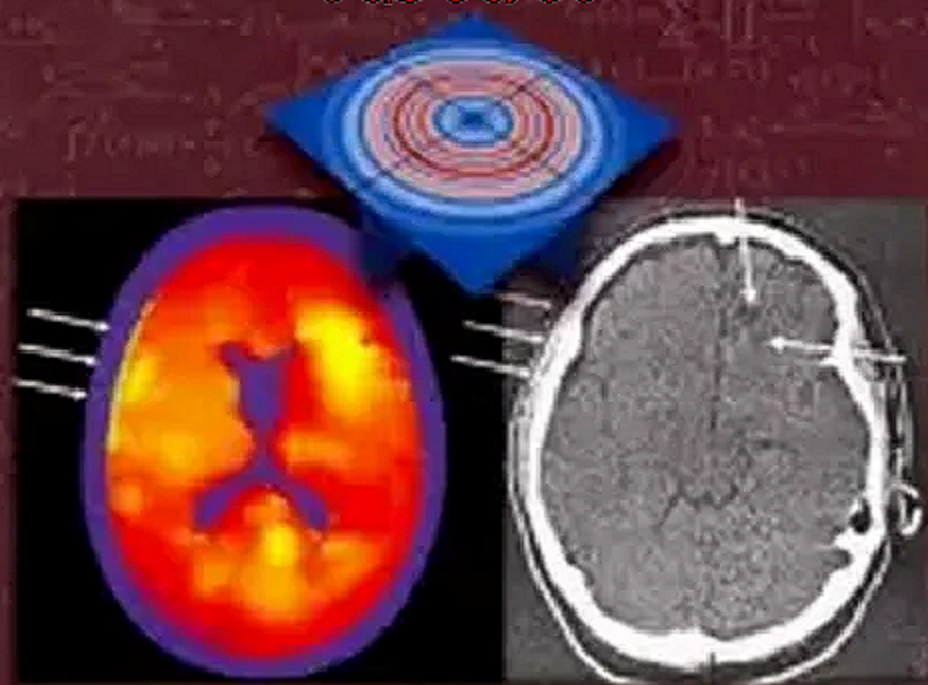


И.Б. ПЕТРОВ

ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА ДЛЯ ФИЗИКОВ

resnet



УДК 519.63 (075.8)

ББК 22.19я73

П 30



Петров И.Б. **Вычислительная математика для физиков.** — М.: ФИЗМАТЛИТ, 2021. — 376 с. — ISBN 978-5-9221-1887-3.

Рассматриваются вычислительные методы решения задач физики (в частности, механики, в том числе механики сплошных сред), а также различных прикладных задач. В книгу включены элементы функционального анализа, методы точных решений разностных уравнений, вопросы теоретического минимума по вычислительной математике для физиков и задачи для вычислительного практикума.

Для студентов университетов (факультетов физико-математического профиля) и технических вузов.



ISBN 978-5-9221-1887-3

© ФИЗМАТЛИТ, 2021

© И. Б. Петров, 2021



ОГЛАВЛЕНИЕ

Предисловие	7
Глава 1. Введение в предмет вычислительной математики	9
1.1. Из истории вычислительной математики.	9
1.2. Вычислительный эксперимент. Высокопроизводительные вычисления.	13
1.3. Особенности вычислительной математики.	18
Список литературы	22
Глава 2. Необходимые сведения из функционального анализа	25
2.1. Метрические пространства	25
2.2. Примеры метрических пространств	26
2.3. Линейные пространства	28
2.4. Примеры линейных пространств	29
2.5. Линейные нормированные пространства	30
2.6. Банаховы и гильбертовы пространства	33
2.7. Линейные операторы	34
2.8. Операторы в гильбертовом пространстве.	37
2.9. Операторные уравнения	38
2.10. Производные Гато и Фреше	39
2.11. Корректность задачи	40
Список литературы	41
Глава 3. Численные методы решения систем линейных алгебраических уравнений (СЛАУ)	42
3.1. Число обусловленности СЛАУ	42
3.2. Обусловленность СЛАУ	45
3.3. Прямые методы численного решения СЛАУ	47
3.4. Метод простых итераций (МПИ)	51
3.5. Сходимость итерационного процесса.	53
3.6. Итерационные вариационные методы последовательных приближений (итераций) численного решения СЛАУ	58
Список литературы	62
Глава 4. Приближение функций (аппроксимация функций в функциональных пространствах). Метод наименьших квадратов (МНК)	64
4.1. Постановка задачи	64
4.2. Существование и единственность полинома наилучшего приближения	66

4.3. Сходимость полинома наилучшего приближения	69
4.4. Полиномы Бернштейна	70
4.5. Аппроксимация тригонометрическими полиномами	72
4.6. Метод наименьших квадратов	72
Список литературы	78
Глава 5. Численные методы решения нелинейных алгебраических уравнений	79
5.1. Введение	79
5.2. Неподвижная точка отображения, сжимающий оператор	80
5.3. Метод простых итераций (МПИ)	82
5.4. Метод Ньютона	85
Список литературы	93
Глава 6. Методы интерполяции функций	94
6.1. Постановка задачи	94
6.2. Интерполяционный полином в форме Лагранжа	95
6.3. Интерполяционный полином в форме Ньютона	98
6.4. Конечные разности	100
6.5. Погрешность интерполяции	101
6.6. Минимизация погрешности интерполяционного процесса	105
6.7. Сходимость интерполяционного процесса	106
6.8. Другие виды интерполяции	109
6.9. Многомерная интерполяция	110
6.10. Интерполяция с кратными узлами	112
6.11. Кусочно-полиномиальная сплайн-интерполяция	113
6.12. В-сплайны	119
Список литературы	121
Глава 7. Численные методы интегрирования функций	122
7.1. Интерполяционные квадратурные формулы	122
7.2. Квадратурные формулы Чебышёва, Гаусса, Гаусса–Кристоффеля	128
7.3. Вычисления кратных интегралов	137
7.4. Вычисления интегралов с особенностями	138
7.5. Апостериорная практическая оценка погрешности квадратурных интерполяционных формул	141
Список литературы	143
Глава 8. Численное решение задач Коши для обыкновенных дифференциальных уравнений (ОДУ)	144
8.1. Методы Рунге–Кутты (нежесткие задачи)	144
8.2. Метод Рунге–Кутты (жесткие задачи)	154
8.3. Барьеры Бутчера	156
Список литературы	160



Глава 9. Численное решение задачи Коши для систем жестких обыкновенных дифференциальных уравнений	161
9.1. Понятие жестких систем ОДУ	161
9.2. Устойчивость жестких систем ОДУ	165
9.3. Нелинейные жесткие системы ОДУ	168
9.4. Численные методы решения жестких систем ОДУ	172
Список литературы	181
Глава 10. Численные методы решения краевых задач для обыкновенных дифференциальных уравнений	183
10.1. Метод фундаментальных систем	183
10.2. Краевые задачи для уравнения второго порядка	187
10.3. Метод прогонки	190
10.4. Нелинейные краевые задачи для обыкновенных дифференциальных уравнений	195
10.5. Метод Фурье	198
10.6. Методы Рунге и Галёркина	200
Список литературы	206
Глава 11. Точные решения разностных уравнений	207
Список литературы	217
Глава 12. Основные понятия теории разностных схем	218
12.1. Сходимость, аппроксимация и устойчивость методов	218
12.2. Построение разностных схем. Исследование на сходимость	222
Список литературы	238
Глава 13. Численные методы решения дифференциальных уравнений в частных производных параболического типа (уравнения диффузии, теплопроводности)	239
13.1. Однородное линейное уравнение теплопроводности	239
13.2. Нелинейное одномерное уравнение теплопроводности	244
13.3. Методы расщепления для численного решения многомерных уравнений параболических типа	247
Список литературы	256
Глава 14. Численное решение дифференциальных уравнений в частных производных гиперболического типа	257
14.1. Двухслойные разностные схемы для численного решения линейного уравнения переноса	257
14.2. Двухслойные разностные схемы для решения нелинейного уравнения переноса	271
14.3. Трехслойные разностные схемы для решения уравнения переноса	275



14.4. Разностные схемы для решения волнового уравнения и акустической системы	277
14.5. Гибридные разностные схемы	282
Список литературы	289
Глава 15. Разностные методы для численного решения уравнений эллиптического типа (уравнения электростатики, Лапласа, Пуассона)	
15.1. Постановка задачи Дирихле для уравнения Пуассона . . .	291
15.2. Итерационные методы решения задачи Дирихле для уравнения Пуассона	295
Список литературы	307
Глава 16 (дополнительная). Математические модели механики сплошных сред (МСС)	
16.1. Вывод уравнений механики сплошных сред	308
16.2. Уравнения МСС в интегральной форме	311
16.3. Система уравнений газовой динамики	312
16.4. Уравнение Навье–Стокса, описывающее течение вязкой жидкости	314
16.5. Система уравнений теории упругости	315
16.6. Нестационарная модель динамики морских и океанических течений	317
16.7. Уравнения магнитной гидродинамики (МГД)	318
16.8. Система уравнений Прандтля ламинарного пограничного слоя в несжимаемой жидкости	322
16.9. Система уравнений теории мелкой воды	323
16.10. Система уравнений акустики	324
16.11. Введение в разностные схемы газодинамики	325
16.12. Уравнение бесстолкновительной плазмы (уравнение Власова)	331
Список литературы	333
Приложение 1. Теоретические вопросы к курсу лекций по вычислительной математике (теоретический минимум) . . .	
Приложение 2. Примеры задач к вычислительному практикуму по курсу	352

Предисловие

Предмет вычислительной математики имеет большую историю. Упоминание о вычислительных методах можно найти у средневековых китайских математиков (например, схема Горнера для вычисления значений полиномов, предложенная Горнером в XIX в., была известна в Китае в XV в.).

Дальнейшее развитие вычислительная математика получила в XVII в., благодаря работам Ньютона, Эйлера, Лейбница, Лагранжа (интерполяционные полиномы, разделенные разности, первые методы численного решения обыкновенных дифференциальных уравнений, вычисления интегралов).

Методы вычислений разрабатывались Гауссом, Эрмитом, Чебышёвым (теория приближения функций в функциональных пространствах, методы интерполяции, решения систем линейных уравнений, высокоточные методы вычисления интегралов). В конце XIX – начале XX вв. бурное развитие получили высокоточные численные методы решения обыкновенных дифференциальных уравнений, разработанные в трудах Галёркина, Бубнова, Ритца, Рунге, Кутты, Крылова, Розенброка, Адамса, Бутчера и других математиков.

Примерно в середине XX в. начали развиваться численные методы решения дифференциальных уравнений в частных производных в работах Куранта, Фридрихса, Лакса, Вендроффа, Харлоу. Большой вклад в развитие этих методов внесли советские (российские) ученые: О. М. Белоцерковский, С. К. Годунов, А. А. Самарский, В. С. Рябенький, Р. П. Федоренко, Н. Н. Яненко, А. С. Холодов, Г. И. Марчук, Н. С. Бахвалов, Б. Н. Четверушкин, А. И. Толстых, В. В. Рusanов.

Актуальность этих работ была обусловлена двумя главными причинами: появлением первых электронно-вычислительных машин, а также ядерного оружия, поскольку было необходимо предсказывать последствия ядерного взрыва и разрабатывать средства его доставки.

Разумеется, в дальнейшем эти численные методы нашли свое применение в решении других (промышленных, медицинских, экологических) задач, например: климатических, геофизических (разведка полезных ископаемых), термодинамики морей

и океанов, арктических, аэрокосмических, химической физики, распространения электромагнитных волн и др.

Огромный вклад в решение сложнейших вычислительных задач внесло развитие высокопроизводительных многопроцессорных систем, для которых необходимо было адаптировать (распараллеливать) известные методы и алгоритмы. Быстрый рост их производительности приводит к возможности решения все более и более сложных задач. В настоящее время уже идет речь о создании эксафлопсного компьютера.

Автор искренне благодарит своих учителей — выдающихся ученых: академиков РАН О. М. Белоцерковского, А. С. Холодова, Б. Н. Четверушкина, докторов физико-математических наук В. С. Рябенского и Р. П. Федоренко за те бесценные знания, которые он получил от них.

Автор выражает благодарность В. С. Ароловичу, В. Д. Иванову, Д. В. Кибардиной и А. В. Фаворской за содействие в написании данной книги.



ВВЕДЕНИЕ В ПРЕДМЕТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ

1.1. Из истории вычислительной математики

Вычислительная математика, как прикладная математическая дисциплина, имеет достаточно долгую историю. По-видимому, простейшие вычислительные алгоритмы были известны еще в античные времена. Трудно представить без предварительных расчетных оценок умение измерять площади, диагонали земельных участков, строить пирамиды в Древнем Египте, огромные сооружения в Элладе, Китае, Индии, Древнем Риме и мн. др. К сожалению, до наших дней дошло немного, однако античная математика, механика, связанные с ними вычисления, создали некоторые предпосылки для развития вычислительных наук в значительно более поздние времена. Нам известны знаменитые ученые древности: Пифагор, Архимед и др., но, по-видимому, многие имена остались в забвении.

Настоящий подъем вычислительной математики происходил примерно начиная с XVII в. Развитие небесной механики, геодезии в связи с потребностями навигации и мореплавания, составлением тригонометрических функций, появление артиллерии диктовали необходимость разработок расчетных методов даже при отсутствии вычислительной техники. В эти времена появляется важнейший математический аппарат для решения многих прикладных задач — интегральное и дифференциальное исчисление, разработанное И. Ньютоном и Г. Лейбницем; появились первые дифференциальные уравнения: сначала обыкновенные, а затем и в частных производных. Для решения многих задач, сводящихся к решению дифференциальных уравнений, взятию интегралов, приближения функций и др. было необходимо разрабатывать как приближенные, так и численные методы. Так появились первые интерполяционные полиномы Лагранжа и Ньютона, первый численный метод Эйлера решения задачи Коши для обыкновенного дифференциального уравнения, формулы Ньютона–Котеса для вычисления определенных интегралов. Позже Гаусс предложил высокоточные методы численного интегрирования для достаточно гладких функций. В связи с развитием небесной механики Лапласа, механики сплошных сред

Эйлера и Лагранжа появилась необходимость в решении уравнений в частных производных, а следовательно, и в численном решении систем линейных и нелинейных алгебраических уравнений. Так появились итерационный метод Ньютона для решения нелинейных алгебраических уравнений, метод исключения Гаусса, итерационные методы решения линейных систем уравнений Якоби, Гаусса–Зейделя, ортогональные полиномы Якоби, Лежандра, Эрмита, Чебышёва.

Важнейшую роль в развитии теории приближения функций, являющейся одной из основных и в функциональном анализе, и в вычислительной математике, сыграли работы П. Л. Чебышёва (чебышёвская система функций, чебышёвские многочлены, теория равномерных приближений и др.). В начале прошлого столетия получили развитие численные методы решения обыкновенных дифференциальных уравнений (работы Галёркина, Рунге, Рунге, Крылова, Кутты, Розенброка, Ван дер Поля, Адамса, Бутчера и др.), что позволило получить численные решения многих важнейших прикладных задач. Эти методы представлены в [1, 2]. В XIX в. появились нелинейные разностные отображения в популяционной динамике (Ферхюльст), заложившие основы теории нелинейных процессов, которая начала бурно развиваться уже в XX в. и потребовала дальнейшего развития вычислительных методов.

В том же XIX в. появились знаменитая теория и соответствующая система дифференциальных уравнений в частных производных Максвелла, описывающая поведение электромагнитных полей во времени и пространстве. Разумеется, появилась и необходимость численного решения этой системы. Надо сказать, что методы решения как систем уравнений механики сплошных сред, так и системы Максвелла серьезно запаздывали, поскольку в те времена не было вычислительной техники; в XIX в. стала использоваться логарифмическая линейка, позже — механические арифмометры. Правда, нет худа без добра: в середине XX в. эти простейшие вычислительные приборы привели к появлению до того неизвестных алгоритмов параллельных вычислений. Действительно, очень нерациональным представлялось выстраивание девушек-операторов в цепочку последовательных расчетчиков на таких арифмометрах [3] (заметим также, что всем нам хорошо известно слово «алгоритм» пришло из средневековой расчетной практики — оно происходит от имени арабского врача, философа, математика Аль-Хорезми). История создания вычислительной техники довольно подробно и интересно описана в [4].

Математические основы численных методов решения уравнений в частных производных были заложены отечественными математиками А. А. Самарским, В. С. Рябеньким, Н. Н. Яненко, американскими учеными Р. Курантом, П. Лаксом, Дж. Нейманом. Ими были введены фундаментальные понятия теории разностных схем: сходимость, аппроксимация, устойчивость, доказана базовая теорема эквивалентности [5, 6]. Отметим, что создание первых разностных схем, правда, низкой точности, связано с именами известнейших отечественных физиков: Л. Д. Ландау, Н. Н. Меймана, И. М. Халатникова, которые, в отсутствие вычислительных методов решения уравнений газодинамики (система дифференциальных уравнений в частных производных) начали сами их разрабатывать. Отметим важнейшую в развитии вычислительных наук роль, которую сыграли создатели первых отечественных электронно-вычислительных машин — С. А. Лебедев и И. С. Брук; без них это развитие было бы невозможным [4].

Однако настоящий подъем вычислительной математики, как прикладной, так и фундаментальной науки, произошел в конце 40-х – начале 50-х гг. XX в. Это было связано со следующими причинами:

- интенсивное развитие новых систем вооружений, что было обусловлено ходом Великой Отечественной войны;
- ядерная программа, которую возглавлял академик И. В. Курчатов, требовавшая проведения многочисленных расчетных работ;
- начавшаяся гонка вооружений между СССР и США;
- начало развития ракетостроения, что связано со сложными аэродинамическими, баллистическими, прочностными численными расчетами;
- развитие электронной техники, систем радиосвязи, для чего необходимо было уметь численно решать систему уравнений Максвелла;
- появление первых электронно-вычислительных машин.

В дальнейшем, разумеется, появились и многие другие вычислительные задачи: климатические, космические, геофизические, задачи сейсморазведки, глобальной сейсмологии, термодинамики морей и океанов, физики атмосферных явлений, радиолокации, акустики, механики грунтов, плавающей и наземной техники, медицины, биологии, химической физики (см., например, [7–23]). Каждый из указанных разделов науки имеет большую вычислительную часть, поэтому вычислительная математика, численные

методы, высокопроизводительные вычисления, информатика уже давно стали столь же важными науками, преподаваемыми в высших учебных заведениях, как и высшая математика и физика. В современном научно-техническом мире эти науки представляются, в определенном смысле, единым циклом, владеть которым необходимо каждому научному работнику и инженеру.

Одними из первых сложнейших вычислительных задач были задачи о ядерном взрыве и об аэродинамическом обтекании затупленных тел потоком сверхзвукового газа (задача об отошедшей ударной волне). Вторую из этих задач первым решил численно, создав новый вычислительный метод, академик О. М. Белоцерковский. В это же время разработкой нового численного метода решения этой же задачи занимался сибирский математик, ныне академик С. К. Годунов. В результате этих исследований появились два численных метода, которыми пользуются до сих пор исследователи во всем мире [7, 8]. Американские вычислители решили эту задачу несколько позже. В конце 50-х гг. XX в. появился совершенно новый метод Харлоу [9], позволявший рассчитывать процессы с сильно изменяющимися границами области интегрирования (например, распыливание капли воды при падении на твердую поверхность, разрушение при взрыве и т. п.) — это был первый метод расщепления по физическим процессам. Разработка численных методов расщепления по координатным направлениям связана с именами советских математиков-вычислителей: А. А. Самарский, О. М. Белоцерковский, Г. И. Марчук, Н. Н. Яненко, В. М. Ковеня. Разработка численных методов решения задач вязкого газа проводилась научными группами академиков А. А. Дородницына, О. М. Белоцерковского, профессора Г. А. Тирского [7] и другими авторами.

Гибридные методы, позволяющие рассчитывать разрывные решения в механике сплошных сред, были предложены в работах Р. П. Федоренко, В. П. Колгана, Ван Лира; затем они получили свое развитие в работах А. Хартена (TVD-схемы), Б. Н. Четверушкина, А. С. Холодова, С. Ошера, Ч.-В. Шу, И. Б. Петрова и др. (см., например, [24–31]). А. С. Холодов предложил математически обоснованную теорию построения квазимонотонных разностных схем в пространстве неопределенных коэффициентов [30].

Численные методы высокого порядка точности были предложены в работах В. В. Русанова, С. Бурштейна, У. Риды и Т. Хилла (разрывный метод Галёркина), Э. Ф. Торо, А. И. Толстых, а также в работах А. С. Холодова, И. Б. Петрова (сеточно-характеристические методы, см., например, [25–28, 30–32]). Большую популярность приобрели методы конечных элементов,

основанные на хорошо известных вариационных методах Галёркина и Ритца [33], полностью консервативные схемы [10], метод конечных объемов [34], разрывный метод Галёркина, метод спектральных элементов (см., например, [35, 36]) для численного решения задач газодинамики, физики плазмы, магнитной гидродинамики, теории упругости.

Значительное развитие получили методы построения расчетных сеток, описание которых можно найти в [37, 52]. Для задач со значительными изменениями границ области интегрирования были предложены бессеточные методы, например, [38, 39]. Отдельную часть вычислительной математики представляют численные методы оптимизации (см., например, [40–42]), методы решения некорректных задач [42, 54]. В 60–80-х гг. XX в. получили развитие итерационные методы решения уравнений нелинейных алгебраических уравнений [50], методы решения уравнений в частных производных эллиптического типа, описанные, например, в монографии [43], среди которых особо отметим методы, разработанные в Институте прикладной математики РАН, в Институте вычислительной математики РАН, многосеточные методы Р. П. Федоренко, а также численные методы решения нелинейных уравнений параболического типа (см., например, [11–13]).

Сегодня особую популярность приобретают методы и алгоритмы, ориентированные на многопроцессорные высокопроизводительные вычислительные системы (распараллеленные алгоритмы), (см., например, [3, 29, 48, 49]). Обзор по работам в области вычислительной математики — это отдельная, довольно трудоемкая работа, поэтому во вводной главе ограничимся очень кратким их описанием.

1.2. Вычислительный эксперимент. Высокопроизводительные вычисления

Важно отметить тот факт, что в последние десятилетия появился относительно новый метод теоретического изучения сложнейших многомерных нелинейных физических процессов — численный эксперимент, т. е. исследование естественно-научных процессов методами вычислительной математики.

Обычно реализация такого эксперимента состоит из следующих этапов:

- формулировка задачи;
- построение (или выбор) математической модели исследуемого явления;

- построение (или выбор) численного метода решения определяющей системы уравнений;
- построение вычислительного алгоритма, в том числе параллельного;
- реализация расчетной программы;
- тестирование и оптимизация расчетной программы;
- проведение расчетов на вычислительных системах;
- анализ полученных расчетных результатов;
- верификация результатов;
- визуализация результатов расчета;
- уточнение математической модели и численного метода, если это необходимо;
- машинное обучение, задачи с большими данными.

Отметим некоторые важнейшие направления вычислительных исследований, в которых использование многопроцессорных высокопроизводительных систем (суперкомпьютеров) оказывает решающее влияние на развитие соответствующей отрасли:

- авиационная и ракетно-космическая техника;
- ядерная и термоядерная энергетика;
- оптимизация сложных систем;
- биотехнологии, медицина;
- разведка углеводородов;
- задачи освоения Арктики;
- создание ситуационных центров в интересах государственных структур;
- фундаментальные исследования (астрофизика, теория турбулентности, квантовая химия и др.);
- машинное обучение, задачи с большими данными.

Отметим также очень быстрый рост производительности суперкомпьютеров. Производительность, которая считалась рекордной 3–5 лет назад, в настоящее время считается уже не рекордной, хотя и весьма серьезной, достаточной для списка TOP500. Сейчас к высокопроизводительным системам предыдущего поколения условно относят компьютеры с производительностью до 100 терафлопсов. Хорошей производительностью считаются компьютеры, приблизительно реализующие 1 петафлопс (1 петафлопс = 10^3 терафлопсов), причем в ближайшие годы парк машин пополнится компьютерами с производительностью 150, 300 петафлопсов [16]. Речь идет о создании эксафлопсных вычислительных систем (2022–2024 гг.).

Однако, хотя перспективы и представляются хорошими, ситуация на самом деле не столь оптимистична. Существующий

программный продукт, как правило, ограничен диапазоном используемой производительности на одну задачу в 10 терафлопсов или 10^3 процессорных ядер. И, при наличии на сегодняшний день в мире более 10 комплексов с производительностью более 1 петафлопса, количество задач на них с одновременным использованием 10^4 ядер (или свыше 100 терафлопсов) на один вариант невелико. Как правило, такие комплексы работают в многозадачном режиме, одновременно производя расчеты 100 и более вариантов. Данная ситуация не случайна, а связана с принципиальными трудностями использования существенно многопроцессорных вычислительных систем. Эти трудности только усугубляются при переходе к вычислительным системам последующих поколений. К таковым относятся: системы на процессорах общего назначения с повышенным числом ядер на процессор; системы, использующие в качестве элементов различного рода ускорители. Наиболее применяемым типом ускорителей являются графические платы. Причинами, вызывающими переход к новым архитектурам, являются высокая стоимость и в первую очередь запредельное энергопотребление. Так, вычислительный комплекс с производительностью 1 петафлопс на четырехъядерных процессорах общего назначения имеет энергопотребление в диапазоне 3–4 МВт. Многопроцессорные высокопроизводительные системы, активное применение которых началось около четверти века тому назад, в значительной мере сместили акценты требований к алгоритмам прикладной математики. К этим требованиям можно отнести, в частности, следующие:

- 1) внутренний параллелизм, позволяющий разбить задачу на равноценные с точки зрения объема вычислений части, число которых должно быть не меньше числа используемых процессоров;
- 2) обеспечение минимизации обмена информацией между процессорами;
- 3) корректность используемых алгоритмов и моделей, что становится особенно актуальным в случае подробной пространственно-временной дискретизации задачи;
- 4) логическая простота алгоритмов;
- 5) равномерность загрузки процессоров.

Быстрый темп развития вычислительной техники приводит к периодической смене приоритетов в области создания вычислительных алгоритмов. Естественно, что адаптация логически несложных алгоритмов к архитектуре многопроцессорных систем более проста. Но это не столь критично для систем, состоящих из относительно небольшого числа процессоров. Главное

достоинство таких алгоритмов состояло в том, что при быстрой смене вычислительной техники, сопровождающейся в той или иной степени ревизией программного продукта, последнюю операцию можно было проводить достаточно быстро и безболезненно. Однако нынешняя ситуация представляется более сложной. Возможность одновременного использования большого количества процессорных ядер ($> 10^4$), усложнение архитектуры вычислительных систем (общая память для ядер внутри процессора и распределенная между процессорными узлами) приводит к тому, что применение сколь-нибудь логически сложных алгоритмов дает слишком малую эффективность параллельной обработки.

С другой стороны, хорошо распараллеливаемые простые алгоритмы обладают низкой эффективностью однопроцессорного расчета, которая к тому же, как правило, резко уменьшается при переходе к более подробной пространственной дискретизации. Эти проблемы усугубляются, когда в качестве ускорителей используются графические платы, обладающие большим количеством процессорных ядер, со сложной иерархической структурой. Рассматривая перспективы развития вычислительной техники, следует обратить внимание на то, что существующие системы, опирающиеся на использование процессоров с небольшим числом ядер, имеют естественный предел по производительности порядка 1 петафлопса. Заметное, например в 10 раз, увеличение производительности приведет к запредельной стоимости проекта, обременительной даже для экономически развитых стран. Однако самой существенной причиной ограничения станет огромное энергопотребление. Так по оценкам энергопотребление системы такого типа, обладающей производительностью 10 петафлопсов, составит около 30 МВт. Решение этой проблемы заключается в разработке и последующем использовании процессоров со все большим числом ядер. При этом соотношение энергопотребления процессора и его цены к пиковой производительности будет падать. Вычислительные системы, опирающиеся на использование существенно многоядерных процессоров общего назначения, будут одним из направлений развития суперкомпьютеров в ближайшем будущем.

Отметим, что уже сейчас существуют процессоры, содержащие несколько сот вычислительных ядер, — это графические платы. Первоначально графические платы предназначались лишь для целей визуализации. Однако в последние годы за счет создания достаточно сложного программного инструментария CUDA удалось приспособить графические платы GPU для решения задач, описываемых уравнениями математической физики. Такие

платы, используемые для решения научно-технических задач, получили название GPGPU (general purpose computing on graphics processing units). Однако применение систем, использующих GPGPU, связано с принципиальными трудностями. Не останавливаясь на архитектуре графических плат, отметим, что они создавались для решения проблем визуализации, которая допускает максимальную независимость работы вычислительных ядер. Одним из путей выхода из создавшейся ситуации является использование гибридных (гетерогенных) узлов, в которых соединены процессоры общего назначения GPU и GPGPU. В этой комбинации процессоры общего назначения берут на себя выполнение сложных логических операций, а GPGPU ведут обработку большого количества однородных потоков информации. Как показывает существующая практика, такие гибридные суперкомпьютеры обладают при той же пиковой производительности приблизительно на порядок меньшими стоимостью и энергопотреблением, чем системы одинаковой производительности на основе процессоров общего назначения. Однако даже в комбинации с процессорами общего назначения использование графических плат в качестве инструмента вычислений сталкивается с серьезными дополнительными трудностями. Во-первых, использование программистских средств CUDA достаточно сложно. Проблема заключается в том, что освоение этих средств достаточно трудоемко, написанные с помощью CUDA программы довольно объемны и непрозрачны. Это вызывает заметные проблемы при переписывании давно работающих и отлаженных программ с целью использования их для расчетов на гибридных вычислительных системах. Вторая трудность состоит в необходимости использования или создания алгоритмов, требующих в основном своем объеме переработки больших массивов однородной информации.

Заметим, что первая трудность может быть преодолена за счет дополнительных трудозатрат программистов. Кроме того, в настоящее время ведутся интенсивные разработки, позволяющие создать языки программирования более высокого уровня для графических плат.

Особую важность приобретают такие направления, как аэромеханика летательных аппаратов [7, 8, 10, 11, 44] прочность аэрокосмической техники, проектирование композиционных покрытий [14, 51], безопасность железных дорог [15], сейсмостойкость объектов атомной энергетики, глобальная сейсмика [16], задачи, связанные с освоением запасов нефти и газа в условиях Севера и Арктики (безопасность и устойчивость ледостойких платформ, нефтегазопроводов, ледоколов и судов ледового

класса, миграция крупных ледовых образований) [17]. Важнейшей проблемой, решаемой с помощью высокопроизводительных вычислительных систем, является сейсмо- и электроразведка углеводородов, особенно в шельфовых зонах российских северных морей (прямые и обратные задачи георазведки [18, 19]). Суперкомпьютерное моделирование также позволяет успешно моделировать сложнейшие процессы в теле человека, происходящие при операциях, травмах, иных процессах в медицинской практике (см., например, [20, 45, 53]), в биологических объектах [46]. Примеры численного решения сложных климатических задач представлены в [21, 47]. Вычислительным методам решения системы уравнений Максвелла (расчет элеткромагнитных полей) посвящена большая работа [22], численному решению задач физики плазмы — сборник статей [23].

1.3. Особенности вычислительной математики

Вычислительная математика является неотъемлемой частью высшей математики, компьютерных наук и науки о математическом моделировании. Однако она имеет свою специфику.

1. Дискретизация области интегрирования, работа как с непрерывными, так и с таблично заданными функциями: $f(x_i)$, $i = 0, \dots, n$.
2. В расчетах используются числа с ограниченным количеством знаков после запятой, т.е. в расчетах всегда присутствует машинная погрешность, округление, чего нет в классической математике.
3. Выбор численного метода влияет на результат решения задачи метод решения;
4. Экономичность вычислений — существенное качество вычислительного метода.
5. Обусловленность задачи, т.е. чувствительность ее решения по отношению к малым изменениям входящих данных.
6. Устойчивость численного алгоритма.
7. Приведем пример влияния конечноразрядной арифметики на результат вычисления.

Требуется вычислить корни алгебраического уравнения:

$$x^4 - 4x^3 + 8x^2 - 16x + 15, \underbrace{9 \dots 9}_8 = (x - 2)^4 - 10^{-8} = 0,$$

$$x_1 = 2,01; \quad x_2 = 1,99; \quad x_{3,4} = 2 \pm 10^2 i.$$

Если компьютер округляет свободный член в рассматриваемом уравнении до 16, то будет решаться уравнение вида

$$(x - 2)^4 = 0,$$

имеющее 4 одинаковых корня: $x_{1,2,3,4} = 2$, которые не совпадают с корнями исходного уравнения.

Рассмотрим пример влияния метода вычисления на количество арифметических действий, требуемое для решения задачи.

Допустим, мы вычисляем значение полинома

$$P_n = \sum_{j=0}^n a_j \cdot x^j$$

самым простым способом, вычисляя значение каждого слагаемого и суммируя. На это, как несложно подсчитать, потребуется $(n^2 + \left\lceil \frac{n}{2} \right\rceil)$ умножений и n сложений.

Очевидное ускорение этого алгоритма — вычисление величины x^{j+1} путем умножения величины x^j на x — приводит к заметному ускорению алгоритма, который уже требует $(2n - 1)$ умножений на n сложений.

Наиболее вычислительно экономичным алгоритмом является схема Горнера

$$P(x) = ((\dots((a_n x + a_{n-1}))x + a_{n-2})x + \dots a_0),$$

для реализации которой требуется n сложений и n умножений.

Приведем пример неустойчивого вычислительного процесса, для чего рассмотрим простую рекуррентную формулу:

$$u_{k+1} = q u_k, \quad k \geq 0, \quad u_0 = a, \quad q > 0, \quad u_k > 0.$$

При проведении вычислений на компьютере на k -м шаге вычислений возникает погрешность округления и реальное значение u_k^m будет следующим:

$$u_k^m = q(u_k + \delta_k),$$

где δ_k — машинная погрешность на k -м шаге. В таком случае вместо u_{k+1}^m имеем

$$u_{k+1}^m = q(u_k + \delta_k) = u_{k+1} + q\delta_k,$$

или

$$\delta_{k+1} = q\delta_k, \quad k = 0, 1, \dots \quad (\delta_{k+1} = u_{k+1}^m - u_{k+1}).$$

Видно, что при $|q| > 1$ машинная погрешность будет только возрастать с ростом k и вычислительный метод оказывается неустойчивым. При $|q| < 1$ этого явления не происходит.

Также приведем пример, демонстрирующий понятие обусловленности задачи. Рассмотрим задачу Коши для обыкновенного дифференциального уравнения:

$$\begin{aligned}\frac{du}{dt} &= 10u, \\ u(t_0) &= u_0, \\ t &\in [0, 1].\end{aligned}$$

Как известно из курса обыкновенных дифференциальных уравнений, его решение имеет вид:

$$u(t) = u_0 e^{10(t-t_0)}.$$

Поскольку начальное значение (u_0^*) известно приближенно, с машинной точностью, то в реальных расчетах

$$u^*(t) = u_0^* e^{10(t-t_0)};$$

соответственно, погрешность имеет вид

$$u^* - u = (u_0^* - u_0) e^{10(t-t_0)}.$$

Пусть задана точность вычисления решения $\varepsilon > 0$ на отрезке $[0, 1]$.

В таком случае

$$|u^* - u| < \varepsilon,$$

откуда получим

$$\max_{t \in [0, 1]} |u^* - u| = |u^*(1) - u(1)| = |u_0^* - u_0| e^{10(1-t_0)}.$$

Из последнего равенства получаем требования к точности задания начальных данных δ :

$$|u_0^* - u_0| < \delta,$$

т. е. при, например, $t_0 = 0$:

$$\delta \leq \varepsilon e^{-10}.$$

Это значит, что требования к точности задания начальных данных превышают требования к точности, предъявляемой к решению задачи, что представляется нереальным. Решение задачи при $t = 1$ оказывается очень чувствительным к малым вариациям начальных данных, т. е. задача плохо обусловлена.

Приведем также пример плохой обусловленности задачи из линейной алгебры. Точным решением системы из двух линейных алгебраических уравнений

$$\begin{cases} x + 10y = 11, \\ 100x + 1001y = 1101 \end{cases}$$

являются числа $x = 1$; $y = 1$.

Внесем небольшое возмущение в правую часть рассматриваемой системы:

$$\begin{cases} x + 10y = 11,01, \\ 100x + 1001y = 1101. \end{cases}$$

Тогда решением будут числа $x = 11,01$; $y = 0,00$, оно значительно отличается от решения невозмущенной системы уравнений вследствие плохой обусловленности исходной задачи.

Заметим, что на плоскости $\{x, y\}$ две прямые, соответствующие двум уравнениям системы, будут почти параллельными.

Классическим примером влияния машинной погрешности является результат вычисления функции $\sin x$ в виде сходящегося при всех x ряда Тейлора:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

При $x_1 \approx 0,5236$ (30°) получим, с точностью до четырех значащих цифр, $\sin x_1 = 0,5000$. При вычислении этой же элементарной функции при $x_2 \approx 25,66$ (1470°) с помощью того же алгоритма, имеем $\sin x_2 = 24, \dots$, что является невозможным результатом.

Выход из данной ситуации очевиден — использование тригонометрических формул приведения, т. е. использование, вообще говоря, другого алгоритма.

Теперь определимся с понятиями абсолютной и относительной погрешностей.

Определение 1.1. Абсолютной погрешностью приближенного значения некоторого приближения u^* величины x назовем величину $\Delta(x^*)$, удовлетворяющую неравенству

$$|x - x^*| \leq \Delta(x^*).$$

Относительной погрешностью называется величина $\delta(u^*)$, удовлетворяющая неравенству

$$\left| \frac{x - x^*}{x^*} \right| \leq \delta(x^*).$$

Определение 1.2. Положим, что некая величина u есть функция n независимых переменных t_1, \dots, t_n ; x^* — ее приближенное значение. *Абсолютной предельной погрешностью* назовем величину

$$A(u^*) = \sup_{\{t_1, \dots, t_n\}} |x(t_1, \dots, t_n) - x^*|.$$

Предельная относительная погрешность определяется в соответствии с формулой

$$\frac{A(x^*)}{|x^*|}.$$

Список литературы

1. Хайпер Э., Нерсет С., Ваннер Г. Решения обыкновенных дифференциальных уравнений. Нежесткие задачи. М.: Мир, 1990. 512 с.
2. Федоренко Р. П. Введение в вычислительную физику. Долгопрудный: Интеллект, 2008. 503 с.
3. Воеводин В. В., Воеводин Вл. В. Параллельные вычисления. СПб.: БХВ-Петербург. 2002.
4. Шилов В. В. Удивительная история информатики и автоматике. М.: ЭНАС, 2011. 216 с.
5. Годунов С. К., Рябенский В. С. Разностные схемы. М.: Наука. 1973. 400 с.
6. Рихтмайер Р., Мортон К. Разностные методы решения краевых задач. М.: Мир, 1972. 418 с.
7. Белоцерковский О. М. Численное моделирование в механике сплошных сред. М.: Наука. Физматлит, 1994. 442 с.
8. Годунов С. К. (ред.) Численное решение многомерных задач газовой динамики. М.: Наука, 1976. 384 с.
9. Харлоу Ф. Х. Численный метод частиц в ячейках для задач гидродинамики // Вычислительные методы в гидродинамике. М.: Мир, 1967. С. 317–342.
10. Utyuzhnikov S. V., Tirskiy G. A. Hypersonic Aerodynamics and Heat Transfer. N. Y.: Begell, 2014. 536 p.
11. Андерсен Д., Таннехилл Дж., Плетчер Р. Вычислительная гидромеханика и теплообмен. Т.1. М.: Мир, 1990. 384 с.
12. Марчук Г. И., Дымников В. П., Галин В. Я. и др. Гидродинамическая модель общей циркуляции атмосферы и океана. Новосибирск, 1975.
13. Самарский А. А. Теория разностных схем. М.: Наука, 1977. 656 с.
14. Беклемишева К. А., Васюков А. В., Ермаков А. С., Петров И. Б. Численное моделирование при помощи сеточно-характеристического метода разрушения композиционных материалов // Матем. моделир. 2016. Т. 28, № 2. С. 97–110.
15. Петров И. Б., Фаворская А. В., Хохлов Н. И. и др. Мониторинг состояния подвижного состава с помощью высокопроизводительных вычислительных систем и высокоточных вычислительных методов // Матем. моделир. 2014. Т. 26, № 7. С. 19–32.

16. Фаворская А. В., Петров И. Б., Голубев В. И., Хохлов Н. И. Численное моделирование сеточно-характеристическим методом воздействия землетрясений на сооружения // Матем. моделир. 2015. Т. 27. С. 109–120.
17. Petrov I. B. Problems of Modeling Natural and Anthropogenics Processes in the Arctic Zone of the Russian Federation // Math. Modeling a. Computer Simulation. 2019. V. 11, № 2. P. 226–246.
18. Жданов М. С. Геофизическая электромагнитная теория и методы. М.: Научный мир. 2012. 680 с.
19. Leviant V., Kvasov I., Petrov I. Numerical Modeling of Seismic Responses from Fractured Reservoirs by the Grid-characteristic Method // Geophysics Developments. 2019. № 17. 256 p.
20. Белоцерковский О. М., Холодов А. С. (отв. ред.). Медицина в зеркале информатики. М.: Наука. 2008. 242 с.
21. Яковлев Н. Г. Математическое моделирование земной системы. М.: МАКС-Пресс, 2016. 328 с.
22. Taflove A., Hagness S. C. Computational electrodynamics. Boston, London.: Artech house. 2005. 1006 p.
23. Олдер Б., Фернбах С., Ротенберг М. Вычислительные методы в физике плазмы. М.: Мир, 1974. 514 с.
24. Куликовский А. Г., Погорелов Н. В., Семёнов А. Ю. Математические вопросы численного решения гиперболических систем уравнений. М.: ФИЗМАТЛИТ, 2012. 656 с.
25. Toro E. F. Riemann Solvers and numerical Methods for Fluid Dynamics. Springer. Berlin. Heidelberg. 1997.
26. Osher S., Chahravarthy S. High resolution schemes for hyperbolic system of conservation laws // Math. Comp. 1982. V. 38. P. 339–374.
27. Shu C.-W. TVD uniformly high-order schemes for conservation law // Math. Comp. 1987. V. 49. P. 501–511.
28. Harten A. On class of high resolution schemes for hyperbolic conservation law // J. Comp. Phys. 1983. V. 49. P. 357–393.
29. Четверушкин Б. Н. Прикладная математика и проблемы использования высокопроизводительных вычислительных систем // Тр. МФТИ. 2011. Т. 3, № 4. С. 55–67.
30. Магомедов К. М., Холодов А. С. Сеточно-характеристические методы. М.: Наука. 1988. 288 с.
31. Петров И. Б., Холодов А. С. О регуляризации некоторых динамических задач механики деформируемого твердого тела сеточно-характеристическим методом // Ж. вычисл. матем. и матем. физ. Т. 27, № 8. С. 1172–1188.
32. Толстых А. И. Компактные и мультиоператорные аппроксимации высокой точности для уравнений в частных производных. М.: Наука. 2015. 350 с.
33. Батэ К.-Ю. Методы конечных элементов. М.: ФИЗМАТЛИТ, 2010. 1022 с.
34. LeVegue R. J. Finite volume Methods for Hyperbolic Problems. Cambridge: Cambridge University Press, 2011. 558 p.
35. Cochburn B. An introduction in the discontinuous Galerkin method for convection-dominated problems // SIAM J. Sci Comput. V. 16. P. 173–261.
36. Patera A. T. A spectral element method fluid dynamic laminar flow in channel expansion // Journal of Computational Physics. V. 54. P. 468–488.

37. Лисейкин В. Д. Разностные сетки. Теория приложения. Новосибирск: Изд. СО РАН, 2014. 253 с.
38. Дьяченко В. Ф. Об одном новом численном методе решения нестационарных задач газовой динамики с двумя независимыми переменными // Ж. вычисл. матем. и матем. физ. 1965., Т. 5, № 4. С. 680–688.
39. Monaghan J. J. Simulation of free surface flows with SPH // J. Comp. Phys. 1994. P. 399–406.
40. Федоренко Р. П. Приближенное решение задач оптимального управления. М.: Наука. 1978. 486 с.
41. Полак Э. Численные методы оптимизации. Единый подход. М.: Мир, 1974. 376 с.
42. Кабанихин С. И. Обратные и некорректные задачи. Новосибирск: Сибирское научное издательство. 2009. 456 с.
43. Самарский А. А., Николаев Е. С. Методы решения сеточных уравнений. М.: Наука, 1978. 591 с.
44. Чернышев С. Л. (ред.) Результаты фундаментальных исследований в прикладных задачах авиастроения М.: Наука, 2016. 511 с.
45. Беклемишева К. А., Васюков А. В., Петров И. Б. Численное моделирование динамических процессов в биомеханике сеточно-характеристическим методом // Ж. вычисл. матем. и матем. физ. 2015., Т. 55, № 8. С. 96–106.
46. Мюррей Дж. Математическая биология. Т. 1. М.: Ижевск. 2001. 774 с.
47. Марчук Г. И. Методы расщепления. М.: Наука. 1988. 263 с.
48. Карпов В. Е., Лобанов А. И. Численные методы, алгоритмы и программы. Введение в распараллеливание. М.: Физматкнига, 2014. 190 с.
49. Яковлевский М. В. Введение в параллельные методы решения задач. М.: МГУ, 2013. 327 с.
50. Хейгеман Л., Янг Д. Прикладные итерационные методы. М.: Мир, 1986. 446 с.
51. Кукуджанов В. Н. Вычислительная механика сплошных сред. М.: ФИЗМАТЛИТ, 2008. 320 с.
52. Василевский Ю. В., Данилов А. А., Липников К. Н., Чугунов В. Н. Автоматизированные технологии построения неструктурированных расчетных сеток. Т. IV. М.: ФИЗМАТЛИТ, 2016. 214 с. (Нелинейная вычислительная механика прочности / Под общ. ред. В. А. Левина: в 5 т. Т. IV).
53. Марчук Г. И. Математические модели в иммунологии. М.: Наука, 1985. 239 с.
54. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. М.: Наука, Физматлит, 1986. 287 с.
55. Petrov I. B., Favorskaya A. V., Favorskaya M. N. et al. Smart Modeling for Engineering Systems. Springer, 2019. 346 p.
56. Паттерсон Дж., Гибсон А. Глубокое обучение с точки зрения практика. М.: ДМК Пресс, 2018. 418 с.



Глава 2

НЕОБХОДИМЫЕ СВЕДЕНИЯ ИЗ ФУНКЦИОНАЛЬНОГО АНАЛИЗА

Определение 2.1. Под *функциональными пространствами* понимают пространства, элементами которых могут быть числовые последовательности или функции.

2.1. Метрические пространства

Определение 2.2. Назовем множество X *метрическим пространством*, если каждой паре его элементов x, y в соответствие поставлено число $\rho(x, y)$, (называемое *расстоянием между элементами x, y* , или *метрикой* пространства X), которое удовлетворяет следующим аксиомам метрик:

$$\rho(x, y) = \rho(y, x); \quad (2.1)$$

$$\rho(x, y) \geq 0; \quad (2.2)$$

$$\rho(x, y) \leq \rho(x, t) + \rho(t, y); \quad (2.3)$$

$$\rho(x, y) = 0 \text{ при } x = y. \quad (2.4)$$

Соотношения (2.1), (2.3) называются *аксиомами симметрии* и *треугольника* соответственно.

Определение 2.3. Величина

$$d(Y) = \sup_{x, y \in Y} \{\rho(x, y)\} \quad (2.5)$$

называется *диаметром множества Y* , где Y — подпространство пространства X , имеющее ту же метрику, что и X . При этом Y называется *ограниченным*, если $d(Y) < \infty$, т. е. \exists элемент $\bar{y} \in X$ и постоянная $C > 0$ такие, что

$$\rho(x, \bar{y}) < C \text{ для } \forall x \in Y.$$

Сферой радиуса $R > 0$ с центром в точке y_0 назовем множество, для которого выполняется $\rho(x, y_0) = R$; *шаром* — для которого выполняется $\rho(x, y_0) < R$; *замкнутым шаром* — $\rho(x, y_0) \leq R$.

2.2. Примеры метрических пространств

2.2.1. Положим $X = R$, где R — множество всех вещественных чисел (числовая прямая); в этом случае

$$\rho(x, y) = |x - y| \quad \text{для } \forall x, y \in R.$$

2.2.2. $X = R^n$, где R^n — n -мерное пространство вещественных векторов $x = \{x_1, \dots, x_n\}$, $y = \{y_1, \dots, y_n\}$, $x, y \in R^n$. В этом случае метрика может быть введена одним из следующих равенств:

$$\rho(x, y) = \max_i |x_i - y_i|, \quad (2.6)$$

$$\rho(x, y) = \sum_{i=1}^n |x_i - y_i|, \quad (2.7)$$

$$\rho(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (2.8)$$

причем каждая метрика порождает свое метрическое пространство. По этой причине правильнее было бы метрическое пространство обозначать как $\{X, \rho\}$.

2.2.3. Для пространства l_n ограниченных числовых последовательностей $x = \{x_1, \dots, x_n, \dots\}$ (для $\forall x \exists C > 0$ такая, что $|x_i| \leq C$ для $\forall i$) определим расстояние между элементами x, y так:

$$\rho(x, y) = \sup_i \{|x_i - y_i|\}. \quad (2.9)$$

2.2.4. Метрику в пространстве Чебышёва $C[a, b]$ непрерывных функций, заданных на отрезке $[a, b]$, введем с помощью равенства

$$\rho(f_1, f_2) = \max_{x \in [a, b]} |f_1(x) - f_2(x)|, \quad (2.10)$$

где $f_1(x), f_2(x) \in X$ — множество всех непрерывных функций, заданных на $[a, b]$.

Метрика в пространстве Чебышёва $C^k[a, b]$ непрерывных вместе с k -ми производными функций вводится по формуле

$$\rho(f_1, f_2) = \sum_{i=0}^k \max_{x \in [a, b]} |f_1^{(i)}(x) - f_2^{(i)}(x)|, \quad 1 \leq i \leq n, \quad (2.11)$$

где $f_1(x), f_2(x) \in X$ — множество всех непрерывных функций на $[a, b]$, имеющих на $[a, b]$ непрерывные производные до k -го порядка ($k \geq 1$) включительно.

2.2.5. Рассмотрим пространства функций, интегрируемых с первой и со второй степенью: $L_1[a, b]$ и $L_2[a, b]$. Для них вводятся метрики, соответственно:

$$\rho(f_1, f_2) = \int_a^b |f_1(x) - f_2(x)| dx, \quad (2.12)$$

$$\rho(f_1, f_2) = \sqrt{\int_a^b [f_1(x) - f_2(x)]^2 dx}, \quad (2.13)$$

где $f_1(x), f_2(x) \in X$ — множество всех непрерывных функций на $[a, b]$.

В случае пространства $L_p[a, b]$, $p \geq 1$, функций, интегрируемых с p -й степенью, имеем

$$\rho(f_1, f_2) = \left[\int_a^b |f_1(x) - f_2(x)|^p dx \right]^{1/p}. \quad (2.14)$$

Для всех введенных метрик справедливы приведенные выше аксиомы.

Определение 2.4. Точку $x \in X$ назовем *пределом бесконечной последовательности* $\{x_n\} \in X$, $n = 1, 2, \dots$ ($x = \lim x_n$, $x_n \rightarrow x$), если

$$\rho(x_n, x) \rightarrow 0 \quad \text{при} \quad n \rightarrow \infty$$

(сходимость по расстоянию).

Можно доказать следующее утверждение: последовательность $\{x_n\}$ метрического пространства сходится только к одному пределу. Две метрики на множестве элементов $x \in X$ называются *эквивалентными*, если сходимость по одной из них означает сходимость и по другой.

Определение 2.5. Точка $x \in A$, $A \subset X$ является *внутренней точкой* множества A , если $\exists \delta > 0$ такое, что шар $D_\delta(x) \subset A$.

Определение 2.6. Точка $x \in X$ является *предельной точкой* множества A , если существует последовательность $u_n \in A$ такая, что $u_n \rightarrow x$.

Множество, полученное присоединением к A всех его предельных точек, называется *замыканием множества* A , обозначается \bar{A} . Множество A называется *замкнутым*, если $\bar{A} = A$; множество A называется *открытым*, если все его точки являются внутренними. *Окрестностью* точки x называется любое открытое множество A , содержащее точку $x \in X$: например, любой шар $D_\delta(x)$.

Определение 2.7. Назовем *расстоянием от точки x до множества Y* число

$$\rho(x, Y) = \inf_{y \in Y} \{\rho(x, y)\}. \quad (2.15)$$

Элемент $u \in A$ называется *элементом наилучшего приближения* для $x \in X$, или *экстремальным элементом*, если

$$u = \arg \inf_{y \in A} \{\rho(x, y)\}.$$

Это определение является самой общей постановкой задачи о наилучшем приближении индивидуального элемента $x \in X$ фиксированным аппроксимирующим множеством A .

Определение 2.8. Последовательность $\{x_n\} \in X$ (X — метрическое пространство) называется *фундаментальной* (или *сходящейся в себе*), если для $\forall \varepsilon > 0 \exists N(\varepsilon)$ такое, что $\rho(x_n, x_m) < \varepsilon$ при $n, m \geq N(\varepsilon)$.

Можно доказать следующие утверждения:

- любая фундаментальная последовательность является ограниченной;
- если последовательность $\{x_n\}$ сходится к некоторому пределу \bar{x} , то она фундаментальна.

Определение 2.9. Метрическое пространство X называется *полным*, если всякая фундаментальная последовательность в нем имеет предел, принадлежащий этому пространству.

Рассмотренные выше пространства R^1 , R^n , $C[a, b]$, $C^k[a, b]$, l_1 , l_2 — полные.

Определение 2.10. Множество $X' \subset X$ называется *компактным*, если из \forall последовательности $\{x_n\} \in X'$ можно выделить фундаментальную подпоследовательность.

Следствие. Всякое компактное множество ограничено.

2.3. Линейные пространства

Определение 2.11. Множество L называется *линейным пространством*, если в нем определены две следующие операции.

1. Каждым двум элементам $x, y \in L$ поставлен в соответствие элемент $(x + y) \in L$ — сумма x плюс y .

2. Каждому элементу $x \in L$ и скаляру λ в соответствие поставлен элемент $\lambda x \in L$ — произведение λ на элемент x .

При этом выполняются следующие аксиоматические свойства суммы и произведения, справедливые для $\forall \lambda$ и $\forall x, y \in L$:

$$\text{коммутативность: } x + y = y + x; \quad (2.16)$$

$$\text{ассоциативность: } x + (y + z) = (x + y) + z. \quad (2.17)$$

Нулевой элемент $0 \in L$:

$$x + 0 = x;$$

$$\lambda_1 (\lambda_2 x) = (\lambda_1 \lambda_2) x;$$

элементы 0 и 1:

$$x \cdot 0 = 0, \quad x \cdot 1 = x;$$

дистрибутивность:

$$\lambda (x + y) = \lambda x + \lambda y;$$

$$(\lambda_1 + \lambda_2) x = \lambda_1 x + \lambda_2 x,$$

где λ_1, λ_2 — скаляры.

В L также определяется противоположный элемент $(-x)$; при этом:

$$-x = (-1)x; \quad x + (-x) = 1 \cdot x + (-1)x = (1 - 1)x = 0;$$

$$x - x = 0; \quad x - y = x + (-y).$$

2.4. Примеры линейных пространств

Линейными пространствами являются уже известные метрические пространства:

$$R^1, R^n, C[a, b], C^k[a, b], l_1, l_2, L_1[a, b], L_2[a, b], L_p[a, b].$$

Линейные пространства образуют также: полиномы $Q_n(x)$ степени не выше n , прямоугольные матрицы A_{ij} порядка $(i \times j)$.

Определение 2.12. Сумма вида

$$\lambda_1 x_1 + \dots + \lambda_n x_n$$

называется *линейной комбинацией элементов* $x_1, \dots, x_n \in L$ (λ_i — скаляры).

Элементы $x_i \in L$, $i = 1, \dots, n$, называются *линейно зависимыми*, если $\exists \lambda_i$, $i = 1, \dots, n$ $\left(\sum_{i=1}^n |\lambda_i| > 0 \right)$, такие, что

$$\sum_{i=1}^n \lambda_i x_i = 0.$$



Если последнее равенство выполняется только при условии $\lambda_i = 0, i = 1, \dots, n$, то элементы $x_i, i = 1, \dots, n$, назовем *линейно независимыми*.

Совокупность всех возможных линейных комбинаций вида $\sum \lambda_i x_{k_i} (x_{k_i} \in \bar{L}, \text{ где } \bar{L} \text{ — некоторое множество в } L)$ называется *линейной оболочкой* множества \bar{L} , или *линейным многообразием*. Замкнутое линейное многообразие называется *подпространством*.

Определение 2.13. Линейное пространство называется *конечномерным* (*n -мерным*), если в нем \exists система из n линейно независимых элементов, линейная оболочка которых совпадает с L ; всякая линейно независимая система из n элементов представляет базис в L ; пространство, не являющееся конечномерным, — *бесконечномерное*.

Пусть $\{e_k\}_1^n$ — базис в n -мерном линейном пространстве; тогда $\forall x \in L$ представляется в виде

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

Это представление называется *разложением элемента по базису* $\{e_k\}_1^n$; разложение является единственным; $\eta_i, i = 1, \dots, n$, — координаты элемента x в базисе $\{e_k\}_1^n$.

Определение 2.14. *Отрезком, соединяющим точки* x_1, x_2 , назовем совокупность всех точек x таких, что

$$x = (1 - t)x_1 + tx_2, \quad (2.18)$$

при всевозможных $t \in [0, 1]$; при $t > 0$ множество (2.18) называется *лучом, исходящим* из точки x_1 .

Определение 2.15. Множество $L' \subset L$ в линейном пространстве L называется *выпуклым*, если для $\forall x_1, x_2 \in L'$ отрезок (2.18) принадлежит L .



2.5. Линейные нормированные пространства

Определение 2.16. Линейное пространство X называется *нормированным*, если на множестве его элементов $x \in X$ определена вещественная функция — *норма* $\|x\|$, удовлетворяющая следующим аксиомам для любых элементов $x, y \in X$ и постоянной λ :

$$\begin{aligned} \|x\| &\geq 0; \\ \|x + y\| &\leq \|x\| + \|y\|; \\ \|\lambda x\| &= |\lambda| \cdot \|x\|; \end{aligned} \quad (2.19a)$$

$$\begin{aligned} x = 0 & \text{ при } \|x\| = 0; \\ \|x\| = 0 & \text{ при } x = 0 \end{aligned} \quad (2.196)$$

(следствие третьей аксиомы).

В нормированном пространстве можно ввести расстояние между элементами:

$$\rho(x, y) = \|x - y\|, \quad (2.20)$$

т. е. любое нормированное пространство является метрическим; следовательно, все определения, введенные для метрических пространств, справедливы и для пространств нормированных.

Приведем примеры некоторых норм, аналогичных метрикам (2.6)–(2.8), соответственно.

Нормы векторов в пространстве R^n :

$$\begin{aligned} \|x\|_1 &= \max |x_i|, \quad 1 \leq i \leq n, \\ \|x\|_2 &= \sum_{i=1}^n |x_i|, \\ \|x\|_3 &= \sqrt{\sum_{i=1}^n x_i^2}. \end{aligned} \quad (2.21)$$

Нормы в пространстве l_n :

$$\|x\| = \sup_i |x_i|, \quad 1 \leq i \leq n. \quad (2.22)$$

В пространствах $C[a, b]$, $C[\Omega]$, $C^k[a, b]$, $C^k[\Omega]$ вводятся следующие нормы соответственно:

$$\begin{aligned} \|f\| &= \max_{x \in [a, b]} |f(x)|; \\ \|f\| &= \max_{x \in \Omega} |f(x)|; \\ \|f\| &= \sum_{i=0}^k \max_{x \in [a, b]} |f^{(i)}(x)|; \\ \|f\| &= \sum_{0 \leq |\alpha| \leq k} \max_{x \in \Omega} |D^\alpha f|, \end{aligned} \quad (2.23)$$

где

$$D^\alpha f(x) = \frac{\partial^{(\alpha)} f(x_1, \dots, x_n)}{\partial x_1^{\alpha_1} \cdot \partial x_2^{\alpha_2} \cdot \dots \cdot \partial x_n^{\alpha_n}}, \quad D^0 f(x) = f(x),$$

α — мультииндекс (вектор) с целыми неотрицательными компонентами:

$$\alpha = \{\alpha_1, \dots, \alpha_n\}, \quad |\alpha| = \sum \alpha_i, \quad \alpha_i \geq 0,$$

или в $C^k[a, b]$:

$$\|f\| = \max \left\{ \max_{x \in [a, b]} |f(x)|, \max_{x \in [a, b]} |f'(x)|, \dots, \max_{x \in [a, b]} |f^k(x)| \right\}. \quad (2.24)$$

Нормы в $L_1[a, b]$, $L_2[a, b]$, $L_p[a, b]$ соответственно будут:

$$\begin{aligned} \|f\|_{L_1} &= \int_a^b |f(x)| dx; \\ \|f\|_{L_2} &= \sqrt{\int_a^b |f(x)|^2 dx}; \\ \|f\|_{L_p} &= \sqrt[p]{\int_a^b |f(x)|^p dx}. \end{aligned} \quad (2.25)$$

В пространстве полиномов

$$P_n = \sum_{k=0}^n a_k x^k$$

степени не выше n может быть введена норма в соответствии с формулой

$$\|P_n\| = \sum_{k=0}^n |a_k|. \quad (2.26)$$

Все эти нормы удовлетворяют введенным выше аксиомам.

Определение 2.17. Нормы $\|x\|_1$, $\|x\|_2$ назовем *эквивалентными*, если \exists постоянные $c_1 \geq 0$ и $c_2 \geq 0$ такие, что для $\forall x \in X$ выполняется

$$C_1 \|x\|_1 \leq \|x\|_2 \leq C_2 \|x\|_1; \quad (2.27)$$

в этом случае

$$C_2^{-1} \|x\|_2 \leq \|x\|_1 \leq C_1^{-1} \|x\|_2; \quad (2.28)$$

если $x_n \xrightarrow[n \rightarrow \infty]{} \bar{x}$ в $\|\cdot\|_1$, то $x_n \xrightarrow[n \rightarrow \infty]{} \bar{x}$ и в $\|\cdot\|_2$, и обратно.

Можно показать, например, что введенные три нормы в пространствах R^n эквивалентны.

2.6. Банаховы и гильбертовы пространства

Определение 2.18. Полное линейное нормированное пространство называется *банаховым* (B).

Поскольку пространства R^1 , R^n , $C[a, b]$, $C^k[a, b]$, l_1 , l_2 , $L_1[a, b]$, $L_2[a, b]$, $L_p[a, b]$ являются полными линейными нормированными (следовательно, и метрическими) пространствами, то они также являются и банаховыми пространствами.

В банаховом пространстве B определена линейная комбинация элементов x_i , $i = 1, \dots, n$:

$$\sum_{i=1}^n \lambda_i x_i,$$

поскольку B — линейное пространство.

Определение 2.19. Линейное вещественное пространство называется *евклидовым*, если любой паре его элементов x, y поставлено в соответствие вещественное число (x, y) , называемое *скалярным произведением*, для которого справедливы следующие аксиомы:

$$(x, x) \geq 0, \quad (x, x) = 0, \text{ если } x = 0; \quad (2.29)$$

$$(x + y, z) = (x, z) + (y, z); \quad (2.30)$$

$$(x, y) = (y, x); \quad (2.31)$$

$$(\lambda x, y) = \lambda (x, y). \quad (2.32)$$

Норма в нормированном евклидовом пространстве определена следующим образом:

$$\|x\| = \sqrt{(x, x)};$$

для нее справедливо неравенство Коши–Буняковского:

$$|(x, y)| \leq \|x\| \cdot \|y\|. \quad (2.33)$$

Определение 2.20. Комплексное линейное пространство \bar{L} называется *унитарным*, если \forall паре его элементов поставлено в соответствие комплексное число (x, y) , которое называется *скалярным произведением*.

При этом выполняются аксиомы (2.27), (2.28) и аксиома $(x, y) = (\overline{y}, \overline{x})$; здесь черта соответствует комплексно сопряженному числу.


Определение 2.21. Полное евклидово или полное унитарное пространство называется *гильбертовым* (H).

H является линейным, нормированным, банаховым (следовательно, и метрическим) одновременно; для него справедливы

все введенные выше определения, понятия и аксиомы для метрических, линейных, нормированных и банаховых пространств.

Определение 2.22. Система $\{x_i\}$, $i = 1, \dots, n$, элементов гильбертового пространства называется *ортонормированной*, если для $\forall i, j$ справедливо равенство $(x_i, x_j) = \delta_{ij}$. Если в H \exists элемента $x \neq 0$, ортогонального всем x_i , то такая система называется *полной*.

Примерами гильбертовых пространств являются:
 R^n со скалярным произведением

$$(x, y) = \sum_{i=1}^n x_i y_i,$$


l_2 со скалярным произведением

$$(x, y) = \sum_{i=1}^{\infty} x_i y_i,$$

$L_2[a, b]$ со скалярным произведением

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

В гильбертовом пространстве имеет место равенство параллелограмма

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2),$$

поскольку

$$\begin{aligned} \|x + y\|^2 + \|x - y\|^2 &= (x + y, x + y) + (x - y, x - y) = \\ &= (x, x) + 2(x, y) + (y, y) + (x, x) - 2(x, y) + (y, y) = \\ &= 2(\|x\|^2 + \|y\|^2). \end{aligned}$$

2.7. Линейные операторы

Определение 2.23. *Линейным отображением (или линейным оператором)* F линейного пространства X в линейное пространство Y ($F: X \rightarrow Y$) называется отображение, для любых двух элементов которого $x, x' \in X$ и постоянных λ, λ' справедливо

$$F(\lambda x + \lambda' x') = \lambda \cdot F(x) + \lambda' \cdot F(x'). \quad (2.34)$$

Множество всех линейных отображений (операторов) обозначим $S(Y, X)$.

При естественном определении сложения элементов этого множества

$$(F_1 + F_2)x = F_1(x) + F_2(x),$$

оно образует линейное пространство (или линейное пространство операторов).

Пусть $F: X \rightarrow Y$ — линейное отображение. Тогда множество элементов $\{x \in X: F(x) = 0\}$ называется *ядром* отображения F :

$$\ker F = \{x \in X: F(x) = 0\} \quad (2.35)$$

Теорема 2.1. Для того чтобы отображение $F: X \rightarrow Y$ было взаимно однозначным, необходимо и достаточно, чтобы его ядро состояло только из нулевого элемента $\ker F = 0$.

Определение 2.24. Говорят, что два линейных пространства *изоморфны*, если между их элементами можно установить взаимно однозначное соответствие, сохраняющее алгебраические операции, т. е. если $x \leftrightarrow y, x' \leftrightarrow y'$, то:

$$\begin{aligned} x + x' &\leftrightarrow y + y', \\ \lambda x &\leftrightarrow \lambda y. \end{aligned}$$

Определение 2.25. Пусть оператор (или отображение) $A: X \rightarrow Y$, где X, Y — два банаховых пространства, ставящий в соответствие каждому элементу $x \in X$ элемент $y \in Y$, определен на множестве $D(A) \in X$ (область определения оператора A). Оператор $A: X \rightarrow Y$ называется *линейным*, если $D(A)$ — линейное пространство:

$$A(\lambda x_1 + \mu x_2) = \lambda Ax_1 + \mu Ax_2 \quad (2.36)$$

для $\forall x_1, x_2 \in D(A)$ и \forall постоянных λ, μ ; если $Ax = 0$, то A — *нулевой оператор*.

При этом если

$$y = Ax, \quad x \in D(A), \quad y \in G(A), \quad (2.37)$$

то множество элементов $G(A) = \{y \in Y, y = Ax, x \in D(A)\}$ называется *множеством значений* оператора A ; y — *образом элемента x* , x — *прообразом элемента y* ; при этом $G(A)$ есть образ $D(A)$: $G(A) = A(D)$.

В пространстве R^n отображение $Ax = y$, где $A = \{a_{ij}\}_{i,j=1}^n$ — квадратная матрица $n \times n$, а x, y — векторы-столбцы из R^n , задает оператор A .

Норма оператора задается следующим соотношением (при этом оператор называется *ограниченным*):

$$\|A\| = \sup_{x \in X} \frac{\|Ax\|}{\|x\|}, \quad (2.38)$$

откуда

$$\|Ax\| \leq \|A\| \cdot \|x\|. \quad (2.39)$$

Здесь (2.39) — согласованная, (2.38) — подчиненная норма оператора $\|A\|$.

Суммой операторов $A + B = C$; $C: X \rightarrow Y$ назовем оператор, определенный для $\forall x \in X$ в соответствии с равенством

$$Cx = Ax + Bx; \quad (2.40)$$

при этом норма C будет

$$\|Cx\| = \|Ax + Bx\| \leq \|Ax\| + \|Bx\| \leq (\|A\| + \|B\|) \cdot \|x\|,$$

т. е. $\|C\| \leq \|A\| + \|B\|$, с учетом (2.38) и (2.39).

Если $A: X \rightarrow Y$, $B: Y \rightarrow Z$, то на множестве $D(A)$ определен оператор $C = BA$; $C: X \rightarrow Z$ — произведение операторов B и A , такой, что

$$\begin{aligned} Cx &= BAx = B(Ax); \\ \|Cx\| &= \|BAx\| \leq \|B\| \cdot \|Ax\| \leq \|B\| \cdot \|A\| \cdot \|x\|, \\ \|C\| &\leq \|B\| \cdot \|A\|. \end{aligned} \quad (2.41)$$

Величина $\rho(A) = \lim_{k \rightarrow \infty} \|A^k\|^{1/k}$ называется *спектральным радиусом* оператора A ; $\rho(A)$ не зависит от выбора норм, причем $\rho(A) = \inf_{\|\cdot\|} \|A\|$.

Определение 2.26 (действия с операторами в B -пространствах). Пусть A, B — множества линейных операторов в нормированном пространстве X , Y — банахово пространство, на котором определены операции

$$(A + B)x = Ax + Bx, \quad (\lambda A)x = \lambda Ax;$$

тогда $(A + B)$, λA также будут линейными операторами, т. е. множество всех линейных операторов является линейным пространством, поскольку все соответствующие аксиомы для него выполняются.

Теорема 2.2. *Множество линейных операторов, определенных всюду в нормированном пространстве X со значениями в банаховом пространстве Y является нормированным пространством.*

Можно доказать, что это пространство $S(X, Y)$ является банаховым.

Сходимостью последовательности операторов A_n в этом пространстве, или равномерной сходимостью к оператору A называется сходимость по норме:

$$\|A_n - A\| \xrightarrow{n \rightarrow \infty} 0$$

В пространстве операторов $\bar{l}(X_0, Y)$ определены *единичный оператор I* такой, что $Ix = x$ для $\forall x \in X$, $\|I\| = 1$, и *степень оператора A^k* :

$$A: A^2 = A \cdot A; \quad A^3 = A \cdot A^2, \dots, A^k = A \cdot A^{k-1}; \quad A^0 = I.$$

Тогда $\|A^k\| \leq \|A\|^k$ и *полиномы от операторов* определяются следующим образом:

$$P_N(A) = \sum_{k=0}^N a_k A^k; \quad (2.42)$$

Определяются *функции от операторов*: например,

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

Определение 2.27. Оператор A^{-1} называется *обратным* к A , а операторы A и A^{-1} *взаимно обратными*, если $A: X \rightarrow Y$ и $\exists A^{-1}: Y \rightarrow X$, определенный на $E(A)$, принимающий значения в $D(A)$ такой, что:

$$A^{-1}Ax = x, \quad x \in D(A)$$

$$AA^{-1}y = y, \quad y \in E(A).$$

При этом:

$$A^{-1}A = I, \quad (AB)^{-1} = B^{-1} \cdot A^{-1}.$$

2.8. Операторы в гильбертовом пространстве

Определение 2.28. Оператор A^* называется *сопряженным* оператору A , если

$$\forall x, y \in H \quad (Ax, y) = (x, A^*y).$$

Для A, A^* выполняется

$$\|A\| = \|A^*\|;$$

оператор A называется *кососимметрическим*, если $A^* = -A$, и *нормальным*, если $AA^* = A^*A$; \forall оператор A представим в виде суммы самосопряженного A' и кососимметрического A'' операторов:

$$\begin{aligned} A &= A' + A'', \\ A' &= \frac{1}{2}(A + A^*), \\ A'' &= \frac{1}{2}(A - A^*), \end{aligned}$$

кроме того: $(Ax, x) = (A'x, x)$, $(A''x, x) = 0$.

Определение 2.29. Числовой радиус оператора A определяется как число:

$$\bar{\rho}(A) = \sup_{\|x\|=1} \{(Ax, x)\} = 1, \quad x \in H.$$

Определение 2.30. Оператор A , действующий в H , называется *положительным*, если $(Ax, x) > 0$; *неотрицательным*, если $(Ax, x) \geq 0$; *положительно определенным*, если $(Ax, x) \geq \delta(x, x)$, $\delta > 0$, $x \in H$; неравенство $A < B$ ($A \leq B$) означает:

$$A - B < 0 \quad (A - B \leq 0).$$

Определение 2.31. Числа $\varepsilon_1 = \inf_{\|x\|=1} \{(Ax, x)\}$ и $\varepsilon_2 = \sup_{\|x\|=1} \{(Ax, x)\}$ называются *границами оператора* A .

2.9. Операторные уравнения

Линейные алгебраические уравнения, линейные интегральные уравнения, линейные дифференциальные уравнения, обыкновенные уравнения, уравнения в частных производных могут быть записаны в операторном виде:

$$Ax = y, \tag{2.43}$$

где A — линейный оператор, $y \in Y$, $x \in X$, $A \in S(X, Y)$.

В этом случае решение (2.43) может быть также записано в операторном виде:

$$x = A^{-1}y. \tag{2.44}$$

Разумеется, здесь возникают вопросы существования, единственности, корректности и разработки методов решения, корректности задачи (2.43). Изучение свойств этого уравнения является одной из основных задач функционального анализа, создание

методов решения — одной из центральных проблем вычислительной математики, породившей широкий спектр численных методов.

Определение 2.32. Если \exists число λ и элемент $x \in X$ такие, что $Ax = \lambda x$ (A — оператор), то x называется *собственным элементом (собственной функцией)* оператора A , λ — *собственным значением*.

Определение 2.33. Пусть A — линейный ограниченный оператор и \exists обратный линейный ограниченный оператор A^{-1} (оба оператора определены в нормированных пространствах).

Тогда

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

называется *числом обусловленности* оператора A .

Определение 2.34. Пусть X — полное метрическое пространство с метрикой $\rho(x, y)$; $x, y \in \Omega$, $\Omega \subset X$; пусть в Ω задан оператор P , переводящий множество Ω в себя ($P: \Omega \rightarrow \Omega$). Элемент $y \in \Omega$ называется *неподвижной точкой* отображения (оператора) P , если

$$y = P(y), \quad (2.45)$$

т. е. неподвижная точка является решением (2.45).

Определение 2.35 (принцип сжимающих отображений). Оператор P назовем *сжимающим отображением (оператором сжатия)* в Ω , если для $\forall x, y \in \Omega$ выполняется условие (условие Липшица):

$$\rho(P(x), P(y)) \leq \alpha \cdot \rho(x, y), \quad (2.46)$$

где $\alpha \in [0, 1)$ — константа.

2.10. Производные Гато и Фреше

Определение 2.36. Пусть X, Y — два нормированных пространства над полем R^1 , $F: X \rightarrow Y$. Если для $x, h \in X$ \exists предел (сходимость по норме Y), т. е.

$$dF(x, h) = \left. \frac{dF(x + t \cdot h)}{dt} \right|_{t=0} = \lim_{t \rightarrow 0} \frac{F(x + t \cdot h) - F(x)}{t},$$

то этот предел называется *дифференциалом Гато* (слабый дифференциал) оператора F в точке x на приращении h .

Ограниченный оператор $F'(x)$, определяемый равенством

$$dF(x; h) = F'(x) \cdot h,$$

называется *производной Гато (слабой производной)* оператора F в точке x .

Определение 2.37. Пусть X, Y — два вещественных банаховых пространства, $F: X \rightarrow Y$ — оператор, действующий из X в Y .

Производной Фреше в точке $x \in X$ назовем линейный оператор $A: X \rightarrow Y$ такой, что для $\forall h \in X$ выполняется

$$F(x+h) - F(x) = Ah + r_0(x, h),$$

где $r_0(x, h)$ — остаточный член, для которого верно соотношение

$$\frac{\|r_0(x, h)\|_Y}{\|h\|_X} \xrightarrow{\|h\|_X \rightarrow 0} 0.$$

Оператор F называется *сильно дифференцируемым*, а линейная часть приращения Ah — *дифференциалом Фреше* функции F .

2.11. Корректность задачи

Задача поиска элемента $x \in X$ в соответствии с данным операторным уравнением $Ax = y$ называется *корректно поставленной по Адамару* (или *корректной*), если:

- 1) для $\forall y \in Y \exists$ решение $x \in X$;
- 2) это решение единственное в X ;
- 3) решение $x \in X$ уравнения $Ax = y$ непрерывно зависит от правой части $y \in Y$ (малые изменения в исходных данных поставленной задачи, т.е. в y , вызывают малые возмущения решения x).

Или: для $\forall \varepsilon > 0 \exists \delta = \delta(\varepsilon)$ такое, что для $\forall y, y' \in Y$ таких, что $\rho(y, y') < \varepsilon$, выполняется $\rho(x, x') < \delta$; $x, x' \in X$.

Пример (корректная и некорректная задачи). Пусть

$$\frac{\partial u}{\partial t} = c \frac{\partial^2 u}{\partial x^2}, \quad 0 \leq x \leq a$$

— одномерное линейное однородное уравнение теплопроводности при нулевых граничных условиях:

$$u(0, t) = u(1, t) = 0.$$

Решение этого уравнения представляется в виде ряда

$$u(t, x) = \sum_{n=1}^{\infty} \alpha_n \exp(-\lambda_n t) \cdot \sin \frac{\pi n x}{\alpha},$$

$$\lambda_n = \frac{c\pi^2 n^2}{\alpha^2},$$

где λ_n — коэффициенты Фурье.

При $t > 0$ гармоники затухают (задача корректная); при $t < 0$ (обратная задача теплопроводности) гармоники растут неограниченно (задача некорректная).

Список литературы

1. *Лебедев В.И.* Функциональный анализ и вычислительная математика. М.: ФИЗМАТЛИТ, 2004. 296 с.
2. *Ильин В.П.* Численный анализ. Ч. 1. Новосибирск: ИВМиМГ СО РАН, 2004. 334 с.
3. *Коллатц Л.* Функциональный анализ и вычислительная математика. М.: Мир, 1969. 448 с.



ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ СИСТЕМ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ (СЛАУ)



3.1. Число обусловленности СЛАУ

Рассматривается система линейных алгебраических уравнений вида:

$$\mathbf{Ax} = \mathbf{f}, \quad (3.1)$$

где

$$\mathbf{x} = \begin{Bmatrix} x_1 \\ \vdots \\ x_n \end{Bmatrix} \in R^n, \quad \mathbf{f} = \begin{Bmatrix} f_1 \\ \vdots \\ f_2 \end{Bmatrix} \in R^n \quad (3.2)$$

— векторы-столбцы (\mathbf{x} — искомое решение, \mathbf{f} — правая часть), принадлежащие n -мерному евклидову пространству R^n , $\mathbf{A} = \{a_{ij}\} \in M(n \times n)$ — квадратная матрица $n \times n$, $M(n \times n)$ — линейное нормированное пространство квадратных матриц.

Теорема 3.1 (Адамара о невырожденности матрицы). Матрица $\mathbf{A}(n \times n)$ является невырожденной, т.е. $\det \mathbf{A} \neq 0$, если для нее выполняется условие диагонального преобладания:

$$|a_{ii}| > \sum_{j=1, j \neq i}^n |a_{ij}|, \quad i = 1, 2, \dots, n.$$

Доказательство (от противного). Пусть \mathbf{A} — вырожденная матрица; в этом случае СЛАУ

$$\mathbf{Ax} = \mathbf{0}$$

имеет ненулевое решение

$$\mathbf{x} = \{x_1, \dots, x_n\},$$

т.е.

$$\sum_{j=1}^n a_{ij} x_j = 0, \quad i = 1, 2, \dots, n.$$

Положим, что \exists элемент x_k такой, что

$$|x_k| = \max_{1 \leq i \leq n} |x_i| > 0.$$



Для строки с номером k имеем:

$$a_{kk}x_k = - \sum_{\substack{j=1, \\ i \neq k}}^n a_{kj}x_j,$$

откуда

$$|a_{kk}| \cdot |x_k| \leq \sum_{\substack{j=1 \\ i \neq k}}^n |a_{kj}| \cdot |x_j| \leq \sum_{\substack{j=1 \\ i \neq k}}^n |a_{kj}| \cdot |x_k|,$$

или

$$|a_{kk}| \leq \sum_{\substack{j=1 \\ i \neq k}}^n |a_{kj}|.$$

Однако последнее неравенство противоречит условию строгого диагонального преобладания; следовательно матрица \mathbf{A} невырождена.

Теорема 3.2 (Гершгорина). *Все собственные значения матрицы \mathbf{A} ($n \times n$) лежат в объединении кругов Гершгорина:*

$$|\lambda - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|, \quad i = 1, 2, \dots, n.$$

Доказательство. Положим, что λ — собственное значение матрицы \mathbf{A} . В этом случае матрица $\mathbf{B} = (\mathbf{A} - \lambda \mathbf{E})$ вырожденная: $\mathbf{A}\omega = \lambda\omega$; $(\mathbf{A} - \lambda \mathbf{E})\omega = \mathbf{0}$, где ω — собственный вектор \mathbf{A} и критерий Адамара для нее не выполняется, т. е.

$$|b_{ii}| = |\lambda - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|,$$

где $b_{ii} = (\lambda - a_{ii})$ — диагональные элементы матрицы $\mathbf{B} = (\mathbf{A} - \lambda \mathbf{E})$. Объединяя весь спектр собственных значений λ_i матрицы \mathbf{A} , получаем утверждение теоремы, которая дает простой, но очень приближенный способ нахождения границ спектра матрицы.

В дальнейшем будем полагать, что $\det \mathbf{A} \neq 0$, решение существует и единственно. Из линейной алгебры известно правило Крамера нахождения решения (3.1). К сожалению, это правило не применимо к системам линейных алгебраических уравнений из-за очень больших затрат машинного времени. Поэтому для численного решения реальных задач используются два типа методов: прямые и итерационные (или методы последовательных приближений).

С помощью прямых методов можно получить точное решение задачи за конечное количество арифметических действий на идеальном компьютере (бесконечноразрядном), который, очевидно, не существует, поэтому численные решения всегда будут приближенными. Итерационные методы позволяют искать решение системы как предел последовательности $x_k \xrightarrow{k \rightarrow \infty} \mathbf{U} \in R^n$ (\mathbf{U} — точное решение (3.1)).

Рассмотрим, в первую очередь, способы оценки погрешностей, образующихся при численном решении систем линейных алгебраических уравнений. В функциональном анализе вводятся нормы векторов в n -мерном линейном векторном пространстве (евклидово пространство R^n):

$$\|\mathbf{x}\|_1 = \max_{1 \leq i \leq n} |x_i|, \quad (3.3)$$

$$\|\mathbf{x}\|_2 = \sum_{i=1}^n |x_i|, \quad (3.4)$$

$$\|\mathbf{x}\|_3 = \sqrt{(\mathbf{x}, \mathbf{x})} = \sqrt{\sum_{i=1}^n x_i^2}, \quad (3.5)$$

а также подчиненные им нормы матриц $\mathbf{A} (n \times n)$ в линейном пространстве матриц, в соответствии с равенством:

$$\|\mathbf{A}\| = \sup_{\mathbf{x} \in R^n} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|}. \quad (3.6)$$

При помощи несложных алгебраических преобразований можно получить эти нормы, в соответствии с (3.4):

$$\|\mathbf{A}\|_1 = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|, \quad (3.7)$$

$$\|\mathbf{A}\|_2 = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|, \quad (3.8)$$

$$\|\mathbf{A}\|_3 = \sqrt{\max_{1 \leq i \leq n} \lambda_i(\mathbf{A}^* \mathbf{A})}, \quad (3.9)$$

При этом в последнем случае для симметрической матрицы \mathbf{A} справедливо

$$\|\mathbf{A}\| = \max_{1 \leq i \leq n} |\lambda_i(\mathbf{A})|,$$

что можно показать самостоятельно.



В качестве примера получим третью норму матрицы \mathbf{A} :

$$\begin{aligned}\|\mathbf{A}\|_3 &= \sup_{\mathbf{x} \in R^n} \left\{ \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} \right\} = \sup_{\mathbf{x} \in R^n} \left\{ \sqrt{\frac{(\mathbf{Ax}, \mathbf{Ax})}{(\mathbf{x}, \mathbf{x})}} \right\} = \\ &= \sup_{\mathbf{x} \in R^n} \left\{ \sqrt{\frac{(\mathbf{A}^* \mathbf{Ax}, \mathbf{x})}{(\mathbf{x}, \mathbf{x})}} \right\} = \sup_{\mathbf{x} \in R^n} \left\{ \sqrt{\frac{\left(\sum_{i=1}^n \lambda_i \xi_i \omega_i, \sum_{i=1}^n \xi_i \omega_i \right)}{\left(\sum_{i=1}^n \xi_i \omega_i, \sum_{i=1}^n \xi_i \omega_i \right)}} \right\} \leq \\ &\leq \sup_{\mathbf{x} \in R^n} \left\{ \sqrt{\frac{\sum_{i=1}^n \lambda_i \xi_i^2}{\sum_{i=1}^n \xi_i^2}} \right\} \leq \sqrt{\max_{1 \leq i \leq n} \lambda_i (\mathbf{A}^* \mathbf{A})}, \quad (3.10)\end{aligned}$$

причем $\sup_{\mathbf{x} \in R^n} \left\{ \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} \right\}$ достигается при $\mathbf{x} = \omega_i$. При выводе (3.10) учтено, что

$$\mathbf{A}\omega_i = \lambda_i \omega_i, \quad i = 1, \dots, n,$$

а также тот факт, что вещественная матрица $\mathbf{A}^* \mathbf{A}$ является симметрической, следовательно, имеет n вещественных чисел λ_i ($i = 1, \dots, n$) и базис из n собственных векторов $\omega_i \in R^n$ ($i = 1, \dots, n$).

3.2. Обусловленность СЛАУ

Теорема 3.1. Пусть матрица \mathbf{A} и правая часть СЛАУ

$$\mathbf{Ax} = \mathbf{f}; \quad \mathbf{x}, \mathbf{f} \in R^n$$

получают приращения $\Delta \mathbf{A}$, $\Delta \mathbf{f} \in R^n$, соответственно, т.е. решается СЛАУ вида

$$(\Delta \mathbf{A} + \mathbf{A})(\mathbf{x} + \Delta \mathbf{x}) = \mathbf{f} + \Delta \mathbf{f}. \quad (3.11)$$

Пусть также $\exists \mathbf{A}^{-1}$, $\|\mathbf{A}\| \neq 0$, $\|\mathbf{A}^{-1}\| \neq 0$, $1 - \mu \frac{\|\Delta \mathbf{A}\|}{\|\mathbf{A}\|} > 0$, $\mu = \|\mathbf{A}^{-1}\| \cdot \|\mathbf{A}\|$, μ — параметр обусловленности СЛАУ.

В таком случае справедливо неравенство

$$\frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\mu}{1 - \mu \frac{\|\Delta \mathbf{A}\|}{\|\mathbf{A}\|}} \left(\frac{\|\Delta \mathbf{f}\|}{\|\mathbf{f}\|} + \frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|} \right). \quad (3.12)$$

Доказательство. Из (3.11) получаем

$$\Delta \mathbf{x} = \mathbf{A}^{-1} (\Delta \mathbf{f} - \Delta \mathbf{A} \cdot \mathbf{x} - \Delta \mathbf{A} \cdot \Delta \mathbf{x}),$$

откуда, переходя к нормам, имеем

$$\|\Delta \mathbf{x}\| \leq \|\mathbf{A}^{-1}\| \cdot \|\Delta \mathbf{f}\| + \|\mathbf{A}^{-1}\| \cdot \|\Delta \mathbf{A}\| \cdot \|\mathbf{x}\| + \|\mathbf{A}^{-1}\| \cdot \|\Delta \mathbf{A}\| \cdot \|\Delta \mathbf{x}\|,$$

или

$$\|\Delta \mathbf{x}\| \leq \|\mathbf{A}^{-1}\| \cdot \|\mathbf{A}\| \cdot \frac{\|\Delta \mathbf{f}\|}{\|\mathbf{f}\|} \cdot \frac{\|\mathbf{f}\|}{\|\mathbf{A}\|} + \|\mathbf{A}^{-1}\| \cdot \|\mathbf{A}\| \cdot \frac{\|\Delta \mathbf{A}\|}{\|\mathbf{A}\|} \cdot \|\mathbf{x}\| + \|\mathbf{A}^{-1}\| \cdot \|\mathbf{A}\| \cdot \frac{\|\Delta \mathbf{A}\|}{\|\mathbf{A}\|} \cdot \|\Delta \mathbf{x}\|.$$

После несложных алгебраических преобразований получим неравенство (3.12).

Теорема доказана.

Отметим, что если положить

$$\Delta \mathbf{A} \cdot \Delta \mathbf{x} \approx 0,$$

т. е. пренебречь малыми величинами, то (3.12) упростится:

$$\frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|} \leq \mu \left(\frac{\|\Delta \mathbf{f}\|}{\|\mathbf{f}\|} + \frac{\|\Delta \mathbf{A}\|}{\|\mathbf{A}\|} \right),$$

или

$$\delta \mathbf{x} \leq \mu (\delta \mathbf{f} + \delta \mathbf{A}),$$

где

$$\delta \mathbf{x} = \frac{\|\Delta \mathbf{x}\|}{\|\mathbf{x}\|}, \quad \delta \mathbf{f} = \frac{\|\Delta \mathbf{f}\|}{\|\mathbf{f}\|}, \quad \delta \mathbf{A} = \frac{\|\Delta \mathbf{A}\|}{\|\mathbf{A}\|}. \quad (3.13)$$

Во многих практически интересных случаях можно положить $\delta \mathbf{A} \approx 0$; тогда

$$\delta \mathbf{x} \leq \mu \cdot \delta \mathbf{f},$$

где $\mu = \mu(\mathbf{A}) = \|\mathbf{A}^{-1}\| \cdot \|\mathbf{A}\|$, причем $\mu \geq 1$, поскольку

$$1 = \|\mathbf{E}\| = \|\mathbf{A}^{-1} \mathbf{A}\| \leq \|\mathbf{A}^{-1}\| \cdot \|\mathbf{A}\| = \mu. \quad (3.14)$$

Можно показать, что для симметрической матрицы \mathbf{A} справедливо:

$$\mu = \frac{\max_i |\lambda_i|}{\min_i |\lambda_i|}. \quad (3.15)$$

Параметр обусловленности СЛАУ является важнейшим показателем «чувствительности» ее решения при малых изменениях

где:

$$Q_2 = \begin{Bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & (-q_{32}) & 1 & 0 & 0 & 0 \\ 0 & (-q_{42}) & 0 & 1 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & (-q_{n2}) & 0 & \dots & \dots & 1 \end{Bmatrix}, \dots,$$

$$A_{n-1} = (Q_{n-1} \dots Q_1) A, \quad f^{n-1} = (Q_{n-1} \dots Q_1) f,$$

откуда: $A = (Q_1^{-1} \dots Q_{n-1}^{-1}) A_{n-1} = L \cdot U$; $L, U, Q_1, \dots, Q_{n-1} \in M(n \times n)$, где U — верхнетреугольная матрица вида (3.20):

$$U = \begin{Bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ 0 & u_{22}^1 & \dots & u_{2n}^1 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & u_{nn}^{n-1} \end{Bmatrix}, \quad (3.22)$$

а L — нижнетреугольная матрица

$$L = \begin{Bmatrix} 1 & 0 & 0 & \dots & 0 \\ q_{21} & 1 & 0 & \dots & 0 \\ q_{31} & q_{32} & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ q_{n1} & q_{n2} & \dots & \dots & 1 \end{Bmatrix}, \quad (3.23)$$

Таким образом, матрица A представляется в виде произведения верхнетреугольной и нижнетреугольной матриц L и U .

Однако существует стандартный метод разложения матрицы A — метод LU -разложения.

Представим систему (3.1) в виде

$$(LU)x = f,$$

или

$$\begin{cases} Ly = f, \\ Ux = y, \end{cases}$$

$L = \{l_{ij}\}$, $U = \{u_{ij}\}$; $i, j = 1, \dots, n$, т.е. в виде двух СЛАУ с нижнетреугольной и верхнетреугольной матрицами:

$$\begin{cases} y_1 = f_1, \\ l_{21}y_1 + y_2 = f_2, \\ \dots \\ l_{n1}y_1 + \dots + l_{n,n-1}y_{n-1} = f_n, \end{cases} \quad (3.24)$$

$$\begin{cases} u_{11}x_1 + u_{12}x_2 + \dots + u_{1n}x_n = y_1, \\ u_{22}x_2 + \dots + u_{2n}x_n = y_2, \\ \dots \\ u_{nn}x_n = y_n, \end{cases}$$

решения которых находятся по известным рекуррентным формулам (3.21). Коэффициенты l_{ij}, u_{ij} находятся из системы линейных уравнений порядка $n \times n$:

$$\begin{Bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{Bmatrix} \stackrel{\text{ЛАНБ}}{=} \begin{Bmatrix} 1 & 0 & 0 & \dots & 0 \\ l_{21} & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ l_{n1} & l_{n2} & \dots & \dots & 1 \end{Bmatrix} \begin{Bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ 0 & u_{22} & \dots & u_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & u_{nn} \end{Bmatrix}. \quad (3.25)$$

При этом: $a_{11} = u_{11}, a_{12} = u_{12}, \dots$;

$$u_{ij} = a_{ij} - \sum_{k=1}^{i-1} l_{ik} u_{kj}, \quad i \leq j;$$

$$l_{ij} = u_{jj}^{-1} - \left(a_{ij} - \sum_{k=1}^{j-1} l_{ik} u_{kj} \right), \quad i > j. \quad (3.26)$$

Как известно из линейной алгебры, такое **LU**-разложение возможно, если главные миноры матрицы **A** отличны от нуля.

Количество арифметических действий в методе **LU**-разложения оценивается как $\approx 2n^3$.

Важным следствием этого метода являются случай симметрической матрицы **A** = **LL***, которую можно представить в виде произведения нижнетреугольной матрицы **L** на транспонированную, т. е. верхнетреугольную матрицу **L***:

$$\mathbf{L} = \begin{Bmatrix} l_{11} & 0 & 0 & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ l_{n1} & l_{n2} & \dots & \dots & l_{nn} \end{Bmatrix}, \quad \mathbf{L}^* = \begin{Bmatrix} l_{11} & l_{21} & \dots & l_{n1} \\ 0 & l_{22} & \dots & l_{n2} \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & l_{nn} \end{Bmatrix}.$$

В этом случае СЛАУ представляется в виде

$$\begin{cases} \mathbf{L}\mathbf{y} = \mathbf{f}, \\ \mathbf{L}^*\mathbf{x} = \mathbf{y}; \end{cases} \quad (3.27)$$

для ее решения требуется приблизительно $\approx 2n^2$ операций.

Алгоритм в этом случае аналогичен алгоритму, используемому в методе **LU**-разложении. Коэффициенты l_{ij} находятся из СЛАУ:

$$\mathbf{A} = \mathbf{LL}^*,$$

или, в развернутом виде:

$$\begin{Bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{Bmatrix} = \begin{Bmatrix} l_{11} & 0 & \dots & 0 \\ l_{21} & l_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ l_{n1} & l_{n2} & \dots & l_{nn} \end{Bmatrix} \begin{Bmatrix} l_{11} & l_{21} & \dots & l_{n1} \\ 0 & l_{22} & \dots & l_{n2} \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & l_{nn} \end{Bmatrix}.$$

В результате получается СЛАУ, которая решается с помощью рекуррентных формул:

$$\begin{aligned}
 l_{11} &= \sqrt{a_{11}}, \quad l_{i1} = \frac{a_{i1}}{l_{11}}, \quad i = 2, \dots, n; \\
 l_{22} &= \sqrt{a_{22} - l_{21}^2}, \quad l_{i2} = \frac{a_{i2} - l_{i1}l_{12}}{l_{22}}, \quad i = 3, \dots, n; \\
 &\dots\dots\dots \\
 l_{kk} &= \sqrt{a_{kk} - l_{k1}^2 - l_{k2}^2 - \dots - l_{k,k-1}^2}, \\
 l_{ik} &= \frac{a_{ik} - l_{i1}l_{k1} - l_{i2}l_{k2} - \dots - l_{i,k-1}l_{k,k-1}}{l_{kk}}, \\
 &i = k + 1, \dots, n.
 \end{aligned} \tag{3.28}$$

Этот метод называется *методом квадратного корня*, или *методом Холецкого*.

3.4. Метод простых итераций (МПИ)

Метод простых итераций получается путем приведения системы линейных алгебраических уравнений

$$\mathbf{Ax} = \mathbf{f} \tag{3.29}$$

к виду

$$\mathbf{x} = \tilde{\mathbf{A}}\mathbf{x} + \mathbf{F}, \tag{3.30}$$

где $\mathbf{x}, \mathbf{f}, \mathbf{F} \in B$ (банаховы, т.е. полные линейные нормированные пространства), $\mathbf{A}, \tilde{\mathbf{A}} \in M(n \times n)$, и введением итерационных индексов (они могут быть как нижними, так и верхними):

$$\mathbf{x}^{k+1} = \tilde{\mathbf{A}}\mathbf{x}^k + \mathbf{F}, \quad \mathbf{x}^0 = \mathbf{a}. \tag{3.31}$$

Решение системы (3.30) вычисляется как предел последовательности $\{\mathbf{x}^k\} \xrightarrow{k \rightarrow \infty} \mathbf{X}$, где \mathbf{X} — точное решение рассматриваемой СЛАУ. При этом $\mathbf{x}^k \in B$ и $\mathbf{X} \in B$, т.е. итерационные процессы рассматриваются в банаховых, либо в гильбертовых пространствах, поскольку в этих процессах важным свойством пространства является его полнота (т.е. всякая фундаментальная последовательность в таком пространстве имеет предел, принадлежащий этому пространству).

Примерами такого итерационного процесса являются методы Якоби и Зейделя. В этом случае матрица \mathbf{A} представляется

при этом матрица $\tilde{\mathbf{A}}$ и правая часть \mathbf{F} будут иметь вид:

$$\tilde{\mathbf{A}} = -(\mathbf{L} + \mathbf{D})^{-1} \mathbf{U} \mathbf{x}, \quad \mathbf{F} = (\mathbf{L} + \mathbf{D})^{-1} \mathbf{f}.$$

Скалярная форма метода Зейделя имеет следующий вид:

$$\begin{aligned} x_1^{k+1} &= -a_{11}^{-1} (a_{12}x_2^k + a_{13}x_3^k + \dots + a_{1n}x_n^k - f_1), \\ x_2^{k+1} &= -a_{22}^{-1} (a_{21}x_1^{k+1} + a_{23}x_3^k + \dots + a_{2n}x_n^k - f_2), \\ x_n^{k+1} &= -a_{nn}^{-1} (a_{n1}x_1^{k+1} + a_{n2}x_2^{k+1} + \dots + a_{n,n-1}x_{n-1}^{k+1} - f_n), \\ x_1^0 &= a_1, \dots, \quad x_n^0 = a_n; \quad k = 0, 1, \dots \end{aligned} \quad (3.38)$$

Обобщением метода Зейделя является метод релаксации (τ – итерационный параметр, позволяющий ускорять итерационный процесс):

$$(\tau \cdot \mathbf{L} + \mathbf{D})^{-1} \frac{\mathbf{x}^{k+1} - \mathbf{x}^k}{\tau} + \mathbf{A} \mathbf{x}^k = \mathbf{f}.$$

Доказано, что итерационный метод сходится к решению СЛАУ, если матрица симметрическая и положительно определенная, а также $0 < \tau < 2$, причем при $0 < \tau < 1$ этот итерационный процесс, называемый *методом нижней релаксации*, не используется в практических вычислениях, в отличие от случая $1 < \tau < 2$ – *метода верхней релаксации*. Для этого метода получим

$$\mathbf{x}^{k+1} = -(\tau \cdot \mathbf{L} + \mathbf{D})^{-1} [(1 - \tau) \mathbf{D} + \tau \mathbf{U}] \mathbf{x}^k + \tau (\tau \mathbf{L} + \mathbf{D})^{-1} \mathbf{f}.$$

3.5. Сходимость итерационного процесса

После построения численного итерационного метода необходимо доказать, что он сходится к точному решению СЛАУ при $k \rightarrow \infty$, и найти условия сходимости.

Достаточное условие сходимости итерационного процесса (в методе простых итераций) дает следующая теорема.

Теорема 3.4. Последовательность $\{\mathbf{x}^k\}$, $k = 0, 1, \dots$, порожденная итерационным процессом

$$\mathbf{x}^{k+1} = \tilde{\mathbf{A}} \mathbf{x}^k + \mathbf{F},$$

сходится к решению \mathbf{U} системы линейных алгебраических уравнений

$$\mathbf{x} = \tilde{\mathbf{A}} \mathbf{x} + \mathbf{F}$$

со скоростью геометрической прогрессии, если выполняется условие

$$\|\tilde{\mathbf{A}}\| \leq q < 1.$$

Доказательство. Вычитая из итерационного уравнения (3.31) равенство

$$\mathbf{X} = \mathbf{A}\mathbf{X} + \mathbf{F},$$

получим:

$$\mathbf{u}^k - \mathbf{X} = \tilde{\mathbf{A}} (\mathbf{u}^{k-1} - \mathbf{X}),$$

или:

$$\boldsymbol{\varepsilon}^k = \tilde{\mathbf{A}} \boldsymbol{\varepsilon}^{k-1},$$

где $\boldsymbol{\varepsilon}^k = \mathbf{u}^k - \mathbf{X}$ — итерационная разность k -го приближения и точного решения системы.

Далее, построив цепочку неравенств, получим следующую оценку:

$$\|\boldsymbol{\varepsilon}_k\| \leq \|\tilde{\mathbf{A}}\| \cdot \|\boldsymbol{\varepsilon}_{k-1}\| < q \|\boldsymbol{\varepsilon}_{k-1}\| < \dots < q^k \|\boldsymbol{\varepsilon}_0\| = q^k \|\mathbf{u}^0 - \mathbf{X}\|,$$

из которой видно, что если $0 < q < 1$, то

$$\{\mathbf{x}^k\} \rightarrow \mathbf{X}, \quad k = 0, 1, \dots$$

Из последнего неравенства получим:

$$\|\boldsymbol{\varepsilon}_k\| < q^k \|\boldsymbol{\varepsilon}_0\|, \quad (3.39)$$

откуда следует оценка количества итераций, требуемого для обеспечения точности решения $\boldsymbol{\varepsilon}$:

$$k \approx \frac{\ln(\boldsymbol{\varepsilon}/\boldsymbol{\varepsilon}_0)}{\ln q}. \quad (3.40)$$

В теории итерационных методов доказывается следующая важная теорема о критерии сходимости метода простой итерации.

Теорема 3.5. Для сходимости итерационного процесса

$$\mathbf{x}^{k+1} = \tilde{\mathbf{A}}\mathbf{x}^k + \mathbf{F}$$

необходимо и достаточно, чтобы все собственные числа матрицы $\tilde{\mathbf{A}}$ были по абсолютной величине строго меньше единицы.

Если сопоставить количества арифметических действий, необходимые для получения численного решения СЛАУ методом Гаусса: $\approx (2/3)n^3$, где n — количество уравнений в системе, и методом простых итераций: $\approx 2n^2 \cdot I$, где I — количество итераций, то окажется, что при

$$I < \frac{n}{3}$$

метод итераций становится более эффективным. В реальных задачах это условие в основном выполняется, и большая часть задач решается итерационными методами.

Следует также заметить, что возможна ситуация, при которой критерий сходимости выполняется, а численное решение возрастает по модулю. Этот эффект не вызван неустойчивостью, поскольку решение сначала возрастает, а затем стремится к точному. Связано это явление с тем, что при выполнении критерия сходимости достаточное условие сходимости может не выполняться, т. е. $\|\tilde{\mathbf{A}}\| > 1$.

Проведем оценку влияния ошибки округления на результаты численного решения СЛАУ.

Представим реальный вычислительный процесс в виде

$$\mathbf{x}_m^k = \tilde{\mathbf{A}} \mathbf{x}_m^{k-1} + \mathbf{F} + \delta^k;$$

где δ^k — суммарная погрешность округления на k -й итерации, \mathbf{x}_m^k — реальное «машинное» значение \mathbf{x}^k .

Вычитая из этого уравнения итерационное соотношение

$$\mathbf{x}^k = \tilde{\mathbf{A}} \mathbf{x}^{k-1} + \mathbf{F},$$

получим:

$$\begin{aligned} \|\mathbf{x}_m^k - \mathbf{x}^k\| &< q \|\mathbf{x}_m^{k-1} - \mathbf{x}^{k-1}\| + \|\delta^k\| \leq \\ &\leq q^2 \|\mathbf{x}_m^{k-2} - \mathbf{x}^{k-2}\| + q \|\delta^{k-1}\| < \dots < q^k \|\mathbf{x}_m^0 - \mathbf{x}^0\| + \\ &+ \left(\max_i \|\delta^i\| \right) (1 + q + \dots + q^{k-1}); \quad i = 1, \dots, k. \end{aligned}$$

Понятно, что так как начальное приближение задается с максимальной точностью, то $\|\mathbf{x}_m^0 - \mathbf{x}^0\| = 0$; положим также

$$\delta = \max_i \|\delta^i\|$$

и оценим сумму полученной геометрической прогрессии:

$$\|\mathbf{x}_m^k - \mathbf{x}^k\| < \delta \frac{q^k - 1}{q - 1} < \frac{\delta}{1 - q}.$$

Мы получили важный результат: при $0 < q < 1$ погрешность не зависит от количества итераций.

Итерационный процесс (3.31) также можно представить в виде

$$\mathbf{x}^{k+1} = (\mathbf{E} - \tau \mathbf{A}) \mathbf{x}^k + \tau \mathbf{f}, \quad (3.41)$$

где

$$\tilde{\mathbf{A}} = \mathbf{E} - \tau \mathbf{A}, \quad \mathbf{F} = \tau \mathbf{f},$$

τ — итерационный параметр, используемый для его ускорения.

Двухслойные итерационные процессы (методы), в которых нужно помнить результаты только одной итерации, представляются в каноническом виде:

$$\mathbf{D}_{k+1} \frac{\mathbf{x}^{k+1} - \mathbf{x}^k}{\tau_{k+1}} + \mathbf{A} \mathbf{x}^k = \mathbf{f}, \quad \mathbf{x}^0 = \mathbf{a}, \quad k = 0, 1, \dots \quad (3.42)$$

Здесь \mathbf{D}_{k+1} — обратимая матрица, задающая итерационный метод, τ_{k+1} — итерационный параметр, вообще говоря, зависящий от номера итерации.

Оператор \mathbf{D}_{k+1} иногда называют *предобусловливателем* СЛАУ; он должен относительно просто вычисляться; тогда (3.42) можно представить в виде:

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \tau_{k+1} \mathbf{D}_{k+1}^{-1} (\mathbf{f} - \mathbf{A} \mathbf{x}^k)$$

— это семейство итерационных методов, зависящих от выбора итерационного параметра τ_k и матрицы \mathbf{D}_{k+1}^{-1} , что можно использовать для ускорения итерационного процесса.

При $\mathbf{D}_{k+1} = \mathbf{E}$ метод называется *явным*, в противном случае — *неявным*; *стационарным*, если \mathbf{D}_{k+1} и τ_{k+1} не зависят от номера итерации k , и *нестационарным* в противном случае.

Теорема 3.6. (*достаточное условие сходимости итерационного процесса Якоби*) — Итерационный процесс Якоби сходится к решению системы линейных уравнений $\mathbf{A} \mathbf{u} = \mathbf{f}$ при выполнении условия диагонального преобладания:

$$|a_{kk}| > \sum_{\substack{j=1 \\ j \neq k}}^n |a_{kj}|, \quad j = 1, \dots, n. \quad (3.43)$$

Доказательство. Если условие диагонального преобладания выполняется, то в любой строке матрицы

$$\tilde{\mathbf{A}} = \begin{pmatrix} 0 & -(a_{12}/a_{11}) & -(a_{13}/a_{11}) & \dots & -(a_{1n}/a_{11}) \\ -(a_{21}/a_{22}) & 0 & -(a_{23}/a_{22}) & \dots & -(a_{2n}/a_{22}) \\ \dots & \dots & \dots & \dots & \dots \\ -(a_{n1}/a_{nn}) & -(a_{n2}/a_{nn}) & \dots & \dots & -(a_{nn-1}/a_{nn}) \end{pmatrix}$$

сумма модулей элементов матрицы меньше единицы, так как из

$$|a_{kk}| > \sum_{\substack{j=1 \\ j \neq k}}^n |a_{kj}|$$

следует

$$\sum_{\substack{j=1 \\ j \neq k}}^n \left| \frac{a_{kj}}{a_{kk}} \right| < 1.$$

т.е. одна из норм матрицы $\tilde{\mathbf{A}}$ меньше 1. При этом достаточное условие сходимости выполняется. Критерий сходимости метода Якоби дается следующий теоремой.

Теорема 3.7. Для того чтобы итерационный процесс Якоби сходил к решению соответствующей СЛАУ, необходимо и достаточно, чтобы все корни уравнения

$$\begin{vmatrix} \lambda a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & \lambda a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & \dots & \dots & \lambda a_{nn} \end{vmatrix} = 0$$

были по модулю меньше единицы.

Доказательство. Покажем, что все корни этого уравнения являются собственными числами матрицы

$$\tilde{\mathbf{A}} = -\mathbf{D}(\mathbf{L} + \mathbf{X}).$$

Если $\boldsymbol{\omega}$ — собственный вектор $\tilde{\mathbf{A}}$, соответствующий собственному значению λ , то: $\tilde{\mathbf{A}}\boldsymbol{\omega} = \lambda\boldsymbol{\omega}$, или $-\mathbf{D}^{-1}(\mathbf{L} + \mathbf{X})\boldsymbol{\omega} = \lambda\boldsymbol{\omega}$, откуда следует

$$(\mathbf{L} + \mathbf{X} + \lambda\mathbf{D})\boldsymbol{\omega} = 0.$$

Эта система линейных алгебраических уравнений имеет нетривиальные решения, если

$$\det(\mathbf{L} + \mathbf{X} + \lambda\mathbf{D}) = 0,$$

т.е. собственные значения матрицы $\tilde{\mathbf{A}} = -\mathbf{D}^{-1}(\mathbf{L} + \mathbf{X})$ — это корни полученного уравнения $\det[(\mathbf{L} + \mathbf{X} + \lambda\mathbf{D})] = 0$, которые должны быть по модулю меньше единицы, в соответствии с критерием сходимости итерационного процесса (МПИ).

Теорема 3.8. Для того чтобы итерационный процесс Зейделя сходил к решению соответствующей СЛАУ, необходимо и достаточно, чтобы корни уравнения

$$\begin{vmatrix} \lambda a_{11} & a_{12} & \dots & a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ \lambda a_{n1} & \dots & \dots & \lambda a_{nn} \end{vmatrix} = 0$$

были по модулю меньше единицы.

Доказательство. Покажем, что корни этого уравнения являются собственными значениями матрицы $\tilde{\mathbf{A}}$. Из $\tilde{\mathbf{A}}\omega = \lambda\omega$ следует

$$-(\mathbf{D} + \mathbf{L})^{-1} \mathbf{X}\omega = \lambda\omega,$$

поскольку в итерационном методе Зейделя $\tilde{\mathbf{A}} = (-\mathbf{D} + \mathbf{L}) \mathbf{X} = \lambda\omega$; тогда СЛАУ

$$(\lambda\mathbf{L} + \mathbf{X} + \lambda\mathbf{D})\omega = 0$$

имеет нетривиальное решение, если $\det(\lambda\mathbf{L} + \mathbf{X} + \lambda\mathbf{D}) = 0$, т.е. корни этого уравнения должны быть по модулю меньше единицы, в соответствии с критерием сходимости Зейделя.

Теорема 3.9 (сходимость метода Зейделя). *Если в СЛАУ*

$$\mathbf{A}\mathbf{x} = \mathbf{f}; \quad \mathbf{f}, \mathbf{x} \in R^n$$

матрица $\mathbf{A} \in M(n \times n)$ — нормальная, то итерационный метод Зейделя сходится (система $\mathbf{A}\mathbf{x} = \mathbf{f}$ называется нормальной, если матрица \mathbf{A} — симметрическая и положительно определенная).

Доказательство этой теоремы сводится к проверке того, что из положительной определенности матрицы $\mathbf{A} = \mathbf{L} + \mathbf{D} + \mathbf{L}^T$ следует выполнение критерия сходимости МПИ (все собственные значения матрицы $\tilde{\mathbf{A}} = -(\mathbf{L} + \mathbf{D})^{-1} \mathbf{L}^T$ меньше единицы по модулю).

Отметим, что любая СЛАУ вида

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

может быть симметризована умножением обеих частей на матрицу \mathbf{A}^* (симметризация по Гауссу):

$$\mathbf{A}^* \mathbf{A} \mathbf{x} = \mathbf{A}^* \mathbf{f}. \quad (3.44)$$

Эта система является нормальной (что доказывается в курсе линейной алгебры), поэтому для ее решения можно использовать метод Зейделя.



3.6. Итерационные вариационные методы последовательных приближений (итераций) численного решения СЛАУ

Рассмотрим квадратичную функцию (функционал) вида:

$$F(\mathbf{x}) = (\mathbf{A}\mathbf{x}, \mathbf{x}) - 2(\mathbf{f}, \mathbf{x}) + c, \quad (3.45)$$

для которого $\mathbf{x}, \mathbf{f} \in H$ (гильбертово пространство), c — скаляр, \mathbf{A} — линейный оператор, действующий в гильбертовом пространстве.

Поскольку

$$(\mathbf{A}\mathbf{x}, \mathbf{x}) = (\mathbf{A}^*\mathbf{x}, \mathbf{x}) = (\mathbf{x}, \mathbf{A}^*\mathbf{x}),$$

то $F(\mathbf{x})$ совпадает с функцией $\Phi(\mathbf{x}) = (\mathbf{A}^*\mathbf{x}, \mathbf{x}) - 2(\mathbf{f}, \mathbf{x}) + c$, т.е. без ограничения общности мы будем полагать, что оператор \mathbf{A} симметричен: $\mathbf{A} = \mathbf{A}^*$; поскольку $F(\mathbf{x})$ можно представить в виде:

$$F(\mathbf{x}) = \left(\frac{\mathbf{A} + \mathbf{A}^*}{2} \mathbf{x}, \mathbf{x} \right) - 2(\mathbf{f}, \mathbf{x}) + c.$$

Покажем, что задачи решения СЛАУ и минимизации рассматриваемой квадратичной функции эквивалентны.

Теорема 3.10. Пусть оператор \mathbf{A} симметричен и положительно определен ($\mathbf{A} = \mathbf{A}^* > 0$). Тогда \exists единственный элемент $\mathbf{Z} \in H$, доставляющий минимум квадратичной функции

$$F(\mathbf{x}) = (\mathbf{A}\mathbf{x}, \mathbf{x}) - 2(\mathbf{f}, \mathbf{x}) + c$$

и являющийся решением системы линейных алгебраических уравнений

$$\mathbf{A}\mathbf{x} = \mathbf{f}. \quad (3.46)$$

Доказательство. Положим, что \mathbf{Z} — решение системы (3.46).

В этом случае будет справедливо следующее неравенство (\mathbf{v} — приращение к \mathbf{Z}):

$$\begin{aligned} F(\mathbf{Z} + \mathbf{v}) &= (\mathbf{A}(\mathbf{Z} + \mathbf{v}), \mathbf{Z} + \mathbf{v}) - 2(\mathbf{f}, \mathbf{Z} + \mathbf{v}) + c = \\ &= (\mathbf{A}\mathbf{Z}, \mathbf{Z}) + (\mathbf{A}\mathbf{Z}, \mathbf{v}) + (\mathbf{A}\mathbf{v}, \mathbf{Z}) + (\mathbf{A}\mathbf{v}, \mathbf{v}) - 2(\mathbf{f}, \mathbf{Z} + \mathbf{v}) + c = \\ &= F(\mathbf{Z}) + 2(\mathbf{A}\mathbf{Z}, \mathbf{v}) - 2(\mathbf{f}, \mathbf{v}) + (\mathbf{A}\mathbf{v}, \mathbf{v}) = \\ &= F(\mathbf{Z}) + 2(\mathbf{A}\mathbf{Z} - \mathbf{f}, \mathbf{v}) + (\mathbf{A}\mathbf{v}, \mathbf{v}) = F(\mathbf{Z}) + (\mathbf{A}\mathbf{v}, \mathbf{v}) > F(\mathbf{Z}), \end{aligned}$$

поскольку

$$(\mathbf{A}\mathbf{Z} - \mathbf{f}, \mathbf{v}) = 0.$$

Теперь докажем обратное: пусть

$$\mathbf{Z} = \arg \min_{\mathbf{x} \in H} F(\mathbf{x}),$$

т.е. \mathbf{Z} доставляет минимум квадратичной функции $F(\mathbf{x})$; в таком случае \mathbf{Z} является решением СЛАУ вида

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

Если это условие выполняется, т. е.

$$\mathbf{Z} = \arg \min_{\mathbf{x} \in H} \mathbf{F}(\mathbf{x}),$$

то:

$$\text{grad } F(\mathbf{x}) = 0,$$

или:

$$\text{grad} [(\mathbf{A}\mathbf{x}, \mathbf{x}) - 2(\mathbf{f}, \mathbf{x})] + c = 0,$$

откуда:

$$\text{grad } F(\mathbf{x}) = 2\mathbf{A}\mathbf{x} - 2\mathbf{f} = 2(\mathbf{A}\mathbf{x} - \mathbf{f}) = 0,$$

т. е. в этом случае \mathbf{Z} также является решением системы линейных алгебраических уравнений

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

Теорема доказана.

Это значит, что задачу решения СЛАУ $\mathbf{A}\mathbf{x} = \mathbf{f}$ можно заменить на задачу о нахождении минимума квадратичной функции

$$F(\mathbf{x}) = (\mathbf{A}\mathbf{x}, \mathbf{x}) - 2(\mathbf{f}, \mathbf{x}) + c,$$

и наоборот.

Представим методы градиентного и наискорейшего спуска для решения системы линейных алгебраических уравнений.

$$\mathbf{A}\mathbf{x} = \mathbf{f}, \quad \mathbf{A} = \mathbf{A}^* > 0,$$

для чего рассмотрим функцию

$$F(\mathbf{x}) = (\mathbf{A}\mathbf{x}, \mathbf{x}) - 2(\mathbf{f}, \mathbf{x}) + c$$

и итерационный процесс для нахождения ее минимума:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{\tau_k}{2} \cdot \text{grad } F(\mathbf{x}). \quad (3.47)$$

В нашем случае

$$\text{grad } F(\mathbf{x}) = 2(\mathbf{A}\mathbf{x} - \mathbf{f}),$$

поэтому

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \tau_k (\mathbf{A}\mathbf{x} - \mathbf{f}). \quad (3.48)$$

Найдем τ_k из условия минимума $F(\mathbf{x})$ по τ_k :

$$\begin{aligned} F'_{\tau_k}(\tau_k, \mathbf{x}_{k+1}) &= (\mathbf{A}\mathbf{x}'_{k+1}, \mathbf{x}_{k+1}) + (\mathbf{A}\mathbf{x}_{k+1}, \mathbf{x}'_{k+1}) - 2(\mathbf{f}, \mathbf{x}'_{k+1}) = \\ &= 2(\mathbf{A}\mathbf{x}_{k+1}, \mathbf{x}'_{k+1}) - 2(\mathbf{f}, \mathbf{x}_{k+1}) = 2(\mathbf{A}\mathbf{x}_{k+1} - \mathbf{f}, \mathbf{x}'_{k+1}) = 0; \end{aligned}$$

все производные в этом выражении берутся по τ_k .

Далее:

$$\begin{aligned} (\mathbf{A}\mathbf{x}_{k+1} - \mathbf{f}, \mathbf{x}'_{k+1}) &= (\mathbf{A}\mathbf{x}_{k+1} - \mathbf{f}, \mathbf{A}\mathbf{x}_k - \mathbf{f}) = \\ &= (\mathbf{A}(\mathbf{x}_k - \tau_k(\mathbf{A}\mathbf{x}_k - \mathbf{f})) - \mathbf{f}, \mathbf{A}\mathbf{x}_k - \mathbf{f}) = \\ &= (\mathbf{A}\mathbf{x}_k - \mathbf{f}, \mathbf{A}\mathbf{x}_k - \mathbf{f}) - \tau_k(\mathbf{A}(\mathbf{A}\mathbf{x}_k - \mathbf{f}), \mathbf{A}\mathbf{x}_k - \mathbf{f}) = 0, \end{aligned}$$

откуда получим выражение τ_k :

$$\tau_k = \frac{(\mathbf{A}\mathbf{x}_k - \mathbf{f}, \mathbf{A}\mathbf{x}_k - \mathbf{f})}{(\mathbf{A}(\mathbf{A}\mathbf{x}_k - \mathbf{f}), \mathbf{A}\mathbf{x}_k - \mathbf{f})} = \frac{(\xi_k, \xi_k)}{(\mathbf{A}\xi_k, \xi_k)}, \quad (3.49)$$

где $\xi_k = \mathbf{A}\mathbf{x}_k - \mathbf{f}$ — невязка.

Этот метод называется *методом наискорейшего спуска*.

Метод минимальных невязок состоит в минимизации функции (ξ_{k+1}, ξ_{k+1}) , где

$$\xi_{k+1} = \mathbf{A}\mathbf{x}_{k+1} - \mathbf{f}; \quad \mathbf{A} = \mathbf{A}^*; \quad \mathbf{x}, \mathbf{f}, \xi_k \in H,$$

с целью вычисления итерационного параметра в итерационном процессе

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \tau_k \xi_k. \quad (3.50)$$

Для этого вычтем из равенства

$$\xi_{k+1} = \mathbf{A}\mathbf{x}_{k+1} - \mathbf{f}$$

равенство

$$\xi_k = \mathbf{A}\mathbf{x}_k - \mathbf{f};$$

получим уравнение

$$\xi_{k+1} - \xi_k = \mathbf{A}(\mathbf{x}_{k+1} - \mathbf{x}_k) = -\tau_k \mathbf{A}\xi_k,$$

которое после возведения в квадрат, в смысле скалярного произведения, будет иметь вид

$$(\xi_{k+1}, \xi_{k+1}) = (\xi_k, \xi_k) - 2\tau_k(\mathbf{A}\xi_k, \xi_k) + \tau_k^2(\mathbf{A}\xi_k, \xi_k),$$

откуда

$$(\xi_{k+1}, \xi_{k+1})'_{\tau_k} = 2(\mathbf{A}\xi_k, \xi_k) + 2\tau_k(\mathbf{A}\xi_k, \xi_k).$$

Из этого выражения получим формулу для определения итерационного параметра

$$\tau_k = \frac{(\mathbf{A}\xi_k, \xi_k)}{(\mathbf{A}\xi_k, \mathbf{A}\xi_k)}. \quad (3.51)$$

В качестве примера построения итерационных процессов Якоби, Зейделя, верхней релаксации рассмотрим СЛАУ вида

$$\begin{cases} 2x + y = 1, \\ x + 2x = -1, \end{cases}$$

или

$$\mathbf{A}\mathbf{u} = \mathbf{f},$$

$$\mathbf{u} = \begin{Bmatrix} x \\ y \end{Bmatrix}, \quad \mathbf{f} = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}, \quad \mathbf{A} = \begin{Bmatrix} 2 & 1 \\ 1 & 2 \end{Bmatrix}.$$

Метод Якоби для этой системы имеет вид

$$\begin{cases} x_{k+1} = -\frac{1}{2}y_k + \frac{1}{2}, \\ y_{k+1} = -\frac{1}{2}x_k - \frac{1}{2}, \end{cases} \quad \{x_0, y_0\} = \{a, b\},$$

или

$$\mathbf{u}_{k+1} = \tilde{\mathbf{A}}\mathbf{u}_k + \mathbf{f},$$

где

$$\tilde{\mathbf{A}} = \begin{Bmatrix} 0 & -1/2 \\ -1/2 & 0 \end{Bmatrix}.$$

Метод Зейделя и верхней релаксации могут быть представлены в следующих видах соответственно:

$$\begin{cases} x_{k+1} = -\frac{1}{2}y_k + \frac{1}{2}, \\ y_{k+1} = -\frac{1}{2}x_{k+1} - \frac{1}{2}; \end{cases}$$

$$\begin{cases} x_{k+1} = (1 - \tau)x_k + \frac{\tau}{2}(1 - y_k), \\ y_{k+1} = (1 - \tau)y_k - \frac{\tau}{2}(1 + x_k). \end{cases}$$

Список литературы

1. *Рябенский В. С.* Введение в вычислительную математику. М.: ФИЗМАТЛИТ, 2008. 288 с.
2. *Бахвалов Н. С., Жидков Н. П., Кобельков Г. М.* Численные методы. М.: ФИЗМАТЛИТ, 2000. 622 с.
3. *Воеводин В. В.* Вычислительные методы линейной алгебры. М.: Наука, 1977. 304 с.
4. *Петров И. Б., Лобанов А. И.* Лекции по вычислительной математике. М.: БИНОМ. Лаборатория знаний, 2006. 522 с.

Дополнительная литература

5. Голуб Дж., Ван Лоун Ч. Матричные вычисления. М.: Мир, 1999. 548 с.
6. Фаддеев Д. К., Фаддеева В. Н. Вычислительные методы линейной алгебры. СПб.: Лань, 2002. 736 с.
7. Коновалов А. Н. Введение в вычислительные методы линейной алгебры. Новосибирск: Наука, 1993. 158 с.
8. Деммель Дж. Вычислительная линейная алгебра. Теория и приложения. М.: Мир, 2001. 429 с.
9. Аристова Е. Н., Завьялова Н. А., Лобанов А. И. Практические занятия по вычислительной математике. Ч. I. М.: МФТИ, 2014. 242 с.
10. Демченко В. В. (ред.). Упражнения и задачи контрольных работ по вычислительной математике. М.: МФТИ, 2017. 203 с.
11. Saad Y. Iterative Methods for Sparse Linear Systems. 2nd Ed. Philadelphia: Society for Industrial and Applied Mathematics, 2003. 547 p.



ПРИБЛИЖЕНИЕ ФУНКЦИЙ (АППРОКСИМАЦИЯ ФУНКЦИЙ В ФУНКЦИОНАЛЬНЫХ ПРОСТРАНСТВАХ). МЕТОД НАИМЕНЬШИХ КВАДРАТОВ (МНК)

4.1. Постановка задачи

Определение 4.1. *Обобщенным полиномом* называется линейная комбинация следующего вида:

$$Q_m(x) = u_0\varphi_0(x) + \dots + u_m\varphi_m(x) = \sum_{i=0}^m u_i\varphi_i(x), \quad (4.1)$$

где $\varphi_i(x)$, $i = 0, \dots, m$, — система базисных функций, обладающих необходимыми свойствами гладкости.

Требуется приближенно заменить (аппроксимировать) заданную функцию $f(x)$ полиномом $Q_m(x)$ так, чтобы отклонение $f(x)$ от $Q_m(x)$ было наименьшим, в некотором заданном смысле, что достигается путем соответствующего выбора коэффициентов u_0, \dots, u_n . При этом $Q_m(x)$ называется *аппроксимирующим полиномом*, x может принадлежать как непрерывному отрезку $[a, b]$, так и точечному множеству

$$\omega_n = \{x_k = kh, h = (b - a)/n, k = 0 \div n\}.$$

В вычислительной практике часто используются в качестве базисных степенные или тригонометрические функции:

$$x^k; \quad \sin kx, \quad \cos kx, \quad k = 0, \dots, n.$$

Если рассматривается точечное множество $\{x_k\}_{k=0}^n$ и выполняется условие, представляющее собой систему линейных алгебраических уравнений:

$$Q_m(x_k) = f(x_k), \quad k = 0, \dots, n,$$

т.е. в узловых точках значения аппроксимируемой функции и аппроксимирующего полинома наиминимальшей возможной степени совпадают, то решается задача интерполяции.

В случае определения коэффициентов обобщенного полинома $Q_m(x)$ из решения задачи минимизации функции (квадратичного отклонения) на точечном множестве $\{x_k\}_{k=0}^n$:

$$\xi_n = \sum_{k=0}^n [Q_m(x_k) - f(x_k)]^2 \quad (4.2)$$

приходим к задаче об аппроксимации функции методом наименьших квадратов; при этом количество разбиений рассматриваемого отрезка превышает степень полинома $Q_m(x)$.

Если мы хотим приблизить непрерывную функцию $f(x)$ на отрезке $[a, b]$ обобщенным полиномом $Q_m(x)$ (4.1), то коэффициенты u_k , $k = 0, \dots, n$, подбираются из условия минимума квадратичного отклонения:

$$\xi_n = \int_a^b [Q_m(x) - f(x)]^2 dx = \int_a^b \left(\sum_{i=0}^m u_i \varphi_i(x) - f(x) \right)^2 dx; \quad (4.3)$$

что, в случае (4.2), в соответствии с МНК, минимум достигается путем приравнивания частных производных ξ_n по u_i к нулю:

$$\frac{\partial \xi_n}{\partial u_i} = 0, \quad i = 0, \dots, m; \quad (4.4)$$

в результате получаем систему линейных алгебраических уравнений порядка m .

Определение 4.2. Среднеквадратичное отклонение функции $Q_m(x)$ от $f(x)$ на отрезке $[a, b]$ определим как величину

$$\bar{\xi} = \frac{1}{b-a} \sqrt{\int_a^b [f(x) - Q_m(x)]^2 dx}; \quad (4.5)$$

среднеквадратичное отклонение $Q_m(x)$ от $f(x)$ на множестве точек $X = \{x_1, \dots, x_n\}$ имеет вид

$$\bar{\xi}_n = \sqrt{n^{-1} \sum_{i=1}^n [f(x_i) - Q_m(x_i)]^2}. \quad (4.6)$$

Формулу (4.5) можно рассматривать как предельный случай (4.6) при $n \rightarrow \infty$.

Во многих случаях при обработке результатов численных, лабораторных или натурных экспериментов квадратичное или среднеквадратичное приближение оказывается вполне приемлемым. Однако в некоторых случаях, при более жестких требованиях

к аппроксимации функции $f(x)$ полиномом $Q_m(x)$, пользуются другой оценкой погрешности, называемой *абсолютным отклонением*:

$$\bar{\xi} = \max_{x \in [a, b]} |f(x) - Q_m(x)|. \quad (4.7)$$

Определение 4.3. Полином $Q_m(x)$, доставляющий минимум невязке (4.7), называется *полиномом наилучшего приближения*.

Если обобщенный полином $Q_m(x)$ аппроксимирует функцию $f(x)$ на системе точек $X = \{x_1, \dots, x_n\}$, то погрешность (абсолютное отклонение) определяется следующим образом:

$$\bar{\xi}_n = \max_{1 \leq k \leq n} |f(x_k) - Q_m(x_k)|. \quad (4.8)$$

Таким образом, задача о наилучшем приближении функции $f(x)$ обобщенным полиномом $Q_m(x)$ состоит в нахождении коэффициентов u_k , ($k = 0, \dots, n$), полинома, при которых величина ξ_n будет минимальной. При выборе ξ_n в виде (4.5), (4.6) имеем задачу о наилучшем среднеквадратичном приближении, при выборе ξ_n в виде (4.7), (4.8) — о наилучшем равномерном приближении.

4.2. Существование и единственность полинома наилучшего приближения

Теорема 4.1 (существование и единственность обобщенного полинома наилучшего равномерного приближения для $\varphi_i = x^i$). Пусть $f(x) \in C[a, b]$. Тогда на $[a, b]$ существует единственный обобщенный полином наилучшего равномерного приближения $\tilde{Q}_m(x) = \sum_{i=0}^m \tilde{u}_i \varphi_i(x)$:

$$\tilde{Q}_m(x) = \arg \left(\inf_{Q_m(x)} \left\{ \max_{x \in [a, b]} |f(x) - Q_m(x)| \right\} \right).$$

Теорема 4.2 (Чебышёва об альтернансе). Пусть функция $f(x)$ непрерывна на отрезке $[a, b]$, содержащем не менее $(n+2)$ точек. В этом случае среди полиномов степени не выше n полином $Q_m(x)$ является полиномом наилучшего равномерного приближения для непрерывной функции $f(x)$ на данном отрезке тогда и только тогда, когда на $[a, b]$ \exists система из по крайней мере $(m+2)$ точек:

$$\zeta_0 < \zeta_1 < \dots < \zeta_{m+1} \quad (4.9)$$

(называемых чебышёвским альтернансом), таких, что разность

$$f(x) - Q_m(x)$$

поочередно принимает в них наибольшие по модулю положительные и отрицательные значения M и $-M$, где

$$M = \max_{x \in [a, b]} |f(x) - Q_m(x)|.$$

Эта теорема говорит о том, что максимальная ошибка аппроксимации функции многочленом наилучшего приближения реализуется в числе точек, на 2 большем, чем степень многочлена при чередовании знаков [2].

В простейшем случае для $f(x) = \cos 2x$, $x \in [0, 2\pi]$, полином 3-й степени наилучшего равномерного приближения имеет вид

$$Q_3(x) = 0,$$

так как в этом случае на всем рассматриваемом отрезке

$$M = \max_{x \in [a, b]} |\cos 2x - 1| = 0,$$

разность

$$f(x) - Q_m(x) = \cos 2x$$

принимает последовательные значения ± 1 в пяти точках:

$$0, \quad \frac{\pi}{2}, \quad \pi, \quad \frac{3}{2}\pi, \quad 2\pi,$$

которые для данного примера являются чебышёвским альтернансом.

Теперь приведем полную постановку задачи о наилучшем приближении функций, предложенной в работах П. Л. Чебышёва и являющейся одной из основных в функциональном анализе.

Задача о наилучшем приближении элемента $x \in X$ аппроксимирующим множеством A (X — метрическое пространство с метрикой $\rho(x, y)$, $x, y \in X$ — элементы X , $A \subset X$) состоит в определении функционала

$$G(x, A, X) = \rho(x, A) = \inf_{y \in A} \{\rho(x, y)\}, \quad (4.10)$$

представляющего собой расстояние между элементом $x \in X$ и множеством $A \subset X$ (наилучшее приближение). Элемент $u \in A$, для которого выполняется

$$G(x, A, X) = \rho(x, u), \quad (4.11)$$

$u \in X$, есть элемент наилучшего приближения для x (или экстремальный элемент):

$$u = \arg \inf_{y \in A} \{\rho(x, y)\}. \quad (4.12)$$

Определение 4.4. Множество $A \subset X$, обладающее тем свойством, что для $\forall x \in X$ в нем всегда существует элемент $u \in A$ наилучшего приближения, называется *множеством существования* (или *чебышёвским множеством*, если этот элемент единственен). В случае чебышёвского множества любому $x \in X$ соответствует ближайший к нему элемент $u \in A$.

Задача наилучшего приближения в линейном нормированном пространстве формулируется следующим образом.

Пусть R^m — m -мерное линейное пространство с нормой $\|\cdot\|_{R^m}$, образованное системой линейно независимых элементов

$$\varphi_0, \dots, \varphi_m \in R^m.$$

Среди всевозможных линейных комбинаций вида

$$Q_m = u_0\varphi_0 + \dots + u_m\varphi_m$$

требуется найти элемент $\varphi \in R^m$, который наименее уклоняется от аппроксимируемой функции $f \in R^m$, т. е. доставляет минимум величине

$$\|f - \varphi\|_R = \min_{\varphi \in R^m} \|f - \varphi\|_R.$$

Совокупность всех таких линейных комбинаций называется *линейной оболочкой* системы линейно независимых элементов $\{\varphi_k\}$, или *линейным многообразием*.

Достаточным условием единственности элемента наилучшего приближения является строгая нормированность пространства R , т. е. $\|f + g\| = \|f\| + \|g\|$, что достигается только тогда, когда $f = \alpha g$, $\alpha > 0$. В функциональном анализе доказывается, что гильбертово пространство H является строго нормированным.

Теорема 4.3. Пусть A — замкнутое выпуклое множество в гильбертовом пространстве H ; $x \notin A$, $x \in H$. В этом случае \exists единственный элемент $u \in A$ такой, что

$$\rho(x, A) = \|x - u\|_H.$$

Иными словами, множество должно содержать все свои предельные точки, т. е. быть полным, причем \forall отрезок, соединяющий две точки $x_1 \in A$ и $x_2 \in A$, должен целиком входить в H , т. е. $\forall \alpha \in [0, 1]$ $x = \alpha x_1 + (1 - \alpha) x_2 \in H$.

Если функции $f(x) \in X$, $X = C[a, b]$, A_m — множество полиномов степени не выше m :

$$A_m = \left\{ Q_m(x); Q_m(x) = \sum_{k=0}^m u_k x^k \right\},$$

то можно доказать существование и единственность экстремального полинома $\overline{Q}_m(x)$ — полинома наилучшего приближения для функции $f(x)$. Нахождение такого полинома представляет собой довольно сложную задачу.

Пусть $f(x) \in X$, $X = L_2(-\pi, \pi)$, где L_2 — пространство функций, интегрируемых с квадратом; A_m — множество тригонометрических полиномов $Q_m(x)$, степени не выше m :

$$A_m = \left\{ Q_m(x); \quad Q_m(x) = \sum_{k=0}^m a_k \cos kx + b_k \sin kx \right\}.$$

В этом случае экстремальным полиномом $\overline{Q}_m(x)$, или полиномом, наименее (в среднем) отклоняющимся от $f(x)$, является тригонометрический полином, коэффициенты a_k , b_k которого являются коэффициентами Фурье функции $f(x)$:

$$a_0 = (2\pi)^{-1} \int_{-\pi}^{\pi} f(x) dx; \quad a_k = \pi^{-1} \int_{-\pi}^{\pi} f(x) \cos kx dx;$$

$$b_k = \pi^{-1} \int_{-\pi}^{\pi} f(x) \sin kx dx; \quad k = 1, \dots, n.$$

Теорема 4.4 [1, гл. 2]. Пусть в гильбертовом пространстве H заданы множество A и точка $x \in H$; расстояние от x до A есть

$$d = \rho(x, A) = \min_{y \in A} \|x - y\|,$$

$$u = \arg \min_{y \in A} \|x - y\|.$$
(4.13)

Тогда выпуклое замкнутое множество A в H является множеством существования и единственности, или чебышёвским множеством (u — проекция элемента $x \notin A$ на A).

4.3. Сходимость полинома наилучшего приближения

Если проблема существования и единственности полинома наилучшего приближения решена, то встает вопрос о его сходимости к аппроксимируемой функции $f(x)$ при определенных предположениях о ее гладкости. В случае, когда базисные функции $\varphi_k(t)$ являются степенными, ответ на этот вопрос дает следующая теорема функционального анализа.



Теорема 4.5 (Вейерштрасса). Если аппроксимируемая функция $f(x) \in C[a, b]$, то для $\forall \varepsilon > 0$ \exists полином $Q_m(x)$ (т.е. m и $a_i, i = 0, \dots, m$) такой, что для $\forall x \in [a, b]$ выполняется

$$|f(x) - Q_m(x)| < \varepsilon.$$

Существуют и более точные оценки наилучших равномерных приближений, требующие определенной степени гладкости $f(x)$.

Теорема 4.6. Если аппроксимируемая функция $f(x)$ непрерывна на отрезке $[-1, 1]$ и имеет непрерывную производную $f^{(k)}(x)$, удовлетворяющую условию Липшица:

$$\left| f_{(x)}^{(k)} - f_{(y)}^{(k)} \right| \leq L |x - y|, \quad x, y \in [-1, 1],$$

то \exists алгебраический полином $P_n(x)$ порядка не выше n такой, что

$$\left\| f_{(x)}^{(k)} - Q_m^{(k)}(x) \right\|_C \leq \frac{M^{k+1} (k+1)^{k+1} L}{(k+1)! m^{k+1}},$$

где M — не зависящая от k, n , L константа, L — постоянная Липшица [2].

Примером полиномов наилучшего приближения являются полиномы Чебышёва первого рода, наименее уклоняющиеся от функции $f(x) = 0$, поскольку нормированный полином

$$\overline{T}_n(x) = 2^{-(n-1)} \cos(n \cdot \arccos x)$$

имеет коэффициент при старшей производной, равный единице, а на отрезке $[-1, 1]$ он имеет экстремальные значения $\pm \frac{1}{2^{n-1}}$, достигаемые в точках

$$x_i = \cos \frac{(i-1)\pi}{n}, \quad i = 0, \dots, n.$$



4.4. Полиномы Бернштейна

Прикладной интерес также представляют полиномы Бернштейна, имеющие вид:

$$B_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) C_n^k x^k (1-x)^{n-k} = \sum_{k=0}^n f\left(\frac{k}{n}\right) B_{n,k}(x),$$

$$C_n^k = \frac{n!}{(n-k)!k!}, \quad x \in [0, 1],$$

$B_{n,k}(x)$ — базисный полином Бернштейна n -го порядка.

Преобразование

$$t = xb - (1 - x)a, \quad x \in [0, 1],$$

переводит единичный отрезок в отрезок $[a, b]$; тогда

$$B_{n,k}(t, a, b) = B_{n,k}\left(\frac{t-a}{b-a}\right) = \frac{1}{b-a} C_n^k (t-a)^k (b-t)^{n-k}. \quad (4.14)$$

При этом имеет место

Теорема 4.7. Если аппроксимируемая функция $f(x)$ на отрезке $[-1, 1]$ удовлетворяет условию Липшица:

$$|f(x) - f(y)| \leq L|x - y|,$$

то имеет место оценка

$$|f(x) - B_n(x)| \leq \frac{L}{2\sqrt{n}}.$$

Эта теорема устанавливает факт равномерной сходимости полиномов Бернштейна при $n \rightarrow \infty$ в случае слабых требований к гладкости $f(x)$. Также показывается, что данная оценка неумлучшаема.

Теорема 4.8. Если функция $f(x)$ имеет всюду на отрезке $[0, 1]$ непрерывную производную k -го порядка $f^{(k)}(x)$, то $B_n^{(k)}(x)$ сходится равномерно к $f^{(k)}(x)$.

Отметим еще одно свойство полинома Бернштейна: он реализует разложение единицы

$$1 = [(1-x) + x]^n = \sum_{k=0}^n B_{n,k}(x) = \sum_{k=0}^n C_n^k x^k (1-x)^{n-k}.$$

Кроме того, имеет место рекуррентное соотношение

$$B_{k,n}(x) = x B_{n-1,x}(x) + (1-x) B_{n-1,k}(x).$$

Для задач построения поверхностей и кривых с заданными свойствами широко используются кривые Безье, представляющие собой специальную форму записи полиномов Бернштейна:

$$B_n(b_k, x) = \sum_{k=0}^n b_k B_{n,k}(x),$$

где $B_{n,k}(x)$ — базисные полиномы Бернштейна, b_k — вещественные коэффициенты.

4.5. Аппроксимация тригонометрическими полиномами

Данный подход позволяет аппроксимировать непрерывные периодические функции периодическими полиномами вида

$$T_m(x) = a_0 + \sum_{k=1}^m (a_k \cos kx + b_k \sin kx),$$

осуществляющими равномерное приближение непрерывной периодической функции $f(x)$.

Теорема 4.9 (вторая теорема Вейерштрасса). *Если $f(x)$ — непрерывная периодическая функция с периодом 2π , то для $\forall \varepsilon > 0 \exists$ тригонометрический полином $T_m(x)$ такой, что для $\forall x \in (-\infty, \infty)$ имеет место неравенство*

$$|f(x) - T_m(x)| < \varepsilon.$$

Иными словами, \forall непрерывная периодическая функция периода 2π может быть представлена как предел равномерно сходящейся последовательности тригонометрических полиномов.

Теорема 4.10 (Джексона). *Если непрерывная периодическая функция $f(x)$ с периодом 2π имеет производную k -го порядка, удовлетворяющую условию Липшица, то \exists тригонометрический полином $T_m(x)$ порядка не более n такой, что*

$$\|f(x) - T_m(x)\| \leq \frac{M^{k+1} \cdot L}{m^{k+1}},$$

где L — константа Липшица, M не зависит от m, k, L .

4.6. Метод наименьших квадратов

Аппроксимацию функций можно реализовывать и в других функциональных пространствах; наиболее удобным в прикладных задачах оказалось пространство $L_2[a, b]$ функций, интегрируемых с квадратом и некоторым весом $\rho(x)$, с нормой

$$\|f\|_{L_2} = \sqrt{\int_a^b \rho(x) f^2(x) dx}. \quad (4.15)$$

В этом пространстве определены скалярное произведение

$$(f_1, f_2) = \int_a^b \rho(x) f_1(x) f_2(x) dx$$

и норма

$$\|\mathbf{f}\|_{L_2}^2 = (\mathbf{f}, \mathbf{f}),$$

поэтому оно является гильбертовым.

Важной прикладной задачей является аппроксимация функции $f(x)$ по известным ее значениям $f_k = f(x_k)$ в точках x_k ; $k = 0, \dots, n$.

В этом случае отклонение определяется суммой

$$\begin{aligned} I_n = \|\mathbf{f} - \mathbf{Q}_m\|_R &= \sqrt{\sum_{k=0}^n \rho_k [f_k - Q_m(x_k)]^2} = \\ &= \sqrt{(\mathbf{f} - \mathbf{Q}_m, \mathbf{f} - \mathbf{Q}_m)}; \quad (4.16) \end{aligned}$$

при этом $\mathbf{f} = \{f_k\}_0^n$, $\mathbf{Q} = \{Q_k\}_0^n$ — векторы $(n+1)$ -мерного пространства, а скалярное произведение определяется как сумма

$$(\mathbf{x}, \mathbf{y}) = \sum_{k=0}^n \rho_k x_k y_k.$$

Построение наилучшего квадратичного приближения \overline{Q} на системе $(m+1)$ линейно независимых элементов $\{\varphi_k\}_0^m \in L_2$ реализуется путем решения задачи о минимизации функционала невязки

$$r = \left\| \mathbf{f} - \sum_{i=0}^m x_i \varphi_i \right\|_{L_2},$$

которая сводится к решению системы линейных алгебраических уравнений относительно x_i следующего вида:

$$\frac{\partial}{\partial x_k} \left(\left(\mathbf{f} - \sum_{i=0}^m x_i \varphi_i \right), \left(\mathbf{f} - \sum_{i=0}^m x_i \varphi_i \right) \right) = 0, \quad k = 0, \dots, m.$$

Можно показать, что вторые производные полученного функционала по x_k положительны, т.е. последнее равенство обеспечивает минимум функционала.

В результате получается система вида

$$\mathbf{D}\mathbf{x} = \tilde{\mathbf{f}},$$

где $\mathbf{x} = \{x_0, \dots, x_m\}^T$,

$$\tilde{\mathbf{f}} = \{(\mathbf{f}, \varphi_0), (\mathbf{f}, \varphi_1), \dots, (\mathbf{f}, \varphi_n)\}^T$$

— векторы-столбцы,

$$\mathbf{D} = \{(\varphi_i, \varphi_j)\} = \left\{ \begin{pmatrix} (\varphi_0, \varphi_0) & (\varphi_1, \varphi_0) & \cdots & (\varphi_m, \varphi_0) \\ (\varphi_0, \varphi_1) & (\varphi_1, \varphi_1) & \cdots & (\varphi_m, \varphi_1) \\ \vdots & \vdots & \ddots & \vdots \\ (\varphi_0, \varphi_m) & (\varphi_1, \varphi_m) & \cdots & (\varphi_m, \varphi_m) \end{pmatrix} \right\}$$

— матрица Грама, симметрическая и положительно определенная;

$$(\varphi_i, \varphi_j) = \sum_{k=0}^n \varphi_i(x_k) \cdot \varphi_j(x_k),$$

(φ_i, φ_j) — скалярные произведения.

Этот метод называется *методом наименьших квадратов*. Он используется также для решения переопределенных систем линейных алгебраических уравнений вида

$$\begin{cases} a_{11}x_1 + \dots + a_{1m}x_m = f_1, \\ \vdots \\ a_{n1}x_1 + \dots + a_{nm}x_m = f_n, \end{cases} \quad (4.17)$$

$$n > m, \quad \mathbf{x} = \{x_1, \dots, x_m\}^T \in R^m, \quad \mathbf{f} = \{f_1, \dots, f_n\}^T \in L^n,$$

где R^m и R^n — линейные векторные нормированные пространства со скалярными произведениями вида

$$(\mathbf{x}, \mathbf{y})^n = \sum_{k=1}^n x_k y_k$$

И

$$[\mathbf{x}, \mathbf{y}]^n = (\mathbf{C}\mathbf{x}, \mathbf{y})^n, \quad \mathbf{C} = \mathbf{C}^* > 0; \quad \mathbf{x}, \mathbf{y} \in R^n,$$

C — весовая матрица.

Перепишем систему (4.17) в более удобном виде:

$$\mathbf{Ax} = \mathbf{f}, \quad \mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix}.$$

К такой системе может привести, например, задача о приближении таблично заданной функции $\{f_k\}_0^n$ степенным полиномом m -й степени. Например, для приближения функции, заданной тремя точками, полиномом первой степени, получим систему

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 = f_1 \\ a_{21}x_1 + a_{22}x_2 = f_2 \\ a_{31}x_1 + a_{32}x_2 = f_3 \end{array} \right\}.$$

Определим *обобщенное решение* полученной системы как элемент гильбертова пространства, доставляющий наименьшее значение функционалу

$$r(\mathbf{x}) = [\mathbf{Ax} - \mathbf{f}, \mathbf{Ax} - \mathbf{f}]^n.$$

Теорема 4.11 [1]. Пусть столбцы матрицы \mathbf{A} системы линейных алгебраических уравнений

$$\mathbf{Ax} = \mathbf{f}$$

линейно независимы, т.е. ранг матрицы равен m . В этом случае \exists единственный элемент пространства $v \in H_m$, являющийся обобщенным решением системы (4.17) и решением системы вида

$$\mathbf{A}^* \mathbf{CAx} = \mathbf{A}^* \mathbf{Cf}, \quad (4.18)$$

решение которой доставляет минимум скалярному произведению

$$r(\mathbf{x}) = [\mathbf{Ax} - \mathbf{f}, \mathbf{Ax} - \mathbf{f}]^n. \quad (4.19)$$

Доказательство. Пусть вектор $\mathbf{q}_k \in R^n$ является k -м столбцом матрицы \mathbf{A} :

$$\mathbf{q}_k = \{a_{1k}, \dots, a_{nk}\}^T; \quad k = 1, \dots, p.$$

Видно, что матрица

$$\mathbf{D} = \mathbf{A}^* \mathbf{CA}$$

квадратная (проверяется непосредственно), $\text{rang } \mathbf{D} = p$.

Элемент d_{ij} этой матрицы, стоящий на пересечении i -й строки и j -го столбца:

$$d_{ij} = (\mathbf{q}_i, \mathbf{Cq}_j)^n = (\mathbf{Cq}_i, \mathbf{q}_j)^n = [\mathbf{q}_i, \mathbf{q}_j]^n.$$

В силу коммутативности скалярного произведения $d_{ij} = d_{ji}$, т.е. матрица \mathbf{D} симметрическая: $\mathbf{D} = \mathbf{D}^*$.

Невырожденность матрицы \mathbf{D} следует из того, что ранг матрицы \mathbf{A} равен m . Также заметим, что $(\mathbf{f}, \mathbf{A}\boldsymbol{\xi})^n = (\mathbf{A}^*\mathbf{f}, \boldsymbol{\xi})^m$; $\mathbf{f} \in R^n$, $\boldsymbol{\xi} \in R^m$, что проверяется непосредственно представлением данного равенства в развернутом виде.

Тогда

$$0 < [\mathbf{Ax}, \mathbf{x}]^n = (\mathbf{CAx}, \mathbf{Ax})^n = (\mathbf{A}^* \mathbf{CAx}, \mathbf{x})^p = (\mathbf{Dx}, \mathbf{x})^p.$$

Так как матрица \mathbf{D} невырождена и положительно определена, то (4.18) имеет решение, которое обозначим $\mathbf{v} \in L^m$.

Покажем, что \mathbf{v} — единственное обобщенное решение (4.19).

Для этого установим справедливость неравенства

$$r(\mathbf{x} + \Delta) > r(\mathbf{x}) \quad (\Delta \neq 0)$$

для

$$\begin{aligned} r(\mathbf{v} + \Delta) &= [\mathbf{A}(\mathbf{v} + \Delta) - \mathbf{f}, \mathbf{A}(\mathbf{v} + \Delta) - \mathbf{f}]^n = \\ &= [(\mathbf{A}\mathbf{v} - \mathbf{f}) + \mathbf{A}\Delta, \mathbf{A}\mathbf{v} - \mathbf{f}, \mathbf{A}\Delta]^n = [\mathbf{A}\mathbf{v} - \mathbf{f}, \mathbf{A}\mathbf{v} - \mathbf{f}]^n - \\ &- 2[\mathbf{A}\mathbf{v} - \mathbf{f}, \mathbf{A}\Delta]^n + [\mathbf{A}\Delta, \mathbf{A}\Delta]^n = r(\mathbf{v}) + 2[\mathbf{A}\mathbf{v} - \mathbf{f}, \mathbf{A}\Delta]^n + \\ &+ (\mathbf{D}\Delta, \Delta)^m = r(\mathbf{v}) + 2(\mathbf{B}\mathbf{A}\mathbf{v} - \mathbf{B}\mathbf{f}, \mathbf{A}\Delta)(\mathbf{D}\Delta, \Delta)^m = \\ &= r(\mathbf{v}) + 2(\mathbf{D}\mathbf{v} - \mathbf{A}^*\mathbf{B}\mathbf{f}, \Delta)^m > r(\mathbf{v}), \end{aligned}$$

что и требовалось доказать. При этом использовалось:

$$\begin{aligned} [\mathbf{A}\mathbf{v} - \mathbf{f}, \mathbf{A}\Delta]^n &= (\mathbf{C}(\mathbf{A}\mathbf{v} - \mathbf{f}), \mathbf{A}\Delta)^n = \\ &= (\mathbf{A}^*\mathbf{C}\mathbf{A}\mathbf{v} - \mathbf{A}^*\mathbf{C}\mathbf{f}, \Delta)^m = 0, \end{aligned}$$

откуда

$$\mathbf{A}^*\mathbf{C}\mathbf{A}\mathbf{v} = \mathbf{A}^*\mathbf{C}\mathbf{f},$$

что и требовалось доказать.

Приведем пример. Пусть переопределенная система имеет следующий вид:

$$\begin{cases} a_{11}u_1 + a_{12}u_2 = f_1, \\ a_{21}u_1 + a_{22}u_2 = f_2, \\ a_{31}u_1 + a_{32}u_2 = f_3; \end{cases} \quad \mathbf{A}\mathbf{x} = \mathbf{f}; \quad \mathbf{A} = \begin{Bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{Bmatrix};$$

$$\mathbf{u} = \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix}; \quad \mathbf{f} = \begin{Bmatrix} f_1 \\ f_2 \\ f_3 \end{Bmatrix}.$$

Система уравнений, полученная в доказанной теореме, будет иметь вид

$$\mathbf{A}^*\mathbf{C}\mathbf{A}\mathbf{u} = \mathbf{A}^*\mathbf{C}\mathbf{f},$$

где

$$\mathbf{A}^* = \begin{Bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \end{Bmatrix}, \quad \mathbf{C} = \mathbf{E};$$

в таком случае для определения коэффициентов u_1, u_2 получим СЛАУ второго порядка

$$\begin{aligned} \begin{Bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \end{Bmatrix} \begin{Bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Bmatrix} \begin{Bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{Bmatrix} \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} &= \\ &= \begin{Bmatrix} a_{11} & a_{21} & a_{13} \\ a_{21} & a_{22} & a_{32} \end{Bmatrix} \begin{Bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{Bmatrix} \begin{Bmatrix} f_1 \\ f_2 \\ f_3 \end{Bmatrix}. \end{aligned}$$

После перемножений матриц получим

$$\begin{pmatrix} a_{11}^2 + a_{21}^2 + a_{31}^2 & a_{11}a_{12} + a_{21}a_{22} + a_{31}a_{32} \\ a_{11}a_{12} + a_{21}a_{22} + a_{31}a_{32} & a_{12}^2 + a_{22}^2 + a_{32}^2 \end{pmatrix} \cdot \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{Bmatrix} a_{11}f_1 + a_{21}f_2 + a_{31}f_3 \\ a_{12}f_1 + a_{22}f_2 + a_{32}f_3 \end{Bmatrix},$$

здесь:

$$\mathbf{D} = \mathbf{A}^* \mathbf{C} \mathbf{A} = \begin{Bmatrix} a_{11}^2 + a_{21}^2 + a_{31}^2 & a_{11}a_{12} + a_{21}a_{22} + a_{31}a_{32} \\ a_{11}a_{12} + a_{21}a_{22} + a_{31}a_{32} & a_{12}^2 + a_{22}^2 + a_{32}^2 \end{Bmatrix};$$

$$\mathbf{u} = \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix}; \quad \mathbf{A}^* \mathbf{C} \mathbf{f} = \begin{Bmatrix} a_{11}f_1 + a_{21}f_2 + a_{31}f_3 \\ a_{12}f_1 + a_{22}f_2 + a_{32}f_3 \end{Bmatrix}.$$

Если система векторов удовлетворяет равенству

$$[\mathbf{q}_i, \mathbf{q}_j]^m = \delta_{ij}; \quad i, j = 1, \dots, m,$$

то матрица $\mathbf{D} = \mathbf{A}^* \mathbf{C} \mathbf{A}$ оказывается единичной, тогда решение системы (4.18) будет иметь простой вид

$$\mathbf{x} = \mathbf{A}^* \mathbf{C} \mathbf{f}.$$

Необходимо сказать, что при $m > 5$, если базисные функции не выбираются специальным образом, система алгебраических уравнений (4.18) часто оказывается плохо обусловленной, например при $\varphi_i(x) = x^i$, $i = 1, \dots, m$. В этом случае получаем систему вида

$$\left\{ \begin{array}{l} u_1 + \left(\sum_{i=0}^n x_i \right) u_2 + \dots + \left(\sum_{i=0}^n x_i^m \right) u_m = \sum_{i=0}^n f(x_i), \\ \left(\sum_{i=0}^n x_i \right) u_1 + \left(\sum_{i=0}^n x_i^2 \right) u_2 + \dots + \left(\sum_{i=0}^n x_i^{m+1} \right) u_m = \sum_{i=0}^n f(x_i) u_m, \\ \dots \dots \dots \\ \left(\sum_{i=0}^n x_i^m \right) u_1 + \left(\sum_{i=0}^n x_i^{m+1} \right) u_2 + \dots + \left(\sum_{i=0}^n x_i^{2m+1} \right) u_m = \\ = \sum_{i=0}^n x_i^m f(x_i). \end{array} \right.$$

Если скалярные произведения выбрать в интегральном виде

$$(\varphi_i, \varphi_j) = \int_0^1 x^i \cdot x^j dx,$$

то после минимизации функционала

$$\Phi(\mathbf{u}) = \int_0^1 [F(x) - g(x)]^2 dx,$$

где $F(x)$ — заданная функция, $g(x) = \sum_{k=0}^m u_k x^k$ — аппроксимирующий полином, получим систему линейных алгебраических уравнений вида

$$\mathbf{H}_{m+1} \cdot \mathbf{u} = \mathbf{G}, \quad \mathbf{H}_{m+1} = \{i + j - 1\}_{i,j=1}^{m+1}.$$

Матрица \mathbf{H}_{m+1} называется *матрицей Гильберта* и является классическим примером плохо обусловленной матрицы;

$$\mathbf{u} = \{u_0, \dots, u_m\}^m; \quad \mathbf{G} = \left\{ \int_0^1 F(x) dx, \dots, \int_0^1 x^m F(x) dx \right\}.$$

При $m = 1$: $\mu = \|\mathbf{H}\| \cdot \|\mathbf{H}^{-1}\| \approx 20$, при $m = 9$: $\mu \approx 10^{13}$.

Если для аналогичной аппроксимации используется не отрезок, а система точек, то при стремлении их количества к ∞ мы получим матрицу Гильберта.

Основная идея так называемого *предобуславливания* матрицы системы линейных алгебраических уравнений, с целью улучшить ее обусловленность, состоит в замене исходной системы $\mathbf{A}\mathbf{u} = \mathbf{f}$ на эквивалентную $\mathbf{C}\mathbf{A}\mathbf{u} = \mathbf{C}\mathbf{f}$, где матрица $\mathbf{C}\mathbf{A}$ будет хорошо обусловленной, либо на систему вида $(\mathbf{C}^{-1}\mathbf{A}\mathbf{C}^{-1})(\mathbf{C}\mathbf{u}) = \mathbf{C}^{-1}\mathbf{f}$.

В последнем случае матрица \mathbf{C} выбирается симметрической, положительно определенной, хорошо обусловленной.

Список литературы

1. Рябенский В. С. Введение в вычислительную математику. М.: ФИЗМАТЛИТ. 2008. 288 с.
2. Ильин В. П. Численный анализ. Ч. 1. Новосибирск: ИВМиМГ СО РАН, 2004. 334 с.

ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ НЕЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ

5.1. Введение

Методы численного решения нелинейных алгебраических уравнений и систем нелинейных алгебраических уравнений СНАУ в основном являются итерационными (или методами последовательных приближений) и имеют много общего с методами решения задач оптимизации (чаще всего это задачи поиска минимума функции от нескольких переменных). Заметим, что слово «итерация» происходит от латинского *iterare* — «еще раз вспахать». В математике это слово означает повторение некоторой заданной математической операции.

Постановка задач о поиске минимума функции многих переменных имеет следующий вид: найти значение \mathbf{t} , доставляющее

$$\inf_{\mathbf{t}} \{ \Phi(\mathbf{t}) \}, \quad \mathbf{t} \in R^n, \quad \mathbf{t} = \begin{Bmatrix} t_1 \\ \vdots \\ t_n \end{Bmatrix} \quad (5.1)$$

при условиях

$$\begin{aligned} \Phi_i(\mathbf{t}) &\geq 0, & i &= 1, \dots, I, \\ \Phi_i(\mathbf{t}) &= 0, & i &= 1, \dots, J; \end{aligned}$$

при этом

$$\mathbf{t} = \arg \min_{\mathbf{t}} \Phi(\mathbf{t}).$$

Такие задачи возникают как при решении задач оптимизации, например, ресурсов в математической экономике, так и при решении вариационных задач математической физики. Функция $\Phi(t)$ предполагается достаточно гладкой, например, имеющей вторые непрерывные производные.

Рассматриваемые системы алгебраических нелинейных уравнений имеют вид

$$\mathbf{F}(\mathbf{t}) = 0,$$

где \mathbf{F} — вектор-столбец:

$$\mathbf{F} = \begin{Bmatrix} f_1(\mathbf{t}) \\ \vdots \\ f_n(\mathbf{t}) \end{Bmatrix}, \quad \mathbf{t} \in R^n.$$



Задачи минимизации функции и решения СНАУ, вообще говоря, сводятся друг к другу. Так, если $\min_t \Phi(t)$ достигается в точке $t_0 \in R^n$ и функция $\Phi(t)$ дифференцируема в этой точке, то эта точка является решением системы уравнений

$$\frac{\partial \Phi}{\partial t_i} = 0, \quad i = 1, \dots, n. \quad (5.2)$$

С другой стороны, если некоторая функция $\Phi(t) \geq 0$ при $\{t_1, \dots, t_n\} \neq \{0, 0, \dots, 0\}$, $\Phi(0, \dots, 0) = 0$, то решение СНАУ

$$F(t) = 0$$

равносильно минимизации функционала

$$\Phi[f_1(t), \dots, f_n(t)],$$

т. е. для поиска системы уравнений строится некий функционал, минимум которого достигается на решениях системы. Обычно эти методы итерационные. *Областью сходимости* метода называется множество начальных условий, при которых итерации сходятся к решению рассматриваемой задачи.

5.2. Неподвижная точка отображения, сжимающий оператор

Определение 5.1. *Оператором (или отображением) F называется закон, по которому каждому элементу $z \in Z$ однозначно ставится в соответствие определенный элемент z' из множества Z' (Z и Z' могут совпадать): $F: Z \rightarrow Z'$ или $z' = F(z)$. Говорят, что отображение F действует из Z в Z' ; отображение $F: Z \rightarrow Z$ — преобразование множества Z в себя. Если Z, Z' — числовые множества, то оператор F называют функцией.*

Определение 5.2. Пусть X — полное метрическое пространство с метрикой (расстоянием).

$$\rho(x, y), \quad x, y \in X,$$

Ω — замкнутое множество в X .

Пусть также на Ω задан оператор P , переводящий Ω в себя, т. е. $P: \Omega \rightarrow \Omega$.

Элемент $z \in \Omega$ называется *неподвижной точкой* оператора P , если имеет место равенство

$$z = P(z), \quad (5.3)$$

— т. е. неподвижная точка является решением уравнения (5.3).

Если в итерационном процессе

$$z_k \xrightarrow[k \rightarrow \infty]{} z,$$

то такая точка называется *притягивающей*, а в противном случае — *отталкивающей*.

Напомним, что полным называется метрическое пространство, в котором всякая фундаментальная последовательность $\{t_n\}_0^\infty$ элементов сходится к своему пределу, принадлежащему этому пространству. Последовательность элементов $\{t_n\}_0^\infty$ называется *фундаментальной* (сходящейся в себе), если $\forall \varepsilon > 0 \exists n_0(\varepsilon)$ такое, что

$$\rho(t_n, t_m) < \varepsilon \quad \text{при} \quad n, m \geq n_0(\varepsilon).$$

Определение 5.3. Оператор (отображение) P называется *сжимающим* (или *оператором сжатия*) на Ω , если при $\forall x, y \in \Omega$ выполняется условие

$$\rho(P(x), P(y)) \leq q \rho(x, y), \quad (5.4)$$

где $0 \leq q < 1$ — коэффициент сжатия.

Если последовательность $x_k \in \Omega$, $k = 0, 1, \dots$, такая, что

$$x_{k+1} = P(x_k), \quad k = 0, 1, \dots,$$

то оператор P задает на Ω *итерационный процесс*, а последовательность $\{x_k\}$ называется *итерационной*.

Итерационный процесс состоит из двух частей: первая — локализация корня, вторая — уточнение корня. Для первой части полезно напомнить теорему из курса математического анализа: если непрерывная функция $f(x)$ принимает значения разных знаков на концах отрезка $[a, b]$, т. е. $f(a) \cdot f(b) < 0$, то внутри этого отрезка содержится по крайней мере один корень уравнения

$$f(x) = 0,$$

т. е. \exists хотя бы одно число $\xi \in [a, b]$ такое, что

$$f(\xi) = 0.$$

При этом $x = \xi$ будет заведомо единственным, если производная $f'(x)$ существует и сохраняет постоянный знак внутри отрезка $[a, b]$, т. е. если $f'(x) > 0$ или $f'(x) < 0$ на отрезке $[a, b]$.

На этой теореме базируется метод деления отрезка пополам (бисекции), применяемый как для локализации (отделения) корней, так и для их уточнения. Пусть корень находится на отрезке $[a, b]$;

границы a_0, b_0 нулевого приближения отрезка локализации корня положим равным координатам его концов:

$$a_0 = a, \quad b_0 = b,$$

первое приближение корня c_1 — их среднему арифметическому:

$$c_1 = \frac{1}{2} (a_0 + b_0).$$

Если $f(c_1) = 0$ (точнее, $f(c_1) = \varepsilon$, где ε — заданная точность численного решения задачи), то решение найдено.

Пусть $f(c_1) f(b_0) > 0$; тогда полагаем:

$$a_1 = a_0, \quad b_1 = c_1,$$

если $f(c_1) f(a_1) > 0$, $a_1 = c_1$, $b_1 = b_0$ и т. д.

На n -м шаге получим отрезок $[a_n, b_n]$ длины

$$b_n - a_n = \frac{b - a}{2^n}, \quad n = 0, 1, \dots$$

Если принять $c_n = \frac{1}{2} \cdot (a_n + b_n)$ за приближенное решение на n -м шаге, то получим (x — точное решение):

$$|x - \xi_n| \leq \frac{1}{2} \cdot \frac{b - a}{2^n}.$$

Тогда для того чтобы вычислить корень с точностью ε , требуется выполнить количество итераций, равное

$$n \approx \left\lceil \frac{\frac{b - a}{2\varepsilon}}{\ln 2} \right\rceil + 1.$$

5.3. Метод простых итераций (МПИ)

Метод простых итераций состоит в следующем. Представим систему уравнений $\mathbf{F}(\mathbf{x}) = 0$ в виде

$$\mathbf{x} = \mathbf{P}(\mathbf{x}),$$

или

$$(5.5)$$

$$x_i = P_i(x_1, \dots, x_n), \quad i = 1, \dots, n,$$

и построим итерационный процесс:

$$\mathbf{x}_{k+1} = \mathbf{P}(\mathbf{x}_k), \quad \mathbf{x}_0 = \mathbf{a}, \quad k = 0, 1, \dots \quad (5.6)$$

для уточнения его решения; $\mathbf{x}_0, \mathbf{x}_k \in \Omega$, $k = 1, \dots$, Ω — полное метрическое пространство.

Теорема 5.1. Пусть P является сжимающим оператором на Ω . Тогда в Ω \exists единственное решение X системы (5.5), являющееся пределом последовательности $\{x_k\}$ из (5.6). При этом скорость сходимости оценивается в соответствии с неравенством

$$\rho(X, x_k) \leq \frac{q^k}{1-q} \rho(x_0, x_1), \quad x_k \in \Omega, \quad (5.7)$$

где q — коэффициент сжатия, $\rho(x_0, x_1) = \rho_0$ — расстояние между первым и начальным приближениями к решению (5.5).

Доказательство. Поскольку (5.5) — сжимающий оператор, то

$$\begin{aligned} \rho(x_{k+1}, x_k) &= \rho[P(x_k), P(x_{k-1})] \leq \\ &\leq q \cdot \rho(x_k, x_{k-1}) \leq \dots \leq q^k \cdot \rho(x_0, x_1) = q^k \cdot \rho_0. \end{aligned}$$

В таком случае можно построить цепочку неравенств при $p > k$:

$$\begin{aligned} \rho(x_p, x_k) &\leq \rho(x_p, x_{p-1}) + \dots + \rho(x_{k+1}, x_k) \leq q^{p-1} \cdot \rho_0 + \\ &+ \dots + q^k \cdot \rho_0 \leq q^k \cdot \rho_0 \sum_{i=0}^{\infty} q^i = \rho_0 \frac{q^k}{1-q}, \end{aligned}$$

откуда видно, что, в соответствии с критерием Коши существования предела последовательности, последовательность $\{x_k\}$, $k = 0, 1, \dots$, стремится к своему пределу X , поскольку правая часть стремится к нулю при $k \rightarrow \infty$ (последовательность $\{x_k\}$, $k = 0, 1, \dots$, сходится, если для $\forall \varepsilon > 0 \exists$ номер $N > 0$ такой, что при $\forall k > N$ и \forall натуральных p, k ($p > k$) выполняется

$$\rho(x_k, x_{k+1}) < \varepsilon).$$

Далее, переходя в полученном неравенстве к пределу при $p \rightarrow \infty$, получим

$$\rho(X, x_k) \leq \rho_0 \frac{q^k}{1-q}.$$

Докажем, что X является корнем рассматриваемого уравнения. Для этого рассмотрим в полном метрическом пространстве расстояние между двумя элементами $X, P(X)$ и воспользуемся неравенством треугольника:

$$\begin{aligned} \rho(X, P(X)) &\leq \rho(X, x_{k+1}) + \rho(x_{k+1}, P(X)) = \\ &= \rho(X, x_{k+1}) + \rho(P(x_k), P(X)) \leq \rho_0 \frac{q^{k+1}}{1-q} + q \rho(x_k, X) \leq \\ &\leq \rho_0 \frac{q^{k+1}}{1-q} + q \rho_0 \frac{q^k}{1-q} = 2 \rho_0 \frac{q^{k+1}}{1-q}. \end{aligned}$$

Так как k — произвольное натуральное число, а левая часть от k не зависит, то

$$\rho(\mathbf{X}, \mathbf{P}(\mathbf{X})) = 0,$$

т. е.

$$\mathbf{X} = \mathbf{P}(\mathbf{X}).$$

Если рассматривается банахово (полное линейное нормированное) пространство B , то

$$\rho(x, y) = \|x - y\|.$$

Обычно $\{x_k\} \in B$; $k = 0, 1, \dots$, т. е. итерационные процессы рассматриваются в банаховых пространствах (С. Л. Соболев: «... теория вычислений, которую сейчас так же невозможно себе представить без банаховых пространств, как и без электронных вычислительных машин.» [5, с. 7]).

Для скалярного нелинейного уравнения получим

$$\rho(x_{k+1}, x_k) = \|x_{k+1} - x_k\| = |x_{k+1} - x_k|; \quad x_k \in [a, b];$$

тогда при $x_k \in [a, b]$:

$$\begin{aligned} |x_{k+1} - x_k| &= |F(x_k) - F(x_{k-1})| \leq \max_{[a, b]} |F'(\theta)| |x_k - x_{k-1}| \leq \dots \\ &\dots \leq \left(\max_{[a, b]} |F'(\theta)| \right)^k |x_1 - x_0|, \quad (5.8) \end{aligned}$$

откуда следует сходимость итераций при $\max_{[a, b]} |F'(\theta)| \leq q < 1$; $x_k \in [a, b]$, $k = 1, 2, \dots$; $\theta \in [a, b]$.

Определение 5.4. Область $\Omega \in X$ называется *выпуклой*, если наряду с любыми двумя точками $\mathbf{x}, \mathbf{y} \in \Omega$ она включает все точки соединяющего их отрезка:

$$\mathbf{z} = \mathbf{x} + t(\mathbf{y} - \mathbf{x}), \quad 0 \leq t \leq 1.$$

Теорема 5.2. Пусть область $\Omega \in R^n$ выпукла, а компоненты функции

$$f(\mathbf{x}) = \begin{cases} f_1(x_1, \dots, x_n) \\ f_2(x_1, \dots, x_n) \\ f_n(x_1, \dots, x_n) \end{cases}$$

имеют непрерывные производные первого порядка в Ω .

Тогда оператор $\mathbf{P}(\mathbf{x})$ является сжимающим в Ω , т. е.

$$\|\mathbf{P}(\mathbf{x}) - \mathbf{P}(\mathbf{y})\| \leq q \|\mathbf{x} - \mathbf{y}\|; \quad \mathbf{x}, \mathbf{y} \in \Omega,$$

если норма матрицы

$$\mathbf{J} = \frac{\partial \mathbf{P}(\mathbf{x})}{\partial \mathbf{x}} = \begin{pmatrix} \frac{df_1}{dx_1} & \cdots & \frac{df_1}{dx_n} \\ \vdots & \ddots & \vdots \\ \frac{df_n}{dx_1} & \cdots & \frac{df_n}{dx_n} \end{pmatrix}$$

не превосходит единицы.

Достаточным условием сходимости итерационного процесса для СНАУ будет неравенство

$$\|\mathbf{J}\| < 1,$$

где \mathbf{J} — матрица Якоби для рассматриваемой системы уравнений.

5.4. Метод Ньютона

Представим функцию $F(x)$, входящую в правую часть уравнения

$$x = F(x); \quad x, F, f \in R,$$

в виде

$$F(x) = x + \tau f(x),$$

где τ — итерационный параметр:

$$x_{k+1} = x + \tau f(x_k), \quad x_0 = a.$$

Значение итерационного параметра выбираем из условия

$$|F'(x)| < 1, \quad F'(x) = 1 + \tau f'(x).$$

Если положить (предельный случай)

$$F'(x) \approx 0, \quad \text{то} \quad \tau = -[f'(x)]^{-1},$$

в результате чего получим итерационный процесс следующего вида:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \quad x_0 = a. \quad (5.9)$$

Проведем линеаризацию функции $f(x)$ путем ее разложения в ряд Тейлора:

$$f(x_k + \Delta x_k) = f(x_k) + f'_k(x_k) \cdot \Delta x_k + O(\Delta^2 x_k).$$

Положив $O(\Delta^2 x_k) \approx 0$, т. е. пренебрегая членами второго порядка малости, получим

$$f(x_k) + f'_x(x_k) \cdot \Delta x_k = 0,$$

Теорема 5.3 (о сходимости метода Ньютона). *Сделаем следующие предположения о функции $f(x)$ и начальном приближении:*

1. $f(x) \in C^2[a, b]$;
2. $\exists [f'_x(u)]^{-1}$;
3. $|f''(x)| \leq D_2$;
4. $|[f'(x)]^{-1}| \leq D_1$ (отображение $f(x)$ равномерно невырождено);
5. $D_1^2 D_2 |f(x_0)| \leq q < 1$.

При этих предположениях метод Ньютона сходится с квадратичной скоростью сходимости.

Доказательство. Разложим функции $f(x)$ в ряд Тейлора:

$$f(x_{k+1}) = f(x_k) + f'_x(x_k) \cdot \Delta x_k + O(\Delta^2 x_k), \\ \Delta x_k = x_{k+1} - x_k.$$

Для метода Ньютона

$$f(x_k) + f'_x(x_k) \cdot \Delta x_k = 0, \quad \text{или} \quad \Delta x_k = \frac{f(x_k)}{f'_x(x_k)}.$$

В таком случае из первого разложения получим

$$|f(x_k)| = O(\Delta^2 x_k) \leq D_2 \cdot \Delta^2 x_k = D_2 \left| \frac{f(x_k)}{f'_x(x_k)} \right|^2 \leq D_1 D_1^2 |f(x_k)|^2,$$

так как по условию

$$|f''(x)| \leq D_2, \quad |[f'(x)]^{-1}| \leq D_1.$$

Невязка ξ_k имеет вид

$$\xi_k = |f(x_k)|;$$

в таком случае $\xi_{k+1} \leq D \xi_k^2$, где $D = D_2 D_1^2$; далее следует цепочка неравенств:

$$\xi_1 \leq D \xi_0^2, \quad \xi_2 \leq D \xi_1^2 \leq D^3 \xi_0^4, \quad \xi_3 \leq D \xi_2^2 \leq D^7 \xi_0^8;$$

в результате будем иметь

$$\xi_{k+1} \leq D^{-1} (D \xi_0)^{2^k}.$$

Полученные неравенства

$$\xi_{k+1} \leq D \xi_k^2, \quad \xi_{k+1} \leq D^{-1} (D \xi_0)^{2^k}$$

позволяют определить порядок скорости сходимости (второй порядок для метода Ньютона).

Также очевидно, что метод Ньютона сходится при следующем условии:

$$D\xi_0 = D|f(x_0)| \leq q < 1,$$

т.е. получим ограничение на начальное приближение. Теорема доказана.

Аналогичным путем можно получить итерационные методы третьего и четвертого порядков сходимости, однако требования к $f(x)$ и x_0 будут существенно более жесткими:

$$\begin{aligned} u_{k+1} &= u_k - \frac{f_k}{f'_k} - \frac{f''_k \cdot f_k^2}{2(f'_k)^3}, \\ u_{k+1} &= u_k - \frac{f_k}{f'_k} - \frac{f''_k \cdot f_k^2}{2(f'_k)^3} - \frac{(f''_k)^2 \cdot f_k^3}{2(f'_k)^5}, \end{aligned} \quad (5.13)$$

где $f_k = f(x_k)$.

Уменьшить количество арифметических действий в методе Ньютона при решении СЛАУ позволяет метод Ньютона–Канторовича, т.е. обратная матрица $[f'_x(x_k)]^{-1}$ вычисляется не на каждой итерации, а один раз в точке $x = x_0$:

$$x_{k+1} = x_k - [f'_x(x_0)]^{-1} \cdot f(x_0), \quad x_0 = a.$$

В случае, когда $f(x)$ задана таблично, можно использовать метод секущих:

$$x_{k+1} = x_k - \frac{f(x_k)}{[f(x_k) - f(x_{k-1})]} \cdot \tau_k, \quad \tau_k = x_k - x_{k-1}, \quad x_0 = a. \quad (5.14)$$

Эта формула получается при замене $f(x)$ ее линейным интерполантом:

$$\tilde{f}(x) = f(x_k) + [f(x_{k+1}) - f(x_k)] \cdot \frac{x - x_k}{x_{k+1} - x_k}.$$

Разумеется, скорость сходимости этого метода уже не квадратичная.

Если функцию $f(x)$ приблизить полиномом второй степени, то получим (метод парабол):

$$\begin{aligned} f(x) &= f(x_k)(x - x_k)f(x_k, x_{k-1}) + \\ &\quad + (x - x_k)(x - x_{k-1}) \cdot f(x_k, x_{k-1}, x_{k-2}); \end{aligned}$$

обозначив $t = x - x_k$, получим квадратное уравнение вида

$$at^2 + bt + c = 0,$$

где

$$a = f(x_k, x_{k-1}, x_{k-2}),$$

$$b = f(x_k, x_{k-1}) + (x_k - x_{k-1}) \cdot f(x_k, x_{k-1}, x_{k-2}), \quad c = f(x_k).$$

В качестве следующего приближения выбирается тот из корней, который ближе к x_k .

Метод Ньютона с параметром, позволяющим ускорить итерационный процесс, имеет вид

$$F'(x_k) \frac{x_{k+1} - x_k}{\tau_{k+1}} + F(x_k) = 0,$$

или

$$x_{k+1} = x_k - \tau_{k+1} \frac{F(x_k)}{F'(x_k)}; \quad x_0 = a.$$

Для численного решения нелинейной системы

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

можно также применить метод Якоби:

$$f_i(x_1^k, \dots, x_{i-1}^k, x_i^{k+1}, x_{i+1}^k, \dots, x_n^k) = 0. \quad (5.15)$$

В этом случае для вычисления x^{k+1} необходимо решить n скалярных уравнений, например, каким-либо итерационным методом.

Модификацией (5.15) является метод Зейделя, учитывающий результаты вычислений на предыдущих итерациях:

$$f_i(x_1^{k+1}, \dots, x_{i-1}^{k+1}, x_i^{k+1}, x_{i+1}^k, \dots, x_n^k) = 0, \quad i = 1, \dots, n. \quad (5.16)$$

Если в (5.16) для определения значения $t_i = x_i^{k+1}$ используется итерационный метод Ньютона, то такой метод называют гибридным. Например, из (5.16) получим, линеаризуя функцию f_i :

$$\begin{aligned} \frac{\partial f_i}{\partial x_i}(x_i^{k+1}, x_i^{k+1}, \dots, x_{i-1}^{k+1}, y_i^s, x_{i+1}^k, \dots, x_n^k) (y_i^{s+1} - y_i^s) + \\ + f(x_1^{k+1}, x_{i-1}^{k+1}, \dots, y_i^s, x_{i+1}^k, \dots, x_n^k) = 0, \end{aligned}$$

где s — итерационный индекс, соответствующий внутренним итерациям по Ньютону, k — внешним по Зейделю, $s > 0, 1, \dots, l$; $y_i^0 = x_i^k$, $y_i^{s+1} = x_i^{k+1}$, $i = 1, \dots, n$. Здесь s — «внутренний», k — «внешний» итерационные индексы.

Итерационный процесс, проводимый по методу Ньютона, с итерационным индексом s для определения x_i^{k+1} , называется *внутренним итерационным процессом*, а процесс, реализуемый

рассмотрим итерационный процесс, который будет иметь следующий вид:

$$\begin{cases} \frac{\partial f_1(x_1^k, x_2^k)}{\partial x_1} (x_1^{k+1} - x_1^k) + f_1(x_1^k, x_2^k) = 0, \\ \frac{\partial f_2(x_1^{k+1}, x_2^k)}{\partial x_2} (x_2^{k+1} - x_2^k) + f_2(x_1^{k+1}, x_2^k) = 0. \end{cases} \quad (5.176)$$

Если не ограничиваться одной итерацией, то итерационный процесс будет иметь вид $(y_1 = x_1^{k+1}, y_2 = x_2^{k+1})$:

$$\begin{cases} \frac{\partial f_1(y_1^s, x_2^k)}{\partial x_1} (y_1^{s+1} - x_1^k) + f_1(y_1^s - x_2^k) = 0, \\ \frac{\partial f_2(x_1^{k+1}, y_2^k)}{\partial x_2} (y_2^{s+1} - x_2^k) + f_2(x_1^{k+1}, y_2^s) = 0. \end{cases}$$

Рассмотрим еще один вариационный подход к итерационным методам решения СНАУ на примере системы из двух уравнений:

$$\begin{cases} f(x, y) = 0, \\ g(x, y) = 0. \end{cases}$$

Построим функционал вида

$$\Phi(x, y) = f^2(x, y) + g^2(x, y).$$

Поскольку $\Phi \geq 0$, то \exists точка (ξ, η) такая, что

$$(\xi, \eta) = \arg \min_{x \in \Omega} \Phi(x, y).$$

При этом $f(x, y) = g(x, y) = 0$, т. е. минимум $\Phi(x, y)$ достигается на решении исходной СНАУ. Построим итерационный процесс (метод градиентного спуска):

$$\begin{Bmatrix} x \\ y \end{Bmatrix}_{k+1} = \begin{Bmatrix} x \\ y \end{Bmatrix}_k - \tau_k \begin{Bmatrix} \Phi'_x(x_{k-1}, y_k) \\ \Phi'_y(x_k, y_k) \end{Bmatrix},$$

где τ_k — итерационный параметр, выбираемый из условия минимальности $\Phi(x_{k+1}, y_{k+1})$ в заданном направлении.

В качестве примера рассмотрим также применение методов простой итерации (МПИ) Ньютона для системы из двух нелинейных уравнений

$$\begin{cases} f(x, y) = 0, \\ g(x, y) = 0. \end{cases}$$

Для МПИ имеем итерационный процесс:

$$\begin{cases} x_{k+1} = F_1(x_k, y_k) \\ y_{k+1} = F_2(x_k, y_k); \end{cases} \quad k = 0, 1, \dots; \quad x_0 = a, \quad y_0 = b.$$

Условие сходимости этих итераций, в соответствии с рассмотренной ранее теоремой ($\|\mathbf{J}\| < 1$, где \mathbf{J} — матрица Якоби) будет иметь вид

$$\begin{cases} \left| \frac{\partial F_1}{\partial x} \right| + \left| \frac{\partial F_2}{\partial x} \right| \leq q < 1, \\ \left| \frac{\partial F_1}{\partial y} \right| + \left| \frac{\partial F_2}{\partial y} \right| \leq q < 1, \end{cases}$$

или

$$\begin{cases} \left| \frac{\partial F_1}{\partial x} \right| + \left| \frac{\partial F_1}{\partial y} \right| \leq q < 1, \\ \left| \frac{\partial F_2}{\partial x} \right| + \left| \frac{\partial F_2}{\partial y} \right| \leq q < 1. \end{cases}$$

В соответствии с методом Ньютона построим итерационный процесс:

$$\begin{aligned} x_{k+1} &= x_k - J^{-1}(x_k, y_k) \begin{vmatrix} F(x_k, y_k) & F'_y(x_k, y_k) \\ G(x_k, y_k) & G'_y(x_k, y_k) \end{vmatrix} = \\ &= x_k - \frac{J_x^k}{J(x_k, y_k)}, \end{aligned}$$

$$\begin{aligned} y_{k+1} &= y_k - J^{-1}(x_k, y_k) \begin{vmatrix} F'_x(x_k, y_k) & F(x_k, y_k) \\ G'_x(x_k, y_k) & G(x_k, y_k) \end{vmatrix} = \\ &= y_k - \frac{J_y^k}{J(x_k, y_k)}. \end{aligned}$$

Здесь:

$$\begin{aligned} J(x, y) &= \begin{vmatrix} F'_x & F'_y \\ G'_x & G'_y \end{vmatrix} \neq 0; & J_x(x, y) &= \begin{vmatrix} F(x_k, y_k) & F'_y(x_k, y_k) \\ G(x_k, y_k) & G'_y(x_k, y_k) \end{vmatrix}; \\ J_y(x, y) &= \begin{vmatrix} F'_x(x_k, y_k) & F(x_k, y_k) \\ G'_x(x_k, y_k) & G(x_k, y_k) \end{vmatrix}. \end{aligned}$$

Отметим еще один важный факт. Существуют отображения, порождающие последовательности $\{x_k\}_0^\infty$, которые могут иметь несколько предельных (неподвижных) точек.

Классический примером такого отображения является логистическое разностное уравнение:

$$\begin{aligned} x_{k+1} &= \lambda x_k (1 - x_k), \quad u_0 = a; \\ x_k &\in [0, 1], \quad 0 < \lambda < 4. \end{aligned}$$

Это отображение порождает такие интересные явления, как бифуркации, притягивающие и отталкивающие циклы. Исследование этого интересного отображения представляет отдельный научный интерес [3].

Список литературы

1. Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. Численные методы. М.: ФИЗМАТЛИТ, 2000. 622 с.
2. Самарский А. А., Гулин А. В. Численные методы. М.: Наука, 1989. 430 с.
3. Шарковский А. Н., Майстренко Ю. А., Романенко Е. Ю. Разностные уравнения и их приложения. Киев: Наук. думка, 1986. 279 с.
4. Хейгеман Л., Янг Д. Прикладные итерационные методы. М.: Мир, 1986. 446 с.

Дополнительная литература

5. Соболев С. Л. Введение в теорию кубатурных формул. М.: Наука, 1974. 808 с.



МЕТОДЫ ИНТЕРПОЛЯЦИИ ФУНКЦИЙ

6.1. Постановка задачи

Рассмотрим линейную комбинацию непрерывных линейно независимых на отрезке $[a, b]$ функций $\varphi_k(x)$:

$$P(x) = \sum_{k=0}^n u_k \varphi_k(x); \quad (6.1)$$

она не равна нулю, если u_k не равны одновременно нулю.

Положим, что в узлах *сетки*

$$\omega_n = \{x_i = a + ih; \quad i = 0, \dots, n; \quad x_0 = a, \quad x_n = b, \\ h = (b - a) / n\} \quad (6.2)$$

заданы значения аппроксимируемой функции $f(x)$:

$$f(x_i) = f_i.$$

Оператор, задающий проекцию функции $f(x)$ на $[a, b]$ на сетку $\{x_i\}_0^n$, т. е. таблицу $f_i = \{f(x_i)\}_{i=0}^n$, называется *оператором ограничения*.

Задача интерполяции функции состоит в определении линейной комбинации (6.1) базисных функций $\varphi_k(t)$, удовлетворяющей уравнениям:

$$\sum_{k=0}^n u_k \varphi_k(x_i) = f_i; \quad i = 0, \dots, n. \quad (6.3)$$

Последнее соотношение называется *условиями интерполяции*, x_i — *узлами интерполяции*, $P(x)$ — *интерполирующей функцией* (интерполянт), $f(x)$ — *интерполируемой функцией*.

Если ввести вектор значений аппроксимируемой функции

$$\mathbf{f} = \{f_0, \dots, f_n\}^T$$

и вектор неизвестных коэффициентов

$$\mathbf{u} = \{u_0, \dots, u_n\}^T$$

то получим систему линейных алгебраических уравнений вида

$$\mathbf{A}\mathbf{u} = \mathbf{f}, \quad (6.4)$$

где

$$\mathbf{A} = \begin{pmatrix} \varphi_0(x_0) & \dots & \varphi_n(x_0) \\ \varphi_0(x_1) & \dots & \varphi_n(x_1) \\ \vdots & \dots & \vdots \\ \varphi_0(x_n) & \dots & \varphi_n(x_n) \end{pmatrix}$$

— квадратная матрица $(n+1) \times (n+1)$.

Для того чтобы решение задачи интерполяции существовало и было единственным, необходимо и достаточно, чтобы определитель матрицы \mathbf{A} был отличен от нуля (система функций $\varphi_k(t)$ должна быть линейно независима). Систему линейно независимых функций, образующих базис в функциональном пространстве $\{\varphi_k(x)\}_{k=0}^n$, называют системой Чебышёва.

К чебышёвским системам относятся, например, функции:

$$\begin{aligned} &\rho(x), x\rho(x), x^2\rho(x), \dots, x^n\rho(x), \quad \rho(x) > 0, \quad x \in [a, b]; \\ &e^x, e^{2x}, \dots, e^{nx}; \\ &1, \cos x, \cos(2x), \dots, \cos(nx) \text{ на } [0, \pi]; \\ &1, \operatorname{ch} x, \operatorname{sh} x, \operatorname{ch}(2x), \operatorname{sh}(2x), \dots, \operatorname{ch}(nx), \operatorname{sh}(nx). \end{aligned} \quad (6.5)$$

Систему можно симметризовать (симметризация по Гауссу); умножив слева и справа на \mathbf{A}^* :

$$\mathbf{A}^* \mathbf{A} \mathbf{u} = \mathbf{A}^* f,$$

где $\mathbf{G} = \mathbf{A}^* \mathbf{A}$ есть матрица Грама с элементами $\gamma_{kj} = (\varphi_k, \varphi_j) = \sum_{i=0}^n \varphi_k(x_i) \cdot \varphi_j(x_i)$; (φ_k, φ_j) — скалярное произведение векторов $\varphi_k, \varphi_j \in H$. В случае выбора системы ортогональных функций, т. е. $(\varphi_k, \varphi_j) = \delta_j^k$, задача интерполяции существенно упрощается: $\mathbf{u} = \mathbf{A}^* f$.

Ошибкой интерполяции называется функция

$$\xi(x) = f(x) - \varphi(x). \quad (6.6)$$

Итак, схема решения задачи интерполяции выглядит следующим образом:

$$f(x) \xrightarrow{r} \{f_k\}_{k=0}^n \xrightarrow{I} F(x),$$

где r — оператор ограничения, I — оператор интерполяции.

6.2. Интерполяционный полином в форме Лагранжа

Выберем в качестве базисных линейно независимых функций степенные:

$$\varphi_0(x) = 1, \quad \varphi_1(x) = x, \quad \varphi_2(x) = x^2, \dots, \quad \varphi_n(x) = x^n. \quad (6.7)$$

В этом случае для системы базисных функций (6.7) определитель системы (6.4) имеет вид

$$\begin{vmatrix} 1 & x_0 & \dots & x_0^n \\ 1 & x_1 & \dots & x_1^n \\ \dots & \dots & \dots & \dots \\ 1 & x_n & \dots & x_n^n \end{vmatrix} = \prod_{\substack{i \neq j \\ i, j=0, \dots, n}} (x_i - x_j) \neq 0. \quad (6.8)$$

Это определитель Вандермонда, который отличен от нуля при условии несовпадения интерполяционных узлов. Неравенство нулю данного определителя означает тот факт, что решение задачи полиномиальной интерполяции со степенными базисными функциями существует и единственно. Однако система оказывается плохо обусловленной для больших n .

При этом

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

если хотя бы один из коэффициентов $a_k \neq 0$; $k = 0, \dots, n$. Число этих базисных функций $(n + 1)$ есть размерность линейного функционального пространства, в котором они образуют базис. Напомним, что вообще говоря, базисом $(n + 1)$ -мерного линейного пространства может быть любая совокупность из $(n + 1)$ линейно независимых функций.

В пространстве Π_n полиномов степени не выше n , базисом может быть также следующая система функций:

$$l_i^n(x) = \prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j}; \quad i = 0, \dots, n, \\ a = x_0 < x_1 < \dots < x_n = b, \quad x \in [a, b].$$

При этом любой полином из пространства Π_n может быть представлен в виде, предложенном Лагранжем:

$$P_n(x) = \sum_{k=0}^n a_k l_k^n(x). \quad (6.9)$$

Поскольку

$$l_k^n(x_k) = \begin{cases} 1, & \text{если } i = k \\ 0, & \text{если } i \neq k, \end{cases}$$

то $a_k = f(x_k)$, $k = 0, \dots, n$; например, при $n = 1$ получим

$$P_1(x) = f_0 \frac{x - x_1}{x_0 - x_1} + f_1 \frac{x - x_0}{x_1 - x_0},$$

где f_0 и f_1 — значения интерполируемой функции в точках x_0 и x_1 ; $f_0 = P_1(x_0)$; $f_1 = P_1(x_1)$. Такое решение сняло проблему численного решения плохо обусловленной системы уравнений, однако появилась проблема устойчивости интерполяционного процесса.

Несложно проверить, что (6.9) удовлетворяет условиям интерполяции. Наиболее простой полином получается в случае равномерной сетки:

$$t = \frac{x - x_0}{h}, \quad h = \frac{b - a}{n}.$$

Тогда:

$$\begin{aligned} P_n(x) &= \sum_{i=0}^n l_i^n(x) f_i = \\ &= (-1)^n \frac{t(t-1)\dots(t-n)}{n!} \sum_{i=0}^n (-1)^i \frac{c_n^i f_i}{x-1}; \quad (6.10) \\ l_i^n(t) &= (-1)^{n+1} C_n^i \frac{t(t-1)\dots(t-n)}{(t-i)n!}, \end{aligned}$$

где

$$C_n^i = \frac{n!}{i!(n-i)!}$$

— число сочетаний из n по i ; l_i^n — базисные функции Лагранжа, не зависящие от $f(x)$ и h .

Теорема 6.1. Пусть дана система узлов $\{x_i\}_0^n$, среди которых нет совпадающих; $\{f(x_i)\}_0^n$ — значения интерполируемой функции $f(x)$ в этих узлах. В таком случае существует единственный интерполяционный полином

$$P_n(x, x_0, \dots, x_n, f(x_0), \dots, f(x_n)) = P_n(x),$$

степени не выше n , принимающий в узлах x_i , где $i = 0, \dots, n$, значения $f(x_i)$, $i = 0, \dots, n$.

Доказательство. Докажем единственность интерполяционного полинома. Пусть существуют два интерполяционных полинома $P_n^1(x)$, $P_n^2(x)$. Тогда их разность $R_n(x) = P_n^1(x) - P_n^2(x)$ также есть полином степени не выше n , имеющий $(n+1)$ корней в точках x_0, \dots, x_n . Однако любой полином, не равный тождественно нулю, имеет число корней, с учетом их кратности, равное его степени. По этой причине $R_n(x) \equiv 0$, тогда $P_n^1(x) = P_n^2(x)$, что и требовалось доказать.

6.3. Интерполяционный полином в форме Ньютона

Введем понятие разделенных разностей первого, второго, ...
..., i -го порядков, составленных соответственно по узлам x_0, \dots, x_n :

$$\begin{aligned} f(x_k, x_{k+1}) &= \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k}, \\ f(x_k, x_{k+1}, x_{k+2}) &= \frac{f(x_{k+1}, x_{k+2}) - f(x_k, x_{k+1})}{x_{k+2} - x_k}, \dots \\ \dots, f(x_k, x_{k+1}, \dots, x_{k+i}) &= \frac{f(x_{k+1}, \dots, x_{k+i}) - f(x_k, \dots, x_{k+i-1})}{x_{k+i} - x_k}. \end{aligned} \quad (6.11)$$

Методом математической индукции доказывается формула

$$f(x_k, x_{k+1}, \dots, x_{k+i}) = \sum_{\substack{r=0 \\ k+r \neq k+j}}^k \frac{f(x_{k+j})}{\prod (x_{k+j} - x_{k+r})}.$$

Интерполяционный многочлен Лагранжа может быть представлен в виде

$$L_n(x) = L_0(x) + [L_1(x) - L_0(x)] + \dots + [L_n(x) - L_{n-1}(x)].$$

Каждая из разностей $L_k(x) - L_{k-1}(x)$ представляет собой многочлен k -го порядка, имеющий корни в точках x_0, \dots, x_{k-1} . Поэтому эти разности определяются с точностью до постоянной c_k :

$$L_k(x) - L_{k-1}(x) = c_k (x - x_0) \dots (x - x_{k-1}).$$

Для того чтобы найти c_k , положим в этом равенстве $x = x_k$ и учтем условия интерполяции

$$L_n(x_k) = f_k.$$

После формальных, хотя и не очень простых, алгебраических преобразований, получим

$$c_k = f(x_0, \dots, x_k),$$

где $f(x_0, \dots, x_k)$ — разделенные разности k -го порядка.

Тогда

$$\begin{aligned} L_n(x) = N_n(x) &= f(x_0) + (x - x_0) f(x_0, x_1) + \\ &+ (x - x_0)(x - x_1) f(x_0, x_1, x_2) + \dots \\ &\dots + (x - x_0) \dots (x - x_{n-1}) f(x_0, \dots, x_n), \end{aligned}$$

где N_n — интерполяционный полином в форме Ньютона.

Интерполяционный полином в форме Ньютона также может быть получен следующим образом.

$$P_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \dots \\ \dots + a_n(x - x_0) \dots (x - x_{n-1}). \quad (6.12)$$
$$\left\{ \begin{array}{l} a_0 = f_0, \\ a_0 + a_1 (x_1 - x_0) = f_1, \\ a_0 + a_1 (x_2 - x_0) + a_2 (x_2 - x_0) (x_2 - x_1) = f_2, \\ \dots\dots\dots \\ a_0 + a_1 (x_n - x_0) + \dots + a_n (x_n - x_0) (x_n - x_{n-1}) = f_n. \end{array} \right. \quad (6.13)$$
$$\begin{aligned} a_0 &= f(x_0) = f_0; & a_1 &= f(x_0, x_1) = \frac{f_1 - f_0}{x_1 - x_0}; \\ a_2 &= f(x_0, x_1, x_2) = \frac{1}{x_2 - x_0} \left(\frac{f_2 - f_1}{x_2 - x_1} + \frac{f_1 - f_0}{x_1 - x_0} \right); \\ a_3 &= f(x_0, x_1, x_2, x_3); \dots & a_n &= f(x_0, x_1, \dots, x_n). \end{aligned}$$
$$N_n(x) = N_{n-1}(x) + f(x_0, \dots, x_n)(x - x_0) \dots (x - x_{n-1}). \quad (6.14)$$

Теорема 6.2. Пусть функция $f(x) \in C^n[x_0, x_n]$; $\{x_k\}_0^n$ — система узлов на отрезке $[x_0, x_n]$. Тогда существует точка η , для которой выполняется

$$f^{(n)}(\eta) = n!f(x_0, \dots, x_n). \quad (6.15)$$

$$\Phi(x) = f(x) - P_n(x, x_0, \dots, x_n, f)$$

обращается в нуль в $(n + 1)$ точках: x_0, \dots, x_n . По теореме Ролля ее производная обращается в нуль хотя бы в одной точке между каждыми x_k , $k = 0, \dots, x_n$. В таком случае функция $\Phi(x)$ обращается в нуль не менее чем в n точках, $\Phi'(x)$ — в $(n - 1)$ точках

и т. д. Тогда n -я производная $\Phi^{(n)}(x)$ имеет по крайней мере один нуль в некоторой точке η :

$$\Phi^{(n)}(\eta) = 0.$$

Продифференцируем (6.15) n раз и положим $x = \eta$:

$$0 = \Phi^{(n)}(\eta) = f^{(x)}(\eta) - \frac{d^n}{dt^n} P_n(x, x_0, \dots, x_n, f) \Big|_{x=\eta}.$$

При этом

$$\begin{aligned} \frac{d^n}{dx^n} P_n &= \frac{d^n}{dx^n} [P_{n-1}(x) + f(x_0, \dots, x_n)(x - x_0) \dots (x - x_{n-1})] = \\ &= 0 + n! f(x_0, \dots, x_n), \end{aligned}$$

откуда и следует (6.15), т. е. разделенные разности позволяют аппроксимировать производную на заданном отрезке $[a, b]$.

Следствие. Для вычисления функции

$$f(x_{n+1}) = P_{n+1}(x)$$

по формуле Ньютона достаточно к функции $f(x_{n+1}) = P_{n+1}(x)$ добавить один член:

$$N_n(x) = N_{n-1}(x) + f(x_0, \dots, x_n)(x - x_0) \dots (x - x_n).$$

В случае записи интерполяционного полинома в форме Лагранжа полином необходимо полностью изменить.

6.4. Конечные разности

Определение 6.1. Пусть сетка, образованная узлами интерполяции $\{x_k\}_0^n$, является равномерной. Конечными разностями первого, второго, третьего, четвертого и т. д. порядков соответственно назовем величины:

$$\Delta f_k = f_{k+1} - f_k;$$

$$\Delta^2 f_k = \Delta f_{k+1} - \Delta f_k = f_{k+2} - 2f_{k+1} + f_k;$$

$$\Delta^3 f_k = \Delta^2 f_{k+1} - \Delta^2 f_k = f_{k+2} + 3f_{k+1} - f_k;$$

$$\Delta^4 f_k = \Delta^3 f_{k+1} - \Delta^3 f_k = f_{k+4} - 4f_{k+3} + 6f_{k+2} - 4f_{k+1} + f_k;$$

$$\dots \dots \dots$$

$$\Delta^n f_k = \Delta^{n-1} f_{k+1} - \Delta^{n-1} f_k; \quad k \geq 1.$$

Будем также считать, что $\Delta^0 f_k = f_k$ — разность нулевого порядка.

Методом математической индукции доказывается формула

$$\Delta^n f_k = \sum_{s=0}^n (-1)^{n-1} C_n^s f_{k+s}, \quad (6.16)$$

где $C_k^s = \frac{k!}{s!(k-s)!}$ — биномиальные коэффициенты.

Теорема 6.3. Пусть $f(x) \in C^n[a, b]$, узлы сетки $x_k \in [a, b]$. Тогда существует точка $\xi \in [a, b]$ такая, что

$$f_{(x)}^{(n)} \approx \Delta^n f_k / h^n, \quad h = x_{k+1} - x_k.$$

Доказательство теоремы проводится с помощью теоремы Лагранжа о среднем.

6.5. Погрешность интерполяции

Теорема 6.4. Пусть $f(x) \in L[x_n, x_{n+1}]$, т.е. $f(x)$ — липшиц-непрерывная функция на отрезке $[x_n, x_{n+1}]$.

В этом случае справедливо неравенство

$$|f(x) - F(x)| \leq C \frac{h}{2},$$

где $F(x)$ — интерполирующая функция:

$$F(x) = \frac{f_{n+1}(x - x_n) + f_n(x_{n+1} - x)}{x_{n+1} - x_n}, \quad x \in [x_n, x_{n+1}].$$

Доказательство. Обозначим шаг сетки

$$h = x_{n+1} - x_n;$$

тогда:

$$x = x_n + \alpha h; \quad F(x) = \alpha f_{n+1} + (1 - \alpha) f_n,$$

в силу линейности $F(x)$; $0 \leq \alpha \leq 1$.

Оценим погрешность:

$$\begin{aligned} |F(x) - f(x)| &= |\alpha f_{n+1} + (1 - \alpha) f_n - \alpha f(x) - (1 - \alpha) f(x)| \leq \\ &\leq \alpha |f_{n+1} - f(x)| + (1 - \alpha) |f_n - f(x)|. \end{aligned}$$

Так как $f_{n+1} = f(x_n + h)$, то

$$\begin{aligned} |f_{n+1} - f(x)| &= |f(x_n + h) - f(x_n + \alpha h)| \leq \\ &\leq |C(1 - \alpha)h| = C(1 - \alpha)h; \end{aligned}$$

аналогично

$$|f_n - f(x)| \leq C\alpha h.$$

Тогда

$$\|F(x) - f(x)\| \leq 2\alpha(1 - \alpha)Ch \leq C \frac{h}{2},$$

что и требовалось доказать.

Заметим, что с виду простой аппарат кусочно-линейной интерполяции позволяет ввести так называемые *конечные элементы* — финитные базисные функции, на которые опирается один из наиболее известных численных методов математической физики — метод конечных элементов. На сетке $\{x_k\}_0^n$ строится набор функций $\varphi_k(x)$, каждая из которых сопоставляется своему узлу x_k так, что

$$\varphi_k(x_i) = \delta_k^i, \quad \varphi_k(x_{i-1}) = \varphi_k(x_{i+1}) = 0, \quad \varphi_k(x_k) = 1;$$

в остальных точках отрезка $[x_{i-1}, x_{i+1}]$ значения функции вычисляются с помощью линейной интерполяции.

На всем отрезке $[a, b]$ функция $F(x)$ представляется в виде

$$F(x) = \sum_{k=0}^n f_k \cdot \varphi_k \in H. \quad (6.17)$$

Это выражение является одним из видов кусочной интерполяции, часто применяемой в вычислительной математике.

Рассмотрим погрешность интерполяции более высокого порядка.

Теорема 6.5. Пусть интерполируемая функция $f(x)$ принадлежит чебышёвскому пространству n раз непрерывно дифференцируемых функций: $f(x) \in C^n[a, b]$. В этом случае для остаточного члена интерполяции справедлива формула:

$$\xi_n(x) = \frac{f^{n+1}(\xi)}{(n+1)!} \cdot \prod_{j=0}^n (x - x_j), \quad \eta \in [a, b].$$

Доказательство. Рассмотрим функцию

$$\varphi(y) = f(y) - L_n(y) - \xi_n(x) \frac{(y - x_0) \dots (y - x_n)}{(x - x_0) \dots (x - x_n)},$$

которая имеет по крайней мере $(n+1)$ -ю производную.

Кроме того, эта функция имеет на $[a, b]$ не менее $(n+2)$ нулей: это точки $y_k = x_k$; $k = 0, \dots, n$ (поскольку $f(x_n) = L(x_n)$), и последнее слагаемое обращается в них в нуль. При этом $(n+2)$ -м нулем является точка $x = t$, в силу определения остаточного члена:

$$\xi_n(x) = f(x) - L_n(x). \quad (6.18)$$

Далее, поскольку между каждыми двумя нулями непрерывно дифференцируемой функции имеется хотя бы один нуль ее производной, то на отрезке $[a, b]$ имеется по крайней мере $(n + 1)$ нулей функции $\xi'(x)$. Аналогичные рассуждения можно провести для ξ'', ξ''', \dots . В конечном счете можно утверждать, что существует точка $\eta \in [a, b]$ такая, что

$$\xi_n^{(n+1)}(\eta) = 0.$$

Продифференцируем $(n + 1)$ раз функцию $\xi_n(x)$ с учетом того, что

$$L_n^{(n+1)}(x) = 0,$$

и вычислим $\xi_n^{(n+1)}(x)$ в точке $x = \eta$.

$$\begin{aligned} \xi_n^{(n+1)}(\eta) &= f^{(n+1)}(\eta) - L_n^{(n+1)}(\eta) - \\ &\quad - \frac{d^{(n+1)}}{dy^{n+1}} \left[\xi_n(t) \frac{(y-x_0) \dots (y-x_n)}{(x-x_0) \dots (x-x_n)} \right] \Big|_{x=\eta} = 0; \\ \frac{d^{(n+1)}}{dy^{n+1}} \left[\frac{(y-x_0) \dots (y-x_n)}{(x-x_0) \dots (x-x_n)} \right] \Big|_{x=\eta} &= \frac{(n+1)!}{\prod_{j=0}^n (x-x_j)}. \end{aligned}$$

В таком случае

$$\xi_n(x) = \frac{f^{(n+1)}(\eta)}{(n+1)!} \prod_{i=0}^n (x-x_i).$$

Теорема 6.6 (о погрешности интерполяции на равномерной сетке). Пусть $\omega_n = \{x_k = kh, k = 0, \dots, n, h = \frac{b-a}{n}\}$ — равномерная сетка.

Тогда имеет место оценка

$$|\xi_n(x)| \leq \frac{h^{n+1}}{n+1} M_n,$$

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

Доказательство. Поскольку шаг h сетки ω_τ постоянен, то

$$x = x_k + \alpha h, \quad 0 \leq \alpha \leq 1, \quad k = 0, \dots, n-1.$$

В таком случае

$$x - x_i = kh + \alpha h - ih = (k + \alpha - i)h,$$



откуда следует

$$\prod_{i=0}^n (x - x_n) = h^{n+1} \prod_{i=0}^n (k + \alpha - i).$$

Простым перебором можно показать, что

$$\prod_{i=0}^n (k + \alpha - i) \leq n!.$$

Тогда остаточный член вычисляется следующим образом:

$$\xi_n(x) = \frac{f^{(n+1)}(\eta)}{(n+1)!} \prod_{i=0}^n (x - x_i), \quad \eta \in [a, b],$$

откуда

$$|\xi_n(x)| \leq \frac{h^{n+1}}{n+1} M_{n+1},$$

что и требовалось доказать.

Следствие. Аналогичным образом можно получить следующие оценки для заданной задачи экстраполяции при удалении точки x от интервала $[x_0, x_n]$. При $x \in [x_n, x_n + h]$ получим

$$|\xi_n(x)| \leq h^{n+1} \cdot \max_{\eta \in [x_n, x_n + h]} |f^{(n+1)}(\eta)|;$$

при $x \in [x_n + h, x_n + 2h]$:

$$|\xi_n(x)| \leq (n+2) h^{n+1} \cdot \max_{\eta \in [x_n, x_n + 3h]} |f^{(n+1)}(\eta)|,$$

при $x \in [x_n + 2h, x_n + 3h]$:

$$|\xi_n(x)| \leq \frac{(n+2)(n+3)}{2!} h^{n+1} \cdot \max_{\eta \in [x_n, x_n + 3h]} |f^{(n+1)}(\eta)|.$$

Отсюда видно, что процесс экстраполяции допустим на интервалах длины $\sim O(h)$.

Погрешность интерполяции может быть выражена и через разделенные разности:

$$\xi_n(x) = f(x) - L_n(x) = f(x, x_0, \dots, x_n) \cdot \prod_{i=0}^n (x - x_i).$$

6.6. Минимизация погрешности интерполяционного процесса

Поскольку остаточный член интерполяции имеет вид


$$|\xi_n(x)| \leq \frac{f^{(n+1)}(\eta)}{(n+1)!} \prod_{i=0}^n (x - x_i),$$

то минимизировать эту величину можно, решив так называемую *минимаксную задачу*, т. е. найдя такое расположение узлов сетки ω_n , чтобы достигнуть минимума функции:

$$\min_{\{x_i\}_0^n} \max_{x \in [-1, 1]} \left| \prod_{i=0}^n (x - x_i) \right|,$$

или

$$\{x_i\}_0^n = \arg \min_{\{x_i\}_0^n} \max_{x \in [-1, 1]} \left| \prod_{i=0}^n (x - x_i) \right|.$$

Решение этой задачи находится с помощью полиномов Чебышёва 1-го рода: 

$$T_0(x) = 1; \quad T_1(x) = x, \quad T_k(x) = \cos k\varphi, \quad (6.19)$$

где $\varphi = \arccos x$, $-1 \leq x \leq 1$, $k = 0, 1, \dots$. При этом:

$$\begin{aligned} T_{k-1}(x) &= \cos(k-1)\varphi, \quad T_{k+1}(x) = \cos(k+1)\varphi, \\ T_{k+1}(x) + T_{k-1}(x) &= \cos(k+1)\varphi + \cos(k-1)\varphi = \\ &= 2 \cos \varphi \cos k\varphi = 2T_1(x)T_k(x), \end{aligned}$$

откуда получим вид полиномов Чебышёва более высоких степеней ($k > 1$):

$$T_2(x) = 2x^2 - 1, \quad T_3(x) = 4x^3 - 3x, \quad \dots$$

Введем также нормированный полином Чебышёва:

$$\overline{T}_n(x) = \frac{1}{2^{n-1}} T_n(x).$$

Теорема 6.7 (Чебышёва). Среди всех полиномов степени $n \geq 1$

$$P_n(x) = u_0 + u_1x + \dots + u_nx^n$$

со старшим коэффициентом a_n , равным единице, наименьшее уклонение от нуля, равное $1/2^{n-1}$, имеет нормированный полином Чебышёва

$$\overline{T}_n(x) = \frac{1}{2^{n-1}} T_n(x), \quad x \in [-1, 1].$$

Иначе говоря, для любого полинома $P_n(x) = u_0 + u_1x + \dots + u_nx^n$, отличного от $\overline{T}_n(x)$, выполняется

$$\frac{1}{2^{n-1}} = \max_{x \in [-1, 1]} |\overline{T}_n(x)| < \max_{t \in [-1, 1]} |P_n(x)|.$$

Выбрав в качестве интерполяционных узлов нули полинома Чебышёва:

$$x_i = \cos \frac{2i+1}{n} \pi, \quad i = 0, \dots, n-1; \quad x \in [-1, 1],$$

или

(6.20)

$$x_i = \frac{a+b}{2} + \frac{b-a}{2} \cos \frac{2i+1}{n} \pi, \quad i = 0, \dots, n-1, \quad x \in [a, b],$$

мы получим минимальный остаточный член интерполяции $\xi_n(x)$.

Для произвольного отрезка $T_{n+1}(x)$ имеет вид:

$$T_{n+1}(x) = \frac{(b-a)^{n+1}}{2^{2n+1}} \cos \left[(n+1) \arccos \frac{2x - (a+b)}{b-a} \right].$$

При этом

$$\max \left| \prod_{i=0}^n (x - x_i) \right| = \frac{(b-a)^{n+1}}{2^{2n+1}}$$

и оценка

$$|\xi_n(x)| = |f(x) - L_n(x)| = \frac{M_{n+1}}{(n+1)!} \prod_{i=0}^n (x - x_i)$$

принимает следующий вид:

$$|\xi_n(x)| = |f(x) - L_n(x)| \leq \frac{M_{n+1}}{(n+1)!} \frac{(b-a)^{n+1}}{2^{2n+1}},$$

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

(6.21)

6.7. Сходимость интерполяционного процесса

После построения интерполяционного полинома возникает очевидный вопрос: будет ли погрешность

$$|\xi_n(x)| = |f(x) - L_n(x)|$$

стремиться к нулю при $n \rightarrow \infty$, ответ на который, вообще говоря, отрицателен. Контпримером является классический пример Бернштейна: последовательность полиномов Лагранжа, интерполирующих функцию $f(x) = |x|$, $x \in [-1, 1]$, на системе узлов

$x_i = -1 + i(2/n)$, не стремится с возрастанием n к интерполируемой функции $f(x)$ ни в одной точке, кроме точек: -1 ; 0 ; 1 . То же утверждение касается и классического примера с функцией Рунге $(1 + x^2)^{-1}$, $x \in [-5, 5]$.

Для приближения непрерывных функций доказана следующая теорема.

Теорема 6.8 (Вейтерштрасса). *Для любой непрерывной на отрезке $[a, b]$ функции $f(x)$ \exists полином $P_n(x)$, приближающий $f(x)$ с любой наперед заданной точностью, т. е. для $\forall \varepsilon > 0 \exists n(\varepsilon)$:*

$$\|f(x) - P_n(x)\|_{C[a,b]} \leq \varepsilon.$$

Отметим, что в данном случае речь идет, вообще говоря, о многочлене наилучшего приближения.

Если говорить об интерполяционном многочлене, то в этом случае потребуются определенные требования к гладкости функции.

Ограничением роста степени интерполяционного полинома является поведение $(n + 1)$ -й производной $M_{n+1} = |f^{(n+1)}(x)|$, присутствующей в выражении для $\xi_n(x)$. По этой причине сходимость интерполяционного процесса может быть установлена лишь для узкого класса функций, упоминаемых в теореме.

Теорема 6.9 [1]. *Положим, что функция $f(x)$ — целая, т. е. может быть представлена в виде степенного ряда*

$$f(x) = a_0 + a_1(x - x_0) + \dots + a_n(x - x_0)^n,$$

сходящегося при всех x .

В этом случае последовательность ее интерполяционных полиномов сходится равномерно на любой совокупности узлов $x_n \in [a, b]$ к $f(x)$ при $n \rightarrow \infty$, т. е. $\lim_{n \rightarrow \infty} |f(x) - L_n(x)| = 0$.

Возможность расходимости (или неустойчивости) интерполяционного процесса подтверждается следующей теоремой.

Теорема 6.10 (Фабера). *Для любой последовательности сеток, составленных из совокупности интерполяционных узлов $\omega_i = \{x_0, \dots, x_i\}$, $i = 0, \dots, n$; $x_i \in [a, b]$, найдется такая непрерывная функция $f(x)$, что построенный по ним интерполяционный полином не будет сходиться равномерно к $f(x)$ на $[a, b]$ при $(\omega_1 = \{x_0, x_1\}, \omega_2 = \{x_0, x_1, x_2\}, \dots)$.*

Ситуация с задачей интерполяции представляется не столь безрадостной, если привести формулировку теоремы Марцинкевича.

Теорема 6.11 (Марцинкевича). Для всякой непрерывной на отрезке $[a, b]$ функции $f(x)$, найдется такая последовательность сеток ω_k , для которой соответствующий интерполяционный процесс сходится равномерно, т.е.

$$\lim_{n \rightarrow \infty} |f(x) - L_n(x)| = 0.$$

Однако необходимо сказать, что построение таких сеток представляет собой довольно трудную задачу, которую, кроме того, придется решать (т.е. находить свою сетку), вообще говоря, для каждой функции. В вычислительной практике обычно избегают интерполяции полиномами высокой степени и используют кусочно-полиномиальную интерполяцию.

Чтобы получить количественные оценки явления неустойчивости интерполяционного процесса, представим полином Лагранжа с учетом погрешности интерполяции в виде

$$L_n(x) = \sum_{k=0}^n f_k \varphi_k^n(x) + \sum_{k=0}^n \delta f_k \varphi_k^n(x),$$

где φ_k^n — фундаментальные полиномы, δf_k — ошибки входных данных (в данном случае — в задании значений интерполируемой функции в узлах интерполяции). Последнее слагаемое, имеющее вид

$$\Delta_n = \sum_{k=0}^n \delta f_k \cdot \varphi_k^n(x),$$

определяет чувствительность $L_n(x)$ к ошибкам входных данных и вычислительным ошибкам. Сделаем простые оценки:

$$\max_{t \in [a, b]} |\Delta_n| \leq l_n \cdot \delta,$$

где

$$l_n = \max_{x \in [a, b]} \left| \sum_{k=0}^n \varphi_k^n(x) \right|; \quad \delta = \max_{x \in [a, b]} |\delta f_k|.$$

Здесь l_n — постоянная Лебега, при этом

$$l_n = \max_{t \in [a, b]} \Lambda_n(x),$$

где $\Lambda_n(x) = \sum_k |\varphi_k^n(x)|$ — функция Лебега.

Поскольку $\Lambda_n(x)$ зависит только от расположения узлов сетки, то и постоянная Лебега зависит только от сетки ω_n . Если оператор интерполяции рассматривать как оператор, переводящий элемент одного банахова пространства (сеточных функций)

в другое (непрерывно дифференцируемых), то постоянная Лебега есть норма такого оператора.

Теорема 6.12 (Бернштейна). Пусть последовательность сеток $\omega_n = \{x_k\}_0^n$ составлена из узлов Чебышёва. В этом случае имеет место оценка

$$l_n \leq 8 + \frac{4}{\pi} \ln(n+1).$$

Заметим, что также можно оценить постоянную Лебега для интерполяции на равномерной сетке:

$$l_n \sim 2^n.$$

Об устойчивости интерполяционного полинома по отношению к ошибкам (вычислений и δf_i) судят по величине постоянной Лебега. Приведем таблицу, демонстрирующую зависимость l_n от n для равномерной и чебышёвской сеток:

n	Равномерная сетка	Чебышёвская сетка
5	3,11	2,10
10	29,89	2,49
15	512,05	2,73
20	10986,53	2,90

Эти цифры не требуют комментария.

6.8. Другие виды интерполяции

Если интерполируемая функция $f(t)$ является периодической с периодом T , то естественна ее аппроксимация с помощью базисных функций вида

$$\varphi_k(t) = a_k \cos \frac{\pi kt}{T} + b_k \sin \frac{\pi kt}{T}, \quad k = 0, \dots, n.$$

При этом интерполяционный полином будет иметь вид

$$P_n(t) = \sum_{k=0}^n \varphi_k(t) = a_0 + \sum_{k=1}^n \left(a_k \cos \frac{\pi kt}{T} + b_k \sin \frac{\pi kt}{T} \right);$$

его коэффициенты находятся из системы линейных уравнений:

$$P_n(t_i) = f(t_i),$$

$$i = 1, 2, \dots, 2n+1, \quad t_{2n+1} - t_0 = T,$$

$\{t_i\}^{2n+1}$ — узлы интерполяции.

Интерполяция Паде (аппроксимация рациональными функциями) может использоваться, например, в случае наличия разрывов в интерполируемой функции:

$$P_{kl} = \frac{a_0 + a_1 t + \dots + a_{k-1} t^{k-1} + a_k t^k}{b_0 + b_1 t + \dots + b_{l-1} t^{l-1} + b_l t^l}. \quad (6.22)$$

Коэффициенты a_i , $i = 0, \dots, k$, и b_j , $j = 0, \dots, l$, находятся из условий интерполяции:

$$P_{kl}(t_i) = f(t_i), \quad i = 0, \dots, n. \quad (6.23)$$

Например, в случае интерполируемой функции вида

$$\varphi(t) = \frac{a_0 + a_1 t}{b_0 + t}$$

из условий $\varphi(t_k) = f_k(t)$, $k = i-1, i, i+1$, — значения в трех узлах t_{i-1}, t_i, t_{i+1} , получим систему из трех уравнений:

$$\begin{cases} a_0 + a_1 t_{i-1} - b_0 f_{i-1} = t_{i-1} f_{i-1}, \\ a_0 + a_1 t_i - b_0 f_i = t_i f_i, \\ a_0 + a_1 t_{i+1} - b_0 f_{i+1} = t_{i+1} f_{i+1}. \end{cases}$$

6.9. Многомерная интерполяция

Пусть интерполяционные узлы $\{x_n, y_m\}_{0,0}^{N,M}$ образованы пересечением прямых $x = x_n$, $n = 0, \dots, N$, и $y = y_m$, $m = 0, \dots, M$.

Построим линейный интерполянт для функции $F(x, y)$ внутри прямоугольника $x \in [x_n, x_{n+1}]$, $y \in [y_m, y_{m+1}]$. Для этого реализуется линейная интерполяция по x на каждой прямой $y = y_m$, затем — по y при $x = x_n$, с учетом значений функции, полученных на первом шаге. В итоге получим

$$\begin{aligned} F(x, y) = & f_{nm} \frac{(x - x_{n+1})(y - y_{m+1})}{(x_n - x_{n+1})(y_m - y_{m+1})} + \\ & + f_{n+1,m} \frac{(x - x_n)(y - y_{m+1})}{(x_{n+1} - x_n)(y_m - y_{m+1})} + \\ & + f_{n+1,m+1} \frac{(x - x_n)(y - y_m)}{(x_{n+1} - x_n)(y_{m+1} - y_m)} + \\ & + f_{n,m+1} \frac{(x - x_{n+1})(y - y_m)}{(x_n - x_{n+1})(y_{m+1} - y_m)}. \end{aligned}$$

Полином Лагранжа $L_{NM}(x, y)$ для функции $f(x, y)$ двух переменных имеет следующий вид:

$$L_{NM}(x, y) = \sum_{n=0}^N \sum_{m=0}^M f_{nm} \prod_{\substack{i=0 \\ j \neq n}}^N \prod_{\substack{j=0 \\ j \neq m}}^M \frac{(x - x_i)(y - y_j)}{(x_n - x_i)(y_m - y_j)}.$$

Для построения интерполяционного полинома для функции двух переменных $f(x, y)$ вводятся понятия частных производных раз-
деленных разностей, по аналогии с частными производными. Например, для разностей 1-го порядка:

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

$$f(x; y_0, y_1) = \frac{f(x, y_0) - f(x, y_1)}{(y_0 - y_1)}.$$

В случае разностей более высоких порядков $f(x_0, \dots, x_i; y_0, \dots, y_i)$ берутся от функции $f(x, y)$ сначала разности $(j - 1)$ -го порядка по y , затем от полученного выражения разности $(i - 1)$ -го порядка по x . Тогда двумерный интерполяционный полином Ньютона может быть записан в виде

$$N_n(x, y) = \sum_{i=0}^n \sum_{k=0}^i f(x_0, \dots, x_{i-k}; y_0, \dots, y_k) \times$$

$$\times \prod_{p=0}^{i-1-k} (x - x_p) \prod_{q=0}^{k-1} (y - y_q).$$

При проведении линейной интерполяции внутри треугольника с вершинами, обозначенными индексами 0, 1, 2 с координатами $\{x_i, y_i\}_{i=0}^2$, решается система трех линейных уравнений, полученная из условий интерполяции:

$$a_0 + a_1 x_i + a_2 y_i = f_i; \quad i = 0, 1, 2.$$

Ее решение представляется в виде полинома функции двух переменных

$$P_1(x, y) = a_0 + a_1 x_i + a_2 y :$$

$$P_1(x, y) = f_0 \frac{(x - x_1)(y_1 - y_2) - (y - y_1)(x_1 - x_2)}{(x_0 - x_1)(y_1 - y_2) - (y_0 - y_1)(x_1 - x_2)} +$$

$$+ f_1 \frac{(x - x_2)(y_2 - y_0) - (y - y_2)(x_2 - x_0)}{(x_1 - x_2)(y_2 - y_0) - (y_1 - y_2)(x_1 - x_2)} +$$

$$+ f_2 \frac{(x - x_0)(y_0 - y_1) - (y - y_0)(x_0 - x_1)}{(x_2 - x_0)(y_0 - y_1) - (y_2 - y_0)(x_0 - x_1)}.$$

6.10. Интерполяция с кратными узлами

Определение 6.2. Пусть в узлах $\{x_k\}_0^m$, $x_k \in [a, b]$, среди которых нет совпадающих, заданы значения интерполируемой функции $f(x_k)$ до порядка $(N_k - 1)$ включительно:

$$f(x_k), \quad f'(x_k), \dots, f^{(N_k-1)}(x_k),$$

т. е. известно $\sum_{j=0}^m N_j$ величин.

Полином $H_n(x)$ степени $n = \sum_{j=0}^m N_j - 1$, для которого выполняются условия интерполяции с кратными узлами $H_n^{(i)}(x_k) = f^{(i)}(x_k)$; $k = 0, \dots, m$; $i = 0, \dots, N_k - 1$, называется *интерполяционным полиномом Эрмита для функции $f(x)$* ; N_k — кратность узла t_k .

Теорема 6.11. Полином Эрмита $H_n(x)$ существует и единственен.

Этот полином имеет следующий общий вид:

$$H_n(x) = \sum_k^m \sum_{i=0}^{N_k-1} \varphi_{ki}(x) f^{(i)}(x_k), \quad (6.24)$$

где φ_{ki} — полиномы степени n (из-за громоздкости общего вида полинома Эрмита мы его не приводим).

Погрешность интерполяции с кратными узлами имеет вид

$$\xi_n(x) = \frac{f^{(n+1)}(\eta)}{(n+1)!} (x-x_0)^{N_0} (x-x_1)^{N_1} \dots (x-x_m)^{N_m}. \quad (6.25)$$

Например, если в трех точках x_0, x_1, x_2 заданы значения

$$f_0, f_1, f'_1, f_2,$$

то вид полинома таков:

$$\begin{aligned} H_3(x) = & f_0 \frac{(x-x_1)(x-x_2)}{(x_0-x_1)^2(x_0-x_2)} + \\ & + f_1 \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} \times \left(1 - \frac{(x-x_1)(2x_1-x_0-x_2)}{(x_1-x_0)(x_1-x_2)} \right) + \\ & + f_2 \frac{(x-x_0)(x-x_1)^2}{(x_2-x_0)(x_2-x_1)^2} + f'_1 \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_1-x_2)(x_1-x_0)}. \end{aligned} \quad (6.26)$$

Полином Эрмита третьего порядка можно получить, например, из решения соответствующей системы четырех линейных уравнений.

Если в двух точках отрезка $[x_0, x_1]$, т.е. на его концах, заданы значения f_0, f_1, f'_0, f'_1 , то кубический интерполяционный полином Эрмита будет иметь вид

$$H_3(x) = f_0 \frac{(x_1 - x)^2 [2(x - x_0) + h]}{h^3} + f'_0 \frac{(x_1 - x)^2 (x - x_0)}{h^2} + \\ + f_1 \frac{(x - x_0)^2 [2(x_1 - x) + h]}{h^3} + f'_1 \frac{(x - x_0)^2 (x - x_1)}{h^2},$$

где $h = x_1 - x_0$.

6.11. Кусочно-полиномиальная сплайн-интерполяция

Основной недостаток глобальной, т.е. на всем отрезке $[a, b]$, интерполяции — явление неустойчивости интерполяционного процесса, т.е. рост постоянной Лебега (нормы оператора интерполяции) с числом узлов и, соответственно, рост погрешности

$$\xi_n(x) = f(x) - F(x).$$

Для устранения этого недостатка интерполяцию можно проводить на каждом из элементарных отрезков $[x_k, x_{k+1}]$, на которые разбивается весь отрезок $[a, b]$. Этот процесс называется кусочно-полиномиальной интерполяцией. Основным видом этой интерполяции является сплайн-интерполяция (сплайн — от *spline*, означает гладкость линии; идея произошла от использования гибких линеек в чертежном деле).

Главное преимущество сплайн-интерполяции — устойчивость вычислительного процесса, сходимость при измельчении сетки.

Определение 6.3. Пусть на отрезке $[a, b]$ задана система узлов

$$\omega_n = \{x_k; k = 0, \dots, n; a = x_0, b = x_n\}.$$

Сплайном $S_{m,d}(x)$ дефекта d называется определенная на $[a, b]$ функция, имеющая l непрерывных производных и являющаяся на каждом элементарном отрезке $[x_{k-1}, x_k]$ полиномом степени m ; при этом

$$d = m - l.$$

Если на $[a, b]$ задана непрерывная функция $f(x)$ и

$$S_{m,d}(x_k) = f(x_k), \quad k = 0, \dots, n,$$

то $S_{m,d}(x_k)$ называется *интерполяционным сплайном*.

В соответствии с этим определением, кусочно-линейная функция является сплайном первой степени дефекта 1: $S_{1,1}(x)$; кусочно-квадратичная первой степени дефекта 1: $S_{2,1}(x)$. В приложениях часто используется кубический сплайн $S_{3,1}(x)$, который обычно обозначают $S(x)$ и называют *естественным сплайном*.

Теорема 6.14 (о построении, существовании и единственности естественного сплайна). Пусть на отрезке $[a, b]$ задана непрерывная функция $f(x)$ и система узлов

$$\{x_k\}_0^m, \quad x_0 = a, \quad x_n = b, \quad h_i = x_i - x_{i-1}.$$

Пусть также выполняются условия:

- 1) $S_k(x) = f(x_k)$, $k = 0, \dots, n$ — условия интерполирования;
- 2) $S_k(x) \in C^2[a, b]$, т.е. $S(x)$ — непрерывная с двумя своими первыми производными функция;
- 3) на каждом элементарном отрезке (элементе) $[x_k, x_{k+1}]$, $k = 1, \dots, n$:

$$S(x) = \sum_{j=0}^3 a_j x^j,$$

т.е. $S(x)$ является полиномом третьей степени;

- 4) краевые условия для $S(x)$ представляются в одном из следующих видов:

$$4.1) S'(a) = f'(a), S'(b) = f'(b);$$

$$4.2) S''(a) = f''(a), S''(b) = f''(b); \text{ иногда полагают } S''(a) = S''(b) = 0 \text{ — так называемый свободный сплайн};$$

$$4.3) S(a) = S(b), S'(a) = S'(b) \text{ — периодические краевые условия}.$$

Д о к а з а т е л ь с т в о теоремы можно проводить несколькими способами.

На каждом элементарном отрезке $[x_{k-1}, x_k]$, $k = 1, \dots, n$, представляем сплайн-интерполянт в виде

$$S_k(x) = a_k + b_k(x - x_k) + \frac{c_k}{2}(x - x_k)^2 + \frac{d_k}{6}(x - x_k)^3,$$

a_k, b_k, c_k, d_k — коэффициенты, которые необходимо определить из условий теоремы.

Имеем:

$$S'_k(x) = b_k + c_k(x - x_k) + \frac{d_k}{2}(x - x_k)^2,$$

$$S''_k(x) = c_k + d_k(x - x_k),$$

$$S'''_k(x) = d_k,$$

откуда:

$$\begin{aligned} a_k &= S_k(x_k), & b_k &= S'_k(x), \\ c_k &= S''_k(x_k), & d_k &= S'''_k(x_k). \end{aligned}$$

Далее используем условия интерполяции

$$a_k = f(x_k); \quad k = 1, \dots, n,$$

доопределенные условия

$$a_0 = f(x_0)$$

и условия непрерывности $S(x)$:

$$S_k(x_k) = S_{k+1}(x_k); \quad k = 1, \dots, n,$$

откуда получим:

$$\begin{aligned} h_k b_k - \frac{h_k^2}{2} c_k + \frac{h_k^3}{6} d_k &= f_k - f_{k-1}, \\ k &= 1, \dots, n, \quad h_k = x_k - x_{k-1}, \end{aligned}$$

а также условия непрерывности первой и второй производных

$$S'_k(x_k) = S'_{k+1}(x_k), \quad S''_k(x_k) = S''_{k+1}(x_k), \quad k = 1, \dots, n-1,$$

откуда получим:

$$\begin{aligned} c_k h_k - \frac{d_k}{2} h_k^2 &= b_k - b_{k-1}, \quad k = 2, \dots, n \\ d_k h_k &= c_k - c_{k-1}, \quad k = 2, \dots, n. \end{aligned}$$

В результате имеем систему $(3n - 2)$ линейных уравнений относительно $3n$ неизвестных $b_k, c_k, d_k, k = 1, \dots, n$. Два недостающих условия берутся из краевых условий для $S(t)$:

$$f''(a) = f''(b) = 0;$$

тогда

$$S''(a) = S''(b) = 0,$$

или

$$S''(x_0) = S''(x_n) = 0,$$

в таком случае:

$$c_k = d_k h_k, \quad c_{k+1} = 0, \quad c_n = 0.$$

Последнее равенство совпадает с полученным выше равенством

$$d_k h_k = c_k - c_{k-1} \quad \text{при} \quad c_0 = 0.$$

Отсюда получаем систему линейных алгебраических уравнений для определения коэффициентов сплайна $S(t)$:

$$\begin{cases} c_0 = c_n = 0, \\ h_k d_k = c_k - c_{k-1}, \quad k = 1, \dots, n, \\ h_k c_k - \frac{1}{2} h_k^2 d_k = b_k - b_{k-1}, \quad k = 2, \dots, n, \\ h_k b_k - \frac{1}{2} h_k^2 c_k + \frac{1}{6} h_k^2 d_k = f_k - f_{k-1}, \quad k = 1, \dots, n. \end{cases}$$

Далее, исключая из полученной системы a_k, b_k , после несложных алгебраических преобразований переходим к системе для определения c_k :

$$h_k c_{k-1} + 2(h_k + h_{k-1}) c_k + h_{k-1} c_{k+1} = 6 \left(\frac{f_{k+1}}{h_{k+1}} - \frac{f_k - f_{k-1}}{h_k} \right), \\ c_0 = c_n = 0, \quad k = 1, \dots, (n-1).$$

Система имеет матрицу трехдиагональной структуры, обладающую свойством диагонального преобладания; следовательно, ее решение существует и единственно. Алгоритм ее численного решения — прогонка — обладает хорошими свойствами устойчивости. Коэффициенты b_k, d_k определяются после решения системы:

$$d_k = h_k^{-1} (c_k - c_{k-1}), \quad b_k = \frac{1}{2} h_k c_k - \frac{1}{6} h_k^2 d_k + \frac{f_k - f_{k-1}}{h_k}, \\ k = 1, \dots, n.$$

Итак, доказана теорема о существовании и единственности интерполяционного кубического сплайна $S''(a) = S''(b) = 0$.

Аналогично доказывается эта же теорема и для других видов краевых условий.

Теорему существования и единственности можно доказать исходя из линейности второй производной на элементарном отрезке $[x_k, x_{k+1}]$, $k = 0, \dots, n-1$:

$$S''(t) = h_k^{-1} [m_k (x_{k+1} - x_k) + m_{k+1} (x - x_k)], \quad (6.27)$$

где m_k — значение второй производной в точке x_k , поскольку $S(x)$ — кубический полином.

Отсюда получаем:

$$S(x) = (6h_k)^{-1} \left[m_k (x_{k+1} - x_k)^3 + m_{k+1} (x - x_k)^3 \right] + \\ + \alpha_k (x_{k+1} - x) + \beta_{k+1} (x - x_k). \quad (6.28)$$

Из условий интерполяции:

$$S(x_k) = f_k; \quad S(x_{k+1}) = f_{k+1},$$

непрерывности первых производных слева и справа от узла x_k :

$$S'_k(x_k + 0) = S'_k(x_k - 0)$$

и краевых условий для свободного сплайна

$$m_0 = m_n = 0,$$

получим систему уравнений с матрицей трехдиагональной структуры, алгоритм решения которой — прогонка:

$$\begin{aligned} \frac{1}{2}m_k h_{k-1} - \frac{1}{2}m_{k-1} h_{k-1} + \frac{f_k - f_{k-1}}{h_{k-1}} - \frac{1}{6}h_{k-1}(m_k - m_{k-1}) = \\ = \frac{1}{2}m_{k+1} h_k - \frac{1}{2}m_k h_k + \frac{f_{k+1} - f_k}{h_k} - \frac{1}{6}h_k(m_{k+1} - m_k), \end{aligned}$$

или, в матричном виде:

$$\mathbf{A}\mathbf{M} = \mathbf{F}, \quad (6.29)$$

где

$$\mathbf{M} = \begin{Bmatrix} m_1 \\ \vdots \\ m_{n-1} \end{Bmatrix}, \quad \mathbf{F} = \begin{Bmatrix} \frac{f_2 - f_1}{h_1} - \frac{f_1 - f_0}{h_0} \\ \dots \\ \frac{f_n - f_{n-1}}{h_n} - \frac{f_{n-1} - f_{n-2}}{h_{n-1}} \end{Bmatrix}$$

— векторы-столбцы искоемых функций и правых частей соответственно, \mathbf{A} — квадратная матрица вида

$$\mathbf{A} = \begin{pmatrix} \frac{h_1 + h_2}{3} & \frac{h_2}{6} & & & \\ \frac{h_2}{6} & \frac{h_2 + h_3}{3} & & & \\ & \frac{h_3 + h_4}{3} & \frac{h_4}{6} & & \\ & & & \ddots & \\ & & & \frac{h_{n-1}}{6} & \frac{h_n + h_{n-1}}{3} \end{pmatrix}.$$

Третье доказательство проводится на основе кубического полинома Эрмита, который для отрезка $[x_k, x_{k+1}]$ имеет вид

$$S(z) = f_k(1-z)^2(1+2z) + f_{k+1}z^2(3-2z) + m_k h_k z(1-z^2) - m_{k+1} h_k z^2(1-z); \quad (6.30)$$

здесь

$$h_k = x_{k+1} - x_k, \quad z = h_k^{-1}(x - x_k), \quad m_k = S'(x_k).$$

Тогда:

$$S''(x) = h_k^{-2} [(f_{k+1} - f_k)(6 - 12z)] m_k \frac{6z - 4}{h_k} + m_{k+1} \frac{6z - 4}{h_k},$$

$$S''(x_k + 0) = 6h_k^{-2} (f_{k+1} - f_k) - h_k^{-1} \cdot 4m_k - h_k^{-1} \cdot 2m_{k+1},$$

$$S''(x_k - 0) = -6h_{k-1}^{-2} (f_k - f_{k-1}) + h_{k-1}^{-1} \cdot 2m_{k-1} + h_{k-1}^{-1} \cdot 4m_k.$$

Из условия непрерывности второй производной получим (с крайевыми условиями первого типа — заданы первые производные при $x_0 = a$, $x_n = b$) получаем

$$\begin{cases} m_0 = f'_0, \\ r_k m_{k-1} + 2m_k + s_k m_{k+1} = c_k, \\ m_n = f'_n, \end{cases} \quad (6.31)$$

где

$$s_k = \frac{h_{k-1}}{h_{k-1} + h_k}, \quad r_k = 1 - s_k.$$

В случае крайевых условий второго типа (заданы вторые производные при $x_0 = a$, $x_n = b$) система будет иметь следующий вид:

$$\begin{cases} 2m_0 + m_1 = 3h_0^{-1} (f_1 - f_0) + \frac{1}{2} h_0 f''_0, \\ r_k m_{k-1} + 2m_k + s_k m_{k+1} = c_n, \\ m_{n-1} + 2m_n = 3h_{n-1}^{-1} (f_n - f_{n-1}) + \frac{1}{2} f''_0 h_n \end{cases} \quad (6.32)$$

Обе эти линейные алгебраические системы также имеют матрицу трехдиагональной структуры и решаются методом прогонки.

Теорема 6.14 (о сходимости кубического сплайна и его двух первых производных). Для интерполируемой функции $f(x) \in C^4[a, b]$ и интерполирующего ее кубического сплайна $S(x)$ на системе узлов $\{x_k\}_{k=0}^n$, $x_0 = a$, $x_n = b$, имеют место неравенства:

$$\begin{aligned} \|f(x) - S(x)\|_{[a,b]} &\leq M_4 h^4, \\ \|f'(x) - S'(x)\|_{[a,b]} &\leq M_4 h^3, \\ \|f''(x) - S''(x)\|_{[a,b]} &\leq M_4 h^2. \end{aligned}$$

Здесь:

$$\begin{aligned} M_4 &= \max_{[a,b]} |f^{(4)}(x)|, \\ h &= \max_k (x_{k+1} - x_k). \end{aligned}$$

Теорема 6.15 (экстремальное свойство кубического сплайна). Пусть сплайн $S(x)$ интерполирует функцию $f(x)$ на системе узлов интерполяции $\{x_k\}_{k=0}^n$, $x_0 = a$, $x_n = b$.

Тогда свободный сплайн $S(t)$ ($S''(a) = S''(b) = 0$) доставляет минимум функционалу

$$\int_a^b [F''(x)]^2 dx$$

среди всех функций $F(x) \in C_2^2[a, b]$, имеющих интегрируемые с квадратом вторые производные, сходящихся на $[a, b]$ и интерполирующих $f(x)$ на этом отрезке.

6.12. В-сплайны

В-сплайны (базисные сплайны) — это не глобальные сплайны, как $S(x)$, а сплайны на конечных носителях, использующиеся как интерполяционные сплайны в машинной графике, компьютерном дизайне, так и при построении численных методов.

Определение 6.4. В-сплайном степени $(n - 1)$ дефекта 1 относительно системы узлов $\{x_k\}_{k=0}^{n=N}$ называется функция

$$B_{N-1,n}(x) = B_{N-1}(x_n, x_{n+1}, \dots, x_{n+N}, x) = \sum_{i=n}^{i=n+N} \frac{(x_i - x)_m^{N-1}}{\prod_{\substack{i=n \\ i \neq j}}^{i=n+N} (x_i - x_j)}, \quad (6.33)$$

где

$$(x_i - x)_m^{N-1} = \begin{cases} (x_i - x)^{N-1}, & x \leq x_i, \\ 0, & x > x_i. \end{cases}$$

Рассмотрим случай $N=2$ на равномерной сетке: $x_{n+i} = x_n + ih$, h — шаг интерполяции.

В этом случае очевидные преобразования приводят к выражению

$$\begin{aligned} B_{1,n} &= B_1(x_n, x_{n+1}, x_{n+2}, x) = 2 \left(\frac{(x_n - x)_m}{(x_n - x_{n+1})(x_n - x_{n+2})} + \right. \\ &+ \frac{(x_{n+1} - x)_m}{(x_{n+1} - x_n)(x_{n+1} - x_{n+2})} + \left. \frac{(x_{n+2} - x)_m}{(x_{n+2} - x_n)(x_{n+2} - x_{n+1})} \right) = \\ &= h^{-2} [(x_n - x)_m - 2(x_n - x)_m + (x_{n+2} - x)_m]; \end{aligned}$$

окончательно получим так называемые базисные «линейные» функции — «крышки», имеющие следующий вид:

$$B_1(x) = \begin{cases} h^{-2}(x_n - x - 2x_{n+1} + 2x + x_{n+2} - x) = 0, & x \leq x_n; \\ h^{-2}(x_n - 2x_{n+1} + 2x + x_{n+2} - x) = h^{-1} + \frac{x - x_{n+1}}{x^2}, & x_n \leq x \leq x_{n+1}; \\ h^{-2}(0 - 0 + x_{n+2} - x) = h^{-1} - \frac{x - x_{N+1}}{h^2}, & x_{n+1} \leq x \leq x_{n+2}; \\ 0, & x \leq x_{n+2}. \end{cases} \quad (6.34)$$

Для случая $N=3$ получим

$$B_2(x) = \begin{cases} z^2, & x = \frac{x - x_{k-2}}{x_{k-1} - x_{k-2}}, & x \in [x_{k-1}, x_{k-2}]; \\ 1 + 2z^2 - z^2, & z = \frac{x - x_{k-1}}{x_k - x_{k-1}}, & x \in [x_{k-1}, x_k]; \\ 2 - z^2, & z = \frac{x - x_k}{x_{k+1} - x_k}, & x \in [x_k, x_{k+1}]; \\ (1 - z)^2, & z = \frac{x - x_k}{x_{k+2} - x_{k+1}}, & x \in [x_{k+1}, x_{k+2}]. \end{cases} \quad (6.35)$$

При $x < x_{k-2}$, $x > x_{k+2}$: $S(x) \equiv 0$.

При $N = 4$ B -сплайн принимает вид

$$B_3(x) = \begin{cases} 0, & x \leq x_n \\ \frac{1}{6\tau^4}(x - x_n)^3, & x_n \leq x \leq x_{n+1}; \\ \frac{1}{6h} + \frac{1}{2h^2}(x - x_{n+1}) + \frac{1}{2h^3}(x - x_{n+1})^2 - \\ - \frac{1}{2h^4}(x - x_{n+1})^3, & x_{n+1} \leq x \leq x_{n+2}; \\ \frac{1}{6h} + \frac{1}{2h^2}(x_{n+3} - x) + \frac{1}{2h^3}(x_{n+3} - x)^2 - \\ - \frac{1}{2h^4}(x_{n+3} - x)^3, & x_{n+2} \leq x \leq x_{n+3}; \\ \frac{1}{6h}(x_{n+3} - x)^3, & x_{n+3} \leq x \leq x_{n+4}; \\ 0, & x \geq x_{n+4}. \end{cases} \quad (6.36)$$

Интерполируемая функция при этом приобретает вид

$$F(x) = \sum_{i=0}^n u_i B_{Ni}(x).$$

Список литературы

1. *Ильин В.П.* Численный анализ. Ч. 1. Новосибирск: ИВМиМГ СО РАН, 2004. 334 с.
2. *Самарский А.А., Гулин А.В.* Численные методы. М.: Наука. 1989. 430 с.
3. *Каханер Д., Моулер К., Нэш С.* Численные методы и программное обеспечение. М.: Мир, 2001. 575 с.
4. *Завьялов Ю.С., Квасов Б.И., Мирошниченко В.Л.* Методы сплайн-функции. М.: Наука, 1980. 350 с.



ЧИСЛЕННЫЕ МЕТОДЫ ИНТЕГРИРОВАНИЯ ФУНКЦИЙ



7.1. Интерполяционные квадратурные формулы

Приближенное численное интегрирование функции одной переменной на $f(t)$ на отрезке $[a, b]$ обычно сводится к вычислению суммы вида

$$I^{(n)} = \sum_{k=0}^n c_k f(x_k). \quad (7.1)$$

При этом полагают, что $I^{(n)}$ — это приближенное значение интеграла

$$I = \int_a^b f(x) dx, \quad (7.2)$$

т. е.

$$I = \int_a^b f(x) dx \approx I^{(n)} = \sum_{k=0}^n c_k f(x_k), \quad f(x) \in L_1[a, b], \quad (7.3)$$

где (7.3) называется *квадратурной формулой*, $c_k \in \mathbb{R}$ — *коэффициентами* квадратурной формулы, $x_k \in [a, b]$ — *узлами* квадратурной формулы, а разность

$$\xi_n = \left| I - I^{(n)} \right| = \left| \int_a^b f(x) dx - \sum_{k=0}^n c_k f(x_k) \right| \quad (7.4)$$

— *погрешностью* квадратурной формулы. Погрешность зависит как от c_k , так и от расположения *узлов*, т. е. от сетки (совокупности, или множества точек):

$$\omega_n = \left\{ x_k = x_0 + kh; \quad k = 0, 1, \dots, n, \quad h = \frac{b-a}{n}, \quad x_0 = a, \quad x_n = b \right\}.$$

Если на отрезке $[a, b]$ ввести совокупность узлов, то интеграл может быть представлен в виде суммы интегралов по элементарным отрезкам $[x_{k-1}, x_k]$:

$$I = \int_a^b f(x) dx \approx I^{(n)} = \sum_{k=1}^n \int_{x_{k-1}}^{x_k} f(x) dx, \quad (7.5)$$

т. е. для вычисления значения I достаточно вычислить интегралы по элементарным отрезкам:

$$I_k = \int_{x_{k-1}}^x f(x) dx \quad (7.6)$$

и просуммировать эти величины от $k=1$ до $k=n$.

Рассмотрим, например, так называемую *формулу средних*, для которой интеграл на каждом элементарном отрезке $[x_{k-1}, x_k]$ вычисляется в соответствии с формулой

$$I_k = \int_{x_{k-1}}^x f(x) dx \approx h_k \cdot f(x_{k-1/2}), \quad (7.7)$$

где $h_k = x_k - x_{k-1}$, $x_{k-1/2} = (1/2)(x_k + x_{k-1})$.

Суммируя эти неравенства от $k=1$ до $k=n$, будем иметь

$$I = \int_a^b f(x) dx \approx \sum_{k=1}^n f(x_{k-1/2}) \cdot h_k. \quad (7.8)$$

В частности, для равномерного разбиения отрезка $[a, b]$, т. е. $h_k = h = (b - a)/n$, получим

$$I \approx h \sum_{k=1}^n f(x_{k-1/2}).$$

Погрешность этой формулы на элементарном отрезке определяется величиной

$$\xi_n^k = \int_{x_{k-1}}^{x_k} |f(x) dx - h \cdot f_{k-1/2}|, \quad (7.9)$$

которую можно оценить, например, с помощью разложения на формуле Тейлора (при этом полагаем $f(x) = f_k$):

$$f(x) = f_{k-1/2} + (x - x_{k-1/2}) f'_{k-1/2} + \frac{(x - x_{k-1/2})^2}{2} f''(\eta_k),$$

где $\eta_k \in [x_{k-1}, x_k]$.

В таком случае приходим к формуле

$$\xi_n^k = \int_{x_{k-1}}^{x_k} \frac{(x - x_{k-1/2})^2}{2} f''(\eta) dx,$$

или

$$|\xi_n^k| \leq M_{2k} \int_{x_{k-1}}^{x_k} \frac{(x - x_{k-1})^2}{2} dx = \frac{h^3}{24} M_{2k}, \quad (7.10)$$

$$M_{2k} = \max_{[x_{k-1}, x_k]} |f''(\xi_n^k)|.$$

Если просуммировать погрешности ξ_n^k по всему отрезку $[a, b]$, то получим оценку погрешности на всем отрезке:

$$\xi_n = \sum_{k=1}^n |\xi_k| = \left| \sum_{k=1}^n \int_{x_{k-1}}^{x_k} \frac{(x - x_{k-1/2})^2}{2} f''(\xi_k) dx \right| \leq$$

$$\leq \frac{M_2}{24} N h \cdot h^2 = \frac{(b-a)}{24} M_2 h^2, \quad (7.11)$$

где $M_2 = \max_{[a,b]} |f''(\eta)|$, т.е. на элементарном отрезке $[x_{k-1}, x_k]$ погрешность формулы средних есть

$$\xi_n^k \leq \frac{h^3}{24} M_{2k}, \quad \text{или} \quad \xi_n^k = O(h^3),$$

а на всем отрезке $[a, b]$:

$$\xi_n \leq \frac{(b-a)}{24} M_2 h^2, \quad \text{или} \quad \xi_n = O(h^2).$$

Таким образом, квадратурная формула средних для приближенного вычисления значения интеграла

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

имеет второй порядок точности.

Формула трапеций, приближающая интеграл I на элементарном отрезке, имеет вид

$$I = \int_{x_{k-1}}^{x_k} f(x) dx \approx h_k \cdot \frac{f_{k-1} + f_k}{2}. \quad (7.12)$$

При этом подынтегральная функция приближается интерполяционным полиномом первой степени:

$$f(x) \approx L_1(x) = f_k \frac{(x - x_{k-1})}{h_k} - f_{k-1} \frac{(x - x_k)}{h_k}.$$

Остаточный член этого интерполянта вычисляется следующим образом:

$$\xi_{1k} = f(x) - L_1(x) = \frac{f''(\xi)}{2} (x - x_{k-1})(x - x_k),$$

откуда

$$\begin{aligned} \xi^k &= \left| \int_{x_{k-1}}^{x_k} f(x) dx - \frac{h_k}{2} (f_k + f_{k-1}) \right| = \left| \int_{x_{k-1}}^{x_k} [f(x) - L_1(x)] dx \right| = \\ &= \left| \int_{x_{k-1}}^{x_k} \frac{(x - x_{k-1})(x - x_k)}{2} f''(\xi) dx \right|, \end{aligned}$$

или

$$|\xi^k| \leq \frac{M_{2k}}{12} h^3.$$

Аналогично предыдущему случаю (формула средних для отрезка $[a, b]$), интеграл по всему отрезку $[a, b]$ будет вычисляться по формуле

$$I = \int_a^b f(x) dx \approx \sum_{k=1}^n h_k \frac{f_k + f_{k-1}}{2} = \frac{1}{2} \sum_{k=1}^n h_k (f_k + f_{k-1}).$$

В случае равномерного разбиения рассматриваемого отрезка, т. е. $h_k = h = (b - a)/n$, получим

$$I = \int_a^b f(x) dx \approx h \left(\frac{1}{2} (f_0 + f_N) + \sum_{k=1}^{N-1} f_k \right). \quad (7.13)$$

Погрешность формулы трапеций на отрезке $[a, b]$ для равномерной сетки оценивается как сумма погрешностей ξ_n^k по элементарным отрезкам $[x_{k-1}, x_k]$:

$$|\xi_n| \leq \frac{(b-a)}{12} M_2 h^2, \quad |\xi_n| = O(h^2),$$

т. е. эта формула имеет такой же порядок точности (второй), как и формула средних, но погрешность при этом вдвое больше последней.

Эту формулу можно уточнить, используя многочлены Эрмита. Если на концах отрезка $[a, b]$ заданы не только значения

функции, но и значения ее производных: $f(a)$ и $f(b)$, $f'(a)$ и $f'(b)$, то мы получаем формулу Эйлера–Маклорена:

$$I = \int_a^b f(x) dx \approx \frac{b-a}{2} [f(a) + f(b)] + \frac{(b-a)^2}{12} [f'(a) - f'(b)].$$

При равномерном разбиении отрезка $[a, b]$ получим

$$I \approx h \left(\frac{1}{2} f_0 + \frac{1}{2} f_n + f_1 + f_2 + \dots + f_{n-1} \right) + \frac{h^2}{12} [f'(a) - f'(b)];$$

причем остаточный член этой формулы имеет вид

$$\xi_n = \frac{h^4}{720} \left| \int_a^b f^{(IV)}(x) dx \right|,$$

т. е. незначительное усложнение формулы трапеции увеличивает порядок точности на два порядка.

Проведем теперь аппроксимацию подынтегральной функции $f(x)$ многочленом второй степени, проходящим через узлы нашей сетки $\{x_i, f_i\}$, $i = k-1, k-1/2, k$:

$$\begin{aligned} f(x) &\approx L_2(x), \quad x \in [x_{k-1}, x_k], \\ L_2(x) &= f_{k-1} \cdot \frac{(x-x_k)(x-x_{k-1/2})}{(x_{k-1}-x_k)(x_{k-1}-x_{k-1/2})} + \\ &+ f_{k-1/2} \frac{(x-x_{k-1})(x-x_k)}{(x_{k-1/2}-x_{k-1})(x_{k-1/2}-x_k)} + \\ &+ f_k \frac{(x-x_{k-1/2})(x-x_{k-1})}{(x_k-x_{k-1/2})(x_k-x_{k-1})} \end{aligned}$$

— интерполяционный полином второй степени в форме Лагранжа.

После интегрирования по элементарному отрезку получим

$$I_k = \int_{x_{k-1}}^{x_k} f(x) dx \approx \int_{x_{k-1}}^{x_k} L_2(x) dx = \frac{h_k}{6} (f_{k-1} + 4f_{k-1/2} + f_k).$$

На всем отрезке $[a, b]$ приближенная формула интегрирования (формула Симпсона) имеет вид

$$I = \int_a^b f(x) dx \approx \frac{1}{6} \sum_{k=1}^n h_k (f_{k-1} + 4f_{k-1/2} + f_k);$$

в случае равномерного разбиения отрезка ($h_k = h = (b - a) / N$) получим

$$I \approx \frac{h}{6} [(f_0 + f_n) + 2(f_1 + f_2 + \dots + f_{n-1}) + 4(f_{1/2} + f_{3/2} + \dots + f_{n-1/2})]. \quad (7.14)$$

Формулу Симпсона можно записать и без дробных индексов:

$$I \approx \frac{h}{6} [(f_0 + f_n) + 2(f_2 + f_4 + \dots + f_{2n-2}) + 4(f_1 + f_3 + \dots + f_{2n-1})]. \quad (7.15)$$

Погрешность формулы (7.15) на элементарном отрезке оценивается так:

$$|\xi^k| \leq \frac{h^5}{290} M_{4k}, \quad M_{4k} = \max_{[x_{k-1}, x_k]} |f^{(IV)}(x)|,$$

а на всем отрезке $[a, b]$ — по формуле

$$|\xi_n| \leq \frac{M_4}{180} (b - a) h^4, \quad M_4 = \max_{[a, b]} |f^{(IV)}(x)|$$

Формулы численного интегрирования (7.8), (7.13), (7.15), основанные на аппроксимации подынтегральной функции интерполяционным многочленом на отрезке $[a, b]$, называются *квадратурными формулами интерполяционного типа*. Важным свойством этих формул является положительность коэффициентов квадратурных формул ($c_k > 0$). Такие формулы называются *правильными*. Квадратурные формулы интерполяционного типа для равномерной сетки называются *формулами Ньютона–Котеса*. Заметим, что если в этих формулах $x_0 = a$, $x_n = b$, то их называют *формулами замкнутого типа*; если же хотя бы один из узлов (x_0 или x_n) не совпадает с соответствующей граничной точкой (a или b) отрезка $[a, b]$, то такие формулы называются *формулами открытого типа*.

Показывается, что при $n \geq 7$ в (7.1) встречаются $c_k < 0$, причем

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n |c_k^{(n)}| \rightarrow \infty,$$

т.е. квадратурные формулы при c_k , имеющие разные знаки, оказываются неустойчивыми, что связано с неустойчивостью интерполяционного процесса при больших n , соответствующих росту постоянных Лебега. По этой причине квадратурные формулы интерполяционного типа для больших n , как правило, не применяются. Из них наиболее часто используются в практике инженерных расчетов формулы Симпсона.

7.2. Квадратурные формулы Чебышёва, Гаусса, Гаусса–Кристоффеля

Однако если выбирать определенным образом узлы сетки, то можно добиться не только того, чтобы все коэффициенты квадратурных формул были положительны, но и повышения точности квадратурных формул. Такие формулы были предложены Гауссом и Чебышёвым.

Отметим еще одну важную особенность квадратурных формул интерполяционного типа: эти формулы точны для полиномов степени n , построенных по $(n + 1)$ узлам $\{x_k\}_{k=0}^n$, т.е. если $f(x)$ — полином степени n , c_k — коэффициенты квадратурной формулы, то мы получаем точную формулу численного интегрирования

$$I = \int_a^b f(x) dx = \sum_{k=0}^n c_k f(x_k). \quad (7.16)$$

Чебышёв предложил в квадратурной формуле

$$I = \int_{-1}^1 f(x) dx \approx \sum_{k=1}^n c_k f(x_k) \quad (7.17)$$

выбирать узлы x_k так, чтобы выполнялись следующие условия:

- коэффициенты квадратурной формулы равны между собой, т.е.

$$c_1 = c_2 = \dots = c_n = c;$$

- квадратурная формула (7.17) точна для всех многочленов до степени n включительно.

Пусть $f(x) = 1$ (многочлен нулевой степени). В таком случае если $c_k = c$ ($k = 1, \dots, n$), то

$$2 = \sum_{k=1}^n c_k, \quad \text{или} \quad c = \frac{2}{n},$$

т.е. квадратурная формула Чебышёва имеет вид

$$I = \int_{-1}^1 f(x) dx \approx \frac{2}{n} \sum_{k=1}^n f(x_k). \quad (7.18)$$

Чтобы определить узлы x_k , заметим, что эта формула должна быть точной для следующих функций:

$$f(x) = x, \quad f(x) = x^2, \quad \dots, \quad f(x) = x^n.$$



при этом из (7.20) получим

$$\begin{cases} y_1 = 0, \\ y_2 = \frac{1}{2} \left[(x_1 + x_2 + x_3)^2 - (x_1^2 + x_2^2 + x_3^2) \right] = \frac{1}{2} (0 - 1) = -\frac{1}{2}, \\ y_3 = \frac{1}{6} \left[(x_1 + x_2 + x_3)^2 - 3(x_1 + x_2 + x_3)(x_1^2 + x_2^2 + x_3^2) + 2(x_1^3 + x_2^3 + x_3^3) \right] = \frac{1}{6} (0 - 0 + 0) = 0. \end{cases}$$

В таком случае получаем систему нелинейных уравнений вида

$$\begin{cases} x_1 + x_2 + x_3 = 0, \\ x_1x_2 + x_1x_3 + x_2x_3 = -\frac{1}{2}, \\ x_1x_2x_3 = 0. \end{cases}$$

Эта система симметрична относительно переменных x_1, x_2, x_3 . Положим, например (см. последнее уравнение системы): $x_3 = 0$. Тогда $x_1 = -x_2 = 1/\sqrt{2}$.

Формула Чебышёва приобретает следующий вид:

$$\int_{-1}^1 f(x) dx = \frac{2}{3} \left[f\left(-\frac{1}{\sqrt{2}}\right) + f(0) + f\left(\frac{1}{\sqrt{2}}\right) \right].$$

Гаусс поставил следующую задачу: найти коэффициенты квадратурной формулы c_k ($k = 1, \dots, n$) и координаты узлов x_i ($k = 1, \dots, n$) так, чтобы квадратурная формула

$$\int_{-1}^1 f(x) dx = \sum_{k=1}^n c_k f(x_k)$$

была точна для полиномов наивысшей возможной степени M .

Отметим, что недостатком квадратурных формул является их относительно невысокий порядок точности, что обусловлено неустойчивостью интерполяционных формул, использующихся для аппроксимации подынтегральной функции. Таким образом, мы имеем ситуацию, при которой подынтегральная функция может иметь высокий порядок гладкости, при этом в выражении для погрешности квадратурной формулы будут присутствовать производные невысокого порядка, т. е. такое свойство, как гладкость подынтегральной функции (если, конечно, это свойство присутствует), недостаточно используется в квадратурных формулах интерполяционного типа, что и было замечено Гауссом. Как уже отмечалось, существуют подынтегральные функции

(полиномы), для которых квадратурные формулы являются точными.

Положим, что подынтегральная функция является полиномом степени n :

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

Аппроксимируем этот интеграл с помощью квадратурной формулы вида

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

или

$$\int_a^b P_n(x) dx = \sum_{k=1}^n c_k P_n(x_k).$$

Представим $P_n(x)$ в виде интерполяционного полинома, записанного в форме Лагранжа:

$$\int_a^b \sum_{k=1}^n P_n(x_k) \prod_{\substack{i=1 \\ i \neq k}}^n \frac{x - x_i}{x_k - x_i} dx = \sum_{k=1}^n P_n(x_k) \cdot c_k.$$

Отсюда получим выражение для коэффициентов c_k :

$$c_k = \int_a^b \prod_{i=1, i \neq k}^n \frac{x - x_i}{x_k - x_i} dx,$$

которые являются интегралами от базисных функций интерполяционного полинома Лагранжа.

Теперь поставим вопрос: можно ли построить квадратурную формулу, точную для полинома $P_m(x)$ степени $m > n$?

Положим:

$$f(x) = P_m(x) = \sum_{k=0}^m a_k x^k,$$

$$I = \int_a^b f(x) dx = \sum_{k=1}^n c_k f(x_k) = \sum_{k=1}^n c_k P_m(x_k).$$

a узлами $\{x\}_k^n$, $k = 1, \dots, n$, в квадратурной формуле являются корни полинома Лежандра $q_k(x)$. В этом случае квадратурная формула

$$\int_{-1}^1 f(x) dx = \sum_{k=1}^n c_k f(x_k)$$

является точной для полиномов степени $(2n - 1)$.

Полиномы Лежандра определяются по рекуррентной формуле

$$(k+1)q_{k+1}(x) = (2k+1)xq_k(x) - kq_{k-1}(x),$$

$$q_0(x) = 1, \quad q_1(x) = x, \quad q_2(x) = \frac{1}{3}(3x^2 - 1), \dots$$

и образуют ортогональную систему полиномов на отрезке $[-1, 1]$:

$$\int_{-1}^1 q_i(x) q_j(x) dx = 0 \text{ при } i \neq j, \quad \int_{-1}^1 q_i(x) q_j(x) dx \neq 0 \text{ при } i = j.$$

Погрешность формулы Гаусса на $[-1, 1]$ имеет вид

$$\xi_n(t) = 2^{2n+1} \cdot \alpha_n \cdot f^{2n}(\xi), \quad \xi \in [-1, 1].$$

При $\xi \in [a, b]$ получаем:

$$\xi_n(t) = (b-a) \cdot \alpha_n \cdot f^{2n}(\xi),$$

$$\alpha_n = \frac{(n!)^4}{[(2n)!]^3 \cdot (2n+1)}.$$

Квадратурные формулы Гаусса являются высокоточными, например, для $n = 2; 3; 4; 5$ получим соответственно:

$$\xi_2 = \frac{2^5 \cdot 2^4}{5 \cdot [4 \cdot 3 \cdot 2]^3} f^{(IV)}(\xi) = \frac{2^9}{5 \cdot [2 \cdot 3]^3} f^{(IV)}(\xi) = \frac{f^{(IV)}(\xi)}{135};$$

$$\xi_3 = \frac{2}{15750} f^{(VI)}(\xi);$$

$$\xi_4 = \frac{2}{3472875} f^{(VIII)}(\xi);$$

$$\xi_5 = \frac{13}{1237732650} f^{(X)}(\xi).$$

Важное свойство этих формул заключается в том, что они являются правильными, т. е. $c_k > 0$ ($k = 0, \dots, n$), что обеспечивает устойчивость вычислительного алгоритма.

Если положить $n = 3$, то корнями полинома Лежандра

$$P_3(x) = \frac{1}{2}(5x^3 - 3x)$$

будут числа:

$$x_1 = -\sqrt{\frac{3}{5}}, \quad x_2 = 0, \quad x_3 = \sqrt{\frac{3}{5}},$$

а для весовых коэффициентов квадратурной формулы получим систему трех нелинейных уравнений:

$$\begin{cases} c_1 + c_2 + c_3 = 2, \\ -\sqrt{\frac{3}{5}} c_1 + \sqrt{\frac{3}{5}} c_3 = 0, \\ \frac{3}{5} c_1 + \frac{3}{5} c_2 = \frac{2}{3}, \end{cases}$$

откуда получим:

$$c_1 = c_3 = \frac{5}{9}, \quad c_2 = \frac{8}{9},$$

следовательно,

$$I = \int_{-1}^1 f(x) dx = \frac{1}{9} \left[5f\left(-\sqrt{\frac{3}{5}}\right) + 8f(0) + 5f\left(\sqrt{\frac{3}{5}}\right) \right].$$

В случае, когда интегрирование проводится по отрезку $[a, b]$, применяется замена переменной:

$$z = \frac{a+b}{2} + \frac{b-a}{2}x,$$

откуда получим

$$\int_a^b f(z) dz = \frac{b-a}{2} \sum_{k=1}^n c_k f(z_k),$$

где

$$z_k = \frac{b+a}{2} + \frac{b-a}{2} \cdot x_k, \quad i = 1, 2, \dots, n,$$

— нули многочлена Лежандра: $P_n(z_k) = 0$.

Квадратурные формулы Гаусса–Кристоффеля, которые иногда называют формулами *наивысшего алгебраического порядка*, имеют вид

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

Здесь введена весовая функция $p(x)$, которая непрерывна и положительна на отрезке $[a, b]$, а также должна быть интегрируема на $[a, b]$, т. е. должен существовать интеграл

$$\int_a^b p(x) dx.$$

При $p(x) = 1$ эта формула является формулой Гаусса, так как функция $p(x)$ удовлетворяет этим требованиям.

Например, положив $a = -1$, $b = 1$, $p(x) = (1 - x^2)^{-1}$ и, взяв в качестве узлов квадратурной формулы $\{x_k\}_0^n$ корни полинома Чебышёва, получим квадратурную формулу Эрмита:

$$\int_a^b \frac{f(x)}{\sqrt{1-x^2}} dx \approx \sum_{k=1}^n c_k f(x_k).$$

Веса этой квадратурной формулы[®] вычисляются следующим образом:

$$c_k = \int_{-1}^1 \frac{\overline{T}_n(x) dx}{\sqrt{1-x^2} T'_n(x_k) (x - x_k)},$$

где $\overline{T}_n(x)$ — нормированные многочлены Чебышёва.

После вычисления этого интеграла получим

$$c_k = c = \frac{\pi}{2}, \quad k = 1, 2, \dots, n.$$

Окончательный вид формулы Эрмита:

$$\int_{-1}^1 \frac{f(x)}{\sqrt{1-x^2}} dx \approx \frac{\pi}{n} \sum_{k=1}^n f(x_k),$$

где x_k — корни многочлена Чебышёва.

К настоящему времени рассчитано большое количество таблиц для формул Гаусса при $a = -1$, $b = 1$, $p(x) = 1$, а также формул с весовыми функциями следующего вида (интегралы Якоби):

$$p(x) = (1+x)^\alpha (1-x)^\beta; \quad \alpha, \beta > -1,$$

$$a_1 = -1, b = 1$$

и

$$p(x) = x^\alpha e^{-x}, \quad \alpha > -1, \quad x \in [0, \infty)$$

(формула Чебышёва–Лагерра).

В этих случаях имеем:

$$I = \int_{-1}^1 (1-x)^\alpha (1+x)^\beta f(x) dx \approx \sum_{k=1}^n c_k f(x_k),$$

$p(x) = e^{-x}$, $p(x) = e^{-x^2}$ квадратурные формулы называются формулами соответственно Лагерра и Эрмита:

$$I = \int_0^\infty e^{-x} f(x) dx \approx \sum_{k=1}^n c_k f(x_k),$$

где x_k — корни многочлена Лагерра.

Многочлены Лагерра $L_n(x)$, $n \in [0, \infty)$, вычисляются по рекуррентной формуле

$$L_{n+1}(x) - 2(n+1-x)L_n(x) + n^2 L_{n-1}(x) = 0;$$

$$L_0 = 1, \quad L_1 = -x + 1.$$

Для них выполняются условия ортогональности:

$$\int_0^\infty e^{-x} \cdot L_k(x) \cdot L_j(x) dx = \begin{cases} 0, & k \neq j, \\ k!, & k = j. \end{cases}$$

Весовые коэффициенты вычисляются по формулам:

$$c_i = \int_0^\infty \frac{L_n(x) \cdot e^{-x} dx}{(x-x_i) \cdot L'_n(x_i)} = \left[\frac{(n-1)!}{n L_{n-1}(x_i)} \right]^2 x_i,$$

$$I = \int_0^\infty e^{-x^2} f(x) dx \approx \sum_{k=1}^n c_k f(x_k),$$

где x_k — корни многочлена Эрмита.

Многочлены Эрмита $H_n(x)$, $x \in (-\infty, \infty)$, вычисляются по рекуррентным формулам

$$H_{n+1}(x) - 2x \cdot H_n(x) + 2n H_{n-1}(x) = 0, \quad H_0 = 1, \quad H_1 = 2x;$$

для них выполняются условия ортогональности:

$$\int_{-\infty}^\infty e^{-x^2} H_k(x) \cdot H_j(x) dx = \begin{cases} 0, & k \neq j, \\ 2^k \cdot k! \sqrt{\pi}, & k = j. \end{cases}$$

Весовые коэффициенты вычисляются по формулам

$$c_i = \int_{-\infty}^{\infty} \frac{H_n(x) e^{-x^2} dx}{(x - x_i) H'_n(x_i)} = \frac{2^{n-1} (n-1) \sqrt{\pi}}{n \cdot H_{n-2}^2(x_i)}.$$

7.3. Вычисления кратных интегралов

Для вычисления кратного интеграла, например, с подынтегральной функцией двух переменных $\{x, y\}$, можно использовать формулу средних:

$$I = \int_a^b \int_c^d f(x, y) dx dy \approx h_x \cdot h_y \cdot f\left(\frac{a+b}{2}, \frac{c+d}{2}\right);$$

$h_x = b - a$, $h_y = d - c$; значение f вычисляется в точке пересечения диагоналей прямоугольника со стороной h_x, h_y .

В случае использования формулы Симпсона представим интеграл в виде

$$I = \int_a^b dx \int_c^d f(x, y) dy$$

и применим эту формулу для вычисления первого (внешнего) интеграла:

$$\begin{aligned} I &\approx \frac{h_x}{6} \left\{ \int_c^d f(a, y) dy + 4 \int_c^d f\left(\frac{a+b}{2}, y\right) dy + \int_c^d f(b, y) dy \right\} = \\ &= \frac{h_x}{6} \left\{ \frac{h_y}{6} \left[f\left(a, \frac{c+d}{2}\right) + 4f\left(a, \frac{c+d}{2}\right) + f(a, d) \right] + \right. \\ &+ \frac{h_y}{6} \left[f\left(\frac{a+b}{2}, c\right) + 4f\left(\frac{a+b}{2}, \frac{c+d}{2}\right) + f\left(\frac{a+b}{2}, d\right) \right] \frac{h_y}{6} + \\ &+ \frac{h_y}{6} \left[f(b, c) + 4f\left(b, \frac{c+d}{2}\right) + f(b, d) \right] + f(b, d) \left. \right\} = \\ &= \frac{h_x h_y}{36} \cdot \left\{ [f(a, c) + f(a, d) + f(b, c) + f(b, d)] + \right. \\ &+ 4 \left[f\left(a, \frac{c+d}{2}\right) + f\left(b, \frac{c+d}{2}\right) + f\left(\frac{c+d}{2}, b\right) + f\left(\frac{a+b}{2}, d\right) \right] + \\ &\quad \left. + 16f\left(\frac{a+b}{2}, \frac{c+d}{2}\right) \right\}. \end{aligned}$$

При разбиении области интегрирования на элементарные прямоугольники:

$$x_k \leq x \leq x_{k+1}, \quad y_k \leq y \leq y_{k+1},$$

квадратурные формулы используются для вычисления интегралов по каждому такому прямоугольнику; далее интеграл по всей области вычисляется как их сумма.

Идея метода Монте-Карло вычисления n -кратных интегралов

$$I = \int \dots \int_{\Omega} f(x_1, \dots, x_n) dx_1 \dots dx_n$$

состоит в следующем.

Сначала генерируется совокупность n_0 случайных точек внутри n -мерного куба $x_1, \dots, x_n \in R^n$, заключающего в себе рассматриваемую область ω , по которой берется n -кратный интеграл. Пусть n_ω точек попали в ω и вычислена сумма

$$\sum_{k=1}^{n_\omega} f(x_k).$$

Вычислим среднее по области ω значение $f(x)$ двумя способами. Возьмем значение

$$\bar{f}(x) = \frac{J_\omega}{V_\omega},$$

где J_ω — искомое значение интеграла по ω , $V_\omega \approx n_\omega/n_0$ — объем ω , и приравняем его к среднему значению функции, вычисленному по другой формуле:

$$\bar{f}(x) \approx \frac{\sum_{k=1}^{n_\omega} f(x_k)}{n_\omega}.$$

В этом случае приближенное значение кратного интеграла будет иметь вид

$$J_\omega \approx \frac{\sum_{k=1}^{n_\omega} f(x_k)}{n_0}.$$

7.4. Вычисления интегралов с особенностями

Рассмотрим способы вычисления несобственного интеграла первого рода

$$I = \int_0^\infty f(x) dx.$$

Первый способ состоит в замене переменной

$$x = \frac{a}{1-t}.$$

При этом

$$I = \int_0^1 \frac{a}{(1-t)^2} \cdot f\left(\frac{a}{1-t}\right) dt.$$

Если подынтегральная функция ограничена, то можно использовать известные квадратурные формулы.

Второй способ состоит в следовании определению несобственного интеграла:

$$I = \int_d^\infty f(x) dx = \lim_{D \rightarrow \infty} \int_d^D f(x) dx.$$

В этом случае вычисляется значение определенного интеграла

$$I_D^1 = \int_d^{D_1} f(x) dx,$$

а затем — значение

$$I_D^2 = \int_d^{D_2} f(x) dx, \quad D_2 > D_1,$$

после чего эти значения сравниваются по модулю, а модуль разности сопоставляется с заданной точностью ε :

$$|I_D^1 - I_D^2| \leq \varepsilon.$$

Если последнее неравенство выполняется, то расчет прекращается; если нет, то вычисляется интеграл при $D = D_3$, который сравнивается по модулю с интегралом при $D = D_2$, и т. д.

Если подынтегральная функция представлена в виде

$$\int_u^\infty p(x) f(x) dx,$$

где $p(x) = e^{-x}$ или $p(x) = e^{-x^2}$, то можно использовать формулы Лагерра или Эрмита соответственно.

Рассмотрим вычисление интеграла второго рода с особенностью в точке $x = c \in [a, b]$:

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

Представим интеграл в виде

$$I = \lim_{\delta \rightarrow 0} \left(\int_a^{c-\delta} f(x) dx + \int_{c+\delta}^b f(x) dx \right),$$

далее вычислим абсолютную величину интеграла I_δ и сравним его значение со значением $I_{\delta/2}$:

$$\Delta = |I_{\delta/2} - I_\delta| \leq \varepsilon.$$

При выполнении последнего неравенства расчет прекращается; если неравенство не выполняется, то уменьшаем δ вдвое и повторяем процедуру расчета:

$$\Delta = |I_{\delta/4} - I_{\delta/2}| \leq \varepsilon \text{ и т. д.}$$

Идея метода выделения особенностей Канторовича состоит в представлении подынтегральной функции $f(x)$ в виде

$$f(x) = F(x) + [f(x) - F(x)],$$

и, соответственно, интеграла —

$$I = \int_a^b f(x) dx = \int_a^b F(x) dx + \int_a^b [f(x) - F(x)] dx.$$

Функция $F(x)$ выбирается таким образом, чтобы она была интегрируемой:

$$F(x) \in L[a, b],$$

а разность

$$|f(x) - F(x)|$$

— ограниченной. Например,

$$I = \int_0^1 \frac{dx}{\sqrt{x(1+x^2)}} = \int_0^1 \frac{dx}{\sqrt{x}} + \int_0^1 \left(\frac{dx}{\sqrt{x(1+x^2)}} - \frac{1}{\sqrt{x}} \right) dx.$$

В этом случае первый интеграл вычисляется аналитически, а второй можно вычислить по квадратурным формулам, так как подынтегральная функция ограничена.

Для вычисления несобственного интеграла также можно использовать разложение в ряд Тейлора: например,

$$\begin{aligned} I = \int_0^1 \frac{\cos x}{\sqrt{x}} dx &\approx \int_0^1 \frac{1 - x^2/2! + x^4/4! + \dots}{x^{1/2}} dx = \\ &= \int_0^1 x^{1/2} dx - \frac{1}{2} \int_0^1 x^{3/2} dx + \int_0^1 \frac{x^{7/2}}{4!} + \dots \end{aligned}$$

Полученные интегралы берутся в квадратурах.

При необходимости вычисления интеграла от быстроосциллирующей функции

$$I = \int_a^b f(t) \sin(\omega t) dx$$

ее можно аппроксимировать интерполяционным полиномом

$$f(t) \approx L_n(t)$$

и далее вычислить полученный интеграл аналитически.

7.5. Апостериорная практическая оценка погрешности квадратурных интерполяционных формул

Пусть $I^h \approx I = \int_a^b f(x) dx$ — приближенное значение интеграла I на отрезке $[a, b]$.

Для погрешности этой формулы справедливо представление вида

$$I - I^h = Ch^k + \xi_k,$$

где $C \neq C(h)$.

Величина Ch^k называется *главным членом погрешности*, а k — *порядком точности* соответствующей квадратурной формулы. Оказывается, что вычислительным путем можно увеличить порядок точности такой формулы и оценить погрешность вычисления интеграла.

Заметим, что приближенная формула

$$I - I^h \approx Ch^k$$

позволяет, несмотря на простоту, сделать вывод о том, что при уменьшении h в k раз погрешность квадратурной формулы также уменьшается примерно в n^k раз.

Пусть $h_n = h/n$. Тогда

$$I - I^{h_n} \approx Ch_n^k = \frac{1}{n^k} Ch^k.$$

Вычитая из предпоследней приближенной формулы последнюю, получим

$$I^{h_n} - I^h \approx \frac{-1}{n^k} Ch^k + Ch^k,$$

откуда определим постоянную C :

$$C = \frac{I^{h_n} - I^h}{h_n^k (n^k - 1)} = h_n^{-k} \frac{I^{h_n} - I^h}{n^k - 1} = \left(\frac{n}{h}\right)^k \frac{I^{h_n} - I^h}{n^k - 1}.$$

В таком случае мы можем вычислить приближенное значение искомого интеграла с шагом $h_n = h/n$:

$$I = I^{h_n} + \frac{1}{n^k} C h^k = I^h + \frac{I^{h_n} - I^h}{n^k - 1}.$$

Например, при уменьшении шага в два раза получим

$$I^{h_2} - I^h \approx \frac{1}{2^k} C h^k (2^k - 1),$$

откуда находим

$$C = \left(\frac{2}{h}\right)^k \cdot \frac{I^{h_2} - I^h}{2^k - 1}.$$

В таком случае апостериорная оценка погрешности квадратурной формулы для шага $h/2$ имеет вид®

$$I - I^{h_2} \approx \frac{I^{h_2} - I^h}{2^k - 1},$$

где величину в правой части приближенного равенства называют *поправкой Ричардсона*.

Эту оценку называют *правилом Рунге*, или *двойным пересчетом*; с ее помощью можно также достаточно простым способом увеличить порядок точности квадратурной формулы:

$$I \approx I^{h_2} + \frac{I^{h_2} - I^h}{2^k - 1}.$$

Заметим, что если при $k = 2$ мы получаем формулу трапеций для I^{h_2} , то формула соответствует формуле Симпсона. Если, например, $k = 4$ (I^{h_2} — значение I , вычисленное по формуле Симпсона), то экстраполяция Ричардсона дает

$$I \approx I^{h_2} = \frac{I^{h_2} - I^h}{15}.$$

Алгоритм Ромберга, позволяющий таким образом увеличить порядок точности квадратурной формулы интегрирования, заключается в следующем.

1. Вычисляем значение интеграла I^0 по формуле трапеций ($h_0 = b - a$).
2. Уменьшаем шаг вдвое: $h_1 = \frac{h_0}{2}$ (далее полагаем $h_i = h_{i-1}/2$).

3. Далее для $i = 1, 2, \dots$ вычисляем поправку Ричардсона:

$$\xi^{(k-1)}(h_i^{\circ}) = \frac{1}{2^k - 1} [I(h_i) - I(h_{i-1})]$$

и приближенное значение интеграла:

$$I^{(k)}(h_i) = I^{(k-1)}(h_i) + \xi^{(k-1)}(h_i); \quad k = 1, \dots, i;$$

i — номер приближения.

Окончание вычислений — при выполнении условия

$$\left| \xi^{(k-1)}(h_i) \right| \leq \delta,$$

где δ — заданная точность. При этом полагаем:

$$I \approx I^{(k)}(h_i).$$

Список литературы

1. Самарский А. А., Гулин А. В. Численные методы. М.: Наука, 1989. 430 с.
2. Вержбицкий В. М. Основы численных методов. М.: Высшая школа, 2002.

Дополнительная литература

3. Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. Численные методы. М.: ФИЗМАТЛИТ, 2000. 622 с.
4. Каханер Д., Моулер К., Нэш С. Численные методы и программное обеспечение. М.: Мир, 2001. 575 с.
5. Амосов А. А., Дубинский Ю. А., Копчёнова Н. В. Вычислительные методы. М.: МЭИ, 2008. 671 с.





ЧИСЛЕННОЕ РЕШЕНИЕ ЗАДАЧ КОШИ ДЛЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ (ОДУ)

8.1. Методы Рунге–Кутты (нежесткие задачи)

Обыкновенные дифференциальные уравнения (ОДУ) были предложены Ньютоном (1671 г.) и Лейбницем (1693 г.) для решения задач небесной механики. Для некоторых из них удавалось найти точные решения; однако их круг был серьезно ограничен, и уже в XVIII в. Эйлер («Интегральные исчисления», 1768 г.) предложил первый численный метод (метод ломаных) для численного решения задачи Коши для ОДУ. Однако из-за отсутствия вычислительной техники развитие вычислительных методов решения ОДУ в течение долгого времени не имело серьезного развития.

Лишь в 1895 г. появилась работа Рунге, в которой был предложен метод численного решения ОДУ, имеющий более высокий порядок точности, чем метод Эйлера, который в настоящее время используется, в основном, в учебно-методических целях. В 1901 г. появился классический четырехстадийный метод Рунге–Кутты. Бутчером была разработана технология построения методов типа Рунге–Кутты, что позволило построить высокоточные вычислительные методы. В конце XIX–начале XX вв. английский математик Адамс разработал семейство многошаговых методов, достоинством которых является возможность сравнительно простого повышения порядка их точности. Эти методы являются рабочими и в наше время.

Рассмотрим систему обыкновенных дифференциальных уравнений (ОДУ)

$$\frac{dx}{dt} = f(t, x), \quad t > 0, \quad x(0) = b, \quad (8.1)$$

где x и f являются векторами-столбцами:

$$x = \begin{Bmatrix} x_1 \\ \vdots \\ x_N \end{Bmatrix}, \quad f = \begin{Bmatrix} f_1 \\ \vdots \\ f_N \end{Bmatrix},$$

и принадлежат евклидову пространству.

В случае необходимости численного решения скалярного ОДУ N -го порядка

$$\begin{aligned}\frac{d^N x}{dt^N} &= f\left(t, x_t, x'_t, \dots, x_t^{(N-1)}\right), \quad t > 0, \\ x(0) &= b_0, \quad x'(0) = b_1, \dots, x_t^{(N-1)} = b_{N-1},\end{aligned}$$

последнее уравнение приводится к системе вида

$$\begin{cases} x_1 = x, \\ x_2 = \frac{dx_1}{dt}, \\ \dots\dots\dots \\ x_{k-1} = \frac{dx_{k-2}}{dt}, \\ \dots\dots\dots \\ x_N = \frac{dx_{N-1}}{dt}, \\ \frac{dx_N}{dt} = f(t, x_1, x_2, \dots, x_N), \end{cases}$$

где: $x_1(0) = b_0, x_2(0) = b_1, \dots, x_N(0) = b_{N-1}$.

Заметим, что встречаются задачи, которые решать численно более экономично, чем искать их аналитическое решение, имеющее неудобный с вычислительной точки зрения вид.

Например, решение обыкновенного дифференциального уравнения

$$\frac{dx}{dt} = \frac{x-t}{x+t}$$

представляется в виде трансцендентного алгебраического уравнения:

$$\frac{1}{2} \ln(t^2 + x^2) + \operatorname{arctg} \frac{x}{t} = \text{const},$$

численное решение которого найти не проще, чем численно решить само дифференциальное уравнение.

Напомним теорему существования и единственности решения ОДУ.

Теорема 8.1. Пусть $\Omega = \{t_0 \leq t \leq T, -\infty < x < \infty\}$ — множество точек $\{t, x\}$, на котором определена непрерывная функция $f(t, x)$, удовлетворяющая условию Липшица

$$|f(t, x_1) - f(t, x_2)| \leq C |x_1 - x_2|$$

для любых $t \in [t_0, T]$ и произвольных x_1, x_2 ; C — постоянная Липшица.

В этом случае для каждого начального значения x_0 существует единственное решение $x(t)$ задачи Коши

$$\frac{dx}{dt} = f(t, x), \quad x(0) = b, \quad t > 0,$$

определенное на отрезке $[t_0, T]$.

Введем на отрезке $[t_0, T]$ расчетную сетку

$$\omega_n = \{t_n = t_0 + nh; n = 0, \dots, N; \tau = (T - t_0) / N, t_0 = b\}$$

и сеточную функцию u_n , значения которой определены в узлах сетки ω_n :

$$u_n = u(t_n).$$

Заметим, что если необходимо вычислять значения между узлами, то можно использовать аппарат интерполяции функций.

Простейший способ приближенного решения скалярного ОДУ вида

$$\frac{dx}{dt} = f(t, x), \quad x(0) = b \quad (8.2)$$

состоит в разложении решения в ряд Тейлора

$$x(t) \approx \sum_{i=0}^n \frac{x^{(i)}(t_0)}{i!} (t - t_0)^i,$$

где производные $x^{(i)}(t)$ находятся из (8.2):

$$x'' = f_t(t, x) + f_x(t, x) x',$$

$$x''' = f_{tt}''(t, x) + 2f_{tx}''(t, x) x' + f_{xx}''(t, x) (x')^2 + f'_x(t, x) x''.$$

Подставив $t = t_0$, $x = x_0$ в (8.2) и в последнее соотношение, получим:

$$x'(t_0), \quad x''(t_0), \dots$$

Следует заметить, что если $|t - t_0|$ больше радиуса сходимости полученного ряда, то погрешность полученного решения не стремится к нулю при $n \rightarrow \infty$. Однако аналитические выражения производных высоких порядков становятся довольно сложными. Кроме того, современные численные методы оказываются более экономичными в машинных расчетах, чем приведенный метод.

Рассмотрим простейшие аппроксимации скалярного ОДУ (простейшие разностные схемы):

$$\frac{x_{n+1} - x_n}{\tau} = f(t_n, x_n). \quad (8.3)$$

Этот метод предложен Эйлером в 1768 г., а в 1820 г. Коши доказал для него теорему о сходимости;

$$\frac{x_{n+1} - x_n}{\tau} = f(t_{n+1}, x_{n+1}) \quad (8.4)$$

— неявный метод Эйлера;

$$\frac{x_{n+1} - x_{n-1}}{2\tau} = f(t_n, x_n) \quad (8.5)$$

— метод второго порядка аппроксимации.

Алгоритмическая реализация первой схемы (схема Эйлера) — «бегущий счет», т.е. рекуррентное соотношение, позволяющее вычислять x_{n+1} по значениям x_n :

$$x_{n+1} = x_n + \tau f(t_n, x_n), \quad x_0 = a.$$

Последовательно вычисляются x_1, x_2, \dots :

$$\begin{aligned} x_1 &= x_0 + \tau f(t_0, x_0) = b + \tau f(t_0, b); \\ x_2 &= x_1 + \tau f(t_1, x_1), \quad x_3 = x_2 + \tau f(t_2, x_2), \dots \end{aligned}$$

Вторая схема (неявная схема Эйлера) представляет собой нелинейное алгебраическое уравнение, которое решается на каждом слое $t = t_n$, причем в качестве начального приближения берется решение с предыдущего слоя, которое чаще всего по норме не сильно отличается от решения на следующем слое. По этой причине итерации для достаточно гладких функций в правых частях обычно сходятся быстро:

$$\begin{aligned} (x_{n+1})^{i+1} - x_n - \tau f[t_{n+1}, (x_{n+1})^i] &= 0; \\ x_{n+1}^i &= x_n, \quad i = 0, 1, \dots \end{aligned}$$

Критерием остановки вычислений может служить неравенство

$$|x_{n+1}^{i+1} - x_{n+1}^i| \leq \varepsilon,$$

где ε — заданная точность. Также для решения этого уравнения можно применить метод Ньютона.

Алгоритмическая реализация третьего метода — «бегущий счет»:

$$x_{n+1} = x_{n-1} + 2\tau f(t_n, x_n), \quad x_0 = b.$$

Однако для начала вычислений необходимо задать, кроме x_0 , еще и x_1 , вычислив его, например, из нелинейного уравнения

$$\frac{x_1 - x_0}{\tau} = \frac{1}{2} [f(t_0, u_0) + f(t_1, u_1)],$$

которое можно решить, к примеру, методом Ньютона или простых итераций.

Другой класс расчетных формул для численного решения ОДУ (8.2) можно получить, аппроксимируя с помощью квадратурных формул интеграл в равенстве

$$x(t_n + \tau) = x(t_n) + \int_{t_n}^{t_n + \tau} x'(\eta) d\eta. \quad (8.6)$$

Так, применяя формулу прямоугольников, получим

$$x(t_n + \tau) = x(t_n) + \tau x'_t(t_n) + O(\tau^2),$$

а так как

$$x'_t(t_n) = f(t_n, x_n),$$

то

$$x(t_n + \tau) = x(t_n) + \tau f(t_n, x_n) + O(\tau^2).$$

Пренебрегая членами второго порядка малости $O(\tau^2)$ и обозначив

$$x_{n+1} = t_n + \tau, \quad x(t_n) = x_n,$$

получим явный метод Эйлера

$$\frac{x_{n+1} - x_n}{\tau} = f(t_n, x_n), \quad x_0 = b.$$

Для получения вычислительного метода более высокого порядка точности, используем формулы трапеций:

$$x(t_n + \tau) = x(t_n) + \frac{\tau}{2} [x'_t(t_n) + x'_t(t_n + \tau)] + O(\tau^3)$$

или

$$x(t_n + \tau) = x(t_n) + \frac{\tau}{2} [x'_t(t_n, x_n) + f(t_n + \tau, x(t_n + \tau))] + O(\tau^3),$$

откуда, отбросив $O(\tau^3)$, получим неявную формулу трапеций, или неявный метод Адамса второго порядка точности:

$$\frac{x_{n+1} - x_n}{\tau} = \frac{1}{2} [f(t_n, x_n) + f(t_{n+1}, x_{n+1})]; \quad x_0 = b. \quad (8.7)$$

Мы получили нелинейное алгебраическое уравнение, которое можно решить итерационным методом. При аппроксимации интеграла с помощью метода средних будем иметь

$$x(t_n + \tau) = x(t_n) + \tau \cdot x'_t\left(t + \frac{\tau}{2}\right) + O(\tau^3),$$

или

$$x(t_n + \tau) = x(t_n) + \tau \cdot f\left[t_n + \frac{\tau}{2}, x\left(t_n + \frac{\tau}{2}\right)\right] + O(\tau^3),$$

откуда получим неявный метод:

$$\frac{x_{n+1} - x_n}{\tau} = f(t_{n+1/2}, x_{n+1/2}), \quad (8.8)$$

$$x_0 = b,$$

где

$$t_{n+1/2} = \frac{t_n + t_{n+1}}{2},$$

$$x_{n+1/2} = x\left(t_n + \frac{\tau}{2}\right).$$

Более точные расчетные формулы также можно получить, используя подход, называемый «*предиктор–корректор*». Например, схеме (8.7) будет соответствовать явный двухэтапный метод:

$$\begin{aligned} \tilde{x}_{n+1} &= x_n + \tau \cdot f(t_n, x_n), \\ x_{n+1} &= x_n + \frac{\tau}{2} \cdot [f(t_n, x_n) + f(t_{n+1}, \tilde{x}_{n+1})]. \end{aligned} \quad (8.9)$$

Схеме (8.8) соответствует метод «предиктор–корректор» следующего вида:

$$\begin{aligned} x_{n+1/2} &= x_n + \frac{\tau}{2} \cdot f(t_n, x_n), \\ x_{n+1} &= x_n + \tau f\left(t_n + \frac{\tau}{2}, x_{n+1/2}\right). \end{aligned} \quad (8.10)$$

Методы (8.9) и (8.10), а также метод Эйлера относятся к классу методов Рунге–Кутты, которые записываются в достаточно общем параметрическом виде:

$$x(t_n + \tau) = x(t_n) + \Delta_n x = x(t_n) + \tau \cdot \sum_{i=1}^r d_i k_i, \quad (8.11)$$

где r — число стадий, коэффициенты k_i вычисляются по формулам:

$$\begin{aligned} k_1 &= f(t_n, x_n), \\ k_2 &= f(t_n + a_2\tau, x_n + \tau b_{21}k_1), \\ k_3 &= f[t_n + a_3\tau, x_n + \tau(b_{31}k_1 + b_{32}k_2)], \\ &\dots\dots\dots \\ k_r &= f[t_n + a_r\tau, x_n + \tau(b_{r1}k_1 + b_{r2}k_2 + \dots + b_{r,r-1}k_{r-1})]. \end{aligned} \quad (8.12)$$

Представим выражения для коэффициентов k_i , $i = 1, \dots, 4$, в последнем случае:

$$\begin{aligned} k_1 &= f(t_n, x_n); \\ k_2 &= f\left(t_n + \frac{1}{2}\tau, x_n + \tau \cdot \frac{1}{2}k_1\right); \\ k_3 &= f\left(t_n + \frac{1}{2}\tau, x_n + \tau \cdot \frac{1}{2}k_2\right); \\ k_4 &= f(t_n + \tau, x_n + \tau k_3). \end{aligned}$$

Таблицы Бутчера для двух неявных методов Рунге–Кутты (метод средних и метод Хаммера–Холлинсворта) имеют следующий вид:

$\frac{1/2}{1/2}$	$\frac{1/2}{1}$	$\frac{1/2 - \sqrt{3}/6}{1/2 + \sqrt{3}/6}$	$\frac{1/4}{1/4 + \sqrt{3}/6}$	$\frac{1/4 - \sqrt{3}/6}{1/4}$
			$\frac{1/2}{1/2}$	$\frac{1/2}{1/2}$

Эти методы относятся к классу многостадийных одношаговых методов (т. е. сеточная функция вычисляется по данным на слое t_n).

Представим также коэффициенты для метода высокого порядка точности (метод Дормана–Принса 5-го порядка):

$$\begin{aligned} d_1 &= \frac{35}{384}; \quad d_2 = 0; \quad d_3 = \frac{500}{1113}; \\ d_4 &= \frac{125}{192}; \quad d_5 = -\frac{2187}{6784}; \quad d_6 = \frac{11}{84}, \\ k_1 &= f(t_n, x_n); \\ k_2 &= f\left(t_n + \frac{1}{5}\tau, x_n + \frac{\tau}{5}k_1\right); \\ k_3 &= f\left[t_n + \frac{3}{10}\tau, x_n + \tau\left(\frac{3}{40}k_1 + \frac{9}{40}k_2\right)\right]; \\ k_4 &= f\left[t_n + \frac{4}{5}\tau, x_n + \tau\left(\frac{44}{45}k_1 - \frac{56}{15}k_2 + \frac{32}{9}k_3\right)\right]; \\ k_5 &= f\left[t_n + \frac{8}{9}\tau, x_n + \right. \\ &\quad \left. + \tau\left(\frac{19472}{6561}k_1 - \frac{25360}{2187}k_2 + \frac{64448}{6561}k_3 - \frac{212}{729}k_5\right)\right]; \\ k_6 &= f[t_n + \tau, x_n + \\ &\quad + \tau\left(\frac{9017}{3168}k_1 - \frac{355}{33}k_2 + \frac{46732}{5247}k_3 + \frac{49}{176}k_4 - \frac{5103}{18656}k_5\right)]; \\ k_7 &= f(t_n + \tau, x_{n+1}). \end{aligned}$$

Рассмотрим получение метода Рунге–Кутты первого порядка (метод Эйлера). Введем погрешность вычисления решения на одном шаге τ :

$$\xi(\tau) = x(t_n + \tau) - \left(x(t_n) + \tau \sum_{i=1}^r d_i k_i \right).$$

Разложим $\xi(\tau)$ в ряд Маклорена:

$$\xi(\tau) = \sum_{i=0}^p \frac{\xi_{\tau}^{(i)}(0)}{i!} \tau^i + \frac{\xi_{\tau}^{(p+1)}(\eta\tau)}{(p+1)!} \tau^{p+1}, \quad 0 \leq \eta \leq 1,$$

и положим:

$$\xi_{\tau}^{(i)}(0) = 0, \quad i = 0, \dots, p,$$

т. е. погрешность на одном шаге вычисляется по формуле

$$\xi(\tau) = \frac{\xi_{\tau}^{(p+1)}(\eta\tau)}{(p+1)!} \tau^{p+1},$$

где p — порядок точности метода.

Рассмотрим простейший случай:

$$r = p = 1.$$

В этом случае

$$\xi(\tau) = x(t + \tau) - x(t) - \tau d_1 k_1.$$

Видно, что $\xi(\tau) = 0$. Для $\xi'_{\tau}(\tau)$ имеем

$$\xi'_{\tau}(0) = x'(t) - d_1 k_1 = (1 - d_1) \cdot f(t_n, x_n),$$

так как

$$x'_t(t) = k_1 = f(t_n, x_n).$$

Поскольку $\xi'_{\tau}(0) = 0$, то $d_1 = 0$, т. е. мы получаем одностадийный ($r = 1$) метод Рунге–Кутты первого порядка ($p = 1$).

Приведем пример исследования разностной задачи

$$\frac{x_{n+1} - x_n}{\tau} + ax_n = 0;$$

$$x_0 = b; \quad n = 0, 1, \dots,$$

аппроксимирующей задачу Коши для обыкновенного дифференциального уравнения

$$\frac{dx}{dt} + ax = 0, \quad t > 0, \quad x(0) = b,$$

на сходимость.

Точные решения дифференциальной и разностной задач соответственно имеют вид:

$$x(t) = be^{-at};$$

$$x_{n+1} = b(1 - \tau a)^n = b(1 - \tau a)^{t_n/\tau},$$

а погрешность ξ вычисляется по формуле

$$\xi_n = b \left| (1 - a\tau)^{t_n/\tau} - e^{-at_n} \right|.$$

Разложим в ряд Тейлора выражение:

$$\begin{aligned} (1 - a\tau)^{t_n/\tau} &= \exp \left[\frac{t_n}{\tau} \ln(1 - a\tau) \right] = \\ &= \exp \left[\frac{t_n}{\tau} \left(-a\tau + \frac{a^2\tau^2}{2} + O(\tau^2) \right) \right] = \\ &= \exp(-at_n) \cdot \exp \left(a^2 \frac{\tau t_n}{2} \right) \cdot \exp [O(\tau^2)] = \\ &= e^{-at_n} \left(1 + \frac{a^2\tau t_n}{2} + O(\tau^2) \right) [1 + O(\tau^2)] = \\ &= e^{-at_n} + \tau \frac{a^2\tau t_n}{2} e^{-at_n} + O(\tau^2). \end{aligned}$$

В таком случае решение разностного уравнения может быть записано в виде

$$u_n = be^{-at_n} + \tau b \frac{a^2 t_n}{2} e^{-at_n} + O(\tau^2),$$

а величина погрешности принимает вид

$$\xi_n = \tau b \frac{a^2 t_n}{2} e^{-at_n} + O(\tau^2) = O(\tau),$$

т.е. полученная погрешность ξ_n стремится к нулю при $\tau \rightarrow 0$ и имеет первый порядок.

Таким образом, мы провели исследование разностного метода Эйлера для линейного обыкновенного дифференциального уравнения. Однако для этого нам потребовалось знание точных решений как дифференциального, так и разностного уравнений, чего не бывает при решении реальных задач. Тем не менее иногда подобные исследования, когда они возможны, могут быть полезны для изучения свойств разностного метода.

В реальных задачах сходимость решений разностных уравнений доказывается путем их исследования на аппроксимацию и устойчивость, о чем будет идти речь ниже.

8.2. Метод Рунге-Кутты

При численном решении задачи Коши для обыкновенного дифференциального уравнения методами Рунге-Кутты можно достаточно просто повысить порядок точности метода, т. е. найти поправку, добавление которой к решению, полученному по схеме p -го порядка, дает численное решение $(p+1)$ -го порядка точности. Этот метод называется *методом Рунге-Кутты*, а оценка точности метода — *оценкой Рунге*.

Пусть X_n — проекция точного решения ОДУ на расчетную сетку на n -м шаге, x_1 — численное решение ОДУ на n -м шаге, вычисленное с постоянным шагом τ методом Рунге-Кутты p -го порядка точности соответственно. В этом случае справедлива оценка

$$X_n - x_1 = C\tau^p + O(\tau^{p+1}), \quad (8.13)$$

$C \neq C(\tau)$ — константа.

Если x_2 — численное решение, полученное тем же методом Рунге-Кутты с шагом $\tau/2$, то

$$X_n - x_2 = C \left(\frac{\tau}{2}\right)^p + O(\tau^{p+1}), \quad (8.14)$$

Вычитая из первого соотношения второе, получим выражение для C :

$$C = \left(\frac{2}{\tau}\right)^p \frac{x_2 - x_1}{2^p - 1} + O(\tau).$$

После подстановки C в (8.14) будем иметь

$$X_n - \left(x_2 + \frac{x_2 - x_1}{2^p - 1}\right) = O(\tau^{p+1}).$$

Величина, стоящая в скобках:

$$\tilde{x}_2 = x_2 + \frac{x_2 - x_1}{2^p - 1},$$

является уточненным численным решением порядка $p+1$. В этом и состоит метод Рунге-Кутты, а способ оценки точности решения называется *методом (оценкой) Рунге*. В данном случае эта оценка (разность между точным и численным решениями) имеет вид

$$X_n - x_2 = \frac{x_2 - x_1}{2^p - 1} + O(\tau^{p+1}).$$

С помощью такого способа оценки погрешности можно выбирать шаг интегрирования τ , необходимый для достижения точности ε на одном шаге:

$$\left| \frac{x_2 - x_1}{2^p - 1} \right| \leq \varepsilon.$$

Если последнее неравенство не выполняется, то шаг уменьшается вдвое; если вновь не выполняется, то шаг вновь уменьшается вдвое и т. д.

Формулы Рунге-Кутты, различающиеся на один порядок аппроксимации, называют *вложенными*. Для них выписываются два соотношения Рунге-Кутты:

$$x_{n+1} = x_n + \tau \sum_{j=1}^r d_j k_j$$

и

$$x_{n+1} = x_n + \tau \sum_{j=1}^r \bar{d}_j k_j.$$

Соответствующая таблица Бутчера будет иметь вид

0				
a_2	b_{21}			
\vdots	\vdots			
a_r	b_{r1}	\dots	$b_{r,r-1}$	
	d_1	\dots	d_{r-1}	d_r
	\bar{d}_1	\dots	\bar{d}_{r-1}	\bar{d}_r

Например, методы Фельберга 2(3) и Кутты-Мерсона 4(5) имеют следующие таблицы Бутчера (здесь запись $N(M)$ означает два порядка точности вложенной формулы: N и M):

0	0		
1	1		
1/2	1/4	1/4	
	1/2	1/2	0
	1/6	1/6	4/6

0	0					
1/3	1/3					
1/3	1/6	1/6				
1/2	1/8	0	3/8	2		
1	1/2	0	-3/2	2	0	
	1/2	0	0	2/3	1/6	

Приведем расчетные формулы для 5-стадийного метода Кутты–Мерсона:

$$\begin{aligned}
 k_1 &= f(t_n, x_n); \\
 k_2 &= f\left(t_n + \frac{\tau}{3}, x_n + \frac{k_1}{5}\right); \\
 k_3 &= f\left(t_n + \frac{\tau}{3}, x_n + \frac{k_1}{6} + \frac{k_2}{6}\right); \\
 k_4 &= f\left(t_n + \frac{\tau}{2}, x_n + \frac{k_1}{8} + \frac{3k_3}{8}\right); \\
 k_5 &= f\left(t_n + \tau, x_n + \frac{k_1}{8} - \frac{3k_2}{8} + 2k_4\right); \\
 x_{n+1} &= x_n + \tau\left(\frac{k_1}{2} - \frac{3k_3}{2} + 2k_4\right); \\
 \tilde{x}_{n+1} &= \tilde{x}_n + \tau\left(\frac{k_1}{6} + \frac{2k_4}{3} - \frac{k_5}{6}\right).
 \end{aligned}$$

8.3. Барьеры Бутчера

Довольно часто в вычислительной практике используются методы более высоких порядков: Фельберга 7(8) и Дормана–Принса 7(8), причем последний метод имеет наименьшую погрешность среди всех методов Рунге–Кутты 8-го порядка точности [6].

Казалось бы, что с увеличением стадий (этапов) в методах Рунге–Кутты на единицу так же будет увеличиваться и порядок точности метода. Этот вопрос был исследован Бутчером. Оказалось, что лишь для явных методов Рунге–Кутты с количеством стадий $r \leq 4$ возможно соотношение

$$p = r.$$

Однако при увеличении количества стадий $r \geq 5$ ситуация изменяется, о чем говорит теорема, доказанная Бутчером (первый барьер Бутчера).

Теорема 8.2. Среди явных методов Рунге–Кутты с числом стадий, равным 5, не существует методов 5-го порядка точности.

Более того, было показано, что при количестве стадий $5 \leq r \leq 7$ максимально возможный порядок точности меньше числа стадий на единицу: $p = r - 1$. Второй барьер Бутчера — это переход от $r = 7$ к $r = 8$. В этом случае оказалось, что $p = r - 2$ (при $r = 8, 9$). При дальнейшем увеличении количества стадий

разность $r - p$ также увеличивается. В частности, для третьего барьера Бутчера ($r = 10, 11$) имеем: $p = r - 3$, и т. д.

Барьеры Бутчера являются следствием роста постоянных Лебега при интерполяции на равномерной сетке, поскольку построение методов Рунге–Кутты связано с квадратурными формулами вычисления интегралов интерполяционного типа. Например, наиболее известный метод Рунге–Кутты 4-го порядка основан на квадратурной формуле Симпсона. Однако с повышением порядка аппроксимации, как известно, в квадратурных формулах для вычисления интегралов

$$\int_a^b f(t) dt = \sum_{k=1}^p c_k f(t_k) + \xi_p$$

не все коэффициенты квадратур c_k будут положительными, т. е. эти формулы становятся неправильными. Это является следствием роста постоянных Лебега, поскольку квадратурные формулы являются интерполяционными, откуда и появляются барьеры Бутчера. Этих недостатков лишены неявные методы, использующие квадратурные формулы Гаусса для вычисления интегралов.

Важнейшим вопросом при численном решении методами Рунге–Кутты является их устойчивость. Представим семейство этих методов численного решения задачи Коши для ОДУ

$$\frac{dx}{dt} = f(t, x), \quad t > 0, \quad x(0) = b \quad (8.15)$$

в виде:

$$\frac{x_{n+1} - x_n}{\tau} = F(t_n, x_n), \quad x_0 = b, \quad (8.16)$$

где $F(t_n, x_n)$ — так называемая *функция приращения* методов Рунге–Кутты, вычисляемая с помощью пересчетов правой части (8.15). Докажем теорему об устойчивости методов Рунге–Кутты, представленных в виде (8.16).

Теорема 8.3. Если функция $f(t, x)$, являющаяся правой частью ОДУ вида

$$\frac{dx}{dt} = f(t, x), \quad x(0) = b; \quad x, f \in R^n,$$

липищ-непрерывна по x :

$$\|f(t, x) - f(t, y)\| \leq C \|x - y\|,$$

причем $C \neq C(\tau)$, $C\tau \ll 1$, где C — постоянная, τ — шаг интегрирования, то разностное уравнение, аппроксимирующее

рассматриваемое ОДУ, представленное в виде

$$\frac{\mathbf{x}_{n+1} - \mathbf{x}_n}{\tau} = \mathbf{F}(t_n, \mathbf{x}_n), \quad \mathbf{x}_0 = \mathbf{b}$$

является устойчивым; при этом выполняется

$$\|\mathbf{x}_n - \mathbf{y}_n\| \leq e^{Ct_n} \left(\|\mathbf{x}_0 - \mathbf{y}_0\| + \frac{2\varepsilon}{C} \right),$$

где ε — малое возмущение правых частей разностного уравнения, а \mathbf{x}, \mathbf{y} — решения двух близких систем разностных уравнений:

$$\frac{\mathbf{x}_{n+1} - \mathbf{x}_n}{\tau} = \mathbf{F}(t_n, \mathbf{x}_n) + \boldsymbol{\eta}_n, \quad (8.17)$$

$$\frac{\mathbf{y}_{n+1} - \mathbf{y}_n}{\tau} = \mathbf{F}(t_n, \mathbf{x}_n) + \boldsymbol{\eta}'_n, \quad (8.18)$$

$$\|\boldsymbol{\eta}'\| \leq \varepsilon, \quad \|\boldsymbol{\eta}\| \leq \varepsilon.$$

Доказательство. Вычитая (8.18) из (8.17) и переходя к неравенству в норме, получим

$$\|\mathbf{x}_{n+1} - \mathbf{y}_{n+1}\| \leq \|\mathbf{x}_n - \mathbf{y}_n\| + \tau \|\mathbf{F}(t, \mathbf{x}_n) - \mathbf{F}(t, \mathbf{y}_n)\| + 2\tau\varepsilon.$$

С учетом условия липшиц-непрерывности $\mathbf{F}(t, \mathbf{x})$ из полученного неравенства следует

$$\|\mathbf{x}_{n+1} - \mathbf{y}_{n+1}\| \leq (1 + C\tau) \|\mathbf{x}_n - \mathbf{y}_n\| + 2\tau\varepsilon.$$

После последовательного применения этого неравенства получим

$$\|\mathbf{x}_1 - \mathbf{y}_1\| \leq (1 + C\tau) \|\mathbf{x}_0 - \mathbf{y}_0\| + 2\tau\varepsilon;$$

$$\|\mathbf{x}_2 - \mathbf{y}_2\| \leq (1 + C\tau) \|\mathbf{x}_1 - \mathbf{y}_1\| + 2\tau\varepsilon \leq$$

$$\leq (1 + C\tau)^2 \|\mathbf{x}_0 - \mathbf{y}_0\| + 2\tau\varepsilon [1 + (1 + C\tau)];$$

$$\|\mathbf{x}_3 - \mathbf{y}_3\| \leq (1 + C\tau) \|\mathbf{x}_2 - \mathbf{y}_2\| + 2\tau\varepsilon \leq$$

$$\leq (1 + C\tau)^3 \|\mathbf{x}_0 - \mathbf{y}_0\| + 2\tau\varepsilon [1 + (1 + C\tau) + (1 + C\tau)^2]; \dots$$

$$\dots, \|\mathbf{x}_n - \mathbf{y}_n\| \leq (1 + C\tau)^n \|\mathbf{x}_2 - \mathbf{y}_2\| + 2\tau\varepsilon [1 + (1 + C\tau) + \dots + (1 + C\tau)^{n-1}],$$

или, после суммирования геометрической прогрессии:

$$\|\mathbf{x}_n - \mathbf{y}_n\| \leq (1 + C\tau)^n \|\mathbf{x}_0 - \mathbf{y}_0\| + 2\tau\varepsilon \frac{(1 + C\tau)^n - 1}{(1 + C\tau) - 1} \leq$$

$$\leq (1 + C\tau)^n \left(\|\mathbf{x}_0 - \mathbf{y}_0\| + \frac{2\varepsilon}{C} \right) =$$

$$\begin{aligned}
 &= (1 + C\tau)^{t_n/\tau} \left(\|x_0 - y_0\| + \frac{2\varepsilon}{C} \right) \approx \\
 &\approx e^{Ct_n} \left(\|x_0 - y_0\| + \frac{2\varepsilon}{C} \right),
 \end{aligned}$$

если $C\tau \ll 1$. Теорема доказана.

Заметим, что для устойчивости численного метода необходимо, чтобы

$$Ct_n = C\tau n = O(1).$$

Это будет выполняться, если $t_n = O(1)$, так как $C = O(1)$, т.е. устойчивость при выполнении условия Липшица доказана для временных интегралов порядка $O(1)$. Для интегрирования на больших временных отрезках необходимо изучение других, более «тонких» свойств правых частей ОДУ. Условие $C\tau \ll 1$ можно переписать в виде $|f'_x| \tau \ll 1$. Это усиленное неравенство является условием выбора шага интегрирования τ при численном решении ОДУ методами Рунге–Кутты.

Следствие. Положим, что для матрицы

$$A(x) = \frac{1}{2} [f'_x(x) + f'^*_x(x)]$$

выполняется

$$(A(x)\eta, \eta) \leq -a(\eta, \eta),$$

т.е. матрицы A строго отрицательны для любых η, x и $a > 0$.

Такие траектории, в окрестности которых выполняется данное неравенство, называются *устойчивыми*. В этом случае при интегрировании ОДУ методом Рунге–Кутты p -го порядка аппроксимации погрешность численного решения есть $O(\tau^p)$ для любых t , $a\tau \ll 1$. При этом выполняется

$$\|x_n - y_n\| \leq (1 + C\tau)^n \|x_0 - y_0\| + \frac{2\varepsilon}{C} \leq \|x_0 - y_0\| + \frac{2\varepsilon}{C}.$$

Так как в правой части неравенства отсутствует экспонента, то утверждение доказано; метод Рунге–Кутты устойчив при любом t . Это важный результат для практических вычислений.

Если $(A(x)\eta, \eta) \leq 0$, т.е. тип траекторий — нейтральные (или не неустойчивые), то в этом случае показывается, что

$$\|x_{n+1} - y_{n+1}\| \leq (1 + C\tau^2)^n \|x_n - y_n\| + 2\tau\varepsilon,$$

или

$$\begin{aligned}
 \|x_n - y_n\| &\leq (1 + C\tau^2)^n \|x_0 - y_0\| + 2\tau\varepsilon \frac{(1 + C\tau^2)^n - 1}{C\tau^2} \leq \\
 &\leq e^{C\tau^2 n} \|x_0 - y_0\| + O(\tau^{p-1}),
 \end{aligned}$$

поскольку $\varepsilon = O(\tau^p)$, $p > 0$.

Из последнего неравенства видно, что при расчетах на временных интервалах $t_n = O(\tau^{-1})$, так как

$$C\tau^2 n = C\tau t_n = O(1);$$

метод устойчив, но его точность понижается до $O(\tau^{p-1})$.

Список литературы



1. Годунов С. К., Рябенький В. С. Разностные схемы. М.: Наука, 1973. 400 с.
2. Федоренко Р. П. Введение в вычислительную физику. Долгопрудный: Интеллект, 2008. 503 с.
3. Петров И. Б., Лобанов А. И. Лекции по вычислительной математике. М.: БИНОМ. Лаборатория знаний, 2006. 522 с.

Дополнительная литература

4. Рунтмайер Р., Мортон К. Разностные методы решения краевых задач. М.: Мир, 1972. 418 с.
5. Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. Численные методы. М.: ФИЗМАТЛИТ, 2000. 622 с.
6. Хайрер Э., Нерсет С., Ваннер Г. Решение обыкновенных дифференциальных уравнений. Нежесткие задачи. М.: Мир, 1990. 512 с.



Глава 9

ЧИСЛЕННОЕ РЕШЕНИЕ ЗАДАЧИ КОШИ ДЛЯ СИСТЕМ ЖЕСТКИХ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ



9.1. Понятие жестких систем ОДУ

Примерно в середине XX в. специалисты по численному решению задач Коши для обыкновенных дифференциальных уравнений встретились с неожиданными трудностями. Задачи решались явными методами Рунге–Кутты с автоматическим выбором шага интегрирования. Однако при численном решении некоторых задач шаг интегрирования становится столь малым, что при использовании первых электронно-вычислительных машин проведение расчетов практически останавливалось. По-видимому, эти проблемы впервые возникли при решении задач химической кинетики, которые можно представить в достаточно общем виде следующим образом:

$$\dot{y}_k = \sum_l B_{kl} y_l + \sum_{l,m} D_{klm} y_l y_m, \quad k, l, m = 1, \dots, M.$$

Здесь y_k — концентрации различных веществ, принимающих участие в реакциях; B_{kl} , D_{klm} — постоянные величины, характеризующие скорость протекания химических реакций.

Специалисты отмечали значительную разницу в значениях этих компонент — они могли различаться на много порядков. Существенная разница могла появляться и в поведении самих концентраций: концентрации различных веществ могли существенно различаться и по величине и по характерным временам заметных их изменений, они могли сильно меняться с течением времени. Приведем один пример такой системы ОДУ, описывающей изменение концентраций трех веществ (x , y , z):

$$\begin{cases} \dot{x} = -4 \cdot 10^{-2} x + 10^4 y z, \\ \dot{y} = 10^{-2} x - 10^4 y z - 3 \cdot 10^7 y^2, \\ \dot{z} = 3 \cdot 10^7 y^2, \end{cases}$$
$$x(0) = 1, \quad y(0) = 0, \quad z(0) = 0.$$

В дальнейшем подобные системы ОДУ встречались при численном решении многих задач биофизики, экономики, радиофизики, астрофизики, механики и др.

Характерным для такого рода задач было наличие участков с быстрым и с медленным изменением искомых параметров исследуемого процесса, которые называли *пограничным слоем* и *квазистационарным слоем* соответственно. В качестве примера можно привести такую систему из двух уравнений:

$$\begin{cases} \dot{x} = \alpha x + \delta^{-1} y, \\ \dot{y} = -\delta^{-1} y, \\ x(0) = A, \quad y(0) = B, \\ A, B, \alpha = O(1), \quad 0 < \delta \ll 1. \end{cases}$$

Точное решение такой системы ОДУ имеет вид:

$$\begin{aligned} x(t) &= A \exp(\alpha t) + B(1 + \alpha\delta)^{-1} \cdot [\exp(\alpha t) - \exp(-\delta^{-1}t)]; \\ y(t) &= B \exp(-\delta^{-1}t). \end{aligned}$$

В данном случае решение состоит из быстро убывающей и медленно изменяющейся компонент. Видна большая разница в коэффициентах α и δ . К тому же если эту же систему ОДУ представить в матричном виде

$$\dot{\mathbf{z}} = \mathbf{A}\mathbf{z},$$

где $\mathbf{z} = \{x, y\}^{-1}$ — вектор-столбец искомых функций, а $\mathbf{A} (2 \times 2)$ — матрица с постоянными заметно различающимися коэффициентами, то собственные числа этой матрицы также будут существенно различными:

$$\lambda_1 \approx \delta^{-1}, \quad \lambda_2 = O(1),$$

т. е. $|\lambda_1| / |\lambda_2| \gg 1$.

При численном решении такого рода задач исследователи столкнулись со следующей вычислительной проблемой. В самом начале процесса искомая функция (или одна из искомых функций) претерпевала значительное изменение, в соответствии с которым выбирался шаг численного интегрирования, который оказывался малым по причине быстрого изменения функции. Однако через небольшой промежуток времени характер процесса существенно не изменился качественно: траектория становится медленно меняющейся, однако расчет продолжается с тем же шагом. Если существенно увеличить шаг интегрирования, в соответствии с медленным изменением одной из искомых функций,

то происходит вычислительная катастрофа: численное решение начинает быстро колебаться и расти по абсолютной величине.

Дело в том, что исходя из условия устойчивости явного метода Рунге–Кутты для численного решения системы ОДУ вида

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}); \quad \mathbf{f}, \mathbf{x} \in R^n, \quad \mathbf{x}(0) = \mathbf{b}$$

шаг интегрирования должен выбираться из неравенства

$$\tau \|\mathbf{f}'_x(\mathbf{x})\| \ll 1.$$

Однако в таком случае этот шаг будет соответствовать самому быстрому из исследуемых процессов; расчеты же медленных процессов приведут к неоправданно большим затратам машинного времени. В этом случае необходимо либо решать рассматриваемую систему ОДУ с малым шагом τ , соответствующим самому быстрому процессу, либо проводить численное интегрирование с большим шагом, соответствующим медленно протекающим процессам. Однако тогда нам придется использовать (или разработать) такой метод, который позволил бы проводить расчет с шагом

$$\tau \gg \|\mathbf{f}'_x(\mathbf{x})\|^{-1}.$$

Поскольку величина $\|\mathbf{f}'_x(\mathbf{x})\|$ соответствует постоянной Липшица, то такие системы ОДУ называют *системами с большой константой Липшица*, или *жесткими системами*.

Дадим строгое определение жесткой системы ОДУ.

Определение 9.1. Задачу Коши для ОДУ вида

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(t, \mathbf{x}), \quad t > 0, \quad \mathbf{x}(0) = \mathbf{b}; \quad \mathbf{x}, \mathbf{f} \in R^N$$

назовем *жесткой*, если спектр матрицы

$$\mathbf{J} = \{\mathbf{f}'_x\}$$

можно разделить на две части:

1) *жесткий спектр*, для которого выполняется:

$$\operatorname{Re} \lambda_j(\mathbf{x}) \leq -\Lambda_0, \quad |\operatorname{Im} \lambda_j(\mathbf{x})| < |\operatorname{Re} \lambda_j(\mathbf{x})|, \quad j = 1, \dots, N_1;$$

2) *мягкий спектр*, для которого

$$|\lambda_j(\mathbf{x})| \leq \lambda_0 \ll \Lambda_0, \quad j = N_1 + 1, \dots, N.$$

Здесь λ_j ($j = 1, \dots, N$) — собственные значения матрицы \mathbf{J} , $\lambda_0 = \min_j |\lambda_j|$, $\Lambda_0 = \max_j |\lambda_j|$.

Отношение Λ_0/λ_0 называется *показателем жесткости* системы ОДУ; при этом обычно $\lambda_0 = O(1)$, величина $\Lambda_0 \gg 1$ в приложениях бывает больше 10^6 .

Что же касается устойчивых численных методов, обеспечивающих интегрирование при больших τ , таких, что

$$\tau \|f'_u(\mathbf{x})\| \gg 1,$$

то такими являются неявные разностные методы: в частности, неявные методы Рунге–Кутты, примеры которых были приведены выше.

Для объяснения этого факта удобно рассмотреть систему линейных обыкновенных дифференциальных уравнений вида

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}, \quad \mathbf{x}(0) = \mathbf{b} \quad (9.1)$$

и ее аппроксимацию с помощью явной и неявной разностных схем.

Представим точное решение этой системы в виде

$$\mathbf{x}(t) = \sum_{j=1}^{N_1} b_j e^{\lambda_j t} \boldsymbol{\omega}_j + \sum_{j=N_1+1}^N b_j e^{\lambda_j t} \boldsymbol{\omega}_j. \quad (9.2)$$

Видно, что в этой сумме значение первого слагаемого экспоненциально убывает, как $\exp(-\Lambda_0 t)$ и вне пограничного слоя является пренебрежительно малой величиной, практически не влияющей на решение. Второе же слагаемое соответствует квазистационарному участку решения системы (9.1).

Точное решение системы явных разностных уравнений, аппроксимирующей исходную систему ОДУ

$$\frac{\mathbf{x}_{n+1} - \mathbf{x}_n}{\tau} = \mathbf{A}\mathbf{x}_n, \quad \mathbf{x}_0 = \mathbf{b}, \quad (9.3)$$

или

$$\mathbf{x}_{n+1} = (\mathbf{E} + \tau \mathbf{A}) \mathbf{x}_n,$$

имеет вид [1]

$$\mathbf{x}_n = \sum_{j=1}^{N_1} b_j (1 + \tau \lambda_j)^n \boldsymbol{\omega}_j + \sum_{j=N_1+1}^N b_j (1 + \tau \lambda_j)^n \boldsymbol{\omega}_j, \quad (9.4)$$

где $\boldsymbol{\omega}_i$ — собственные векторы матрицы \mathbf{A} .

Первое слагаемое полученной суммы (жесткую компоненту решения) можно оценить так:

$$(1 + \tau \lambda_j)^n \sim (-\tau \Lambda_0)^n;$$

очевидно, что при $\tau\Lambda_0 \gg 1$, что реализуется в практических задачах, это слагаемое будет быстро расти, изменяя знак при каждом изменении n , т.е. мы получим осциллирующую функцию с быстро растущей амплитудой.

Для второго слагаемого (мягкой компоненты решения) при $\tau\Lambda_0 \ll 1$ получаем оценку

$$(1 + \tau\lambda_j)^n = e^{\tau n \lambda_j} [1 + O(\tau\lambda_0)], \quad n = [T/\tau], \quad 0 \leq t \leq T,$$

т.е. нежесткая компонента решения аппроксимирует соответствующую компоненту точного решения рассматриваемой системы ОДУ. Таким образом, для численного решения системы (9.1) явная разностная схема (9.3) непригодна по причине ее неустойчивости.

Что касается неявной разностной схемы, аппроксимирующей (9.1), имеющей вид

$$\frac{\mathbf{x}_{n+1} - \mathbf{x}_n}{\tau} = \mathbf{A}\mathbf{x}_{n+1}, \quad \mathbf{x}_0 = \mathbf{b}, \quad (9.5)$$

или

$$\mathbf{x}_{n+1} = (\mathbf{E} - \tau\mathbf{A})^{-1} \mathbf{x}_n,$$

то ее точное решение представляется в виде

$$\mathbf{x}_n = \sum_{j=1}^{N_1} b_j (1 - \tau\lambda_j)^{-n} \boldsymbol{\omega}_j + \sum_{j=N_1+1}^N b_j (1 - \tau\lambda_j)^{-n} \boldsymbol{\omega}_j. \quad (9.6)$$

В данном случае, как хорошо видно, жесткая часть разностного решения (первая сумма) стремится к нулю как $(\tau\Lambda_0)^{-n}$ и качественно верно описывает поведение решения в пограничном слое, а нежесткая часть (вторая сумма) аппроксимирует поведение соответствующей части точного решения рассматриваемой системы ОДУ.

Заметим, что если нас интересует решение исходной системы (9.1) в зоне погранслоя, то его можно получить путем интегрирования с малым (в сравнении с шагом на квазистационарном участке) шагом: $\tau \ll \Lambda_0^{-1}$.

9.2. Устойчивость жестких систем ОДУ

Исследование разностных схем, аппроксимирующих ОДУ, обычно иллюстрируется на тесте Далквиста [5].

Дифференциальное уравнение вида

$$\frac{dx}{dt} = \lambda x, \quad t > 0, \quad x(0) = b, \quad \lambda < 0, \quad (9.7)$$

имеет точное решение

$$x(t) = x_0 e^{\lambda t},$$

которое монотонно убывает при $t \rightarrow \infty$.

Рассмотрим простейшую разностную аппроксимацию (9.7) с помощью явного метода Эйлера:

$$\frac{x_{n+1} - x_n}{\tau} = \lambda x_n, \quad x_0 = a, \quad n = 0, 1, \dots, \quad (9.8)$$

откуда следует

$$x_{n+1} = (1 + \tau\lambda) x_n,$$

т. е. неравенство $|x_{n+1}| \leq |x_n|$, $n = 1, 2, \dots$, означающее устойчивость (9.8), выполняется при $|1 + \tau\lambda| \leq 1$.

При $\lambda < 0$ это условие эквивалентно ограничению на шаг интегрирования τ :

$$0 < \tau \leq \frac{2}{|\lambda|}. \quad (9.9)$$

Функция

$$R(\tau\lambda) = 1 + \tau\lambda$$

называется *функцией устойчивости*.

Определение 9.2. Разностный метод называется *абсолютно устойчивым*, если выполняется условие

$$|R(\tau\lambda)| \leq 1$$

при любых τ , и *условно устойчивым*, если он устойчив при некотором ограничении на τ .

В частности, явный метод Эйлера является условно устойчивым при выполнении условия (9.9), а неявный метод Эйлера:

$$\frac{x_{n+1} - x_n}{\tau} = \lambda x_{n+1}, \quad x_0 = b, \quad n = 0, 1, \dots,$$

для которого

$$R(\tau\lambda) = \left| (1 - \tau\lambda)^{-1} \right| < 1$$

при любом τ , является примером абсолютно устойчивого разностного метода.

Условная устойчивость явных методов является их недостатком вследствие ограничения на шаг τ . Неявные методы лишены этого недостатка, однако при их использовании приходится решать систему алгебраических уравнений, вообще говоря, нелинейную.

Для r -стадийных методов Рунге–Кутты с порядком аппроксимации $p = r$ ($p \leq 4$) функции устойчивости имеют следующий вид:

$$\begin{aligned} R_2(\tau\lambda) &= 1 + (\tau\lambda) + \frac{1}{2}(\tau\lambda)^2, \\ R_3(\tau\lambda) &= 1 + (\tau\lambda) + \frac{1}{2}(\tau\lambda)^2 + \frac{1}{6}(\tau\lambda)^3, \\ R_4(\tau\lambda) &= 1 + (\tau\lambda) + \frac{1}{2}(\tau\lambda)^2 + \frac{1}{6}(\tau\lambda)^3 + \frac{1}{24}(\tau\lambda)^4. \end{aligned}$$

Видно, что эти функции являются частичными суммами ряда Тейлора для функции $e^{\tau\lambda}$.

Определение 9.3.

Схема называется *A-устойчивой*, если $|R(\tau\lambda)| \leq 1$ при $\operatorname{Re}(\tau\lambda) < 0$.

Схема называется *Л-устойчивой*, если она *A-устойчива* и $R(\tau\lambda) \rightarrow 0$ при $\tau\lambda \rightarrow -\infty$.

Схема называется *L_p -устойчивой*, если она *A-устойчива* и $R(\tau\lambda) \rightarrow 0$ при $(\tau\lambda)^p \rightarrow -\infty$.

Таким образом, схема *A-устойчива*, если область устойчивости представляет собой левую полуплоскость комплексной плоскости $\operatorname{Re}(\tau\lambda) < 0$.

В случае если область устойчивости включает в себя угол в левой полуплоскости комплексной плоскости с вершиной в начале координат и с углом полураствора α , то метод называется *$A(\alpha)$ -устойчивым*. Например, *$A(\alpha)$ -устойчивыми* являются некоторые неявные методы, о которых речь пойдет ниже.

Теорема 9.1 (барьер Далквиста). *Не существует A-устойчивых разностных схем порядка аппроксимации выше второго.*

Свойством *A-устойчивости*, например, обладает неявный метод трапеций:

$$\frac{x_{n+1} - x_n}{\tau} = \frac{1}{2}[f(t_n, x_n) + f(t_{n+1}, x_{n+1})].$$

В этом случае для уравнения (9.7) получаем

$$\operatorname{Re}(\tau\lambda) = \left(1 + \frac{1}{2}(\tau\lambda)\right) / \left(1 - \frac{1}{2}(\tau\lambda)\right),$$

т. е. метод является *A-устойчивым*, так как $\operatorname{Re}(\tau\lambda) < 0$.

Таким образом, для повышения порядка аппроксимации необходимо конструировать другие, например, *$A(\alpha)$ -устойчивые* разностные схемы.

При разработке разностных схем для численного решения жестких систем обыкновенных дифференциальных уравнений необходимо учитывать следующие требования:

- схема должна быть аппроксимирующей;
- схема должна обладать устойчивостью, например: A , $A(\alpha)$, $A(0)$, α -устойчивостью;
- схему необходимо верифицировать на известных тестовых задачах.



9.3. Нелинейные жесткие системы ОДУ

Рассмотрим следующую систему двух нелинейных обыкновенных дифференциальных уравнений (система А. Н. Тихонова)

$$\begin{cases} \delta \cdot \dot{x} = F(x, y), \\ \dot{y} = G(x, y), \quad x(0) = a, \quad y(0) = b, \end{cases} \quad (9.10)$$

и перепишем ее в виде:

$$\begin{cases} \dot{x} = \Lambda F(x, y), \\ \dot{y} = G(x, y), \quad \Lambda = \delta^{-1}, \quad x(0) = a, \quad y(0) = b. \end{cases}$$

В этой системе: $0 < \delta \ll 1$ — малый параметр; $\Lambda \gg 1$ — большой параметр; $|a|, |b|, \|F\|, \|G\| = O(1)$.

Из характеристического уравнения

$$\det \mathbf{J} = \det \begin{pmatrix} \Lambda F'_x - \lambda & \Lambda F'_y \\ G'_x & G'_y - \lambda \end{pmatrix} = 0$$

находятся собственные значения матрицы Якоби \mathbf{J} :

$$\lambda_1 \approx \Lambda F'_x, \quad \lambda_2 = O(1), \quad (9.11)$$

откуда видно, что система (9.10) является жесткой при $F'_x < 0$, так как

$$|\lambda_1| / |\lambda_2| \gg 1.$$

Теперь дадим важное для данной темы определение сингулярно возмущенной задачи.

Определение 9.4. Пусть соответственно:

$$Lx = F \quad \text{и} \quad (L + \delta L_\delta) x_\delta = F + \delta F_\delta \quad (9.12)$$

— невозмущенная и возмущенная системы ОДУ, L и $F \in R^n$ — дифференциальный оператор и правая часть невозмущенной системы.

Задачу

$$(L + \delta L_\delta) x_\delta = F + \delta F_\delta$$

называют *регулярно возмущенной*, если выполняется

$$\|x_\delta(t) - x(t)\| \rightarrow \infty \quad \text{при} \quad \delta \rightarrow 0;$$

в противном случае задача называется *сингулярно возмущенной*.

Можно показать, что (9.10) является сингулярно возмущенной.

Рассмотрим в данной системе случай правых частей вида

$$\begin{cases} F = y - \frac{1}{3}x^3 + x, \\ G = -x. \end{cases}$$

График функции $F(x, y)$ представлен на рис. 9.1. Соответствующая кривая делит плоскость $\{x, y\}$ на две части, в которых $F > 0$ и $F < 0$; вне этой кривой поле скоростей близко к горизонтальному, на ней выделяются два участка: A_1A_2 , A_3A_4 (устойчивые) и один A_2A_4 — неустойчивый.

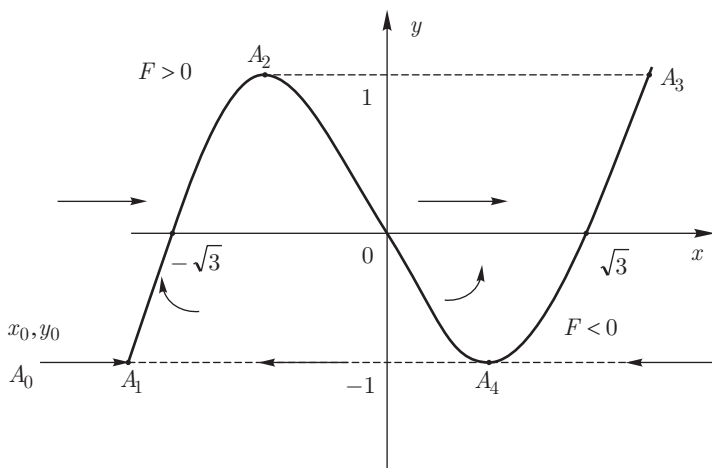


Рис. 9.1

На устойчивых участках выполняется условие устойчивости по Ляпунову

$$F'_x < 0,$$

на неустойчивом участке значение производной F' будет положительным:

$$F'_x > 0.$$

В точках A_2 , A_4 решение теряет устойчивость и происходит его «скачок» на устойчивую ветвь из точки A_2 в точку A_3 .

Качественное поведение решения в плоскости $\{x, y\}$ можно описать следующим образом.

Траектория движения из начальной точки $A_0\{x, y\}$ в некоторую точку устойчивой ветви A_1A_2 кривой $A_1A_2A_3A_4$ представляет собой пограничный слой. Поскольку поле скоростей вне этой кривой почти горизонтально, то A_0A_1 является почти горизонтальным отрезком в плоскости $\{x, y\}$, на котором за некоторый временной интервал $\Delta t_0 = O(\delta)$ траектория из A_0 переходит в малую δ -окрестность рассматриваемой кривой. Поскольку этот участок траектории решения нашей системы почти горизонтален, то решение определяется системой

$$\begin{cases} \dot{x} = \delta^{-1} F(x, y), \\ \dot{y} \approx y_0, \quad x(0) = x_0. \end{cases}$$

В малой δ -окрестности кривой $F(x, y) = 0$ выполняется

$$F'_x < 0;$$

в таком случае можно сделать следующую оценку:

$$F'_t = F'_x \cdot \dot{x} = F'_x \cdot \Lambda F,$$

из которой видно, что $F(x, y)$ на данном отрезке экспоненциально стремится к нулю (как экспонента с показателем $\Lambda F'_x$, $F'_x < 0$) за время $\Delta t_0 = O(\delta)$.

Движение точки $\{x(t), y(t)\}$ по участку A_1A_2 рассматриваемой кривой является квазистационарным; в этом случае оно подчиняется системе ОДУ вида

$$\begin{cases} F(x, y) = 0, \\ \dot{y} = G(x, y). \end{cases}$$

Здесь время движения точки $\{x(t), y(t)\}$ оценивается так:

$$\Delta t_1 = O(1),$$

и поведение системы является устойчивым, после чего она приходит к неустойчивому положению в точке A_2 . Затем движение за малое время $\Delta t_2 = O(\delta)$ переходит в точку A_3 , поскольку траектория, соответствующая части A_2A_4 рассматриваемой кривой, не может быть реализована по причине ее неустойчивости.

Далее реализуется квазистационарный участок кривой A_3A_4 (здесь $F'_x < 0$), где время движения

$$\Delta t_3 = O(1),$$

после чего на участке A_4A_1 происходит быстрое, за время $\Delta t_4 = O(\delta)$, движение («скачок» из точки A_4 в точку A_1). Замкнутая кривая (траектория) $A_1A_2A_3A_4A_1$ называется *предельным циклом*. Графики искомых функций также будут иметь два квазистационарных участка ($\Delta t_1 = O(1)$, $\Delta t_3 = O(1)$) и два быстрых, реализуемых за времена $\Delta t_0 = O(\delta)$ и $\Delta t_4 = O(\delta)$.

Численная реализация решения рассматриваемой системы с помощью явной схемы: например, схемы Эйлера

$$\begin{cases} \frac{x_{n+1} - x_n}{\tau} = \Lambda F(x_n, y_n), \\ \frac{y_{n+1} - y_n}{\tau} = G(x, y_n), \quad x_0 = a, \quad y = b, \end{cases}$$

представляется затруднительной, поскольку в зоне погранслоя должно выполняться условие, существенно ограничивающее шаг по времени:

$$\tau \cdot \Lambda |F'_x| \ll 1.$$

При расчете квазистационарного участка параметр τ можно выбирать много большим, поскольку времена движения на этих двух участках существенно разные: (δ) и (1) .

В этом случае целесообразным было бы использование неявной схемы, например,

$$\begin{cases} \frac{x_{n+1} - x_n}{\tau} = \Lambda F(x_{n+1}, y_{n+1}), \\ \frac{y_{n+1} - y_n}{\tau} = G(x_{n+1}, y_n), \quad x_0 = a, \quad y = b, \end{cases}$$

решение которой может быть реализовано итерационным методом: например, методом простых итераций или методом Ньютона.

Заметим, что в малой окрестности точки A_2 рассматриваемая система алгебраических уравнений может иметь более одного решения, одно из которых может принадлежать участку A_2A_4 кривой $A_1A_2A_3A_4A_1$. Такая ситуация может произойти при выборе большого шага интегрирования, при котором неявная разностная схема не теряет устойчивость. По этой причине при использовании неявных разностных схем для численного решения сингулярных жестких систем обыкновенных дифференциальных уравнений необходимо учитывать возможность получения нефизичного решения на неустойчивых ветвях.

9.4. Численные методы решения жестких систем ОДУ

Представим некоторые наиболее известные в вычислительной практике численные методы для решения жестких систем уравнений.

Неявные методы Рунге–Кутты могут быть представлены в виде

$$x_{n+1} = x_n + \tau \sum_{i=1}^r d_i k_i, \quad (9.13)$$

где коэффициенты k_i рассчитываются по формулам:

$$k_1 = f[t_n + a_1\tau, \quad x_n + \tau(b_{11}k_1 + b_{12}k_2 + \dots + b_{1r}k_r)].$$

.....

$$k_r = f[t_n + a_r\tau, \quad x_n + \tau(b_{r1}k_1 + b_{r2}k_2 + \dots + b_{rr}k_r)],$$

а коэффициенты a_i, b_{ij} ; $i, j = 1, \dots, r$ могут быть представлены в таблице Бутчера:

a_1	b_{11}	b_{12}	\dots	b_{1r}
a_2	b_{21}	b_{22}	\dots	b_{2r}
\vdots	\vdots			
a_r	b_{r1}	b_{r2}	\dots	$b_{r,r}$
	d_1	d_2	\dots	d_r

Такие таблицы для двух методов Гаусса, относящихся к классу неявных методов Рунге–Кутты, были представлены в гл. 8 (метод средних и метод Хаммера–Холлинсворта).

Таблицы для методов Радо, также принадлежащих к классу неявных методов Рунге–Кутты для первого и третьего порядков аппроксимации, имеют вид:

$\begin{array}{c c} 1 & 1 \\ \hline & 1 \end{array}$	$\begin{array}{c cc} 1/3 & 5/12 & -1/12 \\ \hline 1 & 3/4 & 1/4 \\ & 3/4 & 1/4 \end{array}$	(неявный метод Эйлера).
--	---	-------------------------

Представим также таблицы для методов Лобатто второго и четвертого порядков:

$\begin{array}{c cc} 0 & 0 & 0 \\ \hline 1 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array}$	$\begin{array}{c ccc} 0 & 0 & 0 & 0 \\ \hline 1/2 & 5/24 & 1/3 & -1/24 \\ 1 & 1/6 & 2/3 & 1/6 \\ \hline & 1/6 & 2/3 & 1/6 \end{array}$	(неявный метод трапеций).
---	--	---------------------------

Заметим, что наиболее известные неявные методы Рунге–Кутты подразделяются на несколько типов: гауссовы методы, основанные на квадратурных формулах Гаусса; методы Радо IA, Радо IIA, Лобатто IIIA, IIIB, IIIC, основанные на квадратурных формулах Радо и Лобатто. Эти методы подробно описаны в [6]. Получение приведенных расчетных формул можно найти в [1, 5, 6].

Следующий класс методов — безытерационные (полуявные) методы Розенброка, при вычислительной реализации которых не приходится численно решать систему нелинейных алгебраических уравнений. Схема Розенброка выглядит следующим образом [1]:

$$(\mathbf{E} - a\tau\mathbf{D} - b\tau^2\mathbf{D}^2) \frac{\mathbf{x}_{n+1} - \mathbf{x}_n}{\tau} = \mathbf{F}[\mathbf{x}_n + c\tau\mathbf{F}(\mathbf{x}_n)], \quad (9.14)$$

где \mathbf{E} — единичная матрица,

$$\mathbf{D} = \frac{d\mathbf{F}}{d\mathbf{x}}(\mathbf{x}_n)$$

— матрица Якоби, a, b, c — параметры метода, которые подбираются из условий устойчивости и обеспечения заданного порядка аппроксимации. При этом для вычисления решения \mathbf{x}_{n+1} необходимо дважды вычислить значения \mathbf{F} , вычислить компоненты матрицы \mathbf{D} , обратной матрицы ($\mathbf{E} = a\tau\mathbf{D} - b\tau^2\mathbf{D}^2$), а затем — решения. Для метода Розенброка третьего порядка аппроксимации параметры a, b, c имеют следующие значения:

$$a = 1,077; \quad b = -0,372; \quad c = -0,577.$$

Рассмотрим теперь метод неопределенных коэффициентов для численного решения, в первую очередь, жестких систем обыкновенных дифференциальных уравнений.

Для получения численного решения задачи Коши для обыкновенного дифференциального уравнения

$$\frac{dx}{dt} = f(t, x), \quad t > 0, \quad x(0) = X \quad (9.15)$$

рассмотрим разностные схемы вида

$$\frac{\alpha_0 x_n + \alpha_1 x_{n-1} + \dots + \alpha_m x_{n-m}}{\tau} = \beta_0 f_n + \beta_1 f_1 + \dots + \beta_m f_{n-m}$$

или:

$$\frac{\sum_{j=0}^m \alpha_j x_{n-j}}{\tau} = \sum_{j=0}^m \beta_j f_{n-j}; \quad \sum_{j=0}^m (\alpha_j x_{n-j} - \tau \beta_j f_{n-j}) = 0;$$

$$f_{n-j} = f(t_{n-j}, x_{n-j}).$$

Решение такого уравнения начинается при $n = m$ с уравнения

$$\frac{\alpha_0 x_m + \alpha_1 x_{m-1} + \dots + \alpha_m x_0}{\tau} = \beta_0 f_m + \beta_1 f_{m-1} + \dots + \beta_m f_0, \quad (9.16)$$

т. е. для того, чтобы начать расчет, необходимо задать m начальных значений: $j = 0, \dots, m-1$. Обычно задается значение x_0 , а значения u_1, \dots, u_{m-1} вычисляются, например, с помощью методов Рунге–Кутты. Отметим также, что, в отличие от методов Рунге–Кутты, в многошаговых методах правые части вычисляются только в узлах расчетной сетки.

При $\beta_0 = 0$ многошаговый метод называется *явным*, поскольку значение x_n вычисляется явно при известных значениях x_{n-1}, \dots, x_{n-m} . Если же $\beta_0 \neq 0$, то метод называется *неявным*.

В этом случае необходимо численно решать уравнение вида

$$\frac{1}{\tau} \alpha_0 x_n - \beta_0 f(t_n, x_n) = \sum_{j=1}^m \left(\beta_j f_{n-j} - \frac{1}{\tau} \alpha_j x_{n-j} \right)$$

с помощью итерационного метода, обычно для этого используется метод Ньютона при начальном приближении $x_n^{(0)} = x_{n-1}$.

Поскольку коэффициенты в (9.16) определены с точностью до множителя, вводится условие

$$\sum_{j=0}^m \beta_j = 1,$$

означающее, что правая часть разностного уравнения (9.16) аппроксимирует правую часть исходного ОДУ (9.15).

В практике численного решения задач распространение получили многошаговые методы Адамса, для которых:

$$\alpha_0 = 1, \quad \alpha_1 = -1, \quad \alpha_k = 0; \quad k = 2, 3, \dots, m,$$

т. е. в этих методах первая производная аппроксимируется по значениям искомой функции в двух узлах:

$$\frac{x_n - x_{n-1}}{\tau} = \sum_{j=0}^m \beta_j \cdot f_{m-j}. \quad (9.17)$$

При $\beta_0 = 0$ имеем явные методы Адамса, при $\beta_0 \neq 0$ — неявные.

В настоящее время для численного интегрирования жестких систем ОДУ наиболее широко используются чисто неявные

методы, или формулы дифференцирования назад (ФДН), для которых:

$$\begin{aligned} \beta_0 = 1, \beta_j = 0, \quad j = 1, \dots, m; \\ \frac{\alpha_0 x_n + \dots + \alpha_m x_{n-m}}{\tau} = f(t_n, x_n). \end{aligned} \quad (9.18)$$

Последнее уравнение является, вообще говоря, нелинейным:

$$\alpha_0 x - \tau f(t_n, x_n) = - \sum_{j=1}^m \alpha_j x_{n-j}$$

и решается итерационными методами.

Для исследования устойчивости многошагового метода (9.16) рассмотрим однородное разностное уравнение вида

$$\alpha_0 x_n + \alpha_1 x_{n-1} + \dots + \alpha_m x_{n-m} = 0, \quad n = m, m+1, \dots \quad (9.19)$$

Далее будем искать решение этого уравнения в виде

$$x_n = \lambda^n, \quad (9.20)$$

где значения λ находятся из характеристического уравнения, которое получим, если подставим x_n из (9.20) в уравнение (9.19):

$$\alpha_0 \lambda^m + \alpha_1 \lambda^{m-1} + \dots + \alpha_m = 0, \quad n = m. \quad (9.21)$$

Определение 9.5. Говорят, что метод (9.16) *удовлетворяет условию корней*, если все корни λ_j , $j = 1, \dots, m$, характеристического уравнения (9.19) лежат внутри единичного круга на комплексной плоскости и при этом на границе единичного круга нет кратных корней.

Теорема 9.2 (устойчивость однородного разностного уравнения). *Для того чтобы однородное разностное уравнение*

$$\alpha_0 x_n + \alpha_1 x_{n-1} + \dots + \alpha_m x_{n-m} = 0, \quad n = m, m+1, \dots,$$

было устойчиво по начальным данным, необходимо и достаточно, чтобы выполнялось условие корней.

В теории разностных уравнений доказывается теорема существования и единственности решений однородного разностного уравнения [7].

Далее рассмотрим задачу Коши для неоднородного разностного уравнения

$$\alpha_0 x_n + \alpha_1 x_{n-1} + \dots + \alpha_m x_{n-m} = \tau f_{n-m}; \quad (9.22)$$

здесь $n = m, m+1, \dots$; значения x_0, x_1, \dots, x_{m-1} заданы (начальные значения); f_{n-m} — правая часть; $x_k, f_k \in R^n$; $n, m \in N$.

Если $\alpha_0 \neq 0$, то решение задачи Коши неоднородного разностного уравнения существует и единственно; оно может быть найдено по рекуррентной формуле

$$x_n = -\frac{\alpha_m}{\alpha_0} x_{n-m} - \frac{\alpha_{m-1}}{\alpha_0} x_{n-m+1} - \dots - \frac{\alpha_1}{\alpha_0} x_{n-1} + \frac{\tau}{\alpha_0} f_{n-m},$$

если заданы начальные условия: x_k ; $k = 0, 1, \dots, m-1$, и правая часть.

Теорема 9.3 (устойчивость решений неоднородного разностного уравнения). *Если однородное уравнение*

$$\sum_{j=1}^m \alpha_j x_{n-j} = 0$$

устойчиво по начальным данным, то для неоднородного разностного уравнения

$$\sum_{j=1}^m \alpha_j x_{n-j} = \tau f_{n-m}$$

верна оценка

$$\|x_n\| \leq C \cdot \|x_j\| + C_1 \sum_{k=0}^{n-m} \tau \|f_k\|, \quad C \neq C(n), \quad C_1 \neq C_1(n).$$

Выполнение условий этой теоремы означает устойчивость неоднородного разностного уравнения по начальным данным.

Отметим, что все методы Адамса вида

$$\frac{x_n - x_{n-1}}{\tau} = \beta_0 \cdot f_n + \dots + \beta_n \cdot f_{n-m}$$

удовлетворяют условию корней, так как для этих методов выполняется

$$\alpha_0 = 1, \alpha_1 = -1, \quad \text{т. е.} \quad \lambda = \lambda_1 = 1.$$

Приведем пример разностного многошагового метода третьего порядка, который аппроксимирует исходное дифференциальное уравнение, но не удовлетворяет условию корней:

$$\frac{x_n + 4x_{n-1} - 5x_{n-2}}{6\tau} = \frac{2f_{n-1} + f_{n-2}}{3}.$$

Более подробно теория устойчивости многошаговых методов приведена в [3, 7].

Невязкой разностного метода является функция вида

$$\delta_n = -\tau^{-1} \sum_{j=0}^m \alpha_j U_{n-j} + \sum_{j=0}^m \beta_j f(t_{n-j}, U_{n-j}), \quad (9.23)$$

где U_{n-j} — точное решение аппроксимируемой дифференциальной задачи, подставленное в разностное уравнение (проекция точного решения на сетку).

Разложим функции U_{n-j} и f_{n-j} в ряд Тейлора [3]:

$$\begin{aligned} U_{n-j} &= \sum_{s=0}^p \frac{(-j\tau)^s \cdot U^{(s)}(t_n)}{s!} + O(\tau^{p+1}); \\ f(t_{n-j}, x_{n-j}) &= u'(t_n - j\tau) = \sum_{s=0}^p \frac{(-j\tau)^s \cdot U^{(s+1)}(t_n)}{s!} + O(\tau^p). \end{aligned} \quad (9.24)$$

После подстановки полученных разложений (9.24) в (9.23) получим

$$\begin{aligned} \delta_n &= - \sum_{j=0}^m \frac{\alpha_j}{\tau} \left(\sum_{s=0}^p \frac{(-j\tau)^s U^{(s)}(t_n)}{s!} \right) + \\ &+ \sum_{j=0}^m \beta_j \left(\sum_{s=0}^{p-1} \frac{(-j\tau)^s U^{(s+1)}(t_n)}{s!} \right) + O(\tau^p) = \\ &= - \sum_{s=0}^p \left(\sum_{j=0}^m \frac{\alpha_j}{\tau} \cdot \frac{(-j\tau)^s U^{(s)}(t_n)}{s!} \right) + \\ &+ \sum_{s=1}^p \left(\sum_{j=0}^m \beta_j \cdot \frac{(-j\tau)^s U^{(s)}(t_n)}{(s-1)!} \right) + O(\tau^p) = \\ &= - \left(\tau^{-1} \sum_{j=0}^m \alpha_j \right) U(t_n) + \sum_{s=1}^p \left(\sum_{j=0}^m (-j\tau)^{s-1} \left(\frac{j}{s} \alpha_j + \beta_j \right) \right) \times \\ &\quad \times \frac{U^{(s)}(t_n)}{(s-1)!} + O(\tau^p). \end{aligned} \quad (9.25)$$

Из последнего выражения следует, что разностное уравнение имеет порядок аппроксимации p , если:

$$\sum_{j=0}^m \alpha_j = 0, \quad \sum_{j=0}^m j^{s-1} (j\alpha_j + s\beta_j) = 0; \quad s = 1, \dots, p. \quad (9.26)$$

Соотношение (9.26) с условием

$$\sum_{j=1}^m \beta_j = 1 \quad (9.27)$$

образуют систему из $(p + 2)$ алгебраических уравнений с $2(m + 1)$ неизвестными α_j, β_j ; $j = 0, 1, \dots, m$. Если во втором уравнении в (9.26) выделить уравнение с $s = 1$:

$$\sum_{j=0}^m j \alpha_j + \sum_{j=0}^m \beta_j = 0 \quad (9.28)$$

и учесть условие нормирования (9.27), то получим систему линейных алгебраических уравнений с $2m$ неизвестными α_j, β_j ; $j = 0, 1, \dots, m$:

$$\begin{cases} \sum_{j=1}^m j \alpha_j = -1, \\ \sum_{j=1}^m j^{s-1} (j \alpha_j + s \beta_j) = 0; \quad s = 2, \dots, p. \end{cases} \quad (9.29)$$

При этом α_0 и β_0 вычисляются из условий:

$$\alpha_0 = - \sum_{j=1}^m \alpha_j, \quad \beta_0 = 1 - \sum_{j=1}^m \beta_j. \quad (9.30)$$

Очевидно, что система (9.29) не будет переопределенной, если выполняется неравенство

$$p \leq 2m.$$

Откуда следует, что порядок аппроксимации линейных многошаговых (m -шаговых) разностных методов не может быть больше $2m$ (для неявных методов $2m$, для явных m).

В случае $\alpha_0 = -\alpha_1 = 1$ (методы Адамса p -го порядка) получаем:

$$\begin{aligned} s \sum_{j=1}^m j^{s-1} \cdot \beta_j &= 1; \quad s = 2, 3, \dots, p; \\ \beta_0 &= 1 - \sum_{j=1}^m \beta_j \end{aligned} \quad (9.31)$$

откуда следует, что наивысшие порядки неявных и явных методов Адамса равны соответственно $m + 1$ и m .

О МНО
НИЯ
ЛАНЬ®

$$\sum_{j=0}^m \alpha_j \cdot x_{n-j} = \tau f(t_{n+1}, x_{n+1}). \quad (9.32)$$

При выполнении условий p -го порядка аппроксимации система линейных уравнений (9.29), (9.30) приобретает следующий вид:

$$\begin{cases} \alpha_0 = -\sum_{j=1}^m \alpha_j; \\ \sum_{j=1}^m j \alpha_j = -1; \\ \sum_{j=1}^m j^s \alpha_j = 0; \quad s = 2, \dots, p, \end{cases} \quad (9.33)$$

откуда следует, что максимальный порядок аппроксимации чисто m -шагового неявного метода равен m .

Представим (9.33) в виде системы линейных алгебраических уравнений:

[illegible]

при этом α_0 вычисляется так:

$$\alpha_0 = -(\alpha_1 + \dots + \alpha_m). \quad (9.35)$$

Чисто неявные методы обладают хорошими свойствами устойчивости, что делает их использование при численном решении жестких систем ОДУ вполне приемлемыми.

Рассмотрим получение некоторых многошаговых разностных методов, следующих из полученных систем линейных уравнений (9.29), (9.31), (9.34).

Для явных m -шаговых методов Адамса

$$\frac{x_n - x_{n-1}}{\tau} = \beta_1 \cdot f_{n-1} + \dots + \beta_m \cdot f_{n-m}, \quad (9.36)$$

максимальный порядок аппроксимации равен m . В соответствии с (9.31) имеем

$$\sum_{j=1}^m j^{l-1} \cdot \beta_j = l^{-1}, \quad l = 1, 2, \dots, m. \quad (9.37)$$

При решении данной системы алгебраических уравнений для различных m получаем методы Адамса максимального порядка аппроксимации.

При $m = 1$ имеем

$$\frac{x_n - x_{n-1}}{\tau} = f_{n-1} \quad (\text{явный метод Эйлера}).$$

С увеличением m до значений 2, 3, 4 получим методы соответствующих порядков (2, 3, 4, 5):

$$\begin{aligned} \frac{x_n - x_{n-1}}{\tau} &= \frac{3}{2} f_{n-1} - \frac{1}{2} f_{n-2}, \quad p = 2; \\ \frac{x_n - x_{n-1}}{\tau} &= \frac{1}{12} (23f_{n-1} - 16f_{n-2} + 5f_{n-3}), \quad p = 3; \\ \frac{x_n - x_{n-1}}{\tau} &= \frac{1}{24} (55f_{n-1} - 59f_{n-2} + 37f_{n-3} - 9f_{n-4}), \quad p = 4. \end{aligned} \quad (9.38)$$

В случае неявных m -шаговых методов Адамса

$$\frac{x_n - x_{n-1}}{\tau} = \beta_0 \cdot f_n + \beta_1 \cdot f_{n-1} \dots + \beta_m \cdot f_{n-m}; \quad (9.39)$$

мы имеем разностные методы с максимальным порядком аппроксимации $p = m + 1$.

Так, при $m = 1$ получаем неявный метод трапеций:

$$\frac{x_n - x_{n-1}}{\tau} = \frac{1}{2} (f_n + f_{n-1}), \quad p = 2;$$

при $m = 2, 3$ получим соответствующие неявные методы $(m + 1)$ -го порядка:

$$\begin{aligned} \frac{x_n - x_{n-1}}{\tau} &= \frac{1}{12} (5f_n + 8f_{n-1} - f_{n-2}), \quad p = 3; \\ \frac{x_n - x_{n-1}}{\tau} &= \frac{1}{24} (9f_n + 19f_{n-1} - 5f_{n-2} + f_{n-3}), \quad p = 4. \end{aligned} \quad (9.40)$$

Поскольку формулы (9.40) неявные (x_n входит в функцию f_n), то их необходимо решать численно, используя итерационные методы. Так, для метода Адамса 4-го порядка аппроксимации

можно построить следующий итерационный процесс (i — итерационный индекс):

$$\frac{x_n^{(i+1)} - x_{n-1}}{\tau} = \frac{1}{24} \left(9f(t_n, x_n^{(i)}) + 19f(t_{n-1}, x_{n-1}) - 5f(t_{n-2}, x_{n-2}) + f(t_{n-3}, x_{n-3}) \right), \quad i = 0, 1, \dots \quad (9.41)$$

Начальное значение x_n^0 можно получить, используя явный метод Адамса третьего порядка:

$$\frac{x_n^{(0)} - x_{n-1}}{\tau} = \frac{1}{12} [9f(t_{n-1}, x_{n-1}) - 16f(t_{n-2}, x_{n-2}) + 5f(t_{n-3}, x_{n-3})]. \quad (9.42)$$

Заметим, что если в итерационном процессе (9.41) ограничиться одной итерацией ($i = 1$), то из (9.41) и (9.42) получим двухэтапный метод предиктор–корректор.

Чисто неявные методы, или формулы дифференцирования назад для $m = 1, 2, 3$ можно получить из системы линейных алгебраических уравнений (9.34).

Для $m = 1$ получим неявный метод Эйлера:

$$\frac{x_n - x_{n-1}}{\tau} = f(t_{n+1}, x_{n+1});$$

для $m = 2, 3, 4$ получаем:

$$\begin{aligned} \frac{3}{2}x_n - 2x_{n-1} + \frac{1}{2}x_{n-2} &= \tau f(t_{n+1}, x_{n+1}), \quad p = 2; \\ \frac{11}{6}x_n - 3x_{n-1} + \frac{3}{2}x_{n-2} - \frac{1}{3}x_{n-3} &= \tau f(t_{n+1}, x_{n+1}), \quad p = 3; \\ \frac{1}{12}(25x_n - 48x_{n-1} + 36x_{n-2} - 16x_{n-3} + 3x_{n-4}) &= \\ &= \tau f(t_{n+1}, x_{n+1}), \quad p = 4. \end{aligned}$$

Одним из наиболее популярных методов ФДН является метод Гира [5, 6], имеющий наивысший шестой порядок точности и использующий на первых этапах расчета методы ФДН с $p = 1, \dots, 5$.

Список литературы

1. Федоренко Р. П. Введение в вычислительную физику. Долгопрудный: Интеллект, 2008. 503 с.
2. Петров И. Б., Лобанов А. И. Лекции по вычислительной математике. М.: БИНОМ. Лаборатория знаний, 2006. 522 с.
3. Самарский А. А., Гулин А. В. Численные методы. М.: Наука, 1989. 430 с.

Дополнительная литература



4. *Ракитский Ю. В., Устинов С. М., Черноруцкий И. Г.* Численные методы решения жестких систем. М.: Наука, 1979. 208 с.
5. *Уатт Дж., Холл Дж.* Современные численные методы решения обыкновенных дифференциальных уравнений. М.: Мир, 1979. 312 с.
6. *Хайрер Э., Ваннер Г.* Решение обыкновенных дифференциальных уравнений. Жесткие и дифференциально-алгебраические задачи. М.: Мир, 1999. 685 с.
7. *Романко В. К.* Курс разностных уравнений. М.: ФИЗМАТЛИТ, 2012. 199 с.
8. *Васильева А. Б., Бутузов В. Ф.* Асимптотические методы в теории сингулярных возмущений. М.: Высш. шк., 1990. 208 с.



ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ КРАЕВЫХ ЗАДАЧ ДЛЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

10.1. Метод фундаментальных систем

В отличие от задачи для ОДУ, для краевой задачи ставятся условия не на одном, а на двух концах отрезка интегрирования.

В начале главы приведем метод решения краевой задачи для системы линейных обыкновенных дифференциальных уравнений вида

$$\frac{d\mathbf{u}}{dx} = \mathbf{D}(x) \cdot \mathbf{u} + \mathbf{F}(x), \quad (10.1)$$

где $x \in [0, X]$, \mathbf{u}, \mathbf{F} — n -мерные векторы, $\mathbf{D}(x) \in M(n \times n)$ — матрица $n \times n$ ($M(n \times n)$ — пространство матриц размера $n \times n$), $[0, X]$ — интервал интегрирования, x — независимая переменная.

Для замыкания задачи необходимо задать n конечных соотношений (краевых условий):

$$\mathbf{A}\mathbf{u}(0) + \mathbf{B}\mathbf{u}(X) = \mathbf{G}, \quad \mathbf{A}, \mathbf{B} \in M(n \times n), \quad \mathbf{G} \in R^n. \quad (10.2)$$

Из курса обыкновенных дифференциальных уравнений известно, что общее решение системы ОДУ (10.1) задается следующей формулой (метод фундаментальных систем):

$$\mathbf{u}(x) = \mathbf{U}(x) + \sum_{j=1}^n \gamma_j \mathbf{v}^j, \quad (10.3)$$

где $\mathbf{U}(x) \in R^n$ — произвольное (частное) решение неоднородной системы ОДУ

$$\frac{d\mathbf{U}}{dx} = \mathbf{D}\mathbf{U} + \mathbf{F} \quad (10.4)$$

с некоторыми достаточно произвольными, например однородными, краевыми условиями, удобными для построения решения (10.1), а $\mathbf{v}^j(x)$ — линейно независимые решения однородной системы

$$\frac{d\mathbf{v}^j}{dx} = \mathbf{D}\mathbf{v}^j$$

с неоднородными краевыми условиями, обеспечивающими линейную независимость векторов \mathbf{v}^i , $j = 1, \dots, n$ при любых t .

Коэффициентные γ_j находятся из n заданных краевых условий, для чего необходимо решение вида (10.3) подставить в (10.2):

$$\mathbf{A} \left(\mathbf{U}(0) + \sum_{j=1}^n \gamma_j \mathbf{v}^j(0) \right) + \mathbf{B} \left(\mathbf{U}(X) + \sum_{j=1}^n \gamma_j \mathbf{v}^j(X) \right) = \mathbf{G}, \quad (10.5)$$

и полученное соотношение представить в виде

$$\sum_{j=1}^n \gamma_j \{ \mathbf{A} \mathbf{v}^j(0) + \mathbf{B} \mathbf{v}^j(X) \} = \mathbf{G} - \{ \mathbf{A} \mathbf{U}(0) + \mathbf{B} \mathbf{U}(X) \}. \quad (10.6)$$

В результате мы получим систему из n линейных алгебраических уравнений относительно n неизвестных γ_j ; $j = 1, \dots, n$, с матрицей, j -й столбец которой имеет вид

$$\mathbf{A} \mathbf{v}^j(0) + \mathbf{B} \mathbf{v}^j(X).$$

Если эта система имеет единственное решение, то и краевая задача имеет единственное решение. Обычно в реальных задачах это условие выполняется; если решение этой системы не существует или не единственно, то задача считается вырожденной (возможно, необходимо уточнить ее постановку).

Частное решение неоднородной системы ОДУ с однородными начальными данными $\mathbf{U}(0) = 0$ можно получить, используя один из методов Рунге–Кутты, например, метод первого порядка точности:

$$\mathbf{U}_{k+1} = \mathbf{U}_k + h [D(x_k) \cdot \mathbf{U}_k + \mathbf{F}(x_k)],$$

где $h = x_{k+1} - x_k$; $\mathbf{U}(0) = 0$; $k = 0, 1, \dots, n$.

Отметим, что в вычислительной практике, как правило, используются явные методы Рунге–Кутты более высоких порядков точности.

Для получения линейно независимых численных решений $\mathbf{v}^j(x)$ используются естественные начальные данные Коши:

$$\mathbf{v}^j(0) = \left\{ 0, \dots, 1, 0, \dots \right\}_j = \mathbf{e}^j, \quad j = 1, \dots, n,$$

где \mathbf{e}^j — j -й орт n -мерного векторного пространства.

Таким образом, численное решение краевой задачи свелось к решению $(N + 1)$ -кратного решения задачи Коши каким-либо из известных явных методов Рунге–Кутты заданного порядка точности.

Однако можно привести пример системы ОДУ, для которой приведенный метод не даст желаемого результата:

$$\begin{cases} \dot{x} = Ay + G, \\ \dot{y} = Bx + \tilde{G}, \end{cases} \quad x(0) = x_0, \quad y(0) = 0.$$

Решение такой системы можно представить в следующем виде (для простоты положим $G = \tilde{G} = 0$):

$$\begin{aligned} \begin{Bmatrix} x \\ y \end{Bmatrix} &= \begin{Bmatrix} X \\ Y \end{Bmatrix} + \gamma_1 \begin{Bmatrix} x \\ y \end{Bmatrix}_1 + \gamma_2 \begin{Bmatrix} x \\ y \end{Bmatrix}_2 = \begin{Bmatrix} X \\ Y \end{Bmatrix} + \\ &+ \gamma_1 \left[\begin{Bmatrix} \xi \\ \eta \end{Bmatrix}_1 e^{\lambda_1 t} + \begin{Bmatrix} \xi \\ \eta \end{Bmatrix}_2 e^{\lambda_2 t} \right] + \gamma_2 \left[\begin{Bmatrix} \xi \\ \eta \end{Bmatrix}_3 e^{\lambda_1 t} + \begin{Bmatrix} \xi \\ \eta \end{Bmatrix}_4 e^{\lambda_2 t} \right], \end{aligned}$$

где λ_1, λ_2 являются корнями характеристического уравнения

$$\det \begin{Bmatrix} -\lambda & A \\ B & -\lambda \end{Bmatrix} = 0,$$

т. е. $\lambda_{1,2} = \pm \sqrt{AB}$.

Если значения A и B достаточно большие ($\sim 10^2$), то получение решения состоит из двух компонент (как в первой, так и во второй квадратных скобках): одной — быстро возрастающей $\sim e^{100t}$, другой — быстропадающей $\sim e^{-100t}$.

Напомним, что $\begin{Bmatrix} x \\ y \end{Bmatrix}_1$ и $\begin{Bmatrix} x \\ y \end{Bmatrix}_2$ являются решениями двух однородных систем ОДУ следующих видов:

$$\begin{cases} \dot{x}_1 = Ay_1, \\ \dot{y}_1 = Bx_1, \\ x_1(0) = 1, \quad y_1(0) = 0 \end{cases}$$

и

$$\begin{cases} \dot{x}_2 = Ay_2, \\ \dot{y}_2 = Bx_2, \\ x_2(0) = 0, \quad y_2(0) = 1; \end{cases}$$

X, Y — частное решение соответствующей неоднородной системы с однородными краевыми условиями.

Правильное решение данной задачи (соответствующее, например, физике проблемы проникания потоков нейтронов через защиту реактора) соответствует падающей экспоненте, так как защита ослабляет потоки нейтронов; при этом растущие экспоненты должны взаимно уничтожиться. Однако при численном решении экспонентам будут соответствовать большие погрешности.

Пусть, для примера, мы оцениваем машинное значение x_M (δ — машинная ошибка):

$$x_M = x(1 + \delta) \sim e^{100t} (1 + 10^{-10}), \quad t = O(1);$$

в этом случае рост погрешности $\sim e^{100t}$ приводит к получению нефизичного результата и от метода фундаментальных систем трудно ожидать адекватного решения.

Важно отметить, что, как и в случае задачи Коши для обыкновенных дифференциальных уравнений, краевые задачи для ОДУ также могут быть жесткими. В этом случае систему ОДУ вида

$$\frac{du}{dx} = Du + F; \quad F, u \in R^n$$

с краевыми условиями вида

$$\begin{aligned} \sum_{j=1}^n z_{ij} x_j(0) &= y_i; \quad i = 1, \dots, I; \quad I < n; \\ \sum_{j=1}^n z_{ij} x_j(X) &= y_i; \quad i = I + 1, \dots, n, \end{aligned}$$

где z_{ij} — некоторые постоянные, называем *жесткой*, если спектр матрицы D разделяется на три части:

левый жесткий спектр:

$$\operatorname{Re} \lambda_j \leq -L_0, \quad |\operatorname{Im} \lambda_j| < L_0; \quad j = 1, \dots, n_1;$$

правый жесткий спектр:

$$\operatorname{Re} \lambda_j \geq L_0, \quad |\operatorname{Im} \lambda_j| < L_0; \quad j = n_1 + 1, \dots, n_2;$$

мягкий спектр:

$$|\lambda_j| \leq l_0, \quad j = n_2 + 1, \dots, n.$$

Число $\gamma = L_0/l_0$ является параметром, характеризующим жесткость системы ОДУ; полагаем, что в случае жесткой системы

$$X l_0 = O(1), \quad X L_0 \gg 1.$$

Структура рассматриваемой системы следует из общего решения одной задачи:

$$u(x) = \sum_{j=1}^{n_1} \gamma_j e^{\lambda_j x} \cdot \omega_j + \sum_{j=n_1+1}^{n_2} \gamma_j e^{\lambda_j x} \cdot \omega_j + \sum_{j=n_2+1}^n \gamma_j e^{\lambda_j x} \cdot \omega_j.$$

В этом решении первые два слагаемых относятся соответственно к левому и правому пограничным слоям, третье — к мягкому

спектру. Особенностью жестких краевых задач, вследствие ограниченности отрезка интегрирования, является ограниченность их решения:

$$\|\mathbf{u}\| \leq C (\|\mathbf{F}\| + \|\mathbf{Y}\|),$$

$$C = O(1), \quad \mathbf{Y} = \{y_1, \dots, y_N\}.$$

Заметим, что, вообще говоря, значение C может быть $\sim e^{L_0 T}$, однако мы рассматриваем задачи, где $C = O(1) \ll e^{L_0 X}$ (вычислительно корректные задачи). Можно показать, что необходимыми условиями вычислительно корректной задачи являются неравенства:

- $n_1 \geq I$ (количество краевых условий I на левой границе отрезка интегрирования должны быть не меньше количества быстро убывающих вправо решений I);
- $n - I \geq n_2 - n_1$ (количество краевых условий на правой границе не должно быть меньше количества быстро убывающих влево решений J^r).

Краевые задачи, в которых эти условия не выполняются, называются *вычислительно некорректными*, а величина C может достигать очень больших значений: $\sim O(e^{\Lambda_0 T})$.

10.2. Краевые задачи для уравнения второго порядка

Задачи этого класса часто встречаются в физических приложениях. Их также называют *краевыми задачами для уравнения Штурма–Лиувилля* (термин «задача Штурма–Лиувилля» обычно используется для спектральных краевых задач).

Классическая задача Штурма–Лиувилля может быть представлена в виде:

$$\begin{cases} \frac{d}{dx} \left(k(x) \frac{du}{dx} \right) + q(x) \frac{du}{dx} + g(x) \cdot u = f(x), & 0 < x < X; \\ a \frac{du}{dx} + bu = s, & x = 0; \\ a_1 \frac{du}{dx} + b_1 u = s_1, & x = X. \end{cases} \quad (10.7)$$

Введем расчетную сетку

$$\omega_n = \{x_n = hn \mid (n = 0, \dots, N), \quad h = X/N\}$$

и построим аппроксимирующее уравнение следующего вида:

$$\begin{cases} \frac{1}{h} \left(k_{n+1/2} \frac{u_{n+1} - u_n}{h} - k_{n-1/2} \frac{u_n - u_{n-1}}{h} \right) + \\ + q_n \frac{u_{n+1} - u_{n-1}}{2h} + g_n u_n = f_n, & n = 1, \dots, N-1; \\ a \frac{u_1 - u_0}{h} + b u_0 = s_0, & n = 0; \\ a_1 \frac{u_N - u_{N-1}}{h} + b_1 u_N = s_1, & n = N, \end{cases} \quad (10.8)$$

где $k_{n+1/2} = k \left(t_n + \frac{1}{2} h \right)$, $f_n = f(x_n)$, $q_n = q(x_n)$.

Разностная схема во внутренних точках может быть представлена в более компактном операторном виде:

$$h^{-2} (k_{n+1/2} \cdot \Delta_h^+ u_n - k_{n-1/2} \cdot \Delta_h^- u_n) + 2h^{-1} q_n (\Delta_h^+ + \Delta_h^-) u_n + g_n u_n = f_n,$$

где $\Delta_h^+ u_n = u_{n+1} - u_n$, $\Delta_h^- u_n = u_n - u_{n-1}$.

В случае $k(x) = k = \text{const}$, $q_n = 0$, (10.8) преобразуется к виду

$$k \frac{u_{n+1} - 2u_n + u_{n-1}}{h^2} = f_n, \quad n = 1, \dots, N-1,$$

или

$$k \Lambda_{xx} u_n = f_n,$$

с теми же краевыми условиями.

Для построения вычислительного алгоритма запишем систему линейных алгебраических уравнений (10.8) в следующем виде:

$$a_n u_{n-1} - b_n u_n + c_n u_{n+1} = d_n, \quad (10.9)$$

где a_n, b_n, c_n, d_n — локальные коэффициенты разностной схемы, для которых справедливы соотношения:

$$\begin{aligned} a_n &= \frac{1}{h^2} k_{n-1/2} - \frac{1}{2h} q_n, & c_n &= \frac{1}{h^2} k_{n+1/2} + \frac{1}{2h} q_n, \\ b_n &= a_n + c_n - q_n, & d_n &= f_n. \end{aligned} \quad (10.10)$$

Далее, для определенности, положим: $k > 0$, $g > 0$, что в большинстве случаев соответствует физике решаемых задач, и напомним о важном для численного решения систем линейных алгебраических уравнений условии диагонального преобладания [1]:

$$b_n > a_n + c_n, \quad (10.11)$$

которое иногда записывают в несколько ином виде [2]:

$$b_n \geq a_n + c_n + \delta, \quad \delta > 0. \quad (10.12)$$

Проведем аппроксимацию левого и правого краевых условий.
Левое краевое условие

$$a \frac{u_1 - u_0}{h} + bu_0 = d_0$$

может быть представлено в виде, аналогичном (10.9):

$$-b_0 u_0 + c_0 u_1 = d_0; \quad b_0 = \frac{a}{h} - b; \quad c_0 = \frac{a}{h} - b, \quad d_0 = s. \quad (10.13)$$

Аналогично и правое краевое условие

$$a_1 \frac{u_N - u_{N-1}}{h} + b_1 u_N = s_1$$

представимо в виде

$$a_N u_{N-1} - b_N u_N = d_N. \quad (10.14)$$

В дальнейшем для простоты изложения будем полагать: $a > 0$, $b < 0$, откуда $b_0 > c_0$ (эти неравенства соответствуют ряду прикладных задач: например, расчету процессов теплопроводности и диффузии).

Таким образом, получена система линейных алгебраических уравнений специфического вида (с матрицей трехдиагональной структуры):

$$\begin{cases} -b_0 u_0 + c_0 u_1 = d_0; \\ a_n u_{n-1} - b_n u_n + c_n u_{n+1} = d_n, \\ \quad \quad \quad n = 1, 2, \dots, N-1; \\ a_N u_{N-1} - b_N u_N = d_N, \end{cases} \quad (10.15)$$

или

$$\mathbf{A} \mathbf{u} = \mathbf{d}, \quad (10.16)$$

где \mathbf{A} — квадратная матрица $N \times N$, имеющая трехдиагональную (якобиеву) структуру:

$$\mathbf{A} = \begin{pmatrix} -b_0 & c_0 & & & \\ a_1 & -b_1 & c_1 & & 0 \\ & a_2 & -b_2 & c_2 & \\ 0 & a_n & -b_n & c_n & \\ & & a_{N-1} & -b_{N-1} & c_{N-1} \\ & & & a_N & -b_N \end{pmatrix}, \quad (10.17)$$

\mathbf{u} , \mathbf{d} — векторы-столбцы:

$$\mathbf{u} = \begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_N \end{pmatrix}; \quad \mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{pmatrix}.$$

Специфика таких, часто встречающихся в приложениях, систем линейных алгебраических уравнений, состоит в том, что их матрица \mathbf{A} имеет, как правило, высокий порядок ($N \gg 1$, $N = [T/h]$) и в основном состоит из нулей. Использовать для решения такого вида систем, например, стандартную программу для метода Гаусса «в лоб» было бы нерационально. Поэтому для таких систем был разработан метод прогонки (в американской терминологии — алгоритм Томаса), относящийся к классу экономичных алгоритмов, число которых пропорционально количеству уравнений в системе $O(N)$.

10.3. Метод прогонки

В методе прогонки решения системы линейных уравнений

$$\mathbf{A}\mathbf{u} = \mathbf{f} \quad (10.18)$$

будем искать в следующем виде:

$$u_{n-1} = P_n u_n + Q_n; \quad n = 1, \dots, N, \quad (10.19)$$

где P_n, Q_n — так называемые *прогоночные коэффициенты*. Очевидно, что после вычисления прогоночных коэффициентов система (10.15) с трехдиагональной матрицей преобразуется в систему линейных уравнений с двухдиагональной матрицей.

Далее, используя краевые условия и соотношение (10.19), мы находим вектор-столбец решений \mathbf{u} . Для этого приведем левое краевое условие к стандартной форме прогоночного соотношения

$$u_0 = P_1 u_1 - Q_1, \quad (10.20)$$

где $P_1 = c_0/b_0$, $Q_1 = -d_0/b_0$; в силу условия диагонального преобладания $P_1 < 1$. Далее необходимо получить рекуррентные соотношения для определения прогоночных коэффициентов P_2, P_3, \dots, P_N ; Q_2, Q_3, \dots, Q_N . Подставим прогоночное соотношение

$$u_{n-1} = P_n u_n + Q_n$$

в систему (10.9) для $n = 1, 2, \dots, N - 1$:

$$a_n u_{n-1} - b_n u_n + c_n u_{n+1} = d_n,$$

после чего получим

$$a_n (P_n u_n + Q_n) - b_n u_n + c_n u_{n+1} = d_n.$$

Полученное соотношение приводится к виду

$$u_n = P_{n+1} u_{n+1} + Q_{n+1}, \quad (10.21)$$

в котором прогоночные коэффициенты имеют вид

$$P_{n+1} = \frac{c_n}{b_n - a_n P_n}; \quad Q_{n+1} = \frac{a_n Q_n - d_n}{b_n - a_n P_n}, \quad (10.22)$$

а из рекуррентных соотношений (10.22) определяются все прогоночные коэффициенты P_n, Q_n ; $n = 0, 1, \dots, N-1$. Из последнего прогоночного уравнения

$$u_{N-1} = P_N u_N + Q_N$$

и правого краевого условия, или N -го прогоночного уравнения

$$a_N u_{N-1} - b_N u_N = d_N$$

вычисляется значение u_N . При этом говорят о *разрешении правого краевого условия*. Далее, исходя из прогоночного соотношения

$$u_{n-1} = P_n u_n + Q_n \quad (10.23)$$

справа налево вычисляются все компоненты вектора \mathbf{u} при $n = N, N-1, \dots, 1$. На этом алгоритм прогонки завершается.

Далее следует рассматривать вопросы обусловленности и устойчивости разностной задачи. Определение и достаточный признак обусловленности разностной краевой задачи вида

$$\begin{cases} a_n u_{n-1} - b_n u_n + c_n u_{n+1} = d_n, & n = 1, \dots, (N-1); \\ u_0 = \varphi_1, & u_N = \varphi_2 \end{cases} \quad (10.24)$$

были даны в [2].

Определение 10.1. Будем говорить, что разностная краевая задача (10.24) с коэффициентами a_n, b_n, c_n , ограниченными в совокупности

$$|a_n|, |b_n|, |c_n| < \widetilde{M},$$

хорошо обусловлена, если для всех достаточно больших N имеет одно и только одно решение $\{u_n\}$ при произвольных правых частях $\varphi_1, \varphi_2, \{d_n\}$ и если числа u_0, u_1, \dots, u_N , образующие решение, удовлетворяют оценке

$$|u_n| \leqslant \widetilde{M} \cdot \max \left\{ |\varphi_1|, |\varphi_2|, \max_n |d_n| \right\},$$

где \widetilde{M} — некоторое число, не зависящее от N .

Теорема 10.1. Если коэффициенты a_n, b_n, c_n удовлетворяют условию

$$|b_n| \geqslant |a_n| + |c_n| + \delta, \quad \delta > 0, \quad (10.25)$$

то задача (10.24) хорошо обусловлена [2].

Теорема 10.2 (Годунова–Рябенского; критерии хорошей обусловленности краевой задачи с постоянными коэффициентами) [2].
Для хорошей обусловленности краевой задачи

$$au_{n-1} + bu_n + cu_{n+1} = d_n, \quad 0 < n < N, \\ u_0 = \varphi, \quad u_N = \psi$$

с постоянными коэффициентами необходимо и достаточно, чтобы корни q_1 и q_2 характеристического уравнения

$$a + bq + cq^2 = 0$$

были по модулю один больше, другой меньше единицы:

$$|q_1| \geq 1 + \frac{\varepsilon}{2}, \\ |q_2^{-1}| \leq 1 - \frac{\varepsilon}{2}.$$

Теперь проведем исследование вычислительной устойчивости алгоритма прогонки, т. е. покажем, что вычислительные ошибки, обусловленные погрешностью округления машинных чисел, и ошибки, обусловленные машинными вычислениями, не приводят к существенным ошибкам в результатах расчетов.

Пусть ε_n — такая, например, машинная погрешность. Изучим, как она себя ведет в вычислительном процессе (соотношение δ_n и δ_{n+1}). «Машинная», т. е. реальная формула, по которой идет расчет первого прогоночного коэффициента в компьютере, имеет вид [1]

$$P_{n+1} + \varepsilon_{n+1} = \frac{c_n}{b_n - a_n(P_n + \varepsilon_n)} + \Delta_n, \quad (10.26)$$

где Δ_n — вычислительная ошибка, которая появляется при выполнении машинных операций в правой части, связанных с представлением коэффициентов a_n, b_n, c_n ; $P_{n+1}^M = P_{n+1} + \varepsilon_{n+1}$ (P_{n+1}^M — прогоночный коэффициент в машинном представлении). Погрешность ε_{n+1} состоит из наследственной погрешности, в которой суммируются предыдущие погрешности при вычислении $\varepsilon_n, \varepsilon_{n-1}, \dots$, и погрешности, появляющейся в результате одного вычислительного шага $\Delta_n \ll P_n$.

Линеаризация формулы (10.26) дает:

$$P_{n+1} + \varepsilon_{n+1} = \frac{c_n}{b_n - a_n P_n} + \frac{c_n a_n}{(b_n - a_n P_n)^2} \varepsilon_n + \Delta_n, \quad (10.27)$$

откуда следует соотношение для определения эволюции погрешности ε_n :

$$\varepsilon_{n+1} = \frac{a_n}{c_n} (P_{n+1})^2 \varepsilon_n + \Delta_n; \quad n = 0, 1, \dots, N-1, \quad (10.28)$$

поскольку $P_{n+1} = \frac{c_n}{b_n - a_n P_n}$.

Очевидно, что в реальном вычислительном процессе при разбиении отрезка интегрирования значение $N = T/h$ велико, поэтому нас интересует асимптотическое поведение вычислительной погрешности ε_n при $N \rightarrow \infty$.

Теорема 10.3. Пусть коэффициент $k(x)$ является липшиц-непрерывной функцией, локальные коэффициенты разностной схемы a_n, b_n, c_n, d_n удовлетворяют условию диагонального преобладания, $0 \leq P_1 \leq 1$ (для определенности будем полагать: $a_n, b_n, c_n, d_n > 0$).

Тогда:

1) для всех n выполняется

$$0 \leq P_{n+1} \leq 1;$$

2) $|a_n/c_n| \leq 1 + Ch$, $C \neq C(h)$, и

$$\varepsilon_n \leq e^{Cx} \varepsilon_0 + \frac{\Delta}{Ch} e^{Cx}, \quad |\Delta_n| \leq \Delta, \quad C \neq C(h), \quad Ch \ll 1.$$

Доказательство. Предположим, что $0 \leq P_1 \leq 1$. Тогда

$$P_{n+1} = \frac{c_n}{b_n - a_n P_n} \geq \frac{c_n}{b_n - a_n} > 0,$$

так как $b_n > a_n + c_n$; с другой стороны, получим

$$P_{n+1} = \frac{c_n}{b_n - a_n P_n} \leq \frac{c_n}{a_n + c_n - a_n P_n} = \frac{c_n}{c_n + a_n(1 - P_n)} \leq 1,$$

т. е., как и требовалось доказать,

$$\text{если } 0 \leq P_1 \leq 1, \text{ то } 0 \leq P_{n+1} \leq 1. \quad (10.29)$$

Далее:

$$\frac{a_n}{c_n} = \frac{k(x_n - h/2) - hq_n}{k(x_n + h/2) + hq_n} = 1 + O(h) = 1 + Ch; \quad C \neq C(h). \quad (10.30)$$

В таком случае, с учетом полученной оценки для P_{n+1} , получим

$$\varepsilon_{n+1} = \frac{a_n}{c_n} (P_{n+1})^2 \varepsilon_n + \Delta_n \leq (1 + Ch) \varepsilon_n + \Delta; \quad \Delta_n \leq \Delta. \quad (10.31)$$

Из (10.31), используя формулу для суммы геометрической прогрессии, будем иметь

$$\varepsilon_n \leqslant (1 + Ch)^n \varepsilon_0 + \frac{(1 + Ch)^n}{Ch} \Delta. \quad (10.32)$$

Для больших n и $Ch \ll 1$ получим ($nh = T$):

$$\varepsilon_n \leqslant e^{Cx} \varepsilon_0 + \frac{\Delta}{Ch} e^{Cx}, \quad (10.33)$$

что и требовалось доказать.

Отметим следующие особенности полученной оценки. Во втором слагаемом в знаменателе стоит малая величина h , однако, отношение Δ/h обычно мало: $C = O(1)$, поэтому поведение погрешности определяется множителем e^{CX} . При численном решении задач на больших отрезках $[0, X]$ величина $CX \gg 1$, в этом случае возможно появление вычислительных проблем. Однако при решении краевых задач довольно часто $X = O(1)$, что в соответствии с полученной оценкой δ_{n+1} вполне приемлемо. Аналогичные оценки можно провести и для второго прогоночного коэффициента Q_n .

Проанализируем эволюцию погрешности в алгоритм обратной прогонки:

$$u_{n-1} = P_n u_n + Q_n, \quad (10.34)$$

обозначив через ε_n погрешность при вычислении u_n , а машинное значение u_n — через u_n^M :

$$u_{n-1}^M = u_{n-1} + \varepsilon_{n-1} = (P_n \delta_n + \varepsilon_n) (u_n + \varepsilon_n) + (Q_n + \varepsilon_n) + \Delta_n, \quad (10.35)$$

где δ , ε_n — погрешности в вычислении прогоночных коэффициентов P_n и Q_n , Δ_n — суммарная вычислительная погрешность на $(n - 1)$ шагах в алгоритме обратной прогонки.

Из полученного соотношения (10.35) следует:

$$\varepsilon_{n-1} = P_n \varepsilon_n + (u_n \delta_n + \varepsilon_n + \Delta_n). \quad (10.36)$$

В последнем неравенстве главным является величина так называемого «параметра увеличения наследственной погрешности» P_n ; в соответствии с условием теоремы: $|P_n| \leqslant 1$. Случай $|P_n| > 1$ проблематичен, так как погрешность может иметь катастрофический рост, что недопустимо в вычислительной задаче.

Важным классом краевых задач являются задачи, связанные с определением точек спектра для уравнения Штурма–Лиувилля (спектральная задача Штурма–Лиувилля):

$$\frac{d}{dx} \left[k(x) \frac{du}{dx} \right] + g(x) \frac{du}{dx} = \lambda r(x) \cdot u, \quad x \in [0, X], \quad (10.37)$$

где λ — еще один параметр, краевые условия которого могут иметь, например, следующий вид:

$$u(0) = 0; \quad u(X) = U.$$

При некоторых значениях λ (точки спектра) у этого уравнения появляются нетривиальные решения (в остальных точках отрезка $[0, X]$ уравнение имеет тривиальные решения). Эти точки представляют основной интерес для различных приложений; λ подбирается так, чтобы выполнялось второе условие (решается задача Коши для начальных данных Коши: $u(0) = 0; u'_x = 1$), т. е. численно решаются несколько задач Коши, затем численно решается нелинейное уравнение $F(\lambda) = 0$.

10.4. Нелинейные краевые задачи для обыкновенных дифференциальных уравнений

Метод прогонки широко используется при численном линейных краевых задач, а также задач с переменными коэффициентами.

Оказывается, его можно использовать в совокупности с итерационными методами для численного решения нелинейных краевых задач. Рассмотрим в качестве примера следующую задачу:

$$\frac{d^2 u}{dx^2} = g(u), \quad u(0) = a, \quad u(X) = b, \quad 0 \leq x \leq X. \quad (10.38)$$

Ее разностная аппроксимация имеет вид

$$\Lambda u_n = g(u_n); \quad n = 1, 2, \dots, N-1; \quad h = X/N,$$

где

$$\Lambda u_n = \frac{u_{n-1} - 2u_n + u_{n+1}}{h^2}.$$

Вследствие нелинейности задачи (правая часть зависит от решения u), применить метод прогонки для ее решения нельзя.

Пусть $\psi(x_n) = u_n^0$ — начальное приближение решения краевой задачи (10.38). Построим итерационный процесс для численного решения разностной краевой задачи:

$$\frac{u_{n-1}^{i+1} - 2u_n^{i+1} + u_{n+1}^{i+1}}{h^2} = g(u^i), \quad u_n^0 = \psi_n, \quad (10.39)$$

$$i = 0, 1, \dots; \quad n = 1, 2, \dots, N-1; \quad u_0 = a, \quad u_N = b.$$

Его вычислительная реализация — метод прогонки на каждой итерации, т. е. при каждом $i = 0, 1, \dots$:

$$\begin{aligned} \frac{u_{n-1}^1 - 2u_n^1 + u_{n+1}^1}{h^2} &= g(u^0), \quad u_n^0 = \psi_n, \quad i = 0, \\ \frac{u_{n-1}^2 - 2u_n^2 + u_{n+1}^2}{h^2} &= g(u^1), \quad i = 1, \dots \end{aligned}$$

Вычислительный итерационный процесс можно заканчивать, например, по одному из двух критериев (или по обоим):

$$\begin{aligned} |u_n^{i+1} - u_n^i| &\leq \varepsilon; \\ \left| \frac{u_{n-1}^{i+1} - 2u_n^{i+1} + u_{n+1}^{i+1}}{h^2} - g(u^i) \right| &\leq \varepsilon, \end{aligned} \quad (10.40)$$

где ε — заданная точность.

Это метод является методом простых итераций в функциональном пространстве R^n , поскольку наша цель — найти вектор решения

$$u = \begin{Bmatrix} u_0 \\ u_1 \\ \vdots \\ u_N \end{Bmatrix},$$

т. е. функцию $u(x_n)$.

Однако итерационный процесс можно ускорить, применив идею линеаризации правой части $g(x)$ — аналог итерационного метода Ньютона (метод квазилинеаризации):

$$g(u_n^{i+1}) \approx g(u_n^i) + g'(u_n^i)(u_n^{i+1} - u_n^i).$$

В таком случае можно представить следующий итерационный процесс:

$$\begin{aligned} \frac{u_{n-1}^{i+1} - 2u_n^{i+1} + u_{n+1}^{i+1}}{h^2} &= g(u_n^i) + g'(u_n^i)(u_n^{i+1} - u_n^i), \\ u_n^0 &= \varphi_n, \quad i = 0, 1, \dots; \quad n = 1, 2, \dots, N-1; \\ u_0 &= a, \quad u_N = b, \end{aligned} \quad (10.41)$$

вычислительная реализация которого аналогична (10.38):

$$\begin{aligned} \frac{u_{n-1}^1 - 2u_n^1 + u_{n+1}^1}{h^2} &= g(u_n^0) + g'_u(u_n^0)(u_n^1 - u_n^0), \quad i = 0, \\ u_n^0 &= \varphi_n; \\ \frac{u_{n-1}^2 - 2u_n^2 + u_{n+1}^2}{h^2} &= g(u_n^1) + g'_u(u_n^1)(u_n^2 - u_n^1), \quad i = 1, \dots \end{aligned}$$

Критерием окончания итерационного процесса могут быть неравенства (10.40).

Возможно, первым методом численного решения нелинейных краевых задач для обыкновенных дифференциальных уравнений был метод стрельбы (или метод пристрелки), известный еще в артиллерии: стрельба по закрытым (например, холмами) мишеням.

Идея его состоит в следующем.

Предположим, что необходимо найти численное решение следующей краевой задачи:

$$\frac{d^2 u}{dx^2} = g(x, u); \quad x \in [0, T], \quad u(0) = a, \quad u(X) = b. \quad (10.42)$$

Зададим некоторое значение первой производной при $x = 0$:

$$u'(0) = \alpha_1$$

что, разумеется, делается не произвольно, а, например, из физических соображений. Построим численное решение задачи Коши u_h , например, методом Рунге–Кутты:

$$\frac{d^2 u}{dx^2} = g(x, u); \quad x \in [0, X], \quad u(0) = a, \quad u'(0) = \alpha_1, \quad (10.43)$$

и сравним его с точным значением $u(T) = b$:

$$\Delta_1 = |u_h(\alpha_1, X) - b|.$$

Если значение Δ_1 превышает заданную точность ε_1 , то мы ищем численное решение следующей задачи Коши:

$$\frac{d^2 u}{dx^2} = g(x, u); \quad x \in [0, T]; \quad u(0) = a, \quad u'(0) = \alpha_2$$

и вычисляем значение

$$\Delta_2 = |u_h(\alpha_2, X) - b|.$$

и т.д. Тем самым мы получаем серию значений $u(\alpha_i, X)$, $i = 1, 2, \dots$, после чего решаем, например, методом Ньютона нелинейное уравнение

$$F(\alpha) = u(\alpha, X) - b = 0,$$

из которого находим значение пристрелочного параметра α с заданной точностью:

$$|u(\alpha_i, X) - b| \leq \varepsilon,$$

после чего решаем задачу (10.40) и находим искомое решение $u(x)$, $x \in [0, X]$.

Однако отметим, что, в случае жестких краевых задач для ОДУ с наличием погранслоев и малого параметра при старшей производной, использование метода стрельбы может оказаться затруднительным. В качестве примера можно привести краевую задачу

$$\begin{cases} \varepsilon \frac{d^2 u}{dx^2} = u, & x \in [0, 1], \\ u(0) = 1, & u(1) = 2. \end{cases}$$

Точное решение этой задачи имеет вид

$$u(x, \varepsilon) = \left(e^{-1/\sqrt{\varepsilon}} - e^{1/\sqrt{\varepsilon}} \right)^{-1} \left[\left(e^{-1/\sqrt{\varepsilon}} - 2 \right) e^{x/\sqrt{\varepsilon}} + \left(2 - e^{1/\sqrt{\varepsilon}} \right) e^{-x} \right] \approx 2e^{-(1-x)/\sqrt{\varepsilon}} + e^{-x/\sqrt{\varepsilon}}.$$

В этом случае в решении задачи реализуются два пограничных слоя вблизи краев отрезка интегрирования, величина которых при малых ε может оказаться много меньше длины всего отрезка. По этой причине при численном решении таких задач необходимо либо выбирать малый шаг интегрирования, если мы хотим использовать методы, сводящие решение краевой задачи к численному решению нескольких задач Коши (например, метод стрельбы), либо выбирать большой шаг, игнорируя поведение решения в погранслоях. При этом необходимо использовать неявные разностные схемы.

10.5. Метод Фурье

Рассмотрим приближенное решение задачи

$$\frac{u_{n-1} - 2u_n + u_{n+1}}{h^2} = g; \quad n = 1, 2, \dots, (N-1),$$

$$u_0 = 0, \quad u_N = 0, \quad kN = X$$

методом Фурье.

Решение будем искать в виде разложения по базису из собственных функций разностного оператора

$$L_h(u) = h^{-2}(u_{n-1} - 2u_n + u_{n+1}).$$

Этот оператор имеет спектр из собственных значений

$$\lambda_k = \frac{4}{h^2} \sin^2 \frac{\pi kh}{2X}, \quad k = 1, 2, \dots, N-1,$$

и соответствующую ему полную ортонормированную систему из собственных векторов

$$\omega_k = \sqrt{\frac{2}{X}} \sin \frac{\pi k h}{2T}; \quad k = 1, 2, \dots, N-1; \quad t_n = h n,$$

что показывается прямой подстановкой в уравнение

$$L_h \omega_k = \lambda_k \omega_k.$$

Решение в таком случае ищется в виде разложения

$$u(x_n) = \sum_{k=1}^{N-1} C_k \omega_k(x_n); \quad n = 1, 2, \dots, N-1;$$

C_k — подлежащие определению коэффициенты Фурье.

Представим также правую часть разностного уравнения в виде фурье-разложения:

$$g_n = \sum_{k=1}^{N-1} \tilde{C}_k \omega_k(x_n), \quad \tilde{C}_k = \sum_{j=1}^{N-1} g_j \omega_k(x_j),$$

после чего подставим фурье-разложения искомой функции u_n и правой части g_n в исходное разностное уравнение:

$$L_h \left(\sum_{h=1}^{N-1} C_k \omega_k(x_n) \right) = \sum_{k=1}^{N-1} \tilde{C}_k \omega_k(x_n)$$

и учтем другое известное равенство

$$L_h \omega_k = \lambda_k \omega_k,$$

откуда получим выражение для коэффициентов Фурье:

$$\sum_{h=1}^{N-1} C_k [\lambda_k \omega_k(x_n)] = \sum_{h=1}^{N-1} \tilde{C}_k \omega_k(x_n), \quad C_k = \tilde{C}_k / \lambda_k$$

и приближенное решение исходного разностного уравнения:

$$u(x_n) = \sum_{h=1}^{N-1} \frac{\tilde{C}_k}{\lambda_k} \omega_k(x_n).$$

10.6. Методы Ритца и Галёркина

Рассмотрим функционал следующего вида:

$$J(u) = \int_0^1 G\left(x, u, \frac{du}{dx}\right) dx, \quad (10.44)$$

где G — функция, непрерывная по x , u , du/dx со своими производными (задача с закрепленными концами):

$$u(0) = u_0, \quad u(1) = u_1. \quad (10.45)$$

Назовем δ -окрестностью функции $u(x)$ семейство функций $\{v_i(x)\}$, удовлетворяющих на отрезке $[0, 1]$ неравенству

$$|u(x) - v_i(x)| \leq \delta. \quad (10.46)$$

Формулировка задачи вариационного исчисления в этом случае имеет следующий вид: найти экстремум функционала $J(u)$ среди функций, находящихся в δ -окрестности функции $u(x)$ при заданных краевых условиях [5].

Теорема 10.4. Пусть функция $u(x)$ принадлежит указанной δ -окрестности функции $u(x)$. Тогда эта функция доставляет экстремум функционалу $J(u)$ при заданных краевых условиях, если она удовлетворяет уравнению Эйлера:

$$\frac{dG}{du} - \frac{dG'u'}{dx} = 0. \quad (10.47)$$

Справедливо и обратное утверждение: если $u(x)$ является решением задачи (10.47), то она доставляет экстремум функционалу $J(u)$.

Доказательство. Рассмотрим некоторую функцию $g(x)$, удовлетворяющую краевым условиям

$$\eta(0) = \eta(1) = 0, \quad (10.48)$$

такую, что

$$u(x) = u[x + \xi \cdot g(x)], \quad (10.49)$$

где $0 < \xi \ll 1$ — малый параметр, т.е. функции из семейства $u_\xi(x)$ также принадлежат δ -окрестности функции $u(x)$.

Подставим семейство $u_\xi(x)$ в функционал $J(u)$:

$$J(u_\xi) = \int_0^1 G[x, u + \xi g(x), u'(x) + \xi g'(x)] dx, \quad (10.50)$$

и рассмотрим его как функцию от ξ : $J = J(\xi) = \Phi(\xi)$.

Вычислим первую и вторую функционала $J(\xi)$ как соответствующие производные функции $\Phi(\xi)$ в точке $\xi = 0$:

$$\delta J = \frac{\partial \Phi}{\partial \xi} \Big|_{\xi=0} = 0, \quad (10.51)$$

$$\delta^2 J = \frac{\partial^2 \Phi}{\partial \xi^2} \Big|_{\xi=0} = 0, \quad (10.52)$$

откуда следуют соотношения:

$$\delta J = \int_0^1 (G'_u g + G'_u g') dx, \quad (10.53)$$

$$\delta^2 J = \int_0^1 \left(G''_u (g')^2 + 2G''_u \cdot g \cdot g' + G''_u \cdot g^2 \right) dx. \quad (10.54)$$

После интегрирования (10.53) по частям и приравнивания к нулю значения первой производной $\Phi'(0) = 0$ (или $\delta J = 0$) получим

$$\delta J = \int_0^1 (G'_u g + G'_u g') dx = 0. \quad (10.55)$$

Поскольку функция $g(x)$ произвольна, то функция $u(x)$, доставляющая экстремум функционалу $J(u)$ и удовлетворяющая заданным краевым условиям, удовлетворяет уравнению Эйлера

$$\frac{dG}{du} - \frac{dG'_{u'}}{dx} = 0, \quad (10.56)$$

что и требовалось доказать.

Например, если $G(u)$ имеет следующий вид:

$$G(u) = u''_x + q(x)u^2 - 2f(x)u; \quad u(0) = u(1) = 0, \quad (10.57)$$

где $q > 0$, то

$$\frac{dG}{du} = 2qu - 2f; \quad \frac{dG'_{u'}}{dx} = 2u''_x$$

и уравнение Эйлера имеет вид

$$\frac{d^2 u}{dx^2} + q(x)u = f(x). \quad (10.58)$$

Можно сформулировать полученный результат следующим образом: если функция $u(x)$, принадлежащая области определения функционала

$$J(u) = \int_0^1 \left[\left(\frac{du}{dx} \right)^2 + qu^2 - 2fu \right] dx \quad (10.59)$$

и удовлетворяющая краевым условиям $u(0) = u(1) = 0$, доставляет экстремум данному функционалу, то она также является решением краевой задачи

$$\begin{aligned} - \left(\frac{d^2 u}{dx^2} \right) + q(x)u &= f(x); \\ u(0) = u(1) &= 0. \end{aligned}$$

В методе Ритца решение ищется в виде разложения по базису из линейно независимых, дважды непрерывно дифференцируемых по x функций $\{\psi_j\} \in C^2[0, 1]$; $j = 1, 2, \dots, n$:

$$u_n(x) = \sum_{j=1}^n b_j \psi_j, \quad (10.60)$$

где коэффициенты разложения b_j вычисляются из условия минимума функционала $J(u_n)$:

$$\frac{dJ(u_n)}{db_j} = 0; \quad j = 1, 2, \dots, n.$$

Поскольку

$$J(u_n) = \sum_{i=1}^n \sum_{j=1}^n b_i b_j B_{ij} - 2 \sum_{j=1}^n b_j f_j,$$

где

$$B_{ij} = \int_0^1 \left(\frac{\partial \psi_i}{\partial x} \cdot \frac{\partial \psi_j}{\partial x} + q \psi_i \psi_j \right) dx, \quad f_j = \int_0^1 f(x) \cdot \psi_j(x),$$

то в таком случае

$$\frac{dJ(u_n)}{db_i} = 2 \left(\sum_{j=1}^n B_{ij} b_j - f_i \right) = 0; \quad i = 1, 2, \dots, n,$$

откуда получаем систему линейных уравнений:

$$\sum_{j=1}^n B_{ij} b_j = f_i, \quad i = 1, 2, \dots, n. \quad (10.61)$$

Итак, метод Рунца заключается в следующем.

1. Выбирается базис $\{\psi_j\}$, $j = 1, 2, \dots, n$.
2. Решение ищется в виде

$$u_n(x) = \sum_{j=1}^n b_j \psi_j.$$

3. Коэффициенты b_j определяются из системы

$$(Bu_n, \psi_j) = (f, \psi_j); \quad j = 1, 2, \dots, n.$$

Эту же систему уравнений можно получить другим путем (метод Бубнова–Галёркина). Коэффициенты b_j из представления приближенного решения в виде

$$u_n(x) = \sum_{j=1}^n b_j \psi_j$$

будем искать, исходя из условия ортогональности невязки

$$r_n(x) = -(u_n)'' + qu_n - f(x)$$

каждой из базисных функций $\psi_j(x)$:

$$\int_0^1 r_n(x) \cdot \psi_j(x) dx = 0; \quad j = 1, \dots, n. \quad (10.62)$$

Последнее можно представить в виде

$$\sum_{j=1}^n \int_0^1 \left(-\frac{d^2 \psi_j}{dx^2} \psi_i + q \psi_i \psi_j \right) dx \cdot b_j = \int_0^1 f \psi_i dx, \quad i = 1, \dots, n. \quad (10.63)$$

С учетом равенства

$$\int_0^1 \left(-\frac{d^2 \psi_j}{dx^2} \right) \psi_i dx = \int_0^1 \frac{d\psi_j}{dx} \cdot \frac{d\psi_i}{dx} dx \quad i = 1, \dots, n,$$

система уравнений (10.63) может быть переписана следующим образом:

$$\sum_{j=1}^N \int_0^1 \left(\frac{d\psi_j}{dx} \cdot \frac{d\psi_i}{dx} + q \psi_j \cdot \psi_i \right) dx \cdot b_j = \int_0^1 f \psi_i dx, \quad i = 1, \dots, n. \quad (10.64)$$

Следовательно, проецируя невязку $r_n(x)$ на систему линейно независимых базисных функций ψ_i и приравнивая результаты к нулю, мы вновь получаем систему линейных алгебраических уравнений

$$\sum_{j=1}^n B_{ij} b_j = f_i, \quad i = 1, 2, \dots, n,$$

для коэффициентов разложения b_j , которые определяют приближенное решение нашей задачи, представленное в виде суммы

$$u_n(x) = \sum_{j=1}^n b_j \psi_j.$$

Этот метод можно представить в более общем виде. Пусть рассматривается задача, записанная в операторной форме:

$$Au = f, \quad u \in C^2[a, b]. \quad (10.65)$$

Коэффициенты разложения b_j находим из условия ортогональности невязки

$$r_n(x) = Au_n - f$$

некоторым, вообще говоря, отличным от ψ_i базисным функциям φ_i , т. е. исходя из условий

$$\int_0^1 (Au_n - f) \varphi_j dx = 0, \quad j = 1, \dots, n,$$

или же (что одно и то же) из системы уравнений вида

$$\sum_{i=1}^n b_i \int_0^1 \psi_j \left(-\frac{d^2 \psi_i}{dx^2} \psi_i + q \psi_i \right) dx = \int_0^1 f \psi_j dx, \quad j = 1, \dots, n. \quad (10.66)$$

Таким образом, система уравнений (10.66) получена методом проецирования невязки $r_n(x)$ на систему базисных функций $\{\psi_j\}$. В случае если $\varphi_i = \psi_j$, то вновь получаем систему (10.66).

Поскольку данный метод не является методом минимизации некоего функционала, то его называют *проеекционным*, или *проеекционно-сеточным* (метод Рунге называется *вариационным*).

Оказывается, что рассматриваемые методы давали удовлетворительную точность уже при не очень больших значениях n . Сначала использовались базисные функции с носителем, совпадающим со всей областью интегрирования, например:

$$\varphi_i(n) = x_i(1-x); \quad i = 1, \dots, n.$$

Однако использование таких базисных функций приводит к тому, что матрица \mathbf{B}_{ij} становится плотной, большинство ее элементов ненулевые. Следовательно, решать полученную систему при $n \geq 7$ уже затруднительно без использования вычислительной техники. Тем не менее если в качестве базисных функций выбирать так называемые *функции на конечных (финитных) носителях*, т. е. функции, отличные от нуля лишь на небольшой части области интегрирования, то можно добиться того, что матрица системы станет сильно разреженной, точнее, трехдиагональной. Такой подход получил название *метода конечных элементов*. Примером таких базисных финитных функций на отрезке $[0, 1]$ с сеткой $\omega_h = \{x_i = ih, i = 0, \dots, n, h = n^{-1}\}$ могут быть так называемые *функции-крышки*:

$$\psi_i(x) = h^{-1/2} \begin{cases} \frac{x - x_{i-1}}{h}, & x \in (x_{i-1}, x_i), \\ \frac{x_{i+1} - x}{h}, & x \in (x_i, x_{i+1}), \\ 0, & x \notin (x_{i-1}, x_{i+1}), \end{cases} \quad i = 1, \dots, n-1. \quad (10.67)$$

Будем искать решение в виде суммы

$$u_n(x) = \sum_{j=1}^{n-1} b_j \psi_j(x),$$

в которой коэффициенты b_j найдем для нашей краевой задачи с помощью метода Рунца (вариационного метода), т. е. из решения задачи минимизации функционала

$$J(u_n) = \int_0^1 \left[\left(\frac{du}{dx} \right)^2 + qu^2 - 2fu \right] dx,$$

из которой следует система уравнений

$$\sum_{j=1}^{n-1} B_{ij} b_j = f_i; \quad i = 1, \dots, (n-1).$$

После проведения алгебраических преобразований получим вид элементов B_{ij} трехдиагональной матрицы \mathbf{B} :

$$B_{ij} = \begin{cases} \frac{2}{h^2} + \frac{4}{6}, & i = j, \\ -\frac{1}{h^2} + \frac{1}{6}, & j = i-1, i+1, \\ 0, & j-i > 1. \end{cases} \quad (10.68)$$

В таком случае полученная система уравнений

$$\mathbf{B}\mathbf{u} = \mathbf{f}$$

приобретает следующий вид:

$$\begin{pmatrix} \frac{2}{h^2} + \frac{4}{6} & -\frac{1}{h^2} + \frac{1}{6} & 0 \dots & 0 & 0 \\ -\frac{1}{h^2} + \frac{1}{6} & \frac{2}{h^2} + \frac{4}{6} & -\frac{1}{h^2} + \frac{1}{6} & 0 \dots & 0 \\ 0 & -\frac{1}{h^2} + \frac{1}{6} & \frac{2}{h^2} + \frac{4}{6} & -\frac{1}{h^2} + \frac{1}{6} & 0 \dots 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots 0 & -\frac{1}{h^2} + \frac{1}{6} & \frac{2}{h^2} + \frac{4}{6} & \end{pmatrix} \mathbf{u} = \mathbf{f}, \quad (10.69)$$

где

$$f_j = \int_0^1 f \psi_j dx, \quad j = 1, \dots, n-1.$$

Метод численного решения подобных систем — прогонка.

Итак, вычислительный алгоритм, реализующий метод Бубнова–Галёркина, можно вкратце представить следующим образом.

1. Выбирается базис из линейно независимых функций $\{\psi_j\}$, $\psi_j \in H$.
2. Приближенное решение ищется в виде

$$u_n = \sum_{j=1}^n b_j \psi_j.$$

3. Коэффициенты b_j определяются из линейной системы уравнений.

Список литературы

1. Федоренко Р. П. Введение в вычислительную физику. Долгопрудный: Интеллект, 2008. 503 с.
2. Годунов С. К., Рябенкий В. С. Разностные схемы. М.: Наука, 1973. 400 с.

Дополнительная литература

3. Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. Численные методы. М.: ФИЗМАТЛИТ, 2000. 622 с.
4. Петров И. Б., Лобанов А. И. Лекции по вычислительной математике. М.: БИНОМ. Лаборатория знаний, 2006. 522 с.
5. Марчук Г. И., Агошков В. И. Введение в проекционно-сеточные методы. М.: Наука, 1981. 416 с.

Глава 11

ТОЧНЫЕ РЕШЕНИЯ РАЗНОСТНЫХ УРАВНЕНИЙ

В некоторых случаях разностные уравнения имеют точные решения. Знание таких решений имеет большое значение для исследования свойств разностных уравнений, сходимости к точным решениям.

Определение 11.1. *Линейным разностным уравнением порядка n называется уравнение вида*

$$\alpha_0 x_k + \alpha_1 x_{k+1} \dots + \alpha_n x_{k+n} = f_k, \quad (11.1)$$

или

$$\sum_{j=0}^n \alpha_j x_{k+j} = f_k,$$

где $x, \alpha_k, f_k \in R$ в случае одного уравнения и $x, f_k \in R^n$ в случае системы разностных уравнений, α_k — заданные постоянные коэффициенты, f_k — заданные функции (правые части разностного уравнения).

Для решения (11.1) необходимо задать начальные данные в виде

$$x_i = b_i, \quad i = 0, \dots, n-1. \quad (11.2)$$

Теорема 11.1. *Решение задачи Коши (11.1), (11.2) всегда существует, единственно и зависит от начальных значений $x_i, i = 0, \dots, n-1$.*

Будем искать нетривиальное решение однородного разностного уравнения

$$\alpha_0 x_k + \alpha_1 x_{k+1} \dots + \alpha_n x_{k+n} = 0 \quad (11.3)$$

в виде

$$x_k = \lambda^k. \quad (11.4)$$

Заметим, что тривиальное решение (11.3) ($x_k \equiv 0$) всегда существует.

Подставляя $x_k = \lambda^k$ в однородное разностное уравнение (11.3), получим характеристическое уравнение

$$\alpha_0 + \alpha_1 \lambda + \alpha_1 \lambda^2 + \dots + \alpha_n \lambda^n = 0. \quad (11.5)$$

Показывается [1], что если λ_j — вещественные попарно различные корни (11.3), то решения вида $\lambda_1^k, \dots, \lambda_n^k$ будут линейно

независимыми, а следовательно, они образуют фундаментальную систему решений однородного разностного уравнения (11.3).

В таком случае функция

$$x_k = \sum_{j=0}^n C_j \lambda_j^k,$$

где C_j — произвольные вещественные постоянные, будет решением линейного однородного разностного уравнения порядка n .

Теорема 11.2. Если характеристическое уравнение (11.5) имеет корень λ_0 кратности m ($1 \leq m \leq n$), то каждая из функций

$$x_k^j = k^j \lambda_0^k, \quad j = 0, \dots, m-1, \quad (11.6)$$

является решением однородного разностного уравнения.

Если λ_i ($i = 1, \dots, r$) — вещественные попарно различные корни кратностей q_1, \dots, q_r , то общее решение однородного разностного уравнения представимо в виде

$$\begin{aligned} x_k = & (C_{11} \lambda_1^k + C_{12} k \lambda_1^k + \dots + C_{1q_1} k^{q_1-1} \lambda_1^k) + \\ & + (C_{21} \lambda_2^k + C_{22} k \lambda_2^k + \dots + C_{2q_2} k^{q_2-1} \lambda_2^k) + \dots \\ & \dots + (C_{r1} \lambda_r^k + C_{r2} k \lambda_r^k + \dots + C_{rq_r} k^{q_r-1} \lambda_r^k), \end{aligned} \quad (11.7)$$

или

$$x_k = \sum_{n=1}^r \sum_{j=1}^{q_n} C_{nj} k^{q_j-1} \cdot \lambda_n^k.$$

Разностное уравнение (11.1) аппроксимирует обыкновенное дифференциальное уравнение n -го порядка

$$\alpha_0 x(t) + \alpha_1 x'_t(t) + \dots + \alpha_n x_t^{(n)} = f(t), \quad (11.8)$$

решение которого представляется в виде:

$$\begin{aligned} x(t) = & (C_{11} e^{\lambda_1 t} + C_{12} t e^{\lambda_1 t} + \dots + C_{1q_1} t^{q_1-1} e^{\lambda_1 t}) + \\ & + (C_{21} e^{\lambda_2 t} + C_{22} t e^{\lambda_2 t} + \dots + C_{2q_2} t^{q_2-1} e^{\lambda_2 t}) + \dots \\ & \dots + (C_{r1} e^{\lambda_r t} + C_{r2} t e^{\lambda_r t} + \dots + C_{rq_r} t^{q_r-1} e^{\lambda_r t}). \end{aligned}$$

При этом функции $x_k(t) = e^{\lambda_k t}$ являются нетривиальными решениями соответствующего однородного обыкновенного дифференциального уравнения. Подставив $x = e^{\lambda t}$ в однородное ОДУ, получим характеристическое уравнение

$$\sum_{j=0}^n \alpha_j \lambda^j = 0,$$

откуда находятся λ_i .

В случае вещественных попарно независимых собственных значений λ_i ($i = 0, \dots, n$) решение имеет простой вид:

$$x(t) = \sum_{j=0}^n c_j e^{\lambda_j t}.$$

Пример 1. Получим решение однородного разностного уравнения

$$x_{k+2} + x_{k+1} - 2x_k = 0.$$

Характеристическое уравнение имеет вид (ищем решение в виде $u_k = \lambda^k$)

$$\lambda^2 + \lambda - 2 = 0;$$

его корни $\lambda_1 = -2$, $\lambda_2 = 1$; тогда общее решение будет:

$$x_k = C_1 (-2)^k + C_2;$$

C_1, C_2 — произвольные постоянные.

Пример 2. Решить уравнение

$$x_{k+3} - x_{k+2} - x_{k+1} + x_k = 0.$$

Характеристическое уравнение:

$$\begin{aligned} \lambda^3 - \lambda^2 - \lambda + 1 &= \lambda^2(\lambda - 1) - (\lambda - 1) = (\lambda - 1)(\lambda^2 - 1) = \\ &= (\lambda - 1)^2(\lambda + 1) = 0, \end{aligned}$$

корни которого: $\lambda_1 = 1$ (кратности 2) и $\lambda_2 = -1$.

Общее решение представляется в виде

$$x_k = C_1 + C_2 k + C_3 (-1)^k;$$

C_1, C_2, C_3 — произвольные постоянные.

Приведем примеры линейных разностных уравнений:

$$\alpha x_k = f_k, \quad k = 0, 1, \dots \quad (11.9)$$

— уравнение нулевого порядка;

$$\alpha x_k + \beta x_{k+1} = f_k \quad (11.10)$$

— уравнение первого порядка;

$$\alpha x_{k-1} + \beta x_k + \gamma x_{k+1} = f_k \quad (11.11)$$

— уравнение второго порядка.

Рассмотрим решение уравнения первого порядка (11.10). Пусть X_k — частное решение соответствующего однородного

уравнения, удовлетворяющего начальному условию $U_0 = 1$; тогда любое его решение представляется в виде

$$x_k = C X_k,$$

где C — произвольная постоянная.

Общее решение неоднородного уравнения (11.10) складывается из частного решения неоднородного x_k^* и общего решения однородного уравнений при начальном условии $U_0 = 1$:

$$x_k = x_k^* + C \left(-\frac{a}{b}\right)^k,$$

так как $X_k = (-a/b)^k$. Поскольку $\lambda = -a/b$ — решение характеристического уравнения

$$\alpha + \beta \lambda = 0,$$

то $X_k = \lambda^k = (-a/b)^k$.

Частное решение этого разностного уравнения представляется в виде

$$x_k^* = \sum_{k=-\infty}^{\infty} G_{n-k} \cdot f_k, \quad (11.12)$$

где G_{n-k} (функция Грина) — фундаментальное решение неоднородного уравнения, находящееся из уравнения

$$\alpha G_k + \beta G_{k+1} = \delta_0^k,$$

или

$$\begin{cases} \alpha G_0 + \beta G_1 = 1, \\ \alpha G_k + \beta G_{k+1} = 0, \quad k > 0. \end{cases}$$

Представим частное решение неоднородного уравнения в окончательном виде (подробно эта задача рассматривается в [3]):

$$G_{n-k} = \begin{cases} A \left(-\frac{a}{b}\right)^n, & n \leq k; \\ \left(A - \frac{a}{b}\right) \left(-\frac{a}{b}\right)^n, & n \geq k+1. \end{cases}$$

Теорема 11.3. Пусть: $|a/b| \neq 1$; G_n — ограниченное фундаментальное решение; правая часть f_k ограничена по модулю: $|f_k| \leq F$.

Тогда ряд

$$u_n = \sum_{k=-\infty}^{\infty} G_{n-k} \cdot f_k,$$

сходится.

Общее решение неоднородного уравнения второго порядка имеет вид

$$x_k = x_k^* + \alpha X_k + \beta Y_k,$$

где x_k^* — частное решение (11.11), X_k, Y_k — частные решения соответствующего однородного уравнения, общее решение которого есть

$$x_k = \alpha X_k + \beta Y_k,$$

с начальными данными:

$$\begin{cases} X_0 = 1, \\ Y_0 = 0; \end{cases} \quad \begin{cases} X_1 = 0, \\ Y_1 = 1. \end{cases}$$

Подробно решения разностных уравнений (11.10) и (11.11) рассмотрены в [3].

Для уравнения второго порядка (11.11) характеристическое уравнение будет иметь вид

$$\alpha + \beta\lambda + \gamma\lambda^2 = 0,$$

причем его корни могут быть различными или кратными. В первом случае X_n и Y_n имеют вид

$$\begin{cases} X_n = \frac{q_2}{q_2 - q_1} q_1^n - \frac{q_1}{q_2 - q_1} q_2^n, \\ Y_n = \frac{1}{q_2 - q_1} q_1^n + \frac{1}{q_2 - q_1} q_2^n, \end{cases}$$

во втором:

$$\begin{cases} X_n = q_1^n - nq_1^n, \\ Y_n = \frac{1}{q_1} nq_1^n = nq_1^{n-1}. \end{cases}$$

В случае комплексных корней q_1 и q_2 рассматриваемого характеристического уравнения общее решение однородного разностного уравнения (11.11) имеет вид

$$u_n = a_1 \left(\sqrt{\frac{\alpha}{\gamma}} \right)^n \cos(n\varphi) + a_2 \left(\sqrt{\frac{\alpha}{\gamma}} \right)^n \sin(n\varphi),$$

где a_1, a_2 — произвольные постоянные.

Общее решение неоднородного разностного уравнения (11.11) находится по формуле

$$u_n^* = \sum_{k=-\infty}^{\infty} G_{n-k} \cdot f_k,$$

где G_n является фундаментальным решением (11.11) с правой частью вида

$$f_k = \delta_0 = \begin{cases} 0, & n \neq 0, \\ 1, & n = 0. \end{cases}$$

Более подробно такое уравнение представляется в виде:

$$\begin{cases} \alpha G_{-1} + \beta G_n + \gamma G_{n+1} = 0, & n \leq -1, \\ \alpha G_{n-1} + \beta G_0 + \gamma G_1 = 1, \\ \alpha G_{n-1} + \beta G_n + \gamma G_{n+1}, & n \geq 1. \end{cases}$$

Ограниченное фундаментальное решение этого уравнения имеет следующий вид [3]:

$$G_n = \begin{cases} (\alpha q_2^{-1} + \beta + \gamma q_1)^{-1} q_2^n, & n \leq 0, \\ (\alpha q_2^{-1} + \beta + \gamma q_1)^{-1} q_1^n, & n \geq 0; |q_1| < 1, |q_2| > 1; \end{cases}$$

$$G_n = \begin{cases} 0, & n \leq 0, \\ \frac{q_1^n - q_2^n}{\gamma (q_1 - q_2)}, & n \geq 0; |q_1| < 1, |q_2| < 1; \end{cases}$$

$$G_n = \begin{cases} (\alpha q_1^{-1} + \beta + \gamma q_2)^{-1} q_1^n, & n \leq 0, \\ (\alpha q_1^{-1} + \beta + \gamma q_2)^{-1} q_2^n, & n \geq 0; |q_1| > 1, |q_2| < 1; \end{cases}$$

$$G_n = \begin{cases} 0, & n \leq 0, \\ \frac{q_1^n - q_2^n}{\gamma (q_1 - q_2)}, & n \geq 0; |q_1| > 1, |q_2| > 1. \end{cases}$$

Напомним, что условие корней является условием устойчивости решений разностных уравнений (см. гл. 9).

Общее решение линейного неоднородного разностного уравнения с постоянными коэффициентами

$$\alpha_0 x_k + \alpha_1 x_{k+1} + \dots + \alpha_n x_{k+n} = f_{k+n}, \quad k = 0, 1, \dots, \quad (11.13)$$

также ищется в виде

$$x_k = x_k^* + X_k,$$

где x_k^* — частное решение неоднородного уравнения, X_k — общее решение соответствующего однородного уравнения.

Оказывается, что для некоторых конкретных видов правых частей неоднородного уравнения (11.13) его решение представляется достаточно просто в аналитическом виде.

Доказывается теорема о том, что если правая часть (11.13) имеет вид [1]

$$f_k = \alpha^k [P_m(k) \cos(\beta k) + Q_n(k) \sin(\beta k)], \quad (11.14)$$

где $P_m(k)$ и $Q_n(k)$ — полиномы степеней m и n , a и b — некоторые постоянные, то частное решение имеет вид

$$u_k^* = k^s \cdot \alpha^k [\tilde{P}_l(k) \cos(\beta k) + \tilde{Q}_l(k) \sin(\beta k)], \quad (11.15)$$

где $l = \max\{m, n\}$ — степень полиномов $\tilde{P}_l(k)$, $\tilde{Q}_l(k)$; $s = 0$, если α не является корнем характеристического уравнения кратности q .

В случае системы линейных разностных уравнений:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k; \quad \mathbf{x}_k \in R^n, \quad \mathbf{A} \in M(n \times n), \quad (11.16)$$

нетривиальное решение ищется в следующем виде:

$$\mathbf{x}_k = \lambda^k \boldsymbol{\omega}, \quad (11.17)$$

где $\lambda \neq 0$, $\boldsymbol{\omega}$ — ненулевой вектор с n компонентами.

После подстановки \mathbf{x}_k в систему (11.16) получим систему линейных алгебраических уравнений

$$\mathbf{A}\boldsymbol{\omega} = \lambda\boldsymbol{\omega},$$

откуда следует, что λ — собственное значение, $\boldsymbol{\omega}$ — соответствующий ему собственный вектор матрицы \mathbf{A} , причем λ определяется из уравнения

$$|\mathbf{A} - \lambda\mathbf{E}| = 0.$$

Теорема 11.4. Если в R^n существует базис из собственных векторов $\boldsymbol{\omega}_i$, $i = 1, \dots, n$, матрицы \mathbf{A} , а λ_i — соответствующие им попарно независимые собственные значения матрицы \mathbf{A} , то общее решение системы (11.16) представимо в виде

$$\mathbf{x}_k = \sum_{i=1}^n \gamma_i \lambda_i^k \boldsymbol{\omega}_i, \quad (11.18)$$

где γ_i — произвольные постоянные.

В случае неоднородной системы разностных уравнений

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{f}_k \quad (11.19)$$

общее решение ищется в виде

$$\mathbf{x}_k = \mathbf{A} \mathbf{X}_k^* + \alpha \mathbf{X}_k, \quad (11.20)$$

где \mathbf{X}_k — общее решение соответствующей однородной системы, \mathbf{X}_k^* — частное решение (11.19), α — константа.

Если правая часть неоднородной системы имеет вид

$$\mathbf{f}_k = \mu^k \cdot \mathbf{P}_l(k), \quad (11.21)$$

где μ — ненулевое вещественное число, не являющееся собственным значением матрицы \mathbf{A} , $\mathbf{P}_l(k)$ — полином степени l , коэффициентами которого являются вещественные n -мерные векторы, то частное решение (11.19) представляется в следующем виде:

$$\mathbf{x}_k^* = \mu^k \cdot \tilde{\mathbf{P}}_l(k),$$

где $\tilde{\mathbf{P}}_l(k)$ — полином степени l , коэффициенты которого (n -мерные векторы) находятся путем подстановки \mathbf{x}_k^* в систему (11.19).

Пример. Решить неоднородную систему линейных разностных уравнений:

$$\begin{cases} x_{k+1} = x_k - y_k + 3^k, \\ y_{k+1} = -2x_k - 3^k. \end{cases}$$

Решение. Собственные числа и собственные векторы матрицы

$$\mathbf{A} = \begin{Bmatrix} 1 & -1 \\ -2 & 0 \end{Bmatrix}$$

имеют следующие значения:

$$\lambda_1 = -1, \quad \lambda_2 = 2, \quad \omega_1 = \begin{Bmatrix} 1 \\ 2 \end{Bmatrix}, \quad \omega_2 = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}.$$

В таком случае общее решение данной системы имеет вид

$$\begin{Bmatrix} x \\ y \end{Bmatrix} = C_1 (-1)^k + C_2 \cdot 2^k \begin{Bmatrix} 1 \\ -1 \end{Bmatrix} + \begin{Bmatrix} x \\ y \end{Bmatrix}^*,$$

где x^*, y^* — частное решение системы, которое ищется в виде

$$\begin{Bmatrix} x \\ y \end{Bmatrix}^* = 3^k \begin{Bmatrix} a \\ b \end{Bmatrix},$$

причем a и b находятся после подстановки x^*, y^* в исходную систему:

$$\begin{cases} 3^{k+1} \cdot a = 3^k \cdot a - 3^k \cdot b + 3^k, \\ 3^k \cdot b = -2 \cdot 3^k \cdot a - 3^k, \end{cases}$$

откуда получим:

$$\begin{cases} 2a + b = 1, \\ 2a + 3b = -1, \end{cases} \quad \text{или:} \quad a = 1, \quad b = -1.$$

Тогда

$$\begin{Bmatrix} x \\ y \end{Bmatrix}^k = C_1 \cdot (-1)^k \cdot \begin{Bmatrix} 1 \\ 2 \end{Bmatrix} + C_2 \cdot 2^k \cdot \begin{Bmatrix} 1 \\ -1 \end{Bmatrix} + 3^k \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}.$$

Рассмотрим случай кратных собственных значений матрицы \mathbf{A} . В этом случае необходимо использовать жордановы цепочки векторов.

Определение 11.2. Векторы $\omega_2, \dots, \omega_r$ называются *присоединенными к собственному вектору ω_1* матрицы \mathbf{A} , если выполняются соотношения:

$$\mathbf{A}\omega_2 = \lambda\omega_2 + \omega_1, \quad \mathbf{A}\omega_3 = \lambda\omega_3 + \omega_2, \dots, \mathbf{A}\omega_r = \lambda\omega_r + \omega_{r-1}.$$

Такая система векторов $\omega_2, \dots, \omega_r$ называется *жордановой цепочкой для собственного значения λ* матрицы \mathbf{A} , r — длина цепочки.

В случае если λ — собственное значение кратности 1, то жорданова цепочка состоит из одного собственного вектора; если же λ — кратное (кратность больше 1), то для него могут быть несколько жордановых цепочек; если λ — комплексное число, то жордановы цепочки также комплексные.

Теорема 11.5. В комплексном n -мерном линейном пространстве всегда существует базис, составленный из жордановых цепочек для всех собственных значений λ матрицы \mathbf{A} . Если все λ являются вещественными, то и все жордановы цепочки вещественные, базис из этих цепочек также вещественный.

Общее решение линейной однородной системы разностных уравнений имеет вид

$$\mathbf{x}_k = \sum_{j=1}^s [\gamma_{1j} \lambda_j^k \omega_{1j} + \gamma_{2j} (\lambda_j^k \omega_{2j} + C_k^1 \lambda_j^{k-1} \omega_{1j}) + \dots \\ \dots + \gamma_{r_j,j} (\lambda_j^k \omega_{r_j,j} + C_k^1 \lambda_j^{k-1} \omega_{r_j-1,j} + \dots + C_k^{r_j-1} \lambda_j^{k-r_j+1} \omega_{1j})],$$

где γ_{ij} — произвольные постоянные,

$$C_k^m = \frac{k!}{(k-m)!m!}, \quad m = 1, \dots, r-1;$$

$\omega_{1j}, \omega_{2j}, \dots, \omega_{r_j, j}$ — жорданов базис из s жордановых цепочек длины r_j , соответствующих собственным значениям λ_j ; j — номер собственного числа, $j = 1, \dots, s$; $\sum_{j=1}^s r_j = n$.

В случае если $n = 2$, то матрица \mathbf{A} имеет собственное значение λ кратности 2, которому соответствует жорданова цепочка из собственного вектора ω_1 , ее решение рассматриваемой однородной системы разностных уравнений имеет вид

$$\mathbf{x}_k = \gamma_1 \lambda^k \omega_1 + \gamma_2 (\lambda^k \omega_2 + k \lambda^{k-1} \omega_1).$$

Если матрица \mathbf{A} имеет собственное значение кратности 3, которому соответствует жорданова цепочка из собственных векторов ω_2, ω_3 , то общее решение рассматриваемой системы будет

$$\mathbf{x}_k = \gamma_1 \lambda^k \omega_1 + \gamma_2 (\lambda^k \omega_2 + k \lambda^{k-1} \omega_1) + \gamma_3 \left(\lambda^k \omega_3 + k \lambda^{k-1} \omega_2 + \frac{k(k-1)}{2} \lambda^{k-2} \omega_1 \right).$$

Пример 1. Найти общее решение системы разностных уравнений

$$\begin{cases} x_{k+1} = -x_k - 5y_k + z_k, \\ y_{k+1} = -x_k + 3y_k - z_k, \\ z_{k+1} = 4x_k + 5y_k + 2z_k. \end{cases}$$

Собственные числа матрицы $\mathbf{A} = \begin{Bmatrix} -1 & -5 & 1 \\ -1 & 3 & -1 \\ 4 & 5 & 2 \end{Bmatrix}$: $\lambda_1 = -2$; $\lambda_2 = 3$ (кратность 2); собственный вектор, соответствующий λ_1 :

$$\omega_1 = \begin{Bmatrix} -1 \\ 0 \\ 1 \end{Bmatrix}.$$

Собственному значению λ_1 соответствует жорданова цепочка из вектора ω_2 и присоединенного вектора ω_3 :

$$\omega_2 = \begin{Bmatrix} -1 \\ 1 \\ 1 \end{Bmatrix}, \quad \omega_3 = \begin{Bmatrix} -1 \\ 1 \\ 0 \end{Bmatrix}.$$

Тогда общее решение рассматриваемой системы будет иметь вид

$$\begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \gamma_1 (-2)^k \begin{Bmatrix} -1 \\ 0 \\ 1 \end{Bmatrix} + \gamma_2 \cdot 3^k \begin{Bmatrix} -1 \\ 1 \\ 1 \end{Bmatrix} + \gamma_3 \left(3^k \begin{Bmatrix} -1 \\ 1 \\ 0 \end{Bmatrix} + k \cdot 3^{k-1} \begin{Bmatrix} -1 \\ 1 \\ 1 \end{Bmatrix} \right).$$

Пусть характеристическое уравнение имеет комплексные кратные корни. Построим для каждого вещественного корня λ кратности p функции вида

$$\lambda^k, \quad k\lambda^k, \quad \dots, \quad k^{p-1}\lambda^k;$$

для каждого комплексного корня кратности m — функции

$$\lambda = |\lambda| \cdot (\cos \varphi + i \sin \varphi), \quad 0 \leq \varphi \leq 2\pi;$$

для сопряженного к нему корня: $\bar{\lambda} = |\lambda| \cdot (\cos \varphi - i \sin \varphi)$.

В этом случае совокупность функций

$$\begin{aligned} |\lambda|^k \cos k\varphi, \quad k|\lambda|^k \cos k\varphi, \quad \dots, \quad k^{m-1}|\lambda|^k \cos k\varphi, \\ |\lambda|^k \sin k\varphi, \quad k|\lambda|^k \sin k\varphi, \quad \dots, \quad k^{m-1}|\lambda|^k \sin k\varphi \end{aligned}$$

образует вещественную фундаментальную систему решений разностного линейного однородного уравнения.

Пример 2. Рассмотрим разностное уравнение вида

$$u_k + u_{k+2} = 0,$$

характеристическое уравнение для которого

$$\lambda^2 + 1 = 0$$

имеет корни $\lambda = \pm i$, а его общее решение записывается в виде

$$u_k = \bar{c}_1 i^k + \bar{c}_2 (-i)^k,$$

где \bar{c}_1 и \bar{c}_2 — произвольные комплексные постоянные.

Поскольку

$$\lambda = i = \cos \frac{\pi}{2} + i \sin \frac{\pi}{2}, \quad \lambda^k = i^k = \cos \frac{k\pi}{2} + i \sin \frac{k\pi}{2},$$

то общее решение данного уравнения имеет вид

$$u_k = c_1 \cos \frac{k\pi}{2} + c_2 \sin \frac{k\pi}{2},$$

где c_1 и c_2 — произвольные вещественные постоянные.

Список литературы

1. Романко В. К. Курс разностных уравнений. М.: ФИЗМАТЛИТ, 2012. 199 с.
2. Самарский А. А., Гулин А. В. Численные методы. М.: Наука, 1989. 430 с.
3. Годунов С. К., Рябенький В. С. Разностные схемы. М.: Наука, 1973. 400 с.

ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ РАЗНОСТНЫХ СХЕМ

12.1. Сходимость, аппроксимация и устойчивость методов

Важнейшие вопросы при аппроксимации ОДУ разностными уравнениями связаны с понятиями сходимости, аппроксимации и устойчивости численных решений. Для ответов на них введем операторные обозначения дифференциальных и разностных уравнений. Дифференциальные уравнения будем обозначать следующим образом:

$$Pu = f. \quad (12.1)$$

Здесь P — обозначение дифференциального оператора, $u, f \in R^n$ — решение и правая часть ОДУ соответственно. Для обозначения разностного уравнения будем использовать следующее операторное уравнение [1, 2]:

$$P_\tau u_\tau = f_\tau, \quad (12.2)$$

где $P_\tau \in S$ — разностный оператор, принадлежащий пространству линейных операторов S ; $u_\tau, f_\tau \in R^n$ — соответственно решение и правая часть разностного уравнения. Строго говоря, (12.2) является параметрическим семейством разностных уравнений.

Пусть, например, дифференциальное уравнение имеет вид

$$\begin{aligned} \frac{du}{dt} &= \ln t, \quad t > 0; \\ u(0) &= 1, \quad t = 0. \end{aligned}$$

В этом случае его операторная запись будет такова:

$$\begin{aligned} Pu &= \begin{cases} \frac{du}{dt}, & t > 0, \\ u(0), & t = 0, \end{cases} \\ f &= \begin{cases} \ln t, & t > 0, \\ 1, & t = 0. \end{cases} \end{aligned}$$

Допустим, мы аппроксимируем это ОДУ следующей разностной схемой:

$$\begin{aligned} \frac{u_{n+1} - u_n}{\tau} &= \ln t_n, \quad n = 1, 2, \dots, N; \\ u_0 &= 1, \quad n = 0. \end{aligned}$$

Соответствующее операторное уравнение имеет следующий вид:

$$P_\tau u_\tau = \begin{cases} \frac{u_{n+1} - u_n}{\tau}, & n = 1, 2, \dots, N, \\ u_0, & n = 0; \end{cases}$$

$$f_\tau = \begin{cases} \ln t_n, & n = 1, 2, \dots, N \\ 1, & n = 0. \end{cases}$$

Приведем основополагающие определения в теории разностных схем.

Определение 12.1. Решение u_τ разностной задачи

$$P_\tau u_\tau = f_\tau$$

сходится к решению U дифференциальной задачи

$$PU = f,$$

если выполняется условие

$$\|u_\tau - U\| \xrightarrow{\tau \rightarrow 0} 0.$$

Если, кроме этого, имеет место неравенство

$$\|u_\tau - U\| \leq C\tau^p,$$

где $p > 0$, $C \neq C(\tau)$, $C > 0$, то говорят о *сходимости p -го порядка*.

Для определения аппроксимации введем понятие *невязки*:

$$\xi_\tau = P_\tau U_\tau - f_\tau,$$

где ξ_τ — невязка, U_τ — проекция точного решения на расчетную сетку.

Определение 12.2. Разностная схема

$$P_\tau u_\tau = f_\tau$$

аппроксимирует дифференциальное уравнение

$$Pu = f$$

на решении u , если выполняется

$$\|\xi_\tau\| \xrightarrow{\tau \rightarrow 0} 0.$$

Если, кроме этого, имеет место неравенство

$$\|\xi_\tau\| \leq C_1\tau^p,$$

где $p > 0$, $C_1 \neq C_1(\tau)$, $C_1 > 0$ — некоторые постоянные, то имеет место *аппроксимация p -го порядка*.

Определение 12.3 [2]. Разностная задача

$$P_\tau u_\tau = f_\tau$$

называется *устойчивой*, если ее решение существует и единственно, причем из соотношений

$$P_\tau u_\tau = f_\tau + \eta'_\tau,$$

$$P_\tau v_\tau = f_\tau + \eta_\tau,$$

где η'_τ , η_τ — малые возмущения правой части f_τ , следует неравенство вида

$$\|u_\tau - v_\tau\| \leq C_2 (\|\eta'_\tau\| + \|\eta_\tau\|),$$

где $C_2 \neq C_2(\tau)$, $C_2 > 0$.

Понятия сходимости, аппроксимации и устойчивости связывает теорема Рябенского–Лакса.

Теорема 12.1 (Рябенского–Лакса) [1]. Пусть разностная схема

$$P_\tau u_\tau = f_\tau$$

аппроксимирует дифференциальную задачу

$$Pu = f$$

на решении u с порядком p и устойчива. Тогда решение u_τ разностной задачи

$$P_\tau u_\tau = f_\tau$$

сходится к решению дифференциальной задачи

$$Pu = f,$$

причем имеет место оценка:

$$\|u_\tau - U_\tau\| \leq (C_1 C_2) \tau^p = C \tau^p,$$

где U_τ — проекция точного решения на расчетную сетку.

Доказательство. Из определения устойчивости разностной задачи следует: для двух близких решений x_τ, y_τ разностных уравнений

$$\begin{aligned} P_\tau u_\tau &= f_\tau + \eta_\tau, \\ P_\tau v_\tau &= f_\tau + \eta'_\tau \end{aligned} \quad (12.3)$$

выполняется

$$\|u_\tau - v_\tau\| \leq C_2 (\|\eta_\tau\| + \|\eta'_\tau\|). \quad (12.4)$$

Положим, что u_τ есть точное решение разностного уравнения

$$P_\tau u_\tau = f_\tau.$$

В этом случае в (12.3) $\eta_\tau = 0$.

Также положим, что $v_\tau = U_\tau$ есть проекция точного решения дифференциальной задачи. Тогда $\eta_\tau = \xi_\tau$. Значит, (12.3) при этих предположениях можно переписать в виде:

$$\begin{aligned} P_\tau u_\tau &= f_\tau, \\ P_\tau U_\tau &= f_\tau + \xi_\tau. \end{aligned}$$

В этом случае получим:

$$\|u_\tau - U_\tau\| \leq C_2 \|\xi_\tau\| \leq C_2 C_1 \tau^p = C \tau^p.$$

Теорема доказана.

Приведем еще одно определение устойчивости [1], эквивалентное предыдущему.

Определение 12.4 [1]. Разностная схема

$$P_\tau u_\tau = f_\tau$$

устойчива, если существуют постоянные $\tau_0 > 0$ и $\delta > 0$ такие, что при любом $\tau < \tau_0$ и любом ε_τ ($\|\varepsilon_\tau\| < \delta$) «возмущенная» разностная задача

$$P_\tau v_\tau = f_\tau + \varepsilon_\tau$$

имеет единственное решение v_τ , причем имеет место неравенство

$$\|v_\tau - u_\tau\| \leq C \|\varepsilon_\tau\|,$$

где $C \neq C(\tau)$, $C > 0$.

Последнее неравенство означает, что малое возмущение ε_τ правой части рассматриваемой разностной задачи вызывает равномерное относительно τ малое возмущение решения.

В теории разностных схем также вводится понятие корректной задачи.

Определение 12.5 [7]. Семейство разностных уравнений

$$P_\tau u_\tau = f_\tau$$

называется *корректным*, если его решение существует и единственно при любых правых частях f_τ , а также существует постоянная $C \neq C(\tau) > 0$ такая, что при любых f_τ выполняется оценка

$$\|u_\tau\| \leq C \|f_\tau\|.$$

Заметим, что первое условие эквивалентно существованию обратного оператора P_τ^{-1} , а второе — равномерной по τ ограниченности этого оператора (постоянная C является универсальной для всего семейства уравнений).

Также отметим, что условие

$$\|u_\tau\| \leq C \|f_\tau\|$$

означает непрерывную равномерную по τ зависимость решения разностной задачи от правой части, а также является вторым определением устойчивости задачи. Оно получается из условия равномерной по τ ограниченности оператора P_τ^{-1} : из $\|P_\tau^{-1}\| \leq C$ следует

$$\|u_\tau\| \leq \|P_\tau^{-1}\| \cdot \|f_\tau\| \leq C \|f_\tau\|.$$

Несложно показать эквивалентность всех приведенных определений устойчивости.

12.2. Построение разностных схем.

Исследование на сходимость

Рассмотрим основные понятия теории разностных схем на примере уравнения переноса (уравнение гиперболического типа):

$$\begin{aligned} \frac{du}{dt} + a \frac{du}{dx} &= \varphi(x), \\ t \in [0, +\infty); \quad x \in (-\infty, +\infty), \quad a &= \text{const}, \quad a < 0, \end{aligned} \quad (12.5)$$

u — искомая функция; t, x — независимые переменные.

Начальным условием для рассматриваемого уравнения является функция

$$u(0, x) = \psi_0(x), \quad x \in (-\infty, +\infty). \quad (12.6)$$

Задача (12.5), (12.6) называется *задачей Коши* (задача с начальными данными).

В операторном виде эта задача имеет вид

$$Pu = f, \quad (12.7)$$

где

$$\begin{aligned} Pu &= \begin{cases} \frac{du}{dt} + a \frac{du}{dx}; & t \in [0, \infty); \quad x \in (-\infty, +\infty); \\ u(0), & x \in (-\infty, +\infty); \end{cases} \\ f &= \begin{cases} \varphi(t, x); & t \in [0, \infty); \quad x \in (-\infty, +\infty); \\ \psi(x), & x \in (-\infty, +\infty). \end{cases} \end{aligned}$$



Для приближенного решения задачи с помощью метода конечных разностей введем расчетную сетку (совокупность точек):

$$\omega_n = \{t_n = n\tau; x_m = mh; n = 0, 1, \dots, N; \\ m = 0, \pm 1, \pm 2, \dots, \pm M; N = [T/\tau], M = [P/h]\}, \quad (12.8)$$

где $\tau > 0$ — шаг по времени, $h > 0$ — шаг по координате.

В разностной задаче мы полагаем: $0 \leq t \leq T$, $-M \leq x \leq M$, так как в вычислительном процессе интервалы, на которых происходит интегрирование, должны быть ограничены.

Часто полагают: $\tau = rh$, $r = \text{const}$, $r > 0$, т.е. сетка зависит от одного параметра: например, τ .

Приближенное решение ищется в точках пересечения прямых

$$x = mh \quad (m = 0, \pm 1, \pm 2, \dots, \pm M), \\ t = n\tau \quad (n = 0, 1, \dots, N)$$

и обозначается следующим образом:

$$u_m^n = u(n\tau, mh) = u(t_n, x_n).$$

Функция u_m^n от n, m называется *сеточной функцией*.

Теперь перейдем к построению разностной задачи, аппроксимирующей исходную дифференциальную задачу Коши (12.5), которую будет обозначать так:

$$P_\tau u_\tau = f_\tau, \quad (12.9)$$

где

$$P_\tau u_\tau = \begin{cases} \frac{u_m^{n+1} - u_m^n}{\tau} + a \frac{u_{m+1}^n - u_m^n}{h}, & n = 0, 1, \dots, N-1; \\ u_m^0, & m = 0, \pm 1, \pm 2, \dots; \end{cases} \quad (12.10)$$

$$f_\tau = \begin{cases} \varphi(t_n, x_m), & n = 0, 1, \dots, N-1; \\ \psi(x_m), & m = 0, \pm 1, \pm 2, \dots \end{cases}$$

Уравнение

$$\frac{u_m^{n+1} - u_m^n}{\tau} + a \frac{u_{m+1}^n - u_m^n}{h} = f_m^n, \\ n = 0, 1, \dots, (n-1); \quad m = 0, \pm 1, \pm 2, \dots, \pm M,$$

называют *разностным уравнением*, поскольку в нем используется аппроксимация производных с помощью разностных соотношений (конечных или разделенных разностей). Вид расчетной сетки представлен на рис. 12.1.

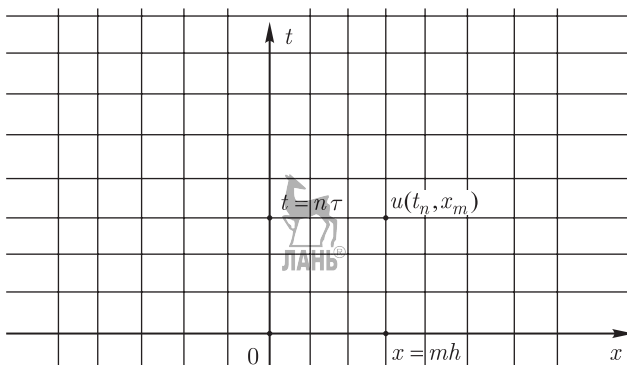


Рис. 12.1

В предположении, что $U(t, x) \in C^2(-\infty, +\infty)$, и используя формулу Тейлора, получим в соответствии с определением аппроксимации (U — проекция точного решения на расчетную сетку):

$$\begin{aligned} \frac{U(t_n, x_m + h) - U(t_n, x_m)}{h} &= \\ &= \frac{\partial U(t_n, x_m)}{\partial x} + \frac{h}{2} \frac{\partial^2 U(t_n, x_m + \xi)}{\partial x^2}; \quad (12.11) \\ \frac{U(t_n + \tau, x_m) - U(t_n, x_m)}{\tau} &= \frac{\partial U(t_n, x_m)}{\partial t} + \frac{\tau}{2} \cdot \frac{\partial^2 U(t_n + \eta, x_m)}{\partial t^2}, \end{aligned}$$

где $0 < \xi < h$, $0 < \eta < \tau$.

В таком случае:

$$P_\tau U_\tau = \begin{cases} \left(\frac{\partial U}{\partial t} + a \frac{\partial U}{\partial x} \right)_{t_n, x_m} + a \frac{\tau}{2} \cdot \frac{\partial^2 U(t_n + \eta, x_m)}{\partial t^2} + \\ + a \frac{h}{2} \cdot \frac{\partial^2 U(t_n, x_m + \xi)}{\partial x^2}, \\ U(0, x_m), \end{cases}$$

или:

$$P_\tau U_\tau = PU + \xi_\tau, \quad (12.12)$$

где невязка ξ_τ и оператор $P_\tau U_\tau$ имеют вид:

$$\begin{aligned} \xi_\tau &= \begin{cases} \frac{\tau}{2} \cdot \frac{\partial^2 U(t_n + \eta, x_m)}{\partial t^2} + a \frac{h}{2} \cdot \frac{\partial^2 U(t_n, x_m + \xi)}{\partial x^2}, \\ 0, \end{cases} \\ P_\tau U_\tau &= \begin{cases} \frac{U_m^{n+1} - U_m^n}{\tau} + a \frac{U_{m+1}^n - U_m^n}{h}, \\ U(0). \end{cases} \end{aligned}$$

В этом случае

$$\|\xi_\tau\| \leq \frac{\tau}{2} \cdot \max_{\Omega_\tau} \left| \frac{\partial^2 U}{\partial t^2} \right| + |a| \frac{h}{2} \max_{\Omega_\tau} \left| \frac{\partial^2 U}{\partial x^2} \right| = \frac{\tau}{2} \|U''_{xx}\| + \frac{|a|h}{2} \|U''_{xx}\|,$$

или $\|r_\tau\| = O(\tau + h)$, т.е. рассматриваемая разностная схема имеет первый порядок аппроксимации, так как в соответствии с определением аппроксимации

$$\|P_\tau U_\tau - Pu\| \leq C(\tau^1 + h^1).$$

Разложим проекцию точного решения на сетку в ряд Тейлора:

$$\begin{aligned} U(t_n + \tau) &= U(t_n) + \tau U'_t + \\ &\quad + \frac{\tau^2}{2} U''_t(t_n) + \frac{\tau^3}{3!} U^{(3)}_t(t_n) + \frac{\tau^4}{4!} U^{(4)}_t(t_n) + O(\tau)^5, \\ U(x_n + h) &= U(x_m) + h U'_x(x_m) + \\ &\quad + \frac{h^2}{2} U''_x(x_m) + \frac{h^3}{3!} U^{(3)}_t(t_n) + \frac{h^4}{4!} U^{(4)}_t(t_n) + O(h)^5, \end{aligned}$$

и подставить их в разностную схему (12.10) при $\varphi \equiv 0$ (при этом $P_\tau U_\tau = 0$), пренебрегая членами второго порядка малости, то получим дифференциальное уравнение в частных производных следующего вида:

$$\frac{du}{dt} + a \frac{du}{dx} + \frac{\tau}{2} \cdot \frac{\partial^2 u}{\partial t^2} + a \cdot \frac{h}{2} \cdot \frac{\partial^2 u}{\partial x^2} = 0. \quad (12.13)$$

Оно называется *первым дифференциальным приближением* (П-форма первого дифференциального приближения). Аналогично, используя разложение функции до малых более высокого порядка, можно получить второе, третье и т.д. дифференциальные приближения исходной задачи (12.5).

Если учесть так называемые *дифференциальные следствия* уравнения (12.5) при $\varphi \equiv 0$:

$$u''_{tt} = -a x_{tx}, \quad u''_{tt} = a^2 u''_{xx},$$

получающиеся при дифференцировании уравнения

$$u'_t = -a u'_x$$

по t :

$$u''_{tt} = -a u''_{xt} = -a u''_{tx} = -a (-a u'_x)_x = a^2 u''_{xx},$$

то получим П-форму первого дифференциального приближения:

$$\frac{du}{dt} + a \frac{du}{dx} - \frac{h}{2} (1 - \sigma) \frac{\partial^2 u}{\partial x^2} = 0, \quad (12.14)$$

где $\sigma = a\tau/h$ — число Куранта, или

$$\frac{du}{dt} + a \frac{du}{dx} - \nu \frac{d^2u}{dx^2} = 0, \quad (12.15)$$

где

$$\nu = |a| \frac{h}{2} (1 - \sigma)$$

— коэффициент аппроксимации вязкости, который, как можно показать, должен быть положительным для того, чтобы решение (12.15) было ограниченным (условие устойчивости), что достигается при

$$\sigma = \frac{|a|\tau}{h} < 1 \quad (12.16)$$

(условие устойчивости Куранта–Фридрихса–Леви).

Например, для разностной схемы «правый уголок»

$$\frac{u_m^{n+1} - u_m^n}{\tau} - a \frac{u_{m+1}^n - u_m^n}{h} = 0, \quad a > 0$$

дифференциальное приближение будет иметь вид

$$u'_t - u'_x = \frac{1}{2!} (h - \tau) u''_x + \frac{1}{3!} (h^2 - \tau^2) u'''_x + \dots \\ \dots + \frac{1}{(n+1)!} (h^n - \tau^n) u_x^{(n+1)} + O(\tau^{n+1} + h^{n+1}),$$

или, опустив $O(\tau^{n+1} + h^{n+1})$, получим

$$u'_t - u'_x = \sum_{k=1}^n h^k \frac{1 - \sigma^k}{(k+1)!} u_x^{(k+1)}.$$

Рассмотрим также аппроксимацию смешанной задачи для уравнения параболического типа:

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f(t, x); & t > 0, \quad 0 < x < X; \\ -a_1 \frac{\partial u}{\partial x} + b_1 u = \varphi_1(t); & x = 0, \quad 0 < t \leq T; \\ -a_2 \frac{\partial u}{\partial x} + b_2 u = \varphi_2(t); & x = X, \quad 0 < t \leq T; \\ u(0, x) = \varphi(x); & t = 0, \quad x \leq x \leq X. \end{cases} \quad (12.17)$$

Для разностной аппроксимации (12.17) введем расчетную сетку (дискретное множество точек либо совокупность точек в рассматриваемой области):

$$\omega_n = \{t_n = n\tau; x_m = mh; n = 0, 1, \dots, Ml; m = 0, \dots, M; \\ \tau = T/N; h = X/M\}. \quad (12.18)$$

Приближенное решение ищем в виде сеточной функции $u(t_n, x_m)$ в узлах расчетной сетки; значения функции между узлами находятся методом интерполяции.

Приведем аппроксимацию с помощью явной разностной схемы:

$$\frac{u_m^{n+1} - u_m^n}{\tau} = \frac{u_{m-1}^n - 2u_m^n + u_{m+1}^n}{h^2} + f_m^n, \quad (12.19)$$

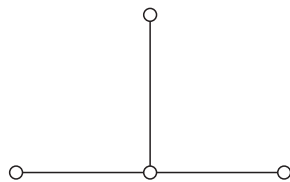


Рис. 12.2

которой соответствует шаблон, состоящий из узловых точек расчетной сетки (рис. 12.2), использующихся в расчете.

Аппроксимация начальных и краевых условий для (12.19) имеет вид:

$$\begin{aligned} u_m^0 &= \varphi(x_m); \quad m = 0, 1, \dots, M; \\ -a_1 \frac{u_1^n - u_0^n}{h} + b_1 u_0^n &= \varphi_1(t_n); \quad n = 1, 2, \dots, N \end{aligned} \quad (12.20)$$

(левое краевое условие);

$$-a_2 \frac{u_M^n - u_{M-1}^n}{h} + b_2 u_M^n = \varphi_2(t_n); \quad n = 1, 2, \dots, N,$$

причем положим для определенности:

$$a_1, a_2, b_1, b_2 \geq 0; \quad a_1 + b_1 > 0; \quad a_2 + b_2 > 0.$$

Вычислительная реализация данной разностной задачи состоит в реализации алгоритма счета по слоям (или алгоритма «бегущего счета»), т.е. по известным значениям на n -м слое находятся значения сеточной функции $u(t_n, x_m)$ на $(n+1)$ -м:

$$\begin{aligned} u_m^{n+1} &= u_m^n + \tau \frac{u_{m-1}^n - 2u_m^n + u_{m+1}^n}{h^2} + \tau f_m^n, \\ m &= 1, 2, \dots, (M-1); \quad n = 1, 2, \dots \end{aligned} \quad (12.21)$$

Начальное значение сеточной функции u_m^n при $n = 0$ задано; из левого и правого краевых условий (при $m = 0$ и $m = M$) находим значения сеточной функции соответственно на левой и на правой границах области интегрирования:

$$\begin{aligned} u_0^{n+1} &= \frac{a_1 u_1^{n+1}}{a_1 + h b_1} + h \frac{\varphi_1(t_{n+1})}{a_1 + h b_1}, \\ u_M^{n+1} &= \frac{a_2 u_{M-1}^{n+1}}{a_2 + h b_2} + h \frac{\varphi_2(t_{n+1})}{a_1 + h b_2}. \end{aligned} \quad (12.22)$$

Вычисления значений сеточной функции на левой и правой границах реализуются после расчетов u_1^{n+1} и u_{M-1}^{n+1} по схеме (12.19). Вычисления на одном шаге требуют $O(M)$ арифметических операций, на N шагах — $O(NM)$. Требуемые ресурсы машинной памяти оцениваются в $(N+1)(M+1)$ ячеек памяти, однако поскольку в расчетах участвуют только два слоя по времени, то можно, как правило, обойтись $2(M+1)$ ячейками.

Неявная схема для численного решения рассматриваемого уравнения будет иметь следующий вид:

$$\frac{u_m^{n+1} - u_m^n}{\tau} = \frac{u_{m-1}^{n+1} - 2u_m^{n+1} + u_{m+1}^{n+1}}{h^2} + f_m^n.$$

Если ввести разностные операторы

$$\Delta_\tau u_m^n = \frac{u_m^{n+1} - u_m^n}{\tau}, \quad \Lambda_{xx} u^n = \frac{u_{m-1}^n - 2u_m^n + u_{m+1}^n}{h^2},$$

$$\Lambda_{xx} u^{n+1} = \frac{u_{m-1}^{n+1} - 2u_m^{n+1} + u_{m+1}^{n+1}}{h^2},$$

то явную и неявную разностные схемы можно записать в виде:

$$\begin{aligned} \frac{\Delta_\tau u_m^n}{\tau} &= \Lambda_{xx} u_m^n + f_m^n, \\ \frac{\Delta_\tau u_m^n}{\tau} &= \Lambda_{xx} u_m^{n+1} + f_m^n. \end{aligned} \quad (12.23)$$

Если в явной схеме координатные производные аппроксимируются на n -м (нижнем) слое, то в неявной — на $(n+1)$ -м (верхнем).

Представим неявную разностную задачу в следующем, удобном для численной реализации, виде:

$$\begin{cases} (a_1 + hb_1) u_0^{n+1} - a_1 u_0^{n+1} = h\varphi_1^{n+1}; \\ \frac{\tau}{h^2} u_{m-1}^{n+1} - \left(1 + 2\frac{\tau}{h^2}\right) u_m^{n+1} + \frac{\tau}{h^2} u_{m+1}^{n+1} = u_{m-1}^{n+1} - \tau f_m^n, \\ \quad m = 1, \dots, M-1; \\ (a_2 + hb_2) u_M^{n+1} - a_2 u_M^{n+1} = h\varphi_2^{n+1}. \end{cases} \quad (12.24)$$

Это система линейных алгебраических уравнений с матрицей трехдиагональной структуры, которая решается методом трехточечной прогонки на каждом временном слое $t = t_n$. Поскольку условие диагонального преобладания выполняется, то данный алгоритм будет устойчивым.

Проведем исследование однородного разностного уравнения (12.24) на аппроксимацию, учтя, что

$$\begin{aligned}\frac{\partial^2 U(x)}{\partial x^2} &\approx \frac{U(x_m + h) - 2U(x_m) + U(x_m - h))}{h^2} = \\ &= U''_{xx}(x_m) + \frac{h^2}{12} U^{(4)}_x(x_m) + O(h^4).\end{aligned}$$

После разложения сеточных функций в окрестности точки $\{t_n, x_m\}$ получим:

$$\begin{aligned}\xi_\tau &= \frac{\tau}{2} \cdot \frac{\partial^2 U}{\partial t^2}(t_n, x_m) - \frac{h^2}{12} \cdot \frac{\partial^{(4)} U}{\partial t^{(4)}}(t_n, x_m) + O(\tau^2, h^4), \\ \|\xi_\tau\| &\leq \frac{\tau}{2} \max_{\Omega_\tau} \left| \frac{\partial^2 U}{\partial t^2} \right| + \frac{h^2}{12} \max_{\Omega_\tau} \left| \frac{\partial^{(4)} U}{\partial t^{(4)}} \right|,\end{aligned}$$

или

$$\|\xi_\tau\| = O(\tau + h^2),$$

т.е. рассматриваемая схема имеет первый порядок аппроксимации по t и второй по x , а первое дифференциальное приближение исходного дифференциального уравнения в частных производных (12.17) будет иметь вид

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = \frac{\tau}{2} \cdot \frac{\partial^2 u}{\partial t^2} - \frac{h^2}{12} \cdot \frac{\partial u^{(IV)}}{\partial x^4}, \quad (12.25)$$

или, с учетом равенства

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^{(4)} u}{\partial x^4},$$

получим

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} - \frac{1}{2} \left(\tau - \frac{h^2}{6} \right) \cdot \frac{\partial u^{(IV)}}{\partial x^4} = 0. \quad (12.26)$$

Разностная схема для численного решения нестационарного уравнения в частных производных может быть представлена в виде [3]:

$$\begin{aligned}B \frac{u_m^{n+1} - u_m^n}{\tau} + Au_n = f_n; \quad u_n \in R^n; \quad f_n \in R^n; \\ n = 0, \dots, (N-1),\end{aligned} \quad (12.27)$$

где $A \in S$ и $B \in S$ — операторы, действующие в линейных n -мерных пространствах R^n , принадлежащие линейному пространству операторов S .

В качестве классического примера такого представления разностной схемы приведем разностную схему с весом $\eta \in [0, 1]$,

аппроксимирующую уравнение в частных производных параболического типа ($u_n \in R^1$):

$$\begin{aligned} \frac{u_m^{n+1} - u_m^n}{\tau} &= \eta \Lambda_{xx} u_m^{n+1} + (1 - \eta) \Lambda_{xx} u_m^n, \\ \Lambda_{xx} u_m^{n+1} &= h^{-2} (u_{m-1}^{n+1} - 2u_m^{n+1} + u_{m+1}^{n+1}); \\ n &= 0, \dots, (N-1); \quad m = 0, \dots, (M-1). \end{aligned} \quad (12.28)$$

При $\eta = 0$ имеем явную схему. Разностное уравнение представляется в виде (12.28), если положить:

$$\begin{aligned} \mathbf{A} \mathbf{u}_n &= -\Lambda_{xx} u_m^n, \quad \mathbf{u}_n = \{u_1^n, \dots, u_{M-1}^n\}^T, \\ \mathbf{B} &= \mathbf{E} + \tau \eta \mathbf{A}. \end{aligned}$$

Определение 12.6. Разностная схема

$$\mathbf{B} \frac{\mathbf{u}_m^{n+1} - \mathbf{u}_m^n}{\tau} + \mathbf{A} \mathbf{u}_n = \mathbf{f}_n$$

называется *равномерно устойчивой по начальным данным*, если \exists постоянные $\rho > 0$, $\rho \neq \rho(\tau, h)$ и $C > 0$, $C \neq C(\tau, h)$ такие, что при $\forall n = 0, \dots, N-1$, $\mathbf{u}_n \in R^n$ выполняется неравенство

$$\|\mathbf{u}_{n+1}\| \leq \rho \|\mathbf{u}_n\|, \quad \rho^n \leq C, \quad (12.29)$$

откуда

$$\|\mathbf{u}_{n+1}\| \leq \rho^n \|\mathbf{u}_0\|, \text{ или } \|\mathbf{u}_n\| \leq C \|\mathbf{u}_0\|.$$

Обычно полагают:

$$\begin{aligned} \rho &= 1; \quad \rho = 1 + C\tau, \quad \rho = e^{C_0\tau}, \quad C_0 > 0, \\ C_0^{\tau} &\ll 1, \quad C \neq C_0(\tau). \end{aligned}$$

Теорема 12.2 [1]. Пусть разностная схема (12.27) равномерно устойчива по начальным данным, а оператор перехода

$$R_{\tau} = E + \tau B^{-1} A \quad (12.30)$$

имеет вещественные собственные значения λ_i и соответствующие собственные векторы ω_i . Тогда

$$\lambda_i \leq 1 + C\tau,$$

где $C \neq C(\tau)$, $C\tau \ll 1$.

Доказательство. Представим разностную схему (12.28) в виде

$$\mathbf{u}_{n+1} = R_{\tau} \mathbf{u}_n + \tau B^{-1} \mathbf{f}_n, \quad n = 0, \dots, N-1,$$

где $R_{\tau} = E + \tau B^{-1} A$ — оператор перехода.

Условие равномерной сходимости по начальным данным эквивалентно условию ограничения нормы оператора перехода

$$\|R_\tau\| \leq \rho,$$

а поскольку $\rho^n \leq C$, то мы получаем условие ограничения норм степеней оператора R_τ :

$$\|R_\tau^n\| \leq C.$$

Для того чтобы оценить норму оператора R_τ , можно воспользоваться известным равенством

$$R_\tau \omega_i = \lambda_i \omega_i,$$

из которого следует

$$\|R_\tau^n\| \cdot \|\omega_i\| \geq \|R_\tau^n \omega_i\| = |\lambda_i^n| \cdot \|\omega_i\|,$$

или

$$\|R_\tau^n\| \geq |\lambda_i|^n.$$

Так как последнее неравенство должно выполняться при любом n и при этом

$$\|R_\tau^n\| \leq C,$$

то величина $|\lambda_i|^n$ не может неограниченно расти с ростом n , чего не произойдет при выполнении условия

$$|\lambda_i| \leq 1 + C\tau, \quad C > 0, \quad C \neq C(\tau), \quad C\tau \ll 1. \quad (12.31)$$

Последнее неравенство называется *спектральным признаком устойчивости*.

Дж. Нейманом был предложен спектральный признак устойчивости для линейных однородных разностных схем, состоящий в следующем.

Рассмотрим разностную задачу Коши, представленную в операторном виде:

$$P_\tau u_\tau = f_\tau. \quad (12.32)$$

Ее можно представить и в скалярном виде: например, одна из различных схем, аппроксимирующих линейное уравнение переноса (12.32), будет иметь следующий вид:

$$P_\tau u_\tau = \begin{cases} \frac{u_m^{n+1} - u_m^n}{\tau} - \frac{u_{m-1}^n - u_m^n}{h}, & n = 0 \div N-1; \\ u_m^0, & m = 0 \div N-1; \end{cases} \quad (12.33)$$

$$f_\tau = \begin{cases} f_m^n, & n = 0 \div N-1; \\ \varphi_m, & m = 0 \div N-1. \end{cases}$$

Условие устойчивости, в соответствии с определением корректности разностной задачи (определение 12.5), может быть записано следующим образом:

$$\|u_\tau\| \leq C \|f\|, \quad (12.34)$$

что означает, в случае однородного разностного уравнения, устойчивость рассматриваемой разностной задачи (12.32) по начальным данным. Поэтому (12.34) мы можем представить в виде

$$\max_m |u_m^n| \leq C \max |u_m^0|. \quad (12.35)$$

Последнее условие должно выполняться и в том случае, если начальное условие в задаче Коши $u_m^0 = \varphi_m$ является гармоникой вида

$$u_m^0 = e^{i\alpha m}, \quad (12.36)$$

где $\alpha \in [0, 2\pi]$ — вещественный параметр.

Решение же разностной однородной задачи (12.32) в этом случае может быть найдено с помощью метода разделений переменных:

$$u_m^n = \lambda^n e^{i\alpha m} \quad (\text{или } u_m^n = \lambda^n u_m^0), \quad (12.37)$$

где первой сомножитель соответствует функции от времени t_n (или от n), второй — функции от координаты x_m (или от m).

После подстановки $\lambda(\alpha)$ в рассматриваемое разностное уравнение, которое можно записать в виде

$$u_m^{n+1} = (1 - \sigma) u_m^n + \sigma u_{m+1}^n, \quad (12.38)$$

получим

$$\lambda(\alpha) = (1 - \sigma) + \sigma e^{i\alpha}.$$

Отметим, что в случае если коэффициент переноса $a \neq 1$ (см. (12.5)), то

$$\sigma = \frac{a\tau}{h}. \quad (12.39)$$

Это число называется *числом Куранта*; оно играет важную роль в исследованиях устойчивости разностных схем, аппроксимирующих дифференциальные уравнения в частных производных гиперболического типа (для его обозначения в литературе также используются обозначения K, r). Основной вопрос, возникающий при исследовании разностной схемы на устойчивость, состоит в том, будет ли начальная гармоника расти со временем (или с увеличением n). Для ответа на него проведем следующие оценки разностного решения u_m^n по норме $\|u_m^n\| = \max_m |u_m^n|$:

$$\|u_m^n\| = \|\lambda^n e^{i\alpha m}\| = |\lambda|^n \cdot \|u_m^0\|,$$

или

$$\max_m |u_m^n| = |\lambda^n| \cdot \max_m |u_m^0|. \quad (12.40)$$

Очевидно, что для выполнения условия устойчивости разностной задачи по начальным данным

$$\|u_m^n\| = C \|u_m^0\|,$$

или

$$\max_m \|u_m^n\| \leq C \max_m |u_m^0|,$$

необходимо выполнение неравенства

$$|\lambda(\alpha)| \leq 1 + C\tau, \quad (12.41)$$

где $C \neq C(\tau)$, что соответствует уже полученному спектральному условию устойчивости разностных задач. Заметим, что линейные однородные разностные уравнения с постоянными коэффициентами, заданные на всем пространстве и ограниченные на бесконечности, имеют частные решения вида

$$u_m^n = \lambda^n e^{i\alpha m},$$

где параметр этого семейства λ зависит от τ, h, α и от вида разностной схемы. При этом λ является собственным числом (или спектром) оператора перехода R_τ , который в случае рассматриваемого разностного уравнения имеет вид

$$u_m^{n+1} = R_\tau u_m^n, \quad (12.42)$$

где

$$R_\tau u_m^n = (1 - \sigma) u_m^n + \sigma u_{m+1}^n.$$

Оператор перехода, определенный на n -м временном слое, ставит в соответствие сеточной функции u_m^n , вычисляемой на n -м слое, сеточную функцию u_m^{n+1} , вычисляемую на $(n+1)$ -м слое. Гармоника $e^{i\alpha m}$ является собственной функцией, а λ — собственным значением оператора R_τ , что видно после подстановки $\omega_m = e^{i\alpha m}$ в уравнение

$$R_\tau \omega_\tau = \lambda \omega_\tau,$$

или

$$(1 - \sigma) e^{i\alpha m} + \sigma e^{i\alpha(m+1)} = \lambda e^{i\alpha m},$$

откуда получим

$$\lambda(\alpha) = (1 - \sigma) + \sigma e^{i\alpha}.$$

Кривая, соответствующая функции $\lambda(\alpha)$ на комплексной плоскости, является спектром оператора перехода.

Таким образом, спектральный признак устойчивости состоит в том, что оператор перехода с n -го по времени слоя на $(n-1)$ -й

должен лежать в круге радиуса $1 + C\tau$ на комплексной плоскости. На нашем примере граница спектра $|\lambda(\alpha)| \leq 1$ представляет собой окружность с центром в точке $1 - \sigma$ и радиусом σ комплексной плоскости. При этом если $\sigma < 1$, то эта окружность лежит внутри единичного круга. Следовательно, наша разностная схема будет устойчивой при выполнении условия

$$\sigma \leq 1.$$

Рассмотрим также условие устойчивости разностной схемы

$$\frac{u_m^{n+1} - u_m^n}{\tau} - k \frac{u_{m-1}^n - 2u_m^n + u_{m+1}^n}{h^2} = 0,$$

аппроксимирующей уравнение теплопроводности.

После подстановки решения, записанного в виде

$$u_m^n = \lambda^n e^{i\alpha m},$$

получим:

$$\lambda(\alpha) = 1 - 4r \sin^2 \frac{\alpha}{2}, \quad r = \frac{k\tau}{h^2}. \quad (12.43)$$

Видно, что при изменении α функции $\lambda(\alpha)$ получают значения:

$$1 - 4rk \leq \lambda(\alpha) \leq 1,$$

т. е. условие устойчивости $|\lambda(\alpha)| \leq 1$ выполняется, если

$$1 - 4rk \leq -1,$$

или же:

$$r \leq \frac{1}{2k}, \quad \tau \leq \frac{h^2}{2k}. \quad (12.44)$$

Отметим, что соответствующая неявная разностная схема вида

$$\frac{u_m^{n+1} - u_m^n}{\tau} - k \frac{u_{m-1}^{n+1} - 2u_m^{n+1} + u_{m+1}^{n+1}}{h^2} = 0$$

будет устойчивой при любых соотношениях τ, h (т. е. при любом r), поскольку подстановка решения в (12.44) дает:

$$\lambda(\alpha) = \left(1 + 4\sigma \sin^2 \frac{\alpha}{2}\right)^{-1},$$

т. е. $0 < \lambda(\alpha) \leq 1$.

Теорема 12.3 [3]. Если разностная схема

$$\mathbf{B} \frac{\mathbf{u}_{n+1} - \mathbf{u}_n}{\tau} + \mathbf{A} \mathbf{u} = \mathbf{f}$$

равномерно устойчива по начальным данным, то она устойчива и по правым частям. При этом выполняется

$$\|\mathbf{u}_n\| \leq C_1 \|\mathbf{u}_0\| + C_2 \cdot \|\mathbf{f}_n\|. \quad (12.45)$$

Доказательство. Представим рассматриваемую разностную схему в виде

$$\mathbf{u}_{n+1} = \mathbf{R}_\tau \mathbf{u}_n + \tau \mathbf{B}^{-1} \mathbf{f}_n. \quad (12.46)$$

Из последнего равенства следует

$$\|\mathbf{u}_{n+1}\| \leq \|\mathbf{R}_\tau\| \cdot \|\mathbf{u}_n\| + \tau \|\mathbf{B}^{-1} \mathbf{f}_n\|,$$

причем если воспользоваться тем, что схема равномерно устойчива по начальным данным, полученное неравенство можно представить в следующем виде:

$$\|\mathbf{u}_{n+1}\| \leq \rho \cdot \|\mathbf{u}_n\| + \tau \|\mathbf{B}^{-1} \mathbf{f}_n\| \leq \rho \|\mathbf{u}_n\| + \tau \|\mathbf{B}^{-1}\| \cdot \|\mathbf{f}\|,$$

где $\|\mathbf{f}\| = \max_n \|\mathbf{f}_n\|$.

Отсюда получим цепочку неравенств:

$$\begin{aligned} \|\mathbf{u}_1\| &\leq \rho \cdot \|\mathbf{u}_0\| + \tau \|\mathbf{B}^{-1} \mathbf{f}_n\|; \\ \|\mathbf{u}_2\| &\leq \rho^2 \cdot \|\mathbf{u}_0\| + \tau \|\mathbf{B}^{-1} \mathbf{f}_n\| + \tau \rho \|\mathbf{B}^{-1} \mathbf{f}_n\|; \\ \|\mathbf{u}_3\| &\leq \rho^3 \cdot \|\mathbf{u}_0\| + \tau \|\mathbf{B}^{-1} \mathbf{f}_n\| + \tau \rho \|\mathbf{B}^{-1} \mathbf{f}_n\| + \tau \rho^2 \|\mathbf{B}^{-1} \mathbf{f}_n\|; \\ &\dots \\ \|\mathbf{u}_{n+1}\| &\leq \rho^{n+1} \cdot \|\mathbf{u}_0\| + \tau \|\mathbf{B}^{-1} \mathbf{f}_n\| + \tau \rho \|\mathbf{B}^{-1} \mathbf{f}_n\| + \\ &\quad + \tau \rho^2 \|\mathbf{B}^{-1} \mathbf{f}_n\| + \dots + \tau \rho^n \|\mathbf{B}^{-1} \mathbf{f}_n\|. \end{aligned}$$

Поскольку $\rho^i \leq C$ для всех i , то получаем следующие оценки:

$$\begin{aligned} \|\mathbf{u}_{n+1}\| &\leq \rho^{n+1} \cdot \|\mathbf{u}_0\| + \tau \|\mathbf{B}^{-1}\| \|\mathbf{f}_n\| + \rho^2 \cdot \tau \|\mathbf{B}^{-1}\| \|\mathbf{f}_n\| + \dots \\ &\quad \dots + \rho^n \|\mathbf{B}^{-1}\| \|\mathbf{f}_n\| \leq C \|\mathbf{u}_0\| + Cn\tau \|\mathbf{B}^{-1}\| \|\mathbf{f}_n\| = \\ &= C_1 \|\mathbf{u}_0\| + C_2 \|\mathbf{f}_n\|, \end{aligned}$$

где $C_2 = Cn\tau \|\mathbf{B}^{-1}\|$.

Теорема доказана.

Наряду со скалярным произведением (\mathbf{x}, \mathbf{y}) и нормой $\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})}$ введем так называемую *энергетическую норму* [3]

$$\|\mathbf{u}\|_P = \sqrt{(P\mathbf{u}, \mathbf{u})}; \quad \mathbf{u}, \mathbf{x}, \mathbf{y} \in R^n, \quad (12.47)$$

порожденную положительно определенным оператором P , для которого

$$(P\mathbf{u}, \mathbf{u}) > 0, \quad P = P^*, \quad P \in S,$$

где S — линейное пространство операторов.

Напомним также, что операторное неравенство $A \geq B$ означает: $((A - B)\mathbf{u}, \mathbf{u}) \geq 0$.

Теорема 12.4 [3]. Пусть $A = A^*$, $A \neq A(n)$.

Для того чтобы разностная схема

$$B \frac{u_{n+1} - u_n}{\tau} + Au = 0, \quad n = 0, \dots, N-1, \quad (12.48)$$

была равномерно устойчивой по начальным данным, необходимо и достаточно, чтобы имело место неравенство

$$B \geq \frac{\tau}{2} A \quad (12.49)$$

и при этом в случае одномерного разностного уравнения выполнялось

$$\|u_{n+1}\|_A \leq \|u\|_A, \quad n = 0, \dots, N-1.$$

Доказательство. Достаточность. Пусть выполняется условие $B \geq (\tau/2)A$. Умножим рассматриваемое разностное уравнение (12.49) на функцию

$$u_t = \frac{u_{n+1} - u_n}{\tau};$$

для краткости обозначим $u = u_n$.

Тогда

$$(Bu_t, u_t) + (Au, u_t) = 0,$$

откуда получим:

$$((B - 0,5\tau A)u_t, u_t) + (0,5\tau Au_t + Au, u_t) = 0. \quad (12.50)$$

Нетрудно показать, что

$$0,5\tau Au_t + Au = 0,5A(u_{n+1} + u_n).$$

В таком случае из (12.51) имеем:

$$((B - 0,5\tau A)u_t, u_t) + 0,5\tau^{-1}(A(u_{n+1} + u_n), u_{n+1} - u_n) = 0. \quad (12.51)$$

Поскольку

$$\begin{aligned} (A(u_{n+1} + u_n), (u_{n+1} - u_n)) &= (Au_{n+1}, u_{n+1}) - (Au_{n+1}, u_n) + \\ &+ (Au_n, u_{n+1}) - (Au_n, u_n) = (Au_{n+1}, u_{n+1}) - (Au_n, u_n) = \\ &= \|u_{n+1}\|_A^2 - \|u_n\|_A^2, \end{aligned}$$

то из (12.51) следует

$$2\tau((B - 0,5\tau A)u_t, u_t) + (\|u_{n+1}\|_A^2 - \|u_n\|_A^2) = 0. \quad (12.52)$$

Из условия теоремы имеем

$$B \geq 0,5\tau A,$$

откуда следует

$$((B + 0, 5\tau A)u_t, u_t) \geq 0,$$

и, с учетом равенства (12.52), получаем искомое неравенство:

$$\|u_{n+1}\|_A^2 - \|u_n\|_A^2 \leq 0 \quad \text{или} \quad \|u_{n+1}\|_A \leq \|u_n\|_A,$$

что и требовалось доказать.

Необходимость доказывается исходя из (12.52).

Пример. Схема Кранка–Никольсон имеет вид

$$\frac{u_{n+1} - u_n}{\tau} = -\frac{1}{2} \Lambda_{xx} u_n - \frac{1}{2} \Lambda_{xx} u_{n+1}. \quad (12.53)$$

Перепишем (12.53) в виде

$$\left(E + \frac{\tau}{2} \Lambda_{xx}\right) \frac{u_{n+1} - u_n}{\tau} + \Lambda_{xx} u_n = 0.$$

Поскольку в этом случае

$$B = E + \frac{\tau}{2} \Lambda_{xx}, \quad A = \Lambda_{xx},$$

а

$$E + \frac{\tau}{2} \Lambda_{xx} > \frac{\tau}{2} \Lambda_{xx},$$

то условие $B \geq (\tau/2)A$ (12.49) выполняется, т. е. рассматриваемая в примере разностная схема устойчива по начальным данным.

Необходимое условие сходимости решения разностного уравнения к решению дифференциального дается теоремой Куранта.

Теорема 12.5 (Куранта). *Рассмотрим краевую задачу*

$$Pu = f,$$

с краевыми условиями

$$\Psi(u) = 0 \text{ на } \Gamma,$$

где Γ — границы области интегрирования, и соответствующую разностную задачу

$$P_\tau u_\tau = f_\tau.$$

Пусть $\Omega(P)$ и $\Omega_\tau(P)$ являются областями зависимости в точке P для решения этих двух задач соответственно. Для того чтобы решение разностной задачи в точке P стремилось к решению дифференциальной задачи в точке P , необходимо, чтобы все точки множества $\Omega(P)$ были предельными точками $\Omega_\tau(P)$, т. е.

$$\Omega(P) \subset \lim_{\tau \rightarrow 0} \Omega_\tau(P).$$

В некоторых случаях проблему устойчивости разностной схемы удастся решить с помощью принципа максимума, который можно продемонстрировать на примере разностной схемы

$$u_m^{n+1} = (1 - \sigma) u_m^n + \sigma u_{m+1}^n + \tau f_m^n \quad (12.54)$$

для численного решения уравнения переноса $u'_t - a u'_x = 0$, для чего оценим (12.54) по норме, предполагая, что $(1 - \sigma) \geq 0$:

$$\begin{aligned} \max_m |u_m^{n+1}| &\leq \|(1 - \sigma) u_m^n + \sigma u_{m+1}^n + \tau f_m^n\| \leq \\ &\leq \left| (1 - \sigma + \sigma) \max_n (|u_m^n|, |u_{m+1}^n|) \right| + \tau |f_m^n|. \end{aligned} \quad (12.55)$$

Здесь: $\|u_m^n\| = \max_n (\sup_m \{|u_m^n|\})$.

Видно, что при $f_m^n = 0$ норма решения $\|u_m^n\|$ не возрастает с ростом n , т.е. схема устойчива по начальным данным, а поскольку правая часть (12.55) не зависит от m , то его можно переписать в виде максимума

$$\max_m |u_m^{n+1}| \leq \max_m |u_m^n| + \tau \max_{n,m} |f_m^n|. \quad (12.56)$$

Список литературы

1. Годунов С. К., Рябенький В. С. Разностные схемы. Москва: Наука, 1973. 400 с.
2. Федоренко Р. П. Введение в вычислительную физику. Долгопрудный: Интеллект, 2008. 503 с.

Дополнительная литература

3. Самарский А. А. Теория разностных схем. М.: Наука, 1983. 656 с.
4. Петров И. Б., Лобанов А. И. Лекции по вычислительной математике. М.: БИНОМ. Лаборатория знаний, 2006. 522 с.
5. Шокин Ю. И., Яненко Н. Н. Метод дифференциального приближения. Применение к газовой динамике. Новосибирск: Наука, 1985. 364 с.
6. Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. Численные методы. М.: ФИЗМАТЛИТ, 2000. 622 с.
7. Рихтмайер Р., Мортон К. Разностные методы решения краевых задач. М.: Мир, 1972. 418 с.

ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ ПАРАБОЛИЧЕСКОГО ТИПА (УРАВНЕНИЯ ДИФФУЗИИ, ТЕПЛОПРОВОДНОСТИ)



13.1. Однородное линейное уравнение теплопроводности

Рассмотрим основные численные методы решения дифференциальных уравнений в частных производных параболического типа на примере одномерного уравнения теплопроводности, которое может быть представлено как в недивергентной, так и в дивергентной (потокковой) форме, соответственно:

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial x} \left(k(x) \frac{\partial T}{\partial x} \right) + f(t, x); \quad (13.1)$$

$$\frac{\partial T}{\partial t} + \frac{\partial \Pi}{\partial x} = f(t, x), \quad \Pi = -k(t, x) \frac{\partial T}{\partial x}; \quad (13.2)$$

$$0 < t \leq \theta; \quad 0 < x < X.$$

Здесь $T(t, x)$ — температура среды, t, x — независимые переменные (время, координата), $k(t, x) > 0$ — коэффициент теплопроводности, Π — тепловой поток, $f(t, x)$ — источники тепла.

К уравнениям (13.1) и (13.2) необходимо добавить начальные и граничные (краевые) условия (смешанная задача о распространении тепла):

$$\begin{aligned} T(0, x) &= \Psi_0(x), \quad t = 0; \\ -a_1 \frac{\partial T}{\partial x} + b_1 T &= \varphi_1(t), \quad x = 0; \\ -a_2 \frac{\partial T}{\partial x} + b_2 T &= \varphi_2(t), \quad x = X. \end{aligned}$$

Для разностной аппроксимации рассматриваемого уравнения, как и выше, проводится дискретизация области интегрирования введением расчетной сетки ω_τ :

$$\begin{aligned} \omega_{NM} = \{t_n = n\tau; \quad n = 0, 1, \dots, N; \quad \tau = \theta/N; \quad x_n = mh; \\ m = 0, 1, \dots, M; \quad h = X/M\}. \end{aligned}$$

В самом общем виде, в случае применения двухслойных разностных схем (т.е. схем, с помощью которых рассчитывается значение искомой функции T_m^{n+1} на верхнем $(n+1)$ -м временном слое по известным значениям на n -м слое), можно записать решение однородного уравнения на верхнем слое в следующем виде:

$$T_m^{n+1} = \sum_i \sum_j \alpha_j^i(\tau, h) \cdot T_{m+j}^{n+i},$$

где α_j^i — коэффициенты разностной схемы. $i = 0, 1$ для двухслойных схем ($i = 0$ — явная схема, $i = 1$ — неявная); $j = 0, \pm 1, \dots$ — номера узловых точек на координатной оси.

Выбор коэффициентов α_j^i разностной схемы зависит от выбранного численного метода и заданных свойств разностной схемы. Начнем построение численных методов для решения уравнения теплопроводности с интегро-интерполяционного метода, который использует интегральную форму записи (13.2).

Аппроксимируем однородное уравнение (13.2) по 6 точкам в плоскости (t, x) : (t_n, x_{m-1}) , (t_n, x_m) , (t_n, x_{m+1}) , (t_{n+1}, x_{m-1}) , (t_{n+1}, x_m) , (t_{n+1}, x_{m+1}) — шаблон представлен на рис. 13.1.

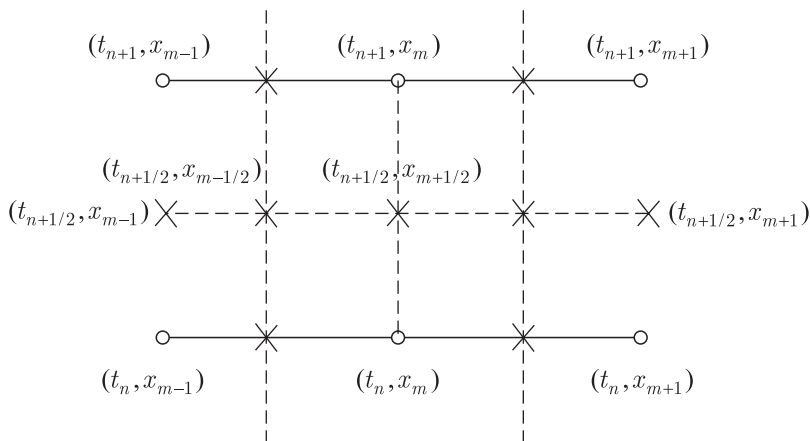


Рис. 13.1

Для этого воспользуемся известной из курса математического анализа формулой

$$\int_S \left(\frac{\partial T}{\partial t} + \frac{\partial \Pi}{\partial x} \right) dt dx = \oint_{\Gamma} (T dx - \Pi dt) = 0 \quad (13.3)$$

и аппроксимируем контурный интеграл в последней формуле по формуле средних:

$$T_m^n \cdot h - \Pi_{m+1/2}^{n+1/2} \cdot \tau - T_m^{n+1} \cdot h + \Pi_{m-1/2}^{n+1/2} \cdot \tau = 0,$$

или:

$$\frac{T_m^{n+1} - T_m^n}{\tau} + \frac{\Pi_{m+1/2}^{n+1/2} - \Pi_{m-1/2}^{n+1/2}}{h} = 0,$$

где аппроксимирующие выражения для тепловых потоков представляются следующим образом:

$$\begin{aligned} \Pi_{m+1/2}^{n+1/2} &= \frac{1}{2} \left(k_{m+1/2}^{n+1} \frac{T_{m+1}^{n+1} - T_m^{n+1}}{h} + k_{m+1/2}^n \frac{T_{m+1}^n - T_m^n}{h} \right), \\ \Pi_{m-1/2}^{n+1/2} &= \frac{1}{2} \left(k_{m-1/2}^{n+1} \frac{T_m^{n+1} - T_{m-1}^{n+1}}{h} + k_{m-1/2}^n \frac{T_m^n - T_{m-1}^n}{h} \right). \end{aligned}$$

Полученную разностную схему можно переписать в виде

$$\begin{aligned} \frac{T_m^{n+1} - T_m^n}{\tau} - \frac{1}{2h} \left[\left(k_{m+1/2} \frac{T_{m+1} - T_m}{h} - k_{m-1/2} \frac{T_m - T_{m-1}}{h} \right)^{n+1} + \right. \\ \left. + \left(k_{m+1/2} \frac{T_{m+1} - T_m}{h} - k_{m-1/2} \frac{T_m - T_{m-1}}{h} \right)^n \right] = 0. \quad (13.4) \end{aligned}$$

Заметим, что в случае постоянного коэффициента теплопроводности эта разностная схема имеет следующий вид:

$$\frac{T_m^{n+1} - T_m^n}{\tau} - k \frac{T_{m-1}^{n+1} - 2T_m^{n+1} + T_{m+1}^{n+1}}{h^2} = 0; \quad (13.5)$$

ее шаблон показан на рис. 13.2. Алгоритм решения такого разностного уравнения — прогонка.

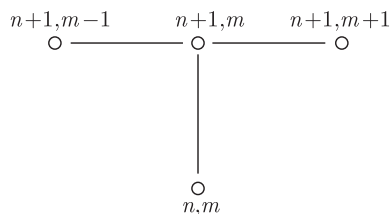


Рис. 13.2

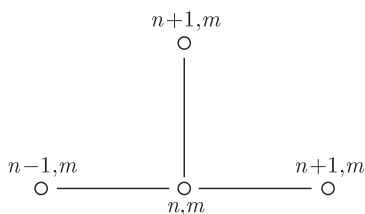


Рис. 13.3

Явная схема для численного решения рассматриваемого уравнения будет иметь вид:

$$\frac{T_m^{n+1} - T_m^n}{\tau} - k \frac{T_{m-1}^n - 2T_m^n + T_{m+1}^n}{h^2} = 0, \quad (13.6)$$

ее шаблон показан на рис. 13.3.

Эти схемы также представимы в операторной форме, соответственно:

$$T_m^{n+1} = T_m^n + \tau k \Lambda_{xx} T_m^{n+1} = (E + \tau k \Lambda_{xx}) T_m^{n+1},$$

$$T_m^{n+1} = T_m^n + \tau k \Lambda_{xx} T_m^n = (E + \tau k \Lambda_{xx}) T_m^n,$$

где Λ_{xx} — разностный оператор вида

$$\Lambda_{xx} T_m^n = \frac{T_{m-1}^n - 2T_m^n + T_{m+1}^n}{h^2}.$$

Алгоритмическая реализация явной разностной схемы — «бегущий счет», т. е. проведение расчета по рекуррентной формуле на каждом временном слое n :

$$T_m^1 = T_m^0 + \tau k \Lambda_{xx} T_m^0, \quad T_m^0 = \Psi_m,$$

$$T_m^2 = T_m^1 + \tau k \Lambda_{xx} T_m^1,$$

$$\dots\dots\dots,$$

$$T_m^{n+1} = T_m^n + \tau k \Lambda_{xx} T_m^n.$$

Неявная схема устойчива при любом соотношении τ, h ; явная схема условно устойчива при $\tau \leq h^2/(2k)$. Последнее условие представляется довольно жестким для выбора временного шага, поэтому для численного решения рассматриваемых задач обычно используются неявные схемы. Приведем еще одну часто используемую шеститочечную разностную схему Кранка–Никольсон:

$$\frac{T_m^{n+1} - T_m^n}{\tau} = \eta \Lambda_{xx} T_m^{n+1} + (1 - \eta) \Lambda_{xx} T_m^n, \quad (13.7)$$

$0 \leq \eta \leq 1$, т. е. при $\eta = 0$ мы имеем явную, при $\eta = 1$ — неявную разностную схему, имеющую второй порядок аппроксимации по координате, и первый — по времени: $O(\tau^2 + h^2)$.

Выражение невязки в этом случае будет иметь следующий вид:

$$\xi_\tau = \left[\tau k^2 \left(\eta - \frac{1}{2} \right) + \frac{k h^2}{12} \right] u_x^{(4)} + \frac{\tau^2}{8} \left(k u_{ttxx}^{(4)} - \frac{1}{3} u_x^{(3)} \right) + O(\tau^3, h^4),$$

причем при $\eta = 1/2$ порядок аппроксимации рассматриваемой схемы: $O(\tau^2 + h^2)$.

Отметим, что при $\eta = \frac{1}{2} - \frac{h^2}{12a\tau}$: $\xi_\tau = O(\tau^2 + h^4)$; при $\eta \neq \frac{1}{2}$, $\eta \neq \frac{1}{2} - \frac{h^2}{12a\tau}$ порядок аппроксимации равен $O(\tau + h^2)$. Если неявную схему с весами (т. е. при $\eta \geq 1/2$) представить в виде

$$a_m T_{m-1}^{n+1} + b_m T_m^{n+1} + c_m T_{m+1}^{n+1} = f_m^n,$$

где

$$a_m = \frac{k\eta\tau}{h^2}, \quad b_m = -1 - \frac{k\eta\tau}{h^2}, \quad c_m = \frac{k\eta\tau}{h^2},$$

то несложно увидеть, что для данной трехдиагональной системы линейных алгебраических уравнений, решаемой методом прогонки, выполняется условие диагонального преобладания при любых τ .

Исследование данной схемы на спектральную устойчивость дает следующее условие устойчивости:

$$\eta \geq \frac{1}{2} - \frac{h^2}{4k\tau}.$$

Две предыдущие схемы (13.5), (13.6) имели второй порядок аппроксимации по координате и первый по времени $O(\tau + h^2)$. Так, исследование на аппроксимацию явной разностной схемы дает невязку следующего вида:

$$L_\tau T_\tau = LT|_{t_n, x_m} + \frac{\tau}{2} T''_{tt} - \frac{ah^2}{12} T^{(4)}_x + O(\tau^2 + h^4). \quad (13.8)$$

Если учесть дифференциальные следствия линейного одномерно-го одномерного уравнения теплопроводности:

$$T'_t = kT''_{tt}; \quad T''_{tt} = k(T'_t)_{xx}; \quad T''_{tt} = k^2 T^{(4)}_x,$$

подставить это выражение в правую часть (13.8) и пренебречь слагаемым $O(\tau^2 + h^4)$, то получим явную разностную схему на пятиточечном шаблоне:

$$\frac{T^{n+1}_m - T^n_m}{\tau} = k\Lambda_{xx} T^n_m - \frac{k^2\tau}{2} \left(1 - \frac{1}{6} \frac{h^2}{\tau}\right) \Lambda_{xxx} T^n_m, \quad (13.9)$$

где

$$\Lambda_{xxx} T^n_m = \tau^{-4} (u^n_{m-2} - 4u^n_{m-1} + 6u^n_m - 4u^n_{m+1} + u^n_{m+2}),$$

имеющую второй порядок точности по времени и четвертый по координате: $O(\tau^2 + h^4)$. Аналогичную процедуру можно провести и для неявной разностной схемы, которая будет устойчивой при любом соотношении τ, h .

Представим также трехслойную разностную схему, аппроксимирующую рассматриваемое уравнение вида

$$\frac{1}{2} \left(\frac{T^{n+1}_m - T^n_m}{\tau} - \frac{T^n_m - T^{n-1}_m}{\tau} \right) = k \left(\frac{T^{n+1}_{m-1} - 2T^{n+1}_m + T^{n+1}_{m+1}}{h^2} \right),$$

устойчивую при любых τ и h , обладающую порядком аппроксимации $O(\tau^2 + h^2)$, монотонную (заметим, что схема Кранка–Никольсон не монотонна); шаблон этой схемы представлен на рис. 13.4.

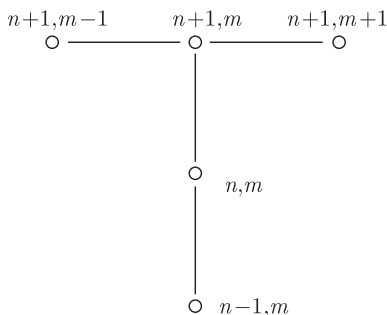


Рис. 13.4

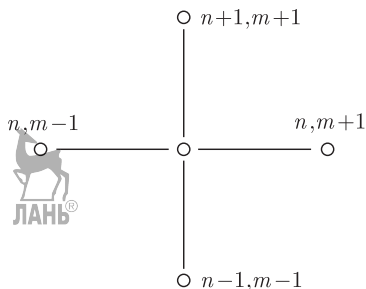


Рис. 13.5

Полезным также представляется исследование на аппроксимацию разностной схемы «крест»:

$$\frac{T_m^{n+1} - T_m^n}{2\tau} = \frac{1}{h^2} \left(T_{m-1}^n - 2 \frac{T_m^{n+1} + T_m^{n-1}}{2} + T_{m+1}^n \right),$$

шаблон которой представлен на рис. 13.5.

Это связано с попытками построения неявных схем бегущего счета для уравнения теплопроводности. Исследование этой схемы на устойчивость приводит к квадратному уравнению для λ вида

$$\lambda^2 - \frac{2 \cos \alpha}{1 + \sigma} \lambda + \frac{1 - \sigma}{1 + \sigma} = 0, \quad \sigma = \frac{h^2}{2\tau},$$

анализ которого приводит к результату:

$$|\lambda_1| \leq 1, \quad |\lambda_2| \leq 1,$$

т.е. схема безусловно устойчива. Однако при ее исследовании на аппроксимацию в выражении для погрешности появляется член $O(\tau^2/h^2)$, что ограничивает ее применение жестким условием на шаг по времени: $\tau \sim h^2$.

13.2. Нелинейное одномерное уравнение теплопроводности

Рассмотрим нелинейное нестационарное уравнение теплопроводности

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial x} \left(k(T) \frac{\partial T}{\partial x} \right) + f(T) \quad (13.10)$$

и аппроксимируем его разностную схему

$$\frac{T_m^{n+1} - T_m^n}{\tau} = h^{-1} \left(k_{m+1/2} \frac{T_{m+1}^{n+1} - T_m^{n+1}}{h} - k_{m-1/2} \frac{T_m^{n+1} - T_{m-1}^{n+1}}{h} \right) + f_m^n, \quad (13.11)$$

где коэффициент теплопроводности может быть вычислен, например, в виде

$$k_{m+1/2} = \frac{1}{2} [k(T_m^n) + k(T_{m+1}^n)], \quad (13.12a)$$

или

$$k_{m+1/2} = k\left(\frac{T_m^n + T_{m+1}^n}{2}\right). \quad (13.12б)$$

Очевидно, что алгоритм прогонки в данном случае применим, поскольку все данные берутся с нижнего временного слоя.

Однако метод простых итераций с нелинейностью на нижнем слое имеет ограничение на шаг по времени:

$$\tau \|f'_u\| \ll 1.$$

Поэтому если значение $\|f'_u\|$ достаточно велико, то в этом случае имеет смысл использовать метод простых итераций с нелинейностью на верхнем временном слое:

$$\frac{T_m^{n+1} - T_m^n}{\tau} = h^{-1} \left(k_{m+1/2}^{n+1} \frac{T_{m+1}^{n+1} - T_m^{n+1}}{h} - k_{m-1/2}^{n+1} \frac{T_m^{n+1} - T_{m-1}^{n+1}}{h} \right) + f_m(T_m^{n+1}). \quad (13.13)$$

Сначала рассмотрим метод простых итераций в функциональном пространстве и разностную схему с нелинейностью на нижнем временном слое, которую представим в следующем виде.

Поскольку $k_{m+1/2} = k(T_{m+1/2}^{n+1})$, $k_{m-1/2} = k(T_{m-1/2}^{n+1})$, $f_m^n = f(T_m^{n+1})$, то метод прогонки, как уже отмечалось, неприменим. Поэтому построим следующий итерационный процесс:

$$\frac{T_m^{i+1} - T_m^i}{\tau} = h^{-1} \left(k T_{m+1/2}^i \frac{T_{m+1}^{i+1} - T_m^{i+1}}{h} - k T_{m-1/2}^i \frac{T_m^{i+1} - T_{m-1}^{i+1}}{h} \right) + f(T_m^i). \quad (13.14)$$

Задав начальное приближение

$$T_m^0 = T_m^n,$$

где T_m^n — численное решение уравнения теплопроводности на нижнем слое ($t = t_n$), можно воспользоваться методом прогонки для $i = 0, 1, \dots$ и т. д. до достижения заданной точности: например,

$$\|T_m^{i+1} - T_m^i\| \leq \varepsilon. \quad (13.15)$$

Поскольку для достаточно гладкой функции $T(t, x)$ разность в значениях искомой функции на нижнем и верхнем временном слоях не очень велика, то начальное приближение оказывается достаточно близким к решению итерационного уравнения, поэтому для выполнения условия не требуется много итераций. Заметим также, что в отличие от итерационного решения алгебраического уравнения, где ищется один корень, в рассматриваемом случае мы вычисляем таблицу $T(t_n, x_m)$, т.е. в конечном итоге, используя интерполяцию значений функций между узлами расчетной сетки, находим функцию $T(t, x)$. Мы также можем построить другой итерационный метод для численного решения рассматриваемой задачи, например:

$$\frac{T_m^{i+1} - T_m^i}{\tau} = h^{-1} \left(k_{m+1/2}^i \frac{T_{m+1}^{i+1} - T_m^{i+1}}{h} - k_{m-1/2}^i \frac{T_m^{i+1} - T_{m-1}^{i+1}}{h} \right) + f(T_m^{i+1}). \quad (13.16)$$

Однако в этом случае метод прогонки оказывается неприменим, поскольку правая часть вычисляется на $(i+1)$ -й итерации. Выходом из данной ситуации оказывается метод квазилинеаризации, или метод Ньютона в функциональных пространствах. Для его реализации достаточно линеаризовать правую часть, что позволяет воспользоваться методом прогонки:

$$f_m^{i+1} = f[T_m^i + (T_m^{i+1} - T_m^i)] \approx f_m^i + f'_T(T_m^i)(T_m^{i+1} - T_m^i), \quad (13.17)$$

где $f_m^i = f(T_m^i)$.

В этом случае итерационный процесс будет иметь вид:

$$\frac{T_m^{i+1} - T_m^i}{\tau} = h^{-1} \left(k_{m+1/2}^i \frac{T_{m+1}^{i+1} - T_m^{i+1}}{h} - k_{m-1/2}^i \frac{T_m^{i+1} - T_{m-1}^{i+1}}{h} \right) + f_m^i + f'_T(T_m^i)(T_m^{i+1} - T_m^i) \quad (13.18)$$

и мы можем применить для вычисления искомой сеточной функции метод прогонки. Заметим, что коэффициент теплопроводности, в случае его существенной зависимости от температуры, также можно линеаризовать:

$$k(T_m^{i+1}) = k[T_m^i + (T_m^{i+1} - T_m^i)] \approx k(T_m^i) + k'_T(T_m^i)(T_m^{i+1} - T_m^i). \quad (13.19)$$

Рассмотренная выше шеститочечная разностная схема Кранка–Никольсон в нелинейном виде будет иметь такой вид:

$$\begin{aligned} \frac{T_m^{i+1} - T_m^n}{\tau} = & \\ = (2h)^{-1} & \left[\left(k_{m+1/2}^i \frac{T_{m+1}^{i+1} - T_m^{i+1}}{h} - k_{m-1/2}^i \frac{T_m^{i+1} - T_{m-1}^{i+1}}{h} \right) + \right. \\ & + \left. \left(k_{m+1/2}^n \frac{T_{m+1}^n - T_m^n}{h} - k_{m-1/2}^n \frac{T_m^n - T_{m-1}^n}{h} \right) \right] + \\ & + \left[f_m^i + (f_T')_m^i (T_m^{i+1} - T_m^n) \right], \end{aligned}$$

где i — итерационный индекс.



13.3. Методы расщепления для численного решения многомерных уравнений параболических типа

В случае двух пространственных переменных уравнение теплопроводности имеет следующий вид:

$$\frac{\partial T}{\partial t} = k \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right), \quad (13.20)$$

$$0 < t \leq \Theta; \quad 0 < x < X, \quad 0 < y < Y.$$

Здесь $T(t, x, y)$ — температура среды, $k = \text{const} > 0$ — коэффициент теплопроводности, начальные и граничные условия в случае прямоугольной области интегрирования представляются в виде (рис. 13.6):

$$\begin{aligned} T(0, x, y) &= \psi(x, y), \quad t = 0; \quad 0 < x < X; \quad 0 < y < Y; \\ T(t, 0, y) &= \varphi_1(t, y), \quad 0 < t \leq \Theta, \quad x = 0, \quad 0 < y < Y; \\ T(t, X, y) &= \varphi_2(t, y), \quad 0 < t \leq \Theta; \quad x = X, \quad 0 < y < Y; \\ T(t, x, 0) &= \varphi_3(t, x), \quad 0 < t \leq \Theta, \quad 0 < x < X = 0; \quad y = 0; \\ T(t, x, Y) &= \varphi_4(t, x), \quad 0 < t \leq \Theta, \quad 0 < x < X = 0; \quad y = Y. \end{aligned} \quad (13.21)$$

Введем для разностной аппроксимации (13.21) расчетную сетку

$$\begin{aligned} \omega_{NM} = \left\{ t = n\tau; \quad n = 0, 1, \dots, N; \quad \tau = \frac{\Theta}{N}; \quad x_m = mh_x; \right. \\ \left. m = 0, 1, \dots, M; \right. \\ \left. h_x = \frac{X}{M}; \quad y_l = lh_y; \quad l = 0, 1, \dots, M; \quad h_y = \frac{Y}{L} \right\} \end{aligned} \quad (13.22)$$



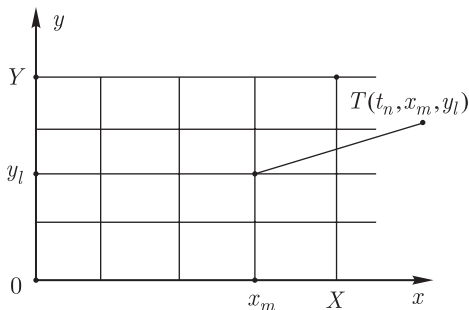


Рис. 13.6

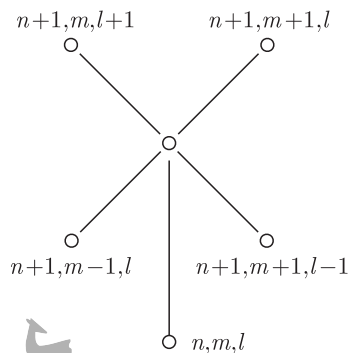


Рис. 13.7

и аппроксимируем двумерное уравнение теплопроводности на шаблоне, который имеет следующий вид (рис. 13.7):

$$\frac{T_{ml}^{n+1} - T_{ml}^n}{\tau} = k \left(\frac{T_{m+1,l}^{n+1} - 2T_{ml}^{n+1} + T_{m-1,l}^{n+1}}{h_x^2} + \frac{T_{m,l+1}^{n+1} - 2T_{ml}^{n+1} + T_{m,l-1}^{n+1}}{h_y^2} \right), \quad (13.23)$$

или, с использованием операторных обозначений:

$$\frac{T_{ml}^{n+1} - T_{ml}^n}{\tau} = k (\Lambda_{xx} T_{ml}^{n+1} + \Lambda_{yy} T_{ml}^{n+1}), \quad (13.24)$$

где

$$\Lambda_{xx} T_{ml}^{n+1} = \frac{T_{m-1,l}^{n+1} - 2T_{ml}^{n+1} + T_{m+1,l}^{n+1}}{h_x^2},$$

$$\Lambda_{yy} T_{ml}^{n+1} = \frac{T_{m,l-1}^{n+1} - 2T_{ml}^{n+1} + T_{m,l+1}^{n+1}}{h_y^2},$$

или

$$\frac{\Delta T_{ml}^{n+1}}{\tau} = k \Lambda T_{ml}^{n+1},$$

где

$$\Delta T_{ml}^{n+1} = T_{ml}^{n+1} - T_{ml}^n, \quad \Lambda = \Lambda_{xx} + \Lambda_{yy}.$$

Явная разностная схема, аппроксимирующая двумерное нестационарное уравнение теплопроводности, представляется в следующем виде:

$$\frac{T_{ml}^{n+1} - T_{ml}^n}{\tau} = k (\Lambda_{xx} T_{ml}^n + \Lambda_{yy} T_{ml}^n). \quad (13.25)$$

Ее шаблон представлен на рис. 13.8.

Исследование явной схемы на устойчивость проводится путем подстановки решения в виде

$$T_{ml}^n = \lambda^n e^{i\alpha m + i\beta l}, \quad (13.26)$$

$$\alpha, \beta \in [0, 2\pi],$$

после чего получим

$$\lambda(\alpha) = 1 - 4 \frac{k\tau}{h_x^2} \sin^2 \frac{\alpha}{2} - 4 \frac{k\tau}{h_y^2} \sin^2 \frac{\beta}{2},$$

откуда следует, что рассматриваемая явная разностная схема будет устойчивой при выполнении условия

$$\tau \leq [2k(h_x^{-2} + h_y^{-2})]^{-1}. \quad (13.27)$$

Для неявной схемы получим

$$\lambda = \left[1 + 4 \frac{k\tau}{h_x^2} \sin^2 \frac{\alpha}{2} + 4 \frac{k\tau}{h_y^2} \sin^2 \frac{\beta}{2} \right]^{-1}, \quad (13.28)$$

откуда следует, что неявная схема, аппроксимирующая нестационарное уравнение теплопроводности, будет устойчивой при любом соотношении сеточных параметров: τ, h_x, h_y .

Что касается условия устойчивости явной схемы, то оно является довольно жестким для выбора временного шага:

$$\tau \sim h_x^2, \quad \tau \sim h_y^2;$$

по этой причине при численном решении как многомерных, так и одномерных дифференциальных уравнений параболического типа, чаще всего используются неявные схемы.

Исследование обеих рассматриваемых разностных схем на аппроксимацию путем разложения сеточных функций в ряд Тейлора дает невязку следующего вида: $O(\tau + h_x^2 + h_y^2)$, т.е. обе схемы имеют первый порядок аппроксимации по времени и второй — по каждой из координат.

Алгоритмическая реализация явной разностной схемы (13.25), как и в одномерном случае, представляет собой «бегущий счет» — вычисление значений сеточной функции по рекуррентной формуле

$$T_{ml}^{n+1} = T_{ml}^n + \tau k \Lambda T_{ml}^n, \quad (13.29)$$

т.е. вычислительный алгоритм использует два цикла: первый — по m (по координате) от левой до правой границы области интегрирования, второй — по n (значение сеточной функции T_{ml}^{n+1} на верхней временном слое вычисляется по известным ее значениям на нижнем).

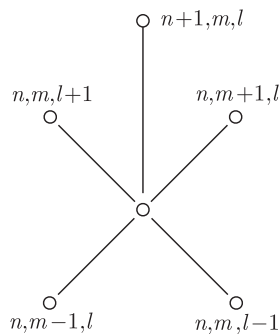


Рис. 13.8

В неявной разностной схеме (13.23)

$$T_{ml}^{n+1} = T_{ml}^n + \tau k \Lambda T_{ml}^{n+1}, \quad (13.30)$$

на верхнем слое имеем 5 неизвестных; причем матрица получившейся системы линейных алгебраических уравнений не является трех- или пятидиагональной, т. е. для ее решения мы не сможем воспользоваться методом прогонки. Поскольку матрица сильно разрежена, то использование метода Гаусса представляется нерациональным.

Писмэном, Рэкфордом и Дугласом в 1955 г. [4, 6] была предложена схема переменных направлений (или продольно-поперечная схема, схема чередующихся направлений), идея которой состоит в реализации вычислительного алгоритма в два этапа (схема типа предиктор-корректор). На первом этапе (полушаге $\tau/2$) реализуется неявная аппроксимация координатной производной по первому направлению и явная — по второму, на втором этапе — наоборот, причем оба этапа аппроксимируют исходное дифференциальное уравнение:

$$\frac{T_{ml}^{n+1/2} - T_{ml}^n}{\tau/2} = \Lambda_{xx} T_{ml}^{n+1/2} + \Lambda_{yy} T_{ml}^n, \quad (13.31)$$

$$\frac{T_{ml}^{n+1} - T_{ml}^{n+1/2}}{\tau/2} = \Lambda_{xx} T_{ml}^{n+1/2} + \Lambda_{yy} T_{ml}^{n+1}. \quad (13.32)$$

В таком случае на каждом этапе можно использовать метод трехточечной прогонки.

Представим полученную двухэтапную разностную схему в операторном виде:

$$\left(E - \frac{\tau}{2} \Lambda_{xx}\right) T_{ml}^{n+1/2} - \left(E + \frac{\tau}{2} \Lambda_{yy}\right) T_{ml}^n = 0, \quad (13.33)$$

$$\left(E - \frac{\tau}{2} \Lambda_{yy}\right) T_{ml}^{n+1} - \left(E + \frac{\tau}{2} \Lambda_{xx}\right) T_{ml}^{n+1/2} = 0, \quad (13.34)$$

где E — единичный (тождественный) оператор. Подействуем на первое из представленных разностных уравнений слева оператором $\left(E + \frac{\tau}{2} \Lambda_{yy}\right)$, а на второе — оператором $\left(E - \frac{\tau}{2} \Lambda_{xx}\right)$ и сложим:

$$\begin{aligned} & \left(E + \frac{\tau}{2} \Lambda_{yy}\right) \left(E - \frac{\tau}{2} \Lambda_{xx}\right) T_{ml}^{n+1} - \left(E + \frac{\tau}{2} \Lambda_{yy}\right) \left(E + \frac{\tau}{2} \Lambda_{xx}\right) T_{ml}^n + \\ & + \left[\left(E + \frac{\tau}{2} \Lambda_{xx}\right) \left(E - \frac{\tau}{2} \Lambda_{xx}\right) - \left(E - \frac{\tau}{2} \Lambda_{xx}\right) \left(E + \frac{\tau}{2} \Lambda_{xx}\right)\right] \times \\ & \times T_{ml}^{n+1/2} = 0. \end{aligned}$$

В предположении коммутативности операторов $\left(E + \frac{\tau}{2}\Lambda_{xx}\right)$ и $\left(E - \frac{\tau}{2}\Lambda_{xx}\right)$ получим равенство

$$\left(E - \frac{\tau}{2}\Lambda_{yy}\right)\left(E - \frac{\tau}{2}\Lambda_{xx}\right)T_{ml}^{n+1} - \left(E + \frac{\tau}{2}\Lambda_{yy}\right)\left(E + \frac{\tau}{2}\Lambda_{xx}\right)T_{ml}^n = 0,$$

из которого следует

$$\frac{T_{ml}^{n+1} - T_{ml}^n}{\tau} = \frac{1}{2}(\Lambda_{xx} + \Lambda_{yy})(T_{ml}^n + T_{ml}^{n+1}) - \frac{\tau^2}{4}\Lambda_{xx}\Lambda_{yy}\frac{(T_{ml}^{n+1} - T_{ml}^n)}{\tau}. \quad (13.35)$$

Последнее разностное уравнение аппроксимирует исходное уравнение теплопроводности со вторым порядком по времени и координатам. Отметим, что каждое из разностных уравнений (13.31) и (13.32) аппроксимирует исходное дифференциальное уравнение в частных производных с первым порядком аппроксимации по τ и вторым — по h . При их совместном использовании порядок аппроксимации будет $O(\tau^2 + h^2)$.

Проведем исследование разностной продольно-поперечной схемы на устойчивость с помощью спектрального признака Неймана, для чего положим:

$$T_{ml}^n = \lambda^{i\alpha m + i\beta l}, \quad (13.36)$$

$$T_{ml}^{n+1/2} = \lambda_1 T_{ml}^n, \quad T_{ml}^{n+1} = \lambda_2 T_{ml}^{n+1/2} = (\lambda_1 \lambda_2) T_{ml}^n.$$

Подставим (13.36) в (13.32), получим выражения для λ_1 и λ_2 :

$$\lambda_1 = \frac{1 - 4\frac{\tau}{h_x^2}\sin^2\frac{\alpha}{2}}{1 + 4\frac{\tau}{h_y^2}\sin^2\frac{\alpha}{2}}, \quad \lambda_2 = \frac{1 - 4\frac{\tau}{h_x^2}\sin^2\frac{\beta}{2}}{1 + 4\frac{\tau}{h_y^2}\sin^2\frac{\beta}{2}},$$

откуда получим

$$\lambda = \lambda_1 \lambda_2 = \frac{\left(1 - 4\frac{\tau}{h_x^2}\sin^2\frac{\alpha}{2}\right)\left(1 - 4\frac{\tau}{h_x^2}\sin^2\frac{\beta}{2}\right)}{\left(1 + 4\frac{\tau}{h_y^2}\sin^2\frac{\alpha}{2}\right)\left(1 + 4\frac{\tau}{h_y^2}\sin^2\frac{\beta}{2}\right)}. \quad (13.37)$$

Видно, что $|\lambda| \leq 1$, т.е. продольно-поперечная схема устойчива при любых соотношениях τ, h_x, h_y .

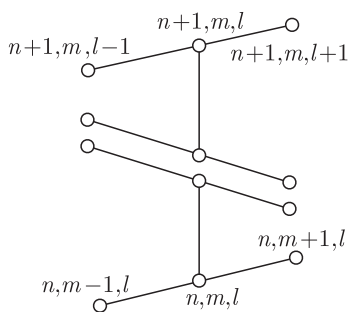


Рис. 13.9

Пространственный шаблон продольно-поперечной схемы имеет вид, представленный на рис. 13.9.

В книге [6] Н. Н. Яненко предложил отказаться от требования аппроксимации исходного дифференциального уравнения на промежуточных шагах. Такие схемы получили название *метода дробных шагов, локально-одномерных схем* (по А. А. Самарскому [4]), или *схем расщепления по направлениям*.

Рассмотрим следующую двухэтапную разностную аппроксимацию нестационарного двумерного уравнения теплопроводности (13.20):

$$\frac{T_{ml}^{n+1/2} - T_{ml}^n}{\tau} = \Lambda_{xx} T_{ml}^{n+1/2}, \quad (13.38)$$

$$\frac{T_{ml}^{n+1} - T_{ml}^{n+1/2}}{\tau} = \Lambda_{yy} T_{ml}^{n+1}, \quad (13.39)$$

и представим ее в виде

$$\begin{cases} \left(E - \frac{\tau}{2} \Lambda_{xx}\right) T_{ml}^{n+1/2} = T_{ml}^n, \\ \left(E - \frac{\tau}{2} \Lambda_{yy}\right) T_{ml}^{n+1} = T_{ml}^{n+1/2}. \end{cases} \quad (13.40)$$

Исключив в (13.40) $T_{ml}^{n+1/2}$, придем к эквивалентной одноэтапной схеме следующего вида:

$$\frac{T_{ml}^{n+1} - T_{ml}^n}{\tau} = (\Lambda_{xx} + \Lambda_{yy}) T_{ml}^{n+1} - \tau \Lambda_{xx} \Lambda_{yy} T_{ml}^{n+1},$$

откуда следует, что порядок аппроксимации исходного дифференциального уравнения будет $O(\tau + h_x^2 + h_y^2)$.

Поскольку схема дробных шагов неявная, то она устойчива при любых τ, h_x, h_y , что также можно показать с помощью спектрального признака Неймана.

Пространственный шаблон схемы представлен на рис. 13.10.

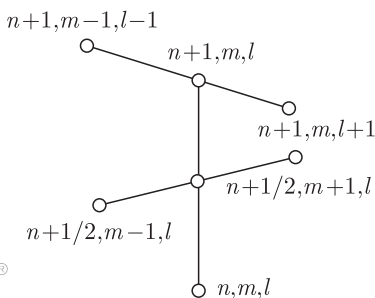


Рис. 13.10

Представим трехмерный вариант схемы дробных шагов:

$$\begin{cases} \frac{u_{mlp}^{n+1/3} - u_{mlp}^n}{\tau} = \Lambda_{xx} u_{mlp}^{n+1/3}, \\ \frac{u_{mlp}^{n+2/3} - u_{mlp}^{n+1/3}}{\tau} = \Lambda_{yy} u_{mlp}^{n+2/3}, \\ \frac{u_{mlp}^{n+1} - u_{mlp}^{n+2/3}}{\tau} = \Lambda_{zz} u_{mlp}^{n+1}. \end{cases} \quad (13.41)$$

Схема устойчива при любых τ, h_x, h_y, h_z и аппроксимирует трехмерное дифференциальное нестационарное уравнение теплопроводности

$$\frac{\partial T}{\partial t} = k \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right) \quad (13.42)$$

с первым порядком по времени и вторым — по координатам: $O(\tau + h_x^2 + h_y^2 + h_z^2)$. Локально-одномерную схему с весовым коэффициентом η (Кранка–Никольсон) можно представить в следующем виде:

$$\begin{cases} \frac{T_{ml}^{n+1/2} - T_{ml}^n}{\tau/2} = \Lambda_{xx} \left(\eta T_{ml}^{n+1/2} + (1-\eta) T_{ml}^n \right), \\ \frac{T_{ml}^{n+1} - T_{ml}^{n+1/2}}{\tau/2} = \Lambda_{yy} \left(\eta T_{ml}^{n+1} + (1-\eta) T_{ml}^{n+1/2} \right). \end{cases} \quad (13.43)$$

Здесь $0 \leq \eta \leq 1$. Эта схема будет устойчивой при любых τ, h_x, h_y ; при $\eta \geq 1/2$ она имеет второй порядок аппроксимации по времени и по координатам: $O(\tau^2 + h_x^2 + h_y^2)$.

Ее шаблон представлен на рис. 13.11.

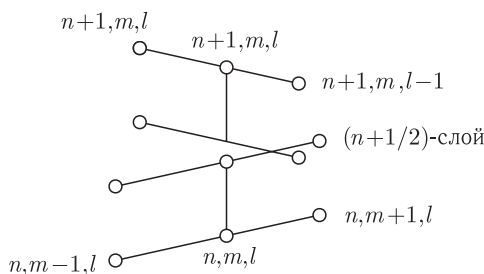


Рис. 13.11

Трехмерный вариант разностной схемы Кранка–Никольсон представляется следующим образом:

$$\begin{cases} \frac{u_{mlp}^{n+1/3} - u_{mlp}^n}{\tau} = \Lambda_{xx} \left(\eta u_{mlp}^{n+1/3} + (1 - \eta) u_{mlp}^n \right), \\ \frac{u_{mlp}^{n+2/3} - u_{mlp}^{n+1/3}}{\tau/2} = \Lambda_{yy} \left(\eta u_{mlp}^{n+2/3} + (1 - \eta) u_{mlp}^{n+1/3} \right), \\ \frac{u_{mlp}^{n+1} - u_{mlp}^{n+2/3}}{\tau/2} = \Lambda_{zz} \left(\eta u_{mlp}^{n+1} + (1 - \eta) u_{mlp}^{n+2/3} \right). \end{cases} \quad (13.44)$$

Аналогично двумерному случаю, схема будет устойчивой при любых шагах по времени и пространству, если $\eta \geq 1/2$, при $\eta = 1/2$ невязка равна $O(\tau^2 + h_x^2 + h_y^2 + h_z^2)$.

Рассмотрим также использующуюся в приложениях трехмерную схему Дугласа–Гана, имеющую следующий вид:

$$\begin{cases} \frac{u_{mlp}^{n+1/3} - u_{mlp}^n}{\tau} = \frac{1}{2} \Lambda_{xx} \left(u_{mlp}^{n+1/3} + u_{mlp}^n \right) + \\ \quad + \Lambda_{yy} u_{mlp}^n + \Lambda_{zz} u_{mlp}^n, \\ \frac{u_{mlp}^{n+2/3} - u_{mlp}^{n+1/3}}{\tau} = \frac{1}{2} \Lambda_{xx} \left(u_{mlp}^{n+2/3} + u_{mlp}^{n+1/3} \right) + \\ \quad + \frac{1}{2} \Lambda_{yy} \left(u_{mlp}^{n+2/3} + u_{mlp}^{n+1/3} \right) + \Lambda_{zz} u_{mlp}^n, \\ \frac{u_{mlp}^{n+1} - u_{mlp}^{n+2/3}}{\tau} = \frac{1}{2} \Lambda_{xx} \left(u_{mlp}^{n+1} + u_{mlp}^{n+2/3} \right) + \\ \quad + \frac{1}{2} \Lambda_{yy} \left(u_{mlp}^{n+1} + u_{mlp}^{n+2/3} \right) + \frac{1}{2} \Lambda_{zz} \left(u_{mlp}^{n+1} + u_{mlp}^{n+2/3} \right). \end{cases} \quad (13.45)$$

Двумерный аналог (рис. 13.12) этой схемы будет таким:

$$\begin{cases} \frac{u_{ml}^{n+1/2} - u_{ml}^n}{\tau} = \frac{1}{2} \Lambda_{xx} \left(u_{ml}^{n+1/2} + u_{ml}^n \right) + \Lambda_{yy} u_{ml}^n, \\ \frac{u_{ml}^{n+1} - u_{ml}^{n+1/2}}{\tau/2} = \frac{1}{2} \Lambda_{xx} \left(u_{ml}^{n+1} + u_{ml}^{n+1/2} \right) + \\ \quad + \frac{1}{2} \Lambda_{yy} \left(u_{ml}^{n+1} + u_{ml}^{n+1/2} \right). \end{cases} \quad (13.46)$$

В заключение заметим, что после появления методов расщепления по направлениям были предложены и получили широкое распространение методы расщепления по физическим процессам, суть которых можно изложить на примере уравнения

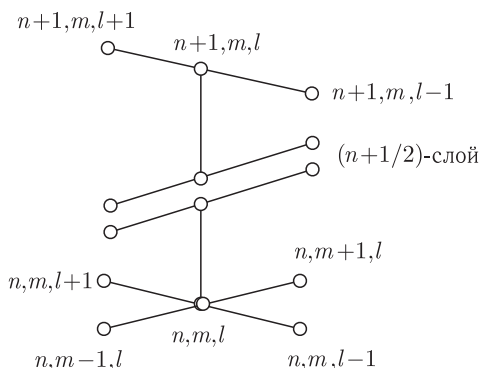


Рис. 13.12

конвекции–диффузии, часто используемом при решении задач экологии, метеорологии и др.:

$$\frac{\partial u}{\partial t} + \mu \left(\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} \right) = \nu \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right), \quad (13.47)$$

где $\mu = \text{const} > 0$, $\nu = \text{const} > 0$.

Положим, что разностная схема

$$\frac{u_{ml}^{n+1/2} - u_{ml}^n}{\tau/2} = \Lambda_1(u^n, u^{n+1/2}) \quad (13.48)$$

аппроксимирует уравнение конвекции вида

$$\frac{\partial u}{\partial t} + \mu \left(\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} \right) = 0,$$

а схема

$$\frac{u_{ml}^{n+1/2} - u_{ml}^n}{\tau/2} = \Lambda_2(u^{n+1}, u^{n+1}) \quad (13.49)$$

аппроксимирует уравнение диффузии

$$\frac{\partial u}{\partial t} = \nu \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right).$$

Можно показать, что двухэтапная схема расщепления (13.48), (13.49) аппроксимирует исходное дифференциальное уравнение в частных производных (13.47). Методы расщепления по физическим параметрам были предложены Ф. Х. Харлоу и О. М. Белоцерковским в [7, 8].

Список литературы

1. *Петров И.Б., Лобанов А.И.* Лекции по вычислительной математике. М.: БИНОМ. Лаборатория знаний, 2006. 522 с.
2. *Федоренко Р.П.* Введение в вычислительную физику. Долгопрудный: Интеллект, 2008. 503 с.
3. *Годунов С.К., Рябенький В.С.* Разностные схемы. М.: Наука, 1973. 400 с.

Дополнительная литература

4. *Самарский А.А.* Теория разностных схем. М.: Наука, 1983. 656 с.
5. *Ворожцов Е.В.* Разностные методы решения задач механики сплошных сред. Новосибирск: НГТУ, 1998. 83 с.
6. *Яненко Н.Н.* Метод дробных шагов решения многомерных задач математической физики. Новосибирск: Наука, 1967. 196 с.
7. *Харлоу Ф.Х.* Численный метод частиц в ячейках для задач гидродинамики // Вычислительные методы в гидродинамике. М.: Мир, 1967. С. 317–342.
8. *Белоцерковский О.М., Давыдов Ю.М.* Метод крупных частиц в газовой динамике. Вычислительный эксперимент. М.: Наука, 1982. 391 с.
9. *Самарский А.А., Гулин А.В.* Численные методы. М.: Наука, 1989. 430 с.





ЧИСЛЕННОЕ РЕШЕНИЕ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ ГИПЕРБОЛИЧЕСКОГО ТИПА

14.1. Двухслойные разностные схемы для численного решения линейного уравнения переноса

Наиболее простым уравнением гиперболического типа является скалярное линейное уравнение переноса

$$u'_t + au'_x = 0, \quad a = \text{const} > 0, \quad (14.1)$$

характеристиками которого являются прямые

$$\frac{dx}{dt} = a; \quad (14.2)$$

при этом само уравнение может быть представлено в виде обыкновенного дифференциального уравнения вдоль характеристического направления:

$$\frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} = \frac{dt}{dt} \cdot \frac{\partial u}{\partial t} + \frac{dx}{dt} \cdot \frac{\partial u}{\partial x} = \left(\frac{du}{dt} \right)_{\frac{dx}{dt}=a} = 0, \quad (14.3)$$

откуда следует, что на характеристиках значение искомой функции постоянно.

Обычно используются следующие постановки задач для уравнения переноса:

- а) задача Коши: $t > 0, -\infty < x < \infty$;
- б) краевая задача: $t > 0, 0 < x < \infty$ с начальным и краевым условиями:

$$\begin{aligned} u(0, x) &= \psi(x), \quad x \in [0, \infty), \\ u(t, 0) &= \varphi(t), \quad t \in [0, \infty), \end{aligned}$$

однако в вычислительной практике область интегрирования всегда конечна: $x \in [0, X], t \in [0, T]$.

Решением задачи Коши, как известно, из курса уравнений математической физики, является функция («бегущая волна»)

$$u(t, x) = \psi(x - at),$$

а краевой задачи — функция следующего вида:

$$u(t, x) = \begin{cases} \psi(x - at), & x \geq at \\ \varphi\left(t - \frac{x}{a}\right), & x < at. \end{cases} \quad (14.4)$$

Как и выше, в рассматриваемой области интегрирования вводится расчетная сетка:

$$\omega_{NM} = \left\{ t_n = n\tau; \quad n = 0, 1, \dots, N; \quad \tau = \frac{T}{N}; \quad x_m = mh, \right. \\ \left. m = 0, 1, \dots, M, \quad h = \frac{X}{M} \right\}.$$

Выбрав для построения аппроксимирующего разностного уравнения двухслойный шаблон

$$\{t^{n+i}, x_{m+j}\}, \quad i = 0, 1; \quad j = 0, \pm 1, \dots, J,$$

представим все возможные на нем линейные разностные схемы в виде

$$u_m^{n+1} = \sum_i \sum_j \alpha_j^i(\tau, h) u_{m+j}^{n+i}, \quad (14.5)$$

где α_j^i — коэффициенты, определяющие ту или иную разностную схему.

В [4] предлагается рассматривать все такие схемы в пространстве неопределенных коэффициентов α_j^i с целью придания им некоторых заданных свойств (например, порядка аппроксимации, монотонности и др.).

Заметим, что уравнение переноса (транспортное уравнение) — одно из фундаментальных уравнений математической физики, использующееся не только как самостоятельное уравнение, но и в системах уравнений механики и электродинамики сплошных сред, экологии, динамики разреженного газа и плазмы; в основном используется нелинейное уравнение вида

$$u'_t + uu'_x = 0, \quad (14.6)$$

или в дивергентной форме:

$$u'_t + f'_x = 0, \quad f = \frac{u^2}{2}. \quad (14.7)$$

Важным свойством для разностных схем, аппроксимирующих уравнение гиперболического типа, является монотонность.

Определение 14.1. Монотонными (по С. К. Годунову) называются линейные схемы, в которых для всех m выполняются неравенства:

$$\begin{aligned} u_{m+1}^{n+1} - u_m^{n+1} &\geq 0, & \text{если} & \quad u_{m+1}^n - u_m^n \geq 0; \\ u_{m+1}^{n+1} - u_m^{n+1} &\leq 0, & \text{если} & \quad u_{m+1}^n - u_m^n \leq 0. \end{aligned} \quad (14.8)$$

Теорема 14.1. Для того чтобы явная двухслойная линейная однородная разностная схема

$$u_m^{n+1} = \sum_j \alpha_j u_{m+j}^n$$

с постоянными коэффициентами α_j была монотонной, необходимо и достаточно, чтобы все ее коэффициенты были неотрицательны, т. е. $\alpha_j \geq 0$.

Теорема 14.2 (Годунова). Двухслойная линейная монотонная разностная схема, аппроксимирующая уравнение переноса

$$u'_t + au'_x = 0, \quad (14.9)$$

не может иметь порядок точности выше первого.

Для доказательства этих теорем необходимо рассмотреть главный член погрешности аппроксимации рассматриваемых разностных схем.

Представим наиболее известные двухслойные разностные схемы для линейного уравнения переноса, имеющие шаблоны, представленные на рис. 14.1–14.12 (14.1 — «левый уголок», схема Куранта–Изаксона–Риса, Годунова, 14.2 — «правый уголок», схема Куранта–Изаксона–Риса, Годунова, 14.3 — центральная четырехточечная схема Лакса–Вендроффа, 14.4 — неявная шеститочечная схема типа Кранка–Никольсона, 14.5 — «левый прямоугольник», схема Бабенко, 14.6 — «правый прямоугольник», схема Бабенко, 14.7 — центральная неявная четырехточечная схема, 14.8 — «неявный левый уголок», схема Карлсона, 14.9 — «неявный правый уголок», схема Ландау–Меймана–Халатникова, 14.10 — схема Бима–Уорминга, 14.11 — схема Фромма, 14.12 — схема Русанова).

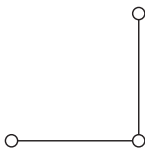


Рис. 14.1

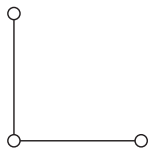


Рис. 14.2

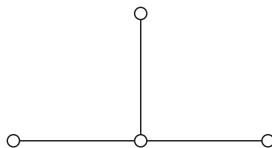


Рис. 14.3

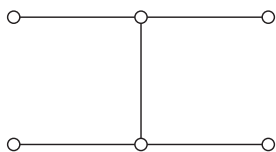


Рис. 14.4

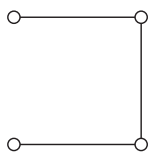


Рис. 14.5

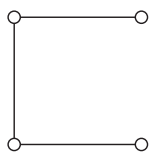


Рис. 14.6

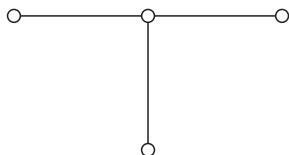


Рис. 14.7

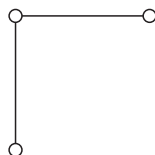


Рис. 14.8

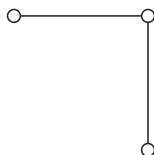


Рис. 14.9

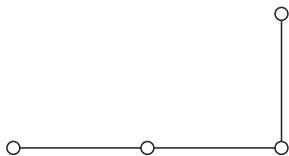


Рис. 14.10

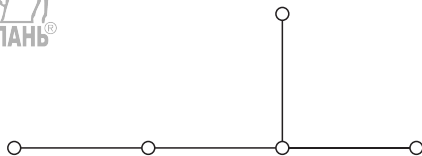


Рис. 14.11

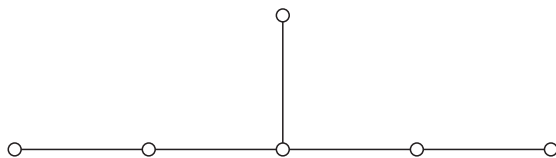


Рис. 14.12

По-видимому, наиболее простым методом построения разностных схем для численного решения уравнения переноса представляется разложение функции $u(t_{n+1}, x_m)$ в ряд Тейлора в окрестности узлов $\{t_n, x_m\}$ расчетной сетки:

$$\begin{aligned} u(t_{n+1} + \tau, x_m) &= \\ &= u(t_n, x_m) + \tau \cdot u'_t(t_n, x_m) + \frac{\tau^2}{2} u''_{tt}(t_n, x_m) + O(\tau^3). \end{aligned} \quad (14.10)$$

Поскольку из уравнения

$$u'_t = -au'_x$$

следует

$$u''_{tt} = -au'_{xt} = -a(u'_t)_x = a^2 u''_{xx}, \quad (14.11)$$

то, подставив (14.11) в разложение (14.10), получим

$$u_m^{n+1} = u_m^n - a\tau (u'_x)_m^n + \frac{a^2\tau^2}{2} (u''_{xx})_m^n + O(\tau^3). \quad (14.12)$$

Отсюда можно, отбросив последнее слагаемое и проведя аппроксимацию производных со вторым порядком точности:

$$(u'_x)_m^n \approx \frac{u_{m+1}^n - u_{m-1}^n}{2h}; \quad (u''_{xx})_m^n \approx \frac{u_{m-1}^n - 2u_m^n + u_{m+1}^n}{h^2},$$

получить схему Лакса–Вендроффа (СЛВ):

$$u_m^{n+1} = u_m^n - \frac{\sigma}{2} (u_{m+1}^n - u_{m-1}^n) + \frac{\sigma^2}{2} (u_{m-1}^n - 2u_m^n + u_{m+1}^n) \quad (14.13)$$

($\sigma = a\tau/h$), имеющую порядок аппроксимации $O(\tau^2 + h^2)$ и устойчивую при выполнении условия Куранта–Фридрихса–Леви (КФЛ):

$$\sigma = \frac{a\tau}{h} \leq 1,$$

шаблон представлен на рис. 14.3.

Эту же схему можно представить в следующем виде (с односторонней производной):

$$u_m^{n+1} = u_m^n - \sigma (u_m^n - u_{m-1}^n) - \frac{\sigma}{2} (1 - \sigma) (u_{m-1}^n - 2u_m^n + u_{m+1}^n), \quad (14.14)$$

или (в приращениях):

$$u_m^{n+1} = u_m^n - \sigma \Delta^- u_m^n - \frac{\sigma}{2} (1 - \sigma) \Delta^2 u_m^n,$$

где

$$\Delta^- u_m^n = u_m^n - u_{m-1}^n, \quad \Delta^+ u_m^n = u_{m+1}^n - u_m^n,$$

$$\Delta^2 u_m^n = u_{m-1}^n - 2u_m^n + u_{m+1}^n.$$

Исследование схемы Лакса–Вендроффа на аппроксимацию приводит к равенству:

$$\begin{aligned} \frac{u_m^{n+1} - u_m^n}{\tau} + \frac{a\tau}{2h} (u_{m+1}^n - u_{m-1}^n) - \frac{a\tau^2}{2h^2} (u_{m-1}^n - 2u_m^n + u_{m+1}^n) = \\ = (u'_t + au'_x)_{t,x,x_n} + \frac{ah^2}{6} (1 - \sigma^2) u''_{xx} + O(h^3). \end{aligned}$$

Если пренебречь членами третьего порядка малости, то получим так называемое *первое дифференциальное приближение*

$$u'_t + au'_x - \frac{ah^2}{6} (\sigma^2 - 1) u'''_x = 0,$$

или

$$Pu + \xi_\tau = 0,$$

где

$$Pu = u'_t + au'_x, \quad \xi_\tau = -\frac{ah^2}{6}(\sigma^2 - 1)u'''_x.$$

Смысл этого приближения в том, что реально решается численно не уравнение переноса, а уравнение, называемое *первым дифференциальным приближением*. Если оставить член третьего порядка малости, то получим второе дифференциальное приближение, четвертого порядка — третье и т.д. То есть первое дифференциальное приближение получается путем прибавления к исходному дифференциальному приближению главного члена ошибки аппроксимации, имеющего минимальный порядок малости. Исследование схемы на спектральную устойчивость дает уравнение для λ :

$$\frac{\lambda - 1}{\tau} - \frac{e^{i\alpha} - e^{-i\alpha}}{2h} - \frac{\tau}{2h}(e^{i\alpha} - 2 + e^{-i\alpha}) = 0,$$

откуда, учитывая известные соотношения:

$$(e^{i\alpha} + e^{-i\alpha})/(2i) = \sin \alpha,$$

$$(e^{i\alpha} - 2 + e^{-i\alpha})/4 = -\left[(e^{i\alpha/2} - e^{-i\alpha/2})/(2i)\right]^2 = -\sin^2(\alpha/2),$$

получим:

$$\lambda = 1 - i\sigma \sin \alpha - 2\sigma^2 \sin^2 \alpha/2,$$

$$|\lambda|^2 = (1 - 2\sigma^2 \sin^2 \alpha/2)^2 + \sigma^2 \sin^2 \alpha.$$

Условие устойчивости $|\lambda| \leq 1$ выполняется при $\sigma \leq 1$.

Если провести аппроксимацию производных следующим образом:

$$(u'_x)_m^n \approx \frac{3u_m^n - 4u_{m-1}^n + u_{m-2}^n}{2h}; \quad (u''_{xx})_m^n \approx \frac{u_m^n - 2u_{m-1}^n + u_{m-2}^n}{h^2},$$

то получим схему «парабола»:

$$u_m^{n+1} = u_m^n - \frac{\sigma}{2}(3u_m^n - 4u_{m-1}^n + u_{m-2}^n) + \frac{\sigma^2}{2}(u_m^n - 2u_{m-1}^n + u_{m-2}^n), \quad (14.15)$$

или схему Бима–Уорминга (СБУ):

$$u_m^{n+1} = u_m^n - \sigma(u_m^n - u_{m-1}^n) - \frac{\sigma}{2}(1 - \sigma)(u_m^n - 2u_{m-1}^n + u_{m-2}^n).$$

Схема имеет порядок аппроксимации $O(\tau^2 + h^2)$ и устойчива при $\sigma \leq 1$, ее шаблон представлен на рис. 14.10.

Первое дифференциальное приближение для нее имеет вид

$$u'_t + au'_x + \frac{ah^2}{6}(\sigma - 1)(2 - \sigma)u_x''' = 0.$$

Схема дает точное решение уравнения переноса при $\sigma = 1$ и при $\sigma = 2$.

Обе схемы можно представить в потоковой форме:

$$u_m^{n+1} = u_m^n - \sigma \left(\Phi_{m+1/2}^n - \Phi_{m-1/2}^n \right), \quad (14.16)$$

где

$$\Phi_{m+1/2}^n = \frac{a}{2} (u_{m-1}^n + u_{m+1}^n) - \frac{a\sigma}{2} (u_m^n - u_{m-1}^n)$$

для СЛВ, и

$$\Phi_{m-1/2}^n = au_{m-1}^n + \frac{a}{2} (1 - \sigma) (u_{m+1}^n - u_{m-1}^n)$$

для СБУ.

Представление разностных схем в потоковой форме важно для создания целого класса численных методов (метод конечных объемов, метод уменьшения полной вариации (TVD) и др.).

Если в (14.10) опустить слагаемое второго порядка малости и аппроксимировать первую производную по координате одним из соотношений:

$$\begin{aligned} (u'_x)_m^n &\approx \frac{u_{m-1}^n - u_m^n}{h} = \frac{\Delta^+ u_m^n}{h} \quad (\text{правая разность}); \\ (u'_x)_m^n &\approx \frac{u_m^n - u_{m-1}^n}{h} = \frac{\Delta^- u_m^n}{h} \quad (\text{левая разность}), \end{aligned}$$

то получим схему соответственно «правый уголок» или «левый уголок» (схема Куранта–Изаксона–Риса, КИР, схема Годунова):

$$u_m^{n+1} = u_m^n - \sigma \begin{cases} u_{m+1}^n - u_m^n, & a < 0; \\ u_m^n - u_{m-1}^n, & a > 0, \end{cases} \quad (14.17)$$

где разность выбирается в соответствии с наклоном характеристики

$$\frac{dx}{dt} = a.$$

Воспользуемся характеристическими свойствами линейного уравнения переноса для построения аппроксимирующей его разностной схемы. Для этого проведем линейную интерполяцию значения $u_m(\tilde{x})$ сеточной функции в точке пересечения характеристики $dx/dt = a$ через узел $\{t_{m+1}, x_m\}$ с координатной

линией, между соседними узлами x_m, x_{m-1} на нижнем временном слое:

$$u_m(\tilde{x}) = \frac{x_m - \tilde{x}}{h} u_{m-1} + \frac{\tilde{x} - x_{m-1}}{h} u_m = \sigma u_{m-1} - (1 - \sigma) u_m,$$

$$\sigma > \frac{a\tau}{h}, \quad a > 0,$$

так как $\tilde{x} = x_m - a\tau$. Поскольку значение функции вдоль характеристики в случае линейного одномерного однородного уравнения переноса остается неизменным, то значение $u_m(\tilde{x})$ будет равно значению u_m^{n+1} .

Эту же схему можно представить в других удобных видах (характеристика проходит через упомянутые две точки):

$$u_m^{n+1} = u_m(\tilde{x}),$$

или, с учетом интерполяционного соотношения:

$$u_m^{n+1} = (1 - \sigma) u_m^n + \sigma u_{m-1}^n.$$

Таким образом, мы пришли к схеме КИР, используя характеристические свойства уравнения переноса (обратный метод характеристик).

Эту же схему можно представить также в других видах:

$$u_m^{n+1} = u_m^n - \frac{\sigma}{2} (u_{m+1}^n - u_{m-1}^n) + \frac{|\sigma|}{2} (u_{m+1}^n - 2u_m^n + u_{m-1}^n), \quad (14.18)$$

$$u_m^{n+1} = u_m^n - \frac{\tau}{h} [a^+ (u_{m+1}^n - u_m^n) + a^- (u_m^n - u_{m-1}^n)], \quad (14.19)$$

где $a^+ = \frac{1}{2} (a + |a|)$; $a^- = \frac{1}{2} (a - |a|)$.

В частности, в (14.18) явно выделен диссипативный член, обеспечивающий устойчивость схеме (14.17). Потокковая форма записи (14.16) обеспечивается, если положить:

$$\Phi_{m+1/2}^n = \frac{1}{2} [a (u_{m+1}^n - u_m^n) + |a| (u_{m+1}^n - u_m^n)],$$

$$\Phi_{m-1/2}^n = \frac{1}{2} [a (u_m^n - u_{m-1}^n) - |a| (u_m^n - u_{m-1}^n)].$$

Схема КИР имеет порядок аппроксимации $O(\tau + h)$, обладает наименьшей погрешностью (невязкой)

$$\xi_\tau = \frac{ah}{2} (1 - \sigma) u_{xx}''$$

среди всех схем первого порядка, условно устойчива при $\sigma \leq 1$.

Первое дифференциальное приближение для этой схемы получается при ее исследовании на аппроксимацию:

$$\frac{u_m^{n+1} - u_m^n}{\tau} + a \frac{u_m^n - u_{m-1}^n}{h} = (u'_t + au'_x) \Big|_{n,m} - \frac{ah}{2} u''_{xx} \Big|_{n,m} + \frac{\tau}{2} (u''_{tt}) \Big|_{n,m} + O(\tau^2 + h^2) = 0.$$

Пренебрегая членами второго порядка малости, получим первое дифференциальное приближение рассматриваемого уравнения:

$$u'_t + au'_x - \frac{ah}{2} u''_{xx} + \frac{\tau}{2} u''_{tt} = 0,$$

или

$$u'_t + au'_x - \frac{ah}{2} (1 - \sigma) u''_{xx} = 0,$$

где слагаемое $\frac{ah}{2} (1 - \sigma) u''_{xx}$ называется *аппроксимационной вязкостью* (или *аппроксимационной диффузией*), а коэффициент $\nu = \frac{ah}{2} (1 - \sigma)$ — *коэффициентом аппроксимационной вязкости* (диффузии).

Схема Лакса вида

$$u_m^{n+1} = \frac{1}{2} (u_{m+1}^n + u_{m-1}^n) - \frac{\sigma}{2} (u_{m+1}^n - u_{m-1}^n) \quad (14.20)$$

является условно аппроксимирующей с порядком аппроксимации $O(\tau + h^2/\tau)$, так как при $\tau \sim h^2$ невязка имеет порядок $\xi_\tau \sim O(1)$. Шаблон этой схемы представлен на рис. 14.13.

Ее первое дифференциальное приближение имеет следующий вид:

$$\frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} + \frac{ah}{2} \cdot \frac{(1 - \sigma^2)}{\sigma} \cdot \frac{\partial^2 u}{\partial x^2} = 0.$$

Ее используют в основном как конструкционный элемент при разработке схем более высокого порядка точности. Напомним, что среди линейных схем только схемы первого порядка являются, в соответствии с теоремой Годунова, монотонными.

В [7] предложен следующий вид семейства разностных схем, объединяющего три первые рассмотренные схемы:

$$u_m^{n+1} = u_m^n - \sigma (u_m^n - u_{m-1}^n) - \frac{\sigma}{2} (1 - \sigma) (\tilde{\Phi}_m - \tilde{\Phi}_{m-1}), \quad (14.21)$$

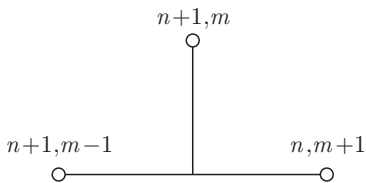


Рис. 14.13

где

$$\tilde{\Phi}_m = 0 \quad \text{для схемы КИР,} \quad (14.22)$$

$$\tilde{\Phi}_m = u_{m+1}^n - u_m^n \quad \text{для схемы Лакса–Вендроффа,} \quad (14.23)$$

$$\tilde{\Phi}_m = u_m^n - u_{m-1}^n \quad \text{для схемы Бима–Уорминга,} \quad (14.24)$$

$$\tilde{\Phi}_m = \frac{1}{2} (u_{m+1}^n - u_{m-1}^n) \quad \text{для схемы Фромма.} \quad (14.25)$$

Схема имеет шаблон, представленный на рис. 14.11, и получается путем сложения разностных схем (14.23) и (14.24):

$$u_m^{n+1} = u_m^n - \frac{\sigma}{4} (u_{m+1}^n + 3u_m^n - 5u_{m-1}^n + u_{m-2}^n) + \frac{\sigma^2}{4} (u_{m+1}^n - u_m^n - u_{m-1}^n + u_{m-2}^n). \quad (14.26)$$

Эта схема также имеет второй порядок аппроксимации $O(\tau^2 + h^2)$, но с меньшей дисперсионной погрешностью, пропорциональной третьей производной по координате.

Заметим, что разностная схема, имеющая тот же шаблон, что и схема Лакса–Вендроффа вида

$$u_m^{n+1} = u_m^n + \sigma \frac{u_m^n - u_{m-1}^n}{2}, \quad (14.27)$$

обладает порядком точности $O(\tau + h^2)$, однако она не является устойчивой, так как для нее

$$|\lambda(\alpha)|^2 = 1 + \sigma^2 \sin^2 \alpha.$$

Исследование на аппроксимацию этой схемы дает невязку вида

$$\frac{u_m^{n+1} - u_m^n}{\tau} + a \frac{u_{m+1}^n - u_m^n}{2h} = (u'_t - au'_x) \Big|_{t_n x_m} + a^2 \frac{\tau}{2} u''_{xx} + O(\tau^2 + h^2). \quad (14.28)$$

Если учесть, что

$$u''_{xx} \approx \frac{u_{m+1}^n - 2u_m^n + u_{m-1}^n}{h^2} = \Lambda_{xx} u_m^n, \quad r_\tau \approx \frac{a^2 \tau}{2} \cdot u''_{xx},$$

пренебречь членами второго порядка малости $O(\tau^2 + h^2)$ и перенести из правой части (14.28) в левую член, равный ξ_τ , то получим уже известную условно устойчивую схему Лакса–Вендроффа, представленную в следующем виде:

$$\frac{u_m^{n+1} - u_m^n}{\tau} + a \frac{u_{m+1}^n - u_{m-1}^n}{2h} - a^2 \frac{\tau}{2} \cdot \frac{u_{m-1}^n - 2u_m^n + u_{m+1}^n}{h^2} = 0. \quad (14.29)$$

Разностные схемы, представленные на рис. 14.8 и 14.9 (неявные левый и правый уголки), имеют порядок аппроксимации $O(\tau + h)$, причем первая схема устойчива при любых σ , а вторая — при $\sigma > 1$.

Неявная разностная схема вида

$$u_m^{n+1} = u_m^n - \sigma (u_{m+1}^{n+1} - u_{m-1}^{n+1}) \quad (14.30)$$

обладает порядком аппроксимации $O(\tau + h^2)$ и устойчива для любых τ, h при решении задачи Коши (вопрос устойчивости при численном решении краевой задачи следует рассматривать отдельно); шаблон представлен на рис. 14.14.

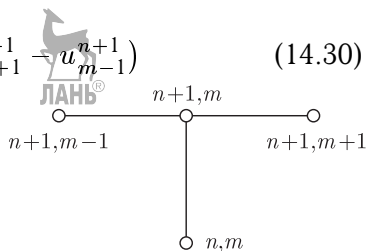


Рис. 14.14

Первое дифференциальное приближение схемы имеет вид

$$u'_t + au'_x - ah\sigma u''_{xx} + \frac{ah^2}{6} (\sigma^2 - 1) u'''_x = 0.$$

Шеститочечная схема типа Кранка–Никольсон

$$u_m^{n+1} = u_m^n - \frac{\sigma}{4} [\eta (u_{m+1}^{n+1} - u_{m-1}^{n+1}) + (1 - \eta) (u_{m+1}^n - u_{m-1}^n)] \quad (14.31)$$

при $\eta = 1/2$ имеет второй порядок аппроксимации по времени и по пространственной координате, устойчива при любом соотношении τ, h . При $\eta = 0$ схема неустойчива, при $\eta = 1$ мы получим неявную схему (14.30).

Первое дифференциальное приближение для этой схемы имеет вид

$$u'_t + au'_x = \frac{ah^2}{2} (1 + \sigma^2) u'''_x. \quad (14.32)$$

Устойчивость исследуется с помощью спектрального признака Неймана, т. е. решение ищется в виде

$$u_m^n = \lambda^n e^{i\alpha m}, \quad (14.33)$$

подстановка этого решения в (14.31) при $\xi = 1/2$ дает:

$$\frac{\lambda - 1}{h} + \frac{a}{4h} [\lambda (e^{i\alpha} - e^{-i\alpha}) + (e^{i\alpha} - e^{-i\alpha})] = 0$$

с учетом $e^{\pm i\alpha} = \cos \alpha \pm i \sin \alpha$;

$$\frac{\lambda - 1}{h} + \frac{a}{4h} (2i\lambda \sin \alpha + 2i \sin \alpha) = 0$$

где

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ a & b & c & 0 & \dots & 0 \\ 0 & a & b & c & 0 & \dots & 0 \\ 0 & \dots & 0 & a & b & c \\ 0 & \dots & \dots & 0 & 1 & -1 \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_0 \\ u_1 \\ u_2 \\ \vdots \\ u_M \end{pmatrix}^{n+1}, \quad \mathbf{f} = \begin{pmatrix} \varphi^{n+1} \\ f_1^n \\ f_2^n \\ \vdots \\ f_{m-1}^n \\ 0 \end{pmatrix},$$

$\mathbf{u}, \mathbf{f} \in R^{M+1}$ — векторы, принадлежащие $(M+1)$ -мерному линейному пространству, $\mathbf{A} \in \bar{M}[(M+1) \times (M+1)]$ — матрица порядка $(M+1) \times (M+1)$, имеющая трехдиагональную структуру (ненулевыми являются только элементы, лежащие на трех главных диагоналях); $\bar{M}[\dots]$ — пространство матриц с постоянными коэффициентами.

Вычислительная реализация метода прогонки была описана в гл. 10: решение ищется в виде

$$u_m^{n+1} = p_m u_{m+1}^{n+1} + q_m; \quad m = 0, 1, \dots (M-1),$$

прогоночные коэффициенты имеют вид

$$p_m = -\frac{c_m}{b_m - ap_{m-1}},$$

$$q_m = -\frac{f_m^n - a_m p_{m-1}}{b_m - ap_{m-1}}.$$

Как уже говорилось, хорошая обусловленность системы алгебраических уравнений достигается при выполнении условия диагонального преобладания:

$$|b_m| \geq |a_m| + |c_m| + \delta, \quad 0 < \delta \ll 1.$$

Подставив выражения для a_m, b_m, c_m в это условие, получим

$$|-\sigma/4| + |\sigma/4| + \delta \leq 1,$$

откуда

$$\frac{\sigma}{2} + \delta \leq 1,$$

т. е. в случае реализации неявных разностных схем для численного решения уравнений гиперболического типа условие диагонального преобладания налагает заметное ограничение на число Куранта $\sigma = \frac{a\tau}{h}$, что существенно уменьшает преимущества использования схем, в отличие от численного решения методом прогонки уравнений параболического типа, где таких жестких ограничений на выбор τ не существует.

Разностная схема «прямоугольник», шаблон которой представлен на рис. 14.5, имеет вид

$$\frac{(u_m^{n+1} - u_{m-1}^{n+1}) - (u_m^n + u_{m-1}^n)}{2\tau} + a \frac{(u_m^{n+1} + u_m^n) + (u_{m-1}^{n+1} - u_{m-1}^n)}{2h} = 0. \quad (14.34)$$

Невязка, получаемая разложением в ряд Тейлора сеточных функций вблизи полуцелого узла $\{t_m + \tau/2; x_m - h/2\}$, в предположении $u(t, x) \in C^3[0, X]$ имеет вид (отметим громоздкость алгебраических выкладок):

$$\xi_\tau = \tau^2 \left(\frac{1}{24} u_{ttt}'' + \frac{a}{8} u_t''' \right) + h^2 \left(\frac{1}{8} u_{txx}''' + \frac{c}{24} u_x''' \right) = O(\tau^2 + h^2),$$

т. е. имеем схему второго порядка аппроксимации по τ, h . Исследование ее на устойчивость, путем представления решения в соответствии со спектральным признаком

$$u_m^n = \lambda^n e^{i\alpha m},$$

дает

$$\lambda = e^{i\alpha m} \cdot \frac{(1 + \sigma) + (1 - \sigma) e^{i\alpha m}}{(1 + \sigma) + (1 - \sigma) e^{i\alpha m}},$$

т. е. $|\lambda| = 1$; при любых соотношениях τ, h схема устойчива.

Первое приближение этой схемы имеет вид

$$u_t + au_x - \frac{ah^2}{12} (1 - \sigma^2) u_{xxx}''' = 0.$$

В [5] был предложен следующий вид записи двухпараметрических разностных схем на шеститочечном шаблоне:

$$u_m^{n+1} = u_m^n - \frac{\sigma}{2} (\Delta^+ + \Delta^-) u_m^\alpha + \frac{h^2 \gamma}{2} (\Delta^+ + \Delta^-) u_m^\alpha,$$

где

$$u_m^\alpha = \alpha u_m^{n+1} + (1 - \alpha) u_m^n;$$

α, γ — параметры, определяющие конкретную схему.

Так, при $\{\alpha = 0, \gamma = 1\}$ мы получаем трехточечную схему Лакса, при $\{\alpha = 0, \gamma = \sigma \operatorname{sign} a\}$ — схему Куранта–Изаксона–Риса (Годунова), при $\{\alpha = 0, \gamma = 1\}$ — четырехточечную явную неустойчивую схему, при $\{\alpha = 0, \gamma = \sigma\}$ — схему Лакса–Вендроффа, при $\{\alpha = 0,5, \gamma = 0\}$ — неявную шеститочечную схему, имеющую второй порядок аппроксимации по времени и по координате, при $\{\alpha = 1, \gamma = 0\}$ — неявную четырехточечную схему, имеющую первый порядок по времени и второй по координате, при $\{\gamma = 1, \gamma = \sigma \operatorname{sign} a\}$ — схему «неявный уголок».

14.2. Двухслойные разностные схемы для решения нелинейного уравнения переноса

Нелинейное уравнение переноса

$$u'_t + u \cdot u'_x = 0 \quad (14.35)$$

может быть представлено в дивергентной форме:

$$u'_t + f'_x = 0, \quad (14.36)$$

где $f = u^2/2$.

Представленные ранее схемы Лакса, Куранта–Изаксона–Риса (Годунова), Лакса–Вендроффа в этом случае имеют следующий вид, соответственно:

$$u_m^{n+1} = \frac{1}{2} (u_{m+1}^n + u_{m-1}^n) - \frac{\tau}{2h} (f_{m+1}^n - f_{m-1}^n), \quad (14.37)$$

$$u_m^{n+1} = u_m^n - \frac{\tau}{h} \begin{cases} f_{m+1}^n - f_m^n, & u_m^n < 0, \\ f_m^n - f_{m-1}^n, & u_m^n > 0, \end{cases} \quad (14.38)$$

$$\begin{cases} u_{m+1/2}^{n+1/2} = \frac{1}{2} (u_{m+1}^n + u_m^n) - \frac{\tau}{2h} (f_{m+1}^n - f_m^n), \\ u_{m-1/2}^{n+1/2} = \frac{1}{2} (u_m^n + u_{m-1}^n) - \frac{\tau}{2h} (f_m^n - f_{m-1}^n) \end{cases} \quad (14.39)$$

— предиктор,

$$u_m^{n+1} = u_m^n - \frac{\tau}{h} (f_{m+1/2}^{n+1/2} - f_{m-1/2}^{n+1/2}) \text{ — корректор.}$$

Все три указанные схемы устойчивы при выполнении условия Куранта–Фридрихса–Леви

$$\frac{\tau}{h} \max_m |u_m^n| \leq 1. \quad (14.40)$$

Отметим, что на этапе «корректор» в схеме Лакса–Вендроффа используются промежуточный слой $n + \frac{1}{2}$ и шаблон «крест» — схема «чехарда» (или «прыгающая лягушка» — «leap-frog»), которая может быть представлена в виде (14.39) или в недивергентной форме:

$$\frac{u_m^{n+1} - u_m^n}{2\tau} + a \frac{(u_m^n + u_{m-1}^n) + (u_m^{n+1} - u_{m-1}^{n+1})}{2h} = 0; \quad (14.41)$$

ее шаблон показан на рис. 14.15.

Схема имеет второй порядок аппроксимации по τ, h , устойчива при выполнении условия Куранта–Фридрихса–Леви. В схеме «предиктор–корректор» (14.39) на этапе «предиктор» используется

схема первого порядка Лакса, на втором этапе — схема второго порядка, итоговая аппроксимация — $O(\tau^2 + h^2)$. Шаблон этой схемы представлен на рис. 14.16.

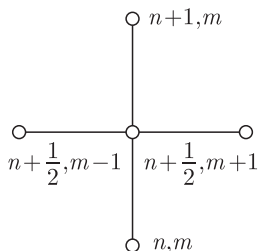


Рис. 14.15

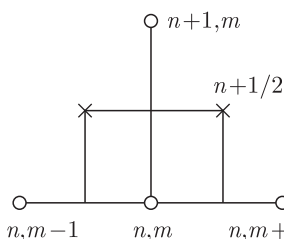


Рис. 14.16

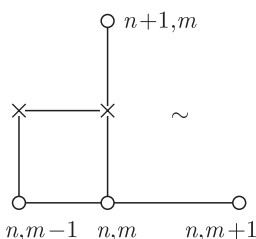


Рис. 14.17

Такие схемы называют *центральными*. Примером нецентральной схемы второго порядка аппроксимации является схема Мак-Кормака:

$$\begin{cases} \tilde{u}_m = u_m^n - \frac{\tau}{h} (f_{m+1}^n - f_m^n); & \tilde{u}_m = u_{m-1}^n - \frac{\tau}{h} (f_m^n - f_{m-1}^n); \\ u_m^{n+1} = \frac{1}{2} (u_m^n + \tilde{u}_m) - \frac{\tau}{2h} (\tilde{f}_m - \tilde{f}_{m-1}), \end{cases} \quad (14.42)$$

ее шаблон показан на рис. 14.17.

Определенным достоинством последней схемы является отсутствие полуцелых индексов, что упрощает постановку граничных условий при численном решении краевой задачи. Схема также устойчива при выполнении условия (14.40).

Центральная 3-этапная разностная схема более высокого, третьего порядка точности по τ, h была предложена В. В. Русановым [6]:

$$\begin{cases} u_{m+1/2}^{n+1/3} = \frac{1}{2} (u_{m+1}^n + u_m^n) - \frac{\tau}{3h} (f_{m+1}^n - f_m^n), & \text{первый этап;} \\ u_m^{n+2/3} = u_m^n - \frac{2\tau}{3h} (f_{m+1/2}^{n+1/3} - f_{m-1/2}^{n+1/3}), & \text{второй этап;} \\ u_m^{n+1} = u_m^n - \frac{3\tau}{8h} (f_{m+1}^{n+2/3} - f_{m-1}^{n+2/3}) - \\ - \frac{\tau}{24h} (-2f_{m+2}^n + 7f_{m+1}^n - 7f_{m-1}^n + 2f_{m-2}^n) - \\ - \omega \frac{\tau}{24} (u_{m+2}^n - 4u_{m+1}^n + 6u_m^n - 4u_{m-1}^n + u_{m-2}^n) = 0, & \text{третий этап;} \\ 4\sigma^2 - \sigma^4 \leq \omega \leq 3, & \sigma = \tau/h. \end{cases} \quad (14.43)$$

Соответствующий данной схеме шаблон представлен на рис. 14.18.

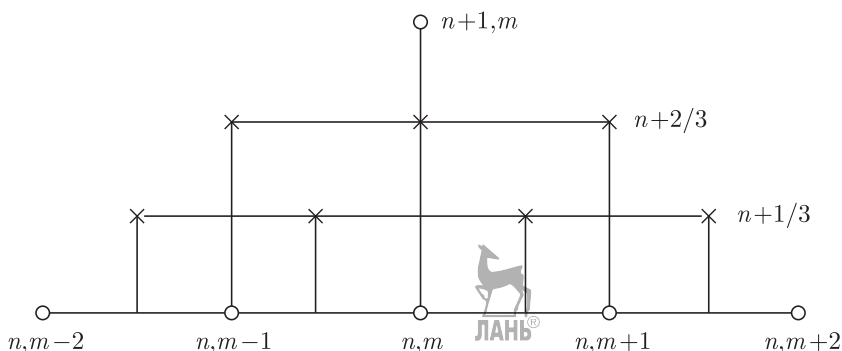


Рис. 14.18

На первом шаге находятся четыре значения сеточной функции по схеме Лакса, на втором — три значения по схеме «чехарда», схема устойчива при выполнении условия $\tau \leq h / \max_m |u_m^n|$.

В случае разностной аппроксимации линейного нелинейного уравнения схема Русанова приобретает следующий вид:

$$\begin{aligned} u_m^{n+1/3} &= \frac{u_{m+1}^n + u_m^n}{2} - \frac{\sigma}{3} (u_{m+1}^n - u_m^n); \\ u_m^{n+2/3} &= u_m^n - \frac{2}{3} \sigma (u_{m+1/2}^{n+1/3} - u_{m-1/2}^{n+1/3}); \\ u_m^{n+1} &= u_m^n - \frac{3}{8} \sigma (u_{m+1}^{n+2/3} - u_{m-1}^{n+2/3}) - \\ &\quad - \frac{\sigma}{3} (-2u_{m+2}^n + 7u_{m+1}^n - 7u_{m-1}^n + 2u_{m-2}^n) - \\ &\quad - \frac{\omega}{24} (u_{m+2}^n - 4u_{m+1}^n + 6u_m^n - 4u_{m-1}^n + u_{m-2}^n). \end{aligned}$$

Нецентральная схема третьего порядка точности была предложена Уормингом, Кутером и Ломаксом [6]:

$$\left\{ \begin{aligned} \tilde{u}_{m+1/2} &= u_m^n - \frac{2\tau}{3h} (f_{m+1}^n - f_m^n), & \text{первый этап;} \\ \tilde{u}_m &= \frac{1}{2} \left[u_{m+1}^n + \tilde{u}_m - \frac{2\tau}{3h} (\tilde{f}_m - \tilde{f}_{m-1}) \right], & \text{второй этап;} \\ u_m^{n+1} &= u_m^n - \frac{3\tau}{8h} (\tilde{f}_{m+1} - \tilde{f}_{m-1}) - \\ &\quad - \frac{\tau}{24h} (-2f_{m+2}^n + 7f_{m+1}^n - 7f_{m-1}^n + 2f_{m-2}^n) - \\ &\quad - \omega \frac{\tau}{24} (u_{m+2}^n - 4u_{m+1}^n + 6u_m^n - 4u_{m-1}^n + u_{m-2}^n), & \text{третий этап;} \\ 4\sigma^2 - \sigma^4 &\leq \omega \leq 3. \end{aligned} \right. \quad (14.44)$$

Схема устойчива при выполнении того же условия, что и в схеме Русанова. Последнее слагаемое в обеих схемах третьего порядка аппроксимации добавляется для их устойчивости. Шаблон данной схемы представлен на рис. 14.19 (крестиками везде обозначены узлы, принадлежащие промежуточным слоям по времени).

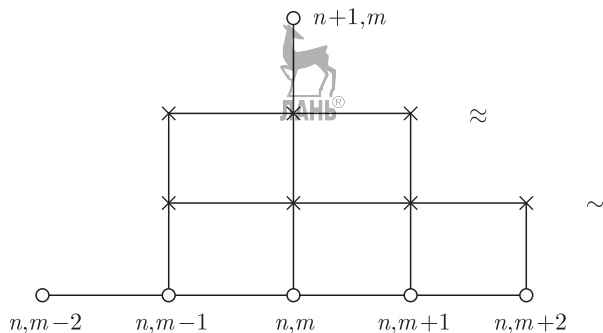


Рис. 14.19

Для аппроксимации линейного уравнения переноса схема Уорминга–Кутлера–Ломакса принимает следующий вид:

$$\begin{aligned}\tilde{u}_m &= u_m^n - \frac{2}{3}\sigma(u_{m-1}^n - u_m^n); \\ \tilde{\tilde{u}}_m &= \frac{u_m^n + \tilde{u}_m}{2} - \frac{\sigma}{3}(\tilde{u}_m - \tilde{u}_{m-1}); \\ u_m^{n+1} &= u_m^n - \frac{3}{8}\sigma(\tilde{\tilde{u}}_{m+1} - \tilde{\tilde{u}}_{m-1}) - \frac{\sigma}{3}(-2u_{m+2}^n + 7u_{m-1}^n - \\ &\quad - 7u_{m-1}^n + 2u_{m-2}^n) - \frac{\omega\tau}{24}(u_{m+2}^n - 4u_{m+1}^n + 6u_m^n - 4u_{m-1}^n + u_{m-2}^n).\end{aligned}$$

Примером неявной двухслойной шеститочечной разностной схемы с шаблоном, изображенным на рис. 14.4, является схема Бима–Уорминга (заметим, что существует и явная схема Бима–Уорминга, которая рассматривалась выше):

$$u_m^{n+1} = u_m^n - \frac{\tau}{2} \left\{ [f'_x]_m^n + [f'_x]_m^{n+1} \right\}, \quad (14.45)$$

где квадратные скобки означают разностную аппроксимацию производных по x .

Проведя линеаризацию сеточной функции f_m^{n+1} в соответствии с формулой

$$f_m^{n+1} \approx f_m^n + \Delta_t u_m^n (f'_u)_m^n = f_m^n + \Delta_t u_m^n \cdot P_m^n,$$

где

$$\Delta_t u_m^n = u_m^{n+1} - u_m^n, \quad P_m^n = (f')_m^n,$$

получим схему вида

$$u_m^{n+1} = u_m^n + \frac{\tau}{2} \left\{ 2[f'_x]_m^n + \left[\frac{\partial}{\partial x} (P_m^n \cdot \Delta_t u_m^n) \right] \right\},$$

или, после разностной аппроксимации производных в квадратных скобках:

$$a_m^n u_{m-1}^{n+1} + b_m^n u_m^{n+1} + c_m^n u_{m+1}^{n+1} = d_m^n, \quad (14.46)$$

где:

$$a = -\frac{\tau}{4h} P_{m-1}^n; \quad b = 1; \quad c = \frac{\tau}{4h} P_{m+1}^n, \\ d_m^n = -\frac{\tau}{2h} (f_{m+1}^n - f_{m-1}^n) - \frac{\tau}{4h} P_{m-1}^n u_{m-1}^n + u_m^n + \frac{\tau}{4h} P_{m+1}^n u_{m+1}^n.$$

В результате мы получили систему линейных алгебраических уравнений с матрицей трехдиагональной структуры.

Получим разностную схему для решения нелинейного уравнения переноса, записанного в дивергентной форме, на шести-точечном шаблоне, аналогично аппроксимации уравнения теплопроводности:

$$\int_S \left(\frac{\partial u}{\partial t} + \frac{\partial f}{\partial x} \right) dt dx = \int_{\Gamma} u dx - f dt = 0.$$

Аппроксимируем контурный интеграл по формуле средних:

$$u_m^n h - f_{m+1/2}^{n+1/2} \tau - u_m^{n+1} h + f_{m-1/2}^{n+1/2} \tau = 0,$$

или:

$$\frac{u_m^{n+1} - u_m^n}{\tau} + \frac{f_{m+1/2}^{n+1/2} - f_{m-1/2}^{n+1/2}}{h} = 0; \quad f = \frac{u^2}{2}.$$

14.3. Трехслойные разностные схемы для решения уравнения переноса

В качестве примера трехслойной разностной схемы (кроме схемы «чехарда») приведем схему «кабаре» [12], ставшую достаточно известной среди исследователей-вычислителей для решения задач газодинамики:

$$\frac{(u_m^{n+1} - u_m^n) + (u_{m-1}^n - u_{m-1}^{n-1})}{2\tau} + \frac{u_m^n - u_{m-1}^n}{h} = 0, \quad (14.47)$$

или

$$u_m^{n+1} = u_m^n - \frac{1}{2\tau} (u_{m-1}^n - u_{m-1}^{n-1}) - \frac{\sigma}{2} (u_m^n - u_{m-1}^n)$$

(шаблон представлен на рис. 14.20).

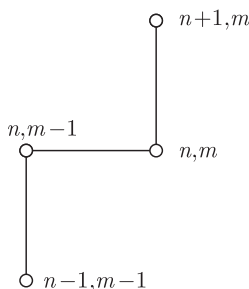


Рис. 14.20

Эта схема является точной для двух значений $\sigma = a\tau/h$: 1 и 0,5. Первое дифференциальное приближение схемы имеет вид (см. шаблон на рис. 14.20)

$$u'_t + au'_x - \frac{ah^2}{12} (1 - 3\sigma + 2\sigma^2) u'''_x = 0.$$

Схема имеет второй порядок аппроксимации по τ, h , устойчива при условии $\sigma \leq 1$.

Примером® компактной трехслойной разностной неявной схемы [8] для аппроксимации линейного уравнения переноса является схема А. И. Толстых

$$\begin{aligned} & \frac{5}{12} \frac{u_m^{n+1} - u_{m-1}^n}{\tau} + \frac{8}{12} \frac{u_m^{n+1} - u_{m-1}^n}{\tau} - \\ & - \frac{1}{12} \frac{u_m^{n+1} - u_{m+1}^n}{\tau} + \frac{5a}{12} \frac{u_m^{n+1} - u_{m-1}^n}{h} + \frac{8a}{12} \frac{u_m^n - u_{m-1}^n}{h} - \\ & - \frac{8a}{12} \frac{u_m^{n-1} - u_m^n}{h} = 0, \quad (14.48) \end{aligned}$$

имеющая порядок аппроксимации $O(\tau^3 + h^3)$; ее шаблон показан на рис. 14.21.

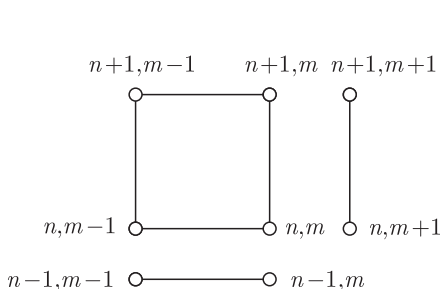


Рис. 14.21

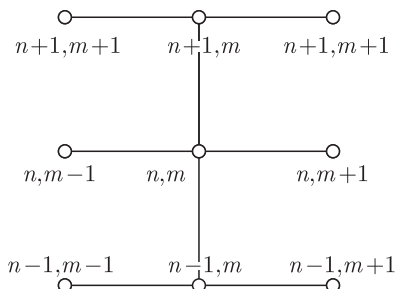


Рис. 14.22

Представим также неявную трехслойную схему Л. А. Тушевой

$$\begin{aligned} & - (1 + \sigma) u_{m+1}^{n+1} - 4u_m^{n+1} - (1 - \sigma) u_{m-1}^{n+1} - 4\sigma u_{m+1}^n + 4\sigma u_{m-1}^n + \\ & + (1 + \sigma) u_{m-1}^{n-1} + 4u_m^{n-1} + (1 - \sigma) u_{m+1}^{n-1} = 0, \quad (14.49) \end{aligned}$$

аппроксимирующую линейное уравнение переноса (14.1) на шаблоне, представленном на рис. 14.22, с четвертым порядком аппроксимации.

14.4. Разностные схемы для решения волнового уравнения и акустической системы

К уравнениям гиперболического типа относится также волновое уравнение, имеющее вид

$$u_{tt}'' = a^2 u_{xx}'', \quad (14.50)$$

которое может быть аппроксимировано на шаблоне «крест» (рис. 14.23) с помощью явной трехслойной разностной схемы следующим образом:

$$\frac{u_m^{n+1} - 2u_m^n + u_m^{n-1}}{\tau^2} - a^2 \frac{u_{m-1}^n - 2u_m^n + u_{m+1}^n}{h^2} = 0, \quad (14.51)$$

или

$$u_m^{n+1} = (2u_m^n - u_{m-1}^n) + \sigma^2 (u_{m-1}^n - 2u_m^n + u_{m+1}^n) = 0.$$

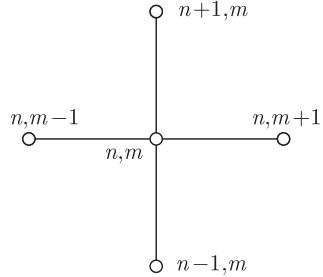


Рис. 14.23

Выражение для невязки получим с помощью разложения сеточной функции в ряд Тейлора

$$\xi_\tau = \frac{\tau^2}{12} u_t^{(4)} - \frac{a^2 h^2}{12} u_x^{(4)} + O(\tau^4, h^4),$$

или, с учетом $u_x^{(4)} = a^4 u_t^{(4)}$:

$$\xi_\tau = \frac{a^2}{12} h^2 (\sigma^2 - 1) u_x^{(4)} + O(\tau^4, h^4),$$

т. е. схема имеет второй порядок аппроксимации по τ, h , а первое дифференциальное приближение имеет вид

$$u_{tt}'' - a^2 u_{xx}'' - \frac{a^2}{12} h^2 (\sigma^2 - 1) u_x^{(4)} = 0. \quad (14.52)$$

Исследование схемы «крест» для численного решения волнового уравнения при помощи спектрального признака дает

$$\lambda^2 - 2 \left(1 - 2\sigma^2 \sin^2 \frac{\alpha}{2} \right) \lambda + 1 = 0.$$

В таком случае, в соответствии с теоремой Виета, произведение корней этого уравнения есть

$$\lambda_1 \cdot \lambda_2 = 1,$$

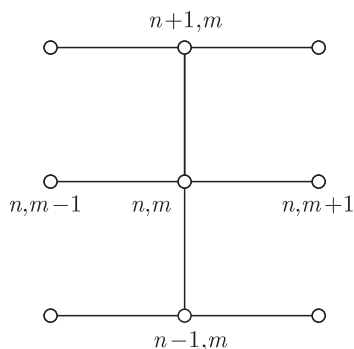


Рис. 14.24

и условие устойчивости выполняется при $|\lambda_1| \cdot |\lambda_2| = 1$, что означает комплексную сопряженность его корней, а это возможно при отрицательном дискриминанте полученного квадратного уравнения:

$$D = 4\sigma^2 \sin^2 \alpha \left(\sigma^2 \sin^2 \frac{\alpha}{2} - 1 \right) < 0.$$

Следовательно, условие устойчивости выполняется для любых α при $\sigma \leq 1$.

Трехслойная разностная схема для рассматриваемого волнового уравнения может быть также представлена в параметрическом виде ($0 \leq \eta \leq 1$) на девятиточечном шаблоне, показанном на рис. 14.24:

$$u_m^{n+1} = u_m^{n-1} + a\sigma \left[\eta (u_{m+1}^{n+1} - u_{m-1}^{n+1}) + (1 - 2\eta) (u_{m+1}^n - u_{m-1}^n) + \eta (u_{m+1}^{n-1} - u_{m-1}^{n-1}) \right]. \quad (14.53)$$

Исследование на устойчивость по признаку Неймана дает следующее условие:

$$\eta \geq \frac{1}{4} - \frac{1}{4\sigma^2}.$$

Невязка в этой схеме имеет вид

$$\xi_\tau = \left(\frac{a^4 \tau}{12} - \frac{a^2 h^2}{12} - \eta a^4 \tau^4 \right) u_x^{(4)} + O(\tau^4, h^4),$$

т. е. при выборе параметра

$$\eta = \frac{1}{12} \frac{1}{\sigma^2}$$

рассматриваемая схема имеет четвертый порядок аппроксимации по τ, h .

Заметим, что численное решение волнового уравнения можно реализовать и на двухслойных разностных схемах, если заметить его на систему двух уравнений переноса первого порядка

$$\left(v = \int_0^x u'_t(y) dy \right):$$

$$\begin{cases} u'_t = v'_x, \\ v'_t = a^2 u'_x. \end{cases} \quad (14.54)$$

Одномерная акустическая система, описывающая распространение плоских звуковых волн, записывается в следующем виде:

$$\begin{cases} \frac{\partial u}{\partial t} + \frac{1}{\rho} \cdot \frac{\partial p}{\partial x} = 0, \\ \frac{\partial p}{\partial t} + \rho c^2 \cdot \frac{\partial u}{\partial x} = 0. \end{cases} \quad (14.55a)$$

Здесь $\rho = \text{const}$ — плотность среды, $c = \text{const}$ — скорость звука в среде, p — давление в среде, u — скорость движения среды в заданной точке пространства x в данный момент времени t .

Если мы умножим (14.55б) на $1/(\rho c)$ и сложим с (14.55а), а затем вычтем из него, то получим систему уравнений акустики, записанную в инвариантах Римана:

$$\begin{cases} \frac{\partial r}{\partial t} + a \frac{\partial r}{\partial x} = 0, \\ \frac{\partial s}{\partial t} - a \frac{\partial s}{\partial x} = 0, \end{cases} \quad (14.56)$$

где $r = u + \frac{p}{\rho a}$, $s = u - \frac{p}{\rho a}$; эта система имеет общее решение вида

$$r = F(x - at), \quad s = G(x + at),$$

откуда получим выражения для u, p :

$$u = \frac{F(x - at) + G(x + at)}{2}; \quad p = \frac{\rho c [F(x - at) - G(x + at)]}{2}.$$

Очевидно, что значения инвариантов Римана остаются постоянными вдоль характеристических прямых $\frac{dx}{dt} = \pm a$. Для разностной аппроксимации (14.56) можно использовать уже известную схему Куранта–Изаксона–Риса (Годунова):

$$\begin{cases} r_m^{n+1} = (1 - \sigma) r_m^n + \sigma r_{m-1}^n, \\ s_m^{n+1} = (1 - \sigma) s_m^n - \sigma s_{m+1}^n, \end{cases} \quad (14.57)$$

которая исследовалась выше.

Полученная схема имеет двойной шаблон (левый и правый уголки; рис. 14.25).

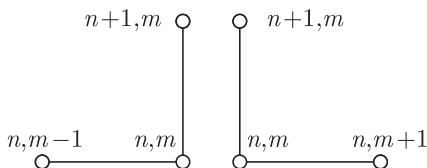


Рис. 14.25

Исследуем на устойчивость три другие разностные схемы, аппроксимирующие акустическую систему (для простоты положим $a = 1$):

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{\partial v}{\partial x}, \\ \frac{\partial v}{\partial t} = \frac{\partial u}{\partial x}, \end{cases} \quad (14.58)$$

записанную в матричном виде:

$$\frac{\partial \mathbf{w}}{\partial t} - \mathbf{B} \frac{\partial \mathbf{w}}{\partial x} = 0, \quad (14.59)$$

где $\mathbf{w} = \begin{Bmatrix} u \\ v \end{Bmatrix}$ — вектор-столбец искомых функций, \mathbf{B} — матрица (2×2) : $\mathbf{B} = \begin{Bmatrix} 0 & 1 \\ 1 & 0 \end{Bmatrix}$.

Первая схема имеет вид

$$\frac{\mathbf{w}_m^{n+1} - \mathbf{w}_m^n}{\tau} - \mathbf{A} \frac{\mathbf{w}_{m+1}^n - \mathbf{w}_m^n}{h} = 0. \quad (14.60)$$

Для исследования данной схемы на спектральную устойчивость будем искать решение разностного уравнения в виде

$$\mathbf{w}_m^n = \lambda^p \begin{Bmatrix} u \\ v \end{Bmatrix} e^{i\alpha m}. \quad (14.61)$$

Подстановка решения в систему дает

$$\frac{\lambda - 1}{\tau} \mathbf{w}^0 - \mathbf{A} \frac{e^{i\alpha} - 1}{h} \mathbf{w}^0 = 0,$$

или

$$(\lambda - 1) \mathbf{w}^0 - \sigma (e^{i\alpha} - 1) \mathbf{A} \mathbf{w}^0 = 0.$$

В развернутом виде полученное уравнение имеет вид

$$\begin{pmatrix} \lambda - 1 & -\sigma (e^{i\alpha} - 1) \\ -\sigma (e^{i\alpha} - 1) & \lambda - 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = 0. \quad (14.62)$$

Данная система линейных алгебраических уравнений имеет решение в случае, если определитель отличен от нуля, т. е.

$$(\lambda - 1)^2 = \sigma^2 (e^{i\alpha} - 1)^2,$$

откуда получим:

$$\lambda_1 = 1 - \sigma + \sigma e^{i\alpha}, \quad \lambda_2 = 1 + \sigma - \sigma e^{i\alpha},$$

т. е. корни $\lambda_1(\alpha)$ и $\lambda_2(\alpha)$ на комплексной плоскости пробегают окружности радиуса σ с центрами в точках $1 - \sigma$ и $1 + \sigma$.

По этой причине спектральное условие устойчивости не выполняется.

Рассмотрим другую схему — Лакса–Вендроффа, имеющую в случае рассматриваемой системы дифференциальных уравнений в частных производных первого порядка (14.59) следующий вид:

$$\frac{\mathbf{w}_m^{n+1} - \mathbf{w}_m^n}{\tau} - \mathbf{B} \frac{\mathbf{w}_{m+1}^n - \mathbf{w}_{m-1}^n}{2h} - \frac{\tau}{2h^2} \mathbf{B}^2 (\mathbf{w}_{m+1}^n - 2\mathbf{w}_m^n + \mathbf{w}_{m-1}^n) = 0. \quad (14.63)$$

Для исследования схемы (14.63) на спектральную устойчивость, поступая, как и в предыдущем случае, получим следующие выражения для двух значений $\lambda(\alpha)$:

$$\begin{aligned} \lambda_1 &= 1 + i\sigma \sin \alpha - 2\sigma^2 \sin^2 \frac{\alpha}{2}; \\ \lambda_2 &= 1 - i\sigma \sin \alpha - 2\sigma^2 \sin^2 \frac{\alpha}{2}, \end{aligned}$$

откуда

$$1 - |\lambda_{1,2}(\alpha)|^2 = 4\sigma^2 \sin^4 \frac{\alpha}{2} (1 - \sigma^2),$$

т. е. рассматриваемая схема будет устойчивой при выполнении условия Куранта–Изаксона–Леви

$$\tau \leq \frac{h}{a}.$$

Третья схема — «шахматная»:

$$\begin{cases} \frac{u_m^{n+1} - u_m^n}{\tau} - \frac{v_{m+1/2}^{n+1/2} - v_{m-1/2}^{n-1/2}}{h} = 0, \\ \frac{v_{m+1/2}^{n+1/2} - v_{m+1/2}^{n-1/2}}{\tau} - \frac{u_{m+1}^n - u_m^n}{h} = 0, \end{cases}$$

в которой введены полуцелые индексы (узлы):

$$\begin{aligned} t_{n+1/2} &= t_n + \tau/2, \\ x_{m+1/2} &= x_m + h/2, \\ v_{m+1/2}^{n+1/2} &= v(t_{n+1/2}; x_{m+1/2}). \end{aligned}$$

Шаблон изображен на рис. 14.26.

Исследование схемы на устойчивость приводит к выражению

$$\begin{Bmatrix} u_m^n \\ v_{m+1/2}^{n+1/2} \end{Bmatrix} = \begin{Bmatrix} U_0 e^{im\alpha} \\ V_0 e^{i(m+1/2)\alpha} \end{Bmatrix},$$

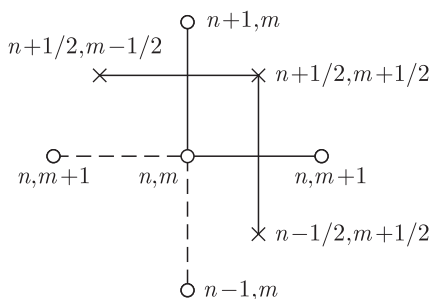


Рис. 14.26

где U_0, V_0 — некоторые постоянные. После подстановки в исходную схему получим систему

$$\begin{cases} U_0 \frac{\lambda - 1}{\tau} + V_0 \frac{e^{i\alpha/2} - e^{-i\alpha/2}}{h} = 0, \\ U_0 \lambda \frac{e^{i\alpha/2} - e^{-i\alpha/2}}{h} + V_0 \frac{\lambda - 1}{\tau} = 0, \end{cases}$$

которая имеет нетривиальное решение (относительно неизвестных U_0, V_0) при условии

$$\det \begin{Bmatrix} \frac{\lambda - 1}{2} & 2\frac{i}{h} \sin \frac{\alpha}{2} \\ 2\lambda \frac{i}{h} \sin \frac{\alpha}{2} & \frac{\lambda - 1}{\tau} \end{Bmatrix} = 0,$$

откуда получим уравнение для определения λ :

$$\left(\frac{\lambda - 1}{\tau} \right)^2 + 4 \frac{\lambda}{h^2} \sin^2 \frac{\alpha}{2} = 0,$$

или

$$\lambda^2 - 2\lambda \left(1 - 2 \frac{\tau^2}{h^2} \sin^2 \frac{\alpha}{2} \right) + 1 = 0.$$

Подобное уравнение уже исследовалось (схема «крест»): схема устойчива при $\tau \leq h$.

14.5. Гибридные разностные схемы

Идею построения гибридных схем для устранения (или минимизации) осцилляций численного происхождения в областях больших градиентов численных решений впервые предложил Р.П. Федоренко в 1962 г. Первая гибридная схема Федоренко

описана в [2, 3] на примере численного решения линейного уравнения переноса

$$u'_t + u'_x = 0 \quad (Pu = 0) \quad (14.64)$$

и разностной схемы КИР («левый уголок»):

$$\frac{u_m^{n+1} - u_m^n}{\tau} + \frac{u_m^n - u_{m-1}^n}{h} = 0, \quad P_\tau u_\tau = 0. \quad (14.65)$$

Исследование данной схемы на аппроксимацию дает

$$\begin{aligned} \frac{u_m^{n+1} - u_m^n}{\tau} + \frac{u_m^n - u_{m-1}^n}{h} = \\ = (u'_t + u'_x) \Big|_{t_n, x_m} + \frac{\tau}{2} u''_{tt} - \frac{h}{2} u''_{xx} + O(\tau^2 + h^2), \end{aligned} \quad (14.66)$$

или

$$P_\tau u_\tau = Lu \Big|_{t_n, x_m} + \xi \tau \Big|_{t_n, x_m} + O(\tau^2 + h^2).$$

Учитывая дифференциальное следствие линейного уравнения переноса

$$u''_{tt} = u''_{xx}$$

и пренебрегая членами второго порядка малости в правой части (14.66) после аппроксимации второй производной u''_{xx} , получим известную нам разностную схему второго порядка аппроксимации по τ и h Лакса–Вендроффа:

$$P_\tau u_\tau - \frac{\tau - h}{2} \cdot \Lambda_{xx} u_m^n = 0, \quad (14.67)$$

или

$$u_m^{n+1} = u_m^n - \sigma \cdot \Delta^- u_m^n - \frac{\sigma}{2} (1 - \sigma) \Delta^2 u_m^n = 0,$$

где

$$\sigma = \frac{\tau}{h}; \quad \Delta^- u_m^n = u_m^n - u_{m-1}^n; \quad \Delta^+ u_m^n = u_{m+1}^n - u_m^n;$$

$$\Delta^2 u_m^n = u_{m-1}^n - 2u_m^n + u_{m+1}^n.$$

Если при исследовании схемы КИР на аппроксимацию в разложении сеточных функций в ряд Тейлора оставить и аппроксимировать разностными соотношениями члены второго порядка малости по τ и h , то после аналогичных алгебраических выкладок получим разностную схему третьего порядка аппроксимации по τ и h следующего вида:

$$u_m^{n+1} = u_m^n - \sigma \Delta^- u_m^n - \frac{\sigma}{2} (1 - \sigma) \Delta^2 u_m^n - \frac{\sigma}{6} (1 - \sigma^2) \Delta^3 u_m^n, \quad (14.68)$$

где $\Delta^3 u_m^n = u_{m+1}^n - 3u_m^n + 3u_{m-1}^n - u_{m-2}^n$.

Если предложить анализатор гладкости сеточной функции в виде

$$\nu_m^n = \begin{cases} 1, & \text{если } |\Delta^2 u_m^n| \leq \lambda \cdot |\Delta^- u_m^n|, \\ 0, & \text{если } |\Delta^2 u_m^n| > \lambda \cdot |\Delta^- u_m^n| \end{cases} \quad (14.69)$$

и сконструировать разностную схему (14.67) с учетом (14.69):

$$u_m^{n+1} = u_m^n - \sigma \cdot \Delta^- u_m^n - \nu_m^n \cdot \frac{\sigma}{2} (1 - \sigma) \Delta^2 u_m^n, \quad (14.70)$$

то получим в областях с большими градиентами численного решения $\nu_m^n = 0$, в областях с гладкими численными решениями $\nu_m^n = 1$.

В первом случае реализуется схема КИР первого порядка аппроксимации ($\lambda = 0$), во втором — схема Лакса–Вендрофа ($\lambda = \infty$).

Такой же анализатор можно предложить и для схемы третьего порядка аппроксимации.

Заметим, что аналогичным путем можно получать разностные схемы более высоких порядков аппроксимации. Действительно, разложение в ряд Тейлора сеточных функций u_m^{n+1}, u_{m+1}^n для линейного уравнения переноса (для простоты выкладок положим в (14.1) $a = -1$)

$$u'_t - u'_x = 0$$

дает следующее выражение для погрешности:

$$u_{m+1}^n = u_m^n + h (u'_x)_m^n + \frac{h^2}{2!} (u''_{xx})_m^n + \dots + \frac{h^n}{n!} \left(u_x^{(n)} \right)_m^n + O(h^{n+1}) = 0,$$

$$u_m^{n+1} = u_m^n + \tau (u'_t)_m^n + \frac{\tau^2}{2!} (u''_{tt})_m^n + \dots + \frac{\tau^n}{n!} \left(u_t^{(n)} \right)_m^n + O(\tau^{n+1}) = 0,$$

$$P_\tau u_\tau = P u_{t_n, x_m} + \xi_\tau,$$

где

$$\begin{aligned} \xi_\tau &= \frac{1}{2!} (\tau - h) u''_x + \frac{1}{3!} (\tau^2 - h^2) u'''_x + \dots \\ &\dots + \frac{1}{(n+1)!} (\tau^n - h^n) u_x^{(n+1)} + O(\tau^{n+1}, h^{n+1}) = \\ &= \sum_{i=1}^n \frac{h^i}{(i+1)!} (\sigma^i - 1) u_x^{(i+1)} + O(\tau^{n+1}, h^{n+1}), \end{aligned}$$

что позволяет построить разностную схему $(n + 1)$ -го порядка аппроксимации

$$u_m^{n+1} = u_m^n + \sigma \cdot \Delta^+ u_m^n + \sigma \sum_{i=1}^n \frac{\sigma^i - 1}{(i + 1)!} \Delta^{i+1} u_m^n = 0, \quad (14.71)$$

где $\Delta_h^{i+1} u_m^n$ — i -я конечная разность для сеточной функции u_m^n . Дифференциальное приближение этой схемы имеет вид

$$u'_t - u'_x = h^{-1} \sum_{i=1}^n \frac{\sigma^i - 1}{(i + 1)!} \Delta^{i+1} u_m^n.$$

Гибридизация (регуляризация) схемы высокого порядка точности проводится аналогичным образом.

В. П. Колганом впервые была предложена гибридная разностная схема, использующая четыре сеточных шаблона (рис. 14.1–14.3, 14.10) [5, 13]:

$$\begin{aligned} \frac{u_m^{n+1} - u_m^n}{\tau} + \frac{\Delta^+ + \Delta^-}{2h} u_m^n &= 0, \\ \text{если } |\Delta^+ u_{m-2}^n| &\geq |\Delta^- u_{m-1}^n| \geq |\Delta^- u_m^n|; \\ \frac{u_m^{n+1} - u_m^n}{\tau} + a \frac{\Delta^-}{h} u_m^n + \frac{ah}{2} \cdot \frac{\Delta^+ \cdot \Delta^-}{h^2} u_{m-1}^n &= 0, \\ \text{если } |\Delta^+ u_{m-2}^n| &\leq |\Delta^+ u_{m-1}^n| \leq |\Delta^+ u_m^n|, \\ \frac{u_m^{n+1} - u_m^n}{\tau} + a \frac{\Delta^-}{h} u_m^n &= 0, \\ \text{если } |\Delta^+ u_{m-2}^n| &\geq |\Delta^+ u_{m-1}^n|, \quad |\Delta^+ u_m^n| \geq |\Delta^+ u_{m-1}^n|, \\ \frac{u_m^{n+1} - u_m^n}{\tau} + a \frac{\Delta^+ + \Delta^-}{2h} u_m^n + \frac{ah}{2} \cdot \frac{\Delta^+ \cdot \Delta^-}{h^2} u_{m-1}^n &= 0, \\ \text{если } |\Delta^+ u_{m-2}^n| &\leq |\Delta^+ u_{m-1}^n|, \quad |\Delta^+ u_m^n| \geq |\Delta^+ u_{m-1}^n|. \end{aligned} \quad (14.72)$$

В зависимости от гладкости численного решения, которая определяется анализатором гладкости, в этом случае используются схемы: 4-точечная с центральными разностями, Бима–Уорминга, Куранта–Изаксона–Риса, Лакса–Вендроффа. Отметим, что из четырех схем только одна является монотонной — схема КИР первого порядка точности.

Развитием идей Федоренко и Колгана о построении гибридных разностных схем, с учетом представления схем в потоковой форме, предложенной Борисом и Буком [6], являются гибридные TVD-схемы Хартена, которые удовлетворяют условиям

уменьшения полной вариации (УПВ, Total Variation Diminishing, TVD) [14]:

$$\text{Var}(u^{n+1}) \leq \text{Var}(u^n), \quad (14.73)$$

где $\text{Var}(u^n) = \sum_{m=-\infty}^{\infty} |u_{m+1}^n - u_m^n|$.

Представим уже известную нам разностную схему Лакса–Вендроффа в потоковом виде:

$$u_m^{n+1} = u_m^n - \sigma \cdot \Delta^- u_m^n - \left(\Phi_{m+1/2}^n - \Phi_{m-1/2}^n \right), \quad (14.74)$$

где $\Phi_{m+1/2}^n = (\sigma/2)(1 - \sigma)(u_{m+1}^n - u_m^n)$.

Эта схема не является монотонной, но ее можно сделать таковой, если ограничить так называемые *антидиффузионные потоки* с помощью функции $\eta(s_m)$:

$$\tilde{\Phi}_{m+1/2}^n = \frac{\eta(s_m)}{2} \sigma (1 - \sigma) (u_{m+1}^n - u_m^n),$$

где $s_m = \Delta^- u_m^n / (\Delta^+ u_m^n)$, η — ограничитель (лимитер).

Оказывается, что полученная схема устойчива при выполнении условия:

$$\begin{cases} 0 < \eta(s_m) \leq \min(2s_m, 2), & \text{если } s_m > 0; \\ \eta(s_m) = 0, & \text{если } s_m \leq 0. \end{cases}$$

Пример функции-ограничителя:

$$\eta(s_m) = \begin{cases} \min(2, s_m), & \text{если } s_m > 1, \text{ градиент численного} \\ & \text{решения убывает;} \\ \min(2s_m, 1), & \text{если } 0 < s_m \leq 1, \text{ градиент решения} \\ & \text{растет;} \\ 0, & \text{если } s_m \leq 0, \text{ численное решение} \\ & \text{осциллирует.} \end{cases}$$

Эта функция является анализатором гладкости численного решения и позволяет регуляризовать численное решение в областях с большими градиентами.

Одним из наиболее эффективных методов построения разностных схем, аппроксимирующих соответствующие дифференциальные уравнения, с заданными свойствами аппроксимации является метод неопределенных коэффициентов.

Построим, например, с помощью метода неопределенных коэффициентов явную разностную схему первого порядка точности, аппроксимирующую неоднородное линейное уравнение переноса

$$u'_t - u'_x = f(t, x). \quad (14.75)$$

Запишем ее в достаточно общем виде (14.4):

$$P_\tau u_\tau = f_\tau, \quad (14.76)$$

где

$$P_\tau u_\tau = \alpha_1 u_m^{n+1} + \alpha_2 u_m^n + \alpha_3 u_{m+1}^n,$$

$$f_\tau = f_m^n,$$

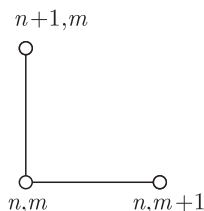


Рис. 14.27

α_i ($i = 1, 2, 3$) — неопределенные коэффициенты, и используем шаблон «правый угол» (рис. 14.27).

Будем искать неопределенные коэффициенты α_i из условия аппроксимации с первым порядком:

$$P_\tau u_\tau = Pu \Big|_{t_n, x_m} + O(\tau + h). \quad (14.77)$$

Разложение двух сеточных функций в ряд Тейлора в окрестности точки $\{t_n, x_m\}$ дает:

$$u_m^{n+1} = u_m^n + \tau (u'_t)_m^n + O(\tau^2),$$

$$u_{m+1}^n = u_m^n + h (u'_x)_m^n + O(h^2).$$

После подстановки этих выражений в уравнение (14.77) с неопределенными коэффициентами получим, положив $\tau = \sigma h$:

$$\begin{aligned} P_\tau u_\tau &= \alpha_1 u_m^{n+1} + \alpha_2 u_m^n + \alpha_3 u_{m+1}^n = (\alpha_1 + \alpha_2 + \alpha_3) u_m^n + \\ &+ \alpha_1 \tau (u'_t)_m^n + \alpha_3 h (u'_x)_m^n + O(\tau^2 + h^2) = (\alpha_1 + \alpha_2 + \alpha_3) u_m^n + \\ &+ \alpha_1 \sigma h (Pu)_m^n + (\alpha_1 \sigma + \alpha_3) h (u'_x)_m^n + O(h^2), \end{aligned}$$

где в последнем равенстве производная по времени $(u'_t)_m^n$ заменена на выражение

$$u'_t = Pu + u'_x.$$

Для того чтобы выполнялись заданные условия аппроксимации с первым порядком по τ и по h , необходимо обеспечить выполнение равенств:

$$\begin{cases} \alpha_1 \sigma h = 1 + O(h), \\ \alpha_1 + \alpha_2 + \alpha_3 = 0 + O(h), \\ h(\alpha_1 \sigma + \alpha_2) = 0 + O(h). \end{cases}$$

Если в правой части опустить члены порядка $O(h)$, то получим систему линейных уравнений следующего вида:

$$\begin{cases} \sigma h \alpha_1 = 1, \\ \alpha_1 + \alpha_2 + \alpha_3 = 0, \\ \alpha_1 \sigma + \alpha_3 = 0, \end{cases}$$

которая имеет единственное решение:

$$\alpha_1 = \frac{1}{\tau}; \quad \alpha_2 = \frac{1}{h} - \frac{1}{\tau}; \quad \alpha_3 = -\frac{1}{h}.$$

Эти коэффициенты соответствуют хорошо известной схеме КИР (Годунова, «явный правый уголок»):

$$\frac{1}{\tau} u_m^{n+1} + \left(\frac{1}{h} - \frac{1}{\tau} \right) u_m^n - \frac{1}{h} u_{m+1}^n = f_m^n,$$

или

$$u_m^{n+1} = u_m^n + \sigma (u_{m+1}^n - u_m^n) + \tau f_m^n.$$

Таким образом, для разностных уравнений, аппроксимирующих дифференциальные уравнения в частных производных, можно использовать следующие методы:

- аппроксимация производных;
- интегро-итерполяционный метод;
- проекционные вариационные методы (Бубнова–Галёркина, Ритца, метод конечных элементов);
- метод неопределенных коэффициентов;
- метод характеристик.

Приведем точные решения одномерного нелинейного уравнения переноса

$$u'_t + a u'_x = 0; \quad u(t, x) = u_0(x); \quad t \geq 0,$$

полезного при тестировании рассмотренных разностных схем для двух начальных данных.

1. Начальное возмущение имеет вид треугольника:

$$u_0(0, x) = \begin{cases} \frac{2(x - x_1)}{x_2 - x_1}, & x \in \left[x_1, \frac{x_1 + x_2}{2} \right]; \\ \frac{2(x_2 - x)}{x_2 - x_1}, & x \in \left[\frac{x_1 + x_2}{2}, x_2 \right]; \\ 0, & x \notin [x_1, x_2]. \end{cases}$$

Точное решение:

$$u(t, x) = \begin{cases} \frac{2(x - x_1)}{(x_2 - x_1) + 2t}, & x \in \left[x_1, \frac{x_1 + x_2}{2} + t \right], \\ & 0 < t \leq \frac{(x_1 - x_2)}{2}; \\ \frac{2(x_2 - x)}{(x_2 - x_1) - 2t}, & x \in \left[x_1, \frac{x_2 + x_1}{2} + t, x_2 \right], \\ & 0 < t \leq \frac{(x_2 - x_1)}{2}; \\ \frac{2(x - x_1)}{(x_2 - x_1) + 2t}, & x \in [x_1, x_1 + C], \quad t > \frac{(x_2 - x_1)}{2}; \\ 0, & x \notin [x_1, x_2 + C], \quad t > \frac{(x_2 - x_1)}{2}. \end{cases}$$

где

$$C = \sqrt{\frac{(x_2 - x_1)(x_2 - x_1) + 2t}{2}}.$$

2. Начальное возмущение имеет вид прямоугольника:

$$u(0, x) = \begin{cases} 1, & x \in [x_1, x_2]; \\ 0, & x \notin [x_1, x_2]. \end{cases}$$

Точное решение:

$$u(t, x) = \begin{cases} \frac{x - x_1}{t}, & x \in [x_1, x_1 + t], \quad 0 < t \leq 2(x_2 - x_1); \\ 1, & x \in [x_1 + t, x_2 + t/2], \quad 0 < t \leq 2(x_2 - x_1); \\ 0, & x \notin [x_1, x_2 + t/2], \quad t \leq 2(x_2 - x_1); \\ \frac{x - x_1}{t}, & x \in [x_1, x_1 + \sqrt{2(x_2 - x_1)}], \\ & t > 2(x_2 - x_1); \\ 0, & x \notin [x_1, x_1 + \sqrt{2(x_2 - x_1)}], \\ & t > 2(x_2 - x_1). \end{cases}$$

Список литературы

1. Годунов С.К., Рябенький В.С. Разностные схемы. М.: Наука. 1973. 400 с.
2. Федоренко Р.П. Введение в вычислительную физику. Долгопрудный, 2008. 503 с.
3. Петров И.Б., Лобанов А.И. Лекции по вычислительной математике. М.: БИНОМ. Лаборатория знаний, 2006. 522 с.

Дополнительная литература

4. Магомедов К. М., Холодов А. С. Сеточно-характеристические методы. М.: Наука, 1988. 288 с.
5. Шокин Ю. И., Яненко Н. Н. Метод дифференциального приближения. Применение к газовой динамике. Новосибирск: Наука, 1985. 364 с.
6. Андерсен Д., Танненхилл Дж., Плетчер Р. Вычислительная гидродинамика и теплообмен. Т. 1. М.: Мир, 1990. 384 с.
7. LeVeque R. J. Finite Volume Methods for Hyperbolic Problems. Cambridge: Cambridge University press, 2011, 558 p.
8. Толстых А. И. Компактные и мультиоператорные аппроксимации высокой точности для уравнений в частных производных. М.: Наука, 2015. 350 с.
9. Белоцерковский О. М. Численное моделирование в механике сплошных сред. М.: Наука. ФИЗМАТЛИТ, 1994. 442 с.
10. Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. Численные методы. М.: ФИЗМАТЛИТ, 2000. 622 с.
11. Холодов Я. А., Уткин П. С., Холодов А. С. Монотонные разностные схемы высокого порядка аппроксимации для систем уравнений гиперболического типа: Учеб. пособие. М.: МФТИ, 2015. 68 с.
12. Головизнин В. М., Самарский А. А. Некоторые свойства разностной схемы «кабаре» // Матем. моделир. 1998. Т. 10, № 1. С. 101–116.
13. Колган В. П. Применение принципа минимальных значений производной к построению конечноразностных схем для расчета разрывных решений газовой динамики // Уч. зап. ЦАГИ. 1972. Т. 3, № 6. С. 68–77.
14. Harten A. High resolution schemes for hyperbolic conservation laws // J. Comput. Phys. 1983. V. 49, Is. 3. P. 357–393.



РАЗНОСТНЫЕ МЕТОДЫ ДЛЯ ЧИСЛЕННОГО РЕШЕНИЯ УРАВНЕНИЙ ЭЛЛИПТИЧЕСКОГО ТИПА (УРАВНЕНИЯ ЭЛЕКТРОСТАТИКИ, ЛАПЛАСА, ПУАССОНА)



15.1. Постановка задачи Дирихле для уравнения Пуассона

В этом разделе будут рассмотрены разностные методы решения задачи Дирихле для уравнения Пуассона в прямоугольной области

$$\tilde{G} = G \cup \Gamma = \{0 \leq x \leq 1; 0 \leq y \leq 1\} \quad (15.1)$$

с границей Γ :

$$\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} = f(x, y),$$

или:

$$\begin{aligned} \Delta \varphi &= f(x, y); \quad \{x, y\} \in G; \\ \varphi|_{\Gamma} &= F(x, y); \quad \{x, y\} \in \Gamma. \end{aligned}$$

Совокупность внутренних точек, принадлежащих прямоугольнику \tilde{G} , будем обозначать \tilde{G}_h , граничных точек — Γ_h .

Аппроксимируем, как и выше, вторые координатные производные разностными соотношениями:

$$\begin{aligned} \Lambda_{xx} \varphi_m &= \Lambda_1 \varphi_m = \frac{\varphi_{m-1,l} - 2\varphi_{m,l} + \varphi_{m+1,l}}{h_x^2} \approx \frac{\partial^2 \varphi}{\partial x^2}, \\ \Lambda_{yy} \varphi_m &= \Lambda_2 \varphi_m = \frac{\varphi_{m,l-1} - 2\varphi_{m,l} + \varphi_{m,l+1}}{h_y^2} \approx \frac{\partial^2 \varphi}{\partial y^2}, \end{aligned}$$

и рассмотрим разностную схему, аппроксимирующую исходное дифференциальное уравнение в частных производных эллиптического типа (см. § 14.1)

$$\frac{\varphi_{m-1,l} - 2\varphi_{m,l} + \varphi_{m+1,l}}{h_x^2} + \frac{\varphi_{m,l-1} - 2\varphi_{m,l} + \varphi_{m,l+1}}{h_y^2} = f_m, \quad \varphi_{m,l} \in G_h, \quad (15.2)$$

которое можно представить в операторном виде

$$\Lambda_{xx} \varphi_{ml} + \Lambda_{yy} \varphi_{ml} = f_{ml},$$

или же

$$\Lambda = f_{ml},$$

где Λ есть оператор следующего вида:

$$\Lambda = \Lambda_{xx} + \Lambda_{yy};$$

граничные условия:

$$\varphi_{ml} = F(x_m, y_l); \quad x_m, y_l \in \Gamma_h.$$

Таким образом, операторная запись сеточной задачи Дирихле будет иметь вид



$$\Lambda \varphi_h = f_h,$$

где

$$\Lambda \varphi_h = \begin{cases} \frac{\varphi_{m+1,l} - 2\varphi_{ml} + \varphi_{m-1,l}}{h_x^2} + \frac{\varphi_{m,l+1} - 2\varphi_{ml} + \varphi_{m,l-1}}{h_y^2}, \\ \{x_m, y_l\} \in G, \\ \varphi_{ml} = F(x_m, y_l); \quad \{x_m, y_l\} \in \Gamma_h. \end{cases} \quad (15.3)$$

Здесь $F(x_m, y_l)$ есть значение функции F_{ml} в точках $\{x_m, y_l\}$, принадлежащих границе Γ_h области интегрирования \tilde{G}_h :

$$f_h = \begin{cases} f(x_m, y_l) = f_{ml}, & \{x_m, y_l\} \in G_h, \\ F(x_m, y_l) = F_{ml}, & \{x_m, y_l\} \in \Gamma_h. \end{cases}$$

Разностная аппроксимация производных в предположении существования четвертых производных дает:

$$\begin{aligned} \varphi_{m+1,l} &= \varphi_{ml} + h_x \varphi'_x + \frac{h_x^2}{2} \varphi''_{xx} + \frac{h_x^3}{6} \varphi'''_{xx} + \frac{h_x^4}{24} \varphi^{(4)}_{xx} + O(h_x^5), \\ \varphi_{m,l+1} &= \varphi_{ml} + h_y \varphi'_y + \frac{h_y^2}{2} \varphi''_{yy} + \frac{h_y^3}{6} \varphi'''_{yy} + \frac{h_y^4}{24} \varphi^{(4)}_{yy} + O(h_y^5). \end{aligned}$$

В таком случае погрешность аппроксимации имеет следующий вид:

$$\xi_h = \frac{h_x^2}{24} \varphi^{(4)}_{xx} + \frac{h_y^2}{24} \varphi^{(4)}_{yy} + O(h_x^4, h_y^4). \quad (15.4)$$

Если в разностном уравнении (15.3) аппроксимировать четвертые производные по x и по y , то получим схему четвертого порядка аппроксимации:

$$\begin{aligned} (\Lambda_{xx} \varphi_{ml} + \Lambda_{yy} \varphi_{ml} - f_{ml}^n) - \frac{1}{12} (h_x^2 \Lambda_{xxxx} \varphi_{ml} + h_y^2 \Lambda_{yyyy} \varphi_{ml}) = \\ = O(h_x^4 + h_y^4), \end{aligned} \quad (15.5)$$

где:

$$\Lambda_{xxxx} \varphi_{ml} = \Lambda_x^{(4)} \varphi_{ml}, \quad \Lambda_{yyyy} \varphi_{ml} = \Lambda_y^{(4)} \varphi_{ml},$$

или же, в более компактном виде,

$$\Lambda \varphi_{ml} - \frac{1}{12} \left(h_x^2 \Lambda_x^{(4)} + h_y^2 \Lambda_y^{(4)} \right) \varphi_{ml} = f_{ml}. \quad (15.6)$$

В дальнейшем нам потребуется знание некоторых важных свойств введенных операторов Λ_{xx} , Λ_{yy} , Λ .

В частности, можно показать, что все они принадлежат пространству линейных операторов S : Λ_{xx} , Λ_{yy} , $\Lambda \in S$, т. е. обладают соответствующими свойствами, являются положительно определенными:

$$(\Lambda_{xx} u_m, u_n) > 0, \quad (\Lambda_{yy} u, u) > 0, \quad (\Lambda u, u) > 0,$$

и самосопряженными:

$$\Lambda_{xx} = \Lambda_{xx}^*, \quad \Lambda_{yy} = \Lambda_{yy}^*, \quad \Lambda = \Lambda^*.$$

Кроме того, эти операторы имеют вещественные собственные значения и соответствующие им собственные функции, которые будут представлены несколько ниже.

Рассмотрим также проблему устойчивости полученной разностной схемы, для чего в пространстве U_h сеточных функций, определенных в \tilde{G} , введем норму

$$\|\varphi_h\|_{U_h} = \max_{x_m, y_l \in G} |\varphi_{ml}|. \quad (15.7)$$

Теорема 15.1 (принцип максимума для сеточных функций). *Каждое численное решение (сеточная функция) разностного уравнения*

$$\Lambda \varphi_{ml} = 0, \quad \{x_m, y_l\} \in G_h,$$

аппроксимирующего дифференциальное уравнение

$$\Delta \varphi = 0, \quad \varphi|_{\Gamma} = F(x, y), \quad \{x, y\} \in G,$$

достигает своих наименьшего и наибольшего значений в точках, принадлежащих границе области интегрирования G_h .

Доказательство.

1. Покажем, что если для сеточной функции φ_{ml} выполняется

$$\Lambda \varphi_{ml} > 0 \quad (15.8)$$

во всех внутренних узлах области интегрирования G , то $\max_{m,l} \varphi_{ml}$ достигается в узле, принадлежащем границе G_h .

Допустим, что этот максимум не достигается на Γ ; в таком случае он достигается в некоторой внутренней точке $\{x_m, y_l\} \in G_h$.

Однако в этой точке выполняется условие (15.8). Если расписать его как разностное, то получим

$$\frac{\varphi_{m-1,l} + u_{m,l-1} - 4u_{m,l} + u_{m+1,l} + u_{m,l+1}}{4} > 0, \quad (15.9)$$

откуда следует

$$\varphi_{ml} < \frac{u_{m-1,l} + u_{m,l-1} + u_{m,l+1} + u_{m+1,l}}{4},$$

что противоречит предположению о том, что сеточная функция φ_{ml} является максимумом (т.е. не меньше, чем каждое из четырех слагаемых).

2. Аналогично устанавливается, что из условия

$$\Delta \varphi_{ml} < 0$$

следует, что минимум сеточной функции φ_{ml} достигается на границе.

Таким образом, из двух доказанных утверждений следует доказательство теоремы.

Из этой теоремы следует факт существования и единственности решения поставленной задачи, так как в силу доказанного принципа максимума эта задача с однородными (нулевыми) граничными условиям имеет только тривиальное решение. Из курса линейной алгебры известно, что если однородная система линейных алгебраических уравнений имеет только тривиальное решение, то решение соответствующей неоднородной системы существует и единственно (задача однозначно разрешена).

Для доказательства устойчивости решения рассматриваемого разностного уравнения рассмотрим вспомогательную функцию (разностную мажоранту Гершгорина) вида

$$U = \frac{1}{4} [R^2 - (x^2 + y^2)] \|f\| + \|F\|,$$

где

$$\|f\| = \max_{\tilde{G}} |f_{ml}|; \quad \|F\| = \max_{\tilde{G}} |f_{ml}|,$$

R — радиус окружности с центром в начале координат $(0,0)$, включающей в себя всю рассматриваемую область \tilde{G} (в нашем случае $R = \sqrt{2}$).

Если к сеточной функции

$$v_{ml} = \varphi_{ml} - U_{ml}$$

применить разностный оператор Лапласа Δ , то получим, что

$$\Delta v_{ml} = f_{ml} + \|f\|$$

во всех внутренних точках области, откуда следует, что в соответствии с доказанной теоремой свое наибольшее значение данная сеточная функция принимает на границе области интегрирования. Однако можно показать, что на границе области интегрирования сеточная функция

$$v_{ml} = \varphi_{ml} - U_{ml}$$

принимает только отрицательные значения, откуда следует, что

$$\varphi_{ml} - U_{ml} \leq 0$$

во всех точках области $\tilde{G} = G \cup \Gamma$.

Если аналогичные рассуждения привести для другой сеточной функции

$$w_{ml} = \varphi_{ml} + U_{ml},$$

то мы получим неравенство

$$\varphi_{ml} + U_{ml} \geq 0$$

также во всей области интегрирования.

В результате объединения двух полученных неравенств будем иметь

$$|\varphi_{ml}| \leq |U_{ml}| \leq \frac{R^2}{4} \|f\| + \|F\|, \quad (15.10)$$

т. е. рассматриваемая разностная схема устойчива.

Рассмотрим теперь более практическую и непростую проблему нахождения решения плохо обусловленной системы линейных алгебраических уравнений высокого порядка с матрицей специальной структуры. Основным вопросом является количество арифметических действий, необходимых для решения СЛАУ с заданной точностью, которое обычно оценивается как $O(N^p)$, где N — порядок системы, $p > 0$.

15.2. Итерационные методы решения задачи Дирихле для уравнения Пуассона

Для численного решения СЛАУ высокого порядка с сильно разреженными матрицами основными методами являются итерационные, в которых по заданному начальному приближению φ_{ml}^0 и алгоритму находят первое приближение u_{ml}^1 , затем по первому — второе u_{ml}^2 и т. д. Итерационный метод является сходящимся, если $u_{ml}^i \xrightarrow{i \rightarrow \infty} U_{ml}$ (U_{ml} — проекция точного решения на расчетную сетку), где $i = 0, 1, \dots$ — номер итерации. Перед реализацией итерационного метода, кроме установления факта

сходимости, необходимо оценить скорость сходимости и количество итераций. Скорость сходимости оценивается из неравенства

$$\|\varphi_{ml}^i - U_{ml}\| \leq Cq^i;$$

здесь C — константа, $0 < q < 1$ — параметр, индивидуальный для каждого метода. Итерации заканчиваются при выполнении условия

$$\|\varphi_{ml}^{i+1} - \varphi_{ml}^i\| \leq \varepsilon,$$

где $\varepsilon \approx Cq^i$, откуда следует оценка количества итераций, необходимых для достижения заданной точности:

$$i(\varepsilon) \approx \ln(\varepsilon/C) / \ln q.$$

Представим метод простых итераций для численного решения рассматриваемого уравнения в следующем виде:

$$\varphi_{ml}^{i+1} = \begin{cases} \varphi_{ml}^i + \tau(\Lambda\varphi_{ml}^i - f_{ml}), & \{x_m, y_l\} \in G_h; \\ F, & \{x_m, y_l\} \in \Gamma_h; \end{cases} \quad (15.11)$$

$$\Lambda\varphi_{ml}^i = \frac{\varphi_{m-1,l}^i - 2\varphi_{m,l}^i + \varphi_{m+1,l}^i}{h_x^2} + \frac{\varphi_{m,l-1}^i - 2\varphi_{m,l}^i + \varphi_{m,l+1}^i}{h_y^2}.$$

Если из (15.11) вычесть очевидное тождество

$$\varphi = \varphi + \tau(\Lambda\varphi - f)$$

во внутренних точках, а из равенства $\varphi_{ml}^{i+1} = \varphi_{ml}^i$ — тождество $\varphi_{ml} = \varphi_{ml}$ в граничных, то для невязки ξ_{ml}^{i+1} получим уравнение

$$\xi_{ml}^{i+1} = \begin{cases} \xi_{ml}^i + \tau\Lambda\xi_{ml}^i, & \{x_m, y_l\} \in G; \\ 0, & \{x_m, y_l\} \in \Gamma, \end{cases} \quad (15.12)$$

которое может быть переписано в виде

$$\xi_{ml}^{i+1} = (E + \tau\Lambda)\xi_{ml}^i. \quad (15.13)$$

Далее переходим к неравенству в нормах:

$$\|\xi_{ml}^{i+1}\| \leq \|E + \tau\Lambda\| \cdot \|\xi_{ml}^i\|, \quad (15.14)$$

или

$$\|\xi_{ml}^i\| \leq \|E + \tau\Lambda\| \cdot \|\xi_{ml}^0\|.$$

Теорема 15.2. Пусть λ_{\min} и λ_{\max} — соответственно минимальное и максимальное собственные числа (границы спектра) оператора $\Lambda = \Lambda^* > 0$ в итерационном процессе

$$\varphi_{ml}^{i+1} = \varphi_{ml}^i + \tau(\Lambda\varphi_{ml}^i - f_{ml}),$$

построенном для решения уравнения в частных производных эллиптического типа:

$$\Delta \varphi = f(x, y). \quad (15.15)$$

Тогда последовательность $\{\varphi_i\}$ сходится к точному решению (15.15), если $0 < \tau < 2/\lambda_{\max}$.

При этом выполняется

$$\|\varphi^i\| \leq q^i \|\varphi^0\|,$$

где $q = \max\{|1 - \tau\lambda_{\min}|, |1 - \tau\lambda_{\max}|\}$ — параметр, принимающий наименьшее значение, равно

$$q_{\min} = \left(1 - \frac{\lambda_{\min}}{\lambda_{\max}}\right) / \left(1 + \frac{\lambda_{\min}}{\lambda_{\max}}\right)$$

при

$$\tau = \tau_{\text{опт}} = \frac{2}{\lambda_{\min} + \lambda_{\max}}. \quad (15.16)$$

Доказательство. Непосредственной подстановкой в известное равенство

$$\Lambda_{xx} \omega_m^{(p)} = -\lambda^{(p)} \omega_m^{(p)}$$

показывается, что числа

$$\lambda^{(p)} = \frac{4}{h^2} \sin^2\left(\frac{p\pi}{2M}\right),$$

где $p = 1, \dots, M-1$ — номер собственного значения, m — номер сеточного узла, являются собственными функциями оператора Λ_{xx} ; $\omega_m^{(p)}$ — соответствующие им собственные функции:

$$\begin{aligned} h^{-2} \left[\sin\left(\frac{m-1}{M}p\pi\right) - 2 \sin\left(\frac{mp\pi}{M}\right) + \sin\left(\frac{m+1}{M}p\pi\right) \right] = \\ = - \left[4h^{-2} \sin^2\left(\frac{p\pi}{M}\right) \cdot \sin\left(\frac{mp\pi}{M}\right) \right]. \end{aligned}$$

Аналогично показывается, что функции

$$\Omega_{ml}^{(pq)} = \omega_m^{(p)} \omega_l^{(q)}$$

являются собственными функциями оператора $\Lambda = \Lambda_{xx} + \Lambda_{yy}$, а соответствующими собственными значениями — величины

$$\lambda^{(pq)} = \lambda^{(p)} + \lambda^{(q)} = 4h^{-2} \left[\sin^2\left(\frac{p\pi}{2M}\right) + \sin^2\left(\frac{q\pi}{2M}\right) \right].$$

Несложно определить границы спектра собственных значений оператора Λ_{xx} :

$$\lambda_{\min} = \lambda^{(1)} = 4h^{-2} \sin^2 \frac{\pi}{2M} \approx \pi^2,$$

поскольку $h = M^{-1}$, $M \gg 1$;

$$\lambda_{\max} = \lambda^{(M-1)} = \frac{4}{h^2} \sin^2 \frac{(M-1)\pi}{2M} \approx \frac{4}{h^2} \sin^2 \frac{\pi}{2} = 4h^{-2} = 4M^2.$$

Соответственно, границами спектра оператора Λ будут значения:

$$\lambda_{\min} \approx 2\pi^2; \quad \lambda_{\max} \approx 8h^{-2} = 8M^2.$$

Число обусловленности системы линейных алгебраических уравнений, аппроксимирующих исходное уравнение в частных производных, вычисляется как отношение

$$\lambda_{\max}/\lambda_{\min} = (2M/\pi)^2,$$

т. е. система плохо обусловлена, поскольку $M^2 \gg 1$.

Функция ξ_{ml} , равная нулю на границах, может быть представлена в виде фурье-разложения по базису из собственных функций $\Omega_{ml}^{(pq)}$ оператора Λ :

$$\xi_{ml}^0 = \sum_{pq} c_{pq} \Omega_{ml}^{(pq)}, \quad c_{pq} = \left(\xi_{ml}^0, \Omega_{ml}^{(pq)} \right),$$

причем

$$\|\xi_{ml}^0\| = (r^0, r^0) = \sqrt{\sum_{pq} c_{pq}^2} = \|c\|$$

— равенство Парсевалю.

Теперь можно провести анализ сходимости итерационного процесса:

$$\begin{aligned} \xi_{ml}^1 &= (E + \tau\Lambda) \xi_{ml}^0 = (E + \tau\Lambda) \sum_{pq} c_{pq} \Omega_{ml}^{(pq)} = \\ &= \sum_{pq} c_{pq} (E + \tau\Lambda) \Omega_{ml}^{(pq)} = \sum_{pq} c_{pq} (1 - \tau\lambda^{(pq)}) \Omega_{ml}^{(pq)}. \end{aligned} \quad (15.17)$$

Последнее равенство объясняется тем, что значения $(1 - \tau\lambda^{(pq)})$ являются собственными числами оператора $E + \tau\Lambda$, что следует из сложения равенств:

$$\tau\Lambda\omega_m^{(p)} = -\tau\lambda\omega_m^{(p)}, \quad E\omega_m^{(p)} = \omega_m^{(p)},$$

откуда имеем

$$(E + \tau\Lambda) \omega_m^{(p)} = (1 - \tau\lambda) \omega_m^{(p)}$$

(для $\Omega_{ml}^{(pq)}$ последнее равенство доказывается аналогично).

Далее можно получить оценку невязки ξ_{ml}^1 по норме:

$$\begin{aligned}\|\xi_{ml}^1\| &= \sqrt{\sum_{pq} c_{pq}^2 (1 - \tau\lambda)^2} \leq \max_{[\lambda_{\min}, \lambda_{\max}]} |1 - \tau\lambda^{(pq)}| \sqrt{\sum_{pq} c_{pq}^2} = \\ &= \max_{[\lambda_{\min}, \lambda_{\max}]} |1 - \tau\lambda^{(pq)}| \cdot \|\xi_{ml}^0\| = q(\tau) \cdot \|\xi_{ml}^0\|,\end{aligned}$$

$$\text{где } q(\tau) = \max_{[\lambda_{\min}, \lambda_{\max}]} |1 - \tau\lambda^{(pq)}|.$$

Аналогичным путем получим оценку для нормы невязки на i -й итерации:

$$\|\xi_{ml}^i\| \leq q^i(\tau) \|\xi_{ml}^0\|. \quad (15.18)$$

Из последнего неравенства видно, что

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

причем итерационный процесс сходится при $0 < q < 1$, т. е. при выполнении условия

$$0 < \tau < \frac{2}{\lambda_{\max}}, \quad (15.19)$$

которое и требовалось доказать.

Будем выбирать итерационный параметр таким образом, чтобы скорость сходимости итерационного процесса была максимальной. При этом мы приходим к типичной «минимаксной» задаче: найти

$$\min_{\tau} \left\{ \max_{[\lambda_{\min}, \lambda_{\max}]} |1 - \tau\lambda| \right\}.$$

Поскольку внутренний максимум

$$\max |1 - \tau\lambda|$$

достигается либо на одной из границ спектра λ (левой λ_{\min} или правой λ_{\max}), то нам необходимо найти минимум

$$\min_{\tau} \{\max |1 - \tau\lambda_{\min}|, \max |1 - \tau\lambda_{\max}|\};$$

который достигается при некотором оптимальном значении $\tau_{\text{опт}}$ параметра τ :

$$\tau_{\text{опт}} = \arg \min_{\tau} \{\max (|1 - \tau\lambda_{\min}|, |1 - \tau\lambda_{\max}|)\}. \quad (15.20)$$

Как видно на графиках двух функций (рис. 15.1):

$$|1 - \tau\lambda_{\min}|, \quad |1 - \tau\lambda_{\max}|,$$

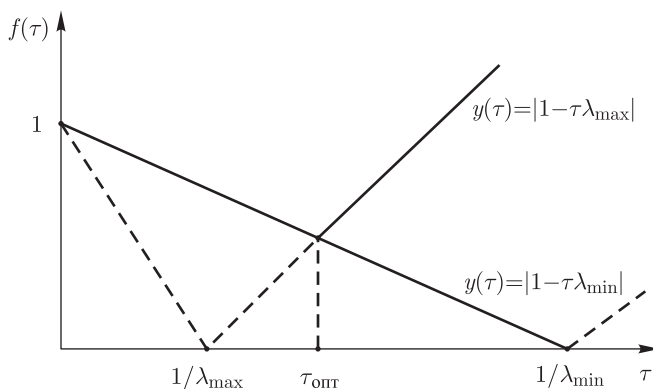


Рис. 15.1

оптимальное значение итерационного параметра $\tau_{\text{опт}}$, при котором достигается минимум в нашей минимаксной задаче, находится из очевидного уравнения (точка пересечения двух графиков)

$$|1 - \tau \lambda_{\min}| = |1 - \tau \lambda_{\max}|,$$

или

$$1 - \tau \lambda_{\min} = -(1 - \tau \lambda_{\max}),$$

откуда получим

$$\tau_{\text{опт}} = \frac{2}{\lambda_{\min} + \lambda_{\max}}. \quad (15.21)$$

Теперь мы сможем оценить скорость сходимости итераций:

$$\begin{aligned} q_{\text{опт}} &= 1 - \tau_{\text{опт}} \lambda_{\min} = 1 - \frac{2}{\lambda_{\min} + \lambda_{\max}} \lambda_{\min} = \frac{\lambda_{\max} - \lambda_{\min}}{\lambda_{\min} + \lambda_{\max}} = \\ &= \frac{1 - \left(\frac{\lambda_{\max}}{\lambda_{\min}}\right)^{-1}}{1 + \left(\frac{\lambda_{\max}}{\lambda_{\min}}\right)^{-1}} = \frac{1 - \left(\frac{\lambda_{\max}}{\lambda_{\min}}\right)^{-1} - 2 \left(\frac{\lambda_{\max}}{\lambda_{\min}}\right)^{-1}}{1 + \left(\frac{\lambda_{\max}}{\lambda_{\min}}\right)^{-1}} \approx 1 - 2 \left(\frac{\lambda_{\max}}{\lambda_{\min}}\right)^{-1}. \end{aligned} \quad (15.22)$$

Теорема доказана.

Заметим, что теперь можно оценить количество итераций, соответствующее полученному итерационному процессу (ε — заданная точность):

$$\begin{aligned} I &= \left\lceil \frac{\ln \varepsilon}{\ln q(\tau)} \right\rceil + 1 = \left\lceil \frac{\ln \varepsilon}{\ln(1 - 2\lambda_{\min}/\lambda_{\max})} \right\rceil + 1 \approx \\ &\approx \left\lceil \frac{\ln \varepsilon}{-2\lambda_{\min}/\lambda_{\max}} \right\rceil + 1 \approx \left\lceil \frac{\lambda_{\max}}{2\lambda_{\min}} \ln(\varepsilon^{-1}) \right\rceil + 1. \end{aligned}$$

Если вычисления проводятся на расчетной сетке $10^2 \times 10^2$ при $\varepsilon = 10^{-5}$ ($\lambda_{\min} = 2\pi^2$, $\lambda_{\max} = 8M^2$), то

$$I = \frac{8M^2}{2 \cdot 2\pi^2} \cdot \ln 10^5 \approx 2 \cdot 10^4; \quad (15.23)$$

при этом

$$q_{\text{опт}} \approx 1 - 2 \left(\frac{\lambda_{\max}}{\lambda_{\min}} \right)^{-1} \approx 0,9995.$$

Напомним, что

$$\mu = \lambda_{\max} / \lambda_{\min}$$

— это число обусловленности, являющееся важной характеристикой матрицы рассматриваемой системы линейных алгебраических уравнений.

Чем больше это число, тем больше требуется вычислений для достижения приемлемой точности. В нашем случае количество итераций $\sim 10^4$ арифметических операций, однако цена каждой итерации приблизительно $10M^2$ арифметических операций, т. е. необходимо количество операций $\sim 10^9$, так что данный метод требует значительных затрат машинного времени. По этой причине были затрачены большие усилия для разработки итерационных методов, существенно уменьшивших количество арифметических операций. Приведенный же метод тем не менее имеет большой методический смысл, необходимый для понимания современных итерационных методов.

Рассмотрим другой итерационный процесс, использующий важное свойство полинома П. Л. Чебышёва (чебышёвское ускорение, итерационный метод с чебышёвскими параметрами).

В предыдущем процессе для погрешности было получено следующее выражение:

$$\xi_{ml}^i = \sum c_{pq} \left(1 - \tau_{\text{опт}} \lambda^{(pq)} \right)^i \omega_{ml}^{(pq)},$$

из которого видно, что ослабление фурье-компонент проходит неравномерно: в «средней» части спектра ($\lambda \approx \lambda_{\max}/2$) заметно быстрее, чем на краях. Логично было бы выбирать итерационные параметры так, чтобы убывание невязки было более равномерным по всем значениям спектра в фурье-разложении. Такой результат может быть достигнут выбором своего итерационного параметра на каждой итерации:

$$\varphi^{i+1} = \varphi^i + \tau_{i+1} (\Lambda \varphi^i - f). \quad (15.24)$$

В этом случае выражение для погрешности будет иметь следующий вид:

$$\xi_{ml}^{i+1} = (E + \tau_{i+1} \Lambda) \xi_{ml}^i, \quad (15.25)$$

причем

$$\xi_{ml}^i = \prod_{j=1}^i (E + \tau_j \Lambda) r_{ml}^0. \quad (15.26)$$

После фурье-разложения ξ_{ml}^i по базису из собственных функций $\Omega_{ml}^{(pq)}$ из (15.17) получим:

$$\begin{aligned} \sum_{pq} c_{pq}^{i+1} \Omega_{ml}^{(pq)} &= \sum_{pq} c_{pq}^i (E + \tau_{i+1} \Lambda) \Omega_{ml}^{(pq)} = \\ &= \sum_{pq} c_{pq}^i (1 - \tau_{i+1}) \lambda^{(pq)} \Omega_{ml}^{(pq)}, \end{aligned}$$

т. е. коэффициенты фурье-разложения на i -й и $(i+1)$ -й итерациях связаны соотношением

$$c_{pq}^{i+1} = c_{pq}^i \left(1 - \tau_{i+1} \lambda^{(pq)} \right),$$

из которого получим

$$c_{pq}^i = \prod_{j=1}^i \left(1 - \tau_j \lambda^{(pq)} \right) c_{pq}^0.$$

В таком случае выражение для невязки будет иметь вид

$$\xi_{ml}^i = \sum_{pq} c_{pq}^i \Omega_{ml}^{(pq)} = \sum_{pq} c_{pq}^0 \prod_j \left(1 - \tau_j \lambda^{(pq)} \right) \Omega_{ml}^{(pq)}. \quad (15.27)$$

Далее оценим величину невязки на i -й итерации по норме:

$$\begin{aligned} \|\xi_{ml}^i\| &= \max_{[\lambda_{\min}, \lambda_{\max}]} \left| \prod_{j=1}^i (1 - \tau_j \lambda) \right| \cdot \left\| \sum c_{pq}^0 \cdot \Omega_{ml} \right\| \leq \\ &\leq \max_{[\lambda_{\min}, \lambda_{\max}]} \left| \prod_{j=1}^i (1 - \tau_j \lambda) \right| \cdot \|\xi_{ml}^0\|. \end{aligned}$$

Желая построить наиболее эффективный, в смысле скорости сходимости, итерационный процесс, мы вновь приходим к минимаксной задаче: определить последовательность итерационных параметров τ_k ($k = 1, \dots, i$) такую, что будет достигнут минимум:

$$\min_{\{\tau_k\}} \max_{[\lambda_{\min}, \lambda_{\max}]} \left| \prod_{k=1}^i (1 - \tau_k \lambda) \right|.$$

Поскольку выражение

$$\prod_{k=1}^i (1 - \tau_k \lambda)$$

представляет собой полином степени i относительно τ , то мы пришли к классической задаче о полиноме, наименее уклоняющемся от нуля на интервале $[\lambda_{\min}, \lambda_{\max}]$. Этот полином, как хорошо известно, есть полином Чебышёва, а итерационные параметры τ_k^i являются величинами, обратными значениям корней этого полинома:

$$\tau_k^i = \left[\frac{\lambda_{\min} + \lambda_{\max}}{2} + \frac{\lambda_{\max} - \lambda_{\min}}{2} \cos\left(\frac{2k-1}{2i}\pi\right) \right]^{-1}, \quad k = 1, 2, \dots, i. \quad (15.28)$$

Опуская промежуточные выкладки, дадим оценку скорости сходимости этого метода:

$$q \approx 1 - 2\sqrt{(\lambda_{\min}/\lambda_{\max})^{-1}} = 1 - 2\sqrt{\mu^{-1}};$$

$$I \approx \left[\frac{1}{2}\sqrt{\mu} \ln \varepsilon^{-1} \right] + 1.$$

Для задачи с теми же параметрами, которые приводились в случае метода с одним оптимальным итерационным параметром $\tau_{\text{опт}}$ ($M = 100$, $\varepsilon = 10^{-5}$), получим:

$$q \approx 0,968, \quad i \approx 360,$$

т.е. чебышёвский метод позволяет сократить количество арифметических операций на два порядка по сравнению с предыдущим методом. Однако попытки применения рассмотренного метода привели к парадоксальному результату — быстрому росту решения задачи в ходе итерационного процесса. Причина оказалась в быстром росте ошибок округления и сгущением величин, обратных корням полинома Чебышёва ($1/\tau_k$), вблизи границ спектра. Оказалось, что это явление можно устранить, если изменить порядок чередования итерационных параметров определенным образом [4]: так, чтобы при любом i частные произведения

$$\prod_{k=1}^i (1 - \tau_k \lambda)$$

не возрастали вблизи границ спектра.

Например, для $i = 4$ получим следующее чередование чебышёвских параметров:

$$\{1, 4, 2, 3\},$$

для $i = 8$:

$$\{1, 8, 4, 5, 2, 7, 3, 6\}$$

и т. д.

Такие итерационные процессы оказываются устойчивыми и существенно быстрее сходящимися: так, в случае $\lambda_{\max} = 8M^2$, $\lambda_{\min} = 2\pi^2$, $M = 100$ за M итераций погрешность убывает в 20 раз, а в случае одного параметра она умножается только на 0,95.

Указанного недостатка лишен трехэтапный метод Чебышёва, который можно представить в следующем виде:

$$\varphi_{ml}^1 = (E - \tau A) \varphi_{ml}^0 + \tau f_{ml},$$

$$\varphi_{ml}^{i+2} = \frac{2\alpha_1\alpha_i}{\alpha_{i+1}} (E - \tau A) \varphi_{ml}^{i+1} - \frac{\alpha_i}{\alpha_{i+1}} \varphi_{ml}^i + \frac{2\alpha_1\alpha_i}{\alpha_{i+1}} f, \quad i = 1, 2, \dots;$$

здесь:

$$\tau = \frac{2}{\lambda_{\min} + \lambda_{\max}}, \quad \alpha_0 = 1, \quad \alpha_1 = \frac{\mu+1}{\mu-1}, \quad \alpha_{i+2} = 2\alpha_1\alpha_{i+1} - \alpha_i.$$

Заметим, что двухслойный итерационный метод может быть записан в канонической форме

$$B \frac{\varphi^{i+1} - \varphi^i}{\tau_{i+1}} + A \varphi^i = f.$$

При $B = E$ такой метод называется *явным*, в противном случае — *неявным*.

Каноническая форма трехэтапного итерационного метода имеет вид

$$B \frac{\tilde{\varphi} - \varphi^i}{\tau_{i+1}} + A \varphi^i = f, \\ \varphi^{i+1} = \alpha_{i+1} \tilde{\varphi} + (1 - \alpha_{i+1}) \varphi^{i-1}.$$

При $\alpha_i = 1$ трехслойная схема переходит в двухслойную. В рассмотренных методах полагалось:

$$A = A^* > 0, \quad 0 < \lambda_{\min} < \lambda_i < \lambda_{\max}.$$

Важным моментом в приведенных процессах является то, что для их реализации необходимо только знание границ спектра.

Добиться более высокого ускорения итерационного процесса для численного решения уравнения Пуассона оказалось возможным, если применить метод установления. Для этого рассматривается нестационарное уравнение

$$\varphi'_t = \Delta \varphi - f \quad (15.29)$$

со стационарными граничными условиями. В этом случае при $t \rightarrow \infty$ решение такого уравнения будет стремиться к решению

стационарного уравнения. Для решения (15.29) воспользуемся методом переменных направлений:

$$\begin{cases} \frac{\varphi_{ml}^{i+1/2} - \varphi_{ml}^i}{\tau} = \Lambda_{xx} \varphi_{ml}^{i+1/2} + \Lambda_{yy} \varphi_{ml}^i - f_{ml}, \\ \frac{\varphi_{ml}^{i+1} - \varphi_{ml}^i}{\tau} = \Lambda_{xx} \varphi_{ml}^{i+1/2} + \Lambda_{yy} \varphi_{ml}^{i+1} - f_{ml}. \end{cases} \quad (15.30)$$

Вычитая из этих уравнений очевидное тождество

$$\frac{\varphi_{ml} - \varphi_{ml}}{\tau} = \Lambda_{xx} \varphi_{ml} + \Lambda_{yy} \varphi_{ml} - f,$$

получим уравнения для погрешности:

$$\begin{cases} \frac{\xi_{ml}^{i+1/2} - \xi_{ml}^i}{\tau} = \Lambda_{xx} \xi_{ml}^{i+1/2} + \Lambda_{yy} \xi_{ml}^i, \end{cases} \quad (15.31a)$$

$$\begin{cases} \frac{\xi_{ml}^{i+1} - \xi_{ml}^{i+1/2}}{\tau} = \Lambda_{xx} \xi_{ml}^{i+1/2} + \Lambda_{yy} \xi_{ml}^{i+1}. \end{cases} \quad (15.31b)$$

Далее представим ξ_{ml}^i и $\xi_{ml}^{i+1/2}$ в виде фурье-разложения, как и выше:

$$\xi_{ml}^i = \sum_{pq} c_{pq}^i \Omega_{ml}^{(pq)}; \quad \xi_{ml}^{i+1/2} = \sum_{pq} c_{pq}^{i+1/2} \Omega_{ml}^{(pq)}. \quad (15.32)$$

Из (15.31a) получим

$$(E - \tau \Lambda_{xx}) \xi_{ml}^{i+1/2} = (E + \tau \Lambda_{yy}) \xi_{ml}^i, \quad (15.33)$$

или, с учетом (15.32),

$$(E - \tau \Lambda_{xx}) \sum_{pq} c_{pq}^{i+1/2} \Omega_{ml}^{(pq)} = (E + \tau \Lambda_{yy}) \sum_{pq} c_{pq}^i \Omega_{ml}^{(pq)}.$$

После введения операторов под знаки сумм будем иметь

$$\sum_{pq} c_{pq}^{i+1/2} (1 + \tau \lambda^{(p)}) \Omega_{ml}^{(pq)} = \sum_{pq} c_{pq}^i (1 - \tau \lambda^{(q)}) \Omega_{ml}^{(pq)}, \quad (15.34)$$

где $\lambda^{(p)}$, $\lambda^{(q)}$ — собственные значения операторов Λ_{xx} и Λ_{yy} соответственно. Из последнего равенства вытекает

$$c_{pq}^{i+1/2} = \frac{1 - \tau \lambda^{(q)}}{1 + \tau \lambda^{(p)}} c_{pq}^i. \quad (15.35)$$

После фурье-разложения на втором этапе итерационного процесса получим

$$c_{pq}^{i+1} = \frac{1 - \tau \lambda^{(q)}}{1 + \tau \lambda^{(q)}} c_{pq}^{i+1/2} = \frac{1 - \tau \lambda^{(q)}}{1 + \tau \lambda^{(q)}} \cdot \frac{1 - \tau \lambda^{(p)}}{1 + \tau \lambda^{(q)}} c_{pq}^{i+1}. \quad (15.36)$$

После введения обозначения

$$v(\tau) = \max_{[\lambda_{\min}, \lambda_{\max}]} \frac{|1 - \tau\lambda|}{|1 + \tau\lambda|} \quad (15.37)$$

получим неравенство

$$|c_{pq}^{i+1}| \leq \nu^2 |c_{pq}^i|. \quad (15.38)$$

В таком случае для нормы погрешности ξ^{i+1} имеем

$$\|\xi^{i+1}\| \leq \mu^2 \cdot \|\xi^i\|, \quad (15.39)$$

так как

$$\begin{aligned} \|\xi^{i+1}\| &= \left\| \sum_{pq} c_{pq}^{i+1} \Omega^{(pq)} \right\| \leq \left\| \sum_{pq} \nu^2 \cdot c_{pq}^{i+1} \cdot \Omega^{(pq)} \right\| \leq \\ &\leq \mu^2 \left\| \sum_{pq} c_{pq}^{i+1} \Omega^{(pq)} \right\| = \mu^2 \|\xi^i\|. \end{aligned}$$

Для того чтобы найти оптимальное значение итерационного параметра τ , доставляющего

$$\min_{\tau} v(\tau),$$

нужно решить знакомую нам минимаксную задачу:

$$\tau = \arg \left\{ \min_{\tau} \max_{[\lambda_{\min}, \lambda_{\max}]} \left| \frac{1 - \tau\lambda}{1 + \tau\lambda} \right| \right\},$$

для которой, как можно увидеть из анализа графика функции $\left| \frac{1 - \tau\lambda}{1 + \tau\lambda} \right|$, выполняется

$$\max_{[\lambda_{\min}, \lambda_{\max}]} \left| \frac{1 - \tau\lambda}{1 + \tau\lambda} \right| = \max_{[\lambda_{\min}, \lambda_{\max}]} \left\{ \left| \frac{1 - \tau\lambda_{\min}}{1 + \tau\lambda_{\min}} \right|, \left| \frac{1 - \tau\lambda_{\max}}{1 + \tau\lambda_{\max}} \right| \right\}.$$

Минимум достигается при выполнении равенства

$$\frac{1 - \tau_{\text{опт}}\lambda_{\min}}{1 + \tau_{\text{опт}}\lambda_{\min}} = -\frac{1 - \tau_{\text{опт}}\lambda_{\max}}{1 + \tau_{\text{опт}}\lambda_{\max}},$$

т. е. при

$$\tau_{\text{опт}} = \frac{1}{\sqrt{\lambda_{\min} \cdot \lambda_{\max}}}.$$

Количество итераций в вышеприведенном примере, требуемое для достижения заданной точности ε , есть

$$i \approx \frac{1}{4} \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} \ln \varepsilon^{-1} = \frac{1}{4} \sqrt{\mu} \ln \varepsilon^{-1}.$$

Обобщением приведенного итерационного метода является введение набора чебышёвских параметров τ_i , что приводит к минимаксной задаче:

$$\min_{\{\tau_i\}} \max_{[\lambda_{\min}, \lambda_{\max}]} \prod_{k=1}^i \left| \frac{1 - \tau_i \lambda}{1 + \tau_i \lambda} \right|^2.$$

Оценка количества итераций в этом случае дает

$$i \approx [(\ln \mu) \cdot \varepsilon^{-1}] + 1.$$

Приведем данные по количеству итераций для различных методов.

1. Метод Якоби: $2M^2/\pi^2$.
2. Метод простых итераций с оптимальным параметром: $2M^2/\pi^2$.
3. Метод Зейделя: M^2/π^2 .
4. Метод верхней релаксации: $2M/\pi$.
5. Метод итераций с чебышёвскими итерационными параметрами: M/π .
6. Метод переменных направлений с оптимальными итерационными параметрами: $(1/2)(M/\pi)$.
7. Метод переменных направлений с чебышёвскими итерационными параметрами: $\frac{2}{\alpha} \ln \left(\frac{2}{\pi} M \right)$, $\alpha \approx 3,2$.

Список литературы

1. Федоренко Р.П. Введение в вычислительную физику. Долгопрудный: Интеллект, 2008. 503 с.
2. Рябенький В.С. Введение в вычислительную математику. М.: ФИЗМАТЛИТ, 2008. 288 с.
3. Петров И.Б., Лобанов А.И. Лекции по вычислительной математике. М.: БИНОМ. Лаборатория знаний, 2006. 522 с.

Дополнительная литература

4. Самарский А.А., Николаев Е.С. Методы решения сеточных уравнений. М.: Наука, 1978. 591 с.
5. Марчук Г.И. Методы вычислительной математики. М.: Наука, 1989. 608 с.
6. Самарский А.А. Теория разностных схем. М.: Наука. 1983. 656 с.

МАТЕМАТИЧЕСКИЕ МОДЕЛИ МЕХАНИКИ СПЛОШНЫХ СРЕД (МСС)

16.1. Вывод уравнений механики сплошных сред

Рассмотрим некоторый объем сплошной среды V_0 и закон сохранения массы для него (уравнения неразрывности) в трехмерном пространстве [1]. Масса жидкости в этом объеме равна

$$\int_{V_0} \rho \cdot dV, \quad (16.1)$$

где ρ — плотность среды, заключенной в рассматриваемом объеме. Через элементы dS поверхности, ограничивающей этот объем, в единицу времени протекает количество $\rho \mathbf{v} dS$ жидкости (вектор dS по модулю равен площади $|dS|$ и направлен по внешней нормали (рис. 16.1)). Тогда

$$\rho \mathbf{v} dS > 0,$$

если жидкость вытекает из объема, и

$$\rho \mathbf{v} dS < 0,$$

если втекает. Полное количество жидкости, вытекающей в единицу времени из V_0 , будет

$$\oint_{S_0} \rho \mathbf{v} dS, \quad (16.2)$$

где интегрирование проводится по всей замкнутой поверхности, охватывающей объем V_0 [1].

С другой стороны, уменьшение количества жидкости в объеме V_0 есть

$$-\frac{\partial}{\partial t} \int_{V_0} \rho dV. \quad (16.3)$$

Приравнявая (16.2) и (16.3), получим

$$-\frac{\partial}{\partial t} \int_{V_0} \rho dV = \oint_{S_0} \rho \mathbf{v} dS. \quad (16.4)$$

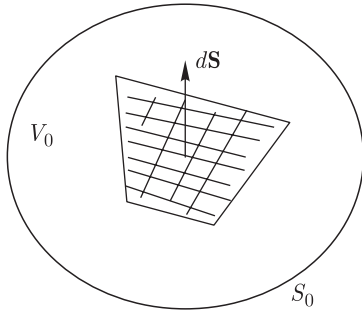


Рис. 16.1

Интеграл по поверхности в (16.4) преобразуется по формуле Остроградского–Гаусса в интеграл по объему:

$$\oint_{S_0} \rho \mathbf{v} d\mathbf{S} = \int_{V_0} \operatorname{div} (\rho \mathbf{v}) dV,$$

после чего получаем закон сохранения массы в произвольном объеме V_0 :

$$\int_{V_0} \left(\frac{\partial \rho}{\partial t} + \operatorname{div} (\rho \mathbf{v}) \right) dV = 0. \quad (16.5)$$

В силу произвольности объема V_0 имеем

$$\frac{\partial \rho}{\partial t} + \operatorname{div} (\rho \mathbf{v}) = 0 \quad (16.6)$$

— уравнение неразрывности; или же, после раскрытия выражения $\operatorname{div} (\rho \mathbf{v})$:

$$\frac{\partial \rho}{\partial t} + \rho \cdot \operatorname{div} \mathbf{v} + \mathbf{v} \cdot \operatorname{grad} \rho = 0 \quad (16.7)$$

(дивергентная и недивергентная формулы).

Рассмотрим закон сохранения энергии (уравнения энергии).

Исследуем, как изменится энергия в неподвижном объеме V_0 . Изменение количества этой энергии, заключенной в объеме, есть

$$\frac{\partial}{\partial t} \int_{V_0} \left(\frac{\rho |\mathbf{v}|^2}{2} + \rho \varepsilon \right) dV; \quad (16.8)$$

здесь

$$e = \frac{\rho |\mathbf{v}|^2}{2} + \rho \varepsilon \quad (16.9)$$

— энергия единицы объема (плотность энергии) (ε — внутренняя энергия единицы массы жидкости; первый член — кинетическая энергия, второй — внутренняя).

Полная энергия проходящей через поверхность S массы жидкости, переносимая в единицу времени, равна

$$\oint_{S_0} \rho \mathbf{v} \left(\varepsilon + \frac{|\mathbf{v}|^2}{2} \right) d\mathbf{S}. \quad (16.10)$$

Работа, производимая силами давления над жидкостью, заключенной в V_0 , есть

$$\oint_{S_0} p \mathbf{v} \cdot d\mathbf{S}. \quad (16.11)$$

Приравниваем изменение энергии единицы объема жидкости к энергии, переносимой в единицу объема через поверхность S , и работе сил давления:

$$\frac{\partial}{\partial t} \int_{V_0} \left(\frac{\rho |\mathbf{v}|^2}{2} + \rho \varepsilon \right) dV = - \oint_{S_0} \rho \mathbf{v} \left(\varepsilon + \frac{|\mathbf{v}|^2}{2} \right) d\mathbf{S} - \oint_{S_0} p \mathbf{v} \cdot d\mathbf{S} \quad (16.12)$$

(знак в правой части связан с направлением внутри $d\mathbf{S}$).

Преобразуем (16.12) по формуле Остроградского–Гаусса:

$$\frac{\partial}{\partial t} \int_{V_0} e dU + \int_{V_0} \operatorname{div} (e \mathbf{v}) dU + \int_{V_0} \operatorname{div} (p \mathbf{v}) dU = 0, \quad (16.13)$$

или

$$\int_{V_0} \left(\frac{\partial e}{\partial t} + \operatorname{div} (e + p) \mathbf{v} \right) dV = 0, \quad (16.14)$$

где $\rho \mathbf{v} (\mathbf{v}^2/2 + p)$ — вектор плотности потока энергии.

В силу произвольности объема V_0 получим

$$\frac{\partial e}{\partial t} + \operatorname{div} (e + p) \mathbf{v} = 0 \quad (16.15)$$

— уравнение энергии.

Закон сохранения движения. Импульс единицы объема жидкости есть $\rho \mathbf{v}$. Его изменение в объеме V равно

$$\frac{\partial}{\partial t} \int_{V_0} \rho \mathbf{v} dV. \quad (16.16)$$

Введем симметричный тензор второго ранга:

$$\Pi_{ik} = p \delta_{ik} + \rho v_i v_k \quad (16.17)$$

— тензор плотности импульса; поток вектора импульса через поверхность, перпендикулярную единичному вектору \mathbf{n} , есть

$$\mathbf{P} = p \mathbf{n} + \rho \mathbf{v} (\mathbf{v} \cdot \mathbf{n}); \quad (16.18)$$

в частности, если \mathbf{n} направлен по \mathbf{v} , то

$$|\mathbf{P}| = p + \rho |\mathbf{v}|^2 \quad (16.19)$$

в направлении, перпендикулярном скорости.

Поток вектора импульса через всю поверхность S есть

$$\int_{S_0} \mathbf{P} d\mathbf{S}, \quad (16.20)$$

или, с учетом (16.16), (16.17):

$$\frac{\partial}{\partial t} \int_{V_0} \rho \mathbf{v} dV = - \oint_{S_0} \mathbf{\Pi} d\mathbf{S}, \quad (16.21)$$

а так как

$$\oint_{S_0} \mathbf{\Pi} d\mathbf{S} = \int_{V_0} \operatorname{div} \mathbf{\Pi} \cdot dV,$$

то

$$\int_{V_0} \left[\frac{\partial}{\partial t} (\rho \mathbf{v}) + \operatorname{div} \mathbf{\Pi} \right] \cdot dV = 0, \quad (16.22)$$

откуда, в силу произвольности выбранного объема, имеем

$$\frac{\partial}{\partial t} (\rho \mathbf{v}) + \operatorname{div} \mathbf{\Pi} = 0 \quad (16.23)$$

— уравнение сохранения импульса объема.

16.2. Уравнения МСС в интегральной форме

Как известно из курса математического анализа, в одномерном случае для двух дифференцируемых функций R и Q справедлива формула Грина

$$\iint_{\Omega} \left(\frac{\partial R}{\partial t} + \frac{\partial Q}{\partial x} \right) dt dx = \int_{\Gamma} R dx - Q dt = 0, \quad (16.24)$$

где Ω — область в системе координат t, x , Γ — ее граница (рис. 16.2). Например, для уравнения неразрывности имеем [2]

$$\begin{aligned} \iint_{\Omega} \left(\frac{\partial \rho}{\partial t} + \frac{\partial (\rho u)}{\partial x} \right) dt dx &= \\ &= \int_{\Gamma} \rho dx - \rho u dt = 0, \end{aligned} \quad (16.25)$$

или

$$\int_{\Gamma} \rho dx - (\rho u) dt = 0 \quad (16.26)$$

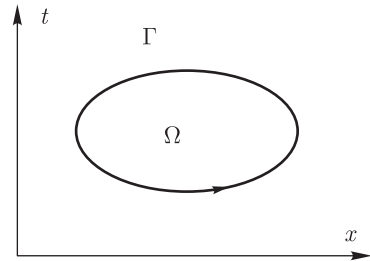


Рис. 16.2

— уравнение неразрывности в интегральной форме.

16.3. Система уравнений газовой динамики

Система уравнений газодинамики в переменных Эйлера (фиксированная в пространстве система координат), записанная в дивергентной форме, может быть представлена в таком виде [3]:

$$\begin{cases} \frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \mathbf{v}) = 0, & \text{уравнение неразрывности;} \\ \frac{\partial e}{\partial t} + \operatorname{div}(e + p) \mathbf{v} = 0, & \text{уравнение энергии;} \\ \frac{\partial(\rho v)}{\partial t} + \operatorname{div} \mathbf{\Pi} = 0, & \text{уравнение движения;} \end{cases} \quad (16.27)$$

$$\begin{cases} \varepsilon = \varepsilon(p, \rho) \\ \varepsilon = \varepsilon(\rho, T), \\ e = \rho \varepsilon + \rho \frac{|v|^2}{2}. \end{cases} \quad (16.28)$$

— уравнения состояния, составляющие систему (16.28).

Система уравнений (16.27) (16.28) в скалярной форме для трехмерного случая имеет следующий вид:

$$\begin{cases} \frac{\partial \rho}{\partial t} + \frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} = 0, & \text{уравнение неразрывности;} \\ \frac{\partial(\rho u)}{\partial t} + \frac{\partial(p + \rho u^2)}{\partial x} + \frac{\partial(\rho uv)}{\partial y} + \frac{\partial(\rho vw)}{\partial z} = 0, & \text{уравнение движения;} \\ \frac{\partial(\rho v)}{\partial t} + \frac{\partial(\rho uv)}{\partial x} + \frac{\partial(p + \rho v^2)}{\partial y} + \frac{\partial(\rho vw)}{\partial z} = 0, & \text{уравнение движения;} \\ \frac{\partial(\rho w)}{\partial t} + \frac{\partial(\rho uw)}{\partial x} + \frac{\partial(\rho vw)}{\partial y} + \frac{\partial(p + \rho w^2)}{\partial z} = 0, & \text{уравнение движения;} \\ \frac{\partial e}{\partial t} + \frac{\partial(e + p)u}{\partial x} + \frac{\partial(e + p)v}{\partial y} + \frac{\partial(e + p)w}{\partial z} = 0, & \text{уравнение энергии} \end{cases} \quad (16.29)$$

— уравнения Эйлера в декартовой системе координат.

Эти же уравнения можно записать в матричной форме:

$$\frac{\partial \mathbf{U}}{\partial t} + \frac{\partial \mathbf{E}}{\partial x} + \frac{\partial \mathbf{F}}{\partial y} + \frac{\partial \mathbf{G}}{\partial z} = 0,$$

$$\mathbf{U} = \begin{Bmatrix} \rho \\ \rho u \\ \rho v \\ \rho w \\ e \end{Bmatrix}, \quad \mathbf{E} = \begin{Bmatrix} \rho \\ \rho u^2 + p \\ \rho uv \\ \rho uw \\ (e + p)u \end{Bmatrix},$$

$$\mathbf{F} = \begin{Bmatrix} \rho v \\ \rho uv \\ \rho v^2 + p \\ \rho vw \\ (e + p)v \end{Bmatrix}, \quad \mathbf{G} = \begin{Bmatrix} \rho w \\ \rho uw \\ \rho vw \\ \rho w^2 + p \\ (e + p)w \end{Bmatrix}.$$

Систему уравнений газодинамики также можно представить в векторной дивергентной форме:

$$\begin{cases} \frac{\partial \rho}{\partial t} + \rho \cdot \operatorname{div} \mathbf{v} + \mathbf{v} \cdot \operatorname{grad} \rho = 0, \\ \frac{\partial(\rho \mathbf{v})}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{1}{\rho} \cdot \operatorname{grad} p. \end{cases} \quad (16.30a)$$

$$(16.30b)$$

Если жидкость находится в поле сил тяжести, то на каждую единицу объема действует сила $\rho \mathbf{g}$, тогда (16.30a) приобретает вид

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{\nabla p}{\rho} + \mathbf{g}. \quad (16.31)$$

Система одномерных уравнений газовой динамики в переменных Лагранжа (поточная система координат) имеет следующий вид [2]:

$$\begin{cases} \frac{du}{dt} + \frac{1}{\rho} \cdot \frac{\partial p}{\partial \xi} \cdot J = 0, \\ \frac{\partial \rho}{\partial t} + \rho \cdot \frac{\partial u}{\partial \xi} \cdot J = 0, \\ \frac{\partial e}{\partial t} + \rho^{-1} \cdot \frac{\partial(pu)}{\partial \xi} \cdot J = 0, \\ \frac{dx}{dt} = u(t, \xi), \end{cases} \quad (16.32)$$

где $J = (\partial x / \partial \xi)^{-1}$ — якобиан перехода от эйлеровых к лагранжевым координатам; здесь x, ξ — соответственно эйлеровы и лагранжевы координаты.

Скалярная недивергентная форма трехмерных уравнений движения (в эйлеровых координатах) имеет такой вид:

$$\frac{dV_i}{dt} + V_k \frac{dV_i}{dx_k} = -\frac{1}{\rho} \cdot \frac{dp}{dx_i}, \quad \mathbf{V} = \{u, v, w\}, \quad i, k = 1, 2, 3. \quad (16.33)$$

16.4. Уравнение Навье–Стокса, описывающее течение вязкой жидкости

Система уравнений, описывающих движение несжимаемой вязкой жидкости (уравнение Навье–Стокса), записанная в переменных Эйлера, может быть представлена в следующем виде [3]:

$$\left\{ \begin{array}{ll} \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = \mathbf{F} - \frac{1}{\rho} \text{grad } p + \frac{\mu}{\rho} \cdot \Delta \mathbf{v}, & \text{уравнение движения;} \\ \text{div } \mathbf{v} = 0, & \text{уравнение неразрывности;} \\ \frac{\partial T}{\partial t} + (\mathbf{v}, \text{grad } T) = \frac{\lambda}{\rho c} \Delta T + \frac{1}{\rho c} \mu \Phi, & \text{уравнение энергии.} \end{array} \right. \quad (16.34)$$

Здесь: Φ — «диссипативная функция», имеющая вид

$$\Phi = 2 \left[\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial v}{\partial y} \right)^2 + \left(\frac{\partial w}{\partial z} \right)^2 \right] + \\ + \left(\frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \right)^2 + \left(\frac{\partial w}{\partial y} + \frac{\partial v}{\partial z} \right)^2 + \left(\frac{\partial w}{\partial z} + \frac{\partial w}{\partial z} \right)^2 + \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \right)^2;$$

μ — коэффициент вязкости.

Граничные условия для системы уравнений динамики вязкой жидкости: $\mathbf{v} = 0$; более общее условие — скольжение с трением. Скалярная форма системы уравнений Навье–Стокса представляется в следующем виде:

$$\left\{ \begin{array}{ll} \frac{du}{dt} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z} = \\ = F_x - \frac{1}{\rho} \cdot \frac{\partial p}{\partial x} + \frac{\mu}{\rho} \Delta u, & \text{уравнение движения;} \\ \frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + w \frac{\partial v}{\partial z} = \\ = F_y - \frac{1}{\rho} \cdot \frac{\partial p}{\partial y} + \frac{\mu}{\rho} \Delta v, & \text{уравнение движения;} \\ \frac{\partial w}{\partial t} + u \frac{\partial w}{\partial x} + v \frac{\partial w}{\partial y} + w \frac{\partial w}{\partial z} = \\ = F_z - \frac{1}{\rho} \cdot \frac{\partial p}{\partial z} + \frac{\mu}{\rho} \Delta w, & \text{уравнение движения;} \\ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, & \text{уравнение неразрывности.} \end{array} \right. \quad (16.35)$$

16.5. Система уравнений теории упругости

Обобщенный закон Гука в теории упругости (реологические соотношения линейно-упругой среды) имеет следующий вид:

$$\sigma_{ij} = \lambda (\varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33}) \delta_{ij} + 2\mu \varepsilon_{ij}; \quad \sum_{k=1}^3 \sigma_{kk} = 3K \sum_{k=1}^3 \varepsilon_{kk}, \quad (16.36)$$

где λ, μ — постоянные Ламе, k — коэффициент всеобщего сжатия, ε_{ij} — компоненты тензора деформаций, σ_{ij} — компоненты тензора напряжений [4], рис. 16.3.

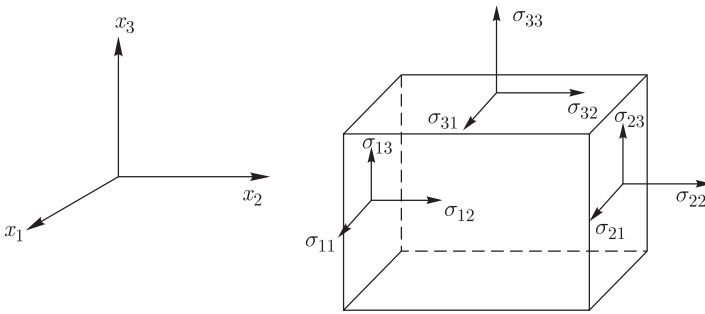


Рис. 16.3

Уравнение Дюгамеля–Неймана (учет температуры):

$$\sigma_{ij} = \left(\lambda \sum_{k=1}^3 \varepsilon_{kk} - \gamma \Delta T \right) \delta_{ij} + 2\mu \varepsilon_{ij}, \quad \varepsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right), \quad (16.37)$$

где u_i — перемещение по x_i ; $\gamma = \alpha_t \cdot 3K = \alpha_t (2\mu + 3\lambda)$; t, x_i — независимые переменные, α_t — коэффициент линейного расширения.

Двумерная система уравнений теории упругости может быть выписана в следующей форме:

$$\left\{ \begin{array}{l} \rho \left(\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} \right) = \frac{\partial \sigma_{11}}{\partial x} + \frac{\partial \sigma_{12}}{\partial y}, \\ \quad \text{уравнение движения;} \\ \rho \left(\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} \right) = \frac{\partial \sigma_{12}}{\partial x} + \frac{\partial \sigma_{22}}{\partial y}, \\ \quad \text{уравнение движения;} \\ \frac{\partial \sigma_{11}}{\partial t} + u \frac{\partial \sigma_{11}}{\partial x} + v \frac{\partial \sigma_{11}}{\partial y} = (\lambda + 2\mu) \frac{\partial u}{\partial x} + \lambda \frac{\partial v}{\partial y}, \\ \quad \text{закон Гука;} \end{array} \right. \quad (16.38a)$$

$$\left\{ \begin{array}{l} \frac{\partial \sigma_{22}}{\partial t} + u \frac{\partial \sigma_{22}}{\partial x} + v \frac{\partial \sigma_{22}}{\partial y} = (\lambda + 2\mu) \frac{\partial v}{\partial x} + \lambda \frac{\partial u}{\partial y}, \\ \frac{\partial \sigma_{33}}{\partial x} + u \frac{\partial \sigma_{33}}{\partial x} + v \frac{\partial \sigma_{33}}{\partial y} = \lambda \frac{\partial u}{\partial x} + \lambda \frac{\partial v}{\partial y}, \\ \frac{\partial \sigma_{12}}{\partial x} + u \frac{\partial \sigma_{12}}{\partial x} + v \frac{\partial \sigma_{12}}{\partial y} = 2\mu \frac{\partial u}{\partial x} + 2\mu \frac{\partial v}{\partial y}, \end{array} \right. \quad (16.386)$$

закон Гука;
закон Гука;
закон Гука;

$c_1 = \sqrt{\frac{\lambda + 2\mu}{\rho}}$; $c_2 = \sqrt{\frac{\mu}{\rho}}$ — скорости звука в линейно-упругой среде.

Эта система уравнений в частных производных может быть переписана в матричной форме:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{A}_1 \frac{\partial \mathbf{u}}{\partial x} + \mathbf{A}_2 \frac{\partial \mathbf{u}}{\partial y} = \mathbf{f}(t, x, y, \mathbf{u});$$

здесь собственные числа матриц \mathbf{A}_1 , \mathbf{A}_2 соответствуют продольным и поперечным скоростям звука c_1 и c_2 ; u, v — компоненты вектора скорости смещения, ρ — плотность среды; \mathbf{A}_1 , \mathbf{A}_2 — матрицы упругих коэффициентов, $\mathbf{u} = \{u, v, \sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{12}\}^T$ — вектор-столбец искомых функций, \mathbf{f} — вектор-столбец правых частей. Приведем также некоторые полезные соотношения теории упругости:

$$\sigma_{ij} = s_{ij} + \frac{1}{3} \left(\sum \sigma_{kk} \right) \delta_{ij}; \quad -p = \frac{1}{3} \sum \sigma_{kk},$$

$$\sum \sigma_{kk} = 3K \varepsilon_{kk}; \quad s_{ij} = 2\mu \varepsilon_{ij}, \quad \sum_{k=1}^3 s_{kk} = 0$$

(s_{ij} — девиатор тензора напряжений);

$$E = \frac{\mu(3\lambda + 2\mu)}{\lambda + \mu}; \quad \nu = \frac{\lambda}{2(\lambda + \mu)}; \quad K = \lambda + \frac{2}{3}\mu,$$

где E — модуль Юнга, ν — коэффициент Пуассона.

Соотношения Гука могут быть представлены в следующем виде через коэффициент Пуассона ν и модуль Юнга:

$$\sigma_{ij} = \frac{E}{1 + \nu} \left(\varepsilon_{ij} + \frac{\nu}{1 - 2\nu} \delta_{ij} \sum_{k=1}^3 \varepsilon_{kk} \right),$$

$$\varepsilon_{ij} = \frac{1}{E} \left((1 + \nu) \sigma_{ij} - \nu \delta_{ij} \sum_{k=1}^3 \sigma_{kk} \right).$$

Уравнение теории упругости также можно представить в дивергентной форме:

$$\left\{ \begin{array}{l} \frac{\partial \rho}{\partial t} + \frac{\partial (\rho u_j)}{\partial x_j} = 0, \quad \text{уравнение неразрывности;} \\ \frac{\partial (\rho u_i)}{\partial t} + \frac{\partial (\rho u_i u_j - \sigma_{ij})}{\partial x_j} = \rho F_i, \quad \text{уравнение движения;} \\ \frac{\partial [\rho (\varepsilon + |u|^2/2)]}{\partial t} + \frac{\partial [\rho u_j (\varepsilon + |u|^2/2) - u_k \sigma_{kj}]}{\partial x_j} = \\ = \frac{\partial}{\partial x_j} \left(x \frac{\partial T}{\partial x_j} \right) + Q + \rho u_k F_k, \quad \text{уравнение энергии,} \end{array} \right.$$

плюс реологические соотношения (закон Гука); используется соглашение о суммировании.

16.6. Нестационарная модель динамики морских и океанических течений

Система уравнений вязкой жидкости, описывающая поведение морских и океанических течений в переменных Эйлера, может быть представлена в следующем виде [8]:

$$\left\{ \begin{array}{l} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z} - l v + \frac{1}{\rho_0} \cdot \frac{\partial p}{\partial x} = \mu \cdot \Delta u + \frac{\partial}{\partial z} \nu \frac{\partial u}{\partial z}, \\ \frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + w \frac{\partial v}{\partial z} - l u + \frac{1}{\rho_0} \cdot \frac{\partial p}{\partial y} = \mu \cdot \Delta v + \frac{\partial}{\partial z} \nu \frac{\partial v}{\partial z}, \\ \frac{\partial u}{\partial z} = g \rho, \quad \text{уравнение движения;} \\ \frac{\partial u}{\partial t} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \quad \text{уравнение неразрывности;} \\ \frac{\partial T}{\partial x} + u \frac{\partial T}{\partial x} + v \frac{\partial T}{\partial y} + \gamma_T w = \mu_T \cdot \Delta T + \frac{\partial}{\partial z} \nu_T \frac{\partial T}{\partial z}, \\ \frac{\partial S}{\partial t} + u \frac{\partial S}{\partial x} + v \frac{\partial S}{\partial y} + \gamma_S w = \mu_S \cdot \Delta S + \frac{\partial}{\partial z} \nu_S \frac{\partial T}{\partial z}, \\ \rho = \alpha_T T + \alpha_S S, \quad \text{уравнение состояния морской воды,} \end{array} \right. \quad (16.39)$$

где $\gamma_T = \frac{\partial T}{\partial z}$, $\gamma_S = \frac{\partial S}{\partial z}$ — заданный параметр стратификации жидкости.

Уравнения записываются для отклонений давления p , плотности ρ , температуры T , солёности S от их средних значений по вертикали: $\{\bar{p}, \bar{\rho}, \bar{T}, \bar{S}\}(z)$.

Пусть область интегрирования представляет собой замкнутый бассейн, ограниченный невозмущенной поверхностью океана: $z = 0$, дном $H = H(x, y)$ и цилиндрической поверхностью σ .

Добавим к системе граничные и начальные условия:

$$\begin{cases} u = v = 0, \\ \frac{\partial T}{\partial v} = \frac{\partial S}{\partial v} = 0 \text{ на } \sigma \end{cases}$$

(\mathbf{v} — вектор внешней нормали к σ);

$$\begin{cases} x \frac{\partial u}{\partial z} = f_1; & x \frac{\partial u}{\partial z} = f_2; & w = 0; \\ a_1 \frac{\partial T}{\partial z} + b_1 T = c_1; & a_2 \frac{\partial S}{\partial z} + b_2 S = c_2 \text{ при } z = 0 \end{cases}$$

($x, f_1, f_2, a_1, b_1, a_2, b_2, c_1, c_2$ — заданные коэффициенты и функции);

$$\begin{cases} x \frac{\partial u}{\partial z} = 0; & x \frac{\partial v}{\partial z} = f; \\ w = u \frac{\partial H}{\partial x} + v \frac{\partial H}{\partial y}; \\ \frac{\partial T}{\partial v} + \frac{\partial S}{\partial v} = 0, \text{ при } z = H. \end{cases}$$

Эта система описывает процессы: перенос количества движения, тепла и слоев, а также их диффузию.

Начальные условия:

$$u = u^0, \quad v = v^0, \quad T = T^0, \quad S = S^0 \text{ при } t = 0.$$

16.7. Уравнения магнитной гидродинамики (МГД)

В этих задачах система уравнений газодинамики дополняется системой уравнений Максвелла [5–7]:

$$\begin{cases} \operatorname{rot} \mathbf{H} = \frac{4\pi}{c} \mathbf{i} + \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t}, \\ \operatorname{rot} \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{H}}{\partial t}, \\ \operatorname{div} \mathbf{E} = 4\pi \rho_e, \\ \operatorname{div} \mathbf{H} = 0. \end{cases}$$

Здесь \mathbf{E}, \mathbf{H} — векторные напряженности электрического и магнитного полей соответственно, \mathbf{i} — плотность электрических токов, ρ_e — плотность электрических зарядов, c — скорость света в пустоте.

Обычно полагают, что:

- диэлектрическая и магнитная проницаемости равны единице (приближенно);
- среда квазинейтральна (т.е. суммарный электрический заряд в любом объеме равен нулю: $\rho_e = 0$);
- членом $\frac{1}{c} \left(\frac{\partial \mathbf{E}}{\partial t} \right)$ — током смещения пренебрегают по сравнению с током проводимости \mathbf{i} ;
- кроме того, полагают, что проводимость среды достаточно велика, а рассматриваемые процессы протекают так медленно, что этим членом можно пренебречь.

$$\begin{cases} \mathbf{i} = \frac{c}{4\pi} \operatorname{rot} \mathbf{H}; \\ \frac{1}{c} \cdot \frac{\partial \mathbf{H}}{\partial t} = -\operatorname{rot} \mathbf{E}; \\ \operatorname{div} \mathbf{H} = 0; \\ \mathbf{i} = \sigma \mathbf{E} \text{ (замыкающее условие)}. \end{cases}$$

В уравнение движения в газодинамической системе войдут объемная электрическая сила, плотность которой равна

$$\mathbf{F} = \frac{1}{c} [\mathbf{i} \times \mathbf{H}],$$

а также объемные источники тепла Q , связанные с нагревом проводящей среды электрическими токами (джоулево тепло); мощность этих источников равна

$$Q = (\mathbf{i} \cdot \mathbf{E}),$$

или, с учетом закона Ома:

$$Q = \mathbf{i}^2 / \sigma = \sigma \mathbf{E}^2.$$

Таким образом, полная система уравнений МГД принимает следующий вид:

$$\begin{cases} \frac{\partial \rho}{\partial t} + \operatorname{div} \mathbf{v} = 0, \text{ уравнение неразрывности;} \\ \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \nabla) \mathbf{v} = -\frac{1}{\rho} \operatorname{grad} p + \mathbf{f} \quad \left(\mathbf{f} = \frac{1}{c} \cdot \frac{[\mathbf{i} \times \mathbf{H}]}{\rho}, \quad \mathbf{i} = \frac{c}{4\pi} \operatorname{rot} \mathbf{H} \right), \\ \frac{1}{c} \frac{\partial \mathbf{H}}{\partial t} = -\operatorname{rot} \mathbf{E}, \text{ закон Фарадея;} \end{cases} \quad \text{уравнение движения;}$$

$$\left\{ \begin{array}{l} \operatorname{div} \mathbf{H} = 0, \text{ закон Гаусса для магнитного поля;} \\ \mathbf{i} = \sigma \mathbf{E}; \\ \frac{\partial}{\partial t} \left(\varepsilon + \frac{\mathbf{v}^2}{2} \right) + (\mathbf{v} \nabla) \left(\varepsilon + \frac{\mathbf{v}^2}{2} \right) - \frac{1}{\rho} \operatorname{div} \rho \mathbf{v} + q + (\mathbf{f} \cdot \mathbf{v}) - \\ - \frac{1}{\rho} \operatorname{div} \mathbf{w} \quad \left(q = \frac{(\mathbf{i} \mathbf{E})}{\rho}, \quad \mathbf{w} = -\kappa \operatorname{grad} T \right), \text{ уравнение энергии;} \\ p = p(\rho, T), \quad \varepsilon = \varepsilon(\rho, T), \quad x = x(\rho, T), \quad \sigma = \sigma(\rho, T), \end{array} \right.$$

уравнения состояния;

здесь \mathbf{f}, q — соответственно плотность силы и мощность тепловых источников на единицу массы, κ — коэффициент теплопроводности.

Выпишем уравнения МГД для одномерного плоского случая. Пусть все компоненты вектора скорости и напряженности магнитного поля $\mathbf{v} = \{v, u, w\}$; $\mathbf{H} = \{H_x, H_y, H_z\}$ отличны от нуля и зависят только от t, x . Пусть продольная компонента магнитного поля и плотность электрического поля имеют лишь поперечные компоненты:

$$\mathbf{i} = \{0, i_y, i_z\}, \quad \mathbf{E} = \{0, E_y, E_z\},$$

$$i_y = -\frac{1}{4\pi} \cdot \frac{\partial H_z}{\partial x}, \quad i_z = \frac{1}{4\pi} \cdot \frac{\partial H_y}{\partial x}, \quad i_y = \sigma E_y, \quad i_z = \sigma E_z.$$

Компоненты вектора плотности электромагнитной силы

$$\mathbf{f} = \frac{1}{\rho} [\mathbf{i} \times \mathbf{H}]$$

вычисляются следующим образом:

$$f_n = \frac{1}{\rho} (i_y H_z - i_z H_y), \quad f_y = \frac{i_z H_{x0}}{\rho}, \quad f_z = \frac{i_y H_{x0}}{\rho},$$

или, учтя предыдущие соотношения для i_y, i_z , представим их в виде

$$f_x = -\frac{1}{\rho} \cdot \frac{\partial}{\partial x} \left(\frac{H_y^2 + H_z^2}{8\pi} \right), \quad f_y = \frac{H_{x0}}{4\pi\rho} \cdot \frac{\partial H_y}{\partial x}, \quad f_z = \frac{H_{x0}}{4\pi\rho} \cdot \frac{\partial H_z}{\partial x}.$$

Джоулево тепло в уравнении энергии есть

$$q = \frac{i_y E_y + i_z E_z}{\rho} = \frac{\sigma}{\rho} (E_y^2 + E_z^2) = \frac{i_y^2 + i_z^2}{\sigma\rho}.$$

Принимая во внимание эти соотношения, получим одномерную нестационарную систему уравнений МГД в переменных Эйлера:

$$\left\{ \begin{array}{l} \frac{\partial \rho}{\partial t} + v \frac{\partial \rho}{\partial x} + \rho \frac{\partial v}{\partial x} = 0, \\ \frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} = -\frac{1}{\rho} \cdot \frac{\partial v}{\partial x} + f_x, \quad \left\{ f_x = -\frac{1}{\rho} \cdot \frac{\partial}{\partial x} \left(\frac{H_y^2 + H_z^2}{8\pi} \right) \right\}, \end{array} \right.$$

$$\left\{ \begin{array}{l} \frac{\partial u}{\partial t} + v \frac{\partial u}{\partial x} = f_y, \quad \left\{ f_y = \frac{H_{x0}}{4\pi\rho} \cdot \frac{\partial H_y}{\partial x} \right\}, \\ \frac{\partial w}{\partial t} + v \frac{\partial w}{\partial x} = f_z, \quad \left\{ f_z = \frac{H_{x0}}{4\pi\rho} \cdot \frac{\partial H_z}{\partial x} \right\}, \\ \frac{\partial v}{\partial t} \left(\frac{H_y}{\rho} \right) + v \frac{\partial}{\partial x} \left(\frac{H_y}{\rho} \right) = \frac{H_{x0}}{\rho} \cdot \frac{\partial u}{\partial x} + \frac{1}{\rho} \cdot \frac{\partial E_z}{\partial x}, \\ \frac{\partial v}{\partial t} \left(\frac{H_z}{\rho} \right) + v \frac{\partial}{\partial x} \left(\frac{H_z}{\rho} \right) = \frac{H_{x0}}{\rho} \cdot \frac{\partial w}{\partial x} - \frac{1}{\rho} \cdot \frac{\partial E_y}{\partial x}, \\ \left\{ i_y = \sigma E_y = -\frac{1}{4\pi} \cdot \frac{\partial H_z}{\partial x}; \quad i_z = \sigma E_z = -\frac{1}{4\pi} \cdot \frac{\partial H_y}{\partial x} \right\}, \\ \frac{\partial}{\partial t} \left(\varepsilon + \frac{v^2 + u^2 + w^2}{2} \right) + v \frac{\partial}{\partial x} \left(\varepsilon + \frac{v^2 + u^2 + w^2}{2} \right) = \\ = -\frac{1}{\rho} \cdot \frac{\partial}{\partial x} (pv) + q + f_x v + f_y u + f_z w, \\ \left\{ q = \frac{1}{\rho} (i_y E_y + i_z E_z), \quad p = p(\rho, T), \quad \varepsilon = \varepsilon(\rho, T), \right. \\ \left. \sigma = \sigma(\rho, T) \right\}, \end{array} \right.$$

где $H_x = H_{x0}$.

Если записать эту систему в матричной форме

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{A} \frac{\partial \mathbf{v}}{\partial x} = 0$$

и решить характеристическое уравнение

$$\det(\mathbf{A} - \lambda \mathbf{E}) = 0,$$

то получим собственные числа:

$$\lambda_1 = 0; \quad \lambda_{2,3} = \pm \frac{H_{x0}}{\sqrt{4\pi\rho}} \rho = \pm a_A$$

$$\lambda_{4,5} = \pm \rho \sqrt{\frac{1}{2} \left[c^2 + \frac{\mathbf{H}^2}{4\pi\rho} + \sqrt{\left(c^2 + \frac{\mathbf{H}^2}{4\pi\rho} \right)^2 - 4c^2 \frac{\mathbf{H}_{x0}^2}{4\pi\rho}} \right]} = \pm a_+$$

$$\lambda_{6,7} = \pm \rho \sqrt{\frac{1}{2} \left[c^2 + \frac{\mathbf{H}^2}{4\pi\rho} - \sqrt{\left(c^2 + \frac{\mathbf{H}^2}{4\pi\rho} \right)^2 - 4c^2 \frac{\mathbf{H}_{x0}^2}{4\pi\rho}} \right]} = \pm a_-.$$

Величина a_A называется *альфвеновской скоростью*, а малые возмущения, распространяемые с этой скоростью — *альфвеновскими волнами*: эти волны являются поперечными; частицы газа (плазма) смещаются в направлении, перпендикулярном к направлению рассматриваемой волны (звуковые же волны в газодинамике — продольные). Величины a_+ и a_- называются

соответственно *быстрой* и *медленной магнитными скоростями звука*, а соответствующие им малые возмущения — *быстрой* и *медленной магнитозвуковыми волнами*. Показывается, что $a_- \leq a_A \leq a_+$.

Если компоненты поля в направлении распространения волны равны нулю, то $c_- = c_A = 0$ и в среде существуют только быстрые магнитозвуковые волны, скорость которых равна $c_+ = \sqrt{c^2 + H^2 / (4\pi\rho)}$.

16.8. Система уравнений Прандтля ламинарного пограничного слоя в несжимаемой жидкости

Рассматривается плоскопараллельное течение жидкости (рис. 16.4), которое описывается системой уравнений Навье–Стокса [3].

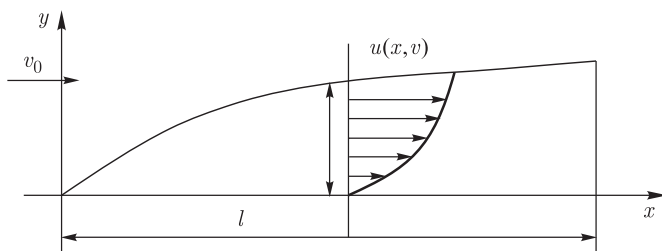


Рис. 16.4

Из эксперимента известно, что при больших значениях числа Рейнольдса $Re = v_0 l / \nu$ существенное влияние на движение жидкости оказывает тонкий пограничный слой. Течение разбивается на две части: идеальная жидкость и тонкий вязкий погранслои вблизи поверхности тела. Удастся упростить систему уравнений Навье–Стокса для плоско-параллельного течения:

$$\frac{\partial u}{\partial z} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} = -\frac{1}{\rho} \frac{\partial p}{\partial x} + \nu \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right), \quad \text{уравнение движения} \quad (16.40a)$$

$$\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} = -\frac{1}{\rho} \frac{\partial p}{\partial y} + \nu \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right), \quad \text{уравнение движения} \quad (16.40б)$$

$$\frac{\partial u}{\partial x} + v \frac{\partial v}{\partial y} = 0, \quad \text{уравнение движения.} \quad (16.40в)$$

Величина δ/l — основная малая величина; в (16.40а)–(16.40в) оставляем члены $\sim O(1)$; $\delta \ll O(1)$; $y \sim \delta$, l — длина пластины.

После проведения преобразований (замена переменных: $x = l\xi$; $y = \delta\eta$; ξ, η изменяются также в конечных пределах) получим:

$$\begin{cases} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} = -\frac{1}{\rho} \frac{\partial p}{\partial y} + v \frac{\partial^2 u}{\partial y^2}, \\ \frac{\partial p}{\partial y} = 0, \\ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0. \end{cases} \quad (16.41)$$

На поверхности обтекаемого тела ставится условие прилипания $u = v = 0$; на границе погранслоя $u = v_0(x, t)$. Отметим, что многие задачи решаются без учета погранслоя, что не всегда оправдано.

16.9. Система уравнений теории мелкой воды

Система двумерных нестационарных уравнений теории мелкой воды (глубина водоема много меньше его характерных размеров) с учетом гравитационных сил имеет вид [3]:

$$\begin{cases} \frac{dh}{dt} + u \frac{\partial(hu)}{\partial x} + \frac{\partial(hv)}{\partial y} = 0, \\ \frac{d(hu)}{dt} + \frac{\partial(hu^2 + gh^2/2)}{\partial x} + \frac{\partial(huv)}{\partial y} = -gh \frac{\partial b}{\partial x}, \\ \frac{\partial(hu)}{\partial t} + \frac{\partial(huv)}{\partial x} + \frac{\partial(hv^2 + gh^2/2)}{\partial y} = -gh \frac{\partial b}{\partial y}, \end{cases} \quad (16.42)$$

$b(x, y)$ — рельеф дна, h — глубина жидкости, $\zeta = h + b$ — уровень жидкости (рис. 16.5), g — ускорение свободного падения; u, v — скорости среды.

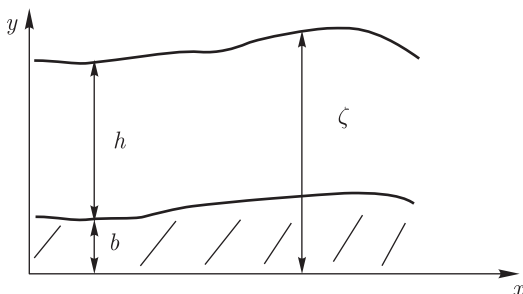


Рис. 16.5

Эта система получается из уравнений Эйлера путем усреднения параметров жидкости по вертикальной координате и используется для расчета параметров вида в мелких водоемах.

16.10. Система уравнений акустики

1. Одномерная акустическая система уравнений путем линеаризации уравнений Эйлера представляется в виде [9]:

$$\left\{ \begin{array}{l} \frac{\partial u}{\partial t} + \frac{1}{\rho_0} \cdot \frac{\partial p}{\partial x} = 0, \end{array} \right. \quad (16.43a)$$

$$\left\{ \begin{array}{l} \frac{\partial p}{\partial t} + \rho_0 c_0^2 \cdot \frac{\partial u}{\partial x} = 0, \end{array} \right. \quad (16.43б)$$

u — скорость среды, p — давление (точнее их малые отклонения от значений в невозмущенной среде), ρ_0 — плотность, c_0 — скорость звука.

После интегрирования (16.43a), (16.43б) по произвольной области с границей Γ на плоскости $\{t, x\}$, переходя к контурным интегралам, получим

$$\left\{ \begin{array}{l} \oint_{\Gamma} \rho_0 u \, dx - p \, dt = 0, \\ \oint_{\Gamma} \frac{p}{c_0^2} \, dx - \rho_0 u \, dt = 0. \end{array} \right. \quad (16.44)$$

Умножая (16.43б) на $(\rho_0 c_0)^{-1}$ и складывая с (16.43a), а затем вычитая из него, получим систему акустики в инвариантах Римана:

$$\left\{ \begin{array}{l} \frac{\partial Y}{\partial t} + c_0 \frac{\partial Y}{\partial x} = 0, \\ \frac{\partial Z}{\partial t} - c_0 \frac{\partial Z}{\partial x} = 0, \end{array} \right. \quad (16.45)$$

где $Y = u + \frac{p}{\rho_0 c_0}$, $Z = u - \frac{p}{\rho_0 c_0}$ — инварианты Римана.

Общее решение имеет вид [9]:

$$Y = f(x - c_0 t), \quad Z = g(x + c_0 t), \quad (16.46)$$

или, с учетом инвариантов Римана:

$$\begin{aligned} u &= \frac{1}{2} [f(x - c_0 t) + g(x + c_0 t)]; \\ p &= \frac{\rho_0 c_0}{2} [f(x - c_0 t) - g(x + c_0 t)], \end{aligned} \quad (16.47)$$

где f, g — функции, определяемые из краевых условий.

Прямые

$$x \pm c_0 t = \text{const} \quad \left(\frac{dx}{dt} = \pm c_0 \right) \quad (16.48)$$

являются характеристиками акустической системы, по которым распространяются звуковые волны.

2. Двумерная система уравнений акустики имеет следующий вид:

$$\begin{cases} \frac{\partial u}{\partial t} + \frac{1}{\rho_0} \cdot \frac{\partial p}{\partial x} = 0, \\ \frac{\partial v}{\partial t} + \frac{1}{\rho_0} \cdot \frac{\partial p}{\partial y} = 0, \\ \frac{\partial p}{\partial t} + \rho_0 c_0^2 \cdot \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = 0. \end{cases} \quad (16.49)$$

Интегральный вид акустической системы:

$$\begin{cases} \oint_S \rho u \, dx \, dy + p \, dy \, dt = 0, \\ \oint_S \rho_0 v \, dx \, dy + p \, dx \, dt = 0, \\ \oint_S \rho u \, dx \, dy + \rho_0 c_0^2 (u \, dy \, dt + v \, dx \, dt) = 0. \end{cases} \quad (16.50)$$

Интегралы берутся по любой замкнутой поверхности в пространстве $\{t, x, y\}$. (16.48) — система законов сохранения, (16.47) — ее следствие.

16.11. Введение в разностные схемы газодинамики

1. Рассмотрим одномерную систему газодинамических уравнений в лагранжевых координатах [5]:

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = 0, \quad (16.51a)$$

$$\frac{\partial v}{\partial t} - \frac{\partial u}{\partial x} = 0, \quad (16.51b)$$

$$\frac{\partial \varepsilon}{\partial t} + \bar{p} \frac{\partial v}{\partial t} = 0. \quad (16.51b)$$

Дивергентный вид уравнения (16.51b):

$$\frac{\partial}{\partial t} \left(\varepsilon + \frac{u^2}{2} \right) + \frac{\partial (\bar{p} u)}{\partial x} = 0 \quad (16.51r)$$

($v = \rho^{-1}$; $p = p(\rho, \varepsilon)$, идеальный газ $p = (\gamma - 1)\varepsilon\rho$; $c = \sqrt{\gamma p/\rho}$, $\bar{p} = p + \omega$, ω — искусственная вязкость или «псевдовязкость»):

$$\omega = -\mu\rho \frac{\partial u}{\partial x} = \gamma \frac{\partial}{\partial x}; \quad \gamma = (\mu\rho). \quad (16.52)$$

В качестве ω можно выбрать величину $\mu = -\mu_0 ch$ («линейная вязкость»), недостатком которой является то, что она действует по всему течению, так что сильное сглаживание ударной волны, соответствующее большому μ_0 , сильно сглаживает решение. По этой причине Дж. Нейман и Р. Рихтмайер в 1950 г. предложили нелинейную вязкость [2]:

$$\mu = -\mu_0 h^2 \rho \frac{\partial u}{\partial x}. \quad (16.53)$$

Можно показать, что ширина ударного перехода для μ равна

$$l = \pi \sqrt{\frac{2\mu_0}{\gamma + 1}} h.$$

Одной из первых разностных схем с псевдовязкостью для численного решения (16.51а)–(16.51в) была схема «крест» Неймана–Рихтмайера (шахматная схема; шаблон представлен на рис. 16.6):

$$\left\{ \begin{array}{l} \frac{u_m^{n+1/2} - u_m^{n-1/2}}{\tau} + \frac{\bar{p}_{m+1/2}^n - \bar{p}_{m-1/2}^n}{h} = 0, \end{array} \right. \quad (16.54a)$$

$$\left\{ \begin{array}{l} \frac{v_{m+1/2}^{n+1} - v_{m+1/2}^n}{\tau} + \frac{u_{m+1}^{n+1/2} - u_m^{n+1/2}}{h} = 0, \end{array} \right. \quad (16.54б)$$

$$\left\{ \begin{array}{l} \frac{\varepsilon_{m+1/2}^{n+1} - \varepsilon_{m+1/2}^n}{\tau} + \frac{\bar{p}_{m+1/2}^n + \bar{p}_{m-1/2}^n}{2} \times \\ \times \frac{v_{m+1/2}^{n+1} - v_{m+1/2}^n}{\tau} = 0, \end{array} \right. \quad (16.54в)$$

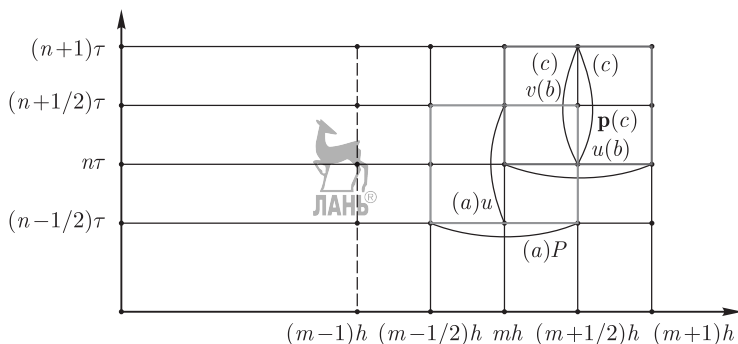


Рис. 16.6

где

$$\begin{cases} \bar{p}_{m+1/2}^n = p_{m+1/2}^n + \omega_{m+1/2}^n, \\ \omega_{m+1/2}^n = -\mu_0 h^2 p_{m+1/2}^n \cdot \frac{|u_{m+1}^{n-1/2} - u_m^{n-1/2}|}{h} \cdot \frac{u_{m+1}^{n-1/2} - u_m^{n-1/2}}{h}, \\ \varepsilon_{m+1/2}^n = \varepsilon(p_{m+1/2}^n, v_{m+1/2}^n) \cdot \frac{x_m^{n+1} - x_m^n}{\tau} = u_m^{n+1}. \end{cases} \quad (16.55)$$

(например, $\varepsilon = [1/(\gamma - 1)] \cdot p/\rho$; для адиабатических течений: $p/\rho^\gamma = \text{const}$).

Чтобы избавиться от полуцелых временных индексов, можно сделать следующую замену переменных:

$$\begin{cases} \frac{u_m^{n+1} - u_m^n}{\tau} + \frac{\bar{p}_{m+1/2}^n - \bar{p}_{m-1/2}^n}{h} = 0, \\ \frac{v_{m+1/2}^{n+1} - v_{m+1/2}^n}{\tau} - \frac{u_{m+1}^{n+1} - u_m^{n+1}}{h} = 0, \\ \omega_{m+1/2}^n = -\mu_0 h^2 \rho_{m+1/2}^n \frac{|u_{m+1}^n - u_m^n|}{h} \cdot \frac{u_{m+1}^n - u_m^n}{h}. \end{cases} \quad (16.56)$$

В 1955 г. Р. Лэттер предложил следующую модификацию метода «псевдовязкости» Неймана–Рихтмайера. Поскольку в газах существуют только ударные сжатия, а волны разрежения отсутствуют (теорема Цемплена), то в расчетах целесообразно исключать «псевдовязкость» в волнах разрежения, т. е. занулять коэффициент «псевдовязкости».

Тогда:

$$\omega = \begin{cases} 0, & \Delta u \geq 0, \\ -\mu_0 h^2 \rho \left| \frac{\Delta u}{\Delta x} \right| \cdot \frac{\Delta u}{\Delta x} \approx \mu_0 \rho (\Delta u)^2, & \Delta u < 0, \end{cases} \quad (16.57)$$

Так как в плоском случае в волнах сжатия и ударных волнах выполняется

$$\frac{\Delta u}{\Delta x} < 0, \quad (16.58)$$

то для волн разрежения

$$\frac{\Delta u}{\Delta x} \geq 0.$$

То же делается и для линейного коэффициента вязкости:

$$\omega = -\mu_0 c h \frac{\partial u}{\partial x},$$

где c — скорость звука.

Схема имеет порядок аппроксимации $O(\tau^2 + h^2)$.

Другим способом построения устойчивой схемы является аппроксимация производных вдоль характеристик (сеточно-характеристический метод).

Выпишем систему уравнений газодинамики в матричной форме:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{A} \frac{\partial \mathbf{u}}{\partial x} + \mathbf{f} = 0; \quad (16.59)$$

$$\mathbf{u} = \begin{Bmatrix} \rho \\ v \\ e \end{Bmatrix}; \quad \mathbf{A} = \begin{Bmatrix} v & \rho & 0 \\ p'_\rho & v & p'_e \\ 0 & p/\rho & v \end{Bmatrix}, \quad \mathbf{f} = \begin{Bmatrix} f_1 \\ f_2 \\ f_3 \end{Bmatrix}.$$

Умножим (16.59) на левый собственный вектор ω_i матрицы \mathbf{A} с учетом соотношений

$$\omega_i \mathbf{A} = \lambda_i \omega_i, \quad (16.60)$$

$$\omega_i u_t + \lambda_i \omega_i u_x + \omega_i \mathbf{f} = 0, \quad i = 1, \dots, I$$

(λ_i — i -е собственное значение матрицы \mathbf{A}) и аппроксимируем (16.60):

$$\omega_i \frac{\mathbf{u}_m^{n+1} - \mathbf{u}_m^n}{\tau} \mp (\lambda_i \omega_i)_m^n \frac{\mathbf{u}_{m \mp 1}^n - \mathbf{u}_m^n}{h} + (\omega_i)_m^n \mathbf{f}_m^n = 0, \quad (16.61)$$

где верхний знак берется при $\lambda_i > 0$, нижний при $\lambda_i < 0$ (случай $\lambda_i = 0$ является вырожденным); шаблон представлен на рис. 16.7.

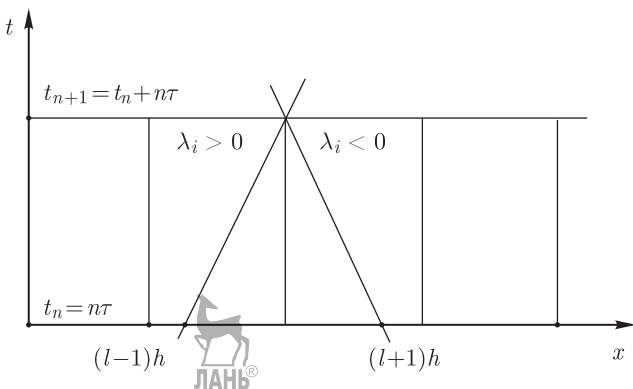


Рис. 16.7

В матричном виде система (16.61) может быть записана в виде

$$\Omega (\mathbf{u}_m^{n+1} - \mathbf{u}_m^n) - \sigma \mathbf{\Lambda}^+ \Omega (\mathbf{u}_{m-1}^n - \mathbf{u}_m^n) + \sigma \mathbf{\Lambda}^- \Omega (\mathbf{u}_{m+1}^n - \mathbf{u}_m^n) + \tau \mathbf{f} = 0, \quad \sigma = \frac{\tau}{h}, \quad (16.62)$$

где

$$\mathbf{\Lambda} = \begin{Bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{Bmatrix}$$

— диагональная матрица из собственных чисел матрицы \mathbf{A} ; $\mathbf{\Lambda}^+$, $\mathbf{\Lambda}^-$ — диагональные матрицы из положительных собственных чисел матрицы \mathbf{A} . $\mathbf{\Omega}$ — матрица, строками которой являются левые собственные векторы матрицы \mathbf{A} . Из (16.62) получаем вид разностной схемы первого порядка аппроксимации (сеточно-характеристический метод) [10, 11]:

$$\mathbf{u}_m^{n+1} = \mathbf{u}_m^n - \sigma [(\mathbf{\Omega}^{-1}\mathbf{\Lambda}^+\mathbf{\Omega})(\mathbf{u}_{m-1}^n - \mathbf{u}_m^n) - (\mathbf{\Omega}^{-1}\mathbf{\Lambda}^-\mathbf{\Omega})(\mathbf{u}_{m+1}^n - \mathbf{u}_m^n)] + \tau \mathbf{f} = 0. \quad (16.63)$$

Для системы уравнений газодинамики (16.59) получим:

$$\begin{aligned} \lambda_1 &= v + c, & \lambda_2 &= v, & \lambda_3 &= v - c, \\ c^2 &= \gamma(\gamma - 1)e, & p &= (\gamma - 1)\rho e, \\ \mathbf{\Omega} &= \begin{Bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{Bmatrix} = \begin{Bmatrix} p'_\rho & \rho c & \rho'_e \\ p & 0 & -\rho^2 \\ p & -\rho c & \rho'_e \end{Bmatrix}. \end{aligned} \quad (16.64)$$

2. Аналог системы (16.60) для одномерного линейного скалярного уравнения переноса — разностная схема Куранта–Изаксона–Риса.

Запишем в матричной форме двумерную систему динамической теории упругости:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{A}_1 \frac{\partial \mathbf{u}}{\partial x_1} + \mathbf{A}_2 \frac{\partial \mathbf{u}}{\partial x_2} = 0, \quad (16.65)$$

где

$$\mathbf{A}_1 = \begin{Bmatrix} 0 & 0 & -\rho^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\rho^{-1} & 0 & 0 \\ -(\lambda + 2\mu) & 0 & 0 & 0 & 0 & 0 \\ 0 & -\mu & 0 & 0 & 0 & 0 \\ -\lambda & 0 & 0 & 0 & 0 & 0 \\ -\lambda & 0 & 0 & 0 & 0 & 0 \end{Bmatrix},$$

$$\mathbf{A}_2 = \begin{Bmatrix} 0 & 0 & 0 & -\rho^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\rho^{-1} & 0 \\ 0 & -\lambda & 0 & 0 & 0 & 0 \\ -\mu & 0 & 0 & 0 & 0 & 0 \\ 0 & -(\lambda + 2\mu) & 0 & 0 & 0 & 0 \\ 0 & -\lambda & 0 & 0 & 0 & 0 \end{Bmatrix},$$

$$\mathbf{u} = \{u, v, \sigma_{11}, \sigma_{12}, \sigma_{22}, \sigma_{33}\};$$

уравнение состояния является следствием уравнения неразрывности и суммирования уравнений для $\sigma_{ii}(\varepsilon_{kl})$ по $i = 1, 2, 3$; оно вытекает из закона Гука:

$$\sigma_{ij} = \lambda \left(\sum \varepsilon_{ii} \right) \delta_{ij} + 2\mu \varepsilon_{ij}. \quad (16.66)$$

Разностная аппроксимирующая (16.65) система уравнений имеет вид:

$$\begin{aligned} \mathbf{u}_m^{n+1} = & \mathbf{u}_m^n + \sigma_1 \left[\left(\boldsymbol{\Omega}_1^{-1} \boldsymbol{\Lambda}_1^+ \boldsymbol{\Omega}_1 \right)_{ml}^n (\mathbf{u}_{m-1,l}^n - \mathbf{u}_{ml}^n) - \right. \\ & \left. - \left(\boldsymbol{\Omega}_1^{-1} \boldsymbol{\Lambda}_1^- \boldsymbol{\Omega}_1 \right)_{ml}^n (\mathbf{u}_{m+1,l}^n - \mathbf{u}_{ml}^n) \right] + \sigma_2 \left[\left(\boldsymbol{\Omega}_2^{-1} \boldsymbol{\Lambda}_2^+ \boldsymbol{\Omega}_2 \right)_{ml}^n \times \right. \\ & \left. \times (\mathbf{u}_{m,l-1}^n - \mathbf{u}_{ml}^n) - \left(\boldsymbol{\Omega}_2^{-1} \boldsymbol{\Lambda}_2^- \boldsymbol{\Omega}_2 \right)_{ml}^n (\mathbf{u}_{m,l+1}^n - \mathbf{u}_{ml}^n) \right], \end{aligned} \quad (16.67)$$

$$\boldsymbol{\Lambda}_k^+ = \frac{1}{2} (\boldsymbol{\Lambda}_k^+ |\boldsymbol{\Lambda}_k|), \quad \boldsymbol{\Lambda}_k^- = \frac{1}{2} (\boldsymbol{\Lambda}_k^- |\boldsymbol{\Lambda}_k|), \quad \det (\boldsymbol{\Omega}_k - \lambda^k \mathbf{E}) = 0,$$

Заметим, что в системе (16.65) также можно провести аппроксимацию координатных производных на верхнем временном слое и тем самым реализовать неявную схему. Однако использование неявных схем приводит к расширению области зависимости решения, что чревато увеличением ошибки аппроксимации.

Рассмотрим другие способы получения неявных разностных схем.

3. Система уравнений газовой динамики для одномерного нестационарного случая может быть представлена в следующем виде (в массовых переменных Лагранжа):

$$\begin{cases} \oint (\rho^{-1} dx + v dt) = 0, & (16.68a) \end{cases}$$

$$\begin{cases} \oint (v dx - p dt) = 0, & (16.68б) \end{cases}$$

$$\begin{cases} \oint \left(\varepsilon + \frac{v^2}{2} \right) dx - (pv) dt = 0. & (16.68в) \end{cases}$$

Получим аппроксимацию (16.68в) с помощью интегро-интерполяционного метода для контура, представленного на рис. 16.8:

$$\int_{x_n}^{x_{m+1}} \varepsilon(t_n, x) dx = \varepsilon_{m+1}^n h, \quad (16.69a)$$

$$\frac{1}{2} \int_{x_n}^{x_{m+1}} v^2(t, x) dx = \frac{1}{4} \left[(v_{m+1}^n)^2 + (v_m^n)^2 \right] h, \quad (16.69б)$$

$$\int_{t_n}^{t_{m+1}} p(t, x_n) \cdot v(t, x_n) dt = \frac{1}{2} \left(p_{m+1/2}^{n+1} + p_{m-1/2}^{n+1} \right) v_m^{n+1} \tau. \quad (16.69\text{в})$$

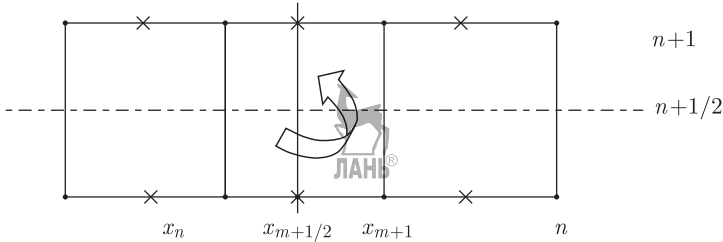


Рис. 16.8

Тогда получаем следующую аппроксимацию (16.69в):

$$\begin{aligned} \frac{\varepsilon_{m+1/2}^{n+1} - \varepsilon_{m+1/2}^n}{\tau} + \frac{1}{4} \frac{(v_{m+1}^{n+1})^2 + (v_m^{n+1})^2 - (v_{m+1}^n)^2 - (v_m^n)^2}{\tau} = \\ = -\frac{1}{2h} \left[\left(p_{m+3/2}^{n+1} + p_{m+1/2}^{n+1} \right) v_{m+1}^{n+1} - \right. \\ \left. - \left(p_{m+1/2}^{n+1} + p_{m-1/2}^{n+1} \right) v_m^{n+1} \right]. \quad (16.70) \end{aligned}$$

Аналогично аппроксимируются и уравнения (16.69а), (16.69б).

16.12. Уравнение бесстолкновительной плазмы (уравнение Власова)

Уравнение Власова не относится к уравнениям МСС; оно описывает движение совокупности большого числа заряженных частиц (ионов или электронов) в условиях, когда можно пренебречь столкновениями частиц и их взаимодействием определяется только электрическими силами; это уравнение описывает события, масштаб которых меньше длины свободного пробега и характерное время много меньше времени свободного пробега. Обычно это процессы, происходящие в сильно разреженной плазме [2].

В первой модели состояние плазмы описывается двумя функциями: $f_e(t, r, v)$, $f_i(t, r, v)$, где $r = \{x, y, z\}$, $v = \{v_x, v_y, v_z\}$ — декартовы координаты точки пространства и трехмерные координаты точки в импульсном пространстве, f_e, f_i — функции распределения электронов и ионов. Значит, если мы выделяем в пространстве маленький кубик $[r, r + \Delta r]$ и интересуемся числом частиц в нем, имеющих скорости в диапазоне $[v, v + \Delta v]$, то

оно выражается величиной $f \cdot \Delta r \cdot \Delta v$. Область фазового пространства $\{r, v\}$ обычно не ограничена по скорости, но f быстро убывает при $|v| \rightarrow \infty$, поэтому можно ограничиться конечной областью $|v| \leq V$, поставив граничное условие $f|_{v=V} = 0$. Будем полагать, что по пространственным переменным все функции периодичны с периодом L (в такой постановке решается большинство задач физики плазмы). Двумерная постановка задач имеет следующий вид:

$$\begin{cases} \frac{\partial f_e}{\partial t} + v_x \frac{\partial f_e}{\partial x} + v_y \frac{\partial f_e}{\partial y} - \frac{q_e}{m_e} \left(\frac{\partial v}{\partial x} \cdot \frac{\partial f_e}{\partial v_x} + \frac{\partial \varphi}{\partial y} \cdot \frac{\partial f_e}{\partial v_y} \right) = 0, \\ \frac{\partial f_i}{\partial t} + v_x \frac{\partial f_i}{\partial x} + v_y \frac{\partial f_i}{\partial y} - \frac{q_i}{m_i} \left(\frac{\partial v}{\partial x} \cdot \frac{\partial f_i}{\partial v_x} + \frac{\partial \varphi}{\partial y} \cdot \frac{\partial f_i}{\partial v_y} \right) = 0. \end{cases} \quad (16.71)$$

Здесь $\varphi(x, y, t)$ — потенциал электронного поля, компоненты $(-E_x, -E_y)$ напряженности электронного поля φ определяются из уравнения Пуассона

$$\Delta \varphi = -4\pi \left(q_e \int_{-\infty}^{\infty} f_e(x, y, v, t) dv + q_i \int_{-\infty}^{\infty} f_i(x, y, v, t) dv \right), \quad (16.72)$$

q_e, q_i — заряды электрона и иона, m_e, m_i — их массы; (16.72) называют *уравнением самосогласованного электронного поля* (в том смысле, что оно не задается расположением каких-то внешних зарядов, а создается участвующими в процессе частицами).

Часто используется идеализированная модель, в которой ионы рассматриваются как нейтрализующий фон с известной плотностью заряда ρ_0 ; рассматривается же система одномерных уравнений

$$\begin{cases} \frac{\partial f}{\partial t} + v \frac{\partial f}{\partial x} + \frac{F}{m} \cdot \frac{\partial f}{\partial v} = 0; \\ \frac{d^2 \varphi}{dx^2} = 4\pi \rho; \quad E = -\frac{d\varphi}{dx}; \quad F = gE; \\ \rho(x, t) = |q| \left(n_0 - \int_{-\infty}^{\infty} f dv \right), \end{cases} \quad (16.73)$$

$\varphi(x_0) = \varphi(x_0 + L)$, либо $\varphi = 0$ в случае заземленных концов.

Хотя система уравнений в частных производных (16.73) есть существенное упрощение полной системы, тем не менее ее решение представляет значительные трудности. Применяя к (16.73) процесс линеаризации, преобразование Фурье по пространству и Лапласа по времени, получают, что характерными частотами

и длинами волн колебаний электростатической плазмы являются плазменная частота

$$\omega_p = \sqrt{\frac{nq^2}{\varepsilon_0 m_e}} \quad (16.74)$$

и дебаевская длина

$$\lambda_D = \sqrt{\frac{\varepsilon_0 k_B T}{nq^2}}. \quad (16.75)$$

Отметим, что существует альтернативная математическая модель, в которой рассматриваются все частицы. Движение каждой из них описывается уравнениями:

$$\frac{dr_k}{dt} = v_k, \quad \frac{dv_k}{dt} = \frac{q_k}{m_k} E(x_k, y_k, t), \quad k = 1, 2, \dots, K, \quad (16.76)$$

где k — номер частицы, q_k , m_k — ее заряд и масса E — напряженность электронного поля:

$$\begin{aligned} E &= -\operatorname{grad} \varphi; \\ \Delta \varphi &= 4\pi \rho(x, y, t), \\ \varphi(r, t) &= \sum_k \frac{q_k}{|r - r_k(t)|}, \end{aligned} \quad (16.77)$$

где q_k — потенциал, создаваемый k -м зарядом, находящимся в точке $r_k(t)$.

Ее основной недостаток — высокий порядок системы обыкновенных дифференциальных уравнений ($K \gg 1$), достоинство — простота их интегрирования.

Список литературы

1. Ландау Л. Д., Лифшиц Е. М. Теоретическая физика: Учеб. пособие. В 10 т. Т. VI. Гидродинамика. 6-е изд. М.: ФИЗМАТЛИТ, 2017. 727 с.
2. Федоренко Р. П. Введение в вычислительную физику. Долгопрудный: Интеллект, 2008. 503 с.
3. Флетчер К. Вычислительные методы в динамике жидкостей: В 2 т. Т. 2. М.: Мир, 1991. 552 с.
4. Новацкий В. Теория упругости. М.: Мир, 1975. 872 с.
5. Самарский А. А., Попов Ю. П. Разностные методы решения задач газовой динамики. М.: Наука, Физматлит. 1992. 423 с.
6. Ландау Л. Д., Лифшиц Е. М. Теоретическая физика: Учеб. пособие. В 10 т. Т. VIII. Электродинамика сплошных сред. 5-е изд. М.: ФИЗМАТЛИТ, 2019. 651 с.

7. Куликовский А. Г., Погорелов Н. В., Семёнов А. Ю. Математические вопросы численного решения гиперболических систем уравнений. М.: ФИЗМАТЛИТ, 2012. 656 с.
8. Марчук Г. И. Математическое моделирование в проблеме окружающей среды. М.: Наука, 1982. 320 с.
9. Рождественский Б. Л., Яненко Н. Н. Системы квазилинейных уравнений. М.: Наука, 1978. 687 с.
10. Магомедов К. М., Холодов А. С. Сеточно-характеристические численные методы. М.: Юрайт, 2018. 313 с.
11. Innovations in Wave Processes Modelling and Decision Making. Grid-Characteristic Method and Applications / Favorskaya, A. V., Petrov, I. B. (Eds.). Switzerland: Springer, 2018. 270 p.





Приложение 1

ТЕОРЕТИЧЕСКИЕ ВОПРОСЫ К КУРСУ ЛЕКЦИЙ ПО ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКЕ (ТЕОРЕТИЧЕСКИЙ МИНИМУМ) ¹⁾

К главе 1

1. Отличие вычислительной математики от классических математических курсов. Понятия обусловленности задачи, устойчивости, алгоритма, погрешности вычислений.

2. Определите погрешность приближенного вычисления производной u'_t по формулам:

$$\frac{u(x+h) - u(x)}{h}, \quad \frac{u(x+h) - u(x-h)}{2h}.$$

3. Оцените оптимальный шаг численного дифференцирования, учитывающий погрешность метода и ошибку округления для приближенной формулы

$$u'_x \approx \frac{u(x+h) - u(x)}{h}.$$

4. Дайте определения абсолютной и относительной погрешностей приближения.

5. Определение предельной абсолютной погрешности.

6. Как оценить погрешность приближения некоторой величины с помощью ее производных по параметрам, от которых она зависит?

7. Приведите формулы для оценки:

- а.п.п. (абсолютной предельной погрешности) суммы величин с известными а.п.п.;
- о.п.п. (относительной предельной погрешности) произведения величин с известными о.п.п.

К главе 3

1. Определение согласованных и подчиненных норм матриц и векторов.

2. Три нормы вектора и соответствующие им три подчиненные нормы матрицы.

¹⁾ Звездочкой отмечены вопросы повышенной сложности.

3. Получите выражение для третьей (спектральной) нормы матрицы.
4. Теорема о погрешности решения СЛАУ; число обусловленности СЛАУ μ .
5. Покажите, что $\mu \geq 1$.
6. Покажите, что $\mu = \max_i |\lambda_i| / \min_i |\lambda_i|$ для симметрической матрицы \mathbf{A} .
7. Алгоритм численного решения СЛАУ с матрицей треугольной структуры (прямой, не итерационный).
8. Алгоритм прямого и обратного хода метода Гаусса.
- 9*. Представьте метод Гаусса через операции с матрицами.
10. Метод Гаусса с выбором главного элемента; условие применимости метода Гаусса.
11. LU-разложение; алгоритм. Оценка количества арифметических действий методов Гаусса и LU-разложения.
12. Метод Холецкого (алгоритм численного решения).
13. Каноническая форма записи итерационного процесса для численного решения СЛАУ.
14. Достаточное условие сходимости метода простой итерации (МПИ) для численного решения СЛАУ.
15. Дайте оценку количества итераций для получения заданной точности ε при численном решении СЛАУ.
16. Критерий сходимости итерационного процесса для численного решения СЛАУ; сравнение количества арифметических действий прямых и итерационных методов.
17. Влияние ошибок округления на результат численного решения СЛАУ.
18. Методы Якоби, Зейделя, релаксации (алгоритмы).
19. Достаточные условия сходимости методов Якоби, Зейделя (получить).
- 20*. Критерий сходимости метода Якоби (получить); условие сходимости метода Зейделя для симметрической матрицы.
21. Связь между вариационной задачей и задачей решения СЛАУ (теорема).
22. Методы градиентного и наискорейшего спусков (вывод).
23. Метод минимальных невязок (вывод).
- 24*. Метод сопряженных градиентов (без вывода).

К главе 4

1. Получите систему из двух линейных алгебраических уравнений для решения методом наименьших квадратов переопределенной системы из трех линейных уравнений.
2. Сформулируйте теорему о методе наименьших квадратов.

3. Проведите прямую, проходящую наиболее близко (в смысле метода наименьших квадратов) к четырем точкам.

4. Что такое обобщенный полином? Что дает использование систем ортогональных функций для приближения функции методом наименьших квадратов?

5*. Как выглядит матрица Гильберта? В чем состоит ее главная особенность?

6*. Изложите идею метода спектральной эквивалентности матриц для численного решения плохо обусловленных систем уравнений.

7*. В чем состоит метод преобуславливания для численного решения плохо обусловленных систем линейных уравнений?

8*. В чем состоит метод ортогонализации для численного решения плохо обусловленных систем алгебраических уравнений?

9. Пусть задана переопределенная система линейных алгебраических уравнений $\mathbf{A}\mathbf{u} = \mathbf{f}$ (3.1).

Как будет выглядеть соответствующая система линейных алгебраических уравнений с квадратной матрицей, полученная методом наименьших квадратов?

К главе 5

1. Что такое сжимающее отображение?

2. Сформулируйте теорему о сжимающем отображении.

3. Получите достаточное условие сходимости итерационного процесса

$$u_{k+1} = F(u_k), \quad u_0 = a$$

для численного решения нелинейного уравнения $u = F(u)$.

4. Как выглядит достаточное условие сходимости итерационного процесса

$$u_{k+1} = F(u_k), \quad u_0 = a,$$

для численного решения системы нелинейных уравнений $\mathbf{u} = \mathbf{F}(\mathbf{u})$?

5. Дайте определение выпуклой области.

6. Сформулируйте теорему о том, в каком случае отображение $v = F(u)$ является сжимающим.

7. Дайте геометрическую интерпретацию: а) монотонной сходимости, б) немонотонной сходимости, в) расходимости итерационного процесса

$$u_{k+1} = F(u_k), \quad u_0 = a.$$

8. Получите итерационную формулу Ньютона для численного решения скалярного нелинейного уравнения $f(x) = 0$.

9. Как выглядит итерационный метод релаксации для численного решения нелинейного уравнения $f(x) = 0$? При каких значениях итерационного параметра τ он сходится?

10. Получите расчетные формулы итерационного метода Ньютона как метода линеаризации для решения системы нелинейных алгебраических уравнений.

11. Дайте графическую интерпретацию метода Ньютона.

12. Какой порядок сходимости имеет метод Ньютона?

13. Сформулируйте теорему о методе Ньютона.

14*. Приведите пример итерационного метода третьего порядка сходимости.

15*. Представьте: а) итерационный метод касательных для численного уравнения скалярного уравнения $f(x) = 0$; б) метод Ньютона–Канторовича для численного решения системы нелинейных уравнений $\mathbf{f}(\mathbf{x}) = \mathbf{0}$.

16. Предложите вариационный итерационный метод для численного решения системы из двух нелинейных уравнений

$$\begin{cases} f(u, v) = 0, \\ g(u, v) = 0. \end{cases}$$

17. Представьте итерационные формулы метода Ньютона для системы из двух уравнений.

18. Постройте итерационный процесс Ньютона для вычисления $\sqrt[n]{a}$, $a > 0$, n — натуральное.

К главе 6

1. Формулировка задачи интерполяции.

2. Теорема о точности кусочно-линейной интерполяции (формулировка).

3. Сформулируйте теорему о существовании и единственности решения задачи интерполяции при приближении функции обобщенным полиномом.

4. Сформулируйте теорему об условии линейной независимости системы функций $\varphi_n(t_k)$, $k = 0 \div n$, $n = 0 \div N$.

5. Почему удобно использовать для интерполяции систему ортогональных функций $\varphi_n(t_k)$, $k = 0 \div n$, $n = 0 \div N$?

6. Как выглядят базисные функции Лагранжа? Как выглядит интерполяционный полином Лагранжа, представленный через эти базисные функции?

7. Сформулируйте теорему об остаточном члене интерполяции.

8. Оцените остаточный член интерполяции для $\tau = \text{const}$ (τ — шаг интерполяции).

9*. Почему экстраполяция является, вообще говоря, неустойчивым процессом?

10. Что такое разделенные разности? Выпишите их с помощью рекуррентных соотношений.

11. Что такое конечные разности? Выпишите соответствующие формулы для первого–четвертого порядков.

12. Представьте интерполяционный полином Ньютона в общем случае, в линейном и в квадратичном случаях. В чем удобство записи интерполяционного полинома в форме Ньютона?

13. Вид полиномов Чебышёва.

14. Постоянная Лебега.

15. Теорема Чебышёва о полиноме, наименее уклоняющемся от нуля.

16. Задача интерполяции с кратными узлами. Приведите пример.

17. Сформулируйте теорему об остаточном члене интерполяционного полинома с кратными узлами.

18. Дайте определение сплайна $S_{m,d}(t)$.

19. Дайте определение кубического сплайна.

20. Докажите теорему о существовании и единственности интерполяционного кубического сплайна.

21. Как строится кубический сплайн?

22. Сформулируйте теорему о точности сплайн-интерполяции.

23. Сформулируйте теорему об экстремальном свойстве кубических сплайнов.

24*. Определение В-сплайна. Пример В-сплайна для $N = 2$ ($(N - 1)$ — степень сплайна).

25. Как выглядит интерполяционный полином Лагранжа первой степени для функции двух переменных?

26. Общая формула интерполяционного полинома Лагранжа для функции двух переменных.

27. Выпишите интерполяционные полиномы Лагранжа первых двух степеней и их остаточные члены.

К главе 7

1. Получите квадратурные формулы численного интегрирования:

- средних;
- трапеций.

2. Получите формулу численного интегрирования Симпсона.
3. Получите формулу для оценки погрешности квадратурных формул численного интегрирования.
4. Получите локальную и глобальную погрешности для численного интегрирования по формуле трапеций.
5. Представьте формулу численного интегрирования методом средних для функции двух переменных

$$\int_a^b \int_d^c f(x, y) dx dy.$$

6. Покажите, что квадратурная формула для численного интегрирования может быть представлена в виде

$$\int_a^b f(t) dt = \sum_{n=0}^N C_n \cdot f(t_n) + r_N.$$

7. Для каких функций эта формула будет точной ($r_N = 0$); какой вид будут иметь весовые коэффициенты c_n в этом случае?

8*. Возможно ли получить точную формулу численного интегрирования на N точках, если подынтегральная функция является полиномом степени $M > N$? Какова максимальная степень такого полинома?

9*. Получите систему нелинейных уравнений для определения коэффициентов c_n и координат узловых точек t_n , при которых квадратурная формула

$$\int_a^b f(t) dt = \sum_{n=0}^N C_n \cdot f(t_n) + r_N$$

будет точна для полиномов степени $2(N + 1)$, где N — количество узловых точек.

10. Напишите формулу для оценки погрешности численного интегрирования по методу Гаусса.

11. В чем состоит метод Канторовича выделения особенностей при численном интегрировании?

12. Предложите метод приближенного вычисления интеграла от быстроосциллирующей функции:

$$\int_a^b f(t) \sin(\omega t) dt; \quad \omega \gg 1.$$

13*. Получите приближенную формулу для вычисления N -мерного интеграла в N -мерной области интегрирования, находящейся в N -мерном кубе.

14*. Получите формулу численного интегрирования Гаусса по двум узлам интегрирования.

К главе 8

1. Представьте ОДУ N -го порядка

$$\frac{d^N u}{dt^N} = f(t, u, \dots, u^{(N-1)}), \quad t > 0,$$

$$u(0) = b_0, \quad u'(0) = b_1, \dots, u^{(N-1)}(0) = b_{N-1}$$

в виде системы уравнений первого порядка.

2. Для ОДУ

$$\frac{du}{dt} = f(t, u), \quad t > 0, \quad u(0) = a$$

представьте методы:

- явный Эйлера;
- неявный Эйлера;
- Эйлера с пересчетом.

3. Для ОДУ

$$\frac{du}{dt} = f(t, u), \quad t > 0, \quad u(0) = a,$$

используя формулу

$$u(t_n + \tau) = x(t_n) + \int_{t_n}^{t_n + \tau} u'(\xi) d\xi,$$

получите:

- явный метод Эйлера;
- неявную формулу трапеций.

4. Для ОДУ

$$\frac{du}{dt} = f(t, u), \quad t > 0, \quad u(0) = a$$

представьте явный метод «предиктор–корректор» второго порядка аппроксимации.

5. Представьте общий вид методов Рунге–Кутты для численного решения ОДУ следующего вида:

$$\frac{du}{dt} = f(t, u), \quad t > 0, \quad u(0) = a.$$

6. Представьте таблицу Бутчера для r -стадийного явного метода Рунге–Кутты.

7. Представьте таблицу Бутчера для:

- явного метода Эйлера;
- метода Эйлера с пересчетом.

8. Получите явный метод Эйлера с помощью метода неопределенных коэффициентов, исследуя выражение для невязки.

9. Используя только определение сходимости, покажите, что метод Эйлера:

$$\frac{u_{n+1} - u_n}{\tau} + au_n = 0; \quad n = 0, 1, \dots, \quad u(0) = a,$$

аппроксимирует ОДУ вида

$$\frac{du}{dt} + au = 0, \quad t > 0, \quad u(0) = a,$$

с первым порядком точности.

10. В чем состоит причина появления барьеров Бутчера?

11. Возможно ли построить явный 5-стадийный метод Рунге–Кутты 5-го порядка точности?

12. Сформулируйте теорему об устойчивости методов Рунге–Кутты.

13. Из каких соображений выбирается шаг по времени τ для метода Рунге–Кутты:

$$\frac{u_{n+1} - u_n}{\tau} = F(u_n), \quad u_0 = a?$$

14*. На каких временах интегрирования гарантируется устойчивость методов Рунге–Кутты, если правая часть удовлетворяет условию Липшица:

- для устойчивых траекторий;
- для нейтральных траекторий?

К главе 9

1. Приведите пример жесткой задачи Коши для ОДУ.

2. Является ли жесткой система ОДУ:

$$\begin{cases} \dot{u} = \alpha u + \varepsilon^{-1} v, \\ \dot{v} = -\varepsilon u, \end{cases}$$

$$0 < \varepsilon \ll 1, \quad u(0) = v(0) = 1,$$

$$\alpha = \text{const}, \quad \alpha = O(1)?$$

3. Дайте определение жесткой задачи Коши для ОДУ.

4. Представьте точное решение задачи Коши для жесткой системы уравнений

$$\dot{\mathbf{u}} = \mathbf{A}\mathbf{u}, \quad \mathbf{u}(0) = \mathbf{a}; \quad t > 0,$$

\mathbf{A} — матрица с постоянными коэффициентами.

5. Представьте точное решение системы разностных уравнений

$$\frac{\mathbf{u}_{n+1} - \mathbf{u}_n}{\tau} = \mathbf{A}\mathbf{u}_n; \quad \mathbf{u}_0 = \mathbf{a}, \quad n = 0, 1, \dots,$$

аппроксимирующих систему ОДУ:

$$\dot{\mathbf{u}} = \mathbf{A}\mathbf{u}, \quad \mathbf{u}(0) = \mathbf{a}; \quad t > 0.$$

6. Представьте точное решение системы разностных уравнений:

$$\frac{\mathbf{u}_{n+1} - \mathbf{u}_n}{\tau} = \mathbf{A}\mathbf{u}_{n+1}; \quad \mathbf{u}_0 = \mathbf{a}, \quad n = 0, 1, \dots,$$

для численного решения системы ОДУ

$$\dot{\mathbf{u}} = \mathbf{A}\mathbf{u}, \quad \mathbf{u}(0) = \mathbf{a}, \quad t > 0;$$

сравните решения разностного и дифференциального уравнений.

7. Чем принципиально различаются явные и неявные методы?

8. Какой метод называется абсолютно устойчивым?

9. Какие методы называются:

- A -устойчивыми;
- L -устойчивыми;
- $A(\alpha)$ -устойчивыми?

10. Получите функцию устойчивости для ОДУ следующего вида:

$$\frac{x_{n+1} - x_n}{\tau} = \lambda x_n; \quad \lambda = \text{const}; \quad t > 0, \quad n = 0, 1, \dots$$

11. Сформулируйте теорему Далквиста (барьер Далквиста).

12. Покажите, что неявный метод Эйлера является L -устойчивым.

13. Является ли жесткой система нелинейных ОДУ:

$$\begin{cases} \varepsilon \dot{u} = F(u, v), \\ \dot{v} = G(u, v), \end{cases}$$

$$0 < \varepsilon \ll 1; \quad F, G = O(1); \quad t > 0;$$

$$u(0) = a; \quad v(0) = b; \quad a, b = O(1)?$$

14. Изобразите на графике (v, u) устойчивые и неустойчивые ветви решения системы ОДУ:

$$\begin{cases} \dot{u} = v - \frac{u^3}{3} + u, \\ \dot{v} = -u; \end{cases}$$

$$u(0) = u_0; \quad v(0) = v_0.$$

15*. Почему, пользуясь неявным численным методом, можно получить решения, соответствующие неустойчивой ветви решения системы ОДУ?

16. Представьте общий вид неявных методов Рунге–Кутты.

17. Представьте общий вид таблицы Бутчера для неявных методов Рунге–Кутты.

18. Представьте таблицы Бутчера для:

- неявного метода Эйлера;
- неявного метода трапеций.

19. Представьте вид полуявного метода Розенброка.

20. Представьте общий вид многошаговых методов.

21. Какие из многошаговых методов называются:

- явными;
- неявными;
- чисто неявными;
- явными методами Адамса?

22. В чем состоит условие корней?

23*. Как получается характеристическое уравнение для одно-родного разностного уравнения?

24. В чем состоит метод неопределенных коэффициентов для получения многошаговых методов?

К главе 10

1. В чем состоит метод фундаментальных решений (МФР) краевых задач для систем ОДУ первого порядка?

2. Приведите пример, когда МФР неприменим для численного решения краевой задачи для системы ОДУ.

3. Дайте определение жесткой краевой задачи для системы ОДУ.

4. Какие краевые задачи для системы ОДУ называются вычислительно корректными?

5. Сформулируйте краевую задачу Штурма–Лиувилля для ОДУ.

6. Проведите разностную аппроксимацию краевой задачи Штурма–Лиувилля с переменными коэффициентами.

7. Получите канонический вид СЛАУ с трехдиагональной матрицей после разностной аппроксимации краевой задачи Штурма–Лиувилля.

8. Представьте алгоритм метода трехточечной прогонки.

9. Условия устойчивости метода прогонки.

10. Для каких задач применим метод прогонки: линейных или нелинейных (обосновать)?

11. Метод стрельбы численного решения краевой задачи для ОДУ второго порядка.

12. Можно ли использовать метод прогонки для численного решения нелинейной задачи Штурма–Лиувилля?

13*. Покажите устойчивость обратной прогонки. При выполнении каких условий обратная прогонка устойчива?

14. Метод Фурье для приближенного решения краевой задачи типа Штурма–Лиувилля.

15. Метод квазилинеаризации Ньютона для численного решения задачи Штурма–Лиувилля.

16. Приведите пример краевой задачи на собственные значения.

17*. Какие собственные значения имеет разностный оператор Λ_{xx} ?

К главе 11

1. Определение линейного разностного уравнения (ЛРУ).

2. Представьте точное решение для однородного ЛРУ первого порядка:

$$\begin{aligned}\alpha u_k + \beta u_{k+1} &= 0; \\ k &= 0, 1, \dots\end{aligned}$$

3. Представьте вид частного решения неоднородного ЛРУ первого порядка.

4. Представьте общий вид решения неоднородного ЛРУ второго порядка.

5. Получите точное решение однородного ЛРУ второго порядка в случаях:

- вещественных различных корней характеристического уравнения;
- вещественных кратных корней характеристического уравнения.

6. Представьте частное решение неоднородного ЛРУ, если его правая часть имеет вид

$$f_k = \alpha^k [P_m(k) \cos(\beta k) + k \sin(\beta k)],$$

где α, β — параметры уравнения.

7. Представьте вид точного решения однородной системы ЛРУ:

$$\mathbf{u}_{n+1} = \mathbf{A} \mathbf{u}_n.$$

8. Представьте частное решение неоднородной системы ЛРУ, если ее правая часть имеет вид

$$\mathbf{f}_n = \mu^k \cdot \mathbf{P}_e(k),$$

μ — параметр системы.

9. Что называется жордановой цепочкой для собственных значений λ матрицы \mathbf{A} ?

10. Представьте точное решение системы ЛРУ второго порядка с вещественным собственным числом λ кратности 2.

11. Решите систему ЛРУ вида

$$\begin{cases} u_{k+1} = u_k - v_k + 3^k, \\ v_{k+1} = -2u_k - 3^k. \end{cases}$$

12. Получите точное решение разностного уравнения вида

$$au_{n-1} + bu_n + cu_{n+1}; \quad n = 1, 2, \dots$$

К главе 12

1. Постановка смешанной задачи для нестационарного уравнения теплопроводности.

2. Дискретизация области интегрирования, сеточная функция.

3. Явная четырехточечная схема для нестационарного уравнения теплопроводности. Постановка разностной задачи, алгоритм ее решения.

4. Неявная четырехточечная схема для уравнения теплопроводности. Постановка разностной задачи, алгоритм решения.

5. Определение корректности разностной задачи.

6. Определения сходимости, аппроксимации, устойчивости разностной задачи.

7. Теорема эквивалентности (формулировка, доказательство).

8. Операторная форма дифференциальной и разностной задач.

9. Необходимое условие сходимости Куранта–Фридрихса–Леви.



10. Исследуйте на аппроксимацию схему «явный правый угол» для численного решения линейного уравнения переноса

$$u'_t - au'_x = 0, \quad a > 0.$$

11. Найдите первое дифференциальное приближение для схемы «явный правый угол» для численного решения уравнения

$$u'_t - au'_x = 0.$$

12. Исследуйте на аппроксимацию 4-точечную явную схему для численного решения линейного уравнения теплопроводности

$$u'_t - au''_{xx} = 0, \quad a > 0.$$

13. Найдите первое дифференциальное приближение для схемы в п. 12.

14. Представьте операторный канонический вид двухслойной разностной схемы для численного уравнения в частных производных. Приведите пример такой записи.

15. Дайте определение равномерной устойчивости.

16. Сформулируйте условие ограниченности норм степеней оператора перехода с нижнего на верхний временной слой для двухслойной разностной схемы.

17. Получите условие устойчивости разностной схемы вида

$$|\lambda| \leq 1 + C\tau.$$

18. В чем состоит признак спектральной устойчивости Неймана?

19. С помощью спектрального признака Неймана получите условие устойчивости схемы

$$\frac{u_m^{n+1} - u_m^n}{\tau} - \frac{u_{m+1}^n - u_m^n}{h} = 0.$$

20. Какое дифференциальное уравнение аппроксимирует схема из вопроса 19?

21. С помощью спектрального признака Неймана получите условие устойчивости схемы

$$\frac{u_m^{n+1} - u_m^n}{\tau} - a \frac{u_{m+1}^n - 2u_m^n + u_{m-1}^n}{h^2} = 0.$$

22. Какое дифференциальное уравнение аппроксимирует схема из вопроса 21?

23*. Покажите, что если разностная схема равномерно устойчива по начальным данным, то она устойчива по правой части.

24. Дайте определение энергетической нормы оператора.

25. Что означает операторное неравенство $A > B$ (A, B — операторы)?

26. Сформулируйте критерий устойчивости разностной схемы по начальным данным.

27*. Покажите с помощью принципа максимума устойчивость схемы «явный уголок», аппроксимирующей линейное одномерное уравнение переноса при выполнении условия Куранта.

28. Получите условие устойчивости явной схемы «уголок», аппроксимирующей линейное уравнение переноса.

29. Получите условие устойчивости явной четырехточечной схемы, аппроксимирующей одномерное уравнение теплопроводности.

30. То же, что в вопросе 29, но для неявной схемы.

31. Дайте определения сходимости, аппроксимации, устойчивости.

32. Докажите теорему эквивалентности.

33. Сформулируйте спектральный признак устойчивости Рябенского–Неймана.

34. Сформулируйте принцип максимума для доказательства устойчивости схем.

35. Что такое равномерная устойчивость по начальным данным?

36. Сформулируйте теорему об устойчивости разностной схемы (теорему Самарского).

37. Сформулируйте критерий устойчивости разностной схемы (по Самарскому).

38. Приведите примеры доказательств устойчивости схемы с помощью:

- спектрального признака Неймана;
- принципа максимума;
- критерия устойчивости.

К главе 13

1. Сформулируйте смешанную задачу для одномерного нестационарного линейного уравнения теплопроводности (ОНУТ).

2. Сформулируйте смешанную задачу для двумерного нестационарного линейного уравнения теплопроводности.

3. Сформулируйте смешанную задачу для трехмерного уравнения теплопроводности.

4. Представьте параметрическую шеститочечную разностную схему Кранка–Никольсон и ее свойства для численного решения ОНУТ с постоянными коэффициентами.

5. Исследуйте на сходимость явную четырехточечную схему для ОНУТ с постоянными коэффициентами.

6. Получите первое дифференциальное приближение четырехточечной явной схемы для ОНУТ с постоянными коэффициентами.

7. Как можно увеличить порядок аппроксимации разностной схемы для ОНУТ с постоянными коэффициентами, используя понятие первого дифференциального приближения?

8. Представьте трехслойную разностную схему (РС) для ОНУТ с постоянными коэффициентами.

9. Нарисуйте шаблоны для двух- и трехслойных РС для ОНУТ с постоянными коэффициентами.

10. Представьте четырехточечную неявную РС для ОНУТ с переменными коэффициентами.

11. Представьте шеститочечную схему Кранка–Никольсон для ОНУТ с нелинейной правой частью и нелинейным коэффициентом теплопроводности, а также алгоритм численного решения задачи.

12. В чем состоит метод квазилинеаризации для численного решения нелинейного ОНУТ?

13. В чем состоит интегро-интерполяционный метод для численного решения ОНУТ (на примере шеститочечной двухслойной РС)?

14. Представьте явную и неявную РС для численного решения двумерного нестационарного уравнения теплопроводности (НУТ), а также постановку разностной задачи.

15. Почему для численного решения двумерного НУТ обычно используют неявные РС?

16. Исследуйте простейшую явную двухслойную РС для численного решения двумерного НУТ на аппроксимацию и устойчивость.

17. Изложите РС расщепления по направлениям для численного решения двумерного НУТ.

18. Изложите РС расщепления по направлениям для численного решения трехмерного НУТ.

19. Представьте продольно-поперечную РС для численного решения двумерного НУТ.

20. Представьте РС Писмэна–Рэкфорда для численного решения трехмерного НУТ.

21. Проведите исследование на устойчивость продольно-поперечной РС для численного решения двумерного НУТ.

22. Исследуйте на аппроксимацию локально-одномерную схему расщепления для двумерного НУТ.

23*. Локально-одномерная схема с весовым коэффициентом типа Кранка–Никольсон: 2D (шаблон, устойчивость, аппроксимация), 3D.

24*. Схема Дугласа–Гана, 3D.

К главе 14

1. Линейные и нелинейные уравнения переноса.

2. Разностные схемы и их исследования на сходимость для линейного одномерного уравнения переноса:

- Лакса;
- Куранта–Изаксона–Риса;
- Бабенко;
- Кранка–Никольсон;
- неявные «уголки»;
- «кабаре»;
- Бима–Уорминга.

3. Разностные схемы для численного решения нелинейного уравнения переноса:

- Лакса;
- Куранта–Изаксона–Риса;
- Лакса–Вендроффа;
- Мак–Кормака;
- * Русанова;
- * Уорминга–Кутлера–Ломакса.

4. Акустическая система. Постановка задачи. Инварианты Римана. Численное решение.

5. Волновые уравнения. Постановка задачи. Численные методы решения.

6. Исследование линейных уравнений гиперболического типа на аппроксимацию, устойчивость.

7. Условие Куранта–Фридрихса–Леви (КФЛ).

К главе 15

1. Уравнение Пуассона. Постановка краевой двумерной задачи.

2. Схема «крест» для численного решения уравнения Пуассона, порядок аппроксимации.

3. Принцип максимума для схемы «крест».

4. Метод простых итераций (МПИ). Оценка их количества.
5. Метод итераций с оптимальным и итерационным параметрами. Оценка их количества.
6. Чебышёвское ускорение для МПИ.
7. Трёхслойный метод Чебышёва.
8. Метод переменных направлений.
9. Методы Якоби, Зейделя, верхней релаксации для решения уравнения Пуассона.
10. Сравнительный анализ итерационных методов (по количеству итераций).



Приложение 2

ПРИМЕРЫ ЗАДАЧ К ВЫЧИСЛИТЕЛЬНОМУ ПРАКТИКУМУ ПО КУРСУ

1. Вокруг Земли вращается спутник по круговой орбите радиуса $r_c = 10^4$ км.

Проработав короткое время, двигатели сообщили спутнику скорость u в направлении, противоположном движению.

Рассчитать новую траекторию спутника. При каком значении u спутник коснется поверхности Земли?

Уравнение движения спутника:

$$\begin{aligned}\ddot{x} &= -\frac{\gamma M}{r^3}x, & \ddot{y} &= -\gamma \frac{M}{r^3}y, \\ r &= \sqrt{x^2 + y^2}, & x(0) &= r_c, & r_c &= 10^4 \text{ км}; \\ y(0) &= 0, & \dot{y}(0) &= v_c - u, & \dot{x}(0) &= 0, \\ v_c &= \sqrt{\frac{\gamma M}{r_c}}\end{aligned}$$

— скорость спутника на круговой орбите.

Масса Земли $M = 5,99 \cdot 10^{24}$ кг, $\gamma = 6,67 \cdot 10^{-11} \text{ м}^3 \text{ кг}^{-1} \text{ с}^{-2}$, $R = 6380$ км (радиус Земли).

- Построить график траектории в плоскости (x, y) ;
- проверить третий закон Кеплера:

$$T = \frac{2\pi a^{3/2}}{\sqrt{\gamma M}}.$$

2. Задача трех тел (Земля, Луна, спутник)

$$\begin{cases} \ddot{x} = 2\dot{y} + x - \frac{\bar{\mu}(x + \mu)}{r_1^3} - \frac{\mu(x - \bar{\mu})}{r_2^3} - k\dot{x}, \\ \ddot{y} = -2\dot{x} + y - \frac{\bar{\mu}y}{r_1^3} - \frac{\mu y}{r_2^3} - k\dot{y}, \end{cases}$$

$\mu = 1/82,45$ (отношение масс Луны и Земли); Земля и Луна — в точках $(1 - \mu, 0)$, $(-\mu, 0)$, масса спутника пренебрежимо мала по сравнению с массами Земли и Луны (его положение — $\{x, y\}$); первые производные появляются вследствие вращения системы

координат и трения, пропорционального скорости с коэффициентом пропорциональности k .

$$\begin{cases} \bar{\mu} = 1 - \mu; & r_1^2 = (x + \mu)^2 + y^2, & r_2^2 = (x - \bar{\mu})^2 + y^2; \\ x(0) = 1,2; & \dot{x}(0) = 0; & y(0) = -1,05. \end{cases}$$

При $k = 0$ — периодическое движение с периодом $T \approx 6,2$. Провести расчеты с $k = 0$; $k = 0,1$; $k = 1$; $0 \leq t \leq 8$.

- Провести расчеты методами Рунге–Кутты первого и второго порядков аппроксимации;
- провести исследования сходимости численного решения по сетке.

3. а) Система ОДУ, описывающая изменение численности популяций двух видов и эволюцию некоего генетического признака α , имеет вид

$$\begin{cases} \dot{x} = x \left(1 - 0,5x - \frac{2}{7\alpha^2}y \right), & (a) \\ \dot{y} = y (2\alpha - 3,5\alpha^2x - 0,5y), & (б) \\ \dot{\alpha} = \varepsilon (2 - 7\alpha x), & (в) \end{cases} \quad (П2.1)$$

$0 < \varepsilon \leq 10^{-2}$, $0 \leq x(0) \leq 3$, $0 \leq y(0) \leq 15$, $\alpha(0) = 0$; из (в) видно, что генетический признак изменяется медленнее, чем численность популяций (решение — релаксационные колебания).

б) Более интересный случай — численность двух популяций зависит от взаимодействия между ними и от двух медленно изменяющихся генетических признаков:

$$\begin{cases} \dot{x} = x (2\alpha_1 - 0,5x - \alpha_1^2\alpha_2^{-2}y), \\ \dot{y} = y (2\alpha_2 - \alpha_1^{-2} \cdot \alpha_2^{-2}x - 0,5y), \\ \dot{\alpha}_1 = \varepsilon (2 - 2\alpha_1\alpha_2^{-2}y), \\ \dot{\alpha}_2 = \varepsilon (2 - 2\alpha_2\alpha_1^{-2}x), \end{cases} \quad (П2.2)$$

$0 < \varepsilon \leq 0,01$; $0 \leq x(0) \leq 40$, $0 \leq y(0) \leq 40$, $\alpha_1(0) = 0$; $\alpha_2(0) = 10$; другой вариант (П2.2) имеет вид

$$\begin{cases} \dot{x} = x (2\alpha_1 - 0,5x - \alpha_1^3 \cdot \alpha_2^{-3}y), \\ \dot{y} = y (2\alpha_2 - \alpha_1^{-3} \cdot \alpha_2^3x - 0,5y), \\ \dot{\alpha}_1 = \varepsilon (2 - 3\alpha_1^2 \cdot \alpha_2^{-3}y), \\ \dot{\alpha}_2 = \varepsilon (2 - 3\alpha_2^2 \cdot \alpha_1^{-3}x), \end{cases} \quad (П2.3)$$

$0 < \varepsilon \leq 0,001$, $0 < x(0) \leq 40$, $0 \leq y(0) \leq 40$, $\alpha_1(0) = 0$, $\alpha_2(0) = 10$.

Нижеследующие три задания следует выполнить для временного интервала $0 \leq t \leq 2000$:

- исследовать изменения двух видов (соответствующие численности — x, y) и их генетических признаков ($\alpha, \alpha_1, \alpha_2$) в зависимости от времени t , построить зависимости $x(t), y(t), \alpha(t), \alpha_2(t), x(y)$;
- исследовать разностные схемы на сходимость по сетке;
- использовать для численного решения (П2.1), (П2.2), (П2.3) явные методы Рунге–Кутты 1-го и 4-го порядков точности и неявный метод Рунге–Кутты (Хаммера–Холлинсворта).

4. Автономные и неавтономные уравнения Ван дер Поля, а также уравнение Рэля, описывающие колебательные процессы в электрических цепях, имеют вид:

$$\begin{cases} \frac{dy_1}{dt} = -a \left(\frac{y_1^3}{3} - y_1 \right) - ay_2, \\ \frac{dy_2}{dt} = -y_1 \end{cases} \quad (\text{П2.4})$$

(уравнение Ван дер Поля);

$$\begin{cases} \frac{dy_1}{dt} = -a \left(\frac{y_1^3}{3} - y_1 \right) + ay_2, \\ \frac{dy_2}{dt} = -y_1 - by_2 + c \end{cases} \quad (\text{П2.5})$$

(уравнение Бонгоффера–Ван дер Поля);

$$\begin{cases} \frac{dy_1}{dt} = -a \left(\frac{y_1^3}{3} - y_1 \right) + ay_2, \\ \frac{dy_2}{dt} = -y_1 + A \cos \omega t \end{cases} \quad (\text{П2.6})$$

(неавтономное уравнение Ван дер Поля, траектория-утка);

$$\ddot{x} - a(1 - \dot{x}^2)\dot{x} + x = 0 \quad (\text{П2.7})$$

(уравнение Рэля).

В (П2.4)–(П2.6) $1 \leq a \leq 10^3$; $y_{10} = 2, y_{20} = 0$; в (П2.5) рассмотреть два случая:

$$0 < A < 1; \quad 1 < A < \sqrt{1 + \frac{1}{64\omega^2}}; \quad 0 \leq t \leq 200, \quad 0 < c < 1.$$

- Провести исследование поведения численных решений (П2.4)–(П2.7) в зависимости от «большого» параметра a ; (П2.6) — в зависимости от ω ;

- построить зависимости $y_1(t), y_2(t), y_2(y_2)$;
- использовать явные методы Рунге–Кутты 1-го и 4-го порядков точности, неявный метод (Хаммера–Холлинсворта);
- исследовать зависимость численного решения от шага интегрирования τ (сходимость в сетке).

5. Изучить поведение концентраций веществ в химических реакциях Белоусова–Жаботинского:

$$\begin{cases} \frac{dy_1}{dt} = 77,27 [y_2 + y_1 (1 - 8,375 \cdot 10^{-6} y_1 - y_2)], \\ \frac{dy_2}{dt} = \frac{1}{77,27} [y_3 - (1 + y_1) y_2], \\ \frac{dy_3}{dt} = 0,16(y_1 - y_3), \end{cases} \quad (\text{П2.8})$$

$$0 \leq t \leq 800, \quad y_1(0) = 1, \quad y_2(0) = 2, \quad y_3(0) = 3.$$

$$\begin{cases} \frac{dy_1}{dt} = -0,04y_1 + 10^4 y_2 y_3, \\ \frac{dy_2}{dt} = 0,04y_1 - 10^4 y_2 y_3 - 3 \cdot 10^7 y_2^2, \\ \frac{dy_3}{dt} = 3 \cdot 10^7 y_2^2, \end{cases} \quad (\text{П2.9})$$

$$0 \leq t \leq 1000, \quad y_1(0) = 1, \quad y_2(0) = y_3(0) = 0.$$

$$\begin{cases} \frac{dy_1}{dt} = -Ay_1 - By_1 y_3, \\ \frac{dy_2}{dt} = Ay_1 - M \cdot Cy_2 y_3, \\ \frac{dy_3}{dt} = Ay_1 - By_1 y_3 - M \cdot Cy_2 y_3 + Cy_4, \\ \frac{dy_4}{dt} = By_1 y_3 - Cy_4, \end{cases} \quad (\text{П2.10})$$

$$0 \leq t \leq 1013, \quad y_1(0) = 1,76 \cdot 10^{-3}, \quad y_2(0) = y_3(0) = y_4(0) = 0;$$

$$A = 7,89 \cdot 10^{-10}; \quad B = 1,1 \cdot 10^7; \quad C = 1,13 \cdot 10^3, \quad M = 10^6.$$

- Построить зависимости параметров от времени и зависимости $y_i(y_j)$, $i \neq j$;
- исследовать сходимость численного решения по сетке;
- использовать явный и неявный методы Рунге–Кутты 4-го порядка точности.

6. Уравнение Ван дер Поля

$$\frac{d^2 y}{dt^2} + a(y^2 - 1) + \frac{dy}{dt} + y = 0,$$

$$y(0) = y_0 > 0; \quad y'(0) = 0; \quad 0 \leq t \leq 30, \quad 1 \leq a \leq 1000, \quad (\text{П2.11})$$

описывает нелинейные колебания в различных системах.

Уравнение Эйлера

$$\ddot{y} + t^{-1} \dot{x} + 10^2 t^{-2} x = 0,$$

$$10 \leq t \leq 10^2, \quad x(1) = 1, \quad \dot{x}(1) = 1, \quad (\text{П2.12})$$

описывает колебания в системе, где возвращающая сила и коэффициент вязкого трения убывают со временем.

Уравнение Капицы

$$L\ddot{\Theta} + (g - A\omega^2 \sin \omega t) \sin \Theta \quad (\text{П2.13})$$

(при $\sin \Theta \approx \Theta$ — уравнение Матьё, при $A = 0$ — уравнение колебания маятника $\ddot{\Theta} = (g/2) \sin \Theta$; L — длина маятника, Θ — угол отклонения от вертикали) описывает колебания «перевернутого» маятника.

Уравнение Минорского

$$\frac{d^2 y}{dt^2} + 2r\dot{y} + \omega^2 y + 2q\dot{y}(t-1) = \varepsilon \dot{y}^3(t-1) \quad (\text{П2.14})$$

встречается в механических и электромеханических задачах с запаздыванием и нелинейностью ($r = -1$, $q = -1$, $\omega = n\pi$; начальные данные задаются для $t \in [-1, 0]$).

- Исследовать зависимость численных решений от параметров процессов;
- исследовать сходимость численного решения по сетке;
- использовать схемы Рунге–Кутты порядка не менее $O(r^4)$, сравнить с численными решениями, полученными по методу Эйлера ($\Theta(\tau)$).

В (П2.13):

L	A	ω	$\Theta(0)$
10	0,5	5,3	3,10
10	10,0	100,0	3,10
10	10,0	100,0	0,10
10	2,0	100,0	0,10
10	0,5	200,0	0,05

- представить зависимости параметров от времени и фазовые портреты.

7. Рассмотрим краевые задачи для ОДУ.

- $\varepsilon y'' = (y')^2$, $y(0) = 1$, $y(1) = 0$, $0 < \varepsilon \ll 1$.
 - а) Получить точное решение, сделав замену переменных $y_x^{-1} = p(y)$.
 - б) Численно исследовать поведение решения при $\varepsilon \rightarrow 0$ ($0 < \varepsilon < 1$), сравнить с точным.
- $\varepsilon y'' = [y - u(x)]^{2q+1}$, $y(-1) = A$, $y(1) = B$, $q \in N$, $0 < \varepsilon \ll 1$. При $u(x) = |x|$ образуется погранслоем вблизи $x = 0$. Рассмотреть случаи: $u(x) = |x|$, $A = 1$, $B > 1$; $u(x) = x^2$, $A > 1$, $B > 1$; $u(x) = |x|$, $A > 1$, $B > 1$.
Что происходит при увеличении q ?
- $\varepsilon y'' = y - y^3$, $y(0) = A$, $y(1) = B$, $|A| < \sqrt{2}$, $|B| < \sqrt{2}$, $0 < \varepsilon \ll 1$, $\varepsilon = 10^{-2}; 10^{-3}; 10^{-4}$ («пиковые» структуры).
- $\varepsilon y'' = y^3 - q$, $y(0) = A < -1$, $y(1) = B > 1$ (внутренний погранслоем $x = 1/2$).
Исследовать толщину погранслоя в зависимости от ε .
- $\varepsilon y'' = -y[y + a(x)]$, $y(0) = y_0$, $y(1) = y_1$, $a(x) = x$.
Рассмотреть поведение решения при $\varepsilon \rightarrow 0$ (удастся ли получить погранслоем типа всплеска?)

8. Численно решить задачу на нахождение собственных значений и функции волнового уравнения методами стрельбы и прогонки: $y'' = -k^2 y$, $y(0) = y(1) = 0$.

Сравнить численные решения между собой и с точными:

$$k_n = n\pi, \quad y_n \approx \sin(n\pi x), \quad n > 0.$$

Рассмотреть случай больших k .

- Использовать явный и неявный методы не ниже четвертого порядка точности, сравнить полученные решения с численным решением, полученным по методу (любому) первого порядка точности;
- для получения численного решения задач из п. 7.1 ÷ 7.5 использовать методы стрельбы и прогонки (квазилинеаризации). Какой из этих двух методов, на ваш взгляд, предпочтительнее?

9. Численно показать, что решение задачи

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{\partial}{\partial x} \left(a u^\alpha \frac{\partial u}{\partial x} \right), & \alpha > 0, \\ u(\infty, t) = 0, & u(0, t) = c t^{1/\alpha}, \quad u(x, 0) = 0 \end{cases} \quad (\text{П2.15})$$

представляет собой бегущую волну, распространяющуюся с конечной скоростью, причем на фронте решение терпит разрыв

первой производной (обобщенное решение). Сравните численное решение с точным:

$$u(t, x) = \left[\frac{\alpha v}{a} (x - vt) \right]^{1/\alpha}, \quad (\text{П2.16})$$

где $v = ac^\alpha / \alpha$. Положить: $\alpha = 1; 3/2; 2$; $a = 0,1; 1; 10$. Использовать схему вида

$$\frac{u_m^{n+1} - u_m^n}{\tau} = \frac{1}{h} \left(k_{m+1/2} \frac{u_{m+1}^{n+1} - u_m^{n+1}}{h} - k_{m-1/2} \frac{u_m^{n+1} - u_{m-1}^{n+1}}{h} \right); \quad (\text{П2.17})$$

проверить численно, какой из вариантов вычисления k предпочтительнее:

$$\begin{aligned} \text{а) } k_{m+1/2} &= \frac{a}{2} [(u_m^n)^\alpha + (u_{m+1}^n)^\alpha], \\ \text{б) } k_{m+1/2} &= a \left(\frac{u_m^n + u_{m+1}^n}{2} \right)^\alpha, \\ \text{в) } k_{m+1/2} &= a \left(\frac{2u_m^n u_{m+1}^n}{u_m^n + u_{m+1}^n} \right)^\alpha, \\ \text{г) } k_{m+1/2} &= a \frac{2(u_m^n)^\alpha (u_{m+1}^n)^\alpha}{(u_m^n)^\alpha + (u_{m+1}^n)^\alpha} \end{aligned} \quad (\text{П2.18})$$

- Построить профили $u(x)$ по времени ($u(t, x)$ — температура внутри сверхновой звезды при взрыве, который инициирует так называемую тепловую волну);
- положив $k = \text{const} = 1$, рассмотреть численное решение, полученное при помощи разностных схем с шаблонами, приведенными на рис. П2.1.

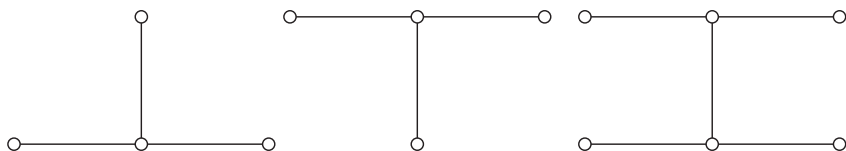


Рис. П2.1

10. Получить численное решение уравнения теплопроводности, описывающего распространение температуры

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}, \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1, \quad (\text{П2.19})$$

$$\begin{aligned} u(0, x, y) &= 0; & u(t, 0, y) &= 0; & u(t, 1, y) &= 1; \\ u(t, x, 0) &= 2; & u(t, x, 1) &= 3, \end{aligned}$$

используя разностные схемы расщепления:

$$\text{а) } \frac{\tilde{u}_{ml} - u_{ml}^n}{\tau} = \Lambda_1 \tilde{u}_{ml}, \quad \frac{u_{ml}^{n+1} - \tilde{u}_{ml}}{\tau} = \Lambda_2 u_{ml}^{n+1}; \quad (\text{П2.20})$$

$$\text{б) } \begin{cases} \frac{u_{ml}^{n+1/2} - u_{ml}^n}{\tau} = \Lambda_1 \left[\xi u_{ml}^{n+1/2} + (1 - \xi) u_{ml}^n \right], \\ \frac{u_{ml}^{n+1/2} - u_{ml}^{n+1}}{\tau} = \Lambda_2 \left[\xi u_{ml}^{n+1} + (1 - \xi) u_{ml}^{n+1/2} \right], \end{cases} \quad \xi = 1/2; \quad (\text{П2.21})$$

$$\text{в) } \begin{cases} \frac{\tilde{u}_{ml}^- u_{ml}^n}{\tau} = \frac{1}{2} (\Lambda_1 \tilde{u}_{ml}^+ \Lambda_2 u_{ml}^n), \\ \frac{u_{ml}^{n+1/2} - u_{ml}^{n+1}}{\tau} = \frac{1}{2} (\Lambda_1 \tilde{u}_{ml}^+ \Lambda_2 u_{ml}^{n+1}); \end{cases} \quad (\text{П2.22})$$

$$\text{г) } \frac{\tilde{u}_{ml} - u_{ml}^n}{\tau} = \Lambda_1 u_{ml}^n, \quad \frac{u_{ml}^{n+1} - \tilde{u}_{ml}}{\tau} = \Lambda_2 \tilde{u}_{ml}, \quad (\text{П2.23})$$

$$\Lambda_1 u_{ml}^{n+1} = \frac{u_{m-1,l}^{n+1} - 2u_{m,l}^{n+1} + u_{m+1,l}^{n+1}}{h_x^2}; \quad \Lambda_2 = \frac{u_{m,l-1}^{n+1} - 2u_{m,l}^{n+1} + u_{m,l+1}^{n+1}}{h_y^2},$$

h_x, h_y — шаги по x, y .

- сравнить их (по u в нескольких точках);
- исследовать сходимость численного решения по сетке.

11. Сравнить численные решения, полученные по разностным схемам Лакса, Куранта–Изаксона–Риса, Лакса–Вендроффа, Уорминга–Кутлера–Ломакса для уравнения переноса в недивергентной и дивергентной формах:

$$\begin{aligned} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} &= 0, \\ \frac{\partial u}{\partial t} + \frac{\partial (u^2/2)}{\partial x} &= 0. \end{aligned}$$

Начальные профили представлены на рис. П2.2.

Исследовать сходимость численных решений по сетке (при $h \rightarrow 0$, h — шаг по координате).

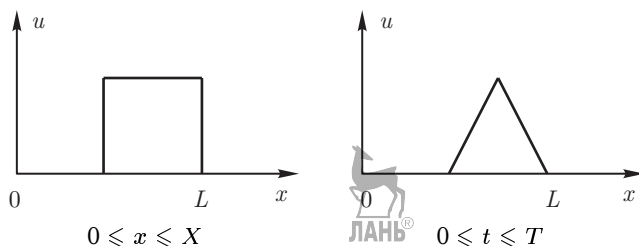


Рис. П2.2

12. Сравнить численные решения, полученные по разностным схемам:

- Куранта–Изаксона–Риса,
- Мак-Кормака,
- гибридной схеме Федоренко,
- TVD,

для линейного одномерного уравнения переноса

$$\frac{\partial u}{\partial t} + c \frac{\partial u}{\partial x} = 0.$$

Начальные профили представлены на рис. П2.3.

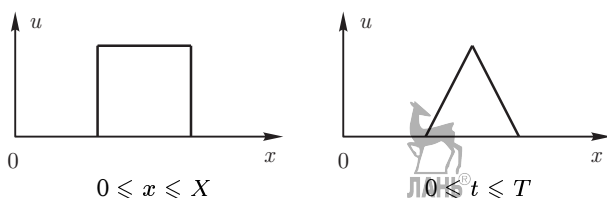


Рис. П2.3

Исследовать сходимость численных решений по сетке (при $h \rightarrow 0$, h — шаг по координате).

13. Рассматривается среда, находящаяся в начальный момент времени в жидком состоянии при температуре $T(x, 0) > T_p$ (T_p — температура плавления). Поверхность среды при $x = 0$ поддерживается при $T(0, t) < T_p$, и при $x = 1$: $T(1, t) > T_p$. В предположении, что плотность среды не изменится при фазовом превращении, процесс затвердевания описывается следующими уравнениями:

$$\begin{cases} \rho c_s \frac{\partial T}{\partial t} = a_s \frac{\partial^2 T}{\partial x^2}, & 0 \leq x \leq y(t), \\ \rho c_f \frac{\partial T}{\partial t} = a_f \frac{\partial^2 T}{\partial x^2}, & y(t) \leq x \leq 1, \end{cases} \quad (\text{П2.24})$$

где $y(t)$ — положение фазового фронта, индексы s и f относятся к твердой и жидкой фазам. (П2.24) дополняется начальными и граничными условиями, а также условиями на фазовом фронте:

$$\begin{cases} u(x, 0) = g(x), & 0 \leq x \leq 1; \\ u(0, t) = f_1(t); & u(1, t) = f_2(t); \\ a_s \frac{\partial T}{\partial x}(y-0) - a_f \frac{\partial T}{\partial x}(y+0) = \rho L \frac{\partial y}{\partial t} \end{cases} \quad (\text{П2.25})$$

(условие баланса энергии при движении фазового фронта).

Примем (вода–лед): $x = 0$ (поверхность водоема), $x = L = 1$ м (дно водоема);

$$g(x) = \frac{7x}{L} + 273 \text{ K};$$

$$f_1(t) = [273 - 13(1 - e^{-10t})] \text{ K};$$

$$f_2(t) = 280 \text{ K}; \quad l = 1 \text{ м}; \quad \rho = 10^3 \frac{\text{кг}}{\text{м}^3};$$

теплоемкость: вода — 4200, лед — 2100 Дж/(кг · К); коэффициент теплопроводности: вода — 0,56, лед — 2,25 Вт/(м · К); коэффициент температуропроводности: вода — $1,33 \cdot 10^{-7}$, лед — $1,08 \times 10^{-6} \text{ м}^2/\text{с}$; удельная теплота плавления: $3,3 \cdot 10^5$ Дж/кг, температура плавления 273 К.

- Рассчитать профили $T(x)$ в различные моменты времени; представить в виде графиков;
- рассчитать и представить в виде графика положение фронта фазового перехода;
- использовать три разностные схемы.

Исследуйте сходимость численного решения по сетке, представленной на рис. П2.4.

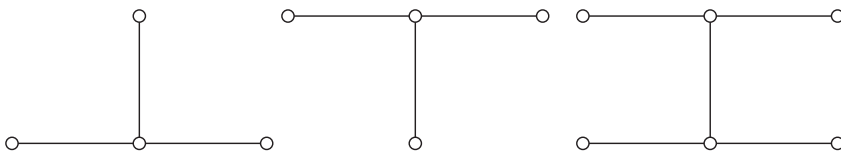


Рис. П2.4

14. Основное уравнение математической экологии — уравнение Бюргерса, описывающее перенос и диффузию загрязнений (в воде или воздухе); его линеаризованный вариант имеет вид:

$$\frac{\partial u}{\partial t} + c \frac{\partial u}{\partial x} = \mu \frac{\partial^2 u}{\partial x^2}, \quad (П2.26)$$

$$u(0, t) = U, \quad u(L, t) = 0, \quad u(x, 0) = 0.$$

Здесь u — концентрация некоторого вещества, μ — коэффициент диффузии, $\{t, x\}$ — независимые переменные, $c = \text{const}$ — постоянная скорость потока (например, реки).

- Предложить явную и неявную схемы для численного решения (П2.1) и получить численное решение;
- представить результаты в виде профилей $u(x)$ в различные моменты времени;

- исследовать поведение численного решения в зависимости от μ ($\mu = 1; 0,5; 0,01; 0,0001$) и c ($c = 1; 0,1; 0,01; 0,0001$), $L = 1$.
- исследовать схемы на сходимость по сетке, т.е. при $h \rightarrow 0$ ($x \in [0, 10]; t \in [0, T]$).

Точное нестационарное решение (П2.26) имеет вид (при $u(x, 0) = \sin kx$ и периодических граничных условиях):

$$u(x, t) = \exp(-k^2 \mu t) \sin k(x - ct). \quad (\text{П2.27})$$

Проверить (численно) формулу (П2.27).

15. Для описания распространения акустических волн в несжимаемой среде можно использовать так называемую акустическую систему:

$$\begin{aligned} \frac{\partial \mathbf{w}}{\partial t} + \mathbf{A} \frac{\partial \mathbf{w}}{\partial x} &= 0, \quad x \in [0, L], \quad t \in [0, T], \\ \mathbf{w} &= \begin{pmatrix} u \\ p \end{pmatrix}; \quad \mathbf{A} = \begin{Bmatrix} u & \rho^{-1} \\ \rho c^2 & 0 \end{Bmatrix}, \end{aligned} \quad (\text{П2.28})$$

или:

$$\begin{aligned} \frac{\partial u}{\partial t} + \frac{1}{\rho} \frac{\partial p}{\partial x} &= 0, \\ \frac{\partial p}{\partial t} + \rho c^2 \frac{\partial u}{\partial x} &= 0, \end{aligned} \quad (\text{П2.28a})$$

где u — скорость частиц среды, p — давление, ρ — плотность среды, c — скорость звука, \mathbf{A} — матрица 2×2 . Система (П2.28) дополняется начальными и граничными условиями:

$$\begin{aligned} u(x, 0) &= q_1(x), \\ p(x, 0) &= q_2(x); \\ \alpha_1 u(0, t) + \beta_1 p(0, t) &= f_1(t) \\ \alpha_2 u(1, t) + \beta_2 p(1, t) &= f_2(t) \end{aligned} \quad (\text{П2.29})$$

а) Введя разностную сетку с шагами τ, h используем для аппроксимации (П2.28), (П2.29) схему Лакса–Вендроффа ($\sigma = ch/\tau$; $T = n\tau$, $L = mh$):

$$\begin{cases} u_m^{n+1} = u_m^n - \frac{\sigma}{2\rho c} (p_{m+1}^n - p_{m-1}^n) + \frac{\sigma^2}{2} (u_{m+1}^n - 2u_m^n + u_{m-1}^n), \\ p_m^{n+1} = p_m^n - \frac{\sigma}{2} \rho c (u_{m+1}^n - u_{m-1}^n) + \frac{\sigma^2}{2} (p_{m+1}^n - 2p_m^n + p_{m-1}^n), \end{cases} \quad (\text{П2.30})$$

или в векторной форме:

$$\mathbf{w}^n = \mathbf{w}^n - \frac{\sigma}{2} \mathbf{A} (\mathbf{w}_{m+1}^n - \mathbf{w}_{m-1}^n) + \frac{\sigma^2}{2} \mathbf{A}^2 (\mathbf{w}_{m+1}^n - 2\mathbf{w}_m^n + \mathbf{w}_{m-1}^n). \quad (\text{П2.31})$$

б) *Инварианты Римана*. Умножим первое уравнение в (П2.28) на $c\rho$, сложим полученные уравнения и вычтем первое из второго, получим:

$$\begin{cases} \frac{\partial}{\partial t} (p + c\rho u) + c \frac{\partial}{\partial x} (p + c\rho u) = 0, \\ \frac{\partial}{\partial t} (p - c\rho u) - c \frac{\partial}{\partial x} (p - c\rho u) = 0, \end{cases} \quad (\text{П2.32})$$

или, в обозначениях $R = p + c\rho u$, $S = p - c\rho u$ (соответственно $u = \frac{R - S}{2\rho c}$, $p = \frac{R + S}{2}$):

$$\frac{\partial R}{\partial t} + c \frac{\partial S}{\partial x} = 0, \quad \frac{\partial S}{\partial t} - c \frac{\partial R}{\partial x} = 0; \quad (\text{П2.33})$$

величины R и S называются *инвариантами Римана*; (П2.33) — уравнение в инвариантах Римана. Решение (П2.33) можно записать в виде:

$$R(x, t) = R(x - ct), \quad S(x, t) = S(x + ct), \quad (\text{П2.34})$$

т. е. R и S сохраняются вдоль характеристик $dx/dt = c$ соответственно.

- Получить численные решения (П2.28) по схеме Лакса-Вендроффа.
- получить численное решение (П2.29) по схеме Рунге (или «кабаре»):

$$\begin{cases} \frac{1}{2} \left(\frac{R_m^{n+1} - R_m^n}{\tau} + \frac{R_{m-1}^n - R_{m-1}^{n-1}}{\tau} \right) + c \frac{R_m^n - R_{m-1}^n}{h} = 0, \\ \frac{1}{2} \left(\frac{S_m^{n+1} - S_m^n}{\tau} + \frac{S_{m+1}^n - S_{m+1}^{n-1}}{\tau} \right) - c \frac{S_{m+1}^n - S_m^n}{h} = 0, \\ R_m^n = p_m^n + c\rho \cdot u_m^n, \quad S_m^n = p_m^n - c\rho \cdot u_m^n, \\ u_m^n = \frac{R_m^n - S_m^n}{2\rho c}, \quad p_m^n = \frac{R_m^n + S_m^n}{2}, \end{cases}$$

предварительно получив соответствующие (П2.32) начальные и граничные условия.

- получить численное решение задачи распада разрыва

$$u(x, 0) = \begin{cases} u_1, & x \leq 0, \\ u_2, & x > 0, \end{cases} \quad u_2 > u_1 > 0,$$

$$p(x, 0) = \begin{cases} p_1, & x \leq 0, \\ p_2, & x > 0; \end{cases} \quad u_2(x, 0) = 2; \quad u_1(x, 0) = 1;$$

$$p_2(x, 0) = \rho c u_2, \quad p_1(x, 0) = \rho c u_1;$$

- сравнить эти решения, представив профили $u(x)$ и $p(x)$ в различные моменты t ;
- показать сходимость численных решений, полученных по обеим схемам, по сетке (т.е. при $\tau \rightarrow 0$);
- получить численное решение (П2.32) с помощью схемы Куранта–Изаксона–Риса; соответствующий шаблон представлен на рис. П2.5.

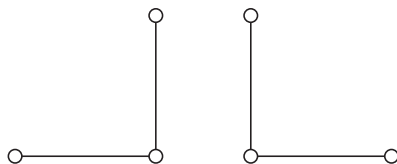


Рис. П2.5

16. Рассмотрим задачу о нагревании балки квадратного сечения, бесконечной по одной оси координат (Oz).

Пусть температура грани $ABCD$ (рис. П2.6) поддерживается постоянной: 1 на AB , 2 на BC , 3 на CD и 4 на DA (температура приведена в относительных единицах; $T_* = 100^\circ\text{C}$). Размер грани $L = 0,1$ м.

Получить численное решение стационарной задачи теплопроводности

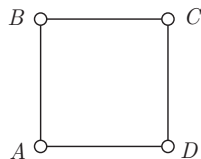


Рис. П2.6

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \quad (\text{П2.35})$$

с приведенными граничными условиями, используя итерационные методы:

а) Якоби:

$$\frac{u_{m-1,l}^{i+1} - 2u_{ml}^{i+1} + u_{m+1,l}^i}{h_x^2} + \frac{u_{m,l-1}^{i+1} - 2u_{ml}^{i+1} + u_{m,l+1}^i}{h_y^2} = f_{m,l}, \quad (\text{П2.36})$$

$m h_x = L$, $l h_y = L$ (h_x, h_y — шаги по координатам x, y);

б) Зейделя:

$$\frac{u_{m-1,l}^{i+1} - 2u_{m,l}^{i+1} + u_{m+1,l}^i}{h_x^2} + \frac{u_{m,l-1}^{i+1} - 2u_{m,l}^{i+1} + u_{m,l+1}^i}{h_y^2} = f_{m,l}; \quad (\text{П2.37})$$

в) верхней релаксации ($h_x = h_y$)

$$\begin{aligned} \frac{u_{m-1,l}^{i+1} + u_{m,l-1}^{i+1}}{h^2} + \frac{u_{m+1,l}^i + u_{m,l+1}^i}{h^2} = \\ = -\frac{4}{h^2} \left[\frac{u_{m,l}^i}{\tau} + \left(1 - \frac{1}{\tau}\right) u_{m,l}^i \right] = f_{ml}. \end{aligned} \quad (\text{П2.38})$$

- Сравнить эти методы по скорости сходимости (численно и теоретически);
- проверить сходимость численного решения по сетке (т. е. при $h_x, h_y \rightarrow 0$);
- исследовать численно скорость сходимости (П2.38) от величины итерационного параметра τ ;
- результаты численного решения представить в виде изолиний $T(x, y) = \text{const}$ и в виде одномерных графиков $T(x)$ при разных значениях y и $T(y)$ при разных значениях x .

17. Получить численное решение одномерных линейного и нелинейного уравнений переноса (в дивергентной и недивергентной формах):

$$\frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} = 0, \quad a = \text{const}; \quad (\text{П2.39})$$

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0, \quad (\text{П2.40})$$

$$\frac{\partial u}{\partial t} + \frac{\partial (u^2/2)}{\partial x} = 0; \quad (\text{П2.41})$$

$$t \in [0, T], \quad x \in [-X, X]; \quad (\text{П2.42})$$

$$u(0, x) = \begin{cases} u_1, & x > 0, \\ u_2, & x \leq 0; \end{cases}$$

$T = 100$; $X = 10$; $a = 1$; $N\tau = T$, $Mh = 2X$, $M = 10^3$, (τ, h — шаги по времени и по координате) с помощью разностных схем:

- Куранта–Изаксона–Риса;
- Лакса–Вендроффа;
- гибридной схемы Федоренко;
- Хартена (TVD);
- Колгана;
- ENO-схемы.

18. Получить численное решение одномерной задачи о распаде разрыва в идеальном газе, используя систему нестационарных уравнений газодинамики

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A} \frac{\partial \mathbf{U}}{\partial x} = 0 \quad (\text{П2.43})$$

$$\mathbf{U} = \{\rho, u, \varepsilon\}^T, \quad \mathbf{A} = \begin{Bmatrix} u & \rho & 0 \\ \frac{1}{\rho} \frac{\partial p}{\partial \rho} & u & \frac{1}{\rho} \frac{\partial p}{\partial \varepsilon} \\ 0 & p/\rho & u \end{Bmatrix},$$

где ρ — плотность, u — скорость газа, ε — удельная внутренняя энергия газа, $\{t, x\}$ — независимые координаты; $\rho(0, x) = \rho_0$, $u(0, x) = 0$, $\varepsilon(x, 0) = \varepsilon_0$; уравнение состояния:

$$p - \rho \varepsilon (\gamma - 1) = 0, \quad \gamma = 1,4;$$

$$\rho(0, x) = \begin{cases} \rho_1, & x \leq 0, \\ \rho_2, & x > 0, \end{cases}$$

$$t \in [0, T]; \quad x \in [-X, X]; \quad N_\tau = T,$$

$$Mh = 2X, \quad M = 20, 100, 1000; \quad h — \text{шаг по } x, \tau — \text{шаг по } t.$$

Использовать сеточно-характеристический метод ($\sigma = \tau/h$):

$$\mathbf{U}_m^{n+1} = \mathbf{U}_m^n - \sigma [(\mathbf{\Omega}^{-1} \mathbf{\Lambda}^+ \mathbf{\Omega})_m^n (\mathbf{U}_{m-1}^n - \mathbf{U}_m^n) - (\mathbf{\Omega}^{-1} \mathbf{\Lambda}^- \mathbf{\Omega})_m^n (\mathbf{U}_{m+1}^n - \mathbf{U}_m^n)]. \quad (\text{П2.44})$$

Здесь: $\mathbf{\Lambda}^\pm = (1/2)(\mathbf{\Lambda} + |\mathbf{\Lambda}|)$, $\mathbf{\Lambda}$ — диагональная матрица: $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \lambda_3)$. $\lambda_1 = u + c$, $\lambda_2 = u$, $\lambda_3 = u - c$ являются собственными числами матрицы \mathbf{A} ;

$$\mathbf{\Omega} = \begin{Bmatrix} \frac{\partial p}{\partial \rho} & \rho c & \frac{\partial p}{\partial \varepsilon} \\ p & 0 & -\rho^2 \\ p & -\rho c & \frac{\partial p}{\partial \rho} \end{Bmatrix} = \begin{Bmatrix} \boldsymbol{\omega}_1 \\ \boldsymbol{\omega}_2 \\ \boldsymbol{\omega}_3 \end{Bmatrix}$$

— матрица, строками которой являются соответствующие собственные векторы \mathbf{A} (причем $\mathbf{A} = \mathbf{\Omega}^{-1} \mathbf{\Lambda} \mathbf{\Omega}$), получаемые из соотношения $\boldsymbol{\omega}_i \mathbf{A} = \lambda_i \boldsymbol{\omega}_i$.

19. Одна из постановок задачи взаимодействия лазерного излучения с веществом имеет следующий вид (задача физики горения, u — температура):

$$\frac{\partial u}{\partial t} = \frac{1}{r} \cdot \frac{\partial}{\partial r} \left(r \frac{\partial u}{\partial r} \right) + \frac{\partial^2 u}{\partial z^2}, \quad r \geq 0, \quad z \geq 0; \quad (\text{П2.45})$$

$$-\frac{\partial u}{\partial z} = I(r) + e^{-1/u} - \eta u; \quad r \geq 0, \quad z \geq 0;$$

$$u(r, z, 0) = u_0(r, z) \geq 0,$$

$$u \rightarrow 0 \text{ при } \sqrt{r^2 + z^2} \rightarrow \infty.$$

$e^{-1/u}$ — описывает энерговыделение реакции на поверхности образца, $-\eta u$ — теплопотери, $I(r) = I_0 e^{(-r^2/r_0^2)}$, $r_0 = 2$ мм.

- Получить численное решение задачи (П2.45) с помощью локально-одномерной разностной схемы;
- исследовать распределение температуры u по r и z в различные моменты времени;
- показать сходимость решения по сетке (т.е. при $h_x \rightarrow 0$, $h_y \rightarrow 0$);
- получить численное решение (П2.45) при помощи явной разностной схемы. Какой шаг по времени необходимо для этого выбрать?
- Исследовать поведение рассматриваемой среды в зависимости от параметров η , I_0 , r_0 .

20. Уравнение, описывающее как конвективные, так и диффузионные процессы, называется *уравнением Бюргерса*:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = \mu \frac{\partial^2 u}{\partial x^2}, \quad \mu = \text{const} > 0. \quad (\text{П2.46})$$

Зададим начальные данные в следующем виде:

$$u(x, 0) = \begin{cases} u_1, & x < x_0, \\ u_2, & x > x_0, \end{cases} \quad u_1 > u_2, \quad x_0 = 0, \quad (\text{П2.47})$$

$u_1 < u_2$; t, x — независимые переменные (положим: $u_1 = 1$, $u_2 = 0$; $x \in [-10, 10]$, $t \in [0, T]$).

Точное решение (П2.46) имеет вид:

$$u(x, t) = u_2 + \frac{u_1 - u_2}{1 + g(x, t) \exp \left\{ \frac{u_1 - u_2}{2\mu} (x - x_0 - Dt) \right\}},$$

$$g(x, t) = \frac{\int_{-(x-x_0-u_2t)/\sqrt{4\mu t}}^{\infty} e^{-\xi^2} d\xi}{\int_{(x-x_0-u_1t)/\sqrt{4\mu t}}^{\infty} e^{-\xi^2} d\xi}, \quad (\text{П2.48})$$

$$D = \frac{u_1 + u_2}{2}.$$

- Получить численное решение (П2.46), (П2.47) при $\mu = 0 \div 1,0$. Есть ли что-нибудь общее во всех решениях при разных μ («центр сглаженных ударных волн»)?

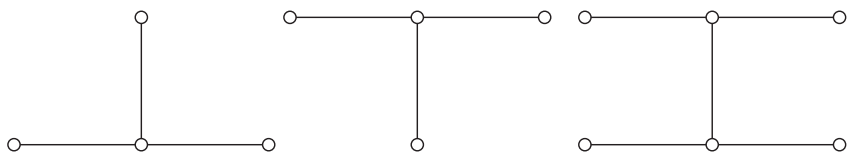


Рис. П2.7

- Использовать для численного решения (П2.46) схемы с шаблонами, приведенными на рис. П2.7;
- проверить сходимость численного решения по сетке (при $h \rightarrow 0$) и сравнить с (П2.48).

21. Для численного решения уравнения Кортевега–де Фриса (КдФ)

$$u_t + 6uu'_x + u'''_x = 0, \quad t > 0, \quad x \in [-10, 10] \quad (\text{П2.49})$$

(на границах области интегрирования ставятся условия периодичности) рассмотреть две разностные схемы (рис. П2.8 и П2.9) с шаблонами (вторая—аналог схемы Саульева для уравнения теплопроводности).

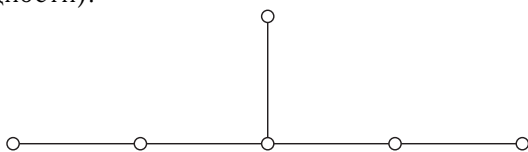


Рис. П2.8

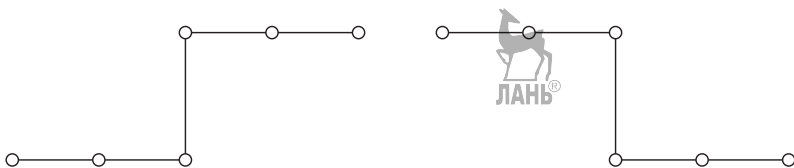


Рис. П2.9

- Сравнить численные решения, полученные по обеим схемам;
- показать сходимость по сетке (при $h \rightarrow 0$);
- рассмотреть начальные условия, представленные на рис. П2.10.

Как изменится численное решение, если к явлениям конвекции и дисперсии, описываемых уравнением КдФ, добавится диссипация

$$u_t - 6uu'_x + u''_x = \mu u''_x, \quad \mu = 10^{-4} \div 1 \quad (\text{П2.50})$$

(показать расчетом по одной из схем на рис. П2.8, П2.9).

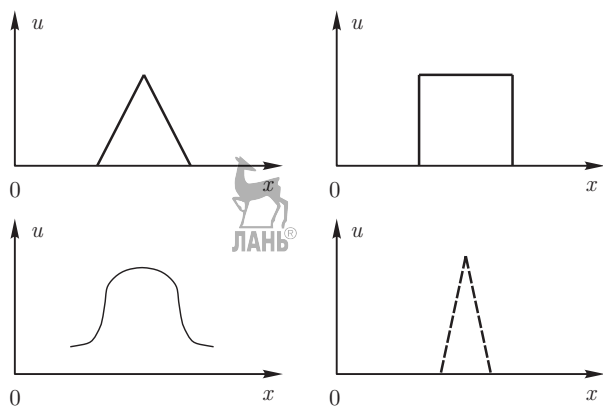


Рис. П2.10

Уравнение (П2.49) имеет бесконечное число законов сохранения:

$$\int_{\Omega} u dx = C_1, \quad (\text{П2.51})$$

$$\int_{\Omega} u^2 dx = C_2, \quad (\text{П2.52})$$

$$\int_{\Omega} \left(\frac{(u'_x)^2}{2} + u^3 \right) dx = C_3, \quad (\text{П2.53})$$

.....

Проверить любой из них.

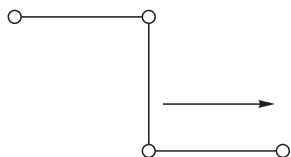


Рис. П2.11

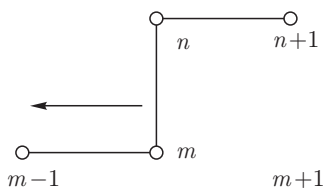


Рис. П2.12

Комментарий. Организация счета по схеме Саульева: на четных слоях счет идет слева направо (рис. П2.11) по формулам

$$\frac{u_m^{n+1} - u_m^n}{\tau} = \frac{1}{k^2} (u_{m-1}^{n+1} - u_m^{n+1} + u_m^n - u_{m+1}^n), \quad (\text{П2.54})$$

на нечетных — справа налево (рис. П2.12):

$$\frac{u_m^{n+1} - u_m^n}{\tau} = \frac{1}{k^2} (u_{m-1}^{n+1} - u_m^{n+1} - u_m^n + u_{m+1}^n). \quad (\text{П2.55})$$

22. Движение частицы заряда q и массы m в магнитном поле описывается системой ОДУ:

$$\begin{aligned}\dot{\mathbf{v}} &= \frac{q}{mc} \mathbf{v} \times \mathbf{B}, \quad \dot{\mathbf{r}} = \mathbf{v}, \\ \mathbf{v} &= v_x \mathbf{i} + v_y \mathbf{j} + v_z \mathbf{k}, \quad \mathbf{r} = x \mathbf{i} + y \mathbf{j} + z \mathbf{k}.\end{aligned}\quad (\text{П2.56})$$

Получить численное решение задачи об отражении заряженной частицы от магнитного зеркала. В этом случае:

$$\begin{aligned}B_x &= -\frac{x}{2} \cdot \frac{B_1 - B_0}{\pi l} \cdot \frac{1}{1 + \frac{(z-L)^2}{l^2}}, \\ B_y &= -\frac{y}{2} \cdot \frac{B_1 - B_0}{\pi l} \cdot \frac{1}{1 + \frac{(z-L)^2}{l^2}}, \\ B_z &= B_0 + (B_1 - B_0) \left(\frac{\pi}{2} + \operatorname{arctg} \frac{z-L}{l} \right) \cdot \frac{1}{\pi}.\end{aligned}$$

Начальные данные:

$$x(0) = \frac{v_1}{\omega_0}, \quad y(0) = 0, \quad z(0) = 0$$

($\omega_0 = qB_0/(mc)$ — ларморова частота),

$$v_x(0) = 0, \quad v_y(0) = -V_1, \quad v_z(0) = -V_1 \cdot \operatorname{ctg} \alpha,$$

$$\omega_0 = 1, \quad \frac{B_1}{B_0} = 2, \quad V_1 = 1, \quad l = 10, \quad L = 40, \quad \alpha \in \left[\frac{\pi}{4}; \frac{\pi}{2} \right].$$

Получить численное решение задачи о движении заряженной частицы в магнитной ловушке. В этом случае:

$$\begin{aligned}B_x &= -\frac{x}{2} \cdot \frac{B_1 - B_0}{2l} \cdot \pi \cdot \sin \frac{\pi z}{l}, \\ B_y &= -\frac{y}{2} \cdot \frac{B_1 - B_0}{2l} \cdot \pi \cdot \sin \frac{\pi z}{l}, \\ B_z &= B_0 + \frac{B_1 - B_0}{2} \left(1 - \cos \frac{\pi z}{l} \right).\end{aligned}$$

Начальные условия:

$$x(0) = \frac{V_1}{\omega_0}, \quad y(0) = 0, \quad z(0) = 0, \quad v_x(0) = 0, \quad v_y = -V_1,$$

$$v_z(0) = -V_1 \cdot \operatorname{ctg} \alpha, \quad \omega_0 = 1, \quad \frac{B_1}{B_0} = 2,$$

$$V_1 = 1, \quad l = 20, \quad \alpha \in \left[\frac{\pi}{4}; \frac{\pi}{2} \right].$$

- Использовать методы Рунге–Кутты 1-го и 4-го порядков точности;
- исследовать сходимость численных решений по сетке (при $\tau \rightarrow 0$, τ — шаг по времени).

23. Некоторые процессы в плазме, в биосистемах и в химических реакциях описываются нелинейным уравнением теплопроводности вида

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial x} \left[k(T) \frac{\partial T}{\partial x} \right] + Q(t), \quad (\text{П2.57})$$

где $T(t, x)$ — температура среды, $\{t, x\}$ — независимые переменные, $k(T)$ — нелинейный коэффициент теплопроводности, $Q(t)$ — нелинейная функция (например, моделирующая процессы горения, детонации); обычно: $k(T) = k_0 T^\alpha$; $Q(T) = q_0 T^\beta$; $k_0, q_0, \alpha > 0$; $\beta > 1$. При $\beta > \alpha + 1$ реализуется так называемый *LS*-режим с обострением, при $\beta < \alpha + 1$ — *HS*-режим с неограниченным ростом температуры, при $\beta = (\alpha + 1)$ — *S*-режим (полуширина профиля температуры постоянна). При $\beta > \alpha + 1$ полуширина профиля сокращается, процесс локализуется, формируется так называемая диссипативная структура, при $\beta < \alpha + 1$ наблюдаются тепловые волны, амплитуда которых растет. Профиль задается в виде, представленном на рис. П2.13.

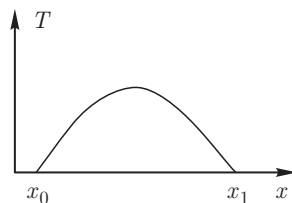


Рис. П2.13

- Проверить эти выводы численно, используя неявную схему. Положить:

$$\{k_0 = 1, \quad q_0 = 1, \quad q_0 = 1; \quad \beta = 3, \quad \alpha = 2\};$$

$$\{k_0 = 1, \quad q_0 = 1; \quad \beta = 3,18; \quad 1,667; \quad \alpha = 2\};$$

- проверить сходимость численных решений по сетке (т. е. при $h \rightarrow 0$, h — шаг по координате);
- вывести профили $T(x)$ в различные моменты времени.

24. Множество точек на фазовой плоскости, к которым стремится решение ОДУ, называется *аттрактором*. Представить численное решение следующих задач на фазовой плоскости и исследовать эти задачи на наличие аттракторов.

Аттрактор Лоренца:

$$\begin{cases} \dot{x} = \sigma y - \sigma x, & x(0) > 0, \quad y(0) = y_0, \quad z(0) = z_0 \\ & (y_0 = 1, z_0 = 1), \\ \dot{y} = -xz + rx - y, & \sigma = 10, \quad r = 28, \quad b = \{8/3; 10; 20\}; \\ \dot{z} = xy - bz, & t_k = 20; \quad \tau = 10^{-3}, \quad 10^{-2}. \end{cases} \quad (\text{П2.58})$$

(σ — число Прандтля, r — число Рэлея).

Аттрактор Реслера:

$$\begin{cases} \dot{x} = -y - z, & x(0) = 0, \quad y(0) = y_0, \quad z(0) = z_0, \\ \dot{y} = x + \frac{y}{5}, & \mu > 0, \quad 0 < \mu \leq 10, \\ \dot{z} = \frac{1}{5} + z(x - \mu). \end{cases} \quad (\text{П2.59})$$

Аттрактор Рикитаци:

$$\begin{cases} \dot{x} = -\mu x + yz, & \gamma_1 \in [0,002; 0,004], \\ \dot{y} = -\mu y + xr, & \gamma_2 = 0,002; \quad \mu \in [0,2; 2], \\ \dot{z} = 1 - xy - \gamma_1 z, & x(0) = 0, \quad y(0) = y_0, \quad z(0) = z_0, \\ \dot{r} = 1 - xy + \gamma_1 r, & r(0) = r(0). \end{cases} \quad (\text{П2.60})$$

Провести исследования свойств систем ОДУ (П2.61)–(П2.63) в зависимости от параметров процессов ($\sigma, r, b, \gamma_1, \gamma_2$):

- исследовать сходимость численного решения по сетке;
- использовать методы Рунге–Кутты первого и четвертого порядков точности.

25. Нелинейное уравнение теплопроводности способно описывать распространение тепловых волн, волн горения и т. п. Рассмотрим следующие уравнения.

а) Уравнение Колмогорова–Пискунова–Петракова (КПП):

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + au(1-u), \quad u(-\infty, t) = 1, \quad u(\infty, t) = 0, \quad (\text{П2.61})$$

$$\text{б) } u \in [0, 1; 10]; \quad k = \text{const} \neq 1, \quad u(x, 0) = \begin{cases} 1, & x \leq k, \\ 0, & x > k. \end{cases} \quad (\text{П2.62})$$

2. Уравнение Зельдовича–Франка–Каменецкого (задача горения):

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + au(u - \varepsilon)(1 - u), \quad 0 < \varepsilon \leq 1, \quad (\text{П2.63})$$

начальные и граничные условия — (П2.61), (П2.62).

- Получить численное решение (П2.61)–(П2.63) методом второго порядка точности;
- исследовать сходимость численного решения по сетке;
- построить профили $u(x)$ в различные моменты времени.

26. Нагревание пластины лазерным излучением описывается нестационарным двумерным уравнением теплопроводности:

$$c \frac{\partial T}{\partial t} = \frac{1}{r^N} \cdot \frac{\partial}{\partial r} \left(r^N k(T) \frac{\partial T}{\partial r} \right) + \frac{\partial}{\partial z} \left[k(T) \frac{\partial T}{\partial z} \right] + q, \quad (\text{П2.64})$$

где $T(t, r, z)$ — температура, c — коэффициент теплоемкости, $k(T)$ — коэффициент теплопроводности, $N = 0$ в плоской и $N = 1$ в цилиндрической геометрии.

а) Начальная температура:

$$T(0, r, z) = T_0 [1 + \cos(2\pi z) \cdot \cos(2\pi r)] + T_1, \\ T_0 = 100, \quad T_1 = 2; \quad k(T) = 1, \quad c = 1, \quad q = 1.$$

В этом случае точное решение имеет вид

$$T = T_0 \left[1 + e^{-8\pi^2 t} \cdot \cos(2\pi z) \cdot \cos(2\pi r) \right] + T_1. \quad (\text{П2.65})$$

б) $q = q_0 e^{-(z/z_0^z)^2}$, $t < \tau$; $q = 0$, $t > \tau$, $q_0 = 10^6$ Вт/см², $k_0 = 5$ мкм, $\tau = 100$ мкс, $T_0 = 300$ К; коэффициент поглощения возрастал от 0,05 для $T = T_0$ до 0,15 для $T = T_{\text{пл}}$ (температура плавления). Для железа $c = 4$ Дж/(см³ · К), $k_{\text{т}} = 0,8$ Вт/(см · К) — твердая фаза; $Q_{\text{пл}} = 2214$ Дж/см³, $k_{\text{ж}} = 0,4$ Вт/(см · К) — расплав. Теплота плавления $Q_{\text{пл}}$ учитывается добавлением к теплоемкости величины $Q_{\text{пл}}/(2 \cdot \Delta T_{\text{пл}})$ при $T_{\text{пл}} - \Delta T_{\text{пл}} < T < T_{\text{пл}} + \Delta T_{\text{пл}} + \Delta T_{\text{пл}}$ ($\Delta T_{\text{пл}} \approx 25 \div 50$ К). Зависимость $c_V(T)$ представлена на рис. П2.14.

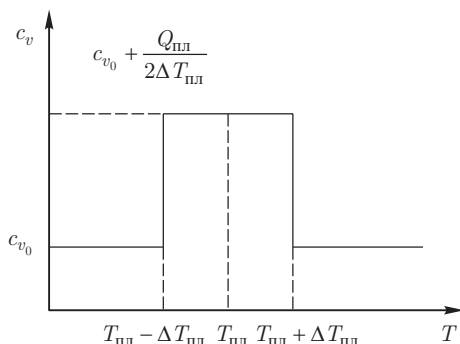


Рис. П2.14

Число частиц, испаренных с единицы поверхности:

$$N_e = \frac{2\pi m}{kT_{\text{пл}}} v_0^3 \alpha \exp \left[- \left(\frac{\lambda_1}{kT} + 1 \right) \right], \quad (\text{П2.66})$$

m — масса молекулы, k — постоянная Больцмана, λ_1 — энергия связи кристаллической решетки, v_0 — дебаевская частота (в качестве λ_1 выбирается работа выхода, соответствующая наиболее легко испаряемой компоненте; $\lambda_1 \approx 4,3$ эВ). $\alpha \in [0,1 \div 0,82]$ — учет обратного потока частиц.

- Получить численное решение (П2.64) с помощью явной и неявной схем;
- сопоставить решение п. а) с точным;
- проверить сходимость решения по сетке.

27. Движение частицы в центрально-симметричном поле с потенциалом $U(r)$ описывается уравнением Шрёдингера

$$\Delta \Psi + \frac{2\mu}{\hbar^2} [E - U(r)] \Psi = 0, \quad (\text{П2.67})$$

где Δ — оператор Лапласа в сферических координатах r, Θ, φ ; μ, \hbar — постоянные; решение ищется в виде

$$\Psi = Y_{\ell m}(\Theta, \varphi) \cdot R(r) r,$$

где $Y_{\ell m}$ — известная сферическая функция; ℓ, m — целые числа. Обозначив

$$\lambda = \frac{2\mu}{\hbar^2} E, \quad V(r) = \frac{2\mu}{\hbar^2} U(r) + \frac{\ell(\ell+1)}{r^2}, \quad (\text{П2.68})$$

получим задачу для определения $R(r)$:

$$R''_r - [V(r) - \lambda] R = 0, \quad r \in [0, \infty]; \quad R(0) = 0;$$

второе граничное условие — нормировки:

$$\int_0^\infty R^2(r) dz = 1. \quad (\text{П2.69})$$

- Получить численное решение задачи методом стрельбы (рис. П2.15). Обычно на бесконечности ставится условие $R(r^*, \lambda) = 0$, где r^* — достаточно большое число. Из этого уравнения методом Ньютона (или просто перебором) находим λ . Положим, что при $\lambda = \lambda_1$ имеем $R(r^*, \lambda) > 0$, при $\lambda = \lambda_2$ будет $R(r^*, \lambda) < 0$; тогда выбираем $\lambda = (\lambda_1 + \lambda_2)/2$ и т. д. Какие трудности встретятся при численной реализации метода стрельбы?

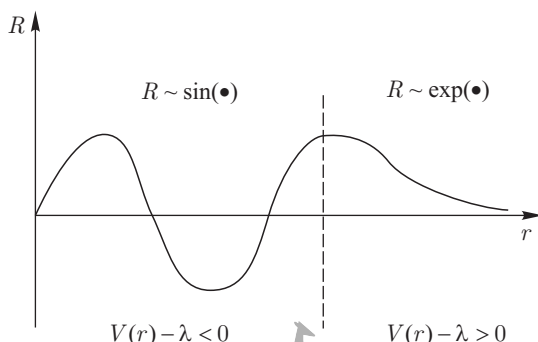


Рис. П2.15

- Получить численное решение задачи методом трехточечной прогонки;
- исследовать сходимость решения по сетке;
- получить численные решения для нескольких λ_i ($i = 1 \div 5$).

28. Для численного решения краевой задачи ОДУ

$$\frac{\partial^2 u}{\partial x^2} = f(u), \quad u(0) = U_1, \quad u(L) = U_2 \quad (\text{П2.70})$$

воспользоваться тремя вариантами трехточечной прогонки (предварительно получив прогоночные соотношения):

- а) прямая прогонка (слева направо);
- б) обратная (справа налево);
- в) встречные прогонки.

Положить:

$$f(u) = e^{\alpha u}; \quad f(u) = \sin(\omega u), \quad U_1 = 0, \quad U_2 = 1, \quad L = 1. \quad (\text{П2.71})$$

29. Пусть в (П2.70) краевые условия являются периодическими. Получить формулы для трехточечной периодической прогонки и численно решить (П2.70).

Исследовать поведение численного решения в зависимости от параметров α и ω в (П2.71).

30. Для аппроксимации краевой задачи

$$u_x^{\text{IV}} = f(x), \quad x \in [0, L], \quad u(0) = 0, \quad u(L) = 0, \quad u'(0) = 1, \quad u'(L) = 1, \quad f(x) = \text{const} \quad (\text{П2.72})$$

получить формулы пятиточечной прогонки и численно решить (П2.72), положив $L = 1$.

Учебное издание


ПЕТРОВ Игорь Борисович



ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА ДЛЯ ФИЗИКОВ

Редактор *В.С. Аролович*
Оригинал-макет: *В.В. Затекин*
Оформление переплета: *В.Ф. Киселёв*

Подписано в печать 08.02.2021. Формат 60×90/16. Бумага офсетная.
Печать офсетная. Усл. печ. л. 23,5. Уч.-изд. л. 25,8. Тираж 700 экз.
Заказ №



Издательская фирма «Физико-математическая литература»
МАИК «Наука/Интерпериодика»
117342, г. Москва, ул. Бутлерова, д. 17 Б
E-mail: porsova@fml.ru, sale@fml.ru
Сайт: <http://www.fml.ru>
Интернет-магазин: <http://www.fmlib.ru>

Отпечатано с электронных носителей издательства
в АО «Первая Образцовая типография»
Филиал «Чеховский Печатный Двор»
142300, Московская область, г. Чехов, ул. Полиграфистов, д. 1
Сайт: www.chpd.ru. E-mail: sales@chpd.ru, тел.: 8 (499) 270-73-59

ISBN 978-5-9221-1887-3



9 785922 118873