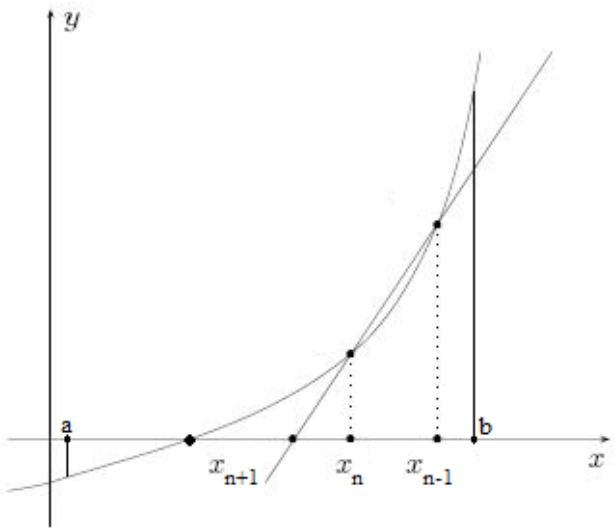


Рис. 4: Геометрическая интерпретация метода подвижных хорд.

Методы неподвижных и подвижных хорд

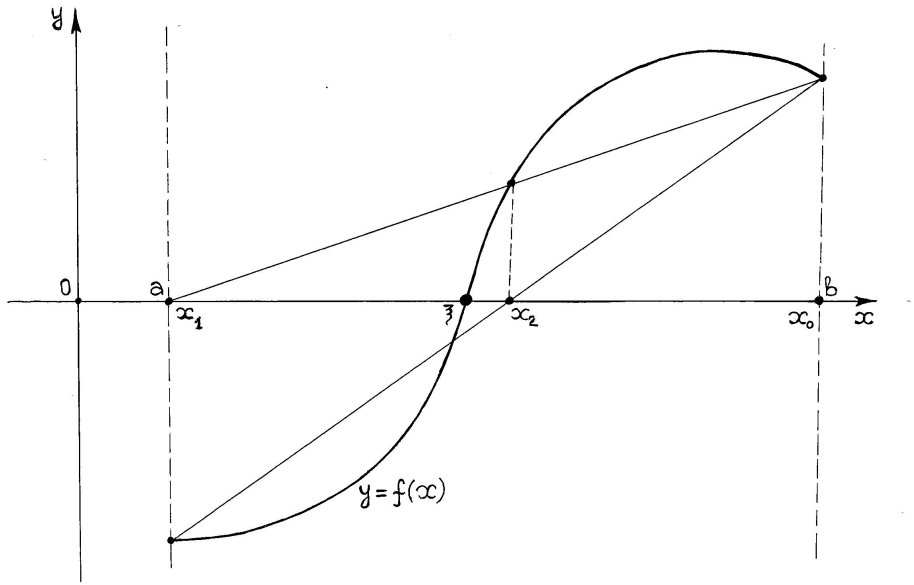


Рис. 5: Зацикливание метода неподвижных хорд.

В методе хорд, в отличие от метода дихотомии, уже используются значения функции f, непрерывной на отрезке $[a, b]$ локализации корня ξ уравнения (3.1)[4].

⊠

[4] Используя геометрическую интерпретацию метода хорд вывести правило выбора одного из концов отрезка локализации корня ξ в качестве его начального приближения x_0.

▲3 III. Метод Ньютона (метод касательных)

$f \in C^1([a,b])$, где $[a,b] \subseteq \mathrm{D}[f]$: $f(a)f(b) < 0$

Пусть $x_0 = a$,[5] $x_1, \ldots, x_n \in [a,b]$ — текущее приближение ξ. Уравнение

$$y = f(x_n) + f'(x_n)(x - x_n) \tag{3.11}$$

геометрически задает прямую, касательную к графику функции $y = f(x)$ в точке $(x_n, f(x_n))$. Точка пересечения этой касательной с осью абсцисс геометрически определяет следующее $(n+1$-ое$)$ приближение x_{n+1} корня ξ. Аналитически значение x_{n+1} определяется как решение уравнения

$$y = f(x_n) - f'(x_n)(x - x_n) = 0. \tag{3.12}$$

Решением уравнения (3.12) является $x = x_n - \dfrac{f(x_n)}{f'(x_n)}$.

[5] Ответ на вопрос какой из концов отрезка $[a,b]$ следует взять в качестве начального приближения x_0 зависит от свойств функции f.

Метод Ньютона (продолжение)

Численная процедура метода Ньютона:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, \ldots \tag{3.13}$$

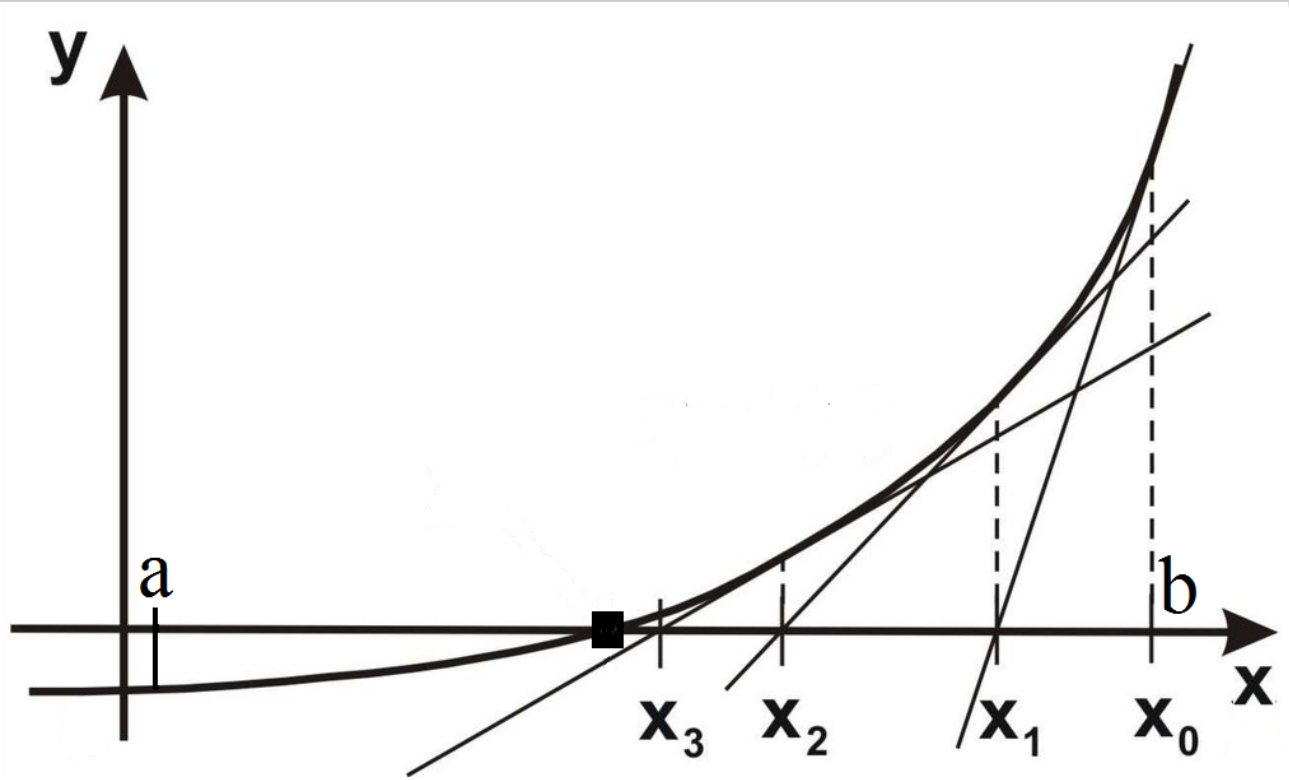


Рис. 6: Геометрическая интерпретация метода Ньютона.

Метод Ньютона (продолжение)

Пусть $f \in C^1([a,b])$ и $m = \min_{x\in[a,b]} |f'(x)| > 0$. Тогда из (3.9) следует, что

$$|\xi - x_{n+1}| \le \frac{|f(x_{n+1})|}{m}.$$

Если $f \in C^2([a,b])$, то

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

где $\theta \in [x_n, x_{n+1}]$.
В силу (3.13), $f(x_n) + f'(x_n)(x_{n+1} - x_n) = 0$. Следовательно,

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

Таким образом,

$$|\xi - x_{n+1}| \le \frac{|f''(\theta)|}{2m}(x_{n+1} - x_n)^2.$$

Метод Ньютона (продолжение)

Оценка погрешности метода

Пусть (a) $f \in C^2([a,b])$, (b) $m = \min_{x\in[a,b]} |f'(x)| > 0$. Тогда

$$|\xi - x_{n+1}| \le \frac{M}{2m}(x_{n+1} - x_n)^2, \tag{3.14}$$

где $M = \max_{x\in[a,b]} |f''(x)|$.

Если

$$\frac{M}{2m}|x_{n+1} - x_n| < 1, \tag{3.15}$$

то $|\xi - x_{n+1}| < |x_{n+1} - x_n|$.
Следовательно, когда значения x_n стабилизируются ($|x_{n+1} - x_n| < \delta$) и начнет выполняться условие (3.15) ($\delta \le \frac{2m}{M}$) тогда $|x_{n+1} - x_n| < \delta \Rightarrow |\xi - x_{n+1}| < \delta$.

Если $\varepsilon > 0$ — заданная точность вычисления корня ξ уравнения (3.1), то условие $|x_{n+1} - x_n| < \delta$ при $0 < \delta \le \min\{\varepsilon, \frac{2m}{M}\}$ можно использовать в качестве условия остановки численной процедуры метода Ньютона.

Метод Ньютона (продолжение)

Для того, чтобы оценить скорость сходимости метода требуется оценить $|\xi - x_{n+1}|$ через $|\xi - x_n|$.
В силу (3.13)

$$\xi - x_{n+1} = \xi - x_n + \frac{f(x_n)}{f'(x_n)} = \frac{f(x_n) + f'(x_n)(\xi - x_n)}{f'(x_n)}. \qquad (3.16)$$

Пусть $f \in C^2([a,b])$. Тогда

$$f(\xi) = f(x_n) + f'(x_n)(\xi - x_n) + \frac{1}{2} f''(\theta)(\xi - x_n)^2 = 0,$$

где $\theta \in [x_n, x_{n+1}]$. Отсюда

$$f(x_n) + f'(x_n)(\xi - x_n) = -\frac{1}{2} f''(\theta)(\xi - x_n)^2. \qquad (3.17)$$

Таким образом, из (3.16) и (3.17) следует

Метод Ньютона (продолжение)

$$\xi - x_{n+1} = -\frac{1}{2}\frac{f''(\theta)}{f'(x_n)}(\xi - x_n)^2 \tag{3.18}$$

Откуда

Оценка скорости сходимости метода

Пусть (a) $f \in C^2([a,b])$, (b) $m = \min_{x\in[a,b]} |f'(x)| > 0$. Тогда

$$|\xi - x_{n+1}| \le \frac{M}{2m}(\xi - x_n)^2, \tag{3.19}$$

где $M = \max_{x\in[a,b]} |f''(x)|$.

В обозначениях $r_n = \frac{M}{2m}|\xi - x_n|$ неравенство (3.19) можно переписать в виде

$$r_{n+1} \le r_n^2. \tag{3.20}$$

Следовательно, если $r_0 < 1$, то $r_n \to 0$ при $n \to \infty$ и при этом, в силу (3.20), скорость сходимости квадратичная. Более того, получено условие на выбор достаточного r_0, обеспечивающего сходимость.

Достаточные условия сходимости методов хорд и Ньютона

Теорема 3.1.

Пусть выполнены следующие условия:

(a) $f \in C^2([a,b])$;

(b) $[a,b] \subseteq \mathrm{D}[f]$: $f(a)f(b) < 0$;

(c) $f'(x) > (<)0 \quad \forall x \in [a,b]$;

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

(e) $x_0 \in [a,b] : \quad f(x_0)f''(x_0) > 0.$

Тогда методы хорд (3.7), (3.10) и Ньютона (3.13) сходятся к корню $\xi \in [a,b]$ уравнения (3.1). Причем сходимость монотонная и методы хорд и Ньютона сходятся к точке ξ с разных сторон.

Можно доказать утверждение теоремы, используя геометрическую интерпретацию методов хорд и Ньютона[6].

[6] Можно ли из условий теоремы исключить условие (c)?

Метод Ньютона

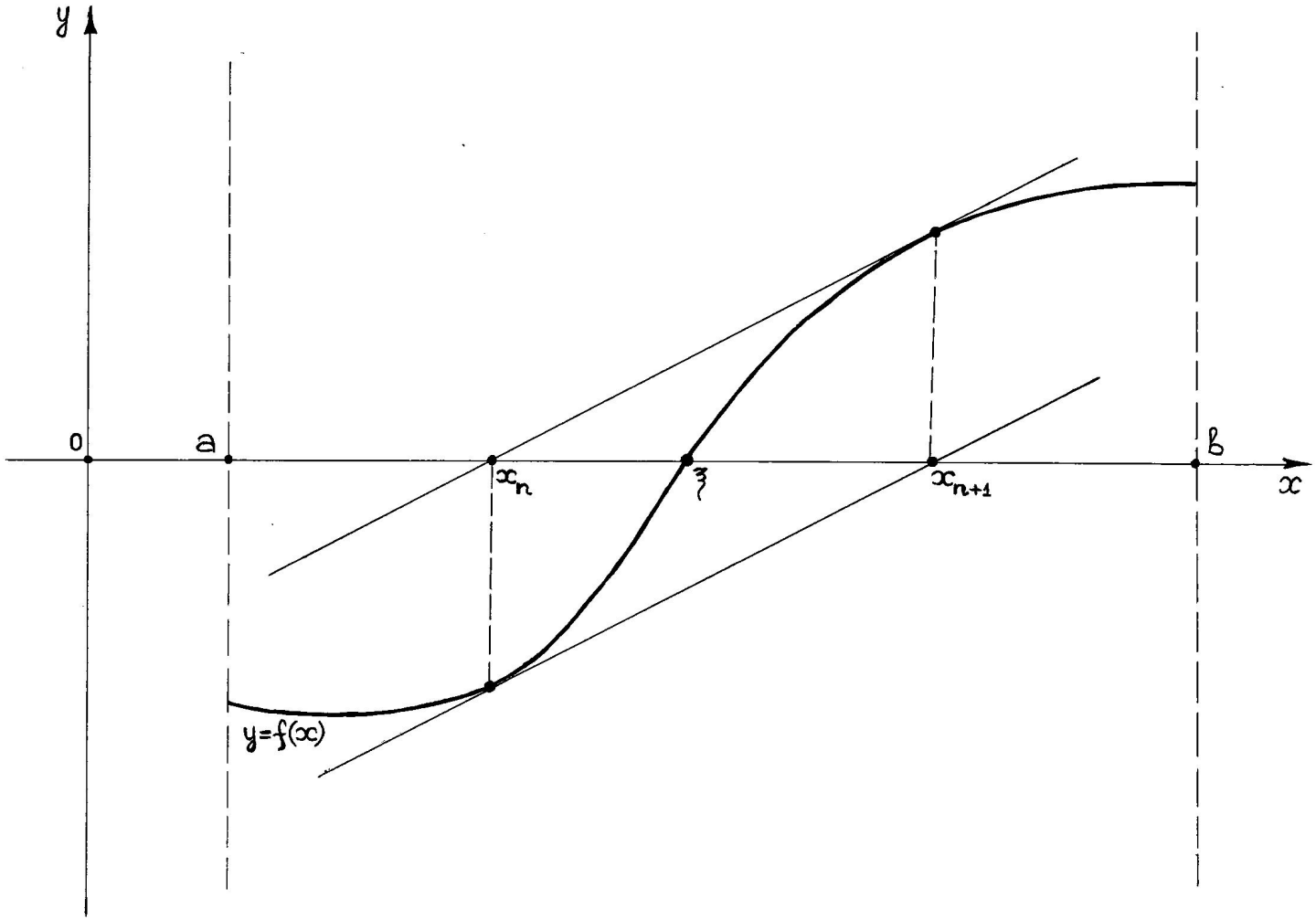


Рис. 7: Зацикливание метода Ньютона.

IV. Метод парабол

$f \in C^2([a,b])$: $M > 0$, где $M = \max_{x\in[a,b]} |f''(x)|$

Пусть $[a,b] \subseteq \mathrm{D}[f]$: $f(a)f(b) < 0$ и $x_n \in [a,b]$ — текущее приближение корня $\xi \in [a,b]$ уравнения (3.1). Например, для определенности можно считать, что $f(x_n) > 0$. Тогда уравнение

$$y = f(x_n) + f'(x_n)(x - x_n) - \frac{M}{2}(x - x_n)^2 \qquad (3.21)$$

геометрически задает параболу[7], которая расположена под графиком функции $y = f(x)$ и касается графика этой функции в точке $(x_n, f(x_n))$. Ветви параболы направлены вниз и пересекают ось абсцисс в двух точках. Эти точки геометрически определяют следующие $(n+1$-ое$)$ приближения x_{n+1} корня ξ.

[7] В случае, если $f(x_n) < 0$, то третье слагаемое в правой части уравнения (3.21) необходимо записывать со знаком “+”.

Метод парабол (продолжение)

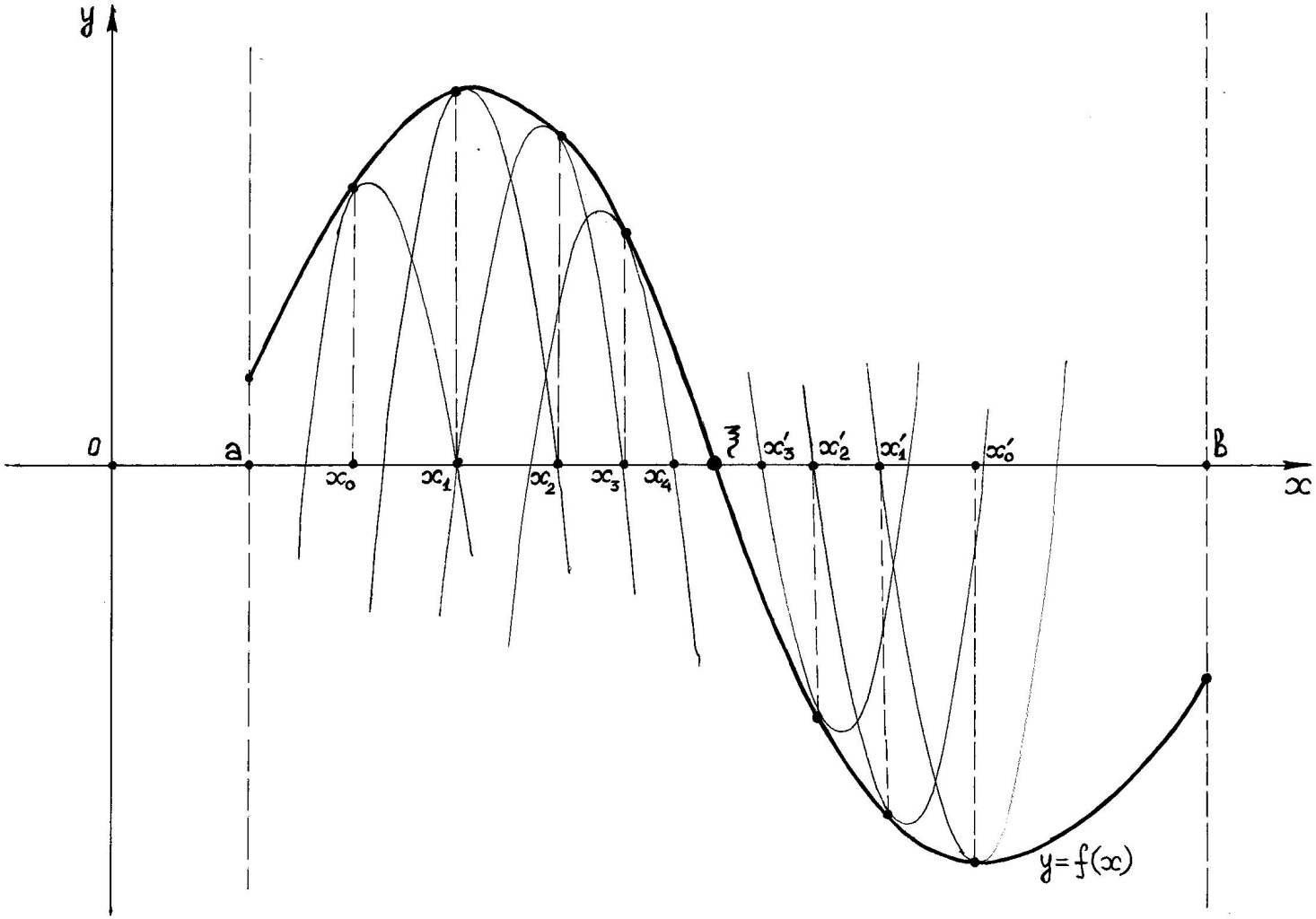


Рис. 8: Геометрическая интерпретация метода парабол.

Метод парабол (продолжение)

Аналитически значения x_{n+1} определяются как решения уравнения

$$y = f(x_n) + f'(x_n)(x - x_n) - \frac{M}{2}(x - x_n)^2 = 0, \tag{3.22}$$

которое удобно записать в виде

$$-\frac{M}{2}\Delta x_n^2 + f'(x_n)\Delta x_n + f(x_n) = 0, \tag{3.23}$$

где $\Delta x_n = x_{n+1} - x_n$.
Решения $\Delta x_n^{(\pm)}$ уравнения (3.23) задаются формулой

$$\Delta x_n^{(\pm)} = \frac{-f'(x_n) \pm \sqrt{\left(f'(x_n)\right)^2 + 2Mf(x_n)}}{-M}. \tag{3.24}$$

Метод парабол (продолжение)

Численная процедура метода парабол:

$$\begin{cases} \Delta x_n^{(\pm)} = \dfrac{-f'(x_n) \pm \sqrt{\left(f'(x_n)\right)^2 + 2M f(x_n)}}{-M} \\ x_{n+1} = x_n + \Delta x_n^{(\pm)} \end{cases} \qquad (3.25)$$

$$n = 0, 1, 2, \ldots$$

В результате последовательного применения (3.25), можно формально построить, по крайней мере, две числовые последовательности $x_n^{(\pm)}$ $(n = 0, 1, 2, \ldots)$, каждая из которых соответствует своему выбору знака ("+" или "-") в числителе дроби в правой части равенства (3.24). В общем случае каждая из этих последовательностей может как сходится к корню ξ уравнения (3.1), так и выходить за пределы отрезка $[a, b]$ его локализации[8].

В частности, если $f(x_0) > 0$ и $f'(x_0) < 0$, то последовательность $x_n^{(-)}$ $(n = 0, 1, 2, \ldots)$ является монотонно-возрастающей и при определенных условиях (см. рис. 8) сходится к искомому корню ξ.

[8] Самостоятельно проиллюстрировать геометрически.

V. Метод простой итерации

$\varphi \in C([a,b])$, $[a,b] \subseteq \mathrm{D}[\varphi]$: $\exists \xi \in [a,b] \quad \xi = \varphi(\xi)$, ξ — корень уравнения

$$x = \varphi(x) \tag{3.26}$$

Уравнение (3.26) может быть получено из уравнения $f(x) = 0$ (см. (3.1)) различными способами. Например, от уравнения (3.1) можно перейти к уравнению (3.26) следующим, вообще говоря, неравносильным образом

$$x = x + g(x)f(x), \tag{3.27}$$

где $g(x)$ — некоторая заданная функция[9].

Численная процедура метода простой итерации:

$$x_{n+1} = \varphi(x_n), \quad n = 0, 1, 2, \ldots \tag{3.28}$$

Метод простой итерации может как сходиться,так и расходиться.

[9] Способы задания функции $g(x)$ будут рассмотрены ниже.

Метод простой итерации

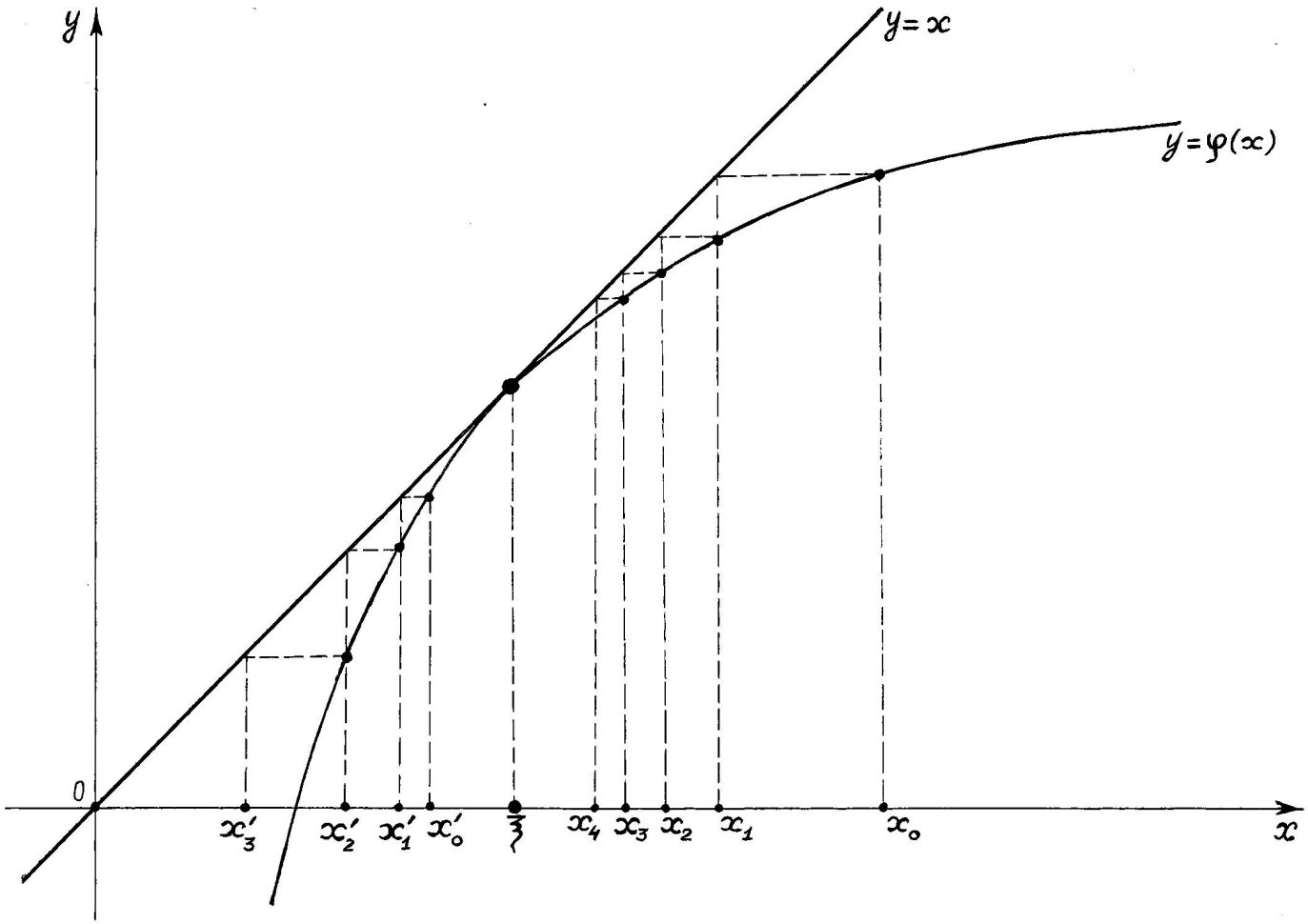


Рис. 9: Геометрическая интерпретация метода простой итерации.

Сходимость метода простой итерации

Определение 3.1.

Пусть R — метрическое пространство с заданной на нем метрикой ρ. Оператор $A : R \to R$ называется сжимающим, если

$$\exists q \in \mathbb{R}\ 0 < q < 1 \ : \forall x, y \in R \ \ \rho(Ax, Ay) \leq q\ \rho(x, y) \tag{3.29}$$

Теорема (Принцип сжимающих отображений)

Пусть R — полное метрическое пространство с заданной на нем метрикой ρ и оператор $A : R \to R$ — сжимающий.
Тогда уравнение $x = Ax$ имеет единственное решение $\xi \in R$:

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

где

$$\forall x_0 \in R, \quad x_{n+1} = Ax_n, \quad n = 0, 1, 2, \ldots.$$

Сходимость метода простой итерации (продолжение)

Теорема 3.2.

Пусть для функции $\varphi(x)$, $x \in [a-r, a+r]$ выполнены следующие условия:

(a) $\exists q \in \mathbb{R}\ 0 < q < 1\ : \forall x, y \in [a-r, a+r] \Rightarrow |\varphi(x) - \varphi(y)| \leq q|x-y|$;[a]

(b) $|\varphi(a) - a| \leq (1-q)r$.

Тогда $\exists! \xi \in [a-r, a+r]$: $\xi = \lim_{n\to\infty} x_n$, где $\forall x_0 \in [a-r, a+r]$, $x_{n+1} = \varphi(x_n)$, $n = 0, 1, 2, \ldots$.

[a] Функция φ — равномерно-липшицевая функция на отрезке $[a-r, a+r]$, q — константа Липшица.

Доказательство: Для того, чтобы для доказательства утверждения теоремы воспользоваться принципом сжимающих отображений достаточно показать, что $\varphi(x) \in [a-r, a+r]\ \ \forall x \in [a-r, a+r]$, то есть $|\varphi(x) - a| \leq r\ \ \forall x \in [a-r, a+r]$.

Действительно,
$|\varphi(x) - a| = |\varphi(x) - \varphi(a) + \varphi(a) - a| \leq |\varphi(x) - \varphi(a)| + |\varphi(a) - a| \leq q|x-a| + (1-q)r \leq qr + (1-q)r = r$.

Сходимость метода простой итерации (продолжение)

Следствие

Пусть $\varphi \in C^1([a,b])$, где $[a,b] \subseteq \mathrm{D}[\varphi]$:

(a) $\exists \xi \in [a,b] \;\; \xi = \varphi(\xi)$;

(b) $|\varphi'(\xi)| < 1$.

Тогда существует такое $\varepsilon > 0$, что для любого начального приближения $x_0 \in [\xi - \varepsilon, \xi + \varepsilon]$ метод простой итерации (3.28) сходится к решению ξ уравнения (3.26).

$\boxtimes$

▲4 Скорость сходимости и погрешность метода простой итерации

Пусть $\xi \in [a,b] \subseteq \mathrm{D}[\varphi]$, где ξ — уравнения (3.26), и функция φ равномерно-липшицева на отрезке $[a,b]$:

$$\exists q \in \mathbb{R}\ 0 < q < 1\ : \forall x, y \in [a,b] \Rightarrow |\varphi(x) - \varphi(y)| \leq q|x - y| \quad (3.30)$$

Тогда, в силу определения корня ξ уравнения (3.26), численной процедуры (3.28) метода простой итерации и (3.30)

$$\begin{aligned} |\xi - x_n| &= |\varphi(\xi) - \varphi(x_{n-1})| \leq q|\xi - x_{n-1}| = \\ &= q|\varphi(\xi) - \varphi(x_{n-2})| \leq q^2|\xi - x_{n-2}| = \\ &= \ldots = q^{n-1}|\varphi(\xi) - \varphi(x_1)| \leq \\ &\leq q^n|\xi - x_0|. \end{aligned} \quad (3.31)$$

Скорость сходимости метода простой итерации

Оценка скорости сходимости метода

Пусть $\xi \in [a, b] \subseteq \mathrm{D}[\varphi]$ и выполнено условие (3.30). Тогда

$$|\xi - x_n| \leq q^n |\xi - x_0|. \tag{3.32}$$

Вывод

Метод простой итерации (3.28) сходится как геометрическая прогрессия со знаменателем q.

Для получения оценки погрешности метода простой итерации будет использоваться следующий прием. При фиксированном n для произвольного натурального m будет построена некоторая оценка

$$|x_{n+m} - x_n| \leq C(q, n, m, x_0, x_1).$$

После чего в этом неравенстве будет осуществлен предельный переход при $m \to \infty$.

Оценка погрешности метода простой итерации

$$
\begin{aligned}
|x_{n+m} - x_n| = |\,&(x_{n+m} - x_{n+m-1}) + \\
&(x_{n+m-1} - x_{n+m-2}) + \ldots + \\
&(x_{n+1} - x_n)\,| \leq \\
\leq\ &|x_{n+m} - x_{n+m-1}| + \\
&|x_{n+m-1} - x_{n+m-2}| + \ldots + \\
&|x_{n+1} - x_n| \qquad\qquad \forall n, m \in \mathbb{N}.
\end{aligned}
\tag{3.33}
$$

Аналогично (3.31) нетрудно показать, что

$$
\begin{aligned}
|x_{n+k} - x_{n+k-1}| &\leq q^k |x_n - x_{n-1}| \leq \\
&\leq q^{n+k-1} |x_1 - x_0| \qquad\qquad (3.34)
\end{aligned}
$$

$\forall n, k \in \mathbb{N}.$

Оценка погрешности метода простой итерации (продолжение)

Тогда из (3.33) с учетом (3.34) следует, что

$$\begin{aligned} |x_{n+m} - x_n| \le q^{n+m-1}|x_1 - x_0| + \\ + q^{n+m-2}|x_1 - x_0| + \ldots + \\ + q^n|x_1 - x_0| = \end{aligned} \tag{3.35}$$

.

$$= q^n|x_1 - x_0|\left(q^{m-1} + q^{m-2} + \ldots + q + 1\right).$$

Предельный переход при $m \to \infty$ в обеих частях неравенства (3.35) приводит к оценке

Оценка погрешности метода

Пусть $\xi \in [a, b] \subseteq \mathrm{D}[\varphi]$ и выполнено условие (3.29). Тогда

$$|\xi - x_n| \le \frac{q^n}{1-q}|x_1 - x_0|. \tag{3.36}$$

Пример

$$f(x) = x^3 - x - 1 = 0, \quad x \in [a, b] = [1, 2]. \tag{3.37}$$

Поскольку

$$f(1) \cdot f(2) = (-1) \cdot 5 < 0$$

и

$$f'(x) = 3x^2 - 1 > 0 \ \ \forall x \in [1, 2],$$

то

$$\exists! \xi \in [1, 2] : \ \ f(\xi) = 0.$$

Уравнение (3.37) может быть преобразовано в равносильное ему уравнение вида $x = \varphi(x)$ несколькими способами.

Пример (продолжение)

Способ A)

$$x = \varphi(x) = x^3 - 1 \tag{3.38}$$

Поскольку $\varphi'(x) = 3x^2 \geq 3 > 1 \quad \forall x \in [1, 2]$, то $\forall x_0 \in [1, 2]$ метод простой итерации (3.28) для уравнения (3.38) расходится[10].

Способ B)

$$x = (x + 1)^{\frac{1}{3}} \tag{3.39}$$

Здесь $\varphi(x) = (x + 1)^{\frac{1}{3}}$ и $\varphi'(x) = \frac{1}{3} \cdot \frac{1}{(x+1)^{\frac{2}{3}}}$.

Очевидно, что $\max_{x \in [1,2]} |\varphi'(x)| = \varphi'(1) = \frac{1}{3} \cdot \frac{1}{(4)^{\frac{1}{3}}} < \frac{1}{3} = q < 1$.

Следовательно, в силу неравенства (3.32), метод простой итерации (3.28) для уравнения (3.39) сходится к ξ для любого его начального приближения $x_0 \in [1, 2]$.

---

[10] Доказать самостоятельно, используя, например, геометрическую интерпретацию метода простой итерации.

Пример (продолжение)

Здесь для $\varphi'(1)$ можно получить более точную оценку

$$\varphi'(1) = \frac{1}{3} \cdot \frac{1}{(4)^{\frac{1}{3}}} = \frac{1}{(27)^{\frac{1}{3}}} \cdot \frac{1}{(4)^{\frac{1}{3}}} \leq \frac{1}{(16)^{\frac{1}{3}}} \cdot \frac{1}{(4)^{\frac{1}{3}}} = \frac{1}{(4)^{\frac{2}{3}}} \cdot \frac{1}{(4)^{\frac{1}{3}}} = \frac{1}{4} = q.$$

Такая оценка позволяет более точно оценить количество итераций метода, которое достаточно для достижения заданной точности $\varepsilon > 0$ вычисления значения корня ξ уравнения (3.39).
Из оценки (3.36) погрешности метода следует, что

$$|\xi - x_n| \leq \frac{\left(\frac{1}{4}\right)^n}{1-\frac{1}{4}} |x_1 - x_0| = \frac{1}{3} \cdot \left(\frac{1}{4}\right)^{n-1} |x_1 - x_0|.$$

Например, если в качестве начального приближения x_0 взять середину отрезка $[1, 2]$, то есть точку $x_0 = \frac{3}{2}$, то $x_1 = (2,5)^{\frac{2}{3}}$.
Нетрудно проверить, что $1,3 \leq x_1 \leq 1,5$. Тогда $|x_1 - x_0| \leq 0,2$.
В этом случае $|\xi - x_n| \leq \frac{1}{15} \cdot \left(\frac{1}{4}\right)^{n-1} \leq \varepsilon$.
В итоге, количество итераций метода, достаточное для достижения заданной точности $\varepsilon > 0$, удовлетворяет неравенству

$$n \geq 1 + \log_4 \left(\frac{1}{15\varepsilon}\right). \tag{3.40}$$

О выборе функции g в (3.27): $x = x + g(x)f(x)$

В заключение обсуждения метода простой итерации следует остановиться на уже упомянутом выше, вообще говоря, неравносильном способе (3.27) построения уравнения вида $x = \varphi(x)$ (см. (3.26)) из уравнения $f(x) = 0$ (см. (3.1)):

$$f(x) = 0 \;\Rightarrow\; x = x + g(x)f(x),$$

где $g(x)$ — некоторая заданная функция, которая определяет функцию φ в уравнении (3.26) следующим образом

$$\varphi(x) = x + g(x)f(x). \tag{3.41}$$

Очевидно, что любой корень уравнения (3.1) является корнем и уравнения (3.27), а обратное в общем случае неверно. Поскольку нули функции g являются корнями уравнения (3.27), а корнями уравнения (3.1), вообще говоря, быть не обязаны.

О выборе функции g в (3.27): $x = x + g(x)f(x)$ (продолжение)

В случае, если требуется вычислить приближенное значение корня ξ уравнения (3.1) ($\xi \in \mathrm{D}[f] : \ f(\xi) = 0$), используя для этого метод простой итерации для уравнения (3.27), то возникает естественный вопрос:

Как в (3.41) следует задать функцию g, чтобы метод простой итерации (3.28) сходился к корню ξ уравнения (3.1) ($\xi \in \mathrm{D}[f] : \ f(\xi) = 0$) и при этом его скорость сходимости была наибольшей?

Пусть $f \in C^1([a,b])$, где $[a,b] \subseteq \mathrm{D}[f]$: $f(a)f(b) < 0$.

Тогда в предположении $g \in C^1([a,b])$ определяемая выражением (3.41) функция φ:

$$\varphi \in C^1([a,b]).$$

О выборе функции g в (3.27): $x = x + g(x)f(x)$ (продолжение)

Локальным достаточным условием сходимости метода простой итерации является $|\varphi'(\xi)| < 1$ (см. следствие теоремы 3.2). При этом, в силу (3.32), скорость сходимости метода зависит от величины q:

$$q = \max_{x \in [\xi - \varepsilon, \xi + \varepsilon]} |\varphi'(x)|,$$

где

$$\varphi'(x) = 1 + g'(x)f(x) + g(x)f'(x). \tag{3.42}$$

В силу непрерывности φ', чем меньше $|\varphi'(\xi)|$, тем меньше и значение q.

Подстановка ξ в (3.42) приводит к равенству $\varphi'(\xi) = 1 + g(\xi)f'(\xi)$. Требование того, чтобы $|\varphi'(\xi)|$ принимало минимально возможное значение ($\varphi'(\xi) = 0$) приводит к выражению

$$g(\xi) = -\frac{1}{f'(\xi)}. \tag{3.43}$$

О выборе функции g в (3.27): $x = x + g(x)f(x)$. Вывод

Продолжение правила (3.43) задания функции g на $[\xi - \varepsilon, \xi + \varepsilon]$ приводит к следующему способу определения этой функции

$$g(x) = -\frac{1}{f'(x)} \quad x \in [\xi - \varepsilon, \xi + \varepsilon]. \tag{3.44}$$

В этом случае[11] для любого начального приближения $x_0 \in [\xi - \varepsilon, \xi + \varepsilon]$ метод простой итерации для уравнения (3.27) выглядит следующим образом

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad n = 0, 1, \ldots. \tag{3.45}$$

Вывод

Метод Ньютона является методом простой итерации для уравнения $f(x) = 0$, который обладает наибольшей локальной скоростью сходимости. $\boxtimes$

[11] Правило (3.44) задания функции g обеспечивает равносильность исходного уравнения (3.1) ($f(x) = 0$) и уравнения (3.27) ($x = x + g(x)f(x)$)?

▲5 ТЕМА 4. Численные методы линейной алгебры

Система линейных алгебраических уравнений

$$A\boldsymbol{x} = \boldsymbol{b}, \tag{4.1}$$

где $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$ $(dim(A) = n \times n)$:

$$det(A) \neq 0. \tag{4.2}$$

Методы решения системы (4.1)

- Точные методы:
  1. Метод, основанный на обращении матрицы A: $\bar{\boldsymbol{x}} = A^{-1}\boldsymbol{b}$;
  2. Правило Крамера: $\bar{x}_i = \frac{\Delta_i}{\Delta}$ $i = 1, 2, \ldots, n$, где $\Delta = det(A)$, Δ_i — определитель матрицы, полученной из матрицы A заменой ее i-го столбца на вектор-столбец $\boldsymbol{b}$ $(i = 1, 2, \ldots, n)$;
  3. Метод исключения Гаусса.
- Приближенные методы — итерационные методы, основанные на процедурах построения последовательностей $\{\boldsymbol{x}^{(i)}\}_{i=0}^{\infty}$: $\lim_{i \to \infty} \boldsymbol{x}^{(i)} = \bar{\boldsymbol{x}}$.

4.1. Метод исключения Гаусса

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \ldots + a_{1n}x_n = a_{1n+1} \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \ldots + a_{2n}x_n = a_{2n+1} \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + \ldots + a_{3n}x_n = a_{3n+1} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + a_{n2}x_2 + a_{n3}x_3 + \ldots + a_{nn}x_n = a_{nn+1} \end{cases} \tag{4.3}$$

С помощью элементарных преобразований система (4.3) преобразуется в равносильную ей систему с верхнетреугольной матрицей

$$\begin{cases} x_1 + c_{12}x_2 + a_{13}x_3 + \ldots + c_{1n}x_n = c_{1n+1} \\ x_2 + a_{23}x_3 + \ldots + c_{2n}x_n = c_{2n+1} \\ x_3 + \ldots + c_{3n}x_n = c_{3n+1} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n = c_{nn+1} \end{cases} \tag{4.4}$$

Метод исключения Гаусса (продолжение)

В матричной форме это преобразование может быть записано в виде

$$A = B \cdot C, \tag{4.5}$$

где

$$B = \begin{pmatrix} b_{11} & 0 & 0 & \cdots & 0 \\ b_{21} & b_{22} & 0 & \cdots & 0 \\ b_{31} & b_{32} & b_{33} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & b_{n3} & \cdots & b_{nn} \end{pmatrix}, \; C = \begin{pmatrix} 1 & c_{12} & c_{13} & \cdots & c_{1n} \\ 0 & 1 & c_{23} & \cdots & c_{2n} \\ 0 & 0 & 1 & \cdots & c_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \tag{4.6}$$

Тогда

$$A\boldsymbol{x} = \boldsymbol{b} \sim (B \cdot C)\,\boldsymbol{x} = \boldsymbol{b} \sim \begin{cases} B\boldsymbol{y} = \boldsymbol{b} \\ C\boldsymbol{x} = \boldsymbol{y} \end{cases} \tag{4.7}$$

Метод исключения Гаусса (продолжение)

Поскольку $det(A) \neq 0$, то $b_{ii} \neq 0 \quad (i = 1, 2, \cdots, n)$. Тогда из (4.3), (4.6), (4.7) следует

$$\begin{cases} y_1 = \frac{a_{1n+1}}{b_{11}} \\ y_2 = \frac{a_{2n+1} - b_{21}y_1}{b_{22}} \\ y_3 = \frac{a_{3n+1} - b_{31}y_1 - b_{32}y_2}{b_{33}} \\ \cdots \\ y_n = \frac{a_{nn+1} - b_{n1}y_1 - b_{n2}y_2 - b_{n3}y_3 - \ldots - b_{nn-1}y_{n-1}}{b_{nn}} \end{cases} \tag{4.8}$$

$$\begin{cases} x_n = y_n \\ x_{n-1} = y_{n-1} - c_{n-1n}x_n \\ \cdots \\ x_3 = y_3 - c_{3n}x_n - c_{3n-1}x_{n-1} - \ldots - c_{34}x_4 \\ x_2 = y_2 - c_{2n}x_n - c_{2n-1}x_{n-1} - \ldots - c_{23}x_3 \\ x_1 = y_1 - c_{1n}x_n - c_{1n-1}x_{n-1} - \ldots - c_{12}x_2 \end{cases} \tag{4.9}$$

Компактная схема Гаусса

Конечношаговая процедура последовательного вычисления значений элементов b_{ij} и c_{ij} матриц B и C соответственно, позволяющая одновременно определить значения компонент вектора $\boldsymbol{y}$

Пример. Пусть $n = 4$. Тогда $A = B \cdot C \sim$

$$\begin{pmatrix} a_{11}| & \underline{a_{12}} & \underline{a_{13}} & \underline{a_{14}} \\ a_{21}| & a_{22}| & \underline{a_{23}} & \underline{a_{24}} \\ a_{31}| & a_{32}| & a_{33}| & \underline{a_{34}} \\ a_{41}| & a_{42}| & a_{43}| & a_{44}| \end{pmatrix} = \begin{pmatrix} b_{11}| & 0 & 0 & 0 \\ b_{21}| & b_{22}| & 0 & 0 \\ b_{31}| & b_{32}| & b_{33}| & 0 \\ b_{41}| & b_{42}| & b_{43}| & b_{44}| \end{pmatrix} \cdot \begin{pmatrix} 1 & \underline{c_{12}} & \underline{c_{13}} & \underline{c_{14}} \\ 0 & 1 & \underline{c_{23}} & \underline{c_{24}} \\ 0 & 0 & 1 & \underline{c_{34}} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

I. b) $a_{i1} = b_{i1} \quad \forall i \geq 1 \quad \Rightarrow \quad b_{i1} = a_{i1} \quad \forall i \geq 1$

c) $a_{1j} = b_{11}c_{1j} \quad \forall j \geq 2 \quad \Rightarrow \quad c_{1j} = \frac{a_{1j}}{b_{11}} \quad \forall j \geq 2$

II. b) $a_{i2} = b_{i1}c_{12} + b_{i2} \quad \forall i \geq 2 \quad \Rightarrow \quad b_{i2} = a_{i2} - b_{i1}c_{12} \quad \forall i \geq 2$

c) $a_{2j} = b_{21}c_{1j} + b_{22}c_{2j} \quad \forall j \geq 3 \quad \Rightarrow \quad c_{2j} = \frac{a_{2j} - b_{21}c_{1j}}{b_{22}} \quad \forall j \geq 3$

Компактная схема Гаусса (продолжение)

$$\begin{pmatrix} a_{11}| & \underline{a_{12}} & \underline{a_{13}} & \underline{a_{14}} \\ a_{21}| & a_{22}| & \underline{a_{23}} & \underline{a_{24}} \\ a_{31}| & a_{32}| & a_{33}| & \underline{a_{34}} \\ a_{41}| & a_{42}| & a_{43}| & a_{44}| \end{pmatrix} = \begin{pmatrix} b_{11}| & 0 & 0 & 0 \\ b_{21}| & b_{22}| & 0 & 0 \\ b_{31}| & b_{32}| & b_{33}| & 0 \\ b_{41}| & b_{42}| & b_{43}| & b_{44}| \end{pmatrix} \cdot \begin{pmatrix} 1 & \underline{c_{12}} & \underline{c_{13}} & \underline{c_{14}} \\ 0 & 1 & \underline{c_{23}} & \underline{c_{24}} \\ 0 & 0 & 1 & \underline{c_{34}} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\begin{aligned} &III.\ b)\ a_{i3} = b_{i1}c_{13} + b_{i2}c_{23} + b_{i3} \quad \forall i \geq 3 \quad \Rightarrow \\ &\qquad b_{i3} = a_{i3} - b_{i1}c_{13} - b_{i2}c_{23} \quad \forall i \geq 3 \end{aligned}$$

$$\begin{aligned} &c)\ a_{3j} = b_{31}c_{1j} + b_{32}c_{2j} + b_{33}c_{3j} \quad \forall j \geq 4 \quad \Rightarrow \\ &\qquad c_{3j} = \frac{a_{3j} - b_{31}c_{1j} - b_{32}c_{2j}}{b_{33}} \quad \forall j \geq 4 \end{aligned}$$

$$\begin{aligned} &IV.\ b)\ a_{i4} = b_{i1}c_{14} + b_{i2}c_{24} + b_{i3}c_{34} + b_{i4} \quad \forall i \geq 4 \quad \Rightarrow \\ &\qquad b_{i4} = a_{i4} - b_{i1}c_{14} - b_{i2}c_{24} - b_{i3}c_{34} \quad \forall i \geq 4 \end{aligned}$$

$$\begin{aligned} &c)\ a_{4j} = b_{41}c_{1j} + b_{42}c_{2j} + b_{43}c_{3j} + b_{44}c_{4j} \quad \forall j \geq 5 \quad \Rightarrow \\ &\qquad c_{4j} = \frac{a_{4j} - b_{41}c_{1j} - b_{42}c_{2j} - b_{43}c_{3j}}{b_{44}} \quad \forall j \geq 5 \end{aligned}$$

Компактная схема Гаусса (продолжение)

В результате при $j = 5$ с учетом (4.8)

$$\begin{cases} c_{15} = \frac{a_{15}}{b_{11}} = y_1 \\ c_{25} = \frac{a_{25} - b_{21}c_{15}}{b_{22}} = y_2 \\ c_{35} = \frac{a_{35} - b_{31}c_{15} - b_{32}c_{25}}{b_{33}} = y_3 \\ c_{45} = \frac{a_{45} - b_{41}c_{15} - b_{42}c_{25} - b_{43}c_{35}}{b_{44}} = y_4 \end{cases} \tag{4.10}$$

Дополнительные возможности:

- Вычисление определителя матрицы A:
  $det(A) = det(B \cdot C) = det(B)det(C) = det(B) = b_{11}b_{22}b_{33}\ldots b_{nn}$;
- Обращение матрицы A:
  $A^{-1} = \left[\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\boldsymbol{x}^{(3)}, \ldots, \boldsymbol{x}^{(n)}\right]$, где $A\boldsymbol{x}^{(i)} = \boldsymbol{b}^{(i)}$ $(i = 1, 2, \ldots, n)$,
  $\boldsymbol{b}^{(i)} = (b_1, b_2, \ldots, b_n)^\top$, $b_j = \delta_{ij}$ $(j = 1, 2, \ldots, n)$;
- Вычисление решений системы $A\boldsymbol{x} = \boldsymbol{b}$ при различных ее правых частях $\boldsymbol{b}$: в силу (4.10) от значений компонент вектора $\boldsymbol{b} = (a_{1n+1}, a_{2n+1}, \ldots, a_{nn+1})^\top$ зависят только значения y_i $(i = 1, 2, \ldots, n)$ — правых частей системы $C\boldsymbol{x} = \boldsymbol{y}$.

4.2. Определение собственных значений и собственных векторов квадратной матрицы

Пусть $A \in \mathbb{R}^{n \times n}$ $(dim(A) = n \times n)$

Определение 4.1.

Комплексное число $\lambda \in \mathbb{C}$ называется собственным значением матрицы A, если существует ненулевой вектор $\boldsymbol{x} \in \mathbb{R}^n$ с комплексными компонентами $x_i \in \mathbb{C}$ $(i = 1, 2, \ldots, n)$, удовлетворяющий уравнению

$$A\boldsymbol{x} = \lambda\boldsymbol{x} \tag{4.11}$$

Собственные значения матрицы A являются корнями ее характеристического уравнения

$$\Delta(\lambda) = det(\lambda E - A) = \lambda^n - p_1\lambda^{n-1} - \ldots - p_{n-1}\lambda - p_n = 0, \tag{4.12}$$

где E — единичная матрица размерности $n \times n$.

Структура и коэффициенты характеристического многочлена $\Delta(\lambda)$

$$\Delta(\lambda) = \lambda^n - p_1\lambda^{n-1} - \ldots - p_{n-1}\lambda - p_n =$$

$$= det\begin{pmatrix} \lambda - a_{11} & \cdots & -a_{1i} & \cdots & \underline{-a_{1j}} & \cdots & -a_{1n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \underline{-a_{i1}} & \cdots & \underline{\lambda - a_{ii}|} & \cdots & \underline{|-a_{ij}|} & \cdots & \underline{-a_{in}} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -a_{j1} & \cdots & -a_{ji} & \cdots & \underline{\lambda - a_{jj}|} & \cdots & -a_{jn} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -a_{n1} & \cdots & -a_{ni} & \cdots & \underline{-a_{nj}} & \cdots & \lambda - a_{nn} \end{pmatrix} =$$

$$= \prod_{k=1}^{n}(\lambda - a_{kk}) + P_{n-2}(\lambda) = \lambda^n - \sum_{k=1}^{n} a_{kk}\lambda^{n-1} + Q_{n-2}(\lambda),$$

Отсюда нетрудно видеть, что

$$p_1 = \sum_{k=1}^{n} a_{kk} = tr(A). \qquad (4.13)$$

Свойства собственных значений и собственных векторов квадратной матрицы

Известно, что алгебраическое уравнение степени n имеет на множестве комплексных чисел ровно n корней $\lambda_1, \lambda_2, \ldots, \lambda_n$ с учетом их кратности. При этом из известной теоремы Виета, дающей связь корней уравнения с его коэффициентами, следует

$$\begin{aligned} p_1 &= \sum_{k=1}^{n} \lambda_k = tr(A), \\ -p_n &= (-1)^n det(A) = \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_n, \end{aligned} \tag{4.14}$$

где λ_k $(k = 1, 2, \ldots, n)$ — корни уравнения (4.12) (собственные значения матрицы A)[12].

Каждая квадратная матрица A размерности $n \times n$ обладает набором из n собственных значений λ_k $(k = 1, 2, \ldots, n)$ и соответствующих им собственных векторов $\boldsymbol{x}^{(k)}$ $(k = 1, 2, \ldots, n)$. Если матрица A симметричная, т.е. $a_{ij} = a_{ji}$ $(i, j = 1, 2, \ldots, n)$, то $\lambda_k \in \mathbb{R}$ $(k = 1, 2, \ldots, n)$, а векторы $\boldsymbol{x}^{(k)}$ ортогональны между собой.

[12] Соотношения (4.14) можно использовать для контроля вычислений собственных значений матрицы.

Численные методы определения собственных значений квадратной матрицы

Численные методы решения проблемы собственных значений условно делятся на две группы:

1. Точные методы, основанные на вычислении коэффициентов характеристического уравнения (4.12) и его последующем решении (методы Леверрье-Фаддеева, Данилевского, Крылова и др.);
2. Итерационные методы, основанные на вычислении части или всех собственных значений без использования характеристического уравнения (4.12), как пределов некоторых числовых последовательностей (степенной метод, QR-алгоритм, метод вращений и др.).

Метод Леверрье

Метод Леверрье построения характеристического уравнения (4.12) матрицы (вычисления его коэффициентов) основан на формулах Ньютона[a] для сумм степеней корней алгебраического уравнения.

[a] Курош А.Г. Курс высшей алгебры.- М.: Наука, 1975.–431 с.

Пусть λ_i $(i = 1, 2, \ldots, n)$ — полная совокупность корней характеристического уравнения (4.12) матрицы A, где каждый корень повторяется столько раз, какова его кратность.

Формула Ньютона

Пусть $S_k = \sum_{i=1}^{n} \lambda_i^k, \quad k = 1, 2, \ldots, n$. Тогда

$$kp_k = S_k - p_1 S_{k-1} - \ldots - p_{k-1} S_1, \quad k = 1, 2, \ldots, n, \tag{4.15}$$

где p_i $(i = 1, 2, \ldots, n)$ — коэффициенты характеристического уравнения (4.12) матрицы A.

Метод Леверрье (продолжение)

Если числа S_k известны, то, решая рекуррентную систему (4.15), можно найти нужные коэффициенты p_k:

$$\begin{cases} p_1 = S_1 \\ p_2 = \frac{1}{2}\left(S_2 - p_1 S_1\right) \\ \dots \\ p_n = \frac{1}{n}\left(S_n - p_1 S_{n-1} - \dots - p_{n-1} S_1\right) \end{cases} \tag{4.16}$$

Из формул (4.11) (или жорданового представления матрицы A) следует, что числа λ_i^k $(i = 1, 2, \dots, n)$ являются собственными значениями матрицы A^k, а из формул (4.14) вытекает, что

$$S_k = tr(A^k), \quad k = 1, 2, \dots, n, \tag{4.17}$$

Степени $A^k = A^{k-1} \cdot A$ $(k = 2, 3, \dots, n)$ матрицы A находятся непосредственным перемножением.

Метод Леверрье (продолжение)

Численная процедура метода Леверрье

1. Вычисляются A^k $(k = 1, 2, \ldots, n)$ — степени матрицы A;
2. Вычисляются $S_k = tr(A^k)$ $(k = 1, 2, \ldots, n)$ — следы матриц A^k;
3. По формулам (4.16) определяются искомые коэффициенты p_k $(k = 1, 2, \ldots, n)$ характеристического уравнения (4.12) матрицы A.

Находятся корни характеристического уравнения (4.12) — собственные значения λ_i $(i = 1, 2, \ldots, n)$ матрицы A.
Для этого могут быть использованы известные методы приближенного решения уравнения (3.1). Представляется, что для вычисления приближенных значений корней характеристического уравнения (4.12) (корней многочлена) наиболее эффективным из них будет метод парабол (3.25).

Локализация собственных значений квадратной матрицы на комплексной плоскости

Теорема Гершгорина

Все собственные значения матрицы A лежат в объединении кругов $S^{(1)}, S^{(2)}, \ldots, S^{(n)}$, где $S^{(i)} = \{z \in \mathbb{C} \ : \ |z - a_{ii}| \leq r_i\}$, $r_i = |a_{i1}| + \ldots + |a_{ii-1}| + |a_{ii+1}| + \ldots + |a_{in}|$ — сумма модулей внедиагональных элементов i-ой строки матрицы A. При этом, если k кругов образуют замкнутую область, изолированную от других кругов, то в этой области находится ровно k собственных значений с учетом их кратности.

Если матрица A — симметричная, то теорема Гершгорина позволяет определить границы вещественных собственных значений, которые будут находиться в объединении интервалов $s^{(i)} = [a_{ii} - r_i, a_{ii} + r_i]$ $(i = 1, 2, ..., n)$.

Примеры локализации собственных значений

Пример 1. $A = \begin{pmatrix} -2 & 0,5 & 0,5 \\ -0,5 & -3,5 & 1,5 \\ 0,8 & -0,5 & 0,5 \end{pmatrix}$

Для матрицы A круги Гершгорина имеют вид:

$$\begin{array}{lll} a_{11} = -2 & r_1 = |0,5| + |0,5| = 1 & S^{(1)} = \{z \in \mathbb{C} : |z+2| \le 1\} \\ a_{22} = -3,5 & r_2 = |-0,5| + |1,5| = 2 & S^{(2)} = \{z \in \mathbb{C} : |z+3,5| \le 2\} \\ a_{33} = 0,5 & r_3 = |0,8| + |-0,5| = 1,3 & S^{(3)} = \{z \in \mathbb{C} : |z-0,5| \le 1,3\} \end{array}$$

Нетрудно построить на комплексной плоскости круги $S^{(1)}$, $S^{(2)}$ и $S^{(3)}$ и убедиться, что $S^{(1)} \cap S^{(2)} \neq \varnothing$, $\left(S^{(1)} \cup S^{(2)}\right) \cap S^{(3)} = \varnothing$.

Таким образом, в объединении кругов $S^{(1)}$ и $S^{(2)}$ находится ровно два собственных значения λ_1 и λ_2, а круг $S^{(3)}$ содержит ровно одно собственное значение λ_3.

Примеры локализации собственных значений

Пример 2. $A = \begin{pmatrix} 2,2 & 1 & 0,5 & 2 \\ 1 & 1,3 & 2 & 1 \\ 0,5 & 2 & 0,5 & 1,6 \\ 2 & 1 & 1,6 & 2 \end{pmatrix}$

Для симметрической матрицы A указанные интервалы имеют вид:

$$
\begin{array}{lll}
a_{11} = 2,2 & r_1 = |1| + |0,5| + |2| = 3,5 & s^{(1)} = [-1,3; 5,7] \\
a_{22} = 1,3 & r_2 = |1| + |2| + |1| = 4 & s^{(2)} = [-2,7; 5,3] \\
a_{33} = 0,5 & r_3 = |0,5| + |2| + |1,6| = 4,1 & s^{(3)} = [-3,6; 4,6] \\
a_{44} = 2 & r_4 = |2| + |1| + |1,6| = 4,6 & s^{(4)} = [-2,6; 6,6]
\end{array}
$$

Все собственные значения матрицы A будут находиться внутри интервала $s = s^{(1)} \cup s^{(2)} \cup s^{(3)} \cup s^{(4)} = [-3,6; 6,6]$.

Дополнительные результаты

1. Определитель матрицы A: $det(A) = (-1)^{n-1} p_n$;
2. Обратная матрица A^{-1}:
   В силу теоремы Гамильтона-Кели

$$A^n - p_1 A^{n-1} - \ldots - p_{n-1} A - p_n E = \mathbf{0},$$

где $\mathbf{0}$ — нуль-матрица размерности $n \times n$.
Умножение обеих частей этого равенства на матрицу A^{-1} приводит к выражению

$$A^{n-1} - p_1 A^{n-2} - \ldots - p_{n-1} E - p_n A^{-1} = \mathbf{0}.$$

Откуда если $p_n \neq 0$, то

$$A^{-1} = \frac{1}{p_n} \left(A^{n-1} - p_1 A^{n-2} - \ldots - p_{n-1} E \right).$$

$\boxtimes$

▲6 Модификация метода Леверрье. Метод Леверрье-Фаддеева.

Д.К.Фаддеев предложил видоизменение метода Леверрье, которое кроме упрощений при вычислении коэффициентов p_k $(k = 1, 2, \ldots, n)$ характеристического уравнения (4.12) матрицы A позволяет более просто определить обратную матрицу A^{-1} и собственные векторы матрицы A.

Вместо набора матриц $A, A^2, A^3, \ldots, A^n$ строится набор матриц $A_1, A_2, A_3, \ldots, A_n$ следующим образом:

$$\begin{cases} A_1 = A, & q_1 = tr(A_1), & B_1 = A_1 - q_1 E \\ A_2 = A \cdot B_1, & q_2 = \frac{1}{2} tr(A_2), & B_2 = A_2 - q_2 E \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ A_n = A \cdot B_{n-1}, & q_n = \frac{1}{n} tr(A_n), & B_n = A_n - q_n E \end{cases} \tag{4.18}$$

Из (4.18) следует

Метод Леверрье-Фаддеева (продолжение)

$$
\begin{aligned}
A_k &= A \cdot B_{k-1} = A\left(A_{k-1} - q_{k-1}E\right) = \\
&= A\left(A \cdot B_{k-2} - q_{k-1}E\right) = A^2 \cdot B_{k-2} - q_{k-1}A = \\
&= A^2\left(A_{k-2} - q_{k-2}E\right) - q_{k-1}A = \\
&= A^2\left(A \cdot B_{k-3} - q_{k-2}E\right) - q_{k-1}A = \\
&= A^3 \cdot B_{k-3} - q_{k-2}A^2 - q_{k-1}A = \\
&= \ldots = \\
&= A^k - q_1 A^{k-1} - q_2 A^{k-2} - \ldots - q_{k-2}A^2 - q_{k-1}A
\end{aligned}
\tag{4.19}
$$

$$
B_k = A^k - q_1 A^{k-1} - q_2 A^{k-2} - \ldots - q_{k-2}A^2 - q_{k-1}A - q_k E
$$

$$
\forall k = 1, 2, \ldots, n
$$

Откуда

$$
kq_k = tr(A_k) = S_k - q_1 S_{k-1} - q_2 S_{k-2} - \ldots - q_{k-2}S_2 - q_{k-1}S_1
\tag{4.20}
$$

$$
\forall k = 1, 2, \ldots, n
$$

Метод Леверрье-Фаддеева (продолжение)

Поскольку $q_1 = tr(A_1) = tr(A) = p_1$, то в силу формулы Ньютона (4.15) и равенств (4.20)

Коэффициенты характеристического уравнения (4.12) для матрицы A

$$p_k = q_k \quad k = 1, 2, \ldots, n \tag{4.21}$$

Обратная матрица

В силу соотношений (4.19), (4.21) и теоремы Гамильтона-Кели

$$B_n = A^n - p_1 A^{n-1} - p_2 A^{n-2} - \ldots - p_{n-2} A^2 - p_{n-1} A - p_n E = \mathbf{0} \tag{4.22}$$

Из формул (4.18) и (4.22) следует, что

$$A \cdot B_{n-1} = A_n = B_n + p_n E = p_n E$$

Откуда, если $p_n \neq 0$, то

$$A^{-1} = \frac{B_{n-1}}{p_n} \tag{4.23}$$

Нахождение собственных векторов

Пусть $\bar{\lambda}$ — собственное значение матрицы A

$$Q(\bar{\lambda}) = \bar{\lambda}^{n-1}E + \bar{\lambda}^{n-2}B_1 + \ldots + \bar{\lambda}B_{n-2} + B_{n-1}, \qquad (4.24)$$

где B_k $(k = 1, 2, \ldots, n-1)$ — матрицы, вычисленные по формулам (4.18).

$$\left(\bar{\lambda}E - A\right) Q(\bar{\lambda}) = \Delta(\bar{\lambda})E, \qquad (4.25)$$

где $\Delta(\bar{\lambda}) = det(\bar{\lambda}E - A)$.

Действительно, в силу (4.18) и (4.21)

$$\begin{aligned}
\left(\bar{\lambda}E-A\right) Q(\bar{\lambda}) &= \left(\bar{\lambda}E-A\right)\left(\bar{\lambda}^{n-1}E+\bar{\lambda}^{n-2}B_1+\ldots+\bar{\lambda}B_{n-2}+B_{n-1}\right) = \\
&= \bar{\lambda}^{n}E+\bar{\lambda}^{n-1}B_1+\ldots+\bar{\lambda}^{2}B_{n-2}+\bar{\lambda}B_{n-1}- \\
&\quad -\bar{\lambda}^{n-1}A-\bar{\lambda}^{n-2}A\cdot B_1-\ldots-\bar{\lambda}A\cdot B_{n-2}-A\cdot B_{n-1}= \\
&= \bar{\lambda}^{n}E+\bar{\lambda}^{n-1}\left(B_1-A\right)+\bar{\lambda}^{n-2}\left(B_2-A\cdot B_1\right)+\ldots+ \\
&\quad +\bar{\lambda}\left(B_{n-1}-A\cdot B_{n-2}\right)-A\cdot B_{n-1}= \\
&= \bar{\lambda}^{n}E-p_1\bar{\lambda}^{n-1}E-p_2\bar{\lambda}^{n-2}E-\ldots-p_{n-1}\bar{\lambda}E-p_nE= \\
&= \Delta(\bar{\lambda})E.
\end{aligned}$$

Нахождение собственных векторов (продолжение)

Поскольку $\bar{\lambda}$ — корень характеристического уравнения (4.12) для матрицы A, то $\Delta(\bar{\lambda}) = 0$. Таким образом,

$$\left(\bar{\lambda}E - A\right) Q(\bar{\lambda}) = \Delta(\bar{\lambda})E = \mathbf{0}$$

Откуда следует, что для любого столбца $\bar{\boldsymbol{x}}$ матрицы $Q(\bar{\lambda})$

$$\left(\bar{\lambda}E - A\right) \bar{\boldsymbol{x}} = \mathbf{0}$$

или

$$A\bar{\boldsymbol{x}} = \bar{\lambda}\bar{\boldsymbol{x}},$$

то есть $\bar{\boldsymbol{x}} \neq \mathbf{0}$ — собственный вектор матрицы A, отвечающий собственному значению $\bar{\lambda}$.

4.3. Неустранимая погрешность при решении систем линейных алгебраических уравнений

Система линейных алгебраических уравнений

$$A\boldsymbol{x} = \boldsymbol{b}, \tag{4.26}$$

где $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times n}$ $(dim(A) = n \times n)$:

$$det(A) \neq 0. \tag{4.27}$$

Пусть $\bar{\boldsymbol{x}} = A^{-1}\boldsymbol{b}$ — точное решение системы (4.26)[a].

[a] Будем считать, что отсутствуют погрешность метода и вычислительная погрешность

Неустранимая погрешность в решении системы (4.26) может быть вызвана неточностью в задании значений элементов матрицы A и вектор-столбца правых частей $\boldsymbol{b}$.

Неустранимая погрешность ... (продолжение)

“Возмущенная” система линейных алгебраических уравнений

$$(A + \Delta A)\,\boldsymbol{x} = \boldsymbol{b} + \Delta \boldsymbol{b}, \tag{4.28}$$

где $\Delta \boldsymbol{b} \in \mathbb{R}^n$, $\Delta A \in \mathbb{R}^{n\times n}$ $(dim(\Delta A) = n \times n)$

ПРИМЕР

$$\begin{cases} 0,2x_1 + x_2 = 1 + \Delta b_1 \\ x_2 = 1 + \Delta b_2 \end{cases} \tag{4.29}$$

$$\begin{cases} -x_1 + x_2 = 1 + \Delta b_1 \\ x_1 + x_2 = 1 + \Delta b_2 \end{cases} \tag{4.30}$$

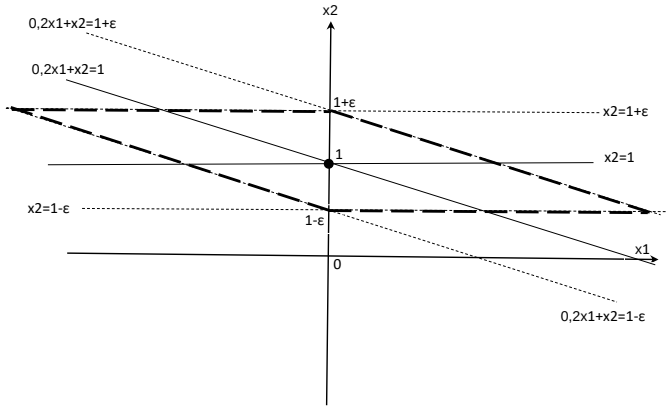
Очевидно, что при $\Delta b_1 = \Delta b_2 = 0$ вектор $\bar{\boldsymbol{x}} = (0,1)^\top$ является точным решением систем (4.29) и (4.30).

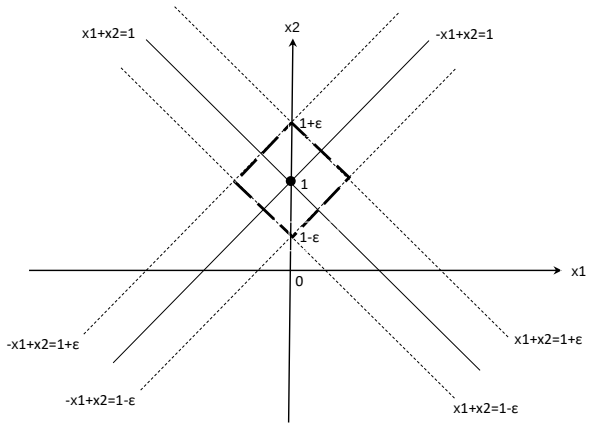
Пусть

$$|\Delta b_1| \le \varepsilon, \quad |\Delta b_2| \le \varepsilon, \tag{4.31}$$

где $\varepsilon > 0$.

Множества возможных решений “возмущенных” систем (4.29) и (4.30)





Хорошо и плохо обусловленные системы

Одно и то же возмущение правых частей в системе (4.29) привело к значительно большему возмущению решения, чем в системе (4.30).

Понятие хорошо и плохо обусловленных систем

Системы, у которых малым возмущениям параметров соответствуют малые возмущения решений, называют хорошо обусловленными. В противном случае — плохо обусловленными.

Система (4.29) является примером плохо обусловленной системы, система (4.30) — хорошо обусловленной.

Нормы векторов и матриц

Определение нормы

Пусть $\boldsymbol{X}$ — линейное (векторное) пространство. Скалярная функция $\|\cdot\| : \boldsymbol{X} \to \mathbb{R}$ называется нормой, если эта функция обладает следующими свойствами:

$$
\begin{aligned}
&1)\ \forall \boldsymbol{x} \in \boldsymbol{X} \quad \|\boldsymbol{x}\| \geq 0, \quad \|\boldsymbol{x}\| = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}, \\
&2)\ \forall \lambda \in \mathbb{R}\ \forall \boldsymbol{x} \in \boldsymbol{X} \quad \|\lambda \boldsymbol{x}\| = |\lambda| \cdot \|\boldsymbol{x}\|, \\
&3)\ \forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{X} \quad \|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|
\end{aligned}
\tag{4.32}
$$

Векторные нормы

$$
\begin{aligned}
&\boldsymbol{x} \in \mathbb{R}^n \\
&1)\ \|\boldsymbol{x}\|_1 = |x_1| + |x_2| + \ldots + |x_n|, \\
&2)\ \|\boldsymbol{x}\|_\infty = \max\nolimits_{i=1,2,\ldots,n} |x_i|, \\
&3)\ \|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}, \\
&4)\ \|\boldsymbol{x}\|_r = \left(|x_1|^r + |x_2|^r + \ldots + |x_n|^r\right)^{\frac{1}{r}}
\end{aligned}
\tag{4.33}
$$

Норма матрицы

Норма матрицы, подчиненная векторной

$A \in \mathbb{R}^{n\times n}$, $\boldsymbol{x} \in \mathbb{R}^n$

$$\|A\| = \sup_{\|\boldsymbol{x}\|=1} \|A\boldsymbol{x}\| = \sup_{\|\boldsymbol{x}\|\neq 0} \frac{\|A\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = \sup_{\|\boldsymbol{x}\|\leq 1} \|A\boldsymbol{x}\| \tag{4.34}$$

Матричные нормы

$$\begin{gathered} A \in \mathbb{R}^{n\times n} \\ 1)\ \|A\|_1 = \max_{j=1,2,\ldots,n} \sum_{i=1}^{n} |a_{ij}|, \\ 2)\ \|A\|_\infty = \max_{i=1,2,\ldots,n} \sum_{j=1}^{n} |a_{ij}|, \\ 3)\ \|A\|_2 = \sqrt{\max_{i=1,2,\ldots,n} \lambda_i\left(A^* A\right)} \end{gathered} \tag{4.35}$$

Свойства подчиненной матричной нормы

$$\begin{gathered} A, B \in \mathbb{R}^{n\times n}, \quad \boldsymbol{x} \in \mathbb{R}^n \\ 1)\ \|A\boldsymbol{x}\| \leq \|A\| \cdot \|\boldsymbol{x}\|, \\ 2)\ \|A \cdot B\| \leq \|A\| \cdot \|B\| \end{gathered} \tag{4.36}$$

Оценки неустранимой погрешности в решениях “возмущенных” систем

Случай I. “Возмущение” значений правых частей уравнений системы

$$A(\bar{\boldsymbol{x}} + \Delta\boldsymbol{x}) = \boldsymbol{b} + \Delta\boldsymbol{b}, \tag{4.37}$$

где $\Delta\boldsymbol{b}, \Delta\boldsymbol{x} \in \mathbb{R}^n$; $A\bar{\boldsymbol{x}} = \boldsymbol{b}$

$$A(\bar{\boldsymbol{x}} + \Delta\boldsymbol{x}) = A\bar{\boldsymbol{x}} + A\Delta\boldsymbol{x} = \boldsymbol{b} + A\Delta\boldsymbol{x} \Rightarrow A\Delta\boldsymbol{x} = \Delta\boldsymbol{b} \Rightarrow \Delta\boldsymbol{x} = A^{-1}\Delta\boldsymbol{b}$$

Отсюда с учетом (4.36) следует:

1) $\|\Delta\boldsymbol{x}\| = \|A^{-1}\Delta\boldsymbol{b}\| \le \|A^{-1}\|\|\Delta\boldsymbol{b}\|$;

2) Так как $A\bar{\boldsymbol{x}} = \boldsymbol{b}$, то $\|A\bar{\boldsymbol{x}}\| = \|\boldsymbol{b}\|$ и $\|\boldsymbol{b}\| \le \|A\|\|\bar{\boldsymbol{x}}\| \Rightarrow \|\bar{\boldsymbol{x}}\| \ge \frac{\|\boldsymbol{b}\|}{\|A\|}$.

$$\frac{\|\Delta\boldsymbol{x}\|}{\|\bar{\boldsymbol{x}}\|} \le \frac{\|A^{-1}\|\|\Delta\boldsymbol{b}\|}{\frac{\|\boldsymbol{b}\|}{\|A\|}} = \|A^{-1}\|\|A\|\frac{\|\Delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}$$

Оценки неустранимой погрешности в решениях “возмущенных” систем (продолжение)

Случай I.

$$\frac{\|\Delta \boldsymbol{x}\|}{\|\bar{\boldsymbol{x}}\|} \leq cond\,(A)\,\frac{\|\Delta \boldsymbol{b}\|}{\|\boldsymbol{b}\|}, \tag{4.38}$$

где $cond\,(A) = \|A^{-1}\|\|A\|$.

Определение 4.2.

Величина

$$cond\,(A) = \|A^{-1}\|\|A\|$$

называется числом обусловленности матрицы A.

Свойство

$$cond\,(A) \geq 1 \tag{4.39}$$

$\boxtimes$

▲7 Оценки неустранимой погрешности в решениях “возмущенных” систем (продолжение)

Случай II. “Возмущение” значений коэффициентов при неизвестных и правых частей уравнений системы

$$(A + \Delta A)(\bar{\boldsymbol{x}} + \Delta \boldsymbol{x}) = \boldsymbol{b} + \Delta \boldsymbol{b}, \tag{4.40}$$

где $\Delta A \in \mathbb{R}^{n\times n}$; $\Delta \boldsymbol{b}, \Delta \boldsymbol{x} \in \mathbb{R}^n$; $A\bar{\boldsymbol{x}} = \boldsymbol{b}$

$$\begin{aligned}
&\boldsymbol{b} + \Delta \boldsymbol{b} = (A + \Delta A)(\bar{\boldsymbol{x}} + \Delta \boldsymbol{x}) = \\
&= A\bar{\boldsymbol{x}} + \Delta A\bar{\boldsymbol{x}} + (A + \Delta A)\Delta \boldsymbol{x} = \\
&= \boldsymbol{b} + \Delta A\bar{\boldsymbol{x}} + (A + \Delta A)\Delta \boldsymbol{x} \Rightarrow
\end{aligned}$$

$$\begin{aligned}
&(A + \Delta A)\Delta \boldsymbol{x} = \Delta \boldsymbol{b} - \Delta A\bar{\boldsymbol{x}} \Rightarrow \\
&A\left(E + A^{-1}\Delta A\right)\Delta \boldsymbol{x} = \Delta \boldsymbol{b} - \Delta A\bar{\boldsymbol{x}} \Rightarrow
\end{aligned}$$

$$\Delta \boldsymbol{x} = \left(E + A^{-1}\Delta A\right)^{-1} A^{-1}(\Delta \boldsymbol{b} - \Delta A\bar{\boldsymbol{x}}) \tag{4.41}$$

Оценки неустранимой погрешности в решениях “возмущенных” систем (продолжение)

Выражение (4.41) с учетом (4.32) и (4.36) приводит к оценке

$$\|\Delta \boldsymbol{x}\| \leq \left\|\left(E + A^{-1}\Delta A\right)^{-1}\right\| \|A^{-1}\| \left(\|\Delta \boldsymbol{b}\| + \|\Delta A\| \|\bar{\boldsymbol{x}}\|\right) \tag{4.42}$$

Утверждение

Пусть $B \in \mathbb{R}^{n\times n}$: $\|B\| < 1$. Тогда

$$\|(E - B)^{-1}\| \leq \frac{1}{1 - \|B\|} \tag{4.43}$$

В предположении, что $\Delta A \in \mathbb{R}^{n\times n}$: $\|\Delta A\| < \frac{1}{\|A^{-1}\|}$ ($\|A^{-1}\Delta A\| \leq \|A^{-1}\| \|\Delta A\| < 1$)[13], из (4.42) следует[14]

$$\|\Delta \boldsymbol{x}\| \leq \frac{1}{1 - \|A^{-1}\| \|\Delta A\|} \|A^{-1}\| \left(\|\Delta \boldsymbol{b}\| + \|\Delta A\| \|\bar{\boldsymbol{x}}\|\right) \tag{4.44}$$

---

[13] Когда ΔA мало, то указанное условие выполняется.

[14] $B = -A^{-1}\Delta A$

Оценки неустранимой погрешности в решениях “возмущенных” систем (продолжение)

Деление обеих частей неравенства (4.44) на $\|\bar{\boldsymbol{x}}\|$ приводит к оценке

$$\frac{\|\Delta \boldsymbol{x}\|}{\|\bar{\boldsymbol{x}}\|} \leq \frac{1}{1 - \|A^{-1}\|\|\Delta A\|} \|A^{-1}\| \left(\frac{\|\Delta \boldsymbol{b}\|}{\|\bar{\boldsymbol{x}}\|} + \|\Delta A\| \right) \tag{4.45}$$

Из (4.45) с учетом того, что $\|\bar{\boldsymbol{x}}\| \geq \frac{\|\boldsymbol{b}\|}{\|A\|}$, следует

$$\frac{\|\Delta \boldsymbol{x}\|}{\|\bar{\boldsymbol{x}}\|} \leq \frac{\|A^{-1}\|\|A\|}{1 - \|A^{-1}\|\|A\| \frac{\|\Delta A\|}{\|A\|}} \left(\frac{\|\Delta \boldsymbol{b}\|}{\|\boldsymbol{b}\|} + \frac{\|\Delta A\|}{\|A\|} \right) \tag{4.46}$$

Случай II.

$$\frac{\|\Delta \boldsymbol{x}\|}{\|\bar{\boldsymbol{x}}\|} \leq \frac{cond\,(A)}{1 - cond\,(A) \frac{\|\Delta A\|}{\|A\|}} \left(\frac{\|\Delta \boldsymbol{b}\|}{\|\boldsymbol{b}\|} + \frac{\|\Delta A\|}{\|A\|} \right) \tag{4.47}$$

ПРИМЕР

Исходная система

$$\begin{cases} 1,01x_1 + 0,99x_2 = 2,00 \\ 0,99x_1 + 0,98x_2 = 1,97 \end{cases} \tag{4.48}$$

“Возмущенная” система

$$\begin{cases} 1,01x_1 + 0,99x_2 = 2,04 \\ 0,99x_1 + 0,98x_2 = 1,99 \end{cases} \tag{4.49}$$

Очевидно, что $A^T = A$ и $A > 0$. Следовательно, $\lambda_i(A) \in \mathbb{R}$ и $\lambda_i(A) > 0$ $(i = 1, 2)$.

Поэтому
1) $\|A\|_2 = \sqrt{\max_{i=1,2} \lambda_i(A^T A)} = \sqrt{\max_{i=1,2} \lambda_i(A^2)} = \max_{i=1,2} \lambda_i(A)$;
2)[15] $\|A^{-1}\|_2 = \max_{i=1,2} \frac{1}{\lambda_i(A)} = \frac{1}{\min_{i=1,2} \lambda_i(A)}$.

---

[15] Если $A^T = A$, то $E = \left(A^{-1}A\right)^T = A^T \left(A^{-1}\right)^T = A \left(A^{-1}\right)^T$. Тогда $\left(A^{-1}\right)^T = A^{-1}$.

ПРИМЕР

$\lambda_1(A) \approx 1,98005,\ \lambda_2(A) \approx 0,00005$

$$cond_2(A) = \|A^{-1}\|\|A\| = \frac{\max_{i=1,2}\lambda_i(A)}{\min_{i=1,2}\lambda_i(A)} \approx \frac{1,98005}{0,00005} \approx 4 \cdot 10^4$$

Вывод

Исходная система (4.48) является плохо обусловленной[a].

[a] Приблизительными значениями числа обусловленности, разделяющими хорошо и плохо обусловленные системы, считается 10-20.

Исходная система

$\boldsymbol{b} = (2,00; 1,97)^\top$; Решение: $\bar{\boldsymbol{x}} = (1,00; 1,00)^\top$

"Возмущенная" система

$\tilde{\boldsymbol{b}} = (2,04; 1,99)^\top = \boldsymbol{b} + \Delta\boldsymbol{b},\ \Delta\boldsymbol{b} = (0,04; 0,02)^\top$;
Решение: $\tilde{\boldsymbol{x}} = (3,00; -1,00)^\top = \bar{\boldsymbol{x}} + \Delta\boldsymbol{x},\ \Delta\boldsymbol{x} = (2,00; -2,00)^\top$

Сравнительно малые вариации значений правых частей исходной системы (4.48) приводят к значительным "возмущениям" ее решения. При этом, как нетрудно видеть, $\lambda_1(A) \gg \lambda_2(A)$.

Геометрическая интерпретация. Частный случай.

$$A\bar{\boldsymbol{x}} = \boldsymbol{b}, \tag{4.50}$$

$$A\left(\bar{\boldsymbol{x}} + \Delta\boldsymbol{x}\right) = \boldsymbol{b} + \Delta\boldsymbol{b}, \tag{4.51}$$

где $A \in \mathbb{R}^{n\times n}$: $A > 0$; $\bar{\boldsymbol{x}}, \boldsymbol{b}, \Delta\boldsymbol{x}, \Delta\boldsymbol{b} \in \mathbb{R}^n$: $A\bar{\boldsymbol{x}} = \boldsymbol{b}$.

В силу (4.50), (4.51)

$$A\Delta\boldsymbol{x} = \Delta\boldsymbol{b}. \tag{4.52}$$

Частный случай: $n = 2$ и $\|\Delta\boldsymbol{b}\| = \varepsilon$, где $\varepsilon > 0$

$\|\Delta\boldsymbol{b}\| = \sqrt{(\Delta\boldsymbol{b}, \Delta\boldsymbol{b})}$.
Из (4.52) следует $(\Delta\boldsymbol{b}, \Delta\boldsymbol{b}) = (A\Delta\boldsymbol{x}, A\Delta\boldsymbol{x}) = (\Delta\boldsymbol{x}, A^*A\Delta\boldsymbol{x})$.

$$(\Delta\boldsymbol{x}, A^*A\Delta\boldsymbol{x}) = \varepsilon^2 \tag{4.53}$$

Поскольку $A^*A > 0$, то уравнение (4.53) определяет в $\mathbb{R}^2$ эллипс.

Геометрическая интерпретация (Продолжение)

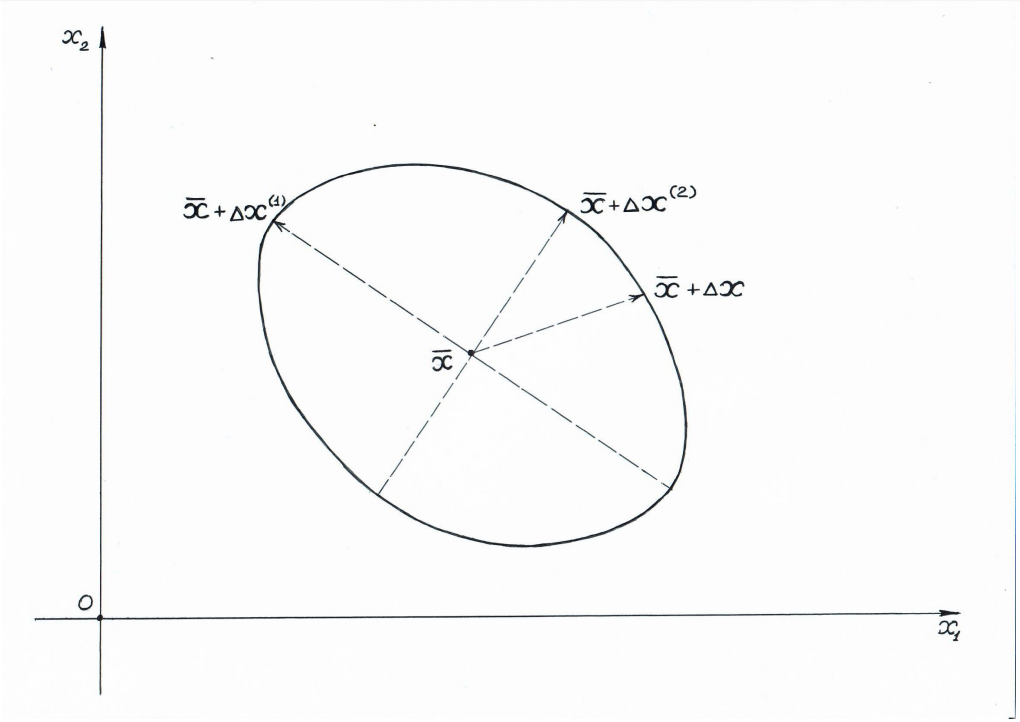
$$\boldsymbol{u}^{(1)},\ \boldsymbol{u}^{(2)} \in \mathbb{R}^2 :$$
$$A\boldsymbol{u}^{(1)} = \lambda_1 \boldsymbol{u}^{(1)}, \quad A\boldsymbol{u}^{(2)} = \lambda_2 \boldsymbol{u}^{(2)},$$
$$\lambda_2 > \lambda_1 > 0, \quad \boldsymbol{u}^{(1)} \perp \boldsymbol{u}^{(2)}$$

Пусть $\Delta\boldsymbol{b}^{(i)} \uparrow\uparrow \boldsymbol{u}^{(i)}$ $(i = 1, 2)$

Тогда $\Delta\boldsymbol{x}^{(i)} = A^{-1}\Delta\boldsymbol{b}^{(i)} = \frac{1}{\lambda_i}\Delta\boldsymbol{b}^{(i)}$: $\quad \|\Delta\boldsymbol{x}^{(i)}\| = \frac{\varepsilon}{\lambda_i}$ $(i = 1, 2)$.

Тогда в ортонормированной системе координат, задаваемой собственными векторами $\boldsymbol{u}^{(1)}$, $\boldsymbol{u}^{(2)}$ матрицы A, длины полуосей эллипса, определяемого уравнением (4.53) равны $\frac{\varepsilon}{\lambda_1}$, $\frac{\varepsilon}{\lambda_2}$ и его эксцентриситет $e = \frac{\lambda_2}{\lambda_1}$.

Геометрическая интерпретация (Продолжение)



Точки этого эллипса геометрически определяют решения $\bar{\boldsymbol{x}} + \Delta \boldsymbol{x}$ "возмущенной" системы (4.51), соответствующие различным "возмущениям" $\Delta \boldsymbol{b}$ ($\|\Delta \boldsymbol{b}\| = \varepsilon$) правых частей $\boldsymbol{b}$ системы (4.50). Таким образом, одно и то же "возмущение" по норме правых частей системы (4.50) приводит к различным "возмущениям" ее решений. Разброс норм этих "возмущений" определяется величиной $\lambda_2 - \lambda_1$.

4.4. Итерационные методы решения систем линейных алгебраических уравнений

Пусть для системы $A\boldsymbol{x} = \boldsymbol{b}$ построена равносильная ей система вида

$$\boldsymbol{x} = B\boldsymbol{x} + \boldsymbol{c}, \tag{4.54}$$

где $B \in \mathbb{R}^{n\times n}$; $\boldsymbol{x}, \boldsymbol{c} \in \mathbb{R}^n$

Численная процедура метода простой итерации

$$\boldsymbol{x}^{(k+1)} = B\boldsymbol{x}^{(k)} + \boldsymbol{c}, \tag{4.55}$$

где $k = 0, 1, 2, \ldots$; $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$

В силу (4.55)

$$\begin{aligned}\boldsymbol{x}^{(k+1)} &= B\boldsymbol{x}^{(k)} + \boldsymbol{c} = B\left(B\boldsymbol{x}^{(k-1)} + \boldsymbol{c}\right) + \boldsymbol{c} = B^2\boldsymbol{x}^{(k-1)} + (B+E)\,\boldsymbol{c} = \\ &= B^2\left(B\boldsymbol{x}^{(k-2)} + \boldsymbol{c}\right) + (B+E)\,\boldsymbol{c} = B^3\boldsymbol{x}^{(k-2)} + \left(B^2+B+E\right)\boldsymbol{c} = \\ &\ldots\ldots\ldots \\ &= B^{k+1}\boldsymbol{x}^{(0)} + \left(B^k + B^{k-1} + \ldots + B^2 + B + E\right)\boldsymbol{c}\end{aligned}$$

Метод простой итерации

$$\boldsymbol{x}^{(k+1)} = B^{k+1}\boldsymbol{x}^{(0)} + \left(\sum_{i=0}^{k} B^i\right)\boldsymbol{c}, \tag{4.56}$$

где $k = 0, 1, 2, \ldots$; $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$

Определение 4.3. Предел последовательности матриц

Пусть $\{A^{(k)}\}_{k=1}^{\infty}$, $\forall k \in \mathbb{N}\ A^{(k)} \in \mathbb{R}^{n\times m}$.

$\lim_{k\to\infty} A^{(k)} = A$, где $A \in \mathbb{R}^{n\times m}$, если

$$\lim_{k\to\infty} a_{ij}^{(k)} = a_{ij} \quad (\forall i = 1, 2, \ldots, n; \forall j = 1, 2, \ldots, m) \tag{4.57}$$

Условие (4.57) равносильно

$$\lim_{k\to\infty} \|A^{(k)} - A\| = 0 \tag{4.58}$$

или

$$\lim_{k\to\infty} A^{(k)}\boldsymbol{x} = A\boldsymbol{x} \quad \forall \boldsymbol{x} \in \mathbb{R}^m \tag{4.59}$$

Метод простой итерации (продолжение)

Свойства сходящихся матричных последовательностей

Пусть $\{A^{(k)}\}_{k=1}^{\infty}$, $\{B^{(k)}\}_{k=1}^{\infty}$, $\forall k \in \mathbb{N}\ A^{(k)}, B^{(k)} \in \mathbb{R}^{n\times n}$:
$\exists \lim_{k\to\infty} A^{(k)} = A$ и $\exists \lim_{k\to\infty} B^{(k)} = B$, где $A, B \in \mathbb{R}^{n\times n}$.
Тогда

$$\begin{aligned} &\exists \lim_{k\to\infty} \left(A^{(k)} + B^{(k)}\right) = A + B, \\ &\exists \lim_{k\to\infty} \left(A^{(k)} \cdot B^{(k)}\right) = A \cdot B \end{aligned} \tag{4.60}$$

Определение 4.4. Сумма матричного ряда

Пусть $\sum_{k=1}^{\infty} A^{(k)}$, $\forall k \in \mathbb{N}\ A^{(k)} \in \mathbb{R}^{n\times m}$.
Ряд $\sum_{k=1}^{\infty} A^{(k)}$ называется сходящимся, если

$\exists \lim_{l\to\infty} S_A^{(l)} = A$, где $S_A^{(l)} = \sum_{k=1}^{l} A^{(k)}$, $A \in \mathbb{R}^{n\times m}$.

Необходимое условие сходимости матричного ряда

Пусть ряд $\sum_{k=1}^{\infty} A^{(k)}$ сходится. Тогда

$$\lim_{k\to\infty} A^{(k)} = \mathbf{0} \tag{4.61}$$

Метод простой итерации (продолжение)

Теорема 4.1.

Пусть $B \in \mathbb{R}^{n\times n}$.
$\lim_{k\to\infty} B^k = \mathbf{0} \quad \Leftrightarrow \quad |\lambda_i(B)| < 1 \ \forall i = 1, 2, \ldots, n$, где
$\lambda_i(B)$ $(i = 1, 2, \ldots, n)$ — собственные значения матрицы B.

Доказательство утверждения теоремы основано на использовании жордановой формы матрицы. Для любой квадратной матрицы B существует невырожденная матрица (преобразование) T такое, что матрица $G = T^{-1} \cdot B \cdot T$ имеет жорданову форму[16].

$$
\begin{aligned}
G^2 &= \left(T^{-1} \cdot B \cdot T\right) \cdot \left(T^{-1} \cdot B \cdot T\right) = T^{-1} \cdot B^2 \cdot T, \\
G^3 &= \left(T^{-1} \cdot B^2 \cdot T\right) \cdot \left(T^{-1} \cdot B \cdot T\right) = T^{-1} \cdot B^3 \cdot T, \\
&\cdots\cdots\cdots \\
G^k &= \left(T^{-1} \cdot B^{k-1} \cdot T\right) \cdot \left(T^{-1} \cdot B \cdot T\right) = T^{-1} \cdot B^k \cdot T.
\end{aligned}
$$

Поскольку

1. $\|G^k\| \le \|T^{-1}\|\|B^k\|\|T\|$,
2. $B^k = T \cdot G^k \cdot T^{-1}$ и $\|B^k\| \le \|T\|\|G^k\|\|T^{-1}\|$,

то $\lim_{k\to\infty} B^k = \mathbf{0} \quad \Leftrightarrow \quad \lim_{k\to\infty} G^k = \mathbf{0}$.

[16] см. теорему о приведении матрицы к жордановой форме

Метод простой итерации (продолжение)

Предлагается ограничиться рассмотрением основной идеи доказательства утверждения теоремы.
В простейшем случае, когда $\lambda_i(B) \in \mathbb{R}$ $(i = 1, 2, \ldots, n)$ и $\lambda_i(B) \neq \lambda_j(B)$ $\forall i \neq j$, утверждение теоремы становится очевидным. В этом случае

$$G = \begin{pmatrix} \lambda_1(B) & 0 & \cdots & 0 \\ 0 & \lambda_2(B) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n(B) \end{pmatrix},$$

$$G^k = \begin{pmatrix} \lambda_1^k(B) & 0 & \cdots & 0 \\ 0 & \lambda_2^k(B) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n^k(B) \end{pmatrix}.$$

Доказательство утверждения теоремы для случая, когда матрица B имеет кратные и/или комплексные собственные значения, является технически существенно более громоздким[17]. $\boxtimes$

[17] Предлагается ознакомиться самостоятельно.

▲8 Метод простой итерации (продолжение)

Теорема 4.2.

Для сходимости ряда $\sum_{k=0}^{\infty} B^k$ необходимо и достаточно, чтобы $|\lambda_i(B)| < 1 \ \forall i = 1, 2, \ldots, n$, где $\lambda_i(B)$ $(i = 1, 2, \ldots, n)$ — собственные значения матрицы B.

Доказательство:

Необходимость следует из необходимого условия (4.61) сходимости ряда $\sum_{i=0}^{\infty} B^k$ и теоремы 4.1.

Достаточность

Пусть $|\lambda_i(B)| < 1 \ \forall i = 1, 2, \ldots, n$. Тогда $\lambda_i(E - B) > 0$ $\forall i = 1, 2, \ldots, n$. Следовательно, матрица $E - B$ невырождена.

Метод простой итерации (продолжение)

Нетрудно убедиться, что

$$\left(E + B + B^2 + \ldots + B^k\right) \cdot (E - B) = E - B^{k+1}. \qquad (4.62)$$

Умножение обеих частей равенства (4.62) на матрицу $(E - B)^{-1}$ справа приводит к выражению

$$E + B + B^2 + \ldots + B^k = \left(E - B^{k+1}\right) \cdot (E - B)^{-1}. \qquad (4.63)$$

Переход к пределу при $k \to \infty$ в (4.63) приводит к формуле

$$\sum_{i=0}^{\infty} B^k = (E - B)^{-1}. \qquad (4.64)$$

Поскольку, в силу теоремы 4.1, $\lim_{k\to\infty} B^k = \mathbf{0}$.
□ Теорема доказана.

Метод простой итерации (продолжение)

Теорема 4.3.

Для сходимости итерационного процесса (4.56) (метода простой итерации) при любых начальном приближении $\boldsymbol{x}^{(0)}$ и векторе $\boldsymbol{c}$ необходимо и достаточно, чтобы $|\lambda_i(B)| < 1 \ \forall i = 1, 2, \ldots, n$, где $\lambda_i(B)$ $(i = 1, 2, \ldots, n)$ — собственные значения матрицы B.

Доказательство:

Необходимость

Если итерационный процесс (4.56) сходится при любых начальном приближении $\boldsymbol{x}^{(0)}$ и векторе $\boldsymbol{c}$, то он сходится и при $\boldsymbol{x}^{(0)} = \boldsymbol{0}$. Для нулевого начального приближения итерационная процедура (4.56) приобретает вид

$$\boldsymbol{x}^{(k+1)} = \left(E + B + B^2 + \ldots + B^k\right)\boldsymbol{c}, \quad k = 0, 1, 2, \ldots \ . \tag{4.65}$$

Поскольку для любого вектора $\boldsymbol{c}$ существует предел при $k \to \infty$ левой части равенства (4.65), то существует $\lim_{k\to\infty}\left(E + B + B^2 + \ldots + B^k\right)\boldsymbol{c}$.

Метод простой итерации (продолжение)

В силу определения сходимости матричной последовательности это означает, что $\exists \lim_{m\to\infty} S_B^{(k)}$, где $S_B^{(k)} = \sum_{m=o}^{k} B^m$, то есть ряд $\sum_{k=0}^{\infty} B^k$ сходится. Тогда условие $|\lambda_i(B)| < 1 \ \forall i = 1, 2, \ldots, n$ следует из теоремы 4.2.

Достаточность следует из теорем 4.1 и 4.2.

□ Теорема доказана.

Теорема 4.4.

Для любой матрицы $B \in \mathbb{R}^{n\times n}$ справедливы неравенства

$$|\lambda_i(B)| \le \|B\| \quad \forall i = 1, 2, \ldots, n, \tag{4.66}$$

где $\lambda_i(B)$ $(i = 1, 2, \ldots, n)$ — собственные значения матрицы B.

Доказательство: Пусть λ — собственное значение матрицы B и $\boldsymbol{x} \neq \mathbf{0}$ — соответствующий ему собственный вектор этой матрицы. Тогда $B\boldsymbol{x} = \lambda\boldsymbol{x}$. В силу свойств (4.33), (4.36) нормы

Метод простой итерации (продолжение)

$$|\lambda| \cdot \|\boldsymbol{x}\| = \|\lambda \boldsymbol{x}\| = \|B\boldsymbol{x}\| \leq \|B\| \cdot \|\boldsymbol{x}\|. \tag{4.67}$$

Поскольку $\|\boldsymbol{x}\| > 0$, то деление обеих частей выражения (4.67) на $\|\boldsymbol{x}\|$ приводит к неравенству $|\lambda| \leq \|B\|$.
□ Теорема доказана.

Следствие

Пусть $B \in \mathbb{R}^{n \times n}$: $\|B\| < 1$. Тогда

$$\|(E - B)^{-1}\| \leq \frac{1}{1 - \|B\|} \tag{4.68}$$

Доказательство: Поскольку $|\lambda_i(B)| \leq \|B\| < 1 \ \forall i = 1, 2, \ldots, n$, то, в силу теоремы 4.2, ряд $\sum_{k=0}^{\infty} B^k$ сходится и его сумма определяется формулой (4.64). Таким образом,

$$\|(E - B)^{-1}\| = \|\sum_{k=0}^{\infty} B^k\| \leq \sum_{k=0}^{\infty} \|B^k\| \leq \sum_{k=0}^{\infty} \|B\|^k = \frac{1}{1 - \|B\|}.$$

□ Следствие доказано.

Оценка скорости сходимости метода простой итерации

Пусть $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ является решением системы (4.54), то есть

$$\bar{\boldsymbol{x}} = B\bar{\boldsymbol{x}} + \boldsymbol{c} \tag{4.69}$$

Тогда, в силу формулы (4.55), задающей процедуру метода простой итерации, и (4.69), а также свойств (4.36) нормы

$$\begin{aligned}
\|\bar{\boldsymbol{x}} - \boldsymbol{x}^{(k+1)}\| &= \|B\bar{\boldsymbol{x}} + \boldsymbol{c} - B\boldsymbol{x}^{(k)} - \boldsymbol{c}\| = \|B(\bar{\boldsymbol{x}} - \boldsymbol{x}^{(k)})\| \leq \\
&\leq \|B\|\|\bar{\boldsymbol{x}} - \boldsymbol{x}^{(k)}\| = \|B\|\|B\bar{\boldsymbol{x}} + \boldsymbol{c} - B\boldsymbol{x}^{(k-1)} - \boldsymbol{c}\| = \\
&= \|B\|\|B(\bar{\boldsymbol{x}} - \boldsymbol{x}^{(k-1)})\| \leq \|B\|^2\|\bar{\boldsymbol{x}} - \boldsymbol{x}^{(k-1)}\| = \\
&\cdots\cdots\cdots\cdots\cdots \\
&\leq \|B\|^{k+1}\|\bar{\boldsymbol{x}} - \boldsymbol{x}^{(0)}\|
\end{aligned}$$

Теорема 4.5. Достаточное условие сходимости

Пусть $\|B\| < 1$. Тогда итерационный процесс (4.55) сходится при любом начальном приближении $\boldsymbol{x}^{(0)}$ и справедлива оценка

$$\|\bar{\boldsymbol{x}} - \boldsymbol{x}^{(k)}\| \leq \|B\|^k\|\bar{\boldsymbol{x}} - \boldsymbol{x}^{(0)}\| \quad \forall k \in \mathbb{N} \tag{4.70}$$

Достаточные условия сходимости метода простой итерации

$$
\begin{aligned}
&1)\ \|B\|_1 = \max_{j=1,2,\ldots,n} \sum_{i=1}^{n} |b_{ij}| < 1, \\
&2)\ \|B\|_\infty = \max_{i=1,2,\ldots,n} \sum_{j=1}^{n} |b_{ij}| < 1, \\
&3)\ \|B\|_2 = \sqrt{\max_{i=1,2,\ldots,n} \lambda_i \left(B^* B\right)} < 1
\end{aligned}
\tag{4.71}
$$

С учетом (4.14)

$$
\begin{aligned}
\left(\|B\|_2\right)^2 &= \max_{i=1,2,\ldots,n} \lambda_i \left(B^* B\right) \leq \\
&\leq \lambda_1 \left(B^* B\right) + \lambda_2 \left(B^* B\right) + \ldots + \lambda_n \left(B^* B\right) = \\
&= tr(B^* B) = \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij}^2
\end{aligned}
$$

$$
4)\ \sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij}^2 < 1 \tag{4.72}
$$

Погрешность метода простой итерации

Пусть $\|B\| < 1$ и $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ — решение системы (4.54), то есть $\bar{\boldsymbol{x}} = B\bar{\boldsymbol{x}} + \boldsymbol{c}$ или

$$\bar{\boldsymbol{x}} = (E - B)^{-1} \boldsymbol{c}. \tag{4.73}$$

Аналогично (4.63) нетрудно показать, что

$$\begin{aligned} \boldsymbol{x}^{(k)} &= B^k \boldsymbol{x}^{(0)} + \left(B^{k-1} + B^{k-2} + \ldots + B^2 + B + E\right) \boldsymbol{c} = \\ &= B^k \boldsymbol{x}^{(0)} + (E - B)^{-1} \left(E - B^k\right) \boldsymbol{c}. \end{aligned} \tag{4.74}$$

Тогда, в силу (4.73) и (4.74),

$$\begin{aligned} \boldsymbol{x}^{(k)} - \bar{\boldsymbol{x}} &= B^k \boldsymbol{x}^{(0)} + (E - B)^{-1} \left(E - B^k\right) \boldsymbol{c} - (E - B)^{-1} \boldsymbol{c} = \\ &= B^k \boldsymbol{x}^{(0)} - (E - B)^{-1} B^k \boldsymbol{c}. \end{aligned}$$

Откуда с учетом (4.68)

Оценка погрешности метода простой итерации

Пусть $\|B\| < 1$. Тогда

$$\|\boldsymbol{x}^{(k)} - \bar{\boldsymbol{x}}\| \le \|B\|^k \|\boldsymbol{x}^{(0)}\| + \frac{\|B\|^k \|\boldsymbol{c}\|}{1 - \|B\|} \quad \forall k \in \mathbb{N}. \tag{4.75}$$

Частные реализации метода простой итерации для системы $A\boldsymbol{x} = \boldsymbol{b}$

$$A\boldsymbol{x} = \boldsymbol{b}, \tag{4.76}$$

где $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$: $det(A) \neq 0$.
Для системы (4.76) справедливо следующее преобразование

$$\begin{aligned} A\boldsymbol{x} = \boldsymbol{b} \;&\sim\; A\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0} \;\sim \\ &\sim\; H\left(A\boldsymbol{x} - \boldsymbol{b}\right) = \boldsymbol{0} \;\sim \\ &\sim\; \boldsymbol{x} = \boldsymbol{x} - H\left(A\boldsymbol{x} - \boldsymbol{b}\right) \;\sim \\ &\sim\; \boldsymbol{x} = \left(E - HA\right)\boldsymbol{x} + H\boldsymbol{b}, \end{aligned} \tag{4.77}$$

где $H \in \mathbb{R}^{n\times n}$ — произвольная невырожденная матрица.

Таким образом, система (4.76) с помощью преобразования (4.77) может быть преобразована в равносильную ей систему вида

$$\boldsymbol{x} = B\boldsymbol{x} + \boldsymbol{c}, \tag{4.78}$$

для приближенного решения которой применим метод простой итерации (4.55). Здесь $B = E - HA$, $\boldsymbol{c} = H\boldsymbol{b}$.

Частные реализации метода простой итерации для системы $Ax = b$

Согласно оценке (4.70) (см. теорему 4.5), для системы (4.78) метод простой итерации (4.55) сходится также, как геометрическая прогрессия со знаменателем $\|B\|$.

Следовательно, чем "ближе" в (4.77) матрица H к A^{-1}, тем быстрее сходится метод простой итерации для системы (4.78) с матрицей $B = E - HA$.

В вычислительной практике для приближенного решения системы (4.76) методом простой итерации используют достаточно много частных способов приведения системы (4.76) к требуемому виду (4.78). Таким образом возникают частные реализации метода простой итерации (4.55). Далее будут рассмотрены:

- Метод Якоби,
- Метод Гаусса-Зейделя.

Метод Якоби

Пусть $A \in \mathbb{R}^{n\times n}$:

$$a_{ii} \neq 0 \quad \forall i = 1, 2, \ldots, n. \tag{4.79}$$

В этом случае из каждого уравнения системы системы (4.76) можно выразить переменную с коэффициентом, задаваемым диагональным элементом матрицы A, через другие переменные. В результате строится равносильная системе (4.76) система (4.78), численная процедура метода простой итерации для которой в покоординатной форме записи имеет вид:

Численная процедура метода Якоби

$$\begin{aligned} & x_i^{(k+1)} = -\frac{1}{a_{ii}} \left(\sum_{j=1}^{i-1} a_{ij} x_j^{(k)} + \sum_{j=i+1}^{n} a_{ij} x_j^{(k)} - b_i \right) \\ & i = 1, 2, \ldots, n, \end{aligned} \tag{4.80}$$

где $k = 0, 1, 2, \ldots$, $\left(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)} \right)^\top \in \mathbb{R}^n$.

Для исследования свойств метода Якоби (4.80) удобно его записать в матричной форме. Для этого предлагается матрицу A представить в виде суммы трех слагаемых:

Метод Якоби (продолжение)

$$A = L + D + R, \tag{4.81}$$

где L — нижнетреугольная матрица с нулевой главной диагональю, D — диагональная матрица, R — верхнетреугольная матрица с нулевой главной диагональю.

Тогда метод Якоби (4.80) в матричной форме может быть представлен в следующем виде:

Численная процедура метода Якоби

$$\boldsymbol{x}^{(k+1)} = -D^{-1}(L + R)\,\boldsymbol{x}^{(k)} + D^{-1}\boldsymbol{b}, \tag{4.82}$$

где $k = 0, 1, 2, \ldots$, $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$.

Согласно теореме 4.3, критерий сходимости метода Якоби (4.82) может быть сформулирован в терминах корней уравнения

$$det(\lambda E - B) = 0, \tag{4.83}$$

в котором $B = -D^{-1}(L + R)$.

Критерий сходимости метода Якоби

Левая часть равенства (4.83) может быть преобразована следующим образом

$$\begin{aligned} det(\lambda E - B) &= det(\lambda E + D^{-1}(L+R)) = \\ &= det(D^{-1}(\lambda D + L + R)) = \\ &= det(D^{-1}) det(\lambda D + L + R) \end{aligned} \tag{4.84}$$

Поскольку в силу сделанных предположений (4.79)[18] $det(D^{-1}) \neq 0$, то с учетом (4.84) уравнение (4.83) равносильно следующему

$$det(\lambda D + L + R) = 0. \tag{4.85}$$

Теорема 4.6.

Для сходимости метода Якоби (4.82) необходимо и достаточно, чтобы абсолютные значения всех корней уравнения (4.85) были меньше единицы.

[18] $a_{ii} \neq 0 \ \forall i = 1, 2, \ldots, n$

Достаточное условие сходимости метода Якоби

Достаточное условие сходимости метода Якоби сформулировано в теореме 4.5. Для рассматриваемого метода предлагается его конкретизировать в терминах свойств матрицы A, используя "построчную" норму матрицы $B = -D^{-1}(L+R)$:

$$B = \{b_{ij}\}_{i,j=1}^{n}, \quad b_{ij} = \begin{cases} -\frac{a_{ij}}{a_{ii}}, & i \neq j \\ 0, & i = j \end{cases}$$

Для того, чтобы $\|B\|_\infty < 1$ необходимо и достаточно, чтобы

$$\sum_{j=1}^{n} |b_{ij}| = \sum_{j=1, j\neq i}^{n} \frac{|a_{ij}|}{|a_{ii}|} < 1 \quad \forall i = 1, 2, \ldots, n. \tag{4.86}$$

Условие (4.86) равносильно следующему

$$\sum_{j=1, j\neq i}^{n} |a_{ij}| < |a_{ii}| \quad \forall i = 1, 2, \ldots, n. \tag{4.87}$$

Достаточное условие сходимости метода Якоби (продолжение)

Определение 4.5.

Матрица $A \in \mathbb{R}^{n \times n}$ обладает свойством строгого диагонального преобладания[a], если она удовлетворяет условию (4.87).

[a] Если в условиях (4.87) знак неравенства "$\leq$", то говорят о нестрогом диагональном преобладании.

Теорема 4.7. Достаточное условие сходимости

Если матрица A системы (4.76) обладает свойством строгого диагонального преобладания, то метод Якоби (4.82) сходится для любого начального приближения $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$.

Метода Якоби. Замечания.

Замечание 4.1.

Если матрица A системы (4.76) обладает свойством нестрогого диагонального преобладания, но хотя бы для одной ее строки выполнено условие строгого диагонального преобладания, то при выполнении для матрицы A еще одного дополнительного условия[a] метод Якоби (4.82) сходится для любого начального приближения $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$.

[a] Самостоятельно ознакомиться в литературе.

Замечание 4.2.

Перестановка уравнений в системе $A\boldsymbol{x} = \boldsymbol{b}$ влияет на сходимость метода Якоби (4.82).

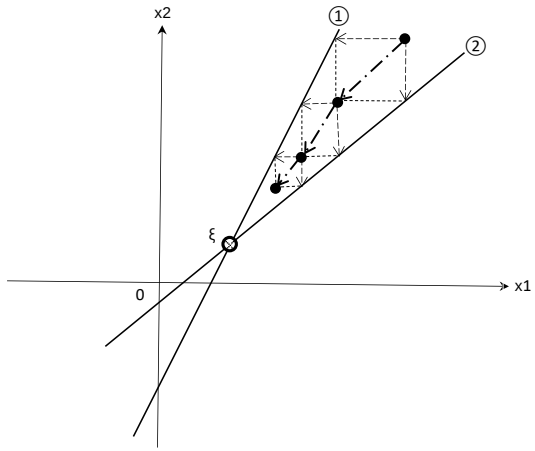
Пример

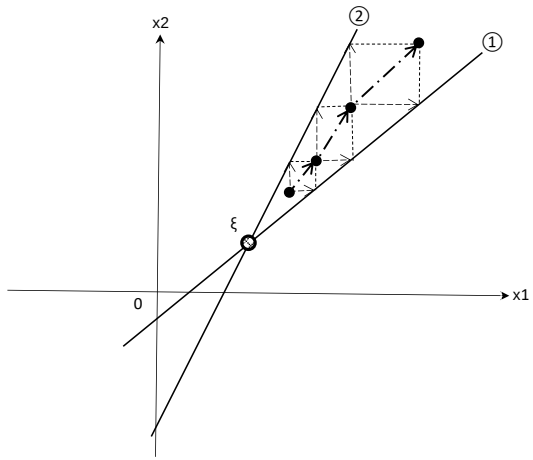
Пусть $n = 2$. Численная процедура метода Якоби

$$\begin{cases} a_{11}x_1^{(k+1)} + a_{12}x_2^{(k)} = b_1 \\ a_{21}x_1^{(k)} + a_{22}x_2^{(k+1)} = b_2 \end{cases} \tag{4.88}$$

Геометрическая интерпретация метода Якоби

Пусть $\xi \in \mathbb{R}^2$ — точное решение системы (4.88). $\boxtimes$





▲9 Метод Гаусса-Зейделя

Пусть выполнено условие (4.79), то есть $A \in \mathbb{R}^{n\times n}$:
$a_{ii} \neq 0 \quad \forall i = 1, 2, \ldots, n$.

Численная процедура метода Гаусса-Зейделя

$$x_i^{(k+1)} = -\frac{1}{a_{ii}} \left(\sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} + \sum_{j=i+1}^{n} a_{ij} x_j^{(k)} - b_i \right) \qquad (4.89)$$
$$i = 1, 2, \ldots, n,$$

где $k = 0, 1, 2, \ldots,$ $\left(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)} \right)^\top \in \mathbb{R}^n$.

Метод Гаусса-Зейделя в матричной форме имеет вид

$$(L + D)\boldsymbol{x}^{(k+1)} + R\boldsymbol{x}^{(k)} = \boldsymbol{b}$$

Численная процедура метода Гаусса-Зейделя в матричной форме

$$\boldsymbol{x}^{(k+1)} = -(L + D)^{-1} R\boldsymbol{x}^{(k)} + (L + D)^{-1}\boldsymbol{b}, \qquad (4.90)$$

где $k = 0, 1, 2, \ldots,$ $\left(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)} \right)^\top \in \mathbb{R}^n$.

Критерий сходимости метода Гаусса-Зейделя

Согласно теореме 4.3, критерий сходимости метода Гаусса-Зейделя (4.90) может быть сформулирован в терминах корней уравнения

$$det(\lambda E - B) = 0, \tag{4.91}$$

в котором $B = -(L + D)^{-1} R$.
Левая часть равенства (4.91) может быть преобразована следующим образом

$$\begin{aligned} det(\lambda E - B) &= det(\lambda E + (L + D)^{-1} R) = \\ &= det((L + D)^{-1} (\lambda L + \lambda D + R)) = \\ &= det((L + D)^{-1}) det(\lambda L + \lambda D + R) \end{aligned} \tag{4.92}$$

Поскольку $det((L + D)^{-1}) \neq 0$, то уравнение (4.91) равносильно

$$det(\lambda L + \lambda D + R) = 0. \tag{4.93}$$

Теорема 4.8.

Для сходимости метода Гаусса-Зейделя (4.90) необходимо и достаточно, чтобы абсолютные значения всех корней уравнения (4.93) были меньше единицы.

Достаточное условие сходимости метода Гаусса-Зейделя

Теорема 4.9. Достаточное условие сходимости

Если матрица A системы (4.76) является положительно-определенной, то метод Гаусса-Зейделя (4.90) сходится для любого начального приближения $\boldsymbol{x}^{(0)} \in \mathbb{R}^n$.

Доказательство.

Этап 1. Вспомогательная экстремальная задача

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} F(\boldsymbol{x}), \tag{4.94}$$

где $F(\boldsymbol{x}) = (\boldsymbol{x}, A\boldsymbol{x}) - 2\,(\boldsymbol{x}, \boldsymbol{b})$.

Необходимое условие экстремума для функции F имеет вид $\nabla F(\boldsymbol{x}) = 2A\boldsymbol{x} - 2\boldsymbol{b} = 0$, то есть $A\boldsymbol{x} = \boldsymbol{b}$.
Пусть $F_0(\boldsymbol{x}) = F(\boldsymbol{x}) - F(\bar{\boldsymbol{x}})$, где $\bar{\boldsymbol{x}} \in \mathbb{R}^n$: $A\bar{\boldsymbol{x}} = \boldsymbol{b}$.
Тогда, поскольку $A > 0$, то для любого $\boldsymbol{x} \in \mathbb{R}^n$

Достаточное условие сходимости метода Гаусса-Зейделя (продолжение)

$$F_0(\boldsymbol{x}) = F(\boldsymbol{x}) - F(\bar{\boldsymbol{x}}) = (\boldsymbol{x}, A\boldsymbol{x}) - 2(\boldsymbol{x}, \boldsymbol{b}) - (\bar{\boldsymbol{x}}, A\bar{\boldsymbol{x}}) + 2(\bar{\boldsymbol{x}}, \boldsymbol{b}) = \\ = (\boldsymbol{x}, A\boldsymbol{x}) - 2(\boldsymbol{x}, A\bar{\boldsymbol{x}}) - (\bar{\boldsymbol{x}}, A\bar{\boldsymbol{x}}) + 2(\bar{\boldsymbol{x}}, A\bar{\boldsymbol{x}}) = \\ = (\boldsymbol{x} - \bar{\boldsymbol{x}}, A(\boldsymbol{x} - \bar{\boldsymbol{x}})) \geq 0. \tag{4.95}$$

Таким образом, для любого $\boldsymbol{x} \in \mathbb{R}^n$ $F(\boldsymbol{x}) \geq F(\bar{\boldsymbol{x}})$.
Следовательно, $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ — единственная точка минимума[19] непрерывной функции F.

Этап 2. Метод покоординатного спуска

Пусть $\boldsymbol{x}^{(k)}$ — текущее приближение $\bar{\boldsymbol{x}}$. Тогда следующее приближение $\boldsymbol{x}^{(k+1)}$ на $k+1$-ой итерации метода покоординатного спуска определяется в результате выполнения следующей n-шаговой процедуры

[19] Матрицей Гессе для функции F является матрица $H(F) = A > 0$.

Достаточное условие сходимости метода Гаусса-Зейделя (продолжение)

$$
\begin{aligned}
&x_1^{(k+1)} = Arg\{\min_{x_1 \in \mathbb{R}} F\left(x_1, x_2^{(k)}, x_3^{(k)}, \ldots, x_n^{(k)}\right)\},\\
&x_2^{(k+1)} = Arg\{\min_{x_2 \in \mathbb{R}} F\left(x_1^{(k+1)}, x_2, x_3^{(k)}, \ldots, x_n^{(k)}\right)\},\\
&\ldots\ldots\ldots\ldots\\
&x_i^{(k+1)} = Arg\{\min_{x_i \in \mathbb{R}} F\left(x_1^{(k+1)}, \ldots, x_{i-1}^{(k+1)}, x_i, x_{i+1}^{(k)}, \ldots, x_n^{(k)}\right)\},\\
&\ldots\ldots\ldots\ldots\\
&x_n^{(k+1)} = Arg\{\min_{x_n \in \mathbb{R}} F\left(x_1^{(k+1)}, x_2^{(k+1)}, \ldots, x_{n-1}^{(k+1)}, x_n\right)\}.
\end{aligned}
\tag{4.96}
$$

Необходимым условием экстремума на i-ом шаге этой процедуры является следующее равенство

$$
\frac{\partial F}{\partial x_i}\left(x_1^{(k+1)}, \ldots, x_{i-1}^{(k+1)}, x_i, x_{i+1}^{(k)}, \ldots, x_n^{(k)}\right) = 0.
$$

Следовательно, экстремальная точка $x_i^{(k+1)}$ удовлетворяет равенству

Достаточное условие сходимости метода Гаусса-Зейделя (продолжение)

$$\frac{\partial F}{\partial x_i}\left(x_1^{(k+1)},\ldots,x_{i-1}^{(k+1)},x_i^{(k+1)},x_{i+1}^{(k)},\ldots,x_n^{(k)}\right)=$$
$$=2(a_{i1}x_1^{(k+1)}+\ldots+a_{ii-1}x_{i-1}^{(k+1)}+a_{ii}x_i^{(k+1)}+ \tag{4.97}$$
$$+a_{ii+1}x_{i+1}^{(k)}+\ldots+a_{in}x_n^{(k)}-b_i)=0.$$

Нетрудно видеть, что формулы (4.97) идентичны формулам (4.89), задающим численную процедуру Гаусса-Зейделя. Таким образом, метод покоординатного спуска как метод приближенного решения вспомогательной экстремальной задачи (4.94) реализуется как численная процедура метода Гаусса-Зейделя.

Этап 3. Сходимость метода покоординатного спуска

По построению процедуры (4.96) для любого $\boldsymbol{x}^{(k)}\neq\bar{\boldsymbol{x}}$ выполняется неравенство $F\left(\boldsymbol{x}^{(k+1)}\right)<F\left(\boldsymbol{x}^{(k)}\right)$.

Достаточное условие сходимости метода Гаусса-Зейделя (продолжение)

Тогда и $F_0\left(\boldsymbol{x}^{(k+1)}\right) < F_0\left(\boldsymbol{x}^{(k)}\right)\ \forall \boldsymbol{x}^{(k)} \neq \bar{\boldsymbol{x}}$. Учитывая (4.95), это неравенство равносильно

$$\frac{F_0\left(\boldsymbol{x}^{(k+1)}\right)}{F_0\left(\boldsymbol{x}^{(k)}\right)} = \frac{\left(\boldsymbol{x}^{(k+1)} - \bar{\boldsymbol{x}}, A(\boldsymbol{x}^{(k+1)} - \bar{\boldsymbol{x}})\right)}{\left(\boldsymbol{x}^{(k)} - \bar{\boldsymbol{x}}, A(\boldsymbol{x}^{(k)} - \bar{\boldsymbol{x}})\right)} < 1. \tag{4.98}$$

$\boldsymbol{x}^{(k+1)} = B\boldsymbol{x}^{(k)} + \boldsymbol{c}$, где $B = -\left(L + D\right)^{-1} R$ и $\boldsymbol{c} = -\left(L + D\right)^{-1} \boldsymbol{b}$,
$\boldsymbol{x}^{(k+1)} - \bar{\boldsymbol{x}} = B\boldsymbol{x}^{(k)} + \boldsymbol{c} - B\bar{\boldsymbol{x}} - \boldsymbol{c} = B\left(\boldsymbol{x}^{(k)} - \bar{\boldsymbol{x}}\right)$.
Тогда неравенство (4.98) можно записать в терминах функции $\varphi(\boldsymbol{r}) = \frac{(B\boldsymbol{r}, AB\boldsymbol{r})}{(\boldsymbol{r}, A\boldsymbol{r})}$, где $\boldsymbol{r} = \boldsymbol{x}^{(k)} - \bar{\boldsymbol{x}}$, следующим образом

$$\varphi(\boldsymbol{r}) < 1 \quad \forall \boldsymbol{r} \neq \boldsymbol{0}. \tag{4.99}$$

Поскольку функция φ непрерывна и $\|\boldsymbol{r}\| = 1$ — компактная сфера, то достигается

$$\sup_{\|\boldsymbol{r}\| \neq 0} \varphi(\boldsymbol{r}) = \sup_{\|\boldsymbol{r}\| = 1} \varphi(\boldsymbol{r}) = q < 1. \tag{4.100}$$

Достаточное условие сходимости метода Гаусса-Зейделя (продолжение)

Из (4.99) и (4.100) следует, что $\dfrac{F_0\left(\boldsymbol{x}^{(k+1)}\right)}{F_0\left(\boldsymbol{x}^{(k)}\right)} \leq q < 1 \quad \forall \boldsymbol{x}^{(k)} \neq \bar{\boldsymbol{x}}.$

Откуда

$$F_0\left(\boldsymbol{x}^{(k+1)}\right) \leq q^{k+1} F_0\left(\boldsymbol{x}^{(0)}\right).$$

Следовательно, при $k \to \infty$

$$F_0\left(\boldsymbol{x}^{(k)}\right) = \left(\boldsymbol{x}^{(k)} - \bar{\boldsymbol{x}}, A(\boldsymbol{x}^{(k)} - \bar{\boldsymbol{x}})\right) \to 0.$$

Отсюда с учетом $A > 0$ следует, что при $k \to \infty$

$$\boldsymbol{x}^{(k)} - \bar{\boldsymbol{x}} \to \mathbf{0}.$$

$\square$ Теорема доказана.

Метод Гаусса-Зейделя. Замечания.

Замечание 4.3.

Для системы $A\boldsymbol{x} = \boldsymbol{b}$ крамеровского типа можно построить равносильную ей систему

$$A^*A\boldsymbol{x} = A^*\boldsymbol{b} \tag{4.101}$$

с неотрицательно-определенной матрицей A^*A. Если матрица A^*A является положительно-определенной, то метод Гаусса-Зейделя (4.90) приближеннного решения системы (4.101) сходится.

Замечание 4.4.

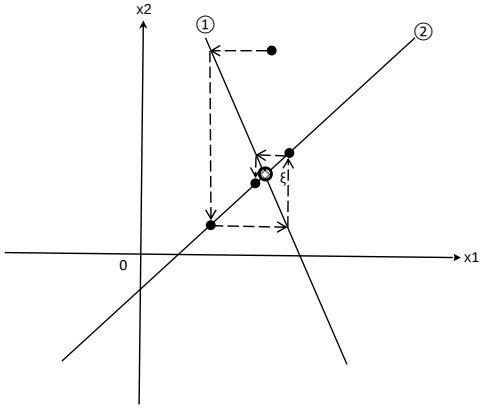
Перестановка уравнений в системе $A\boldsymbol{x} = \boldsymbol{b}$ влияет на сходимость метода Гаусса-Зейделя (4.90).

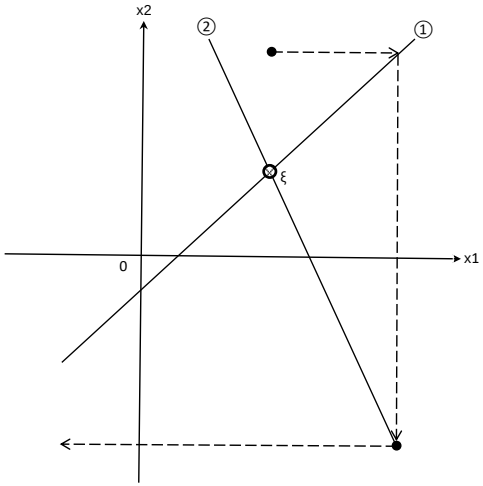
Пример

Пусть $n = 2$. Численная процедура метода Гаусса-Зейделя

$$\begin{cases} a_{11}x_1^{(k+1)} + a_{12}x_2^{(k)} = b_1 \\ a_{21}x_1^{(k+1)} + a_{22}x_2^{(k+1)} = b_2 \end{cases} \tag{4.102}$$

Геометрическая интерпретация метода Гаусса-Зейделя





О скорости сходимости метода Гаусса-Зейделя

Как было показано при доказательстве теоремы 4.9, численная процедура метода Гаусса-Зейделя реализуется как численная процедура метода покоординатного спуска в задаче минимизации функции

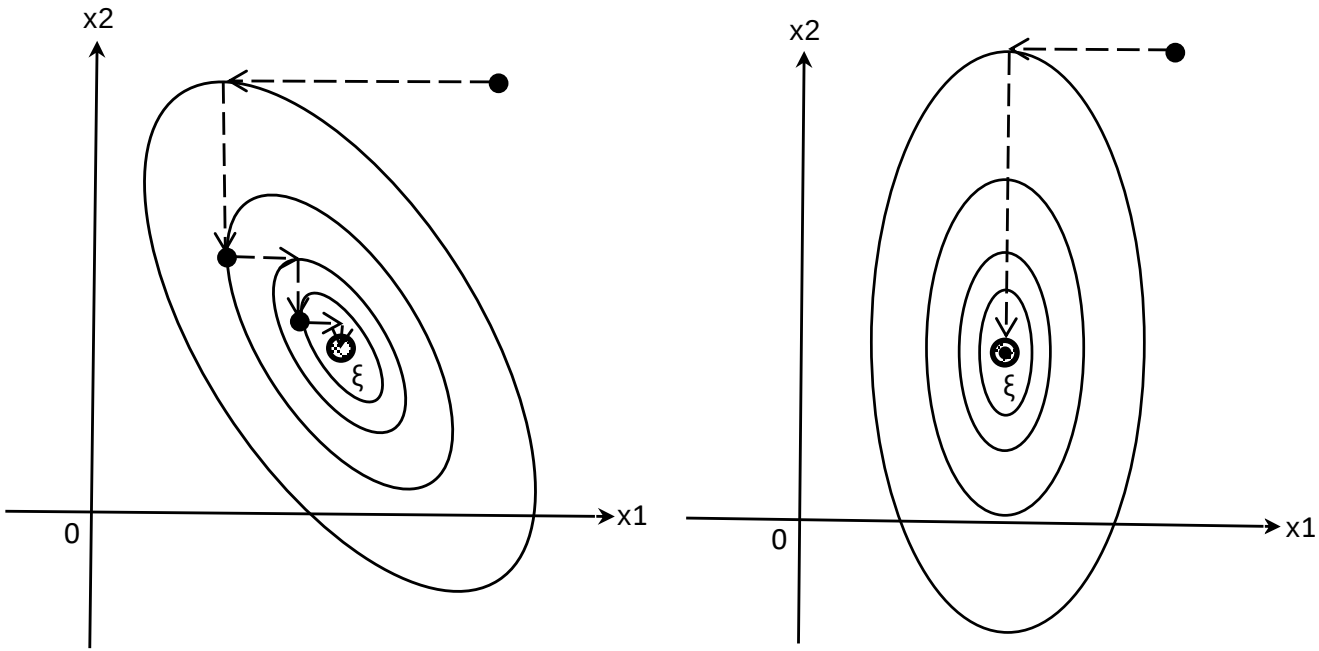
$$F_0(\boldsymbol{x}) = (\boldsymbol{x} - \bar{\boldsymbol{x}}, A(\boldsymbol{x} - \bar{\boldsymbol{x}}))\,,$$

где $A\bar{\boldsymbol{x}} = \boldsymbol{b},\ A > 0$.

В двумерном случае ($n = 2$) фазовый портрет функции F_0 (набор линий уровня этой функции) представляет собой набор концентрических эллипсов, главные полуоси которых задаются двумя ортогональными собственными векторами матрицы A, выпущенными из точки $\bar{\boldsymbol{x}}$.

Реализация процедуры метода покоординатного спуска для различных вариантов ориентации собственных векторов матрицы A может проиллюстрирована следующим образом.

О скорости сходимости метода Гаусса-Зейделя. Геометрическая интерпретация метода покоординатного спуска.



ТЕМА 5. Численные методы решения систем нелинейных уравнений

Система нелинейных уравнений

$$f_i(x_1, x_2, \ldots, x_n) = 0, \quad i = 1, 2, \ldots, n \tag{5.1}$$

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}, \tag{5.2}$$

где $\boldsymbol{f} = (f_1, f_2, \ldots, f_n)^\top$, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top \in \mathbb{R}^n$.

Задачи

1. Отделение корней;
2. Уточнение корней.

Методы уточнения корней

1. Метод Ньютона;
2. Метод простой итерации.

5.1. Метод Ньютона

Определение 5.1.

Отображение (вектор-функция) $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ называется дифференцируемым по Фреше в точке $\boldsymbol{x} \in \mathrm{D}[\boldsymbol{f}]$, если

$$\exists P(\boldsymbol{x}) \in \mathbb{R}^{n\times n} : \ \forall \boldsymbol{x} + \Delta\boldsymbol{x} \in \mathrm{D}[\boldsymbol{f}] \Rightarrow$$
$$\|\boldsymbol{f}(\boldsymbol{x} + \Delta\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}) - P(\boldsymbol{x})\Delta\boldsymbol{x}\| = o(\|\Delta\boldsymbol{x}\|) \tag{5.3}$$

Определение 5.2.

Отображение (вектор-функция) $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$ называется дифференцируемым по направлению $\boldsymbol{z} \in \mathbb{R}^n$ в точке $\boldsymbol{x} \in \mathrm{D}[\boldsymbol{f}]$, если

$$\exists P(\boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R}^{n\times n} : \quad \lim_{t\to 0} \frac{\boldsymbol{f}(\boldsymbol{x} + t\boldsymbol{z}) - \boldsymbol{f}(\boldsymbol{x})}{t} = P(\boldsymbol{x}, \boldsymbol{z})\,\boldsymbol{z}. \tag{5.4}$$

Если матрица P не зависит от вектора $\boldsymbol{z}$, то вектор-функцию называют дифференцируемой по Гато в точке $\boldsymbol{x}$.

Дифференцируемость вектор-функции

Из дифференцируемости вектор-функции $\boldsymbol{f}$ по Фреше в некоторой точке $\boldsymbol{x} \in \mathrm{D}[\boldsymbol{f}]$ следует ее дифференцируемость по Гато в этой точке.

В случае дифференцируемости вектор-функции $\boldsymbol{f}$ по Гато матрицу P можно сформировать, взяв в качестве направлений $\boldsymbol{z}^{(j)} = (\underbrace{0, \ldots, 0}_{j-1}, 1, 0, \ldots, 0)^\top \in \mathbb{R}^n$ — j-ые единичные орты $(j = 1, 2, \ldots, n)$. Тогда в качестве производной вектор-функции $\boldsymbol{f}$ в точке $\boldsymbol{x}$ выступает матрица Якоби

$$\boldsymbol{f}'(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{x}) & \frac{\partial f_1}{\partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\boldsymbol{x}) \\ \frac{\partial f_2}{\partial x_1}(\boldsymbol{x}) & \frac{\partial f_2}{\partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial f_2}{\partial x_n}(\boldsymbol{x}) \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial f_n}{\partial x_1}(\boldsymbol{x}) & \frac{\partial f_n}{\partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial f_n}{\partial x_n}(\boldsymbol{x}) \end{pmatrix} \tag{5.5}$$

Дифференцируемость вектор-функции по Фреше

Достаточное условие дифференцируемости

Если f_i $(i = 1, 2, \ldots, n)$ определены и непрерывны со своими частными производными в некоторой окрестности точки $\boldsymbol{x}$, то вектор-функция $\boldsymbol{f}$ дифференцируема по Фреше в точке $\boldsymbol{x}$ и ее производная в этой точке

$$P(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x}) \tag{5.6}$$

20

Общее предположение

Пусть вектор-функция $\boldsymbol{f}$ дифференцируема по Фреше на области своего определения

---

[20] В отличие от производной по Гато (5.5) в производной по Фреше (5.6) все частные производные $\frac{\partial f_i}{\partial x_j}$ $(i, j = 1, 2, \ldots, n)$ должны быть непрерывны в окрестности рассматриваемой точки.

Метод Ньютона. Численная процедура метода.

Пусть $\boldsymbol{\xi}$ — решение системы (5.2), то есть $\boldsymbol{f}(\boldsymbol{\xi}) = \mathbf{0}$, $\boldsymbol{x}^{(k)}$ — текущее приближение $\boldsymbol{\xi}$. Тогда

$$\|\boldsymbol{f}(\boldsymbol{\xi}) - \boldsymbol{f}(\boldsymbol{x}^{(k)}) - \boldsymbol{f}'(\boldsymbol{x}^{(k)})\left(\boldsymbol{\xi} - \boldsymbol{x}^{(k)}\right)\| = o\left(\|\boldsymbol{\xi} - \boldsymbol{x}^{(k)}\|\right).$$

Отсюда

$$\boldsymbol{\xi} \approx \boldsymbol{x}^{(k)} - [\boldsymbol{f}'(\boldsymbol{x}^{(k)})]^{-1}\boldsymbol{f}(\boldsymbol{x}^{(k)}).$$

Численная процедура метода Ньютона

$$\begin{gathered}\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - [\boldsymbol{f}'(\boldsymbol{x}^{(k)})]^{-1}\boldsymbol{f}(\boldsymbol{x}^{(k)}), \\ k = 0, 1, 2, \ldots\end{gathered} \tag{5.7}$$

где $\boldsymbol{x}^{(0)} \in \mathrm{D}[\boldsymbol{f}]$; $\boldsymbol{f}(\boldsymbol{x}^{(k)}) \in \mathbb{R}^n$; $[\boldsymbol{f}'(\boldsymbol{x}^{(k)})]^{-1} \in \mathbb{R}^{n\times n}$.

Реализация численной процедуры (5.8):

$$\begin{cases} \boldsymbol{f}'(\boldsymbol{x}^{(k)})\Delta\boldsymbol{x}^{(k)} = -\boldsymbol{f}(\boldsymbol{x}^{(k)}) \\ \boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \Delta\boldsymbol{x}^{(k)} \end{cases} \tag{5.8}$$

$k = 0, 1, 2, \ldots$.

Сходимость метода Ньютона

Теорема 5.1. Достаточные условия локальной сходимости

Пусть выполнены следующие условия:
a) функции f_i $(i = 1, 2, \ldots, n)$ непрерывно-дифференцируемы в некоторой окрестности решения $\boldsymbol{\xi}$ системы (5.2);
b) у функций f_i $(i = 1, 2, \ldots, n)$ существуют и равномерно-ограничены частные производные второго порядка в некоторой окрестности $\boldsymbol{\xi}$;
c) матрица $\boldsymbol{f}'(\boldsymbol{\xi})$ обратима.
Тогда для любого начального приближения $\boldsymbol{x}^{(0)}$ из некоторой окрестности точки $\boldsymbol{\xi}$ метод Ньютона (5.8) сходится и справедлива следующая оценка

$$\exists C \in \mathbb{R}: \quad \|\boldsymbol{x}^{(k+1)} - \boldsymbol{\xi}\| \leq C\|\boldsymbol{x}^{(k)} - \boldsymbol{\xi}\|^2 \quad \forall k = 0, 1, 2, \ldots. \tag{5.9}$$

Доказательство

Построим искомую окрестность $U(\boldsymbol{\xi})$ точки $\boldsymbol{\xi}$.

Доказательство ...

Поскольку $\forall \boldsymbol{x}, \boldsymbol{y} \in U(\boldsymbol{\xi})\ \forall i = 1, 2, \ldots, n$

$$f_i(\boldsymbol{y}) - f_i(\boldsymbol{x}) = \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(\boldsymbol{x})\,(y_j - x_j) + O(\|\boldsymbol{y} - \boldsymbol{x}\|^2),$$

то силу дифференцируемости вектор-функции $\boldsymbol{f}$ и равномерной ограниченности частных производных второго порядка функций f_i $(i = 1, 2, \ldots, n)$ в некоторой окрестности $U(\boldsymbol{\xi})$ для любых $\boldsymbol{x}, \boldsymbol{y} \in U(\boldsymbol{\xi})$ справедливо неравенство

$$\|\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}'(\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x})\| \le K_1 \|\boldsymbol{y} - \boldsymbol{x}\|^2. \tag{5.10}$$

Поскольку матрица $\boldsymbol{f}'(\boldsymbol{\xi})$ обратима (см. условие (c)), то из условия (a) следует, что эта матрица обратима в малой окрестности $O_\delta(\boldsymbol{\xi}) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{\xi}\| \le \delta\} \subseteq U(\boldsymbol{\xi})$ и

$$\|[\boldsymbol{f}'(\boldsymbol{x})]^{-1}\| \le K_2 \quad \forall \boldsymbol{x} \in O_\delta(\boldsymbol{\xi}). \tag{5.11}$$

Здесь K_1 и K_2 — положительные константы. Тогда

$$\begin{aligned} \boldsymbol{x}^{(1)} - \boldsymbol{\xi} &= \boldsymbol{x}^{(0)} - [\boldsymbol{f}'(\boldsymbol{x}^{(0)})]^{-1} \boldsymbol{f}(\boldsymbol{x}^{(0)}) - \boldsymbol{\xi} = \\ &= [\boldsymbol{f}'(\boldsymbol{x}^{(0)})]^{-1} \left(\boldsymbol{f}(\boldsymbol{\xi}) - \boldsymbol{f}(\boldsymbol{x}^{(0)}) - \boldsymbol{f}'(\boldsymbol{x}^{(0)})(\boldsymbol{\xi} - \boldsymbol{x}^{(0)})\right). \end{aligned} \tag{5.12}$$

Доказательство ...

Если $\boldsymbol{x}^{(0)} \in O_\delta(\boldsymbol{\xi})$, то в силу (5.10), (5.11), (5.12) получаем оценку

$$\|\boldsymbol{x}^{(1)} - \boldsymbol{\xi}\| \le K_2 K_1 \|\boldsymbol{x}^{(0)} - \boldsymbol{\xi}\|^2, \tag{5.13}$$

которую можно переписать в следующем виде

$$K_2 K_1 \|\boldsymbol{x}^{(1)} - \boldsymbol{\xi}\| \le \left(K_2 K_1 \|\boldsymbol{x}^{(0)} - \boldsymbol{\xi}\|\right)^2. \tag{5.14}$$

Рассмотрим некоторую окрестность
$O_\varepsilon(\boldsymbol{\xi}) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{\xi}\| \le \varepsilon\} \subseteq O_\delta(\boldsymbol{\xi})$. Для любого начального приближения $\boldsymbol{x}^{(0)} \in O_\varepsilon(\boldsymbol{\xi})$, то из (5.13), (5.15) имеем

$$\|\boldsymbol{x}^{(1)} - \boldsymbol{\xi}\| \le K_2 K_1 \varepsilon^2.$$

Требование, чтобы $K_2 K_1 \varepsilon^2 < \varepsilon$ ($\boldsymbol{x}^{(1)} \in O_\varepsilon(\boldsymbol{\xi})$), приводит к ограничению на ε

$$\varepsilon < \frac{1}{K_2 K_1}. \tag{5.15}$$

Доказательство ...

Введем обозначение $q_k = K_2 K_1 \|\boldsymbol{x}^{(k)} - \boldsymbol{\xi}\|$, $k = 0, 1, 2, \ldots$.
Из (5.15) следует, что

$$q_0 = K_2 K_1 \|\boldsymbol{x}^{(0)} - \boldsymbol{\xi}\| \leq K_2 K_1 \varepsilon < 1. \tag{5.16}$$

Таким образом[21], в результате приходим к неравенствам

$$q_k \leq q_0^{2^k}, \quad k = 1, 2, 3, \ldots. \tag{5.17}$$

Из (5.16), (5.17) следует, что

$$q_k \to 0 \quad \text{при} \quad k \to \infty.$$

□ Теорема доказана.

Фигурирующая в (5.9) константа $C = K_2 K_1$, а радиус ε искомой окрестности $U(\boldsymbol{\xi})$ должен удовлетворять условию (5.15).

[21] Здесь с учетом (5.14) $q_1 \leq q_0^2 < 1$.

5.2. Метод простой итерации

Пусть система (5.1) равносильна системе

$$x_i = \varphi_i(x_1, x_2, \ldots, x_n), \quad i = 1, 2, \ldots, n \tag{5.18}$$

$$\boldsymbol{x} = \boldsymbol{\varphi}(\boldsymbol{x}), \tag{5.19}$$

где $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \ldots, \varphi_n)^\top$, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top \in \mathbb{R}^n$.

Численная процедура метода простой итерации

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{\varphi}\left(\boldsymbol{x}^{(k)}\right), \tag{5.20}$$

где $k = 0, 1, 2, \ldots$; $\boldsymbol{x}^{(0)} \in \mathrm{D}[\boldsymbol{\varphi}]$.

Матрица Якоби отображения $\boldsymbol{\varphi}$

$$\boldsymbol{\varphi}'(\boldsymbol{x}) = \begin{pmatrix} \frac{\partial \varphi_1}{\partial x_1}(\boldsymbol{x}) & \frac{\partial \varphi_1}{\partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial \varphi_1}{\partial x_n}(\boldsymbol{x}) \\ \frac{\partial \varphi_2}{\partial x_1}(\boldsymbol{x}) & \frac{\partial \varphi_2}{\partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial \varphi_2}{\partial x_n}(\boldsymbol{x}) \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial \varphi_n}{\partial x_1}(\boldsymbol{x}) & \frac{\partial \varphi_n}{\partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial \varphi_n}{\partial x_n}(\boldsymbol{x}) \end{pmatrix} \tag{5.21}$$

Достаточное условие сходимости метода простой итерации

Теорема 5.2. Достаточное условие сходимости

Пусть функции φ_i $(i = 1, 2, \ldots, n)$ непрерывно-дифференцируемы в окрестности решения $\boldsymbol{\xi}$ системы (5.19) и

$$\|\boldsymbol{\varphi}'(\boldsymbol{\xi})\| < 1 \tag{5.22}$$

для подчиненной нормы.
Тогда метод простой итерации (5.20) сходится для любого начального приближения $\boldsymbol{x}^{(0)}$ из некоторой окрестности точки $\boldsymbol{\xi}$.

Доказательство

Рассмотрим разности скалярной функции от векторного аргумента $\varphi_i(\boldsymbol{y}) - \varphi_i(\boldsymbol{x})$, $i = 1, 2, \ldots, n$.
Сведем разности значений функции векторного аргумента к разности значений функции скалярного аргумента.

Доказательство ...

Введем вспомогательную векторную функцию скалярного аргумента:

$$\boldsymbol{x}(t) = \boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}).$$

Тогда

$$\begin{aligned} &\boldsymbol{x}(0) = \boldsymbol{x}, \quad \boldsymbol{x}(1) = \boldsymbol{y}, \\ &\varphi_i(\boldsymbol{y}) - \varphi_i(\boldsymbol{x}) = \varphi_i(\boldsymbol{x}(1)) - \varphi_i(\boldsymbol{x}(0)) = \psi_i(1) - \psi_i(0), \end{aligned}$$

где $\psi_i(t) = \psi_i(\boldsymbol{x}(t))$ — сложная функция.
Применим к функции $\psi_i(t)$ формулу конечных приращений Лагранжа:

$$\psi_i(1) - \psi_i(0) = \psi_i'(c_i)(1 - 0), \quad c_i \in [0, 1].$$

Вычисляя производную функции $\psi_i(t)$, получаем

$$\psi_i(1) - \psi_i(0) = \sum_{j=1}^{n} \frac{\partial \varphi_i}{\partial x_j}(\boldsymbol{x}(c_i))\,(y_j - x_j).$$

Доказательство ...

Возвращаясь к векторной функции векторного аргумента, получаем

$$\boldsymbol{\varphi}(\boldsymbol{y}) - \boldsymbol{\varphi}(\boldsymbol{x}) = B(c_1, c_2, ..., c_n)(\boldsymbol{y} - \boldsymbol{x}),$$

где

$$B(c_1, c_2, ..., c_n) = \begin{pmatrix} \frac{\partial \varphi_1}{\partial x_1}(\boldsymbol{x}(c_1)) & \frac{\partial \varphi_1}{\partial x_2}(\boldsymbol{x}(c_1)) & \cdots & \frac{\partial \varphi_1}{\partial x_n}(\boldsymbol{x}(c_1)) \\ \frac{\partial \varphi_2}{\partial x_1}(\boldsymbol{x}(c_2)) & \frac{\partial \varphi_2}{\partial x_2}(\boldsymbol{x}(c_2)) & \cdots & \frac{\partial \varphi_2}{\partial x_n}(\boldsymbol{x}(c_2)) \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial \varphi_n}{\partial x_1}(\boldsymbol{x}(c_n)) & \frac{\partial \varphi_n}{\partial x_2}(\boldsymbol{x}(c_n)) & \cdots & \frac{\partial \varphi_n}{\partial x_n}(\boldsymbol{x}(c_n)) \end{pmatrix}.$$

В силу непрерывности частных производных выполняется $\|\boldsymbol{\varphi}'(\boldsymbol{x})\| \leq q < 1$ в некоторой окрестности $U(\boldsymbol{\xi})$ точки $\boldsymbol{\xi}$ и $\|B\| \leq q < 1$, если $\boldsymbol{x}$ и $\boldsymbol{y}$ принадлежат этой окрестности. Тогда

$$\|\boldsymbol{\varphi}(\boldsymbol{y}) - \boldsymbol{\varphi}(\boldsymbol{x})\| \leq q\|\boldsymbol{y} - \boldsymbol{x}\|.$$

Сузим $U(\boldsymbol{\xi})$ до его подмножества — некоторого шара в той метрике, в которой рассматривается подчиненная норма, с центром в $\boldsymbol{\xi}$, который будем обозначать так же.

Доказательство ...

Докажем, что если некоторое приближение $\boldsymbol{x}^{(k)} \in U(\boldsymbol{\xi})$, то $\boldsymbol{x}^{(k+1)} \in U(\boldsymbol{\xi})$.
В самом деле,

$$\|\boldsymbol{x}^{(k+1)} - \boldsymbol{\xi}\| = \|\boldsymbol{\varphi}(\boldsymbol{x}^{(k)}) - \boldsymbol{\varphi}(\boldsymbol{\xi})\| \leq q\|\boldsymbol{x}^{(k)} - \boldsymbol{\xi}\| < \|\boldsymbol{x}^{(k)} - \boldsymbol{\xi}\|.$$

Таким образом, если взять начальное приближение $\boldsymbol{x}^{(0)}$ из этой окрестности, то все последующие приближения также будут в этой окрестности. Таким образом, имеем

$$\|\boldsymbol{x}^{(k)} - \boldsymbol{\xi}\| \leq q\|\boldsymbol{x}^{(k-1)} - \boldsymbol{\xi}\| \leq \ldots \leq q^k\|\boldsymbol{x}^{(0)} - \boldsymbol{\xi}\|,$$

откуда и следует сходимость $\boldsymbol{x}^{(k)} \to \boldsymbol{\xi}$ при $k \to \infty$.
□ Теорема доказана.
⊠

Постановка задачи

Пусть заданы набор данных

$$\begin{matrix} x_0, x_1, x_2, \ldots, x_n \in [a,b]: & x_i \neq x_j \quad \forall i \neq j \\ y_0, y_1, y_2, \ldots, y_n \end{matrix} \tag{6.1}$$

и класс функций Φ.

$$\varphi \in \Phi: \ \varphi(x_i) = y_i \quad \forall i = 0, 1, 2, \ldots, n. \tag{6.2}$$

Числа x_i $(i = 0, 1, 2, \ldots, n)$ называют узлами интерполяции[22].

Примеры Φ:

- $\Phi = \langle \sin kx, \cos kx \rangle, \ k = 1, 2, \ldots;$
- $\Phi = \langle \exp^{kx} \rangle, \ k = 1, 2, \ldots;$
- $\Phi = \langle x^k \rangle, \ k = 0, 1, 2, \ldots$

[22] На классе функций Φ требуется определить функцию, график которой проходит через заданные точки $(x_i, y_i) \quad i = 0, 1, 2, \ldots, n.$

6.1. Решение задачи численной интерполяции

Пусть

$$\Phi = \langle \varphi_0(x), \varphi_1(x), \varphi_2(x), \ldots, \varphi_n(x) \rangle \quad x \in [a,b] \subset \mathrm{D}[\varphi_i] \quad (i = 0, 1, 2, \ldots, n). \tag{6.3}$$

$$\forall \varphi \in \Phi \quad \exists (a_0, a_1, a_2, \ldots, a_n)^\top \in \mathbb{R}^{n+1} \tag{6.4}$$

$$\varphi(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + a_2\varphi_2(x) + \ldots + a_n\varphi_n(x).$$

Тогда, в силу (6.3) и (6.4), задача численной интерполяции (6.2) сводится к решению следующей системы линейных алгебраических уравнений относительно неизвестных $a_0, a_1, a_2, \ldots, a_n$

$$a_0\varphi_0(x_i) + a_1\varphi_1(x_i) + a_2\varphi_2(x_i) + \ldots + a_n\varphi_n(x_i) = y_i \quad i = 0, 1, 2, \ldots, n \tag{6.5}$$

Возникает вопрос о существовании и единственности решения системы (6.5) для заданных набора данных (6.1) и класса функций Φ.

Чебышевская система функций

Пример

Пусть $x_0 = -1$, $x_1 = 1$; $\Phi = \langle 1, x^2 \rangle$.
Тогда $\forall y_0, y_1$ определитель матрицы системы (6.5)

$$det \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = 0.$$

Определение 6.1.

Система функций $\varphi_0, \varphi_1, \varphi_2, \ldots, \varphi_n$ называется чебышевской на отрезке $[a, b] \subset \mathrm{D}[\varphi_i]$ $(i = 0, 1, 2, \ldots, n)$, если любая нетривиальная линейная комбинация

$$a_0\varphi_0(x) + a_1\varphi_1(x) + a_2\varphi_2(x) + \ldots + a_n\varphi_n(x)$$

этих функций имеет на отрезке $[a, b]$ не более, чем n нулей.

Разрешимость задачи численной интерполяции

Примеры

- Система функций
  $\varphi_0(x) \equiv 1, \varphi_1(x) = x, \varphi_2(x) = x^2, \ldots, \varphi_n(x) = x^n$
  является чебышевской на любом отрезке $[a, b]$;
- Система функций
  $\varphi_0(x) = x, \varphi_1(x) = x^2, \varphi_2(x) = x^3, \ldots, \varphi_n(x) = x^{n+1}$
  не является чебышевской на любом отрезке $[a, b]$, который содержит 0.

Теорема 6.1.

Для того, чтобы определитель системы (6.5) был отличен от нуля для любого набора узлов интерполяции $x_0, x_1, x_2, \ldots, x_n \in [a, b]$ необходимо и достаточно, чтобы система функций

$$\varphi_0, \varphi_1, \varphi_2, \ldots, \varphi_n$$

была чебышевской на отрезке $[a, b]$.

Доказательство критерия разрешимости задачи численной интерполяции

Необходимость. От противного.

Пусть определитель системы (6.5) отличен от нуля для любого набора узлов интерполяции $x_0, x_1, x_2, \ldots, x_n \in [a, b]$, а система функций $\varphi_0, \varphi_1, \varphi_2, \ldots, \varphi_n$ не является чебышевской на отрезке $[a, b]$. Это означает, что существует такая нетривиальная линейная комбинация

$$\bar{a}_0\varphi_0(x) + \bar{a}_1\varphi_1(x) + \bar{a}_2\varphi_2(x) + \ldots + \bar{a}_n\varphi_n(x)$$

этих функций, которая на отрезке $[a, b]$ имеет, по крайней мере, $n + 1$ нулей $\bar{x}_0, \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n \in [a, b]$. Тогда система (6.5) становится однородной

$$\begin{gathered}\bar{a}_0\varphi_0(\bar{x}_i) + \bar{a}_1\varphi_1(\bar{x}_i) + \bar{a}_2\varphi_2(\bar{x}_i) + \ldots + \bar{a}_n\varphi_n(\bar{x}_i) = 0 \\ i = 0, 1, 2, \ldots, n,\end{gathered} \tag{6.6}$$

которая имеет нетривиальное решение $(\bar{a}_0, \bar{a}_1, \bar{a}_2, \ldots, \bar{a}_n) \neq \mathbf{0}$. Следовательно, определитель матрицы системы (6.6) равен нулю. Возникает противоречие.

Доказательство критерия разрешимости задачи численной интерполяции

Достаточность. От противного.

Пусть система функций $\varphi_0, \varphi_1, \varphi_2, \ldots, \varphi_n$ является чебышевской на отрезке $[a, b]$, но существует набор узлов интерполяции $\bar{x}_0, \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n \in [a, b]$, для которого определитель системы (6.5) равен нулю. Тогда для этого набора узлов соответствующая системе (6.5) однородная система (6.6) имеет нетривиальное решение $(\bar{a}_0, \bar{a}_1, \bar{a}_2, \ldots, \bar{a}_n) \neq \mathbf{0}$. Следовательно, для нетривиальной линейной комбинации

$$\bar{a}_0\varphi_0(x) + \bar{a}_1\varphi_1(x) + \bar{a}_2\varphi_2(x) + \ldots + \bar{a}_n\varphi_n(x)$$

$n + 1$ узлов $\bar{x}_0, \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n \in [a, b]$ являются нулями. Получено противоречие с чебышевостью системы функций $\varphi_0, \varphi_1, \varphi_2, \ldots, \varphi_n$ на отрезке $[a, b]$.
□ Теорема доказана.

Решение задачи численной интерполяции на классе многочленов степени не более, чем n

$\Phi = \langle 1, x, x^2, \ldots, x^n \rangle$

$\varphi(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \quad \forall \varphi \in \Phi$

Система (6.5) имеет вид

$$\begin{cases} a_0 + a_1 x_0 + a_2 x_0^2 + \ldots + a_n x_0^n = y_0 \\ a_0 + a_1 x_1 + a_2 x_1^2 + \ldots + a_n x_1^n = y_1 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_0 + a_1 x_n + a_2 x_n^2 + \ldots + a_n x_n^n = y_n \end{cases} \tag{6.7}$$

Так как система функций $1, x, x^2, \ldots, x^n$ является чебышевской на любом отрезке $[a, b]$, то для любых $x_0, x_1, x_2, \ldots, x_n \in [a, b]$ система (6.7) имеет единственное решение[23].

Кроме того, известно, что определитель системы (6.7)[24] отличен от нуля для любого набора различных точек $x_0, x_1, x_2, \ldots, x_n$.

[23] Через любые $n + 1$ точек можно провести единственный многочлен n-ой степени.

[24] Определитель системы (6.7) называется определителем Вандермонда.

6.2. Интерполяционный многочлен Лагранжа

Постановка задачи

Пусть задан набор данных (6.1). Требуется построить многочлен степени n

$$\begin{aligned} &P_n(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n : \\ &P_n(x_i) = y_i \quad \forall i = 0, 1, 2, \ldots, n. \end{aligned} \tag{6.8}$$

Частная задача

$$\begin{aligned} &l_k(x) \in \Phi = \langle 1, x, x^2, \ldots, x^n \rangle : \\ &l_k(x_i) = \delta_{ik} = \begin{cases} 0, & i \neq k \\ 1, & i = k \end{cases} \quad \forall i = 0, 1, 2, \ldots, n \\ &k = 0, 1, 2, \ldots, n. \end{aligned} \tag{6.9}$$

Решение частной задачи (6.9)

$$\begin{aligned} &l_k(x) = \frac{(x-x_0)(x-x_1)\ldots(x-x_{k-1})(x-x_{k+1})\ldots(x-x_n)}{(x_k-x_0)(x_k-x_1)\ldots(x_k-x_{k-1})(x_k-x_{k+1})\ldots(x_k-x_n)} \\ &k = 0, 1, 2, \ldots, n. \end{aligned} \tag{6.10}$$

Интерполяционный многочлен Лагранжа ...

Решение общей задачи (6.8)

$$L_n(x) = \sum_{k=0}^{n} y_k l_k(x) \tag{6.11}$$

Классическая форма интерполяционного многочлена Лагранжа

Пусть для некоторой функции f известны ее значения в $n+1$-ой точке $x_i \in \mathrm{D}[f]$ $(i = 0, 1, 2, \ldots, n)$.

$$L_n(x) = \sum_{k=0}^{n} f(x_k) \Pi_{i=0, i\neq k}^{n} \frac{x - x_i}{x_k - x_i} \tag{6.12}$$

Основным недостатком формулы (6.12) при ее практическом использовании является необходимость заново пересчитывать все коэффициенты интерполяционного многочлена Лагранжа при изменении набора данных (x_i, y_i) $(i = 0, 1, 2, \ldots, n)$.

6.3. Погрешность интерполяционного многочлена Лагранжа

Определение 6.2.

Пусть для некоторой функции f известны ее значения в $n+1$-ой точке $x_i \in [a,b] \subseteq \mathrm{D}[f]$ $(i = 0, 1, 2, \ldots, n)$. Тогда под погрешностью интерполяционного многочлена Лагранжа (6.12), построенного для функции f, (погрешностью метода) понимается

$$R_n(x) = f(x) - L_n(x) \quad x \in [a,b] \tag{6.13}$$

В предположении, что $f \in C^{(n+1)}([a,b])$ требуется определить структуру погрешности (6.13). Предлагается искать $R_n(x)$ в следующем виде

$$R_n(x) = g(x)\omega_n(x) \quad x \in [a,b], \tag{6.14}$$

где

$$\omega_n(x) = (x - x_0)(x - x_1)(x - x_2)\ldots(x - x_n). \tag{6.15}$$

Здесь $g(x)$ — некоторая функция, которую и следует определить.

Структура погрешности интерполяционного многочлена Лагранжа

Вспомогательная функция

$$F(x) = f(x) - L_n(x) - g(\bar{x})\omega_n(x) \quad x \in [a, b], \tag{6.16}$$

где $\bar{x} \in [a, b]$ — произвольно-выбранная и зафиксированная точка.

Свойства функции F:

1. $F \in C^{(n+1)}([a, b])$, так как $f \in C^{(n+1)}([a, b])$;
2. По построению F: $F(\bar{x}) = 0$, $F(x_i) = 0$ $(i = 0, 1, 2, \ldots, n)$.

Таким образом, функция гладкая функция F обращается в ноль в $n + 2$ точках отрезка $[a, b]$. Здесь без потери общности можно считать, что

$$a \leq x_0 < x_1 < x_2 < \ldots < x_n < \bar{x} \leq b$$

Последовательное применение теоремы Ролля приводит к следующим заключениям:

Структура погрешности интерполяционного многочлена Лагранжа ...

1) $F'(x_i^{(1)}) = 0 \ (i = 1, 2, \ldots, n+1)$:
$$x_j^{(1)} \in [x_{j-1}, x_j] \ \ (j = 1, 2, \ldots, n), \ \ x_{n+1}^{(1)} \in [x_n, \bar{x}];$$
2) $F''(x_i^{(2)}) = 0 \ (i = 1, 2, \ldots, n)$:
$$x_j^{(2)} \in [x_j^{(1)}, x_{j+1}^{(1)}] \ (j = 1, 2, \ldots, n);$$
3) $F^{(3)}(x_i^{(3)}) = 0 \ (i = 1, 2, \ldots, n-1)$:
$$x_j^{(3)} \in [x_j^{(2)}, x_{j+1}^{(2)}] \ (j = 1, 2, \ldots, n-1);$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
n) $F^{(n)}(x_i^{(n)}) = 0 \ (i = 1, 2)$:
$$x_j^{(n)} \in [x_j^{(n-1)}, x_{j+1}^{(n-1)}] \ (j = 1, 2);$$
$n+1$) $F^{(n+1)}(\xi) = 0$:
$$\xi \in [x_1^{(n)}, x_2^{(n)}].$$

Структура погрешности интерполяционного многочлена Лагранжа ...

Поскольку ${\omega_n}^{(n+1)}(x) \equiv (n+1)!$ и ${L_n}^{(n+1)}(x) \equiv 0$, то из (6.16) следует, что

$$F^{(n+1)}(x) = f^{(n+1)}(x) - (n+1)!g(\bar{x}).$$

Отсюда с учетом того, что $F^{(n+1)}(\xi) = 0$ и $\xi = \xi(\bar{x})$, следует, что

$$g(\bar{x}) = \frac{f^{(n+1)}(\xi(\bar{x}))}{(n+1)!}.$$

Поскольку $\bar{x} \in [a, b]$ — произвольно-выбранная точка, то

$$g(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!} \quad \forall x \in [a, b]. \tag{6.17}$$

Формулы (6.15) и (6.17) приводят к выражению

Структура погрешности интерполяционного многочлена Лагранжа ...

Погрешность интерполяционного многочлена Лагранжа

Пусть $f \in C^{(n+1)}([a,b])$. Тогда

$$R_n(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!}\omega_n(x) \quad \forall x \in [a,b]. \tag{6.18}$$

Из (6.18) следует

Оценка погрешности интерполяционного многочлена Лагранжа

Пусть $f \in C^{(n+1)}([a,b])$. Тогда

$$|R_n(x)| \le \frac{M_{n+1}}{(n+1)!}|\omega_n(x)| \quad \forall x \in [a,b], \tag{6.19}$$

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

Многочлен ω_n

Случай равноотстоящих узлов

$x_i \in [a,b]$ $(i = 0,1,2,\ldots,n)$: $x_{i+1} = x_i + h \quad \forall i = 0,1,2,\ldots,n-1$.
Величины $|\omega_n(\bar{x})| = |(\bar{x}-x_0)(\bar{x}-x_1)(\bar{x}-x_2)\ldots(\bar{x}-x_n)|$ тем больше, чем дальше значение $\bar{x}$ от середины $\frac{a+b}{2}$ отрезка $[a,b]$.

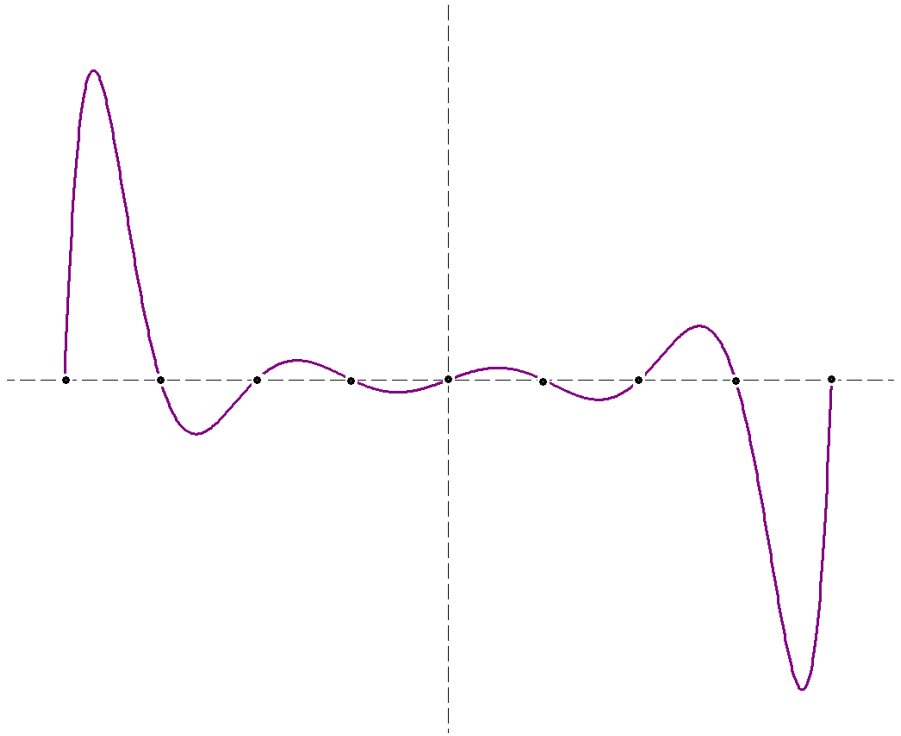


Рис. 10: График $\omega_n(x)$ $x \in [a,b]$, $[a,b] = [-1.6, 1.6]$, $n = 8$.

6.4. Задача минимизации погрешности численной интерполяции

$$\omega_n(x) = (x - x_0)(x - x_1)(x - x_2)\dots(x - x_n) = \\ = x^{n+1} + a_1 x^n + a_2 x^{n-1} + \dots + a_{n-1}x^2 + a_n x + a_{n+1},$$

$$a_i \in \mathbb{R} \quad (i = 1, 2, \dots, n+1)$$

Тогда $\omega_n \in \Upsilon$, где Υ — множество многочленов степени $n+1$, приведенных над полем вещественных чисел.

Постановка задачи

$$p^* = Arg\{ \min_{p \in \Upsilon} V(p) \}, \tag{6.20}$$

где $V(p) = \max_{x \in [a,b]} |p(x)|, \quad p \in \Upsilon.$

$\boxtimes$

▲12 Многочлены Чебышева

Определение 6.3.

Многочленом Чебышева n-ой степени называется

$$T_n(x) = \cos(n \arccos x) \quad x \in [-1, 1],\ n \in \mathbb{N} \cup \{0\} \tag{6.21}$$

Известно, что

$$\cos((n+1)y) + \cos((n-1)y) = 2\cos(y)\cos(ny). \tag{6.22}$$

Тогда равенство (6.22) в обозначениях (6.21) ($y = \arccos x$) может быть записано в виде

$$T_{n+1}(x) + T_{n-1}(x) = 2T_1(x)T_n(x)\ x \in [-1, 1],\ n \in \mathbb{N}$$

Откуда следует рекуррентная формула

$$T_{n+1}(x) = 2T_1(x)T_n(x) - T_{n-1}(x)\ x \in [-1, 1],\ n \in \mathbb{N} \tag{6.23}$$

В силу (6.21) и (6.23)

$$T_0(x) \equiv 1,\ T_1(x) = x,$$
$$T_2(x) = 2x^2 - 1,\ T_3(x) = 2x\left(2x^2 - 1\right) - x = 4x^3 - 3x,$$
$$\ldots\ldots\ldots$$

Экстремальное свойство многочленов Чебышева

Таким образом, $T_n(x) = 2^{n-1}x^n + \ldots \; \forall n \in \mathbb{N}$. Следовательно,

$$p^*_{n+1}(x) = \frac{1}{2^n} T_{n+1}(x) \in \Upsilon, \; x \in [-1, 1] \tag{6.24}$$

Теорема 6.2.

Многочлен $p^*_{n+1}(x)$ $(x \in [-1, 1])$, задаваемый формулой (6.24), имеет $n + 1$ различных корней, принадлежащих отрезку $[-1, 1]$.

Доказательство. $p^*_{n+1}(x) = 0 \;\Leftrightarrow\; T_{n+1}(x) = 0.$

$T_{n+1}(x) = \cos((n+1)\arccos x) = 0 \;\Leftrightarrow$
$(n+1)\arccos x = \frac{\pi}{2} + \pi k, \quad k \in \mathbb{Z} \;\Leftrightarrow\; \arccos x = \frac{(2k+1)\pi}{2(n+1)}, \quad k \in \mathbb{Z} \;\Leftrightarrow$

$$\bar{x}_k = \cos \frac{(2k+1)\pi}{2(n+1)}, \quad k = 0, 1, 2, \ldots, n \tag{6.25}$$

Нетрудно проверить, что для любого целого значения индекса k: $k < 0$ или $k > n$ набор, задаваемый (6.25), уже содержит соответствующую такому значению индекса точку.

□ Теорема доказана.

Экстремальное свойство многочленов Чебышева ...

Теорема 6.3.

Многочлен $p^*_{n+1}(x)$ $(x \in [-1,1])$, задаваемый формулой (6.24), является единственным решением экстремальной задачи (6.20) на отрезке $[-1,1]$.

Оптимальные узлы численной интерполяции на отрезке $[-1,1]$

При решении задачи численной интерполяции (6.1) по $n+1$-ому узлу на отрезке $[-1,1]$ с помощью интерполяционного многочлена Лагранжа (6.12) оптимальными[a] узлами численной интерполяции являются корни $\bar{x}_k$ $(k = 0,1,2,\ldots,n)$ многочлена Чебышева, определяемые формулами (6.25). При этом справедливо неравенство

$$|\omega_n(x)| = |(x-\bar{x}_0)(x-\bar{x}_1)(x-\bar{x}_2)\ldots(x-\bar{x}_n)| \le \frac{1}{2^n} \quad \forall x \in [-1,1] \qquad (6.26)$$

[a] В смысле решения экстремальной задачи (6.20).

Оптимальные узлы численной интерполяции на произвольном отрезке $[a, b] \subset \mathbb{R}$

Пусть задана некоторая биекция $\psi : [a, b] \to [-1, 1]$. Наиболее простой является линейная функция

$$x = \frac{2t - (b + a)}{b - a} \quad t \in [a, b] \tag{6.27}$$

Тогда с помощью биекции (6.27) многочлен Чебышева $n + 1$-ой степени можно определить на отрезке $[a, b]$ следующим образом

$$T_{n+1}(x) = T_{n+1}\left(\frac{2t-(b+a)}{b-a}\right) = 2^n \frac{2^{n+1}}{(b-a)^{n+1}} t^{n+1} + \ldots \quad t \in [a, b] \tag{6.28}$$

Многочлен $p^*_{n+1}(t)$ $(t \in [a, b])$ — решение задачи (6.20) на отрезке $[a, b]$ имеет вид

$$p^*_{n+1}(t) = \omega_n(t) = \frac{T_{n+1}\left(\frac{2t-(b+a)}{b-a}\right)}{\frac{2^{2n+1}}{(b-a)^{n+1}}} \quad t \in [a, b] \tag{6.29}$$

Оптимальные узлы численной интерполяции на произвольном отрезке ...

Таким образом, если в качестве узлов интерполяции на отрезке $[a, b]$ взять точки

$$\bar{t}_k = \frac{(b-a)\bar{x}_k + (b+a)}{2} \quad k = 0, 1, 2, \ldots, n, \tag{6.30}$$

где $\bar{x}_k$ $(k = 0, 1, 2, \ldots, n)$ — корни многочлена Чебышева $T_{n+1}(x)$, определяемые формулами (6.25), то оценка погрешности численной интерполяции на отрезке $[a, b]$ с помощью интерполяционного многочлена Лагранжа задается неравенством

$$|R_n(t)| \leq \frac{(b-a)^{n+1}}{2^{2n+1}} \frac{M_{n+1}}{(n+1)!} \quad \forall t \in [a, b], \tag{6.31}$$

где $M_{n+1} = \max_{t\in[a,b]} |f^{(n+1)}(t)|$.

$$\max_{t\in[a,b]} |\omega_n(t)| = \frac{(b-a)^{n+1}}{2^{2n+1}}. \tag{6.32}$$

6.5. Разделенные разности

Определение 6.4.

Пусть $x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+k} \in [a,b] \subset \mathrm{D}[f]$: $x_{i+j} \neq x_{i+s}\ \forall j \neq s$. Разделенной разностью k-го порядка функции f, построенной по набору узлов $x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+k}$, называется число

$$f(x_i, x_{i+1}, \ldots, x_{i+k-1}, x_{i+k}) = \\ = \frac{f(x_i, x_{i+1}, \ldots, x_{i+k-1}) - f(x_{i+1}, x_{i+2}, \ldots, x_{i+k})}{x_i - x_{i+k}} \tag{6.33}$$

Пусть заданы значения функции f в узлах интерполяции $x_0, x_1, x_2, \ldots, x_n \in [a,b] \subset \mathrm{D}[f]$: $x_j \neq x_s\ \forall j \neq s$. Тогда, в силу (6.33), разделенными разностями 0-го порядка функции f являются

$$f(x_0), f(x_1), f(x_2), \ldots, f(x_n).$$

Примерами разделенных разностей 1-го порядка функции f, построенных по указанному выше набору узлов, являются числа

$$f(x_i, x_{i+1}) = \frac{f(x_i) - f(x_{i+1})}{x_i - x_{i+1}} \quad \forall i = 0, 1, \ldots, n-1,$$

Разделенные разности ...

разделенных разностей 2-го порядка —

$$f(x_i, x_{i+1}, x_{i+2}) = \frac{f(x_i, x_{i+1}) - f(x_{i+1}, x_{i+2})}{x_i - x_{i+2}} \quad \forall i = 0, 1, \ldots, n-2;$$

разделенных разностей 3-го порядка —

$$f(x_i, x_{i+1}, x_{i+2}, x_{i+3}) = \frac{f(x_i, x_{i+1}, x_{i+2}) - f(x_{i+1}, x_{i+2}, x_{i+3})}{x_i - x_{i+3}} \quad \forall i = 0, 1, \ldots, n-3$$

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

Здесь разделенная разность наибольшего, n-го порядка, функции f, построенная по набору узлов $x_0, x_1, x_2, \ldots, x_n$ — это число

$$f(x_0, x_1, \ldots, x_{n-1}, x_n) = \frac{f(x_0, x_1, \ldots, x_{n-1}) - f(x_1, \ldots, x_{n-1}, x_n)}{x_0 - x_n}.$$

Все приведенные выше разделенные разности порядков от 0-го до n-го, построенные для функции f по набору узлов $x_0, x_1, x_2, \ldots, x_n$, формируют треугольник Паскаля.

Свойства разделенных разностей

Теорема 6.4.

Разделенная разность k-го порядка $f\left(x_i, x_{i+1}, \ldots, x_{i+k-1}, x_{i+k}\right)$ функции f, построенная по набору узлов $x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+k}$, представима в виде

$$f\left(x_i, x_{i+1}, \ldots, x_{i+k-1}, x_{i+k}\right) = \\ = \sum_{s=0}^{n} \frac{f(x_{i+s})}{\prod_{j=0, j\neq s}^{n}(x_{i+s}-x_{i+j})} \tag{6.34}$$

Утверждение теоремы можно доказать методом математической индукции по порядку k разделенной разности[25].

Из формулы (6.34) следуют несколько полезных свойств разделенных разностей.

[25] Доказать самостоятельно.

Свойства разделенных разностей

Свойство 1.

При фиксированном наборе узлов $x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+k}$ разделенная разность k-го порядка $f(x_i, x_{i+1}, \ldots, x_{i+k-1}, x_{i+k})$ может быть рассмотрена как отображение, действующее из некоторого пространства функций Φ в множество вещественных чисел $\mathbb{R}$, то есть как функционал.
Тогда такой функционал является линейным и однородным, то есть

$$\begin{aligned}\forall f, g \in \Phi\ \forall \alpha, \beta \in \mathbb{R} \quad & (\alpha f + \beta g)(x_i, x_{i+1}, \ldots, x_{i+k-1}, x_{i+k}) = \\ & = \alpha f(x_i, x_{i+1}, \ldots, x_{i+k-1}, x_{i+k}) + \\ & + \beta g(x_i, x_{i+1}, \ldots, x_{i+k-1}, x_{i+k})\end{aligned} \tag{6.35}$$

Свойство 2.

Разделенная разность $f(x_i, x_{i+1}, \ldots, x_{i+k-1}, x_{i+k})$ не зависит от упорядочения узлов в наборе, по которому она построена.

Свойства разделенных разностей ...

Свойство 3.

Для многочлена $P_n(x)$ степени n разделенная разность n-го порядка является константой, которая не зависит от набора узлов[a].

[a] Следствие. Для многочлена $P_n(x)$ степени n любая разделенная разность $n+1$-го порядка равна нулю.

Доказательство.

Пусть $x_0, x_1, \ldots, x_{n-1}, x_n$ — произвольный набор узлов. Тогда[26]

$$P_n(x) \equiv L_n(x) = \sum_{k=0}^{n} P_n(x_k)\Pi_{i=0, i\neq k}^{n} \frac{x - x_i}{x_k - x_i} = a_0 x^n + \ldots,$$

$$a_0 = \frac{P_n(x_0)}{(x_0-x_1)(x_0-x_2)\ldots(x_0-x_n)} + \frac{P_n(x_1)}{(x_1-x_0)(x_1-x_2)\ldots(x_1-x_n)} + \ldots + \frac{P_n(x_i)}{(x_i-x_0)\ldots(x_i-x_{i-1})(x_i-x_{i+1})\ldots(x_i-x_n)} + \ldots + \frac{P_n(x_n)}{(x_n-x_0)(x_n-x_1)\ldots(x_n-x_{n-1})} =$$
$$= P_n(x_0, x_1, \ldots, x_{n-1}, x_n)$$

□ Свойство доказано.

[26] Через $n+1$-у точку можно провести единственный многочлен степени n.

$(n+1)$-шаговая процедура для произвольной точки $x \in [a,b]$

Пусть $x_0, x_1, \ldots, x_{n-1}, x_n \in [a,b]$: $[a,b] \subset \mathrm{D}[f]$.

0) $f(x,x_0)=\frac{f(x)-f(x_0)}{x-x_0} \Rightarrow$
$f(x)=f(x_0)+f(x,x_0)(x-x_0);$

1) $f(x,x_0,x_1)=\frac{f(x,x_0)-f(x_0,x_1)}{x-x_1} \Rightarrow$
$f(x,x_0)=f(x_0,x_1)+f(x,x_0,x_1)(x-x_1) \Rightarrow$
$f(x)=f(x_0)+f(x_0,x_1)(x-x_0)+f(x,x_0,x_1)(x-x_0)(x-x_1);$
$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$

$n)$ $f(x,x_0,x_1,\ldots,x_{n-1},x_n)=\frac{f(x,x_0,x_1,\ldots,x_{n-1})-f(x_0,x_1,\ldots,x_{n-1},x_n)}{x-x_n} \Rightarrow$
$f(x,x_0,x_1,\ldots,x_{n-1})=f(x_0,x_1,\ldots,x_{n-1},x_n)+$
$+f(x,x_0,x_1,\ldots,x_{n-1},x_n)(x-x_n) \Rightarrow$
$f(x)=f(x_0)+f(x_0,x_1)(x-x_0)+f(x_0,x_1,x_2)(x-x_0)(x-x_1)+\ldots+$
$+f(x_0,x_1,\ldots,x_{n-1},x_n)(x-x_0)(x-x_1)\ldots(x-x_{n-1})+$
$+f(x,x_0,x_1,\ldots,x_{n-1},x_n)(x-x_0)(x-x_1)\ldots(x-x_{n-1})(x-x_n)$
(6.36)

Из (6.36) следует, что

$$f(x) = P_n(x) + R_n(x),$$

где

$$\begin{aligned} P_n(x) = & f(x_0) + \\ & + f(x_0, x_1)(x - x_0) + \\ & + f(x_0, x_1, x_2)(x - x_0)(x - x_1) + \\ & \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & + f(x_0, x_1, \ldots, x_{k-1}, x_k)(x - x_0)(x - x_1)\ldots(x - x_{k-1}) + \\ & + f(x_0, x_1, \ldots, x_k, x_{k+1})(x - x_0)(x - x_1)\ldots(x - x_k) + \\ & \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & + f(x_0, x_1, \ldots, x_{n-1}, x_n)(x - x_0)(x - x_1)\ldots(x - x_{n-1}), \end{aligned} \tag{6.37}$$

$$R_n(x) = f(x, x_0, x_1, \ldots, x_n)(x - x_0)(x - x_1)\ldots(x - x_n). \tag{6.38}$$

Несложно убедиться, что задаваемый формулой (6.37) многочлен n-ой степени решает задачу численной интерполяции, то есть $P_n(x_k) = f(x_k)\ \forall k = 0, 1, 2, \ldots, n$.

Действительно, подстановка точки x_k $(k = 0, 1, 2, \ldots, n)$ в обе части выражения (6.37) приводит к равенству

$$\begin{aligned} P_n(x_k) = & f(x_0) + \\ & + f(x_0, x_1)(x_k - x_0) + \\ & + f(x_0, x_1, x_2)(x_k - x_0)(x_k - x_1) + \\ & \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & + f(x_0, x_1, \ldots, x_{k-1}, x_k)(x_k - x_0)(x - x_1)\ldots(x_k - x_{k-1}). \end{aligned}$$

В силу свойства 2 разделенных разностей $f(x_0, x_1, \ldots, x_{k-1}, x_k) = f(x_k, x_0, x_1, \ldots, x_{k-1})$. Таким образом,

$$\begin{aligned} P_n(x_k) = & f(x_0) + \\ & + f(x_0, x_1)(x_k - x_0) + \\ & + f(x_0, x_1, x_2)(x_k - x_0)(x_k - x_1) + \\ & \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & + f(x_k, x_0, x_1, \ldots, x_{k-1})(x_k - x_0)(x - x_1)\ldots(x_k - x_{k-1}). \end{aligned} \tag{6.39}$$

С другой стороны, $(k-2)$-й и $(k-1)$-й шаги в процедуре (6.36):

$$k-2)\ f(x)=f(x_0)+f(x_0,x_1)(x-x_0)+\ldots+$$
$$+f(x_0,x_1,\ldots,x_{k-3},x_{k-2})(x-x_0)(x-x_1)\ldots(x-x_{k-3})+$$
$$+f(x,x_0,x_1,\ldots,x_{k-2})(x-x_0)(x-x_1)\ldots(x-x_{k-2})$$

$$k-1)\ f(x,x_0,x_1,\ldots,x_{k-2},x_{k-1})=\frac{f(x,x_0,x_1,\ldots,x_{k-2})-f(x_0,x_1,\ldots,x_{k-2},x_{k-1})}{x-x_{k-1}}\Rightarrow$$
$$f(x,x_0,x_1,\ldots,x_{k-2})=f(x_0,x_1,\ldots,x_{k-2},x_{k-1})+$$
$$+f(x,x_0,x_1,\ldots,x_{k-2},x_{k-1})(x-x_{k-1})\Rightarrow$$
$$f(x)=f(x_0)+f(x_0,x_1)(x-x_0)+f(x_0,x_1,x_2)(x-x_0)(x-x_1)+\ldots+$$
$$+f(x_0,x_1,\ldots,x_{k-2},x_{k-1})(x-x_0)(x-x_1)\ldots(x-x_{k-2})+$$
$$+f(x,x_0,x_1,\ldots,x_{k-1})(x-x_0)(x-x_1)\ldots(x-x_{k-1})\Rightarrow$$

$$f(x_k)=f(x_0)+$$
$$+f(x_0,x_1)(x_k-x_0)+$$
$$+f(x_0,x_1,x_2)(x_k-x_0)(x_k-x_1)+ \qquad (6.40)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$+f(x_k,x_0,x_1,\ldots,x_{k-1})(x_k-x_0)(x-x_1)\ldots(x_k-x_{k-1}).$$

Из (6.39) и (6.40) следует, что $P_n(x_k)=f(x_k)$ $(k=0,1,\ldots,n)$.

Интерполяционный многочлен Лагранжа в форме Ньютона

Поскольку через $(n+1)$-у точку можно провести единственный многочлен n-ой степени, то $P_n(x) \equiv L_n(x)$ $x \in \mathbb{R}$.

Определение 6.5.

Интерполяционным многочленом Лагранжа в форме Ньютона, построенным для функции $f(x)$ $(x \in [a,b] \subset \mathrm{D}[f])$ по $(n+1)$-у узлу $x_0, x_1, \ldots, x_n \in [a,b]$: $x_j \neq x_s$ $\forall j \neq s$, называется многочлен степени n

$$
\begin{aligned}
L_n(x) = f(x_0) + \\
&+ f(x_0, x_1)(x - x_0) + \\
&+ f(x_0, x_1, x_2)(x - x_0)(x - x_1) + \\
&\quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&+ f(x_0, x_1, \ldots, x_{k-1}, x_k)(x - x_0)(x - x_1)\ldots(x - x_{k-1}) + \\
&+ f(x_0, x_1, \ldots, x_k, x_{k+1})(x - x_0)(x - x_1)\ldots(x - x_k) + \\
&\quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&+ f(x_0, x_1, \ldots, x_{n-1}, x_n)(x - x_0)(x - x_1)\ldots(x - x_{n-1}).
\end{aligned}
\tag{6.41}
$$

Погрешность интерполяционного многочлена Лагранжа в форме Ньютона

В силу (6.38) погрешность интерполяционного многочлена Лагранжа в форме Ньютона (6.41) представима в виде

$$R_n(x) = f(x, x_0, x_1, \ldots, x_n)\omega_n(x). \tag{6.42}$$

При этом ранее установлено (см. (6.18)), что, если функция $f \in C^{(n+1)}([a,b])$, то $R_n(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!}\omega_n(x)$ $(x \in [a,b])$.
Следовательно, на классе $(n+1)$ раз непрерывно-дифференцируемых на отрезке $[a,b]$ функций

$f \in C^{(n+1)}([a,b])$

$$f(x, x_0, x_1, \ldots, x_n) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!}, \tag{6.43}$$

где $\xi(x) \in [x, x_0, x_1, \ldots, x_n]$.

Свойство разделенных разностей

Обобщением (6.43) является следующее свойство разделенных разностей

Свойство 4

Пусть $f \in C^{(k)}([a,b])$. Тогда для любого набора узлов $x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+k} \in [a,b] \subset \mathrm{D}[f]$: $x_{i+j} \neq x_{i+s}\ \forall j \neq s$

$$f(x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+k}) = \frac{f^{(k)}(\eta)}{k!}, \tag{6.44}$$

где $\eta \in [x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+k}]$.

$\boxtimes$

▲13 ПРИМЕР: $f(1) = -2, f(2) = 3, f(4) = 1$

Интерполяционный многочлен Лагранжа в классической форме

В силу формулы (6.12)

$$L_2(x) = f(1)\frac{(x-2)(x-4)}{(1-2)(1-4)} + f(2)\frac{(x-1)(x-4)}{(2-1)(2-4)} + f(4)\frac{(x-1)(x-2)}{(4-1)(4-2)} =$$

$$= -2\frac{(x-2)(x-4)}{(-1)(-3)} + 3\frac{(x-1)(x-4)}{(1)(-2)} + 1\frac{(x-1)(x-2)}{(3)(2)} =$$

$$= -\frac{2}{3}(x-2)(x-4) - \frac{3}{2}(x-1)(x-4) + \frac{1}{6}(x-1)(x-2) =$$

$$= -\frac{2}{3}(x^2-6x+8) - \frac{3}{2}(x^2-5x+4) + \frac{1}{6}(x^2-3x+2) =$$

$$= -2x^2 + 11x - 11$$

Проверка

Нетрудно убедиться, что $L_2(1) = -2, L_2(2) = 3, L_2(4) = 1$.

ПРИМЕР: $f(1) = -2, f(2) = 3, f(4) = 1$

Интерполяционный многочлен Лагранжа в форме Ньютона (6.41)

$$L_2(x)=f(x_0)+f(x_0,x_1)(x-x_0)+f(x_0,x_1,x_2)(x-x_0)(x-x_1), \quad (6.45)$$

где $x_0 = 1, x_1 = 2, x_2 = 4$

Таблица разделенных разностей (6.33)

$$\begin{bmatrix} x_0=1 & f(x_0)=-2 & & \\ & & f(x_0,x_1)=\ 5 & \\ x_1=2 & f(x_1)=\ 3 & & f(x_0,x_1,x_2)=-2 \\ & & f(x_1,x_2)=-1 & \\ x_2=4 & f(x_2)=\ 1 & & \end{bmatrix}, \quad (6.46)$$

где

$$f(x_0,x_1)=\frac{f(x_0)-f(x_1)}{x_0-x_1}=\frac{-2-3}{1-2}=5,\ f(x_1,x_2)=\frac{f(x_1)-f(x_2)}{x_1-x_2}=\frac{3-1}{2-4}=-1,$$

$$f(x_0,x_1,x_2)=\frac{f(x_0,x_1)-f(x_1,x_2)}{x_0-x_2}=\frac{5-(-1)}{1-4}=-2.$$

Результат подстановки (6.46) в (6.45)

$$L_2(x)=-2+5(x-1)-2(x-1)(x-2)=-2x^2+11x-11$$

ПРИМЕР: $f(1) = -2, f(2) = 3, f(4) = 1$

Оценка погрешности

Пусть $f \in C^{(3)}([1,4])$ и $M_3 = \max_{x\in[1,4]} |f^{(3)}(x)|$. Тогда в силу (6.19)

$$|R_2(x)| \le \frac{M_3}{6}|(x-1)(x-2)(x-4)| \quad \forall x \in [1,4]$$

Таблица разделенных разностей (треугольник Паскаля)

$$\begin{bmatrix}
x_0 & \underline{f(x_0)} & & & & \\
 & & \underline{f(x_0,x_1)} & & & \\
x_1 & f(x_1) & & \underline{f(x_0,x_1,x_2)} & & \\
 & & f(x_1,x_2) & & \cdots & \\
x_2 & f(x_2) & & f(x_1,x_2,x_3) & & \\
 & & f(x_2,x_3) & & & \underline{f(x_0,x_1,\ldots,x_n)} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \\
x_{n-1} & f(x_{n-1}) & & f(x_{n-2},x_{n-1},x_n) & & \\
 & & f(x_{n-1},x_n) & & & \\
x_n & f(x_n) & & & &
\end{bmatrix} \tag{6.47}$$

Использование таблицы разделенных разностей для построения интерполяционного многочлена Лагранжа в форме Ньютона

$$
\begin{aligned}
L_n(x) = & \underline{f(x_0)} + \\
& + \underline{f(x_0, x_1)}(x - x_0) + \\
& + \underline{f(x_0, x_1, x_2)}(x - x_0)(x - x_1) + \\
& \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
& + \underline{f(x_0, x_1, \ldots, x_{n-1}, x_n)}(x - x_0)(x - x_1)\ldots(x - x_{n-1}).
\end{aligned} \tag{6.48}
$$

Практическое использование интерполяционного многочлена Лагранжа в форме Ньютона (см. (6.48)) позволяет в значительной степени преодолеть основной алгоритмический недостаток его классической формы (см. (6.12)). Например, добавление в набор данных $(x_i, f(x_i))$ $(i = 0, 1, 2, \ldots, n)$ очередной точки $(x_{n+1}, f(x_{n+1}))$ приводит только к вычислению в таблице (6.47) дополнительных $(n+1)$-ой разделенной разности и появлению в правой части равенства (6.48) еще одного слагаемого $\underline{f(x_0, x_1, \ldots, x_{n-1}, x_n, x_{n+1})}(x - x_0)(x - x_1)\ldots(x - x_{n-1})(x - x_n)$.

6.6. Разделенные разности с кратными узлами

Определение 6.6.

Пусть задан набор узлов $x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+k} \in [a,b] \subset \mathrm{D}[f]$: среди x_{i+j} $(j = 0, 1, \ldots, k)$ не все узлы различные.
Разделенной разностью k-го порядка функции f, построенной по набору узлов $x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+k}$, называется число

$$f\left(x_i, x_{i+1}, \ldots, x_{i+k-1}, x_{i+k}\right) = \\ = \lim_{\varepsilon \to 0} f\left(x_i^\varepsilon, x_{i+1}^\varepsilon, \ldots, x_{i+k-1}^\varepsilon, x_{i+k}^\varepsilon\right), \tag{6.49}$$

где $\forall \varepsilon > 0$ $x_i^\varepsilon, x_{i+1}^\varepsilon, \ldots, x_{i+k-1}^\varepsilon, x_{i+k}^\varepsilon \in [a,b]$:

$$x_{i+j}^\varepsilon \neq x_{i+s}^\varepsilon \ \forall j \neq s;$$
$$\lim_{\varepsilon \to 0} x_{i+j}^\varepsilon = x_{i+j} \quad \forall j = 0, 1, \ldots, k.$$

И при этом предел в правой части в (6.49) не зависит от способа задания узлов x_{i+j}^ε $(j = 0, 1, \ldots, k)$.

Свойства разделенных разностей с кратными узлами

Пусть $f \in C^{(k)}([a,b])$, $[a,b] \subset \mathrm{D}[f]$.

Свойство 1

Для любого набора узлов $x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+k} \in [a,b]$

$$f(x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+k}) = \frac{f^{(k)}(\eta)}{k!}, \tag{6.50}$$

где $\eta \in [x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+k}]$.

Свойство 2

При фиксированном наборе узлов $x_i, x_{i+1}, x_{i+2}, \ldots, x_{i+k}$ разделенная разность k-го порядка $f(x_i, x_{i+1}, \ldots, x_{i+k-1}, x_{i+k})$ является линейным и однородным функционалом.

Свойства разделенных разностей ...

Свойство 3

Разделенная разность $f(x_i, x_{i+1}, \ldots, x_{i+k-1}, x_{i+k})$ не зависит от упорядочения узлов в наборе, по которому она построена.

Свойство 4. Непрерывность.

Разделенная разность $f(x_i, x_{i+1}, \ldots, x_{i+k-1}, x_{i+k})$ является непрерывной по совокупности переменных функцией $k+1$-х переменных.

Свойство 5. Дифференцируемость.

Пусть $f \in C^{(k+2)}([a, b])$. Тогда

$$\begin{aligned} \frac{d}{dx} f(x, x_i, x_{i+1}, \ldots, x_{i+k}) = \\ = f(x, x, x_i, x_{i+1}, \ldots, x_{i+k}). \end{aligned} \tag{6.51}$$

Свойства разделенных разностей ...

Доказательство.

$$\frac{d}{dx} f(x, x_i, x_{i+1}, \ldots, x_{i+k}) =$$

$$= \lim_{\widetilde{x} \to x} \frac{f(\widetilde{x}, x_i, x_{i+1}, \ldots, x_{i+k}) - f(x, x_i, x_{i+1}, \ldots, x_{i+k})}{\widetilde{x} - x} =\Big|_{(\text{св.3})}$$

$$= \lim_{\widetilde{x} \to x} \frac{f(\widetilde{x}, x_i, x_{i+1}, \ldots, x_{i+k}) - f(x_i, x_{i+1}, \ldots, x_{i+k}, x)}{\widetilde{x} - x} =\Big|_{(\text{св.1,опр.})}$$

$$= \lim_{\widetilde{x} \to x} f(\widetilde{x}, x_i, x_{i+1}, \ldots, x_{i+k}, x) =\Big|_{(\text{св.4})}$$

$$= f(x, x_i, x_{i+1}, \ldots, x_{i+k}, x) =\Big|_{(\text{св.3})}$$

$$= f(x, x, x_i, x_{i+1}, \ldots, x_{i+k})$$

□ Свойство доказано.

Следствие. Пусть $f \in C^{(k+3)}([a, b])$. Тогда

$$\frac{d^2}{dx^2} f(x, x_i, x_{i+1}, \ldots, x_{i+k}) = \\ = 2f(x, x, x, x_i, x_{i+1}, \ldots, x_{i+k}). \tag{6.52}$$

6.7. Интерполяция с кратными узлами

Постановка задачи

Пусть $f \in C^{(n+1)}([a,b])$, $[a,b] \subset \mathrm{D}[f]$ $(n = m_1 + m_2 + \ldots + m_s - 1)$ и $x_1, x_2, \ldots, x_s \in [a,b]$: $x_i \neq x_j \ \forall i \neq j$.

$$\begin{array}{l} f(x_1), f'(x_1), f''(x_1), \ldots, f^{(m_1-1)}(x_1); \\ f(x_2), f'(x_2), f''(x_2), \ldots, f^{(m_2-1)}(x_2); \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ f(x_s), f'(x_s), f''(x_s), \ldots, f^{(m_s-1)}(x_s). \end{array} \tag{6.53}$$

Требуется построить многочлен $L_n(x)$ степени n, который удовлетворяет следующим условиям:

$$\begin{array}{l} L_n(x_1)=f(x_1), L_n'(x_1)=f'(x_1), \ldots, L_n^{(m_1-1)}(x_1)=f^{(m_1-1)}(x_1); \\ L_n(x_2)=f(x_2), L_n'(x_2)=f'(x_2), \ldots, L_n^{(m_2-1)}(x_2)=f^{(m_2-1)}(x_2); \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ L_n(x_s)=f(x_s), L_n'(x_s)=f'(x_s), \ldots, L_s^{(m_s-1)}(x_s)=f^{(m_s-1)}(x_s). \end{array} \tag{6.54}$$

Существование и единственность решения задачи численной интерполяции с кратными узлами

Теорема 6.5.

Существует единственный многочлен $L_n(x)$ степени n, который является решением задачи (6.53),(6.54).

Доказательство.

Единственность

От противного. Пусть существует два различных многочлена $P_n(x)$ и $L_n(x)$ $(P_n(x) \neq L_n(x))$ степени n, которые являются решениями задачи (6.54). Тогда можно построить вспомогательный многочлен $S_n(x) = P_n(x) - L_n(x)$, для которого узлы $x_1, x_2, \ldots, x_s$ являются корнями кратностей $m_1, m_2, \ldots, m_s$ соответственно. Таким образом, многочлен $S_n(x)$ степени n имеет $n+1$ корней с учетом их кратностей $(n+1 = m_1 + m_2 + \ldots + m_s)$. Возникает противоречие.
Следовательно, $S_n(x) \equiv 0$, то есть $P_n(x) \equiv L_n(x)$.

Существование и единственность решения задачи ...

Существование

Для построения многочлена $L_n(x)$ степени n, который является решением задачи (6.54), предлагается следующая процедура. Для каждого из узлов x_i $(i = 1, 2, \ldots, s)$ строится однопараметрическое семейство наборов различных узлов x_{ij}^{ε} $(j = 1, 2, \ldots, m_i,\ \varepsilon \in \mathbb{R})$:

$$\begin{array}{l} x_{ij}^{\varepsilon} = x_i \pm \varepsilon(j-1) \\ \forall i = 1, 2, \ldots, s \quad \forall j = 1, 2, \ldots, m_i \end{array} \tag{6.55}$$

где

$$0 < \varepsilon < \bar{\varepsilon},\ \bar{\varepsilon} = \frac{1}{2\bar{m}} \min_{i,j=1,2,\ldots,s: i \neq j} |x_i - x_j|,\ \bar{m} = \max_{i=1,2,\ldots,s} m_i. \tag{6.56}$$

Очевидно, что в силу (6.56)

$$x_{ij}^{\varepsilon} \in [a, b] \quad \forall 0 < \varepsilon < \bar{\varepsilon} \quad \forall i = 1, 2, \ldots, s \quad \forall j = 1, 2, \ldots, m_i$$

при подходящем выборе знака в правой части выражения (6.55) и

$$\lim_{\varepsilon \to 0} x_{ij}^{\varepsilon} = x_i \quad \forall i = 1, 2, \ldots, s \quad \forall j = 1, 2, \ldots, m_i.$$

Существование и единственность решения задачи ...

Считая узлы x_{ij}^{ε} $(i = 1, 2, \ldots, s\ j = 1, 2, \ldots, m_i)$ однократными, для функции f для любого $0 < \varepsilon < \bar{\varepsilon}$ по формуле (6.41) можно построить интерполяционный многочлен Лагранжа $L_n^{\varepsilon}(x)$ степени $n = m_1 + m_2 + \ldots + m_s - 1$:

$$
\begin{aligned}
L_n^{\varepsilon}(x) = & f(x_{11}^{\varepsilon}) + f(x_{11}^{\varepsilon}, x_{12}^{\varepsilon})(x - x_{11}^{\varepsilon}) + \ldots + \\
& + f(x_{11}^{\varepsilon}, \ldots, x_{1m_1}^{\varepsilon}) \Pi_{j=1}^{m_1 - 1}(x - x_{1j}^{\varepsilon}) + \\
& + f(x_{11}^{\varepsilon}, \ldots, x_{1m_1}^{\varepsilon}, x_{21}^{\varepsilon}) \Pi_{j=1}^{m_1}(x - x_{1j}^{\varepsilon}) + \\
& + f(x_{11}^{\varepsilon}, \ldots, x_{1m_1}^{\varepsilon}, x_{21}^{\varepsilon}, x_{22}^{\varepsilon}) \left(\Pi_{j=1}^{m_1}(x - x_{1j}^{\varepsilon}) \right) (x - x_{21}^{\varepsilon}) + \ldots + \\
& + f(x_{11}^{\varepsilon}, \ldots, x_{1m_1}^{\varepsilon}, x_{21}^{\varepsilon}, \ldots, x_{2m_2}^{\varepsilon}) \Pi_{j=1}^{m_1}(x - x_{1j}^{\varepsilon}) \Pi_{j=1}^{m_2 - 1}(x - x_{2j}^{\varepsilon}) + \\
& \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
& + f(x_{11}^{\varepsilon}, \ldots, x_{21}^{\varepsilon}, \ldots, x_{s1}^{\varepsilon}, \ldots, x_{sm_s}^{\varepsilon}) \Pi_{j=1}^{m_1}(x - x_{1j}^{\varepsilon}) \ldots \Pi_{j=1}^{m_s - 1}(x - x_{sj}^{\varepsilon}).
\end{aligned}
\tag{6.57}
$$

Существование и единственность решения задачи ...

Переход в обеих частях равенства (6.57) к пределу при $\varepsilon \to 0$, учитывая свойства разделенных разностей с кратными узлами для достаточно гладкой функции f, приводит к выражению

$$
\begin{aligned}
L_n(x) = {} & f(x_1) + f(x_1, x_1)(x - x_1) + \ldots + f(\underbrace{x_1, \ldots, x_1}_{m1})(x - x_1)^{m_1 - 1} + \\
& + f(\underbrace{x_1, \ldots, x_1}_{m1}, x_2)(x - x_1)^{m_1} + \\
& + f(\underbrace{x_1, \ldots, x_1}_{m1}, x_2, x_2)(x - x_1)^{m_1}(x - x_2) + \ldots + \\
& + f(\underbrace{x_1, \ldots, x_1}_{m1}, \underbrace{x_2, \ldots, x_2}_{m2})(x - x_1)^{m_1}(x - x_2)^{m_2 - 1} + \\
& \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
& + f(\underbrace{x_1, \ldots, x_1}_{m1}, \underbrace{x_2, \ldots, x_2}_{m2}, \ldots, \underbrace{x_s, \ldots, x_s}_{ms}) \cdot \\
& \cdot (x - x_1)^{m_1} \ldots (x - x_{s-1})^{m_{s-1}} (x - x_s)^{m_s - 1}.
\end{aligned}
\tag{6.58}
$$

Существование и единственность решения задачи ...

Многочлен $L_n(x)$, задаваемый выражением (6.58), решает задачу численной интерполяции (6.53), (6.54). Действительно, с учетом гладкости функции f и свойства 1 разделенных разностей с кратными узлами, подстановка x_1 в правую часть выражения (6.58) и ее производные до (m_1-1)-го порядка приводит к

$$
\begin{aligned}
&L_n(x_1)=f(x_1),\\
&L_n{}'(x_1)=f(x_1,x_1)=f'(x_1),\\
&L_n{}''(x_1)=2!f(x_1,x_1,x_1)=2!\frac{f''(x_1)}{2!}=f''(x_1),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&L_n{}^{(m_1-1)}(x_1)=(m_1-1)!f(\underbrace{x_1,\ldots,x_1}_{m1})=(m_1-1)!\frac{f^{(m_1-1)}(x_1)}{(m_1-1)!}=f^{(m_1-1)}(x_1).
\end{aligned}
\tag{6.59}
$$

Аналогичные (6.59) равенства можно получить и для любого из узлов x_k $(k=2,3,\ldots,s)$ с учетом их кратности. Для этого достаточно узлы x_1 и x_k поменять местами и повторить приведенные выше рассуждения.

□ Теорема доказана.

Интерполяционный многочлен Эрмита

Определение 6.7.

Пусть $f \in C^{(n+1)}([a,b])$ $(n = m_1 + m_2 + \ldots + m_s - 1)$. Интерполяционным многочленом Эрмита, построенным по набору данных (6.53), называется

$$
\begin{aligned}
L_n(x) = {} & f(x_1) + f'(x_1)(x-x_1) + \tfrac{1}{2} f''(x_1)(x-x_1)^2 + \ldots + \\
& + \tfrac{1}{(m_1-1)!} f^{(m_1-1)}(x_1)(x-x_1)^{m_1-1} + \\
& + f(\underbrace{x_1, \ldots, x_1}_{m1}, x_2)(x-x_1)^{m_1} + \\
& + f(\underbrace{x_1, \ldots, x_1}_{m1}, x_2, x_2)(x-x_1)^{m_1}(x-x_2) + \ldots + \\
& + f(\underbrace{x_1, \ldots, x_1}_{m1}, \underbrace{x_2, \ldots, x_2}_{m2})(x-x_1)^{m_1}(x-x_2)^{m_2-1} + \\
& \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
& + f(\underbrace{x_1, \ldots, x_1}_{m1}, \underbrace{x_2, \ldots, x_2}_{m2}, \ldots, \underbrace{x_s, \ldots, x_s}_{ms}) \cdot \\
& \quad \cdot (x-x_1)^{m_1} \ldots (x-x_{s-1})^{m_{s-1}} (x-x_s)^{m_s-1}.
\end{aligned}
\tag{6.60}
$$

Погрешность интерполяционного многочлена Эрмита

Погрешность интерполяционного многочлена Эрмита

$$R_n(x) = f(x, \underbrace{x_1, \ldots, x_1}_{m1}, \underbrace{x_2, \ldots, x_2}_{m2}, \ldots, \underbrace{x_s, \ldots, x_s}_{ms})\omega_n(x), \tag{6.61}$$

$$\omega_n(x) = (x-x_1)^{m_1}(x-x_2)^{m_2}\ldots(x-x_s)^{m_s}, \tag{6.62}$$

где $n = m_1 + m_2 + \ldots + m_s - 1$.

Оценка погрешности интерполяционного многочлена Эрмита

Пусть $f \in C^{(n+1)}([a,b])$. Тогда

$$R_n(x) = \frac{f^{(n+1)}(\xi(x))}{(n+1)!}\omega_n(x), \tag{6.63}$$

$$|R_n(x)| \le \frac{M_{n+1}}{(n+1)!}|\omega_n(x)| \quad \forall x \in [a,b], \tag{6.64}$$

где $\xi(x) \in [x, x_1, x_2, \ldots, x_s]$, $M_{n+1} = \max_{x\in[a,b]} |f^{(n+1)}(x)|$.

ПРИМЕР

Пусть $x_1, x_2 \in [a,b] \subset \mathrm{D}[f]$: $f(x_1), f'(x_1), f''(x_1)$; $f(x_2), f'(x_2)$.

$$
\begin{bmatrix}
x_1 & \underline{f(x_1)} & & & & & \\
 & & \underline{f(x_1,x_1)} & & & & \\
x_1 & f(x_1) & & \underline{f(x_1,x_1,x_1)} & & & \\
 & & f(x_1,x_1) & & \underline{f(x_1,x_1,x_1,x_2)} & & \\
x_1 & f(x_1) & & f(x_1,x_1,x_2) & & \underline{f(x_1,x_1,x_1,x_2,x_2)} & \\
 & & f(x_1,x_2) & & f(x_1,x_1,x_2,x_2) & & \\
x_2 & f(x_2) & & f(x_1,x_2,x_2) & & & \\
 & & f(x_2,x_2) & & & & \\
x_2 & f(x_2) & & & & &
\end{bmatrix}
$$

Здесь $f(x_1,x_1)=f'(x_1)$, $f(x_1,x_1,x_1)=\frac{f''(x_1)}{2}$, $f(x_2,x_2)=f'(x_2)$.

$$
L_4(x)=f(x_1)+f'(x_1)(x-x_1)+\frac{f''(x_1)}{2}(x-x_1)^2+
$$

$$
+f(x_1,x_1,x_1,x_2)(x-x_1)^3+f(x_1,x_1,x_1,x_2,x_2)(x-x_1)^3(x-x_2)
$$

6.8. Сходимость интерполяционного процесса

Определение 6.8.

Под интерполяционным процессом понимают бесконечную функциональную последовательность, элементами которой являются интерполяционные многочлены Лагранжа $L_n(x)$, каждый из которых построен по своему набору узлов $x_0^{(n)}, x_1^{(n)}, \ldots, x_n^{(n)} \in [a,b] \subset \mathrm{D}[f]$, где $n \to \infty$.

Пусть $f \in C([a,b])$. Тогда можно задать такую систему узлов $\{x_0^{(n)}, x_1^{(n)}, \ldots, x_n^{(n)}\}_{n=0}^{\infty}$, что интерполяционный процесс будет равномерно на отрезке $[a,b]$ сходиться к функции f[a].

[a] Например, для построения такой системы узлов могут использоваться корни многочленов Чебышева различных степеней.

Для любой системы узлов $\{x_0^{(n)}, x_1^{(n)}, \ldots, x_n^{(n)}\}_{n=0}^{\infty}$ можно на отрезке $[a,b]$ задать такую функцию f, что интерполяционный процесс к этой функции f сходиться не будет.

$\boxtimes$

▲14 6.9. Приближенное решение нелинейного уравнения с помощью интерполяционного многочлена Лагранжа

Постановка задачи

$f \in C\left(\mathrm{D}[f]\right), \mathrm{D}[f] \subseteq \mathbb{R}$

$$f(x) = 0 \tag{6.65}$$

$$\xi \in \mathrm{D}[f]: \quad f(\xi) = 0 \tag{6.66}$$

Корень уравнения (6.65) $\xi \in [a, b] \subseteq \mathrm{D}[f]$.

Пусть функция f обратима на отрезке $[a, b]$ и задан набор данных

$$\begin{array}{ll} x_0, x_1, x_2, \ldots, x_n \in [a, b]: & x_i \neq x_j \ \forall i \neq j \\ y_0, y_1, y_2, \ldots, y_n: & y_i = f(x_i) \ \forall i = 0, 1, \ldots, n \end{array} \tag{6.67}$$

По набору данных (6.67) для функции f^{-1} можно построить интерполяционный многочлен Лагранжа $L_n(y)$. Тогда

$$\xi \approx L_n(0) \tag{6.68}$$

ТЕМА 7. Численное дифференцирование

7.1. Постановка задачи

$f \in C^{(1)}\left(\mathrm{D}[f]\right), \mathrm{D}[f] \subseteq \mathbb{R}$

$$\begin{aligned} &x_0, x_1, x_2, \ldots, x_n \in [a,b] \subseteq \mathrm{D}[f]: \quad x_i \neq x_j \ \forall i \neq j \\ &y_0, y_1, y_2, \ldots, y_n: \quad y_i = f(x_i) \ \forall i = 0, 1, \ldots, n \end{aligned} \tag{7.1}$$

Требуется, используя набор данных (7.1), определить приближенное значение производной функции f в заданной точке из области ее определения.

7.2. Алгоритм решения (операция численного дифференцирования)

$$\begin{aligned} &f(x) = L_n(x) + R_n(x), \\ &f'(x) = L_n{}'(x) + R_n{}'(x), \\ &f'(x) \approx L_n{}'(x) \end{aligned} \tag{7.2}$$

Численное дифференцирование ...

7.3. Погрешность численного дифференцирования

$$R_n{}'(x) = \frac{d}{dx}\left(f(x, x_0, x_1, \ldots, x_n)\omega_n(x)\right) \tag{7.3}$$

Из того, что значение $R_n(x)$ мало не следует, что значение $R_n{}'(x)$ также мало.

ПРИМЕР

Пусть

$$\varphi(x) = \frac{1}{N}\sin(N^2 x),$$

где $N \gg 1$.

$$\varphi'(x) = N\cos(N^2 x).$$

Этот пример свидетельствует о неустойчивости операции численного дифференцирования.

7.4. Построение формул численного дифференцирования

$$f(x) = L_n(x) + f(x, x_0, x_1, \ldots, x_n)\omega_n(x)$$

Пусть $f \in C^{(n+2)}(\mathrm{D}[f])$. Тогда

$$\begin{aligned}
R_n{}'(x) &= \tfrac{d}{dx}\left(f(x, x_0, x_1, \ldots, x_n)\omega_n(x)\right) = \\
&= \left(\tfrac{d}{dx} f(x, x_0, x_1, \ldots, x_n)\right)\omega_n(x) + \\
&\quad + f(x, x_0, x_1, \ldots, x_n)\tfrac{d}{dx}\omega_n(x) = \\
&= f(x, x, x_0, x_1, \ldots, x_n)\omega_n(x) + \\
&\quad + f(x, x_0, x_1, \ldots, x_n)\tfrac{d}{dx}\omega_n(x) = \\
&= \frac{f^{(n+2)}(\xi_1(x))}{(n+2)!}\omega_n(x) + \frac{f^{(n+1)}(\xi_2(x))}{(n+1)!}\frac{d}{dx}\omega_n(x),
\end{aligned} \tag{7.4}$$

где $\xi_1(x), \xi_2(x) \in [x, x_0, x_1, \ldots, x_n]$.

Построение формул численного дифференцирования

Пусть $f \in C^{(n+3)}(\mathrm{D}[f])$. Тогда

$$
\begin{aligned}
R_n''(x) &= \frac{d^2}{dx^2}\left(f(x, x_0, x_1, \ldots, x_n)\omega_n(x)\right) = \\
&= \frac{d}{dx}\left[\left(\frac{d}{dx} f(x, x_0, x_1, \ldots, x_n)\right)\omega_n(x) + \right. \\
&\quad \left. + f(x, x_0, x_1, \ldots, x_n)\frac{d}{dx}\omega_n(x)\right] = \\
&= \frac{d}{dx}\left[f(x, x, x_0, x_1, \ldots, x_n)\omega_n(x) + \right. \\
&\quad \left. + f(x, x_0, x_1, \ldots, x_n)\frac{d}{dx}\omega_n(x)\right] = \\
&= f(x, x, x, x_0, x_1, \ldots, x_n)\omega_n(x) + \\
&\quad + 2f(x, x, x_0, x_1, \ldots, x_n)\frac{d}{dx}\omega_n(x) + \\
&\quad + f(x, x_0, x_1, \ldots, x_n)\frac{d^2}{dx^2}\omega_n(x) = \\
&= \frac{f^{(n+3)}(\eta_1(x))}{(n+3)!}\omega_n(x) + 2\frac{f^{(n+2)}(\eta_2(x))}{(n+2)!}\frac{d}{dx}\omega_n(x) + \frac{f^{(n+1)}(\eta_3(x))}{(n+1)!}\frac{d^2}{dx^2}\omega_n(x),
\end{aligned}
\tag{7.5}
$$

где $\eta_1(x), \eta_2(x), \eta_3(x) \in [x, x_0, x_1, \ldots, x_n]$.

Оценки погрешностей формул численного дифференцирования

Пусть $f \in C^{(n+2)}\,(\mathrm{D}[f])$. Тогда

$$|R_n{}'(x)| \le \frac{M_{n+2}}{(n+2)!}|\omega_n(x)| + \frac{M_{n+1}}{(n+1)!}|\frac{d}{dx}\omega_n(x)|. \qquad (7.6)$$

Пусть $f \in C^{(n+3)}\,(\mathrm{D}[f])$. Тогда

$$|R_n{}''(x)| \le \frac{M_{n+3}}{(n+3)!}|\omega_n(x)| + 2\frac{M_{n+2}}{(n+2)!}|\frac{d}{dx}\omega_n(x)| + \frac{M_{n+1}}{(n+1)!}|\frac{d^2}{dx^2}\omega_n(x)|. \qquad (7.7)$$

Здесь

$$x \in [x_0, x_1, \ldots, x_n],$$
$$M_k = \max\nolimits_{x\in[x_0,x_1,\ldots,x_n]} |f^{(k)}(x)| \quad (k = n+1, n+2, n+3)$$

Формула численного дифференцирования по двум узлам

Исходные данные

$f \in C^{(3)}([a,b]), [a,b] \subseteq \mathrm{D}[f]$

$$\begin{array}{l} x_0, x_1 \in [a,b]: \quad x_1 = x_0 + h, \ h > 0 \\ y_0, y_1: \quad y_i = f(x_i) \ \forall i = 0,1 \end{array} \tag{7.8}$$

$$f(x) = f(x_0) + f(x_0, x_1)(x - x_0) + f(x, x_0, x_1)\omega_1(x)$$

Формула численного дифференцирования

$$f'(x) \approx f(x_0, x_1) = \frac{f(x_0) - f(x_1)}{x_0 - x_1} \tag{7.9}$$

Погрешность формулы численного дифференцирования по двум узлам

$$
\begin{aligned}
R_1'(x) &= \frac{d}{dx}\left(f(x, x_0, x_1)\omega_1(x)\right) = \\
&= f(x, x, x_0, x_1)\omega_1(x) + f(x, x_0, x_1)\frac{d}{dx}\omega_1(x) = \\
&= \frac{f^{(3)}(\xi_1(x))}{6}\omega_1(x) + \frac{f^{(2)}(\xi_2(x))}{2}\frac{d}{dx}\omega_1(x),
\end{aligned}
\tag{7.10}
$$

где $\xi_1(x), \xi_2(x) \in [x, x_0, x_1]$.

$$
\omega_1(x) = (x - x_0)(x - x_1),
$$

$$
\frac{d}{dx}\omega_1(x) = 2x - (x_0 + x_1)
$$

$$
\begin{aligned}
&\max_{x\in[x_0,x_1]} |\omega_1(x)| = |\omega_1(\tfrac{x_0+x_1}{2})| = \frac{h^2}{4}, \\
&\max_{x\in[x_0,x_1]} |\frac{d}{dx}\omega_1(x)| = |\frac{d}{dx}\omega_1(x_i)| = h, \ \ i = 0, 1.
\end{aligned}
\tag{7.11}
$$

Частные случаи погрешности формулы численного дифференцирования по двум узлам

Случай I. Дифференцирование на середину

Пусть $x = \frac{x_0+x_1}{2}$. Тогда, в силу (7.10) и (7.11)

$$|R_1{}'(x)| = |\frac{f^{(3)}(\xi_1(x))}{6}||\omega_1(x)| \leq \frac{M_3}{24}h^2, \qquad (7.12)$$

где $M_3 = \max_{x\in[x_0,x_1]} |f^{(3)}(x)|$.

Случай II. Дифференцирование на край

Пусть $x = x_i,\ i = 0, 1$. Тогда, в силу (7.10) и (7.11)

$$|R_1{}'(x)| = |\frac{f^{(2)}(\xi_2(x))}{2}||\frac{d}{dx}\omega_1(x)| \leq \frac{M_2}{2}h, \qquad (7.13)$$

где $M_2 = \max_{x\in[x_0,x_1]} |f^{(2)}(x)|$.

Неустранимая погрешность формулы численного дифференцирования по двум узлам

Пусть $f(x_i) \approx f_i^\star$: $A_{f_i^\star} = \varepsilon > 0$ $(i = 0, 1)$. Тогда

$$|f(x_0, x_1) - f^\star(x_0, x_1)| = \left|\frac{f(x_0) - f(x_1)}{x_0 - x_1} - \frac{f_0^\star - f_1^\star}{x_0 - x_1}\right| =$$

$$= \left|\frac{f(x_0) - f_0^\star}{x_0 - x_1} - \frac{f(x_1) - f_1^\star}{x_0 - x_1}\right| \leq$$

$$\leq \frac{|f(x_0) - f_0^\star|}{h} + \frac{|f(x_1) - f_1^\star|}{h} \leq$$

$$\leq \frac{\varepsilon}{h} + \frac{\varepsilon}{h} = \frac{2\varepsilon}{h}$$

Неустранимая погрешность

$$A_{\text{н}}(h) = \frac{2\varepsilon}{h} \tag{7.14}$$

Задача определения оптимального шага численного дифференцирования по двум узлам

$$|f'(x)-f^\star(x_0,x_1)|=|f'(x)-f(x_0,x_1)+f(x_0,x_1)-f^\star(x_0,x_1)|\leq$$

$$\leq|f'(x)-f(x_0,x_1)|+|f(x_0,x_1)-f^\star(x_0,x_1)|\leq$$

$$\leq A_{\text{М}}(h)+A_{\text{Н}}(h)$$

Полная погрешность

$$A_{\text{П}}(h) = A_{\text{М}}(h)+A_{\text{Н}}(h) \tag{7.15}$$

Постановка задачи

$$A_{\text{П}}(h) \to \min_{h\geq 0} \tag{7.16}$$

ПРИМЕР: Оптимальный шаг численного дифференцирования на середину

Пусть $f \in C^{(3)}([x_0, x_1])$: $M_3 = \max_{x \in [x_0, x_1]} |f^{(3)}(x)| > 0$, $x = \frac{x_0 + x_1}{2}$.
Здесь $A_{\text{М}}(h) = \frac{M_3}{24} h^2$.

Постановка задачи

$$A_{\text{П}}(h) = \frac{M_3}{24} h^2 + \frac{2\varepsilon}{h} \to \min_{h \geq 0} \tag{7.17}$$

$A_{\text{П}}(h)$ — строго-выпуклая на $(0, +\infty)$ функция.

$$A'_{\text{П}}(h^\star) = \frac{M_3}{12} h^\star - \frac{2\varepsilon}{(h^\star)^2} = 0$$

Оптимальный шаг

$$h^\star = \left(\frac{24\varepsilon}{M_3} \right)^{\frac{1}{3}} \tag{7.18}$$

$\boxtimes$

▲15. ТЕМА 8. Численное интегрирование

8.1. Постановка задачи

Пусть функция f ограничена на отрезке $[a, b] \subseteq \mathrm{D}[f]$.

$$\begin{aligned} &x_0, x_1, x_2, \ldots, x_n \in [a, b]: \quad x_i \neq x_j \ \forall i \neq j \\ &y_0, y_1, y_2, \ldots, y_n: \quad y_i = f(x_i) \ \forall i = 0, 1, \ldots, n \end{aligned} \tag{8.1}$$

$$I[f] = \int_a^b f(x)dx$$

8.2. Алгоритм решения (операция численного интегрирования)

$$\begin{aligned} f(x) &= L_n(x) + R_n(x), \\ I[f] &\approx \int_a^b L_n(x)dx \end{aligned} \tag{8.2}$$

8.3. Погрешность численного интегрирования

$$R_n[f] = \int_a^b R_n(x)dx = \int_a^b f(x, x_0, x_1, \ldots, x_n)\omega_n(x)dx, \tag{8.3}$$

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

8.4. Квадратурные формулы численного интегрирования

$$\int_a^b L_n(x)dx = \int_a^b \sum_{k=0}^n f(x_k)\Pi_{i=0,i\neq k}^n \frac{x-x_i}{x_k-x_i}dx =$$

$$= \sum_{k=0}^n f(x_k) \int_a^b \Pi_{i=0,i\neq k}^n \frac{x-x_i}{x_k-x_i}dx =$$

$$= \sum_{k=0}^n A_k f(x_k)$$

$$A_k = \int_a^b \Pi_{i=0,i\neq k}^n \frac{x-x_i}{x_k-x_i}dx, \quad k = 0,1,2,\ldots,n \tag{8.4}$$

Квадратурная сумма

$$S_n[f] = \sum_{k=0}^n A_k f(x_k), \tag{8.5}$$

где $A_k \in \mathbb{R}$ — квадратурные коэффициенты, x_k — квадратурные узлы $(k = 0,1,2,\ldots,n)$.

Квадратурные формулы ...

Определение 8.1.

Формула численного интегрирования

$$I[f] \approx S_n[f] \tag{8.6}$$

называется квадратурной, если ее коэффициенты A_k и узлы x_k (k=0, 1, 2, . . ., n) не зависят от функции f.

Определение 8.2.

Квадратурная формула (8.6) называется интерполяционной, если ее коэффициенты A_k (k=0, 1, 2, . . ., n) вычисляются по формуле (8.4).

ПРИМЕР. $f \in C([a,b])$. Формула (см. Теорему о среднем)

$$\int_a^b f(x)dx = (b-a)f(c), \quad c \in [a,b]$$

не является квадратурной[27].

[27] Пояснить самостоятельно.

Характеристическое свойство интерполяционной квадратурной формулы

Теорема 8.1. Критерий интерполяционности квадратурной формулы

Для того, чтобы квадратурная формула (8.6) была интерполяционной необходимо и достаточно, чтобы она была точна для любого многочлена степени n и ниже, то есть

$$I[P_n] = S_n[P_n] \quad \forall P_n(x). \tag{8.7}$$

Доказательство.

Необходимость

Пусть квадратурная формула (8.6) является интерполяционной. Требуется доказать, что она точна для любого многочлена $P_n(x)$ степени n.

Характеристическое свойство ...

Действительно,

$$S_n[P_n] = \sum_{k=0}^{n} A_k P_n(x_k) =$$

$$= \sum_{k=0}^{n} P_n(x_k) \int_a^b \Pi_{i=0, i\neq k}^{n} \frac{x-x_i}{x_k-x_i} dx =$$

$$= \int_a^b \sum_{k=0}^{n} P_n(x_k) \Pi_{i=0, i\neq k}^{n} \frac{x-x_i}{x_k-x_i} dx =$$

$$= \int_a^b P_n(x) dx = I[P_n],$$

так как $\sum_{k=0}^{n} P_n(x_k) \Pi_{i=0, i\neq k}^{n} \frac{x-x_i}{x_k-x_i} = P_n(x)$.

Достаточность

Пусть квадратурная формула (8.6) точна для любого многочлена $P_n(x)$ степени n. Требуется доказать, что ее коэффициенты A_k (k=0, 1, 2, . . ., n) вычисляются по формуле (8.4).

Характеристическое свойство ...

В частности, формула (8.6) точна для всех многочленов $Q_n^{(j)}(x) = \Pi_{i=0, i\neq j}^{n} \frac{x-x_i}{x_j-x_i}$ $(j = 0, 1, 2, \ldots, n)$. Следовательно,

$$\int_a^b Q_n^{(j)}(x)dx = \sum_{k=0}^{n} A_k Q_n^{(j)}(x_k) = A_j,$$

поскольку

$$Q_n^{(j)}(x_k) = \delta_{jk} = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases}$$

$$\forall j = 0, 1, 2, \ldots, n.$$

Таким образом, коэффициенты A_k $(k{=}0, 1, 2, \ldots, n)$ квадратурной формулы (8.6) вычисляются по формуле (8.4), то есть квадратурная формула (8.6) является интерполяционной.

□ Теорема доказана.

Алгебраическая степень точности квадратурной формулы

Определение 8.3.

Алгебраической степенью точности квадратурной формулы (8.6) называется неотрицательное целое число N такое, что

1. Формула (8.6) точна для любого многочлена $P_N(x)$;
2. Существует многочлен $Q_{N+1}(x)$, для которого формула (8.6) не точна, то есть $I[Q_{N+1}] \neq S_n[Q_{N+1}]$.

Из характеристического свойства интерполяционной квадратурной формулы (теоремы 8.1) следует, что, если квадратурная формула (8.6) является интерполяционной, то ее алгебраическая степень точности $N \geq n$, где $n+1$ — количество узлов, по которым формула (8.6) построена.

Примеры интерполяционных квадратурных формул

Случай $n = 0$: $x_0 \in [a, b]$. Формула прямоугольников.

$$I[f] \approx S_0[f] = (b - a) f(x_0) \tag{8.8}$$

- Если $x_0 = a$, то формула (8.8) называется формулой левых прямоугольников, для которой $N = 0$ (см. рис. 11(I));
- Если $x_0 = b$, то формула (8.8) называется формулой правых прямоугольников, для которой $N = 0$ (см. рис. 11(II));

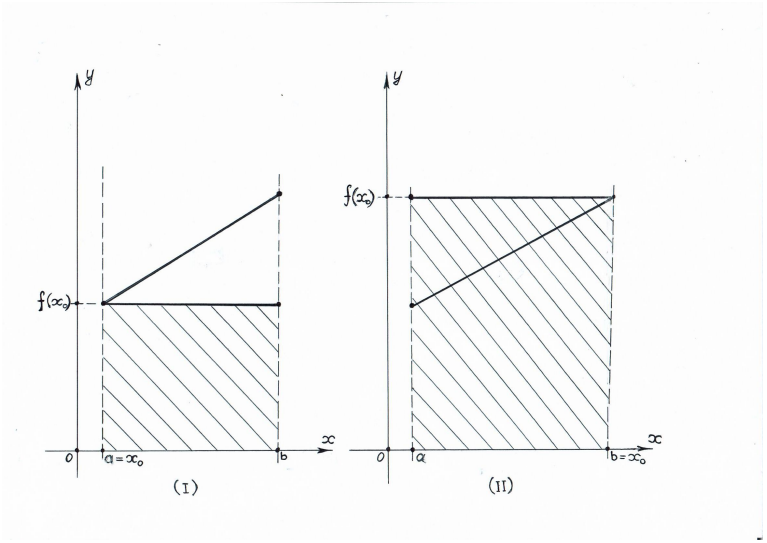


Рис. 11: Формулы левых (I) и правых (II) прямоугольников. Геометрическая интерпретация.

Примеры интерполяционных квадратурных формул

- Если $x_0 = \frac{a+b}{2}$, то формула (8.8) называется формулой средних прямоугольников, для которой $N = 1$ (см. рис. 12).

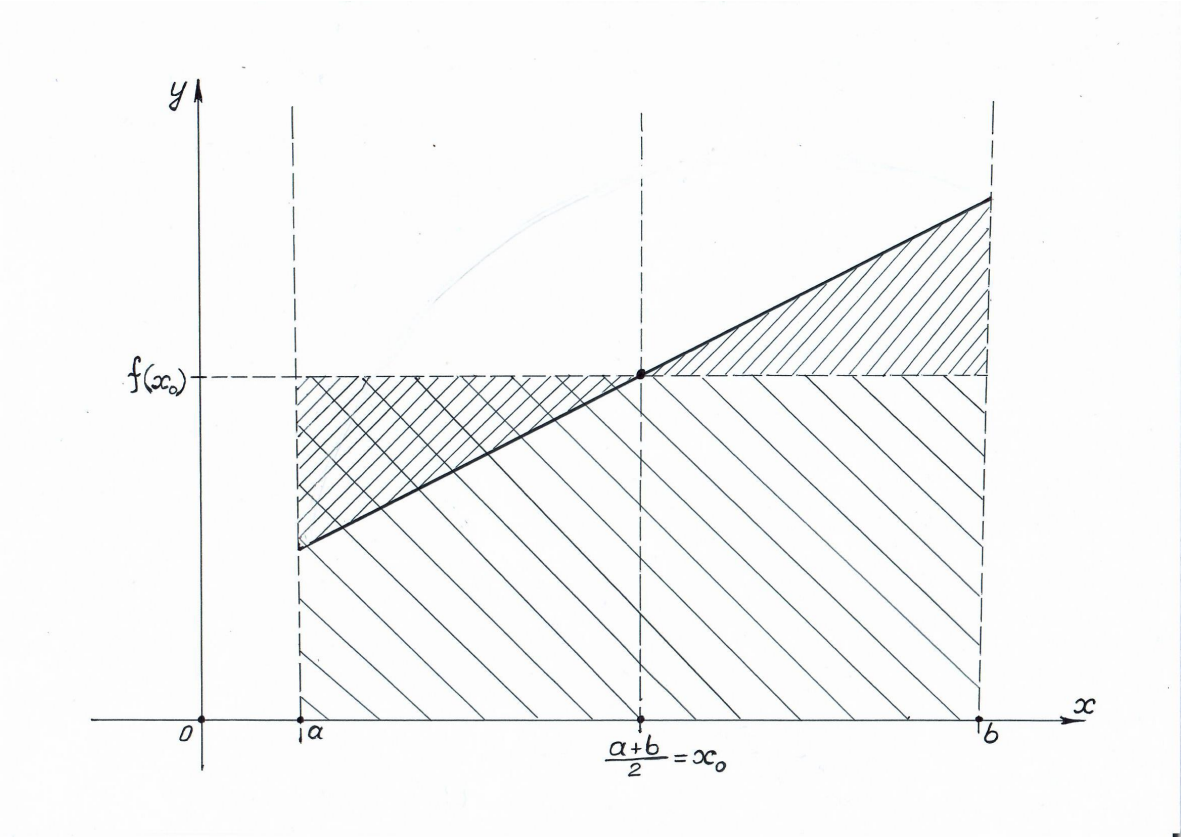


Рис. 12: Формула средних прямоугольников. Геометрическая интерпретация.

Примеры интерполяционных квадратурных формул

Случай $n = 1$: $x_0 = a, x_1 = b$. Формула трапеций.

$$I[f] \approx S_1[f] = \frac{b-a}{2}\left(f(a) + f(b)\right) \tag{8.9}$$

Алгебраическая степень точности $N = 1$ (см. рис. 13)

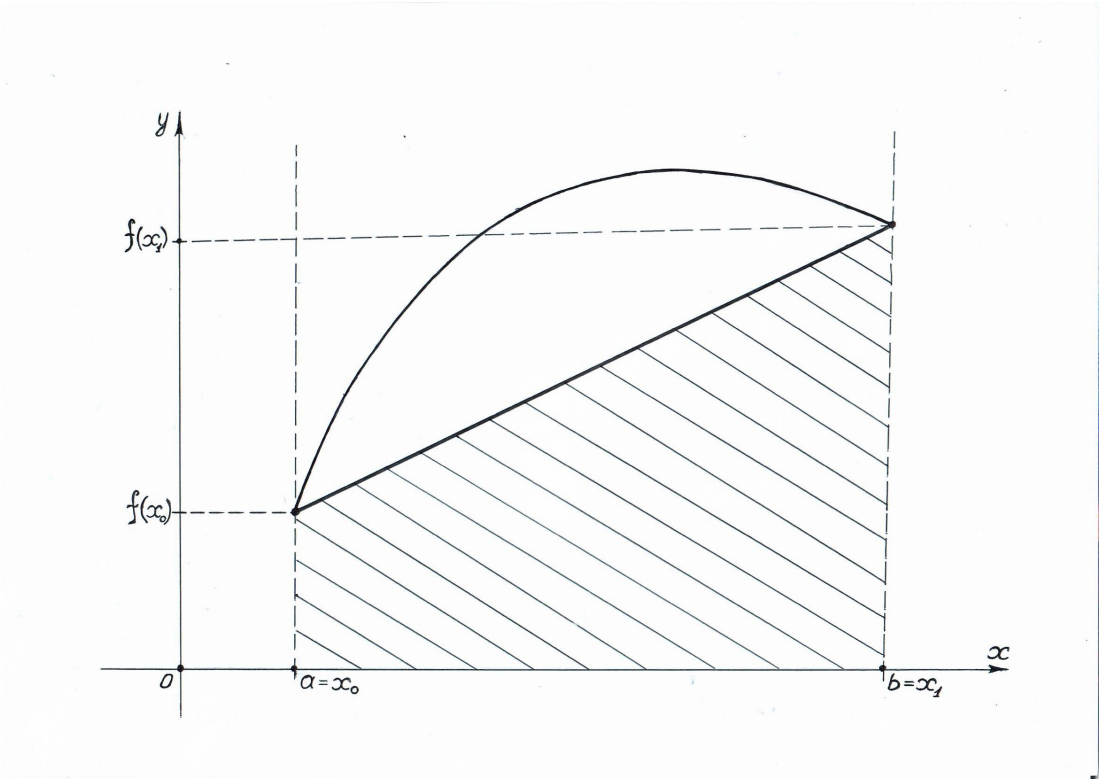


Рис. 13: Формула трапеций. Геометрическая интерпретация.

Примеры интерполяционных квадратурных формул ...

Случай $n = 2$: $x_0 = a, x_1 = \frac{a+b}{2}, x_2 = b$. Формула Симпсона.

$$I[f] \approx S_2[f] = A_0 f(a) + A_1 f(\frac{a+b}{2}) + A_2 f(b) \qquad (8.10)$$

Предлагается, используя характеристическое свойство интерполяционной квадратурной формулы (теорему 8.1), определить в квадратурной формуле (8.10) значения ее коэффициентов A_0, A_1, A_2 так, чтобы формула (8.10) имела наибольшую алгебраическую степень точности.
Из требования точности формулы (8.10) для многочленов степеней $0, 1, 2$: $Q_0(x) \equiv 1$, $Q_1(x) = x - a$, $Q_2(x) = (x - a)^2$ возникает следующая система линейных алгебраических уравнений относительно A_0, A_1, A_2:

$$\begin{cases} A_0 + A_1 + A_2 = b - a \\ \frac{b-a}{2} A_1 + (b-a) A_2 = \frac{(b-a)^2}{2} \\ (\frac{b-a}{2})^2 A_1 + (b-a)^2 A_2 = \frac{(b-a)^3}{3} \end{cases} \qquad (8.11)$$

Примеры интерполяционных квадратурных формул ...

Решением системы (8.11) является:

$$A_0 = \frac{b-a}{6}, A_1 = \frac{4(b-a)}{6}, A_2 = \frac{b-a}{6}.$$

Поскольку указанные выше многочлены $Q_0(x)$, $Q_1(x)$, $Q_2(x)$ образуют базис в пространстве многочленов степени не выше второй, а определенный интеграл $I[f]$ и квадратурная сумма $S_n[f]$ (см. (8.5)) обладают свойствами линейности и однородности

$$\forall \alpha, \beta \in \mathbb{R}, \quad \forall f, g$$

$$I[\alpha f + \beta g] = \alpha I[f] + \beta I[g], \quad S_n[\alpha f + \beta g] = \alpha S_n[f] + \beta S_n[g],$$

то квадратурная формула

$$I[f] \approx S_2[f] = \frac{b-a}{6} f(a) + \frac{4(b-a)}{6} f(\frac{a+b}{2}) + \frac{b-a}{6} f(b) \qquad (8.12)$$

точна для любого многочлена второй степени. Следовательно, в силу теоремы 8.1, эта формула является интерполяционной.

Примеры интерполяционных квадратурных формул ...

Непосредственными вычислениями значений определенного интеграла и квадратурной суммы в (8.12) несложно убедиться, что формула (8.12) точна для многочлена $Q_3(x) = (x-a)^3$ и неточна, например, для многочлена $Q_4(x) = (x-a)^4$. Следовательно, ее алгебраическая степень точности $N = 3$.[28]

Таким образом, интерполяционная квадратурная формула Симпсона имеет вид

$$I[f] \approx S_2[f] = \frac{b-a}{6}\left(f(a) + 4f(\frac{a+b}{2}) + f(b)\right) \qquad (8.13)$$

Алгебраическая степень точности $N = 3$, где $N > n = 2$.

---

[28] Более изящным способом показать, что формула (8.12) не точна для многочленов четвертой степени, является следующий. Рассмотрим $Q_2(x) = (x - \frac{a+b}{2})^2$ и $Q_4^{(\alpha)}(x) = \alpha(x - \frac{a+b}{2})^4$, где $\alpha \in \mathbb{R}$. Здесь $\forall\alpha$ $Q_2(\frac{a+b}{2}) = Q_4^{(\alpha)}(\frac{a+b}{2}) = 0$ и $Q_2(a) = Q_2(b)$, $Q_4^{(\alpha)}(a) = Q_4^{(\alpha)}(b)$. При этом $\exists!\alpha^\star > 0$: $Q_2(a) = Q_4^{(\alpha^\star)}(a)$ и $Q_2(b) = Q_4^{(\alpha^\star)}(b)$. В силу (8.12), $S_2[Q_2] = S_2[Q_4^{(\alpha^\star)}]$. Однако, очевидно (см. рис. 14), что $I[Q_2] \neq I[Q_4^{(\alpha^\star)}]$.

Примеры интерполяционных квадратурных формул ...

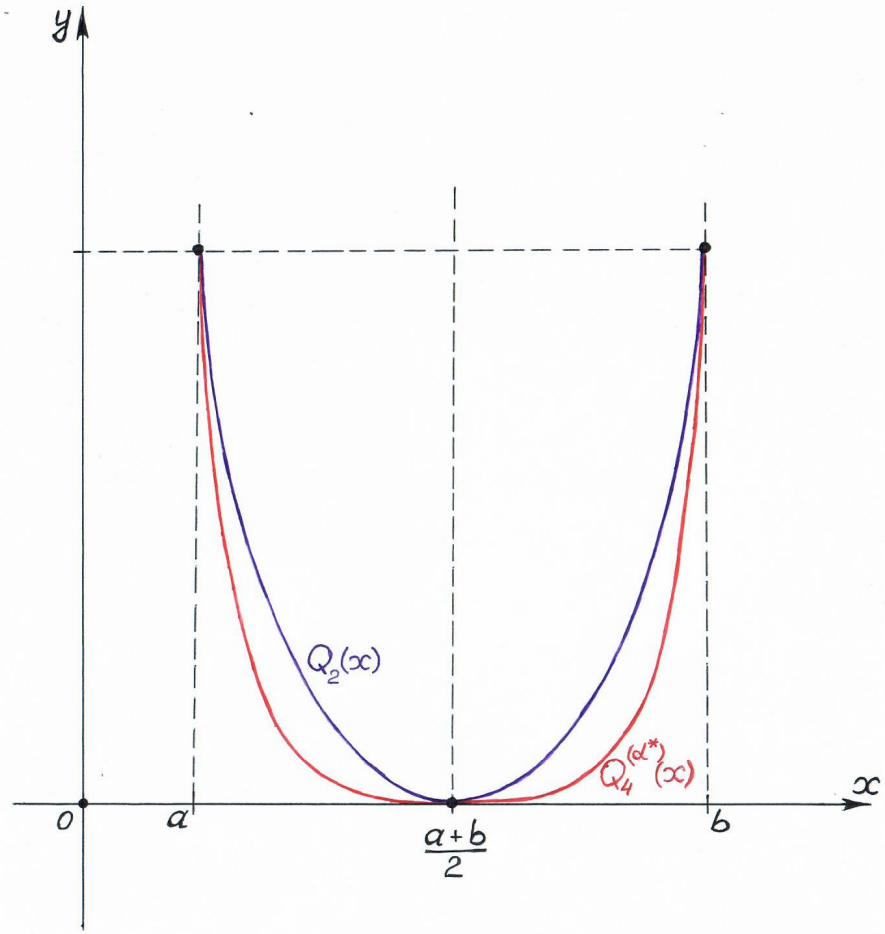


Рис. 14: Графики многочленов $Q_2(x)$ и $Q_4^{(\alpha^\star)}(x)$, $x \in [a, b]$.

Формулы Ньютона-Котеса

Метод неопределенных коэффициентов

Численная процедура построения квадратурной формулы, имеющую наибольшую алгебраическую степень точности, в основу которой положено характеристическое свойство интерполяционной квадратурной формулы (теорема 8.1), называют методом неопределенных коэффициентов.

Случай $n = 3$: $x_0 = a, x_1 = a + \frac{b-a}{3}, x_2 = a + \frac{2(b-a)}{3}, x_3 = b$.

Соответствующая этому случаю интерполяционная квадратурная формула называется формулой "$\frac{3}{8}$-ых"[a].

[a] Построить самостоятельно, используя метод неопределенных коэффициентов.

Определение 8.4.

Интерполяционные квадратурные формулы, построенные для наборов равноотстоящих соседних узлов, называются формулами Ньютона-Котеса.

Погрешность интерполяционной квадратурной формулы

В силу (8.2) и (8.3)

$$R_n[f] = \int_a^b R_n(x)dx = \int_a^b f(x, x_0, x_1, \ldots, x_n)\omega_n(x)dx.$$

Оценка погрешности

Пусть $f \in C^{(n+1)}([a, b])$. Тогда

$$|R_n[f]| \le \int_a^b |R_n(x)|dx = \int_a^b |f(x, x_0, x_1, \ldots, x_n)||\omega_n(x)|dx =$$

$$= \int_a^b \frac{|f^{(n+1)}(\xi(x))|}{(n+1)!}|\omega_n(x)|dx \le$$

$$\le \frac{M_{n+1}}{(n+1)!}\int_a^b |\omega_n(x)|dx, \tag{8.14}$$

где $\xi(x) \in [x, x_0, x_1, \ldots, x_n]$, $M_{n+1} = \max_{x\in[a,b]} |f^{(n+1)}(x)|$.

Оценки погрешностей некоторых формул Ньютона-Котеса

Формула средних прямоугольников

$$R_0[f] = \int_a^b f(x, \frac{a+b}{2})(x - \frac{a+b}{2})dx \tag{8.15}$$

$$\int_a^b |x - \frac{a+b}{2}|dx = 2\int_{\frac{a+b}{2}}^b \left(x - \frac{a+b}{2}\right)dx = \frac{2\left(x - \frac{a+b}{2}\right)^2}{2}|_{\frac{a+b}{2}}^b = \frac{(b-a)^2}{4}$$

Пусть $f \in C^{(1)}([a,b])$. Тогда

$$|R_0[f]| \leq \frac{M_1}{4}(b-a)^2, \tag{8.16}$$

где $M_1 = \max_{x\in[a,b]} |f'(x)|$.

Оценки погрешностей ...

Формула трапеций

$$R_1[f] = \int_a^b f(x,a,b)(x-a)(x-b)dx \tag{8.17}$$

$\int_a^b |(x-a)(x-b)|dx = -\int_a^b (x-a)(x-b)dx$

$\int_a^b (x-a)(x-b)dx = \frac{1}{2}\int_a^b (x-a)((x-b)^2)'dx =$

$= -\frac{1}{2}\int_a^b (x-b)^2 dx = -\frac{(x-b)^3}{6}|_a^b = -\frac{(b-a)^3}{6}$

Пусть $f \in C^{(2)}([a,b])$. Тогда

$$|R_1[f]| \le \frac{M_2}{12}(b-a)^3, \tag{8.18}$$

где $M_2 = \max_{x\in[a,b]} |f''(x)|$.

$\boxtimes$

▲16. 8.5. Оценка погрешности интерполяционной квадратурной формулы для случая $N > n$

Пусть $Q_N(x)$ — интерполяционный многочлен Эрмита, построенный для функции f по набору кратных узлов $x_0, x_1, x_2, \ldots, x_n \in [a, b]$.
Поскольку $I[Q_N] = S_n[Q_N]$, то

$$R_n[f] = I[f] - S_n[f] = (I[f] - I[Q_N]) - (S_n[f] - S_n[Q_N]).$$

Откуда, в силу свойств линейности определенного интеграла и квадратурной суммы,

$$R_n[f] = I[f - Q_N] - S_n[f - Q_N],$$

где $S_n[f - Q_N] = 0$, так как $Q_N(x_i) = f(x_i)\ \forall i = 0, 1, \ldots, n$.

$$R_n[f] = I[f - Q_N] \tag{8.19}$$

[29]

[29] Формула (8.19) справедлива для любого интерполяционного многочлена Q_m степени $m \leq N$.

ПРИМЕР. Формула средних прямоугольников

$n = 0$, $x_0 = \frac{a+b}{2}$, $N = 1$

Пусть $Q_1(x)$ — интерполяционный многочлен Эрмита, построенный по двукратному узлу $x_0 = \frac{a+b}{2}$, то есть $Q_1(x_0) = f(x_0)$, $Q_1'(x_0) = f'(x_0)$.
Тогда, в силу (8.19),

$$R_0[f] = \int_a^b f(x, \tfrac{a+b}{2}, \tfrac{a+b}{2})(x - \tfrac{a+b}{2})^2 dx = \\ = f(\eta, \tfrac{a+b}{2}, \tfrac{a+b}{2}) \int_a^b (x - \tfrac{a+b}{2})^2 dx = \frac{f''(\xi)}{2} \frac{(b-a)^3}{12},$$

где $f \in C^{(2)}([a,b])$, $\eta \in [a,b]$ (см. теорему о среднем), $\xi \in [a,b]$.

Оценка погрешности

$$|R_0[f]| \le \frac{M_2}{24}(b-a)^3, \tag{8.20}$$

где $f \in C^{(2)}([a,b])$, $M_2 = \max_{x\in[a,b]} |f''(x)|$ [a].

[a] Если $f \in C^{(1)}([a,b])$, то $|R_0[f]| \le \frac{M_1}{4}(b-a)^2$, где $M_1 = \max_{x\in[a,b]} |f'(x)|$ (см. (8.16)).

Оценка погрешности формулы Симпсона

$n = 2,\ x_0 = a, x_1 = \frac{a+b}{2}, x_2 = b,\ N = 3$

Пусть $Q_3(x)$ — интерполяционный многочлен Эрмита, построенный по набору узлов x_0, x_1, x_2, где $x_1 = \frac{a+b}{2}$ — двукратный узел, то есть $Q_3(x_1) = f(x_1)$, $Q_3'(x_1) = f'(x_1)$. Тогда, в силу (8.19),

$$
\begin{aligned}
R_2[f] &= \int_a^b (f(x) - Q_3(x))dx = \\
&= \int_a^b f(x, a, \tfrac{a+b}{2}, \tfrac{a+b}{2}, b)(x-a)(x-\tfrac{a+b}{2})^2(x-b)dx = \\
&= f(\eta, a, \tfrac{a+b}{2}, \tfrac{a+b}{2}, b) \int_a^b (x-a)(x-\tfrac{a+b}{2})^2(x-b)dx = \\
&= -\frac{f^{(4)}(\xi)}{4!}\frac{1}{120}(b-a)^5 = -\frac{f^{(4)}(\xi)}{2880}(b-a)^5,
\end{aligned}
\tag{8.21}
$$

где $f \in C^{(4)}([a,b])$, $\eta \in [a,b]$ (см. теорему о среднем), $\xi \in [a,b]$.

Оценка погрешности формулы Симпсона ...

$$\int_a^b (x-a)(x-\tfrac{a+b}{2})^2(x-b)dx = \tfrac{1}{3}\int_a^b (x-a)(x-b)d(x-\tfrac{a+b}{2})^3 =$$

$$= -\tfrac{1}{3}\int_a^b (2x-a-b)(x-\tfrac{a+b}{2})^3 dx = -\tfrac{2}{3}\int_a^b (x-\tfrac{a+b}{2})^4 dx =$$

$$= -\tfrac{2}{3}\frac{\left(x-\frac{a+b}{2}\right)^5}{5}\Big|_a^b = -\tfrac{4}{3\cdot 5}\left(\tfrac{b-a}{2}\right)^5 = -\tfrac{1}{120}(b-a)^5$$

Из (8.21) следует

Оценка погрешности формулы Симпсона

Пусть $f \in C^{(4)}([a,b])$. Тогда

$$|R_2[f]| \le \frac{M_4}{2880}(b-a)^5, \tag{8.22}$$

где $M_4 = \max_{x\in[a,b]} |f^{(4)}(x)|$.

30

---

[30] Если $f \in C^{(3)}([a,b])$, то
$R_2[f] = \int_a^b f(x,a,\frac{a+b}{2},b)(x-a)(x-\frac{a+b}{2})(x-b)dx = D(b-a)^4$.

8.6. Квадратурный процесс

Создается впечатление, что для того, чтобы увеличить точность квадратурной формулы нужно в ней увеличивать количество узлов.

Определение 8.5.

Пусть задана система узлов $x_0^{(n)}, x_1^{(n)}, \ldots, x_n^{(n)} \in [a,b]$. Под квадратурным процессом для некоторой функции $f(x)$ ($[a,b] \subseteq \mathrm{D}[f]$) понимают числовую последовательность $\{S_n[f]\}_{n=0}^{\infty}$: $I[f] \approx S_n[f] = \sum_{k=0}^{n} A_k^{(n)} f(x_k^{(n)})\ \forall n \geq 0$.

Определение 8.6.

Квадратурный процесс $\{S_n[f]\}_{n=0}^{\infty}$ для функции f называется сходящимся, если

$$\exists \lim_{n\to\infty} S_n[f] = I[f].$$

Сходимость квадратурного процесса

Теорема 8.2. Критерий сходимости интерполяционного квадратурного процесса

Для сходимости интерполяционного квадратурного процесса для любой непрерывной функции f необходимо и достаточно, чтобы

$$\exists M > 0 : \forall n \geq 0 \Rightarrow \sum_{k=0}^{n} |A_k^{(n)}| \leq M. \tag{8.23}$$

Доказательство.

Необходимость[a].

[a] Березин И. С., Жидков Н. П. Методы вычислений. 1962.

Достаточность

Пусть выполнено условие (8.23).
Зафиксируем функцию $f \in C([a, b])$.

Критерий сходимости интерполяционного квадратурного процесса ...

Поскольку пространство многочленов плотно в пространстве непрерывных функций, то

$$\forall \varepsilon > 0 \quad \exists Q_{m(\varepsilon)} : \quad \max_{x \in [a,b]} |f(x) - Q_{m(\varepsilon)}(x)| < \varepsilon \tag{8.24}$$

Так как формула $I[f] \approx S_n[f]$ является интерполяционной для всех $n \geq 0$, то, в силу (8.7) (см. критерий интерполяционности квадратурной формулы), эта формула будет точна для многочлена $Q_{m(\varepsilon)}$ для любого $n \geq m(\varepsilon)$, то есть

$$I[Q_{m(\varepsilon)}] = S_n[Q_{m(\varepsilon)}] \quad \forall n \geq m(\varepsilon) \tag{8.25}$$

Следовательно, учитывая линейность определенного интеграла и квадратурной суммы, $\forall n \geq m(\varepsilon)$ справедливо равенство

$$\begin{aligned} I[f] - S_n[f] &= \big(I[f] - I[Q_{m(\varepsilon)}]\big) - \big(S_n[f] - S_n[Q_{m(\varepsilon)}]\big) = \\ &= I[f - Q_{m(\varepsilon)}] - S_n[f - Q_{m(\varepsilon)}] \end{aligned} \tag{8.26}$$

Критерий сходимости интерполяционного квадратурного процесса ...

Таким образом, из (8.24), (8.25), (8.26) следует, что $\forall n \geq m(\varepsilon)$ справедлива оценка

$$|I[f] - S_n[f]| \leq |I[f - Q_{m(\varepsilon)}]| + |S_n[f - Q_{m(\varepsilon)}]| =$$

$$= |\int_a^b \left(f(x) - Q_{m(\varepsilon)}(x)\right) dx| +$$

$$+ |\sum_{k=0}^n A_k^{(n)} \left(f(x_k^{(n)}) - Q_{m(\varepsilon)}(x_k^{(n)})\right)| \leq$$

$$\leq \int_a^b |f(x) - Q_{m(\varepsilon)}(x)| dx +$$

$$+ \sum_{k=0}^n |A_k^{(n)}| \cdot |f(x_k^{(n)}) - Q_{m(\varepsilon)}(x_k^{(n)})| \leq$$

$$\leq \varepsilon \int_a^b 1 dx + \varepsilon \sum_{k=0}^n |A_k^{(n)}| \leq \varepsilon \left((b-a) + M\right).$$

□ Теорема доказана.

Критерий сходимости интерполяционного квадратурного процесса ...

Следствие

Если $A_k^{(n)} \geq 0 \ \forall n \geq 0$ и $\forall k = 0, 1, \ldots, n$, то интерполяционный квадратурный процесс для любой $f \in C([a, b])$ сходится.

Доказательство. Поскольку интерполяционная квадратурная формула точна для $Q_0(x) \equiv 1$, то

$$\sum_{k=0}^{n} |A_k^{(n)}| = \sum_{k=0}^{n} A_k^{(n)} \cdot 1 = \int_a^b 1 dx = b - a = M.$$

□ Следствие доказано.

Поскольку в формулах Ньютона-Котеса, начиная с $n = 8$, среди коэффициентов $A_k^{(n)}$ появляются отрицательные числа и, более того, для этих формул последовательность $\{\sum_{k=0}^{n} |A_k^{(n)}|\}_{n=0}^{\infty}$ расходится, то с помощью формул Ньютона-Котеса невозможно вычислить $I[f]$ с заранее заданной точностью.

8.7. Составные квадратурные формулы

Пусть $x_0, x_1, \ldots, x_m$ — разбиение отрезка интегрирования $[a, b]$ на m элементарных отрезков $[x_i, x_{i+1}]$ $(i = 0, 1, \ldots, m-1)$:

$$x_{i+1} = x_i + h \quad i = 0, 1, \ldots, m-1 : \quad x_0 = a, h = \frac{b-a}{m}.$$

Составная квадратурная формула

$$I[f] = \int_a^b f(x)dx = \sum_{i=0}^{m-1} \int_{x_i}^{x_{i+1}} f(x)dx \approx \sum_{i=0}^{m-1} S_n[f|x_i, x_{i+1}], \quad (8.27)$$

где $S_n[f|x_i, x_{i+1}]$ — квадратурная сумма для функции f, которая построена по $n+1$-ому узлу, принадлежащим отрезку элементарному $[x_i, x_{i+1}]$ $(i = 0, 1, \ldots, m-1)$.

Примеры составных квадратурных формул

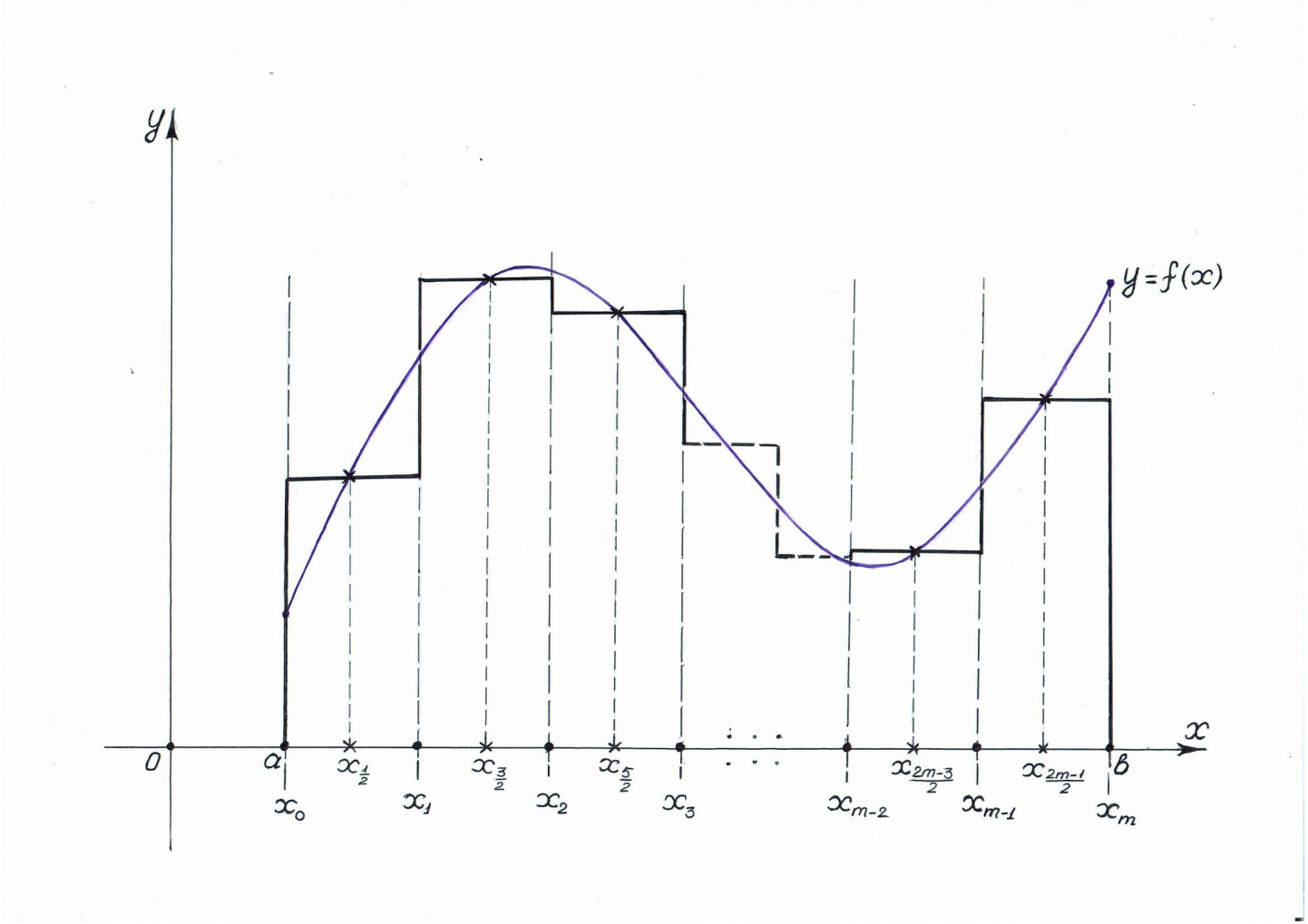


Рис. 15: Геометрическая интерпретация составной формулы средних прямоугольников.

Примеры составных квадратурных формул

Составная формула средних прямоугольников

$$I[f] \approx \sum_{i=0}^{m-1} h f(x_{i+\frac{1}{2}}) = \frac{b-a}{m}\left(f(x_{\frac{1}{2}}) + f(x_{\frac{3}{2}}) + \ldots + f(x_{\frac{2m-1}{2}})\right), \tag{8.28}$$

где $x_{i+\frac{1}{2}} = x_i + \frac{h}{2}$ $(i = 0, 1, \ldots, m-1)$.

Пусть $f \in C^{(1)}([a,b])$ и $M_1 = \max_{x\in[a,b]} |f'(x)|$. Тогда (см. (8.16))

$$|R_m[f]| \le \sum_{i=0}^{m-1} |R_0[f|x_i, x_{i+1}]| \le \sum_{i=0}^{m-1} \frac{M_1}{4} h^2 =$$

$$= \frac{M_1}{4} h \sum_{i=0}^{m-1} h = \frac{M_1}{4} \cdot \frac{b-a}{m} \cdot (b-a) = \frac{M_1 (b-a)^2}{4m}$$

Погрешность составной формулы средних прямоугольников

$$|R_m[f]| \le \frac{M_1}{4} \cdot \frac{(b-a)^2}{m} \tag{8.29}$$

Примеры составных квадратурных формул ...

Пусть $f \in C^{(2)}([a,b])$ и $M_2 = \max_{x\in[a,b]} |f''(x)|$. Тогда (см. (8.20))

$$|R_m[f]| \le \sum_{i=0}^{m-1} |R_0[f|x_i, x_{i+1}]| \le$$

$$\le \sum_{i=0}^{m-1} \frac{M_2}{24} h^3 =$$

$$= \frac{M_2}{24} h^2 \sum_{i=0}^{m-1} h = \frac{M_2}{24} \cdot \frac{(b-a)^2}{m^2} \cdot (b-a) =$$

$$= \frac{M_2}{24} \cdot \frac{(b-a)^3}{m^2}$$

Погрешность составной формулы средних прямоугольников

Если $f \in C^{(2)}([a,b])$ и $M_2 = \max_{x\in[a,b]} |f''(x)|$, то

$$|R_m[f]| \le \frac{M_2}{24} \cdot \frac{(b-a)^3}{m^2} \tag{8.30}$$

Примеры составных квадратурных формул ...

Составная формула трапеций

$$I[f] \approx \sum_{i=0}^{m-1} \frac{h}{2} \left(f(x_i) + f(x_{i+1})\right) = \\ = \frac{b-a}{2m} \left(f(x_0) + 2f(x_1) + \ldots + 2f(x_{m-1}) + f(x_m)\right) \tag{8.31}$$

Используя оценку (8.18), аналогично (8.29), (8.30) нетрудно показать, что

Погрешность составной формулы трапеций

Пусть $f \in C^{(2)}([a,b])$ и $M_2 = \max_{x\in[a,b]} |f''(x)|$. Тогда

$$|R_m[f]| \leq \frac{M_2}{12} \cdot \frac{(b-a)^3}{m^2} \tag{8.32}$$

Примеры составных квадратурных формул ...

Составная формула Симпсона

$$I[f] \approx \sum_{i=0}^{m-1} \frac{h}{6}\left(f(x_i) + 4f(x_{i+\frac{1}{2}}) + f(x_{i+1})\right) =$$

$$= \frac{b-a}{6m}[f(x_0) + 4f(x_{\frac{1}{2}}) + 2f(x_1) + 4f(x_{\frac{3}{2}}) + 2f(x_2) + \tag{8.33}$$

$$+ \ldots + 2f(x_{m-1}) + 4f(x_{m-\frac{1}{2}}) + f(x_m)],$$

Используя оценку (8.22), аналогично (8.29), (8.30) нетрудно показать, что

Погрешность составной формулы Симпсона

Пусть $f \in C^{(4)}([a,b])$ и $M_4 = \max_{x\in[a,b]} |f^{(4)}(x)|$. Тогда

$$|R_m[f]| \le \frac{M_4}{2880} \cdot \frac{(b-a)^5}{m^4} \tag{8.34}$$

Сходимость квадратурного процесса для составных квадратурных формул

В силу (8.29), (8.30), при $m \to \infty$

$$|R_m[f]| \leq \frac{M_1}{4} \cdot \frac{(b-a)^2}{m} \to 0,$$
$$|R_m[f]| \leq \frac{M_2}{24} \cdot \frac{(b-a)^3}{m^2} \to 0.$$

Вывод

Следовательно, квадратурный процесс для составной квадратурной формулы средних прямоугольников сходится. Учитывая структуру правых частей неравенств (8.29) и (8.30), погрешность составной квадратурной формулы средних прямоугольников монотонно уменьшается при увеличении количества узлов в разбиении отрезка интегрирования $[a, b]$.

Аналогичный вывод можно сделать и для составных квадратурных формул трапеций и Симпсона на основе соответствующих оценок (8.32) и (8.34) этих формул.

$\boxtimes$

▲17. 8.8. Метод Рунге практической оценки погрешности составной квадратурной формулы

Составная формула Симпсона

Если $f \in C^{(4)}([a,b])$, то с учетом (8.21)

$$R_m[f] = \sum_{i=0}^{m-1} R_2[f|x_i, x_{i+1}] = \\ = -\sum_{i=0}^{m-1} \frac{f^{(4)}(\xi_i)}{2880} h^5 = -\frac{h^4}{2880} \sum_{i=0}^{m-1} f^{(4)}(\xi_i) h, \tag{8.35}$$

где $\xi_i \in [x_i, x_{i+1}]$ $(i = 0, 1, \ldots, m-1)$.

$$I[f^{(4)}] = \int_a^b f^{(4)}(x)dx = \sum_{i=0}^{m-1} f^{(4)}(\xi_i)h + \varepsilon_m(h), \tag{8.36}$$

где $\varepsilon_m(h) \to 0$ при $m \to \infty$ $(h \to 0)$.

Пусть $c = -\frac{1}{2880} \int_a^b f^{(4)}(x)dx$.

Метод Рунге практической оценки погрешности составной формулы Симпсона ...

Тогда в силу (8.35) и (8.36) погрешность составной формулы Симпсона может быть записана в виде

$$R_m[f] = c \cdot h^4 + o(h^4) \tag{8.37}$$

Из (8.37) следует, что

$$R_{2m}[f] = c \cdot \left(\frac{h}{2}\right)^4 + o\left(\left(\frac{h}{2}\right)^4\right) = \left(\frac{1}{2}\right)^4 \cdot c \cdot h^4 + o(h^4).$$

Следовательно, с точностью до $o(h^4)$ справедливо

$$R_{2m}[f] \approx \left(\frac{1}{2}\right)^4 R_m[f]. \tag{8.38}$$

Откуда

$$\begin{aligned} I[f] &= S_m[f] + R_m[f], \\ I[f] &= S_{2m}[f] + R_{2m}[f] \approx S_{2m}[f] + \tfrac{1}{16} R_m[f], \end{aligned} \tag{8.39}$$

где $S_k[f]$ — квадратурная сумма в правой части равенства (8.33), которое определяет составную формулу Симпсона для k элементарных отрезков в разбиении отрезка $[a, b]$.

Метод Рунге практической оценки погрешности составной формулы Симпсона ...

Из (8.39) следует

$$R_m[f] \approx \frac{16}{15} \left(S_{2m}[f] - S_m[f] \right), \tag{8.40}$$

Учитывая (8.38),

Формула Рунге для составной формулы Симпсона

$$R_{2m}[f] \approx \frac{1}{15} \left(S_{2m}[f] - S_m[f] \right). \tag{8.41}$$

Метод Рунге практической оценки погрешности составной формулы Симпсона ...

Из (8.38) и (8.41) можно получить приближенное условие стабилизации величины погрешности составной формулы Симпсона

$$\frac{S_{4m}[f] - S_{2m}[f]}{S_{2m}[f] - S_m[f]} \approx \frac{R_{4m}[f]}{R_{2m}[f]} \approx \frac{1}{16}. \tag{8.42}$$

Если условие (8.42) достаточно точно выполняется, то отброшенная в равенствах (8.37) величина $o(h^4)$ уже не имеет принципиального значения в сравнении с погрешностью $R_{4m}[f]$.

Если же при увеличении m стабилизация не происходит (условие (8.42) не выполняется), то интегрируемая функция f не является достаточно гладкой.

Формула Рунге для составной формулы трапеций

В силу (8.17), (8.18)

$$R_m[f] = \sum_{i=0}^{m-1} R_1[f|x_i, x_{i+1}] = -\sum_{i=0}^{m-1} \frac{f''(\xi_i)}{12} h^3 = \\ = -\frac{1}{12} h^2 \sum_{i=0}^{m-1} f''(\xi_i) h = c \cdot h^2 + o(h^2), \tag{8.43}$$

где $\xi_i \in [x_i, x_{i+1}]$ $(i = 0, 1, \ldots, m-1)$, $c = -\frac{1}{12} \int_a^b f''(x)dx$.
Аналогично (8.40) и (8.41), для составной формулы трапеций

$$R_{2m}[f] \approx \left(\frac{1}{2}\right)^2 R_m[f]. \tag{8.44}$$

Формула Рунге для составной формулы трапеций

$$R_{2m}[f] \approx \frac{1}{3} \left(S_{2m}[f] - S_m[f]\right). \tag{8.45}$$

Формула Рунге для составной квадратурной формулы. Общий случай.

Пусть k — порядок точности составной квадратурной формулы на одном шаге (степень длины h отрезка интегрирования в выражении для погрешности на элементарном отрезке интегрирования $[x_i, x_{i+1}]$ (см. (8.35),(8.43))[31]. Тогда

$$R_m[f] = c \cdot h^{k-1} + o(h^{k-1}), \quad R_{2m}[f] \approx \left(\frac{1}{2}\right)^{k-1} R_m[f];$$

$$I[f] = S_m[f] + R_m[f], \quad I[f] = S_{2m}[f] + R_{2m}[f] \approx S_{2m}[f] + \frac{1}{2^{k-1}} R_m[f].$$

Откуда $R_m[f] \approx \frac{1}{1-\frac{1}{2^{k-1}}} (S_{2m}[f] - S_m[f])$ и

Формула Рунге для составной квадратурной формулы k-го порядка точности на одном шаге

$$R_{2m}[f] \approx \frac{S_{2m}[f] - S_m[f]}{2^{k-1} - 1}. \tag{8.46}$$

[31] Для составной формулы Симпсона $k = 5$, для составной формулы трапеций $k = 3$.

Неустранимая погрешность квадратурных формул

Пусть

1. $A_k^{(n)} \geq 0 \ \forall n \geq 0, \ \forall k = 0, 1, \ldots, n$[32];
2. $f \in C([a,b])$: $\forall x \in [a,b] \ |f(x) - \widetilde{f}(x)| \leq \varepsilon$.

Тогда

Оценка неустранимой погрешности квадратурной интерполяционной формулы

$$|S_n[f] - S_n[\widetilde{f}]| = |\sum_{k=0}^{n} A_k^{(n)} \left(f(x_k^{(n)}) - \widetilde{f}(x_k^{(n)}) \right)| \leq$$
$$\leq \varepsilon \cdot \sum_{k=0}^{n} |A_k^{(n)}| = \varepsilon \cdot \sum_{k=0}^{n} A_k^{(n)} = \varepsilon(b-a)$$

Вывод

С ростом n неустранимая погрешность квадратурной интерполяционной формулы не увеличивается.

[32] Интерполяционный квадратурный процесс сходится.

8.9. Вычисление определенных интегралов с весом

Постановка задачи

Пусть функции f и p ограничены на отрезке $[a,b] \subseteq \mathrm{D}[f] \cap \mathrm{D}[p]$: $\int_a^b p(x)x^m dx$ вычисляются аналитически $(m = 0, 1, 2, \ldots)$.

$$\begin{gathered} x_0, x_1, x_2, \ldots, x_n \in [a,b]: \quad x_i \neq x_j \ \forall i \neq j \\ y_0, y_1, y_2, \ldots, y_n: \quad y_i = f(x_i) \ \forall i = 0, 1, \ldots, n \end{gathered} \tag{8.47}$$

$$I[f] = \int_a^b p(x)f(x)dx$$

Алгоритм решения

$$\begin{gathered} f(x) = L_n(x) + R_n(x), \\ I[f] = \int_a^b p(x)L_n(x)dx + \int_a^b p(x)R_n(x)dx \end{gathered} \tag{8.48}$$

$$I[f] \approx \int_a^b p(x)L_n(x)dx = S_n[f]$$

Квадратурные формулы наивысшей алгебраической степени точности. Формула Гаусса.

Квадратурная интерполяционная формула

$$S_n[f] = \int_a^b p(x) \sum_{k=0}^{n} f(x_k) \Pi_{i=0, i\neq k}^{n} \frac{x - x_i}{x_k - x_i} dx = \sum_{k=0}^{n} A_k f(x_k),$$

$$A_k = \int_a^b p(x) \Pi_{i=0, i\neq k}^{n} \frac{x - x_i}{x_k - x_i} dx, \quad k = 0, 1, 2, \ldots, n \tag{8.49}$$

Постановка задачи

Пусть квадратурная формула

$$I[f] = \int_a^b p(x) f(x) dx \approx \sum_{k=0}^{n} A_k f(x_k), \tag{8.50}$$

строится по $n+1$-му различному узлу $x_0, x_1, x_2, \ldots, x_n \in [a, b]$. Каким образом следует выбирать ее узлы x_k и вычислять коэффициенты A_k $(k = 0, 1, 2, \ldots, n)$, чтобы формула (8.50) имела наивысшую алгебраическую степень точности?

Квадратурные формулы наивысшей алгебраической степени точности ...

Теорема 8.3.

Для того, чтобы квадратурная формула (8.50) была точна для всех многочленов степени $2n+1$ и ниже необходимо и достаточно, чтобы:

a) формула (8.50) была интерполяционной, т.е. ее коэффициенты A_k $(k = 0, 1, 2, \ldots, n)$ были вычислены по формулам (8.49);

b) узлы $x_0, x_1, x_2, \ldots, x_n \in [a, b]$ были таковыми, чтобы многочлен $\omega_n(x) = (x - x_0)(x - x_1)\ldots(x - x_n)$ был ортогонален с весом p любому многочлену $Q_m(x)$ $(m \leq n)$, то есть

$$\int_a^b p(x)Q_n(x)\omega_n(x)dx = 0 \quad \forall\, Q_n(x) \tag{8.51}$$

Доказательство

Необходимость

Пусть квадратурная формула (8.50) точна для всех многочленов степени $2n+1$ и ниже.
Справедливость условия (a) следует из критерия интерполяционности квадратурной формулы (см. Теорему 8.1).
Для любого многочлена $Q_n(x)$ справедливо

$$\int_a^b p(x)Q_n(x)\omega_n(x)dx = \int_a^b p(x)P_{2n+1}(x)dx = \sum_{k=0}^n A_k P_{2n+1}(x_k) =$$

$$= \sum_{k=0}^n A_k Q_n(x_k)\omega_n(x_k) = 0,$$

где $P_{2n+1}(x) = Q_n(x)\omega_n(x)$. Таким образом, выполняется и условие (b).

Достаточность

Пусть для квадратурной формулы (8.50) выполнены условия (a) и (b).
Произвольный многочлен $Q_{2n+1}(x)$ можно представить в виде

Доказательство ...

$$Q_{2n+1}(x) = G_n(x)\omega_n(x) + H_m(x) \quad (m \le n).$$

Тогда

$$\int_a^b p(x)Q_{2n+1}(x)dx = \int_a^b p(x)\left(G_n(x)\omega_n(x) + H_m(x)\right)dx =$$

$$= \int_a^b p(x)G_n(x)\omega_n(x)dx + \int_a^b p(x)H_m(x)dx = |_{(b)}$$

$$= \int_a^b p(x)H_m(x)dx = |_{(a)} \sum_{k=0}^n A_k H_m(x_k) =$$

$$= \sum_{k=0}^n A_k H_m(x_k) + \sum_{k=0}^n A_k G_n(x_k) \underbrace{\omega_n(x_k)}_{0} =$$

$$= \sum_{k=0}^n A_k \left(H_m(x_k) + G_n(x_k)\omega_n(x_k)\right) =$$

$$= \sum_{k=0}^n A_k Q_{2n+1}(x_k) = S_n[Q],$$

то есть формула (8.50) точна для любого многочлена степени $2n+1$ и ниже.

□ Теорема доказана.

О существовании квадратурной формулы алгебраической степени точности $N \geq 2n+1$

Узлы $x_0, x_1, x_2, \ldots, x_n$ являются корнями многочлена $\omega_n(x) = (x - x_0)(x - x_1)\ldots(x - x_n)$, который в общем виде может быть записан следующим образом

$$G_{n+1}(x) = x^{n+1} + a_1 x^n + a_2 x^{n-1} + \ldots + a_{n-1}x^2 + a_n x + a_{n+1}, \quad (8.52)$$

где $a_i \in \mathbb{R}$ $(i = 1, 2, \ldots, n+1)$.

Задача

Требуется построить многочлен $G_{n+1}(x)$ вида (8.52), который бы был ортогонален с весом p любому многочлену $Q_m(x)$ $(m \leq n)$

Для того, чтобы многочлен $G_{n+1}(x)$ вида (8.52) был ортогонален с весом p любому многочлену $Q_m(x)$ $(m \leq n)$ необходимо и достаточно, чтобы

$$\int_a^b p(x) G_{n+1}(x) x^m dx = 0 \quad \forall m = 0, 1, 2, \ldots, n. \quad (8.53)$$

О существовании квадратурной формулы алгебраической степени точности $N \geq 2n+1$ ...

Равенства (8.53) образуют неоднородную систему, содержащую $(n+1)$-о линейное алгебраическое уравнение относительно неизвестных $a_i \in \mathbb{R}$ $(i = 1, 2, \ldots, n+1)$

$$\int_a^b p(x) \left(a_1 x^n + a_2 x^{n-1} + \ldots + a_{n-1} x^2 + a_n x + a_{n+1}\right) x^m dx =$$

$$= -\int_a^b p(x) x^{n+m+1} dx \qquad (8.54)$$

$$\forall m = 0, 1, 2, \ldots, n.$$

Неоднородная система (8.54) имеет единственное решение тогда и только тогда, когда соответствующая ей однородная система (8.55) имеет лишь тривиальное решение.

$$\int_a^b p(x) \left(a_1 x^n + a_2 x^{n-1} + \ldots + a_{n-1} x^2 + a_n x + a_{n+1}\right) x^m dx = 0 \qquad (8.55)$$

$$\forall m = 0, 1, 2, \ldots, n.$$

О существовании квадратурной формулы алгебраической степени точности $N \geq 2n+1$ ...

Достаточными условиями того, что однородная система (8.55) имеет лишь тривиальное решение являются следующие:

$$\begin{array}{l} a)\ p(x) \geq 0 \quad \forall x \in [a,b], \\ b)\ \int_a^b p(x)dx > 0. \end{array} \tag{8.56}$$

Справедливость этого может быть доказана методом от противного.
Если предположить, что однородная система (8.55) имеет нетривиальное решение $(\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_{n+1}) \neq 0$, то суммирование равенств

$$\bar{a}_{n-m+1} \int_a^b p(x) \left(\bar{a}_1 x^n + \bar{a}_2 x^{n-1} + \ldots + \bar{a}_{n-1} x^2 + \bar{a}_n x + \bar{a}_{n+1}\right) x^m dx = 0$$
$$\forall m = 0, 1, 2, \ldots, n.$$

приводит к выражению

$$\int_a^b p(x) \left(\bar{a}_1 x^n + \bar{a}_2 x^{n-1} + \ldots + \bar{a}_{n-1} x^2 + \bar{a}_n x + \bar{a}_{n+1}\right)^2 dx = 0,$$

которое противоречит условиям (8.56).

О существовании квадратурной формулы алгебраической степени точности $N \geq 2n+1$ ...

Таким образом, доказана следующая теорема

Теорема 8.4.

Пусть весовая функция $p(x)$ $(x \in [a,b])$ удовлетворяет следующим условиям:

$$a)\ p(x) \geq 0 \quad \forall x \in [a,b],$$

$$b)\ \int_a^b p(x)dx > 0, \tag{8.57}$$

$$c)\ \exists \int_a^b p(x)x^k dx \quad \forall k = 0,1,2,\ldots,2n+1.$$

Тогда существует единственный многочлен

$$G_{n+1}(x)=x^{n+1}+a_1x^n+a_2x^{n-1}+\ldots+a_{n-1}x^2+a_nx+a_{n+1},$$

который ортогонален с весом p любому многочлену $Q_n(x)$.

О существовании квадратурной формулы алгебраической степени точности $N \geq 2n+1$ ...

Теорема 8.5.

Пусть выполнены условия (8.57). Тогда многочлен

$$G_{n+1}(x)=x^{n+1}+a_1x^n+a_2x^{n-1}+\ldots+a_{n-1}x^2+a_nx+a_{n+1},$$

который ортогонален с весом p любому многочлену $Q_m(x)$ $(m \leq n)$, имеет на отрезке $[a,b]$ $n+1$ различных корней.

Доказательство.

Принадлежность корней многочлена $G_{n+1}(x)$ отрезку $[a,b]$

От противного. Пусть отрезку $[a,b]$ принадлежат $s+1$ корней $x_0, x_1, x_2, \ldots, x_s$ $(s<n)$ корней многочлена $G_{n+1}(x)$. Этот многочлен можно представить в виде

$$G_{n+1}(x)=D_{s+1}(x)R_{n-s}(x),$$

где $D_{s+1}(x) = (x-x_0)(x-x_1)\ldots(x-x_s)$.

Доказательство...

Поскольку многочлен $G_{n+1}(x)$ ортогонален с весом p любому многочлену $Q_n(x)$, а $(s+1) \leq n$, то, с одной стороны,

$$\int_a^b p(x)G_{n+1}(x)D_{s+1}(x)dx = 0.$$

С другой стороны,

$$\int_a^b p(x)G_{n+1}(x)D_{s+1}(x)dx = \int_a^b p(x)\left(D_{s+1}(x)\right)^2 R_{n-s}(x)dx > 0,$$

поскольку

a) $p(x) \geq 0 \quad \forall x \in [a,b]$;

b) $\int_a^b p(x)dx > 0$;

c) $\left(D_{s+1}(x)\right)^2 \geq 0 \quad \forall x \in [a,b]$;

d) многочлен $R_{n-s}(x)$ на отрезке $[a,b]$ корней не имеет и, следовательно, для определенности $R_{n-s}(x) > 0 \quad \forall x \in [a,b]$.

Возникает противоречие.

Доказательство...

Все корни многочлена $G_{n+1}(x)$ простые

От противного. Пусть у многочлена $G_{n+1}(x)$ существует кратный корень x_s $(0 \le s \le n)$. Тогда $G_{n+1}(x)$ можно представить в виде

$$G_{n+1}(x) = (x - x_s)^2 R_{n-1}(x).$$

Аналогично, так как многочлен $G_{n+1}(x)$ ортогонален с весом p любому многочлену $Q_n(x)$, то, с одной стороны,

$$\int_a^b p(x) G_{n+1}(x) R_{n-1}(x) dx = 0.$$

С другой стороны,

$$\int_a^b p(x) G_{n+1}(x) R_{n-1}(x) dx = \int_a^b p(x) (x - x_s)^2 R_{n-1}(x) R_{n-1}(x) dx =$$

$$= \int_a^b p(x) (x - x_s)^2 (R_{n-1}(x))^2 dx > 0$$

в силу указанных выше свойств (8.56) весовой функции $p(x)$.
Возникает противоречие.
□ Теорема доказана.
⊠

▲18. Алгоритм построения квадратурной формулы алгебраической степени точности $N \geq 2n+1$

Пусть выполнены условия (8.57) (см. теорему 8.4). Требуется по $n+1$-у узлу построить квадратурную формулу (8.50) алгебраической степени точности $N \geq 2n+1$.

Численная процедура

- Строится неоднородная система (8.54) линейных алгебраических уравнений относительно неизвестных коэффициентов a_i $(i = 1, 2, \ldots, n+1)$ многочлена $\omega_n(x)$:

$$\int_a^b p(x)\left(a_1 x^n + a_2 x^{n-1} + \ldots + a_{n-1}x^2 + a_n x + a_{n+1}\right) x^m dx =$$

$$= -\int_a^b p(x) x^{n+m+1} dx \qquad \forall m = 0, 1, 2, \ldots, n$$

- Находится решение $\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_{n+1}$ системы (8.54) и определяется многочлен

$$\omega_n(x) = x^{n+1} + \bar{a}_1 x^n + \bar{a}_2 x^{n-1} + \ldots + \bar{a}_{n-1}x^2 + \bar{a}_n x + \bar{a}_{n+1};$$

Численная процедура...

- Вычисляются корни $\bar{x}_i$ $(i = 0, 1, 2, \ldots, n)$ многочлена $\omega_n(x)$;
- В силу интерполяционности искомой квадратурной формуры (см. теорему 8.3) по формулам (8.49) вычисляются ее коэффициенты $\bar{A}_k$ $(k = 0, 1, 2, \ldots, n)$:

$$\bar{A}_k = \int_a^b p(x)\Pi_{i=0, i\neq k}^n \frac{x - \bar{x}_i}{\bar{x}_k - \bar{x}_i} dx, \quad k = 0, 1, 2, \ldots, n;$$

- Строится квадратурная интерполяционная формула (8.50) алгебраической степени точности $N \geq 2n + 1$:

$$I[f] = \int_a^b p(x)f(x)dx \approx \sum_{k=0}^n \bar{A}_k f(\bar{x}_k).$$

Определение 8.7.

Квадратурная интерполяционная формула (8.50) алгебраической степени точности $N \geq 2n + 1$, построенная по $n + 1$-у узлу с помощью описанной выше численной процедуры, называется квадратурной формулой Гаусса.

ПРИМЕР

Постановка задачи

Пусть $p(x) \equiv 1$ $x \in [-1, 1]$, $n = 1$.
Требуется по 2-м узлам построить квадратурную формулу алгебраической степени точности $N \geq 3$.

1. Многочлен $\omega_1(x) = (x - x_0)(x - x_1) = x^2 + a_1 x + a_2$;
   Система линейных алгебраических уравнений относительно коэффициентов a_1 и a_2:
   $$\int_{-1}^{1} (a_1 x + a_2)\, x^m dx = -\int_{-1}^{1} x^{m+2} dx \qquad m = 0, 1$$
   или
   $$\begin{cases} 2a_2 = -\frac{2}{3} \\ \frac{2}{3} a_1 = 0 \end{cases}; \qquad (8.58)$$
2. Решение системы (8.58): $a_1 = 0$, $a_2 = -\frac{1}{3}$;
3. Многочлен $\omega_1(x) = x^2 - \frac{1}{3}$, корни которого $x_0 = -\frac{\sqrt{3}}{3}$, $x_1 = \frac{\sqrt{3}}{3}$;
4. В силу формул (8.49) $A_0 = A_1 = 1$;
5. Интерполяционная квадратурная формула:
   $\int_{-1}^{1} 1 \cdot f(x) dx \approx f(-\frac{\sqrt{3}}{3}) + f(\frac{\sqrt{3}}{3})$, $N = 3$.

Свойства квадратурной формулы Гаусса

Теорема 8.6.

Квадратурная формула Гаусса, построенная по $n+1$ узлам, имеет алгебраическую степень точности $N = 2n+1$.

Доказательство.
Пусть $Q_{2n+2}(x) = (\omega_n(x))^2$. Поскольку весовая функция p удовлетворяет условиям (8.57), то

$$I[Q_{2n+2}] = \int_a^b p(x)\,(\omega_n(x))^2 dx > 0.$$

При этом

$$S_n[Q_{2n+2}] = \sum_{k=0}^{n} A_k\,(\omega_n(x_k))^2 = 0.$$

Таким образом,

$$I[Q_{2n+2}] \neq S_n[Q_{2n+2}].$$

□ Теорема доказана.

Свойства квадратурной формулы Гаусса ...

Теорема 8.7.

Наивысшая алгебраическая степень точности квадратурной формулы, построенной по $n+1$ узлам, равна $2n+1$.

Доказательство. Пусть существует такая квадратурная формула, у которой $N > 2n+1$. Тогда, в силу теоремы 8.3, многочлен $\omega_n(x)$ должен удовлетворять условию ортогональности с весом p любому многочлену $Q_n(x)$. Согласно теореме 8.4, указанному условию удовлетворяет единственный многочлен

$$G_{n+1}(x)=x^{n+1}+a_1x^n+a_2x^{n-1}+\ldots+a_{n-1}x^2+a_nx+a_{n+1}.$$

Этот многочлен имеет единственный набор корней

$$x_0, x_1, x_2, \ldots, x_n \in [a,b]: \quad x_i \neq x_j \ \forall i \neq j,$$

которые являются узлами квадратурной формулы. Ее коэффициенты A_k $(k = 0, 1, 2, \ldots, n)$ вычисляются по формулам (8.49) единственным образом. Поэтому такая квадратурная формула совпадает с квадратурной формулой Гаусса, у которой, согласно теореме 8.6, $N = 2n+1$.

□ Теорема доказана.

Свойства квадратурной формулы Гаусса ...

Теорема 8.8.

Коэффициенты A_k $(k = 0, 1, 2, \ldots, n)$ квадратурной формулы Гаусса положительны.

Доказательство.

Пусть $Q_{2n}^{(k)}(x) = \left(\Pi_{i=0, i\neq k}^{n} \frac{x-x_i}{x_k-x_i}\right)^2$, $k = 0, 1, 2, \ldots, n$.

Очевидно, по построению, $\forall x \in [a, b]$ $Q_{2n}^{(k)}(x) \geq 0$ и

$$Q_{2n}^{(k)}(x_j) = \delta_{jk} = \begin{cases} 0, & j \neq k \\ 1, & j = k \end{cases} \quad \forall j = 0, 1, 2, \ldots, n$$

При этом, поскольку весовая функция p удовлетворяет условиям (8.57), то

$$I[Q_{2n}^{(k)}] = \int_a^b p(x) Q_{2n}^{(k)}(x) dx > 0.$$

Свойства квадратурной формулы Гаусса ...

Для всех многочленов $Q_{2n}^{(k)}(x)$ $(k = 0, 1, 2, \ldots, n)$ квадратурная формула Гаусса точна. Следовательно, $\forall k = 0, 1, 2, \ldots, n$

$$I[Q_{2n}^{(k)}] = \int_a^b p(x) Q_{2n}^{(k)}(x) dx = \sum_{j=0}^{n} A_j Q_{2n}^{(k)}(x_j) = A_k > 0.$$

□ Теорема доказана.

Теорема 8.9.

Квадратурный процесс Гаусса сходится.

Доказательство.
Поскольку у квадратурной формулы Гаусса $A_k > 0$ $(k = 0, 1, 2, \ldots, n)$ (см. теорему 8.8) и эта формула является интерполяционной (см. теорему 8.3), то справедливость утверждения теоремы следует из следствия из теоремы 8.2 (критерий сходимости интерполяционного квадратурного процесса).
□ Теорема доказана.

Погрешность квадратурной формулы Гаусса

Теорема 8.10.

Пусть $f \in C^{(2n+2)}([a,b])$. Тогда

$$R_n[f] = \frac{f^{(2n+2)}(\xi)}{(2n+2)!} \int_a^b p(x) \left(\omega_n(x)\right)^2 dx \tag{8.59}$$

и справедлива оценка

$$|R_n[f]| \le \frac{M_{2n+2}}{(2n+2)!} \int_a^b p(x) \left(\omega_n(x)\right)^2 dx, \tag{8.60}$$

где $\xi \in [a,b]$, $M_{2n+2} = \max_{x\in[a,b]} |f^{(2n+2)}(x)|$.

Доказательство.

Пусть $L_{2n+1}(x)$ — интерполяционный многочлен Эрмита, построенный для функции f по набору двукратных узлов $x_i \in [a,b]$ $(i = 0, 1, \ldots, n)$.

Доказательство...

Тогда

$$f(x) = L_{2n+1}(x) + f(x, x_0, x_0, \ldots, x_n, x_n) \left(\omega_n(x)\right)^2.$$

Следовательно,

$$R_n[f] = \int_a^b p(x) f(x, x_0, x_0, \ldots, x_n, x_n) \left(\omega_n(x)\right)^2 dx =$$

$$= f(\eta, x_0, x_0, \ldots, x_n, x_n) \int_a^b p(x) \left(\omega_n(x)\right)^2 dx =$$

$$= \frac{f^{(2n+2)}(\xi)}{(2n+2)!} \int_a^b p(x) \left(\omega_n(x)\right)^2 dx,$$

где $\eta \in [a, b]$ (по теореме о среднем), $\xi \in [a, b]$ (в силу свойств разделенных разностей с кратными узлами для гладких функций).

$\square$ Теорема доказана.

Вычисление определенных интегралов с особенностями

Под особенностью понимается ситуация, в которой стандартные приемы численного интегрирования приводят к неудовлетворительному результату[a].

[a] В частности, при вычислении несобственных интегралов.

Методы устранения особенностей

1. Аналитический;
2. Усечения;
3. Мультипликативный;
4. Аддитивный.

I. Аналитический метод устранения особенности

Пример

$$\int_0^{\frac{\pi}{2}} \ln(\sin(x))dx$$

Здесь $\lim_{x\to 0} \ln(\sin(x)) = -\infty$.

$$\int_0^{\frac{\pi}{2}} \ln(\sin(x))dx = \int_0^{\frac{\pi}{2}} \ln(\frac{\sin(x)}{x})dx + \int_0^{\frac{\pi}{2}} \ln(x)dx,$$

где $\int_0^{\frac{\pi}{2}} \ln(\frac{\sin(x)}{x})dx$ — интеграл без особенностей, а интеграл $\int_0^{\frac{\pi}{2}} \ln(x)dx$ может быть вычислен аналитически:

$$\int_0^{\frac{\pi}{2}} \ln(x)dx = \int_0^{\frac{\pi}{2}} x' \ln(x)dx = x\ln(x)|_0^{\frac{\pi}{2}} - \int_0^{\frac{\pi}{2}} x \cdot \frac{1}{x}dx =$$

$$= \frac{\pi}{2}\ln(\frac{\pi}{2}) - \lim_{x\to 0} x\ln(x) - x|_0^{\frac{\pi}{2}} = \frac{\pi}{2}\left(\ln(\frac{\pi}{2}) - 1\right),$$

где $\lim_{x\to 0} x\ln(x) = \lim_{x\to 0} \frac{\ln(x)}{\frac{1}{x}} = \lim_{x\to 0} \frac{\frac{1}{x}}{-\frac{1}{x^2}} = 0.$

II. Метод усечения

Пример

Требуется вычислить несобственный интеграл

$$\int_0^{\infty} \frac{1}{1+x^3} \arctan(x) dx$$

с заданной точностью $\varepsilon > 0$.

$$\int_0^{\infty} \frac{\arctan(x)}{1+x^3} dx = \int_0^{A} \frac{\arctan(x)}{1+x^3} dx + \int_A^{\infty} \frac{\arctan(x)}{1+x^3} dx.$$

$$\int_0^{\infty} \frac{1}{1+x^3} \arctan(x) dx \approx \int_0^{A} \frac{1}{1+x^3} \arctan(x) dx$$

$$A > 0 : \ \left| \int_A^{\infty} \frac{1}{1+x^3} \arctan(x) dx \right| \leq \varepsilon_1,$$

Здесь $\varepsilon = \varepsilon_1 + \varepsilon_2$, где $\varepsilon_1 > 0$ — погрешность метода усечения, $\varepsilon_2 > 0$ — погрешность вычисления интеграла $\int_0^{A} \frac{1}{1+x^3} \arctan(x) dx$.

Метод усечения. Пример...

Поскольку для любой константы $A > 0$ справедливо

$$\frac{1}{1+x^3}\arctan(x) > 0, \quad \arctan(x) < \frac{\pi}{2} \quad \forall x \in [A, \infty),$$

то

$$\left|\int_A^\infty \frac{1}{1+x^3}\arctan(x)dx\right| = \int_A^\infty \frac{1}{1+x^3}\arctan(x)dx \leq$$

$$\leq \frac{\pi}{2}\int_A^\infty \frac{1}{1+x^3}dx \leq$$

$$\leq \frac{\pi}{2}\int_A^\infty \frac{1}{x^3}dx =$$

$$= \frac{\pi}{2}\left(-\frac{1}{2x^2}\right)\Big|_A^\infty = \frac{\pi}{4A^2} \leq \varepsilon_1.$$

Откуда

$$A \geq \sqrt{\frac{\pi}{4\varepsilon_1}} > 0$$

III. Мультипликативный метод устранения особенности

$$\int_a^b p(x)f(x)dx,$$

где функция $f(x)$ без особенностей на $[a,b]$, а функция $p(x)$ обладает особенностью на $[a,b]$.

При этом у функции $p(x)$ существует аналитическая первообразная $s(x)$ ($x \in [a,b]$), которая обратима на $[a,b]$.

$$s(x) = \int_a^x p(t)dt \quad x \in [a,b]: \quad \exists r = s^{-1}: \; x = r(s) \quad s \in [s(a), s(b)].$$

Замена $s = s(x)$: $x = r(s)$, $ds = s'(x)dx = p(x)dx$

$$\int_a^b p(x)f(x)dx = \int_{s(a)}^{s(b)} f(r(s))ds \quad -$$

интеграл без особенностей.

Мультипликативный метод устранения особенности

Пример

$$\int_{-1}^{1} \frac{1}{\sqrt{1-x^4}} dx$$

$$\int_{-1}^{1} \frac{1}{\sqrt{1-x^4}} dx = \int_{-1}^{1} \frac{1}{\sqrt{1-x^2}} \frac{1}{\sqrt{1+x^2}} dx$$

Здесь $f(x) = \frac{1}{\sqrt{1+x^2}}$, $p(x) = \frac{1}{\sqrt{1-x^2}}$ $x \in [-1, 1]$:

$$s(x) = \int_0^x p(t)dt = \int_0^x \frac{1}{\sqrt{1-t^2}} dt = \arcsin(x) \quad x \in [-1, 1]$$

Замена $s = \arcsin(x)$: $x = \sin(s)$, $ds = s'(x)dx = \frac{1}{\sqrt{1-x^2}} dx$

$$\int_{-1}^{1} \frac{1}{\sqrt{1-x^4}} dx = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{\sqrt{1+\sin^2(s)}} ds \quad -$$

интеграл без особенностей.

IV. Аддитивный метод устранения особенности

$$\int_a^b f(x)dx,$$

где производные $f^{(k)}(x)$ не ограничены на $[a, b]$ для всех $k \geq K$.

Особенность

Пусть для приближенного вычисления интеграла требуется использовать метод, для оценки погрешности которого требуется ограниченность производной K-го порядка у подинтегральной функции.

$$f(x) = \varphi(x) + \psi(x),$$

где функция ψ без особенности, а функция φ с особенностью, но $I[\varphi]$ вычисляется аналитически.

Аддитивный метод устранения особенности...

ПРИМЕР

$$f(x) = (x-c)^{\alpha} g(x),$$

где $c \in [a,b]$, $-1 < \alpha < K$[a], функция g непрерывно-дифференцируема на $[a,b]$ достаточное число раз.

[a] Достаточное условие сходимости интеграла $I[f]$

$$g(x) = g(c) + g'(c)(x-c) + \ldots + \frac{g^{(K)}(c)}{K!}(x-c)^K + \\ + \left(g(x) - g(c) - g'(c)(x-c) - \ldots - \frac{g^{(K)}(c)}{K!}(x-c)^K \right),$$

$$f(x) = (x-c)^{\alpha} \left(g(c) + g'(c)(x-c) + \ldots + \frac{g^{(K)}(c)}{K!}(x-c)^K \right) + \\ + (x-c)^{\alpha} \left(g(x) - g(c) - g'(c)(x-c) - \ldots - \frac{g^{(K)}(c)}{K!}(x-c)^K \right) = \\ = \varphi(x) + \psi(x),$$

где

Аддитивный метод устранения особенности...

$$\varphi(x) = (x-c)^\alpha \left(g(c) + g'(c)(x-c) + \ldots + \frac{g^{(K)}(c)}{K!}(x-c)^K \right)$$
$$\psi(x) = (x-c)^\alpha \left(g(x) - g(c) - g'(c)(x-c) - \ldots - \frac{g^{(K)}(c)}{K!}(x-c)^K \right)$$

Здесь $I[\varphi]$ вычисляется аналитически, функцию ψ можно представить в виде

$$\psi(x) = (x-c)^\alpha g(x, \underbrace{c, c, \ldots, c}_{K+1})(x-c)^{K+1} = (x-c)^{\alpha+K+1} g(x, \underbrace{c, c, \ldots, c}_{K+1}) :$$

$\psi \in C^{(K)}([a,b])$.

В результате особенность сдвинута в производную $K+1$-го порядка.

$$I[f] \approx I[\varphi] + S_n[\psi],$$

где функция ψ имеет ограниченную на отрезке $[a,b]$ производную K-го порядка

$\boxtimes$

▲19.ТЕМА 9. Численные методы решения задачи Коши для обыкновенного дифференциального уравнения первого порядка

Постановка задачи Коши

$$y' = f(x, y), \quad x \in [x_0, x_0 + X], \tag{9.1}$$

$$y(x_0) = y_0, \tag{9.2}$$

где функция f непрерывна по каждому своему аргументу и имеет непрерывную частную производную f'_y [a].

[a] Достаточное условие существования и единственности классического решения $y = y(x|x_0, y_0)$ задачи Коши (9.1),(9.2), которое является непрерывно- дифференцируемой на отрезке $[x_0, x_0 + X]$ функцией.

Приближенное решение задачи Коши (9.1),(9.2)

$$x_0, x_1, x_2, \ldots, x_n \in [x_0, x_0 + X] : \; x_{i+1} = x_i + h_i \; \forall i = 0, 1, \ldots, n-1$$
$$y_0, y_1, y_2, \ldots, y_n : \; y_i \approx y(x_i) \; \forall i = 1, \ldots, n \tag{9.3}$$

9.1. Методы, основанные на разложении решения задачи Коши в ряд Тейлора

Пусть функция f такова, что у решения $y = y(x|x_0, y_0)$ задачи Коши (9.1), (9.2) существуют производные до достаточно большого порядка и
$x_i \in [x_0, x_0 + X]: \quad x_{i+1} = x_i + h \ \forall i = 0, 1, \ldots, n-1, \quad h = \frac{X}{n}$.
Разложение функции $y(x|x_0, y_0)$ в ряд Тейлора в окрестности узла x_0 имеет вид

$$y(x_0+h)=y(x_0)+y'(x_0)h+\frac{1}{2}y''(x_0)h^2+\ldots+\frac{1}{k!}y^{(k)}(x_0)h^k+O(h^{k+1}). \tag{9.4}$$

Равенство (9.4) для $k = 2$ можно записать в виде

$$\begin{aligned} y(x_0+h) = y(x_0) + f(x_0, y(x_0))h + \tfrac{h^2}{2}[f'_x(x_0, y(x_0))+ \\ +f'_y(x_0, y(x_0))f(x_0, y(x_0))] + O(h^3), \end{aligned} \tag{9.5}$$

где, в силу дифференциального уравнения (9.1),

$$\begin{aligned} y'(x_0) &= f(x_0, y(x_0)), \\ y''(x_0) &= f'_x(x_0, y(x_0)) + f'_y(x_0, y(x_0))y'(x_0) = \\ &= f'_x(x_0, y(x_0)) + f'_y(x_0, y(x_0))f(x_0, y(x_0)). \end{aligned}$$

Методы, основанные на разложении решения задачи Коши в ряд Тейлора ...

Отбрасывание из правой части равенства (9.5) слагаемого $O(h^3)$ приводит к приближенному равенству

$$y(x_1) \approx y(x_0) + f(x_0, y(x_0))h + \\ + \frac{h^2}{2}\left(f'_x(x_0, y(x_0)) + f'_y(x_0, y(x_0))f(x_0, y(x_0))\right), \tag{9.6}$$

которое связывает точные значения $y(x_0)$ и $y(x_1)$ искомого решения задачи Коши в соответствующих узлах.
Требование точности этого равенства для приближенных значений y_i и y_{i+1} в произвольных двух соседних узлах x_i и x_{i+1} $(0 \le i < n)$ приводит к реккурентной формуле

$$y_{i+1} = y_i + f(x_i, y_i)h + \frac{h^2}{2}\left(f'_x(x_i, y_i) + f'_y(x_i, y_i)f(x_i, y_i)\right) \\ \forall i = 0, 1, \ldots, n-1, \tag{9.7}$$

которая задает итерационный метод 3-го[33] порядка точности на одном шаге (итерации), так как $|y(x_1) - y_1|_{(9.7)} = O(h^3)$ при $i = 0$.

[33] Определяется степенью h в слагаемом $O(h^3)$.

9.2. Метод ломанных Эйлера

Для случая $k = 1$ в равенстве (9.4) аналогично (9.5)-(9.7) можно построить метод, имеющий второй[34] порядок точности на одном шаге

Явный метод Эйлера

$$y_{i+1} = y_i + f(x_i, y_i)h \quad \forall i = 0, 1, \ldots, n-1 \tag{9.8}$$

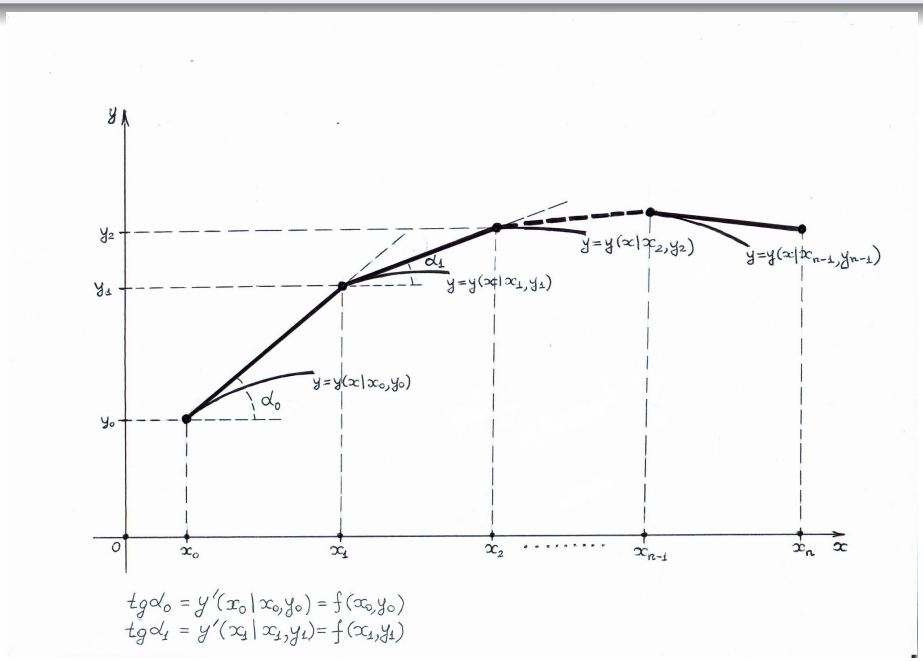


Рис. 16: Геометрическая интерпретация явного метода Эйлера (9.8).

[34] Поскольку $|y(x_1) - y_1|_{(9.8)} = O(h^2)$

Явный метод Эйлера...

Теорема 9.1.

Пусть функция f непрерывно-дифференцируема по каждому своему аргументу. Тогда погрешность явного метода Эйлера (9.8) на всем промежутке $[x_0, x_0 + X]$ является величиной первого порядка по h, то есть

$$\exists C > 0 : \quad |y(x_i) - y_i|_{(9.8)} \leq Ch \quad \forall i = 0, 1, 2, \ldots, n. \tag{9.9}$$

Доказательство.
Для точных значений $y(x_i)$ $(i = 0, 1, \ldots, n-1)$ решения задачи Коши (9.1),(9.2) справедливо следующее равенство

$$y(x_{i+1}) = y(x_i) + \int_{x_i}^{x_{i+1}} f(x, y(x))dx \quad \forall i = 0, 1, \ldots, n-1. \tag{9.10}$$

Разность соответствующих частей равенств (9.10) и (9.8) с учетом

$$h = \int_{x_i}^{x_{i+1}} 1dx$$

приводит к выражению

Доказательство ...

$$y(x_{i+1}) - y_{i+1} = y(x_i) - y_i + \int_{x_i}^{x_{i+1}} (f(x, y(x)) - f(x_i, y_i))dx =$$

$$= y(x_i) - y_i + \int_{x_i}^{x_{i+1}} (f(x, y(x)) - f(x_i, y(x_i)))dx +$$

$$+ \int_{x_i}^{x_{i+1}} (f(x_i, y(x_i)) - f(x_i, y_i))dx.$$

Откуда

$$|y(x_{i+1}) - y_{i+1}| \leq |y(x_i) - y_i| + \int_{x_i}^{x_{i+1}} |f(x, y(x)) - f(x_i, y(x_i))|dx +$$

$$+ \int_{x_i}^{x_{i+1}} |f(x_i, y(x_i)) - f(x_i, y_i)|dx. \tag{9.11}$$

Используя свойства функции $f(x, y)$, можно оценить каждый из определенных интегралов в правой части неравенства (9.11):

I) Пусть $F(x)=f(x, y(x))$ $x \in [x_0, x_0 + X]$. Поскольку функция $f(x, y)$ непрерывно-дифференцируема по всем своим аргументам, то функция $F(x)$ является непрерывно-дифференцируемой на отрезке $[x_0, x_0 + X]$.

Доказательство ...

Следовательно,

$$\exists K > 0: \quad |F'(x)| \le K \quad \forall x \in [x_0, x_0 + X],$$

где $F'(x) = f'_x(x, y(x)) + f'_y(x, y(x))f(x, y(x))$[35]. Тогда

$$\int_{x_i}^{x_{i+1}} |f(x, y(x)) - f(x_i, y(x_i))|dx = \int_{x_i}^{x_{i+1}} |F(x) - F(x_i)|dx =$$

$$= \int_{x_i}^{x_{i+1}} |F'(\xi(x))| \cdot |x - x_i|dx \le$$

$$\le K \int_{x_i}^{x_{i+1}} (x - x_i)dx = K\frac{h^2}{2}, \tag{9.12}$$

где $\xi(x) \in [x_i, x]$, $x \in [x_i, x_{i+1}]$.

II) В силу непрерывности функции $f'_y(x, y)$

$$\exists L > 0: \quad |f'_y(x, y)| \le L \quad \forall (x, y) \in [x_0, x_0 + X] \times \mathrm{H}[y_h],$$

[35] При $K = 0$ задача Коши (9.1),(9.2) имеет очевидное аналитическое решение $y(x) \equiv y_0$ ($x \in [x_0, x_0 + X]$). Этот случай здесь не рассматривается.

Доказательство ...

где $\mathrm{H}[y_h]$ — ограниченное замкнутое множество допустимых значений $y(x)$ и $y_i(h)$ ($x \in [x_0, x_0 + X]$, $h \in [0, X]$). Следовательно,

$$\int_{x_i}^{x_{i+1}} |f(x_i, y(x_i)) - f(x_i, y_i)|dx = \int_{x_i}^{x_{i+1}} |f'_y(x_i, \eta_i)| \cdot |y(x_i) - y_i|dx \leq$$

$$\leq L \int_{x_i}^{x_{i+1}} |y(x_i) - y_i|dx =$$

$$= Lh|y(x_i) - y_i|, \tag{9.13}$$

где $\eta_i \in [y_i, y(x_i)]$.

Пусть $\varepsilon_i = |y(x_i) - y_i|$, $i = 0, 1, 2, \ldots, n-1$.

Из (9.11), (9.12), (9.13) следует

$$\varepsilon_{i+1} \leq \varepsilon_i + K\frac{h^2}{2} + Lh\varepsilon_i = A(h)\varepsilon_i + B(h), \tag{9.14}$$

где $A(h) = 1 + Lh > 0$, $B(h) = K\frac{h^2}{2} \geq 0$ $\forall h \in [0, X]$.

Доказательство ...

Так как $y(x_0) = y_0$ (см. начальное условие (9.2)), то $\varepsilon_0 = 0$.
Тогда, в силу (9.14),

$$\varepsilon_1 \le A(h)\varepsilon_0 + B(h) = B(h),$$

$$\varepsilon_2 \le A(h)\varepsilon_1 + B(h) \le A(h)B(h) + B(h) = (A(h)+1)\,B(h),$$

$$\varepsilon_3 \le A(h)\varepsilon_2 + B(h) \le A(h)\,(A(h)+1)\,B(h) + B(h) = \\ = \left(A^2(h) + A(h) + 1\right) B(h),$$

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

$$\varepsilon_{i+1} \le A(h)\varepsilon_i + B(h) = \left(A^i(h) + A^{i-1}(h) + \ldots + A(h) + 1\right) B(h) = \\ = B(h) \cdot \frac{A^{i+1}(h)-1}{A(h)-1} \le \\ \le B(h) \cdot \frac{A^n(h)-1}{A(h)-1}$$

$$\forall i \le n-1. \qquad (9.15)$$

Доказательство ...

$$
\begin{aligned}
B(h)\cdot\frac{A^n(h)-1}{A(h)-1} &= \frac{Kh^2}{2}\cdot\frac{(1+Lh)^n-1}{Lh} = \\
&= \frac{Kh}{2L}\cdot\left(\left(1+L\frac{X}{n}\right)^n-1\right) = \\
&= \frac{K}{2L}\cdot\left(\left(\left(1+\frac{LX}{n}\right)^{\frac{n}{LX}}\right)^{LX}-1\right)h \le \\
&\le \frac{K}{2L}\cdot\left(e^{LX}-1\right)h = Ch,
\end{aligned}
\tag{9.16}
$$

где $C = \frac{K}{2L}\cdot\left(e^{LX}-1\right)$.
Таким образом, из (9.15) и (9.16) следует, что

$$\varepsilon_i = |y(x_i)-y_i|_{(9.8)} \le Ch \quad \forall i = 0,1,2,\ldots,n.$$

□ Теорема доказана.

Порядок точности явного метода Эйлера (9.8) на всем промежутке $[x_0, x_0+X]$ на единицу меньше его порядка точности на одном шаге.

Явный метод Эйлера...

Сходимость явного метода Эйлера (9.8)

Из (9.9) следует, что при $h \to 0$ функциональная последовательность сеточных функций $\{y_i(h)\}_{i=1}^n$, построенных в силу формулы (9.8) для различных значений параметра $0 < h \leq X$, равномерно на отрезке $[x_0, x_0 + X]$ сходится к точному решению $y = y(x|x_0, y_0)$ задачи Коши (9.1),(9.2).

Практическое преимущество явного метода Эйлера (9.8) перед другими численными процедурами типа (9.7), построенными на основе разложения решения задачи Коши в ряд Тейлора, заключается в следующем.

При численной реализации этого метода не требуется вычислять значения частных производных различных порядков функции $f(x, y)$ — правой части дифференциального уравнения (9.1).

Замечание

Для всех итерационных численных методов приближенного решения задачи Коши (9.1),(9.2), в результате реализации которых формируется числовая последовательность $\{y_i(h)\}_{i=1}^{n}$, определяющая приближенное решение указанной задачи, справедлива следующая закономерность.

Если погрешность такого метода на одном шаге является величиной $O(h^{k+1})$, то его погрешность на всем промежутке $[x_0, x_0 + X]$ будет величиной $O(h^k)$.

9.3. Методы, основанные на интегральном представлении решения задачи Коши

Решение задачи Коши (9.1),(9.2) может быть представлено в виде

$$y(x) = y(x_0) + \int_{x_0}^{x} f(\tau, y(\tau))d\tau \quad \forall x \in [x_0, x_0 + X] \tag{9.17}$$

Из (9.17) следует, что для любого узла x_k $(k = 0, 1, 2, \ldots, n-1)$ справедливо равенство

$$y(x_{k+1}) = y(x_k) + \int_{x_k}^{x_{k+1}} F(x)dx, \tag{9.18}$$

где $F(x) = f(x, y(x)) \quad x \in [x_0, x_0 + X]$.
Для вычисления приближенного значения интеграла $I[F|[x_k, x_{k+1}]]$ в правой части равенства (9.18) может быть использована любая формула численного интегрирования. Тогда замена в (9.18) этого интеграла квадратурной суммой $S_m[F|[x_k, x_{k+1}]]$ приводит к приближенному равенству

Методы, основанные на интегральном представлении решения задачи Коши ...

$$y(x_{k+1}) \approx y(x_k) + S_m[F|[x_k, x_{k+1}]]$$

или

$$y(x_{k+1}) \approx y(x_k) + \sum_{i=0}^{m} \bar{A}_i f(\bar{x}_i, y(\bar{x}_i)),$$

где $\bar{x}_i \in [x_k, x_{k+1}]$ $(i = 0, 1, \ldots, m)$. Требование точности этого равенства для приближенных значений функции y в соответствующих узлах приводит к следующей численной процедуре

$$y_{k+1} = y_k + \sum_{i=0}^{m} \bar{A}_i f(\bar{x}_i, \bar{y}_i) \qquad (9.19)$$

$$k = 0, 1, 2, \ldots, n-1$$

В частности, формула (9.8) явного метод Эйлера может быть получена с помощью этой методики, если для вычисления приближенного значения интеграла $I[F|[x_k, x_{k+1}]]$ в (9.18) воспользоваться формулой левых прямоугольников (8.8).

9.4. Метод Эйлера с пересчетом

В случае, если для вычисления приближенного значения интеграла $I[F|[x_k, x_{k+1}]]$ в (9.18) используется формула трапеций (8.9), то численная процедура (9.19) задается следующей неявной формулой

$$y_{k+1} = y_k + \frac{h}{2}[f(x_k, y_k) + f(x_{k+1}, y_{k+1})] \qquad (9.20)$$
$$k = 0, 1, 2, \ldots, n-1$$

На практике для преодоления вычислительных трудностей, обусловленных наличием искомой величины y_{k+1} в правой части равенства (9.20) в качестве аргумента функции f, предлагается использовать следующую численную схему

Численная процедура метода Эйлера с пересчетом

$$\bar{y}_{k+1} = y_k + hf(x_k, y_k)$$
$$y_{k+1} = y_k + \frac{h}{2}[f(x_k, y_k) + f(x_{k+1}, \bar{y}_{k+1})] \qquad (9.21)$$
$$k = 0, 1, 2, \ldots, n-1$$

Погрешность метода Эйлера с пересчетом

Для определения порядка точности метода Эйлера с пересчетом (9.21) на одном шаге используется следующее разложение функции $y(x|x_0, y_0)$ в ряд Тейлора в окрестности узла x_0

$$\begin{aligned} y(x_1) = y(x_0) + f(x_i, y(x_0))h + \\ + \frac{h^2}{2}\left(f'_x(x_0, y(x_0)) + f'_y(x_0, y(x_0))f(x_0, y(x_0))\right) + \\ + O(h^3), \end{aligned} \tag{9.22}$$

В силу (9.21), справедливо

$$\begin{aligned} y_1 &= y_0 + \frac{h}{2}[f(x_0, y_0) + f(x_1, \bar{y}_1)] = \\ &= y_0 + \frac{h}{2}[f(x_0, y_0) + f(x_0 + h, y_0 + hf(x_0, y_0))] = \\ \\ &= y_0 + \frac{h}{2}f(x_0, y_0) + \\ &\quad + \frac{h}{2}[f(x_0, y_0) + f'_x(x_0, y_0)h + f'_y(x_0, y_0)hf(x_0, y_0) + \bar{O}(h^2)] \\ \\ &= y_0 + hf(x_0, y_0) + \\ &\quad + \frac{h}{2}[f'_x(x_0, y_0)h + f'_y(x_0, y_0)hf(x_0, y_0)] + \bar{O}(h^3). \end{aligned} \tag{9.23}$$

Погрешность метода Эйлера с пересчетом ...

Нетрудно убедиться, что с учетом начального условия (9.1) ($y(x_0) = y_0$) разность соответствующих левых и правых частей выражений (9.22) и (9.23) приводит к равенству

$$y(x_1) - y_1 = O(h^3) - \bar{O}(h^3) = \tilde{O}(h^3) \qquad (9.24)$$

Вывод

Из (9.24) следует, что погрешность метода Эйлера с пересчетом (9.21) на одном шаге является величиной $O(h^3)$, то есть этот метод имеет третий порядок точность на одном шаге. В силу выше сделанного Замечания, погрешность метода Эйлера с пересчетом (9.21) на всем промежутке $[x_0, x_0 + X]$ будет величиной $O(h^2)$.

[36] Самостоятельно представить геометрическую интерпретацию метода Эйлера с пересчетом (9.21).

9.5. Метод Коши

Если для вычисления приближенного значения интеграла $I[F|[x_k, x_{k+1}]]$ в (9.18) применяется формула средних прямоугольников (8.8), то численная процедура (9.19) задается формулой

$$y_{k+1} = y_k + hf(x_{k+\frac{1}{2}}, y_{k+\frac{1}{2}}) \quad k = 0, 1, 2, \ldots, n-1, \tag{9.25}$$

где $x_{k+\frac{1}{2}} = x_k + \frac{h}{2}$, $y_{k+\frac{1}{2}} \approx y(x_k + \frac{h}{2})$

Численная процедура метода Коши

$$\bar{y}_{k+\frac{1}{2}} = y_k + \frac{h}{2} f(x_k, y_k)$$

$$y_{k+1} = y_k + hf(x_{k+\frac{1}{2}}, \bar{y}_{k+\frac{1}{2}}) \tag{9.26}$$

$$k = 0, 1, 2, \ldots, n-1$$

a

---

[a] Самостоятельно доказать, что метод Коши (9.26) имеет третий порядок точности на одном шаге и представить его геометрическую интерпретацию.

▲20. 9.6. Семейство явных методов Рунге-Кутты второго порядка точности на всем промежутке

Предлагается построить семейство методов, в которых на каждом шаге требуется последовательно вычислять два значения правой части $f(x, y)$ дифференциального уравнения (9.1).

$$
\begin{aligned}
&y_{i+1} = y_i + p_1 K_1 + p_2 K_2, \\
&K_1 = hf(x_i, y_i), \\
&K_2 = hf(x_i + \alpha h, y_i + \beta K_1), \\
&i = 0, 1, 2, \ldots, n-1,
\end{aligned}
\tag{9.27}
$$

где $p_1, p_2, \alpha, \beta \in \mathbb{R}$.

Семейство (9.27) содержит четыре параметра: p_1, p_2, α, β. В частности, для явного метода Эйлера (9.8) $p_1 = 1, p_2 = 0$, для метода Эйлера с пересчетом (9.21) $p_1 = \frac{1}{2}, p_2 = \frac{1}{2}, \alpha = 1, \beta = 1$, для метода Коши (9.26) $p_1 = 0, p_2 = 1, \alpha = \frac{1}{2}, \beta = \frac{1}{2}$.

Семейство явных методов Рунге-Кутты второго порядка точности на всем промежутке ...

В (9.27) требуется определить значения параметров p_1, p_2, α, β так, чтобы любой метод из семейства (9.27) имел второй порядок точности на всем промежутке $[x_0, x_0 + X]$. Для этого искомые значения указанных параметров должны обеспечивать третий порядок точности этих методов на одном шаге.

Как было ранее показано (см. (9.5), разложение точного решения $y(x|x_0, y_0)$ задачи Коши (9.1), (9.2) в ряд Тейлора в окрестности узла x_0 можно записать следующим образом

$$y(x_0+h) = y(x_0) + f(x_0, y(x_0))h + \frac{h^2}{2}[f'_x(x_0, y(x_0)) + \\ + f'_y(x_0, y(x_0))f(x_0, y(x_0))] + O(h^3), \tag{9.28}$$

В силу (9.27), справедливо равенство

$$y_1 = y_0 + p_1 h f(x_0, y_0) + p_2 h f\left(x_0 + \alpha h, y_0 + \beta h f(x_0, y_0)\right). \tag{9.29}$$

Требуемое разложение функции $f(x, y)$ в ряд Тейлора в окрестности точки (x_0, y_0) имеет вид

Семейство явных методов Рунге-Кутты второго порядка точности на всем промежутке ...

$$f\left(x_0+\alpha h, y_0+\beta h f(x_0,y_0)\right) = f(x_0,y_0) + f'_x(x_0,y_0)\alpha h + \\ + f'_y(x_0,y_0)\beta h f(x_0,y_0) + O(h^2) \tag{9.30}$$

Подстановка правой части равенства (9.30) в (9.29) приводит к выражению

$$y_1 = y_0 + p_1 h f(x_0,y_0) + p_2 h f\left(x_0+\alpha h, y_0+\beta h f(x_0,y_0)\right) =$$

$$= y_0 + p_1 h f(x_0,y_0) + p_2 h[f(x_0,y_0) + f'_x(x_0,y_0)\alpha h + \\ + f'_y(x_0,y_0) f(x_0,y_0)\beta h + O(h^2)] = \tag{9.31}$$

$$= y_0 + (p_1+p_2)\, h f(x_0,y_0) + p_2\alpha h^2 f'_x(x_0,y_0) + \\ + p_2\beta h^2 f'_y(x_0,y_0) f(x_0,y_0) + \widetilde{O}(h^3).$$

Тогда равенство разностей соответствующих частей выражений (9.28) и (9.31) с учетом начального условия (9.2) ($y(x_0)=y_0$) имеет вид

Семейство явных методов Рунге-Кутты второго порядка точности на всем промежутке ...

$$y(x_1)-y_1=(1-p_1-p_2)\,hf(x_0,y_0)+\left(\frac{1}{2}-p_2\alpha\right)h^2f'_x(x_0,y_0)+ \\ +\left(\frac{1}{2}-p_2\beta\right)h^2f'_y(x_0,y_0)f(x_0,y_0)+O(h^3). \tag{9.32}$$

Из (9.32) следует, что для того чтобы $y(x_1)-y_1=O(h^3)$ значения параметров p_1, p_2, α, β должны удовлетворять системе уравнений

$$\begin{cases} p_1+p_2=1 \\ p_2\alpha=\frac{1}{2} \\ p_2\beta=\frac{1}{2} \end{cases} \tag{9.33}$$

Очевидно, что система (9.33) совместна и имеет однопараметрическое семейство решений.

Таким образом, (9.27), (9.33) задают однопараметрическое семейство явных методов Рунге-Кутты второго порядка точности на всем промежутке $[x_0, x_0+X]$.

9.7. Семейство явных методов Рунге-Кутты четвертого порядка точности на всем промежутке

$$y_{i+1} = y_i + p_1 K_1 + p_2 K_2 + p_3 K_3 + p_4 K_4,$$

$$\begin{aligned} K_1 &= hf(x_i, y_i), \\ K_2 &= hf(x_i + \alpha_1 h, y_i + \beta_1 K_1), \\ K_3 &= hf(x_i + \alpha_2 h, y_i + \beta_2 K_1 + \beta_3 K_2), \\ K_4 &= hf(x_i + \alpha_3 h, y_i + \beta_4 K_1 + \beta_5 K_2 + \beta_6 K_3), \end{aligned} \tag{9.34}$$

$$i = 0, 1, 2, \ldots, n-1,$$

где $p_i, \alpha_j, \beta_s \in \mathbb{R}$ $(i = 1, 2, 3, 4;\ j = 1, 2, 3;\ s = 1, 2, 3, 4, 5, 6)$.

Семейство методов (9.34) содержит тринадцать параметров: $p_i, \alpha_j, \beta_s \in \mathbb{R}$ $(i = 1, 2, 3, 4;\ j = 1, 2, 3;\ s = 1, 2, 3, 4, 5, 6)$.

Требование того, чтобы методы (9.34) обладали пятым порядком точности на одном шаге приводит к совместной системе из одиннадцати уравнений, которая имеет двупараметрическое семейство решений.

Семейство явных методов Рунге-Кутты четвертого порядка точности на всем промежутке...

Метод Рунге-Кутты четвертого порядка точности на всем промежутке

$$
\begin{aligned}
y_{i+1} &= y_i + \tfrac{1}{6}\left(K_1 + 2K_2 + 2K_3 + K_4\right), \\
K_1 &= hf(x_i, y_i), \\
K_2 &= hf(x_i + \tfrac{h}{2}, y_i + \tfrac{K_1}{2}), \\
K_3 &= hf(x_i + \tfrac{h}{2}, y_i + \tfrac{K_2}{2}), \\
K_4 &= hf(x_i + h, y_i + K_3), \\
i &= 0, 1, 2, \ldots, n-1.
\end{aligned}
\tag{9.35}
$$

Методы Рунге-Кутты. Заключительные замечания.

Метод Рунге-Кутты (9.35) на каждом шаге требует четырех вычислений значений правой части $f(x, y)$ дифференциального уравнения (9.1), но так как он четвертого порядка точности на всем промежутке $[x_0, x_0 + X]$, то там, где явный метод Эйлера (9.8) для достижения определенной точности требует 10000 вычислений значений правой части дифференциального уравнения, метод Рунге-Кутты (9.35) — только 40, что ведет к уменьшению вычислительной погрешности.

Барьер Бутчера

При пяти вычислениях правой части $f(x, y)$ дифференциального уравнения (9.1) не существует явного метода Рунге-Кутты пятого порядка точности на всем промежутке $[x_0, x_0 + X]$, этот факт называется барьером Бутчера. Для построения такого метода нужно как минимум шесть вычислений значений правой части.

9.8. Методы Адамса

Общая идея

На каждом (m+1-ом) шаге метода Адамса используются значения правой части $f(x, y)$ дифференциального уравнения (9.1) в текущем и предыдущих узлах x_{m-j} $(j = 0, 1, \ldots, k-1)$. Его численная процедура строится на основе интегрального представления точного решения задачи Коши (9.1),(9.2)

$$y(x_{m+1}) = y(x_m) + \int_{x_m}^{x_{m+1}} F(x)dx, \tag{9.36}$$

где $F(x) = f(x, y(x))$ $\;x \in [x_0, x_0 + X]$.

Для функции F на основе данных $(x_{m-j}, y(x_{m-j}))$ $j = 0, 1, \ldots, k-1$ строится интерполяционный многочлен Лагранжа $L_{k-1}(x)$. Затем в определенном интеграле в (9.36) функция F заменяется этим многочленом. Полученное приближенное равенство заменяется на точное для приближенных значений $y_{m+1} \approx y(x_{m+1})$, $y_{m-j} \approx y(x_{m-j})$ $(j = 0, 1, \ldots, k-1)$ решения задачи Коши в соответствующих узлах.

Явные (экстраполяционные) методы Адамса

Пусть $f_{m-j} = f(x_{m-j}, y_{m-j})$ $(j = 0, 1, \ldots, k-1)$. По этим данным строится интерполяционный многочлен Лагранжа $L_{k-1}(x)$ степени $k-1$.

Определение 9.1.

Явным методом Адамса k-го порядка называется метод

$$y_{m+1} = y_m + \int_{x_m}^{x_{m+1}} L_{k-1}(x)dx, \qquad m = k-1, k, \ldots, n-1. \tag{9.37}$$

Порядок точности явного метода Адамса k-го порядка на одном шаге: $y_{m-j} = y(x_{m-j})$ $(j = 0, 1, \ldots, k-1)$

$$O(h^{k+1}) \sim \frac{F^{(k)}(\xi_m)}{k!} \int_{x_m}^{x_{m+1}} (x - x_m)(x - x_{m-1}) \ldots (x - x_{m-k+1})dx, \tag{9.38}$$

где $\xi_m \in [x_{m-k+1}, x_{m+1}]$

Примеры явных методов Адамса

$k = 1$. Явный метод Эйлера. $O(h^2)$

$$L_0(x) = F(x_m) = f_m,$$

$$y_{m+1} = y_m + \int_{x_m}^{x_{m+1}} L_0(x)dx = y_m + f_m h,$$

$$m = 0, 1, \ldots, n-1$$

$k = 2$. $O(h^3)$

$$L_1(x) = F(x_m) + F(x_m, x_{m-1})(x - x_m) = f_m + \frac{f_m - f_{m-1}}{h}(x - x_m),$$

$$y_{m+1} = y_m + \int_{x_m}^{x_{m+1}} L_1(x)dx = y_m + \frac{h}{2}\left(3f_m - f_{m-1}\right),$$

$$m = 1, 2, \ldots, n-1$$

Примеры явных методов Адамса ...

$k = 3$. Метод Адамса-Бэшфорта. $O(h^4)$

$$L_2(x) = F(x_m) + F(x_m, x_{m-1})(x - x_m) + \\ + F(x_m, x_{m-1}, x_{m-2})(x - x_m)(x - x_{m-1}) =$$

$$= f_m + \frac{f_m - f_{m-1}}{h}(x - x_m) +$$

$$+ \frac{f_m - 2f_{m-1} + f_{m-2}}{2h^2}(x - x_m)(x - x_{m-1}),$$

$$y_{m+1} = y_m + \int_{x_m}^{x_{m+1}} L_2(x)dx =$$

$$= y_m + \frac{h}{12}\left(23f_m - 16f_{m-1} + 5f_{m-2}\right),$$

$$m = 2, 3, \ldots, n - 1$$

Неявные (интерполяционные) методы Адамса

Формально строится интерполяционный многочлен Лагранжа $L_k(x)$ степени k. Для этого используются данные:
$f_{m+1} = f(x_{m+1}, y_{m+1})$ и $f_{m-j} = f(x_{m-j}, y_{m-j})$ $(j = 0, 1, \ldots, k-1)$.

Определение 9.2.

Неявным методом Адамса k-го порядка называется метод

$$y_{m+1} = y_m + \int_{x_m}^{x_{m+1}} L_k(x)dx, \qquad m = \{k-1\}k, \ldots, n-1. \tag{9.39}$$

Порядок точности неявного метода Адамса k-го порядка на одном шаге: $y_{m-j} = y(x_{m-j})$ $(j = 0, 1, \ldots, k-1)$

$$O(h^{k+2}) \sim \frac{F^{(k+1)}(\xi_m)}{(k+1)!} \int_{x_m}^{x_{m+1}} (x - x_{m+1})(x - x_m)\ldots(x - x_{m-k+1})dx, \tag{9.40}$$

где $\xi_m \in [x_{m-k+1}, x_{m+1}]$

Примеры неявных методов Адамса

$k = 0$. Неявный метод Эйлера. $O(h^2)$

$$L_0(x) = F(x_{m+1}) = f_{m+1},$$

$$y_{m+1} = y_m + \int_{x_m}^{x_{m+1}} L_0(x)dx = y_m + f_{m+1}h,$$

$$m = 0, 1, \ldots, n-1$$

$k = 1$. $O(h^3)$

$$L_1(x) = F(x_{m+1}) + F(x_{m+1}, x_m)(x - x_{m+1}) =$$

$$= f_{m+1} + \frac{f_{m+1} - f_m}{h}(x - x_{m+1}),$$

$$y_{m+1} = y_m + \int_{x_m}^{x_{m+1}} L_1(x)dx = y_m + \frac{h}{2}\left(f_{m+1} + f_m\right),$$

$$m = 0, 1, 2, \ldots, n-1$$

Примеры неявных методов Адамса ...

$k = 2$. Метод Адамса-Мултона. $O(h^4)$

$$L_2(x) = F(x_{m+1}) + F(x_{m+1}, x_m)(x - x_{m+1}) + \\ + F(x_{m+1}, x_m, x_{m-1})(x - x_{m+1})(x - x_m) =$$

$$= f_{m+1} + \frac{f_{m+1} - f_m}{h}(x - x_{m+1}) +$$

$$+ \frac{f_{m+1} - 2f_m + f_{m-1}}{2h^2}(x - x_{m+1})(x - x_m),$$

$$y_{m+1} = y_m + \int_{x_m}^{x_{m+1}} L_2(x)dx =$$

$$= y_m + \frac{h}{12}\left(5f_{m+1} + 8f_m - f_{m-1}\right),$$

$$m = 1, 2, \ldots, n - 1$$

Методы Адамса. Заключительные замечания.

Разгон

Для всех методов Адамса, начиная с двухшаговых, кроме начального значения y_0 требуется знание стартовых значений $y_1, \ldots, y_{k-1}$. Процедура вычисления стартовых значений называется разгоном. Рекомендуется делать разгон методами (например, Рунге-Кутты) не меньшего порядка точности на одном шаге, а лучше на единицу большего.

Сравнивая явные и неявные методы Адамса, следует отметить:

1. Недостаток неявных методов состоит в необходимости на каждом шаге решать уравнение относительно неизвестной величины y_{m+1}.
2. Некоторое преимущество неявных методов состоит в точности: при одном и том же порядке k неявные методы имеют порядок точности на одном шаге $k+2$, в отличие от явных, у которых порядок точности на одном шаге $k+1$.
3. Главное преимущество неявных методов состоит в возможности решать жесткие системы.

9.9. Общий класс многошаговых методов

Определение 9.3.

k-шаговым разностным методом называется

$$y_m = \sum_{i=1}^{k} \alpha_i y_{m-i} + h \sum_{j=0}^{k} \beta_j f_{m-j}, \qquad m = k, \ldots, n-1, \tag{9.41}$$

где $\alpha_i \in \mathbb{R}$ $(i = 0, 1, \ldots, k)$; $\beta_j \in \mathbb{R}$, $f_{m-j} = f(x_{m-j}, y_{m-j})$ $(j = 0, 1, \ldots, k)$.

Если $\beta_0 = 0$, то метод (9.41) называется явным.
Если $\beta_0 \neq 0$, то метод (9.41) называется неявным.

Определение 9.4.

Невязкой метода (9.41) на одном шаге называется величина

$$z_m = y(x_m) - \sum_{i=1}^{k} \alpha_i y(x_{m-i}) - h \sum_{j=0}^{k} \beta_j f(x_{m-j}, y(x_{m-j})) \tag{9.42}$$

k-шаговые разностные методы

Определение 9.5.

Пусть $y_{m-i} = y(x_{m-i})$ $(j = 1, \ldots, k)$. Погрешностью метода (9.41) на одном шаге (локальной погрешностью) называется величина

$$r_m = y(x_m) - y_m|_{(9.41)} = y(x_m) - \sum_{i=1}^{k} \alpha_i y_{m-i} - h \sum_{j=0}^{k} \beta_j f_{m-j} \quad (9.43)$$

Теорема 9.2.

Невязка z_m и погрешность r_m на одном шаге метода (9.41) — величины одного порядка по h.

Доказательство.
Если $\beta_0 = 0$, то $z_m = r_m$.
Пусть $\beta_0 \neq 0$ и функция $f(x, y)$ непрерывно-дифференцируема по второму аргументу.

Доказательство ...

Поскольку

$$y(x_m) = z_m + \sum_{i=1}^{k} \alpha_i y(x_{m-i}) + h\beta_0 f(x_m, y(x_m)) + \\ + h\sum_{j=1}^{k} \beta_j f(x_{m-j}, y(x_{m-j})),$$

$$y_m = \sum_{i=1}^{k} \alpha_i y(x_{m-i}) + h\beta_0 f(x_m, y_m) + \\ + h\sum_{j=1}^{k} \beta_j f(x_{m-j}, y(x_{m-j})),$$

то

$$r_m = z_m + h\beta_0 \left(f(x_m, y(x_m)) - f(x_m, y_m)\right) = z_m + h\beta_0 f'_y(x_m, \xi_m) r_m,$$

где $\xi_m \in [y_m, y(x_m)]$. Откуда

$$z_m = \left(1 - h\beta_0 f'_y(x_m, \xi_m)\right) r_m. \qquad (9.44)$$

Так как $0 \le h \le X$ и $\exists D \ge 0 : \ |f'_y(x, y)| \le D$, то из (9.44) следует, что

$$\exists 0 < C_1 \le C_2 : \ C_1|r_m| \le |z_m| \le C_2|r_m|.$$

□ Теорема доказана.

k-шаговые разностные методы ...

Определение 9.6.

Неотрицательное целое число N называется алгебраической степенью точности формулы (9.41), если:
1) формула (9.41) точна для всех многочленов степени N и ниже;
2) среди многочленов степени $N+1$ найдется хотя бы один, для которого формула (9.41) неточна.

Теорема 9.3.

Пусть $N \geq 0$ — алгебраическая степень точности формулы (9.41). Тогда невязка метода (9.41) на одном шаге $z_m = O(h^{N+1})$.

Класс явных k-шаговых методов имеет $2k$ параметров, а класс неявных k-шаговых методов имеет $2k+1$ параметров. Поэтому за счет выбора этих параметров можно получить локальную погрешность наибольшего порядка малости.
Подбор параметров следует проводить так, чтобы обеспечить формуле метода (9.41) наибольшую алгебраическую степень точности.

$\boxtimes$

▲21. Пример k-шаговых разностных методов

Класс явных двухшаговых методов

$$y_m = \alpha_1 y_{m-1} + \alpha_2 y_{m-2} + h\left(\beta_1 f_{m-1} + \beta_2 f_{m-2}\right) \qquad (9.45)$$

Требуется на классе методов (9.45) построить метод с наивысшей алгебраической степенью точности.

Последовательная подстановка в (9.45) вместо функции y многочленов[37]

$$Q_0(x) \equiv 1, Q_1(x) = x - x_{m-1}, Q_2(x) = (x - x_{m-1})^2, Q_3(x) = (x - x_{m-1})^3,$$

то есть требование точности формулы (9.45) для указанных многочленов, приводит к системе уравнений

$$\begin{cases} 1 = \alpha_1 + \alpha_2 \\ h = -\alpha_2 h + h(\beta_1 + \beta_2) \\ h^2 = \alpha_2 h^2 + h(-2h\beta_2) \\ h^3 = -\alpha_2 h^3 + h(3h^2\beta_2) \end{cases}$$

[37] Здесь следует учитывать, что $f_{m-i} = f(x_{m-i}, y_{m-i}) = y'_{m-i}$.

Пример...

Решением этой системы является

$$\begin{cases} \alpha_1 = -4 \\ \alpha_2 = 5 \\ \beta_1 = 4 \\ \beta_2 = 2 \end{cases}$$

подстановка которого в (9.45) приводит к методу

$$y_m = -4y_{m-1} + 5y_{m-2} + h\left(4f_{m-1} + 2f_{m-2}\right), \qquad (9.46)$$

который имеет алгебраическую степень точности $N = 3$ (как нетрудно проверить, для многочлена $Q_4(x) = (x - x_{m-1})^4$ формула (9.46) неточна).

Согласно теоремам 9.2 и 9.3, метод (9.46) имеет невязку (локальную погрешность) $O(h^4)$, которая на порядок выше явного двухшагового метода Адамса из того же класса методов.

Однако метод (9.46), несмотря на его достаточно высокий порядок точности, расходится. В этом можно убедиться на тестовом примере.

Пример...

Тестовая задача Коши

$$y' = 0, \quad x \in [0, X], \tag{9.47}$$

$$y(0) = y_0 = 0. \tag{9.48}$$

Задача (9.47), (9.48) имеет тривиальное решение

$$y(x) \equiv 0 \quad x \in [0, X].$$

Метод (9.46) для дифференциального уравнения (9.47) имеет вид

$$y_m = -4y_{m-1} + 5y_{m-2}. \tag{9.49}$$

Чтобы воспользоваться этим методом для приближенного решения задачи (9.47), (9.48) необходимо выполнить его разгон, то есть определить y_1. Реализация разгона с помощью любого численного метода приводит к тому, что y_1 вычисляется с некоторой малой ошибкой $y_1 = \varepsilon > 0$. В этом случае дальнейшее использование метода (9.49) приводит к следующему результату:

$$y_2 = -4\varepsilon, y_3 = 21\varepsilon, y_4 = 104\varepsilon, y_5 = 521\varepsilon, \ldots \tag{9.50}$$

Пример...

Как видно из (9.50), при использовании метода (9.49) от шага к шагу происходит достаточно быстрое увеличение величины ошибки $\varepsilon_m = |y(x_m) - y_m|$, $(m = 0, 1, 2, \ldots)$:

$$\varepsilon_0 = 0, \quad \varepsilon_1 = \varepsilon, \quad \varepsilon_2 = 4\varepsilon, \quad \varepsilon_3 = 21\varepsilon, \quad \varepsilon_4 = 104\varepsilon, \quad \varepsilon_5 = 521\varepsilon, \ldots,$$

что свидетельствует о расходимости метода (9.49).
Такое явление называется неустойчивостью метода:

Неустойчивость k-шагового метода по начальным условиям

Малая ошибка в начальных условиях $y_1, y_2, \ldots, y_{k-1}$, для k-шагового метода приводит к ее увеличению в приближенных значениях y_m $(m \geq k)$ решения задачи Коши на его последующих итерациях.

Чтобы получить условия, которые гарантируют устойчивость и в конечном итоге — сходимость многошаговых методов, требуется рассмотреть элементы теории линейных разностных уравнений.

9.10. Линейные разностные уравнения и их устойчивость

Линейное однородное разностное уравнение k-го порядка с постоянными коэффициентами

$$y_m + a_1 y_{m-1} + a_2 y_{m-2} + \ldots + a_k y_{m-k} = 0, \qquad (9.51)$$

где $a_i \in \mathbb{R}$ $(i = 1, 2, \ldots, k)$.

Пример: Линейное однородное разностное уравнение 2-го порядка

$$y_m + 4y_{m-1} - 5y_{m-2} = 0.$$

Если заданы начальные условия — числа $y_0, y_1, \ldots, y_{k-1}$, то следующие члены последовательности $y_k, y_{k+1}, \ldots$ определяются по явной формуле

$$y_m = -\left(a_1 y_{m-1} + a_2 y_{m-2} + \ldots + a_k y_{m-k}\right)$$

и решением уравнения (9.51) является числовая последовательность $\{y_m\}_{m=0}^{\infty}$.

Линейные разностные уравнения...

Свойства решений линейного разностного уравнения

1. Среди решений уравнения (9.51) есть тривиальное $\{y_m = 0\}_{m=0}^{\infty}$;
2. Любая линейная комбинация решений уравнения (9.51) является его решением;
3. Общее решение уравнения (9.51) представляет собой линейную комбинацию его линейно-независимых решений, образующих фундаментальную систему решений.

Для нахождения фундаментальной системы решений уравнения (9.51) предлагается искать его линейно-независимые решения в виде $y_s = \rho^s$. Подстановка $y_m = \rho^m$ в (9.51) приводит к уравнению

$$\rho^m + a_1\rho^{m-1} + a_2\rho^{m-2} + \ldots + a_k\rho^{m-k} = 0,$$

или, после сокращения на ρ^{m-k}, алгебраическое уравнение

Характеристическое уравнение

$$\rho^k + a_1\rho^{k-1} + a_2\rho^{k-2} + \ldots + a_{k-1}\rho + a_k = 0 \qquad (9.52)$$

Линейные разностные уравнения...

Если ρ — простой корень характеристического уравнения (9.52) (возможно, комплексный), то ему соответствует решение уравнения (9.51) — последовательность $y = \{1, \rho, \rho^2, \ldots, \rho^m, \ldots\}$, которая остается ограниченной тогда и только тогда, когда $|\rho| \leq 1$.

Если же корень ρ имеет кратность $l > 1$, то ему соответствует l линейно независимых решений вида

$$\begin{aligned} y^{(1)} &= \{1, \rho, \rho^2, \ldots, \rho^m, \ldots\}, \\ y^{(2)} &= \{1, 1\rho, 2\rho^2, \ldots, m\rho^m, \ldots\}, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y^{(l)} &= \{1, 1^{l-1}\rho, 2^{l-1}\rho^2, \ldots, m^{l-1}\rho^m, \ldots\}, \end{aligned}$$

которые ограничены тогда и только тогда, когда $|\rho| < 1$.

Линейное неоднородное разностное уравнение k-го порядка с постоянными коэффициентами:

$$y_m + a_1 y_{m-1} + a_2 y_{m-2} + \ldots + a_k y_{m-k} = b_0, \qquad (9.53)$$

где $b_0, a_i \in \mathbb{R}$ $(i = 1, 2, \ldots, k)$, $b_0 \neq 0$. Общее решение уравнения (9.53) записывается в виде суммы его частного решения и общего решения соответствующего ему однородного уравнения (9.51).

Устойчивость линейного разностного уравнения

Определение 9.7.

Тривиальное решение разностного уравнения (9.51) называется устойчивым, если[a]

$$\forall \varepsilon > 0 \;\; \exists \delta > 0 : \;\; \forall \{y_i\}_{i=0}^{k-1} : \\ |y_i| < \delta, \;\; i = 0, 1, \ldots, k-1 \Rightarrow |y_j| < \varepsilon, \;\; j = k, k+1, \ldots \tag{9.54}$$

[a] В этом случае говорят также, что разностное уравнение (9.51) устойчиво.

Теорема 9.4. Критерий устойчивости

Для того чтобы разностное уравнение (9.51) было устойчивым, необходимо и достаточно, чтобы для всех корней характеристического уравнения (9.52) выполнялось условие

$$|\rho| \leq 1, \tag{9.55}$$

причем если $|\rho| = 1$, то такой корень не является кратным.

9.11. 0-устойчивость и сходимость k-шаговых разностных методов

Определение 9.8.

k-шаговый разностный метод (9.41) называется 0-устойчивым, если разностное уравнение

$$y_m = \sum_{i=1}^{k} \alpha_i y_{m-i}, \ m = k, \ldots, n-1, \tag{9.56}$$

получающееся из формулы (9.41) при h $= 0$, является устойчивым.

ПРИМЕР 1. Явный двушаговый разностный метод

$$y_m = -4y_{m-1} + 5y_{m-2} + h\left(4f_{m-1} + 2f_{m-2}\right),$$

который имеет четвертый порядок точности на одном шаге (третий — на всем промежутке) (см. (9.46)).
Разностное уравнение, получающееся из (9.46) при $h = 0$, имеет вид $y_m = -4y_{m-1} + 5y_{m-2}$ (см. (9.49)).

0-устойчивость и сходимость k-шаговых разностных методов ...

Соответствующее (9.49) характеристическое уравнение

$$\rho^2 + 4\rho - 5 = 0$$

имеет корни $\rho_1 = 1$, $\rho_2 = -5$. Таким образом, условие устойчивости $|\rho| \leq 1$ не выполняется. Следовательно, метод (9.46) не является 0-устойчивым.

ПРИМЕР 2. Все методы Адамса (9.37), (9.39) являются 0-устойчивыми, так как при $h = 0$ соответствующее им разностное уравнение имеет вид $y_{m+1} = y_m$, характеристическое уравнение $\rho = 1$ имеет единственный корень, удовлетворяющий условию устойчивости (9.55).

Теорема 9.5. "Аппроксимация + Устойчивость = Сходимость"

Если k-шаговый разностный метод (9.41) имеет невязку порядка $p + 1$, разгон порядка p и является 0-устойчивым, то он сходится с порядком p.

9.12. Жесткие системы. A-устойчивость.

Свойство жесткости системы, задаваемой решением задачи Коши (9.1),(9.2), создает большие проблемы при численном решении этой задачи.

Пример жесткой системы

$$y' = -\alpha y, \quad x \in [0, X], \tag{9.57}$$

$$y(0) = 1, \tag{9.58}$$

где $\alpha \gg 1$

Точное решение задачи (9.57), (9.58)

$$y(x) = e^{-\alpha x} > 0 \quad \forall x \in [0, X] \tag{9.59}$$

Пример жесткой системы ...

Применение явного метода Эйлера (9.8)

$$y_{m+1} = y_m - h\alpha y_m = (1 - h\alpha)\, y_m, \quad m = 0, 1, \ldots, n-1 \qquad (9.60)$$

Выводы

1. Если $h > \frac{1}{\alpha}$ $(1 - h\alpha < 0)$, то приближенное решение задачи (9.57), (9.58), полученное явным методом Эйлера (9.60), не сохраняет такое важное свойство ее точного решения (9.59), как положительность. Таким образом, допустимые значения h в (9.60) должны удовлетворять неравенству $h < \frac{1}{\alpha}$.
2. Разностное уравнение (9.60) устойчиво, только если $h \leq \frac{2}{\alpha}$ $(|1 - h\alpha| \leq 1)$. Но малый шаг ведет к накоплению вычислительной погрешности и метод (9.60) как с большими, так и с малыми шагами h может дать очень неточные результаты.

В данном примере эту проблему можно решить с помощью переменного шага h, малого на почти вертикальном участке и большого на почти горизонтальном участке, но если решается система уравнений, то переменный шаг h не поможет.

Пример жесткой системы ...

Применение неявного метода Эйлера

$$y_{m+1} = y_m - h\alpha y_{m+1}$$

или

$$y_{m+1} = \frac{1}{1+h\alpha} y_m, \quad m = 0, 1, \ldots, n-1 \tag{9.61}$$

Выводы

1. Приближенное решение задачи (9.57), (9.58), полученное неявным методом Эйлера (9.60), сохраняет свойство положительности ее точного решения (9.59) при любом шаге $h > 0$.
2. Разностное уравнение (9.61) устойчиво, так как корень $\rho = \frac{1}{1+h\alpha}$ его характеристического уравнения удовлетворяет условию $0 < \rho < 1$ при любом шаге $h > 0$.
3. Задачу (9.57), (9.58) можно решать неявным методом Эйлера с любым шагом $h > 0$.

Пример жесткой системы ...

На рис. 17 жирной линией показано точное решение задачи Коши (9.57), (9.58) при $\alpha = 10$, ломаными линиями показаны приближенные решения, полученные явным (кружки) и неявным (квадратики) методом Эйлера при шаге $h = 0.2$.

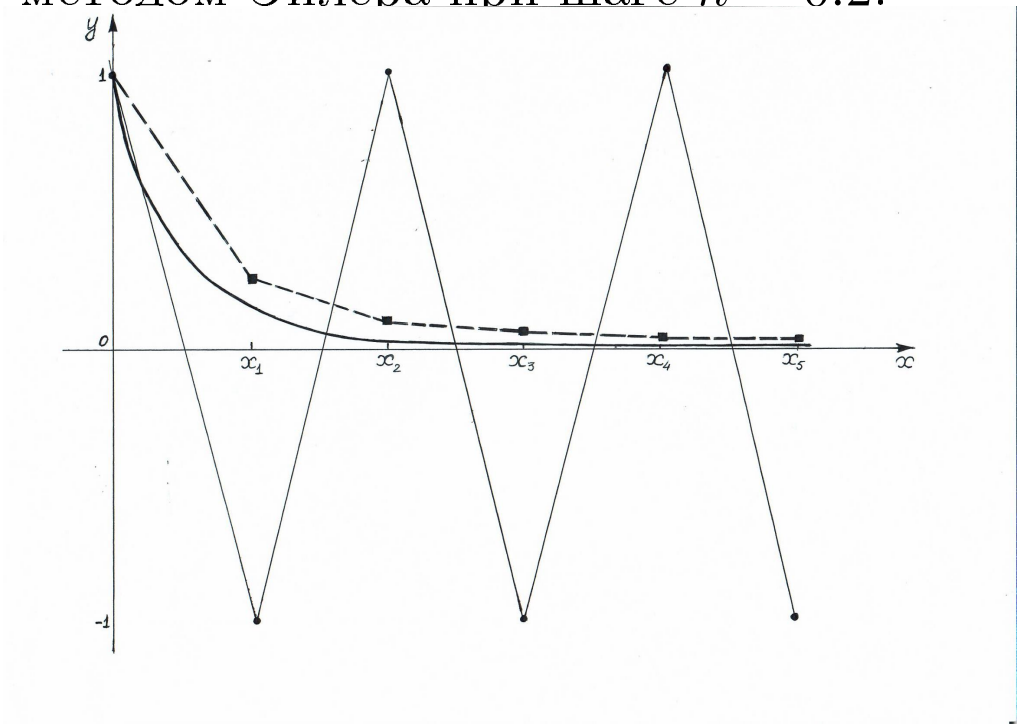


Рис. 17: Графики точного решения при $\alpha = 10$, а также приближенного, полученного явным и неявным методом Эйлера при $h = 0.2$.

Определение жесткой системы

Не существует единого определения жестких систем, некоторые из них:

1. Жесткие системы — системы с наличием почти вертикальных и почти горизонтальных участков решения.
2. Жесткие системы — системы с большой константой Липшица.
3. Жесткие системы — системы, которые нельзя решать явными методами.

Определение жесткой системы ...

Система дифференциальных уравнений в нормальной форме

$$\boldsymbol{y}' = \boldsymbol{f}(x, \boldsymbol{y}), \quad \boldsymbol{y} \in \mathbb{R}^n, \ x \in [x_0, x_0 + X], \tag{9.62}$$

$$\boldsymbol{y}(x_0) = \boldsymbol{y}_0. \tag{9.63}$$

Пусть $\boldsymbol{y}(x)$ — точное решение задачи Коши (9.62), (9.63).
Система первого приближения вдоль этого решения

$$\boldsymbol{y}' = A(x)\boldsymbol{y}, \quad x \in [x_0, x_0 + X].$$

Пусть $\lambda_i = \lambda_i(x)$ — собственные числа матрицы $A(x)$.
Если для некоторого $x \in [x_0, x_0 + X]$ выполняются условия:

$$1) \ Re(\lambda_i(x)) < 0;$$

$$2) \ \frac{\max Re(\lambda_i(x))}{\min Re(\lambda_i(x))} \geq 1,$$

то система (9.62), (9.63) называется жесткой. $\boxtimes$

▲22. A-устойчивость

Для решения жестких систем рекомендуется применять методы со специальным свойством A-устойчивости.

Пусть к решению тестового уравнения

$$y' = \lambda y, \quad x \in [0, X], \tag{9.64}$$

где λ — комплексное число, применяется k-шаговый разностный метод (9.41)

$$y_m = \sum_{i=1}^{k} \alpha_i y_{m-i} + h \sum_{j=0}^{k} \beta_j f_{m-j}, \quad m = k, \ldots, n-1.$$

В результате получается линейное однородное разностное уравнение k-го порядка

$$y_m = \sum_{i=1}^{k} \alpha_i y_{m-i} + h\lambda \sum_{j=0}^{k} \beta_j y_{m-j} \sim$$

$$(1 - h\lambda\beta_0)\, y_m - \sum_{i=1}^{k} (\alpha_i + h\lambda\beta_i)\, y_{m-i} = 0, \tag{9.65}$$

$$m = k, \ldots, n-1.$$

A-устойчивость ...

Определение 9.9.

Областью устойчивости метода (9.41) называется множество параметров $h\lambda$ на комплексной плоскости, таких, что соответствующее ему разностное уравнение (9.65) устойчиво ($|\rho(h\lambda)| < 1$).

Определение 9.10.

Метод (9.41) называется A-устойчивым, если область устойчивости содержит всю левую комплексную полуплоскость $Re(h\lambda) < 0$.

ПРИМЕР.
Явный метод Эйлера $y_{m+1} = y_m + hf_m, \quad m = 0, 1, \ldots, n-1.$

Результатом его применения к тестовому уравнению (9.64) является разностное уравнение

$$y_{m+1} = y_m + h\lambda y_m,$$

A-устойчивость ...

характеристическое уравнение $\rho - (1 + h\lambda) = 0$ которого имеет единственный корень $\rho = 1 + h\lambda$.
Область устойчивости $|1 + h\lambda| < 1$ представляет собой на комплексной плоскости внутренность единичного круга с центром в точке $(-1, 0)$ и не содержит всю левую комплексную полуплоскость (рис. 18.а). Метод не является A-устойчивым.

ПРИМЕР. Неявный метод Эйлера
$y_{m+1} = y_m + hf_{m+1}, \quad m = 0, 1, \ldots, n - 1.$

Результатом его применения к тестовому уравнению (9.64) является разностное уравнение

$$y_{m+1} = y_m + h\lambda y_{m+1},$$

корнем характеристического уравнения которого является $\rho = \frac{1}{1-h\lambda}$. Область устойчивости $|1 - h\lambda| > 1$ представляет собой на комплексной плоскости внешность единичного круга с центром в точке $(1, 0)$ и содержит всю левую комплексную полуплоскость (рис. 18.б). Метод является A-устойчивым.

A-устойчивость ...

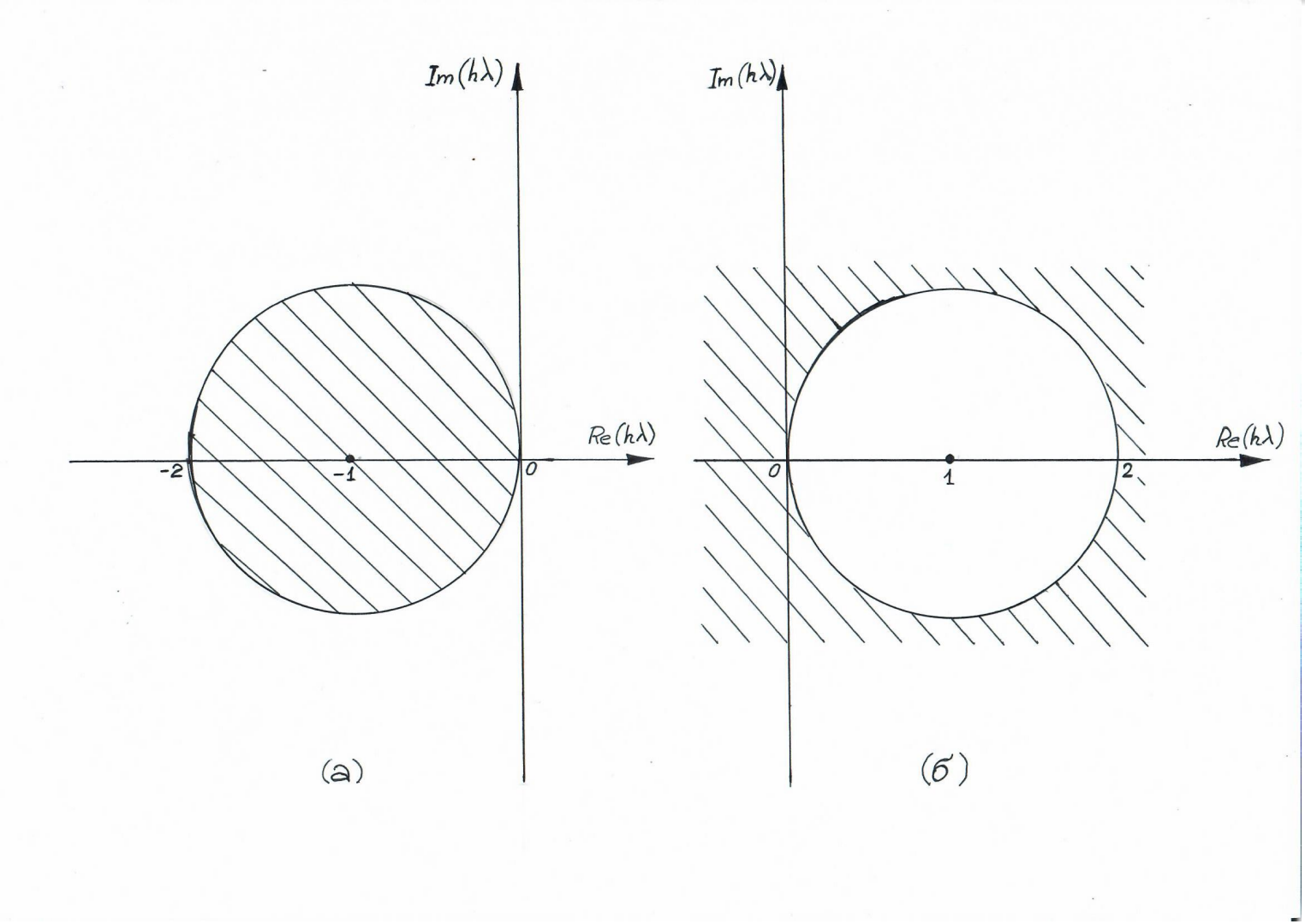


Рис. 18: Область устойчивости: а — явного метода Эйлера; б — неявного метода Эйлера

ТЕМА 10. Численные методы решения краевых задач для дифференциальных уравнений второго порядка

10.1. Метод стрельбы решения краевой задачи.

Постановка краевой задачи

$$y'' = f(x, y, y'), \quad x \in [a, b], \tag{10.1}$$

$$y(a) = \alpha, \quad y(b) = \beta \tag{10.2}$$

Решение краевой задачи (10.1),(10.2) сводится к решению вспомогательной задачи Коши с параметром и нелинейного уравнения, корень которого определяет искомое значение параметра

Однопараметрическое семейство вспомогательных задач Коши

$$y'' = f(x, y, y'), \quad x \in [a, b], \tag{10.3}$$

$$y(a) = \alpha, \quad y'(a) = \mu, \tag{10.4}$$

где $\mu \in \mathbb{R}$ — параметр.

Метод стрельбы ...

Задачу Коши (10.3),(10.4) при фиксированном значении параметра μ можно решать численно [38]. Тогда, если предположить, что решение $y(x, \mu)$ задачи Коши (10.3),(10.4) найдено, то нужно подобрать значение μ^* параметра μ, чтобы выполнялось условие

$$y(b, \mu^*) = \beta.$$

Для этого нужно решить нелинейное уравнение

$$\varphi(\mu) = y(b, \mu) - \beta = 0, \tag{10.5}$$

что также можно сделать численными методами. Эта методика решения краевых задач называется методом стрельбы (рис. 19).

Метод стрельбы представляет собой пару вложенных методов:

Внешний — для решения нелинейного уравнения (10.5), внутренний — для решения задачи Коши (10.3),(10.4) (например, метод дихотомии и явный метод Эйлера).

[38] При этом следует только учитывать, что дифференциальное уравнение второго порядка нужно предварительно свести к системе дифференциальных уравнений первого порядка.

Метод стрельбы ...

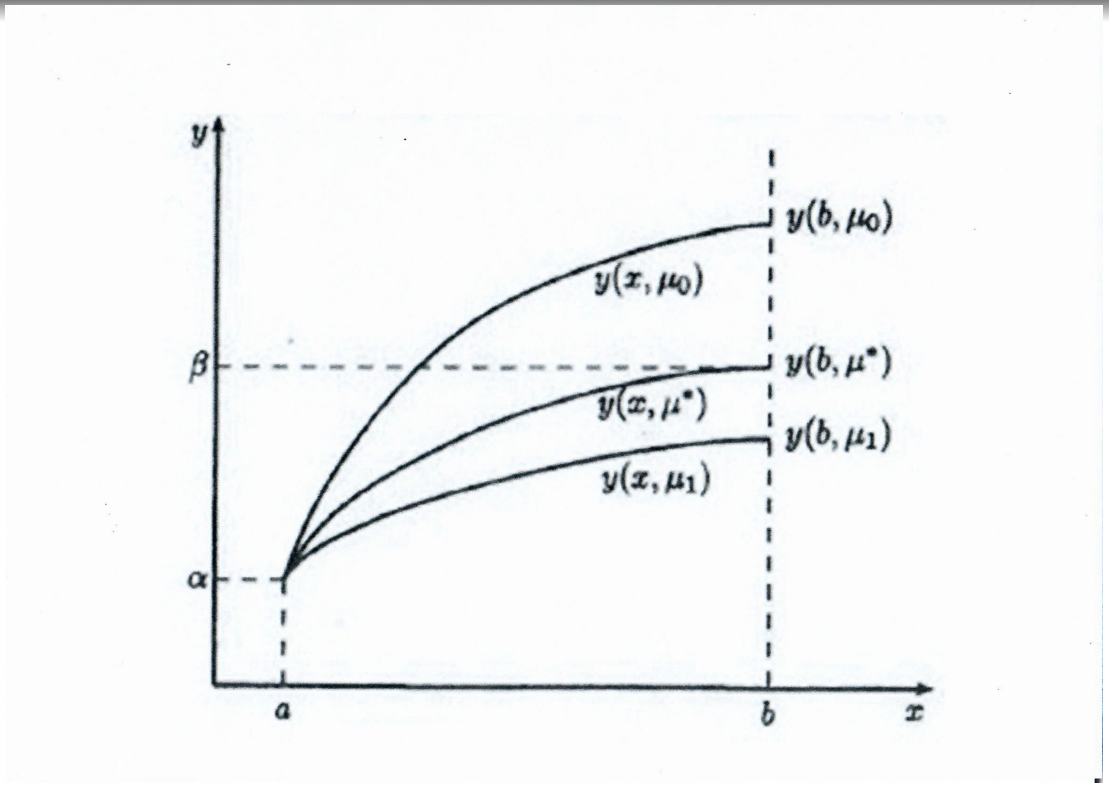


Рис. 19: Геометрическая интерпретация метода стрельбы.

Требуется определить такое значение μ^* параметра μ, чтобы решение $y(x, \mu^*)$ задачи Коши (10.3),(10.4) являлось решением краевой задачи (10.1),(10.2).

Метод стрельбы ...

Особый интерес представляет алгоритм в случае, когда для решения нелинейного уравнения (10.5) применяется метод Ньютона

$$\mu_{k+1} = \mu_k - \frac{\varphi(\mu_k)}{\varphi'(\mu_k)}, \quad k = 0, 1, 2, \ldots, \tag{10.6}$$

где на каждом шаге необходимо вычислять величину $\varphi'(\mu_k)$.
Для того чтобы вычислить эту величину, предлагается воспользоваться уравнением в вариациях.
В предположении достаточной гладкости функций $f(x, y, y')$ и $y(x, \mu)$ дифференцирование уравнения (10.3) по μ приводит к

$$y'''_{x\mu} = f'_y y'_\mu + f'_{y'} y''_{x\mu},$$

которое в терминах вспомогательной функции $z(x, \mu) = y'_\mu(x, \mu)$ может быть записано в виде

$$z'' = f'_y(x, y(x, \mu), y'_x(x, \mu))z + f'_{y'}(x, y(x, \mu), y'_x(x, \mu))z'_x.$$

Метод стрельбы ...

В результате с учетом начальных условий (10.4) вспомогательная функция z является решением следующей задачи Коши

$$z'' = f'_y z + f'_{y'} z'_x, \quad x \in [a, b], \tag{10.7}$$

$$z(a) = 0, \quad z'(a) = 1 \tag{10.8}$$

Тогда требуемая в формуле (10.6) метода Ньютона величина

$$\varphi'(\mu_k) = z(b, \mu_k)$$

Численная реализация метода стрельбы

Таким образом, метод стрельбы с применением метода Ньютона (10.6) сводит решение краевой задачи (10.1),(10.2) к последовательному численному решению двух вспомогательных задач Коши (10.3),(10.4) и (10.7),(10.8) на каждом шаге метода Ньютона.

10.2. Метод разностной прогонки решения линейной краевой задачи

Постановка краевой задачи

$$y'' = p(x)y + q(x), \quad p(x) \geq \bar{p} > 0 \quad x \in [a, b] \tag{10.9}$$

краевые условия первого рода

$$y(a) = \alpha, \quad y(b) = \beta; \tag{10.10}$$

краевые условия третьего рода

$$y'(a) = \alpha_0 y(a) + \alpha_1, \quad \alpha_0 > 0, \tag{10.11}$$

$$y'(b) = -\beta_0 y(b) + \beta_1, \quad \beta_0 > 0. \tag{10.12}$$

Приближенное решение задачи (10.9)-(10.12)

$$\begin{array}{l} x_i \in [a, b]: \ x_i = a + ih, \ \forall i = 0, 1, \ldots, n, \quad h = \frac{b-a}{n} \\ y_i \approx y(x_i), \ \forall i = 0, 1, \ldots, n \end{array} \tag{10.13}$$

Дискретизация краевой задачи

Пусть $y \in C^{(4)}[a,b]$

$$y''(x) = \frac{y(x-h)-2y(x)+y(x+h)}{h^2} - \frac{h^2}{12} y^{(4)}(\xi(x)),$$

$$\xi(x) \in [x-h, x+h] \tag{10.14}$$

$$\forall x \in [a+h, b-h]$$

Дискретизация дифференциального уравнения (10.9)

$$\frac{y_{i-1} - 2y_i + y_{i+1}}{h^2} = p_i y_i + q_i, \quad i = 1, 2, \ldots, n-1, \tag{10.15}$$

где $p_i = p(x_i),\ q_i = q(x_i)$.

Дискретизация краевой задачи ...

Дискретизация краевых условий (10.10) и(10.11),(10.12):

краевых условий первого рода

$$y_0 = \alpha, \quad y_n = \beta; \tag{10.16}$$

краевых условий третьего рода

$$\frac{y_1 - y_0}{h} = \alpha_0 y_0 + \alpha_1, \tag{10.17}$$

$$\frac{y_n - y_{n-1}}{h} = -\beta_0 y_n + \beta_1; \tag{10.18}$$

Из (10.15)-(10.18) следует, что приближенное решение $\{y_i\}_{i=0}^n$ краевой задачи может быть построено как решение системы линейных алгебраических уравнений

Приближенное решение краевой задачи (10.9)-(10.12)

Система линейных алгебраических уравнений для краевых условий первого рода (10.10)

$$\begin{aligned} &y_0 = \alpha, \\ &y_{i-1} - (2 + h^2 p_i) y_i + y_{i+1} = h^2 q_i, \quad i = 1, 2, \ldots, n-1, \\ &y_n = \beta \end{aligned} \tag{10.19}$$

Система линейных алгебраических уравнений для краевых условий третьего рода (10.11),(10.12)

$$\begin{aligned} &-(1 + h\alpha_0) y_0 + y_1 = h\alpha_1, \\ &y_{i-1} - (2 + h^2 p_i) y_i + y_{i+1} = h^2 q_i, \quad i = 1, 2, \ldots, n-1, \\ &y_{n-1} - (1 + h\beta_0) y_n = -h\beta_1 \end{aligned} \tag{10.20}$$

Используя обозначения:
$A_i = 2 + h^2 p_i$, $B_i = h^2 q_i$ $(i = 1, 2, \ldots, n-1)$;
$A_0 = 1$, $B_0 = \alpha$, $A_n = 1$, $B_n = \beta$ (для краевых условий (10.10));
$A_0 = 1 + h\alpha_0$, $B_0 = h\alpha_1$, $A_n = 1 + h\beta_0$, $B_n = -h\beta_1$ (для краевых условий (10.11),(10.12)),
системы (10.19) и (10.20) можно записать в матричной форме:

Приближенное решение краевой задачи (10.9)-(10.12) ...

для краевых условий первого рода (10.10)

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & -A_1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -A_2 & 1 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -A_{n-1} & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \cdots \\ y_{n-1} \\ y_n \end{pmatrix} = \begin{pmatrix} B_0 \\ B_1 \\ B_2 \\ \cdots \\ B_{n-1} \\ B_n \end{pmatrix} \tag{10.21}$$

для краевых условий третьего рода (10.11),(10.12)

$$\begin{pmatrix} -A_0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & -A_1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -A_2 & 1 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & -A_{n-1} & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -A_n \end{pmatrix} \cdot \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \cdots \\ y_{n-1} \\ y_n \end{pmatrix} = \begin{pmatrix} B_0 \\ B_1 \\ B_2 \\ \cdots \\ B_{n-1} \\ B_n \end{pmatrix} \tag{10.22}$$

Приближенное решение краевой задачи (10.9)-(10.12)

Вопросы

1. Существование и единственность решений систем (10.19) и (10.20);
2. Методы решения систем (10.19) и (10.20);
3. Сходимость приближенных решений $\boldsymbol{y}(h)$ к точному решению $\boldsymbol{y}$ краевой задачи (10.9)-(10.12) при $h \to 0$.

Существование и единственность решений систем (10.19) и (10.20)

Матрицы систем (10.19) и (10.20) для любого $h > 0$ обладают свойством строгого диагонального преобладания, поскольку $p(x) \geq \bar{p} > 0 \quad x \in [a, b]$ и $\alpha_0 > 0$, $\beta_0 > 0$. Следовательно, эти матрицы при любом $h > 0$ являются невырожденными и у систем (10.19) и (10.20) существуют соответствующие единственные решения.

Диагональное преобладание матрицы является достаточным условие сходимости метода Якоби — итерационного метода приближенного решения систем линейных алгебраических уравнений.$\boxtimes$

▲23. Методы решения систем (10.19) и (10.20)

Метод разностной прогонки

Матрицы систем (10.19) и (10.20) являются трехдиагональными. Это свойство позволяет использовать для точного решения таких систем некоторую модификацию метода Гаусса, которую называют методом разностной прогонки.

В основе метода разностной прогонки лежит следующая особенность реализации метода Гаусса решения системы с трехдиагональной матрицей.
Метод Гаусса как процедура последовательного исключения переменных, приводящая расширенную матрицу системы к верхнетреугольному виду, применительно к системе к трехдиагональной матрицей позволяет построить равносильную исходной систему линейных алгебраических уравнений, у которой каждое уравнение содержит только две переменные, коэффициенты при которых расположены на главной диагонали матрицы и диагонали, расположенной над ней.

Метод разностной прогонки решения системы с трехдиагональной матрицей

Пусть $i-1$-ое уравнение системы имеет вид

$$y_{i-1} = \lambda_i y_i + \mu_i, \quad i \geq 1. \tag{10.23}$$

Для краевых условий первого рода (10.10) $y(a) = \alpha$
$\lambda_1 = 0$, $\mu_1 = \alpha$;
для краевых условий третьего рода (10.11) $-A_0 y_0 + y_1 = B_0$
$\lambda_1 = \frac{1}{A_0}$, $\mu_1 = -\frac{B_0}{A_0}$.
Подстановка правой части равенства (10.23) вместо y_{i-1} в i-ое уравнение

$$y_{i-1} - A_i y_i + y_{i+1} = B_i$$

преобразует его к виду

$$\lambda_i y_i + \mu_i - A_i y_i + y_{i+1} = B_i$$

или

$$y_i = \frac{1}{A_i - \lambda_i} y_{i+1} + \frac{\mu_i - B_i}{A_i - \lambda_i}. \tag{10.24}$$

Метод разностной прогонки ...

Таким образом выражение (10.24) определяет рекуррентную связь

Прямая прогонка

$$\lambda_{i+1} = \frac{1}{A_i - \lambda_i},$$

$$\mu_{i+1} = \frac{\mu_i - B_i}{A_i - \lambda_i}, \qquad (10.25)$$

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

с помощью которой определяются значения прогоночных коэффициентов (λ_i, μ_i), $i = 1, 2, \ldots, n+1$ — прямая прогонка. При этом справедливо выражение

$$y_n = \mu_{n+1}, \qquad (10.26)$$

которое очевидно для краевых условий первого рода (10.10) $y(b) = \beta = \mu_{n+1}$.

Метод разностной прогонки ...

Для краевых условий третьего рода (10.12) последнее ($n+1$-ое) уравнение в системе (10.20) в принятых выше обозначениях имеет вид

$$y_{n-1} - A_n y_n = B_n, \tag{10.27}$$

а предпоследнее (n-ое) уравнение в результате прямой прогонки записывается в виде

$$y_{n-1} = \lambda_n y_n + \mu_n.$$

Последним шагом прямой прогонки является подстановка правой части этого равенства вместо y_{n-1} в уравнение (10.27). В результате возникает выражение

$$\lambda_n y_n + \mu_n - A_n y_n = B_n.$$

Откуда с учетом формул (10.25)

$$y_n = \frac{\mu_n - B_n}{A_n - \lambda_n} = \mu_{n+1}.$$

Метод разностной прогонки ...

Заключительным этапом решения системы методом разностной прогонки является обратная прогонка, которая осуществляется по формулам (10.23)

Обратная прогонка

$$y_{i-1} = \lambda_i y_i + \mu_i, \qquad i = n, n-1, \ldots, 1. \tag{10.28}$$

Осуществимость прямой прогонки (10.25)

Требуется доказать, что при условии строгого диагонального преобладания матрицы системы (10.19) ((10.20))

$$A_i - \lambda_i \neq 0, \quad \forall i = 1, 2, \ldots, n. \tag{10.29}$$

[39] Самостоятельно, например методом математической индукции, показать, что $\lambda_i < 1, \quad \forall i = 1, 2, \ldots, n$. Тогда $A_i - \lambda_i > 1, \quad \forall i = 1, 2, \ldots, n$.

Сходимость метода разностной прогонки

Требуется доказать, что при $h \to 0$

$$y_i(h) \to y(x_i)$$

Для этого достаточно оценить погрешности

$$\varepsilon_i = y(x_i) - y_i, \quad i = 0, 1, 2, \ldots, n. \tag{10.30}$$

Учитывая, что для внутренних узлов выполняется

$$y''(x_i) = \frac{y(x_{i-1}) - 2y(x_i) + y(x_{i+1})}{h^2} + R_i, \quad i = 1, 2, \ldots, n-1, \tag{10.31}$$

где для погрешности этой формулы численного дифференцирования для второй производной справедлива оценка

$$|R_i| \le \frac{1}{12} M_4 h^2, \quad M_4 = \max_{x \in [a,b]} |y^{(4)}(x)|. \tag{10.32}$$

Сходимость метода разностной прогонки ...

Подстановка правой части выражения (10.31) в дифференциальное уравнение (10.9) приводит к равенствам

$$\frac{y(x_{i-1}) - 2y(x_i) + y(x_{i+1})}{h^2} + R_i = p_i y(x_i) + q_i, \quad i = 1, 2, \ldots, n-1.$$

Вычитание из них соответствующих (по индексу i) равенств (10.15) приводит к выражениям

$$\frac{(y(x_{i-1}) - y_{i-1}) - 2(y(x_i) - y_i) + (y(x_{i+1}) - y_{i+1})}{h^2} + R_i = \\ = p_i \left(y(x_i) - y_i\right),$$

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

которые в терминах погрешностей ε_i можно переписать в виде

$$\frac{\varepsilon_{i-1} - 2\varepsilon_i + \varepsilon_{i+1}}{h^2} + R_i = p_i \varepsilon_i, \quad i = 1, 2, \ldots, n-1$$

или

$$\varepsilon_{i-1} - (2 + h^2 p_i)\varepsilon_i + \varepsilon_{i+1} = -R_i h^2, \quad i = 1, 2, \ldots, n-1.$$

Сходимость метода разностной прогонки. Случай краевых условий первого рода (10.10).

В случае краевых условий первого рода (10.10) (см. систему (10.19)) для погрешностей (10.30) система уравнений имеет вид

$$\begin{aligned} &\varepsilon_0 = 0, \\ &\varepsilon_{i-1} - (2 + h^2 p_i)\varepsilon_i + \varepsilon_{i+1} = -R_i h^2, \quad i = 1, 2, \ldots, n-1, \\ &\varepsilon_n = 0. \end{aligned} \tag{10.33}$$

Из (10.33) следует, что

$$(2 + h^2 p_i)\varepsilon_i = \varepsilon_{i-1} + \varepsilon_{i+1} + R_i h^2, \quad i = 1, 2, \ldots, n-1.$$

Тогда

$$\begin{aligned} &(2 + h^2 p_i)|\varepsilon_i| \le 2|\varepsilon_{i^*}| + |R_i| h^2, \quad i = 1, 2, \ldots, n-1, \\ &(2 + h^2 p_{i^*})|\varepsilon_{i^*}| \le 2|\varepsilon_{i^*}| + |R_{i^*}| h^2, \\ &h^2 p_{i^*} |\varepsilon_{i^*}| \le |R_{i^*}| h^2, \end{aligned}$$

где $|\varepsilon_{i^*}| = \max_{i=1,2,\ldots,n-1} |\varepsilon_i|$.

Сходимость метода разностной прогонки. Случай краевых условий первого рода (10.10). ...

Откуда

$$|\varepsilon_{i^*}| \leq \frac{|R_{i^*}|}{p_{i^*}}.$$

В результате с учетом оценки (10.32) и свойств функции p ($p(x) \geq \bar{p} > 0 \quad x \in [a, b]$)

Случай краевых условий первого рода (10.10)

$$|y(x_i) - y_i| \leq \frac{1}{12\bar{p}} M_4 h^2, \quad i = 0, 1, 2, \ldots, n. \tag{10.34}$$

Вывод

В силу (10.34), для краевых условий первого рода (10.10) метод разностной прогонки (10.19) сходится со вторым порядком.

Сходимость метода разностной прогонки. Случай краевых условий третьего рода (10.11),(10.12).

В этом случае для построения оценки погрешности (10.30) метода разностной прогонки (см. систему (10.20)) придется учитывать не только погрешности формулы численного дифференцирования для второй производной во внутренних узлах, но и погрешности формул численного дифференцирования для первой производной на краях:

$$y'(a) = \frac{y(x_1) - y(x_0)}{h} + r_0, \quad y'(b) = \frac{y(x_n) - y(x_{n-1})}{h} + r_n,$$

где для погрешности этих формул численного дифференцирования для первой производной справедливы оценки

$$|r_0| \le \frac{1}{2} M_2 h, \quad |r_n| \le \frac{1}{2} M_2 h, \quad M_2 = \max_{x \in [a,b]} |y''(x)|.$$

Тогда в случае краевых условий третьего рода (10.11),(10.12) (см. систему (10.20)) для погрешностей (10.30) система уравнений имеет вид

Сходимость метода разностной прогонки. Случай краевых условий третьего рода (10.11),(10.12). ...

$$
\begin{aligned}
&-(1+h\alpha_0)\varepsilon_0+\varepsilon_1=r_0,\\
&\varepsilon_{i-1}-(2+h^2p_i)\varepsilon_i+\varepsilon_{i+1}=-R_ih^2,\quad i=1,2,\ldots,n-1,\\
&\varepsilon_{n-1}-(1+h\beta_0)\varepsilon_n=-r_n.
\end{aligned}
\tag{10.35}
$$

Аналогично (10.34) нетрудно доказать, что

Случай краевых условий третьего рода (10.11),(10.12)

$$
|y(x_i)-y_i|\le\max\{\frac{1}{2\alpha_0}M_2h,\frac{1}{2\beta_0}M_2h,\frac{1}{12\bar{p}}M_4h^2\},\quad i=0,1,2,\ldots,n. \tag{10.36}
$$

Вывод

В силу (10.36), для краевых условий третьего рода (10.11),(10.12) метод разностной прогонки (10.20) сходится с первым[a] порядком.

---

[a] Для достижения второго порядка сходимости используются разные способы: формулы дифференцирования для первой производной по трем узлам на край или метод фиктивного узла.

10.3. Более точная аппроксимация краевых условий третьего рода (10.11),(10.12).

Аппроксимация первой производной в краевом условии (10.11)

$$y'(a) = \frac{-3y(x_0) + 4y(x_1) - y(x_2)}{2h} + O(h^2). \qquad (10.37)$$

Тогда первое уравнение в системе (10.20) записывается следующим образом

$$\frac{-3y_0 + 4y_1 - y_2}{2h} = \alpha_0 y_0 + \alpha_1$$

или, если привести подобные,

$$-(3 + 2\alpha_0 h)y_0 + 4y_1 - y_2 = 2\alpha_1 h. \qquad (10.38)$$

Свойство трехдиагональности матрицы системы (10.20) утрачено. Однако его можно восстановить, если сложить уравнение (10.38) со вторым уравнением

$$y_0 - A_1 y_1 + y_2 = B_1$$

этой системы. В результате первое уравнение принимает вид

10.4. Более точная аппроксимация краевых условий третьего рода (10.11), (10.12). ...

$$-(2+2\alpha_0 h)y_0 + (4 - A_1)y_1 = 2\alpha_1 h + B_1.$$

Таким образом, матрица системы (10.20) становится трехдиагональной.

Более точную аппроксимацию краевых условий третьего рода (10.11),(10.12) можно получить используя фиктивные узлы

Аппроксимация второй производной в краевом условии (10.11) с фиктивным узлом $x_{-1} = a - h$

$$y'(a) = \frac{y(a+h) - y(a-h)}{2h} + O(h^2) = \frac{y(x_1) - y(x_{-1})}{2h} + O(h^2). \tag{10.39}$$

Тогда первое уравнение в системе (10.20) записывается следующим образом

Аппроксимация первой производной в краевом условии (10.11) с использованием фиктивного узла

$$\frac{y_1 - y_{-1}}{2h} = \alpha_0 y_0 + \alpha_1$$

или, если привести подобные,

$$-y_{-1} - 2\alpha_0 h y_0 + y_1 = 2\alpha_1 h. \tag{10.40}$$

При этом, если предположить, что функции p и q продолжимы влево по непрерывности, то можно выписать дискретный аналог (10.15) дифференциального уравнения (10.9) в узле x_0:

$$\frac{y_{-1} - 2y_0 + y_1}{h^2} = p_0 y_0 + q_0,$$

который приводит к еще одному уравнению, содержащему неизвестную величину y_{-1}

$$y_{-1} - \left(2 + p_0 h^2\right) y_0 + y_1 = q_0 h^2. \tag{10.41}$$

Аппроксимация первой производной в краевом условии (10.11) с использованием фиктивного узла ...

В результате сложения уравнений (10.40) и (10.41) удается исключить из рассмотрения величину y_{-1}, тем самым построив первое уравнение системы (10.20)

$$-\left(2+2\alpha_0 h+p_0 h^2\right) y_0+2y_1=2\alpha_1 h+q_0 h^2. \qquad (10.42)$$

Таким образом матрица системы (10.20) не только становится трехдиагональной, но и приобретает свойство диагонального преобладания (см. (10.42)).

Аналогичным образом можно построить более точную аппроксимацию краевого условия (10.12) (на правом конце).

Порядок сходимости метода разностной прогонки

Результатом применения описанных выше подходов к аппроксимации краевых условий третьего рода (10.11), (10.12) является второй порядок сходимости метода разностной прогонки.

10.5. Вариационные методы решения краевой задачи

Постановка краевой задачи

$$y'' = f(x, y, y'), \quad x \in [a, b], \tag{10.43}$$

$$y(a) = \alpha, \quad y(b) = \beta. \tag{10.44}$$

Основная идея вариационных методов решения краевой задачи (10.43), (10.44) заключается в определении ее решения как решения некоторой вспомогательной экстремальной задачи

Постановка экстремальной задачи

$$J[y] \to \min_{y \in Y}, \tag{10.45}$$

где

$$J[y] = \int_a^b F(x, y, y')dx, \tag{10.46}$$

$$Y = \{y \in C^{(2)}([a, b]) \mid y(a) = \alpha, \quad y(b) = \beta\}. \tag{10.47}$$

Уравнение Эйлера в задаче (10.45)

Пусть $y^* \in Y$ — экстремаль в задаче (10.45), то есть функция, доставляющая минимальное значение функционалу (10.46) на функциональном пространстве Y (см. (10.47)).

Необходимое условие экстремума в задаче (10.45)

$$\frac{\partial}{\partial \varepsilon} J[y^* + \varepsilon\varphi]|_{\varepsilon=0} = 0 \quad \forall \varphi \in \Phi, \tag{10.48}$$

где $\varepsilon \in \mathbb{R}$; $\Phi = \{\varphi \in C^{(2)}([a,b]) \mid \varphi(a) = \varphi(b) = 0\}$

$$J[y^* + \varepsilon\varphi] = \int_a^b F(x, y^* + \varepsilon\varphi, (y^*)' + \varepsilon\varphi')dx,$$

$$\frac{\partial}{\partial \varepsilon} J[y^* + \varepsilon\varphi]|_{\varepsilon=0} = \int_a^b [F'_y(x, y^*, (y^*)')\varphi + F'_{y'}(x, y^*, (y^*)')\varphi']dx =$$

$$= \int_a^b F'_y \varphi dx + F'_{y'}\varphi|_a^b - \int_a^b \frac{d}{dx}\left(F'_{y'}\right)\varphi dx$$

Уравнение Эйлера ...

Поскольку $\varphi(a) = \varphi(b) = 0$, то

$$\frac{\partial}{\partial \varepsilon} J[y^* + \varepsilon\varphi]|_{\varepsilon=0} = \int_a^b F'_y \varphi dx - \int_a^b \frac{d}{dx}\left(F'_{y'}\right)\varphi dx =$$

$$= \int_a^b [F'_y - \frac{d}{dx}\left(F'_{y'}\right)]\varphi dx$$

Вариационный принцип

$$\int_a^b g(x)\varphi(x)dx = 0 \quad \forall\varphi \Leftrightarrow g(x) \equiv 0, \quad x \in [a, b] \tag{10.49}$$

Тогда, в силу (10.49),

$$\int_a^b [F'_y - \frac{d}{dx}\left(F'_{y'}\right)]\varphi dx = 0 \quad \forall\varphi \Leftrightarrow F'_y - \frac{d}{dx}\left(F'_{y'}\right) \equiv 0, \quad x \in [a, b]$$

Следовательно, необходимое условие экстремума (10.48) в задаче (10.45) равносильно следующему тождеству

Уравнение Эйлера ...

Уравнение Эйлера

$$\frac{d}{dx}\left(F'_{y'}(x, y^*, (y^*)')\right) - F'_y(x, y^*, (y^*)') \equiv 0, \quad x \in [a, b] \qquad (10.50)$$

Как нетрудно видеть, уравнение Эйлера (10.50) является дифференциальным уравнением второго порядка относительно функции y.

Связь между краевой задачей (10.43), (10.44) и экстремальной задачей (10.45)

Для того, чтобы экстремаль y^* задачи (10.45) являлась и решением краевой задачи (10.43), (10.44) достаточно выбрать функцию $F(x, y, y')$ так, чтобы дифференциальное уравнение $y'' = f(x, y, y')$ (см. (10.43)) и уравнение Эйлера (10.50) были идентичны.

$\boxtimes$

▲24. Общая схема построения приближенного решения экстремальной задачи (10.45)

I. Переход от задачи (10.45) к вспомогательной экстремальной задаче на конечно-мерном подпространстве Y_n бесконечно-мерного пространства Y

Пусть

$$Y_n = \{\ y(x, c_1, c_2, \ldots, c_n) \in C^{(2)}([a,b])\ \mid \\ y(a, c_1, c_2, \ldots, c_n) = \alpha,\quad y(b, c_1, c_2, \ldots, c_n) = \beta \\ \forall (c_1, c_2, \ldots, c_n)^\top \in \mathbb{R}^n\ \}, \tag{10.51}$$

где $Y_n \subset Y \quad \forall n \in \mathbb{N}$.

Вспомогательная экстремальная задача:

$$J[y] \to \min_{y \in Y_n}. \tag{10.52}$$

Общая схема построения приближенного решения экстремальной задачи (10.45) ...

II. Переход от вспомогательной экстремальной задачи (10.52) к задаче минимизации функции нескольких переменных

Пусть

$$G(c_1, c_2, \ldots, c_n) = J[y_n], \tag{10.53}$$

где $y_n(x) = y(x, c_1, c_2, \ldots, c_n) \in Y_n$.

Задача минимизации функции нескольких переменных:

$$G(c_1, c_2, \ldots, c_n) \to \min_{(c_1, c_2, \ldots, c_n)^\top \in \mathbb{R}^n}. \tag{10.54}$$

III. Решение системы алгебраических уравнений относительно n неизвестных

Необходимое условие в задаче минимизации функции нескольких переменных (10.54):

$$\frac{\partial}{\partial c_i} G(c_1^*, c_2^*, \ldots, c_n^*) = 0 \quad \forall i = 1, 2, \ldots, n. \tag{10.55}$$

Вопросы

1. Существование и единственность решения системы (10.55);
2. Является ли решение системы (10.55) точкой экстремума для функции G (см. (10.53);
3. В предположении, что система (10.55) имеет единственное решение $(c_1^*, c_2^*, \ldots, c_n^*)^\top \in \mathbb{R}^n$, которое является точкой минимума функции G, пусть $y_n^*(x) = y_n(x, c_1^*, c_2^*, \ldots, c_n^*) \in Y_n$ — экстремаль во вспомогательной задаче (10.52).

   Существует ли предел

   $$\lim_{n\to\infty} J[y_n^*] = J[y^*];$$

4. Существует ли предел

   $$\lim_{n\to\infty} y_n^* = y^*.$$

Общий ответ на 3-ий вопрос

Определение 10.1.

Последовательность подпространств $Y_n \subset Y$ плотна в пространстве Y, если

$$\forall y \in Y \quad \forall \varepsilon > 0 \quad \exists n \in \mathbb{N} \quad \exists y_n \in Y_n :$$

$$|y(x) - y_n(x)| < \varepsilon, \quad |y'(x) - y'_n(x)| < \varepsilon \tag{10.56}$$

$$\forall x \in [a, b].$$

Теорема 10.1.

Пусть последовательность подпространств $Y_n \subset Y$ плотна в пространстве Y и функция $F(x, y, y')$ ($J[y] = \int_a^b F(x, y, y')dx$) — равномерно Липшицева по аргументам y и y'. Тогда

$$J[y_n^*] \to J[y^*] \text{ при } n \to \infty.$$

Доказательство.

Доказательство

Поскольку $\forall n \in \mathbb{N} \;\; Y_n \subset Y$, то $\min_{y \in Y_n} J[y] \geq \min_{y \in Y} J[y]$. Следовательно, $J[y_n^*] \geq J[y^*]$. Тогда

$$0 \leq J[y_n^*] - J[y^*] \leq J[y_n] - J[y^*] = \\ = \int_a^b \big(F(x, y_n, y_n') - F(x, y^*, (y^*)')\big)dx. \tag{10.57}$$

Здесь

$$\int_a^b [F(x, y_n, y_n') - F(x, y^*, (y^*)')]dx = \\ = \int_a^b [F(x, y_n, y_n') - F(x, y^*, y_n')]dx + \\ + \int_a^b [F(x, y^*, y_n') - F(x, y^*, (y^*)')]dx \leq \tag{10.58}$$

$$\leq \int_a^b |F(x, y_n, y_n') - F(x, y^*, y_n')|dx + \\ + \int_a^b |F(x, y^*, y_n') - F(x, y^*, (y^*)')|dx.$$

Доказательство

В силу равномерной липшицевости функции $F(x, y, y')$ по аргументам y и y', существуют такие две константы $L > 0$ и $K > 0$, что

$$\begin{aligned} &|F(x, y_n, y_n') - F(x, y^*, y_n')| \le L|y_n(x) - y^*(x)|, \\ &|F(x, y^*, y_n') - F(x, y^*, (y^*)')| \le K|y_n'(x) - (y^*)'(x)| \\ &\quad \forall x \in [a, b]. \end{aligned} \tag{10.59}$$

Таким образом, из (10.57), (10.58) и (10.59) следует, что

$$0 \le J[y_n^*] - J[y^*] \le L \int_a^b |y_n(x) - y^*(x)|dx + K \int_a^b |y_n'(x) - (y^*)'(x)|dx. \tag{10.60}$$

Далее, поскольку последовательность подпространств $Y_n \subset Y$ плотна в пространстве Y, то

$$\forall \varepsilon > 0 \; \exists n \in \mathbb{N} \; \exists y_n \in Y_n : \; |y_n(x) - y^*(x)| < \varepsilon, \; |y_n'(x) - (y^*)'(x)| < \varepsilon \; \forall x \in [a, b]. \tag{10.61}$$

В результате (10.60), (10.61) приводят к неравенствам

$$0 \le J[y_n^*] - J[y^*] \le (b - a)(L + K)\varepsilon. \tag{10.62}$$

□ Теорема доказана.

10.6. Метод Ритца

Постановка линейной краевой задачи

$$(p(x)y')' - q(x)y = f(x), \quad x \in [a,b], \tag{10.63}$$

$$y(a) = \alpha, \quad y(b) = \beta, \tag{10.64}$$

где $p \in C^{(1)}([a,b])$, $q, f \in C([a,b])$: $p(x) \geq p_0 > 0$, $q(x) \geq 0$, $x \in [a,b]$.

Решение краевой задачи (10.63), (10.64) определяется как решение следующей экстремальной задачи

Постановка экстремальной задачи

$$J[y] \to \min_{y \in Y}, \tag{10.65}$$

где

$$J[y] = \int_a^b F(x,y,y')dx, \tag{10.66}$$

$$F(x,y,y') = p(x)[y']^2 + q(x)[y]^2 + 2f(x)y, \quad x \in [a,b], \tag{10.67}$$

$$Y = \{y \in C^{(2)}([a,b]) \mid y(a) = \alpha, \quad y(b) = \beta\}. \tag{10.68}$$

Метод Ритца ...

Нетрудно проверить, что уравнение Эйлера (10.50)

$$\frac{d}{dx}\left(F'_{y'}(x,y,y')\right) - F'_y(x,y,y') \equiv 0, \quad x \in [a,b]$$

для функции $F(x,y,y')$, определяемой выражением (10.67), идентично дифференциальному уравнению (10.63). Действительно, здесь

$$F'_{y'}(x,y,y') = 2p(x)y', \quad F'_y(x,y,y') = 2q(x)y + 2f(x).$$

Следовательно, уравнение Эйлера (10.50) для функции рассматриваемой $F(x,y,y')$ имеет вид

$$(2p(x)y')' - 2q(x)y - 2f(x) = 0, \quad x \in [a,b],$$

то есть является дифференциальным уравнением (10.63)

$$(p(x)y')' - q(x)y = f(x), \quad x \in [a,b].$$

Метод Ритца. Вспомогательная экстремальная задача на конечно-мерном подпространстве Y_n.

$$Y_n = \{\ y_n \in C^{(2)}([a,b]) \mid y_n(x) = \psi(x) + \sum_{i=1}^{n} c_i \varphi_i(x) :$$

$$\psi, \varphi_i \in C^{(2)}([a,b])(i = 1, 2, \ldots, n) : \qquad (10.69)$$

$$\psi(a) = \alpha,\ \psi(b) = \beta,$$
$$\varphi_i(a) = \varphi_i(b) = 0\ (i = 1, 2, \ldots, n)\ \}$$

Примеры

$$\psi(x) = \alpha + \tfrac{x-a}{b-a}(\beta - \alpha);$$

1) $\varphi_i(x) = (x-a)^i(x-b)\ (i = 1, 2, \ldots, n),$
2) $\varphi_i(x) = \sin\left(i\pi \tfrac{x-a}{b-a}\right)\ (i = 1, 2, \ldots, n).$

Условие плотности последовательности подпространств $Y_n \subset Y$ в пространстве Y выполняется.

Метод Ритца. Задача минимизации функции нескольких переменных.

Определение функции нескольких переменных

$$
\begin{aligned}
&J[y_n] = \\
&= \int_a^b \left(p[\psi' + \textstyle\sum_{i=1}^n c_i \varphi_i{}']^2 + q[\psi + \sum_{i=1}^n c_i \varphi_i]^2 + 2f[\psi + \sum_{i=1}^n c_i \varphi_i]\right) dx = \\
&= \textstyle\sum_{i=1}^n \sum_{j=1}^n A_{ij} c_i c_j + 2 \sum_{k=1}^n A_k c_k + A_0 = \\
&= G(c_1, c_2, \ldots, c_n),
\end{aligned}
\tag{10.70}
$$

$$
\begin{aligned}
A_{ij} &= \int_a^b \left(p\varphi_i'\varphi_j' + q\varphi_i\varphi_j\right) dx, \\
A_k &= \int_a^b \left(p\psi'\varphi_k' + q\psi\varphi_k + f\varphi_k\right) dx, \\
A_0 &= \int_a^b \left(p[\psi']^2 + q\psi^2 + 2f\psi\right) dx.
\end{aligned}
\tag{10.71}
$$

Метод Ритца. Задача минимизации функции нескольких переменных...

Необходимое условие экстремума функции $G(c_1, c_2, \ldots, c_n)$

$$\frac{\partial}{\partial c_i} G(c_1^*, c_2^*, \ldots, c_n^*) = 2 \sum_{j=1}^{n} A_{ij} c_j^* + 2A_i = 0 \quad \forall i = 1, 2, \ldots, n,$$

которое равносильно системе линейных алгебраических уравнений

$$\sum_{j=1}^{n} A_{ij} c_j^* = -A_i, \forall i = 1, 2, \ldots, n. \tag{10.72}$$

Теорема 10.2.

Пусть функции $\varphi_1', \varphi_2', \ldots, \varphi_n'$ — линейно-независимы. Тогда система (10.72) имеет единственное решение $(c_1^*, c_2^*, \ldots, c_n^*)^\top \in \mathbb{R}^n$, которое доставляет минимальное значение функции $G(c_1, c_2, \ldots, c_n)$.

Доказательство

Доказательство.

Достаточно показать, что квадратичная форма $\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}c_ic_j$ является положительно-определенной.

Действительно, в этом случае, во-первых, матрица $\{A_{ij}\}_{i,j=1}^{n}$ является положительно-определенной и, поэтому, невырожденной. Следовательно, у системы (10.72) существует единственное решение $(c_1^*, c_2^*, \ldots, c_n^*)^\top \in \mathbb{R}^n$. Во-вторых, элементы A_{ij} этой матрицы вычисляются как значения частных производных второго порядка функции $G(c_1, c_2, \ldots, c_n)$ по переменным c_i и c_j, умноженные на коэффициент 2, и не зависят от значений переменных c_i $(i = 1, 2, \ldots, n)$. Таким образом, для любой точки $(c_1, c_2, \ldots, c_n)^\top \in \mathbb{R}^n$ матрица $\{A_{ij}\}_{i,j=1}^{n}$ для функции G является ее матрицей Гессе[40] $H(G)$. Свойство положительной определенности этой матрицы является достаточным условием того, что решение системы (10.72) является единственной точкой минимума функции G.

[40] с точностью до постоянного положительного множителя

Доказательство

Используя (10.71), нетрудно убедиться, что

$$\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}c_i c_j = \int_a^b \left(p\left(\sum_{i=1}^{n} c_i \varphi_i'\right)^2 + q\left(\sum_{i=1}^{n} c_i \varphi_i\right)^2 \right) dx. \quad (10.73)$$

С учетом свойств функций p и q ($p(x) \geq p_0 > 0$, $q(x) \geq 0$, $x \in [a, b]$) из (10.73) следует, что для любого вектора $(c_1, c_2, \ldots, c_n)^\top \neq \mathbf{0}$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}c_i c_j > 0.$$

Предположение, что

$$\sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}c_i c_j = 0,$$

Доказательство ...

с необходимостью приводит к тождеству

$$p(x)\left(\sum_{i=1}^{n} c_i\varphi_i'(x)\right)^2 + q(x)\left(\sum_{i=1}^{n} c_i\varphi_i(x)\right)^2 \equiv 0, \quad x \in [a,b].$$

Откуда

$$\sum_{i=1}^{n} c_i\varphi_i'(x) \equiv 0, \quad x \in [a,b],$$

что, в силу линейной независимости функций $\varphi_1', \varphi_2', \ldots, \varphi_n'$, означает, что $c_1 = c_2 = \ldots = c_n = 0$.
Возникает противоречие с $(c_1, c_2, \ldots, c_n)^\top \neq \mathbf{0}$.
□ Теорема доказана.

Для метода Ритца теоремы 10.1 и 10.2 дают положительные ответы на поставленные выше первые три вопроса, возникающие при реализации предложенной общей схемы приближенного решения экстремальной задачи (10.63), (10.64).

⊠

▲25. Метод Ритца. Сходимость.

Ответ на четвертый вопрос о существовании предела $\lim_{n\to\infty} y_n^* = y^*$ дает следующая теорема.

Теорема 10.3.

Пусть

$$1)\ p \in C^{(1)}([a,b]):\ p(x) \geq p_0 > 0,\ x \in [a,b];$$
$$2)\ q, f \in C([a,b]):\ q(x) \geq 0,\ x \in [a,b];$$
$$3)\ \forall n \in \mathbb{N}\ \exists y_n \in Y_n:\ J[y_n] \to J[y^*] \text{ при } n \to \infty.$$

Тогда при $n \to \infty$ $y_n(x) \rightrightarrows y^*(x)$ на отрезке $[a,b]$.

Доказательство.

Учитывая, что $J[y_n] \geq J[y^*]$, достаточно показать, что

$$|y_n(x) - y^*(x)| \leq \sqrt{\frac{b-a}{p_0}}\,(J[y_n] - J[y^*])^{\frac{1}{2}} \quad \forall x \in [a,b], \qquad (10.74)$$

где

$J[y_n] - J[y^*] = \int_a^b \left(p[(y_n')^2 - ((y^*)')^2] + q[(y_n)^2 - (y^*)^2] + 2f(y_n - y^*)\right)dx.$

Доказательство ...

Действительно, для любого $x \in [a,b]$ справедливо выражение $y_n(x) - y^*(x) = \int_a^x \left(y_n'(t) - (y^*)'(t)\right)dt$. Откуда

$$|y_n(x)-y^*(x)|\leq \int_a^x |y_n'(t)-(y^*)'(t)|dt\leq \int_a^b |y_n'(t)-(y^*)'(t)|dt. \tag{10.75}$$

С учетом известного неравенства Гельдера

$$\int_a^b f(x)g(x)dx \leq \sqrt{\int_a^b f^2(x)dx}\sqrt{\int_a^b g^2(x)dx}$$

и свойств функций p, q ($p(x) \geq p_0 > 0$, $q(x) \geq 0$, $x \in [a,b]$) из (10.75) следует, что

$$|y_n(x) - y^*(x)| \leq \sqrt{b-a}\sqrt{\int_a^b [y_n'(x) - (y^*)'(x)]^2dx} \leq$$

$$\leq \sqrt{\tfrac{b-a}{p_0}}\sqrt{\int_a^b p(x)[y_n'(x) - (y^*)'(x)]^2dx} \leq \tag{10.76}$$

$$\leq D\sqrt{\int_a^b \left(p[y_n'-(y^*)']^2+q[y_n-y^*]^2\right)dx} = D\sqrt{\int_a^b R(x)dx},$$

Доказательство ...

где $D = \sqrt{\frac{b-a}{p_0}}$, $R(x) = p[y_n' - (y^*)']^2 + q[y_n - y^*]^2$, $x \in [a, b]$.
Далее, если воспользоваться очевидным равенством
$(c-d)^2 = c^2 - d^2 - 2d(c-d)$, то

$$\int_a^b R(x)dx = \int_a^b \left(p[y_n' - (y^*)']^2 + q[y_n - y^*]^2\right)dx =$$

$$= \int_a^b \left(p[(y_n')^2 - ((y^*)')^2] + q[(y_n)^2 - (y^*)^2]\right)dx - \qquad (10.77)$$

$$-2\int_a^b \left(p(y^*)'[y_n' - (y^*)'] + qy^*[y_n - y^*]\right)dx,$$

где

$$\int_a^b p(y^*)'[y_n' - (y^*)']dx = \int_a^b p(y^*)'d(y_n - y^*) =$$

$$= p(y^*)'[y_n - y^*]|_a^b - \int_a^b \left(p(y^*)'\right)'[y_n - y^*]dx =$$

$$= -\int_a^b \left(p(y^*)'\right)'[y_n - y^*]dx,$$

так как $y_n(a) = y^*(a)$ и $y_n(b) = y^*(b)$.

Доказательство ...

Следовательно,

$$\int_a^b \left(p(y^*)'[y_n' - (y^*)'] + qy^*[y_n - y^*]\right)dx =$$

$$= -\int_a^b \left(p(y^*)'\right)'[y_n - y^*]dx + \int_a^b qy^*[y_n - y^*]dx = \qquad (10.78)$$

$$= \int_a^b \left(-\left(p(y^*)'\right)' + qy^*\right)[y_n - y^*]dx.$$

Поскольку функция y^* как решение экстремальной задачи (10.65) является и решением исходной краевой задачи (10.63), (10.64), то есть удовлетворяет дифференциальному уравнению $(py')' - qy = f$, то $-\left(p(y^*)'\right)' + qy^* = -f$. В результате равенство (10.78) принимает вид

$$\int_a^b \left(p(y^*)'[y_n' - (y^*)'] + qy^*[y_n - y^*]\right)dx = -\int_a^b f[y_n - y^*]dx. \qquad (10.79)$$

Тогда из (10.77) и (10.79) следует, что

Доказательство ...

$$\int_a^b R(x)dx = \int_a^b \left(p[y_n' - (y^*)']^2 + q[y_n - y^*]^2\right)dx =$$

$$= \int_a^b \left(p[(y_n')^2 - ((y^*)')^2] + q[(y_n)^2 - (y^*)^2] + 2f[y_n - y^*]\right)dx. \tag{10.80}$$

Подстановка правой части равенства (10.80) в (10.76) приводит к требуемой оценке (10.74)

$$|y_n(x) - y^*(x)| \leq$$

$$\leq D\sqrt{\int_a^b \left(p[(y_n')^2 - ((y^*)')^2] + q[(y_n)^2 - (y^*)^2] + 2f[y_n - y^*]\right)dx} =$$

$$= \sqrt{\frac{b-a}{p_0}}\left(J[y_n] - J[y^*]\right)^{\frac{1}{2}}$$

$$\forall x \in [a, b].$$

□ Теорема доказана.

10.7. Вариационные методы приближенного решения краевой задачи для линейного дифференциального уравнения второго порядка в общем виде

Постановка краевой задачи

$$L[y] = f(x), \quad x \in [a, b], \tag{10.81}$$

$$y(a) = \alpha, \quad y(b) = \beta, \tag{10.82}$$

где L — линейный дифференциальный оператор второго порядка[a].

[a] Примеры: 1) $L[y] = y''$; 2) $L[y] = (py')' - qy$.

Пусть $\varphi \in C^2([a, b])$: $\varphi(a) = \alpha, \quad \varphi(b) = \beta$. Тогда введение в рассмотрение вспомогательной функции

$$z(x) = y(x) - \varphi(x), \quad x \in [a, b]$$

позволяет сформулировать для краевой задачи (10.81), (10.82) равносильную ей краевую задачу с однородными краевыми условиями:

Вариационные методы приближенного решения линейной краевой задачи с однородными краевыми условиями

Постановка однородной краевой задачи

$$L[z] = \bar{f}(x), \quad x \in [a, b], \tag{10.83}$$

$$z(a) = 0, \quad z(b) = 0, \tag{10.84}$$

где, в силу линейности оператора L,

$$\begin{aligned} &L[z + \varphi] = f(x), \\ &L[z] + L[\varphi] = f(x), \\ &L[z] = f(x) - L[\varphi] = \bar{f}(x), \\ &x \in [a, b]. \end{aligned}$$

Без потери общности можно считать, что краевые условия (10.82) в исходной краевой задаче (10.81), (10.82) являются однородными, то есть $\alpha = 0, \quad \beta = 0$.

Метод Галеркина

Пусть Y_n — конечномерное подпространство пространства Y, которое образовано линейными комбинациями базисных (линейно-независимых) функций:

$$Y_n = \{\, y_n \in C^{(2)}([a,b]) \mid y_n(x) = \sum_{i=1}^{n} c_i \varphi_i(x) : \quad \varphi_i \in C^{(2)}([a,b]) \;: \quad \varphi_i(a) = \varphi_i(b) = 0, \quad i = 1, 2, \ldots, n \,\} \tag{10.85}$$

Предлагается находить приближенное решение y_n краевой задачи(10.81), (10.82) в этом пространстве Y_n. Невязкой называют величину

$$\mu(x) = L[y_n](x) - f(x), \quad x \in [a,b]. \tag{10.86}$$

Так как невязка — функция, то можно предложить разные подходы к уменьшению невязки.
Метод Галеркина основал на том, чтобы сделать невязку μ ортогональной некоторой (вообще говоря, другой) системе базисных функций $\{\psi_i,\ i = 1, 2, \ldots, n\}$ из пространства Y_n:

Метод Галеркина ...

$$\langle L[y_n] - f, \psi_i \rangle = 0, \quad i = 1, 2, \ldots, n.$$

Подстановка вместо функции y_n ее разложения по системе базисных функций $\{\varphi_i,\ i = 1, 2, \ldots, n\}$ с учетом линейной независимости функций ψ_j $(j = 1, 2, \ldots, n)$ приводит к системе линейных уравнений относительно коэффициентов разложения c_k $(k = 1, 2, \ldots, n)$:

$$\sum_{k=1}^{n} \langle L[\varphi_k], \psi_i \rangle c_k = \langle f, \psi_i \rangle, \quad i = 1, 2, \ldots, n. \tag{10.87}$$

Замечание

Метод Ритца является частным случаем метода Галеркина, если, во-первых, уравнение самосопряженное, а, во-вторых, системы базисных функций $\{\varphi_i,\ i = 1, 2, \ldots, n\}$ и $\{\psi_i,\ i = 1, 2, \ldots, n\}$ совпадают.

Метод наименьших квадратов

Приближенное решение y_n краевой задачи(10.81), (10.82) ищется в пространстве Y_n (см. (10.85)). Для определения искомого элемента этого пространства предлагается минимизировать квадрат нормы невязки (10.86), то есть скалярное произведение невязки (10.86) самой на себя

$$G(c_1, c_2, \ldots, c_n) = \|L[y_n] - f\|^2 = \langle L[y_n] - f, L[y_n] - f\rangle =$$

$$= \langle \sum_{k=1}^{n} c_k L[\varphi_k] - f, \sum_{i=1}^{n} c_i L[\varphi_i] - f\rangle =$$

$$= \sum_{k=1}^{n}\sum_{i=1}^{n} \langle L[\varphi_k], L[\varphi_i]\rangle c_k c_i - 2\sum_{k=1}^{n} \langle L[\varphi_k], f\rangle c_k + \langle f, f\rangle. \tag{10.88}$$

Выражение (10.88) представляет собой линейно-квадратичную форму относительно коэффициентов $c_1, c_2, \ldots, c_n$. Необходимое условие экстремума функции $G(c_1, c_2, \ldots, c_n)$

$$\frac{\partial}{\partial c_i} G(c_1, c_2, \ldots, c_n) = 0 \quad \forall i = 1, 2, \ldots, n$$

может быть записано в виде системы линейных уравнений

Метод наименьших квадратов ...

$$\sum_{j=1}^{n} a_{ij} c_j = b_i, \ \ i = 1, 2, \ldots, n, \tag{10.89}$$

где

$$a_{ij} = \langle L[\varphi_i], L[\varphi_j] \rangle, \ \ b_i = \langle L[\varphi_i], f \rangle, \ \ i, j = 1, 2, \ldots, n.$$

При этом матрица $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \cdots & \cdots & \cdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$ системы (10.89) является положительно-определенной тогда и только тогда, когда $L[\varphi_1], L[\varphi_2], \ldots, L[\varphi_n]$ линейно-независимы. Если это условие выполняется, то метод наименьших квадратов сводится к решению системы (10.89).

Замечание

Метод наименьших квадратов является частным случаем метода Галеркина, если в качестве системы базисных функций $\{\psi_i, \ i = 1, 2, \ldots, n\}$ выбрать $\{L[\varphi_i], \ i = 1, 2, \ldots, n\}$.

Метод коллокации

Приближенное решение y_n краевой задачи(10.81), (10.82) ищется в пространстве Y_n (см. (10.85)). Для определения искомого элемента этого пространства предлагается занулить невязку (10.86) в некоторых узлах x_i $(i = 1, 2, \ldots, n)$, называемых узлами коллокации.

Пусть узлы коллокации $a \le x_1 \le x_2 \le \ldots \le x_n \le b$. Тогда

$$L[y_n](x_i) = f(x_i), \ \forall i = 1, 2, \ldots, n. \tag{10.90}$$

Равенства (10.90) образуют систему линейных алгебраических уравнений относительно коэффициентов $c_1, c_2, \ldots, c_n$

$$\sum_{k=1}^{n} c_k L[\varphi_k](x_i) = f(x_i), \ i = 1, 2, \ldots, n. \tag{10.91}$$

Замечание

Основная проблема в методе коллокации состоит в таком выборе узлов, чтобы матрица системы (10.91) была невырожденной, а также в обосновании сходимости метода при данном способе увеличения количества узлов.

Постановка задачи

Пусть заданы набор данных

$$\begin{matrix} x_0, x_1, x_2, \ldots, x_n \in [a,b]: & x_i \neq x_j \quad \forall i \neq j \\ y_0, y_1, y_2, \ldots, y_n \end{matrix} \tag{11.1}$$

и класс функций Φ.

$$\varphi \in \Phi: \ \varphi(x_i) = y_i \quad \forall i = 0, 1, 2, \ldots, n. \tag{11.2}$$

Задача численной интерполяции (11.2) на классе многочленов степени n может быть решена с помощью интерполяционного многочлена Лагранжа $L_n(x)$, задаваемого с помощью формулы (6.12). Однако такая конструкция обладает рядом существенных недостатков, которые проявляются в приложениях при больших значениях n.

11.1. Интерполяционный кубический сплайн.

Сплайнами называют функции, заданные кусочно многочленами на отрезках $[x_{i-1}, x_i]$, $i = 1, 2, \ldots, n$. Среди всех сплайнов наибольшее распространение получили интерполяционные кубические сплайны.

Определение 11.1.

Интерполяционным кубическим сплайном $S(x)$ называется дважды непрерывно-дифференцируемая внутри отрезка $[a, b]$ функция, заданная на каждом отрезке $[x_{i-1}, x_i]$ $(i = 1, 2, \ldots, n)$ кубическим многочленом и удовлетворяющая интерполяционным условиям (11.2).

Интерполяционный кубический сплайн $S(x)$ определяется на отрезке $[a, b]$ следующими формулами:

$$S(x) = p_i(x) = A_i x^3 + B_i x^2 + C_i x + D_i, \quad x \in [x_{i-1}, x_i], \quad i = 1, 2, \ldots, n. \tag{11.3}$$

Интерполяционный кубический сплайн ...

Таким образом, интерполяционный кубический сплайн $S(x)$ задается на отрезке $[a, b]$ с помощью $4n$ параметров: (A_i, B_i, C_i, D_i), $i = 1, 2, \ldots, n$. Значения этих параметров определяются следующими условиями:

$$p_i(x_{i-1}) = y_{i-1},\ p_i(x_i) = y_i, \quad i = 1, 2, \ldots, n; \tag{11.4}$$

$$p_i'(x_i) = p_{i+1}'(x_i), \quad i = 1, 2, \ldots, n-1; \tag{11.5}$$

$$p_i''(x_i) = p_{i+1}''(x_i), \quad i = 1, 2, \ldots, n-1. \tag{11.6}$$

В результате, для определения значений $4n$ параметров имеется всего $2n + 2(n-1) = 4n - 2$ условий. Следовательно для однозначного определения сплайна S к (11.4), (11.5), (11.6) требуется добавить еще два дополнительных условия. На концах отрезка задают краевые условия[41], например:

$$S'(a) = y_0',\ S'(b) = y_n' \tag{11.7}$$

или

$$S''(a) = y_0'',\ S''(b) = y_n''. \tag{11.8}$$

[41] Условия (11.7) называют краевыми условиями первого типа, (11.8) — второго типа.

11.2. Эффективный способ построения интерполяционного кубического сплайна.

Так как сплайн на каждом отрезке $[x_{i-1}, x_i]$ задается кубической функцией, то его вторая производная является линейной функцией. Пусть $M_i = S''(x_i)$ и $h_i = x_i - x_{i-1}$, $i = 1, 2, \ldots, n$. Тогда

$$S''(x) = M_{i-1}\frac{x_i - x}{h_i} + M_i\frac{x - x_{i-1}}{h_i}, \ x \in [x_{i-1}, x_i]. \qquad (11.9)$$

Интегрирование обеих частей равенства (11.9) приводит к выражениям

$$S'(x) = -M_{i-1}\frac{(x_i-x)^2}{2h_i} + M_i\frac{(x-x_{i-1})^2}{2h_i} + R_i,$$

$$S(x) = M_{i-1}\frac{(x_i-x)^3}{6h_i} + M_i\frac{(x-x_{i-1})^3}{6h_i} + R_i x + H_i, \ x \in [x_{i-1}, x_i],$$

где R_i, H_i — константы интегрирования. Для их определения из интерполяционных условий (11.4) удобно выполнить невырожденную замену констант

Эффективный способ построения интерполяционного кубического сплайна ...

$$R_i x + H_i = a_i(x_i - x) + b_i(x - x_{i-1}).$$

Здесь $\begin{cases} -a_i + b_i = R_i, \\ x_i a_i - x_{i-1} b_i = H_i, \end{cases}$, где определитель матрицы этой системы равен $det\begin{pmatrix} -1 & 1 \\ x_i & -x_{i-1} \end{pmatrix} = x_{i-1} - x_i = -h_i \neq 0$. Тогда

$$S(x) = p_i(x) = M_{i-1}\frac{(x_i - x)^3}{6h_i} + M_i\frac{(x - x_{i-1})^3}{6h_i} + \\ + a_i(x_i - x) + b_i(x - x_{i-1}), \quad (11.10)$$

$x \in [x_{i-1}, x_i]$.

Из (11.10) с учетом условий интерполяции (11.4) следует

$$\begin{cases} p_i(x_{i-1}) = M_{i-1}\frac{h_i^2}{6} + a_i h_i = y_{i-1}, \\ p_i(x_i) = M_i\frac{h_i^2}{6} + b_i h_i = y_i. \end{cases}$$

Эффективный способ построения интерполяционного кубического сплайна ...

Откуда

$$\begin{cases} a_i = \frac{y_{i-1}}{h_i} - M_{i-1}\frac{h_i}{6}, \\ b_i = \frac{y_i}{h_i} - M_i\frac{h_i}{6}. \end{cases}$$

В результате

$$p_i(x) = M_{i-1}\frac{(x_i-x)^3}{6h_i} + M_i\frac{(x-x_{i-1})^3}{6h_i} +$$

$$+\left(\frac{y_{i-1}}{h_i} - M_{i-1}\frac{h_i}{6}\right)(x_i - x) + \left(\frac{y_i}{h_i} - M_i\frac{h_i}{6}\right)(x - x_{i-1}),$$

$$x \in [x_{i-1}, x_i],\ i = 1, 2, \ldots, n. \tag{11.11}$$

Для того чтобы получить уравнения для M_i $(i = 0, 1, 2, \ldots, n)$, предлагается воспользоваться условиями (11.5)

$$p_i'(x_i) = p_{i+1}'(x_i), \quad i = 1, 2, \ldots, n-1.$$

Эффективный способ построения интерполяционного кубического сплайна ...

Из (11.11) следует, что

$$p_i'(x_i) = \frac{h_i}{2} M_i + \frac{y_i - y_{i-1}}{h_i} - \frac{h_i}{6} M_i + \frac{h_i}{6} M_{i-1},$$

$$p_{i+1}'(x_i) = -\frac{h_{i+1}}{2} M_i + \frac{y_{i+1} - y_i}{h_{i+1}} - \frac{h_{i+1}}{6} M_{i+1} + \frac{h_{i+1}}{6} M_i.$$

Подстановка правых частей этих равенств в (11.5) после приведения подобных приводит к линейной системе из $n-1$-го уравнения относительно $n+1$-ой неизвестной

$$\frac{h_i}{6} M_{i-1} + \frac{h_i + h_{i+1}}{3} M_i + \frac{h_{i+1}}{6} M_{i+1} = \frac{y_{i+1} - y_i}{h_{i+1}} - \frac{y_i - y_{i-1}}{h_i}, \tag{11.12}$$

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

В результате добавления к (11.12) краевых условий (11.8)

$$S''(a) = y_0'', \; S''(b) = y_n''$$

Эффективный способ построения интерполяционного кубического сплайна ...

возникает система крамеровского типа

$$\begin{cases} M_0 = y_0'', \\ \frac{h_i}{6} M_{i-1} + \frac{h_i + h_{i+1}}{3} M_i + \frac{h_{i+1}}{6} M_{i+1} = \frac{y_{i+1} - y_i}{h_{i+1}} - \frac{y_i - y_{i-1}}{h_i}, \ i = 1, 2, \ldots, n-1 \\ M_n = y_n'', \end{cases} \tag{11.13}$$

матрица которой обладает свойством строго диагонального преобладания.

Единственное решение системы (11.13) для исходных данных (11.1) и краевых условий (11.8) определяет единственный интерполяционный кубический сплайн $S(x)$, $x \in [a, b]$, который на отрезке $[a, b]$ задается формулами (11.11).

11.3. Экстремальное свойство интерполяционного кубического сплайна

Постановка экстремальной задачи

$$J[y] \to \min_{y \in Y}, \tag{11.14}$$

где

$$J[y] = \int_a^b [f''(x)]^2 dx, \tag{11.15}$$

$$Y = \{y \in C^{(2)}([a,b]) \mid y(x_i) = y_i, \quad i = 0, 1, 2, \ldots, n\}, \tag{11.16}$$

$$a = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b.$$

Механическая интерпретация задачи (11.14)-(11.16) состоит в том, что нужно минимизировать потенциальную энергию упругого гибкого тела, например, металлической линейки, с закрепленными точками.

Экстремальное свойство интерполяционного кубического сплайна ...

Теорема 11.1. Теорема Холлидея

Единственным решением экстремальной задачи (11.14)-(11.16) является интерполяционный кубический сплайн $S(x)$, $x \in [a, b]$, с дополнительными условиями

$$S''(a) = S''(b) = 0.$$

$\boxtimes$

▲26. Метод наименьших квадратов

12.1. Метод наименьших квадратов в задаче решения линейных систем.

Система линейных алгебраических уравнений

$$A\boldsymbol{x} = \boldsymbol{b}, \tag{12.1}$$

где $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$ $(dim(A) = m \times n)$: $m \neq n$.

Задача минимизации квадрата невязки

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} J(\boldsymbol{x}), \tag{12.2}$$

где

$$J(\boldsymbol{x}) = \|A\boldsymbol{x} - \boldsymbol{b}\|^2. \tag{12.3}$$

Используя свойства скалярного произведения, минимизируемую величину можно представить в виде

$$J(\boldsymbol{x}) = \|A\boldsymbol{x} - \boldsymbol{b}\|^2 = \langle A\boldsymbol{x} - \boldsymbol{b}, A\boldsymbol{x} - \boldsymbol{b}\rangle = \langle A\boldsymbol{x}, A\boldsymbol{x}\rangle - 2\langle A\boldsymbol{x}, \boldsymbol{b}\rangle + \langle \boldsymbol{b}, \boldsymbol{b}\rangle = \\ = \langle \boldsymbol{x}, A^\top A\boldsymbol{x}\rangle - 2\langle \boldsymbol{x}, A^\top \boldsymbol{b}\rangle + \langle \boldsymbol{b}, \boldsymbol{b}\rangle.$$

Метод наименьших квадратов в задаче решения линейных систем ...

Тогда необходимое условие экстремума для функции $J(\boldsymbol{x})$

$$\frac{\partial}{\partial x_i} J(\boldsymbol{x}^*) = 0 \quad \forall i = 1, 2, \ldots, n$$

равносильно следующей системе линейных алгебраических уравнений (МНК-системе)

$$A^\top A \boldsymbol{x} = A^\top \boldsymbol{b}. \tag{12.4}$$

Здесь матрица $A^\top A \in \mathbb{R}^{n \times n}$ и является неотрицательно-определенной.

Невырожденный метод наименьших квадратов

Пусть $A^\top A > 0$. Тогда система (12.4) имеет единственное решение — МНК-решение системы (12.1)

$$\boldsymbol{x}^* = \left(A^\top A\right)^{-1} A^\top \boldsymbol{b} \tag{12.5}$$

Невырожденный метод наименьших квадратов ...

Теорема 12.1.

Матрица $A^\top A$ является положительно определенной тогда и только тогда, когда столбцы $\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_n$ матрицы A линейно независимы, т. е. ранг матрицы A равен n.[a]

[a] Это условие называется невырожденным МНК.

Доказательство.
Вектор $A\boldsymbol{x}$ можно записать в следующем виде $A\boldsymbol{x} = \sum_{i=1}^{n} \boldsymbol{h}_i x_i$.
Тогда $\langle \boldsymbol{x}, A^\top A\boldsymbol{x} \rangle = \langle A\boldsymbol{x}, A\boldsymbol{x} \rangle = 0$ тогда и только тогда, когда

$$\sum_{i=1}^{n} \boldsymbol{h}_i x_i = \mathbf{0}.$$

А это, в свою очередь, означает, что матрица $A^\top A$ является положительно определенной тогда и только тогда, когда столбцы $\boldsymbol{h}_1, \boldsymbol{h}_2, \ldots, \boldsymbol{h}_n$ матрицы A линейно независимы, т. е. ранг матрицы A равен n.
□ Теорема доказана.

Невырожденный метод наименьших квадратов ...

Пример

Переопределенная несовместная[a] система

$$\begin{cases} x_1 + x_2 = 1 \\ x_1 = 1 \\ x_2 = 1 \end{cases} \tag{12.6}$$

[a] в классическом смысле

Здесь

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Тогда

$$A^\top = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, A^\top A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, A^\top \boldsymbol{b} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}.$$

МНК-система для (12.6) имеет вид

$$\begin{cases} 2x_1 + x_2 = 2 \\ x_1 + 2x_2 = 2. \end{cases} \tag{12.7}$$

Невырожденный метод наименьших квадратов ...

Нетрудно видеть, что система (12.7) невырождена и ее решение

МНК-решение системы (12.6)

$$\begin{cases} x_1^* = \frac{2}{3} \\ x_2^* = \frac{2}{3}. \end{cases}$$

Сумма квадратов расстояний от точки $\boldsymbol{x}^* = \left(\frac{2}{3}, \frac{2}{3}\right)^\top$ до прямых, задаваемых уравнениями системы (12.6), — наименьшая среди всех возможных точек (см. рис. 20).

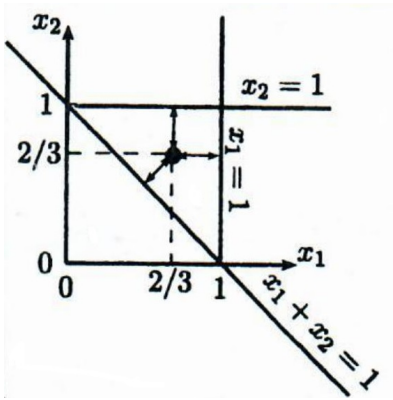


Рис. 20: Геометрическая интерпретация МНК-решения системы.

Вырожденный метод наименьших квадратов

Пусть теперь наблюдается вырождение МНК, т.е. $rang(A) < n$. В частности, вырожденный МНК имеет место, когда $m < n$, т.е. число уравнений в системе (12.1) меньше числа неизвестных, что приводит к неединственности решения. Задача относится к числу некорректных и регуляризированный МНК позволяет отыскать решение с минимальной нормой.

Алгоритм вырожденного метода наименьших квадратов

К минимизируемой функции $\langle A\boldsymbol{x} - \boldsymbol{b}, A\boldsymbol{x} - \boldsymbol{b}\rangle$ (квадрату невязки) добавляется квадрат нормы $\boldsymbol{x}$ с некоторым положительным коэффициентом α (параметр регуляризации). Таким образом, минимизируемая величина представляется в виде

$$J_\alpha(\boldsymbol{x}) = \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \alpha\|\boldsymbol{x}\|^2 = \langle A\boldsymbol{x} - \boldsymbol{b}, A\boldsymbol{x} - \boldsymbol{b}\rangle + \alpha\langle \boldsymbol{x}, \boldsymbol{x}\rangle =$$

$$= \langle A\boldsymbol{x}, A\boldsymbol{x}\rangle - 2\langle A\boldsymbol{x}, \boldsymbol{b}\rangle + \langle \boldsymbol{b}, \boldsymbol{b}\rangle + \alpha\langle \boldsymbol{x}, \boldsymbol{x}\rangle =$$

$$= \langle \boldsymbol{x}, \left(A^\top A + \alpha E\right)\boldsymbol{x}\rangle - 2\langle \boldsymbol{x}, A^\top \boldsymbol{b}\rangle + \langle \boldsymbol{b}, \boldsymbol{b}\rangle.$$

Алгоритм вырожденного метода наименьших квадратов ...

Аналогично (12.4) строится МНК-система

$$\left(A^\top A + \alpha E\right) \boldsymbol{x} = A^\top \boldsymbol{b}, \tag{12.8}$$

которая имеет единственное решение

$$\boldsymbol{x}_\alpha^* = \left(A^\top A + \alpha E\right)^{-1} A^\top \boldsymbol{b}$$

при любом $\alpha > 0$.
Если существует предел последовательности векторов $\boldsymbol{x}_\alpha^*$ при $\alpha \to 0$, то он называется вырожденным МНК-решением линейной системы (12.1).
Это решение формально записывается в виде

$$\boldsymbol{x}^* = \lim_{\alpha \to 0} \left(A^\top A + \alpha E\right)^{-1} A^\top \boldsymbol{b}. \tag{12.9}$$

Матрица псевдообратная к матрице A

По аналогии с обратной матрицей A^{-1}, с помощью которой можно записать решение системы $\boldsymbol{x} = \boldsymbol{b}$ в случае невырожденной матрицы A в виде $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ в вырожденном методе наименьших квадратов можно определить матрицу A^+, которую называют псевдообратной к матрице A

$$A^+ = \lim_{\alpha \to 0} \left(A^\top A + \alpha E\right)^{-1} A^\top \tag{12.10}$$

Свойство A^+

Матрица A^+A представляет собой диагональную матрицу, у которой на главной диагонали стоит несколько единиц, а все остальные элементы — нули.

МНК-решение в вырожденном случае

$$\boldsymbol{x}^* = A^+\boldsymbol{b}$$

12.2. Метод наименьших квадратов в задаче приближения функций (дискретный вариант)

Постановка задачи

$$x_1, x_2, \ldots, x_m \\ y_1, y_2, \ldots, y_m \tag{12.11}$$

$$\Phi = \{\ \varphi \mid \varphi(x) = \sum_{i=1}^{n} c_i \varphi_i(x)\ \} \tag{12.12}$$

Требуется в классе функций Φ выбрать функцию f, наилучшим образом приближающую значения в узлах x_i, $i = 1, 2, \ldots, m$.

Если в Φ существует функция f, в точности удовлетворяющая интерполяционным условиям $f(x_j) = y_j$, $j = 1, 2, \ldots, m$, то решается задача интерполяции. В более общем случае, когда интерполяционные условия не выполняются, то рассматривается задача приближения, в которой минимизируются отклонения от этих условий. В методе наименьших квадратов в качестве критерия используется сумма квадратов отклонений в узлах

Дискретный вариант...

$$J(c_1, c_2, \ldots, c_n) = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} c_i \varphi_i(x_j) - y_j \right)^2. \qquad (12.13)$$

Вводится скалярное произведение в пространстве функций, определенных на узлах

$$\langle f(x), g(x) \rangle = \sum_{j=1}^{m} f(x_j) g(x_j).$$

Тогда минимизируемый критерий (12.13) можно переписать в виде

$$J(c_1, c_2, \ldots, c_n) = \langle \textstyle\sum_{i=1}^{n} c_i \varphi_i(x) - Y, \sum_{i=1}^{n} c_i \varphi_i(x) - Y \rangle =$$

$$= \textstyle\sum_{i=1}^{n} \sum_{k=1}^{n} c_i c_k \langle \varphi_i, \varphi_k \rangle - \qquad (12.14)$$

$$-2 \textstyle\sum_{i=1}^{n} c_i \langle \varphi_i, Y \rangle + \langle Y, Y \rangle,$$

где $Y = (y_1, y_2, \ldots, y_m)^\top \in \mathbb{R}^m$.

Дискретный вариант...

Таким образом, минимизируемый критерий (12.13) представим в виде суммы квадратичной формы с симметрической матрицей $A = \{\langle \varphi_i, \varphi_j \rangle\}_{i,j=1}^{n}$, однородной формы и константы и может рассматриваться как функция n переменных.

Необходимое условие экстремума функции $J(c_1, c_2, \ldots, c_n)$

$$\frac{\partial}{\partial c_i} J(c_1, c_2, \ldots, c_n) = 0 \quad \forall i = 1, 2, \ldots, n,$$

равносильно системе линейных алгебраических уравнений

$$\sum_{j=1}^{n} \langle \varphi_i, \varphi_j \rangle c_j = \langle \varphi_i, Y \rangle \quad \forall i = 1, 2, \ldots, n. \tag{12.15}$$

Если матрица A системы (12.15) невырождена, то ее единственное решение $c_1^*, c_2^*, \ldots, c_n^*$ определяет искомую функцию

$$f^*(x) = \sum_{i=1}^{n} c_i^* \varphi_i(x).$$

Дискретный вариант...

Замечание

Удобство такого варианта задачи приближения состоит в том, что элементы матрицы и вектора правых частей системы (12.15) легко считать, так они представляют собой конечномерные скалярные произведения, т.е. конечные суммы. В то же время этот метод имеет существенный недостаток: при увеличении числа n базисных функций φ_i, $i = 1, 2, \ldots, n$, метод наименьших квадратов вырождается.

$\boxtimes$

▲27. Численные методы решения интегральных уравнений

Задача решения интегральных уравнений возникает как вспомогательная при решении краевых задач для дифференциальных уравнений с частными производными.

Как самостоятельная задача решение интегральных уравнений возникает во многих прикладных задачах. Например, исследования работы ядерных реакторов, при решении обратных задач геофизики, при обработке результатов наблюдений и т.п.

Типы интегральных уравнений

Интегральное уравнение Фредгольма

первого рода

$$Iy = \int_a^b K(x,s)y(s)ds = f(x) \tag{13.1}$$

и второго рода

$$y - \lambda Iy = y(x) - \lambda \int_a^b K(x,s)y(s)ds = f(x). \tag{13.2}$$

Интегральное уравнение Вольтерра

первого рода

$$Iy = \int_a^x K(x,s)y(s)ds = f(x) \tag{13.3}$$

и второго рода

$$y - \lambda Iy = y(x) - \lambda \int_a^x K(x,s)y(s)ds = f(x). \tag{13.4}$$

Функция $K(x,s)$ в (13.1)-(13.3) называется ядром интегрального оператора I.

Постановка задачи

Пусть заданы некоторые вещественные функции $f(x)$ ($x \in [a, b] \subseteq \mathrm{D}[f]$) и $K(x, s)$ ($(x, s) \in [a, b] \times [a, b] \subseteq \mathrm{D}[K]$), а также — некоторое число $\lambda \in \mathbb{R}$. Требуется найти функцию $y(x)$ ($x \in [a, b]$), которая является решением интегрального уравнения[a]:

[a] Подстановка функции y в соответствующее уравнение равенство приводит к верному по переменной x тождеству.

Задача на собственные значения и собственные функции ядра $K(x,s)$

Определение 13.1.

Если в равенствах (13.1)-(13.3) правая часть $f(x) \equiv 0$, то соответствующие интегральные уравнения называются однородными. В противном случае — неоднородными.

Для однородных интегральных уравнений Фредгольма второго рода (13.2) можно рассмотреть задачу на собственные значения и собственные функции ядра $K(x,s)$.

Постановка задачи на собственные значения и собственные функции ядра $K(x,s)$

$$y - \lambda I y = 0 \tag{13.5}$$

Требуется определить числа λ, при которых интегральное уравнение (13.5) имеет нетривиальное решение $y(x)$ $(x \in [a,b])$[a].

[a] Множество собственных значений ядра $K(x,s)$ называется спектром интегрального оператора I.

Метод разложения интегрального оператора по его спектру

Рассматривается неоднородное интегральное уравнение Фредгольма второго рода (13.2), которое в операторной форме имеет следующий вид

$$(E - \lambda I)\, y = f, \tag{13.6}$$

где E — единичный оператор $(Ey = y)$.

Пусть $\lambda : ||\lambda I|| < 1$.

Тогда $\exists\, (E - \lambda I)^{-1} = \sum_{i=0}^{\infty} (\lambda I)^k$ (см. (4.64)).
В этом случае решение уравнения (13.1) определяется формулой

$$y = (E - \lambda I)^{-1} f = \sum_{i=0}^{\infty} (\lambda I)^k f \tag{13.7}$$

Решение уравнений Фредгольма методом замены интеграла конечной суммой

Пусть ядро $K(x,s)$ и правая часть $f(x)$ имеют непрерывные производные до некоторого порядка[a].

[a] Тогда и решение интегрального уравнения имеет производные до того же порядка.

Для решения интегральных уравнений Фредгольма (13.1) и (13.2) можно применить метод замены интеграла, входящего в уравнения, конечной суммой, воспользовавшись для этого теми или иными квадратурными формулами.
Пусть за основу принята некоторая квадратурная формула

$$\int_a^b F(x)dx = \sum_{j=0}^{n} A_j F(x_j) + R_n[F], \qquad (13.8)$$

где узлы $x_0, x_1, x_2, \ldots, x_n \in [a,b]$ $(x_i \neq x_j \ \forall i \neq j)$ и коэффициенты $A_0, A_1, A_2, \ldots, A_n$ не зависят от выбора функции $F(x)$, а $R_n[F]$ – остаточный член квадратурной формулы.

Решение уравнения Фредгольма второго рода методом замены интеграла конечной суммой ...

Если в интегральном уравнении Фредгольма второго рода (13.2) положить $x = x_i$ $(i = 0, 1, 2, \ldots, n)$, то

$$y(x_i) - \lambda \int_a^b K(x_i, s) y(s) ds = f(x_i) \quad i = 0, 1, 2, \ldots, n. \qquad (13.9)$$

Замена в (13.9) интеграла с помощью квадратурной формулы (13.8) приводит в равенствам

$$y(x_i) - \lambda \sum_{j=0}^{n} A_j K(x_i, x_j) y(x_j) = f(x_i) + \lambda R_n^{(i)},$$

$$R_n^{(i)} = R_n[K(x_i, s) y(s)] \qquad (13.10)$$

$$i = 0, 1, 2, \ldots, n.$$

Решение уравнения Фредгольма второго рода методом замены интеграла конечной суммой ...

Отбрасывание в равенствах (13.10) слагаемых $\lambda R_n^{(i)}$ приводит к системе линейных алгебраических уравнений относительно неизвестных приближенных значений Y_i искомого решения $y(x)$ в узлах $x_0, x_1, x_2, \ldots, x_n \in [a, b]$

$$Y_i - \lambda \sum_{j=0}^{n} A_j K_{ij} Y_j = f_i \qquad (13.11)$$

$$i = 0, 1, 2, \ldots, n,$$

где $K_{ij} = K(x_i, x_j)$, $f_i = f(x_i)$ $\forall i, j = 0, 1, 2, \ldots, n$.

Приближенное решение интегрального уравнения Фредгольма второго рода на всем отрезке $[a, b]$

Приближенное решение интегрального уравнения (13.2) на всем отрезке $[a, b]$ можно построить в результате интерполяции по решению $Y_0, Y_1, Y_2, \ldots, Y_n$ системы (13.11).

Решение уравнения Фредгольма второго рода методом замены интеграла конечной суммой ...

Аналитическое выражение приближенного решения интегрального уравнения Фредгольма второго рода на всем отрезке $[a, b]$

За аналитическое выражение приближенного решения интегрального уравнения (13.2) можно принять функцию

$$Y(x) = f(x) + \lambda \sum_{j=0}^{n} A_j K(x, x_j) Y_j \quad x \in [a, b], \tag{13.12}$$

принимающую в узлах $x_0, x_1, x_2, \ldots, x_n$ значения $Y_0, Y_1, Y_2, \ldots, Y_n$.

В случае уравнений Фредгольма первого рода (13.1) система (13.11) имеет вид

$$\sum_{j=0}^{n} A_j K_{ij} Y_j = f_i \quad i = 0, 1, 2, \ldots, n. \tag{13.13}$$

Задача на собственные значения и собственные функции ядра $K(x,s)$ для однородного интегрального уравнения Фредгольма второго рода

В случае однородного интегрального уравнения Фредгольма второго рода (13.2) ($f(x) \equiv 0$) система (13.11) является однородной

$$Y_i - \lambda \sum_{j=0}^{n} A_j K_{ij} Y_j = 0 \quad i = 0, 1, 2, \ldots, n. \tag{13.14}$$

Однородная система (13.14) будет иметь нетривиальное решение тогда и только тогда, когда ее определитель равен нулю. Приравнивание к нулю определителя системы (13.14) приводит к алгебраическому уравнению степени $n+1$ относительно λ. Корни $\bar{\lambda}_0, \bar{\lambda}_1, \bar{\lambda}_2, \ldots, \bar{\lambda}_n$ этого уравнения определяют приближенные значения первых $n+1$ собственных значений ядра $K(x,s)$. Подставляя в однородную систему (13.14) одно из из найденных значений $\bar{\lambda}_i$ и находя линейно-независимые решения этой системы, можно получить приближения к линейно-независимым собственным функциям ядра $K(x,s)$, соответствующим данному собственному значению.

Решение уравнения Фредгольма второго рода методом замены интеграла конечной суммой ...

Заключительные замечания

При выборе квадратурной формулы в методе замены интеграла конечной суммой нужно иметь в виду, что чем более точная формула применяется, тем большая гладкость ядра $K(x,s)$ и правой части $f(x)$ интегрального уравнения Фредгольма требуется. При несоблюдении этого условия попытка применения более точных квадратурных формул для получения более точного приближения искомого решения может привести к совсем обратному результату.

Может оказаться полезным следующий прием. Если ядро $K(x,s)$ гладкое, а правая часть $f(x)$ имеет особенности, то можно вместо $y(x)$ ввести новую неизвестную функцию

$$z(x) = y(x) - f(x).$$

Подстановка ее в исходное уравнение, например, в интегральное уравнение Фредгольма второго рода (13.2), дает

Решение уравнения Фредгольма второго рода методом замены интеграла конечной суммой. Заключительные замечания.

$$z(x) - \lambda \int_a^b K(x,s)z(s)ds = \lambda \int_a^b K(x,s)f(s)ds.$$

В результате строится интегральное уравнение такого же вида, в котором правая часть $\lambda \int_a^b K(x,s)f(s)ds$ будет уже более гладкой, а следовательно, и решение $z(x)$ будет более гладким.
Решив это интегральное уравнение, то есть построив его решение $\bar{z}(x)$, можно затем найти и искомое решение $\bar{y}(x)$ исходного уравнения

$$\bar{y}(x) = \bar{z}(x) + f(x).$$

$\boxtimes$