УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ
«Брестский государственный университет имени А.С. Пушкина»

МЕТОДЫ ЧИСЛЕННОГО АНАЛИЗА

*Электронный курс лекций
для студентов специальностей
1-31 03 03-01 «Прикладная математика» и
1-31 03 06-01 «Экономическая кибернетика»
физико-математического факультета*

Брест
БрГУ имени А.С. Пушкина
2013

*Кафедры
ПМиТП
АиГ*

Рецензенты:

заведующий кафедрой интеллектуальных информационных технологий учреждения образования «Брестский государственный технический университет», доктор технических наук, профессор

В.А. Головко

профессор кафедры теоретической физики и астрономии учреждения образования «Брестский государственный университет имени А.С. Пушкина», доктор физико-математических наук, профессор

В.А. Плетюхов

Савчук, В.Ф.

Методы численного анализа : электрон. курс лекций для студ. специальностей 1-31 03 03-01 «Прикладная математика» и 1-31 03 06-01 «Экономическая кибернетика» физ.-мат. фак. / В.Ф. Савчук, О.В. Матысик ; Брест. гос. ун-т им. А.С. Пушкина, каф. ПМ и ТП, каф. алгебры и геометрии. – Брест : электрон. издание БрГУ, 2013. – 403 с.

Электронный курс лекций написан в соответствии с действующей типовой программой по дисциплине «Методы численного анализа» и ставит своей целью облегчить самостоятельную работу студентов с теоретическим материалом при подготовке к лекциям, практическим занятиям и экзамену.

Предназначено для студентов специальностей 1-31 03 03-01 «Прикладная математика» и 1-31 03 06-01 «Экономическая кибернетика» физико-математического факультета.

СОДЕРЖАНИЕ

Предисловие 11

Глава 1 **РЕШЕНИЕ НЕЛИНЕЙНЫХ УРАВНЕНИЙ** **12**
§1. Итерационные методы. Исследование уравнения 12
§2. Метод простой итерации для решения нелинейных уравнений. Теорема о сходимости 16
2.1 Геометрический смысл 16
§3. Ускорение сходимости метода итераций 22
§4. Метод Ньютона (касательных) решения уравнений с одним неизвестным 25
4.1 Геометрический смысл метода Ньютона (касательных) 27
§5. Решение систем нелинейных уравнений. Метод простых итераций. Метод Зейделя 29
§6. Метод Ньютона для нелинейных систем уравнений 34
§7. Сведение решения системы нелинейных уравнений к решению вариационной задачи. Метод покоординатного спуска 37
7.1 Метод покоординатного спуска 38
7.2 Метод градиентного спуска 39

Глава 2 **ПРИБЛИЖЕНИЕ ФУНКЦИЙ** **40**
§8. Интерполирование функций 40

Кафедры
ПМиТП
АиГ

§9. Интерполяционный многочлен Лагранжа 42
§10. Конечные разности. Разделённые разности 45
10.1 Свойства конечных разностей 46
§11. Интерполяционные многочлены Ньютона 48
§12. Интерполирование внутри таблицы. Интерполяционная формула Стирлинга 52
§13. Интерполирование с кратными узлами 55
§14. Многочлены Чебышева 57
14.1 Приложение многочленов Чебышева 59
§15. Численное дифференцирование. (Применение интерполирования к вычислению производных) 60
§16. Некоторые частные формулы вычисления производных . . 63
§17. Интерполяционные методы решения нелинейных уравнений 66
§18. Интерполяция и приближение сплайнами 68
§19. Построение кубического сплайна 71
§20. Многомерная интерполяция 75
20.1 Двумерная интерполяция 75
§21. Наилучшее приближение функции в линейном нормированном пространстве 78
§22. Наилучшие приближения функций в гильбертовом пространстве 81
22.1 Сведение к алгебраической задаче о минимуме квадратичного функционала 82

§23. Эмпирические формулы. Метод наименьших квадратов . . 86
23.1 Метод наименьших квадратов 87
23.2 Определение параметров эмпирических формул по методу наименьших квадратов в случае квадратичной зависимости 88
§24. Наилучшие равномерные приближения 91
24.1 Примеры наилучшего равномерного приближения . 92

Глава 3 **ЧИСЛЕННОЕ ИНТЕГРИРОВАНИЕ** **94**
§25. Приближенное вычисление определенных интегралов. Квадратурные формулы Ньютона-Котеса 94
25.1 Формулы Ньютона-Котеса 95
§26. Формула трапеций 99
§27. Формула Симпсона (парабол) 103
§28. Формулы прямоугольников 108
§29. Формула «трех восьмых» 112
§30. Выбор шага интегрирования. Правило Рунге 115
§31. Квадратурные формулы интерполяционного типа 117
31.1 Оценка погрешности квадратурной формулы интерполяционного типа 120
§32. Квадратурные формулы наивысшей алгебраической степени точности (НАСТ) 122
§33. Основная теорема 125

§34. Существование и единственность квадратурных формул наивысшей алгебраической степени точности 129
§35. О положительности квадратурных коэффициентов 132
§36. Погрешность квадратуры наивысшей степени точности . . 134
§37. Связь с ортогональной системой многочленов 137
§38. Квадратурные формулы, отвечающие простейшим весовым функциям 139
38.1 Постоянная весовая функция 140
38.2 Формула численного интегрирования Эрмита 142
38.3 Интегралы вида $\int\limits_a^b (b-x)^\alpha (x-a)^\beta f(x)\,dx$. 143
38.4 Интегралы вида $\int\limits_0^\infty x^\alpha e^{-x} f(x) dx$ 146
38.5 Интегралы вида $\int\limits_{-\infty}^\infty e^{-x^2} f(x) dx$ 147
§39. Формулы численного интегрирования, содержащие заранее предписанные узлы 150
39.1 Содержание задачи и общие теоремы 150
§40. Квадратурные формулы с равными коэффицинтами 155
40.1 Построение формулы 155
40.2 Случай постоянной весовой функции 158
§41. Интерполяционные кубатурные формулы 161
§42. Кубатурная формула трапеций 166

§43. Кубатурная формула Симпсона 170

Глава 4 МЕТОДЫ ЧИСЛЕННОГО РЕШЕНИЯ ЗАДАЧИ КОШИ ДЛЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ **176**

§44. Аналитические методы решения задачи Коши 176
§45. Понятие одношаговых и многошаговых методов 181
§46. Построение одношаговых методов способом разложения решение в ряд тейлора 183
§47. Методы типа Рунге-Кутта 190
§48. Случай уравнений высших порядков 200
§49. Оценка погрешности (сходимость) одношаговых методов . 202
§50. Правило Рунге-Кутта 209
§51. Многошаговые методы 211
§52. Экстраполяционные методы Адамса 213
§53. Интерполяционные методы Адамса 219
§54. Понятие жесткой системы дифференциальных уравнений . 225
§55. Нелинейные системы дифференциальных уравнений 229
§56. Условно устойчивые и абсолютно устойчивые разностные методы 232
§57. Специальные определения устойчивости 236
§58. Чисто неявные разностные методы 240

Глава 5 **РЕШЕНИЕ ГРАНИЧНЫХ ЗАДАЧ ДЛЯ ОДУ** **246**
§59. Метод стрельбы (пристрелки) 246
§60. Метод редукции к задачам Коши 249
§61. Метод прогонки для граничных задач 251
§62. Метод моментов 256
§63. Метод Галеркина 262
§64. Методы моментов и Галеркина для операторов, заданных в гильбертовом пространстве H 265
§65. Метод наименьших квадратов 268
§66. Метод Ритца 277
§67. Метод сеток для решения линейных граничных задач . . . 282
67.1 Постановка задачи. Идея метода сеток 282
67.2 Методы замены обыкновенных дифференциальных уравнений и граничных условий системой алгебраических уравнений. 284
§68. Оценка погрешности и сходимость метода сеток 292
§69. Сходимость и аппроксимация разностных схем 297
§70. Однородные разностные схемы для дифференциальных уравнений 303
§71. Интегро-интерполяционный метод 307
§72. Метод аппроксимации квадратичного функционала 310
§73. Метод аппроксимации интегрального тождества 312

§74. Метод Ритца и Бубнова-Галеркина (вариационно-разностные методы) 314
§75. Монотонные разностные схемы 318

Глава 6 **ЧИСЛЕННОЕ РЕШЕНИЕ ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ** **322**
§76. Некоторые предварительные определения 322
§77. Метод механических квадратур решения уравнения Фредгольма II рода 325
77.1 Сходимость метода квадратур 335
§78. Метод последовательных приближений решения интегрального уравнения Фредгольма второго рода 338
§79. Интерполяционный квадратурный метод 342
§80. Метод замены ядра уравнения на вырожденное ядро для решения уравнений Фредгольма второго рода 348
80.1 Применение степенного ряда 352
80.2 Использование интерполяционных методов 353
§81. Оценка близости между решениями уравнений в зависимости от близости самих уравнений 354
§82. Метод моментов 358
82.1 Связь метода Галеркина с задачей замены ядра на вырожденное 360
§83. Метод коллокации 365

§84. Метод квадратур решения интегрального уравнения Вольтерра второго рода 370
§85. Метод последовательных приближений решения интегрального уравнения Вольтерра второго рода 376
§86. Решение нелинейных уравнений вида Вольтерра 379
§87. Корректно поставленные и некорректно поставленные задачи 381
§88. Метод регуляризации 387
Литература 402

ПРЕДИСЛОВИЕ

Настоящий электронный курс лекций предназначен для студентов специальностей 1-31 03 03-01 «Прикладная математика» и 1-31 03 06-01 «Экономическая кибернетика» физико-математического факультета. Он написан в соответствии с действующей типовой программой по дисциплине «Методы численного анализа», утверждённой первым заместителем Министра образования Республики Беларусь.

В электронном издании излагается теоретический материал, содержащий вопросы: решение нелинейных уравнений, приближение функций, численное интегрирование, методы численного решения задачи Коши для обыкновенных дифференциальных уравнений (ОДУ), решение граничных задач для ОДУ, численное решение интегральных уравнений. Теоретический материал иллюстрируется примерами решения задач.

Курс лекций ставит своей целью облегчить самостоятельную работу студентов с теоретическим материалом при подготовке к лекциям, лабораторным занятиям и экзамену.

Авторы.

ГЛАВА 1
РЕШЕНИЕ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

§1. Итерационные методы. Исследование уравнения

Численные методы делятся на точные и приближенные. Точные методы дают решение задачи за конечное число арифметических действий. Решение получается точным только в том случае, если исходные данные заданы точно и промежуточные вычисления выполняются без округлений.

К точным методам относятся метод Гаусса, метод квадратного корня, метод оптимального исключения, метод ортогонализации.

Итерационные методы дают бесконечную последовательность приближений, предел которой (если он существует) является решением задачи.

Итерационные методы применяются для приближенного решения алгебраических и трансцендентных уравнений, систем уравнений и др.

Решение уравнения или системы уравнений итерационным методом получается как предел последовательности приближений, вычисляемых в ходе процесса итераций.

К итерационным методам относятся метод простых итераций, метод Ньютона, хорд, секущих решения уравнений, метод простой итерации и метод Ньютона решения систем уравнений.

При решении задачи методом итераций нужно сначала задать начальное приближение, из этого начального приближения получаются новые «улучшенные» приближенные значения. Новые приближенные значения «улучшаются» и т. д. При определенных условиях построенная таким образом последовательность приближений сходится к точному решению. Множество начальных приближений, при которых последовательность приближений сходится к решению задачи, называется областью сходимости метода.

Кроме сходимости к решению, при использовании итерационных методов существенной является и скорость сходимости, которая определяется так:

$$|x_{n+1} - x^*| = \Theta|x^n - x^*|^k,$$

где Θ – символ Ландау, некоторая константа. При $k = 1$ – линейная скорость сходимости, $k = 2$ – квадратичная скорость сходимости, $k = 3$ – кубическая скорость сходимости. Любая скорость свыше линейной – сверхлинейная.

В основе многих итерационных методов лежит так называемый принцип сжимающих отображений.

Исследование уравнения

Пусть задана непрерывная функция $f(x)$ и требуется найти все или некоторые корни уравнения

$$f(x) = 0. \tag{1}$$

Корнем уравнения (1) называется такое значение $x = x^*$, что $f(x^*) = 0$.

Если функция $f(x)$ непрерывна на $[a, b]$ и ее значения на концах отрезка имеют разные знаки $f(a)f(b) < 0$ то на $[a, b]$ найдется, по крайней мере, один корень уравнения.

Определить корень уравнения – значит, найти такой интервал, внутри которого имеется корень данного уравнения и притом единственный на данном интервале. Для отделения корней уравнения (1) применяют следующий критерий: если на $[a, b]$ функция $f(x)$ непрерывна и монотонна, а её значения на концах отрезка имеют разные знаки, то на рассматриваемом отрезке существует и только один корень данного уравнения. Достаточный признак монотонности функции на отрезке – сохранение знака производной функции.

Отделение корней уравнения можно выполнить графически. Для этого надо построить график функции $y = f(x)$, по которому можно судить, в каких интервалах находятся его точки пересечения с осью Ox. Наиболее совершенным способом отделения корней является метод Штурма.

Теорема 1. ***(принцип сжимающих отображений)***. *Пусть R – полное метрическое пространство. Если отображение $f : R \to R$ – сжатие, то оно имеет в R единственную неподвижную точку, которая является пределом последовательности, полученной по фор-*

муле

$$x_{n+1} = f(x_n), \tag{2}$$

где x_0 – произвольный элемент из R.

Доказательство этой теоремы приведено в курсе «Вычислительные методы алгебры».

§2. Метод простой итерации для решения нелинейных уравнений. Теорема о сходимости

Применим принцип сжатых отображений для исследования сходимости итерационного метода решения нелинейного уравнения

$$F(x) = 0, \tag{1}$$

где $F(x)$ – вещественная функция вещественного аргумента.

Сначала уравнение (1) приводят к виду, удобному для итераций

$$x = \varphi(x), \tag{2}$$

($\varphi(x) = x - \psi(x)F(x)$, $\psi(x)$ – знакопостоянная функция), где искомый корень x^* уравнения (1) является и корнем уравнения (2). Допустим, что для x^* каким-то способом указано начальное приближение x_0, а затем дальнейшие приближения строятся по формуле

$$x_{n+1} = \varphi(x_n), \quad n = 0, 1, 2, \ldots \tag{3}$$

Этот процесс называется простой одношаговой итерацией (методом простой итерации).

2.1 Геометрический смысл

Здесь $y = x$ – биссектриса первого координатного угла. Решение уравнения x^* – точка пересечения биссектрисы и графика функции $y = \varphi(x)$ и $M_0(x_0, \varphi(x_0))$, $M_1(x_1, \varphi(x_2)), \ldots$.

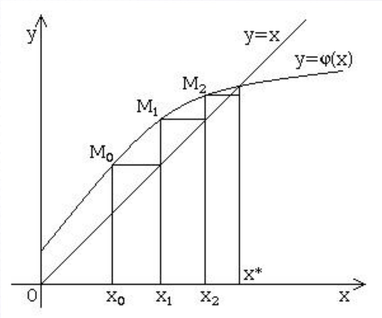

Рисунок 1.

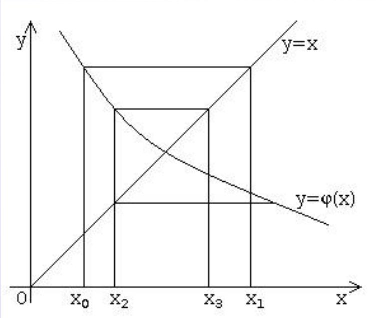

Рисунок 2.

Найдем достаточные условия сходимости метода простой итерации.

Теорема 1. *Пусть выполняются условия*

1. *функция определена на отрезке*

$$|x - x_0| \le \delta, \tag{4}$$

непрерывна там и удовлетворяет условию Липшица с постоянным коэффициентом, меньшим единицы

$$|\varphi(x) - \varphi(x')| \le q|x - x'|, \ (0 < q < 1); \tag{5}$$

2. *для исходного приближения x_0 верно равенство*

$$|x_0 - \varphi(x_0)| \le m; \tag{6}$$

3. *числа δ, q, m удовлетворяют условию*

$$\frac{m}{1-q} \le \delta. \tag{7}$$

Тогда

1. *уравнение (2) в области (4) имеет решение;*
2. *последовательность x_n приближений, построенных по правилу (3), принадлежит отрезку (4), является сходящейся ($\lim x_n = x^*$), и предел последовательности x^* удовлетворяет уравнению (2);*

3. скорость сходимости x_n к x^* оценивается неравенством

$$|x^* - x_n| \le \frac{m}{1-q} q^n, \ n = 1, 2, \ldots \tag{8}$$

Доказательство. Покажем по индукции, что приближения x_n лежат на отрезке (4) и для них верно неравенство

$$|x_{n+1} - x_n| \le mq^n. \tag{9}$$

При $n = 0$ неравенство проверяется просто. Приближение $x_1 = \varphi(x_0)$, очевидно, может быть найдено, так как $x = x_0$ принадлежит отрезку (4). Кроме того, $|x_1 - x_0| = |\varphi(x_0) - x_0| \le m$ по второму условию теоремы, и неравенство (9) для $n = 0$ выполнено. Наконец, так как $m \le \dfrac{m}{1-q} \le \delta$, то x_1 принадлежит отрезку (4).

Предположим, что $x_0, x_1, \ldots, x_n$ принадлежит области (4) и выполняются условия

$$|x_{k+1} - x_k| \le mq^k, \quad k = 0, 1, 2, \ldots, n-1.$$

Так как x_n, по предположению, принадлежит области (4), то приближение $x_{n+1} = \varphi(x_n)$ может быть построено. По сделанному допущению $|x_n - x_{n-1}| \le mq^{n-1}$, поэтому

$$|x_{n+1} - x_1| = |\varphi(x_n) - \varphi(x_{n-1})| \le q|x_n - x_{n-1}| \le qmq^{n-1} = mq^n$$

и для приближений x_{n+1} и x_n неравенство (9) выполнено. Осталось еще проверить принадлежность x_{n+1} области (4).

$$|x_{n+1}-x_0| = \left|(x_{n+1}-x_n)+(x_n-x_{n-1})+(x_{n-1}-x_{n-2})+\ldots+(x_1-x_0)\right| \leq$$

$$\leq mq^n + mq^{n-1} + \ldots + m = \frac{m-mq^{n+1}}{1-q} < \frac{m}{1-q} \leq \delta.$$

Этим завершается индукция.

Покажем теперь, что для последовательности x_n выполняется условие Больцано-Коши

$$|x_{n+p}-x_n| = \left|(x_{n+p}-x_{n+p-1})+(x_{n+p-1}-x_{n+p-2})+\ldots+(x_{n+1}-x_n)\right| \leq$$

$$\leq mq^{n+p-1}+mq^{n+p-2}+\ldots+mq^n = m\frac{q^n-q^{n+p}}{1-q} < \frac{m}{1-q}q^n.$$

Последняя часть цепочки неравенства не зависит от p и, так как $0 < q < 1$, то при достаточно больших n будет меньше любого заданного заранее числа, т. е. последовательность x_n – фундаментальная (удовлетворяет условию Больцано-Коши). Следовательно, существует $\lim\limits_{n\to\infty} x_n = x^*$.

Принадлежность x^* замкнутому отрезку $|x-x_0| \leq \delta$ следует из того, что ему принадлежат все x_n.

Покажем, что x^* – есть решение уравнения (2). Для этого в правиле вычислений $x_{n+1} = \varphi(x_n)$ устремим n к бесконечности. Тогда $x_{n+1} \to x^*$ и $x_n \to x^*$, ввиду же непрерывности $\varphi(x)$ во всех точках отрезка (4) будет $\varphi(x_n) \to \varphi(x^*)$. В пределе получим равенство $x^* = \varphi(x^*)$, т. е. x^* – решение уравнения (2). В неравенстве $|x_{n+p} - x_n| = \dfrac{mq^n}{1-q}$ устремим $p \to \infty$, тогда $x_{n+p} \to x^*$ и $|x^* - x_n| \le \dfrac{mq^n}{1-q}$. Теорема 1 доказана.

Теорема 2. (о единственности решения). *На всяком множестве точек, где для функции $\varphi(x)$ выполняется условие $|\varphi(x) - \varphi(y)| < |x - y|$, $x \ne y$, уравнение $x = \varphi(x)$ может иметь не более одного решения.*

Доказательство. Допустим противоположное и будем считать, что на указанном множестве существуют два разных решения x и y, $x \ne y$. Оценим разность $x - y : |x - y| = |\varphi(x) - \varphi(y)| < |x - y|$, приходим к невозможному неравенству $|x - y| < |x - y|$. Поэтому предположение о существовании двух разных решений неверно. Теорема 2 доказана.

Замечание 1. *Условие Липшица с константой $q < 1$ выполняется для функции $\varphi(x)$ на (4), если эта функция имеет на (4) производную $\varphi'(x)$, удовлетворяющую неравенству $\left|\varphi'(x)\right| \le q < 1$.*

Скорость сходимости метода итерации – линейная.

Метод простой итерации является самоисправляющимся: допущенная при вычислении ошибка (не выводящая за пределы отрезка (4)) будет исправлена.

Начало
Содержание
Страница 21 из 403
Назад
На весь экран
Закрыть

§3. Ускорение сходимости метода итераций

В методе простой итерации $x_{n+1} = \varphi(x_n)$ последовательность x_n сходится к решению x^*, если $\left|\varphi'(x^*)\right| < 1$ и если начальное приближение взято достаточно близко к x^*. Но скорость сходимости зависит от $|\varphi'(x^*)|$. Если $|\varphi'(x^*)|$ близок к единице, то сходимость может быть очень медленной и для получения нужной точности требуется проделать много шагов вычислений.

Как всякий процесс последовательных приближений, метод простой итерации можно улучшать, преследуя при этом две цели: ускорение сходимости процесса и ослабление условий сходимости. Для осуществления этих целей можно воспользоваться двумя средствами: изменять заданное уравнение и изменять итерационный процесс. Чтобы перейти от уравнения $F(x) = 0$ к $x = \varphi(x)$ в качестве $\varphi(x)$ выбирают функцию $\varphi(x) = x - \psi(x)F(x)$, где $\psi(x)$ – произвольная непрерывная в окрестности x^* функция.

1. Возьмем в качестве $\psi(x)$ функцию $\psi(x) = \dfrac{x_n - x_{n-1}}{F(x_n) - F(x_{n-1})}$, тогда метод $x_{n+1} = \varphi(x_n)$ запишется в виде

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

Это *метод секущих.* Его скорость сходимости выше, чем для метода простой итерации $|x_{n+1} - x^*| = \Theta|x_n - x^*|^k$. Здесь $k = \dfrac{\sqrt{5}+1}{2} > 1$, скорость сходимости сверхлинейная. Начальное приближение x_0 выбираем достаточно близким к x^* и таким, чтобы $F(x_0)F''(x_0) > 0$. Это двухшаговый метод.

2. *Метод Стеффенсона* имеет вид

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

где, $x_n' = \varphi(x_n)$, $x_n'' = \varphi(x_n') = \varphi\big[\varphi(x_n)\big]$.

Отсюда, $x_{n+1} = \dfrac{x_n\varphi\big[\varphi(x_n)\big] - \big[\varphi(x_n)\big]^2}{\varphi\big[\varphi(x_n)\big] - 2\varphi(x_n) + x_n}$.

Полученное равенство называют *итерационной формулой Стеффенсона*, она является одношаговой и требует вычисления двух значений φ на каждом шаге. Обозначим $\varepsilon_n = x_n - x^*$ – погрешность.

Тогда

$$\varphi(x_n) = \varphi(x^* + \varepsilon_n) = \varphi(x^*) + \alpha\varepsilon_n + \beta\varepsilon_n^2 + \gamma\varepsilon_n^3 + \ldots = x^* + \alpha\varepsilon_n + \beta\varepsilon_n^2 + \gamma\varepsilon_n^3 + \ldots\,.$$

Здесь $x^* = \varphi(x^*)$, $\alpha = \varphi'(x^*)$, $\beta = \dfrac{1}{2}\varphi''(x^*)$, $\gamma = \dfrac{1}{6}\varphi'''(x^*)$. Нетрудно получить, что $\varepsilon_{n+1} \approx \dfrac{1}{2}\cdot\dfrac{\alpha\beta}{\alpha - 1}\varepsilon_n^2 = B\varepsilon_n^2$. Здесь $k = 2$ и скорость сходимости метода Стеффенсона квадратичная.

Если погрешность $\varepsilon_0 = x_0 - x^*$ начального приближения настолько мала, что для нее выполняется неравенство

$$|B\varepsilon_0| = \left|\frac{1}{2}\frac{\varphi'(x^*)\varphi''(x^*)}{\varphi'(x^*)-1}\varepsilon_0\right| < 1,$$

то метод Стеффенсона сходится к решению x^* с квадратичной скоростью: $\varepsilon_n \approx B^{-1}(B\varepsilon_0)^{2^n}$.

3. *Метод Ньютона.*

Возьмем в качестве $\psi(x) = \dfrac{1}{F'(x)}$, тогда $x_{n+1} = x_n - \dfrac{F(x_n)}{F'(x_n)}$ – метод Ньютона. Скорость сходимости метода Ньютона – квадратичная

$$|x_{n+1} - x^*| = \Theta|x_n - x^*|^2,$$

т. е. $k = 2$. Начальное приближение x_0 следует выбирать достаточно близким к x^* и таким, чтобы $F(x_0)F''(x_0) > 0$.

Видоизменение метода Ньютона с постоянным значением производной:

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

Скорость сходимости более медленная, чем для классического метода Ньютона.

Если взять $F'(x) \approx \dfrac{F(x_n) - F(x_{n-1})}{x_n - x_{n-1}}$, то получим видоизменение метода Ньютона – *метод секущих:*

$$x_{n+1} = x_n - \frac{x_n - x_{n-1}}{F(x_n) - F'(x_{n-1})}F(x_n).$$

§4. Метод Ньютона (касательных) решения уравнений с одним неизвестным

Предположим, что на некотором отрезке $[a, b]$ функция $F(x)$ дважды непрерывно дифференцируема, т. е. $F(x) \in C^2[a, b]$. Кроме того, $F'(x)$ и $F''(x)$ не обращаются в нуль на $[a, b]$ и $F(a)F(b) < 0$. Это означает, что на $[a, b]$ существует единственный корень x^* уравнения

$$F(x) = 0. \tag{1}$$

Если функция $\psi(x)$ непрерывна в некоторой окрестности x^*, то уравнение

$$x = \varphi(x), \tag{2}$$

где $\varphi(x) = x - \psi(x)F(x)$ также имеет своим корнем x^*.

В качестве $\psi(x) = \dfrac{1}{F'(x)}$, тогда (2) запишется в виде

$$x = x - \frac{F(x)}{F'(x)}. \tag{3}$$

Заметим, что (1) и (3) эквивалентны. Выберем начальное приближение x_0 так, чтобы оно было близко к точному решению x^* уравнения (1).

Получим итерационный процесс

$$x_{n+1} = x_n - \frac{F(x_n)}{F'(x_n)}. \tag{4}$$

Покажем, что последовательность $\{x_n\}$ сходится к решению уравнения (1), т. е. $\lim\limits_{n\to\infty} x_n = x^*$.

Для этого покажем, что $\varphi(x)$ – сжатие. Возьмем производную от правой части (3)

$$\varphi'(x) = 1 - \frac{\left[F'(x)\right]^2 - F(x)F''(x)}{\left[F'(x)\right]^2}.$$

Очевидно, что $\varphi'(x^*) = 0$.

Действительно, $\varphi'(x^*) = 1 - \dfrac{\left[F'(x^*)\right]^2 - \left(F'(x^*)\right)^2 + F(x^*)F''(x^*)}{\left[F'(x)\right]^2} = 0,$

так как $F(x^*) = 0$.

В силу предположения функция $\varphi'(x)$ непрерывна на $[a, b]$ и в точке x^* обращается в нуль. Следовательно, существует такая окрестность U_{x^*} точки x^*, для которой $|\varphi'(x)| \leq q < 1 \Rightarrow \varphi(x)$ – сжатие.

Тогда, если выбрать x_0 так, чтобы $x_0 \in U_{x^*}$, то по замечанию о сходимости метода простой итерации можно сделать вывод, что метод Ньютона сходится, т. е. $\lim\limits_{n\to\infty} x_n = x^*$. Заметим, что x_0 следует выбирать так, чтобы $x_0 \in U_{x^*}$ и $F(x_0)F''(x_0) > 0$.

По формуле Тейлора

$$F(x) = F(x_n) + (x - x_n)F'(x_n) + \frac{(x - x_n)^2}{2!}F''(\xi), \quad \xi \in (x, x_n).$$

Так как $F(x^*) = 0$, то $\dfrac{F(x_n)}{F'(x_n)} = x_n - x^* - \dfrac{1}{2}\dfrac{F''(\xi)}{F'(x_n)}(x^* - x_n)^2$.

Следовательно, $x_{n+1} - x^* = x_n - x^* - \dfrac{F(x_n)}{F'(x_n)} = \dfrac{1}{2}\dfrac{F''(\xi)}{F'(x_n)}(x^* - x_n)^2$.

Если $m_1 = \min\left|F'(x)\right|$, $M_2 = \max\limits_{[a,b]}|F''(x)|$, то

$$|x_{n+1} - x^*| \le \frac{M_2}{2m_1}|x_n - x^*|^2.$$

Остюда следует квадратичная скорость сходимости.

Потребуем, чтобы начальное приближение x_0 было выбрано так, чтобы выполнялось неравенство

$$\frac{M_2}{2m_1}|x_n - x^*| \le q < 1. \tag{5}$$

Тогда $|x_1 - x^*| \le \dfrac{M_2}{2m_1}|x_0 - x^*|^2 \le \dfrac{2m_1}{M_2}q^2, \ \ldots, \ |x_n - x^*| \le \dfrac{2m_1}{M_2}q^{2^n}$.

4.1 Геометрический смысл метода Ньютона (касательных)

Пусть на $[a, b]$ находится единственный корень уравнения $F(x) = 0$. Проведем касательную к кривой $y = F(x)$ в точке $A[a, F(a)]$ до пересечения с осью Ox. Уравнение касательной, проходящей через точку A будет иметь вид $y = F(a) + F'(a)(x - a)$. Если $F'(a) \ne 0$, то находим абсциссу x_1 точки пересечения касательной с Ox $x_1 = a - \dfrac{F(a)}{F'(a)}$.

Значение x_1 можно взять в качестве приближения к корню. Проводим касательную через точку $A_1\big(x_1, F(x_1)\big)$ и найдем точку пересечения с осью Ox, получим второе приближение x_2 к корню уравнения $x_2 = x_1 - \dfrac{F(x_1)}{F'(x_1)}$. Применяя метод касательных, n–ое приближение получим $x_n = x_{n-1} - \dfrac{F(x_{n-1})}{F'(x_{n-1})}$, $n = 1, 2, 3, \ldots$, причем x_0 удовлетворяет условию $F(x_0)F''(x_0) > 0$.

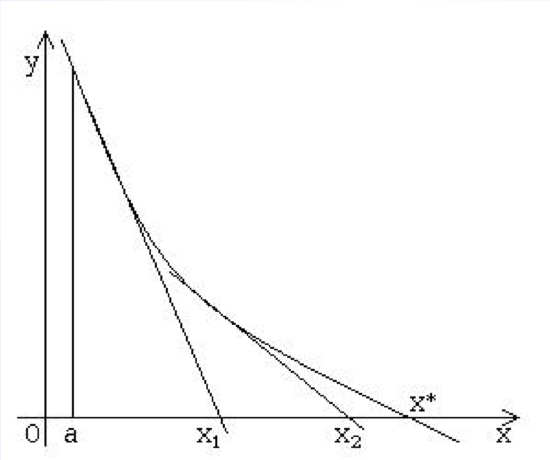

Рисунок 1.

§5. Решение систем нелинейных уравнений. Метод простых итераций. Метод Зейделя

Систему нелинейных уравнений можно кратко записать в виде

$$f(x) = 0 \tag{1}$$

или более подробно в координатном виде

$$f_i(x_1, x_2, \dots, x_m) = 0, \quad 1 \le i \le m. \tag{2}$$

Такие системы решают практически только итерационными методами. Для такой системы используют метод простых итераций. Для этого систему (1) приводят к виду

$$\begin{gathered} x_1 = \varphi_1(x_1, x_2, \dots, x_m), \\ \dots \\ x_m = \varphi_m(x_1, x_2, \dots, x_m). \end{gathered} \tag{3}$$

Для ее записи в m–мерном числовом пространстве R_m введем два вектора: один из них x для изображения совокупности неизвестных $(x_1, x_2, \dots, x_m)$, второй φ будет обозначать совокупность значений функций $(\varphi_1, \varphi_2, \dots, \varphi_m)$. Система (3) запишется в краткой векторной форме

$$x = \varphi(x). \tag{4}$$

Пусть как-то выбрано начальное приближение к решению $x^0 = = (x_1^0, x_2^0, \dots, x_m^0)$. Все следующие приближения будут находиться по правилу

$$x^{n+1} = \varphi(x^n), \quad n = 0, 1, 2, \dots,$$

иначе $x_i^{n+1} = \varphi_i(x_1^n, x_2^n, \dots, x_m^n), \quad i = \overline{1, m}$.

Подойдем к изучению этого метода с более общих позиций. Пусть X – полное метрическое пространство, а оператор $y = \varphi(x)$ отображает X в себя. Рассмотрим итерационный процесс $x^{n+1} = \varphi(x^n)$ решения уравнения $x = \varphi(x)$.

Определение 1. *Если при некотором $q < 1$ отображение $y = \varphi(x)$ удовлетворяет условию*

$$\rho\big[\varphi(x_1), \varphi(x_2)\big] \le q\rho(x_1, x_2) \tag{5}$$

при всех $x_1, x_2 \in X$, то такое отображение называется сжатием.

Теорема 1. *Если отображение $y = \varphi(x)$ – сжатие, то уравнение $x = \varphi(x)$ имеет решение x^* и $\rho(x^*, x^n) < \dfrac{mq^n}{1-q}$, где $m = \rho(x^1, x^0)$.*

Доказательство. Согласно (5), имеем

$$\rho(x^{n+1}, x^n) = \rho\big[\varphi(x^n), \varphi(x^{n-1})\big] \le q\rho(x^n, x^{n-1}),$$

поэтому $\rho(x^{n+1}, x^n) \le q^n \rho(x^1, x^0) = mq^n$.

При $p > n$ имеем цепочку неравенств

$$\rho(x^p, x^n) \le \rho(x^p, x^{p-1}) + \rho(x^{p-1}, x^{p-2}) + \ldots + \rho(x^{n+1}, x^n) \le$$

$$\le mq^{p-1} + mq^{p-2} + \ldots + mq^n \le mq^n \sum_{i=0}^{\infty} q^i = \frac{mq^n}{1-q}. \tag{6}$$

Итак, последовательность x^n является фундаментальной. Согласно критерию Коши, последовательность x^n имеет некоторый предел x^*. Переходя к пределу в (6) при $p \to \infty$, получим $\rho(x^*, x^n) \le \dfrac{mq^n}{1-q}$. Очевидно,

$$\rho\big(x^*, \varphi(x^*)\big) < \rho(x^*, x^{n+1}) + \rho\big(x^{n+1}, \varphi(x^*)\big) = \rho(x^*, x^{n+1}) +$$

$$+\rho\big(\varphi(x^n), \varphi(x^*)\big) \le \rho(x^*, x^{n+1}) + q\rho(x^n, x^*) \le 2\frac{mq^{n+1}}{1-q}.$$

Поскольку n произвольное, то $\rho\big(x^*, \varphi(x^*)\big) = 0$, иначе $x^* = \varphi(x^*)$. Теорема доказана.

В случае системы нелинейных уравнений $f(x) = 0$ аналогом метода Зейделя является итерационный процесс, где компоненты приближений определяются из соотношений

$$f_1(x_1^{n+1}, x_2^n, \ldots, x_m^n) = 0,$$

$$f_2(x_1^{n+1}, x_2^{n+1}, x_3^n, \ldots, x_m^n) = 0,$$

$$\dots \tag{7}$$

$$f_m(x_1^{n+1}, x_2^{n+1}, \dots, x_m^{n+1}) = 0.$$

Промежуточное место между итерационными методами (3) и (7) занимает метод, где компоненты приближений определяются из соотношений

$$x_1^{n+1} = \varphi_1(x_1^n, \dots, x_m^n),$$

$$x_2^{n+1} = \varphi_2(x_1^{n+1}, x_2^n, \dots, x_m^n),$$

$$\dots \tag{8}$$

$$x_m^{n+1} = \varphi_m(x_1^{n+1}, x_2^{n+1}, \dots, x_{m-1}^{n+1}, x_m^n).$$

Методы (7) и (8) особенно широко используются в различных моделирующих устройствах.

В достаточно малой окрестности решения x^* системы уравнений для приближений методом простой итерации имеем

$$x^{n+1} - x^* = \varphi(x^n) - \varphi(x^*) \approx B(x^n - x^*),$$

где $B = \left[\dfrac{d\varphi_i}{dx_j}\right]_{x^*}$.

Удается показать, что при достаточно хорошем начальном приближении итерационный процесс $x^{n+1} = \varphi(x^n)$ сходится при тех же условиях на матрицу B, что и метод простых итераций для линейных систем.

Сходимость метода линейная, сами вычисления итераций просты. Но зато сложно найти такую систему $x = \varphi(x)$, которая была бы эквивалентна исходной системе $f(x) = 0$ и одновременно обеспечивала бы сходимость.

§6. Метод Ньютона для нелинейных систем уравнений

Пусть требуется решить систему нелинейных уравнений

$$
\begin{aligned}
f_1(x_1, x_2, \ldots, x_m) &= 0,\\
&\ldots\\
f_m(x_1, x_2, \ldots, x_m) &= 0.
\end{aligned}
\tag{1}
$$

Рассмотрим m-мерное векторное пространство R_m, $x = (x_1, x_2, \ldots, x_m)$. Введем вектор-функцию $f(x) = \big[f_1(x), f_2(x), \ldots, f_m(x)\big]$, тогда система (1) запишется в виде

$$
f(x) = 0. \tag{2}
$$

Пусть известно некоторое исходное приближение $x^0 = (x_1^0, x_2^0, \ldots, x_m^0)$ к решению системы $x^* = (x_1^*, x_2^*, \ldots, x_m^*)$.

Для выделения главной линейной части из системы (2) удобно рассмотреть не точное решение x^*, а вектор-погрешность $x^* - x^0 = (x_1^* - -x_1^0, \ldots, x_m^* - x_m^0) = \varepsilon = (\varepsilon_1, \ldots, \varepsilon_m)$.

Уравнение для ε получается, если в равенстве $f(x^*) = 0$ заменить x^* на $x^0 + \varepsilon$, т.е. $f(x^0 + \varepsilon) = 0$. Предполагая все составляющие вектора ε малыми величинами, выделим в системе (1) главную линейную часть. Для этого рассмотрим уравнение любого номера $f_i(x^0 + \varepsilon) = 0$ и разложим f_i в ряд Тейлора по степеням погрешности $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m$ и сохраним в разложении линейную часть, отбросив все члены более высокого порядка. После этого получим линейную систему уравнений относительно погрешностей, приближенно заменяющую систему (1).

$$0 = f_i(x^0 + \varepsilon) \approx f_i(x^0) + \sum_{j=1}^{m} \frac{\partial f_i(x^0)}{\partial x_j}\varepsilon_j.$$

$$\sum_{j=1}^{m} \varepsilon_j \frac{\partial f_i(x^0)}{\partial x_j} \approx -f_i(x^0), \quad i = \overline{1, m}. \tag{3}$$

Решая ее относительно ε_i, например, методом Гаусса, получим приблизительные значения для ε_i, например, ε_i^0, $i = \overline{1, m}$. Мы улучшим исходные значения неизвестных x_i^0, если к ним прибавим ε_i^0, т.е. $x_1^1 = x_1^0 + \varepsilon_1^0$, $x_2^1 = x_2^0 + \varepsilon_2^0$, $\ldots$, $x_m^1 = x_m^0 + \varepsilon_m^0$. Новые значения x_i^1 аналогичными вычислениями могут быть улучшены и т.д.

В результате для каждого значения x_i^* получится последовательность приближений x_i^n такая, что каждое следующее приближение x_i^{n+1} будет находиться из линейной системы по предыдущему приближению x_i^n:

$$\sum_{j=1}^{m} (x_j^{n+1} - x_j^n) \frac{\partial}{\partial x_j} f_i(x^n) = -f_i(x^n), \quad i = \overline{1, m}, \quad n = 0, 1, 2, \ldots \tag{4}$$

Рассмотрим матрицу Якоби системы функций f_i, $i = \overline{1,m}$

$$\begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \dots & \dfrac{\partial f_1}{\partial x_m} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \dots & \dfrac{\partial f_2}{\partial x_m} \\ \dots & \dots & \dots & \dots \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \dots & \dfrac{\partial f_m}{\partial x_m} \end{bmatrix} = f'(x).$$

Значение ее при $x = x^n$ есть матрица системы (4). Система будет иметь единственное решение, если ее определитель отличен от нуля $D\big[f'(x^n)\big] \neq \neq 0$. Итак, метод Ньютона сходится, если в достаточно малой окрестности корня $D\left[\dfrac{\partial f}{\partial x}\right] \neq 0$, причем сходимость квадратичная. Вычисления по методу Ньютона несколько сложнее, чем при простых итерациях тем, что на каждом шаге надо находить матрицу производных и решать систему линейных уравнений. Поэтому в некоторых учебниках рекомендуют вычислить матрицу производных только на начальной итерации и использовать её во всех остальных итерациях.

§7. Сведение решения системы нелинейных уравнений к решению вариационной задачи. Метод покоординатного спуска

Пусть требуется решить систему нелинейных уравнений $f(x) = 0$ или в координатном виде

$$f_1(x_1, x_2, \ldots, x_m) = 0,$$

$$\ldots \tag{1}$$

$$f_m(x_1, x_2, \ldots, x_m) = 0.$$

Рассмотрим функцию $\Phi(x) = \sum\limits_{i=1}^{m} \left| f_i(x) \right|^2$. Она неотрицательна и обращается в нуль в том и только том случае, когда $f(x) = 0$. Таким образом, решение исходной системы уравнений (1) будет одновременно нулевым минимумом скалярной функции многих переменных $\Phi(x_1, x_2, \ldots, x_m)$. Иногда бывает проще искать такой минимум, чем решать систему уравнений.

Мы свели решение системы нелинейных уравнений к решению вариационной задачи, которую можно решить, например, методами покоординатного спуска или методом градиентного спуска.

7.1 Метод покоординатного спуска

Он является общим методом приближения к точке минимума функции нескольких аргументов. Пусть имеется приближение $x^\circ = (x_1^\circ, x_2^\circ, \ldots, x_m^\circ)$ к точке минимума Φ. Рассмотрим функцию одного аргумента x_1 : $\Phi(x_1, x_2^\circ, \ldots, x_m^\circ)$ и найдём точку её минимума. Для этого нужно решить уравнение $\frac{\partial}{\partial x_1}\Phi(x_1, x_2^\circ, \ldots, x_m^\circ) = 0$. Обозначим найденное x_1 через x_1^1 и рассмотрим функцию от x_2 : $\Phi(x_1^1, x_2, x_3^\circ, \ldots, x_m^\circ)$ и найдём точку её минимума, т.е. решим уравнение $\frac{\partial}{\partial x_2}\Phi(x_1^1, x_2, x_3^\circ, \ldots, x_m^\circ) = 0$. Продолжим этот процесс до нахождения x_m^1. После этого повторим цикл вычислений, отправляясь от значений $(x_1^1, x_2^1, \ldots, x_m^1)$ и так далее. Даже если на каждом шаге отыскивается абсолютный экстремум функции $\Phi(x_1, x_2, \ldots, x_m)$ по соответствующей координате, уже при $m = 2$ можно построить пример, когда итерационный процесс сходится при некотором начальном приближении не к искомой точке абсолютного экстремума, а к некоторой точке локального экстремума. Спуск в циклическом порядке необязателен. Если из рассмотрения проводившихся ранее вычислений видно, что спуск по каким-либо координатам обеспечивает наибольшее убывание $\Phi(x)$, то иногда целесообразен частый спуск по этим координатам.

7.2 Метод градиентного спуска

Используется для нахождения точки минимума любого функционала, для которого имеет смысл понятие градиента.

Пусть требуется найти минимум функции нескольких аргументов $\Phi(x_1, x_2, \ldots, x_m)$. Предположим, что известно приближенное положение $x^\circ = (x_1^\circ, x_2^\circ, \ldots, x_m^\circ)$ этой точки. Если мы хотим возможно быстро, отправляясь от x°, дойти до места минимума, то разумно выйти из x° в направлении, противоположном градиенту Φ в точке x°. Следующее приближение отыскивается в виде $x^{n+1} = x^n - t \cdot grad\Phi(x^n)$.

Значение t определяется из условия $\varphi'(t) = 0$, где $\varphi(t) = \Phi\big(x^\circ - t \cdot grad\Phi(x^\circ)\big)$. Находится точка минимума t^* функции $\varphi(t)$, затем находится следующее приближение x^1, $x^1 = x^0 - t^* \cdot grad\Phi(x^0)$. После этого, отправляясь от x^1, выполняем вычисления так же, как и для x°.

Напомним, что $grad\Phi = \left(\frac{\partial\Phi}{\partial x_1}, \frac{\partial\Phi}{\partial x_2}, \ldots, \frac{\partial\Phi}{\partial x_m}\right)$.

ГЛАВА 2
ПРИБЛИЖЕНИЕ ФУНКЦИЙ

§8. Интерполирование функций

Пусть функция $f(x)$ задана таблицей значений для конечного множества x

x	x_0	x_1	x_2	$\dots$	x_n
$f(x)$	$f(x_0)$	$f(x_1)$	$f(x_2)$	$\dots$	$f(x_n)$

Такая таблица может быть получена в результате наблюдений за ходом некоторого процесса. Если необходимо найти значение $f(x)$ для промежуточного значения аргумента, то строят функцию $\varphi(x)$, достаточно простую для вычислений, которая в заданных точках $x_0, x_1, \dots, x_n$ принимает значения $f(x_0), f(x_1), \dots, f(x_n)$, а в остальных точках отрезка $[a, b]$, принадлежащего области определения $f(x)$, приближенно представляет функцию $f(x)$ с той или иной степенью точности и при решении задачи вместо функции $f(x)$ используется $\varphi(x)$. Задача построения функции $\varphi(x)$ называется *задачей интерполирования*. Чаще всего интерполяционную функцию представляют в виде алгебраического многочлена некоторой степени.

К интерполированию прибегают и в том случае, когда для функции $f(x)$ известно аналитическое выражение, с помощью которого можно вычислить её значения x из отрезка, в котором она определена, но вычисление каждого значения сопряжено с большим объёмом вычислений. Если надо найти значение функции для большого количества значений аргумента, то прибегают к интерполированию, т. е. вычисляют несколько значений $f(x_i)$, $i = 0, 1, 2, \ldots, n$ и по ним строят простую интерполирующую функцию $\varphi(x)$, посредством которой и вычисляются приближенные значения $f(x)$ в других точках.

В качестве функции $\varphi(x)$ будем брать в дальнейшем алгебраический многочлен степени n, интерполяция в этом случае называется алгебраической. Алгебраическая интерполяция функции $y = f(x)$ на $[a, b]$ состоит в приближенной замене этой функции на данном отрезке многочленом $P_n(x)$ степени n, то есть $f(x) \approx P_n(x)$, причем $P_n(x)$ принимает в точках $x_0, x_1, \ldots, x_n$ те же значения, что и $f(x)$, т.е. $f(x_i) = P_n(x_i)$, $i = 0, 1, \ldots, n$. Отметим, что двух различных интерполяционных многочленов одной и той же степени n существовать не может для данной функции $f(x)$.

§9. Интерполяционный многочлен Лагранжа

Поставим следующую задачу: построить многочлен $P_n(x)$ степени n, который в $n+1$ данных точках $x_0, x_1, \ldots, x_n$ (узлах интерполяции) принимает значения $y_0, y_1, \ldots, y_n$. Для решения задачи определим так называемые фундаментальные многочлены $Q_n^k(x)$, т. е. многочлены n-ой степени относительно x, удовлетворяющие следующим условиям:

$$Q_n^k(x_i) = \begin{cases} 0, & \text{при } i \neq k, \\ 1, & \text{при } i = k. \end{cases} \tag{1}$$

Искомый многочлен запишется в виде суммы при помощи фундаментальных многочленов

$$P_n(x) = y_0 Q_n^0(x) + y_1 Q_n^1(x) + \ldots + y_n Q_n^n(x). \tag{2}$$

Поскольку $x_0, x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n$ – нули многочлена $Q_n^k(x)$, то

$$Q_n^k(x) = \frac{(x-x_0)(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}. \tag{3}$$

Формула (2) с учетом (3) запишется

$$P_n(x) = \sum_{k=0}^{n} y_k \frac{(x-x_0)(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}. \tag{4}$$

Многочлен, определяемый формулой (4), называется интерполяционным многочленом Лагранжа, а фундаментальные многочлены (3) коэффициентами Лагранжа. Формула

$$f(x) \approx P_n(x) = \sum_{k=0}^{n} y_k \frac{(x-x_0)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)} \tag{5}$$

называется интерполяционной формулой Лагранжа. Формулы (3) – (5) можно записать более компактно, если ввести обозначение $\omega(x) = (x-x_0)(x-x_1)\ldots(x-x_n)$.

Поскольку $\omega'(x_k) = (x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)$, то

$$Q_n^k(x) = \frac{\omega(x)}{(x-x_k)\omega'(x_k)},$$

$$P_n(x) = \omega(x)\sum_{k=0}^{n}\frac{y_k}{(x-x_k)\omega'(x_k)},$$

$$f(x) \approx \omega(x)\sum_{k=0}^{n}\frac{y_k}{(x-x_k)\omega'(x_k)}.$$

Полином $P_n(x)$ совпадает с $f(x)$ в $n+1$ точке, а в остальных точках отрезка $[a,b]$ разность $R_n(x) = f(x) - P_n(x)$ называется остаточным членом интерполяции. Она не равна нулю и представляет собой погрешность метода. Справедлива теорема для оценки остаточного члена.

Теорема 1. *Если функция $f(x)$ в промежутке $[a, b]$ имеет непрерывные производные до $n+1$-го порядка, то остаточный член интерполяции $R_n(x)$ определяется формулой $R_n(x) = f^{(n+1)}(\xi)\dfrac{\omega(x)}{(n+1)!}$, где ξ – точка промежутка $[a, b]$, зависящая от x.*

§10. Конечные разности. Разделённые разности

Пусть $y_i = f(x_i)$ – значение функции $y = f(x)$ в точках x_i, $i = 0, 1, \ldots, n$, тогда разности $y_1 - y_0$, $y_2 - y_1, \ldots$, $y_n - y_{n-1}$ называются конечными разностями 1-го порядка. Обозначим $\Delta y_i = y_{i+1} - y_i$. Разности второго порядка получаются по формулам $\Delta^2 y_0 = \Delta y_1 - \Delta y_0$, $\Delta^2 y_1 = \Delta y_2 - \Delta y_1, \ldots$ Аналогично определяются последующие разности $\Delta^{n+1} y_0 = \Delta^n y_1 - \Delta^n y_0$, $\Delta^{n+1} y_1 = \Delta^n y_2 - \Delta^n y_1, \ldots$ Для работы с конечными разностями удобно использовать таблицу

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
x_0	y_0				
		Δy_0			
x_1	y_1		$\Delta^2 y_0$		
		Δy_1		$\Delta^3 y_0$	
x_2	y_2		$\Delta^2 y_1$		$\Delta^4 y_0$
		Δy_2		$\Delta^3 y_1$	
x_3	y_3		$\Delta^2 y_2$		
		Δy_3			
x_4	y_4				
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$

10.1 Свойства конечных разностей

1. Конечные разности суммы (разности) функций равны сумме (разности) конечных разностей этих функций.

2. При умножении функции на постоянный множитель конечные разности умножаются на этот множитель.

3. Конечные разности n-го порядка от многочленов степени n постоянны, а конечные разности $n+1$-го порядка равны нулю.

Разделенные разности первого порядка определяются формулами

$$f(x_1, x_0) = \frac{y_1 - y_0}{x_1 - x_0}, \quad f(x_2, x_1) = \frac{y_1 - y_0}{x_2 - x_1}, \dots, \quad f(x_n, x_{n-1}) = \frac{y_n - y_{n-1}}{x_n - x_{n-1}}.$$

Разделенные разности второго порядка получаются из разделенных разностей первого порядка по формулам

$$f(x_2, x_1, x_0) = \frac{f(x_2, x_1) - f(x_1, x_0)}{x_2 - x_0},$$

$$f(x_3, x_2, x_1) = \frac{f(x_3, x_2) - f(x_2, x_1)}{x_3 - x_1}.$$

Разделенные разности n-го порядка получаются из разделенных разностей $n-1$-го порядка по формулам

$$f(x_n, x_{n-1}, \dots, x_0) = \frac{f(x_n, x_{n-1}, \dots, x_1) - f(x_{n-1}, x_{n-2}, \dots, x_0)}{x_n - x_0}.$$

Разделенные разности в случае равностоящих узлов с шагом h, $(x_k = x_0 + kh,\ k = 0, 1, \ldots, n)$ выражаются следующим образом

$$f(x_1, x_0) = \frac{\Delta y_0}{h},\ f(x_2, x_1) = \frac{\Delta y_1}{h}, \ldots,\ f(x_n, x_{n-1}) = \frac{\Delta y_{n-1}}{h}, \ldots,$$

$$f(x_2, x_1, x_0) = \frac{f(x_2, x_1) - f(x_1, x_0)}{x_2 - x_0} = \frac{\dfrac{\Delta y_1}{h} - \dfrac{\Delta y_0}{h}}{2h} = \frac{\Delta y_1 - \Delta y_0}{2h^2} = \frac{\Delta^2 y_0}{2!h^2},$$

$$f(x_3, x_2, x_1) = \frac{\Delta^2 y_1}{2!h^2}, \ldots,\ f(x_n, x_{n-1}, \ldots, x_0) = \frac{\Delta^2 y_0}{n!h^n}.$$

Тем самым установлена связь между конечными и разделёнными разностями.

§11. Интерполяционные многочлены Ньютона

Интерполяционный многочлен Лагранжа, который можно построить при любом расположении узлов интерполяции, имеет один существенный недостаток. Если понадобится увеличить число знаков (следовательно, и степень многочлена) прибавлением нового узла, многочлен Лагранжа придётся вычислять заново, так как каждый член его зависит от узлов интерполирования. Указанным недостатком не обладает интерполяционный многочлен Ньютона.

Пусть дана функция $y = f(x)$, $y_k = f(x_k)$ – значения ее в точках x_k, $(k = 0, 1, \ldots, n)$. Из первой разделённой разности $f(x, x_0) = = \dfrac{f(x) - y_0}{x - x_0}$ получим $f(x) = y_0 + (x - x_0)f(x, x_0)$. Поскольку $f(x, x_0, x_1) = = \dfrac{f(x, x_0) - f(x_0, x_1)}{x - x_1}$, то $f(x, x_0) = f(x_0, x_1) + (x - x_1)f(x, x_0, x_1)$. Следовательно,

$$f(x) = y_0 + (x - x_0)f(x, x_0) = y_0 + (x - x_0)f(x_0, x_1) +$$

$$+ (x - x_0)(x - x_1)f(x, x_0, x_1).$$

Продолжая процесс, получим

$$f(x) = y_0 + (x - x_0)f(x_0, x_1) + (x - x_0)(x - x_1)f(x_0, x_1, x_2) +$$

$$+ \ldots + (x - x_0)(x - x_1)\cdots(x - x_{n-1})f(x_0, x_1, \ldots, x_n) +$$

$$+(x-x_0)(x-x_1)\cdots(x-x_n)f(x,x_0,x_1,\cdots,x_n) \qquad (1)$$

или $f(x)=P_n(x)+(x-x_0)\ldots(x-x_n)f(x,x_0,x_1,\ldots,x_n)$, где

$$P_n(x)=y_0+(x-x_0)f(x_0,x_1)+\ldots+$$

$$+(x-x_0)(x-x_1)\cdots(x-x_{n-1})f(x_0,x_1,\ldots,x_n). \qquad (2)$$

Полагая в (1) $x=x_k$, получим $y_k=f(x_k)=P_n(x_k)$, $k=0,1,\ldots,n$, следовательно, многочлен (2) – интерполяционный многочлен для функции $y=f(x)$, построенный по $n+1$ узлам $x_0,\ldots,x_n$. Многочлен (2) называется интерполяционным многочленом Ньютона. Подставляя его в общую интерполяционную формулу, получим

$$f(x)\approx y_0+(x-x_0)f(x_0,x_1)+(x-x_0)(x-x_1)f(x_0,x_1,x_2)+$$

$$+\ldots+(x-x_0)(x-x_1)\cdots(x-x_{n-1})f(x_0,x_1,\ldots,x_n). \qquad (3)$$

В случае равноотстоящих узлов интерполяции $x_1=x_0+h$, $x_2=x_0+$ $+2h,\ldots,$ $x_n=x_0+nh$ из интерполяционной формулы Ньютона с учетом равенств $f(x_k,x_{k-1},\ldots,x_0)=\dfrac{\Delta^k y_0}{k!h^k}$, $k=1,2,\ldots,n$, получается интерполяционная формула Ньютона «интерполирования вперёд»

$$f(x)\approx y_0+\frac{\Delta y_0}{h}(x-x_0)+\frac{\Delta^2 y_0}{2!h}(x-x_0)(x-x_1)+$$

$$+\ldots+\frac{\Delta^n y_0}{n!h^n}(x-x_0)(x-x_1)\ldots(x-x_{n-1}). \qquad (4)$$

«Интерполирование вперёд» объясняется тем, что формула содержит заданные значения функции, соответствующие узлам интерполяции, находящимся только вправо от x_0, формула (4) удобна при интерполировании функций для значений x, близких к наименьшему узлу x_0. Положим $x = x_0 + ht$. Тогда

$$\frac{x-x_0}{h} = t, \quad \frac{(x-x_0)(x-x_1)}{h^2} = t(t-1),$$

$$\frac{(x-x_0)(x-x_1)\cdots(x-x_{n-1})}{h^n} = t(t-1)\cdots(t-n+1),$$

формула (3) примет вид

$$f(x) = f(x_0+th) \approx y_0 + t\Delta y_0 + \frac{t(t-1)}{2!}\Delta^2 y_0 +$$

$$+ \frac{t(t-1)(t-2)}{3!}\Delta^3 y_0 + \ldots + \frac{t(t-1)\cdots(t-n+1)}{n!}\Delta^n y_0. \tag{5}$$

Остаточный член для полинома (5) имеет вид

$$R_n(x) = \frac{h^{n+1}f^{(n+1)}(\xi)}{(n+1)!}t(t-1)\cdots(t-n), \text{ где } \xi \in [a,b].$$

Абсолютная погрешность метода по формуле Ньютона «интерполирование вперед» определяется неравенством

$$\left|R_n(x)\right| \le \frac{M_{n+1}h^{n+1}\left|t(t-1)\cdots(t-n)\right|}{(n+1)!},$$

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

Интерполяционную формулу Ньютона (3) можно записать так:

$$f(x) \approx y_n + (x - x_n) f(x_n, x_{n-1}) + (x - x_n)(x - x_{n-1}) f(x_n, x_{n-1}, x_{n-2}) +$$
$$+ \ldots + (x - x_n)(x - x_{n-1}) \cdots (x - x_1) f(x_n, \ldots, x_0).$$

В случае равноотстоящих узлов из неё аналогично формуле Ньютона «интерполирование вперёд» можно получить формулу Ньютона «интерполирование назад».

$$f(x) \approx y_n + t\Delta y_{n-1} + \frac{t(t+1)}{2!} \Delta^2 y_{n-2} + \frac{t(t+1)(t+2)}{3!} \Delta^3 y_{n-2} + \ldots +$$
$$+ \frac{t(t+1) \cdots (t+n-1)}{n!} \Delta^n y_0, \quad t = \frac{x - x_n}{h}.$$

Формулу «интерполирование назад» используют при интерполировании функций в точках x, близких к наибольшему узлу x_n.

Абсолютная погрешность метода «интерполирования назад» определяется формулой

$$\left| R_n(x) \right| \le \frac{M_{n+1} h^{n+1} \left| t(t+1) \cdots (t+n) \right|}{(n+1)!},$$

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

§12. Интерполирование внутри таблицы. Интерполяционная формула Стирлинга

Предположим, что точка x лежит вблизи внутреннего узла x_k таблицы с любой стороны от него. Тогда табличные узлы целесообразно привлекать для интерполирования в порядке удалённости от x_k, то есть взять сначала узел x_k и присоединить к нему пары узлов $(x_k + h, x_k - h)$, $(x_k + 2h, x_k - 2h), \ldots,$ $(x_k + nh, x_k - nh)$. При таком порядке узлов интерполирования формула Ньютона будет иметь вид

$$f(x) = y(x) = f(x_k) + (x - x_k)f(x_k, x_k + h) + (x - x_k)(x - x_k - h)\times$$

$$\times f(x_k, x_k + h, x_k - h) + (x - x_k)(x - x_k - h)(x - x_k + h)\times$$

$$\times f(x_k, x_k + h, x_k - h, x_k + 2h) + \ldots + (x - x_k)(x - x_k - h)\times$$

$$\times \ldots \times (x - x_k - nh) f(x_k, x_k + h, x_k - h, \ldots, x_k - nh) + R_{2n}(x),$$

где $R_{2n}(x) = \dfrac{(x - x_k)(x - x_k - h)\cdots(x - x_k + nh)}{(2n+1)!} f^{2n+1}(\xi)$, где ξ – есть точка отрезка, содержащего $x_k - nh$, $x_k + nh$ и x.

Заменим разностные отношения их выражениями через конечные разности

$$f(x_k) = y_k, \quad f(x_k, x_k + h) = \frac{\Delta y_k}{1!h},$$

$$f(x_k, x_k + h, x_k - h) = f(x_k - h, x_k, x_k + h) = \frac{\Delta^2 y_{k-1}}{2!h^2}, \ldots$$

и введём переменную $t = \dfrac{x - x_k}{h}$, тогда получим

$$y(x_k + th) = y_k + \frac{t}{1!}\Delta y_k + \frac{t(t-1)}{2!}\Delta^2 y_{k-1} + \frac{t(t-1)(t+1)}{3!}\Delta^3 y_{k-1} + \ldots +$$
$$+\frac{1}{(2n-1)!}(t-n+1)\cdots t(t+1)\cdots(t+n-1)\Delta^{2n-1} y_{k-n+1} +$$
$$+\frac{1}{(2n)!}(t-n)\cdots t(t+1)\cdots(t+n-1)\Delta^{2n} y_{k-n+1} + R_{2n}(x).$$

Для придания правой части симметричного вида, перепишем равенство в форме

$$y(x_k + th) = y_k + t\left[\Delta y_k - \frac{1}{2}\Delta^2 y_{k-1}\right] + \frac{t^2}{2!}\Delta^2 y_{k-1} +$$
$$+\frac{t\left(t^2 - 1^2\right)}{3!}\left[\Delta^3 y_{k-1} - \frac{1}{2}\Delta^4 y_{k-2}\right] +$$
$$+\ldots+\frac{t\left(t^2-1^2\right)\cdots\left(t^2-(n-1)^2\right)}{(2n-1)!}\left[\Delta^{2n-1} y_{k-n+1} - \frac{1}{2}\Delta^{2n} y_{k-n}\right] +$$
$$+\frac{t^2\left(t^2-1^2\right)\cdots\left(t^2-(n-1)^2\right)}{(2n)!}\Delta^{2n} y_{k-n} + R_{2n}(x).$$

Если из квадратных скобок исключить конечные разности четного порядка, пользуясь равенствами $\Delta^2 y_{k-1} = \Delta y_k - \Delta y_{k-1}$, $\Delta^4 y_{k-2} =$

$= \Delta^3 y_{k-1} - \Delta^3 y_{k-2}, \ldots,$ то получим интерполяционную формулу Ньютона-Стирлинга

$$y(x_k + th) = y_k + \frac{t}{1!}\frac{\Delta y_{k-1} + \Delta y_k}{2!} + \frac{t^2}{2!}\Delta^2 y_{k-1} +$$

$$+ \frac{t\left(t^2 - 1^2\right)}{3!}\frac{\Delta^3 y_{k-2} + \Delta^3 y_{k-1}}{2} + \frac{t^2\left(t^2 - 1^2\right)}{4!}\Delta^4 y_{k-2} +$$

$$+ \ldots + \frac{t\left(t^2 - 1\right)\cdots\left(t^2 - (n-1)^2\right)}{(2n-1)!}\frac{\Delta^{2n-1} y_{k-n} + \Delta^{2n-1} y_{k-n+1}}{2} +$$

$$+ \frac{t^2\left(t^2 - 1^2\right)\cdots\left(t^2 - (n-1)^2\right)}{(2n)!}\Delta^{2n} y_{k-n} + R_{2n}(x),$$

$$R_{2n}(x) = h^{2n+1}\frac{t^2\left(t^2 - 1^2\right)\cdots\left(t^2 - n^2\right)}{(2n+1)!}y^{(2n+1)}(\xi).$$

§13. Интерполирование с кратными узлами

Пусть на отрезке $[a, b]$ даны m различных узлов x_i, $(i = 1, 2, \ldots, m)$. Предположим, что в точке x_1 известны значения $f(x_1)$, $f'(x_1), \ldots,$ $f^{(\alpha_1-1)}(x_1)$, в точке x_2 известны значения $f(x_2)$, $f'(x_2), \ldots,$ $f^{(\alpha_2-1)}(x_2)$ и так далее. Числа $\alpha_1, \alpha_2, \ldots, \alpha_m$ называются кратностями узлов $x_1, x_2,$ $\ldots, x_m$. Общее число известных данных о функции f обозначим $n+1$: $\alpha_1+\alpha_2+\ldots+\alpha_m = n+1$. Возьмем многочлен степени n с произвольными коэффициентами

$$P_n(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0. \tag{1}$$

Выберем его коэффициенты a_j так, чтобы выполнялись условия

$$P_n^{(i)}(x_k) = f^{(i)}(x_k), \quad k = \overline{1, m}, \quad i = \overline{0, \alpha_k - 1}, \tag{2}$$

которые дают для a_j, $j = \overline{0, n}$ систему $n+1$ линейных уравнений.

Многочлен $P_n(x)$, удовлетворяющий условиям (2), называется многочленом Эрмита для функций $f(x)$. (В [7, с. 136] $P_n(x)$ обозначен $H_n(x)$). Нетрудно доказать, что система (2) имеет единственное решение. Для этого достаточно показать, что однородная система

$$P_n^{(i)}(x_k) = 0, \quad k = \overline{1, m}, \quad i = \overline{0, \alpha_k - 1} \tag{3}$$

имеет только нулевое решение. Действительно, система говорит о том, что для многочлена $P_n(x)$ каждый узел x_k является корнем кратности не меньше α_k. Значит для многочлена $P_n(x)$, степень которого равна n, сумма кратностей его корней не меньше, чем $\alpha_1+\alpha_2+\ldots+\alpha_m = n+1$. Такой многочлен тождественно равен 0, и поэтому его коэффициенты равны 0. Однородная система (3) имеет, следовательно, только нулевое решение, и существует лишь один многочлен (1), удовлетворяющий условиям (2).

Пусть отрезок $[a, b]$, содержащий узлы и точку x, конечен, а функция f – аналитическая в замкнутой области D, ограниченной контуром l и содержащей $[a, b]$ внутри.

Из системы (2) следует, что коэффициенты a_j будут линейными функциями свободных членов системы $f^{(i)}(x_k)$. Поэтому интерполирующий многочлен (1) также будет линейно выражаться через $f^{(i)}(x_k)$, и может быть записан в виде $P_n(f, x) = \sum\limits_{k=1}^{m}\sum\limits_{i=1}^{\alpha_{k-1}} L_{k,i}(x)f^{(i)}(x_k)$.

Справедлива

Теорема 1. *Пусть узлы x_k, $k=\overline{1,m}$ и точка x принадлежат отрезку $[a, b]$ и функция f имеет на $[a, b]$ непрерывную производную порядка $n+1$. Тогда на $[a, b]$ существует точка ξ такая, что для погрешности $R_n(x)$ интерполирования верно равенство $R_n(x) = \dfrac{A_n(x)}{(n+1)!}\times$ $\times f^{(n+1)}(\xi)$, где $A_n(x) = (x-x_1)^{\alpha_1}\cdots(x-x_m)^{\alpha_m}$. Здесь $R_n(x) = f(x)-$ $-P_n(x)$.*

§14. Многочлены Чебышева

Определение 1. *Многочлены Чебышева* $T_n(x)$, *где* $n \geq 0$, *определяются соотношениями*

$$T_0(x) = 1, \;\; T_1(x) = x, \;\; T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \text{ при } n \geq 1.$$

Например, $T_2(x) = 2x^2-1$, $T_3(x) = 4x^3-3x$, $T_4(x) = 8x^4-8x^2+1, \ldots$ Старший член $T_{n+1}(x)$ получается из старшего члена $T_n(x)$ умножением на $2x$ и, следовательно, старший член $T_n(x)$ при $n > 0$ есть $2^{n-1}x^n$. Все многочлены $T_{2n}(x)$ – чётные, а $T_{2n+1}(x)$ – нечётные.

Пример. $T_n(x) = \cos(n \arccos x)$. $T_0(x) = \cos(0 \cdot \arccos x) = \cos 0 = 1$, $T_1(x) = \cos(1 \cdot \arccos x) = x$, $\cos\big((n+1)\theta\big) + \cos\big((n-1)\theta\big) = 2\cos n\theta \cdot \cos\theta$. При $\forall\theta$ имеем $\cos\big((n+1)\theta\big) = 2\cos\theta\cos n\theta - \cos\big((n-1)\theta\big)$. Полагая $\theta = \arccos x$, получим

$$\cos\big((n+1)\theta\big) = \cos\big((n+1)\arccos x\big) =$$

$$= 2x\cos(n\arccos x) - \cos\big((n-1)\arccos x\big).$$

Следовательно, $T_n(x) = \cos(n\arccos x)$ – многочлен Чебышева.

Определение 2. *Многочлены* $\overline{T_n(x)} = 2^{1-n}T_n(x) = x^n + \ldots$. *называют многочленами, наименее уклоняющимися от нуля.*

Это определение объясняется следующей леммой.

Лемма 1. *Если $P_n(x)$ – многочлен степени n со старшим коэффициентом 1, то*

$$\max_{[-1,1]}\left|P_n(x)\right| \geq \max_{[-1,1]}\left|\overline{T_n(x)}\right| = 2^{1-n}.$$

Линейной заменой $x' = \dfrac{b+a}{2} + \dfrac{b-a}{2}x$ отрезок $[-1,1]$ можно перевести в $[a,b]$. Многочлен $T_n(x)$ при этом преобразуется в $T_n\left(\dfrac{2x-(b+a)}{b-a}\right)$ со старшим коэффициентом $\left(\dfrac{2}{b-a}\right)^n$, т.е. $\overline{T_n}^{[a,b]}(x) = (b-a)^n 2^{1-2n} \times$ $\times T_n\left(\dfrac{2x-(b+a)}{b-a}\right)$ со старшим коэффициентом 1 является многочленом, наименее уклоняющимся от нуля на $[a,b]$, то есть

$$\max_{[a,b]}\left|P_n(x)\right| \geq \max_{[a,b]}\left|\overline{T_n}^{[a,b]}(x)\right| = (b-a)^n 2^{1-2n}. \quad (1)$$

Отметим, что многочлены $\widetilde{T}_0(x) = \dfrac{1}{\sqrt{2}}$ и $\widetilde{T}_n(x) = T_n(x)$ при $n \geq 1$ образуют на $[-1,1]$ ортонормированную систему с весом $\dfrac{2}{\pi\sqrt{1-x^2}}$.

14.1 Приложение многочленов Чебышева

Пусть $f(x)$ на $[a,b]$ приближается с помощью полинома $L_{n-1}(x)$ степени $n-1$ с узлом интерполяции $x_1, x_2, \ldots, x_n$ и $f(x) - L_{n-1}(x) = = \dfrac{f^{(n)}(\xi)\omega_n(x)}{n!}$, где $\xi \in [a,b]$. Тогда

$$\|f - L_{n-1}\| \leq \frac{\left\|f^{(n)}\right\|\|\omega_n\|}{n!}, \text{ где } \left\|f^{(n)}\right\| = \sup_{[a,b]}\left\|f^{(n)}(x)\right\|. \tag{2}$$

Минимизируем правую часть оценки за счет выбора узлов $x_1, x_2, \ldots, x_n$, для этой цели и были введены многочлены Чебышева, наименее уклоняющиеся от нуля. Многочлены $\omega_n(x) = (x - x_1)(x - x_2)\ldots(x - x_n)$ имеют старший коэффициент 1 и поэтому $\|\omega_n\| \geq (b-a)^n 2^{1-2n}$ согласно (1). Если взять в качестве узлов интерполирования $x_m = \dfrac{b+a}{2} + + \dfrac{b-a}{2}\cos\left(\dfrac{\pi(2m+1)}{2n}\right)$, $m = 0, 1, \ldots, n-1$, то

$$\omega_n(x) = (b-a)^n 2^{1-2n} T_n\left(\frac{2x-(b+a)}{b-a}\right) \text{ и } \|\omega_n\| = (b-a)^n 2^{1-2n}.$$

Следовательно, при таком расположении узлов справедлива наилучшая из оценок, которая получается из оценки (2)

$$\|f - L_{n-1}\| \leq \frac{\left\|f^{(n)}\right\|(b-a)^n 2^{1-2n}}{n!}.$$

§15. Численное дифференцирование. (Применение интерполирования к вычислению производных)

Численное дифференцирование применяется, если

1) функция задана таблично,

2) функция задана неудобным для дифференцирования аналитическим выражением.

Задача численного дифференцирования некорректна, так как нарушается условие 3 корректности (решение непрерывно зависит от входных данных).

При численном дифференцировании функцию $f(x)$ заменяют интерполяционным многочленом $P_n(x)$ и приближенно полагают $f'(x) = P_n'(x)$.

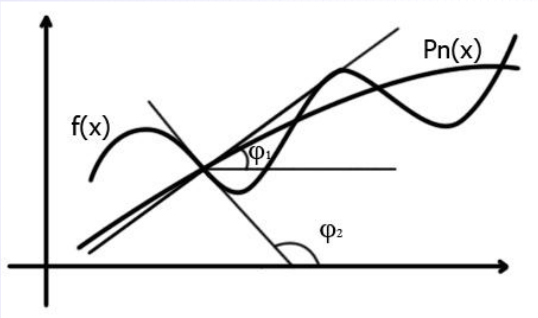

Рисунок 1.

Близость значений функции $f(x)$ и полинома $P_n(x)$ не гарантирует близости их угловых коэффициентов φ_2 и φ_1.

Пусть на отрезке $[a, b]$ рассматривается функция $f(x)$, имеющая непрерывную производную порядка $n+1$. Возьмём на $[a, b]$ $n+1$ различных узлов $x_0, x_1, x_2, \ldots, x_n$. Для упрощения записи предположим, что они перенумерованы слева направо так, что $x_0 < x_1 < x_2 < \ldots < x_n$. Интерполируем $f(x)$ по её значениям $f(x_0), f(x_1), \ldots, f(x_n)$ в узлах $x_0, x_1, \ldots, x_n$ посредством многочлена $P_n(x)$ степени n и обозначим $R_n(x)$ погрешность интегрирования:

$$f(x) = P_n(x) + R_n(x),$$

причем $P_n(x_i) = f(x_i), \quad i = \overline{0, n}$.

Вычислим производную от f порядка m:

$$f^{(m)}(x) = P_n^{(m)}(x) + R_n^{(m)}(x). \tag{1}$$

Пренебрегая величиной $R_n^{(m)}(x)$, получим формулу для приближенного вычисления производной:

$$f^{(m)}(x) \approx P_n^{(m)}(x). \tag{2}$$

Ее погрешность равна $R_n^{(m)}(x)$.

Пользоваться ею целесообразно при небольших порядках m производной, во всяком случае, когда $m \leq n$, так как все производные от $P_n(x)$ выше n тождественно равны 0.

Будем считать, что в точке x $\omega^{(m)}(x) \neq 0$ и на отрезке $[\alpha, \beta]$ производная $\varphi^{(m)}(t) = R_n^{(m)}(t) - \dfrac{K}{(n+1)!}\omega^{(m)}(t)$ имеет $n+2-m$ нулей (отрезок $[\alpha, \beta]$ – наименьший отрезок, содержащий точки x_0, x_n, x). K выбирается из условия так, чтобы точка x была нулем функции $\varphi^{(m)}(t)$, т.е. $\varphi^{(m)}(x) = 0$. Тогда для погрешности $R_n^{(m)}(x)$ вычислительной формулы (2) верно представление:

$$R_n^{(m)}(x) = \frac{\omega^{(m)}(x)}{(n+1)!} f^{(n+1)}(\xi), \quad \xi \in [\alpha, \beta],$$

$$\omega(x) = (x - x_0)(x - x_1) \cdots (x - x_n).$$

§16. Некоторые частные формулы вычисления производных

Каждая из формул для интерполяционного многочлена (Лагранжа, Ньютона) может служить источником для получения формулы вычисления производных. Таких формул можно получить большое число, но для выяснения идеи их построения достаточно ограничиться несколькими примерами. Возьмем формулу Ньютона:

$$P_n(x) = f(x_0) + (x - x_0)f(x_0, x_1) + (x - x_0)(x - x_1)f(x_0, x_1, x_2)+$$

$$+ \ldots + (x - x_0)(x - x_1)\cdots(x - x_{n-1})f(x_0, x_1, \ldots, x_n). \qquad (1)$$

Введем сокращенные обозначения $x - x_k = \alpha_k$, тогда (1) запишется в виде:

$$P_n(x) = f(x_0) + \alpha_0 f(x_0, x_1) + \alpha_0\alpha_1 f(x_0, x_1, x_2)+$$

$$+ \ldots + \alpha_0\alpha_1\cdots\alpha_{n-1} f(x_0, x_1, \ldots, x_n).$$

Последовательное дифференцирование этого равенства дает следующие выражения производных от f:

$$f'(x) \approx P_n'(x) = f(x_0, x_1) + (\alpha_0 + \alpha_1)f(x_0, x_1, x_2)+$$

$$+(\alpha_0\alpha_1 + \alpha_0\alpha_2 + \alpha_1\alpha_2)f(x_0, x_1, x_2, x_3) + \ldots,$$

$$\frac{1}{2!}f''(x) \approx \frac{1}{2!}P_n''(x) = f(x_0, x_1, x_2) + (\alpha_0 + \alpha_1 + \alpha_2)f(x_0, x_1, x_2, x_3) + \ldots,$$

$$\frac{1}{3!}f^{(3)}(x) \approx \frac{1}{3!}P_n^{(3)}(x) = f(x_0, x_1, x_2, x_3)+$$

$$+(\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3)f(x_0, x_1, x_2, x_3, x_4) + \ldots,$$

$$\frac{1}{4!}f^{(4)}(x) \approx \frac{1}{4!}P_n^{(4)}(x) = f(x_0, x_1, x_2, x_3, x_4)+$$

$$+(\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)f(x_0, x_1, x_2, x_3, x_4, x_5) + \ldots.$$

Аналогично можно получить формулы для вычисления производных в случае равноотстоящих узлов. Возьмем формулу Ньютона для интерполирования вперед (в начале таблицы).

Пусть $x = x_0 + th$, $t = \dfrac{x - x_0}{h}$, тогда $\dfrac{dt}{dx} = \dfrac{1}{h}$,

$$y(x) = P_n(x) = y(x_0 + th) = y_0 + \frac{t}{1!}\Delta y_0 + \frac{t(t-1)}{2!}\Delta^2 y_0 + \ldots +$$

$$+\frac{t(t-1)\cdots(t-n+1)}{n!}\Delta^n y_0 + R_n(x)$$

и вычислим производную по переменной t, то получим:

$$P_n'(t) = \left(y_0 + \frac{t}{1!}\Delta y_0 + \frac{t^2 - t}{2!}\Delta^2 y_0 + \frac{t^3 - 3t^2 + 2t}{3!}\Delta^3 y_0 + \ldots +\right.$$

$$\left. + \frac{t(t-1)\ldots(t-n+1)}{n!}\Delta^n y_0\right)';$$

$$\frac{dP}{dx} = \frac{dP}{dt}\frac{dt}{dx} = P'_t\left(\frac{x - x_0}{h}\right)'_x = \frac{1}{h}P'_t = \frac{1}{h}\left(\Delta y_0 + \frac{2t-1}{2!}\Delta^2 y_0 + \right.$$

$$\left. + \frac{3t^2 - 6t + 2}{3!}\Delta^3 y_0 + \frac{4t^3 - 18t^2 + 22t - 6}{4!}\Delta^4 y_0 + \ldots\right).$$

Найдем вторую производную:

$$\frac{d^2P}{dx^2} = P''_{x^2} = \frac{d}{dx}\left(\frac{dP}{dx}\right) = \frac{d}{dt}\left(\frac{dP}{dx}\right)\frac{dt}{dx} =$$

$$= \frac{1}{h}\frac{1}{h}\left(\Delta^2 y_0 + \frac{6t-6}{3!}\Delta^3 y_0 + \frac{12t^2 - 36t + 22}{4!}\Delta^4 y_0 + \ldots\right) =$$

$$= \frac{1}{h^2}\left(\Delta^2 y_0 + (t-1)\Delta^3 y_0 + \frac{6t^2 - 18t + 11}{12}\Delta^4 y_0 + \ldots\right).$$

Найдем третью производную:

$$P'''_{x^3} = \frac{1}{h^3}\left(\Delta^3 y_0 + \frac{2t-3}{2}\Delta^4 y_0 + \ldots\right).$$

§17. Интерполяционные методы решения нелинейных уравнений

Идея интерполяционных методов состоит в том, что нахождение корней уравнения

$$f(x) = 0 \tag{1}$$

заменяется нахождением корней интерполяционного многочлена, построенного для $f(x)$. Интерполяционный многочлен Ньютона 1-го порядка приводит к методу секущих. Интерполяционный многочлен Ньютона 2-го порядка называют методом парабол. Метод Ньютона можно получить, заменяя $f(x)$ интерполяционным многочленом Эрмита первой степени.

Получим формулы *метода парабол.* Пусть приближения x_{k-2}, x_{k-1}, x_k известны. Построим интерполяционный многочлен Ньютона:

$$P_2(x) = f(x_k) + (x - x_k)f(x_k, x_{k-1}) + (x - x_k)(x - x_{k-1})f(x_k, x_{k-1}, x_{k-2})$$

и обозначим $z = x - x_k$. Тогда уравнение $P_2(x) = 0$ примет вид:

$$az^2 + bz + c = 0, \tag{2}$$

где $x - x_{k-1} = (x - x_k) + (x_k - x_{k-1})$, $a = f(x_k, x_{k-1}, x_{k-2})$, $b = f(x_k, x_{k-1}) + (x_k - x_{k-1})f(x_k, x_{k-1}, x_{k-2})$, $c = f(x_k)$.

Решая уравнение (2), получим два, может быть, комплексных корня $z^{(1)}$ и $z^{(2)}$, по которым вычислим $x^{(1)} = x_k + z^{(1)}$, $x^{(2)} = x_k + z^{(2)}$.

В качестве следующего приближения x_{k+1} в методе парабол выбирается то из значений $x^{(1)}$, $x^{(2)}$, которое ближе к x_k, т.е. отвечающее минимальному по модулю корню уравнения (2). Метод парабол удобен тем, что позволяет получить комплексные корни уравнения (1), пользуясь вещественными начальными приближениями x_0, x_1, x_2.

§18. Интерполяция и приближение сплайнами

Интерполирование многочленом Лагранжа или Ньютона на $[a, b]$ с использованием большого числа узлов часто приводит к плохому приближению, что объясняется сильным накоплением погрешности в процессе вычислений.

Кроме того, *из-за расходимости процесса интерполяции увеличение числа узлов не обязано приводить к повышению точности.* Для того, чтобы избежать больших погрешностей, весь отрезок $[a, b]$ разбивают на частичные отрезки и на каждом из частичных отрезков приближенно заменяют функцию $f(x)$ многочленом невысокой степени (так называемая кусочно-полиномиальная интерполяция). Одним из способов такой интерполяции является интерполирование с помощью сплайн-функций.

Для определенности будем говорить о приближении функции $f(x)$ на [0,1]. Отрезок разбиваем на части $[x_0, x_1], \ldots, [x_{N-1}, x_N]$, $x_0 = 0$, $x_N = 1$. Обозначим это разбиение Δ.

Назовем *сплайном* $S^m_\Delta(f, x)$ порядка m функцию, являющуюся многочленом степени m на каждом из отрезков $[x_{n-1}, x_n]$:

$$S^m_\Delta(f, x) = P_{nm}(x) = a_{n0} + \ldots + a_{nm}x^m, \tag{1}$$

при $x_{n-1} \leq x \leq x_n$, и удовлетворяющую условиям непрерывности производных до порядка $m-1$ в точках $x_1, \ldots, x_{N-1}$.

$$P^{(k)}_{nm}(x_n) = P^{(k)}_{n+1,m}(x_n), \tag{2}$$

при $k = \overline{0, m-1}$, $n = 1, \ldots, N-1$.

Всего имеется в распоряжении $Q = N(m+1)$ неизвестных a_{nm}, соотношения (2) образуют систему из $(N-1)m$ уравнений. Другие уравнения для коэффициентов получаются из условия близости функции и из некоторых дополнительных условий ($N+1$ условие за счет совпадения сплайна с $f(x)$ в точках $x_0, x_1, \ldots, x_N$).

Рассмотрим простейшую задачу приближения линейными сплайнами $(m = 1)$. Тогда общее число неизвестных $Q = N(m+1) = 2N$. Построим сплайн $S^1_\Delta(f, x) = P_{n1}(x) = a_{n0} + a_{n1}x$, совпадающий с $f(x)$ в точках $x_0, \ldots, x_N$. Получится система уравнений:

$$P_{n1}(x_{n-1}) = f(x_{n-1}), \quad n = \overline{1, N},$$
$$P_{n1}(x_n) = f(x_n), \quad n = \overline{1, N}.$$

Эта система распадается на системы уравнений относительно коэффициентов отдельных многочленов:

$$P_{n1}(x_{n-1}) = a_{n0} + a_{n1}x_{n-1} = f(x_{n-1}),$$
$$P_{n1}(x_n) = a_{n0} + a_{n1}x_n = f(x_n).$$

Отсюда: $a_{n1} = \dfrac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}$, $a_{n0} = f(x_{n-1}) - a_{n1}x_{n-1}$.

Многочлен $P_{n1}(x)$ – интерполяционный многочлен первой степени с узлами интерполирования x_{n-1}, x_n. Получаемый сплайн совпадает с $f(x)$ во всех узлах.

Преимуществом сплайнов перед обычной интерполяцией является, во-первых, их сходимость, а, во-вторых, устойчивость процесса вычислений.

§19. Построение кубического сплайна

Пусть на отрезке $[a, b]$ задана непрерывная функция $f(x)$. Введем сетку $a = x_0 < x_1 < \ldots < x_{N-1} < x_N = b$ и обозначим $f_i = f(x_i)$, $i = 0, 1, \ldots, N$.

Определение 1. *Сплайном, соответствующим функции $f(x)$ и данным узлам $\{x_i\}$, $i = \overline{0, N}$, называется функция $S(x)$, удовлетворяющая следующим условиям:*

а) на каждом сегменте $[x_{i-1}, x_i]$, $i = 1, 2, \ldots, N$, функция $S(x)$ является многочленом третьей степени;

б) функция $S(x)$, а также её первая и вторая производные непрерывны на $[a, b]$;

в) $S(x_i) = f(x_i)$, $i = 0, 1, \ldots, N$.

Последнее условие называется *условием интерполирования*, а сплайн, определяемый условиями а) – в) называется также *интерполяционным кубическим сплайном*.

Докажем *существование и единственность сплайна*, определяемого условиями а) – в), т.е. покажем способ построения кубического сплайна.

На каждом из отрезков $[x_{i-1}, x_i]$, $i = 1, 2, \ldots, N$ будем искать функцию $S(x) = S_i(x)$ в виде многочлена третьей степени

$$S_i(x) = a_i + b_i(x - x_i) + \frac{c_i}{2}(x - x_i)^2 + \frac{d_i}{6}(x - x_i)^3, \qquad (1)$$

Начало
Содержание
◄ ►
◄◄ ►►
Страница 71 из 403
Назад
На весь экран
Закрыть

$x_{i-1} \le x \le x_i$, $i = 1, 2, \ldots, N$, где a_i, b_i, c_i, d_i – коэффициенты, подлежащие определению. Поясним смысл введенных коэффициентов. Имеем:

$$S_i'(x) = b_i + c_i(x - x_i) + \frac{d_i}{2}(x - x_i)^2; \quad S_i''(x) = c_i + d_i(x - x_i); \quad S_i'''(x) = d_i,$$

поэтому $a_i = S_i(x_i)$, $b_i = S_i'(x_i)$, $c_i = S_i''(x_i)$, $d_i = S_i'''(x_i)$.

Из условий интерполирования $S(x_i) = f(x_i)$, $i = 1, 2, \ldots, N$, получаем, что $a_i = f(x_i)$, $i = 1, 2, \ldots, N$. Доопределим, кроме того, $a_0 = f(x_0)$. Далее, требование непрерывности функции $S(x)$ приводит к условиям: $S_i(x_i) = S_{i+1}(x_i)$, $i = 1, 2, \ldots, N-1$. Отсюда, учитывая выражение для функций $S_i(x)$, получаем при $i = 0, 1, 2, \ldots, N-1$ уравнения:

$$a_i = a_{i+1} + b_{i+1}(x_i - x_i + 1) + \frac{c_{i+1}}{2}(x_i - x_{i+1})^2 + \frac{d_{i+1}}{6}(x_i - x_{i+1})^3. \tag{2}$$

Обозначим $h_i = x_i - x_{i-1}$ и перепишем это уравнение в виде:

$$h_i b_i - \frac{h_i^2}{2}c_i + \frac{h_i^3}{6}d_i = f_i - f_{i-1}, \quad i = \overline{1, N}. \tag{3}$$

Условия непрерывности первой производной $S_i'(x_i) = S_{i+1}'(x_i)$, $i = \overline{1, N-1}$, приводят к уравнениям:

$$c_i h_i - \frac{d_i}{2}h_i^2 = b_i - b_{i-1}, \quad i = \overline{2, N}. \tag{4}$$

Из условий непрерывности второй производной получаем уравнения:

$$d_i h_i = c_i - c_{i-1}, \quad i = \overline{2, N}. \tag{5}$$

Объединяя (3) – (5), получаем систему $3N - 2$ уравнений относительно $3N$ неизвестных b_i, c_i, d_i, $i = 1, 2, \ldots, N$. Два недостающих уравнения получают, задавая те или иные граничные условия для $S(x)$. Предположим, например, что функция $f(x)$ удовлетворяет условиям $f''(a) = f''(b) = 0$, т.е. $S''(a) = S''(b) = 0$. Отсюда получаем $S_1''(x_0) = 0$, $S_N''(x_N) = 0$, т.е. $c_1 - d_1 h_1 = 0$, $c_N = 0$. Заметим, что условие $c_1 - d_1 h_1 = 0$ совпадает с уравнением (5) при $i = 1$, если положить $c_0 = 0$. Таким образом, приходим к замкнутой системе уравнений для определения коэффициентов кубического сплайна:

$$h_i d_i = c_i - c_{i-1}, \quad i = \overline{2, N}, \quad c_0 = c_N = 0, \tag{6}$$

$$h_i c_i - \frac{h_i^2}{2} d_i = b_i - b_{i-1}, \quad i = \overline{2, N}, \tag{7}$$

$$h_i b_i - \frac{h_i^2}{2} c_i + \frac{h_i^3}{6} d_i = f_i - f_{i-1}, \quad i = \overline{1, N}. \tag{8}$$

Эта система имеет единственное решение (см. [7, с. 143]). Следовательно, существует единственный кубический сплайн, определяемый условиями а) – в) и граничными условиями $S''(a) = S''(b) = 0$. Интерполирование кубическими сплайнами является *сходящимся процессом*, т.е. при неограниченном увеличении числа узлов N соответствующая последовательность сплайн-функций сходится к интерполируемой функции $f(x)$. Оценки погрешности интерполяции зависят от выбора сеток и от гладкости функции $f(x)$. Справедлива

Теорема 1. *Для* $f \in C^{(4)}[a,b]$ *справедливы оценки:*

$$\left\|f(x) - S_h(x)\right\|_{C[a,b]} \le M_4 h^4, \quad \left\|f'(x) - S_h'(x)\right\|_{C[a,b]} \le M_4 h^3,$$

$$\left\|f''(x) - S_h''(x)\right\|_{C[a,b]} \le M_4 h^2,$$

где $S_h(x)$ – *кубический сплайн, построенный для функции* $f(x)$ *на сетке* $\omega_h = \{x_i = a + ih,\ i = \overline{0, N}\}$, *а* $M_4 = \left\|f^{(4)}(x)\right\|_{C[a,b]}$.

Из этих оценок следует, что при $h \to 0$ (т.е. при $N \to \infty$) последовательности $S_h^{(i)}(x)$, $i = 0, 1, 2$, сходятся соответственно к функциям $f^{(i)}(x)$, $i = 0, 1, 2$.

§20. Многомерная интерполяция

Двумерные таблицы широко распространены в физике и технике, например, таковыми являются *таблицы термодинамических функций газов*, где независимыми переменными обычно являются температура и плотность. Трехмерные таблицы составляют и используют значительно реже, но не потому что таких зависимостей нет, а потому, что таблицы слишком громоздки. Четырехмерных таблиц практически нет, хотя в физике немало задач с большим объемом параметров; так *проводимость плазмы* $\sigma(T, \rho, E, H)$ зависит от температуры и плотности, напряженностей электрического и магнитного полей. Для простоты рассмотрим двумерные таблицы (двумерную интерполяцию); обобщить все результаты на большее число измерений нетрудно.

20.1 Двумерная интерполяция

На плоскости xOy заданы три точки $A_i(x_i, y_i)$, $i = 1, 2, 3$, не лежащие на одной прямой. Требуется, используя значения $u_i = u(x_i, y_i)$ функции $u(x, y)$ в этих точках, построить аппроксимацию производных $\dfrac{du}{dx}, \dfrac{du}{dy}$. Для решения этой задачи воспользуемся линейной интерполяцией, т.е. будем считать, что

$$u(x, y) = a(x - x_1) + b(y - y_1) + c. \qquad (1)$$

Тогда получим, что $\dfrac{du}{dx} = a$, $\dfrac{du}{dy} = b$, т.е. при интерполяции функции $u(x, y)$ с помощью линейной функции производные заменяются константами. Явные выражения для коэффициентов a, b, c нетрудно найти из условия интерполирования $u(x_i, y_i) = u_i$. Действительно, из условия $u(x_1, y_1) = u_1$ получаем, что $c = u_1$. Далее, решая систему:

$$a(x_2 - x_1) + b(y_2 - y_1) = u_2 - u_1, \quad a(x_3 - x_1) + b(y_3 - y_1) = u_3 - u_1,$$

получим:

$$a = \frac{1}{\Delta}\begin{vmatrix} u_2 - u_1 & y_2 - y_1 \\ u_3 - u_1 & y_3 - y_1 \end{vmatrix}, \quad b = \frac{1}{\Delta}\begin{vmatrix} x_2 - x_1 & u_2 - u_1 \\ x_3 - x_1 & u_3 - u_1 \end{vmatrix},$$

$$\Delta = \begin{vmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \end{vmatrix}. \qquad (2)$$

Выражения (2) и задают искомые приближения к производным $\dfrac{du}{dx}$, $\dfrac{du}{dy}$. Определитель Δ данной системы не равен 0, так как по условию точки A_1, A_2, A_3 не лежат на одной прямой.

Замечание 1. *Чтобы объем таблиц был приемлем, приходится шаги по аргументам брать довольно большими. Это предъявляет жесткие требования к способу интерполяции.*

Замечание 2. *Не любое число узлов интерполяции выгодно.* Если для одной переменной степень многочлена была взаимно-однозначно связана с числом узлов, то для двух переменных многочлен n-ой степени $P_n(x,y) = \sum\limits_{k+m=0}^{n} a_{km}x^k y^m$ имеет $\dfrac{(n+1)(n+2)}{2}$ узлов. Если число узлов не соответствует этой формуле, то часть коэффициентов при высших степенях должна задаваться принудительно (в частности, нулями), для выбора этих коэффициентов редко есть разумные основания.

Замечание 3. *Не всякое расположение узлов допустимо.* В одномерном случае узлы не должны совпадать. В двумерном случае узлы не должны лежать на одной прямой в плоскости (x,y), потому что тогда $\Delta = 0$ (для интерполяции многочленом $P_1(x,y)$). При интерполяции многочленом $P_2(x,y)$ требуется, чтобы узлы не лежали на кривой второго порядка и т.д.

§21. Наилучшее приближение функции в линейном нормированном пространстве

Рассмотрим следующую задачу. Имеется элемент f линейного нормированного пространства L. Требуется найти его наилучшее приближение линейной комбинацией $\sum\limits_{i=1}^{n} c_i g_i$, где $g_1, g_2, \ldots, g_n \in L$ – линейно независимы. Это значит, нужно найти такой элемент $\sum\limits_{i=1}^{n} c_i^0 g_i$, что

$$\left\| f - \sum_{i=1}^{n} c_i^0 g_i \right\| = \Delta = \inf_{c_1, c_2, \ldots, c_n} \left\| f - \sum_{i=1}^{n} c_i g_i \right\|.$$

Если такой элемент существует, то он называется *элементом наилучшего приближения*.

Нетрудно доказать (см. 1, с. 211), что элемент наилучшего приближения существует. Но таких элементов, вообще говоря, может быть несколько.

Определение 1. *Линейное нормированное пространство называется строго нормированным, если равенство* $\|f_1 + f_2\| = \|f_1\| + \|f_2\|$ *возможно тогда и только тогда, когда* $f_2 = \alpha f_1$, $\alpha > 0$.

Справедлива

Теорема 1. *Если пространство L строго нормировано, то элемент наилучшего приближения единственен.*

Гильбертово пространство строго нормировано. Пространство $L_p(0,1)$ с нормой $\|f\|_{L_p} = \sqrt[p]{\int\limits_0^1 \left|f(x)\right|^p dx}$ строго нормировано при $1 < p < \infty$. Пространство непрерывных функций $C[0,1]$ с нормой $\|f\|_C = \max\limits_{[0,1]} \left|f(x)\right|$ является строго нормированным.

Пусть функция $f(x)$ задана таблично. Обозначим через $g(x) = \sum\limits_{i=1}^{n} c_i g_i(x)$ и пусть значения функций $f(x)$ и $g(x)$ известны в точках $x_k \in [a,b]$, $k = 0,1,\ldots,m$. Обозначим $r_k = g(x_k) - f(x_k)$, $k = 0,1,\ldots,m$.

Для вектора погрешностей $r = (r_0, r_1, \ldots, r_m)^T$ можно ввести ту или иную норму, например:

$$\|r\| = \left(\sum_{k=0}^{m} r_k^2\right)^{\frac{1}{2}} = \left[\sum_{k=0}^{m} \left(g(x_k) - f(x_k)\right)^2\right]^{\frac{1}{2}} \qquad (1)$$

или

$$\|r\| = \max_{0\le k\le m} |r_k| = \max_{0\le k\le m} \left|g(x_k) - f(x_k)\right|. \qquad (2)$$

Задача о наилучшем приближении функции $f(x)$, заданной таблично состоит в нахождении коэффициентов $c_1, c_2, \ldots, c_n$, минимизиру-

ющих норму вектора r. В зависимости от выбора нормы получим различные задачи. Так норме (1) соответствует задача о наилучшем квадратичном приближении, а норме (2) – задача о наилучшем равномерном приближении функции, заданной таблично.

§22. Наилучшие приближения функций в гильбертовом пространстве

Пусть дано гильбертово пространство H, может быть бесконечномерное, и в нем задана конечная система линейно-независимых элементов

$$\varphi_k \in H, \quad k = \overline{0, n}. \tag{1}$$

Требуется приближенно заменить заданный элемент $f \in H$ линейной комбинацией

$$\varphi = c_0\varphi_0 + c_1\varphi_1 + \ldots + c_n\varphi_n. \tag{2}$$

Элемент φ, определенный согласно (2), называется обобщенным многочленом, построенным по системе элементов (1). Будем рассматривать задачу о наилучшем приближении, состоящую в том, чтобы для заданного элемента $f \in H$ среди всех линейных комбинаций вида (2) найти такой обобщенный многочлен φ, для которого отклонение

$$\left\| f - (c_0\varphi_0 + c_1\varphi_1 + \ldots + c_n\varphi_n) \right\| \tag{3}$$

было бы минимальным. Элемент

$$\overline{\varphi} = \overline{c_0}\varphi_0 + \overline{c_1}\varphi_1 + \ldots + \overline{c_n}\varphi_n, \tag{4}$$

дающий решение этой задачи, называется элементом наилучшего приближения.

Пусть H – вещественное гильбертово пространство со скалярным произведением (f, g) и нормой $\|f\| = \sqrt{(f, f)}$. В данном случае задача о наилучшем приближении состоит в нахождении элемента $\overline{\varphi} \in H$, для которого отклонение

$$\left\|f - \overline{\varphi}\right\|_H = \left(f - \overline{\varphi}, f - \overline{\varphi}\right)^{\frac{1}{2}} \tag{5}$$

является минимальным среди всех многочленов вида $\varphi = c_0\varphi_0 + c_1\varphi_1 + \ldots + c_n\varphi_n$.

22.1 Сведение к алгебраической задаче о минимуме квадратичного функционала

Перепишем равенство (5) в виде:

$$\left\|f - \varphi\right\|^2 = \left(f - \sum_{i=0}^{n} c_i\varphi_i, f - \sum_{i=0}^{n} c_i\varphi_i\right) =$$

$$= \sum_{l,k=0}^{n} c_k c_l \left(\varphi_k, \varphi_l\right) - 2\sum_{k=0}^{n} c_k \left(f, \varphi_k\right) + \|f\|^2. \tag{6}$$

Пусть $A = \left[a_{kl}\right]$ – матрица с элементами

$$a_{kl} = \left(\varphi_k, \varphi_l\right), \quad k, l = 0, 1, \ldots, n \tag{7}$$

и c, $\widehat{f}$ – векторы $c = (c_0, c_1, \ldots, c_n)^T$, $\widehat{f} = (f_0, f_1, \ldots, f_n)^T$, где $f_i = (f, \varphi_i)$, $i = \overline{0, n}$. Обозначая через $(u, v) = \sum\limits_{i=0}^{n} u_i v_i$ скалярное произведение векторов u и v, можно записать тождество в виде:

$$\left\|f - \overline{\varphi}\right\|^2 = (Ac, c) - 2(\widehat{f}, c) + \left\|f\right\|^2. \tag{8}$$

Отсюда видно, что задача о нахождении наилучшего приближения в гильбертовом пространстве H сводится к минимизации функционала

$$F(c) = (Ac, c) - 2(\widehat{f}, c). \tag{9}$$

Здесь A – матрица симметрическая, так как скалярное произведение симметрично: $a_{ki} = (\varphi_k, \varphi_i) = (\varphi_i, \varphi_k) = a_{ik}$. Кроме того, A – положительно определенная матрица, т.е. $(Ac, c) > 0$, для всех $c \neq 0$.

Следующая теорема сводит проблему минимизации квадратичного функционала (9) к решению некоторой системы линейных алгебраических уравнений.

Теорема 1. *Пусть A – симметрическая положительно определенная матрица и $\widehat{f}$ – заданный вектор. Тогда функционал (9) имеет единственную точку минимума $\overline{c}$. Вектор $\overline{c}$ удовлетворяет системе уравнений*

$$A\overline{c} = \widehat{f}. \tag{10}$$

Более подробно систему (10) можно записать в виде:

$$\sum_{i=0}^{n} \overline{c_i}(\varphi_k, \varphi_i) = (f, \varphi_k), \quad k = 0, 1, \dots, n. \tag{11}$$

Таким образом, *элемент наилучшего приближения в пространстве H имеет вид* (4), *где коэффициенты* $\overline{c_k}$, $k = 0, 1, \dots, n$, *отыскиваются из системы* (11). Итак, алгоритм построения элемента наилучшего приближения в гильбертовом пространстве состоит в следующем:

1) вычисляются элементы $a_{kl} = (\varphi_k, \varphi_l)$, $k, l = 0, 1, \dots, n$, матрицы A;

2) вычисление правых частей (f, φ_k), $k = 0, 1, \dots, n$;

3) решение системы (11);

4) вычисление суммы $\overline{\varphi} = \sum\limits_{k=0}^{n} \overline{c_k}\varphi_k$.

Наиболее часто среднеквадратичное приближение применяется в том случае, когда система $\varphi_0, \varphi_1, \dots, \varphi_k$ ортонормирована, т.е.

$$(\varphi_k, \varphi_l) = \begin{cases} 0, & \text{если } k \neq l, \\ 1, & \text{если } k = l. \end{cases}$$

Тогда система (11) решается в явном виде

$$\overline{c_k} = (f, \varphi_k), \quad k = 0, 1, \dots, n. \tag{12}$$

А погрешность приближения определяется формулой:

$$\left\|f-\overline{\varphi}\right\|^2 = \|f\|^2 - \sum_{k=0}^{n} \overline{c_k}^2.$$

Определение 1. *Числа $\overline{c_k}$, определенные (12), называются коэффициентами Фурье элемента $f \in H$ по ортонормированной системе φ_0, $\varphi_1, \ldots, \varphi_n$, а обобщенный многочлен $\overline{\varphi} = \sum_{k=0}^{n} \overline{c_k} \varphi_k$ называется многочленом Фурье.*

§23. Эмпирические формулы. Метод наименьших квадратов

Нередко при обработке результатов наблюдений встречаются со следующей задачей: в итоге опыта получен ряд значений x и y, однако характер функциональной зависимости между ними остается неизвестным. Требуется по полученным данным найти аналитическое выражение зависимости между x и y. Формулы, полученные в результате решения такого рода задач, называются эмпирическими.

Использование интерполяционных многочленов для этой цели не всегда целесообразно, так как совпадение значений полученной функции с табличными значениями в узлах интерполяции не гарантирует достаточно малого различия указанных значений в других точках, отличных от узлов.

Задача о построении эмпирической формулы состоит в следующем. Пусть результаты измерений представлены таблицей:

x	x_1	x_2	$\dots$	x_k	x_{k+1}	$\dots$	x_n
y	y_1	y_2	$\dots$	y_k	y_{k+1}	$\dots$	y_n

Пусть $y = \varphi(x; a_1, a_2, \dots, a_m)$ – искомая эмпирическая формула, где функция $y = \varphi(x; a_1, a_2, \dots, a_m)$ зависит от некоторых параметров. Разности $\varphi(x_k, a_1, a_2, \dots, a_m) - y_k = \varepsilon_k$, $k = 1, 2, \dots, n$, где y_k – числа из второй строки таблицы, $\varphi(x_k, a_1, a_2, \dots, a_m)$ – значения функции $y = \varphi(x; a_1, a_2, \dots, a_m)$ при соответствующих значениях x_k из верхней строки таблицы, называются уклонениями или погрешностями. Требуется так подобрать параметры $a_1, a_2, \ldots, a_m$ функции $y = \varphi(x; a_1, a_2, \ldots, a_m)$, чтобы уклонения ε_k оказались наименьшими (в каком-то смысле).

Для определения параметров эмпирической формулы $y = \varphi(x; a_1, a_2, \ldots, a_m)$ используются методы:

1) метод выбранных точек;
2) метод средних;
3) метод наименьших квадратов.

23.1 Метод наименьших квадратов

Пусть известен вид эмпирической формулы

$$y = \varphi(x; a_1, a_2, \ldots, a_m) \tag{1}$$

и

$$\varepsilon_k = \varphi(x_i; a_1, a_2, \ldots, a_m) - y_i, \quad i = \overline{1, n} \tag{2}$$

– уклонения эмпирической формулы (1) от исходных данных (x_i, y_i). По методу наименьших квадратов наилучшими коэффициентами $a_1, a_2, \ldots, a_m$ считаются те, для которых сумма квадратов уклонений $\varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2$ минимальна, т.е.

$$S(a_1, a_2, \ldots, a_m) = \sum_{i=1}^{n} \left[\varphi(x_i; a_1, a_2, \ldots, a_m) - y_i\right]^2 \tag{3}$$

минимальна. Отсюда, используя необходимые условия экстремума функции нескольких переменных, получаем так называемую нормальную систему для определения коэффициентов a_i, $i = \overline{1, m}$

$$\frac{\partial S}{\partial a_1} = 0, \ \frac{\partial S}{\partial a_2} = 0, \ldots, \ \frac{\partial S}{\partial a_m} = 0. \tag{4}$$

система (4) имеет единственное решение, то оно будет искомым.

Метод наименьших квадратов обладает тем преимуществом, что если сумма S квадратов уклонений мала, то сами эти уклонения также малы по абсолютной величине. Для метода средних такого вывода сделать нельзя. Недостатком метода наименьших квадратов является громоздкость вычислений.

23.2 Определение параметров эмпирических формул по методу наименьших квадратов в случае квадратичной зависимости

Дана таблица

x	x_1	x_2	$\ldots$	x_n
y	y_1	y_2	$\ldots$	y_n

Рассмотрим пары (x_i, y_i) как прямоугольные координаты на плоскости. Предположим, что точки $M_i(x_i, y_i)$, $i = \overline{1, n}$ почти лежат на параболе.

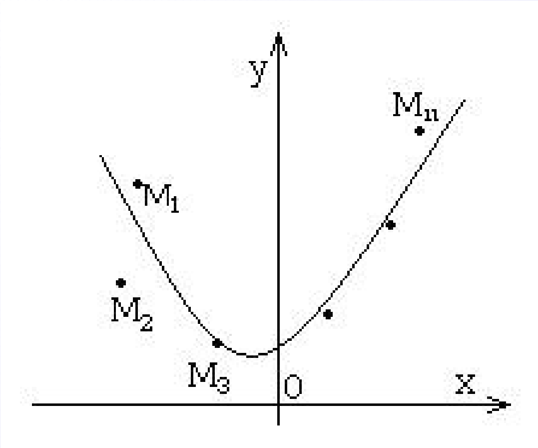

Рисунок 1.

В этом случае естественно предположить, что между x и y существует квадратичная зависимость $z_i = ax_i^2 + bx_i + c,\ i = \overline{1,n}$.

Обозначим $\varepsilon_i = z_i - y_i,\ i = \overline{1,n}$.

Параметры a, b, c выберем так, чтобы сумма квадратов уклонений

$$S = \varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 = (z_1 - y_1)^2 + (z_2 - y_2)^2 + \ldots + (z_n - y_n)^2 =$$

$$= \left(ax_1^2 + bx_1 + c - y_1\right)^2 + \ldots + \left(ax_n^2 + bx_n + c - y_n\right)^2$$

была наименьшей. Для этого необходимо, чтобы $\frac{\partial S}{\partial a} = 0$, $\frac{\partial S}{\partial b} = 0$, $\frac{\partial S}{\partial c} = 0$. Находя выражения для частных производных функции S по переменным a, b, c, получим так называемую нормальную систему уравнений.

$$\begin{cases} a\sum\limits_{i=1}^{n} x_i^4 + b\sum\limits_{i=1}^{n} x_i^3 + c\sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} y_i x_i^2, \\ a\sum\limits_{i=1}^{n} x_i^3 + b\sum\limits_{i=1}^{n} x_i^2 + c\sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} x_i y_i, \\ a\sum\limits_{i=1}^{n} x_i^2 + b\sum\limits_{i=1}^{n} x_i + nc = \sum\limits_{i=1}^{n} y_i. \end{cases}$$

Из этой системы определяются параметры a, b, c эмпирической формулы. Систему можно решить, например, методом Гаусса.

§24. Наилучшие равномерные приближения

Если норма в линейном нормированном пространстве определяется не через скалярное произведение, то нахождение элемента наилучшего приближения существенно усложняется. Рассмотрим типичную задачу, встречающуюся, в частности, при составлении стандартных программ вычисления функций.

Пусть L-пространство ограниченных вещественных функций, определенных на отрезке $[a,b]$ вещественной оси, с нормой $\|f\| = \sup\limits_{[a,b]}\bigl|f(x)\bigr|$.

Ищем наилучшие приближения в виде $Q_n(x) = \sum\limits_{j=0}^{n} a_j x^j$.

В §21. говорится о том, что такой элемент приближения существует (см. [1, теорема 1, с. 210]), т.е. существует многочлен $Q_n^0(x)$ такой, что $E_n(f) = \|f - Q_n^0\| \le \|f - Q_n\|$ при любом $Q_n(x)$ степени n. Такой многочлен $Q_n^0(x)$ называется многочленом наилучшего равномерного приближения. Справедлива

Теорема Чебышева. *Чтобы многочлен $Q_n(x)$ был многочленом наилучшего приближения непрерывной функции $f(x)$, необходимо и достаточно существования на $[a,b]$ по крайней мере $m = n+2$ точек $x_0 < \ldots < x_{n+1}$ таких, что $f(x_i) - Q_n(x_i) = \alpha(-1)^i\|f - Q_n\|$, $i = 0,1,\ldots,n+1$, $\alpha = 1$ (или $\alpha = -1$) одновременно для всех i.*

Определение 1. *Точки $x_0,\ldots,x_{n+1}$, удовлетворяющие условиям* *теоремы, принято называть точками чебышевского альтернанса.*

Эта теорема дает необходимые и достаточные условия того, чтобы многочлен являлся многочленом наилучшего приближения для непрерывной функции.

Теорема единственности. *Многочлен наилучшего равномерного приближения непрерывной функции единственен* [1, с. 227].

24.1 Примеры наилучшего равномерного приближения

1^0. Непрерывная на $[a,b]$ функция $f(x)$ приближается многочленом нулевой степени. Пусть $\sup\limits_{[a,b]} f(x) = f(x_1) = M$, $\inf\limits_{[a,b]} f(x) = f(x_0) = m$. Многочлен $Q_0(x) = \dfrac{M+m}{2}$ является многочленом наилучшего приближения, а точки x_1, x_0 – точками чебышевского альтернанса.

2^0. Выпуклая на $[a,b]$ функция $f(x)$ приближается многочленом первой степени $Q_1(x) = a_0 + a_1 x$. Вследствие выпуклости $f(x)$ разность $f(x) - (a_0 + a_1 x)$ может иметь только одну внутреннюю точку экстремума. Поэтому точки a, b являются точками чебышевского альтернанса. Пусть d – третья точка альтернанса. Согласно теореме Чебышева, имеем равенства.

$$f(a) - (a_0 + a_1 a) = \alpha L,$$

$$f(d) - (a_0 + a_1 d) = -\alpha L,$$

Начало
Содержание
Страница 92 из 403
Назад
На весь экран
Закрыть

$$f(b) - (a_0 + a_1 b) = \alpha L,$$

где $L = \|f - Q_1^0\|$. Вычитая первое уравнение из третьего, получим $f(b) - f(a) = a_1(b-a)$. Отсюда находим $a_1 = \dfrac{f(b)-f(a)}{b-a}$. Для определения неизвестных d, L, a_0, a_1 и $\alpha = \pm 1$ получено всего три уравнения. Однако следует вспомнить, что точка d является точкой экстремума разности $f(x) - (a_0 + a_1 x)$. Находим значение d. Если $f(x)$ – дифференцируема, то уравнение для определения d будет $f'(d) - a_1 = 0$. Теперь определим a_0 из уравнения получающегося сложением первого и второго уравнений.

ГЛАВА 3

ЧИСЛЕННОЕ ИНТЕГРИРОВАНИЕ

§25. Приближенное вычисление определенных интегралов. Квадратурные формулы Ньютона-Котеса

Точное вычисление определенных интегралов по формуле Ньютона-Лейбница

$$\int\limits_a^b f(x)\,dx = F(x)\Big|_a^b = F(b) - F(a), \text{ где } F'(x) = f(x)$$

не всегда представляется возможным (так как первообразная подынтегральной функции иногда не выражается в элементарных функциях – неберущийся интеграл) или нецелесообразным (поскольку нахождение первообразной часто связано с громоздкими преобразованиями). В этих случаях, а также в случае, когда подынтегральная функция задана таблично способом, целесообразно определенные интегралы вычислять приближенно. Существуют различные методы численного интегрирования.

25.1 Формулы Ньютона-Котеса

Пусть требуется вычислить $\int\limits_a^b f(x)\,dx$. Заменим подынтегральную функцию ее интерполяционным полиномом $P_n(x)$, $f(x) = P_n(x)+r_n(x)$, тогда

$$\int\limits_a^b f(x)\,dx = \int\limits_a^b P_n(x)\,dx + \int\limits_a^b r_n(x)\,dx =$$

$$= \int\limits_a^b P_n(x)\,dx + R_n = \int\limits_a^b \sum_{i=0}^{n} \frac{\omega(x)f(x_i)}{(x-x_i)\omega'(x_i)}\,dx + R_n =$$

$$= \sum_{i=0}^{n} f(x_i) \int\limits_a^b \frac{\omega(x)}{(x-x_i)\omega'(x_i)}\,dx + R_n = \sum_{i=0}^{n} A_i f(x_i) + R_n,$$

где $R_n = \int\limits_a^b r_n(x)\,dx$, $A_i = \int\limits_a^b \frac{\omega(x)dx}{(x-x_i)\omega'(x_i)}$ и $\omega(x) = \prod\limits_{i=0}^{n}(x-x_i)$.
A_i – коэффициенты, зависящие от узлов интерполяции x_i. В формуле

$$\int\limits_a^b f(x)\,dx \approx \sum_{i=0}^{n} A_i f(x_i) \tag{1}$$

правая часть называется квадратурной суммой, A_i – квадратурные коэффициенты, x_i – квадратурные узлы.

$$\int\limits_a^b f(x)\,dx = \sum_{i=0}^{n} A_i f(x_i) + R_n \qquad (2)$$

– формула механических квадратур.

Введем новую переменную $t = \dfrac{x - x_0}{h}$, $\quad dx = hdt$,

$$x - x_i = x - (x_0 + ih) = (x - x_0) - ih = th - ih = (t - i)h,$$

$$\omega(x) = (x - x_0)(x - x_1)\cdots(x - x_n) = th(t-1)h\cdots(t-n)h =$$

$$= h^{n+1}t(t-1)(t-2)\cdots(t-n) = h^{n+1}\prod(t),$$

где $\prod(t) = \prod\limits_{i=0}^{n}(t - i)$,

$$\omega'(x_i) = (x_i - x_0)(x_i - x_1)\cdots(x_i - x_{i-1})(x_i - x_{i+1})\cdots(x_i - x_n) =$$

$$= \Big(h^i i(i-1)(i-2)\cdots 1\Big)\cdot\Big(1\cdot 2\cdots(n-i)h^{n-i}(-1)^{n-i}\Big) =$$

$$= h^i i! h^{n-i}(n-i)!(-1)^{n-i} = (-1)^{n-i}h^n i!(n-i)!$$

Отсюда,

$$A_i = \int\limits_a^b \frac{\omega(x)}{(x - x_i)\omega'(x_i)}\,dx = \int\limits_0^n \frac{h^{n+1}\prod(t)}{h(t-i)(-1)^{n-i}h^n i!(n-i)!}\,hdt =$$

$$= \frac{h(-1)^{n-i}}{i!(n-i)!}\int\limits_0^n \frac{\prod(t)}{t-i}\,dt. \qquad (3)$$

Формула (3) показывает, что еще квадратурные коэффициенты A_i зависят от отрезка интегрирования, так как в (3) присутствует h.

Введем B_i:

$$B_i = \frac{(-1)^{n-i}}{ni!(n-i)!}\int\limits_0^n \frac{\prod(t)dt}{t-i} \qquad (4)$$

и установим связь между коэффициентами (3) и (4), $h = \dfrac{b-a}{n}$.

$$A_i = (b-a)B_i. \qquad (5)$$

Коэффициенты B_i не зависят от отрезка интегрирования, лишь от числа точек разбиения n.

$$\int\limits_a^b f(x)\,dx = (b-a)\sum_{i=0}^{n} B_i f(x_i) + R_n. \qquad (6)$$

Определение 1. *Коэффициенты B_i носят название коэффициентов Ньютона-Котеса, а квадратурная формула (6) называется квадратурной формулой Ньютона-Котеса.*

§26. Формула трапеций

Найдем явный вид коэффициентов B_i. При $n = 1$ имеем 2 узла $x_0 = a,\ x_1 = b$.

$$B_0 = \frac{(-1)^1}{1 \cdot 0! \cdot 1!} \int\limits_0^1 \frac{t(t-1)}{t-0}\,dt = -\int\limits_0^1 (t-1)\,dt =$$

$$= -\left(\frac{t^2}{2} - t\right)\Big|_0^1 = -\left(\frac{1}{2} - 1\right) = \frac{1}{2};$$

$$B_1 = \frac{(-1)^0}{1 \cdot 1! \cdot 0!} \int\limits_0^1 \frac{t(t-1)}{t-1}\,dt = \int\limits_0^1 t\,dt = \frac{t^2}{2}\Big|_0^1 = \frac{1}{2}.$$

Отсюда и из (6) (из §25)

$$\int\limits_a^b f(x)\,dx = (b-a)\left(\frac{1}{2}f(x_0) + \frac{1}{2}f(x_1)\right) + R_n = \frac{b-a}{2}\left[f(x_0) + f(x_1)\right] + R_n \tag{1}$$

– *формула трапеций.*

Выведем остаточный член для формулы трапеций. При выводе формулы трапеций мы пользовались двумя точками $x_0 = a$ и $x_1 = b$, поэтому $n = 1$. Заменим подынтегральную функцию ее интерполяционным полиномом, остаток интерполирования в общем виде имеет вид

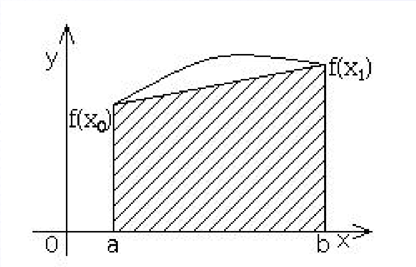

Рисунок 1.

$r_n(x) = \dfrac{\omega(x) f^{(n+1)}(\xi)}{(n+1)!}$. В нашем случае $r_1(x) = \dfrac{(x-x_0)(x-x_1)}{2!} f''(\xi)$. Тогда

$$\int\limits_a^b f(x)\,dx = \int\limits_a^b \big(P_n(x) + r_n(x)\big)\,dx = \int\limits_a^b P_n(x)\,dx + \int\limits_a^b r_n(x)\,dx.$$

Следовательно, $R_n = \int\limits_a^b r_n(x)\,dx = \{n=1\} = \int\limits_a^b \dfrac{(x-x_0)(x-x_1)}{2!} f''(\xi)\,dx$.

Если вторая производная f'' непрерывна на $[a,b]$, то, так как множитель $(x-a)(x-b)$ сохраняет знак на $[a,b]$, то существует на $[a,b]$ такая точка η, для которой $n = 1$

$$R_n(f) = f''(\eta) \int\limits_a^b \frac{(x - x_0)(x - x_1)}{2}\, dx =$$

$$= f''(\eta) \int\limits_a^b \frac{(x - a)(x - b)}{2}\, dx = -\frac{(b - a)^3}{12} f''(\eta),$$

где $a \le \eta \le b$.

Итак, остаточный член для формулы трапеции имеет вид

$$R_n(f) = -\frac{(b - a)^3}{12} f''(\eta). \qquad (2)$$

Отсюда, $\left|R_n(f)\right| \le \dfrac{(b - a)^3}{12} M_2$, где $M_2 = \max\limits_{a \le x \le b} \left|f''(x)\right|$.

Для увеличения точности формулы трапеций (1), разделим $[a, b]$ на n частей длины $h = \dfrac{b - a}{n}$ и рассмотрим $x_k = a + kh$, $(k = 0, 1, 2, \ldots, n)$ и $y_k = f(x_k)$ – значение подынтегральной функции в этих точках. Тогда формула (1) запишется в виде:

$$\int\limits_a^b f(x)\, dx \approx h \left[\frac{y_0}{2} + y_1 + y_2 + \cdots + y_{n-1} + \frac{y_n}{2}\right] \text{ или}$$

$$\int\limits_a^b f(x)\,dx \approx \frac{h}{2}\left[y_0 + 2(y_1 + y_2 + \cdots + y_{n-1}) + y_n\right]$$

– *общая квадратурная формула трапеций.* Ее остаточный член имеет вид $R_n(f) = -\dfrac{(b-a)^3}{12n^2} f''(\xi)$. Следовательно, справедлива оценка для остаточного члена

$$\left|R_n(f)\right| \le \frac{(b-a)^3}{12n^2} M_2, \text{ где } M_2 = \max_{a\le x\le b}\left|f''(x)\right|. \qquad (3)$$

Формула трапеций имеет второй порядок точности, т.е. $O(h^2)$.

§27. Формула Симпсона (парабол)

Рассмотрим $n = 2$. Тогда

$$B_0 = \frac{(-1)^{2-0}}{2\cdot 0!\cdot 2!}\int\limits_0^2 \frac{t(t-1)(t-2)}{t}\,dt = \frac{1}{4}\int\limits_0^2 \left(t^2-3t+2\right)dt =$$

$$= \frac{1}{4}\left(\frac{t^3}{3}-\frac{3t^2}{2}+2t\right)\Bigg|_0^2 = \frac{1}{4}\left(\frac{8}{3}-6+4\right) = \frac{1}{6};$$

$$B_1 = \frac{(-1)^{2-1}}{2\cdot 1!\cdot 1!}\int\limits_0^2 \frac{t(t-1)(t-2)}{t-1}\,dt = -\frac{1}{2}\int\limits_0^2 t(t-2)\,dt =$$

$$= -\frac{1}{2}\int\limits_0^2 \left(t^2-2t\right)dt = -\frac{1}{2}\left(\frac{t^3}{3}-\frac{2t^2}{2}\right)\Bigg|_0^2 = -\frac{1}{2}\left(\frac{8}{3}-4\right) = \frac{2}{3};$$

$$B_2 = \frac{(-1)^{2-2}}{2\cdot 2!\cdot 0!}\int\limits_0^2 \frac{t(t-1)(t-2)}{t-2}\,dt = \frac{1}{4}\int\limits_0^2 t(t-1)\,dt = \frac{1}{4}\int\limits_0^2 \left(t^2-t\right)dt =$$

$$= \frac{1}{4}\left(\frac{t^3}{3}-\frac{t^2}{2}\right)\Bigg|_0^2 = \frac{1}{4}\left(\frac{8}{3}-2\right) = \frac{1}{4}\cdot\frac{2}{3} = \frac{1}{6};$$

Следовательно, из (6):

$$\int\limits_a^b f(x)\,dx = \frac{b-a}{6}\big[f(x_0) + 4f(x_1) + f(x_2)\big] + R_n \qquad (1)$$

– *формула парабол* или *формула Симпсона.*

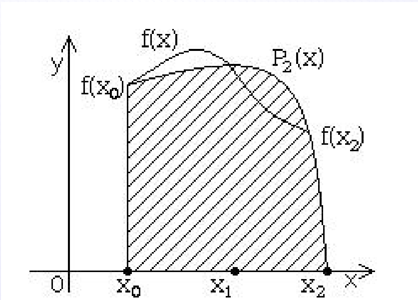

Рисунок 1.

Дуга графика функции $f(x)$ заменяется дугой параболы. Здесь $P_n(x) = P_2(x)$ – многочлен второй степени, поэтому график $P_2(x)$ – парабола.

Остаточный член формулы парабол (Симпсона) имеет вид

$$R(f) = -\frac{1}{90}\left(\frac{b-a}{2}\right)^5 + f^{(4)}(\xi). \qquad (2)$$

Докажем формулу (2).

Для нахождения погрешности формулы (1) рассмотрим многочлен $P_3(x)$ третьей степени, удовлетворяющий условиям: $\left(c = \dfrac{a+b}{2}\right)$, $P_3(a) = f(a)$, $P_3(c) = f(c)$, $P_3'(c) = f'(c)$, $P_3(b) = f(b)$. Он интерполирует $f(x)$ по значениям в двух однократных узлах a и b, и по значениям $f(c)$, $f'(c)$ в двойном узле c: $f(x) = P_3(x) + r(x)$. Для $P_3(x)$ равенство $\int\limits_a^b f(x)\,dx \approx \dfrac{b-a}{6}\left[f(a) + 4f(c) + f(b)\right]$ является точным и поэтому

$$\int\limits_a^b f(x)\,dx = \int\limits_a^b P_3(x)\,dx + \int\limits_a^b r(x)\,dx =$$

$$= \frac{b-a}{6}\left[P_3(a) + 4P_3(c) + P_3(b)\right] + \int\limits_a^b r(x)\,dx =$$

$$= \frac{b-a}{6}\left[f(a) + 4f(c) + f(b)\right] + \int\limits_a^b r_n(x)\,dx.$$

Погрешность формулы парабол имеет вид $R_n(f) = \int\limits_a^b r_n(x)\,dx$.

Если считать, что f имеет на $[a,b]$ непрерывную производную четвертого порядка, то для $r(x)$ справедливо равенство

$$r(x) = \frac{1}{4!}(x-a)(x-c)^2(x-b)f^{(4)}(\xi), \quad (a \le \xi \le b).$$

Поэтому $R(f) = \dfrac{1}{24}\int\limits_a^b (x-a)(x-c)^2(x-b)f^{(4)}(\xi)\,dx.$

Так как множитель $(x-a)(x-c)^2(x-b)$ не изменяет знак на отрезке $[a,b]$ и $f^{(4)}(x)$ есть непрерывная функция при $a \le x \le b$, то на $[a,b]$ существует точка η такая, что

$$R(f) = \frac{1}{24}f^{(4)}(\eta)\int\limits_a^b (x-a)(x-c)^2(x-b)\,dx = -\frac{1}{90}\left(\frac{b-a}{2}\right)^5 f^{(4)}(\eta) \quad (3)$$

Формула (2) доказана.

Получим теперь общую формулу Симпсона. Для этого отрезок $[a,b]$ разобьем на четное число $n = 2m$ равных отрезков длины $h = \dfrac{b-a}{2m}$. Получим точки $x_0 = a$, $x_1 = x_0 + h$, $x_2 = x_0 + 2h, \dots,$ $x_{2m} = b$ и пусть $y_0, y_1, \dots, y_{2m}$ – значения функции $y = f(x)$ в этих точках. Тогда *общая* *квадратурная формула Симпсона* запишется в виде:

$$\int\limits_a^b f(x)\,dx \approx \frac{h}{3}\left[y_0+4(y_1+y_3+\cdots+y_{2m-1})+2(y_2+y_4+\cdots+y_{2m-2})+y_{2m}\right]. \tag{4}$$

Остаточный член формулы парабол (4) имеет вид

$$R(f) = -\frac{(b-a)^5}{180n^4} f^{(4)}(\xi), \text{ где } a \le \xi \le b. \tag{5}$$

Справедлива следующая оценка для остаточного члена

$$\left|R_n(f)\right| = \frac{(b-a)^5}{180n^4} M_4, \text{ где } M_4 = \max_{[a,b]}\left|f^{(4)}(x)\right|. \tag{6}$$

Формула Симпсона существенно точнее, чем формула трапеций, на $[a, b]$ она имеет точность $O\left(h^4\right)$.

Сравнивая между собой формулы (3) (из §26) и (6) для остаточных членов метода трапеций и метода парабол заключаем, что с увеличением n остаточный член формулы трапеций уменьшается пропорционально величине $\frac{1}{n^2}$, а формулы парабол – $\frac{1}{n^4}$, т.е. заданная точность для гладких функций достигается быстрее по методу парабол, хотя с точки зрения техники вычислений оба метода одинаковы.

§28. Формулы прямоугольников

Метод прямоугольников основан на определении определенного интеграла и его геометрическом значении. Формула прямоугольников имеет вид

$$\int\limits_a^b f(x)\,dx \approx h\sum_{k=0}^{n-1} y_k = h[y_0 + y_1 + \cdots + y_{n-1}] \tag{1}$$

$$\int\limits_a^b f(x)\,dx = h\sum_{k=1}^{n} y_k = h[y_1 + y_2 + \cdots + y_n], \tag{2}$$

где $h = \dfrac{b-a}{n}$, $y_k = f(x_k)$, $x_k = a + kh$.

Формула (1) называется *формулой левых прямоугольников*, а формула (2) – *формулой правых прямоугольников*.

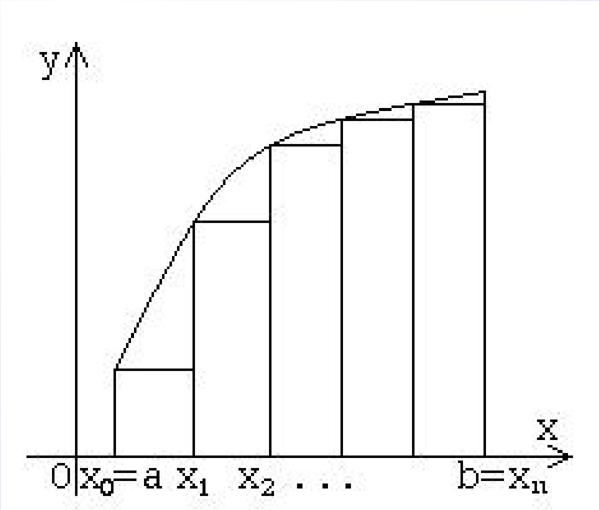

Рисунок 1.

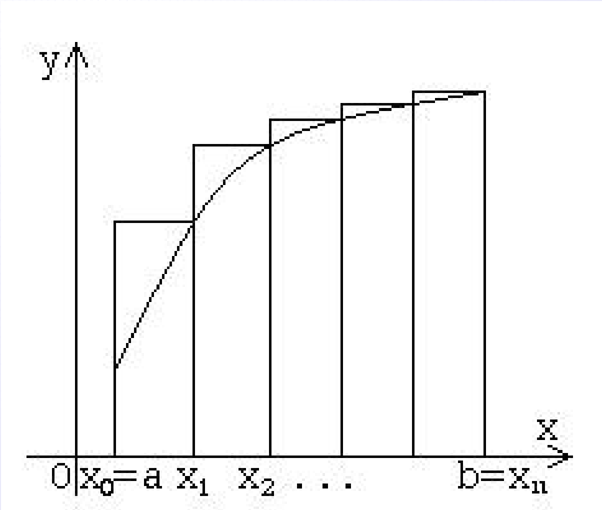

Рисунок 2.

Правая часть формулы (1) выражает площадь ступенчатой фигуры, состоящей из прямоугольников, основание каждого из которых равно h, а высота равна левой ординате.

Правая часть формулы (2) выражает площадь ступенчатой фигуры, состоящей из прямоугольников, основание каждого из которых равно h, а высота равна правой ординате. Абсолютная погрешность метода прямоугольников определяется формулой

$$\left|R_n(f)\right| \le \frac{(b-a)^2}{2n} M_1, \text{ где } M_1 = \max_{[a,b]} \left|f'(x)\right|.$$

Из-за нарушения симметрии погрешность этих формул является величиной $O(h)$.

Формула средних прямоугольников имеет вид:

$$\int_a^b f(x)\,dx = \frac{b-a}{n}\left[f\left(a+\frac{h}{2}\right) + f\left(a+\frac{3h}{2}\right) + \right.$$

$$\left. \cdots + f\left(a + \left(n-\frac{1}{2}\right)h\right)\right] + R_n,$$

где остаточный член $R_n = \dfrac{(b-a)^3}{24n^2} f''(\xi)$, $a \le \xi \le b$.

Справедлива оценка: $\left|R_n(f)\right| \le \dfrac{(b-a)^3}{24n^2} M_2$, где $M_2 = \max\limits_{[a,b]}\left|f''(x)\right|$, т.е. погрешность формулы средних прямоугольников на $[a,b]$ есть величина $O(h^2)$, как и у формулы трапеций.

§29. Формула «трех восьмых»

Она получается из формулы Ньютона-Котеса при $n = 3$, интерполирование выполняется для функции $f(x)$ в точках a, $a + \dfrac{H}{3}$, $a + \dfrac{2H}{3}$, b, где $H = b - a$.

Формула будет следующей:

$$\int\limits_a^b f(x)\,dx \approx H\left[\frac{1}{8}f(a) + \frac{3}{8}f\left(a + \frac{H}{3}\right) + \frac{3}{8}f\left(a + \frac{2H}{3}\right) + \frac{1}{8}f(b)\right]. \quad (1)$$

Ее часто называют *формулой «трех восьмых»*. Можно показать, что погрешность формулы (1) может быть, если $f^{(4)}$ есть непрерывная и мало меняющаяся на $[a, b]$ функция, представлена в виде

$$R_n \approx -\frac{(b-a)^5}{6480} f^{(4)}(\eta), \quad a \le \eta \le b. \quad (2)$$

Применим формулу трех восьмых к построению формул приближенного интегрирования при большем числе равноотстоящих узлов.

Пусть n есть число, кратное трем. Разделим $[a, b]$ на n равных частей длины $h = \dfrac{b-a}{n}$. Возьмем строенный отрезок $[a + kh, a + (k + 3)h]$ и к интегрированию по нем применим правило (1):

$$\int\limits_{a+kh}^{a+(k+3)h} f(x)\,dx = \frac{3h}{8}\{f(a+kh) + 3f[a+(k+1)h] + 3f[a+(k+2)h]+$$

$$+f[a+(k+3)h]\} + R_{k+1}(f), \quad R_{k+1}(f) \approx -\frac{(3h)^5}{6480} f^{(4)}(\eta_{k+1}).$$

Если такие равенства записать для всех строенных отрезков $[a, a+3h]$, $[a+3h, a+6h], \ldots$ и сложить их почленно, построим общую формулу трех восьмых:

$$\int\limits_a^b f(x)\,dx = \frac{3h}{8}\big[(f_0 + f_n) + 2(f_3 + f_6 + \cdots + f_{n-3})+$$

$$+ 3(f_1 + f_2 + f_4 + f_5 + \cdots + f_{n-2} + f_{n-1})\big] + R(f), \qquad (3)$$

где $f_k = f(a+kh)$.

Погрешность $R(f)$ имеет следующее значение

$$R(f) = R_1(f) + R_2(f) + \ldots + R_{n/3}(f) \approx -\frac{(3h)^5}{6480}\left[f^{(4)}(\eta_1) + \ldots + f^{(4)}(\eta_{n/3})\right] =$$

$$= -\frac{(b-a)^5}{80n^4} \cdot \frac{3}{n}\left[f^{(4)}(\eta_1) + \ldots + f^{(4)}(\eta_{n/3})\right].$$

Множитель $\dfrac{3}{n}\left[f^{(4)}(\eta_1)+\ldots+f^{(4)}(\eta_{n/3})\right]$ есть среднее арифметическое, состоящее из $\dfrac{n}{3}$ значений четвертой производной, и так как $f^{(4)}$ считается непрерывной функцией, то на $[a,b]$ существует точка ξ такая, что этот множитель равен $f^{(4)}(\xi)$. Для погрешности $R(f)$ окончательно получается выражение: $R(f)=-\dfrac{(b-a)^5}{80n^4}f^{(4)}(\xi),\ a\le\xi\le b.$

Погрешность формулы имеет такой же порядок, как и для формулы парабол, т.е. $O(h^4)$ на $[a,b]$.

§30. Выбор шага интегрирования. Правило Рунге

Для выбора шага интегрирования можно воспользоваться выражением остаточного члена. Введем, например, остаточный член формулы Симпсона:

$$R_2(h) = -\frac{(b-a)^5}{180n^4} f^{(4)}(\xi) = -\frac{h^4(b-a)}{180} f^{(4)}(\xi).$$

Если $\max\limits_{[a,b]} \left| f^{(4)}(x) \right| = M_4$, то $|R_2(h)| \leq \dfrac{h^4(b-a)M_4}{180}$. По заданной точности ε метода интегрирования из последнего неравенства определяем подходящий шаг, $\dfrac{h^4(b-a)M_4}{180} \leq \varepsilon$, $h \leq \sqrt[4]{\dfrac{180\varepsilon}{(b-a)M_4}}$.

Однако такой способ требует оценки $\max\limits_{[a,b]} \left| f^{(4)}(x) \right|$ (что на практике не всегда возможно). Поэтому пользуются другими приемами определения оценки точности, которые по ходу вычисления позволяют выбрать нужный шаг.

Разберем один из таких приемов на примере метода Симпсона. Пусть $\int\limits_a^b f(x)\,dx = I_h - \dfrac{h^4(b-a)}{180} f^{(4)}(\eta)$, где I_h – приближенное значение интеграла, вычисленное с шагом h. Уменьшим шаг h в два раза, разбив отрезок $[a, b]$ на две равные части $[a, c]$ и $[c, b]$, $c = \dfrac{a+b}{2}$. Тогда

$$\int\limits_a^b f(x)\,dx = I_{h/2} - \left(\frac{h}{2}\right)^4 \frac{(c-a)}{180} f^{(4)}(\eta_1) - \left(\frac{h}{2}\right)^4 \frac{(b-c)}{180} f^{(4)}(\eta_2),$$

$a \le \eta_1 \le c$, $c \le \eta_2 \le b$.

Предположим теперь, что $f^{(4)}$ меняется не слишком быстро, так что $f^{(4)}$ почти постоянна: $f^{(4)}(\eta_1) = f^{(4)}(\eta_2) = f^{(4)}(\eta) = K$. Тогда $\int\limits_a^b f(x)\,dx = = I_h - \dfrac{h^4(b-a)}{180}K$ и $\int\limits_a^b f(x)\,dx = I_{h/2} - \dfrac{h^4(b-a)}{2^4 \cdot 180}K$, откуда $I_h + R_2(h) = = I_{h/2} + \dfrac{R_2(h)}{2^4}$, т.е. $R_2(h) = \dfrac{I_{h/2} - I_h}{1 - \dfrac{1}{2^4}}$. Отсюда можно сделать такой вывод: если $|R_2(h)| = \dfrac{2^4}{15}\left|I_{h/2} - I_h\right| \le \varepsilon$, т.е. $\left|I_{h/2} - I_h\right| \le \varepsilon_1$, где $\varepsilon_1 = \dfrac{15\varepsilon}{2^4}$, а ε – требуемая точность, то шаг h подходит для вычисления интеграла с достаточной точностью. Если же $\left|I_{h/2} - I_h\right| > \varepsilon$, то расчет повторяют с шагом $h/4$ и затем сравнивают $I_{h/2}$ и $I_{h/4}$, и т.д. Это правило называется *правилом Рунге.*

§31. Квадратурные формулы интерполяционного типа

Пусть $[a, b]$ есть любой конечный или бесконечный интервал числовой оси, и рассматривается интеграл $\int\limits_a^b f(x)\,dx$. Если $[a, b]$ – конечный и интегрируемая функция $F(x)$ имеет высокую гладкость, то можно рассчитывать хорошо приблизить ее многочленом невысокой степени или несложной рациональной функцией. Если же сама функция $F(x)$ или ее производные невысоких порядков имеют особенности или даже обращаются в бесконечность, то это затруднит приближение F или сделает его вообще невозможным. В этом случае мы должны заранее освободиться от таких особенностей путем их выделения. Делается это при помощи разложения F на два сомножителя $F(x) = p(x)f(x)$, где $p(x)$ имеет такие же особенности, как и $F(x)$, а $f(x)$ есть достаточно гладкая функция, и интеграл рассматривается в виде

$$\int\limits_a^b p(x)f(x)\,dx. \tag{1}$$

Такое представление важно также, когда вычисляется несобственный интеграл по бесконечным отрезкам, например, $\int\limits_a^\infty F(x)\,dx$.

Определение 1. *Функция $p(x)$ в (1) называется весовой функцией или весом.*

При построении вычислительного правила для (1) она считается фиксированной. Обычно считают, весовая функция $p(x) > 0$ на $[a, b]$, для которой $\int\limits_a^b p(x)\,dx > 0$.

Если получить квадратурные формулы путем замены подынтегральной функции $f(x)$ на интерполяционный многочлен $P_n(x)$ сразу на всем отрезке $[a, b]$ (не разбивая $[a, b]$ на частичные отрезки), то такие квадратурные формулы называются квадратурными формулами интерполяционного типа. Формулы прямоугольников, трапеций, Симпсона – частные случаи квадратурных формул интерполяционного типа, когда $n = 0, 1, 2$ и $p(x) \equiv 1$.

Рассматриваемые далее квадратурные формулы имеют вид

$$\int\limits_a^b p(x) f(x)\,dx \approx \sum_{k=0}^{n} c_k f(x_k), \tag{2}$$

где $x_k \in [a, b]$ и c_k – числа, $k = 0, 1, \ldots, n$. Получим выражения для коэффициентов c_k квадратурных формул интерполяционного вида. Заменяя в интеграле (1) функцию $f(x)$ интерполяционным полиномом Лагранжа

$$P_n(x) = \sum_{k=0}^{n} \frac{\omega(x)}{(x - x_k)\omega'(x_k)} f(x_k), \tag{3}$$

где $\omega(x) = \prod\limits_{i=0}^{n}(x - x_i)$, $\omega'(x_k) = \prod\limits_{i \neq k}(x_k - x_i)$, получим приближенную формулу (2),

$$\int\limits_a^b p(x)f(x)\,dx \approx \int\limits_a^b p(x)\sum_{k=0}^{n}\frac{\omega(x)}{(x-x_k)\omega'(x_k)}f(x_k)\,dx =$$

$$=\sum_{k=0}^{n} f(x_k)\int\limits_a^b \frac{p(x)\omega(x)}{(x-x_k)\omega'(x_k)}\,dx = \sum_{k=0}^{n} f(x_k)c_k,$$

где

$$c_k = \int\limits_a^b \frac{p(x)\omega(x)}{(x-x_k)\omega'(x_k)}\,dx, \quad k=\overline{0,n}. \tag{4}$$

Таким образом, формула (2) является *квадратурной формулой интерполяционного типа* тогда и только тогда, когда ее коэффициенты вычисляются по правилу (4).

31.1 Оценка погрешности квадратурной формулы интерполяционного типа

Пусть $f(x) = P_n(x) + r_{n+1}(x)$, где $r_{n+1}(x)$ – остаточный член интерполирования. Тогда

$$\int\limits_a^b p(x)f(x)\,dx = \int\limits_a^b p(x)P_n(x)\,dx + \int\limits_a^b r_{n+1}(x)p(x)\,dx =$$

$$= \sum_{k=0}^{n} c_k f(x_k) + \int\limits_a^b r_{n+1}(x)p(x)\,dx.$$

Таким образом, погрешность R_{n+1} квадратичной формулы (2), (4) равна $R_{n+1} = \int_a^b p(x)r_{n+1}(x)dx$. Вспоминая выражение для погрешности интерполирования, получаем $r_{n+1} = \dfrac{f^{(n+1)}\big(\xi(x)\big)\omega(x)}{(n+1)!}$, тогда

$$R_{n+1} = \frac{1}{(n+1)!}\int_a^b p(x)\omega(x)f^{(n+1)}\big(\xi(x)\big)dx.$$

Откуда приходим к следующей оценке погрешности квадратурной формулы интерполяционного типа

$$|R_{n+1}| \le \frac{M_{n+1}}{(n+1)!}\int\limits_a^b p(x)|\omega(x)|\,dx, \text{ где } M_{n+1} = \max_{[a,b]}\left|f^{(n+1)}(x)\right|. \qquad (5)$$

Из формулы (5) следует

Теорема 1. *Квадратурная формула интерполяционного типа, построенная по $n+1$ узлу $x_0, x_1, \ldots, x_n$, является точной для любого многочлена $f(x)$ степени n (здесь $f(x)$ – многочлен степени n и c_k – коэффициенты, вычисленные по формуле (4)), т.е. имеет место точное равенство*

$$\int\limits_a^b p(x)f(x)\,dx = \sum_{k=0}^{n} c_k f(x_k). \tag{6}$$

Справедливо и обратное утверждение.

Теорема 2. *Если квадратурная формула* $\int\limits_a^b p(x)f(x)\,dx \approx \sum\limits_{k=0}^{n} d_k f(x_k)$ *точна для любого многочлена степени n, то она является квадратурной формулой интерполяционного типа.*

§32. Квадратурные формулы наивысшей алгебраической степени точности (НАСТ)

В предыдущих параграфах предполагалось, что узлы квадратурных формул заданы заранее. Причем, если использовать узлов интерполирования, то получались квадратурные формулы, точные для алгебраических многочленов степени $n-1$. Оказывается, что за счет выбора узлов можно получить квадратурные формулы, которые будут точными и для многочленов степени выше $n-1$.

Рассмотрим следующую задачу: построить квадратурную формулу

$$\int_a^b p(x)f(x)\,dx \approx \sum_{k=1}^{n} c_k f(x_k), \tag{1}$$

которая при заданном n была бы точна для алгебраического многочлена возможно большей степени. Здесь для удобства изложения нумерация узлов начинается с $k=1$.

Формула (1) имеет $2n$ параметров c_k и x_k, и можно ожидать, что при помощи выбора их можно сделать равенство (1) точным для всяких алгебраических многочленов степени $2n-1$ или, что то же самое, чтобы оно было точным для степеней x от нулевой до $2n-1$.

$$\int\limits_a^b p(x)x^m\,dx = \sum_{k=1}^{n} c_k x_k^m, \quad m=\overline{0,2n-1}.$$

Будет показано, что такие квадратурные формулы существуют. Они называются *квадратурными формулами наивысшей алгебраической точности* или *формулами Гаусса*. Эти формулы точны для любого алгебраического многочлена степени $2n-1$.

Итак, потребуем, чтобы квадратурная формула (1) была точна для любого алгебраического многочлена степени m. Это эквивалентно требованию, чтобы формула была точна для функций $f(x) = x^\alpha$, $\alpha = 0,1,\ldots,m$. Отсюда получаем условия

$$\sum_{k=1}^{n} c_k x_k^\alpha = \int\limits_a^b p(x)x^\alpha\,dx, \quad \alpha = 0,1,\ldots,m, \tag{2}$$

которые представляют собой нелинейную систему $m+1$ уравнений относительно $2n$ неизвестных: $c_1, c_2, \ldots, c_n;\ x_1, x_2, \ldots, x_n$.

Для того, чтобы число уравнений равнялось числу неизвестных, надо потребовать $m+1=2n$, т.е. $m=2n-1$. В дальнейшем будет показано, что система (2) при $m=2n-1$ имеет единственное решение. Однако сначала рассмотрим несколько частных случаев, когда решение системы (2) можно найти непосредственно.

Пусть $p(x) = 1$, $a = -1$, $b = 1$. При $n = 1$ получаем $m = 1$ и система (2) примет вид при $\alpha = 0$ $c_1 = \int\limits_{-1}^{1} dx = 2$, при $\alpha = 1$ $c_1 x_1 = \int\limits_{-1}^{1} x dx = 0$, $\Rightarrow x_1 = 0$, т.е. приходим к известной формуле прямоугольников $\int\limits_{-1}^{1} f(x)\,dx \approx 2f(0)$, которая точна для любого многочлена 1-ой степени.

При $n = 2$, $m = 3$ система (2) записывается в виде

$$c_1 + c_2 = 2; \quad c_1 x_1 + c_2 x_2 = 0;$$

$$c_1 x_1^2 + c_2 x_2^2 = \frac{2}{3}; \quad c_1 x^3 + c_2 x^3 = 0.$$

Отсюда находим $c_1 = c_2 = 1$, $x_1 = -\dfrac{1}{\sqrt{3}}$, $x_2 = \dfrac{1}{\sqrt{3}}$, т.е. получаем квадратурную формулу $\int\limits_{-1}^{1} f(x)\,dx \approx f\left(-\dfrac{1}{\sqrt{3}}\right) + f\left(\dfrac{1}{\sqrt{3}}\right)$, которая точна для любого алгебраического многочлена третьей степени.

§33. Основная теорема

Возвращаемся к рассмотрению квадратурных формул общего вида

$$\int_a^b p(x)f(x)\,dx \approx \sum_{k=1}^{n} c_k f(x_k). \tag{1}$$

Введем многочлен

$$\varpi(x) = (x - x_1)\dots(x - x_n). \tag{2}$$

Полагаем $p(x) > 0$.

Теорема 1. *Квадратурная формула (1) точна для любого многочлена степени $m = 2n-1$ тогда и только тогда, когда выполнены два условия:*

1. *многочлен $\varpi(x)$ ортогонален с весом $p(x)$ любому многочлену $q(x)$ степени меньше n, т.е.*

$$\int_a^b p(x)\varpi(x)q(x)\,dx = 0; \tag{3}$$

2. *формула (1) является квадратурной формулой интерполяционного типа, т.е.*

$$c_k = \int\limits_a^b p(x)\frac{\varpi(x)}{(x-x_k)\varpi'(x_k)}\,dx, \quad k=\overline{1,n}. \qquad (4)$$

Доказательство.

Необходимость. Пусть формула (1) точна для любого многочлена степени $m = = 2n-1$. Это означает, что она точна и для многочлена $\varpi(x)q(x)$, имеющего степень $2n-1$, т.е. $\int\limits_a^b p(x)\varpi(x)q(x)\,dx = = \sum\limits_{k=1}^{n} c_k\varpi(x_k)q(x_k) = 0$, так как $\varpi(x_k) = 0$. Требование (4) выполняется в силу теоремы 2 (из §31) (т.е. если квадратурная формула (1) точна для любого многочлена степени $n-1$, то она является формулой интерполяционного типа). Необходимость доказана.

Достаточность. Пусть $f(x)$ – любой многочлен степени $2n-1$. Согласно теореме о делении многочленов с остатком, его можно представить в виде $f(x) = \varpi(x)q(x) + r(x)$, где $q(x)$ и $r(x)$ – многочлены, имеющие степень не выше $n-1$. Тогда

$$\int\limits_a^b p(x)f(x)\,dx = \int\limits_a^b p(x)\varpi(x)q(x)\,dx + \int\limits_a^b p(x)r(x)\,dx = \int\limits_a^b p(x)r(x)\,dx.$$

Последнее равенство справедливо в силу предположения (3). Далее, поскольку $r(x)$ – многочлен степени не выше $n-1$ и формула (1) является формулой интерполяционного типа, то она точна для $r(x)$, т.е.

$$\int\limits_a^b p(x)r(x)\,dx = \sum_{k=1}^{n} c_k r(x_k) =$$

$$= \sum_{k=1}^{n} c_k\big(f(x_k) - \varpi(x_k)q(x_k)\big) = \sum_{k=1}^{n} c_k f(x_k),$$

так как $\varpi(x_k) = 0$.

Таким образом, $\int\limits_a^b p(x)f(x)\,dx = \sum\limits_{k=1}^{n} c_k f(x_k)$, т.е. формула (1) точна для любого многочлена степени $2n-1$. Теорема доказана.

Отметим, что использование приведенной теоремы существенно упрощает построение формул Гаусса. Условие (3) эквивалентно требованиям

$$\int\limits_a^b p(x)\varpi(x)x^{\alpha}\,dx = 0, \quad \alpha = 0, 1, \ldots, n-1, \tag{5}$$

которые представляют собой систему n уравнений относительно n неизвестных x_1, $x_2, \ldots, x_n$. Таким образом, для построения формул Гаусса достаточно найти узлы $x_1, x_2, \ldots, x_n$ из условий ортогональности (5), а затем вычислить коэффициенты c_k согласно (4).

Теорема не гарантирует существование и единственности решения системы (5). Надо доказать ещё существование и единственность многочлена $\varpi(x)$ степени n, ортогонального всем многочленам степени меньшей n, а также убедиться в том, что все корни такого многочлена расположены на отрезке $[a, b]$.

§34. Существование и единственность квадратурных формул наивысшей алгебраической степени точности

Представим искомый многочлен (2) (из §33) в виде

$$\varpi(x) = a_0 + a_1x + a_2x^2 + \ldots + a_{n-1}x^{n-1} + x^n. \qquad (1)$$

Тогда условия ортогональности (5) (из §33) примут вид

$$\int\limits_a^b p(x)\big(a_0+a_1x+\ldots+a_{n-1}x^{n-1}+x^n\big)x^\alpha\,dx = 0, \quad \alpha = 0, 1, \ldots, n-1. \qquad (2)$$

Условия (2) представляют собой систему линейных алгебраических уравнений относительно коэффициентов $a_0, a_1, \ldots, a_{n-1}$. Покажем, что соответствующая однородная система уравнений

$$\int\limits_a^b p(x)\big(a_0 + a_1x + \ldots + a_{n-1}x^{n-1}\big)x^\alpha\,dx = 0, \quad \alpha = 0, 1, \ldots, n-1 \qquad (3)$$

имеет единственное решение $a_0 = a_1 = \ldots = a_{n-1} = 0$. Для этого умножим уравнение (3) на a_α и просуммируем по всем α. Тогда получим

$$\sum_{\alpha=0}^{n-1} a_\alpha \int\limits_a^b \left(p(x)\sum_{k=0}^{n-1} a_kx^k\right)x^\alpha\,dx = \int\limits_a^b p(x)\left[\sum_{\alpha=0}^{n-1}\sum_{k=0}^{n-1} a_\alpha a_k x^k x^\alpha\right]dx = 0,$$

то есть

$$\int\limits_a^b p(x)\left(\sum_{i=0}^{n-1} a_i x^i\right)^2 dx = 0. \qquad (4)$$

Если хотя бы один из коэффициентов $a_i,\ i=\overline{0,n-1}$ отличен от 0, то функция $\left(\sum\limits_{i=0}^{n-1} a_i x^i\right)^2$ может обратиться в нуль на $[a,b]$ лишь в конечном числе точек. Отсюда и из условия $p(x) > 0$ следует, что равенство (4) возможно лишь в случае, когда $a_0 = a_1 = \ldots = a_{n-1} = 0$. Тем самым неоднородная система (2) имеет единственное решение. Следовательно, существует единственный многочлен $\varpi(x)$ степени n со старшим коэффициентом 1, ортогональный с весом $p(x) > 0$ любому многочлену степени $n-1$.

Теорема 1. *Если $\varpi(x)$ – многочлен степени n, ортогональный на $[a,b]$ с весом $p(x) > 0$ любому многочлену степени меньше n, то все его корни различны и расположены на $[a,b]$.*

Доказательство. Предположим, что многочлен $\varpi(x)$ имеет $m \geq 0$ различных корней нечетной кратности на $[a,b]$. Очевидно, что $m \leq n$. Теорема будет доказана, если покажем, что $m = n$. Обозначая эти корни через $\xi_1, \xi_2, \ldots, \xi_m$, представим $\varpi(x)$ в виде

$$\varpi(x) = (x-\xi_1)^{\alpha_1}(x-\xi_2)^{\alpha_2}\ldots(x-\xi_m)^{\alpha_m} r(x),$$

где $\alpha_1, \alpha_2, \ldots, \alpha_m$ – нечетные числа и функция $r(x)$ не меняет знак на $[a, b]$ (так как не имеет корней). Вычислим интеграл

$$I = \int\limits_a^b p(x)\varpi(x)(x-\xi_1)\cdots(x-\xi_m)\,dx =$$

$$= \int\limits_a^b p(x)(x-\xi_1)^{\alpha_1+1}\cdots(x-\xi_m)^{\alpha_m+1} r(x)\,dx. \qquad (5)$$

Поскольку $\alpha_1+1, \ldots, \alpha_m+1$ – четные числа и $r(x)$ – знакопостоянная на $[a, b]$, то интеграл (5) отличен от нуля. С другой стороны, если $m < n$, то $q(x) = (x-\xi_1)\cdots(x-\xi_m)$ – многочлен степени меньше n и по условию теоремы имеем $I = 0$. Следовательно, $m = n$, что и доказывает теорему.

Из теорем §33 и §34 следует, что для любого n существует, при том единственная квадратурная формула, точная для любого многочлена, степени $2n-1$.

§35. О положительности квадратурных коэффициентов

Дана квадратурная формула

$$\int_a^b p(x)f(x)\,dx \approx \sum_{k=1}^{n} c_k f(x_k). \tag{1}$$

Знаки квадратурных коэффициентов устанавливаются на основании следующей теоремы.

Теорема 1. *Если $p(x) \geq 0$, $x \in [a,b]$ и квадратурная формула (1) имеет наивысшую алгебраическую степень точности $2n-1$, то все ее коэффициенты c_k положительны.*

Сформулированная теорема является следствием леммы.

Лемма 1. *Если $p(x) \geq 0$, $x \in [a,b]$ и равенство (1) является точным для следующих многочленов степени $2n-2$: $f_i(x) = \left[\dfrac{\varpi(x)}{(x-x_i)}\right]^2$, $i = 1, 2, \ldots, n$, то все квадратурные коэффициенты c_k равенства (1) положительны.*

Доказательство теоремы. Так как в узлах x_k многочлены $f_i(x)$ принимают значения

$$f_i(x_k) = \begin{cases} \left[\varpi'(x_i)\right]^2, & k = i, \\ 0, & k \neq i, \end{cases}$$

то равенство (1) для $f_i(x)$ приводит к результату

$$0 < \int\limits_a^b p(x) f_i(x)\, dx = \sum_{k=1}^{n} c_k f_i(x_k) = c_i \big[\varpi'(x_i)\big]^2.$$

Отсюда $c_i > 0$. Теорема доказана.

§36. Погрешность квадратуры наивысшей степени точности

Будет получена лишь простейшая теорема о представлении погрешности квадратуры наивысшей степени точности, предполагающая достаточно высокий порядок дифференцируемости f.

Теорема 1. *Пусть весовая функция сохраняет знак на* $[a,b]$ *и* f *имеет на* $[a,b]$ *непрерывную производную порядка* $2n$. *Тогда на* $[a,b]$ *существует такая точка* ξ, *что для погрешности* $R(f)=\int\limits_a^b p(x)f(x)\,dx-$ $-\sum\limits_{k=1}^{n} c_k f(x_k)$ *квадратуры наивысшей алгебраической степени точности верно представление*

$$R(f)=\frac{1}{(2n)!}f^{(2n)}(\xi)\int\limits_a^b p(x)\varpi^2(x)\,dx. \qquad (1)$$

Доказательство. Пусть $H(x)$ есть многочлен степени $2n-1$, интерполирующий $f(x)$ на условиях $H(x_k)=f(x_k)$, $H'(x_k)=f'(x_k)$, $k-\overline{1,n}$. Это есть интерполирование с двукратными узлами. В нашей задаче, если предположить, что f имеет на $[a,b]$ непрерывную производную порядка $2n$, то остаточный член такой интерполяции может быть записан в виде

$$r(x)=\frac{1}{(2n)!}\varpi^2(x)f^{(2n)}(\xi),$$

где ξ – есть некоторая точка отрезка, содержащего x и x_k, $k = 1, 2, \dots, n$. Преобразуем искомый интеграл

$$\int\limits_a^b p(x)f(x)\,dx = \int\limits_a^b p(x)H(x)\,dx + \frac{1}{(2n)!}\int\limits_a^b p(x)\omega^2(x)f^{(2n)}(\xi)\,dx. \quad (2)$$

Так как квадратурная формула верна для всяких многочленов степени $2n-1$ и $H(x_k) = f(x_k)$, то

$$\int\limits_a^b p(x)f(x)\,dx = \int\limits_a^b p(x)H(x)\,dx + \frac{1}{(2n)!}\int\limits_a^b p(x)\omega^2(x)f^{(2n)}(\xi)\,dx =$$

$$= \sum_{k=1}^n c_k H(x_k) + \frac{1}{(2n)!}\int\limits_a^b p(x)\varpi^2(x)f^{(2n)}(\xi)\,dx = \sum_{k=1}^n c_k f(x_k) +$$

$$+ \frac{1}{(2n)!}\int\limits_a^b p(x)\varpi^2(x)f^{(2n)}(\xi)\,dx.$$

Итак, $\int\limits_a^b p(x)H(x)\,dx = \sum\limits_{k=1}^n c_k H(x_k) = \sum\limits_{k=1}^n c_k f(x_k)$.

Следовательно, погрешность приближения интегрирования имеет значение

$R(f) = \frac{1}{(2n)!} \int\limits_a^b p(x)\varpi^2(x) f^{(2n)}(\xi)\, dx$. Отсюда можно сделать вывод, что существует точка $\eta \in [a, b]$ такая, что

$$R(f) = \frac{1}{(2n)!} f^{(2n)}(\eta) \int\limits_a^b p(x)\varpi^2(x)\, dx.$$

§37. Связь с ортогональной системой многочленов

Говорят, что последовательность многочленов

$$p_0(x),\ p_1(x), \ldots,\ p_n(x), \ldots,$$

где $p_n(x) = a_n x^n + a_n + x^{n+1} + \ldots + a_0,\ a_n \neq 0$, образует ортогональную систему по весу $p(x)$ на $[a, b]$, если

$$\int\limits_a^b p(x) P_m(x) P_n(x)\, dx = 0, \quad m \neq n. \qquad (1)$$

Определение 1. *Многочлен $P_n(x)$ называется <u>нормированным</u>, если*

$$\int\limits_a^b p(x) {P_n}^2(x)\, dx = 1,\ a_n > 0.$$

Если все многочлены ортогональной системы нормированы, то систему называют ортонормированной.

В условии ортогональности (1) ввиду равноправности индексов m и n, достаточно требовать его выполнения для $m < n$. С другой стороны, так как всякий многочлен $Q(x)$ степени не выше $n-1$ может быть разложен по многочленам $P_m(x)$ $(m = 0, 1, \ldots, n-1)$, условие ортогональности

(1) равносильно требованию, чтобы каждый многочлен $P_n(x)$ системы был ортогонален к любым многочленам степени $m < n$.

Возвратимся к квадратурным формулам наивысшей степени точности. Такая формула может быть построена для каждого $n = 1, 2, \ldots$. Когда n фиксировано, ей будет отвечать свой многочлен ϖ_n и свои квадратурные узлы ${x_k}^n$ $(k = \overline{1, n})$ $\varpi_n(x) = (x - x_1^n) \cdots (x - x_n^n)$. По теореме из §33 он должен быть ортогонален ко всяким многочленам $Q(x)$ степени $m < n$, следовательно, он может отличаться от $P_n(x)$ только численным множителем, который, очевидно, должен быть равен коэффициенту при старшей степени $P_n(x)$: $P_n(x) = a_n \varpi_n(x)$, $n = 1, 2, \ldots$ Поэтому узлы ${x_k}^n$ квадратурной формулы должны быть корнями $P_n(x)$.

§38. Квадратурные формулы, отвечающие простейшим весовым функциям

Здесь будет рассмотрены квадратурные формулы наивысшей алгебраической степени точности, отвечающие классическим весам:

1) постоянному весу $p(x) \equiv 1$;

2) весу Якоби $p(x) = (b-x)^\alpha (x-a)^\beta$, $a < x < b$;

3) весу Лагерра $p(x) = x^\alpha e^{-x}$, $0 < x < \infty$;

4) весу Эрмита $p(x) = e^{-x^2}$ $(-\infty < x < \infty)$.

Эти весовые функции позволяют заранее учитывать наиболее часто встречающиеся в приложениях особенности у интегрируемых функций $F(x)$, путем представления их в виде произведения $F(x) = p(x)f(x)$.

Постоянная весовая функция выбирается в том случае, когда $F(x)$ не имеет особенностей и является достаточно гладкой как внутри отрезка $[a, b]$, так и на его концах. Функция Якоби позволяет учитывать степенные особенности $F(x) = (b-x)^\alpha (x-a)^\beta f(x)$ на концах a и b отрезка интегрирования или на одном из его концов при достаточной гладкости F внутри отрезка.

Вес Лагерра учитывает степенную особенность в точке $x = 0$ и связан со скоростью убывания $F(x) = e^{-x} x^\alpha f(x)$ при $x \to \infty$.

Вес Эрмита связан со скоростью убывания $F(x) = e^{-x^2} f(x)$ при стремлении x к $+\infty$ и $-\infty$ в случае интерполирования по всей числовой оси.

38.1 Постоянная весовая функция

Отрезок интегрирования $[a, b]$ предполагается конечным. Линейным преобразованием аргумента $x' = \dfrac{b+a}{2} + \dfrac{b-a}{2}x$ его можно перевести в $[-1, 1]$. Интеграл берется в форме

$$\int\limits_{-1}^{1} f(x)\,dx, \tag{1}$$

при этом f предполагается достаточно гладкой функцией всюду на $[-1, 1]$, включая его концы.

Известно, что ортогональную на $[-1, 1]$ систему многочленов с постоянным весом образуют многочлены Лежандра

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n = \frac{(2n)!}{2^n(n!)^2}x^n - \dots\ . \tag{2}$$

Они ортонормированны, но не на единицу, так как

$$\int\limits_{-1}^{1} {P_n}^2(x)\,dx = \frac{2}{2n+1}. \tag{3}$$

В формуле квадратур с n узлами

$$\int\limits_{-1}^{1} f(x)\,dx \approx \sum_{k=1}^{n} c_k f(x_k), \tag{4}$$

имеющей наивысшую степень точности $2n-1$ и впервые полученной Гауссом, узлы x_k должны быть корнями многочлена Лежандра степени $n: P_n(x_k)=0,\ k=\overline{1,n}$. Коэффициенты c_k могут быть вычислены при помощи равенства

$$c_k=\int\limits_a^b p(x)\frac{\varpi(x)}{(x-x_k)\varpi'(x_k)}\,dx,$$

но для них существует более простое выражение через многочлен Лежандра, которое мы приведем без вывода $c_k=\dfrac{2(1-{x_k}^2)}{n^2p_{n-1}^2(x_k)}$. Соответствующий формуле (3) многочлен $\varpi(x)$ отличается от $p_n(x)$ постоянным множителем, равным обратной величине старшего коэффициента в $p_n(x):\ \varpi(x)=\dfrac{2^2(n!)^2}{(2n)!}p_n(x)$.

Если воспользоваться этим равенством и (3), то в случае существования у f непрерывной на $[-1,1]$ производной порядка $2n$ погрешность формулы (4) $R(f)$ представима в виде

$$R(f)=\frac{2^{2n+1}}{(2n+1)(2n)!}\left[\frac{(n!)^2}{(2n)!}\right]^2 f^{(2n)}(\xi),\quad -1\le\xi\le1. \tag{5}$$

В настоящее время рассчитано много таблиц формул Гаусса, в частности, при:

$[a,b]=[-1,1]$	$p(x)=1$
$[a,b]=[0,1]$	$p(x)=x^\alpha$
$[a,b]=[0,1]$	$p(x)=(x(1-x))^\alpha$
$[a,b]=[0,+\infty)$	$p(x)=x^s e^{-x}$

38.2 Формула численного интегрирования Эрмита

Рассмотрим формулу Гаусса для вычисления интегралов $\int\limits_{-1}^{1}\frac{f(x)}{\sqrt{1-x^2}}dx$, где $p(x)=\frac{1}{\sqrt{1-x^2}}$:

$$\int\limits_{-1}^{1}\frac{f(x)}{\sqrt{1-x^2}}dx=\sum_{k=1}^{n}{c_k}^n f({x_k}^{(n)})+R(f).$$

Узлы $x^{(n)}$ должны быть корнями многочлена

$$\varpi_n(x)=\frac{1}{2^{n-1}}T_n(x)=\frac{1}{2^{n-1}}\cos(n\arccos x),$$

где $T_n(x)$ – многочлен Чебышева степени n. Тогда

$$\int\limits_{-1}^{1}\frac{f(x)}{\sqrt{1-x^2}}dx=\frac{\pi}{n}\sum_{k=1}^{n}f(x_k)+\frac{f^{(2n)}(\xi)}{(2n)!}\int\limits_{-1}^{1}\frac{{T_n}^2(x)dx}{2^{2n-2}\sqrt{1-x^2}}.$$

Здесь $c_k = \dfrac{\pi}{n}$, $k = \overline{1, n}$, $x_k = \cos\dfrac{(2k-1)\pi}{2n}$. Остаточный член можно упростить $R_n(f) = \dfrac{\pi}{(2n)!2^{n-1}} f^{(2n)}(\xi)$.

38.3 Интегралы вида $\int\limits_a^b (b-x)^\alpha (x-a)^\beta f(x)\,dx$.

Линейным преобразованием $x' = \dfrac{a+b}{2} + \dfrac{b-a}{2}x$, $-1 \le x \le 1$ отрезок интегрирования $[a, b]$ приводится к отрезку $[-1, 1]$, и достаточно рассмотреть интеграл вида

$$\int\limits_{-1}^{1} (1-x)^\alpha (1+x)^\beta f(x)\,dx, \quad \alpha, \beta > -1.$$

Ортогональными на $[-1, 1]$ с весом $p(x) = (1-x)^\alpha (1+x)^\beta$ являются многочлены Якоби

$$P_n^{(\alpha,\beta)}(x) = \frac{(-1)^n}{2^n n!}(1-x)^{-\alpha}(1+x)^{-\beta}\frac{d^n}{dx^n}\left[(1-x)^{\alpha+n}(1+x)^{\beta+n}\right] =$$

$$= \frac{\Gamma(\alpha+\beta+2n+1)}{2^n n!\Gamma(\alpha+\beta+n+1)}x^n - \ldots . \qquad (6)$$

Здесь буквой Γ обозначена гамма-функция. *Гамма–функция* – это несобственный интеграл, зависящий от параметра $\Gamma(\alpha) = \int\limits_0^{+\infty} x^{\alpha-1}e^{-x}\,dx$. В квадратурной формуле наивысшей точности

$$\int\limits_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}f(x)dx \approx \sum_{k=1}^{n} c_k f(x_k) \tag{7}$$

узлы располагаются в корнях многочлена Якоби степени n :

$$P_n^{(\alpha,\beta)}(x_k) = 0,\ k = 1, 2, \ldots, n.$$

Коэфиценты c_k могут быть найдены по общей формуле, но для них известно более удобное для вычислений выражения через многочлен Якоби $P_n^{(\alpha,\beta)}(x)$:

$$c_k = 2^{\alpha+\beta+1}\frac{\Gamma(\alpha+n+1)\Gamma(\beta+n+1)}{n!\Gamma(\alpha+\beta+n+1)(1-x_k^2)\left[P_n^{(\alpha,\beta)'}(x_k)\right]^2}.$$

Соответствующий формуле (7) многочлен $\varpi(x)$ связан с $P_n^{(\alpha,\beta)}(x)$ равенством $\varpi(x) = \dfrac{2^n n!\Gamma(\alpha+n+1)}{\Gamma(\beta+2n+1)}P_n^{(\alpha,\beta)}(x)$. Известно, что для многочлена Якоби имеет место соотношение

$$\int\limits_{-1}^{1}(1-x)^{\alpha}(1+x)^{\beta}\left[P_n^{(\alpha,\beta)}(x)\right]^2 dx=\frac{2^{(\alpha+\beta+1)}\Gamma(\alpha+n+1)\Gamma(\beta+n+1)}{(\alpha+\beta+2n+1)\Gamma^2(\alpha+\beta+2n+1)}.$$

Два последних равенства позволяют вычислить для формулы (7) интеграл, который входит в общее представление погрешности приближенной квадратуры наивысшей степени точности. После простых вычислений получим

$$R(f)=\frac{f^{(2n)}(\xi)}{(2n)!}\frac{2^{\alpha+\beta+2n+1}\Gamma(\alpha+n+1)\Gamma(\beta+n+1)}{(\alpha+\beta+2n+1)\Gamma^2(\alpha+\beta+2n+1)}\Gamma(\alpha+\beta+n+1),$$

$$-1\leq\xi\leq 1. \tag{8}$$

Равенство (7) содержит два произвольных параметра α и β и является, по сути дела, записью семейства квадратурных формул. Придавая α и β численные значения, будем из (7) получать частные формулы, учитывающие те или иные особенности поведения интегрируемой функции в точках $x=\pm1$.

38.4 Интегралы вида $\int\limits_0^\infty x^\alpha e^{-x} f(x)dx$

Ортогональную систему на полуоси $[0, +\infty)$ с весом $p(x) = x^\alpha e^{-x}$, $\alpha > -1$ образуют многочлены Чебышева-Лагерра

$$L_n^{(a)}(x) = (-1)^n x^{-\alpha} e^x \frac{d^n}{dx^n}(x^{\alpha+n} e^{-x}) = x^n - \frac{n(n+\alpha)}{1!} x^{n-1} + \dots .$$

В формуле приближенного интегрирования наивысшей степени точности

$$\int\limits_0^\infty x^\alpha e^{-x} f(x)dx = \sum_{k=1}^{n} c_k f(x_k) + R(f) \qquad (9)$$

за узлы x_k должны быть взяты корни многочлена $L_n^{(\alpha)}(x)$

$$L_n^{(\alpha)}(x) = 0, \;\; k = 1, 2, \dots, n.$$

Для вычисления коэффициентов c_k здесь применяется формула

$$c_k = \frac{\Gamma(n+1)\Gamma(\alpha+n+1)}{x_k \left[L_n^{(\alpha)\prime}(x_k)\right]^2} = \frac{x_k \Gamma(n)\Gamma(\alpha+n)}{n(\alpha+n)\left[L_{n-1}^{(\alpha)}(x_k)\right]^2}.$$

Укажем еще представление погрешности $R(f)$. Для этого воспользуемся ее общим выражением (6). Для формулы (9) многочлены $\varpi(x)$ и $L_n^{(\alpha)}(x)$ совпадают, так как старшие коэффициенты у них равны 1. Корни x_k у них одинаковы. Кроме того, $p(x) = x^\alpha e^{-x}$. Поэтому

$$\int_o^\infty p(x)\varpi^2(x)dx = \int_0^\infty x^\alpha e^{-x}\left[L_n^{(\alpha)}(x)\right]^2 dx.$$

Численное значение последнего интеграла находится легко, и в теории многочленов Лагерра известно, что оно равно $n!\Gamma(\alpha+n+1)$. Поэтому, если f имеет непрерывную производную порядка $2n$ на $[0,+\infty)$, то для $R(f)$ верно соотношение

$$R(f) = \frac{n!\Gamma(\alpha+n+1)}{(2n)!} f^{(2n)}(\xi),\ 0 \le \xi \le \infty. \tag{10}$$

38.5 Интегралы вида $\int\limits_{-\infty}^{\infty} e^{-x^2} f(x)dx$

На всей числовой оси с весом $p(x) = e^{-x^2}$ ортогональными являются многочлены Чебышева-Эрмита: $H_n(x) = (-1)^n e^{x^2}\dfrac{d^n}{dx^n}e^{-x^2} = 2^n x^n - \ldots$.
В формуле приближенного интегрирования

$$\int\limits_{-\infty}^{\infty} e^{-x^2} f(x)dx = \sum_{k=1}^{n} c_k f(x_k) + R(f), \tag{11}$$

имеющей степень точности $2n-1$, узлы x_k должны быть корнями многочлена $H_n(x) : H_n(x_k) = 0$. Многочлен $\varpi(x)$ для формулы (11) отличается от $H_n(x)$ численным множителем 2^{-n} : $\varpi(x) = 2^{-n} H_n(x)$.
Для вычисления коэффициентов c_k используется представление

$$c_k = \frac{2^{n+1} n! \sqrt{\Pi}}{[H_n'(x_k)]^2} = \frac{2^{n-1}(n-1)! \sqrt{\Pi}}{n H_{n-1}^2(x_k)}.$$

Для нахождения выражения погрешности воспользуемся известным в теории многочленов $H_n(x)$ соотношением $\int\limits_{-\infty}^{\infty} e^{-x^2} H_n^2(x)dx = 2^n n! \sqrt{\pi}$. Тогда

$$\int\limits_{-\infty}^{\infty} p(x)\varpi^2(x)dx = 2^{-2n} \int\limits_{-\infty}^{\infty} e^{-x^2} H_n^2(x)dx = \frac{n! \sqrt{\Pi}}{2^n};$$
$$R(f) = \frac{n! \sqrt{\Pi}}{2^n (2n)!} f^{(2n)}(\xi), \tag{12}$$

где $-\infty < \xi < \infty$.

Для формул (9) и (11) таблицы значение узлов x_k и коэффициентов c_k можно найти в книге [9].

§39. Формулы численного интегрирования, содержащие заранее предписанные узлы

С такими формулами приходится встречаться, например, при решении следующих задач. Рассмотрим для дифференциального уравнения граничную или даже многоточечную задачу с заданными значениями функций в нескольких точках рассматриваемого отрезка числовой оси.

Такие задачи во многих случаях могут быть сведены к решению интегральных уравнений. К последним же для приведения их к системе численных уравнений может быть применен метод квадратур, состоящий в том, что интегралы, входящие в уравнение, вычисляются посредством какой-либо из квадратурных формул. Но тогда при построении квадратурной формулы естественно воспользоваться теми точками, в которых известны значения функции, приняв эти точки за узлы формулы и взяв еще несколько узлов, выбором которых можно распоряжаться для увеличения точности.

39.1 Содержание задачи и общие теоремы

Будет рассматриваться квадратурная формула

$$\int\limits_a^b p(x)f(x)dx \approx \sum_{k=1}^{n} c_k f(x_k) + \sum_{c=1}^{m} B_i f(a_i), \tag{1}$$

которая содержит m фиксированных узлов a_i, $(i = \overline{1,m})$. В нее входят также $2n+m$ произвольных параметров c_k, x_k, $(k = \overline{1,m})$ и B_i, $(i = = \overline{i,m})$. Их можно надеяться выбрать так, чтобы формула (1),была точной для любых многочленов степени меньшей $2n+m$. Известно, см. §31 (теорема 1), что при всяких x_k и a_i равенство (1) можно сделать точным для всех многочленов степени $n+m+1$ при помощи выбора c_k и B_i; для этого необходимо, чтобы формула (1) была интерполяционной. Для (1) это означает, что ее коэффициенты должны иметь следующие значения:

$$c_k = \int_a^b p(x) \frac{\omega(x)\Omega(x)}{(x-x_k)\omega'(x_k)\Omega(x_k)} dx, \ \ k = \overline{1,n},$$

$$B_i = \int_a^b p(x) \frac{\omega(x)\Omega(x)}{(x-a_i)\omega(a_i)\Omega'(a_i)} dx, \ \ i = \overline{1,m},$$

$$\omega(x) = (x-x_1)\dots(x-x_n), \ \Omega(x) = (x-a_1)\dots(x-a_m). \tag{2}$$

Теорема 1. *Чтобы формула (1) была точной для многочленов степени*
$2n+m+1$, *необходимо и достаточно выполнение условий:*

1. *Формула является интерполяционной, т.е. ее коэффициенты c_k и B_i имеют значения (2);*

2. *узлы* x_k, $(k = \overline{1,n})$ *таковы, что соответствующий им многочлен* $\omega(x)$ *ортогонален с весом* $p(x)\Omega(x)$ *на* $[a,b]$ *ко всякому многочлену* $Q(x)$ *степени меньшей* n :

$$\int\limits_a^b p(x)\Omega(x)\omega(x)Q(x)dx = 0. \tag{3}$$

Для проверки справедливости утверждений теоремы достаточно повторить с несущественными изменениями – с заменой веса $p(x)$ на вес $p(x)\Omega(x)$ и квадратурной суммы $\int\limits_a^b p(x)f(x)dx \approx \sum\limits_{k=1}^{n} c_k f(x_k)$ на сумму (1) доказательства теоремы 1 из §33 (основной теоремы).

Рассмотренная теорема требует существования многочлена $\omega(x)$, обладающего свойством ортогональности (3). Кроме того, по существу задачи корни $\omega(x)$ должны все принадлежать отрезку $[a,b]$.

Наличие обоих факторов можно гарантировать в том случае, когда весовая функция (в нашей задаче это есть $p(x)\Omega(x)$) сохраняет знак на отрезке $[a,b]$. Последнее же может произойти в том и только том случае, когда точки перемены знака $p(x)$, лежащей внутри $[a,b]$ совпадают с такими же точками $\Omega(x)$, т.е. с фиксированными узлами a_i, $i = \overline{1,m}$.

Наибольший интерес в приложениях имеет случай, когда $p(x) \geq 0$ и, следовательно, внутри $[a,b]$ весовая функция перемен знака не имеет. Поэтому особое значение имеют случаи, когда фиксированными узлами являются концы отрезка, оба или один из них. Предположим теперь, что многочлен $\omega(x)$ существует и формула (1), имеющая степень точности $2n+m-1$, может быть построена. Рассмотрим погрешность этой формулы и получим для нее представление, рассчитанное на функции высокого порядка гладкости.

Построим интерполирование функции f при помощи многочлена $H(x)$ степени $2n+m-1$ по следующим условиям

$$H(x_k) = f(x_k),\ \ H'(x_k) = f'(x_k), k = \overline{1,n},$$

$$H(a_i) = f(a_i),\ \ i = \overline{1,m}.$$

Это и есть интерполирование с n двукратными узлами x_k и m простыми узлами a_i.

Если функция f имеет на отрезке $[a,b]$, где располагаются x, x_k, a_i, непрерывную производную порядка $2n+m$, то остаточный член интеполирования $r(x)$ имеет следующее представление

$$r(x) = w^2(x)\Omega(x)\frac{f^{(2n+m)}(\xi)}{(2n+m)!},\ \ a \le \xi \le b.$$

Так как $f(x) = H(x)+r(x)$, то для погрешности $R(f)$ верно равенство $R(f) = \int\limits_a^b p(x)r(x)dx - \sum\limits_{k=1}^{n} c_k r(x_k) - \sum\limits_{i=1}^{m} B_i r(a_i)$ или, по причине $r(x_k)=0$ и $r(a_i)=0$,

$$R(f)=\int\limits_a^b p(x)r(x)dx=\frac{1}{(2n+m)!}\int\limits_a^b p(x)\Omega(x)\omega^2(x)f^{(2n+m)}(\xi)dx. \quad (4)$$

Формула (4) для $R(f)$ позволяет ответить на вопрос о степени алгебраической точности формулы (1) в наиболее важном случае.

Теорема 2. *Если узлы x_k и a_i формулы (1) таковы, что*

$$\int\limits_a^b p(x)\Omega(x)\omega^2(x)dx\neq 0, \quad (5)$$

то формула (1) не может быть точной для многочленов степени $2n+m$.

Доказательство. Оно очевидным образом вытекает из (5), так как если f есть многочлен степени $2n+m$, то для f производная $f^{(2n+m)}$ есть величина постоянная и; поэтому в (1) левая и правая часть не $R(f)=$ $=\dfrac{f^{(2n+m)}(\xi)}{(2n+m)!}\int\limits_a^b p(x)Q(x)\omega^2(x)dx\neq 0$ могут совпадать (остаток $\neq 0$).
Теорема 2 доказана.

§40. Квадратурные формулы с равными коэффицинтами

40.1 Построение формулы

Формулы квадратур с одинаковыми коэффициентами

$$\int_a^b p(x)f(x)dx \approx c_n \sum_{k=1}^{n} f(x_k) \tag{1}$$

особенно удобны при работе с чертежами, когда ординаты легко снимаются с чертежа и столь же просто суммируются. Формула (1) содержит $n+1$ параметров c_k, x_k, $(k = \overline{1, n})$ и их естественно выбрать так, чтобы равенство (1) выполнялось точно для все многочленов степени n или, что равносильно, выполнялось для степеней x от нулевой до n :

$$\int_a^b p(x)x^i dx = c_n \sum_{k=1}^{n} x_k^i, i = 0,\ 1, \dots, n. \tag{2}$$

К принятому ранее предположению об абсолютной интегрируемости на $[a, b]$ произведений $p(x)x^m$, $(m = 0,\ 1, \dots)$ мы сделаем еще одно предположение о весе $p(x)$, естественное в рассматриваемой задаче. Будем считать, что выполняется условие

$$I_0 = \int\limits_a^b p(x)dx \neq 0. \tag{3}$$

Если окажется, что $I_0 = 0$ и если, кроме того, что равенство (1) выполняется точно в случае, когда $f(x) \equiv 1$, т.е. что верно равенство

$$\int\limits_a^b p(x)dx = nc_n, \tag{4}$$

то должно быть $c_n = 0$, и формула (1) тогда теряет всякое значение.

Выясним возможность решения системы (2) относительно c_n и x_i. При $i = 0$ получится равенство (4), из него найдем

$$c_n = \frac{1}{n}\int\limits_a^b p(x)dx. \tag{5}$$

Полагая последовательно $i = 1, 2, \dots, n$, получим систему уравнений для нахождения x_i

$$\begin{cases} s_1 = x_1 + x_2 + \ldots + x_n = c_n^{-1} \int\limits_a^b px dx = c_n^{-1} \mu_1, \\ s_2 = x_1^2 + x_2^2 + \ldots + x_n^2 = c_n^{-1} \int\limits_a^b px^2 dx = c_n^{-1} \mu_2, \\ \ldots \\ s_n = x_1^n + x_2^n + \ldots + x_n^n = c_n^{-1} \int\limits_a^b px^n dx = c_n^{-1} \mu_n. \end{cases} \tag{6}$$

Рассмотрим многочлен $\omega(x)$, для которого x_i являются корнями:

$$\omega(x) = (x - x_1) \ldots (x - x_n) = x^n + a_1 x^{n-1} + \ldots + a_n. \tag{7}$$

Уравнения (6) дают численные значения сумм степеней корней от s_1 до s_n. В алгебре многочленов известны соотношения между коэффициентами многочлена $a_1, \ldots, a_n$ и суммами степеней корней

$$\begin{gathered} s_1 + a_1 = 0, \\ s_2 + a_1 s_1 + 2a_2 = 0, \\ s_3 + a_1 s_2 + a_2 s_1 + 3a_3 = 0, \\ \ldots \\ s_n + a_1 s_{n-1} + s_2 s_{n-2} + \ldots + n a_n = 0. \end{gathered} \tag{8}$$

Значения $s_i, i = \overline{1,n}$ известны, с их помощью мы можем, и притом единственным образом, найти коэффициенты a_i многочлена. После этого, решая уравнение $\omega(x) = 0$, найдем узлы x_i квадратурной формулы (1). Но необходимо заметить, что корни многочлена $\omega(x)$ могут оказаться комплексными или выходить за границы $[a, b]$.

Все изложенное позволяет сформулировать теорему.

Теорема 1. *Если $\int\limits_a^b p(x)dx \neq 0$, квадратурная формула вида (1) с действительными или комплексными узлами x_k, точная для любых алгебраических многочленов степени n, может быть построена, и при этом единственным образом, при всяких $n = 1, 2, \ldots$.*

Определение 1. *Формулы квадратур типа (1), обладающие свойством, чтобы равенство (1) выполнялось точно для всех многочленов степени n (для степеней x от нулевой до n), называются формулами Чебышева.*

40.2 Случай постоянной весовой функции

Положим $p(x) \equiv 1$ и отрезок интегрирования приведенным к $[-1, 1]$. Тогда

$$\int\limits_{-1}^{1} f(x)dx \approx c_n \sum_{k=1}^{n} f(x_k). \tag{9}$$

Коэффициент c_n определяется при помощи равенства (8), т.е. $c_n =$

$= \frac{1}{n}\int\limits_{-1}^{1} dx = \frac{2}{n}$. Так как $\int\limits_{-1}^{1} x^i dx = [1-(-1)^{i+1}]/(i+1)$, то уравнения (6) дадут для сумм степеней узлов x_i значения

$$s_1 = \frac{n}{2}\int\limits_{-1}^{1} xdx = 0,\;\; s_2 = \frac{n}{2}\int\limits_{-1}^{1} x^2dx = \frac{n}{3}, \dots, s_n = \frac{n}{2}\frac{1}{n+1}[1-(-1)^{n+1}].$$

Поэтому система уравнений для коэффициентов a_k многочленов $\omega(x)$ имеет вид:

$$\begin{cases} a_1 = 0, \\ \frac{n}{3} + 2a_2 = 0, \\ a_3 = 0, \\ \frac{n}{5} + \frac{n}{3}a_2 + 4a_4 = 0, \\ a_5 = 0, \\ \frac{n}{7} + \frac{n}{5}a_2 + \frac{n}{3}a_4 + 6a_6 = 0, \\ \dots \end{cases}$$

Отсюда находим $a_1, a_2, a_3, \dots, a_n$.

При $n = 1$ $\omega(x) = x$, $x_1 = 0$, $c_1 = 2$, поэтому $\int\limits_{-1}^{1} f(x)dx \approx 2f(0)$. Это есть формула прямоугольников с высотой, равной ординате в середине отрезка интерполирования.

Для $n = 2$ $\omega(x) = x^2 - \frac{1}{3}$, $x_1 = -\frac{1}{\sqrt{3}}$, $x_2 = \frac{1}{\sqrt{3}}$ $c_3 = 1$, откуда $\int\limits_{-1}^{1} f(x)dx \approx f(-\frac{1}{\sqrt{3}})+f(\frac{1}{\sqrt{3}})$. Эта формула совпадает с формулой Гаусса для двух узлов.

§41. Интерполяционные кубатурные формулы

Требуется вычислить интеграл

$$I(f) = \int\limits_{\Omega} p(x)f(x)dx.$$

Область интегрирования $\Omega \subset R^n$ (евклидову пространству), $p(x)$ – весовая функция – предполагается заданной и такой, что существуют ее моменты – интегралы

$$P_\alpha = \int\limits_{\Omega} p(x)x^\alpha dx, \tag{1}$$

где $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ – мультииндекс $a_i \geq 0$. В некоторых случаях удобно пользоваться обозначением

$$P_i = \int\limits_{\Omega} p(x)\varphi_i(x)dx, \; i =, 1, 2 \ldots \tag{2}$$

Определение 1. *Кубатурной формулой для вычисления интеграла $I(f)$ называется приближенное равенство*

$$\int\limits_{\Omega} p(x)f(x)dx \cong \sum_{j=1}^{N} c_j f(x^{(j)}). \tag{3}$$

Точки $x^{(j)}$ считаются попарно различными и называются узлами кубатурной формулы, а c_j – ее коэффициенты.

Мы не предполагаем, что узлы принадлежат области интегрирования Ω, более того допускаем, что среди узлов и коэффициентов могут быть комплексные.

Сумма в правой части (3) называется кубатурной.

$R(f) = \int\limits_{\Omega} p(x)f(x)dx - \sum\limits_{j=1}^{N} c_j f(x^{(j)})$ – погрешность кубатурной формулы.

Погрешность $R(f)$ является однородным и аддитивным функционалом в векторном пространстве всех многочленов.

Один из способов построения кубатурных формул основан на интерполировании. Выберем точки $x^{(i)}, i = 1, 2, \ldots, \mu, \mu = M(n, m) = \dfrac{(n+m)!}{n!m!}$, не лежащие на алгебраической гиперповерхности порядка m ($n-$ размерность R^n), и построим интерполяционный многочлен $P(x)$ функции $f(x)$ степени не выше m ($\leq m$). Многочлен $P(x)$ удовлетворяет условиям $P(x^{(i)}) = f(x^{(i)}), i = 1, 2, \ldots, \mu$. Получим приближенное равенство

$$f(x) \approx \sum_{j=1}^{\mu} L_j(x) f(x^{(j)}). \tag{4}$$

умножим обе части равенства на $p(x)$ и проинтегрируем по Ω, получим

$$\int\limits_{\Omega} p(x)f(x)dx \cong \sum_{j=1}^{\mu} c_j f(x^{(j)}), \tag{5}$$

где

$$c_j = \int\limits_{\Omega} p(x)L_j(x)dx, \ j = 1, \ldots, \mu. \tag{6}$$

Определение 2. *Кубатурная формула (5), у которой:*

1. *число узлов равно μ;*
2. *узлы не лежат на алгебраической гиперповерхности порядка m;*
3. *коэффициенты определяются равенствами (6) называется интерполяционной кубатурной формулой.*

Интерполяционная кубатурная формула однозначно определяется заданием узлов, так как ее коэффициенты определяются единственым образом равенствами (6). Справедлива теорема.

Теорема 1. *Если $\mu = M(n, m)$ попарно различных точек $x^{(i)}, i = \overline{1, \mu}$ не лежат на алгебраической гиперповерхности порядка m, то существует единственная интерполяционная кубатурная формула, узлами которой являются эти точки.*

Если узлы интерполяционной кубатурной формулы вещественны, то вещественны ее коэффициенты.

Интерполяционная формула обладает $m-$ свойством: она обращает в точное равенство, если $f(x)$ является любым многочленом степени не выше m. Действительно, в этом случае $f(x)$ совпадает со своим интерполяционным многочленом, так что равенство (4) является точным при всех $x \in R^n$, и следовательно (5) точное.

Коэффициенты интерполяционной кубатурной формулы можно находить, решая линейную алгебраическую систему. Именно, запишем, что кубатурная формула (5) точна для всех одночленов степени не выше $m: \ c_1\varphi_i(x^{(1)})+\ldots+c_\mu\varphi_i(x^{(\mu)}) = p_i, \ i=\overline{1,\mu}$, где p_i из (2). Эти равенства представляют собой линейную алгебраическую систему относительно коэффициентов c_i. Матрицей системы будет

$$V'_m = \left[\varphi_i(x^{(1)}),\ldots,\varphi_i(x^{(\mu)})\right]_{i=1}^{\mu},$$

она неособенная, транспонированная к матрице Вандермонда, определяемой точками $x^{(i)}, i=\overline{1,\mu}$. Так как точки не лежат на алгебраической гиперповерхности порядка m, то эта матрица неособенная $(D \neq 0)$

$$\omega_i(x) = \begin{vmatrix} \varphi_1(x) & \varphi_2(x) & \dots & \varphi_\mu(x) \\ \varphi_1(x^{(1)}) & \varphi_2(x^{(1)}) & \dots & \varphi_\mu(x^{(1)}) \\ \dots & & & \\ \varphi_1(x^{(\mu)}) & \varphi_2(x^{(\mu)}) & \dots & \varphi_\mu(x^{(\mu)}) \end{vmatrix}$$

– определитель Вандермонда.

Очевидно, $\omega_i(x^{(j)}) = 0$ при $i \neq j$, $\omega_i(x^{(i)}) = (-1)^{i-1} D(V_m) \neq 0$.

§42. Кубатурная формула трапеций

Процедура вычисления двойного интеграла называется кубатурой. Обычно мы двойной интеграл заменяем повторным интегралом.

$$\iint\limits_D f(x,y)dxdy = \int\limits_a^b \left[\int\limits_c^d f(x,y)dy\right] dx.$$

Получим кубатурную формулу трапеций для вычисления двойного интеграла.

Прежде всего отметим формулу для вычисления объема усеченной призмы, в основании которой прямоугольный треугольник ABC, $AC = a$, $BC = b$. Пусть $AD = z_3, BE = z_1, CF = z_2$, тогда объем усеченной призмы равен $V = \dfrac{ab}{6}(z_1 + z_2 + z_3)$.

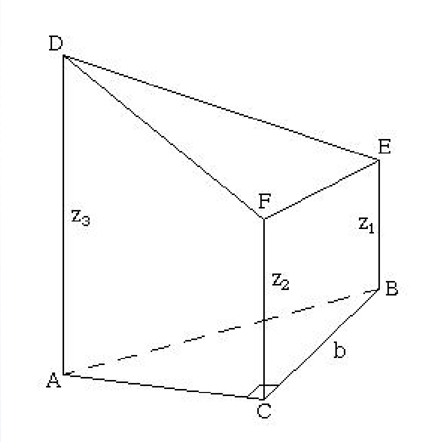

Рисунок 1.

Пусть требуется вычеслить двойной интеграл $\iint\limits_D f(x,y)dxdy$. Область D будем предполагать прямоугольной, а функцию $f(x,y)$ положительной. В этом случае двойной интеграл численно равен объему тела, основанием которого является область D, а сверху тело ограничено поверхностью $z = f(x,y)$.

Прямоугольник D, служащий основанием тела, разобьем на прямоугольник со сторонами h и k, а каждый прямоугольник диагонально разделяем на треугольники.

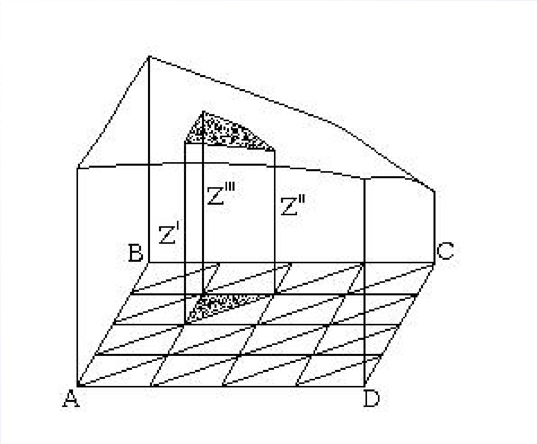

Рисунок 2.

Через прямые, разделяющие основание тела на прямоугольники проведем плоскости, перпендикулярные основанию. Такие же плоскости проведем через диагонали частичных прямоугольников. Данное тело окажется разбитым на элементарные тела (элементарные криволинейные призмы).

Обозначим вершины частичных прямоугольников, на которые разделена область D двойными индексами i и j.. Соответствующее значение функции $z = f(x, y)$ в этой точке обозначим z_{ij}. Заменим все криволинейные призмы усеченными прямолинейными с теми же боковыми ребрами. Объем такой прямолинейной усеченной призмы равен

$$\triangle V = \frac{hk}{6}(z' + z'' + z'''), \qquad (1)$$

где z', z'', z''' – длины боковых ребер.

За приближенное значение объема (и, значит, приближенное значение интеграла) примем сумму объемов всех таких прямолинейных усеченных призм.

При суммировании следует учесть, что некоторые ординаты z_{ij} будут повторяться два, три, шесть раз, так как соответствующая точка является вершиной нескольких треугольников. В результате получим

$$V = \frac{hk}{6}\left(\sum{}_1 z_{ij} + 2\sum{}_2 z_{ij} + 3\sum{}_3 z_{ij} + 6\sum{}_4 z_{ij}\right), \qquad (2)$$

где $\sum_1$ – сумма, относящаяся к тем двум вершинам прямоугольника D, которые не являются вершинами проведенных диагоналей (это B, D); $\sum_2$ – это сумма, относящаясе к тем двум вершинам прямоугольника D, которые являются вершинами диагоналей (это A и C); $\sum_3$ – сумма, относящаяся к вершинам частичных прямоугольников, лежащих строго внутри сторон прямоугольника D; $\sum_4$ – сумма, относящаяся к вершинам частичных прямоугольников, лежащих строго внутри прямоугольника D (см. рис. 3).

Формула (2) – кубатурная формула трапеций.

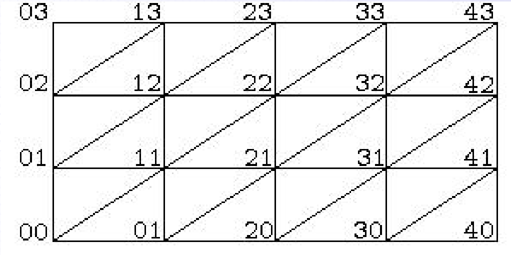

Рисунок 3.

§43. Кубатурная формула Симпсона

Для приближенного вычисления двойного интеграла получим формулу Симпсона, более точную, чем формула трапеци. Будем вычислять двойной интеграл $I = \iint\limits_D f(x,y)dxdy$ по прямоугольной области со сторонами, параллельными координатным осям. Обозначим длины сторон прямоугольника $2h$ и $2k$. Разобьем этот прямоугольник на четыре равных прямоугольника (стороны каждого из них равны h и k). Вершины малых прямоугольников обозначим двойными индексами.

Они будут служить узлами в формуле приближенного вычисления функции $z = f(x,y)$ в этих узлах обозначим: $z_{00}, z_{10}, z_{20}, z_{01}, z_{11}, z_{21}, z_{02}, z_{12}, z_{22}$.

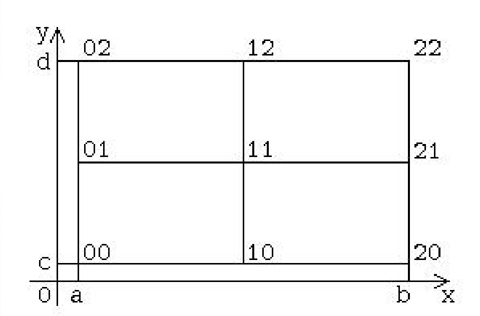

Рисунок 1.

Двойной интеграл представим в виде повторного интеграла:

$$I=\int\limits_{c}^{c+2k}\left[\int\limits_{a}^{a+2h} f(x,y)dx\right]dy.$$

Рассмотрим внутренний интеграл

$$F(y)=\int\limits_{a}^{a+2h} f(x,y)dx.$$

Это будет функция переменной y, определенная на отрезке $c\leq y\leq$ $\leq c+2k$. Найдем ее значения $F_0=F(c)$, $F_1=F(c+k)$, $F_2=F(c+2k)$, применяя формулу Симпсона для однократного интеграла:

$$F_0 = F(c) = \int\limits_a^{a+2h} f(x,c)dx \approx \frac{h}{3}[z_{00} + 4z_{10} + z_{20}],$$

$$F_1 = F(c+k) = \int\limits_a^{a+2h} f(x,c+k)dx \approx \frac{h}{3}[z_{01} + 4z_{11} + z_{21}],$$

$$F_2 = F(c+2k) = \int\limits_a^{a+2h} f(x,c+2k)dx \approx \frac{h}{3}[z_{02} + 4z_{12} + z_{22}].$$

Теперь вычислим $I = \int\limits_c^{c+2k} F(y)dy$. Мы опять имеем однократный интеграл (по формуле Симпсона) $I \approx \frac{k}{3}[F_0 + 4F_1 + F_2]$.

Подставим сюда найденные значения F_0, F_1, F_2, получим окончательную формулу

$$\iint\limits_D f(x,y)dxdy \approx \frac{hk}{9}[z_{00} + z_{02} + z_{20} + \\ + z_{22} + 4(z_{01} + z_{10} + z_{21} + z_{12}) + 16z_{11}]. \quad (1)$$

Заметим, что значение функции в вершинах прямоугольника взяты с коэффициентом 1, значение функции в центре прямоугольника с коэф-

фициентом 16, а значения функции в серединах сторон прямоугольника D с коэффициентами 4.
Формула (1) называется элементарной кубатурной формулой Симпсона.

Чтобы получить более точное значение интеграла, разобьем прямоугольник на любое (достаточно большое) число m равных прямоугольников со сторонами $2h$ и $2k$. К каждому из них применим формулу (1). Потом в каждом из этих прямоугольников отметим точки: вершины прямоугольников, середины сторон и центр. Применим к кажому прямоугольнику формулу (1) и найдем сумму результатов вычеслений для всех прямоугольников. Получим формулу

$$\iint\limits_D f(x,y)dxdy \approx$$

$$\approx \frac{hk}{9}\left[\sum{}_1 z_{ij} + 2\sum{}_2 z_{ij} + 4\sum{}_3 z_{ij} + 4\sum{}_4 z_{ij} + 8\sum{}_5 z_{ij} + 16\sum{}_6 z_{ij}\right], \tag{2}$$

где z_{ij} – значение функции $f(x,y)$ в узлах.
Здесь

1. $\sum_1$ – сумма, значений в вершинах основного прямоугольника (A, B, C, D);

2. $\sum_2$ – сумма, значений в вершинах частичных прямоугольников, лежащих внутри сторон основного прямоугольника;

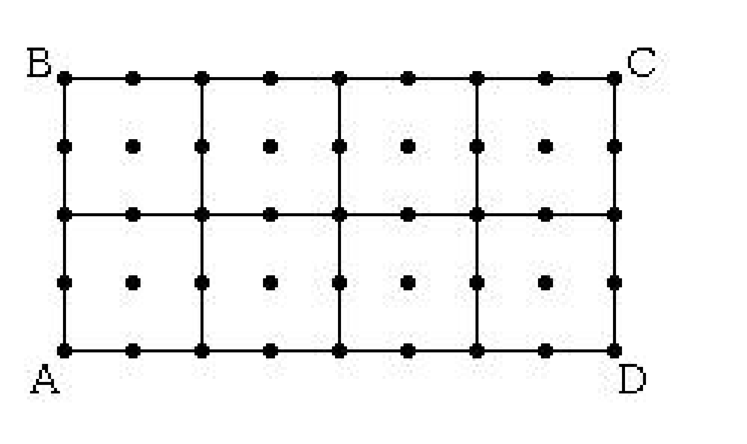

Рисунок 2.

3. $\sum_3$ – сумма, значений в вершинах частичных прямоугольников, которые лежат внутри основного прямоугольника;

4. $\sum_4$ – сумма, значений в серединах сторон частичных прямоугольников, лежащих на сторонах основного прямоугольника;

5. $\sum_5$ – сумма, значений в серединах частичных прямоугольников, лежащих внутри основного прямоугольника;

6. $\sum_6$ – сумма, значений в центрах частичных прямоугольников.

Формула (2) – кубатурная (объемная) формула Симпсона.

ГЛАВА 4
МЕТОДЫ ЧИСЛЕННОГО РЕШЕНИЯ ЗАДАЧИ КОШИ ДЛЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

§44. Аналитические методы решения задачи Коши

Среди задач, с которыми приходиться иметь дело в вычислительной практике, значительную часть составляют различные задачи для обыкновенных дифференциальных уравнений. В курсе дифференциальных уравнений изучаются методы интегрирования простейших видов этих уравнений. К сожалению, дифференциальные уравнения, которые можно проинтегрировать известными методами, встречаются очень редко. В связи с этим приобретают особое значение приближенные методы решения дифференциальных уравнений. Такие методы в зависимости от того, ищется ли приближенное решение в аналитическом виде или в виде таблицы чисел, часто разделяют на аналитические и численные. Аналитические методы дают приближенное решение в виде аналитического выражения, численные методы в виде таблицы численных значений.

В случае обыкновенных дифференциальных уравнений в зависимости от того, ставятся ли дополнительные условия в одной или нескольких точках отрезка изменения независимой переменной, задачи обычно подразделяют на одноточечные (задачи с начальными условиями, или задачи Коши) и многоточечные. Среди многоточечных задач наиболее часто в прикладных задачах встречаются так называемые граничные задачи, когда дополнительные условия ставятся на концах рассматриваемого отрезка.

Наиболее простой из указанных выше задач является задача Коши.

Пусть на отрезке $x_0 \le x \le X$ требуется найти решения $y(x)$ дифференциального уравнения

$$y' = f(x, y), \tag{1}$$

удовлетворяющего начальному условию:
при

$$x = x_0 \quad y(x_0) = y_0. \tag{2}$$

Условия существования и единственности задачи Коши будем считать выполненными. Будем считать также, что функция $f(x, y)$ в некоторой области изменения ее аргументов сколь угодно раз дифференцируема по x и y (обладает нужной гладкостью).

Метод последовательных приближений (метод Пикара) приближенного решения задачи (1), (2)

Этот метод позволяет получить в аналитичеком виде последовательность приближений $y_m(x)$, $m = 1, 2, \ldots$ к решению $y(x)$ рассматриваемой задачи Коши по следующему итерационному правилу

$$y_m(x) = y_0 + \int\limits_{x_0}^{x} f\left[t, y_{m-1}(t)\right] dt,\ x_0 \le x \le X,\ y(x_0) = y_0,\ x_0 \le t \le X,$$

$$m = 1, 2, \ldots$$

Метод Пикара, однако редко используется на практике. Одним из его существенных недостатков, препятствующих широкому практическому применению, является необходимость выполнения операции интегрирования при осуществлении каждой итерации.

Несколько более широкое распространение в вычислительной практике получил другой аналитический метод, основанный на идее разложения в ряд решения рассматриваемой задачи Коши. Особенно часто для этих целей используют ряд Тейлора.

Метод степенных рядов

Приближенное решение $y_m(x)$ исходной задачи (1), (2) ищут в виде

$$\begin{aligned} y_m(x) &= \sum_{i=0}^{m} \frac{(x-x_0)^i}{i!} y^{(i)}(x_0) = \\ &= y(x_0) + \frac{y'(x_0)}{1!}h + \frac{y''(x_0)}{2!}h^2 + \ldots + \frac{y^m(x_0)}{m!}h^m, \end{aligned} \qquad (3)$$

где $h = x - x_0$. Здесь $y^{(0)}(x_0) = y(x_0) = y_0,\ y^{(1)}(x_0) = y'(x_0) = f(x_0, y_0)$, а значения $y^{(i)}(x_0), i = 2, 3, \dots, m$ находятся по формулам, полученным последовательным дифференцированием уравнения (1):

$$\begin{gathered} y^{(2)}(x_0) = y''(x_0) = f_x(x_0,y_0) + f(x_0, y_0)f_y(x_0, y_0), \\ y^{(3)}(x_0) = y'''(x_0) = f_{x^2}(x_0,y_0) + 2f(x_0, y_0)f_{xy}(x_0, y_0)+ \\ +f^2(x_0, y_0)f_{y^2}(x_0, y_0)+ \\ +f_y(x_0, y_0)[f_x(x_0, y_0) + f(x_0, y_0)f_y(x_0, y_0)], \ \dots \ . \end{gathered} \tag{4}$$

Для значений x близких к x_0, метод степенных рядов (3) при достаточно больших m дает обычно хорошее приближение к точному решению $y(x)$ задачи (1), (2). Однако, с увеличением расстояния $|x - x_0|$ погрешность приближенного равенства $y(x) \approx y_m(x)$, вообще говоря, возрастает по абсолютной величине и правило (3) становится вовсе неприемлемым, когда x выходит из области сходимости соответствующего (3) ряда Тейлора.

Предпочтительными в таких случая будут, например, численные методы решения задачи Коши, позволяющие в некоторых попарно близких друг другу фиксированных точках (узлах)

$$x_0 < x_1 < x_2 < \dots < x_n = X \tag{5}$$

последовательно находить значения $y_n \approx y_n(x)$, $n = 1, 2, \dots, N$, приближенного решения. Сходимость методов пособного типа не так жестко связана с длиной отрезка $[x_0, X]$, и их чаще кладут в основу стандартных программ для ПЭВМ.

§45. Понятие одношаговых и многошаговых методов

Большинство численных методов решения рассматриваемой задачи Коши можно привести к виду

$$y_{n+1} = F(y_{n-q}, y_{n-q+1}, \ldots, y_n, y_{n+1}, \ldots, y_{n+s}), \tag{1}$$

где F – некоторая известная функция указанных аргументов, определяемая способом построения метода и зависящая от вида уравнения (1) и избранной сетки (5) (из §44). При $q = 0,\ 0 \le s \le 1$ такие вычислительные методы обычно называют одношаговыми, а при $q \ge 1$ или $s > 1$ – многошаговыми. Как одношаговые, так и многошаговые методы вида (1) называют явными в случае $s = 0$ и неявными при $s = 1$. В случае $s > 1$ многошаговые правила часто называют методами с забеганием вперед.

Если правило (1) является одношаговым, то вычисления по этому правилу можно начинать со значения $n = 0$ и проводить до значения $n = N - 1$ включительно. В случае же многошаговых методов указанного вида дискретный аргумент n может пробегать значения лишь от $n = q$ до $n = N - s$, что, вообще говоря, влечет за собой нарушение однородности вычислительного процесса и требует применения специальных вычислительных правил для нахождения первых q значений $y_1, y_2, \ldots, y_q$ приближенного решения и последних его $s-1$ значений $y_{N-S+2}, y_{N-S+3}, \ldots, y_N$. Одношаговые методы в этом отношении являются предпочтительнее. Удоблее пользоваться ими и в том случае, когда шаг сетки $h_n = x_{n+1} - x_n$ не является постоянным для всех значений n. Вид одношаговых методов не связан с величиной шага на предыдущем этапе вычислительного процесса и поэтому эти правила легко допускают изменение шага численного интегрирования, в то время, как многошаговые методы нужно специальным образом приспосабливать для этих целей. Основной недостаток одношаговых методов связан с их трудоемкостью по сравнению с многошаговыми методами. В случае одношаговых методов информация о решаемой задаче используется лишь в пределах одного шага и при переходе от шага к шагу ее, вообще говоря, надо заново получать. В то время как в случае многошаговых методов предоставляется возможным повторно использовать части такой информации на нескольких соседних этапах вычислений, что позволяет уменьшить затраты вычислительного труда на каждый шаг численного интегрирования.

§46. Построение одношаговых методов способом разложения решение в ряд тейлора

Будем считать, что процесс решения задачи Коши (1), (2) (из §44) доведен до точки $x_n, (0 \le n \le N)$ и известно соответствующее приближение $y(x_n)$ искомого решения. Построим вычислительное правило для нахождения решения в очередной узловой точке $x_{n+1} = x_n + h_n$ сетки (5) (из §44).

Поскольку при построении одношаговых методов используется информация о решаемой задаче лишь в пределах одного шага интегрирования, то можно без ущерба для понимания не писать индекс n, обозначающий номер процесса. Итак, нужно по известному значению $y(x)$ в узловой точке $x \ge x_0$ найти значение этого решения в очередной точке $x + h$. Воспользуемся вычислительным правилом типа (3) (из §44).

$$y_m(x+h) \approx \sum_{i=0}^{m} \frac{h^i}{i!} y^{(i)}(x), \tag{1}$$

которое может быть положено в основу вычислительного одношагового метода, если для вычисления значений $y^{(i)}(x)$, $i = 2, 3, \ldots, m$ использовать формулы типа (3) (из §44). При условии, что данное решение уравнения (1) имеет на рассматриваемом отрезке непрерывную производную порядка $m+1$, погрешность приближенного равенства (1) будет, очевидно, величиной порядка h^{m+1} и при малых h и больших m построенный метод будет давать, как правило, достаточно хорошее приближение к искомому значению решения.

Однако такой одношаговый метод интегрирования дифференциальных уравнений при $m > 1$ все же редко используется в практике вычислений, т.к. его применение требует на каждом шаге нахождения $\dfrac{m(m+1)}{2}$ различных производных $f, f_x, f_y, f_{x^2}, f_{xy}, \ldots, f_{y^{m-1}}$. Естественно поставить задачу о таком усовершенствовании приведенного выше одношагового метода, которое сохраняло бы основные его достоинства, но не было связано с нахождением значений производных правой части уравнения (1). Чтобы выполнить последнее условие, производные $y^{(i)}(x)$, $i = 2, 3, \ldots, m$, входящие в правую часть равенства , можно заменить по формулам численного дифференцирования их приближенными выражениями через значения функции y' и учесть, что $y' = f[x, y(x)]$. Ниже на конкретных примерах будет рассмотрен один из возможных подходов к решению поставленной задачи.

Случай $m = 1$. Тогда приближенное равенство (1) не требует вычисления производных правой части уравнений и позволяет с погрешностью порядка h^2 находить значение $y(x_n + h)$ решения этого уравнения по известному его значению $y(x_n)$. Соответствующее одношаговое правило можно записать в виде

$$y_{n+1} = y_n + hf_n, \tag{2}$$

использовав широко применяемое ниже обозначение $f(x_n + \alpha h, y_{n+\alpha}^{[k]}) =$

$= f^{[k]}_{n+\alpha}$, где $y^{[k]}_{n+\alpha}$ – найденное с ошибкой порядка h^k приближенное значение решения в точке $x_n + \alpha h$. Здесь $f_n = f(x_n, y_n)$.

Правило (2) впервые построено Эйлером и носит его имя. Иногда его называют правилом ломаных или методом касательных, чем подчеркивают простой геометрический смысл формулы. Погрешность r_{n+1} этой формулы можно, очевидно, записать в виде $r_{n+1} = \dfrac{h^2}{2} y''(x_n + \Theta h), 0 <$ $< \Theta < 1$.

Случай $m = 2$. Здесь равенство (1) требует вычисления производной $y''(x_n + h)$ с локальной ошибкой порядка h^3. Чтобы не понизить порядок погрешности приближенного равенства

$$y(x_n + h) \approx y(x_n) + hy'(x_n) + \frac{h^2}{2} y''(x_n), \tag{3}$$

значение производной $y''(x_n)$ необходимо найти по крайней мере не хуже, чем с ошибкой порядка h для чего, очевидно, достаточно иметь два значения функции $y'(x) = f[x, y(x)]$ на отрезке $x_n \le x \le x_n + h, 0 < h < 1$.

В точке x_n значением $f^{[3]}_n$ этой функции мы уже располагаем, так как, по предположению, на предыдущем шаге было найдено $y^{[3]}_n$. Найдем теперь еще значение функции y' в точке $x_n + h^i, i \ge 1$. Для этого с учетом уравнения (1) достаточно указать правило вычисления $y(x_n + h^i)$. Очевидно, $y(x_n + h^i) = y(x_n) + h^i y'(x_n) + O(h^{2i})$.

Поэтому справедлива следующая расчетная формула

$$y^{[2i]}_{n+h^{i-1}} = y^{[3]}_n + h^i f^{[3]}_n. \tag{4}$$

Так как

$$y''(x_n) = \frac{y'(x_n + h^i) - y'(x_n)}{h^i} + O(h^i), y'(x_n + h^i) = f[x_n + h^i, y(x_n, h^i)],$$

то на основании (3) можно записать

$$\begin{aligned} y^{[3]}_{n+1} &= y^{[3]}_n + hf^{[3]}_n + \frac{h^{2-i}}{2}(f^{[2i]}_{n+h^{i-1}} - f^{[3]}_n) = \\ &= y^{[3]}_n + h(1-\frac{h^{i-1}}{2})f^{[3]}_n + \frac{h^{2-i}}{2} f^{[2i]}_{n+h^{i-1}}. \end{aligned} \tag{5}$$

Формулы (4), (5) можно рассматривать как семейство (зависящих от параметра $i \geq 1$) одношаговых методов решения задачи Коши (1), (2)(из §44) с локальной погрешностью порядка h^3.

При $i = 1$ эти формулы принимают вид

$$y^{[2]}_{n+1} = y^{[3]}_n + hf^{[3]}_n, \tag{6}$$

$$y^{[3]}_{n+1} = y^{[3]}_n + \frac{h}{2}(f^{[3]}_n + f^{[2]}_{n+1}). \tag{7}$$

Они имеют предсказывающе-исправляющий характер: формула (6) служит для получения грубого приближения искомой величины $y(x_n + h)$, а по формуле (7) происходит уточнение полученного значения.

В случае $i = 2$ формулы (4), (5) имеют вид

$$y_{n+h}^{[4]} = y_n^{[3]} + h^2 f_n^{[3]}, \tag{8}$$

$$y_{n+1}^{[3]} = y_n^{[3]} + h f_n^{[3]} + \frac{1}{2}(f_{n+h}^{[4]} - f_n^{[3]}) = y_n^{[3]} + (h - \frac{1}{2}) f_n^{[3]} + \frac{1}{2} f_{n+h}^{[4]}. \tag{9}$$

Увеличение значения i на единицу позволило несколько улучшить структуру остаточного члена. Если в случае правила (6), (7) погрешность складывалась из одного слагаемого вида $\frac{h^3}{6} y'''(x_n + \Theta h), 0 < \Theta < 1$, представляющего собой ошибку приближенного равенства (3), и еще двух слагаемых также порядка h^3, порождаемых неточностью замены производной y'' и ошибкой формулы Эйлера (4), то в случае (8), (9), два последних слагаемых остаточного члена будут уже, очевидно, величинами порядка h^4.

При $i = 3$ соответствующие слагаемые станут величинами порядка h^5 и т.д.

Случай $m = 3$. Соответствующее разложение решения по формуле Тейлора будет сейчас обрываться на члене, содержащем вторую производную функции y'. Чтобы построить приближенное выражение для этой производной через значений функции, нужно иметь по меньшей мере три таких значения. Возьмем в качестве них $y'(x_n), y'(x_n + h^i)$ и $y'(x_n + 2h^i)$, где $i \geq 2$ при $0 < h \leq \frac{1}{2}$. Чтобы воспользоваться простейшими симметричными приближенными выражениями для первой и второй производной функции y' через эти значения, будем исходить из разложений решения по формуле Тейлора не около точки x_n, а около точки $x_n + h^i$:

$$y(x_n+h) = y[x_n+h^i+h(1-h^{i-1})] = y(x_n+h^i)+h(1-h^{i-1})y'(x_n+h^i)+$$

$$+\frac{h^2(1-h^{i-1})^2}{2}y''(x_n+h^i)+\frac{h^3(1-h^{i-1})^3}{6}y'''(x_n+h^i)+O(h^4).$$

Опираясь на это представление можно также, как и для $m = 2$, выписать семейство вычислительных правил.

При $i = 2$, например, имеем расчетное правило

$$y^{[4]}_{n+h} = y^{[4]}_n + h^2 f^{[4]}_n,$$

$$y^{[6]}_{n+h} = y^{[4]}_n + \frac{h^2}{2}(f^{[4]}_n - f^{[4]}_{n+h}),$$

$$y^{[6]}_{n+2h} = y^{[6]}_{n+h} + \frac{h^2}{2}(3f^{[6]}_{n+h} - f^{[4]}_n),$$

$$y^{[4]}_{n+1} = y^{[6]}_{n+h} + \frac{(1-h)^2}{2}(\frac{1-h}{3h} - \frac{1}{2})f^{[4]}_n+$$

$$+(1-h)\left[h - \frac{(1-h)^2}{3h}\right]f^{[6]}_{n+h} + \frac{(1-h)^2}{2}\left[\frac{1-h}{3h}+\frac{1}{2}\right]f^{[6]}_{n+2h}.$$

Наиболее часто на практике используют следующие методы:

1. *Метод Эйлера*

$$y_{n+1} = y_n + hf(x_n, y_n);$$

2. *Модифицированный метод Эйлера*

$$y_{n+\frac{1}{2}} = y_n + \frac{h}{2}f(x_n, y_n), y_{n+1} = y_n + hf\left(x_n + \frac{h}{2}, y_{n+\frac{1}{2}}\right);$$

3. *Метод Эйлера-Коши*

$$\tilde{y}_{n+1} = y_n + hf(x_n, y_n), \tilde{f}_{n+1} = f(x_{n+1}, \tilde{y}_{n+1}), y_{n+1} = y_n + h\frac{f_n + \tilde{f}_{n+1}}{2}.$$

Достоинства метода Эйлера:

1. простота;
2. малый объем вычислений;
3. наглядность.

Недостатки:

1. малая точность;
2. работает на малых интервалах.

Он дает весьма грубое приближение решения задачи Коши. Он обычно используется, когда требуется получить примерное представление о решении на небольшом промежутке.

§47. Методы типа Рунге-Кутта

Для удобства записи используем обозначение $\Delta y = y(x+h) - y(x)$, тогда равенство (1) (из §46) запишется

$$\Delta y = h \int\limits_x^{x+h} f(t, y(t))dt = (t = x + \alpha h) = h \int\limits_0^1 f[x+\alpha h, y(x+\alpha h)]d\alpha. \quad (1)$$

Чтобы на основе (1) построить одношаговый метод численного интегрирования уравнения (1) (из §46) введем три набора параметров:

$\alpha_1, \alpha_2, \ldots, \alpha_q;$ (α)

$\beta_{10},$

$\beta_{20}, \beta_{21},$ (β)

$\ldots$

$\beta_{q0}, \beta_{q1}, \ldots, \beta_{q,q-1};$

$A_0, A_1, \ldots, A_q,$ (A)

выбором которых распорядимся в дальнейшем. При помощи двух первых наборов составим величины

$\varphi_0 = hf(x, y),$

$\varphi_1 = hf(x + \alpha_1 h, y + \beta_{10}\varphi_0),$

$\varphi_2 = hf(x + \alpha_2 h, y + \beta_{20}\varphi_0 + \beta_{21}\varphi_1),$

$\ldots$

$$\varphi_q = hf(x+\alpha_q h, y+\beta_{q0}\varphi_0+\beta_{q1}\varphi_1+\cdots+\beta_{q,q-1}\varphi_{q-1}),$$

которые при заданных (α) и (β) могут быть вычислены последовательно.

Хотя $\varphi_i = hf(x+\alpha_i h, y+\beta_{i0}\varphi_0+\cdots+\beta_{i,i-1}\varphi_{i-1}), i = 1, 2, \ldots, q$ вообще говоря, не равны значениям $hf[x+\alpha_i h, y(x+\alpha_i h)]$, однако при соответствующем выборе параметров (β) их можно трактовать как приближенные значения интегрируемой функции $f[x+\alpha h, y(x+\alpha h)]$, умноженные на h. Это дает основание надеяться при помощи параметров (A) составить такую линейную комбинацию величин $\varphi_i, i = 0, 1, \ldots, q$, которая будет являться аналогом квадратурной суммы и позволит вычислить приближенное значение приращения Δy:

$$\Delta y \approx \sum_{i=0}^{q} A_i\varphi_i. \tag{2}$$

Тем самым параметрам (α), (β) и (A) можно придать некоторый квадратурный смысл. Рассмотрим теперь задачу выбора этих параметров. Введем величину

$$r_q(h) = \Delta y - \sum_{i=0}^{q} A_i\varphi_i, \tag{3}$$

представляющую собой погрешность приближенного равенства (2). В предположении, что правая часть уравнения (8) (из §46) является достаточно гладкой функцией, запишем следующее разложение этой величины

$$r_q(h) = \sum_{j=0}^{k} \frac{h^j}{j!} r_q^{(j)}(0) + \frac{h^{k+1}}{(k+1)!} r_q^{k+1}(\Theta h), 0 < \Theta < 1.$$

На основании этого разложения можно утверждать, что если параметры (α), (β) и (A) подобрать так, чтобы выполнялись условия

$$r_q^{(j)}(0) = 0, j = 0, 1, \ldots, k, \tag{4}$$

то погрешность (3) приближенного равенства (2) будет величиной порядка не ниже h^{k+1}:

$$r_q(h) = \frac{h^{k+1}}{(k+1)!} r_q^{(k+1)}(\Theta h). \tag{5}$$

Число k при этом обычно называют порядком точности соответствующего метода.

Для выполнения условий (4) при возможно большем значении k величины $r_q^{(j)} = 0, j = 0, 1, \ldots, k$, выражают через значения функции $f(x, y)$ и ее частных производных и требуют обращения в нуль возможно большего числа этих величин для любой достаточно гладкой функции f. Иными словами, выбор параметров (α), (β), (A) осуществляется на основании требования, чтобы разложение

$$\Delta y = y(x+h) - y(x) = \frac{h}{1!} y'(x) + \frac{h^2}{2!} y''(x) + \frac{h^3}{3!} y'''(x) + \ldots \ . \tag{6}$$

и разложение по степеням h линейной комбинации $\sum_{i=0}^{q} A_i\varphi_i$ совпадали до членов с возможно большими степенями h в случае любой правой части уравнения (2).

Метод первого порядка точности.

Пусть $q = 0$, что равнозначно введению лишь одного параметра A_0. Приближенное равенство (2) в этом случае примет вид $\Delta y \approx A_0\varphi_0 = = hA_0 f(x, y)$ и погрешность (3) может быть записана в виде

$$r_0(h) = y(x + h) - y(x) - hA_0 f(x, y).$$

Тогда $r_0'(h) = y'(x + h) - A_0 f(x, y), r_0''(h) = y''(x + h)$.

Так как $r_0''(h)$ не зависит от A_0, то уже при $j = 2$ условие (4) в случае произвольной функции f удовлетворено быть не может. Поэтому $k = 1$ и система (4) принимает вид $(1 - A_0)f(x, y) = 0$. Отсюда находим, что $A_0 = 1$. Следовательно, из (2) $\Delta y = hf(x, y)$ и из (5)

$$r_0(h) = \frac{h^2}{2} r_0''(\Theta h) = \frac{h^2}{2} y''(x + \Theta h), 0 < \Theta < 1.$$

В простейшем случае $q = 0$, таким образом, метод Рунге-Кутта приводит к известному методу Эйлера $y_{n+1} = y_n + hf(x_n, y_n)$.

Методы второго порядка точности.

При $q = 1$ имеем

$$\Delta y = A_0\varphi_0 + A_1\varphi_1 = hA_0 f(x, y) + hA_1 f[x + \alpha_1 h, y + h\beta_{10} f(x, y)].$$

С целью выбора введенных параметров $\alpha_1, \beta_{10}, A_0, A_1$ разложим Δy и $A_0\varphi_0 + A_1\varphi_1$ по степеням h.

Разложение (6) для Δy с учетом (1) (из §46) можно записать в виде

$$\Delta y = y(x+h) - y(x) = hf + \frac{h^2}{2}(f_x + f \cdot f_y) + \\ + \frac{h^3}{6}[f_{xx} + 2f \cdot f_{xy} + f^2 f_{yy} + f_y(f_x + f \cdot f_y)] + O(h^4). \tag{7}$$

Используя формулу Тейлора, для линейной комбинации $A_0\varphi_0 + A_1\varphi_1$ можно дать следующее представление:

$$A_0\varphi_0 + A_1\varphi_1 = h(A_0 + A_1)f + h^2 A_1(\alpha_1 f_x + \beta_{10} f \cdot f_y) + \\ + \frac{h^3}{2} A_1(\alpha_1^2 f_{xx} + 2\alpha_1\beta_{10} f \cdot f_{xy} + \beta_{10}^2 f^2 \cdot f_{yy}) + O(h^4). \tag{8}$$

Сравним в разложениях (7) и (8) коэффициенты при $hf, h^2 f_x, h^2 f f_y$. Тем самым на выбор четырех параметров $\alpha_1, \beta_{10}, A_0, A_1$ будут наложены три условия: $A_0 + A_1 = 1, A_1\alpha_1 = \dfrac{1}{2}, A_1\beta_{10} = \dfrac{1}{2}$.

Непосредственно из (7), (8) следует, что в случае $q = 1$ для произвольных f нельзя добиться совпадения всех членов с множителем h^3 за счет выбора введенных параметров. Поэтому при $q = 1$ вычислительные правила Рунге-Кутта будут иметь лишь второй порядок точности. Параметры $\alpha_1, \beta_{10}, A_0$ могут быть выражены через A_1 по формулам $\alpha_1 = \beta_{10} = \dfrac{1}{2A_1}, A_0 = 1 - A_1$.

В качестве A_1 может быть взято, вообще говоря, произвольное отличное от нуля число. Например, при $A_1 = \frac{1}{2}$ будем иметь

$$\Delta y = \frac{1}{2}(\varphi_0 + \varphi_1) + O(h^3),$$

$$\varphi_0 = hf(x, y), \varphi_1 = hf(x + h, y + \varphi_0). \tag{9}$$

Вычислительное правило, построенное на основе равенство (9) имеет следующий квадратурный смысл.

Так при $\varphi_1 = hf(x + h, y + hy') \approx hf[x + h, y(x + h)]$, то линейная комбинация $\frac{1}{2}(\varphi_0 + \varphi_1)$ является аналогом квадратурной суммы формулы трапеций при вычислении интеграла $\int\limits_{x_n}^{x_{n+1}} f(t, y(t))dt$ в правой части равенства (1) (из §46).

Выбрав $A_1 = 1$ получим еще одно из широко известных вычислительных правил типа Рунге-Кутта второго порядка точности $(A_0 = 0)$

$$\Delta y \approx \varphi_1,$$

$$\varphi_0 = hf(x, y), \varphi_1 = hf\left(x + \frac{h}{2}, y + \frac{\varphi_0}{2}\right). \tag{10}$$

Формула (10) является, очевидно, аналогом одноточечной квадратурной формулы средних прямоугольников.

Локальная погрешность любого из методов Рунге-Кутта второго порядка точности, как следует из (7), (8), может быть представлена в виде

$$r_1(h) = \frac{h^3}{6}[f_{xx}(1-3\alpha_1^2 A_1)+2f\cdot f_{xy}(1-3\alpha_1\beta_{10}A_1)+ \\ +f^2\cdot f_{yy}(1-3\beta_{10}^2 A_1)+f_y(f_x+f\cdot f_y)]+O(h^4). \qquad (11)$$

Иногда свободный параметр A_1 выбирают так, чтобы в этом представлении можно было обратить в нуль хотя бы часть членов. Например, если учесть, что $\alpha_1 = \beta_{10} = \dfrac{1}{2A_1}$ и положить $A_1 = \dfrac{3}{4}$, то правая часть равенства (11) существенно упроститься

$$r_1(h) = \frac{h^3}{6}f_y(f_x+f\cdot f_y)+O(h^4)$$

При таком выборе A_1 будем иметь

$$\Delta y \approx \frac{1}{4}(\varphi_0+3\varphi_1),$$

$$\varphi_0 = hf(x,y), \varphi_1 = hf\left(x+\frac{2}{3}h, y+\frac{2}{3}\varphi_0\right). \qquad (12)$$

Методы третьего порядка точности.

С повышением требований к точности вычислительных правил Рунге-Кутта очень быстро возрастает громоздкость необходимых построений, хотя общая схема таких построений и не претерпевает существенных изменений. Поэтому в случае $q = 2$ мы не станем производить все выкладки и напишем лишь ту систему уравнений, которым должны удовлетворять параметры (α), (β), и (A) в методах типа Рунге-Кутта третьего порядка точности:

$$A_0 + A_1 + A_2 = 1; A_1\alpha_1 + A_2\alpha_2 = \frac{1}{2}; A_1\alpha_1^2 + A_2\alpha_2^2 = \frac{1}{3};$$

$$A_2\alpha_1\beta_{21} = \frac{1}{6}; \beta_{20} + \beta_{21} = \alpha^2; \beta_{10} = \alpha_1.$$

Одно из решений этой системы шести уравнений с восьмью неизвестными приводит к следующим формулам:

$$\Delta y \approx \frac{1}{6}(\varphi_0 + 4\varphi_1 + \varphi_2),$$

$$\varphi_0 = hf(x, y), \varphi_1 = hf\left(x + \frac{h}{2}, y + \frac{\varphi_0}{2}\right), \varphi_2 = hf(x + h, y - \varphi_0 + 2\varphi_1). \tag{13}$$

Это вычислительное правило является аналогом квадратурной формулы Симпсона.

Часто встречается в практике вычислений и такой метод типа Рунге-Кутта третьего порядка точности:

$$\Delta y \approx \frac{1}{4}(\varphi_0 + 3\varphi_2),$$

$$\varphi_0 = hf(x,y), \varphi_1 = hf\left(x+\frac{h}{3}, y+\frac{\varphi_0}{3}\right), \varphi_2 = hf\left(x+\frac{2}{3}h, y\frac{2}{3}\varphi_1\right).$$

Методы четвертого порядка точности.

В случае $q = 3$ система ограничений на выбор параметров (α), (β), (A) вычислительных правил Рунге-Кутта четвертого порядка точности может быть приведена к виду

$$A_0 + A_1 + A_2 + A_3 = 1; A_1\alpha_1 + A_2\alpha_2 + A_3\alpha_3 = \frac{1}{2}; A_1\alpha_1^2 + A_2\alpha_2^2 + A_3\alpha_3^2 = \frac{1}{3};$$

$$A_1\alpha_1^3 + A_2\alpha_2^3 + A_3\alpha_3^3 = \frac{1}{4};$$

$$A_2\beta_{21}\alpha_1 + A_3\beta_{31}\alpha_1 + A_3\beta_{32}\alpha_2 = \frac{1}{6};$$

$$A_2\beta_{21}\alpha_2\alpha_1 + A_3\beta_{31}\alpha_3\alpha_1 + A_3\beta_{32}\alpha_3\alpha_2 = \frac{1}{8};$$

$$A_2\beta_{21}\alpha_1^2 + A_3\beta_{31}\alpha_1^2 + A_3\beta_{32}\alpha_2^2 = \frac{1}{12};$$

$$A_3\beta_{32}\beta_{21}\alpha_1 = \frac{1}{24};$$

$$\beta_{30} + \beta_{31} + \beta_{32} = \alpha_3; \beta_{20} + \beta_{21} = \alpha_2; \beta_{10} = \alpha_1.$$

Одним из методов, удовлетворяющим этим требованиям, является следующий аналог четырехточечной квадратурной формулы «трех восьмых»:

$$\Delta y \approx \frac{1}{8}(\varphi_0 + 3\varphi_1 + 3\varphi_2 + \varphi_3),$$

$$\varphi_0 = hf(x,y), \varphi_1 = hf\left(x + \frac{h}{3}, y + \frac{\varphi_0}{3}\right), \varphi_2 = hf\left(x + \frac{2}{3}h, y - \frac{\varphi_0}{3} + \varphi_1\right),$$
$$\varphi_3 = hf\left(x + h, y + \varphi_0 - \varphi_1 + \varphi_2\right).$$

Особенно широко известно другое вычислительное правило типа Рунге-Кутта четвертого порядка точности

$$\Delta y \approx \frac{1}{6}[\varphi_0 + 2\varphi_1 + 2\varphi_2 + \varphi_3],$$

$$\varphi_0 = hf(x,y), \varphi_1 = hf\left(x + \frac{h}{2}, y + \frac{\varphi_0}{2}\right),$$

$$\varphi_2 = hf\left(x + \frac{h}{2}, y + \frac{\varphi_1}{2}\right), \varphi_3 = hf(x + h, y + \varphi_2). \qquad (14)$$

Можно также построить правило Рунге-Кутта пятого порядка точности, если взять $q > 3$. Дальнейшее повышение порядка точности методов типа Рунге-Кутта связано, как правило, с быстрым возрастанием их трудоемкости. Однако нужно учесть, что методы более высокого порядка точности обычно допускают использование большего шага h, что может снизить общие вычислительные затраты.

§48. Случай уравнений высших порядков

Изложенные выше способы построения вычислительных методов решения задачи Коши (1), (2) (из §44) легко могут быть обобщены на случай уравнений высших порядков.

Рассмотрим задачу Коши для уравнения второго порядка

$$y'' = f(x, y, y'), \tag{1}$$

$$y(x_0) = y_0, y'(x_0) = y_0'. \tag{2}$$

Подобно (1), (2), здесь можно записать, что

$$\begin{gathered} y'(x_n + h) = y'(x_n) + h\int_0^1 y''(x_n + \alpha h)d\alpha \approx y'(x_n)+ \\ +h\sum_{i=0}^{q_1} A_i f[x_n + \alpha_i h, y(x_n + \alpha_i h), y'(x_n + \alpha_i h)]. \end{gathered} \tag{3}$$

И аналогично

$$\begin{gathered} y(x_n + h) = y(x_n) + hy'(x_n) + h^2\int_0^1 (1-\beta)y''(x_n + \beta h)d\beta \approx \\ \approx y(x_n) + hy'(x_n) + h^2\sum_{i=0}^{q_0} B_i f[x_n + \beta_i h, y(x_n + \beta_i h), y'(x_n + \beta_i h)]. \end{gathered} \tag{4}$$

Подберем параметры $A_i, \alpha_i, i = 0, 1, \dots, q_1, B_i, \beta_i, i = 0, 1, \dots, q_0$ этих приближенных равенств так, чтобы их погрешности были величинами порядка $h^{k+1}(1 < k < min\{2q_1 + 2, 2q_0 + 3\})$. Эти требования приводят к следующим системам, вообще говоря, нелинейных уравнений относительно искомых параметров:

$$\sum_{i=0}^{q_1} A_i = 1; \sum_{i=0}^{q_1} A_i \alpha_i^j = \frac{1}{j+1}, j = 1, 2, \dots, k-1,$$

$$\sum_{i=0}^{q_0} B_i = \frac{1}{2}; \sum_{i=0}^{q_0} B_i \beta_i^{j-1} = \frac{1}{j(j+1)}, j = 2, 3, \dots, k-1. \tag{5}$$

Подобрав подходящие значения параметров, мы построим приближенные формулы (13), (14) (из §47). Процесс дальнейшего построения вычислительных правил совершенно аналогичен случаю задачи (1) и (2) (из §44).

§49. Оценка погрешности (сходимость) одношаговых методов

Рассмотрим множество всевозможных методов интегрирования задачи $y' = f(x, y)$, $y(x_0) = y_0$, где последовательно получаются приближения y_j к значениям $y(x_j), x_0 < x_1 < \cdots < x_N = x_0 + X$.

Пусть в процессе численного интегрирования k фиксировано и при $j \geq k$ значение y_j определяется как значение некоторого функционала

$$y_j = F(f; x_j, \ldots, x_{j-k}, y_{j-1}, \ldots, y_{j-k}). \tag{1}$$

Такой способ численного интегрирования называется k-шаговым. Все рассмотренные выше методы одношаговые, для них $k = 1$. При $k = 1$ иногда удобнее записать формулу (1) в виде

$$y_{j+1} = F(f, x_j, x_{j+1} - x_j, y_j). \tag{2}$$

Получаемые в процессе реальных вычислений приближения к значениям $y(x_j)$ связаны не соотношениями (2), а некоторыми соотношениями

$$y_{j+1} = F(f, x_j, x_{j+1} - x_j, y_j) + \delta_{(j+1)}. \tag{3}$$

Наличие слагаемого δ_{j+1} обусловлено следующими причинами:

1. округлением чисел при вычислениях;

2. ошибками в значениях правой части $f(x, y)$; эти ошибки связаны с тем, что рассматриваемая нами функция $f(x, y)$ является некоторым приближением к правой части реального дифференциального уравнения;

3. в некоторых случаях значение y_{j+1} определяется из уравнения, эквивалентного (1), но неразрешенного относительно переменной y_{j+1}; тогда величина δ_{j+1} содержит составляющую, являющуюся следствием приближенного решения этого уравнения.

Довольно часто бывает справедлива оценка:

$$|\delta_{j+1}| \leq C(|y_j| + 1)2^{-t},$$

где t – разрядность чисел в машине, а постоянная C не очень большая.

Точно также начальное условие y_0 отличается от значения отыскиваемого решения задачи $y(x_0)$ вследствие ошибки в определении исходных данных и округлений.

Пусть $y(x)$ – искомое решение дифференциального уравнения, а $y_j(x)$ – решение, удовлетворяющее условиям $y_j(x_j) = y_j$. Ошибку $R_n = y_n - y(x_n)$ можно представить в виде

$$\begin{aligned} R_n &= (y_n(x_n) - y_0(x_n)) + (y_0(x_n) - y(x_n)) = \\ &= \sum_{j=1}^{n} (y_j(x_n) - y_{j-1}(x_n)) + (y_0(x_n) - y(x_n)). \end{aligned} \tag{4}$$

Разность решений дифференциального уравнения в одной точке может быть выражена через разность их в другой точке следующим образом:

Лемма 1. *Пусть $Y_1(x)$ и $Y_2(x)$ – решения дифференциального уравнения $y' = f(x, y)$. Тогда*

$$Y_2(\beta) - Y_1(\beta) = (Y_2(\alpha) - Y_1(\alpha)) \exp\left(\int_\alpha^\beta f_y(x, \tilde{y}(x))dx\right). \tag{5}$$

где $\tilde{y}(x)$ заключено между $Y_1(x)$ и $Y_2(x)$.

Доказательство. Вычтем друг из друга равенства $Y_2' = f(x, Y_2)$, $Y_1' = f(x, Y_1)$. Согласно формуле Лагранжа, разность $f(x, Y_2) - f(x, Y_1)$ может быть представлена в виде $f_y = f(x, \tilde{y})(Y_2 - Y_1)$, где $\tilde{y}$ заключено между Y_1 и Y_2.

В результате получается линейное дифференциальное уравнение относительно $Y_2 - Y_1$: $(Y_2 - Y_1)' = f_y(x, \tilde{y})(Y_2 - Y_1)$, откуда $(ln(Y_2 - Y_1))' = f_y(x, \tilde{y})$. Подставляя пределы α и β и потенцируя, получим

$$Y_2(\beta) - Y_1(\beta) = (Y_2(\alpha) - Y_1(\alpha))e^{\int\limits_\alpha^\beta f_y(x,\tilde{y}(x))dx}.$$

Лемма доказана.

Пусть $\alpha = x_j$, $\beta = x_n$, $Y_1(x) = y_{j-1}(x)$, $Y_2(x) = y_j(x)$, тогда

$$y_j(x_n) - y_{j-1}(x_n) = (y_j(x_j) - y_{j-1}(x_j))e^{\int\limits_{x_j}^{x_n} f_y(x,\tilde{y}_j(x))dx},$$

где $\tilde{y}_j(x)$ заключено между $y_{j-1}(x)$ и $y_j(x)$.

Точно также $y_0(x_n) - y(x_n) = (y_0(x_0) - y(x_0))e^{\int\limits_{x_0}^{x_n} f_y(x,\tilde{y}_0(x))dx}$.

Теперь равенство (4) можно записать в виде

$$R_n = \sum_{j=1}^{n} \omega_j \exp \int\limits_{x_j}^{x_n} f_y(x, \tilde{y}_j(x))dx + R_0 \exp\left(\int_{x_0}^{x_n} f_y(x, \tilde{y}_0(x))dx\right), \qquad (6)$$

где $\omega_j = y_j(x_j) - y_{j-1}(x_j), j = 1, 2, \ldots ;\ R_0 = y_0(x_0) - y(x_0)$.

Из (3) вытекает соотношение $\omega_j = y_l - y_{j-1}(x_j) = \rho_j + \delta_j$, где $\rho_j =$ $= F(f, x_{j-1}, x_j - x_{j-1}, y_{j-1}) - y_{j-1}(x_j)$.

Посмотрим, какой смысл имеет величина ρ_j. $F(f, x_{j-1}, x_j - x_{j-1}, y_{j-1})$ есть число, получаемое по расчетной формуле (2), а $y_{j-1}(x_j)$ – значение в точке x_j точного решения дифференциального уравнения, удовлетворяющее условию $y_{j-1}(x_{j-1}) = y_{j-1}$. Таким образом, ρ_j есть ошибка одного шага рассматриваемого метода, если вычисления начинать с точки (x_{j-1}, y_{j-1}) и производить без округлений, а шаг равен $x_j - x_{j-1}$.

Предположим, что при всех j, соответствующих рассматриваемому отрезку интегрирования $x_0 < x_j \leq x_0 + X$, выполняется неравенство

$$|\rho_j| \leq C_1(x_j - x_{j-1})^{s+1}. \qquad (7)$$

Пусть $L = \sup\limits_{x_0 \leq x \leq z_0 + X} |f_y| < \infty$, $H = \max\limits_{0<j\leq N} |x_j - x_{j-1}|$. Тогда из (7) имеем

$$|\rho_j| \leq C_1 H^s (x_j - x_{j-1}). \qquad (8)$$

При $x_0 \leq x_j \leq x_n \leq x_0 + X$ справедливы неравенства

$$\exp\left(\int\limits_{x_j}^{x_n} f_y(x, \tilde{y}_j(x))dx\right) \leq \exp(L(x_n - x_j)) \leq \exp(LX).$$

Воспользуемся этим неравенством для оценки правой части (6), получим $|R_n| \leq \exp(LX)\left(\sum_{j=1}^{n}(|\rho_j| + |\delta_j|) + |R_0|\right)$.

Применим теперь оценку (8):

$$|R_n| \leq \exp(LX)(\sum_{j=1}^{n}(C_1 H^s(x_j - x_{j-1}) + |\delta_j|) + |R_0|) \leq \exp(LX)\times \\ \times(C_1 H^s(x_n - x_0) + n\delta + |R_0|) \leq \exp(LX)(C_1 H^s X + N\delta + |R_0|) \quad (9)$$

Здесь $\delta = \max\limits_{j} |\delta_j|$. Из этого соотношения следует, что $\max\limits_{x_0 < x_n \leq z_0 + X} |R_n| \to$ $\to 0$ при $H \to 0$, если одновременно $N\delta \to 0$, $|R_0| \to 0$.

Мы показали, что при достаточно мелком шаге интегрирования и малой вычислительной погрешности приближенное решение, получаемое по методу Рунге-Кутта, будет близко к точному решению.

В ряде случаев решение дифференциального уравнения отыскивается на большом промежутке. Тогда в полученные нами оценки погрешности входит как множитель очень большое число $\exp(LX)$. Поэтому при LX

большом может оказаться, что достижение требуемой точности требует столь мелких шагов и столь малой величины вычислительной погрешности на шаге, что это может поставить под сомнение целесообразность использования рассматриваемого метода.

Если $f_y(x,y) \le -b < 0$, то в оценке (9) можно избавиться от множителя, резко растущего с X.

Рассмотрим случай постоянного шага $x_j - x_{j-1} = h$. Тогда

$$\exp\left(\int\limits_{x_j}^{x_n} f_y(x,\tilde{y}(x))dx\right) \le \exp(-b(x_n - x_j)) \le \exp(-b(n-j)h).$$

Пусть $|\omega_j| \le C_1 h^{s+1} + \delta$. Оценивая правую часть (6), получим

$$|R_n| \le \sum_{j=0}^{n} (C_1 h^{s+1} + \delta)\exp(-b(n-j)h) + |R_0|\exp(-bnh). \tag{10}$$

Имеем $\sum\limits_{j=0}^{n} \exp(-b(n-j)h) \le \sum\limits_{k=0}^{\infty} \exp(-bkh) = \dfrac{1}{1-\exp(-bh)}$. (бесконечно убывающая геометрическая прогрессия).

Итак, получаем окончательную оценку погрешности

$$R_n \le \frac{C_1 h^{s+1} + \delta}{1-\exp(-bh)} + |R_0|\exp(-bnh). \tag{11}$$

Поскольку $1-\exp(-bh) \sim |b|h$, то $R_n \leq C_2 \left(h^s + \dfrac{\delta}{h}\right) + |R_0| \exp(-bnh)$.

§50. Правило Рунге-Кутта

Пусть $y(x)$ – точное решение уравнения $y' = f(x, y)$, $y(x_0) = y_0$. При численном его решении пошаговым методом происходит постепенное накопление ошибки. Если y_0 задано точно, то $\tilde{y}_1$ получается с некоторой ошибкой, которая имеет порядок h^{s+1} (где s - порядок метода Рунге-Кутта; $s = r$ для формул Рунге-Кутта с $r = 1, 2, 3$ или 4). Отклонение $\tilde{y}_2$ от y_2 является результатом ошибки в определении $\tilde{y}_2$ и ошибки в определении $\tilde{y}_1$. Можно показать, что для k-го шага верно

$$\tilde{y}_k - y(x_k) = p(x_k)h^s + o(|h^s|). \tag{1}$$

При малых h и $p(x) \neq 0$ первое слагаемое в правой части равенства (1) преобладает, оно называется главным членом погрешности и полагают, если $p(x) \neq 0$,

$$\tilde{y}_k - y(x_k) \approx p(x_k)h^s. \tag{2}$$

На этом соотношении основан метод Рунге приближенной оценки точности численного решения уравнения $y' = f(x, y)$. Он состоит в следующем.

Пусть $\tilde{y}^h(x)$ и $\tilde{y}^{2h}(x)$ – приближенные решения уравнения $y' = f(x, y)$, полученные по методу Рунге-Кутта, с шагом h и $2h$ соответственно. Из условия (2)

$$\tilde{y}^h(x) - y(x) \approx p(x)h^s,$$
$$\tilde{y}^{2h}(x) - y(x) \approx p(x)h^s 2^s.$$

Следовательно,
$\tilde{y}^{2h}(x)-\tilde{y}^h(x)\approx p(x)h^s(2^s-1)$, откуда $p(x)\approx$
$\approx\dfrac{\tilde{y}^{2h}(x)-\tilde{y}^h(x)}{h^s(2^s-1)}$, $y(x)-\tilde{y}^h(x)\approx -p(x)h^s=\dfrac{\tilde{y}^h(x)-\tilde{y}^{2h}(x)}{h^s(2^s-1)}h^s=$
$=\dfrac{\tilde{y}^h(x)-\tilde{y}^{2h}(x)}{2^s-1}$. Это соотношение можно использовать и для практического определения главного члена погрешности, и для уточнения построенного приближенного решения $\tilde{y}^h$:

$$y(x)\approx\tilde{y}^h(x)+\frac{\tilde{y}^h(x)-\tilde{y}^{2h}(x)}{2^s-1}. \tag{3}$$

Из (1) следует что погрешность соотношения (3) равна $o(|h|^s)$.

Заметим, что способ Рунге оценки точности численного решения дифференциальных уравнений сходен со способом Рунге определения погрешности численного интегрирования. Из (3) следует, что если $\tilde{y}^h$ и $\tilde{y}^{2h}$ совпадают в пределах заданной точности, то $y(x)\approx\tilde{y}^h$. На этом и основывается применяемый на практике метод подбора шага. на каждом i-том этапе (вычислении $\tilde{y}_i$), исходя из известного значения $\tilde{y}_{i-1}=\tilde{y}(x_{i-1})$, вычисляют $\tilde{y}(x_{i-1}+2h)$ с шагом h и двойным шагом $2h$ соответственно. Если расхождение полученных значений не превышает допустимой погрешности, то считают, что шаг h выбран правильно. В противном случае шаг уменьшают в два раза.

§51. Многошаговые методы

При использовании одношаговых методов предполагается привлекать информацию о решаемой задаче только на отрезке длиной в один шаг, поэтому подобная информация на каждом шаге процесса должна быть, вообще говоря, получена заново, что предопределяет большую трудоемкость соответствующих вычислительных правил. Отказавшись от требования одношаговости можно вычислительные методы строить так, чтобы часть полученной информации могла быть использована повторно на нескольких соседних шагах вычислительного процесса. При этом иногда оказывается целесообразным привлекать также и информацию с забеганием вперед за ту точку, значение решения в которой в настоящий момент мы вычисляем.

Для решения задачи Коши (см. §48) рассмотрим методы типа

$$y_{n+1} = \sum_{i=0}^{p} a_i y_{n-i} + h \sum_{j=-s}^{q} A_j f(x_{n-j}, y_{n-j}), \tag{1}$$

позволяющие искать приближенное решение y_{n+1} в точке x_{n+1} сетки в виде линейной комбинации нескольких известных приближенных значений y_{n-i} решения в точках x_{n-i} этой сетки с коэффициентами $a_i, i = \overline{0, p}$ и нескольких приближенных значений $f(y_{n-j}, x_{n-j})$ производной $y'(x) =$

$= f[x, y(x)]$ искомого решения в точках x_{n-j} с коэффициентами hA_j,

$j = -s, -s+1, \dots, q$. При этом среди указанных значений производной могут быть и неизвестные (при $s \geq 1$).

Если $s < 1$, то вычислительные методы вида (1) обычно называют явными, при $s = 1$ неявными (если, конечно, $A_{-1} \neq 0$), а при $s > 1$ (при условии $\sum\limits_{j=-s}^{-2} A_j^2 > 0$) такие методы называют методами с забеганием вперед.

Далее, ограничимся рассмотрением вычислительных методов вида

$$y_{n+1} = y_n + h \sum_{i=-s}^{q} A_i f(x_{n-i}, y_{n-i}), \qquad (2)$$

т.е. наиболее часто встречающихся в практике многошаговых методов. Среди вычислительных правил вида (2) особенно широко известны экстраполяционные ($s = 0$) и интерполяционные ($s = 1$, $A_{-1} \neq 0$) методы Адамса.

§52. Экстраполяционные методы Адамса

Будем строить многошаговый метод вида

$$y_{n+1} = y_n + h\sum_{i=0}^{q} A_i f(x_{n-i}, y_{n-i}). \tag{1}$$

Для его построения воспользуемся равенством

$$y(x_n + h) = y(x_n) + h\int_0^1 f[x_n + \alpha h, y(x_n + \alpha h)]d\alpha.$$

Заменим интеграл квадратурной суммой, получим

$$y(x_n + h) \approx y(x_n) + h\sum_{i=0}^{q} A_i f[x_n + \alpha_i h, y(x_n + \alpha_i h)]. \tag{2}$$

Выбор параметров $A_i, \alpha_i, i = \overline{0, q}$, в этом приближенном равенстве будем осуществлять, например, на основании требования, чтобы квадратурная формула

$$\int_0^1 z_n(\alpha)d\alpha \approx \sum_{i=0}^{q} A_i z_n(\alpha_i), \tag{3}$$

где $z_n(\alpha) = f[x + \alpha h, y(x_n + \alpha h)]$ была точной для всевозможных алгебраических многочленов до степени $k-1, (0 < k \leq 2q+2)$ включительно.

Это приводит к следующей системе k уравнений с $2q+2$ неизвестными $A_i, \alpha_i, i = 0, 1, \ldots, q$:

$$\sum_{i=0}^{q} A_i = 1, \sum_{i=0}^{q} A_i \alpha_i^j = \frac{1}{j+1}, j = 1, 2, \ldots, k-1. \tag{4}$$

Зададим заранее параметрам α_i, $i = 0, 1, \ldots, q$ нужные для (1) значения. $\alpha_i = -i, i = 0, 1, \ldots, q$. Соответствующие параметрам $A_i, i = 0, 1, \ldots, q$ могут быть найдены из следующей системы линейных алгебраических уравнений:

$$\sum_{i=0}^{q} A_i = 1, \sum_{i=0}^{q} A_i (-i)^j = \frac{1}{j+1}, j = 1, 2, \ldots, q. \tag{5}$$

Так как определитель этой системы есть определитель Вандермонда, а все $\alpha_i = -i, i = 0, 1, \ldots, q$ различны, то значения параметров $A_i, i = 0, 1, \ldots, q$ из (5) могут быть выбраны для любого $q \geq 0$, притом единственным образом. При заданном q тем самым будет построен соответствующий экстраполяционный метод Адамса (1).

Погрешность формулы (1) может быть записана в виде

$$r_{n+1} = h^{q+2} y^{(q+2)}(x_n)[\frac{1}{(q+2)!} - \frac{1}{(q+1)!} \sum_{i=0}^{q} A_i (-i)^{q+1}] + o(h^{q+3}). \tag{6}$$

При $q = 0$ из (1) получаем известный метод Эйлера

$$y_{n+1} = y_n + hf_n,$$

где

$$f_n = f(x_n, y_n). \qquad (7)$$

Из (6) получаем $r_{n+1} = \frac{1}{2}h^2 y''(x_n) + o(h^3)$. При $q = 1$ система (5) примет вид $A_0 + A_1 = 1, A_1 = -\frac{1}{2}$, откуда находим $A_0 = \frac{3}{2}$, $A_1 = -\frac{1}{2}$.

Метод Адамса (1) примет вид

$$y_{n+1} = y_n + \frac{h}{2}(3f_n - f_{n-1}). \qquad (8)$$

Погрешность формулы (8) в силу (6) задается равенством $r_{n+1} = \frac{5}{12}h^3 y'''(x_n) + o(h^4)$.

В отличие от метода Эйлера (7), который не требует предварительного построения начала таблицы значений приближенного решения, экстраполяционный метод Адамса (8) уже предполагает предварительно нахождение величины y_1.

При $q = 2$ имеем $A_0 + A_1 + A_2 = 1$, $A_1 + 2A_2 = -\frac{1}{2}$, $A_1 + 4A_2 = \frac{1}{3}$, откуда находим $A_0 = \frac{23}{12}$, $A_1 = -\frac{4}{3}$, $A_2 = \frac{5}{12}$, что позволяет записать следующий метод

$$y_{n+1} = y_n + \frac{h}{12}(23f_n - 16f_{n-1} + 5f_{n-2}). \qquad (9)$$

Его локальная погрешность задается формулой $r_{n+1} = \frac{3}{8}h^4 y^{(4)}(x_n) + +o(h^5)$. Вычислительное правило (9) может применяться, начиная с $n = = 2$. (Т.е. нужно заранее знать y_1 и y_2).

При $q = 3$ формулы (1) будут иметь вид

$$y_{n+1} = y_n + \frac{h}{24}(55f_n - 59f_{n-1} + 37f_{n-2} - 9f_{n-3}). \tag{10}$$

Погрешность формулы (10) задается равенством $r_{n+1} = \frac{251}{720}h^5 y^{(5)}(x_n) + + o(h^6)$. (Заранее нужно знать y_1, y_2, y_3).

С помощью экстраполирования в x по $q+1$ значениям функции $y'(x)$ в точках x_n, x_{n-1}, x_{n-q} получим формулу

$$\begin{aligned} y(x_{n+1}) = y(x_n) + h[y'(x_n) + \frac{1}{2}\Delta y'(x_{n-1}) + \frac{5}{12}\Delta^2 y'(x_{n-2}) + \\ + \frac{3}{8}\Delta^3 y'(x_{n-3}) + \dots + C_q \Delta^q y'(x_{n-q})] + r_{n+1}. \end{aligned} \tag{11}$$

Предварительно проведем замену переменной $x = x_n + uh$, тогда $y(x_{n+1}) = y(x_n) + h\int_0^1 y'(x_n + uh)du$.

$$\begin{aligned} y'(x_n + uh) = y'(x_n) + \frac{u}{1!}\Delta y'(x_{n-1}) + \frac{u(u+1)}{2!}\Delta^2 y'(x_{n-2}) + \dots + \\ + \frac{u(u+1)\dots(u+q-1)}{q!}\Delta^q y'(x_{n-q}) + r_q(u). \end{aligned} \tag{12}$$

Подставим в (12) значение $y'(x_n+uh)$ и выполним интегрирование, получим формулу (11). Здесь

$$C_q = \frac{1}{q!}\int\limits_0^1 u(u+1)\ldots(u+q-1)du,$$

$$r_{n+1} = h^{q+2}\int\limits_0^1 \frac{u(u+1)\ldots(u+q)}{(q+1)!}y^{(q+2)}(\xi)du =$$

$$= \frac{h^{q+2}}{(q+1)!}y^{(q+2)}(\xi)\int_0^1 u(u+1)\ldots(u+q)du,$$

$$x_{n-q} \le \xi \le x_{n+1}, u = \frac{x-x_n}{h}, 0 \le u \le 1, |r_{n+1}| \le h^{q+2}C_{q+1}M_{q+2},$$

где $M_{q+2} = \max\limits_{x_0\le x\le X}|y^{(q+2)}(x)|$.

Если шаг $h>0$ мал, а решения задачи обладают непрерывной производной до $q+2$ порядка, то величиной $r_{n+1}=o(h^{q+2})$ можно пренебречь и тогда, обозначив $\varphi_i = hf_i = hy_i' = hf(x_i,y_i)$, получим расчетное правило экстраполяционного многочлена Адамса:

$$\begin{aligned} y_{n+1} = y_n + \varphi_n + \frac{1}{2}\Delta\varphi_{n-1} + \frac{5}{12}\Delta^2\varphi_{n-2} + \frac{3}{8}\Delta^3\varphi_{n-3} + \frac{251}{720}\Delta^4\varphi_{n-4} + \\ + \frac{95}{288}\Delta^5\varphi_{n-5} + \frac{19087}{60480}\Delta^6\varphi_{n-6} + \cdots + C_q\Delta^q\varphi_{n-q}, \end{aligned} \tag{13}$$

$$|r_{n+1}| = h^{q+2} y^{(q+2)}(x_n) C_{q+1} + O(h^{q+3}). \tag{14}$$

Чтобы начать расчет по методу Адамса недостаточно знать $y(x_0)$, а надо знать величину решения в q точках, если формула q-того порядка точности. Поэтому недостающие значения нужно вычислить другим методом, например, Рунге-Кутта.

§53. Интерполяционные методы Адамса

При $s = 1$ формула (2) (из §51) примет вид

$$y_{n+1} = y_n + h \sum_{i=-1}^{q} A_i f(x_{n-i}, y_{n-i}). \qquad (1)$$

Построение таких вычислительных методов может быть осуществлено совершенно аналогично случаю экстраполяционных формул Адамса.

Положим $\alpha_i = -i, i = -1, 0, 1, \dots, q$, а для нахождения соответствующих значений параметров $A_i, i = -1, 0, 1, \dots, q$, получим систему линейных алгебраических уравнений

$$\sum_{i=-1}^{q} A_i = 1, \sum_{i=-1}^{q} A_i(-i)^j = \frac{1}{j+1}, j = 1, 2, \dots, q+1,$$

которая однозначно разрешима при любом $q \geq 1$.

Для коэффициентов $A_i, i = -1, 0, 1, \dots, q$ вычислительных методов вида (1), подобно случаю экстраполяционных методов Адамса, можно записать следующие явные выражения

$$A_i = \frac{(-1)^{i+1}}{(i+1)!(q-i)!} \int\limits_0^1 \frac{(\alpha-1)\alpha(\alpha+1)\dots(\alpha+q)}{\alpha+i} d\alpha, i = -1, 0, 1, \dots, q.$$

Погрешность формулы (1) может быть записана в виде

$$r_{n+1} = h^{q+3} y^{(q+3)}(x_n) \left[\frac{1}{(q+3)!} - \frac{1}{(q+2)!} \sum_{i=-1}^{q} A_i (-i)^{q+2} \right] + O(h^{q+4}). \quad (2)$$

Запишем несколько примеров интерполяционных методов Адамса (1). При $q = -1$ будем иметь

$$y_{n+1} = y_n + h f_{n+1},$$

$$r_{n+1} = -\frac{1}{2} h^2 y''(x_n) + O(h^3). \quad (3)$$

При $q = 0$ находим

$$y_{n+1} = y_n + \frac{h}{2}(f_{n+1} + f_n),$$

$$r_{n+1} = -\frac{1}{12} h^3 y'''(x_n) + O(h^4). \quad (4)$$

При $q = 1$ получаем

$$y_{n+1} = y_n + \frac{h}{12}(5 f_{n+1} + 8 f_n - f_{n-1}),$$

$$r_{n+1} = -\frac{1}{24} h^4 y^{(4)}(x_n) + O(h^5). \quad (5)$$

При $q = 2$ найдем

$$y_{n+1} = y_n + \frac{h}{24}(9f_{n+1} + 19f_n - 5f_{n-1} + f_{n-2}),$$

$$r_{n+1} = -\frac{19}{720}h^5 y^{(5)}(x_n) + O(h^6). \qquad (6)$$

Приведем без предварительных выкладок иную форму записи интерполяционных методов Адамса, использующую конечные разности функции $f(x, y)$:

$$y_{n+1} = y_n + \varphi_{n+1} - \frac{1}{2}\Delta\varphi_n - \frac{1}{12}\Delta^2\varphi_{n-1} - \frac{1}{24}\Delta^3\varphi_{n-2} - \frac{19}{720}\Delta^4\varphi_{n-3} -$$
$$-\frac{3}{160}\Delta^5\varphi_{n-4} - \frac{863}{60480}\Delta^6\varphi_{n-5} - \frac{275}{24195}\Delta^7\varphi_{n-6} - \dots - C_{q+1}\Delta^{q+1}\varphi_{n-q}. \qquad (7)$$

Здесь $\varphi_i = hf_i = hy_i' = hf(x_i, y_i)$, $C_{q+1} = \frac{1}{(q+1)!}\int\limits_{-1}^{0} \alpha(\alpha+1)\dots(\alpha+q)d\alpha.$

Погрешность может быть записана в виде

$$r_{n+1} = h^{q+3}y^{(q+3)}(x_n)C_{q+2} + O(h^{q+4}); \qquad (8)$$

$r_{n+1} = h^{q+3}C_{q+2}y^{(q+3)}(\xi)$, где $x_{n-q} \le \xi \le x_{n+1}$;

$$r_{n+1} = \frac{h^{q+3}}{(q+2)!}y^{(q+3)}(\xi)\int\limits_{-1}^{0} \alpha(\alpha+1)\dots(\alpha+q+1)d\alpha, x = x_{n+1} + \alpha h,$$

$$\alpha = \frac{x - x_{n+1}}{h}.$$

Заметим, что построенные интерполяционные методы Адамса не дают явных выражений для нахождения y_{n+1}, а представляют собой лишь уравнение относительно этой неизвестной. Это значительно усложняет процесс вычисления y_{n+1} и является существенным недостатком метода. Обычно значение y_{n+1} находят лишь приближено, пользуясь каким либо методом последовательных приближений. Как в форме (1), так и в форме (7) интерполяционный метод Адамса уже приведен к виду, удобному для итераций. Кроме того, заданием достаточно малого h при удачном выборе начального приближения можно обеспечить сходимость соответствующего итерационного процесса. В самом деле, уравнение относительно y_{n+1} можно (см. (1), (7)) записать в виде $y_{n+1} = \varphi(y_{n+1})$, где $\varphi(y_{n+1}) = h\left(1 + \sum\limits_{i=1}^{q+1} C_i\right) f(x_{n+1}, y_{n+1}) + F(h, x_n, y_n, y_{n-1}, \ldots, y_{n-q})$, а через $F(h, x_n, y_n, y_{n-1}, \ldots, y_{n-q})$ обозначена известная функция указанных аргументов. Если при $x \in [x_0, X]\, f(x, y)$ в достаточно широкой окрестности решения имеет непрерывную производную $\dfrac{\partial f(x,y)}{\partial y}$, то при удачном выборе исходного приближения $y_{n+1}^{(0)}$ за счет уменьшения шага $h > 0$ можно заведомо обеспечить выполнение условий сходимости процесса итераций $y_{n+1}^{(i+1)} = \varphi(y_{n+1}^{(i)}), i = 0, 1, 2, \ldots$.

При использовании интерполяционного метода Адамса (1) обычно в качестве начального приближения к y_{n+1} будет соответствующее значение, полученное экстраполяционным методом, который, вообще говоря, дает более грубый результат, чем интерполяционная формула Адамса. Такая организация вычислений требует поочередного использования экстраполяционного и интерполяционного методов Адамса и носит предсказывающее-исправляющий характер. По формуле (1) находится приближенное значение величины $y_(x_{n+1})$ с локальной погрешностью h^{q+2}, которое затем уточняется на порядок на основании формулы типа (13) (из §52).

Погрешность и сходимость.

Пусть α_n – погрешность округления, r_n – погрешность формулы, ε_n – погрешность приближенного решения, т.е. $\varepsilon_n = y(x_n) - \tilde{y}_n$, $n = q+1, q+ + 2, \ldots, N$, $\varepsilon_{n+1} = \varepsilon_n + h \sum\limits_{j=-1}^{q} \beta_j [f(x_{n-1}, \tilde{y}_{n-j} + \varepsilon_{n-j}) - f(x_{n-j}, \tilde{y}_{n-j})] + + r_n + \alpha_n$.

Предположим, что для погрешностей начальных данных, погрешностей формулы и округления справедливы следующие оценки $|\varepsilon_i| \leq \varepsilon$, $i = -1, 0, 1, \ldots, q$, $|r_m| \leq r$, $\alpha_m \leq \alpha$, $m = q, q+1, \ldots, N$, тогда приближенное решение задачи Коши, получаемое по формуле Адамса на отрезке интегрирования будет равномерно сходиться к точному решению этой задачи, если выполняются условия

$$1)\varepsilon \to 0, h \to 0,$$

$$2)\frac{r}{h} \to 0, h \to 0,$$

$$3)\frac{\alpha}{h} \to 0, h \to 0.$$

Кафедры
ПМиТП
АиГ

§54. Понятие жесткой системы дифференциальных уравнений

Многие из рассмотренных численных методов интегрирования обыкновенных дифференциальных уравнений переносятся без изменений на системы дифференциальных уравнений. Однако в случае численного решения систем уравнений могут появиться дополнительные трудности, связанные с разномасштабностью процессов, описываемых данной системой.

Поясним характер возникающих трудностей на примере системы состоящей из двух независимых уравнений

$$\frac{du_1}{dt} + a_1 u_1 = 0, \frac{du_2}{dt} + a_2 u_2 = 0, t > 0, \tag{1}$$

где a_1 и a_2 – положительные постоянные. Система (1) имеет решение $u_1(t) = u_1(0)e^{-a_1 t}, u_2(t) = u_2(0)e^{-a_2 t}$ монотонно убывающие с ростом t. Предположим, что a_2 гораздо больше, чем a_1. Тогда компонента $u_2(t)$ затухает гораздо быстрее, чем $u_1(t)$, и, начиная с некоторого t, поведение решения $u(t) = (u_1(t), u_2(t))$ почти полностью определяется компонентой $u_1(t)$. Однако оказывается, что при решении системы (1) разностным методом шаг интегрирования τ определяется, как правило, компонентой $u_2(t)$, не существенной с точки зрения поведения решения системы. Например, метод Эйлера

$$\frac{u_1^{n+1} - u_1^n}{\tau} + a_1 u_1^n = 0, \frac{u_2^{n+1} - u_2^n}{\tau} + a_2 u_2^n = 0, \tag{2}$$

где $u_i^n = u_i(t_n), i = 1, 2$, будет устойчив, если шаг τ удовлетворяет одновременно двум неравенствам $\tau a_1 \leq 2, \tau a_2 \leq 2$ Поскольку $a_2 >>> a_1 > 0$, то условие устойчивости приводит к ограничению $\tau \leq 2/a_2$.

Аналогичные трудности, возникают при решении любой системы обыкновенных дифференциальных уравнений

$$\frac{du}{dt} = Au, \tag{3}$$

если матрица этой системы имеет большой разброс собственных чисел.

Сформулируем теперь определение жесткой системы уравнений. Рассмотрим сначала систему (3) с постоянной, т. е. независящей от t матрицей A, $A = [a_{ij}]$.

Определение 1. *Система дифференциальных уравнений (3) с постоянной матрицей $A(m \times m)$ называется жесткой, если*

1. $Re\lambda_k < 0, k = 1, 2, \ldots, m$, *(т. е. система асимптотически устойчива по Ляпунову);*
2. *отношение* $s = \dfrac{\max\limits_{1 \leq k \leq m} |Re\lambda_k|}{\min\limits_{1 \leq k \leq m} |Re\lambda_k|}$ *велико.*

Число s называется *числом жесткости системы* (3). Второе требование не указывает границу для s, начиная с которой система становится жесткой.

Если матрица A – симметрична, то все ее собственные значения вещественны и жесткость системы означает, что матрица A – положительна и что система плохо обусловлена, т. е. $s = \dfrac{\max\limits_{1\le k\le m} \lambda_k}{\min\limits_{1\le k\le m} \lambda_k} >> 1.$

Напомним следующее определение из матричного анализа.

Симметричная матрица A называется неотрицательной, если для $\forall x \in R^n, x \neq 0, x'Ax \geq 0$, положительно определенной, если для $\forall x \in R^n, x \neq 0, x'Ax > 0$.

Если матрица A зависит от t, то $\lambda_k = \lambda_k(t), k = 1, 2, \ldots, m$. При каждом t можно определить число жесткости $s = \dfrac{\max\limits_{1\le k\le m} |Re\lambda_k|}{\min\limits_{1\le k\le m} |Re\lambda_k|}$. В этом случае свойство жесткости может зависеть от длины отрезка интегрирования.

Определение 2. *Система $\dfrac{du}{dt} = A(t)u$ называется жесткой на интервале $(0, T)$, если $Re\lambda_k(t) < 0, k = 1, 2, \ldots, m$ для всех $t \in (0, T)$ и число $\sup\limits_{t\in(0,T)} s(t)$ велико.*

Решение жесткой системы содержит как быстро убывающие, так и медленно убывающие составляющие. Начиная с некоторого $t > 0$ решение системы почти полностью определяется медленно убывающей составляющей. Однако при использовании явных разностных методов быстро убывающая составляющая отрицательно влияет на устойчивость, что вынуждает брать шаг интегрирования слишком мелким, поэтому явные методы оказались непригодными для решения жестких систем.

Выход из этой парадоксальной ситуации был найден в применении неявных абсолютно устойчивых разностных методов. Например, систему (1) можно решать с помощью неявного метода Эйлера

$$\frac{u_1^{n+1} - u_1^n}{\tau} + a_1 u_1^{n+1} = 0, \frac{u_2^{n+1} - u_2^n}{\tau} + a_2 u_2^{n+1} = 0,$$

который устойчив при всех $\tau > 0$. Поэтому шаг интегрирования τ здесь можно выбирать, руководствуясь лишь соображениями точности, а не устойчивости.

§55. Нелинейные системы дифференциальных уравнений

Обобщим понятие жесткости на случай нелинейной системы

$$\frac{du}{dt} = f(t,u), t > 0, \tag{1}$$

где $u(t) = (u_1(t), u_2(t), \ldots, u_m(t))^T$, $f(t,u) = (f_1(t,u), f_2(t,u), \ldots, f_m(t,u))^T$.

Зафиксируем какое-либо решение $v(t)$ системы (1) и образуем разность $z(t) = u(t) - v(t)$ между произвольным решением системы (1) и данным решением $v(t)$. Эта разность удовлетворяет следующей системе уравнений:

$$\frac{dz_k}{dt} = f_k(t, v(t) + z(t)) - f_k(t, v(t)), k = 1, 2, \ldots, m. \tag{2}$$

Эта система (2) получается так:

$$z_k(t) = u_k(t) - v_k(t);$$

$$u_k(t) = z_k(t) + v_k(t);$$

$$\frac{dz_k}{dt} = \frac{du_k}{dt} - \frac{dv_k}{dt};$$

$$\frac{dz_k}{dt} = f_k(t, u(t)) - f_k(t, v(t));$$

$$\frac{dz_k}{dt} = f_k(t, z(t) + v(t)) - f_k(t, v(t)), k = \overline{1, m}.$$

Будем рассматривать $z(t)$ как малое возмущение, внесенное в основное решение $v(t)$. Проведем разложение по формуле Тейлора в правой части системы (2). Так как $f_k(t, u) = f_k(t, u_1, u_2, \ldots, u_m)$ то имеем

$$f_k(t, v + z) - f_k(t, v) = \sum_{j=1}^{m} \frac{\partial f_k(t, v)}{\partial u_j} z_j(t) + O(|z|),$$

где через $O(|z|)$ обозначены величины более высокого, чем первый, порядка малости по z. В результате разложения система (2) примет вид

$$\frac{dz}{dt} = A(t, v(t))z(t) + O(|z|), \tag{3}$$

где $A(t, v(t)) = \dfrac{\partial f_i(t, v(t))}{\partial u_j}, i, j = 1, 2, \ldots, m.$

Отбрасывая в (3) величины $O(|z|)$, получим так называемую систему уравнений первого приближения

$$\frac{d\omega(t)}{dt} = A(t, v(t))\omega(t). \tag{4}$$

Система (4) является системой линейных дифференциальных уравнений относительно $\omega(t)$, так как функция $v(t)$ задана.

Определение жесткости системы нелинейных дифференциальных уравнений связано как с так называемым фиксированным решением $v(t)$, так и с длиной отрезка интегрирования.

Пусть $\lambda_k(t), k = 1, 2, \ldots, m$ – собственные числа матрицы $A(t, v(t))$.

Число жесткости $s(t)$ определяется как $s = \dfrac{\max\limits_{1\le k\le m} |Re\lambda_k(t)|}{\min\limits_{1\le k\le m} |Re\lambda_k(t)|}$.

Определение 1. *Система (1) называется жесткой на решении $v(t)$ и на данном интервале $0 < t < T$, если:*

1. $Re\lambda_k(t) < 0, k = \overline{1, m}$ *для всех* $t \in (0, T)$;
2. *число* $\sup\limits_{t\in(0,T)} s(t)$ *велико.*

§56. Условно устойчивые и абсолютно устойчивые разностные методы

Для задачи Коши

$$\frac{du}{dt} = f(t,u), t > 0, u(0) = u_0 \tag{1}$$

будем рассматривать разностные методы вида

$$\sum_{k=0}^{m} \frac{a_k}{\tau} y_{n-k} = \sum_{k=0}^{m} b_k f(t_{n-k}, y_{n-k}), n = m, m+1, \dots\ . \tag{2}$$

Устойчивость и сходимость метода определяются расположением корней характеристического уравнения

$$\sum_{k=0}^{m} a_k \lambda^{m-k} = 0. \tag{3}$$

А именно, требуется, чтобы все корни удовлетворяли условию $|\lambda| \leq 1$ причем корни λ, для которых $|\lambda| = 1$ не должны быть кратными.

Эти условия устойчивости являются очень общими и не могут учесть многие характерные свойства решений исходной дифференциальной задачи (1) и аппроксимирующего его разностного метода (2). Они означают лишь, что все решения однородного разностного уравнения, соответствующего (2), остаются ограниченными при $n \longrightarrow \infty$.

Предположим, однако, что заранее известна та или иная характерная особенность в поведении решения исходной дифференциальной задачи. Тогда естественно требовать, чтобы эта особенность сохранялась и у решения разностного уравнения. Такое требование приведет к сужению класса допустимых разностных методов. В настоящем параграфе будут рассмотрены методы, предназначенные для расчета асимптотически устойчивых решений уравнения (1).

Рассмотрим сначала характерный пример. Уравнение

$$\frac{du}{dt} = \lambda u, t > 0, u(0) = u_0, \tag{4}$$

где $\lambda < 0$ имеет решение $u(t) = u_0 e^{\lambda t}$,монотонно убывающее при $t \to \infty$.

При любых $\tau > 0$ для решения этого уравнения справедливо неравенство

$$|u(t+\tau)| \le |u(t)|, \tag{5}$$

означающее устойчивость решения $u(t)$ уравнения.

Естественно требовать, чтобы и для решения разностной задачи, аппроксимирующей (4), выполнялось бы неравенство, аналогичное (5). Рассмотрим с этой точки зрения метод Эйлера

$$\frac{y_{n+1} - y_n}{\tau} = \lambda y_n, n = 0, 1 \ldots \ . \tag{6}$$

Из уравнения (6) получаем $y_{n+1} = p y_n$, $p = 1 + \tau\lambda$. Оценка вида (5), т. е. неравенство

$$|y_{n+1}| \le |y_n|, n = 1, 2 \ldots \qquad (7)$$

для метода (6) будет выполнено тогда и только тогда, когда $|p| \le 1$. В случае $\lambda < 0$ это условие эквивалентно следующему ограничению на шаг τ :

$$0 < \tau \le \frac{2}{|\lambda|}. \qquad (8)$$

Определение 1. *Разностный метод (2) называется абсолютно устойчивым, если он устойчив при любых $\tau > 0$ и условно устойчивым, если он устойчив при некоторых ограничениях на шаг τ .*

Метод Эйлера условно устойчив при условии (8). Примером абсолютно устойчивого метода для уравнения (4) с $\lambda < 0$ является неявный метод Эйлера: $\frac{y_{n+1} - y_n}{\tau} = \lambda y_{n+1}$ для которого, $|p| = |(1 - \tau\lambda)^{-1}| < 1$ для любых $\tau > 0$. Приведенные здесь простые примеры характерны тем, что и для более общих асимптотически устойчивых систем дифференциальных уравнений явные разностные схемы являются условно устойчивыми, а среди неявных методов существуют абсолютно устойчивые.

Условная устойчивость является недостатком явного метода, так как вынуждает брать слишком мелкий шаг . Неявный метод лишен этого недостатка, однако его применение к задаче (1) приводит к необходимости решения на каждом шаге системы алгебраических уравнений, в общем случае нелинейной.

§57. Специальные определения устойчивости

При исследовании разностных методов для жестких систем уравнений обычно рассматривают уравнение

$$\frac{du}{dt} = pu, \tag{1}$$

где p – произвольное комплексное число. Свойства различных разностных методов изучают и сопоставляют на примере модельного уравнения (1). Для того, чтобы уравнение (1) действительно моделировало систему $\dfrac{du}{dt} = Au$, необходимо рассматривать его при таких p, которые являются собственными значениями матрицы A.

Разностный метод (2) из предыдущего параграфа, примененный к уравнению (1), имеет вид

$$\sum_{k=0}^{m}(a_k - \mu b_k)\lambda^{m-k} = 0, \tag{2}$$

которое отличается от уравнения (3) (из §56) тем, что его коэффициенты зависят от параметра $\mu = \tau p$. При малых μ корни уравнений (3) (из §56) и (2) близки.

Кроме обычного определения устойчивости разностного метода (все корни характеристического уравнения (2) не превосходят по модулю 1), в случае жестких систем используют и другие, более узкие определения устойчивости. Здесь мы рассмотрим два таких определения: A - устойчивый метод и $A(\alpha)$ - устойчивый метод.

Определение 1. *Областью устойчивости разностного метода (2) (из §56) называется множество всех точек комплексной плоскости для которых данный метод, примененный к уравнению (1), является устойчивым.*

Рассмотрим, например, явный метод Эйлера

$$\frac{y_{n+1} - y_n}{\tau} = f(t_n, y_n).$$

В применении к уравнению (1) этот метод принимает вид

$$y_{n+1} = (1 + \mu)y_n, \mu = \tau p.$$

Условие устойчивости $|1 + \mu| \leq 1$ для комплексного $\mu = \mu_0 + i\mu_1$ означает, что $(\mu_0+1)+\mu_1^2 \leq 1$. Тем самым область устойчивости данного метода представляет собой круг единичного радиуса с центром в точке $(-1, 0)$. Для неявного метода Эйлера $\dfrac{y_{n+1} - y_n}{\tau} = f(t_{n+1}, y_{n+1})$ областью устойчивости является внешность круга единичного радиуса с центром в точке $(1, 0)$.

Определение 2. *Разностный метод называется A - устойчивым, если область его устойчивости содержит левую полуплоскость* $\mathrm{Re}\mu < 0$.

Отметим, что уравнение (1) асимптотически устойчиво при $\mathrm{Re}p < 0$. Поэтому A - устойчивый разностный метод является абсолютно устойчивым (при любых $\tau > 0$), если устойчиво решение исходного дифференциального уравнения.

Нетрудно видеть, что невный метод Эйлера является A - устойчивым, а явный метод Эйлера не является.

Рассмотрим еще одношаговый метод второго порядка точности

$$\frac{y_{n+1} - y_n}{\tau} = 0,5(f(t_{n+1}, y_{n+1}) + f(t_n, y_n)). \qquad (3)$$

Для уравнения (1) метод (3) примет вид: $y_{n+1} = \lambda y_n$, $\lambda = \dfrac{1 + 0,5\mu}{1 - 0,5\mu}$. Действительно,

$$\frac{y_{n+1} - y_n}{\tau} = 0,5f(py_{n+1} + py_n); y_{n+1} - y_n = 0,5\mu y_{n+1} + 0,5\mu y_n;$$

$$y_{n+1}(1 - 0,5\mu) = y_n(1 + 0,5\mu); y_{n+1} = \frac{1 + 0,5\mu}{1 - 0,5\mu} y_n.$$

Отсюда видно, что $|\lambda| \leq 1$ тогда и только тогда, когда $\mathrm{Re}\mu < 0$ т. е. метод (3) является A - устойчивым.

При решении жестких задач было бы желательно пользоваться именно A - устойчивыми методами, так как условия их устойчивости не накладывают ограничений на шаг τ. Оказывается, однако, что класс A -

устойчивых методов весьма узок. В частности, среди методов вида (2) из предыдущего параграфа не существует A - устойчивых методов.

Доказано, что среди неявных линейных многошаговых методов нет A - устойчивых методов, имеющих порядок точности выше второго. Примером A - устойчивых методов второго порядка точности является симметричная схема (3). В связи с этим было введено еще несколько определений устойчивости, которые являются менее ограничительными, чем определение A - устойчивости.

Определение 3. *Разностный метод называется $A(\alpha)$ - устойчивым, если область его устойчивости содержит угол $|\arg(-\mu)| < \alpha$, $\mu = \tau p$.*

В частности $A\left(\dfrac{\pi}{2}\right)$ - устойчивость совпадает с A - устойчивостью.

Доказано, что ни для какого α не существует явного $A(\alpha)$ - устойчивого линейного двушагового метода. Построены $A(\alpha)$ - устойчивые неявные методы третьего и четвертого порядка точности. Например, схема

$$\frac{25y_n - 48y_{n-1} + 36y_{n-2} - 16y_{n-3} + 3y_{n-4}}{12\tau} = f(t_n, y_n) \tag{4}$$

имеет четвертый порядок точности и $A(\alpha)$ - устойчива при некотором $\alpha > 0$.

§58. Чисто неявные разностные методы

В настоящее время при интегрировании жестких систем уравнений широко используется метод Гира, в основу которого положены чисто неявные разностные методы высокого порядка точности.

Разностный метод

$$\sum_{k=0}^{m} a_k y_{n-k} = \tau f(t_n, y_n) \tag{1}$$

называется чисто неявным. Он является частным случаем метода (3) (из §57), когда $b_1 = b_2 = \ldots = b_m = 0, b_0 = 1$. Для отыскания y_n получаем из (1) нелинейное уравнение

$$a_0 y_n - \tau f(t_n, y_n) = -\sum_{k=1}^{m} a_k y_{n-k}, \tag{2}$$

которое можно решать тем или иным итерационным методом. Условие p-го порядка аппроксимации в случае метода (1) принимают вид

$$a_0 = -\sum_{k=1}^{m} a_k, \sum_{k=1}^{m} k a_k = -1, \sum_{k=1}^{m} k^l a_k = 0, l = 2, 3, \ldots, p. \tag{3}$$

Отсюда видно, что наивысший достижимый порядок аппроксимации чисто неявного m - шагового метода равен m. Метод Гира использует чисто неявные схемы наивысшего порядка аппроксимации. Система уравнений (3) для определения коэффициентов $a_1, \ldots, a_m$ метода наивысшего порядка имеет вид

$$\begin{aligned} a_1 + 2a_2 + \ldots + ma_m &= -1, \\ a_1 + 2^2 a_2 + \ldots + m^2 a_m &= 0, \\ &\ldots \\ a_1 + 2^m a_2 + \ldots + m^m a_m &= 0. \end{aligned} \tag{4}$$

Эта система однозначно разрешима, так как её определитель отличен от нуля.

При $m = 1$ метод (2), (4) совпадает с неявным методом Эйлера. При $m = 2$ получаем метод

$$\frac{3}{2} y_n - 2y_{n-1} + \frac{1}{2} y_{n-2} = \tau f(t_n,\ y_n). \tag{5}$$

При $m = 3$ получаем метод

$$\frac{11}{6} y_n - 3y_{n-1} + \frac{3}{2} y_{n-2} - \frac{1}{3} y_{n-3} = \tau f(t_n,\ y_n). \tag{6}$$

Метод (5) имеет второй порядок точности, а метод (6) – третий. При $m = 4$ получаем схему (4) (из §57). Для практических расчётов используются аналогичные методы вплоть до десятого порядка точности.

Важно отметить, что чисто неявные разностные методы обладают хорошими свойствами устойчивости, позволяющими использовать их для решения жестких систем уравнений.

Рассмотрим более подробно метод второго порядка (5) и найдем область его устойчивости. Для модельного уравнения метод (5) примет вид

$$\frac{3}{2}y_n - 2y_{n-1} + \frac{1}{2}y_{n-2} = \mu y_n, \tag{7}$$

где $\mu = \tau p$. Ему соответствует характеристическое уравнение

$$\left(\frac{3}{2} - \mu\right)\lambda^2 - 2\lambda + \frac{1}{2} = 0. \tag{8}$$

Нам нужно найти множество точек G комплексной плоскости $\mu = \mu_0 +$ $+ i\mu_1$, для которых оба корня $\lambda_{1,2}(\mu)$ уравнения (8) не превосходят по модулю 1. Границей области G является множество таких точек, для которых $|\lambda| = 1$. Выразим из уравнения (8) параметр μ через переменную λ, т. е.

$$\mu = \frac{3}{2} - \frac{2}{\lambda} + \frac{1}{2\lambda^2}. \tag{9}$$

Отсюда видно, что если $|\lambda| = 1$, т. е. $\lambda = e^{-i\varphi}$ то

$$\mu = \frac{3}{2} - 2e^{i\varphi} + \frac{1}{2}e^{2i\varphi}. \tag{10}$$

При изменении аргумента φ от 0 до 2π точка μ описывает замкнутую кривую Γ, симметричную относительно действительной оси. Для точек $\mu(\lambda)$, расположенных снаружи от этой кривой выполнено условие

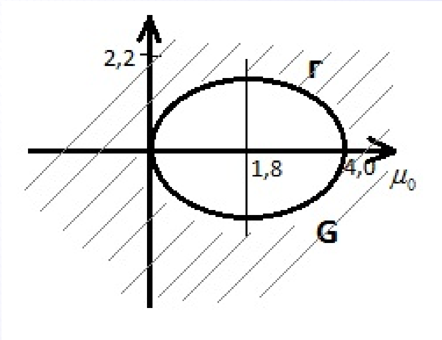

Рисунок 1.

$|\lambda| < 1$ поэтому область устойчивости G метода (7) представляет собой внешность кривой Г. Точки, расположенные внутри Г, составляют область неустойчивости. Обозначим $x = \cos\varphi$, тогда (10) запишется в виде $\mu = (1-x)^2 \pm i\sqrt{1-x^2}(2-x)$, откуда следует, что вся кривая Г расположена в правой полуплоскости. Поэтому область устойчивости метода (5) целиком содержит левую полуплоскость и тем самым метод (5) является A - устойчивым.

Исследуем аналогичным образом область устойчивости метода 4-го порядка.

Запишем для него характеристическое уравнение

$$\mu = \frac{1}{12}(25 - 48\lambda^{-1} + 36\lambda^{-2} - 16\lambda^{-3} + 3\lambda^{-4})$$

и полагаем $\lambda = e^{-i\varphi}$, находим уравнение границы, разделяющей области устойчивости и неустойчивости:

$$\mu = -\frac{2}{3}(1 - x)^3(3x + 1) \pm \frac{i\sqrt{1 - x^2}}{3}(6x^3 - 16x^2 + 15x - 8),$$

где $x = \cos\varphi$. В отличии от предыдущего метода, имеются точки границы, расположенные в левой полуплоскости. Поэтому метод (4) (из §57) не является A - устойчивым.

Найдем теперь значение α, при котором метод (5) (из §57) $A(\alpha)$ - устойчив. Для этого достаточно найти угол α, который образует касательная γ, проходящая через точку (0,0), с отрицательным направлением оси μ_0. Обозначим

$$\mu_0(x) = -\frac{2}{3}(1-x)^3(3x+1); \mu_1(x) = \frac{\sqrt{1 - x^2}}{3}(6x^3 - 16x^2 + 15x - 8). \quad (11)$$

Условия касания γ с границей определяется следующим уравнением относительно параметра x:

$$\frac{\mu_1(x)}{\mu_0(x)} = \frac{\mu_1'(x)}{\mu_0'(x)}. \quad (12)$$

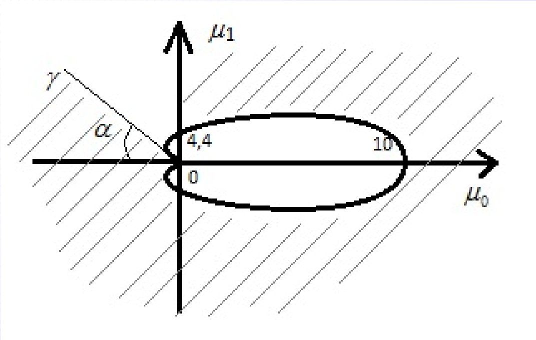

Рисунок 2.

Из (11) получаем

$$\mu_0'(x) = 8x(1-x)^2; mu_1'(x) = \frac{-8x^4+16x^3-4x^2-8x+5}{\sqrt{1-x^2}}. \qquad (13)$$

Уравнение (12) после подстановки в него выражений для μ_0, μ_0', μ_1, μ_1' из (11), (13) и приведения подобных членов сводится к линейному уравнению $x - 0,2 = 0$. При $x = 0,2$ из (11) получим $\mu_0 = \dfrac{4^5}{3*5^4}$, $\mu_1 = $ $= \dfrac{\sqrt{24}*699}{3*5^4}$, так что $\operatorname{tg}\alpha = \dfrac{\mu_1}{\mu_0} = \dfrac{\sqrt{24}*699}{512} > \sqrt{6}$.

Поэтому метод (5) из §57 $A(\alpha)$ - устойчив, где $\alpha = \operatorname{arctg}\sqrt{6} \approx 68°$.

ГЛАВА 5
РЕШЕНИЕ ГРАНИЧНЫХ ЗАДАЧ ДЛЯ ОДУ

Существуют краевые (граничные) условия различных видов:
1) $y(a) = A$, $y(b) = B$.
2) Краевые условия 2-го рода
$y'(a) = A$, $y'(b) = B$.
3) $\begin{cases} \alpha_1 y(a) + \alpha_2 y'(a) = A \\ \beta_1 y(b) + \beta_2 y'(b) = B. \end{cases}$

§59. Метод стрельбы (пристрелки)

Среди граничных задач для ОДУ существенную часть составляют задачи для уравнений и систем второго порядка. В частности, такие задачи возникают в баллистике, теории упругости и т. д.

Рассмотрим частную, но довольно распространенную граничную задачу. Ищется решение уравнения

$$L_y \equiv y'' - p(x)y = f(x) \text{ на } [0, X] \tag{1}$$

при граничных условиях

$$y(0) = a, y(0) = b. \tag{2}$$

Зададимся шагом $h = \dfrac{X}{N}$, N – целое, точки $x_j = jh$ примем за узлы сетки, y_j – приближения к значениям $y(x_j)$. После замены производной $y''(x_j)$ на разностное отношение $\dfrac{\Delta^2 y_j}{h^2} = \dfrac{y_{j+1} - 2y_j + y_{j-1}}{h^2}$ получим систему уравнений

$$l(y_j) = \frac{y_{j+1} - 2y_j + y_{j-1}}{h^2} - p_j y_j = f_j, j = 1, 2, \ldots, N - 1. \qquad (3)$$

Здесь $p_j = p(x_j)$, $f_j = f(x_j)$, граничные условия заменим соотношениями

$$y_0 = a, y_N = b. \qquad (4)$$

Можно доказать, что при $p(x) \geq 0$ система уравнений (3), (4) имеет решение. Однородная система, соответствующая (3) и (4), имеет только нулевое решение. Так как число уравнений в (3), (4) равно числу неизвестных, то неоднородная система (3), (4) имеет единственное решение.

При численном решении задачи Коши значения решения определяются в последовательных узлах по реккурентным формулам, а в случае граничной задачи такой возможности нет, поскольку значения решения зависят от граничных условий на обоих концах отрезка интегрирования.

Метод пристрелки.

Возьмем частное решение $y_0(x)$ неоднородного уравнения

$$y_0'' - p(x)y_0 = f(x)$$

и два линейнонезависимых решения однородного уравнения

$$y_i'' - p(x)y_i = 0, i = 1, 2.$$

Общее решение неоднородного уравнения запишется в виде

$$y_0(x) + C_1 y_1(x) + C_2 y_2(x);$$

постоянные C_1 и C_2 определяем из граничных условий. Приближения к функциям $y_1(x)$, $y_2(x)$ находим каким-либо численным методом задачи Коши; затем находим C_i, $i = 1, 2$ и получаем нужное решение.

Экономнее поступить следующим образом. Находим частное решение неоднородного уравнения $y_0'' - p(x)y_0 = f(x)$, удовлетворяющее условию $y_0(0) = a$, и частное решение однородного уравнения $y_1(x)$, удовлетворяющее условию $y_1(0) = 0$. Общее решение неоднородного уравнения, удовлетворяющее условию $y(0) = a$ имеет вид $y_0(x) + Cy_1(x)$; значение C определяется из условия $y_0(X) + Cy_1(X) = b$. Метод решения граничной задачи, следующий этой схеме, принято называть *методом стрельбы* или *методом пристрелки*.

Сеточный аналог этого метода заключается в следующем. Задаемся $y_0^0 = a$, $y_0^1 = 0$, произвольными $y_1^1 \neq 0$ и y_1^0, и из уравнений

$$\frac{y_{n+1}^0 - 2y_n^0 + y_{n-1}^0}{h^2} - p_n y_n^0 = f_n; \tag{5}$$

$$\frac{y_{n+1}^1 - 2y_n^1 + y_{n-1}^1}{h^2} - p_n y_n^1 = 0 \tag{6}$$

последовательно определяем $y_2^0, \ldots, y_N^0$, $y_2^1, \ldots, y_N^1$. Затем находим C из уравнения $y_N^0 + Cy_N^1 = b$ и полагаем $y_n = y_n^0 + Cy_n^1$; функции y_n являются требуемым решением.

§60. Метод редукции к задачам Коши

Пусть на $[a, b]$ задано линейное обыкновенное дифференциальное уравнение порядка n с непрерывными коэффициентами $p_i(x)$, $i = \overline{0, n}$ и свободным членом $f(x)$:

$$L_n(y) = p_0(x)y^{(n)} + p_1(x)y^{(n-1)} + \ldots + p_n(x)y = f(x). \tag{1}$$

Предположим, что $p_0(x)$ не обращается в нуль на $[a, b]$. Всегда можно считать $p_0(x) > 0$. Допустим, что даны n условий вида

$$\begin{gathered} l_j(y(x_1), y'(x_1), \ldots, y^{(n-1)}(x_1), \ldots, y(x_k), y'(x_k), \ldots, y^{(n-1)}(x_k)) = 0, \\ a \leq x_1 \leq x_2 \leq \ldots \leq x_k \leq b, j = 1, 2, \ldots, n. \end{gathered} \tag{2}$$

Будем считать l_j линейными функциями, хотя можно применить редукцию и тогда, когда l_j – нелинейные.

Требуется найти на $[a, b]$ решение, которое удовлетворило бы условиям (2). Предполагаем, что оно единственное.

Суть метода к задачам Коши состоит в следующем. Известно, что искомое решение задачи (1) и (2) содержится во множестве решений, определяемых формулой общего решения уравнения (1)

$$y(x) = Y(x) + c_1 Z_1 + c_2 Z_2 + \ldots + c_n Z_n, \tag{3}$$

где $Y(x)$ – частные решения неоднородного уравнения $L_n(Y) = f(x)$, а $Z_i(x)$ – фундаментальная система решений соответствующего однородного уравнения $L_n(Z_i) = 0$, c_i – произвольные постоянные. Для выбора c_j воспользуемся условиями (2). Если предположить, что известны функции $Y(x)$ и $Z_i(x)$ и потребовать, чтобы решение $y(x)$ из (3) удовлетворяло условию (2), то это дает нам линейную систему n уравнений для определения величин c_1, c_2, ..., c_n. Если эта система имеет единственное решение, то разыскиваемая функция будет найдена по формуле (3) единственным образом. Осталось выяснить, как найти функции $Y(x)$ и $Z_i(x)$. Чаще всего функцию $Y(x)$ определяют по уравнению $L_n(Y) = f(x)$ с нулевыми, например, начальными условиями $Y(a) = Y'(a) = \ldots = Y^{(n-1)}(a) = 0$, а функцию $Z_i(x)$, $i = 1, 2, \ldots, n$ по уравнению $L_n(Z_i) = 0$ с начальными условиями:

$$Z_i^{(v)}(a) = \begin{cases} 1, v = i - 1, \\ 0, v \neq i - 1, v = 1, 2, \ldots, n - 1. \end{cases}$$

Таким образом, определенные решения $Z_i(x)$ образуют полную систему линейно независимых решений уравнения $L_n z = 0$. Итак, в методе редукции порядок вычислений следующий:

1) вычисляем на $[a, b]$ $Y(x)$ и $Z_i(x)$, привлекая любой метод решения задачи Коши;

2) находим величины $c_1, c_2, \ldots, c_n$, используя вид общего решения $y(x) =$

$= Y(x) + \sum_{i=1}^{n} c_i Z_i(x)$ и условия (2);

3) вычисляем искомое решение по формуле (3).

§61. Метод прогонки для граничных задач

Дано линейное дифференциальное уравнение 2-го порядка

$$L(y) = y'' + p(x)y' + q(x)y = f(x), a \le x \le b \tag{1}$$

с граничными условиями:

$$\alpha_1 y'(a) + \alpha_0 y(a) = A, \alpha_1^2 + \alpha_0^2 > 0, \tag{2}$$

$$\beta_1 y'(b) + \beta_0 y(b) = B, \beta_1^2 + \beta_0^2 > 0. \tag{3}$$

Ищем функцию $y(x)$, удовлетворяющую условиям (2)-(3). Рассмотрим общее решение дифференциального уравнения (1)

$$y(x) = Y(x) + c_1 Z_1(x) + c_2 Z_2(x), \tag{4}$$

где $Y(x)$ – частное решение уравнения (1), определяемое уравнениями $L(Y) = f(x)$, $Y(a) = Y'(a) = 0$, $Z_1(x)$ и $Z_2(x)$ – линейно независимые решения соответствующего однородного уравнения, определяемые условиями:

$$L(Z_1) = 0, Z_1(a) = 1, Z_1'(a) = 0,$$

$$L(Z_2) = 0, Z_2(a) = 0, Z_2'(a) = 1,$$

где c_1 и c_2 – произвольные постоянные. Рассмотрим 1-ое граничное условие $\alpha_1 y'(a) + \alpha_0 y(a) = A$ и воспользуемся им для определения c_1 и c_2. Имеем: $y(a) = Y(a) + c_1 Z_1(a) + c_2 Z_2(a) = c_1$,

$$y'(a) = Y'(a) + c_1 Z_1'(a) + c_2 Z_2'(a) = c_2.$$

Так как $y(a) = c_1$, $y'(a) = c_2$, то $\alpha_1 c_2 + \alpha_0 c_1 = A$.

Пусть, например, $\alpha_1 \neq 0$, тогда $c_2 = -\dfrac{\alpha_0}{\alpha_1} c_1 + \dfrac{1}{\alpha_1} A$.

Подставим c_2 в (4), получим функцию $y(x)$, которая уже при $\forall c_1$ удовлетворяет граничному условию (2). Запишем $y(x) = Y(x) + c_1 Z_1(x) + \dfrac{A}{\alpha_1} Z_2(x) - \dfrac{\alpha_0}{\alpha_1} c_1 Z_2(x) = \psi(x) + c_1 \varphi(x)$. Отсюда

$$y(x) = \psi(x) + c_1 \varphi(x), \tag{5}$$

где $\psi(x) = Y(x) + \dfrac{1}{\alpha_1} A Z_2(x)$, $\varphi(x) = Z_1(x) - \dfrac{\alpha_0}{\alpha_1} Z_2(x)$.

Заметим, что: $L(\psi) = f(x)$, $\psi(a) = 0$, $\psi'(a) = \dfrac{1}{\alpha_1} A$, так как $L(\psi) = L(Y) + \dfrac{A_0}{\alpha_1} L(Z_2) = f(x)$, $\psi(a) = Y(a) + \dfrac{A}{\alpha_1} Z_2(a) = 0$, $\psi'(a) = Y'(a) + \dfrac{A}{\alpha_1} Z_2'(a) = \dfrac{A}{\alpha_1}$, $L(\psi) = L(Y(x)) + \dfrac{A}{\alpha_1} L(Z_2(a)) = f(x)$. $L(\varphi) = 0$, $\varphi(a) = 1$, $\varphi'(a) = -\dfrac{\alpha_0}{\alpha_1}$, так как $\varphi(a) = Z_1(a) - \dfrac{\alpha_0}{\alpha_1} Z_2(a) = 1$, $\varphi'(a) = Z_1'(a) - \dfrac{\alpha_0}{\alpha_1} Z_2'(a) = -\dfrac{\alpha_0}{\alpha_1}$, $L(\varphi) = L(Z_1) - \dfrac{\alpha_0}{\alpha_1} L(Z_2) = 0$. Формула (5) определяет линейное однопараметрическое семейство функций. Его можно рассматривать как общее решение линейного дифференциального уравнения 1-го порядка

$$y' + \alpha(x) y = \beta(x), \tag{6}$$

где

$$\alpha(x) = -\frac{\varphi'(x)}{\varphi(x)}, \beta(x) = \psi'(x) - \frac{\varphi'(x)}{\varphi(x)}\psi(x). \tag{7}$$

Чтобы убедиться в этом, достаточно подставить $y = \psi + c_1\varphi$ в уравнение (6): $\psi' + c_1\varphi' + \alpha(x)(\psi + c_1\varphi) = \beta(x)$ и потребовать, чтобы полученное равенство выполнялось при всех c_1 и x, что дает условия $\psi' + \alpha(x)\psi + c_1(\varphi' + \alpha(x)\varphi) = \beta(x)$, $\varphi' + \alpha(x)\varphi = 0$, $\psi' + \alpha(x)\psi = \beta$. Отсюда следует (7). Используя (7), получим значения $\alpha(x)$ и $\beta(x)$ в точке $x = a$. Получим $\alpha(a) = -\frac{\varphi'(a)}{\varphi(a)} = -(-\frac{\alpha_0}{\alpha_1}) : 1 = \frac{\alpha_0}{\alpha_1}$,
$\beta(a) = \psi'(a) - \frac{\varphi'(a)}{\varphi(a)}\psi(a) = \frac{A}{\alpha_1} - \frac{\alpha_0}{\alpha_1} \cdot 0 = \frac{A}{\alpha_1}$. Следовательно,

$$\alpha(a) = \frac{\alpha_0}{\alpha_1}, \beta(a) = \frac{A}{\alpha_1}. \tag{8}$$

Для определения значений $\alpha(x)$ и $\beta(x)$ в других точках $[a, b]$ построим дифференциальное уравнение для $\alpha(x)$ и $\beta(x)$. Будем считать, что y – произвольное решение уравнения (6), которое запишется в виде $y' = \beta(x) - \alpha(x)y$. Отсюда

$$y'' = \beta'(x) - \alpha'(x)y - \alpha(x)y' = \beta'(x) - \alpha'(x)y - \alpha(x)[\beta(x) - \alpha(x)y].$$

Функция y должна удовлетворять (1). Подставим y' и y'' в (1), получим $L(y) = \beta' - p(x)\beta - \alpha\beta - y[\alpha' + p(x)\alpha - \alpha^2 - q] = f(x)$. Полученное равенство должно выполняться при всех значениях x и для всех решений уравнения (6). Но при каждом фиксированном значении x существует решение этого уравнения, принимающее наперед заданное значение y. Поэтому равенство должно выполняться тождественно относительно y, поэтому

$$\alpha' + p(x)\alpha - \alpha^2 = q(x), \tag{9}$$

$$\beta' + p(x)\beta - \alpha\beta = f(x). \tag{10}$$

Теперь $\alpha(x)$ и $\beta(x)$ можно найти как решение задачи Коши соответственно (9) и $\alpha(a) = \dfrac{\alpha_0}{\alpha_1}$, (10) и $\beta(a) = \dfrac{1}{\alpha_1}A$. Задача Коши для определения $\alpha(x)$:

$$\alpha' + p(x)\alpha - \alpha^2 = q(x), \alpha(a) = \frac{\alpha_0}{\alpha_1}.$$

Задача Коши для определения $\beta(x)$:

$$\beta' + p(x)\beta - \alpha\beta = f(x), \beta(a) = \frac{A}{\alpha_1}.$$

Процесс определения $\alpha(x)$ и $\beta(x)$ получил название *процесса прогонки прямой*.

Предположим, что $\alpha(x)$ и $\beta(x)$ найдены, тогда искомое решение находится из уравнения (6) и начального условия

$$\begin{cases} y'(b) + \alpha(b)y(b) = \beta(b), \\ \beta_1 y'(b) + \beta_0 y(b) = B. \end{cases}$$

Из этой системы уравнений можно определить $y(b)$ и $y'(b)$. Нам достаточно знать $y(b)$. Если $D = \begin{vmatrix} 1 & \alpha(b) \\ \beta_1 & \beta_0 \end{vmatrix} \neq 0$, то (единственное решение)

$$y(b) = \frac{1}{D}(B - \beta_1\beta(b)). \qquad (11)$$

Действительно, $D_2 = \begin{vmatrix} 1 & \beta(b) \\ \beta_1 & B \end{vmatrix} = B - \beta_1\beta(b)$, $y(b) = \dfrac{D_2}{D} = \dfrac{\beta - \beta_1\beta(b)}{D}$, $D = \beta_0 - \beta_1\alpha(b)$.

Искомое решение задачи (1)–(3) может быть найдено как решение задачи Коши для уравнения (6) и начального значения (11). Процесс определения $y(x)$ по уравнению (6) и начальному условию (11) называется *обратной прогонкой*.

Если $D = 0$, то по методу прогонки граничная задача будет иметь бесконечное множество решений. Таким образом, справедлива

Теорема 1. *Для того, чтобы метод прогонки был применим к граничной задаче (1)–(3) необходимо и достаточно:*

1) решение $\varphi(x)$ задачи Коши для уравнения $\varphi'' + p(x)\varphi' + q(x)\varphi = 0$ с начальными условиями $\varphi(a) = 1$, $\varphi'(a) = -\dfrac{\alpha_0}{\alpha_1}$ не обращается в нуль ни в одной точке $[a, b]$,

2) определитель $\Delta = \begin{vmatrix} 1 & \alpha(b) \\ \beta_1 & \beta_0 \end{vmatrix} \neq 0$.

§62. Метод моментов

Решение ряда задач физики и механики иногда более выгодно находить в аналитической форме. Этим свойствам обладают методы Галеркина, моментов, наименьших квадратов и др. По универсальности и простоте реализации упомянутые методы часто уступают методу сеток. Вместе с тем они имеют и определённое преимущество перед методом сеток, поскольку позволяют находить решение в аналитическом виде.

Рассмотрим метод Галеркина и метод моментов на примере граничной задачи для линейного дифференциальное уравнения второго порядка.

Пусть при $a \le x \le b$ задано дифференциальное уравнение

$$F(x, y, y', y'') = 0 \tag{1}$$

с граничными условиями

$$l_\alpha(y) = \alpha_0 y(a) + \alpha_1 y'(a) = A, \tag{2}$$

$$l_b(y) = \beta_0 y(a) + \beta_1 y'(b) = B. \tag{3}$$

Будем считать, что граничная задача (1)-(3) имеет на $[a, b]$ единственное решение и это решение непрерывно дифференцируемо до второго порядка включительно.

В методе моментов применительно к граничной задачи (1)-(3) рассматриваются две системы функций:
1) система функций $\{\psi_k(x)\}$, подчинённая условиям:

а) $\psi_k(x) \in C[a,b]$, $k = 0, 1, 2, \ldots$;

б) функции $\psi_k(x)$ образует замкнутую на $[a,b]$ систему;

2) система функций $\{\varphi(x)\}$ удовлетворяющая условиям:

а) $\varphi_k(x) \in C_2[a,b]$, $k = 0, 1, 2, \ldots$;

б) при любом конечном n функции $\varphi_1(x)$, $\varphi_2(x)$, $\ldots$, $\varphi_n(x)$ линейно-независимы на $[a,b]$;

в) функция $\varphi_0(x)$ удовлетворяет граничным условиям (2), (3), а функция $\varphi_k(x)$, $k = 1, 2, \ldots$ удовлетворяет требованию $l_a(\varphi_k) = 0$, $l_b(\varphi_k) = 0$, $k = 1, 2, \ldots$;

г) функции $\varphi_k(x)$ образуют в классе функций $C_2[a,b]$ удовлетворяющих условиям (2), (3), полную систему.

Определение 1. *Систему функций $\{\psi_k(x)\}$ будем называть замкнутой на множестве интегрируемых на $[a,b]$ функций, если не существует такой непрерывной функции, кроме тождественного нуля, которая была бы ортогональной на $[a,b]$ ко всем функциям $\psi_k(x)$.*

Таким образом, если система функций $\{\psi_k(x)\}$ замкнута, то из условия $\int\limits_a^b f(x)\psi_k(x)dx = 0$, $k = 0, 1, 2, \ldots$, где $f(x)$ непрерывная на $[a,b]$ функций следует, что $f(x) \equiv 0$.

Свойство полноты понимается так. Обозначим через G класс функций $y(x)$, принадлежащих $C_2[a,b]$ и удовлетворяющих граничным условиям (2), (3).

Определение 2. *Говорят, что система функций* $\{\varphi_k(x)\}$ *<u>полна в классе</u>* G, *если для* $\forall\varepsilon > 0$ *и любой функции* $y(x) \in G$ *можно указать такое* n *и такие параметры* $a_1, a_2, \ldots, a_n$, *что имеет место неравенство* $|y^{(i)}(x) - y_n^{(i)}(x)| < \varepsilon$, $i = 0, 1, 2$, $a \le x \le b$, *где* $y_n(x) = \varphi_0(x) + \sum_{k=1}^{n} a_k\varphi_k(x)$.

Это означает, что для любой допустимой функции $y(x) \in G$ найдётся такая функция $y_n(x)$, которая на $[a, b]$ будет сколь угодно точно приближать функцию $y(x)$ вместе с её производными $y'(x)$ и $y''(x)$. Перейдём к построению приближённого решения граничной задачи (1)-(3) методом моментов.

Будем считать функцию $F(x, y, z, t)$ непрерывной по всем аргументам в области $\Omega = \{a \le x \le b, -\infty < y, z, t < +\infty\}$. Составим линейную комбинацию из $n + 1$ первых функций $\varphi_k(x)$:

$$y_n(x) = \varphi_0(x) + \sum_{k+1}^{n} a_k\varphi_k(x), \quad (4)$$

где a_k – некоторые численные параметры, выбор которых будет указан ниже. Отметим, что функция $y_n(x)$ удовлетворяет граничным условиям (2) и (3) при любых значениях a_k. Действительно, в силу условия в) и линейности граничных условий имеем $l_a(y_n) = l_a(\varphi_0) + \sum_{k=1}^{n} a_k l_a(\varphi_k) = A$, ибо $l_a(\varphi_0) = A$, а $l_a(\varphi_k) = 0$ при всех $k = 1, 2, \ldots$. Аналогично

Кафедры ПМиТП АиГ

можно показать, что $l_b(y_n) = B$ при любых $a_1, a_2, \ldots, a_n$. Действительно, $l_b(y_n) = l_b(\varphi_0) + \sum\limits_{k=1}^{n} a_k l_b(\varphi_k) = B$, так как $l_b(\varphi_0) = B$ и $l_b(\varphi_k) = 0$, $k = 1, 2, \ldots$.

Будем выбирать параметры $a_1, a_2, \ldots, a_n$ таким образом, чтобы функция $y_n(x)$ в указанном ниже смысле возможно лучше удовлетворяла дифференциальному уравнению (1).

Подставим в $F(x, y, y', y'')$ вместо y, y', y'' соответственно выражения $y_n(x)$, $y'_n(x)$, $y''_n(x)$, получим $F(x, y_n, y'_n, y''_n) = F_n(x)$. Образуем обобщенные моменты функции $F_n(x)$: $\mu_k = \int\limits_a^b F_n(x)\psi_k(x)dx$, $k = 0, 1, 2, \ldots$.

По своему смыслу момент μ_k является некоторой функцией параметров $a_1, a_2, \ldots, a_n$. Выберем их так, чтобы первые n моментов обратились в нуль:

$$\mu_k = 0, k = 0, 1, \ldots, n-1. \tag{5}$$

Эти равенства образуют систему n уравнений, из которой должны быть найдены коэффициенты $a_1, a_2, \ldots, a_n$. Пусть мы решили систему (5) и нашли значения $a_1^*, a_2^*, \ldots, a_n^*$.

В методе моментов в качестве n-го приближения к решению граничной задачи (1)-(3) принимают функцию $y_n^*(x) = \varphi_0(x) + \sum_{k=1}^{n} a_k^* \varphi_k(x)$.

Почему функцию $y_n^*(x)$ можно рассматривать как приближённое решение задачи (1)-(3)?. Поясним это наглядными соображениями. Если бы оказалось, что равенство $\mu_k = 0$ выполняются при всяких $k =$

$0, 1, 2, \ldots$, то ввиду замкнутости системы функций $\{\psi_k(x)\}$ отсюда следовало бы, что $F_n^*(x) = F_n(x, y_n^*, (y_n^*)', (y_n^*)'') = 0$ и функция $y_n^*(x)$ была бы точным решением граничной задачи (1)-(3). Но в нашем распоряжении имеется только конечное число параметров $a_1, a_2, \ldots, a_n$ и выполнить бесконечное число условий ($\mu_k = 0, k = 0, 1, \ldots$) при помощи их выбора можно только в исключительных случаях. Например, в тех случаях, когда решение граничной задачи (1)-(3) является линейной комбинацией вида (4). Можно надеяться, вообще говоря, на выполнение лишь n условий $\mu_k = 0$, $k = 0, 1, \ldots, n-1$. При этом можно ожидать, что если n достаточно большое число, то функция $F_n^*(x)$, будет близкой к нулю: $F_n^*(x) = \varepsilon_n(x)$, $\max\limits_{a \le x \le b} |\varepsilon_n(x)| < \varepsilon$ и $\varepsilon > 0$ – достаточно малое число. В этом случае функция $y_n^*(x)$ будет удовлетворять дифференциальному уравнению

$$F(x, y_n^*, (y_n^*)', (y_n^*)'') = \varepsilon_n(x), \tag{6}$$

которое мало отличается от заданного уравнения (1). Можно ожидать, что для широкого класса функций F из условия (6) будет следовать достаточная малость на $[a, b]$ разности $|y(x) - y_n^*(x)|$, где $y(x)$ – решение граничной задачи (1)-(3).

Изложим алгоритм метода применительно к граничной задаче (1)-(3):

1) в начале вычислений выбираем системы функций $\{\psi_k(x)\}$ и $\{\varphi_k(x)\}$, подчиняя их выбор сформулированным выше условиям;

2) затем решаем систему нелинейных численных уравнений (5) и выписываем её решения $a_1^*, a_2^*, \dots, a_n^*$;

3) значения параметров a_k^* подставляем в формулу (4) и полученную функцию $y_n^*(x)$ принимают за n-ое приближение к искомому решению $y(x)$.

Если уравнение (1) линейно, то запись системы уравнений для определения параметров a_k упростится.

Рассмотрим, например, уравнение $F(x, y, y', y'') = L(y) - f(y) = 0$, где $L(y) = \dfrac{d}{dx}(p(x)y') - q(x)y$, $p(x) \geq p_0 > 0$; $p(x) \in C_1[a, b]$, $q(x) \geq 0$, $a \leq x \leq b$ с граничными условиями $y(a) = A$, $y(b) = B$. В этом случае мы получим систему линейных алгебраических уравнений вида

$$\sum_{k=1}^{n} c_{ki}a_k - D_i = 0, i = \overline{0, n-1} \tag{7}$$

где $c_{ki} = \int\limits_a^b L(\varphi_k)\psi_i(x)dx$, $D_i = \int\limits_a^b [f(x) - L(\varphi_0)]\psi_i(x)dx$.

§63. Метод Галеркина

В методе Галеркина обе системы функций $\{\varphi_k(x)\}$ и $\{\psi_k(x)\}$ совпадают, поэтому система уравнений (5) из §62 примет вид:

$$\int\limits_a^b F(x, y_n, y_n', y_n'')\varphi_k(x)dx = 0, k = \overline{1, n}. \quad (1)$$

Здесь мы считаем, что свойством замкнутости обладает система функций $\{\varphi_k(x)\}$, $k = 1, 2, \ldots$.

Если исходное дифференциальное уравнение было линейным самосопряженным, то в методе Галеркина получим следующую систему уравнений:

$$\sum_{k=1}^{n} c_{ki}a_k - D_i = 0, i = 1, 2, \ldots, n, \quad (2)$$

где $c_{ki} = \int\limits_a^b L(\varphi_k)\varphi_i(x)dx$, $D_i = \int\limits_a^b [f(x) - L(\varphi_0)]\varphi_i(x)dx$.

Запись коэффициентов C_{ki} можно упростить следующим образом:

$$C_{ki} = \int\limits_a^b \left(\frac{d}{dx}(p(x)\varphi_k') - q(x)\varphi_k(x)\right)\varphi_i(x)dx = p(x)\varphi_k'(x)\varphi_i(x)|_a^b -$$

$$-\int_a^b (p(x)\varphi_k'(x)\varphi_i'(x) + q(x)\varphi_k(x)\varphi_i(x))dx =$$

$$= -\int_a^b (p(x)\varphi_k'(x)\varphi_i'(x) + q(x)\varphi_k(x)\varphi_i(x))dx,$$

так как $\varphi_k'(a)\varphi_i(a) = 0$ и $\varphi_k'(b)\varphi_i(b) = 0$. Отметим, что $C_{ki} = C_{ik}$ и матрица системы уравнений (2) симметричная. Аналогично для D_i получим $D_i = \int_a^b (p(x)\varphi_0'(x)\varphi_i'(x) + q(x)\varphi_0(x)\varphi_i(x) + f(x)\varphi_i(x))dx.$

Нетрудно показать, что система (2) однозначно разрешима. Для этого достаточно показать, что соответствующая ей однородная система $\sum_{k=1}^{n} c_{ki}a_k = 0$ имеет только нулевое решение. Выбор $\varphi_0(x)$, $\varphi_i(x)$, $i = \overline{1,n}$ удовлетворяет следующим условиям: $l_a(\varphi_0) = A$, $l_b(\varphi_0) = B$. Для этого можно $\varphi_0(x)$ определить как многочлен первой степени $\varphi_0(x) = \alpha + \beta x$ и коэффициенты α и β выбрать как решение системы уравнений

$$\alpha_0(\alpha + \beta a) + \alpha_1\beta = A,$$

$$\beta_0(\alpha + \beta b) + \beta_1\beta = B.$$

Другие функции $\varphi_k(x)$, можно, например, в случае, когда $\alpha_1 = \beta_1 = 0$, $\alpha_0 = \beta_0 = 1$ вычислять по правилу $\varphi_k(x) = (x-a)^k(x-b)$, $k = 1, 2, \ldots, n$ или $\varphi_k(x) = (x-a)(x-b)^k$, $k = 1, 2, \ldots, n$.

Чтобы иметь представление о скорости сходимости метода Галеркина хотя бы для некоторых частных задач, приведем известный результат для следующей граничной задачи:

$$y'' - q(x)y = f(x), 0 \leq x \leq 1, y(0) = y(1) = 0. \tag{3}$$

Пусть $\varphi_k(x) = \sin \pi kx$. Обозначим через $y_n^*(x) = \sum\limits_{k=1}^{n} a_k^* \varphi_k(x)$ – приближенное решение, полученное по методу Галеркина. Если $0 < q(x) < q$, $f(x)$, $q(x)$, $\dfrac{f(x)}{\sqrt{q(x)}} \in L_2[0,1]$, то имеет место оценка

$$|y(x) - y_n^*(x)| \leq \left[\frac{q}{4(n+1)^2} + \frac{\sqrt{2}}{\left(n+\frac{1}{2}\right)^{3/2}}\right] \cdot \frac{\sqrt{q}}{\sqrt{3}\pi^2} \cdot \left\| \frac{f(x)}{\sqrt{q(x)}} \right\| = O(n^{-3/2}),$$

где через $y(x)$ обозначено решение граничной задачи (3) и

$$\left\| \frac{f(x)}{\sqrt{q(x)}} \right\| = \left(\int\limits_0^1 \frac{f^2(x)}{q(x)} dx \right)^{\frac{1}{2}}.$$

§64. Методы моментов и Галеркина для операторов, заданных в гильбертовом пространстве H

Пусть линейный оператор A задан на множестве $D(A)$, всюду плотном в H со значениями, принадлежащими тому же пространству H. Предположим, что пространство H обладает счетной и полной системой линейно независимых элементов. Рассмотрим уравнение

$$Ay = f, \tag{1}$$

где y – искомый элемент из $D(A)$ подлежащий определению, f – некоторый элемент из H.

Метод Галеркина.

Выберем последовательность элементов $\{\varphi_k(x)\}$, $\varphi_k \in D(A)$. Предположим, что любой элемент y из $D(A)$ однозначно представим сходящимся рядом $y = \sum\limits_{k=1}^{\infty} a_k \varphi_k$, где a_k – некоторые коэффициенты, зависящие от y. Обозначим через y_n – n-ое приближение к решению уравнения (1) и будем разыскивать это приближение в виде

$$y_n = \sum_{k=1}^{n} a_k \varphi_k, \tag{2}$$

где a_k – параметры, которые мы определим из условия ортогональности выражения $Ay_n - f$ к элементам $\varphi_1, \varphi_2, \ldots, \varphi_n$. Будем считать, что имеют место соотношения

$$(Ay_n - f, \varphi_k) = 0, k = 1, 2, \ldots, n. \tag{3}$$

Подставляя сюда вместо y_n его выражение (2), получим систему линейных алгебраических уравнений

$$\sum_{k=1}^{n} A_{ki}a_k - B_i = 0, i = 1, 2, \ldots, n, \tag{4}$$

где $A_{ki} = (A\varphi_k, \varphi_i)$, $B_i = (f, \varphi_i)$.

Пусть система уравнений (4) однозначно разрешима и a_1^*, a_2^*, ..., a_n^* – ее решение. Тогда n-ное приближение к элементу y-решению уравнения (1) в методе Галеркина вычисляют по формуле $y_n^* = \sum\limits_{k=1}^{n} a_k^*\varphi_k$. Если бы выражение $Ay_n^* - f$ оказалось ортогональным ко всем элементам φ_k,$(k = 1, 2, \ldots, n, \ldots)$, то это бы означало, что $Ay_n^* - f = 0$ и y_n^* было бы точным решением уравнения (1). Если же условие ортогональности (3) выполнено для достаточно больших n, то можно надеяться, что y_n^* близко к точному решению уравнения (1).

В методе моментов параметры a_k определяются по-иному. Требуют выполнения условий ортогональности элемента $Ay_n - f$ к элементам последовательности $\{\psi_k\}$.

$$(Ay_n - f, \psi_k) = 0, k = 1, 2, \ldots, n. \tag{5}$$

При этом предполагается, что элементы $\psi_1, \psi_2, \ldots, \psi_n$ линейно независимы при любом конечном n и последовательность $\{\psi_k\}$ полна в H. Формула (5) есть система линейных алгебраических уравнений вида

$$\sum_{k=1}^{n} c_{ki} a_k - D_i = 0, i = 1, 2, \ldots, n, \tag{6}$$

где $c_{ki} = (A\varphi_k, \psi_i)$, $D_i = (f, \psi_i)$. Пусть a_1^*, a_2^*, ..., a_n^* – единственное решение системы уравнений (6). В методе моментов, как и в методе Галеркина, n-ое приближение к решению уравнения (1) вычисляется по формуле $y_n^* = \sum\limits_{k=1}^{n} a_k^* \varphi_k$. Легко видеть, что при $\psi_k = \varphi_k$ метод моментов совпадает с методом Галеркина. Отметим, что оба этих метода применимы и к решению нелинейных уравнений.

§65. Метод наименьших квадратов

Рассмотрим алгоритм метода наименьших квадратов применительно к операторным уравнениям.

Пусть $L_2[a, b]$ – пространство вещественных функций, интегрируемых с квадратом на $[a, b]$, и A – линейный оператор со значениями из $L_2[a, b]$. Пусть A определяет на множестве $D(A)$, всюду плотном в $L_2[a, b]$.

Предположим, что требуется решить уравнение

$$Ay = f, \tag{1}$$

где $f \in L_2[a, b]$, $y \in D(A)$. Будем считать, что уравнение (1) имеет единственное решение. Поставим в соответствие уравнению (1) функционал

$$J(y) = \|Ay - f\|^2. \tag{2}$$

Здесь знаком $\|\|$ помечена норма в пространстве $L_2[a, b]$. Если в L_2 определено скалярное произведение и $z \in L_2$, то

$$\|z\| = \sqrt{(z, z)} = \left\{ \int\limits_a^b z^2(x) dx \right\}^{1/2}.$$

Задачу отыскания решения уравнения (1) заменим эквивалентной задачей об определении элемента, минимизирующего в $D(A)$ функционал

(2). Очевидно, что $\min\limits_{y\in D(a)} J(y) = J(y^*) = 0$ и элемент y^* является решением уравнения $Ay = f$. Метод решения уравнения (1), основанный на минимизации функционала (2), называется методом наименьших квадратов.

Функционал $J(y)$ можно минимизировать различными способами. Можно это выполнить так. Выбираем последовательность линейно независимых элементов $\{\varphi_k\}$, $\varphi_k \in D(A)$. Определяем элемент y_n – n-ное приближение к решению уравнения (1) по формуле

$$y_n = \sum_{k=1}^{n} a_k\varphi_k, \tag{3}$$

где a_k – некоторые параметры. Они определяются из условия, чтобы функционал $J(y_n) = \|Ay_n - f\|^2$ принимал минимальное значение на пространстве $D_n(A) \in D(A)$, натянутом на элементы $\varphi_1, \varphi_2, \ldots, \varphi_n$. Это требование приводит к системе линейных алгебраических уравнений для определения $a_1, a_2, \ldots, a_n$:

$$\frac{\partial J(y_n)}{\partial a_k} = 0, k = \overline{1, n}. \tag{4}$$

Если $a_1^*, a_2^*, \ldots, a_n^*$ – решение системы (4), то n-ное приближение к решению y дается формулой (3) при подстановке в нее вместо величин a_k величин a_k^*.

Дадим более подробное описание метода наименьших квадратов на примере линейной граничной задачи для дифференциального уравнения второго порядка

$$L(y) = y'' + p(x)y' + q(x)y = f(x) \tag{5}$$

с граничными условиями

$$l_a(y) = \alpha_0 y(a) + \alpha_1 y'(a) = A, \tag{6}$$

$$l_b(y) = \beta_0 y(b) + \beta_1 y'(b) = B. \tag{7}$$

Будем считать, что задача (5)-(7) имеет единственное решение и это решение вместе со своими производными до второго порядка включительно непрерывно на $[a, b]$.

В методе наименьших квадратов применительно к граничной задаче (5)-(7) рассматривается система функций $\{\varphi_k(x)\}$, которая выбирается так же, как это мы делали в методе моментов.

Приближенное решение задачи (5)-(7) ищется в виде

$$y_n(x) = \varphi_0(x) + \sum_{k=1}^{n} a_k \varphi_k, \tag{8}$$

где a_k – некоторые параметры. Справедливо

$$l_a(y_n) = l_a(\varphi_0) + \sum_{k=1}^{n} a_k l_a(\varphi_k) = A, l_b(y_n) = l_b(\varphi_0) + \sum_{k=1}^{n} a_k l_b(\varphi_k) = B.$$

Функция $y_n(x)$ удовлетворяет граничным условиям (6), (7) при любых значениях параметров a_k. Выберем их так, чтобы $y_n(x)$ возможно лучше удовлетворяла уравнению (5). Подставим $y_n(x)$ в уравнение (5), получим $L(y_n) - f(x) = \varepsilon_n(x)$. Почти всегда $\varepsilon_n(x) \neq 0$. Мы заинтересованы в том, чтобы погрешность $\varepsilon_n(x)$ по абсолютной величине была возможно меньшей. Введем в рассмотрение величину δ_n^2, определив её так $\delta_n^2 = \int\limits_a^b \varepsilon_n^2(x)dx$. Параметры $a_1, a_2, \ldots, a_n$ выберем так, чтобы величина δ_n^2 принимала наименьшее значение.

Заметим, что $\delta_n^2 = I(y_n)$, если $I(y_n)$ определить по правилу $I(y) = \|L(y) - f\|^2 = (L(y) - f, L(y) - f)$. Здесь скалярное произведение имеет следующий смысл $(g, f) = \int\limits_a^b g(x)f(x)dx$ для любых функций $g, f \in L_2[a, b]$. Имеем

$$L(y_n) = L(\varphi_0) + \sum_{k=1}^{n} a_k L(\varphi_k), \delta_n^2 = \int\limits_a^b \left[\sum_{k=1}^{n} a_k L(\varphi_k) - f_0\right]^2 dx,$$

где $f_0 = f - L(\varphi_0)$.

Для отыскания искомых значений параметров записываем систему линейных алгебраических уравнений

$$\frac{1}{2}\frac{\partial\delta_n^2}{\partial a_i}=\int\limits_a^b\left[\sum_{k=1}^{n}a_kL(\varphi_k)-f_0\right]L(\varphi_i)dx=0, i=1,2,\ldots,n,$$

где $\delta_n^2=\int\limits_\infty^b\left[\sum\limits_{k=1}^{n}a_kL(\varphi_k)-f_0\right]^2dx$. Эту запись можно упростить. Имеем

$$\sum_{k=1}^{n}A_{ik}a_k-B_i=0, i=\overline{1,n}, \tag{9}$$

где $A_{ik}=A_{ki}=\int\limits_a^b L(\varphi_l)L(\varphi_i)dx$, $B_i=\int\limits_a^b f_0(x)L(\varphi_i)dx$.

Отметим, что матрица системы уравнений (9) симметрична и ее определитель есть определитель Грамма для функций $L(\varphi_1)$, $L(\varphi_2)$, $\ldots$, $L(\varphi_n)$. Предположим, что система уравнений (9) однозначно разрешима и a_1^*, a_2^*, $\ldots$, a_n^* – ее решение. Тогда n-ое приближение к решению $y(x)$ задачи (5)-(7) вычисляется по формуле (8) , если в нее вместо a_k подставлены значения a_k^*.

Выясним, когда можно надеяться *на однозначную разрешимость* системы уравнений (9). Ответ на этот вопрос зависит не только от свойств системы функций $\{\varphi_k(x)\}$, но также и от природы рассматриваемой граничной задачи (5)-(7), в частности от того, имеет ли однородная granичная задача

$$L(z) = 0, l_a(z) = 0, l_b(z) = 0 \qquad (10)$$

только нулевое решение.

Будем далее предполагать, что $\alpha_1 = \beta_1 = 0$, $\alpha_0 = \beta_0 = 1$. Пусть $D = (A_{ik})_1^n$ – матрица системы уравнений (9), а $|D|$ – ее определитель.

Теорема 1. *Если граничная задача (10) имеет только нулевое решение $z(x) = 0$, то $|D| \neq 0$ и система уравнений (9) имеет единственное решение.*

Доказательство. Рассмотрим однородную систему уравнений $\sum\limits_{k=1}^{n} A_{ik} * b_k = 0$, $i = \overline{1,n}$. Докажем, что эта система имеет только нулевое решение $b_k = 0$. Умножим уравнение номера i на b_i и просуммируем полученные выражения по i от 1 до n. Имеем

$$\sum_{i=1}^{n} b_i \sum_{k=1}^{n} A_{ik} b_k = \int_a^b \sum_{i=1}^{n} \sum_{k=1}^{n} b_i L(\varphi_i) b_k L(\varphi_k) dx = 0. \qquad (11)$$

Введем обозначение $z_n(x) = \sum\limits_{k=1}^{n} b_k \varphi_k(x)$.

Теперь из соотношения (11), получим $\int\limits_a^b L^2(z_n) dx = 0$. Отсюда следует, что $L(z_n) = 0$. В силу выбора функций $\varphi_k(x)$ имеем

$z_n(a) = z_n(b) = 0$. Действительно, $z_n(a) = \sum_{k=1}^{n} b_k\varphi_k(a) = 0$, $z_n(b) = \sum_{k=1}^{n} b_k\varphi_k(b) = 0$, так как $l_a(z_n) = \sum_{k=1}^{n} b_k l_a(\varphi_k) = 0$, $l_b(z_n) = \sum_{k=1}^{n} b_k l_b(\varphi_k) = 0$. Таким образом, функция $z_n(x)$ удовлетворяет граничной задаче (10) и по условию теоремы должно быть $z_n(x) \equiv 0$ или $\sum_{k=1}^{n} b_k\varphi_k(x) \equiv 0$. Поскольку функция $\varphi_1(x), \varphi_2(x), \ldots, \varphi_n(x)$ линейно независимы, то из условия $z_n \equiv 0$ следует, что $b_k \equiv 0$. Значит, $\mid D \mid \neq 0$ и система уравнений (9) имеет единственное решение. Теорема 1 доказана.

Изучим *вопрос о сходимости метода наименьших квадратов* на примере граничной задачи (5)–(7) в частном случае, когда

$$l_a(y) = y(a) = 0, l_b(y) = y(b) = 0. \tag{12}$$

Пусть $y^*(x)$ – точное решение задачи (5), (12), а $y_n^*(x)$ – n-ое приближение к $y^*(x)$, получаемое по методу наименьших квадратов. Заметим, что в случае граничных условий (12) $\varphi_0(x) \equiv 0$ и $y_n^*(x) = \sum_{k=1}^{n} a_k^*\varphi_k(x)$.

Теорема 2. *Последовательность функций $\{y_n^*(x)\}$, получаемых по методу наименьших квадратов, сходится в метрике L_2 к точному решению $y^*(x)$, если выполняются следующие условия:*

1. *граничная задача (5), (12) имеет единственное решение $y^*(x)$;*

2. *существует такая постоянная M, что для любой дважды непрерывно дифференцируемой функции $y(x)$, обращающейся в нуль на концах отрезка $[a,b]$, имеет место неравенство* $\| L(y) \| \geq$
$\geq \dfrac{1}{M}\left(\int_a^b y^2(x)dx\right)^{\frac{1}{2}}$.

Доказательство. В силу условия (1) и теоремы 1 последовательность решений $\{y_n^*(x)\}$ может быть построена. Последовательность $\{\varphi_k(x)\}$ полна в классе $C_2[a,b]$, поэтому для любого $\varepsilon > 0$ можно указать такое N и такие параметры $a_1, a_2, \ldots, a_N$, что $\| L(y^*) - \sum\limits_{k=1}^{N} a_k L(\varphi_k) \| \leq \dfrac{\varepsilon}{M}$. Это неравенство можно переписать в виде

$$\| f - L(y_N) \| \leq \frac{\varepsilon}{M} \tag{13}$$

так как $L(y^*) = f$, $\sum\limits_{k=1}^{N} a_k L(\varphi_k) = L\left(\sum\limits_{k=1}^{N} a_k \varphi_k\right) = L(y_N)$.

Неравенство (13) тем более будет выполняться, если в нем функцию $y_N(x)$ заменить на $y_N^*(x)$, так как последняя строится из условия минимизации квадратичной формы $I(y_N) = \| L(y_N) - f \|^2$ на множестве параметров $a_1, a_2, \ldots, a_N$. Значит, $\| L(y^*) - L(y_N^*) \| \leq \dfrac{\varepsilon}{M}$.

Очевидно неравенство $\| L(y^*) - L(y_n^*) \| = \| L(y^* - y_n^*) \| \leq \dfrac{\varepsilon}{M}$. Отсюда, используя условие 2 теоремы 2, получим

$$\| y^* - y_n^* \| \leq M \| L(y^* - y_n^*) \| < \varepsilon. \tag{14}$$

Поскольку ε может быть сколь угодно малым числом, то из (14) следует, что $\lim\limits_{n\to\infty} \| y_n^* - y^* \|$. Теорема 2 доказана.

Неравенство (14) можно переписать в виде $\| y^* - y_n^* \| \leq M \| f - L(y_n^*) \|$. По этому неравенству можно судить, с какой скоростью $y_n^*(x)$ сходится при $n \to \infty$ к точному решению $y^*(x)$.

§66. Метод Ритца

Изложим вычислительную схему метода Ритца на частном примере граничной задачи для уравнения

$$(p(x)y')' - q(x)y = f(x) \tag{1}$$

с условиями

$$y(a) = A, y(b) = B, \tag{2}$$

где $p(x) \geq p_0 > 0$, $p(x) \in C_1[a, b]$, $q(x) \geq 0$, причем $q(x)$ и $f(x) \in C[a, b]$. Будем предполагать, что задача (1), (2) имеет *единственное решение.*

Рассмотрим функционал

$$I(y) = \int_a^b [p(x)(y')^2 + q(x)y^2 + 2f(x)y]dx \tag{3}$$

и попытаемся найти его минимальное значение на классе допустимых функций, т.е. функций, удовлетворяющих двум условиям:

1. эти функции непрерывно дифференцируемы на [a, b];
2. они удовлетворяют граничным условиям (2).

Справедлива

Теорема 1. *Если функция $y^*(x)$ доставляет минимальное значение функционалу $I(y)$ среди всех допустимых функций, то она является решением граничной задачи (1), (2).*

Доказательство. Так как функция $y^*(x)$ является допустимой, то она удовлетворяет граничному условию (2) и остается доказать, что она удовлетворяет уравнению (1).

Для этой цели воспользуемся одной из теорем вариационного исчисления. Пусть рассматривается функционал вида $G(y) = \int_a^b F(x,y,y')dx$ и требуется в упомянутом выше классе допустимых функций найти наименьшее значение $G(y)$. Доказывается, что при достаточной гладкости F минимизирующая функция должна удовлетворять *дифференциальному уравнению* Эйлера $\dfrac{d}{dx}F_{y'} - F_y = 0$.

В вариационной задаче $F \equiv p(x)(y')^2 + q(x)y^2 + 2f(x)y$. Для нее $F_{y'} = 2p(x)y'$, $F_y = 2q(x)y + 2f(x)$. Поэтому уравнение Эйлера для минимизирующей функции $y^*(x)$ будет иметь вид $2\left(p(x)(y^*)'\right)' - 2q(x)y^* - 2f(x) = 0$ или $(p(x)(y^*)')' - q(x)y^* = f(x)$, что и требовалось доказать. Теорема 1 доказана.

Таким образом, задачу вычисления решения $y^*(x)$ граничной задачи (1), (2) можно заменить задачей отыскания минимума функционала (3) в классе непрерывно дифференцируемых функций на $[a,b]$, принимающих заданные значения $y(a) = A$ и $y(b) = B$ на концах отрезка. Это вторая задача в вычислительном смысле во многих случаях оказывается *предпочтительнее первой.*

В методе Ритца n-ое приближение к минимизирующей функции $y^*(x)$

вычисляется в виде

$$y_n(x) = \varphi_0(x) + \sum_{k=1}^{n} a_k \varphi_k(x) \tag{4}$$

где a_k – численные параметры, φ_k – некоторые известные функции. Выбор этих функций подчиняют условиям:

1. $\varphi_k(x) \in C_1[a,b].k = 0, 1, 2, \dots;$
2. $\varphi_0(a) = A, \varphi_0(b) = B, \varphi_k(a) = \varphi_k(b) = 0, k = 0, 1, 2, \dots;$
3. при любом конечном n функции $\varphi_1(x)$, $\varphi_2(x)$, …, $\varphi_n(x)$ линейно независимы;
4. образованное по функциям $\varphi_k(x)$ семейство функций $\{y_n(x)\}$ обладает свойством C_1 – полноты.

Отметим, что при любом выборе параметров a_k функция $y_n(x) \in C_1[a,b]$, $y_n(a) = A$, $y_n(b) = B$. Функции $y_n(x)$ являются допустимыми. Подставив $y_n(x)$ в функционал $I(y)$, получим

$$I(y_n) = A_0 + 2\sum_{k=1}^{n} A_k a_k + \sum_{i=1}^{n}\sum_{k=1}^{n} A_{ik} a_i a_k,$$

где $A_0 = \int_a^b [p(x)(\varphi_0')^2 + q(x)\varphi_0^2 + 2f(x)\varphi_0]dx$,

$$A_k = \int_a^b [p(x)\varphi_0'\varphi_k' + q(x)\varphi_0\varphi_k + f(x)\varphi_k]dx$$

$$A_{ik} = A_{ki} = \int_a^b [p(x)\varphi_k^{'}\varphi_i^{'} + q(x)\varphi_k\varphi_i]dx, i, k = 1, 2, \cdots, n.$$

Параметры $a_1, a_2, \cdots, a_n$ определяем из системы линейных алгебраических уравнений

$$\frac{1}{2}\frac{\partial I(y_n)}{\partial a_k} = \sum_{i=1}^{n} A_{ik}a_i + A_k = 0, k = 1, 2, \cdots, n. \quad (5)$$

Система уравнений (5) имеет *единственное решение*, это можно доказать аналогично тому, как доказывалась однозначная разрешимость системы (9) из § 65.

Пусть $a_1^*, a_2^*, \ldots, a_n^*$ – решение системы уравнений (5), тогда вычисляемая в методе Ритца функция $y_n^*(x) = \varphi_0(x) + \sum\limits_{k=1}^{n} a_k^*\varphi_k(x)$ будет n-ым приближением к решению $y^*(x)$ граничной задачи (1), (2).

Введем следующее обозначение

$$\lim_{n\to\infty} I(y_n^*) = m = I(y^*). \quad (6)$$

Последовательность функций $\{y_n(x)\}$ называется минимизирующей, если она удовлетворяет условию (6). Отметим, что из (6) не всегда следует то, что $\lim_{n\to\infty} y_n^*(x) = y^*(x)$. Т.е., из того, что последовательность является минимизирующей, не всегда следует ее сходимость. Сходимость имеет место только тогда, когда выполняются определенные условия.

Теорема 2. *Пусть выполняются условия:*

1. $p(x) \geq p_0 > 0,\ p(x) \in C_1[a,b]$;
2. $q(x) \geq 0$ *и* $q(x),\ f(x) \in C[a,b]$;
3. *последовательность функций* $\{y_n^*(x)\}$ *является минимизирующей. Тогда эта последовательность функций будет сходящейся равномерно на* $[a,b]$ *к* $y^*(x)$ *– решению граничной задачи (1) –(2).*

§67. Метод сеток для решения линейных граничных задач

Этот метод является наиболее универсальным как для обыкновенных дифференциальных уравнений, так и для уравнений в частных производных. Мы проведем изложение метода сеток на примере простейших линейных граничных задач для дифференциальных уравнений второго порядка.

67.1 Постановка задачи. Идея метода сеток

Пусть при $a \le b$ рассматривается граничная задача для дифференциального уравнения

$$L(y) \equiv y'' + p(x)y' + q(x)y = f(x) \tag{1}$$

с условиями

$$l_a(y) = \alpha_0 y(a) + \alpha_1 y'(a) = A, \tag{2}$$

$$l_b(y) = \beta_0 y(b) + \beta_1 y'(b) = B. \tag{3}$$

Будем считать, что граничная задача (1)–(3) имеет *единственное решение*, это решение непрерывно на $[a, b]$ и имеет *непрерывные производные* на этом отрезке до *четвертого порядка включительно*.

Метод сеток для решения граничной задачи (1)–(3), равно как и для многих других задач, состоит в следующем.

1. Область задания дифференциального уравнения (1) – отрезок $[a, b]$ заменяется некоторой дискретной (сеточной) областью. Это означает, что на отрезке $[a, b]$ выбирается некоторая система точек. Совокупность этих точек называется сеткой. Если положение каждой точки определяется по правилу $x_k = a + kh$, $k = 0, 1, \ldots, N$, – целое число, то сетку называют *равномерной*. Точки x_k называют *узлами сетки*.

2. Граничная задача (1)–(3) на множестве узлов, принадлежащих сетке, заменяются некоторой сеточной задачей. Под термином сеточная задача мы будем понимать некоторые соотношения между приближенными значениями решения граничной задачи (1)–(3) в узлах сетки. В рассматриваемом случае это будет система линейных алгебраических уравнений.

3. Полученная сеточная задача решается по какому либо численному методу и тем самым находятся приближенные значения решения граничной задачи в узлах сетки. Это и является конечной целью метода сеток.

Можно указать как на основные в методе сеток следующие вопросы:

1. Как заменить область задания дифференциальных уравнений, а в случае дифференциальных уравнений в частных производных еще и границу области некоторой сеточной областью?

2. Как заменить дифференциальное уравнение и граничные условия некоторыми сеточными соотношениями?

3. Будет ли полученная сеточная задача однозначно разрешимой, будет ли она устойчивой, сходящейся?

67.2 Методы замены обыкновенных дифференциальных уравнений и граничных условий системой алгебраических уравнений.

Рассмотрим граничную задачу (1)–(3). Выберем равномерную сетку: $x_k = a + hk,\ k = 0, 1, \ldots, N,\ h = \dfrac{b-a}{N}$. Дифференциальное уравнение (1) будем рассматривать только во внутренних узлах сетки, т.е. будем полагать, что $x = x_k,\ \ k = 0, 1, \ldots, N-1$. Граничные условия (2), (3) рассмотрим при $x_0 = a, x_N = b$.

Положим в (1) $x = x_k$:

$$L(y)_{x=x_k} = y''(x_k) + p(x_k)y'(x_k) + q(x_k)y(x_k) = f(x_k), \tag{4}$$

$$k = 0, 1, \ldots, N-1.$$

Выразим $y'(x_k)$ и $y''(x_k)$ через значения функции $y(x)$ в узлах x_{k-1}, x_k, x_{k+1}, т.е. через значения $y(x_{k-1}), y(x_k), y(x_{k+1})$. Для этой цели мы воспользуемся интерполированием и численным дифференцированием на основе интерполяционных полиномов.

Имеем

$$y'(x_k) = \frac{y(x_k) - y(x_{k-1})}{h} + r_k^{(1)}(h), \tag{5}$$

$$r_k^{(1)}(h) = \frac{h}{2} y''\left(x_k^{(1)}\right), \; x_{k-1} < x_k^{(1)} < x_k;$$

$$y'(x_k) = \frac{y(x_{k+1}) - y(x_k)}{h} + r_k^{(2)}(h), \tag{6}$$

$$r_k^{(2)}(h) = -\frac{h}{2} y''\left(x_k^{(2)}\right), \; x_k < x_k^{(2)} < x_{k+1};$$

$$y'(x_k) = \frac{y(x_{k+1}) - y(x_{k-1})}{2h} + r_k^{(3)}(h), \tag{7}$$

$$r_k^{(3)}(h) = -\frac{h^2}{6} y'''\left(x_k^{(3)}\right), \; x_{k-1} < x_k^{(3)} < x_{k+1};$$

$$y''(x_k) = \frac{y(x_{k+1}) - 2y(x_k) + y(x_{k-1})}{h^2} + r_k^{(4)}(h), \tag{8}$$

$$r_k^{(4)}(h) = -\frac{h^2}{12} y^{IV}\left(x_k^{(4)}\right), \; x_{k-1} < x_k^{(4)} < x_{k+1}.$$

Подставим в формулу (4) выражения (7) и (8) для $y'(x_k)$ и $y''(x_k)$, получим

$$L(y)_{x=x_k} = L_h(y(x_k) + R_k(h)) = f(x_k), \tag{9}$$

где $L_h(y(x_k)) = \dfrac{y(x_{k+1}) - 2y(x_k) + y(x_{k-1})}{h^2} + p(x_k)\dfrac{y(x_{k+1}) - y(x_{k-1})}{2h} +$ $+q(x_k)y(x_k)$, $R_k(h) = r_k^{(4)}(h) + p(x_k)r_k^{(3)}(h)$.

Выражение $L_h(y(x_k))$ называется *разностным оператором второго порядка*, а величина $R_k(h)$ – погрешностью аппроксимации дифференциального оператора $L(y)$ разностным оператором $L_h(y(x_k))$ на решении. Если для $R_k(h)$ выполняется условие $|R_k(h)| \le Mh^2$, $k = \overline{1, N-1}$, где M–const, не зависящая от h, то говорят, что разностный оператор L_h аппроксимирует на решении дифференциальный оператор L с погрешностью второго порядка относительно h.

Пусть h достаточно мало, тогда в формуле (9) величиной $R_k(h)$ можно пренебречь и мы получим

$$L_h(y_k) = f(x_k), \ k = 0, 1, \ldots, N-1, \tag{10}$$

где при выполнении некоторых условий можно предположить, что $y_i \approx$ $\approx y(x_i)$, $i = 0, 1, 2, \ldots, N$. Вообще говоря, всегда $y_i \ne y(x_i)$, и только при $R_k(h) = 0$ можно ожидать, что будет $y_i = y(x_i)$, $i = 0, 1, 2, \ldots, N$. Равенство (10) мы будем называть *разностной схемой, аппроксимиру-*

ющей уравнение $L(y) = f(x)$.

Отметим еще, что (10) есть система линейных алгебраических уравнений, число таких уравнений $N-1$, а матрица этой системы – *трехдиагональная*. Неизвестными являются $y_0, y_1, \ldots, y_N$. Число этих неизвестных в системе равно $N+1$.

Обратимся к граничным условиям (2), (3). Используя (6) при $k=0$, из (2) получим

$$l_a(y) = l_a^{(h)}(y(x_0)) + R_0(h) = A, \tag{11}$$

где $l_a^{(h)}(y(x_0)) = \alpha_0 y(x_0) + \alpha_1 \left(\dfrac{y(x_1) - y(x_0)}{h}\right)$, $R_0^{(}h) = \alpha_1 r_0^{(2)}(h)$.

Аналогично, используя (5) при $k=N$, из (3) получим

$$l_b(y) = l_b^{(h)}(y(x_N)) + R_N(h) = B, \tag{12}$$

где $l_b^{(h)}(y(x_N)) = \beta_0 y(x_N) + \beta_1 \left(\dfrac{y(x_N) - y(x_{N-1})}{h}\right)$, $R_N(h) = \beta_1 r_N^{(1)}(h)$.

При достаточно малом h величинами $R_0(h)$ и $R_N(h)$, *имеющими первый порядок малости относительно* h, в выражениях (11) и (12) можно пренебречь. Тогда вместо (11) и (12) будем иметь

$$l_a^{(h)}(y_0) = A. \tag{13}$$

$$l_b^{(h)}(y_N) = B. \tag{14}$$

Операторы $l_a^{(h)}(y_0)$ и $l_b^{(h)}(y_N)$ аппроксимируют соответственно граничные операторы $l_a(y)$ и $l_b(y)$ *с погрешностью* $O(h)$. Формулы (10), (13), (14) образуют в совокупности систему $N+1$ линейных алгебраических уравнений с неизвестными $y_0, y_1, \ldots, y_N$. В методе сеток эту систему решают по какому-либо численному алгоритму и после этого полагают $y(x_k) = y_k$, $k = 0, 1, \ldots, N$. Обратим внимание, что граничные условия (2), (3) при необходимости можно аппроксимировать разностными условиями с погрешностью второго порядка малости относительно. Для этого достаточно, например, воспользоваться вместо (5) и (6) следующими формулами:

$$y'(x_0) = \frac{-y(x_2) + 4y(x_1) - 3y(x_0)}{2h} + O\left(h^2\right),$$

$$y'(x_N) = \frac{3y(x_N) - 4y(x_{N-1}) + y(x_{N-2})}{2h} + O\left(h^2\right).$$

<u>О разрешимости систем разностных уравнений.</u>

В методе сеток приближенное решение граничной задачи (1)-(3) вычисляется как решение системы линейных алгебраических уравнений, образуемых формулами (10), (13), (14). *Выясним условия, при которых такая система уравнений будет однозначно разрешимой.* С целью упрощения доказательств будем считать, что в граничных условиях (2),

(3) $\alpha_0 = \beta_0 = 1$, $\alpha_1 = \beta_1 = 0$. В этом случае сеточная задача (10), (13), (14) может быть записана в таком виде:

$$L_h(y_k) = \frac{1}{h^2}\Lambda_h(y_k) = f_k, k = \overline{1, N-1}, y_0 = A, y_N = B, \qquad (15)$$

где $\Lambda_h(y_k) = A_k y_{k+1} + 2B_k y_k + C_k y_{k-1}$, $A_k = 1 + \frac{h}{2}p_k$, $2B_k = -2 + h^2 q_k$, $C_k = 1 - \frac{h}{2}p_k$, $p_k = p(x_k)$, $q_k = q(x_k)$, $f_k = f(x_k)$.

Перепишем систему уравнений (15) в ином виде, нормировав уравнения для внутренних точек сетки. Получим

$$\Lambda_h(y_k) = h^2 f_k, k = \overline{1, N-1}, y_0 = A, y_N = B. \qquad (16)$$

Чтобы доказать однозначную разрешимость системы уравнений (16), достаточно показать, что соответствующая однородная система уравнений

$$\Lambda_h(z_k) = 0, k = \overline{1, N-1}, z_0 = 0, z_N = 0 \qquad (17)$$

имеет только нулевое решение, т.е. $z_k \equiv 0$. Тогда определитель системы (17) отличен от нуля, и так как он совпадает с определителем системы (16), то последняя будет однозначно разрешима в силу правила Крамера.

Доказательство разрешимости системы уравнений (16) будем вести при двух предположениях относительно шага сетки h и коэффициента $q(x)$ уравнения (1). Будем считать, что выполняются условия:

$$1)\ \frac{h}{2}\max_{a \le x \le b}|p(x)| < 1, \qquad (18)$$

$$2)\ q(x) \le 0 \text{ при } a \le x \le b. \tag{19}$$

Предположим, что заданы некоторые числа Y_0, Y_1, $\dots$, Y_N, среди которых есть неравные. Имеет место

Лемма 1. *Пусть выполняются условия (18), (19) и*

$$\Lambda_h(Y_k) \ge 0, (\Lambda_h(Y_k) \le 0), \tag{20}$$

при $k = 1, 2, \dots, N-1$. Среди чисел Y_0, Y_1, $\dots$, Y_N наибольшее положительное значение (наименьшее отрицательное значение) принимает или Y_0 или Y_N.

Эту лемму часто называют *принципом максимума.*

Доказательство. Будем доказывать только первое утверждение при условии, что имеет место неравенство $\Lambda_h(Y_k) \ge 0$, $k = 1, 2, \dots, N-1$.

Предположим противное и будем считать, что наибольшее положительное значение числа Y_k принимают при некотором k $(k = \overline{1, N-1})$, и пусть $Y_s = M > 0$ есть одно из таких значений, для которого Y_{s+1} или Y_{s-1} строго меньше M.

Тогда $\Lambda_h(Y_s) = A_s Y_{s+1} + 2b_s y_s + C_s Y_{s-1} < (A_s + C_s)M + 2B_s M$, ибо A_s и C_s строго положительны в силу условия (18). В силу условия (19) $\Lambda_h(Y_s) < (A_s + C_s + 2B_s)M = h^2 q_s M \le 0$.

Действительно, $A_s = 1 + \dfrac{h}{2}p_s$, $2B_s = -2 + h^2 q_s$, $C_s = 1 - \dfrac{h}{2}p_s$. $A_s +$ $+ C_s + 2B_s = 1 + \dfrac{h}{2}p_s + 1 - \dfrac{h}{2}p_s - 2 + h^2 q_s = h^2 q_s$.

Таким образом, имеем $\Lambda_h(Y_s) < 0$. Но это противоречит условию (20). Значит, наше предположение неверно и наибольшее положительное значение будет иметь Y_0 или Y_N. Аналогично доказывается утверждение леммы и в случае условия $\Lambda_h(Y_k) \leq 0$. Лемма доказана.

Теорема 1. *Пусть выполняются условия (18), (19). Тогда система уравнений (17) имеет только нулевое решение.*

Доказательство. В силу формулы (17) $\Lambda_h(Z_k) = 0$, $k = 1, 2, \ldots, N - 1$. Поэтому можно считать, что выполняются оба условия (20) леммы, т.е. $\Lambda_h(Z_k) \geq 0$ и $\Lambda_h(Z_k) \leq 0$. В этом случае по лемме наибольшее положительное значение и наименьшее отрицательное значение среди чисел Z_0, Z_1, $\ldots$, Z_N будет иметь числа Z_0 и Z_N, но в уравнениях (17) $Z_0 = Z_N = 0$, значит, все числа $Z_k = 0$. Теорема доказана.

Из теоремы следует однозначная разрешимость системы уравнений (16).

§68. Оценка погрешности и сходимость метода сеток

Эти вопросы мы будем изучать на примере задачи (1)-(3) из §67. Будем считать, что граничные условия (2), (3) из §67 заданы в форме $y(a) = A$, $y(b) = B$.

Учитывая (9) и (15) (см. §67), запишем сеточную задачу для точных значений граничной задачи (1)-(3) из предыдущего параграфа в следующем виде:

$$\frac{1}{h^2}\Lambda_h(y(x_k)) + R_k(h) = f(x_k), k = \overline{1, N-1}, y(x_0) = A, y(x_N) = B. \quad (1)$$

В случае, когда решение граничной задачи (1)-(3) (см. §67) непрерывно и имеет непрерывные на $[a, b]$ производные до четвертого порядка включительно, для погрешности аппроксимации $R_k(h)$ верна оценка

$$|R_k(h)| \leq Mh^2, k = 1, 2, \ldots, N-1, \quad (2)$$

где M – постоянная, не зависящая от h.

Реально из-за наличия погрешностей округления сеточные значения будут вычисляться как решение системы линейных алгебраических уравнений

$$\frac{1}{h^2}\Lambda_h(y_k) = f_k - \delta_k, k = \overline{1, N-1}, y_0 = A, y_N = B, \quad (3)$$

где через δ_k обозначены погрешности округлений. Будем считать, что для δ_k указана некоторая оценка сверху:

$$|\delta_k| \leq \delta, k = 1, 2, \ldots, N-1. \quad (4)$$

Введем следующие обозначения: $\varepsilon_k = y(x_k) - y_k$, $\varepsilon(h) = \max\limits_{0\le k\le N} |\varepsilon_k|$.

Метод сеток, определяемый формулой (3), называется равномерно сходящимся, если при $h \to 0$ выполняется условие: $\varepsilon(h) \to 0$.

Дадим оценку для $|\varepsilon_k|$ и установим условия сходимости метода сеток. Используя формулы (1) и (3), получим

$$\Lambda_h(\varepsilon_k) = h^2(-R_k(h) + \delta_k), k = 1, 2, \ldots, N-1, \varepsilon_0 = 0, \varepsilon_N = 0. \quad (5)$$

Действительно,

$$\frac{\frac{1}{h^2}\Lambda_h(y(x_k) + R_k(h)) = f_k}{\frac{1}{h^2}\Lambda_h(y_k) = f_k - \delta_k},$$

$$\frac{1}{h^2}\Lambda_h(\varepsilon_k) + R_k(h) = \delta_k,$$

$$\Lambda_h(\varepsilon_k) = +h^2(-R_k(h) + \delta_k).$$

Метод получения оценки для $|\varepsilon_k|$ будет заключаться в построении некоторой функции $u(x)$, которая будет мажорантой для $|\varepsilon_k|$, т.е. будет выполнятся условие $|\varepsilon_k| \le u(x_k)$, $u(x) \ge 0$, $k = 0, 1, \ldots, N$.

Для построения мажоранты нам понадобится следующая лемма

Лемма 1. *Пусть $t_0, t_1, \ldots, t_N$ и $s_0, s_1, \ldots, s_N$ – некоторые числа, причем $s_k \ge 0$, $k = 0, 1, \ldots, N$. Пусть выполняются неравенства (18), (19) из §67 и*

$$1)\ \Lambda_h(s_k) \le -|\Lambda_h(t_k)|, k = \overline{1, N-1},$$

Кафедры

ПМиТП

АиГ

$$2)\ |t_0| \le s_0, |t_N| \le s_N.$$

Тогда $|t_k| \le s_k$, $k = \overline{1, N-1}$.

Доказательство. Рассмотрим сумму $s_k + t_k$. В силу первого условия леммы имеем $\Lambda_h(s_k + t_k) = \Lambda_h(s_k) + \Lambda_h(t_k) \le 0$.

По лемме 1 из §67 среди чисел $s_k + t_k$ наименьшее отрицательное значение могут иметь или $s_0 + t_0$, или $s_N + t_N$. Но по второму условию леммы эти числа неотрицательны, поэтому все $s_k + t_k \ge 0$. Аналогично можно показать, что также и $s_k - t_k \ge 0$. Последние два неравенства означают, что $-s_k \le t_k \le s_k$ или $|t_k| \le s_k$. Лемма доказана.

Рассмотрим вспомогательную задачу

$$L(v) \equiv v'' + p(x)v' + q(x)v = -1, v(a) = 0, v(\varepsilon) = 0, \qquad (6)$$

где $q(x) < 0$. Ее решение обозначим через $v(x)$. Покажем, что $v(x) > 0$ для всех $a < x < b$. Действительно, если бы функция $v(x)$ принимала на $[a, b]$ отрицательные значения, то существовала бы точка $\overline{x}$ такая, что $v(\overline{x}) < 0$, $v'(\overline{x}) = 0$ и $v''(\overline{x}) > 0$. Но тогда $L(v)|_{x=\overline{x}} \equiv v''(\overline{x}) + p(\overline{x})v'(\overline{x}) + q(\overline{x})v(\overline{x}) > 0$, что противоречит (6), ибо $L(v(x)) = -1$ при всех $a < x < b$.

Определим $u(x)$ по формуле $u(x) = \alpha v(x)$, где α - положительное число, правило выбора которого мы укажем позже.

Имеем $\Lambda(v(x_k)) = h^2(-1 - R_k(v, h))$, где через $R_k(v, h)$ мы обозначили погрешность аппроксимации дифференциального оператора $L(v)$

разностным $L_h(v(x_k))$. Очевидно, что

$$|R_k(v,h)| \le M^* h^2, \tag{7}$$

где M^* - постоянная, не зависящая от h. Запишем сеточную задачу для функции $u(x)$:

$$\Lambda_h(u(x_k)) = -\alpha h^2(1 + R_k(v,h)), k = \overline{1, N-1}, u(x_0) = 0, u(x_N) = 0. \tag{8}$$

Сравним задачи (5) и (8). Если нам удастся определить α так, чтобы было

$$\Lambda_h(u(x_k)) \le -|\Lambda_h(\varepsilon_k)| \tag{9}$$

при $k = \overline{1, N-1}$, то тогда в силу леммы будет выполняться неравенство $|\varepsilon_k| \le |u(x_k)|$. Используя оценку (2), имеем

$$|\Lambda_h(\varepsilon_k)| \le h^2(Mh_\delta^2). \tag{10}$$

Пусть h настолько мало, что $1 - M^* h^2 > 0$. Очевидно, что

$$\Lambda_h(u(x_k)) \le -\alpha h^2(1 - m^* h^2). \tag{11}$$

Из (10) и (11) следует, что неравенство (9) будет выполняться, если α выбрано из условия $-\alpha h^2(1 - M^* h^2) \le -h^2(Mh^2 + \delta)$.

Положим $\alpha = \dfrac{Mh^2 + \delta}{1 - M^* h^2}$, тогда

$$|\varepsilon_k| \le u(x_k) = \alpha v(x_k) = \frac{Mh^2 + \delta}{1 - M^* h^2} v(x_k), k = 0, 1, \ldots, N. \tag{12}$$

Оценка (12) является искомой оценкой погрешности метода сеток, определяемого формулой (3). Пусть $V \geq \max\limits_{a \leq x \leq b} v(x)$, тогда на основании (12) можно записать, что

$$\varepsilon(h) \leq \frac{Mh^2 + \delta}{1 - M^* h^2} V. \tag{13}$$

Из этого неравенства видно, что $\varepsilon(h) \to 0$ при $h \to 0$ со скоростью $O(h^2)$, если $\delta \leq \delta_0 h^2$. Таким образом, установлена оценка погрешности (12), равномерная сходимость и скорость сходимости метода сеток. Доказана

Теорема 1. *Пусть решение граничной задачи (1)-(3) из предыдущего параграфа с условиями частного вида $y(a) = A$, $y(b) = B$, единственно и непрерывно дифференцируемо на $[a, b]$ до четвертого порядка включительно. Если выполняются условия (18), (19) (см. §67), то метод сеток для рассматриваемой задачи будет равномерно сходящимся со скоростью $O(h^2)$, если $\delta \leq \delta_0 h^2$.*

Теорема имеет место для рассматриваемой граничной задачи и при более общих граничных условиях, в частности она верна в случае, когда $l_a(y) = \alpha_0 y(a) + \alpha_1 y'(a) = A$, $l_b(y) = \beta_0 y(b) + \beta_1 y'(b) = B$, $\alpha_0 \alpha_1 \leq 0$ и $\beta_0 \beta_1 \geq 0$.

§69. Сходимость и аппроксимация разностных схем

Пусть D - некоторая область изменения независимых переменных x, y, ограниченная контуром Γ. Рассмотрим уравнение

$$L(u) \equiv a(x,y)\frac{\partial^2 u}{\partial x^2} + 2b(x,y)\frac{\partial^2 u}{\partial x \partial y} + c(x,y)\frac{\partial^2 u}{\partial y^2} + \\ +2d(x,y)\frac{\partial u}{\partial x} + 2e(x,y)\frac{\partial u}{\partial y} + g(x,y)u = f(x,y), \tag{1}$$

где a, b, c, d, e, g – коэффициенты, а f – свободный член. Эти функции известны, их считают определенными в замкнутой области $D + \Gamma$.

Решение $u(x,y)$ подлежит определению, (1) – линейное дифференциальное уравнение второго порядка.

Пусть u - есть решение дифференциального уравнения

$$L(u) = f, \tag{2}$$

заданного в некоторой области D, ограниченной контуром Γ.

Рассмотрим некоторое множество $D_h = M_h$, состоящее из изолированных точек M_h, принадлежащих области $\overline{D} = D + \Gamma$. Число точек во множестве D_h будем характеризовать величиной h, чем меньше h, тем большим будет число точек в D_h. Множество D_h называется *сеткой*, а точки M_h из этого множества - *узлами сетки*. Функция, определенная в узлах сетки, называется *сеточной функцией*.

Обозначим через U пространство непрерывных в D функций $v(x,y)$. Через U_h обозначим пространство, образованное совокупностью сеточных функций $v_h(x,y)$, определенное на D_h. В методе сеток осуществляется замена пространства U на пространство U_h.

Пусть $u(x,y)$ – точное решение уравнения (2) и $u(x,y) \in U$. Поставим задачу: вычислить значения $u_h(x,y)$. Эти значения в совокупности образуют таблицу, в которой число значений равно числу точек в D_h. Точно поставленную задачу удается решить лишь в нескольких случаях. Как правило, можно вычислить некоторые сеточные значения $u^{(h)}$, относительно которых при выполнении дополнительных условий (о которых мы еще скажем) можно утверждать, что $u^{(h)} = u_h(x,y)$.

Величины u_h называются *приближенными сеточными значениями решения* $u(x,y)$. Для их вычисления строят систему численных уравнений, которую мы будем записывать в виде:

$$L_h\left(u^{(h)}\right) = f^{(h)}, \tag{3}$$

где L_h есть разностный оператор, соответствующий оператору L, $f^{(h)} \in$ $\in F_h$. Если $f(x,y) \in F$, то F_h образуется по F аналогично тому, как U_h образовывалось по U. Формулу (3) будем называть *разностной схемой.*

Пусть в линейных пространствах U_h и F_h введены соответственно нормы $\|\cdot\|_{U_h}$ и $\|\cdot\|_{F_h}$, которые являются сеточными аналогами норм $\|\cdot\|_U$ и $\|\cdot\|_F$ в исходных пространствах U и F.

Определение 1. *Будем говорить, что разностная схема (3) является сходящейся, если при $h \to 0$ выполняется условие $\|u_h(x,y) - u\|_{u_h} \to 0$.*

Если выполняется условие $\|u_h(x,y) - u^{(h)}\|_{u_h} \le ch^5$, где c - постоянная, не зависящая от h, $s > 0$, то говорят, что сходимость имеет место со скоростью порядка s относительно h.

Определение 2. *Говорят, что разностная схема (3) аппроксимирует задачу (2) на решении $u(x,y)$, если $L_h(u_h(x,y)) = f^{(h)} + \delta f^{(h)}$ и $\|\delta f^{(h)}\|_{F_h} \to 0$ при $h \to 0$. Величина $\delta f^{(h)}$ называется погрешностью аппроксимации.*

Если $\|\delta f^{(h)}\|_{F_h} \le Mh^\sigma$, где M– постоянная, не зависящая от h, $\sigma > 0$, то говорят, что разностная схема (3) аппроксимирует задачу (2) на решении $u(x,y)$ с погрешностью порядка σ относительно h.

Введем важное понятии устойчивости разностной схемы (3).

Путь L_h – линейный разностный оператор.

Определение 3. *Разностная схема (3) называется устойчивой, если существует такое $h_0 > 0$, что для всех $h < h_0$ и любых $f^{(h)} \in F_h$ выполняются условия:*

1. *разностная схема (3) имеет единственное решение;*
2. $\|u^{(h)}\|_{U_h} \le M\|f^h\|_{F_h}$, *где M– постоянная, не зависящая от h и $f^{[h]}$.*

Из этого определения следует, что устойчивость разностной схемы существенным образом связывается с понятием норм, вводимых в пространствах U_h и F_h.

Возможны случаи, когда условие 2) будет выполняться для некоторых норм и не будет выполняться для других норм. В каждом отдельном случае важно выяснить, по какой причине для рассматриваемой задачи условие 2) не выполняется. Если оно не выполняется из-за того, что мы неудачно ввели определение норм в пространствах U_h и F_h, то следует испробовать другие способы введения норм в U_h и F_h с тем, чтобы добиться выполнения условия 2). Но если условие 2) не выполняется ни при каком выборе норм, то это означает, что разностная задача неустойчива. Отметим, что устойчивость является внутренним свойством разностной схемы, поэтому на устойчивость можно влиять, изменяя некоторые параметры, входящие в разностную схему, или вообще видоизменяя всю разностную схему.

Связь аппроксимации и устойчивости со сходимостью.

Между введенными выше понятиями сходимости, аппроксимации и устойчивости существует связь. Она состоит в том, что из аппроксимации и устойчивости следует сходимость.

Теорема 1. *Пусть разностная схема* $L_h\left(u^{(h)}\right) = f^{(h)}$ *аппроксимирует задачу* $L(u) = f$ *на решение* $u(x, y)$ *с порядком* $s > 0$ *относительно* h *и устойчива. Тогда эта схема будет сходящейся и порядок* *ее сходимости будет совпадать с порядком аппроксимации, т.е. будет справедлива оценка*

$$\|u_h(x,y) - u^{(h)}\|_{U_h} \leq Kh^s, \tag{4}$$

где K – постоянная, не зависящая от h.

Доказательство. По определению аппроксимации имеем $\|\delta f^{(h)}\|_{F_h} \leq$ $\leq ch^s$, где $\delta f^h = L_h(u_h(x,y)) - f^{(h)}$. Обозначим $\varepsilon_h(x,y) = u_h(x,y) - u^{(h)}$. Легко видеть, что в силу линейности L_h для $\varepsilon_h(x,y)$ имеет место формула $L_h(\varepsilon_h(x,y)) = \delta f^{(h)}$. Действительно,

$$\delta f^{(h)} = L_h(u_h(x,y)) - f^{(h)} = L_h(u_h(x,y)) - L_h(u^{(h)}) = L_h(u_h(x,y) - u^{(h)}) =$$

$$= L_h(\varepsilon_h(x,y)).$$

Отсюда, используя определение устойчивости, получим $\|\varepsilon_h(x,y)\|_{U_h} \leq$ $\leq M\|\delta f^{(h)}\|_{F_h} \leq M(ch^s) = Kh^s$, где $K = Mc^h$. Таким образом, оценка (4) установлена и теорема доказана. Итак, при изучении вопросов построения и изучения разностных схем, следует поступать так:

1. Сначала указывается правило выбора сетки, т.е. указывается правило замены области D и контура Γ некоторой сеточной областью.

2. Указывается и строится конкретно одна или несколько разностных схем. Проверяется условие аппроксимации разностных схем и устанавливается порядок аппроксимации.

3. Доказывается устойчивость построенных разностных схем. Если разностные схемы обладают аппроксимацией и устойчивостью, то о сходимости разностных схем судят по доказанной выше теореме.

4. Рассматривается вопрос численного решения разностных схем. В случае линейных разностных схем это будет система линейных алгебраических уравнений. Уже в двумерном случае порядок этих систем может быть очень большим. Это делает задачу численного решения упомянутых систем во многих случаях весьма трудной. Поэтому для решения систем уравнений, возникающих в методе сеток, разработаны и разрабатываются специальные методы решения, учитывающие особенности таких задач.

§70. Однородные разностные схемы для дифференциальных уравнений

Известно, что дифференциальные уравнения математической физики являются следствиями интегральных законов сохранения.

Определение 1. *Разностные схемы, удовлетворяющие этому требованию, называются консервативными.*

Оказалось, что при одном и том же числе точек сетки консервативные разностные схемы более правильно отражают поведение решения исходной задачи, чем неконсервативные схемы.

Определение 2. *Будем говорить, что разностная схема однородна, если ее коэффициенты во всех узлах сетки вычисляются по одним и тем же формулам.*

Вид однородных разностных схем не зависит ни от выбора конкретной задачи из данного класса, ни от выбора разностной сетки. Во всех узлах сетки для любой задачи из данного класса разностные уравнения имеют один и тот же вид. Коэффициенты однородной разностной схемы определяются как функционалы коэффициентов дифференциального уравнения.

Однородные трехточечные разностные схемы.

Рассмотрим граничную задачу для обыкновенного дифференциального уравнения 2-го порядка

$$L(u)=\frac{d}{dx}\left(p(x)\frac{du}{dx}\right)-q(x)u=-f(x),0<x<1, \quad (1)$$

где $p(x)\geq p_0>0, q(x)\geq 0, u(0)=\mu_1, u(1)=\mu_2$.

Это стационарное уравнение теплопроводности. Задача (1) имеет единственное решение, если $p(x), q(x), f(x)$– кусочно-непрерывные функции.

На отрезке $[0,1]$ введем равномерную сетку $\overline{\omega_h}=\{x_i=ih, i=\overline{0,N}\}$ с шагом $h=\dfrac{1}{N}$ и выберем трехточечный шаблон x_{i-1}, x_i, x_{i+1}, на котором и будем писать разностную схему, аппроксимирующую задачу (1). Любые разностные уравнения на этом шаблоне имеют вид

$$b_iy_{i+1}-c_iy_i+a_iy_{i-1}=-h^2\phi_i, \quad (2)$$

где a_i, b_i, c_i – коэффициенты, зависящие от $p(x), q(x), h$. Они пока не определены. Перепишем (2) иначе

$$\frac{1}{h}\left(b_i\frac{y_{i+1}-y_i}{h}-a_i\frac{y_i-y_{i-1}}{h}\right)-d_iy_i=-\phi_i, \quad (3)$$

где $d_i=\dfrac{(c_i-a_i-b_i)}{h^2}$.

Полученная разностная схема (3) однородна, так как ее коэффициенты во всех узлах сетки вычисляются по одним и тем же формулам. Так, если ввести в рассмотрение функционалы $A[\overline{p}(s)]$, $B[\overline{p}(s)]$, $D[\overline{p}(s)]$,

$F[\overline{f}(s)]$, определенные для любых кусочно-непрерывных функций на отрезке $-1 \leq s \leq 1$ и вычислить коэффициенты схемы по формулам $a_i = A[p(x_i+sh)]$, $b_i = B[p(x_i+sh)]$, $d_i = D[p(x_i+sh)]$, $\phi_i = F[f(x_i+sh)]$, $\overline{p}(s) = p(x_i + sh)$, то такая схема будет однородной.

Приведем простейшие формулы:

$$A[\overline{p}(s)] = \overline{p}(-0,5) = p(x_i - 0,5h), a_i = p_{i-\frac{1}{2}} = p(x_i - 0,5h),$$

$$F[\overline{f}(s)] = f(0)], \phi_i = f_i = f(x_i), b_i = p_{i+\frac{1}{2}} = p(x_i + 0,5h), d_i = q_i.$$

Если система однородна, то удобно пользоваться безиндексной системой обозначений:

$$\Lambda_y = \frac{1}{h}(by_x - ay_{\overline{x}}) - dy = -\phi, x \in \omega_h, y(0) = \mu_1, y(1) = \mu_2, \qquad (4)$$

где $a = a(x)$, $b = b(x)$, $y = y(x)$, $x = ih \in \omega_h$, $y_x = \dfrac{[y(x+h) - y(x)]}{h}$, $y_{\overline{x}} = \dfrac{[y(x) - y(x-h)]}{h}$.

Для разрешимости задачи (4) достаточно, чтобы выполнялось $a > 0, b > 0, d \geq 0$. При этом решение можно найти методом прогонки. Полученная схема имеет 2-ой порядок аппроксимации, если выполняется условия

$$\frac{b-a}{h} = p'(x) + O(h^2), \frac{b+a}{h} = p(x) + O(h^2),$$

$$d(x) = q(x) + O(h^2), \phi = f(x) + O(h^2).$$

§71. Интегро-интерполяционный метод

Основные способы построения разностных схем: интегро - интерполяционный метод, метода Ритца, метод Галеркина, методы аппроксимации квадратичного функционала и сумматорного тождества. Обычно дифференциальное уравнение выражает некоторый физический закон сохранения. Этот закон можно записать в интегральной форме для интервала (ячейки) сетки – уравнение баланса. Дифференциальное уравнение получается из уравнения баланса при стремлении шага сетки к нулю в предположении существования непрерывных производных, входящих в уравнение. Входящие в уравнение баланса интегралы и производные следует заменить приближенными выражениями на сетке. В результате получим однородную схему. Такой метод называется интегро-интерполяционным методом или методом баланса.

Проиллюстрируем его на примере задачи

$$(pu')' - qu = -f(x) 0 < x < 1,$$

$$(pu') - \sigma_1 u = -\mu_1 x = 0 u(1) = \mu_2. \tag{1}$$

Напишем уравнение баланса тепла на $0 \le x \le 1$.

$$\omega_{i+\frac{1}{2}} - \omega_{i-\frac{1}{2}} + \int\limits_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} f(x)dx = \int\limits_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q(x)u(x)dx, \tag{2}$$

где $\omega = pu'(-\omega(x)$ – поток тепла, $q(x)u(x)$ – мощность стоков тепла, пропорциональная температуре, $f(x)$ – плотность распределения внешних источников тепла).

Чтобы получить из (2) трехточечное разностное уравнение, заменим $\omega_{i-\frac{1}{2}}, \omega_{i+\frac{1}{2}}$ и интегралы в уравнении (2) линейной комбинацией значений подынтегральных функций в узлах сетки (x_{i-1}, x_i, x_{i+1}), например, $\frac{1}{h}\int\limits_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q(x)u(x)dx \approx d_i u_i, d_i = \frac{1}{h}\int\limits_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} q(x)dx.$

Проинтегрируем равенство $u' = \dfrac{\omega}{p}$ по x от x_{i-1} до x_i:

$$u_i - u_{i-1} = \int\limits_{x_{i-1}}^{x_i} \frac{\omega}{p(x)}dx \approx h\omega_{i-\frac{1}{2}}\frac{1}{a_i}, \qquad (3)$$

где $a_i = \left[\dfrac{1}{h}\int\limits_{x_{i-1}}^{x_i}\dfrac{dx}{p(x)}\right]^{-1}, \Rightarrow \omega_{i-\frac{1}{2}} = \dfrac{a_i(u_i - u_{i-1})}{h}, \omega_{i+\frac{1}{2}} = \dfrac{a_{i+1}(u_{i+1} - u_i)}{h}.$

В результате получим из (2) *однородную консервативную схему*

$$\frac{1}{h}[a_{i+1}\frac{y_{i+1} - y_i}{h} - a_i\frac{y_i - y_{i-1}}{h}] - d_i y_i = -\varphi_i,$$

где $\varphi_i = \dfrac{1}{h}\int\limits_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} f(x)dx$, y – искомое неизвестное, т. е. u.

Напишем *разностную аппроксимацию для краевого условия* третьего рода $(pu')_0 = \sigma_1 u_0 - \mu_1$ при $x = 0$.

Если схемы выражают на сетке соответствующий закон сохранения, то они называются *консервативными (дивергентными)*. Схемы, нарушающие закон сохранения, называются *неконсервативными* или *дисбалансными*.

Для этого воспользуемся уравнением баланса при $0 \le x \le x_{\frac{1}{2}} = \dfrac{h}{2}$,

$$\omega_{\frac{1}{2}} - \omega_0 - \int\limits_0^{x_{\frac{1}{2}}} qu dx = -\int\limits_0^{x_{\frac{1}{2}}} f(x) dx.$$

Подставим сюда из (3) $\omega_{\frac{1}{2}} = a_1 u_{\overline{x},1}$, $\omega_0 = (pu')_0$, $\int\limits_0^{x_{\frac{1}{2}}} qu dx \approx q_0 u_0 \dfrac{h}{2}$, $\int\limits_0^{x_{\frac{1}{2}}} f(x) dx \approx f_0 \dfrac{h}{2}$ и заменяя всюду u на y, получим разностное краевое условие

$$a_1 y_{\overline{x},1} - \sigma_1 y_0 + \mu_1 - \frac{h q_0 y_0}{2} = -\frac{h f_0}{2}. \quad (4)$$

Разностное краевое условие (4) аппроксимирует условие с точностью $O(h^2)$.

§72. Метод аппроксимации квадратичного функционала

Краевая задача

$$Lu = (pu')' - qu = -f(x), 0 < x < 1,$$

$$u(0) = 0, u(1) = 0, p(x) \geq p_0 > 0, q(x) \geq 0 \qquad (1)$$

эквивалентна задаче об отыскании минимизирующего элемента квадратичного функционала $J[u] = \int\limits_0^1 (p(u')^2 + qu^2)dx - \int\limits_0^1 fudx$ на множестве функций $u(x)$, заданных на $[0,1]$ и удовлетворяющих условиям $u(0) = u(1) = 0$.

Введем на $[0,1]$ сетку $\overline{\omega}_h = \{x_i = ih, i = 0, 1, \ldots, N\}$ и аппроксимируем функционал. Для этого сначала представим его в виде суммы интегралов по интегралам сетки $J[u] = \sum\limits_{i=1}^{N} J_i[u]$, $J_i[u] = \int\limits_{x_{i-1}}^{x_i} (p(u')^2 + qu^2 - 2fu)dx$, после чего аппроксимируем J_i так: $\int\limits_{x_{i-1}}^{x_i} p(u')^2dx \approx a_i(u_{\overline{x},1})^2 h$, $u_{\overline{x_i}} = \dfrac{u_i - u_{i-1}}{h}$, $\int\limits_{x_{i-1}}^{x_i} (qu^2 - 2fu)dx \approx \dfrac{h}{2}\{[qu^2 - 2gu]_i + (qu^2 - 2fu)_{i-1}\}$, где a_i – некоторый коэффициент, например, $a_i = \dfrac{1}{h}\int\limits_{x_{i-1}}^{x_i} p(x)dx$. В результате получим функционал $J_h[y] = \sum\limits_{k=1}^{N} ha_k(y_{\overline{x},k})^2 + \sum\limits_{k=1}^{N-1} (q_k y_k^2 - 2f_k y_k)h$, где

$y_i = y(i)$ – произвольная сеточная функция, обращающаяся в нуль при $i = \overline{0, N}$. Уравнение или $Ay = \varphi$, или, $\sum\limits_{i=1}^{N} a_{ij} y_j = \varphi_i, i = \overline{1, N}, A = A^* > 0$ имеет решение, минимизирующее функционал

$$J_A[y] = (Ay, y) - 2(\varphi, y) = \sum_{i,j=1}^{N} a_{ij} y_j y_i - 2 \sum_{i=1}^{N} \varphi_i y_i.$$

В этом нетрудно убедиться, приравняв нулю производную $\dfrac{\partial J_A[y]}{\partial y_{i_0}} = $ $= 2 \sum\limits_{j=0}^{N} a_{i_0 j} y_j - 2\varphi_{i_0} = 0, \dfrac{\partial^2 J_A}{\partial y_{i_0}^2} > 0$, так как $a_{ij} > 0, i = \overline{1, N}$ в силу положительности $A > 0$.

Вычисляя производные

$$\frac{\partial J_h}{\partial y_i} = 2a_i y_{\overline{x},i} - 2a_{i+1} y_{\overline{x},i+1} + h(2q_i y_i - 2f_i) = 0,$$

$$\frac{\partial^2 J_h}{\partial y_i^2} = \frac{2a_i}{h} + \frac{2a_{i+1}}{h} + 2q_i h > 0,$$

$$u_{\overline{x},i+1} = \frac{u_{i+1} - u_i}{h}, u_{\overline{x},i} = \frac{u_i - u_{i-1}}{h},$$

убеждаемся, что элемент $y = y(x) \in H_h$, минимизирующий квадратичный функционал, является решением разностной задачи

$$(ay_{\overline{x}})_{x,i} - q_i y_i = -f_i, i = 1, 2, \ldots, N, y_0 = y_N = 0.$$

§73. Метод аппроксимации интегрального тождества

Пусть дана граничная задача

$$(pu')' - qu + f(x) = 0, 0 < x < 1, u(0) = u(1) = 0. \quad (1)$$

Умножаем уравнение (1) на произвольную дифференцируемую функцию $\nu(x)$, обращающуюся в нуль при $x = 0, x = 1$, и интегрируя по x от 0 до 1, получим тождество $J(u\nu) = \int\limits_0^1 (pu'\nu' + qu\nu - f\nu)dx = 0$. Для его получения использовали формулы:

$$\int u d\nu = u\nu - \int \nu du;$$

$$\int\limits_0^1 (pu')'\nu = pu'\nu|_0^1 = -\int \nu'(pu')dx. \quad (2)$$

Заменяя по аналогии с предыдущим параграфом интеграл и производные u', ν', напишем *сумматорное тождество*

$$J_h[y,\nu] = \sum_{i=1}^{N} a_i y_{\overline{x},i} h + \sum_{i=1}^{N-1} (q_i y_i - f_i)\nu_i h = 0$$

Полагая, например, $\nu_i = \delta_{i,i_0}, 0 < i_0 < N$ и учитывая, что $\nu_{\overline{x},i} = 0$ при $i < i_0$ и $i > i_0 + 1$; получим $h(\frac{1}{h}a_{i+1}y_{x,i} - \frac{1}{h}a_i y_{\overline{x},i}) + (q_i y_i - f_i)h = 0$ при $i = i_0$, т. е. $(ay_{\overline{x}})_x - qy = -f$.

При этом учли, что

$$\nu_i = \delta_{i,i_0} = \begin{cases} 1, i = i_0, \\ 0, i \neq i_0; \end{cases}$$

$$\nu_{\overline{x},i} = \frac{\nu(x_i) - \nu(x_{i-1})}{h};$$

при $i = i_0$ $\nu(x_i) = 1 \neq 0$, при $i = i_0 + 1$ $\nu(x_{i-1}) = 1 \neq 0$,

$$\nu_{\overline{x},\ i_{0+1}} = -\frac{1}{h}, \nu_{\overline{x},i_0} = \frac{1}{h}.$$

§74. Метод Ритца и Бубнова-Галеркина (вариационно-разностные методы)

Задача о минимуме функционала

$$J[u] = (Au, u) - 2(u, f) \tag{1}$$

где A – самосопряженный и положительно определенный линейный оператор в гильбертовом пространстве H со скалярным произведением (x, y), эквивалентна задаче о решении уравнения $Au = f$.

Вводится последовательность конечномерных пространств V_n с базисом $[\varphi_i^{(n)}], i = 1, 2, \ldots, n$. Метод Ритца заключается в том, что ищется элемент $u \in V_n$, минимизирующий функционал $J[u]$ в V_n. Приближенное решение u_n ищется в виде суммы

$$u_n = \sum_{j=1}^{n} y_j \varphi_j \tag{2}$$

где $y_1, y_2, \ldots, y_n$ – неизвестные коэффициенты. Вычисления дают:

$$J[u_n] = \sum_{i,j=1}^{n} \alpha_{ij} y_i y_j - 2 \sum_{i=1}^{n} \beta_i y_i, \alpha_{ij} = \alpha_{ji} = (A\varphi_i, \varphi_j), \beta = (f, \varphi_i).$$

Здесь $J[u_n] = \Phi(y_1, y_2, \ldots, y_n)$ – функция от n коэффициентов y_i. Приравняв нулю производные $\dfrac{\partial J[u_n]}{\partial y_i}$, получим систему n уравнений

$$2\sum_{j=1}^{n}\alpha_{ij}y_j - 2\beta_i = 0,\ \ i = 1, 2, \ldots, n$$

для определения $y_1, y_2, \ldots, y_n$.

Пример. Применим метод Ритца к уравнению

$$Lu = (pu')' - qu = -f(x),\ \ 0 < x < 1,\ \ u(0) = 0,\ \ u(1) = 0. \qquad (3)$$

В качестве $\varphi_i(x)$ возьмем $\varphi_i(x) = \eta\left(\dfrac{x - x_i}{h}\right) = \eta_i(x)$, где

$$\eta(s) = \begin{cases} 0,\ \ s < -1,\ \ s > 1, \\ 1 + s <\ \ -1 < s < 0, \\ 1 - s,\ \ 0 < s < 1. \end{cases}$$

Подставим $A\varphi_i = -(p\varphi_i')' + q\varphi_i$ в формулу для α_{ij}

$$(A\varphi_i, \varphi_j) = -\int_l imits_0^1 (p\varphi_i')'\varphi_j + \int_0^1 q\varphi_i\varphi_j = \int_0^1 \varphi_j'(p\varphi_i') + \int_0^1 q\varphi_i\varphi_j, \qquad (4)$$

получим

$$\alpha_{ij} = (A\varphi_i, \varphi_j) = \int_0^1 \left(p\frac{d\eta_i}{dx}\frac{d\eta_j}{dx} + q\eta_i\eta_j\right)dx,\ \ \beta_i = \int_0^1 f(x)\eta_i(x)dx. \qquad (5)$$

Вычисления дают

$$\frac{d\eta_i}{dx} = 0 \text{ при } x < x_{i-1},\ x > x_{i+1};\ \frac{d\eta_i}{dx} = \begin{cases} \dfrac{1}{h}, x_{i-1} < x < x_i, \\ -\dfrac{1}{h}, \quad x_i < x < x_i + 1. \end{cases}$$

Отсюда и из (4) видно, что матрица $[\alpha_{ij}]$ трехдиагональная, так как от нуля отличны лишь те α_{ij}, для которых $j = i-1, i, i+1$. Поэтому для y_i получаем систему $\alpha_{i,i-1} i_{i-1} + \alpha_i, y_i + \alpha_{i,i+1} y_{i+1} - \beta_i = 0$. Вводя обозначения $a_i = -h\alpha_{i,i-1} + h^2\alpha_1 = h(\alpha_{i,i-1} + \alpha_{i,i+1}) + hd_i;\ \beta_i = -h^2\varphi_i$ и замечая, что $\alpha_{i,i+1} = \alpha_{i+1,i}$ получим схему

$$a_i y_{i-1} - (\alpha_i + \alpha_{i+1} + h^2 d_i) y_i + a_{i+1} y_{i+1} + h^2\varphi_i = 0 \tag{6}$$

или $(ay_{\overline{x}})_x - dy + \varphi = 0$, где

$$a_i = \int\limits_{-1}^{0} p(x_i + sh)ds + h^2 \int\limits_{-1}^{0} q(x_i + sh)s(1+s)ds,$$

$$d_i = \int\limits_{-1}^{0} q(x_i + sh)(1+s)ds + \int\limits_{0}^{1} q(x_i + sh)(1-s)ds,$$

$$\varphi_i = \int\limits_{-1}^{0} f(x_i + sh)(1+s)ds + \int\limits_{0}^{1} f(x_i + sh)(1-s)ds.$$

Формула (6) – это схема второго порядка аппроксимации.

В методе Бубнова–Галеркина решение u_n также ищется в виде (2), однако коэффициенты y_i находятся из условия ортогональности невязки $Au_n - f$ к базисным функциям $\varphi_i(x)$:

$$(Au_n - f, \varphi_i) = 0, i = 1, 2, \ldots, n. \tag{7}$$

При этом самосопряженность оператора A не требуется. Для задачи (3) снова выбираем те же базисные функции. Подставляя (2) в (7) получим систему уравнений для y_i . Вычисляя α_{ij} и β_i, приходим к той же самой схеме (6), которую мы получили методом Ритца. При указанном выборе координатных функций $\varphi_i(x) = \eta\left(\dfrac{x - x_i}{h}\right)$ методы Ритца и Бубнова–Галеркина совпадают с методом конечных разностей.

§75. Монотонные разностные схемы

Пусть дана разностная схема

$$Ly(x) = F(x), x \in \Omega, \tag{1}$$

где

$$Ly(x) = A(x)y(x) - \sum_{\xi \in Ш'(x)} \beta(x,\xi)y(\xi). \tag{2}$$

Каждой точке $x \in \Omega$ сопоставим один и только один шаблон $Ш(x)$ – любое подмножество Ω, содержащее данную точку.

Определение 1. *Окрестностью точки x назовем множество* $Ш'(x) = Ш(x)/\{x\}$.
Заметим, что $Ш'(x)$ может быть и пустым множеством.

Причем оператор L удовлетворяет условию положительности коэффициентов

$$A(x) > 0, B(x,\xi) > 0, D(x) = A(x) - \sum_{\xi \in Ш'(x)} B(x,\xi) \geq 0. \tag{3}$$

Определение 2. *Линейный оператор L называется монотонным оператором, если из условия $Ly(x) \geq 0$ для всех $x \in \Omega$ следует, что $y(x) \geq 0$ для всех $x \in \Omega$.*

Поэтому разностные схемы, удовлетворяющие при всех $x \in \Omega$ условиям (3), называются монотонными разностными схемами. Схемы, для которых условия (3) не выполнены хотя бы в одной точке $x \in \Omega$, называются немонотонными. Оказывается, что условия (3) обеспечивают монотонность оператора L, выполнение принципа максимума и корректность разностной задачи (1) в сеточной норме C:

$$\| y \|_{C(\Omega)} = \max_{x \in \Omega} |y(x)|.$$

Разумеется, отсюда не следует, что немонотонная задача обязательно некорректна.

Рассмотрим обыкновенное дифференциальное уравнение второго порядка

$$u''(x) + r(x)u'(x) = -f(x), 0 < x < 1, u(0) = u(1) = 0 \tag{4}$$

и поставим задачу построить для нее разностную схему, имеющую второй порядок аппроксимации и монотонную при любых шагах сетки h . Очевидная схема второго порядка аппроксимации, которая получается заменой $u'(x)$ центральной разностной производной, является монотонной лишь при достаточно малых h. Действительно, такая схема имеет вид

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} + r_i \frac{y_{i+1} - y_{i-1}}{2h} = -f_i$$

или $\dfrac{2}{h^2}y_i = \left(\dfrac{1}{h^2}+\dfrac{r_i}{2h}\right)y_{i+1}+\left(\dfrac{1}{h^2}-\dfrac{r_i}{2h}\right)y_{i-1}+f_i$. Условия положительности коэффициентов сводятся к неравенствам $0,5h|r_i|<1$ и выполняются, если $h \le \dfrac{2}{\max\limits_{0\le i\le N}|r_i|}$. Схема будет монотонной при любых h только в случае, когда . Прежде, чем построить требуемую схему для уравнения (4), рассмотрим несколько частных случаев. Предположим, что $r(x)\ge 0$ для всех $x\in(0,1)$ и рассмотрим схему с односторонней разностью

$$\frac{y_{i+1}-2y_i+y_{i-1}}{h^2}+r_i\frac{y_{i+1}-y_i}{h}=-f_i.$$

Эта схема имеет первый порядок аппроксимации и монотонна при $\forall h$. Действительно, записывая ее в виде $\left(\dfrac{2}{h^2}+\dfrac{r_i}{h}\right)y_i=\left(\dfrac{1}{h^2}+\dfrac{r_i}{h}\right)y_{i+1}+$ $+\dfrac{1}{h^2}y_{i-1}+f_i$ и учитывая неотрицательность $r(x)$, убеждаемся в том, что условие положительности коэффициентов выполнены при всех h. Точно также, если $r(x)\le 0$, то схема $\dfrac{y_{i+1}-2y_i+y_{i-1}}{h^2}+r_i\dfrac{y_i-y_{i-1}}{h}=-f_i$ монотонна при $\forall h$ и имеет первый порядок аппроксимации.

В общем случае представим функцию $r(x)$ в виде суммы

$$r(x)=r_+(x)+r_-(x),$$

где $r_+(x)=0,5(r(x)+|r(x)|)\ge 0$, $r_-(x)=0,5(r(x)-|r(x)|)\le 0$. Схема с направленными разностями

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} + r_+(x_i)\frac{y_{i+1} - y_i}{h} + r_-(x_i)\frac{y_i - y_{i-1}}{h} = -f_i \qquad (5)$$

является, как не трудно видеть, монотонной при любых h, но имеет первый порядок аппроксимации. Несколько измененная по сравнению с (5) схема

$$(1 - 0,5h|r(x_i)|)\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} + r_+(x_i)\frac{y_{i+1} - y_i}{h} + r_-(x_i)\frac{y_i - y_{i-1}}{h} = -f_i$$

имеет второй порядок аппроксимации.

ГЛАВА 6
ЧИСЛЕННОЕ РЕШЕНИЕ ИНТЕГРАЛЬНЫХ УРАВНЕНИЙ

§76. Некоторые предварительные определения

Идеи методов численного решения интегральных уравнений мы постараемся выяснить на примере простого случая, когда неизвестная функция зависит от одного аргумента, а интегралы, в которых она содержится, являются однократными и неособенными. В достаточно общем виде эти интегралы можно записать в форме

$$F(x, \varphi(x), z) = 0, \tag{1}$$

где $\varphi(x)$ есть искомая функция, а буква z обозначает один из следующих интегралов $z = \int\limits_a^b f(x, s, \varphi(s))ds$ или

$$z = \int\limits_a^x f(x, s, \varphi(s))ds. \tag{2}$$

С формальной точки зрения второй из интегралов является частным случаем первого, когда функция f такова, что $f(x, s, \varphi) = 0$ при $s > x$ и любых φ . Но это свойство функции f имеет настолько глубокие следствия, что делает теории уравнений с интегралами первого и второго рода (2) принципиально различными.

Уравнения, содержащие интегралы первого из видов (2), называются уравнениями Фредгольма, а интегралы с уравнениями второго вида – уравнениями Вольтерра.

Ограничимся рассмотрением линейных уравнений. В Фредгольмовом случае линейные уравнения имеют вид

$$A(x)\varphi(x) + \int\limits_a^b K(x,s)\varphi(s)ds = F(x), \tag{3}$$

а в случае Вольтерра

$$A(x)\varphi(x) + \int\limits_a^x K(x,s)\varphi(s)ds = F(x). \tag{4}$$

Когда коэффициент $A(x)$ тождественно равен нулю, уравнения (3) и (4) называются уравнениями первого рода. Они имеют вид

$$\int\limits_a^b K(x,s)\varphi(s)ds = F(x)$$

для случая уравнения Фредгольма, $\int\limits_a^x K(x,s)\varphi(s)ds = F(x)$ для случая уравнения Вольтерра.

Если же коэффициент $A(x)$ отличен от нуля всюду на отрезке $[a, b]$, то уравнения (3), (4) называются уравнениями второго рода. Они, очевидно, могут быть приведены к следующим каноническим формам

$$\varphi(x) - \lambda \int\limits_a^b K(x,s)\varphi(x)ds = f(x), a \le x \le b, \lambda = const \qquad (5)$$

в случае уравнения Фредгольма, и

$$\varphi(x) - \lambda \int\limits_a^x K(x,s)\varphi(x)ds = f(x), a \le x \le b, \lambda = const \qquad (6)$$

в случае уравнения Вольтерра. В дальнейшем будем рассматривать только эти уравнения. Функции $f(x)$ и $K(x,s)$ предполагаются непрерывными соответственно в областях $a \le x \le b$ и $a \le x, s \le b$. Функцию $K(x,s)$ называют ядром интегрального уравнения.

§77. Метод механических квадратур решения уравнения Фредгольма II рода

Пусть дано интегральное уравнение второго рода вида Фредгольма

$$\varphi(x) = f(x) + \lambda \int\limits_a^b K(x,s)\varphi(x)ds. \qquad (1)$$

На отрезке интегрирования $[a,b]$ возьмем произвольно n точек $a \le x_1 < < x_2 < \ldots \le x_n |leqb$ и для нахождения приближенных значений функции φ в них построим систему уравнений. Такую систему уравнений можно получить, если к вычислению интеграла, входящего в уравнение, применить какое-либо правило приближенного интегрирования. Интеграл зависит от параметра x, поэтому коэффициенты правила и его остаточный член будут, вообще говоря, некоторыми функциями от x:

$$\int\limits_a^b K(x,s)\varphi(x)ds = \sum_{k=1}^{n} c_k(x)\varphi(x_k) + \rho(x). \qquad (2)$$

Подставляя (2) в (1) и положив там последовательно $x = x_1, x_2, \ldots, x_n$, получим для точных значений $\varphi(x)$ систему уравнений

$$\varphi(x_i) = f_i + \lambda \sum_{k=1}^{n} c_k(x_i)\varphi(x_k) + \lambda\rho(x_i). \qquad (3)$$

Остаток $\rho(x)$ в квадратурном правиле (2) по сравнению с квадратурной суммой имеет обычно малую величину. Поэтому, если в равенствах (3) пренебречь малыми величинами $\lambda\rho(x_i)$, то можно рассчитывать на то, что получится алгебраическая система уравнений, решение которой $\varphi_i (i = 1, \ldots, n)$ будет близким к значениям $\varphi(x_i)$ точного решения интегрального уравнения

$$\varphi_i - \lambda \sum_{k=1}^{n} c_k(x_i)\varphi_k = f_i, i = \overline{1, n}. \qquad (4)$$

Определитель системы

$$\Delta(\lambda) = D(E - \lambda C), \qquad (5)$$

где $C = \begin{bmatrix} c_1(x_1) & \ldots & c_n(x_1) \\ & \ldots & \\ c_1(x_n) & \ldots & c_n(x_n) \end{bmatrix}$ есть многочлен от λ степени n. Предположим, что λ не есть собственное значение уравнения (1) и уравнение, следовательно, имеет только одно решение. Тогда можно ожидать, что определитель системы $\Delta(\lambda)$ отличен от нуля и система (4) также имеет единственное решение. Но это надлежит проверить в каждой частной задаче.

Будем считать, что $\Delta(\lambda) \neq 0$. При решении системы каждое уравнение удовлетворяется не вполне точно, а с некоторым округлением чисел.

Если погрешность, вызванную округлениями, обозначить – δ_i, то в действительности для φ_i будут выполняться равенства

$$\varphi_i - \lambda \sum_{k=1}^{n} c_k(x_i)\varphi_k = f_i - \delta_i \qquad (6)$$

и, следовательно,

$$\varphi_i = \frac{1}{\Delta(\lambda)} \sum_{k=1}^{n} \Delta_{ki}[f_k - \delta_k], \qquad (7)$$

где Δ_{ki} есть алгебраическое дополнение элемента определителя $\Delta(\lambda)$ с индексами k, i.

Теперь рассмотрим погрешность приближенного решения $\varepsilon_i = \varphi(x_i) - \varphi_i$ и получим ее оценку. Точные значения $\varphi(x_i)$ решения удовлетворяют уравнениям (3), которые отличаются от уравнений (4) для φ_i только заменой – δ_i на $\lambda\rho(x_i)$, и поэтому для $\varphi(x_i)$ верны выражения вида (7) с такой же заменой:

$$\varphi(x_i) = \frac{1}{\Delta(\lambda)} \sum_{k=1}^{n} \Delta_{k_i}[f_k + \lambda\rho(x_k)]. \qquad (8)$$

Действительно, точное решение $\varphi(x_i)$ получается из уравнения

$$\varphi(x_i) - \lambda \sum_{k=1}^{n} c_k(x_i)\varphi(x_k) = f_i + \lambda\rho(x_i).$$

Кафедры ПМиТП АиГ

Начало

Содержание

Назад

На весь экран

Закрыть

Приближенное решение φ_i является решением уравнения $\varphi_i - \lambda \sum\limits_{k=1}^{n} c_k(x_i)\varphi_k = f_i - \delta_i$, поэтому $\varphi(x_i) = \dfrac{1}{\Delta(\lambda)} \sum\limits_{k=1}^{n} \Delta_{ki}[f_k + \lambda\rho(x_k)]$. Если теперь отсюда вычесть (7), то получится представление для ε_i:

$$\varepsilon_i = \frac{1}{\Delta(\lambda)} \sum_{k=1}^{n} \Delta_{ki}[\pi\rho(x_k) + \delta_k]. \tag{9}$$

Действительно,

$$\varepsilon_i = \varphi(x_i) - \varphi_i = \frac{1}{\Delta(\lambda)} \sum_{k=1}^{n} \Delta_{ki}[f_k + \lambda\rho(x_k)] - \frac{1}{\Delta(x)} \sum_{k=1}^{n} \Delta_{ki}[f_k - \delta_k] =$$

$$= \frac{1}{\Delta(\lambda)} \sum_{k=1}^{n} \Delta_{ki}[\lambda\rho(x_k) + \delta_k].$$

Оно дает возможность получить оценку ε_i . Предположим, что для погрешности $\rho(x)$ квадратурного правила (2) верно неравенство $|\rho(x)| \le \rho = \rho(n)$. В качестве $\rho(n)$ может быть взята величина $\rho(n) = \max\limits_{x} |\rho(x)|$. Предположим также, что для погрешности δ_k, вызванной округлениями, выполняется неравенство $|\delta_k| \le \delta = \delta(n), k = \overline{1, n}$.

Величина $\delta(n)$ определяется принятым вычислительным процессом. Из (9) вытекает следующая оценка для $|\varepsilon_i|$

$$|\varepsilon_i| \le B(|\lambda|\rho + \delta), \tag{10}$$

где $\dfrac{1}{|\Delta(\lambda)|}\sum\limits_{k=1}^{n}|\Delta_{ki}| \le B, i = \overline{1,n}.$

Отметим попутно, что значения $\Delta(\lambda)$ и Δ_{ki} находятся в процессе вычислений и поэтому является вычислимой величиной. В оценке (10) остается еще найти величину ρ. С этой целью возвратимся к остаточному члену $\rho(x)$ квадратурного правила (2) при произвольном значении x:

$$\rho(x) = \int\limits_a^b K(x,s)\varphi(s)ds - \sum_{k=1}^{n} c_k(x)\varphi(x_k).$$

Введем в него вместо $\varphi(x)$ выражение, стоящее в правой части уравнения (1):

$$\rho(x) = \int\limits_a^b K(x,s)\left[f(s) + \lambda\int\limits_a^b K(s,u)\varphi(u)du\right]ds - \sum_{k=1}^{n} c_k(x)[f(x_k)+$$

$$+\lambda\int\limits_a^b K(x_k,u)\varphi(u)du] = \int\limits_a^b K(x,s)f(s)ds - \sum_{k=1}^{n} c_k(x)f(x_k)+$$

$$+\lambda\left[\int\limits_a^b\int\limits_a^b K(x,s)K(s,u)\varphi(u)duds - \sum_{k=1}^{n} c_k\int\limits_a^b K(x_k,u)\varphi(u)du\right] = \rho_f(x)+$$

$$+\lambda\int_a^b \rho_K(x,u)\varphi(u)du, \tag{11}$$

где

$$\rho_f(x)=\int_a^b K(x,s)f(s)ds-\sum_{k=1}^n c_k(x)f(x_k),$$

$$\rho_K(x,u)=\int_a^b K(x,s)K(s,u)ds-\sum_{k=1}^n c_k(x)K(x_k,u).$$

Очевидно, $\rho_f(x)$ есть остаток приближенного интегрирования по правилу (2) свободного члена f заданного уравнения, а $\rho_K(x,u)$ – остаток интегрирования по тому же правилу ядра $K(s,u)$. Обе эти функции известны и поэтому две следующие величины:

$$\rho_f=\max_x|\rho_f(x)| \text{ и } \rho_k=\max_x\int_a^b|\rho_K(x,u)|du$$

мы вправе считать также известными.

Из (11) получается следующая оценка для ρ:

$$|\rho(x_i)\leq\rho\leq\max|\rho(x)|\leq\rho_f+|\lambda|\rho_K H,$$

$$H=\max_x|\varphi(x)|. \tag{12}$$

Подставим (12) в (10). Полученное неравенство совместно с соотношением (10) приводит к следующей оценке погрешности ε_i:

$$|\varepsilon_i| \le B\left[|\lambda|(\rho_f + |\lambda|\rho_K H) + \delta\right], i = \overline{1,n}. \tag{13}$$

Недостаток оценки (13) в том, что неизвестно число H , являющиеся наибольшим абсолютным значением неизвестной функции. Этот недостаток можно устранить и оценить H через вычисляемые величины.

Для этого понадобится определить приближенное решение не только в точках x_i , но и на всем отрезке $[a, b]$. Такую интерполяцию приближенного решения с сетки на отрезок можно выполнить многими способами, например, при помощи алгебраического интерполирования. Мы воспользуемся «естественным» в рассматриваемой задаче интерполированием через квадратурное правило. Возвратимся к уравнению (1) и вычислим в нем интеграл при помощи квадратурного правила (2)

$$\varphi(x) = f(x) + \lambda \sum_{k=1}^{n} c_k(x)\varphi(x_k) + \lambda\rho(x). \tag{14}$$

Заменим в правой части точные значения $\varphi(x_k)$ функции φ в узлах x_k на вычисленные приближенные значения φ_k и отбросим «малый» остаток $\lambda\rho(x)$:

$$\widetilde{\varphi}(x) = f(x) + \lambda \sum_{k=1}^{n} c_k(x)\varphi_k. \tag{15}$$

Полученную так функцию $\widetilde{\varphi}(x)$ примем за приближенное решение на всем отрезке $[a,b]$. Получаемая погрешность имеет вид

$$\varepsilon(x)=\varphi(x)-\widetilde{\varphi}(x)=f(x)+\lambda\sum_{k=1}^{n}c_k(x)\varphi(x_k)+\lambda\rho(x)-f(x)-\lambda\sum_{k=1}^{n}c_k(x)\varphi_k=$$

$$=\lambda\sum_{k=1}^{n}c_k(x)[\varphi(x_k-\varphi_k)+\lambda\rho(x)]=\lambda\sum_{k=1}^{n}c_k(x)\varepsilon_k+\lambda\rho(x).$$

Отсюда $\varepsilon(x)=\varphi(x)-\widetilde{\varphi}(x)=\lambda\left[\sum\limits_{k=1}^{n}C_k(x)\varepsilon_k+\rho(x)\right]$. Отсюда, если ввести обозначение $\max\limits_{x}\sum\limits_{k=1}^{n}|C_k(x)|=C$, то для $\varepsilon(x)$ с помощью оценок (12) и (13) получается неравенство $|\varepsilon(x)|\leq|\lambda|\{B[|\lambda|(\rho_f+|\lambda|\rho_K H)+\delta]C+$ $+(\rho_f+|\lambda|\rho_K h)\}$.

Введем величину $\widetilde{H}=\max\limits_{x}|\widetilde{\varphi}(x)|$. Она может быть вычислена, так как $\widetilde{\varphi}(x)$ выражается только через величины, находящиеся в процессе вычислений.

Из равенства $\varphi(x)=\widetilde{\varphi}(x)+\varepsilon(x)$ следует $H\leq\widetilde{H}+\max\limits_{x}|\varepsilon(x)|$.

Воспользуемся полученной выше оценкой для $|\varepsilon(x)|$, получим

$$H\leq\widetilde{H}+|\lambda|\{BC[|\lambda|(\rho_f+|\lambda|\rho_K H)+\delta]+(\rho_f+|\lambda|\rho_K H)\}.$$

Таким образом,

$$H-|\lambda|^2BC|\lambda|\rho_K H-|\lambda^2|\rho_k H\leq\widetilde{H}+|\lambda|^2BC\rho_f+|\lambda|BC\delta+|\lambda|\rho_f,$$

Кафедры
ПМиТП
АиГ
Начало
Содержание
Страница 332 из 403
Назад
На весь экран
Закрыть

$$H[1-|\lambda|^2\rho_K(BC|\lambda|+1)] \le \widetilde{H}+\rho_f|\lambda|(1+|\lambda|BC)+\delta BC|\lambda|.$$

Отсюда при условии, что $|\lambda|^2\rho_K(1+|\lambda|BC)<1$ вытекает неравенство для H:

$$H \le \frac{\widetilde{H}+|\lambda|[\rho_f(1+|\lambda|BC)+\delta BC]}{1-|\lambda|^2\rho_K(1+|\lambda|BC)}. \tag{16}$$

Оно дополняет оценку (13) погрешности ε_i для тех случаев, когда необходимо знать границу для численного значения величины H.

В предыдущем изложении узлы x_k считались произвольными. Сделаем сейчас несколько замечаний об их выборе и соответствующим им квадратурных правилах. Когда необходимо составить таблицу значений решения φ, то наибольший интерес представляет случай равноотстоящих значений аргумента $s_k = a+hk, k=\overline{0,n},\ h=\dfrac{b-a}{n}$. Тогда для вычисления интеграла могут быть применены, например, правила трапеций, парабол, трех восьмых и т.д.

В случае трапеций формула (2) будет иметь вид

$$\int\limits_a^b K(x,s)\varphi(s)ds = h\left[\frac{1}{2}K(x,s_0)+K(x,s_0)\varphi(s_1)+\ldots+\right.$$

$$\left.+K(x,s_{n-1})\varphi(s_{n-1})+\frac{1}{2}K(x,b)\varphi(b)\right]-\frac{(b-a)^3}{12n^2}\left[\frac{\partial^2}{\partial s^2}(K\varphi)\right]_{s=\xi}, a\le\xi\le b.$$

В случае парабол

$$\int\limits_a^b K(x,s)\varphi(s)ds = h\left\{\frac{1}{3}[K(x,s_0)\varphi(s_0)+K(x,s_n)\varphi(x_n)]+\right.$$

$$\left.+\frac{4}{3}\sum_m K(x,s_{2m+1})\varphi(s_{2m+1})+\frac{2}{3}\sum_m K(x,s_{2m})\varphi(s_{2m})\right\}-$$

$$-\frac{1}{180}\frac{(b-a)^5}{n^4}\left[\frac{\partial^4}{\partial s^4}(K\varphi)\right]_{s=\xi}, a\le\xi\le b.$$

Правила приближенного интегрирования этого вида имеют невысокую степень точности и для получения малой погрешности часто бывает необходимо брать малое значение h и, следовательно, большое число узлов x_k. Тогда система вида (4) будет иметь большое число уравнений и ее решение потребует большого числа вычислений. Ввиду этого часто предпочитают применять квадратурные правила, имеющие высокую степень точности, например, правило Гаусса или его аналоги.

77.1 Сходимость метода квадратур

Вычислительный процесс метода квадратур определяется двумя таблицами: бесконечной таблицей узлов квадратурных правил

$$X = \begin{bmatrix} x_1^1 & & & \\ x_1^2 & x_2^2 & & \\ \dots & \dots & \dots & \\ x_1^n & x_2^n & \dots & x_n^n \\ \dots & \dots & \dots & \dots \end{bmatrix} \qquad (17)$$

и таблицей коэффициентов $c_k^i(x)$ этих правил

$$A = \begin{bmatrix} c_1^1(x) & & & \\ c_1^2(x) & c_2^2(x) & & \\ \dots & \dots & \dots & \\ c_1^n(x) & c_2^n(x) & \dots & c_n^n(x) \\ \dots & \dots & \dots & \dots \end{bmatrix}. \qquad (18)$$

Таблицы составлены так, что в строке номера n помещены узлы и коэффициенты квадратурного правила вида (2), которое применяется для приближенного вычисления интеграла на шаге номера n процесса

$$\int_a^b K(x,s)\varphi(s)ds = \sum_{k=1}^{n} c_k^n(x)\varphi(x_k^n) + \rho_\varphi^n(x). \qquad (19)$$

В вычислительном процессе все величины, зависящие от c_k^n и x_k^n будут некоторыми функциями от n, например, остаточный член $\rho^n(x)$ квадратурного правила будет зависеть как от x, так и от n. В оценке (13) для $|\varepsilon_i|$ от номера n шага будут зависеть как сама погрешность ε_i , так и величины B, ρ_f, ρ_K. От n будет зависеть и погрешность δ, так как, если мы хотим сделать процесс сходящимся к точному решению, то с ростом n должны неограниченно увеличивать точность решений системы вида (4). Указанная зависимость от n отличена при помощи верхнего индекса n. Для процесса оценка вида (13) запишется в следующей форме:

$$|\varepsilon_i^n| \leq B^n[|\lambda|(\rho_f^n + |\lambda|\rho_k^n H) + \delta^n], i = \overline{1, n}. \tag{20}$$

Такая оценка погрешности позволяет считать доказанной приводимую ниже теорему о сходимости вычислительного процесса метода квадратур.

Теорема 1. *Предположим, что λ не есть собственное значение ядра $K(x, s)$ и уравнение (1), следовательно, имеет единственное решение. Пусть для вычислительного процесса метода квадратур, определяемого равенством (19), выполняются условия:*

1. *погрешности $\rho_{f(x)}^n(x)$ и $\rho_{K(s,u)}^n(x)$ квадратурного правила (19) для функций $\varphi(s) = f(x)$ и $\varphi(s) = K(s, u)$ таковы, что при неограниченном росте будет соответственно*

$$B^n \rho_{f(x)}^n(x) \to 0 \tag{21}$$

равномерно относительно x и

$$\rho^n_{K(s,u)}(x) \to 0 \qquad (22)$$

равномерно относительно x и u,

2. *погрешности $-\delta_i^n$ решения уравнений системы вида (4) уменьшаются настолько быстро при увеличении n, что имеет место соотношения*

$$B^n\delta^n \to 0, \delta^n = \max_i |\delta_i^n|. \qquad (23)$$

Тогда для всякого положительного ε существует $N(\varepsilon)$ такое, что при $n > N(\varepsilon)$ будут выполняться неравенства

$$|\varphi(x_i^n) - \varphi_i^n| < \varepsilon, i = \overline{1, n}. \qquad (24)$$

Приведенная теорема имеет ту особенность, что ее условия (21) (22) должны проверяться в процессе вычислений, а условие (23) должны учитываться при определении точности решения системы уравнений для нахождения $\varphi_i^n, i = \overline{1, n}$.

§78. Метод последовательных приближений решения интегрального уравнения Фредгольма второго рода

В этом методе решение уравнения

$$\varphi(x) - \lambda \int_a^b K(x,s)\varphi(s)ds = f(x), a \le x \le b, \lambda = const, \tag{1}$$

ищется в виде степенного ряда λ, т.е

$$\varphi(x) = \sum_{k=0}^{\infty} \lambda^k \varphi_k(x), \tag{2}$$

где $\varphi_k(x)$ – некоторая система функций, которая будет рекуррентно определена из следующих соображений. Подставив (2) в уравнение (1), получим

$$\sum_{k=0}^{\infty} \lambda^k \varphi_k(x) = \lambda \int_a^b K(x,s) \sum_{k=0}^{\infty} \lambda^k \varphi_k(s)ds + f(x). \tag{3}$$

Для определения $\varphi_k(x)$ будем сравнивать коэффициенты при одинаковых степенях λ в соотношении (3). Получим

$$\lambda^0 : \varphi_0(x) = f(x),$$

$$\lambda^1 : \varphi_1(x) = \int\limits_a^b K(x,s)\varphi_0(s)ds,$$

$$\lambda^2 : \varphi_2(x) = \int\limits_a^b K(x,s)\varphi_1(s)ds,$$

$$\dots\dots\dots$$

$$\lambda^m : \varphi_m(x) = \int\limits_a^b K(x,s)\varphi_{m-1}(s)ds,$$

$$\dots\dots\dots \tag{4}$$

Система (4) и есть система рекуррентного определения $\varphi_k(x)$, $k = 0, 1, \ \dots$. Тогда приближенное решение определяется формулой

$$y_n(x) = \sum_{k=0}^{n} \lambda^k \varphi_k(x). \tag{5}$$

Для исследования вопроса о сходимости метода (5) проведем предварительные оценки в системе (4). Рассмотрим равномерную метрику и пусть

$$||f(x)|| = \max_{a \le x \le b} |f(x)| = ||f(x)||_c.$$

Тогда из (4) можно получить следующие оценки:

$$||\varphi_0(x)|| = ||f(x)||;$$

$$||\varphi_1(x)|| \leq K(x,s)||\ ||\varphi_0(x)||(b-a) = (b-a)||K(x,s)||\ ||f(x)||;$$
$$||\varphi_2(x) \leq ||K(x,s)||||\varphi_1(x)||(b-a) = (b-a)^2||K(x,s)||^2||f(x)||;$$

$$\ldots$$

$$||\varphi_n(x)|| \leq (b-a)^n||K(x,s)||^n||f(x)||;$$

$$\ldots,$$

где $||K(x,s)|| = \max\limits_{a\leq x,s\leq b}|K(x,s)|$. Обозначим $z_n(x) = \varphi(x) = y_n(x) = = \sum\limits_{k=0}^{\infty}\lambda^k\varphi_k(x) - \sum\limits_{k=0}^{n}\lambda^k\varphi_k(x) = \sum\limits_{k=n+1}^{\infty}\lambda^k\varphi_k(x)$. Тогда

$$||z_n(x)|| \leq \sum_{k=n+1}^{\infty}|\lambda|^k||\varphi_k(x)|| \leq \sum_{k=n+1}^{\infty}(b-1)^k|\lambda|^k||K(x,s)||^k||f|| =$$

$$= (b-a)^{n+1}|\lambda|^{n+1}||K(x,s)||^{n+1}||f|| + (b-a)^{n+2}|\lambda|^{n+2}||K(x,s)||^{n+2}||f|| +$$
$$+ (b-a)^{n+3}|\lambda|^{n+3}||K(x,s)||^{n+3}||f|| + \ldots = (b-a)^{n+1}|\lambda|^{n+1}||K(x,s)||^{n+1}\times$$
$$\times||f||(1+(b-a)|\lambda|||K(x,s)|| + (b-a)^2|\lambda|^2||K(x,s)||^2 + \ldots) =$$

$$= (b-a)^{n+1}|\lambda|^{n+1}||K(x,s)||^{n+1}||f||\sum_{k=0}^{\infty}(b-a)^k|\lambda|^k||K(x,s)||^k.$$

Пусть $q = (b-a)|\lambda|||K(x,s)|| < 1$. Тогда метод последовательных приближений сходится равномерно по x, причем сходимость линейная. В этом случае $\sum\limits_{k=0}^{\infty}(b-a)^k|\lambda|^k||K(x,s)||^k = (1-|\lambda|(b-a)||K(x,s)||)^{-1}$ и для погрешности имеет место оценка

$$||z_n(x)|| \leq \frac{(b-a)^{n+1}|\lambda|^{n+1}||K(x,s)||^{n+1}||f||}{1-(b-a)|\lambda|\ ||K(x,s)||}.$$

§79. Интерполяционный квадратурный метод

Этот метод является частным случаем общего квадратурного метода и основан на интерполировании неизвестной функции φ. Для определенности изложения будем иметь в виду алгебраическое интерполирование.

Выберем на отрезке $[a, b]$ n различных точек $a \leq x_1 \leq x_2 \leq \ldots \leq \leq x_n \leq b$ и выполним интерполирование функции φ по ее значениям в этих точках. Воспользуемся лагранжевым представлением интерполяционного многочлена

$$\varphi(x) = \sum_{k=1}^{n} \frac{\omega(x)}{(x - x_k)\omega'(x_k)}\varphi(x_k) + r(x) = \sum_{k=1}^{n} l_k(x)\varphi(x_k) + r(x). \qquad (1)$$

Внесем это представление φ в интегральный член уравнения:

$$\varphi(x) = f(x) + \lambda \sum_{k=1}^{n} L_k(x)\varphi(x_k) + \lambda\rho(x),$$

$$L_k(x) = \int_a^b K(x, s) l_k(s) ds,$$

$$\rho(x) = \int_a^b K(x, s) r(s) ds. \qquad (2)$$

Чтобы получить систему уравнений для $\varphi(x_k)$ отбросим в (2) остаточный член $\lambda\rho(x)$ и в полученном приближенном равенстве положим x равным $x_1, x_2, \dots, x_n$:

$$\varphi_i = f_i + \lambda \sum_{k=1}^{n} L_k(x_i)\varphi_k, i = \overline{1,n}. \tag{3}$$

Эта система является частным случаем более общей системы, когда $C_k(x_i) = L_k(x_i)$.

После решения системы (3) мы будем знать приближенные значения φ_i решения φ в точках x_i, $i = \overline{1,n}$. Чтобы найти приближенные значения φ всюду на $[a,b]$ можно выполнить алгебраическое интерполирование φ по значениям φ_i. Но можно ожидать, что лучшая точность будет достигнута, если воспользоваться «естественным» интерполированием φ через ядро $K(t,s)$. Оно получается, если в равенстве (2) отбросить остаточный член $\lambda\rho(x)$:

$$\varphi = f(x) + \lambda \sum_{k=1}^{n} L_k(x)\varphi(x_k). \tag{4}$$

Остановимся на выборе узлов x_k в интерполяционном методе. Если точное решение $\varphi(x)$ является многочленом от x степени не выше $n-1$, то интерполирование $\varphi(x)$ будет точным и остаточный член $r(x)$ в (1) тождественно равным нулю. В соответствие с этим остаточный член $\lambda\rho(x)$ в (2) также будет тождественным нулем, и системе (3) будут удовлетворять точные значения решения $\varphi_i = \varphi(x_i)$. Эти значения могут быть найдены из системы и по ним построено точное решение $\varphi(x)$.

Это верно при всех узлах $x_k, k = \overline{1,n}$, лишь бы они были различными. Поэтому говорят, что интерполяционный метод при любых узлах x_k имеет степень точности не меньше $n-1$. Естественно попытаться выяснить, можно ли путем выбора узлов сделать интерполяционный метод точным для любых многочленов $\varphi(x)$ степени n, или, может быть, для всяких многочленов степени $m > n$. Остановимся на требовании точности метода для многочленов степени n. Это требование эквивалентно тому, чтобы выполнялись точно равенства

$$\int_a^b K(x_i,s)\varphi(s)ds = \sum_{k=1}^{n} L_k(x_i)\varphi(x_k), i = \overline{1,n} \tag{5}$$

для любых многочленов $\varphi(s)$ степени n.

Теорема 1. *Для того, чтобы равенства (5) были точными для всякого многочлена $\varphi(s)$ степени n необходимо и достаточно выполнение условий:*

1. *коэффициенты $L_k(x_i)$ должны быть значениями при $x = x_i$ функции $L_k(x)$, указанной в равенствах (2);*

2. *многочлен $\omega(s) = (s-x_1)\ldots(s-x_n)$ должен удовлетворять сле-*

дующим условиям ортогональности:

$$\int\limits_a^b K(x_i,s)\omega(s)ds = 0, i = \overline{1,n}. \tag{6}$$

Замечание 1. *Равенство (6) дает систему n уравнений, из которой должны быть найдены узлы $x_i, i = \overline{1,n}$.*

Доказательство. Проверим *необходимость* условий теоремы. Предположим, что равенство (5) верно для всякого многочлена $\varphi(s)$ степени n. Рассмотрим коэффициент Лагранжа $l_m(s)$, отвечающий узлу x_m:
$l_m(s) = \dfrac{\omega(s)}{(s-x_m)\omega'(x_m)}$.

Это есть многочлен степени $n-1$, поэтому для него равенство (5) должно быть точным, т.е.

$$\int\limits_a^b K(x_i,s)l_m(s)ds = \sum_{k=0}^{n} L_k(x_i)l_m(x_k).$$

С другой стороны, $l_m(x_k) = 0, k \neq m$ и $l_m(x_m) = 1$ и в сумме соотношений (5) сохранится только слагаемое, отвечающее $k = m$. Тогда равенство (5) примет вид:

$$\int\limits_a^b K(x_i,s)l_m(s)ds = L_m(x_i),$$

Кафедры ПМиТП АиГ

что доказывает необходимость первого условия.

Положим теперь $\varphi(s) = \omega(s)$. Многочлен $\omega(s)$ имеет степень n, и для него (5) выполняется точно:

$$\int_a^b K(x_i, s)\omega(s)ds = \sum_{k=0}^{n} L_k(x_i)\omega(x_k).$$

Но так как $\omega(x_k) = 0, k = \overline{1,n}$, то $\int_a^b K(x_i, s)\omega(s)ds = 0$, $i = \overline{1,n}$, тогда равенство (5) совпадает с (6). Этим доказывается необходимость второго условия теоремы.

Достаточность. Пусть $\varphi(x)$ – произвольный многочлен степени n. Разделим его на $\omega(s)$. В частном получится постоянная величина, которую обозначим через α, а в остатке некоторый многочлен $Q(x)$ степени не больше $n-1$: $\varphi(s) = \alpha\omega(s) + Q(s)$.

Так как $\omega(x_k) = 0$, то $Q(x_k) = \varphi(x_k)$ и

$$\int_a^b K(x_i, s)\varphi(s)ds = \alpha \int_a^b K(x_i, s)\omega(s)ds + \int_a^b K(x_i, s)Q(s)ds.$$

По второму условию теоремы первый из интегралов, стоящих справа, равен нулю. Так как $Q(s)$ имеет степень $\leq n-1$, то для него интерполяционное правило интегрирования

$$\int_a^b K(x,s)\varphi(s)ds = \sum_{k=1}^{n} L_k(x)\varphi(x_k) + \rho(x),$$

$$L_k(x) = \int_a^b K(x,s)l_k(s)ds$$

выполняется при всяких x точно (без остатка $\rho(x)$). В частности, при $x = x_i$ имеем $\int_a^b K(x_i,s)Q(s)ds = \sum_{k=1}^{n} L_k(x_i)Q(x_k) = \sum_{k=1}^{n} L_k(x_i)\varphi(x_k)$.

Поэтому $\int_a^b K(x_i,s)\varphi(s)ds = \sum_{k=1}^{n} L_k(x_i)\varphi(x_k)$ и равенства (5) действительно точны для всякого многочлена $\varphi(x)$ степени n. Теорема доказана.

§80. Метод замены ядра уравнения на вырожденное ядро для решения уравнений Фредгольма второго рода

Ядро $K(t, s)$ интегрального уравнения

$$\varphi(x) - \lambda \int_a^b K(x, s)\varphi(s)ds = f(x), a \le x \le b, \lambda = const \qquad (1)$$

называется вырожденным, если оно представимо в виде

$$K(x, s) = \sum_{i=1}^{n} \alpha_i(x)\beta_i(s). \qquad (2)$$

Если в интегральном уравнении (1) ядро $K(x, s)$ имеет вид (2), то его решение, если оно существует, представимо в следующем виде

$$\varphi(x) = f(x) + \sum_{j=1}^{n} A_j \alpha_j(x), \qquad (3)$$

где коэффициенты A_j есть некоторые неизвестные постоянные

$$A_j = \lambda \int_a^b \beta_j(s)\varphi(s)ds.$$

Действительно,

$$\varphi(x) = f(x) + \lambda \int\limits_a^b K(x,s)\varphi(s)ds = f(x) + \lambda \int\limits_a^b \sum_{i=1}^n \alpha_i(x)\beta_i(s)\varphi(s)ds =$$

$$= f(x) + \lambda \sum_{i=1}^n \alpha_i(x) \int\limits_a^b \beta_i(s)\varphi(s)ds = f(x) + \lambda \sum_{j=1}^n A_j\alpha_j(x).$$

Для нахождения этих коэффициентов может быть просто построена система уравнений.

Введем следующие обозначения: $\beta_{ij} = \int\limits_a^b \beta_i(s)\alpha_j(s)ds$, $f_i = \int\limits_a^b f(s)\beta_i(s)ds$. Подставляя в исходное уравнение (1) вместо $\varphi(x)$ его представление (3), получим

$$f(x) + \sum_{i=1}^n A_i\alpha_i(x) = f(x) + \lambda \int\limits_a^b \sum_{i=1}^n \alpha_i(x)\beta_i(s) \left(f(s) + \sum_{j=1}^n A_j\alpha_j(x) \right) ds.$$

Функции $\alpha_i(x), i = 1, 2, \ldots, n$ будем считать линейно независимыми. Аналогичное требование наложим и на $\beta_i(x), i = \overline{1, n}$. Сравнивая коэффициенты при соответствующих $\alpha_i(x)$ в левой и правой части последнего равенства, получим следующую систему для A_i:

$$A_i = \lambda f_i + \lambda \sum_{j=1}^{n} \beta_{ij} A_j, i = \overline{1, n}. \tag{4}$$

Если ввести в рассмотрение векторы $A = (A_1, A_2, \ldots, A_n)$, $f = (f_1, f_2, \ldots, f_n)$ и матрицу

$$B = \begin{bmatrix} \beta_{11} & \beta_{12} & \ldots & \beta_{1n} \\ \beta_{21} & \beta_{22} & \ldots & \beta_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ \beta_{n1} & \beta_{n2} & \ldots & \beta_{nn} \end{bmatrix},$$

то систему (4) можно записать в виде

$$A = \lambda f + \lambda BA. \tag{5}$$

Будем считать, что главный определитель этой системы $\neq 0$ т.е.

$$\Delta(\lambda) = D(E - \lambda B) = \begin{bmatrix} 1 - \lambda\beta_{11} & -\lambda\beta_{12} & \ldots & -\lambda\beta_{1n} \\ -\lambda\beta_{21} & 1 - \lambda\beta_{22} & \ldots & -\lambda\beta_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ -\lambda\beta_{n1} & -\lambda\beta_{n2} & \ldots & 1 - \lambda\beta_{nn} \end{bmatrix} \neq 0.$$

Решением системы (5) является вектор

$$A = (E - \lambda B)^{-1} \lambda f. \tag{6}$$

Чтобы дать явное выражение для составляющих вектора A, обозначим $\Delta_{ij}(\lambda)$ – алгебраическое дополнение элемента определителя $\Delta(\lambda)$, стоящего в строке номера i и j-том столбце. Тогда составляющая A_i вектора может быть записана в виде $A_i = \lambda \dfrac{1}{\Delta(\lambda)} \sum\limits_{j=1}^{n} \Delta_{ij} f_j$.

Подставим A_i в (3), получим

$$\varphi(x) = f(x) + \lambda \frac{1}{\Delta(\lambda)} \sum_{i=1}^{n} \sum_{j=1}^{n} \Delta_{ji} \alpha_i(x) \int\limits_a^b f(s) \beta_j(s) ds =$$

$$= f(x) + \lambda \int\limits_a^b \frac{\Delta(x, s; \lambda)}{\Delta(\lambda)} f(s) ds.$$

Резольвента интегрального уравнения здесь имеет простое явное выражение: $\Gamma(x, s; \lambda) = \frac{1}{\Delta(\lambda)} \Delta(x, s; \lambda) = \frac{1}{\Delta(\lambda)} \sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} \Delta_{ji} \alpha_i(x) \beta_j(s)$.

Построение решения $\varphi(x)$, $\varphi(x) = f(x) + \sum\limits_{j=1}^{n} A_j \alpha_j(x)$ приводится в случае вырожденного ядра $K(x, s)$ к вычислению величин f_i, β_{ij} и решению системы линейных алгебраических уравнений (4).

Метод приближенного решения, который рассматривается в этом параграфе основан на простом соображении: если в заданном интегральном уравнении произвольное ядро заменить на близкое к нему вырожденное ядро, то решение нового вспомогательного уравнения будет близким к решению заданного уравнения. Построение же решения вспомогательного уравнения будет намного проще ввиду вырожденности ядра. Более того, решение вспомогательного уравнения может быть сделано сколь угодно близким к решению заданного уравнения, если во-первых, число λ не есть собственное значение заданного уравнения и, во-вторых, если вырожденное ядро вспомогательного уравнения взято достаточно близким к заданному ядру. Укажем некоторые примеры построения вырожденных ядер.

80.1 Применение степенного ряда

Этот метод применим в том случае, когда ядро $K(x,s)$ есть аналитическая функция от s в области $|s-c| \le R$, где c есть середина отрезка $[a,b]$ и $R > \frac{1}{2}(b-a)$. В этом случае $K(x,s)$ можно разложить в степенной ряд по степеням $s-c$, сходящийся в круге $|s-c| < R$. Коэффициенты этого ряда будут зависеть от x:

$$K(x,s) = c_0(x) + c_1(x)(s-c) + \ldots + c_n(x)(s-c)^n + \ldots,$$

где $c_n(x) = \frac{1}{n!}\dfrac{\partial^n K(x,c)}{\partial x^n}$. В качестве вырожденного ядра $K(x,s)$ здесь может быть взят конечный отрезок степенного ряда $\widetilde{K}(x,s) = c_0(x) + {}$ $+ c_1(x)(s-c) + \ldots + c_n(x)(s-c)^n$. Аналогично можно поступить и в тех случаях, когда ядро есть аналитическая функция от x или от обоих аргументов.

80.2 Использование интерполяционных методов

Выберем на отрезке $[a, b]$ n узлов $s_1, s_2, \ldots, s_n$ и проинтерполируем по аргументу s функцию $K(x, s)$ по ее значениям во взятых точках. При этом будет получено приближенное значение $K(x, s)$

$$\widetilde{K}(x, s) = \sum_{i=1}^{n} \frac{\omega(s)}{(s - s_k)\omega'(s_k)} K(x, s_k), \left(\omega(s) = \prod_{i=1}^{n} (s - s_i)\right),$$

которое может быть принято за вырожденное ядро.

§81. Оценка близости между решениями уравнений в зависимости от близости самих уравнений

Пусть задано интегральное уравнение

$$\varphi(x) = f(x) + \lambda \int\limits_a^b K(x,s)\varphi(s)ds \tag{1}$$

и одновременно с ним рассматривается измененное уравнение

$$\widetilde{\varphi}(x) = \widetilde{f}(x) + \lambda \int\limits_a^b \widetilde{K}(x,s)\widetilde{\varphi}(s)ds. \tag{2}$$

Здесь роль измененного уравнения будет играть уравнение с вырожденным ядром.

При сравнении уравнений (1) и (2) возникают два вопроса:

1) что можно сказать о разрешимости заданного уравнения, если известна разрешимость измененного уравнения и как можно оценить разность между решениями по различию между самими уравнениями;

2) как можно ответить на сходные вопросы об измененном уравнении по известным свойствам заданного уравнения.

В постановке вопросов и в доказываемой ниже теореме заданное и измененное уравнения легко можно поменять местами, поэтому достаточно получить ответ на первый вопрос.

Теорема 1. *Предположим, что для уравнений (1) и (2) выполняются условия:*

1)для ядер $K(x,s)$ и $\widetilde{K}(x,s)$ при всяких $x \in [a,b]$ выполняется неравенство

$$\int_a^b |K(x,s) - \widetilde{K}(x,s)|ds \le h; \tag{3}$$

2) при всяких x и s из $[a,b]$ для резольвенты $\widetilde{\Gamma}(x,s;\lambda)$ уравнения (2) верна оценка

$$\int_a^b |\widetilde{\Gamma}(x,s;\lambda)ds| \le B; \tag{4}$$

3) для свободных членов f и $\widetilde{f}$ уравнений верно неравенство

$$|f(x) - \widetilde{f}(x)| \le \eta, x \in [a,b]; \tag{5}$$

4) числа λ, h, η, B удовлетворяют условию

$$1 - |\lambda|h(1 + |\lambda|B) > 0. \tag{6}$$

Тогда уравнение (1) имеет единственное решение $\varphi(x)$ и для решений уравнений (1) и (2) верна оценка

$$|\varepsilon(x)| = |\varphi(x) - \widetilde{\varphi}(x)| \le (\eta + |\lambda|hM)(1 + |\lambda|B), \tag{7}$$

где $M = \max\limits_x |\varphi(x)|$, *при этом для* M *выполняется неравенство*

$$M \le \frac{\widetilde{M} + \eta(1 + |\lambda|B)}{1 - |\lambda|h(1 + |\lambda|B)}, \widetilde{M} = \max_x |\widetilde{\varphi}(x)|.$$

Доказательство. Вычтем (2) из (1), получим

$$\varepsilon(x) = \varphi(x) - \widetilde{\varphi}(x) =$$

$$= f(x) - \widetilde{f}(x) + \lambda \int\limits_a^b K(x,s)\varphi(s)ds - \lambda \int\limits_a^b \widetilde{K}(x,s)\widetilde{\varphi}(s)ds =$$

$$= f(x) - \widetilde{f}(x) + \lambda \int\limits_a^b [K(x,s) - \widetilde{K}(x,s)]\varphi(s)ds + \lambda \int\limits_a^b \widetilde{K}(x,s)[\varphi(s) - \widetilde{\varphi}(s)]ds =$$

$$= f^*(x) + \lambda \int\limits_a^b \widetilde{K}(x,s)[\varphi(s) - \widetilde{\varphi}(s)]ds.$$

Это равенство можно рассматривать как интегральное уравнение для $\varphi - \widetilde{\varphi}$ со свободным членом f^* и ядром $\widetilde{K}(x,s)$. При помощи резольвенты $\widetilde{\Gamma}$ из него получается явное выражение для $\varphi - \widetilde{\varphi}$:

$$\varphi(x) - \widetilde{\varphi}(x) = f^*(x) + \lambda \int\limits_a^b \widetilde{\Gamma}(x,s;\lambda) f^*(s)ds. \qquad (8)$$

Оно дает возможность получить оценку для $\varphi - \widetilde{\varphi}$. Так как

$$|f^*(x)| \le |f(x) - \widetilde{f}(x)| + |\lambda| \left| \int_a^b [K(x,s) - \widetilde{K}(x,s)]\varphi(s)ds \right| \le \eta + |\lambda| hM,$$

то верно неравенство, вытекающее из (8):

$$\begin{aligned} |\varepsilon(x)| = |\varphi(x) - \widetilde{\varphi}(x)| &\le \eta + |\lambda| hM + |\lambda| B(\eta + |\lambda| hM) \\ &= (\eta + |\lambda| hM)(1 + |\lambda| B). \end{aligned} \tag{9}$$

В правую часть (9) входит величина $M = \max\limits_x |\varphi(x)|$, которая является неизвестной. Она может быть легко оценена в условиях теоремы через $\widetilde{M}$. Действительно, из (9) следует $|\varphi(x)| \le \widetilde{M} + (\eta + |\lambda| hM)(1 + |\lambda| B)$, и так как неравенство выполняется при любых $x \in [a, b]$, то $M \le \widetilde{M} + (\eta + |\lambda| hM)(1 + |\lambda| B)$. Следовательно,

$$\max_x |\varphi(x)| = M \le \frac{\widetilde{M} + \eta(1 + |\lambda| B)}{1 - |\lambda| h(1 + |\lambda| B)}. \tag{10}$$

Отметим, что ввиду условия (6) знаменатель дроби положителен. Из неравенства (10) следует, что все решения уравнения (1) ограничены. Поэтому число λ не является собственным значением ядра $K(x, s)$, так как в противном случае к $\varphi(x)$ можно было бы прибавить решение однородного уравнения, умноженное на любую постоянную, и неравенство (10) не могло бы выполняться. Следовательно, уравнение (1) имеет решение, и притом единственное.

§82. Метод моментов

Как и выше, будем считать свободный член $f(x)$ и ядро $K(x,s)$ интегрального уравнения

$$L(\varphi) = \varphi(x) - f(x) - \lambda \int_a^b K(x,s)\varphi(s)ds = 0 \qquad (1)$$

непрерывными функциями.

Нам придется иметь дело с двумя системами функций, предназначенных для разных целей. Первая из систем $\psi_i(x), i = 1, 2, \ldots,$ будет служить для составления приближенного представления решения $\varphi(x)$ уравнения и вторая, обозначенная $\chi_j(x), j = 1, 2, \ldots,$ будет применяться для возможно лучшего в определяемом ниже смысле выполнения уравнения (1).

Функции $\psi_i(x)$ считаются непрерывными на $[a,b]$, линейно независимыми там, и система их предполагается обладающей свойством *C-полноты* на множестве непрерывных на $[a,b]$ функций $F(x)$.
Последнее означает, что для каждой функции $F(x)$ из этого множества и для произвольного $\varepsilon > 0$ существует такое n и такие коэффициенты $c_1, c_2, \ldots, c_n$, что выполняется неравенство

$$\left| F(x) - \sum_{i=1}^{n} c_i\psi_i(x) \right| < \varepsilon, a < x < b.$$

Функции $\chi_i(x)$ предположим непрерывными, линейно независимыми и систему $\chi_i(x)$ – замкнутой во множестве непрерывных на $[a, b]$ функций.

Определение 1. *Система функций* $\{\chi_i(x)\}$ *замкнута на множестве интегрируемых функций на* $[a, b]$, *если не существует такой непрерывной функции, кроме тождественного нуля, которая была бы ортогональной на* $[a, b]$ *ко всем функциям* $\chi_i(x)$. *Т.е., если система* $\chi_i(x)$ *замкнута, то из условия* $\int\limits_a^b f(x)\chi_i(x)dx = 0$, $i = 0, 1, 2, \dots$, *где* $f(x) \in C[a, b]$ *следует, что* $f(x) = 0$.

Составим линейную неоднородную комбинацию

$$\varphi_n(x) = f(x) + \sum_{i=1}^{n} c_i\psi_i(x). \tag{2}$$

Она содержит n неопределенных коэффициентов c_i , и мы их выберем при помощи приводимых ниже наглядных соображений.

Требование, чтобы функция $\varphi_n(x)$ была решением заданного уравнения, т.е. чтобы было верно тождество $L(\varphi_n) \equiv 0$, равносильно выполнению бесконечного множества равенств $\int\limits_a^b L(\varphi_n)\chi_j(x)dx = 0, j = 1, 2, \dots$. Но в нашем распоряжении имеется лишь n коэффициентов c_i, и мы имеем возможность выполнить только n первых условий

$$\int_a^b L(\varphi_n)\chi_j(x)dx =$$

$$= \int_a^b \left[\varphi_n(x) - f(x) - \lambda \int_a^b K(x,s)\varphi_n(s)ds \right] \chi_j(x)dx = 0, \ \ j = \overline{1,n}. \quad (3)$$

Такие равенства означают, что первые n моментов функции $L(\varphi_n)$ по системе функций $\chi_j(x)$, $j = 1, 2, \ldots$ равны нулю. Условия (3) дают линейную систему уравнений для коэффициентов c_i. Если решить ее и подставить найденные значения в (2), то получим вообще говоря, не точное значение решения $\varphi(x)$ уравнения (1), а лишь некоторое приближение к нему $\varphi_n(x)$. Отметим, что в вычислениях часто пользуются не двумя системами функций, а только одной системой, полагая, $\chi_j(x) = \psi_j(x)$, $j = 1, 2, \ldots$. Такой частный случай метода моментов носит название *метода Галеркина.*

82.1 Связь метода Галеркина с задачей замены ядра на вырожденное

Покажем, что метод Галеркина равносилен методу замены ядра на вырожденное при некотором простом частном правиле построения последнего.

Не уменьшая общности изложения, всегда можно считать систему функций $\psi_i(x)$ ортонормированной, ибо, если она не является такой, то ее можно привести к ортонормированной с помощью известного процесса ортогонализации. Решается уравнение

$$L(\varphi) = \varphi(x) - \lambda \int\limits_a^b K(x,s)\varphi(s)ds = 0.$$

По методу моментов ищем решение уравнения в виде $\varphi_n(x) = f(x) + \sum\limits_{j=1}^{n} c_j\psi_j(x)$, где c_j находятся из (3). При условии $\chi(x) = \psi_j(x)$ уравнения (3) примут вид

$$\int\limits_a^b L(\varphi_n)\psi_j(x)dx = \int\limits_a^b L(\varphi_n)\psi_j(x)dx =$$
$$= \int\limits_a^b \left[\varphi_n(x) - f(x) - \lambda \int\limits_a^b K(x,s)\varphi_n(s)ds\right]\psi_j(x)dx = 0. \qquad (4)$$

Решим уравнение $L(\varphi) = \varphi(x) - \lambda \int\limits_a^b K(x,s)\varphi(s)ds = 0$ методом вырожденного ядра. Для построения вырожденного ядра разложим $K(x,s)$

по первому аргументу x в ряд по функциям $\psi_j(x)$, возьмем отрезок ряда из n первых членов

$$\breve{K}_n(x,s)=\sum_{j=1}^{n}u_j(s)\psi_j(x),\, u_j(s)=\int\limits_a^b K(x,s)\psi_j(x)dx \tag{5}$$

и рассмотрим интегральное уравнение с вырожденным ядром $\breve{K}_n(x,s)$.

$$\breve{L}(\breve{\varphi})=\breve{\varphi}(x)-f(x)-\lambda\int\limits_a^b \breve{K}_n(x,s)\breve{\varphi}(s)ds=0. \tag{6}$$

Если в (6) вместо $\breve{K}_n$ внести его представление (5), то будет ясно, что решение $\breve{\varphi}$ имеет следующее выражение:

$$\breve{\varphi}(x)=f(x)+\sum_{j=1}^{n}c_j\psi_j(x), \tag{7}$$

где $c_j=\lambda\int\limits_a^b u_j(s)\breve{\varphi}(s)ds$, совпадающее по форме с (2). Коэффициенты c_j здесь неизвестны и их можно найти по методу моментов. Подставляя в уравнение вместо φ выражение (6), умноженное на $\chi_j(x)$, и интегрируя по x, получим следующую систему уравнений для c_i:

$$\int_a^b \breve{L}(\breve{\varphi})\psi_j(x)dx =$$

$$= \int_a^b \left[\breve{\varphi} - f(x) - \lambda \int_a^b \breve{K}_n(x,s)\breve{\varphi}(s)ds\right] \psi_j(x)dx = 0, \; j = \overline{1,n}. \qquad (8)$$

По способу ее получения ясно, что она является следствием уравнения (6). Но верно и обратное. В самом деле, если $\breve{\varphi}(x)$ имеет форму (7), то в прямых скобках под знаком интеграла стоит линейная комбинация $\psi(x)_j(x)$, $i = \overline{1,n}$, и так как она ортогональна ко всем функциям $\chi_i(x)$, то уравнения (8) выполняются только в том случае, когда выражение в скобке есть тождественный нуль и $\breve{\varphi}(x)$ является, следовательно, решением интегрального уравнения (6). Таким образом, решение системы (8) и уравнения (6) есть операции равносильные.

Если же сравнить уравнения (4) и (8), то легко убедиться, что они совпадают. Действительно, уравнения (8) отличаются от (4) лишь тем, что в них под знаком двойного интеграла стоит ядро $\breve{K}_n(x,s)$, а не заданное ядро $K(x,s)$. Однако можно просто говорить, что слагаемые с двойными интегралами совпадают, для чего достаточно сначала выполнить интегрирование по x и воспользоваться тем, что

$$\int_a^b \overset{\smile}{K}_n(x,s)\psi_j(x)dx = c_j(x) = \int_a^b K_n(x,s)\psi_j(x)dx.$$

Изложенное приводит к заключению: если в методе моментов $\psi_i(x) = \chi_i(x)$, $i = 1, 2, \ldots$, то применение его к решению уравнения (1) равносильно замене ядра $K(x,s)$ этого уравнения на вырожденное ядро $\overset{\smile}{K}_n(x,s)$, определенное равенством (5).

Этот результат позволяет для определения погрешности метода применять оценки, которые были получены для разности решений близких интегральных уравнений.

§83. Метод коллокации

Как и прежде, будем считать свободный член $f(x)$ и ядро $K(x,s)$ уравнения непрерывными, но запишем уравнение в иной, более удобной для нас форме с другим оператором:

$$\Lambda[\varphi(x)] = \varphi(x) - \lambda \int_a^b K(x,s)\varphi(s)ds = f(x). \tag{1}$$

Возьмем на $[a,b]$ n различных точек $a \le x_1 \le x_2 \le \ldots \le x_n \le b$. Рассмотрим, кроме того, n функций $\psi_i(x)$, $i = 1, 2, \ldots, n$, которые удовлетворяют двум условиям: 1) функции непрерывны на $[a,b]$; 2) линейно независимы на $[a,b]$. Дальнейшие условия на $\psi(x)$ будут указаны ниже.

Образуем линейную комбинацию

$$\varphi_n(x) = f(x) + \sum_{j=1}^{n} c_j\psi_j(x). \tag{2}$$

Коэффициенты c_j выберем так, чтобы $\varphi_n(x)$ удовлетворяла уравнению (1) в точках x_i, $i = \overline{1,n}$. Такое требование приведет к следующей системе уравнений для x_i:

$$\sum_{j=1}^{n} c_j \Lambda\left[\psi_j(x_i)\right] = \lambda \int_a^b K(x_i, s) f(s) ds, i = \overline{1, n}, \qquad (3)$$

Система (3) получится, если подставить (2) в (1). Действительно,

$$f(x) + \sum_{j=1}^{n} c_j \psi_j(x) - \lambda \int_a^b K(x, s) \left[f(s) + \sum_{j=1}^{n} c_j \psi_j(s) \right] ds = f(x),$$

$$\sum_{j=1}^{n} c_j \psi_j(x) - \lambda \int_a^b K(x, s) f(s) ds - \lambda \int_a^b K(x, s) \sum_{j=1}^{n} c_j \psi_j(s) ds = 0,$$

$$\sum_{j=1}^{n} c_j \psi_j(x) - \lambda \sum_{j=1}^{n} c_j \int_a^b K(x, s) \psi_j(s) ds = \lambda \int_a^b K(x, s) f(s) ds,$$

$$\sum_{j=1}^{n} c_j \left[\psi_j(x) - \lambda \int_a^b K(x, s) \psi_j(s) ds \right] = \lambda \int_a^b K(x, s) f(s) ds,$$

$$\sum_{j=1}^{n} c_j \Lambda\left[\psi_j(x)\right] = \lambda \int_a^b K(x,s) f(s) ds.$$

Если ввести обозначения $\Lambda\left[\psi_j(x)\right] = \Phi_j(x)$ и $\lambda \int_a^b K(x_i, s) f(s) ds = F(x)$, то система уравнений (3) запишется в виде:

$$\sum_{j=1}^{n} c_j \Phi_j(x_i) = F(x_i), i = \overline{1, n}. \tag{4}$$

Чтобы эта система имела единственное решение, нужно потребовать, чтобы ее определитель был отличен от нуля

$$D_n = D_n\left[\Phi_1, \Phi_2, \ldots, \Phi_n\right] = \begin{bmatrix} \Phi_1(x_1) & \ldots & \Phi_n(x_1) \\ \vdots & \ddots & \vdots \\ \Phi_1(x_n) & \ldots & \Phi_n(x_n) \end{bmatrix} \neq 0. \tag{5}$$

Усилим это требование и будем считать, что неравенство (5) выполняется при всяких $x_1, x_2, \ldots, x_n$, различных между собой. Напомним, что система функций $\Phi_j(x)$, $j = \overline{1, n}$, обладающая таким свойством, называется системой Чебышева. Изложенное приводит к третьему требованию, которому мы подчиняем выбор ψ_i:

3)$\psi_i(x)$ таковы, что соответствующие им функции $\Phi_i(x)$ образуют на $[a, b]$ систему Чебышева.

Решив уравнения (4) и подставив найденные значения в (2), найдем функцию $\varphi_n(x)$, которую примем за приближение к точному решению $\varphi(x)$. Его погрешность $\varepsilon_n(x) = \varphi(x) - \varphi_n(x)$ зависит как от свойств системы $\psi_j(x)$, так и от выбора узлов x_i, $i = \overline{1, n}$. Оценка $\varepsilon_n(x)$, также как и условия сходимости $\varphi_n(x) \to \varphi(x)$, $n \to \infty$, выяснена в небольшом числе случаев.

Последнее условие, которому обычно подчиняют выбор функций $\psi_i(x)$ связано с вопросом о сходимости последовательности приближений $\varphi_n(x)$, $n = 1, 2, \ldots$ к точному решению $\varphi(x)$. Чтобы определить вычислительный процесс, необходимо, прежде всего, задать либо бесконечную треугольную таблицу функций $\psi_j(x)$

$$\begin{Vmatrix} \psi_1^{(1)}(x) & & \\ \psi_1^{(2)}(x) & \psi_2^{(2)}(x) & \\ \psi_1^{(3)}(x) & \psi_2^{(3)}(x) & \psi_3^{(3)}(x) \\ \ldots & \ldots & \ldots \end{Vmatrix},$$

либо бесконечную последовательность этих функций $\psi_j(x)$, $j = 1, 2, \ldots$. Кроме того, должна быть задана треугольная таблица узлов $x_1^{(n)}, x_2^{(n)}, \ldots, x_n^{(n)}$, $n = 1, 2, \ldots$, на которых нужно выполнять равенство (4).

Если не делать никаких специальных предположений о $f(x)$ и $K(x,s)$ и считать их произвольными непрерывными функциями, то для того, чтобы надеяться на равномерную сходимость $\varphi_n(x)$ к $\varphi(x)$, $n \to \infty$, мы должны предположить, что последовательность $\psi_i(x)$ обладает $C-$полнотой на множестве непрерывных на $[a,b]$ функций. Это требование и принимают в качестве четвертого условия, налагаемого на выбор последовательности функций $\psi_i(x)$.

§84. Метод квадратур решения интегрального уравнения Вольтерра второго рода

Будем рассматривать интегральное уравнение Вольтерра второго рода

$$\varphi(x) - \lambda \int\limits_a^x K(x,s)\varphi(s)ds = f(x), a \le x \le b, \lambda = const. \qquad (1)$$

Пусть $x \in [a,b]$ и на $[a,b]$ взята сетка равностоящих узлов $x_n = a+nh$, $n = 0, 1, \ldots, N$. Положим в (1) $x = x_n$, $\lambda = 1$ и рассмотрим равенство

$$\varphi(x_0) = f(x_0), \varphi(x_n) - \lambda \int\limits_a^{x_n} K(x_n,s)\varphi(s)ds = f(x_n), n = \overline{1,N}. \qquad (2)$$

Пусть для вычисления интеграла в (2) взята квадратурная формула с узлами в точках $x_0, x_1, x_2 \ldots, x_n$, т.е.

$$\lambda \int\limits_a^{x_n} K(x_n,s)\varphi(s)ds = h\sum_{k=0}^{n} K(x_n,x_k)\varphi(x_k)A_{nk} + r_n. \qquad (3)$$

Здесь r_n – остаток квадратурной формулы. Подставим вместо интегралов в (2) их вид через (3), получим

$$\varphi(x_0) = f(x_0),$$

$$\varphi(x_n) - h\sum_{k=0}^{n} K(x_n, x_k)\varphi(x_k)A_{nk} + r_n = f(x_n), n = \overline{1, N}. \qquad (4)$$

Отбрасывая остаток r_n, получим уравнение для приближенных значений $\varphi(x_n)$, которые мы обозначим через φ_n, $K_{nk} = K(x_n, x_k)$.

$$\varphi_n - h\sum_{k=0}^{n} K_{nk}\varphi_k A_{nk} = f_n, n = \overline{1, N}, \qquad (5)$$

где использованы введенные ранее обозначения. Из последних соотношений при ограниченных K_{nk}, A_{nk} и h таких, что $1 - hK_{nn}, A_{nn} \neq 0$, будем иметь

$$\varphi_n = \frac{1}{1 - hA_{nn}K_{nn}}\left(h\sum_{k=0}^{n-1} K_{nk}\varphi_k A_{nk} + f_n\right), n = \overline{1, N} \qquad (6)$$

Действительно, из (5)

$$\varphi_n - hK_{nn}\varphi_n A_{nn} - h\sum_{k=1}^{n-1} K_{nk}A_{nk}\varphi_k = f_n,$$

$$\varphi_n(1 - hK_{nn}A_{nn}) = h\sum_{k=1}^{n-1} K_{nk}A_{nk}\varphi_k + f_n.$$

Если $1 - hK_{nn}A_{nn} \neq 0$, то $\varphi_n = \dfrac{1}{1 - hK_{nn}A_{nn}}\left[h\sum\limits_{k=1}^{n-1} K_{nk}A_{nk}\varphi_k + f_n\right]$, $n = \overline{1, N}$.

Из этих соотношений рекуррентно получим приближенные значения решения (1) в узлах сетки. Более того, если для всех n выполняется неравенство $h \mid K_{nn}A_{nn} \mid \leq q < 1$, то этот квадратурный процесс сходится.

Укажем на необходимость составления начала расчетной таблицы при применении правила (5). Коэффициенты A_{nk} обычно выбираются так, чтобы локальная погрешность r_n правила имела во всех узлах один и тот же порядок малости относительно h, т.е. чтобы погрешность r_n при всех n имела представление вида $r_n = h^k C_n(h)$, $C_n(h) \to C_n \neq 0$ при $h \to 0$.

При численном решении системы уравнения будут выполнены неточно и, если погрешность их выполнения обозначить через δ_n, то в действительности точные равенства будут иметь вид

$$\varphi_0 = f_0 + \delta_0, \varphi_n - h\sum_{k=0}^{n} K_{nk}A_{nk}\varphi_k = f_n + \delta_n, n = \overline{1, N}. \qquad (7)$$

Справедлива

Теорема 1. *Пусть для вычислительного процесса (7) выполняются условия:*

1. *погрешность* r_n *квадратурной формулы* (3) *стремится к нулю равномерно относительно* n, *когда* $h \to 0$;

2. *существует число* A *такое, что при всех значениях* n, k, h *выполняется неравенство* $\mid A_{nk} \mid \leq A < \infty$;

3. *погрешность* δ_n *вычислительного процесса* (7) *стремится к нулю равномерно относительно* n *при* $h \to 0$.

Тогда при всяких достаточно малых h *приближенное решение* φ_n *может быть построено по правилу* (7) *и для него будет выполняться неравенство* $\mid \varphi_n - \varphi(x_n) \mid \leq \varepsilon$, $n = \overline{0, N}$, *при любом положительном заданном* ε, *если* h *будет достаточно малым.*

Предположим теперь, что нам нужно вычислить φ_1. Для этого рассмотрим квадратурную формулу (3) при $n = 1$:

$$\int\limits_a^{a+h} K(x_1, s)\varphi(s)ds = h[A_{10}K_{10}\varphi(a) + A_{11}K_{11}\varphi(a+h)] + r_1.$$

Здесь $x_0 = a$, $x_1 = a+h$. Она содержит два численных параметра A_{10} и A_{11}, которые мы можем выбрать так, чтобы формула имела возможно большую степень точности. Максимум, что мы можем получить в этом отношении – это сделать формулу точной для многочленов первой степени. Тогда формула станет известным правилом трапеций, для которого $A_{10} = A_{11} = \frac{1}{2}$: $\int\limits_a^{a+h} K(x_1, s)\varphi(s)ds = \frac{h}{2}[K_{10}\varphi(a)+K_{11}\varphi(a+h)]+r_1$.

Остаточный член $r_1 = -\frac{h^3}{12}\left(\frac{\partial^2(K\varphi)}{\partial s^2}\right)_{s=\xi}$, $a \le \xi \le a + h$.

Отсюда видно, что r_1 имеет относительно h третий порядок малости.

Если окажется, что порядок k малости r_1 больше трех, то формулу (7) нельзя применять к нахождению φ_1. Тогда φ_1 должно быть вычислено заранее до применения правила (7).

Аналогично, чтобы воспользоваться правилом (7) для нахождения φ_2, нужно положить в (3) $n = 2$ и рассмотреть квадратурную формулу

$$\int\limits_a^{a+2h} K(x_2, s)\varphi(s)ds =$$

$$= h[A_{20}K_{20}\varphi(a) + A_{21}K_{21}\varphi(a + h) + A_{22}K_{22}\varphi(a + 2h)] + r_2$$

.

Наивысшая степень алгебраической точности, которую можно достичь при выборе коэффициентов A_{20}, A_{21}, A_{22} равна 3 и достигается она в правиле парабол, для которого $A_{20} = A_{22} = \frac{1}{3}$, $A_{21} = \frac{4}{3}$.

$$\int\limits_a^{a+2h} K(x_2, s)\varphi(s)ds = \frac{h}{3}[K_{20}\varphi(a) + 4K_{21}\varphi(a+h) + K_{22}\varphi(a+2h)] + r_2.$$

Здесь r_2 – остаточный член формулы парабол,

$$r_2 = -\frac{1}{90}h^5\left(\frac{\partial^4(K\varphi)}{\partial s^4}\right)_{s=\xi}, a \leq \xi \leq a+2h.$$

Остаток r_2 имеет пятый порядок малости относительно h, и если окажется, что $k > 5$, то правило вычислений (7) нецелесообразно применять для нахождения φ_2, и это значение, как и φ_1, должно быть вычислено заранее до применения (7).

Значения φ_n, которые должны быть найдены предварительно составят начало расчетной таблицы. Их можно находить, решая интегральное уравнение в окрестности точки a при помощи степенного ряда, если функция $f(x)$ и ядро $K(x, s)$ являются аналитическими функциями, или применяя для вычисления значений решения формулу трапеций с уменьшенным шагом.

§85. Метод последовательных приближений решения интегрального уравнения Вольтерра второго рода

Решение интегрального уравнения Вольтерра второго рода

$$\varphi(x) - \lambda \int\limits_a^{x_n} K(x,s)\varphi(s)ds = f(x) \tag{1}$$

будем искать в виде

$$\varphi(x) = \sum_{k=0}^{\infty} \lambda^k \varphi_k(x), \tag{2}$$

где $\varphi_k(x)$ – система функций, подлежащая определению. Для их определения подставим (2) в (1). Будем иметь

$$\sum_{k=0}^{\infty} \lambda^k \varphi_k(x) - \lambda \int\limits_a^{x_n} K(x,s) \sum_{k=0}^{\infty} \lambda^k \varphi_k(s)ds = f(x). \tag{3}$$

В (3) приравняем коэффициенты при одинаковых степенях λ.

$\lambda^0 : \varphi_0(x) = f(x);$

$\lambda^1 : \varphi_1(x) = \int\limits_a^x K(x,s)\varphi_0(s)ds;$

$\lambda^2 : \varphi_2(x) = \int\limits_a^x K(x,s)\varphi_1(s)ds;$

$\ldots\ldots$

$\lambda^n : \varphi_n(x) = \int\limits_a^x K(x,s)\varphi_{n-1}(s)ds.$

$\ldots\ldots$ (4)

Система соотношений (4) является системой рекуррентного определения $\varphi_n(x)$. Приближенное значение ищется в виде

$$y_n(x) = \sum_{k=0}^{n} \lambda^k \varphi_k(x). \tag{5}$$

Рассмотрим вопрос о сходимости этого процесса: будет ли $y_n(x) \to \varphi(x)$ при $n \to \infty$ или, что тоже самое, будет ли $\varphi(x) - y_n(x) \to 0$ при $n \to \infty$. Обозначим погрешность $\varepsilon_n(x) = \varphi(x) - y_n(x) =$ $= \sum\limits_{k=n+1}^{\infty} \lambda^k \varphi_k(x)$. Тогда

$\|\varepsilon_n(x)\|_C = \max\limits_{[a,b]} |\varepsilon(x)| \leq \sum\limits_{k=n+1}^{\infty} |\lambda|^k \|\varphi_k(x)\|_C.$

Рассмотрим вопрос о поведении $\|\varepsilon_n(x)\|$ при условии, что $\|K(x,s)\|_C \leq K$, $\|f(x)\|_C \leq F$. Из (4) будем иметь $\|\varphi_0(x)\| = \|f(x)\| \leq$ $\leq F$,

$$\|\varphi_1(x)\| \leq \|\int_a^x K(x,s)\varphi_0(s)ds\| \leq$$

Кафедры ПМиТП АиГ

Начало

Содержание

Назад

На весь экран

Закрыть

$$\leq \int_a^x \|K(x,s)\| \|\varphi_0(s)\| ds \leq KF(x-a) \leq KF(b-a),$$

$$\|\varphi_2(x)\| \leq | \int_a^x K(x,s)\varphi_1(s)ds\| \leq$$

$$\leq \int_a^x \|K(x,s)\| \|\varphi_1(s)\| ds \leq K \int_a^x KF(s-a)ds \leq$$

$$\leq K^2 F\frac{(x-a)^2}{2!} \leq \frac{FK^2(b-a)^2}{2!},$$

$$\ldots$$

$$\|\varphi_n(x)\| \leq \frac{K^n F(b-a)^n}{n!}.$$

Тогда $\|\varepsilon_n(x)\| \leq \sum\limits_{k=n+1}^{\infty} |\ \lambda\ |^k \ \|\varphi_k(x)\| \leq \sum\limits_{k=n+1}^{\infty} |\ \lambda\ |^k \ \dfrac{K^k F(b-a)^k}{k!} =$ $= \dfrac{|\ \lambda\ |^{n+1}\ K^{n+1}(b-a)^{n+1}F}{(n+1)!} \sum\limits_{k=0}^{\infty} |\ \lambda\ |^k \ \dfrac{K^k(b-a)^k}{k!}$. При этом последний ряд при сделанных ограничениях всегда сходится. Следовательно, метод последовательных приближений (5) решения интегрального уравнения Вольтерра второго рода сходится всегда.

§86. Решение нелинейных уравнений вида Вольтерра

Идеи численного решения, о которых говорилось выше для линейных уравнений Вольтерра, могут быть перенесены на нелинейные уравнения весьма общего вида.

Рассмотрим нелинейное уравнение Вольттера второго рода

$$\varphi(x) = f(x) + \int\limits_a^x F(x, s, \varphi(s))ds, \tag{1}$$

и предположим, что нужно, как и выше, найти значения решения в равноотстоящих точках $x_k = a + kh$, $k = \overline{1, N}$, $x_N \le b \le x_{N+1}$.

Положим в уравнении $x = x_n$:

$$\varphi(x_n) = f(x_n) + \int\limits_a^{x_n} F(x_n, s, \varphi(s))ds. \tag{2}$$

Применим для вычисления интеграла какое-либо квадратурное правило с узлами $x_0, x_1, \ldots, x_n$:

$$\int\limits_{x_0}^{x_n} F(x_n, s, \varphi(s))ds = h\sum_{i=0}^{n} B_{ni}F_{ni} + r_n, F_{ni} = F(x_n, x_i, \varphi(x_i)). \tag{3}$$

Если внести это выражение интеграла в (2) и отбросить «малый» остаточный член r_n квадратурной формулы (3), то получим систему уравнений для нахождения приближенных значений $\varphi(x_n)$, которые обозначены φ_n:

$$\varphi_0 = f_0; \varphi_n + h\sum_{i=0}^{n} B_{ni}\overset{\smile}{F}_{ni}; n = 1, 2, \dots N; \overset{\smile}{F}_{ni} = F_{ni}(x_n, x_i, \varphi_i); f_n = f(x_n).$$

Система имеет рекурсивную форму и дает, если $B_{nn} \neq 0$, уравнение для нахождения φ_n по ранее найденным $\varphi_0, \dots, \varphi_{n-1}$. Если же $B_{nn} = 0$, то она дает явное выражение φ_n через $\varphi_0, \dots, \varphi_{n-1}$.

§87. Корректно поставленные и некорректно поставленные задачи

При приближенном решении задач важно знать, корректна ли решаемая задача.

Большое число некорректных задач может быть приведено к уравнению первого рода, имеющему вид

$$Ax = y, x \in X, y \in Y, \tag{1}$$

в котором по заданному не обязательно линейному оператору A, действующему из пространства X в пространство Y, и по заданному элементу y из пространства Y требуется определить решение x в пространстве X. Пространства X и Y всюду ниже будем считать метрическими, а в особо оговариваемых случаях – банаховыми или даже гильбертовыми.

Определение 1. *Следуя Ж.Адамару, задачу отыскания $x \in X$ из уравнения (1) называют корректной (корректно поставленной), если при любой фиксированной правой части $y = y_0 \in Y$ ее решение*

1. *существует в пространстве X;*
2. *единственно в X;*
3. *устойчиво в X, т.е. непрерывно зависит от $y \in Y$.*

Если хотя бы одно из этих условий не выполняется, задачу называют некорректной(некорректно поставленной).

Таким образом, корректность задачи связана с наличием обратного оператора A^{-1}, определенного и непрерывного на всем пространстве Y.

Определение 2. *Задача отыскания $x = R(y)$ из X по известной правой части $y \in Y$ называется устойчивой на пространствах X и Y, если для $\forall \varepsilon > 0$ $\exists \varepsilon > 0$ такое, что из условия $\rho_Y(y_1, y_2) \leq \delta$ следует $\rho_X(x_1, x_2) \leq \varepsilon$, где $x_1 = R(y_1)$, $x_2 = R(y_2)$, $x_1, x_2 \in X$, $y_1, y_2 \in Y$.*

По другому, задача (1) устойчива, если бесконечно малым изменениям правой части уравнения (1) соответствуют бесконечно малые изменения решения.

Хорошо известен построенный Адамаром пример некорректной задачи, представляющий собой задачу Коши для уравнения Лапласа.

В качестве другого примера можно рассмотреть интегральное уравнение Фредгольма первого рода

$$\int_0^1 A(t, s)x(s)ds = y(t), 0 \leq t \leq 1. \tag{2}$$

Будем считать $X = Y = L_2(0, 1)$, а вещественное ядро $A(t, s)$ не только квадратично суммируемо, но и непрерывно. Такое ядро, очевидно, преобразует любую функцию $x(t) \in L_2(0,1)$ в непрерывную функцию, и, следовательно, не для всякой правой части $y(t) \in L_2(0,1)$ решение уравнения (2) существует в пространстве $L_2(0,1)$, т.е. не выполняется первое условие из определения 1.

Если $\lambda = 0$ – собственное значение оператора, т.е. ядро $A(t,s)$ – неполное, то решение в случае его существования неединственно в $L_2(0,1)$, т.е. не выполняется второе условие из определения 1.

Наконец, неустойчивость решения покажем для полного симметричного ядра. Пусть вещественные λ_i – собственные значения ядра и $z_i(t)$, $i = 1, 2, \ldots$ – соответствующие им ортонормированные собственные функции ядра. Тогда, как известно из теории интегральных уравнений, $\lambda_i \neq 0$ и $\lambda_i \to 0$, $i \to \infty$, а решение выражается через правую часть уравнения, представимую в виде $y(t) = \sum\limits_{i=1}^{\infty} y_i z_i(t)$, $y_i = (y, z_i) = \int\limits_0^1 y(t) z_i(t) dt$ с помощью ряда $x(t) = \sum\limits_{i=1}^{\infty} \dfrac{y_i}{\lambda_i} z_i(t)$. Если правые части $y^{(n)}(t)$, $n = 1, 2, \ldots$, сходятся к функции $y(t)$ в пространстве $L_2(0,1)$, т.е.

$$\|y^n - y\|^2 = \sum_{i=1}^{\infty} (y_i^n - y_i)^2 \to 0, n \to \infty,$$

то норма разности соответствующих решений уравнения (2), выражаемая равенством $\|x^n - x\|^2 = \sum\limits_{i=1}^{\infty} \dfrac{(y_i^n - y_i)^2}{\lambda_i^2}$ не только не обязана стремиться к нулю, но и может быть бесконечно большой. В этом легко убедиться, положив $y^n(t) = y(t) + \sqrt{|\ \lambda_n\ |} z_n(t)$. Хотя $\|y^n - y\|^2 =|\ \lambda_n\ |\to 0$, $n \to \infty$, тем не менее $\|x^n - x\|^2 = \dfrac{1}{|\ \lambda_n\ |} \to \infty$, $n \to \infty$, и, следовательно, устойчивость решений отсутствует. Принадлежность точки $\lambda = 0$ предельному множеству точек спектра оператора A, как это было в последнем примере, весьма характерна для линейных неустойчивых задач.

Условия корректности задачи и особенно устойчивости решения представляются настолько естественными с точки зрения приложений математики, что долгое время некорректные задачи считались не имеющими физического смысла и поэтому не изучались. Между тем практика и научные исследования стали одну за одной выдвигать некорректные задачи. Так задача Коши для уравнения Лапласа оказалась важной для геофизических методов разведки полезных ископаемых, а задача Коши для эллиптических уравнений и систем – для сверхзвуковой аэродинамики. Уравнения Фредгольма первого рода приходится решать в спектроскопии и в обратных задачах теории потенциала. Некорректна и задача теплопроводности с обращенным временем, в которой требуется определить распределение температур в прошлом по данным, относящимся к настоящему. Вообще некорректно большинство так называемых обратных задач, в которых по результатам действия какого-либо поля или процесса определяются первоначальные характеристики самого этого поля или процесса.

Если не изменить постановку неустойчивых задач, то обычные методы, применяемые для решения корректных задач, оказываются, естественно, непригодными для задач некорректных, так как сколь бы малой ни была погрешность исходных данных, нельзя быть уверенным в малости погрешности решения. Поэтому потребности практики в решении некорректных задач привели к необходимости пересмотреть классическое понятие корректности и выработать более широкий и приспособленный к реальным нуждам подход. Начало этому было положено в 1943 году А.Н. Тихоновым.

Будем обозначать образ множества $M \subset X$ в пространстве Y при отображении A через $N = AM$.

Определение 3. *Задачу (1) называют корректной по Тихонову на множестве $M \subset X$, а само множество M называют ее множеством (классом) корректности, если:*

1. *точное решение задачи существует и принадлежит множеству M;*
2. *принадлежащее множеству M решение единственно для любой правой части y из множества N;*
3. *принадлежащее множеству M решение устойчиво (непрерывно) относительно $y \in N$.*

Отсюда видно, что при $M = X$ и $N = Y$ корректность по Тихонову совпадает с корректностью по Адамару. Следовательно, корректность по Тихонову достигается за счет сужения рассматриваемого множества решений до класса корректности.

§88. Метод регуляризации

Этот метод был предложен А.Н. Тихоновым почти одновременно с методом квазирешений В.К. Иванова и оказался достаточно удобным в практических вычислениях.

Пусть требуется решить уравнение

$$Ax = y, x \in X, y \in Y, \tag{1}$$

где по заданному не обязательно линейному оператору A и элементу $y \in Y$ требуется найти решение $x \in X$.

Предполагается дополнительно, что A−непрерывен, взаимнооднозначен и, возможно, нелинеен. Предполагаем, что точное решение существует, и подберём регуляризующий (стабилизирующий) функционал $\Omega(x)$ обладающий следующими свойствами:

1. точное решение принадлежит области определения $D(\Omega)$ функционала $\Omega(x)$;

2. на области определения функционал $\Omega(x)$ принимает вещественные неотрицательные решения $(\Omega(x) \geq 0, x \in D(\Omega))$;

3. все множества $M_C = \{x : \Omega(x) \leq C\}, C \geq 0$ являются компактными в пространстве X.

Определение 1. *Множество $M \subset X$ называется компактным, если любая бесконечная последовательность его элементов имеет предельные точки. Если множество M кроме того, замкнуто, то его называют компактом (т.е. компакту принадлежат все предельные точки последовательности).*

Идея метода регуляризации состоит в том, чтобы разыскать минимизирующий элемент некоторого функционала, но не функционала $\rho(Ax, y)$—такая задача была бы эквивалентной уравнению (1) и поэтому тоже некорректной, а несколько «исправленного» и обладающего стабилизирующими свойствами функционала

$$f^{\alpha}(x, y) = \rho^2(Ax, y) + \alpha\Omega(x), x \in D(\Omega),$$

с регуляризующим параметром $\alpha > 0$. Минимизацию будем вести на множестве $D(\Omega)$. Применение метода регуляризации основано на следующих предложениях.

Лемма 1. *Если пространства X и Y – метрические, оператор A непрерывен и взаимно однозначен, а функционал $\Omega(x)$ удовлетворяет указанным выше требованиям, то для $\forall y \in Y$ существует элемент $x^{\alpha} \in D(\Omega)$, минимизирующий на $D(\Omega)$ функционал $f^{\alpha}(x)$ с $\alpha > 0$.*

Доказательство. Так как $\rho^2 \geq 0, \Omega(x) \geq 0, \alpha > 0$, то f^{α}– сглаживающий функционал является ограниченным снизу (т.е. просто неотрицателен). Следовательно, существует $\inf\limits_{x\in D(A)} f^\alpha(x) = m$. По определению инфимума существует минимизирующая последовательность $\{x_n\} \in D(\Omega)$, для которой $f^\alpha(x_n) \to m = \inf\limits_{x\in D(\Omega)} f^\alpha(x), n \to \infty$.

Можно считать, что $f^\alpha(x_{n+1}) < f^\alpha(x_n)$ для всех $n \geq 1$. Тогда $\rho^2(Ax_n, y)+\alpha\Omega(x_n) \leq f^\alpha(x_1)$. Отбросив положительный член $\rho^2(Ax_n, y)$, получим $\alpha\Omega(x_n) \leq f^\alpha(x_1), \Omega(x_n) \leq \dfrac{1}{\alpha} f^\alpha(x_1), n \geq 1$, и вся последовательность $\{x_n\}$ лежит в компакте $M_C, C = \dfrac{1}{\alpha} f^\alpha(x_1)$. Поэтому существует предельная точка $x^\alpha \in M_C \subset D(\Omega)$, которую можно считать пределом последовательности $\{x_n\}$. Покажем, что x^α– искомый элемент. Для этого нужно установить, что $f^\alpha(x^\alpha) = m$, т.е. $\rho^2(Ax^\alpha, y) + \alpha\Omega(x^\alpha) = m$, или что $\Omega(x^\alpha) = \dfrac{1}{\alpha}[m - \rho^2(Ax^\alpha, y)]$.

Так как оператор A непрерывен, а $x_n \to x^\alpha, n \to \infty$ то $Ax_n \to Ax^\alpha$, следовательно, $\rho(Ax_n, y) \to \rho(Ax^\alpha, y), n \to \infty$. Поэтому $\Omega(x_n) = \dfrac{1}{\alpha}[f^\alpha(x_n) - \rho^2(Ax_n, y)] \to \dfrac{1}{\alpha}[m - \rho^2(Ax^\alpha, y)], n \to \infty$ Обозначим $\dfrac{1}{\alpha}[m - \rho^2(Ax^\alpha, y)] = a$, следовательно, $\Omega(x_n) \to a$, значит, для $\forall\varepsilon > 0$ для достаточно больших номеров n $\Omega(x_n) < a + \varepsilon$ по определению предела, но этим равенством определяется компакт $M_{a+\epsilon}$, следовательно, предельная точка последовательности $\{x_n\}$ должна принадлежать компакту $M_{a+\varepsilon}, \forall\varepsilon > 0$, т. е. $x^\alpha \in M_\alpha, \Rightarrow \Omega(x^\alpha) \leq a$. Лемма 1 доказана.

Единственность минимизирующего элемента.

Лемма 2. *Если в дополнение к условиям леммы 1 пространства $X, Y-$ банаховы, оператор $A-$аддитивный, а функционал $\Omega(x)-$строго выпуклый, то минимизирующий элемент x^α единственен.*

Доказательство.

Определение 2. *Множество M банахова пространства X называется выпуклым, если для $\forall x_1, x_2 \in M$ соединяющий их отрезок $\alpha x_1 + (1-\alpha)x_2$, $0 \le \alpha \le 1$ тоже принадлежит M.*

Определение 3. *Заданный на выпуклом множестве M вещественный функционал $\Omega(x)$ называется строго выпуклым (вниз), если для $\forall x_1, x_2 \in M, x \neq x_2$ выполняется строгое неравенство*
$$\Omega\left(\frac{x_1+x_2}{2}\right) < \frac{1}{2}\Omega(x_1) + \frac{1}{2}\Omega(x_2).$$

Предположим противное, что имеются два элемента x_1 и x_2, для которых $f^\alpha(x_1) = f^\alpha(x_2) = m$. Покажем, что тогда должно быть
$f^\alpha\left(\frac{x_1+x_2}{2}\right) < m$. Сначала докажем неравенство $(a+b)^2 \le 2(a^2+b^2)$.
Действительно, $a^2+2ab+b^2 \le 2a^2+2b^2$; $a^2-2ab+b^2 \ge 0$; $(a-b)^2 \ge 0$.

Так как пространства $X, Y-$ банаховы, то в них введена норма, следовательно,

$$\|A\frac{x_1+x_2}{2}-y\|^2=\|\frac{1}{2}(Ax_1-y)+\frac{1}{2}(Ax_2-y)\|^2\leq$$

$$\leq\left(\frac{1}{2}\|Ax_1-y\|+\frac{1}{2}\|Ax_2-y\|\right)^2\leq 2\times\frac{1}{4}\left(\|Ax_1-y\|^2+\|Ax_2-y\|^2\right)=$$

$$=\frac{1}{2}\left(\|Ax_1-y\|^2+\|Ax_2-y\|^2\right)\qquad(2)$$

В силу строгой выпуклости стабилизирующего функционала получим

$$\Omega\left(\frac{x_1+x_2}{2}\right)<\frac{1}{2}\Omega(x_1)+\frac{1}{2}\Omega(x_2).\qquad(3)$$

Сложив (2) и (3), получим (умножив предварительно (3) на α):

$$f^\alpha\left(\frac{x_1+x_2}{2}\right)=\|A\frac{x_1+x_2}{2}-y\|^2+\alpha\Omega\left(\frac{x_1+x_2}{2}\right)<\frac{1}{2}f^\alpha(x_1)+\frac{1}{2}f^\alpha(x_2)=m.$$

Поэтому $f^\alpha\left(\frac{x_1+x_2}{2}\right)<m$, что противоречит определению величины m. Таким образом, двух минимизирующих элементов быть не может и лемма 2 доказана.

Лемма 3. *Если для $y_0\in Y$ существует точное решение уравнения (1) $x_0\in D(\Omega)$, то в условиях леммы 1 с $y=y_0=Ax_0$ имеет место сходимость $x^\alpha\to x_0, a\to 0$.*

Доказательство. Так как x^α–минимизирует функционал $f^\alpha = \alpha\Omega(x_0)$, то $f^\alpha(x^\alpha, y_0) \le f^\alpha(x_0, y_0) = \rho^2(Ax_0, y_0) + \alpha\Omega(x_0) = \alpha\Omega(x_0)$.

Следовательно, $\rho^2(Ax^\alpha, y_0) + \alpha\Omega(x^\alpha) \le \alpha\Omega(x_0)$. Отсюда $\alpha\Omega(x^\alpha) \le f^\alpha(x^\alpha, y_0) \le \alpha\Omega(x_0)$. Так как $\alpha > 0$, то $\Omega(x^\alpha) \le \Omega(x_0)$. Следовательно, все $x^\alpha \in M_C, C = \Omega(x_0)$, т.е. все x^α принадлежат компакту. Кроме того, так как $\rho^2(Ax^\alpha, y_0) + \alpha\Omega(x^\alpha) \le \alpha\Omega(x_0)$ и $\alpha\Omega(x^\alpha) \ge 0$, то $\rho^2(Ax^\alpha, y_0) \le \alpha\Omega(x^\alpha) \to 0, \alpha > 0$, $(\Omega(x_0) = const)$. Поэтому $\rho^2(Ax^\alpha, y_0) \to 0, \alpha \to 0$. Значит, $y_\alpha = Ax^\alpha \to y_0$ на образе компакта при непрерывном отображении A. Но на образе компакта непрерывное и взаимно однозначное отображение имеет непрерывное обратное отображение A^{-1}, т.е. прообразы образов тоже сходятся, т.е. $A^{-1}(Ax^\alpha) \to A^{-1}y_0$, т.е. $x^\alpha \to x_0$. Лемма 3 доказана.

Теорема 1. *Пусть пространства X и Y – банаховы, оператор A– аддитивен, непрерывен и взаимно однозначен, функционал $\Omega(x)$– строго выпуклый и удовлетворяет требованиям 1) – 3) и пусть для $y_0 \in Y$ существует точное решение уравнения (1) $x_0 \in D(\Omega)$. Если вместо точной правой части уравнения $y_0 \in Y$ известны приближения $y_\delta \in Y$ такие, что $\rho(y_\delta, y_0) \le \delta$, и значения параметра α выбирается так, чтобы*

$$\alpha(\delta) \to 0, \delta \to 0, \qquad (4)$$

$$\overline{\lim_{\delta\to 0}}\frac{\delta^2}{\alpha(\delta)} \le \gamma < \infty, \tag{5}$$

то элементы $x_\delta^{\alpha(\delta)}$, *минимизирующие функционал* $f^{\alpha(\delta)}(x; y_\delta)$ *на* $D(\Omega)$, *сходятся к точному решению* x_0 *в пространстве* X *при* $\delta \to 0$.

Доказательство. Существование и единственность каждого минимизирующего элемента $x_\delta^{\alpha(\delta)}$ доказаны в леммах 1 и 2. Сходимость $x_\delta^{\alpha(\delta)} \to x_0, \delta \to 0$ покажем аналогично тому, как это сделано при доказательстве леммы 3. Имеем по определению, что $x_\delta^{\alpha(\delta)}$– минимизирующий элемент:

$$f^{\alpha(\delta)}(x_\delta^{\alpha(\delta)}, y_\delta) \le f^{\alpha(\delta)}(x_0, y_\delta) = \rho^2(Ax_0, y_\delta) + \alpha(\delta)\Omega(x_0) \le$$
$$\le \delta^2 + \alpha(\delta)\Omega(x_0) = \alpha(\delta)\left(\Omega(x_0) + \frac{\delta^2}{\alpha(\delta)}\right). \tag{6}$$

Отсюда видно, что $\Omega(x_\delta^{\alpha(\delta)}) \le \Omega(x_0) + \dfrac{\delta^2}{\alpha(\delta)}$.

Для любых достаточно малых δ в силу $\overline{\lim\limits_{\delta\to 0}}\dfrac{\delta^2}{\alpha(\delta)} \le \gamma$ можно утверждать, что $\Omega(x_\delta^{\alpha(\delta)}) \le \Omega(x_0)+\gamma+\varepsilon$, где $\varepsilon > 0$. Следовательно, наши минимизирующие элементы принадлежат компакту M_C, где $C = \Omega(x_0)+\gamma+\varepsilon$. Посмотрим, как ведут себя образы этих минимизирующих элементов.

$$\rho(Ax_\delta^{\alpha(\delta)}, y_0) \le \rho(y_\delta^{\alpha(\delta)}, y_\delta) + \rho(y_\delta, y_0) \le \rho(y_\delta^{\alpha(\delta)}, y_\delta) + \delta.$$

Так как из (6) $\rho^2(Ax_\delta^{\alpha(\delta)}, y_\delta) \le \delta^2 + \alpha(\delta)\Omega(x_0)$, то
$\rho(Ax_\delta^{\alpha(\delta)}, y_0) \le \delta + \sqrt{\delta^2 + \alpha(\delta)\Omega(x_0)} \to 0, \delta \to 0$,
поскольку $\alpha(\delta) \to 0, \delta \to 0$.

Итак, образы минимизирующих элементов стремятся к y_0, поэтому их прообразы $x_\delta^{\alpha(\delta)} \to x_0, \delta \to 0$. Теорема 1 доказана.

Эта теорема показывает, что для регуляризации нужно выбирать параметр α стремящимся к нулю, но не быстрее δ^2.

Не следует думать, что более быстрое убывание α обязательно неприемлемо: легко построить пример, в котором и более высокая, и даже любая скорость стремления к нулю параметра α обеспечивает сходимость метода регуляризации. Однако в условиях теоремы 1 этого доказать нельзя, так как существуют задачи, в которых нарушение условия (5) ведет к расходимости элементов $x_\delta^{\alpha(\delta)}$ так что в классе задач, описываемых теоремой 1, это условие необходимо.

В гильбертовом пространстве X функционал $\Omega(x)$ можно брать более простой, в виде $\|x\|^2$, и хотя множества M_C в этом случае лишь слабо компактны, сходимость регуляризованных решений получается сильная. Такой выбор стабилизирующего функционала, кроме своей простоты, удобен еще тем, что область его определения совпадает со всем пространством X. Для регуляризуемости уравнения достаточно лишь фак-та существования точного решения. Но условия на параметр α при этом несколько жестче: α должно стремиться к нулю медленнее, чем δ^2.

Теорема 2. *Пусть X – гильбертово пространство, $\Omega(x) = \|x\|^2$ и выполнены остальные условия теоремы 1. Тогда при $\alpha(\delta)$, удовлетворяющих соотношениям (4) и (5) с $\gamma = 0$, регуляризованные элементы $x_\delta^{\alpha(\delta)}$ сходятся при $\delta \to 0$ к точному решению по норме пространства X.*

Доказательство. При $\forall\gamma$ слабая сходимость имеет место $x_\delta^{\alpha(\delta)} - \to x_0, \delta \to 0$. Но из слабой сходимости следует, что $\|x_0\| \le$ $\le \varliminf\limits_{\delta\to 0} \|x_\delta^{\alpha(\delta)}\|$. Поскольку $\Omega(x) = \|x\|^2$ и $\Omega(x_\delta^{\alpha(\delta)}) \le \Omega(x_0) + \dfrac{\delta^2}{\alpha(\delta)}$ (из теоремы 1, из (6)), следовательно, $\|x_\delta^{\alpha(\delta)}\|^2 \le \|x_0\|^2 + \dfrac{\delta^2}{\alpha(\delta)}$.

Тогда $\varlimsup\limits_{\delta\to 0}\|x_{\delta^2}^{\alpha(\delta)}\|^2 \le \varlimsup\limits_{\delta\to 0}\left(\|x_0\|^2 + \dfrac{\delta^2}{\alpha(\delta)}\right) = \|x_0\|^2$.

Следовательно, $\|x_\delta^{\alpha(\delta)}\| \to \|x_0\|, \delta \to 0$, а это вместе со слабой сходимостью дает ввиду свойств гильбертова пространства и сильную сходимость: $\|x_\delta^{\alpha(\delta)} - x_0\| \to 0, \delta \to 0$. Теорема 2 доказана.

В случае, когда $X = Y$ – гильбертово пространство, то выбор $\Omega(x) = \|x\|^2$ дает для функционала $f^\alpha(x, y)$ уравнение Эйлера вида

$$A^*A + \alpha x = A^*y. \tag{7}$$

Таким образом, мы получили в явной форме линейный регуляризатор $R_\alpha = (A^*A + \alpha E)^{-1}A^*$, E– тождественный оператор, и тем самым теорема 2 дает обоснование для перехода от уравнения I рода (1) к уравнению II рода (7). Это обоснование можно получить и иначе, если учесть, что в наших условиях уравнение (1) эквивалентно уравнению
$A^*Ax = A^*y$ с положительным оператором A^*A. По сравнению с этим последним оператором оператор уравнения (7) имеет спектр, сдвинутый вправо на $\alpha > 0$, и, следовательно, не содержащий нуля. Поэтому оператор $A^*A + \alpha E$ заведомо имеет ограниченный обратный, и регуляризация сводится к разумному согласованию параметра α со значением погрешности δ. Правила такого согласования дает теорема 2.

Получим априорную оценку погрешности для метода регуляризации, которая позволяет рационально выбирать значение параметра α по заданному фиксированному δ.

Теорема 3. *Пусть $X = Y$– гильбертовы пространства, $\Omega(x) = \|x\|^2$ и выполнены остальные условия теоремы 2. Тогда, если точное решение x_0 является истокопредставимым с оператором A^*, т.е. $x_0 = A^*z_0, \|z_0\| \le C_0$, то справедливо следующее неравенство*

$$\|x_\delta^\alpha - x_0\| \le \frac{\delta}{\sqrt{\alpha}} + \sqrt{\alpha}C_0.$$

Оптимальная оценка получается при выборе $\alpha = \dfrac{\delta}{C_0}$.

Доказательство. Очевидно, что $\|x_\delta^\alpha - x_0\| \le \|x_\delta^\alpha - x_0^\alpha\| + \|x_0^\alpha - x_0\|$, где x_0^α– минимизирующий элемент того самого функционала, у которого в правой части стоит не y_δ, а y_0: $f^\alpha(x, y_0) = \|Ax - y_0\|^2 + \alpha\|x\|^2$.

Оценим первое слагаемое справа. Как отмечалось выше, для функционала $f^\alpha(x, y) = \|Ax - y\|^2 + \alpha\|x\|^2$ уравнением Эйлера служит уравнение $A^*Ax + \alpha x = A^*y$. Поэтому элементы x_δ^α и x_o^α удовлетворяют соответствующим равенствам

$$A^*Ax_\delta^\alpha + \alpha x_\delta^\alpha = A^*y_\delta, \;\; A^*Ax_0^\alpha + \alpha x_0^\alpha = A^*y_0.$$

Следовательно, разность $x_\delta^\alpha - x_0^\alpha$ есть минимизирующий элемент функционала $f^\alpha(x, y_\delta - y_0)$. Это получается, если из первого уравнения вычесть второе: $A^*A(x_\delta^\alpha - x_0^\alpha) + \alpha(x_\delta^\alpha - x_0^\alpha) = A^*(y_\delta - y_0)$.

По определению минимизирующего элемента
$f^\alpha(x_\delta^\alpha - x_0^\alpha, y_\delta - y_0) \le f^\alpha(0, y_\delta - y_0) = \|y_\delta - y_0\|^2 \le \delta^2$

и поскольку $\alpha\|x_\delta^\alpha - x_0^\alpha\|^2 \le f^\alpha(x_\delta^\alpha - x_0^\alpha, y_\delta - y_0)$, то $\|x_\delta^\alpha - x_0^\alpha\|^2 \le \dfrac{\delta}{\sqrt{\alpha}}$.

Переходим ко второму слагаемому. Предположим, что $x_0 = A^*z_0$, где x_0– решение уравнения $A^*Ax_0 = A^*y_0$, а x_0^α– решение уравнения $A^*Ax_0^\alpha + \alpha x_0^\alpha = A^*y_0$, Рассмотрим разность этих уравнений:

$A^*A(x_0^\alpha - x_0) + \alpha x_0^\alpha = 0$ или $A^*A(x_0^\alpha - x_0) + \alpha(x_0^\alpha - x_0) = -\alpha x_0$.

$A^*A(x_0^\alpha - x_0) + \alpha(x_0^\alpha - x_0) = -\alpha A^*z_0$ – это уравнение Эйлера для функционала $f^\alpha(x, -\alpha z_0)$. Отсюда следует, что элемент $x_0^\alpha - x_0$ минимизирует функционал $f^\alpha(x, -\alpha z_0)$. Очевидно поэтому, что

$f^\alpha(x_0^\alpha - x_0, -\alpha z_0) \le f^\alpha(0, -\alpha z_0) = \alpha^2\|z_0\|^2$. Так как $\alpha\|x_0^\alpha - x_0\|^2 \le f^\alpha(x_0^\alpha - x_0, -\alpha z_0)$, то $\|x_0^\alpha - x_0\| \le \sqrt{\alpha}\|z_0\|$.

Поэтому $\|x_\delta^\alpha - x_0\| \le \|x_\delta^\alpha - x_0^\alpha\| + \|x_0^\alpha - x_0\| \le \dfrac{\delta}{\sqrt{\alpha}} + \sqrt{\alpha}C_0$.

Оптимизируем полученную оценку по α. Обозначим $\sqrt{\alpha} = t$, тогда $f(t) = \dfrac{\delta}{t} + C_0 t, f'(t) = -\dfrac{\delta}{t^2} + C_0 . f'(t) = 0, \Rightarrow t^2 = \dfrac{\delta}{C_0}$, т. е. $t^* = \sqrt{\dfrac{\delta}{C_0}}$, $\alpha = \dfrac{\delta}{C_0}, f''(t) = \dfrac{2\delta}{t^3}, f''(t^*) = \dfrac{2\delta}{\left(\sqrt{\dfrac{\delta}{C_0}}\right)^3} > 0$.

Следовательно, t^* – точка минимума. Отсюда

$$\|x_\delta^\alpha - x_0\|_{\text{опт}} \le \frac{\delta}{\sqrt{\dfrac{\delta}{C_0}}} + \sqrt{\frac{\delta}{C_0}}C_0 = 2\sqrt{\delta C_0}.$$

Теорема 3 доказана.

Как видно из доказанной теоремы, априорное значение истообразности точного решения $x_0 = A^* z_0$ с оценкой $\|z_0\| \le C_0$ весьма важно для получения оценки погрешности метода регуляризации и для разумного выбора значения регуляризующего параметра α.

Пример. Рассмотрим в пространстве $L_2(0,1)$ интегральное уравнение Фредгольма I рода.

$$\int_0^1 A(t,s)x(s)ds = y(t), 0 \le t \le 1 \tag{8}$$

с симметричным положительным ядром

$$A(t,s) = \begin{cases} t(1-s), 0 \le t \le s \le 1 \\ s(1-t), 0 \le s \le t \le 1 \end{cases} \tag{9}$$

и точной правой частью

$$y(t) = \frac{t(t-1)}{12}(t^2 - t - 1). \tag{10}$$

Точным решением задачи (8) является $x(t) = t(1-t)$. Нетрудно проверить, что это решение истокопредставимо $x(t) = \int\limits_0^1 A(t,s)z_0(s)ds$, где $z_0(t) = 2$, причем $\|z_0(t)\| = 2$. Оператор уравнения (1) аддитивен, взаимно однозначен и непрерывен.

Обычно на практике точное значение $y(t)$ неизвестно, а известно δ-приближение y_δ. Чтобы имитировать эту ситуацию, будем считать заданными значения $\tilde{y}_i, i = \overline{1,m}$, полученные следующим образом: в точках $t_i = h\left(t - \dfrac{1}{2}\right)$, $i = \overline{1,m}$, при $h = \dfrac{1}{m}$ вычислим значения y_i по формулам (10) и округлим их до трех значащих цифр после запятой, получим $\widetilde{y}_i$. При этом величину погрешности δ приближенно оценим так $\int\limits_0^1 [y(t) - \widetilde{y}(t)]^2 dt \approx \sum\limits_{i=1}^{m} h[y(t_i) - \widetilde{y}_i]^2 \leq mh(10^{-3})^2 = 10^{-6}$. Следовательно, $\delta = 10^{-3}$. Для решения уравнения (8) применим метод регуляризации со стабилизирующим функционалом $\Omega(x) = \|x\|^2$. Запишем уравнение Эйлера

$$\int\limits_0^1 K(t,s)x(s)ds + \alpha x(t) = \int\limits_0^1 A(s,t)y(s)ds, \tag{11}$$

где ядро $K(t,s)$ оператора A^*A легко вычисляется

$$K(t,s) = \begin{cases} \dfrac{t(s-1)}{6}(t^2+s^2-2s), 0 \leq t \leq s \leq 1 \\ \dfrac{s(t-1)}{6}(t^2+s^2-2t), 0 \leq s \leq t \leq 1. \end{cases}$$

Задача (11) при $\alpha > 0$ уже корректна, и для её решения используем метод замены ядра квадратурной суммой. Выберем формулу средних прямоугольников, получим

$$\sum_{j=1}^{m} K(t_i, s_j)hx_j + \alpha x_i = Y_i, i = \overline{1,m}, \tag{12}$$

где $Y_i = \sum_{j=1}^{m} hA(s_j, t_i)\bar{y}_j, i = \overline{1, m}$. Здесь $t_i = h\left(i - \frac{1}{2}\right), s_j = h\left(j - \frac{1}{2}\right)$, $i, j = \overline{1, m}, h = \frac{1}{m}$.

Исходная задача свелась к решению системы уравнений (12) с подходящим значением параметра α. Выбирать α будем из следующих соображений. Будем считать априори известным, что точное решение исток представимо через некоторую функцию $z_0(t)$ с оценкой $\|z_0\| \leq 4$. Следовательно, $c_0 = 4$ и $\alpha_{\text{опт}} = \frac{\delta}{c_0} = \frac{10^{-3}}{4} = 0,00025$.

Литература

1. Бахвалов, Н. С. Численные методы / Н. С. Бахвалов, Н. П. Жидков, Г. М. Кобельков. – М., 2004.

2. Калиткин, Н. Н. Численные методы: учеб. пособие / Н. Н. Калиткин. – М., 1978.

3. Крылов, В. И. Вычислительные методы: учеб. пособие: в 2 т. / В. И. Крылов, В. В. Бобков, П. И. Монастырный. – М., 1976. – Т. 1, 1977. – Т. 2.

4. Крылов, В. И. Приближенное вычисление интегралов / В. И. Крылов. – М., 1967.

5. Марчук, Г. И. Методы вычислительной математики: учеб. пособие / Г. И. Марчук. – М., 1989.

6. Самарский, А. А. Введение в численные методы: учеб. пособие / А. А. Самарский. – М., 1983.

7. Самарский, А.А. Численные методы: учеб. пособие /А. А.Самарский, А. В. Гулин. – М., 1989.

8. Тихонов, А. Н. Методы решения некорректных задач: учеб. пособие / А. Н. Тихонов, В. Я. Арсенин. – М., 1986.

9. Крылов, В. И. Справочная книга по численному интегрированию / В. И. Крылов, Л.Т. Шульгина. – М., 1966.

10. Крылов, В. И. Интегральные уравнения, некорректные задачи и улучшение сходимости / В. И. Крылов, В. В. Бобков, П. И. Монастырный. – Минск : Наука и техника, 1984.

11. Вакульчик, П.И. Методы численного анализа / П. И. Вакульчик – Минск : БГУ, 2002.

12. Мысовских, И. П. Лекции по методам вычислений / И. П. Мысовских. – СПб., 1998.

13. Вержбицкий, В.М. Основы численных методов: учебник для вузов / В.М. Вержбицкий. – М., 2002.