СОДЕРЖАНИЕ

Предисловие..... 7

Лабораторная работа 1. **Погрешности вычислений**..... 9

Введение (9). Погрешности вычислений. (9). Вычисление значения функции с помощью разложения ее в ряд Тейлора (13). Вычисление производной (15). Формула первого порядка аппроксимации (15). Формула второго порядка аппроксимации (17). Формула четвертого порядка аппроксимации (18). Контрольные вопросы (19). Порядок выполнения работы (21). Библиографическая справка (23).

Лабораторная работа 2. **Табличное задание и интерполирование функций**..... 24

Введение (24). Задача интерполяции (24). Алгебраическая интерполяция (25). Непосредственное вычисление коэффициентов интерполяционного полинома (25). Интерполяционный полином в форме Лагранжа. Интерполяционный многочлен в форме Ньютона (26). Формула для погрешности алгебраической интерполяции (27). О сходимости интерполяционного процесса (28). Обусловленность задачи построения интерполяционного многочлена для функции, заданной таблицей (29). Классическая кусочно-многочленная интерполяция (31). Оценка неустранимой погрешности при приближении функции по ее значениям в узлах интерполяции. Выбор степени кусочно-многочленной интерполяции (31). Насыщаемость (гладкостью) кусочно-многочленной интерполяции (32). Кусочно-многочленная гладкая интерполяция (сплайны). Локальные сплайны (32). Нелокальная гладкая кусочно-многочленная интерполяция (34). Тригонометрическая интерполяция (35). Многочлены Чебышёва (38). Алгебраический интерполяционный полином на сетке из нулей полинома Чебышёва (38). Алгебраический интерполяционный полином на сетке из экстремумов полинома Чебышёва (39). Чувствительность интерполяционного тригонометрического многочлена к погрешностям задания функции в узлах интерполяции (39). Контрольные вопросы (40). Порядок выполнения работы (41). Библиографическая справка (44).

Лабораторная работа 3. **Численное интегрирование**..... 45

Введение (45). Способы конструирования квадратурных формул (45). Погрешность квадратурных формул (48). Приемы вычисления несобственных интегралов (52). Контрольные

вопросы (56). Порядок выполнения работы (56). Библиографическая справка (57).

Лабораторная работа 4. **Численное решение систем линейных уравнений** 58

Введение (58). Обусловленность систем линейных уравнений (59). Метод Гаусса (61). Метод сопряженных градиентов (63). Метод простых итераций (65). Метод Зейделя и метод релаксации (67). Метод простых итераций с оптимальным параметром (68). Трехслойный метод Чебышева (70). Метод с оптимальным набором параметров (71). Метод минимальных невязок (72). Метод скорейшего спуска (73). Контрольные вопросы (74). Порядок выполнения работы (75). Некоторые рекомендации по работе с программой (76). Библиографическая справка (78).

Лабораторная работа 5. **Численное решение нелинейных уравнений** 79

Введение (79). Нелинейные уравнения. Теоретическая справка (79). Метод простой итерации (80). Метод Ньютона (81). Метод секущих (83). Мера погрешности (84). Сходимость по аргументу (84). Сходимость по функции (85). Контрольные вопросы (85). Порядок выполнения работы (86). Библиографическая справка (87).

Лабораторная работа 6. **Переопределенные системы линейных уравнений. Метод наименьших квадратов** 88

Введение (88). Переопределенная система линейных алгебраических уравнений (88). Геометрический смысл метода наименьших квадратов (90). Оценка обусловленности матрицы системы МНК (91). Метод Гаусса (92). Метод сопряженных градиентов (92). Полиномы Лежандра (92). Порядок выполнения работы (93). Некоторые рекомендации по работе с программой (95). Библиографическая справка (96).

Лабораторная работа 7. **Численное решение обыкновенных дифференциальных уравнений. Задача Коши** 97

Введение (97). Численные методы решения задачи Коши для обыкновенных дифференциальных уравнений (98). Устойчивость (99). Дифференциальная задача (99). Сеточная область (100). Разностная задача (100). Погрешность метода (100). Явные методы Рунге–Кутты (100). Метод Рунге–Кутты первого порядка точности (101). Метод Рунге–Кутты второго порядка точности (101). Метод Рунге–Кутты третьего порядка точности (102). Метод Рунге–Кутты четвертого

порядка точности (102). Экстраполяция Ричардсона (102). Схема второго порядка с центральной разностью (103). Теоремы об устойчивости методов Рунге–Кутты (103). Контрольные вопросы (104). Порядок выполнения работы (105). Библиографическая справка (108).

Лабораторная работа 8. **Численное решение обыкновенных дифференциальных уравнений. Краевая задача** 109

Введение (109). Пример краевой задачи (109). Линейная краевая задача (110). Метод численного построения общего решения (110). Конечно-разностный метод (прогонки) (111). Нелинейная краевая задача (112). Метод стрельбы (112). Вычислительная неустойчивость задачи Коши (114). Метод линеаризации (115). Порядок выполнения работы (116). Библиографическая справка (118).

Лабораторная работа 9. **Численное решение дифференциальных уравнений в частных производных гиперболического типа. Уравнение переноса** 119

Введение (119). Дифференциальная задача (119). Сеточная область (120). Пример разностной задачи (120). Шаблон разностной схемы (120). Погрешность метода (120). Невязка (121). Спектральный признак устойчивости (122). Явный левый уголок (122). Явная четырехточечная схема (123). Явная центральная трехточечная схема (124). Гибридная схема (схема Федоренко) (124). Схема «чехарда» (126). Неявный левый уголок (126). Неявный правый уголок (127). Неявная четырехточечная схема (127). Схема «прямоугольник» (128). Неявная шеститочечная схема (129). Точное решение задачи Коши для однородного уравнения (129). Порядок выполнения работы (130). Библиографическая справка (131).

Лабораторная работа 10. **Численное решение дифференциальных уравнений в частных производных гиперболического типа. Волновое уравнение** 132

Введение (132). Дифференциальная краевая задача (132). Сеточная область (133). Разностная задача (133). Шаблон разностной схемы (133). Ошибка аппроксимации (133). Спектральный признак устойчивости (134). Способы конструирования разностных схем (135). Сведение задачи (10.1) к задаче для системы двух уравнений первого порядка (138). Контрольные вопросы (149). Порядок выполнения работы (150). Библиографическая справка (150).

Лабораторная работа 11. **Численное решение дифференциальных уравнений в частных производных параболического типа. Уравнение теплопроводности** 151

Введение (151). Дифференциальная краевая задача (152). Сеточная область (152). Пример разностной задачи (152). Шаблон разностной схемы (153). Спектральный признак устойчивости (153). Шеститочечная параметрическая схема (154). Схема Франкела–Дюфорта (155). Схема Ричардсона (155). Явная центральная четытрехточечная схема (156). Схема Алена–Чена (157). Нецентральная явная схема (157). Схема Саульева (158). Точные решения тестовых краевых задач для одномерного линейного уравнения теплопроводности (159). Порядок выполнения работы (160). Библиографическая справка (161).

Лабораторная работа 12. **Численное решение уравнений эллиптического типа. Уравнение Пуассона** 162

Введение (162). Аппроксимация и устойчивость простейшей разностной схемы (163). Обусловленность систем линейных уравнений (167). Метод дискретного преобразования Фурье (167). Метод сопряженных градиентов (169). Метод простых итераций (169). Метод с оптимальным параметром (169). Трехслойный метод Чебышева (169). Метод спектрально-эквивалентных операторов (169). Контрольные вопросы (170). Порядок выполнения работы (171). Библиографическая справка (171).

Лабораторная работа 13. **Метод разностных потенциалов** 172

Введение (172). Форма области и сетка (173). Сеточные множества (173). Разностная вспомогательная задача (174). Разностный потенциал (175). Граничный проектор (175). Решение краевой задачи (176). Оператор вычисления плотности (177). Вычисление нормальной производной (178). Кусочно-кубическая интерполяция (179). Порядок выполнения работы (180). Библиографическая справка (181).

Приложение 1. **Теоретическая справка к работам 9–12** 182

Разностные методы (182). Спектральный признак устойчивости для эволюционных уравнений (184).

Приложение 2. **Некоторые рекомендации при работе с системой ОВМ** 188

Редактирование (188). Вывод списка графиков на экран (189).

Список литературы 191

ПРЕДИСЛОВИЕ

Предлагаемое учебное пособие включает описание лабораторных работ по вычислительной математике с использованием разработанного на кафедре вычислительной математики МФТИ практикума «Основы вычислительной математики». Теоретической основой практикума служит книга В. С. Рябенького «Введение в вычислительную математику» [1]. При подготовке практикума нашли свое отражение и другие учебники, созданные на кафедре [2–4]. Конечно, предлагаемое пособие не заменяет собой учебник, так как содержит лишь краткие теоретические справки по темам предлагаемых работ. Для более подробного изучения материала требуется знакомство с другими источниками. Список рекомендованной литературы приведен в конце книги.

Изучение вычислительной математики в последнее время тесно связано с практикой на ЭВМ. Примером таких «машинно-ориентированных» курсов служат недавние переводы книг [5, 6]. Можно отметить, что выбор основных тем для практикума вполне соответствует современным тенденциям [5].

Предлагаемое пособие в корне отличается от зарубежных аналогов. Так, в книге Дж. Каханера, К. Моулера и С. Нэша [5] в качестве вычислительной основы использованы программы из научной библиотеки SLATEK министерства энергетики США, написанной на фортране. Книга Дж. Голуба и Ч. Ван Лоуна [6], как и большинство других зарубежных учебников, ориентирована на использовании MATLAB. С одной стороны, это является достоинством «компьютеризированных» курсов, рассчитанных на профессионалов (не обязательно вычислителей) — знакомство студентов с широко распространенными профессиональными пакетами. С другой стороны, при таком подходе страдает методическая сторона, поскольку ни один прикладной пакет не содержит неустойчивый метод, не слишком удачную аппроксимацию и т. п. Вместе с тем, эффекты, проявляющиеся при неудачном выборе метода, незнание границ его применимости могут привести к «открытиям» в соответствующих предметных областях. Вопросы, которые по определению не подлежат реализации в прикладных пакетах, находят свое место в рамках лабораторного практикума.

Хорошо подобранные примеры и задания, прямое моделирование изучаемых процессов дают возможность освоить наиболее известные методы, традиционно используемые при решении научных и практических задач на ЭВМ, понять границы их применимости. Графические возможности ЭВМ позволяют в понятной и наглядной форме познакомиться с характерными эффектами, возникающими при численном решении задач. Практикум можно использовать для проведения лекционных и семинарских занятий, практических работ на ЭВМ, при самостоятельном изучении вычислительной математики. Отличительными чертами пакета являются: наличие контекстно-зависимой подсказки, гипертекстовой системы помощи, интерактивного графического интерфейса, прямое моделирование исследуемых задач с возможностью интерактивного изменения параметров моделирования.

Пакет состоит из 13 работ, каждая из которых содержит краткую справку, контрольные вопросы, порядок выполнения лабораторной работы и краткие рекомендации по работе с программой. К большинству работ также приложена краткая библиографическая справка по теме работы. Практикум разработан в 1992 г. на кафедре вычислительной математики МФТИ и с тех пор успешно применяется при проведении занятий по вычислительной математике в МФТИ, МЭИ и других вузах. Он послужил основой первого в России электронного учебника, созданного под эгидой Российского НИИ информационных систем [7].

Авторы выражают благодарность выпускникам МФТИ, которые в свои студенческие годы отдали много времени и сил на написание программ практикума. Это В. В. Байков, Д. Л. Будько, К. Б. Бухаров, А. Ю. Езерский, А. Б. Константинов, С. А. Корытник, Ю. П. Кравченко, Ю. Д. Крикунов, Д. В. Лунев, В. А. Торгашев, А. А. Тренихин, Г. Л. Химичев.

Авторы благодарны издательству МЗ-Пресс за возможность осуществления данного проекта.

Лабораторный практикум доступен по адресу в Интернете: http://cs.mipt.ru/nummeth.

Желаем Вам успешного освоения курса вычислительной математики и надеемся, что наш практикум поможет сделать его более приятным и увлекательным.

ЛАБОРАТОРНАЯ РАБОТА 1

ПОГРЕШНОСТИ ВЫЧИСЛЕНИЙ

1.1. Введение

В этой работе Вы познакомитесь с основными источниками возникновения погрешности. На специально подобранных примерах изучите влияние конечной арифметики на достоверность результатов, получаемых при численном решении задачи. В частности, для функции, представляемой сходящимся рядом Тейлора с теоретически бесконечным радиусом сходимости, вычислить ее значение с заданной точностью путем суммирования ряда удается лишь для сравнительно небольших значений аргумента. Реальный «радиус сходимости» весьма невелик, и он сильно зависит от числа значащих цифр, используемых для представления чисел в ЭВМ.

Вычисление производной с использованием формул численного дифференцирования также таит в себе много интересного. Все это Вы узнаете, если проделаете предлагаемую работу, но прежде чем Вы начнете ее выполнять, советуем ознакомиться с данной теоретической справкой, которая, конечно же, ни в коей мере не заменяет учебника.

Выполнение этой работы необходимо для понимания реальной ситуации, в которой используются рассматриваемые в других работах численные методы решения задач.

1.2. Погрешности вычислений. Теоретическая справка

Напомним некоторые понятия, связанные с погрешностями. Если a — точное значение некоторой величины, a^* — ее приближенное значение, то *абсолютной погрешностью* величины a^* обычно называют наименьшую величину $\Delta(a^*)$, про которую известно, что

$$|a^* - a| \le \Delta(a^*).$$

Относительной погрешностью приближенного значения называют наименьшую величину $\delta(a^*)$, про которую известно, что

$$\left| \frac{(a^* - a)}{a^*} \right| \le \delta(a^*).$$

В любой вычислительной задаче по некоторым входным данным требуется найти ответ на поставленный вопрос. Для вычисления значения функции $y = f(x)$ при $x = t$ входными данными задачи служат число x и закон f, по которому каждому значению аргумента x ставится в соответствии значение функции $y = f(x)$.

Если ответ можно дать с любой точностью, то погрешность отсутствует. Но обычно ответ удается найти лишь приближенно. Погрешность задачи вызывается тремя причинами.

Первая — неопределенность при задании входных данных, которая приводит к неопределенности в ответе. Ответ может быть указан лишь с погрешностью, которая называется *неустранимой.*

Проиллюстрируем понятие неустранимой погрешности на примере. Пусть функция $f(x)$ известна приближенно, например, она отличается от $\sin x$ не более чем на величину $\varepsilon > 0$:

$$\sin(x) - \varepsilon \le f(x) \le \sin(x) + \varepsilon. \qquad (1.1)$$

Кроме того, пусть значение аргумента $x = t$ получается приближенным измерением, в результате которого получаем $x = t^*$, причем известно, что t лежит в пределах

$$t^* - \delta \le t \le t^* + \delta, \qquad (1.2)$$

где $\delta > 0$ — число, характеризующее точность измерения (для определенности будем считать, что функция $\sin t$ на отрезке (1.2) монотонно возрастает).

Величиной $y = f(t)$ может оказаться любая точка отрезка $y \in [a,\ b]$ (см. рис. 1), где $a = \sin(t^* - \delta) - \varepsilon,\quad b = \sin(t^* + \delta) + \varepsilon.$ Понятно, что, приняв за приближенное значение величины $y = f(x)$ любую точку y^* отрезка $[a, b]$, можно гарантировать оценку погрешности:

$$|y - y^*| \le |b - a|. \qquad (1.3)$$

Эту гарантированную оценку погрешности нельзя существенно улучшить при имеющихся неполных входных данных.

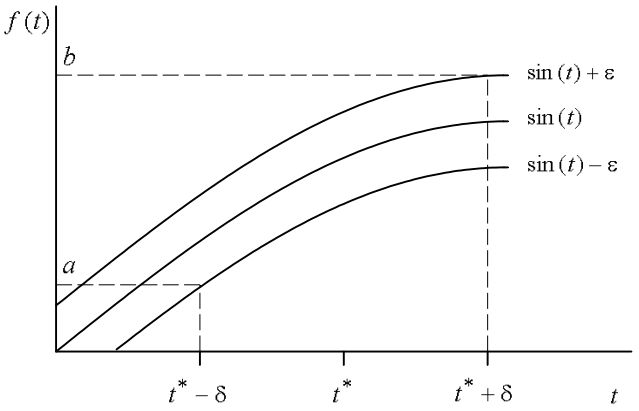


Рис. 1

Самая малая погрешность, получается, если принять за y середину отрезка [*a*, *b*], положив

$$y^* = y^*_{\text{опт}} = \frac{|b+a|}{2}.$$

Тогда справедливая оценка

$$|y - y^*_{\text{опт}}| \le \frac{|b-a|}{2}. \qquad (1.4)$$

Таким образом, $0{,}5|b-a|$ и есть та *неустранимая* (не уменьшаемая) *погрешность*, которую можно гарантировать при имеющихся неопределенных входных данных в случае самого удачного выбора приближенного решения $y^*_{\text{опт}}$. Оптимальная оценка (1.4) ненамного лучше оценки (1.3). Поэтому не только о точке $y^*_{\text{опт}}$, но и о любой точке $y^* \in [a,\ b]$ условимся говорить, что она является приближенным решением задачи вычисления числа $y(t)$, найденным с неустранимой погрешностью, а вместо $0{,}5|b-a|$ из (1.4) за величину неустранимой погрешности примем (условно) число $|b-a|$.

Вторая причина возникновения погрешности состоит в том, что при фиксированных входных данных ответ вычисля-

ется с помощью приближенного метода. Возникает погрешность, связанная с выбором метода — *погрешность метода вычислений*. Проиллюстрируем это понятие на следующем простом примере.

Положим $y^* = \sin t^*$. Точка y^* выбрана среди других точек отрезка [*a*, *b*] (см. выше по поводу неустранимой погрешности), так как она задается при помощи удобной для дальнейшего формулы.

Воспользуемся разложением функции $\sin t$ в ряд Тейлора:

$$\sin t = t - \frac{t^3}{3!} + \frac{t^5}{5!} - \ldots \tag{1.5}$$

Для вычисления значения y^* можно выбрать одно из следующих выражений:

$$y^* \approx y_1^* = t^*,$$

$$y^* \approx y_2^* = t^* - \frac{t^{*3}}{3!}, \tag{1.6}$$

$$y^* \approx y_n^* = \sum_{k=0}^{n} (-1)^k \frac{t^{*(2k+1)}}{(2k+1)!}.$$

Выбирая для приближенного вычисления y^* одну из формул (1.6), тем самым выбираем метод вычисления.

Величина $|y^* - y_n^*|$ — *погрешность метода вычисления*.

Фактически выбранный метод вычисления зависит от параметра *n* и позволяет добиться, чтобы погрешность метода была меньше любой наперед заданной величины за счет выбора этого параметра.

Очевидно, нет смысла стремиться, чтобы погрешность метода была существенно (во много раз) меньше неустранимой погрешности. Поэтому число *n* не стоит выбирать слишком большим. Однако, если *n* слишком мало и погрешность метода существенно больше неустранимой погрешности, то избранный способ не полностью использует информацию о решении, содержащуюся во входных данных. Часть этой информации теряется.

Наконец, сам выбранный приближенный метод реализуется неточно из-за ошибок округления при вычислениях на реаль-

ном компьютере. Так, при вычислении y_n^* по одной из формул (1.6) на реальном компьютере в результате ошибок округления мы получим значение $\tilde{y}_n^*$.

Величину $|y_n^* - \tilde{y}_n^*|$ называют *погрешностью округления*. Она не должна быть существенно больше погрешности метода. В противном случае произойдет потеря точности метода за счет ошибок округления. Точность метода вычислений также целесообразно согласовывать с величиной ожидаемых ошибок округления.

Погрешность результата складывается, таким образом, из неустранимой погрешности, погрешности метода и погрешности округления. Рассмотрим несколько простых примеров.

1.3. Вычисление значения функции с помощью разложения ее в ряд Тейлора

Пусть требуется вычислить значения $y = \sin t$. Воспользуемся разложением функции $\sin t$ в окрестности нуля в ряд Тейлора, радиус сходимости которого для данной функции равен бесконечности:

$$\sin t = t - \frac{t^3}{3!} + \frac{t^5}{5!} - \ldots$$

Для вычисления y можно воспользоваться одним из приближенных выражений:

$$y^* \approx y_1^* = t^*,$$

$$y^* \approx y_2^* = t^* - \frac{t^{*3}}{3!},$$

$$y^* \approx y_n^* = \sum_{k=0}^{n} (-1)^k \frac{t^{*(2k+1)}}{(2k+1)!}.$$

Выбирая для вычисления y одну из приведенных формул, мы тем самым выбираем приближенный метод вычисления, точность которого определяется числом привлекаемых членов ряда n.

Ряд Тейлора для функции $\sin t$ является знакопеременным,

сходится для любого значения t, а его частичная сумма отличается от точного значения функции не более, чем на величину первого отброшенного члена ряда. Выбирая n так, чтобы

$$\frac{t^{2n+1}}{(2n+1)!} \le \varepsilon,$$

можно добиться любой наперед заданной точности ε.

Однако при вычислениях на реальном компьютере получить результат с требуемой точностью для t (которое существенно больше единицы) не удается из-за быстрого роста ошибок округления. Последние тем больше, чем больше t. Это связано с различным характером поведения величины членов ряда Тейлора при $t > 1$ и $t < 1$. При $t < 1$ члены ряда по абсолютной величине монотонно убывают в зависимости от n. При $t > 1$ члены ряда по модулю сначала растут (тем сильнее, чем больше t) и только потом, достигнув при некотором $k = m$ максимума, начинают убывать и стремиться к нулю при $n \to \infty$. Для того, чтобы обеспечить при вычислении, например, a_m-го (максимального по модулю) члена ряда абсолютную погрешность, не превосходящую ε, необходимо вычислить его с относительной погрешностью, не хуже чем

$$\delta(a_m) \le \frac{\Delta a_m}{|a_m|} \le \frac{\varepsilon}{|a_m|}.$$

Требуемая относительная точность тем выше, чем больше $|a_m|$, что можно обеспечить только увеличением длины мантиссы.

Величина погрешности округления зависит также от того, как алгоритмически реализован приближенный метод. Например, частичную сумму ряда для функции $\sin t$ можно подсчитывать, суммируя члены ряда в их естественном порядке; можно суммировать в обратном порядке (с конца); можно рассматривать отрезок ряда как полином и использовать для его вычисления *схему Горнера*; можно просуммировать отдельно положительные и отрицательные члены ряда и затем вычесть из первой суммы вторую и т. д. (из перечисленных алгоритмов последний наиболее чувствителен к ошибкам округления).

Схемой Горнера называют запись полинома n-й степени в следующем виде

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

Формальное использование этой схемы для вычисления значений отрезка ряда без учета специфики вычислений на ЭВМ приведет к неверным результатам уже при сравнительно небольших n. Сохранив идею, необходимо внести в нее соответствующие коррективы.

1.4. Вычисление производной

Пусть задана функция $f(x)$. Необходимо вычислить ее первую производную в некоторой точке x. Воспользуемся для этого формулами численного дифференцирования различного порядка аппроксимации.

1.5. Формула первого порядка аппроксимации

$$f^{(\mathrm{I})}(x) \approx \frac{f(x+h) - f(x)}{h}. \tag{1.7}$$

Пусть известно, что $|f^{(\mathrm{II})}(\xi)| \le M_2$; тогда погрешность метода для этой формулы имеет первый порядок по h:

$$|r_1| = \left| f^{(\mathrm{I})}(x) - \frac{f(x+h) - f(x)}{h} \right| \le \frac{M_2\ h}{2}. \tag{1.8}$$

Пусть значения функции $f(x)$ известны с погрешностью $\varepsilon(x)$, $|\varepsilon(x)| \le \mathrm{E}$. Даже в случае отсутствия неустранимой погрешности f, при вычислении значения функции на ЭВМ возникает погрешность за счет ошибок округления, и ее величина в этом случае зависит от представления чисел в машине. Тогда при вычислении производной по формуле (1.7) возникает погрешность r_2, причем

$$|r_2| \le \frac{2\mathrm{E}}{h}. \tag{1.9}$$

Для суммарной погрешности r имеем оценку

$$|r| \le |r_1| + |r_2| \le g(h) = \frac{M_2\ h}{2} + \frac{2\ \mathrm{E}}{h}. \tag{1.10}$$

Для уменьшения погрешности метода необходимо, со-

гласно оценке (1.8), уменьшить шаг h, но при этом растет второе слагаемое в (1.10).

На рис. 2 представлен характер зависимости погрешности метода, погрешности вычисления функции и суммарной погрешности в зависимости от шага h. Минимум суммарной погрешности достигается в точке h^* экстремума функции $q(h)$: $q'(h) = 0,$ причем в ней $r_1 = r_2$.

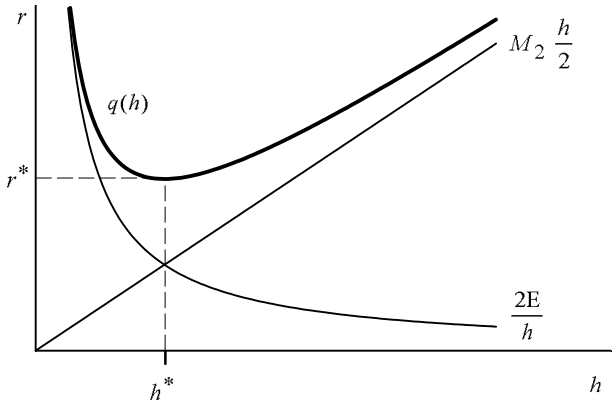


Рис. 2

Тогда имеем для оптимального шага дифференцирования:

$$\frac{dq(h)}{dh} = 0, \qquad h^* = 2\sqrt{\frac{\mathrm{E}}{M_2}}. \tag{1.11}$$

При использовании формулы (1.7) нельзя рассчитывать на точность более высокую, чем

$$r^* = \sqrt{\mathrm{E}\, M_2}, \tag{1.12}$$

которая является следствием (1.10) при $h = h^*$.

Если погрешность при вычислении функции связана лишь с ошибками округления, то в этом случае $\mathrm{E} \cong 2^{-t} |f|,$ где t — число разрядов, отводимых под хранение мантиссы числа. Следовательно, производную можно вычислить, в лучшем случае, с половиной верных знаков (если M_2 и $|f| \cong 1$).

Рассмотрим теперь как изменятся результаты в случае ис-

пользования формулы численного дифференцирования второго порядка аппроксимации.

1.6. Формула второго порядка аппроксимации

$$f^{(\mathrm{I})}(x) \approx \frac{f(x+h) - f(x-h)}{2h}. \tag{1.13}$$

Пусть известно, что $|f^{(\mathrm{III})}(\xi)| < M_3$; тогда погрешность метода для этой формулы имеет второй порядок по h:

$$|r_1| = \left| f^{(\mathrm{I})}(x) - \frac{f(x+h) - f(x-h)}{2h} \right| \le \frac{M_3 h^2}{6}. \tag{1.14}$$

Пусть значения функции $f(x)$ известны с погрешностью $\varepsilon(x)$, $|\varepsilon(x)| \le \mathrm{E}$. Тогда при вычислении производной по формуле (1.13) возникает погрешность $|r_2|$, причем

$$|r_2| \le \frac{\mathrm{E}}{h}. \tag{1.15}$$

Для суммарной погрешности r имеем оценку:

$$|r| = |r_1| + |r_2| \le q(h) = \frac{M_3 h^2}{6} + \frac{\mathrm{E}}{h}. \tag{1.16}$$

Для уменьшения погрешности метода необходимо, согласно оценке (1.14), уменьшить шаг h, но при этом растет второе

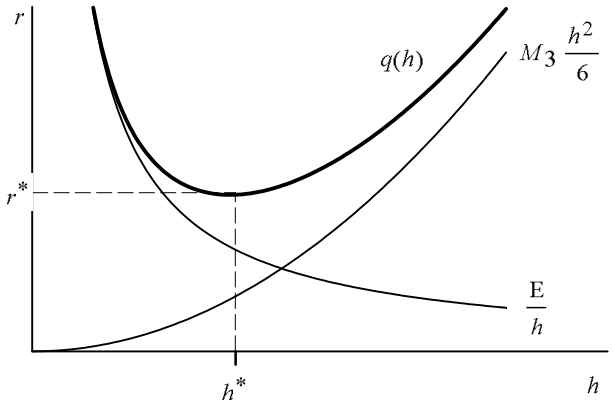


Рис. 3

слагаемое в (1.16). На рис. 3 представлен характер зависимости погрешности метода, погрешности вычисления функции и суммарной погрешности в зависимости от шага h. Минимум погрешности достигается в точке h — экстремума функции $q(h)$: $q'(h) = 0$. Оптимальное значение шага численного дифференцирования есть:

$$h^* = \sqrt[3]{\frac{3\,\mathrm{E}}{M_3}}. \tag{1.17}$$

Таким образом, при использовании формулы (1.13) нельзя рассчитывать на точность более высокую, чем

$$r^* = \sqrt[3]{\frac{9\,\mathrm{E}^2\,M_3}{8}}. \tag{1.18}$$

Ниже рассматривается формула четвертого порядка аппроксимации.

1.7. Формула четвертого порядка аппроксимации

$$f^{(\mathrm{I})} \approx \frac{f(x-2h) - 8f(x-h) + 8f(x+h) - f(x+2h)}{12h}. \tag{1.19}$$

Пусть известно, что $|f^{(5)}(\xi)| \le M_5$; тогда погрешность метода для этой формулы имеет четвертый порядок по h:

$$|r_1| = \left| f^{(\mathrm{I})} - \frac{f(x-2h) - 8f(x-h) + 8f(x+2h) - f(x+2h)}{12h} \right| \le \frac{M_5\,h^4}{30}. \tag{1.20}$$

Пусть значения функции $f(x)$ известны с погрешностью $\varepsilon(x)$, $|\varepsilon(x)| \le \mathrm{E}$. Тогда при вычислении производной по формуле (1.19) возникает погрешность $|r_2|$, причем

$$|r_2| \le \frac{3\,\mathrm{E}}{2h}. \tag{1.21}$$

Для суммарной погрешности r имеем оценку

$$|r| \le |r_1| + |r_2| \le q(h) = \frac{M_5\,h^4}{30} + \frac{3\,\mathrm{E}}{2h}. \tag{1.22}$$

Для уменьшения погрешности метода необходимо, согласно оценке (1.20), уменьшить шаг h, но при этом растет второе слагаемое в (1.22).

На рис. 4 представлен характер зависимости погрешности метода, погрешности вычислений и суммарной погрешности в зависимости от шага h.

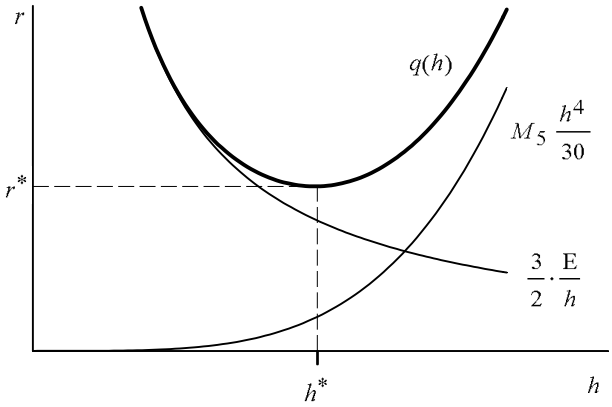


Рис. 4

Минимум погрешности достигается в точке h^* экстремума функции $q(h)$: $q'(h) = 0$. Имеем для оптимального шага численного дифференцирования

$$h^* = \sqrt[5]{\frac{45\,\mathrm{E}}{4M_5}}. \qquad (1.23)$$

Таким образом, при использовании формулы (1.19) нельзя рассчитывать на точность более высокую, чем

$$r^* = \frac{15}{8}\sqrt[5]{\frac{4\,\mathrm{E}^4\,M_5}{15}}.$$

1.8. Контрольные вопросы

1. Как известно, для вычисления функции $\ln x$ можно использовать следующий ряд по x:

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots + (-1)^{k+1}\frac{x^k}{k} + \ldots \qquad \text{(а)}$$

Можно представить $1+x$ в виде $1+x = 2^m \cdot z$, где $z \in [0{,}5;\ 1]$, положив далее

$$y = \frac{1-z}{1+z},$$

для представления логарифма получаем ряд

$$\ln x = m \ln 2 - 2\left(y + \frac{y^3}{3} + \ldots + \frac{y^{2k-1}}{2k-1} + \ldots \right). \qquad \text{(б)}$$

В чем преимущества и недостатки использования ряда (б)? Как оценить погрешность метода при использовании каждого из этих разложений?

2. Какова относительная погрешность округления при представлении действительного числа в ЭВМ, если под хранение мантиссы отводится p бит? (Ответ: 2^{-p}.)

Указание: Рассмотрите представление произвольного действительного числа в виде бесконечной двоичной дроби:

$$a = \operatorname{sign} a \cdot 2^q \cdot \left(\frac{a_1}{2} + \frac{a_2}{2^2} + \ldots + \frac{a_p}{2^p} + \frac{a_{p+1}}{2^{p+1}} + \ldots \right),$$

где a_l равно 0 или 1, и соответствующее ему округленное представление:

$$a = \operatorname{sign} a \cdot 2^q \cdot \left(\frac{a_1}{2} + \frac{a_2}{2^2} + \ldots + \frac{a_p}{2^p} \right).$$

3. Пусть функция $f(x)$ задана таблично: заданы значения аргументов $x_0 < x_1 < x_2 < \ldots < x_N$ (расстояние между двумя соседними точками h) и значения функции в них $f_0,\ f_1,\ \ldots\ f_N$.

Самостоятельно выведите формулу вычисления *односторонней производной* для приближенного вычисления $f'(x)$ в точках x_0 и x_N с точностью до $O(h^2)$ и $O(h^3)$. Найдите оп-

тимальные шаги численного дифференцирования. Сравните их с оценками для центральных разностей.

Указание: Для вывода формул используйте *метод неопределенных коэффициентов*, а именно, равенство

$$f'(x_0) \approx \frac{\alpha_0 f(x_0) + \alpha_1 f(x_1) + \alpha_2 f(x_2)}{h}.$$

Подберите α_0, α_1 и α_2 так, чтобы равенство выполнялось с точностью до $O(h^2)$.

4. Вторая и третья производные функции вычисляются по приближенным формулам

$$f''(x) = \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}$$

и

$$f'''(x) = \frac{f(x+2h) - 2f(x+h) + 2f(x-h) - f(x-2h)}{2h^3}.$$

Найдите погрешность метода и неустранимую погрешность при вычислениях по этим формулам. Найдите оптимальные шаги численного дифференцирования и минимально возможную ошибку.

1.9. Порядок выполнения работы

Начните выполнение работы с вычисления функции $y = \sin t$ с помощью ряда Тейлора (пункт меню «*Методы*» к разделу «*Ряды*»). Задайте максимальную длину мантиссы (K = 52), обычно используемую при работе с переменными типа double; задайте начальный интервал изменения аргумента $t \in [0, 10]$. Последовательно увеличивая число членов ряда, привлекаемых для вычисления суммы, визуально убедитесь в том, что для фиксированного *t* точность метода растет с ростом *n*. Отметьте, что чем больше *t*, тем большее число членов ряда необходимо привлекать для обеспечения необходимой точности.

Отодвиньте правую границу интервала *t* вправо настолько, чтобы наблюдаемое отклонение от точного значения функции $\sin t$ нельзя было устранить увеличением точности метода. Уста-

новите режим «*Вычисление ряда с заданной точностью* ε» и убедитесь, что результат будет тем же. Попытайтесь объяснить наблюдаемое явление. В случае затруднений обратитесь к разделу меню «*Учебник*». Меняя длину мантиссы (число К), убедитесь, что наблюдаемый эффект возникает тем раньше, чем меньше это число.

Установите режим «*Вычисление ряда с заданной точностью* ε» и задайте точность очень грубую, например $\varepsilon = 3$. Объясните наблюдаемую картину.

Задайте интервал для $t \in [0, 20]$, длину мантиссы K = 20, число членов ряда $n = 24$. Оцените, какой вклад в наблюдаемую погрешность вносит ошибка метода и какая ошибка возникает из-за мантиссы конечной длины. Задайте режим «*Вычисление ряда с заданной точностью* ε» (в этом режиме программа сама выбирает минимально необходимое число членов ряда для вычисления значения функции при каждом t с заданной точностью) и задайте достаточно высокую точность. Объясните, почему наблюдаемая погрешность намного превышает заданную точность.

Используя информацию с экрана, оцените величину максимального по модулю члена ряда для $t = 20$, приняв в качестве гипотезы, что наблюдаемая погрешность возникла при вычислении только одного максимального члена ряда.

Найдите номер m максимального по модулю члена ряда. Для этого можно воспользоваться связью между a_{n+1}-м и a_n-м членами ряда:

$$|a_{n+1}| = |a_n| \frac{t^2}{2n\,(2n+1)}.$$

Найдя m, оцените $|a_m|$. Сравните полученный результат с ранее найденным значением $|a_m|$.

Установите $\varepsilon = 10^{-5}$. Почему в этом случае погрешность носит пилообразный характер?

Перейдите к пункту меню «*Методы–Дифференцирование*». Установите длину мантиссы (начните с максимальной). Задайте начальный шаг h для вычисления производной. Выберите формулу первого порядка аппроксимации и функцию, для которой будет вычисляться производная. Последовательно уменьшая шаг h, проследите, как ведет себя погрешность метода и по-

грешность, связанная с использованием конечной арифметики. Уменьшайте шаг до тех пор, пока на графике производной не появятся аномальные эффекты; посмотрите, что произойдет при дальнейшем уменьшении шага. Постарайтесь объяснить наблюдаемые явления. Изменяя длину мантиссы, исследуйте, какое влияние оказывает на них эта характеристика ЭВМ.

Установив режим «*Выбор точки*», получите на экране зависимость погрешности вычисления производной от шага, сравните ее с теоретической. С помощью инструментария программы Вы можете вывести на экран отдельные фрагменты этой зависимости в более крупном масштабе (пункт меню «*Запуск–Масштабирование*»). Определите, при каком шаге h погрешность минимальна. Установите, какова эта погрешность. Как влияет на эти величины длина мантиссы? Сравните результаты с теоретическими оценками.

Изменение масштабов в окне, где отображена погрешность, осуществляется либо с помощью мыши, либо с использованием клавиатуры. При этом рамка лупы управляется кнопками *Home*, *End*, *PageUp*, *PageDown*. Перемещается рамка лупы с помощью стрелок.

Проделайте это же задание, используя формулы численного дифференцирования второго и четвертого порядка точности. Сравните все три метода по оптимальному шагу, при котором достигается минимум погрешности, по величине этой погрешности. Сравните с теоретическими оценками.

1.10. Библиографическая справка

Подробнее элементарная теория погрешностей рассмотрена в [1, 32]. Некоторые аспекты вычислений с конечной арифметикой можно найти в [5]. Анализ влияния конечноразрядной арифметики на результаты вычислений в задачах линейной алгебры проведен в [6].

ЛАБОРАТОРНАЯ РАБОТА 2

ТАБЛИЧНОЕ ЗАДАНИЕ И ИНТЕРПОЛИРОВАНИЕ ФУНКЦИЙ

2.1. Введение

Работа позволяет изучить основные свойства процесса интерполяции функций, заданных таблицей. Для функций с различными дифференциальными свойствами иллюстрируются особенности процесса глобальной алгебраической и тригонометрической интерполяции. Изучается погрешность алгебраической и тригонометрической интерполяции, сходимость, устойчивость и насыщаемость гладкостью интерполяционного процесса на различных системах узлов интерполяции: равноотстоящие узлы, корни полиномов Чебышева, экстремумы полиномов Чебышева. Рассматривается кусочно-многочленная гладкая интерполяция двух типов — локальные и нелокальные сплайны, а также негладкая кусочно-линейная, кусочно-квадратичная, и кусочно-кубическая интерполяция.

2.2. Задача интерполяции

Задача интерполяции состоит в нахождении обобщенного многочлена

$$P_n(x) = \sum_{k=0}^{n} c_k \varphi_k(x), \tag{2.1}$$

где $\varphi_k(x)$ — фиксированные функции, а значения коэффициентов определяются из условия равенства со значением приближаемой функции в узлах интерполяции

$$P_n(x_k) = f_k, \qquad k = 0, 1, \ldots, n. \tag{2.2}$$

Набор точек x_j на интервале $[a, b]$, $a \le x_0 < x_1 < \ldots$ $\ldots < x_n \le b$, в которых заданы значения функции $f(x_j)$, назы-

вают *сеткой*. Множество точек x_j иногда также называют *узлами* сетки или *узлами интерполяции*.

Мы будем называть сетку *равномерной*, если

$$x_{j+1} - x_j = \text{const}, \qquad j = 0, 1, \ldots, n-1;$$

$$a = x_0, \qquad b = x_n.$$

Если $\varphi_k(x) = x^k$, то соответствующая интерполяция называется *алгебраической*, если φ_k — тригонометрические функции, то говорят о *тригонометрической интерполяции.*

Если построенный многочлен (2.2) используется для восстановления функции на всем отрезке [*a*, *b*], то говорят о *глобальной* интерполяции. Если же для восстановления функции между *каждыми двумя* соседними узлами строится многочлен заданной невысокой степени, то говорят о *кусочно-многочленной* интерполяции.

Если значения функции *f*(*x*) заданы в узлах x_j на интервале [*a*, *b*], $a \le x_0 < x_1 < \ldots < x_n \le b$, то говорят, что функция *f*(*x*) задана *таблицей*.

2.3. Алгебраическая интерполяция

Теорема 1. *Пусть задан* $n+1$ *узел* x_0, x_1, …, x_n, *среди которых нет совпадающих, и значения функции в этих узлах* $f(x_0)$, $f(x_1)$, …, $f(x_n)$. *Тогда существует один и только один многочлен* $P_n(x) = P_n(x, f, x_0, x_1, \ldots, x_n)$ *степени не выше* n, *принимающий в узлах* x_k *заданные значения* $f(x_k)$.

Интерполяционный многочлен можно записать (и соответственно вычислить) различными способами представляя его в виде *разложения по степеням x* (в *форме Лагранжа* и в *форме Ньютона*), или в виде *разложения по ортогональным многочленам*.

2.4. Непосредственное вычисление коэффициентов интерполяционного полинома

Полином степени *n* можно записать в виде

$$P_n(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n, \tag{2.3}$$

где a_0, …, a_n — неопределенные коэффициенты. Их можно определить из n+1 условия:

$$\begin{aligned} &a_0 + a_1 x_0 + a_2 x_0^2 + \ldots + a_n x_0^n = f(x_0), \\ &a_0 + a_1 x_1 + a_2 x_1^2 + \ldots + a_n x_1^n = f(x_1), \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &a_0 + a_1 x_n + a_2 x_n^2 + \ldots + a_n x_n^n = f(x_n). \end{aligned} \tag{2.4}$$

Определитель системы (2.4) есть детерминант Вандермонда, известный из курса линейной алгебры. Его значение в случае, когда выполняются условия теоремы 1, отлично от нуля, что доказывает существование и единственность полинома. Эта линейная система во многих случаях является *плохо обусловленной*. Последнее связано с тем, что последовательные степени 1, x, x^2, …, x^n «почти линейно зависимы» на интервале $0 < x < 1$. Напомним, что *обусловленность* линейной системы $\mathbf{Ay} = \mathbf{b}$ определяется числом

$$\mu = \| \mathbf{A} \| \cdot \| \mathbf{A}^{-1} \|, \tag{2.5}$$

которое определяет относительную погрешность решения системы в зависимости от относительной погрешности правой части **b**:

$$\frac{\| \boldsymbol{\delta}\mathbf{y} \|}{\| \mathbf{y} \|} \le \mu \frac{\| \boldsymbol{\delta}\mathbf{b} \|}{\| \mathbf{b} \|}. \tag{2.6}$$

Матрицу **A** будем называть *сингулярной*, если в рамках системы вычислений с плавающей точкой на данной машине выполняется равенство $\mu = \mu + 1$.

2.5. Интерполяционный полином в форме Лагранжа. Интерполяционный многочлен в форме Ньютона

Введем вспомогательные многочлены

$$l_k = \frac{(x - x_0) \ldots (x - x_{k-1})(x - x_{k+1}) \ldots (x - x_n)}{(x_k - x_0) \ldots (x_k - x_{k-1})(x_k - x_{k+1}) \ldots (x_k - x_n)}. \tag{2.7}$$

Многочлен $P_n(x)$, заданный равенством

$$P_n(x) = P_n(x, f, x_0, x_1, \ldots, x_n) = f(x_0)\, l_0(x) +$$

$$+f(x_1)\,l_1(x)+\ldots+f(x_n)\,l_n(x), \qquad (2.8)$$

есть интерполяционный многочлен в *форме Лагранжа*.

Употребляются и другие виды записи интерполяционного многочлена. Часто используется запись *в форме Ньютона*.

Определим *разностные отношения* (иногда употребляется термин «*разделенные разности*»). Пусть функция $f(x)$ в точках x_a, x_b, x_c, x_d и т. д. принимает значения $f(x_a)$, $f(x_b)$, $f(x_c)$, $f(x_d)$.

Разностные отношения нулевого порядка $f(x_k)$ функции $f(x)$ в точке x_k определяют, как значение функции в этой точке $f(x_k)=f(x_k)$, $k=a,b,c,d\ldots$ Разностные отношения первого порядка $f(x_k,x_t)$ функции $f(x)$ для произвольной пары точек x_k и x_t определим через разностные отношения нулевого порядка:

$$f(x_k,x_t)=\frac{f(x_t)-f(x_k)}{x_t-x_k}. \qquad (2.9)$$

Разностное отношение $f(x_0,x_1,\ldots,x_n)$ порядка n определим через разностное отношение порядка $n-1$, положив:

$$f(x_0,x_1,\ldots,x_n)=\frac{f(x_1,\ldots,x_n)-f(x_0,\ldots,x_{n-1})}{x_n-x_0}. \qquad (2.10)$$

Интерполяционный многочлен в *форме Ньютона* $P_n(x,f,x_0,x_1,\ldots,x_n)$ с использованием введенных разностных отношений может быть записан как:

$$\begin{aligned}P_n(x,f,x_0,x_1,\ldots,x_n)&=f(x_0)+f(x_0,x_1)(x-x_0)+\\&+f(x_0,x_1,x_2)(x-x_0)(x-x_1)+\ldots+\\&+f(x_0,x_1,\ldots,x_n)(x-x_0)\ldots(x-x_{n-1}).\end{aligned} \qquad (2.11)$$

Если $x_0<x_1<\ldots<x_n$, то соответствующую интерполяцию называют *интерполяцией вперед*; в случае $x_0>x_1>\ldots>x_n$ интерполяцию называют *интерполяцией назад*.

2.6. Формула для погрешности алгебраической интерполяции

Оценим погрешность $R_s(x) = f(x) - P_s(x, f), \quad x_k < x < x_{k+1},$ возникающую при приближенной замене $f(x)$ алгебраическим многочленом $P_s(x, f)$. В основе оценки лежит следующая общая теорема о формуле погрешности.

Теорема 2. *Пусть* $f(t)$ *— функция, определенная на некотором отрезке* $\alpha < t < \beta$ *и имеющая производные до некоторого порядка* $s + 1$ *включительно.*

Пусть t_0, t_1, …, t_s *— произвольный набор попарно различных точек из отрезка* $[\alpha, \beta]$; $f(t_0)$, $f(t_1)$, …, $f(t_s)$ *— значения функции* $f(t)$ *в этих точках*; $P_s(t)$ *— интерполяционный многочлен степени не выше s, построенный по этим значениям. Тогда погрешность интерполяции* $R_s(t) = f(t) - P_s(t)$ *представляется формулой*:

$$R_s(t) = \frac{f^{(s+1)}(z)}{(s+1)!}(t - t_0)(t - t_1)\dots(t - t_s), \tag{2.12}$$

где $z = z(t)$ *— некоторая точка интервала* $[\alpha, \beta]$.

2.7. О сходимости интерполяционного процесса

На отрезке $a \le x \le b$ будем рассматривать бесконечную последовательность узлов интерполяции

$$\begin{array}{l} x_1^1 \\ x_1^2, x_2^2 \\ x_1^3, x_2^3, x_3^3 \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ x_1^n, x_2^n, x_3^n, \ldots, x_n^n \\ \ldots\ldots\ldots\ldots\ldots\ldots \end{array} \tag{2.13}$$

и соответствующую последовательность интерполяционных многочленов $P_n(x, f)$, построенную для некоторой функции $f(x)$, принимающей конечное значение.

Теорема 3. (Фабера). *Какова бы ни была последователь-*

ность узлов интерполяции, существует непрерывная функция f, для которой последовательность интерполяционных многочленов расходится.

Теорема 4. *Для каждой функции f, непрерывной на конечном отрезке, существует такая последовательность узлов интерполяции, что соответствующий ей интерполяционный процесс равномерно сходится к f.*

Теорема 5. *Не существует последовательности узлов, для которой интерполяционный процесс был бы равномерно сходящимся для всякой непрерывной на отрезке функции.*

Теорема 6. *Если функция f имеет ограниченную производную на отрезке, то интерполяционный процесс, в котором за узлы принимаются корни многочленов Чебышёва, сходится равномерно к f.*

Основные термины. Интерполирующую функцию иногда называют *интерполянтом.*

Гладкий кусочно-многочленный интерполянт называется *сплайном*.

2.8. Обусловленность задачи построения интерполяционного многочлена для функции, заданной таблицей

Интерполяционный многочлен

$$P_n(x) = \sum_{k=0}^{n} c_k \varphi_k(x), \tag{2.14}$$

где $\varphi_k(x)$ — фиксированные функции, а значения коэффициентов c_k определяются из условия совпадения со значениями приближаемой функции в узлах интерполяции, можно записать в виде

$$P_n(x) = \sum_{k=0}^{n} f_k l_k(x), \tag{2.15}$$

для $l_k(x_i) = 0, \quad k \neq i; \quad l_k(x_k) = 1; \quad k, i = 0, 1, \ldots, n.$ $l_k(x)$ иногда называют *фундаментальными полиномами*.

Придадим значениям функции $f(x_j)$ возмущения

$\delta f(x_j)$. Интерполяционный многочлен $P_n(x, f)$ заменится многочленом $P_n(x, f + \delta f)$.

Так как $P_n(x, f + \delta f) = P_n(x, f) + P_n(x, \delta f)$ в силу линейности (2.15) по f, то возмущение $P_n(x, \delta f)$, которое претерпевает интерполяционный многочлен, можно оценить как:

$$| P_n(x, \delta f) | \le \max | \delta f(x_j) | \sum_{k=0}^{n} | l_k(x) |. \qquad (2.16)$$

Это возмущение при заданных узлах интерполяции и фиксированных базисных функциях $\varphi_k(x)$ зависит только от δf.

Введем в рассмотрение функцию $L_n(x) = \sum_{k=0}^{n} | l_k(x) |$, которая называется *функцией Лебега*.

За меру чувствительности интерполяционного многочлена к возмущениям задания функции в узлах δf принимается наименьшее число L_n, при котором для каждого δf выполнено неравенство

$$\max_{a<x<b} | P_n(x, \delta f) | \le \mathrm{L}_n \max_j | \delta f(x_j) |. \qquad (2.17)$$

Очевидно, что $\mathrm{L}_n = \max_{a<x<b} L_n(x)$.

Числа $\mathrm{L}_n = \mathrm{L}_n(x_0, x_1, \ldots, x_n, a, b), \quad n = 0, 1, \ldots$ называют *константами Лебега*. Эти числа растут с ростом n. Их поведение при возрастании n существенно зависит от отрезка $[a, b]$ и от расположения узлов интерполяции на этом отрезке.

Для алгебраической интерполяции $(\varphi_k(x) = x^k)$ в случае равномерно расположенных узлов

$$\mathrm{L}_n > \mathrm{const} \frac{2^n}{\sqrt{n}},$$

т. е. чувствительность результата интерполяции к погрешностям задания функции в узлах резко возрастает с ростом n. Такие погрешности неизбежны как при получении табличных значений в результате измерений, так и в результате округлений.

Если узлами интерполяции являются корни полинома Чебышёва или точки, где этот полином достигает экстремумов, то

$L_n \le \text{const} \cdot \ln n$, т. е. с ростом n константы Лебега растут очень медленно. В этом случае вычислительная неустойчивость не является препятствием для использования интерполяционных многочленов высокой степени.

2.9. Классическая кусочно-многочленная интерполяция

Пусть функция $f(x)$ задана таблицей. Для восстановления функции между узлами можно воспользоваться функцией, которая между каждыми двумя соседними узлами является многочленом заданной невысокой степени, например, первой, второй, третьей и т. д.

Соответствующая интерполяция называется *кусочно-линейной*, *кусочно-квадратичной* и т. п.

2.10. Оценка неустранимой погрешности при приближении функции по ее значениям в узлах интерполяции. Выбор степени кусочно-многочленной интерполяции

Пусть функция $f(x)$ определена на отрезке $[0, \pi]$ и пусть заданы ее значения в узлах равномерной сетки $x_k = k\pi / n$, $k = 0, 1, \ldots, n$. По таблице $f(x_0)$, $f(x_1)$, …, $f(x_n)$, в принципе, нельзя восстановить функцию $f(x)$ точно, потому что значения различных функций могут совпадать в точках x_k, $k = 1, \ldots, n$, т. е. различные функции могут иметь одинаковую таблицу.

Если, о функции известно лишь то, что она непрерывна, то ее нельзя восстановить в точке $x \ne x_k$, $k = 0, 1, \ldots, n$, ни с какой гарантированной точностью.

Пусть о функции $f(x)$ известно, что она имеет производные порядка $s + 1$, причем

$$\max_x | f^{(s+1)}(x) | \le M_s = \text{const}. \qquad (2.18)$$

Укажем две функции

$$f_{(\text{I})}(x) = \frac{\sin nx}{n^{s+1}}, \qquad f_{(\text{II})}(x) = -\frac{\sin nx}{n^{s+1}}, \qquad (2.19)$$

для которых таблицы $f_{(\mathrm{I})}(x_k) = f_{(\mathrm{II})}(x_k)$, $k = 0, 1, \ldots, n$, совпадают (обе таблицы содержат лишь нули) Эти функции уклоняются друг от друга на величину порядка h^{s+1}:

$$\max_x |f_{(\mathrm{I})}(x) - f_{(\mathrm{II})}(x)| = 2 \max_x \left| \frac{\sin nx}{n^{s+1}} \right| = 2h^{s+1}. \qquad (2.20)$$

Таким образом, зная лишь оценку $s + 1$ производной, в принципе нельзя восстановить табличную функцию с точностью, большей, чем $\mathrm{O}(h^{s+1})$. Данная погрешность *неустранима*.

2.11. Насыщаемость (гладкостью) кусочно-многочленной интерполяции

Пусть функция $f(x)$ определена на отрезке $[a, b]$, и задана ее таблица $f(x_k)$ в равноотстоящих узлах x_k, $k = 0, 1, \ldots, n$; с шагом $h = (b - a)/n$.

Погрешность кусочно-многочленной интерполяции степени s (с помощью интерполяционных многочленов $P_s(x, f_{kj})$ на отрезке $x_k \le x \le x_{k+1}$) в случае, если на $[a, b]$ существует и ограничена $f^{(s+1)}(x)$, имеет порядок $\mathrm{O}(h^{s+1})$.

Если о функции $f(x)$ известно лишь, что она имеет ограниченную производную до некоторого порядка q, $q < s$, то неустранимая погрешность при ее восстановлении по таблице есть $\mathrm{O}(h^{q+1})$. Можно показать, что при интерполяции с помощью $P_s(x, f_{kj})$ порядок $\mathrm{O}(h^{q+1})$ достигается.

Если $f(x)$ имеет ограниченную производную порядка $q + 1$, $q > s$, то погрешность интерполяции с помощью $P_s(x, f_{kj})$ остается $\mathrm{O}(h^{s+1})$, т. е. порядок погрешности не реагирует на дополнительную, сверх $s + 1$ производной, гладкость функции $f(x)$. Это свойство кусочно-многочленной интерполяции называют свойством насыщаемости (гладкостью).

2.12. Кусочно-многочленная гладкая интерполяция (сплайны). Локальные сплайны

Классическая кусочно-линейная, кусочно-квадратичная и, вообще, кусочно-многочленная интерполяция заданной степени приводят к интерполирующей функции (интерполянту), которая в узлах интерполяции, вообще говоря, не имеет производной даже первого порядка. Существует два типа гладких кусочно-многочленных интерполянтов — *локальные* и *нелокальные сплайны*. Они обладают заданным числом производных всюду, включая узлы интерполяции.

Рассмотрим *локальные сплайны*. Пусть заданы узлы интерполяции x_l : $x_l < x_{l+1}$ и значения функции $f(x_l)$ в них. Зададим натуральное число s, фиксируем натуральное число j: $j \le s$. Каждой точке x_l сопоставим интерполяционный многочлен $P_s(x, f)$, построенный по значениям $f(x_{l-j})$, $f(x_{l-j+1})$, …, $f(x_{l-j+s})$ в узлах x_{l-j}, x_{l-j+1}, …, x_{l-j+s}. Кусочно-многочленную интерполирующую функцию $\varphi(x, s)$, имеющую непрерывные производные порядка s, определим равенствами

$$\varphi(x, s) = Q_{2s+1}(x, k),$$
$$x_k \le x \le x_{k+1}, \qquad k = 0, \pm 1, \pm 2, \ldots \tag{2.21}$$

где $Q_{2s+1}(x, k)$ — многочлен степени не выше $2s + 1$, определяемый равенствами

$$\frac{d^m Q_{2s+1}(x, k)}{dx^m} = \begin{cases} \dfrac{d^m P_s(x, f)}{dx^m}, & \text{при } x = x_k, \quad m = 0, 1, \ldots, s; \\ \dfrac{d^m P_s(x, f)}{dx^m}, & \text{при } x = x_{k+1}, \ m = 0, 1, \ldots, s. \end{cases} \tag{2.22}$$

Верны следующие теоремы.

Теорема 7. *Существует один и только один многочлен степени не выше* $2s + 1$, *удовлетворяющий* (2.22).

Теорема 8. *Пусть* $f(x)$ *— многочлен степени не выше s. Тогда интерполянт* $\varphi(x, s)$ *совпадает с этим многочленом.*

Теорема 9. *Кусочно-многочленная интерполирующая функция* $\varphi(x, s)$, *определяемая равенством* (2.22), *в узлах интерполяции* x_l *совпадает с заданным в них значением* $f(x_l)$, $l = 0, 1, \ldots$

Кроме того, $\varphi(x, s)$ *имеет всюду в области своего определения непрерывную производную порядка* s.

Многочлен $Q_{2s+1}(x, k)$ можно записать в виде

$$Q_{2s+1}(x, k) = P_s(x, f) + R_{2s+1}(x, k), \qquad (2.23)$$

обозначив через $R_{2s+1}(x, k)$ поправку к классическому интерполяционному многочлену $P_s(x, f)$.

Рассмотрим здесь наиболее интересный для приложений случай, когда $s = 2$, $j = 1$, а узлы интерполяции функции составляют равномерную сетку. Тогда

$$R_5(x, k) = \frac{h^3}{2!}\,\frac{f(x_{k+2}) - 3f(x_{k+1}) + 3f(x_k) - f(x_{k-1})}{h^3} \cdot \frac{(x - x_k)^3}{h^3} \times$$
$$\times \frac{x - x_{k+1}}{h} \cdot \left(3 - \frac{2(x - x_k)}{h}\right). \qquad (2.24)$$

Эта формула справедлива только при $0 < k < n$. На отрезках $a =$
$= x_0 \le x \le x_1$ и $x_{n-1} \le x \le x_n = b$ поправка: $R_5(x, k) \equiv 0$.

2.13. Нелокальная гладкая кусочно-многочленная интерполяция

Пусть задана таблица. Поставим задачу найти на каждом отрезке $x_k \le x \le x_{k+1}$ кубический многочлен $P_3(x, k)$ так, чтобы возникающая при этом на отрезке $a \le x \le b$ кусочно-многочленная функция совпадала с заданной функцией в узлах и имела непрерывные производные до порядка $s = 2$. Общее число неизвестных — $4n$. Число дополнительных условий равно $4n - 2$. Недостающие условия можно задавать различными способами. Наиболее употребляемыми являются следующие два:

$$\frac{d^2P_3(x, 0)}{dx^2} = \frac{d^2P_3(x, n)}{dx^2} = 0 \text{ — «свободный сплайн»;} \qquad (2.25)$$

$$\frac{d^3P_3(x, 0)}{dx^3} = \frac{d^3c_0}{dx^3}, \qquad \frac{d^3P_3(x, n)}{dx^3} = \frac{d^3c_n}{dx^3}. \qquad (2.26)$$

Здесь $c_0(x)$, $c_n(x)$ — единственные кубические кривые, которые проходят соответственно через четыре первые и четыре по-

следние из заданных точек.

Построенная таким образом функция называется *кубическим сплайном Шонберга.* Если интерполируемая функция имеет ограниченную производную третьего порядка, то непрерывный с производными до второго порядка сплайн Шонберга сохраняет неулучшаемые аппроксимационные свойства классической, а также локальной гладкой кусочно-многочленной интерполяции.

Однако сплайны Шонберга теряют свойства локальности, присущие как классической кусочно-многочленной интерполяции, так и локальной гладкой интерполяции: коэффициенты многочлена, задающие интерполянт на каком-либо отрезке $x_k \le x \le x_{k+1}$, зависят от значений функции во всех узлах сетки.

2.14. Тригонометрическая интерполяция

Задача (линейной) тригонометрической интерполяции состоит в нахождении тригонометрического многочлена вида

$$Q_n\left(\cos\frac{2\pi\,(x-x_0)}{L},\ \sin\frac{2\pi\,(x-x_0)}{L}\right)=$$

$$=\sum_{k=0}^{n} a_k\cos\frac{2\pi k\,(x-x_0)}{L}+\sum_{k=1}^{n} b_k\sin\frac{2\pi k\,(x-x_0)}{L}. \qquad (2.27)$$

Здесь k и n — натуральные числа, $L = x_N - x_0$ — положительное число, $[x_0, x_N]$ — отрезок интерполяции, a_k и b_k — числовые коэффициенты.

Теорема 10. (Первый вариант задания узлов интерполяции). *Пусть* $N = 2(n+1)$, *n — натуральное число. При произвольном задании значений функции* f_m, *периодической с периодом L, в узлах сетки*

$$x_m = x_0 + \frac{Lm}{N} + \frac{L}{2N}, \qquad m = 0, 1, \ldots, N-1$$

существует один и только один интерполяционный тригонометрический многочлен

$$Q_n\left(\cos\frac{2\pi\,(x-x_0)}{L},\ \sin\frac{2\pi\,(x-x_0)}{L},\ f\right)=$$

$$= \sum_{k=0}^{n} a_k \cos \frac{2\pi k\,(x-x_0)}{L} + \sum_{k=1}^{n+1} b_k \sin \frac{2\pi k\,(x-x_0)}{L}, \qquad (2.28)$$

удовлетворяющий равенствам $Q_n(x_m) = f_m$, $m = 0, \ldots, N-1$.

Коэффициенты этого многочлена задаются формулами

$$a_0 = \frac{1}{N} \sum_{m=0}^{N-1} f_m, \qquad b_{n+1} = \frac{1}{N} \sum_{m=0}^{N-1} (-1)^m f_m,$$

$$a_k = \frac{2}{N} \sum_{m=0}^{N-1} f_m \cos k \left(\frac{2\pi m}{N} + \frac{\pi}{N} \right), \quad k = 1, 2, \ldots, n, \qquad (2.29)$$

$$b_k = \frac{2}{N} \sum_{m=1}^{N-1} f_m \sin k \left(\frac{2\pi m}{N} + \frac{\pi}{N} \right), \quad k = 1, 2, \ldots, n.$$

Теорема 11. (Второй вариант задания узлов интерполяции). *Пусть* $N = 2n$. *При произвольном задании значений функции* f_m, *периодической с периодом* L, *в узлах сетки*

$$x_m = x_0 + \frac{Lm}{N}, \qquad m = 0, \pm 1, \ldots, \pm(N-1)$$

существует один и только один интерполяционный тригонометрический многочлен

$$Q_n\left(\cos \frac{2\pi\,(x-x_0)}{L}, \ \sin \frac{2\pi\,(x-x_0)}{L}, \ f \right) =$$

$$= \sum_{k=0}^{n} a_k \cos \frac{2\pi k\,(x-x_0)}{L} + \sum_{k=1}^{n-1} b_k \sin \frac{2\pi k\,(x-x_0)}{L}, \qquad (2.30)$$

удовлетворяющий равенствам $Q_n(x_m) = f_m$, $m = 0, \pm 1, \pm 2, \ldots, \pm\,(N-1)$.

Коэффициенты этого многочлена задаются формулами

$$a_0 = \frac{1}{N} \sum_{m=0}^{N-1} f_m, \qquad a_n = \frac{1}{N} \sum_{m=0}^{N-1} (-1)^m f_m,$$

$$a_k = \frac{2}{N} \sum_{m=0}^{N-1} f_m \cos \frac{2\pi km}{N}, \qquad k = 1, 2, \ldots, n-1, \qquad (2.31)$$

$$b_k = \frac{2}{N}\sum_{m=1}^{N-1} f_m \sin\frac{2\pi km}{N}, \qquad k = 1, 2, \ldots, n-1.$$

Теорема 12. (Произвольное расположение узлов интерполяции на отрезке периодичности). *Пусть заданы значения* f_i : $i = 1, \ldots, N$, *периодической с периодом L функции* $f(x)$ *в* $N = 2n$ *несовпадающих точках* x_i : $i = 1, \ldots, N$, *принадлежащих отрезку* $[a, b]$, $b - a = L$, $f(a) = f_1 = f_N = f(b)$.

Тогда существует один и только один интерполяционный тригонометрический многочлен

$$Q_n\left(\cos\frac{2\pi(x-a)}{L},\ \sin\frac{2\pi(x-a)}{L},\ f\right) = \sum_{k=0}^{N-1} f_k \cdot l_k(x), \qquad (2.32)$$

$$l_k(x) = \frac{\sin\frac{\pi(x-x_1-a)}{L}\ldots\sin\frac{\pi(x-x_i-a)}{L}\ldots\sin\frac{\pi(x-x_{N-1}-a)}{L}}{\sin\frac{\pi(x_k-x_1-a)}{L}\ldots\sin\frac{\pi(x_k-x_i-a)}{L}\ldots\sin\frac{\pi(x_k-x_{N-1}-a)}{L}}. \qquad (2.33)$$

(*В произведениях отсутствует сомножитель, соответствующий i = k, так что* $l_k(x_k) = 1$).

Построенные тригонометрические многочлены обладают определенными преимуществами перед алгебраическим многочленом, построенным по значениям функции в узлах x_m.

Во-первых, при $N \to \infty$ *погрешность тригонометрической интерполяции*

$$R_N(x, f) = f(x) - Q\left(\cos\frac{2\pi x}{L},\ \sin\frac{2\pi x}{L},\ f\right)$$

равномерно стремится к нулю, если $f(x)$ имеет хотя бы вторую производную, причем скорость убывания погрешности автоматически учитывает гладкость $f(x)$, т. е. возрастает с ростом числа (r+1) производных:

$$\max_x |R_N(x)| = \mathrm{O}\left(M_{r+1}\frac{\ln N}{N^r}\right), \qquad M_{r+1} = \max_x \left|\frac{d^{r+1}f(x)}{dx^{r+1}}\right|. \qquad (2.34)$$

Во-вторых, чувствительность тригонометрического интерполяционного многочлена к погрешности задания значений f_m в узлах с ростом числа узлов «почти» не возрастает.

Эти два положительных свойства тригонометрической интерполяции, а именно, возрастание точности при увеличении гладкости и вычислительную устойчивость, можно придать и алгебраической интерполяции функций на отрезке за счет специального выбора узлов интерполяции и использования алгебраических многочленов Чебышева, обладающих многими замечательными свойствами.

2.15. Многочлены Чебышёва

Многочлены Чебышёва можно ввести по формуле $T_k(x) = \cos(k \arccos x)$, $k = 0, 1, \ldots$. Функции $T_k(x)$ суть многочлены степени $k = 0, 1, \ldots$ При этом $T_0(x) = 1$, $T_1 = x$; $T_2(x)$, $T_3(x)$ … вычисляются по рекуррентной формуле

$$T_{k+1}(x) = 2x\, T_k(x) - T_{k-1}(x). \qquad (2.35)$$

Нули $T_k(x)$ определяются из уравнения:

$$T_k(x) = \cos(k \arccos x) = 0, \qquad (2.36)$$

т. е. $x_m = \cos \dfrac{\pi(2m+1)}{2k}$, $m = 0, 1, \ldots, k-1$.

Точки экстремума $T_k(x)$ определяются как:

$$x_l = \cos \frac{\pi l}{k}, \qquad l = 0, 1, \ldots, k.$$

Таким образом, $T_k(x)$ на интервале $-1 \le x \le 1$ имеет k вещественных нулей и $k+1$ точку экстремума. На оси x эти точки получаются проекцией пересечения полуокружности с множеством лучей, имеющих между собой равные углы.

2.16. Алгебраический интерполяционный полином на сетке из нулей полинома Чебышёва

При выборе в качестве узлов интерполяции нулей полинома Чебышева интерполяционный полином можно записать в виде:

$$P_n(x, f) = \sum_{k=0}^{n} a_k T_k(x), \tag{2.37}$$

где $a_0 = \sum_{m=0}^{n} f_m T_0(x_m)/(n+1)$; $a_k = 2\sum_{m=0}^{n} f_m T_k(x_m)/(n+1)$;

$n + 1$ — число узлов интерполяции.

2.17. Алгебраический интерполяционный полином на сетке из экстремумов полинома Чебышёва

При выборе в качестве узлов интерполяции экстремумов полинома Чебышева интерполяционный полином можно записать в следующем виде:

$$P_n(x, f) = \sum_{k=0}^{n} a_k T_k(x), \tag{2.38}$$

где

$$a_0 = \frac{1}{2n}(f_0 + f_n) + \frac{1}{n}\sum_{m=1}^{n-1} f_m,$$

$$a_k = \frac{1}{n}(f_0 + (-1)^k f_n) + \frac{2}{n}\sum_{m=1}^{n-1} f_m T_k(x_m),$$

$$a_n = \frac{1}{2n}(f_0 + (-1)^n f_n).$$

Здесь $n + 1$ — число узлов интерполяции.

2.18. Чувствительность интерполяционного тригонометрического многочлена к погрешностям задания функции в узлах интерполяции

Чувствительность интерполяционного тригонометрического многочлена к погрешности задания значений f_m оценивается следующим образом. Пусть вместо $f = [f_m]$ задана сеточная функция $f + \delta f = \{f_m + \delta f_m\}$. Тогда возникающая погрешность

$$\delta Q_n = Q_n\left(\cos\frac{2\pi x}{L},\ \sin\frac{2\pi x}{L},\ \delta f\right) \tag{2.39}$$

и, следовательно, мерой чувствительности интерполяционного тригонометрического многочлена к возмущению δf входных данных могут служить числа L_n, называемые *константами Лебега* (см. п. 2.8):

$$\max_x |\,\delta Q_n\,| \le \mathrm{L}_n \max_m |\,f_m\,|. \tag{2.40}$$

Теорема 13. *Константы Лебега тригонометрического интерполяционного многочлена удовлетворяют оценке* $\mathrm{L}_n \le 2n$.

Интерполяционный полином на сетке из нулей или экстремумов полиномов Чебышёва наследует от тригонометрической интерполяции слабый рост константы Лебега при увеличении n. Для этого случая справедлива оценка

$$\mathrm{L}_n \approx \frac{2}{\pi}\ln n + 1.$$

2.19. Контрольные вопросы

1. Докажите, что при выборе в качестве узлов интерполяции нулей полинома Чебышева, алгебраический интерполяционный полином можно записать в виде

$$P_n(x, f) = \sum_{k=0}^{n} a_k T_k(x),$$

где

$$a_0 = \frac{1}{n+1}\sum_{m=0}^{n} f_m T_0(x_m), \qquad a_k = \frac{2}{n+1}\sum_{m=0}^{n} f_m T_k(x_m).$$

Здесь $n + 1$ — число узлов интерполяции.

2. Докажите, что при выборе в качестве узлов интерполяции экстремумов полинома Чебышева алгебраический интерполяционный полином можно записать в виде

$$P_n(x, f) = \sum_{k=0}^{n} a_k T_k(x),$$

где коэффициенты

$$a_0 = \frac{1}{2n}(f_0 + f_n) + \frac{1}{n}\sum_{m=1}^{n-1} f_m T_0(x_m),$$

$$a_k = \frac{1}{n}(f_0 + (-1)^k f_n) + \frac{2}{n}\sum_{m=1}^{n-1} f_m T_k(x_m),$$

$$a_n = \frac{1}{n}(f_0 + (-1)^n f_n) + \frac{1}{n}\sum_{m=1}^{n-1} f_m (-1)^m T_n(x_m).$$

Здесь $n + 1$ — число узлов интерполяции.

3. Выведите формулы для интерполяции табличной функции кубическим сплайном (Шонберга):

а) на равномерной сетке;

б) на неравномерной сетке.

Указание. Пусть m_i — значение второй производной в i-м узле, m_{i+1} — значение второй производной в $i + 1$-м узле. На отрезке $[x_i, x_{i+1}]$ $P_3''(x)$ — линейная функция. На этом отрезке $P_3(x)$ — непрерывная функция, которую можно получить, дважды интегрируя $P_3''(x)$ по x. При этом возникают две константы интегрирования. Они находятся из условий: $P_3(x_i) = f_i$, $P_3(x_{i+1}) = f_{i+1}$. Для значения вторых производных m_i строится система алгебраических уравнений из условия $P_3'(x_i + 0) = P_3'(x_i - 0)$.

2.20. Порядок выполнения работы

1. Войдите в меню «*Параметры*/*Таблица*» и выберите из предложенного списка функций в пункте «*Функции*» какую-нибудь гладкую (т. е. имеющую непрерывные производные) функцию. В меню «*Сетка*» выберите пункт равномерная, число узлов, равное двум. Алгебраический интерполяционный полином какой степени может быть построен по этим данным?

Выберите в подменю «*Метод*/*Глобальная*» строку «*Ин-*

терполяционный полином в форме Лагранжа», степень которого 3, 4, 5, 6, 7, …, 50. Постройте интерполяционный полином. Теоретически оцените погрешность интерполяции и сравните ее с фактической (т. е. разностью между исходной функцией и ее интерполянтом по данной таблице) погрешностью, выбрав пункт меню «*Окна*/*Ошибка*». Сформируйте таблицу, выбрав функцию, имеющую:

а) только две непрерывных производных ($y = \text{sign}\ (x)\ x^3 / 3$);

б) одну непрерывную производную ($y = \text{sign}\ (x)\ x^2 / 2$);

в) не имеющую непрерывной производной ($y = |\ x\ |$).

Как будет меняться фактическая погрешность восстановления функции с ростом степени полинома для равномерной сетки? Почему в пункте а) для нечетных степеней k интерполяционного полинома ошибка меньше, чем для $k - 1$ и $k + 1$? То же для сетки из нулей полинома Чебышева. Чем объяснить существенное различие в поведении погрешности?

2. Выберите целую (разлагающуюся в сходящийся степенной ряд для любого конечного x) функцию (например, $y = e^{-x^2}$) и постройте на сетке из равноотстоящих узлов глобальный алгебраический интерполянт. Сравните различные способы вычисления интерполяционного полинома:

а) находя его коэффициенты по базису $\{x^k\}$, решая соответствующую линейную систему;

б, в) записывая интерполяционный многочлен в форме Лагранжа; в форме Ньютона.

Какой способ требует меньшего числа арифметических действий?

Выберите число узлов $n = 10, 20, 30, \ldots, 90$. Объясните проблемы, возникающие при численном решении линейной системы. Почему результаты использования формы Ньютона и Лагранжа в начале совпадают, а затем начинают различаться?

Проанализируйте накопление погрешности при вычислении полинома в форме Ньютона при перенумерации в порядке возрастания и в порядке убывания (интерполяция вперед и назад)?

Как изменятся результаты, если вычисления производить с двойной (мантисса 52 бита) и стандартной (мантисса 24 бита) точностью?

3. Проанализируйте влияние погрешности задания функции в

узлах на величину фактической погрешности восстановления функции. В силу линейности интерполяции эти эффекты удобно изучать на функции $y = 0$. Как задать распределение погрешности для получения максимального отклонения между узлами интерполяции? Рассмотрите случай равноотстоящих узлов и сетку из нулей полинома Чебышева. Сравните с теоретической оценкой.

4. Сравните фактическую погрешность интерполяции функций $y = e^{-x^2}$ и $y = 1/(1 + 25x^2)$ (функция Рунге) на отрезке $-5 \le x \le 5$ на равномерной сетке при числе узлов n = 11, 21, 31, 41, 51. В чем причина отсутствия сходимости интерполяционного процесса для функции Рунге? Задайте теперь отрезок интерполяции $0 \le x \le 5$ и повторите вычисления. Попытайтесь объяснить полученный результат.

Указание. Обратите внимание на то, что функция $y = e^{-x^2}$ целая, а функция Рунге имеет полюсы при $x = \pm 0{,}2i$.

5. Посмотрите, как ведет себя глобальный интерполянт вне отрезка, на котором расположены узлы интерполяции. Для этого в меню «*Окно*» установите соответствующие границы изменения аргумента x для отображения $f(x)$ и $P_n(x, f)$. Что лучше использовать для экстраполяции: глобальный интерполянт или кусочно-многочленный?

6. Сравните ошибку при алгебраической интерполяции и интерполяции сплайнами на равномерной сетке из 3, 4, …, 25 узлов на отрезке [–1, 1] следующих функций:

а) $\frac{x^3}{3}\operatorname{sign} x$; б) e^{-x^2}; в) функции Рунге.

Объясните результаты.

7. Интерполирующим кубическим сплайном (Шонберга) приближается функция

$$y(x) = \begin{cases} 0, & x < 0; \\ 1, & x \ge 0. \end{cases}$$

на отрезке $x \in [-1,\ 1]$.

Почему в окрестности разрыва возникают осцилляции? Как

меняется их амплитуда в зависимости от изменения числа узлов сетки? Объясните эффект. Ответьте на те же вопросы для локального сплайна.

8. Исследуйте, как изменяется ошибка интерполяции какой-либо гладкой функции при кусочно-линейном восполнении и при глобальной интерполяции с увеличением числа узлов.

9. Выбирая функции, имеющие одну, две и т. д. непрерывные на отрезке производные, убедитесь, что кусочно-линейная интерполяция насыщается гладкостью, в то время как ошибка глобальной интерполяции на сетке из нулей полинома Чебышева тем меньше, чем больше непрерывных производных имеет $f(x)$ (не насыщающийся гладкостью алгоритм).

2.20. Библиографическая справка

Теория интерполяции — обширный раздел вычислительной математики. В данной работе изложение ведется по книге [1]. Другие аспекты тории изложены в [9, 10], в частности, подробно рассмотрено свойство насыщаемости алгоритмов. О построении глобального сплайна см. [11]. О локальных сплайнах, кроме [1], см. также [12].

В связи с развитием алгоритмов машинной графики бурно развивается теория сплайн-интерполяции. О применении сплайнов и *кривых Безье* см. также [5]. Различным приложениям сплайнов посвящены книги [13–15]. В [5, 14]дается понятие о *B-сплайнах*, нашедших широкое применение как в алгоритмах машинной графики, так и в развитии численных методов. О применении B-сплайнов в инженерных расчетах см. в [16]. В [15] описаны алгоритмы построения *сглаживающих сплайнов*, находящие применение, в частности, в обработке экспериментальных данных.

ЧИСЛЕННОЕ ИНТЕГРИРОВАНИЕ

3.1. Введение

Работа предоставляет возможность наглядно исследовать свойства основных квадратурных формул для вычисления определенного интеграла

$$I = \int_a^b f(x)\,dx.$$

Поясняются способы конструирования квадратурных формул, методы оценки погрешности, к которым они приводят. Демонстрируются приемы вычисления несобственных интегралов.

3.2. Способы конструирования квадратурных формул

Рассмотрим простейшие, но широко используемые в практических вычислениях формулы: прямоугольников (с центральной точкой), трапеций, Симпсона. Способ их получения состоит в следующем. Разобьем отрезок интегрирования $[a, b]$ на N частей точками x_n $(n = 0, 1, \ldots, N)$.

Положим $h_n = x_{n+1} - x_n$, так что

$$\sum_{n=0}^{N-1} h_n = b - a.$$

В дальнейшем будем называть x_n — узлами, h_n — шагами интегрирования. Иногда отрезок от x_n до x_{n+1} будем именовать элементарным отрезком. В частном случае шаг интегрирования может быть постоянным: $h = (b-a)/N$. Также будем использовать обозначение $f_n = f(x_n)$.

После введения шагов интегрирования искомый интеграл можно представить в виде

$$I = \sum_{n=0}^{N-1} \int_{x_n}^{x_{n+1}} f(x)\, dx = \sum_{n=0}^{N-1} I_n, \tag{3.1}$$

где $I_n = \int_{x_n}^{x_{n+1}} f(x)\, dx.$

3.2.1. Формула прямоугольников. Считая h_n малым параметром, заменим I_n в (3.1) площадью прямоугольника с основанием h_n и высотой $f_{n+1/2} = f(x_n + h_n / 2)$. Тогда придем к локальной формуле прямоугольников

$$\widetilde{I}_n = h_n f_{n+1/2}. \tag{3.2}$$

Суммируя в соответствии с (3.1) приближенные значения по всем элементарным отрезкам, получаем формулу прямоугольников для вычисления приближения к I:

$$\widetilde{I} = \sum_{n=0}^{N-1} h_n f_{n+1/2}. \tag{3.3}$$

В частном случае, когда $h_n = h = \text{const}$, формула прямоугольников принимает вид

$$\widetilde{I} = h \sum_{n=0}^{N-1} f_{n+1/2}. \tag{3.3а}$$

Замечание. Можно конструировать аналогичные формулы, используя в качестве высоты элементарных прямоугольников значение $f(x)$ не в середине отрезка, а на границе (левой или правой). Но в этом случае существенно ухудшается точность приближения вычисляемого интеграла.

3.2.2. Формула трапеций. На элементарном отрезке $[x_n, x_{n+1}]$ заменим подынтегральную функцию интерполяционным полиномом первой степени:

$$f(x) \approx f_n + \frac{f_{n+1} - f_n}{x_{n+1} - x_n}(x - x_n).$$

Выполняя интегрирование на отрезке, приходим к локальной формуле трапеций:

$$\widetilde{I}_n = \frac{1}{2}(x_{n+1} - x_n)(f_{n+1} + f_n) = \frac{1}{2} h_n (f_{n+1} + f_n). \qquad (3.4)$$

Замечание. Название формулы связано с тем, что интеграл по элементарному отрезку заменяется площадью трапеции с основаниями, равными значениям $f(x)$ на краях отрезка, и высотой, равной h_n.

Суммируя (3.4) по всем отрезкам, получаем формулу трапеций для вычисления приближения к I:

$$\widetilde{I} = \frac{1}{2} \sum_{n=0}^{N-1} h_n (f_n + f_{n+1}). \qquad (3.5)$$

В случае постоянного шага интегрирования формула принимает вид:

$$\widetilde{I} = \frac{h}{2} \sum_{n=0}^{N-1} (f_n + f_{n+1}) = \frac{h}{2} [f_0 + 2f_1 + 2f_2 + \ldots + 2f_{N-1} + f_N]. \qquad (3.5a)$$

О точности приближения $\widetilde{I}$ к I см. п. 3.3.

3.2.3. Формула Симпсона. На элементарном отрезке $[x_n, x_{n+1}]$, используя значение функции в центре отрезка, заменим подынтегральную функцию $f(x)$ интерполяционным полиномом второй степени:

$$f(x) \approx P_2(x) = f_{n+1/2} + \frac{f_{n+1} - f_n}{h_n} \left[x - \frac{x_{n+1} + x_n}{2} \right] + \\ + \frac{f_{n+1} - 2f_{n+1/2} + f_n}{2(h_n/2)^2} \left[x - \frac{x_{n+1} + x_n}{2} \right]^2.$$

Напомним, что мы обозначили: $h_n = x_{n+1} - x_n$, $f_n = f(x_n)$, а значение в полуцелой точке $f_{n+1/2} = f([x_n + x_{n+1}]/2)$.

Вычисляя интеграл от полинома на отрезке $[x_n, x_{n+1}]$, приходим к локальной формуле Симпсона:

$$\widetilde{I}_n = \frac{h_n}{6} (f_n + 4f_{n+1/2} + f_{n+1}). \qquad (3.6)$$

Суммируя (3.6) по всем отрезкам, получаем формулу Симпсона для вычисления приближения к I:

$$\widetilde{I} = \frac{1}{6} \sum_{n=0}^{N-1} h_n (f_n + 4f_{n+1/2} + f_{n+1}). \qquad (3.7)$$

Для постоянного шага интегрирования $h_n = \text{const} = h = (b-a)/N$ формула Симпсона принимает вид

$$\widetilde{I} = \frac{h}{6} \sum_{n=0}^{N-1} (f_n + 4f_{n+1/2} + f_{n+1}) = \frac{h}{6} (f_0 + 4f_{1/2} + 2f_1 + \ldots + 4f_{3/2} + 2f_{N-1} + 4f_{N-1/2} + f_N). \qquad (3.8)$$

Замечание. Последнюю формулу иногда записывают без использования дробных индексов, в виде

$$\widetilde{I} = \frac{h}{3} (f_0 + 4f_1 + 2f_2 + 4f_3 + \ldots + 2f_{N-2} + 4f_{N-1} + f_N). \qquad (3.8а)$$

К этой записи приходим, если под локальной формулой понимать результат интегрирования по паре элементарных отрезков:

$$\widetilde{I}_n = \int_{x_{n-1}}^{x_{n+1}} \widetilde{P}_2(x)\, dx = \frac{h}{3} (f_{n-1} + 4f_n + f_{n-1}),$$

где $\widetilde{P}_2(x)$ — интерполяционный полином второй степени для $f(x)$ на $[x_{n-1}, x_{n+1}]$, построенный по значениям в точках x_{n-1}, x_n, x_{n+1}. Суммируя локальные приближения по всем парам, получим (3.8а). Разумеется, число пар на $[a, b]$ в этом случае должно быть целым, т. е. N — четным.

Формулы, используемые для приближенного вычисления интеграла, называются *квадратурными.*

3.3. Погрешность квадратурных формул

Один из возможных способов оценки точности построенных формул состоит в следующем. Рассмотрим интеграл по элементарному отрезку:

$$I_n = \int_{x_n}^{x_{n+1}} f(x)\, dx.$$

Выберем на этом отрезке какую-либо «опорную» точку $x = z$ и разложим подынтегральную функцию в ряд по формуле Тейлора относительно этой точки:

$$f(x) = f(z) + f'(z)(x-z) + \frac{1}{2} f''(z)\,(x-z)^2 + \ldots + R(x-z),$$

$R(x-z)$ — остаточный член используемой формулы Тейлора.

Вычисляя интеграл от последней суммы, получаем представление I_n в виде:

$$I_n = f(z)h_n + Ah_n^2 + Bh_n^3 + \ldots + \int_{x_n}^{x_{n+1}} R(x-z)\,dx, \qquad (3.9)$$

где коэффициенты A, B, … зависят от значения производных в точке z: $f'(z)$, $f''(z)$, …

Заметим далее, что каждая из рассматриваемых квадратурных формул (прямоугольников, трапеций и Симпсона) в пределах элементарного отрезка $[x_n, x_{n+1}]$ может быть представлена следующим образом:

$$\widetilde{I}_n = h_n[r f_n + s f_{n+1/2} + q f_{n+1}] \qquad (3.10)$$

со своими коэффициентами r, s, q.

Заменяя в (3.10) каждое из значений функции f по формуле Тейлора относительно той же точки z, получим представление приближенного значения $\widetilde{I}_n$ в виде, аналогичном (3.9):

$$\widetilde{I}_n = f(z)h_n + A_1 h_n^2 + B_1 h_n^3 + \ldots + \widetilde{R}. \qquad (3.11)$$

Сравнивая представления (3.9) и (3.11), обнаруживаем, что кроме первых слагаемых в (3.9), (3.11) совпадает еще некоторое количество $(p-1)$ слагаемых, так что $A = A_1$, $B = B_1$, … Несовпадающие слагаемые характеризуют ошибку квадратурной формулы. Оценивая величины этих слагаемых, приходим к оценке для локальной (на интервале $[x_n, x_{n+1}]$) погрешности

$$|\, I_n - \widetilde{I}_n \,| \le D \max_{[x_n,\, x_{n+1}]} |\, f^{(p)} \,|\; h_n^{p+1},$$

где D — числовая константа, а $f^{(p)}$ — p-я производная функции $f(x)$.

Суммируя локальные погрешности по всем интервалам, получим требуемую оценку погрешности рассматриваемой формулы интегрирования:

$$|\widetilde{I} - I| \le D(b-a)M_p h^p, \qquad (3.12)$$

где $M_p = \max |f^{(p)}|$ по всему отрезку $[a, b]$ (если шаг интегрирования не постоянен, т. е. $h_n \neq \text{const}$, то $h = \max\limits_n h_n$).

Степень p в (3.12) *принято называть порядком точности квадратурной формулы.*

Для рассмотренных квадратурных формул полученные таким образом оценки погрешности имеют вид:

$|\widetilde{I} - I| \le \dfrac{M_2 h^2}{24}(b-a)$ — для формулы прямоугольников;

$|\widetilde{I} - I| \le \dfrac{M_2 h^2}{12}(b-a)$ — для формулы трапеций;

$|\widetilde{I} - I| \le \dfrac{M_4 h^4}{2880}(b-a)$ — для формулы Симпсона в случае, когда используются узлы с дробным индексом (3.8) и

$|\widetilde{\mathrm{I}} - \mathrm{I}| \le \dfrac{M_4 h^4}{180}(b-a)$ — для (3.8а).

Замечание 1. Для формулы трапеций приведенную оценку можно было бы получить, интегрируя по элементарным отрезкам выражение для остаточного члена соответствующего интерполяционного полинома.

Замечание 2. Полученные оценки погрешности, как следует из их вывода, зависят от гладкости подынтегральной функции. Например, при наличии 4-х (и выше) производных у $f(x)$ формула Симпсона обеспечивает 4-й порядок точности. Если же $f(x)$ только трижды непрерывно дифференцируема на $[a, b]$, то точность формулы Симпсона на порядок уменьшается.

Если известны оценки для абсолютных величин соответствующих производных, то, используя (3.12), можно apriori (до проведения расчета) определить шаг интегрирования $h = \text{const}$, при котором погрешность вычисленного результата гарантировано не превышает допустимого уровня погрешности ε. Для это-

го, как следует из (3.12), достаточно решить относительно h неравенство $D\,(b-a)\,M_p h^p \le \varepsilon$.

Однако типичной является ситуация, когда величины нужных производных не поддаются оценке. Тогда контроль за точностью вычисляемых результатов можно организовать, проводя вычисления на последовательно сгущающейся сетке узлов интегрирования.

3.3.1. Контроль за точностью вычисляемого значения интеграла. Пусть шаг интегрирования $h = \text{const}$, $I^{(h)}$ — вычисленное с шагом h приближение к I.

Если, далее вычислено также приближенное значение $I^{(h/2)}$ с шагом $h/2$, то в качестве приближенной оценки погрешности последнего вычисленного значения можно рассматривать величину

$$|I^{(h/2)} - I| \approx |I^{(h/2)} - I^{(h)}|.$$

На практике, при необходимости вычислить результат с требуемой точностью (ε) вычисления повторяются с последовательно уменьшающимся (вдвое) шагом до тех пор, пока не выполнится условие

$$|I^{(h/2)} - I^{(h)}| \le \varepsilon.$$

3.3.1а. Экстраполяция Ричардсона. Пусть, как и в предыдущем пункте, $h = \text{const}$, $I^{(h)}$ — вычисленное с шагом h приближение к I. Пусть использован метод порядка p, тогда можно оценить значение интеграла по элементарному отрезку:

$$I = I^{(h)} + ch^p + \mathrm{O}(h^{p+1}).$$

Измельчив шаг вдвое, получаем

$$I = I^{(h/2)} + 2c\left(\frac{h}{2}\right)^p + \mathrm{O}(h^{p+1}),$$

откуда главный член погрешности в первой формуле можно оценить как:

$$ch^p = \frac{I^{(h)} - I^{(h/2)}}{2^{1-p} - 1},$$

а для приближения интеграла I с порядком $\mathrm{O}(h^{p+1})$ имеем:

$$I = \frac{2^{p-1} I^{(h/2)} - I^{(h)}}{2^{p-1} - 1} + \mathrm{O}(h^{p+1}).$$

На основе алгоритма экстраполяции Ричардсона возможен алгоритм автоматического выбора шага, несколько отличный от описанного ниже.

3.3.2. Счет с автоматическим выбором шагов интегрирования. Можно применять указанное правило для контроля локальной погрешности на каждом элементарном интервале. При этом длина очередного интервала $h_n = x_{n+1} - x_n$, посредством последовательного уменьшения (или увеличения!) начальной длины вдвое, устанавливается такой, чтобы выполнялось условие

$$|\widetilde{I}_n - I_n| \approx |I_n^{(h)} - I_n^{(h/2)}| \le \frac{\varepsilon h_n}{(b-a)},$$

так что

$$\sum_n \frac{\varepsilon h_n}{b-a} = \frac{\varepsilon}{b-a} \sum_n h_n = \varepsilon.$$

Преимущество способа вычисления интеграла с автоматическим выбором шага состоит в том, что он приспосабливается к особенностям подынтегральной функции: в областях резкого изменения функции шаг уменьшается, а там, где функция меняется слабо, — увеличивается.

3.4. Приемы вычисления несобственных интегралов

Рассмотрим сходящиеся интегралы следующих двух типов:

$$\int_a^b f(x)\,dx, \text{ причем } f(x) \to \infty \text{ при } x \to a \text{ (первый тип);}$$

$$\int_a^\infty f(x)\,dx \text{ (второй тип).}$$

Замечание. Второй интеграл, вообще говоря, может быть све-

ден к первому заменой переменной интегрирования $t = 1/x$. Поэтому пока будем говорить об интегралах первого типа.

Очевидно, непосредственное использование квадратурных формул трапеций и Симпсона для вычисления таких интегралов невозможно (так как точка $x = a$ является для этих формул узлом интегрирования). По методу прямоугольников вычисления формально провести можно, но результат будет сомнительным, так как оценка погрешности теряет смысл (производные подынтегральной функции не ограничены).

Продемонстрируем приемы, которые позволяют получать в подобных случаях надежные результаты, на примере интеграла

$$I = \int_0^1 \frac{\cos x}{\sqrt{x}}\, dx.$$

а) Иногда подходящая замена переменной интегрирования позволяет вообще избавиться от особенности.

В рассматриваемом примере после замены $x = t^2$ получаем

$$I = 2\int_0^1 \cos t^2\, dt,$$

и интеграл вычисляется с требуемой точностью по любой из квадратурных формул.

б) Та же цель (избавление от особенности) достигается иногда предварительным интегрированием по частям:

$$I = \int_0^1 \frac{\cos x}{\sqrt{x}}\, dx = 2\sqrt{x}\cos x \Big|_0^1 + 2\int_0^1 \sqrt{x}\sin x\, dx.$$

Последний интеграл формально может быть вычислен стандартным образом, но оценка погрешности для любой квадратурной формулы будет иметь лишь первый порядок, так как при $x = 0$ не существует вторая производная от подынтегральной функции. Проводя еще раз интегрирование по частям, придем к интегралу от дважды непрерывно дифференцируемой функции, который с гарантированной точностью может быть вычислен по формулам трапеций или прямоугольников.

в) Если упомянутыми простыми средствами избавиться от особенности не удается, то прибегают к универсальному методу

выделения особенности. В рассматриваемом случае представим интеграл в виде суммы двух интегралов:

$$I = I_1 + I_2, \qquad I_1 = \int_0^{\delta} \frac{\cos x}{\sqrt{x}}\, dx, \qquad I_2 = \int_{\delta}^{1} \frac{\cos x}{\sqrt{x}}\, dx.$$

Второй интеграл особенности не содержит и вычисляется по любой квадратурной формуле. Вопрос о выборе величины δ обсуждается ниже.

Первый интеграл с требуемой точностью вычисляем аналитически, используя представление подынтегральной функции в окрестности особой точки ($x = 0$) в виде отрезка ряда по степеням x, который получим после замены $\cos x$ соответствующим рядом Тейлора:

$$I_1 = \int_0^{\delta} \frac{1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \ldots + (-1)^m \frac{x^{2m}}{(2m)!}}{\sqrt{x}}\, dx = 2\sqrt{\delta} - \frac{1}{2!}\frac{2}{5}\delta^{5/2} +$$

$$+ \frac{1}{4!}\frac{2}{9}\delta^{9/2} + \ldots + (-1)^m \frac{1}{(2m)!}\frac{1}{2m+1/2}\delta^{2m+1/2}.$$

Важно, что подобное аналитическое представление в малой окрестности особой точки можно получить практически во всех конкретных случаях. Как это сделать — зависит от квалификации вычислителя.

Допустим, что мы решили ограничиться в полученном представлении первыми m слагаемыми. При этом для данного примера мы допускаем погрешность, которая не превосходит последнего приведенного в записи для I_1 слагаемого в силу того, что ряд для δ — знакопеременный. Следовательно, для выбора двух параметров (δ и m) имеем следующий критерий

$$\frac{1}{(2m)!}\frac{1}{2m+1/2}\delta^{2m+1/2} \le \frac{\varepsilon}{2} \qquad (3.13)$$

($\varepsilon/2$ отводится в качестве допустимого уровня погрешности при вычислении I_2).

Таким образом, один из параметров (m или δ) можно задавать по своему усмотрению, второй – определяется из неравенства (3.13). При этом нужно принять в расчет следую-

щее соображение.

Если $\delta << 1$, то существенно ухудшается оценка погрешности для любой квадратурной формулы, которую мы предполагаем использовать для вычисления I_2, так как в качестве коэффициента при h^p (где p — порядок точности выбранной формулы) фигурирует максимальное на отрезке $[\delta, 1]$ значение p-й производной от подынтегральной функции, которое при $x = \delta$ в рассматриваемом случае имеет порядок $\delta^{-(p+0,5)}$.

Кроме того, при вычислении интеграла I_2 придется вычислять подынтегральную функцию $f(x)$ от аргумента либо равного δ (для формул трапеций и Симпсона), либо очень близкого к δ (для формулы прямоугольников с центральной точкой). Но значение $f(\delta)$ при малом δ может быть настолько большим (в рассматриваемом случае $f(\delta) \sim 1/\sqrt{|\delta|}$), что абсолютная погрешность функции $f(\delta)$ не позволит вычислить интеграл с требуемой точностью при заданной длине мантиссы и выбранном шаге интегрирования.

Следовательно, целесообразно задать «не слишком малое» δ (например, $\delta = 0,1$), а затем m найти из условия (3.13).

Замечание. Разумеется, если поиск последовательных членов разложения подынтегральной функции затруднителен, то приходится ограничиваться доступными членами. В этом случае из условия типа (3.13) находится параметр δ.

Рассмотрим пример вычисления интеграла второго типа:

$$\int_0^{\infty} e^{-x^2}\, dx.$$

Можно, как уже отмечалось, свести его к интегралу первого типа. Но мы воспользуемся универсальным приемом выделения особенности. Особенность состоит в том, что верхний предел интегрирования — бесконечность. Представим интеграл в виде суммы двух интегралов: $I = I_1 + I_2$, где I_1 — интеграл по конечному отрезку $[a, A]$; I_2 — интеграл по $[A, \infty]$. Вычисление I_1 при заданном A затруднений не вызывает.

Выберем теперь A так, чтобы в пределах допустимой по-

грешности вторым интегралом можно было пренебречь, т. е. так, чтобы $|I_2| \le \varepsilon/2$. Например, учитывая, что при $A \ge 1$

$$\int_A^\infty e^{-x^2}\,dx \le \int_A^\infty x\,e^{-x^2}\,dx = \frac{1}{2}e^{-A^2},$$

и требуя, чтобы выполнялось условие

$$\frac{1}{2}e^{-A^2} \le \frac{1}{2}\varepsilon,$$

найдем $A \ge \sqrt{|\ln\varepsilon|}$.

3.5. Контрольные вопросы и упражнения

1.1. Получить оценки для погрешности квадратурной формулы трапеций.

1.2. Выполнить то же задание для формулы прямоугольников (с центральной точкой).

1.3. То же для формулы Симпсона.

2. Описать алгоритм автоматического выбора шага, основанный на экстраполяции по Ричардсону.

3.6. Порядок выполнения работы

1.1. Вычислить с постоянным шагом h = 0,1; 0,02 приближенное значение интеграла

$$I = \int_0^1 x\,(10x+1)\,(10x+2)\,dx$$

по формулам:

а) прямоугольников;
б) трапеций;
в) Симпсона.

1.2. Сравнить фактическую погрешность с теоретической (точное значение интеграла I = 36.)

1.3. Теоретически оценить шаг h, при котором погрешность результата для используемой квадратурной формулы не превышает $\varepsilon = 10^{-4}$. Сравнить найденное значение h с шагом, который вырабатывается по заданному ε автоматически при счете с

уменьшающимся шагом.

2. Выполнить задания п. 1.1–1.3 для

$$I = \int_{0,01}^{1} \ln x \, dx.$$

(Точное значение: $I = 0{,}01\,(1 + \ln 100) - 1 \approx -0{,}94395$).

3. То же задание для интеграла

$$I = \int_{0}^{1} e^{4x} \sin 40\pi x \; dx.$$

Объяснить результат, полученный при счете с автоматическим контролем точности (режимы с уменьшающимся шагом и с автоматическим выбором шага), когда начальное значение $h = 0{,}1$.

Каким должно быть начальное значение h в этом случае? (Точное значение интеграла $I = -0{,}426089$.)

4. По основным квадратурным формулам (прямоугольников, трапеций, Симпсона) вычислить интеграл

$$I = \int_{-1}^{2} x \,|\, x \,|\; dx$$

с шагом $h = 0{,}1$; 0,05; 0,02 и с шагом $h = 1/4$; 1/8; 1/16, 1/32.

5. Вычислить интеграл $I = \int_{0}^{1} (1 + x^{3/2})^{-1} \ln x \; dx.$

Указание. См. п. 3.4.

3.7. Библиографическая справка

О методах численного интегрирования см. [1–3, 5, 8, 9, 10]. Отметим, что большей точности метода при небольшом количестве точек (узлов сетки) позволяют достигать квадратурные формулы Гаусса. Про них можно прочитать в [5], теоретические основы методов Гаусса разбираются в [9, 10].

ЧИСЛЕННОЕ РЕШЕНИЕ СИСТЕМ ЛИНЕЙНЫХ УРАВНЕНИЙ

4.1. Введение

В этой работе Вы познакомитесь с численными методами решения систем линейных алгебраических уравнений. Все рассматриваемые методы ориентированы на решение систем уравнений большой размерности. Подобные системы возникают на практике, например, при интегрировании уравнений в частных производных эллиптического типа.

Везде далее будем рассматривать линейное пространство R^m. Будем также считать, что $\mathbf{x}$, $\mathbf{x}^k$, $k = 0, 1, 2, \ldots$, являются элементами пространства R^m. Кроме того, под $\mathbf{A}$ будем понимать тот или иной линейный оператор, действующий из R^m в R^m.

Запишем систему линейный уравнений в следующем виде:

$$\mathbf{A}\mathbf{x} = \mathbf{f}. \tag{4.1}$$

Предполагается, что определитель матрицы $\mathbf{A}$ отличен от нуля, так что решение существует и оно единственно.

Численные методы решения системы (4.1) делятся на две группы: *прямые* и *итерационные*.

В прямых методах точное решение находится за конечное число арифметических действий. Примерами прямых методов являются метод Гаусса и метод сопряженных градиентов.

Каждый итерационный метод состоит в том, что при решении системы (4.1) указывается рекуррентное соотношение, которое по заданному произвольно «нулевому» приближению $\mathbf{x}^0$ решения $\mathbf{x}$ позволяет вычислить первое, второе, …, p-е приближение $\mathbf{x}^p$ ($p = 1, 2, 3, \ldots$) решения $\mathbf{x}$.

Иногда необходимо решить уравнение (4.1), в котором $\mathbf{A}$ — линейный оператор, не заданный явно в виде матрицы. В

этом случае прямые методы в отличие от итерационных, могут быть неприменимы.

В данной работе для *итерационных* методов дается наглядная иллюстрация изменения нормы разности двух последовательных приближений. Эффективность различных методов (в том числе прямых) можно сравнить по затратам машинного времени (необходимая информация выводится на экран), которое необходимо для достижения заданной точности, т. е. до тех пор, пока не будет выполнена оценка

$$\| \mathbf{x} - \mathbf{x}^n \| < \varepsilon.$$

Задача отыскания точного решения уравнения (4.1) не диктуется, как правило, запросами приложений. Обычно допустимо использование приближенного решения, известного с достаточной точностью. Поэтому во многих случаях для вычисления решения уравнения (4.1) точным методам целесообразно предпочесть тот или иной итерационный метод.

Итерационный процесс должен быть построен так, чтобы последовательность приближений $\mathbf{x}^k$ стремилась к решению $\mathbf{x}$ уравнения (4.1). Тогда для любого $\varepsilon > 0$ существует номер $n = n(\varepsilon)$ такой, что $\| \mathbf{x} - \mathbf{x}^n \| < \varepsilon$. Задавая $\varepsilon > 0$ достаточно малым, можно воспользоваться n-м приближением $\mathbf{x}$.

Представлены следующие девять методов:

1) Гаусса;
2) сопряженных градиентов;
3) простых итераций;
4) с оптимальным параметром;
5) с оптимальным набором параметров;
6) Зейделя;
7) трехслойный метод Чебышева;
8) минимальных невязок;
9) скорейшего спуска.

4.2. Обусловленность систем линейных уравнений

Две на первый взгляд похожие системы линейных уравнений могут обладать различной чувствительностью к погрешностям задания входных данных. Это свойство связано с понятием *обусловленности системы уравнений*.

Числом обусловленности линейного оператора **A**, дейст-

вующего в нормированном пространстве R^m, а также числом обусловленности системы линейных уравнений $\mathbf{Ax} = \mathbf{f}$ назовем величину

$$\mu(\mathbf{A}) = \| \mathbf{A} \| \cdot \| \mathbf{A}^{-1} \| .$$

Таким образом, появляется связь числа обусловленности с выбором нормы.

Предположим, что матрица и правая часть системы заданы неточно. При этом погрешность матрицы составляет $\delta\mathbf{A}$, а правой части — $\delta\mathbf{f}$. Можно показать, что для погрешности $\delta\mathbf{x}$ имеет место следующая оценка ($\| \mathbf{A}^{-1} \| \cdot \| \delta\mathbf{A} \| < 1$):

$$\frac{\| \delta\mathbf{x} \|}{\| \mathbf{x} \|} \le \frac{\mu(\mathbf{A})}{1 - \mu(\mathbf{A}) \dfrac{\| \delta\mathbf{A} \|}{\| \mathbf{A} \|}} \left(\frac{\| \delta\mathbf{A} \|}{\| \mathbf{A} \|} + \frac{\| \delta\mathbf{f} \|}{\| \mathbf{f} \|} \right).$$

В частности, если $\delta\mathbf{A} = 0$, то

$$\frac{\| \delta\mathbf{x} \|}{\| \mathbf{x} \|} \le \mu(\mathbf{A}) \frac{\| \delta\mathbf{f} \|}{\| \mathbf{f} \|} .$$

При этом решение уравнения $\mathbf{Ax} = \mathbf{f}$ не при всех $\mathbf{f}$ одинаково чувствительно к возмущению $\delta\mathbf{f}$ правой части.

Свойства числа обусловленности линейного оператора:

$$1.\ \mu(\mathbf{A}) = \frac{\max \| \mathbf{Ax} \|}{\min \| \mathbf{Ax} \|},$$

причем максимум и минимум берутся для всех таких $\mathbf{x}$, что $\| \mathbf{x} \| = 1$. Как следствие,

$$2.\ \mu(\mathbf{A}) \ge 1.$$

$$3\ \mu(\mathbf{A}) \ge \frac{| \lambda_{\max} |}{| \lambda_{\min} |},$$

где $\lambda_{\max}$ и $\lambda_{\min}$ — соответственно минимальное и максимальное по модулю собственные значения матрицы $\mathbf{A}$. Равенство достигается для самосопряженных матриц в случае использования евклидовой нормы в пространстве R^m.

$$4.\ \mu(\mathbf{AB}) \le \mu(\mathbf{A})\mu(\mathbf{B}).$$

Матрицы с большим числом обусловленности (ориентиро-

вочно $\mu(\mathbf{A}) > 10^3$) называются *плохо обусловленными матрицами*. При численном решении систем с плохо обусловленными матрицами возможно сильное накопление погрешностей, что следует из оценки для погрешности $\delta\mathbf{x}$. Исследуем вопрос о погрешности решения, вызванной ошибками округления в ЭВМ при вычислении правой части. Пусть t — двоичная разрядность чисел в ЭВМ. Каждая компонента f_i вектора правой части округляется с относительной погрешностью $O(2^{-t})$. Следовательно,

$$\frac{\| \delta\mathbf{x} \|}{\| \mathbf{x} \|} \le \mu(\mathbf{A})\, O(2^{-t}).$$

Таким образом, погрешность решения, вызванная погрешностями округления, может быть недопустимо большой в случае плохо обусловленных систем.

4.3. Метод Гаусса

Метод Гаусса является *прямым* методом. Запишем систему линейных уравнений в виде:

$$\begin{aligned} &a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m = f_1, \\ &a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m = f_2, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mm}x_m = f_m. \end{aligned} \qquad (4.1a)$$

Пусть $a_{11} \neq 0$. Выполнения этого условия всегда можно добиться перенумерацией компонент вектора **x**, либо перестановками уравнений. Тогда запишем первое из уравнений (4.1a) в виде

$$x_1 + \frac{a_{12}}{a_{11}} x_2 + \ldots + \frac{a_{1m}}{a_{11}} x_m = \frac{f_1}{a_{11}}. \qquad (4.2)$$

Вычтем это уравнение, умноженное на соответствующий коэффициент a_{i1}, из i-го уравнения системы (4.1a), где $i > 1$. Тогда эти уравнения преобразуются к виду

$$\begin{aligned} &a'_{22}x_2 + a'_{23}x_3 + \ldots + a'_{2m}x_m = f'_2, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &a'_{m2}x_2 + a'_{m3}x_3 + \ldots + a'_{mm}x_m = f'_m. \end{aligned} \qquad (4.3)$$

Первая компонента вектора **x** не входит в подсистему (4.3). Выполним с (4.3) те же операции, что и ранее с системой уравнений (4.1а). В результате получим новую подсистему уравнений, в которую уже не будут входить x_1 и x_2. Она дополняется уравнением (4.2) и первым уравнением системы (4.3), не содержащим x_1. Уже после $m - 3$ подобных циклов мы получаем подсистему из одного уравнения с одним неизвестным. При этом матрица системы будет иметь треугольный вид. Совокупность операций по приведению системы уравнений к такому виду называется *прямым ходом метода Гаусса*. Решение системы с треугольной матрицей не вызывает затруднений. Совокупность операций по нахождению решения системы с треугольной матрицей называется *обратным ходом метода Гаусса.*

Общее число арифметических операций при решении системы (4.1а) методом Гаусса составляет $O(m^3)$.

В приложениях матрица **A** часто имеет трехдиагональный вид, т. е. ненулевые элементы матрицы располагаются на главной диагонали и двух близлежащих к ней. Применение метода Гаусса к такой системе уравнений называется *методом прогонки.*

Метод Гаусса может оказаться неустойчивым по отношению к росту вычислительной погрешности.

Теорема 1. *Для устойчивости метода Гаусса достаточно диагонального преобладания, т. е. выполнения неравенств*

$$|a_{ii}| > |a_{i1}| + |a_{i2}| + \ldots + |a_{ii-1}| + |a_{ii+1}| + \ldots + |a_{im}| + \delta,$$

$\delta > 0$, $i = 1, 2, \ldots, m$.

Вычислительную погрешность метода Гаусса можно уменьшить, если применить модификацию метода, называемую *методом Гаусса с выделением главного элемента.* Суть этой модификации заключается в следующем. Нумерация компонент вектора **x** и уравнений выбирается так, чтобы a_{11} являлся максимальным по модулю элементом матрицы **A**. Затем, после исключения x_1, перенумерацией строк и столбцов добиваются того, чтобы a'_{22} в (4.3) являлся максимальным по модулю элементом матрицы системы (4.3). Подобная процедура продолжается и далее.

При расчете на реальной ЭВМ с заданным числом разрядов, наряду с влиянием неточного задания входных данных на каждой арифметической операции, вносятся ошибки округления. Влияние последних на результат зависит не только от раз-

рядности машины, но и от числа обусловленности матрицы системы, а также от выбранного алгоритма. Существуют алгоритмы, учитывающие влияние ошибок округления и позволяющие получить результат с гарантированной точностью, если система не обусловлена настолько плохо, что при расчете с заданной разрядностью эта точность не может быть гарантирована.

4.4. Метод сопряженных градиентов

Пусть матрица **A** системы (4.1) самосопряженная и положительно-определенная:

$$\mathbf{A} = \mathbf{A}^* > 0.$$

Запись $\mathbf{A} > 0$ означает, что для любого $x \in \mathrm{R}^m$, такого что $\| \mathbf{x} \| \neq 0$, выполнено

$$(\mathbf{Ax}, \mathbf{x}) \geq \alpha\, (\mathbf{x}, \mathbf{x}),$$

где $\alpha > 0$.

Метод сопряженных градиентов может применяться и как прямой, и как итерационный. Итерационный метод не уступает по скорости сходимости методу Чебышева, который будет рассмотрен ниже, но выгодно отличается от последнего тем, что не требует знания границ спектра. В то же время, метод сопряженных градиентов уступает методу Чебышева, поскольку является неустойчивым для плохо обусловленных матриц высокой размерности.* В точной арифметике этот метод дает точное решение не позднее p итераций, где p — число различных собственных значений. Наиболее благоприятная ситуация для применения метода сопряженных градиентов имеет место, если границы спектра неизвестны, а порядок m системы много больше числа итераций k, при котором погрешность на k-й итерации ε^k удовлетворяет поставленному требованию точности.

Метод сопряженных градиентов можно рассматривать как модифицированный вариант метода наискорейшего спуска (см. ниже).

Рассмотрим квадратичную форму

---

* В настоящее время существуют модификации метода сопряженных градиентов, для которых повышается скорость сходимости, и улучшается устойчивость, — см. [6].

$$F(\mathbf{x}) = \frac{1}{2}(\mathbf{x}, \mathbf{Ax}) - (\mathbf{f}, \mathbf{x}).$$

Поскольку

$$F(\mathbf{x}) = F(\bar{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}, \mathbf{A}(\mathbf{x} - \bar{\mathbf{x}})),$$

где $\bar{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{f}$ — решение системы (4.1), и $\mathbf{A} = \mathbf{A}^* > 0,$ то решение задачи (4.1) эквивалентно минимизации $F(\mathbf{x})$. Для градиента $F(\mathbf{x})$ справедливо выражение

$$\nabla F(\mathbf{x}) = \mathbf{f} - \mathbf{Ax} = -\mathbf{r}.$$

В этом направлении функционал $F(\mathbf{x})$ обладает наибольшей мгновенной скоростью изменения.

Для метода сопряженных градиентов следующее приближение к решению выбирается по формуле

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \tau_k \mathbf{p}^k,$$

где $\mathbf{p}^k$ — «вектор направления». Параметр τ_k выбирается таким образом, чтобы вектор $\mathbf{p}^k$ был **A**-сопряженным с вектором $\mathbf{p}^{k-1}$: $(\mathbf{p}^k, \mathbf{A}\mathbf{p}^{k-1}) = 0,$ а значение τ_k вычисляется из условия минимума $F(\mathbf{x}^{k+1})$:

$$\tau_k = \frac{(\mathbf{p}^k, \mathbf{r}^k)}{(\mathbf{p}^k, \mathbf{A}\mathbf{p}^k)}.$$

Вычислительный алгоритм состоит в следующем. Зададим произвольный вектор $\mathbf{x}^0 \in \mathrm{R}^m$ и построим последовательность

$$\mathbf{x}^1 = (\mathbf{E} - \tau_1 \mathbf{A})\,\mathbf{x}^0 + \tau_1 \mathbf{f},$$
$$\mathbf{x}^{k+1} = \alpha_{k+1}(\mathbf{E} - \tau_{k+1}\mathbf{A})\,\mathbf{x}^k + (1 - \alpha_{k-1})\,\mathbf{x}^{k-1} + \alpha_{k+1}\tau_{k+1}\mathbf{f}, \tag{4.4}$$

где

$$\tau_{k+1} = \frac{(\mathbf{r}_k, \mathbf{r}_k)}{(\mathbf{A}\mathbf{r}_k, \mathbf{r}_k)}, \qquad \mathbf{r}_k \equiv \mathbf{A}\mathbf{x}^k - \mathbf{f}, \qquad k = 0, 1, 2, \ldots, \tag{4.5}$$

$$\alpha_1 = 1,$$

$$\alpha_{k+1} = \left(1 - \frac{\tau_{k+1}}{\tau_k} \frac{(\mathbf{r}_k, \mathbf{r}_k)}{(\mathbf{A}\mathbf{r}_{k-1}, \mathbf{r}_{k-1})} \frac{1}{\alpha_k} \right)^{-1}, \quad k = 1, 2, \ldots$$

Оказывается, что существует номер k_0, $k_0 < m$, такой, что член $\mathbf{x}^{k_0}$ последовательности (4.4) совпадает с точным решением **x**:

$$\mathbf{x} = \mathbf{x}^{k_0}, \qquad k_0 \le m. \tag{4.6}$$

Элементы последовательности $\mathbf{x}^1$, $\mathbf{x}^2$, …, $\mathbf{x}^n$ являются уточняющимися с ростом номера k приближениями к решению; при заданном малом ε, ε > 0, погрешность $\| \mathbf{x} - \mathbf{x}^k \|$ приближения $\mathbf{x}^k$ при некоторых условиях может стать меньше ε даже при значениях $k << m$.

Можно показать, что «вектор направления» $\mathbf{p}^k$ с точностью до скалярного множителя представляет собой проекцию градиента $\mathbf{r}^k = \nabla F(\mathbf{x}^k) = \mathbf{f} - \mathbf{A}\mathbf{x}^k$ на пространство, натянутое на векторы $\mathbf{p}^k$, $\mathbf{p}^{k+1}$, …, $\mathbf{p}^{N-1}$, а векторы $\mathbf{p}^0$, …, $\mathbf{p}^{N-1}$ являются взаимно **A**-сопряженными. Отсюда следует название метода — «сопряженных градиентов».

В настоящее время метод сопряженных градиентов при «умеренном» числе обусловленности и больших m используется обычно именно как метод последовательных приближений.

Попытка вычисления точного решения при большом числе обусловленности и большом m с помощью последовательности (4.4) и с учетом равенства (4.6) может натолкнуться на препятствие, состоящее в возможной потере вычислительной устойчивости при нахождении членов $\mathbf{x}^k$ последовательности (4.4).

Заметим, что для применения метода не обязательно, чтобы система была задана в каноническом виде $\sum a_{ij} x_j = f_i$, $i = 1, 2, \ldots, m$, или приведена к такому виду. К тому же нет необходимости хранить матрицу **A**, которая имела бы m^2 элементов, в то время как векторы $\mathbf{y} \in \mathrm{R}^m$ и $\mathbf{z} = \mathbf{A}\mathbf{y} \in \mathrm{R}^m$, записанные в координатной форме, задаются лишь m числами каждый.

4.5. Метод простых итераций

Заметим, что систему линейных уравнений

$$\mathbf{A}\mathbf{x} = \mathbf{f} \tag{4.7}$$

можно преобразовать к виду

$$\mathbf{x} = (\mathbf{E} - \tau\mathbf{A})\,\mathbf{x} + \tau\mathbf{f}, \tag{4.8}$$

причем новое уравнение (4.8) равносильно исходному при любом значении τ. Вообще, (4.7) многими способами можно заменить равносильной системой вида

$$\mathbf{x} = \mathbf{B}\,\mathbf{x} + \varphi, \qquad \mathbf{x} \in \mathrm{R}^m, \qquad \boldsymbol{\varphi} \in \mathrm{R}^m, \tag{4.9}$$

частным случаем которой является (4.8).

Итерационная схема при заданном произвольно $\mathbf{x}^0$ имеет следующий вид

$$\mathbf{x}^{p+1} = \mathbf{B}\mathbf{x}^p + \boldsymbol{\varphi}, \qquad p = 0, 1, \ldots \tag{4.10}$$

при заданном произвольно $\mathbf{x}^0$.

Ниже приведены условия, при которых последовательность (4.10) сходится к решению системы (4.7).

Рассмотрим частный случай итерационной формулы (4.10):

$$\mathbf{x}^{p+1} = (\mathbf{E} - \tau\mathbf{A})\,\mathbf{x}^p + \tau\mathbf{f}, \qquad p = 0, 1, \ldots \tag{4.11}$$

Для вычисления по этой формуле достаточно уметь по заданному $\mathbf{x}^p \in \mathrm{R}^m$ находить элемент $\mathbf{A}\mathbf{x}^p$, получающийся в результате действия оператора $\mathbf{A}$.

Таким образом, итерационный процесс вычисления решения системы $\mathbf{A}\mathbf{x} = \mathbf{f}$, в отличие от метода Гаусса, можно реализовать и в случае операторной формы задания системы линейных уравнений, не выделяя какой-либо базис в R^m и не приводя к каноническому виду

$$\sum_{j=1}^{m} a_{ij} x_j = f_i, \qquad i = 1, 2, \ldots, m. \tag{4.12}$$

Для некоторых классов систем линейных уравнений число арифметических действий, необходимых для получения решения с разумной точностью итерационными методами много меньше, чем $O(m^3)$.

Теорема 2. (Достаточное условие сходимости). *Пусть в* R^m *фиксирована некоторая норма, причем соответствующая*

норма оператора **B** *равносильной системы* (4.9) *оказалась меньше единицы*:

$$\| \mathbf{B} \| \le q < 1. \tag{4.13}$$

Тогда система (4.7) *имеет одно и только одно решение* **x**; *при любом* $\mathbf{x}^0$ *из* R^m *последовательность* (4.10) *сходится к решению* **x**, *причем погрешность p-го приближения* (*или p*-й *итерации*)

$$\boldsymbol{\varepsilon}^p \equiv \mathbf{x} - \mathbf{x}^p$$

удовлетворяет оценке

$$\| \boldsymbol{\varepsilon}^p \| \le q^p \| \boldsymbol{\varepsilon}^0 \|. \tag{4.14}$$

Тем самым, норма погрешности $\| \boldsymbol{\varepsilon}^p \|$ стремится к нулю с ростом p не медленнее геометрической прогрессии q^p.

Замечание. Условие (4.13) может нарушаться при каком-нибудь другом выборе нормы $\| \mathbf{x} \|$. Однако сходимость сохранится, причем оценка (4.14) заменится оценкой

$$\| \boldsymbol{\varepsilon}^p \|' \le C q^p \| \boldsymbol{\varepsilon}^0 \|',$$

где C — некоторая постоянная, зависящая от новой нормы, а знаменатель прогрессии q — прежний.

Теорема 3. (Необходимое и достаточное условие сходимости). *Для того*, *чтобы итерационный процесс* (4.10) *сходился при любом начальном приближении, необходимо и достаточно, чтобы все собственные значения* **B** *лежали внутри единичного круга.*

Замечание. В условиях теоремы при проведении вычислений в реальной арифметике (с ограниченным числом значащих цифр) метод простых итераций может оказаться неустойчивым к росту ошибок округления. Например, если спектральный радиус матрицы **B** меньше единицы, а $\| \mathbf{B} \| > 1$. (См. пример из задания.)

4.6. Метод Зейделя и метод релаксации

Итерационная схема имеет вид

$$\mathbf{B}\mathbf{x}^{n+1} + \mathbf{C}\mathbf{x}^n = \mathbf{f}.$$

Здесь **B** — треугольная матрица, содержащая выше главной диагонали нули, а на главной диагонали и ниже ее — элементы матрицы **A**:

$$\mathbf{C} = \mathbf{A} - \mathbf{B}.$$

Теорема 4. (Достаточное условие сходимости). *Пусть* **A** — *вещественная симметричная положительно определенная матрица. Тогда метод сходится при любом начальном приближении.*

Если для отыскания следующего приближения используется итерационная схема вида

$$(\mathbf{B} - \mathbf{D})\mathbf{x}^{n+1} + \frac{1}{\omega}\mathbf{D}\,(\mathbf{x}^{n+1} + (1-\omega)\mathbf{x}^{n}) + \mathbf{C}\mathbf{x}^{n} = \mathbf{f},$$

где **D** — диагональная матрица с элементами a_{ii} на главной диагонали, то такой метод называется *методом релаксации*.

В случае $\omega > 1$ метод называется *методом верхней релаксации*. Обычно полагают $1 < \omega < 2$. Очевидно, что в случае $\omega = 1$ метод релаксации совпадает с методом Зейделя.

4.7. Метод простых итераций с оптимальным параметром

Пусть матрица **A** в (4.1) самосопряженная и положительно-определенная: $\mathbf{A} = \mathbf{A}^{*} > 0$. Собственные числа λ в этом случае действительные положительные числа.

Пусть известны наименьшее $\lambda_{\min}$ и наибольшее $\lambda_{\max}$ собственные значения **A**. Зададим произвольное приближение $\mathbf{x}^{0}$ и рассмотрим последовательность простых итераций:

$$\mathbf{x}^{p+1} = (\mathbf{E} - \tau\mathbf{A})\,\mathbf{x}^{p} + \tau\mathbf{f}, \qquad p = 0, 1, \ldots$$

Справедлива следующая теорема.

Теорема 5. 1) *Если* τ *достаточно мало, а именно, удовлетворяет неравенствам*

$$0 < \tau < \frac{2}{\lambda_{\max}}, \tag{4.15}$$

то последовательность $\mathbf{x}^{p}$ *сходится к решению* **x** *системы уравнений* $\mathbf{A}\mathbf{x} = \mathbf{f}$, *причем гарантировано убывание нормы по-*

грешности $\| \mathbf{x} - \mathbf{x}^p \|$ *при возрастании* p *в соответствии с оценкой*

$$\| \boldsymbol{\varepsilon}^p \| \le q^p \| \boldsymbol{\varepsilon}^0 \|, \qquad p = 0, 1, \ldots \tag{4.16}$$

Здесь $q < 1$ *и дается выражением*:

$$q = q(\tau) = \max \{ |1 - \tau\lambda_{\min}|, \ |1 - \tau\lambda_{\max}| \}, \tag{4.17}$$

т. е. q *— наибольшее из чисел* $|1 - \tau\lambda_{\min}|$ и $|1 - \tau\lambda_{\max}|$, *каждое из которых при условии* (4.15) *строго меньше единицы.*

2) *Пусть* τ *— произвольное число, удовлетворяющее* (4.15). *Существует начальное приближение* $\mathbf{x}^0$, *при котором оценку* (4.16) *улучшить нельзя, так как соотношение* (4.16) *превращается в точное равенство.*

3) *Если условие* (4.15) *не выполняется, то существует* $\mathbf{x}^0$, *при котором последовательность* $\mathbf{x}^p$ *не сходится с ростом* p *к решению* $\mathbf{x}$.

4) *Значение* $\tau = \tau_{\text{опт}}$, *при котором* q, *задаваемое формулой* (4.17), *принимает наименьшее значение*[*] $q_{\text{опт}} = q(\tau_{\text{опт}})$, *есть*

$$\tau = \tau_{\text{опт}} = \frac{2}{\lambda_{\min} + \lambda_{\max}}.$$

В этом случае q *принимает наименьшее значение*

$$q = q_{\text{опт}} = \frac{\mu(\mathbf{A}) - 1}{\mu(\mathbf{A}) + 1},$$

где $\mu(\mathbf{A}) = \lambda_{\max} / \lambda_{\min}$ *— число обусловленности оператора* $\mathbf{A}$.[**]

Замечание 1. Во многих случаях (например, при приближенной замене некоторых эллиптических краевых задач разностными) оператор $\mathbf{A}$ оказывается положительно определенным и самосопряженным в смысле некоторого естественного скалярного умножения. Однако обычно не удается точно указать его

---

[*] Такой вариант метода простых итераций часто называют методом *простых итераций с оптимальным параметром.*

[**] Отметим, что $\mu(\mathbf{A}) = \lambda_{\max} / \lambda_{\min}$ верно для самосопряженной матрицы, если в R^m выбрана норма вектора $\| \mathbf{x} \| = \sqrt{(\mathbf{x}, \mathbf{x})}$.

наибольшее и наименьшее собственные значения. Можно указать лишь общие оценки границ спектра, т. е. числа a и b такие, что выполняются неравенства

$$0 < a \le \lambda_{\min} \le \lambda_{\max} \le b.$$

В этом случае также можно воспользоваться методом простых итераций. Сходимость окажется тем медленнее, чем хуже известны границы спектра a и b.

Замечание 2. Чем больше число обусловленности матрицы $\mathbf{A}$, тем больше значение $q_{\text{опт}}$ и тем хуже скорость сходимости. В некоторых случаях сходимость метода простых итераций для плохо обусловленных систем уравнений можно существенно улучшить.

4.7.1. Переход к лучше обусловленной системе с помощью энергетически эквивалентного оператора. В случае плохо обусловленной системы $\mathbf{Ax} = \mathbf{f}$ иногда удается перейти к равносильной системе с оператором, который имеет меньшее число обусловленности, а затем решить эту систему методом итераций.

Пусть $\mathbf{B} = \mathbf{B}^* > 0$ — пока произвольный оператор. Умножим обе части уравнения $\mathbf{Ax} = \mathbf{f}$ на $\mathbf{B}^{-1}$. В результате получим равносильное уравнение

$$\mathbf{Cx} = \mathbf{g}, \qquad \mathbf{C} = \mathbf{B}^{-1}\mathbf{A}, \qquad \mathbf{g} = \mathbf{B}^{-1}\mathbf{f}.$$

Оператор $\mathbf{C}$ является самосопряженным и положительно определенным в смысле скалярного произведения $(\mathbf{x}, \mathbf{y})_{\mathbf{B}} = (\mathbf{Bx}, \mathbf{y})$. Имеет смысл выбирать оператор $\mathbf{B}$ лишь среди тех операторов, для которых вычисление $\mathbf{B}^{-1}\mathbf{z}$ по заданному $\mathbf{z}$ существенно проще, чем вычисление $\mathbf{A}^{-1}\mathbf{z}$. Если при этом удается выбрать $\mathbf{B}$ так, чтобы он был «похож» на оператор $\mathbf{A}$, то можно надеяться, что оператор $\mathbf{B}^{-1}\mathbf{A}$ будет «похож» на единичный, а его максимальное и минимальное собственные числа и число обусловленности будут «ближе» к единице.

4.7.2. Масштабирование как средство улучшения числа обусловленности. Прием масштабирования заключается в следующем. Каждое из скалярных уравнений системы умножается на такой множитель, чтобы максимальный коэффициент нового уравнения оказался равным единице. От этой новой, масштабированной системы уравнений вида $\mathbf{A}'\mathbf{x} = \mathbf{f}'$, переходим к равно-

сильной системе с симметричной и положительно определенной матрицей путем умножения обеих частей системы на $(\mathbf{A}')^*$.

4.8. Трехслойный метод Чебышева

Пусть $\mathbf{A} = \mathbf{A}^* > 0$, а также известны числа $a > 0$ и $b > 0$, являющиеся границами спектра оператора $\mathbf{A}$. Зададим произвольное нулевое приближение $\mathbf{x}^0$ и будем вычислять последующие приближения по формулам:

$$\mathbf{x}^1 = (\mathbf{E} - \tau\,\mathbf{A})\,\mathbf{x}^0 + \tau\,\mathbf{f},$$

$$\mathbf{x}^{p+1} = \frac{2\gamma_1\gamma_p}{\gamma_{p+1}}(\mathbf{E} - \tau\mathbf{A})\,\mathbf{x}^p - \frac{\gamma_{p-1}}{\gamma_{p+1}}\mathbf{x}^{p-1} + \frac{2\gamma_1\gamma_p}{\gamma_{p+1}}\tau\mathbf{f},$$

где $p = 1, 2, \ldots$; $\tau = \dfrac{2}{a+b}$, $\gamma_k = \gamma_k(\xi)$, $\xi = a/b$ заданы как:

$$\gamma_0 = 1, \qquad \gamma_1 = \frac{1+\xi}{1-\xi}, \qquad \gamma_{p+1} = 2\gamma_1\gamma_p - \gamma_{p-1}, \quad p = 1, 2, \ldots$$

Можно показать, что погрешность $\boldsymbol{\varepsilon}^p \equiv \mathbf{x} - \mathbf{x}^p$ приближения $\mathbf{x}^p$ удовлетворяет оценке

$$\|\boldsymbol{\varepsilon}^p\| \le \frac{2q^p}{1+q^{2p}}\,\|\boldsymbol{\varepsilon}^0\|, \qquad q = \frac{1-\sqrt{\xi}}{1+\sqrt{\xi}},$$

$$\xi = a/b, \qquad p = 1, 2, \ldots$$

Метод сходится тем быстрее, чем точнее определены границы спектра a и b.

Рассмотренный метод часто называют *трехслойным методом Чебышева* в отличие от двухслойного метода Чебышева (см. ниже). Трехслойный метод *вычислительно устойчив*, не уступает *двухслойному методу* по скорости сходимости но, в отличие от последнего, не требует оптимизации набора параметров γ_k. Трехслойный метод Чебышева уступает двухслойному методу по затратам памяти ЭВМ.

4.9. Метод с оптимальным набором параметров

Этот метод является развитием метода простых итераций, когда параметр τ зависит от номера итерации. Пусть $\mathbf{A} = \mathbf{A}^* > 0$. Бу-

дем решать уравнение $\mathbf{Ax} = \mathbf{f}$ с помощью итерационного метода

$$\mathbf{x}^{p+1} = \mathbf{x}^p - \tau_{p+1} (\mathbf{A}\mathbf{x}^p - \mathbf{f}), \quad p = 0, 1, 2, \ldots, \tag{4.18}$$

где $\mathbf{x}^0$ — некоторое начальное приближение.

Зафиксируем некоторое натуральное n. Поставим задачу выбрать такой набор итерационных параметров $\tau_1, \tau_2, \ldots, \tau_n$, при котором норма погрешности $\| \mathbf{x} - \mathbf{x}^n \|$ на n-й итерации минимальна. Такими являются следующие параметры:

$$\tau_{k+1} = \frac{\tau_0}{1 + \rho_0 \tau_k}, \qquad k = 1, 2, \ldots, n, \tag{4.19}$$

где

$$\tau_0 = \frac{2}{\lambda_{\min} + \lambda_{\max}}, \qquad \rho_0 = \frac{1 - \xi}{1 + \xi},$$
$$\xi = \frac{\lambda_{\min}}{\lambda_{\max}}, \qquad \tau_k = \cos \frac{(2k - 1)\, \pi}{2n}. \tag{4.20}$$

Здесь $\lambda_{\min}$ и $\lambda_{\max}$ — соответственно минимальное и максимальное собственные значения.

Теорема 6. *Если выбрать* τ_k *согласно* (4.19) *и* (4.20), *то для погрешности* $\varepsilon_n \equiv \| \mathbf{x} - \mathbf{x}^n \|$ *справедлива оценка*

$$\varepsilon_n \le q_n \| \mathbf{x}^0 - \mathbf{x} \|,$$

где

$$q_n = \frac{2\rho_1^n}{1 + \rho_1^{2n}}, \qquad \rho_1 = \frac{1 - \sqrt{\xi}}{1 + \sqrt{\xi}}.$$

Итерационный метод (4.18)–(4.20) иногда называют *двухслойным итерационным методом Чебышева.* Отметим, что τ_k связано с k-м нулем многочленов Чебышева.

Как уже отмечалось выше, при проведении вычислений с ограниченным числом значащих цифр двухслойный метод может оказаться неустойчивым к росту ошибок округления. Для устойчивости метода необходимо упорядочить набор параметров τ_k. Этой проблемы можно избежать, применяя трехслойный метод Чебышева, описанный выше. В настоящее время двухслойный

метод применяется редко.

4.10. Метод минимальных невязок

Пусть по-прежнему матрица $\mathbf{A} = \mathbf{A}^* > 0$. Обозначим через

$$\mathbf{r}^p = \mathbf{A}\mathbf{x}^p - \mathbf{f} \tag{4.21}$$

невязку, получающуюся при подстановке некоторого значения $\mathbf{x}^p$ в уравнение $\mathbf{Ax} = \mathbf{f}$. Итерационный алгоритм запишем в виде

$$\mathbf{x}^{p+1} = \mathbf{x}^p - \tau_{p+1}\mathbf{r}^p, \qquad p = 0, 1, 2, \ldots, \tag{4.22}$$

где $\mathbf{x}^0$ — некоторое начальное приближение.

Методом минимальных невязок называется итерационный метод (4.22), в котором τ_{p+1} выбирается из условия минимума $\| \mathbf{r}^{p+1} \|$ при заданной норме $\| \mathbf{r}^p \|$. Минимум нормы невязки $\| \mathbf{r}^{p+1} \|$ достигается, если

$$\tau_{p+1} = \frac{(\mathbf{A}\mathbf{r}^p, \mathbf{r}^p)}{\| \mathbf{A}\mathbf{r}^p \|^2}.$$

Метод минимальных невязок (4.22), (4.23) сходится с той же скоростью, что и метод простой итерации с оптимальным параметром.

Метод минимальных невязок по своей идее близок к методу сопряженных градиентов и отличается от последнего функционалом, который минимизируется на каждом шаге итераций.

4.11. Метод скорейшего спуска

Пусть матрица $\mathbf{A} = \mathbf{A}^* > 0$. Рассмотрим итерационный алгоритм (4.22) из предыдущего раздела:

$$\mathbf{x}^{p+1} = \mathbf{x}^p - \tau_{p+1}\mathbf{r}^p, \qquad p = 0, 1, 2, \ldots$$

Здесь $\mathbf{r}^p = \mathbf{A}\mathbf{x}^p - \mathbf{f}$, $\mathbf{x}^0$ — некоторое начальное приближение.

Определим вектор $\mathbf{z}^p = \mathbf{x}^p - \mathbf{x}$. Введем норму

$$\| \mathbf{r} \|_{\mathbf{A}} = \sqrt{(\mathbf{A}\mathbf{r}, \mathbf{r})}.$$

В методе *скорейшего спуска* τ_{p+1} находится из условия минимума $\| \mathbf{z}^{p+1} \|_{\mathbf{A}}$ при заданном векторе $\mathbf{z}^p$. Оказывается, что

$$\tau_{k+1} = \frac{(\mathbf{r}^p, \mathbf{r}^p)}{(\mathbf{A}\mathbf{r}^p, \mathbf{r}^p)}.$$

Метод скорейшего спуска сходится с такой же скоростью, что и метод простой итерации с оптимальным параметром.

4.12. Контрольные вопросы

1. Выпишите расчетные формулы для метода Зейделя и метода верхней релаксации.

2. Покажите, что для самосопряженной матрицы **A** число обусловленности

$$\mu(\mathbf{A}) = \frac{|\lambda_{\max}|}{|\lambda_{\min}|},$$

где λ — собственное значение матрицы **A**. Покажите, что в этом случае $\mu(\mathbf{A}^2) = \mu^2(\mathbf{A})$.

3. Как изменится число обусловленности матрицы, если умножить ее на ненулевую константу? Покажите, что для произвольных невырожденных матриц **A** и **B** выполняется неравенство

$$\mu(\mathbf{AB}) \le \mu(\mathbf{A})\ \mu(\mathbf{B}).$$

4. Дана система

$$\begin{cases} 10x_1 + x_2 = 1, \\ x_1 + 10x_2 + x_3 = 2, \\ \dots \quad \dots \quad \dots \quad \dots \quad \dots \\ x_{98} + 10x_{99} + x_{100} = 99, \\ x_1 + x_2 \quad \dots \quad + x_{100} = a. \end{cases}$$

Предложите алгоритм, позволяющий экономно вычислять совокупность решений, отвечающих различным значениям a.

Указание. Алгоритм разобран в [17].

5. Рассмотрим матрицу **A** размером 30×30, где

$$\mathbf{A} = \begin{Vmatrix} 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{Vmatrix}.$$

Оцените $\mu(\mathbf{A})$.

Наряду с **A** рассмотрите матрицу **B**, являющуюся возмущением **A**:

$$\mathbf{B} = \mathbf{A} + \mathbf{\Omega},$$

где матрица $\mathbf{\Omega}$ такая, что $\Omega_{30,1} = 2^{-29}$, а остальные ее элементы равны нулю. В этом случае $\| \mathbf{\Omega} \| = 2^{-29}$. Как изменился определитель при переходе от матрицы **A** к матрице **B**?

5. Для системы

$$\begin{cases} 10^{-3} x_1 + x_2 = b_1, \\ x_1 - x_2 = b_2, \end{cases}$$

ответьте на следующие вопросы:

а) каково число обусловленности системы;

б) какова допустимая относительная погрешность $\delta\mathbf{b}$ при заданном фиксированном векторе $\mathbf{b} = (b_1, b_2)^{\mathrm{T}}$, при которой относительная погрешность не превосходит 10^{-2};

в) каково наименьшее число μ, при котором независимо от **b** и $\delta\mathbf{b}$ выполняется оценка

$$\frac{\| \delta\mathbf{x} \|}{\| \mathbf{x} \|} \le \mu \frac{\| \delta\mathbf{b} \|}{\| \mathbf{b} \|}.$$

4.13. Порядок выполнения работы

1. Численно решите систему уравнений $\mathbf{Ax} = \mathbf{f}$ размером $N \times N$ методом простой итерации. Матрица **A** является треугольной. По диагонали у матрицы **A** расположены числа, принимающие значение *r*, за исключением элемента a_{11}, который равен еди-

нице. На одной диагонали выше главной элементы матрицы принимают значение $-r$, остальные элементы равны нулю. Правую часть системы считать нулевой.

Попытайтесь применить метод простой итерации с параметром, равным 1, для значений $N = 5, 10, 20$ и $r = 3/2$.

Как видим, необходимые и достаточные условия сходимости метода простой итерации выполнены (теорема 3 из п. 4.5 — метод простой итерации). Тем не менее, результат может оказаться отличным от ожидаемого. Подумайте, как это можно объяснить.

2. Попытайтесь решить методом сопряженных градиентов систему линейных уравнений с матрицей **A** и правой частью

$$\mathbf{b} = (1, 10^3, 10^6, 10^9, 10^{12})^{\mathrm{T}},$$

$$\mathbf{A} = \mathrm{diag}\,(1, 10^3, 10^6, 10^9, 10^{12}).$$

Объясните полученный результат.

При решении систем уравнений с плохо обусловленной матрицей иногда помогает введение масштабирования. Нетрудно догадаться, к какому эквивалентному виду надо привести систему уравнений в задаче 2, чтобы можно было с успехом применить метод сопряженных градиентов. Аналогичный переход помогает и при решении системы задачи 3.

3. Методом простых итераций со значением итерационного параметра $\tau = 0{,}3$ и начальным приближением $\mathbf{x}^0 = (1, 1)$ решите систему с нулевой правой частью и с матрицей:

$$\begin{Vmatrix} 2 & 1 \\ 1 & 2 \end{Vmatrix}.$$

Определите оптимальное значение τ и сравните требуемое для сходимости число итераций с предыдущим. Объясните результат.

4. Найдите решение системы

$$\begin{cases} 10^{-3}x + y = 5, \\ x - y = 6, \end{cases}$$

простым методом Гаусса и с выбором главного элемента. Вычисления вести с двумя знаками. Объясните результаты.

4.14. Некоторые рекомендации по работе с программой

Поясним основную схему работы с программой. Прежде всего необходимо задать условие задачи. Определение условия задачи равносильно заданию:

1) матрицы;
2) вектора (правая часть);
3) целого числа — размерности матрицы;
4) двух чисел с плавающей точкой — границ спектра;
5) параметра.

Последние три числа входят в стандартное условие задачи, и их нужно задавать. Введя условие (с клавиатуры или из заранее созданного файла), рекомендуется выбрать режим вывода на экран, после чего можно запускать программу на счет с помощью пункта меню «*Запуск*». По мере необходимости можно записывать введенные с клавиатуры матрицы в файлы или в стек.

Работа программы. После начала работы программы на экране будет демонстрироваться невязка по компонентам, если включен соответствующий режим, а по достижении заданного значения невязки программа выведет на экран полученное решение и график зависимости нормы невязки от номера итерации.

В любой момент можно остановить счет, нажав произвольную клавишу на клавиатуре или на мыши. Тогда после утвердительного ответа на вопрос, хотите ли Вы остановить программу, выдается вся информация, известная на текущий момент (текущее приближение) и т. д.

Имеется возможность ввода матрицы (и параметров) из пяти файлов matrix1.dat, …, matrix5.dat.

Формат файлов matrix следующий:

N

$$\begin{array}{ccccc} a_{11} & a_{12} & \cdots & a_{1N-1} & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N-1} & a_{2N} \\ \multicolumn{5}{c}{\dots\dots\dots\dots\dots\dots\dots\dots} \\ a_{N1} & a_{N2} & \cdots & a_{NN-1} & a_{NN} \\ b_1 & b_2 & \cdots & b_{N-1} & b_N \\ \tau & Y_1 & Y_2 & & \end{array}$$

Здесь N — размерность матрицы; a_{ij} — элемент матрицы **A**; Y_1 — нижняя граница спектра; Y_2 — верхняя граница спектра; τ — параметр.

Пользователю предоставляется возможность производить с матрицами следующие действия:

1) умножить текущую матрицу на сопряженную ей;
2) вычислить матрицу, обратную текущей;
3) запомнить текущую матрицу в стеке;
4) заменить текущую матрицу извлеченной из стека;
5) вычислить спектр текущей матрицы.

Редактирование текущей матрицы осуществляется аналогично вводу с клавиатуры, при этом у пользователя не запрашивается размерность матрицы.

В программе предусмотрен стек данных. Элементом стека является структура, состоящая из матрицы, вектора, значений границ спектра и параметра τ. В стек помещаются все перечисленные выше текущие данные. Емкость стека ограничена только размером свободной памяти.

4.15. Библиографическая справка

Прикладная (машинная) линейная алгебра — обширная, бурно развивающаяся отрасль математики. Для знакомства с ней рекомендуем книгу [5], в которой существует также обзор работ по данной тематике. Полезно также ознакомиться с работами [17–19], где обсуждаются различные аспекты линейной алгебры.

Отдельной темой является работа с матрицами специального вида — так называемыми *разреженными матрицами*, см. книги [5, 20, 21] и библиографии в них.

ЛАБОРАТОРНАЯ РАБОТА 5

ЧИСЛЕННОЕ РЕШЕНИЕ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

5.1. Введение

В этой работе Вы познакомитесь с итерационными методами решения нелинейных уравнений. Предусматривается возможность на характерных примерах рассмотреть и сравнить различные численные методы решения нелинейных уравнений. Сравнение проводится по числу итераций и затратам времени ЭВМ.

В предложенной работе используются следующие методы решения нелинейных уравнений:

1) метод простой итерации;
2) метод Ньютона;
3) метод секущих.

Первые два метода можно перенести на системы нелинейных алгебраических уравнений, а также на уравнения в произвольных метрических пространствах.

5.2. Нелинейные уравнения. Теоретическая справка

Нелинейные уравнения и системы уравнений решаются с применением итерационных методов. *Итерационные методы* (их называют также *методами последовательных приближений*) состоят в том, что решение x^* находится как предел последовательных приближений x^n при числе итераций n, стремящемся к бесконечности. Обычно задаются числом $\varepsilon > 0$ и вычисления проводят до тех пор, пока не будет выполнена в норме оценка

$$\| x^* - x^n \| < \varepsilon.$$

Так как точное решение x^* неизвестно, то это условие на практике часто заменяют лишь необходимым, но легко проверяемым

условием

$$\| x^n - x^{n-1} \| < \varepsilon.$$

Выполнение этого «условия сходимости» еще не гарантирует, что итерационный процесс сходится.

Текущее значение $\| x^n - x^{n-1} \|$ на экране монитора при выполнении работы называется термином «*погрешность*». Такое название является условным, так как на самом деле погрешностью является $\| x^* - x^n \|$.

При решении нелинейных уравнений возникают две задачи: указание областей, в которых находится по одному решению (задача локализации корней), и задача отыскания корней с заданной точностью (задача уточнения корней). Для локализации корней не существует общих приемов. Можно использовать построение графиков функции, отыскание участков ее монотонности, участков на которых функция меняет знак, и другие частные приемы.

5.3. Метод простой итерации

Пусть известно, что интересующий нас корень x^* уравнения $F(x) = 0$ лежит в интервале $\mathrm{Y} = \{x \mid a < x < b\}$. Приведем уравнение $F(x) = 0$ к равносильному уравнению вида $x = f(x)$ на интервале $\mathrm{U} \subseteq \mathrm{Y}$ таком, что $a < x < b$. Можно положить

$$f(x) = x - \alpha F(x),$$

где $\alpha = \mathrm{const}$. Такой вариант метода простых итераций иногда также называют методом релаксации. Для отыскания решения x^*, принадлежащего интервалу Y, зададим x^0, а затем вычислим последующие x^n по формуле

$$x^{n+1} = f(x^n), \qquad n = 0, 1, 2, \ldots \tag{5.1}$$

Теорема 1. *Пусть функция* $F(x)$ *непрерывна и итерационный процесс* (5.1) *сходится к значению* x^*. *Тогда* x^* *— корень уравнения* $F(x) = 0$.

Теорема 2. *Пусть функция* $f(x)$ *имеет производную во всех точках области* U (*т. е. интервала* $a < x < b$), *и пусть существует* q, $0 \le q < 1$, $q = \text{const}$ *такое*, что $\| f'_x \| \le q$ *всюду в* U. *Тогда существует такая окрестность корня*, *что при любом* x^0 *из этой окрестности метод простых итераций сходится*, *причем имеет место оценка*:

$$\| x^* - x^n \| \le q^n \| x^* - x^0 \|, \qquad n > 0.$$

Замечание. Практический смысл теоремы 2 заключается в следующем. Она утверждает, что существует такое достаточно мелкое разбиение U, при котором любая из точек одного из элементов разбиения может быть выбрана как x^0. При таком выборе начального приближения итерационный процесс с необходимостью сходится. Таким образом, поиск начального приближения, при котором итерационный процесс сходится, можно передать машине. При этом для погрешности $\varepsilon^n = x^* - x^n$ на каждой итерации выполнена оценка

$$\| \varepsilon^n \| \le q^n \| \varepsilon^0 \|, \qquad n = 1, 2, \ldots$$

5.4. Метод Ньютона

Пусть приближение x^n к корню x^* уравнения $F(x) = 0$ уже найдено. Воспользуемся приближенной формулой

$$F(x) \approx F(x^n) + F'_x \cdot (x - x^n),$$

точность которой возрастает при приближении x^n к x^*. Вместо исходного уравнения $F(x) = 0$ воспользуемся линейным уравнением

$$F(x^n) + F'_x(x^n) \cdot (x - x^n) = 0.$$

Решение этого уравнения примем за приближение x^{n+1} :

$$x^{n+1} = x^n - [F'_x(x^n)]^{-1} \cdot F(x^n), \qquad n = 0, 1, 2, \ldots \qquad (5.2)$$

Метод линеаризации Ньютона допускает простую геометрическую интерпретацию (рис. 5). График функции $F(x)$ за-

меняется касательной к нему в точке $(x^n, F(x^n))$. За приближение x^{n+1} принимается точка пересечения полученной прямой с осью абсцисс.

Формулу (5.2) можно интерпретировать как метод простой итерации с функцией $f(x) = x - [F'_x]^{-1} \cdot F(x)$. В точке корня x^* уравнения $F(x) = 0$ выполняется равенство $f'_x = 0$, поэтому неравенство $|f'_x| < q$ верно для любого положительного фиксированного значения q в достаточно малой окрестности корня.

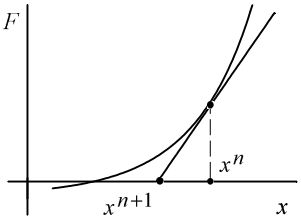


Рис. 5

Следовательно, асимптотически последовательность погрешностей $\varepsilon^n = |x^* - x^n|$ метода Ньютона убывает быстрее последовательности членов геометрической прогрессии. Справедлива *теорема о квадратичной сходимости метода Ньютона*.

Теорема 3. *Пусть функция* $F(x)$ *задана на интервале*

$$y - r < x < y + r, \qquad r > 0$$

и удовлетворяет следующим условиям:

1) $F(x)$ *дважды непрерывно дифференцируема на этом интервале*;

2) *для всех точек интервала* $F'(x) \neq 0$ *и существуют конечные значения*

$$M_1 = \sup |[F'(x)]^{-1}|,$$

$$M_2 = \sup |F''(x)|, \qquad M_2 > 0;$$

3) *уравнение* $F(x) = 0$ *имеет корень* ξ:

$$y - r \le \xi - \frac{2}{M} < \xi < \xi + \frac{2}{M} \le y + r,$$

где $M = M_1 M_2$.

Тогда для любого значения x^0 :

$$\xi - \frac{2}{M} \le x^0 \le \xi + \frac{2}{M},$$

итерационный процесс сходится к ξ, *причем*

$$|x^n - \xi| \le \left(\frac{M}{2}\right)^{2^n - 1} |x^0 - \xi|^{2^n}.$$

На практике более привлекательна такая формулировка условий сходимости метода Ньютона, для которой не нужна никакая информация о решении уравнения. Примером формулировки может служить следующая теорема.

Теорема 4. *Пусть функция* $F(x)$ *определена и дважды непрерывно дифференцируема на интервале* $|x - x^0| < r \ \ (r > 0)$. *Пусть также* $F(x^0) \ne 0, \quad F'(x^0) \ne 0,$ *существует конечное значение* $M = \sup |[F'(x^0)]^{-1} F''(x)| > 0$ *и*

$$2M \cdot \left|\frac{F(x^0)}{F'(x^0)}\right| < 1, \qquad 2 \cdot \left|\frac{F(x^0)}{F'(x^0)}\right| < r.$$

Тогда итерации процесса Ньютона сходятся к некоторому решению уравнения ξ, *для погрешности справедлива оценка*

$$|x^n - \xi| \le \frac{1}{2^n M} \cdot \left(2 \left|\frac{F(x^0)}{F'(x^0)}\right| M\right)^{2^n}.$$

5.5. Метод секущих

Зададим начальные значения x^0 и x^1. Последующие значения x^n вычисляем по формуле

$$x^{n+1} = x^n - r^n \cdot F(x^n), \qquad n = 1, 2, \ldots,$$

где $r^n = \dfrac{x^n - x^{n-1}}{F(x^n) - F(x^{n-1})}$.

Метод секущих является разностным аналогом метода Ньютона. Он применяется в тех случаях, когда вычисление производной $F'_x(x)$ является затруднительным.

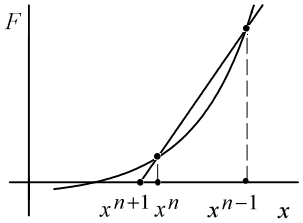


Рис. 6

Геометрическая интерпретация метода секущих состоит в следующем. Через две точки $(x^{n-1}, F(x^{n-1}))$ и $(x^n, F(x^n))$ проводится прямая. Абсцисса точки пересечения полученной таким образом прямой с осью x и является новым приближением x^{n+1} к решению нелинейного уравнения (см. рис. 6).

5.6. Мера погрешности

Условием окончания итерационного процесса является выполнение одного из двух условий (выбор условия определяется соответствующим пунктом меню):

$$\| x^n - x^{n-1} \| < \varepsilon \qquad \text{или} \qquad \| F(x^n) \| < \varepsilon.$$

Значение ε будем называть *мерой погрешности*.

Следует заметить, что выполнение условия сходимости не гарантирует, что последнее приближение x^n находится достаточно близко от корня.

В настоящей работе сходимость итерационного процесса фиксируется следующими способами: сходимость по аргументу и сходимость по функции.

5.7. Сходимость по аргументу

Считается, что итерационный процесс сошелся, если выполнено условие

$$\| x^n - x^{n-1} \| < \varepsilon$$

где ε — мера погрешности.

5.8. Сходимость по функции

Считается, что итерационный процесс сошелся, если выполнено условие

$$\| F(x^n) \| < \varepsilon.$$

5.9. Контрольные вопросы

1. Требуется найти оба корня уравнения $x = \ln(x+2)$.

1.1. Покажите, что для отыскания положительного корня можно воспользоваться итерационным процессом $x^{n+1} = \ln(x^n + 2)$, где $x^0 > 0$ —произвольно.

1.2. Можно ли указать x^0, не совпадающее с отрицательным корнем заданного уравнения, таким образом, чтобы итерационный процесс $x^{n+1} = \ln(x^n + 2)$ сходился к отрицательному корню?

1.3. Предложите способ вычисления отрицательного корня.

2. Выпишите формулы подходящего способа последовательных приближений для нахождения положительного корня нелинейного уравнения:

$$x - x^3 + 0{,}1 = 0.$$

Оцените необходимое число итераций для достижения точности $\varepsilon = 10^{-8}$ и сравните с тем числом, которое Вы получили при расчетах на ЭВМ.

3. Пусть уравнение $f(x) - g(x) = 0$, где $f(x)$ и $g(x)$ — заданные функции, решается методом Ньютона. Покажите, что приближение x^{n+1} имеет геометрический смысл абсциссы точки пересечения касательных к графикам $y = f(x)$ и $y = g(x)$, проведенным при $x = x^n$.

4. Пронумеруем корни $x(n)$, где $n = 0, 1, 2, \ldots$ уравнения $e^{-x} = \cos x$ в порядке возрастания. Покажите, что при решении уравнения $e^{-x} - \cos x = 0$ методом Ньютона, итерации сходятся

к корню $x(n)$, если за $x^0(n)$ принять число $x^0(n) = \pi n/2$.

5.10. Порядок выполнения работы

1. Решите уравнение $x = \operatorname{tg} x$ методом Ньютона. Как изменится характер сходимости с увеличением номера корня?

2. Покажите, что для решения методом Ньютона следующих уравнений за x^0 можно принять любое положительное число:

1) $e^{-x} = \frac{1}{x}$;

2) $e^{-x} + x - 2 = 0$.

Решите предложенные уравнения численно.

3. Отделите корни следующих уравнений, а затем уточните один из них с помощью итерационного процесса:

1) $\operatorname{arctg}(x-1) + 2x = 0$; 2) $\ln x + (x-1)^3 = 0$;

3) $2 \operatorname{tg} x - \frac{x}{2} + 1 = 0$; 4) $\sqrt{x+1} = \frac{1}{x}$.

4. Уравнение

$$t \cdot x^3 + x^2 - 1 = 0$$

зависит от времени *t*. Предложите итерационный алгоритм отыскания положения этих корней в зависимости от времени *t* за время от $t = 0$ до $t = 1$.

Выясните, при каком значении *t* эволюция отрицательного корня заканчивается его исчезновением.

5. Решите каждое уравнение различными методами с точностью до 10^{-6} и сравните их по эффективности. Объясните полученный результат.

1) $x \ln x = 1$;

2) $\cos^5 x = x^2$;

3) $\ln |x| + (x+1)^3 = 0$;

4) $\operatorname{tg} x = \operatorname{th} x$;

5) $3 \operatorname{arctg} \frac{1}{x} - \frac{1}{2} \operatorname{sh} x = 0$ при $x > 0$.

6. Методом Ньютона найдите корень уравнения $x^7 = 0{,}5$ с точностью до 10^{-6}. Рассмотрите отдельно критерии сходимости по функции и по аргументу. Сравните результат и число итераций, требуемое для сходимости.

5.11. Библиографическая справка

Итерационным методам решения нелинейных уравнений и систем посвящена обширная литература. Для выполнения работы вполне достаточно ознакомиться с основными идеями и теоремами по книгам [1–3]. Более полные сведения о методе можно получить из [8–10, 22], см. также [23] и библиографию в ней.

С итерационными методами решения нелинейных систем тесно связаны различные дискретные отображения. О них лучше прочитать в [24, 25], а на более серьезном уровне в [26].

ПЕРЕОПРЕДЕЛЕННЫЕ СИСТЕМЫ ЛИНЕЙНЫХ УРАВНЕНИЙ. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

6.1. Введение

Работа позволяет исследовать и продемонстрировать особенности нахождения обобщенного решения *переопределенных* систем линейных уравнений. Рассматриваются переопределенные системы линейных уравнений, возникающие, например, при обработке результатов физических экспериментов. Исследуется влияние выбора базиса (*степенные* и *тригонометрические* многочлены, *многочлены Лежандра*), степени многочлена на свойства вычислительного алгоритма и обусловленность возникающей задачи.

6.2. Переопределенная система линейных алгебраических уравнений

Каноническая запись переопределенной системы линейных алгебраических уравнений имеет следующий вид:

$$\begin{aligned} &a_{11}b_1 + \ldots + a_{1s}b_s = f_1, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad n > s. \\ &a_{n1}b_1 + \ldots + a_{ns}b_s = f_n, \end{aligned} \tag{6.1}$$

Введем пространства R^s и R^n состоящие из элементов вида $\mathbf{b} = (b_1, \ldots, b_s)^{\mathrm{T}}$, $\mathbf{f} = (f_1, \ldots, f_n)^{\mathrm{T}}$ и имеющие размерности s и $n > s$ соответственно. Обозначим через **A** прямоугольную матрицу системы (6.1):

$$\mathbf{A} = \begin{Vmatrix} a_{11} & \ldots & a_{1s} \\ a_{21} & \ldots & a_{2s} \\ \ldots & \ldots & \ldots \\ a_{n1} & \ldots & a_{ns} \end{Vmatrix}. \tag{6.2}$$

Тогда систему (6.1) можно записать в виде

$$\mathbf{Ab} = \mathbf{f}, \qquad \mathbf{b} \in \mathrm{R}^s, \qquad \mathbf{f} \in \mathrm{R}^n \tag{6.3}$$

Введем в R^n «основное» скалярное произведение, положив

$$(\mathbf{f}, \mathbf{g})^{(n)} = \sum_{k=1}^{n} f_k g_k. \tag{6.4}$$

Скалярное произведение в R^n можно ввести множеством других способов. Именно, произвольной симметричной и положительно определенной матрице $\mathbf{B} = \mathbf{B}^* > 0$, т. е. $(\mathbf{Bf}, \mathbf{f}) > 0$, для любого вектора $\mathbf{f} \neq \mathbf{0}$, соответствует скалярное умножение

$$(\mathbf{f}, \mathbf{g})_{\mathbf{B}} = (\mathbf{Bf}, \mathbf{g}); \qquad \mathbf{f}, \mathbf{g} \in \mathrm{R}^n. \tag{6.5}$$

Известно, что любое скалярное произведение в пространстве R^n можно записать формулой (6.5), подобрав соответствующий самосопряженный оператор $\mathbf{B} = \mathbf{B}^* > 0$.

Система (6.1), как правило, не имеет классического решения, т. е. не существует такого набора чисел $b_1, \ldots, b_s$, который обращает каждое из *n* уравнений (6.1) в тождество.

Определение. *Фиксируем* $\mathbf{B} = \mathbf{B}^* > 0$, B: $\mathrm{R}^n \to \mathrm{R}^n$. *Введем функцию от* $\mathbf{b} \in \mathrm{R}^s$, *положив*

$$\Phi(\mathbf{b}) = (\mathbf{Ab} - \mathbf{f}, \mathbf{Ab} - \mathbf{f})_{\mathbf{B}}. \tag{6.6}$$

Примем за обобщенное решение системы (6.1) *вектор* $\mathbf{b} \in \mathrm{R}^s$, *придающий наименьшее значение квадратичной форме* (6.6).

Замечание. Выбор $\mathbf{B} = \mathbf{B}^* > 0$ зависит от исследователя. Матрица **B** имеет смысл «весовой» матрицы и выбирается из тех или иных соображений о том, какую цену придать невязке системы (6.1) при заданном $(b_1, b_2, \ldots, b_s)$.

Теорема 1. *Пусть столбцы матрицы* **A** *линейно независимы, т. е. ранг матрицы* **A** *равен s. Тогда существует одно и только одно обобщенное решение* **b** *системы* (6.1). *Обобщенное решение системы* (6.1) *является классическим решением системы уравнений*

$$\mathbf{A}^*\mathbf{B}\mathbf{A}\,\mathbf{b} = \mathbf{A}^*\mathbf{B}\,\mathbf{f}, \qquad (6.7)$$

которая содержит s скалярных уравнений относительно s неизвестных $b_1, b_2, \ldots, b_s$.

В дальнейшем будем иногда использовать обозначение

$$\mathbf{C} = \mathbf{A}^*\mathbf{B}\mathbf{A}.$$

6.3. Геометрический смысл метода наименьших квадратов

Переопределенную систему $\mathbf{Ab} = \mathbf{f}$, где $\mathbf{A} = \| a_{ij} \|$, $1 \le i \le n$, $1 \le j \le s$, $n > s$, можно записать в виде:

$$b_1\mathbf{V}_1 + b_2\mathbf{V}_2 + \ldots + b_s\mathbf{V}_s = \mathbf{f},$$

где $\mathbf{V}_i \in \mathrm{R}^n$ — i-й столбец матрицы $\mathbf{A}$, $\mathbf{f} = (f_1, f_2, \ldots, f_n)^{\mathrm{T}} \in \mathrm{R}^n$, а вектор $\mathbf{b} = (b_1, b_2, \ldots, b_s)^{\mathrm{T}} \in \mathrm{R}^s$.

Требуется найти коэффициенты $b_1, b_2, \ldots, b_s$ линейной комбинации $b_1\mathbf{V}_1 + b_2\mathbf{V}_2 + \ldots + b_s\mathbf{V}_s$ так, чтобы эта линейная комбинация наименее отличалась от **f**:

$$\left\| \mathbf{f} - \sum b_k V_k \right\|_{\mathbf{B}} \Rightarrow \min.$$

Обозначим через $\mathrm{R}^s(\mathbf{V}) \subset \mathrm{R}^n$ подпространство размерности s пространства R^n, состоящее из всевозможных линейных комбинаций векторов $\mathbf{V}_1$, …, $\mathbf{V}_s$.

Пусть $b_1, b_2, \ldots, b_s$ — обобщенное решение переопределенной системы. Тогда линейная комбинация $\sum b_k \mathbf{V}_k$ — ортогональная в смысле скалярного умножения $(\cdot, \cdot)_{\mathbf{B}}$ проекция вектора $\mathbf{f} \in \mathrm{R}^n$ на подпространство $\mathrm{R}^s(\mathbf{V})$, так как любой вектор из $\mathrm{R}^s(\mathbf{V})$ имеет вид:

$$\mathbf{A}\boldsymbol{\delta} = \delta_1\mathbf{V}_1 + \ldots + \delta_1\mathbf{V}_s \in R^s(\mathbf{V}), \qquad \boldsymbol{\delta} \in \mathrm{R}^s.$$

Наименее уклоняется от **f** элемент $\sum b_k\mathbf{V}_k$ подпространства $\mathrm{R}^s(\mathbf{V})$, имеющий вид $\mathbf{Ab}_{\mathbf{B}}$, где $\mathbf{b}_{\mathbf{B}}$ — решение системы (6.7) методом наименьших квадратов (МНК).

В силу $(\mathbf{Ab_B} - \mathbf{f}, \mathbf{A\delta})_{\mathbf{B}} = (\mathbf{B}(\mathbf{Ab_B} - \mathbf{f}), \mathbf{A\delta})^{(n)} = 0$ элемент $\mathbf{f} - \mathbf{Ab_B}$ ортогонален любому элементу $\mathbf{A\delta} \in \mathrm{R}^s(\mathbf{V})$.

Если в пространстве R^s вместо базиса $\mathbf{V}_1, \ldots, \mathbf{V}_s$ выбрать какой-либо другой базис $\mathbf{V}'_1, \ldots, \mathbf{V}'_s$, то система (6.7) заменится системой

$$\mathbf{C}'\mathbf{b}' = \mathbf{f}' \tag{6.8}$$

с матрицей $\mathbf{C}' = \| c'_{ij} \|$, где $c'_{ij} = (\mathbf{V}'_i, \mathbf{V}'_j)_{\mathbf{B}}$, $i, j = 1, 2, \ldots, s$, и правой частью i-я компонента которой $\mathbf{f}'_i = (\mathbf{f}, \mathbf{V}'_i)_{\mathbf{B}}$.

Вместо решения $\mathbf{b_B}$ системы (6.7) получим новое решение $\mathbf{b}'_{\mathbf{B}}$ системы (6.8), но проекция $\mathbf{f}$ на R^s останется прежней.

Если нас интересует проекция заданного вектора $\mathbf{f}$ на заданное подпространство $\mathrm{R}^s \in \mathrm{R}^n$, то естественно стремиться к выбору базиса $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_s$ этого подпространства, по возможности мало отличающегося от ортонормированного. Искомая проекция от выбора базиса в R^s не зависит, а система МНК (6.7) в случае такого базиса будет иметь хорошо обусловленную матрицу.

6.4. Оценка обусловленности матрицы системы МНК

Обусловленность линейной системы (6.7) определяется числом $\mu = \| \mathbf{C} \| \cdot \| \mathbf{C}^{-1} \|$, которое определяет относительную погрешность решения системы в зависимости от погрешности правой части $\mathbf{f}$:

$$\frac{\| \delta\mathbf{x} \|}{\| \mathbf{x} \|} \le \mu \frac{\| \delta\mathbf{f} \|}{\| \mathbf{f} \|},$$

$$\| \mathbf{C} \|_2 = \max_j \sum_{i=1}^{s} | c_{ij} |,$$

$$\| \mathbf{C}^{-1} \|_2 \approx \frac{\| \mathbf{z} \|_2}{\| \mathbf{y} \|_2}, \qquad \| \mathbf{z} \|_2 = \sum_{i=1}^{s} | z_i |,$$

где векторы $\mathbf{y}$ и $\mathbf{z}$ определяются из решения следующих двух систем уравнений:

$$\mathbf{C}^{\mathrm{T}}\mathbf{y} = \mathbf{e}, \qquad \mathbf{C}\mathbf{z} = \mathbf{y},$$

где $\mathbf{C}^{\mathrm{T}}$ — транспонированная матрица, $\mathbf{e}$ — вектор с компонентами ± 1, выбираемые так, чтобы обеспечить максимальный рост $\| \mathbf{y} \|_2$ на этапе *обратной подстановки.*

6.5. Метод Гаусса

Реализованный в программе прямой метод решения системы МНК (6.7) является вариантом метода Гаусса последовательного исключения неизвестных с частичным выбором ведущего элемента (по столбцу). Исключение по методу Гаусса состоит из двух этапов: *прямого хода* и *обратной подстановки.* В прямом ходе на k-м шаге k-е уравнение вычитается из оставшихся с целью исключения k-ого неизвестного. Обратная подстановка состоит в решении последнего уравнения относительно x_n, предпоследнего — относительно x_{n-1} и т. д. до x_1.

6.6. Метод сопряженных градиентов

Это прямой метод, но может применяться и как итерационный. Подробнее см. работу [4] и библиографию к ней. Используется, если матрица линейной системы самосопряженная и положительно определенная. Матрица системы МНК как раз такая (см. п. 6.2). Метод неустойчив, если обусловленность системы достаточно велика. Обозначим матрицу системы МНК через $\mathbf{C}$.

$$\mathbf{s}_1 = \mathbf{r}_1 = \mathbf{C}\mathbf{b}_0 - \mathbf{f},$$

$$\mathbf{r}_n = \mathbf{r}_{n-1} - a_n \mathbf{C}\mathbf{s}_n, \qquad \mathbf{b}_n = \mathbf{b}_{n-1} - a_n \mathbf{s}_n,$$

$$\mathbf{s}_{n+1} = \mathbf{r}_n + \mathbf{g}_n \mathbf{s}_n,$$

$$\mathbf{a}_n = \frac{(\mathbf{r}_{n-1}, \mathbf{r}_{n-1})}{(\mathbf{C}\mathbf{s}_n, \mathbf{s}_n)}, \qquad \mathbf{g}_n = \frac{(\mathbf{r}_n, \mathbf{r}_n)}{(\mathbf{r}_{n-1}, \mathbf{r}_{n-1})}.$$

Существенным преимуществом метода является отсутствие необходимости знания границ спектра. В точной арифметике метод сходится не более чем за n итераций, где n — размерность матрицы.

6.7. Полиномы Лежандра

Многочлены

$$P_0(x)=1, \qquad P_1(x)=x, \qquad P_2(x)=\frac{3x^2-1}{2}, \qquad P_3(x)=\frac{5x^3-3x}{2}, \ \ldots,$$

$$P_{i+1}=\frac{1}{i+1}\left((2i+1)x\,P_i(x)-iP_{i-1}(x)\right), \qquad i=3,\,4,\,\ldots \qquad (6.10)$$

называют *многочленами Лежандра.* Они ортогональны на отрезке $-1\le x\le 1,$ т. е.

$$\int_{-1}^{1} P_k(x)\,P_l(x)\,dx=0, \qquad k\ne l;$$

$$\int_{-1}^{1} P_k^2(x)\,dx=\frac{2}{2k+1}.$$

В программе используются многочлены, ортогональные на отрезке $[a,\,b],$ $a=\min\limits_i x_i,$ $b=\max\limits_i x_i,$ получающиеся из (6.10.) соответствующей заменой переменных; $x_i,$ $i=1,\,\ldots,\,N$ — абсциссы точек исходных данных.

6.8. Порядок выполнения работы

1. В окне «*Данные и их аппроксимация*» маркерами отмечены точки с координатами $(x_k,\,y_k),$ $k=1,\,2,\,\ldots,\,5,$ записанные в файле данных. По умолчанию это файл DATA\LSQ1.DAT.

В этом файле записаны результаты замеров какой-либо зависимости $y(x)$. Расположение точек позволяет приближенно считать, что зависимость $y(x)$ является линейной, т. е. имеет вид $y(x)=b_0+b_1x.$ Числа $b_0,$ b_1 желательно подобрать так, чтобы при $x=x_k$ получались значения $y_k.$ Так как измерений больше, чем неизвестных $b_0,$ $b_1,$ то соответствующая система линейных уравнений будет *переопределенной* и в общем случае не имеет классического решения, поскольку не существует прямой, проходящей сразу через все экспериментальные точки.

Для аппроксимации данной зависимости в меню «*Параметры*/*Выбор системы функций*» выберите строку «*Полиномы* 1, x, x^2, …» и в строке «*количество базисных функций*» установите число 2.

Если считать все измерения равноправными, то весовую матрицу **B** следует выбрать единичной. Для этого в меню «*Параметры*/*Выбор Евклидовой нормы*» выберите строку «*Матрица* **B** *Единичная*».

Если Вы сочтете разумным придать каждому измерению свой вес, то в меню «*Параметры*/*Выбор Евклидовой нормы*» выберите строку *Матрица* **B** *произвольная* и строку «*редактирование матрицы*». Задайте диагональным элементам **B** неравные значения. Запустите задачу на счет и сравните результаты расчетов. Коэффициенты b_0, b_1 можно посмотреть, активизировав меню «*Окна*/*Полная информация*».

Если заранее известен вид зависимости $y = y(x)$, то тот же набор данных можно интерпретировать как результат нескольких измерений для уточнения коэффициентов b_0, b_1.

2. Выберите базис из полиномов Лежандра («*Параметры*/*Выбор системы функций*/*Полиномы Лежандра*») для представления того же набора экспериментальных данных (x_k, y_k), т. е. будем искать зависимость $y(x)$ в виде $y(x) = b_0 P_0(x) + b_1 P_1(x)$. Установите, зависит ли аппроксимирующий полином от выбора базиса (см. п. 6.2).

3. Введите набор данных из файла LSQ2.DAT («*Параметры*/*Данные*/*Ввод данных из файла*»). Он отвечает более сложной, чем в предыдущем задании, зависимости $y(x)$ и для хорошей аппроксимации нужно выбрать полином более высокой степени. Установите число базисных функций 10 и базис-полином 1, x, x^2, … в меню «*Параметры*/*Выбор системы функций*».

Для решения системы МНК в меню «*Метод*» выберите «*Метод сопряженных градиентов*» и установите число итераций 8.

Следующий расчет выполните для базиса из полиномов Лежандра. Сравните полученные результаты. Обратите внимание на оценку числа обусловленности в обоих случаях. Опишите, как изменяются результаты, если уменьшить точность про-

межуточных вычислений. Для этого в меню «*Параметры*» в строке «*Длина мантиссы*» введите вместо установленной по умолчанию (53) цифру 24, а затем 12.
4. Данные в LSQ3.DAT соответствуют замерам зашумленного сигнала

$$y(x) = c_1 \sin 2\pi x + c_2 \sin 4\pi x.$$

Найти c_1, c_2.

Указание. Используйте базис из тригонометрических полиномов («*Параметры*/*Выбор системы функций*/*Полиномы тригонометрические*»).

5. Введите таблицу функции из файла LSQ4.DAT. Приблизьте эту функцию тригонометрическими многочленами степени m и $m + r$, $r > 0$, $m + r < N$, где N — число точек в таблице. Сравните первые m коэффициентов. Объясните полученный результат.
6. Используя мышь, введите произвольный набор данных с экрана терминала («*Параметры*/*Данные*/*Ввод данных с экрана*»). Аппроксимируйте их полиномами различной степени. Проанализируйте влияние выбора базиса, выбора матрицы B, задающей скалярное произведение, способы решения системы МНК *(«Метод Гаусса*, *Метод сопряженных градиентов*»).

6.9. Некоторые рекомендации по работе с программой

Главное меню/*Окна*/*График функции*. При нажатии клавиши *Enter* открывается или закрывается окно с графиком функции («*Данные и их аппроксимация*»), на котором отображаются данные и построенные графики.

Главное меню/*Окна*/*График невязки*. При нажатии клавиши *Enter* открывается или закрывается окно с графиком невязки, на котором отображаются в виде вертикальных прямых разность между значением аппроксимирующего полинома и ординатой точки данных. Цвет прямых соответствует цвету графика полинома в окне «*График функции*».

Главное меню/*Окна*/*Масштабирование*. Здесь можно задать вертикальные и горизонтальные масштабы для окон «*График функции*» и «*График невязки*».

Главное меню/*Окна*/*Очистить экран*. Удаляет все графики из окна «*График функции*» и все графики невязок из окна «*График невязки*», а также информацию из окна «*Полная информа-*

ция».

Главное меню/*Окна*/*Полная информация*. В окне содержится информация обо всех графиках в окне «*График функции*». Вы также можете посмотреть *коэффициенты аппроксимирующего полинома.*

Главное меню/*Окна*/*Полная информация*/*Коэффициенты*. Нажав клавишу *Enter* можно посмотреть коэффициенты аппроксимирующего полинома.

Подготовка файла данных. Для решения своей задачи можно подготовить данные в отдельном файле и записать его в директорию DATA. *Формат данных* должен иметь следующий вид:

1-я строка — комментарий;

2-я строка — если начинается с @, то границы окна с данными: $x_{\min}$, $x_{\max}$, $y_{\min}$, $y_{\max}$, иначе — комментарий;

3-я строка — количество точек в файле, далее данные:

x^1 y^1

x^2 y^2

.....

x^n y^n

Пример:

```
DATA FILE FOR LSQ
файл данных для МНК
3
1. –1e4
–5
–23 10
```

6.10. Библиографическая справка

Изложение теоретических основ использования МНК ведется в соответствии с [1], см. также [8, 11, 27].

ЛАБОРАТОРНАЯ РАБОТА 7

ЧИСЛЕННОЕ РЕШЕНИЕ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ. ЗАДАЧА КОШИ

7.1. Введение

В этой работе Вы познакомитесь с численными методами решения задачи Коши для обыкновенных дифференциальных уравнений (ОДУ):

$$\frac{d\mathbf{u}}{dx} - \mathbf{G}(x, \mathbf{u}) = \mathbf{0}, \qquad x > x_0,$$

$$\mathbf{u}(x_0) = \mathbf{U}_0.$$

В работу включены следующие явные методы численного решения задачи Коши для обыкновенных дифференциальных уравнений:

1) Рунге–Кутты первого порядка точности (метод Эйлера);
2) Рунге–Кутты второго порядка точности;
3) метод второго порядка с центральной разностью;
4) Рунге–Кутты третьего порядка точности (метод Хойна);
5) Рунге–Кутты четвертого порядка точности;
6) экстраполяция Ричардсона второго порядка для метода Рунге–Кутты первого порядка точности.

В процессе работы Вы сможете получать приближенные решения различных обыкновенных дифференциальных уравнений (или систем обыкновенных дифференциальных уравнений) с помощью реализованных численных методов, исследовать их поведение в зависимости от величины шага расчетной сетки, параметров задачи, сравнивать их между собой и с некоторыми точными решениями, получать фазовые портреты.

7.2. Численные методы решения задачи Коши для обыкновенных дифференциальных уравнений

Пусть

$$\mathbf{L}\,\mathbf{u} = \mathbf{f} \tag{7.1}$$

— краткое символьное обозначение исходной дифференциальной задачи.

Пусть, далее, введена в рассмотрение сеточная область W^h и пространство U^h сеточных функций $\mathbf{u}^h$:

$$\mathbf{L}_h\,\mathbf{u}^h = \mathbf{f}^h \tag{7.1a}$$

— соответствующее символьное обозначение разностной задачи (схемы), которой заменяется задача (7.1), и решение которой может быть вычислено с помощью ЭВМ.

Замечание. Решение (7.1а) рассматривается в качестве приближенного решения задачи (7.1) в узлах сетки. Ошибка этого приближения определяется как сеточная функция

$$\delta\mathbf{u}^h = [\mathbf{u}]^h - \mathbf{u}^h,$$

где $[\mathbf{u}]^h$ — значения точного решения в узлах сетки.

Введем в пространстве U^h сеточных функций $\mathbf{u}^h$ какую-либо норму $\|\cdot\|$.

Определение. *Погрешностью метода численного решения задачи* (7.1) *в смысле выбранной нормы принято называть величину* $\delta = \|\,\delta\mathbf{u}^h\,\|$.

Если выполнено условие

$$\delta = \mathrm{O}(h^p),$$

то говорят, что схема имеет *p*-й порядок точности.

Введем в пространстве F^h правых частей $\mathbf{f}^h$ некоторую норму для элементов этого пространства:

$$\|\,\mathbf{f}^h\,\| = \|\,\mathbf{f}^h\,\|_{\mathrm{F}^h}\ .$$

Далее индекс F^h будем иногда опускать.

Из теории известно, что величина δ имеет тот же порядок,

что и *погрешность аппроксимации* $\| \delta \mathbf{f}^h \|$, где

$$\delta \mathbf{f}^h = \mathbf{L}_h [\mathbf{u}]^h - \mathbf{f}^h.$$

Иными словами, $\mathbf{f}^h$ — невязка, характеризующая, насколько нарушаются сеточные уравнения (7.1а) при подстановке в них проекции (ограничения на сетку) решения исходной задачи (7.1).

Замечание 1. Теорема о том, что погрешность δ, т. е. погрешность метода (7.1а) для задачи (7.1), имеет тот же порядок, что и погрешность аппроксимации $\| \delta \mathbf{f}^h \|$, справедлива в предположении, что разностная задача (7.1а) устойчива.

Замечание 2. Вопрос о сходимости метода (7.1а) сводится к исследованию разностной схемы (7.1а) на *аппроксимацию* и *устойчивость*.

7.3. Устойчивость

Устойчивость разностной схемы означает, что решение (7.1а) существует, единственно, непрерывно зависит от входных данных равномерно по *h*.

Для линейных задач последнее требование означает, что существует число $C = \text{const}$, не зависящее от параметра *h* такое, что для любых **f** выполнено

$$\| \mathbf{u}^h \| < C \| \mathbf{f}^h \|.$$

7.4. Дифференциальная задача

Запишем дифференциальную задачу в следующем виде:

$$\frac{d\mathbf{u}}{dx} - \mathbf{G}(x, \mathbf{u}) = \mathbf{0}, \qquad x > x_0,$$
$$\mathbf{u}(x_0) = \mathbf{U}_0. \tag{7.2}$$

7.5. Сеточная область

Без ограничения общности положим $x_0 = 0$. Кроме того, будем рассматривать задачу при значении аргумента $x \in [0,\ 1]$.

Сеточной областью назовем множество узлов $\mathrm{W}^h = \{x_n;\ n = 0, 1, \ldots, N\}$ (рис. 7), $x_n = nh$, где h — шаг сетки, $Nh = 1$. Шаг сетки может быть неравномерным:

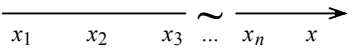


Рис. 7

$$h_n = x_{n+1} - x_n .$$

На сетке определена сеточная функция $u^h = \{u_n;\ n = 0, 1, 2, \ldots, N\}$, здесь u_n — ее компонента, относящаяся к узлу x_n.

7.6. Разностная задача

В качестве простейшего примера разностной схемы, аппроксимирующей дифференциальную задачу Коши, приведем следующую схему (метод Эйлера) для одного уравнения вида (7.2):

$$\frac{u_{n+1} - u_n}{h} = G(x_n, u_n), \qquad n = 0,\ 1,\ 2,\ \ldots,\ N - 1.$$

$$u_0 = U_0 .$$

7.7. Погрешность метода

Величину погрешности для скалярного аналога (7.2) можно определить, например, следующим образом:

$$\delta = \max_n |\,[u]^h - u^h\,| .$$

Напомним, что если $\delta = \mathrm{O}(h^p)$, то говорят, что метод имеет p-й порядок точности.

7.8. Явные методы Рунге–Кутты

Пусть значение u_n приближенного решения в точке x_n уже найдено и требуется вычислить u_{n+1} в точке $x_{n+1} = x_n + h$.

Задаем натуральное m и выписываем выражения вида:

$$k_1 = G(x_n, u_n),$$

$$k_2 = G(x_n + a_1 h,\ u_n + h \cdot b_{21} k_1),$$

$$k_3 = G(x_n + a_2 h,\ u_n + h \cdot (b_{31} k_1 + b_{32} k_2)),$$

.....

$$k_m = G(x_n + a_{m-1} h,\ u_n + h \sum_{i=1}^{m-1} b_{mi} k_i).$$

Затем полагаем

$$L_h\, u^h = \frac{u_{n+1} - u_n}{h} - (p_1 k_1 + \ \dots\ + p_m k_m) = 0, \qquad n = 0, 1, \dots, N-1.$$

$$u_0 = U_0.$$

Неопределенные коэффициенты a_1, a_2, …, a_{m-1}; b_{21}; b_{31}, b_{32}; …; b_{m1}, …, b_{mm-1}; p_1, p_2, …, p_m подбираем так, чтобы получить при заданном m аппроксимацию возможно более высокого порядка.

7.9. Метод Рунге–Кутты первого порядка точности (метод Эйлера)

Частный случай методов Рунге–Кутты при $m = 1$:

$$\frac{u_{n+1} - u_n}{h} - G_n = 0, \qquad n = 0, 1, 2, \dots, N-1.$$

$$u_0 = U_0.$$

7.10. Метод Рунге–Кутты второго порядка точности

Один из возможных методов, реализованных в этой лабораторной работе:

$$\frac{u_{n+1} - u_n}{h} - k_2 = 0, \qquad n = 0, 1, 2, \dots, N-1.$$

$$u_0 = U_0,$$

$$k_1 = G(x_n,\ u_n), \qquad k_2 = G\left(x_n + \frac{h}{2},\ u_n + k_1 \frac{h}{2}\right).$$

В [1] приведено однопараметрическое семейство методов Рунге–Кутты второго порядка.

7.11. Метод Рунге–Кутты третьего порядка точности (метод Хойна)

Наиболее употребительный метод Рунге–Кутты третьего порядка (метод Хойна) имеет вид:

$$\frac{u_{n+1}-u_n}{h}=\frac{k_1+3k_3}{4}, \qquad n=0,\ 1,\ 2,\ \ldots,\ N-1.$$

$$u_0=U_0,$$

$$k_1=G(x_n,\ u_n),$$

$$k_2=G\left(x_n+\frac{h}{3},\ u_n+k_1\frac{h}{3}\right), \quad k_3=G\left(x_n+\frac{2h}{3},\ u_n+k_2\frac{2h}{3}\right).$$

7.12. Метод Рунге–Кутты четвертого порядка точности

Приведем формулы классического метода Рунге–Кутты четвертого порядка:

$$\frac{u_{n+1}-u_n}{h}=\frac{k_1+2k_2+2k_3+k_4}{6}, \qquad n=0,\ 1,\ 2,\ \ldots,\ N-1.$$

$$u_0=U_0,$$

$$k_1=G(x_n,\ u_n),$$

$$k_2=G\left(x_n+\frac{h}{2},\ u_n+k_1\frac{h}{2}\right), \qquad k_3=G\left(x_n+\frac{h}{2},\ u_n+k_2\frac{h}{2}\right),$$

$$k_4=G(x_n+h,\ u_n+hk_3).$$

Замечание. Известно, что существует большое число методов третьего и четвертого порядков [28–30]. В работе приводятся те, которые требуют минимального количества вычислений правой части.

7.13. Экстраполяция Ричардсона

Пусть в точке x известно значение решения $u(x)$. Пусть методом Рунге–Кутты порядка p в результате выполнения численно-

го интегрирования на двух шагах величины h найдено численное значение u в точке $x+2h$, а в результате выполнения одного шага $2h$ получено значение u_{2h} (в той же точке). Тогда выражение

$$u' = u_h + \frac{u_h - u_{2h}}{2^p - 1}$$

аппроксимирует величину $u(x+2h)$ с порядком $p+1$. Другими словами, экстраполяция Ричардсона позволяет увеличивать на единицу точность метода.

7.14. Схема второго порядка с центральной разностью

Приведем пример разностной схемы на трехточечном шаблоне:

$$\frac{u_{n+1} - u_{n-1}}{2h} - G_n = 0, \qquad n = 1, 2, \ldots, N-1.$$

$$u_0 = U_0.$$

$$u_1 = u_0 + hG_0.$$

Заметим, что схемы Рунге–Кутты относятся к *двухточечным*, т. е. для вычисления значения функции в точке x_{n+1} требуется лишь значение в точке x_n. В схеме с центральной разностью требуется знать значение в двух точках x_n и x_{n-1}. Говорят, что дифференциальное уравнение 1-го порядка в этом случае приближено разностным уравнением 2-го порядка.

Формально порядок разностного уравнения определяется тем, какое количество начальных условий для его решения необходимо поставить.

Подробнее о линейных разностных уравнениях разного порядка см. в [31].

7.15. Теоремы об устойчивости методов Рунге–Кутты

Приведем без доказательств три теоремы об устойчивости методов Рунге–Кутты, доказательства см. в [2, 29].

Теорема 1. *Пусть функция G в (7.2) удовлетворяет условиям Липшица по аргументу и с постоянной C*:

$$\| G(x,\ u) - G(x,\ v) \| \le C \| u - v \|$$

(*эта оценка не зависит от сеточного параметра* h). *Пусть также* $CT << 1$. *Тогда метод Рунге–Кутты устойчив и имеет место оценка*

$$\| u_n - v_n \| \le e^{CT} \| u_0 - v_0 \| + 2\varepsilon \frac{e^{CT}}{C}.$$

Здесь ε *— максимальная ошибка округления на данной ЭВМ,* $\{u_n\}$ *— «точное» сеточное решение задачи,* $\{v_n\}$ *— решение возмущенной задачи,* T *— длина отрезка интегрирования.*

Замечание. Данный вывод не зависит от порядка метода Рунге–Кутты. Более тонкие оценки получаются с учетом информации о характере решения [2].

Введем обозначение

$$\mathbf{A} = \frac{1}{2}\left\{ \frac{\partial \mathbf{G}}{\partial \mathbf{u}} + \left(\frac{\partial \mathbf{G}}{\partial \mathbf{u}} \right)^* \right\}.$$

Пусть система (7.2) такова, что для всех $x \in [0,\ T]$ и любого вектора **y** выполнено

$$(\mathbf{Ay}, \mathbf{y}) \le -a\,(\mathbf{y}, \mathbf{y}); \qquad a = \text{const} > 0.$$

(Такие траектории называются *устойчивыми*.)

Теорема 2. *При численном интегрировании устойчивой траектории методом Рунге–Кутты порядка* k *при всех* $x > 0$ *погрешность метода есть* $\mathrm{O}(h^k)$.

Пусть теперь $(\mathbf{Ay}, \mathbf{y}) \le 0$ для любого вектора **y**. Такие траектории называются «*не устойчивыми*» (нейтральными).

Теорема 3. *При численном интегрировании нейтральной системы методом Рунге–Кутты порядка* $k \ge 2$ *точность метода падает на порядок при* $x = \mathrm{O}(1/h)$.

7.16. Контрольные вопросы

1. Рассматривается задача Коши:

$$y' = ay, \qquad a = \text{const},$$

$$y(0) = 1.$$

Ее точное решение есть $y = e^{ax}$.

Исследуйте схемы из п. 7.9, 7.14 данной работы на аппроксимацию на решении данной задачи и на сходимость.

2. Рассматривается задача Коши для системы уравнений:

$$u' = v,$$

$$v' = -u,$$

$$u(0) = v(0) = 1.$$

Система имеет первый интеграл (закон сохранения энергии)

$$\frac{u^2}{2} + \frac{v^2}{2} = 1.$$

Что будет происходить с энергией системы в случае применения явного метода Эйлера из п. 7.5? Будет ли сохраняться энергия при использовании схемы из п. 7.14?

Что произойдет при использовании неявного метода Эйлера

$$\frac{u_{n+1} - u_n}{h} = v_{n+1}; \qquad \frac{v_{n+1} - v_n}{h} = -u_{n+1}.$$

3. Показать, что система из предыдущего вопроса «не устойчивая» (нейтральная).

7.17. Порядок выполнения работы

1. Изучите поведение численных решений задачи Коши для обыкновенного дифференциального уравнения, полученных разными численными методами.

Для одного дифференциального уравнения

$$u_x + Au = 0, \qquad A = \pm 3,$$

$$u(0) = 1$$

на отрезке [0, 2] получите численные решения с помощью всех указанных в меню методов; сравните их между собой и с точным решением (шаг интегрирования $h = 0{,}05$). Получите численные решения для методов Рунге–Кутты первого и четвертого порядка при различных значениях шага интегрирования h.

Проведите численное решение задачи (при $A = 20$) двумя

методами (с центральной точкой и Рунге–Кутты четвертого порядка) с $h = 0{,}1;\ 0{,}06;\ 0{,}01;\ 0{,}001$ на отрезке [0, 3]. Объясните результаты, полученные с помощью метода с центральной точкой.

Точное решение: $x(t) = x(0) \cdot e^{\pm 3t}$.

2. Проведите аналогичные расчеты для системы двух уравнений

$$u' = \upsilon,$$

$$\upsilon' = -u,$$

$$u(0) = \upsilon(0) = 1$$

сравните с точным решением системы ($h = 0{,}01$).

Точное решение: $u(t) = \sin t + \cos t,\ \ \upsilon(t) = \cos t - \sin t.$

3. Получите численное решение «жесткой» системы уравнений

$$u' = 98u + 198\upsilon,$$

$$\upsilon' = -99u - 199\upsilon$$

$$u(0) = \upsilon(0) = 1.$$

и сравните с точным решением. Какой шаг интегрирования необходимо взять, чтобы численное решение было устойчивым? О жестких системах см. [2, 28, 30].

Точное решение: $u(t) = 4e^{-t} - 3e^{-100t}$,

$$\upsilon(t) = -2e^{-t} + 3e^{-100t}.$$

4. Получите численное решение системы двух ОДУ

$$u' = A + u^2\upsilon - (B+1)\,\upsilon, \qquad u(0) = 1,$$

$$\upsilon' = Bu - u^2\upsilon, \qquad \upsilon(0) = 1,$$

$$A = 1,\ \ B \in [1,\ 5]$$

двумя методами: Рунге–Кутты первого и четвертого порядка. Изучите фазовые портреты. Удалось ли Вам получить предельные циклы и бифуркацию Хопфа (при которой предельный цикл вырождается в точку; при этом $B \to A \cdot (A+1)$)?

Эта система — модель Лефевра–Пригожина «брюсселя-

тор». Подробнее о ней см. в [29, 30].

5. Изучите поведение численного решения ОДУ второго порядка (уравнения Ван-дер-Поля):

$$y'' + e\,(y^2 - 1)y' + y = 0,$$

представленного в виде системы двух ОДУ первого порядка

$$x' = z,$$

$$z' = e\,(1 - x^2)z - x, \qquad e > 0,$$

или в представлении Льенара

$$z' = -y,$$

$$y' = z - e\left(\frac{y^3}{3} - y\right); \qquad e > 0,$$

$$x(0) = 2, \qquad z(0) = 0, \qquad 0 < t \le 100$$

в зависимости от изменения параметра e ($0{,}01 < e < 100$).

6. Исследуйте поведение фазовых траекторий для системы ОДУ

$$x' = y,$$

$$y' = x^2 - 1$$

вблизи особых точек (1, 0) и (–1, 0) с помощью двух методов Рунге–Кутты (первого и четвертого порядка точности). Объясните их поведение. Значения $x(0)$ и $y(0)$ варьируйте самостоятельно.

7. Получите траекторию движения спутника вокруг планеты, проведя численное решение задачи двух тел

$$x' = z,$$

$$y' = u,$$

$$z' = -\frac{x}{(x^2 + y^2)^{3/2}},$$

$$u' = -\frac{y}{(x^2 + y^2)^{3/2}},$$

$$x(0) = 0{,}5; \qquad y(0) = z(0) = 0, \qquad u(0) = \sqrt{3} \approx 1{,}73$$

на интервале времени $0 < t \le 20$ двумя методами (Рунге–Кутты первого и второго порядка точности). Исследуйте зависимость численного решения от шага интегрирования.

8*. Получите численное решение ОДУ с особенностью

$$u' = \frac{1}{2\sqrt{x}} + u^2(x), \qquad u(0) = 0.$$

9. Методами разных порядков аппроксимации численно решить систему Лоренца:

$$x' = -\sigma(x - y),$$

$$y' = -xz + rx - y,$$

$$z' = xy - bz,$$

$$x(0) = y(0) = z(0) = 1$$

при $b = 8/3$, $\sigma = 10$, $r = 28$. Считаем, что $0 < t \le 50$. Объяснить полученные результаты.

Указание. О системе Лоренца см. [29].

7.18. Библиографическая справка

Численному решению ОДУ посвящена обширная литература. Мы можем рекомендовать для первоначального ознакомления книги [2, 3, 4, 31]. Современные методы численного решения описаны в [5, 29]. Теории численных методов решения жестких систем методами Рунге–Кутты и исследованию их на устойчивость посвящена книга [28], а обзор современных методов численного решения жестких систем см. в [5, 30]. Конечно, это лишь малая часть литературы по численному решению задач Коши, и заинтересованный читатель сможет пополнить свои знания, используя приведенную в упомянутых книгах библиографию.

ЛАБОРАТОРНАЯ РАБОТА 8

ЧИСЛЕННОЕ РЕШЕНИЕ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ. КРАЕВАЯ ЗАДАЧА

8.1. Введение

Эта работа знакомит с различными методами решения линейных и нелинейных краевых задач. Отличие краевой задачи от задачи Коши (задачи с начальными условиями) состоит в том, что решение дифференциального уравнения должно удовлетворять граничным условиям, связывающим значения искомой функции более чем в одной точке.

Простейшим представителем краевой задачи является двухточечная граничная задача, для которой граничные условия задаются в двух точках, как правило, на концах интервала, на котором ищется решение. Двухточечные граничные задачи встречаются во всех областях науки и техники. На примерах таких задач и будет рассмотрено применение методов, обсуждаемых в настоящей работе. В случае задания краевых условий в более общем виде использование этих методов не представит принципиальных затруднений.

8.2. Пример краевой задачи

Примером двухточечной краевой задачи является задача:

$$y'' = f(x, y, y'), \qquad 0 < x \le 1,$$
$$y(0) = Y_0, \qquad y(1) = Y_1. \tag{8.1}$$

с граничными условиями на обоих концах отрезка $0 \le x \le 1$, на котором надо найти решение $y = y(x)$. На этом примере мы схематически изложим некоторые способы численного решения краевых задач.

Если функция $f(x, y, y')$ в (8.1) линейна по аргументам y и

y', то мы имеем линейную краевую задачу, иначе — нелинейную краевую задачу.

8.3. Линейная краевая задача

Рассмотрим частную, но довольно распространенную краевую задачу следующего вида:

$$\begin{aligned} &Ly = y'' - p(x)\, y = f(x), \qquad 0 < x \le 1, \\ &y(0) = Y_0, \qquad y(1) = Y_1. \end{aligned} \tag{8.2}$$

Для этой задачи проиллюстрируем два способа решения: один основан на идее численного построения общего решения линейного дифференциального уравнения, другой (конечно-разностный) сводит исходную дифференциальную краевую задачу к системе линейных алгебраических уравнений, решение которых находится методом прогонки.

8.4. Метод численного построения общего решения

Для нахождения решения краевой задачи (8.2) можно численно построить решение дифференциального уравнения, представимое в виде

$$y(x) = C_1 y_1(x) + C_2 y_2(x) + y_0(x),$$

где $y_0(x)$ — какое-либо решение неоднородного уравнения

$$y'' - p(x)\, y = f(x),$$

а $y_1(x)$ и $y_2(x)$ — два любые линейно независимые решения однородного уравнения $y'' - p(x)\, y = 0$. Постоянные C_1 и C_2 находятся из граничных условий задачи (8.2).

Так как решения $y_0(x)$, $y_1(x)$, $y_2(x)$ произвольны, то их можно построить различными способами. Например, можно задать какие-то начальные условия и решить одну задачу Коши для неоднородного и две задачи Коши для однородного уравнений. Эти условия, в частности, могут быть такими:

$y_0(0) = 0, \quad y_0'(0) = 0$ — для неоднородного уравнения;

$y_1(0) = 1, \quad y_1'(0) = 0;$

$y_2(0) = 0, \qquad y_2'(0) = 1$ — для однородного уравнения.

Однако при реализации этого способа, например, в случае $p(x) >> 1$ для рассматриваемого уравнения могут возникнуть трудности, связанные с неустойчивостью задачи Коши. В этом случае можно попытаться построить $y_0(x)$, $y_1(x)$, $y_2(x)$ с помощью решения одной краевой задачи для неоднородного уравнения и двух краевых задач для однородного уравнения. Краевые условия для этих задач могут быть, например, следующими:

$y_0(0) = 0, \qquad y_0'(1) = 0$ — для неоднородного уравнения;

$y_1(0) = 1, \qquad y_1'(1) = 0;$

$y_2(0) = 0, \qquad y_2'(1) = 1$ — для однородного уравнения.

Эти задачи могут быть решены методом прогонки. Условия устойчивости метода прогонки при $p(x) > 0$, как легко проверить, выполнены. Этот подход может оказаться полезным, если краевые условия таковы, что для исходной задачи (8.2) метод прогонки применен быть не может.

Отметим, что с учетом специфики краевых условий исходной задачи можно строить общее решение вида

$$y(x) = y_0(x) + C y_1(x),$$

где $y_0(x)$ — некоторое решение неоднородного уравнения, а $y_1(x)$ — некоторое решение однородного уравнения.

8.5. Конечно-разностный метод (прогонки)

При нахождении решения линейной краевой задачи:

$$y'' - p(x)\, y = f(x), \qquad 0 < x \le 1,$$

$$y(0) = Y_0, \qquad y(1) = Y_1.$$

для $p(x) >> 1$ методом построения общего решения, если оно находится с помощью решения задач Коши, могут возникнуть трудности, связанные с вычислительной неустойчивостью задачи Коши.

Для решения поставленной задачи можно воспользоваться разностной схемой:

$$\frac{y_{m+1} - 2y_m + y_{m-1}}{h^2} - p(x_m)\, y_m = f(x_m),$$

$$0 < m < M, \qquad Mh = 1, \qquad y_0 = Y_0, \qquad y_M = Y_1$$

и решить разностную задачу методом прогонки. Условия применимости метода прогонки при $p(x) > 0$, как легко проверить, выполнены. Подробнее о методе прогонки см. в [1–4, 17, 31]. В [17] рассмотрены различные варианты метода прогонки.

8.6. Нелинейная краевая задача

Краевая задача

$$y'' = f(x, y, y'), \qquad 0 < x \le 1,$$
$$y(0) = Y_0, \qquad y(1) = Y_1. \tag{8.3}$$

является нелинейной краевой задачей, если функция $f(x, y, y')$ нелинейна хотя бы по одному из аргументов y или y'.

В настоящей работе реализованы два способа решения нелинейных краевых задач: *метод стрельбы* и *метод линеаризации* (метод Ньютона), который сводит решение нелинейной краевой задачи к решению серии линейных краевых задач.

8.7. Метод стрельбы

Метод стрельбы для решения краевой задачи (8.3) базируется на том, что имеются удобные способы численного решения задачи Коши, т. е. задачи следующего вида

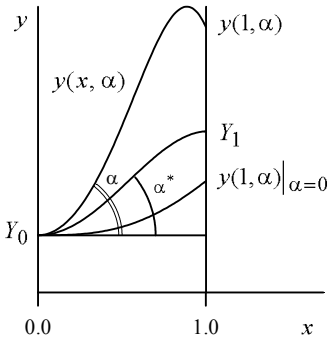


Рис. 8

$$y'' = f(x, y, y'), \quad 0 < x \le 1,$$
$$y(0) = Y_0, \tag{8.4}$$
$$y'(0) = \operatorname{tg} \alpha,$$

где Y_0 — ордината точки $(0, Y_0)$, из которой выходит интегральная кривая; α — угол на-

клона интегральной кривой к оси x при выходе из точки $(0, Y_0)$ (рис. 8). При фиксированном Y_0 решение задачи (8.4) имеет вид $y = y(x, \alpha)$. При $x = 1$ решение $y(x, \alpha)$ зависит только от α:

$$y(x, \alpha)\big|_{x=1} = y(1, \alpha).$$

Используя указанное замечание о решении задачи Коши (8.4), можно задачу (8.3) переформулировать следующим образом: найти такой угол $\alpha = \alpha^*$, при котором интегральная кривая, выходящая из точки $(0, Y_0)$ под углом α к оси абсцисс, попадет в точку $(1, Y_1)$:

$$y(1, \alpha) = Y_1. \qquad (8.5)$$

Решение задачи (8.4) при этом $\alpha = \alpha^*$ совпадает с искомым решением задачи (8.3). Таким образом, дело сводится к решению уравнения (8.5) (рис. 9). Уравнение (8.5) — это уравнение вида

$$F(\alpha) = 0,$$

где $F(\alpha) = y(1, \alpha) - Y_1$.

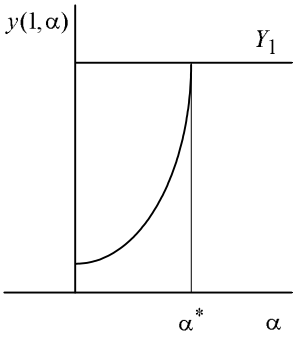


Рис. 9

Оно отличается от привычных уравнений лишь тем, что функция $F(\alpha)$ задана не аналитическим выражением, а с помощью алгоритма численного решения задачи (8.4).

Для решения уравнения (8.5) можно использовать любой метод, пригодный для уточнения корней нелинейного уравнения, например, метод деления отрезка пополам, метод Ньютона (касательных) и др. Метод Ньютона здесь предпочтительнее (если имеется достаточно хорошее начальное приближение) из-за высокой стоимости вычисления одного значения функции $F(\alpha)$ (нужно решить задачу Коши (8.4) с данным α).

Метод стрельбы, сводящий решение краевой задачи (8.3) к вычислению решений задачи Коши (8.4), хорошо работает в том случае, если решение $y(x, \alpha)$ «не слишком сильно» зависит от α. В противном случае он становится вычислительно неустой-

чивым, даже если решение задачи (8.3) зависит от входных данных «умеренно».

При решении уравнений $F(\alpha) = 0,$ методом деления отрезка пополам, мы задаем α_0 и α_1 так, чтобы разности $y(1, \alpha_0) - Y_1$ и $y(1, \alpha_1) - Y_1$ имели разные знаки. Затем полагаем

$$\alpha_2 = \frac{\alpha_0 + \alpha_1}{2}.$$

Вычисляем $y(1, \alpha_2)$. Затем вычисляем α_3 по одной из формул:

$$\alpha_3 = \frac{\alpha_1 + \alpha_2}{2} \qquad \text{или} \qquad \frac{\alpha_0 + \alpha_2}{2}$$

в зависимости от того, имеют ли разности $y(1, \alpha_2) - Y_1$ и $y(1, \alpha_1) - Y_1$ соответственно разные или одинаковые знаки. Затем вычисляем $y(1, \alpha_3)$. Процесс продолжаем до тех пор, пока не будет достигнута требуемая точность $\left| y(1, \alpha_n) - Y_1 \right| < \varepsilon.$

В случае использования для решения уравнения $F(\alpha) = 0$ метода Ньютона задаем α_0, а затем последующие α_n вычисляем по рекуррентной формуле

$$\alpha_{n+1} = \alpha_n - \frac{F(\alpha_n)}{F'(\alpha_n)}, \qquad n = 0, 1, \ldots$$

Производная $F'(\alpha_n)$ может быть вычислена по одной из формул численного дифференцирования, например, первого порядка аппроксимации:

$$F'(\alpha_n) = \frac{F(\alpha_n + h) - F(\alpha_n)}{h}.$$

8.8. Вычислительная неустойчивость задачи Коши

Поясним причину возникновения вычислительной неустойчивости на примере следующей линейной краевой задачи:

$$y'' - p^2 \, y = 0, \qquad 0 < x \le 1,$$

$$y(0) = Y_0, \qquad y(1) = Y_1. \tag{8.6}$$

при постоянном p^2. Выпишем решение этой задачи:

$$y(x) = \frac{e^{-px} - e^{-p(2-x)}}{1 - e^{-2p}} Y_0 + \frac{e^{-p(1-x)} - e^{-p(1+x)}}{1 - e^{-2p}} Y_1.$$

Коэффициенты при Y_0 и Y_1 с ростом p остаются ограниченными на отрезке $0 \le x \le 1$ функциями; при всех $p > 0$ они не превосходят единицу. Поэтому небольшие ошибки при задании Y_0 и Y_1 ведут к столь же небольшим погрешностям в решении, т. е. краевая задача является «хорошей».

Рассмотрим теперь задачу Коши:

$$y'' - p^2\, y = 0, \qquad 0 < x \le 1,$$
$$y(0) = Y_0, \qquad y(0) = \operatorname{tg} \alpha. \tag{8.7}$$

Ее решение имеет вид:

$$y(x) = \frac{pY_0 + \operatorname{tg} \alpha}{2p} e^{px} + \frac{pY_0 - \operatorname{tg} \alpha}{2p} e^{-px}.$$

Если при задании $\operatorname{tg} \alpha$ допущена погрешность ε, то значение решения при $x = 1$ получит приращение

$$\Delta\, y(1) = \frac{\varepsilon}{2p} e^{p} - \frac{\varepsilon}{2p} e^{-p}. \tag{8.8}$$

При больших p вычитаемое в равенстве (8.8) пренебрежимо мало, но коэффициент в первом слагаемом $e^p / 2p$ становится большим. Поэтому метод стрельбы при решении задачи (8.6), будучи формально приемлемой процедурой, при больших p становится практически непригодным. Подробнее о возникновении неустойчивостей см. [1, 2].

8.9. Метод линеаризации (метод Ньютона)

Метод Ньютона сводит решение нелинейной краевой задачи к решению серии линейных краевых задач и состоит в следующем.

Пусть для нелинейной краевой задачи (8.3) известна функция $y_0(x)$, удовлетворяющая граничным условиям и грубо при-

ближенно равная искомому $y(x)$. Положим

$$y(x) = y_0(x) + \upsilon(x), \tag{8.9}$$

где $\upsilon(x)$ — поправка к нулевому приближению $y_0(x)$. Подставим (8.9) в уравнение (8.8) и линеаризуем задачу, используя следующие равенства:

$$y''(x) = y_0''(x) + \upsilon''(x),$$

$$f(x, y_0 + \upsilon, y_0' + \upsilon') = f(x, y_0, y_0') + \\ + \frac{\partial f(x, y_0, y_0')}{\partial y}\upsilon + \frac{\partial f(x, y_0, y_0')}{\partial y'}\upsilon' + \mathrm{O}(\upsilon^2 + |\upsilon'|^2).$$

Отбрасывая остаточный член $\mathrm{O}(\upsilon^2 + |\upsilon'|^2)$, получим линейную краевую задачу для нахождения поправки $\tilde{\upsilon}(x)$:

$$\tilde{\upsilon}'' = p(x)\,\tilde{\upsilon}' + g(x)\,\tilde{\upsilon} + r(x), \\ \tilde{\upsilon}(0) = 0, \qquad \tilde{\upsilon}(1) = 0, \tag{8.10}$$

где

$$p(x) = \frac{\partial f(x, y_0, y_0')}{\partial y'}, \qquad q(x) = \frac{\partial f(x, y_0, y_0')}{\partial y},$$

$$r(x) = f(x, y_0, y_0') - y_0''.$$

Решая линейную краевую задачу (8.10) каким-либо численным методом найдем поправку $\tilde{\upsilon}$ и примем за первое приближение

$$y_1(x) \equiv y_0(x) + \tilde{\upsilon}.$$

Аналогично, зная приближение $y_1(x)$, положим $y(x) = y_1(x) + \tilde{\upsilon}_1$ и найдем следующее приближение. Продолжая процесс до тех пор, пока не будут выполнены неравенства

$$\max|\tilde{\upsilon}(x)| \le \varepsilon, \qquad x \le [0,\ 1],$$

где ε — требуемая точность, найдем приближенное решение исходной нелинейной задачи.

8.10. Порядок выполнения работы

1. Начните выполнение работы с темы «*Линейная краевая задача*». Выбрав с помощью меню один из методов решения линейной краевой задачи, перейдите к пункту меню «*Параметры*». Наберите следующую краевую задачу:

$$y'' - py = -p, \qquad 0 < x \le 1,$$

$$y(0) = 1, \qquad y(1) = 1.$$

для $p = \text{const} > 0$. Решением этой задачи является функция $y \equiv 1$. Установите значение шага сетки $h = 0{,}05$.

1.1. Найдите решение этой задачи методом построения общего решения и методом прогонки для разных p, начиная с умеренных значений и увеличивая их до величины порядка 1200. Сравните получаемые решения с точным и объясните наблюдаемые эффекты. Попытайтесь найти решение этой же задачи методом стрельбы. Проанализируйте, как влияет при разных p точность задания недостающего начального условия на левом конце интервала на успешное решение задачи методом стрельбы.

1.2. Объясните полученные результаты. Замените левое краевое условие (положите, например, $y'(0) = 0$) и посмотрите, как изменится характер решения.

1.3. Выполните п. 1.1, 1.2 для задачи:

$$y'' + py = p, \qquad 0 < x \le 1,$$

$$y(1) = 1, \qquad y'(0) = 0.$$

Ее точное решение $y \equiv 1$. Объясните полученные результаты. Найдите условие устойчивости метода прогонки для данной задачи.

2. Получите численное решение следующих нелинейных краевых задач:

2.1. $y'' + px\cos y = 0, \qquad 0 < x \le 1,$

$$y'(0) = 0, \qquad y(1) = 0, \qquad p = 1, 4, 7, 25, 50, 100;$$

2.2. $y'' + \dfrac{0{,}5}{1 - 0{,}5y}\, y'^2 = 0, \qquad 0 < x \le 1,$

$$y(0) = y_0, \quad y(1) = 0,$$

$y_0 = 0{,}25;\ 0{,}5;\ 1;\ 1{,}5;\ 1{,}8;\ 1{,}9;\ 1{,}95;$

2.3. $y'' + \sin y = 0, \qquad 0 < x \le x_k,$

$y(0) = 0, \qquad y(x_k) = \pi,$

$x_k = 0{,}5;\ 1;\ 2;\ 4;\ 6.$

3. Рассмотрите следующие краевые задачи:

3.1. $y'' = e^y, \qquad 0 < x \le 1,$

$y(0) = 1, \qquad y(1) = a;$

3.2. $y'' = -e^y, \qquad 0 < x \le 1,$

$y(0) = 1, \qquad y(1) = a.$

Параметр a меняется от 0 до 2. Что при этом происходит с решением задач? Почему в задаче 3.2 при значениях $a > 1{,}4999\ldots$ не работает метод линеаризации?

Замечание. Задача 3 подробно рассмотрена в [29].

4. Рассмотреть две сингулярно-возмущенные задачи (с малым параметром при старших производных):

$$\varepsilon y'' = y(y^2 - 1), \qquad -1 < x \le 1,$$

$$y(-1) = y(1) = \sqrt{2}$$

и

$$\varepsilon y'' = -y(y^2 - 1), \qquad -1 < x \le 1,$$

$$y(-1) = y(1) = \sqrt{2}.$$

Считаем $\varepsilon = 10^{-3}$. Какие численные методы позволят получить решение каждой из этих задач? Почему?

8.11. Библиографическая справка

Более детальную теоретическую справку о методах решения краевых задач можно получить, используя книги [1–4, 7, 27, 32]. Подробнее о различных вариантах метода прогонки см. в [17].

ЛАБОРАТОРНАЯ РАБОТА 9

ЧИСЛЕННОЕ РЕШЕНИЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ ГИПЕРБОЛИЧЕСКОГО ТИПА. УРАВНЕНИЕ ПЕРЕНОСА

9.1. Введение

В этой работе Вы познакомитесь с численными методами решения дифференциальных уравнений в частных производных гиперболического типа на примере одномерного линейного уравнения переноса, для которого рассматривается краевая задача:

$$u_t + u_x = f(t, x), \qquad t > 0, \qquad 0 \le x < \infty, \tag{9.1}$$

$$u(0, x) = \varphi(x), \qquad 0 \le x < \infty,$$

$$u(t, 0) = \psi(t), \qquad t > 0.$$

На предложенных примерах Вы изучите характерные черты поведения численных решений этого уравнения в зависимости от входных данных задачи, выбора разностного метода, сеточных параметров (шагов по времени и координате).

Полученные приближенные (численные) решения однородного уравнения Вы сможете сравнить с аналитическим.

9.2. Дифференциальная задача

В работе рассматривается следующая дифференциальная задача:

$$u_t + a u_x = f, \qquad 0 \le t \le T, \qquad 0 \le x \le X, \tag{9.1a}$$

$$u(0, x) = \varphi(x), \qquad 0 \le x \le X,$$

$$u(t, 0) = \psi(t), \qquad t > 0.$$

9.3. Сеточная область

Для рассмотренной задачи введена равномерная сетка

$$\mathrm{W}^h = [(t_p, x_m)], \qquad p = 0, 1, \ldots, P; \quad m = 0, 1, \ldots, M;$$

в узлах которой определена сеточная функция u^h :

$$u^h = [u_m^p]; \qquad p = 0, 1, \ldots, P; \quad m = 0, 1, \ldots, M;$$

где u_m^p — компонента сеточной функции, относящаяся к узлу (t_p, x_m), $t_p = p\tau$, τ — шаг по времени, $P\tau = T$, $x_m = mh$, h — шаг по координате, $Mh = X$.

9.4. Пример разностной задачи (схемы)

Для рассмотренной дифференциальной задачи одна из возможных разностных схем («явный правый уголок») имеет вид:

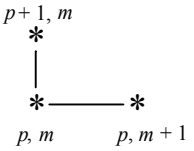
$$\frac{u_m^{p+1} - u_m^p}{\tau} + a\,\frac{u_{m+1}^p - u_m^p}{h} = f_m^p,$$

$$p = 0, 1, \ldots, P-1; \qquad m = 0, 1, \ldots, M-1;$$

$$u_m^0 = \varphi_m, \qquad m = 0, 1, \ldots, M;$$

$$u_0^p = \psi^p, \qquad p = 1, 2, \ldots, P.$$

9.5. Шаблон разностной схемы



Рассмотренная разностная схема при заданных m и p связывает значения решения в трех точках сетки, которые образуют конфигурацию «правый уголок», называемую *шаблоном* этой схемы.

9.6. Погрешность метода

Введем в пространстве решений норму, положив например,

$$\| u^h \| = \sup_{p,\,m} | u_m^p |, \qquad m = 0, 1, \ldots, M; \quad p = 0, 1, \ldots, P.$$

Тогда погрешность выражается формулой

$$\delta = \max_{p,\, m} |[u]_m^p - u_m^p| ,$$

где под $[u]_m^p$ подразумевается значение точного решения исходной дифференциальной задачи в узле (t_p, x_m).

9.7. Невязка

Для рассмотренной разностной задачи соотношения при $p = 0$ (начальные данные) при подстановке в них решения (9.1), удовлетворяются точно, а соотношения, которые получены из (9.1) заменой производных по формулам численного дифференцирования в произвольном (p, m) узле, дают компоненту ошибки аппроксимации:

$$\delta f_m^p = \frac{[u]_m^{p+1} - [u]_m^p}{\tau} + a\,\frac{[u]_{m+1}^p - [u]_m^p}{h} - f_m^p .$$

Представление о величине δf_m^p легко получить, если задать опорную точку и представить значения $[u(t, x)]$, входящие в выражение для δf_m^p, в виде разложения в ряды Тейлора относительно этой точки.

Выбирая в качестве опорной точку (t_p, x_m), получим:

$$\delta f_m^p = \frac{[u]_m^p + \tau[u'_t]_m^p + 0{,}5\,\tau^2\,[u''_{tt}(mh,\, p\tau + \xi\tau)] - [u]_m^p}{\tau} +$$

$$+ a([u]_m^p + h[u'_x]_m^p + \frac{h^2}{2}[u''_{xx}(mh + \zeta h,\, p\tau)] - [u]_m^p) - f_m^p =$$

$$= ([u'_t]_m^p + a[u'_x]_m^p - f_m^p) + \tau u''_{tt}(mh,\, p\tau + \xi\tau) + \frac{ah}{2}\,u''_{xx}(mh + \zeta h,\, p\tau),$$

где $0 \le \xi,\ \ \zeta \le 1$.

Для решения (9.1) выражение в первых скобках в последнем равенстве равно нулю. Поскольку выкладки проводились

для произвольного узла (t_p, x_m), получаем (в предположении ограниченности вторых производных u''_{tt} и u''_{xx}):

$$\| \delta f^h \| = \mathrm{O}(\tau + h).$$

Если τ и h имеют одинаковый порядок, то $\| \delta f^h \| = \mathrm{O}(h)$.

9.8. Спектральный признак устойчивости

Для однородной разностной задачи ищем решение в виде

$$u_m^p = \lambda^p \, e^{im\omega}.$$

Подставляя его в разностные уравнения, получим

$$\frac{\lambda - 1}{\tau} + a \, \frac{e^{i\omega h} - 1}{h} = 0,$$

или

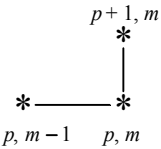
$$\lambda = 1 + \frac{a\tau}{h} - a\tau \, \frac{e^{i\omega h}}{2h}.$$

В силу произвольности ω отсюда получаем, что $|\lambda| \le 1$ при всех ω тогда и только тогда, когда $-1 \le a\tau / h \le 0$, т. е. схема устойчива только, если в (9.1) $a < 0$ и $\{\tau, h\}$ выбраны так, что $|a\tau| / h \le 1$.

Подробнее о *спектральном признаке устойчивости* см. Приложение 1.

9.9. Явный левый уголок

Сеточный шаблон:



Разностная схема (далее полагаем $a = 1$):

$$\frac{u_m^{p+1} - u_m^p}{\tau} + \frac{u_m^p - u_{m-1}^p}{h} = f_m^p, \quad p = 0, 1, \ldots, P-1; \qquad m = 1, \ldots, M;$$

$$u_m^0 = \varphi_m, \qquad m = 0, 1, \ldots, M;$$

$$u_0^p = \psi^p, \qquad p = 1, 2, \ldots, P.$$

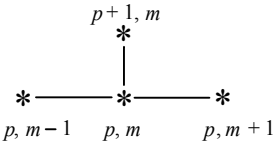
Значение сеточной функции на верхнем временном слое $p + 1$ рассчитывается по ее значениям на нижнем слое p:

$$u_m^{p+1} = u_m^p - \frac{\tau}{h}(u_m^p - u_{m-1}^p) + \tau f_m^p.$$

Порядок аппроксимации: $O(\tau + h)$.
Схема устойчива при $\tau / h \le 1$.

9.10. Явная четырехточечная схема

Сеточный шаблон:



Разностная схема:

$$\frac{u_m^{p+1} - u_m^p}{\tau} + \frac{u_{m+1}^p - u_{m-1}^p}{2h} - \tau q \frac{u_{m-1}^p - 2u_m^p + u_{m+1}^p}{h^2} = f_m^p,$$

$p = 0, 1, \ldots, P - 1; \qquad m = 1, 2, \ldots, M - 1;$

где q — коэффициент искусственной (схемной) вязкости;

$$u_m^0 = \varphi_m, \qquad m = 0, 1, \ldots, M;$$

$$u_0^p = \psi^p, \qquad p = 1, 2, \ldots, P.$$

Значение сеточной функции на верхнем временном слое $p + 1$ при $m = 1, 2, ..., M - 1$ рассчитывается по ее значениям на нижнем слое p:

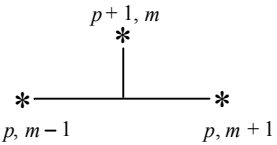
$$u_m^{p+1} = u_m^p - \frac{\tau}{2h}(u_{m+1}^p - u_{m-1}^p) + \\ + \frac{\tau^2 q}{h^2}(u_{m-1}^p - 2u_m^p + u_{m+1}^p) + \tau f_m^p.$$

Порядок аппроксимации ($q = 1/2$ и $f = 0$): $O(\tau^2 + h^2)$.

Схема (при $q = 1/2$) устойчива при $\tau / h \le 1$.

9.11. Явная центральная трехточечная схема

Сеточный шаблон:



Разностная схема:

$$\frac{u_m^{p+1} - 0{,}5(u_{m+1}^p + u_{m-1}^p)}{\tau} + \frac{u_{m+1}^p - u_{m-1}^p}{2h} = f_m^p,$$

$p = 0, 1, \ldots, P-1;$ $\quad m = 1, 2, \ldots, M-1;$

$u_m^0 = \varphi_m,$ $\quad m = 0, 1, \ldots, M;$

$u_0^p = \psi^p,$ $\quad p = 1, 2, \ldots, P.$

Значение сеточной функции на верхнем временном слое $p + 1$ находится по ее значениям на нижнем слое p (значения сеточной функции в точках x_M рассчитываются по схеме «явный левый уголок»):

$$u_m^{p+1} = u_m^p - \frac{\tau}{h}(u_m^p - u_{m-1}^p) + \tau f_m^p.$$

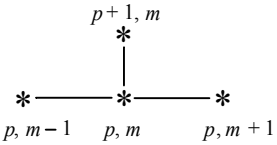
Порядок аппроксимации:

$$O\left(\frac{h^2}{\tau} + \tau + h^2\right).$$

Схема устойчива при $\tau / h \le 1$.

9.12. Гибридная схема (схема Федоренко)

Сеточный шаблон:



Разностная схема:

$$\frac{u_m^{p+1}-u_m^p}{\tau}+\frac{u_m^p-u_{m-1}^p}{h}+\gamma\left(\frac{\tau}{h}-\frac{\tau^2}{h^2}\right)\frac{u_{m-1}^p-2u_m^p+u_{m+1}^p}{2\tau}=f_m^p,$$

$$p = 0, 1, \ldots, P-1; \qquad m = 1, 2, \ldots, M-1;$$

$$u_m^0 = \varphi_m, \qquad m = 0, 1, \ldots, M;$$

$$u_0^p = \psi^p, \qquad p = 1, 2, \ldots, P.$$

Здесь $\gamma = 1$ при $|u_{m-1}^p - 2u_m^p + u_{m+1}^p| \le \lambda |u_m^p - u_{m-1}^p|$ и $\gamma = 0$ в противном случае (λ — численно подбираемый параметр гибридной схемы).

Гибридные схемы используются при расчетах процессов, имеющих особенности разрывного характера (или большие градиенты искомых функций).

Вблизи областей с большими градиентами искомого решения используется схема первого порядка аппроксимации, обладающая сглаживающими свойствами (см., например, поведение численного решения, полученного по схеме «явный левый уголок» при начальном условии третьего типа). В «гладких» областях расчет ведется по немонотонной схеме второго порядка аппроксимации (см., например, поведение численного решения, полученного по явной четырехточечной схеме при том же начальном условии).

Значение сеточной функции на верхнем временном слое $p + 1$ рассчитывается по ее значениям на нижнем слое p:

$$u_m^{p+1} = u_m^p - \frac{\tau}{h}(u_m^p - u_{m-1}^p) -$$

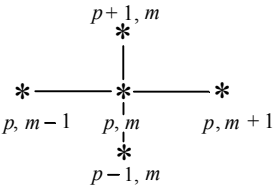
$$-\gamma\left(\frac{\tau}{h}-\frac{\tau^2}{h^2}\right)\frac{u_{m-1}^p-2u_m^p+u_{m+1}^p}{2}+\tau f_m^p.$$

Порядок аппроксимации: $O(\tau + h)$.

Схема устойчива при $\tau / h \le 1$.

9.13. Схема «Чехарда»

Сеточный шаблон:



Разностная схема:

$$\frac{u_m^{p+1}-u_m^{p-1}}{2\tau}+\frac{u_{m+1}^p-u_{m-1}^p}{2h}=f_m^p,$$

$p = 1, \ldots, P-1;$ $\qquad m = 1, 2, \ldots, M-1;$

$u_m^0 = \varphi_m,$ $\qquad m = 0, 1, \ldots, M;$

$u_0^p = \psi^p,$ $\qquad p = 1, 2, \ldots, P.$

Алгоритм численного решения задачи: значения сеточной функции на верхнем временном слое $p + 1$ рассчитываются по ее значениям на двух нижних слоях:

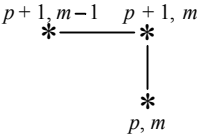
$$u_m^{p+1} = u_m^{p-1} - \frac{\tau}{h}(u_{m+1}^p - u_{m-1}^p) + 2\tau f_m^p.$$

Порядок аппроксимации: $O(\tau^2 + h^2)$.

Схема устойчива при $\tau / h \le 1$.

9.14. Неявный левый уголок

Сеточный шаблон:



Разностная схема:

$$\frac{u_m^{p+1}-u_m^p}{\tau}+\frac{u_m^{p+1}-u_{m-1}^{p+1}}{h}=f_m^{p+1},\quad p=0,\ldots,P-1;\qquad m=1,\ldots,M;$$

$$u_m^0=\varphi_m,\qquad m=0,1,\ldots,M;$$

$$u_0^p=\psi^p,\qquad p=1,2,\ldots,P.$$

Значение сеточной функции на верхнем временном слое $p+1$ рассчитывается по ее значениям в точках верхнего и нижнего p временных слоев:

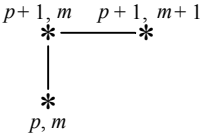
$$u_m^{p+1}=\frac{u_m^p+\tau u_{m-1}^{p+1}\big/h+\tau f_m^{p+1}}{1+\tau/h}.$$

Порядок аппроксимации: $O(\tau+h)$.

Схема устойчива при любых соотношениях между шагами сетки τ и h.

9.15. Неявный правый уголок

Сеточный шаблон:



Разностная схема:

$$\frac{u_m^{p+1}-u_m^p}{\tau}+\frac{u_{m+1}^{p+1}-u_m^{p+1}}{h}=f_m^{p+1},\quad p=0,\ldots,P-1;\quad m=0,\ldots,M-1;$$

$$u_m^0=\varphi_m,\qquad m=0,1,\ldots,M;$$

$$u_0^p=\psi^p,\qquad p=1,2,\ldots,P.$$

Значение сеточной функции на верхнем временном слое $p+1$ рассчитывается по ее значениям в двух точках верхнего и нижнего p-слоев:

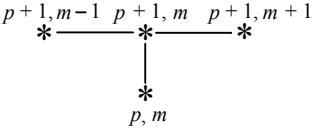
$$u_{m+1}^{p+1}=u_m^{p+1}-\frac{h}{\tau}(u_m^{p+1}-u_m^p)-hf_m^{p+1}.$$

Порядок аппроксимации: $O(\tau+h)$.

Схема устойчива при $\tau / h \geq 1$.

9.16. Неявная четырехточечная схема

Сеточный шаблон:



Разностная схема:

$$\frac{u_m^{p+1} - u_m^p}{\tau} + \frac{u_{m+1}^{p+1} - u_{m-1}^{p+1}}{2h} = f_m^{p+1},$$

$p = 0, 1, \ldots, P-1;$ $\qquad m = 1, 2, \ldots, M-1;$

$u_m^0 = \varphi_m,$ $\qquad m = 0, 1, \ldots, M;$

$u_0^p = \psi^p,$ $\qquad p = 1, 2, \ldots, P.$

Алгоритм численного решения полученной системы линейных уравнений с матрицей трехдиагональной структуры — метод прогонки.

Для предложенного метода рассматриваются пять типов условий на правой границе области интегрирования:

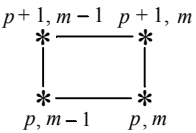
1) схема «явный левый уголок»;
2) схема «неявный левый уголок»;
3) схема «неявный правый уголок»;
4) условие «сноса» (производная сеточной функции по переменной x равна нулю);
5) значение сеточной функции равно нулю.

Порядок аппроксимации: $O(\tau + h^2)$.

Схема устойчива при любых соотношениях между шагами сетки τ и h.

9.17. Схема «прямоугольник»

Сеточный шаблон:



Разностная схема:

$$\frac{u_{m-1}^{p+1} - u_{m-1}^{p} + u_{m}^{p+1} - u_{m}^{p}}{2\tau} + \frac{u_{m}^{p+1} - u_{m-1}^{p+1} + u_{m}^{p} - u_{m-1}^{p}}{2h} = f_{m-1/2}^{p+1/2},$$

$p = 0, 1, \ldots, P-1;$ $\quad m = 1, 2, \ldots, M;$

$u_m^0 = \varphi_m,$ $\quad m = 0, 1, \ldots, M;$

$u_0^p = \psi^p,$ $\quad p = 1, 2, \ldots, P.$

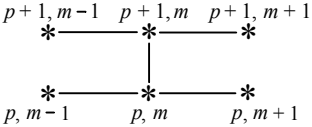
Значение сеточной функции на верхнем временном слое в точке $(p+1, m)$ рассчитывается по известным ее значениям в трех точках $(p+1, m-1)$, (p, m), $(p, m-1)$.

Порядок аппроксимации: $O(\tau^2 + h^2)$.

Схема устойчива при любых соотношениях между шагами сетки τ и h.

9.18. Неявная шеститочечная схема

Сеточный шаблон:



Разностная схема:

$$\frac{u_{m}^{p+1} - u_{m}^{p}}{\tau} + \frac{u_{m+1}^{p+1} - u_{m-1}^{p+1} + u_{m+1}^{p} - u_{m-1}^{p}}{4h} = f_{m}^{p+1/2},$$

$p = 0, 1, \ldots, P-1;$ $\quad m = 1, 2, \ldots, M-1;$

$u_m^0 = \varphi_m,$ $\quad m = 0, 1, \ldots, M;$

$u_0^p = \psi^p,$ $\quad p = 1, 2, \ldots, P.$

Значения сеточной функции в точках x_M рассчитываются по схеме «явный левый уголок».

Порядок аппроксимации: $O(\tau^2 + h^2)$.

Схема устойчива при любых соотношениях между шагами сетки τ и h.

9.19. Точное решение задачи Коши для однородного уравнения

Укажем случай, когда известно решение краевой задачи (9.1а), и можно сравнивать приближенное решение с точным.

Зададим функцию $\varphi(x)$: $-\infty < x < \infty$.

Непосредственно проверяется, что при условиях

$$u(x, 0) = \varphi(x), \qquad 0 \le x \le X,$$

$$u(0, t) = \varphi(-t), \qquad 0 \le t \le T$$

точное решение рассматриваемой краевой задачи задается следующей формулой:

$$u(x, t) = \varphi(x - t).$$

9.20. Порядок выполнения работы

1. Исследуйте поведение численного решения в зависимости от изменения числа Куранта $r = \tau / h$.

Для однородного уравнения переноса (с нулевой правой частью) с установленным разбиением отрезка $[0, T]$ получите численное решение задачи (9.1а) с третьим типом начальных условий (см. пункт меню «*Параметры*»), используя следующие разностные схемы (см. пункт меню «*Методы*»):

1) явный левый уголок ($r = 1$; 0,5; 1,01);
2) явная четырехточечная схема ($q = 0{,}5$; $r = 1$; 0,5; 1,01);
3) неявная четырехточечрая схема ($r = 1$; 0,5; 1,01);
4) неявная шеститочечная схема ($r = 1$; 1,01);
5) явный правый уголок ($r = 0{,}5$);
6) неявный правый уголок ($r = 0{,}5$; 1,01).

Сравните полученные численные решения с точным. Объясните разницу в поведении численных решений при различных значениях числа Куранта. Расчеты проводите последовательно по разным схемам с одним установленным значением числа Куранта, затем установите следующее значение r и т. д.

2. Исследуйте поведение численного решения в зависимости от изменения сеточных параметров (τ, h). Для однородного уравнения и второго типа начальных данных проведите расчеты по следующим разностным схемам (N — количество разбиений отрезка интегрирования):

1) явный левый уголок ($N = 16$; $r = 0{,}5$);

2) явная четырехточечная схема ($q = 0{,}5$; $N = 16$; $r = 0{,}5$); для ($q = 0{,}5$; $N = 30$) расчетным путем подберите шаг по времени ($\tau = rh$) такой, чтобы эта схема была устойчивой;
3) неявный левый уголок ($N = 16$; $r = 0{,}5$);
4) неявная шеститочечная схема ($N = 16$; $r = 0{,}5$);
5) явная центральная трехточечная схема ($N = 80$; $r = 1$; 0,1; 0,01; 0,001);

Обратите внимание на отсутствие сходимости к точному решению, если при измельчении сетки выполняется соотношение $\tau = h^2$. Объясните это явление, исследовав схему на аппроксимацию.

3. Исследуйте поведение численных решений уравнения переноса в зависимости от типа начальных данных. Для всех четырех типов начальных данных проведите расчеты по схемам:

1) явный левый уголок;
2) явная четырехточечная схема;
3) схема «чехарда»;
4) неявный левый уголок;
5) неявная четырехточечная схема;
6) неявная шеститочечная схема.

Объясните поведение этих схем при использовании различных типов начальных данных.

4. Исследуйте влияние на численное решение дополнительного краевого условия на правой границе области интегрирования ($x = X$). Для задачи 2 проведите расчеты (при $r = 0{,}9$; $r = 2$) с использованием неявной четырехточечной схемы и всех указанных в меню краевых условиях.

5. Методы регуляризации численных решений с большими градиентами по координатной переменной:

5.1. *Гибридная схема*. Для установленного разбиения отрезка $[0, T]$ и начальных условий третьего и четвертого типов расчетным путем подберите наилучшее значение (в смысле близости к точному решению) коэффициента λ в формуле перехода от одной схемы к другой (явной четырехточечной и левого уголка). Предварительно проведите численные расчеты при $\lambda = 0$ и $\lambda = 100$ (предельные случаи). Объясните поведение гибридной схемы.

5.2. *Схема с искусственной вязкостью*. Проведя исследование явной четырехточечной схемы на устойчивость, численно опре-

делите наилучшее значение коэффициента q (в смысле близости к точному решению). Используйте третий тип начальных данных.

9.21. Библиографическая справка

Подробнее о разностных схемах для решения модельного уравнения переноса см. в [1–4, 27, 31]. Уравнение переноса является хорошей моделью уравнений газовой динамики, поэтому на его примере проводится изучение свойств разностных схем газовой динамики и тестирование задач механики сплошных сред [33–37].

Гибридные схемы были предложены Р.П. Федоренко (см. [2] и библиографическую справку в ней). В настоящее время подходы, связанные с применением гибридных схем, активно развиваются.

ЧИСЛЕННОЕ РЕШЕНИЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ ГИПЕРБОЛИЧЕСКОГО ТИПА. ВОЛНОВОЕ УРАВНЕНИЕ

10.1. Введение

Работа предоставляет возможность познакомиться с разностными методами решения типичных задач для одномерного гиперболического уравнения второго порядка (на примере задачи о малых колебаниях тонкой струны):

$$\begin{aligned} &u''_{tt} - a^2 u''_{xx} = f(t, x), \qquad 0 \le x \le 1, \quad t \ge 0, \\ &u(x, 0) = \varphi_1(x), \qquad u'_t(0, x) = \varphi_2(x), \\ &u(t, 0) = \psi_1(t), \qquad u(t, 1) = \psi_2(t). \end{aligned} \tag{10.1}$$

Обсуждаются:

1) способы конструирования разностных схем для решения задачи (10.1);
2) вопросы аппроксимации начальных и краевых условий;
3) последовательность вычислений.

10.2. Дифференциальная краевая задача

В работе рассматривается следующая задача:

$$\begin{aligned} &u''_{tt} - a^2 u''_{xx} = f(t, x), \qquad 0 \le x \le 1, \quad t \ge 0, \\ &u(0, x) = \varphi_1(x), \qquad u'_t(0, x) = \varphi_2(x), \\ &u(t, 0) = \psi_1(t), \qquad u(t, 1) = \psi_2(t). \end{aligned}$$

10.3. Сеточная область

Для рассмотренной задачи:

$$\mathrm{W}^h = [(t_p, x_m)], \quad p = 0, 1, ..., P, \quad m = 0, 1, \ldots, M,$$

$$u^h = [u_m^p], \; p = 0, 1, ..., P, \quad m = 0, 1, \ldots, M,$$

где u_m^p — компонента сеточной функции, относящаяся к узлу (t_p, x_m), $t_p = p\tau$, τ — шаг по времени, $P\tau = T$, $x_m = mh$, h — шаг по координате, $Mh = 1$.

10.4. Разностная задача (разностная схема)

Для рассмотренной дифференциальной задачи одна из возможных разностных схем имеет вид:

$$\frac{u_m^{p+1} - 2u_m^p + u_m^{p-1}}{\tau^2} - a^2 \frac{u_{m-1}^p - 2u_m^p + u_{m+1}^p}{h^2} = f_m^p,$$

$$p = 1, 2, \ldots, P-1; \qquad m = 1, 2, \ldots, M-1;$$

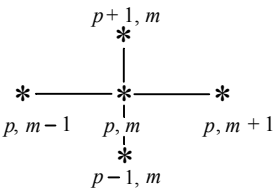
$$u_m^0 = \varphi_{1m}, \qquad \frac{u_m^1 - u_m^0}{\tau} - = \varphi_{2m}, \qquad m = 0, 1, \ldots, M;$$

$$u_0^p = \psi_1^p, \qquad p = 1, 2, \ldots, P;$$

$$u_M^p = \psi_2^p, \qquad p = 1, 2, \ldots, P.$$

10.5. Шаблон разностной схемы

Рассмотренная разностная схема при заданных m и p связывает значения решения в четырех точках сетки, которые образуют конфигурацию, называемую *шаблоном* рассматриваемой схемы.



10.6. Ошибка аппроксимации (невязка)

Представление о величине невязки легко получить, если задать опорную точку и представить значения $[u(t, x)]$, входящие в

выражение для δf^h (или δf_m^p), в виде разложения в ряды Тейлора относительно этой точки. Например, выбирая в качестве опорной точку (t_p, x_m), для рассмотренной в п. 10.4 разностной схемы получим

$$\| \delta f^h \| = O(\tau^2 + h^2).$$

Замечание. Схема «крест» для волнового уравнения имеет порядок аппроксимации, определяемый порядком аппроксимации начальных условий. Он может быть и ниже второго порядка по обоим переменным, достигаемого во внутренних точках.

10.7. Спектральный признак устойчивости

Подробнее о спектральном признаке устойчивости см. в теоретической справке к работам 9–11.

Для однородной разностной задачи ищем решение в виде

$$u_m^p = \lambda^p e^{im\omega}.$$

Подставляя это решение в разностные уравнения, получим

$$\frac{\lambda - 2 + 1/\lambda}{\tau^2} - a^2 \frac{e^{-i\omega} - 2 + e^{i\omega}}{h^2} = 0,$$

или

$$\lambda^2 - 2\left(1 - \frac{2\tau^2 a^2}{h^2} \sin^2 \frac{\omega}{2}\right)\lambda + 1 = 0.$$

Произведение корней этого уравнения равно единице. Если дискриминант

$$D(\omega) = \frac{4\tau^2 a^2}{h^2}\left(\frac{\tau^2 a^2}{h^2} \sin^2 \frac{\omega}{2} - 1\right) \sin^2 \frac{\omega}{2}$$

квадратного уравнения отрицателен, то корни $\lambda_1(\omega)$, $\lambda_2(\omega)$ комплексно-сопряженные и равны единице по модулю.

Разностная схема устойчива, если выполнено неравенство $|\lambda| \le 1$, т. е. когда $\{\tau, h\}$ выбраны так, что

$$\frac{a^2\tau^2}{h^2} \le 1.$$

10.8. Способы конструирования разностных схем для задачи (10.1)

10.8.1. Непосредственная аппроксимация задачи (10.1) на сеточной области. Сеточная область

$$\mathrm{W}^{(h)} = \{(t_p, x_m), \quad p = 0, 1, \ldots, M = 1/h\}, \qquad t_p = p\tau, \qquad x_m = mh,$$

τ — шаг по времени, h — шаг по координате x;

$$u^{(h)} = \{u_m^p, \quad p = 0, 1, \ldots; \quad m = 0, 1, \ldots, M\}$$

— искомая сеточная функция; u_m^p — значение сеточной функции, относящееся к узлу (t_p, x_m).

Схема «*Крест*». Разностные уравнения для внутренних узлов сетки:

$$\frac{u_m^{p+1} - 2u_m^p + u_m^{p-1}}{\tau^2} - a^2 \frac{u_{m-1}^p - 2u_m^p + u_{m+1}^p}{h^2} = f_m^p, \tag{10.2}$$

$p = 1, 2, \ldots, P-1; \qquad m = 1, 2, \ldots, M-1.$

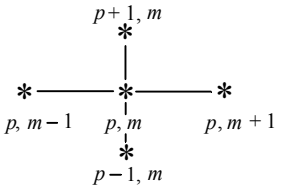
Аппроксимация краевых условий:

$$u_0^p = \psi_1(t^p), \qquad u_m^p = \psi_2(t^p), \qquad p = 1, 2, \ldots, P.$$

Аппроксимация начальных условий:

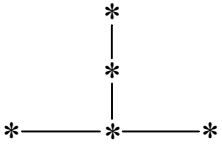
$$u_m^0 = \varphi_1(x_m), \qquad \frac{u_m^1 - u_m^0}{\tau} = \varphi_2(x_m), \qquad m = 0, 1, \ldots, M.$$

Во внутренних узлах уравнения (10.2) аппроксимируют исходное дифференциальное уравнение со вторым порядком точности. Однако в данном варианте схемы второе начальное условие аппроксимируется простейшим образом — с первым порядком точности (по τ). Поэтому в целом это схема первого порядка.



Условие устойчивости численного решения — число Куранта $\mathrm{Cu} = a\tau/h \le 1$. (По поводу отмеченных фактов см. п. 10.4.–10.7). О проведении расчетов по этой схеме см. п. 10.8.

Явная схема. Шаблон схемы имеет вид:



Разностные уравнения для внутренних узлов сетки:

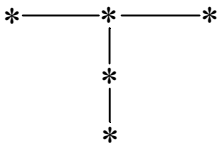
$$\frac{u_m^{p+1} - 2u_m^p + u_m^{p-1}}{\tau^2} - a^2 \frac{u_{m-1}^{p-1} - 2u_m^{p-1} + u_{m+1}^{p-1}}{h^2} = f_m^{p-1}, \tag{10.3}$$

$p = 1, 2, \ldots, P-1;$ $\qquad m = 1, 2, \ldots, M-1.$

Начальные и краевые условия аппроксимируются как и в схеме «крест». Таким образом, данная схема состоит из уравнений (10.3) и начальных и краевых условий из схемы «крест».

Это схема первого порядка точности, но абсолютно неустойчива! Для решения конкретных задач она не используется и приводится здесь лишь как пример абсолютно неустойчивой разностной схемы.

Неявная схема (1). Шаблон схемы имеет вид:



Разностные уравнения для внутренних узлов сетки:

$$\frac{u_m^{p+1} - 2u_m^p + u_m^{p-1}}{\tau^2} - a^2 \frac{u_{m-1}^{p+1} - 2u_m^{p+1} + u_{m+1}^{p+1}}{h^2} = f_m^{p+1}, \tag{10.4}$$

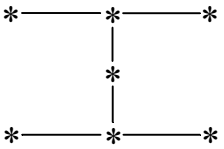
$p = 1, 2, \ldots, P-1;$ $\qquad m = 1, 2, \ldots, M-1.$

Начальные и краевые условия аппроксимируются также как в схеме «крест». Таким образом, данная схема состоит из приведенных здесь уравнений (10.4) и начальных и краевых условий из схемы «крест».

Это схема первого порядка точности абсолютно устойчива. Однако для решения конкретных задач она практически не

используется, так как дает слишком плохие результаты. В этом можно убедиться на методических примерах этой лабораторной работы.

Неявная схема (2). Шаблон схемы имеет вид:



Разностные уравнения для внутренних узлов сетки:

$$\frac{u_m^{p+1} - 2u_m^p + u_m^{p-1}}{\tau^2} -$$

$$- a^2 \frac{(u_{m-1}^{p+1} - 2u_m^{p+1} + u_{m+1}^{p+1}) + (u_{m-1}^{p-1} - 2u_m^{p-1} + u_{m+1}^{p-1})}{2h^2} = f_m^p, \qquad (10.5)$$

$p = 1, 2, \ldots, P-1;$ $\qquad m = 1, 2, \ldots, M-1.$

Начальные и краевые условия аппроксимируются как в схеме «крест». Таким образом, данная схема состоит из уравнений (10.5) и начальных и краевых условий из схемы «крест».

Это схема первого порядка точности по t (за счет грубой аппроксимации начальных данных, как и схема «крест»), абсолютно устойчива.

10.8.2. Об аппроксимации начальных данных. В пояснениях к схеме «крест» приведена простейшая аппроксимация условия $u_t'(0, x) = \varphi_2(t)$, приводящая к погрешности $\mathrm{O}(\tau)$, в то время как во внутренних узлах сетки погрешность аппроксимации для некоторых из рассматриваемых здесь схем является величиной второго порядка как по τ, так и по h.

Можно без труда обеспечить второй порядок и при аппроксимации данного начального условия, так что соответствующая схема станет схемой второго порядка точности.

Представим значение решения в точках первого слоя по времени в виде ряда Тейлора по степеням τ:

$$u(\tau, x_m) = u(0, x_m) + \tau u_t'(0, x_m) + \frac{\tau^2}{2} u_{tt}''(0, x_m) + \mathrm{O}(\tau^3).$$

Замечаем, что из дифференциального уравнения следует

$$u''_{tt} = a^2 u''_{xx} + f(t, x).$$

Таким образом,

$$u(\tau, x_m) = u(0, x_m) + \tau u'_t(0, x_m) + \\ + \frac{\tau^2}{2}\{f(0, x_m) - a^2 u''_{xx}(0, x_m)\} + \mathrm{O}(\tau^3).$$

Отсюда получаем расчетную формулу для данных на первом слое по времени:

$$u^1_m = \varphi_1(x_m) + \tau\varphi_2(x_m) + \frac{\tau^2}{2}\{-a^2\varphi''_{1\,xx}(x_m) + f(0, x_m)\}.$$

Последнее соотношение, переписанное в виде

$$\frac{u^1_m - u^0_m}{\tau} = \varphi_2(x_m) + \frac{\tau}{2}\{-a^2\varphi''_{1\,xx}(x_m) + f(0, x_m)\},$$

очевидно, аппроксимирует условие $u'_t(0, x) = \varphi_2(t)$ со вторым порядком точности.

10.8.3. О последовательности вычислений. Сначала из соотношений для начальных данных схемы «крест» находятся значения: u^0_m ($m = 0, \ldots, M$) и u^1_m ($m = 1, 2, \ldots, M-1$). Недостающие компоненты сеточной функции на первом слое u^1_0 и u^1_M определяются из условий на границах. Затем для $p = 1, 2, \ldots, P-1$ из соотношения (10.2) отыскиваются u^{p+1}_m ($m = 0, 1, \ldots, M$). Для неявных схем необходимо решать систему линейных уравнений с трехдиагональной матрицей.

10.9. Сведение задачи (10.1) к задаче для системы двух уравнений первого порядка

Замечание. В этом пункте демонстрируется, как с помощью искусственного приема рассматриваемую задачу можно свести к другой — известной как задача для уравнений акустики. Последние — ничто иное, как уравнения переноса. Таким образом, этот раздел представляет собой связующее звено между данной лабораторной работой и работой «Численное решение диффе-

ренциальных уравнений в частных производных гиперболического типа. Уравнение переноса».

Задача (10.1) эквивалентна задаче:

$$u'_t - a\upsilon'_x = F(t, x), \qquad \upsilon'_t - au'_x = 0, \qquad t \geq 0, \qquad 0 \leq x \leq 1,$$

где функция, стоящая в правой части первого уравнения

$$F(t, x) = \int_0^t f(\xi, x)\, d\xi + \varphi_2(x),$$

$$u(0, x) = \varphi_1(x), \qquad \upsilon(0, x) = 0, \qquad (10.6)$$

$$u(t, 0) = \psi_1(t), \qquad u(t, 1) = \psi_2(t).$$

10.9.1. Обоснование. Введем в рассмотрение новую функцию $\upsilon(t, x)$ связав ее с $u(t, x)$ уравнением $\upsilon'_t - au'_x = 0$ и полагая $\upsilon(0, x) = 0$. Тогда волновое уравнение можно записать в виде

$$u''_{tt} - a\upsilon''_{tx} - f(t, x) = 0.$$

Интегрируем последнее соотношение по t, получаем:

$$\int_0^t [(u'_t - a\upsilon'_x)'_t - f(t, x)]\, dt = 0.$$

После выполнения интегрирования

$$u'_t - a\upsilon'_x - \int_0^t f(t, x)\, dt = [u'_t - a\upsilon'_x]\Big|_{t=0},$$

т. е. $u'_t - a\upsilon'_x = F(t, x)$, где $F(t, x) = \int_0^t f(\xi, x)\, d\xi + \varphi_2(x)$.

Таким образом, задача (10.1) сведена к задаче (10.6) для двух дифференциальных уравнений первого порядка.

10.9.2. Дополнительные замечания. Легко проверить непосредственно, что дифференциальные уравнения (10.6) можно записать в виде:

$$q'_t - aq'_x = F(t, x), \qquad r'_t + a'r_x = F(t, x),$$

где $q = u - \upsilon$, $r = u + \upsilon$ — *инварианты Римана*.

Особенность последних уравнений состоит в том, что в каждом из них дифференцируется лишь одна неизвестная функция (q или r). Отметим, что возможность записи исходной системы типа (10.6) в виде системы для инвариантов (функций, для которых уравнения разделяются так же, как и здесь) является признаком гиперболичности системы.

Очевидно, что рассматриваемая задача может быть сформулирована как задача для инвариантов:

$$q'_t - aq'_x = F(t, x), \qquad r'_t + ar'_x = F(t, x), \qquad t \geq 0, \qquad 0 \leq x \leq 1,$$

$$q(0, x) = \varphi_1(x), \qquad r(0, x) = \varphi_1(x), \tag{10.7}$$

$$q(t, 0) + r(t, 0) = 2\psi_1(t), \qquad q(t, 1) + r(t, 1) = 2\psi_2(t).$$

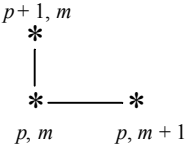
Уравнения для инвариантов, к которым мы перешли, являются известными уравнениями переноса (см. лабораторную работу «Численное решение дифференциальных уравнений в частных производных гиперболического типа. Уравнение переноса»). Последнее обстоятельство позволит распространить разностные схемы для уравнения переноса на наш случай.

Таким образом, вместо исходной задачи (10.1) можно решать задачу (10.6) или задачу (10.7) для инвариантов. В последнем случае исходная искомая функция u вычисляется через q и r в нужных точках по формуле $u = (q + r)/2$.

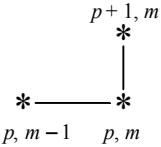
10.9.3. Варианты разностных схем для задачи (10.6). *Предварительные замечания.* Как известно, неявные схемы практически всегда устойчивы. Однако, применительно к задаче (10.6) использование их связано с определенными трудностями, так как расчет данных на очередном слое по времени требует решения линейной системы уравнений (более общего вида, нежели система с трехдиагональной матрицей). Поэтому в рамках данной работы мы сосредоточим внимание на способах построения более простых (с точки зрения реализации) явных разностных схем для задачи (10.6). С ними, однако, применительно к задаче (10.6) снова не все ясно. (Попробуйте по произвольному «явному» шаблону построить устойчивую схему!). Поэтому сначала разберемся как строятся подходящие явные схемы для задачи (10.7), сформулированной для инвариантов q и r.

Так как каждое из дифференциальных уравнений (10.7) есть уравнение переноса, то соответствующие явные схемы для последнего с некоторыми уточнениями, касающимися расчета в точках правой и левой границы, легко обобщаются на задачу (10.7).

СХИ1 — схема для инвариантов (первого порядка точности). Эта схема является естественным обобщением на случай системы для инвариантов схем типа «явный угол» для уравнения переноса. С учетом характеристических направлений (направлений переноса инвариантов) первое из уравнений (10.9) аппроксимируется по шаблону



шаблон для второго уравнения



Разностные уравнения для внутренних узлов сетки:

$$\frac{q_m^{p+1} - q_m^p}{\tau} - a\,\frac{q_{m+1}^p - q_m^p}{h} = F_m^p,$$

$$p = 1, 2, \ldots, P-1; \qquad m = 0, 1, \ldots, M-1;$$

$$\frac{r_m^{p+1} - r_m^p}{\tau} + a\,\frac{r_m^p - r_{m-1}^p}{h} = F_m^p, \tag{10.8}$$

$$p = 1, 2, \ldots, P-1; \qquad m = 1, 2, \ldots, M.$$

Аппроксимация краевых условий:

$$q(t,0) + r(t,0) = 2\psi_1(t),$$

$$q(t,1) + r(t,1) = 2\psi_2(t), \qquad p = 1, 2, \ldots, P-1. \tag{10.9}$$

Начальные данные:

$$q_m^0 = \varphi_1(x_m), \qquad r_m^0 = \varphi_1(x_m), \quad m = 0, 1, \ldots, M. \tag{10.10}$$

Схема имеет первый порядок точности, устойчива при выполнении условия Куранта $\mathrm{Cu} = a\tau/h \le 1$.

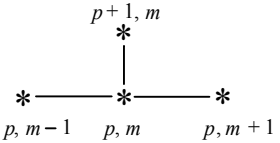
Последовательность вычислений для схемы СХИ1. Из соотношений (10.10) находятся q и r во всех точках начального слоя. Затем в рамках стандартного программного цикла для

$p = 0, 1, 2, \ldots, P-1$ осуществляется расчет данных на $(p+1)$-м слое по времени. При этом:

1) из соотношений (10.8) находятся q_m^{p+1} для $m = 0, \ldots, M-1$ и r_m^{p+1} для $m = 1, 2, \ldots, M$;

2) из (10.9) с использованием найденных значений q_0^{p+1} и r_M^{p+1} отыскиваются r_0^{p+1} и q_M^{p+1}.

СХИ2 — схема для инвариантов (второго порядка точности). В этой схеме используется шаблон



Замечание. Данная схема является обобщением на случай системы (10.7) явной четырехточечной схемы второго порядка для уравнения переноса.

Разностные уравнения для внутренних узлов сетки:

$$\frac{q_m^{p+1} - q_m^p}{\tau} - a\,\frac{q_{m+1}^p - q_{m-1}^p}{2h} = \\ = F_m^p + \frac{\tau}{2}\left[(F_t')_m^p + a\,(F_x')_m^p + a^2\,\frac{q_{m+1}^p - 2q_m^p + q_{m-1}^p}{h^2}\right],$$

$$\frac{r_m^{p+1} - r_m^p}{\tau} + a\,\frac{r_{m+1}^p - r_{m-1}^p}{2h} = \\ = F_m^p + \frac{\tau}{2}\left[(F_t')_m^p - a\,(F_x')_m^p + a^2\,\frac{r_{m+1}^p - 2r_m^p + r_{m-1}^p}{h^2}\right],$$

$$p = 1, 2, \ldots, P-1; \qquad m = 1, 2, \ldots, M-1. \tag{10.11}$$

Аппроксимация краевых условий:

$$q(t,0) + r(t,0) = 2\psi_1(t), \\ q(t,1) + r(t,1) = 2\psi_2(t), \quad p = 1, 2, \ldots, P-1. \tag{10.12}$$

Начальные данные:

$$q_m^0 = \varphi_1(x_m), \qquad r_m^0 = \varphi_1(x_m), \qquad m = 0, 1, \ldots, M. \qquad (10.13)$$

В отличие от схемы СХИ1 данная система соотношений пока не замкнута. Подходящий способ замыкания данной схемы рассматривается ниже. При выбранном там способе замыкания схема имеет второй порядок точности и устойчива при выполнении условия Куранта: $\mathrm{Cu} = a\tau / h \le 1$.

Дополнительные соотношения схемы СХИ2. Заметим, что дифференциальное уравнение для инварианта q — ничто иное, как характеристическое соотношение

$$\left.\frac{dq}{dt}\right|_{c_-} = F,$$

выполняющееся вдоль характеристик c_- : $dx/dt = -a$, которые представляют собой семейство прямых $x + at = \mathrm{const}$.

В частности, это характеристическое соотношение связывает значение q в точках левой границы со значениями q внутри расчетной области (на c_- — характеристике, выпущенной из рассматриваемой точки левой границы). Поэтому логично для получения одного недостающего уравнения аппроксимировать с желаемым порядком точности дифференциальное уравнение для q из (10.7) в ячейках сетки, примыкающих к левой границе. Аналогичным образом второе дополнительное соотношение получается аппроксимацией уравнения для r из (10.7) по ячейке, примыкающей к правой границе.

Замечание. В схеме СХИ1 разностные уравнения для q при $m = 0$ и для r при $m = M$, можно рассматривать как аппроксимации первого порядка характеристических соотношений: для q вблизи левой границы и для r вблизи правой.

В этом случае разностные уравнения (10.11) аппроксимируют соответствующие дифференциальные уравнения для инва-

риантов со вторым порядком точности. Поэтому желательно, чтобы необходимые замыкающие соотношения обеспечивали ту же точность.

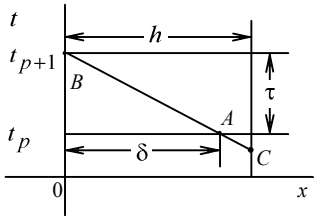


Рис. 10

Приведем соответствующий пример явной аппроксимации дифференциального уравнения для q вблизи левой границы.

На рис. 10 изображен фрагмент сеточной области вблизи левой границы $(x = 0)$. Точка A является точкой пересечения c_- — характеристики, выпущенной из узла B с координатами $(t_{p+1}, x_0 = 0)$, с предыдущим слоем по времени.

Интегрируя характеристическое соотношение

$$\left.\frac{dq}{dt}\right|_{c_-} = F$$

по отрезку AB, получим

$$q_B = q_A + \int\limits_{AB} F\, dt.$$

Заменим интеграл, например, по формуле трапеций. Имеем:

$$q_B = q_A + \frac{\tau}{2}(F_A + F_B) + \mathrm{O}(\tau^3).$$

Значение q_A определим по известным значениям q в узлах p-го слоя по времени с помощью интерполяционной формулы:

$$q_A = q_0^p + \delta\,\frac{q_1^p - q_0^p}{h} + \frac{1}{2}\,\delta\,(\delta - h)\,\frac{q_2^p - 2q_1^p + q_0^p}{h^2} + \mathrm{O}(\delta^3),$$

здесь $\delta = a\tau$ — расстояние точки A от левой границы.

Подставляя выражение для q_A в предыдущую формулу, получаем одно из недостающих дополнительных соотношений:

$$q_0^{p+1} = q_B = q_0^p + \delta\,\frac{q_1^p - q_0^p}{h} + \frac{1}{2}\,\delta\,(\delta - h)\,\frac{q_2^p - 2q_1^p + q_0^p}{h^2} +$$

$$+\frac{\tau}{2}[F(t_{p+1},0)+F(t_p,\tau a)]+\mathrm{O}(\tau^3+\delta^3).$$

Другое недостающее уравнение получим, аппроксимируя аналогичным образом характеристическое соотношение для инварианта r вблизи правой границы:

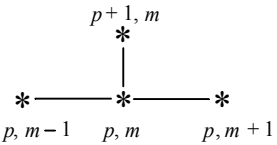
$$r_M^{p+1}=r_M^p-\delta\frac{r_M^p-r_{M-1}^p}{h}+\frac{1}{2}\delta(\delta-h)\frac{r_M^p-2r_{M-1}^p+r_{M-2}^p}{h^2}+$$
$$+\frac{\tau}{2}[F(t_{p+1},1)+F(t_p,1-\tau a)]+\mathrm{O}(\tau^3+\delta^3).$$

Расчет по схеме СХИ2. Из соотношений (10.13) находятся q и r во всех точках начального слоя. Затем в рамках стандартного программного цикла (для $p = 0, 1, 2, \ldots, P-1$) осуществляется расчет данных на $(p+1)$-м слое по времени. При этом:

1) из уравнений (10.11) находится q_m^{p+1}, а также r_m^{p+1} для $m = 1, 2, \ldots, M-1$;
2) из уравнений (10.12) с использованием найденных из дополнительных соотношений q_0^{p+1} и r_M^{p+1} отыскиваются r_0^{p+1} и q_M^{p+1}.

Теперь рассмотрим примеры явных разностных схем непосредственно для задачи (10.6).

Схема СХ1. В этой схеме используется шаблон



Разностные уравнения для внутренних узлов сетки:

$$\frac{u_m^{p+1}-u_m^p}{\tau}-a\frac{\upsilon_{m+1}^p-\upsilon_{m-1}^p}{2h}=F_m^p+\frac{a^2\tau}{2}\frac{u_{m+1}^p-2u_m^p+u_{m-1}^p}{h^2},$$

$$\frac{\upsilon_m^{p+1}-\upsilon_m^p}{\tau}-a\frac{u_{m+1}^p-u_{m-1}^p}{2h}=\frac{a^2\tau}{2}\frac{\upsilon_{m+1}^p-2\upsilon_m^p+\upsilon_{m-1}^p}{h^2},$$

$$p = 1, 2, \ldots, P-1; \qquad m = 1, 2, \ldots, M-1. \tag{10.14}$$

Аппроксимация краевых условий:

$$u(t, 0) = \psi_1(t), \qquad u(t, 1) = \psi_2(t), \qquad p = 1, 2, \ldots, P-1. \tag{10.15}$$

Начальные данные:

$$u_m^0 = \varphi_1(x_m), \qquad v_m^0 = 0, \qquad m = 0, 1, \ldots, M. \tag{10.16}$$

Как и в случае схемы СХИ2 данная система соотношений пока не замкнута. Подходящий способ замыкания данной схемы обсуждается ниже. При разумном способе замыкания схема имеет первый порядок точности и устойчива, если выполнено условие Куранта: $\mathrm{Cu} = a\tau / h \le 1$.

Дополнительные соотношения для схемы СХ1. Недостающие для вычисления значений в точках левой и правой границы соотношения можно получить, аппроксимируя, например, дифференциальное уравнение для v по тому или иному шаблону вблизи левой и правой границ. Возможны следующие варианты.

А) При $m = 0$: шаблон «явный правый уголок», уравнение

$$\frac{v_0^{p+1} - v_0^p}{\tau} - a\,\frac{u_1^p - u_0^p}{h} = 0.$$

При $m = M$: шаблон «явный левый уголок», уравнение

$$\frac{v_M^{p+1} - v_M^p}{\tau} - a\,\frac{u_M^p - u_{M-1}^p}{h} = 0.$$

Дополнительные соотношения:

$$v_0^{p+1} = v_0^p + \frac{a\tau}{h}(u_1^p - u_0^p) \quad \text{и} \quad v_M^{p+1} = v_M^p + \frac{a\tau}{h}(u_M^p - u_{M-1}^p).$$

В) При $m = 0$: шаблон «неявный левый уголок», уравнение

$$\frac{v_0^{p+1} - v_0^p}{\tau} - a\,\frac{u_1^{p+1} - u_0^{p+1}}{h} = 0.$$

При $m = M$: шаблон «явный правый уголок», уравнение

$$\frac{\upsilon_M^{p+1} - \upsilon_M^p}{\tau} - a\,\frac{u_M^{p+1} - u_{M-1}^{p+1}}{h} = 0.$$

Дополнительные соотношения:

$$\upsilon_0^{p+1} = \upsilon_0^p + \frac{a\tau}{h}(u_1^{p+1} - u_0^{p+1}),$$

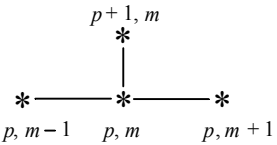
$$\upsilon_M^{p+1} = \upsilon_M^p + \frac{a\tau}{h}(u_M^{p+1} - u_{M-1}^{p+1}).$$

Замечание 1. Возникают естественные вопросы: почему выбрано дифференциальное уравнение для υ при конструировании дополнительных соотношений? Почему выбраны те или иные шаблоны? Точного ответа на первый вопрос нет. Что касается шаблона, то тут руководствуемся тем, чтобы не ухудшить аппроксимацию и устойчивость схемы в целом. Вариант А), по-видимому, непригоден, так как аппроксимация уравнений (10.7) по таким шаблонам, вообще говоря, неустойчива.

Замечание 2. Данная схема является «близкой родственницей» схемы СХИ1 для инвариантов. В самом деле, уравнения группы (10.14) этой схемы являются линейной комбинацией (суммой и разностью) соответствующих уравнений схемы СХИ1. Тем самым проясняется вопрос о происхождении данной схемы.

Учитывая отмеченное обстоятельство, можно сделать вывод, что естественным способом замыкания данной схемы является привлечение в качестве дополнительных соотношений уравнений для q (при $m = 0$) и r (при $m = M$) из группы уравнений (10.14) схемы СХИ1. В этом случае, очевидно, рассматриваемая схема алгебраически эквивалентна схеме СХИ1 для инвариантов.

Схема СХ2. В этой схеме используется шаблон



Разностные уравнения для внутренних узлов сетки представляют собой линейные комбинации (полусумму и полуразность) уравнений для q и r схемы СХИ2 для инвариантов:

$$\frac{u_m^{p+1}-u_m^p}{\tau}-a\,\frac{v_{m+1}^p-v_{m-1}^p}{2h}=$$

$$=F_m^p+\frac{\tau}{2}\left[(F_t')_m^p+a^2\,\frac{u_{m+1}^p-2u_m^p+u_{m-1}^p}{h^2}\right],$$

$$\frac{v_m^{p+1}-v_m^p}{\tau}-a\,\frac{u_{m+1}^p-u_{m-1}^p}{2h}=\frac{a\tau}{2}\left[(F_x')_m^p+a\,\frac{v_{m+1}^p-2v_m^p+v_{m-1}^p}{h^2}\right],$$

$$p=1,\,2,\,\ldots,\,P-1;\qquad m=1,\,2,\,\ldots,\,M-1. \tag{10.17}$$

Краевые условия (10.15) и начальные данные (10.16) записываются так же как в схеме СХ1.

Как и в случае схемы СХИ2 для инвариантов, данная система соотношений пока не замкнута. Подходящий способ замыкания обсуждается ниже. При соответствующем способе замыкания схема имеет второй порядок точности и устойчива, если выполнено условие Куранта: $\mathrm{Cu}=a\tau/h\le 1$.

Дополнительные соотношения для схемы СХ2. Недостающие для вычисления значений v в точках левой и правой границы соотношения можно получить, как и для схемы СХ1, аппроксимируя, например, дифференциальное уравнение для v по тому или иному шаблону вблизи левой и правой границ.

Способы, использованные для замыкания схемы СХ1, приводят к погрешности аппроксимации первого порядка, в то время, как уравнения для внутренних узлов данной схемы обеспечивают второй порядок аппроксимации.

В качестве подходящего способа получения дополнительных уравнений, не портящих второй порядок точности схемы в целом, применим аппроксимацию уравнения для v системы (10.6) вблизи левой и правой границы по прямоугольному шаблону.

При $m=0$ имеем

$$\frac{v_0^{p+1}-v_0^p}{\tau}+\frac{v_1^{p+1}-v_1^p}{\tau}-a\,\frac{u_1^{p+1}-u_0^{p+1}}{h}-a\,\frac{u_1^p-u_0^p}{h}=0.$$

Отсюда получаем дополнительную расчетную формулу, из которой находится υ_0^{p+1}.

Аналогичным образом, при $m = M$ получаем еще одну недостающую формулу (для υ_M^{p+1}).

Применительно к данной схеме справедливы замечания, сделанные при обсуждении способов замыкания схемы СХ1. Естественным способом представляется привлечение характеристических соотношений. В частности, если использовать в качестве дополнительных уравнений аппроксимацию характеристических соотношений вблизи границ со вторым порядком точности, то данная схема будет алгебраически эквивалентна СХ1.

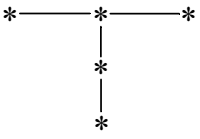
Расчет по схемам СХ1 *и* СХ2. Из соотношений (10.16) находятся u и υ во всех точках начального слоя. Затем осуществляется расчет данных на $(p + 1)$-м слое по времени для значений $p = 0, \ldots, P - 1$. При этом:

1) из уравнений (10.14) находятся величины u_m^{p+1} и υ_m^{p+1} для $m = 1, 2, \ldots, M - 1$;
2) из (10.15) с использованием дополнительных соотношений отыскиваются u_0^{p+1}, υ_0^{p+1} и u_M^{p+1}, υ_M^{p+1}.

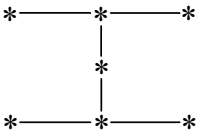
10.10. Контрольные вопросы

1. Исследуйте на аппроксимацию и устойчивость разностные схемы, рассматриваемые в данной лабораторной работе:

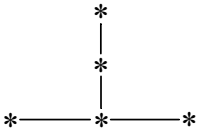
1) схему «крест»;
2) неявную схему с шаблоном



3) неявную схему с шаблоном



4) явную схему с шаблоном



5) схему СХИ1 для инвариантов;
6) схему СХИ2 для инвариантов;
7) схему СХ1 для системы (10.6);
8) схему СХ2 для системы (10.6).

10.11. Порядок выполнения работы

1. Проследите, как воспроизводится по различным разностным схемам решение «эталонных» задач (с известным точным решением).

Замечание. Точное решение можно либо получить в аналитической форме самостоятельно, либо можно посмотреть на графическое его воспроизведение, выбрав нужную задачу по пунктам меню: «*Методы*/*Аналитическое решение*», и осуществив затем пуск соответствующей программы визуализации из пакета меню «*Запуск*».

2. Объясните поведение численного решения «эталонной задачи» номер 2, полученного по разным схемам.

3. Подберите параметры исходной задачи так, чтобы решение представляло собой стоячую волну. Проследите, как это решение воспроизводится по различным разностным схемам.

Под стоячей волной здесь подразумевается стационарное решение волнового уравнения. Например, для задачи

$$u''_{tt} - u''_{xx} = \pi^2 \sin \pi x,$$
$$u(0, t) = u(1, t) = 0, \qquad u(x, 0) = \sin \pi x$$

можно рассмотреть решение $u(x, t) = \sin \pi x$.

10.12. Библиографическая справка

О простейших схемах для решения волнового уравнения можно прочитать в [1–4, 27, 36]. О сеточно-характеристических методах см. [35].

ЧИСЛЕННОЕ РЕШЕНИЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ ПАРАБОЛИЧЕСКОГО ТИПА. УРАВНЕНИЕ ТЕПЛОПРОВОДНОСТИ

11.1. Введение

Эта работа знакомит с численными методами решения дифференциальных уравнений в частных производных параболического типа на примере одномерного линейного уравнения теплопроводности. Рассматривается следующая краевая задача:

$$u'_t - a^2 u''_{xx} = f(t, x), \qquad t > 0, \qquad x_1 \le x \le x_2 ,$$

$$u(0, x) = \varphi(x), \qquad x_1 \le x \le x_2 ,$$

$$u(t, x_1) = \alpha_1(t), \qquad t > 0,$$

$$u(t, x_2) = \alpha_2(t), \qquad t > 0.$$

На примерах рассматриваются особенности поведения численных решений этого уравнения в зависимости от входных данных задачи, выбора разностного метода, сеточных параметров (шагов по времени и координате).

Реализованы следующие разностные схемы для численного решения одномерного уравнения теплопроводности:

1) шеститочечная параметрическая схема;
2) схема Франкела–Дюфорта;
3) схема Ричардсона;
4) явная центральная четырехточечная схема;
5) схема Алена–Чена;
6) нецентральная явная четырехточечная схема;
7) схема Саульева.

Полученные приближенные (численные) решения однородного уравнения можно сравнить с точными для нескольких тестовых задач.

11.2. Дифференциальная краевая задача

Как уже было отмечено, в работе рассматривается задача:

$$u'_t - a^2 u''_{xx} = f(t, x), \qquad t > 0, \qquad x_1 \le x \le x_2 ,$$

$$u(0, x) = \varphi(x), \qquad x_1 \le x \le x_2 ,$$

$$u(t, x_1) = \alpha_1(t), \qquad t > 0,$$

$$u(t, x_2) = \alpha_2(t), \qquad t > 0.$$

11.3. Сеточная область

Для рассмотренной задачи

$$\mathrm{W}^h = [(t_p, x_m)], \qquad p = 0, 1, ..., P, \qquad m = 0, 1, \ldots, M,$$

$$u^h = [u_m^p], \qquad p = 0, 1, ..., P, \qquad m = 0, 1, \ldots, M,$$

где u_m^p — компонента сеточной функции, относящаяся к узлу (t_p, x_m), $t_p = p\tau$, τ — шаг по времени, $P\tau = T$, $x_m = x_1 + mh$, h — шаг по координате, $Mh = x_2 - x_1$.

11.4. Пример разностной задачи (разностной схемы)

Для рассмотренной дифференциальной задачи одна из возможных разностных схем имеет следующий вид:

$$\frac{u_m^{p+1} - u_m^p}{\tau} - a^2 \frac{u_{m-1}^p - 2u_m^p + u_{m+1}^p}{h^2} = f_m^p ,$$

$$p = 0, 1, \ldots, P-1; \qquad m = 1, 2, \ldots, M-1;$$

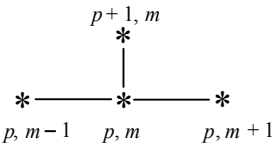
$$u_m^0 = \varphi_m , \qquad m = 0, 1, \ldots, M; \qquad (11.1a)$$

$$u_0^p = \alpha_1^p , \qquad p = 1, 2, \ldots, P;$$

$$u_M^p = \alpha_2^p , \qquad p = 1, 2, \ldots, P.$$

11.5. Шаблон разностной схемы

Рассмотренная разностная схема при заданных m и p связывает значения решения в четырех точках сетки, которые образуют конфигурацию, называемую *шаблоном* схемы.



11.6. Спектральный признак устойчивости

Для широкого класса эволюционных (зависящих от времени) задач исследование устойчивости можно осуществить с помощью спектрального признака, который в случае разностной задачи с постоянными коэффициентами, состоит в следующем.

Заменяем правую часть разностного уравнения в (11.1а) нулем, краевую задачу — задачей Коши, функцию φ_m — гармоникой $e^{i\omega m}$ и ищем решение в виде

$$u_m^p = \lambda^p \, e^{i\omega m}$$

(для задач с одной пространственной переменной), ω — произвольное число, $0 \le \omega \le 2\pi$.

Для устойчивости разностной схемы необходимо, чтобы спектр $\lambda = \lambda(\omega)$ лежал в круге $|\lambda| \le 1 + c\tau$, где c не зависит от τ. Подставляя $u_m^p = \lambda^p \, e^{i\omega m}$ в рассмотренное разностное уравнение, получим:

$$\frac{\lambda - 1}{\tau} + a^2 \, \frac{e^{-i\omega m} - 2 + e^{i\omega m}}{h^2} = 0,$$

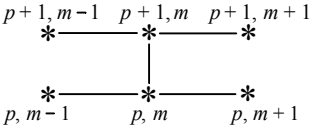
или

$$\lambda = 1 - \frac{4a^2\tau}{h^2} \sin^2 \frac{\omega}{2}.$$

Разностная схема устойчива, если выполнено неравенство $|\lambda| \le 1$, т. е. когда τ, h выбраны так, что $\tau a^2 / h^2 \le 0{,}5$.

11.7. Шеститочечная параметрическая схема

Сеточный шаблон:



Разностная схема:

$$\frac{u_m^{p+1}-u_m^p}{\tau}-\frac{a^2}{h^2}[\sigma\,(u_{m-1}^{p+1}-2u_m^{p+1}+u_{m+1}^{p+1})+$$

$$+(1-\sigma)(u_{m-1}^{p}-2u_m^{p}+u_{m+1}^{p})]=f_m^{p+1/2},$$

$p = 0, 1, \ldots, P-1;$ $\quad m = 1, 2, \ldots, M-1;$

$u_m^0 = \varphi_m,$ $\quad m = 0, 1, \ldots, M;$

$u_0^p = \alpha_1^p,$ $\quad p = 1, 2, \ldots, P;$

$u_M^p = \alpha_2^p,$ $\quad p = 1, 2, \ldots, P.$

где $0 \le \sigma \le 1$ — параметр схемы.

$\sigma = 0$ — явная четырехточечная схема;

$\sigma = 1$ — неявная четырехточечная схема;

$\sigma = 1/2$ — схема Кранка–Николсона.

Метод решения полученной системы линейных уравнений с матрицей трехдиагональной структуры — прогонка.

Порядок аппроксимации:

$\sigma = 1/2$: $\quad O(\tau^2 + h^2),$

$\sigma = 0;\ 1$: $\quad O(\tau + h^2),$

$\sigma = 1/6$: $\quad O(\tau + h^4).$

Введем обозначения

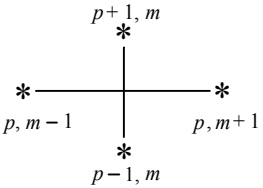
$$K = a^2 r, \qquad r = \frac{\tau}{h^2}.$$

Схема устойчива при любых K, если $\sigma \ge 1/2$; при $0 \le \sigma < 1/2$ схема устойчива, если

$$\tau < \frac{h^2}{2a^2(1-2\sigma)}.$$

11.8. Схема Франкела–Дюфорта

Сеточный шаблон:



Разностная схема:

$$\frac{u_m^{p+1} - u_m^{p-1}}{2\tau} - a^2 \frac{u_{m-1}^{p} - [u_m^{p+1} + u_m^{p-1}] + u_{m+1}^{p}}{h^2} = f_m^p,$$

$p = 1, 2, \ldots, P-1;$ $\qquad m = 1, 2, \ldots, M-1;$

$u_m^0 = \varphi_m,$ $\qquad m = 0, 1, \ldots, M;$

$u_0^p = \alpha_1^p,$ $\qquad p = 1, 2, \ldots, P;$

$u_M^p = \alpha_2^p,$ $\qquad p = 1, 2, \ldots, P.$

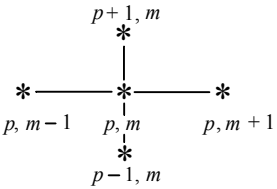
Значения функции на втором слое по времени рассчитываются по явной центральной четырехточечной схеме. Значение сеточной функции на верхнем временном слое $p + 1$ рассчитывается по ее значениям на двух предыдущих нижних слоях: p и $p - 1$.

Порядок аппроксимации: $O(\tau^2 + h^2 + \tau^2 / h^2)$.

Схема устойчива при любых $K = a^2 r, \quad r = \tau / h^2$.

11.9. Схема Ричардсона

Сеточный шаблон:



Разностная схема:

$$\frac{u_m^{p+1} - u_m^{p-1}}{2\tau} - a^2 \frac{u_{m-1}^p - 2u_m^p + u_{m+1}^p}{h^2} = f_m^p,$$

$p = 1, 2, \ldots, P-1;$ $\quad m = 1, 2, \ldots, M-1;$

$u_m^0 = \varphi_m,$ $\quad m = 0, 1, \ldots, M;$

$u_0^p = \alpha_1^p,$ $\quad p = 1, 2, \ldots, P;$

$u_M^p = \alpha_2^p,$ $\quad p = 1, 2, \ldots, P.$

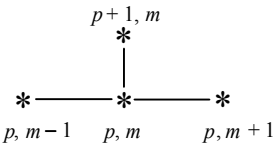
Значения сеточной функции на втором слое по времени рассчитываются по явной центральной четырехточечной схеме. Значение сеточной функции на верхнем временном слое $p + 1$ рассчитывается по ее значениям на двух предыдущих нижних слоях p и $p - 1$.

Порядок аппроксимации: $O(\tau^2 + h^2)$.

Схема неустойчива при любых K.

11.10. Явная центральная четырехточечная схема

Сеточный шаблон:



Разностная схема:

$$\frac{u_m^{p+1} - u_m^p}{\tau} - a^2 \frac{u_{m-1}^p - 2u_m^p + u_{m+1}^p}{h^2} = f_m^p,$$

$p = 0, 1, \ldots, P-1;$ $\quad m = 1, 2, \ldots, M-1;$

$u_m^0 = \varphi_m,$ $\quad m = 0, 1, \ldots, M;$

$u_0^p = \alpha_1^p,$ $\quad p = 1, 2, \ldots, P;$

$u_M^p = \alpha_2^p,$ $\quad p = 1, 2, \ldots, P.$

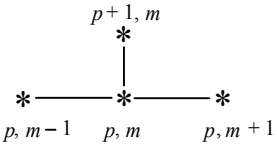
Значение сеточной функции на верхнем временном слое $p+1$ рассчитывается по ее значениям на нижнем слое p.

Порядок аппроксимации: $O(\tau + h^2)$.

Схема устойчива при $K \le 1/2$.

11.11. Схема Алена–Чена

Сеточный шаблон:



Разностная схема:

$$\frac{u_m^{p+1} - u_m^p}{\tau} - a^2 \frac{u_{m-1}^p - 2u_m^{p+1} + u_{m+1}^p}{h^2} = f_m^p,$$

$p = 0, 1, \ldots, P-1;$ $\quad m = 1, 2, \ldots, M-1;$

$u_m^0 = \varphi_m,$ $\quad m = 0, 1, \ldots, M;$

$u_0^p = \alpha_1^p,$ $\quad p = 1, 2, \ldots, P;$

$u_M^p = \alpha_2^p,$ $\quad p = 1, 2, \ldots, P.$

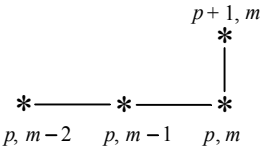
Значения сеточной функции на верхнем временном слое находятся по ее значениям на нижнем слое, поскольку разностное уравнение разрешается относительно u_m^{p+1}.

Порядок аппроксимации: $O(\tau + h^2 + \tau/h^2)$.

Схема устойчива при любых K.

11.12. Нецентральная явная схема

Сеточный шаблон:



Разностная схема:

$$\frac{u_m^{p+1} - u_m^p}{\tau} - a^2 \frac{u_{m-2}^p - 2u_{m-1}^p + u_m^p}{h^2} = f_m^p,$$

$p = 0, 1, \ldots, P-1;$ $\quad m = 2, 3, \ldots, M;$

$u_m^0 = \varphi_m,$ $\quad m = 0, 1, \ldots, M;$

$u_0^p = \alpha_1^p,$ $\quad p = 1, 2, \ldots, P;$

$u_M^p = \alpha_2^p,$ $\quad p = 1, 2, \ldots, P.$

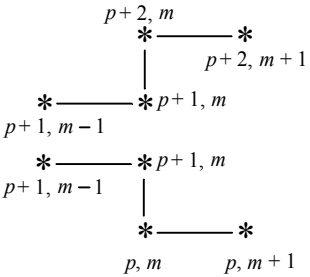
Значение сеточной функции на верхнем временном слое $p + 1$ рассчитывается по ее значениям на нижнем слое p (значения сеточной функции в точках $\{m = 1;\ p = 1, 2, \ldots, P\}$ рассчитываются по шеститочечной параметрической схеме при $\sigma = 1$).

Порядок аппроксимации: $O(\tau + h)$.

Схема неустойчива при любых K.

11.13. Схема Саульева

Сеточный шаблон:



Разностная схема:

$$\frac{u_m^{p+1} - u_m^p}{\tau} - a^2 \frac{u_{m-1}^{p+1} - [u_m^{p+1} + u_m^p] + u_{m+1}^p}{h^2} = f_m^{p+1/2},$$

$$\frac{u_m^{p+2} - u_m^{p+1}}{\tau} - a^2 \frac{u_{m-1}^{p+1} - [u_m^{p+2} + u_m^{p+1}] + u_{m+1}^{p+2}}{h^2} = f_m^{p+3/2},$$

$p = 0, 1, \ldots, P-2;$ $\quad m = 1, 2, \ldots, M-1;$

начальные и граничные условия в такой схеме реализуют следующим образом:

$$u_m^0 = \varphi_m, \qquad m = 0, 1, \ldots, M;$$

$$u_0^p = \alpha_1^p, \qquad p = 1, 2, \ldots, P;$$

$$u_M^p = \alpha_2^p, \qquad p = 1, 2, \ldots, P.$$

Алгоритм численного решения задачи — «бегущий счет»: слева направо — первый этап, справа налево — второй.

Порядок аппроксимации: $O(\tau^2 + h^2 + \tau^2 / h^2)$.

Схема устойчива при любых K.

11.14. Точные решения тестовых краевых задач для одномерного линейного уравнения теплопроводности

В дальнейшем мы будем сравнивать численные решения с точными решениями следующих задач.

Задача 1:

$$u'_t = u''_{xx}, \qquad 0 < t < T, \quad 0 < x < 1,$$

$$u(0, x) = 0, \qquad 0 < x < 1,$$

$$u(t, 0) = 0, \qquad 0 < t < T,$$

$$u(t, 1) = 1, \qquad 0 < t < T.$$

Точное решение:

$$u(t, x) = x + 2 \sum_{n=1}^{\infty} (-1)^n \cdot (\pi n)^{-1} \cdot e^{-n^2 \pi^2 t} \sin n\pi x.$$

Задача 2:

$$u'_t = u''_{xx}, \qquad 0 < t < T, \quad 0 < x < 1,$$

$$u(0, x) = \sin \pi x, \qquad 0 < x < 1,$$

$$u(t, 0) = 0, \qquad 0 < t < T,$$

$$u(t, 1) = 0, \qquad 0 < t < T.$$

Точное решение:

$$u(t, x) = e^{-\pi^2 t} \sin \pi x.$$

Задача 3:

$$u'_t = u''_{xx}, \qquad 0 < t < T, \quad 0 < x < 1,$$
$$u(0, x) = 0, \qquad 0 < x < 1,$$
$$u(t, 0) = 0, \qquad 0 < t < T,$$
$$u(t, 1) = 0, \qquad 0 < t < T.$$

Точное решение:

$$u(t, x) = \frac{1}{\pi^2}[1 - e^{-\pi^2 t}] \sin \pi x.$$

11.15. Порядок выполнения работы

1. Исследуйте поведение численного решения линейного уравнения теплопроводности при изменении параметра $K = a^2 \tau / h^2$ (начальные и краевые условия задачи 1).

Проведите расчеты по всем указанным в меню разностным схемам при следующих значениях параметров:

1) $K = 0{,}5$; $N = 50$;
2) $K = 20$; $N = 50$;
3) $K = 1{,}01$; $N = 50$.

Сравните полученные численные решения с аналитическим, исследуйте эти схемы на сходимость.

2. Исследуйте поведение численного решения линейного уравнения теплопроводности при изменении шага интегрирования.

Проведите численные расчеты по устойчивым разностным схемам при следующих значениях параметров:

1) $N = 6$, $K = 1$;
2) $N = 12$, $K = 1$;
3) $N = 60$, $K = 1$;

(начальные и краевые условия задачи 1). Сравните полученные численные решения с аналитическим.

3. Получите численные решения задачи 2 при всех указанных в меню начальных условиях. Сравните численные решения, найденные по указанным в меню разностным схемам с точным решением

задачи 2; исследуйте расчетным путем сходимость численного решения (при измельчении расчетной сетки) к точному.

4. Получите с помощью любой из указанных разностных схем численное решение однородного уравнения теплопроводности при следующих начальных краевых условиях:

$$u(0, x) = 0, \qquad u(t, 0) = 0,$$

$$u(t, 1) = \sin(\omega t) + 1,$$

при различных значениях ω ($N = 50$, $10 < \omega < 100\,000$).

При каком значении ω решение существенно отличается от нуля только вблизи границы?

5. Получите численные решения одномерного уравнения теплопроводности при следующих правых частях:

1) см. задачу 3;

2) $f(t, x) = -|x - 1| + 1, \qquad -5 \le x \le 5;$

3) $f(t, x) = e^{x^3} - 1, \qquad 0 \le x \le 1.$

6. *Появление паразитных осцилляций в численных решениях*. Получите численные решения уравнения теплопроводности при следующих параметрах задачи ($K = 25$):

а) $a = 1, \qquad u(t, 0) = u(t, 1) = 0, \qquad f(t, x) = 0;$

$$u(0, x) = \frac{1}{\operatorname{ch}\{99(x - 0{,}5)\}}.$$

б) см. задачу 1.

Используйте следующие разностные схемы:

1) шеститочечная параметрическая: σ = 0,5; 1; 0;
2) схема Франкела–Дюфорта;
3) схема Алена–Чена;
4) схема Саульева.

Какие из перечисленных разностных схем приводят к появлению паразитических осцилляций в численных решениях?

11.16. Библиографическая справка

О разностных схемах для решения уравнения теплопроводности см. в [1–4, 27, 31, 32, 35, 36]. Линейное уравнение теплопроводности — хорошая тестовая задача для рассмотрения устойчивости разностных схем — как спектрального метода [1, 31, 39], так и энергетических методов [40–42].

ЧИСЛЕННОЕ РЕШЕНИЕ УРАВНЕНИЙ ЭЛЛИПТИЧЕСКОГО ТИПА. УРАВНЕНИЕ ПУАССОНА

12.1. Введение

Простейшим уравнением в частных производных эллиптического типа является уравнение Пуассона

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \varphi(x, y). \tag{12.1}$$

Пусть задана некоторая область D на плоскости, а на ее границе $\Gamma = \partial D$ поставлено краевое условие вида

$$\left(au - b \frac{\partial u}{\partial n} \right) \Bigg|_{\Gamma} = \psi(s), \tag{12.2}$$

где $\frac{\partial}{\partial n}$ — производная в направлении внутренней нормали, $a \geq 0, \quad b \geq 0, \quad a^2 + b^2 = 1$ — некоторые числа, s — длина дуги, отсчитывается вдоль границы Г. Функции $\varphi(x, y)$, $\psi(s)$ считаются заданными. Требуется найти численное решение краевой задачи (12.1), (12.2). В случае $a = 1$, $b = 0$ возникающая задача называется *первой краевой задачей* или *задачей Дирихле*. В случае $a = 0, \quad b = 1$ — это *вторая краевая задача* или *задача Неймана*. В случае $a > 0, \quad b > 0$ — это *третья краевая задача*.*

Перечисленные краевые задачи являются основными для уравнения Пуассона. Наряду с уравнением Пуассона могут рассматриваться уравнения с переменными коэффициентами сле-

* В зарубежной литературе она иногда называется задачей Робена.

дующего вида:

$$\frac{\partial}{\partial x}\left(a\frac{\partial u}{\partial x}\right)+\frac{\partial}{\partial y}\left(b\frac{\partial u}{\partial y}\right)=\varphi(x, y),$$

где $a=a(x, y)>0;\ \ b=b(x, y)>0;$ для которых также ставится первая, вторая или третья краевая задача.

Эллиптические уравнения применяются для описания многих стационарных состояний. Так, например, уравнение Пуассона может описывать потенциал электрического поля, потенциал скоростей установившегося потока несжимаемой жидкости, установившуюся температуру в однородном теплопроводном теле. Система уравнений упругости Ляме описывает смещения в находящемся под действием стационарных сил твердом теле и т. д.

Численное решение краевых задач для эллиптических уравнений во многих случаях осуществляется с помощью разностных схем. В простейших случаях, когда решение во всей рассматриваемой области меняется более или менее равномерно, а сама область не имеет узких «перешейков», можно использовать регулярные сетки, а для получения разностных схем заменять производные разностными отношениями.

В противном случае регулярная сетка становится практически непригодной, так как места быстрого изменения решения или места, где область D имеет узкие «горловины», диктуют слишком мелкую сетку. В случае нерегулярной сетки построение разностной схемы путем замены производных разностными соотношениями становится сложной процедурой. В этом случае большое распространение получили так называемые *вариационно-разностные* и *проекционно-разностные* схемы. «Прикладники» соответствующие схемы обычно называют *методом конечных элементов*.

О схемах конечных элементов см. в [1, 11, 16, 32, 43].

12.2. Аппроксимация и устойчивость простейшей разностной схемы

Рассмотрим задачу Дирихле для уравнения Пуассона (12.1) в квадратной области $\mathrm{D}=\{0\le x, y\le 1\}$ с границей Г. Совокупность точек $(x, y)=(mh, nh)$ сетки (где $h = 1/M$, M — целое), попавших внутрь квадрата или на его границу, обозначим D_h.

Точки D_h, лежащие строго внутри квадрата D, будем называть *внутренними точками сеточного квадрата* D_h. Совокупность внутренних точек обозначим D_h^0. Совокупность точек D_h, попавших на границу квадрата D, будем обозначать Γ_h.

Рассмотрим разностную схему

$$L_h u^{(h)} = f^{(h)}, \tag{12.3}$$

здесь

$$L_h u^{(h)} = \begin{cases} \dfrac{u_{m+1,\,n} - 2u_{mn} + u_{m-1,\,n}}{h^2} + \dfrac{u_{m,\,n+1} - 2u_{mn} + u_{m,\,n-1}}{h^2} \\ \qquad\qquad \text{при } (x_m, y_n) \in D_h^0, \\ u_{mn} \qquad\qquad \text{при } (x_m, y_n) \in \Gamma_h. \end{cases} \tag{12.4}$$

Правая часть $f^{(h)}$ разностной схемы (12.3) принимает вид

$$f^{(h)} = \begin{cases} \varphi(x_m, y_n) & \text{при} \quad (x_m, y_n) \in D_h^0, \\ \psi(s_{mn}) & \text{при} \quad (x_m, y_n) \in \Gamma_h. \end{cases} \tag{12.5}$$

где $\psi(s_{mn})$ — значение функции $\psi(s)$ в точке (x_m, y_m), принадлежащей границе Γ_h.

Обозначим через $[u]_h$ проекцию точного решения задачи на пространство сеточных функций. Например, это может быть сеточная функция, численно совпадающая с решением задачи в узлах сетки. Определим

$$\| f^{(h)} \| = \max_{(mh,\, nh) \in D_h^0} | \varphi_{mn} | + \max_{(mh,\, nh) \in \Gamma_h} | \psi_{mn} | .$$

Можно показать, что норма невязки $\| \delta f^{(h)} \|$, возникающая при подстановке $[u]_h$ в левую часть разностной схемы (12.3) составляет $O(h^2)$.

Таким образом, разностная краевая задача (12.3) аппрокси-

мирует задачу Дирихле со вторым порядком относительно h.

Определим норму в пространстве U_h функций, заданных на сетке D_h, положив

$$\| u^{(h)} \|_{\mathrm{U}_h} = \max_{(mh,\, nh)\in \mathrm{D}_h} | u_{mn} | .$$

Перейдем к исследованию устойчивости. Это исследование опирается на следующий факт, представляющий самостоятельный интерес.

Теорема 1. (Принцип максимума). *Каждое решение разностного уравнения*

$$\Delta_h \, v^{(h)}\Big|_{(mh,\, nh)} = 0, \qquad (mh, nh) \in \mathrm{D}_h^0,$$

$$\Delta_h v^{(h)}\Big|_{(mh,\, nh)} \equiv \frac{v_{m+1,\, n} - 2v_{mn} + v_{m-1,\, n}}{h^2} + \frac{v_{m,\, n+1} - 2v_{mn} + v_{m,\, n-1}}{h^2}$$

достигает своих наибольшего и наименьшего значений в некоторых точках Γ_h.

Теорема 2. *Задача* (12.3) *однозначно разрешима при произвольной правой части* $f^{(h)}$, *причем это свойство не зависит от выбора нормы, и разностная схема* (12.3) *устойчива, т. е. выполнена оценка вида*

$$\| u^{(h)} \|_{\mathrm{U}_h} \le c \, \| f^{(h)} \|_{\mathrm{F}_h},$$

где c *не зависит ни от* h, *ни от* $f^{(h)}$.

В случае задачи Дирихле для эллиптического уравнения с переменными коэффициентами

$$\frac{\partial}{\partial x}\left(a \frac{\partial u}{\partial x} \right) + \frac{\partial}{\partial y}\left(b \frac{\partial u}{\partial y} \right) = \varphi(x, y), \qquad (x, y) \in \mathrm{D},$$

$$u\big|_{\Gamma} = \psi(s),$$

где $a = a(x, y) > 0$; $b = b(x, y) > 0$ — положительные в прямоугольнике D гладкие функции, разностную схему можно построить аналогично.

На практике при решении конкретных задач обычно огра-

ничиваются обоснованиями принципиального характера на модельных задачах типа приведенной выше. Конкретные рассуждения о погрешности получаются, как правило, не из теоретических оценок, а из сравнения между собой результатов расчетов, выполненных на сетках с различными значениями шага h.

После того, как разностная краевая задача, аппроксимирующая дифференциальную, построена, нужно указать не слишком трудоемкий способ ее решения. Ведь при малом h задача (12.3) представляет собой систему скалярных уравнений очень высокого порядка.

Численные методы решения систем линейных уравнений большой размерности делятся на две группы: *прямые* и *итерационные*. В прямых методах точное решение находится за конечное число арифметических действий. Примерами прямых методов являются метод дискретного преобразования Фурье и метод сопряженных градиентов.

Каждый итерационный метод состоит в том, что при решении системы уравнений указывается рекуррентное соотношение, которое по заданному произвольно «нулевому» приближению u^0 решения u позволяет вычислить первое, второе, p-е, $p = 1, 2, 3, \ldots$ приближение u^p решения u.

В работе для итерационных методов дается наглядная иллюстрация изменения «невязки» решения в зависимости от итерации. Эффективность различных методов (в том числе прямых) можно сравнить по затратам машинного времени (необходимая информация выводится на экране) для достижения заданной точности. В случае итерационных методов вычисления проводятся до тех пор, пока не будет выполнена оценка

$$\| u - u^p \| < \varepsilon.$$

Здесь u — вектор точного решения, u^p — решение, полученное на p-й итерации, ε — наперед заданное число.

Задача отыскания точного решения не диктуется, как правило, запросами приложений. В приложениях обычно допустимо использование приближенного решения, известного с достаточной точностью. Поэтому во многих случаях для вычисления решения точным методам целесообразно предпочесть тот или иной итерационный метод. Итерационный процесс должен быть построен так, чтобы последовательность приближений u^p стреми-

лась к решению u. Тогда для любого $\varepsilon > 0$ существует номер $n = n(\varepsilon)$ такой, что $\| u - u^n \| < \varepsilon$. . Задавая $\varepsilon > 0$ достаточно малым, можно воспользоваться n-м приближением u.

В работе рассматриваются следующие методы:

1) дискретного преобразования Фурье;
2) сопряженных градиентов;
3) простых итераций;
4) трехслойный итерационный метод Чебышева;
5) спектрально эквивалентных операторов.

Предусмотрены возможности сравнения методов по затратам времени ЭВМ и проведения расчетов на различных разностных сетках.

12.3. Обусловленность систем линейных уравнений

Общие сведения об обусловленности линейных систем см. в справке к работе 4, п. 4.2.

Рассмотрим задачу Дирихле для уравнения (12.1) в квадрате $\mathrm{D} = \{0 \le x, y \le 1\}$ с однородными граничными условиями. Обозначим через λ_i собственные значения оператора $L_h' = -L_h$ (см. (12.3)–(12.4)) при однородных граничных условиях. Тогда можно показать, что имеет место следующая оценка для λ_i :

$$\lambda_{\min} \le \lambda_i \le \lambda_{\max},$$

где $\lambda_{\min} = \dfrac{8}{h^2} \sin^2 \dfrac{\pi h}{2}$, $\lambda_{\max} = \dfrac{8}{h^2} \cos^2 \dfrac{\pi h}{2}$.

Легко видеть, что оператор L_h в (12.3) при малых h обладает плохой обусловленностью:

$$\mu(L_h') = \mathrm{tg}^{-2} \frac{\pi h}{2} \sim \frac{4}{\pi^2 h^2}.$$

12.4. Метод дискретного преобразования Фурье

Пусть решается задача Дирихле для уравнения Пуассона в единичном квадрате с однородными граничными условиями.

Рассмотрим следующую одномерную задачу на собствен-

ные функции и собственные значения:

$$\frac{\mu(n+1)-2\mu(n)+\mu(n-1)}{h^2}=\lambda\mu(n), \qquad n=1,\,2,\,3,\,\ldots,\,M-1, \qquad M=\frac{1}{h};$$

$$\mu_0=\mu_N=0.$$

Задача имеет следующие решения:

$$\mu_k(n)=\sqrt{2}\,\sin\pi knh,$$

$$\lambda_k=-\frac{4}{h^2}\sin^2\frac{\pi kh}{2}, \qquad k,\,n=1,\,2,\,3,\,\ldots,\,M-1, \qquad M=\frac{1}{h}.$$

Зафиксируем m ($0<m<M$) в (12.3)–(12.5). Тогда можно разложить u_{mn} и φ_{mn} по собственным функциям $\mu_k(n)$:

$$u_{mn}=\sum_{k=1}^{M-1}c_k(m)\,\mu_k(n),$$

$$\varphi_{mn}=\sum_{k=1}^{M-1}f_k(m)\,\mu_k(n), \qquad n=1,\,2,\,\ldots,\,M-1,$$

где

$$f_k(m)=\sum_{j=1}^{M-1}h\,\varphi_{mj}\,\mu_k(j),$$

а относительно $c_k(m)$ получаем систему разностных уравнений второго порядка с трехдиагональной матрицей

$$\frac{c_k(m-1)-2c_k(m)+c_k(m+1)}{h^2}+\lambda_k c_k(m)=f_k(m),$$

$$c_k(0)=c_k(M), \qquad k=1,\,2,\,\ldots,\,M-1,$$

решаемую методом прогонки.

Алгоритм реализации метода Фурье требует $O(M^3)$ арифметических операций. Это число для больших $M=2^n$ (n — натуральное число) можно значительно сократить до $O(M^2\ln M)$, если применить алгоритм быстрого преобразования Фурье.

Недостатком изложенного метода является ограниченность

области его применимости, предполагающая знание собственных функций и собственных значений одномерной задачи.

В данной работе решается задача Дирихле в прямоугольнике, и решение, полученное методом дискретного преобразования Фурье принимается за точное.

12.5. Метод сопряженных градиентов

См. лабораторную работу 4, п. 4.4.

12.6. Метод простых итераций

См. лабораторную работу 4, п. 4.5.

12.7. Метод с оптимальным параметром

См. лабораторную работу 4, п. 4.7.

12.7.1. Переход к лучше обусловленной системе с помощью энергетически эквивалентного оператора. См. лабораторную работу 4, п. 4.6.1.

12.7.2. Масштабирование как средство улучшения числа обусловленности. См. лабораторную работу 4, п. 4.6.2.

12.8. Трехслойный метод Чебышева

См. лабораторную работу 4, п. 4.8.

12.9. Метод спектрально-эквивалентных операторов

Для решения системы $\mathbf{Ax} = \mathbf{b}$ рассмотрим итерационный процесс вида

$$\mathbf{B}\mathbf{x}^{n+1} = \mathbf{B}\mathbf{x}^{n} - \alpha\,(\mathbf{A}\mathbf{x}^{n} - \mathbf{b}),$$

где матрица $\mathbf{B} \neq \mathbf{E}$.

Пусть $\mathbf{A}$, $\mathbf{B} > 0$ и матрица $\mathbf{A}$ обладает плохой обусловленностью. Тогда, если удается найти такую матрицу $\mathbf{B}$, что система уравнений $\mathbf{Bx} = \mathbf{c}$ может быть легко решена, и величина

$$\frac{\sup\limits_{\mathbf{x}} \dfrac{(\mathbf{Ax}, \mathbf{x})}{(\mathbf{Bx}, \mathbf{x})}}{\inf\limits_{\mathbf{x}} \dfrac{(\mathbf{Ax}, \mathbf{x})}{(\mathbf{Bx}, \mathbf{x})}}$$

много меньше, чем число обусловленности матрицы **A**, то такой итерационный процесс сходится значительно быстрее, чем метод простой итерации.

Так, например, можно при численном решении уравнения эллиптического типа с переменными коэффициентами в качестве матрицы B выбрать матрицу, соответствующую оператору, возникающему при аппроксимации оператора Лапласа. В этом случае матрицу можно обращать, например, методом преобразования Фурье.

12.10. Контрольные вопросы

1. Доказать, что, если во внутренних точках области D_h функция $u^{(h)}$ удовлетворяет уравнению

$$L_h u^{(h)}\Big|_{(mh,\, nh)} = 0, \qquad m, n = 1, 2, \ldots, M-1, \qquad Mh = 1,$$

то, либо $u^{(h)}$ принимает всюду на D_h одинаковые значения, либо наибольшее и наименьшее значения функции $u^{(h)}$ не достигаются ни в одной внутренней точке сетки D_h (усиленный принцип максимума).

2. Если во внутренних точках области D_h выполнено условие $L_h u^{(h)} \geq 0$, причем хотя бы в одной точке неравенство строгое, то $u^{(h)}$ не достигнет своего наибольшего значения ни в одной внутренней точке.

3. Рассмотрим разностную схему $L_h u^{(h)} = f^{(h)}$ вида

$$L_h u^{(h)} \equiv \begin{cases} L_h u \equiv \dfrac{u_{m+1,\,n} + u_{m,\,n+1} + u_{m-1,\,n} + u_{m,\,n-1} - 4u_{mn}}{h^2} = \varphi(mh, nh), \\ \qquad\qquad (mh, nh) \in D_h^0; \\ u_{mn} = \psi_1(s_{mn}), \qquad m = M, \quad n = 0, M; \\ \dfrac{u_{1,\,n} - u_{0,\,n}}{h} = \psi_2(n, h), \qquad n = 1, \ldots, M-1. \end{cases}$$

Эта разностная схема аппроксимирует задачу

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=\varphi(x, y), \quad (x, y)\in \mathrm{D};$$

$$u(x, y)=\psi_1(s), \qquad x=1, \quad y=0;\ 1;$$

$$\frac{\partial u}{\partial x}=\psi_2(s), \qquad x=0.$$

а) Доказать, что при любых $\varphi(mh, nh)$, $\psi_1(s_{mn})$, $\psi_2(s_{mn})$ задача $L_h u^{(h)}=f^{(h)}$ имеет единственное решение.

б) Доказать, что если $\varphi(mh, nh)$ неотрицательно, а $\psi_1(s_{mn})$, $\psi_2(s_{mn})$ не положительны, то $u^{(h)}$ не положительно.

в) Доказать, что при любых $\varphi(mh, nh)$, $\psi_1(s_{mn})$, $\psi_2(s_{mn})$ имеет место оценка вида

$$\max_{(mh, nh)\in \mathrm{D}_h} |u_{mn}| \le C\left(\max_{(mh, nh)\in \mathrm{D}_h^0} |\varphi_{mn}| + \right.$$

$$\left. + \max_{m, n} |\psi_1(s_{mn})| + \max_{n} |\psi_2(s_{0n})|\right),$$

C — некоторая постоянная, не зависящая от величины шага h. Найти C.

12.11. Порядок выполнения работы

1. Задайте правую часть уравнения, равную нулю. Получите численно решение задачи Дирихле для уравнения Лапласа всеми методами и проведите сравнение по их быстродействию.

2. Задавая различные правые части уравнения (12.1), сравните методы решения задачи Дирихле для уравнения Пуассона.

12.12. Библиографическая справка

О простейших схемах решения эллиптических уравнений см. в [1–4, 27, 32]. О других методах (расщепления, попеременно-треугольных и т. д.) см. книги [16, 17, 40–42, 44, 45]. Там же приводится исследование устойчивости некоторых методов. О вариационно-разностных схемах — в [1, 11, 16, 32, 43].

МЕТОД РАЗНОСТНЫХ ПОТЕНЦИАЛОВ

13.1. Введение

Метод разностных потенциалов (МРП) предназначен для численного решения дифференциальных и разностных краевых задач. Теоретические основы МРП изложены в [1, гл. 13]. Рассмотрим основные конструкции МРП и сформулируем задания, которые можно выполнить на компьютере самостоятельно, обращаясь к заготовленным в этой работе подпрограммам.

Для определенности и краткости изложим вариант МРП, ориентированный на вычисление решения *внутренней* задачи:

$$u_{xx} + u_{yy} - \mu u = 0, \qquad (x, y) \in \mathrm{D}^+,$$

$$u\big|_{\Gamma} = \xi(\varphi), \qquad \Gamma = \partial \mathrm{D}^+,$$

где D^+ — ограниченная область с гладкой границей Γ, заданной параметрически:

$$x(\varphi) = \rho(\varphi)\cos\varphi, \qquad y(\varphi) = \rho(\varphi)\sin\varphi;$$

(ρ, φ) — полярные координаты, $\xi(\varphi)$ — заданная функция, а параметр уравнения $\mu \geq 0$ — заданное число.

Рассмотрим также вариант МРП для вычисления решения следующей задачи (будем называть ее *внешней*):

$$u_{xx} + u_{yy} - \mu u = 0, \qquad (x, y) \in \mathrm{D}^-,$$

$$u\big|_{\Gamma} = \xi(\varphi), \qquad u\big|_{\partial D^0} = 0,$$

определенной в области $\mathrm{D}^- = \mathrm{D}^0 \setminus \overline{\mathrm{D}^+}$, где D^0 — квадрат, содержащий область D^+ строго внутри.

При больших размерах этого квадрата последняя задача аппроксимирует внешнюю задачу:

$$u_{xx} + u_{yy} - \mu u = 0, \qquad (x, y) \notin \mathrm{D}^+,$$

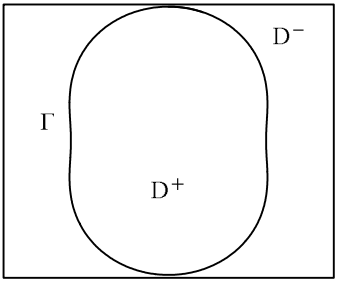
$$u\big|_\Gamma = \xi(\varphi), \qquad u(x, y) \to 0 \text{ при } x^2 + y^2 \to \infty.$$

13.2. Форма области и сетка

В этой работе для определенности используется область D^+, ограниченная кривой Γ вида

$$\rho(\varphi) = 0{,}8 + 0{,}2 \cos k\varphi,$$

где k — параметр границы. При $k = 0$ кривая Γ является единичной окружностью.



Квадрат

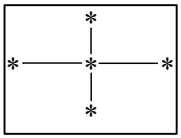
$$\mathrm{D}^0 = \{-L < x, y < L\},$$

где L — размер квадрата, содержит область D^+ строго внутри, а область D^- определяется как разность:

$$\mathrm{D}^- = \mathrm{D}^0 \setminus \overline{\mathrm{D}^+}.$$

13.3. Сеточные множества

Пятиточечный шаблон N_m с центром в точке $m = (m_1 h, m_2 h)$ представляет совокупность пяти точек квадратной сетки: $((m_1 \pm 1)h, m_2 h)$, $(m_1 h, (m_2 \pm 1)h)$ и $(m_1 h, m_2 h)$. Введем

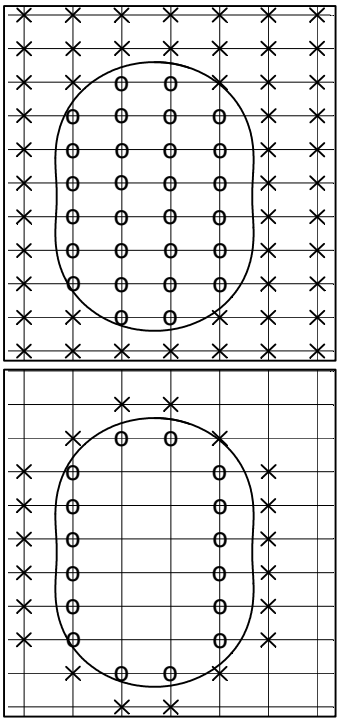


$$\mathrm{M}^0 = \{m: \ (m_1 h, m_2 h) \in \mathrm{D}^0\}$$

— множество точек квадратной сетки, которые попали внутрь квадрата D^0. Разобьем его на два подмножества:

$$\mathrm{M}^+ = \{m: \ (m_1 h, m_2 h) \in \mathrm{D}^+\}$$

— подмножество точек, попавших внутрь области D^+, и $M^- = M^0 \setminus M^+$ — подмножество точек, попавших внутрь области D^- и на кривую Г.



$$N^0 = \bigcup N_m, \quad m \in M^0$$

— множество точек квадратной сетки, которые заметаются пятиточечным шаблоном N_m при пробегании его центра по множеству M^0.

Множества N^+ и N^- строятся аналогично множеству N^0 :

$$N^+ = \bigcup N_m, \quad m \in M^+;$$

$$N^- = \bigcup N_m, \quad m \in M^-.$$

Сеточная граница определяется как пересечение $\gamma = N^+ \bigcap N^-$ и распадается на два отдельных слоя: $\gamma^+ = M^+ \bigcap \gamma$ — внутренний слой и $\gamma^- = M^- \bigcap \gamma$ — внешний слой.

13.4. Разностная вспомогательная задача

Разностная вспомогательная задача (РВЗ) имеет вид:

$$\sum_{n \in N_m} a_{mn} u_n = f_m, \qquad m \in M^0,$$

$$u_n = 0, \qquad n \in N^0 \setminus M^0,$$

коэффициенты суммирования

$$a_{mn} = \begin{cases} -\dfrac{4}{h^2} - \mu, & n = m, \\ \dfrac{1}{h^2}, & n \neq m. \end{cases}$$

Эта задача имеет единственное решение при произвольном задании правой части f_m, $m \in \mathrm{M}^0$. Решение вычисляется методом разделения переменных.

13.5. Разностный потенциал

Разностный потенциал $u_{\mathrm{N}^+} = P_{\mathrm{N}^+\gamma}\upsilon_\gamma$ $(u_{\mathrm{N}^-} = P_{\mathrm{N}^-\gamma}\upsilon_\gamma)$ с плотностью υ_γ есть сеточная функция, определенная на множестве N^+ (N^-) и равная решению РВЗ с правой частью f_m^+ (f_m^-) следующего вида:

$$f_m^+ = \begin{cases} 0, & m \in M^+, \\ \sum\limits_{n \in \mathrm{N}_m} a_{mn}\,\upsilon_n, & m \in M^-, \end{cases} \qquad f_m^- = \begin{cases} \sum\limits_{n \in \mathrm{N}_m} a_{mn}\,\upsilon_n, & m \in M^+, \\ 0, & m \in M^-, \end{cases}$$

где $\upsilon_n = \begin{cases} \upsilon_\gamma, & n \in \gamma, \\ 0, & n \notin \gamma. \end{cases}$

Если сеточная функция υ_{N^+} $(\upsilon_{\mathrm{N}^-})$ является решением однородного разностного уравнения

$$\sum_{n \in \mathrm{N}_m} a_{mn}\,\upsilon_n = 0, \qquad m \in \mathrm{M}^+ \ (\mathrm{M}^-),$$

то разностный потенциал u_{N^+} (u_{N^-}) с плотностью

$$\upsilon_\gamma = \upsilon_{\mathrm{N}^+}(\upsilon_{\mathrm{N}^-})\big|_\gamma$$

равен функции υ_{N^+} $(\upsilon_{\mathrm{N}^-})$.

13.6. Граничный проектор

Граничный проектор P_γ^+ (P_γ^-) сопоставляет сеточной функции υ_γ некоторую другую функцию $u_\gamma = P_\gamma^+\upsilon_\gamma$ $(P_\gamma^-\upsilon_\gamma)$, совпадающую со следом разностного потенциала u_{N^+} (u_{N^-}) с плотностью υ_γ.

Основные свойства граничного проектора:

$$(P_\gamma^+)^2 = P_\gamma^+ \qquad \text{и} \qquad (P_\gamma^-)^2 = P_\gamma^-.$$

Сеточную функцию υ_γ можно доопределить на всем множестве N^+ (N^-) до решения однородного разностного уравнения

$$\sum_{n\in \mathrm{N}_m} a_{mn}\,\upsilon_n = 0, \qquad m \in \mathrm{M}^+ \ \ (\mathrm{M}^-)$$

(удовлетворяющего условию $\upsilon\big|_{\mathrm{N}^0\setminus\mathrm{M}^0} = 0$) в том и только том случае, когда $P_\gamma^+\upsilon_\gamma = \upsilon_\gamma$ $(P_\gamma^-\upsilon_\gamma = \upsilon_\gamma)$.

13.7. Решение краевой задачи

Ограничимся решением уравнения Лапласа $\Delta u = u''_{xx} + u''_{yy} = 0$ внутри или вне единичной окружности. В методических целях краевые условия будем задавать такие, для которых решение заранее известно и оно совпадает с одной из следующих гармонических функций (последние две гармоничны вне точки (0, 0)):

$$u^{(1)} = -x,$$

$$u^{(2)} = x^2 - y^2,$$

$$u^{(3)} = x^3 - 3xy^2,$$

$$u^{(4)} = \frac{y}{x^2 + y^2},$$

$$u^{(5)} = \frac{xy}{(x^2 + y^2)^2}.$$

В зависимости от выбора в меню «*Данные*» точного решения $u(x, y)$ будем решать внутреннюю или внешнюю задачу. Решение разностной задачи представим в виде разностного потенциала с некоторой неизвестной плотностью. В идеале желательно найти такую плотность, которая совпадала бы со следом точного решения $\upsilon_\gamma = u(x, y)\big|_\gamma$. Такое задание плотности в работе предусмотрено и реализуется выбором в меню «*Данные*» пункта «*Сеточная плотность*». Если же выбран пункт «*Данные*

Коши», то известной считается вектор-функция

$$\mathbf{u}_\Gamma = \left(u\big|_\Gamma, \quad \frac{\partial u}{\partial n}\bigg|_\Gamma \right) = (\xi(\varphi), \ \zeta(\varphi)),$$

где

$$\xi(\varphi) = u(x(\varphi), y(\varphi)),$$

$$\zeta(\varphi) = u'_x(x(\varphi), y(\varphi))\, y'_\varphi - u'_y(x(\varphi), y(\varphi))\, x'_\varphi,$$

а плотность находится с помощью оператора вычисления плотности по формуле:

$$\upsilon_\gamma = \Pi_{\gamma\Gamma}\, \mathbf{u}_\Gamma.$$

И, наконец, когда выбран пункт меню «*Данные Дирихле*», нам известна функция $u\big|_\Gamma = \xi(\varphi)$, нужно вычислить решение задачи Дирихле. Для этого применим алгоритм МРП вычисления нормальной производной

$$\frac{\partial u}{\partial n}\bigg|_\Gamma = \zeta(\varphi),$$

а затем вычислим плотность $\upsilon_\gamma = \Pi_{\gamma\Gamma}\, \mathbf{u}_\Gamma.\ (x, y^0)$

13.8. Оператор вычисления плотности

Оператор $\Pi_{\gamma\Gamma}$ по известным данным Коши функции $u(x, y)$

$$\mathbf{u}_\Gamma = \left(u\big|_\Gamma, \quad \frac{\partial u}{\partial n}\bigg|_\Gamma \right) = (\xi(\varphi), \ \zeta(\varphi))$$

вычисляет значение плотности $\upsilon_\gamma = \Pi_{\gamma\Gamma}\, \mathbf{u}_\Gamma$.

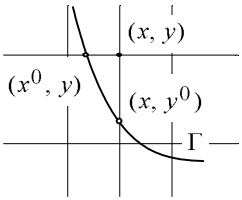


Рис. 11

1. Чтобы посчитать значение плотности в точке $(x, y) \in \gamma$, нужно на кривой Γ выбрать точку (x^0, y) или (x, y^0), из которой будет переноситься значение (см. рис. 11).

2. В выбранной точке записывается система пяти линейных уравнений относительно неизвестных значений u'_x, u'_y, а также u''_{xx} и u''_{yy}:

$$u''_{xx} + u''_{yy} = 0,$$

$$\xi'_\varphi = u'_x x_\varphi + u'_y y_\varphi,$$

$$\zeta(\varphi) = u'_x y'_\varphi - u'_y x'_\varphi,$$

$$\xi''_{\varphi\varphi} = u''_{xx} x'^2_\varphi + 2u''_{xy} x'_\varphi y'_\varphi + u''_{yy} y'^2_\varphi + u'_x x''_{\varphi\varphi} + u'_y y''_{\varphi\varphi},$$

$$\zeta'_\varphi = (u''_{xx} - u''_{yy})\, x'_\varphi y'_\varphi - u''_{xy}(x'^2_\varphi - y'^2_\varphi) + u'_x y''_{\varphi\varphi} - u'_y x''_{\varphi\varphi},$$

из которой находится значение производных u'_x, u'_y, u''_{xx} и u''_{yy} в этой точке.

3. Значение плотности вычисляется по одной из формул разложения в ряд Тейлора:

$$\upsilon_\gamma\Big|_{(x,\,y)} = u(x^0,\,y) + (x - x^0)\,u_x(x^0,\,y) + \frac{1}{2}(x - x^0)^2\,u_{xx}(x^0,\,y),$$

$$\upsilon_\gamma\Big|_{(x,\,y)} = u(x,\,y^0) + (y - y^0)\,u_y(x,\,y^0) + \frac{1}{2}(y - y^0)^2\,u_{yy}(x,\,y^0).$$

13.9. Вычисление нормальной производной

При решении внутренней задачи для вычисления нормальной производной $\left.\dfrac{\partial u}{\partial n}\right|_\Gamma = \zeta(\varphi)$ используем равенство $P_\gamma^- \upsilon_\gamma = 0$, которому должна удовлетворять плотность $\upsilon_\gamma = \Pi_{\gamma\Gamma}\,(\xi(\varphi), \zeta(\varphi))$.

1. На кривой Γ зададим N опорных узлов в точках

$$\varphi_k = \frac{2\pi}{N}\,k, \qquad (1 \le k \le N, \qquad 6 \le N \le 60).$$

Представим функцию $\zeta(\varphi)$ через значения в узлах $\zeta_k = \zeta(\varphi_k)$, используя кусочно-кубическую интерполяцию

$$\zeta(\varphi) = \sum_{k=1}^{N} \zeta_k \lambda^k(\varphi).$$

2. Преобразуем равенство $P_\gamma^- v_\gamma = 0,$ в левую часть которого неявно входят ζ_k, в систему линейных уравнений:

$$\sum_{k=1}^{N} a_\gamma^k \zeta_k = f_\gamma,$$

где $a_\gamma^k = P_\gamma^- \Pi_{\gamma\Gamma}(0, \lambda^k(\varphi)), \quad f_\gamma = -P_\gamma^- \Pi_{\gamma\Gamma}(\xi(\varphi), 0).$

3. Число уравнений этой системы, равное количеству точек сеточной границы, вообще говоря, превышает число неизвестных, которое равно количеству опорных узлов. Поэтому решение системы будем понимать в смысле метода наименьших квадратов, т. е. как набор чисел ζ_k, придающих наименьшее значение сумме

$$\sum_{n\in\gamma} \left[\sum_{k=1}^{N} a_n^k \zeta_k - f_n \right]^2.$$

13.10. Кусочно-кубическая интерполяция

Значение функции $\zeta(\varphi)$ интерполируется кубическим многочленом по значениям ζ_k в четырех ближайших к φ точках:

$$\zeta(\varphi) = \sum_{k=1}^{N} \zeta_k \lambda^k(\varphi),$$

где

$$\lambda^k(\varphi) = \frac{(\varphi-\varphi_{k-3})(\varphi-\varphi_{k-2})(\varphi-\varphi_{k-1})}{(\varphi_k-\varphi_{k-3})(\varphi_k-\varphi_{k-2})(\varphi_k-\varphi_{k-1})}, \quad \varphi_{k-2} < \varphi \le \varphi_{k-1},$$

$$\lambda^k(\varphi) = \frac{(\varphi-\varphi_{k-2})(\varphi-\varphi_{k-1})(\varphi-\varphi_{k+1})}{(\varphi_k-\varphi_{k-2})(\varphi_k-\varphi_{k-1})(\varphi_k-\varphi_{k+1})}, \quad \varphi_{k-1} < \varphi \le \varphi_k,$$

$$\lambda^k(\varphi) = \frac{(\varphi-\varphi_{k-1})(\varphi-\varphi_{k+1})(\varphi-\varphi_{k+2})}{(\varphi_k-\varphi_{k-1})(\varphi_k-\varphi_{k+1})(\varphi_k-\varphi_{k+2})}, \quad \varphi_k < \varphi \le \varphi_{k+1},$$

$$\lambda^k(\varphi) = \frac{(\varphi-\varphi_{k+1})(\varphi-\varphi_{k+2})(\varphi-\varphi_{k+3})}{(\varphi_k-\varphi_{k+1})(\varphi_k-\varphi_{k+2})(\varphi_k-\varphi_{k+3})}, \quad \varphi_{k+1} < \varphi \le \varphi_{k+2},$$

$\lambda^k(\varphi) = 0,$ если $\varphi \le \varphi_{k-2}$ или $\varphi > \varphi_{k+2}$.

13.11. Порядок выполнения работы

1. «*Сеточные множества*». Изменяя значения параметров h, L и k, а также задавая в меню «*Данные*» различные сеточные множества, четко уясните, каким образом строятся эти множества.

2. «*Разностные потенциалы*»

2.1. Задайте в меню «*Данные*» функцию $f(x, y)$ которая определит сеточную плотность $\upsilon_\gamma\big|_n = f(n_1 h, n_2 h),\ n \in \gamma$.

Вычислите с помощью меню «*Вычисления*» разностные потенциалы

$$u_{\mathrm{N}+} = P_{\mathrm{N}+\gamma}\upsilon_\gamma \qquad \text{и} \qquad u_{\mathrm{N}-} = P_{\mathrm{N}-\gamma}\upsilon_\gamma$$

с этой плотностью и просуммируйте их. Обратите внимание на то, что на сеточной границе γ выполняется равенство

$$u_{\mathrm{N}+}\big|_\gamma + u_{\mathrm{N}-}\big|_\gamma = \upsilon_\gamma.$$

Случайно ли это? Докажите, что это всегда так.

2.2. Выберите в меню «*Данные*/*След потенциала* $u_{\mathrm{N}+}$», т. е. результат действия граничного проектора P_γ^+ на плотность υ_γ. Тем смым задается другая сеточная плотность $w_\gamma = P_\gamma^+\upsilon_\gamma$. Вычислите разностные потенциалы с этой плотностью:

$$u_{\mathrm{N}+} = P_{\mathrm{N}+\gamma}w_\gamma \qquad \text{и} \qquad u_{\mathrm{N}-} = P_{\mathrm{N}-\gamma}w_\gamma$$

и объясните, почему выполняются равенства

$$u_{\mathrm{N}+}\big|_\gamma = w_\gamma \qquad \text{и} \qquad u_{\mathrm{N}-} = 0.$$

2.3. Теперь вернитесь к старой плотности υ_γ, выбрав в меню «*Данные*» пункт «$f(x, y) = \ldots$» и вычислите разностный потенциал $u_{\mathrm{N}-}$ с этой плотностью $u_{\mathrm{N}-} = P_{\mathrm{N}-\gamma}\upsilon_\gamma$.

Выберите в меню «*Данные*/*След потенциала* $u_{\mathrm{N}-}$», задав тем самым новую сеточную плотность $w_\gamma = P_\gamma^-\upsilon_\gamma$. Вычислите разностные потенциалы с этой плотностью

$$u_{\mathrm{N}+} = P_{\mathrm{N}+\gamma}w_\gamma \qquad \text{и} \qquad u_{\mathrm{N}-} = P_{\mathrm{N}-\gamma}w_\gamma$$

и объясните, почему теперь $u_{\mathrm{N}-}\big|_{\gamma} = w_{\gamma}$, а $u_{\mathrm{N}^+} = 0$.

2.4. Все это можно проделать с различными значениями параметра границы k и параметра уравнения μ, изменяя их в меню «*Параметры*». Как влияет на потенциалы u_{N^+} и $u_{\mathrm{N}-}$ увеличение параметра μ?

3. «*Краевая задача*»

3.1. Выберите в меню «*Данные*/*Сеточная плотность*» и посчитайте для всех функций $u^{(j)}(x, y)$ разностные потенциалы.

В окне «*Информация*» будет выводиться значение $\varepsilon^{(j)}$ максимума модуля отклонения потенциала от точного решения в процентах.

Покажите, что независимо от параметров сетки h и L, значения $\varepsilon^{(1)} = \varepsilon^{(2)} = \varepsilon^{(3)} = 0$. Как влияют h и L на $\varepsilon^{(4)}$ и $\varepsilon^{(5)}$?

3.2. Выберите пункт «*Данные Коши*». Посчитайте сеточную плотность и разностный потенциал для каждой $u^{(j)}(x, y)$.

Докажите, что и в этом случае $\varepsilon^{(1)} = \varepsilon^{(2)} = 0$ при любых h и L, однако теперь $\varepsilon^{(3)} \neq 0$, но уменьшается при уменьшении h, а от L не зависит. Почему это так?

3.3. И наконец, выберите пункт «*Данные Дирихле*» и для любой из функций $u^{(j)}(x, y)$ вычислите сеточную плотность. В левом нижнем окне синим цветом будет изображаться точная нормальная производная, а красным — вычисленная. Как изменяется значение $\varepsilon^{(j)}$ в зависимости от числа опорных узлов?

13.12. Библиографическая справка

О методе разностных потенциалов см. учебник [1, гл. 13] и библиографию в нем, а также монографию [12].

ПРИЛОЖЕНИЕ 1

ТЕОРЕТИЧЕСКАЯ СПРАВКА К РАБОТАМ 9–12

Разностные методы

Пусть

$$Lu = f \qquad \text{(П.1)}$$

— краткое символьное обозначение исходной дифференциальной задачи.

Пусть, далее, в рассмотрение введена сеточная область W^h и пространство U^h сеточных функций u^h.

$$L_h u^h = f^h \qquad \text{(П.1а)}$$

— соответствующее символьное обозначение разностной задачи (схемы), которой заменяется задача (П.1). Решение задачи (П.1а) может быть вычислено с помощью ЭВМ.

Замечание 1. Разностная задача (схема) может быть получена из дифференциальной посредством замены производных разностными соотношениями. Это довольно распространенный способ конструирования задачи (П.1а) и в общем случае.

Замечание 2. Узлы сетки, значения компонент сеточной функции в которых использованы при записи соотношения (П.1а), заменяющего дифференциальное уравнение (П.1), образуют конфигурацию, которую называют *шаблоном схемы* (П.1а).

Замечание 3. Индекс h в записи сеточной задачи (П.1а) обозначает в общем случае совокупность сеточных параметров, а не только пространственный шаг сетки, для которого далее использовано то же обозначение.

Решение (П.1а) рассматривается в качестве приближенного решения задачи (П.1) в узлах сетки. Ошибка этого приближения определяется как сеточная функция

$$\delta u^h = [u]^h - u^h,$$

где $[u]^h$ — значения точного решения в узлах сетки.

Введем в пространстве сеточных функций какую-либо норму $\| \, . \, \|$. *Погрешностью метода численного решения* задачи (П.1) в смысле выбранной нормы принято называть величину

$$\delta = \| \, \delta u^h \, \| .$$

Если выполнено условие $\delta = \mathrm{O}(h^p)$, то говорят, что схема имеет p-й порядок *сходимости*.

Введем в пространстве F^h правых частей f^h норму. (Индекс F^h далее будем иногда опускать). Имеем

$$\| \, f^h \, \| = \| \, f^h \, \|_{\mathrm{F}^h} \, .$$

Известно, что величина δ имеет тот же порядок, что и *погрешность аппроксимации* $\| \, \delta f^h \, \|$, где

$$\delta f^h = L_h \, [u]^h - f^h .$$

Другими словами, δf^h — невязка, характеризующая насколько не удовлетворяется задача (П.1а) при подстановке в нее решения исходной задачи (П.1).

Замечание 4. Теорема о том, что погрешность δ, т. е. погрешность метода (П.1а) для задачи (П.1), имеет тот же порядок, что и погрешность аппроксимации $\| \, \delta f^h \, \|$, справедлива в предположении, что разностная задача (П.1а) *устойчива*, т. е. ее решение существует, единственно и непрерывно зависит от возмущений входных данных равномерно по h.

Для линейных задач последнее требование означает, что существует $C = \mathrm{const}$, не зависящая от параметра h, такая, что для любых f^h

$$\| \, u^h \, \| \le C \, \| \, f^h \, \| .$$

Таким образом, вопрос о сходимости метода (П.1а) сводится к исследованию разностной схемы (П.1а) на аппроксимацию и устойчивость.

Спектральный признак устойчивости для эволюционных уравнений

Для широкого класса эволюционных (зависящих от времени) задач исследование устойчивости можно осуществить с помощью спектрального признака, который в случае разностной задачи с постоянными коэффициентами, состоит в следующем. Заменяем правую часть разностного уравнения нулем, краевую задачу — задачей Коши, функцию φ_m — гармоникой $e^{i\omega m}$ и ищем решение в следующем виде

$$u_m^p = \lambda^p \, e^{i\omega m} \tag{П.2}$$

(для задач с одной пространственной переменной), где ω — произвольное число, $0 \le \omega \le 2\pi$.

Для устойчивости разностной схемы необходимо, чтобы спектр $\lambda = \lambda(\omega)$ лежал в круге $|\lambda| \le 1 + c\tau$, где c не зависит от шага интегрирования по времени τ.

Рассматривается линейное уравнение эволюционного типа

$$\frac{\partial u}{\partial t} = Au, \tag{П.3}$$

где A — дифференциальный оператор с постоянными коэффициентами. Для (П.3) решается задача Коши: при $t = 0$ задается $u_0(x)$ и ищется решение $u(t, x)$, $t \ge 0$.

Пусть $\varphi(t, k)$ — преобразование Фурье функции $u(t, x)$, $\alpha(k)$ — частотная характеристика оператора A. Беря преобразование Фурье от правой и левой частей (П.3), получим уравнение вида

$$\frac{\partial \varphi}{\partial t} = \alpha(k)\, \varphi, \tag{П.4}$$

решение которого есть

$$\varphi(t, k) = e^{\alpha(k)\, t}\ \varphi_0(k),$$

$$\varphi_0(k) = \varphi(0, k).$$

Переход от начальных условий $u_0(x)$ к решению задачи (П.3) осуществляется по формуле

$$u(t, x) = \int\limits_{-\infty}^{+\infty} e^{\alpha(k)\, t - ikx}\ \varphi_0(k)\, dk = F_t u_0(x),$$

где F_t — оператор (типа свертки) с частотной характеристикой $\mu(t) = e^{\alpha(k)\, t}$. Задача Коши для (П.3) будет поставлена корректно в том и только том случае, когда $\mu(t)$ ограничено при всех $t > 0$.

Заменим теперь (П.3) его разностной аппроксимацией:

$$u^{n+1} = \sum_m q_m\, u^n(x - mh), \qquad \text{(П.5)}$$

где суммирование ведется по шаблону разностной схемы. Коэффициенты q_m зависят от шагов разностной сетки τ и h.

Пусть, кроме того,

$$\sum_m |\, q_m \,| < \infty$$

(что почти всегда выполняется в силу аппроксимации на решении уравнения (П.3)).

Соотношение (П.5) задает разностный оператор послойного перехода:

$$u^{n+1} = Gu^n. \qquad \text{(П.6)}$$

В силу (П.5) имеем

$$\| u^{n+1} \| \le \sum_m \| q_m\, u^n(x - mh) \| = \sum_m |\, q_m \,| \cdot \| u^n(x - mh) \|. \qquad \text{(П.7)}$$

Из этого равенства следует ограниченность разностного оператора G. Но для функции $u^n(x)$ можно написать

$$u^n(x) = G^n u_0(x).$$

Поскольку $\| G^n \| \le \| G \|^n$, то оператор, переводящий $u_0(x)$ в u^n, ограничен. Однако для устойчивости разностной задачи необходимо, чтобы метод «выдерживал» счет со сколь угодно малыми шагами сетки, т. е. при $\tau, h \to 0$.

При $\tau \to 0$ величина $n = T/\tau$ стремится к бесконечности. Если при этом норма G^n неограниченно возрастает, то сколь угодно малые ошибки в задании $u_0(x)$ могут привести к сколь угодно большим возмущениям функции $u^n(x)$, т. е. возникает вычислительная неустойчивость.

Устойчивость задачи Коши означает, что непрерывная зависимость $u^n(x)$ от $u_0(x)$ *равномерна* по τ, h.

Считаем, что фиксирована некоторая функция $h = h(\tau)$, такая, что выполняется следующее условие:

$$\lim_{\tau \to 0} h(\tau) = 0.$$

Тогда q_m зависят лишь от τ, а $G = G_\tau$.

Положим $n = T/\tau$.

Разностное уравнение (П.5) или (П.6) устойчиво, если

$$\| G_\tau^n \| \le M,$$

где M не зависит от n (но может зависеть от T).

Достаточное условие устойчивости (Неймана) есть $\| G_\tau \| \le 1 + c\tau$, где c не зависит от сеточных параметров.

Действительно, тогда

$$\| G_\tau^n \| \le \| G_\tau \|^n \le (1 + c\tau)^n = \left(1 + c\frac{T}{n}\right)^n \le e^{cT}.$$

Более сильное требование

$$\| G_\tau \| \le 1 \qquad \text{(П.8)}$$

иногда называют условием *строгой устойчивости*.

Строго устойчивые схемы могут существовать лишь для дифференциальных уравнений, решения которых подчиняются принципу максимума:

$$\| u(t, x) \| \le \| u_0(x) \|. \qquad \text{(П.9)}$$

В пространстве L_2 вычисление нормы оператора основано на равенстве

$$\| G_\tau \| = \max | \lambda(k) | ,$$

где λ — частотная характеристика оператора. Она ищется следующим образом. Подставляем $u^n(x) = e^{ikx}$ в правую часть (П.5), тогда $u^{n+1}(x) = \lambda(k)\, e^{ikx}$. Из (П.5) сразу получаем

$$\lambda(k) = \sum_m q_m\, e^{-imhk}. \qquad (П.10)$$

Обозначив $hk = \varphi$, получаем следующие условия устойчивости, основанные на вычислении спектра оператора послойного перехода:

$$| \lambda | = \left| \sum_m q_m\, e^{-im\varphi} \right| \le 1 + c\tau$$

для всех значений φ, $0 \le \varphi \le 2\pi$ (условие Неймана) или

$$| \lambda | = \left| \sum_m q_m\, e^{-im\varphi} \right| \le 1.$$

Подробнее о спектральном признаке устойчивости см. [1–4, 31, 39]. В иностранной литературе он иногда называется признаком фон Неймана [36].

ПРИЛОЖЕНИЕ 2

НЕКОТОРЫЕ РЕКОМЕНДАЦИИ ПРИ РАБОТЕ С СИСТЕМОЙ ОВМ

Редактирование

Ваша функция может содержать:

1. Числа, переменные, константу Pi = 3.1415926.

2. Алгебраические действия:

+	плюс,
–	минус,
*	умножить,
/	разделить,
^	возведение в степень.

3. Математические функции:

abs	модуль,
acos	арккосинус,
asin	арксинус,
atan	арктангенс,
cos	косинус,
cosh	косинус гиперболический,
exp	экспонента,
log	логарифм натуральный,
lg10	логарифм десятичный,
sin	синус,
sinh	синус гиперболический,
sqrt	корень квадратный,
tan	тангенс,
tanh	тангенс гиперболический.

Например:

$$F(x, y) = -1.5\text{e}{-4} / \cos(Pi * x) + \text{abs}(\lg 10(y + 1))$$

Программа использует следующие клавиши

«F1»	вызывается окно помощь;
«F2»	контекстная помощь по пункту;
«F3»	вызов предыдущего окна помощи
«F4»	список тем по алфавиту;
«F5»	учебный маршрут;
«Esc»	выход из текущего окна, выход из редактора без сохранения;
«Пробел»	редактирование строки, где это возможно
«Enter»	начало счета, вход в окно нижнего уровня, фиксация выбранного условия или метода, поставить/убрать звездочку (*) в пункте, выход из редактора с сохранением строки;
«Alt» + «x»	выход из программы.

Для перемещения по окнам, пунктам меню и редактируемой строке используйте кнопки со стрелками «←», «↑», «→», «↓» и мышь. Использовать выделенные красным цветом буквы можно следующим образом:

1) загрузите знакогенератор русского алфавита;
2) переключитесь на русский алфавит;
3) если буква заглавная, то нажмите «Shift».

Вывод списка графиков на экран в лабораторных работах 9–11

Для того, чтобы вывести одновременно несколько графиков на экран необходимо проделать следующие операции:

1. Установить метод решения задачи и все необходимые параметры для вывода 1-го графика.
2. Добавить график к списку графиков, используя меню «*Окна*»/«*Список графиков*»/«*Добавить график*», или же используйте для этой цели, находясь в данном подменю, клавишу «Ins».
3. Для каждого следующего графика повторить последовательность действий, описанную в пунктах 1 и 2.
4. Для того, чтобы вывести на экран все графики, занесенные в список, следует использовать пункт «*Перевывод списка графиков*» в меню «*Запуск*».

Если Вы находитесь в меню «*Окна*»/«*Список графиков*», Вам предоставляются следующие возможности:

добавление в список графиков текущей задачи,

вывод информации о графике с помощью подменю «*Невидимость*», «*Удаление*», «*Цвет*».

«*Невидимость*» означает, что график будет игнорироваться при выводе списка графиков. «*Удаление*» предоставляет возможность удалить график из списка, а «*Цвет*» позволяет выбирать цвет графика.

СПИСОК ЛИТЕРАТУРЫ

1. *В. С. Рябенький.* Введение в вычислительную математику. — М.: Наука – Физматлит, 2000. — 296 с.
2. *Р. П. Федоренко.* Введение в вычислительную физику. — М.: Изд-во МФТИ, 1994. — 528 с.
3. *В. И. Косарев.* 12 лекций по вычислительной математике. — М.: Изд-во МФТИ, 2000. — 224 с.
4. *В. М. Пасконов*, *В. И. Полежаев*, *Л. А. Чудов.* Численное моделирование процессов тепло- и массообмена. — М.: Наука, 1984.
5. *Д. Каханер*, *К. Моулер*, *С. Нэш.* Численные методы и программное обеспечение. — М.: Мир, 1998. — 575 с.
6. *Дж. Голуб*, *Ч. Ван Лоун.* Матричные вычисления. — М.: Мир, 1999. — 548 с.
7. Высшая математика для инженерных специальностей: Учеб. пособие для ВУЗов. — М.: РосНИИС, 1997. — кн. 3. — 125 с; кн. 4. — 98 с.
8. Сборник задач по методам вычислений / Под ред. *П. И. Монастырского*. — М.: Наука–Физматлит, 1994. — 320 с.
9. *К. И. Бабенко.* Основы численного анализа. — М.: Наука, 1986. — 528 с.
10. *О. В. Локуциевский*, *М. Б. Гавриков.* Начала численного анализа. — М.: ТОО Янус, 1995. — 581 с.
11. *Г. И. Марчук.* Методы вычислительной математики. — М.: Наука, 1989. — 608 с.
12. *В. С. Рябенький.* Метод разностных потенциалов и его приложения. — М.: Физматлит, 2002. — 496 с.
13. *Ю. С. Завьялов*, *Б. И. Квасов*, *В. Л. Мирошниченко.* Методы сплайн-функций. — М.: Наука, 1980. — 352 с.
14. *Ю. С. Завьялов*, *В. А. Леус*, *В. А. Скороспелов.* Сплайны в инженерной геометрии. — М.: Машиностроение, 1985. — 224 с.

15. *В. В. Вершинин*, *Ю. С. Завьялов*, *Н. Н. Павлов.* Экстремальные свойства сплайнов и задача сглаживания. — Новосибирск: Наука, 1988. — 102 с.

16. *Д. Ши.* Математическое моделирование задач тепло- и массообмена. — М.: Мир, 1988. — 544 с.

17. *А. А. Самарский*, *Е. С. Николаев.* Методы решения сеточных уравнений. — М.: Наука, 1978. — 592 с.

18. *В. В. Воеводин*, *Ю. А. Кузнецов.* Матрицы и вычисления. — М.: Наука, 1984. — 320 с.

19. Вычислительные процессы и системы. вып. 1. / Под ред. *Г. И. Марчука.* — М.: Наука, 1988. — 320 с.

20. *Х. Д. Икрамов.* Численные методы для симметричных линейных систем. — М.: Наука, 1988. — 160 с.

21. *Х. Д. Икрамов.* Несимметричная проблема собственных значений. — М.: Наука, 1991. — 160 с.

22. *В. И. Крылов*, *В. В. Бобков*, *П. И. Монастырский.* Вычислительные методы. Т. 1. — М.: Наука, 1976. — 304 с.

23. *В. М. Вержбицкий.* Численные методы. Линейная алгебра и нелинейные уравнения. — М.: Высшая школа, 2000. — 266 с.

24. *Т. С. Ахромеева*, *С. П. Курдюмов*, *Г. Г. Малинецкий*, *А. А. Самарский.* Нестационарные структуры и диффузионный хаос. — М.: Наука, 1992. — 544 с.

25. *Х. – О. Пайтген*, *П. Х. Рихтер.* Красота фракталов. — М.: Мир, 1993. — 176 с.

26. *А. Н. Шарковский*, *Ю. А. Майстренко*, *Е. Ю. Романенко.* Разностные уравнения и их приложения. — Киев: Наукова думка, 1986. — 279 с.

27. *Н. Н. Калиткин.* Численные методы. — М.: Наука, 1978. — 521 с.

28. *К. Деккер*, *Я. Вервер.* Устойчивость методов Рунге–Кутты для жестких нелинейных дифференциальных уравнений. — М.: Мир, 1988.

29. *Э. Хайрер*, *С. Нёрсетт*, *Г. Ваннер.* Решение обыкновенных дифференциальных уравнений. Нежёсткие задачи. — М.: Мир, 1990. — 512 с.

30. *Э. Хайрер*, *Г. Ваннер*. Решение обыкновенных дифференциальных уравнений. Жёсткие и дифференциально-алгебраические системы. — М.: Мир, 1999. — 685 с.
31. *С. К. Годунов*, *В. С. Рябенький.* Разностные схемы. — М.: Наука, 1977. — 439 с.
32. *Н. С. Бахвалов*, *Н. П. Жидков*, *Г. Л. Кобельков.* Численные методы. — М.: Наука, 1987. — 600 с.
33. *А. А. Самарский*, *Ю. П. Попов.* Разностные методы решения задач газовой динамики. — М.: Наука, 1980. — 352 с.
34. *С. К. Годунов*, *А. В. Забродин*, *М. Я. Иванов*, *А. Н. Крайко*, *Г. П. Прокопов.* Численное решение многомерных задач газовой динамики. — М.: Наука, 1976. — 400 с.
35. *М. – К. М. Магомедов*, *А.С. Холодов.* Сеточно-характеристические сеточные методы. — М.: Наука, 1984. — 320 с.
36. *Д. Андерсон*, *Дж. Таннехилл*, *Р. Плетчер.* Вычислительная гидродинамика и теплообмен. — М.: Мир, 1990. — 728 с.
37. *Э. Оран*, *Дж. Борис.* Численное моделирование реагирующих потоков. — М.: Мир, 1990. — 528 с.
38. *У. Г. Пирумов*, *Г. С. Росляков.* Численные методы газовой динамики. — М.: Высшая школа, 1987. — 232 с.
39. *А. И. Жуков.* Метод Фурье в вычислительной математике. — М.: Наука, 1992. — 176 с.
40. *А. А. Самарский.* Теория разностных схем. — М.: Наука, 1994. — 616 с.
41. *А. А. Самарский*, *А. В. Гулин.* Устойчивость разностных схем. — М.: Наука, 1973. — 240 с.
42. *А. А. Самарский.* Введение в численные методы. — М.: Наука, 1987. — 288 с.
43. *К. Ректорис*. Вариационные методы в математической физике и технике. — М.: Мир, 1985. — 590 с.
44. *А. А. Самарский*, *В. Б. Андреев.* Разностные методы для эллиптических уравнений. — М.: Наука, 1976. — 352 с.
45. *А. А. Самарский*, *Р. Д. Лазоров*, *В.Л. Макаров.* Разностные схемы для дифференциальных уравнений с обобщенными решениями. — М.: Высшая школа, 1987. — 296 с.