ОСНОВЫ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ

Н.Г. Бураго

Книга написана по материалам лекций по численным методам механики сплошной среды, читанных автором студентам 5-х курсов кафедры прикладной математики МГТУ им. Н.Э. Баумана, кафедры сопротивления материалов МГСУ-МИСИ, кафедры физики МАТИ. Целью лекций является освещение идей, лежащих в основе вычислительной механики.

Оглавление

Введение

Эта книга содержит сведения, которые могут быть интересны тем, кто хочет получить представление о методах и алгоритмах решения задач механики сплошной среды.

Предполагается, что читатель уже освоил курсы математики высшей школы.

Порядок изложения численных методов, принятый в книге, является общепринятым: интерполяция, численное интегрирование и дифференцирование, решение систем алгебраических уравнений, обыкновенных дифференциальных уравнений и, наконец, уравнений в частных производных.

Ключом к пониманию работы большинства имеющихся численных методов является теория проекционных методов. Поэтому краткий и очень простой пересказ этой теории приведен уже в самом начале книги для того, чтобы в дальнейшем при изложении методов активно пользоваться терминологией и результатами этой теории.

Основные методы интерполяции, численного интегрирования и дифференцирования описаны для случая многих независимых переменных.

Поскольку задачи численного анализа сводятся к системам алгебраических уравнений, то освещена тема о решении линейных и нелинейных алгебраических уравнений. Наряду с традиционными вариантами метода исключения описаны эффективные итерационные безматричные методы, приводящие к точному решению за конечное число операций.

В общей форме, применимой как к задачам в интегро-дифференциальной форме, так и к дискретизированным задачам, описаны основные методы решения нелинейных задач: ква-

зилинеаризация, дифференцирование (продолжение) по параметру и установление. Описаны приемы исследования вопросов существования, единственности и ветвления решений нелинейных уравнений в процессе их численного решения.

Так как системы алгебраических уравнений часто выражают условия экстремальности некоторых функционалов, то рассмотрены задачи и методы поиска экстремальных точек функционалов в соответствии с теорией математического программирования. Описаны основные методы безусловной минимизации функционалов и сведения задач условной минимизации к задачам безусловной минимизации.

В связи с эволюционными задачами рассмотрены основные методы численного решения задач Коши для систем обыкновенных дифференциальных уравнений.

Дальнейшее описание методов вычислительной механики привязано к иерархии постепенно усложняющихся постановок задач для уравнений в частных производных. Последовательно рассмотрены основные методы рещения классических многомерных эллиптических, параболических и гиперболических многомерных задач. Описаны основные варианты метода конечных разностей, метода конечных объемов, метода конечных элементов, метода граничных элементов, бессеточных методов. Показана важная роль свойств консервативности, согласованности (аппроксимации) и устойчивости для сходимости решений дискретных задач.

Основная трудность в отборе материала для данного курса состояла в том, чтобы НЕ начать излагать отдельные более знакомые автору вопросы особенно обстоятельно и не превращать курс в руководство при всем разделам вычислительной математики. Это нарушило бы баланс тем и убило бы саму затею создания ознакомительного курса численных методов. Поэтому к знатокам просьба не возмущаться, если изложение какого-либо вопроса покажется неполным или слишком поверхностным.

Особенностью принятого в книге стиля изложения является

минимальное использование математических формул. Во всех случаях по мере возможности предпочтение отдано словесному описанию численных методов, поскольку автор по собственному опыту знает, что за лесом формул очень часто теряется смысл и предназначение выкладок. Чтобы удержать интерес читателя к изложению автор сознательно упрощал изложение и сокращал объем книги за счет тех мест в изложении, написание которых вызывало скуку. Скучным, как правило, является ненужное, излагаемое "для порядка".

В Приложении приведены некоторые полезные справочные материалы: сведения из функционального анализа, запись основных операторов в абстрактной тензорной нотации, в криволинейных координатах, краткое толкование основных свойств разностных схем.

Книга написана по материалам лекций, читавшихся автором студентам 5-х курсов МГТУ им. Н. Э. Баумана, МГСУ-МИСИ и РГТУ-МАТИ на протяжении ряда лет.

Глава 1

Элементарное введение в численные методы

1.1 Балансное уравнение

В механике сплошных сред математическое описание явлений опирается на законы сохранения, представленные в виде так называемых балансных уравнений. В качестве примера типичного балансного уравнения рассмотрим уравнение закона сохранения тепла

$$\int\limits_{\tilde{V}} \rho \frac{dU}{dt} dV = -\int\limits_{\tilde{S}} \mathbf{q} \cdot \mathbf{n} dS + \int\limits_{\tilde{V}} \rho f_T dV \qquad (1)$$

где $\tilde{V}$ - произвольный объем, $\tilde{S}$ - поверхность объема $\tilde{V}$, ρ - плотность, U - удельная тепловая энергия единицы массы, t - время, $\mathbf{q}$ - диффузионный тепловой поток, $\mathbf{n}$ - внешняя единичная нормаль к поверхности, f_T - внешний массовый источник тепла, точка умножения означает скалярное умножение векторов. Из уравнения (1) видно, что тепло в произвольной области прирастает или убывает благодаря теплопроводности или диффузии тепла через границу (первое слагаемое правой части) и за счет распределенных источников или стоков тепла (второе слагаемое правой части). Напомним, что для жидкостей и газов тепло является мерой энергии хаотичного (броуновского) движения атомов и молекул, а для структурированных сред тепло является мерой энергии колебаний атомов и молекул около равновесных состояний.

С помощью теоремы Остроградского-Гаусса

$$\int\limits_{\tilde{S}} \mathbf{q} \cdot \mathbf{n} dS = \int\limits_{\tilde{V}} \nabla \cdot \mathbf{q} dV$$

преобразуем уравнение (1) к виду

$$\int\limits_{\tilde{V}} \rho \frac{dU}{dt} dV = -\int\limits_{\tilde{V}} \nabla \cdot \mathbf{q} dV + \int\limits_{\tilde{V}} \rho f_T dV$$

откуда в силу произвольности объема $\tilde{V}$ получаем дифференциальную форму закона сохранения тепла

$$\rho \frac{dU}{dt} = -\nabla \cdot \mathbf{q} + \rho f_T$$

где ∇ - оператор пространственного дифференцирования.

Используя калорическое уравнение

$$dU = c_V dT$$

и закон теплопроводности Фурье

$$\mathbf{q} = -k_T \nabla T$$

приводим (1) к дифференциальной форме, называемой уравнением теплопроводности

$$\rho c_V \frac{dT}{dt} = \nabla \cdot (k_T \nabla T) + \rho f_T$$

Общая задача теплопроводности формулируется так. В области $\{t \geq 0, \quad \mathbf{x} \in V\}$ требуется решить уравнение теплопроводности

$$\rho c_V dT/dt = \nabla \cdot (k_T \nabla T) + \rho f_T \qquad (2a)$$

при следующих граничных:

$$t \geq 0, \quad \mathbf{x} \in S_T : \; T = T_S(t, \mathbf{x}) \qquad (2b)$$

$$t \geq 0, \quad \mathbf{x} \in S \backslash S_T : \; -(k_T \nabla T) \cdot \mathbf{n} = q_S(t, \mathbf{x}) \qquad (2c)$$

и начальных условиях:

$$t = 0 \; , \quad \mathbf{x} \in V : \quad T = T^0(\mathbf{x}) \qquad (2d)$$

здесь $\mathbf{x} = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z$ - радиус вектор, $\nabla = \mathbf{e}_x \partial/\partial x + \mathbf{e}_y \partial/\partial y + \mathbf{e}_z \partial/\partial z$ - векторный оператор пространственного дифференцирования "набла", $\mathbf{e}_x$, $\mathbf{e}_y$, $\mathbf{e}_z$ - орты (единичные ортогональные базисные векторы) декартовой системы координат x, y, z. Для определения оператора "набла" достаточно указать его вид в какой-либо системе координат (в данном случае это сделано в декартовой системе координат), его представление в других системах координат устанавливается аккуратной заменой базисных векторов и пространственных переменных.

Формулировка задачи (2) справедлива для любой системы координат.

1.2 Конвективные потоки

Упорядоченное движение сплошной среды называется конвективным. Если материальная сплошная среда, обладающая тепловой энергией, в упорядоченном движении перетекает через границы элементарных объемов, то тепло, переносимое конвективным потоком, также учитывается в уравнении баланса тепловой энергии. Произвольные объемы, для которых формулировалась интегральная форма закона сохранения, могут иметь произвольно подвижные границы. При этом конвективные потоки характеризуют перетекание сплошной среды через границы этих объемов.

В записанном выше уравнении (1) конвективные потоки "спрятаны" в членах с материальными производными по времени d/dt. По определению материальное дифференцирование по времени производится вдоль траекторий бесконечно малых материальных объемов сплошной среды ("материальных частиц") и связано с

дифференцированием по времени $\partial/\partial t$ при фиксированных значениях пространственных переменных соотношением

$$d/dt = \partial/\partial t + (\mathbf{v} - \mathbf{w}) \cdot \nabla \tag{5}$$

здесь первое слагаемое в правой части описывает изменение искомой величины во времени при фиксированных значениях используемых пространственных независимых переменных (подвижных или неподвижных, любых), а второе слагаемое описывает изменение искомой величины во времени за счет конвективного движения, которое определяется разностью скоростей материальной среды ($\mathbf{v}$) и "координатной среды" ($\mathbf{w}$).

При $\mathbf{w} = \mathbf{v}$ конвективные члены зануляются и имеем лагранжевы координаты. При $\mathbf{w} = 0$ имеем эйлеровы координаты. В общем случае при $\mathbf{w} \neq \mathbf{v}$ имеем произвольно подвижные координаты.

Скорости $\mathbf{v}$ и $\mathbf{w}$ зависят от координат и времени, поэтому в общем случае одна и та же точка, определяемая фиксированными значениями независимых пространственных координат в различные моменты времени может быть лагранжевой, эйлеровой или произвольно подвижной.

1.3 Источниковый член

По поводу источникового члена f_T заметим, что он учитывает производство или поглощение тепла во внешних процессах, которые рассматриваются как заданные. В случае отдельно рассматриваемого закона сохранения тепла это может быть прирост или убыль тепла за счет деформирования и разрушения среды, внутреннего трения, турбулентности, химических и ядерных реакций, облучения. При явном рассмотрении некоторого дополнительного процесса в постановку задачи вводится соответствующая данному процессу зависимая переменная или величина, определяющее соотношение для диффузионного потока этой величины и

соответствующее балансное уравнение. При этом тепловой источниковый член от такого теперь явно рассматриваемого процесса выделяется из "копилки" f_T и записывается в балансе тепла явно и отдельно, а член f_T по-прежнему обозначает оставшиеся источники тепла от внешних процессов.

Например, если в дополнение к балансу тепла рассматриваются деформации вязкой жидкости, то вводится новая зависимая переменная - вектор скорости материальной среды, записывается соответствуюшее балансное соотношение для импульса - уравнение количества движения - и определяющее соотношение - закон связи вязких напряжений с градиентами скорости. При этом в уравнении баланса тепла явно учитывается источник тепла от вязкого трения, равный скорости работы вязких напряжений. Источниковый член f_T теперь не содержит составляющих, связанных с внутренним вязким трением.

Число балансных уравнений вида (1) в постановке задачи может исчисляться сотнями. Например, в задачах многокомпонентной гидродинамики при описании протекающих химических реакций для каждой реагирующей компоненты вводится ее концентрация и соответствующее балансное уравнение.

1.4 Примеры балансных уравнений.

Рассмотрим упомянутый выше пример расширенной постановки задачи для совместного учета распространения тепла в движущейся вязкой эидкости. В этом случае в постановке задачи участвуют балансные уравнения для массы, импульса и энергии

$$d\rho/dt = \nabla \cdot \mathbf{q}_\rho - \rho \nabla \cdot \mathbf{v} \tag{6}$$

$$\rho d\mathbf{v}/dt = \nabla \cdot \sigma + \rho \mathbf{g} \tag{7}$$

$$\rho c_V dT/dt = \nabla \cdot \mathbf{q} + \sigma : \nabla \mathbf{v} + \rho f_T \tag{8}$$

Они дополняются определяющими соотношениями для диффузионных потоков массы, импульса и тепла

$$\mathbf{q}_\rho = \nu_\rho \nabla \rho$$

$$\sigma = -p(\rho, T)\mathbf{I} + \lambda_v(\nabla \cdot \mathbf{v})\mathbf{I} + \mu_v(\nabla \mathbf{v} + (\nabla \mathbf{v})^T)$$

$$\mathbf{q} = k_T \nabla T$$

где ν_ρ, λ_v, μ_v, k_T - коэффициенты диффузии, $p(\rho, T)$ - давление, $\rho\mathbf{g}$ - массовые силы (объемный источник импульса).

В балансном соотношении для массы (6) по "принципу равноприсутствия"дописан член самодиффузии плотности (первое слагаемое в правой части), которым обычно пренебрегают.

Надо заметить, что самодиффузия плотности определяется из опытов с использованием изотопов, но для большинства случаев коэффициенты самодиффузии оказываются пренебрежимо малыми. Заметным влияние самодиффузии может быть для разреженных сред.

Далее при анализе численных методов будет показано, что при численном решении член самодиффузии появляется в уравнении баланса массы явно в виде искусственной вязкости или неявно в виде аппроксимационной вязкости и служит регуляризатором, то есть дополнительным членом, улучшающим численное решение.

1.5 Вариационная формулировка задачи

Вернемся к исходной задаче (1)-(4). При решении этой задачи методом конечных элементов используется вариационная галеркинская формулировка.

Для этого вводится понятие множества пробных функций, в котором разыскивается решение. Для исходной формулировки задачи элементами такого множества являются дважды дифференцируемые функции, определенные в области решения и

удовлетворяющие главным граничным условиям (2), задающим значения искомой функции на участке границы S_T.

Далее вводится понятие вариации решения как разности двух произвольных элементов T_1^* и T_2^* множества пробных функций:

$$\delta T = T_1^* - T_2^* \tag{9}$$

для которых очевидно выполнено однородное граничное условие

$$x = x_e \ : \quad \delta T = 0 \tag{10}$$

или, в многомерном случае

$$x \in S_T \ : \quad \delta T = 0 \tag{10a}$$

В основе вариационной формулировки лежит следующее простое соображение: если для любого a произведение ab равно нулю, то b равно нулю и наоборот. Умножим уравнение (1) на произвольную вариацию искомой функции δT и проинтегрируем результат по пространственной области решения, получим

$$\int\limits_{x_b}^{x_e} c_V \frac{dT}{dt} \delta T dx = \int\limits_{x_b}^{x_e} \frac{\partial}{\partial x} \left(k_T \frac{\partial T}{\partial x} \right) \delta T dx + \int\limits_{x_b}^{x_e} f_T \delta T dx$$

Используя правило дифференцирования произведения функций, преобразуем подынтегральное выражение первого слагаемого правой части

$$\frac{\partial}{\partial x} \left(k_T \frac{\partial T}{\partial x} \right) \delta T = \frac{\partial}{\partial x} \left(k_T \frac{\partial T}{\partial x} \delta T \right) - k_T \frac{\partial T}{\partial x} \frac{\partial \delta T}{\partial x}$$

тогда после подстановки получим наше уравнение в следующем виде

$$\int\limits_{x_b}^{x_e} \left(c_V \frac{dT}{dt} \delta T + k_T \frac{\partial T}{\partial x} \frac{\partial \delta T}{\partial x} - f_T \delta T \right) dx = \left(k_T \frac{\partial T}{\partial x} \delta T \right) \Bigg|_{x_b}^{x_e}$$

откуда после учета граничных условий (2), (3) и (10) следует искомое вариационное уравнение:

$$\int_{x_b}^{x_e} \left(c_V \frac{dT}{dt} \delta T + k_T \frac{\partial T}{\partial x} \frac{\partial \delta T}{\partial x} - f_T \delta T \right) dx = q_S \delta T(x_e) \qquad (11)$$

дополняемое граничным условием

$$x = x_b \ : \quad T = T_S(t) \qquad (12)$$

и начальным условием

$$t = 0 \ : \quad T = T^0(x) \qquad (13)$$

В многомерном случае вариационная формулировка задачи имеет вид

$$\int_V [c_V \frac{dT}{dt} \delta T + k_T \nabla T \cdot \nabla \delta T - f_T \delta T] dV = \int_{S \setminus S_T} q_N \delta T dS \qquad (11a)$$

$$\mathbf{x} \in S_T \ : \quad T = T_S(t, \mathbf{x}) \qquad (12a)$$

$$t = 0 \ : \quad T = T^0(t, \mathbf{x}) \qquad (13a)$$

Заметим, что вариационную формулировку несложно получить и с использованием интегрирования по пространственно-временной области решения, что иногда делается. Большого смысла в этом, однако, нет, так как это не дает никакой выгоды из-за того, что решение по времени выстраивается временными слоями, отвечающими последовательно возрастающим моментам времени, а шаг по времени, как правило, является единым для всей пространственной области решения.

1.6 Следствия вариационного уравнения

Если теперь подвергнуть вариационное уравнение обратному преобразованию, то есть, используя интегрирование по частям,

"перебросить" дифференцирование с вариации искомой функции на ее поток:

$$k_T \nabla T \cdot \nabla \delta T = \nabla \cdot (k_T \nabla T \delta T) - \nabla \cdot (k_T \nabla T) \delta T$$

то вариационное уравнение примет вид:

$$\int_V (c_V dT/dt - \nabla \cdot (k_T \nabla T) - f_T) \delta T dV = \int_{S \setminus S_T} (k_T \nabla T - q_S) \delta T dS$$

откуда по основной лемме вариационного исчисления (лемма Эйлера) следуют уравнение Эйлера (исходное дифференциальное уравнение задачи (1a)) и граничные условия для потока (3a), называемые, именно поэтому, естественными граничными условиями. Граничные условия для искомой функции (2a) называют главными.

1.7 Метод множителей Лагранжа

В некоторых вариационных формулировках и методах решения (в МКЭ тоже, в частности) главные граничные условия учитываются в вариационном уравнении с помощью метода штрафных функций или метода множителей Лагранжа. При использовании метода множителей Лагранжа вариационная формулировка задачи принимает вид:

$$\int_V [c_V \frac{dT}{dt} \delta T + k_T \nabla T \cdot \nabla \delta T - f_T \delta T] dV =$$

$$= \int_{S_T} (\delta q_L (T - T_S)) + q_L \delta T) dS + \int_{S \setminus S_T} q_N \delta T dS \qquad (11b)$$

$$t = 0 \ : \quad T = T^0(t, \mathbf{x}) \qquad (13b)$$

где q_L и δq_L - множитель Лагранжа и его вариация. Множитель Лагранжа представляет собой дополнительную искомую функцию, которая вводится в вариационное уравнение так, чтобы множителем при ее произвольной вариации являлось главное граничное условие. Нетрудно проверить, как это было сделано в предыдущем разделе, что главное граничное условие будет следствием модифицированного вариационного уравнения.

С физической точки зрения введенный выше множитель Лагранжа является диффузионным потоком через границу, на которой заданы главные граничные условия. Обратим внимание на то, что интеграл в члене с множителями Лагранжа берется по участку границы с заданными главными граничными условиями.

1.8 Метод штрафных функций

В методе штрафных функций для искомого граничного диффузионного потока q_L принимается следующее выражение $q_L = -\lambda_*(T - T_S)$ с большим коэффициентом штрафа ($\lambda_* >> 1$). Чем больше коэффициент штрафа, тем точнее удовлетворяется главное граничное условие (2a). Для метода штрафных функций вариационная формулировка принимает вид

$$\int_V [c_V \frac{dT}{dt}\delta T + k_T \nabla T \cdot \nabla \delta T - f_T \delta T] dV =$$

$$= -\int_{S_T} 2\lambda_*(T - T_S)\delta T dS + \int_{S \setminus S_T} q_N \delta T dS \qquad (11c)$$

$$t = 0 \ : \quad T = T^0(t, \mathbf{x}) \qquad (13c)$$

При реализации метода штрафных функций значение коэффициента штрафа подбирается эмпирически (методом проб и

ошибок). Коэффициент штрафа должен быть достаточно большим, чтобы главное граничное условие выполнялось с достаточной точностью. Но одновременно он не должен быть слишком большим, так как в этом случае приближенное решение может быть испорчено ошибками в решении системы дискретных уравнений из-за ухудшения ее обусловленности.

1.9 Задача для скалярного балансного уравнения

Рассмотрим численное решение одномерной задачи теплопроводности в области $\{x_b \le x \le x_e, t \ge 0\}$:

$$\rho c_V \frac{dT}{dt} = \frac{\partial}{\partial x}\left(k_T(t,x,T)\frac{\partial T}{\partial x}\right) + \rho f_T(t,x,T) \tag{1}$$

$$t \ge 0, x = x_b : \; T = T_S(t) \tag{2}$$

$$t \ge 0, x = x_e : \; -k_T \frac{\partial T}{\partial x} = q_S(t,T) \tag{3}$$

$$t = 0, x_b \le x \le x_e : \; T = T^0(x) \tag{4}$$

где x - пространственная независимая переменная, t - время, T - искомая температура, c_V - коэффициент теплоемкости при постоянном объеме, k_T - коэффициент теплопроводности, f_T - заданная функция источника тепла, T_S - заданная граничная температура, q_S - заданная функция граничного теплового потока.

Требуется найти решение уравнения (1) с граничными (2), (3) и начальным (4) условиями.

1.10 Конечно-элементная сетка

При решении задач методом конечных элементов пространственная область решения разбивается на маленькие подобласти простой формы, представляющие собой в простейших случаях

двухузловые отрезки в одномерных задачах, трехузловые треугольники в двумерных задачах и четырехузловые тетраэдры в трехмерных задачах. Эти маленькие подобласти называют ячейками или конечными элементами. Предполагается, что состоящая из таких ячеек сетка воспроизводит геометрию области решения, не содержит наложенных друг на друга ячеек, ячеек неположительного объема и пустот.

В реальных прикладных расчетах геометрия областей решения описывается с помощью CAD-программ (Computer Aided Design codes) и используеться в виде CAD-файлов для задания исходных данных для программ генерации сеток (сеточных генераторов). Если возможности этих программ устраивают, то их используют как "черные ящики", в противном случае эти программы разрабатываются специально.

1.11 Внутренние элементы.

В рассматриваемой одномерной задаче введем разбиение области решения $[x_b, x_e]$ на два конечных элемента $[x_1, x_2]$ и $[x_2, x_3]$, где $x_1 = x_b$, $x_3 = x_e$ и $x_b < x_2 < x_e$. Таким образом, число узлов конечно-элементной сетки N_1 равно 3, число элементов N_2 равно 2, Число узлов в элементе M равно 2. Полагается, что массив координат узлов сетки $\mathbf{x}_i$, $(i = 1, ..., N_1)$ задан. Каждая ячейка сетки определена массивом номеров принадлежащих этой ячейке узлов $J1(i, j)$, $i = 1, ..., N_2$; $j = 1, ..., M$. Здесь i нумерует элементы, j локально нумерует узлы в элементах. Локальному номеру узла j в элементе i отвечает глобальный номер этого узла $J1(i, j)$. Информационный массив $J1$ в нашем случае имеет вид:

$$J1(1,1) = 1 \ , \quad J1(1,2) = 2 \ , \quad J1(2,1) = 2 \ , \quad J1(2,2) = 3$$

1.12 Граничные элементы

Аналогично, для геометрического описания границы вводятся граничные элементы. В одномерном случае граница представлена точками $x = x_b$ и $x = x_e$, поэтому надобности в граничных элементах нет. В двумерном случае граничными элементами служат отрезки, на которые разбивается контур границы. В трехмерном случае граничные элементы представлены поверхностными треугольными ячейками. Граничные элементы заданы номерами образующих узлов, которые хранятся в целочисленном массиве $J2(i,j)$, $\quad i = 1, ..., N_3;\ j = 1, ..., M_g$, где N_3 - число граничных элементов, M_g - число узлов в граничном элементе. Локальному номеру узла j в граничном элементе i отвечает глобальный номер этого узла $J2(i,j)$.

1.13 Конечно-элементная аппроксимация

По времени обычно вводится конечно-разностная аппроксимация решения и решение ищется последовательно для возрастающих моментов времени t^n , $\quad n = 1, 2, ...$. Величина $\Delta t_n = t_n - t_{n-1}$ называется шагом по времени и может быть переменной. Множество узлов, отвечающих фиксированному моменту времени t^n, образует временной слой n. Решение на слое $n = 0$ задано начальными условиями. Для узловых значений искомой функции используется обозначение T_i^n, где i - пространственный номер узла, n - номер временного слоя.

В пределах ячейки (конечного элемента) решение определяется по его значениям в узлах интерполяцией. В простейших случаях используется линейная интерполяция. Например, в одномерном конечном элементе $[x_1, x_2]$ нашей задачи решение ищется в виде

$$x \in [x_1, x_2] \ : \quad T^n(x) = d_{11}^{(0)} T^n_{J1(1,1)} + d_{12}^{(0)} T^n_{J1(1,2)} \qquad (14)$$

а в одномерном конечном элементе $[x_2, x_3]$

$$x \in [x_2, x_3] \ : \quad T^n(x) = d_{21}^{(0)}(x) T^n_{J1(2,1)} + d_{22}^{(0)}(x) T^n_{J1(2,2)} \qquad (15)$$

где

$$d_{11}^{(0)} = \frac{x_2 - x}{x_2 - x_1} \ , \quad d_{12}^{(0)} = \frac{x - x_1}{x_2 - x_1} \ , \quad d_{21}^{(0)} = \frac{x_3 - x}{x_3 - x_2} \ , \quad d_{22}^{(0)} = \frac{x - x_2}{x_3 - x_2}$$

Введенное с помошью кусочно-линейной интерполяции приближенное решение непрерывно в области решения. Первые производные от приближенного решения на границе между конечными элементами имеют конечный разрыв. Вторые производные от приближеного решения, вычисленные дифференцированием интерполянтов, тождественно равны нулю.

Для вариационной формулировки задачи (11)-(13) требования к гладкости искомого решения по сравнению с исходной формулировкой (1)-(4) ослаблены, существование вторых производных не требуется, так как они в формулировке задачи не участвуют. Достаточно, чтобы решение имело интегрируемые первые производные. Поэтому вариационная формулировка называется слабой. Ясно, что пониженные требования к гладкости решения облегчают конструирование численных алгоритмов решения, так как допускают более простые методы интерполяции решения.

Аналогично вводится кусочно-линейная интерполяция в двумерном и трехмерном случаях

$$\mathbf{x} \in V_i \ : \quad T^n(\mathbf{x}) = \sum_{j=1}^{M} d_{ij}^{(0)}(\mathbf{x}) T^n_{J1(i,j)} \qquad (16)$$

где V_i - конечный элемент номер i, $d_{ij}^{(0)}(\mathbf{x})$ - функции формы в элементе i для узла j, равная единице в этом узле и равная нулю в остальных узлах, j - локальный номер узла в элементе, $J1(i,j)$ - глобальный номер узла j.

Более подробное описание дано в главе про интерполяцию.

1.14 Определение производных

Формулы для определения производных имеют вид:

$$\mathbf{x} \in V_i \ : \quad \nabla T^n(\mathbf{x}) = \sum_{j=1}^{M} \mathbf{d}_{ij}^{(x)}(\mathbf{x}) T^n_{J1(i,j)} \tag{17}$$

где

$$\mathbf{d}_{ij}^{(x)}(\mathbf{x}) = \nabla d_{ij}^{(0)}(\mathbf{x}) \tag{18}$$

Поскольку производные от константы равны нулю, то коэффициенты для вычисления производных обладают важным свойством

$$\sum_{j=1}^{M} \mathbf{d}_{ij}^{(x)}(\mathbf{x}) = 0 \tag{19}$$

которое обеспечивает консервативность конечно-элементной дискретизации вариационного уравнения (отсутствие вычислительных источников и стоков).

В рассматриваемой одномерной задаче коэффициенты дифференцирования равны

$$x \in [x_1, x_2] \ : \quad d_{11}^{(x)} = \frac{-1}{x_2 - x_1} \ , \quad d_{12}^{(x)} = \frac{1}{x_2 - x_1} \tag{20}$$

$$x \in [x_2, x_3] \ : \quad d_{21}^{(x)} = \frac{-1}{x_3 - x_2} \ . \quad d_{22}^{(x)} = \frac{1}{x_3 - x_2} \tag{21}$$

1.15 Аппроксимация вариационного уравнения

Представим пространственные интегралы в вариационном уравнении суммой интегралов по конечныи элементам. Аппроксимацию решения по времени пока вводить не будем, полагая, что искомые узловые значения решения являются функциями времени $T_n(t)$. Хотя в пределах конечных элементов во многих случаях подынтегральные выражения являются полиномами невысоких

степеней, интегрирование по отдельному конечному элементу выполняют численно, так как в общем случае такое интегрирование аналитически выполнить невозможно из-за того, что коэффициенты уравнений могут не иметь аналитического представления и нередко определяются алгоритмически. Алгоритмическое определение означает, что задается алгоритм (набор математических операций) вычисления коэффициентов по значениям аргументов. Для численного интегрирования применяют квадратурные формулы.

В соответствии с теоремой сходимости приближенных решений МКЭ минимально необходимая точность численного интегрирования определяется требованием точного интегрирования объема конечного элемента и входящих в вариационное уравнение производных от решения. Подробное теоретическое обсуждение этих требований можно найти в книге Стренга и Фикса по теории метода конечных элементов (1979).

В нашем случае для численного интегрирования применяются простейшие квадратурные формулы прямоугольников, подынтегральное выражение вычисляется в центре элемента и умножается на объем конечного элемента (в нашем одномерном случае - на его длину). Для нестационарного члена, содержащего производные по времени, применяются квадратурные формулы с точками численного интегрирования в узлах конечноэлементой сетки (формула трапеций), что приводит к дискретным уравнениям, разрешенным относительно производных от решения по времени.

Вариации решения δT, так же как и решение, представляются при дискретизации их узловыми значениями. Для вариаций также используется кусочно-линейная аппроксимация вида (16)-(17).

Дискретизированное вариационное уравнение имеет вид

$$(c_V \frac{dT_1}{dt} - f_T)\delta T_1 \frac{x_2 - x_1}{2} + (c_V \frac{dT_2}{dt} - f_T)\delta T_2 \frac{x_3 - x_1}{2} +$$

$$+(c_V\frac{dT_3}{dt}-f_T)\delta T_3\frac{x_3-x_2}{2}+k_T\frac{T_2-T_1}{x_2-x_1}\frac{\delta T_2-\delta T_1}{x_2-x_1}(x_2-x_1)+$$

$$+k_T\frac{T_3-T_2}{x_3-x_2}\frac{\delta T_3-\delta T_2}{x_3-x_2}(x_3-x_2)=q_S\delta T_3 \qquad (22)$$

и дополняется главным граничным условием (12)

$$x=x_1\ :\quad T_1(t)=T_S(t) \qquad (23)$$

и начальным условием (13)

$$T_j(0)=T_j^*\ (j=1,2,3)$$

1.16 Матричный МКЭ

Обычно следующий шаг в построении численного алгоритма МКЭ состоит в приведении дискретизированного вариационного уравнения (22) к виду

$$\sum_{i=1}^{N_1}(\sum_{j=1}^{N_1}M_{ij}dT_j/dt+\sum_{j=1}^{N_1}K_{ij}T_j-F_i)\delta T_i=0 \qquad (24)$$

где M_{ij} - "матрица масс", K_{ij} - "матрица жесткости", F_i - "вектор внешних сил". Термины, взятые в кавычки, пришли из тех времен, когда первоначально метод конечных элементов применялся для расчета деформируемых стержневых конструкций. Термины взяты в кавычки, так как теперь, когда МКЭ стал общим методом, использование этих терминов является данью традиции и в контексте рассматриваемой задачи эти традиционные термины звучат странно, так как ни масса, ни жесткость, ни силы в задаче не фигурируют. Тем не менее эти термины нередко используются, причем кавычки не употребляются.

Поскольку дискретные вариации решения могут принимать произвольные значения, а вариационное выражение левой части

(24) при этом остается равным нулю, то это значит, что множители при дискретных вариациях равны нулю

$$\sum_{j=1}^{N_1} M_{ij} dT_j/dt + \sum_{j=1}^{N_1} K_{ij} T_j - F_i = 0 \ , \quad (i = 1, ..., N) \qquad (25)$$

Соотношения (25) совместно с главными и начальными условиями образуют систему уравнений метода конечных элементов. Решение $T_j(t)$ определяется последовательным интегрированием данных уравнений по времени. Заметим, что в задачах, для которых решение, коэффициенты и свободные члены не зависят от времени, члены с производными по времени зануляются. Для лагранжева описания движения полагается $dT/dt = 0$, конвекция отсутствует и такие задачи называются статическими. Для нелагранжева описания движения конвективные члены сохраняются, зануляются производные $\partial T/\partial t$, а задачи называются стационарными.

Для рассматриваемого примера в вариационном уравнении (22) приведем подобные члены при дискретных вариациях решения, тогда вариационное уравнение (22) примет вид

$$\delta T_1((c_V \frac{dT_1}{dt} - f_T)\frac{x_2 - x_1}{2} - k_T \frac{T_2 - T_1}{x_2 - x_1}) +$$
$$+\delta T_2((c_V \frac{dT_2}{dt} - f_T)\frac{x_3 - x_1}{2} + k_T \frac{T_2 - T_1}{x_2 - x_1} - k_T \frac{T_3 - T_2}{x_3 - x_2}) +$$
$$+\delta T_3((c_V \frac{dT_3}{dt} - f_T)\frac{x_3 - x_2}{2} + k_T \frac{T_3 - T_2}{x_3 - x_2} - q_S) = 0 \qquad (26)$$

Отсюда в силу произвольности дискретных вариаций решения δT_i следует, что суммы подобных членов в скобках, являющиеся множителями при этих произвольных вариациаях, равны нулю. Это и дает искомую систему уравнений МКЭ.

Заметим, что в узлах, принадлежащих участкам границы с заданными главными граничными условиями (с заданной искомой функцией), дискретные вариации решения не произвольны,

а тождественно равны нулю. Поэтому на таких участках границы уравнениями служат сами главные граничные условия. Поэтому в нашем случае выражение в скобках в первом слагаемом уравнения (26) нулю не равно и его надо заменить главным граничным условием (23). Если этого не сделать, то матрица жесткости будет вырожденной. Вырождение матрицы жесткости как раз и устраняется наложением главных граничных условий. В нестационарных задачах главные граничные условия могут отсутствовать без ущерба для корректности начально-краевой задачи..

С учетом главных граничных условий система дискретных уравнений (26) принимает вид

$$T_1 = T_S(t)$$

$$(c_V\frac{dT_2}{dt} - f_T)\frac{x_3 - x_1}{2} + k_T\frac{T_2 - T_S(t)}{x_2 - x_1} - k_T\frac{T_3 - T_2}{x_3 - x_2} = 0$$

$$(c_V\frac{dT_3}{dt} - f_T)\frac{x_3 - x_2}{2} + k_T\frac{T_3 - T_2}{x_3 - x_2} - q_S(t) = 0 \qquad (27)$$

При записи этой системы уравнений в виде (25) коэффициенты M_{ij}, K_{ij} и F_i имеют вид:

$$M_{11} = 0, \quad M_{22} = c_V\frac{x_3 - x_1}{2}\ , \quad M_{33} = c_V\frac{x_3 - x_2}{2}$$

$$M_{ij} = 0,\ (i, j = 1, 2, 3\ .\quad i \neq j)$$

$$K_{11} = 1\ , \quad K_{12} = 0\ , \quad K_{13} = 0$$

$$K_{21} = 0\ ,\ K_{22} = \frac{k_T}{x_2 - x_1} + \frac{k_T}{x_3 - x_2}\ ,\ K_{23} = -\frac{k_T}{x_3 - x_2}$$

$$K_{31} = 0\ , \quad K_{32} = -\frac{k_T}{x_3 - x_2}\ , \quad K_{33} = \frac{k_T}{x_3 - x_2}\ , \quad F_1 = T_S(t^n)$$

$$F_2 = f_T\frac{x_3 - x_1}{2} + \frac{k_T T_S(t)}{x_3 - x_2}, \quad F_3 = q_S + f_T\frac{x_3 - x_2}{2} \qquad (28)$$

1.17 Безматричный МКЭ

У матричного метода конечных элементов, рассмотренного в предыдущем разделе (формулы (23)-(28)) имеются следующие недостатки. Во-первых, для задач с большим числом узлов конечно-элементной сетки $N_1 >> 1$ хранение матриц жесткости и их обращение требует больших затрат ресурсов ЭВМ.

Во-вторых, вычисление коэффициентов матриц жесткости путем приведения подобных членов при общих множителях вида $T_i \delta T_j$, часто служит источником ошибок, требует трудоемкого тестирования и делает программы для ЭВМ трудно модифицируемыми при изменениях формулировки исходной задачи или при изменениях метода интегрирования по времени.

Имеется свободный от этих недостатков и более простой способ реализации метода конечных элементов. Этот способ, называемый безматричным методом конечных элементов, основан на непосредственном использовании дискретизированного вариационного уравнения вида (22). При этом весь материал предыдущего раздела (формулы (23)-(28)) становится абсолютно ненужным.

Идея безматричного МКЭ основана на применении итерационных методов решения дискретных уравнений, реализуемых с помощью алгоритма вычисления невязок g_i дискретных уравнений для заданного пробного решения T_i.

Рассмотрим безматричный алгоритм вычисления невязок рассмотрим на примере неявной схемы Эйлера интегрирования по времени дискретной системы уравнений МКЭ

$$\sum_{j=1}^{N_1} M_{ij}^n \frac{T_j^n - T_j^{n-1}}{\Delta t_n} + \sum_{j=1}^{N_1} K_{ij}^n T_j^n - F_i^n = 0 \qquad (29)$$

здесь значения решения T_j^{n-1} на предыдущем временном слое $(t = t^{n-1})$ считаются известными, а значения решения T_j^n на воеменном слое $t = t^n$ требуется определить. При подстановке

в левую часть уравнения (29) пробного решения $T_j^{n(k)}$, где $k = 0, 1, 2, ...$ (k - номер итерации), она будет равна отличной от нуля невязке $g_i^{(k)}$

$$g_i^{(k)} = \sum_{j=1}^{N_1} M_{ij}^n \frac{T_j^{n(k)} - T_j^{n-1}}{\Delta t_n} + \sum_{j=1}^{N_1} K_{ij}^n T_j^{n(k)} - F_i^n \qquad (30)$$

Номер итерации взят в круглые скобки. Сходящийся итерационный процесс позволяет по найденной невязке $g_i^{(k)}$ определить улучшенное значение пробного решения $T_i^{n(k+1)}$, так что $g_i^{(k)} \to 0$ при $k \to \infty$ и $T_i^{n(k)} \to T_i^n$, где T_i^n - искомое решение.

Алгоритм вычисления невязок $g_i^{(k)}$ по заданному пробному решению $T_i^{n(k)}$ на основе дискретизированного вариационного уравнения (22)

$$(c_V \frac{T_1^{n(k)} - T_1^{n-1}}{\Delta t_n} - f_{T1}^n)\delta T_1 \frac{x_2 - x_1}{2} + (c_V \frac{T_2^{n(k)} - T_2^{n-1}}{dt} - f_{T2}^n)\delta T_2 \frac{x_3 - x_1}{2} +$$

$$+ (c_V \frac{T_3^{n(k)} - T_3^{n-1}}{\Delta t_n} - f_{T3}^n)\delta T_3 \frac{x_3 - x_2}{2} + k_T \frac{T_2^{n(k)} - T_1^{n(k)}}{x_2 - x_1} \frac{\delta T_2 - \delta T_1}{x_2 - x_1}(x_2 - x_1) +$$

$$+ k_T \frac{T_3^{n(k)} - T_2^{n(k)}}{x_3 - x_2} \frac{\delta T_3 - \delta T_2}{x_3 - x_2}(x_3 - x_2) = q_e^n \delta T_3 \qquad (22')$$

заключается в следующем.

Начало алгоритма вычисления невязок.

1) Пробное решение подчиняется главным граничным условиям (в нашем случае это узел 1)

$$T_1^{nk} = T_b^n$$

2) Зануляются дискретные значения невязки $g_i^k := 0$ $(i = 1, 2, 3)$.

3) В цикле по узлам (i=1,2,3) вычисляются вклады в невязку от нестационарных членов и объемных источников и исправляются значения невязок (в нашем случае этому циклу соответствуют

первые три слагаемых в вариационном уравнении)

$$g_i^{n(k)} := g_i^{n(k)} + (c_V \frac{T_1^{n(k)} - T_1^{n-1}}{\Delta t_n} - f_{T1}^n) V_i^n$$

где V_i - приузловые объемы:

$$V_1 = \frac{x_2 - x_1}{2} \ , \quad V_2 = \frac{x_3 - x_1}{2} \ , \quad V_3 = \frac{x_3 - x_2}{2}$$

4) В цикле по конечным элементам вычисляются вклады в невязку от потоковых членов и исправляются значения невязок (в нашем случае двум элементам соответствуют четвертое и пятое слагаемое в вариационном уравнении)

$$g_1^{n(k)} := g_1^{n(k)} + q_{[1]}^{n(k)} d_{11} V_{[1]} \ , \quad g_2^{n(k)} := g_2^{n(k)} + q_{[1]}^{n(k)} d_{12} V_{[1]}$$

$$g_2^{n(k)} := g_2^{n(k)} + q_{[2]}^{n(k)} d_{21} V_{[2]} \ , \quad g_3^{n(k)} := g_3^{n(k)} + q_{[2]}^{n(k)} d_{22} V_{[2]}$$

где первая строка отвечает первому элементу, а вторая строка отвечает второму элементу. Объемы элементов $V_{[1]}$ и $V_{[2]}$ в данном одномерном случае равны их длинам. Номера элементов взяты в квадратные скобки. Потоки имеют вид

$$q_{[1]}^{n(k)} = k_T \frac{T_2^{n(k)} - T_1^{n(k)}}{x_2 - x_1} \ , \quad q_{[2]}^{n(k)} = k_T \frac{T_3^{n(k)} - T_2^{n(k)}}{x_3 - x_2}$$

Обратим внимание на то, что структура вкладов в невязку от потоковых членов везде одинакова: значение потока в элементе умножается на коэффициент пространственного дифференцирования для данного узла и на объем элемента.

При использовании более точных квадратурных формул с несколькими точками численного интегрирования по конечному элементу структура формул сохранится. Только суммирование будет по точкам численного интегрирования, а не по элементам. Значения подынтегральных выражений также будут отвечать точкам численного интегрирования, а вместо объемов элементов

множителями будут стоять объемы элементов, умноженные на весовые коэффициенты квадратцрных формул.

5) В цикле по граничных элементам вычисляются вклады в невязку от граничных потоков и исправляются значения невязок (в нашем случае влияние граничных потоков представлено одним членом в правой части (22')))

$$g_3^{n(k)} := g_3^{n(k)} - q_e^n$$

В общем случае строение вкладов в невязку от граничных потоков определяется аппроксимацией граничного интеграла

$$\int\limits_{S \setminus S_T} q_N^n \delta T dS$$

и имеет вид ("вклад в невязку в узле J2(i,j) от локального узла j в граничном элементе i")

$$g_{J2(i,j)}^{n(k)} := g_{J2(i,j)}^{n(k)} + q_N^{n(k)} L_{ij} S_{[i]}$$

где $S_{[i]}$ - площадь граничного элемента, L_ij - функция формы локального узла j в граничном элементе i, $J2(i,j)$ - глобальный номер локального узла j в граничном элементе i.

6) Значения невязок в узлах, в которых заданы главные граничные условия, зануляются (в нашем случае это узел 1)

$$g_1^{n(k)} = 0$$

Конец алгоритма вычисления невязок.

По сравнению с матричным вариантом метода конечных элементов безматричный вариант значительно проще. В алгоритме безматричного МКЭ трудно сделать ошибку, так как практически все формулы отвечают исходной вариационной постановке задачи.

Для завершения конструирования алгоритма безматричного МКЭ остается только подобрать эффективный итерационный

процесс, опирающийся на вычисление невязок. Таким эффективным итерационным методом является метод сопряженных градиентов.

Рассмотрим алгоритм метода сопряженных градиентов. В формулах алгоритма номер временного слоя n опустим, из узловых значений T_j^n искомого решения составим вектор $\mathbf{T}$ и введем обозначения $\mathbf{g}$ для вектора узловых невязок, $\mathbf{g_0}$ для однородной части вектора невязок, $\mathbf{s}$ - для вспомогательного вектора направления поиска решения. Точки между векторами в записываемых формулах означают скалярное произведение векторов размерности N_1 (число неизвестных). Буквой k обозначим счетчик итераций.

Начало алгоритма метода сопряженных градиентов

1) Для $k = 0$ задаем начальное приближение к решению $\mathbf{T}^{(0)}$ и вычисляем вектор невязок $\mathbf{g}^{(0)}(\mathbf{T}^{(0)})$, полагаем $\mathbf{s}^{(0)} = \mathbf{g}^{(0)}$.

2) Для итераций $k = 0, 1, ...$ делаем следующее.

Проверяем критерий окончания итераций по малости невязки

$$\mathbf{g}^{(k)} \cdot \mathbf{g}^{(k)} < \epsilon^2$$

здесь ϵ - машинное эпсилон, то есть максимальное число, добавление которого к единице компьютер не чувствует: добавляй его хоть миллион раз, результатом будет единица. Для четырехбайтовой (32-битной) арифметики $\epsilon \approx 0.000001$. Если невязка удовлетворяет критерию малости, то решение считается найденным.

Если критерий окончания итераций не выполнен, то по направлению поиска решения $\mathbf{s}^{(k)}$ вычисляем однородную часть вектора невязок $\mathbf{g}_0(\mathbf{s}^{(k)})$, Затем проверяем невырожденность задачи. Если

$$\mathbf{g}_0(\mathbf{s}^{(k)}) \cdot \mathbf{s}^{(k)} < \epsilon^2$$

то задача считается вырожденной и решение прекращается, выводится соответствующее сообщение. Обычно это происходит из-за ошибок в исходных данных.

Если задача невырождена, то находим коэффициент $\alpha^{(k)}$

$$\alpha^{(k)} = \frac{\mathbf{g}^{(k)} \cdot \mathbf{s}^{(k)}}{\mathbf{g}_0(\mathbf{s}^{(k)}) \cdot \mathbf{s}^{(k)}}$$

уточняем решение

$$\mathbf{T}^{(k+1)} = \mathbf{T}^{(k)} - \alpha^{(k)}\mathbf{s}^{(k)}$$

Определяем новое значение невязки, для этого вместо алгоритма определения невязки используется формула

$$\mathbf{g}^{(k+1)} = \mathbf{g}^{(k)} - \alpha^{(k)}\mathbf{g}_0(\mathbf{s}^{(k)})$$

Определяем новое направление поиска решения

$$\mathbf{s}^{(k+1)} = \mathbf{g}^{(k+1)} - \beta^{(k)}\mathbf{s}^{(k)}$$

где коэффициент $\beta^{(k)}$ определяется по формуле

$$\beta^{(k)} = \frac{\mathbf{g}^{(k+1)} \cdot \mathbf{g}_0(\mathbf{s}^{(k)})}{\mathbf{g}_0(\mathbf{s}^{(k)}) \cdot \mathbf{s}^{(k)}}$$

Далее проверяем критерий окончания итераций по малости поправки к решению. Если

$$(\alpha^{(k)}\mathbf{s}^{(k)}) \cdot (\alpha^{(k)}\mathbf{s}^{(k)}) < \epsilon^2$$

то решение считается найденным.

Далее проверяется число итераций. Если k превысило число неизвестных N_1, а решение не найдено, то задача считается плохо обусловленной, процесс итераций прерывается и выводится соответствующее сообщение.

Если критерий окончания итераций не выполнен и число итераций меньше числа неизвестных, то увеличиваем счетчик итераций $k := k+1$ и переходим к выполненю следующей итерации.

Конец алгоритма метода сопряженных градиентов.

Безматричный вариант МКЭ требует для реализации 4 вспомогательных массива длиной N_1: $\mathbf{T}^{(k)}$, $\mathbf{s}^{(k)}$, $\mathbf{g}^{(k)}$, $\mathbf{g}_0^{(k)}$. В самом худшем случае число операций пропорционально N_1^2, так как полное число итераций не более N_1 и на каждой итерации число операций пропорционально N_1. Практическая оценка дает число операций, пропорциональное $N^{3/2}$, а при хорошем начальном приближении число операций на каждом временном шаге может стать пропорциональным N_1, то есть, скорость роста числа операций в неявных схемах для малых временных шагов становится такой же, как для явных схем.

Поясним понятие обусловленности задачи. Пусть минимальное и максимальное собственные числа матрицы A системы линейных алгебраических уравнений задачи $A\mathbf{x} = \mathbf{b}$ обозначены λ_{min} и λ_{max}. Пусть в векторе правой части $\mathbf{b}$ допущена погрешность $\Delta\mathbf{b}$. Тогда для погрешности решения $\Delta\mathbf{x}$ справедлива оценка

$$max|\Delta\mathbf{x}| \approx Cmax|\Delta\mathbf{b}|$$

где $C = \lambda_{max}/\lambda_{min}$ - число обусловленности задачи. Если $C >> 1$, то незначительная погрешность в задании правой части, возникшая, например, из-за ограниченной разрядности представления чисел на ЭВМ, может вызвать катастрофические ошибки в определении решения. В этом случае говорят, что задача плохо обусловлена. Улучшить обусловленность задачи можно, если умножить систему уравнений на обратную матрицу A^{-1}, при этом число обусловленности станет равным единице, а система уравнений предстанет в разрешенном виде. Конечно, обратная матрица неизвестна, поэтому для улучшения обусловленности линейных алгебраических задач используют так называемое предобусловливание, заключающееся в умножении системы уравнений на приближенную обратную матрицу дискретизированной задачи.

Для успешной работы метода сопряженных градиентов при-

ближенную обратную матрицу можно определить, обращая диагональную матрицу, составленную из диагональных элементов матрицы системы, которые нетрудно определить перед началом итераций. Предобусловливание сведется к умножению вектора невязки на каждой итерации на эту приближенную обратную матрицу.

Глава 2

Проекционные методы

2.1 Схема проекционных методов

В общей операторной форме исходная задача может быть представлена так:

$$\mathbf{F}(\mathbf{y}) = 0 \tag{1}$$

где $\mathbf{y}$ - искомый элемент (бесконечномерного) пространства решений Y, $\mathbf{F}$ - (нелинейный) оператор исходного уравнения со значениями в пространстве G. Решение данной задачи получают, как правило, по следующей схеме, называемой проекционным методом. А именно, решение $\mathbf{y}$ ищется в виде разложения по аппроксимационному базису $\{\varphi_j\}_{j=1}^{k}$, определяющему k-мерное пространство приближенных решений $Y^{(k)}$ с элементами следующего вида:

$$\mathbf{y}^{(k)} = \sum_{j=1}^{k} \tilde{y}_j \varphi_j \tag{2}$$

где набор искомых коэффициентов разложения $\{\tilde{y}_j\}_{j=1}^{k}$ иногда называют каркасом приближенного решения и рассматривают как элемент пространства каркасов приближенных решений. Y^k. Результатом подстановки приближенного решения в исходное уравнение является отличная от нуля невязка

$$\mathbf{R}^{(k)} = \mathbf{F}(\mathbf{y}^{(\mathbf{k})}) \tag{3}$$

Поскольку приближенное решение определяется конечным числом (k) коэффициентов, то исходное уравнение можно удовлетворить только приближенно, требуя обращения в нуль проекций

невязки на k-мерное проекционное пространство $G^{(k)}$, определяемое своим базисом ψ_i $(i = 1, ..., k)$. Для этого вводится понятие скалярного произведения элементов $((a, b))$ и ортогональности $((a, b) = 0)$, после чего условия ортогональности невязки и проекционного пространства принимают вид:

$$(\mathbf{R}^{(\mathbf{k})}, \psi_i) = 0 \tag{4}$$

где $i = 1, ..., k$. Условия (4) являются системой (нелинейных) алгебраических уравнений и определяют дискретизированную задачу проекционного метода относительно каркасов приближенных решений. Размерности аппроксимационного и проекционного пространств должны совпадать для того, чтобы число уравнений равнялось числу неизвестных.

Расчетные схемы вида (1)-(4) называются методами Галеркина. Имеется бесконечное разнообразие способов построения аппроксимационного и проекционного базисов, каждому из которых отвечает свой вариант проекционного метода. Когда хотят подчеркнуть различие аппроксимационного и проекционного базисов, говорят о методах Галеркина-Петрова, наоборот, при совпадении базисов говорят о методах Бубнова-Галеркина.

Если уравнение (1) выражает условия минимума некоторого функционала $\Phi(\mathbf{y})$

$$\delta\Phi = (\frac{\partial\Phi}{\partial\mathbf{y}}, \delta\mathbf{y}) = 0 \quad \Longrightarrow \quad \frac{\partial\Phi}{\partial\mathbf{y}} = \mathbf{F}(\mathbf{y}) = 0 \tag{6}$$

где $\delta\mathbf{y}$ является вариацией решения, то есть разностью двух произвольных элементов пространства решений. В этом случае приближенное решение $\mathbf{y}^{(\mathbf{k})}$ подставляется непосредственно в функционал Φ и приближенное решение находится из условий минимума дискретизированного функционала

$$\delta\Phi(\mathbf{y}^{(\mathbf{k})}) = \sum_{j=1}^{k} \frac{\partial\Phi(\mathbf{y}^{(\mathbf{k})})}{\partial\tilde{y}_j}\delta\tilde{y}_j = 0 \tag{7}$$

такой метод решения называется методом Ритца.

Функционалы Φ, имеющие минимум на решении исходной задачи (1) можно построить искусственно, воспользовавшись методом наименьших квадратов:

$$\Phi(\mathbf{y}) = (\mathbf{F}(\mathbf{y}), \mathbf{F}(\mathbf{y})) \tag{8}$$

2.2 Теоремы о сходимости

Некоторые полезные теоретические положения можно сформулировать для следующей задачи, полученной линеаризацией исходного уравнения около произвольного элемента $\mathbf{y}_0 \in Y$

$$\mathbf{F}(\mathbf{y}_0) + \mathbf{F}'_y(\mathbf{y}_0)(\mathbf{y} - \mathbf{y}_0) \approx 0$$

которую перепишем так:

$$Ay = g \tag{10}$$

где $\mathbf{A} = \mathbf{F}'_y(\mathbf{y}_0)$ - оператор линеаризованной задачи, $\mathbf{g} = \mathbf{F}'_y(\mathbf{y}_0)\mathbf{y}_0 - \mathbf{F}(\mathbf{y}_0)$ - известная правая часть. Отвечающая этому уравнению система линейных алгебраических уравнений метода Галеркина имеет вид:

$$A_k y^k = g^k \tag{11}$$

где $y^k = \{\tilde{y}_i\}_{i=1}^k$ - каркас приближенного решения, выражения для матрицы A_k и вектора правой части $g^k = \{\tilde{g}_i\}_{i=1}^k$ имеют вид

$$A_k = \{(\psi_i, A\varphi_j)\}_{i,j=1}^k \ , \quad g^k = \{(\psi_i, g)\}_{i=1}^k \tag{12}$$

Обозначим операции дискретизации решения и уравнения задачи операторами проектирования $p_k \colon Y \to Y^k$ и $P_k : G \to G^k$. Здесь Y^k и G^k - пространства коэффициентов разложения решения $\{\tilde{y}_i\}_{i=1}^k$ и правой части $\{\tilde{g}_i\}_{i=1}^k$ по базисам $\{\varphi_i\}_{i=1}^k$ и $\{\psi_i\}_{i=1}^k$:

$$y^k = p_k y^{(k)} \ , \quad g^k = P_k g^{(k)} \tag{13}$$

Наоборот, переход от дискретного представления к элементам функциональных пространств реализуется операторами восполнения $\tilde{p}_k : Y^k \to Y^{(k)}$ и $\tilde{P}_k : G^k \to G^{(k)}$. Здесь $Y^{(k)}$ и $G^{(k)}$ пространства приближенных решений и правых частей:

$$y^{(k)} = \sum_{i=1}^{k} \tilde{y}_i \varphi_i \ , \quad g^{(k)} = \sum_{i=1}^{k} \tilde{g}_i \psi_i \tag{14}$$

или

$$y^{(k)} = \tilde{p}_k y^k \ , \quad g^{(k)} = \tilde{P}_k g^k \tag{15}$$

Заметим, что операторы проектирования и восполнения не являются взаимно обратными.

Здесь и далее полагаем, что введенные пространства являются нормированными. Решение дискретизированной задачи y^k отличается от дискретной проекции точного решения $p_k y$ на величину:

$$\tau_k = ||y^k - p_k y||$$

называемую ошибкой приближенного решения в пространстве каркасов приближенных решений Y^k.

Близость исходного и дискретизированного уравнений характеризуется мерой аппроксимации, которая сравнивает результаты двух возможных путей преобразования решения исходной задачи в дискретный вектор правой части g^k. Первый путь состоит в подстановке решения в исходный оператор задачи и проектировании результата на проекционное пространство G^k:

$$y \to Ay \to P_k(Ay)$$

Второй путь состоит в проектировании решения на аппроксимационное пространство с последующей подстановкой результата в дискретный аналог исходного оператора задачи:

$$y \to p_k y \to A_k p_k y$$

где A_k: $Y^k \to G^k$ - дискретный аналог исходного оператора задачи. Хотя результаты вычислений по двум путям принадлежат проекционному пространству G^k, они не совпадают. Их отличие определяет величина

$$\gamma_k = ||A_k p_k y - P_k A y||$$

которая называется мерой аппроксимации.

Пусть для меры аппроксимации $\gamma_k = ||A_k p_k y - P_k A y||$ имеет место оценка $\gamma_k = Ck^{-N}$, где число $N > 0$ возможно является дробным и характеризует скорость убывания меры аппроксимации с ростом размерности k проекционного и аппроксимационного пространств. Это число N называется порядком аппроксимации.

Под устойчивостью конечномерной задачи понимается существование ограниченного обратного оператора дискретизированной задачи $||A_k^{-1}|| \leq M < \infty$ При этом возмущения решения малы, если малы возмущения правых частей и оператора задачи.

Теорема о сходимости каркасов приближенных решений.*Из аппроксимации* $\gamma_k \to 0$ *и устойчивости* $||A_k^{-1}|| \leq M < \infty$ *следует сходимость:* $\tau_k \to 0$.

Доказательство: поскольку $A_k y^k = g^k = P_k g = P_k A y$, то

$$||p_k y - y^k|| = ||A_k^{-1}(A_k p_k y - A_k y^k)|| \leq$$

$$\leq ||A_k^{-1}||||A_k p_k y - P_k g|| \leq M\gamma_k \to 0$$

Теорема о сходимости приближенных решений. *Если каркасы приближенных решений сходятся (* $\beta_k = ||A_k^{-1}||\gamma_k \to 0$ *), а оператор восполнения корректен (* $||\tilde{p}_k p_k y - y|| \to 0$ *) и ограничен (* $||\tilde{p}_k|| \leq P < \infty$ *), то приближенные решения сходятся:* $y^{(k)} \to y$.

Доказательство:

$$||y^{(k)} - y|| = ||\tilde{p}_k y^k - y|| \leq$$

$$\leq ||\tilde{p}_k y^k - \tilde{p}_k p_k y|| + ||\tilde{p}_k p_k y - y|| \leq$$

$$\leq ||\tilde{p}_k||||p_k y - y^k|| + ||\tilde{p}_k p_k y - y|| =$$

$$= ||\tilde{p}_k||||p_k y - A_k^{-1} A_k y^k|| + ||\tilde{p}_k p_k y - y|| =$$

$$= ||\tilde{p}_k||||A_k^{-1} A_k p_k y - A_k^{-1} P_k A y|| + ||\tilde{p}_k p_k y - y|| \leq$$

$$\leq ||\tilde{p}_k||\beta_k| + ||\tilde{p}_k p_k y - y|| \to 0$$

2.3 Ошибки проекционных методов

При численной реализации различают: ошибку в задании оператора задачи ΔA, ошибку в задании правой части Δg и ошибку в вычислении невязки уравнения Δ_s.

Суммарная ошибка Δy определяется из уравнения:

$$(A + \Delta A)(y + \Delta y) = g + \Delta g + \Delta_s$$

и подчинена следующему неравенству, вывод которого можно найти в книге Гавурина (1970):

$$||\Delta y|| \leq \frac{cond(||A||)}{1 - cond(A)\frac{||\Delta A||}{||A||}} \left[\frac{||\Delta A||}{||A||} + \frac{||\Delta g|| + ||\Delta_s||}{||g||} \right]$$

где $condA = ||A^{-1}||||A||$ - число обусловленности оператора A. Аналогичный результат имеет место и для дискретных уравнений.

Оценки показывают, что при больших значениях числа обусловленности ($cond(A) >> 1$) влияние ошибок приводит к потере точности. Задачи при больших значениях числа обусловленности являются плохо обусловленными. Они нуждаются в регуляризации, то есть, в улучшении свойств путем преобразования уравнений к виду с меньшим значением числа обусловленности.

2.4 Нестационарные задачи

Рассмотрим схему проекционных методов в случае эволюционных уравнений. В этом случае запись исходной задачи в операторной форме содержит нестационарный член с производной по времени:

$$\partial_t y = Ay - g$$

и начальные условия

$$y|_{t=0} = y^0$$

Как и для стационарных задач по методу Галеркина решение ищется в виде разложения по базисным функциям, но с коэффициентами, зависящими от времени

$$y^{(k)} = \sum_{i=1}^{k} \tilde{y}_i(t)\varphi_i(r)$$

Разрешающие уравнения в этом случае как и для стационарных задач выражают ортогональность невязки к проекционному пространству, но из-за нестационарных членов принимают вид системы обыкновенных дифференциальных уравнений по времени

$$\sum_{j=1}^{k}(\psi_i, \varphi_j)\partial_t\tilde{y}_j = (\psi_i, g) - \sum_{j=1}^{k}(\psi_i, A\varphi_j)\tilde{y}_j \ , \quad i = 1, 2, ..., k.$$

дополненной начальными условиями, которые также скалярным умножением на элементы проекционного базиса приводятся к начальным условиям для каркасов приближенных решений

$$\sum_{j=1}^{k}(\psi_i, \varphi_j)\tilde{y}_j(0) = (\psi_i, y^0)$$

Нестационарные задачи можно решать как последовательность вспомогательных стационарных задач. Для этого вводится

разностная аппроксимация производных по времени, например, такая:

$$\frac{y^{n+1} - y^n}{\Delta t_n} = Ay^{n+1} - g^{n+1}$$

где индекс n нумерует временные слои, Δt_n - шаг по времени. Для величин на новом временном слое (n+1) возникает вспомогательная стационарная задача, так что все ранее рассмотренные проекционные методы можно применять и в этом случае.

2.5 Задачи на собственные значения

Задачи на собственные значения возникают во многих практических приложениях в связи с определением собственных частот и форм колебаний, критических нагрузок и форм потери устойчивости, построением спектральных базисов, а также в связи с определением точек ветвления решений нелинейных задач (см. далее главу про ветвление решений нелинейных уравнений).

Операторная запись задачи на собственные значения имеет вид:

$$Ay = \lambda By$$

Тривиальное решение $y = 0$ имеет место для любых значений числа λ и интереса не представляет. Требуется определить нетривиальные решения (собственные функции) и соответствующие значения параметра λ (собственные значения). Проекционные методы решения основаны на представлении решения в виде линейных комбинаций функций аппроксимационного базиса:

$$y^{(k)} = \sum_{i=1}^{k} \tilde{y}_i \varphi_i(r)$$

Для метода Галеркина-Петрова дискретные уравнения задачи на собственные значения выражают ортогональность невязки к

проекционному пространству с базисом $\{\psi_i\}_{i=1}^k$ и имеют вид:

$$A_k y^k = \lambda B_k y^k$$

где

$$A_k = \{(\psi_i, A\varphi_j)\}_{i,j=1}^k \ , \quad B_k = \{(\psi_i, B\varphi_j)\}_{i,j=1}^k$$

Решение полученной алгебраической задачи на собственные значения получается далее методами линейной алгебры.

Другой эффективный метод отыскания собственных решений, основан на минимизации функционалов. Минимальное собственное значение и соответствующая собственная функция определяются путем решения задачи минимизации

$$\lambda_1 = \min_{y \in Y} \frac{(Ay, y)}{(By, y)}$$

Следующие собственные значения λ_m , $m = 2, 3, ...$ также определяются задачами минимизации

$$\lambda_m = \min_{y \in Y \setminus Y_{m-1}} \frac{(Ay, y)}{(By, y)}$$

где Y_{m-1} - оболочка, натянутая на (m-1) собственных функций, отвечающих первым (m-1) собственным числам. Подробное практическое описание упомянутых алгоритмов для линейных задач на собственные значения можно найти в книгах (Михлин, 1970) и (Уилкинсон, Райнш, 1976).

Глава 3

Интерполяция

3.1 Задание функций

Известны следующие способы задания функций: аналитический способ подразумевает, что имеется формула для вычисления значения функции по значению аргумента; алгоритмический способ использует последовательность математических действий (алгоритм) вычисления функции по значению аргумента и, наконец, табличный способ, который определяет интерполяцией значение функции f(x) по ее значениям в конечном числе точек (то есть по таблице): $(x_k, f_k)_{k=1}^{N}$.

Интерполяция это аналитическое или алгоритмическое приближенное представление таблично заданной функции, позволяющее определить ее значение в любой точке ее области определения.

Экстраполяция это применение интерполяционных формул или алгоритмов для определения значений функции за пределами ее области определения.

Различают следующие основные типы интерполяции. Глобальная интерполяция использует базисные функции, отличные от нуля во всей области определения интерполируемой функции. Примером может служить интерполяция степенными или тригонометрическими функциями. Глобальная интерполяция часто является бессеточной.

Локальная интерполяция использует базисные функции, отличные от нуля в малой окрестности данной точки. Такие интерполяции используются при численном моделировании с применением сеток, частиц или свободных узлов. Простейшим примером

локальной интерполяции является одномерная кусочно-линейная интерполяция.

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

где $x \in [x_i, x_{i+1}]$, $i = 1, 2, ..., N - 1$, x_i и x_{i+1} - соседние узлы сетки.

3.2 Полиномы Лагранжа

Функции полиномиального базиса следующего вида называются полиномами Лагранжа

$$\varphi^{(i)}(x) = \frac{\prod\limits_{k=1,k\neq i}^{N}(x - x_k)}{\prod\limits_{k=1,k\neq i}^{N}(x_i - x_k)}$$

где i-й полином принимает значения 1 в точке x_i и 0 во всех остальных табличных точках, то есть

$$\varphi^{(i)}(x_k) = \delta_{ki}$$

где δ_{ki} - индексная функция Кронекера, равная единице, если индексы совпадают и нулю в противном случае. Иногда ее называют символом Кронекера или дельтой Кронекера.

Решение интерполяционной проблемы Лагранжа имеет следующий очень простой вид

$$f^{(N)}(x) = \sum_{i=1}^{N} f_i \varphi^{(i)}(x)$$

то есть табличные значения функции служат коэффициентами разложения.

Если интерполируемая функция имеет производные до N-го порядка включительно и N-я производная ограничена

$$|f^{(N)}(\xi)| < M < \infty$$

то оценка погрешности интерполяции имеет вид:

$$|f(x) - f^{(N)}(x)| \leq \frac{1}{n!} M |x - x_1| \cdots |x - x_N|$$

3.3 Степенные функции

Выбор базиса исключительно важен для успеха численных методов. Например, в том же функциональном пространстве степенных полиномов, которому принадлежат полиномы Лагранжа, можно воспользоваться другим базисом и столкнуться с вычислительной катастрофой.

Действительно, попробуем поискать решение задачи интерполяции в виде разложения по глобальному базису степенных функций

$$f(x) = \sum_{i=1}^{N} c_i x^{i-1}$$

Ошибка запишется так:

$$E = \int_{x_1}^{x_N} \left[\sum_{i=1}^{N} c_i x^{i-1} - f(x) \right]^2 dx$$

Коэффициенты разложения определяем из условия минимума ошибки:

$$\frac{\partial E}{\partial c_i} = 0 \Rightarrow \sum_{j=1}^{N} h_{ij} c_j = b_i$$

где

$$b_i = \int_{x_1}^{x_N} f(x) x^{i-1} dx$$

$$h_{ij} = \int_{x_1}^{x_N} x^{i+j-2} dx = \frac{1}{i+j-1}$$

Матрица Гильберта $H = \{h_{ij}\}$ очень плоха для вычислений, что сейчас станет видно.

3.4 Ошибки и число обусловленности

Уравнение для собственных значений матрицы Гильберта H имеет вид

$$\det(h_{ij} - \lambda\delta_{ij}) = 0$$

где δ_{ij} - дельта Кронекера.

Числом обусловленности симметричной вещественной положительной матрицы H называется величина

$$cond(H) = ||H||||H^{-1}|| = \frac{\lambda_{\max}}{\lambda_{\min}}$$

равная отношению максимального и минимального собственных чисел матрицы. Число обусловленности больше или равно единице.

В случае знаконеопределенных матриц A, у которых собственные числа могут принимать положительные, отрицательные и даже комплексные значения, число обусловленности определяется как отношение максимального и минимального сингулярных чисел матрицы, являющихся квадратными корнями собственных чисел симметризованной положительной матрицы AA^T .

Ошибка решения $||\delta c||$ системы линейных алгебраических уравнений

$$Hc = b$$

возникаюшая из-за погрешности правой части растет пропорционально числу обусловленности (см. Форсайт, Молер, 1967).

$$||\delta c||/||c|| \leq cond(H)||\delta b||/||b||$$

Числа обусловленности $cond(H)$ для матрицы Гильберта быстро стремятся к бесконечности с ростом числа используемых базисных функций N, что показано в таблице 2.4.1.

Таблица 2.4.1.

N	cond(H)
2	$1.9_{10}1$
3	$5.2_{10}3$
4	$1.6_{10}4$
5	$4.8_{10}5$
6	$1.5_{10}7$
7	$4.8_{10}8$
8	$1.5_{10}10$
...	...

Большие значения числа обусловленности делают невозможным определение коэффициентов интерполяции уже при приближении числа базисных функций N к 10,. так как небольшие возмущения в правой части вызывают огромные изменения в решении. Отсюда следует вывод о том, что степенные функции образуют очень плохой глобальный базис, который приводит к очень плохо обусловленной задаче для определения коэффициентов интерполяции.

Хотя полиномы Лагранжа являются линейными комбинациями степенных функций и принадлежат тому же функциональному пространству, они представляют наилучший базис в этом пространстве, поскольку система уравнений для коэффициентов разложения Лагранжа характеризуется единичной матрицей и имеет число обусловленности равное единице, что представляет идеальный случай.

Таким образом, в одном и том же функциональном пространстве эффективность интерполяции определяется выбором базиса. Выбор базиса играет важнейшую роль и в общем случае применения проекционных методов.

3.5 Сплайны

В основе сплайн-аппроксимации лежит идея приближения функции полиномами невысокого порядка, каждый из которых действует на своей ячейке сетки. Коэффициенты таких полиномов определяются условиями коллокации (совпадения значений) этих полиномов и интерполируемой функции в точках коллокации и условиями сопряжения полиномов между собой по значению функции и ее нескольких низших производных на границах между ячейками. Для замыкания системы алгебраических уравнений на границах области изменения функции сплайны подчиняются некоторым дополнительным граничным условиям (выражающим, например, равенство нулю старших производных). Сплайны применяются на регулярных сетках, имеющих ijk покоординатную нумерацию узлов, так как при этом запись условий непрерывности на границах ячеек не вызывает затруднений.

Кусочно-полиномиальная аппроксимация сплайнами приводит к хорошо обусловленным системам алгебраических уравнений относительно коэффициентов разложения. Для приближения сплайнами функций со сложным поведением не требуется повышать порядок полиномов, а достаточно увеличить число ячеек сетки.

Во многих случаях сплайны показывают очень хорошие результаты. Так, кубические сплайны, образованные набором полиномов третьей степени, позволяют интерполировать табличные данные так, что человеческий глаз не замечает каких-либо изломов на получающихся графиках. Для точного воспроизведения окружности радиуса R достаточно использовать параметрическое представление окружности ($x(\xi) = Rcos(\xi)$, $y(\xi) = Rsin(\xi)$) и представить функции x и y кубическими сплайнами на четырех одномерных ячейках по параметрической координате $0 \leq \xi \leq 2\pi$.

Подробное изложение теории сплайн-аппроксимации с прак-

тическими примерами дано в монографии (Алберг Дж., Нильсон Э., Уолш Дж., 1973).

Применение сплайнов ограничено случаями, когда область решения представима регулярными ijk сетками, то есть может быть отображена на отрезок (одномерный случай), квадрат (двумерный случай) или куб (трехмерный случай). При этом, хотя базис сплайн-аппроксимации является глобальным, системы уравнений для коэффициентов сплайна характеризуются ленточными матрицами.

3.6 Многомерная сеточная интерполяция

3.6.1 Типы сеток

Далее будет использоваться стандартная терминология для характеристики свойств используемых сеток. Говорят, что сетка задана, если ее узлы пронумерованы, координаты узлов заданы и для каждого узла сетки определены его соседи. Область определения заданной на сетке функции при этом аппроксимирована (приближенно представлена) объединением ячеек сетки, для которых указаны номера образующих эти ячейки узлов.

Регулярная (структурированная) сетка это такая сетка, для которой имеется правило для определения соседства узлов. Примером может служить ijk-сетка с координатами узлов $x_i = ih_x$, $y_j = jh_y$, $z_k = kh_z$ В такой сетке для узла $(i, j, k\ )$ соседями являются узлы $(i \pm 1, j \pm 1, k \pm 1\ )$.

В нерегулярных (неструктурированных) сетках соседство узлов определяется информационными массивами соседства, содержащими для каждого узла номера соседних узлов или для каждой ячейки номера образующих ее узлов.

В равномерной сетке все ячейки имеют одинаковую форму и размер. *В неравномерных сетках* имеются ячейки разных размеров.

В однородных сетках все ячейки имеют одинаковое число узлов. *В неоднородных сетках* содержатся ячейки с разным числом узлов. *Ребром* назыается линия, соединяющая два соседних узла. *Гранью* называется поверхностная ячейка, служащая границей для объемной ячейки. Заметим, что регулярная сетка вполне может быть неравномерной и непрямоугольной при использовании криволинейных координатных линий, отвечающих постоянным значениям индексов i, j и k.

3.6.2 Покоординатная интерполяция

Для регулярных ijk-сеток чаще всего используется набор одномерных интерполяций по координатным направлениям i, j, k. Пусть положение узлов такой криволинейной сетки определяются отображением $x = x(\xi)$. Здесь $\xi = (\xi_1, \xi_2, \xi_3))$ - декартовы прямоугольные координаты в трехмерном пространстве прообраза Ξ, в котором в кубе $(0 \le \xi_1 \le 1;\ 0 \le \xi_2 \le 1;\ 0 \le \xi_3 \le 1)$ задана равномерная ijk-сетка. Прямоугольные декартовы координаты $x = (x_1, x_2, x_3)$ заданы в трехмерном пространстве образа X, в котором рассматриваемое отображение определяет криволинейный куб-образ с наведенной в нем криволинейной ijk-сеткой. Это отображение характеризуется матрицей Якоби $\partial x/\partial \xi$ с якобианом $det(\partial x/\partial \xi) > 0$.

Формула связи интерполянтов на исходной прямоугольной сетке (прообразе) и криволинейной сетке (образе) имеет вид

$$f_{(x)}(x) = f_{(\xi)}(\xi(x))$$

3.6.3 L-координаты

Нерегулярные сетки, содержащие ячейки переменной формы и размера, а также ячейки с различным числом узлов, строятся обычно без использования отображений сразу в пространстве образа (в актуальной области решения). Для нерегулярных сеток

применяется кусочно-полиномиальная (конечно-элементная) интерполяция. Для этого на простейших одномерных (отрезок), двумерных (треугольник) или трехмерных (тетраэдр) конечных элементах используются так называемые L-координаты.

Одномерные L-координаты. В одномерном случае сетка представлена конечными элементами, являющимися отрезками. Значение интерполируемой функции f в точке p с координатой x по ее значениям в узлах определяется по следующей интерполяционной формуле

$$f(x) = f_1 L_1(x) + f_2 L_2(x)$$

где функции формы (L координаты) определяются отношениями длин отрезков

$$L_1 = \frac{l_{2p}}{l_{21}}\ , \quad L_2 = \frac{l_{p1}}{l_{21}}$$

где $l_{ij} = x_i - x_j$.

Двумерные L-координаты. В двумерном случае для треугольного конечного элемента значение интерполируемой функции f в точке p с координатами (x, y) по значениям ее в узлах определяется по следующей интерполяционной формуле

$$f(x, y) = f_1 L_1(x, y) + f_2 L_2(x, y) + f_3 L_3(x, y)$$

где L координаты определяются отношениями площадей треугольников (см. рис. 3.1)

$$L_1 = \frac{\Delta_{p23}}{\Delta_{123}}\ , \quad L_2 = \frac{\Delta_{1p3}}{\Delta_{123}}\ , \quad L_3 = \frac{\Delta_{12p}}{\Delta_{123}}$$

Замечание. Площадь треугольника Δ_{ijk} , где i, j, k - номера вершин, определяется половиной векторного произведения векторов, представляющих смежные стороны треугольника:

$$\Delta_{ijk} = [(y_j - y_i)(x_k - x_i) - (y_k - y_i)(x_j - x_i)]/2$$

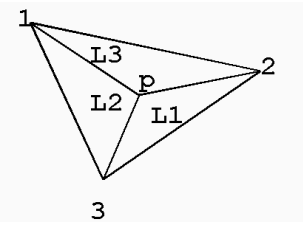

Рис. 3.1: L-координаты для треугольной ячейки

или в другой записи

$$\Delta_{ijk} = \frac{1}{2} \begin{vmatrix} 1 & x_1 - x_m & y_1 - y_m \\ 1 & x_2 - x_m & y_2 - y_m \\ 1 & x_3 - x_m & y_3 - y_m \end{vmatrix}$$

где

$$x_m = (x_1 + x_2 + x_3)/3$$
$$y_m = (y_1 + y_2 + y_3)/3$$

Локальную нумерацию узлов определяют так, чтобы площадь треугольника была бы положительной.

Трехмерные L-координаты. В трехмерном случае для тетраэдрального конечного элемента значение интерполируемой функции f в точке p с координатами (x, y) по значениям ее в узлах определяется по следующей интерполяционной формуле

$$f(x, y, z) = f_1 L_1(x, y, z) + f_2 L_2(x, y, z) + f_3 L_3(x, y, z) + f_4 L_4(x, y, z)$$

где функции формы (L координаты) определяются отношениями объемов тетраэдров

$$L_1 = \frac{V_{p234}}{V_{1234}}, \quad L_2 = \frac{V_{1p34}}{V_{1234}}, \quad L_3 = \frac{V_{12p4}}{V_{1234}}, \quad L_4 = \frac{V_{123p}}{V_{1234}}$$

Объем тетраэдра , где i, j, k, l - номера вершин, определяется формулой

$$V_{ijkl} = \frac{1}{6} \begin{vmatrix} 1 & x_1 - x_m & y_1 - y_m & z_1 - z_m \\ 1 & x_2 - x_m & y_2 - y_m & z_2 - z_m \\ 1 & x_3 - x_m & y_3 - y_m & z_3 - z_m \\ 1 & x_4 - x_m & y_4 - y_m & z_4 - z_m \end{vmatrix}$$

где

$$x_m = (x_1 + x_2 + x_3 + x_4)/4$$
$$y_m = (y_1 + y_2 + y_3 + y_4)/4$$
$$z_m = (z_1 + z_2 + z_3 + z_4)/4$$

Приведенные простейшме интерполяционные формулы используют кусочно-линейную аппроксимацию и имеют второй порядок точности, то есть ошибка интерполяции пропорциональна квадрату характерного размера ячейки: $\varepsilon = O(h_\Delta^2)$. С помощью L-координат строятся интерполяции и более высоких порядков

Описанные способы интерполяции далеко не исчерпывают их множества и разнообразия. Более подробное описание основных способов интерполяции можно найти в специальной литературе по численным методам, в частности, для метода конечных элементов смотри книгу Сегерлинда (1979).

Глава 4

Численное интегрирование

При реализации проекционных методов матрицы систем уравнений и векторы правых частей определяются скалярными произведениями, реализуемыми с помошью интегралов. Поскольку подынтегральные выражения в практических задачах нелинейны, приходится использовать численное интегрирование, реализуемое с помощью квадратурных формул приближенного вычисления интегралов. Рассмотрим задачу численного интегрирования.

4.1 Простейшие квадратурные формулы

Простейшей квадратурной формулой является формула прямоугольников:

$$\int_{x_i}^{x_{i+1}} f(x)dx = f(\tilde{x})(x_{i+1} - x_i)$$

которая просто обобщается на двумерный и трехмерный случаи

$$\int_{S_k} f(x)dS = f(\tilde{x})S_k \ , \quad \int_{V_k} f(x)dS = f(\tilde{x})V_k$$

где S_k площадь поверхностной и V_k объем пространственной ячейки, $\tilde{x}$ - некоторая точка, принадлежащая ячейке. В одномерном случае оценка локальной ошибки квадратурной формулы

прямоугольников выполняется так

$$\varepsilon = \left| \int_{x_i}^{x_{i+1}} f(x)dx - f(\tilde{x})(x_{i+1} - x_i) \right| \leq$$

$$\leq \left| f'(\tilde{x}) \int_{x_i}^{x_{i+1}} (x - \tilde{x})dx \right| + \frac{1}{2} \left| f''(\tilde{x}) \int_{x_i}^{x_{i+1}} (x - \tilde{x})^2 dx \right| + (h_i^4)$$

где $h_i = x_{i+1} - x_i$. Во всех точках кроме центра интервала $x \neq x_{i+1/2} = 0.5(x_{i+1} + x_i)$ локальная ошибка пропорциональна квадрату шага сетки:

$$\varepsilon = 0.5 M_i^{(1)} h_i^2 \ , \quad |f'(x)| < M_i^{(1)}$$

а в центре интервала $x = x_{i+1/2}$ она пропорциональна третьей степени шага сетки:

$$\varepsilon = \frac{1}{24} M_i^{(2)} h_i^3 \ , \quad |f''(x)| < M_i^{(2)}$$

В середине интервала асимптотическая скорость убывания погрешности скачком возрастает. Такие точки называются точками сверхсходимости.

Пример квадратурной формулы повышенной точности дает формула Симпсона:

$$\int_{x_i}^{x_{i+1}} f(x)dx = \frac{h}{3}(f_i + 4f_{i+1} + f_{i+2})$$

где

$$h = x_{i+2} - x_{i+1} = x_{i+1} - x_i$$

Ошибка формулы Симпсона записывается так:

$$\varepsilon \approx M^{(4)} h^5 \ , \quad |f^{(4)}(x)| < M^{(4)}$$

4.2 Квадратуры Гаусса

В многомерном случае применяются квадратурные формулы Гаусса

$$\int_V f(x)dV = \sum_{i=1}^{N} f(x_i)\omega_i V + \varepsilon_N$$

где V - n-мерная ячейка (отрезок, треугольник, четырехугольник, тетраэдр, топологический куб и т.д.), N - количество гауссовых точек интегрирования x_i , ω_i - весовые коэффициенты, обладающие свойством

$$\sum_{i=1}^{N} \omega_i = 1$$

гарантирующим точное интегрирование функции-константы, ε_N - погрешность, зависящая от числа гауссовых точек. Число и координаты гауссовых точек интегрирования для каждой квадратуры зависят от типа ячейки (линейная, плоская, объемная, треугольная, четырехугольная, тетраэдральная и так далее) и желаемой точности интегрирования. Таблицы широко используемых гауссовых квадратур приведены ниже.

4.2.1 Одномерное интегрирование

Рассмотрим квадратуры Гаусса для вычисления интеграла

$$\int_{-1}^{1} f(x)dx = \sum_{i=1}^{n} \omega_i f(a_i)$$

где координаты точек интегрирования $a_i = \pm a$, число точек N и весовые коэффициенты ω_i даны ниже в таблице

Таблица 1.3.2.1.

	±a	ω
$N = 2$	0.577360	1.0
$N = 3$	0.774591	0.(5)
	0.0	0.(8)
$N = 4$	0.861136	0.347865
	0.339981	0.652145
$N = 5$	0.906180	0.236927
	0.538470	0.478629
	0.0	0.56(8)
$N = 6$	0.932470	0.171324
	0.661210	0.360762
	0.238619	0.467914
$N = 7$	0.949110	0.129485
	0.741531	0.279705
	0.405845	0.381830
	0.0	0.417959
$N = 8$	0.960290	0.101228
	0.796666	0.222381
	0.525532	0.313707
	0.183435	0.362684
	±a	ω
$N = 9$	0.968160	0.081274
	0.836031	0.180648
	0.013371	0.260611
	0.324253	0.312347
	0.0	0.330239
$N = 10$	0.973906	0.066671
	0.865063	0.149451
	0.679410	0.219086
	0.433395	0.269267
	0.148874	0.295524

4.2.2 Двумерное интегрирование

Квадратуры Гаусса для треугольных ячеек имеют вид

$$\int_S f(x,y)dS = S\sum_{i=1}^{n}\omega_i f(L_1,L_2,L_3) + R$$

где S_Δ - площадь треугольника. В таблице 1.3.2.2. даны значения L-координат точек численного интегрирования, соответствующие значения весовых коэффициентов ω_i и погрешности R

Таблица 1.3.2.2

N	ω	L_1	L_2	L_3	Кратность
$N=3$	0.(3)	0.(6)	0.1(6)	0.1(6)	3
$N=3$	0.(3)	0.5	0.5	0.0	3
$N=4$	-0.56250	0.(3)	0.(3)	0.(3)	1
	0.5208(3)	0.6	0.2	0.2	3
$N=6$	0.1(6)	0.659028	0.231933	0.109039	6
$N=6$	0.109952	0.816848	0.091576	0.091576	3
	0.223381	0.108103	0.445948	0.445948	3
$N=7$	0.375	0.(3)	0.(3)	0.(3)	1
	0.1041(6)	0.736712	0.237932	0.025355	6
$N=7$	0.225033	0.(3)	0.(3)	0.(3)	1
	0.125939	0.797427	0.101286	0.101286	3
	0.132394	0.470142	0.470142	0.059716	3
$N=9$	0.205950	0.124950	0.437525	0.437525	3
	0.063691	0.797112	0.165410	0.037477	6
$N=12$	0.050845	0.873822	0.063089	0.063089	3
	0.116786	0.501426	0.249287	0.249287	3
	0.082851	0.636502	0.310352	0.053145	6
$N=13$	-0.149570	0.(3)	0.(3)	0.(3)	1
	0.175615	0.479308	0.260346	0.260346	3
	0..053347	0.869740	0.065130	0.065130	3
	0.077114	0.638444	0.312865	0.048690	6

Приведенные формулы (Стренг и Фикс, 1977) симметричны относительно пространственных переменных, поэтому если встречается квадратурный узел (L_1 ,L_2 ,L_3), то обязательно встречаются и все его перестановки. Если все L-координаты различны, то таких узлов в квадратуре 6, если две L -координаты совпадают, то таких узлов три, если используется центральная точка (все L -координаты совпада.ют), то лишь один раз. В выражении для погрешности R характерный размер ячейки обозначен h.

4.2.3 Трехмерное интегрирование

Квадратуры Гаусса для тетраэдральной ячейки имеют вид

$$\int_V f(x,y)dV = V\sum_{i=1}^{n}\omega_i f(L_1,L_2,L_3) + R$$

Весовые коэффициенты и координаты даны в таблице 1.3.2.3:

Таблица 1.3.2.3.

Точки	L-координаты	ω
$N=1,\ R=O(h^2)$		
a	0.25 0.25 0.25 0.25	1.0
$N=4,\ R=O(h^3)$ $\alpha=0.585410,\ \beta=0.138197$		
a	α,β,β,β	1/4
b	α,β,β,β	1/4
c	α,β,β,β	1/4
d	α,β,β,β	1/4
$N=5,\ R=O(h^4)$		
a	0.250.250.250.25	-4/5
b	1/31/61/61/6	9/20
c	1/61/31/61/6	9/20
d	1/61/61/31/6	9/20
e	1/61/61/61/3	9/20

4.3 Бессеточное интегрирование

Нередко возникает необходимость численного интегрирования функций многих переменных в областях сложной формы в условиях, когда никакой сетки нет. Например, такая ситуация создается при реализации бессеточных методов Галеркина, в которых решение ищется в виде разложения по некоторому, не связанному с какой-либо сеткой, набору базисных функций. В таких случаях используется каноническая вспомогательная область, которая включает в себя рассматриваемую область интегрирования сложной формы. В канонической области вводится равномерная регулярная ijk сетка прямоугольных ячеек (параллелепипедов или квадратов). Вычисление интеграла проводится суммированием интегралов по ячейкам, центры которых принадлежит исходной области интегрирования. Интегралы в ячейках аппроксимируются по какой-либо квадратурной формуле, например, по формуле прямоугольников.

При увеличении числа ячеек интегрирования погрешности, возникающие из-за несогласованности ячеек сетки с границей заданной области интегрирования, стремятся к нулю вместе с обычными ошибками аппроксимации интегралов. Поскольку запоминать вспомогательную сетку не требуется, то реализация вычисления интегралов в бессеточных методах Галеркина по сути также является бессеточной.

Глава 5

Численное дифференцирование

5.1 Использование интерполянтов

Один из наиболее очевидных способов численного дифференцирования заключается в построении интерполирующей функции и в ее последующем обычном дифференцировании. Пусть $f_h(x)$ - интерполянт функции $f(x)$, аппроксимирующий ее с ошибкой $O(h^m)$, тогда интерполянты производных (то есть функции, интерполирующие производные) вычисляются дифференцированием

$$\frac{df}{dx} = \frac{df_h}{dx} + O(h^{m-1})$$

в результате порядок аппроксимации производной (показатель степени шага в ошибке) оказывается на единицу меньше, чем для самой функции.

5.2 Метод неопределенных коэффициентов

Формулы для вычисления производных можно получить методом неопределенных коэффициентов. В соответствии с этим методом в окрестности данного узла сетки интерполянт функции ищется в виде полинома. Окрестность узла определяется шаблоном, то есть набором узлов, включающих данный узел и его соседние узлы. Коэффициенты полинома определяются из системы алгебраических уравнений, выражающих требование равенства значений полинома и функции в узлах шаблона (условия

коллокации). Число узлов шаблона должно быть равным числу искомых коэффициентов полинома, то есть порядок полинома и шаблон взаимосвязаны требованием равенства числа неизвестных и числа уравнений. Пока полиномы имеют невысокий порядок, вычислительная катастрофа, описанная в разделе про интерполяцию степенными функциями, места не имеет.

Ниже приводятся наиболее распространенные формулы численного дифференцирования для одномерного случая. Выкладки по выводу формул опущены.

Простейшая формула для производной первого порядка имеет вид:

$$f'_{hi} = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}$$

Оценим ошибку аппроксимации, используя разложение Тейлора в окрестности точки x

$$f'_{hi} = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} =$$

$$= \frac{1}{x_{i+1} - x_i}[f(x) + f'(x)(x_{i+1} - x) +$$

$$+\frac{1}{2}f''(x)(x_{i+1} - x)^2 + O((x_{i+1} - x)^3) -$$

$$-f(x) - f'(x)(x - x_i) - \frac{1}{2}f''(x)(x - x_i)^2 + O((x - x_i)^3)]$$

откуда следует

$$f'_{hi} = f'(x) + \frac{f''(x)[(x_{i+1}) - x)^2 - (x_i - x)^2]}{2(x_{i+1} - x_i)} + O((x_{i+1} - x_i)^2)$$

Ошибка имеет первый порядок для всех точек, кроме середины ячейки $(x_{i+1} + x_i)/2$ и второй порядок в середине. Точки ячеек, в которых производные имеют повышенный порядок точности называают точками суперсходимости или сверхсходимости.

Для равномерной сетки с шагом h формулы второго порядка точности для первой производной в точке x_i имеют вид:

$$f'_h = \frac{f(x_{i+1}) - f(x_{i-1})}{2h}$$

$$f'_h = \frac{4f(x_{i+1}) - f(x_{i+2}) - 3f(x_i)}{2h}$$

$$f'_h = \frac{3f(x_i) - 4f(x_{i-1}) - f(x_{i-2})}{2h}$$

Формула для второй производной в точке x_i на неравномерной сетке имеет первый порядок точности и выглядит так:

$$f''_{hi} = \frac{2}{x_{i+1} - x_{i-1}} \left[\frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} - \frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}} \right]$$

Для равномерной сетки эта формула имеет второй порядок точности.

Метод неопределенных коэффициентов можно применять и в общем многомерном случае, при этом в памяти машины формируется система линейных алгебраических уравнений, выражающих условия коллокации для интерполяционного полинома в узлах шаблона. Решение такой системы уравнений можно предоставить вычислительной машине.

5.3 Естественная аппроксимация

Метод естественной аппроксимации основан на исходных интегральных определениях операторов дифференцирования, известных из курса математического анализа, и на использовании простейших квадратурных формул. Например, для вычисления производных применяется формула Остроградского-Гаусса (теорема

о дивергенции) для преобразования интеграла по ячейке объема V_i в интеграл по ее поверхности S_i

$$\int\limits_{V_i} \nabla \cdot \mathbf{F} dV = \int\limits_{S_i} \mathbf{F} \cdot \mathbf{n} dS$$

где $\mathbf{n}$ - внешняя единичная нормаль к границе. Учитывая теорему Ролля, получаем формулу для вычисления производных

$$[\nabla \cdot \mathbf{F}]_i = \frac{\int\limits_{V_i} \mathbf{F} \cdot \mathbf{n} dS}{\int\limits_{V_i} dV}$$

Дискретные формулы для $\partial f/\partial x, \partial f/\partial y, \partial f/\partial z$ получатся путем подстановки в это соотношение выражений $F = (f, 0, 0)$, $F = (0, f, 0)$, $F = (0, 0, f)$ и заменой интегралов в правой части их аппроксимациями в соответствии с квадратурными формулами. При интегрировании поверхность пространственной ячейки представляется набором треугольных или четырехугольных плоских ячеек.

Для двумерного случая формулы метода естественной аппроксимации принимают вид

$$[\frac{\partial f}{\partial x}]_i = \frac{\int\limits_{l_i} f dy}{\int\limits_{l_i} x dy}, \quad [\frac{\partial f}{\partial y}]_i = \frac{\int\limits_{l_i} f dx}{\int\limits_{l_i} y dx}$$

В отечественной литературе метод естественной аппроксимации производных носит название интегро-интерполяционного метода. Описание этого метода и его оформление в виде теорем можно найти в в сборнике статей ученых Лос-Аламосской лаборатории "Вычислительные методы в гидродинамике"(1967) и в учебнике Годунова и Рябенкого (1968). Этот метод часто используется для построения консервативных аппроксимаций интеральных законов сохранения.

5.4 Метод отображений

Метод отображений, называемый также методом якобианов или изопараметрическим методом, позволяет использовать простейшие аппроксимации производных для прямоугольной равномерной сетки и в случае криволинейных неравномерных сеток. Для этого шаблон или ячейка неравномерной криволинейной сетки отображается на шаблон или ячейку равномерной прямоугольной сетки, на которой производится простейшее численное дифференцирование, а затем с результатом совершается обратное преобразование координат к исходной неравномерной сетке. Поскольку шаблоны и ячейки имеют простую форму, то преобразование координат также является простым и его легко построить.

Рассмотрим, например, определение методом отображений производных для тетраэдральной ячейки, которая определяется 4 узлами $(x_i, y_i, z_i),\ i = 1, 2, 3, 4$. Отобразим ее на каноническую ячейку в трехмерном параметрическом пространстве (декартовых) координат ξ, η, θ так, что узел 1 находится в начале координат, а узлы 2,3,4 находятся на декартовых осях на единичном расстоянии от начала координат. Для канонической ячейки операция дифференцирования тривиальна:

$$\partial_\xi f = f_2 - f_1\ , \quad \partial_\eta f = f_3 - f_1\ , \quad \partial_\theta f = f_4 - f_1$$

По цепному правилу дифференцирования легко найти связь производных в физическом и параметрическом пространствах:

$$\partial_\xi f = \partial_x f \partial_\xi x + \partial_y f \partial_\xi y + \partial_z f \partial_\xi z$$

$$\partial_\eta f = \partial_x f \partial_\eta x + \partial_y f \partial_\eta y + \partial_z f \partial_\eta z$$

$$\partial_\theta f = \partial_x f \partial_\theta x + \partial_y f \partial_\theta y + \partial_z f \partial_\theta z$$

Имеем три уравнения для определения трех искомых производных ($\partial_x f$, $\partial_y f$, $\partial_z f$) в физическом пространстве. Коэффициенты

при неизвестных вычисляются так же легко как и левые части. Получаемые таким способом производные выражаются по правилу Крамера через детерминанты матрицы системы уравнений. Поскольку эта матрица является матрицей Якоби преобразования координат, то способ дифференцирования назван методом якобианов. На одинаковых шаблонах и ячейках формулы метода якобианов могут совпадать с формулами естественной аппроксимации.

Глава 6

Прямые методы решения СЛАУ

При реализации численных методов важным является вопрос о том, как решать возникающие системы алгебраических уравнений. В общем случае такие системы уравнений нелинейны. Решение нелинейных уравнений как правило получается как предел последовательности решений вспомогательных линеаризованных уравнений. Поэтому сначала рассмотривается решение систем линейных алгебраических уравнений (СЛАУ) вида

$$Ax - b = 0$$

где A - матрица системы уравнений, x - вектор неизвестных, b - вектор правой части. Ниже дается описание наиболее важных для практического применения методов.

Под прямыми методами здесь подразумеваются различные варианты метода Гауссова исключения. Такие методы являются точными, поскольку они позволяют в принципе получить точное решение за конечное число операций.

6.1 Подготовка к решению

Симметризация СЛАУ. Симметризация СЛАУ необходима для функционирования некоторых методов решения и заключается в переходе к симметризованной системе уравнений

$$A^T(Ax - b) = 0$$

с симметричной матрицей $A^T A$. Симметризация увеличивает число ненулевых элементов и увеличивает ширину ленты для

ленточных матриц. Обусловленность системы при этом ухудшается, так как

$$cond(A^T A) = (cond(A))^2$$

Несмотря на требующуюся дополнительную вычислительную работу, симметризация часто производится, поскольку задачи с симетричными и положительными матрицами предпочтительны для численного решения (для таких задач решение заведомо существовует и единственно).

Предобусловливание Еще до решения СЛАУ число обусловленности ее матрицы можно уменьшить и тем самым уменьшить чувствительность решения данной алгебраической задачи к возмущениям компонентов матрицы и правой части, а также к ошибкам округления в процессе численного решения. Для этого можно умножить рассматриваемую СЛАУ на приближенную обратную матрицу системы. Такая операция называется предобусловливанием и приводит к новой системе, имеющей то же решение, но лучшие свойства:

$$A_0^{-1}(Ax - b) = 0\ , \quad 1 \le cond(A_0^{-1}A) \le cond(A)$$

Масштабирование неизвестных является простейшим частным случаем предобусловливания, когда приближенная обратная матрица выбирается диагональной, составленной из обратных диагональных элементов исходной матрицы. Подробные примеры масштабирования приведены в книге Форсайта и Молера (1967).

Оптимизация структуры матриц СЛАУ. Устройство матриц СЛАУ, порождаемых при решении задач механики проекционными методами, зависит от выбора базисных функций.

В проекционных методах, использующих глобальные базисные функции, матрицы для коэффициентов разложения решения по базису получаются полностью заполненными, хотя в большинстве случаев абсолютная величина элементов матрицы убывает

по мере удаления от главной диагонали. Иногда удаленными от главной диагонали элементами можно пренебрегать без заметной потери точности решения.

Для локальных (финитных) базисных функций, матрицы получаются редкозаполненными, имеющими большое число нулевых элементов, поскольку скалярные произведения базисных функций, носители которых не пересекаются, равны нулю (носителем функции называется область, в которой она отлична от нуля). Редкозаполненные матрицы свойственны большинству сеточных методов.

Имеется связь между структурой матриц (расположением ненулевых элементов) и нумерацией узлов сетки. Для применения методов исключения желательно иметь ленточные матрицы, так как процесс исключения можно организовать так, что вне ленты ненулевые элементы появляться не будут. Тогда в памяти ЭВМ достаточно будет держать только ненулевую ленточную часть матрицы системы. Поэтому много работ посвящено поискам алгоритмов оптимальной перенумерации узлов для обеспечения минимальной ширины ленты ненулевых элементов.

В настоящее время разработаны безматричные итерационные методы решения, сходящиеся к решению за конечное число итераций и являющиеся точными. Для таких методов проблемы формирования, хранения матриц, оптимизации их структуры, оптимальной нумерации узлов не существуют, поскольку в них матрицы вообще не используются. Поэтому интерес к прямым методам и к методам оптимизации матриц заметно угасает. Даже в задачах на собственные значения наблюдается та же тенденция. Тем не менее, многие современные алгоритмы используют прямые методы и понимание принципов их работы является необходимым.

6.2 Правило Крамера

Известное из курса линейной алгебры правило Крамера имеет вид

$$x_i = \det A_i / \det A$$

где матрицы A_i получается из матрицы A заменой ее i-го столбца столбцом свободных членов.

Правило Крамера дает пример очень неэффективного метода решения СЛАУ с большим числом неизвестных, который характеризуется неприемлемо большим объемом операций, пропорциональным четвертой степени числа неизвестных.

6.3 Методы исключения

В *методе Гаусса* исключение неизвестных производится путем комбинирования уравнений (сложения с умножением на некоторое число) и применяется с учетом структуры матрицы СЛАУ так, чтобы минимизировать число операций с нулевыми элементами и не плодить по возможности, новых ненулевых элементов. Сначала система уравнений преобразуется к виду с нижней треугольной матрицей (прямой ход), а затем она преобразуется к виду с единичной матрицей (обратный ход), в результате решение дается вектором правой части преобразованной системы. Описанная процедура исключения соответствует разложению матрицы A на нижнюю L и верхнюю U треугольные матрицы

$$A = LU$$

При этом прямой ход отвечает умножению исходной СЛАУ слева на обратную к L матрицу

$$U\mathbf{x} = L^{-1}\mathbf{b} = \mathbf{c}$$

а обратный ход состоит в умножении полученного матричного уравнения слева на обратную к U матрицу

$$\mathbf{x} = U^{-1}\mathbf{c}$$

Число операций в метоле Гаусса растет пропорционально кубу числа неизвестных.

Метод Гаусса с выбором главного элемента используется для снижения влияния ошибок округления на решение. Для этого среди элементов $a_{ij}(i, j = 1, ...n)$ выбирается максимальный по модулю, например, a_{pq}, называемый главным элементом. Строка с номером p, содержащая главный элемент, называется главной строкой.

Далее из каждой i-й неглавной строки расширенной матрицы (со столбцом правой части) вычитается главная строка, умноженная на $m_i = a_{iq}/a_{pq}$. В результате получается матрица, у которой все элементы q-го столбца, за исключением a_{pq} , равны нулю. Отбрасывая этот столбец и главную строку, получим новую матрицу с меньшим на единицу числом строк м столбцов. С полученной матрицей описанная выше операция повторяется, пока не получится матрица, содержащая одну строку. Затем все главные строки подвергаются перестановке, приводящей систему уравнений к виду с треугольной матрицей. На этом оканчивается этап прямого хода. Решение полученной системы с треугольной матрицей составляет алгоритм обратного хода.

Прогонка. Для СЛАУ с ленточными матрицами метод Гаусса называется прогонкой. Например, для системы уравнений с трехдиагональной матрицей

$$a_i x_{i-1} + b_i x_i + c_i x_{i+1} = d_i \quad (i = 1, ..., N)$$

$$x_0 = U_0 \ , \quad x_{N+1} = U_1$$

формулы метода прогонки имеют вид:

$$x_i = X_i x_{i+1} + Y_i \quad (i = 1, ..., N)$$

где

$$X_i = \frac{-c_i}{b_i + a_i X_{i-1}}\ , \quad Y_i = \frac{d_i - a_i Y_{i-1}}{b_i + a_i X_{i-1}} \quad (i = 1, ..., N)$$

$$X_0 = 0 \qquad Y_0 = U_0$$

На прямом ходе прогонки определяются коэффициенты X_i и Y_i ($i = 1, ..., N$), а затем на обратном ходе для $i = N, ..., 1$ по формулам метода прогонки определяется искомое решение.

Матричная прогонка. Прогонка для СЛАУ с блочно-ленточными матрицами называется матричной. При этом коэффициенты a_i, b_i, c_i, d_i из предыдущего примера являются квадратными матрицами порядка m, а искомые неизвестные x_i являются векторами размерности m. Формулы метода прогонки сохраняют свой вид, только деление надо понимать как умножение на обратную матрицу. Алгоритм метода прогонки устойчив для матриц с диагональным преобладанием, в которых модуль диагонального элемента в строке больше суммы модулей остальных элементов данной строки. Число операций в методе прогонки растет пропорционально $n^2 m_1$, где m_1 - ширина ленты.

Экономичное вычисление определителей. Для эффективного вычисления определителя $det(A)$ достаточно выполнить прямой ход метода Гаусса и затем найти произведение ведущих (главных) элементов

$$det(A) = a_{11} a_{22}^{(1)} a_{nn}^{(n-1)}$$

где $a_{ii}^{(i-1)}$- значение главного элемента в i-й строке после использования первых (i-1) строк в прямом ходе процедуры исключения.

6.4 Метод квадратного корня

Метод квадратного корня эффективно реализует гауссово исключение для СЛАУ с симметричными положительно опреде-

ленными матрицами, не меняя при этом ширину ленты исходной матрицы СЛАУ (см., например, Копченова и Марон, 1972; Уилкинсон и Райнш, 1976). Положительная симметричная матрица A представляется произведением взаимно транспонированных треугольных матриц:

$$A = \tilde{L}^T \tilde{L}$$

где компоненты матрицы $\tilde{L} = |\tilde{l}_{ij}|$ определяются формулами

$$\tilde{l}_{11} = \sqrt{a_{11}}\ , \quad \tilde{l}_{1j} = \frac{1}{\tilde{l}_{11}} a_{1j} \qquad (j = 2, ..., n);$$

$$\tilde{l}_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} (\tilde{l}_{ki})^2}\ , \quad \tilde{l}_{ij} = \frac{1}{\tilde{l}_{ii}} \left(a_{ij} - \sum_{k=1}^{i-1} \tilde{l}_{ki}^2 \right) \qquad (j = i+1, ..., n)$$

Решение системы уравнений по методу квадратного корня сводится к обращению двух треугольных матриц.

6.5 Метод Холецкого

Иногда метод квадратного корня называют методом Холецкого, хотя в методе Холецкого используется другое разложение, а именно

$$A = LU$$

где отличные от нуля компоненты матриц U и L определяются так:

$$u_{i1} = a_{i1}$$

$$u_{ij} = a_{ij} - \sum_{k=1}^{j-1} u_{ik} l_{kj} \qquad (i \geq j > 1)$$

$$l_{ij} = 1\ , \quad l_{1j} = \frac{a_{1j}}{u_{11}}$$

$$l_{ij} = \frac{1}{u_{ii}}(a_{ij} - \sum_{k=1}^{i-1} u_{ik} l_{kj}) \quad (1 < i < j)$$

а искомый вектор $\mathbf{x}$ вычисляется из уравнений с треугольными матрицами

$$U\mathbf{y} = \mathbf{b} \quad L\mathbf{x} = \mathbf{y}$$

Неполное разложение Холецкого, а также приближенная версия метода квадратного корня используют приближенные треугольные матрицы, вычисленные с пренебрежением компонентами матрицы , расположенными вне ленты заданной ширины. Приближенные обратные матрицы, полученные методами Холецкого и квадратного корня, часто используются для предобусловливания систем алгебраических уравнений.

6.6 Фронтальный метод

Для экономии оперативной памяти ЭВМ метод исключения Гаусса можно реализовать так, что на каждои этапе прямого и обратного хода процесса исключения в оперативной памяти ЭВМ будет находится лишь активная часть матрицы СЛАУ, а остальная часть при этом будет храниться во внешней (дисковой) пямяти. Этот способ решения воплощен во фронтальном методе решения конечноэлементных СЛАУ путем последовательного обхода конечноэлементной сетки элемент за элементом (отсюда произошло название метода - "фронтальный"). Конечно, такая экономия оперативной памяти замедляет процесс решения за счет обменов информацией с внешней памятью. Подробное описание метода можно найти в книге Норри и де Фриза (1981).

6.7 Итерационное уточнение

Из-за плохой обусловленности СЛАУ решение, полученное прямыми методами, содержит погрешности, которые можно уменьшить посредством итерационного уточнения решения. Пусть $\tilde{x}$ - полученное прямым методом приближенное решение СЛАУ. Используя арифметику с двойной точностью, вычисляют невязку

$$\mathbf{r} = \mathbf{b} - A\tilde{\mathbf{x}}$$

а затем решают уравнение

$$A\mathbf{y} = \mathbf{r}$$

относительно y и определяют уточненное решение

$$\mathbf{x} = \tilde{\mathbf{x}} + \mathbf{y}$$

Этот процесс повторяется пока поправка не станет достаточно малой. Если поправка у мала, то можно ожидать, что полученное решение обладает достаточной точностью, в противном случае СЛАУ плохо обусловлена. Более подробно итерационное уточнение обсуждается в книге (Форсайт и Молер, 1969).

Если задача хорошо предобусловлена и метод решения устойчив, то итерационное уточнение не потребуется.

Глава 7

Итерационные методы решения СЛАУ

Значительные упрощения в алгоритмах решения СЛАУ возможны при использовании итерационных методов решения. Современные итерационные методы сильно потеснили прямые методы гауссова исключения, особенно при решении задач с очень большим числом неизвестных, для которых итерационные методы решения не имеют альтернативы.

7.1 Метод простой итерации

Простейший итерационный процесс решения системы алгебраических уравнений носит название метода простой итерации и имеет следующий вид:

$$x^{n+1} = x^n - A_0^{-1}(Ax - b)$$

где A_0 - некоторая легко обратимая матрица, аппроксимирующая матрицу системы уравнений A. Для ошибки $e^n = x^n - x^*$ процесс имеет вид

$$e^n = (I - A_0^{-1}A)^n e^0$$

Условие сходимости, называемое принципом сжимающих отображений, имеет вид

$$|I - A_0^{-1}A|| < 1 \Rightarrow ||e^n|| \leq ||e^0||||I - A_0^{-1}A||^n \to 0$$

Отображение $\Psi(x) = x - A_0^{-1}(Ax - b)$ преобразует решение рассматриваемой системы уравнений в себя $x = \Psi(x)$, поэтому решение называют неподвижной точкой этого отображения.

7.2 Метод Гаусса-Зейделя

Алгоритм метода Гаусса-Зейделя, называемого также методом Либмана или методом последовательных смещений, имеет следующий вид:

1. Задается начальное приближение $x_i^{(0)}$.

2. Реализуется цикл по уравнениям для i=1,2,..,N:

$$x_i^{(n+1)} = a_{ii}^{-1}(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(n+1)} - \sum_{j=i+1}^{N} a_{ij}x_j^{(n)})$$

3. Если $max|x_i^{(n+1)} - x_i^{(n)}| > \varepsilon$, то повторить цикл 2.

Пусть D - диагональная матрица, составленная из диагональных элементов матрицы A, L - нижняя треугольная матрица, составленная из элементов матрицы A исключая главную диагональ, а U - верхняя треугольная матрица из оставшихся элементов A

$$A = L + D + U$$

тогда рассмотренный процесс можно записать кратко так:

$$(D + U)\mathbf{x}^{(n+1)} + L\mathbf{x}^{(n)} = \mathbf{b}$$

Для сходимости матрица $(D + U)^{-1}L$ должна удовлетворять принципу сжимающих отображений.

7.3 Методы последовательной релаксации

Для ускорения сходимости процесс последовательных смещений модифицируется.

$$(D + U)\mathbf{x}^{(n+1)} + \omega L\mathbf{x}^{(n)} = \mathbf{b}$$

Если параметр релаксации ω принимает значения $1 \leq \omega \leq 2$, то имеем метод последовательной верхней релаксации, если же

$0 \le \omega \le 1$, то имеем метод последовательной нижней релаксации. Методы последовательных смещений и релаксации использовались довольно часто на начальной стадии развития численных алгоритмов в 50-60-70 годы 20-го столетия, пока не были вытеснены более эффективными методами исключения и сопряженных градиентов.

7.4 Градиентные методы

Можно построить функционалы, для которых рассматриваемая система уравнений будет выражать условия их минимума. Для положительно определенных симметричных матриц A (то есть таких, что для любого $\mathbf{x} \neq 0$ $A\mathbf{x} \cdot \mathbf{x} > 0$) существует функционал энергии

$$\Psi(\mathbf{x}) = \frac{1}{2} A\mathbf{x} \cdot \mathbf{x} - \mathbf{b} \cdot \mathbf{x}$$

здесь и далее в этой главе точка между векторами будет обозначать скалярное произведение: $\mathbf{a} \cdot \mathbf{b} = \sum a_i b_i$. Для произвольной невырожденной матрицы можно построить положительно определенный функционал нормы невязок

$$\Psi(\mathbf{x}) = (A\mathbf{x} - \mathbf{b}) \cdot (A\mathbf{x} - \mathbf{b})$$

Поскольку умножение системы уравнений на произвольную невырожденную матрицу не меняет решения, то число таких функционалов бесконечно. Метод наискорейшего спуска для минимизации функционала энергии имеет вид

$$\mathbf{x}^{n+1} = \mathbf{x}^n - \alpha_n \mathbf{g}^n \ , \quad \mathbf{g}^n = A\mathbf{x}^n - \mathbf{b} \ , \quad \alpha_n = \frac{\mathbf{g}^n \cdot \mathbf{g}^n}{A\mathbf{g}^n \cdot \mathbf{g}^n}$$

Вещественный параметр $\alpha = \alpha_n$ обеспечивает минимум одномерному функционалу $\Phi(\mathbf{x}) = 0.5A\mathbf{x} \cdot \mathbf{x} - \mathbf{b} \cdot \mathbf{x}$ вдоль линии $\mathbf{x} = \mathbf{x}^n - \alpha \mathbf{g}^n$, то есть определяется из условия $\partial\Phi/\partial\alpha = 0$. Аналогичный метод минимизации функционала невязок называется

методом минимальных невязок и имеет вид

$$\mathbf{x}^{n+1} = \mathbf{x}^n - \alpha_n \mathbf{g}^n \ , \quad \mathbf{g}^n = A\mathbf{x}^n - \mathbf{b} \ , \quad \alpha_n = \frac{\mathbf{g}^n \cdot A\mathbf{g}^n}{A\mathbf{g}^n \cdot A\mathbf{g}^n}$$

Коэффициент $\alpha = \alpha_n$ обеспечивает минимум одномерному функционалу $\Phi(\mathbf{x}) = (A\mathbf{x}-\mathbf{b})\cdot(A\mathbf{x}-\mathbf{b})$ вдоль линии $\mathbf{x} = \mathbf{x}^n - \alpha\mathbf{g}^n$.

Оба описанных градиентных метода очень быстро минимизируют функционалы на первых итерациях, а потом начинают "буксовать", то есть дальнейшее итерирование показывает очень медленную сходимость, делающую применение градиентных методов неэффективным. Это особенно проявляется в случае, когда собственные значения матрицы A сильно различны.

7.5 Метод сопряженных градиентов

Недостаток эффективности градиентных методов устранен в методе сопряженных градиентов, первый вариант которого был предложен Хестенесом и Штифелем в 1952 году. Алгоритмы метода сопряженных градиентов относятся к числу наиболее эффективных методов рещении СЛАУ большой размерности, возникающих при численом решении задач механики сплошных сред. Они решают систему уравнений за конечное число операций. Рассматрим систему линейных алгебраических уравнений

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

Итерационный процесс метода сопряженных градиентов имеет вид

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} - \alpha_n \mathbf{s}^{(n)}$$
$$\mathbf{s}^{(n+1)} = \mathbf{g}^{(n+1)} - \beta_n \mathbf{s}^{(n)}$$

где $n = 0, 1, ..$ и вектор невязки (градиента) определяется соотношениями

$$\mathbf{g}^{(n+1)} = \mathbf{g}^{(n)} - \alpha_n A\mathbf{s}^{(n)}$$

$$\mathbf{g}^{(0)} = A\mathbf{x}^{(0)} - \mathbf{b}$$

коэффициенты α_n и β_n определяются формулами

$$\alpha_n = \frac{\mathbf{g}^{(n)} \cdot \mathbf{s}^{(n)}}{A\mathbf{s}^{(n)} \cdot \mathbf{s}^{(n)}}$$

$$\beta_n = \frac{Ag^{(n+1)} \cdot \mathbf{s}^{(n)}}{A\mathbf{s}^{(n)} \cdot \mathbf{s}^{(n)}}$$

если решение представлено проекциями на A-ортогональный базис $A\mathbf{s}^{(n+1)} \cdot \mathbf{s}^{(n)} = 0$, $\mathbf{g}^{(n+1)} \cdot \mathbf{s}^{(n)} = 0$, и формулами

$$\alpha_n = \frac{\mathbf{g}^{(n)} \cdot A\mathbf{s}^{(n)}}{A\mathbf{s}^{(n)} \cdot A\mathbf{s}^{(n)}}$$

$$\beta_n = \frac{A\mathbf{g}^{(n+1)} \cdot A\mathbf{s}^{(n)}}{A\mathbf{s}^{(n)} \cdot A\mathbf{s}^{(n)}}$$

если решение представлено проекциями на A^TA-ортогональный базис $A\mathbf{s}^{(n+1)} \cdot A\mathbf{s}^{(n)} = 0$, $\mathbf{g}^{(n+1)} \cdot A\mathbf{s}^{(n)} = 0$

В первом случае метод минимизирует функционал энергии, во втором случае - функционал невязок. В первом случае матрица A должна быть знакоопределенной (положительной или отрицательной). Во втором случае достаточно, чтобы матрица A была невырожденной. Свойство симметричности матрицы A в обоих случаях не требуется. Классические формулы для коэффициентов α_n и β_n , предложенные Хестенсом и Штифелем, имеют вид

$$\alpha_n = \frac{\mathbf{g}^{(n)} \cdot \mathbf{s}^{(n)}}{A\mathbf{s}^{(n)} \cdot \mathbf{s}^{(n)}}$$

$$\beta_n = \frac{\mathbf{g}^{(n+1)} \cdot A\mathbf{s}^{(n)}}{A\mathbf{s}^{(n)} \cdot \mathbf{s}^{(n)}}$$

Эти формулы получаются в результате решения на каждой итерации двухпараметрической задачи минимизации функционала и приводит к дополнительным требованиям положительности и симметричности матрицы A. В первых двух вариантах

метода на каждой итерации требуется два умножения матрицы A на вектор (плата за несимметричность), в третьем (классическом) методе только одно, но матрица должна быть симметричной и положительной.

Методы сопряженных градиентов обеспечивают решение задачи за число итераций, не превосходящее числа неизвестных, поскольку одновременно и вырабатывают базис в конечномерном пространстве решения, и находят проекцию на этот базис. При хорошем начальном приближении и при хорошем предобусловливании число итераций, нужных для решения задачи, резко сокращается, так как подобно всем градиентным методам невязка уравнений резко убывает уже в самом начале итерационного процесса.

7.6 Безматричные итерации

Итерационные методы, основанные на вычислении невязок или градиентов, не требуют вычисления матриц СЛАУ. Основной проблемно-ориентированной операцией метода является вычисление невязки условий стационарности минимизируемого функционала, которая реализуется без формирования матрицы системы так же, как вычисляются правые части дифференциальных уравнений при использовании явных схем интегрирования задач Коши. Поэтому проблемы, связанные с хранением матриц и оптимизацией их структуры путем оптимальной перенумерации узлов сетки, в таких методах не возникают вообще, алгоритмы сильно упрощаются по сравнению с прямыми методами. При этом достигается большая экономия в использовании машинной памяти, высокая эффективность и обеспечивается основное свойство, присущее прямым методам - конечность числа операций, необходимых для решения СЛАУ. Для задач высокой размерности как правило применяются безматричные итерационные методы.

Итерационные методы вообще тесно связаны с явными схемами для нестационарных задач: каждый итерационный процесс можно трактовать как некоторый явный метод установления.

Заметим, что неявные схемы для нестационарных задач часто эффективно реализуются с использованием итерационных методов. Наличие хороших начальных приближений (решение на предыдущем временном слое) делает так реализуемые неявные схемы экономичными и асимптотически столь же быстрыми как явные схемы, то есть показывающими сходную скорость роста числа операций в зависимости от размерности задачи (от числа неизвестных).

Глава 8

Нелинейные уравнения

Рассмотрим способы решения нелинейного операторного уравнения:

$$\mathbf{g}(\mathbf{x}) = 0$$

которое является условной записью некоторой системы уравнений. Если это алгебраические уравнения, то $\mathbf{x}$ является искомым вектором с числовыми компонентами, если же операторное уравнение является интегро-дифференциальным, то $\mathbf{x}$ является вектором искомых функций, зависящих от пространственных переменных.

8.1 Метод Ньютона

Итерационный метод Ньютона для нелинейных уравнений основан на разложении нелинейных членов уравнений в ряд Тейлора в окрестности известного приближенного решения $\mathbf{x}^{(n)}$ с удержанием линейной части разложения. Полученная в результате линеаризованная система уравнений относительно нового приближенного решения $\mathbf{x}^{(n+1)}$ имеет вид:

$$\mathbf{g}(\mathbf{x}^{(n+1)}) \approx \mathbf{g}(\mathbf{x}^{(n)}) + \left.\frac{\partial \mathbf{g}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^{(n)}} (\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}) = 0$$

Оператор линеаризованной задачи $\partial \mathbf{g}/\partial \mathbf{x}$ изменяется на каждой итерации.

Модифицированный метод Ньютона подразумевает проведение итераций с использованием постоянного оператора линеаризованной задачи, отвечающего начальному приближению:

$$\mathbf{g}(\mathbf{x}^{(n+1)}) \approx \mathbf{g}(\mathbf{x}^{(n)}) + \left.\frac{\partial \mathbf{g}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^{(0)}} (\mathbf{x}^{(n+1)} - \mathbf{x}^{(n)}) = 0$$

этот метод имеет ограниченное применение для задач с малой нелинейностью.

Линеаризацию нелинейных уравнений для получения решений итерациями по нелинейности обычно проводят на уровне исходной (дифференциальной, интегральной, вариационной) формулировки с использованием дифференциалов Гато (см. книгу Колмогорова и Фомина (1972)):

$$\mathbf{g}'_x(\mathbf{x}_0)(\mathbf{x}_2 - \mathbf{x}_1) = (\partial/\partial\xi[\mathbf{g}(\mathbf{x}_0 + \xi(\mathbf{x}_2 - \mathbf{x}_1))])|_{\xi=0}$$

Применительно
к вариационным и интегро-дифференциальным уравнением метод Ньютона называют методом квазилинеаризации или методом Ньютона-Канторовича. Применительно к нелинейным системам алгебраических уравнений метод Ньютона называют также методом Ньютона-Рафсона.

Обобщающий термин квазиньютоновские методы используется для обозначения семенйства итерационных методов, которые используют лиеаризацию путем разложения нелинейностей в ряд Тейлора в окрестности некоторого элемента в пространстве решений с удержанием линейных членов. Подробности можно найти в книгах Коллатца(1969), Беллмана и Калабы (1968).

8.2 Метод дифференцирования по параметру

В нелинейное уравнение можно ввести внешний параметр λ:

$$\mathbf{g}(\mathbf{x}, \lambda) = 0$$

и искать решение в зависимости от этого параметра $\mathbf{x} = \mathbf{x}(\lambda)$. Для этого продифференцируем исходное уравнение по параметру λ и получим дифференциальное уравнение метода дифференцирования по параметру (Давиденко, 1953):

$$\frac{\partial \mathbf{g}}{\partial \mathbf{x}} \frac{dx}{d\lambda} + \frac{\partial \mathbf{g}}{\partial \lambda} = 0$$

В методе дифференцирования по параметру полагается, что при некотором значении параметра $\lambda = \lambda^{(0)}$ решение задачи $\mathbf{x}^{(0)}$ известно

$$\mathbf{g}(\mathbf{x}^{(0)}, \lambda^{(0)}) = 0$$

и, таким образом, для данного дифференциального уравнения формулируется задача Коши с начальным условием

$$\mathbf{x}|_{\lambda=\lambda^{(0)}} = \mathbf{x}^{(0)}$$

которая может быть проинтегрирована аналитически или численно.

8.3 Метод погружения

Метод погружения заключается в том, что вводится дополнительная эволюционная переменная (фиктивное время) t и соответствующий нестационарный член, который добавляется в исходное уравнение (задача "погружается"в пространство дополнительного измерения, играющего роль времени). Решение ищется как стационарное (установившееся) решение полученной вспомогательной эволюционной задачи

$$\mathbf{g}(\mathbf{x}) = B\frac{\partial \mathbf{x}}{\partial t}\ , \quad \mathbf{x}|_{t=0} = \mathbf{x}_0$$

где B - некоторая положительная невырожденная, легко обращаемая матрица. Правая часть начального условия часто полагается равной нулю. Уравнение метода погружения должно принадлежать параболическому типу, что обеспечивает сходимость

к стационарному решению. Метод погружения нередко называют методом установления.

При определения установившихся решений на основе физических эволюционных уравнений можно ускорить выход на установление с помощью входных параметров, управляющих процессом установления, например, путем выбора матрицы B и начального значения x_0. Если интересует только установившееся решение, то можно не заботиться о соответствии промежуточных решений физической реальности.

При решении уравнений метода погружения нелинейные члены уравнения $\mathbf{g}$ нередко линеаризуются по методу Ньютона на каждом шаге по времени относительно решения на старом временном слое.

Глава 9

Единственность и ветвление решений

9.1 Теорема о неявной функции

В курсах функционального анализа доказывается теорема о неявной функции: неявная функция $x(\lambda)$ определяемая из решения нелинейного уравнения

$$\mathbf{g}(\mathbf{x}, \lambda) = 0$$

имеет единственное продолжение в малой окрестности точки $(\mathbf{x}_0, \lambda_0)$, где $\mathbf{g}(\mathbf{x}_0, \lambda_0) = 0$, если оператор ($\mathbf{g}'_{\mathbf{x}}$) линеаризованной задачи

$$\mathbf{g}(\mathbf{x}_0, \lambda_0) + \mathbf{g}'_{\mathbf{x}}|_{(\mathbf{x}_0,\lambda_0)} (\mathbf{x} - \mathbf{x}_0) + \mathbf{g}'_{\lambda}|_{(\mathbf{x}_0,\lambda_0)} (\lambda - \lambda_0) = 0$$

невырожден.

Приведенная теорема имеет место в общем случае функциональных уравнений, так что под нелинейным уравнением, о котором идет речь, можно понимать нелинейную начально-краевую задачу, систему нелинейных интегральных уравнений, или, например, систему нелинейных алгебраических уравнений. При этом x обозначает набор искомых функций рассматриваемой задачи или их дискретный аналог.

Самой простой для понимания излагаемых далее основ теории ветвления решений нелинейных уравнений является алгебраическая трактовка теоремы о неявной функции.

Параметр λ связывается с интенсивностью процессов в сплошной среде. Например, в механике деформируемых твердых тел

роль параметра λ исполняет параметр нагрузки, в задачах механики жидкости роль λ отводится числу Рейнольдса. Если оператор линеаризованной задачи $\mathbf{g}'_x(\mathbf{x_0}, \lambda_0)$ вырождается, то точка $(\mathbf{x_0}, \lambda_0)$ называется особой точкой и вопрос о возможном продолжении решения нетривиален и рассматривается далее. В континуальной механике анализ поведения решения в особых точках и их обнаружение составляет предмет теории устойчивости тонкостенных конструкций и теории гидродинамической устойчивости.

9.2 Особые точки и продолжение решений

Особыми называются точки $(\mathbf{x}, \lambda)$ в пространстве решений, в которых одно или несколько собственных значений оператора линеаризованной задачи $\mathbf{g}'_x(\mathbf{x}, \lambda)$ обращаются в нуль.

Если имеется только одно, равное нулю, собственное значение и соответствующая собственная функция удовлетворяет неоднородной линеаризованной задаче, то неявная функция $\mathbf{x}(\lambda)$ имеет единственное продолжение и эта собственная функция указывает направление продолжения. Соответствующая точка называется предельной точкой. Таковы точки, описывающие состояние цилиндрических панелей перед прощелкиванием к новому устойчивому положению равновесия при действии внешнего давления. Соответствующее явление хорошо известно всем, кто когда-либо играл, щелкая кусочком фотопленки.

Если же собственная функция не удовлетворяет неоднородной линеаризованной задаче, то неявная функция $\mathbf{x}(\lambda)$ не имеет продолжения в направлении данной собственной функции. При отсутствии возможных продолжений решения особая точка называется точкой умирания решения или тупиковой точкой.

Если оператор линеаризованной задачи имеет несколько нулевых собственных чисел, то соответствующие собственные решения могут указывать несколько направлений продолжения

неявной функции при условии, что они удовлетворяют неоднородной линеаризованной задаче. Соответствующая особая точка называется точкой ветвления решений нелинейной задачи. Такие случаи встречаются и в теории устойчивости и выпучивания деформируемых тел, и в теории устойчивости гидродинамических течений. Хорошим примером может служить случай потери устойчивости сжатого стержня, когда при достижении критической нагрузки наряду с прямолинейной формой равновесия становится возможной и изогнутая форма равновесия.

Заметим, что могут существовать и изолированные ветви решения задачи о неявной функции, на которые нельзя попасть непрерывно продолжая решение. Иногда их удается обнаружить в численных экспериментах "методом тыка"(перебором вероятных значений). Для нелинейных задач теории тонких упругих оболочек такие решения были найдены и описаны в монографии [Валишвили. 1979].

Подробный анализ ветвления решений нелинейных уравнений имеется в сборнике статей [Келлер, 1974].

Глава 10

Методы минимизации функционалов

Системы алгебраических уравнений часто представляют условия минимума или стационарности (условия Эйлера) для некоторых функционалов, возникающих в задачах оптимизации процессов и конструкций. Такие задачи составляют предмет теории математического (линейного и нелинейного) программирования. Подробное изложение теории математического программирования можно найти, например, в книгах Полака (1974) и Пшеничного,Данилина (1979), а также в сборнике статей американских специалистов (Методы условной минимизации, 1977). Ниже приводятся основные положения этой теории.

10.1 Условная минимизация линейных функционалов

Сначала рассмотрим задачи минимизации линейных функционалов. Несмотря на линейность рассматриваемых функционалов их минимизация является существенно нелинейной задачей, так как область допустимых решений определяется набором некоторых ограничений типа равенств и неравенств.

Например, вес летательного аппарата представляется суммой весов его составных частей. Аргументами функционала веса являются параметры геометрии, удельные веса материалов и тому подобные характеристики, каждая из которых имеет определенные границы изменения, зависящие возможно от значений других параметров. Минимум функционала может быть неединственным, задача его поиска является нетривиальной и, конечно же, не сводится к решению системы линейных уравнений.

Другой пример приложения теории линейного программирования дает задача поиска равнопрочных конструкций, все составляющие которых имеют одинаковый запас прочности, т.е. отношение предела прочности к максимальному напряжению при эксплуатации. Такие функционалы как правило не имеют аналитического представления и определяются алгоритмически по результатам решения вспомогательных краевых задач, описывающих напряженно-деформированное состояние элементов конструкции.

Каноническая форма задачи линейного программирования имеет вид: найти минимум линейного функционала F

$$\min_{\mathbf{x}} F(\mathbf{x}) = \mathbf{c}^T \cdot \mathbf{x}$$

при ограничениях равенствах

$$A\mathbf{x} = \mathbf{b} \ , \quad \mathbf{x} \geq 0$$

где $\mathbf{c}^T$ - заданный вектор размерности n, $\mathbf{x}$ - вектор неизвестных размерности n, - матрица размерности $m \times n$ $(m < n)$, m - число ограничений. Заметим, что любое ограничение неравенство

$$B_{i1}x_1 + ... + B_{in}x_n > 0$$

введением дополнительной переменной $x_{n+1} \geq 0$ сводится к ограничению равенству

$$B_{i1}x_1 + ...B_{in}x_n - x_{n+1} = 0$$

поэтому полагаем, что все ограничения имеют вид равенств. Векторы неизвестных $\mathbf{x}$, удовлетворяющие ограничениям, образуют допустимое множество пробных решений R

$$R = \{\mathbf{x} : A\mathbf{x} = \mathbf{b} \ , \quad \mathbf{x} \geq 0\}$$

Каждое из равенств, выражающих ограничения, определяет в n-мерном пространстве решений (-1)-мерную гиперплоскость, пересечение которой с допустимым множеством R дает выпуклый

$(n-1)$ - мерный многогранник G. Минимальное значение целевого функционала достигается в некоторой вершине многогранника G. При вырождении оно может достигаться во всех точках некоторого ребра или некоторой грани многогранника. Для решения задачи линейного программирования надо найти значение $\mathbf{x}$ в вершине с наименьшим значением целевого функционала F.

Отыскание решения перебором пробных значений $\mathbf{x}$ в вершинах многогранников G неэффективно, так как вершин слишком много. Эффективный способ решения носит название симплекс-метода (см. Методы условной минимизации, 1977) и состоит в следующем. Сначала значение функционала F определяется в произвольной вершине произвольного многогранника G и сравнивается со значениями F в соседних вершинах на концах ребер, выходящих из данной вершины. Далее вдоль ребра, по которому функционал F убывает, попадают в следующую вершину и процесс повторяют. Если вдоль всех выходящих из вершины ребер функционал F возрастает, то минимум считается достигнутым в данной вершине. Описанный процесс резко сокращает число вершин, в которых надо вычислить значения целевого функционала для решения задачи.

10.2 Минимизация нелинейных функционалов

Методы безусловной минимизации квадратичных функционалов, имеющих системы линейных алгебраических уравнений в качестве условий минимума, уже рассмотрены выше в разделах по градиентным методам решения систем линейных алгебраических уравнений.

Методы минимизации нелинейных функционалов общего вида при наличии ограничений составляют предмет теории нелинейного програмирования. Постановка общей задачи нелинейного программирования имеет вид:

найти точку $\mathbf{x}$ минимума функционала F

$$\min_x F(\mathbf{x})$$

при ограничениях неравенствах

$$c_i(\mathbf{x}) > 0\ , \quad i = 1, ..., m_1$$

и при ограничениях равенствах

$$d_i(\mathbf{x}) = 0, \quad i = 1, ..., m_2$$

где размерность вектора неизвестных равна $n > m = m_1 + m_2$. При числе ограничений $m = 0$ имеем задачу безусловной минимизации, в противном случае решаем задачу условной минимизации. Векторы неизвестных $\mathbf{x}$, удовлетворяющие ограничениям, образуют допустимое множество R пробных решений.

$$R = \{\mathbf{x} : c_i(\mathbf{x}) \geq 0\ (i = 1, ..., m_1)\ \ or\ \ d_i(\mathbf{x}) = 0\ (i = 1, ..., m_2)\}$$

Рассмотрим основные методы сведения задач условной минимизации к последовательности задач безусловной минимизации: метод множителей Лагранжа, метод штрафных функций и барьерный метод.

10.3 Метод множителей Лагранжа

Теорема Куна-Таккера. Если $\mathbf{x}^*$ является решением задачи нелинейного программирования, то найдется вектор λ^* размерности m такой, что

$$\mathbf{g}(\mathbf{x}^*) - \sum_{i=1}^{m_1} \lambda_1^* \mathbf{a}_i(\mathbf{x}^*) - \sum_{i=1}^{m_2} \lambda_2^* \mathbf{b}_i(\mathbf{x}^*) = 0$$

$$\lambda_1^* \geq 0$$

где через $\mathbf{g}$, $\mathbf{a}_i$ и $\mathbf{b}_i$ обозначены градиенты функций F, c_i и d_i соответственно. При взаимно независимых градиентах функций ограничений векторы множителей Лагранжа $\lambda_\mathbf{1}$ и $\lambda_\mathbf{2}$ определяются однозначно. Учет ограничений с помощью множителей Лагранжа носит название метода множителей Лагранжа.

10.4 Методы штрафных и барьерных функций

В методе штрафных функций вместо исходной задачи условной минимизации рассматривается модифицированная задача безусловной минимизации функционала

$$T(\mathbf{x}, \mathbf{r}) = F(\mathbf{x}) + \Phi(c(\mathbf{x}), \mathbf{r})$$

где r - вектор управляющих параметров, Φ - положительная штрафная функция, регулируемая вектором $\mathbf{r}$. Безусловный локальный минимум функционала T по $\mathbf{x}$ обозначается $\mathbf{x}(\mathbf{r})$. Различные методы штрафных функций отличаются выбором штрафного функционала Φ и последовательности управлений $\mathbf{r}^{(k)}$, обеспечивающих сходимость $\mathbf{x}(\mathbf{r}^{(k)})$ к $\mathbf{x}^*$ при $k \to \infty$.

В методе штрафных функций приближенные решения могут не принадлежать допустимому множеству пробных решений, то есть ограничения типа равенств выполняются с погрешностью, которая постепенно стремится к нулю с ростом k.

Метод барьерных функций применяется для выполнения ограничений типа неравенств, функционал модифицируется так, чтобы на границе допустимого множества "построить барьер", препятствующий нарушению ограничений в процессе безусловной минимизации целевого функционала T по $\mathbf{x}$, и чтобы точки $\mathbf{x}(\mathbf{r}^{(k)})$ сходились к $\mathbf{x}^*$.

Конкретные примеры штрафных и барьерных функций приводятся далее при рассмотрении способов учета ограничений в задачах о несжимаемых средах, в контактных задачах, в задачах построения сеток с выпуклыми ячейками.

10.5 Метод локальных вариаций

Для минимизации негладких функционалов был разработан метод локальных вариаций, широко используемый в механике деформируемого тела, теории управления и оптимизации (Баничук, Картвелишвили, Черноусько, 1973).

На каждой итерации метода локальных вариаций функционалы минимизируются последовательно по отдельным компонентам вектора неизвестных. Значение каждого отдельного неизвестного увеличивается и уменьшается на некоторую фиксированную величину приращения (проводится "локальная вариация" функционала). За новое значение этого неизвестного принимается то, которое приводит к уменьшению функционала. Если перебор новых значений неизвестных не уменьшает функционал, то величина приращения уменьшается вдвое и процесс локальных вариаций продолжается до тех пор, пока миниммум функционала не будет найден для достаточно малого приращения неизвестных. Для экономии вычислений при минимизации интегральных функционалов континуальной механики методом локальных вариаций используется тот факт, что вариация узлового неизвестного меняет только ту часть функционала, которая определяется интегрированием по окрестности этого узла. Метод локальных вариаций применим к произвольным негладким положительным функционалам, поскольку не содержит операций дифференцирования функционала. К сожалению, скорость сходимости метода локальных вариаций невысока, что является платой за универсальность.

Глава 11

Методы решения задач Коши

11.1 Постановка задач Коши

Рассмотрим задачу Коши для системы обыкновенных дифференциальных уравнений (ОДУ) первого порядка

$$d\mathbf{y}/dt = \mathbf{f}(\mathbf{y}, t) \ , \quad \mathbf{y}|_{t=0} = \mathbf{y}^0$$

где t - временная координата (или независимая переменная), $\mathbf{y}$ - искомые функции, функция $\mathbf{f}$ и значение $\mathbf{y}^0$ заданы. Численное решение ишетса путем пошагового интегрирования уравнений для дискретных значений времени $t_0 < t_1 < ... < t_n < ...$, представляющих временные слои.

О свойствах системы ОДУ судят по поведению решений однородной линеаризованной системы дифференциальных уравнений, полученной из исходной нелинейной системы уравнений с помощью операции квазилинеаризации путем разложения нелинейных членов в ряд Тейлора в окрестности некоторого приближенного решения $\mathbf{y} = \mathbf{y}^n$ с удержанием линейной части разложения. Линеаризованная система уравнений имеет вид

$$d\mathbf{y}/dt = \mathbf{f}(\mathbf{y}_n, t) + \partial\mathbf{f}/\partial\mathbf{y}|_{\mathbf{y}=\mathbf{y}^n}(\mathbf{y} - \mathbf{y}^n)$$

Взаимно независимые решения однородной линеаризованной системы уравнений

$$d\mathbf{y}/dt = \partial\mathbf{f}/\partial\mathbf{y}|_{\mathbf{y}=\mathbf{y}^n}\mathbf{y}$$

называются фундаментальными и ищутся в виде экспонент $\mathbf{y} = \mathbf{y}^* e^{\lambda t}$. В результате подстановки этого выражения в однородную

систему линеаризованных дифференциальных уравнений получаем однородную систему линейных алгебраических уравнений для определения вектора $\mathbf{y}^*$

$$(\partial f/\partial \mathbf{y}|_{\mathbf{y}=\mathbf{y}^n} - E\lambda)\mathbf{y}^* = 0$$

Показатели экспонент λ определяются из решения характеристического уравнения:

$$\det(\partial f/\partial \mathbf{y}|_{\mathbf{y}=\mathbf{y}^n} - \lambda E) = 0$$

являющегося условием существования нетривиального решения алгебраической системы уравнений.

Если все фундаментальные решения убывающие, то есть, если все показатели экспонент отрицательны, то исходная система уравнений является устойчивой. В противном случае среди фундаментальных решений имеются неограниченно возрастающие и решение исходной системы уравнений имеет смысл только на ограниченном интервале времени, для которого малым изменениям в начальных данных будут отвечать достаточно малые изменения в решении (требование устойчивости задачи).

Разностные методы решения задач Коши для системы ОДУ, служашие для определения значений искомых функций для набора лискретных значений аргумента, отвечающих узлам сетки, называются разностными схемами. Приведем несколько важных определений, характеризующих основные разновидности разностных схем.

Явная схема представляется системой уравнений относительно величин на новом временном слое $t = t_{n+1}$, которая характеризуется диагональной матрицей и легко (явно) разрешается.

Неявная схема содержит значения функции правой части на новом временном слое $t = t_{n+1}$ и требует для определения величин на новом временном слое решения системы алгебраических уравнений.

Двух-, трех-,..., много- слойная схема использует соответствующее число временных слоев для аппроксимации временных производных.

Одно-, двух-, ..., много- шаговая схема использует соответствующее число промежуточных вспомогательных шагов (промежуточных вычислений функции правой части) на каждом шаге по времени.

11.2 Явные методы Рунге-Кутты

Методы Рунге-Кутты реализуют повышение точности аппроксимации дифференциального уравнения на шаге по времени ($\Delta t_n = t_{n+1} - t_n$) за счет увеличения числа промежуточных вычислений функции правой части. Ниже приводятся варианты методов Рунге-Кутты, расположенные по порядку точности.

В явной схеме Эйлера (одношаговый метод Рунге-Кутты)

$$\mathbf{y}^{n+1} = \mathbf{y}^n + \mathbf{f}^n \Delta t_n$$

ошибка убывает со скоростью $O(\Delta t)$.

В явной схеме Эйлера с пересчетом (двухшаговый метод, Рунге- Кутты, называемый: в западной литературе методом Хойна)

$$\tilde{\mathbf{y}}^{n+1} = \mathbf{y}^n + \mathbf{f}^n \Delta t_n$$

$$\mathbf{y}^{n+1} = \mathbf{y}^n + (\mathbf{f}^n + \tilde{\mathbf{f}}^{n+1}) \Delta t_n / 2$$

ошибка убывает со скоростью $O(\Delta t^2)$.

В явной схеме предиктор-корректор второго порядка точности (двухшаговая схема Рунге-Кутты)

$$\tilde{\mathbf{y}}^{n+1} = \mathbf{y}^n + \mathbf{f}^n \Delta t_n / 2$$

$$\mathbf{y}^{n+1} = \mathbf{y}^n + \mathbf{f}^{n+1/2} \Delta t_n$$

ошибка убывает со скоростью $O(\Delta t^2)$. Классический метод Рунге-Кутты четвертого порядка точности выражается формулой (четырехшаговый метод)

$$\mathbf{y}^{n+1} = \mathbf{y}^n + \frac{1}{6}(\mathbf{k}_0 + 2\mathbf{k}_1 + 2\mathbf{k}_2 + \mathbf{k}_3)$$

где

$$\mathbf{k}_0 = \Delta t_n \mathbf{f}(\mathbf{y}^n, t_n)$$

$$\mathbf{k}_1 = \Delta t_n \mathbf{f}(\mathbf{y}^n + \mathbf{k}_0/2, t_n + \Delta t_n/2)$$

$$\mathbf{k}_2 = \Delta t_n \mathbf{f}(\mathbf{y}^n + \mathbf{k}_1/2, t_n + \Delta t_n/2)$$

$$\mathbf{k}_3 = \Delta t_n \mathbf{f}(\mathbf{y}^n + \mathbf{k}_2, t_n + \Delta t_n)$$

Методы Рунге-Кутты более высоких порядков точности приводятся в справочниках.

Все показанные выше двухслойные многошаговые схемы Рунге- Кутты выводятся применением квадратурных формул численного интегрирования к формуле аналитического представления решения задачи Коши на шаге по времени

$$\mathbf{y}^{n+1} = \mathbf{y}^n + \int_{t_n}^{t_{n+1}} \mathbf{f} dt$$

Например, схема Эйлера отвечает квадратурной формуле прямоугольников, схема Эйлера с пересчетом - квадратурной формуле трапеций, схему третьего порядка точности можно получить, применяя квадратурную формулу Симпсона, и так далее.

11.3 Явные методы Адамса

В схемах Адамса повышение точности достигается за счет увеличения числа временных слоев, используемых для аппроксимации дифференциального уравнения. Они более экономичны, так

как используют уже вычисленные значения функции правой части, но требуют постоянного шага по времени. Простейшей схемой этой группы является явная двухслойная схема Эйлера первого порядка точности. Следующей является трехслойная схема квазивторого порядка точности:

$$\mathbf{y}^{n+1} = \mathbf{y}^n + ((1+\alpha)\mathbf{f}^n - \alpha\mathbf{f}^{n-1})\Delta t_n$$

которая при $\alpha = 0$ отвечает схеме Эйлера первого порядка точности, а при $\alpha = 0.5$ имеет второй порядок точности и называется схемой Адамса-Башфорта. Схемы более высокого порядка точности описаны в справочниках (см. книгу Камке) Схема Адамса 4-го порядка имеет следующий вид:
предиктор

$$\tilde{\mathbf{y}}^{n+1} = \mathbf{y}^n + \Delta t_n/24(55\mathbf{f}^n - 59\mathbf{f}^{n-1} + 37\mathbf{f}^{n-2} - 9\mathbf{f}^{n-3})$$

корректор

$$\mathbf{y}^{n+1} = \mathbf{y}^n + \Delta t_n/24(9\tilde{\mathbf{f}}^{n+1} + 19\mathbf{f}^n - 5\mathbf{f}^{n-1} + f^{n-2})$$

Схемы Адамса выводятся с использованием интерполяции Лагранжа подинтегральной функции f

$$\mathbf{y}^{n+1} = \mathbf{y}^n + \int_{t_n}^{t_{n+1}} \mathbf{f}dt$$

по ее k значениям, $\mathbf{y}^{n-k+1}, \mathbf{y}^{n-k+2}, ..., \mathbf{y}^n$, предшествующих искомому $\mathbf{y}^{n+1}$.

11.4 Неявные схемы для жестких задач

Неявные аппроксимации применяются для жестких систем ОДУ, характеризующихся тем, что матрицы $\partial f/\partial \mathbf{y}|_{\mathbf{y}=\mathbf{y}_n}$ соответсвующих линеаризованных систем ОДУ плохо обусловлены

и, следовательно, такие системы ОДУ имеют сильно различающиеся по величине скорости изменения фундаментальных решений (даже в случае устойчивых систем уравнений) или просто очень быстро меняющиеся фундаментальные решения ($\mathbf{y} = ce^{\lambda t}$, $max|\lambda|T >> 1$, [0,T] - интервал интегрирования).

Явные схемы для жестких систем уравнений требуют очень сильных ограничений на шаг по независимой переменной ($\Delta t < 1/|\lambda|$), диктуемых быстро меняющимися фундаментальными решениями, и .неэффективны, если надо получить решение на больших интервалах времени [0,T], описываемое в основном медленно меняющимися фундаментальными решениями и заданными правыми частями. Проведение расчета по явной схеме с шагом, превышающим упомянутое ограничение на шаг по времени, немедленно приводит к неустойчивости.

Простейший пример задачи Коши для жесткого дифференциального уравнения можно искусственно построить так. Пусть решением является функция $y = \cos(t)$. Хорошо обусловленное уравнение для этого решения получается непосредственным дифференцированием принятого решения: $dy/dt = -\sin(t)$. Сделаем это уравнение жестким добавив член, равный на решении нулю, с большим коэффициентом:

$$\frac{dy}{dt} = -100(y - \cos(t)) - \sin(t)$$

и дополним полученное уравнение начальным условием, вид которого также диктуется желанием сделать функцию $\cos(t)$ решением рассматриваемой задачи:

$$y|_{t=0} = 1$$

Фундаментальное решение данной задачи характеризуется показателем роста $\lambda = -100$ и на интервале $t \in [0, T]$ $(T = 1)$ является быстро изменяющимся ($|\lambda|T = 100 >> 1$). Это фундаментальное решение является убывающим и, следовательно, задача Коши устойчива. Однако она является жесткой. Попытка

решения такой задачи по явной схеме с шагои по времени, превышающим $1/|\lambda| = 0.01$ обречена на неудачу: небольшим изменениям в начальных данных будут отвечать громадные изменения в значении решения в конце интервала изменения независимой координаты t. Это легко проверяется непосредственным вычислением.

Ярким примером жестких систем уравнений является система обыкновенных дифференциальных уравнений по времени, возникающая для каркасов приближенных решений гиперболических систем уравнений в частных производных при использованиии проекционных методов. В частности, для явных сеточных методов решения гиперболичнских уравнений шаг интегрирования по времени ограничен условием устойчивости Куранта, которое требует, чтобы за один шаг по времени сигнал от

данного узла не вышел бы за пределы его окрестности, образованной соседними узлами пространственной сетки. В этом примере можно теоретически обосновать и определить ограничение на шаг по времени, гарантирующее устойчивый расчет по явным схемам ($\Delta t < h/c$, где - скорость распространения малых возмущений).

Однако, во многих случаях, возникающих в приложениях, применение явных схем невозможно, так как требуется определить решение на итервале времени, значительно превышающем ограничение на шаг по времени в явных схемах. В этих случаях имеется потребность в специальных методах интегрирования, реализующих безусловно устойчивый расчет решений жестких систем ОДУ. Ключевым средством обеспечения безусловной устойчивости является применение неявных схем решения. Типичными примерами методов для жестких уравнений служат следующие:

1) неявная схема Эйлера (простейший вариант метода Гира, на-

зываемый также обратным методом Эйлера):

$$\mathbf{y}^{n+1} = \mathbf{y}^n + \mathbf{f}(\mathbf{y}^{n+1}, t_{n+1})\Delta_n$$

2) неявная схема Кранка-Николсона:

$$\mathbf{y}^{n+1} = \mathbf{y}^n + ((1-\alpha)\mathbf{f}(\mathbf{y}^n, t_n) + \alpha\mathbf{f}(\mathbf{y}^{n+1}, t_{n+1}))\Delta_n$$

3) неявная квазиньютоновская схема (линеаризованная схема Кранка-Николсона)

$$\mathbf{y}^{n+1} = \mathbf{y}^n + (\mathbf{f}(\mathbf{y}^n, t_n) + \alpha \left.\frac{\partial \mathbf{f}}{\partial \mathbf{y}}\right|_n (\mathbf{y}^{n+1} - \mathbf{y}^n))\Delta_n$$

Для жестких дифференциальных уравнений применяются также неявные многослойные схемы Адамса-Башфорта.

Схемы Кранка-Николсона безусловно устойчивы при $\alpha > 0.5$ и переходят в явную схему Эйлера при $\alpha \to 0$. Подробнее о методах решения жестких систем ОДУ можно прочитать в книге Форсайта и др. (1980).

Системы нелинейных алгебраических уравнений, к которым приводят неявные схемы, решаются обычно с помошью каких либо вариантов итерационного метода Ньютона. Если шаг по времени мал, то часто хватает одной итерации по методу Ньютона на каждом временном шаге, как это и делается в записанной выше квазиньютоновской схеме.

Помимо ограничений, связанных со скоростью изменения решений, шаг интегрирования для явных и неявных схем подчиняется условию точности путем сравнения результатов расчетов на вложенных сетках (то есть на сетках с шагами Δt_n и $\Delta t_n/2$) и поддержания разности таких решений достаточно малой за счет уменьшения шага Δt_n.

Глава 12

Решение эллиптических уравнений

12.1 Формулировка задачи

В этом разделе рассмотрим основные способы решения краевых задач для эллиптических уравнений в частных производных второго порядка. Постановка типичной задачи имеет следующий вид. В некоторой пространственной области V с границей S требуется найти функцию $\mathbf{u}(\mathbf{x})$, удовлетворяющую уравнению

$$\nabla \cdot (\mathbf{A}(\mathbf{u}) \cdot \nabla \mathbf{u}) + C(\mathbf{u}) = 0 \tag{1}$$

и граничным условиям

$$\mathbf{x} \in S_u \ : \quad \mathbf{u} = \mathbf{u}_*(\mathbf{x}) \tag{2}$$

$$\mathbf{x} \in S \backslash S_u \ : \quad \mathbf{n} \cdot \mathbf{A} \cdot \nabla \mathbf{u} = P_n^*(\mathbf{x}) \tag{3}$$

где тензор коэффициентов $\mathbf{A}(\mathbf{u})$ и источниковый член $C(\mathbf{u})$ являются заданными функциями искомого решения $\mathbf{u}$, вектор $\mathbf{n}$ является единичной внешней нормалью к границе S.

Граничные условия (2), заданные на S_u, выражают ограничения на искомую величину $\mathbf{u}$ и называются условиями Дирихле, а условия (3), заданные на оставшейся части границы $S \backslash S_u$, определяют граничные значения нормальной составляющей (диффузионного) потока $\mathbf{A} \cdot \nabla \mathbf{u}$ и называются условиями Неймана. Если в формулировке задачи участвуют оба типа условий, то краевая задача является смешанной, если $S_u = S$, то имеем задачу Дирихле, если же $S_u = \emptyset$, то имеем задачу Неймана.

Искомая функция $\mathbf{u}$ может трактоваться как скаляр, вектор или тензор второго ранга (и так далее). Задача поставлена корректно, если выполнено условие Адамара, требующее положительности оператора A:

$$\forall \nabla \mathbf{u} \neq 0 \ : \quad (\mathbf{A} \cdot \nabla \mathbf{u}) \cdot \nabla \mathbf{u} > 0 \tag{4}$$

В случае задачи Неймана ($S_u = \emptyset$) дополнительно требуется, чтобы правая часть граничного условия для потока (3) была бы согласована со свободным членом

$$\int\limits_{S \setminus S_u} P_n dS + \int\limits_V C dV = 0 \tag{5}$$

Это ограничение является следствием интегрального закона сохранения величины $\mathbf{u}$, получаемого интегрированием исходного уравнения по области решения и преобразованием интеграла по области от потокового члена к интегралу по границе с помощью теоремы о дивергенции Остроградского-Гаусса.

Здесь и далее используется сокращенная запись уравнений. Уравнение (1)

$$\nabla \cdot (\mathbf{A}(\mathbf{u}) \cdot \nabla \mathbf{u}) + C(\mathbf{u}) = 0 \tag{1}$$

в развернутой форме выглядит так

$$\sum_{i=1}^{3} \partial_i \Big(\sum_{j=1}^{3} \sum_{\beta=1}^{N} A^{ij}_{\alpha\beta}(\mathbf{u}) \partial_j u_\beta \Big) + C_\alpha(\mathbf{u}) = 0 \tag{1'}$$

где N - число искомых функций u_α $(\alpha = 1, ..., N)$, $\partial_i = \partial / \partial x_i$. Сравнивая (1) и (1') видим, что использование сокращенной записи делает изложение более ясным.

Здесь и далее выражения вида $\mathbf{a} \cdot \mathbf{b}$ используются для обозначения скалярного (внутреннего) произведения в смысле тензорного анализа.

Характерным свойством решений эллиптических краевых задач является то, что изменение в условиях задачи приводит к изменению решения сразу повсюду в области решения.

Следующие примеры показывают, что критерий корректности Адамара является вполне естественным с физической точки зрения.

Если u - температура, то рассматриваемая задача описывает процесс теплопроводности, а критерий Адамара сводится к требованию положительности тензора коэффициентов теплопроводности. Роль потоков играют тепловые потоки.

Если $\mathbf{u}$ - перемещение, то рассматриваемая задача описывает равновесие упругого тела, а критерий Адамара сводится к требованию положительности тензора модулей упругости. Роль потоков играют упругие (консервативные) напряжения.

Если $\mathbf{u}$ - скорость, то рассматриваемая задача описывает стационарное течение вязкой жидкости в приближении Стокса, а критерий Адамара сводится к требованию положительности тензора коэффициентов вязкости. Роль потоков играют вязкие (диссипативные) напряжения.

Если $\mathbf{u}$ - концентрация, то рассматриваемая задача описывает процесс диффузии примеси, соответственно критерий Адамара заключается в требовании положительности тензора коэффициентов диффузии. Роль потоков играют диффузионные потоки.

Список можно легко расширить, так как очень многие физические и геометрические задачи относятся к эллиптическому типу. Кроме того, эллиптические члены содержатся во многих более общих уравнениях. Нередко они вводятся в уравнения в качестве регуляризаторов для обеспечения свойств существования и единственности решений. Поэтому методы решения эллиптических задач заслуживают пристального внимания. Поскольку методов много, то здесь описываются лишь их схемы реализации, вопросы же исследования и обоснования сознательно опускаются или освещаются качественно без математических выкладок.

Это позволяет рассмотреть идеи многих возможных подходов к решению без утомительного погружения в детали.

Если в исходном уравнении дописать члены с первыми пространственными производными от искомой функции, умноженными на заданные коэффициенты, или даже члены, являющиеся нелинейными функциями от первых производных искомой функции, то формально тип уравнения не изменится. Однако, делать этого в данном разделе не будем, поскольку такая модификация может радикально изменить поведение решений и даже сделать рассматриваемую задачу некорректной. Модифицированные таким образом задачи потребуют специальных методов решения.

Имеется важная особенность формулировки исходной задачи, связанная с консервативной и неконсервативной формами записи. Консервативная запись исходного уравнения (1)

$$\nabla \cdot (\mathbf{A} \cdot \nabla \mathbf{u}) + C(\mathbf{u}) = 0 \tag{6}$$

для произвольного объема $\tilde{V}$ с поверхностью $\tilde{S}$ поддерживает закон сохранения величины $\mathbf{u}$:

$$\int\limits_{\tilde{S}} \mathbf{n} \cdot (\mathbf{A} \cdot \nabla \mathbf{u}) dS + \int\limits_{\tilde{V}} C dV = 0 \tag{7}$$

Интегральное уравнение (7) получается интегрированием уравнения (6) по объему и заменой интеграла по объему интегралом по границе в соответствии с теоремой о дивергенции Остроградского-Гаусса.

Дифференциальная запись закона сохранения, содержащая дивергенцию потока сохраняемой величины, называется дивергентной.

Неконсервативная форма записи получается дифференцированием в уравнении (6) сомножителей, определяющих поток,

$$(\nabla \cdot \mathbf{A}) \cdot \nabla \mathbf{u} + \mathbf{A} : (\nabla \otimes \nabla)\mathbf{u} + C = 0 \tag{8}$$

Неприятные сюрпризы при использовании неконсервативной формы записи (8) заключаются в следующем.

1) На дифференциальном уровне на гладких решениях консервативная и неконсервативная формы записи эквивалентны и переходят одна в другую с помощью тождественных преобразований. На дискретном уровне эта эквивалентность может нарушаться и в дискретных уравнениях, полученных из неконсервативной формы исходного уравнения, могут возникать нефизические источниковые члены (добавки к C), искажающие решение. Это искажение решения называется ошибкой по консервативности. На гладких решениях ошибки по консервативности с ростом размерности дискретной задачи стремятся к нулю. Но при наличии зон всплеска производных эти ошибки могут нарушить сходимость приближенных решений.

2) Неконсервативная запись исходного уравнения (8) для постоянного в пространстве тензора коэффициентов диффузии $\mathbf{A}$ принимает вид

$$\mathbf{A} : (\nabla \otimes \nabla)\mathbf{u} + C = 0 \tag{9}$$

Эта запись называется лапласовой формой эллиптического уравнения и нередко используется. Серьезная и довольно часто встречающаяся ошибка состоит в том, что при учете переменности коэффициентов диффузии $\mathbf{A}$ продолжают пользоваться формой записи (9). В этом случае закон сохранения (7) уже не может быть получен тождественными преобразованиями уравнения (9) на дифференциальном уровне и, тем более, нарушается на дискретном уровне. Ошибки по консервативности в этом случае не стремятся к нулю при увеличении размерности дискретной задачи (при измельчении шагов сетки), а аппроксимация эллиптического оператора отсутствует.

12.2 Метод конечных разностей

Метод конечных разностей состоит в следующем.

1) Область решения аппроксимируется сеткой узлов, для которых определено отношение соседства с помощью шаблонов. Шаблоном узла сетки называется набор его соседних узлов.

2) Дискретное решение представляется узловыми значениями искомых функций. Для заданного шаблона методом неопределенных коэффициентов (или методом разложения решения в ряд Тейлора) выводятся конечно-разностные формулы аппроксимации производных. Затем, производные в исходном дифференциальном уравнении и граничных условиях заменяются конечными разностями и тем самым реализуется переход к дискретным уравнениям. В более общем случае метод неопределенных коэффициентов используется не для приближения каждой производной в отдельности, а сразу для всего дифференциального оператора.

3) Полученная система алгебраических уравнений решается прямыми или итерационными методами.

4) Окончательно, с помощью интерполяции совершается переход от дискретного к непрерывному решению (операция восполнения).

Данная схема реализации метода конечных разностей может быть видоизменена так, чтобы стало ясно, что метод конечных разностей является вариантом проекционного метода. Действительно, базисными функциями аппроксимационного базиса для метода конечных разностей являются полиномы, принимающие значение единица в центральном узле шаблона и нуль во всех остальных узлах шаблона. Если число членов полинома равно числу узлов в шаблоне, то задача определения коэффициентов таких полиномов разрешается методом неопределенных коэффициентов. Число базисных функций равно числу узлов конечно-разностной сетки. В качестве базисных функций проекционного базиса принимаются дельта-функции $\delta(\mathbf{x}-\mathbf{x_i})$, определяемые со-

отношением

$$\int_V f\delta(\mathbf{x}-\mathbf{x_i})dV = f(\mathbf{x_i})$$

где i - индекс центрального узла шаблона, f - любая функция. В результате уравнения метода Галеркина-Петрова в точности соответствуют уравнениям метода конечных разностей.

В подавляющем большинстве случаев используются регулярные сетки, с раздельной нумерацией узлов по координатным направлениям: каждый узел имеет мультииндекс (i, j, k), в котором каждый индекс отвечает своей пространственной независимой переменной. Соответственно, в число соседей узла $\mathbf{u}_{ijk}$ вовлекаются все узлы вида $\mathbf{u}_{i\pm\alpha,j\pm\beta,k\pm\gamma}$, где параметры α, β, γ принимают значения $1, 2, ..., M$. Точность аппроксимации производных конечными разностями повышается за счет увеличения числа соседей.

Если область решения является прямой (одномерный случай), прямоугольником (двумерный случай) или параллелепипедом (трехмерный случай), то говорят, что область решения имеет каноническую форму. Такая область легко задается в декартовой прямоугольной системе координат ($\mathbf{x} = (x_1, x_2, x_3)$)

$$V = [(x_1, x_2, x_3) \ : \quad 0 \le x_1 \le 1 \ , \quad 0 \le x_2 \le 1 \ , \quad 0 \le x_3 \le 1] \quad (10)$$

и покрывается регулярной равномерной сеткой узлов $\mathbf{x}_{ijk}$:

$$x_{1i} = i/N_1 \ , \quad x_{2j} = j/N_2 \ , \quad x_{3k} = k/N_3$$

$$(i = 0, ..., N_1 \ ; \quad j = 0, ..., N_2 \ ; \quad k = 0, ..., N_3)$$

Приведем в качестве примера разностную аппроксимацию *уравнения Пуассона* на семиточечном шаблоне (рис. 12.1)

$$\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} + \frac{\partial^2 u}{\partial x_3^2} + C = 0 \qquad (11)$$

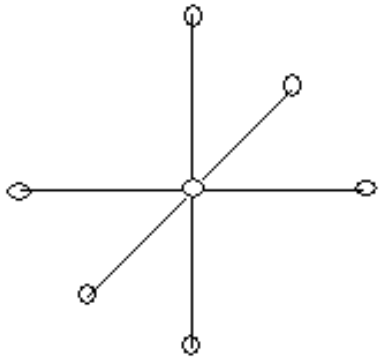

Рис. 12.1: 7-точечный шаблон

Подставляя в уравнение Пуассона разностные аппроксимации производных

$$\frac{\partial^2 u}{\partial x_1^2} \approx \frac{u_{i-1,j,k} - 2u_{i,j,k} + u_{i+1.j,k}}{h^2}$$

$$\frac{\partial^2 u}{\partial x_2^2} \approx \frac{u_{i,j-1,k} - 2u_{i,j,k} + u_{i.j+1,k}}{h^2}$$

$$\frac{\partial^2 u}{\partial x_3^2} \approx \frac{u_{i,j,k-1} - 2u_{i,j,k} + u_{i.j,k+1}}{h^2}$$

получаем следующую разностную схему

$$6u_{i,j,k} - u_{i+1,j,k} - u_{i-1.j,k} - u_{i,j+1,k} -$$

$$-u_{i,j-1,k} - u_{i,j,k+1} - u_{i,j,k-1} = h^2 C_{i,j.k} \qquad (12)$$

где интервал значений индексов i, j, k от 1 до $N_1 - 1$ отвечает внутренним точкам, сетка имеет равное количество узлов $N_1 + 1$ вдоль каждого из координатных направлений x_1 x_2 и x_3, поэтому $h_1 = h_2 = h_3 = h = 1/N_1$. Для замыкания этой системы уравнений надо добавить разностные выражения граничных условий. В случае условий Дирихле это будут просто заданные граничные узловые значения искомой функции, а в случае условий Неймана надо будет заменить первые производные по нормали к границе

односторонними конечно-разностными выражениями этих производных. Например, для границы $x_1 = 0$ условия Неймана можно записать так

$$\frac{u_{1,j,k} - u_{0,j,k}}{h}(-1) = (P_n)_{0,j,k} \tag{13}$$

где множитель (-1) учитывает знак проекции внешней нормали на ось x_1. Вместо двухузловой формулы первого порядка точности, можно использовать трехузловую формулу второго порядка точности

$$\frac{4u_{1,j,k} - u_{2,j,k} - 3u_{0,j,k}}{2h}(-1) = (P_n)_{0,j,k} \tag{14}$$

где граничные значения P_n снабжены тремя индексами для единообразия записи.

Выписанная разностная схема аппроксимирует уравнение Пуассона со вторым порядком точности во внутренних точках и имеет первый или второй порядок точности при аппроксимации условий Неймана. Условия Дирихле учитываются точно.

Схемы повышенного порядка точности содержат большее количество соседних узлов и у границ используют специальные несимметричные разностные формулы повышенного порядка точности. Не приводя громоздких формул, покажем, например, 19-точечный шаблон внутреннего узла для трехмерной задачи (Рис. 12.2).

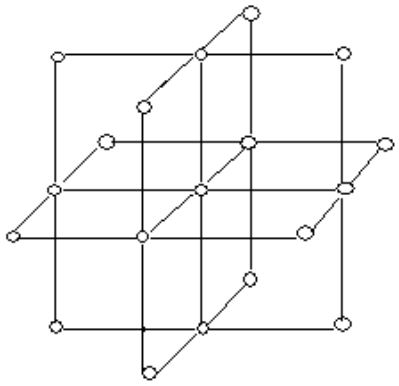

Рис. 12.2: 19-точечный шаблон

Метод конечных разностей нередко реализуется на криволинейных регулярных (структурированных) сетках в областях сложной формы. Такие сетки получаются с помощью отображений области решения на каноническую область $x_i = x_i(\xi_1, \xi_2, \xi_3), \quad (i = 1, 2, 3)$. В исходных уравнениях производят замену независимых пространственных переменных, после чего вычисление производных по переменным ξ_i $(i = 1, 2, 3)$ проводится с использованием обычных шаблонов прямоугольной сетки в канонической области. Методом неопределенных коэффициентов несложно вывести конечно-разностные аппроксимации производных и на неструктурированных сетках и переменных от узла к узлу шаблонах. Системы уравнений метода неопределенных коэффициентов при этом придется решать численно (аналитически не получится). Этого, однако, почти никто не делает, поскольку для неструктурированных сеток есть методы попривлекательнее.

12.3 Метод контрольных объемов

Обеспечение консервативности, то есть выполнение интегрального балансного соотношения (7) на дискретном уровне в конечно-разностных схемах автоматически не реализуется. В случае гладких решений нарушения консервативности не создают проблем, поскольку ошибки по консервативности, как и ошибки аппроксимации в целом, при измельчении сеток стремятся к нулю. Однако при наличии зон всплеска значений производных (при сквозном счете разрывных решений) ошибки по консервативности способны испортить численное решение даже на самых мелких сетках. Поэтому часто дискретизация проводится непосредственно на основе интегральной формы исходного уравнения (7). Для этого каждый узел сетки $\mathbf{x}_i$ окружается своим контрольным объемом V_i. Балансное соотношение (7) записывается для

каждого контрольного объема сетки

$$\int\limits_{\tilde{S}_i} \mathbf{n} \cdot (\mathbf{A} \cdot \nabla \mathbf{u}) dS + \int\limits_{\tilde{V}_i} C dV = 0 \tag{15}$$

и затем интегрируется численно с использованием гауссовых квадратурных формул и конечно-разностной аппроксимации подынтегральных выражений. Для пограничных контрольных объемов при этом принимаются во внимание граничные условия. Из полученной системы алгебраических уравнений находятся узловые значения искомых функций, по которым интерполяцией определяется искомое решение.

Метод контрольных объемов называется также методом конечных объемов, интегро-интерполяционным методом, методом баланса. Он был предложен в середине двадцатого века практически одновременно в работах ряда ученых у нас и за рубежом.

Метод контрольных объемов является вариантом проекционного метода Галеркина-Петрова. В качестве аппроксимационного базиса используются полиномы метода конечных разностей. А в качестве проекционного базиса применяются кусочно-постоянная базисные функции, каждая из которых равна единице в окружающем узел контрольном объеме и нулю вне этого объема. В результате проектирования исходное дивергентное дифференциальное уравнение порождает набор интегральных уравнений, выражающих закон сохранения величины $\mathbf{u}$ для каждой ячейки. Число таких уравнений равно числу узлов сетки, то есть равно числу искомых узловых значений решения.

Под названием метод контрольных объемов скрывается множество методов, различающихся способами построения контрольных объемов, способами численного интегрирования по границам контрольных объемов, способами аппроксимации подынтегральных выражений, способами решения дискретных уравнений, методами восполнения приближенного решения.

В отличие от метода конечных разностей метод контрольных объемов автоматически консервативен, так как по построению поддерживает законы сохранения на дискретном уровне. К достоинствам метода контрольных объемов следует отнести также то, что он пригоден к использованию на неструктурированных сетках.

12.4 Метод конечных элементов

В методе конечных элементов используется галеркинская вариационная форма исходного уравнения, которая получается скалярным умножением исходного уравнения на вариацию решения $\delta\mathbf{u}$, полученную разностью двух произвольных функций из линейного пространства искомых решений, с последующим интегрированием по области определения решения и применением теоремы о дивергенции для понижения порядка производных.

Запись исходного вариационного уравнения имеет вид:

$$\int\limits_V (\mathbf{A}\cdot\nabla u)\cdot\nabla\delta u dV = \int\limits_{S_p} P_n\delta u dS + \int\limits_V C\delta u dV \qquad (16)$$

где $S_p = S\backslash S_u$. Условия Неймана уже учтены в интеграле по границе области решения, а условия Дирихле, называемые по традициям вариационного исчисления главными, дополняют постановку краевой задачи:

$$\mathbf{x}\in S_u \;:\quad u = u^*(\mathbf{x}) \qquad (17)$$

Пространственная область решения V аппроксимируется набором непересекающихся ячеек, называемых конечными элементами и построенных так, чтобы ребра (линии, соединяющие соседние узлы) принадлежали границам ячеек. Заметим, что *ячейки конечных элементов в отличие от ячеек контрольных объемов не окружают узлы сетки. То есть разбиение пространственной области на ячейки в методах контрольных объемов и*

в методах конечных элементов принципиально различны.

Непонимание этого факта привело к распространению в научной литературе (в основном в зарубежной) мифа о неконсервативности метода конечных элементов, поскольку интегральные балансные соотношения (7) на ячейках конечных элементов не выполняются, а потоки $\mathbf{A} \cdot \nabla u$ на границах конечных элементов могут претерпевать разрыв. На самом деле, доказательство консервативности метода конечных элементов, как и многих других методов (в частности, бессеточных) проводится без привлечения представлений о контрольных объемах.

Для определения сетки конечных элементов: 1) задаются координаты узлов $\mathbf{x}_i$ $(i = 1, ..., N_1)$ 2) задаются конечные элементы, для этого используется список Ω_{jk}, который для каждого конечного элемента j $(j = 1, ..., N_2)$ задает глобальные номера образующих его узлов k $(k = 1, ..., M_j^{(g)})$; 3) задаются граничные элементы, для этого используется список Γ_{jk}, который для каждого граничного элемента j $(j = 1, ..., N_3)$ задает глобальные номера образующих его узлов k $(k = 1, ..., M_j^{(g)})$.

Для дальнейшей дискретизации надо представить приближенное решение в виде разложения по функциям аппроксимационного базиса, а его вариации в виде разложения по функциям проекционного базиса. После выполнения интегрирования и приведения подобных членов при дискретных вариациях дискретные уравнения метода конечных элементов получается с помощью основной леммы вариационного исчисления. В силу произвольности вариаций решения их множители приравниваются нулю, что и дает искомую систему уравнений.

Рассмотрим простейшие конечные элементы, использующие узловые значения искомых функций в качестве искомого дискретного решения. В этом случае приближенное решение ищется

в виде разложения по базисным функциям $\phi_i(\mathbf{x})$

$$u^{(N_1)} = \sum_{i=1}^{N_1} u_i \phi_i(\mathbf{x}) \tag{18}$$

где u_i - узловые значения искомой функции. Каждая базисная функция $\phi_i(\mathbf{x})$ ассоциируется со своим узлом i, в котором она равна единице, а в остальных узлах она равна нулю. Вариации искомых функций чаще всего ищутся в виде разложения по тому же самому базису $\phi_i(\mathbf{x})$

$$\delta u^{(N_1)} = \sum_{i=1}^{N_1} \delta u_i \phi_i(\mathbf{x}) \tag{19}$$

После подстановки этих разложений вариационное уравнение приводится к дискретной форме

$$\sum_{i=1}^{N_1} (\sum_{j=1}^{N_1} B_{ij} u_j - b_i) \delta u_i = 0 \tag{20}$$

где

$$B_{ij} = \int\limits_V (\mathbf{A} \cdot \nabla \phi_i) \cdot \nabla \phi_j dV \tag{21}$$

$$b_i = \int\limits_{S_p} P_n \phi_i dS + \int\limits_V C \phi_i dV \tag{22}$$

из которой в силу произвольности вариаций дискретного решения δu_i получаем систему алгебраических уравнений метода конечных элементов

$$\sum_{j=1}^{N_1} B_{ij} u_j = b_i \ , \quad (i = 1, ..., N_1) \tag{23}$$

Интегралы в выражениях для компонент матрицы жесткости B_{ij} и вектора правой части b_i обычно берутся численно.

Детальное описание численного алгоритма решения краевой задачи для уравнения Пуассона с помощью МКЭ дано ниже в подразделе про применение безматричных итерационных процессов решения.

12.5 Метод граничных элементов

Исходная краевая задача. Рассмотрим метод граничных интегральных уравнений на примере смешанной краевой задачи для уравнения Лапласа

$$\mathbf{x} \in \Omega_\varphi \ : \quad \Delta\varphi = 0 \tag{24}$$

с граничными условиями

$$\mathbf{x} \in \partial\Omega_\varphi \ : \quad \varphi = f_* \tag{25}$$

и

$$\partial\varphi/\partial n = g_* \qquad \partial\Omega_g = \Omega \backslash \Omega_\varphi \tag{26}$$

Уравнение Лапласа имеет сингулярное решение

$$\varphi = 1/(4\pi r)$$

Здесь $r = |\mathbf{x} - \mathbf{x}'|$ является расстоянием между произвольными точками $\mathbf{x}$ и $\mathbf{x}'$ в области решения. Сингулярное решение удовлетворяет исходному уравнению Лапласа при наличии точечного источника единичной интенсивности, расположенного в точке $\mathbf{x}'$:

$$\Delta\varphi = \delta(r)$$

где $\delta(r)$ - дельта-функция Дирака, определяемая соотношением

$$\int\limits_\Omega \Phi(\mathbf{x})\delta(|\mathbf{x} - \mathbf{x}'|)d\Omega = \Phi(\mathbf{x}')$$

Это проверяется подстановкой сингулярного решения в уравнение Лапласа, записанное в сферической системе координат

$$\frac{\partial^2\varphi}{\partial r^2}+\frac{2}{r}\frac{\partial\varphi}{\partial r}-\delta(r)=0$$

При $r=0$ сингулярное решение имеет особенность и трактуется в смысле обобщенного решения в соответствии с определением дельта-функции.

Граничное интегральное уравнение. Сингулярное решение используется для приведения исходной краевой задачи для уравнения Лапласа к соответствующему граничному интегральному уравнению. Для этого уравнение Лапласа умножается на произвольную функцию w и дважды интегрируется по частям, результат имеет вид

$$0=\int\limits_{\Omega}\nabla^2\varphi w d\Omega=\int\limits_{\Omega}\nabla\cdot(\nabla\varphi w)d\Omega-\int\limits_{\Omega}\nabla\varphi\cdot\nabla w d\Omega=$$

$$=\int\limits_{\partial\Omega}\mathbf{n}\cdot\nabla\varphi w ds-\int\limits_{\Omega}\nabla\cdot(\varphi\nabla w)d\Omega+\int\limits_{\Omega}\varphi\nabla^2 w d\Omega$$

или

$$\int\limits_{\partial\Omega}\mathbf{n}\cdot\nabla\varphi w ds-\int\limits_{\partial\Omega}\varphi\mathbf{n}\cdot\nabla w d\Omega+\int\limits_{\Omega}\varphi\nabla^2 w d\Omega=0$$

После подстановки сюда вместо произвольной функции w сингулярного решения получается так называемое исходное интегральное тождество метода граничных интегральных уравнений

$$\varphi(\mathbf{x})=\frac{1}{4\pi}\int\limits_{\partial\Omega}\left\{g(\mathbf{x}')\frac{1}{|\mathbf{x}-\mathbf{x}'|}-f(\mathbf{x}')\frac{\partial}{\partial n}\left[\frac{1}{|\mathbf{x}-\mathbf{x}'|}\right]\right\}ds' \qquad (27)$$

где $\mathbf{x}\in\Omega$, $\mathbf{x}'\in\partial\Omega$, $g(\mathbf{x}')=\mathbf{n}\cdot\nabla\varphi\big|_{\mathbf{x}'}$, $f(\mathbf{x}')=\varphi\big|_{\mathbf{x}'}$. Устремлением внутренней точки $\mathbf{x}$ к точке границы $\mathbf{x}''$ $(\mathbf{x}\in\Omega\to\mathbf{x}''\in\partial\Omega)$

это тождество сводится к интегральному граничному уравнению

$$f(\mathbf{x}'') = \frac{1}{2\pi} \int\limits_{\partial\Omega} \left\{ g(\mathbf{x}') \frac{1}{|\mathbf{x}'' - \mathbf{x}'|} - f(\mathbf{x}') \frac{\partial}{\partial n'} \left[\frac{1}{|\mathbf{x}'' - \mathbf{x}'|} \right] \right\} ds' \quad (28)$$

где $f(\mathbf{x}'') = \lim \varphi(\mathbf{x})$ при $\mathbf{x} \in \Omega \to \mathbf{x}'' \in \partial\Omega$. Это граничное интегральное уравнение устанавливает связь между граничными значениями искомой функции f и ее потока g. В каждой точке границы задана одна из функций f и g, а другая подлежит определению. После решения этого уравнения искомое решение φ в области Ω определяется с помошью исходного интегрального тождества.

Численная реализация. Граница представляется набором N граничных элементов (ячеек). Значения искомой функции f и ее потока g на границе ищутся в классе кусочно-постоянных функций, принимающих постоянные значения на каждом из граничных элементов. Интегральное уравнение записывается для каждого из граничных элементов. Для кусочно постоянной аппроксимации значения искомых функций f_i и g_i относятся обычно к центру элемента i. Для каждого граничного элемента i дискретизированное интегральное уравнение принимает вид

$$\sum_{j=1}^{N} A_{ij} f_j = \sum_{j=1}^{N} B_{ij} g_j$$

где матрицы A_{ij} и B_{ij} размера $N \mathrm{x} N$ определяются по формулам:

$$A_{ij} = \delta_{ij} + \frac{1}{2\pi} \int\limits_{\partial\Omega_j} \frac{\partial}{\partial n'} \left[\frac{1}{|\mathbf{x}_\mathbf{i}'' - \mathbf{x}'|} \right] ds'$$

$$B_{ij} = \frac{1}{2\pi} \int\limits_{\partial\Omega_j} \frac{1}{|\mathbf{x}_\mathbf{i}'' - \mathbf{x}'|} ds'$$

здесь $\delta_{ij} -$, f_j и g_j - значения искомых функций в граничном элементе $\partial\Omega_j$, $i, j = 1, ..., N$, ds' - бесконечно малая часть

граничного элемента $\partial\Omega_j$, содержащая точку $\mathbf{x}'$, точка $\mathbf{x}''_\mathbf{i}$ расположена в центре граничного элемента i. На каждом элементе искомым является значение одной из функций f и g, в то время как значение другой функции задано граничными условиями. Таким образом число неизвестных совпадает с числом уравнений.

После решения системы уравнений искомая функция φ_i в любой внутренней точке $\mathbf{x}$ определяется подстановкой полученных граничных значений f_i и g_i в исходное интегральное тождество (27). Производные от решения определяются непосредственным дифференцированием исходного интегрального тождества (27) по $\mathbf{x}$ (дифференцируются подынтегральные выражения, содержашие $|\mathbf{x} - \mathbf{x}'|$).

Описанный способ решения носит название метода граничных интегральных уравнений (метода ГИУ) или метода граничных элементов (МГЭ).

Плюсы и минусы МГЭ. Для большинства задач МГЭ хорошо работает при малом числе граничных элементов N даже для самых простых кусочно-постоянных аппроксимаций функций f и g на граничных элементах. Повышение точности достигается или за счет увеличения числа граничных элементов N, или за счет применения на элементах аппроксимаций более высокого порядка точности (кусочно-линейных, кусочно-квадратичных и так далее). Интегралы в выражениях для коэффициентов матриц системы алгебраических уравнений определяются численно с использованием квадратурных формул. Матрицы МГЭ являются полностью заполненными с диагональным преобладанием.

МГЭ позволяет понизить на единицу пространственную размерность задачи, так как разрешающая система уравнений формулируется относительно граничных значений неизвестных. Это избавляет от необходимости строить сетки внутри областей решения, но, к сожалению, в ограниченном числе случаев. Ненулевые свободные члены в правой части исходного дифференциального уравнения приводят к появлению в исходном интегральном тож-

дестве членов с интегралами по объему. Для вычисления этих интегралов приходится вводить сетку не только на границе, но и в области. Объемные интегралы неизбежно возникают в МГЭ при его применении к нелинейным задачам и к линейным задачам, для которых не удается получить исходное интегральное тождество.

В случае задач для уравнений с переменными коэффициентами и, тем более, в случае нелинейных задач исходные интегральные тождества МГЭ построить невозможно. Поэтому в этих случаях для реализации МГЭ применяются внешние итерации по нелинейности, которые строятся методом простой итерации путем выделения линейного оператора, для которого исходные интегральные тождества можно вывести. Оставшиеся после выделения линейного оператора нелинейные члены относятся в правую часть как дополнение к свободному члену. Записывая исходную нелинейную краевую задачу в операторной форме

$$D(u) = d$$

процесс решения методом простой итерации можно представить так

$$Lu^{(n+1)} = d + Lu^{(n)} - D(u^{(n)}) = \tilde{d}(u^{(n)})$$

где L - оператор линейной краевой задачи, для которого имеется исходное интегральное тождество, D - нелинейный оператор краевой задачи, которую надо решить, d - заданная правая часть исходной краевой задачи. На каждой итерации значения правых частей $\tilde{d}$ линеаризованной краевой задачи определяются по решению на предыдущей итерации $u^{(n)}$. С ростом влияния нелинейности сходимость таких итераций замедляется и вообще может отсутствовать. Это ограничивает применение МГЭ к нелинейным задачам и задачам с переменными коэффициентами. Поэтому МГЭ не применяется в задачах удара и волновой динамики, в которых эволюция (развитие, изменение во времени) решения

происходит часто вдали от границ и не зависит напрямую от граничных условий. Тем не менее, у МГЭ есть достаточно обширная область применения, в которой он находится вне конкуренции (линейные квазистатические задачи без объемных источников).

12.6 Бессеточные методы

Бессеточные методы решения краевых задач для уравнений с частными производными используют разложение решения по функциям аппроксимационного базиса, которые строятся без введения сеток. Примером бессеточных методов являются проекционные методы, использующие глобальные базисные функции. Из-за трудностей выполнения граничных условий глобальные аппроксимационные базисы, составленные из функций, отличных от нуля во всей области решения, удается применить только к областям решения простой формы (прямая, квадрат, куб) (см. Михлин [1950, 1970]).

Метод R-функций. Рвачевым (1978) предложен метод модификации глобальных базисов, называемый методом R-функций и позволяющий выполнить граничные условия в областях решения произвольной формы. Ключевым приемом в методе Рвачева является построение функции области решения, которая положительна внутри нее и отрицательна снаружи, а на границе обращается в нуль. С помощью такой функции любой глобальный базис модифицируется так, чтобы базисные функции принимали на границе заданные значения. Обзор приложений этого метода к решению конкретных задач сделан в работе Рвачева (1995).

Метод фиктивных областей. В методах фиктивных областей граничные условия трактуются как ограничения и, пользуясь методами математического программирования, включаются в вариационное уравнение рассматриваемой краевой задачи. Во многих современных бессеточных методах это делается с помощью

вариантов метода штрафных функций. При этом никаких требований к граничным значениям функций аппроксимационного базиса предъявлять не требуется.

Методы локальных базисных функций. Начиная с последней четверти 20-го века бурно развиваются бессеточные методы, использующие локальные аппроксимационные базисы, для построения которых в области решения распределяются так называемые "свободные узлы, точки или частицы"[1]. С каждым свободным узлом $\mathbf{x_i}$ связывается своя локальная базисная функция $\phi_i(\mathbf{x})$, равная 1 в этом узле и обращающаяся в нуль вне заданной окрестности Ω_i этого узла. Граничные условия учитываются в вариационном уравнении Галеркина, так что на граничные значения базисных функций ограничений не накладывается. Расположение свободных узлов и размер окрестностей задаются так, чтобы обеспечить взаимное пересечение окрестностей соседних узлов. Пересечение окрестностей свободных узлов необходимо для аппроксимации решения, иначе дискретная модель распадается на множество невзаимодействующих свободных узлов и аппроксимация решения исходной краевой задачи отсутствует.

Далее в соответствии с методом Галеркина вводится проекционный базис $\psi_i(\mathbf{x})$ и формируется система алгебраических уравнений относительно коэффициентов разложения решения (см. главу про проекционные методы). Основная трудность при формировании систем дискретных уравнений метода Галеркина состоит в выполнении операции численного интегрирования, для которой разработан ряд специальных приемов бессеточного интегрирования, наиболее распространенными из которых являются метод коллокации (если выбирается проекционный базис из дельта-функций) и метод наложения вспомогательных равномерных сеток (которые запоминать не нужно).

[1]Первая реализация метода свободных точек принадлежит Дьяченко(1973)

12.7 Итерации по нелинейности

В линейных задачах систему дискретных уравнений можно сформировать и ввести в память ЭВМ путем вычисления и запоминания коэффициентов матрицы и вектора правой части СЛАУ. *В нелинейных задачах систему алгебраических уравнений сформировать и ввести в память ЭВМ в общем случае невозможно*, так как сплошь и рядом она оказывается определенной алгоритмически и не имеет аналитического представления. Поэтому в общем случае нелинейные уравнения представляются алгоритмами вычисления их невязок.

Решения нелинейных задач

$$F(u) = 0$$

как правило находятся итерациями как предел последовательности решений вспомогательных линейных задач, которые представляют внешний итерационный процесс и называются итерациями по нелинейности. Итерации по нелинейности реализуются методами квазилинеаризации,

$$F(u^n) + F'_u(u^n)(u^{n+1} - u^n) = 0$$

методами дифференцирования по параметру

$$F'_\lambda + F'_u u'_\lambda = 0 \ , \quad u|_{\lambda=0} = u^{(0)}$$

или методами установления (нефизическими простыми итерациями)

$$Lu'_t = F(u) \ , \quad u|_{t=0} = u^{(0)}$$

Приведем типичные примеры итераций по нелинейности для рассматриваемых нелинейных эллиптических задач (1).

В случае метода квазилинеаризации последовательность линеаризованных задач, отвечающих задаче (1), может иметь вид

$$\begin{array}{l}\mathbf{x}\in V \;:\quad \nabla\cdot(\mathbf{A}(u^n)\cdot\nabla u^n)+C(u^n)+\\ \quad+\nabla\cdot\big((\mathbf{A}'_u(u^{n+1}-u^n))\cdot\nabla u^n+\mathbf{A}(u^n)\cdot\nabla(u^{n+1}-u^n)\big)+\\ \qquad+C'_u(u^{n+1}-u^n)=0\\ \mathbf{x}\in S_u \;:\quad u^{n+1}=u_*(\mathbf{x})\\ \mathbf{x}\in S\backslash S_u \;:\quad \mathbf{n}\cdot(\mathbf{A}(u^n)\cdot\nabla u^n+\\ \quad+(\mathbf{A}'_u(u^{n+1}-u^n))\cdot\nabla u^n+\mathbf{A}(u^n)\cdot\nabla(u^{n+1}-u^n)\big)=P_n^*(\mathbf{x})\end{array}$$

В случае метода дифференцирования по параметру роль параметра λ в задачах о деформациях тел исполняет параметр нагрузки или параметр граничного перемещения. В задачах гидродинамики параметром интенсивности процесса может быть характерная величина заданной граничной скорости, в задачах теплопередачи - граничная температура, интенсивность источника тепла и так далее. Дифференцируя исходные уравнения и граничные условия по выбранному параметру получаем

$$\begin{array}{l}\mathbf{x}\in V \;:\quad \nabla\cdot\big((\mathbf{A}'_u\frac{du}{d\lambda}\cdot\nabla u+\mathbf{A}(u)\cdot\nabla\frac{du}{d\lambda}\big)++C'_u\frac{du}{d\lambda}=0\\ \mathbf{x}\in S_u \;:\quad \frac{du}{d\lambda}=\frac{d}{d\lambda}u_*(\mathbf{x})\\ \mathbf{x}\in S\backslash S_u \;:\quad \mathbf{n}\cdot\big(\mathbf{A}'_u\frac{du}{d\lambda}\cdot\nabla u+\mathbf{A}(u)\cdot\nabla\frac{du}{d\lambda}\big)=\frac{d}{d\lambda}P_n^*(\mathbf{x})\end{array}$$

Если подходящего параметра λ в задаче нет, то его можно ввести искусственно, например, так. Исходная задача переписывается в виде

$$\begin{array}{l}\mathbf{x}\in V \;:\quad \nabla\cdot(\mathbf{A}(u)\cdot\nabla u)+\lambda C(u)=0\\ \mathbf{x}\in S_u \;:\quad u=\lambda u_*(\mathbf{x})\\ \mathbf{x}\in S\backslash S_u \;:\quad \mathbf{n}\cdot(\mathbf{A}(u)\cdot\nabla u)=\lambda P_n^*(\mathbf{x})\end{array}$$

тогда при $\lambda=0$ она имеет тривиальное решение $u=0$, а при $\lambda=1$ решение исходной задачи. Вспомогательная линейная задача Коши метода дифференцирования по параметру имеет вид

$$\begin{array}{l}\mathbf{x}\in V \;:\quad \nabla\cdot\big((\mathbf{A}'_u\frac{du}{d\lambda}\cdot\nabla u+\mathbf{A}(u)\cdot\nabla\frac{du}{d\lambda}\big)+C(u)=0\\ \mathbf{x}\in S_u \;:\quad \frac{du}{d\lambda}=u_*(\mathbf{x})\\ \mathbf{x}\in S\backslash S_u \;:\quad \mathbf{n}\cdot\big(\mathbf{A}'_u\frac{du}{d\lambda}\cdot\nabla u+\mathbf{A}(u)\cdot\nabla\frac{du}{d\lambda}\big)=P_n^*(\mathbf{x})\end{array}$$

решая для заданых u эти линейные задачи относительно $du/d\lambda$ получаем возможность реализовать процесс численного интегрирования задачи Коши $\{du/d\lambda = u'_\lambda(u),\ u = 0(\lambda = 0)\}$ на интервале $\lambda \in [0,1]$ по какой-либо из схем Рунге-Кутта или Адамса.

В случае метода установления последовательность линеаризованных задач выглядит так

$$\begin{array}{ll} \mathbf{x} \in V & : \quad u^{n+1} = u^n + \Delta t_n \left(\nabla \cdot (\mathbf{A}(u^n) \cdot \nabla u^n) + C(u^n)\right) \\ \mathbf{x} \in S_u & : \quad u^{n+1} = u_*(\mathbf{x}) \\ \mathbf{x} \in S \backslash S_u & : \quad \mathbf{n} \cdot (\mathbf{A}(u^n) \cdot \nabla u^n) = P_n^*(\mathbf{x}) \end{array}$$

Во многих руководствах по численным методам, включая и данное, решение нелинейных задач поясняется так: "... исходная нелинейная краевая задача методом ... сводится к системе нелинейных алгебраических уравнений, которая затем линеаризуется...". Читая такие в принципе верные описания, надо понимать, что в общем случае нет способа введения систем нелинейных уравнений в память ЭВМ. В реальности нелинейные уравнения задаются детальным описанием алгоритма подсчета невязки нелинейных уравнений для любого пробного решения. Поэтому необходимая в итерациях по нелинейности линеаризация нелинейных краевых задач проводится, как правило, еще на исходных интегро-дифференциальных формулировках. Тем не менее, упомянутый вариант объяснения численных методов для нелинейных задач использовался, используется и будет использоваться, так как он удобен для общих рассуждений о методах решения и в принципе правилен, хотя и не вполне точно отражает действительность.

12.8 Безматричные двухшаговые итерации

Рассмотрим теперь решение вспомогательных линейных эллиптических краевых задач безматричными двухшаговыми итерационными методами. Почему нужно ориентироваться на без-

матричные итерационные двухшаговые методы? Ответ таков. Потому, что они 1) очень простые; 2) не требуют много памяти; 3) более эффективны, чем матричные методы или одношаговые итерационные методы; 4) требуют для решения задачи конечного числа операций; 5) вне конкуренции при решении задач большой размерности; 6) они вне конкуренции при векторизации/распараллеливании. А как же недостатки? - А недостатков нет.

Рассмотрим применение безматричного двухшагового итерационного метода сопряженных градиентов к решению смешанной краевой задачи для уравнения Пуассона в трехмерной области произвольной геометрии.

Аппроксимируем область решения сеткой тетраэдральных конечных элементов, границу области опишем сеткой граничных треугольных элементов. Пусть $\{\mathbf{x_i}\}_{i=1}^{N_v}$ - радиус-векторы узлов, N_v - число узлов; $\{J(k,l)\}_{k=1,l=1}^{N_e,e}$ - массив номеров узлов в тетраэдральных элементах, N_e - число тетраэдральных элементов, $M_e = 4$ - число узлов в тетраэдральном элементе; $\{G(k,l)\}_{k=1,l=1}^{N_g,M_g}$ - массив номеров узлов в треугольных граничных элементах, N_g - число граничных элементов, $M_g = 3$ - число узлов в граничном элементе.

Для каждого тетраэдрального конечного элемента k ($k = 1, ..., N_e$) вычислим объем

$$V_k = \frac{1}{6} det \begin{vmatrix} x_{(k)21} & y_{(k)21} & z_{(k)21} \\ x_{(k)31} & y_{(k)31} & z_{(k)31} \\ x_{(k)41} & y_{(k)41} & z_{(k)41} \end{vmatrix}$$

где

$$x_{(k)ij} = x_{J(k,i)} - x_{J(k,j)}\ , \quad y_{(k)ij} = y_{J(k,i)} - y_{J(k,j)}\ , \quad z_{(k)ij} = z_{J(k,i)} - z_{J(k,j)}$$

и определим коэффициенты в формулах для интерполяции иско-

мой функции

$$[\phi(\mathbf{x_m}]_k = \sum_{l=1}^{M_e} d_{kl}^{(0)} \phi_{J(k,l)}$$

где

$$d_{k1}^{(0)} = d_{k2}^{(0)} = d_{k3}^{(0)} = d_{k4}^{(0)} = 1/4$$

а также в формулах для вычисления производных от искомой функции:

$$[\partial_x \phi(\mathbf{x_m}]_k = \sum_{l=1}^{M_e} d_{kl}^{(x)} \phi_{J(k,l)}$$

$$[\partial_y \phi(\mathbf{x_m}]_k = \sum_{l=1}^{M_e} d_{kl}^{(y)} \phi_{J(k,l)}$$

$$[\partial_z \phi(\mathbf{x_m}]_k = \sum_{l=1}^{M_e} d_{kl}^{(z)} \phi_{J(k,l)}$$

где

$$d_{k2}^{(x)} = \frac{y_{(k)31} z_{(k)41} - z_{(k)31} y_{(k)41}}{6V_k}$$

$$d_{k3}^{(x)} = \frac{y_{(k)41} z_{(k)21} - z_{(k)41} y_{(k)21}}{6V_k}$$

$$d_{k4}^{(x)} = \frac{y_{(k)21} z_{(k)31} - z_{(k)21} y_{(k)31}}{6V_k}$$

$$d_{k1}^{(x)} = -(d_{k2}^{(x)} + d_{k3}^{(x)} + d_{k4}^{(x)})$$

$$d_{k2}^{(y)} = \frac{z_{(k)31} x_{(k)41} - x_{(k)31} z_{(k)41}}{6V_k}$$

$$d_{k3}^{(y)} = \frac{z_{(k)41} x_{(k)21} - x_{(k)41} z_{(k)21}}{6V_k}$$

$$d_{k4}^{(y)} = \frac{z_{(k)21} x_{(k)31} - x_{(k)21} z_{(k)31}}{6V_k}$$

$$d_{k1}^{(y)} = -(d_{k2}^{(y)} + d_{k3}^{(y)} + d_{k4}^{(y)})$$

$$d_{k2}^{(z)} = \frac{x_{(k)31}y_{(k)41} - y_{(k)31}x_{(k)41}}{6V_k}$$

$$d_{k3}^{(z)} = \frac{x_{(k)41}y_{(k)21} - y_{(k)41}x_{(k)21}}{6V_k}$$

$$d_{k4}^{(z)} = \frac{x_{(k)21}y_{(k)31} - y_{(k)21}x_{(k)31}}{6V_k}$$

$$d_{k1}^{(z)} = -(d_{k2}^{(z)} + d_{k3}^{(z)} + d_{k4}^{(z)})$$

Определим дискретный аналог оператора пространственного дифференцирования

$$[\nabla\phi]_k = \sum_{l=1}^{M_e} \nabla_{kl}\phi_{J(k.l)}$$

который в базисе декартовой системы координат имеет следующее представление

$$\nabla_{kl} = \mathbf{e_x}d_{kl}^{(x)} + \mathbf{e_y}d_{kl}^{(y)} + \mathbf{e_z}d_{kl}^{(z)}$$

Интегралы в вариационном уравнении Галеркина (16) представим суммой интегралов по внутренним (тетраэдральным) и граничным (треугольным) конечным элементам. Тогда, применяя простейшую квадратурную формулу прямоугольников, получим дискретный аналог этого уравнения

$$\sum_{k=1}^{N_e} \left(([A]_k \cdot [\nabla\phi]_k) \cdot \sum_{l=1}^{M_e} \nabla_{kl}\delta\phi_{J(k,l)} \right) V_k =$$

$$= \sum_{k=1}^{N_g} \left([P_n]_k \sum_{l=1}^{M_g} \delta\phi_{G(k,l)}/3 \right) S_k$$

В рассматриваемом здесь в качестве примера алгоритме вариации решения и решение аппроксимированы единообразно, так

что имеем дело с вариантом проекционного метода Бубнова-Галеркина.

Полученное дискретное вариационное уравнение смешанной краевой задачи для уравнения Пуассона пока записано в виде, в котором подобные члены при дискретных вариациях искомой функции еще не приведены.

Сделаем замечание о смысле вариаций. Вариации решения $\delta\phi$ (непрерывные) и $\delta\phi_i$ (дискретные) никаких конкретных значений не принимают и в ходе решения задачи не определяются. Вариации служат общими множителями. Суммы подобных членов при этих общих множителях в силу основной леммы вариационного исчисления представляют уравнения задачи, которые должны на искомом решении обращаться в нуль.

Вариационное исчисление (см. например, Цлаф(1970)) определяет вариацию как разность двух произвольных пробных решений. Должно быть ясно, что встречающееся иногда выражение "бесконечно малая вариация" лишено смысла. Операцию варьирования нелинейных операторов и функционалов проводят в пространстве решений, применяя технику дифференцирования в функциональных пространствах в соответствии с определением дифференциала Гато:

$$\delta F(\mathbf{u}) = F'_u(\mathbf{u})\delta\mathbf{u} = \lim_{\alpha\to 0}\frac{\partial}{\partial\alpha}F(\mathbf{u}+\alpha\delta\mathbf{u})$$

Отсюда видно, что нельзя определять вариацию как дифференциал решения в пространстве независимых переменных задачи, что иногда встречается в литературе.

Рассмотрим простой пример. Вариационное уравнение

$$(x+y-2)\delta x + (x-y)\delta y = 0$$

должно выполняться для любых значений вариаций δx и δy, поэтому оно эквивалентно системе уравнений

$$x+y-2=0\ , \quad x-y=0$$

В руководствах по МКЭ рекомендуется сначала сформировать матрицу жесткости путем приведения подобных членов при множителях $\phi_j \delta\phi_i$. Процесс приведение этих подобных членов реализуется последовательныи суммированием вкладов в матрицу жесткости от каждого из конечных элементов и называется "конденсацией"(см. например, Зенкевич, 1975).

В итерационных безматричных методах операция "конденсации" в том виде, как это описано выше, не требуется. Для проведения итераций надо уметь подсчитывать невязки уравнений. Для этого совершенно нет нужды определять, да еще и где-то хранить матрицу системы уравнений. Чтобы подсчитать невязку не надо определять N_v^2 компонент матрицы, приводя подобные члены при $\phi_j \delta\phi_i$, а вполне достаточно подсчитать N_v невязок, приводя подобные члены при $\delta\phi_i$, что является значительно более простой операцией. В результате подсчета невязок дискретное вариационное уравнение приводится к виду

$$\sum_{i=1}^{N_v} [g(\phi)]_i \delta\phi_i = 0$$

где невязки $[g]_i$ дискретных уравнений определяются выражением

$$[g(\phi)]_i = \sum_{k=1}^{N_e} \sum_{l=1}^{M_e} [g_c(\phi)]_{kl} \tilde{H}(i - J(k,l)) - \sum_{k=1}^{N_g} \sum_{l=1}^{M_g} [g_b(\phi)]_{kl} \tilde{H}(i - G(k,l))$$

в котором функция $\tilde{H}$ равна единице, если аргумент равен нулю, и равна нулю в противном случае.

Величины $[g_c(\phi)]_{kl}$ и $[g_b(\phi)]_{kl}$ являются вкладами в невязку от интегралов по области решения V и по границе S соответственно:

$$[g_c(\phi)]_{kl} = V_k([A]_k \cdot [\nabla\phi]_k) \cdot \nabla_{kl} \ , \quad [g_b(\phi)]_{kl} = S_k [P_n]_k / 3$$

В тех граничных узлах, в которых искомая функция задана, надо невязку занулить.

Невязка $[g(\phi)]_i$ соответствует уравнению

$$[g(\phi)]_i = \sum_{j=1}^{N_v} [A]_{ij}[\phi]_j - b_i$$

но вычислена напрямую из вариационного уравнения, минуя стадию формирования матрицы жесткости $[A]_{ij}$ и вектора правой части b_i. Если надо вычислить однородную часть невязки

$$[g_0(\phi)]_i = \sum_{j=1}^{N_v} [A]_{ij}[\phi]_j$$

то из выражений для компонент вектора полной невязки исключаются члены, не содержащие искомую функцию в качестве множителя.

Рассмотрим теперь итерационный процесс решения, использующий алгоритм вычисления невязок. Если краевая задача характеризуется положительным самосопряженным оператором[1], как в нашем случае, то имеет смысл воспользоваться классической версией метода сопряженных градиентов с предобусловливанием, который уже был рассмотрен в разделе про итерационные методы для линейных алгебраических уравнений.

Итак, в соответствии с методом сопряженных градиентов сначала задается некоторое приближение к решению $\phi^{(0)}$, по нему вычисляется соответствующая невязка $\mathbf{g}^{(0)} = \mathbf{g}(\phi^{(0)})$ и назначается начальное направление поиска решения $\mathbf{s}^{(0)} = \mathbf{g}^{(0)}$. Далее в итерационном процессе для значений счетчика итераций $n = 0, 1, ..$ требуется только вычисление однородной части невязки $\mathbf{g}_0(\mathbf{s}^{(\mathbf{n})})$ с использованием в качестве аргумента текущего направления поиска $\mathbf{s}^{(n)}$. Итерационный процесс имеет вид

$$\phi^{(n+1)} = \phi^{(n)} - \alpha^{(n)}\mathbf{s}^{(n)}$$

[1] дискретные аналоги краевых задач с самосопряженными положительными операторами характеризуются СЛАУ с симметричными положительными матрицами.

$$\mathbf{g}^{(n+1)} = \mathbf{g}^{(n)} - \alpha^{(n)}\mathbf{g}_0(\mathbf{s}^{(n)})$$
$$\mathbf{s}^{(n+1)} = \mathbf{g}^{(n+1)} - \beta^{(n)}\mathbf{s}^{(n)}$$

где коэффициенты $\alpha^{(n)}$ и $\beta^{(n)}$ определяются формулами

$$\alpha^{(n)} = \frac{\mathbf{g}^{(n)} \cdot \mathbf{s}^{(n)}}{\mathbf{g}_0(\mathbf{s}^{(n)}) \cdot \mathbf{s}^{(n)}}$$

$$\beta^{(n)} = \frac{\mathbf{g}^{(n+1)} \cdot \mathbf{g}_0(\mathbf{s}^{(n)})}{\mathbf{g}_0(\mathbf{s}^{(n)}) \cdot \mathbf{s}^{(n)}}$$

здесь точки означают скалярное произведение векторов размерности N_v (число неизвестных). Процесс итераций считается законченным, если нормы невязки и поправки к решению становятся достаточно малыми

$$\mathbf{g}^{(n)} \cdot \mathbf{g}^{(n)} < \epsilon^2 \quad \vee \quad (\alpha^{(n)}\mathbf{s}^{(n)}) \cdot (\alpha^{(n)}\mathbf{s}^{(n)}) < \epsilon^2$$

здесь ϵ - машинное эпсилон, то есть максимальное число, добавление которого к единице компьютер не чувствует: добавляй его хоть миллион раз, результатом будет единица. Для четырехбайтовой арифметики $\epsilon \approx 0.000001$.

Если число итераций превысило число неизвестных, а решение не найдено, то задача является плохо обусловленной. Если в процессе итераций нарушилось свойство положительной определенности дискретного оператора задачи

$$\mathbf{g}_0(\mathbf{s}^{(n)}) \cdot \mathbf{s}^{(n)} < \epsilon^2$$

то задача является некорректной. Эти случаи могут иметь место из-за ошибок в исходных данных.

Для реализации процесса решения краевой задачи помимо данных о сетке, данных о физических коэффициентах уравнений и граничных условий дополнительно требуется всего-навсего 4 рабочих массива ϕ, $\mathbf{g}$, $\mathbf{g}_0$, $\mathbf{s}$ размерности N_v (где N_v - число неизвестных), так что все числовые данные можно держать в оперативной памяти современного (2008) персонального компьютера

даже для задач с сотнями тысяч неизвестных. Число итераций, затрачиваемых на решение, зависит от начального приближения и используемого предобусловливателя, оно, как правило, не превышает $\sqrt{N_v}$.

Для устойчивости метода вполне достаточно использовать простейшее предобусловливание с помощью приближенной обратной матрицы, полученной обращением диагональной составляюшей матрицы дискретных уравнений. Формула для вычисления диагональных элементов имеет вид:

$$D_{ii} = \sum_{k=1}^{N_e} \sum_{l=1}^{M_e} V_k([A]_k \cdot \nabla_{kl}) \cdot \nabla_{kl} \tilde{H}(i - J(k,l))$$

При программировании суммам будут соответствовать циклы. Для более точных конечно-элементных аппроксимаций приведенные выше формулы для вычисления невязок сохранят свою структуру, первый цикл будет соответствовать обходу всех гауссовых точек численного интегрирования, второй цикл по прежнему будет отвечать перебору узлов конечного элемента, к которому относится гауссова точка. Конечно, значения коэффициентов в формулах интерполяции решения и его производных (дискретный набла-оператор) в данной гауссовой точке изменяться, но структура интерполяционных формул сохранится, поэтому алгоритм решения в целом останется прежним.

12.9 Обоснование консервативности МКЭ

Рассмотрим вопрос о консервативности метода конечных элементов. Нетрудно увидеть, рассматривая дискретное вариационное уравнение, что оно состоит из групп членов, которые для каждой гауссовой точки k описывают распределение потоков величины ϕ между соседними узлами $J(k,l)$, $(l = 1, ..., M_e)$. При

этом на долю каждого соседа $J(k,l)$ гауссовой точки k приходится следующий вклад

$$V_k([A]_k \cdot [\nabla\phi]_k) \cdot \nabla_{kl}$$

В силу следующего свойства дискретного оператора дифференцирования

$$\sum_{l=1}^{M_e} \nabla_{kl} = 0$$

означающего равенство нулю производной от константы, сумма вкладов по всем соседним узлам для каждой гауссовой точки k равна нулю, что и означает *локальную консервативность* метода конечных элементов.

Глобальная консервативность метода конечных элементов обеспечивается, во-первых, его локальной консервативностью и, во-вторых, корректным заданием граничных условий. Последнее означает, что граничные условия должны удовлетворять интегральный закон сохранения (7), записанный для всей области решения.

Отметим, что изначально метод конечных элементов сложился в результате прямого применения закона сохранения импульса (уравнений равновесия сил) к расчету напряженно-деформированного состояния сложных дискретных стержневых систем минуя стадию формулировки и дискретизации краевых задач. То есть требование консервативности с самого начала было заложено в алгоритмы МКЭ. Безусловно, при неудачной или неаккуратной реализации консервативность МКЭ, как и любого другого метода, может быть нарушена, поэтому контроль соблюдения свойства консервативности в процессе численного решения необходим.

В литературе (особенно в зарубежной), ориентированной на решение задач гидродинамики методами контрольных объемов,

в обзорах нередко встречаются утверждения о неконсервативности МКЭ, сопровождающиеся рекламой метода контрольных объемов, как чуть ли не единственного истинно консервативного локально и глобально метода. В приводимых "доказательствах" неконсервативности МКЭ либо критикуется абсолютно далекая от практики неконсервативная формулировка МКЭ, либо проверяется выполнение интегральных законов сохранения путем некорректного применения концепции контрольных объемов непосредственно к конечным элементам. При этом всегда игнорируется тот факт, что конечные элементы отнюдь не являются контрольными объемами, так как не могут быть истолкованы как окрестности каких-либо узлов. Таким образом проявляется непонимание того, что разбиение области решения на контрольные объемы и конечные элементы принципиально различны. В конечных элементах узлы, с которыми ассоциированы искомые значения решения, лежат на границах ячеек в то время, как в случае контрольных объемов каждый узел находится в центре своей ячейки.

Как было показано выше, для обоснования консервативности МКЭ понятие контрольных объемов не требуется. Так же обстоит дело и с обоснованием консервативности бессеточных методов и методов частиц.

12.10 Двухточечные краевые задачи

Важным частным случаем эллиптических краевых задач являются двухточечные краевые задачи, то есть, другими словами, одномерные задачи эллиптического типа. Для таких задач разработаны специализированные численные методы решения, называемые дифференциальными прогонками. Эти методы не требуют решения систем алгебраических уравнений высокого порядка и эффективно реализуются на компьютерах низкой производительности. Эти методы позволяют эффективно решать плохо обуслов-

ленные задачи с узкими пограничными слоями, например, задачи моментных теорий тонких непологих оболочек. Приведем ниже краткие сведения об этих методах.

Двухточечная краевая задача для искомых функций $y_i(x)$ имеет вид [1]:

$$x \in [x_a, x_b] \ : \quad \frac{dy_i}{dx} = F_i(x, y) \quad i = 1, ..., 2N$$

$$x = x_a \ : \quad \sum_{j=1}^{2N} B_{ij}^{(a)} y_j(x_a) - \alpha_i = 0 \quad i = 1, ..., N$$

$$x = x_b \ : \quad \sum_{j=1}^{2N} B_{ij}^{(b)} y_j(x_b) - \beta_i = 0 \quad i = 1, ..., N$$

где функции F_i, матрицы $B_{ij}^{(a)}$, $B_{ij}^{(b)}$ и векторы α_i, β_i являются заданными.

Простейшим способои решения данной задачи является метод пристрелки или метод стрельбы. Пусть $y_i^{(k,a)}$ $(i = 1, ..., 2N; k = 0, 1, ...)$ - значения пробного решения номер k при $x = x_a$.

Определим это пробное решение как решение вспомогательной задачи Коши

$$x \in [x_a, x_b] \ : \quad \frac{dy_i}{dx} = F_i(x, y) \quad i = 1, ..., 2N$$

$$x = x_a \ : \quad y_i(x_a) = y_i^{(k,a)} \quad i = 1, ..., 2N$$

и найдем тем самым значения этого пробного решения $y_i^{(k,b)}$ $(i = 1, ..., 2N)$ при $x = x_b$. Определенные таким образом пробные решения удовлетворяют системе дифференциальных уравнений исходной задачи, но в общем случае не удовлетворяют граничным

[1] Здесь рассмотрены обычные одномерные эллиптические краевые задачи с равным количеством граничных условий на концах отрезка, представляющего область решения.

условиям. Метод пристрелки решения исходной задачи состоит в минимизации функционала невязок граничных условий

$$F(y_1^{(k,a)},...,y_N^{(k,a)}) = \sum_{i=1}^{N}(\sum_{j=1}^{2N} B_{ij}^{(a)} y_j^{(k,a)} - \alpha_i)^2 +$$

$$+ \sum_{i=1}^{N}(\sum_{j=1}^{2N} B_{ij}^{(b)} y_j^{(k,b)} - \beta_i)^2$$

по значениям пробных решений на левом краю $y_i^{(k,a)}$ $(i = 1, ...2N)$.

Полученная задача на минимум функционала является непростой. Во-первых, вычисление значений функционала проводится с помошью решений вспомогательных задач Коши. Во-вторых, в общем случае решение этой задачи минимизации может быть неединственным. В третьих, задача относительно значений решения на левом краю $(x = x_a)$ может быть плохо обусловленной, то есть небольшие возмущения правых частей уравнений и граничных условий могут вызывать катастрофически большие искажения решения.

Поэтому для улучшения обусловленности отрезок $[x_a, x_b]$ области решения разбивают на M достаточно коротких участков $[x_a^{(m)}, x_b^{(m)}]$ $(l = 1, ..., M)$, где $x_a^{(1)} = x_a$, $x_b^{(m)} = x_a^{(m+1)}$ $(m = 1, ..., M-1)$, $x_b^{(M)} = x_b$ и минимизируется функционал

$$F(y_1^{(k,a)},...,y_N^{(k,a)}) = \sum_{i=1}^{N}(\sum_{j=1}^{2N} B_{ij}^{(a)} y_j^{(k,a)} - \alpha_i)^2 +$$

$$+ \sum_{i=1}^{N}(\sum_{j=1}^{2N} B_{ij}^{(b)} y_j^{(k,b)} - \beta_i)^2 +$$

$$+ \sum_{m=1}^{M-1}\sum_{j=1}^{2N}(y_j^{(k,b,m)} - y_j^{(k,a,m+1)})^2$$

где в третьей строке добавлены условия сопряжения пробных решений на соседних участках. Значения пробных решений на левых и правых границах участков $[x_a^{(m)}, x_b^{(m)}]$ $(l = 1, ..., M)$, а именно, $y_i^{(k,a,m)}$ и $y_i^{(k,b,m)}$, связаны между собой через решения вспомогательных задач Коши для исходного уравнения на этих участках.

Подробный анализ метода пристрелки можно найти в книге Бахвалова, Жидкова, Кобелькова [1987]. Примеры применения этого метода к прикладным задачам о ветвлении решений задач нелинейной теории оболочек даны в монографии Валишвили [1976].

Гораздо чаще рассматриваемая задача решается с использованием метода Ньютона, который сводит исходную нелинейную краевую задачу к последовательности вспомогательных линейных задач

$$x \in [x_a, x_b] \ : \quad \frac{dy_i^{(k+1)}}{dx} = F_i(x, y^{(k)}) + \frac{\partial F_i}{\partial y_j}(x, y^{(k)})(y_j^{(k+1)} - y_j^{(k)})$$

где $i = 1, ..., 2N$ и $k = 0, 1, ...$

$$x = x_a \ : \quad \sum_{j=1}^{2N} B_{ij}^{(a)} y_j^{(k+1)}(x_a) - \alpha_i = 0 \quad i = 1, ..., N$$

$$x = x_b \ : \quad \sum_{j=1}^{2N} B_{ij}^{(b)} y_j^{(k+1)}(x_b) - \beta_i = 0 \quad i = 1, ..., N$$

$$k = 0 \ : \quad y_i^{(k)} = y_{i(0)}(x)$$

Часто применяется и метод дифференцирования по параметру, который сводит исходную нелинейную задачу к задаче Коши

$$\frac{dy_i}{dt} = \dot{y}_i(x)$$

$$t = 0 \ : \ y_i = y_{i(0)}(x)$$

где $t \in [0, \infty]$ и "скорости" изменения решения $\dot{y}_i(x,t)$ определяются следующей линейной двухточечной краевой задачей:

$$x \in [x_a, x_b] \ : \quad \frac{d}{dx}\dot{y}_i = \dot{F}_i + \frac{\partial F_i}{\partial y_j}(x,y)\dot{y}_j \ ; \quad i = 1, ..., 2N$$

$$x = x_a \ : \quad \sum_{j=1}^{2N} B_{ij}^{(a)} \dot{y}_j(x_a) - \dot{\alpha}_i = 0 \ ; \quad i = 1, ..., N$$

$$x = x_b \ : \quad \sum_{j=1}^{2N} B_{ij}^{(b)} \dot{y}_j(x_b) - \dot{\beta}_i = 0 \ ; \quad i = 1, ..., N$$

Задача Коши решается с помощью какого-либо метода пошагового интегрирования по "времени" t. Заметим, что роль параметра t может играть время, параметр нагружения. скорость набегающего потока и так далее. Этот параметр может быть также искусственно введен в исходные уравнения.

И в методе Ньютона, и в методе дифференцирования по параметру основной операцией является решение линейной двухточечной краевой задачи, которую для дальнейшего изложения запишем в виде

$$x \in [x_a, x_b] \ : \quad \frac{dy_i}{dx} = \sum_{j=1}^{2N} A_{ij}(x) y_j(x) + f_i(x) \ ; \quad i = 1, ..., 2N$$

$$x = x_a \ : \quad \sum_{j=1}^{2N} B_{ij}^{(a)} y_j(x_a) - \alpha_i = 0 \quad i = 1, ..., N$$

$$x = x_b \ : \quad \sum_{j=1}^{2N} B_{ij}^{(b)} y_j(x_b) - \beta_i = 0 \quad i = 1, ..., N$$

Общее решение линейной двухточечной краевой задачи имеет вид

$$y_i(x) = y_{i(0)}(x) + \sum_{i=1}^{2N} c_i y_{i(j)}(x)$$

где $y_{i(0)}(x)$ решение вспомогательной задачи Коши для неоднородного уравнения

$$x \in [x_a, x_b] \ : \quad \frac{dy_{i(0)}}{dx} = \sum_{j=1}^{2N} A_{ij}(x) y_{j(0)}(x) + f_i(x) \ ; \quad i = 1, ..., 2N$$

$$x = x_a \ : \quad y_{i(0)} = 0 \quad (i = 1, ..., 2N)$$

а $y_{i(j)}(x)$ являются решениями вспомогательных задач Коши для однородного уравнения

$$x \in [x_a, x_b] \ : \quad \frac{dy_{i(j)}}{dx} = \sum_{k=1}^{2N} A_{ik}(x) y_{k(j)}(x) \ ; \quad i = 1, ..., 2N$$

$$x = x_a \ : \quad y_{i(j)} = \delta_{ij} \quad (i, j = 1, ..., 2N)$$

где $\{\delta_{ij}\}$ - единичная матрица 2Nx2N. Коэффициенты $c_i \ \ (i = 1, ..., 2N)$ определяются из граничных условий.

Для улучшения обусловленности системы алгебраических уравнений для коэффициентов $c_i \ \ (i = 1, ..., 2N)$, как и в методе пристрелки, отрезок интегрирования делится на участки, на каждом из которых решение ищется в виде своего разложения по решениям вспомогательных задач Коши и граничные условия дополняются условиями непрерывности решения на стыках участков. Система уравнений для коэффициентов разложения решения на участках имеет размерность $2NM$, где M - число участков. Она характеризуется блочно-ленточной матрицей и решается методом матричной прогонки. Этот способ решения носит

название метода Калнинса (1964). Поскольку решение на участках определяется интегрированием дифференциального уравнения, данный способ решения классифицируется как простейший метод дифференциальной прогонки.

Другие более сложные варианты дифференциальных прогонок предложены Абрамовым (1961) и Годуновым (1961) и подробно описаны в книге Бахвалова (1973).

Наконец, двухточечные краевые задачи можно решать и с использованием общих методов решения, разработанных для многомерных задач, что чаще всего и делается в настоящее время (2008г.), так как экономить количество операций в одномерных задачах на мощных ЭВМ смысла не имеет, а с плохой обусловленностью так или иначе все равно приходится бороться и в многомерном случае.

Глава 13

Решение параболических уравнений

13.1 Формулировка задачи

В этом разделе рассмотрим основные способы решения начально-краевых задач для параболических уравнений в частных производных второго порядка.

Постановка типичной начально-краевой задачи для параболического уравнения имеет следующий вид. В некоторой пространственно-временной области

$$V_t = \{(\mathbf{x}, t) \mid \mathbf{x} \in V \ , \ \ t \geq 0\}$$

с границей

$$S_t = \{(\mathbf{x}, t) \mid \mathbf{x} \in \partial V \ , \ \ t \geq 0\}$$

требуется найти функцию $\mathbf{u}(\mathbf{x}, \mathbf{t})$, удовлетворяющую уравнению

$$(\mathbf{x}, t) \in V_t \ : \ \partial_t \mathbf{u} = \nabla \cdot (\mathbf{A}(\mathbf{u}) \cdot \nabla \mathbf{u}) + C(\mathbf{u}) \tag{1}$$

граничным условиям

$$(\mathbf{x}, t) \in S_t^{(u)} \ : \quad \mathbf{u} = \mathbf{u}_*(\mathbf{x}, t)$$

$$(\mathbf{x}, t) \in S_t \backslash S_t^{(u)} \ : \quad \mathbf{n} \cdot (\mathbf{A} \cdot \nabla \mathbf{u}) = \mathbf{P}_n^*(\mathbf{x}, t) \tag{2}$$

и начальным условиям

$$t = 0 \ : \quad \mathbf{u} = \mathbf{u}_*^{(0)}(\mathbf{x}) \tag{3}$$

где тензор коэффициентов диффузии $\mathbf{A}(\mathbf{u}, \mathbf{x}, t)$ и источниковый член $\mathbf{C}(\mathbf{u}, \mathbf{x}, t)$ являются заданными функциями искомого решения $\mathbf{u}$, координат и времени, вектор $\mathbf{n}$ является единичной

внешней нормалью к границе S пространственной области V, часть границы, на которой заданы значения искомой функции, обозначена $V_t^{(u)}$, заданные функции в правых частях граничных и начальных условий отмечены звездочками, $\mathbf{P}_n$ - поток величины $\mathbf{u}$ через границу.

Здесь и далее используется сокращенная запись уравнений. Уравнение (1) в развернутой форме выглядит так

$$\partial_t u_\alpha = \sum_{i=1}^{3} \partial_i \Big(\sum_{j=1}^{3} \sum_{\beta=1}^{N} A_{\alpha\beta}^{ij}(\mathbf{u}) \partial_j u_\beta \Big) + C_\alpha(\mathbf{u}) \qquad (1')$$

где N - число искомых функций u_α $(\alpha = 1, ..., N)$, $\partial_t = \partial/\partial t$, $\partial_i = \partial/\partial x_i$. Сравнивая (1) и (1') видим, что использование сокращенной записи делает изложение более ясным.

Для корректной постановки задачи необходимо выполнение условия Адамара, требующего положительности оператора A:

$$\forall \nabla \mathbf{u} \neq 0 \ : \quad (\mathbf{A} \cdot \nabla \mathbf{u}) \cdot \nabla \mathbf{u} > 0 \qquad (4)$$

Характерным свойством решений параболических начально-краевых задач является то, что, так же как и в эллиптических краевых задачах, возмущения в условиях задачи сразу оказывают воздействие на решение повсюду в области решения, но любое возмущение решения затухает с течением времени и его влияние на решение ослабевает по мере удаления от места его приложения. Поэтому малые параболические члены (диффузионные члены) нередко вводятся в эволюционные уравнения как средство сглаживания решений.

Если в исходное уравнение (1) дописать слагаемые с первыми пространственными производными от искомой функции, то формально тип уравнения не изменится (постановка граничных и начальных условий сохраняется). Однако модифицированные таким образом задачи могут содержать пограничные слои ("погранслои"), то есть зоны всплеска градиентов решения и потребуют специальных методов решения задач с малым параметром

при старших производных.

Консервативная запись исходного уравнения (1)

$$\partial_t \mathbf{u} = \nabla \cdot (\mathbf{A} \cdot \nabla \mathbf{u}) + \mathbf{C} \tag{5}$$

для произвольного объема $\tilde{V}$ с поверхностью $\tilde{S}$ поддерживает закон сохранения величины $\mathbf{u}$:

$$\int\limits_{\tilde{V}} \partial_t \mathbf{u} dV = \int\limits_{\tilde{S}} \mathbf{n} \cdot (\mathbf{A} \cdot \nabla \mathbf{u}) dS + \int\limits_{\tilde{V}} \mathbf{C} dV \tag{6}$$

который получается интегрированием исходного уравнения и заменой интеграла по объему интегралом по границе в соответствии с теоремой о дивергенции, обозначение $\mathbf{n}$ использовано для единичной внешней нормали к границе.

Из уравнения (6) видно, что в случае краевых условий Неймана, заданных на всей границе ($S_u = \emptyset$), для существования установившегося решения ($\partial_t \mathbf{u} = 0$) требуется согласование правой части граничного условия для потока (2) P_n со свободным членом

$$\int\limits_{S \setminus S_u} \mathbf{P}_n dS + \int\limits_{V} \mathbf{C} dV = 0 \tag{7}$$

Установившееся решение отвечает в пределе при $t \to \infty$ решению эллиптического уравнения, в которое превращается рассматриваемое параболическое уравнение при $\partial_t \mathbf{u} = 0$.

Неконсервативная форма записи получается дифференцированием в уравнении (5) сомножителей, определяющих поток. Она имеет вид [1]

$$\partial_t \mathbf{u} = (\nabla \cdot \mathbf{A}) \cdot \nabla \mathbf{u} + \mathbf{A} : (\nabla \otimes \nabla) \mathbf{u} + \mathbf{C} \tag{8}$$

или, в случае уравнений с постоянными коэффициентами

$$\partial_t \mathbf{u} = \mathbf{A} : (\nabla \otimes \nabla) \mathbf{u} + \mathbf{C} \tag{9}$$

[1] Внешнее произведение $\otimes$ образует из двух векторов тензор второго ранга $\mathbf{a} \otimes \mathbf{b} = a_i b_j \mathbf{e}_i \mathbf{e}_j$ и так далее.

Как и в случае эллиптических уравнений использование неконсервативных форм записи (8) и (9) чревато неприятными сюрпризами, которые уже рассматривались в главе про эллиптические уравнения.

Решения начально-краевых задач для параболических уравнений удовлетворяют принципу максимума [69], в соответствии с которым при отсутствии источников максимальное и минимальное значения решения достигаются на границах пространственно-временной области решения V_t.

13.2 Методы для параболических задач

Рассмотрим методы решения начально-краевых задач для параболических уравнений.

Если каким-либо из способов, описанных в главе про эллиптические краевые задачи, ввести пространственную дискретизацию искомого решения, то для параболических начально-краевых задач получается система обыкновенных дифференциальных уравнений по времени

$$\partial_t[\mathbf{u}] = [\mathbf{B}][\mathbf{u}] + [\mathbf{b}]$$

где матрица $[\mathbf{B} = \mathbf{B}(\mathbf{u})]$ и вектор свободных членов $[\mathbf{b} = \mathbf{b}(\mathbf{u})]]$ правой части вычисляются так же, как и в случае эллиптических уравнений (см. формулы (20-23) раздела 11). Эту систему уравнений можно решать какими-либо пошаговыми явными или неявными методами решения задач Коши. Например, можно использовать следующую двухслойную схему

$$\frac{\mathbf{u}_i^{n+1} - \mathbf{u}_i^n}{\tau^n} = \{[\mathbf{D}][\mathbf{u}] + [\mathbf{b}]\}_i^{n+\alpha}$$

где $\tau^n = t^{n+1} - t^n$ - шаг по времени, верхний индекс показывает номер временного слоя, нижний индекс показывает номер узла. Значения переменных на промежуточном временном слое $t^{n+\alpha}$ определены по правилу: $\mathbf{u}_i^{n+\alpha} = (1-\alpha)\mathbf{u}_i^n + \alpha\mathbf{u}_i^{n+1} \quad (0 \le \alpha \le 1)$.

Решения нелинейных задач по неявным схемам определяются итерациями по нелинейности как предел последовательности решений вспомогательных линеаризованных задач.

При $\alpha = 0$ имеем явный метод Эйлера, при $\alpha = 1$ имеем неявный метод Эйлера, при $\alpha = 1/2$ - неявный метод Кранка-Николсона[1]. Методы Эйлера имеют первый, а схема Кранка-Николсона второй порядки точности по времени.

Явный метод Эйлера условно устойчив. При аппроксимации эллиптического оператора центральными разностями второго порядка точности, условие устойчивости явного метода Эйлера имеет вид

$$\tau^n < \min_i(h_i^2/((d+1)!\nu_i^n))$$

где h_i - минимальный шаг по пространственным переменным в окрестности узла i, ν_i^n максимальное собственное число матрицы коэффициентов диффузии $\mathbf{A}$ в окрестности узла i, d - число независимых пространственных переменных.

Неявные разностные схемы при $\alpha \geq 0.5$ безусловно устойчивы и при их применении шаг по времени ограничен только условием точности

$$\tau^n < \gamma||\mathbf{u}^n||/||\partial_t[\mathbf{u}]^n||$$

где $0 < \gamma << 1$. Условие точности означает, что норма приращения решения $\Delta\mathbf{u}^n$ на шаге по времени должна быть малой по сравнению с нормой самого решения $\mathbf{u}^n$. Переход к неявным аппроксимациям диффузионного оператора производится, если диффузионное условие устойчивости для явных схем слишком ограничивает величину шага по времени.

Если коэффициент диффузии ν мал, но величина $||(\partial[\mathbf{b}]/\partial\mathbf{u})_i^{n+\alpha}||$ велика, то условие точности может быть более ограничительным, чем диффузионное условие устойчивости. В этих условиях схема с явной аппроксимацией

[1]Иногда применяют термин "метод Кранка-Николсон".

диффузионного оператора и неявной аппроксимацией свободного члена

$$\frac{\mathbf{u}_i^{n+1} - \mathbf{u}_i^n}{\tau^n} = [\mathbf{D}]_i^n + [\mathbf{b}]_i^n + \frac{\partial [\mathbf{b}]_i^n}{\partial \mathbf{u}_i^n} \mathbf{u}^{n+\alpha}$$

позволяет получать достаточно точные решения при удовлетворении необременительного в этом случае диффузионного условия устойчивости. Если линеаризовать свободный член, как это уже сделано в выписанной формуле, то матрица системы уравнений относительно величин на новом временном слое останется диагональной, как в случае явной схемы.

Существует явная трехслойная схема, которая позволяет ослабить обременительное диффузионное ограничение шага по времени. Это схема Дюфорта-Франкела, которая имеет вид:

$$\frac{\mathbf{u}_i^{(n+1)} - \mathbf{u}_i^{(n)}}{\tau_n} + \left(\frac{\alpha^* \tau_n}{h_i}\right)^2 \frac{\mathbf{u}_i^{(n+1)} - 2\mathbf{u}_i^{(n)} + \mathbf{u}_i^{(n-1)}}{\tau_n^2} = \{[\mathbf{D}][\mathbf{u}] + [\mathbf{b}]\}_i^n$$

Эта схема не нарушает диагональности матрицы СЛАУ относительно новых значений на (n+1)-м слое и, таким образом, остается явной. Для того, чтобы дискретное уравнение аппроксимировало исходное уравнение, требуется выполнить условие

$$\tau_n \leq h_i^2/(2\alpha^*)$$

где α^* - произвольная постоянная, имеющая размерность коэффициента диффузии. Это условие обеспечивает малость второго члена в левой части дискретного уравнения. Хотя условие устойчивости схемы Дюфорта-Франкела по форме совпадает с диффузионным, оно не так ограничительно, так как значение коэффициента α^* можно взять меньшим, чем физический коэффициент диффузии.

Отметим специфические для параболических задач особенности основных уравнений для построения методов решения.

Балансное соотношение (6) метода конечных объемов (МКО) для параболических задач принимает вид

$$\int\limits_{\tilde{V}_i} \partial_t \mathbf{u} dV = \int\limits_{\tilde{S}_i} \mathbf{n} \cdot \mathbf{A} \cdot \nabla \mathbf{u} dS + \int\limits_{\tilde{V}_i} \mathbf{C} dV$$

Запись исходного вариационного уравнения метода конечных элементов (МКЭ) для параболических краевых задач принимает вид:

$$\int\limits_{V} \left(\partial_t \mathbf{u} \cdot \delta \mathbf{u} + (\mathbf{A} \cdot \nabla \mathbf{u}) \cdot \nabla \delta \mathbf{u}\right) dV = \int\limits_{S_p} \mathbf{P}_n \cdot \delta \mathbf{u} dS + \int\limits_{V} \mathbf{C} \cdot \delta \mathbf{u} dV$$

Подчеркнем, что в случае контрольных объемов интегрирование проводится по произвольному объему, а в случае конечных элементов интегрирование выполняется по всей области решения при произвольных вариациях решения.

Способы аппроксимации диффузионного оператора и свободного члена в МКО и МКЭ рассмотрены ранее в главе про решение краевых задач для эллиптических уравнений. После введения аппроксимаций временных производных в случае явных схем устойчивость численных решений МКО и МКЭ обеспечивается тем же диффузионным ограничением на шаг по времени, что и в случае конечно-разностных схем. Все рассмотренные ранее в главе про эллиптические уравнения методы итераций по нелинейности и итерационные методы решения вспомогательных линеаризованных задач применяются для реализации вычислений на каждом временном шаге неявных схем для параболических уравнений без каких-либо изменений.

Отметим, что в пределе при $t \to \infty$ решения параболических начально-краевых задач стремятся к решению соответствующих эллиптических краевых задач. Это используется в методах установления для получения решений задач эллиптического типа.

Применение метода граничных элементов (МГЭ) к решению параболических задач основывается на сведении параболических задач к последовательности вспомогательных эллиптических задач, которые решается так, как описано в главе про эллиптические задачи. Сведение параболической задачи к последовательности вспомогательных краевых задач эллиптического типа реализуется или использованием какой-либо неявной схемы интегрирования по времени, или, в случае параболических уравнений с постоянными коэффициентами, с использованием интегральных преобразований (по времени) Фурье или Лапласа. Подробное описание вариантов МГЭ для параболических задач можно найти в книге (Бреббия, Уокер,1982)

Аналогично в параболических задачах используются также и описанные в разделе эллиптических задач бессеточные методы.

Глава 14

Решение гиперболических уравнений

14.1 Формулировка задачи

В этом разделе рассмотрим основные способы решения начально-краевых задач для гиперболических уравнений.

Постановка типичной начально-краевой задачи для гиперболического уравнения имеет следующий вид. В некоторой пространственно-временной области

$$V_t = \{(\mathbf{x}, t) \mid \mathbf{x} \in V \ , \quad t \geq 0\}$$

с границей

$$S_t = \{(\mathbf{x}, t) \mid \mathbf{x} \in V \ , \quad t \geq 0\}$$

требуется найти вектор-функцию $\mathbf{u}(\mathbf{x}, \mathbf{t})$, удовлетворяющую уравнению

$$(\mathbf{x}, t) \in V_t \ : \ \partial_t^2 \mathbf{u} = \nabla \cdot (\mathbf{A}(\mathbf{u}) \cdot \nabla \mathbf{u}) + C(\mathbf{u}) \tag{1}$$

граничным условиям

$$(\mathbf{x}, t) \in S_t^{(u)} \ : \quad \mathbf{u} = \mathbf{u}_*(\mathbf{x}, t)$$

$$(\mathbf{x}, t) \in S_t \backslash S_t^{(u)} \ : \quad \mathbf{n} \cdot (\mathbf{A} \cdot \nabla \mathbf{u}) = P_n^*(\mathbf{x}, t) \tag{2}$$

и начальным условиям

$$t = 0 \ : \quad \mathbf{u} = \mathbf{u}_*^{(0)}(\mathbf{x}), \quad \partial_t \mathbf{u} = \mathbf{v}_*^{(0)}(\mathbf{x}) \tag{3}$$

где тензор коэффициентов эллиптического оператора правой части $\mathbf{A}(\mathbf{u}, \mathbf{x}, t)$ и источниковый член $C(\mathbf{u}, \mathbf{x}, t)$ являются заданными функциями искомого решения $\mathbf{u}$, координат и времени, вектор

$\mathbf{n}$ является единичной внешней нормалью к границе S пространственной области V, часть границы, на которой заданы значения искомой функции, обозначена $V_t^{(u)}$, заданные функции в правых частях граничных и начальных условий отмечены звездочками.

Здесь и далее используется сокращенная запись уравнений. Уравнение (1) в развернутой форме выглядит так

$$\partial_t^2 u_\alpha = \sum_{i=1}^{3} \partial_i \Big(\sum_{j=1}^{3} \sum_{\beta=1}^{N} A_{\alpha\beta}^{ij}(\mathbf{u}) \partial_j u_\beta \Big) + C_\alpha(\mathbf{u}) \qquad (1')$$

где N - число искомых функций u_α $(\alpha = 1, ..., N)$, $\partial_t = \partial/\partial t$, $\partial_i = \partial/\partial x_i$. Сравнивая (1) и (1') видим, что использование сокращенной записи делает изложение более ясным.

Задача поставлена корректно, если выполнено условие Адамара, требующее положительности оператора A:

$$\forall \nabla \mathbf{u} \neq 0 \ : \quad (\mathbf{A} \cdot \nabla \mathbf{u}) \cdot \nabla \mathbf{u} > 0 \qquad (4)$$

Если в исходном уравнении (1) дописать члены с первыми пространственными и временными производными от искомой функции, умноженными на заданные коэффициенты, или даже члены, являющиеся нелинейными функциями от первых производных искомой функции, то формально тип уравнения не изменится. Модифицированные таким образом задачи требуют специальных методов решения, рассматриваемых отдельно.

Консервативная запись исходного уравнения (1)

$$\partial_t^2 \mathbf{u} = \nabla \cdot (\mathbf{A} \cdot \nabla \mathbf{u}) + C(\mathbf{u}) \qquad (6)$$

для произвольного объема $\tilde{V}$ с поверхностью $\tilde{S}$ поддерживает закон сохранения величины $\mathbf{u}$:

$$\int\limits_{\tilde{V}} \partial_t^2 \mathbf{u} dV = \int\limits_{\tilde{S}} \mathbf{n} \cdot (\mathbf{A} \cdot \nabla \mathbf{u}) dS + \int\limits_{\tilde{V}} C dV \qquad (7)$$

который получается интегрированием исходного уравнения и заменой интеграла по объему интегралом по границе в соответствии с теоремой о дивергенции, обозначение $\mathbf{n}$ использовано для единичной внешней нормали к границе.

Неконсервативная форма записи получается дифференцированием в уравнении (6) сомножителей, определяющих поток $\mathbf{A} \cdot \nabla \mathbf{u}$, и имеет вид

$$\partial_t^2 \mathbf{u} = (\nabla \cdot \mathbf{A}) \cdot \nabla \mathbf{u} + \mathbf{A} : (\nabla \otimes \nabla)\mathbf{u} + C \tag{8}$$

или, в случае уравнений с постоянными коэффициентами

$$\partial_t^2 \mathbf{u} = \mathbf{A} : (\nabla \otimes \nabla)\mathbf{u} + C \tag{9}$$

Использование неконсервативных форм записи (8) и (9) при численном решении задач чревато неприятными сюрпризами, которые уже рассматривались в главе про эллиптические уравнения.

Рассмотрим методы решения начально-краевых задач для гиперболических уравнений с частными производными второго порядка.

Как и в случае параболических задач можно ввести пространственную дискретизацию решения, тогда придем к системе обыкновенных дифференциальных уравнений по времени

$$\partial_t^2 [\mathbf{u}] = [\mathbf{B}][\mathbf{u}] + [\mathbf{b}]$$

где $[\mathbf{u}]$ - вектор дискретного решения, матрица $[\mathbf{B}]$ и вектор свободных членов $[\mathbf{b}]$ правой части, отвечают какому-либо из уже рассмотренных методов дискретизации пространственного эллиптического оператора задачи.

Можно с самого начала ввести дискретизацию решения по времени, тогда для неявных аппроксимаций для определения искомого решения на каждом новом временном слое будем иметь

вспомогательные эллиптические краевые задачи, способы решения которых уже рассмотрены. Примером может служить семейство трехслойных схем

$$\frac{[\mathbf{u}]_i^{n+1} - 2[\mathbf{u}]_i^n + [\mathbf{u}]^{n-1}}{\tau^n} = \{[\mathbf{D}][\mathbf{u}] + [\mathbf{b}]\}_i^{n+\alpha}$$

где верхний индекс показывает номер временного слоя, нижний индекс показывает номер узла, $\tau^n = t^{n+1} - t^n$ - шаг по времени. Значения переменных на промежуточном временном слое $t^{n+\alpha}$ определены по правилу: $[\mathbf{u}]_i^{n+\alpha} = (1-\alpha)[\mathbf{u}]_i^n + \alpha[\mathbf{u}]_i^{n+1}$, $(0 \le \alpha \le 1)$.

При $\alpha = 0$ имеем явную схему "крест", при $0 < \alpha \le 1$ имеем варианты неявных схем.

14.2 Характеристическая форма гиперболических уравнений

Гиперболические системы уравнений обладают рядом специфических свойств, которые следует принимать во внимание в процессе конструирования численных методов решения. В отличие от уравнений эллиптического и параболического типов уравнения гиперболического типа можно представить в виде системы уравнений в частных производных первого порядка, разрешенных относительно временных производных.

$$\partial_t \mathbf{w} + \sum_{k=1}^{3} A_{(k)}(\mathbf{w}) \partial_{x_k} \mathbf{w} + \mathbf{a}(\mathbf{w}) = 0 \qquad (10)$$

где матрицы $A_{(k)}$ и вектор свободных членов $\mathbf{a}$ являются заданными функциями от искомого решения $\mathbf{w}$, координат x_k $(k = 1, 2, 3)$ и времени t. Обозначение для вектора искомых функций изменено не случайно, а намеренно, чтобы подчеркнуть

отличие этого набора искомых функций от набора искомых функций $\mathbf{u}$ из предыдущего раздела, где рассматривалась система гиперболических уравнений второго порядка.

Для гиперболических уравнений в любой точке четырехмерного пространства $(\mathbf{x}, t)$ существует по меньшей мере одна поверхность $\varphi(\mathbf{x}, t) = 0$, с которой решение не может быть продолжено в область. Чтобы продолжить заданное на некоторой поверхности решение в область надо по заданному на ней решению определить его временную и пространственные производные, а затем использовать разложение решения в ряд Тейлора в окрестности рассматриваемой точки поверхности.

Система уравнений для определения временной и пространственных производных по заданному на поверхности решению состоит из исходной квазилинейной системы уравнений для $\mathbf{w}$ и выражений для дифференциалов решения вдоль линий пересечения этой поверхности с координатными плоскостями (t, x_k):

$$d_{(k)}\mathbf{w} - \partial_t \mathbf{w} dt - \partial_{x_k} \mathbf{w} dx_k = 0 \quad (k = 1, 2, 3) \tag{11}$$

где суммирования по k нет.

Для каждой из трех координатных плоскостей (t, x_k) разрешающие системы уравнений для определения производных $\partial_{x_k}\mathbf{u}$ получается исключением временных производных в соотношениях (11) с помощью (10) и имеют вид

$$(A_{(k)} - \frac{dx_k}{dt} E) \cdot \partial_{x_k} \mathbf{w} = -(\mathbf{a} + \sum_{\substack{j=1,2,3 \\ j \neq k}} A_{(j)} \partial_{x_j} \mathbf{w} - \frac{d_{(k)}\mathbf{w}}{dt}) \tag{12}$$

где $k = 1, 2, 3$, E - единичная матрица. Производные по направлениям, не лежащим в плоскости (t, x_k), не влияют на разрешимость задачи определения производных $\partial_{x_k}\mathbf{w}$, поэтому соответствующие члены рассматриваются как заданные и отнесены в правую часть разрешающих уравнений.

Искомые линии пересечения характеристической поверхности $\varphi(\mathbf{x}, t) = 0$ с координатными плоскостями (t, x_k)

$$\frac{dx_k}{dt} = \lambda_{(k)} = -\frac{\partial\varphi}{\partial t}/\frac{\partial\varphi}{\partial x_k}$$

определяются из условий вырождения матриц $A_{(k)} - \lambda_{(k)}E$ разрешающих систем уравнений

$$\det(A_{(k)} - \lambda^{(k)}E) = 0$$

Если собственные числа матриц $A_{(k)}$ вещественны, то характеристические поверхности в данной пространственно-временной точке существуют и, следовательно, исходная система уравнений является гиперболической. Вдоль характеристических линий (или, короче, вдоль характеристик) решение подчиняется так называемым характеристическим соотношениям. Они выписываются с помощью левых собственных векторов $\mathbf{l}_i^{(k)}$ матрицы $A_{(k)}$, которые определяются как нетривиальные решения следующих уравнений:

$$(\mathbf{l}_i^{(k)})^T(A_{(k)} - \lambda_i^{(k)}E) = 0$$

где $k = 1, 2, 3$. Умножая исходную систему уравнений, записанную в нормальной форме, слева на собственные векторы, получаем характеристические соотношения

$$\mathbf{l}_i^{(k)} \cdot (\partial_t\mathbf{w} + \lambda_i^{(k)}\partial_{x_k}\mathbf{w} + \sum_{j\neq k} A_{(j)}\partial_{x_j}\mathbf{w} + \mathbf{a}) = 0$$

выполняющиеся вдоль i-й характеристической линии

$$\frac{dx_k}{dt} = \lambda_i^{(k)}$$

в координатной плоскости (t, x_k).

Если матрицы $A_{(k)}$ имеют постоянные коэффициенты, то соотношения на характеристике можно переписать так:

$$\Delta(\mathbf{l}_i^{(k)} \cdot \mathbf{w}) = -\mathbf{l}_i^{(k)} \cdot (\sum_{j\neq k} A_{(j)}\partial_j\mathbf{w} + \mathbf{a})\Delta t$$

где стоящие в левой части под знаком приращения Δ величины $r_i^{(k)} = \mathbf{l}_i^{(k)} \cdot \mathbf{w}$ называются инвариантами Римана. При равенстве нулю правых частей характеристических соотношений инварианты Римана сохраняются вдоль характеристик.

Общее число характеристических соотношений, выполняющихся на характеристиках, проходящих через данную точку, равно числу искомых функций в системе гиперболических уравнений. В каждой граничной точке пространственно-временной области решения надо задать столько граничных условий, сколько характеристических соотношений определено на характеристиках, уходящих из данной граничной точки за пределы области решения. Эти свойства используются для того, чтобы построить для гиперболических уравнений численные методы характеристик.

14.3 Метод характеристик

Пусть в пространственной области решения определена сетка расчетных узлов, то есть заданы координаты узлов и определено отношение соседства узлов с помощью группирования их в ячейки или шаблоны. Пусть дискретное решение гиперболической задачи, представленное узловыми значениями искомой вектор-функции $\mathbf{w_i^n}$, в некоторый момент времени t^n известно.

Если характеристики, выпущенные из узла x_i на новом временном слое $t = t^{n+1}$ пересекают старый временной слой $t = t^n$ в точках x_i^α, где $\alpha = 1, ..., N$ и N - число искомых функций, то решение в этом узле определяется из системы N характеристических соотношений, связывающих искомые значения решения $\mathbf{w_i^{n+1}}$ с известными значениями решения в точках x_i^α старого временного слоя.

В общем случае точки x_i^α не совпадают с узлами используемой сетки, поэтому значения решения в этих точках определяются интерполяцией. Таким образом, для каждой внутренней точки x_i на

новом временном слое решение определяется путем формирования и решения системы характеристических соотношений.

В граничных точках часть характеристических соотношений недоступна для использования, так как соответствующие характеристики уходят за пределы области решения и не имеют точек пересечения со старым временным слоем. Вместо таких соотношений надо использовать граничные условия, которые замыкают задачу определения решения в граничных узлах нового временного слоя $t = t^{n+1}$. Описанный метод решения называется обратно-характеристическим.

Термин "прямой метод характеристик"используется для обозначения редкого частного случая обратно-характеристического метода, при котором характеристики проходят через узлы сетки и образуют сетку характеристик. Это бывает только в двумерных задачах, например, в теории жесткопластического течения или в одномерной линейной теории распространения волн.

Вдоль характеристик характеристические соотношения представляют собой обыкновенные дифференциальные уравнения по времени. Для их интегрирования в пределах шага по времени можно использовать большое разнообразие явных и неявных схем, описанных в главе по задачи Коши. Соответственно, получается большое разнообразие схем метода характеристик.

Отметим, что шаг по времени (точнее, по гиперболической координате) при использовании метода характеристик должен быть ограничен с тем, чтобы точки пересечения характеристик со старым временным слоем лежали бы внутри окрестности, определяемой шаблоном рассчитываемого узла или содержащими узел ячейками. Это условие называется условием Куранта-Фридрихса-Леви (КФЛ-условие или просто условие Куранта) и имеет вид

$$\tau_n \leq \min_i h_i/c_i$$

где h_i - максимальный радиус вписанной в окрестность узла i

окружности, c_i - скорость распространения возмущений, являющаяся максимальным из собственных чисел $\lambda^{(k)}$, определяющих характеристики для узла i. При нарушении КФЛ-условия метод характеристик становится неустойчивым и сходимость дискретного решения к решению исходной задачи отсутствует.

14.4 Соотношения на сильных разрывах

Рассмотрим систему интегральных уравнений, представляющих законы сохранения

$$\int\limits_{V_t} [\partial_t \mathbf{u} + \nabla \cdot \mathbf{F}(\mathbf{u}) + B(\mathbf{u})] dx dt = 0$$

где $\mathbf{u}$ - консервативная переменная, F - поток величины $\mathbf{u}$, B - источник величины $\mathbf{u}$, интегрирование проводится по произвольному гиперобъему $V_t = V \times [t_1, t_2]$. С использованием преобразования объемного интеграла в поверхностный получаем слабую интегральную формулировку закона сохранения

$$\int\limits_{\partial V_t} (\mathbf{u} n_t + \mathbf{F}(\mathbf{u}) \cdot \mathbf{n}) dx dt + \int\limits_{V_t} B(\mathbf{u}) dx dt = 0$$

которая в отсутствие диффузии не содержит операций дифференцирования и допускает разрывные решения. Здесь n_t представляет единичную внешнюю нормаль к гиперобъему V_t на поверхностях $t = t_1$ и $t = t_2$ и принимает значения ± 1, вектор $\mathbf{n}$ обозначает внешнюю единичную нормаль к объему V.

В областях гладкости интегральное уравнение эквивалентно дифференциальному уравнению

$$\partial_t \mathbf{u} + \nabla \cdot \mathbf{F}(\mathbf{u}) + B(\mathbf{u}) = 0$$

В случае, если область V_t содержит разрывные решения, переход от интегрального уравнения к дифференциальному в окрестности разрыва невозможен и справедливы соотношения, связывающие значения искомых функций по обе стороны поверхности разрыва, называемые соотношениями на скачке. Справедлива следующая теорема. о соотношениях на скачке: *в случае разрывных решений на заранее неизвестных поверхностях разрыва, определяемых уравнением* $\phi(x,t)=0$ *, выполняются соотношения на скачке*

$$[\mathbf{u}]\varphi_t + [\mathbf{F}(\mathbf{u})]\cdot\nabla\varphi = 0$$

или

$$[\mathbf{u}]n_t + [\mathbf{F}(\mathbf{u})]\cdot\mathbf{n} = 0$$

где $[f]=f^+-f^-$ *- скачок величины* f *при переходе через поверхность разрыва,* $n_t=\partial_t\varphi/|\nabla\varphi|$ *- скорость движения поверхности разрыва по нормали,* $\mathbf{n}=\nabla\varphi/|\nabla\varphi|$ *- пространственная единичная нормаль к поверхности разрыва.*

Доказательство. Пусть гиперобъем V_t содержит поверхность разрыва, которая делит его на две подобласти $V_t = V_t^+ \cup V_t^-$

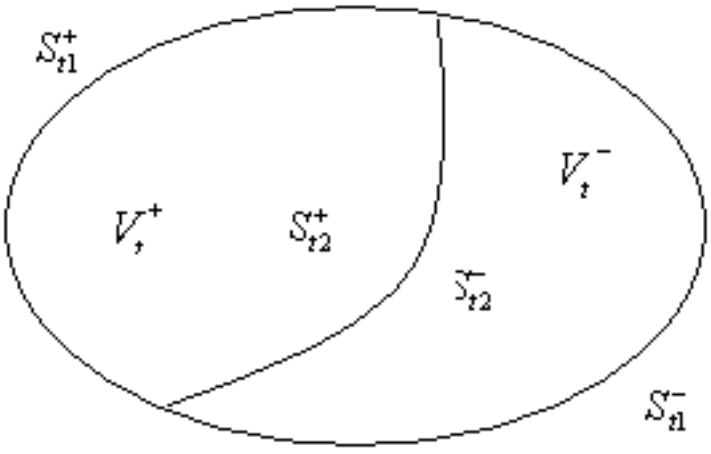

Рис. 5. Гиперобъем с поверхностью разрыва

Поверхность каждой из этих двух подобластей частично совпадает с поверхностью исходного гиперобъема $\partial V_t = S_{t1}^+ \cup S_{t1}^-$ и

содержит участки поверхности S_2^+ и S_2^- , принадлежащие поверхности разрыва: $\partial V_t^+ = S_1^+ \cup S_2^+$, $\partial V_t^- = S_1^- \cup S_2^-$. Подвергнем интегральное уравнение цепочке преобразований с учетом аддитивности операции интегрирования и теоремы Остроградского-Гаусса:

$$0 = \int\limits_{V_t} (\partial_t \mathbf{u} + \nabla \cdot \mathbf{F}(\mathbf{u}) + B(\mathbf{u}))dxdt =$$

$$= \int\limits_{V_t^1 \cup V_t^-} (\partial_t \mathbf{u} + \nabla \cdot \mathbf{F}(\mathbf{u}) + B(\mathbf{u}))dxdt =$$

$$= \int\limits_{S_{t1}^+ \cup S_{t2}^+} (\mathbf{u} n_t + \mathbf{F} \cdot \mathbf{n})dxdt + \int\limits_{\partial S_{t1}^- \cup S_{t2}^-} (\mathbf{u} n_t + \mathbf{F} \cdot \mathbf{n})dxdt +$$

$$+ \int\limits_{V_t} B(\mathbf{u})dxdt =$$

$$= \int\limits_{V_t} (\partial_t \mathbf{u} + \nabla \cdot \mathbf{F} + B(\mathbf{u}))dxdt + \int\limits_{S_2} ([\mathbf{u}] n_t + [\mathbf{F}] \cdot \mathbf{n})dxdt = 0$$

где явно выделелены составляющие поверхностных интегралов относящиеся к поверхности разрыва. В силу произвольности рассматриваемого гиперобъема отсюда следуют соотношения на сильном разрыве.

Следствие. Для частного случая систем гиперболических уравнений с постоянными коэффициентами

$$A_{(t)} \partial_t \mathbf{u} + A_{(k)} \cdot \nabla_k \mathbf{u} + B(\mathbf{u}) = 0$$

необходимым условием существования скачка является условие

$$\det(A_{(t)} \varphi_t + A_{(k)} \varphi_{x_k}) = 0$$

иначе уравнение имеет лишь тривиальное решение и разрыв отсутствует.

В общем случае положение скачков и скорости их распространения для нелинейных систем гиперболических уравнений зависят от интенсивности (величины) разрыва. Недифференциальные члены исходной системы уравнений не влияют на соотношения на сильном разрыве.

Глава 15

Сходимость приближенных решений

Важную роль в понимании поведения численных решений играет априорное (до решения) теоретическое исследование приближенных методов решения. В реальных условиях такое исследование удается проводить, как правило, на упрощенных модельных задачах, которые передают некоторые основные черты решаемой общей задачи. Теоретический анализ позволяет понять, почему работает или не работает тот или иной приближенный метод, а также использовать это понимание для повышения эффективности и конструирования приближенных методов. Не менее важен апостериорный (после решения) анализ численных решений для выяснения вопросов о точности и достоверности получаемых результатов. Есть некоторые общие подходы к решению данных вопросов, которые и рассматриваются ниже.

15.1 Теоремы о сходимости

Пусть дискретизированная задача (разностная схема) представлена системой алгебраических уравнений

$$L_h \mathbf{u}_h = \mathbf{f}_h$$

где L_h - матрица системы алгебраических уравнений, $\mathbf{u}_h$ - сеточные значения искомой функции, $\mathbf{f}_h$ - вектор правых частей. При более общем рассмотрении под $\mathbf{u}_h$ следовало бы понимать каркас приближенного решения, то есть набор дискретных параметров, которые не обязательно являются сеточными значениями искомой функции, а далее при сравнении приближенного и точного

решений надо было бы перейти от каркаса приближенного решения к самому приближенному решению с помощью оператора восполнения. Такое более общее изложение теории приближенных методов было сделано в главе 1. Здесь изложение упрощено для краткости в расчете на то, что читатель, желающий строгости, легко сопоставит это изложение с материалом главы 1 и внесет необходимые поправки в изложение самостоятельно.

Аппроксимация Говорят, что разностная схема аппроксимирует дифференциальное уравнение

$$L\mathbf{u} = \mathbf{f}$$

с порядком аппроксимации $n > 0$, если на решении исходной задачи $\mathbf{u}$ выполнено условие

$$||L_h P_h \mathbf{u} - P_h \mathbf{f}|| = O(h^n)$$

где P_h - сеточный оператор проектирования.

Устойчивость Под устойчивостью разностной схемы понимается ограниченность обратного оператора дискретизированной задачи $||L_h|| \leq O(h^{-k})$.

Теорема Лакса о сходимости: *Решение разностного уравнения сходится к сеточной проекции решения дифференциального уравнения*

$$||P_h \mathbf{u} - \mathbf{u}_h|| = O(h^m)$$

если разностное уравнение аппроксимирует дифференциальное

$$||L_h P_h \mathbf{u} - P_h \mathbf{f}|| = O(h^n)$$

разностный оператор имеет ограниченный обратный

$$||L_h^{-1}|| = O(h^{-k})$$

и m=n-k>0.

Доказательство. Учитывая, что

$$P_h \mathbf{f} = \mathbf{f}_h = L_h \mathbf{u}_h$$

получаем

$$\varepsilon = ||P_h\mathbf{u} - \mathbf{u}_h|| = ||L + h^{-1}L_h(P_h\mathbf{u} - \mathbf{u}_h)|| \leq$$

$$\leq ||L_h^{-1}||||L_hP_h\mathbf{u} - P_h\mathbf{f}|| = O(h^{m-k})$$

При $m - k > 0$ и $h \to 0$ ошибка $\varepsilon \to 0$.

Теорема о сходимости метода конечных элементов формулируется так (см. Стренг и Фикс, 1978): *Приближенное решение метода конечных элементов сходится к решению исходной вариационной задачи, если система пробных функций полна в пространстве решений, аппроксимация вариационного уравнения согласованна (т.е. интегрирование обеспечивает точное вычисление объемов, площадей и производных от базисных функций, входящих в вариационное уравнение) и матрица разрешающей системы алгебраических уравнений имеет ограниченную обратную.*

Теорема Лакса-Вендроффа о сходимости разрывных решений: *консервативность является достаточным условием сходимости устойчивой аппроксимирующей конечно-разностной схемы к слабому решению нелинейной системы уравнений.*

Разностная схема, аппроксимирующая закон сохранения, обладает свойством консервативности, если она поддерживает этот закон на дискретном уровне для каждого малого дискретного объема сетки.

Можно показать (см. обсуждение в статье Азаренка, 2006), что для консервативных устойчивых аппроксимирующих разностных схем приближенные решения вблизи скачков удовлетворяют соотношениям на скачках. Неконсервативные схемы таким свойством не обладают, поэтому приводят к искаженным картинам расположения разрывов и к неверным значениям разрывов решения. Для гладких решений с устойчивыми аппроксимирующими

неконсервативными схемами все в порядке. Таким образом, теорема Лакса-Вендроффа обосновывает корректность применения схем сквозного счета разрывных решений.

Подробнее о теореме Лакса-Вендроффа можно прочитать в оригинальной статье авторов (Lax, Wendroff (1960)).

15.2 Априорное исследование устойчивости

Ниже на примере ВВЦП-схемы (Вперед по Времени, Центральные разности по Пространству) для основного модельного уравнения

$$\frac{u_i^{n+1}-u_i^n}{\tau_n}+U\frac{u_{i+1}^n-u_{i-1}^n}{2h}=\nu\frac{u_{i+1}^n-2u_i^n+u_{i-1}^n}{h^2}+C^n$$

рассмотрим основные приемы исследования устойчивости разностных схем.

15.2.1 Метод дискретных возмущений

При исследовании устойчивости схем для линейных уравнений в частных производных в произвольном узле сетки в некоторый момент времени вводится малое возмущение решения и прослеживается его влияние на решение во времени. Если возмущение растет, то схема неустойчива, если оно остается ограниченным, то устойчива. Если возмущение дополнительно сохраняет свой знак от шага к шагу, то схема называется монотонной. Рассмотрим пример

$$\frac{u_i^{n+1}-u_i^n}{\tau_n}=\nu\frac{u_{i+1}^n-2u_i^n+u_{i-1}^n}{h^2}$$

Если ввести начальное возмущение δu^n в точке $x=x_i$ при $t=t_n$, то разностное соотношение для возмущения δu^{n+1} в той же пространственной точке на новом временном слое при $t=t_{n+1}$

определяется из уравнения

$$\frac{(u_i^{n+1} + \delta u_i^{n+1}) - (u_i^n + \delta u_i^n)}{\tau_n} =$$

$$= \nu \frac{(u_{i+1}^n + \delta u_{i+1}^n) - 2(u_i^n + \delta u_i^n) + (u_{i-1}^n + \delta u_{i-1}^n)}{h^2}$$

Вычитая первое уравнение из второго, получаем

$$\frac{\delta u_i^{n+1} - \delta u_i^n}{\tau_n} = \nu \frac{-2\delta u_i^n}{h^2}$$

или

$$\frac{\delta u_i^{n+1}}{\delta u_i^n} = 1 - \frac{2\nu\tau_n}{h^2}$$

Для устойчивости необходимо

$$\left| \frac{\delta u_i^{n+1}}{\delta u_i^n} \right| \leq 1$$

Для монотонности необходимо

$$\frac{\delta u_i^{n+1}}{\delta u_i^n} > 0$$

Подставляя в эти условия соотношение для возмущений получаем окончательные условия устойчивости и монотонности

$$0 \leq 1 - \frac{2\nu\tau_n}{h^2} \leq 1$$

или

$$\tau_n \leq \frac{h^2}{2\nu}$$

15.2.2 Метод гармонических возмущений

Метод гармонических возмущений Фурье-Неймана применяется при исследовании разностных схем для линейных уравнений в частных производных. Для этого 1) произвольное возмущение подставляется в каждую точку сетки в некоторый момент

времени 2) возмущение раскладывается в ряд Фурье 3) отдельно прослеживается эволюция каждой гармоники Фурье. Если какая-либо гармоника растет, то схема неустойчива; если ни одна гармоника не растет, то схема устойчива. Пример. Рассмотрим явную схему для уравнения теплопроводности

$$\frac{u_i^{n+1} - u_i^n}{\tau_n} = \nu \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{h^2}$$

где $i = 1, 2, ..., N+1$. Вводим возмущения

$$\frac{(u+\delta u)_i^{n+1} - (u+\delta u)_i^n}{\tau_n} =$$

$$= \nu \frac{(u+\delta u)_{i+1}^n - 2(u+\delta u)_i^n + (u+\delta u)_{i-1}^n}{h^2}$$

вычитание уравнений дает уравнение для возмущений

$$\frac{(\delta u)_i^{n+1} - (\delta u)_i^n}{\tau_n} =$$

$$= \nu \frac{(\delta u)_{i+1}^n - 2(\delta u)_i^n + (\delta u)_{i-1}^n}{h^2}$$

Разложение возмущения в ряд Фурье имеет вид

$$(\delta u)_i^n = \sum_{j=-m}^{j=m} \delta u_j^n exp(Ik_j x_i) \ , \quad m = N/2, \quad k_j = \frac{\pi j}{(mh)}, \quad I = \sqrt{-1}$$

где волновому числу k_j соответствует длина волны

$$\lambda_j = \frac{2\pi}{k_j}$$

Подставляя это разложение в уравнение для возмущений и учитывая, что

1) $e^{Ik_j x_{i\pm1}} = e^{Ik_j x_i \pm h} = e^{Ik_j e^{x_i \pm h}}$;

2) знак суммы можно опустить, так как уравнение линейно и отдельные гармоники между собой не взаимодействуют,

3) $e^{I\phi} = cos\phi + Isin\phi$, где $I = \sqrt{-1}$, получаем

$$\delta u_i^{n+1} = \delta u_i^n \left(1 - 4\frac{\nu\tau_n}{h^2} sin^2 \frac{k_j h}{2}\right)$$

Условие устойчивости имеет вид

$$\left|\frac{\delta u_i^{n+1}}{\delta u_i^n}\right| \leq 1 \quad \Rightarrow -1 \leq 1 - 4\frac{\nu\tau_n}{h^2} sin^2 \frac{k_j h}{2} \leq 1$$

для всех j, или

$$0 \leq \frac{2\nu\tau_n}{h^2} \leq 1$$

15.2.3 Спектральный метод

Разностные схемы для линейных уравнений в частных производных исследуются также матричным или, в другой терминологии, спектральным методом. В отличие от метода гармонических возмущений, этот метод позволяет учитывать влияние граничных условий, которые также предполагаются линейными. В соответствии с матричным методом

1) Разностная схема записывается в матричной форме

$$\mathbf{u}^{n+1} = L\mathbf{u}^n + \mathbf{b}$$

где L - матрица перехода (с n-го временного слоя на $(n+1)$-й), $\mathbf{u}^n$ - вектор сеточных значений искомой функции на n-м временном слое, $\mathbf{b}$ - вектор правых частей. По рекурсии получаем другую запись разностной схемы:

$$\mathbf{u}^{(n+1)} = L^{n+1}\mathbf{u}^{(0)} + (L^n + L^{n-1} + ... + I)\mathbf{b}$$

где L^n обозначает n-ю степень матрицы L, I - единичная матрица.

2) В начальный момент времени в каждую точку сетки вводится возмущение, связь возмущенных решений на n-м и (n+1)-м временных слоях устанавливается в соответствии с рассматриваемой разностной схемой

$$(\mathbf{u}+\delta\mathbf{u})^{(n+1)} = L^{n+1}(\mathbf{u}+\delta\mathbf{u})^{(0)} + (L^n + L^{n-1} + ... + I)\mathbf{b}$$

после вычитания невозмущенного уравнения для возмущения получается уравнение

$$(\delta\mathbf{u})^{(n+1)} = L^{n+1}(\delta\mathbf{u})^{(0)}$$

Если матрица L сжимающая ($||L|| < 1$), то схема устойчива. Условие устойчивости может быть ослаблено

$$||L|| < 1 + O(\tau)$$

где τ - шаг по времени.

15.2.4 Метод дифференциальных приближений

В соответствии с методом дифференциальных приближений (Хирт, 1968; Шокин, 1974) значения сеточной функции в разностной схеме заменяются на значения соответствующей непрерывной функции в узлах сетки, а затем с помощью разложений в ряд Тейлора в окрестности рассматриваемого узла сетки восстанавливается дифференциальное уравнение. Восстановленное дифференциальное уравнение не совпадает с исходным уравнением, для которого была записана разностная схема, и содержит дополнительные члены ошибок аппроксимации. По свойствам восстановленного уравнения, называемого дифференциальным приближением разностной схемы, судят о свойствах разностной схемы.

Например, в случае центрально-разностной схемы

$$\frac{u_i^{n+1} - u_i^n}{\tau_n} + U\frac{u_{i+1}^n - u_{i-1}^n}{2h} = \nu\frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{h^2} + C^n$$

делается следующач замена

$$u_{i\pm1}^{n\pm1} = u(x_{i\pm1}, t^{n\pm1})$$

где

$$u(x_{i\pm1}, t^{n\pm1}) = u|_{x=x_n}^{t=t^n} \pm \left.\frac{\partial u}{\partial t}\right|_{x=x_i}^{t=t^n} \Delta t_n \pm \left.\frac{\partial u}{\partial x}\right|_{x=x_i}^{t=t^n} h +$$

$$+ \left.\frac{\partial^2 u}{\partial t^2}\right|_{x=x_i}^{t=t^n} \frac{\Delta t_n^2}{2} \pm \left.\frac{\partial^2 u}{\partial x^2}\right|_{x=x_i}^{t=t^n} \frac{h^2}{2} + O(h^3, \Delta t_n^3)$$

В результате подстановки данного разложения в разностное уравнение получается гиперболическая форма (Г-форма) восстановленного дифференциального уравнения задачи. В нашем примере Г-форма уравнения имеет вид

$$\frac{\partial u}{\partial t} + \frac{\partial^2 u}{\partial t^2}\frac{\Delta t}{2} + U\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2} + O(h^2, \Delta t^2)$$

Затем с помощью исходного дифференциального уравнения исключаются все временные производные более высокого порядка, нежели присутствующие в исходном уравнении. В нашем случае вторая производная по времени выражается через пространственные производные так

$$\frac{\partial^2 u}{\partial t^2} = -U\frac{\partial}{\partial x}\left(-U\frac{\partial u}{\partial x} + \nu\frac{\partial^2 u}{\partial x^2}\right) + \nu\frac{\partial^2}{\partial x^2}\left(-U\frac{\partial u}{\partial x} + \nu\frac{\partial^2 u}{\partial x^2}\right)$$

или

$$\frac{\partial^2 u}{\partial t^2} = U^2\frac{\partial^2 u}{\partial x^2} - 2U\nu\frac{\partial^3 u}{\partial x^3} + \nu^2\frac{\partial^4 u}{\partial x^4}$$

В результате замены получается параболическая форма (П-форма) восстановленного дифференциального уравнения задачи

$$\frac{\partial u}{\partial t} + U\frac{\partial u}{\partial x} = \left(\nu - \frac{U^2\Delta t}{2}\right)\frac{\partial^2 u}{\partial x^2} +$$

$$+U\nu\frac{\partial^3 u}{\partial x^3}\Delta t - \nu^2\frac{\partial^4 u}{\partial x^4}\frac{\Delta t}{2} + O(h^2, \Delta t^2)$$

Восстановленные уравнения задачи в Г- и П- форме называются дифференциальными приближениями разностной схемы. Разностная схема наследует свойства ее дифференциального приближения, которое и подлежат анализу.

В общем случае П-форма дифференциального приближения имеет вид

$$\frac{\partial u}{\partial t} = \sum_{p=0}^{\infty} \mu_{2p+1} \frac{\partial^{2p+1} u}{\partial x^{2p+1}} + \sum_{p=0}^{\infty} \mu_{2p} \frac{\partial^{2p} u}{\partial x^{2p}}$$

Решение восстановленного (возмущенного заменой производных разностями) уравнения ищется в виде некоторой Фурье компоненты

$$u(x,t) = exp(at)exp(Ikx)$$

Подставляя это решение в П-форму и приравнивая мнимые и вещественные части, получим

$$Re(a) = f(\mu_{2p}) = \sum_{p=1} \infty (-1)^p k^{2p} \mu_{2p}$$

$$Im(a) = f(\mu_{2p+1})$$

Члены с четными пространственными производными представляют диффузию $(2p = 2)$ и диссипацию. Члены с нечетными производными отвечают за конвекцию $(2p + 1 = 1)$ и дисперсию $(2p + 1) = 3$. Так как рост решения зависит только от $Re(a)$, эта величина должна быть неположительной. Значит П-форма устойчива, если

$$Re(a) = \sum_{p=1} \infty (-1)^p k^{2p} \mu_{2p} < 0$$

для любого k. Упрощенный критерий устойчивости требует выполнения данного неравенства только для первого ненулевого члена низшего четного порядка.

Поясним, что диффузия сглаживает искомые функции. В реальных средах диффузия обусловлена хаотическим движением

атомов и молекул в жидкостях и газах и межатомными взаимодействиями в (структурированных) твердых деформируемых средах. Диссипацией называется необратимый процесс рассеяния энергии. Конвекцией (или адвекцией) называется перенос характеристик сплошной среды с упорядоченным потоком массы. Дисперсией называется зависимость скорости распространения отдельной фазы гармонической волны (то есть, синусоидальной волны) от ее частоты.

В нашем примере для устойчивости центрально-разностной схемы необходимо, чтобы эффективный коэффициент диффузии был неотрицателен, то есть

$$\tau_n \leq \frac{2\nu}{U^2}$$

Критерии устойчивости схемы, определяемые по методу дифференциальных приближений представляют необходимые, но недостаточные условия.

15.2.5 "Замораживание" коэффициентов

В случае нелинейных уравнений и уравнений с переменными коэффициентами изложенные выше методы исследования устойчивости применяются не к исходным, а к соответствующим приближенным разностным уравнениям с постоянными коэффициентами. Такие приближенные уравнения имеют <замороженные> постоянные значения коэффициентов, которые определяются по их значениям в окрестности данной точки для локальных методов исследования или по некоторым наиболее неудачным для численного решения значениям этих коэффициентов в области решения. Обычно эти значения равны максимальным значениям коэффициентов и поэтому называются <мажорирующими>.

15.2.6 Использование расщепления

В практических задачах исследование устойчивости для простоты часто проводится отдельно для различных подпроцессов задачи. Например, можно раздельно анализировать устойчивость аппроксимации уравнений движения, теплопроводности, переноса примеси и тому подобных. .

15.2.7 Влияние свободных членов

В задачах с большими по норме свободными (недифференциальными) членами шаг по времени в явных схемах приходится дополнительно ограничивать. Такие свободные члены появляются в законах сохранения в виде источников импульса, энергии, примеси и так далее, а также, например, играют роль вязких членов в определяющих соотношениях теорий упруговязкопластичности и ползучести. Распространенным приемом, позволяющим избавиться от таких обременительных ограничений, является неявная аппроксимация свободных членов. Поскольку свободные члены не содержат пространственного дифференцирования, то такие неявные аппроксимации не нарушают порядок вычислений по явным схемам, так как каждое дискретизированное уравнение может быть отдельно от остальных явно разрешено относительно значения искомой функции на новом временном слое. Если свободные члены нелинейны, то они при этом квазилинеаризуются относительно приращений искомого решения на шаге по времени.

15.2.8 Коэффициент запаса

Поскольку в практических задачах исследование устойчивочти всегда проводится приближенно, то приближенность априорного анализа компенсируется дополнительным уменьшением

шага по времени путем умножения его приближенного теоретического значения на положительный и меньший единицы коэффициент запаса, значение которого подбирается эмпирически и лежит обычно в пределах 0.1-0.9. Этот коэффициент является характеристикой приближенности анализа устойчивости: чем точнее априорный анализ устойчивости, предсказывающий величины допустимого шага по времени, тем ближе коэффициент запаса к единице. Если схема плохая, то введение коэффициента запаса бесполезно.

15.2.9 Условие точности

Независимо от типа задачи и помимо условий устойчивости шаг по времени ограничивается еще и условиями точности. Условия точности заключаются в требовании малости изменения нормы решения на шаге по времени по сравнению с нормой самого решения. Часто условие точности можно увязать с априорными теоретическими оценками скорости изменения термомеханических параметров состояния моделируемых процессов. Например, в задачах механики деформируемого твердого тела роль условия точности испольняет требование малости приращений деформации на шаге по времени. При наличии источников большой интенсивности шаг по времени из соображений точности может быть во много раз меньше шага, обеспечиваюшего устойчивость, даже при использовании "безусловно устойчивых"неявных аппроксимаций.

15.2.10 Оценка шага по пространству

Рассмотрим вопрос об определении шагов неравномерной и нерегулярной пространственной сетки для использования в условиях устойчивости. Этих шаги можно определить с помощью формул дифференцирования. Пусть, например, дискретизированная формула дифференцирования по координате x для узла

k или для ячейки k имеет вид

$$\left[\frac{\partial f}{\partial x}\right]_k = \sum_{j\in\omega_k} d_{xj} f_j$$

где суммирование производится по узлам $j \in \omega_k$ шаблона для узла k или по узлам $j \in \omega_k$ ячейки k, тогда роль пространственного шага сетки по координате x в условиях устойчивости может с успехом играть величина

$$h_{x(k)} = \left(\sum_{j\in\omega_k} \max(d_{xj}, 0)\right)^{-1}$$

Расположение ребер сетки в пространстве для данной оценки пространственного шага роли не играет. Оценка записана без вывода, интуитивно.

15.3 Апостериорное исследование численного решения

В этом разделе опишем основные приемы, которые применяются при проведении расчетов для оценки точности и достоверности численных решений.

15.3.1 Обезразмеривание переменных и уравнений

Для любой вычислительной машины имеется самое маленькое, отличное от нуля, число и самое большое число, с которым может оперировать данная машина. Например, для четырехбайтовых ЭВМ, к которым относятся обычные персональные компьютеры, минимальное и максимальное по модулю значения вещественных чисел равны: $\pm 8.43_{10} - 37$ и $\pm 3.37_{10}38$, а диапазон представления целых чисел определяется значениями ± 2147483647, соответственно. Необходимо, чтобы диапазон изменения решения и входной информации находился бы в области представимых на

ЭВМ значений. Отметим, что при выполнении арифметических действий ограничения, накладываемые на величины операндов ограниченной разрядностью представления чисел, являются более жесткими. Так, "машинное эпсилон", то есть минимальное положительное вещественное число, добавление которого к единице приводит к результату, отличному от единицы, для четырехбайтовых ЭВМ равно примерно $1_{10} - 6$.

"Машинное эпсилон"играет важную роль при реализации неравенств в программах для ЭВМ. Если суммируются числа, величины которых отличаются более, чем на шесть порядков, то точность будет потеряна. То есть, например, добавление к единице бесконечного числа слагаемых, меньших "машинного эпсилон", будет иметь своим результатом единицу! Желательно поэтому, чтобы функции, описывающие решение, не слишком бы отличались от единицы. Это достигается масштабированием (обезразмериванием) искомых функций.

Из теории размерности и подобия (см. Седов, 1962) известно, что числовые значения искомых переменных и коэффициентов уравнений зависят от выбора масштабов (размерностей или характерных значений). Неудачный выбор размерностей из-за ограниченного числа разрядов для представления чисел в вычислительных машинах может приводить к потере точности при выполнении арифметических операций с очень большими и очень маленькими числами, Поэтому важно хорошо отмасштабировать искомые переменные, то есть перейти от размерных к безразмерным переменным с разумным выбором масштабов размерных переменных. Выбор масштабов или, другими словами, выбор характерных значений физических величин производится так, чтобы безразмерные переменные не слишком отличались от единицы.

Разумно отмасштабированные переменные приобретают ясный физический смысл. Например, сообщение о том, что рассматривается соударение тел со скоростью 100 метров в секунду практически недостаточно для того, чтобы судить об интенсивности

удара. Напротив, сообщение о том, что скорость соударения тел составляет одну десятую от скорости звука, говорит о том, что соударение является высокоскоростным и будет сопровождаться заметными деформациями соударяющихся тел. Аналогично, в гидродинамике для определения режима течения важно знать безразмерное значение скорости набегающего потока по отношению к скорости звука, простое же сообщение размерной величины скорости потока в метрах в секунду является бесполезным.

Участвующие в безразмерных уравнениях безразмерные коэффициенты, составленные из размерных масштабов, называются параметрами подобия. Количество независимых параметров подобия определяется π-теоремой о параметрах подобия: **π-теорема**: *Пусть среди размерных масштабов величин* $\{a_i\}_{i=1}^n$, *характеризующих некоторый процесс, первые* k *имеют независимые размерности. Тогда с помощью этих* k *независимых размерных масштабов из остальных масштабов можно образовать систему* n-k *безразмерных параметров подобия*

$$\Pi_i = \frac{a_{k+i}}{a_1^{q_1} a_2^{q_2} \cdots a_k^{q_k}}$$

где $i=1,\ldots,n$-k.

Значения безразмерных параметров подобия ответственны за математические свойства уравнений. Знание ожидаемого диапазона изменения параметров подобия важно как на стадии конструирования численного метода, так и в процессе нахождения численного решения. Например, сказать, что на графике показано решение задачи в момент времени, равный 15 секундам с начала процесса, это все равно, что не сказать ничего. Напротив, если сказано, что график отвечает безразмерному моменту времени, равному 0.5, где масштабом времени является время распространения возмущения по области решения, то такое высказывание уже вполне конкретно и полезно. Из него следует, что фронт волны возмущения должен находиться на расстоянии

половины максимального размера области решения от точки первоначального возмущения.

Явления, характеризуемые одними и теми же значениями параметров подобия являются подобными. Это означает, что в безразмерной форме подобные явления описываются одними и теми же значениями безразмерных физических переменных.

Рассмотрим пример обезразмеривания. В случае начально-краевой задачи для модельного уравнения конвекции-диффузии

$$\frac{\partial u}{\partial t} + U \cdot \nabla u = \nabla \cdot (\nu \nabla u) + C$$

безразмерные переменные можно ввести так:

$$\tilde{x} = x/x_* \ , \quad \tilde{t} = tu_*/x_* \quad \tilde{U} = U/u_* \ , \quad \tilde{\nabla} = \nabla x_* \ , \quad \tilde{u} = u/u_*$$

где звездочки отмечают размерные константы, используемые в качестве масштабов переменных задачи, а тильды отмечают вводимые безразмерные переменные. Подставляя вместо размерных переменных их выражения, после несложных преобразований получаем запись уравнения в безразмерном виде:

$$\frac{\partial \tilde{u}}{\partial \tilde{t}} + \tilde{U} \cdot \tilde{\nabla} \tilde{u} = \tilde{\nabla} \cdot \left(\frac{1}{\mathrm{Re}} \tilde{\nabla} \tilde{u} \right) + \tilde{C}$$

где $\mathrm{Re} = \frac{U_* x_*}{\nu}$ - параметр подобия, называемый числом Рейнольдса, а $\tilde{C} = C \frac{x_*}{U_* u_*}$ - безразмерный источниковый член. Начальные и граничные условия аналогично преобразуются к безразмерному виду. Значки "тильда"над безразмерными переменными в дальнейшем, как правило, опускаются.

В безразмерных переменных уравнения сохраняют свою форму. Поэтому при написании алгоритмов и программ можно использовать исходную размерную форму уравнений, а безразмерные переменные использовать при проведении расчетов путем задания входных данных для коэффициентов уравнений и краевых

условий в соответствии с принятым вариантом обезразмеривания переменных.

При написании научных отчетов и статей по результатам исследований хорошей практикой является представление числовых данных в безразмерной форме. Если способ обезразмеривания указан, то восстановление размерных значений величин для использования в технических приложениях не составляет большого труда.

15.3.2 Искусственные аналитические решения

Имеется следующий простой способ получения аналитических решений для дальнейшего их использования при тестировании. Для того, чтобы произвольная достаточно гладкая функция $\tilde{u}(x,t)$ являлась аналитическим решением краевой задачи, то есть, например, удовлетворяла бы модельному уравнению

$$\frac{\partial u}{\partial t} + U \cdot \nabla u = \nabla \cdot (\nu \nabla u) + C$$

а также начальным

$$t = 0 \ , \quad : u = u^0(x)$$

и граничным условиям

$$x \in S_u : u(x,t) = u^*(x,t)$$

$$x \in S \backslash S_u : n \cdot \nabla u(x,t) = B_n^*(x,t)$$

достаточно задать функции правых частей уравнения, начальных и граничных условий в виде:

$$C(x,t) = \frac{\partial \tilde{u}}{\partial t} + U \cdot \nabla \tilde{u} - \nabla \cdot (\nu \nabla \tilde{u})$$

$$x \in S_u : u^*(x,t) = \tilde{u}$$

$$x \in S \backslash S_u : B_n^*(x,t) = n \cdot \nabla \tilde{u}$$

15.3.3 Тестирование численных алгоритмов

Для того, чтобы убедиться в достоверности численных решений и оценить их точность, необходимо протестировать численные алгоритмы, а именно сравнить численные решения с известными аналитическими или численными решениями, точность которых уже установлена. При отсутствии аналитических решений, имеющих физическое содержание, всегда можно получить искусственные аналитические решения и использовать их для сравнений и оценки погрешности численных решений.

При проверках используется последовательность тестовых задач, расположенных по порядку от более простых к более сложным. Первые проверки проводятся, как правило, при минимально возможном числе расчетных дискретных параметров. При проведении испытаний сложного алгоритма имеет смысл отключать отдельные рассчитываемые процессы и соответствующие части алгоритма, тестировать поочередно отдельные процессы.

Например, при отладке алгоритмов решения многомерных упругопластических задач можно вначале проверить расчет простых упругих одномерных задач отдельно по каждому из трех пространственных направлений, затем проверить работу части алгоритма, ответственной за расчет пластических деформаций, тоже сначала для одномерных процессов. Отдельно тестируются случаи динамических процессов распространения волн и случаи квазистатического нагружения. Далее тестируется случай различного рода симметрии, плоской, аксиальной, сферической. Затем переходят к тестированию двумерных процессов и только после успешного завершения этих испытаний переходят к тестам трехмерных процессов.

Как правило, для каждого класса задач подбирается свое множество характерных тестовых задач, которые приходится постоянно решать повторно после очередных изменений в алгоритме решения. Изменения в алгоритмы нередко приходится вносить

в связи с расширением класса решаемых задач, при совершенствовании численного алгоритма или при проведении испытаний какого-либо нового приема или метода.

Хорошей практикой является хранение предшествующих версий разработываемого алгоритма и программы для ЭВМ, в серьезных разработках число таких версий может исчисляться сотнями. Наличие архива версий позволяет быстро находить и устранять ошибки, вносимые при модификациях метода, а также сравнивать эффективность различных вариантов реализации численных решений.

Для оценки погрешности решения новой сложной задачи решение вычисляют на последовательности вложенных сеток, получаемых дроблением ребер сетки пополам. Модуль разности решений в общих узлах последовательных вложенных сеток дает функцию распределения погрешности, которую можно использовать в алгоритмах адаптации сеток к решению, хотя ее величина не совпадает с реальной погрешностью решения, а лишь воспроизводит ее изменение по области решения. На основе правдоподобных предположений об асимптотической скорости стремления погрешности к нулю ($|\Delta u| = c_u h^k$) с использованием экстраполяции Ричардсона получают значения уточненного решения, комбинируя решения в общих узлах вложенных сеток. Например, пусть $u_i^{(1)}$ и $u_i^{(2)}$ - решения, полученные методом первого порядка точности на вложенных сетках с шагами h и $h/2$, соответственно. Для метода первого порядка точности эти решения отличаются от неизвестного точного решения $u^{(\infty)}$ на величину погрешности, определяемую неизвестным параметром $c(x)$:

$$u_i^{(1)} = u_i^{(\infty)} + c(x)h$$

$$u_i^{(2)} = u_i^{(\infty)} + c(x)h/2$$

В соответствии с методом Ричардсона умножаем второе уравнение на 2 и вычитаем из него первое уравнение, получаем уточ-

ненное решение

$$u_i^{(\infty)} = 2u_i^{(2)} - u_i^{(1)}$$

При этом разность решений на вложенных сетках дает оценку погрешности

$$|c(x)h/2| = |u_i^{(1)} - u_i^{(2)}|$$

Важную роль в обосновании достоверности решений и оценке их качества играют диагностические функционалы и различного рода индикаторы состояния моделируемых процессов. Например, это могут быть числа Маха, определяющие дозвуковой и сверхзвуковой режимы течения, дивергенция скорости, определяющая зоны разрежения и сжатия в сжимаемой среде, а также определяющая погрешность условия несжимаемости в несжимаемой среде. Диагностические функционалы, определяющие количество массы, импульса и энергии в области решения могут использоваться для контроля выполнения законов сохранения численным алгоритмом. Наблюдение за распределением значений коэффициентов искусственной вязкости может быть полезным для оптимизации диссипативных свойств используемых методов. Полезно реализовать наблюдение за диагностическими функционалами, характеризующими выполнение условий пластичности, разгрузки и активного нагружения, а также условий разрушения.

Подчеркнем особо, что сравнение численных решений с результатами физических экспериментов нельзя использовать в качестве обоснования достоверности численных решений и для оценки их точности. Такие сравнения приобретают смысл при условии, что точность численных решений уже установлена чисто математическими средствами. Тогда сравнение с физическими экспериментами дают информацию о пригодности применяемой математической теории физического процесса.

Приложения

П1. Сведения из алгебры и функционального анализа

Ниже напоминаются некоторые основные понятия линейной алгебры и функционального анализа, использованные в изложении. Для более подробного чтения рекомендуются книги Колмогорова и Фомина (1972) и Михлина (1950) [30, ?].

Линейное множество (линейное пространство, пространство) это множество элементов, сложение которых и умножение на число дает элемент того же множества. Педантичное (строгое) определение формулируется в виде восьми аксиом линейного пространства и приводится в курсах линейной алгебры. Эти аксиомы выражают коммутативность и ассоциативность сложения элементов, существование нулевого и противоположного элементов, дистрибутивность умножения элементов на числа. В контексте настоящей книги вполне достаточно краткого интуитивного определения, приведенного в первой фразе данного абзаца.

Примеры: 1) множество векторов определенной размерности. 2) множество функций, имеющих общую область определения и непрерывные производные до какого-то определенного порядка.

Метрическое пространство это линейное множество, в котором для любых двух элементов x и y определена вещественная функция расстояния $\rho(x, y)$ (метрика) такая, что

1) если $\rho(x, y) = 0$, то $x = y$,

2) $\rho(x, y) = \rho(y, x)$,

и для произвольного элемента z этого же пространства выполнено неравенство треугольника

3) $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$

Норма вводится для измерения величин элементов линейного множества. Норма превращает линейное множество в норми-

рованное пространство. Свойства нормы имеют вид

$$||f|| \geq 0$$

$$||\lambda f|| = \lambda ||f||$$

$$||f + g|| \leq ||f|| + ||g||$$

где последнее неравенство называется неравенством треугольника. Из равенства нормы элемента нулю ($||f|| = 0$) следует равенство элемента нулю ($f = 0$), в противном случае величина $||f||$ называется полунормой. Норма используется в качестве функции расстояния $\rho(x, y) = ||x - y||$.

Примеры. В функциональном пространстве непрерывных функций для любого элемента норма определяется так

$$\forall f \in C : \quad ||f|| = \max_x |f(x)|$$

где значок $\forall$ читается "для всех" или "для любого", значок $\in$ читается "принадлежащий", значок : можно прочитать "выполнено".

В функциональном пространстве L_2 "интегрируемых с квадратом" функций норма вводится так

$$\forall f \in L_2 : \quad ||f|| = \left(\int_\Omega f^2 dx \right)^{1/2}$$

где Ω - область определения функций пространства L_2.

Примером нормы векторного пространства может служить максимум модуля компонент вектора.

Примером полунормы векторного пространства может служить максимум модуля части компонент вектора.

Скалярное произведение является операцией, которая любым двум элементам пространства a и b ставит в соответствие число (a,b), обладает следующими свойствами

$$(f, g) = (g, f)$$

$$\alpha(f,g) = (\alpha f, g)$$

$$(f, g_1 + g_2) = (f, g_1) + (f, g_2)$$

$$\forall f \neq 0: \quad (f,f) > 0$$

$$(f,f) = 0: \quad f = 0$$

и удовлетворяет неравенству Коши-Буняковского

$$|(a,b)| \leq ||a||||b||$$

Гильбертовы пространства. Нормированное пространство со скалярным произведением называют гильбертовым, если нормой служит корень квадратный от скалярного произведения элемента на себя. Конечномерное гильбертово пространство называют евклидовым.

Элементы гильбертова пространства ортогональны, если их скалярное произведение равно нулю.

Примеры. Если, например, скалярное произведение определить так

$$(f_1, f_2) = \int_\Omega f_1 f_2 dx$$

то имеем гильбертово функциональное пространство интегрируемых с квадратом функций H_2, в котором норма определена так

$$||f|| = (f,f)^{1/2}$$

Другим примером может служить N-мерное пространство векторов со скалярным произведением

$$(a,b) = \sum_{i=1}^{N} a_i b_i$$

и нормой

$$||a|| = (a,a)^{1/2}$$

называемое гильбертовым или евклидовым пространством N-мерных векторов.

Линейная зависимость. Говорят, что элемент $f_{(N)}$ линейно зависит от элементов u_i или является их линейной комбинацией (суперпозицией), если найдутся вещественные числа α_i такие, что

$$f_{(N)} = \sum_{i=1}^{N} \alpha_i u_i$$

Базис Набор линейно-независимых элементов u_i называется базисом линейного пространства, если он является полным, то есть, для произвольного элемента f найдется целое положительное число N и набор чисел α_i, при которых справедливо неравенство

$$||f - f_{(N)}|| < \epsilon$$

для любого наперед заданного малого числа $\epsilon > 0$. При этом числа α_i называются коэффициентами Фурье.

Важными являются следующие определения и утверждения:

- а) Базис это полный набор линейно-независимых элементов пространства (базисных элементов).
- б) Число базисных элементов определяет размерность пространства.
- в) Любой элемент пространства есть линейная комбинация базисных элементов. Коэффициенты Фурье такой линейной комбинации называются коэффициентами разложения этого элемента по базисным элементам.
- г) В нормированном базисе норма каждого из базисных элементов равна единице.
- д) В ортогональном базисе скалярные произведения базисных векторов между собой равны нулю.

Оператор. Оператор преобразует один элемент пространства в элемент того же или другого пространства. Примерами операторов являются операции дифференцирования $f(x) \to df/dx$ и умножения на число $f \to \alpha f$.

Положительно определенный оператор A для любого элемента $x \neq 0$ удовлетворяет неравенству

$$(Ax, x) > 0$$

Самосопряженный оператор A удовлетворяет равенству

$$(Ax, y) = (x, Ay)$$

для любых элементов x и y.

Функционал. Функционалом называют оператор, который преобразует элемент(ы) пространства в число. Примерами могут служить норма элемента $f \to ||f||$, скалярное произведение элемента на некоторый фиксированный элемент $f \to (f, f_0)$.

Предел f последовательности элементов f_i $(i = 1, 2, ...)$ существует тогда и только тогда, когда для любого $\epsilon > 0$ найдется такой номер i_*, что для всех больших номеров $i > i_*$ будет выполнено неравенство $|f_i - f| < \epsilon$.

Полнота пространства означает, что пределы последовательностей элементов этого пространства принадлежат этому же пространству.

Подпространство. Подпространство H' пространства H определяется отношенем $H' \subset H$, означающим, что $f \in H' \to f \in H$. Значок $\to$ читается "влечет".

Оболочкой называют конечномерное подпространство $H^{(N)}$, образованное линейными комбинациями N линейно независимых элементов u_i. Это определение выражается следующей формулой:

$$H^{(N)} = \{f | f = \sum_{i=1}^{N} \alpha_i u_i \ , \quad \forall \alpha_i \in R^{(1)}\}$$

где $R^{(1)}$ - одномерное арифметическое пространство (множество вещественных чисел). Значок | читается "такие,что". Значок $\forall$ читается "для всех".

П2. Абстрактная тензорная нотация

Абстрактной тензорной нотацией называют способ краткой записи тензорных соотношений, не зависящий от выбора систем координат.

При записи формул приняты следующие обозначения: f, g - скаляры, $\mathbf{a}, \mathbf{b}, ...$ - векторы, $\mathbf{T}$ - тензор второго ранга, $\mathbf{I}$ - единичный тензор второго ранга, $\mathbf{e}_i$ $(i = 1, 2, 3)$ - ортонормированный базис декартовой системы координат, по повторяющимся индексам подразумевается суммирование, заключенный в скобки индекс не суммируется, $\mathbf{r} = x_i\mathbf{e}_i$ - радиус-вектор, V - область трехмерного пространства, ограниченная поверхностью S, $\mathbf{n}$ - вектор единичной внешней нормали. В ряде формул рассмотрена плоская поверхность S , ограниченная контуром C с элементом dl. По повторяющимся индексам подразумевается суммирование от 1 до 3 (правило Эйнштейна).

Подчеркнем, что, вопреки распространенному заблуждению, приводимая ниже запись выражений в абстрактной тензорной нотации, кроме выражений, явно использующих индексно-компонентную нотацию, справедливы для любой системы координат, а не только для декартовой.

Для вычисления компонент некоторого тензора (в частности, вектора) в криволинейной системе координат, надо воспользоваться его диадным определением в декартовой системе координат и по правилам тензорного анализа произвести замену базиса и компонент в соответствии с заданным преобразованием декартовых координат в криволинейные.

Краткое, но достаточно ясное и простое изложение основ тензорного анализа в связи с механикой сплошной среды и объясне-

ние абстрактной тензорной нотации можно найти в книгах Астариты и Маруччи (1979), а также Коларова, Балтова и Бончевой (1980).

Справочные формулы:

$$\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}=\mathbf{a}\times\mathbf{b}\cdot\mathbf{c}=\mathbf{b}\cdot\mathbf{c}\times\mathbf{a}=\mathbf{b}\times\mathbf{c}\cdot\mathbf{a}=\mathbf{c}\cdot\mathbf{a}\times\mathbf{b}=\mathbf{c}\times\mathbf{a}\cdot\mathbf{b}$$

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c})=(\mathbf{c}\times\mathbf{b})\times\mathbf{a}=(\mathbf{a}\cdot\mathbf{c})\mathbf{b}-(\mathbf{a}\cdot\mathbf{b})\cdot\mathbf{c}$$

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c})+\mathbf{b}\times(\mathbf{c}\times\mathbf{a})+(\mathbf{c}\times\mathbf{a})\times\mathbf{b}=0$$

$$(\mathbf{a}\times\mathbf{b})\cdot(\mathbf{c}\times\mathbf{d})=(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})-(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})$$

$$(\mathbf{a}\times\mathbf{b})\times(\mathbf{c}\times\mathbf{d})=(\mathbf{a}\times\mathbf{b}\cdot\mathbf{d})\mathbf{c}-(\mathbf{a}\times\mathbf{b}\cdot\mathbf{c})\mathbf{d}$$

$$\nabla=\mathbf{e}_i\partial/\partial\mathbf{x}_i$$

$$\nabla(fg)=\nabla(gf)=f\nabla g+g\nabla f$$

$$\nabla\cdot(f\mathbf{a})=f\nabla\cdot\mathbf{a}+\mathbf{a}\cdot\nabla f$$

$$\nabla\times(f\mathbf{a})=f\nabla\times\mathbf{a}+\nabla f\times\mathbf{a}$$

$$\nabla\cdot(\mathbf{a}\times\mathbf{b})=\mathbf{b}\cdot\nabla\times\mathbf{a}-\mathbf{a}\cdot\nabla\times\mathbf{b}$$

$$\nabla\times(\mathbf{a}\times\mathbf{b})=\mathbf{a}(\nabla\cdot\mathbf{b})-\mathbf{b}(\nabla\cdot\mathbf{a})+(\mathbf{b}\cdot\nabla)\mathbf{a}-(\mathbf{a}\cdot\nabla)\mathbf{b}$$

$$\mathbf{a}\times(\nabla\times\mathbf{b})=(\nabla\mathbf{b})\cdot\mathbf{a}-(\mathbf{a}\cdot\nabla)\mathbf{b}$$

$$\mathbf{a}\times(\nabla\times\mathbf{b})=(\nabla\mathbf{b})\cdot\mathbf{a}+\mathbf{b}\times(\nabla\times\mathbf{a})+(\mathbf{a}\cdot\nabla)\mathbf{b}+(\mathbf{b}\times\nabla)\mathbf{a}$$

$$\nabla^2 f=\nabla\cdot\nabla f$$

$$\nabla^2\mathbf{a}=\nabla(\nabla\cdot\mathbf{a})-\nabla\times\nabla\times\mathbf{a}$$

$$\nabla\times\nabla f=0$$

$$\nabla\cdot\nabla\times\mathbf{a}=0$$

$$\mathbf{T}=T_{ij}\mathbf{e}_i\mathbf{e}_j$$

$$\nabla\cdot\mathbf{T}=\partial T_{ij}/\partial x_j\mathbf{e}_i$$

$$\nabla\cdot(\mathbf{ab})=(\nabla\cdot\mathbf{a})\mathbf{b}+(\mathbf{a}\cdot\nabla)\mathbf{b}$$

$$\nabla\cdot(f\mathbf{T})=(\nabla\cdot f)\mathbf{T}+f\nabla\cdot\mathbf{T}$$

$$\nabla \cdot \mathbf{r} = 3$$

$$\nabla \times \mathbf{r} = 0$$

$$\nabla |\mathbf{r}| = \mathbf{r}/|\mathbf{r}|$$

$$\nabla (1/|\mathbf{r}|) = -\mathbf{r}/|\mathbf{r}|^3$$

$$\nabla \otimes \mathbf{r} = \mathbf{I}$$

$$\int\limits_V \nabla f dV = \int\limits_S f \mathbf{n} dS$$

$$\int\limits_V \nabla \cdot \mathbf{a} dV = \int\limits_S \mathbf{a} \cdot \mathbf{n} dS$$

$$\int\limits_V \nabla \cdot \mathbf{T} dV = \int\limits_S \mathbf{T} \cdot \mathbf{n} dS$$

$$\int\limits_V \nabla \times \mathbf{a} dV = \int\limits_S \mathbf{a} \times \mathbf{n} dS$$

$$\int\limits_V (f \nabla^2 g - g \nabla^2 f) dV = \int\limits_S (f \nabla g - g \nabla f) \cdot \mathbf{n} dS$$

$$\int\limits_V (a \cdot \nabla \times \nabla \times \mathbf{b} - \mathbf{b} \cdot \nabla \times \nabla \times \mathbf{a}) dV =$$

$$= \int\limits_S (\mathbf{b} \times \nabla \times \mathbf{a} - \mathbf{a} \times \nabla \times \mathbf{b}) \cdot \mathbf{n} dS$$

$$\int\limits_S \nabla f \times \mathbf{n} dS = \int\limits_C f \mathbf{n} dl$$

$$\int\limits_S (\nabla \times \mathbf{a}) \cdot \mathbf{n} dS = \int\limits_C \mathbf{a} \cdot \mathbf{n} dl$$

$$\int\limits_S (\nabla f \times \nabla g) \cdot \mathbf{n} dS = \int\limits_C f dg = -\int\limits_C g df$$

П3. Операторы в криволинейных координатах

Цилиндрические координаты

Дивергенция

$$\nabla \cdot \mathbf{a} = \frac{1}{r}\frac{\partial}{\partial r}(r a_r) + \frac{1}{r}\frac{\partial a_\theta}{\partial \theta} + \frac{\partial a_z}{\partial z}$$

Градиент

$$(\nabla f)_r = \frac{\partial f}{\partial r}$$

$$(\nabla f)_\theta = \frac{1}{r}\frac{\partial f}{\partial \theta}$$

$$(\nabla f)_z = \frac{\partial f}{\partial z}$$

Вихрь (Ротор)

$$(\nabla \times \mathbf{a})_r = \frac{1}{r}\frac{\partial a_z}{\partial \theta} - \frac{\partial a_\theta}{\partial z}$$

$$(\nabla \times \mathbf{a})_\theta = \frac{\partial a_r}{\partial z} - \frac{\partial a_z}{\partial r}$$

$$(\nabla \times \mathbf{a})_z = \frac{1}{r}\frac{\partial (r a_\theta)}{\partial r} - \frac{1}{r}\frac{\partial a_r}{\partial \theta}$$

Лапласиан

$$\nabla^2 f = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial f}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{\partial^2 f}{\partial z^2}$$

Лапласиан вектора

$$(\nabla^2 \mathbf{a})_r = \nabla^2 a_r - \frac{2}{r^2}\frac{\partial a_\theta}{\partial \theta} - \frac{a_r}{r^2}$$

$$(\nabla^2 \mathbf{a})_\theta = \nabla^2 a_\theta + \frac{2}{r^2}\frac{\partial a_r}{\partial \theta} - \frac{a_\theta}{r^2}$$

$$(\nabla^2\mathbf{a})_z = \nabla^2 a_z$$

Компоненты конвективного члена $(\mathbf{a}\cdot\nabla)\mathbf{b}$

$$((\mathbf{a}\cdot\nabla)\mathbf{b})_r = a_r\frac{\partial b_r}{\partial r} + \frac{a_\theta}{r}\frac{\partial b_r}{\partial\theta} + a_z\frac{\partial b_r}{\partial z} - \frac{a_\theta b_\theta}{r}$$

$$((\mathbf{a}\cdot\nabla)\mathbf{b})_\theta = a_r\frac{\partial b_\theta}{\partial r} + \frac{a_\theta}{r}\frac{\partial b_\theta}{\partial\theta} + a_z\frac{\partial b_\theta}{\partial z} + \frac{a_\theta b_r}{r}$$

$$((\mathbf{a}\cdot\nabla)\mathbf{b})_z = a_r\frac{\partial b_z}{\partial r} + \frac{a_\theta}{r}\frac{\partial b_z}{\partial\theta} + a_z\frac{\partial b_z}{\partial z}$$

Дивергенция тензора

$$(\nabla\cdot\mathbf{T})_r = \frac{1}{r}\frac{\partial}{\partial r}(rT_{rr}) + \frac{1}{r}\frac{\partial T_{\theta r}}{\partial\theta} + \frac{\partial T_{zr}}{\partial z} - \frac{T_{\theta\theta}}{r}$$

$$(\nabla\cdot\mathbf{T})_\theta = \frac{1}{r}\frac{\partial}{\partial r}(rT_{r\theta}) + \frac{1}{r}\frac{\partial T_{\theta\theta}}{\partial\theta} + \frac{\partial T_{z\theta}}{\partial z} + \frac{T_{\theta r}}{r}$$

$$(\nabla\cdot\mathbf{T})_z = \frac{1}{r}\frac{\partial}{\partial r}(rT_{rz}) + \frac{1}{r}\frac{\partial T_{\theta z}}{\partial\theta} + \frac{\partial T_{zz}}{\partial z}$$

Сферические координаты

Дивергенция вектора

$$(\nabla\cdot\mathbf{a}) = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 a_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta a_\theta) + \frac{1}{r\sin\theta}\frac{\partial a_\varphi}{\partial\varphi}$$

Градиент

$$(\nabla f)_r = \frac{\partial f}{\partial r}$$

$$(\nabla f)_\theta = \frac{1}{r}\frac{\partial f}{\partial\theta}$$

$$(\nabla f)_\varphi = \frac{1}{r\sin\theta}\frac{\partial f}{\partial\varphi}$$

Вихрь

$$(\nabla\times\mathbf{a})_r = \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta a_\varphi) - \frac{1}{r\sin\theta}\frac{\partial a_\theta}{\partial\varphi}$$

$$(\nabla \times \mathbf{a})_\theta = \frac{1}{r\sin\theta}\frac{\partial a_r}{\partial \varphi} - \frac{1}{r}\frac{\partial}{\partial r}(r a_\varphi)$$

$$(\nabla \times \mathbf{a})_\varphi = \frac{1}{r}\frac{\partial}{\partial r}(r a_\theta) - \frac{1}{r}\frac{\partial a_r}{\partial \theta}$$

Лапласиан

$$\nabla^2 f = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2\frac{\partial f}{\partial r}) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial \theta}(\sin\theta\frac{\partial f}{\partial \theta}) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 f}{\partial \varphi^2}$$

Лапласиан вектора

$$(\nabla^2\mathbf{a})_r = \nabla^2 a_r - \frac{2a_r}{r^2} - \frac{2}{r^2}\frac{\partial a_\theta}{\partial\theta} - \frac{2\cot\theta a_\theta}{r^2} - \frac{2}{r^2\sin\theta}\frac{\partial a_\varphi}{\partial\varphi}$$

$$(\nabla^2\mathbf{a})_\theta = \nabla^2 a_\varphi + \frac{2}{r^2}\frac{\partial a_r}{\partial\theta} - \frac{a_\theta}{r^2\sin^2\theta} - \frac{2\cos\theta}{r^2\sin^2\theta}\frac{\partial a_\varphi}{\partial\varphi}$$

$$(\nabla^2\mathbf{a})_\varphi = \nabla^2 a_\varphi - \frac{a_\varphi}{r^2\sin^2\theta} + \frac{2}{r^2\sin\theta}\frac{\partial a_r}{\partial\varphi} + \frac{2\cos\theta}{r^2\sin^2\theta}\frac{\partial a_\theta}{\partial\varphi}$$

Компоненты конвективного члена $(\mathbf{a}\cdot\nabla)\mathbf{b}$

$$(\mathbf{a}\cdot\nabla b)_r = a_r\frac{\partial b_r}{\partial r} + \frac{a_\theta}{r}\frac{\partial b_r}{\partial\theta} + \frac{a_\varphi}{r\sin\theta}\frac{\partial b_r}{\partial\varphi} - \frac{a_\theta b_\theta + a_\varphi b_\varphi}{r}$$

$$(\mathbf{a}\cdot\nabla b)_\theta = a_r\frac{\partial b_\theta}{\partial r} + \frac{a_\theta}{r}\frac{\partial b_\theta}{\partial\theta} + \frac{a_\varphi}{r\sin\theta}\frac{\partial b_\theta}{\partial\varphi} + \frac{a_\theta b_r}{r} - \frac{\cot\theta a_\varphi b_\varphi}{r}$$

$$(\mathbf{a}\cdot\nabla b)_\varphi = a_r\frac{\partial b_\varphi}{\partial r} + \frac{a_\theta}{r}\frac{\partial b_\varphi}{\partial\theta} + \frac{a_\varphi}{r\sin\theta}\frac{\partial b_\varphi}{\partial\varphi} + \frac{a_\varphi b_r}{r} + \frac{\cot\theta a_\varphi b_\theta}{r}$$

Дивергенция тензора

$$(\nabla\cdot\mathbf{T})_r = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 T_{rr}) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta T_{\theta r}) +$$

$$+\frac{1}{r\sin\theta}\frac{\partial T_{\varphi r}}{\partial\varphi} - \frac{T_{\theta\theta}+T_{\varphi\varphi}}{r}$$

$$(\nabla\cdot\mathbf{T})_\theta = \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 T_{r\theta}) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta T_{\theta\theta})$$

$$+\frac{1}{r\sin\theta}\frac{\partial T_{\varphi\theta}}{\partial\varphi}+\frac{T_{\theta r}}{r}-\frac{\cot\theta T_{\varphi\varphi}}{r}$$

$$(\nabla\cdot\mathbf{T})_{\varphi}=\frac{1}{r^2}\frac{\partial}{\partial r}(r^2T_{r\varphi})+\frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta T_{\theta\varphi})$$

$$+\frac{1}{r\sin\theta}\frac{\partial T_{\varphi\varphi}}{\partial\varphi}+\frac{T\varphi r}{r}-\frac{\cot\theta T_{\varphi\theta}}{r}$$

П4. Об использовании криволинейных координат

Сделаем важное для практики вычислений замечание. Приведенные в предыдущем разделе формулы записаны не столько для практического использования, сколько для наглядной демонстрации громоздкости записи уравнений, если в качестве зависимых (искомых) переменных выбраны компоненты тензоров относительно базиса криволинейной системы координат.

Здесь еще не выписаны соотношения для произвольных криволинейных координат, которые выглядят еще более громоздко.

Использование проекций искомых тензоров на базис криволинейной системы координат имеет смысл только, если область решения ограничена координатными поверхностями и если имеется возможность использования симметрии искомого решения (осевой или центральной).

При использовании компонент тензоров (векторов) в базисе произвольной криволинейной системы координат придется

* различать математические и физические компоненты тензоров, иначе возникнут осложнения, связанные с размерностью искомых зависимых переменных;

* придется учитывать переменность базиса произвольной системы координат в пространстве и времени;

* использовать при записи уравнений ковариантное дифференцирование;

* вычислять переменные во времени и пространстве символы Кристоффеля, определяемые вторыми пространственными производными от радиус-вектора;

* придется повысить требования к гладкости решения из-за символов Кристоффеля;

* возрастет вероятность внесения ошибок в алгоритмы решения;

* возрастут затраты времени на составление и отладку программ для ЭВМ.

Поэтому, если в задаче симметрии нет и, тем более, если границы области решения не являются координатными поверхностями, то нет никакой уважительной причины для использования в качестве зависимых переменных компонент тензоров, отнесенных к базису криволинейной системы координат. Такой выбор зависимых переменных только усложнит формулировки задач и алгоритмы их решения.

Все упомянутые сложности исчезают, если при использовании независимых криволинейных координат продолжить пользоваться зависимыми переменными в виде компонент искомых векторов и тензоров, отнесенных к неизменному во времени и пространстве глобальному базису. Это не запрещено и является очень удобным. При этом уравнения упрошаются, нет нужды в ковариантном дифференцировании, в вычислении символов Кристоффеля, в повышении требований к гладкости решения, в разделении на математические и физические компоненты (они совпадут). Упрощается при этом и формулировка граничных условий для произвольно ориентированных и, возможно, подвижных границ.

П5. Определения основных свойств разностных схем

Аппроксимация это приближенное представление функций и уравнений для их определения.

Ошибка аппроксимации это норма разности между точным и приближенным выражениями математического объекта (оператора, уравнения, функции и так далее).

Порядок аппроксимации это показатель степенной функции, характеризующей скорость убывания ошибки аппроксимации с ростом размерности аппроксимирующего пространства.

Устойчивость это существование и ограниченность оператора, обратного исходному оператору задачи, обеспечивающие непрерывную зависимость решения от входных данных.

Сходимость это стремление нормы ошибки (разности между точным и приближенным решениями) к нулю при наращивании числа используемых базисных элементов аппроксимирующего пространства.

Транспортивность это правильное описание распространения возмущений в процессе конвекции. (Конвекция это процесс переноса субстанции вместе с упорядоченным потоком сплошной среды).

Диссипация это наличие диффузии в дифференциальном приближении схемы.

Диффузия это вязкое сглаживание (консервативный обмен сохраняемой величиной без упорядоченного потока сплошной среды). Диффузия обеспечивается хаотичным движением и взаимодействием молекул, диффузионный поток пропорционален антиградиенту диффундирующей величины с коэффициентом пропорциональности, называемым коэффициентом диффузии.

Дисперсия это зависимость фазовой скорости распространения малых возмущений от частоты.

Основными типами вязкости являются: физическая вязкость, аппроксимационная (схемная) вязкость, явная искусственная вязкость, вязкость сглаживания, эффективная (суммарная) вязкость.

Консервативность схемы это выполнение схемой законов сохранения на дискретном уровне.

Дивергентная форма записи закона сохранения это дифференциальная запись соответствующего балансного соотношения, в которой скорость изменения сохраняемой величины определяется дивергенцией ее потока.

Монотонность это свойство схем не производить новые минимумы и максимумы для рассчитываемой функции при отсутствии источников и стоков.

Однородность это одинаковость алгоритма для всех дискретных элементов численной модели.

Робастность это способность алгоритма найти решение при произвольных входных данных

Эффективность это способность схемы выдать решение в разумный срок.

Экономичность это уменьшенный объем вычислений для достижения определенной точности. Более экономичная схема требует меньшего объема вычислений для достижения той же точности. Для оценки экономичности важна асимптотическая скорость возрастания обьема вычислений при увеличении числа базисных элементов.

Точность характеризует величину ошибки приближенного решения при ограничении на число базисных элементов (членов ряда, числа конечных элементов/объемов, числа узлов сетки). Более точная схема имеет меньшую ошибку решения при том же числе базисных элементов.

Корректность задачи подразумевает выполнение следующих условий (условий корректности): 1) задача имеет решение при любых допустимых исходных данных (существование решения); 2) каждым исходным данным соответствует только одно решение (однозначность решения); 3) решение устойчиво (имеет место непрерывная зависимость решения от входных данных).

Литература

1. *Абрамов А.А.* Вариант метода прогонки. Ж. вычисл. матем. и матем. физ., 1, N2, 1961, с. 349-351.

2. Алберг Дж., Нильсон Э., Уолш Дж. Теория сплайнов и ее приложения. М.: Мир, 1972.

3. Астарита Дж., Марруччи Дж. Основы гидромеханики неньютоновских жидкостей. М. : Мир, 1978. 309 с.

4. Бабенко К.И. (Ред.) Теоретические основы и конструирование численных алгоритмов задач математической физики. М.: Наука, 1979. 295 с.

5. Баничук Н. В., Картвелишвили В. А., Черноусько Ф. Л. Вариационные задачи механики и управления. М.:Наука, 1973.

6. Бате К., Вилсон Е. Численные методы анализа и метод конечных элементов /Пер. с англ. А.С. Алексеева и др.; Под ред. А.Ф. Смирнова. М.: Стройиздат, 1982

7. Бахвалов Н.С. Численные методы, Т. 1. М.: Наука, 1973, 631 с.

8. Бенерджи П., Баттерфилд Р. Методы граничных элементов в прикладных науках. М.: Мир, 1984. 494 с.

9. Вайнберг М.М., Треногин В.А. Теория ветвлений решений нелинейных уравнений. М.: Наука, 1969. 527 с.

10. Валишвили Н.В. Методы расчета оболочек вращения на ЭЦВМ. М.: Машиностроение. 1976. - 278 с.

11. Ванько В. И., Ермошина О. В., Кувыркин Г. Н. Вариационное исчисление и оптимальное управление. М.: МГТУ, 1999. Т. 15. 487 с.

12. Воеводин В.В. Вычислительные основы линейной алгебры. М.: Наука, 1977.

13. Воеводин В.В., Кузнецов Ю.А. Матрицы и вычисления. М.: Наука, 1984.

14. Гавурин М. К. Лекции по методам вычислений. М.: Наука, 1971. 248 с.

15. Галлагер Р. МКЭ: Основы /Пер. с англ. - М.: Мир, 1984. - 215с.

16. Годунов С. К. Метод ортогональной прогонки. Ж. вычисл. матем. и матем. физ., 2, N6, 1962, s. 972-982.

17. Годунов С. К., Рябенький В. С. Разностные схемы. М.: Наука, 1973, 400 с.

18. Григолюк Э. И., Шалашилин В. И. Проблемы нелинейного деформирования. Метод продолжения по параметру в нелинейных задачах механики твердого деформируемого тела. М.: Наука. Гл. ред. Физматлит, 1988. 232с

19. Джордж А., Лю Дж. Численное решение больших разреженных систем уравнений. М.: Мир, 1984, 333 с.

20. Дьяченко В. Ф. Основные понятия вычислительной математики. М., Наука, 1972, 119 с.

21. Дьяченко В.Ф. Об одном новом методе численного решения задач газовой динамики с двумя пространственными переменными. ЖВМиМФ. 1965. Т.5. No. 4. С. 680-688.

22. Еремин А.Ю. и Марьяшкин Н.Я. Метод сопряженных градиентов с неполным разложением Холецкого для решения систем линейных алгебраических уравнений, // М., Препринт ВЦ АН СССР, 1978, б/н, с. 1-14.

23. Завьялов Ю.С., Квасов Б.И., Мирошниченко В.Л. Методы сплайн-функций, Москва: Наука 1980.

24. Зенкевич О. Метод конечных элементов в технике. М. Мир. 1975. 541 с. (O. C. Zienkiewicz. The Finite Element Method in Engineering Science. McGraw-Hill, London, 1971)

25. Иваненко С. А. Адаптивно-гармонические сетки. М.: Изд-во ВЦ РАН, 1997. 181 с.

26. Ильин В. П. Численные методы решения задач строительной механики СПбГАСУ,АСВ 2005 425с

27. Калиткин Н. Н. Численные методы. -М.:Наука,1978. 512с.

28. Калнинс А. Исследование оболочек вращения при действии симметричной и несимметричной нагрузок // Тр. амер. общ. ииж. мех. Прикладная механика. 1964. Т. 31. No. 3

29. Коларов Д., Балтов А., Бончева Н. Механика пластических сред. М.: Мир. 1979. 302 с.

30. Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. М., Наука,1972, 496 с.

31. Копченова Н. В., Марон И. А. Вычислительная математика в примерах и задачах. М.: Наука, 1972. 367 с.

32. Косарев В.И. 12 лекций по вычислительной математике. 2-е изд. - М.: Изд-во МФТИ, 2000. - 224 с.

33. Крылов В.И., Бобков В.В., Монастырский П.И. Вычислительные методы, Т. 1., Т. 2., Москва: Наука 1977.

34. Кукуджанов В.Н. Вычислительная механика сплошных сред. М.: Физматлит, 2008.

35. Куликовский А. Г., Погорелов Н. В., Семенов А. Ю. Математические вопросы численного решения гиперболических систем уравнений.М.: Физматлит, 2001. 608 с.

36. Кунин В. Вычислительная физика. М., Мир, 1979.

37. Магомедов К. М., Холодов А. С. Сеточно-характеристические методы. М.: Наука. 1988. 287с.

38. Марчук Г. И. Методы вычислительной математики. М.: Наука, 1977. 456 с.; 1989. 608 с.

39. Митчелл Э., Уэйт Р. Метод конечных элементов для уравнений с частными производными. М.:Мир. 1981. 216 с. (Mitchell A. R. and Wait R. The finite element method in partial differential equations. Wiley. N.-Y. 1977)

40. Михлин С. Г. Прямые методы в математической физике, М.-Л.: ГИТТЛ, 1950, 452 с.

41. Норри Д., де Фриз Ж Введение в метод конечных элементов. М.: "Мир", 1981. 304 с.

42. Оден Дж. Конечные элементы в нелинейной механике сплошных сред. М.: Мир, 1976. - 358 с.

43. Ортега Дж., Рейнболдт В. Итерационные методы решения нелинейных систем уравнений со многими неизвестными. М.: Мир, 1975, 558 с.

44. Петров И.Б., Лобанов А.И. Лекции по вычислительной математике: Учебное пособие. М.:Интернет-Университет информационных технологий, 2009. 522 с.

45. Полак Э. Численные методы оптимизации. М.:Мир, 1974, 376 с.

46. Поттер Д. Вычислительные методы в физике. М.:Мир, 1975, 392с.

47. Приклонский В.И. Численные методы. МГУ.:Физфак, 1999. 146с.

48. Пшеничный, Б. Н., Данилин, Ю. М. Численные методы в экстремальных задачах. М.: Наука, 1979, 319 с.

49. Рихтмайер Р.Д. Разностные методы решения краевых задач, М., ИЛ, 1960.

50. Рихтмайер Р.Д., Мортон К.В. Разностные методы решения краевых задач, М., Мир, 1972.

51. Росляков Г. С., Чудов Л. А. Численные методы в механике сплошных сред. Ч. 1-3, М.: ВЦ МГУ, 1968.

52. Роуч П. Вычислительная гидродинамика. М.: Мир, 1980, 616 с.

53. Самарский А. А. Введение в теорию разностных схем. М.: Наука, 1971.

54. Самарский, А.А. и Попов, Ю.П. (1980) Разностные методы решения задач газовой динамики. М.: Наука, 352 с.

55. Самарский А.А. Введение в численные методы.- М.: Наука, 1987. 459с.

56. Самарский А.А., Гулин А.В. Численные методы. М.:Наука, 1989. 432с.

57. Сборник задач для упражнений по курсу: Основы вычислительной математики. / Под ред. Рябенького В.С. - М.: МФТИ, 1988.

58. Сегерлинд Л. Применение метода конечных элементов. М.: Мир, 1979, 392 с

59. Современные численные методы решения обыкновенных дифференциальных уравнений / Ред. Дж. Холл, Дж. Уатт // М.: Мир, 1979. 312 с

60. Стренг Г. Фикс Г. Теория метода конечных элементов. М.: Мир, 1980.

61. Теория ветвления и нелинейные задачи на собственные значения. / Редакторы Дж. Б. Келлер, С. Антман.М.: Мир, 1974. 254 с.

62. Турчак Л.И. Основы численных методов. М.: Наука, 1987. 250 с.

63. Уилкинсон, Райнш, Справочник алгоритмов на языке АЛГОЛ. Линейная алгебра, М. Машиностроение, 1976

64. Федоренко Р. П. Введение в вычислительную физику. Учебное пособие для ВУЗов. М.: Изд-во МФТИ, 1994. 528 с.

65. Флетчер К. Численные методы на основе метода Галеркина. М.: Мир, 1988. 352 с.

66. Флетчер К. Вычислительные методы в динамике жидкостей, Мир, 1991 (т.1, 2)

67. Фокс А., Пратт М. Вычислительная геометрия. М.: Мир, 1982. 304 с.

68. Форсайт Дж., Малькольм М., Моулер К. Машинные методы математических вычислений. М.: Мир, 1980. 280 с.

69. А. Фридман. Уравнения с частными производными параболического типа. ? М.: Мир, 1968. 427 с.

70. Хокни Р., Иствуд Дж., Численное моделирование методом частиц, Мир, 1987

71. Шокин Ю. И. Метод дифференциального приближения.Новосибирск: Наука, 1979.

72. Шокин Ю. И., Яненко Н. Н., Метод дифференциального приближения. Приложения к газовой динамике. Нск.: Наука Сиб. Отд. 1985. 357с.