МИНИСТЕРСТВО ОБЩЕГО И ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ЯРОСЛАВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМ. П. Г. ДЕМИДОВА

М. В. НЕВСКИЙ, И. П. ИРОДОВА

НЕКОТОРЫЕ ВОПРОСЫ ТЕОРИИ ПРИБЛИЖЕНИЯ ФУНКЦИЙ

Учебное пособие

ЯРОСЛАВЛЬ 1999

ББК В161.54я73
Н40
УДК 517.518.82/.84

Рецензенты:

кафедра математического анализа Ярославского государственного педагогического университета им. К.Д. Ушинского;

С.Ф. Булычев, канд. техн. наук, ст. науч. сотр. Института проблем вычислительной техники РАН

Невский М.В., Иродова И.П.

Н40 Некоторые вопросы теории приближения функций: Учеб. пособие / Яросл. гос. ун-т. Ярославль, 1999. 92 с.

Учебное пособие предназначено для студентов старших курсов университетов специальностей "Математика" и "Прикладная математика", изучающих спецкурсы по теории приближения функций. Рассматриваются вопросы, связанные с приближением функций одного и двух переменных с помощью многочленов, кусочно-полиномиальных функций и сплайнов. Ряд результатов публикуется в учебной литературе впервые. Пособие содержит большой набор упражнений и задач по всем разделам.

Табл. 1. Ил. 12. Библиогр.: 55 назв.

© Ярославский государственный университет, 1996

© Невский М.В., Иродова И.П., 1996

Содержание

Предисловие..... 4

Часть 1. Модули непрерывности и полиномиальная аппроксимация функций одного переменного..... 5

§1. Модули непрерывности первого порядка и их свойства..... 5
§2. Модули непрерывности порядка k 13
§3. Соотношения между модулями непрерывности различного.... порядка и вида..... 19
§4. Теорема Уитни. Константы Уитни..... 25
§5. Кусочно-полиномиальная и сплайн-аппроксимация функций... 32
§6. Одномерные интерполяционные проекторы и их оценки..... 43

Часть 2. Некоторые методы полиномиальной и кусочно-полиномиальной аппроксимации функций двух переменных..... 51

§7. Разностные операторы и модули непрерывности функций двух переменных..... 51
§8. Задачи двумерной интерполяции. Интерполяция с помощью многочленов из $\mathcal{P}_n$ 57
§9. Линейная и квадратичная интерполяция на квадрате..... 61
§10. Оценки для нормы интерполяционного проектора. Способы выбора узлов интерполяции на подмножестве квадрата..... 69
§11. Интерполяция функций двух переменных с помощью многочленов из $\mathcal{P}_{m,n}$ 72
§12. Приближение кусочно-полиномиальными функциями..... 75

Комментарии и библиографические замечания..... 84

Литература..... 88

ПРЕДИСЛОВИЕ

Пособие базируется на лекциях авторов по теории функций и приближений, прикладной теории приближения, математическому моделированию и др. в рамках соответствующих спецкурсов, читаемых ими в течение последних 13 - 15 лет в Ярославском государственном университете. По своей идеологии, рассматриваемым задачам и стилю изложения оно близко к учебным пособиям по теории приближения функций, написанным начиная с 1981 г. на кафедре теории функций и функционального анализа ЯрГУ ([2 - 5] в списке литературы). Автором (или соавтором) упомянутых книг является Ю. А.Брудный, основатель и руководитель (в 1974 - 1991 гг.) ярославской школы по теории функций. Настоящее пособие написано благодарными учениками Юрия Абрамовича Брудного.

Пособие целиком посвящено полиномиальной и кусочно-полиномиальной аппроксимации функций одного и двух переменных в равномерной метрике и близким вопросам (модули непрерывности функций, теорема Уитни, интерполяция, алгоритмы кусочно-полиномиальной аппроксимации).

Как правило, рассматриваемые вопросы совсем не отражены или недостаточно отражены в учебной литературе. Ряд результатов изложен в специальной литературе, недоступной студентам.

Авторы стремились сделать изложение материала максимально понятным студентам старших курсов; с этой целью в пособии также разобран ряд примеров и приведён большой набор задач и упражнений по всем разделам книги. Это делает возможным, по мысли авторов, использование пособия для активной самостоятельной работы студентов. В конце книги даются комментарии и библиографические замечания для заинтересованных читателей.

Весь материал пособия разбит на две части и двенадцать параграфов. Нумерация утверждений и формул осуществляется независимо в пределах каждого параграфа; это учитывается при ссылках на них.

Авторы

Часть 1. Модули непрерывности и полиномиальная аппроксимация функций одного переменного

В первой части пособия $\mathcal{P}_k$ обозначает совокупность алгебраических многочленов от одного переменного степени $\leq k$, $k=0,1,\ldots$ Через $C(a,b)$, как обычно, обозначается совокупность непрерывных функций $f:[a,b]\longrightarrow\mathbb{R}$ с нормой $\|f\|_{C(a,b)}:=\max\limits_{a\leq x\leq b}|f(x)|$.

§1. Модули непрерывности первого порядка и их свойства

Модули непрерывности порядка k являются так называемыми разностными характеристиками поведения функций. Эти характеристики часто используются в анализе и его приложениях. В теории приближения, вычислительной математике модули непрерывности применяются, в частности, при оценке точности или скорости приближения. Важно, что эти характеристики непрерывности (или гладкости) функций не используют их дифференцируемости. Однако ряд свойств модулей непрерывности аналогичен свойствам производных соответствующего порядка.

Мы начнем с введения модулей непрерывности первого порядка непрерывных на отрезке функций.

<u>Определение.</u> Модулем непрерывности первого порядка (или просто модулем непрерывности) функции $f\in C(a,b)$ называется функция $\omega(f;t)=\omega(t)$ аргумента $t\in[0,b-a]$, определенная равенством

$$\omega(f;t):=\sup_{\substack{a\leq x,y\leq b\\ |x-y|\leq t}}|f(x)-f(y)| \tag{1}$$

или, что то же самое,

$$\omega(f;t):=\sup_{0\leq h\leq t}\ \sup_{a\leq x\leq b-h}|f(x+h)-f(x)|. \tag{2}$$

Таким образом, значение $\omega(f;t)$ равно максимальному колебанию функции f на отрезке длины $\leq t$.

Можно дать эквивалентное определение модуля непрерывности через

оператор Δ_h - так называемый оператор взятия разности с шагом h. Именно, пусть τ_h - оператор сдвига на h : $\tau_h f(x) := f(x+h)$, J - тождественный оператор; тогда $\Delta_h := \tau_h - J$. Это означает, что для всех f

$$\Delta_h f(x) = f(x+h) - f(x).$$

Нетрудно показать, что τ_h и Δ_h при всех $h \in [0, b-a]$ - линейные непрерывные операторы, действующие из $C(a,b)$ в $C(a,b-h)$; при этом их нормы равны соответственно

$$\|\tau_h\|_{C(a,b)\to C(a,b-h)} = 1 \; ; \; \|\Delta_h\|_{C(a,b)\to C(a,b-h)} = 2 .$$

В соответствии с формулой (2) для $f \in C(a,b)$

$$\omega(f;t) = \sup_{0\le h\le t} \|\Delta_h f\|_{C(a,b-h)} \qquad (3)$$

В правой части (3) вместо равномерной нормы можно использовать некоторую другую $\|\cdot\|_X$; это дает возможность ввести модули непрерывности для функций из других функциональных пространств X, например, пространств L_p, $1 \le p \le \infty$. Мы ограничимся лишь случаем $X = C(a,b)$.

Модуль непрерывности функции $f: \mathbb{R} \to \mathbb{R}$, равномерно непрерывной на всей прямой, определяется равенством

$$\omega(f;t) := \sup_{|x-y|\le t} |f(x) - f(y)| = \sup_{\substack{0\le h\le t \\ x\in\mathbb{R}}} |\Delta_h f(x)| .$$

Отметим, что из определения $\omega(t)$ следует неравенство

$$|f(x) - f(y)| \le \omega(|x-y|).$$

<u>Примеры</u>.

1. Если $f(x) = Ax + B$, $x \in \mathbb{R}$, то

$$\omega(f;t) = \sup_{\substack{0\le h\le t \\ x\in\mathbb{R}}} |A(x+h) + B - Ax - B| = \sup_{0\le h\le t} |Ah| = |A|t .$$

2. Если $f(x) = \sin x$, $x \in \mathbb{R}$, то

$$\omega(f;t) = \sup_{\substack{0\le h\le t\\ x\in R}} |\sin(x+h)-\sin x| = 2\sup_{\substack{0\le h\le t\\ x\in R}} \left|\cos\left(x+\frac{h}{2}\right)\sin\frac{h}{2}\right| =$$

$$= 2\sup_{0\le h\le t}\left|\sin\frac{h}{2}\right| = \begin{cases} 2\sin\frac{t}{2}, & t\le\pi \\ 2, & t>\pi \end{cases}.$$

3. Пусть $f(x)$ имеет график, изображенный на рис. 1. Тогда $\omega(f,t) = \omega(t)$ имеет график, показанный на рис. 2.

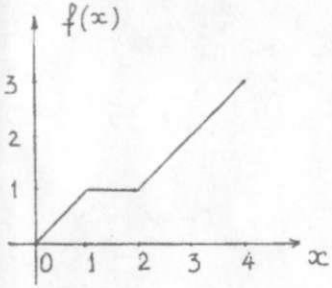

Рис. 1.

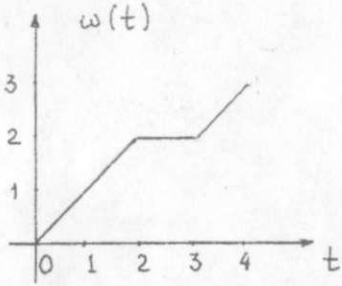

Рис. 2.

Свойства модуля непрерывности (первого порядка)

1. $\omega(0) = 0$.

2. $\omega(t)$ не убывает по t.

3. $\omega(t)$ непрерывна.

4. $\omega(t)$ субаддитивна (полуаддитивна) по t, то есть
$$\omega(t_1+t_2) \le \omega(t_1) + \omega(t_2).$$

5. Для $n\in N$, $nt\in[0,b-a]$ $\omega(nt)\le n\,\omega(t)$.

6. Для $\lambda\ge 0$, $(\lambda+1)t\in[0,b-a]$ $\omega(\lambda t)\le[\lambda+1]\,\omega(t)$.

7. $\omega(f;t)\equiv 0 \iff f\in\mathcal{P}_0$.

8. $\omega(f;t) \geq ct$, причем для $f \notin \mathcal{P}_0$ $c > 0$.

9. Если $f \in C^1(a,b)$, то $\omega(f;t) \leq t\|f'\|_{C(a,b)}$.

Поэтому для $f \in C^1(a,b)$ $c_1 t \leq \omega(t) \leq c_2 t$.

10. $\omega(f;t) \leq 2\|f\|_{C(a,b)}$.

<u>Доказательство</u>. Свойства 1,2,7 и 10 очевидны. Свойства 3,5 и 6 следуют из субаддитивности $\omega(t)$. Поэтому докажем сначала свойство 4.

$$\omega(t_1+t_2) = \sup_{0\leq h\leq t_1+t_2} \|\Delta_h f\|_{C(a,b-h)} = \sup_{0\leq h_i\leq t_i} \|\Delta_{h_1+h_2} f\|_{C(a,b-h_1-h_2)} \leq$$

$$\leq \sup_{0\leq h_i\leq t_i} \|f(x+h_1+h_2)-f(x+h_2)\|_{C(a,b-h_1-h_2)} + \sup_{0\leq h_i\leq t_i} \|f(x+h_2)-f(x)\|_{C(a,b-h_1-h_2)} \leq$$

$$\leq \sup_{0\leq h_1\leq t_1} \|\Delta_{h_1} f\|_{C(a,b-h_1)} + \sup_{0\leq h_2\leq t_2} \|\Delta_{h_2} f\|_{C(a,b-h_2)} = \omega(t_1)+\omega(t_2).$$

Из субаддитивности $\omega(t)$ следует, что при $t_2 \geq t_1$

$$\omega(t_2) - \omega(t_1) \leq \omega(t_2 - t_1).$$

Поэтому при любых t_1, t_2

$$|\omega(t_2) - \omega(t_1)| \leq \omega(|t_2 - t_1|). \qquad (4)$$

Как известно, функция, непрерывная на отрезке, является равномерно непрерывной на нем. Это в точности означает, что для любой $f \in C(a,b)$

$$\omega(f;t) \to 0, \quad t \to +0.$$

Отсюда и из неравенства (4) следует непрерывность $\omega(t)$ при всех t (свойство 3).

Неравенства свойств 5 и 6 легко следуют из субаддитивности и неубывания ω (для свойства 6). Далее, при $t > 0$

$$\omega(b-a)=\omega\left(t\cdot\frac{b-a}{t}\right)\leq\left(\frac{b-a}{t}+1\right)\omega(t)\leq 2\,\frac{b-a}{t}\,\omega(t),$$

то есть

$$\omega(t)\geqslant\frac{\omega(b-a)}{2(b-a)}\,t$$

- неравенство свойства 8 с константой $c:=\dfrac{\omega(b-a)}{2(b-a)}$, отличной от нуля для непостоянной f.

Наконец, если $f\in C^1(a,b)$, то по формуле конечных приращений для $x\leq\xi\leq x+h$

$$|\Delta_h f(x)|\leq|f'(\xi)|h\leq\|f'\|_{C(a,b)}\,h.$$

Взятие верхней грани по $h\in[0,t]$ дает неравенство свойства 9.

Замечание. Свойства 1-4 полностью характеризуют модуль непрерывности. Именно, функция $f:[0,T]\to R_+$, удовлетворяющая свойствам 1-4, является модулем непрерывности некоторой непрерывной функции, а именно самой себя, то есть $\omega(f;t)=f(t)$ (докажите!). В связи с этим любую функцию ω, удовлетворяющую свойствам 1-4, коротко также называют **модулем непрерывности**. Обратим особое внимание на проверку субаддитивности. Часто бывает полезным такой признак субаддитивности функции f: если f - вогнутая, $f(0)=0$, то f субаддитивна. Действительно, из этих условий следует, что $f(\alpha s)\geqslant\alpha f(s)$ при $0<\alpha\leq 1$, или

$$\frac{f(\alpha s)}{\alpha s}\geqslant\frac{f(s)}{s}.$$

В связи с этим $f(s)/s$ убывает, и поэтому при $t_1,t_2>0$

$$f(t_1+t_2)=\frac{t_1}{t_1+t_2}\,f(t_1+t_2)+\frac{t_2}{t_1+t_2}\,f(t_1+t_2)\leq$$
$$\leq t_1\,\frac{f(t_1)}{t_1}+t_2\,\frac{f(t_2)}{t_2}=f(t_1)+f(t_2).$$

Пример. Функция $f(x)=x^\alpha$, $x\in[0,1]$, $0<\alpha\leq 1$ является, как нетрудно показать, модулем непрерывности. Поэтому $\omega(f;t)=t^\alpha$.

Отметим, что модуль непрерывности часто бывает вогнутой функцией, однако это не всегда так (см. пример 3). Оказывается, что существует вогнутая функция $\tilde{\omega}(t)$, такая, что

$$\omega(t) \le \tilde{\omega}(t) \le 2\omega(t) \ ;$$

$\tilde{\omega}(t)$ - так называемая наименьшая вогнутая мажоранта $\omega(t)$, (см. [9]).

Модули непрерывности первого порядка лежат в основе определения важных функциональных классов, встречающихся в анализе. Мы остановимся на классах Липшица и Дини-Липшица (R. Lipschitz, U. Dini). Условие Липшица рассматривалось впервые в 1864 году в качестве достаточного условия для сходимости ряда Фурье функции f.

<u>Определение</u>. Классы функций Липшица и Дини-Липшица на отрезке $[a,b]$ есть соответственно

$$Lip\,\alpha := \{ f \in C(a,b) : \ \omega(f;t) \le ct^{\alpha} \}, \ \alpha > 0;$$

$$DL := \{ f \in C(a,b) : \ \omega(f;t)\ln t \to 0, \ t \to +0 \}.$$

Нетрудно показать, что при $\alpha > 1$ $Lip\,\alpha$ содержит лишь константы. Кроме того, для $\alpha > \beta$ $Lip\,\alpha \subset Lip\,\beta$; в частности, $Lip\,1 \subset Lip\,\alpha, \alpha < 1$. В силу свойства 9 модулей непрерывности $C^1(a,b) \subset Lip\,1$. Кроме того, при всех $\alpha > 0$ $Lip\,\alpha \subset DL$; это связано с тем, что

$$\lim_{t \to +0} t^{\alpha} \ln t = 0, \ \alpha > 0.$$

Таким образом, при $0 < \beta < \alpha < 1$ имеют место включения

$$C^1 \subset Lip\,1 \subset Lip\,\alpha \subset Lip\,\beta \subset DL.$$

Все эти включения являются строгими.

Наиболее важным является класс функций $Lip\,1$. По определению, $f \in Lip\,1$ (или f удовлетворяет **условию Липшица**), если

$$|f(x) - f(y)| \le c|x-y|, \ a \le x, y \le b. \qquad (5)$$

В курсах по теории функций доказывают <u>теорему</u>: функция f удовлетворяет условию (5) тогда и только тогда, когда f абсолютно непрерывна и почти всюду на $[a,b]$ производная f' удовлетворяет неравенству $|f'(x)| \le c$, где c -константа из (5). Доказательство этого утверждения имеется, например, в монографии [9].

Чтобы установить принадлежность данной функции f тому или иному классу Липшица, достаточно получить подходящую оценку модуля

непрерывности $\omega(f;t)$.

Пример. Пусть $f(x) := x\sin\frac{1}{x}$ для $0 < x \le 1$, $f(0) := 0$. Функция f непрерывна на $[0,1]$, причем f дифференцируема при $x > 0$, но $f'(0)$ не существует (почему?). Покажем, что $f \in Lip\frac{1}{2}$.

Пусть $x \geqslant \sqrt{h} > 0$. Так как

$$f'(x) = \sin\frac{1}{x} - \frac{1}{x}\cos\frac{1}{x} ,$$

то $|f'(x)| \le 1 + \frac{1}{x}$. По формуле конечных приращений (теореме о среднем) для $\xi \in [x, x+h]$

$$|f(x+h) - f(x)| = |f'(\xi)|h \le \left(1 + \frac{1}{x}\right)h \le \left(1 + \frac{1}{\sqrt{h}}\right)h =$$
$$= h + \sqrt{h} \le 2\sqrt{h} .$$

Если $0 \le x \le \sqrt{h}$, используем оценку $|f(t)| \le t$:

$$|f(x+h) - f(x)| \le |f(x+h)| + |f(x)| \le 2x + h \le 3\sqrt{h} .$$

Итак, при всех $0 < h < 1$, $0 \le x \le 1-h$

$$|f(x+h) - f(x)| \le 3\sqrt{h} .$$

Это означает, что $\omega(f;t) \le 3\sqrt{t}$ и $f \in Lip\frac{1}{2}$. Заметим, что $f \notin Lip1$ (для любой константы $C > 0$ неравенство $|f'| > C$ выполнено на множестве положительной меры).

Задачи и упражнения к §1.

1. Показать, что нормы τ_h и Δ_h как операторов из $C(a,b)$ в $C(a,b-h)$ равны соответственно 1 и 2.

2. Вычислить модули непрерывности функций:

(a) $f(x) = e^x$, $a \le x \le b$;

(b) $f(x) = |x-1|$;

(c) $f(x) = x^k$, $0 \le x \le 1$, $k \in \mathbb{N}$.

3. Вычислить модули непрерывности функций, графики которых даны на рис. 3-6.

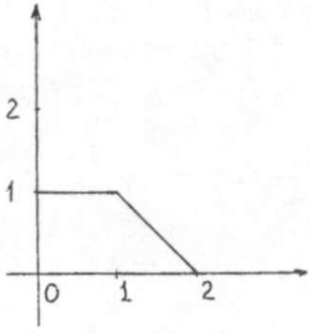

Рис. 3.

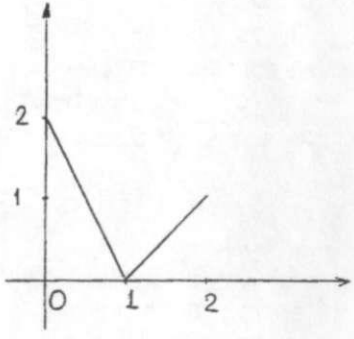

Рис. 4.

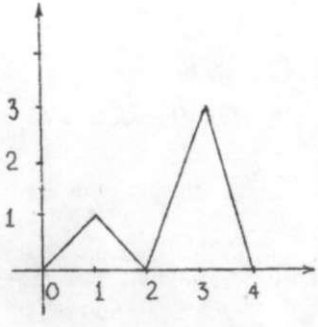

Рис. 5.

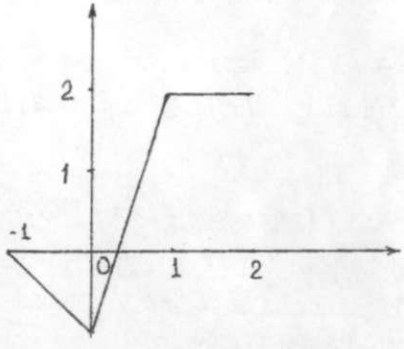

Рис. 6.

4. Показать, что если функция f обладает свойствами 1-4, то $\omega(f;t) = f(t)$.

5. Показать, что $f \in Lip\,\alpha \iff |f(x) - f(y)| \le c|x-y|^{\alpha}$.

6. Показать, что $Lip\,\alpha = \mathcal{P}_0$ при $\alpha > 1$.

7. Почему $Lip\,\alpha \subset Lip\,\beta$ при $\alpha > \beta$?

8. Показать, что функции из примеров 2,3 принадлежат классу $Lip\,1$.

9. Показать, что включение $Lip\,1 \subset DL$ является строгим. Рассмотреть для $0 \le x \le 1$ функцию $f(x) = x^{\alpha}$, $0 < \alpha < 1$.

§2. Модули непрерывности порядка k

В этом параграфе мы рассмотрим разностные характеристики функций более высокого порядка, чем первый; эти характеристики называются модулями непрерывности порядка k, $k \in \mathbb{N}$, или (в случае $k > 1$) модулями гладкости функций.

Пусть $\Delta_h = \tau_h - J$ – оператор взятия конечной разности с шагом h, введенный в §1. Напомним, что для действительной функции f $\Delta_h f(x) = f(x+h) - f(x)$. Для $k \in \mathbb{N}$ положим

$$\Delta_h^k := (\Delta_h)^k = (\tau_h - J)^k, \tag{1}$$

или, что то же самое,

$$\Delta_h^k = \Delta_h(\Delta_h^{k-1}); \quad \Delta_h^0 := J \tag{2}$$

Формула (1) задает **оператор взятия k-й разности с шагом h**. Применяя формулу для бинома к степени разности операторов $\tau_h - J$ и пользуясь тем, что для оператора сдвига $(\tau_h)^j = \tau_{jh}$, получаем эквивалентное определение:

$$\Delta_h^k = \sum_{j=0}^{k} (-1)^{k-j} C_k^j \tau_{jh}.$$

Таким образом, для функции f, по определению,

$$\Delta_h^k f(x) := \sum_{j=0}^{k} (-1)^{k-j} C_k^j f(x+jh). \tag{3}$$

Например, $\Delta_h^2 f(x) = f(x+2h) - 2f(x+h) + f(x)$. Итак, значение $\Delta_h^k f(x)$ есть линейная комбинация значений f в $(k+1)$-й точке вида $x+jh$, $j=0,\dots,k$; все эти точки, естественно, должны принадлежать области определения f. Заметим, что шаг h в (1)–(3) может быть и отрицательным. Если f определена на отрезке $[a,b]$, то $\Delta_h^k f$ как функция x определена при $h \neq 0$ на меньшем отрезке, а именно на $[a, b-kh]$ при $h>0$ или $[a-kh, b]$ при $h<0$; нужно брать $|h| \le \frac{b-a}{k}$. Как правило, мы будем рассматривать случай $0 \le h \le \frac{b-a}{k}$.

Реймона (P. Du Bois Reymond, 1879).

5. Если f обладает k - й суммируемой производной, то

$$\Delta_h^k f(x) = \int_0^h \dots \int_0^h f^{(k)}(x+u_1+\dots+u_k)\,du_1\dots du_k . \quad (4)$$

<u>Доказательство</u> осуществляется индукцией по k. При $k=1$ имеем:

$$\Delta_h f(x) = f(x+h)-f(x) = \int_x^{x+h} f'(t)\,dt = \int_0^h f'(x+u_1)\,du_1 .$$

Если аналог равенства (4) справедлив для $(k-1)$-й разности, то это равенство верно и для k -й разности.

$$\Delta_h^k f(x) = \Delta_h(\Delta_h^{k-1} f(x)) = \Delta_h\Big(\int_0^h \dots \int_0^h f^{(k-1)}(x+u_1+\dots+u_{k-1})\,du_1\dots du_{k-1}\Big) =$$

$$= \int_0^h \dots \int_0^h \Delta_h f^{(k-1)}(x+u_1+\dots+u_{k-1})\,du_1\dots du_{k-1} =$$

$$= \int_0^h \dots \int_0^h f^{(k)}(x+u_1+\dots+u_k)\,du_1\dots du_k .$$

В последнем переходе используется интегральное тождество для $\Delta_h f^{(k-1)}$.

6. Если $f \in C^k(a,b)$, то для $x \in [a, b-kh]$ найдется точка $\xi \in [x, x+kh]$ такая, что

$$\Delta_h^k f(x) = h^k f^{(k)}(\xi) . \quad (5)$$

Поэтому $|\Delta_h^k f(x)| \le h^k \|f^{(k)}\|_{C(a,b)}$ и

$$f^{(k)}(x) = \lim_{h\to 0} \Delta_h^k f(x) / h^k . \quad (6)$$

Равенство (5) получается из (4) после k -кратного применения

Свойства k -х разностей

1. Пусть $0 \le h \le \frac{b-a}{k}$. Тогда Δ_h^k - линейный оператор из $C(a,b)$ в $C(a, b-kh)$; при этом

$$\| \Delta_h^k \|_{C(a,b) \to C(a, b-kh)} = 2^k .$$

Докажите самостоятельно по аналогии со случаем $k=1$.

2. $\Delta_h^{k+j} = \Delta_h^k \, \Delta_h^j$.

Следует из определения по формуле (1)

3. Δ_h^k аннулирует многочлены степени $\le k-1$.

Действительно, если $p \in \mathcal{P}_m$, то $\Delta_h p \in \mathcal{P}_{m-1}$:

$$\Delta_h (c x^m) = c \{ (x+h)^m - x^m \} \in \mathcal{P}_{m-1} .$$

Теперь отмеченное свойство получается индукцией из (2).

4. Если $f \in C(a,b)$ и при всех $0 \le h \le \frac{b-a}{k}$ и $x \in [a, b-kh]$ $\Delta_h^k f(x) = 0$, то $f \in \mathcal{P}_{k-1}$.

<u>Доказательство</u>. Для $\varphi \in C_0^\infty(a,b)$ с носителем $E \subset [a,b]$ при достаточно малых h

$$0 = \int_E \varphi(x) \, \Delta_h^k f(x) \, dx = \int_a^b f(x) \, \Delta_{-h}^k \varphi(x) \, dx$$

(замена переменных). Разделим это равенство на h^k и перейдем к пределу при $h \to 0$. Тогда

$$\lim_{h \to 0} \frac{\Delta_{-h}^k \varphi(x)}{h^k} = \varphi^{(k)}(x)$$

(см. ниже (6)). Поэтому для всех $\varphi \in C_0^\infty(a,b)$

$$\int_a^b f \, \varphi^{(k)} \, dx = 0.$$

Это означает, что k -я обобщенная производная f равна нулю. Поэтому $f \in \mathcal{P}_{k-1}$. Для $k=1$ это результат известной леммы Дюбуа-

теоремы о среднем.

7. Имеет место тождество

$$\Delta_{nh}^{k} f(x) = \sum_{i_1=0}^{n-1} \dots \sum_{i_k=0}^{n-1} \Delta_h^k f(x + i_1 h + \dots + i_k h) \qquad (7)$$

(индукция по k).

Пусть теперь $f \in C(a,b)$, $k \in \mathbb{N}$.

<u>Определение</u>. Модулем непрерывности функции f порядка k называется функция $\omega_k(f;t) = \omega_k(t)$ аргумента $t \in [0, \frac{b-a}{k}]$, определяемая равенством

$$\omega_k(f;t) := \sup_{0 \le h \le t} \| \Delta_h^k f \|_{C(a,b-kh)} .$$

Часто полагают $\omega_k(t) := \omega_k(\frac{b-a}{k})$ при $t > \frac{b-a}{k}$.

<u>Примеры</u>. 1. Пусть $f(x) = x^k$, $x \geq 0$. Тогда $f^{(k)} = k!$ и в соответствии с (5) $\Delta_h^k f(x) = k!\, h^k$. Поэтому $\omega_k(f;t) = k!\, t^k$.

2. Для функции $f(x) = e^x$, $a \le x \le b$,

$$\Delta_h^k f(x) = (e^h - 1)^k e^x ,$$

поэтому при $t \in [0, \frac{b-a}{k}]$

$$\omega_k(f;t) = \sup_{0 \le h \le t} \sup_{a \le x \le b-kh} (e^h - 1)^k e^x =$$
$$= e^b \sup_{0 \le h \le t} (1 - e^{-h})^k = e^b (1 - e^{-t})^k .$$

Свойства модулей непрерывности порядка k .

1. $\omega_k(0) = (0)$.

2. $\omega_k(t)$ не убывает по t .

3. $\omega_k(t)$ непрерывна по t .

4. Если $n \in \mathbb{N}$, $nt \le \frac{b-a}{k}$, то $\omega_k(nt) \le n^k \omega_k(t)$.

5. Если $\lambda \in \mathbb{R}$, то $\omega_k(\lambda t) \le (\lambda+1)^k \omega_k(t)$.

6. $\omega_k(t_1+t_2) \le 2^k(\omega_k(t_1)+\omega_k(t_2))$.

7. $\omega_k(f;t)=0$ при всех $t\in[0,\frac{b-a}{k}] \Leftrightarrow f\in\mathcal{P}_{k-1}$.

8. Для $t\in[0,\frac{b-a}{k}]$ $\omega_k(t)\ge ct^k$; при этом $c\ne 0$, если $f\notin\mathcal{P}_{k-1}$

9. Если $f\in C^k(a,b)$, то $\omega_k(f;t)\le t^k\|f^{(k)}\|_{C(a,b)}$

10. $\omega_k(\alpha f+\beta g;t)\le|\alpha|\omega_k(f;t)+|\beta|\omega_k(g;t)$, $\alpha,\beta\in\mathbb{R}$.

11. $\omega_k(\alpha f;t)=|\alpha|\,\omega_k(f;t)$, $\alpha\in\mathbb{R}$.

12. $\omega_k(f;t)\le 2^k\|f\|_{C(a,b)}$.

Некоторые другие свойства модулей непрерывности отмечаются в §3.

Остановимся коротко на <u>доказательствах</u> свойств 1-12 (аккуратное обоснование предоставляется читателю). Первые два свойства очевидны. Третье свойство устанавливается с помощью оценки

$$\omega_k(t_2)-\omega_k(t_1)\le\gamma(k)\,\omega_1(t_2-t_1),\quad t_2\ge t_1, \tag{8}$$

и свойств $\omega_1(t)$. Свойство 4 легко получается из тождества (7); из монотонности $\omega_k(t)$ теперь следует свойство 5.

Установим свойство 6. Пусть $t_2\ge t_1$, тогда

$$\omega_k(t_2+t_1)\le\omega_k(2t_2)\le 2^k\omega_k(t_2)\le 2^k(\omega_k(t_1)+\omega_k(t_2)).$$

Для доказательства важной эквивалентности свойства 7 нужно использовать свойства 3-4 k-х разностей.

Далее, если $t\in(0,\frac{b-a}{k}]$, то

$$\omega_k\left(\frac{b-a}{k}\right)=\omega_k\left(\frac{b-a}{kt}t\right)\le\left(\frac{b-a}{kt}+1\right)^k\omega_k(t)\le$$
$$\le\left(\frac{2(b-a)}{kt}\right)^k\omega_k(t),$$

что дает оценку $\omega_k(t) \geq c t^k$ с константой

$$c = \left(\frac{k}{2(b-a)} \right)^k \omega_k\left(\frac{b-a}{k} \right)$$

(свойство 8).

Оценка свойства 9 сразу получается из неравенства

$$|\Delta_h^k f(x)| \leq h^k \, \| f^{(k)} \|_{C(a,b)} .$$

Наконец, свойства 10-12 следуют из определения ω_k.

Заметим, что при любом фиксированном $t \in (0, \frac{b-a}{k}]$ величина

$$|f|_t := \omega_k(f;t) , \quad f \in C(a,b),$$

задает, как говорят, норму на факторпространстве $C/\mathcal{P}_{k-1}$. Последнее в точности означает, что выполнены следующие условия:

1. $|f|_t \geq 0 \; ; \; |f|_t = 0 \iff f \in \mathcal{P}_{k-1}$

2. $|\alpha f|_t = |\alpha| \, |f|_t \quad , \alpha \in \mathbb{R}$.

3. $|f+g|_t \leq |f|_t + |g|_t$.

Подобным условиям удовлетворяет и величина наилучшего (локального) приближения с помощью многочленов степени $\leq k-1$:

$$e_{k-1}(f)_Q := \inf_{p \in \mathcal{P}_{k-1}} \| f - p \|_{C(Q)} ,$$

Q - произвольный фиксированный отрезок, $Q \subset [a,b]$. Как мы покажем в §4, для $t = \frac{b-a}{k}$ и $Q = [a,b]$ эти нормы эквивалентны.

Модули непрерывности порядка k лежат в основе определения многих важных функциональных классов, изучаемых в теории функций, функциональном анализе и их приложениях; мы не будем касаться этой обширной тематики.

Задачи и упражнения к §2

1. Доказать первое свойство k-х разностей.

2. Доказать тождество $\sum_{j=0}^{k}(-1)^{k-j} C_k^j = 0$, используя свойства k -х разностей.

3. Доказать тождество (7) для k -х разностей.

4. Найти $\omega_k(f;t)$ для $f(x) = \sin x$, $x \in \mathbb{R}$.

5. Вычислить $\omega_2(f;t)$ для $f(x) = |x|$, $x \in [-1,1]$.

6. Установить неравенство (8) с $\gamma(k) = \sum_{i=0}^{k} i\, C_k^i$.

7. Доказать свойство 4 k - х модулей.

8. Пусть $Lip^{(k)}\alpha = \{ f \in C(a,b) : \omega_k(f,t) \le ct^{\alpha} \}$. Показать, что

(a) $Lip^{(k)}\alpha \subset Lip^{(k)}\beta$ при $\alpha \ge \beta$,

(b) $Lip^{(k)}\alpha = \mathcal{P}_{k-1}$ при $\alpha > k$.

§3. Соотношения между модулями непрерывности различного порядка и вида

Как связаны между собой модули непрерывности одной и той же функции f различных порядков? Простое неравенство оценивает "старший" модуль непрерывности через "младший".

Утверждение 1. Пусть $k > j \ge 1$, $f \in C(a,b)$. Тогда

$$\omega_k(f;t) \le 2^{k-j}\, \omega_j(f;t), \quad 0 \le t \le \frac{b-a}{k} . \qquad (1)$$

Доказательство. Из свойства 2 k -х разностей (см. §2) следует, что при $a \le x \le b - kh$ ($h \le \frac{b-a}{k}$)

$$|\Delta_h^k f(x)| = |\Delta_h^{k-j}\, \Delta_h^j f(x)| =$$

$$= \left|\sum_{i=0}^{k-j}(-1)^{k-j-i}\, C_{k-j}^i\, \Delta_h^j f(x+ih)\right| \le$$

$$\leq \left(\sum_{i=0}^{k-j} C_{k-j}^{i} \right) \| \Delta_h^j f \|_{C(a,b-jh)} .$$

Здесь используется то, что для $i = 0, \dots, k-j$

$$\sup_{a \leq x \leq b-kh} | \Delta_h^j f(x+ih) | \leq \sup_{a \leq y \leq b-jh} | \Delta_h^j f(y) | .$$

Итак,

$$\| \Delta_h^k f \|_{C(a,b-kh)} \leq 2^{k-j} \| \Delta_h^j f \|_{C(a,b-jh)} ,$$

что дает оценку (1).

Неравенство (1) означает, в частности, что модули непрерывности 2, 3 и т.д. порядков оцениваются сверху через модуль непрерывности первого порядка.

Оценка "младшего" модуля непрерывности через "старший" выглядит (и получается) существенно сложнее. Справедлива следующая теорема, доказанная в 1927 году Маршо (A. Marchaud).

<u>Теорема 1</u>. Для $f \in C(a,b)$, $k \in \mathbb{N}$ и $j = 1, \dots, k$ имеет место неравенство

$$\omega_j(f;t) \leq A_k t^j \left(\frac{\| f \|_{C(a,b)}}{(b-a)^j} + \int_t^{\frac{b-a}{2j}} \frac{\omega_{k+1}(f;u)}{u^{j+1}} du \right) . \quad (2)$$

Здесь A_k - постоянная, зависящая только от k .

Мы приведем здесь <u>доказательство</u> неравенства Маршо (2) для случая $j = k = 1$; в общем случае используются более сложные разностные тождества (см., например, [9]).

Прежде всего заметим, что

$$\Delta_{2h}^1 f(x) = \Delta_h^1 f(x) + \Delta_h^1 f(x+h) .$$

Это равенство верно при всех $x \in [a, \frac{a+b}{2}]$ и $h \in [0, \frac{b-a}{4}]$. Поэтому

$$| \Delta_{2h}^1 f(x) - 2 \Delta_h^1 f(x) | = | \Delta_h^1 f(x+h) - \Delta_h^1 f(x) | =$$
$$= | \Delta_h^2 f(x) | \leq \omega_2(f;h) .$$

Имеем оценку

$$|\Delta_h^1 f(x)| \le \frac{1}{2}|\Delta_{2h}^1 f(x)| + \frac{1}{2}\omega_2(f;h)$$

Эту оценку можно продолжить, если учесть, что

$$|\Delta_{2h}^1 f(x)| \le \frac{1}{2}|\Delta_{4h}^1 f(x)| + \frac{1}{2}\omega_2(f;2h).$$

Таким образом, для любого натурального m

$$|\Delta_h^1 f(x)| \le \frac{1}{2^m}|\Delta_{2^m h}^1 f(x)| + \frac{1}{2}\sum_{i=0}^{m-1}\frac{\omega_2(f;2^i h)}{2^i}.$$

Первое слагаемое оценим величиной $2^{-m+1}\|f\|_{C(a,b)}$; чтобы оценить второе слагаемое, воспользуемся результатом задачи 3 §3, в силу которого от суммы можно перейти к интегралу. Действительно,

$$\frac{1}{2}\sum_{i=0}^{m-1}\frac{\omega_2(f;2^i h)}{2^i} \le h\sum_{i=0}^{m-1}\int_{2^i h}^{2^{i+1}h}\frac{\omega_2(f;u)}{u^2}du = h\int_h^{2^m h}\frac{\omega_2(f;u)}{u^2}du.$$

Итак, если m выбрать так, что $\frac{b-a}{4} < 2^m h < \frac{b-a}{2}$, то

$$|\Delta_h^1 f(x)| \le \frac{16h}{b-a}\|f\|_{C(a,b)} + 2h\int_h^{\frac{b-a}{2}}\frac{\omega_2(f;u)}{u^2}du$$

Это неравенство сохранится также при всех $x \in [\frac{a+b}{2}, b]$. Для этого нужно на $[-b,-a]$ рассмотреть функцию $f(-x)$. Неравенство (2) для $j=k=1$ и $t \le \frac{b-a}{4}$ следует из полученной выше оценки для $|\Delta_h^1 f(x)|$. Осталось доказать неравенство при $\frac{b-a}{2} > t > \frac{b-a}{4}$. Из пятого свойства модуля непрерывности следует, что

$$\omega_1(f;t) \le 2\omega_1(f;\tfrac{t}{2}),$$

поэтому все сводится к случаю, который мы уже рассмотрели. Таким образом, неравенство Маршо (2) для $j=k=1$ полностью доказано.

Другое доказательство неравенства Маршо в общем случае предла-

гается в §6.

Наиболее просто неравенство Маршо выглядит для $f \in C(0,1)$. Считая, что $\omega_k(f;t)=\omega_k(f;\frac{1}{k})$ при $t>\frac{1}{k}$, запишем неравенство (2) в виде

$$\omega_j(f;t) \le A_k t^j \Big(\|f\|_{C(0,1)} + \int_t^1 \frac{\omega_{k+1}(f;u)}{u^{j+1}}\, du \Big). \qquad (3)$$

Отметим здесь одно следствие неравенства (3), на основе которого мы получим используемый в §4 критерий предкомпактности подмножеств $C(0,1)$.

<u>Утверждение 2</u>. Пусть $k \in \mathbb{N}$, $\mathcal{A}$ - подмножество $C(0,1)$, для которого $\sup\limits_{f\in\mathcal{A}} \|f\|_{C(0,1)} = M < \infty$. Тогда два условия

$$\omega_1(\mathcal{A};t) := \sup_{f\in\mathcal{A}} \omega_1(f;t) \to 0, \qquad (4)$$

$$\omega_k(\mathcal{A};t) := \sup_{f\in\mathcal{A}} \omega_k(f;t) \to 0 \qquad (5)$$

$(t \to +0)$ эквивалентны.

<u>Доказательство</u>. Ясно, что из (4) следует (5) в силу неравенства $\omega_k(f;t) \le 2^{k-1}\omega_1(f;t)$, $0<t\le\frac{1}{k}$ (см. утверждение 1). Пусть теперь имеет место (5). Применяя неравенство (3) для $j=1$ к функциям $f\in\mathcal{A}$, получим

$$\omega_1(\mathcal{A};t) \le A_{k-1}\, t \Big(M + \int_t^1 \frac{\omega_k(\mathcal{A};u)}{u^2}\, du \Big). \qquad (6)$$

Правая часть последнего соотношения стремится к нулю при $t\to+0$. Действительно, если $\varepsilon>0$ - произвольно, то для достаточно малого $\delta=\delta(\varepsilon)$ и $0<u\le\delta$ будет $\omega_k(\mathcal{A};u)<\varepsilon$. Считая $t<\delta$, имеем:

$$t\int_t^1 \frac{\omega_k(\mathcal{A};u)}{u^2}\, du = t\Big(\int_t^\delta + \int_\delta^1\Big) \frac{\omega_k(\mathcal{A};u)}{u^2}\, du \le$$

$$\leq \varepsilon t \int_t^{\delta} \frac{du}{u^2} + 2^k M t \int_{\delta}^{1} \frac{du}{u^2} = \varepsilon t \left(\frac{1}{t} - \frac{1}{\delta} \right) +$$

$$+ 2^k M t \left(\frac{1}{\delta} - 1 \right) \leq \varepsilon + \frac{2^k M}{\delta} t .$$

Поэтому

$$\overline{\lim_{t \to +0}} \; t \int_t^1 \frac{\omega_k(A; u)}{u^2} du \leq \varepsilon .$$

В силу произвольности ε отсюда следует, что последний предел равен нулю. Тогда в силу (6)

$$\lim_{t \to +0} \omega_1(A; t) = 0 ,$$

что и требовалось доказать.

Результат утверждения 2 справедлив и для функций из $C(a,b)$.

Подмножество $A \subset C(a,b)$ называется **предкомпактным**, если его замыкание $\bar{A}$ компактно в $C(a,b)$. Последнее эквивалентно тому, что произвольная бесконечная последовательность элементов A имеет предельную точку, возможно, не принадлежащую A (иначе: для любой последовательности функций $f_n \in A$ существует подпоследовательность $\{f_{n_i}\}$, сходящаяся равномерно на $[a,b]$ к некоторой $f \in C(a,b)$). Как известно, критерий предкомпактности подмножества A дается следующей **теоремой Арцела** (C. Arzela, 1893; см., например, [10], стр. 110). A предкомпактно в $C(a,b)$ тогда и только тогда, когда A равномерно ограничено (то есть $\sup_{f \in A} \|f\|_{C(a,b)} < \infty$) и равностепенно непрерывно. Последнее условие в терминах модуля непрерывности первого порядка означает, что

$$\lim_{t \to 0} \sup_{f \in A} \omega_1(f; t) = 0 .$$

Если A равномерно ограничено, то его равностепенная непрерывность характеризуется аналогичным условием для k - го модуля непрерывности, см. утверждение 2 (там $[a,b] = [0,1]$). Таким образом, справедливо следующее обобщение теоремы Арцела.

Теорема 2. Подмножество $\mathcal{A}$ предкомпактно в $C(a,b)$ тогда и только тогда, когда одновременно выполнены условия :

1. $\sup\limits_{f\in\mathcal{A}} \|f\|_{C(a,b)} < \infty$.

2. Для некоторого $k\in\mathbb{N}$ $\lim\limits_{t\to+0} \sup\limits_{f\in\mathcal{A}} \omega_k(f;t) = 0$.

Одно из приложений этого критерия предкомпактности мы отметим в следующем параграфе.

Обратим внимание на другую возможность для определения модулей непрерывности функций $f\in C(a,b)$. Операцию взятия верхней грани по h в определении $\omega_k(f;t)$ можно заменить интегрированием (точнее усреднением) по h ; так возникает следующий **интегральный модуль непрерывности** функции $f\in C(a,b)$ порядка $k\in\mathbb{N}$:

$$\mathcal{J}_k(f;t) := \int_0^1 \|\Delta_{th}^k f\|_{C(a,b-kth)}\, dh . \tag{7}$$

Очевидно, при всех $t\in[0,\frac{b-a}{k}]$

$$\mathcal{J}_k(f;t) \le \sup_{0\le h\le 1} \|\Delta_{th}^k f\|_{C(a,b-kth)} =$$

$$= \sup_{0\le y\le t} \|\Delta_y^k f\|_{C(a,b-ky)} = \omega_k(f;t).$$

Оказывается, справедливо и противоположное неравенство: существует не зависящая от $f\in C(a,b)$ и $t\in[0,\frac{b-a}{k}]$ константа $\gamma(k)$ такая, что

$$\omega_k(f;t) \le \gamma(k)\, \mathcal{J}_k(f;t). \tag{8}$$

Неравенство (8) вместе с оценками для константы $\gamma(k)$ можно получить, используя некоторое разностное тождество. Доказательство является техническим и здесь не приводится.

В заключение параграфа отметим, что приведенные в нем утверждения переносятся на случай функциональных пространств, отличных от $C(a,b)$, и на случай функций многих переменных.

Задачи и упражнения к §3

1. Доказать, что

$$\omega_{k+1}(f;2^i h) \le 2k\cdot 2^{ik} h^k \int_{2^i h}^{2^{i+1}h} \frac{\omega_{k+1}(f;u)}{u^{k+1}}\,du$$

2. Доказать неравенство Маршо для $j=k$. Указание. Воспользовавшись свойством 7 §2 для разностей, показать, что

$$\Delta_{2h}^k f(x) = \sum_{i=0}^{k} C_k^i\, \Delta_h^k f(x+ih).$$

Затем оценить величину $|\Delta_{2h}^k f(x) - 2^k \Delta_h^k f(x)|$ через $(k+1)$-й модуль непрерывности.

3. Пусть $\gamma(k)$ - наименьшая константа, с которой (8) выполняется для всех $f\in C(0,1)$ и $t\in[0,\frac{1}{k}]$. Показать, что $\gamma(k)\to\infty$ при $k\to\infty$.

4. Для $f\in C(0,1)$ получить более слабое, чем (8), неравенство

$$\omega_k(f;t) \le c(k)\int_0^1 \omega_k(f;th)\,dh .$$

Использовать свойства $\omega_k(f;t)$ (в частности, $\omega_k(2t)\le 2^k\omega_k(t)$).

§4. Теорема Уитни. Константы Уитни

Пусть функция f определена и непрерывна на отрезке $I:=[0,1]$. Если f имеет ограниченную производную порядка k, то для ее тейлоровского многочлена $p_{k-1}\in\mathcal{P}_{k-1}$ выполнено

$$|f(x)-p_{k-1}(x)| \le \frac{1}{k!}\sup|f^{(k)}|$$

для всех $x\in I$. Аналог этого свойства для функций, не имеющих производных, устанавливает замечательная теорема американского математика Хасслера Уитни (H.Whitney, 1907-1989). Х.Уитни оставил яркий след в ряде различных областей математики. Теорема, о которой пойдет речь, опубликована в 1957 году, см. [54]; она стала классическим результатом теории приближения функций.

Теорема 1. Для любого натурального k существует положительное число W_k такое, что если $f \in C(I)$, то найдется некоторый многочлен $p = p_f$ степени $\leq k-1$, для которого при всех $x \in I$ будет выполнено

$$|f(x) - p(x)| \leq W_k \sup_{y,h} |\Delta_h^k f(y)|. \qquad (1)$$

Верхняя грань в правой части (1) взята по множеству всех $h \in [0, \frac{1}{k}]$ и $y \in I$ таких, что $y + kh \in I$, то есть $y \in [0, 1-kh]$.

Наименьшее число W_k, удовлетворяющее условию теоремы, называется **k -й константой Уитни**. Вопрос о значениях W_k обсуждается в конце параграфа.

К настоящему времени известны различные доказательства теоремы 1, использующие, например, регуляризацию f или интерполяционные многочлены. Мы приведем короткое доказательство теоремы Уитни, использующее критерий предкомпактности подмножеств $C(I)$, формулируемый в терминах k-го модуля непрерывности. Идея использования соображений компактности в доказательстве теоремы Уитни и ряда ее аналогов принадлежит Ю. А. Брудному.

Предварительно переформулируем теорему 1 в терминах наилучшего приближения. Положим

$$e_k(f) := e(f; \mathcal{P}_k)_{C(I)} = \inf_{g \in \mathcal{P}_k} \|f - g\|_{C(I)}.$$

Как известно, $e_k(f)$ называется **величиной наилучшего приближения** (или коротко **наилучшим приближением**) f с помощью $\mathcal{P}_k$ в метрике пространства $C(I)$, см. [2] по поводу свойств. Напомним, что для произвольной $f \in C(I)$ и $k = 0, 1, \ldots$ существует единственный **многочлен наилучшего приближения (м. н. п.)** $p^* \in \mathcal{P}_k$, то есть такой, что

$$\|f - p^*\|_{C(I)} = e_k(f).$$

Этот многочлен характеризуется **теоремой Чебышева об альтернансе**. Практическое нахождение p^* по заданной f осуществляется с помощью **алгоритма Ремеза** и его модификаций, подробно разобранных в [4]. С вычислительной точки зрения задача нахождения p^* является весьма трудоемкой.

Утверждение теоремы Уитни в терминах наилучшего приближения $e_{k-1}(f)$ и модуля непрерывности $\omega_k(f;t)$ выглядит так: существует константа W_k такая, что для всех $f \in C(I)$

$$e_{k-1}(f) \le W_k\, \omega_k(f). \qquad (2)$$

Здесь и далее $\omega_k(f)=\omega_k(f;\frac{1}{k})$ - максимальное значение модуля непрерывности (эту величину также называют **колебанием порядка** k функции f).

Отметим сначала, что неравенство, противоположное (2), устанавливается достаточно просто. Действительно, пусть p^* - м.н.п. из $\mathcal{P}_{k-1}$ для f . Так как $\Delta_h^k p^*=0$ при всех h , то

$$\omega_k(f) = \omega_k(f-p^*) \le 2^k \|f-p^*\|_{C(I)} = 2^k e_{k-1}(f) \qquad (3)$$

для любой $f \in C(I)$. Соединенные вместе, неравенства (2) и (3) означают, что величины $e_{k-1}(f)$ и $\omega_k(f)$ эквивалентны с константами, не зависящими от f :

$$e_{k-1}(f) \asymp \omega_k(f). \qquad (4)$$

Запись $A \asymp B$ для $A=A(t)$ и $B=B(t)$ здесь и далее означает, что с независящими от t постоянными c_1, c_2

$$c_1 A(t) \le B(t) \le c_2 A(t).$$

Эквивалентности типа (4) позволяют переходить в различных соотношениях от модулей непрерывности к наилучшим приближениям (или, шире, от разностных к аппроксимационным характеристикам) функций, и наоборот. Этот плодотворный подход является одним из основных методов современной теории приближения функций.

Перейдем к <u>доказательству</u> теоремы Уитни в форме неравенства (2). Предположим, что не существует не зависящей от f константы W_k , с которой (2) выполнено для всех $f \in C(I)$. Тогда существует последовательность $\{f_n\} \subset C(I)$ такая, что

$$e_{k-1}(f_n) > n\, \omega_k(f_n).$$

Пусть p_n - м.н.п. для f_n из $\mathcal{P}_{k-1}$: положим $\psi_n := \dfrac{f_n - p_n}{\|f_n - p_n\|}$, $\|\cdot\| := \|\cdot\|_{C(I)}$.

Тогда по свойствам наилучших приближений и модулей непрерывности

$$e_{k-1}(\psi_n) = \frac{e_{k-1}(f_n)}{\|f_n - p_n\|} = \|\psi_n\| = 1 \; ; \tag{5}$$

$$\omega_k(\psi_n) = \frac{\omega_k(f_n)}{\|f_n - p_n\|} = \frac{\omega_k(f_n)}{e_{k-1}(f_n)} < \frac{1}{n} \; . \tag{6}$$

Пусть $\mathcal{A} = \{\psi_n\}$. Покажем, что $\mathcal{A}$ предкомпактно в $C(I)$, т.е. содержит последовательность, сходящуюся к элементу из $C(I)$. Используем критерий предкомпактности подмножеств $C(I)$, полученный в §3, см. теорему 2 на стр. 24. Для применения этого критерия проверим, что

$$\sup_n \omega_k(\psi_n; t) \longrightarrow 0 \; , \quad t \to 0 \; . \tag{7}$$

Пусть $\varepsilon > 0$. Выберем $N = N(\varepsilon)$ так, чтобы было

$$\sup_{n > N} \omega_k(\psi_n) < \varepsilon \; .$$

Это возможно в силу (6). Так как $\omega_k(f, t) \to 0$ при $t \to 0$ для каждой $f \in C(I)$, то и

$$\max_{1 \le n \le N} \omega_k(\psi_n; t) \to 0 \; , \quad t \to 0 \; .$$

Поэтому существует t^* такое, что

$$\max_{1 \le n \le N} \omega_k(\psi_n; t^*) < \varepsilon \; .$$

Тогда при $t < \min\left(t^*; \frac{1}{k}\right)$

$$\sup_{n \in \mathbb{N}} \omega_k(\psi_n; t) < \varepsilon \; .$$

Это и означает, что выполняется (7).

Так как

$$\sup_n \|\psi_n\| = 1 \; ,$$

то $\mathcal{A}$ предкомпактно в $C(I)$. Пусть ψ - предельная точка $\mathcal{A}$: $\psi_{n_i} \to \psi$ (в $C(I)$). Переходя в соотношениях (5), (6) к пределу при $i \to \infty$, получаем, что для этой непрерывной функции ψ одновременно

$$e_{k-1}(\psi) = 1 \quad , \quad \omega_k(\psi) = 0.$$

Второе из равенств означает, что $\psi \in \mathcal{P}_{k-1}$ (свойство 7 §2). Это противоречит первому равенству. Полученное противоречие доказывает теорему.

Замечание. В случае $k=1$ (приближение константами) эквивалентность (4) совпадает с равенством

$$e_0(f) = \frac{1}{2}\,\omega_1(f);$$

читателю предлагается убедиться в том, что оно имеет ясный геометрический смысл. Для $k=2$ (приближение линейными функциями) точное на классе $C(I)$ неравенство Уитни имеет вид

$$e_1(f) \le \frac{1}{2}\,\omega_2(f). \tag{8}$$

Покажем, как (8) можно легко получить с помощью теоремы Чебышева об альтернансе. Пусть $\rho(x) = ax+b$ - м.н.п. для f. Тогда существует альтернанс разности $f-\rho$, состоящий из трех точек x_1, $x_2, x_3 \in [0,1]$. Напомним, что x_i обладают свойствами:

$$\|f-\rho\|_{C(I)} = e_1(f) = |(f-\rho)(x_i)|,$$

$$(f-\rho)(x_1) = -(f-\rho)(x_2) = (f-\rho)(x_3).$$

Выберем из x_1, x_2, x_3 две точки t', t'', чтобы $t', t''=t'+h$, $t'+2h \in [0,1]$. Тогда

$$\Delta_h^2 f(t') = \Delta_h^2 (f-\rho)(t') = (f-\rho)(t'+2h) - 2(f-\rho)(t'+h) +$$

$$+(f-\rho)(t') = (f-\rho)(t'+2h) - 3(f-\rho)(t'+h) =$$

$$= (f-\rho)(t'+2h) \pm 3e_1(f).$$

Так как $|(f-\rho)(t'+2h)| \le \|f-\rho\|_{C(I)} = e_1(f)$, то для этих t' и h

$$|\Delta_h^2 f(t')| = |(f-\rho)(t'+2h) \pm 3e_1(f)| \geqslant 2e_1(f).$$

Если $h<0$, то точки, по которым берется вторая разность, нужно взять в обратном порядке. Можно считать поэтому, что $h>0$. Тогда

последнее неравенство гарантирует (8).

Константа $\frac{1}{2}$ в неравенстве (8) не может быть уменьшена (см. упр. 1 в конце параграфа).

Для приложений важным является вопрос о точном значении или хороших оценках констант Уитни W_k, $k \in \mathbb{N}$. Такие оценки зависят от способа доказательства теоремы 1. В своей работе [54] Х. Уитни использовал многочлены, интерполирующие f на равномерной сетке; им были получены оценки W_k для небольших k:

$$W_1 = W_2 = \frac{1}{2}\ ;\ \frac{8}{15} \le W_3 \le \frac{7}{10}\ ;\ \frac{1}{2} \le W_4 \le 3.2425\ ;\ \frac{1}{2} \le W_5 \le 10.4.$$

Первые оценки W_k при всех k были получены Ю. А. Брудным. Лучшая из опубликованных к 1983 году оценок имела вид $W_k \le (k+1)k^k$. Константам Уитни посвящен цикл работ Б. Сендова и других математиков болгарской школы по теории функций - последовательные оценки W_k в виде $O(k \ln k)$ (К. Иванов, М. Такев), $O(k)$ (П. Бинев). Важный шаг вперед сделан Б. Сендовым [52], доказавшим ограниченность констант Уитни числом 6 (см. также монографию [15]). Затем эта граница понижалась рядом авторов. В 1993 году Ю. В. Крякин доказал [31], что для многочлена $p \in \mathcal{P}_{k-1}$, интерполирующего $f \in C(I)$ "в среднем", то есть такого, что

$$\int_0^{i/k} (f(x) - p(x))\, dx = 0\ ,\ i = 1, \dots, k, \tag{9}$$

выполнено

$$|f(x) - p(x)| \le 2\, \omega_k(f),\ x \in I\ ;$$

Это, в частности, означает, что

$$W_k \le 2\ ,\ k \in \mathbb{N}, \tag{10}$$

- наилучшая оценка констант Уитни на настоящий момент. Заметим, что в 1982 году Б. Сендов опубликовал гипотезу о том, что $W_k \le 1$. Вопрос о точном значении W_k сам Х. Уитни считал очень трудным.

Отметим в этом параграфе одно из применений теоремы Уитни - оценку величины приближения $f \in C(I)$ с помощью так называемых проекционных операторов на пространства многочленов.

Линейный непрерывный оператор $P: C(I) \to \mathcal{P}_k$ называется проекционным оператором на $\mathcal{P}_k$ (или коротко **проектором**), если $Pg = g$ для всех $g \in \mathcal{P}_k$. Важный пример проекторов дают интерполяционные операторы (их оценки даны в §6). Если P - проектор на $\mathcal{P}_k$, то для фиксированной f многочлен $p = Pf$ часто используют как приближающий объект. Для приложений важно, что в отличие от оператора метрической проекции $f \longmapsto p^*$, сопоставляющего функции $f \in C(I)$ ее м.н.п. из $\mathcal{P}_k$, оператор P является линейным и обычно просто реализуемым. Естественно, при замене p^* на Pf мы, вообще говоря, проигрываем в точности приближения. Тем не менее, если $P: C(I) \to \mathcal{P}_k$ - проектор, то для $f \in C(I)$ справедливо двойное неравенство

$$e_k(f) \le \| f - Pf \|_{C(I)} \le (1 + \|P\|)\, e_k(f). \qquad (11)$$

Правое неравенство в (11) называется **неравенством Лебега**. Здесь $\|P\|$ - норма P как оператора из $C(I)$ в $C(I)$. Ясно, что для любого проектора $\|P\| \ge 1$; проектор P с $\|P\| = 1$ будем называть **наилучшим**. Неравенство Лебега нетрудно доказать непосредственно из определения проектора.

Для оценки $e_k(f)$ в правой части (11) мы можем применить неравенства (2), (10) этого параграфа. Это дает возможность оценить точность приближения функции f с помощью Pf через $\omega_k(f)$:

$$\| f - Pf \|_{C(I)} \le 2\,(1 + \|P\|)\, \omega_k(f). \qquad (12)$$

Отметим, что с помощью замены переменных результаты этого параграфа переносятся на случай $C(a, b)$ с теми же постоянными; в этой ситуации $\omega_k(f) = \omega_k(f; \frac{b-a}{k})$ и т.д. Кроме того, аналог теоремы 1 справедлив и для функций $f \in C(\mathbb{R})$ или $f \in C(0, +\infty)$; соответствующие константы Уитни W_k^{**} и W_k^* удовлетворяют неравенствам

$$W_k^{**} \le W_k^* \le W_k \;; \quad \frac{1}{2} \le W_k^* \le 1 \;;$$

$$W_k^{**} \le \left(C_k^{[k/2]} \right)^{-1} \approx 2^{-k} \sqrt{\frac{\pi k}{2}}.$$

Задачи и упражнения к §4

1. Показать, что константа $\frac{1}{2}$ в неравенстве (8) не может быть уменьшена. Для этого рассмотреть при $\varepsilon \to 0$ последовательность функций

$$f_\varepsilon(x) := \begin{cases} \frac{1-\varepsilon}{\varepsilon} x, & 0 \le x \le \varepsilon \\ 1-x, & \varepsilon < x \le 1. \end{cases}$$

2. Проверить, что из справедливости теоремы 1 для отрезка $[0,1]$ следует ее справедливость для отрезка $[a,b]$ и соответствующие константы W_k совпадают.

3. Показать, что $\|P\| \geqslant 1$ для любого проектора P на линейное подпространство $X \subset C(a,b)$.

4. Доказать неравенство Лебега (см. (11)).

5. Доказать, что существует единственный многочлен $p \in \mathcal{P}_{k-1}$, интерполирующий функцию f "в среднем", и, таким образом, равенство (9) задает проектор $\widetilde{P}_{k-1}: C(I) \to \mathcal{P}_{k-1}$ (полагаем $\widetilde{P}_{k-1} f := p$).

6. Показать, что $\|\widetilde{P}_1\| = 2$.

§5. Кусочно-полиномиальная и сплайн-аппроксимация функций

Теорема Уитни утверждает, что скорость приближения непрерывной функции многочленами зависит от ее модуля непрерывности:

$$e_{k-1}(f)_{C(a,b)} \le c\, \omega_k\left(f; \frac{b-a}{k}\right)_{C(a,b)}.$$

Существует много функций, которые хорошо приближаются многочленами, например функции без особенностей. Если нужно уменьшить погрешность аппроксимации, то (при заданном отрезке $[a,b]$) добиться этого можно увеличением степени многочлена k. Однако в некоторых случаях приближение многочленами будет неэффективно. Классическим примером является функция $f(x)=\sqrt{|x|}$, которая плохо приближается многочленами на отрезке $[-1,1]$. Тем не менее, взяв достаточно

большую степень аппроксимационного многочлена ($k \geqslant 20$) и используя разложение по многочленам Чебышева для уменьшения вычислительной погрешности (см., например, [8]), и в этом случае можно получить неплохой результат. Более убедительный пример - случай, когда f " склеена" из кусков многочленов. Самый разумный способ приближения состоит в следующем: разбить отрезок $[a, b]$ на отрезки меньшей длины (в соответствии с узлами склейки) и на каждом из них приблизить f многочленом. Таким образом, иногда f удобно приближать не многочленом, а функцией, составленной из "кусков" многочленов. Если эти куски склеены достаточно гладко, то функция называется **сплайном**.

Перейдем к строгим математическим определениям и формулировкам результатов о скорости кусочно-полиномиальной аппроксимации.

Разобьем отрезок $[a, b]$ точками x_i ($i = 0, \dots, n$) на n частей: $a = x_0 < x_1 < \dots < x_n = b$. Обозначим полученное разбиение через π . Будем писать π_n в том случае, когда отрезок $[a, b]$ разбит на n равных частей.

<u>Определение</u>. $\mathcal{P}_k(\pi)$ есть пространство кусочно-полиномиальных функций степени $\leq k$, подчиненных разбиению π , то есть таких функций g , которые на каждом интервале $[x_i, x_{i+1})$ совпадают с некоторым многочленом из $\mathcal{P}_k$.

Наилучшее приближение f с помощью элементов пространства $\mathcal{P}_k(\pi)$ определяется равенством

$$e_k(f; \pi) := \inf_{g \in \mathcal{P}_k(\pi)} \| f - g \|_{C(a,b)} .$$

<u>Замечание</u>. Функция $f - g$ является кусочно-непрерывной, поэтому для вычисления ее нормы в пространстве $C(a,b)$ нужно выделить отрезки $[x_i, x_{i+1})$ и найти наибольшую из величин $\| f - g \|_{C(x_i, x_{i+1})}$ *). Иначе говоря,

$$e_k(f; \pi) = \inf_{g \in \mathcal{P}_k(\pi)} \max_i \| f - g \|_{C(x_i, x_{i+1})} .$$

---

*) Всюду ниже для $f : \bar{A} \to \mathbb{R}$ будем считать, что

$$\| f \|_{C(A)} := \| f \|_{C(\bar{A})} .$$

Теорема 1. Для любой непрерывной на $[a,b]$ функции f и любого разбиения π имеет место неравенство

$$e_{k-1}(f;\pi) \le c\,\omega_k(f;|\pi|);$$

здесь через $|\pi|$ обозначена длина наибольшего интервала в разбиении π.

Доказательство. Из монотонности модуля непрерывности сразу следует, что

$$\omega_k(f;|\pi|) = \max_i \omega_k(f;|I_i|), \qquad (1)$$

где через $I_i = [x_i, x_{i+1}]$ обозначены отрезки, из которых состоит разбиение π; $|I_i| = x_{i+1} - x_i$. Кроме того,

$$e_{k-1}(f;\pi) = \max_i e_{k-1}(f)_{C(I_i)}. \qquad (2)$$

Действительно, для любого многочлена g_i степени не более $k-1$ выполняется неравенство

$$e_{k-1}(f)_{C(I_i)} \le \|f - g_i\|_{C(I_i)}.$$

Взяв максимум по всем отрезкам I_i и учитывая замечание, получим

$$\max_i e_{k-1}(f)_{C(I_i)} \le \|f - g\|_{C(a,b)}$$

для кусочно-полиномиальной функции g из $\mathcal{P}_{k-1}(\pi)$. Так как последнее неравенство выполняется для любой функции $g \in \mathcal{P}_{k-1}(\pi)$, то

$$\max_i e_{k-1}(f)_{C(I_i)} \le e_{k-1}(f;\pi).$$

Противоположное неравенство следует из определения $e_{k-1}(f;\pi)$, так как

$$e_{k-1}(f;\pi) \le \|f - g\|_{C(a,b)}$$

для любой $g \in \mathcal{P}_{k-1}(\pi)$. Если g состоит из многочленов наилучшего приближения, то есть

$$\| f - g \|_{C(I_i)} = e_{k-1}(f)_{C(I_i)},$$

то

$$e_{k-1}(f;\pi) \leq \max_i e_{k-1}(f)_{C(I_i)}.$$

Таким образом, равенство (2) доказано.

Воспользуемся теперь теоремой Уитни (см. §4), согласно которой

$$e_{k-1}(f)_{C(I_i)} \leq c\, \omega_k(f; |I_i|)_{C(I_i)}.$$

Очевидно, что

$$\omega_k(f; |I_i|)_{C(I_i)} \leq \omega_k(f; |I_i|)_{C(a,b)}.$$

Следовательно,

$$e_{k-1}(f)_{C(I_i)} \leq c\, \omega_k(f; |I_i|)_{C(a,b)}.$$

Взяв максимум по всем отрезкам и учитывая равенства (1) и (2), получим утверждение теоремы 1. Теорема доказана.

В отличие от приближения многочленами, обратное неравенство нетривиально. Ю.А.Брудным в [25] доказано утверждение об "атомном" разложении модуля непрерывности (см. также библиографические замечания).

<u>Теорема 2</u>. Для любой $f \in C(a,b)$ имеет место соотношение

$$\omega_k(f;t) \asymp \sup_{\pi} \sup_{I_i \in \pi} e_{k-1}(f)_{C(I_i)}; \qquad (3)$$

здесь верхняя грань взята по всем разбиениям π, состоящим из отрезков I_i, длина которых не превосходит t.

Формула (3) имеет чисто теоретическое значение, однако ее можно уточнить, заменив произвольное разбиение π разбиениями специального вида. Для формулировки соответствующего результата обозначим через π'_n сдвинутые на $\frac{1}{2n}$ отрезки разбиения π_n, то есть

$I' \in \pi_n'$, если

$$I' = [(i+\tfrac{1}{2})n^{-1}, (i+\tfrac{3}{2})n^{-1}].$$

<u>Следствие 1</u>. Для любой $f \in C(a,b)$ имеет место соотношение

$$\omega_k(f; \tfrac{1}{n}) \asymp \sup_{I \in \pi_n} e_{k-1}(f)_{C(I)} + \sup_{I' \in \pi_n'} e_{k-1}(f)_{C(I')}. \qquad (4)$$

При доказательстве этого утверждения используются некоторые факты теории локальных приближений (см. задачу 2 этого параграфа).

Формула (4) может быть использована для приближенного вычисления модуля непрерывности. Для этого нужно на каждом отрезке $I \in \pi_n$ и $I' \in \pi_n'$ построить интерполяционный многочлен степени $\leq k-1$. Подробнее об интерполяции многочленами см. §6.

Перейдем от приближения кусочно-полиномиальными функциями к приближению сплайнами, то есть функциями, которые состоят из "гладко склеенных кусков" многочленов.

Введем пространство $S_k^\ell(\pi)$. Ему принадлежат функции, которые имеют производные до порядка ℓ включительно и, кроме того, являются элементами пространства $\mathcal{P}_k(\pi)$, то есть

$$S_k^\ell(\pi) := \mathcal{P}_k(\pi) \cap C^\ell(a,b);$$

здесь и далее π - разбиение отрезка $[a,b]$. Положим $S_k^{-1}(\pi) = \mathcal{P}_k(\pi)$. Заметим, что $S_k^k(\pi) = \mathcal{P}_k$. В дальнейшем мы будем работать с пространством $S_k^\ell(\pi)$ при $\ell = k-1$ (так называемое пространство сплайнов дефекта 1), поэтому для удобства обозначим $S_k(\pi) := S_k^{k-1}(\pi)$. По аналогии с величиной $e_k(f;\pi)$ введем величину $s_k(f;\pi)$ равенством

$$s_k(f;\pi) := \inf_{s \in S_k(\pi)} \|f - s\|_{C(a,b)}.$$

Так как $S_k(\pi)$ является подпространством пространства $\mathcal{P}_k(\pi)$, то

$$e_k(f;\pi) \leq s_k(f;\pi).$$

Докажем обобщение теоремы Уитни для сплайнов.

<u>Теорема 3</u>. Для любой непрерывной функции f и любого разбиения π верно неравенство

$$S_{k-1}(f;\tilde{\pi}) \le c\,\omega_k(f;|\pi|)$$

Здесь через $\tilde{\pi}$ обозначено разбиение, которое получается из π делением каждого отрезка на $k-1$ равных частей.

Доказательство. Пусть $I_i=[x_i,x_{i+1}]$ и заданы два многочлена p_i и p_{i+1}. Разобьем отрезок I_i на $k-1$ равных частей. Построим сплайн S степени $k-1$ дефекта 1, подчиненный этому равномерному разбиению и удовлетворяющий граничным условиям

$$S^{(j)}(x_i)=p_i^{(j)}(x_i),\ j=0,\dots,k-2;$$
$$S^{(j)}(x_{i+1})=p_{i+1}^{(j)}(x_{i+1}),\ j=0,\dots,k-2.$$

Существует единственный сплайн, удовлетворяющий этим условиям (см. задачу 4). Кроме того, для некоторой постоянной $C=C(k)$ выполнено неравенство

$$\|S\|_{C(I_i)} \le c\left(\|p_i\|_{C(I_i)}+\|p_{i+1}\|_{C(I_i)}\right). \qquad (5)$$

Доказательство неравенства (5) в силу соображений подобия достаточно провести для отрезка $I=[0,1]$. В этом случае из формулы Крамера следует, что

$$\sup_{x\in[0,1]}|S(x)| \le c(k)\sum_{j=0}^{k-2}\left(|p_i^{(j)}(0)|+|p_{i+1}^{(j)}(1)|\right).$$

Чтобы продолжить оценку, используем следующее **неравенство Маркова** для многочленов степени $\le k-1$:

$$\|p^{(j)}\|_{C(0,1)} \le c(k)\,\|p\|_{C(0,1)}.$$

В силу этого неравенство (5) выполняется для отрезка [0,1], а значит, и для произвольного отрезка I .

Сконструируем теперь сплайн, который удовлетворяет условию теоремы 3. Для этого на каждом отрезке $[x_i,x_{i+2}]$ построим многочлен наилучшего приближения $p_{i+1}\in\mathcal{P}_{k-1}$ для функции f , то есть такой, что

$$\| f - p_{i+1} \|_{C(x_i, x_{i+2})} = e_{k-1}(f)_{C(x_i, x_{i+2})}.$$

Таким образом каждой точке x_{i+1} соответствует свой многочлен p_{i+1}. Исключение составляют точки a и b, им будут соответствовать те же многочлены, что и соседним с ними точкам. Пусть $I_i \in \pi$. Разобьем I_i на $k-1$ равную часть и по многочленам p_i и p_{i+1} построим сплайн s так, как это было проделано выше. Заметим, что на первом и последнем отрезке сплайн совпадает с многочленом. Получим сплайн, подчиненный разбиению $\widetilde{\pi}$. Остается доказать, что он удовлетворяет условию теоремы. Оценим расстояние от f до s на отрезке I_i. По неравенству треугольника

$$\| f - s \|_{C(I_i)} \leq \| f - p_i \|_{C(I_i)} + \| p_i - s \|_{C(I_i)}.$$

Применим к сплайну $s - p_i$ неравенство (5), в силу которого

$$\| s - p_i \|_{C(I_i)} \leq c \, \| p_i - p_{i+1} \|_{C(I_i)}.$$

Заметим, что по определению многочленов p_i и p_{i+1}

$$\| f - p_i \|_{C(I_i)} \leq e_{k-1}(f)_{C(x_{i-1}, x_{i+1})},$$

$$\| f - p_{i+1} \|_{C(I_i)} \leq e_{k-1}(f)_{C(x_i, x_{i+2})}.$$

Объединяя все эти оценки, имеем :

$$\| f - s \|_{C(I_i)} \leq c \, e_{k-1}(f)_{C(x_{i-1}, x_{i+1})} + e_{k-1}(f)_{C(x_i, x_{i+2})}.$$

Если I_i - первый или последний отрезок, то

$$\| f - s \|_{C(I_i)} = \| f - p_i \|_{C(I_i)}.$$

Итак,

$$\| f - s \|_{C(a,b)} \leq c \sum_i \left[e_{k-1}(f)_{C(x_{i-1}, x_{i+1})} + e_{k-1}(f)_{C(x_i, x_{i+2})} \right].$$

Отрезки $[x_{i-1}, x_{i+1}]$ образуют разбиение отрезка $[a,b]$, поэтому по теореме 1 для кусочно-полиномиальных функций

$$\sum_i e_{k-1}(f)_{C(x_{i-1}, x_{i+1})} \le c\, \omega_k(f; |\pi|).$$

Аналогично для отрезков $[x_i, x_{i+2}]$

$$\sum_i e_{k-1}(f)_{C(x_i, x_{i+2})} \le c\, \omega_k(f; |\pi|).$$

Таким образом,

$$\| f - s \|_{C(a,b)} \le c\, \omega_k(f; |\pi|).$$

Остается заметить, что $s \in S_{k-1}(\tilde{\pi})$, и теорема 3 полностью доказана.

<u>Следствие 2</u>. Для любой функции $f \in C(a,b)$ верно неравенство

$$s_{k-1}(f; \pi_n) \le c\, \omega_k(f; \tfrac{1}{n}). \qquad (6)$$

Для <u>доказательства</u> этого утверждения достаточно заметить, что $\tilde{\pi} = \pi_{n(k-1)}$, и воспользоваться свойствами модуля непрерывности.

В заключение этого параграфа еще раз докажем неравенство Марше (см. §3), используя обобщенное неравенство Уитни для сплайнов. Это доказательство почти без изменений переносится на случай пространств L_p, $p>0$, и многомерную ситуацию.

<u>Утверждение</u>. Для любой $f \in C(0,1)$ имеет место неравенство $(j \le k)$

$$\omega_j(f,t) \le c(k)\, t^j \left(\|f\|_{C(0,1)} + \int_t^{\frac{1}{2}} \frac{\omega_k(f;s)}{s^{j+1}}\, ds \right).$$

<u>Доказательство</u>. Согласно (6) для любой функции $f \in C(0,1)$ существует последовательность сплайнов $g_n \in S_k(\pi_{2^n})$ такая, что

$$\| f - g_n \|_{C(0,1)} \le c\, \omega_{k+1}(f; 2^{-n}). \qquad (7)$$

Обозначим правую часть неравенства (7) через ω_n. Представим f в виде ряда

$$f = (f - g_M) + g_0 + \sum_{i=1}^{M} (g_i - g_{i-1}).$$

Число M будет выбрано позднее. Учитывая свойство 10 модуля непрерывности, имеем:

$$\omega_j(f;t) \le \omega_j(f-g_M;t) + \omega_j(g_0;t) + \sum_{i=1}^{M} \omega_j(g_i - g_{i-1};t). \quad (8)$$

Оценим каждое из трех слагаемых в правой части (8). Для оценки первого слагаемого используем свойство 12 §2:

$$\omega_j(f-g_M;t) \le 2^j \|f-g_M\|_{C(0,1)} \le c(k)\,\omega_M. \quad (9)$$

Для того, чтобы оценить следующие слагаемые в (8), нам потребуется

<u>Лемма</u>. Если $g \in S_k(\pi_n)$, то

$$\omega_j(g;t) \le c(k)\,(tn)^j \|g\|_{C(0,1)}.$$

Для доказательства см. задачу 9.

Если применить лемму к многочлену g_0, то получится неравенство

$$\omega_j(g_0;t) \le c(k)\,t^j \|g_0\|_{C(0,1)}.$$

Продолжим оценку, воспользовавшись неравенством

$$\|g_0\|_{C(0,1)} \le \|f-g_0\|_{C(0,1)} + \|f\|_{C(0,1)}.$$

Кроме того, как следует из (7),

$$\|f-g_0\|_{C(0,1)} \le c(k)\,\|f\|_{C(0,1)}.$$

Итак,

$$\omega_j(g_0;t) \le c(k)\,t^j \|f\|_{C(0,1)}. \quad (10)$$

Оценим третье слагаемое в (7). Прежде всего заметим, что пространство $S_k(\pi_{2^{i-1}})$ вложено в пространство $S_k(\pi_{2^i})$, так как разбиение π_{2^i} является измельчением разбиения $\pi_{2^{i-1}}$. Поэтому

сплайн $g_i - g_{i-1}$ принадлежит пространству $S_k(\pi_{2^i})$ и к нему можно применить утверждение леммы. Имеем:

$$\omega_j(g_i - g_{i-1}; t) \le c(k)(t \cdot 2^i)^j \, \|g_i - g_{i-1}\|_{C(0,1)}.$$

Далее из (7) и монотонности модуля непрерывности следует неравенство

$$\|g_i - g_{i-1}\|_{C(0,1)} \le c \cdot (\omega_i + \omega_{i-1}) \le c(k)\,\omega_i.$$

Итак, для $i = 1, \dots, M$ получаем оценку

$$\omega_j(g_i - g_{i-1}; t) \le c(k)(t \cdot 2^i)^j \, \omega_i. \tag{11}$$

Если выбрать M из условия $2^{-M} \le t \le 2^{-M+1}$, то из (8), (9), (10), (11) получим соотношение

$$\omega_j(f; t) \le c(k)\, t^j \left(\|f\|_{C(0,1)} + \sum_{i=1}^{M} 2^{ij} \omega_i \right). \tag{12}$$

Остается перейти от суммирования в правой части (12) к интегрированию. Для этого используем неравенство (см. задачу 1 §3)

$$2^{ij} \omega_i \le c(k) \int_{2^{-i}}^{2^{-i+1}} \frac{\omega_{k+1}(f; s)}{s^{j+1}} \, ds$$

Просуммировав по i, получим:

$$\omega_j(f; t) \le c(k)\, t^j \left(\|f\|_{C(0,1)} + \int_{2^{-M+1}}^{2^{-1}} \frac{\omega_{k+1}(f; s)}{s^j} \, ds \right).$$

Оценку можно продолжить, увеличив отрезок интегрирования, так как в силу выбора M выполняется неравенство $2^{-M+1} \geqslant t$. Неравенство Маршо доказано.

Задачи и упражнения к §5

1. Доказать, что множество $\mathcal{P}_k(\pi)$ является линейным пространст-

вом, найти его базис и размерность.

2. Доказать соотношение (4). Использовать тот факт, что

$$e_k(f)_{C(I_1 \cup I_2)} \le c(r)\left(e_k(f)_{C(I_1)} + e_k(f)_{C(I_2)}\right),$$

если

$$\frac{|I_1 \cap I_2|}{|I_i|} \geqslant r \quad ; \; i = 1, 2 . \tag{13}$$

3. Найти базис и размерность пространства $S_k^{\ell}(\pi)$.

4. Пусть сплайн $s \in S_{k-1}(\pi_{k-1})$ и выполняются следующие граничные условия:

$$s^{(j)}(a) = s^{(j)}(b) = 0 \;, \; j = 0, 1, \ldots, k-2 .$$

Доказать, что $s \equiv 0$.

5. Доказать, что если неравенство (4) выполняется для отрезка $[0,1]$, то оно выполняется для любого отрезка I.

6. Пользуясь процедурой склейки, описанной в теореме 3, построить параболический сплайн s_2, для функции f на отрезке $[0,1]$.

(a) $f(x) = \sqrt{x}$;
(b) $f(x) = e^x$;
(c) $f(x) = \sin 10x$.

7. Оценить константу c в неравенстве (5).

8. Написать программу вычисления модуля непрерывности на ЭВМ по формуле (4). Сравнить с точными значениями.

9. Доказать, что если $g \in S_k(\pi_n)$, то для всех $j \le k-1$ имеет место неравенство

$$\omega_j(g; t) \le c(k)(tn)^j \, \|g\|_{C(0,1)} .$$

<u>Указание</u>. Воспользовавшись девятым свойством модуля непрерывности, перейти от функции g к функции $g^{(j)}$. Затем использовать неравенство Маркова

$$\|g^{(j)}\|_{C(x_i, x_{i+1})} \le \left(\frac{2k}{x_{i+1} - x_i}\right)^j \|g\|_{C(x_i, x_{i+1})} .$$

§6. Одномерные интерполяционные проекторы и их оценки

Интерполяция представляет собой один из простых способов "восстановления" неизвестной функции f на некотором множестве E по набору известных значений в отдельных точках этого множества. Более точно, неизвестная функция f приближается на E с помощью интерполянта p из некоторого класса X ; p определяется условиями $p(x_j) = f(x_j)$ в узлах $x_j \in E$. Если в некоторых x_j ставятся дополнительные условия на равенство производных, то такие узлы называются кратными. В качестве X обычно выбирают совокупность алгебраических или тригонометрических многочленов, рациональных функций или сплайнов (простейшие варианты - кусочно-постоянные функции или интерполирующие ломаные). Соответствующие вопросы подробно описаны в литературе, в том числе учебной (см., например, [4] - [8]). В этом параграфе мы остановимся на некоторых оценках для интерполяционных алгебраических многочленов и смежных вопросах, как правило, не изложенных в учебниках.

Единственное решение интерполяционной задачи

$$p(x_0) = f(x_0) = f_0 \ , \ \dots \ , \ p(x_n) = f(x_n) = f_n \quad ; \ p \in \mathcal{P}_n$$

для попарно различных узлов x_i дает известная формула Лагранжа (J. L. Lagrange, 1795):

$$p(x) = \sum_{i=0}^{n} f_i \, l_i(x) \ ; \quad l_i(x) := \prod_{j \neq i} \frac{x - x_j}{x_i - x_j} \ . \qquad (1)$$

Многочлены $l_i(x)$ образуют специальный (зависящий от узлов) базис $\mathcal{P}_n$; они выбраны так, что

$$l_i(x_k) = \delta_{ik} := \begin{cases} 1 \ , & i = k \\ 0 \ , & i \neq k \end{cases}$$

Будем считать, что $x_0, \dots, x_n \in [a,b]$ и $f \in C(a,b)$. В этом случае формула (1) задает интерполяционный проектор $P : C(a,b) \to \mathcal{P}_n$, если положить $Pf(x) = p(x)$. Пусть $\| P \|$ - норма P как оператора из $C(a,b)$ в $C(a,b)$.

<u>Теорема 1</u>. $\| P \| = \max\limits_{a \le x \le b} \sum\limits_{i} | l_i(x) | \ .$

Доказательство. Положим $\lambda(x) := \sum |\ell_i(x)|$. Пусть $x^* \in [a,b]$ такова, что $\lambda(x^*) = \|\lambda\|_C$. Рассмотрим любую непрерывную функцию φ такую, что $\varphi(x_i) = \operatorname{sign} \ell_i(x^*)$, $\|\varphi\|_C = 1$. Для нее

$$\|P\varphi\|_C = \max_{a \le x \le b} |P\varphi(x)| \geqslant P\varphi(x^*) = \sum \varphi(x_i)\ell_i(x^*) =$$
$$= \sum |\ell_i(x^*)| = \|\lambda\|_C \|\varphi\|_C .$$

Это означает, что $\|P\| \geqslant \|\lambda\|_C$. В то же время для любой $f \in C(a,b)$

$$\|Pf\|_C \le \|f\|_C \max_{a \le x \le b} \sum |\ell_i(x)| = \|\lambda\|_C \|f\|_C ,$$

что гарантирует оценку $\|P\| \le \|\lambda\|_C$. В связи с этим $\|P\| = \|\lambda\|_C$, и теорема доказана.

Ясно, что $\|P\|$ зависит от расположения узлов x_i. Так как $\|P\|$ не меняется при сдвиге и растяжении отрезка, можно считать, что $[a,b] = [0,1]$ или $[-1,1]$. В соответствии со сказанным в §4 (см. там неравенство (11)) возникает задача минимизации $\|P\|$ за счет выбора узлов. Проектор P, соответствующий минимальному значению $\|P\|$ при данном n, называется минимальным. Отметим, что вопрос об оптимальном выборе узлов полностью не решен.

Наиболее естественным является равномерный выбор узлов интерполяции : $x_j = a + jh$, $j = 0,\dots,n$; $h := (b-a)/n$. При небольших n такой выбор вполне допустим; отметим, что равномерные узлы порождают минимальные проекторы для $n = 1, 2$ (их нормы равны соответственно 1 и 1.25). Однако с увеличением n многочлен $Pf = P_n f$ может оказаться полностью непригодным для приближения f по той причине, что $\|P_n f\|$ неограниченно возрастает. Именно, для равномерно распределенных узлов интерполяции

$$\|P_n\| \geqslant const \cdot e^{n/2}$$

(см., например, [8]). Асимптотика $\|P_n\|$ может быть улучшена при использовании в качестве узлов интерполяции нулей многочлена Чебышева степени $n+1$, ассоциированного с отрезком $[a,b]$, т.е. точек вида

$$x_i = \frac{b+a}{2} + \frac{b-a}{2} \cos\left(\frac{\pi(2i+1)}{2(n+1)}\right); \; i = 0,\dots,n . \qquad (2)$$

Точки (2) получаются построением на отрезке $[a,b]$ полуокружности, делением ее на $n+1$ равных дуг и проектированием середин каждой из дуг на $[a,b]$. В этом случае, как показано в [14],

$$\| P_n \| = \frac{1}{n+1} \sum_{k=1}^{n+1} tg \frac{(2k-1)\pi}{4(n+1)} . \qquad (3)$$

Формула (3) позволяет вычислить норму P_n для данного n; некоторые вычисления даны в следующей таблице.

n	0	1	2	3	4	...	18	19	20
$\| P_n \|$	1	1.414	1.667	1.848	1.989	...	2.837	2.870	2.901

Таким образом, при использовании чебышевских узлов в соответствии с неравенством Лебега (см. §4)

$$\| f - P_n f \| \le (1 + \| P_n \|)\, e(f;\mathcal{P}_n) \le 3\, e(f;\mathcal{P}_n) , \quad n \le 4 ;$$

$$\| f - P_n f \| \le 4\, e(f;\mathcal{P}_n) , \quad n \le 20 .$$

С помощью формулы (3) получается асимптотическая оценка

$$\| P_n \| \approx \frac{2}{\pi} \ln n + const , \qquad (4)$$

которая является наилучшей по порядку при $n \to \infty$ (запись $A \approx B$ при $n \to \infty$ означает, что $A/B \to 1$).

Поясним последнее формулировкой точного результата. Для этого введем в рассмотрение понятие интерполяционного массива $\mathcal{X}$:

$$\mathcal{X} : \quad \begin{array}{l} x_0^{(1)} \\ x_0^{(2)}, x_1^{(2)} \\ x_0^{(3)}, x_1^{(3)}, x_2^{(3)} \\ \dots\dots\dots \\ x_0^{(n)}, x_1^{(n)}, \dots, x_{n-1}^{(n)} \\ \dots\dots\dots \\ \dots\dots\dots \\ \dots\dots\dots \end{array}$$

n -я строка бесконечной вниз треугольной матрицы $\mathcal{X}$ содержит n различных чисел $x_0^{(n)}, x_1^{(n)}, \ldots, x_{n-1}^{(n)} \in [a,b]$; по ним строится интерполяционный проектор $P_n : C(a,b) \to \mathcal{P}_{n-1}$.Последовательность многочленов $p_n = P_n f$ для фиксированной f задает, как говорят, интерполяционный процесс. Оказывается (см., например, [19]), что для любого интерполяционного массива $\mathcal{X}$ с независящей от $\mathcal{X}$ константой c выполняется

$$\| P_n \| > \frac{2}{\pi} \ln n + c,$$

так что порядок асимптотики (4) уменьшить нельзя.

Для $f \in C^{n+1}(-1,1)$ точность интерполяции с помощью $\mathcal{P}_n$ дается известной оценкой

$$\| f - Pf \|_{C(-1,1)} \leq \frac{\| f^{(n+1)} \|_{C(-1,1)}}{(n+1)!} \| w \|_{C(-1,1)} , \qquad (5)$$

$$w(x) := (x - x_0)(x - x_1) \ldots (x - x_n) .$$

Норма $\| w \|_C$ является минимальной, если в качестве узлов x_i выбрать корни многочлена Чебышева $T_{n+1}(x) = \cos[(n+1)\arccos x]$ (это гарантируется экстремальным свойством многочленов Чебышева - наименьшим уклонением от нуля, см. [2]). Старший коэффициент T_{n+1} равен 2^n , поэтому для чебышевских узлов

$$\| w \|_{C(-1,1)} = \frac{1}{2^n} \| T_{n+1} \|_{C(-1,1)} = \frac{1}{2^n} .$$

Таким образом, для узлов Чебышева $x_0, \ldots, x_n$

$$\| f - Pf \|_{C(-1,1)} \leq \frac{1}{2^n (n+1)!} \| f^{(n+1)} \|_{C(-1,1)} \qquad (6)$$

- лучшая оценка, получающаяся из (5).

Коротко остановимся на вопросах о принципиальной ограниченности полиномиальной интерполяции при возрастании n . Если $\{p_n^*\}$ - последовательность многочленов наилучшего приближения для $f \in C(a,b)$,

$p_n^* \in \mathcal{P}_n$, то по теореме Вейерштрасса

$$\| f - p_n^* \|_C \to 0 \quad , n \to \infty , \tag{7}$$

что дает, естественно, $p_n^*(x) \to f(x)$ во всех точках $x \in [a,b]$. Если узлы интерполяции адаптировать к f и выбирать как точки совпадения f и p_n^* , то будет, естественно, $P_n f = p_n^*$ и $\|f - P_n f\|_C \to 0$. Удобно, конечно, использовать один и тот же набор узлов для всех f . В этом случае ни равномерной, ни поточечной сходимости $P_n f$ к f гарантировать нельзя. Приведем без доказательства следующее утверждение, известное как **теорема Фабера** (G.Faber, 1914).

Теорема 2. Для любого интерполяционного массива $\mathcal{X}$ найдется функция $f \in C(a,b)$ такая, что $P_n f$ не сходится равномерно к f .

Существуют непрерывные функции, для которых интерполяционный процесс с равномерными узлами не сходится ни в одной точке! Опасности, связанные с полиномиальной интерполяцией, впервые заметил К.Рунге (C.Runge, 1901). При интерполировании функции $f(x) = (1+25x^2)^{-1}$ полиномом p_n с равномерными узлами выяснилось, что $\{p_n(x)\}$ расходится при всех x таких, что $0.726... \le |x| \le 1$ ($[a,b] = [-1,1]$).

При использовании узлов Чебышева равномерная (значит, и поточечная) сходимость $p_n \to f$ имеет место для всех функций f из класса Дини - Липшица (см. ниже упр.16). Таким образом, интерполяция по чебышевским узлам гарантирует лучшую оценку точности для гладких функций (6), наилучшую асимптотическую оценку нормы проектора (4) и поточечную (даже равномерную) сходимость последовательности интерполянтов к f для широкого класса непрерывных функций f (для $f \in DL$). Следует помнить, однако, что при небольших n норма проекторов по равномерным узлам несколько меньше.

Несмотря на то, что интерполяционная форма Лагранжа, основанная на использовании точек Чебышева, является почти наилучшей, она обладает и рядом недостатков. Ниже мы рассмотрим интерполяционные многочлены Ньютона. Как считает К.Де Бор [8], эта форма интерполяции "представляет собой наилучший компромисс между простотой структуры и удобством вычислений. Кроме того, она позволяет легко оценить погрешность интерполяции и реализовать кратную интерполяцию без привлечения дополнительных средств".

Итак, пусть, как и ранее, функция f приближается на E с

помощью многочлена $p \in \mathcal{P}_n$; p находится из условий $p(x_j)=f(x_j)$; $x_j \in E$, $j=0,\dots, n$. Если раньше в качестве базиса пространства $\mathcal{P}_n$ мы выбирали многочлены Лагранжа, то теперь выберем многочлены p_i , которые задаются формулами

$$p_0(x)=1 \ ; \ p_i(x)=(x-x_0)(x-x_1)\dots(x-x_{i-1}) \ , \ i=1,\dots,n \ .$$

Тогда решение интерполяционной задачи дает формула Ньютона (I. Newton, опубликовано в 1736 г.):

$$p(x) = \sum_{i=0}^{n} f(x_0,\dots,x_i)\, p_i(x) \ ;$$

здесь $f(x_0,\dots,x_i)$ - i -я разделенная разность, которую удобнее всего (учитывая вычисления с помощью ЭВМ) считать по формулам:

$$f(x_0,\dots,x_i) = \frac{f^{(i)}(x_0)}{i!} \ , \quad x_0=x_1=\dots=x_i \ , \ f \in C^i \ ;$$

$$f(x_0,\dots,x_i) = \frac{f(x_0,\dots,x_{t-1},x_{t+1},\dots,x_i)-f(x_0,\dots,x_{s-1},x_{s+1},\dots,x_i)}{x_s - x_t}$$

для любых двух различных точек x_t, x_s последовательности $x_0,\dots,x_i$.

Заметим, что если все точки последовательности $\{x_j\}$ совпадают, то интерполяционный многочлен имеет вид

$$p(x) = \sum_{i=0}^{n} \frac{f^{(i)}(x_0)}{i!}(x-x_0)^{i-1} \ .$$

Таким образом, интерполяционный многочлен - это сумма $n+1$ членов разложения функции f в ряд Тейлора в т. x_0 .

Для оценки погрешности вычислений можно воспользоваться следующим результатом.

<u>Утверждение</u>. Если $f \in C^{n+1}(a,b)$ и $x_0 \le \dots \le x_n$ - последовательность точек из отрезка $[a,b]$, то для всех $x \in [a,b]$ имеет место равенство

$$f(x)-p(x) = (x-x_0)\dots(x-x_n)\, f(x_0,\dots,x_n,x) \ .$$

Задачи и упражнения к §6

1. Показать, что функции $l_i(x)$ из формулы (1) образуют базис

пространства $\mathcal{P}_n$.

2. Найти минимальные интерполяционные проекторы на $\mathcal{P}_0$ и $\mathcal{P}_1$.

3. Пусть $P_{x_0,x_1} : C(0,1) \to \mathcal{P}_1$ - проектор, построенный по точкам $x_0, x_1 \in [0,1]$. Найти $\| P_{x_0,x_1} \|$ и обсудить результат.

4. Найти норму проектора на $\mathcal{P}_2$ по точкам равномерного разбиения. (Заметим, что соответствующий проектор является минимальным).

5. Найти приближенно $\| P \|$ с помощью ЭВМ.

6. Доказать, что $\| P \|$ не меняется при сдвиге и растяжении $[a,b]$.

7. Для функции Рунге $f(x)=(1+25x^2)^{-1}$ на отрезке $[-1,1]$ построить интерполяционные многочлены

а) по чебышевским узлам;

б) по равномерным узлам.

Сравнить максимальные погрешности.

8. Построить интерполяционный многочлен p для функции $f(x)=\ln x$ по узлам 1, 2, 3, 4. Доказать, что $p(x) > \ln x$ для $2 < x < 3$.

9. Построить интерполяционный многочлен для функции $f(x)=\cos x$ по последовательности узлов $x_1=0$, $x_2=\frac{\pi}{2}$, $x_3=\pi$, $x_4=\frac{3\pi}{2}$, $x_5=2\pi$. Оценить погрешность вычислений и сравнить ее с максимальной погрешностью в точках

10. Доказать, что многочлены Ньютона p_i образуют базис в пространстве многочленов.

11. Сравнить трудоемкость вычислений по формулам Ньютона и Лагранжа.

12. Построить интерполяционный многочлен Ньютона для функции $f(x)=\cos x$ по последовательности узлов $x_1=x_2=0$, $x_3=x_4=\pi$, $x_5=x_6=2\pi$. Сравнить результаты вычислений с задачей 9.

13. Обозначим через $p(x)$ полином степени n такой, что $p(k)=k/(k+1)$, $k=0,1,\dots,n$. Найти $p(n+1)$.

<u>Указание.</u> Рассмотрите отдельно случаи четного и нечетного n .

14. В ходе эксперимента получены следующие семь пар данных:

t	-1.000	-0.960	-0.860	-0.790	0.220	0.500	0.930
y	-1.000	-0.151	0.894	0.986	0.895	0.500	-0.306

Оценить значение $y(t)$ для $-1 \le t \le 1$, , построив интерполяционный многочлен 6 - й степени. Предполагается, что y - очень гладкая функция. Хорошо ли отражает полином характер изменения данных?

15. Используя интерполяционную формулу Лагранжа, доказать, что если все корни x_i $(i=1,\dots,n)$ многочлена p_n ($n \geqslant 2$) различны, то

$$\sum_{i=1}^{n} \frac{1}{p_n'(x_i)} = 0 .$$

16. Показать, что последовательность интерполяционных многочленов P_n по чебышевским узлам равномерно сходится к интерполируемой функции f , если $f \in DL$.

Указание. Воспользоваться неравенством Лебега, логарифмической оценкой (4) для нормы соответствующего проектора и следующим **неравенством Джексона** (D. Jackson, 1911): если $f \in C(a,b)$, то

$$e(f; \mathcal{P}_n)_{C(a,b)} \leq \gamma\, \omega_1\left(f; \frac{b-a}{n}\right)_{C(a,b)} ,$$

γ - абсолютная постоянная.

Часть 2. Некоторые методы полиномиальной и кусочно-полиномиальной аппроксимации функций двух переменных

§7. Разностные операторы и модули непрерывности функций двух переменных

Пусть f - действительная функция двух действительных переменных x и y. Чаще всего мы будем считать, что f определена в точках квадрата $Q:=[0,1]^2$. Во второй части пособия рассматриваются некоторые вопросы, связанные с приближением f с помощью многочленов и кусочно-полиномиальных функций. По сравнению с аппроксимацией функций одного переменного задачи приближения функций двух (и, вообще, многих) переменных не только усложняются, но и часто приобретают специфические черты; многие вопросы в одномерной ситуации просто не возникают. Двумерные задачи прикладного анализа часто имеют ясное естественное происхождение - они встают, например, при обработке изображений.

Прежде всего мы введем в рассмотрение разностные характеристики функций двух переменных; эти характеристики (модули непрерывности), как и в одномерной ситуации, часто используются в оценках точности приближения функций. В основе их определения лежит понятие разностного оператора Δ_h, $h=(h_1,h_2)\in\mathbb{R}^2$; мы считаем, что

$$\Delta_h f(x,y) := f(x+h_1, y+h_2) - f(x,y).$$

Пусть $k\in\mathbb{Z}_+$. Положим, как и ранее (см. §2), $\Delta_h^k := (\Delta_h)^k$.

Это означает, что

$$\Delta_h^k f(x,y) = \sum_{j=0}^{k} (-1)^{k-j} C_k^j f(x+jh_1, y+jh_2). \qquad (1)$$

Особую роль играют операторы взятия разности по координатным направлениям $\Delta^k_{\sigma e^{(i)}}$, $i=1,2$. Здесь и ниже $\sigma\in\mathbb{R}$; $e^{(1)}=(1,0)$, $e^{(2)}=(0,1)$

- канонический базис $\mathbb{R}^2$; $e:=(1,1)$. Положим для $\alpha=(\alpha_1,\alpha_2)\in\mathbb{Z}_+^2$

$$\Delta_h^\alpha := \Delta_{h_1 e^{(1)}}^{\alpha_1}\ \Delta_{h_2 e^{(2)}}^{\alpha_2}\ ,\quad \tilde{\Delta}_h^\alpha := \Delta_{h_1 e^{(1)}}^{\alpha_1} + \Delta_{h_2 e^{(2)}}^{\alpha_2}\ .$$

Оператор Δ_h^α называют оператором взятия **смешанной разности порядка** α .

В дальнейшем мы будем иметь дело с двумя линейными пространствами алгебраических многочленов от двух переменных - $\mathcal{P}_n$ (общей степени по x , y не выше n) и $\mathcal{P}_{m,n}$ (степени не выше m по x и не выше n по y). Если $\alpha=(m,n)$, то $\mathcal{P}_\alpha:=\mathcal{P}_{m,n}$. Таким образом, $\mathcal{P}_n$ состоит из функций вида

$$g(x,y)=\sum_{0\le i+j\le n} a_{ij}\,x^i y^j\ ,\quad a_{ij}\in\mathbb{R}\ ,$$

а $\mathcal{P}_{m,n}$ - из функций вида

$$g(x,y)=\sum_{\substack{0\le i\le m\\ 0\le j\le n}} a_{ij}\,x^i y^j\ ,\quad a_{ij}\in\mathbb{R}\ .$$

Так, например,

$$\mathcal{P}_1=\{ax+by+c\}\ ,\quad \mathcal{P}_{1,1}=\{axy+bx+cy+d\}$$

и т.д.

Введем в рассмотрение также классы **квазимногочленов** $\mathcal{P}_m^{(1)}$ и $\mathcal{P}_n^{(2)}$; они состоят из многочленов по одному переменному, коэффициенты которых являются (непрерывными) функциями другого переменного:

$$\mathcal{P}_m^{(1)}=\Big\{\sum_{j=0}^{m} a_j(y)x^j\Big\}\ ,\quad \mathcal{P}_n^{(2)}=\Big\{\sum_{j=0}^{n} b_j(x)y^j\Big\}\ .$$

Будем считать $\mathcal{P}_{-1}^{(i)}:=\{0\}$. Нетрудно видеть, что

$$\mathcal{P}_\alpha=\mathcal{P}_{\alpha_1}^{(1)}\cap\mathcal{P}_{\alpha_2}^{(2)}\ . \tag{2}$$

Наконец, пусть $\Pi_\alpha:=\mathcal{P}_{\alpha_1}^{(1)}+\mathcal{P}_{\alpha_2}^{(2)}$. Это означает, что $g\in\Pi_\alpha$, если $g=p_1+p_2$, $p_i\in\mathcal{P}_{\alpha_i}^{(i)}$, $i=1,2$. Отметим, что классы квазимногочле-

нов являются бесконечномерными линейными подпространствами $C(Q)$.

Приведем ряд свойств разностных операторов, введенных выше.

1. Δ_h^k, $0 \le h_1, h_2 \le \frac{1}{k}$ - линейный оператор из $C(Q)$ в $C(\Pi)$, $\Pi := [0, 1-kh_1] \times [0, 1-kh_2]$; при этом

$$\| \Delta_h^k \|_{C(Q) \to C(\Pi)} = 2^k. \tag{3}$$

Доказательство использует формулу (1).

2. Если $p \in \mathcal{P}_{k-1}, k \in \mathbb{N}$, то $\Delta_h^k p(x,y) \equiv 0$ при всех h.

3. Если $p \in \mathcal{P}_{k-1}^{(i)}$, то $\Delta_{\sigma e^{(i)}}^k p(x,y) \equiv 0$ при всех σ.

4. Если $p \in \mathcal{P}_{\alpha - e}, \alpha \in \mathbb{N}^2$, то $\tilde{\Delta}_h^\alpha p(x,y) \equiv 0$ при всех h.

5. Если $p \in \Pi_{\alpha - e}$, то $\Delta_h^\alpha p(x,y) \equiv 0$ при всех h.

Отметим (также без доказательства) некоторые разностные тождества для функций двух переменных; читателю предлагается разобрать их некоторые частные случаи (см. упр. 5-7).

Пусть $|\alpha| = \alpha_1 + \alpha_2$ - порядок мультииндекса α; $\alpha! := \alpha_1! \alpha_2!$. Тогда

$$\Delta_h^k f(x,y) = \sum_{|\alpha| = k} \frac{k!}{\alpha!} \Delta_h^\alpha f(x, y + \alpha_1 h_2). \tag{4}$$

В случае $k = 1$ тождество (4) имеет следующий простой вид:

$$\Delta_h f(x,y) = \Delta_h^{(1,0)} f(x, y + h_2) + \Delta_h^{(0,1)} f(x,y)$$

и достаточно наглядный геометрический смысл (какой?).

Представление смешанной разности через полные дает следующее тождество (J. H. Kemperman):

$$\Delta_{h^{(1)}} \dots \Delta_{h^{(k)}} f(z) = \sum_{\mathcal{D}} (-1)^{k - |\mathcal{D}|} \Delta_{h_{\mathcal{D}}}^k f\Big(z + \sum_{m \notin \mathcal{D}} h^{(m)}\Big). \tag{5}$$

Здесь $h^{(j)} \in R^2$, $j = 1, \dots, k$; $h_{\mathcal{D}} := \sum_{j \in \mathcal{D}} j^{-1} h^{(j)}$; $\mathcal{D}$ - непустое подмно-жество $\{1, \dots, k\}$; $z = (x, y)$. Тождество верно и для $h^{(j)}, z \in R^n$. Ясно, что (5) можно использовать для представления разнос

Менее явным, но часто более удобным для получения ну является представление смешанной разности, полученное бо математиками П. Биневым и К. Ивановым (1985):

$$\Delta_{h^{(1)}} \dots \Delta_{h^{(k)}} f(z) = \sum_{s=1}^{r} b_s \, \Delta^k_{h'} f(z + h'') ;$$

$$h' = u_1^{(s)} h^{(1)} + \dots + u_k^{(s)} h^{(k)} ; \; h'' = v_1^{(s)} h^{(1)} + \dots + v_k^{(s)} h^{(k)} .$$

Натуральное r , числа b_s, $u_j^{(s)}$, $v_j^{(s)}$ из (6), (7) зав от k , причем

$$0 \le v_j^{(s)} , \; v_j^{(s)} + k u_j^{(s)} \le 1 , \; j = 1, \dots, k .$$

Существование этих чисел гарантируется. Последнее услови ется наиболее существенным для приложений (см., например,

С помощью введенных разностных операторов можно опре дули непрерывности функций f двух переменных. Будем сч $f \in C(Q)$, $\| f \|_{C(Q)} := \max_{(x,y) \in Q} | f(x, y) |$.

Введем в рассмотрение следующие модули непрерывности

- **полный** порядка $k \in \mathbb{N}$:

$$\omega_k(f; \tau) := \sup_{0 \le h_1, h_2 \le \tau} \| \Delta^k_h f \| , \quad \tau \ge 0$$

- **частные** по x и по y порядка k :

$$\omega_k^{(i)}(f; \tau) := \sup_{0 \le \sigma \le \tau} \| \Delta^k_{\sigma e^{(i)}} f \| , \quad i = 1, 2$$

- сумма частных модулей:

$$\Omega_\alpha(f; t) := \omega^{(1)}_{\alpha_1}(f; t_1) + \omega^{(2)}_{\alpha_2}(f; t_2) , \; t = ($$

- **смешанный** порядка $\alpha \in \mathbb{Z}_+^2$:

$$\omega_\alpha(f; t) := \sup_{0 \le h_i \le t_i} \| \Delta^\alpha_h f \| , \quad t = (t_1, t_2)$$

В правых частях этих равенств норма ассоциирована с областью определения функции, стоящей под знаком разностного оператора; например,

$$\| \Delta_h^k f \| = \| \Delta_h^k f \|_{C(\Pi)} \ , \quad \Pi = [0, 1-kh_1] \times [0, 1-kh_2].$$

Эти разностные характеристики имеют ряд свойств, которые аналогичны свойствам одномерных модулей непрерывности (см. §1-3). Приведем для примера следующий аналог неравенства Маршо ($n \le k$):

$$\omega_n(f;\tau) \le c\tau^n \left(\int_\tau^1 \frac{\omega_k^{(1)}(f;s) + \omega_k^{(2)}(f;s)}{s^{n+1}} ds + \| f \|_{C(Q)} \right).$$

Новой тематикой (по сравнению с одномерной ситуацией) является связь между модулями непрерывности различного вида. Мы приведем два соотношения такого типа.

<u>Теорема 1</u>. Для $f \in C(Q)$, $k \in \mathbb{Z}_+$, $t = (\tau, \tau)$

$$\max_{|\alpha| = k} \omega_\alpha(f;t) \asymp \omega_k(f;\tau), \tag{9}$$

$$\omega_k(f;\tau) \le c\,\Omega_\alpha(f;t), \ k \ge |\alpha| - 1. \tag{10}$$

Доказательство эквивалентности (9) использует тождество (4) и предоставляется читателю. Оценка (10) может быть получена более сложным путем с использованием теорем о скорости кусочно-полиномиальной аппроксимации.

Модули непрерывности функции двух (и многих) переменных используются в оценках для скорости приближения функций. Аналогом теоремы Уитни (см. §4) для функций $f(x,y)$ является следующее утверждение.

<u>Теорема 2</u>. Пусть $k \in \mathbb{N}$, $\alpha \in \mathbb{N}^2$, $\frac{1}{\alpha} := \left(\frac{1}{\alpha_1}, \frac{1}{\alpha_2}\right)$. Для $f \in C(Q)$ имеют место неравенства:

$$e(f; \mathcal{P}_{k-1})_{C(Q)} \le \gamma_1(k)\, \omega_k(f; \tfrac{1}{k});$$

$$e(f; \mathcal{P}_{\alpha-e})_{C(Q)} \le \gamma_2(\alpha)\, \Omega_\alpha(f; \tfrac{1}{\alpha});$$

$$e(f;\Pi_{\alpha-e})_{C(Q)} \le \gamma_3(\alpha)\,\omega_\alpha(f;\tfrac{1}{\alpha}).$$

Первые два соотношения можно установить с помощью подходящих критериев предкомпактности подмножеств $C(Q)$ (см. §4 для функций одного переменного). Третье неравенство (как частный случай более общего результата) получено в 1970 году Ю.А.Брудным.

Все определения и основные результаты этого параграфа переносятся на функции $f(x_1,\dots,x_n)$ многих переменных и некоторые функциональные пространства, отличные от C.

Задачи и упражнения к §7.

1. Получить явную формулу для смешанной разности $\Delta_h^\alpha f(x,y)$, аналогичную (1). Рассмотреть частные случаи $\alpha=(1,1)$, $\alpha=(2,1)$.

2. Обосновать представление $\mathcal{P}_\alpha = \mathcal{P}_{\alpha_1}^{(1)} \cap \mathcal{P}_{\alpha_2}^{(2)}$.

3. Доказать, что $\|\Delta_h^k\| = 2^k$. Чему равна норма оператора Δ_h^α?

4. Доказать свойства 2-5 разностных операторов.

5. Доказать тождество $\Delta_h^k f(x,y) = \sum\limits_{|\alpha|=k} \frac{k!}{\alpha!}\Delta_h^\alpha f(x,y+\alpha_1 h_2)$ для $k=1$ и $k=2$.

6. Выписать и доказать тождество Кемпермана (5) для $k=2$ и затем для разностей $\Delta^\alpha{}_h f$, $|\alpha|=2$.

7. Получить вариант представления (6) с ограничением (8) для $k=2$.

8. Вычислить модули непрерывности $\omega_1(f;\tau)$, $\omega_1^{(1)}(f;\tau)$, $\omega_\alpha(f;t)$ с $\alpha=(1,1)$ для функций

а) $f(x,y)=xy$;

в) $f(x,y)=x^i y^j$, $i,j\in\mathbb{Z}_+$;

с) $f(x,y)=\sqrt{xy}$

Во всех случаях $0\le x,y\le 1$.

9. Установить эквивалентность (9) теоремы 1. Использовать для этого тождества (4) и (6).

10. Доказать для функций $f\in C(Q)$ неравенство

$$\omega_\alpha(f;t) \le c(\alpha)\int_Q \|\Delta_{t*h}^\alpha f\|\,dh,\quad t*h:=(t_1h_1,t_2h_2).$$

<u>Указание</u>. Использовать неравенство (8) §3.

11. Доказать, что $\omega_1(f;\tau) \asymp \omega_1^{(1)}(f;\tau)+\omega_1^{(2)}(f;\tau)$.

12. Оценить константы в неравенствах теоремы 2 для $k=1,2;\ |\alpha|=1$.

§8. Задачи двумерной интерполяции. Интерполяция с помощью многочленов из $\mathcal{P}_n$

Основной вопрос, рассматриваемый во второй части пособия - приближение функции f двух переменных x и y с помощью алгебраических многочленов и кусочно - полиномиальных функций. Будем считать, что $f\in C(Q)$, $Q=[0,1]^2=\{(x,y):\ 0\le x,y\le 1\}$. Пусть X обозначает одно из пространств алгебраических многочленов $\mathcal{P}_n$ или $\mathcal{P}_{m,n}$ от двух переменных, введенных в §7; $p^*\in X$ - многочлен наилучшего приближения f с помощью X в метрике $C(Q)$, т. е. такой многочлен, для которого

$$e(f;X)_{C(Q)} := \inf_{p\in X} \|f-p\|_{C(Q)} = \|f-p^*\|_{C(Q)}.$$

Существование p^* для каждой $f\in C(Q)$ следует из общих результатов теории приближения (см., например, [2], стр. 8). Однако для функций двух переменных эффективные методы поиска p^* (типа одномерного алгоритма Ремеза) неизвестны. В связи с этим вместо p^* на практике часто используют многочлен $p=Pf$, где $P: C(Q)\to X$ - некоторый линейный проектор на X, т. е. $Pg=g$ для $g\in X$. Так как

$$e(f;X)_{C(Q)} \le \|f-Pf\|_{C(Q)} \le (1+\|P\|)\,e(f;X)_{C(Q)} \quad (1)$$

(правое неравенство в (1) есть неравенство Лебега), то $\|f-Pf\|$ тем меньше отличается от $\|f-p^*\|=e(f;X)$, чем меньше $\|P\|$. Итак, как и в случае одного переменного, предпочтительно иметь дело с проектором наименьшей нормы. В § 8-11 мы рассмотрим задачу минимизации $\|P\|$ для простейших проекторов - интерполяционных.

Рассмотрим сначала задачу интерполяции f с помощью многочленов из $\mathcal{P}_n$, $n\in\mathbb{Z}_+$. Так как $\dim\mathcal{P}_n=\frac{1}{2}(n+1)(n+2)=N$ (почему?), то для однозначной разрешимости интерполяционной задачи требуется, как ми-

нимум, N условий в (простых) узлах. Поэтому общая постановка задачи следующая.

Пусть $A_1=(x_1,y_1), \dots, A_N=(x_N,y_N)$ - фиксированный набор (различных) точек квадрата Q. Для $f\in C(Q)$ требуется найти такой многочлен $p\in \mathcal{P}_n$, для которого

$$p(A_1)=f_1, \dots, p(A_N)=f_N, \tag{2}$$

$f_j := f(A_j)$. Равенство (2) эквивалентно системе N линейных уранений с N неизвестными - коэффициентами a_{ik} многочлена p:

$$\sum_{0\le i+k\le n} a_{ik}\, x_s^i\, y_s^k = f_s, \quad s=1,\dots,N, \tag{3}$$

или в матричном виде (мономы $x^i y^k$ упорядочиваются по убыванию общей степени)

$$\begin{pmatrix} x_1^n & x_1^{n-1}y_1 & \dots & y_1^n & \dots & x_1 & y_1 & 1 \\ x_2^n & x_2^{n-1}y_2 & \dots & y_2^n & \dots & x_2 & y_2 & 1 \\ \vdots & & & & & & & \\ x_N^n & x_N^{n-1}y_N & \dots & y_N^n & \dots & x_N & y_N & 1 \end{pmatrix} \begin{pmatrix} a_{n0} \\ a_{n-1,1} \\ \vdots \\ a_{0n} \\ \vdots \\ a_{10} \\ a_{01} \\ a_{00} \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix}. \tag{4}$$

Обозначим через A матрицу системы (4), а ее определитель через Δ. Условие $\Delta=0$ эквивалентно тому, что точки $A_1,\dots,A_N$ лежат на одной алгебраической линии порядка n; в этом случае задача некорректна. Поэтому при выборе узлов интерполяции нужно проследить, чтобы они не попали на одну линию порядка n. В случае $n=1, N=3$ (интерполяция с помощью $\mathcal{P}_1$) неравенство

$$\Delta = \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} \neq 0$$

означает, что A_1, A_2, A_3 не лежат на одной прямой. Заметим, что в этом случае $\Delta = 2S$, S - площадь треугольника с вершинами A_j.

Если $\Delta \neq 0$, то неизвестные коэффициенты a_{ik} единственным образом находятся из (4) с помощью метода Гаусса. Их можно также получить после обращения матрицы $\mathbb{A}$:

$$\begin{pmatrix} a_{n0} \\ a_{n-1,1} \\ \vdots \\ a_{10} \\ a_{01} \\ a_{00} \end{pmatrix} = \mathbb{A}^{-1} \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_N \end{pmatrix} \qquad (5)$$

Обозначим через p_i многочлен из $\mathcal{P}_n$, коэффициенты которого образуют i-й столбец матрицы $\mathbb{A}^{-1}$. Равенство (5) означает, что $p_i(x,y)$ является решением интерполяционной задачи, соответствующей случаю $f_i = 1$, $f_j = 0$, $j \neq i$: i-й столбец $\mathbb{A}^{-1}$ получается после ее умножения на вектор

$$(0 \ldots\ldots \overset{(i)}{1} \ldots\ldots 0)^T$$

Итак, многочлены $p_1, \ldots, p_N$ удовлетворяют условиям:

$$p_i(x,y) \in \mathcal{P}_n \quad ; \quad p_i(A_j) = \delta_{ij} = \begin{cases} 1, & j = i \\ 0, & j \neq i \end{cases} \qquad (6)$$

Поэтому интерполяционный многочлен p, удовлетворяющий (2), имеет вид

$$p(x,y) = \sum_{i=1}^{N} f_i \, p_i(x,y). \qquad (7)$$

Формула (6) является аналогом интерполяционной формулы Лагранжа.

Итак, интерполяционный проектор $P: C(Q) \to \mathcal{P}_n$, построенный по системе узлов $A_1, \dots, A_N$, имеет вид

$$Pf(x,y) = \sum_{i=1}^{N} f(A_i)\, p_i(x,y) \quad ; \tag{8}$$

многочлены p_i удовлетворяют условиям (6); их коэффициенты составляют i-й столбец матрицы A^{-1}.

Можно дать другое описание базисных многочленов $p_i(x,y)$. Обозначим через $\Delta_i(x,y)$ определитель порядка N, получающийся из Δ путем замены i-й строки (она зависит от x_i, y_i) на строку упорядоченных мономов $x^s y^t$, $0 \le s+t \le n$. Тогда для всех $i=1,\dots,N$

$$p_i(x,y) = \frac{1}{\Delta}\, \Delta_i(x,y). \tag{9}$$

Действительно, $p_i(x,y) \in \mathcal{P}_n$ (надо разложить Δ_i по i-й строке); $p_i(A_j) = \delta_{ij}$ (т.к. $\Delta_i(A_i) = \Delta$, а определитель $\Delta_i(A_j)$ содержит две одинаковые строки и равен 0). Поэтому

$$Pf(x,y) = \frac{1}{\Delta} \sum_{i=1}^{N} f(A_i)\, \Delta_i(x,y). \tag{10}$$

Из формулы (8) получаем, как и в одномерной ситуации, выражение для нормы P как оператора из $C(Q)$ в $C(Q)$:

$$\|P\|_{C(Q)\to C(Q)} = \max_{(x,y)\in Q} \sum_{i=1}^{N} |p_i(x,y)|, \tag{11}$$

или с учетом (9)

$$\|P\|_{C(Q)\to C(Q)} = \frac{1}{|\Delta|} \max_{(x,y)\in Q} \sum_{i=1}^{N} |\Delta_i(x,y)|. \tag{12}$$

Задачи и упражнения к §8

1. Показать, что $\dim \mathcal{P}_n = \frac{1}{2}(n+1)(n+2)$.

2. Получить неравенство (1).

3. Доказать справедливость формул (11)-(12) для нормы проектора P.

4. Составить программу, реализующую вычисление $\|P\|$ на ЭВМ.

§9. Линейная и квадратичная интерполяция на квадрате

В этом параграфе мы рассмотрим простые, но важные для приложений случаи линейной и квадратичной интерполяции на квадрате, т. е. интерполяции с помощью многочленов из $\mathcal{P}_1$ и $\mathcal{P}_2$. Основное внимание мы уделим примерам.

§9.1. Линейная интерполяция на квадрате

Задача интерполяции состоит в том, чтобы для данного набора точек A_1, A_2, A_3 и значений $f_1 = f(A_1)$, $f_2 = f(A_2)$, $f_3 = f(A_3)$ указать многочлен $p(x,y) = ax + by + c$ такой, что $p(A_i) = p(x_i, y_i) = f_i$. Мы считаем, что узлы A_1, A_2, A_3 не лежат на одной прямой, т. е. (см. §8)

$$\Delta = \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} \neq 0.$$

В этом случае явное решение задачи записывается в виде

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \mathbb{A}^{-1} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix}, \quad \mathbb{A} = \begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{pmatrix},$$

или в другой форме

для некоторых естественных расположений узлов интерполяции.

A_1	A_2	A_3	$p(x,y) = Pf(x,y)$	$\|P\|$
$(0,0)$	$(1,0)$	$(0,1)$	$(f_2-f_1)x+(f_3-f_1)y+f_1$	3
$(0,0)$	$(1,0)$	$(\frac{1}{2},\frac{1}{2})$	$(f_2-f_1)x+(2f_3-f_1-f_2)y+f_1$	3
$(0,0)$	$(1,0)$	$(\frac{1}{2},1)$	$(f_2-f_1)x+(f_3-\frac{1}{2}f_1-\frac{1}{2}f_2)y+f_1$	2
$(0,0)$	$(1,\frac{1}{2})$	$(\frac{1}{2},1)$	$\frac{4}{3}\left[\left(-\frac{1}{2}f_1+f_2-\frac{1}{2}f_3\right)x + \left(-\frac{1}{2}f_1-\frac{1}{2}f_2+f_3\right)y+\frac{3}{4}f_1\right]$	$\frac{7}{3}$
$(0,0)$	$(1,0)$	(α,β), $\beta\neq 0$	$(f_2-f_1)x+\frac{1}{\beta}(f_3-\alpha f_2-(1-\alpha)f_1)y+f_1$	$F(\alpha,\beta)$

Пятая строка таблицы обобщает три первых случая (две точки фиксированы в смежных вершинах, третья меняется по квадрату). Норма интерполяционного проектора в этом случае зависит от α,β и выглядит следующим образом:

$$F(\alpha,\beta)=\begin{cases}\dfrac{2+\beta-2\alpha}{\beta}, & (\alpha,\beta)\in \mathrm{I}\\[2ex] \dfrac{2\alpha+\beta}{\beta}, & (\alpha,\beta)\in \mathrm{II}\\[2ex] \dfrac{2-\beta}{\beta}, & (\alpha,\beta)\in \mathrm{III}\end{cases} \qquad (1)$$

$$p(x,y) = \frac{1}{\Delta}\left(f_1 \Delta_1(x,y) + f_2 \Delta_2(x,y) + f_3 \Delta_3(x,y)\right)$$

Многочлен $\Delta_i(x,y)$ получается из Δ заменой i -й строки на строку $(\ x \quad y \quad 1\ )$. Норма соответствующего проектора $P: C(Q) \to \mathcal{P}_1$ (считаем $Pf = p$) равна

$$\| P \| = \frac{1}{|\Delta|} \max_{0 \le x,y \le 1} \sum_1^3 |\Delta_i(x,y)| .$$

Норму P можно в каждом конкретном случае найти иначе - надо использовать то, что максимум и минимум линейной функции $p = Pf$ на квадрате Q достигается в одной из его вершин.

<u>Пример 1</u>. Пусть $A_1 = (0,0)$, $A_2 = (1,0)$, $A_3 = (0,1)$. Тогда

$$\mathbb{A} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad \mathbb{A}^{-1} = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},$$

и $p(x,y) = (-f_1+f_2)x + (-f_1+f_3)y + f_1$; здесь $p_1(x,y) = -x-y+1$, $p_2(x,y) = x$, $p_3(x,y) = y$ (коэффициенты этих многочленов стоят по столбцам $\mathbb{A}^{-1}$). В данном случае $\Delta = |\mathbb{A}| = 1$.

$$\sum_1^3 |p_i(x,y)| = \begin{cases} 1, & 0 \le x+y \le 1 \\ 2(x+y)-1, & 1 \le x+y \le 2 \end{cases}$$

Поэтому $\| P \| = \max\limits_{x,y} \sum\limits_i |p_i| = 3$. С другой стороны,

$$\| P \| = \max_{-1 \le f_i \le 1} \max_{(x,y) \in Q} |p(x,y)| = \max_{-1 \le f_i \le 1} \max \{ |p(0,0)|, |p(1,0)|,$$

$$|p(0,1)|, |p(1,1)| \} = \max_{-1 \le f_i \le 1} \max \{ |f_1|, |f_2|, |f_3|, |-f_1+f_2+f_3| \} = 3.$$

Приведем здесь таблицу, в которой представлены значения $\| P \|$

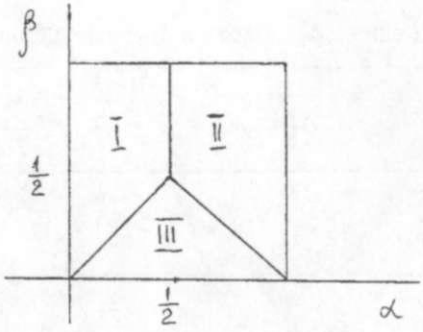

Рис. 7

Отметим, что расположения узлов в первом и третьем случаях дают максимальное значение площади треугольника $A_1A_2A_3$.

Итак, случай выбора $A_1=(0,0)$, $A_2=(1,0)$, $A_3=(1/2,1/2)$ дает $\|P\|=2$ и гарантирует оценку для точности приближения

$$\|f-Pf\|_{C(Q)} \leq 3\,e(f;\mathcal{P}_1)_{C(Q)},$$

вполне пригодную для практических вычислений.

Отметим, что за счет выбора узлов интерполяции последнюю оценку можно несколько улучшить. Пусть $A_1=(0,0)$, $A_2=(t,1)$, $A_3=(1,t)$, $t\neq 1$. Значения $t=0$ и $t=\frac{1}{2}$ дают $\|P\|=3$ и $\|P\|=\frac{7}{3}=2.33\ldots$ соответственно. Оказывается (весьма неожиданно), что для некоторого промежуточного значения $t^*\in[0,\frac{1}{2}]$ $\|P\|<2$! В общем случае

$$\|P\| = \varphi(t) = \max\left(\frac{3-t}{1+t}\,,\ \frac{1+2t-t^2}{1-t^2}\right). \qquad (2)$$

Формула (2) получается по общей схеме. Простой анализ показывает, что функция $\varphi_1(t)=\frac{3-t}{1+t}$ убывает, а функция $\varphi_2(t)=\frac{1+2t-t^2}{1-t^2}$ возрастает; следовательно, $\min_t \varphi(t)$ достигается в точке $t^*\in[0,1]$, которая является решением уравнения

$$\frac{3-t}{1+t} = \frac{1+2t-t^2}{1-t^2}.$$

или $t^2-3t+1=0$. Это дает значение $t^*=\frac{3-\sqrt{5}}{2}=0.38196\ldots$
Так как

$$\varphi(t^*)=\frac{3-t^*}{1+t^*}=\frac{3+\sqrt{5}}{5-\sqrt{5}}=\frac{(3+\sqrt{5})(5+\sqrt{5})}{20}=1+\sqrt{0.8},$$

то минимальное значение $\|P\|=\varphi(t^*)=1+\sqrt{0.8}=1.8944271\ldots$

Интересно заметить, что число t^* является "магическим" числом "золотого сечения": действительно, уравнение

$$\frac{t^*}{1-t^*}=\frac{1-t^*}{1} \qquad (3)$$

эквивалентно $t^{*2}-3t^*+1=0$. Равенство (3) как раз и содержит определение "золотого сечения": отношение меньшей части к большей равно отношению большей части к целому. Геометрически (3) означает, что при соответствующем выборе точек A_1, A_2, A_3 площади трех треугольников, примыкающих к вершинам квадрата, равны.

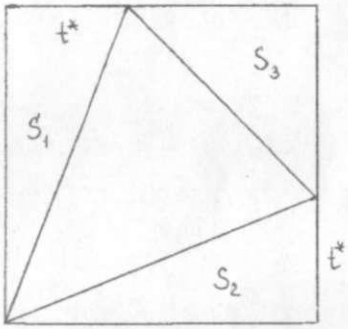

Рис. 8. $S_1=S_2=S_3$.

§9.2. Квадратичная интерполяция на квадрате

Задача интерполяции $f\mapsto p=Pf$ состоит в том, чтобы для фиксированного набора точек $A_1,\ldots,A_6\in Q=[0,1]^2$, $A_i=(x_i,y_i)$, не лежащих на линии второго порядка, и значений $f_i=f(A_i)$ указать многочлен $p(x,y)=ax^2+bxy+cy^2+dx+ey+h$ такой, что

$p(A_i)=f_i$, $i=1,\dots,6$. Условие, накладываемое на точки A_i , является весьма ограничительным; в частности не подходят многие симметричные расположения узлов, например такие:

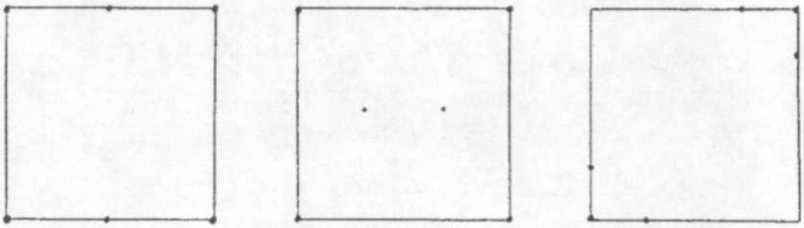

Рис. 9

Если A_i не лежат ни на какой линии второго порядка, то 6x6 - матрица

$$\mathbb{A} = \begin{pmatrix} x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_6^2 & x_6y_6 & y_6^2 & x_6 & y_6 & 1 \end{pmatrix}$$

является невырожденной, и решение задачи дается равенством

$$\begin{pmatrix} a \\ \vdots \\ h \end{pmatrix} = \mathbb{A}^{-1} \begin{pmatrix} f_1 \\ \vdots \\ f_6 \end{pmatrix},$$

или

$$p(x,y) = \sum_1^6 f_i p_i(x,y),$$

где p_i - многочлены, коэффициенты которых составляют столбцы матрицы $\mathbb{A}^{-1}$

Пример 2. Пусть $A_1=(0,0)$, $A_2=(1,0)$, $A_3=(0,1)$, $A_4=(1,1)$, $A_5=(\frac{1}{2},\frac{1}{2})$, $A_6=(\frac{1}{2},0)$. Точки $A_1,\dots,A_5$ расположены симметрично, выбор A_6 связан с указанным ограничением ($A_1,\dots,A_5$ лежат на паре прямых).

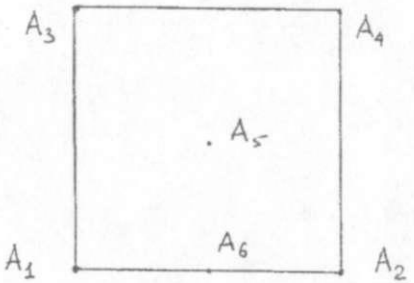

Рис. 10

В этом случае

$$\mathbb{A}=\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} & \frac{1}{2} & 1 \\ \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 & 1 \end{pmatrix}, \quad \mathbb{A}^{-1}=\begin{pmatrix} 2 & 2 & 0 & 0 & 0 & -4 \\ 1 & -1 & -1 & 1 & 0 & 0 \\ -1 & -1 & 1 & 1 & -4 & 4 \\ -3 & -1 & 0 & 0 & 0 & 4 \\ 0 & 1 & 0 & -1 & 4 & -4 \\ 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

и

$$p(x,y)=Pf(x,y)=(2f_1+2f_2-4f_6)x^2+(f_1-f_2-f_3+f_4)xy+$$
$$+(-f_1-f_2+f_3+f_4-4f_5+4f_6)y^2+(-3f_1-f_2+4f_6)x+$$
$$+(f_2-f_4+4f_5-4f_6)y+f_1.$$

Многочлен $p(x,y)$ можно записать также в виде $p(x,y)=\sum_1^6 f_i p_i(x,y)$, где

$$p_1(x,y)=2x^2+xy-y^2-3x+1\ ;\quad p_2(x,y)=2x^2-xy-y^2-x+y\ ;$$
$$p_3(x,y)=-xy+y^2\ ;\quad p_4(x,y)=xy+y^2-y\ ;$$

$$p_5(x,y) = -4y^2 + 4y \;;\quad p_6(x,y) = -4x^2 + 4y^2 + 4x - 4y .$$

Нетрудно показать, что для всех $i = 1, \dots, 6$ и всех $(x,y) \in Q$

$$|p_i(x,y)| \le 1 .$$

Поэтому для проектора P по узлам $A_1, \dots, A_6$

$$\|P\| = \max_{(x,y)\in Q} \sum_1^6 |p_i(x,y)| \le 6 . \qquad (4)$$

Оценку (4) можно улучшить, например, с помощью ЭВМ. Во всяком случае, для точности приближения гарантировано неравенство

$$\|f - Pf\|_{C(Q)} \le 7\, e(f; \mathcal{P}_2)_{C(Q)} . \qquad (5)$$

Задачи и упражнения к §9

1. Почему функция $p \in \mathcal{P}_1$ принимает максимальное (минимальное) значение в одной из вершин квадрата?

2. Показать, что максимальное значение площади треугольника, содержащегося в квадрате $[0,1]^2$, равно $\frac{1}{2}$.

3. Получить результаты, содержащиеся в 2-4 строках табл. 1.

4. Получить для функции $F(\alpha,\beta)$ формулу (1).

5. Доказать, что для нормы интерполяционного проектора P, построенного по точкам $A_1 = (0,0)$, $A_2 = (t,1)$, $A_3 = (1,t)$, справедлива формула (2).

6. Рассмотрим известную последовательность Фибоначчи:

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \dots$$

Доказать, что для нее $\lim\limits_{n\to\infty} a_{n-1} / a_n = 1 - t^* = 0.6180\dots$

7. Чему равна $\|P\|_{C(\Pi)\to C(\Pi)}$ при интерполяции по трем вершинам прямоугольника $\Pi=[0,10]\times[0,1]$? Любого прямоугольника $\Pi=[a,b]\times[c,d]$?

8. Аккуратно обосновать непригодность выбора точек рис. 9 в качестве узлов для квадратичной интерполяции.

9. Улучшить оценки (4), (5) с помощью ЭВМ.

10. Рассмотреть вопрос об эффективной линейной интерполяции функций трех переменных на единичном кубе $[0,1]^3$. Определить норму проектора по четырем вершинам куба, каждые две из которых принадлежат общей грани.

§10. Оценки для нормы интерполяционного проектора. Способы выбора узлов интерполяции на подмножестве квадрата

Пусть G - замкнутое подмножество квадрата $Q=[0,1]^2$, однозначно определяющее любой многочлен из $\mathcal{P}_n$ (если $p(x,y)=q(x,y)$ для $p,q\in\mathcal{P}_n$ и всех $(x,y)\in G$, то $p=q$ всюду). В простейшей ситуации G - весь квадрат Q. Пусть, как и ранее, $N=\dim\mathcal{P}_n=\frac{1}{2}(n+1)(n+2)$. Если точки $A_i=(x_i,y_i)\in G$, $i=1,\dots,N$, таковы, что определитель

$$\Delta=\begin{vmatrix} x_1^n & x_1^{n-1}y_1 & \dots & y_1^n & \dots & x_1 & y_1 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_N^n & x_N^{n-1}y_N & \dots & y_N^n & \dots & x_N & y_N & 1 \end{vmatrix}$$

отличен от 0, то формула

$$p(x,y)=Pf(x,y)=\sum_{i=1}^{N} f_i p_i(x,y)=\frac{1}{\Delta}\sum_{i=1}^{N} f_i\,\Delta_i(x,y), \qquad (1)$$

$f_i=f(A_i)$, задает интерполяционный проектор $P: C(G)\to\mathcal{P}_n$.

такой что $p(A_i)=f_i$. Здесь Δ_i - определитель, получающийся из Δ заменой i - й строки на строку ($x^n\ x^{n-1}y\ \dots\ x\ \ y\ \ 1$). Норма P как оператора из $C(G)$ в $C(G)$ равна

$$\|P\| = \frac{1}{|\Delta|} \max_{(x,y)\in G} \sum_{i=1}^{N} |\Delta_i(x,y)| . \qquad (2)$$

Доказательства формул (1) и (2) аналогичны случаю $G=Q$ (см. §8) и предоставляются читателю.

Выберем точки $A_1,\dots,A_N$ специальным образом, а именно так, чтобы величина $|\Delta|$ имела максимальное возможное значение при всех вариантах $A_j\in G$. Тогда, очевидно, для всех $i=1,\dots,N$ и $(x,y)\in G$ будет выполняться неравенство

$$|\Delta_i(x,y)| \le |\Delta| .$$

В связи с этим при таком выборе A_j

$$\|P\|_{C(G)\to C(G)} \le N. \qquad (3)$$

Итак, мы обосновали такой способ выбора узлов интерполяции.

<u>Первый способ</u>. Узлы интерполяции $A_1=(x_1,y_1),\dots,A_N=(x_N,y_N)$ выбираются как решение задачи максимизации

$$|\Delta| \to \max \quad \text{при условии} \quad (x_i,y_i)\in G . \qquad (4)$$

Точное и эффективное решение задачи (4) для произвольного G указать трудно. В реальной ситуации G - конечное (многоточечное) множество; если G состоит из M точек, то полный перебор требует вычисления C_M^N определителей Δ для всевозможных наборов из N точек. В ряде случаев (при небольших n ; для симметричных или специальных G) решение упрощается. Можно также потребовать, чтобы величина $|\Delta|$ была "почти максимальна": если $|\Delta_i(x,y)|\le\gamma|\Delta|$, $(x,y)\in G$, $i=1,\dots,N$, с некоторой $\gamma\ge 1$, то это дает оценку нормы соответствующего проектора в виде $\|P\|\le\gamma N$. Отметим также, что хорошие результаты дают некоторые эвристические алгоритмы решения задачи (4).

Приведем другой метод выбора точек A_j. Важным достоинством этого метода является то, что набор точек $A_1,\dots,A_N$ строится последовательно. Следует, однако, отметить, что известная теоретическая оценка $\|P\|$ в этом случае по порядку n хуже, чем (3). Этот способ описан А.Йонссоном (A.Jonsson, 1986).

Второй способ. Обозначим через T взаимно-однозначное отображение множества G на подмножество $\mathbb{R}^N$, осуществляемое по правилу

$$(x,y) \xmapsto{T} X := (x^n, x^{n-1}y, \dots, y^n, \dots, x, y, 1) \in \mathbb{R}^N$$

Пусть $T(G)$ - образ G. Точки $X_1,\dots,X_N$ (и их прообразы (x_i, y_i)) ищутся так. Первые две точки X_1, X_2 выбираются как решение экстремальной задачи

$$|X - Y|_{\mathbb{R}^N} \to max \quad , \quad X, Y \in T(G).$$

Здесь $|X-Y|_{\mathbb{R}^N}$ - евклидово расстояние в $\mathbb{R}^N$. Таким образом, X_1, X_2 - две максимально удаленные одна от другой точки из $T(G)$. На этом этапе в дискретном случае надо перебрать все пары точек; впрочем, решение может быть очевидным.

Далее, X_3 выбираем как наиболее удаленную точку прямой, проходящей через X_1, X_2. Точка X_4 должна быть максимально удалена от плоскости $X_1X_2X_3$, и т. д.. Итак, $X_k \in T(G)$ - точка, наиболее удаленная от $(k-2)$-мерного линейного многообразия, проходящего через $X_1,\dots,X_{k-1}$.

Расстояние d от точки X до линейного многообразия $M = L + X^*$, где L - подпространство с базисом $V_1,\dots,V_s$, равно

$$d = \left(\frac{Gr(V_1, V_2, \dots, V_s, X - X^*)}{Gr(V_1, V_2, \dots, V_s)} \right)^{\frac{1}{2}} ;$$

$Gr(\cdot)$ обозначает определитель Грама системы векторов $(\cdot)$. В связи с этим (т.к. знаменатель фиксирован) на k - м шаге мы рассматриваем задачу

$$Gr(X_2 - X_1, \dots, X_{k-1} - X_1 ; X - X_1) \to max ,$$
$$X \in T(G),$$

решение которой дает X_k. Если G состоит из M точек, то эта схема требует $\leq MN$ вычислений определителей Грама нарастающего порядка. При вычислении симметричной матрицы Грама следует учесть ее преемственность при переходе к следующему порядку (окаймление строчкой и столбцом).

В заключение отметим, что интерполяционный многочлен Pf, построенный по точкам $(x_i, y_i) \in G$, может служить распространением (продолжением) f на весь квадрат.

Задачи и упражнения к §10

1. К каким результатам приводят действия описанных методов для $N = 3$ (т.е. $n = 1$), если G -

а) весь квадрат;

в) вписанный круг;

с) вписанный треугольник ?

Разобрать другие простые случаи.

2. Придумать эвристический алгоритм решения задачи (4).

3. Реализовать выбор узлов по способам 1 и 2 на ЭВМ.

§11. Интерполяция функций двух переменных с помощью многочленов из $\mathcal{P}_{m,n}$

В §8-10 мы рассмотрели вопросы, связанные с интерполяцией функций двух переменных с помощью многочленов из $\mathcal{P}_n$, т. е. многочленов от двух переменных общей степени не выше n. Другой естественный класс интерполирующих многочленов - совокупность $\mathcal{P}_{m,n}$ многочленов от x и y степени не выше m по x и не выше n по y; про такие многочлены говорят иногда, что они имеют "распадающуюся" или "векторную" степень. Мы покажем, что при определенном подходе интерполяция многочленами векторной степени базируется на одномерной интерполяции. Фактически, описываемый метод интерполяции получается комбинацией интерполяции по двум направлениям (по x и по y).

По определению, $p \in \mathcal{P}_{m,n}$, если $p(x,y)$ имеет вид

$$p(x,y)=\sum_{\substack{0\le i\le m\\ 0\le j\le n}} a_{ij}\, x^i y^j \quad , \quad a_{ij}\in R .$$

Ясно, что $D=\dim \mathcal{P}_{m,n}=(m+1)(n+1)$. Минимальное число узлов интерполяции, гарантирующее единственность решения задачи, равно D . Простой и естественный вариант выбора D узлов на квадрате $Q=[0,1]^2$ получается так. Пусть $x_0,\dots,x_m$; $y_0,\dots,y_n$ - два набора (различных) точек отрезка $[0,1]$. Положим

$$A_{ij}:=(x_i,y_j) \quad ; \; i=0,\dots,m \; ; \; j=0,\dots,n .$$

(A_{ij} - узлы сетки, получающейся при разбиении координатных сторон квадрата). Пусть $f\in C(Q)$, $f_{ij}=f(A_{ij})=f(x_i,y_j)$. Задача интерполяции заключается в нахождении такого многочлена $p\in\mathcal{P}_{m,n}$. для которого

$$p(x_i,y_j)=f_{ij} \quad ; \; i=0,\dots,m \; ; \; j=0,\dots,n . \tag{1}$$

В случае $m=n=1$ многочлен $p(x,y)=axy+bx+cy+d$ является линейным по x при фиксированном y , и наоборот. Таким образом, интерполяцию с помощью $\mathcal{P}_{1,1}$ можно назвать **билинейной** (для нее $D=4$). Интерполяция с помощью $\mathcal{P}_{2,2}$ называется **биквадратной** (здесь $D=9$). Решение задачи (1) дается следующим аналогом формулы Лагранжа. Положим

$$L_i(x):=\prod_{k\ne i}\frac{x-x_k}{x_i-x_k} \quad , \; i=0,\dots,m ;$$

$$M_j(y):=\prod_{k\ne j}\frac{y-y_k}{y_j-y_k} \quad , \; j=0,\dots,n .$$

Тогда многочлен

$$p(x,y)=\sum_{i=0}^{m}\sum_{j=0}^{n} f_{ij}\, L_i(x)\, M_j(y) \tag{2}$$

принадлежит $\mathcal{P}_{m,n}$; кроме того,

$$p(x_s, y_t) = \sum_{i=0}^{m} \sum_{j=0}^{n} f_{ij} L_i(x_s) M_j(y_t) = f_{st} . \tag{3}$$

Следовательно, формула (2) дает решение задачи (1). Другая запись интерполяционного многочлена

$$Pf(x,y) := \sum_{i=0}^{m} \sum_{j=0}^{n} f(x_i, y_j) L_i(x) M_j(y) \tag{4}$$

определяет проектор $P: C(Q) \to \mathcal{P}_{m,n}$. Нетрудно показать, что оператор P представляется в виде суперпозиции операторов $P^{(1)}$ и $P^{(2)}$, которые здесь обозначают одномерные интерполяционные проекторы по системам точек $x_0, \dots, x_m$ (направление оси x) и $y_0, \dots, y_n$ (направление оси y), т. е.

$$P = P^{(1)} \circ P^{(2)} . \tag{5}$$

Запись $Pf(x,y) = P^{(1)} P^{(2)} f(x,y)$ означает, что сначала (при фиксированном x) f интерполируется по y с помощью $P^{(2)}$; затем результат интерполируется по x с помощью $P^{(1)}$.

Так же, как в одномерном случае, получается равенство

$$\| P \|_{C(Q) \to C(Q)} = \max_{(x,y) \in Q} \sum_{i,j} | L_i(x) M_j(y) | . \tag{6}$$

Если преобразовать правую часть (6), то окажется, что

$$\| P \|_{C(Q) \to C(Q)} = \| P^{(1)} \| \cdot \| P^{(2)} \| ; \tag{7}$$

здесь $\| P^{(i)} \|$ - норма одномерного интерполяционного проектора $P^{(i)}$ по соответствующей системе точек как оператора из $C(0,1)$ в $C(0,1)$. Таким образом, оптимальность двумерной интерполяции обусловливается оптимальностью выбора узлов по координатным направлениям. В частности, например, интерполяционный проектор $P: C(Q) \to \mathcal{P}_{1,1}$ по узлам, совпадающим с вершинами квадрата, является наилучшим - для него $\| P \| = \| P^{(i)} \| = 1$, $i = 1, 2$ (см. упр. 2-3 после §6). Проектор

$P: C(Q) \to \mathcal{P}_{2,2}$, построенный по симметричному набору узлов: вершины - середины сторон - центр квадрата, имеет норму $\|P\| = (1.25)^2 = 1.5625$.

Задачи и упражнения к §11

1. Почему $p(x_s, y_t) = f_{st}$, см. равенство (3)?
2. Почему имеет место представление (5)?
3. Показать, используя (5), что P является проектором на $\mathcal{P}_{m,n}$.
4. Доказать равенство (6) для нормы проектора. Проверить, что $\|P\| = \|P^{(1)}\| \cdot \|P^{(2)}\|$.
5. Определить явный вид билинейного и биквадратного проекторов по точкам равномерного разбиения сторон квадрата.
6. Показать, не используя равенства (7), что интерполяционный проектор на $\mathcal{P}_{1,1}$ по вершинам квадрата является наилучшим.
7. Сравнить экспериментальным путем точности приближения функций с помощью интерполяционных проекторов на $\mathcal{P}_{2,2}$ по равномерным узлам и узлам Чебышева.

§12. Приближение кусочно-полиномиальными функциями

Пусть функция $f(x,y)$ задана на единичном квадрате $Q_0 = [0,1]^2$. По аналогии с одномерным случаем обозначим через π разбиение этого квадрата на прямоугольники, стороны которых параллельны сторонам квадрата. Кроме того, если все прямоугольники равны между собой и длины их сторон в направлениях осей равны $\frac{1}{k}$ и $\frac{1}{\ell}$, то такое разбиение будем обозначать $\pi_{k\ell}$. Наконец, в случае $k = \ell$ (разбиение на k^2 равных квадратов) будем просто писать π_k. В дальнейшем, говоря о прямоугольниках из разбиения π, мы будем исключать их верхние и правые границы, если они не совпадают с границей единичного квадрата. Таким образом, прямоугольники из разбиения π не пересекаются и их объединение дает единичный квадрат.

Введем теперь множество $\mathcal{P}_n(\pi)$ (соответственно $\mathcal{P}_{m,n}(\pi)$) кусочно-полиномиальных функций степени $\leq n$ (соответственно $\leq (m,n)$), подчиненных разбиению π. Каждая функция из этого множества на

прямоугольнике из разбиения π совпадает с некоторым многочленом p из пространства $\mathcal{P}_n$ (или $\mathcal{P}_{m,n}$). Очевидно, что $\mathcal{P}_n(\pi)$ и $\mathcal{P}_{m,n}(\pi)$ являются линейными пространствами.

Рассмотрим вопрос о нахождении кусочно-полиномиальной функции g^* наилучшего приближения. Такая функция определяется соотношением

$$\|f - g^*\|_{C(Q_0)} = \inf_{g \in \mathcal{P}_{m,n}(\pi)} \|f - g\|_{C(Q_0)}.$$

Теоретически вопрос решается очень просто. Нужно на каждом прямоугольнике из разбиения π построить для функции f многочлен наилучшего приближения из пространства $\mathcal{P}_{m,n}$. Однако практически построить такой многочлен часто бывает невозможно. Из этого затруднительного положения можно выйти, если использовать интерполяционную технику, рассмотренную в § 8-11 части 2, т. е. на каждом прямоугольнике из разбиения π построить интерполяционный многочлен. Таким образом, если разбиение π заранее задано, то нетрудно построить кусочно-полиномиальную функцию, которая дает приближение почти как наилучшая.

Перейдем к более сложной ситуации, когда разбиение заранее не задано; при этом нужно найти кусочно-полиномиальную функцию, приближающую заданную функцию f с точностью ε . Разбиение, которому будет подчинена кусочно-полиномиальная функция, зависит и от функции f , и от точности ε . Такой процесс приближения называется **адаптивным**, так как разбиение подстраивается под характер особенностей функции. В частности, если у функции нет особенностей, то разбиение будет состоять из почти равных прямоугольников. Ниже мы опишем три таких адаптивных алгоритма.

Уточним постановку задачи. На квадрате $Q_0=[0,1]^2$ задана непрерывная функция f ; кроме того, заданы числа $m,n \in \mathbb{Z}_+$ и $\varepsilon > 0$. Нужно построить такое разбиение единичного квадрата $\pi=\pi(f,\varepsilon)$, что

$$\inf_{g \in \mathcal{P}_{m,n}(\pi)} \|f - g\|_{C(Q_0)} < \varepsilon. \qquad (1)$$

<u>Первый алгоритм</u>. На квадрате Q_0 построим интерполяционный многочлен p_0 из пространства $\mathcal{P}_{m,n}$ и найдем погрешность приближения. Если погрешность меньше ε , то p_0 - искомая функция. В этом

случае разбиение π состоит из единственного квадрата Q_0. Пусть погрешность приближения (т.е. $\|f-p_0\|_{C(Q_0)}$) больше ε. Разделим квадрат Q_0 на четыре равных квадрата $Q_{00}=[0,\frac{1}{2}]^2$, $Q_{10}=[\frac{1}{2},1]\times[0,\frac{1}{2}]$, $Q_{01}=[0,\frac{1}{2}]\times[\frac{1}{2},1]$, $Q_{11}=[\frac{1}{2},1]^2$.
На каждом из этих квадратов построим интерполяционный многочлен p_{ij} ($i,j=0,1$). Вычислим погрешности

$$\varepsilon_{ij} = \|f - p_{ij}\|_{C(Q_{ij})}, \quad i,j=0,1.$$

Возможны две ситуации: $\varepsilon_{ij}<\varepsilon$ и $\varepsilon_{ij}\geqslant\varepsilon$. В первом случае соответствующий квадрат Q_{ij} не подвергается дальнейшему делению, а входит в искомое разбиение. Во втором случае квадрат Q_{ij} делится на четыре равных квадрата, и весь описанный процесс повторяется. Одна из возможных ситуаций изображена на рис. 11.

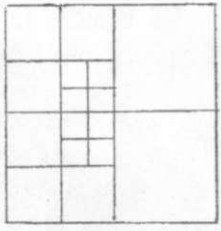

Рис. 11

Измельчение квадратов продолжается до тех пор, пока погрешность приближения не становится меньше ε. Искомое разбиение состоит из так называемых **диадических квадратов** разных размеров. Квадрат Q будем называть диадическим, если найдутся целые числа i,j,k, такие что $Q=[i2^{-k},(i+1)2^{-k}]\times[j2^{-k},(j+1)2^{-k}]$.

<u>Второй алгоритм</u>. Первые два шага такие же, как и в первом алгоритме, а именно: строим многочлен $p_0\in\mathcal{P}_{m,n}$ и, если погрешность приближения больше $\frac{\varepsilon}{2}$, то строим многочлены $p_{ij}^{(1)}\in\mathcal{P}_{m,n}$ $(i,j=0,1)$. Кусочно-полиномиальную функцию, состоящую из $p_{ij}^{(1)}$, обозначим g_1. Очевидно, что $g_1\in\mathcal{P}_{m,n}(\pi_2)$. Вычислим погрешность приближения

$$\varepsilon_1 = \|f-g_1\|_{C(Q_0)} = \max_{i,j=0,1} \|f-p_{ij}^{(1)}\|_{C(Q_{ij})}.$$

Если $\varepsilon_1 \geqslant \frac{\varepsilon}{2}$, то продолжим процесс деления: все квадраты Q_{ij} ($i,j=0,1$) разделим на четыре равные квадрата и на каждом из 16 более мелких квадратов построим многочлены $p_{ij}^{(2)}$ ($i,j=0,1,2,3$). Кусочно-полиномиальную функцию, состоящую из $p_{ij}^{(2)}$, обозначим g_2. Нетрудно видеть, что $g_2 \in \mathcal{P}_{m,n}(\pi_4)$. Вычислим ε_2. Если $\varepsilon_2 \geqslant \frac{\varepsilon}{2}$, то построим функцию g_3. Продолжим процесс деления до тех пор, пока не найдем кусочно-полиномиальной функции $g_N \in \mathcal{P}_{m,n}(\pi_{2^N})$ такой, что

$$\| f - g_N \|_{C(Q_0)} < \frac{\varepsilon}{2}$$

Мы могли бы сказать, что функция g_N является искомой (см. постановку задачи (1)). Однако этот способ построения хуже рассмотренного нами в первом алгоритме. Мы построили равномерное разбиение π_{2^N}, которое состоит из большего количества квадратов, чем в первом алгоритме. Действительно, на этом пути мы делили все квадраты, даже те, на которых точность приближения была достигнута. Поэтому построение функции g_N можно считать промежуточным шагом. Мы перешли от функции f к функции g_N. Итак, нужно уточнить постановку задачи. Нам не только нужно построить кусочно-полиномиальную функцию, которая удовлетворяет (1), но и по возможности минимизировать количество квадратов, составляющих π.

Представим функцию g_N в виде ряда

$$g_N(x) = \sum_{i=0}^{N} (g_i(x) - g_{i-1}(x));$$

здесь $g_0(x) = p_0(x)$, $g_{-1}(x) \equiv 0$. Учитывая специфику разбиения π_{2^i}, заметим, что разбиение π_{2^i} "вписано" в разбиение $\pi_{2^{i-1}}$, а это значит (см. задачу 3 §12), что $\mathcal{P}_{m,n}(\pi_{2^{i-1}}) \subset \mathcal{P}_{m,n}(\pi_{2^i})$. Отсюда сразу следует, что функция $g_i - g_{i-1}$ является элементом пространства $\mathcal{P}_{m,n}(\pi_{2^i})$. Выберем в этом пространстве базис, состоящий из B-сплайнов, и разложим $g_i - g_{i-1}$ по базису (см. задачу 4 §12):

$$(g_i - g_{i-1})(x) = \sum_{j=1}^{K_i} \lambda_j^i B_j^i(x);$$

здесь $K_i := \dim \mathcal{P}_{m,n}(\pi_{2^i}) = 2^i(m+1)\cdot 2^i(n+1)$.
Имеем равенство

$$g_N(x) = \sum_{i=1}^{N} \sum_{j=1}^{K_i} \lambda_j^i B_j^i(x). \qquad (2)$$

Обозначим $T_i := g_i - g_{i-1}$. Построим две последовательности сплайнов S_i ($i=0,\dots,N$) и T_i' ($i=1,\dots,N$), используя последовательность T_i. Положим $S_0 = T_0$ и $T_i' = T_i + \dots + T_0 - S_{i-1} - \dots - S_0$. Пусть сплайны $S_0,\dots,S_{i-1}, T_i'$ уже построены. Покажем, как строить сплайн S_i. Для этого разложим T_i' по базису пространства $\mathcal{P}_{m,n}(\pi_{2^i})$. Пусть

$$T_i'(x) = \sum_{j=1}^{K_i} c_j^i B_j^i(x).$$

Выделим самые большие коэффициенты в этом разложении. Пусть

$$\mathcal{L}_i := \left\{ j=1,\dots,K_i : |c_j^i| > \frac{i\varepsilon}{2N} \right\},$$

тогда

$$S_i(x) := \sum_{j\in\mathcal{L}_i} c_j^i B_j^i(x).$$

Как доказано в [43],

$$\| g_N - s \|_{C(Q_0)} < \frac{\varepsilon}{2},$$

где $s = S_0 + \dots + S_N$. Таким образом, мы нашли функцию s такую, что

$$\| f - s \|_{C(Q_0)} < \varepsilon.$$

<u>Третий алгоритм</u> ввиду его сложности опишем только для случая приближения кусочно-постоянными функциями.

Начнем с построения функции $g_N \in \mathcal{P}_0(\pi_{2^N})$ такой, что

$$\| f - g_N \|_{C(Q_0)} < \frac{\varepsilon}{2}.$$

Для этого повторим то, что мы делали в предыдущем алгоритме. Имеем

представление

$$g_N = \sum_{i=0}^{N} (g_i - g_{i-1}).$$

Учтем, что базис пространства $\mathcal{P}_0(\pi_{2^i})$ образуют характеристические функции квадратов, составляющих π_{2^i}. Для $Q \in \pi$ эти функции определяются равенством

$$\chi_Q(x) = \begin{cases} 1, & x \in Q \\ 0, & x \notin Q \end{cases}.$$

Тогда разложение функции $g_i - g_{i-1}$ по базису имеет вид

$$(g_i - g_{i-1})(x) = \sum_{Q \in \pi_{2^i}} \lambda_Q \chi_Q(x).$$

Если обозначить множество всех диадических квадратов через $\mathcal{D} := \bigcup_{i=1}^{N} \pi_{2^i}$, то

$$g_N(x) = \sum_{Q \in \mathcal{D}} \lambda_Q \chi_Q(x).$$

Это равенство является частным случаем (2). Всюду ниже для функции φ, имеющей разложение по характеристическим функциям

$$\varphi(x) = \sum_{Q \in \mathcal{D}} c_Q \chi_Q(x),$$

и любого диадического квадрата I полагаем

$$\lambda(\varphi, I) := \Big(\sum_{Q \subseteq I} |c_Q|^3 \Big)^{1/3}.$$

Заметим, что если $I_1 \subset I_2$, то $\lambda(\varphi, I_1) \le \lambda(\varphi, I_2)$. Обозначим $\Lambda = \lambda(g_N, Q_0)$. Теперь мы индуктивно определим последовательность функций φ_j и последовательность диадических квадратов I_j, $j = 0, \dots, m+1$; $I_{m+1} := Q_0$. Пусть $I_0 = \varnothing$, $\varphi_0 = g_N$ и $\lambda_Q^0 = \lambda_Q$. Предположим, что φ_j уже определена:

$$\varphi_j(x) = \sum_{Q \in \mathcal{D}} \lambda_Q^j f_Q(x).$$

Выберем минимальный квадрат I_{j+1}, для которого $\lambda(\varphi_j, I_{j+1}) < \varepsilon'$ и $\lambda(\varphi_j, I'_{j+1}) > \varepsilon'$. Здесь через I' обозначен "предшественник" квадрата I, то есть тот диадический квадрат, при делении которого получился квадрат I. Кроме того, $\varepsilon' := \varepsilon(2\Lambda)^{-1}$. Если такого минимального квадрата не существует, то положим $I_{j+1} = Q_0$, и нужная последовательность квадратов построена. После того, как квадрат I_{j+1} найден, положим

$$\lambda_Q^{j+1} = \begin{cases} 0, & Q \subseteq I_{j+1} \\ \lambda_Q^j & \text{в противном случае}. \end{cases}$$

Таким образом,

$$\varphi_{j+1} := \sum_{Q \in \mathcal{D}} \lambda_Q^{j+1} f_Q.$$

Отметим важное свойство построенных квадратов I_j, $j = 0, \dots, m+1$: если $j < k$, то или I_j не пересекается с I_k, или $I_j \subset I_k$. Учитывая это, построим множества

$$A_j := I_j \setminus \Big(\bigcup_{k<j} I_k \Big), \quad j = 0, \dots, m+1.$$

Эти множества не пересекаются, и их объединение дает весь квадрат Q_0. Обозначим через a_j константу наилучшего приближения функции φ_j на I_{j+1}, то есть предположим, что

$$\| \varphi_j - a_j \|_{C(I_{j+1})} = \inf_{a \in \mathbb{R}} \| \varphi_j - a \|_{C(I_{j+1})}$$

Тогда искомая функция s задается равенством:

$$s(x) := \Lambda a_j, \quad \text{если} \quad x \in A_{j+1}, \ j = 0, \dots, m.$$

<u>Замечание</u>. Хотя множества A_{j+1} не являются квадратами, можно показать, что

$$s(x) = \sum_{j=0}^{m} b_j \chi_{I_{j+1}}(x).$$

Подробнее о коэффициентах b_j см. задачу 5.

Задачи и упражнения к §12

1. Доказать, что $\mathcal{P}_{m,n}(\pi_{k\ell})$ является линейным пространством. Найти его базис и размерность.

2. Показать, что базис $\{\varphi_{ij}\}$ пространства $\mathcal{P}_{m,n}(\pi_{k\ell})$ может быть получен из базисов одномерных "координатных" пространств $\mathcal{P}_m(\pi_k)$ и $\mathcal{P}_n(\pi_\ell)$.

3. Напомним, что разбиение π "вписано" в разбиение π', если для каждого квадрата $Q \in \pi$ существует квадрат $Q' \in \pi'$ такой, что $Q \subset Q'$. Доказать, что в этом случае $\mathcal{P}_{m,n}(\pi') \subset \mathcal{P}_{m,n}(\pi)$.

4. Для того, чтобы построить базис из B-сплайнов в пространстве $\mathcal{P}_{m,n}(\pi_{k\ell})$, достаточно построить базис в одномерном случае и воспользоваться результатом задачи 2. Кроме того, так как пространство $\mathcal{P}_m(\pi_k)$ (и $\mathcal{P}_n(\pi_\ell)$) состоит из кусочно-полиномиальных функций, то достаточно построить базис из многочленов на каждом из интервалов, составляющих π_k (соответственно π_ℓ). Таким образом, задача свелась к построению специального базиса в пространстве многочленов $\mathcal{P}_n$. Построим этот базис для $n = 0, 1, 2$. При $n = 0$ базис состоит из функции тождественно равной 1. Если $n = 1$, то на отрезке $[a, b]$ базисные функции имеют вид $b_1(x) = (x-a)/(b-a)$, $b_2(x) = (b-x)/(b-a)$. Для построения этих функций мы воспользовались рекуррентным соотношением для B - сплайнов ([8], см. также [5])

$$B_{ik}(x) = \frac{x - t_i}{t_{i+k} - t_i} B_{i,k-1}(x) + \frac{t_{i+k+1} - x}{t_{i+k+1} - t_{i+1}} B_{i+1,k-1}(x) \quad (3)$$

и

$$B_{i0}(x) = \begin{cases} 1, & t_i \le x < t_{i+1} \\ 0 & \text{в других случаях.} \end{cases}$$

Если в (3) взять $k = 1$, $t_i = a$, $t_{i+1} = t_{i+2} = b$, то получим функцию

b_1 , а если взять $t_i = t_{i+1} = a,\ t_{i+2} = b$, то получим функцию b_2.

Пусть теперь $n = 2$. Мы определим базисные функции b_1, b_2, b_3, если в формуле (3) возьмем $k = 2$ и $(t_i, t_{i+1}, t_{i+2}, t_{i+3})$ заменим на $(a, a, a, b), (a, a, b, b), (a, b, b, b)$. Графики этих функций представлены на рис. 12.

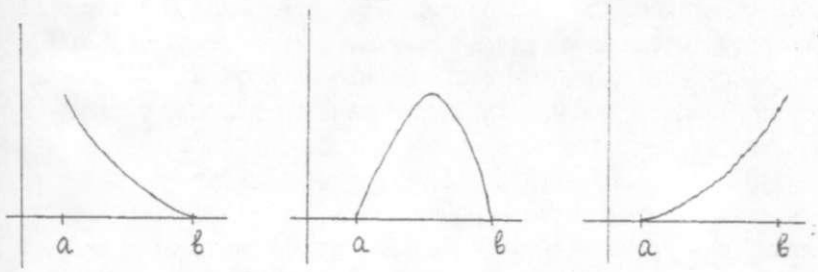

Рис. 12

Отметим, что функции b_1, b_2, b_3 образуют разбиение единицы, то есть $b_1(x) + b_2(x) + b_3(x) \equiv 1$.

Найти специальный базис пространств $\mathcal{P}_{2,1}(\pi_{2,2})$, $\mathcal{P}_{3,1}(\pi_{1,1})$, $\mathcal{P}_{3,3}(\pi_{2,1})$.

5. Доказать, что функция из третьего алгоритма допускает следующее представление:

$$s(x) = a_m \chi_{Q_0} + \sum_{j=0}^{m-1} (a_j - a_{r(j)}) \chi_{I_{j+1}} .$$

Здесь через $r(j)$ обозначено наименьшее целое число такое, что $r > j$ и I_{j+1} содержится в I_{r+1} .

6. Написать программы, реализующие описанные алгоритмы. Сравнить алгоритмы по трудоемкости, по точности приближения, по коэффициенту сжатия информации.