



ՀԱՅ-ՌՈՒՄԱԿԱՆ ՀԱՄԱԼՍԱՐԱՆ

ՅՄ.Ռ. ՀԱԿՈԲՅԱՆ

ՎԵՐՁԱՎՈՐ ՏԱՐԲԵՐԻ
ՄԵԹՈԴԻ ՆԵՐԱԾՈՒԹՅՈՒՆ

Ուսումնական ձեռնարկ

ԵՐԵՎԱՆ
ՀՌՀ հրատարակչություն
2022

РОССИЙСКО–АРМЯНСКИЙ УНИВЕРСИТЕТ

Ю.Р. АКОПЯН

**ВВЕДЕНИЕ В МЕТОД
КОНЕЧНЫХ ЭЛЕМЕНТОВ**

Учебное пособие

ЕРЕВАН
Издательство РАУ
2022

УДК 517.96(07)
ББК 22.161.6я7
А 406

*Печатается по решению Ученого совета и РИС РАУ и
кафедры математики и математического моделирования*

Рецензенты: член-корр. НАН РА, профессор С.М. Мхитарян
доктор физ.-мат. наук, профессор Х.А. Хачатрян

Акопян Ю.Р.

А 406 Введение в метод конечных элементов/Ю.Р. Акопян. - Ер.: Изд-во РАУ,
2022.-142с.

Предлагаемая книга посвящена методу конечных элементов, который получил широкое распространение как один из наиболее эффективных инструментов численного решения дифференциальных уравнений в частных производных. Издание может рассматриваться как введение в современные исследования по изучаемому вопросу. На примере уравнения Пуассона излагаются основные принципы, методы и техника построения систем конечноэлементных уравнений.

Учебное пособие предназначено для студентов старших курсов, магистрантов и аспирантов математических специальностей университетов. Она будет полезна также научным работникам, применяющим метод конечных элементов на практике.

УДК 517.96(07)
ББК 22.161.6я7

ISBN 978-9939-67-294-6

© Издательство РАУ, 2022
© Акопян Ю.Р., 2022

ОГЛАВЛЕНИЕ

Предисловие	7
Глава 1. ВАРИАЦИОННЫЕ И ПРОЕКЦИОННЫЕ МЕТОДЫ	9
1.1 Гильбертово пространство	9
1.2 Линейные операторы	13
1.3 Энергетическое пространство	15
1.4 Задача о минимуме функционала энергии	19
1.5 Дифференциальные операторы краевых задач	22
1.6 Метод Рунта	28
1.7 Метод Бубнова-Галеркина	32
Глава 2. МЕТОД КОНЕЧНЫХ ЭЛЕМЕНТОВ ДЛЯ УРАВНЕНИЯ ПУАССОНА 35	
2.1 О методе конечных элементов	35
2.2 Двухточечная краевая задача	37
2.3 Пространства функций, области, границы областей	45
2.4 Сеточные области и восполнения функций	48
2.5 Конечноэлементное решение уравнения Пуассона	58
2.6 Теоремы аппроксимации	65
2.7 Неравенства по приграничным полоскам	74
2.8 Теоремы аппроксимации (продолжение)	80
2.9 Оценки погрешности приближенного решения	85
Глава 3. ПЕРЕОБУСЛАВЛИВАНИЕ КОНЕЧНОЭЛЕМЕНТНЫХ СИСТЕМ	89
3.1 Ансамблирование матрицы жесткости	89
3.2 Сеточные уравнения и задача переобуславливания	97
3.3 Иерархические треугольные сетки	102
3.4 Спектрально эквивалентная матрица	105
3.5 О дополнении Шура	110
3.6 Двухсеточные переобуславливатели	116

3.7 О чебышевском итерационном методе	125
3.8 Многосегочный переобуславливатель	128
Литература	137
Предметный указатель	140

Предисловие

Математическое моделирование многих естественнонаучных явлений и процессов приводит к необходимости решения дифференциальных уравнений в частных производных. Однако в большинстве случаев найти точное решение дифференциального уравнения, которое выражалось бы в привычных для нас элементарных функциях, не удастся. Поэтому важное значение приобрели численные методы решения уравнений. К таким методам относится *метод конечных элементов*. Этот метод получил широкое распространение на практике как эффективный инструмент решения краевых задач, особенно в областях сложной геометрической формы, где применение разностных методов часто бывает затруднительным. На его основе строятся также и сеточные методы решения нестационарных задач математической физики, в которых конечноэлементная дискретизация по пространственным переменным сочетается с разностной дискретизацией по времени.

На сегодняшний день существует довольно обширная литература по методу конечных элементов — статьи, учебники и монографии. При этом многие книги — при всей полноте исследования изучаемого вопроса и высоком научном уровне — отличаются большим объемом и разнообразием предоставляемого в них материала; это создает определенные трудности при первоначальном ознакомлении с предметом и использовании их в учебном процессе.

Главной целью настоящего сравнительно небольшого по объему учебного пособия является систематическое и, по возможности, несложное изложение основ метода конечных элементов применительно к двумерным эллиптическим уравнениям второго порядка; для этого изложение ведется на примере наиболее известного и простого по формулировке представителя указанного класса уравнений — уравнения Пуассона. Вместе с тем ход рассуждений и практически все полученные результаты могут быть по аналогии перенесены на эллиптические уравнения общего вида. Книга рассчитана на студентов, магистрантов и аспирантов математических специальностей университетов. Для ее понимания требуется знание математического анализа и линейной алгебры в рамках университетского курса, а также знакомство с основами функционального анализа и уравнений математической физики.

Основной материал книги распределен следующим образом.

Глава 1 носит вводный характер. Она посвящена описанию вариационных и проекционных методов решения линейных операторных уравнений в гильбертовом пространстве. Сначала рассматривается вариационный *метод Рунца* — как метод построения мини-

мизирующей последовательности, сходящейся к обобщенному решению задачи о минимуме функционала энергии положительно определенного оператора. Затем излагается имеющий более широкую область применения *метод Бубнова-Галеркина*.

В главе 2 метод конечных элементов характеризуется как результат синтеза метода конечных разностей и методов Рунге и Бубнова-Галеркина. Даются основные принципы, методы и техника построения систем конечноэлементных уравнений на треугольных сетках. Значительное место в главе отводится выводу оценок погрешности приближенного решения. Для этого доказываются теоремы об аппроксимации функций их кусочно-линейными восполнениями.

В главе 3 обсуждаются вопросы, связанные с численной реализацией метода конечных элементов. Рассматривается важная задача переобуславливания систем конечноэлементных уравнений. Описывается одна из версий построения так называемых *многосеточных переобуславливателей*, в основе которой лежит идея использования иерархически измельчающихся сеток и внутренних чебышевских итерационных процедур.

Нумерация определений, теорем, лемм, замечаний, рисунков и формул — автономная в каждом параграфе. Доказательства утверждений завершается значком \square .

Ю. Р. Акопян

Глава 1

ВАРИАЦИОННЫЕ И ПРОЕКЦИОННЫЕ МЕТОДЫ

Настоящая глава носит вводный характер. Она посвящена описанию *вариационных* и *проекционных* методов решения линейных операторных уравнений в гильбертовых пространствах. В основе *вариационных* методов — применительно к симметричным положительно определенным операторам — лежит идея сведения исходной задачи к равносильной задаче отыскания минимума специальным образом определенного линейного функционала. Последующее численное решение вариационной задачи осуществляется путем решения аналогичных задач на последовательности вложенных конечномерных подпространств (*метод Рунца*). Другой подход к построению вычислительных процессов для решения операторных уравнений реализуется в так называемых *проекционных методах*, основанных на проектировании уравнения на конечномерные подпространства. Эти методы имеют более широкую область применения, так как в них, в принципе, можно отказаться от требования симметричности и положительной определенности оператора задачи. Наиболее известным методом такого типа является *метод Бубнова-Галеркина*, изложенный в последнем параграфе. В несколько видоизмененном виде указанные методы применяются в последующих главах при численном решении дифференциальных уравнений в частных производных эллиптического типа.

1.1 Гильбертово пространство

Понятие нормированного пространства является одним из основных в математике. Напомним его определение. При этом мы полагаем, что читатель знаком с определением линейного пространства (см., например, [6, 14]). В нашем учебном курсе мы ограничимся рассмотрением *вещественных* линейных пространств.

Определение 1.1.1 Линейное пространство L называется *нормированным пространством*, если каждому элементу $x \in L$ поставлено в соответствие некоторое вещественное число $\|x\|$, называемое *нормой* этого элемента, так, что выполняются условия:

1. $\|x\| \geq 0$, $\|x\| = 0 \Leftrightarrow x = \mathbf{0}$,
2. $\|\lambda x\| = |\lambda| \|x\|$,
3. $\|x + y\| \leq \|x\| + \|y\|$ (неравенство треугольника)

для любых элементов $x, y \in L$ и любого вещественного числа λ .

Приведем несколько примеров нормированных пространств:

- пространство векторов $x = [x_1 \ x_2 \ \dots \ x_n]^T \in R^n$ с нормой

$$\|x\| \equiv \max_{1 \leq i \leq n} |x_i|;$$

- пространство векторов $x = [x_1 \ x_2 \ \dots \ x_n]^T \in R^n$ с нормой

$$\|x\| \equiv \left(\sum_{i=1}^n x_i^2 \right)^{1/2};$$

- пространство непрерывных на отрезке $[a, b]$ функций с нормой

$$\|f\| \equiv \max_{x \in [a, b]} |f(x)|.$$

Элемент x называется *нормированным*, если $\|x\| = 1$.

Далее, введем понятие сходимости в нормированном пространстве. Говорят, что последовательность элементов $\{x_n\}_{n=1}^{\infty}$ нормированного пространства L *сходится* к элементу $x \in L$, если

$$\|x_n - x\| \rightarrow 0 \quad \text{при} \quad n \rightarrow \infty.$$

Для указания сходимости последовательности используется символика

$$x_n \rightarrow x \quad \text{или} \quad \lim_{n \rightarrow \infty} x_n = x.$$

Для сходящейся последовательности верно, что

$$\|x_n\| \rightarrow \|x\| \quad \text{при} \quad n \rightarrow \infty.$$

Это утверждение очевидно следует из неравенства

$$|\|x_n\| - \|x\|| \leq \|x_n - x\|,$$

которое, в свою очередь, легко следует из неравенства треугольника.

Последовательность элементов $\{x_n\}_{n=1}^{\infty}$ нормированного пространства L называется *фундаментальной*, если

$$\|x_m - x_n\| \rightarrow 0 \quad \text{при} \quad m, n \rightarrow \infty.$$

Определение 1.1.2 Нормированное пространство называется *полным*, если в нем всякая фундаментальная последовательность имеет предел. Это пространство называется также *банаховым пространством*¹ или короче: *B-пространством*.

Рассмотрим еще один тип пространства.

Определение 1.1.3 Линейное пространство L называется *евклидовым*, если каждой паре элементов $x, y \in L$ поставлено в соответствие некоторое вещественное число (x, y) , называемое *скалярным произведением* этих элементов, так, что выполнены условия

1. $(x, x) \geq 0$, $(x, x) = 0 \Leftrightarrow x = 0$,
2. $(x, y) = (y, x)$,
3. $(\lambda x, y) = \lambda (x, y)$,
4. $(x + y, z) = (x, z) + (y, z)$

для любых элементов $x, y, z \in L$ и произвольного вещественного числа λ .

Например, линейное пространство векторов $x = [x_1 \ x_2 \ \dots \ x_n]^T \in R^n$ становится евклидовым, если скалярное произведение в нем определить следующим образом:

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

Элементы x и y евклидова пространства называются *ортгоналными*, если $(x, y) = 0$.

Евклидово пространство можно превратить в нормированное, определив норму элемента следующим образом:

$$\|x\| \equiv \sqrt{(x, x)}. \quad (1.1.1)$$

Определение 1.1.4 Евклидово пространство, в котором введена норма посредством соотношения (1.1.1), называется *унитарным пространством*.

Тем самым мы подошли к определению *гильбертова пространства*, в названии которого увековечена память о крупнейшем немецком ученом Давиде Гильберте (1862–1943), оказавшем большое влияние на развитие многих разделов математики.

Определение 1.1.5 Полное бесконечномерное унитарное пространство называется *гильбертовым пространством*.

В качестве примера гильбертова пространства приведем пространство l_2 бесконечных последовательностей $x = (x_1, x_2, \dots)$ вещественных чисел, удовлетворяющих условию

$$\sum_{i=1}^{\infty} x_i^2 < \infty.$$

¹Название пространства связано с именем польского математика Стефана Банаха (1892–1945) — одного из создателей функционального анализа.

Скалярное произведение и норма в этом пространстве определяются следующим образом:

$$(x, y) \equiv \sum_{i=1}^{\infty} x_i y_i, \quad \|x\| \equiv \sqrt{\sum_{i=1}^{\infty} x_i^2}.$$

Другим известным примером гильбертова пространства является пространство $L_2[a, b]$ функций $f(x)$, суммируемых с квадратом на отрезке $[a, b]$, то есть таких, что

$$\int_a^b f^2(x) dx < \infty.$$

Скалярное произведение и норма в пространстве $L_2[a, b]$ вводятся по формулам:

$$(f, g) \equiv \int_a^b f(x)g(x) dx, \quad \|f\| \equiv \sqrt{\int_a^b f^2(x) dx}.$$

В гильбертовом пространстве справедливо *неравенство Коши²-Буняковского³*

$$|(x, y)| \leq \|x\| \|y\| \quad \forall x, y. \quad (1.1.2)$$

Доказывается оно достаточно просто (см., например, [8, 13]). В упомянутых выше пространствах l_2 и $L_2[a, b]$ неравенство (1.1.2) принимает вид

$$\left| \sum_{i=1}^{\infty} x_i y_i \right| \leq \sqrt{\sum_{i=1}^{\infty} x_i^2} \sqrt{\sum_{i=1}^{\infty} y_i^2} \quad (1.1.3)$$

и

$$\left| \int_a^b f(x)g(x) dx \right| \leq \sqrt{\int_a^b f^2(x) dx} \sqrt{\int_a^b g^2(x) dx}, \quad (1.1.4)$$

соответственно.

В заключение параграфа введем еще несколько важных для дальнейшего изложения материала понятий.

Пусть множество M есть некоторое подмножество гильбертова пространства H .

²Коши Огюстен Луи (1789–1857) — французский математик, внес фундаментальный вклад в развитие математического анализа, теории функций и математической физики.

³Буняковский Виктор Яковлевич (1804–1889) — русский математик, основные работы относятся к теории чисел и теории вероятностей.

Точка $x \in H$ называется *точкой прикосновения* множества M , если любая ее окрестность содержит хотя бы одну точку из M . Совокупность всех точек прикосновения множества M обозначается $[M]$ и называется *замыканием* этого множества. Таким образом, запись $x \in [M]$ означает, что либо $x \in M$, либо $x = \lim_{n \rightarrow \infty} x_n$, где $x_n \in M$.

Множество $M \subset H$ называется *замкнутым*, если $[M] = M$.

Пусть A и B есть два подмножества гильбертова пространства H . Множество A называется *плотным* в B , если $B \subseteq [A]$. В частности, множество A называется *всюду плотным* в H , если $[A] = H$.

1.2 Линейные операторы

Пусть на множестве D_A элементов гильбертова пространства H определен *оператор* A , который каждому элементу $u \in D_A$ ставит в соответствие по некоторому правилу один и только один элемент $v \in H$. При этом мы будем писать $v = Au$ и называть элемент v *образом* элемента u , а элемент u — *прообразом* элемента v . Множество D_A называется *областью определения* оператора A . Множество $R_A = \{v \in H : v = Au, u \in D_A\}$ называется *областью значений* оператора. Фактически, оператор A есть отображение $A : D_A \rightarrow R_A$.

Мы будем полагать, что область определения D_A есть *линейное множество*, то есть если $u_1, u_2 \in D_A$, то и $\alpha u_1 + \beta u_2 \in D_A$ для всех вещественных чисел α, β .

Определение 1.2.1 Оператор A называется *линейным*, если

$$A(\alpha u_1 + \beta u_2) = \alpha Au_1 + \beta Au_2$$

для всех $u_1, u_2 \in D_A$ и любых вещественных чисел α, β .

Определение 1.2.2 Линейный оператор A называется *непрерывным в точке* $u \in D_A$, если $Au_n \rightarrow Au$ при $u_n \rightarrow u$. Оператор A называется *непрерывным*, если он непрерывен в каждой точке $u \in D_A$.

Дадим еще одно определение, тесно связанное с понятием непрерывности оператора.

Определение 1.2.3 Линейный оператор A называется *ограниченным*, если существует такая постоянная C , что

$$\|Au\| \leq C \|u\| \tag{1.2.1}$$

для всех $u \in D_A$.

Наименьшее из чисел C , для которых справедливо неравенство (1.2.1), называется *нормой* оператора A и обозначается $\|A\|$. В курсе функционального анализа (см., например, [13]) доказывается, что

$$\|A\| = \sup_{u \neq 0} \frac{\|Au\|}{\|u\|} = \sup_{\|u\| \leq 1} \|Au\|. \tag{1.2.2}$$

Какова связь между непрерывностью оператора и его ограниченностью? Оказывается, что *линейный оператор непрерывен тогда и только тогда, когда он ограничен*. Доказательство этого утверждения можно найти, например, в [13, 17].

Далее мы будем рассматривать только линейные непрерывные или, что то же самое, ограниченные операторы. Более того, мы будем рассматривать линейные непрерывные операторы с всюду плотными в H областями определения.

Определение 1.2.4 Линейный непрерывный оператор A называется *симметричным*, если

$$(Au, v) = (u, Av)$$

для всех $u, v \in D_A$.

Определение 1.2.5 Симметричный оператор A называется *положительно определенным*, если существует такая положительная постоянная γ , что

$$(Au, u) \geq \gamma \|u\|^2 \quad (1.2.3)$$

для всех $u \in D_A$.

В заключение настоящего параграфа приведем пример положительно определенного оператора.

Пусть оператор A определен в пространстве $L_2[0, 1]$ на множестве функций, которые имеют непрерывные вторые производные на отрезке $[0, 1]$ и удовлетворяют краевым условиям

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

Оператор A действует по формуле

$$Au = -u''.$$

Заметим, что область определения оператора является всюду плотной в пространстве $L_2[0, 1]$. Доказательство этого можно найти, например, в книге [21].

Установим, прежде всего, симметричность оператора. Применяя формулу интегрирования по частям, получим

$$(Au, v) = - \int_0^1 u''v \, dx = -u'v \Big|_0^1 + \int_0^1 u'v' \, dx = \int_0^1 u'v' \, dx = (u, Av).$$

Далее, так как $u(0) = 0$, то

$$u(x) = \int_0^x u'(t) \, dt.$$

Воспользуемся неравенством Коши-Буняковского (1.1.4):

$$u^2(x) = \left(\int_0^x u'(t) dt \right)^2 \leq \int_0^x 1^2 dt \int_0^x (u'(t))^2 dt = x \int_0^x (u'(t))^2 dt.$$

Усилим последнее неравенство, заменив верхний предел интегрирования единицей:

$$u^2(x) \leq x \int_0^1 (u'(t))^2 dt.$$

Интегрируя обе части этого неравенства в пределах от нуля до единицы, получим

$$\int_0^1 u^2(x) dx \leq \frac{1}{2} \int_0^1 (u'(t))^2 dt. \quad (1.2.4)$$

Последнее неравенство можно записать в виде

$$(Au, u) \geq 2 \|u\|^2.$$

Таким образом, по определению 1.2.5 наш оператор A положительно определенный.

1.3 Энергетическое пространство

Пусть A — положительно определенный оператор, действующий в гильбертовом пространстве H . Напомним, что в силу нашего соглашения область определения D_A оператора всюду плотна в H (см. определение 1.2.4). На множестве элементов D_A введем новое скалярное произведение

$$[u, v] \equiv (Au, v) \quad \forall u, v \in D_A \quad (1.3.1)$$

(для обозначения нового скалярного произведения мы будем использовать квадратные скобки). Нетрудно проверить выполнимость всех условий скалярного произведения, указанных в определении 1.1.3. Далее, определим новую норму

$$\| [u] \| \equiv \sqrt{[u, u]} = \sqrt{(Au, u)}. \quad (1.3.2)$$

Вводя новое скалярное произведение и соответствующую норму, мы переходим от линейного множества D_A к унитарному пространству. При этом, пользуясь неравенством (1.2.3), получим

$$\| [u] \| = \sqrt{(Au, u)} \geq \sqrt{\gamma \|u\|^2} = \sqrt{\gamma} \|u\|,$$

откуда следует неравенство

$$\|u\| \leq \frac{1}{\sqrt{\gamma}} \|[u]\|, \quad (1.3.3)$$

верное для всех $u \in D_A$.

Однако наше унитарное пространство D_A может оказаться неполным. В таком случае мы можем превратить его в гильбертово, присоединив к нему некоторые новые элементы (подобно тому, как множество рациональных чисел дополняется до всей числовой оси введением иррациональных чисел). Реализуется это следующим образом. Пусть последовательность $\{u_n\}_{n=1}^{\infty}$ элементов из D_A является фундаментальной по норме (1.3.2), но не имеет предела в D_A . Припишем этой последовательности в качестве предела новый элемент u , называемый *предельным*. Тем самым мы осуществим *пополнение* унитарного пространства D_A до гильбертова пространства, которое обозначим через H_A .

Как определяются скалярное произведение и норма предельных элементов? Пусть элементы $u, v \in H_A$. Это означает, что существуют такие фундаментальные последовательности $\{u_n\}_{n=1}^{\infty}$ и $\{v_n\}_{n=1}^{\infty}$ элементов из D_A , для которых элементы u и v приписаны в качестве предельных, соответственно. Тогда мы полагаем

$$[u, v] \equiv \lim_{n \rightarrow \infty} [u_n, v_n] \quad (1.3.4)$$

и

$$\|[u]\| \equiv \lim_{n \rightarrow \infty} \|[u_n]\| = \sqrt{[u, u]}. \quad (1.3.5)$$

Существование пределов (1.3.4) и (1.3.5) нетрудно доказать. Остановимся на пределе (1.3.4). Рассмотрим последовательность вещественных чисел $\delta_n \equiv [u_n, v_n]$, $n = 1, 2, \dots$. Покажем сначала, что эта последовательность фундаментальная. Пользуясь определением (1.3.1) нового скалярного произведения и принимая во внимание симметричность оператора A , имеем

$$\begin{aligned} |\delta_m - \delta_n| &= |[u_m, v_m] - [u_n, v_n]| = |(Au_m, v_m) - (Au_n, v_n)| \\ &= |(A(u_m - u_n), v_m) + (A(v_m - v_n), u_n)| \\ &\leq |(A(u_m - u_n), v_m)| + |(A(v_m - v_n), u_n)|. \end{aligned}$$

Далее, воспользуемся неравенством Коши-Буняковского (1.1.2). Получим

$$|\delta_m - \delta_n| \leq \|A(u_m - u_n)\| \|v_m\| + \|A(v_m - v_n)\| \|u_n\|.$$

Так как оператор A ограниченный (см. неравенство (1.2.1)), то

$$|\delta_m - \delta_n| \leq C(\|u_m - u_n\| \|v_m\| + \|v_m - v_n\| \|u_n\|).$$

Наконец, применяя неравенство (1.3.3), получим

$$|\delta_m - \delta_n| \leq \frac{C}{\gamma} (\|[u_m - u_n]\| \|[v_m]\| + \|[v_m - v_n]\| \|[u_n]\|).$$

Так как последовательности $\{u_n\}_{n=1}^{\infty}$ и $\{v_n\}_{n=1}^{\infty}$ являются фундаментальными, а тем самым и ограниченными, то $|\delta_m - \delta_n| \rightarrow 0$ при $n \rightarrow \infty$. В силу полноты пространства вещественных чисел последовательность δ_n сходится.

Итак, если $u \in H_A$, то, как следует из определения, либо $u \in D_A$, либо существует такая последовательность $\{u_n\}_{n=1}^{\infty}$ элементов из D_A , что $\|u_n - u\| \rightarrow 0$ при $n \rightarrow \infty$. Отсюда, в частности, следует, что D_A плотно в H_A .

Гильбертово пространство H_A называют *энергетическим пространством* оператора A . Величины $[u, v]$ и $\|u\|$ называются, соответственно, *энергетическим произведением* элементов $u, v \in H_A$ и *энергетической нормой* элемента $u \in H_A$.

Замечание 1.3.1 Выбор терминов *энергетическое пространство*, *энергетическая норма* и т.п. связан со следующим обстоятельством. Пусть некоторая физическая система под действием внешней причины P , характеризуемой функцией $f(P)$, получает смещение $u(P)$, и эти две функции связаны соотношением $Au = f$, где A есть положительно определенный оператор. Оказывается, что величина (Au, u) пропорциональна величине энергии, которую нужно затратить, чтобы сообщить системе смещение u .

Докажем одно важное утверждение.

Теорема 1.3.1 *Все элементы пространства H_A принадлежат также и пространству H . При этом каждому элементу из H_A соответствует один и только один элемент из H , причем разным элементам из H_A соответствуют разные элементы из H .*

Доказательство. Утверждение очевидно, если $u \in D_A$. В этом случае достаточно элементу u привести в соответствие его самого. Остается рассмотреть случай, когда u есть предельный элемент пространства H_A .

Пусть $u \in H_A \setminus D_A$. В этом случае, по определению, существует фундаментальная последовательность $u_n \in D_A$, сходящаяся к элементу u в энергетической норме. По неравенству (1.3.3) имеем

$$\|u_m - u_n\| \leq \frac{1}{\sqrt{\gamma}} \|u_m - u_n\|.$$

Поэтому $\|u_m - u_n\| \rightarrow 0$ при $m, n \rightarrow \infty$. То есть u_n есть фундаментальная последовательность в H . А так как пространство H — полное, то существует такой элемент $u' \in H$, что $\|u_n - u'\| \rightarrow 0$ при $n \rightarrow \infty$. отождествляя u и u' , мы установим первую часть нашего утверждения.

Докажем теперь, что соответствие между элементами пространств H_A и H линейно. Действительно, пусть

$$H_A \ni u_1 \implies u'_1 \in H$$

(знак \implies означает соответствие) и

$$H_A \ni u_2 \implies u'_2 \in H.$$

Иначе говоря, существуют такие две последовательности $u_{1,n} \in D_A$ и $u_{2,n} \in D_A$, что

$$\begin{aligned} u_{1,n} &\rightarrow u_1 \quad (\text{в норме } H_A), & u_{1,n} &\rightarrow u'_1 \quad (\text{в норме } H), \\ u_{2,n} &\rightarrow u_2 \quad (\text{в норме } H_A), & u_{2,n} &\rightarrow u'_2 \quad (\text{в норме } H). \end{aligned}$$

Тогда для любых постоянных α и β , при $n \rightarrow \infty$, имеем

$$\begin{aligned} |[(\alpha u_{1,n} + \beta u_{2,n}) - (\alpha u_1 + \beta u_2)]| &\leq |\alpha| \|u_{1,n} - u_1\| + |\beta| \|u_{2,n} - u_2\| \rightarrow 0, \\ |[(\alpha u_{1,n} + \beta u_{2,n}) - (\alpha u'_1 + \beta u'_2)]| &\leq |\alpha| \|u_{1,n} - u'_1\| + |\beta| \|u_{2,n} - u'_2\| \rightarrow 0. \end{aligned}$$

Это означает, что

$$H_A \ni \alpha u_1 + \beta u_2 \implies \alpha u'_1 + \beta u'_2 \in H.$$

Допустим, что элементам $v \in H_A$ и $w \in H_A$ соответствует один и тот же элемент $u' \in H$. Тогда, в силу линейности соответствия, элементу $z = v - w$ соответствует нулевой элемент пространства H . Докажем, что z — нулевой элемент из H_A .

Пусть z есть предел (в смысле сходимости в H_A) последовательности $u_n \in D_A$, то есть $\|u_n - z\| \rightarrow 0$ при $n \rightarrow \infty$. Так как элементу z в пространстве H соответствует нулевой элемент, то, по правилу соответствия, $\|u_n\| \rightarrow 0$. Рассмотрим произвольный элемент $f \in D_A$. По неравенству Коши-Буняковского

$$|(Af, u_n)| \leq \|Af\| \|u_n\|.$$

Поэтому $(Af, u_n) \rightarrow 0$ при $n \rightarrow \infty$. Но $(Af, u_n) = [f, u_n]$, так как $f, u_n \in D_A$. Следовательно, $[f, u_n] \rightarrow 0$ при $n \rightarrow \infty$. Переходя к пределу, по определению (1.3.4), получим

$$[f, z] = \lim_{n \rightarrow \infty} [f, u_n] = 0.$$

Таким образом, элемент $z \in H_A$ ортогонален всем элементам множества D_A , плотного в H_A . Тогда, по известному свойству (см., например, [13]), z есть нулевой элемент пространства H_A . Тем самым мы доказали, что $v = w$. \square

В заключение параграфа вернемся к установленному нами выше неравенству (1.3.3). На самом деле это неравенство выполняется не только в D_A , но и во всем пространстве H_A . Действительно, пусть $u \in H_A \setminus D_A$. Это означает, что существует последовательность $u_n \in D_A$ такая, что $\|u_n - u\| \rightarrow 0$. Тогда, как это следует из доказательства теоремы 1.3.1, $\|u_n - u\| \rightarrow 0$ (напомним, что мы отождествляем u и u'). Следовательно, $\|u_n\| \rightarrow \|u\|$ и $\|u_n\| \rightarrow \|u\|$. Так как $u_n \in D_A$, то

$$\|u_n\| \leq \frac{1}{\sqrt{\gamma}} \|u_n\|.$$

Переходя к пределу при $n \rightarrow \infty$, получим неравенство

$$\|u\| \leq \frac{1}{\sqrt{\gamma}} \|u\|, \tag{1.3.6}$$

верное для всех элементов $u \in H_A$.

1.4 Задача о минимуме функционала энергии

Рассмотрим в гильбертовом пространстве H уравнение

$$Au = f, \quad (1.4.1)$$

где A — положительно определенный оператор с областью определения D_A . Нетрудно показать, что если существует решение уравнения, то оно единственно. В самом деле, пусть уравнение (1.4.1) имеет два решения, u_1 и u_2 , так что $Au_1 = f$ и $Au_2 = f$. Тогда, в силу линейности оператора, $A(u_1 - u_2) = 0$. Умножив скалярно на $u_1 - u_2$, получим $(A(u_1 - u_2), (u_1 - u_2)) = 0$. А так как наш оператор положительно определенный, то из неравенства (1.2.3) следует, что $u_1 - u_2 = 0$, то есть $u_1 = u_2$.

Напомним определение функционала. Будем говорить, что на линейном множестве M гильбертова пространства H задан *линейный функционал* l , если каждому элементу $u \in M$ ставится в соответствие некоторое вещественное число $l(u)$, причем

$$l(\alpha u_1 + \beta u_2) = \alpha l(u_1) + \beta l(u_2)$$

для всех $u_1, u_2 \in M$ и любых вещественных чисел α, β . Линейный функционал l называется *ограниченным*, если существует такая постоянная C , что

$$|l(u)| \leq C \|u\| \quad (1.4.2)$$

для всех $u \in M$.

Рассмотрим, наряду с операторным уравнением (1.4.1), следующий функционал:

$$F(u) \equiv (Au, u) - 2(f, u), \quad u \in D_A. \quad (1.4.3)$$

Этот функционал называют *функционалом энергии* (относительно названия см. приведенное выше замечание 1.3.1).

Теорема 1.4.1 *Если уравнение (1.4.1) имеет решение, то это решение сообщает наименьшее значение функционалу (1.4.3). Обратно, элемент гильбертова пространства, реализующий минимум функционала (1.4.3), удовлетворяет уравнению (1.4.1).*

Доказательство. Пусть элемент $u_0 \in D_A$ является решением уравнения (1.4.1). Возьмем произвольный элемент $u \in D_A$ и положим $z = u - u_0$. Имеем

$$\begin{aligned} F(u) &= F(u_0 + z) = (A(u_0 + z), (u_0 + z)) - 2(f, u_0 + z) \\ &= F(u_0) + 2(Au_0 - f, z) + (Az, z) = F(u_0) + (Az, z). \end{aligned}$$

Если $u \neq u_0$, то $z \neq 0$ и, как следует из неравенства (1.2.3), $(Az, z) > 0$. Тем самым $F(u) > F(u_0)$.

Обратно, пусть элемент $u_0 \in D_A$ реализует минимум функционала $F(u)$. Возьмем произвольный элемент $z \in D_A$. Для любого вещественного числа λ , в силу предположения,

$$F(u_0 + \lambda z) \geq F(u_0).$$

Это неравенство, используя линейность оператора A , можно записать в виде

$$F(u_0) + 2\lambda(Au_0 - f, z) + \lambda^2(Az, z) \geq F(u_0),$$

откуда очевидно следует, что

$$2\lambda(Au_0 - f, z) + \lambda^2(Az, z) \geq 0$$

для любого вещественного λ . Это возможно, если $(Au_0 - f, z) = 0$. Получим, что элемент $Au_0 - f$ ортогонален всем элементам всюду плотного в H множества D_A . Тем самым $Au_0 - f = 0$, то есть $Au_0 = f$. \square

Таким образом, доказанная теорема позволяет утверждать, что задача решения уравнения (1.4.1) равносильна задаче отыскания минимума функционала (1.4.3). Задачи такого типа называются *вариационными*. Однако наше утверждение предполагает существование элемента, удовлетворяющего уравнению или реализующего минимум функционала. Оказывается, что это не всегда так. Например, рассмотрим уравнение

$$-u'' = f, \quad x \in [a, b]$$

с краевыми условиями $u(a) = u(b) = 0$. При этом пусть f — функция, терпящая разрыв в некоторой точке $c \in (a, b)$. Областью определения оператора задачи являются дважды непрерывно дифференцируемые функции, удовлетворяющие указанным краевым условиям. Ясно, что решения задачи — в обычном понимании — не существует.

Причина отсутствия решения заключается в том, что оператор A не всегда определен на достаточно широком линейном множестве D_A . Вместе с тем оказывается, что в определенных случаях задача о минимуме функционала энергии может иметь решение, если область D_A , — а с ней и сам оператор, — должным образом расширить. Иными словами, необходимо изменить постановку вариационной задачи так, чтобы можно было гарантировать существование ее решения.

Принимая во внимание вышесказанное, мы переходим к определению так называемого *обобщенного решения* задачи о минимуме функционала энергии.

Запишем функционал $F(u)$ из (1.4.3) в виде

$$F(u) = [u, u] - 2(f, u), \quad u \in D_A. \quad (1.4.4)$$

Расширим этот функционал на все энергетическое пространство H_A , то есть рассмотрим функционал

$$\Phi(u) \equiv [u, u] - 2(f, u), \quad u \in H_A. \quad (1.4.5)$$

Заметим, что используя запись (f, u) при $u \in H_A$, мы основываемся на утверждении теоремы 1.3.1. Говоря о расширении оператора, мы имеем в виду то, что $\Phi(u) = F(u)$ для $u \in D_A$.

Для фиксированного элемента $f \in H$ рассмотрим функционал

$$l(u) \equiv (f, u), \quad u \in H_A, \quad (1.4.6)$$

являющийся линейным ограниченным функционалом. Действительно, пользуясь неравенством Коши-Буняковского (1.1.2) и неравенством (1.3.6), имеем

$$|l(u)| = |(f, u)| \leq \|f\| \|u\| \leq \frac{\|f\|}{\sqrt{\gamma}} |[u]|.$$

По известной теореме Рисса⁴ существует единственный элемент $u_0 \in H_A$ — такой, что для любого элемента $u \in H_A$ справедливо равенство

$$l(u) = [u_0, u]. \quad (1.4.7)$$

Теперь функционал $\Phi(u)$ из (1.4.5) можно представить в виде

$$\Phi(u) = [u, u] - 2[u_0, u], \quad u \in H_A. \quad (1.4.8)$$

Таким образом, вместо нахождения минимума функционала $F(u)$ из (1.4.4) на множестве D_A мы будем искать минимум функционала $\Phi(u)$ из (1.4.8) в энергетическом пространстве H_A . Как увидим ниже, видоизмененная таким образом вариационная задача решается достаточно просто. Для этого преобразуем выражение в правой части (1.4.8) следующим образом:

$$\begin{aligned} \Phi(u) &= [u, u] - 2[u_0, u] + [u_0, u_0] - [u_0, u_0] \\ &= [u - u_0, u - u_0] - [u_0, u_0] = |[u - u_0]|^2 - |[u_0]|^2, \end{aligned}$$

то есть получим

$$\Phi(u) = |[u - u_0]|^2 - |[u_0]|^2, \quad u \in H_A. \quad (1.4.9)$$

Отсюда ясно, что минимум функционала $\Phi(u)$ достигается при $u = u_0$, а именно:

$$\Phi_{min} = \Phi(u_0) = -|[u_0]|^2. \quad (1.4.10)$$

Элемент $u_0 \in H_A$ называется *обобщенным решением* операторного уравнения (1.4.1). Заметим, что если обобщенное решение принадлежит D_A , то оно является *классическим решением* уравнения (1.4.1).

⁴Рисс Фридьеш (1880–1956) — венгерский математик, внес фундаментальный вклад в развитие функционального анализа.

Приведем полезное для дальнейших выкладок тождество, которому удовлетворяет обобщенное решение u_0 . Из (1.4.6) и (1.4.7) очевидно следует, что

$$[u_0, u] = (f, u) \quad \forall u \in H_A. \quad (1.4.11)$$

В связи с рассматриваемой нами вариационной задачей введем одно важное понятие. Определив минимальное значение (1.4.10) функционала $\Phi(u)$, назовем последовательность элементов $u_n \in H_A$ *минимизирующей*, если

$$\lim_{n \rightarrow \infty} \Phi(u_n) = -|[u_0]|^2. \quad (1.4.12)$$

Справедливо следующее легко доказываемое утверждение.

Теорема 1.4.2 *Минимизирующая последовательность сходится к обобщенному решению уравнения (1.4.1) — как в энергетической норме, так и в метрике исходного гильбертова пространства.*

Доказательство. Пусть $u_n \in H_A$ есть произвольная минимизирующая последовательность. Из (1.4.9) имеем

$$\Phi(u_n) = |[u_n - u_0]|^2 - |[u_0]|^2.$$

Тогда, в силу (1.4.12), получим, что $|[u_n - u_0]| \rightarrow 0$ при $n \rightarrow \infty$. Сходимость в норме пространства H следует из неравенства (1.3.6). \square

Эта теорема лежит в основе многих приближенных методов решения операторных уравнений (1.4.1). А именно, чтобы решить уравнение (в обобщенном смысле), достаточно построить минимизирующую последовательность для функционала $\Phi(u)$. Любой член этой последовательности можно рассматривать в качестве приближенного решения данного уравнения.

1.5 Дифференциальные операторы краевых задач

Задачи математической физики сводятся к решению дифференциальных уравнений (обычно к уравнениям с частными производными, реже — к обыкновенным дифференциальным уравнениям) с дополнительными условиями на границе рассматриваемой области. Эти условия называются *краевыми* или *граничными* условиями. Мы можем применить к решению уравнений описанный в предыдущем параграфе энергетический метод, если установим положительную определенность оператора задачи, определяемого левой частью дифференциального уравнения и краевыми условиями. Приведем несколько примеров.

- Уравнение Штурма⁵-Лиувилля⁶

Рассмотрим уравнение

$$Lu \equiv -(p(x)u'(x))' + q(x)u(x) = f(x), \quad a < x < b \quad (1.5.1)$$

с однородными краевыми условиями

$$u(a) = u(b) = 0. \quad (1.5.2)$$

Коэффициенты уравнения $p(x)$ и $q(x)$ непрерывны на отрезке $[a, b]$, а функция $p(x)$ еще и непрерывно дифференцируема. Кроме того, предполагается, что

$$p(x) \geq p_0 = \text{const} > 0, \quad q(x) \geq 0, \quad \forall x \in [a, b]. \quad (1.5.3)$$

Областью определения D_L дифференциального оператора L являются функции, непрерывные на отрезке $[a, b]$ вместе со своими производными первого и второго порядка, удовлетворяющие краевым условиям (1.5.2). Доказано, что область D_L всюду плотна в пространстве $L_2[a, b]$ (см., например, [21]). Заметим, что частный случай дифференциального оператора ($p(x) \equiv 1, q(x) \equiv 0$) был рассмотрен в параграфе 1.2.

Пользуясь формулой интегрирования по частям, для всех $u, v \in D_L$ получим

$$(Lu, v) = \int_a^b (-(pu')'v + quv) dx = \int_a^b (pu'v' + quv) dx = (u, Lv). \quad (1.5.4)$$

Таким образом, оператор L – симметричный. Принимая во внимание условия (1.5.3), из (1.5.4) имеем

$$(Lu, u) \geq \int_a^b p(u')^2 dx \geq p_0 \int_a^b (u')^2 dx. \quad (1.5.5)$$

Далее, повторяя рассуждения, в результате которых было получено неравенство (1.2.4), можно аналогично доказать справедливость неравенства

$$\int_a^b u^2 dx \leq \frac{(b-a)^2}{2} \int_a^b (u')^2 dx. \quad (1.5.6)$$

Из (1.5.5) и (1.5.6) следует неравенство

$$(Lu, u) \geq \frac{2p_0}{(b-a)^2} \int_a^b u^2 dx = \frac{2p_0}{(b-a)^2} \|u\|^2,$$

⁵Штурм Шарль Франсуа (1803–1855) – французский математик, основные научные труды посвящены дифференциальным уравнениям и алгебраическим многочленам.

⁶Лиувиль Жозеф (1809–1882) – французский математик, известен своими работами по дифференциальным уравнениям и теории функций.

которое означает, что оператор L положительно определенный.

Описанное в параграфе 1.3 пополнение области определения D_L оператора приводит к энергетическому пространству $H_L \subset L_2[a, b]$, скалярное произведение и норма в котором определяются следующим образом:

$$[u, v] \equiv \int_a^b (p u'v' + quv) dx, \quad |[u]| \equiv \sqrt{[u, u]}. \quad (1.5.7)$$

Отыскание обобщенного решения задачи (1.5.1), (1.5.2), согласно (1.4.5), сводится к нахождению минимума функционала

$$\Phi(u) = \int_a^b (p(u')^2 + qu^2 - 2fu) dx, \quad u \in H_L. \quad (1.5.8)$$

• Уравнение Пуассона: задача Дирихле

Пусть Ω есть конечная односвязная область в плоскости координат x и y , ограниченная гладкой кривой S (определение гладкой кривой можно найти, например, в [11, 30]). Рассмотрим в области Ω уравнение

$$-\Delta u = f, \quad (1.5.9)$$

где Δ есть оператор Лапласа⁷, определяемый следующим выражением:

$$\Delta u \equiv \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}. \quad (1.5.10)$$

Уравнение (1.5.9) называется *уравнением Пуассона*⁸; однородное уравнение Пуассона (когда $f(x) \equiv 0$) называют также *уравнением Лапласа*.

*Задача Дирихле*⁹ для уравнения Пуассона заключается в нахождении функции $u(x, y)$, удовлетворяющей в области Ω уравнению (1.5.9), а на границе области — краевому условию

$$u(x, y) = \varphi(x, y), \quad (x, y) \in S. \quad (1.5.11)$$

Здесь мы ограничимся рассмотрением задачи Дирихле с однородным краевым условием, то есть когда $\varphi(x, y) \equiv 0$. Будем рассматривать оператор Лапласа как оператор,

⁷Лаплас Пьер Симон (1749–1827) — французский математик, физик и астроном, известен работами в области дифференциальных уравнений и небесной механики, один из создателей теории вероятностей.

⁸Пуассон Симеон Дени (1781–1840) — французский математик, механик и физик, основные научные труды относятся к математической физике и небесной механике.

⁹Дирихле Петер Густав (1805–1859) — немецкий математик, внесший существенный вклад в математический анализ, теорию функций и теорию чисел.

действующий в пространстве $L_2(\Omega)$ (см., например, [13, 14]). Напомним, что скалярное произведение и норма в этом пространстве определяются следующим образом:

$$(u, v) \equiv \int_{\Omega} uv \, d\Omega, \quad \|u\| \equiv \left(\int_{\Omega} u^2 \, d\Omega \right)^{1/2}.$$

За область определения D_L оператора $L \equiv -\Delta$ задачи Дирихле примем множество функций, непрерывных вместе со своими первыми и вторыми производными в замкнутой области $\bar{\Omega} = \Omega \cup S$ и равных нулю на границе S . Доказывается, что область D_L всюду плотна в пространстве $L_2(\Omega)$ (см., например, [21]).

Для последующего изложения материала нам понадобятся некоторые понятия и обозначения. *Градиентом* функции $u(x, y)$ в точке (x, y) называется вектор

$$\nabla u \equiv \begin{bmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial u}{\partial y} \end{bmatrix}. \quad (1.5.12)$$

Запись $\nabla u \nabla v$ означает скалярное произведение векторов ∇u и ∇v , а именно,

$$\nabla u \nabla v = \frac{\partial u}{\partial x} \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \frac{\partial v}{\partial y}. \quad (1.5.13)$$

При этом

$$|\nabla u| = \sqrt{\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2} \quad (1.5.14)$$

есть, фактически, длина вектора.

Пусть \mathbf{n} – единичная внешняя нормаль к кривой S . Выражение

$$\frac{\partial u}{\partial \mathbf{n}} = \frac{\partial u}{\partial x} \cos(\mathbf{n}, x) + \frac{\partial u}{\partial y} \cos(\mathbf{n}, y) \quad (1.5.15)$$

определяет производную функции $u(x, y)$ по направлению \mathbf{n} (см., например, [11, 30]).

Для функции двух переменных $u(x_1, x_2)$ приведем формулу интегрирования по частям

$$\int_{\Omega} \frac{\partial u}{\partial x_i} v \, d\Omega = - \int_{\Omega} u \frac{\partial v}{\partial x_i} \, d\Omega + \int_S uv \cos(\mathbf{n}, x_i) \, dS, \quad i = 1, 2, \quad (1.5.16)$$

которую мы будем неоднократно использовать в дальнейшем.

Итак, возвратимся к нашей задаче Дирихле. Для любых функций $u, v \in D_L$, пользуясь формулой (1.5.16), имеем

$$(Lu, v) = - \int_{\Omega} \Delta u \cdot v \, d\Omega = - \int_{\Omega} \frac{\partial^2 u}{\partial x^2} v \, d\Omega - \int_{\Omega} \frac{\partial^2 u}{\partial y^2} v \, d\Omega = \\ \int_{\Omega} \frac{\partial u}{\partial x} \frac{\partial v}{\partial x} \, d\Omega - \int_S \frac{\partial u}{\partial x} v \cos(\mathbf{n}, x) \, dS + \int_{\Omega} \frac{\partial u}{\partial y} \frac{\partial v}{\partial y} \, d\Omega - \int_S \frac{\partial u}{\partial y} v \cos(\mathbf{n}, y) \, dS.$$

Отсюда, принимая во внимание (1.5.13) и (1.5.15), получим

$$(Lu, v) = \int_{\Omega} \nabla u \nabla v \, d\Omega - \int_S \frac{\partial u}{\partial \mathbf{n}} v \, dS. \quad (1.5.17)$$

А так как на границе S функция v обращается в нуль, то

$$(Lu, v) = \int_{\Omega} \nabla u \nabla v \, d\Omega = (u, Lv). \quad (1.5.18)$$

Таким образом, наш оператор L является симметричным.

Положительная определенность оператора следует из неравенства

$$\int_{\Omega} u^2 \, d\Omega \leq C \int_{\Omega} |\nabla u|^2 \, d\Omega, \quad (1.5.19)$$

где $C = const > 0$, справедливое для функций, удовлетворяющих нулевому граничному условию (см., например, [13, 20]). Это неравенство называется *неравенством Фридриха*¹⁰. Применяя это неравенство, из (1.5.18) получим

$$(Lu, u) \geq \int_{\Omega} |\nabla u|^2 \, d\Omega \geq \frac{1}{C} \|u\|^2. \quad (1.5.20)$$

Пополнение области определения D_L оператора L приводит к энергетическому пространству $H_L \subset L_2(\Omega)$ со скалярным произведением и нормой

$$[u, v] \equiv \int_{\Omega} \nabla u \nabla v \, d\Omega, \quad \|u\| \equiv \sqrt{[u, u]}. \quad (1.5.21)$$

Нахождение обобщенного решения уравнения (1.5.9) с однородным граничным условием, согласно (1.4.5), сводится к отысканию минимума функционала

$$\Phi(u) = \int_{\Omega} (|\nabla u|^2 - 2fu) \, d\Omega, \quad u \in H_L. \quad (1.5.22)$$

¹⁰Фридрихс Курт Отто (1901–1982) – немецкий и американский математик, основные научные труды относятся к теории уравнений с частными производными.

• **Уравнение Пуассона: третья краевая задача**

Рассмотрим теперь уравнение Пуассона (1.5.9) с так называемым *третьим краевым условием*

$$\left. \left(\frac{\partial u}{\partial \mathbf{n}} + \sigma u \right) \right|_S = 0, \quad (1.5.23)$$

где $\sigma(x, y)$ — заданная на кривой S ограниченная функция, причем $\sigma(x, y) \geq \sigma_0 > 0$.

Областью определения D_L оператора $L \equiv -\Delta$ рассматриваемой нами задачи является множество функций, непрерывных вместе со своими первыми и вторыми производными в замкнутой области $\bar{\Omega} = \Omega \cup S$ и удовлетворяющих краевому условию (1.5.23).

Симметричность оператора L легко следует из равенства (1.5.17). Действительно, принимая во внимание краевое условие (1.5.23), имеем

$$(Lu, v) = \int_{\Omega} \nabla u \nabla v \, d\Omega + \int_S \sigma uv \, dS = (u, Lv). \quad (1.5.24)$$

Положительная определенность оператора доказывается также несложно, если воспользоваться неравенством

$$\int_{\Omega} u^2 \, d\Omega \leq C \left(\int_{\Omega} |\nabla u|^2 \, d\Omega + \int_S u^2 \, dS \right), \quad (1.5.25)$$

где C есть положительная постоянная, также приписываемое Фридрихсу (см., например, [20]). Применяя это неравенство, из (1.5.24) получим

$$\begin{aligned} (Lu, u) &= \int_{\Omega} |\nabla u|^2 \, d\Omega + \int_S \sigma u^2 \, dS \geq \int_{\Omega} |\nabla u|^2 \, d\Omega + \sigma_0 \int_S u^2 \, dS \\ &\geq \min(1, \sigma_0) \left(\int_{\Omega} |\nabla u|^2 \, d\Omega + \int_S u^2 \, dS \right) \geq \frac{\min(1, \sigma_0)}{C} \|u\|^2. \end{aligned}$$

Пополнение области определения D_L оператора L приводит к энергетическому пространству $H_L \subset L_2(\Omega)$, скалярное произведение и норма в котором определяются следующим образом:

$$[u, v] \equiv \int_{\Omega} \nabla u \nabla v \, d\Omega + \int_S \sigma uv \, dS, \quad |[u]| \equiv \sqrt{[u, u]}.$$

Таким образом, задача сводится к нахождению минимума функционала

$$\Phi(u) = \int_{\Omega} (|\nabla u|^2 - 2fu) \, d\Omega + \int_S \sigma u^2 \, dS, \quad u \in H_L.$$

1.6 Метод Ритца

Как было сказано в параграфе 1.4, задача нахождения обобщенного решения уравнения

$$Au = f \quad (1.6.1)$$

с положительно определенным оператором A , действующим в гильбертовом пространстве H , сводится к задаче построения минимизирующей последовательности для функционала

$$\Phi(u) = [u, u] - 2(f, u), \quad u \in H_A. \quad (1.6.2)$$

Одним из методов построения минимизирующей последовательности является *метод Ритца*¹¹, который мы изложим ниже.

Определим в энергетическом пространстве H_A последовательность элементов

$$\varphi_1, \varphi_2, \dots, \varphi_n, \dots, \quad (1.6.3)$$

удовлетворяющих следующим условиям:

1. при любом n элементы $\varphi_1, \varphi_2, \dots, \varphi_n$ линейно независимы;
2. последовательность (1.6.3) полна в пространстве H_A ; полнота последовательности понимается в следующем смысле: для любого элемента $u \in H_A$ и любого $\varepsilon > 0$ найдутся такое натуральное число $n = n(u, \varepsilon)$ и такой набор коэффициентов $\alpha_1, \alpha_2, \dots, \alpha_n$, что

$$\left\| u - \sum_{i=1}^n \alpha_i \varphi_i \right\| < \varepsilon.$$

Элементы φ_i будем называть *базисными*.

Для всех значений $n = 1, 2, \dots$ образуем линейную оболочку первых n элементов последовательности (1.6.3), то есть

$$H_n \equiv \text{span}\{\varphi_1, \varphi_2, \dots, \varphi_n\}.$$

Тем самым мы имеем последовательность вложенных подпространств

$$H_1 \subset H_2 \subset \dots \subset H_n \subset \dots \subset H_A. \quad (1.6.4)$$

Поставим следующую задачу: *найти элемент $v_n \in H_n$, реализующий минимум функционала $\Phi(u)$ в подпространстве H_n , то есть такой, что*

$$\Phi(v_n) = \min_{u \in H_n} \Phi(u). \quad (1.6.5)$$

Любой элемент $u_n \in H_n$ представим в виде разложения по базисным элементам:

$$u_n = \sum_{i=1}^n a_i \varphi_i.$$

¹¹Ритц Вальтер (1878–1909) — швейцарский и немецкий физик и математик, основные работы относятся к решению вариационных задач математической физики.

Подставив такую запись элемента в выражение (1.6.2) функционала $\Phi(u)$, получим

$$\begin{aligned}\Phi(u_n) &= [u_n, u_n] - 2(f, u_n) = \left[\sum_{j=1}^n a_j \varphi_j, \sum_{i=1}^n a_i \varphi_i \right] - 2 \left(f, \sum_{i=1}^n a_i \varphi_i \right) \\ &= \sum_{i,j=1}^n [\varphi_j, \varphi_i] a_j a_i - 2 \sum_{i=1}^n (f, \varphi_i) a_i.\end{aligned}\tag{1.6.6}$$

Итак, ограничение нашего функционала $\Phi(u)$ на подпространство H_n приводит к функции n независимых переменных a_1, a_2, \dots, a_n . Как известно, необходимыми условиями минимума являются

$$\frac{\partial \Phi(u_n)}{\partial a_i} = 0, \quad i = 1, 2, \dots, n$$

(см., например, [11, 30]). Эти частные производные легко вычисляются:

$$\frac{\partial \Phi(u_n)}{\partial a_i} = 2 \sum_{j=1}^n [\varphi_j, \varphi_i] a_j - 2(f, \varphi_i).$$

Приравнявая их к нулю, получим систему линейных алгебраических уравнений

$$\sum_{j=1}^n [\varphi_j, \varphi_i] a_j = (f, \varphi_i), \quad i = 1, 2, \dots, n,\tag{1.6.7}$$

называемую *системой уравнений Ритца*. Матрица системы

$$G = \left[[\varphi_j, \varphi_i] \right]_{i,j=1}^n\tag{1.6.8}$$

есть *матрица Грама*¹² линейно независимой системы элементов $\varphi_1, \varphi_2, \dots, \varphi_n$; известно, что эта матрица положительно определенная (см., например, [2, 4]). Поэтому система (1.6.7) однозначно разрешима. Если ввести в рассмотрение вектор коэффициентов и вектор правых частей, а именно:

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \quad r = \begin{bmatrix} (f, \varphi_1) \\ (f, \varphi_2) \\ \vdots \\ (f, \varphi_n) \end{bmatrix},\tag{1.6.9}$$

то систему (1.6.7) можно записать в компактном виде

$$Ga = r.\tag{1.6.10}$$

¹²Грама Йорген Педерсен (1850–1916) — датский математик, основные труды относятся к теории приближения функций, теории чисел и математической статистике.

Пусть $a_1^*, a_2^*, \dots, a_n^*$ — решение системы (1.6.7). Тогда элемент

$$v_n = \sum_{i=1}^n a_i^* \varphi_i \quad (1.6.11)$$

и будет решением вариационной задачи (1.6.5). В этом нетрудно убедиться. Действительно, как следует из записей (1.6.6), (1.6.8) и (1.6.9), для любого элемента $u_n \in H_n$ верно представление

$$\Phi(u_n) = (Ga, a) - 2(r, a).$$

Имеем

$$\Phi(u_n) - \Phi(v_n) = (Ga, a) - 2(r, a) - (Ga^*, a^*) + 2(r, a^*),$$

где a^* — вектор коэффициентов разложения (1.6.11). Так как a^* является решением системы (1.6.10), то

$$\Phi(u_n) - \Phi(v_n) = (Ga, a) - 2(Ga^*, a) + (Ga^*, a^*) = (G(a - a^*), (a - a^*)).$$

Поэтому, если $u_n \neq v_n$ (или, что то же самое, $a \neq a^*$), то в силу положительной определенности матрицы Грама $(G(a - a^*), (a - a^*)) > 0$; тем самым $\Phi(u_n) > \Phi(v_n)$.

Заметим, что если базисные элементы выбраны из области определения D_A оператора A , то согласно (1.3.1) можем записать $[\varphi_j, \varphi_i] = (A\varphi_j, \varphi_i)$, и система Ритца (1.6.7) примет вид

$$\sum_{j=1}^n (A\varphi_j, \varphi_i) a_j = (f, \varphi_i), \quad i = 1, 2, \dots, n. \quad (1.6.12)$$

Имея последовательность (1.6.4) вложенных подпространств H_n и решая в каждом из них задачу (1.6.5), мы приходим к последовательности элементов v_n , $n = 1, 2, \dots$. Очевидно, что

$$\Phi(v_1) \geq \Phi(v_2) \geq \dots \geq \Phi(v_n) \geq \dots \geq \Phi(u_0),$$

где $u_0 \in H_A$ есть обобщенное решение операторного уравнения (1.6.1).

Докажем, что построенная нами последовательность элементов $v_n \in H_A$ является минимизирующей.

Ортогонализируя систему базисных элементов (1.6.3), перейдем к ортонормированной системе $\{\psi_i\}_{i=1}^{\infty}$, то есть такой, что

$$[\psi_i, \psi_j] = \begin{cases} 1, & \text{если } i = j, \\ 0, & \text{если } i \neq j. \end{cases} \quad (1.6.13)$$

Для этого можно применить хорошо известный процесс ортогонализации Грама-Шмидта¹³ (см., например, [2, 6, 14]). Полнота системы $\{\psi_i\}_{i=1}^{\infty}$ вытекает из полноты

¹³Шмидт Эрхард (1876–1959) — немецкий математик, основные работы относятся к теории функций, интегральным уравнениям и функциональному анализу.

исходной системы (1.6.3). При этом из принципа построения процесса Грама-Шмидта следует, что

$$H_n \equiv \text{span}\{\psi_1, \psi_2, \dots, \psi_n\}.$$

Разложим обобщенное решение $u_0 \in H_A$ по элементам нового базиса:

$$u_0 = \sum_{i=1}^{\infty} \alpha_i \psi_i,$$

где $\alpha_i = [u_0, \psi_i]$, что легко получить из (1.6.13). Положим

$$z_n = \sum_{i=1}^n \alpha_i \psi_i, \quad n = 1, 2, \dots$$

Так как $z_n \in H_n$ и $\lim_{n \rightarrow \infty} z_n = u_0$, то $\Phi(z_n) \rightarrow \Phi(u_0)$ при $n \rightarrow \infty$. Имеем

$$\Phi(v_n) = \min_{u \in H_n} \Phi(u), \quad \Phi(u_0) = \min_{u \in H_A} \Phi(u);$$

следовательно,

$$\Phi(u_0) \leq \Phi(v_n) \leq \Phi(z_n)$$

и, тем самым, $\Phi(v_n) \rightarrow \Phi(u_0)$ при $n \rightarrow \infty$. Таким образом, мы показали, что последовательность v_n минимизирующая.

Итак, на основании доказанной выше теоремы 1.4.2 мы можем сформулировать следующее утверждение.

Теорема 1.6.1 *Построенная по методу Ритца последовательность элементов v_n , $n = 1, 2, \dots$ сходится к обобщенному решению уравнения (1.6.1) как в энергетической норме, так и в норме исходного гильбертова пространства.*

Получим одно полезное тождество, которому удовлетворяет приближенное решение $v_n \in H_n$ уравнения (1.6.1). Так как коэффициенты разложения (1.6.11) элемента v_n по базисным элементам являются решением системы Ритца (1.6.7), то

$$\sum_{j=1}^n [\varphi_j, \varphi_i] a_j^* = (f, \varphi_i), \quad i = 1, 2, \dots, n.$$

Отсюда следует, что

$$[v_n, \varphi_i] = (f, \varphi_i), \quad i = 1, 2, \dots, n.$$

Умножая обе части этих равенств на произвольные числа α_i и суммируя по всем i от 1 до n , получим тождество

$$[v_n, u_n] = (f, u_n) \quad \forall u_n \in H_n. \quad (1.6.14)$$

В заключение сформулируем и докажем утверждение, которое согласуется как с самой постановкой задачи о минимуме функционала энергии, так и с идеей метода Ритца ее решения.

Теорема 1.6.2 Пусть $u_0 \in H_A$ есть обобщенное решение операторного уравнения (1.6.1), а $v_n \in H_n$ – приближенное решение, полученное методом Рунца. Тогда

$$|[u_0 - v_n]| = \min_{u_n \in H_n} |[u_0 - u_n]|. \quad (1.6.15)$$

Доказательство. Как очевидно следует из (1.4.11),

$$[u_0, u_n] = (f, u_n) \quad \forall u_n \in H_n.$$

Имея еще и тождество (1.6.14), получим, что $[u_0 - v_n, u_n] = 0$ для любого элемента $u_n \in H_n$. В частности,

$$[u_0 - v_n, u_n - v_n] = 0.$$

Преобразуем это равенство, записав его в виде

$$[u_0 - v_n, u_0 - v_n] = [u_0 - v_n, u_0 - u_n]$$

или

$$|[u_0 - v_n]|^2 = [u_0 - v_n, u_0 - u_n].$$

Для оценки правой части равенства воспользуемся неравенством Коши-Буняковского – применительно к энергетической норме:

$$|[u_0 - v_n]|^2 \leq |[u_0 - v_n]| |[u_0 - u_n]|.$$

В итоге получим, что

$$|[u_0 - v_n]| \leq |[u_0 - u_n]|$$

для любого элемента $u_n \in H_n$, откуда и следует равенство (1.6.15). \square

Доказанная нами теорема, как мы увидим в следующей главе, указывает путь получения оценки разности между обобщенным и приближенным решениями операторного уравнения (1.6.1).

1.7 Метод Бубнова-Галеркина

Основным недостатком метода Рунца является то, что он применим лишь для уравнений с симметричными положительно определенными операторами. От этого недостатка свободен другой вычислительный алгоритм, называемый *методом Бубнова¹⁴-Галеркина¹⁵* (его часто называют также *методом Галеркина*).

¹⁴Бубнов Иван Григорьевич (1872–1919) – российский корабельный инженер, математик и механик.

¹⁵Галеркин Борис Григорьевич (1871–1945) – российский механик и математик, известный своими работами в области теории упругости.

Изложим основную идею метода и дадим его краткое описание. Рассмотрим уравнение

$$Au = f \quad (1.7.1)$$

с линейным непрерывным оператором A , действующим в некотором гильбертовом пространстве H . Как было уже сказано в параграфе 1.2, предполагается, что область определения оператора D_A всюду плотна в H .

Зададим в множестве D_A бесконечную последовательность *базисных* элементов

$$\varphi_1, \varphi_2, \dots, \varphi_n, \dots, \quad (1.7.2)$$

удовлетворяющих следующим условиям:

1. при любом n элементы $\varphi_1, \varphi_2, \dots, \varphi_n$ линейно независимы;
2. последовательность (1.7.2) полна в пространстве H в том смысле, что для любого элемента $u \in H$ и любого $\varepsilon > 0$ найдутся такое натуральное число $n = n(u, \varepsilon)$ и такой набор коэффициентов $\alpha_1, \alpha_2, \dots, \alpha_n$, что $\|u - \sum_{i=1}^n \alpha_i \varphi_i\| < \varepsilon$.

Для каждого значения n приближенное решение уравнения (1.7.1) ищется в виде

$$u_n = \sum_{i=1}^n a_i \varphi_i, \quad (1.7.3)$$

где коэффициенты a_1, a_2, \dots, a_n определяются из условия ортогональности так называемой *невязки* $Au_n - f$ и базисных функций:

$$(Au_n - f, \varphi_i) = 0, \quad i = 1, 2, \dots, n. \quad (1.7.4)$$

Эти условия приводят к системе линейных алгебраических уравнений

$$\sum_{j=1}^n (A\varphi_j, \varphi_i) a_j = (f, \varphi_i), \quad i = 1, 2, \dots, n, \quad (1.7.5)$$

называемой *системой уравнений Бубнова-Галеркина*.

Итак, основная идея метода заключается в следующем: в силу плотности множества D_A в пространстве H , в пределе (при $n \rightarrow \infty$), как следует из (1.7.4), элемент $Au_\infty - f$ обращается в нуль.

Отметим одно важное обстоятельство. Уравнения Бубнова-Галеркина (1.7.5) по форме совпадают с соответствующими уравнениями (1.6.12) метода Рунца, когда $\varphi_i \in D_A$. Таким образом, если оператор A положительно определен, то методы Бубнова-Галеркина и Рунца совпадают.

Метод Рунца, как было сказано в параграфе 1.4, является методом *вариационного* типа. Метод же Бубнова-Галеркина не связан непосредственно с минимизацией какого-либо функционала. Он основан на проектировании невязки $Au_n - f$ на каждый из базисных

элементов φ_i и приравнявании результата к нулю. Тем самым метод Бубнова-Галеркина относят к классу так называемых *проекционных* методов.

Здесь мы не будем заниматься обоснованием метода в общем случае. В следующей главе — при рассмотрении краевых задач для дифференциальных уравнений — этот вопрос будет достаточно подробно исследован.

Глава 2

МЕТОД КОНЕЧНЫХ ЭЛЕМЕНТОВ ДЛЯ УРАВНЕНИЯ ПУАССОНА

В настоящей главе обсуждаются основные принципы, методы и техника построения систем конечноэлементных уравнений для двумерных эллиптических уравнений второго порядка. Значительное место отводится выводу оценок погрешности приближенного решения, основанных на соответствующих теоремах аппроксимации. Преследуя цель ознакомить читателя с основами метода конечных элементов — без привлечения излишних при таком подходе деталей чисто технического характера — изложение материала ведется на примере уравнения Пуассона. Однако ход рассуждений и практически все полученные результаты могут быть по аналогии перенесены на эллиптические уравнения общего вида.

2.1 О методе конечных элементов

Одним из наиболее распространенных численных методов решения линейных дифференциальных уравнений с частными производными является *метод конечных разностей*. В основе этого метода лежит идея замены производных, входящих в уравнение, конечными разностями на выбранной сетке узлов. В результате дискретизации задача сводится к решению системы линейных алгебраических уравнений — *системы сеточных уравнений*, неизвестные которой трактуются как приближенные значения искомого решения в узлах сетки. Метод конечных разностей обладает одним важным достоинством, которое заключается в том, что матрица системы сеточных уравнений является *разреженной*, то есть содержит большое количество нулей, число которых сопоставимо с порядком матрицы. Метод хорошо зарекомендовал себя при решении краевых задач в областях "правильной" формы, например, в областях прямоугольного типа. Однако в областях произвольной геометрической формы применение метода сопряжено с определенными трудностями, связанными как с записью разностных уравнений вблизи границы области, так и с аппроксимацией самих краевых условий. Могут быть нарушены

(как правило, это и происходит!) такие важные характеристики оператора исходной краевой задачи, как его симметричность и положительная определенность — матрица системы сеточных уравнений очень часто не наследует эти свойства. А это влечет за собой ряд трудностей при численном решении систем сеточных уравнений.

В начале XX-го столетия появились новые методы численного решения дифференциальных уравнений с частными производными, основанные на *вариационных* и *проеекционных* принципах: это методы Рунца и Бубнова-Галеркина, о которых говорилось в предыдущей главе. Привлекательность этих методов заключалась, прежде всего, в том, что они сохраняли важные свойства оператора рассматриваемой краевой задачи — симметричность и положительную определенность. Вместе с тем, возникли проблемы другого рода. Вначале в вариационных и проекционных методах в качестве базисных функций использовались алгебраические и тригонометрические многочлены. При этом матрицы получающихся систем линейных алгебраических уравнений оказывались практически "забитыми" ненулевыми элементами. Было утрачено — по сравнению с методом конечных разностей — такое важное свойство, как разреженность матрицы системы. Это обстоятельство затрудняло применение вариационных и проекционных методов при большом числе базисных функций.

Во второй половине XX-го столетия отношение к проекционным методам в корне изменилось. Началось все с того, что в 1943 году была опубликована статья Р. Куранта¹[41], в которой автор предлагал использовать в методе Рунца специальные базисные функции. Кратко поясним, в чем заключалась основная специфика этих функций. Если бы базисные функции φ_i метода Рунца или метода Бубнова-Галеркина были "ортогональны" в том смысле, что $(A\varphi_j, \varphi_i) = 0$ при $i \neq j$, то мы пришли бы к системам (1.6.12) и (1.7.5) с диагональными матрицами. В действительности построение таких функций нереально. Однако можно брать базисные функции, которые отличны от нуля лишь в сравнительно малой области изменения аргументов. Такие функции получили название *финитных функций* или *функций с локальными носителями*. Это приведет к тому, что каждая базисная функция окажется "ортогональной" почти всем остальным, за исключением конечного их числа, не зависящего от общего количества функций. В результате матрица системы линейных алгебраических уравнений станет разреженной.

Идея Куранта не была по достоинству воспринята в течение долгого времени, вплоть до 1964 года, когда появилась статья Л.А. Оганесяна²[22], в которой были заложены основы нового направления в численных методах решения дифференциальных уравнений с частными производными — *метода конечных элементов*. Эта статья стала первой работой, где была дана современная формулировка метода и получены первые результаты по его математическому обоснованию.

Метод конечных элементов является, по существу, сеточным методом, в котором

¹ Курант Рихард (1888–1972) — немецкий и американский математик польского происхождения, известен своими работами в области дифференциальных уравнений с частными производными.

² Оганесян Леонард Амаякович (1925–2013) — советский математик, крупный специалист по вычислительной математике и ее приложениям, один из создателей метода конечных элементов.

дискретизация краевой задачи производится на основе вариационных или проекционных методов с использованием специальных базисных функций с локальными носителями, представляющими собой объединение небольшого числа ячеек сетки. Обычно в качестве базисных берутся кусочно-полиномиальные функции, которые на каждой ячейке сетки являются интерполяционными многочленами. Таким образом, метод конечных элементов есть результат синтеза метода конечных разностей и метода Бубнова-Галеркина.

Вместе со ставшим теперь общепринятым названием метода употребляются также термины *вариационно-разностный*, *проекционно-сеточный* и др. Следует заметить, что название *метод конечных элементов* происходит, по всей видимости, от применяемого инженерами подхода к решению задач строительной механики и теории упругости, основанного на разбиении конструкции или тела на подструктуры (элементы) и установлению взаимодействующих связей между ними.

Метод конечных элементов получил широкое распространение на практике как эффективное средство решения многих стационарных задач и, в частности, эллиптических уравнений. Применяется он и при решении нестационарных уравнений. Ему посвящена обширная научная литература – огромное число статей, множество учебников и монографий.

2.2 Двухточечная краевая задача

Рассмотрим обыкновенное дифференциальное уравнение

$$-u''(x) + qu(x) = f(x), \quad a < x < b, \quad (2.2.1)$$

где $q = \text{const} \geq 0$, с однородными граничными условиями

$$u(a) = u(b) = 0. \quad (2.2.2)$$

Предположим, что правая часть уравнения (2.2.1) есть непрерывная функция. В курсе обыкновенных дифференциальных уравнений доказывается, что решение поставленной задачи существует и единственно (см., например, [24, 32]). Говоря о решении, мы имеем в виду решение из пространства $C^2[a, b]$ – пространства дважды непрерывно дифференцируемых на отрезке $[a, b]$ функций. Уравнение имеет определенный физический смысл – оно описывает, например, стационарное распределение температуры в однородном стержне. Заметим также, что уравнение (2.2.1) является частным случаем рассмотренного нами выше уравнения Штурма-Лиувилля (1.5.1), когда $p(x) \equiv 1$ и $q(x) \equiv q \geq 0$.

Краевая задача (2.2.1), (2.2.2) является простой, но, вместе с тем, и достаточно содержательной одномерной моделью более сложных – в смысле численного решения – двумерных эллиптических уравнений, на примере которой мы постараемся представить основные аспекты метода конечных элементов.

Для решения задачи (2.2.1), (2.2.2) мы применим описанный в параграфе 1.6 метод Рунге со специальными базисными функциями. С учетом сделанного выше замечания

относительно того, что уравнение (2.2.1) есть частный случай уравнения Штурма-Лиувилля, скалярное произведение и норма в энергетическом пространстве H_L оператора $Lu \equiv -u'' + qu$ задачи, согласно (1.5.7), определяются следующим образом:

$$[u, v] \equiv \int_a^b (u'v' + quv) dx, \quad \|[u]\| \equiv \sqrt{[u, u]}. \quad (2.2.3)$$

Введем на отрезке $[a, b]$ сетку узлов

$$x_i = a + ih, \quad i = 0, 1, \dots, n+1, \quad (2.2.4)$$

где $h = (b - a)/(n + 1)$ есть шаг сетки. Узлы x_i , $i = 1, 2, \dots, n$ назовем *внутренними*; узлы $x_0 = a$ и $x_{n+1} = b$ — *граничные узлы*. Каждому внутреннему узлу x_i поставим в соответствие *базисную* непрерывную кусочно-линейную функцию $\varphi_i(x)$, равную единице в узле x_i и нулю во всех остальных узлах сетки (Рис. 2.2.1).

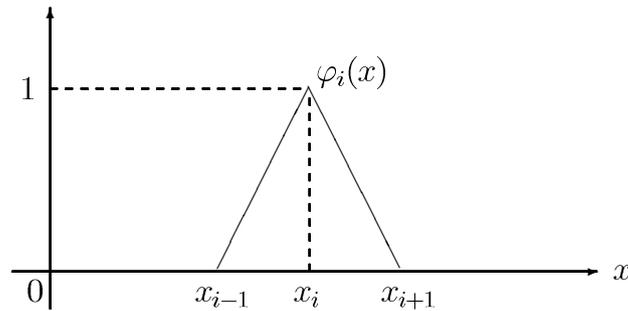


Рис. 2.2.1. Базисная функция.

Легко проверить, что

$$\varphi_i(x) = \begin{cases} \frac{x - x_{i-1}}{h}, & \text{если } x \in [x_{i-1}, x_i], \\ \frac{x_{i+1} - x}{h}, & \text{если } x \in [x_i, x_{i+1}], \\ 0, & \text{если } x \notin [x_{i-1}, x_{i+1}]. \end{cases} \quad (2.2.5)$$

Каждая из функций $\varphi_i(x)$ отлична от нуля лишь на отрезке $[x_{i-1}, x_{i+1}]$ длины $2h$. Совокупность базисных функций $\varphi_i(x)$, $i = 1, 2, \dots, n$ образует линейно независимую систему; это следует из определения самих функций, а именно, из того, что

$$\varphi_i(x_j) = \delta_{ij}, \quad i, j = 1, 2, \dots, n,$$

где δ_{ij} есть символ Кронекера ($\delta_{ij} = 0$, если $i \neq j$, и $\delta_{ii} = 1$).

Приближенное решение задачи (2.2.1), (2.2.2) будем искать в виде линейной комбинации введенных базисных функций, то есть как элемент подпространства

$$H_n \equiv \text{span}\{\varphi_1, \varphi_2, \dots, \varphi_n\} \quad (2.2.6)$$

энергетического пространства H_L . Функции из H_n (мы будем обозначать их с "шапочкой") представимы в виде

$$\hat{u}(x) = \sum_{i=1}^n u_i \varphi_i(x). \quad (2.2.7)$$

Очевидно, что $\hat{u}(x_i) = u_i$, $i = 1, 2, \dots, n$ и $\hat{u}(a) = \hat{u}(b) = 0$. Фактически, непрерывная кусочно-линейная функция $\hat{u}(x)$ есть результат линейной интерполяции по значениям u_1, u_2, \dots, u_n во внутренних узлах сетки (Рис. 2.2.2).

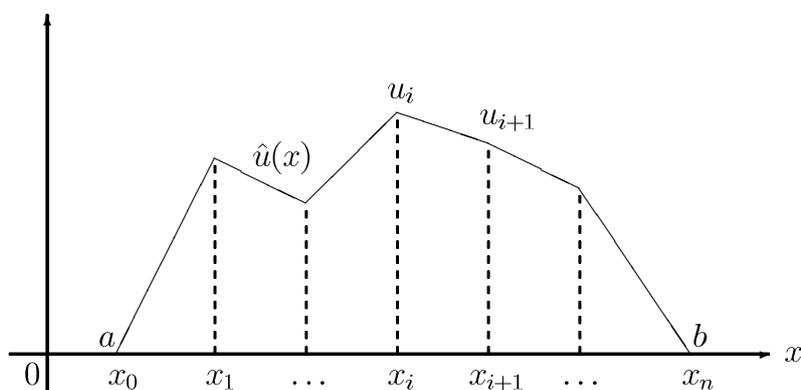


Рис. 2.2.2. Кусочно-линейная функция.

Согласно методу Рунге, для нахождения приближенного решения задачи нам нужно решить систему линейных алгебраических уравнений

$$\sum_{j=1}^n a_{ij} u_j = (f, \varphi_i), \quad i = 1, 2, \dots, n, \quad (2.2.8)$$

где

$$a_{ij} = \int_a^b (\varphi_j'(x) \varphi_i'(x) + q \varphi_j(x) \varphi_i(x)) dx$$

и

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

(см. систему (1.6.7) и определение энергетического произведения (2.2.3)). Коэффициенты системы (2.2.8) легко вычисляются. При этом подчеркнем, что в силу финитности базисных функций $\varphi_i(x)$ матрица системы (2.2.8) является трехдиагональной. Пользуясь выражениями (2.2.5) для базисных функций, находим:

$$a_{ii-1} = a_{ii+1} = -\frac{1}{h} + q\frac{h}{6}, \quad a_{ii} = \frac{2}{h} + q\frac{2h}{3}.$$

В результате приходим к линейной системе

$$-\frac{u_{i-1} - 2u_i + u_{i+1}}{h^2} + q\frac{u_{i-1} + 4u_i + u_{i+1}}{6} = r_i, \quad i = 1, 2, \dots, n, \quad (2.2.9)$$

где

$$r_i = \frac{1}{h} \int_{x_{i-1}}^{x_{i+1}} f(x)\varphi_i(x) dx; \quad (2.2.10)$$

при этом в граничных узлах мы полагаем $u_0 = u_{n+1} = 0$.

Дадим некоторые комментарии к системе (2.2.9). По своему виду она очень напоминает систему сеточных уравнений, получаемую методом конечных разностей. Небольшие различия наблюдаются лишь в следующих двух позициях. В методе конечных разностей — при аппроксимации в узле x_i слагаемого $qu(x)$ — в левой части уравнения (2.2.1) обычно используется значение приближенного решения в этом узле; здесь участвует выпуклая комбинация значений искомой кусочно-линейной функции в окрестности рассматриваемого узла. Далее, в качестве правых частей системы конечно-разностных уравнений берутся значения функции $f(x)$ в узлах сетки; здесь мы имеем "усредненное" значение (2.2.10) правой части дифференциального уравнения в окрестности узла. Если, например, в промежутке $[x_{i-1}, x_{i+1}]$ функция $f(x)$ меняется незначительно, то

$$r_i \approx f(x_i) \frac{1}{h} \int_{x_{i-1}}^{x_{i+1}} \varphi_i(x) dx = f(x_i).$$

Перепишем трехдиагональную систему (2.2.9) в "стандартном" виде

$$\begin{aligned} -(6 - qh^2)u_{i-1} + 4(3 + qh^2)u_i - (6 - qh^2)u_{i+1} &= 6h^2r_i, \\ i = 1, 2, \dots, n; \quad u_0 = u_{n+1} &= 0. \end{aligned} \quad (2.2.11)$$

Нетрудно заметить, что она имеет диагональное преобладание. Для ее решения можно применить эффективный *метод прогонки* (см., например, [1, 25]).

Пусть совокупность значений v_1, v_2, \dots, v_n есть решение системы сеточных уравнений (2.2.11). Тем самым, мы имеем приближенное решение

$$\hat{v}(x) = \sum_{i=1}^n v_i \varphi_i(x)$$

задачи (2.2.1), (2.2.2) из пространства кусочно-линейных функций H_n (см. (2.2.6) и (2.2.7)). Рассмотрим вопрос об оценке разности между точным и приближенным решениями задачи. Для этого воспользуемся утверждением теоремы 1.6.2, а точнее, равенством (1.6.15), переписав его в обозначениях настоящего параграфа:

$$||u - \hat{v}|| = \min_{\hat{w} \in H_n} ||u - \hat{w}||. \quad (2.2.12)$$

Далее, пусть

$$\hat{u}(x) = \sum_{i=1}^n u_i \varphi_i(x),$$

где $u_i = u(x_i)$. Функция $\hat{u}(x) \in H_n$ совпадает с точным решением $u(x)$ в узлах сетки. Назовем эту функцию *кусочно-линейным исполнением* функции $u(x)$. Из равенства (2.2.12) очевидно следует, что

$$||u - \hat{v}|| \leq ||u - \hat{u}||. \quad (2.2.13)$$

Таким образом, вопрос об оценке погрешности приближенного решения в энергетической норме сводится к оценке разности между функцией $u(x)$ и ее кусочно-линейным исполнением $\hat{u}(x)$.

Перейдем к выводу этой оценки. Напомним, что гильбертовым пространством, ассоциируемым с нашей задачей, является пространство $L_2[a, b]$, норма в котором определяется следующим образом:

$$||f|| \equiv \left(\int_a^b f^2(x) dx \right)^{1/2}.$$

При проведении оценок мы будем неоднократно использовать неравенство типа

$$\left(\int_{\alpha}^{\beta} f(x) dx \right)^2 \leq |\beta - \alpha| \left| \int_{\alpha}^{\beta} f^2(x) dx \right|, \quad (2.2.14)$$

которое легко следует из неравенства Коши-Буняковского (1.1.4).

Лемма 2.2.1 Для любой функции $u(x) \in C^2[a, b]$ справедливо неравенство

$$\int_a^b (u'(x) - \hat{u}'(x))^2 dx \leq h^2 \int_a^b (u''(x))^2 dx. \quad (2.2.15)$$

Доказательство. Рассмотрим некоторый промежуток $[x_i, x_{i+1}]$ разбиения (2.2.4) отрезка $[a, b]$. Так как функция $\hat{u}(x)$ на этом промежутке линейная, то

$$u'(x) - \hat{u}'(x) = u'(x) - \frac{u(x_{i+1}) - u(x_i)}{h} = \frac{1}{h} \int_{x_i}^{x_{i+1}} (u'(x) - u'(\xi)) d\xi.$$

Это равенство можно записать также в виде

$$u'(x) - \hat{u}'(x) = \frac{1}{h} \int_{x_i}^{x_{i+1}} \int_{\xi}^x u''(\eta) d\eta d\xi.$$

Возведем обе части полученного равенства в квадрат. Дважды применяя неравенство (2.2.14) и расширяя пределы интегрирования, найдем, что

$$(u'(x) - \hat{u}'(x))^2 \leq h \int_{x_i}^{x_{i+1}} (u''(\eta))^2 d\eta.$$

Далее, проинтегрируем полученное неравенство:

$$\int_{x_i}^{x_{i+1}} (u'(x) - \hat{u}'(x))^2 dx \leq h^2 \int_{x_i}^{x_{i+1}} (u''(x))^2 dx.$$

В результате получим

$$\begin{aligned} \int_a^b (u'(x) - \hat{u}'(x))^2 dx &= \sum_{i=0}^n \int_{x_i}^{x_{i+1}} (u'(x) - \hat{u}'(x))^2 dx \leq \\ &h^2 \sum_{i=0}^n \int_{x_i}^{x_{i+1}} (u''(x))^2 dx = h^2 \int_a^b (u''(x))^2 dx, \end{aligned}$$

и тем самым мы приходим к требуемому неравенству (2.2.15). \square

Лемма 2.2.2 Для любой функции $u(x) \in C^2[a, b]$ справедливо неравенство

$$\int_a^b (u(x) - \hat{u}(x))^2 dx \leq h^4 \int_a^b (u''(x))^2 dx. \quad (2.2.16)$$

Доказательство. Как и в случае предыдущей леммы, рассмотрим некоторый промежуток $[x_i, x_{i+1}]$ разбиения (2.2.4) отрезка $[a, b]$. Так как в узлах сетки функции $u(x)$ и $\hat{u}(x)$ совпадают, то

$$u(x) - \hat{u}(x) = \int_{x_i}^x (u'(\xi) - \hat{u}'(\xi)) d\xi.$$

Возводя обе части равенства в квадрат и используя неравенство (2.2.14), получим

$$(u(x) - \hat{u}(x))^2 \leq (x - x_i) \int_{x_i}^x (u'(\xi) - \hat{u}'(\xi))^2 d\xi \leq h \int_{x_i}^{x_{i+1}} (u'(x) - \hat{u}'(x))^2 dx.$$

Проинтегрируем это неравенство:

$$\int_{x_i}^{x_{i+1}} (u(x) - \hat{u}(x))^2 dx \leq h^2 \int_{x_i}^{x_{i+1}} (u'(x) - \hat{u}'(x))^2 dx.$$

Далее, с учетом последнего неравенства находим, что

$$\begin{aligned} \int_a^b (u(x) - \hat{u}(x))^2 dx &= \sum_{i=0}^n \int_{x_i}^{x_{i+1}} (u(x) - \hat{u}(x))^2 dx \leq \\ &h^2 \sum_{i=0}^n \int_{x_i}^{x_{i+1}} (u'(x) - \hat{u}'(x))^2 dx = h^2 \int_a^b (u'(x) - \hat{u}'(x))^2 dx. \end{aligned}$$

Отсюда, воспользовавшись доказанным выше неравенством (2.2.15), убеждаемся в справедливости неравенства (2.2.16). \square

Имея неравенства (2.2.15) и (2.2.16), нетрудно оценить норму $\|u - \hat{u}\|$. Действительно, согласно (2.2.3),

$$\|u - \hat{u}\|^2 = \int_a^b ((u' - \hat{u}')^2 + q(u - \hat{u})^2) dx.$$

Тогда

$$\|u - \hat{u}\|^2 \leq h^2(1 + qh^2)\|u''\|^2.$$

Так как величина $1 + qh^2$ ограничена сверху, то

$$\|u - \hat{u}\| \leq Ch\|u''\|, \quad (2.2.17)$$

где C есть положительная постоянная, не зависящая от h .

Теперь уже можно вывести в окончательном виде оценку погрешности приближенного решения.

Теорема 2.2.1 Пусть $u(x)$ есть точное решение задачи (2.2.1), (2.2.2), а $\hat{v}(x)$ — приближенное решение. Тогда

$$\|u - \hat{v}\| \leq Ch\|f\|, \quad (2.2.18)$$

где C есть положительная постоянная, не зависящая от h .

Доказательство. Во-первых, как следует из неравенств (2.2.13) и (2.2.17),

$$\|u - \hat{v}\| \leq Ch \|u''\|. \quad (2.2.19)$$

Далее, нетрудно показать, что

$$\|u''\| \leq \|f\|. \quad (2.2.20)$$

Действительно, возведем обе части уравнения (2.2.1) в квадрат и проинтегрируем:

$$\int_a^b ((u''(x))^2 - 2qu''(x)u(x) + q^2u^2(x)) dx = \int_a^b f^2(x) dx.$$

Применяя формулу интегрирования по частям и принимая во внимание граничные условия (2.2.2), легко убедиться в том, что

$$- \int_a^b u''(x)u(x) dx = \int_a^b (u'(x))^2 dx.$$

В результате получим равенство

$$\int_a^b ((u''(x))^2 + 2q(u'(x)) + q^2u^2(x)) dx = \int_a^b f^2(x) dx.$$

Неравенство (2.2.20) очевидно следует из последнего равенства, а из (2.2.19) и (2.2.20) приходим к оценке (2.2.18). \square

На примере двухточечной краевой задачи мы получили первое реальное представление о методе конечных элементов. Укажем, что в рассматриваемом нами одномерном случае *конечными элементами* являются элементарные промежутки $[x_i, x_{i+1}]$ в совокупности с определенными на них линейными функциями.

В заключение параграфа дадим некоторые уточнения и комментарии к примененному нами методу. При внимательном рассмотрении можно заметить, что этот метод несколько отличается от методов Рунге и Бубнова-Галеркина — в их классической постановке. Как в методе Рунге, так и в методе Бубнова-Галеркина сначала задается бесконечная последовательность базисных элементов (см. (1.6.3) и (1.7.2)). Затем, при переходе от подпространства H_n к подпространству H_{n+1} сохраняются все использованные ранее базисные элементы $\varphi_1, \varphi_2, \dots, \varphi_n$, и к ним добавляется лишь один новый элемент φ_{n+1} . В нашей же задаче при уменьшении шага сетки мы практически полностью меняем набор узлов (2.2.4), а, тем самым, и все базисные финитные функции (2.2.5). Это неизбежно, поскольку метод конечных элементов, как уже было сказано в параграфе 2.1, является в своей основе сеточным методом.

пространство функций $u(x, y)$, непрерывных в области Ω , за исключением конечного числа точек и кусочно-гладких линий, и таких, что

$$\int_{\Omega} u^2(x, y) d\Omega < \infty.$$

Норму в этом пространстве определим следующим образом:

$$\|u\|_{0,\Omega} \equiv \left(\int_{\Omega} u^2 d\Omega \right)^{1/2}. \quad (2.3.1)$$

Условимся впредь не различать функции, не совпадающие только в точках и на линиях, на которых они терпят разрыв.

Пространство $W_2^1(\Omega)$ есть множество непрерывных в области Ω функций, частные производные первого порядка которых принадлежат $L_2(\Omega)$. Норму в этом пространстве определим так:

$$\|u\|_{1,\Omega} \equiv \left(\int_{\Omega} [u^2 + |\nabla u|^2] d\Omega \right)^{1/2}. \quad (2.3.2)$$

Через $\overset{\circ}{W}_2^1(\Omega)$ обозначим подпространство пространства $W_2^1(\Omega)$, которое состоит из функций, обращающихся в нуль на границе S области Ω .

Пространство $W_2^2(\Omega)$ есть множество непрерывно дифференцируемых в области Ω функций, частные производные второго порядка которых принадлежат $L_2(\Omega)$. Норма в этом пространстве определяется следующим образом:

$$\|u\|_{2,\Omega} \equiv \left(\int_{\Omega} \left[u^2 + |\nabla u|^2 + \left(\frac{\partial^2 u}{\partial x^2} \right)^2 + \left(\frac{\partial^2 u}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 u}{\partial y^2} \right)^2 \right] d\Omega \right)^{1/2}. \quad (2.3.3)$$

Заметим, что введенные нами пространства $W_2^1(\Omega)$ и $W_2^2(\Omega)$ являются всюду плотными множествами в соответствующих пространствах Соболева³[26]; то же самое можно сказать и о пространстве $L_2(\Omega)$.

При выводе оценок мы будем пользоваться неравенством Коши-Буняковского вида

$$\int_{\Omega} uv d\Omega \leq \left(\int_{\Omega} u^2 d\Omega \right)^{1/2} \left(\int_{\Omega} v^2 d\Omega \right)^{1/2}, \quad (2.3.4)$$

³Соболев Сергей Львович (1908–1989) — выдающийся советский математик, автор фундаментальных работ по функциональному анализу и уравнениям в частных производных.

а также неравенством, которое легко следует из последнего:

$$\left(\int_{\Omega} u \, d\Omega \right)^2 \leq \text{mes } \Omega \cdot \int_{\Omega} u^2 \, d\Omega, \quad (2.3.5)$$

где $\text{mes } \Omega$ — мера множества (области) Ω .

Теперь перейдем к рассмотрению кривых и границ областей. Пусть S есть некоторая кривая в плоскости переменных x, y и $z_0 = (x_0, y_0)$ — произвольная точка этой кривой. Введем локальную систему координат ξ, η с началом в точке z_0 ; например, если существует касательная к кривой S в точке z_0 , то ось ξ направим по этой касательной, а ось η — по нормали в точке z_0 (Рис. 2.3.1).

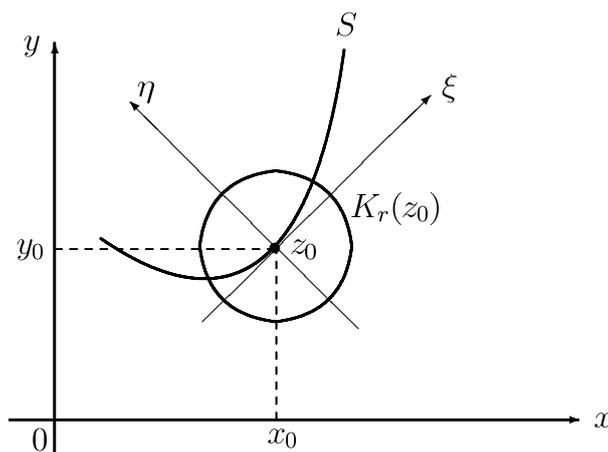


Рис. 2.3.1. Локальная система координат.

Выберем некоторое число $r > 0$ и рассмотрим шар этого радиуса с центром в точке z_0 :

$$K_r(z_0) = \{z = (x, y) : |z - z_0| \leq r\}$$

(Рис. 2.3.1). Далее, пусть $\sigma_r(z_0)$ есть проекция $S \cap K_r(z_0)$ на ось ξ . При достаточно малом радиусе шара $K_r(z_0)$ уравнение части $S \cap K_r(z_0)$ кривой S в системе координат ξ, η запишется в виде $\eta = g(\xi)$, где $\xi \in \sigma_r(z_0)$.

Говорят, что S есть кривая класса C^k , $k \geq 1$, если существует такое число r , что для произвольной точки $z_0 \in S$ функция $g(\xi)$ непрерывно дифференцируема k раз на отрезке $\sigma_r(z_0)$. Будем говорить, что S есть кусочно-гладкая кривая класса C^k , если она состоит из конечного числа гладких кривых, пересекающихся под углами, отличными от 0 и 2π .

Определим на кусочно-гладкой кривой пространство $L_2(S)$, состоящее из функций, непрерывных на S везде, за исключением конечного числа точек; норму в этом

пространстве определим через криволинейный интеграл следующим образом:

$$\|u\|_{0,S} \equiv \left(\int_S u^2 dS \right)^{1/2}. \quad (2.3.6)$$

В нашем курсе мы будем рассматривать, в общем случае, области, ограниченные кусочно-гладкими замкнутыми кривыми.

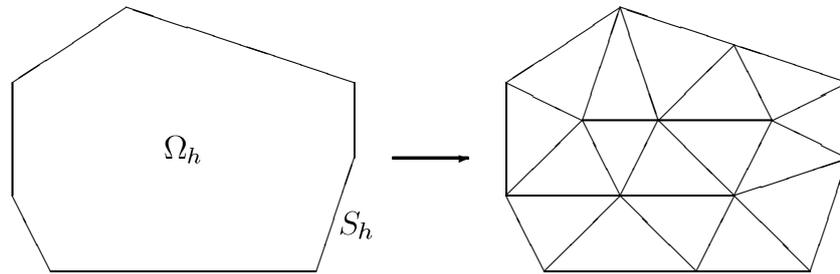
2.4 Сеточные области и восполнения функций

В методе конечных разностей, при аппроксимации дифференциального уравнения системой разностных уравнений, в области, где рассматривается уравнение, прежде всего строится *сетка*, то есть выбирается дискретное множество точек, называемых *узлами*, в которых разыскивается приближенное решение. В методе же конечных элементов, являющимся по сути также сеточным методом, выбор сетки неразрывно сопряжен с построением так называемой *сеточной области*, которая представляет собой многоугольную аппроксимацию исходной области. Затем сеточная область, в соответствии с общей концепцией метода конечных элементов, разбивается на малые подструктуры – треугольники, прямоугольники и т.п. Вопрос аппроксимации двумерной области с криволинейной границей многоугольной областью мы обсудим в дальнейшем, когда будем определять конечноэлементную систему сеточных уравнений, а сейчас перейдем к описанию сеточных областей и введению связанных с ними понятий.

Задав некоторый достаточно малый параметр $h > 0$, который в дальнейшем будет трактоваться как "усредненный" шаг сетки, рассмотрим в плоскости переменных x, y многоугольную область Ω_h с границей S_h ; предполагается, что длины звеньев ломаной S_h ограничены снизу величиной lh , где l – некоторая положительная постоянная, не зависящая от h . Далее область Ω_h разбивается на треугольники так, чтобы:

1. любые два треугольника либо не пересекались, либо имели только общую вершину, либо только общую сторону;
2. треугольники имели длины сторон, лежащие в пределах $[l_1h, l_2h]$, где l_1 и l_2 – некоторые положительные постоянные, не зависящие от h ;
3. площади треугольников лежали в пределах $[s_1h^2, s_2h^2]$, где s_1 и s_2 – некоторые положительные постоянные, не зависящие от h ;

Процесс разбиения области на треугольники называется *триангуляцией* (Рис. 2.4.1). Совокупность вершин и сторон треугольников образует *сетку*, которую обозначим через ω_h ; вершины треугольников – *узлы* сетки. Триангулированная область Ω_h и есть наша *сеточная область*. Через τ_h обозначим множество треугольников триангуляции.

Рис. 2.4.1. Триангуляция области Ω_h .

Наши следующие обозначения:

\bar{N}_h — множество узлов сетки ω_h ;

\bar{n}_h — число узлов в множестве \bar{N}_h .

Будем считать, что все узлы множества \bar{N}_h перенумерованы в некотором порядке, и пусть $\xi_i = (x_i, y_i)$ есть i -ый узел.

\bar{G}_h — пространство *сеточных функций*, определенных на множестве узлов \bar{N}_h ; при заданной нумерации узлов сеточную функцию $v = \{v_i\}_{i=1}^{\bar{n}_h}$, где v_i есть значение функции в узле ξ_i , можно рассматривать также как вектор длины \bar{n}_h :

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_{\bar{n}_h} \end{bmatrix}.$$

Поставим в соответствие каждому узлу $\xi_i \in \bar{N}_h$ сетки ω_h *базисную* непрерывную функцию $\varphi_i(x, y)$, которая равна единице в узле ξ_i , нулю в остальных узлах множества \bar{N}_h и линейна на каждом треугольнике (Рис. 2.4.2). Обозначим через $\text{supp}(\varphi_i)$ объединение треугольников триангуляции, имеющих узел ξ_i своей вершиной (от английского *support* — опора, подставка). Вне $\text{supp}(\varphi_i)$ базисная функция тождественно равна нулю. Подобласть $\text{supp}(\varphi_i)$ сеточной области Ω_h есть *локальный носитель* функции $\varphi_i(x, y)$.

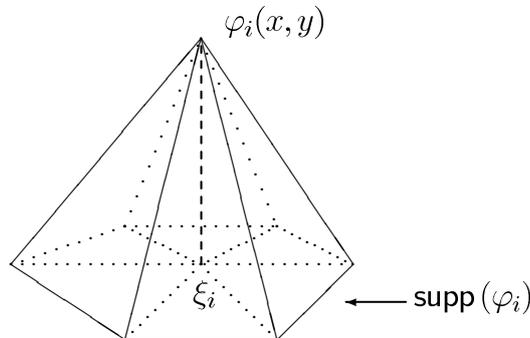


Рис. 2.4.2. Базисная функция.

Запишем аналитическое выражение базисной функции $\varphi_i(x, y)$ в треугольнике $\Delta \in \mathcal{T}_h$ с вершинами $\xi_i = (x_i, y_i)$, $\xi_j = (x_j, y_j)$ и $\xi_k = (x_k, y_k)$, полагая, что вершины треугольника перечислены в направлении против часовой стрелки. Непосредственной проверкой убеждаемся в том, что

$$\varphi_i(x, y) = \frac{1}{2S_\Delta} [(y_j - y_k)x + (x_k - x_j)y + (x_j y_k - x_k y_j)], \quad (2.4.1)$$

где S_Δ — площадь рассматриваемого треугольника, которую можно вычислить по формуле

$$S_\Delta = \frac{1}{2} \begin{vmatrix} 1 & x_i & y_i \\ 1 & x_j & y_j \\ 1 & x_k & y_k \end{vmatrix}. \quad (2.4.2)$$

Выражения для базисных функций $\varphi_j(x, y)$ и $\varphi_k(x, y)$ — в том же самом треугольнике — могут быть получены из формул (2.4.1) и (2.4.2) путем циклической перестановки индексов по схеме: $i \rightarrow j \rightarrow k$, $j \rightarrow k \rightarrow i$, $k \rightarrow i \rightarrow j$.

Нетрудно убедиться в том, что совокупность функций $\varphi_i(x, y)$, $i = 1, 2, \dots, \bar{n}_h$ образует линейно независимую систему.

Итак, мы подошли к определению пространства аппроксимирующих функций.

\bar{V}_h — пространство кусочно-линейных в области Ω_h функций; иными словами, пространство непрерывных в области Ω_h функций, линейных на каждом треугольнике триангуляции.

Для любой функции $\tilde{v} \in \bar{V}_h$ имеет место представление

$$\tilde{v}(x, y) = \sum_{i=1}^{\bar{n}_h} v_i \varphi_i(x, y) \quad \left(\text{или} \quad \sum_{\xi_i \in \bar{N}_h} v_i \varphi_i(x, y) \right), \quad (2.4.3)$$

где $v_i = \tilde{v}(x_i, y_i)$ (или, иначе, $v_i = \tilde{v}(\xi_i)$). На Рис. 2.4.3 дан фрагмент кусочно-линейной функции $\tilde{v}(x, y)$.

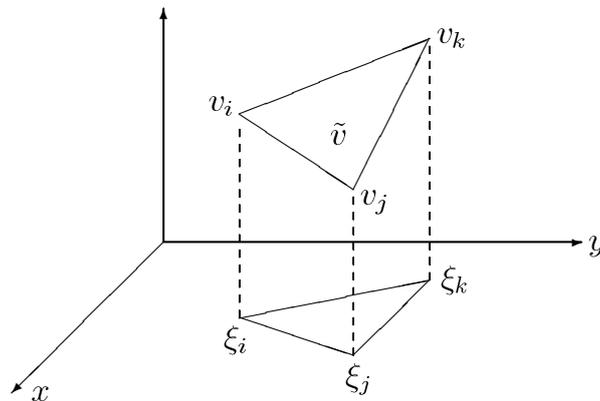


Рис. 2.4.3. Фрагмент кусочно-линейной функции.

Пространство \bar{V}_h является подпространством размерности \bar{n}_h пространства $W_2^1(\Omega_h)$.

Функцию \tilde{v} назовем *кусочно-линейным восполнением* сеточной функции $v = \{v_i\}_{i=1}^{\bar{n}_h}$ и запишем этот факт так:

$$\tilde{v} = \text{prol}(v \in \bar{G}_h : \bar{V}_h).$$

Замечание 2.4.1 В дальнейшем мы будем часто иметь дело с операцией *восполнения функций* (как сеточных, так и непрерывных) в пространствах кусочно-линейных функций. Обозначать эту операцию будем символом prol (от английского *prolongation* – продолжение, восполнение). Сама операция будет записываться в виде

$$\tilde{v} = \text{prol}(\mathcal{F} : \mathcal{L}), \quad (2.4.4)$$

где \mathcal{F} – восполняемая функция (либо сеточная, либо непрерывная – по своим значениям в узлах сетки), а \mathcal{L} – пространство кусочно-линейных функций, которому принадлежит восполнение.

Между кусочно-линейными функциями из \bar{V}_h и сеточными функциями из \bar{G}_h устанавливается естественное взаимно-однозначное соответствие. А именно, функции $\tilde{v} \in \bar{V}_h$ ставится в соответствие сеточная функция $v \in \bar{G}_h$, i -ая компонента которой равна значению функции \tilde{v} в i -ом узле множества \bar{N}_h .

Треугольники триангуляции, в совокупности с определенными на них линейными функциями, называются *линейными треугольными элементами*. В более широком смысле под *треугольными элементами* будем понимать также сами треугольники триангуляции.

Получим некоторые важные неравенства, устанавливающие соотношения между нормами сеточных функций и их кусочно-линейных восполнений.

Рассмотрим в плоскости переменных ξ, η стандартный прямоугольный треугольник Δ с вершинами $(0, 0)$, $(h, 0)$ и $(0, h)$, перенумерованными цифрами 1, 2 и 3, соответственно, как показано на Рис. 2.4.4.

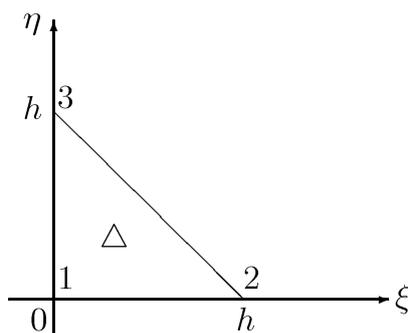


Рис. 2.4.4. Стандартный прямоугольный треугольник Δ .

Пусть \tilde{v} есть произвольная линейная функция, заданная на треугольнике Δ . Ее можно записать в виде

$$\tilde{v}(\xi, \eta) = \sum_{i=1}^3 v_i \varphi_i(\xi, \eta), \quad (2.4.5)$$

где v_i есть значение функции $\tilde{v}(\xi, \eta)$ в вершине треугольника с номером i , а $\varphi_i(\xi, \eta)$ — соответствующая базисная функция линейного восполнения. Нетрудно проверить, что

$$\varphi_1(\xi, \eta) = -\frac{\xi}{h} - \frac{\eta}{h} + 1, \quad \varphi_2(\xi, \eta) = \frac{\xi}{h}, \quad \varphi_3(\xi, \eta) = \frac{\eta}{h}. \quad (2.4.6)$$

Лемма 2.4.1 *Справедливы неравенства*

$$\frac{h^2}{24} \sum_{i=1}^3 v_i^2 \leq \int_{\Delta} \tilde{v}^2 d\Delta \leq \frac{h^2}{6} \sum_{i=1}^3 v_i^2. \quad (2.4.7)$$

Доказательство. В силу представления (2.4.5) функции \tilde{v} имеем

$$\int_{\Delta} \tilde{v}^2 d\Delta = \sum_{i,j=1}^3 v_i v_j \int_{\Delta} \varphi_i \varphi_j d\Delta.$$

Пользуясь выражениями (2.4.6), вычислим интегралы:

$$\int_{\Delta} \varphi_i^2 d\Delta = \frac{h^2}{12}, \quad i = 1, 2, 3; \quad \int_{\Delta} \varphi_i \varphi_j d\Delta = \frac{h^2}{24}, \quad i \neq j.$$

Тогда

$$\int_{\Delta} \tilde{v}^2 d\Delta = \frac{h^2}{24} [2(v_1^2 + v_2^2 + v_3^2) + 2v_1v_2 + 2v_2v_3 + 2v_1v_3]. \quad (2.4.8)$$

Отсюда, применяя известное неравенство Коши $2ab \leq a^2 + b^2$, находим, что

$$\int_{\Delta} \tilde{v}^2 d\Delta \leq \frac{h^2}{6} (v_1^2 + v_2^2 + v_3^2).$$

Далее, выделяя в правой части равенства (2.4.8) полный квадрат, получим неравенство

$$\int_{\Delta} \tilde{v}^2 d\Delta = \frac{h^2}{24} [(v_1 + v_2 + v_3)^2 + (v_1^2 + v_2^2 + v_3^2)] \geq \frac{h^2}{24} (v_1^2 + v_2^2 + v_3^2).$$

Тем самым завершается доказательство утверждения. \square

Лемма 2.4.2 *Справедливо неравенство*

$$\int_{\Delta} |\nabla \tilde{v}|^2 d\Delta \leq 2 \sum_{i=1}^3 v_i^2. \quad (2.4.9)$$

Доказательство. Исходя из представления (2.4.5) функции \tilde{v} и выражений (2.4.6) для базисных функций, имеем

$$|\nabla \tilde{v}|^2 = \left(\frac{\partial \tilde{v}}{\partial \xi} \right)^2 + \left(\frac{\partial \tilde{v}}{\partial \eta} \right)^2 = \left(v_1 \frac{\partial \varphi_1}{\partial \xi} + v_2 \frac{\partial \varphi_2}{\partial \xi} + v_3 \frac{\partial \varphi_3}{\partial \xi} \right)^2 + \left(v_1 \frac{\partial \varphi_1}{\partial \eta} + v_2 \frac{\partial \varphi_2}{\partial \eta} + v_3 \frac{\partial \varphi_3}{\partial \eta} \right)^2 = \frac{(v_2 - v_1)^2 + (v_3 - v_1)^2}{h^2}.$$

Отсюда

$$\int_{\Delta} |\nabla \tilde{v}|^2 d\Delta = \frac{(v_2 - v_1)^2 + (v_3 - v_1)^2}{2} \leq 2 \sum_{i=1}^3 v_i^2,$$

чем и завершается доказательство утверждения. \square

Для вывода аналогичных неравенств во всей сеточной области Ω_h мы воспользуемся отображением треугольников триангуляции на стандартный прямоугольный треугольник.

Рассмотрим некоторый треугольник $\Delta' \in \tau_h$. Пусть вершины треугольника, перенумерованные цифрами 1, 2 и 3, имеют координаты (x_1, y_1) , (x_2, y_2) и (x_3, y_3) , соответственно. Этот треугольник отображается на прямоугольный треугольник Δ в плоскости переменных ξ, η , соответствующими вершинами которого являются $(0, 0)$, $(h, 0)$ и $(0, h)$ (Рис. 2.4.5).

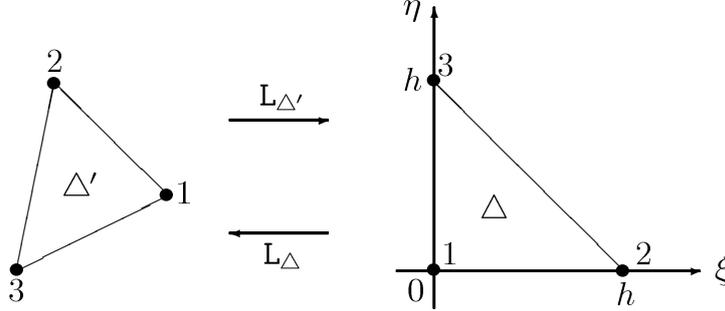


Рис. 2.4.5. Отображение на стандартный треугольник.

Линейное отображение $L_{\Delta'} : \Delta' \rightarrow \Delta$ осуществляется по формулам

$$\begin{aligned} \xi &= A_1 x + B_1 y + C_1, \\ \eta &= A_2 x + B_2 y + C_2 \end{aligned} \quad (2.4.10)$$

с коэффициентами

$$\begin{aligned} A_1 &= h \frac{y_3 - y_2}{2S_{\Delta'}}, \quad B_1 = h \frac{x_2 - x_3}{2S_{\Delta'}}, \quad C_1 = h \frac{x_1(y_2 - y_3) + y_1(x_3 - x_2)}{2S_{\Delta'}}, \\ A_2 &= h \frac{y_1 - y_2}{2S_{\Delta'}}, \quad B_2 = h \frac{x_2 - x_1}{2S_{\Delta'}}, \quad C_2 = h \frac{x_1 y_2 - x_2 y_1}{2S_{\Delta'}}, \end{aligned} \quad (2.4.11)$$

где $S_{\Delta'}$ есть площадь треугольника $S_{\Delta'}$, вычисляемая по формуле (2.4.2), а именно,

$$S_{\Delta'} = \frac{1}{2} \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}.$$

Принимая во внимание свойства треугольников триангуляции (см. параграф 2.4, пункты 1, 2 и 3), можем дать оценку коэффициентов линейного отображения (2.4.10); из формул (2.4.11) следует, что

$$|A_i|, |B_i| \leq \frac{l_2}{2s_1}, \quad i = 1, 2. \quad (2.4.12)$$

Якобиан отображения $L_{\Delta'}$, то есть

$$\frac{D(\xi, \eta)}{D(x, y)} = \begin{vmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial \xi}{\partial y} \\ \frac{\partial \eta}{\partial x} & \frac{\partial \eta}{\partial y} \end{vmatrix} = \begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix} = A_1 B_2 - A_2 B_1$$

легко вычисляется; пользуясь формулами (2.4.11), находим

$$A_1 B_2 - A_2 B_1 = \frac{h^2}{2S_{\Delta'}}.$$

Отсюда получаем двустороннюю оценку якобиана:

$$\frac{1}{2s_2} \leq \frac{D(\xi, \eta)}{D(x, y)} \leq \frac{1}{2s_1}. \quad (2.4.13)$$

Нам понадобится также и обратное отображение $L_{\Delta} : \Delta \rightarrow \Delta'$, записываемое в виде

$$\begin{aligned} x &= Q_1 \xi + R_1 \eta + T_1, \\ y &= Q_2 \xi + R_2 \eta + T_2, \end{aligned} \quad (2.4.14)$$

где

$$\begin{aligned} Q_1 &= \frac{x_2 - x_1}{h}, & R_1 &= \frac{x_3 - x_2}{h}, & T_1 &= x_1, \\ Q_2 &= \frac{y_2 - y_1}{h}, & R_2 &= \frac{y_3 - y_2}{h}, & T_2 &= y_1. \end{aligned} \quad (2.4.15)$$

Как очевидно следует из последних формул,

$$|Q_i|, |R_i| \leq l_2, \quad i = 1, 2. \quad (2.4.16)$$

Двусторонняя оценка якобиана этого отображения, основываясь на (2.4.13), такова:

$$2s_1 \leq \frac{D(x, y)}{D(\xi, \eta)} \leq 2s_2. \quad (2.4.17)$$

Теорема 2.4.1 Для любой кусочно-линейной функции $\tilde{v}(x, y) \in \bar{V}_h$ справедливы неравенства

$$C_1 h^2 \sum_{i=1}^{\bar{n}_h} v_i^2 \leq \|\tilde{v}\|_{0, \Omega_h}^2 \leq C_2 h^2 \sum_{i=1}^{\bar{n}_h} v_i^2, \quad (2.4.18)$$

где v_i есть значение функции $\tilde{v}(x, y)$ в i -ом узле сетки ω_h , а C_1 и C_2 — положительные постоянные, не зависящие от h .

Доказательство. Полагая, согласно отображению (2.4.14), $\tilde{v}(\xi, \eta) = \tilde{v}(x(\xi, \eta), y(\xi, \eta))$, для произвольного треугольника триангуляции $\Delta' \in \tau_h$ запишем равенство

$$\int_{\Delta'} \tilde{v}^2(x, y) d\Delta' = \int_{\Delta} \tilde{v}^2(\xi, \eta) \frac{D(x, y)}{D(\xi, \eta)} d\Delta'.$$

Отсюда, имея двустороннюю оценку якобиана (2.4.17), получим

$$2s_1 \int_{\Delta} \tilde{v}^2 d\Delta \leq \int_{\Delta'} \tilde{v}^2 d\Delta' \leq 2s_2 \int_{\Delta} \tilde{v}^2 d\Delta.$$

Обратимся к полученным выше неравенствам (2.4.7). Тогда

$$\frac{s_1}{12} h^2 \sum_{\Delta'} v_i^2 \leq \int_{\Delta'} \tilde{v}^2 d\Delta' \leq \frac{s_2}{3} h^2 \sum_{\Delta'} v_i^2, \quad (2.4.19)$$

где $\sum_{\Delta'} v_i^2$ означает сумму квадратов значений сеточной функции в узлах треугольника Δ' .

Левое из неравенств (2.4.18) легко устанавливается. Действительно,

$$\int_{\Omega_h} \tilde{v}^2 d\Omega = \sum_{\Delta' \in \tau_h} \int_{\Delta'} \tilde{v}^2 d\Delta' \geq \frac{s_1}{12} h^2 \sum_{\Delta' \in \tau_h} \sum_{\Delta'} v_i^2 \geq \frac{s_1}{12} h^2 \sum_{i=1}^{\bar{n}_h} v_i^2. \quad (2.4.20)$$

Продолжим доказательство, указав предварительно на одно важное свойство триангуляции сеточной области. Площадь треугольника $\Delta' \in \tau_h$ может быть вычислена по хорошо известной формуле

$$S_{\Delta'} = \frac{1}{2} ab \sin \alpha,$$

где a и b есть стороны треугольника Δ' , а α — угол между ними. Тогда

$$\sin \alpha = \frac{2S_{\Delta'}}{ab} \geq \frac{2s_1 h^2}{l_2^2 h^2} = \frac{2s_1}{l_2^2}.$$

Последнее означает, что углы треугольников находятся в пределах $\alpha_0 \leq \alpha \leq \pi - \alpha_0$, где

$$\alpha_0 = \arcsin \frac{2s_1}{l_2^2}. \quad (2.4.21)$$

Тем самым, мы можем утверждать, что число треугольников триангуляции, имеющих один и тот же узел сетки своей общей вершиной, не превосходит величины

$$\frac{2\pi}{\alpha_0}. \quad (2.4.22)$$

С учетом сказанного, из (2.4.19) получим, что

$$\int_{\Omega_h} \tilde{v}^2 d\Omega = \sum_{\Delta' \in \tau_h} \int_{\Delta'} \tilde{v}^2 d\Delta' \leq \frac{s_2}{3} h^2 \sum_{\Delta' \in \tau_h} \sum_{\Delta'} v_i^2 \leq \frac{2\pi s_2}{3\alpha_0} h^2 \sum_{i=1}^{\bar{n}_h} v_i^2. \quad (2.4.23)$$

Итак, наше утверждение доказано. Из (2.4.20) и (2.4.23) следует, что в качестве постоянных C_1 и C_2 в неравенствах (2.4.18) могут рассматриваться величины

$$C_1 = \frac{s_1}{12} \quad \text{и} \quad C_2 = \frac{2\pi s_2}{3\alpha_0}, \quad (2.4.24)$$

где α_0 вычисляется согласно (2.4.21). \square

Теорема 2.4.2 Для любой кусочно-линейной функции $\tilde{v}(x, y) \in \bar{V}_h$ справедливо неравенство

$$\|\|\nabla \tilde{v}\|\|_{0, \Omega_h}^2 \leq C \sum_{i=1}^{\bar{n}_h} v_i^2, \quad (2.4.25)$$

где v_i есть значение функции $\tilde{v}(x, y)$ в i -ом узле сетки ω_h , а C — положительная постоянная, не зависящая от h .

Доказательство. Рассмотрим произвольный треугольник триангуляции $\Delta' \in \tau_h$. Пусть, согласно отображению (2.4.14), $\tilde{v}(\xi, \eta) = \tilde{v}(x(\xi, \eta), y(\xi, \eta))$. Исходя из равенств

$$\frac{\partial \tilde{v}}{\partial x} = \frac{\partial \tilde{v}}{\partial \xi} \frac{\partial \xi}{\partial x} + \frac{\partial \tilde{v}}{\partial \eta} \frac{\partial \eta}{\partial x} = A_1 \frac{\partial \tilde{v}}{\partial \xi} + A_2 \frac{\partial \tilde{v}}{\partial \eta},$$

$$\frac{\partial \tilde{v}}{\partial y} = \frac{\partial \tilde{v}}{\partial \xi} \frac{\partial \xi}{\partial y} + \frac{\partial \tilde{v}}{\partial \eta} \frac{\partial \eta}{\partial y} = B_1 \frac{\partial \tilde{v}}{\partial \xi} + B_2 \frac{\partial \tilde{v}}{\partial \eta}$$

(см. отображение (2.4.10)), имеем

$$\begin{aligned} \int_{\Delta'} |\nabla \tilde{v}|^2 d\Delta' &= \int_{\Delta'} \left[\left(\frac{\partial \tilde{v}}{\partial x} \right)^2 + \left(\frac{\partial \tilde{v}}{\partial y} \right)^2 \right] d\Delta' \\ &= \int_{\Delta} \left[\left(A_1 \frac{\partial \tilde{v}}{\partial \xi} + A_2 \frac{\partial \tilde{v}}{\partial \eta} \right)^2 + \left(B_1 \frac{\partial \tilde{v}}{\partial \xi} + B_2 \frac{\partial \tilde{v}}{\partial \eta} \right)^2 \right] \frac{D(x, y)}{D(\xi, \eta)} d\Delta. \end{aligned}$$

Применяя известное неравенство $(a + b)^2 \leq 2(a^2 + b^2)$, получим

$$\int_{\Delta'} |\nabla \tilde{v}|^2 d\Delta' \leq 2 \int_{\Delta} \left[(A_1^2 + B_1^2) \left(\frac{\partial \tilde{v}}{\partial \xi} \right)^2 + (A_2^2 + B_2^2) \left(\frac{\partial \tilde{v}}{\partial \eta} \right)^2 \right] \frac{D(x, y)}{D(\xi, \eta)} d\Delta.$$

Неравенства

$$A_i^2 + B_i^2 \leq \frac{l_2^2}{2s_1^2}, \quad i = 1, 2$$

следуют непосредственно из оценок (2.4.12). Тогда

$$\int_{\Delta'} |\nabla \tilde{v}|^2 d\Delta' \leq \frac{l_2^2}{s_1^2} \int_{\Delta} |\nabla \tilde{v}|^2 \frac{D(x, y)}{D(\xi, \eta)} d\Delta.$$

Отсюда, имея оценку якобиана (2.4.17), получим, что

$$\int_{\Delta'} |\nabla \tilde{v}|^2 d\Delta' \leq 2 \frac{s_2 l_2^2}{s_1^2} \int_{\Delta} |\nabla \tilde{v}|^2 d\Delta.$$

Далее, если обратиться к полученному выше неравенству (2.4.9), то

$$\int_{\Delta'} |\nabla \tilde{v}|^2 d\Delta' \leq 4 \frac{s_2 l_2^2}{s_1^2} \sum_{\Delta'} v_i^2,$$

где $\sum_{\Delta'} v_i^2$ означает сумму квадратов значений сеточной функции в узлах треугольника Δ' .

Перейдем к заключительному этапу доказательства. Пользуясь последним неравенством и принимая во внимание верхнюю оценку (2.4.22) числа треугольников триангуляции, имеющих общую вершину, получим

$$\int_{\Omega_h} |\nabla \tilde{v}|^2 d\Omega = \sum_{\Delta' \in \pi_h} \int_{\Delta'} |\nabla \tilde{v}|^2 d\Delta' \leq 4 \frac{s_2 l_2^2}{s_1^2} \sum_{\Delta' \in \pi_h} \sum_{\Delta'} v_i^2 \leq \frac{8\pi s_2 l_2^2}{s_1^2 \alpha_0} \sum_{i=1}^{\bar{n}_h} v_i^2.$$

Таким образом, наше утверждение доказано. Как и в предыдущей теореме заметим, что в качестве постоянной C в неравенстве (2.4.25) может рассматриваться величина

$$C = \frac{8\pi s_2 l_2^2}{s_1^2 \alpha_0}, \quad (2.4.26)$$

где α_0 определяется согласно (2.4.21). \square

Полученные в теоремах 2.4.1 и 2.4.2 неравенства будут существенно использованы позднее при оценке погрешности приближенного решения и при изучении вопросов, связанных с численной реализацией систем сеточных уравнений.

2.5 Конечноэлементное решение уравнения Пуассона

Рассмотрим в конечной односвязной области Ω , которая ограничена кривой S , уравнение Пуассона

$$-\Delta u = f \quad (2.5.1)$$

с однородным граничным условием

$$u|_S = 0. \quad (2.5.2)$$

Напомним, что эту краевую задачу мы уже обсуждали выше, в параграфе 1.5.

Умножим обе части уравнения (2.5.1) на произвольную функцию $\varphi \in \overset{\circ}{W}_2^1(\Omega)$ и проинтегрируем по области Ω . Применяя формулу интегрирования по частям (1.5.16), с учетом граничного условия (2.5.2) получим так называемое *интегральное тождество*

$$\int_{\Omega} \nabla u \nabla \varphi \, d\Omega = \int_{\Omega} f \varphi \, d\Omega \quad \forall \varphi \in \overset{\circ}{W}_2^1(\Omega). \quad (2.5.3)$$

Заметим, что, по существу, это тождество уже было получено нами ранее, в (1.5.18).

Функцию $u \in \overset{\circ}{W}_2^1(\Omega)$, удовлетворяющую интегральному тождеству (2.5.3), назовем *обобщенным решением* задачи Дирихле (2.5.1), (2.5.2).

Вопрос существования и единственности обобщенного решения досконально изучен в научной литературе (см., например, [15, 16]). Пусть правая часть f уравнения (2.5.1) есть ограниченная в области Ω функция, принадлежащая пространству $L_2(\Omega)$. Предположим, что граница S является кусочно-гладкой кривой класса C^2 . Если внутренние углы области Ω в точках стыка гладких кусков границы лежат в пределах $(0, \pi]$, то обобщенное решение задачи обладает большей гладкостью, а именно, принадлежит пространству $W_2^2(\Omega)$ и выполняется неравенство

$$\|u\|_{2,\Omega} \leq C \|f\|_{0,\Omega}, \quad (2.5.4)$$

где C есть положительная постоянная, не зависящая от функции f .

Отметим следующее. Определенное нами посредством интегрального тождества (2.5.3) обобщенное решение полностью согласуется с понятием обобщенного решения задачи о минимуме функционала энергии, введенного в параграфе 1.4. Покажем, что обобщенное решение $u \in \overset{\circ}{W}_2^1(\Omega)$ из (2.5.3) доставляет минимум функционалу

$$\Phi(w) = \int_{\Omega} (|\nabla w|^2 - 2fw) \, d\Omega, \quad w \in \overset{\circ}{W}_2^1(\Omega) \quad (2.5.5)$$

(см. (1.5.22)). Действительно, так как

$$\int_{\Omega} \nabla u \nabla w \, d\Omega = \int_{\Omega} f w \, d\Omega,$$

то функционал (2.5.5) можно записать в виде

$$\Phi(w) = \int_{\Omega} (|\nabla w|^2 - 2\nabla w \nabla u) d\Omega.$$

В частности,

$$\Phi(u) = - \int_{\Omega} |\nabla u|^2 d\Omega.$$

Тогда

$$\Phi(w) - \Phi(u) = \int_{\Omega} (|\nabla w|^2 - 2\nabla w \nabla u + |\nabla u|^2) d\Omega = \int_{\Omega} |\nabla(w - u)|^2 d\Omega \geq 0,$$

то есть $\Phi(w) \geq \Phi(u)$.

Полученное выше интегральное тождество (2.5.3) играет важную роль в определении приближенного решения. Начнем с построения сеточной области Ω_h , о которой уже говорилось в параграфе 2.4. Прежде всего, задавшись некоторым достаточно малым шагом h , построим замкнутую ломаную S_h , обладающую следующими свойствами:

1. область Ω_h , ограниченная ломаной S_h , содержится в области Ω ;
2. наикратчайшее расстояние точек ломаной S_h от кривой S не превосходит величины dh^2 , где d — некоторая положительная постоянная, не зависящая от h ;
3. длины звеньев ломаной S_h ограничены снизу величиной lh , где l — некоторая положительная постоянная, не зависящая от h ;

Возможность построения ломаной S_h с указанными свойствами, вообще говоря, нуждается в обосновании; вместе с тем, эти геометрические положения наглядно ясны, но требуют довольно громоздких выкладок, которые мы приводить не будем.

Далее проводится триангуляция сеточной области Ω_h согласно требованиям, приведенным в параграфе 2.4. Перечислим эти требования еще раз; а именно, триангуляция осуществляется так, чтобы:

4. любые два треугольника либо не пересекались, либо имели только общую вершину, либо только общую сторону;
5. треугольники имели длины сторон, лежащие в пределах $[l_1h, l_2h]$, где l_1 и l_2 — некоторые положительные постоянные, не зависящие от h ;
6. площади треугольников лежали в пределах $[s_1h^2, s_2h^2]$, где s_1 и s_2 — некоторые положительные постоянные, не зависящие от h .

Итак, мы имеем сеточную область $\Omega_h \subseteq \Omega$. Для этой области остаются в силе все обозначения, введенные в параграфе 2.4. В дополнение к ним, с учетом однородности граничного условия, введем еще и следующие обозначения:

N_h — множество внутренних узлов сетки ω_h (иначе говоря, множество узлов, не принадлежащих границе S_h);

n_h — число узлов в множестве N_h ;

G_h — пространство сеточных функций, определенных на множестве узлов N_h ; при заданной нумерации узлов множества N_h сеточную функцию $v = \{v_i\}_{i=1}^{n_h}$, где v_i есть значение функции в узле $\xi_i \in N_h$, мы можем рассматривать также как вектор

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n_h} \end{bmatrix};$$

V_h — пространство кусочно-линейных в области Ω_h функций, обращающихся в нуль на границе S_h . Для любой функции $\hat{v} \in V_h$ имеет место представление

$$\hat{v}(x, y) = \sum_{i=1}^{n_h} v_i \varphi_i(x, y) \quad \left(\text{или} \quad \sum_{\xi_i \in N_h} v_i \varphi_i(x, y) \right), \quad (2.5.6)$$

где $v_i = \hat{v}(x_i, y_i)$ (иначе, $v_i = \hat{v}(\xi_i)$). Таким образом, $\hat{v} = \text{prol}(v \in G_h : V_h)$.

Так как $\Omega_h \subseteq \Omega$ и $\hat{u}|_{S_h} = 0$, то мы можем продолжить функцию $\hat{v}(x, y)$ нулем на приграничную подобласть $\Omega \setminus \Omega_h$. С учетом этого дополнения пространство V_h можно рассматривать как подпространство размерности n_h пространства $\overset{\circ}{W}_2^1(\Omega)$.

Перейдем к определению приближенного решения задачи. Так как $V_h \subset \overset{\circ}{W}_2^1(\Omega)$, то из интегрального тождества (2.5.3) следует, что

$$\int_{\Omega} \nabla u \nabla \hat{w} \, d\Omega = \int_{\Omega} f \hat{w} \, d\Omega \quad \forall \hat{w} \in V_h. \quad (2.5.7)$$

Функцию $\hat{v} \in V_h$, удовлетворяющую интегральному тождеству

$$\int_{\Omega} \nabla \hat{v} \nabla \hat{w} \, d\Omega = \int_{\Omega} f \hat{w} \, d\Omega \quad \forall \hat{w} \in V_h, \quad (2.5.8)$$

назовем приближенным *конечноэлементным решением* задачи (2.5.1), (2.5.2).

Как следует из представления (2.5.6) кусочно-линейных функций,

$$\nabla \hat{v} = \sum_{j=1}^{n_h} v_j \nabla \varphi_j, \quad \nabla \hat{w} = \sum_{i=1}^{n_h} w_i \nabla \varphi_i.$$

Подставляя эти разложения в левую часть интегрального тождества (2.5.8), получим, что

$$\int_{\Omega} \nabla \hat{v} \nabla \hat{w} \, d\Omega = \sum_{i,j=1}^{n_h} v_j w_i \int_{\Omega_h} \nabla \varphi_i \nabla \varphi_j \, d\Omega_h.$$

Введем в рассмотрение матрицу

$$A = [a_{ij}]_{n_h \times n_h}, \quad \text{где} \quad a_{ij} = \int_{\Omega_h} \nabla \varphi_i \nabla \varphi_j d\Omega_h. \quad (2.5.9)$$

Тогда

$$\int_{\Omega} \nabla \hat{v} \nabla \hat{w} d\Omega = \sum_{i,j=1}^{n_h} a_{ij} v_j w_i = w^T A v; \quad (2.5.10)$$

напомним, что $\hat{v} = \text{prol}(v \in G_h : V_h)$ и $\hat{w} = \text{prol}(w \in G_h : V_h)$. Далее,

$$\int_{\Omega} f \hat{w} d\Omega = \sum_{i=1}^{n_h} w_i \int_{\Omega_h} f \varphi_i d\Omega_h.$$

Вводя в рассмотрение сеточную функцию

$$g = \{g_i\}_{i=1}^{n_h}, \quad \text{где} \quad g_i = \int_{\Omega_h} f \varphi_i d\Omega_h, \quad (2.5.11)$$

получим, что

$$\int_{\Omega} f \hat{w} d\Omega = \sum_{i=1}^{n_h} g_i w_i = w^T g. \quad (2.5.12)$$

В итоге, имея (2.5.10) и (2.5.12), получим матрично-векторную запись интегрального тождества (2.5.8), а именно:

$$w^T A v = w^T g \quad \forall w \in G_h.$$

В силу произвольности сеточной функции $w \in G_h$, мы приходим к системе сеточных уравнений

$$A v = g \quad (2.5.13)$$

для вычисления конечноэлементного решения задачи (2.5.1), (2.5.2).

Из предыдущих рассуждений легко следует, что

$$w^T A u = \int_{\Omega_h} \nabla \hat{u} \nabla \hat{w} d\Omega_h \quad \forall u, w \in G_h, \quad (2.5.14)$$

где $\hat{u} = \text{prol}(u \in G_h : V_h)$, $\hat{w} = \text{prol}(w \in G_h : V_h)$.

Вопросы, связанные с численной реализацией системы уравнений (2.5.13), мы обсудим в следующей главе. Укажем лишь, что матрица A системы является симметричной, что легко заметить из определения (2.5.9) элементов этой матрицы. Кроме того, матрица

системы *разреженная*, то есть содержит большое число нулей; действительно, как видно опять же из (2.5.9), если узлы ξ_i и ξ_j сетки не являются соседними, то $a_{ij} = 0$. Пользуясь терминологией, принятой в механике, матрицу A часто называют *матрицей жесткости*.

Чтобы яснее представить структуру конечноэлементных уравнений, рассмотрим простой случай, когда область Ω есть квадрат. Например, пусть $\Omega = \{(x, y) : 0 < x, y < 1\}$. Сначала построим в этой области равномерную квадратную сетку с шагом h ; затем каждую ячейку сетки посредством диагонали разобьем на два прямоугольных треугольника, как показано на Рис. 2.5.1.

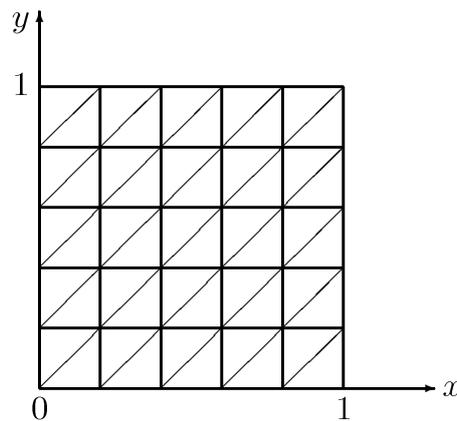


Рис. 2.5.1. Триангуляция квадрата.

Рассмотрим некоторый внутренний i -ый узел сетки с координатами x_i, y_i ; условно обозначим этот узел буквой O . Пусть $\varphi_O(x, y)$ есть базисная кусочно-линейная функция, соответствующая этому узлу. Указанная базисная функция тождественно равна нулю вне объединения треугольников триангуляции $\Delta_k, k = 1, 2, \dots, 6$ (Рис. 2.5.2).

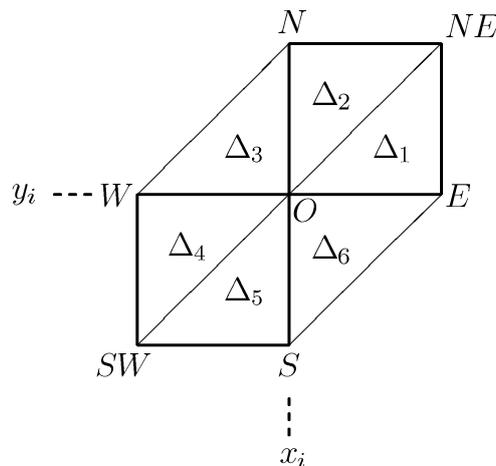


Рис. 2.5.2. Окрестность рассматриваемого узла O .

Связанным с узлом O соседним узлам сетки припишем условные обозначения E, W, N, S, NE и SW ; по тому же принципу будем обозначать и соответствующие этим узлам базисные кусочно-линейные функции: $\varphi_E(x, y), \varphi_W(x, y)$ и т.д. Легко проверить, что в отмеченных треугольниках триангуляции базисная функция $\varphi_O(x, y)$ имеет следующий вид:

$$\varphi_O(x, y) = \begin{cases} -(x - x_i)/h + 1 & \text{в } \Delta_1, \\ -(y - y_i)/h + 1 & \text{в } \Delta_2, \\ (x - x_i)/h - (y - y_i)/h + 1 & \text{в } \Delta_3, \\ (x - x_i)/h + 1 & \text{в } \Delta_4, \\ (y - y_i)/h + 1 & \text{в } \Delta_5, \\ -(x - x_i)/h + (y - y_i)/h + 1 & \text{в } \Delta_6. \end{cases}$$

Далее, выпишем остальные базисные функции:

$$\begin{aligned} \varphi_E(x, y) &= \begin{cases} (x - x_i)/h - (y - y_i)/h & \text{в } \Delta_1, \\ (x - x_i)/h & \text{в } \Delta_6, \end{cases} \\ \varphi_W(x, y) &= \begin{cases} -(x - x_i)/h & \text{в } \Delta_3, \\ -(x - x_i)/h + (y - y_i)/h & \text{в } \Delta_4, \end{cases} \\ \varphi_N(x, y) &= \begin{cases} -(x - x_i)/h + (y - y_i)/h & \text{в } \Delta_2, \\ (y - y_i)/h & \text{в } \Delta_3, \end{cases} \\ \varphi_S(x, y) &= \begin{cases} (x - x_i)/h - (y - y_i)/h & \text{в } \Delta_5, \\ -(y - y_i)/h & \text{в } \Delta_6, \end{cases} \\ \varphi_{NE}(x, y) &= \begin{cases} (y - y_i)/h & \text{в } \Delta_1, \\ (x - x_i)/h & \text{в } \Delta_2, \end{cases} \\ \varphi_{SW}(x, y) &= \begin{cases} -(y - y_i)/h & \text{в } \Delta_4, \\ -(x - x_i)/h & \text{в } \Delta_5. \end{cases} \end{aligned}$$

Согласно введенным нами локальным обозначениям, i -ое уравнение системы сеточных уравнений (2.5.13) записывается как

$$a_{OO}v_O + a_{OE}v_E + a_{OW}v_W + a_{ON}v_N + a_{OS}v_S + a_{ONE}v_{NE} + a_{OSW}v_{SW} = g_O, \quad (2.5.15)$$

где

$$\begin{aligned} g_O &= \int_{\bigcup_{k=1}^6 \Delta_k} f \varphi_O d\Omega_h, & a_{OO} &= \int_{\bigcup_{k=1}^6 \Delta_k} |\nabla \varphi_O|^2 d\Omega_h, \\ a_{OE} &= \int_{\Delta_1 \cup \Delta_6} \nabla \varphi_O \nabla \varphi_E d\Omega_h, & a_{OW} &= \int_{\Delta_3 \cup \Delta_4} \nabla \varphi_O \nabla \varphi_W d\Omega_h, \end{aligned}$$

$$a_{ON} = \int_{\Delta_2 \cup \Delta_3} \nabla \varphi_O \nabla \varphi_N d\Omega_h, \quad a_{OS} = \int_{\Delta_5 \cup \Delta_6} \nabla \varphi_O \nabla \varphi_S d\Omega_h,$$

$$a_{ONE} = \int_{\Delta_1 \cup \Delta_2} \nabla \varphi_O \nabla \varphi_{NE} d\Omega_h, \quad a_{OSW} = \int_{\Delta_4 \cup \Delta_5} \nabla \varphi_O \nabla \varphi_{SW} d\Omega_h$$

(см. (2.5.9) и (2.5.11)). Вычисляя интегралы, находим, что

$$a_{OO} = 4, \quad a_{OE} = a_{OW} = a_{ON} = a_{OS} = -1, \quad a_{ONE} = a_{OSW} = 0.$$

Тем самым, из (2.5.15) получаем уравнение

$$4v_O - v_E - v_W - v_N - v_S = g_O.$$

Если переписать это уравнение в виде

$$-\frac{v_E - 2v_O + v_W}{h^2} - \frac{v_N - 2v_O + v_S}{h^2} = \frac{1}{h^2} g_O,$$

то нетрудно заметить, что оно фактически является уравнением обычной пятиточечной разностной схемы. Отличие состоит лишь в том, что правая часть равна не $f_O \equiv f(x_i, y_i)$, а некоторому усредненному значению правой части в окрестности узла O ; при этом

$$\frac{1}{h^2} g_O = \frac{1}{h^2} \int_{\bigcup_{k=1}^6 \Delta_k} f \varphi_O d\Omega_h \approx \frac{1}{h^2} f_O \int_{\bigcup_{k=1}^6 \Delta_k} \varphi_O d\Omega_h = f_O.$$

Далее нам нужно оценить погрешность приближенного решения. Укажем путь получения оценки. Из интегральных тождеств (2.5.7) и (2.5.8) следует, что

$$\int_{\Omega} \nabla(u - \hat{v}) \nabla \hat{w} d\Omega = 0 \quad \forall \hat{w} \in V_h. \quad (2.5.16)$$

Возьмем в качестве функции \hat{w} разность $\hat{u} - \hat{v}$, где $\hat{u} = \text{prol}(u \in W_2^2(\Omega) : V_h)$. Получим

$$\int_{\Omega} \nabla(u - \hat{v}) \nabla(\hat{u} - \hat{v}) d\Omega = 0.$$

Перепишем это равенство в виде

$$\int_{\Omega} |\nabla(u - \hat{v})|^2 d\Omega = \int_{\Omega} \nabla(u - \hat{v}) \nabla(u - \hat{u}) d\Omega. \quad (2.5.17)$$

При выводе оценок в дальнейшем мы будем часто пользоваться неравенством Коши-Буняковского вида

$$\left| \int_{\Omega} fg \, d\Omega \right| \leq \left(\int_{\Omega} f^2 \, d\Omega \right)^{1/2} \left(\int_{\Omega} g^2 \, d\Omega \right)^{1/2}, \quad (2.5.18)$$

а также легко получаемым из него неравенством

$$\left(\int_{\Omega} f \, d\Omega \right)^2 \leq \text{mes}(\Omega) \int_{\Omega} f^2 \, d\Omega, \quad (2.5.19)$$

где $\text{mes}(\Omega)$ есть площадь области Ω .

Правая часть равенства (2.5.17) оценивается следующим образом:

$$\int_{\Omega} \nabla(u - \hat{v}) \nabla(u - \hat{u}) \, d\Omega \leq \left(\int_{\Omega} |\nabla(u - \hat{v})|^2 \, d\Omega \right)^{1/2} \left(\int_{\Omega} |\nabla(u - \hat{u})|^2 \, d\Omega \right)^{1/2}.$$

Тем самым, приходим к неравенству

$$\left(\int_{\Omega} |\nabla(u - \hat{v})|^2 \, d\Omega \right)^{1/2} \leq \left(\int_{\Omega} |\nabla(u - \hat{u})|^2 \, d\Omega \right)^{1/2}$$

или — в эквивалентной записи,

$$\| |\nabla(u - \hat{v})| \|_{0,\Omega} \leq \| |\nabla(u - \hat{u})| \|_{0,\Omega}. \quad (2.5.20)$$

Таким образом, оценка погрешности приближенного решения задачи (2.5.1), (2.5.2) сводится к оценке нормы, стоящей в правой части неравенства (2.5.20). Оценку погрешности в окончательном виде мы получим ниже, в параграфе 2.9. А теперь мы займемся оценкой разности функций $u \in W_2^2(\Omega)$ и их кусочно-линейных восполнений в определенных нормах.

2.6 Теоремы аппроксимации

Начнем с получения некоторых вспомогательных оценок. Рассмотрим в плоскости переменных ξ, η стандартный прямоугольный треугольник Δ с вершинами $(0, 0)$, $(h, 0)$ и $(0, h)$ (Рис. 2.6.1). Пусть $\tilde{u}(\xi, \eta)$ есть линейное восполнение некоторой непрерывной функции $u(\xi, \eta)$ в треугольнике Δ , а именно:

$$\tilde{u}(\xi, \eta) = u(0, 0) + \frac{u(h, 0) - u(0, 0)}{h} \xi + \frac{u(0, h) - u(0, 0)}{h} \eta, \quad (\xi, \eta) \in \Delta. \quad (2.6.1)$$

Обозначим через Π квадрат с вершинами $(0, 0)$, $(0, h)$, (h, h) и $(h, 0)$ — как показано ниже на Рис. 2.6.1. В дальнейшем будем предполагать, что $u(\xi, \eta) \in W_2^2(\Pi)$.

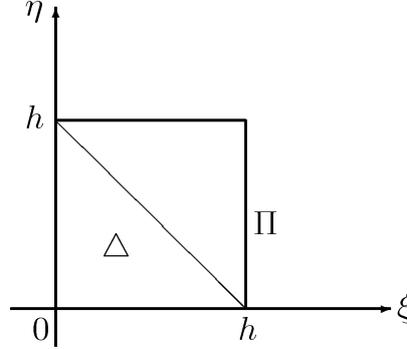


Рис. 2.6.1 Стандартный треугольник Δ .

Лемма 2.6.1 *Справедливо неравенство*

$$\|\nabla(u - \tilde{u})\|_{0,\Delta}^2 \leq 4h^2 \int_{\Pi} \left[\left(\frac{\partial^2 u}{\partial \xi^2} \right)^2 + \left(\frac{\partial^2 u}{\partial \xi \partial \eta} \right)^2 + \left(\frac{\partial^2 u}{\partial \eta^2} \right)^2 \right] d\Pi. \quad (2.6.2)$$

Доказательство. Из представления (2.6.1) функции $\tilde{u}(\xi, \eta)$ имеем

$$\frac{\partial(u - \tilde{u})}{\partial \xi} = \frac{\partial u(\xi, \eta)}{\partial \xi} - \frac{u(h, 0) - u(0, 0)}{h}.$$

Отсюда, путем несложных преобразований, получим следующее выражение для этой производной:

$$\begin{aligned} \frac{\partial(u - \tilde{u})}{\partial \xi} &= \frac{\partial u(\xi, \eta)}{\partial \xi} - \frac{1}{h} \int_0^h \frac{\partial u(\alpha, 0)}{\partial \alpha} d\alpha = \frac{1}{h} \int_0^h \left[\frac{\partial u(\xi, \eta)}{\partial \xi} - \frac{\partial u(\alpha, 0)}{\partial \alpha} \right] d\alpha \\ &= \frac{1}{h} \int_0^h \left[\frac{\partial u(\xi, \eta)}{\partial \xi} - \frac{\partial u(\alpha, \eta)}{\partial \alpha} \right] d\alpha + \frac{1}{h} \int_0^h \left[\frac{\partial u(\alpha, \eta)}{\partial \alpha} - \frac{\partial u(\alpha, 0)}{\partial \alpha} \right] d\alpha \\ &= \frac{1}{h} \int_0^h \int_{\alpha}^{\xi} \frac{\partial^2 u(\beta, \eta)}{\partial \beta^2} d\beta d\alpha + \frac{1}{h} \int_0^h \int_0^{\eta} \frac{\partial^2 u(\alpha, \mu)}{\partial \alpha \partial \mu} d\mu d\alpha. \end{aligned}$$

Далее, применяя неравенство Коши-Буняковского и расширяя пределы интегрирования, найдем, что

$$\left(\frac{\partial(u - \tilde{u})}{\partial \xi} \right)^2 \leq 2h \int_0^h \left(\frac{\partial^2 u(\beta, \eta)}{\partial \beta^2} \right)^2 d\beta + 2 \int_0^h \int_0^h \left(\frac{\partial^2 u(\alpha, \mu)}{\partial \alpha \partial \mu} \right)^2 d\mu d\alpha.$$

Проинтегрируем последнее неравенство по треугольнику Δ . Опять же, расширяя пределы интегрирования, получим неравенство

$$\int_{\Delta} \left(\frac{\partial(u - \tilde{u})}{\partial \xi} \right)^2 d\Delta \leq 2h^2 \int_0^h \int_0^h \left[\left(\frac{\partial^2 u(\xi, \eta)}{\partial \xi^2} \right)^2 + \left(\frac{\partial^2 u(\xi, \eta)}{\partial \xi \partial \eta} \right)^2 \right] d\eta d\xi. \quad (2.6.3)$$

Аналогично оценивается производная по η . Из представления (2.6.1) имеем

$$\frac{\partial(u - \tilde{u})}{\partial \eta} = \frac{\partial u(\xi, \eta)}{\partial \eta} - \frac{u(0, h) - u(0, 0)}{h}.$$

Далее,

$$\begin{aligned} \frac{\partial(u - \tilde{u})}{\partial \eta} &= \frac{\partial u(\xi, \eta)}{\partial \eta} - \frac{1}{h} \int_0^h \frac{\partial u(0, \mu)}{\partial \mu} d\mu = \frac{1}{h} \int_0^h \left[\frac{\partial u(\xi, \eta)}{\partial \eta} - \frac{\partial u(0, \mu)}{\partial \mu} \right] d\mu \\ &= \frac{1}{h} \int_0^h \left[\frac{\partial u(\xi, \eta)}{\partial \eta} - \frac{\partial u(\xi, \mu)}{\partial \mu} \right] d\mu + \frac{1}{h} \int_0^h \left[\frac{\partial u(\xi, \mu)}{\partial \mu} - \frac{\partial u(0, \mu)}{\partial \mu} \right] d\mu \\ &= \frac{1}{h} \int_0^h \int_{\mu}^{\eta} \frac{\partial^2 u(\xi, \nu)}{\partial \nu^2} d\nu d\mu + \frac{1}{h} \int_0^h \int_0^{\xi} \frac{\partial^2 u(\alpha, \mu)}{\partial \mu \partial \alpha} d\alpha d\mu. \end{aligned}$$

По неравенству Коши-Буняковского

$$\left(\frac{\partial(u - \tilde{u})}{\partial \eta} \right)^2 \leq 2h \int_0^h \left(\frac{\partial^2 u(\xi, \nu)}{\partial \nu^2} \right)^2 d\nu + 2 \int_0^h \int_0^{\xi} \left(\frac{\partial^2 u(\alpha, \mu)}{\partial \mu \partial \alpha} \right)^2 d\alpha d\mu.$$

Интегрируя по треугольнику Δ , приходим к неравенству

$$\int_{\Delta} \left(\frac{\partial(u - \tilde{u})}{\partial \eta} \right)^2 d\Delta \leq 2h^2 \int_0^h \int_0^h \left[\left(\frac{\partial^2 u(\xi, \eta)}{\partial \eta^2} \right)^2 + \left(\frac{\partial^2 u(\xi, \eta)}{\partial \xi \partial \eta} \right)^2 \right] d\eta d\xi. \quad (2.6.4)$$

Имея неравенства (2.6.3) и (2.6.4), получим

$$\begin{aligned} \|\nabla(u - \tilde{u})\|_{0, \Delta}^2 &= \int_{\Delta} \left[\left(\frac{\partial(u - \tilde{u})}{\partial \xi} \right)^2 + \left(\frac{\partial(u - \tilde{u})}{\partial \eta} \right)^2 \right] d\Delta \\ &\leq 2h^2 \int_0^h \int_0^h \left[\left(\frac{\partial^2 u}{\partial \xi^2} \right)^2 + 2 \left(\frac{\partial^2 u}{\partial \xi \partial \eta} \right)^2 + \left(\frac{\partial^2 u}{\partial \eta^2} \right)^2 \right] d\eta d\xi, \end{aligned}$$

откуда следует неравенство (2.6.2). \square

Лемма 2.6.2 *Справедливо неравенство*

$$\|u - \tilde{u}\|_{0,\Delta}^2 \leq 15h^4 \int_{\Pi} \left[\left(\frac{\partial^2 u}{\partial \xi^2} \right)^2 + \left(\frac{\partial^2 u}{\partial \xi \partial \eta} \right)^2 + \left(\frac{\partial^2 u}{\partial \eta^2} \right)^2 \right] d\Pi. \quad (2.6.5)$$

Доказательство. Пользуясь записью (2.6.1) функции $\tilde{u}(\xi, \eta)$, имеем

$$\begin{aligned} u(\xi, \eta) - \tilde{u}(\xi, \eta) &= u(\xi, \eta) - u(0, 0) - \frac{u(h, 0) - u(0, 0)}{h} \xi - \frac{u(0, h) - u(0, 0)}{h} \eta \\ &= u(\xi, \eta) - u(\xi, 0) + u(\xi, 0) - u(0, 0) - \frac{u(h, 0) - u(0, 0)}{h} \xi - \frac{u(0, h) - u(0, 0)}{h} \eta \\ &= \int_0^\eta \left[\frac{\partial u(\xi, \mu)}{\partial \mu} - \frac{1}{h} \int_0^h \frac{\partial u(0, \nu)}{\partial \nu} d\nu \right] d\mu + \int_0^\xi \left[\frac{\partial u(\alpha, 0)}{\partial \alpha} - \frac{1}{h} \int_0^h \frac{\partial u(\beta, 0)}{\partial \beta} d\beta \right] d\alpha \\ &= \frac{1}{h} \int_0^\eta \int_0^h \left[\frac{\partial u(\xi, \mu)}{\partial \mu} - \frac{\partial u(0, \nu)}{\partial \nu} \right] d\nu d\mu + \frac{1}{h} \int_0^\xi \int_0^h \left[\frac{\partial u(\alpha, 0)}{\partial \alpha} - \frac{\partial u(\beta, 0)}{\partial \beta} \right] d\beta d\alpha. \end{aligned}$$

Продолжим преобразования:

$$\begin{aligned} u(\xi, \eta) - \tilde{u}(\xi, \eta) &= \frac{1}{h} \int_0^\eta \int_0^h \left[\frac{\partial u(\xi, \mu)}{\partial \mu} - \frac{\partial u(\xi, \nu)}{\partial \nu} + \frac{\partial u(\xi, \nu)}{\partial \nu} - \frac{\partial u(0, \nu)}{\partial \nu} \right] d\nu d\mu \\ &+ \frac{1}{h} \int_0^\xi \int_0^h \left[\frac{\partial u(\alpha, 0)}{\partial \alpha} - \frac{\partial u(\alpha, \eta)}{\partial \alpha} + \frac{\partial u(\alpha, \eta)}{\partial \alpha} - \frac{\partial u(\beta, \eta)}{\partial \beta} + \frac{\partial u(\beta, \eta)}{\partial \beta} - \frac{\partial u(\beta, 0)}{\partial \beta} \right] d\beta d\alpha \\ &= \frac{1}{h} \int_0^\eta \int_0^h \int_\nu^\mu \frac{\partial^2 u(\xi, \rho)}{\partial \rho^2} d\rho d\nu d\mu + \frac{1}{h} \int_0^\eta \int_0^h \int_0^\xi \frac{\partial^2 u(\gamma, \nu)}{\partial \nu \partial \gamma} d\gamma d\nu d\mu \\ &+ \int_0^\xi \int_\eta^0 \frac{\partial^2 u(\alpha, \rho)}{\partial \alpha \partial \rho} d\rho d\alpha + \frac{1}{h} \int_0^\xi \int_0^h \int_\beta^\alpha \frac{\partial^2 u(\gamma, \eta)}{\partial \gamma^2} d\gamma d\beta d\alpha + \frac{1}{h} \int_0^\xi \int_0^h \int_0^\eta \frac{\partial^2 u(\beta, \rho)}{\partial \beta \partial \rho} d\rho d\beta d\alpha. \end{aligned}$$

Далее, применяя неравенство Коши-Буняковского и расширяя пределы интегрирования, получим

$$\begin{aligned} (u(\xi, \eta) - \tilde{u}(\xi, \eta))^2 &\leq 5h^3 \int_0^h \left(\frac{\partial^2 u(\xi, \eta)}{\partial \eta^2} \right)^2 d\eta + 15h^2 \int_0^h \int_0^h \left(\frac{\partial^2 u(\xi, \eta)}{\partial \xi \partial \eta} \right)^2 d\eta d\xi \\ &+ 5h^3 \int_0^h \left(\frac{\partial^2 u(\xi, \eta)}{\partial \xi^2} \right)^2 d\xi. \end{aligned}$$

Остается проинтегрировать полученное неравенство по треугольнику Δ ; расширяя при этом пределы интегрирования, приходим к неравенству

$$\int_{\Delta} (u - \tilde{u})^2 d\Delta \leq 5h^4 \int_0^h \int_0^h \left[\left(\frac{\partial^2 u}{\partial \xi^2} \right)^2 + 3 \left(\frac{\partial^2 u}{\partial \xi \partial \eta} \right)^2 + \left(\frac{\partial^2 u}{\partial \eta^2} \right)^2 \right] d\eta d\xi,$$

из которого следует неравенство (2.6.5). \square

Итак, в доказанных выше леммах нами получены оценки приближения функции ее линейным восполнением для стандартного прямоугольного треугольника. Для вывода аналогичных оценок во всей сеточной области Ω_h мы воспользуемся техникой отображения треугольников триангуляции на стандартный прямоугольный треугольник, которая уже применялась нами выше, в параграфе 2.4.

Рассмотрим некоторый треугольник $\Delta' \in \tau_h$. Пусть вершины треугольника, перенумерованные цифрами 1, 2 и 3, имеют координаты (x_1, y_1) , (x_2, y_2) и (x_3, y_3) , соответственно. Этот треугольник отображается на прямоугольный треугольник Δ в плоскости переменных ξ, η , соответствующими вершинами которого являются $(0, 0)$, $(h, 0)$ и $(0, h)$ (Рис. 2.6.2). Формулы, описывающие как прямое отображение $L_{\Delta'} : \Delta' \rightarrow \Delta$, так и обратное отображение $L_{\Delta} : \Delta \rightarrow \Delta'$, даны в параграфе 2.4 (см. (2.4.10) – (2.4.17)).

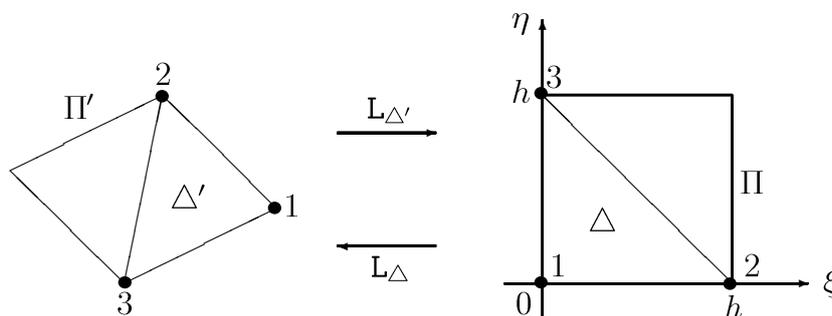


Рис. 2.6.2 Отображение на стандартный треугольник.

При обратном отображении L_{Δ} квадрат Π в плоскости переменных ξ, η преобразуется в параллелограмм Π' в плоскости переменных x, y (Рис. 2.6.2). Это очевидно, так как при линейном отображении параллельность прямых сохраняется. В связи с этим отображением получим одно неравенство, которое будет использовано в последующих оценках.

Пусть $u(x, y) \equiv u(\xi(x, y), \eta(x, y))$. Из (2.4.14) имеем

$$\frac{\partial u}{\partial \xi} = \frac{\partial u}{\partial x} \frac{\partial x}{\partial \xi} + \frac{\partial u}{\partial y} \frac{\partial y}{\partial \xi} = Q_1 \frac{\partial u}{\partial x} + Q_2 \frac{\partial u}{\partial y},$$

$$\frac{\partial u}{\partial \eta} = \frac{\partial u}{\partial x} \frac{\partial x}{\partial \eta} + \frac{\partial u}{\partial y} \frac{\partial y}{\partial \eta} = R_1 \frac{\partial u}{\partial x} + R_2 \frac{\partial u}{\partial y}.$$

Запишем вторые производные:

$$\begin{aligned}\frac{\partial^2 u}{\partial \xi^2} &= Q_1^2 \frac{\partial^2 u}{\partial x^2} + 2Q_1 Q_2 \frac{\partial^2 u}{\partial x \partial y} + Q_2^2 \frac{\partial^2 u}{\partial y^2}, \\ \frac{\partial^2 u}{\partial \xi \partial \eta} &= Q_1 R_1 \frac{\partial^2 u}{\partial x^2} + (Q_1 R_2 + Q_2 R_1) \frac{\partial^2 u}{\partial x \partial y} + Q_2 R_2 \frac{\partial^2 u}{\partial y^2}, \\ \frac{\partial^2 u}{\partial \eta^2} &= R_1^2 \frac{\partial^2 u}{\partial x^2} + 2R_1 R_2 \frac{\partial^2 u}{\partial x \partial y} + R_2^2 \frac{\partial^2 u}{\partial y^2}.\end{aligned}$$

Далее, осуществляя замену переменных под знаком интеграла, находим

$$\begin{aligned}\int_{\Pi} \left[\left(\frac{\partial^2 u}{\partial \xi^2} \right)^2 + \left(\frac{\partial^2 u}{\partial \xi \partial \eta} \right)^2 + \left(\frac{\partial^2 u}{\partial \eta^2} \right)^2 \right] d\Pi &= \int_{\Pi'} \left[\left(Q_1^2 \frac{\partial^2 u}{\partial x^2} + 2Q_1 Q_2 \frac{\partial^2 u}{\partial x \partial y} + Q_2^2 \frac{\partial^2 u}{\partial y^2} \right)^2 \right. \\ &+ \left(Q_1 R_1 \frac{\partial^2 u}{\partial x^2} + (Q_1 R_2 + Q_2 R_1) \frac{\partial^2 u}{\partial x \partial y} + Q_2 R_2 \frac{\partial^2 u}{\partial y^2} \right)^2 \\ &+ \left. \left(R_1^2 \frac{\partial^2 u}{\partial x^2} + 2R_1 R_2 \frac{\partial^2 u}{\partial x \partial y} + R_2^2 \frac{\partial^2 u}{\partial y^2} \right)^2 \right] \frac{D(\xi, \eta)}{D(x, y)} d\Pi' \leq \\ &3 \int_{\Pi'} \left[(Q_1^4 + Q_1^2 R_1^2 + R_1^4) \left(\frac{\partial^2 u}{\partial x^2} \right)^2 + (Q_2^4 + Q_2^2 R_2^2 + R_2^4) \left(\frac{\partial^2 u}{\partial y^2} \right)^2 \right. \\ &+ \left. (4Q_1^2 Q_2^2 + (Q_1 R_2 + Q_2 R_1)^2 + 4R_1^2 R_2^2) \left(\frac{\partial^2 u}{\partial x \partial y} \right)^2 \right] \frac{D(\xi, \eta)}{D(x, y)} d\Pi' .\end{aligned}$$

Если воспользоваться оценкой (2.4.16), то

$$\begin{aligned}\int_{\Pi} \left[\left(\frac{\partial^2 u}{\partial \xi^2} \right)^2 + \left(\frac{\partial^2 u}{\partial \xi \partial \eta} \right)^2 + \left(\frac{\partial^2 u}{\partial \eta^2} \right)^2 \right] d\Pi &\leq \\ &3 \int_{\Pi'} \left[3l_2^4 \left(\frac{\partial^2 u}{\partial x^2} \right)^2 + 3l_2^4 \left(\frac{\partial^2 u}{\partial y^2} \right)^2 + 12l_2^4 \left(\frac{\partial^2 u}{\partial x \partial y} \right)^2 \right] \frac{D(\xi, \eta)}{D(x, y)} d\Pi' .\end{aligned}$$

Окончательно, имея оценку якобиана (2.4.13), получим следующее неравенство:

$$\begin{aligned}\int_{\Pi} \left[\left(\frac{\partial^2 u}{\partial \xi^2} \right)^2 + \left(\frac{\partial^2 u}{\partial \xi \partial \eta} \right)^2 + \left(\frac{\partial^2 u}{\partial \eta^2} \right)^2 \right] d\Pi &\leq \\ &18 \frac{l_2^4}{s_1} \int_{\Pi'} \left[\left(\frac{\partial^2 u}{\partial x^2} \right)^2 + \left(\frac{\partial^2 u}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 u}{\partial y^2} \right)^2 \right] d\Pi' .\end{aligned} \tag{2.6.6}$$

Перейдем к выводу основных оценок аппроксимации функции ее кусочно-линейным восполнением.

Теорема 2.6.1 Пусть $u \in W_2^2(\Omega')$, где Ω' есть достаточно широкая область, включающая сеточную область Ω_h , и $\tilde{u} = \mathbf{prol}(u \in W_2^2(\Omega') : \bar{V}_h)$. Тогда

$$\| |\nabla(u - \tilde{u})| \|_{0,\Omega_h} \leq Ch \|u\|_{2,\Omega'}, \quad (2.6.7)$$

где C есть положительная постоянная, не зависящая от h .

Доказательство. Запишем равенство

$$\| |\nabla(u - \tilde{u})| \|_{0,\Omega_h}^2 = \sum_{\Delta' \in \tau_h} \| |\nabla(u - \tilde{u})| \|_{0,\Delta'}^2. \quad (2.6.8)$$

Для отдельного треугольника $\Delta' \in \tau_h$, при отображении (2.4.10) на стандартный прямоугольный треугольник Δ в плоскости переменных ξ, η , имеем

$$\begin{aligned} \frac{\partial(u - \tilde{u})}{\partial x} &= A_1 \frac{\partial(u - \tilde{u})}{\partial \xi} + A_2 \frac{\partial(u - \tilde{u})}{\partial \eta}, \\ \frac{\partial(u - \tilde{u})}{\partial y} &= B_1 \frac{\partial(u - \tilde{u})}{\partial \xi} + B_2 \frac{\partial(u - \tilde{u})}{\partial \eta}. \end{aligned}$$

Отсюда следуют неравенства

$$\begin{aligned} \left(\frac{\partial(u - \tilde{u})}{\partial x} \right)^2 &\leq 2A_1^2 \left(\frac{\partial(u - \tilde{u})}{\partial \xi} \right)^2 + 2A_2^2 \left(\frac{\partial(u - \tilde{u})}{\partial \eta} \right)^2, \\ \left(\frac{\partial(u - \tilde{u})}{\partial y} \right)^2 &\leq 2B_1^2 \left(\frac{\partial(u - \tilde{u})}{\partial \xi} \right)^2 + 2B_2^2 \left(\frac{\partial(u - \tilde{u})}{\partial \eta} \right)^2. \end{aligned}$$

Тогда

$$\begin{aligned} \| |\nabla(u - \tilde{u})| \|_{0,\Delta'}^2 &= \int_{\Delta'} \left[\left(\frac{\partial(u - \tilde{u})}{\partial x} \right)^2 + \left(\frac{\partial(u - \tilde{u})}{\partial y} \right)^2 \right] d\Delta' \leq \\ &2 \int_{\Delta} \left[(A_1^2 + B_1^2) \left(\frac{\partial(u - \tilde{u})}{\partial \xi} \right)^2 + (A_2^2 + B_2^2) \left(\frac{\partial(u - \tilde{u})}{\partial \eta} \right)^2 \right] \frac{D(x, y)}{D(\xi, \eta)} d\Delta. \end{aligned}$$

Имея оценку (2.4.12) коэффициентов отображения, получим

$$\| |\nabla(u - \tilde{u})| \|_{0,\Delta'}^2 \leq \frac{l_2^2}{s_1^2} \int_{\Delta} \left[\left(\frac{\partial(u - \tilde{u})}{\partial \xi} \right)^2 + \left(\frac{\partial(u - \tilde{u})}{\partial \eta} \right)^2 \right] \frac{D(x, y)}{D(\xi, \eta)} d\Delta,$$

а если воспользоваться оценкой якобиана (2.4.17), то

$$\| |\nabla(u - \tilde{u})| \|_{0,\Delta'}^2 \leq 2 \frac{s_2 l_2^2}{s_1^2} \| |\nabla(u - \tilde{u})| \|_{0,\Delta}^2.$$

Далее, согласно неравенству (2.6.2),

$$\| |\nabla(u - \tilde{u})| \|_{0,\Delta'}^2 \leq 8 \frac{s_2 l_2^2}{s_1^2} h^2 \int_{\Pi} \left[\left(\frac{\partial^2 u}{\partial \xi^2} \right)^2 + \left(\frac{\partial^2 u}{\partial \xi \partial \eta} \right)^2 + \left(\frac{\partial^2 u}{\partial \eta^2} \right)^2 \right] d\Pi.$$

Применяя теперь неравенство (2.6.6), получим

$$\| |\nabla(u - \tilde{u})| \|_{0,\Delta'}^2 \leq 144 \frac{s_2 l_2^6}{s_1^3} h^2 \int_{\Pi'} \left[\left(\frac{\partial^2 u}{\partial x^2} \right)^2 + \left(\frac{\partial^2 u}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 u}{\partial y^2} \right)^2 \right] d\Pi'.$$

Итак, исходя из равенства (2.6.8), приходим к неравенству

$$\| |\nabla(u - \tilde{u})| \|_{0,\Omega_h}^2 \leq 144 \frac{s_2 l_2^6}{s_1^3} h^2 \sum_{\Delta' \in \tau_h} \int_{\Pi'} \left[\left(\frac{\partial^2 u}{\partial x^2} \right)^2 + \left(\frac{\partial^2 u}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 u}{\partial y^2} \right)^2 \right] d\Pi'. \quad (2.6.9)$$

В правой части последнего неравенства, при суммировании по всем треугольникам $\Delta' \in \tau_h$, происходит "налегание" параллелограммов Π' друг на друга. Нетрудно показать, что число таких "налеганий" ограничено некоторой постоянной, не зависящей от h . Попробуем дать грубую оценку. Действительно, любой параллелограмм Π' можно заключить в круг радиуса $r = l_2 h$. Число "налеганий" не превосходит числа треугольников триангуляции, заключенных в круге радиуса $3r$, а именно, числа

$$\frac{\pi(3r)^2}{s_1 h^2} = \frac{9\pi l_2^2}{s_1}. \quad (2.6.10)$$

Тем самым из (2.6.9) получим неравенство

$$\| |\nabla(u - \tilde{u})| \|_{0,\Omega_h}^2 \leq 1296 \pi \frac{s_2 l_2^8}{s_1^4} h^2 \| |u| \|_{2,\Omega'}^2,$$

из которого очевидно следует оценка (2.6.7); при этом в качестве постоянной C в неравенстве (2.6.7) может рассматриваться величина

$$C \approx 64 \frac{\sqrt{s_2} l_2^4}{s_1^2}. \quad (2.6.11)$$

Заметим, что эта величина, по всей видимости, несколько завышена. Ее можно уменьшить, если проводить более тонкие оценки в процессе доказательства. Однако мы этого не делали чтобы излишне не загромождать изложение. \square

Замечание 2.6.1 Как нетрудно проследить по ходу вывода оценки (2.6.7), область $\Omega' \supset \Omega_h$, о которой говорится в условии теоремы 2.6.1, "шире" сеточной области Ω_h на величину порядка h . Иными словами, в качестве области Ω' можно понимать область, ограниченную гладкой кривой, и такую, что "ширина" полосы $\Omega' \setminus \Omega_h$ есть величина порядка h .

Сформулируем и докажем еще одно утверждение.

Теорема 2.6.2 Пусть $u \in W_2^2(\Omega')$, где Ω' есть достаточно широкая область, включающая сеточную область Ω_h , и $\tilde{u} = \text{prol}(u \in W_2^2(\Omega') : \bar{\Omega}_h)$. Тогда

$$\|u - \tilde{u}\|_{0,\Omega_h} \leq Ch^2 \|u\|_{2,\Omega'}, \quad (2.6.12)$$

где C есть положительная постоянная, не зависящая от h

Доказательство. Имеем равенство

$$\|u - \tilde{u}\|_{0,\Omega_h}^2 = \sum_{\Delta' \in \tau_h} \|u - \tilde{u}\|_{0,\Delta'}^2. \quad (2.6.13)$$

Для каждого треугольника $\Delta' \in \tau_h$, при его отображении (2.4.10) на стандартный прямоугольный треугольник Δ в плоскости переменных ξ, η , можем записать

$$\|u - \tilde{u}\|_{0,\Delta'}^2 = \int_{\Delta'} (u - \tilde{u})^2 d\Delta' = \int_{\Delta} (u - \tilde{u})^2 \frac{D(x,y)}{D(\xi,\eta)} d\Delta.$$

Пользуясь оценкой (2.4.17) якобиана, получим

$$\|u - \tilde{u}\|_{0,\Delta'}^2 \leq 2s_2 \|u - \tilde{u}\|_{0,\Delta}^2.$$

Далее, принимая во внимание неравенство (2.6.5),

$$\|u - \tilde{u}\|_{0,\Delta'}^2 \leq 30s_2 h^4 \int_{\Pi} \left[\left(\frac{\partial^2 u}{\partial \xi^2} \right)^2 + \left(\frac{\partial^2 u}{\partial \xi \partial \eta} \right)^2 + \left(\frac{\partial^2 u}{\partial \eta^2} \right)^2 \right] d\Pi,$$

откуда, применяя неравенство (2.6.6), получим

$$\|u - \tilde{u}\|_{0,\Delta'}^2 \leq 540 \frac{s_2 l_2^4}{s_1} h^4 \int_{\Pi'} \left[\left(\frac{\partial^2 u}{\partial x^2} \right)^2 + \left(\frac{\partial^2 u}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 u}{\partial y^2} \right)^2 \right] d\Pi'.$$

Таким образом, исходя из равенства (2.6.13), приходим к неравенству

$$\|u - \tilde{u}\|_{0,\Omega_h}^2 \leq 540 \frac{s_2 l_2^4}{s_1} h^4 \sum_{\Delta' \in \tau_h} \int_{\Pi'} \left[\left(\frac{\partial^2 u}{\partial x^2} \right)^2 + \left(\frac{\partial^2 u}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 u}{\partial y^2} \right)^2 \right] d\Pi'.$$

С учетом оценки (2.6.10) числа "налеганий" параллелограммов Π' друг на друга, получим неравенство

$$\|u - \tilde{u}\|_{0,\Omega_h}^2 \leq 4860 \pi \frac{s_2 l_2^6}{s_1^2} h^4 \|u\|_{2,\Omega'}^2,$$

из которого и следует оценка (2.6.12). В качестве постоянной C может рассматриваться величина

$$C \approx 124 \frac{\sqrt{s_2} l_2^3}{s_1}; \quad (2.6.14)$$

здесь также уместно замечание, сделанное в конце доказательства теоремы 2.6.1. \square

Отметим, что для вновь доказанной теоремы остается в силе приведенное выше замечание 2.6.1.

Имея в распоряжении оценки, полученные в теоремах 2.6.1 и 2.6.2, ниже мы установим аналогичные оценки аппроксимации функций, обращающихся в нуль на границе области, их кусочно-линейными восполнениями из пространства V_h . Для этого нам понадобятся некоторые вспомогательные неравенства.

2.7 Неравенства по приграничным полоскам

В настоящем параграфе будут выведены некоторые неравенства для норм функций в приграничных полосках малой ширины.

Прежде напомним определение кривизны плоской кривой (см., например, [11, 30]). Пусть гладкая кривая S представлена в параметрическом виде

$$x = x(s), \quad y = y(s),$$

где s есть переменная длина дуги. Обозначим через $\beta(s)$ угол, образуемый касательным вектором $\vec{t}(s) = (x'(s), y'(s))$ и положительным направлением оси Ox . Величина

$$\kappa(s) = \beta'(s) \quad (2.7.1)$$

называется *кривизной* кривой S в точке $(x(s), y(s))$. *Радиусом кривизны* называется величина

$$R(s) = \frac{1}{|\kappa(s)|}. \quad (2.7.2)$$

Рассмотрим некоторую область Ω , ограниченную гладкой кривой S длины L . Пусть $K \equiv \max_{0 \leq s \leq L} |\kappa(s)|$. Тогда радиус кривизны кривой в любой точке ограничен снизу величиной $1/K$. Действительно, как следует из (2.7.2),

$$R(s) = \frac{1}{|\kappa(s)|} \geq \frac{1}{K}.$$

Далее, пусть $\rho > 0$ таково, что круги радиуса ρ , касательные к S внутренним образом, не имеют с S иных общих точек. Выбрав положительную величину

$$\delta < \frac{1}{2} \min \left(\frac{1}{K}, \rho \right), \quad (2.7.3)$$

рассмотрим внутреннюю приграничную полоску ω_δ ширины δ , как показано на Рис. 2.7.1.

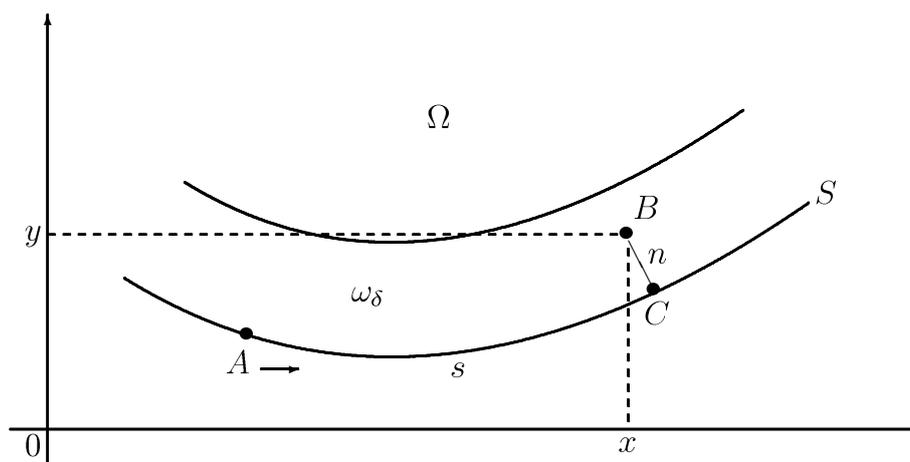


Рис. 2.7.1. Приграничная система координат.

Введем в полоске ω_δ так называемую *приграничную систему координат* (s, n) . Для этого выберем на кривой S некоторую точку A и направление отсчета длины дуг, например, против часовой стрелки. Пусть точка $B(x, y) \in \omega_\delta$. Опустим из точки B перпендикуляр на границу S . Тогда положение точки B однозначно определяется двумя параметрами: длиной n нормали BC и длиной s дуги AC ; при этом $0 \leq n \leq \delta$, $0 \leq s \leq L$.

Пусть

$$x = g_1(s), \quad y = g_2(s)$$

есть параметрическое представление кривой S . В каждой точке границы S определим единичный вектор внутренней нормали $\vec{n}(s) = (n_1(s), n_2(s))$. Тогда координаты точек из ω_δ в системах координат (x, y) и (s, n) связаны соотношениями

$$x = g_1(s) + n_1(s)n, \quad y = g_2(s) + n_2(s)n. \quad (2.7.4)$$

Сформулируем и докажем несколько вспомогательных утверждений.

Лемма 2.7.1 Для якобиана перехода от одной системы координат к другой справедливо равенство

$$\frac{D(x, y)}{D(s, n)} = 1 - \kappa(s)n. \quad (2.7.5)$$

Доказательство. Обозначим через $\alpha(s)$ угол между вектором $\vec{n}(s)$ и положительным направлением оси Ox . Очевидно, что

$$n_1(s) = \cos \alpha(s), \quad n_2(s) = \sin \alpha(s). \quad (2.7.6)$$

Пользуясь соотношениями (2.7.4), имеем

$$\begin{aligned} \frac{D(x, y)}{D(s, n)} &= \begin{vmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial n} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial n} \end{vmatrix} = \begin{vmatrix} g'_1(s) + n'_1(s)n & n_1(s) \\ g'_2(s) + n'_2(s)n & n_2(s) \end{vmatrix} \\ &= (g'_1(s)n_2(s) - g'_2(s)n_1(s)) + (n'_1(s)n_2(s) - n'_2(s)n_1(s))n. \end{aligned}$$

Отсюда, с учетом (2.7.6) находим, что

$$\frac{D(x, y)}{D(s, n)} = (g'_1(s) \sin \alpha(s) - g'_2(s) \cos \alpha(s)) - \alpha'(s)n. \quad (2.7.7)$$

Выше мы дали определение кривизны кривой (см. (2.7.1)). В любой точке границы S касательный вектор $\vec{t}(s)$ и вектор внутренней нормали $\vec{n}(s)$ взаимно перпендикулярны. Нетрудно убедиться в том, что $\kappa(s) = \alpha'(s)$ и $\cos \beta(s) = \sin \alpha(s)$, $\sin \beta(s) = -\cos \alpha(s)$. Далее, исходя из определения производной находим, что

$$\begin{aligned} g'_1(s) &= \lim_{\Delta s \rightarrow 0} \frac{g_1(s + \Delta s) - g_1(s)}{\Delta s} = \cos \beta(s) = \sin \alpha(s), \\ g'_2(s) &= \lim_{\Delta s \rightarrow 0} \frac{g_2(s + \Delta s) - g_2(s)}{\Delta s} = \sin \beta(s) = -\cos \alpha(s). \end{aligned} \quad (2.7.8)$$

В итоге, из (2.7.7) получаем равенство (2.7.5). \square

Как следствие равенства (2.7.5), принимая во внимание выбор (2.7.3) ширины приграничной полосы, легко получить двустороннюю оценку якобиана:

$$\frac{1}{2} \leq \frac{D(x, y)}{D(s, n)} \leq \frac{3}{2}. \quad (2.7.9)$$

Лемма 2.7.2 *Справедливо равенство*

$$|\nabla u|^2 = (1 - \kappa(s)n)^{-2} \left(\frac{\partial u}{\partial s} \right)^2 + \left(\frac{\partial u}{\partial n} \right)^2, \quad (2.7.10)$$

где $u(s, n) = u(x(s, n), y(s, n))$.

Доказательство. Запишем равенства

$$\frac{\partial u}{\partial s} = \frac{\partial u}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial u}{\partial y} \frac{\partial y}{\partial s}, \quad \frac{\partial u}{\partial n} = \frac{\partial u}{\partial x} \frac{\partial x}{\partial n} + \frac{\partial u}{\partial y} \frac{\partial y}{\partial n}. \quad (2.7.11)$$

Имея соотношения (2.7.4) и пользуясь равенствами (2.7.6), (2.7.8), находим, что

$$\frac{\partial x}{\partial s} = \sin \alpha(s)(1 - \kappa(s)n), \quad \frac{\partial y}{\partial s} = -\cos \alpha(s)(1 - \kappa(s)n)$$

и

$$\frac{\partial x}{\partial n} = \cos \alpha(s), \quad \frac{\partial y}{\partial n} = \sin \alpha(s).$$

Подстановка этих выражений в равенства (2.7.11) дает:

$$\begin{aligned} \frac{\partial u}{\partial s} &= (1 - \kappa(s)n) \left(\frac{\partial u}{\partial x} \sin \alpha(s) - \frac{\partial u}{\partial y} \cos \alpha(s) \right), \\ \frac{\partial u}{\partial n} &= \frac{\partial u}{\partial x} \cos \alpha(s) + \frac{\partial u}{\partial y} \sin \alpha(s). \end{aligned}$$

Отсюда легко получаем, что

$$(1 - \kappa(s)n)^{-2} \left(\frac{\partial u}{\partial s} \right)^2 + \left(\frac{\partial u}{\partial n} \right)^2 = \left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial y} \right)^2.$$

Тем самым убеждаемся в справедливости равенства (2.7.10). \square

Перейдем к основным утверждениям этого параграфа. Мы установим ряд неравенств для функций из пространства $W_2^1(\Omega)$.

Теорема 2.7.1 *Справедливо неравенство*

$$\|u\|_{0,S} \leq 2(\delta^{1/2} \|\nabla u\|_{0,\omega_\delta} + \delta^{-1/2} \|u\|_{0,\omega_\delta}). \quad (2.7.12)$$

Доказательство. Запишем тождество

$$u(s, 0) = u(s, n) - \int_0^n \frac{\partial u(s, \nu)}{\partial \nu} d\nu. \quad (2.7.13)$$

Возведем обе части равенства в квадрат и воспользуемся известным неравенством (2.2.14).

В результате получим

$$u^2(s, 0) \leq 2u^2(s, n) + 2n \int_0^n \left(\frac{\partial u(s, \nu)}{\partial \nu} \right)^2 d\nu \leq 2u^2(s, n) + 2\delta \int_0^\delta \left(\frac{\partial u(s, n)}{\partial n} \right)^2 dn.$$

Проинтегрируем это неравенство по приграничной полоске ω_δ . Тогда

$$\int_0^L \int_0^\delta u^2(s, 0) dn ds \leq 2 \int_0^L \int_0^\delta u^2(s, n) dn ds + 2\delta^2 \int_0^L \int_0^\delta \left(\frac{\partial u(s, n)}{\partial n} \right)^2 dn ds,$$

или

$$\delta \int_0^L u^2(s, 0) ds \leq 2 \int_0^L \int_0^\delta u^2 dn ds + 2\delta^2 \int_0^L \int_0^\delta \left(\frac{\partial u}{\partial n} \right)^2 dn ds.$$

Разделим обе части неравенства на величину δ . В результате получим, что

$$\int_0^L u^2(s, 0) ds \leq \frac{2}{\delta} \int_0^L \int_0^\delta u^2 dn ds + 2\delta \int_0^L \int_0^\delta \left(\frac{\partial u}{\partial n} \right)^2 dn ds.$$

Неравенство

$$\int_0^L u^2(s, 0) ds \leq \frac{2}{\delta} \int_0^L \int_0^\delta u^2 dn ds + 2\delta \int_0^L \int_0^\delta \left[(1 - \kappa(s)n)^{-2} \left(\frac{\partial u}{\partial s} \right)^2 + \left(\frac{\partial u}{\partial n} \right)^2 \right] dn ds$$

очевидно следует из предыдущего. Осуществим обратный переход к системе координат (x, y) . С учетом равенства (2.7.10) получим

$$\int_S u^2 dS \leq \frac{2}{\delta} \int_{\omega_\delta} u^2 \frac{D(s, n)}{D(x, y)} d\omega_\delta + 2\delta \int_{\omega_\delta} |\nabla u|^2 \frac{D(s, n)}{D(x, y)} d\omega_\delta.$$

Как следует из (2.7.9),

$$\frac{2}{3} \leq \frac{D(s, n)}{D(x, y)} \leq 2. \quad (2.7.14)$$

Поэтому

$$\int_S u^2 dS \leq \frac{4}{\delta} \int_{\omega_\delta} u^2 d\omega_\delta + 4\delta \int_{\omega_\delta} |\nabla u|^2 d\omega_\delta,$$

и, тем самым, мы приходим к неравенству (2.7.12). \square

Замечание 2.7.1 При выводе неравенства (2.7.12), считая, что $u \in W_2^1(\Omega)$, мы начали с тождества (2.7.13), которое не является корректным при тех s , при которых не определена производная $\frac{\partial u(s, \nu)}{\partial \nu}$. Однако с помощью более тщательных рассуждений можно убедиться в том, что указанное неравенство все же верно. Это утверждение основано на том, что функции из пространства $W_2^1(\Omega)$ могут быть получены как предел последовательности непрерывно дифференцируемых функций. Мы не даем более подробного доказательства, чтобы излишне не загромождать выкладки.

Теорема 2.7.2 *Справедливо неравенство*

$$\|u\|_{0,S} \leq C \|u\|_{1,\Omega}, \quad (2.7.15)$$

где C есть положительная постоянная, не зависящая от функции u .

Доказательство. Требуемое неравенство (2.7.15) легко следует из неравенства (2.7.12). Действительно, зафиксируем некоторое δ_0 , удовлетворяющее условию (2.7.3). Тогда

$$\begin{aligned} \|u\|_{0,S} &\leq 2 \left(\delta_0^{1/2} \|\nabla u\|_{0,\omega_{\delta_0}} + \delta_0^{-1/2} \|u\|_{0,\omega_{\delta_0}} \right) \\ &\leq 2 \max(\delta_0^{1/2}, \delta_0^{-1/2}) \left(\|\nabla u\|_{0,\omega_{\delta_0}} + \|u\|_{0,\omega_{\delta_0}} \right) \\ &\leq 2\sqrt{2} \max(\delta_0^{1/2}, \delta_0^{-1/2}) \|u\|_{1,\Omega}. \end{aligned}$$

При этом в качестве постоянной C в неравенстве (2.7.15) можно рассматривать величину $C = 2\sqrt{2} \max(\delta_0^{1/2}, \delta_0^{-1/2})$. \square

Теорема 2.7.3 *Справедливо неравенство*

$$\|u\|_{0,\omega_\delta} \leq \sqrt{6} \delta \|\nabla u\|_{0,\omega_\delta} + \sqrt{3} \delta^{1/2} \|u\|_{0,S}; \quad (2.7.16)$$

в частности, если $u|_S = 0$, то

$$\|u\|_{0,\omega_\delta} \leq \sqrt{6} \delta \|\nabla u\|_{0,\omega_\delta}. \quad (2.7.17)$$

Доказательство. Перепишем тождество (2.7.13) в виде

$$u(s, n) = \int_0^n \frac{\partial u(s, \nu)}{\partial \nu} d\nu + u(s, 0).$$

Возводя обе части равенства в квадрат и расширяя пределы интегрирования, получим неравенство

$$u^2(s, n) = 2\delta \int_0^\delta \left(\frac{\partial u(s, n)}{\partial n} \right)^2 dn + 2u^2(s, 0).$$

Проинтегрируем это неравенство по приграничной полоске ω_δ . В результате,

$$\begin{aligned} \int_0^L \int_0^\delta u^2(s, n) dn ds &\leq 2\delta^2 \int_0^L \int_0^\delta \left(\frac{\partial u(s, n)}{\partial n} \right)^2 dn ds + 2\delta \int_0^L u^2(s, 0) ds \\ &\leq 2\delta^2 \int_0^L \int_0^\delta \left[(1 - k(s)n)^{-2} \left(\frac{\partial u}{\partial s} \right)^2 + \left(\frac{\partial u}{\partial n} \right)^2 \right] dn ds + 2\delta \int_0^L u^2(s, 0) ds. \end{aligned}$$

Осуществляя обратный переход к системе координат (x, y) , с учетом равенства (2.7.10) получим

$$\int_{\omega_\delta} u^2 \frac{D(s, n)}{D(x, y)} d\omega_\delta \leq 2\delta^2 \int_{\omega_\delta} |\nabla u|^2 \frac{D(s, n)}{D(x, y)} d\omega_\delta + 2\delta \int_S u^2 dS.$$

Воспользуемся теперь двусторонней оценкой якобиана (2.7.14). Тогда

$$\frac{2}{3} \int_{\omega_\delta} u^2 d\omega_\delta \leq 4\delta^2 \int_{\omega_\delta} |\nabla u|^2 d\omega_\delta + 2\delta \int_S u^2 dS$$

или

$$\int_{\omega_\delta} u^2 d\omega_\delta \leq 6\delta^2 \int_{\omega_\delta} |\nabla u|^2 d\omega_\delta + 3\delta \int_S u^2 dS.$$

Тем самым, приходим к неравенству (2.7.16). \square

Теорема 2.7.4 *Справедливо неравенство*

$$\|u\|_{0, \omega_\delta} \leq C\delta^{1/2} \|u\|_{1, \Omega}, \quad (2.7.18)$$

где C есть положительная постоянная, не зависящая от функции u .

Доказательство. Неравенство (2.7.18) является простым следствием неравенств (2.7.15) и (2.7.16). \square

Замечание 2.7.2 При выводе неравенств предполагалось, что граница S области Ω является гладкой кривой. Однако нетрудно показать, что эти оценки, возможно с несколько иными положительными постоянными в качестве коэффициентов, остаются верными также и для случая кусочно-гладких границ. Для этого достаточно разбить приграничную полосу на части, и в каждой части ввести свою приграничную систему координат.

2.8 Теоремы аппроксимации (продолжение)

Как было сказано в параграфе 2.6, для получения оценки погрешности конечноэлементного решения задачи Дирихле для уравнения Пуассона (2.5.1), (2.5.2) нам понадобятся теоремы аппроксимации для функций из пространства $W_2^2(\Omega)$, обращающихся в нуль на границе области. Точнее, мы будем иметь дело с сеточной областью $\Omega_h \subseteq \Omega$, построение которой описано в параграфе 2.5; функции $u \in W_2^2(\Omega)$ аппроксимируются их кусочно-линейными восполнениями из пространства V_h (см. (2.5.6)).

Чтобы излишне не загромождать изложение, при выводе неравенств одной и той же буквой C мы будем обозначать различные положительные постоянные, не зависящие от шага сетки.

Итак, пусть $u \in W_2^2(\Omega)$ и $u|_S = 0$. Согласно принятым в параграфе 2.5 обозначениям, $\hat{u} = \text{prol}(u : V_h)$. Запишем равенства

$$\|u - \hat{u}\|_{0,\Omega}^2 = \|u\|_{0,\Omega \setminus \Omega_h}^2 + \|u - \hat{u}\|_{0,\Omega_h}^2, \quad (2.8.1)$$

$$\|\nabla(u - \hat{u})\|_{0,\Omega}^2 = \|\nabla u\|_{0,\Omega \setminus \Omega_h}^2 + \|\nabla(u - \hat{u})\|_{0,\Omega_h}^2; \quad (2.8.2)$$

здесь мы учли то обстоятельство, что функции из V_h равны тождественно нулю в приграничной области $\Omega \setminus \Omega_h$. По свойству ломаной S_h , ограничивающей сеточную область Ω_h , область $\Omega \setminus \Omega_h$ может быть заключена в приграничную полосу ω_δ , где δ есть величина порядка h^2 . Тогда, пользуясь неравенством (2.7.17), получим

$$\|u\|_{0,\Omega \setminus \Omega_h}^2 \leq \|u\|_{0,\omega_\delta}^2 \leq Ch^4 \|\nabla u\|_{0,\omega_\delta}^2.$$

Далее, имея неравенство (2.7.18), нетрудно убедиться в том, что

$$\|\nabla u\|_{0,\Omega \setminus \Omega_h}^2 \leq \|\nabla u\|_{0,\omega_\delta}^2 \leq Ch^2 \|u\|_{2,\Omega}^2. \quad (2.8.3)$$

В результате, от равенств (2.8.1) и (2.8.2) переходим к неравенствам

$$\|u - \hat{u}\|_{0,\Omega}^2 \leq Ch^6 \|u\|_{2,\Omega}^2 + \|u - \hat{u}\|_{0,\Omega_h}^2, \quad (2.8.4)$$

$$\|\nabla(u - \hat{u})\|_{0,\Omega}^2 \leq Ch^2 \|u\|_{2,\Omega}^2 + \|\nabla(u - \hat{u})\|_{0,\Omega_h}^2. \quad (2.8.5)$$

Далее мы обратимся к теоремам аппроксимации из параграфа 2.6. Действительно, запишем очевидные неравенства

$$\|u - \hat{u}\|_{0,\Omega_h} \leq \|u - \tilde{u}\|_{0,\Omega_h} + \|\tilde{u} - \hat{u}\|_{0,\Omega_h},$$

$$\|\nabla(u - \hat{u})\|_{0,\Omega_h} \leq \|\nabla(u - \tilde{u})\|_{0,\Omega_h} + \|\nabla(\tilde{u} - \hat{u})\|_{0,\Omega_h},$$

где $\tilde{u} = \text{prol}(u : \bar{V}_h)$. Пользуясь оценками (2.6.7) и (2.6.12), находим, что

$$\|u - \hat{u}\|_{0,\Omega_h} \leq Ch^2 \|u\|_{2,\Omega'} + \|\tilde{u} - \hat{u}\|_{0,\Omega_h},$$

$$\|\nabla(u - \hat{u})\|_{0,\Omega_h} \leq Ch \|u\|_{2,\Omega'} + \|\nabla(\tilde{u} - \hat{u})\|_{0,\Omega_h};$$

напомним, что Ω' есть достаточно широкая область, включающая сеточную область Ω_h (см. замечание 2.6.1). Итак, из неравенств (2.8.4) и (2.8.5) получаем неравенства

$$\|u - \hat{u}\|_{0,\Omega}^2 \leq Ch^4 \|u\|_{2,\Omega'}^2 + 2\|\tilde{u} - \hat{u}\|_{0,\Omega_h}^2, \quad (2.8.6)$$

$$\|\nabla(u - \hat{u})\|_{0,\Omega}^2 \leq Ch^2 \|u\|_{2,\Omega'}^2 + \|\nabla(\tilde{u} - \hat{u})\|_{0,\Omega_h}^2. \quad (2.8.7)$$

Согласно определению кусочно-линейных функций \tilde{u} и \hat{u} , их разность $\tilde{u} - \hat{u}$ равна нулю во внутренних узлах сеточной области Ω_h . Тогда, в силу неравенств (2.4.18) и (2.4.25), можем записать, что

$$\|\tilde{u} - \hat{u}\|_{0,\Omega_h}^2 \leq Ch^2 \sum_{\xi_i \in S_h} u_i^2, \quad \|\nabla(\tilde{u} - \hat{u})\|_{0,\Omega_h}^2 \leq C \sum_{\xi_i \in S_h} u_i^2.$$

Тем самым, из (2.8.6) и (2.8.7) получаем неравенства

$$\|u - \hat{u}\|_{0,\Omega}^2 \leq Ch^4 \|u\|_{2,\Omega'}^2 + Ch^2 \sum_{\xi_i \in S_h} u_i^2, \quad (2.8.8)$$

$$\|\nabla(u - \hat{u})\|_{0,\Omega}^2 \leq Ch^2 \|u\|_{2,\Omega'}^2 + C \sum_{\xi_i \in S_h} u_i^2. \quad (2.8.9)$$

Таким образом, нам осталось оценить сумму квадратов значений функции u в граничных узлах сеточной области Ω_h . Для этого в полоске ω_δ перейдем к приграничной системе координат (s, n) . Будем считать, что граница S является гладкой кривой.

Рассмотрим некоторый граничный узел $\xi_i \in S_h$; через $\bar{\xi}_i$ обозначим следующий за этим узлом граничный узел — как показано на Рис. 2.8.1. Пусть узлы ξ_i и $\bar{\xi}_i$ имеют координаты (s_i, n_i) и (\bar{s}_i, \bar{n}_i) , соответственно. Введем обозначение $\Delta s_i \equiv \bar{s}_i - s_i$. Отметим, что в силу свойств треугольников триангуляции Δs_i есть величина порядка h .

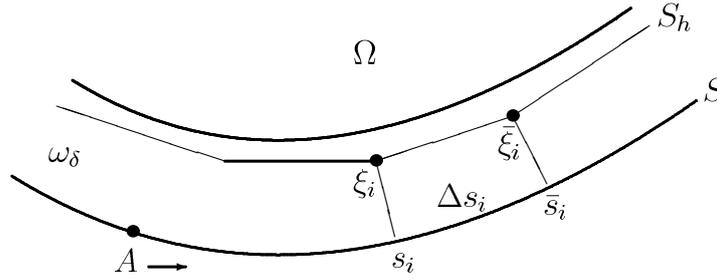


Рис. 2.8.1. Окрестность граничного узла ξ_i .

Принимая во внимание то, что функция u равна нулю на границе S , запишем

$$\begin{aligned} u_i = u(s_i, n_i) &= \int_0^{n_i} \frac{\partial u(s_i + \sigma, n)}{\partial n} dn - \int_{s_i}^{s_i + \sigma} \frac{\partial u(s, n_i)}{\partial s} ds = \int_0^{n_i} \frac{\partial u(s_i + \sigma, n)}{\partial n} dn - \\ &\int_{s_i}^{s_i + \sigma} \left[\frac{\partial u(s, n_i)}{\partial s} - \frac{\partial u(s, 0)}{\partial s} \right] ds = \int_0^{n_i} \frac{\partial u(s_i + \sigma, n)}{\partial n} dn - \int_{s_i}^{s_i + \sigma} \int_0^{n_i} \frac{\partial^2 u(s, n)}{\partial s \partial n} dn ds, \end{aligned}$$

где $0 \leq \sigma \leq \Delta s_i$. Интегрируя это равенство по σ , получим

$$u_i \Delta s_i = \int_0^{\Delta s_i} \int_0^{n_i} \frac{\partial u(s_i + \sigma, n)}{\partial n} dn d\sigma - \int_0^{\Delta s_i} \int_{s_i}^{s_i + \sigma} \int_0^{n_i} \frac{\partial^2 u(s, n)}{\partial s \partial n} dn ds d\sigma,$$

после чего разделим равенство на Δs_i :

$$u_i = \frac{1}{\Delta s_i} \int_0^{\Delta s_i} \int_0^{n_i} \frac{\partial u(s_i + \sigma, n)}{\partial n} dn d\sigma - \frac{1}{\Delta s_i} \int_0^{\Delta s_i} \int_{s_i}^{s_i + \sigma} \int_0^{n_i} \frac{\partial^2 u(s, n)}{\partial s \partial n} dn ds d\sigma.$$

Возводя обе части последнего равенства в квадрат и расширяя пределы интегрирования, получим неравенство

$$\begin{aligned} u_i^2 &\leq \frac{2}{\Delta s_i} \delta \int_{s_i}^{\bar{s}_i} \int_0^\delta \left(\frac{\partial u(s, n)}{\partial n} \right)^2 dn ds + 2\delta \Delta s_i \int_{s_i}^{\bar{s}_i} \int_0^\delta \left(\frac{\partial^2 u(s, n)}{\partial s \partial n} \right)^2 dn ds \\ &\leq C \left(h \int_{s_i}^{\bar{s}_i} \int_0^\delta \left(\frac{\partial u}{\partial n} \right)^2 dn ds + h^3 \int_{s_i}^{\bar{s}_i} \int_0^\delta \left(\frac{\partial^2 u}{\partial s \partial n} \right)^2 dn ds \right). \end{aligned}$$

Суммируя это неравенство по граничным узлам сеточной области Ω_{h_i} , находим, что

$$\sum_{\xi_i \in S_h} u_i^2 \leq C \left(h \int_0^L \int_0^\delta \left(\frac{\partial u}{\partial n} \right)^2 dn ds + h^3 \int_0^L \int_0^\delta \left(\frac{\partial^2 u}{\partial s \partial n} \right)^2 dn ds \right). \quad (2.8.10)$$

Оценим интегралы, стоящие в правой части неравенства (2.8.10). Неравенство

$$\int_0^L \int_0^\delta \left(\frac{\partial u}{\partial n} \right)^2 dn ds \leq 2 \int_{\omega_\delta} |\nabla u|^2 d\omega_\delta$$

фактически уже получено нами при доказательстве теоремы 2.7.1. То же самое можно сказать и о неравенстве

$$\int_{\omega_\delta} |\nabla u|^2 d\omega_\delta \leq Ch^2 \|u\|_{2, \Omega}^2$$

(см. неравенство (2.8.3)). Таким образом,

$$\int_0^L \int_0^\delta \left(\frac{\partial u}{\partial n} \right)^2 dn ds \leq Ch^2 \|u\|_{2, \Omega}^2. \quad (2.8.11)$$

Далее, пользуясь формулами, приведенными в параграфе 2.7, нетрудно показать, что

$$\int_0^L \int_0^\delta \left(\frac{\partial^2 u}{\partial s \partial n} \right)^2 dn ds \leq C \|u\|_{2, \omega_\delta}^2. \quad (2.8.12)$$

Оставляем это читателю в качестве упражнения. В результате, из (2.8.10) – (2.8.12) получаем оценку

$$\sum_{\xi_i \in S_h} u_i^2 \leq Ch^3 \|u\|_{2,\Omega}^2.$$

Замечание 2.8.1 При выводе последней оценки мы предполагали, что граница S области Ω является гладкой кривой. Однако можно показать, что эта оценка остается верной и для случая кусочно-гладких границ. Для этого достаточно разбить приграничную полосу на части, и в каждой части ввести свою систему координат.

Итак, имея неравенства (2.8.8) и (2.8.9), можем записать, что

$$\|u - \hat{u}\|_{0,\Omega}^2 \leq C (h^4 \|u\|_{2,\Omega'}^2 + h^5 \|u\|_{2,\Omega}^2), \quad (2.8.13)$$

$$\|\nabla(u - \hat{u})\|_{0,\Omega}^2 \leq C (h^2 \|u\|_{2,\Omega'}^2 + h^3 \|u\|_{2,\Omega}^2). \quad (2.8.14)$$

Для получения окончательных оценок возвратимся к неравенствам (2.6.7) и (2.6.12), которые сыграли существенную роль в получении неравенств (2.8.13) и (2.8.14). Речь идет об области $\Omega' \supset \Omega_h$ (см. замечание 2.6.1). Внесем некоторые уточнения. Теперь мы будем предполагать, что область Ω' такова, что $\Omega_h \subseteq \Omega \subset \Omega'$. Здесь возникает необходимость в продолжении функции за пределы ее области задания. Пусть область Ω ограничена кусочно-гладкой кривой класса C^2 . Оказывается, что тогда функцию $u \in W_2^2(\Omega)$ можно продолжить с сохранением класса и нормы на более широкую область $\Omega' \supset \Omega$. Иными словами, существует функция $u \in W_2^2(\Omega')$, совпадающая с функцией u в области Ω , и такая, что

$$\|u\|_{2,\Omega'} \leq C \|u\|_{2,\Omega},$$

где C есть некоторая положительная постоянная, не зависящая от функции $u \in W_2^2(\Omega)$. Продолжение функций из различных функциональных пространств рассмотрено в ряде учебников и монографий (см., например, [23, 30]). С учетом сказанного из (2.8.13) и (2.8.14) приходим к оценкам

$$\begin{aligned} \|u - \hat{u}\|_{0,\Omega}^2 &\leq Ch^4 \|u\|_{2,\Omega}^2, \\ \|\nabla(u - \hat{u})\|_{0,\Omega}^2 &\leq Ch^2 \|u\|_{2,\Omega}^2. \end{aligned}$$

Подводя итог, сформулируем основные утверждения настоящего параграфа.

Теорема 2.8.1 Пусть функция $u \in W_2^2(\Omega)$ и $u|_S = 0$. Тогда

$$\|u - \hat{u}\|_{0,\Omega} \leq Ch^2 \|u\|_{2,\Omega}, \quad (2.8.15)$$

где C есть положительная постоянная, не зависящая от функции u .

Теорема 2.8.2 Пусть функция $u \in W_2^2(\Omega)$ и $u|_S = 0$. Тогда

$$\|\nabla(u - \hat{u})\|_{0,\Omega} \leq Ch \|u\|_{2,\Omega}, \quad (2.8.16)$$

где C есть положительная постоянная, не зависящая от функции u .

Теперь мы можем завершить оценку погрешности приближенного конечноэлементного решения задачи Дирихле для уравнения Пуассона, начатую в параграфе 2.5.

2.9 Оценки погрешности приближенного решения

Пусть $u \in \overset{\circ}{W}_2^1(\Omega)$ есть обобщенное решение задачи (2.5.1), (2.5.2), удовлетворяющее интегральному тождеству (2.5.3). Предположим, что выполнены условия, обговоренные в параграфе 2.5, при которых это обобщенное решение принадлежит пространству $W_2^2(\Omega)$ и справедливо неравенство (2.5.4). Далее, пусть $\hat{v} \in V_h$ есть приближенное конечноэлементное решение задачи, определяемое посредством интегрального тождества (2.5.8). В конце параграфа 2.5 мы пришли к неравенству (2.5.20), а именно:

$$\|\nabla(u - \hat{v})\|_{0,\Omega} \leq \|\nabla(u - \hat{u})\|_{0,\Omega}.$$

Для завершения оценки остался всего один шаг. Действительно, пользуясь неравенством (2.8.16), получим

$$\|\nabla(u - \hat{v})\|_{0,\Omega} \leq Ch\|u\|_{2,\Omega},$$

откуда, с учетом неравенства (2.5.4), приходим к следующему утверждению.

Теорема 2.9.1 *Справедлива оценка*

$$\|\nabla(u - \hat{v})\|_{0,\Omega} \leq Ch\|f\|_{0,\Omega}, \quad (2.9.1)$$

где C есть положительная постоянная, не зависящая от f .

Можно оценить погрешность также и в норме пространства $L_2(\Omega)$.

Теорема 2.9.2 *Справедлива оценка*

$$\|u - \hat{v}\|_{0,\Omega} \leq Ch^2\|f\|_{0,\Omega}, \quad (2.9.2)$$

где C есть положительная постоянная, не зависящая от f .

Доказательство. Рассмотрим — наряду с задачей (2.5.1), (2.5.2) — краевую задачу

$$-\Delta w = u - \hat{v}, \quad w|_S = 0. \quad (2.9.3)$$

Так как изменилась лишь правая часть уравнения, то обобщенное решение задачи (2.9.3) также принадлежит пространству $W_2^2(\Omega)$, и справедливо неравенство

$$\|w\|_{2,\Omega} \leq C\|u - \hat{v}\|_{0,\Omega} \quad (2.9.4)$$

(см. неравенство (2.5.4)).

Умножим обе части уравнения (2.9.3) на функцию $u - \hat{v}$ и проинтегрируем полученное равенство по области Ω . Применение формулы интегрирования по частям (1.5.16) дает нам равенство

$$\|u - \hat{v}\|_{0,\Omega}^2 = \int_{\Omega} \nabla w \nabla(u - \hat{v}) d\Omega.$$

Перепишем это равенство в виде

$$\|u - \hat{v}\|_{0,\Omega}^2 = \int_{\Omega} \nabla(w - \hat{w}) \nabla(u - \hat{v}) d\Omega + \int_{\Omega} \nabla \hat{w} \nabla(u - \hat{v}) d\Omega,$$

где $\text{prol}(w \in W_2^2(\Omega) : V_h)$. Отсюда, с учетом равенства (2.5.16), следует, что

$$\|u - \hat{v}\|_{0,\Omega}^2 = \int_{\Omega} \nabla(w - \hat{w}) \nabla(u - \hat{v}) d\Omega.$$

Далее, применяя неравенство Коши-Буняковского, получим

$$\|u - \hat{v}\|_{0,\Omega}^2 \leq \| |\nabla(w - \hat{w})| \|_{0,\Omega} \| |\nabla(u - \hat{v})| \|_{0,\Omega}.$$

Теперь воспользуемся неравенством (2.8.16) — применительно к функции $w \in W_2^2(\Omega)$, а также неравенством (2.9.1). Тогда

$$\|u - \hat{v}\|_{0,\Omega}^2 \leq Ch^2 \|w\|_{2,\Omega} \|f\|_{0,\Omega}.$$

Наконец, имея неравенство (2.9.4), находим, что

$$\|u - \hat{v}\|_{0,\Omega}^2 \leq Ch^2 \|u - \hat{v}\|_{0,\Omega} \|f\|_{0,\Omega},$$

откуда и следует требуемая оценка (2.9.2). \square

Объединяя оценки (2.9.1) и (2.9.2), приходим к следующему утверждению.

Теорема 2.9.3 *Справедлива оценка*

$$\|u - \hat{v}\|_{1,\Omega} \leq Ch \|f\|_{0,\Omega}, \quad (2.9.5)$$

где C есть положительная постоянная, не зависящая от f .

Замечание 2.9.1 Оценки погрешности, приведенные в теоремах 2.9.1, 2.9.2 и 2.9.3, являются точными по порядку. Иначе говоря, если обобщенное решение задачи принадлежит пространству $W_2^2(\Omega)$, то эти оценки нельзя улучшить за счет иного выбора того же числа базисных функций. Обоснование этому можно найти в монографии [23].

В заключение этого параграфа, а также и всей главы, кратко остановимся на некоторых иных типах граничных условий. Выше мы рассматривали уравнение Пуассона (2.5.1) с однородным граничным условием (2.5.2). В случае неоднородного граничного условия

$$u|_S = g$$

построение конечноэлементного решения осуществляется следующим образом. Приближенное решение $\tilde{v} \in V_h$ ищется в виде линейной комбинации

$$\tilde{v}(x, y) = \sum_{\xi_i \in N_h} v_i \varphi_i(x, y) + \sum_{\xi_i \in S_h} v_i \varphi_i(x, y), \quad (2.9.6)$$

в которой коэффициенты v_i , когда $\xi_i \in N_h$, подлежат определению, а в узлах $\xi_i \in S_h$ мы полагаем $v_i = g(\xi_i^c)$, где ξ_i^c есть точка на границе S области Ω , являющаяся основанием нормали, опущенной из узла ξ_i . Фактически, происходит "снос" граничного условия с S на S_h , что, по существу, делается и в случае однородного граничного условия. Во внутренних узлах сетки записывается система уравнений

$$\int_{\Omega_h} \nabla \tilde{v} \nabla \varphi_i d\Omega = \int_{\Omega_h} f \varphi_i d\Omega, \quad i = 1, 2, \dots, n_h,$$

решая которую, находится приближенное решение (2.9.6).

Рассмотрим еще и *третью краевую задачу* для уравнения Пуассона (2.5.1), в которой граничное условие записывается в виде

$$\left(\frac{\partial u}{\partial \mathbf{n}} + \sigma u \right) \Big|_S = 0, \quad (2.9.7)$$

где \mathbf{n} есть единичный вектор внешней нормали к кривой S (эта задача уже обсуждалась нами в параграфе 1.5). Получим интегральное тождество, аналогичное (2.5.3). Для этого умножим обе части уравнения (2.5.1) на произвольную функцию $\varphi \in W_2^1(\Omega)$ и проинтегрируем полученное равенство по всей области:

$$-\int_{\Omega} \Delta u \varphi d\Omega = \int_{\Omega} f \varphi d\Omega.$$

Применяя формулу интегрирования по частям (1.5.16), получим равенство

$$\int_{\Omega} \nabla u \nabla \varphi d\Omega - \int_S \frac{\partial u}{\partial \mathbf{n}} \varphi dS = \int_{\Omega} f \varphi d\Omega$$

(см. вывод равенства (1.5.17)). Отсюда, с учетом граничного условия (2.9.7), приходим к интегральному тождеству

$$\int_{\Omega} \nabla u \nabla \varphi d\Omega + \int_S \sigma u \varphi dS = \int_{\Omega} f \varphi d\Omega \quad \forall \varphi \in W_2^1(\Omega). \quad (2.9.8)$$

Обобщенное решение задачи (2.5.1), (2.9.7) определяется как функция $u \in W_2^1(\Omega)$ для которой справедливо тождество (2.9.8).

Для определения конечноэлементного решения строится сеточная область $\Omega_h \supseteq \Omega$, которая обладает теми же свойствами, что и сеточная область для задачи Дирихле, о которой говорилось в параграфе 2.5 — с той лишь разницей, что область Ω_h включает в себя область Ω . Кусочно-линейную функцию $\tilde{v} \in \bar{V}_h$, удовлетворяющую интегральному тождеству

$$\int_{\Omega} \nabla \tilde{v} \nabla \tilde{w} \, d\Omega + \int_S \sigma \tilde{v} \tilde{w} \, dS = \int_{\Omega} f \tilde{w} \, d\Omega \quad \forall \tilde{w} \in \bar{V}_h, \quad (2.9.9)$$

будем рассматривать как приближенное конечноэлементное решение третьей краевой задачи.

Интегральное тождество (2.9.9) приводит к системе сеточных уравнений

$$Av = g$$

с матрицей $A = [a_{ij}]_{\bar{n}_h \times \bar{n}_h}$, элементами которой являются

$$a_{ij} = \int_{\Omega} \nabla \varphi_i \nabla \varphi_j \, d\Omega + \int_S \sigma \varphi_i \varphi_j \, dS$$

(см. вывод системы (2.5.13) для задачи Дирихле).

В следующей главе мы более подробно изучим структуру и свойства систем конечно-элементных уравнений, а также обсудим некоторые вопросы, связанные с численной реализацией метода.

Глава 3

ПЕРЕОБУСЛАВЛИВАНИЕ КОНЕЧНОЭЛЕМЕНТНЫХ СИСТЕМ

Применение метода конечных элементов к решению дифференциальных уравнений в частных производных порождает системы линейных алгебраических уравнений (системы сеточных уравнений) большой размерности и специальной структуры. Обычно эти системы решаются с помощью итерационных методов. Как правило, системы сеточных уравнений *плохо обусловлены*, что сказывается в быстром росте числа итераций при уменьшении шага сетки. Чтобы исправить это положение, в итерационных методах используются так называемые *переобуславливающие матрицы* или *переобуславливатели*. Основные требования, предъявляемые к переобуславливателям, таковы: во-первых, они должны хорошо аппроксимировать матрицу жесткости исходной конечноэлементной системы и, во-вторых, процедура их обращения должна быть легко осуществимой.

Существуют разные подходы к построению переобуславливателей для эллиптических уравнений второго порядка. В настоящей главе мы изложим один из методов построения так называемых *многосеточных* переобуславливателей, в основе которого лежит идея использования иерархически измельчающихся сеток и внутренних итерационных процедур.

3.1 Ансамблирование матрицы жесткости

В предыдущей главе нами было показано, что при заданной нумерации узлов множества N_h нахождение приближенного конечноэлементного решения задачи Дирихле для уравнения Пуассона (2.5.1) сводится к решению системы (2.5.13) сеточных уравнений

$$Av = g \quad (3.1.1)$$

с матрицей жесткости $A = [a_{ij}]_{n_h \times n_h}$, элементы которой вычисляются следующим образом:

$$a_{ij} = \int_{\Omega_h} \nabla \varphi_i \nabla \varphi_j \, d\Omega_h, \quad i, j = 1, 2, \dots, n_h \quad (3.1.2)$$

(см. (2.5.9)). Там же было сказано, что матрица A является разреженной; действительно, как легко заметить из (3.1.2), если узлы сетки ξ_i и ξ_j не являются соседними, то $a_{ij} = 0$.

Наиболее эффективный способ построения матрицы A , которую называют также *глобальной* матрицей жесткости, базируется на сборке этой матрицы из матриц жесткости отдельных треугольных элементов, вычисляемых достаточно просто. Эта операция носит название *ансамблирования* (от английского *assembling* — сборка, монтаж). Сначала дадим ее общее описание.

Пусть сеточная область Ω_h есть объединение непересекающихся связанных подобластей $\Omega_h^{(k)}$, $k = 1, 2, \dots, p$, каждая из которых, в свою очередь, является объединением некоторого числа треугольников триангуляции. Точнее, пусть

$$\overline{\Omega_h} = \bigcup_{k=1}^p \overline{\Omega_h^{(k)}}; \quad \Omega_h^{(k)} \cap \Omega_h^{(m)} = \emptyset, \quad \text{если } k \neq m. \quad (3.1.3)$$

Обозначим через $N_h^{(k)}$ множество внутренних узлов сетки, принадлежащих замкнутой подобласти $\overline{\Omega_h^{(k)}}$, а через $n_h^{(k)}$ — число узлов в этом множестве. В соответствии с определением, множество $N_h^{(k)}$ можно рассматривать как *сужение* множества узлов N_h на подобласть $\Omega_h^{(k)}$. В связи с этим новым понятием введем в обращение символ *restr* (от английского *restriction* — сужение, ограничение) и запишем, что

$$N_h^{(k)} = \text{restr}(N_h : \Omega_h^{(k)}).$$

Нумерация узлов множества N_h однозначно определяет номера узлов множества $N_h^{(k)}$; пусть это будут узлы с номерами $i_1^{(k)}, i_2^{(k)}, \dots, i_{n_h^{(k)}}^{(k)}$. Ради простоты изложения мы будем предполагать, что $i_1^{(k)} < i_2^{(k)} < \dots < i_{n_h^{(k)}}^{(k)}$.

Через $G_h^{(k)}$ обозначим пространство сеточных функций, определенных на множестве узлов $N_h^{(k)}$. В силу нашего соглашения относительно нумерации узлов, сеточную функцию $v = \{v_{i_r^{(k)}}\}_{r=1}^{n_h^{(k)}} \in G_h^{(k)}$, где $v_{i_r^{(k)}}$ есть значение функции в узле $\xi_{i_r^{(k)}} \in N_h^{(k)}$, мы можем рассматривать также как вектор

$$v = \begin{bmatrix} v_{i_1^{(k)}} \\ v_{i_2^{(k)}} \\ \vdots \\ v_{i_{n_h^{(k)}}^{(k)}} \end{bmatrix}$$

длины $n_h^{(k)}$. Таким образом, $G_h^{(k)} = \text{restr}(G_h : \Omega_h^{(k)})$ есть *сужение* пространства G_h на подобласть $\Omega_h^{(k)}$.

Далее, пусть $V_h^{(k)} \equiv \text{restr}(V_h : \Omega_h^{(k)})$. Для кусочно-линейных функций $\hat{v} \in V_h^{(k)}$ имеет место представление

$$\hat{v}(x, y) = \sum_{r=1}^{n_h^{(k)}} v_{i_r^{(k)}} \varphi_{i_r^{(k)}}(x, y), \quad \text{где } v_{i_r^{(k)}} = \hat{v}(\xi_{i_r^{(k)}}).$$

После введения необходимых понятий и обозначений, связанных с разбиением сеточной области, вернемся к описанию операции ансамблирования. Из формулы (3.1.2) очевидно следует, что

$$a_{ij} = \sum_{k=1}^p \int_{\Omega_h^{(k)}} \nabla \varphi_i \nabla \varphi_j d\Omega_h^{(k)}, \quad i, j = 1, 2, \dots, n_h. \quad (3.1.4)$$

Последняя запись означает, что каждый элемент глобальной матрицы жесткости A может быть получен путем суммирования соответствующих элементов матриц жесткости по отдельным подобластям, а именно, матриц

$$A_k = \left[a_{i_r^{(k)} i_s^{(k)}}^{(k)} \right]_{r,s=1}^{n_h^{(k)}}, \quad (3.1.5)$$

где

$$a_{i_r^{(k)} i_s^{(k)}}^{(k)} = \int_{\Omega_h^{(k)}} \nabla \varphi_{i_r^{(k)}} \nabla \varphi_{i_s^{(k)}} d\Omega_h^{(k)}. \quad (3.1.6)$$

Уточним сказанное на уровне матричных операций.

Прежде всего определим матрицу $Q_k = [q_{ij}^{(k)}]_{n_h \times n_h^{(k)}}$ с элементами

$$q_{ij}^{(k)} = \begin{cases} 1, & \text{если } i = i_j^{(k)}, \\ 0, & \text{если } i \neq i_j^{(k)}, \end{cases} \quad j = 1, 2, \dots, n_h^{(k)}. \quad (3.1.7)$$

Для наглядности выпишем матрицы Q_k и Q_k^T :

$$Q_k = \begin{array}{cccc|c} & 1 & 2 & \dots & n_h^{(k)} & \\ \hline & & & & & 1 \\ & & & & & \vdots \\ 1 & & & & & i_1^{(k)} \\ & & & & & \vdots \\ & 1 & & & & i_2^{(k)} \\ & & & & & \vdots \\ & & & & 1 & i_{n_h}^{(k)} \\ & & & & & \vdots \\ & & & & & n_h \end{array}, \quad Q_k^T = \begin{array}{cccc|c} & 1 & \dots & i_1^{(k)} & \dots & i_2^{(k)} & \dots & i_{n_h}^{(k)} & \dots & n_h & \\ \hline & & & 1 & & & & & & & 1 \\ & & & & & 1 & & & & & 2 \\ & & & & & & & & & & \vdots \\ & & & & & & & 1 & & & n_h^{(k)} \end{array}.$$

Далее, составим произведение $A_k Q_k^T$. Эта матрица размера $n_h^{(k)} \times n_h$ имеет следующую структуру:

$$A_k Q_k^T = \begin{array}{cccccccc} & 1 & \dots & i_1^{(k)} & \dots & i_2^{(k)} & \dots & i_{n_h^{(k)}}^{(k)} & \dots & n_h \\ \begin{array}{c} 1 \\ 2 \\ \vdots \\ n_h^{(k)} \end{array} & \begin{array}{c} \times \\ \times \\ \times \\ \times \end{array} & & & & & & & & \end{array}$$

Меткой \times обозначены элементы (3.1.6) матрицы A_k . Точнее, на пересечении r -ой строки, $1 \leq r \leq n_h^{(k)}$ и столбца с номером $i_s^{(k)}$, где $1 \leq s \leq n_h^{(k)}$, находится элемент $a_{i_r^{(k)}, i_s^{(k)}}^{(k)}$. Все остальные элементы матрицы $A_k Q_k^T$ заведомо равны нулю. Наконец, рассмотрим матрицу $Q_k A_k Q_k^T$ размера $n_h \times n_h$. Эта матрица такова:

$$Q_k A_k Q_k^T = \begin{array}{cccccccc} & 1 & \dots & i_1^{(k)} & \dots & i_2^{(k)} & \dots & i_{n_h^{(k)}}^{(k)} & \dots & n_h \\ \begin{array}{c} 1 \\ \vdots \\ i_1^{(k)} \\ \vdots \\ i_2^{(k)} \\ \vdots \\ i_{n_h^{(k)}}^{(k)} \\ \vdots \\ n_h \end{array} & \begin{array}{c} \\ \\ \times \\ \\ \times \\ \\ \times \\ \\ \times \\ \\ \end{array} & & & & & & & & \end{array}$$

На пересечении строки с номером $i_r^{(k)}$, $1 \leq r \leq n_h^{(k)}$ и столбца с номером $i_s^{(k)}$, $1 \leq s \leq n_h^{(k)}$ расположен элемент $a_{i_r^{(k)}, i_s^{(k)}}^{(k)}$ матрицы A_k . Все остальные элементы матрицы $Q_k A_k Q_k^T$ равны нулю. Фактически, умножение матрицы A_k справа на Q_k^T и слева на Q_k распределяет элементы этой матрицы по соответствующим местам глобальной матрицы жесткости.

Итак, на основании разложения (3.1.4) элементов матрицы A на составляющие, получаем равенство

$$A = \sum_{k=1}^p Q_k A_k Q_k^T. \quad (3.1.8)$$

В этом и заключается суть операции ансамблирования. Будем обозначать эту операцию символом *assem*. В рассмотренном нами случае запишем, что

$$A = \text{assem}\{A_k : 1 \leq k \leq p\}. \quad (3.1.9)$$

В результате умножения матрицы Q_k^T на сеточную функцию $v \in G_h$ получим сеточную функцию $v^{(k)} = Q_k^T v \in G_h^{(k)}$. Действительно,

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n_h} \end{bmatrix} \in G_h \quad \longrightarrow \quad v^{(k)} = \begin{bmatrix} v_{i_1}^{(k)} \\ v_{i_2}^{(k)} \\ \vdots \\ v_{i_{n_h}^{(k)}}^{(k)} \end{bmatrix} \in G_h^{(k)},$$

то есть $v^{(k)} = \text{restr}(v \in G_h : \Omega_h^{(k)})$.

В дальнейшем нам понадобится одно полезное равенство. Пользуясь представлением (3.1.8) глобальной матрицы жесткости A , для любых сеточных функций $v, w \in G_h$ имеем

$$w^T A v = \sum_{k=1}^p w^T Q_k A_k Q_k^T v = \sum_{k=1}^p (Q_k^T w)^T A_k (Q_k^T v) = \sum_{k=1}^p w^{(k)T} A_k v^{(k)}.$$

Таким образом,

$$w^T A v = \sum_{k=1}^p w^{(k)T} A_k v^{(k)}, \quad (3.1.10)$$

где $v^{(k)} = \text{restr}(v \in G_h : \Omega_h^{(k)})$, $w^{(k)} = \text{restr}(w \in G_h : \Omega_h^{(k)})$.

Как было сказано в начале настоящего параграфа, глобальная матрица жесткости A может быть получена путем ансамблирования матриц жесткости отдельных треугольных элементов. После довольно подробного описания самой операции — в общем виде — это утверждение становится полностью понятным. А именно, если в качестве подобластей (3.1.3) рассматривать отдельные треугольники триангуляции, то, согласно (3.1.9),

$$A = \text{assem}\{A_e : e \in \tau_h\}, \quad (3.1.11)$$

где A_e есть матрица жесткости треугольного элемента $e \in \tau_h$

На элементном уровне разбиения сеточной области мы будем использовать все обозначения, связанные с ансамблированием, заменяя лишь метку "(k)" (или "k"), указывающую на подобласть, на метку "e", указывающую на элемент. Например,

$$A_e = [a_{i_r^e i_s^e}^e]_{r,s=1}^{n_h^e},$$

где

$$a_{i_r^e i_s^e}^e = \int_e \nabla \varphi_{i_r^e} \nabla \varphi_{i_s^e} de.$$

(см. (3.1.5) и (3.1.6)). Заметим, что $n_h^e = 3$, если все три вершины треугольника триангуляции e являются внутренними узлами сетки, $n_h^e = 2$, если одна из вершин находится на границе S_h и $n_h^e = 1$, если две вершины лежат на S_h .

Выведем формулы, с помощью которых можно легко вычислять матрицы жесткости треугольных элементов. Рассмотрим треугольник триангуляции $e \in \mathcal{T}_h$, все три вершины которого являются внутренними узлами сетки. Эти вершины перенумерованы цифрами 1, 2 и 3 (ради упрощения записи формул здесь мы используем локальную нумерацию узлов). Соответственно, координаты вершин есть (x_1, y_1) , (x_2, y_2) и (x_3, y_3) . Далее, пусть a, b и c — длины сторон треугольника, а θ_1, θ_2 и θ_3 — величины его углов, как показано на Рис. 3.1.1.

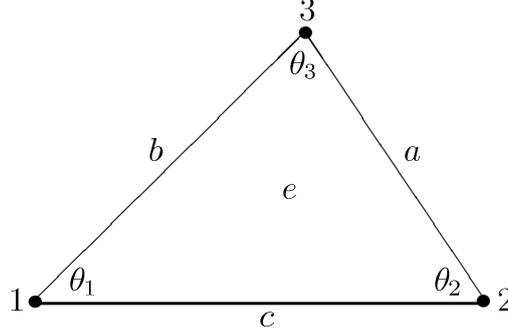


Рис. 3.1.1. Треугольник триангуляции $e \in \mathcal{T}_h$.

Пользуясь выражением (2.4.1) для базисных кусочно-линейных функций, имеем

$$\begin{aligned}\varphi_1(x, y) &= \frac{1}{2S_e} [(y_2 - y_3)x + (x_3 - x_2)y + (x_2y_3 - x_3y_2)], \\ \varphi_2(x, y) &= \frac{1}{2S_e} [(y_3 - y_1)x + (x_1 - x_3)y + (x_3y_1 - x_1y_3)], \\ \varphi_3(x, y) &= \frac{1}{2S_e} [(y_1 - y_2)x + (x_2 - x_1)y + (x_1y_2 - x_2y_1)],\end{aligned}$$

где S_e — площадь треугольника e . Матрица жесткости нашего элемента e есть

$$A_e = [a_{ij}^e]_{i,j=1}^3, \quad (3.1.12)$$

где

$$a_{ij}^e = \int_e \nabla \varphi_i \nabla \varphi_j \, de, \quad i, j = 1, 2, 3. \quad (3.1.13)$$

Вычисляя интегралы, находим

$$\begin{aligned}\int_e |\nabla \varphi_1|^2 &= \frac{a^2}{4S_e}, & \int_e \nabla \varphi_1 \nabla \varphi_2 \, de &= \frac{c^2 - a^2 - b^2}{8S_e}, \\ \int_e |\nabla \varphi_2|^2 &= \frac{b^2}{4S_e}, & \int_e \nabla \varphi_2 \nabla \varphi_3 \, de &= \frac{a^2 - b^2 - c^2}{8S_e}, \\ \int_e |\nabla \varphi_3|^2 &= \frac{c^2}{4S_e}, & \int_e \nabla \varphi_3 \nabla \varphi_1 \, de &= \frac{b^2 - a^2 - c^2}{8S_e}.\end{aligned} \quad (3.1.14)$$

В частности, если треугольник e равносторонний, то

$$A_e = \frac{\sqrt{3}}{6} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

Замечание 3.1.1 Когда $n_h^e = 2$ или $n_h^e = 1$, матрица жесткости элемента получается из матрицы (3.1.12) путем удаления строк и столбцов, номера которых соответствуют узлам, попадающим на границу S_h сеточной области.

В заключение параграфа получим еще одно используемое в дальнейшем равенство. Вернемся к треугольнику триангуляции $e \in \tau_h$ с обозначениями, указанными на Рис. 3.1.1.

Лемма 3.1.1 Для любых функций u и v , линейных на треугольнике e , выполняется равенство

$$\int_e \nabla u \nabla v \, de = \frac{1}{2} [(u_2 - u_1)(v_2 - v_1) \operatorname{ctg} \theta_3 + (u_3 - u_2)(v_3 - v_2) \operatorname{ctg} \theta_1 + (u_1 - u_3)(v_1 - v_3) \operatorname{ctg} \theta_2], \quad (3.1.15)$$

где u_i и v_i , $i = 1, 2, 3$, есть значения соответственно функций u и v в вершинах треугольника.

Доказательство. Линейные функции u и v представимы в виде

$$u(x, y) = \sum_{i=1}^3 u_i \varphi_i(x, y), \quad v(x, y) = \sum_{j=1}^3 v_j \varphi_j(x, y),$$

поэтому

$$\int_e \nabla u \nabla v \, de = \sum_{i,j=1}^3 u_i v_j \int_e \nabla \varphi_i \nabla \varphi_j \, de.$$

В более подробной записи

$$\begin{aligned} \int_e \nabla u \nabla v \, de &= u_1 v_1 \int_e |\nabla \varphi_1|^2 \, de + u_2 v_2 \int_e |\nabla \varphi_2|^2 \, de + u_3 v_3 \int_e |\nabla \varphi_3|^2 \, de \\ &+ (u_1 v_2 + u_2 v_1) \int_e \nabla \varphi_1 \nabla \varphi_2 \, de + (u_2 v_3 + u_3 v_2) \int_e \nabla \varphi_2 \nabla \varphi_3 \, de \\ &+ (u_3 v_1 + u_1 v_3) \int_e \nabla \varphi_3 \nabla \varphi_1 \, de. \end{aligned}$$

Подставляя значения интегралов из (3.1.14), получим

$$\int_e \nabla u \nabla v \, de = \frac{1}{8S_e} [2a^2 u_1 v_1 + 2b^2 u_2 v_2 + 2c^2 u_3 v_3 + (c^2 - a^2 - b^2)(u_1 v_2 + u_2 v_1) + (a^2 - b^2 - c^2)(u_2 v_3 + u_3 v_2) + (b^2 - a^2 - c^2)(u_3 v_1 + u_1 v_3)].$$

После несложных преобразований приходим к равенству

$$\int_e \nabla u \nabla v \, de = \frac{1}{8S_e} [(a^2 + b^2 - c^2)(u_2 - u_1)(v_2 - v_1) + (b^2 + c^2 - a^2)(u_3 - u_2)(v_3 - v_2) + (a^2 + c^2 - b^2)(u_1 - u_3)(v_1 - v_3)].$$

По хорошо известной теореме косинусов

$$a^2 = b^2 + c^2 - 2bc \cos \theta_1, \quad b^2 = a^2 + c^2 - 2ac \cos \theta_2, \quad c^2 = a^2 + b^2 - 2ab \cos \theta_3.$$

Следовательно,

$$\int_e \nabla u \nabla v \, de = \frac{ab}{4S_e} \cos \theta_3 (u_2 - u_1)(v_2 - v_1) + \frac{bc}{4S_e} \cos \theta_1 (u_3 - u_2)(v_3 - v_2) + \frac{ac}{4S_e} \cos \theta_2 (u_1 - u_3)(v_1 - v_3).$$

Осталось воспользоваться также хорошо известными формулами вычисления площади треугольника, а именно,

$$S_e = \frac{1}{2} ab \sin \theta_3, \quad S_e = \frac{1}{2} bc \sin \theta_1, \quad S_e = \frac{1}{2} ac \sin \theta_2,$$

и мы получим требуемое равенство (3.1.15). \square

Следствие 3.1.1 Если треугольник e равносторонний, то

$$\int_e \nabla u \nabla v \, de = \frac{\sqrt{3}}{6} \phi_e(u, v), \tag{3.1.16}$$

где

$$\phi_e(u, v) \equiv (u_2 - u_1)(v_2 - v_1) + (u_3 - u_2)(v_3 - v_2) + (u_1 - u_3)(v_1 - v_3). \tag{3.1.17}$$

Равенство (3.1.16) будет существенно использовано нами ниже при рассмотрении вопросов, связанных с численным решением систем сеточных уравнений.

3.2 Сеточные уравнения и задача переобуславливания

Вернемся к конечноэлементной системе (2.5.13) сеточных уравнений

$$Av = g \quad (3.2.1)$$

с симметричной матрицей. Покажем, что матрица A является положительно определенной. Выше нами было получено равенство (2.5.14), верное для любых сеточных функций и их кусочно-линейных восполнений. Тогда для любой ненулевой сеточной функции $u \in G_h$ имеем

$$u^T Au = \int_{\Omega_h} |\nabla \hat{u}|^2 d\Omega_h > 0.$$

Как известно, при численном решении систем линейных алгебраических уравнений важную роль играет число обусловленности матрицы системы (см., например, [1, 4, 5]). Для симметричных положительно определенных матриц определим это число как

$$\kappa(A) \equiv \frac{\lambda_{max}}{\lambda_{min}}, \quad (3.2.2)$$

где λ_{max} и λ_{min} есть, соответственно, наибольшее и наименьшее из собственных значений матрицы A .

Замечание 3.2.1 Дадим краткое пояснение к определению (3.2.2). Пусть в пространстве матриц порядка n введена некоторая норма. Числом обусловленности невырожденной матрицы называется величина

$$\kappa(A) \equiv \|A\| \|A^{-1}\|.$$

Как видно, значение этой величины зависит от выбора матричной нормы. Норма матрицы, подчиненная некоторой векторной норме, вводится следующим образом:

$$\|A\| = \max_{u \neq 0} \frac{\|Au\|}{\|u\|}$$

(см., например, [1, 7, 29]). Если взять векторную норму

$$\|u\| = \left(\sum_{i=1}^n |u_i|^2 \right)^{1/2},$$

то оказывается, что

$$\|A\| = \sqrt{\lambda_{max}(A^T A)}.$$

Если матрица A симметричная и положительно определенная, то $\|A\| = \lambda_{max}(A)$; в этом случае $\|A^{-1}\| = 1/\lambda_{min}(A)$, и мы получим выражение (3.2.2).

Дадим оценку числа обусловленности матрицы нашей конечноэлементной системы (3.2.1). Для наибольшего и наименьшего собственных значений матрицы имеют место соотношения

$$\lambda_{max} = \max_{u \neq 0} \frac{u^T Au}{u^T u}, \quad \lambda_{min} = \min_{u \neq 0} \frac{u^T Au}{u^T u} \quad (3.2.3)$$

(см., например, [7, 9, 29]).

Для оценки наибольшего собственного значения воспользуемся равенством (2.5.14) и неравенством (2.4.25). С учетом того, что кусочно-линейные функции из пространства V_h равны нулю на границе S_h сеточной области Ω_h , получим неравенство

$$u^T Au = \|\nabla \hat{u}\|_{0, \Omega_h}^2 \leq C \sum_{i=1}^{n_h} u_i^2 = C u^T u,$$

где C — положительная постоянная, не зависящая от h . Отсюда следует, что для любой ненулевой сеточной функции

$$\frac{u^T Au}{u^T u} \leq C.$$

Тем самым, в соответствии с (3.2.3),

$$\lambda_{max} \leq C. \quad (3.2.4)$$

Теперь оценим снизу наименьшее собственное значение. Для этого нам понадобится неравенство Фридрихса (1.5.19). Согласно этому неравенству

$$\|\hat{u}\|_{0, \Omega_h}^2 \leq C \|\nabla \hat{u}\|_{0, \Omega_h}^2$$

(напомним, что буквой C , с индексами и без них, мы обозначаем различные положительные постоянные в неравенствах). Тогда

$$u^T Au = \|\nabla \hat{u}\|_{0, \Omega_h}^2 \geq C \|\hat{u}\|_{0, \Omega_h}^2.$$

Далее воспользуемся неравенством, установленным в (2.4.18). Получим

$$u^T Au \geq Ch^2 \sum_{i=1}^{n_h} u_i^2 = Ch^2 u^T u,$$

откуда следует, что для любой ненулевой сеточной функции

$$\frac{u^T Au}{u^T u} \geq Ch^2.$$

Итак, согласно (3.2.3),

$$\lambda_{min} \geq Ch^2. \quad (3.2.5)$$

В результате, из (3.2.2), (3.2.4) и (3.2.5) получаем оценку

$$\kappa(A) \leq Ch^{-2}. \quad (3.2.6)$$

числа обусловленности матрицы жесткости A . Заметим, что эта оценка точна по порядку величины h^{-1} (см., например, [2, 25]).

Системы сеточных уравнений, к которым сводится численное решение дифференциальных уравнений, как правило, имеют высокий порядок, пропорциональный числу узлов сетки. Это связано с тем, что для получения высокой точности аппроксимации надо брать малый шаг сетки. В рассматриваемой нами задаче это видно из оценок (2.9.1) и (2.9.2).

Алгебраические системы линейных уравнений с большим числом обусловленности матрицы называют *плохо обусловленными*.

Итак, перечислим, какими характерными особенностями обладает наша конечно-элементная система сеточных уравнений:

- система имеет высокий порядок;
- матрица системы разреженная;
- система, согласно оценке (3.2.6), плохо обусловлена.

Поясним, какие вопросы, связанные с высоким порядком и плохой обусловленностью, возникают при численном решении задачи. Для решения систем сеточных уравнений применяются в основном итерационные методы. Например *чебышевский метод* или *метод сопряженных градиентов* [25, 33, 47]. Но здесь мы изложим суть проблемы на более простом методе. Пусть система уравнений (3.2.1) решается с помощью итерационного метода

$$\frac{v^{m+1} - v^m}{\tau} = -Av^m + g, \quad m = 0, 1, 2, \dots \quad (3.2.7)$$

с параметром

$$\tau = \frac{2}{\lambda_{max} + \lambda_{min}}, \quad (3.2.8)$$

где λ_{max} и λ_{min} – соответственно наибольшее и наименьшее из собственных значений матрицы A . Этот метод часто называют *методом Рундсона*¹. Показано, что для получения решения с точностью $\varepsilon > 0$, то есть $\|v^m - v\| < \varepsilon$, надо проделать

$$m \simeq \frac{\kappa(A)}{2} \ln \frac{1}{\varepsilon} \quad (3.2.9)$$

итераций (см., например, [1, 25]). Как видно из оценки (3.2.6), число обусловленности, а вместе с ним и число итераций, быстро растет с уменьшением шага сетки. А это приводит к увеличению числа арифметических операций и, в конечном итоге, к увеличению машинного времени, требуемого для решения системы.

¹Ричардсон Льюис Фрай (1881–1953) – английский геофизик и математик, основные научные труды посвящены математической физике и вычислительной математике.

Для исправления представленной ситуации в рассмотрение вводится так называемая *переобуславливающая матрица* или *переобуславливатель*, и вместо системы (3.2.1) решается система

$$P^{-1}Av = P^{-1}g. \quad (3.2.10)$$

Предполагается, что матрица P также симметричная и положительно определенная. Соответственно, метод Ричардсона для системы (3.2.10) записывается как

$$\frac{v^{m+1} - v^m}{\tau} = -P^{-1}Av^m + P^{-1}g, \quad m = 0, 1, 2, \dots \quad (3.2.11)$$

или, в более приемлемом, с точки зрения организации вычислений, виде

$$P \frac{v^{m+1} - v^m}{\tau} = -Av^m + g, \quad m = 0, 1, 2, \dots \quad (3.2.12)$$

Здесь следует дать некоторые комментарии. Матрицы A и P – симметричные и положительно определенные, но матрица $P^{-1}A$, вообще говоря, не является симметричной. Однако легко показать, что задача на собственные значения для матрицы $P^{-1}A$ равносильна *обобщенной задаче на собственные значения*

$$Au = \lambda Pu, \quad (3.2.13)$$

собственные числа которой действительны и положительны, причем

$$\lambda_{max} = \max_{u \neq 0} \frac{u^T Au}{u^T Pu}, \quad \lambda_{min} = \min_{u \neq 0} \frac{u^T Au}{u^T Pu} \quad (3.2.14)$$

(см., например, [7, 9]). При этом для всех $u \in G_h$ справедливы неравенства

$$\lambda_{min} u^T Pu \leq u^T Au \leq \lambda_{max} u^T Pu. \quad (3.2.15)$$

На основе вышеизложенного под числом обусловленности матрицы $P^{-1}A$ мы будем понимать величину

$$\kappa(P^{-1}A) \equiv \frac{\lambda_{max}}{\lambda_{min}}, \quad (3.2.16)$$

где λ_{max} и λ_{min} есть, соответственно, наибольшее и наименьшее из собственных чисел задачи (3.2.13).

В переобусловленном методе Ричардсона (3.2.12) параметр τ выбирается опять же по формуле (3.2.8), – но уже с собственными числами обобщенной задачи. Для получения решения с точностью $\varepsilon > 0$, по аналогии с (3.2.9), потребуется уже

$$m \simeq \frac{\kappa(P^{-1}A)}{2} \ln \frac{1}{\varepsilon} \quad (3.2.17)$$

итераций. Остановимся еще на одном важном моменте. Здесь переход от приближения v^m к приближению v^{m+1} осуществляется в следующей последовательности:

$$\begin{aligned}
 & \bullet \quad r^m = -Av^m + g, \\
 & \bullet \quad Pz^m = r^m, \\
 & \bullet \quad v^{m+1} = v^m + \tau z^m.
 \end{aligned} \tag{3.2.18}$$

Как видим, возникает задача — решение системы с матрицей P . Поэтому к переобуславливателю P предъявляются два основных требования:

— матрица P должна хорошо аппроксимировать матрицу A в том смысле, что число обусловленности $\kappa(P^{-1}A)$ должно быть существенно меньше числа обусловленности $\kappa(A)$ (см. оценки числа итераций (3.2.9) и (3.2.17));

— в процессе реализации итерационного метода с переобуславливателем приходится неоднократно решать системы с матрицей P (см. (3.2.18)), поэтому эта операция должна быть легко осуществимой.

В силу перечисленных требований, матрицу P рассматривают в качестве *оптимального переобуславливателя* для матрицы A , если:

— матрица P *спектрально эквивалентна* матрице A , то есть существуют такие постоянные c_1 и c_2 , не зависящие от порядка матрицы, что

$$c_1 u^T P u \leq u^T A u \leq c_2 u^T P u \quad \forall u \in G_h; \tag{3.2.19}$$

тогда, как следует из (3.2.15) и (3.2.16), $\kappa(P^{-1}A) \leq c_2/c_1$, и число итераций, необходимых для получения решения системы (3.2.10) с заданной точностью, не зависит от порядка этой системы;

— число арифметических операций, требуемых для решения одной системы с матрицей P , пропорционально числу неизвестных.

Методы и техника построения переобуславливателей для матриц, возникающих при конечноэлементной дискретизации эллиптических уравнений второго порядка, довольно разнообразны. Этим вопросам посвящена обширная литература — статьи, обзоры, учебники и монографии (см., например, [33, 39, 40, 43, 48]). Для построения оптимальных переобуславливателей хорошо приспособлены *многоуровневые/многосеточные процедуры с внутренними итерациями*, один из вариантов которых мы изложим в последующих параграфах. Реализующие такой подход методы были предложены практически одновременно в работах У.Аксельсона и П.Вассилевского [36, 37] и Ю.А.Кузнецова [44]. Направление, представленное в статьях [36, 37], и развитое в последующих публикациях этих же авторов, получило название *алгебраических методов многоуровневых итераций* или *AMLI-методов* (Algebraic Multilevel Iteration Methods). Второе направление, начало которому было положено в [44], — это *алгебраические многосеточные методы с разбиениями на подструктуры* или *AM/S-методы* (Algebraic Multigrid/Substructuring Methods). Отличительной чертой обоих направлений, между которыми существует

тесная внутренняя связь, является возможность построения переобуславливателей чисто алгебраическим путем, используя строящуюся по определенному принципу иерархическую последовательность сеток. Свое дальнейшее развитие оба направления получили в работах Ю.А.Кузнецова [45], Ю.Р.Акопяна и Ю.А.Кузнецова [42], У.Аксельсона, Ю.Р.Акопяна и Ю.А.Кузнецова [35] и др.

Итак, перейдем к изложению AM/S-версии построения многосеточных переобуславливателей.

3.3 Иерархические треугольные сетки

В параграфе 2.5 нами было дано построение сеточной области $\Omega_h \subseteq \Omega$, триангулированной с некоторым достаточно малым "усредненным" шагом h , который обеспечивал нужную точность приближенного конечноэлементного решения задачи Дирихле для уравнения Пуассона – в соответствии с оценками, полученными в теоремах 2.9.1 и 2.9.2. Однако здесь мы подойдем к построению сетки с несколько иных позиций.

Пусть область Ω , где рассматривается задача (2.5.1), (2.5.2), достаточно хорошо аппроксимирована некоторой многоугольной областью $\Omega' \subseteq \Omega$, которая разбита на треугольники Δ_m , $m = 1, 2, \dots, l$. Предполагается, что любые два треугольника либо не пересекаются, либо имеют только общую вершину, либо только общую сторону. Процесс построения *иерархической* последовательности сгущающихся треугольных сеток основан на следующей процедуре измельчения: каждый из треугольников, образующих уже имеющуюся триангуляцию, с помощью отрезков, соединяющих середины его сторон, разбивается на четыре треугольника – как показано на Рис. 3.3.1.

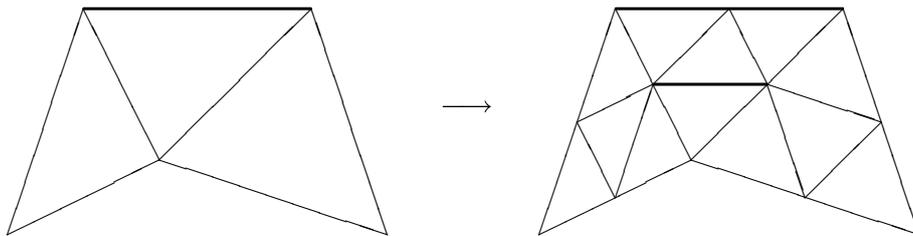


Рис. 3.3.1. Процедура измельчения сетки.

Имея разбиение многоугольной области Ω' на треугольники Δ_m , $m = 1, 2, \dots, l$, мы фактически имеем некоторую триангуляцию области и соответствующую сетку. Начнем процесс измельчения, руководствуясь сформулированным выше правилом. Возьмем сетку ω_0 , полученную после нескольких шагов, в качестве начальной в наших рассуждениях. Мы будем предполагать, что сетка ω_0 содержит небольшое число внутренних узлов. Далее, ограничившись некоторым целым числом $p \geq 1$, продолжим процесс и построим конечную последовательность измельчающихся сеток ω_k , $k = 0, 1, \dots, p$. При этом будем говорить,

что сетка ω_k соответствует k -му уровню измельчения. Таким образом, на p -ом уровне имеем самую мелкую сетку ω_p , а на нулевом уровне — самую грубую сетку ω_0 . Заметим, что сетка ω_p — это та сетка, на которой мы будем решать конечноэлементную задачу.

Для всех значений $k = 0, 1, \dots, p$ введем следующие обозначения:

τ_k — множество треугольников триангуляции k -го уровня измельчения сетки;

N_k — множество внутренних узлов сетки ω_k ;

n_k — число узлов в множестве N_k ;

G_k — пространство сеточных функций, определенных на множестве узлов N_k ;

V_k — пространство кусочно-линейных на сетке ω_k функций, обращающихся в нуль на границе многоугольной области Ω' .

В работе [38] показано, что при каждом измельчении сетки число внутренних узлов увеличивается приблизительно в четыре раза. Этим мы воспользуемся ниже — при оценке вычислительной сложности реализации многосеточного переобуславливателя.

В связи с иерархической структурой сеток уточним порядок нумерации их узлов. По построению

$$N_k \supset N_{k-1}, \quad k = 1, 2, \dots, p. \quad (3.3.1)$$

Тем самым, для значений $k = 1, 2, \dots, p$ мы имеем естественное разбиение множества узлов N_k на два подмножества:

$$N_k = N_k^{(1)} \cup N_k^{(2)}, \quad (3.3.2)$$

где

$$N_k^{(1)} \equiv N_{k-1}, \quad N_k^{(2)} \equiv N_k \setminus N_{k-1} \quad (3.3.3)$$

(см. Рис. 3.3.2).

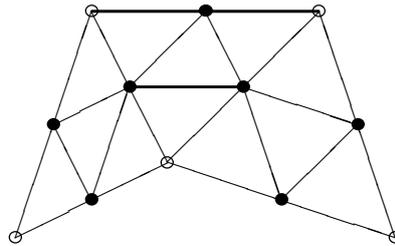


Рис. 3.3.2. Разбиение множества N_k : \circ — узлы подмножества $N_k^{(1)}$,
 \bullet — узлы подмножества $N_k^{(2)}$.

Обозначим через $n_k^{(i)}$, $i = 1, 2$ число узлов в множестве $N_k^{(i)}$; согласно (3.3.3), можем записать, что

$$n_k^{(1)} = n_{k-1}, \quad n_k^{(2)} = n_k - n_{k-1}. \quad (3.3.4)$$

Основываясь на разбиении (3.3.2), установим следующий порядок нумерации узлов множества N_k : сначала числами от 1 до n_{k-1} нумеруются узлы множества $N_k^{(1)}$, а затем числами от $n_{k-1} + 1$ до n_k — узлы множества $N_k^{(2)}$.

Для значений $k = 1, 2, \dots, p$, в соответствии с принятым правилом нумерации узлов, произвольную сеточную функцию $u \in G_k$ — рассматриваемую как вектор — можно представить в виде

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \begin{matrix} \} N_k^{(1)} \\ \} N_k^{(2)} \end{matrix}; \quad u_i \in G_k^{(i)}, \quad i = 1, 2, \quad (3.3.5)$$

где $G_k^{(i)}$ есть пространство сеточных функций, определенных на множестве узлов $N_k^{(i)}$.

Как уже было сказано выше в настоящем параграфе, мы будем решать задачу Дирихле для уравнения Пуассона на p -ом уровне измельчения, который соответствует самой мелкой сетке ω_p . В связи с этим уточним определение (2.5.8) конечноэлементного решения задачи — применительно к новым обозначениям. Функцию $\hat{v} \in V_p$, удовлетворяющую интегральному тождеству

$$\int_{\Omega'} \nabla \hat{v} \nabla \hat{w} \, d\Omega' = \int_{\Omega'} f \hat{w} \, d\Omega' \quad \forall \hat{w} \in V_p, \quad (3.3.6)$$

и продолженную нулем на приграничную подобласть $\Omega \setminus \Omega'$, будем рассматривать как конечноэлементное решение задачи (2.5.1), (2.5.2). Интегральное тождество (3.3.6) приводит к системе сеточных уравнений

$$Av = g \quad (3.3.7)$$

с симметричной положительно определенной матрицей A порядка n_p и правой частью $g \in G_p$. Элементами матрицы жесткости

$$A = [a_{ij}]_{n_p \times n_p}$$

являются

$$a_{ij} = \int_{\Omega'} \nabla \varphi_i^{(p)} \nabla \varphi_j^{(p)} \, d\Omega', \quad i, j = 1, 2, \dots, n_p,$$

где $\varphi_i^{(p)}$ есть базисные кусочно-линейные функции пространства V_p (см. (2.5.9)). При этом, как следует из (2.5.14), справедливо равенство

$$w^T Au = \int_{\Omega'} \nabla \hat{u} \nabla \hat{w} \, d\Omega' \quad \forall u, w \in G_p, \quad (3.3.8)$$

где $\hat{u} = \text{prol}(u \in G_p : V_p)$ и $\hat{w} = \text{prol}(w \in G_p : V_p)$.

Основная цель настоящей главы — построение оптимального переобуславливателя для матрицы A системы сеточных уравнений (3.3.7). Эта задача решается нами в два этапа. Вначале мы перейдем к специальным образом определенной матрице, спектрально эквивалентной матрице A (см. определение (3.2.19)); затем для новой матрицы, используя иерархическую структуру сеток, строится многосеточный переобуславливатель, который будет рассматриваться в качестве переобуславливателя и для исходной матрицы.

3.4 Спектрально эквивалентная матрица

Начнем с того, что мы имеем начальное разбиение многоугольной области Ω' на треугольники Δ_m , $m = 1, 2, \dots, m$. Предлагаемый нами способ построения переобуславливателей основан на линейных отображениях треугольников Δ_m , осуществляемых независимо друг от друга, на стандартный единичный равносторонний треугольник. Предварительно введем некоторые вспомогательные понятия и обозначения. В качестве иллюстрации ниже приводится Рис. 3.4.1.

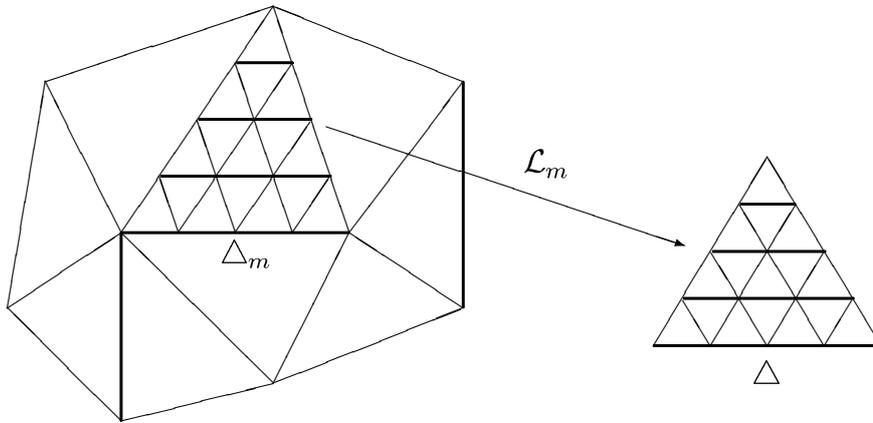


Рис. 3.4.1. Линейное отображение \mathcal{L}_m .

В силу правила генерирования последовательности иерархических сеток каждая сетка ω_k , $0 \leq k \leq p$ задает сетку ω_k^m в треугольнике Δ_m . Это есть не что иное, как сужение сетки ω_k на треугольник Δ_m (см. параграф 3.1). Узлами сетки ω_k^m являются узлы сетки ω_k , принадлежащие $\overline{\Delta_m}$. Таким образом,

$$\omega_k^m = \text{restr}(\omega_k : \Delta_m).$$

Далее, пусть:

τ_k^m — сужение множества треугольников триангуляции k -го уровня на треугольник Δ_m , то есть

$$\tau_k^m = \text{restr}(\tau_k : \Delta_m);$$

N_k^m — подмножество множества N_k , состоящее из узлов, принадлежащих $\overline{\Delta_m}$; иначе говоря,

$$N_k^m = \text{restr}(N_k : \Delta_m);$$

мы будем считать, что нумерация узлов множества N_k^m наследует нумерацию узлов множества N_k ;

n_k^m — число узлов в множестве N_k^m ;

G_k^m — пространство сеточных функций, заданных на множестве узлов N_k^m , то есть

$$G_k^m = \text{restr}(G_k : \Delta_m);$$

V_k^m — сужение пространства V_k на треугольник Δ_m :

$$V_k^m = \text{restr}(V_k : \Delta_m)$$

(кусочно-линейные на сетке ω_k^m функции, обращающиеся в нуль на $\partial\Delta_m \cap \partial\Omega'$).

Следуя результатам параграфа 3.1, укажем, что матрица жесткости A конечноэлементной системы (3.3.7) может быть получена как результат ансамблирования матриц жесткости для отдельных треугольников, в совокупности образующих область Ω' . А именно, определим для каждого треугольника Δ_m , $m = 1, 2, \dots, m$ матрицу A_m порядка n_p^m с помощью соотношения

$$w^T A_m u = \int_{\Delta_m} \nabla \hat{u} \nabla \hat{w} d\Delta_m \quad \forall u, w \in G_p^m, \quad (3.4.1)$$

где $\hat{u} = \text{prol}(u \in G_p^m : V_p^m)$, $\hat{w} = \text{prol}(w \in G_p^m : V_p^m)$. Тогда

$$A = \text{assem}\{A_m : 1 \leq m \leq l\}. \quad (3.4.2)$$

Заметим, что матрица A_m является положительно полуопределенной, если треугольник Δ_m не имеет с границей многоугольной области Ω' общих точек; в противном случае матрица A_m — положительно определенная.

Итак, рассмотрим некоторый треугольник Δ_m , $1 \leq m \leq l$. Осуществим линейное отображение \mathcal{L}_m , которое переводит Δ_m в равносторонний треугольник Δ с длиной стороны, равной единице. На каждом уровне измельчения это отображение преобразует сетку ω_k^m в регулярную сетку $\omega_k^m(\Delta)$, ячейками которой являются равносторонние треугольники с длиной стороны h_k (Рис. 3.4.1). По построению

$$h_k = 2^{-k} h_0, \quad k = 1, 2, \dots, p. \quad (3.4.3)$$

Триангуляция τ_k^m треугольника Δ_m определяет регулярную триангуляцию $\tau_k^m(\Delta)$ треугольника Δ . Заметим, что триангуляции $\tau_k^m(\Delta)$ — при различных значениях m — геометрически ничем не отличаются; однако, ради строгости изложения, мы будем сохранять в обозначениях указание на прообразы этих триангуляций. То же самое можно сказать и о сетке $\omega_k^m(\Delta)$.

Далее, линейное отображение \mathcal{L}_m переводит множество узлов N_k^m сетки ω_k^m в множество узлов $N_k^m(\Delta)$ сетки $\omega_k^m(\Delta)$. Особо отметим, что при этом сохраняется нумерация узлов, то есть узлы множества $N_k^m(\Delta)$ имеют те же номера, что и их прообразы в N_k^m . Тогда пространство сеточных функций G_k^m , заданных на множестве N_k^m , можно рассматривать одновременно как пространство сеточных функций, заданных на $N_k^m(\Delta)$.

Наконец, определим пространство $V_k^m(\Delta)$ как образ пространства V_k^m при отображении \mathcal{L}_m . Иными словами, $V_k^m(\Delta)$ есть пространство кусочно-линейных на сетке $\omega_k^m(\Delta)$ функций, обращающихся в нуль на образе пересечения $\partial\Delta_m \cap \partial\Omega'$.

Перейдем к построению матрицы, спектрально эквивалентной матрице A системы сеточных уравнений (3.3.7). Вначале для всех $m = 1, 2, \dots, l$ определим матрицы L_m порядка n_p^m с помощью соотношения

$$w^T L_m u = \int_{\Delta} \nabla \hat{u} \nabla \hat{w} d\Delta \quad \forall u, w \in G_p^m, \quad (3.4.4)$$

где $\hat{u} = \text{prol}(u \in G_p^m : V_p^m(\Delta))$, $\hat{w} = \text{prol}(w \in G_p^m : V_p^m(\Delta))$. Так же, как и в случае с матрицей A_m , матрица L_m положительно полуопределенная, если треугольник Δ_m не имеет с границей $\partial\Omega'$ общих точек; в противном случае матрица L_m — положительно определенная. Далее, взяв некоторый набор параметров σ_m , $m = 1, 2, \dots, l$, с помощью операции ансамблирования построим матрицу

$$L = \text{assem}\{\sigma_m L_m : 1 \leq m \leq l\} \quad (3.4.5)$$

порядка n_p . Матрица L является симметричной и положительно определенной. Прежде чем перейти к вопросу выбора параметров σ_m , сформулируем несколько вспомогательных утверждений, справедливость которых легко устанавливается.

Лемма 3.4.1 Пусть неотрицательные числа a_i и b_i , $i = 1, 2, \dots, s$ удовлетворяют следующим условиям: 1) $b_1 + b_2 + \dots + b_s > 0$; 2) если $b_i = 0$ для некоторого i , то и $a_i = 0$. Тогда

$$\frac{\sum_{i=1}^s a_i}{\sum_{i=1}^s b_i} \leq \max_{\substack{1 \leq i \leq s, \\ b_i \neq 0}} \frac{a_i}{b_i}. \quad (3.4.6)$$

Лемма 3.4.2 Пусть неотрицательные числа a_i и b_i , $i = 1, 2, \dots, s$ удовлетворяют следующим условиям: 1) $b_1 + b_2 + \dots + b_s > 0$; 2) если $a_i = 0$ для некоторого i , то и $b_i = 0$. Тогда

$$\frac{\sum_{i=1}^s a_i}{\sum_{i=1}^s b_i} \geq \min_{\substack{1 \leq i \leq s, \\ b_i \neq 0}} \frac{a_i}{b_i}. \quad (3.4.7)$$

Мы оставляем доказательство неравенств (3.4.6) и (3.4.7) читателю в качестве упражнения. Как очевидное следствие лемм 3.4.1 и 3.4.2 приведем следующее утверждение.

Лемма 3.4.3 Пусть неотрицательные числа a_i и b_i , $i = 1, 2, \dots, s$ удовлетворяют следующим условиям: 1) $b_1 + b_2 + \dots + b_s > 0$; 2) для всех значений i числа a_i и b_i равны или не равны нулю одновременно. Тогда

$$\min_{\substack{1 \leq i \leq s, \\ b_i \neq 0}} \frac{a_i}{b_i} \leq \frac{\sum_{i=1}^s a_i}{\sum_{i=1}^s b_i} \leq \max_{\substack{1 \leq i \leq s, \\ b_i \neq 0}} \frac{a_i}{b_i}. \quad (3.4.8)$$

Для решения задачи о выборе параметров σ_m в (3.4.5) установим сначала границы спектра матрицы $L^{-1}A$. Из (3.4.1) имеем соотношение

$$u^T A_m u = \int_{\Delta_m} |\nabla \hat{u}|^2 d\Delta_m \quad \forall u \in G_p^m, \quad (3.4.9)$$

где $\hat{u} = \text{prol}(u \in G_p^m : V_p^m)$. Аналогично, из (3.4.4) следует, что

$$u^T L_m u = \int_{\Delta} |\nabla \hat{u}|^2 d\Delta \quad \forall u \in G_p^m, \quad (3.4.10)$$

где $\hat{u} = \text{prol}(u \in G_p^m : V_p^m(\Delta))$. Можно показать, что существуют такие положительные постоянные $\delta_m^{(1)}$ и $\delta_m^{(2)}$, что

$$\delta_m^{(1)} \int_{\Delta} |\nabla \hat{u}|^2 d\Delta \leq \int_{\Delta_m} |\nabla \hat{u}|^2 d\Delta_m \leq \delta_m^{(2)} \int_{\Delta} |\nabla \hat{u}|^2 d\Delta.$$

Оставляем вывод этих неравенств читателю в качестве упражнения. При этом заметим, что постоянные $\delta_m^{(1)}$ и $\delta_m^{(2)}$ определяются лишь геометрическими параметрами треугольника Δ_m . Таким образом, мы получаем отношение эквивалентности

$$\delta_m^{(1)} u^T L_m u \leq u^T A_m u \leq \delta_m^{(2)} u^T L_m u \quad \forall u \in G_p^m. \quad (3.4.11)$$

Исходя из представлений (3.4.2) и (3.4.5) — согласно свойству (3.1.10) операции ансамблирования — имеем равенства

$$u^T A u = \sum_{m=1}^l u^{mT} A_m u^m, \quad u^T L u = \sum_{m=1}^l \sigma_m u^{mT} L_m u^m,$$

где $u^m = \text{restr}(u \in G_p : \Delta_m)$. Тем самым для любой ненулевой сеточной функции $u \in G_p$ имеем равенство

$$\frac{u^T A u}{u^T L u} = \frac{\sum_{m=1}^l u^{mT} A_m u^m}{\sum_{m=1}^l \sigma_m u^{mT} L_m u^m}.$$

Нетрудно установить, что при одном и том же значении m ядра матриц A_m и L_m совпадают, то есть $\ker A_m = \ker L_m$; это легко следует из (3.4.9) и (3.4.10). Тогда – в силу утверждения леммы 3.4.3 – можем записать неравенства

$$\min_{\substack{1 \leq m \leq l, \\ u^m \notin \ker L_m}} \frac{u^{mT} A_m u^m}{\sigma_m u^{mT} L_m u^m} \leq \frac{u^T A u}{u^T L u} \leq \max_{\substack{1 \leq m \leq l, \\ u^m \notin \ker L_m}} \frac{u^{mT} A_m u^m}{\sigma_m u^{mT} L_m u^m}.$$

Отсюда, если воспользоваться отношением эквивалентности (3.4.11), следуют неравенства

$$\min_{1 \leq m \leq l} \frac{\delta_m^{(1)}}{\sigma_m} \leq \frac{u^T A u}{u^T L u} \leq \max_{1 \leq m \leq l} \frac{\delta_m^{(2)}}{\sigma_m},$$

справедливые для всех ненулевых сеточных функций $u \in G_p$. Имея это, на основании рассуждений параграфа 3.2 (см. (3.2.13) – (3.2.15)), получаем границы спектра матрицы $L^{-1}A$, а именно:

$$\text{sp}(L^{-1}A) \subset \left[\min_{1 \leq m \leq l} \frac{\delta_m^{(1)}}{\sigma_m}, \max_{1 \leq m \leq l} \frac{\delta_m^{(2)}}{\sigma_m} \right]. \quad (3.4.12)$$

В соответствии с (3.2.16), из (3.4.12) следует оценка числа обусловленности:

$$\kappa(L^{-1}A) \leq \frac{\max_{1 \leq m \leq l} \frac{\delta_m^{(2)}}{\sigma_m}}{\min_{1 \leq m \leq l} \frac{\delta_m^{(1)}}{\sigma_m}} \equiv \chi(\sigma_1, \sigma_2, \dots, \sigma_l). \quad (3.4.13)$$

Поставим следующий вопрос: *при каких значениях параметров $\sigma_1, \sigma_2, \dots, \sigma_l$ выражение $\chi(\sigma_1, \sigma_2, \dots, \sigma_l)$ в правой части неравенства (3.4.13) принимает наименьшее значение?* Ответ дает следующее утверждение.

Лемма 3.4.4 *Функция $\chi(\sigma_1, \sigma_2, \dots, \sigma_l)$ принимает свое наименьшее значение*

$$\chi_{\min} = \max_{1 \leq m \leq l} \frac{\delta_m^{(2)}}{\delta_m^{(1)}} \quad (3.4.14)$$

при

$$\sigma_m = \sqrt{\delta_m^{(1)} \delta_m^{(2)}}, \quad m = 1, 2, \dots, l. \quad (3.4.15)$$

Доказательство. Для любых $\sigma_m > 0$, $m = 1, 2, \dots, l$ справедливо неравенство

$$\chi(\sigma_1, \sigma_2, \dots, \sigma_l) \geq \max_{1 \leq m \leq l} \frac{\delta_m^{(2)}}{\delta_m^{(1)}}.$$

Действительно, это неравенство является очевидным следствием неравенства

$$\frac{\delta_m^{(2)}}{\delta_m^{(1)}} = \frac{\delta_m^{(2)}/\sigma_m}{\delta_m^{(1)}/\sigma_m} \leq \chi(\sigma_1, \sigma_2, \dots, \sigma_l),$$

которое верно для всех $m = 1, 2, \dots, l$. Выберем теперь величины σ_m согласно (3.4.15). В этом случае

$$\chi(\sigma_1, \sigma_2, \dots, \sigma_l) = \frac{\max_{1 \leq m \leq l} \sqrt{\frac{\delta_m^{(2)}}{\delta_m^{(1)}}}}{\min_{1 \leq m \leq l} \sqrt{\frac{\delta_m^{(1)}}{\delta_m^{(2)}}}} = \max_{1 \leq m \leq l} \frac{\delta_m^{(2)}}{\delta_m^{(1)}}. \quad \square$$

Итак, подводя итог предыдущим рассуждениям, приходим к следующему утверждению.

Теорема 3.4.1 Пусть параметры σ_m в определении (3.4.5) матрицы L выбраны в соответствии с формулами (3.4.15). Тогда

$$\text{sp}(L^{-1}A) \subset \left[\frac{1}{\sqrt{\chi_{\min}}}, \sqrt{\chi_{\min}} \right],$$

где величина χ_{\min} вычисляется по формуле (3.4.14). При этом

$$\kappa(L^{-1}A) \leq \chi_{\min}. \quad (3.4.16)$$

Таким образом, определена матрица L , спектрально эквивалентная матрице жесткости A конечноэлементной системы уравнений (3.3.7). Наша следующая задача заключается в построении переобуславливателя для матрицы L , который будет являться одновременно переобуславливателем и для матрицы A .

3.5 О дополнении Шура

Пусть матрица A порядка n представлена в блочном виде

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad (3.5.1)$$

с блоками A_{ij} размера $n_i \times n_j$, $i, j = 1, 2$, где $n_1 + n_2 = n$.

Определение 3.5.1 В предположении, что $\det A_{22} \neq 0$, матрица

$$S_{11} = A_{11} - A_{12}A_{22}^{-1}A_{21} \quad (3.5.2)$$

называется *дополнением Шура*² блока A_{22} . Сходным образом, в предположении, что $\det A_{11} \neq 0$, матрица

$$S_{22} = A_{22} - A_{21}A_{11}^{-1}A_{12} \quad (3.5.3)$$

называется *дополнением Шура* блока A_{11} .

²Шур Иссай (1875–1941) – немецкий математик, внесший большой вклад в теорию матриц и ее приложения.

Вновь определенное понятие играет важную роль как в самой теории матриц, так и в разработке вычислительных алгоритмов линейной алгебры. В качестве примера рассмотрим следующий способ решения невырожденной системы линейных алгебраических уравнений

$$Av = b$$

с блочной матрицей (3.5.1). Запишем эту систему в виде

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad (3.5.4)$$

где v_1, v_2 и b_1, b_2 есть подвекторы векторов v и b соответствующей длины. Фактически, мы имеем систему двух матрично-векторных уравнений

$$\begin{aligned} A_{11}v_1 + A_{12}v_2 &= b_1, \\ A_{21}v_1 + A_{22}v_2 &= b_2. \end{aligned} \quad (3.5.5)$$

При условии, что матрица A_{22} невырожденная, из второго уравнения системы находим выражение вектора v_2 через вектор v_1 , то есть

$$v_2 = A_{22}^{-1}(b_2 - A_{21}v_1).$$

Подставляя это выражение в первое уравнение системы (3.5.5), последовательно получаем

$$\begin{aligned} A_{11}v_1 + A_{12}A_{22}^{-1}(b_2 - A_{21}v_1) &= b_1, \\ (A_{11} - A_{12}A_{22}^{-1}A_{21})v_1 &= b_1 - A_{12}A_{22}^{-1}b_2, \\ S_{11}v_1 &= b_1 - A_{12}A_{22}^{-1}b_2. \end{aligned}$$

В результате мы приходим к следующей процедуре решения блочной системы (3.5.4).

ПРОЦЕДУРА $\langle 2 \times 2$ block system \rangle

1. вычисляется вектор

$$c_1 = b_1 - A_{12}A_{22}^{-1}b_2;$$

2. решается система

$$S_{11}v_1 = c_1;$$

3. вычисляется вектор

$$v_2 = A_{22}^{-1}(b_2 - A_{21}v_1).$$

КОНЕЦ ПРОЦЕДУРЫ

Такой подход к решению системы (3.5.4) имеет смысл применять в том случае, если мы можем эффективно решать системы с матрицами A_{22} и S_{11} . Заметим, что в определенных случаях этого удастся достигнуть. В этом мы убедимся ниже.

При наличии соответствующих условий дополнение Шура вида (3.5.3) также может быть использовано для решения системы (3.5.4).

Установим некоторые свойства дополнения Шура. Изложение будем вести относительно дополнения S_{11} ; все результаты могут быть легко перенесены по аналогии на случай дополнения S_{22} .

В дальнейшем мы будем предполагать, что матрица A — симметричная. Это означает, что в блочном представлении (3.5.1) этой матрицы $A_{11}^T = A_{11}$, $A_{22}^T = A_{22}$ и $A_{12}^T = A_{21}$.

Лемма 3.5.1 *Если матрица A — положительно определенная (положительно полуопределенная), то блоки A_{11} и A_{22} также являются положительно определенными (положительно полуопределенными) матрицами.*

Доказательство. Возьмем произвольные ненулевые векторы u_1 и u_2 длины n_1 и n_2 , соответственно. Определим векторы

$$w = \begin{bmatrix} u_1 \\ 0 \end{bmatrix} \quad \text{и} \quad z = \begin{bmatrix} 0 \\ u_2 \end{bmatrix}$$

длины n . Утверждение леммы следует из легко получаемых равенств

$$w^T A w = u_1^T A_{11} u_1 \quad \text{и} \quad z^T A z = u_2^T A_{22} u_2. \quad \square$$

Пусть матрица A — положительно определенная. Рассмотрим произвольный вектор u длины n , разбитый на подвекторы u_1 и u_2 длины n_1 и n_2 , соответственно:

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

Лемма 3.5.2 *Справедливо равенство*

$$u^T A u = u_1^T S_{11} u_1 + (u_2 + z_2)^T A_{22} (u_2 + z_2), \quad (3.5.6)$$

где

$$z_2 \equiv A_{22}^{-1} A_{21} u_1. \quad (3.5.7)$$

Доказательство. Пользуясь блочными представлениями матрицы A и вектора u , имеем

$$\begin{aligned} u^T A u &= \begin{bmatrix} u_1^T & u_2^T \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} u_1^T & u_2^T \end{bmatrix} \begin{bmatrix} A_{11} u_1 + A_{12} u_2 \\ A_{21} u_1 + A_{22} u_2 \end{bmatrix} \\ &= u_1^T A_{11} u_1 + u_1^T A_{12} u_2 + u_2^T A_{21} u_1 + u_2^T A_{22} u_2. \end{aligned}$$

Из условия (3.5.7), записанного в виде равенства

$$A_{22} z_2 = A_{21} u_1,$$

с учетом симметричности матрицы A следует, что

$$z_2^T A_{22} = u_1^T A_{12}.$$

Тогда

$$u^T Au = u_1^T A_{11} u_1 + z_2^T A_{22} u_2 + u_2^T A_{22} (u_2 + z_2).$$

Далее, так как $A_{11} = S_{11} + A_{12} A_{22}^{-1} A_{21}$, то

$$\begin{aligned} u^T Au &= u_1^T S_{11} u_1 + u_1^T A_{12} A_{22}^{-1} A_{21} u_1 + z_2^T A_{22} u_2 + u_2^T A_{22} (u_2 + z_2) \\ &= u_1^T S_{11} u_1 + z_2^T A_{22} z_2 + z_2^T A_{22} u_2 + u_2^T A_{22} (u_2 + z_2) \\ &= u_1^T S_{11} u_1 + z_2^T A_{22} (u_2 + z_2) + u_2^T A_{22} (u_2 + z_2) \\ &= u_1^T S_{11} u_1 + (u_2 + z_2)^T A_{22} (u_2 + z_2). \quad \square \end{aligned}$$

Следующее утверждение является очевидным следствием лемм 3.5.1 и 3.5.2.

Следствие 3.5.1 *Справедливо неравенство*

$$u^T Au \geq u_1^T S_{11} u_1. \quad (3.5.8)$$

Оказывается, что дополнение Шура наследует важное свойство матрицы.

Теорема 3.5.1 *Если матрица A — положительно определенная, то дополнение Шура S_{11} также является положительно определенной матрицей.*

Доказательство. Симметричность матрицы S_{11} следует из ее определения (3.5.2). Далее, возьмем произвольный ненулевой вектор u_1 и определим вектор u_2 из условия

$$A_{21} u_1 + A_{22} u_2 = 0,$$

то есть положим

$$u_2 = -A_{22}^{-1} A_{21} u_1.$$

Тем самым, получим ненулевой вектор

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}.$$

По определению $u_2 = -z_2$, где z_2 — вектор из (3.5.7). Тогда $u_2 + z_2 = 0$, и из равенства (3.5.6) получим, что

$$u_1^T S_{11} u_1 = u^T Au > 0. \quad \square$$

Пусть имеются две симметричные положительно определенные матрицы A и B порядка n со следующей блочной структурой:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad B = \begin{bmatrix} B_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}; \quad (3.5.9)$$

фактически, эти матрицы отличаются лишь одним диагональным блоком.

Лемма 3.5.3 Число $\lambda = 1$ является решением обобщенной задачи на собственные значения

$$Au = \lambda Bu. \quad (3.5.10)$$

Доказательство. Рассмотрим вектор

$$u = \begin{bmatrix} 0 \\ u_2 \end{bmatrix},$$

где u_2 — некоторый ненулевой вектор длины n_2 . Тогда, в силу блочного вида (3.5.9) матриц,

$$Au = \begin{bmatrix} A_{12}u_2 \\ A_{22}u_2 \end{bmatrix}, \quad Bu = \begin{bmatrix} A_{12}u_2 \\ A_{22}u_2 \end{bmatrix}.$$

Тем самым $Au = Bu$. \square

Теорема 3.5.2 Решения обобщенной задачи на собственные значения с дополнительным условием

$$Au = \lambda Bu, \quad \lambda \neq 1 \quad (3.5.11)$$

являются собственными числами обобщенной задачи

$$S_{11}v = \lambda \hat{S}_{11}v, \quad (3.5.12)$$

где S_{11} и \hat{S}_{11} есть дополнения Шура блока A_{22} в матрицах A и B , соответственно.

Доказательство. Если λ — решение задачи (3.5.11), то существует ненулевой вектор

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

такой, что

$$A_{11}u_1 + A_{12}u_2 = \lambda(B_{11}u_1 + A_{12}u_2), \quad (3.5.13)$$

$$A_{21}u_1 + A_{22}u_2 = \lambda(A_{21}u_1 + A_{22}u_2). \quad (3.5.14)$$

Так как $\lambda \neq 1$, то из (3.5.14) следует, что

$$A_{21}u_1 + A_{22}u_2 = 0.$$

Ясно, что $u_1 \neq 0$; в противном случае $A_{22}u_2 = 0$, а так как матрица A_{22} положительно определенная (лемма 3.5.1), мы получили бы, что и $u_2 = 0$. Тогда

$$u_2 = -A_{22}^{-1}A_{21}u_1.$$

Подставляя это выражение в (3.5.13), получим

$$(A_{11} - A_{12}A_{22}^{-1}A_{21})u_1 = \lambda(B_{11} - A_{12}A_{22}^{-1}A_{21})u_1,$$

или

$$S_{11}u_1 = \lambda \hat{S}_{11}u_1,$$

что доказывает утверждение теоремы. \square

Имея лемму 3.5.3 и теорему 3.5.2, приходим к следующему утверждению.

Следствие 3.5.2 Если собственные числа задачи (3.5.12) принадлежат отрезку $[\alpha, \beta]$, где $\alpha > 0$, то собственные числа задачи (3.5.10) принадлежат отрезку $[a, b]$, где $a = \min(1, \alpha)$, $b = \max(1, \beta)$.

В заключение параграфа укажем один из путей построения переобуславливателей, основанный на использовании дополнения Шура.

Запишем симметричную положительно определенную матрицу A , имеющую блочную структуру (3.5.1), в виде

$$A = \begin{bmatrix} S_{11} + A_{12}A_{22}^{-1}A_{21} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad (3.5.15)$$

где S_{11} — дополнение Шура, определенное в (3.5.2). Предположим, что матрица A_{22} легко обратима; например, это может быть диагональная или трехдиагональная матрица. Заменим дополнение Шура S_{11} на некоторую другую симметричную положительно определенную матрицу R_{11} . Получим матрицу

$$P = \begin{bmatrix} R_{11} + A_{12}A_{22}^{-1}A_{21} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad (3.5.16)$$

которую будем рассматривать в качестве переобуславливателя для матрицы A . При этом R_{11} есть не что иное, как дополнение Шура блока A_{22} в матрице P . Построенная нами матрица P — симметричная и положительно определенная. Симметричность очевидна, а положительная определенность следует из неравенства (3.5.8), примененного к матрице P , то есть неравенства

$$u^T P u \geq u_1^T R_{11} u_1.$$

Выбор матрицы R_{11} в блочном представлении (3.5.16) матрицы P обусловлен характером, свойствами и структурой исходной матрицы A . Это не тривиальная задача. Здесь мы укажем пока два основных требования, предъявляемых к матрице R_{11} .

Во-первых, матрица R_{11} должна быть легко обратимой. Иными словами, решение системы с матрицей R_{11} не должно вызывать затруднений. Это существенно, так как согласно приведенной выше вычислительной процедуре (2×2 block system) решение системы с матрицей P сводится к решению системы с матрицей R_{11} .

Во-вторых, матрица R_{11} должна хорошо аппроксимировать дополнение Шура S_{11} в том смысле, что диапазон спектра матрицы $R_{11}^{-1}S_{11}$ должен быть достаточно узким. Пусть собственные числа матрицы $R_{11}^{-1}S_{11}$ лежат в промежутке $[\alpha, \beta]$. Тогда, согласно следствию 3.5.2, собственные числа матрицы $P^{-1}A$ будут принадлежать отрезку $[a, b]$, где $a = \min(1, \alpha)$, $b = \max(1, \beta)$; в результате $\kappa(P^{-1}A) \leq b/a$ (см. (3.2.16)).

В последующих параграфах для построения переобуславливателей нами будет применен сходный по существу, однако несколько отличающийся по реализации самой идеи подход.

3.6 Двухсеточные переобуславливатели

Рассмотрим k -ый уровень измельчения сетки, где $0 \leq k \leq p$. Для каждого треугольника Δ_m , $m = 1, 2, \dots, l$ определим матрицу $L_m^{(k)}$ порядка n_k^m посредством соотношения

$$w^T L_m^{(k)} u = \int_{\Delta} \nabla \hat{u} \nabla \hat{w} d\Delta \quad \forall u, w \in G_k^m, \quad (3.6.1)$$

где $\hat{u} = \text{prol}(u \in G_k^m : V_k^m(\Delta))$, $\hat{w} = \text{prol}(w \in G_k^m : V_k^m(\Delta))$. Далее, выбрав параметры σ_m по формулам (3.4.15), с помощью операции ансамблирования построим матрицу

$$L^{(k)} = \text{assem}\{\sigma_m L_m^{(k)} : 1 \leq m \leq l\} \quad (3.6.2)$$

порядка n_k . По построению матрица $L^{(p)}$ совпадает с матрицей L , спектрально эквивалентной матрице жесткости A конечноэлементной системы сеточных уравнений (3.3.7). Это нетрудно заметить, сравнивая (3.4.4), (3.4.5) с (3.6.1), (3.6.2).

Таким образом, на последовательности иерархических сеток ω_k , $k = 0, 1, \dots, p$ нами определена последовательность матриц конечноэлементного типа

$$L^{(0)}, L^{(1)}, \dots, L^{(p)} = L. \quad (3.6.3)$$

Для значений $k \geq 1$, в соответствии с правилом нумерации узлов множества N_k , принятом в параграфе 3.3, матрица $L^{(k)}$ может быть представлена в блочном виде

$$L^{(k)} = \begin{bmatrix} L_{11}^{(k)} & L_{12}^{(k)} \\ L_{21}^{(k)} & L_{22}^{(k)} \end{bmatrix} \begin{matrix} \} N_k^{(1)} \\ \} N_k^{(2)} \end{matrix} \quad (3.6.4)$$

с $n_k^{(i)} \times n_k^{(j)}$ -блоками $L_{ij}^{(k)}$, $i, j = 1, 2$. При этом блок $L_{11}^{(k)}$ – диагональная матрица (это следует из того, что в сетке ω_k узлы множества $N_k^{(1)}$ не являются соседними).

Оказывается, что соотношения (3.6.1), с помощью которых определяются матрицы $L_m^{(k)}$, можно записать в несколько иной форме. Для этого обратимся к следствию 3.1.1, а именно, к равенству (3.1.16). Согласно этому равенству,

$$\int_{\Delta} \nabla \hat{u} \nabla \hat{w} d\Delta = \sum_{e' \in \tau_k^m(\Delta)} \int_{e'} \nabla \hat{u} \nabla \hat{w} de' = \frac{\sqrt{3}}{6} \sum_{e' \in \tau_k^m(\Delta)} \phi_{e'}(\hat{u}, \hat{w}), \quad (3.6.5)$$

где билинейная форма $\phi_{e'}(\hat{u}, \hat{w})$ по существу определена нами в (3.1.17). Точнее,

$$\phi_{e'}(\hat{u}, \hat{w}) = (u_2 - u_1)(w_2 - w_1) + (u_3 - u_2)(w_3 - w_2) + (u_1 - u_3)(w_1 - w_3), \quad (3.6.6)$$

где u_i, w_i ($i = 1, 2, 3$) есть значения линейных на треугольнике e' функций $\hat{u}, \hat{w} \in V_k^m(\Delta)$, соответственно, в вершинах 1, 2, 3 – как показано на Рис. 3.6.1 (здесь ради удобства мы используем локальную нумерацию вершин треугольника).

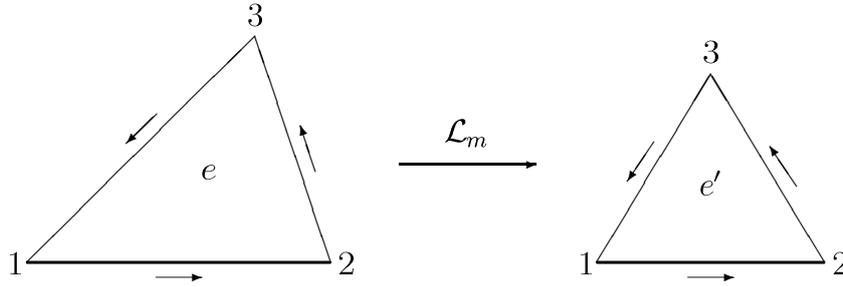


Рис. 3.6.1. Отображение элемента $e \in \tau_k^m$ на элемент $e' \in \tau_k^m(\Delta)$.

Продолжим выкладки. Пусть $e \in \tau_k^m$ есть прообраз равностороннего треугольного элемента $e' \in \tau_k^m(\Delta)$ при линейном отображении \mathcal{L}_m (Рис. 3.6.1). Для треугольника e , по аналогии с (3.6.6), введем в рассмотрение билинейный функционал

$$\phi_e(\hat{u}, \hat{w}) \equiv (u_2 - u_1)(w_2 - w_1) + (u_3 - u_2)(w_3 - w_2) + (u_1 - u_3)(w_1 - w_3), \quad (3.6.7)$$

где u_i, w_i ($i = 1, 2, 3$) есть уже значения линейных на треугольнике e функций $\hat{u}, \hat{w} \in V_k^m$ в соответствующих вершинах. Из (3.6.6) и (3.6.7) очевидно следует, что

$$\phi_{e'}(\hat{u}, \hat{w}) = \phi_e(\hat{u}, \hat{w}).$$

Тем самым, мы можем записать равенство (3.6.5) в следующем виде:

$$\int_{\Delta} \nabla \hat{u} \nabla \hat{w} d\Delta = \frac{\sqrt{3}}{6} \sum_{e \in \tau_k^m} \phi_e(\hat{u}, \hat{w}).$$

Таким образом, как следует из (3.6.1), матрица $L_m^{(k)}$ удовлетворяет соотношению

$$w^T L_m^{(k)} u = \frac{\sqrt{3}}{6} \sum_{e \in \tau_k^m} \phi_e(\hat{u}, \hat{w}) \quad \forall u, w \in G_k^m, \quad (3.6.8)$$

где $\hat{u} = \text{prol}(u \in G_k^m : V_k^m)$, $\hat{w} = \text{prol}(w \in G_k^m : V_k^m)$.

Рассмотрим некоторый треугольный элемент $e \in \tau_{k-1}$, где $1 \leq k \leq p$. На следующем уровне измельчения сетки он разбивается на четыре треугольных элемента (Рис. 3.6.2). В результате на k -ом уровне элемент e превращается в суперэлемент E . Для значений $k = 1, 2, \dots, p$ пусть:

T_k — множество суперэлементов k -го уровня;

T_k^m — сужение множества T_k на треугольник Δ_m , то есть $T_k^m = \text{restr}(T_k : \Delta_m)$.

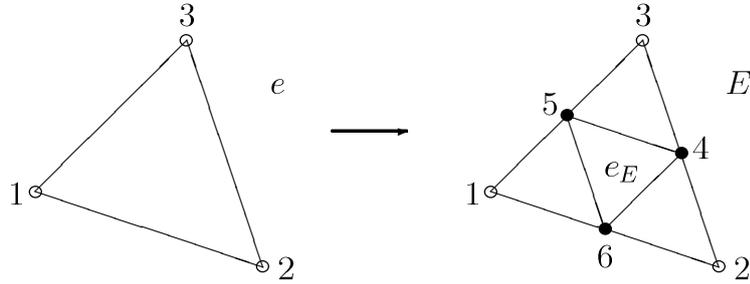


Рис. 3.6.2. Элемент $e \in \tau_{k-1}$ и соответствующий суперэлемент $E \in T_k$.

С каждым суперэлементом $E \in T_k$ будем ассоциировать билинейный функционал

$$\begin{aligned} \Phi_E(u, w) \equiv & (u_4 - u_2)(w_4 - w_2) + (u_4 - u_3)(w_4 - w_3) + \\ & (u_5 - u_3)(w_5 - w_3) + (u_5 - u_1)(w_5 - w_1) + \\ & (u_6 - u_1)(w_6 - w_1) + (u_6 - u_2)(w_6 - w_2), \end{aligned} \quad (3.6.9)$$

где u_i и w_i , $i = 1, 2, \dots, 6$ есть значения функций u и w , соответственно, в узлах суперэлемента (см. локальную нумерацию узлов суперэлемента на Рис. 3.6.2).

Пусть $1 \leq k \leq p$. Осуществим в правой части соотношения (3.6.8) группировку элементов из множества τ_k^m в суперэлементы множества T_k^m . В результате получим, что матрица $L_m^{(k)}$ удовлетворяет соотношению

$$w^T L_m^{(k)} u = \frac{\sqrt{3}}{6} \sum_{E \in T_k^m} [\Phi_E(\hat{u}, \hat{w}) + 2\phi_{e_E}(\hat{u}, \hat{w})] \quad \forall u, w \in G_k^m, \quad (3.6.10)$$

где $e_E \in \tau_k^m$ есть треугольный элемент, вершинами которого являются середины сторон суперэлемента E (Рис. 3.6.2).

Определим матрицу $B_m^{(k)}$ порядка n_k^m с помощью соотношения

$$w^T B_m^{(k)} u = \frac{\sqrt{3}}{6} \sum_{E \in T_k^m} \Phi_E(\hat{u}, \hat{w}) \quad \forall u, w \in G_k^m, \quad (3.6.11)$$

где $\hat{u} = \text{prol}(u \in G_k^m : V_k^m)$, $\hat{w} = \text{prol}(w \in G_k^m : V_k^m)$. Затем, для значений $k = 1, 2, \dots, p$ с помощью операции ансамблирования построим матрицы

$$B^{(k)} = \text{assem}\{\sigma_m B_m^{(k)} : 1 \leq m \leq l\} \quad (3.6.12)$$

с теми же параметрами σ_m , что и в (3.6.2). Сравнивая соотношения (3.6.10) и (3.6.11), видим, что в каждом суперэлементе E был исключен функционал, связанный с внутренним треугольным элементом e_E . Благодаря этому в блочном представлении

$$B^{(k)} = \left[\begin{array}{cc|c} L_{11}^{(k)} & L_{12}^{(k)} & \} N_k^{(1)} \\ L_{21}^{(k)} & B_{22}^{(k)} & \} N_k^{(2)} \end{array} \right] \quad (3.6.13)$$

матрицы $B^{(k)}$ нижний правый блок $B_{22}^{(k)}$ — диагональная матрица; остальные блоки $L_{11}^{(k)}$, $L_{12}^{(k)}$ и $L_{21}^{(k)}$ те же, что и в блочном представлении (3.6.4) матрицы $L^{(k)}$.

Матрицу $B^{(k)}$ будем рассматривать в качестве *переобуславливателя* для матрицы $L^{(k)}$.

Перейдем к исследованию построенных переобуславливателей. Сначала сформулируем и докажем одно вспомогательное утверждение.

Лемма 3.6.1 Пусть сеточная функция

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \begin{matrix} \} N_k^{(1)} \\ \} N_k^{(2)} \end{matrix}; \quad u_1 \in G_k^{(1)}, \quad u_2 \in G_k^{(2)}$$

удовлетворяет соотношению

$$L_{21}^{(k)} u_1 + B_{22}^{(k)} u_2 = 0, \quad (3.6.14)$$

и $\hat{u} = \text{prol}(u \in G_k : V_k)$. Если узел β сетки ω_k не является узлом сетки ω_{k-1} , то

$$\hat{u}(\beta) = \frac{1}{2} (\hat{u}(\alpha_1) + \hat{u}(\alpha_2)), \quad (3.6.15)$$

где α_1 и α_2 — соседние с узлом β узлы сетки ω_k , являющиеся одновременно узлами сетки ω_{k-1} .

Доказательство. Узел β , в силу предположения, является серединой стороны некоторого суперэлемента из множества T_k . Точнее, таких суперэлементов может быть либо два (если узел β лежит внутри области Ω'), либо он один (если узел β лежит на границе области Ω'). Рассмотрим, для определенности, первый случай (второй случай более простой). На Рис. 3.6.3 через $E' \in T_k^q$ и $E'' \in T_k^r$ обозначены упомянутые суперэлементы.

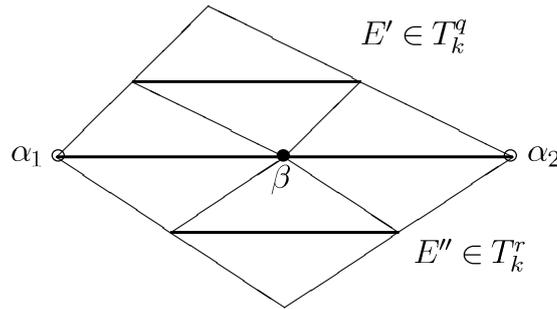


Рис. 3.6.3. Суперэлементы E' и E'' .

Выберем сеточную функцию $w_2 \in G_k^{(2)}$ следующим образом: положим ее равной единице в узле $\beta \in N_k^{(2)}$ и равной нулю в остальных узлах множества $N_k^{(2)}$. Затем определим сеточную функцию

$$w = \begin{bmatrix} 0 \\ w_2 \end{bmatrix} \in G_k.$$

Пусть $\hat{w} = \text{prol}(w \in G_k : V_k)$. С учетом (3.6.14) имеем

$$w^T B^{(k)} u = w_2^T (L_{21}^{(k)} u_1 + B_{22}^{(k)} u_2) = 0. \quad (3.6.16)$$

Воспользовавшись свойством (3.1.10) операции ансамблирования, запишем равенство

$$w^T B^{(k)} u = \sum_{m=1}^l \sigma_m w^{mT} B_m^{(k)} u^m,$$

где $u^m = \text{restr}(u \in G_k : \Delta_m)$, $w^m = \text{restr}(w \in G_k : \Delta_m)$. Отсюда, в силу выбора сеточной функции w , получим

$$\begin{aligned} w^T B^{(k)} u &= \frac{\sqrt{3}}{6} (\sigma_q \Phi_{E'}(\hat{u}, \hat{w}) + \sigma_r \Phi_{E''}(\hat{u}, \hat{w})) \\ &= \frac{\sqrt{3}}{6} (\sigma_q + \sigma_r) (2\hat{u}(\beta) - \hat{u}(\alpha_1) - \hat{u}(\alpha_2)). \end{aligned} \quad (3.6.17)$$

Сопоставляя (3.6.16) и (3.6.17), приходим к равенству

$$2\hat{u}(\beta) - \hat{u}(\alpha_1) - \hat{u}(\alpha_2) = 0,$$

откуда и следует (3.6.15). \square

Рассмотрим дополнение Шура

$$S_{11}^{(k)} = L_{11}^{(k)} - L_{12}^{(k)} B_{22}^{(k)-1} L_{21}^{(k)} \quad (3.6.18)$$

блока $B_{22}^{(k)}$ в блочном представлении (3.6.13) матрицы $B^{(k)}$.

Теорема 3.6.1 Для значений $k = 1, 2, \dots, p$ справедливы равенства

$$S_{11}^{(k)} = \frac{1}{2} L^{(k-1)}. \quad (3.6.19)$$

Доказательство. Возьмем произвольные сеточные функции $u_1, w_1 \in G_k^{(1)}$. Пусть

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \in G_k, \quad w = \begin{bmatrix} w_1 \\ 0 \end{bmatrix} \in G_k,$$

где $u_2 \in G_k^{(2)}$ определяется из условия (3.6.14). В силу выбора функций u и w верно равенство

$$w_1^T S_{11}^{(k)} u_1 = w^T B^{(k)} u. \quad (3.6.20)$$

С учетом соотношения (3.6.11) имеем равенство

$$w^T B^{(k)} u = \sum_{m=1}^l \sigma_m w^{mT} B_m^{(k)} u^m = \sum_{m=1}^l \sigma_m \left(\frac{\sqrt{3}}{6} \sum_{E \in T_k^m} \Phi_E(\hat{u}, \hat{w}) \right),$$

где $\hat{u} = \text{prol}(u \in G_k : V_k)$, $\hat{w} = \text{prol}(w \in G_k : V_k)$. Пользуясь утверждением леммы 3.6.1, с помощью несложных преобразований получим, что

$$\Phi_E(\hat{u}, \hat{w}) = \frac{1}{2} \phi_e(\hat{u}_1, \hat{w}_1),$$

где $e \in \tau_{k-1}$ есть треугольный элемент $(k-1)$ -го уровня, который на следующем уровне измельчения сетки переходит в суперэлемент E ; при этом $\hat{u}_1 = \text{prol}(u_1 \in G_k^{(1)} : V_{k-1})$, $\hat{w}_1 = \text{prol}(w_1 \in G_k^{(1)} : V_{k-1})$. Таким образом,

$$\begin{aligned} w^T B^{(k)} u &= \frac{1}{2} \sum_{m=1}^l \sigma_m \left(\frac{\sqrt{3}}{6} \sum_{e \in \tau_{k-1}^m} \phi_e(\hat{u}_1, \hat{w}_1) \right) \\ &= \frac{1}{2} \sum_{m=1}^l \sigma_m w_1^{mT} L_m^{(k-1)} u_1^m = \frac{1}{2} w_1^T L^{(k-1)} u_1 \end{aligned} \quad (3.6.21)$$

(мы воспользовались соотношением (3.6.8)). Из (3.6.20) и (3.6.21) получим равенство

$$w_1^T S_{11}^{(k)} u_1 = \frac{1}{2} w_1^T L^{(k-1)} u_1,$$

которое выполняется для всех сеточных функций $u_1, w_1 \in G_k^{(1)} = G_{k-1}$. Тем самым приходим к равенству (3.6.19). \square

Имея в виду равенство (3.6.19), матрицу $B^{(k)}$ назовем *двухсеточным переобуславливателем* для матрицы $L^{(k)}$. Для значений $k = 1, 2, \dots, p$ блочное представление (3.6.13) матрицы $B^{(k)}$ можно теперь записать в виде

$$B^{(k)} = \begin{bmatrix} \frac{1}{2} L^{(k-1)} + L_{12}^{(k)} B_{22}^{(k)-1} L_{21}^{(k)} & L_{12}^{(k)} \\ L_{21}^{(k)} & B_{22}^{(k)} \end{bmatrix}. \quad (3.6.22)$$

Основываясь на приведенной в параграфе 3.5 вычислительной процедуре $\langle 2 \times 2 \text{ block system} \rangle$, дадим алгоритм решения системы сеточных уравнений

$$B^{(k)} v = g,$$

где

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \quad g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}; \quad v_i, g_i \in G_k^{(i)}, \quad i = 1, 2.$$

ПРОЦЕДУРА $\langle \text{TG/PREC} \rangle$

1. вычисляется сеточная функция

$$z_1 = 2 \begin{pmatrix} g_1 - L_{12}^{(k)} B_{22}^{(k)-1} \\ g_2 \end{pmatrix};$$

2. решается система

$$L^{(k-1)} v_1 = z_1;$$

3. вычисляется сеточная функция

$$v_2 = B_{22}^{(k)-1} \begin{pmatrix} g_2 - L_{21}^{(k)} v_1 \end{pmatrix}.$$

КОНЕЦ ПРОЦЕДУРЫ

Таким образом, решение системы с матрицей $B^{(k)}$, имеющей порядок n_k , в итоге сводится к решению системы с матрицей $L^{(k-1)}$ порядка n_{k-1} . При этом напомним, что матрица $B_{22}^{(k)}$ является диагональной.

Перейдем теперь к вопросу об установлении границ спектра матрицы $B^{(k)-1} L^{(k)}$. Рассмотрим обобщенную задачу на собственные значения

$$L^{(k)} u = \lambda B^{(k)} u. \quad (3.6.23)$$

Обозначим через $\lambda_{\min}^{(k)}$ и $\lambda_{\max}^{(k)}$, соответственно, наименьшее и наибольшее из собственных чисел задачи (3.6.23).

Для всех $u \in G_k$ справедливо неравенство

$$u^T L^{(k)} u \geq u^T B^{(k)} u.$$

Оно следует непосредственно из соотношений (3.6.10) и (3.6.11). Это означает, что все собственные числа задачи (3.6.23) не меньше единицы. Кроме того, из леммы 3.5.3 следует, что $\lambda = 1$ – собственное число. Получаем, что

$$\lambda_{\min}^{(k)} = 1. \quad (3.6.24)$$

Оценим теперь $\lambda_{\max}^{(k)}$. Для этого применим технику перехода на суперэлементный уровень. Пусть $G_k^E = \text{restr}(G_k : E \in T_k)$ и $V_k^E = \text{restr}(V_k : E \in T_k)$. Для каждого суперэлемента $E \in T_k$ определим матрицы $L_E^{(k)}$ и $B_E^{(k)}$ с помощью соотношений

$$w^T L_E^{(k)} u = \frac{\sqrt{3}}{6} [\Phi_E(\hat{u}, \hat{w}) + 2\phi_{e_E}(\hat{u}, \hat{w})] \quad \forall u, w \in G_k^E \quad (3.6.25)$$

и

$$w^T B_E^{(k)} u = \frac{\sqrt{3}}{6} \Phi_E(\hat{u}, \hat{w}) \quad \forall u, w \in G_k^E \quad (3.6.26)$$

где $\hat{u} = \text{prol}(u \in G_k^E : V_k^E)$, $\hat{w} = \text{prol}(w \in G_k^E : V_k^E)$. Из (3.6.2) и (3.6.10) следует, что

$$L^{(k)} = \text{assem}\{\sigma_m L_E^{(k)} : E \in T_k^m, 1 \leq m \leq l\}, \quad (3.6.27)$$

а из (3.6.11) и (3.6.12) получаем

$$B^{(k)} = \text{assem}\{\sigma_m B_E^{(k)} : E \in T_k^m, 1 \leq m \leq l\}. \quad (3.6.28)$$

Имеем равенство

$$\lambda_{max}^{(k)} = \frac{w^T L^{(k)} w}{w^T B^{(k)} w}$$

(см. (3.2.13), (3.2.14)), где $w \in G_k$ есть соответствующая ненулевая собственная сеточная функция. Согласно (3.6.27) и (3.6.28), последнее равенство можно переписать в виде

$$\lambda_{max}^{(k)} = \frac{\sum_{m=1}^l \sum_{E \in T_k^m} \sigma_m w^{E^T} L_E^{(k)} w^E}{\sum_{m=1}^l \sum_{E \in T_k^m} \sigma_m w^{E^T} B_E^{(k)} w^E},$$

где $w^E = \text{restr}(w \in G_k : E)$. Можно показать, что $\ker L_E^{(k)} = \ker B_E^{(k)}$ для всех суперэлементов $E \in T_k$ (это следует из определений (3.6.25) и (3.6.26) этих матриц). Отсюда, применяя лемму 3.4.3, получим неравенство

$$\lambda_{max}^{(k)} \leq \max_{\substack{E \in T_k, \\ w^E \notin \ker B_E^{(k)}}} \frac{w^{E^T} L_E^{(k)} w^E}{w^{E^T} B_E^{(k)} w^E}.$$

Пусть максимум в правой части этого неравенства достигается на некотором суперэлементе $E \in T_k$. Тогда, для этого суперэлемента

$$\lambda_{max}^{(k)} \leq \frac{w^{E^T} L_E^{(k)} w^E}{w^{E^T} B_E^{(k)} w^E} \leq \max_{u \notin \ker B_E^{(k)}} \frac{u^T L_E^{(k)} u}{u^T B_E^{(k)} u}. \quad (3.6.29)$$

Очевидно, что выражение

$$\max_{u \notin \ker B_E^{(k)}} \frac{u^T L_E^{(k)} u}{u^T B_E^{(k)} u} \quad (3.6.30)$$

принимает одинаковые значения на всех "внутренних" суперэлементах $E \in T_k$, то есть на тех суперэлементах, все узлы которых являются внутренними узлами сетки ω_k . Более того, на "граничных" суперэлементах выражение (3.6.30) принимает заведомо не большее значение, чем на "внутренних". Таким образом, принимая во внимание сделанное замечание, и, исходя из неравенства (3.6.29), можем записать, что

$$\lambda_{max}^{(k)} \leq \lambda_E^{(k)}, \quad (3.6.31)$$

где $\lambda_E^{(k)}$ есть наибольшее собственное число задачи

$$L_E^{(k)} u = \lambda B_E^{(k)} u, \quad u \notin \ker B_E^{(k)} \quad (3.6.32)$$

для некоторого "внутреннего" суперэлемента.

Проведем разбиение матриц $L_E^{(k)}$ и $B_E^{(k)}$ на блоки в соответствии с нашим правилом разбиения множества узлов на два подмножества — вершины суперэлементов и середины их сторон:

$$L_E^{(k)} = \begin{bmatrix} L_{11,E}^{(k)} & L_{12,E}^{(k)} \\ L_{21,E}^{(k)} & L_{22,E}^{(k)} \end{bmatrix}, \quad B_E^{(k)} = \begin{bmatrix} L_{11,E}^{(k)} & L_{12,E}^{(k)} \\ L_{21,E}^{(k)} & B_{22,E}^{(k)} \end{bmatrix}. \quad (3.6.33)$$

Так как нас интересуют собственные числа $\lambda \neq 1$, то нетрудно показать (см. в связи с этим теорему 3.5.2), что задача (3.6.32) сводится к задаче на собственные значения

$$S_E^{(k)}(L)v = \lambda S_E^{(k)}(B)v, \quad v \notin \ker(S_E^{(k)}(B)), \quad (3.6.34)$$

где $S_E^{(k)}(L)$ и $S_E^{(k)}(B)$ есть дополнения Шура в матрицах $L_E^{(k)}$ и $B_E^{(k)}$, соответственно:

$$S_E^{(k)}(L) = L_{22,E}^{(k)} - L_{21,E}^{(k)}L_{11,E}^{(k)-1}L_{12,E}^{(k)}, \quad S_E^{(k)}(B) = B_{22,E}^{(k)} - L_{21,E}^{(k)}L_{11,E}^{(k)-1}L_{12,E}^{(k)}. \quad (3.6.35)$$

Имея соотношения (3.6.25) и (3.6.26), запишем матрицы $L_E^{(k)}$ и $B_E^{(k)}$ согласно нумерации узлов суперэлемента, данной на Рис. 3.6.2:

$$L_E^{(k)} = \frac{\sqrt{3}}{6} \left[\begin{array}{ccc|ccc} 2 & 0 & 0 & 0 & -1 & -1 \\ 0 & 2 & 0 & -1 & 0 & -1 \\ 0 & 0 & 2 & -1 & -1 & 0 \\ \hline 0 & -1 & -1 & 6 & -2 & -2 \\ -1 & 0 & -1 & -2 & 6 & -2 \\ -1 & -1 & 0 & -2 & -2 & 6 \end{array} \right], \quad B_E^{(k)} = \frac{\sqrt{3}}{6} \left[\begin{array}{ccc|ccc} 2 & 0 & 0 & 0 & -1 & -1 \\ 0 & 2 & 0 & -1 & 0 & -1 \\ 0 & 0 & 2 & -1 & -1 & 0 \\ \hline 0 & -1 & -1 & 2 & 0 & 0 \\ -1 & 0 & -1 & 0 & 2 & 0 \\ -1 & -1 & 0 & 0 & 0 & 2 \end{array} \right].$$

Матрицы разбиты на блоки согласно (3.6.33). По формулам (3.6.35) находим:

$$S_E^{(k)}(L) = \frac{5\sqrt{3}}{12} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \quad S_E^{(k)}(B) = \frac{\sqrt{3}}{12} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.$$

Отсюда делаем заключение, что единственным собственным числом задачи (3.6.34) является $\lambda = 5$. Таким образом, наибольшее собственное число задачи (3.6.32) есть $\lambda_E^{(k)} = 5$, и из (3.6.31) получаем, что

$$\lambda_{max}^{(k)} \leq 5. \quad (3.6.36)$$

Итак, нами оценены наименьшее и наибольшее собственные числа задачи (3.6.23). Это оценки (3.6.24) и (3.6.36). Подводя итог вышеизложенному, приходим к следующему результату.

Теорема 3.6.2 Для значений $k = 1, 2, \dots, p$ собственные числа матрицы $B^{(k)-1}L^{(k)}$ принадлежат промежутку $[1, 5]$, то есть

$$\text{sp}(B^{(k)-1}L^{(k)}) \subset [1, 5], \quad (3.6.37)$$

и, тем самым,

$$\kappa(B^{(k)-1}L^{(k)}) \leq 5. \quad (3.6.38)$$

На основе двухсеточных переобуславливателей ниже нами будет построен многосеточный переобуславливатель. Получается он путем замены на каждом уровне измельчения сетки дополнения Шура в блочном представлении (3.6.22) двухсеточного переобуславливателя на рекурсивно задаваемые матрицы, при определении которых используются итерации чебышевского типа.

3.7 О чебышевском итерационном методе

Этот параграф носит вспомогательный характер. Здесь мы кратко опишем чебышевский итерационный метод с переобуславливателем и приведем некоторые формулы и соотношения, которые будут использованы при построении многосеточного переобуславливателя для матрицы жесткости конечноэлементной системы (3.3.7). Более подробное изложение этого метода можно найти в монографиях [25, 33].

Рассмотрим систему линейных алгебраических уравнений

$$Av = b \quad (3.7.1)$$

порядка n с симметричной положительно определенной матрицей A . Чебышевский итерационный метод для этой системы записывается в виде

$$P \frac{v^j - v^{j-1}}{\theta_j} = -Av^{j-1} + b, \quad j = 1, 2, \dots, s. \quad (3.7.2)$$

Дадим некоторые пояснения. Переобуславливатель P также является симметричной положительно определенной матрицей. Число итераций в (3.7.2) ограничено некоторым числом $s > 1$. Итерационные параметры θ_j выбираются следующим образом. Пусть собственные числа матрицы $P^{-1}A$ принадлежат отрезку $[\alpha, \beta]$, где $0 < \alpha < \beta$, то есть

$$\text{sp}(P^{-1}A) \subset [\alpha, \beta]. \quad (3.7.3)$$

Параметры θ_j определяются по формулам

$$\theta_j = \frac{2}{(\beta + \alpha) + (\beta - \alpha)t_j}, \quad j = 1, 2, \dots, s, \quad (3.7.4)$$

где

$$t_j = \cos \frac{2j-1}{2s} \pi, \quad j = 1, 2, \dots, s$$

есть корни многочлена Чебышева первого рода степени s . Напомним, что эти многочлены задаются следующим образом:

$$T_s(x) = \frac{(x + \sqrt{x^2 - 1})^s + (x - \sqrt{x^2 - 1})^s}{2}, \quad -\infty < x < \infty. \quad (3.7.5)$$

При этом для значений $|x| \leq 1$ многочлены допускают и другое представление:

$$T_s(x) = \cos(s \arccos x).$$

Введем векторную норму

$$\|u\|_P \equiv \sqrt{u^T P u}.$$

На s -ом шаге итераций (3.7.2) справедлива оценка погрешности

$$\|v - v^s\|_P \leq \gamma \|v - v^0\|_P, \quad (3.7.6)$$

где

$$\gamma = \frac{2q^s}{1 + q^{2s}}, \quad q = \frac{\sqrt{c} - 1}{\sqrt{c} + 1}, \quad c = \frac{\beta}{\alpha}. \quad (3.7.7)$$

Доказано, что

$$\gamma = \max_{\alpha \leq x \leq \beta} \left| \prod_{j=1}^s (1 - \theta_j x) \right|. \quad (3.7.8)$$

Для получения приближенного решения v^s системы (3.7.1) с точностью $\varepsilon > 0$, то есть

$$\|v - v^s\|_P < \varepsilon \|v - v^0\|_P,$$

число итераций s в методе (3.7.2) — согласно оценке (3.7.6) — определяется из условия $\gamma < \varepsilon$. Путем несложных выкладок можно показать, что s следует выбрать как

$$s > \frac{\sqrt{\kappa(P^{-1}A)}}{2} \ln \frac{1}{\varepsilon}$$

(в связи с этим см. (3.2.16) и (3.2.17)). При этом из (3.7.3) и (3.7.7) следует, что

$$\kappa(P^{-1}A) \leq c. \quad (3.7.9)$$

Основываясь на приведенных выше сведениях о чебышевском итерационном методе, перейдем к выводу необходимых нам в дальнейшем формул. Из (3.7.2) имеем рекуррентное соотношение

$$v^j = (I - \theta_j P^{-1}A)v^{j-1} + \theta_j P^{-1}b, \quad j = 1, 2, \dots, s,$$

где I — единичная матрица. Отсюда последовательно находим:

$$v^1 = (I - \theta_1 P^{-1}A)v^0 + [I - (I - \theta_1 P^{-1}A)] A^{-1}b,$$

$$v^2 = (I - \theta_2 P^{-1}A)(I - \theta_1 P^{-1}A)v^0 + [I - (I - \theta_2 P^{-1}A)(I - \theta_1 P^{-1}A)] A^{-1}b$$

и т.д. В результате, проделав s шагов, получим

$$v^s = \prod_{j=1}^s (I - \theta_j P^{-1}A)v^0 + \left[I - \prod_{j=1}^s (I - \theta_j P^{-1}A) \right] A^{-1}b.$$

Если в качестве начального приближения взять $v^0 = 0$, то

$$v^s = \left[I - \prod_{j=1}^s (I - \theta_j P^{-1}A) \right] A^{-1}b.$$

Это равенство можно записать в виде

$$A \left[I - \prod_{j=1}^s (I - \theta_j P^{-1} A) \right]^{-1} v^s = b.$$

Введем в рассмотрение матрицу

$$R \equiv A \left[I - \prod_{j=1}^s (I - \theta_j P^{-1} A) \right]^{-1}. \quad (3.7.10)$$

В результате приходим к тому, что

$$Rv^s = b.$$

Таким образом, на основе приведенных рассуждений делаем вывод, что решение системы

$$Rv = b \quad (3.7.11)$$

с матрицей R равносильно осуществлению s шагов чебышевского метода (3.7.2) с начальным приближением $v^0 = 0$. Определенная в (3.7.10) матрица R будет существенно использована при построении многосегочного переобуславливателя.

В заключение параграфа установим границы спектра матрицы $R^{-1}A$. Для этого рассмотрим задачу на собственные значения

$$R^{-1}Au = \lambda u. \quad (3.7.12)$$

Пользуясь определением (3.7.10) матрицы R , запишем эту задачу в виде

$$\left[I - \prod_{j=1}^s (I - \theta_j P^{-1} A) \right] u = \lambda u$$

или

$$\prod_{j=1}^s (I - \theta_j P^{-1} A) u = (1 - \lambda)u.$$

Следовательно, собственными числами задачи (3.7.12) являются $\lambda = 1 - \mu$, где μ — собственные числа задачи

$$\prod_{j=1}^s (I - \theta_j P^{-1} A) u = \mu u.$$

Эти числа легко вычисляются:

$$\mu = \prod_{j=1}^s (I - \theta_j \eta),$$

где η есть собственные числа матрицы $P^{-1}A$. Таким образом,

$$\lambda = 1 - \prod_{j=1}^s (I - \theta_j \eta). \quad (3.7.13)$$

Имея выражение (3.7.8) для величины γ , в силу предположения (3.7.3) выводим:

$$\max_{\eta \in \text{sp}(P^{-1}A)} \left| \prod_{j=1}^s (I - \theta_j \eta) \right| \leq \max_{\alpha \leq x \leq \beta} \left| \prod_{j=1}^s (I - \theta_j x) \right| = \gamma. \quad (3.7.14)$$

Окончательно, из (3.7.13) и (3.7.14) приходим к тому, что $1 - \gamma \leq \lambda \leq 1 + \gamma$, то есть

$$\text{sp}(R^{-1}A) \subset [1 - \gamma, 1 + \gamma]. \quad (3.7.15)$$

Располагая необходимыми для нас сведениями из теории чебышевского итерационного метода, перейдем к непосредственному построению многосеточного переобуславливателя.

3.8 Многосеточный переобуславливатель

В параграфе 3.6 мы построили последовательность (3.6.3) матриц конечноэлементного типа (3.6.2), а именно, последовательность

$$L^{(0)}, L^{(1)}, \dots, L^{(p)} = L, \quad (3.8.1)$$

где L — построенная в параграфе 3.4 матрица, спектрально эквивалентная матрице жесткости A конечноэлементной системы сеточных уравнений (3.3.7). Там же была построена последовательность

$$B^{(1)}, B^{(2)}, \dots, B^{(p)} \quad (3.8.2)$$

двухсеточных переобуславливателей (3.6.12) для матриц $L^{(k)}$, $k = 1, 2, \dots, p$ и получено их блочное представление (3.6.22).

Выберем некоторое целое число $s \geq 1$ и для значений $k = 1, 2, \dots, p$ последовательно определим матрицы

$$M^{(k)} = \begin{bmatrix} \frac{1}{2}R^{(k-1)} + L_{12}^{(k)} B_{22}^{(k)-1} L_{21}^{(k)} & L_{12}^{(k)} \\ L_{21}^{(k)} & B_{22}^{(k)} \end{bmatrix}, \quad (3.8.3)$$

где

$$\begin{cases} R^{(0)} = L^{(0)}, & \text{для } k = 1, \\ R^{(k-1)} = L^{(k-1)} \left[I - \prod_{j=1}^s \left(I - \theta_j^{(k-1)} M^{(k-1)-1} L^{(k-1)} \right) \right]^{-1}, & \text{для } 2 \leq k \leq p \end{cases} \quad (3.8.4)$$

(см. определение (3.7.10) матрицы R); в качестве параметров $\theta_j^{(k-1)}$ берутся величины

$$\theta_j^{(k-1)} = \frac{2}{(\beta_{k-1} + \alpha_{k-1}) + (\beta_{k-1} - \alpha_{k-1})t_j}, \quad j = 1, 2, \dots, s$$

(см. выражения (3.7.4)), где

$$t_j = \cos \frac{2j-1}{2s} \pi, \quad j = 1, 2, \dots, s$$

есть корни многочлена Чебышева первого рода степени s , а $0 < \alpha_{k-1} < \beta_{k-1}$ — концы промежутка, содержащего все собственные числа матрицы $M^{(k-1)-1}L^{(k-1)}$.

Матрицы $M^{(k)}$ рассматриваются в качестве *переобуславливателей* для матриц $L^{(k)}$. Как видно из определения (3.8.3) этих матриц, они получены путем замены дополнений Шура в блочных представлениях (3.6.22) двухсеточных переобуславливателей $B^{(k)}$ на рекурсивно определяемые матрицы (3.8.4). Эта рекурсия использует чебышевские итерации на более грубой сетке (см. ниже (3.8.15)).

Выведем правило, по которому вычисляются границы спектра матриц $M^{(k)-1}L^{(k)}$, $k = 1, 2, \dots, p$. По определению $M^{(1)} = B^{(1)}$, что следует из (3.8.3), (3.8.4) и (3.6.22). Поэтому, согласно (3.6.37),

$$\text{sp}(M^{(1)-1}L^{(1)}) \subset [\alpha_1, \beta_1], \quad (3.8.5)$$

где

$$\alpha_1 = 1, \quad \beta_1 = 5. \quad (3.8.6)$$

Пусть $k \geq 2$. Предположим, что

$$\text{sp}(M^{(k-1)-1}L^{(k-1)}) \subset [\alpha_{k-1}, \beta_{k-1}].$$

Тогда на основе результатов параграфа 3.7 можем утверждать, что

$$\text{sp}(M^{(k)-1}B^{(k)}) \subset [1 - \gamma_{k-1}, 1 + \gamma_{k-1}], \quad (3.8.7)$$

где

$$\gamma_{k-1} = \frac{2q_{k-1}^s}{1 + q_{k-1}^{2s}}, \quad q_{k-1} = \frac{\sqrt{c_{k-1}} - 1}{\sqrt{c_{k-1}} + 1}, \quad c_{k-1} = \frac{\beta_{k-1}}{\alpha_{k-1}}; \quad (3.8.8)$$

это следует из (3.7.15), (3.7.7) и определения (3.8.4) матрицы $R^{(k-1)}$.

Имеем равенство

$$M^{(k)-1}L^{(k)} = (M^{(k)-1}B^{(k)})(B^{(k)-1}L^{(k)}).$$

Отсюда, принимая во внимание (3.6.37) и (3.8.7), нетрудно получить, что

$$\text{sp}(M^{(k)-1}L^{(k)}) \subset [\alpha_k, \beta_k], \quad (3.8.9)$$

где

$$\alpha_k = 1 - \gamma_{k-1}, \quad \beta_k = 5(1 + \gamma_{k-1}). \quad (3.8.10)$$

Таким образом, мы построили последовательность матриц

$$M^{(1)}, M^{(2)}, \dots, M^{(p)} \equiv M. \quad (3.8.11)$$

Матрицу M назовем *многосеточным переобуславливателем* для матрицы L из (3.8.1). Она же будет рассматриваться в качестве многосеточного переобуславливателя и для матрицы A конечноэлементной системы сеточных уравнений (3.3.7). Поясним название переобуславливателя M — *многосеточный*.

При решении системы с матрицей M необходимо решать системы с матрицами $M^{(k)}$, $k = 1, 2, \dots, p$. Рассмотрим систему

$$M^{(k)}v = g, \quad (3.8.12)$$

где

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \quad g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}; \quad v_i, g_i \in G_k^{(i)}, \quad i = 1, 2.$$

ПРОЦЕДУРА (MG/PRES)

1. вычисляется сеточная функция

$$z_1 = 2 \left(g_1 - L_{12}^{(k)} B_{22}^{(k)-1} g_2 \right); \quad (3.8.13)$$

2. решается система

$$R^{(k-1)}v_1 = z_1; \quad (3.8.14)$$

при $2 \leq k \leq p$:

решение системы (3.8.14) равносильно выполнению s шагов чебышевского итерационного метода

$$M^{(k-1)} \frac{v_1^{(j)} - v_1^{(j-1)}}{\theta_j^{(k-1)}} = -L^{(k-1)}v_1^{(j-1)} + z_1, \quad (3.8.15)$$

$$j = 1, 2, \dots, s; \quad v_1^{(0)} = 0; \quad v_1 = v_1^{(s)};$$

при $k = 1$:

решается система

$$L^{(0)}v_1 = z_1; \quad (3.8.16)$$

3. вычисляется сеточная функция

$$v_2 = B_{22}^{(k)-1} \left(g_2 - L_{21}^{(k)}v_1 \right). \quad (3.8.17)$$

КОНЕЦ ПРОЦЕДУРЫ

Дадим некоторые комментарии к процедуре. Матрицы $B_{22}^{(k)}$ являются диагональными. Поэтому вычисления по формулам (3.8.13) и (3.8.17) не вызывают затруднений. Предполагается, что на нулевом уровне, который соответствует самой грубой сетке, система сеточных уравнений (3.8.16) с матрицей $L^{(0)}$ решается с помощью некоторого прямого метода с затратами $O(1)$ арифметических операций.

Итак, для того, чтобы решить систему с матрицей $M = M^{(p)}$, следует решить на более грубой сетке s систем с матрицей $M^{(p-1)}$. В свою очередь, каждая из систем с матрицей

$M^{(p-1)}$ сводится к решению s систем с матрицей $M^{(p-2)}$ и т.д. Иначе говоря, многосеточный переобуславливатель M строится рекурсивно — от одного уровня к другому, вплоть до самой грубой сетки, где система сеточных уравнений решается с помощью некоторого прямого метода (обычно эта система имеет небольшой порядок).

На Рис. 3.8.1 дана графическая схема решения системы сеточных уравнений с матрицей M — согласно вычислительной процедуре (MG/PREC).

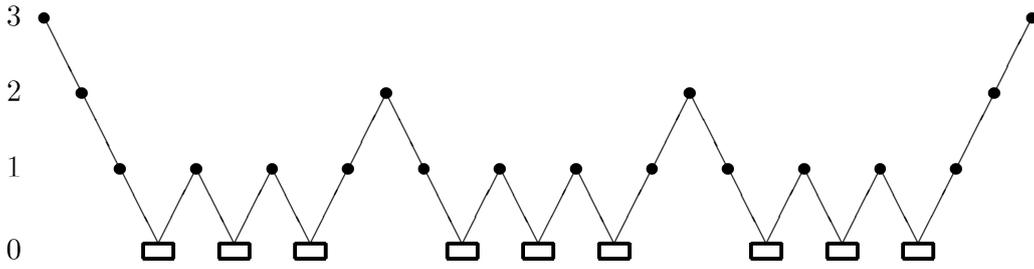


Рис. 3.8.1. Схема решения системы с матрицей M ($p = 3, s = 3$).

Перейдем к решению важного вопроса о выборе числа чебышевских итераций s на промежуточных уровнях измельчения сетки. Рассмотрим величины

$$c_k = \frac{\beta_k}{\alpha_k}, \quad k = 1, 2, \dots, p \quad (3.8.18)$$

(по существу, мы уже ввели их в (3.8.8)). Эти величины, согласно (3.8.9), являются верхними оценками чисел обусловленности матриц $M^{(k)-1}L^{(k)}$, то есть

$$\kappa(M^{(k)-1}L^{(k)}) \leq c_k, \quad k = 1, 2, \dots, p. \quad (3.8.19)$$

Как следует из (3.8.6) и (3.8.10),

$$c_1 = 5; \quad c_k = \frac{1 + \gamma_{k-1}}{1 - \gamma_{k-1}}, \quad k = 1, 2, \dots, p.$$

Пользуясь выражениями из (3.8.8), получим

$$\frac{1 + \gamma_{k-1}}{1 - \gamma_{k-1}} = \frac{1 + 2q_{k-1}^s + q_{k-1}^{2s}}{1 - 2q_{k-1}^s + q_{k-1}^{2s}} = \left(\frac{1 + q_{k-1}^s}{1 - q_{k-1}^s} \right)^2 = \left[\frac{(\sqrt{c_{k-1}} + 1)^s + (\sqrt{c_{k-1}} - 1)^s}{(\sqrt{c_{k-1}} + 1)^s - (\sqrt{c_{k-1}} - 1)^s} \right]^2.$$

Таким образом, введенные в (3.8.18) величины c_k могут быть получены с помощью рекуррентной процедуры

$$c_1 = 5, \\ c_k = 5 \left[\frac{(\sqrt{c_{k-1}} + 1)^s + (\sqrt{c_{k-1}} - 1)^s}{(\sqrt{c_{k-1}} + 1)^s - (\sqrt{c_{k-1}} - 1)^s} \right]^2, \quad k = 2, 3, \dots, p. \quad (3.8.20)$$

Для установления характера поведения величин c_k рассмотрим нелинейное уравнение

$$c = \varphi(c), \quad (3.8.21)$$

где $\varphi(c)$ есть функция, определяемая следующим образом:

$$\varphi(c) = 5 \left[\frac{(\sqrt{c_{k-1}} + 1)^s + (\sqrt{c_{k-1}} - 1)^s}{(\sqrt{c_{k-1}} + 1)^s - (\sqrt{c_{k-1}} - 1)^s} \right]^2.$$

Лемма 3.8.1 Если $s^2 > 5$, то уравнение (3.8.21) имеет единственное положительное решение.

Доказательство. Рассматривая положительные значения c , сделаем замену переменной $z = 1/\sqrt{c}$. В результате от уравнения (3.8.21) мы перейдем к уравнению

$$z = \psi(z), \quad (3.8.22)$$

где

$$\psi(z) = \frac{1}{\sqrt{5}} \frac{(1+z)^s - (1-z)^s}{(1+z)^s + (1-z)^s}.$$

Как видно из выражения для этой функции, $\psi(0) = 0$, $\psi(1) = 1/\sqrt{5}$. Прежде всего заметим, что для значений $z \geq 1$ верно неравенство

$$\frac{(1+z)^s - (1-z)^s}{(1+z)^s + (1-z)^s} \leq z,$$

и, тем самым, $\psi(z) < z$. Поэтому положительные решения уравнения (3.8.22) нам следует искать в интервале $(0, 1)$.

Вычисляя производную функции $\psi(z)$, получим

$$\psi'(z) = \frac{4s}{\sqrt{5}} \frac{(1-z^2)^{s-1}}{[(1+z)^s + (1-z)^s]^2}.$$

Очевидно, что $\psi'(z) > 0$ при $0 < z < 1$, причем $\psi'(0) = s/\sqrt{5} > 1$ и $\psi'(1) = 0$. Далее, вычислим вторую производную:

$$\psi''(z) = -\frac{8s}{\sqrt{5}} \frac{(1-z^2)^{s-2} ((s-1)z[(1+z)^s + (1-z)^s] + s(1-z^2)[(1+z)^s - (1-z)^s])}{[(1+z)^s + (1-z)^s]^3}.$$

Как видим, $\psi''(z) < 0$ при $0 < z < 1$.

Последующие рассуждения таковы. Рассмотрим функцию

$$f(z) = z - \psi(z), \quad z \in [0, 1].$$

Очевидно, что $f(0) = 0$ и $f(1) = 1 - 1/\sqrt{5} > 0$. Для производной $f'(z) = 1 - \psi'(z)$ этой функции $f'(0) = 1 - s/\sqrt{5} < 0$ и $f'(1) = 1$. Так как $f''(z) = -\psi''(z) > 0$, то функция $f'(z)$ в интервале $(0, 1)$ монотонно возрастает. Следовательно, существует единственная

точка $\bar{z} \in (0, 1)$ — такая, что $f'(\bar{z}) = 0$. При этом $f'(z) < 0$ для $z < \bar{z}$ и $f'(z) > 0$, если $z > \bar{z}$. То есть в интервале $(0, \bar{z})$ функция $f(z)$ монотонно убывает, а в интервале $(\bar{z}, 1)$ — монотонно возрастает. Далее, так как $f(\bar{z}) = f(0) + f'(\xi)\bar{z}$, где $\xi \in (0, \bar{z})$, то $f(\bar{z}) = f'(\xi)\bar{z} < 0$. Следовательно, существует единственная точка $z_* \in (\bar{z}, 1)$ — такая, что $f(z_*) = 0$, то есть $z_* = \psi(z_*)$. Таким образом, если $s^2 > 5$, то уравнение (3.8.22) имеет единственное положительное решение z_* , заключенное в интервале $(0, 1)$. Тем самым уравнение (3.8.21) имеет единственное положительное решение $c_* = 1/z_*^2 > 1$. \square

Возвратимся к рекуррентной процедуре (3.8.20). Пусть

$$z_k = \frac{1}{\sqrt{c_k}}, \quad k = 1, 2, \dots, p.$$

Согласно проведенной при доказательстве леммы 3.8.1 замене переменной, эти величины рекуррентным соотношениям

$$z_1 = \frac{1}{\sqrt{5}}; \quad z_k = \psi(z_{k-1}), \quad k = 2, 3, \dots, p. \quad (3.8.23)$$

Ниже, в качестве иллюстрации, приведен график функции $\psi(z)$ при условии, что $s^2 > 5$.

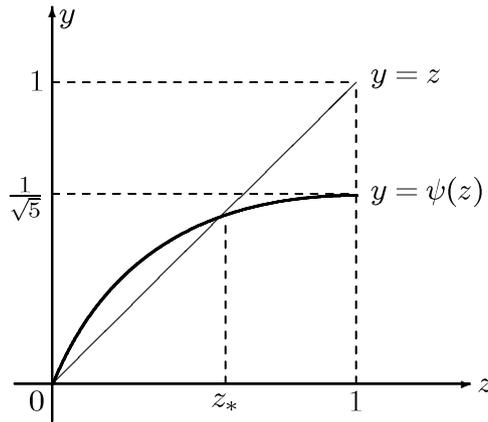


Рис. 3.8.2. Иллюстрация к рекуррентной процедуре (3.8.23).

Докажем, что

$$z_* < z_k, \quad k = 1, 2, \dots, p.$$

На отрезке $[0, 1]$ функция $\psi(z)$ монотонно возрастает. Поэтому $z_* < z_1$. Действительно, так как $z_* < 1$ (см. доказательство леммы 3.8.1), то

$$z_* = \psi(z_*) < \psi(1) = \frac{1}{\sqrt{5}} = z_1.$$

Далее рассуждения проводятся по индукции. Пусть $z_* < z_{k-1}$ для некоторого $k \geq 2$. Тогда

$$z_* - z_k = \psi(z_*) - \psi(z_{k-1}) = \psi'(\xi)(z_* - z_{k-1}) < 0,$$

то есть $z_* < z_k$.

Покажем теперь, что последовательность z_k монотонно убывает, то есть

$$z_k < z_{k-1}, \quad k = 2, 3, \dots, p.$$

Просмотрев доказательство леммы 3.8.1, приходим к заключению, что $\psi'(z) < 1$ при $z \geq z_*$. Тогда

$$\begin{aligned} z_k &= \psi(z_{k-1}) = \psi(z_{k-1}) - \psi(z_*) + z_* - z_{k-1} + z_{k-1} = \psi'(\eta)(z_{k-1} - z_*) + \\ &+ (z_* - z_{k-1}) + z_{k-1} = (1 - \psi'(\eta))(z_* - z_{k-1}) + z_{k-1} < z_{k-1}. \end{aligned}$$

Получили, что последовательность величин z_k монотонно убывает и ограничена снизу (при неограниченном росте p) величиной z_* , то есть

$$z_* < z_p < \dots < z_2 < z_1.$$

Переходя к величинам c_k , приходим к заключению, что

$$c_1 < c_2 < \dots < c_p < c_*.$$

Таким образом, вычисляемая по рекуррентным формулам (3.8.20) последовательность c_k является монотонно возрастающей и ограниченной сверху (при неограниченном росте уровней измельчения сетки) величиной c_* — единственным положительным решением уравнения (3.8.21). Имея неравенства (3.8.19), приходим к следующему утверждению.

Теорема 3.8.1 Если $s^2 > 5$, то

$$\kappa(M^{-1}L) \leq c_*$$

где c_* — единственное положительное решение уравнения (3.8.21).

Итак, число чебышевских итераций на промежуточных уровнях измельчения сетки надо брать исходя из условия $s^2 > 5$. Наименьшим целым положительным числом, удовлетворяющим этому условию, является $s = 3$. На этом мы и остановимся. Ниже будет показано, что такой выбор числа s — наилучший, с точки зрения минимизации объема вычислительных затрат.

При $s = 3$ уравнение (3.8.21) принимает вид

$$c = 5c \left(\frac{c+3}{3c+1} \right)^2.$$

Его единственным положительным решением является $c_* = 3 + 2\sqrt{5} \leq 7.48$. Как следствие теоремы 3.8.1, сформулируем следующее утверждение.

Теорема 3.8.2 Если $s = 3$, то

$$\kappa(M^{-1}L) \leq 3 + 2\sqrt{5}. \quad (3.8.24)$$

Наконец, дадим оценку числа обусловленности матрицы A конечноэлементной системы сеточных уравнений (3.3.7) — по отношению к многосеточному переобуславливателю M . Из равенства

$$M^{-1}A = (M^{-1}L)(L^{-1}A)$$

следует неравенство

$$\kappa(M^{-1}A) \leq \kappa(M^{-1}L)\kappa(L^{-1}A).$$

Числа обусловленности, стоящие в правой части этого неравенства, у нас уже оценены. Это оценки (3.8.24) и (3.4.16). Тем самым приходим к следующему утверждению.

Теорема 3.8.3 Если $s = 3$, то

$$\kappa(M^{-1}A) \leq (3 + 2\sqrt{5}) \max_{1 \leq m \leq l} \frac{\delta_m^{(2)}}{\delta_m^{(1)}}, \quad (3.8.25)$$

где $\delta_m^{(1)}$ и $\delta_m^{(2)}$ — положительные постоянные из отношений эквивалентности (3.4.11).

Замечание 3.8.1 Построение многосеточного переобуславливателя было осуществлено алгебраическим путем, который не имеет, вообще говоря, геометрического аналога в смысле отображения исходной области Ω' на область, составленную из равносторонних треугольников. Линейные отображения, переводящие треугольники Δ_m в один и тот же единичный равносторонний треугольник Δ , осуществляются независимо друг от друга. Поэтому не имеет значения количество треугольников Δ_m , имеющих общей вершиной одну и ту же точку плоскости.

Замечание 3.8.2 Если многоугольная область Ω' представляет собой объединение равносторонних треугольников Δ_m , то отпадает необходимость в осуществлении линейных отображений на единичный равносторонний треугольник. В этом случае, фактически, $L = A$. В итоге оценка (3.8.25) принимает вид

$$\kappa(M^{-1}A) \leq (3 + 2\sqrt{5}).$$

В заключение дадим оценку числа арифметических операций, затрачиваемых на выполнение одного шага переобуславливания. Введем следующие обозначения:

$A_{ops}^{(k)}$ — число арифметических операций, требуемых для решения системы (3.8.12) с матрицей $M^{(k)}$ ($k = 1, 2, \dots, p$);

$A_{ops}^{(0)}$ — число арифметических операций, требуемых для решения системы (3.8.16) с матрицей $L^{(0)}$; предполагается, что $A_{ops}^{(0)} = O(1)$.

Чтобы излишне не загромождать анализ, будем считать, что n_k — общее число узлов сетки ω_k . В работе [38] устанавливается связь между числом узлов на двух соседних уровнях:

$$4 - O(2^{-k}) \leq \frac{n_k}{n_{k-1}} \leq 4, \quad k = 1, 2, \dots, p. \quad (3.8.26)$$

Путем прямых вычислений выводятся выражения, ограничивающие сверху введенные выше сложностные характеристики:

$$A_{ops}^{(k)} \leq (6 + 4s)n_k - 8n_{k-1} + sA_{ops}^{(k-1)}, \quad 2 \leq k \leq p,$$

$$A_{ops}^{(1)} \leq 10n_1 - 9n_0 + A_{ops}^{(0)}.$$

Если воспользоваться правым из неравенств (3.8.26), то получим:

$$A_{ops}^{(k)} \leq 4(1 + s)n_k + sA_{ops}^{(k-1)}, \quad 2 \leq k \leq p,$$

$$A_{ops}^{(1)} \leq 7.75 n_1 + A_{ops}^{(0)}.$$

Отсюда следует неравенство

$$A_{ops}^{(p)} \leq 4(1 + s) [n_p + sn_{p-1} + \dots + s^{p-2}n_2] + 7.75 s^{p-1}n_1 + s^{p-1}A_{ops}^{(0)},$$

а так как $7.75 \leq 4(1 + s)$, то

$$A_{ops}^{(p)} \leq 4(1 + s) [n_p + sn_{p-1} + \dots + s^{p-1}n_1] + s^{p-1}A_{ops}^{(0)}.$$

Из этой оценки, принимая во внимание левое из неравенств (3.8.26), находим, что

$$A_{ops}^{(p)} \simeq 4(1 + s) \left[1 + \left(\frac{s}{4}\right) + \left(\frac{s}{4}\right)^2 + \dots + \left(\frac{s}{4}\right)^{p-1} \right] n_p + s^{p-1}A_{ops}^{(0)}. \quad (3.8.27)$$

В построенном нами многосеточном переобуславливателе M мы выбрали $s = 3$. Тогда

$$A_{ops}^{(p)} \simeq 64n_p + 3^{p-1}A_{ops}^{(0)},$$

а так как очевидно $3^{p-1} \leq n_p$, то

$$A_{ops}^{(p)} \simeq (64 + A_{ops}^{(0)})n_p. \quad (3.8.28)$$

Наконец, принимая во внимание предположение, что $A_{ops}^{(0)} = O(1)$, получим оценку

$$A_{ops} \equiv A_{ops}^{(p)} \leq Cn_p \quad (3.8.29)$$

числа арифметических операций, затрачиваемых на один шаг переобуславливания, где C есть положительная постоянная, не зависящая от шага h_p самой мелкой сетки.

Таким образом, с учетом оценок (3.8.25) и (3.8.29) многосеточный переобуславливатель M можно рассматривать как оптимальный — в смысле определения, данного в параграфе 3.2.

Замечание 3.8.3 Если в многосеточном переобуславливателе положить $s = 4$, то оценка (3.8.24) числа обусловленности улучшится:

$$\kappa(M^{-1}L) \leq 5.57.$$

Соответственно улучшится и оценка (3.8.25). Однако, в этом случае, согласно (3.8.27),

$$A_{ops} \simeq n_p \log_2 n_p,$$

что хуже, чем в (3.8.28).

Литература

- [1] **Акопян Ю.Р.** Основы численных методов, ч.1.-Ер.: Изд-во РАУ, 2005.
- [2] **Акопян Ю.Р.** Основы численных методов, ч.2.-Ер.: Изд-во РАУ, 2007.
- [3] **Акопян Ю.Р., Кузнецов Ю.А.** Алгебраический многосеточный метод решения конечно-элементных уравнений на иерархических треугольных сетках.- Препринт №.250, ОБМ АН СССР, М, 1990.
- [4] **Бахвалов Н.С., Жидков Н.П., Кобельков Г.М.** Численные методы.-М: БИНОМ: Лаборатория Знаний, 2004.
- [5] **Вержбицкий В.М.** Основы численных методов.-М: Высшая школа, 2002.
- [6] **Воеводин В.В.** Линейная алгебра.-М: Наука, 1987.
- [7] **Воеводин В.В., Кузнецов Ю.А.** Матрицы и вычисления.-М: Наука, 1984.
- [8] **Вулих Б.З.** Введение в функциональный анализ.-М: Наука, 1967.
- [9] **Гантмахер Ф.Р.** Теория матриц.-М: Наука, 1967.
- [10] **Зенкевич О., Морган К.** Конечные элементы и аппроксимация.-М: Мир, 1986.
- [11] **Ильин В.А., Позняк Э.Г.** Основы математического анализа, ч.1,2.-М: Наука-Физматлит, 2000.
- [12] **Канторович Л.В., Акилов Г.П.** Функциональный анализ.-М: Наука, 1977.
- [13] **Колмогоров А.Н., Фомин С.В.** Элементы теории функций и функционального анализа.-М: Наука, 1989.
- [14] **Курош А.Г.** Курс высшей алгебры.-М: Наука, 1971.
- [15] **Ладыженская О.А., Уральцева Н.Н.** Линейные и квазилинейные уравнения эллиптического типа.-М: Наука, 1964.
- [16] **Ладыженская О.А.** Краевые задачи математической физики.-М: Наука, 1973.
- [17] **Лебедев В.И.** Функциональный анализ и вычислительная математика.-М: ВИНТИ, 1994.
- [18] **Марчук Г.И., Агошков В.И.** Введение в проекционно-сеточные методы.-М: Наука, 1981.
- [19] **Митчелл Э., Уэйт Р.** Метод конечных элементов для уравнений с частными производными.- М: Мир, 1981.
- [20] **Михлин С.Г.** Вариационные методы в математической физике.-М: Наука, 1970.
- [21] **Михлин С.Г.** Линейные уравнения в частных производных.-М: Высшая школа, 1977.

- [22] **Оганесян Л.А.** Численные расчет плит. // В сб. "Решение инженерных задач на электронно-вычислительных машинах". - Ленинград, ЦБТИ, ЛСНХ, 1963, 85 – 97.
- [23] **Оганесян Л.А., Руховец Л.А.** Вариационно-разностные методы решения эллиптических уравнений. - Ер., АН Арм.ССР, 1979.
- [24] **Понтрягин Л.С.** Обыкновенные дифференциальные уравнения. - М.: Наука, 1974.
- [25] **Самарский А.А., Николаев Е.С.** Методы решения сеточных уравнений. - М.: Наука, 1978.
- [26] **Соболев С.Л.** Некоторые применения функционального анализа в математической физике. - М.: Наука, 1988.
- [27] **Стрэнг Г., Фикс Дж.** Теория метода конечных элементов. - М.: Мир, 1977.
- [28] **Сьярле Ф.** Метод конечных элементов для эллиптических задач. - М.: Мир, 1980.
- [29] **Фаддеев Д.К., Фаддеева В.Н.** Вычислительные методы линейной алгебры. - М.: ФМ, 1960.
- [30] **Фихтенгольц Г.М.** Основы математического анализа, т. 1, 2. - М.: Наука, 1968.
- [31] **Шайдуров В.В.** Многосеточные методы конечных элементов. - М.: Наука, 1989.
- [32] **Эльсгольц Л.Э.** Дифференциальные уравнения и вариационное исчисление. - М.: Наука, 1969.
- [33] **Axelsson O.** Iterative Solution Methods. - Cambridge University Press, 1994.
- [34] **Axelsson O. and V.A. Barker.** Finite Element Solution of Boundary Value Problems. - Academic Press, 1984.
- [35] **Axelsson O., Hakopian Yu.R. and Yu.A. Kuznetsov.** Multilevel preconditioning for perturbed finite element matrices. - *IMA Journal of Numerical Analysis*, v. **17**, 1997, 125 – 149.
- [36] **Axelsson O. and P.S. Vassilevski.** Algebraic multilevel preconditioning methods. I. - *Numerische Mathematik*, v. **56**, 1989, 157 – 177.
- [37] **Axelsson O. and P.S. Vassilevski.** Algebraic multilevel preconditioning methods. II. - *SIAM Journal of Numerical Analysis*, v. **27**, No. 6, 1990, 1569 – 1590.
- [38] **Axelsson O. and P.S. Vassilevski.** Asymptotic work estimates for AMLI methods. - *Applied Numerical Mathematics*, v. **7**, 1991, 437 – 451.
- [39] **Benzi M.** Preconditioning techniques for large linear systems: A survey. - *Journal of Computational Physics*, **182**, 2002, 418 – 477.
- [40] **Bramble J.H., Pasciak J.E. and J. Xu.** Parallel multilevel preconditioners. - *Mathematics of Computation*, v. **55**, No. 191, 1990, 1 – 22.
- [41] **Courant R.** Variational methods for the solution of problems of equilibrium and vibrations. - *Bull. Amer. Math. Soc.*, **49**, No. 1, 1943, 1 – 23.
- [42] **Hakopian Yu.R. and Yu.A. Kuznetsov.** Algebraic multigrid/substructuring preconditioners on triangular grids. - *Soviet Journal of Numerical Analysis and Mathematical Modelling*, v. **6**, No. 6, 1991, 453 – 483.
- [43] **Ke Chen.** Matrix Preconditioning Techniques and Applications. - Cambridge University Press, 2005.
- [44] **Kuznetsov Yu.A.** Algebraic multigrid domain decomposition methods. - *Soviet Journal of Numerical Analysis and Mathematical Modelling*, v. **4**, No. 5, 1989, 351 – 379.

-
- [45] **Kuznetsov Yu.A.** A new parallel algebraic preconditioner.-*Journal of Numerical Linear Algebra with Applications*, v. **1**, No. 2 ,1992, 215–225.
- [46] **Meyer C.D.** Matrix Analysis and Applied Linear Algebra.-SIAM, 2000.
- [47] **Saad Y.** Iterative Methods for Sparse Linear Systems.-PWS Publishing, 1996.
- [48] **Vassilevski P.S.** Multilevel Block Factorization Preconditioners.-Springer, 2008.

Предметный указатель

- градиент функции 25
- дополнение Шура 110
- задача
 - Дирихле 24
 - на собственные значения
 - - обобщенная 100
 - третья краевая 27, 87
- иерархические сетки 102
- кривая
 - класса C^k 47
 - кусочно-гладкая 47
- матрица
 - Грама 29
 - жесткости 62
 - спектрально эквивалентная 101, 107
- метод
 - Бубнова-Галеркина 32
 - вариационно-разностный 37
 - вариационный 33, 36
 - конечных элементов 36
 - проекционно-разностный 37
 - проекционный 34, 36
 - Рунге 28
 - чебышевский 125
- множество
 - всюду плотное 13
 - замкнутое 13
 - линейное 13
 - плотное 13
- неравенство
 - Коши-Буняковского 12
 - Фридрихса 26, 27
- норма
 - матрицы 97
 - оператора 13
 - элемента 10
 - энергетическая 17
- оператор
 - линейный 13
 - - непрерывный 13
 - - ограниченный 13
 - - положительно определенный 14
 - - симметричный 14
- операция
 - ансамблирования 90
 - восполнения 51
 - сужения 100
- переобуславливатель 100
 - двухсеточный 121
 - многосеточный 130
 - оптимальный 101
- последовательность
 - минимизирующая 22
 - фундаментальная 10
- произведение
 - скалярное 11, 25
 - - энергетическое 17
- пространство
 - банахово 11
 - гильбертово 11
 - евклидово 11
 - нормированное 9
 - унитарное 11
 - энергетическое 17

решение

- конечноэлементное 60
- обобщенное 22, 58

сеточная область 48

система

- координат
- - локальная 47
- - приграничная 75
- ортогональная 30
- ортонормированная 30
- уравнений
- - Бубнова-Галеркина 33
- - Ритца 29

суперэлемент 117

триангуляция 48

уравнение

- Лапласа 24
- Пуассона 24
- Штурма-Лиувилля 23

функционал

- линейный 19
- - ограниченный 19
- энергии 19

функция

- кусочно-линейная 50
- - базисная 49
- сеточная 49
- с локальным носителем 36
- финитная 36

число обусловленности 97

элемент

- базисный 28, 33
- конечный 36
- - треугольный 51
- - - линейный 51
- нормированный 10

Юрий Рубенович Акопян

Введение в метод
конечных элементов

учебное пособие

Յուրի Ռուբենի Հակոբյան

Վերջավոր տարրերի
մեթոդի ներածություն

ուսումնական ձեռնարկ